Gradient flows associated with Hessian Riemannian metrics induced by Legendre functions in constrained optimization

Felipe Alvarez* Jérôme Bolte† Olivier Brahic.‡

Abstract Motivated by a constrained minimization problem, it is studied the gradient flows with respect to Hessian Riemannian metrics induced by convex functions of Legendre type. The first result characterizes Hessian Riemannian structures on convex sets as those metrics that have a specific integration property with respect to variational inequalities, giving a new motivation for the introduction of Bregman-type distances. Then, the general evolution problem is introduced and a differential inclusion reformulation is given. A general existence result is proved and global convergence is established under quasi-convexity conditions, with interesting refinements in the case of convex minimization. Some explicit examples of these gradient flows are discussed. Dual trajectories are identified and sufficient conditions for dual convergence are examined for a convex program with positivity and equality constraints. Some convergence rate results are established. In the case of a linear objective function, several optimality characterizations of the orbits are given: optimal path of viscosity methods, continuous-time model of Bregman-type proximal algorithms, geodesics for some adequate metrics and projections of $\dot{q}$-trajectories of some Lagrange equations and completely integrable Hamiltonian systems.

Keywords Gradient flow, Hessian Riemannian metric, Legendre type convex function, existence, global convergence, Bregman distance, Liapounov functional, quasi-convex minimization, convex and linear programming, Legendre transform coordinates, Lagrange and Hamilton equations.

AMS classification: 34G20, 34A12, 34D05, 90C25.

*Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático (CNRS UMR 2071), Universidad de Chile, Blanco Encalada 2120, Santiago, Chile. `falvarez@dim.uchile.cl`. Supported by Fondecyt 1020610, Fondap en Matemáticas Aplicadas and Programa Iniciativa Científica Milenio.

†ACSIOM-CNRS FRE 2311, Département de Mathématiques, case 51, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 5, France. Partially supported by Ecos-Conicyt C00E05.

‡GTA-CNRS UMR 5030, Département de Mathématiques, case 51, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 5, France.

1 Introduction

The aim of this paper is to study the existence, global convergence and geometric properties of gradient flows with respect to a specific class of Hessian Riemannian metrics on convex sets. Our work is indeed deeply related to the constrained minimization problem

$$(P) \qquad \min\{f(x) \mid x \in \overline{C},\ Ax = b\},$$

where $\overline{C}$ is the closure of a nonempty, *open* and convex subset C of $\mathbb{R}^n$, A is a $m < n$ real matrix with $m \leq n$, $b \in \mathbb{R}^m$ and $f \in C^1(\mathbb{R}^n)$. A strategy to solve (P) consists in endowing C with a Riemannian structure $(\cdot,\cdot)^H$, to restrict it to the relative interior of the feasible set $\mathcal{F} := C \cap \{x \mid Ax = b\}$, and then to consider the trajectories generated by the steepest descent vector field $-\nabla_H f_{|\mathcal{F}}$. This leads to the initial value problem

$$(H\text{-}SD) \qquad \dot{x}(t) + \nabla_H f_{|\mathcal{F}}(x(t)) = 0,\ x(0) \in \mathcal{F},$$

where $(H\text{-}SD)$ stands for H-steepest descent. We focus on those metrics that are induced by the Hessian $H = \nabla^2 h$ of a *Legendre type* convex function h defined on C (cf. Def. 3.1).

The use of Riemannian methods in optimization has increased recently: in relation with Karmarkar algorithm and linear programming see Karmarkar [29], Bayer-Lagarias [5]; for continuous-time models of proximal type algorithms and related topics see Iusem-Svaiter-Da Cruz [27], Bolte-Teboulle [6]. For a systematic dynamical system approach to constrained optimization based on double bracket flows, see Brockett [8, 9], the monograph of Helmke-Moore [22] and the references therein. On the other hand, the structure of $(H\text{-}SD)$ is also at the heart of some important problems in applied mathematics. For connections with population dynamics and game theory see Hofbauer-Sygmund [25], Akin [1], Attouch-Teboulle [3]. We will see that $(H\text{-}SD)$ can be reformulated as the differential inclusion $\frac{d}{dt}\nabla h(x(t)) + \nabla f(x(t)) \in \operatorname{Im} A^T$, $x(t) \in \mathcal{F}$, which is formally similar to some evolution problems in infinite dimensional spaces arising in thermodynamical systems, see for instance Kenmochi-Pawlow [30] and references therein.

A classical approach in the asymptotic analysis of dynamical systems consists in exhibiting attractors of the orbits by using Liapounov functionals. Our choice of Hessian Riemannian metrics is based on this idea. In fact, we consider first the important case where f is convex, a condition that permits us to reformulate (P) as a variational inequality problem: find $a \in \overline{\mathcal{F}}$ such that $(\nabla_H f_{|\mathcal{F}}(x), x - a)^H_x \geq 0$ for all x in $\mathcal{F}$. In order to identify a suitable Liapounov functional, this variational problem is met through the following integration problem: *find the metrics* $(\cdot,\cdot)^H$ *for which the vector fields* $V^a : \mathcal{F} \to \mathbb{R}^n$, $a \in \mathcal{F}$, *defined by* $V^a(x) = x - a$, *are* $(\cdot,\cdot)^H$*-gradient vector fields.* Our first result (cf. Theorem 3.1) establishes that such metrics are given by the Hessian of strictly convex functions, and in that case the vector fields V^a appear as gradients with respect to the second variable of some distance-like functions that are called D-functions. Indeed, if $(\cdot,\cdot)^H$ is induced by the Hessian $H = \nabla^2 h$ of $h : \mathcal{F} \mapsto \mathbb{R}$, we have for all a, x in $\mathcal{F}$: $\nabla_H D_h(a,.)(x) = x - a$, where $D_h(a,x) = h(a) - h(x) - dh(x)(x-a)$. For another characterization of Hessian metrics, see Duistermaat [17].

Motivated by the previous result and with the aim of solving (P), we are then naturally led to consider Hessian Riemannian metrics that cannot be smoothly extended out of $\mathcal{F}$. Such a requirement is fulfilled by the Hessian of a *Legendre (convex) function* h, whose definition is recalled in section 3. We give then a differential inclusion reformulation of $(H\text{-}SD)$, which permits to show that in the case of a linear objective function f, the flow of $-\nabla_H f_{|\mathcal{F}}$ stands at the crossroad of many optimization methods. In fact, following [27], we prove that viscosity methods and Bregman proximal algorithms produce their paths or iterates in the orbit of $(H\text{-}SD)$. The D-function of h plays an essential role for this. In section 4.4 it is given a systematic method to construct Legendre functions based on barrier functions for convex inequality problems, which is illustrated with some examples; relations to other works are discussed.

Section 4 deals with global existence and convergence properties. After having given a non trivial well-posedness result (cf. Theorem 4.1), we prove in section 4.2 that $f(x(t)) \to \inf_{\overline{\mathcal{F}}} f$ as $t \to +\infty$ whenever f is convex. A natural problem that arises is the trajectory convergence to a critical point. Since one expects the limit to be a (local) solution to (P), which may belong to the boundary of C, the notion of critical point must be understood in the sense of the optimality condition for a local minimizer a of f over $\overline{\mathcal{F}}$:

$$(\mathcal{O}) \qquad \nabla f(a) + N_{\overline{\mathcal{F}}}(a) \ni 0,\ a \in \overline{\mathcal{F}},$$

where $N_{\overline{\mathcal{F}}}(a)$ is the normal cone to $\overline{\mathcal{F}}$ at a, and ∇f is the Euclidean gradient of f. This involves an asymptotic singular behavior that is rather unusual in the classical theory of dynamical systems, where the critical points are typically supposed to be in the manifold. In section 4.3 we assume that the Legendre type function h is a *Bregman function with zone* C and prove that under a quasi-convexity assumption on f, the trajectory converges to some point a satisfying $(\mathcal{O})$. When f is convex, the preceding result amounts to the convergence of $x(t)$ toward a global minimizer of f over $\overline{\mathcal{F}}$. We also give a variational characterization of the limit and establish an abstract result on the rate of convergence under uniqueness of the solution. We consider in section 4.5 the case of linear programming, for which asymptotic convergence as well as a variational characterization are proved without the Bregman-type condition. Within this framework, we also give some estimates on the convergence rate that are valid for the specific Legendre functions commonly used in practice. In section 4.6, we consider the interesting case of positivity and equality constraints, introducing a *dual* trajectory $\lambda(t)$ that, under some appropriate conditions, converges to a solution to the dual problem of (P) whenever f is convex, even if primal convergence is not ensured.

