Legendre Transform and Applications to Finite and Infinite Optimization

Cristopher Hermosilla

LEGENDRE TRANSFORM AND APPLICATIONS TO FINITE AND INFINITE OPTIMIZATION

CRISTOPHER HERMOSILLA

Department of Mathematics, Louisiana State University, LA 70803–4918

ABSTRACT. We investigate convex constrained nonlinear optimization problems and optimal control with convex state constraints in the light of the so-called Legendre transform. We use this change of coordinate to propose a gradient-like algorithm for mathematical programs, which can be seen as a search method along geodesics. We also use the Legendre transform to study the value function of a state constrained Mayer problem and we show that it can be characterized as the unique viscosity solution of the Hamilton-Jacobi-Bellman equation.

1. INTRODUCTION

In this paper we report some applications to constrained optimization problems with a particular emphasis on convex constraints. Our study is based on the so-called *Legendre transform* induced by a penalization function $\mathbf{g} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ associated with a given convex set $\mathcal{K} \subseteq \mathbb{R}^n$.

At a first stage we are concerned with nonlinear (finite dimensional) optimization problems of the form $\inf\{f(x) \mid x \in \mathcal{K}\}$. The main contribution in this context is the study of an interior point method written as:

$$x_{k+1} = \nabla\mathbf{g}^*(\nabla\mathbf{g}(x_k) + t_k d_k), \qquad \forall k \in \mathbb{N}.$$

This algorithm can be interpreted as a geodesic search method provided that $\mathbf{int}\,(\mathcal{K})$ is endowed with a Riemannian metric induced by the squared Hessian of the penalization function $\mathbf{g}$. The analysis provided for the method is based on a dual lecture given to the original problem thanks to the Legendre transform.

To the best of our knowledge, the algorithm we propose has not been directly studied before. However, there are some works of similar nature when seeing it as a penalization method; see for instance [4, 29]. Nevertheless, our approach, specially the choice of the descent directions, seems to be new.

At a second stage we use the Legendre transform to study the value function of the Mayer problem with convex state constraints:

$$\inf\{f(p(T)) \mid \dot{p} \in F(p) \text{ a.e. on } [0,T], \ p(t) \in \mathcal{K} \text{ for each } t \in [0,T]\}.$$

In particular, we provide a characterization of the optimal cost map as the unique viscosity solution to the Hamilton-Jacobi-Bellman equation on $\mathbf{int}(\mathcal{K})$. It is well-known that the value function verifies this property, when no state constraints are considered; cf. [5]. However, in presence of these restrictions the situation is different and without additional compatibility assumptions, this might not be longer true. Regarding this issue, the most common hypotheses are the so-called *pointing conditions*, which are, roughly speaking, assumptions over the direction to where the dynamics points to; we refer for more details to [34, 20, 15, 19].

In this paper, we consider control systems for which the boundary of the state constraints is absorbing, and so the pointing conditions are not satisfied; precise statements can be found at Section 4. We also mention that there are other ways, different from the proposed here, to deal with the value function overcoming the usual compatibility assumptions; see for example [7, 21].

The underlying idea behind the applications to finite and infinite optimization is that $\mathbf{int}(\mathcal{K})$ can be endowed with the structure of Riemannian manifold by means 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 under suitable assumptions, it provides an isometry between $\mathbf{int}(\mathcal{K})$ and $\mathbb{R}^n$. This fact is essential for understanding the nature of the paper's exposition, because it allows us to transform constrained problems into unconstrained ones.

We finally mentioned that the interest in optimization problems on manifolds has been increasing in the last years; for instance [33, 1] in nonlinear optimization, [29, 35] in convex optimization and [13, 11] in optimal control. The present work contributes also in this direction by providing an explicit study for the manifolds that are convex open sets on the ambient space, which can be endowed with a Riemannian metric induced by the squared Hessian of $\mathbf{g}$; this is the subject of discussion in Section 5.

1.1. **Notation.** Throughout this paper $\mathbb{R}$ denotes the sets of real numbers, $|\cdot|$ and $\langle\cdot,\cdot\rangle$ are the Euclidean norm and inner product on $\mathbb{R}^n$, $\mathbb{B}$ stands for the unit open ball and $\mathbb{B}(x,r) = x + r\mathbb{B}$. For a given set S, $\mathbf{int}(S), \overline{S}$ and $\mathbf{bdry}(S)$ stand for the interior, closure, and boundary of S, respectively. The effective domain of a given function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is $\mathbf{dom}\,\varphi = \{x \in \mathbb{R}^n \mid \varphi(x) \in \mathbb{R}\}$. If $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued map, $\mathbf{dom}\,\Phi$ is $\{x \in \mathbb{R}^n \mid \Phi(x) \neq \emptyset\}$.

We denote by $\mathcal{C}^k(\mathcal{O})$ the set of k-times continuously differentiable functions on the open set $\mathcal{O} \subseteq \mathbb{R}^n$. When appropriate, $\nabla\varphi(x)$ and $\nabla^2\varphi(x)$ stand for the gradient and Hessian matrix of φ at x, respectively.

In all the paper, the abbreviations "l.s.c." and "u.s.c." stand respectively for "lower semicontinuous" and "upper semicontinuous".

2. Preliminaries.

2.1. **Some notions of Convex Analysis.** We begin with a short review of some concepts of Convex Analysis ([31, 10]). We denote by $\Gamma_0(\mathbb{R}^n)$ the class of all convex and l.s.c. functions defined on $\mathbb{R}^n$ with values in $\mathbb{R} \cup \{+\infty\}$ whose effective domain is nonempty. For any $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$, we write $\mathbf{g}^*$ for its *Legendre-Fenchel conjugate*

$$\mathbf{g}^*(y) = \sup\{\langle x, y \rangle - \mathbf{g}(x) \mid x \in \mathbf{dom}\ \mathbf{g}\}, \qquad \forall y \in \mathbb{R}^n,$$

and we denote by $\partial \mathbf{g}$ its *subdifferential*

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

Accordingly to [31, Section 26], $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ is called *essentially smooth* if

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

This property has a dual interpretation in terms of the Legendre-Fenchel conjugate. Indeed, $\mathbf{g}$ is essentially smooth if and only if $\mathbf{g}^*$ is strictly convex on every convex subset of $\mathbf{dom}\ \partial \mathbf{g}$; see [31, Theorem 26.3]. The last property is known as *essential strictly convexity*. In other words, the following holds true

$\mathbf{g}$ is essentially smooth if and only if $\mathbf{g}^*$ is essentially strictly convex.

This last relation motivates the next definition taken from [31, Section 26].

Definition 2.1. *A function* $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ *is called of Legendre type or simply Legendre function provided it is essentially smooth and essentially strictly convex.*

Notice that $\mathbf{g}$ is a Legendre function if and only if $\mathbf{g}^*$ it is as well. In this case, by [31, Theorem 26.5], $\nabla \mathbf{g} : \mathbf{int}\ (\mathbf{dom}\ \mathbf{g}) \to \mathbf{int}\ (\mathbf{dom}\ \mathbf{g}^*)$ is a homeomorphism known as the *Legendre transform*. Furthermore, we have

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

Remark 2.1. *A well-known example of Legendre function on* $\mathbb{R}$ *is the Boltzmann-Shannon entropy:*

$$\mathbf{g}_{ent}(x) := \begin{cases} x \log(x) - x & x \ge 0, \\ +\infty & x < 0. \end{cases}$$

In this case $\mathbf{g}_{ent}^*(y) = \exp(y)$ *for any* $y \in \mathbb{R}$. $\qquad \square$

2.2. **Legendre zone consistent pairs.** We now introduce a definition that plays a fundamental role throughout this manuscript. Indeed, this is the essential tool we use to provide all our results. From now on, the symbol $\mathbf{\Omega}$ is reserved to denote a given open nonempty convex set of $\mathbb{R}^n$ and $\mathbf{g}$ is a given Legendre function.

Definition 2.2. *For any $\mathbf{g} \in \Gamma_0\left(\mathbb{R}^n\right)$ of Legendre type and any nonempty open convex set $\mathbf{\Omega} \subseteq \mathbb{R}^n$, we say that the pair $(\mathbf{\Omega}, \mathbf{g})$ is Legendre zone consistent if $\boldsymbol{int}\left(\mathbf{dom}\ \mathbf{g}\right) = \mathbf{\Omega}$ and $\mathbf{dom}\ \mathbf{g}^* = \mathbb{R}^n$.*

The preceding definition essentially implies that the function $\mathbf{g}$ is a suitable penalization function for the set $\mathbf{\Omega}$. In a general situation, the open convex set is given and we look for a Legendre function such that the pair $(\mathbf{\Omega}, \mathbf{g})$ 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 constructive way to find one of such Legendre functions; see also [10, Chapter 7.4].

Before continuing with the exposition, we define a *Riemannian metric* on $\mathbf{\Omega}$ and introduce some notation, in order to simplify the forthcoming discussion.

Definition 2.3. *A function $\rho : \mathbf{\Omega} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called a $\mathcal{C}^k$-Riemannian metric on $\mathbf{\Omega}$ if the following condition are met:*

- *For any $x \in \mathbf{\Omega}$, $\rho_x(\cdot, \cdot) := \rho(x, \cdot, \cdot)$ is an inner product on $\mathbb{R}^n$.*
- *For any $\Psi_1, \Psi_2 \in \mathcal{C}^k(\mathbf{\Omega})$, the function $\rho_x(\Psi_1(x), \Psi_2(x))$ also belongs to $\mathcal{C}^k(\mathbf{\Omega})$.*

The idea of Legendre zone consistent pairs is also useful to establish an isometry between $\mathbf{\Omega}$ (endowed with an appropriate Riemannian metric) and the Euclidean space $\mathbb{R}^n$. A discussion about this interpretation has been postponed to Section 5.

We write, when appropriate,

$$(2) \qquad \nabla_{\mathbf{g}} f(x) = [\nabla^2 \mathbf{g}(x)]^{-1} \nabla f(x), \qquad \forall x \in \mathbf{\Omega},$$

$$(3) \qquad \mathrm{d}_{\mathbf{g}}(x, \tilde{x}) = |\nabla \mathbf{g}(x) - \nabla \mathbf{g}(\tilde{x})|, \qquad \forall x, \tilde{x} \in \mathbf{\Omega}.$$

On the one hand, (2) corresponds to the so-called *Riemannian gradient* of f when $\mathbf{\Omega}$ is endowed with a Riemannian metric induced by $\nabla^2 \mathbf{g}$; see [2] for details. On the other hand, as we will see in Section 5, (3) is the so-called *geodesic distance* on $\mathbf{\Omega}$, when the latter has an appropriate Riemannian structure; see Corollary 5.1.

2.3. **Dual unconstrained problem and first applications.** We finish the preliminary part by showing how to link an optimization problem with constraints represented by $\mathcal{K} = \overline{\mathbf{\Omega}}$ with an auxiliary unconstrained problem. This, as we have already mentioned, is the key feature of the exposition.

Let $\boldsymbol{\Omega}$ be given together with a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. We assume that there is $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ such that $(\boldsymbol{\Omega}, \mathbf{g})$ is Legendre zone consistent. We consider the problem

$$(P_o) \qquad\qquad\qquad \inf\{f(x) \mid x \in \boldsymbol{\Omega}\}.$$

For the purposes we will see shortly, we introduce an auxiliary cost $\varphi_{\mathbf{g}} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined via

$$\varphi_{\mathbf{g}}(y) = f \circ \nabla\mathbf{g}^*(y), \qquad \forall y \in \mathbb{R}^n.$$

The function $\varphi_{\mathbf{g}}$ is well-defined on $\mathbb{R}^n$ because $\mathbf{dom}\ \mathbf{g}^* = \mathbb{R}^n$, and by virtue of the Legendre transform $y = \nabla\mathbf{g}(x)$ we have

$$(4) \qquad\qquad \forall x \in \boldsymbol{\Omega},\ \exists y \in \mathbb{R}^n \text{ such that } f(x) = \varphi_{\mathbf{g}}(y),$$

$$(5) \qquad\qquad \forall y \in \mathbb{R}^n,\ \exists x \in \boldsymbol{\Omega} \text{ such that } \varphi_{\mathbf{g}}(y) = f(x).$$

Therefore, by proceeding in this way we can associate (P_o) with an auxiliary optimization problem where no constraints are involved:

$$(D_o) \qquad\qquad\qquad \inf\{\varphi_{\mathbf{g}}(y) \mid y \in \mathbb{R}^n\}.$$

The notation (D_o) has been used to emphasize the idea of duality between $\boldsymbol{\Omega}$ and $\mathbb{R}^n$ given by the Legendre transform. This makes us refer to (P_o) as the primal problem, to f as the primal function and to $\varphi_{\mathbf{g}}$ as the dual function.

Remark 2.2. *Several notions of duality exist in the literature of convex optimization and optimal control; see for instance [32]. Note that, we do not need the convexity of any of the objective functions (neither primal nor dual) to make sense to our duality scheme. Therefore, our notion differs from the classical conjugate duality and works for nonconvex problems as well. We stress that, the dual problem (D_o) is concerned with the minimization, of a given cost, over $y = \nabla g(x)$, instead of the primal variable x. In this context, $\nabla g(x)$ plays the role of dual variable.*

It is also worthy to notice that, our notion of duality strongly relies on the convexity of the constraints $\boldsymbol{\Omega}$ and the existence of a penalization function $\mathbf{g}$. We refer to [23, 38], for other suitable notions of duality, specially tailored for infinite dimensional nonconvex problems with nonconvex constraints. $\square$

The next result summarizes the relations between the problems (P_o) and (D_o). Furthermore, it also identifies a simple no-gap condition over the cost f.

Theorem 2.1. *Let $(\boldsymbol{\Omega}, \mathbf{g})$ be a Legendre zone consistent pair and $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be bounded from below and continuous on its domain. Suppose $\{x_k\} \subseteq \boldsymbol{\Omega}$, then it is a minimizing sequence for (P_o) if and only if $\{\nabla\mathbf{g}(x_k)\}$ is a minimizing sequence for (D_o). In any case $\mathrm{val}(P_o)=\mathrm{val}(D_o)$. Moreover, $x \in \boldsymbol{\Omega}$ is a local minimizer (maximizer) of f if and only if $\nabla\mathbf{g}(x)$ is a local minimizer (maximizer) of $\varphi_{\mathbf{g}}$.*

Proof. Note that by (4) and (5) the affirmation for the minimizing sequence holds true. Remark that $\mathrm{val}(P_o) \in \mathbb{R}$, so there exists a minimizing sequence $\{x_k\}$ such that $f(x_k) \to \mathrm{val}(P_o)$. By continuity of f on its domain we can find, for every $k \in \mathbb{N}$, a point $\tilde{x}_k \in \boldsymbol{\Omega}$ so that $f(\tilde{x}_k) \le f(x_k) + \frac{1}{k}$. Hence, noticing that $f(\tilde{x}_k) \to \mathrm{val}(P_o)$ and using that $\{\nabla \mathbf{g}(\tilde{x}_k)\}$ is minimizing sequence for (D_o) we get the result.

Finally, since $\nabla \mathbf{g}$ is a homeomorphism, it is an open mapping and so, from (4) and (5) the affirmation about the local extremals is straightforward. □

□

Example 2.1. *Consider the function*

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

Let $\boldsymbol{\Omega} = (0, +\infty) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $\mathbf{g}(x_1, x_2) = x_1 \log x_1 - x_1 + \frac{1}{2}\tan^2 x_2$. We see that $(\boldsymbol{\Omega}, \mathbf{g})$ is a Legendre zone consistent pair. Let us consider the problem

(P₁) $\min\left\{f(x_1, x_2) \mid 0 < x_1, -\pi \le 2x_2 < \pi\right\}.$

Hence, the dual problem is

(D₁) $\min\left\{y_1^2 + \exp y_2 \mid y_1, y_2 \in \mathbb{R}\right\}.$

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

□

On the other hand, one may expect that differentiability properties on the primal function are inherited by the dual function. This is indeed the case as we will see shortly. It is worthy to note that further regularity on $\mathbf{g}$ is required.

Proposition 2.1. *Let $(\boldsymbol{\Omega}, \mathbf{g})$ be a Legendre zone consistent pair. Suppose that f is continuously differentiable and $\mathbf{g}$ is twice continuously differentiable in a neighborhood of $x \in \boldsymbol{\Omega}$ with $\nabla^2 \mathbf{g}(x)$ being nonsingular. Then $\varphi_{\mathbf{g}}$ is continuously differentiable in a neighborhood of $y = \nabla \mathbf{g}(x)$ and moreover*

$$\nabla \varphi_{\mathbf{g}}(y) = \nabla_{\mathbf{g}} f(x).$$

Proof. It is a general fact that (cf. [16]), if $\mathbf{g} \in \Gamma_0\,(\mathbb{R}^n)$ is twice continuously differentiable in a neighborhood of $x \in \mathbb{R}^n$ with $\nabla^2 \mathbf{g}(x)$ being positive definite, then $\mathbf{g}^*$ is twice continuously differentiable around $\nabla \mathbf{g}(x)$. In particular

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

Therefore, by the chain rule the conclusion follows easily.   □

Now, this type of result can also be stated for nondifferentiable functions. Indeed, it is possible to characterize the viscosity subgradients (resp. supergradients) of $\varphi_{\mathbf{g}}$ when it is only l.s.c. (resp. u.s.c.). Before going further, we recall the definition of the viscosity subgradients and supergradients.

Definition 2.4. *Let $\omega : \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ be a given function and $z \in \mathbf{dom}\,\omega$.*

  a. *If ω is l.s.c., a vector $\zeta \in \mathbb{R}^k$ is called a viscosity subgradient of ω at z if and only if there is a function ψ, continuously differentiable around z, so that $\nabla \psi(z) = \zeta$ and $\omega - \psi$ has a local minimum at z.*
  b. *If ω is u.s.c., a vector $\zeta \in \mathbb{R}^k$ is called a viscosity supergradient of ω at x if and only $-\zeta$ is a viscosity subgradient of $-\omega$ at z.*

The viscosity subdifferential (resp. superdifferential) is set of viscosity subgradients (resp. supergradients) of ω at z and it is denoted by $\partial_V \omega(z)$ (resp. $\partial^V \omega(z)$).

The following result is going to be helpful in Section 4 where we use the viscosity subgradients and supergradients to investigate the value function of a Mayer problem in an application to optimal control.

Proposition 2.2. *Let $(\mathbf{\Omega}, \mathbf{g})$ be a Legendre zone consistent pair with $\mathbf{g}$ being twice continuously differentiable in a neighborhood of $x \in \mathbf{\Omega}$ and $\nabla^2 \mathbf{g}(x)$ being nonsingular. Suppose that f is l.s.c., then*

$$\partial_V \varphi_{\mathbf{g}}(\nabla \mathbf{g}(x)) = [\nabla^2 \mathbf{g}(x)]^{-1} \partial_V f(x).$$

If on the other hand, f is u.s.c., then

$$\partial^V \varphi_{\mathbf{g}}(\nabla \mathbf{g}(x)) = [\nabla^2 \mathbf{g}(x)]^{-1} \partial^V f(x).$$

Proof. Suppose f is l.s.c., the other case is analogous. Let $y = \nabla \mathbf{g}(x)$ and take $\zeta \in \partial_V \varphi_{\mathbf{g}}(y)$, then there exist $\delta > 0$ and a function ψ, continuously differentiable on $\mathbb{B}(y, \delta)$, such that $\nabla \psi(y) = \zeta$ and

$$\varphi_{\mathbf{g}}(\tilde{y}) - \psi(\tilde{y}) \geq \varphi_{\mathbf{g}}(y) - \psi(y), \quad \forall \tilde{y} \in \mathbb{B}(y, \delta).$$

Set $\mathcal{O} = \nabla \mathbf{g}^*(\mathbb{B}(y, \delta)) \subseteq \mathbf{\Omega}$ and consider $\psi_{\mathbf{g}} : \mathcal{O} \to \mathbb{R}$ given by

$$\psi_{\mathbf{g}}(\tilde{x}) = \psi \circ \nabla \mathbf{g}(\tilde{x}), \quad \forall \tilde{x} \in \mathcal{O}.$$

This function is continuous differentiable function in a neighborhood of x, and by Proposition 2.1, $\nabla \psi(y) = [\nabla^2 \mathbf{g}(x)]^{-1} \nabla \psi_{\mathbf{g}}(x)$. Furthermore, by Theorem 2.1, x is a local minimizer of $f - \psi_{\mathbf{g}}$. Hence $\psi_{\mathbf{g}}$ is an admissible test function, which means that $\nabla \psi_{\mathbf{g}}(x) \in \partial_V f(x)$, and consequently $\partial_V \varphi_{\mathbf{g}}(\nabla \mathbf{g}(x)) \subseteq [\nabla^2 \mathbf{g}(x)]^{-1} \partial_V f(x)$. The other inclusion is similar, therefore the conclusion follows. □ □

3. Gradient-like algorithms for solving optimization problems.

3.1. **Curvilinear search methods.** Inspired by the dual scheme proposed in the previous section, we study an interior point algorithm for the constrained optimization problem (P_o). This strategy can be seen as a line

search algorithm on $\mathbb{R}^n$ transported to the original space $\mathbf{\Omega}$ by means of the Legendre change of coordinates $y = \nabla\mathbf{g}(x)$.