Finally, inspired by the seminal work [5], we define in section 5 a change of coordinates called *Legendre transform coordinates*, which permits to show that the orbits of $(H\text{-}SD)$ may be seen as straight lines in a positive cone. This leads to additional geometric interpretations of the flow of $-\nabla_H f_{|\mathcal{F}}$. On the one hand, the orbits are geodesics with respect to an appropriate metric and, on the other hand, they may be seen as $\dot{q}$-trajectories of some Lagrangian, with consequences in terms of integrable Hamiltonians.

Notations. $\operatorname{Ker} A = \{x \in \mathbb{R}^n \mid Ax = 0\}$. The orthogonal complement of $\mathcal{A}_0$ is denoted by $\mathcal{A}_0^\perp$, and $\langle \cdot, \cdot \rangle$ is the standard Euclidean scalar product of $\mathbb{R}^n$. Let us denote by $\mathbb{S}^n_{++}$ the cone of real symmetric definite positive matrices. Let $\Omega \subset \mathbb{R}^n$ be an open set. If $f : \Omega \to \mathbb{R}$ is differentiable then ∇f stands for the Euclidean gradient of f. If $h : \Omega \mapsto \mathbb{R}$ is twice differentiable then its Euclidean Hessian at $x \in \Omega$ is denoted by $\nabla^2 h(x)$ and is defined as the endomorphism of $\mathbb{R}^n$ whose matrix in canonical coordinates is given by $[\frac{\partial^2 h(x)}{\partial x_i \partial x_j}]_{i,j\in\{1,..,n\}}$. Thus, $\forall x \in \Omega$, $d^2h(x) = \langle \nabla^2 h(x) \cdot, \cdot \rangle$.

2 Preliminaries

2.1 The minimization problem and optimality conditions

Given a positive integer $m < n$, a full rank matrix $A \in \mathbb{R}^{m\times n}$ and $b \in \mathrm{Im}\, A$, let us define

$$\mathcal{A} = \{x \in \mathbb{R}^n \mid Ax = b\}. \tag{1}$$

Set $\mathcal{A}_0 = \mathcal{A} - \mathcal{A} = \mathrm{Ker}\, A$. Of course, $\mathcal{A}_0^\perp = \mathrm{Im}\, A^T$ where A^T is the transpose of A. Let C be a nonempty, open and convex subset of $\mathbb{R}^n$, and $f : \mathbb{R}^n \to \mathbb{R}$ a $\mathcal{C}^1$ function. Consider the constrained minimization problem

$$(P) \qquad \inf\{f(x) \mid x \in \overline{C},\ Ax = b\}.$$

The set of optimal solutions of (P) is denoted by $S(P)$. We call f the *objective function* of (P). The *feasible set* of (P) is given by $\overline{\mathcal{F}} = \{x \in \mathbb{R}^n \mid x \in \overline{C},\ Ax = b\} = \overline{C} \cap \mathcal{A}$, and $\mathcal{F}$ stands for the *relative interior* of $\overline{\mathcal{F}}$, that is

$$\mathcal{F} = \mathrm{ri}\, \overline{\mathcal{F}} = \{x \in \mathbb{R}^n \mid x \in C,\ Ax = b\} = C \cap \mathcal{A}. \tag{2}$$

Throughout this article, we assume that

$$\mathcal{F} \neq \emptyset. \tag{3}$$

It is well known that a necessary condition for a to be locally minimal for f over $\overline{\mathcal{F}}$ is $(\mathcal{O}) :\ -\nabla f(a) \in N_{\overline{\mathcal{F}}}(a)$, where $N_{\overline{\mathcal{F}}}(x) = \{\nu \in \mathbb{R}^n \mid \forall y \in \overline{\mathcal{F}},\ \langle y - x, \nu\rangle \leq 0\}$ is the *normal cone* to $\overline{\mathcal{F}}$ at $x \in \overline{\mathcal{F}}$ ($N_{\overline{\mathcal{F}}}(x) = \emptyset$ when $x \notin \overline{\mathcal{F}}$); see for instance [37, Theorem 6.12]. By [36, Corollary 23.8.1], $N_{\overline{\mathcal{F}}}(x) = N_{\overline{C}\cap\mathcal{A}}(x) = N_{\overline{C}}(x) + N_{\mathcal{A}}(x) = N_{\overline{C}}(x) + \mathcal{A}_0^\perp$, for all $x \in \overline{\mathcal{F}}$. Therefore, the necessary optimality condition for $a \in \overline{\mathcal{F}}$ is

$$-\nabla f(a) \in N_{\overline{C}}(a) + \mathcal{A}_0^\perp. \tag{4}$$

If f is convex then this condition is also sufficient for $a \in \overline{\mathcal{F}}$ to be in $S(P)$.

2.2 Riemannian gradient flows on the relative interior of the feasible set

Let M be a smooth manifold. The tangent space to M at $x \in M$ is denoted by T_xM. If $f : M \mapsto \mathbb{R}$ is a $\mathcal{C}^1$ function then $df(x)$ denotes its differential or tangent map $df(x) :$

$T_xM \to \mathbb{R}$ at $x \in M$. A $\mathcal{C}^k$ metric on M, $k \geq 0$, is a family of scalar products $(\cdot,\cdot)_x$ on each T_xM, $x \in M$, such that $(\cdot,\cdot)_x$ depends in a $\mathcal{C}^k$ way on x. The couple $M, (\cdot,\cdot)_x$ is called a $\mathcal{C}^k$ Riemannian manifold. This structure permits to identify T_xM with its dual, i.e. the cotangent space T_xM^*, and thus to define a notion of gradient vector. Indeed, given f in M, the gradient of f is denoted by $\nabla_{(\cdot,\cdot)} f$ and is uniquely determined by the following conditions:

(g$_1$) tangency condition: for all $x \in M$, $\nabla_{(\cdot,\cdot)} f(x) \in T_xM^* \simeq T_xM$,

(g$_2$) dualility condition: for all $x \in M$, $v \in T_xM$, $df(x)(v) = (\nabla_{(\cdot,\cdot)} f(x), v)_x$.

We refer the reader to [16, 33] for further details.

Let us return to the minimization problem (P). Since C is open, we can take $M = C$ with the usual identification $T_xC \simeq \mathbb{R}^n$ for every $x \in C$. Given a continuous mapping $H : C \to \mathbb{S}^n_{++}$, the metric defined by

$$\forall x \in C,\ \forall u, v \in \mathbb{R}^n,\ (u,v)^H_x = \langle H(x)u, v\rangle, \tag{5}$$

endows C with a $\mathcal{C}^0$ Riemannian structure. The corresponding Riemannian gradient vector field of the objective function f restricted to C, which we denote by $\nabla_H f_{|C}$, is given by

$$\nabla_H f_{|C}(x) = H(x)^{-1}\nabla f(x). \tag{6}$$

Next, take $N = \mathcal{F} = C \cap \mathcal{A}$, which is a smooth submanifold of C with $T_x\mathcal{F} \simeq \mathcal{A}_0$ for each $x \in \mathcal{F}$. Definition (5) induces a metric on $\mathcal{F}$ for which the gradient of the restriction $f_{|\mathcal{F}}$ is denoted by $\nabla_H f_{|\mathcal{F}}$. Conditions (g_1) and (g_2) imply that for all $x \in \mathcal{F}$

$$\nabla_H f_{|\mathcal{F}}(x) = P_x H(x)^{-1}\nabla f(x), \tag{7}$$

where, given $x \in C$, $P_x : \mathbb{R}^n \to \mathcal{A}_0$ is the $(\cdot,\cdot)^H_x$-orthogonal projection onto the linear subspace $\mathcal{A}_0$. Since A has full rank, it is easy to see that

$$P_x = I - H(x)^{-1}A^T(AH(x)^{-1}A^T)^{-1}A, \tag{8}$$

and we conclude that for all $x \in \mathcal{F}$

$$\nabla_H f_{|\mathcal{F}}(x) = H(x)^{-1}[I - A^T(AH(x)^{-1}A^T)^{-1}AH(x)^{-1}]\nabla f(x). \tag{9}$$

Given $x \in \mathcal{F}$, the vector $-\nabla_H f_{|\mathcal{F}}(x)$ can be interpreted as that direction in $\mathcal{A}_0$ such that f decreases the most steeply at x with respect to the metric $(\cdot,\cdot)^H_x$. The *steepest descent method* for the (local) minimization of f on the Riemannian manifold $\mathcal{F}, (\cdot,\cdot)^H_x$ consists in finding the solution trajectory $x(t)$ of the vector field $-\nabla_H f_{|\mathcal{F}}$ with initial condition $x^0 \in \mathcal{F}$:

$$\begin{cases} \dot{x} + \nabla_H f_{|\mathcal{F}}(x) = 0, \\ x(0) = x^0 \in \mathcal{F}. \end{cases} \tag{10}$$

3 Legendre gradient flows in constrained optimization

3.1 Liapounov functionals, variational inequalities and Hessian metrics

This section is intended to motivate the particular class of Riemannian metrics that is studied in this paper in view of the asymptotic convergence of the solution to (10).