From this point onward, we assume that the following holds:

$$(H_0) \qquad \begin{cases} (i) & (\mathbf{\Omega}, \mathbf{g}) \text{ is a Legendre zone consistent pair.} \\ (ii) & \mathbf{g} \in \mathcal{C}^2(\mathbf{\Omega}) \text{ with } \nabla^2\mathbf{g}(x) \text{ positive definite on } \mathbf{\Omega}. \end{cases}$$

We recall that the general scheme we are interested in is the following:

$$(6) \qquad\qquad x_{k+1} = \nabla\mathbf{g}^*(\nabla\mathbf{g}(x_k) + t_k d_k) \qquad \forall k \geq 0,$$

where $d_k \in \mathbb{R}^n$ and $t_k > 0$ are appropriate descent directions and step lengths. It is important to note that each iteration of (6) belongs to $\mathbf{\Omega}$, and so it is an interior point method. This strategy can be classified as well as a Geodesic search method (cf. [33, 39]) because, as we will see in Section 5, the map $t \mapsto \nabla\mathbf{g}^*(\nabla\mathbf{g}(x_k) + t d_k)$ defines a geodesic on $\mathbf{\Omega}$ when the latter is endowed with a suitable Riemannian metric.

Remark 3.1. *The line search strategy is a particular case of* (6). *Indeed, if* $\mathbf{\Omega} = \mathbb{R}^n$, *then* $\mathbf{g}(x) = \frac{1}{2}|x|^2$ *is an admissible Legendre function for* (H_0) *to hold.* $\qquad\square$

3.2. **Descent directions.** The interest of this algorithm lies in the dual lecture that we can give to (6). First of all, note that if we transport the equation (6) to the dual space $\mathbb{R}^n$ by means of the Legendre transform and we set $y_k = \nabla\mathbf{g}(x_k)$ for any $k \in \mathbb{N}$, we get

$$(7) \qquad\qquad y_{k+1} = y_k + t_k d_k, \quad \forall k \geq 0.$$

This sequence corresponds to a line search strategy on the Euclidean space $\mathbb{R}^n$. 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 (D_o).

The next table shows some suitable choices for the descent directions. By Proposition 2.1 we have that $\nabla\varphi_\mathbf{g}(y_k) = \nabla_\mathbf{g} f(x_k)$, and so we can also exhibit the expressions in terms of the dual variable y_k.

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

	d_k in $\mathbf{\Omega}$	d_k in $\mathbb{R}^n$
Steepest descent	$-\nabla_\mathbf{g} f(x_k)$	$-\nabla\varphi_\mathbf{g}(y_k)$
Quasi-Newton	$-M_k^{-1}\nabla_\mathbf{g} f(x_k)$	$M_k^{-1}\nabla\varphi_\mathbf{g}(y_k)$

In Table 1, M_k is a real nonsingular matrix. Note that the directions of Table 1 yield to the steepest descent and the quasi-Newton method for the dual problem (D_o), respectively. In particular, the way in which M_k is chosen can be based on conditions involving the dual function $\varphi_\mathbf{g}$ as well. We refer the reader to [28, Chapter 6] for more details about the Quasi-Newton methods.

Remark 3.2. *Other choices for the descent directions d_k, not based on the dual problem, are also possible. In* [4] *the authors considered the iteration*

$$(8) \qquad x_{k+1} \in \arg\min\{t_k \langle \nabla f(x_k), x \rangle + D_{\mathbf{g}}(x, x_k) : \ x \in \mathbf{\Omega}\}, \quad \forall k \in \mathbb{N},$$

where $D_{\mathbf{g}}$ denotes the Bregman pseudo-distance *associated with $\mathbf{g}$:*

$$D_{\mathbf{g}}(x, \tilde{x}) := \mathbf{g}(x) - \mathbf{g}(\tilde{x}) - \langle \nabla\mathbf{g}(\tilde{x}), x - \tilde{x} \rangle, \quad \forall x, \tilde{x} \in \mathbf{\Omega}.$$

If (H_0) holds, then using the optimality condition and the strict convexity of $\mathbf{g}$, it is not difficult to see that x_{k+1} in (8) *is characterized by*

$$\lambda_k \nabla f(x_k) + \nabla\mathbf{g}(x_{k+1}) - \nabla\mathbf{g}(x_k) = 0.$$

Therefore, by (1) *we see that* (8) *corresponds to* (6) *with $d_k = -\nabla f(x_k)$.* □

We now show that (6) is well-defined for a class of descent directions.

Proposition 3.1. *Assume that (H_0) holds and $f \in \mathcal{C}^1(\mathbf{\Omega})$. Let us define*

$$(9) \qquad\qquad \Delta(x) := \{d \in \mathbb{R} \mid \ \langle \nabla_{\mathbf{g}} f(x), d \rangle < 0\}, \quad \forall x \in \mathbf{\Omega}.$$

Then, for any $x \in \mathbf{\Omega}$ and any $d \in \Delta(x)$ there exists $t > 0$ such that

$$f(x) > f(\nabla\mathbf{g}^*(\nabla\mathbf{g}(x) + td)).$$

In particular, the sequence (6) *is well-defined whenever $d_k \in \Delta(x_k)$ and moreover, $\{f(x_k)\}_{k \in \mathbb{N}}$ is nonincreasing.*

Proof. Let $x \in \mathbf{\Omega}$ and $d \in \Delta(x)$, we set $X(t) := \nabla\mathbf{g}^*(\nabla\mathbf{g}(x) + td)$ for any $t \in \mathbb{R}$, this is well-defined because $\mathbf{dom}\ \mathbf{g}^* = \mathbb{R}^n$. Let $\phi = f \circ X$, hence

$$\frac{d}{dt}\phi(t) = -\langle \nabla f(X(t)), \nabla^2 \mathbf{g}^*(\nabla\mathbf{g}(x) + td)d \rangle, \quad \forall t \in \mathbb{R}.$$

Recall from the proof of Proposition 2.1 that $\nabla^2 \mathbf{g}^*(\nabla\mathbf{g}(\tilde{x})) = [\nabla^2 \mathbf{g}(\tilde{x})]^{-1}, \quad \forall \tilde{x} \in \mathbf{\Omega}$. Therefore, we get that $\frac{d}{dt}\phi(t) = \langle \nabla_{\mathbf{g}} f(X(t)), d \rangle$. Consequently, since $\phi(0) < 0$ and $t \mapsto \nabla_{\mathbf{g}} f(X(t))$ is continuous, there exists $\delta > 0$ such that $\frac{d}{dt}\phi(t) < 0$ for any $t \in [0, \delta)$. So the conclusion follows easily. □ □

3.3. **Step lengths.** We now discuss how to set rules for choosing the step length by transferring the information with the help of the Legendre transform.

Definition 3.1. *Let $(\mathbf{\Omega}, \mathbf{g})$ be a Legendre zone consistent pair and $f \in \mathcal{C}^1(\mathbf{\Omega})$. Take $x \in \mathbf{\Omega}$ and $d \in \mathbb{R}^n$, consider $X(t) = \nabla\mathbf{g}^*(\nabla\mathbf{g}(x) + td)$ for any $t \in \mathbb{R}$. We say that $t_0 \in \mathbb{R}$ satisfies:*

    a. *the curvilinear Cauchy rule if $t_0 \in \arg\min\{f(X(t)) \mid \ t > 0\}$.*
    b. *the curvilinear Wolfe rule if there exist $0 < c_1 < c_2 < 1$ so that*

$$f(X(t_0)) \le f(x) + c_1 t_0 \langle \nabla_{\mathbf{g}} f(x), d \rangle \quad and \quad \langle \nabla_{\mathbf{g}} f(X(t_0)), d \rangle \ge c_2 t_0 \langle \nabla_{\mathbf{g}} f(x), d \rangle.$$

Note that the curvilinear Cauchy and Wolfe rules correspond to the exact line search and the Wolfe rules applied to (D_o), respectively; see for example [28, Chapter 3] for a concise exposition about there rules. Hence, it is not difficult to see that in our setting these rules are well defined; cf. [28, Lemma 3.1].