Let us consider the minimization problem (P) and assume that C is endowed with some Riemannian metric $(\cdot,\cdot)^H_x$ as defined in (5). Recall that $V : \mathcal{F} \mapsto \mathbb{R}$ is a *Liapounov functional* for the vector field $-\nabla_H f_{|\mathcal{F}}$ if $\forall x \in \mathcal{F}$, $(-\nabla_H f_{|\mathcal{F}}(x), \nabla_H V(x))^H_x \leq 0$. If $x(t)$ is a solution to (10), this implies that $t \mapsto V(x(t))$ is nonincreasing. Although $f_{|\mathcal{F}}$ is indeed a Liapounov functional for $-\nabla_H f_{|\mathcal{F}}$, this does not ensure the convergence of $x(t)$ (see for instance the counterexample of Palis-De Melo [35] in the Euclidean case).

Suppose that the objective function f is convex. For simplicity, we also assume that $A = 0$ so that $\mathcal{F} = C$. In the framework of convex minimization, the set of minimizers of f over $\overline{C}$, denoted by $\operatorname{Argmin}_{\overline{C}} f$, is characterized in variational terms as follows:

$$a \in \operatorname{Argmin}_{\overline{C}} f \quad \Leftrightarrow \quad \forall x \in \overline{C},\ \langle \nabla f(x), x - a\rangle \geq 0. \tag{11}$$

Setting $q_a(x) = \frac{1}{2}|x-a|^2$ for all $a \in \operatorname{Argmin}_{\overline{C}}$, one observes that $\nabla q_a(x) = x - a$ and thus, by (11), q_a is a Liapounov functional for $-\nabla f$. This key property allows one to establish the asymptotic convergence as $t \to +\infty$ of the corresponding steepest descent trajectories; see [10] for more details in a very general non-smooth setting. To use the same kind of arguments in a non Euclidean context, observe that by (6) together with the continuity of ∇f, the following variational Riemannian characterization holds

$$a \in \operatorname{Argmin}_{\overline{C}} f \quad \Leftrightarrow \quad \forall x \in C,\ (\nabla_H f(x), x - a)^H_x \geq 0. \tag{12}$$

We are thus naturally led to the problem of *finding the Riemannian metrics on C for which the mappings $C \ni x \mapsto x - y \in \mathbb{R}^n$, $y \in C$, are gradient vector fields.* The next result gives a characterization of such metrics: they are induced by Hessian of strictly convex functions.

Theorem 3.1. *Assume that $H \in \mathcal{C}^1(C; \mathbb{S}^n_{++})$, or in other words that $(\cdot,\cdot)^H_x$ is a $\mathcal{C}^1$ metric. The family of vector fields $\{V^y : C \ni x \mapsto x - y \in \mathbb{R}^n\}$, $y \in C$ is a family of $(\cdot,\cdot)^H$-gradient vector fields if and only if there exists a strictly convex function $h \in \mathcal{C}^3(C)$ such that $\forall x \in C$, $H(x) = \nabla^2 h(x)$. Besides, defining $D_h : C \times C \mapsto \mathbb{R}$ by*

$$D_h(y,x) = h(y) - h(x) - \langle \nabla h(x), x - y\rangle, \tag{13}$$

we obtain $\nabla_H D_h(y,\cdot)(x) = x - y$.

Proof. The set of metrics complying with the "gradient" requirement is denoted by $\mathcal{M}$, that is, $(\cdot,\cdot)^H_x \in \mathcal{M} \Leftrightarrow H \in \mathcal{C}^1(C; \mathbb{S}^n_{++})$ and $\forall y \in C$, $\exists \varphi_y \in \mathcal{C}^1(C; \mathbb{R})$, $\nabla_H \varphi_y(x) = x - y$.. Let $(x_1, .., x_n)$ denote the canonical coordinates of $\mathbb{R}^n$ and write $\sum_{i,j} H_{ij}(x) dx_i dx_j$ for $(\cdot,\cdot)^H_x$. By (6), the mappings $x \mapsto x - y$, $y \in C$, define a family of $(\cdot,\cdot)^H_x$ gradients iff k_y :

$x \mapsto H(x)(x-y)$, $y \in C$, is a family of Euclidean gradients. Setting $\alpha^y(x) = \langle k_y(x), \cdot\rangle$, $x, y \in C$, the problem amounts to find necessary (and sufficient) conditions under which the 1-forms α^y are all exact. Let $y \in C$. Since C is convex, the Poincaré lemma [33, Theorem V.4.1] states that α^y is exact iff it is closed. In canonical coordinates we have $\alpha^y(x) = \sum_i \left(\sum_k H_{ik}(x)(x_k - y_k)\right) dx_i,\ x \in C$, and therefore α^y is exact iff for all $i, j \in \{1, .., n\}$ we have $\frac{\partial}{\partial x_j}\sum_k H_{ik}(x)(x_k - y_k) = \frac{\partial}{\partial x_i}\sum_k H_{jk}(x)(x_k - y_k)$, which is equivalent to $\sum_k \frac{\partial}{\partial x_j} H_{ik}(x)(x_k - y_k) + H_{ij}(x) = \sum_k \frac{\partial}{\partial x_i} H_{jk}(x)(x_k - y_k) + H_{ji}(x)$. Since $H_{ij}(x) = H_{ji}(x)$, this gives the following condition: $\sum_k \frac{\partial}{\partial x_j} H_{ik}(x)(x_k - y_k) = \sum_k \frac{\partial}{\partial x_i} H_{jk}(x)(x_k - y_k)$, $\forall i, j \in \{1, .., n\}$. If we set $V_x = (\frac{\partial}{\partial x_j} H_{i1}(x), .., \frac{\partial}{\partial x_j} H_{in}(x))^T$ and $W_x = (\frac{\partial}{\partial x_i} H_{j1}(x), .., \frac{\partial}{\partial x_i} H_{jn}(x))^T$, the latter can be rewritten $\langle V_x - W_x, x - y\rangle = 0$, which must hold for all $(x, y) \in C \times C$. Fix $x \in C$. Let $\epsilon_x > 0$ be such that the open ball of center x with radius ϵ_x is contained in C. For every ν such that $|\nu| = 1$, take $y = x + \epsilon_x/2\nu$ to obtain that $\langle V_x - W_x, \nu\rangle = 0$. Consequently, $V_x = W_x$ for all $x \in C$. Therefore, $(\cdot, \cdot)^H_x \in \mathcal{M}$ iff

$$\forall x \in C,\ \forall i, j, k \in \{1, .., n\},\ \frac{\partial}{\partial x_i} H_{jk}(x) = \frac{\partial}{\partial x_j} H_{ik}(x). \tag{14}$$

Lemma 3.1. *If $H : C \mapsto \mathbb{S}^n_{++}$ is a differentiable mapping satisfying* (14)*, then there exists $h \in \mathcal{C}^3(C)$ such that $\forall x \in C$, $H(x) = \nabla^2 h(x)$. In particular, h is strictly convex.*

of Lemma 3.1. For all $i \in \{1, .., n\}$, set $\beta^i = \sum_k H_{ik} dx_k$. By (14), β^i is closed and therefore exact. Let $\phi_i : C \mapsto \mathbb{R}$ be such that $d\phi_i = \beta^i$ on C, and set $\omega = \sum_k \phi_k dx_k$. We have that $\frac{\partial}{\partial x_j}\phi_i(x) = H_{ij}(x) = H_{ji}(x) = \frac{\partial}{\partial x_i}\phi_j(x)$, $\forall x \in C$. This proves that ω is closed, and therefore there exists $h \in \mathcal{C}^2(C, \mathbb{R})$ such that $dh = \omega$. To conclude we just have to notice that $\frac{\partial}{\partial x_i} h(x) = \phi_i$, and thus $\frac{\partial^2 h}{\partial x_j \partial x_i}(x) = H_{ji}(x)$, $\forall x \in C$. □

To finish the proof, remark that taking $\varphi_y = D_h(y, \cdot)$ with D_h being defined by (13), we obtain $\nabla \varphi_y(x) = \nabla^2 h(x)(x - y)$, and therefore $\nabla_H \varphi_y(x) = x - y$ in virtue of (6). □

Remark 3.1. (a) In the theory of Bregman proximal methods for convex optimization, the distance-like function D_h defined by (13) is called the *D-function* of h. Theorem 3.1 is a new and surprising motivation for the introduction of D_h in relation with variational inequality problems. (b) For a geometrical approach to Hessian Riemannian structures the reader is referred to the recent work of Duistermaat [17].