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

Proposition 3.2. *Suppose that (H_0) holds and that $f \in \mathcal{C}^1(\mathbf{\Omega})$. Take $x_0 \in \mathbf{\Omega}$ and consider the sequence given by (6) with $d_k \in \Delta(x_k)$ (see (9)) and t_k satisfying the curvilinear Wolfe rule. Then $\nabla_{\mathbf{g}} f(x_k) \to 0$ as long as $k \to +\infty$ provided that*

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

$$(10) \qquad\qquad -\frac{\langle \nabla_{\mathbf{g}} f(x_k), d_k \rangle}{|\nabla_{\mathbf{g}} f(x_k)||d_k|} > \delta, \quad \forall k \in \mathbb{N}.$$

    ii. *$x \mapsto \nabla_{\mathbf{g}} f(x)$ verifies a Lipschitz estimate: $\exists L > 0$ so that $\forall x, \tilde{x} \in \mathbf{\Omega}$ we have*

$$(11) \qquad f(x), f(\tilde{x}) \leq f(x_0) \quad \Rightarrow \quad |\nabla_{\mathbf{g}} f(x) - \nabla_{\mathbf{g}} f(\tilde{x})| \leq L \, \mathrm{d}_{\mathbf{g}}(x, \tilde{x}).$$

Proof. By Proposition 2.1 we get that (11) implies that $y \mapsto \nabla \varphi_{\mathbf{g}}(y)$ is L-Lipschitz on $\{y \in \mathbb{R}^n \mid \varphi_{\mathbf{g}}(y) \leq f(x_0) = \varphi_{\mathbf{g}}(y_0)\}$. Since the curvilinear Wolfe rule corresponds to apply the Wolfe rule to (D_o), we have that the so-called Zoutendijk condition holds for the dual function $\varphi_{\mathbf{g}}$; see for instance [28, Theorem 3.2]. Therefore,

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

Finally, by (10) and Proposition 2.1 we get that $\sum_{k=0}^{\infty} |\nabla \varphi_{\mathbf{g}}(y_k)|^2 < +\infty$. Consequently, $\nabla_{\mathbf{g}} f(x_k) = \nabla \varphi_{\mathbf{g}}(y_k) \to 0$ as long as $k \to +\infty$. $\qquad \square$

3.4. **Convexity properties and value convergence.** In this section we investigate the special case when the function $\varphi_{\mathbf{g}}$ turns out to be convex. This situation can also be understood in the light of variable metrics. Indeed, the convexity of $\varphi_{\mathbf{g}}$ yields to the convexity of f but in the sense of Riemannian Geometry, i.e. convexity along geodesics; see for example [35]. A short discussion about this fact has also been included in Section 5; see Remark 5.2.

Remark 3.3. *Consider f as in Example 2.1 and $\mathbf{\Omega} = (0, +\infty) \times (-\frac{\pi}{2}, \frac{\pi}{2})$. We take $\mathbf{g}(x_1, x_2) = x_1 \log x_1 - x_1 + \frac{1}{2} \tan^2 x_2$ as well. Under these circumstances, we have that dual function is convex because it is given by*

$$\varphi_{\mathbf{g}}(y_1, y_2) = y_1^2 + \exp y_2, \quad (y_1, y_2) \in \mathbb{R}^2.$$

$\square$

To prove a convergence theorem for convex function in the Riemannian sense (Theorem 3.1 below), we require a convergence result for the classical convex case. The next statement is well-known (cf. [12]) and so we skip its proof.

Lemma 3.1. *Let $\varphi \in \Gamma_0(\mathbb{R}^n) \cap \mathcal{C}^1(\mathbb{R}^n)$. Suppose that φ is bounded from below and that $\nabla\varphi$ is L-Lipschitz on $\{y \in \mathbb{R}^n \mid \varphi(y) \leq \varphi(y_0)\}$. Let $\{y_k\} \subseteq \mathbb{R}^n$ be the sequence defined via (7) starting from $y_0 \in \mathbb{R}^n$ and such that $\{\varphi(y_k)\}$ is nonincreasing. Suppose that $t_k \in [0, t_{max}]$ with $0 < t_{max} < \frac{2}{L}$ and $\sum t_k = \infty$. Then*

    i. *$\varphi(y_k) \to \inf\varphi$ as long as $k \to \infty$.*
    ii. *If $\arg\min\varphi \neq \emptyset$, $\exists y_\infty \in \arg\min\varphi$ such that $y_k \to y_\infty$.*

The convergence theorem reads as follows.

Theorem 3.1. *Assume that (H_0) holds and take $f \in \mathcal{C}^1(\mathbf{\Omega})$ bounded from below and continuous on its domain. Let $\{x_k\}$ be the sequence induced by (6) with $x_0 \in \mathbf{\Omega}$. Suppose that $f \circ \nabla\mathbf{g}^*$ is convex and (11) holds as well. Take $t_k \in [0, t_{max}]$ with $0 < t_{max} < \frac{2}{L}$ and $\sum t_k = \infty$ where $L > 0$ is the constant given by (11) on $\{x \in \mathbf{\Omega} \mid f(x) \leq f(x_0)\}$. Consider the descent directions in (6) as $d_k = -\nabla_{\mathbf{g}}f(x_k)$. Then*

$$\lim_{k \to +\infty} f(x_k) = \mathrm{val}\,(P_o).$$

Furthermore, if $\{x \in \mathbf{\Omega} \mid f(x) = \mathrm{val}\,(P_o)\} \neq \emptyset$, then the sequence induced by (6) converges to some $x_\infty \in \{x \in \mathbf{\Omega} \mid f(x) = \mathrm{val}\,(P_o)\}$.

Proof. By the statement, $\varphi_{\mathbf{g}} \in \mathcal{C}^1(\mathbb{R}^n)$ and it is also convex. Moreover, $\nabla\varphi_{\mathbf{g}}$ is L-Lipschitz continuous on $\{y \in \mathbb{R}^n \mid \varphi_{\mathbf{g}}(y) \leq \varphi_{\mathbf{g}}(y_0)\}$. Indeed, by (11) we have that for any $y, \tilde{y} \in \mathbb{R}^n$ so that $\varphi_{\mathbf{g}}(y), \varphi_{\mathbf{g}}(\tilde{y}) \leq \varphi_{\mathbf{g}}(y_0)$

$$|\nabla\varphi_{\mathbf{g}}(y) - \nabla\varphi_{\mathbf{g}}(\tilde{y})| = |\nabla_{\mathbf{g}}f(\nabla\mathbf{g}^*(y)) - \nabla_{\mathbf{g}}f(\nabla\mathbf{g}^*(\tilde{y}))| \leq L\,\mathrm{d}_{\mathbf{g}}(\nabla\mathbf{g}^*(y), \nabla\mathbf{g}^*(\tilde{y})).$$

Furthermore, by (3) we get the Lipschitz continuity of $\nabla\varphi_{\mathbf{g}}$. Therefore, by **i.** in Lemma 3.1, $\varphi_{\mathbf{g}}(y_k) \to \inf\varphi_{\mathbf{g}}$ and since f is continuous on $\overline{\mathbf{\Omega}}$, by Theorem 2.1 we get that $\mathrm{val}(P_o) = \mathrm{val}(D_o)$ and so the first conclusion follows. Besides, if $\{x \in \mathbf{\Omega} \mid f(x) = \mathrm{val}\,(P_o)\} \neq \emptyset$ then $S := \arg\min\varphi_{\mathbf{g}} \neq \emptyset$. Consequently, by **ii.** in Lemma 3.1 there is $y_\infty \in S$ such that $y_k \to y_\infty$, and thus x_k is a minimizing sequence for (P_o) (see Theorem 2.1), which in addition converges to $x_\infty := \nabla\mathbf{g}^*(y_\infty)$. $\quad\square$ $\qquad\qquad\qquad\square$

4. Application to Optimal Control problems.

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

It is known that, unless some compatibility condition between constraints and dynamics holds, the value function has not enough regularity and can fail to be the unique viscosity solution to the Hamilton-Jacobi-Bellman (HJB) equation. In this section we explore the fact that the original problem

can be transported into an unconstrained one, and we use this to characterize the original value function in terms of the HJB equation.

The main theorem we study is about the characterization of the value function with convex state constraints $\mathcal{K} = \overline{\mathbf{\Omega}}$ for a system where no pointing condition holds. For simplicity the analysis is focused on the Mayer problem

$$(P_c) \qquad \vartheta(t,x) := \inf_{p \in \mathbb{S}_t^T(x)} f(p(T)), \quad \forall x \in \mathcal{K}, t \in [0,T],$$

where $\mathbb{S}_t^T(x)$ denotes the set of absolutely continuous arcs defined on $[t,T]$ verifying

$$\dot{p}(s) \in F(p(s)) \quad \text{a.e. } s \in [t,T], \quad p(t) = x, \quad p(s) \in \mathcal{K}, \ \forall s \in [t,T].$$

We recall that given $\mathcal{O} \subseteq \mathbb{R}^n$ and a Hamiltonian function $(x,\eta) \mapsto \mathbb{H}(x,\eta)$ (not necessarily associated with an optimal control problem), a *viscosity solution* of the Hamilton-Jacobi (HJ) equation (see for instance [5, Chapter 2])

$$\mathbb{H}(x, \nabla\omega(t,x)) = 0, \quad \forall (t,x) \in (0,T) \times \mathcal{O}$$

is a continuous function that is a super and subsolution in the next sense:

- An l.s.c. function ω is a *viscosity supersolution* of the HJ equation if

$$\mathbb{H}(x,(\theta,\zeta)) \geq 0, \quad \forall (t,x) \in (0,T) \times \mathcal{O}, \ (\theta,\zeta) \in \partial_V \omega(t,x).$$

- An u.s.c. function ω is a *viscosity subsolution* of the HJ equation if

$$\mathbb{H}(x,(\theta,\zeta)) \leq 0, \quad \forall (t,x) \in (0,T) \times \mathcal{O}, \ (\theta,\zeta) \in \partial^V \omega(t,x).$$

In this paper, the hypotheses required for the characterization of the value function ϑ, differ from the standing assumptions done in the literature; see Theorem 4.1 for details. Actually, our hypotheses are due to the penalization technique we are studying. It is worthy to notice that, our assumptions are not covered by the current literature, which is mainly done for Lipschitz continuous dynamics. One of the main features of our result concerns dynamics that are *absorbing* at $\mathbf{bdry}\,(\mathbf{\Omega})$ in the sense that

$$(H_1) \qquad \exists c > 0 \text{ so that } \max_{v \in F(x)} |v| \leq c \left(\frac{1 + |\nabla \mathbf{g}(x)|}{|\nabla^2 \mathbf{g}(x)|} \right), \quad \forall x \in \mathbf{\Omega},$$

where $(\mathbf{\Omega}, \mathbf{g})$ is a Legendre zone consistent pair that verified (H_0).

Before going further, let us present some examples to illustrate the potential of the theory we develop in this paper, and how this complements the existing literature of HJB theory and optimal control.