Theorem 3.1 suggests to endow C with a Riemannian structure associated with the Hessian $H = \nabla^2 h$ of a strictly convex function $h : C \mapsto \mathbb{R}$. As we will see under some additional conditions, the D-function of h is essential to establish the asymptotic convergence of the trajectory. On the other hand, if it is possible to replace h by a sufficiently smooth strictly convex function $h' : C' \mapsto \mathbb{R}$ with $C' \supset\supset C$ and $h'_{|C} = h$, then the gradient flows for h and h' are the same on C but the steepest descent trajectories associated with the latter may leave the feasible set of (P) and in general they will not converge to a solution of (P). We shall see that to avoid this drawback it is sufficient to require that $|\nabla h(x^j)| \to +\infty$ for all sequences (x^j) in C converging to a boundary point of C. This may be interpreted as a sort of *barrier technique*, a classical strategy to enforce feasibility in optimization theory.

3.2 Legendre type functions and the $(H\text{-}SD)$ dynamical system

In the sequel, we adopt the standard notations of convex analysis theory; see [36]. Given a closed convex subset S of $\mathbb{R}^n$, we say that an extended-real-valued function $g : S \mapsto \mathbb{R} \cup \{+\infty\}$ belongs to the class $\Gamma_0(S)$ when g is lower semicontinuous, proper ($g \not\equiv +\infty$) and convex. For such a function $g \in \Gamma_0(S)$, its *effective domain* is defined by $\operatorname{dom} g = \{x \in S \mid g(x) < +\infty\}$. When $g \in \Gamma_0(\mathbb{R}^n)$ its *Legendre-Fenchel conjugate* is given by $g^*(y) = \sup\{\langle x, y\rangle - g(x) \mid x \in \mathbb{R}^n\}$, and its *subdifferential* is the set-valued mapping $\partial g : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ given by $\partial g(x) = \{y \in \mathbb{R}^n \mid \forall z \in \mathbb{R}^n,\ f(x) + \langle y, z - x\rangle \leq f(z)\}$. We set $\operatorname{dom} \partial g = \{x \in \mathbb{R}^n \mid \partial g(x) \neq \emptyset\}$.

Definition 3.1. [36, Chapter 26] *A function $h \in \Gamma_0(\mathbb{R}^n)$ is called:*
(i) essentially smooth*, if h is differentiable on* $\operatorname{int} \operatorname{dom} h$*, with moreover $|\nabla h(x^j)| \to +\infty$ for every sequence $(x^j) \subset$* $\operatorname{int}\operatorname{dom} h$ *converging to a boundary point of* $\operatorname{dom} h$ *as $j \to +\infty$;*
(ii) of Legendre type *if h is essentially smooth and strictly convex on* $\operatorname{int}\operatorname{dom} h$*.*

Remark that by [36, Theorem 26.1], $h \in \Gamma_0(\mathbb{R}^n)$ is essentially smooth iff $\partial h(x) = \{\nabla h(x)\}$ if $x \in \operatorname{int}\operatorname{dom} h$ and $\partial h(x) = \emptyset$ otherwise; in particular, $\operatorname{dom} \partial h = \operatorname{int}\operatorname{dom} h$.

Motivated by the results of section 3.1, we define a Riemannian structure on C by introducing a function $h \in \Gamma_0(\mathbb{R}^n)$ such that:

$$(H_0) \qquad \begin{cases} \text{(i)} & h \text{ is of Legendre type with } \operatorname{int}\operatorname{dom} h = C. \\ \text{(ii)} & h_{|C} \in \mathcal{C}^2(C;\mathbb{R}) \text{ and } \forall x \in C, \nabla^2 h(x) \in \mathbb{S}^n_{++}. \\ \text{(iii)} & \text{The mapping } C \ni x \mapsto \nabla^2 h(x) \text{ is locally Lipschitz continuous.} \end{cases}$$

Here and subsequently, we take $H = \nabla^2 h$ with h satisfying (H_0). The Hessian mapping $C \ni x \mapsto H(x)$ endows C with the (locally Lipschitz continuous) Riemannian metric

$$\forall x \in C,\ \forall u, v \in \mathbb{R}^n,\ (u, v)^H_x = \langle H(x)u, v\rangle = \langle \nabla^2 h(x)u, v\rangle, \tag{15}$$

and we say that $(\cdot,\cdot)^H_x$ is the *Legendre metric* on C induced by the Legendre type function h, which also defines a metric on $\mathcal{F} = C \cap \mathcal{A}$ by restriction. In addition to $f \in \mathcal{C}^1(\mathbb{R}^n)$, we suppose that the objective function satisfies

$$\nabla f \text{ is locally Lipschitz continuous on } \mathbb{R}^n. \tag{16}$$

The corresponding steepest descent method in the manifold $\mathcal{F}, (\cdot,\cdot)^H_x$, which we refer to as $(H\text{-}SD)$ for short, is then the following continuous dynamical system

$$(H\text{-}SD) \qquad \begin{cases} \dot{x}(t) + \nabla_H f_{|\mathcal{F}}(x(t)) = 0,\ t \in (T_m, T_M), \\ x(0) = x^0 \in \mathcal{F}, \end{cases}$$

with $H = \nabla^2 h$ and where $-\infty \leq T_m < 0 < T_M \leq +\infty$ define the interval corresponding to the unique maximal solution of $(H\text{-}SD)$. Given an initial condition $x^0 \in \mathcal{F}$, we shall say that $(H\text{-}SD)$ is *well-posed* when its maximal solution satisfies $T_M = +\infty$. In section 4.1 we will give some sufficient conditions ensuring the well-posedness of $(H\text{-}SD)$.

3.3 Differential inclusion formulation of $(H\text{-}SD)$ and some consequences

It is easily seen that the solution $x(t)$ of $(H\text{-}SD)$ satisfies:

$$\begin{cases} \dfrac{d}{dt}\nabla h(x(t)) + \nabla f(x(t)) & \in & \mathcal{A}_0^{\perp} \text{ on } (T_m, T_M), \\ x(t) & \in & \mathcal{F} \text{ on } (T_m, T_M), \\ x(0) & = & x^0 \in \mathcal{F}. \end{cases} \tag{17}$$

This differential inclusion problem makes sense even when $x \in W^{1,1}_{loc}(T_m, T_M; \mathbb{R}^n)$, the inclusions being satisfied almost everywhere on (T_m, T_M). Actually, the following result establishes that $(H\text{-}SD)$ and (17) describe the same trajectory.

Proposition 3.1. *Let $x \in W^{1,1}_{loc}(T_m, T_M; \mathbb{R}^n)$. Then, x is a solution of* (17) *iff x is the solution of* $(H\text{-}SD)$. *In particular,* (17) *admits a unique solution of class $\mathcal{C}^1$.*

Proof. Assume that x is a solution of (17), and let I' be the subset of (T_m, T_M) on which $t \mapsto (x(t), \nabla h(x(t))$ is derivable. We may assume that $x(t) \in \mathcal{F}$ and $\frac{d}{dt}\nabla h(x(t)) + \nabla f(x(t)) \in \mathcal{A}_0^{\perp}$, $\forall t \in I'$. Since x is absolutely continuous, $\dot{x}(t) + H(x(t))^{-1}\nabla f((x(t)) \in H(x(t))^{-1}\mathcal{A}_0^{\perp}$ and $\dot{x}(t) \in \mathcal{A}_0$, $\forall t \in I'$. But the orthogonal complement of $\mathcal{A}_0$ with respect to the inner product $\langle H(x)\cdot, \cdot\rangle$ is exactly $H(x)^{-1}\mathcal{A}_0^{\perp}$ when $x \in \mathcal{F}$. It follows that $\dot{x} + P_x H(x)^{-1}\nabla f(x) = 0$ on I'. This implies that x is the $\mathcal{C}^1$ solution of $(H\text{-}SD)$. □

Suppose that f is convex. On account of Proposition 3.1, $(H\text{-}SD)$ can be interpreted as a continuous-time model for a well-known class of iterative minimization algorithms. In fact, an implicit discretization of (17) yields the following iterative scheme: $\nabla h(x^{k+1}) - \nabla h(x^k) + \mu_k \nabla f(x^{k+1}) \in \text{Im } A^T$, $Ax^{k+1} = b$, where $\mu_k > 0$ is a step-size parameter and $x^0 \in \mathcal{F}$. This is the optimality condition for

$$x^{k+1} \in \text{Argmin}\left\{ f(x) + 1/\mu_k D_h(x, x^k) \mid Ax = b \right\}, \tag{18}$$

where D_h is given by

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y\rangle, \; x \in \text{dom}\, h, \; y \in \text{dom}\, \partial h = C. \tag{19}$$

The above algorithm is accordingly called the *Bregman proximal minimization* method; for an insight of its importance in optimization see for instance [12, 13, 26, 32].

Next, assume that $f(x) = \langle c, x\rangle$ for some $c \in \mathbb{R}^n$. As already noticed in [5, 21, 34] for the log-metric and in [27] for a fairly general h, in this case the $(H\text{-}SD)$ gradient trajectory can be viewed as a *central optimal path*. Indeed, integrating (17) over $[0, t]$ we obtain $\nabla h(x(t)) - \nabla h(x^0) + tc \in \mathcal{A}_0^{\perp}$. Since $x(t) \in \mathcal{A}$, it follows that

$$x(t) \in \text{Argmin}\left\{ \langle c, x\rangle + 1/t D_h(x, x^0) \mid Ax = b \right\}, \tag{20}$$

which corresponds to the so-called *viscosity method* relative to $g(x) = D_h(x, x^0)$; see [2, 4, 27] and Corollary 4.1. Remark now that for a linear objective function, (18) and (20) are essentially the same: the sequence generated by the former belongs to the optimal path defined by the latter. Indeed, setting $t_0 = 0$ and $t_{k+1} = t_k + \mu_k$ for all $k \geq 0$ ($\mu_0 = 0$) and integrating (17) over $[t_k, t_{k+1}]$, we obtain that $x(t_{k+1})$ satisfies the optimality condition for (18). The following result summarizes the previous discussion.

Proposition 3.2. *Assume that f is linear and that the corresponding $(H\text{-}SD)$ dynamical system is well-posed. Then, the viscosity optimal path $\widetilde{x}(\varepsilon)$ relative to $g(x) = D_h(x, x^0)$ and the sequence (x^k) generated by* (18) *exist and are unique, with in addition $\widetilde{x}(\varepsilon) = x(1/\varepsilon)$, $\forall \varepsilon > 0$, and $x^k = x(\sum_{l=0}^{k-1} \mu_l)$, $\forall\ k \geq 1$, where $x(t)$ is the solution of $(H\text{-}SD)$.*

Remark 3.2. In order to ensure asymptotic convergence for proximal-type algorithms, it is usually required that the step-size parameters satisfy $\sum \mu_k = +\infty$. By Proposition 3.2, this is necessary for the convergence of (18) in the sense that when $(H\text{-}SD)$ is well-posed, if x^k converges to some $x^* \in S(P)$ then either $x^0 = x^*$ or $\sum \mu_k = +\infty$.

4 Global existence, asymptotic analysis and examples

4.1 Well-posedness of $(H\text{-}SD)$

In this section we establish the well-posedness of $(H\text{-}SD)$ (i.e. $T_M = +\infty$) under three different conditions. In order to avoid any confusion, we say that a set $E \subset \mathbb{R}^n$ is *bounded* when it is so for the usual Euclidean norm $|y| = \sqrt{\langle y, y \rangle}$. First, we propose the condition:

(WP_1) The lower level set $\{y \in \overline{\mathcal{F}} \mid f(y) \leq f(x^0)\}$ is bounded.

Notice that (WP_1) is weaker than the classical assumption imposing f to have bounded lower level sets in the H metric sense. Next, let D_h be the D-function of h that is defined by (19) and consider the following condition:

$$(WP_2)\quad \begin{cases} \text{(i) dom } h = \overline{C} \text{ and } \forall a \in \overline{C},\ \forall \gamma \in \mathbb{R},\ \{y \in \mathcal{F} \mid D_h(a, y) \leq \gamma\} \text{ is bounded.} \\ \text{(ii) } S(P) \neq \emptyset \text{ and } f \text{ is quasi-convex (i.e. the lower level sets of } f \text{ are convex).} \end{cases}$$

When $\overline{\mathcal{F}}$ is unbounded (WP_1) and (WP_2) involve some a priori properties on f. This is actually not necessary for the well-posedness of $(H\text{-}SD)$. Consider:

(WP_3) $\exists\ K \geq 0$, $L \in \mathbb{R}$ such that $\forall x \in C$, $||H(x)^{-1}|| \leq K|x| + L$.

This property is satisfied by relevant Legendre type functions; take for instance (33).

Theorem 4.1. *Assume that* (16) *and* (H_0) *hold and additionally that either* (WP_1), (WP_2) *or* (WP_3) *is satisfied. If $\inf_{\mathcal{F}} f > -\infty$ then the dynamical system $(H\text{-}SD)$ is well-posed. Consequently, the mapping $t \mapsto f(x(t))$ is nonincreasing and convergent as $t \to +\infty$.*

Proof. When no confusion may occur, we drop the dependence on the time variable t. By definition,

$$T_M = \sup\{T > 0 \,|\exists!\text{ solution } x \text{ of } (H\text{-}SD) \text{ on } [0,T) \text{ s.t. } x([0,T)) \subset \mathcal{F}\}.$$

We have that $T_M > 0$. The definition (8) of P_x implies that for all $y \in \mathcal{A}_0$, $(H(x)^{-1}\nabla f(x) + \dot{x}, y + \dot{x})_x^H = 0$ on $[0, T_M)$ and therefore

$$\langle \nabla f(x) + H(x)\dot{x}, y + \dot{x}\rangle = 0 \text{ on } [0, T_M). \tag{21}$$

Letting $y = 0$ in (21), yields

$$\frac{d}{dt}f(x) + \langle H(x)\dot{x}, \dot{x}\rangle = 0. \tag{22}$$

By (3)(ii), $f(x(t))$ is convergent as $t \to T_M$. Moreover

$$\langle H(x(\cdot))\dot{x}(\cdot), \dot{x}(\cdot)\rangle \in L^1(0, T_M; \mathbb{R}). \tag{23}$$

Suppose that $T_M < +\infty$. To obtain a contradiction, we begin by proving that x is bounded. If (WP_1) holds then x is bounded because $f(x(t))$ is non-increasing so that $x(t) \in \{y \in \overline{\mathcal{F}} | f(y) \leq f(x^0)\}$, $\forall t \in [0, T_M)$. Assume now that f and h comply with (WP_2), and let $a \in \overline{\mathcal{F}}$. For each $t \in [0, T_M)$ take $y = x(t) - a$ in (21) to obtain $\langle \nabla f(x) + \frac{d}{dt}\nabla h(x), x - a + \dot{x}\rangle = 0$. By (22), this gives $\langle \frac{d}{dt}\nabla h(x), x - a\rangle + \langle \nabla f(x), x - a\rangle = 0$, which we rewrite as

$$\frac{d}{dt}D_h(a, x(t)) + \langle \nabla f(x(t)), x(t) - a\rangle = 0, \ \forall t \in [0, T_M). \tag{24}$$

Now, let $a \in \overline{F}$ be a minimizer of f on $\overline{\mathcal{F}}$. From the quasi-convexity property of f, it follows that $\forall t \in [0, T_M)$, $\langle \nabla f(x(t)), x(t) - a\rangle \geq 0$. Therefore, $D_h(a, x(t))$ is non-increasing and (WP_2)(ii) implies that x is bounded. Suppose that (WP_3) holds and fix $t \in [0, T_M)$, we have $|x(t) - x^0| \leq \int_0^t |\dot{x}(s)|ds \leq \int_0^t ||\sqrt{H(x(s))^{-1}}|||\sqrt{H(x(s))}\,\dot{x}(s)|ds \leq (\int_0^t ||H(x(s))^{-1}||ds)^{1/2}(\int_0^t \langle H(x(s))\dot{x}(s), \dot{x}(s)\rangle ds)^{1/2}$. The latter follows from the Cauchy-Schwartz inequality together with the fact that $\|H(x)\|^2$ is the biggest eigenvalue of $H(x)$. Thus $|x(t) - x^0| \leq 1/2[\int_0^t ||H(x(s))^{-1}||ds + \int_0^t \langle H(x(s))\dot{x}(s), \dot{x}(r)\rangle ds]$. Combining (WP_3) and (23), Gronwall's lemma yields the boundedness of x.