Example 4.1. *Let us consider the control system $\dot{x} = xu$ with $u \in [-1,1]$ and the state constraints $x \geq 0$. Note that here $F(x) = [-x,x]$. We can check that (H_1) holds by taking the Boltzmann-Shannon entropy $\mathbf{g}_{\text{ent}}$ given in Remark 2.1. Let us point out that, since $F(x) = \{0\}$ at $x = 0$, this simple dynamics does not satisfy the pointing qualification conditions done in the literature (cf. [34, 15, 19]). It is also worthy to mention that, this dynamical system can be treated in the framework proposed in [21].* $\qquad \square$

Example 4.2. *Let us consider the two-dimensional control system $\dot{x}_1 = x_1 \log(x_2)$ $\dot{x}_2 = x_2 u$ with $u \in [-1, 1]$ and the state constraints $x_1, x_2 \geq 0$. Note that here $F(x_1, x_2) = \{x_1 \log(x_2)\} \times [-x_2, x_2]$. Under these circumstances, we can also check that (H_1) holds by taking the Boltzmann-Shannon entropy $\mathbf{g}_{\mathrm{ent}}$ in $\mathbb{R}^2$, that is,*

$$\mathbf{g}_{\mathrm{ent}}(x_1, x_2) = x_1 \log(x_1) - x_1 + x_2 \log(x_2) - x_2.$$

In this example the dynamics at the boundary of $\mathbf{\Omega}$ satisfies:

$$F(0, x_2) = \{0\} \times [-x_2, x_2], \quad \forall x_2 > 0 \qquad and \qquad F(x_1, 0) = \emptyset, \quad \forall x_1 \geq 0.$$

Hence, we have that, if η is a normal vector to $\overline{\mathbf{\Omega}}$ at $x \in \mathbf{bdry}\,(\mathbf{\Omega})$, the following statement holds true (sometimes by vacuity):

$$\langle v, \eta \rangle = 0, \quad \forall v \in F(x).$$

Therefore, no pointing qualification conditions (cf. [34, 15, 19]) can be satisfied by this dynamical system. Furthermore, this dynamical system can not be treated by the theory proposed in [21], because the dynamics F can not be defined in a neighborhood of the state constraints $\mathcal{K} = \overline{\mathbf{\Omega}}$. □

Remark 4.1. *We say that (H_1) is an absorbing condition at $\mathbf{bdry}\,(\mathbf{\Omega})$ because, for any $x \in \mathbf{bdry}\,(\mathbf{\Omega})$, $v \in F(x)$ and any normal vector η to $\overline{\mathbf{\Omega}}$ at x, it is likely to have $\langle v, \eta \rangle = 0$; this can be verified in Example 4.1 and Example 4.2.*

To see this in a more general setting, suppose that $\mathbf{\Omega} = \{h(x) < 0\}$ and that $\mathbf{g}(x) = \theta(h(x))$, where h belongs to $\mathcal{C}^2(\mathbf{\Omega})$ and $\theta \in \Gamma_0(\mathbb{R})$ is such that $((0, +\infty), \theta)$ is a Legendre zone consistent pair with $\theta'''(r) < 0$ for any $r > 0$ small enough. By the Fundamental Theorem of Calculus, we have that $\frac{\theta'(r)}{\theta''(r)} \to 0$ as $r \to 0$, and so $[\nabla^2 \mathbf{g}(x)]^{-1} \nabla \mathbf{g}(x) \to 0$ as x approaches to $\partial \mathbf{\Omega}$. Therefore, since any normal to $\overline{\mathbf{\Omega}}$ can be written as $\eta = -\frac{t}{|\nabla h(x)|} \nabla h(x)$ for some $t \geq 0$ and $\frac{1}{|\nabla h(x)|} \nabla h(x) = \frac{1}{|\nabla \mathbf{g}(x)|} \nabla \mathbf{g}(x)$ for any $x \in \mathbf{\Omega}$, we get

$$t|\langle v, \eta \rangle| = \left| \left\langle v, \frac{\nabla \mathbf{g}(x)}{|\nabla \mathbf{g}(x)|} \right\rangle \right| \leq c \left(\frac{1}{|\nabla \mathbf{g}(x)|} + 1 \right) \frac{|\nabla \mathbf{g}(x)|}{|\nabla^2 \mathbf{g}(x)|}, \quad \forall x \in \mathbf{\Omega}.$$

Since the righthand side converges to 0 as x approaches to $\mathbf{bdry}\,(\mathbf{\Omega})$, we obtain that $\langle v, \eta \rangle = 0$ for any $x \in \mathbf{bdry}\,(\mathbf{\Omega})$ and $v \in F(x)$. □

The proof of the characterization of the value function defined in (P_c) (Theorem 4.1 in Section 4.2) is based on the Legendre transform $\nabla \mathbf{g} : \mathbf{\Omega} \to \mathbb{R}^n$ as well. It turns out that, the problem (P_c) can be associated with another optimal control problem, that doesn't have state constraints. We study this fact in the next subsection.

4.1. **Dual optimal control problem.** We now introduce an unconstrained auxiliary problem which is intrinsically related to the original problem (P_c). We begin with the definition of an auxiliary differential inclusion.

Take $(t, x) \in [0, T] \times \mathbf{\Omega}$ and let $p \in \mathbb{S}_t^T(x)$ satisfying $p(s) \in \mathbf{\Omega}$ for any $s \in [t, T]$. Then the arc $s \mapsto q(s) := \nabla \mathbf{g}(p(s))$ is a solution of the differential inclusion

$$(12) \qquad \dot{q}(s) \in \Phi_{\mathbf{g}}(q(s)) \quad \text{a.e. } s \in [t, T], \quad q(t) = \nabla \mathbf{g}(x),$$

where $\Phi_{\mathbf{g}}(y) = \left[\nabla^2 \mathbf{g}^*(y) \right]^{-1} F(\nabla \mathbf{g}^*(y))$ for any $y \in \mathbb{R}^n$.

Remark 4.2. *Let us pick up the dynamical systems presented in Example 4.1 and Example 4.2. In the first case, the dual dynamics is $\Phi_{\mathbf{g}}(y) = [-1, 1]$ and, in the second one, is $\Phi_{\mathbf{g}}(y_1, y_2) = \{y_2\} \times [-1, 1]$.*

Moreover, suppose that there exists a Legendre function $\mathbf{g}$ together with two matrices A and B of dimensions $n \times n$ and $n \times m$, respectively, for which

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

In this setting, the dual differential equation (12) is a linear system because

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

Furthermore, if there exist two vector fields $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}^n$ so that

$$\nabla^2 \mathbf{g}(x) F(x) = \{f_1(\nabla \mathbf{g}(x)) + f_2(\nabla \mathbf{g}(x))u : \ u \in [-1, 1]\}, \quad \forall x \in \mathbf{\Omega},$$

the dual differential equation (12) is a control-affine system. $\qquad\square$

Notice that since $\mathbf{dom}\ \mathbf{g}^* = \mathbb{R}^n$, then $q(\cdot)$ remains in $\mathbb{R}^n$, and so we can associate to (12) an unconstrained Mayer process. More precisely, let $\vec{S}_t^T(y\ ; \mathbf{g})$ indicate the set of absolutely continuous curves defined on $[t, T]$ satisfying (12) with initial condition $q(t) = y$. The auxiliary problem at issue is:

$$(D_c) \quad \varpi_{\mathbf{g}}(t, y) := \inf \left\{ \varphi_{\mathbf{g}}(q(T)) \mid q \in \vec{S}_t^T(y\ ; \mathbf{g}) \right\}, \ \forall (t, q) \in [0, T] \times \mathbb{R}^n.$$

We remind that $\varphi_{\mathbf{g}} : \mathbb{R}^n \to \mathbb{R}$ is the dual function or *auxiliary cost* defined via

$$\varphi_{\mathbf{g}}(y) = f \circ \nabla \mathbf{g}^*(y), \qquad \forall y \in \mathbb{R}^n.$$

The map $\varpi_{\mathbf{g}} : [0, T] \times \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is the value function of the auxiliary Mayer problem.

On the other hand, let $q \in \vec{S}_t^T(y\ ; \mathbf{g})$, then thanks to the Legendre transform, $\nabla \mathbf{g}^*(q) \in \mathbb{S}_t^T(\nabla \mathbf{g}^*(y))$ and so, there is an one-to-one correspondence between $\vec{S}_t^T(\nabla \mathbf{g}(x)\ ; \mathbf{g})$ and the trajectories of $\mathbb{S}_t^T(x)$ that live in $\mathbf{\Omega}$. This means that

$$\vartheta(t, x) \leq \varpi_{\mathbf{g}}(t, \nabla \mathbf{g}(x)), \quad \forall (t, x) \in [0, T] \times \mathbf{\Omega}.$$

Without additional hypotheses the equality may not hold. However, under an interior approximation hypothesis, as (H_3), the equality is reached. Interior approximation hypotheses, as the recently mentioned, are made in order to ensure that any feasible trajectory can be approximated by a sequence of trajectories lying on the interior of the state constraint.

Proposition 4.1. *Suppose* (H_0) *holds. Assume that* f *is uniformly contin-uous on* $\overline{\boldsymbol{\Omega}}$ *of modulus* $\alpha_f(\cdot)$ *and that the following holds:*

$$(H_3) \qquad \begin{cases} \forall \varepsilon > 0, \ \forall t \in [0,T], \ \forall x \in \boldsymbol{\Omega}, \ \forall p \in \mathbb{S}_t^T(x), \ \exists p_\varepsilon \in \mathbb{S}_t^T(x) \\ such\ that: \ p_\varepsilon(s) \in \boldsymbol{\Omega}, \ \forall s \in [t,T] \ and \ |p(T) - p_\varepsilon(T)| \le \varepsilon. \end{cases}$$

Then, the value functions given by (P_c) *and* (D_c) *satisfy:*

$$\varpi_{\mathbf{g}}(t, \nabla \mathbf{g}(x)) = \vartheta(t,x), \quad \forall (t,x) \in [0,T] \times \boldsymbol{\Omega}.$$

Proof. If $\vartheta(t,x) = +\infty$, we get that $\varpi_{\mathbf{g}}(t, \nabla \mathbf{g}(x)) = +\infty$ as well, so under these circumstances, there is nothing to be proved.