Let $\omega(x^0)$ be the set of limit points of x, and set $K = x([0, T_M)) \cup \omega(x^0)$. Since x is bounded, $\omega(x^0) \neq \emptyset$ and K is compact. If $K \subset C$ then the compactness of K implies that x can be extended beyond T_M, which contradicts the maximality of T_M. Let us prove $K \subset C$. We argue again by contradiction. Assume that $x(t_j) \to x^*$, with $t_j < T_M$, $t_j \to T_M$ as $j \to +\infty$ and $x^* \in \operatorname{bd} C = \overline{C} \setminus C$. Since h is of Legendre type, we have $|\nabla h(x(t_j))| \to +\infty$, and we may assume that $\nabla h(x(t_j))/|\nabla h(x(t_j))| \to \nu \in \mathbb{R}^n$ with $|\nu| = 1$.

Lemma 4.1. *If $(x^j) \subset C$ is such that $x^j \to x^* \in \operatorname{bd} C$ and $\nabla h(x^j)/|\nabla h(x^j))| \to \nu \in \mathbb{R}^n$, h being a function of Legendre type with $C = \operatorname{int} \operatorname{dom} h$, then $\nu \in N_{\overline{C}}(x^*)$.*

of Lemma 4.1. . By convexity of h, $\langle \nabla h(x^j) - \nabla h(y), x^j - y\rangle \geq 0$ for all $y \in C$. Dividing by $|\nabla h(x^j)|$ and letting $j \to +\infty$, we get $\langle \nu, y - x^*\rangle \leq 0$ for all $y \in C$, which holds also for $y \in \overline{C}$. Hence, $\nu \in N_{\overline{C}}(x^*)$. □

Therefore, $\nu \in N_{\overline{C}}(x^*)$. Let $\nu_0 = \Pi_{\mathcal{A}_0}\nu$ be the Euclidean orthogonal projection of ν onto $\mathcal{A}_0$, and take $y = \nu_0$ in (21). Using (22), integration gives

$$\langle \nabla h(x(t_j)), \nu_0 \rangle = \langle \nabla h(x^0) - \int_0^{t_j} \nabla f(x(s))ds, \nu_0 \rangle. \tag{25}$$

By (H_0) and the boundedness property of x, the right-hand side of (25) is bounded under the assumption $T_M < +\infty$. Hence, to draw a contradiction from (25) it suffices to prove $\langle \nabla h(x(t_j)), \nu_0 \rangle \to +\infty$. Since $\langle \nabla h(x(t_j))/|\nabla h(x(t_j))|, \nu_0 \rangle \to |\nu_0|^2$, the proof of the result is complete if we check that $\nu_0 \neq 0$. This is a direct consequence of the following

Lemma 4.2. *Let C be a nonempty open convex subset of $\mathbb{R}^n$ and $\mathcal{A}$ an affine subspace of $\mathbb{R}^n$ such that $C \cap \mathcal{A} \neq \emptyset$. If $x^* \in (\text{bd}\, C) \cap \mathcal{A}$ then $N_{\overline{C}}(x^*) \cap \mathcal{A}_0^{\perp} = \{0\}$ with $\mathcal{A}_0 = \mathcal{A} - \mathcal{A}$.*

of Lemma 4.2. . Let us argue by contradiction and suppose that we can pick some $v \neq 0$ in $\mathcal{A}_0^{\perp} \cap N_{\overline{C}}(x^*)$. For $y_0 \in C \cap \mathcal{A}$ we have $\langle v, x^* - y_0 \rangle = 0$. For $r \geq 0$, $z \in \mathbb{R}^n$, let $B(z, r)$ denote the ball with center z and radius r. There exists $\epsilon > 0$, such that $B(y_0, \epsilon) \subset C$. Take w in $B(0, \epsilon)$ such that $\langle v, w \rangle < 0$, then $y_0 + w \in C$, yet $\langle v, x^* - (y_0 + w) \rangle = \langle v, w \rangle < 0$. This contradicts the fact that v is in $N_{\overline{C}}(x^*)$. □

This completes the proof of the theorem. □

4.2 Value convergence for a convex objective function

As a first result concerning the asymptotic behavior of $(H\text{-}SD)$, we have the following:

Proposition 4.1. *If $(H\text{-}SD)$ is well-posed and f is convex then $\forall a \in \mathcal{F}$, $\forall t > 0$, $f(x(t)) \leq f(a) + \frac{1}{t} D_h(a, x^0)$, where D_h is defined by (19), hence $\lim_{t \to +\infty} f(x(t)) = \inf_{\overline{\mathcal{F}}} f$.*

Proof. We begin by noticing that $f(x(t))$ converges as $t \to +\infty$ (see Theorem 4.1). Fix $a \in \mathcal{F}$. By (24), we have that the solution $x(t)$ of $(H\text{-}SD)$ satisfies $\frac{d}{dt} D_h(a, x(t)) + \langle \nabla f(x(t)), x(t) - a \rangle = 0$, $\forall t \geq 0$. The convex inequality $f(x) + \langle \nabla f(x), x - a \rangle \leq f(a)$ yields $D_h(a, x(t)) + \int_0^t [f(x(s)) - f(a)]ds \leq D_h(a, x^0)$. Using that $D_h \geq 0$ and since $f(x(t))$ is non-increasing, we get the estimate. Letting $t \to +\infty$, it follows that $\lim_{t \to +\infty} f(x(t)) \leq f(a)$. Since $a \in \mathcal{F}$ was arbitrary chosen, the proof is complete. □

4.3 Bregman metrics and trajectory convergence

In this section we establish the convergence of $x(t)$ under some additional properties on the D-function of h. Let us begin with a definition.

Definition 4.1. *A function $h \in \Gamma_0(\mathbb{R}^n)$ is called* Bregman function with zone C *when the following conditions are satisfied:*
(i) dom $h = \overline{C}$, *h is continuous and strictly convex on $\overline{C}$ and $h_{|C} \in \mathcal{C}^1(C; \mathbb{R})$.*
(ii) *$\forall a \in \overline{C}$, $\forall \gamma \in \mathbb{R}$, $\{y \in C | D_h(a, y) \leq \gamma\}$ is bounded, where D_h is defined by* (19).
(iii) *$\forall y \in \overline{C}$, $\forall y^j \to y$ with $y^j \in C$, $D_h(y, y^j) \to 0$.*

Observe that this notion slightly weakens the usual definition of Bregman function that was proposed by Censor and Lent in [11]; see also [7]. Actually, a Bregman function in the sense of Definition 4.1 belongs to the class of B-functions introduced by Kiwiel (see [31, Definition 2.4]). Recall the following important asymptotic separation property:

Lemma 4.3. [31, Lemma 2.16] *If h is a Bregman function with zone C then $\forall y \in \overline{C}$, $\forall (y^j) \subset C$ such that $D_h(y, y^j) \to 0$, we have $y^j \to y$.*

Theorem 4.2. *Suppose that (H_0) holds with h being a Bregman function with zone C. If f is quasi-convex satisfying* (16) *and $S(P) \neq \emptyset$ then $(H\text{-}SD)$ is well-posed and its solution $x(t)$ converges as $t \to +\infty$ to some $x^* \in \overline{\mathcal{F}}$ with $-\nabla f(x^*) \in N_{\overline{C}}(x^*) + \mathcal{A}_0^{\perp}$. If in addition f is convex then $x(t)$ converges to a solution of (P).*