Assume that $\vartheta(t,x) < +\infty$ and take $\{\varepsilon_k\} \subseteq (0,1)$ a sequence such that $\varepsilon_k \to 0$ as $k \to +\infty$. Thus, for any $k \in \mathbb{N}$, there exists $p_k \in \mathbb{S}_t^T(x)$ for which $f(p_k(T)) \le \vartheta(t,x) + \varepsilon_k$. Besides, by (H_3), for any $k \in \mathbb{N}$ we can find $\tilde{p}_k \in \mathbb{S}_t^T(x)$ such that

$$\tilde{p}_k(s) \in \boldsymbol{\Omega}, \ \forall s \in [t,T] \quad \text{and} \quad |\tilde{p}_k(T) - p_k(T)| \le \varepsilon_k.$$

By gathering the last inequalities we get

$$f(\tilde{p}_k(T)) \le f(p_k(T)) + \alpha_f(\varepsilon_k) = \vartheta(t,x) + \varepsilon_k + \alpha_f(\varepsilon_k).$$

Moreover, since $q_k := \nabla \mathbf{g}(\tilde{p}_k) \in \vec{S}_t^T(\nabla \mathbf{g}(x)\ ; \mathbf{g})$

$$f(\tilde{p}_k(T)) = \varphi_{\mathbf{g}}(q_k(T)) \ge \varpi_{\mathbf{g}}(t, \nabla \mathbf{g}(x)) \ge \vartheta(t,x).$$

So, letting $k \to +\infty$, the conclusion follows. $\qquad\square \qquad\qquad\square$

The importance of the previous statement lies in the fact that, for optimal control problems without state constraints, the relation between the value function and the HJB equation is well-understood, and no additional com-patibility assumptions are required. Indeed, under the assumptions we have done so far, we can identify $\varpi_{\mathbf{g}}$ as the viscosity solution of a HJB equation.

Proposition 4.2. *Suppose that* (H_0) *and* (H_1) *hold. Assume that* f *is continuous and* F *has nonempty convex compact images on* $\boldsymbol{\Omega}$. *If* F *satisfies the following Lipschitz-like estimate: for each* $r > 0$, *there is* $L > 0$ *for which*

$$(H_2) \qquad \nabla^2 \mathbf{g}(x) F(x) \subseteq \nabla^2 \mathbf{g}(\tilde{x}) F(\tilde{x}) + L\, d_{\mathbf{g}}(x, \tilde{x}) \mathbb{B}, \quad \forall x, \tilde{x} \in \boldsymbol{\Omega} \cap \mathbb{B}(0, r).$$

Then, the value function $\varpi_{\mathbf{g}}$, *defined in* (D_c), *is the unique viscosity solution of*

$$-\partial_t \omega(t,y) + \widetilde{H}(y, \nabla_y \omega(t,y)) = 0, \quad (t,y) \in (0,T) \times \mathbb{R}^n,$$

that satisfies $\omega(T,y) = \varphi_{\mathbf{g}}(y)$ *for any* $y \in \mathbb{R}^n$. *Here, the Hamiltonian is given by* $\widetilde{H}(y, \xi) = \sup\{-\langle v, \xi \rangle \mid v \in \Phi_{\mathbf{g}}(y)\}$ *for any* $y, \xi \in \mathbb{R}^n$.

Proof. Since F has nonempty convex compact images on $\boldsymbol{\Omega}$, we can easily check that $\Phi_{\mathbf{g}}$ has nonempty convex compact images on $\mathbb{R}^n$. By the absorb-ing property (H_1), we have that $\Phi_{\mathbf{g}}$ has linear growth on $\mathbb{R}^n$ and thanks to (H_2) it is also locally Lipschitz continuous on $\mathbb{R}^n$. Moreover, since f is continuous so does $\varphi_{\mathbf{g}}$.

Finally, since the auxiliary problem has no state constraints, it is a classical result that under these circumstances the value function is the unique continuous viscosity solution of the HJB equation; see for instance [5, Theorem 3.7], [14, Proposition 4.7.10] or [37, Theorem 12.3.7]. Therefore, the result follows.   □   □

4.2. Characterization of the value function of the primal problem.
Now, with all these tools at hand, we are now in position to state and provide the proof of the principal result of this section, which reads as follows.

Theorem 4.1. *Suppose (H_0) holds. Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued maps having nonempty compact convex images on $\mathbf{\Omega}$. Assume that f is uniformly continuous on $\overline{\mathbf{\Omega}}$, that the dynamics is absorbing at $\mathbf{bdry}\,(\mathbf{\Omega})$, that is, (H_1) holds, and that the Lipschitz-like condition (H_2) holds as well.*

Then, the value function ϑ, defined in (P_c), is the unique uniformly continuous function on $\mathbf{\Omega}$, that satisfies the final time condition $\vartheta(T,x) = f(x)$ for any $x \in \mathbf{\Omega}$ and that is a viscosity solution of

$$-\partial_t \omega(t,x) + H(x, \nabla_x \omega(t,x)) = 0, \quad (t,x) \in (0,T) \times \mathbf{\Omega},$$

where $H(x,\xi) = \sup\{-\langle v, \xi \rangle \mid v \in F(x)\}$ for any $x, \xi \in \mathbb{R}^n$

Proof. First of all, note that (H_3) holds trivially because for any $x \in \mathbf{\Omega}$, $t \in [0,T]$ and $p \in \mathbb{S}_t^T(x)$, we have $p(s) \in \mathbf{\Omega}$ as long as $s \in [t,T]$. Indeed, by shifting t if necessary, we can assume that $T = \inf\{s \in [t,T] \mid p(s) \notin \mathbf{\Omega}\}$. Due to the Gronwall's Lemma (cf. [14, Proposition 4.1.4]), for any $\tau \in [t,T)$, each $q \in \vec{S}_t^\tau(\nabla \mathbf{g}(x)\,;\mathbf{g})$ satisfies

$$|q(\tau)| \leq (1 + |\nabla \mathbf{g}(x)|)e^{c(\tau - t)}.$$

Notice that for any $\tau \in [0,T)$, $\nabla \mathbf{g}(p) \in \vec{S}_t^\tau(\nabla \mathbf{g}(x)\,;\mathbf{g})$ because $p(s) \in \mathbf{\Omega}$ on $[t,T]$. Therefore,

$$|\nabla \mathbf{g}(p(\tau))| \leq (1 + |\nabla \mathbf{g}(x)|)e^{c(\tau - t)}, \quad \forall \tau \in [t,T).$$

Since $\mathrm{dist}(p(\tau), \mathbf{bdry}\,(\mathbf{\Omega})) \to 0$ as $\tau \to T$, we get a contradiction with the fact that $\mathbf{g}$ is essentially smooth on $\mathbf{\Omega}$, which implies that (H_3) holds.

On the one hand, by Proposition 4.2 we have $\varpi_{\mathbf{g}}$ is the unique continuous function that satisfies

$$(13) \qquad -\theta + \widetilde{H}(y, \varrho) \geq 0, \quad \forall(t,y) \in (0,T) \times \mathbb{R}^n, \ \forall(\theta, \varrho) \in \partial_V \omega(t,y),$$

$$(14) \qquad -\theta + \widetilde{H}(y, \varrho) \leq 0, \quad \forall(t,y) \in (0,T) \times \mathbb{R}^n, \ \forall(\theta, \varrho) \in \partial^V \omega(t,y),$$

$$(15) \qquad \omega(t,y) = \varphi_{\mathbf{g}}(y), \quad \forall y \in \mathbb{R}^n.$$

On the other hand, Proposition 4.1 implies that $\vartheta(t,x) = \varpi_{\mathbf{g}}(t, \nabla \mathbf{g}(x))$ for any $(t,x) \in [0,T] \times \mathbf{\Omega}$. By Proposition 2.2 we get that $(\theta, \nabla \mathbf{g}(x)\zeta) \in \partial_V \vartheta(t,x)$ if and only if $(\theta, \zeta) \in \partial_V \varpi_{\mathbf{g}}(t, \nabla \mathbf{g}(x))$. A similar relation also holds

for the viscosity superdifferential. Hence, the value function ϑ satisfies

$$-\theta + H(x, \zeta) \geq 0, \quad \forall (t, x) \in (0, T) \times \mathbf{\Omega}, \ \forall (\theta, \zeta) \in \partial_V \vartheta(t, x),$$
$$-\theta + H(x, \zeta) \leq 0, \quad \forall (t, x) \in (0, T) \times \mathbf{\Omega}, \ \forall (\theta, \zeta) \in \partial^V \vartheta(t, x),$$
$$\vartheta(T, x) = f(x), \quad \forall x \in \mathbf{\Omega}.$$

The previous arguments show that the value function ϑ is a viscosity solution of the HJB equation on $(0, T) \times \mathbf{\Omega}$. The uniqueness comes from the fact that, if $\alpha : [0, T] \times \mathbf{\Omega} \to \mathbb{R}$ is a viscosity solution of the preceding HJB equation, Proposition 2.2 implies that $\beta(t, y) = \alpha(t, \nabla \mathbf{g}^*(y))$ defined on $[0, T] \times \mathbb{R}^n$ is a viscosity solution of (13)-(14)-(15), so by Proposition 4.2 we obtain that $\beta = \varpi_{\mathbf{g}}$ and wherefore $\alpha = \vartheta$. $\qquad\qquad \square \qquad\qquad \square$

4.3. **Discussion.** In this section we have investigated the HJB approach for problems with state constraints from a point of view that seems to be quite new for optimal control. The methodology used has been widely studied in mathematical programming theory in the so-called *interior-point methods*, where a suitable barrier function is introduced in order to construct algorithms, whose iterations are strictly feasible on the interior of the constraints; see for instance [28, 30] among many others. The technique has also been employed to study continuous versions of numerical methods, usually referred as central path methods; see for instance [6, 26, 22, 2, 8].

Here we have taken the idea of penalizing the state constraints and make use of it in order to study the value function of an optimal control problem with convex state constraints. The outcome is that we have identified a class of problems, neither covered by the current literature nor by the more recent works ([7, 21]).

5. A Riemannian manifolds interpretation

We finish the paper with an interpretation of the results in terms of Riemannian geometry; for a deeper introduction into the subject we refer to [17, 24].

We evoke from the classical theory of differential geometry (cf. [25, Proposition 3.10]) that the tangent space $T_x M$ to a smooth ($\mathcal{C}^\infty$) manifold M is a finite dimensional vector space for any $x \in M$. This fact allows to endow each $T_x M$ with an inner product $(\cdot, \cdot)_x$. Roughly speaking, when this procedure can be done in such a way that the inner dot depends on x in an appropriate smooth manner, the manifold is called a Riemannian manifold. When M is an open set of $\mathbb{R}^n$, tangent space $T_x M$ can be identified with the whole space $\mathbb{R}^n$, and so Definition 2.3 provides a suitable notion of Riemannian manifold when $M = \mathbf{\Omega}$.

One of the main purposes 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 $(\mathbf{\Omega}, \mathbf{g})$ be a Legendre zone consistent pair with $\mathbf{g} \in \mathcal{C}^2(\mathbf{\Omega})$. For any $x \in \mathbf{\Omega}$ consider the bilinear mapping defined on $\mathbb{R}^n \times \mathbb{R}^n$

as follows

$$\rho_x(u,v) := \langle \nabla^2 \mathbf{g}(x)u, \nabla^2 \mathbf{g}(x)v \rangle, \qquad \forall u, v \in \mathbb{R}^n. \tag{16}$$

It turns out that the preceding bilinear map can define a Riemannian metric, in the sense of Definition 2.3, if the hypothesis (H_0) is slightly strengthen.

Lemma 5.1. *Let $\mathbf{\Omega} \subseteq \mathbb{R}^n$ be convex and consider $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ so that (H_0) holds with $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$. Then, the family $\{\rho_x\}_{x \in \mathbf{\Omega}}$ is a $\mathcal{C}^k$-Riemannian metric on $\mathbf{\Omega}$.*

Proof. We have that $[\nabla^2 \mathbf{g}(x)]^2$ is symmetric and positive definite for any $x \in \mathbf{\Omega}$, so, ρ_x is an inner product on $\mathbb{R}^n$. Besides, since $\mathbf{\Omega}$ is an open set, the tangent space to $\mathbf{\Omega}$ at x can be identified with $\mathbb{R}^n$ for any $x \in \mathbf{\Omega}$. Furthermore, given that $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$, for any $\Psi_1, \Psi_2 \in \mathcal{C}^k(\mathbf{\Omega})$, the map $x \mapsto \rho_x(\Psi_1(x), \Psi_2(x))$ belongs to $\mathcal{C}^k(\mathbf{\Omega})$. Hence, $\{\rho_x\}_{x \in \mathbf{\Omega}}$ is a $\mathcal{C}^k$-Riemannian metric on $\mathbf{\Omega}$. $\qquad\qquad\square$

Definition 5.1. *Let $\mathbf{\Omega} \subseteq \mathbb{R}^n$ be convex and $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ so that (H_0) holds with $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$. The family of inner products given by (16) is called the squared Hessian Riemannian metric on $\mathbf{\Omega}$ induced by the Legendre function $\mathbf{g}$, and it is denoted by $(\cdot, \cdot)_x^{\mathbf{g}}$, that is, $\forall x \in \mathbf{\Omega}$,*

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

Consequently, with a slight abuse of notation, we say that $(\mathbf{\Omega}, \mathbf{g})$ is a squared Hessian Riemannian manifold.

Remark 5.1. *Take $\mathbf{\Omega} = \mathbb{B}$ in $\mathbb{R}^n$, a suitable choice of Legendre function that makes $(\mathbf{\Omega}, \mathbf{g})$ a squared Hessian Riemannian manifolds is*

$$\mathbf{g}(x) = \begin{cases} -\log\left(\cos\left(\frac{\pi}{2}|x|\right)\right) & |x| < 1 \\ +\infty & |x| \le 1. \end{cases}$$

$\square$

In the light of the foregoing definition, we can see that all along this paper we have basically studied a nonlinear optimization problem and a Mayer problem on a squared Hessian Riemannian manifold. This explain the form of the hypotheses we have required. Actually, as we will see later, the assumption **dom** $\mathbf{g}^* = \mathbb{R}^n$ is nothing else that the completeness of the Riemannian manifold as a metric space endowed with the distance given by (3).

As we have claimed before, $\mathbf{\Omega}$ endowed with the squared Hessian Riemannian metric is isomorphic to the euclidean space $\mathbb{R}^n$. To give a precise statement of this result, we recall first that if M and N are two Riemannian manifolds, a diffeomorphism $\Psi : M \to N$ is called an *isometry between M and N* provided

$$(u,v)_x = (d_x\Psi(u), d_x\Psi(v))_{\Psi(x)}, \quad \forall x \in M, \ \forall u, v \in T_x M. \tag{17}$$

Lemma 5.2. *Let $\mathbf{\Omega} \subseteq \mathbb{R}^n$ be convex and consider $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ so that (H_0) holds with $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$. Then $\nabla \mathbf{g} : \mathbf{\Omega} \longrightarrow \mathbb{R}^n$ is a $\mathcal{C}^k$ isometry between the Riemannian manifold $(\mathbf{\Omega}, \mathbf{g})$ induced by $\mathbf{g}$ and $\mathbb{R}^n$ endowed with the Euclidean metric.*

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

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

Therefore, (17) is verified and so $\nabla \mathbf{g}$ is an isometry between $(\mathbf{\Omega}, \mathbf{g})$ and $\mathbb{R}^n$ endowed with the Euclidean metric. $\qquad\qquad \square \qquad\qquad \square$

In the light of the previous lemma, we can fully characterize the geodesics on a squared Hessian Riemannian manifold. The next result, and in particular the formula (18), explains why the algorithm (6) is a geodesic search method.

Proposition 5.1 (Geodesics)**.** *Let $\mathbf{\Omega} \subseteq \mathbb{R}^n$ be convex and $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ so that (H_0) holds with $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$. The geodesic on $(\mathbf{\Omega}, \mathbf{g})$ that satisfies $X(0) = x_0 \in \mathbf{\Omega}$ and $\dot{X}(0) = v_0 \in \mathbb{R}^n$ is given by*

$$(18) \qquad X(t) = \nabla \mathbf{g}^*(\nabla \mathbf{g}(x_0) + t \nabla^2 \mathbf{g}(x_0) v_0), \ \forall t \in (T_{min}, T_{max}).$$

Proof. Recall that by the Fundamental Lemma of Riemannian Geometry [24, Theorem 5.4] the Levi-Civita is the unique linear connection on $(\mathbf{\Omega}, \mathbf{g})$, so by the Existence and Uniqueness of Geodesics Theorem [24, Theorem 4.10] such geodesic is uniquely determined on an interval containing zero. Let $X(\cdot)$ be the geodesic defined on a maximal interval (T_{min}, T_{max}). By virtue of Lemma 5.2 and [24, Proposition 5.6], $Y(t) = \nabla \mathbf{g}(X(t))$ is a geodesic in $\mathbb{R}^n$ endowed with the Euclidean metric, this means that $Y(t) = y_0 + tw_0$ for some $\exists y_0, w_0 \in \mathbb{R}^n$. Therefore, $y_0 = \nabla \mathbf{g}(x_0)$ and since $\dot{Y}(t) = \nabla^2 \mathbf{g}(X(t)) \dot{X}(t)$, we deduce that $w_0 = \nabla^2 \mathbf{g}(x_0) v_0$. Hence, thanks to (1) we finally obtain (18). $\qquad\qquad \square \qquad\qquad \square$

Remark 5.2. *Notice that by (18), there is a unique geodesic $X : [0, 1] \to \mathbf{\Omega}$ defined on $(\mathbf{\Omega}, \mathbf{g})$ joining $x, \tilde{x} \in \mathbf{\Omega}$, which is defined via*

$$X(t) = \nabla g^*(\nabla g(\tilde{x}) + t(\nabla g(x) - \nabla g(\tilde{x}))).$$

Consequently, setting $y = \nabla g(x)$ and $\tilde{y} = \nabla g(\tilde{x})$ we get

$$f \circ X(t) = \varphi_{\mathbf{g}}(\tilde{y} + t(y - \tilde{y})), \quad \forall t \in [0, 1].$$

Hence, since $t f(\tilde{x}) + (1 - t) f(x) = t \varphi_{\mathbf{g}}(\tilde{y}) + (1 - t) \varphi_{\mathbf{g}}(y)$ for any $t \in [0, 1]$, it is recognizable that $\varphi_{\mathbf{g}}$ being convex, as in Section 3.4, implies that f is convex along geodesics. Some authors call this property convexity in the Riemannian sense; see for instance [35]. $\qquad \square$

Remark 5.3. *Consider the data used in Remark 5.1. Basic calculus rules for the Legendre-Fenchel transform yield to*

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

Consequently, by Proposition 5.1 the geodesic of $(\mathbf{\Omega}, \mathbf{g})$ starting from $x_0 \in \mathbf{\Omega}$ with initial velocity $\nabla^2 \mathbf{g}(x_0)^{-1} d_0$ is determined by

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

$\square$

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 $\mathbf{g}$ (see Remark 3.2). In this case, the variable t acts like a penalization parameter. For more details about central path methods we refer the reader to [36, Chapter 25].