Proof. Notice first that (WP_2) is satisfied. By Theorem 4.1, $(H\text{-}SD)$ is well-posed, $x(t)$ is bounded and for each $a \in S(P)$, $D_h(a, x(t))$ is non-increasing and hence convergent Set $f_\infty = \lim_{t\to+\infty} f(x(t))$ and define $L = \{y \in \overline{\mathcal{F}} \mid f(y) \leq f_\infty\}$. The set L is nonempty and closed. Since f is supposed to be quasi-convex, L is convex, and similar arguments as in the proof of Theorem 4.1 under (WP_2) show that $D_h(a, x(t))$ is convergent for all $a \in L$. Let $x^* \in L$ denote a cluster point of $x(t)$ and take $t_j \to +\infty$ such that $x(t_j) \to x^*$. Then, by (iii) in Definition 4.1, $\lim_t D_h(x^*, x(t)) = \lim_j D_h(x^*, x(t_j)) = 0$. Therefore, $x(t) \to x^*$ thanks to Lemma 4.3. Let us prove that x^* satisfies the optimality condition $-\nabla f(x^*) \in N_{\overline{C}}(x^*) + \mathcal{A}_0^{\perp}$. Fix $z \in \mathcal{A}_0$, and for each $t \geq 0$ take $y = -\dot{x}(t) + z$ in (21) to obtain $\langle \frac{d}{dt}\nabla h(x(t)) + \nabla f(x(t)), z\rangle = 0$. This gives

$$\frac{1}{t}\int_0^t \langle \nabla f(x(s)), z\rangle ds = \langle s(t), z\rangle, \tag{26}$$

where $s(t) = [\nabla h(x^0) - \nabla h(x(t))]/t$. If $x^* \in \mathcal{F}$ then $\nabla h(x(t)) \to \nabla h(x^*)$, hence $\langle \nabla f(x^*), z\rangle = \lim_{t\to+\infty} \frac{1}{t}\int_0^t \langle \nabla f(x(s)), z\rangle ds = \lim_{t\to+\infty}\langle s(t), z\rangle = 0$. Therefore, $\Pi_{\mathcal{A}_0}\nabla f(x^*) = 0$. But $N_{\overline{\mathcal{F}}}(x^*) = \mathcal{A}_0^{\perp}$ when $x^* \in \mathcal{F}$, which proves our claim in this case. Assume now that $x^* \notin \mathcal{F}$, which implies that $x^* \in \partial C \cap \mathcal{A}$. By (26), we have that $\langle s(t), z\rangle$ converges to $\langle \nabla f(x^*), z\rangle$ as $t \to +\infty$ for all $z \in \mathcal{A}_0$, and therefore $\Pi_{\mathcal{A}_0} s(t) \to \Pi_{\mathcal{A}_0}\nabla f(x^*)$ as $t \to +\infty$. On the other hand, by Lemma 4.1, we have that there exists $\nu \in -N_{\overline{C}}(x^*)$ with $|\nu| = 1$ such that $\nabla h(x(t_j))/|\nabla h(x(t_j))| \to \nu$ for some $t_j \to +\infty$. Since $N_{\overline{C}}(x^*)$ is positively homogeneous, we deduce that $\exists\ \bar{\nu} \in -N_{\overline{C}}(x^*)$ such that $\Pi_{\mathcal{A}_0}\nabla f(x^*) = \Pi_{\mathcal{A}_0}\bar{\nu}$. Thus, $-\nabla f(x^*) \in -\Pi_{\mathcal{A}_0}\bar{\nu} + \mathcal{A}_0^{\perp} \subseteq N_{\overline{C}}(x^*) + \mathcal{A}_0^{\perp}$, which proves the theorem. □

Following [27], we remark that when f is linear, the limit point can be characterized as a sort of "D_h-projection" of the initial condition onto the optimal set $S(P)$. In fact, we have:

Corollary 4.1. *Under the assumptions of Theorem 4.2, if f is linear then the solution $x(t)$ of $(H\text{-}SD)$ converges as $t \to +\infty$ to the unique optimal solution x^* of*

$$\min_{x\in S(P)} D_h(x, x^0). \tag{27}$$

Proof. Let $x^* \in S(P)$ be such that $x(t) \to x^*$ as $t \to +\infty$. Let $\bar{x} \in S(P)$. Since $x(t) \in \mathcal{F}$, the optimality of $\bar{x}$ yields $f(x(t)) \geq f(\bar{x})$, and it follows from (20) that $D_h(x(t), x^0) \leq D_h(\bar{x}, x^0)$. Letting $t \to +\infty$ in the last inequality, we deduce that x^* solves (27). Noticing that $D_h(\cdot, x^0)$ is strictly convex due to Definition 4.1(i), we conclude the result. □

We finish this section with an abstract result concerning the rate of convergence under uniqueness of the optimal solution. We will apply this result in the next section. Suppose that f is convex and satisfies (3) and (16), with in addition $S(P) = \{a\}$. Given a Bregman function h complying with (H_0), consider the following growth condition:

$$(GC) \qquad f(x) - f(a) \geq \alpha D_h(a, x)^\beta, \ \forall x \in U_a \cap \overline{C},$$

where U_a is a neighborhood of a and with $\alpha > 0$, $\beta \geq 1$. The next abstract result gives an estimation of the convergence rate with respect to the D-function of h.

Proposition 4.2. *Assume that f and h satisfy the above conditions an let $x : [0, +\infty) \to \mathcal{F}$ be the solution of $(H\text{-}SD)$. Then we have the following estimations:*

- *If $\beta = 1$ then there exists $K > 0$ such that $D_h(a, x(t)) \leq K e^{-\alpha t}$, $\forall t > 0$.*
- *If $\beta > 1$ then there exists $K' > 0$ such that $D_h(a, x(t)) \leq K'/t^{\frac{1}{\beta - 1}}$, $\forall t > 0$.*

Proof. The assumptions of Theorem 4.2 are satisfied, this yields the well-posedness of $(H\text{-}SD)$ and the convergence of $x(t)$ to a as $t \to +\infty$. Besides, from (24) it follows that for all $t \geq 0$, $\frac{d}{dt} D_h(a, x(t)) + \langle \nabla f(x(t)), x(t) - a \rangle = 0$. By convexity of f, we have $\frac{d}{dt} D_h(a, x(t)) + f(x(t)) - f(a) \leq 0$. Since $x(t) \to a$, there exists t_0 such that $\forall t \geq t_0$, $x(t) \in U_a \cap \mathcal{F}$. Therefore by combining (GC) and the last inequality it follows that

$$\frac{d}{dt} D_h(a, x(t)) + \alpha D_h(a, x(t))^\beta \leq 0, \ \forall t \geq t_0. \tag{28}$$

In order to integrate this differential inequality, let us first observe that we have the following equivalence: $D_h(a, x(t)) > 0$, $\forall t \geq 0$ iff $x^0 \neq a$. Indeed, if $a \in \overline{\mathcal{F}} \setminus \mathcal{F}$ then the equivalence follows from $x(t) \in \mathcal{F}$ together with Lemma 4.3; if $a \in \mathcal{F}$ then the optimality condition that is satisfied by a is $\Pi_{\mathcal{A}_0} \nabla f(a) = 0$, and the equivalence is a consequence of the uniqueness of the solution $x(t)$ of $(H\text{-}SD)$. Hence, we can assume that $x^0 \neq a$ and divide (28) by $D_h(a, x(t))^\beta$ for all $t \geq t_0$. A simple integration procedure then yields the result. □

4.4 Examples: interior point flows in convex programming

This section gives a systematic method to construct explicit Legendre metrics on a quite general class of convex sets. By so doing, we will also show that many systems studied earlier by various authors [5, 29, 18, 21, 34] appears as particular cases of $(H\text{-}SD)$ systems.