Proposition 5.2 (Central Path). *Let $\mathbf{\Omega} \subseteq \mathbb{R}^n$ be convex and $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ so that (H_0) holds with $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$. Let $t \mapsto X(t)$ be the geodesic on $(\mathbf{\Omega}, \mathbf{g})$ that satisfies $X(0) = x_0 \in \mathbf{\Omega}$ and $\dot{X}(0) = -\nabla^2\mathbf{g}(x_0)^{-1}d_0$ with $d_0 \in \mathbb{R}^n$. Then*

$$(19) \qquad X(t) = \underset{x \in \mathbf{\Omega}}{\operatorname{argmin}}\left\{\langle d_0, x\rangle + \frac{1}{t}D_{\mathbf{g}}(x, x_0)\right\}, \quad \forall t \in (0, T_{max}).$$

Proof. Let $t \mapsto X(t)$ be the maximal geodesic starting from x_0 with initial velocity v_0. By (18) and (1), $X(t)$ solves

$$(20) \qquad\qquad td_0 + \nabla\mathbf{g}(x) - \nabla\mathbf{g}(x_0) = 0, \quad \forall t \in (T_{\min}, T_{\max}).$$

Since, $X(t) \in \mathbf{\Omega}$ for any $t \in (T_{\min}, T_{\max})$, the equation (20) corresponds to the first-order optimality condition for the following minimization problem:

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

which is equivalent to the righthand side of (19) for any $t \in (0, T_{max})$, in terms of optimal solutions. Note that both problems are convex, so $X(t)$ solves problem on the righthand side of (19). By strict convexity of $\mathbf{g}$ on $\mathbf{\Omega}$, the solution is unique, so (19) holds true. $\qquad\square\qquad\qquad\square$

Since Riemannian isometries are arc-length preserving (cf. [24, Proposition 5.6]), it is possible to characterize the Riemannian distance on a squared Hessian Riemannian manifold as well. The distance on a Riemannian manifold M gives to the manifold the topological structure of metric space and

it is defined as follows:

$$\text{dist}_M(x, \tilde{x}) := \inf_{\gamma \in AC([0,1];M)} \{\ell(\gamma) \mid \ell(0) = x \text{ and } \ell(1) = \tilde{x}\}, \quad \forall x, \tilde{x} \in M,$$

where $AC([0,1]; M)$ stands for the set of all the absolutely continuous curves $\gamma : [0,1] \to M$ and $\ell(\gamma)$ is the length of $\gamma \in AC([0,1]; M)$, which is given by

$$\ell(\gamma) = \int_0^1 \sqrt{(\dot{\gamma}(t), \dot{\gamma}(t))_{\gamma(t)}} dt.$$

Corollary 5.1 (Geodesic distance). *Let $\mathbf{\Omega} \subseteq \mathbb{R}^n$ be convex and $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ so that (H_0) holds with $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$. Then, the geodesic distance $\mathbf{dist}_{\mathbf{\Omega}}(\cdot, \cdot)$ on the squared Hessian Riemannian manifold associated with a Legendre zone consistent pair $(\mathbf{\Omega}, \mathbf{g})$ coincides with the expression given by (3), that is,*

$$(21) \qquad \mathbf{dist}_{\mathbf{\Omega}}(x, \tilde{x}) = |\nabla \mathbf{g}(x) - \nabla \mathbf{g}(\tilde{x})|, \qquad \forall x, \tilde{x} \in \mathbf{\Omega}.$$

Proof. Set $y = \nabla \mathbf{g}(x)$ and $\tilde{y} = \nabla \mathbf{g}(\tilde{x})$, then

$$Y(t) = t\tilde{y} + (1 - t)y \in \mathbb{R}^n, \quad \forall t \in [0, 1].$$

This curve is a geodesic in $\mathbb{R}^n$ and so, by virtue of [24, Proposition 5.6], the curve $X(t) = \nabla \mathbf{g}^*(Y(t))$ is a geodesic on $\mathbf{\Omega}$. Hence, Lemma 5.2 implies that $\ell(X) = \ell(Y)$, and this yields to $\mathbf{dist}_{\mathbf{\Omega}}(x, \tilde{x}) \leq \ell(X) = \ell(Y) = |y - \tilde{y}|$.

On the other hand, for any $\varepsilon > 0$, there exists an absolutely continuous curve $X_\varepsilon : [0, 1] \to \mathbf{\Omega}$ joining x and $\tilde{x}$ such that $\ell(X_\varepsilon) \leq \mathbf{dist}_{\mathbf{\Omega}}(x, \tilde{x}) + \varepsilon$.

Since $Y_\varepsilon = \nabla \mathbf{g} \circ X_\varepsilon$ is an absolutely continuous curve joining y and $\tilde{y}$ we have that $\ell(Y_\varepsilon) \geq |y - \tilde{y}|$. Therefore, by Lemma 5.2 the length of X_ε and Y_ε has the same value, so letting $\varepsilon \to 0$ we get (21). $\square$ $\square$

On the other hand, the completeness of a squared Hessian Riemannian manifold can also be studied. There are some conditions that ensure this property in the *self-concordant case*; see for instance [27]. We present a sort of dual characterization for the case of study of this paper. The next result implies in particular that $T_{\min}$ can be taken as $-\infty$ and $T_{\max}$ as $+\infty$ in Proposition 5.1 and Proposition 5.2.

Corollary 5.2 (Completeness). *Let $\mathbf{\Omega} \subseteq \mathbb{R}^n$ be convex and $\mathbf{g} \in \Gamma_0(\mathbb{R}^n)$ so that (H_0) holds with $\mathbf{g} \in \mathcal{C}^{k+2}(\mathbf{\Omega})$. Then, the squared Hessian Riemannian manifold $(\mathbf{\Omega}, \mathbf{g})$ is geodesically complete and $(\mathbf{\Omega}, \mathbf{dist}_{\mathbf{\Omega}})$ is a complete metric space.*

Proof. Note that (18) provides an explicit formula for the geodesic starting from $x_0 \in \mathbf{\Omega}$ with initial velocity $v_0 = -\nabla^2 \mathbf{g}(x_0)^{-1} d_0$. Therefore, we can check that the geodesic is defined for every $t \in \mathbb{R}$ if and only if $\nabla \mathbf{g}(x_0) - td_0 \in \mathbf{int}(\mathbf{dom}\, \mathbf{g}^*)$ for every $t \in \mathbb{R}$. Since $\mathbf{\Omega}$ is an open set, we have no restriction over the vector d_0, but this is not an issue because $\mathbf{dom}\, \mathbf{g}^* = \mathbb{R}^n$, so the first conclusion follows. The last affirmation is direct from the Hopf-Rinow's Theorem ([24, Theorem 6.13]). $\square$ $\square$

6. Conclusions and outlook

Finally, we would like to stress that, our results only require the existence of a Legendre function **g** for a given convex domain **Ω**. It is suitable then to wonder if a particular choice of **g** provides any advantage over any other choice.

Preliminary results show that the rate of convergence of the method (6) doesn't seem to change when different Legendre functions are considered. However, this fact still needs to be studied in details, as well as a proper comparison with the long list of algorithms designed to solve constrained mathematical programs.

In a similar way, different choices of Legendre functions can yield to different dual Mayer problems (D_c). It is clear that the characterization of the value function defined in (P_c), is not affected by this choice. Nevertheless, a numerical scheme to construct the value function can be.

These are subjects of ongoing works that deserve to be further investigated before presenting it to the scientific community.

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 Optim.*, 43(2):477–501, 2004.

[3] H. Attouch, J. Bolte, P. Redont and M. Teboulle. Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization. *Optimization*, 53(5-6):435–454, 2004.

[4] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM J. 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] O. Bokanowski, N. Forcadel and H. Zidani. Deterministic state-constrained optimal control problems without controllability assumptions. *ESAIM: Control, Optimization and Calculus of Variations*, 17(04):995–1015, 2011.

[8] J. Bolte and M. Teboulle. Barrier operators and associated gradient-like dynamical systems for constrained minimization problems. *SIAM J. Control Optim.*, 42(4):1266–1292, 2003.

[9] J. M. Borwein and J. Vanderwerff. Convex functions of Legendre type in general Banach spaces. *J. Convex Analysis*, 8(2):569–582, 2001.

[10] J. M. Borwein and J. D. Vanderwerff. *Convex functions: constructions, characterizations and counterexamples*. Cambridge University Press, 2010.

[11] U. Boscain and B. Piccoli. *Optimal syntheses for control systems on 2-D manifolds* 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. Chryssochoos and R. Vinter. Optimal control problems on manifolds: a dynamic programming approach. *J. mathematical analysis and applications*, 287(1):118–140, 2003.

[14] F. Clarke, Y. Ledyaev, R. Stern and P. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer-Verlag New York, 1998.

[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] J. Crouzeix. A relationship between the second derivatives of a convex function and of its conjugate. *Mathematical Programming*, 13(1):364–365, 1977.

[17] M. P. do Carmo. *Riemannian Geometry*. Birkhäuser Boston, 1992.

[18] A. V. Fiacco. Perturbed variations of penalty function methods. Example: projective SUMT. *Ann. Oper. Res.*, 27(1-4):371–380, 1990.

[19] H. Frankowska and M. Mazzola. Discontinuous solutions of Hamilton–Jacobi–Bellman equation under state constraints. *Calc. Var. Partial Differ. Equ.*, 46(3-4):725–747, 2013.

[20] H. Frankowska and R. B. Vinter. Existence of neighboring feasible trajectories: Applications to dynamic programming for state-constrained optimal control problems. *J. Optimization Theory and Applications*, 104(1):20–40, 2000.

[21] C. Hermosilla and H. Zidani. Infinite horizon problems on stratifiable state-constraints sets. *J. Differential Equations*, 258(4):1430–1460, 2015.

[22] 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.

[23] R. Klötzler. On a general conception of duality in optimal control. In Equadiff IV. Springer Berlin Heidelberg, 189–196, 1979.

[24] J. M. Lee. *Riemannian manifolds: an introduction to curvature*. Springer-Verlag New York, 1997.

[25] J. M. Lee. *Introduction to smooth manifolds*. Springer-Verlag New York, 2012.

[26] G. P. McCormick. The projective SUMT method for convex programming. *Math. Oper. Res.*, 14(2):203–223, 1989.

[27] 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.

[28] J. Nocedal and S. Wright. *Numerical Optimization*. Springer-Verlag New York, 2000.

[29] E. P. Quiroz and P. R. Oliveira. Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. *J. Convex Analysis*, 16(1):49–69, 2009.

[30] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. Society for Industrial and Applied Mathematics, 2001.

[31] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[32] R. T. Rockafellar. *Conjugate duality and optimization*. Society for Industrial and Applied Mathematics, 1974.

[33] S. T. Smith. Optimization techniques on Riemannian manifolds. *Fields institute communications*, 3(3):113–135, 1994.

[34] H. Soner. Optimal control with state-space constraint I. *SIAM J. Control Optim.*, 24(3):552–561, 1986.

[35] C. Udriste. *Convex Function and Optimization Methods on Riemannian Manifolds*. Springer Netherlands, 1994.

[36] R. J. Vanderbei. Linear programming. *Foundations and extensions, International Series in Operations Research & Management Science*, 37, 2001.

[37] R. Vinter. *Optimal control.* Springer, 2010.

[38] R. Vinter. Convex duality and nonlinear optimal control. *SIAM J. Control Optim.*, 31(2): 518–538, 1993.

[39] Y. Yang. Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *J. Optimization Theory and Applications*, 132(2):245–265, 2007.

E-mail address: `chermosilla@lsu.edu`