Let $p \geq 1$ be an integer and set $I = \{1, \ldots, p\}$. Let us assume that to each $i \in I$ there corresponds a $\mathcal{C}^3$ concave function $g_i : \mathbb{R}^n \to \mathbb{R}$ such that

$$\exists x^0 \in \mathbb{R}^n, \ \forall i \in I, \ g_i(x^0) > 0. \tag{29}$$

Suppose that the open convex set C is given by

$$C = \{x \in \mathbb{R}^n \mid g_i(x) > 0, i \in I\}. \tag{30}$$

By (29) we have that $C \neq \emptyset$ and $\overline{C} = \{x \in \mathbb{R}^n \mid g_i(x) \geq 0, i \in I\}$. Let us introduce a class of convex functions of Legendre type $\theta \in \Gamma_0(\mathbb{R})$ satisfying

$$(H_1) \qquad \begin{cases} \text{(i) } (0,\infty) \subset \operatorname{dom}\theta \subset [0,\infty). \\ \text{(ii) } \theta \in \mathcal{C}^3(0,\infty) \text{ and } \lim_{s\to 0^+} \theta'(s) = -\infty. \\ \text{(iii) } \forall s > 0,\ \theta''(s) > 0. \\ \text{(iv) Either } \theta \text{ is non-increasing or } \forall i \in I, g_i \text{ is an affine function.} \end{cases}$$

Proposition 4.3. *Under* (29) *and* (H_1)*, the function* $h \in \Gamma_0(\mathbb{R}^n)$ *defined by*

$$h(x) = \sum_{i\in I} \theta(g_i(x)). \tag{31}$$

is essentially smooth with $\operatorname{int}\operatorname{dom} h = C$ *and* $h \in \mathcal{C}^3(C)$*, where* C *is given by* (30)*. If we assume in addition the following non-degeneracy condition:*

$$\forall x \in C,\ \operatorname{span}\{\nabla g_i(x) \mid i \in I\} = \mathbb{R}^n, \tag{32}$$

then $H = \nabla^2 h$ *is positive definite on* C*, and consequently* h *satisfies* (H_0)*.*

Proof. Define $h_i \in \Gamma_0(\mathbb{R}^n)$ by $h_i(x) = \theta(g_i(x))$. We have that $\forall i \in I$, $C \subset \operatorname{dom} h_i$. Hence $\operatorname{int}\operatorname{dom} h = \bigcap_{i\in I} \operatorname{int}\operatorname{dom} h_i \supseteq C \neq \emptyset$, and by [36, Theorem 23.8], we conclude that $\partial h(x) = \sum_{i\in I} \partial h_i(x)$ for all $x \in \mathbb{R}^n$. But $\partial h_i(x) = \theta'(g_i(x))\nabla g_i(x)$ if $g_i(x) > 0$ and $\partial h_i(x) = \emptyset$ if $g_i(x) \leq 0$; see [24, Theorem IX.3.6.1]. Therefore $\partial h(x) = \sum_{i\in I} \theta'(g_i(x))\nabla g_i(x)$ if $x \in C$, and $\partial h(x) = \emptyset$ otherwise. Since ∂h is a single-valued mapping, it follows from [36, Theorem 26.1] that h is essentially smooth and $\operatorname{int}\operatorname{dom} h = \operatorname{dom}\partial h = C$. Clearly, h is of class $\mathcal{C}^3$ on C. Assume now that (32) holds. For $x \in C$, we have $\nabla^2 h(x) = \sum_{i\in I} \theta''(g_i(x))\nabla g_i(x)\nabla g_i(x)^T + \sum_{i\in I} \theta'(g_i(x))\nabla^2 g_i(x)$. By (H_1)(iv), it follows that for any $v \in \mathbb{R}^n$, $\sum_{i\in I} \theta'(g_i(x))\langle \nabla^2 g_i(x)v, v\rangle \geq 0$. Let $v \in \mathbb{R}^n$ be such that $\langle \nabla^2 h(x)v, v\rangle = 0$, which yields $\sum_{i\in I} \theta''(g_i(x))\langle v, \nabla g_i(x)\rangle^2 = 0$. According to (H_1)(iii), the latter implies that $v \in \operatorname{span}\{\nabla g_i(x) | i \in I\}^\perp = \{0\}$. Hence $\nabla^2 h(x) \in \mathbb{S}^n_{++}$ and the proof is complete. □

If h is defined by (31) with $\theta \in \Gamma_0(\mathbb{R})$ satisfying (H_1), we say that θ is the *Legendre kernel* of h. Such kernels can be divided into two classes. The first one corresponds to those kernels θ for which $\operatorname{dom}\theta = (0,\infty)$ so that $\theta(0) = +\infty$, and are associated with *interior barrier* methods in optimization as for instance : the log-barrier $\theta_1(s) = -\ln(s)$, $s > 0$ and the inverse barrier $\theta_2(s) = 1/s$, $s > 0$. The kernels θ belonging to the second class satisfy $\theta(0) < +\infty$, and are connected with the notion of *Bregman function* in proximal algorithms theory. Here are some examples: the Boltzmann-Shannon entropy $\theta_3(s) = s\ln(s) - s$, $s \geq 0$ (with $0\ln 0 = 0$); $\theta_4(s) = -\frac{1}{\gamma}s^\gamma$ with $\gamma \in (0,1)$, $s \geq 0$ (Kiwiel [31]); $\theta_5(s) = (\gamma s - s^\gamma)/(1-\gamma)$ with $\gamma \in (0,1)$, $s \geq 0$ (Teboulle [38]); the "$x\log x$" entropy $\theta_6(s) = s\ln s$, $s \geq 0$. In relation with Theorem 4.2 given in the previous section, note that the Legendre kernels θ_i, $i = 3, ..., 6$, are all Bregman functions with zone $\mathbb{R}_+$. Moreover, it is easily seen that each corresponding Legendre function h defined by (31) is indeed a Bregman function with zone C.

In order to illustrate the type of dynamical systems given by (H-SD), consider the case of positivity constraints where $p = n$ and $g_i(x) = x_i$, $i \in I$. Thus $C = \mathbb{R}^n_{++}$ and $\overline{C} = \mathbb{R}^n_+$. Let us assume that $\exists x^0 \in \mathbb{R}^n_{++}$, $Ax^0 = b$. Recall that the corresponding minimization problem is $(P)\ \ \min\{f(x) \mid x \geq 0,\ Ax = b\}$ and take first the kernel θ_3 from above. The associated Legendre function (31) is given by

$$h(x) = \sum_{i=1}^{n} x_i \ln x_i - x_i,\ x \in \mathbb{R}^n_+, \tag{33}$$

and the differential equation in (H-SD) is given by

$$\dot{x} + [I - XA^T(AXA^T)^{-1}A]X\nabla f(x) = 0. \tag{34}$$

where $X = \operatorname{diag}(x_1, ..., x_n)$. If $f(x) = \langle c, x\rangle$ for some $c \in \mathbb{R}^n$ and in absence of linear equality constraints, then (34) is $\dot{x}+Xc = 0$. The change of coordinates $y = \nabla h(x) = (\ln x_1, ..., \ln x_n)$ gives $\dot{y} + c = 0$. Hence, $x(t) = (x^0_1 e^{-c_1 t}, ..., x^0_n e^{-c_n t})$, $t \in \mathbb{R}$, where $x^0 = (x^0_1, ..., x^0_n) \in \mathbb{R}^n_{++}$. If $c \in \mathbb{R}^n_+$ then $\inf_{x\in\mathbb{R}^n_+}\langle c, x\rangle = 0$ and $x(t)$ converges to a minimizer of $f = \langle c, \cdot\rangle$ on $\mathbb{R}^n_+$; if $c_{i_0} < 0$ for some i_0, then $\inf_{x\in\mathbb{R}^n_+}\langle c, x\rangle = -\infty$ and $x_{i_0}(t) \to +\infty$ as $t \to +\infty$. Next, take $A = (1, \ldots, 1) \in \mathbb{R}^{1\times n}$ and $b = 1$ so that the feasible set of (P) is given by $\overline{\mathcal{F}} = \Delta_{n-1} = \{x \in \mathbb{R}^n \mid x \geq 0, \sum_{i=1}^n x_i = 1\}$, that is the $(n-1)$-dimensional simplex. In this case, (34) corresponds to $\dot{x} + [X - xx^T]\nabla f(x) = 0$, or componentwise

$$\dot{x}_i + x_i \left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^{n} x_j \frac{\partial f}{\partial x_j} \right) = 0, \quad i = 1, \ldots, n. \tag{35}$$

For suitable choices of f, this is a *Lotka-Volterra* type equation that naturally arises in population dynamics theory and, in that context, the structure $(\cdot, \cdot)^H$ with h as in (33) is usually referred to as the *Shahshahani* metric; see [1, 25] and the references therein. The figure 1 gives a numerical illustration of system (35) for $n = 3$ and with $f(x) = x_3 - x_2$. Karmarkar studied (35) in [29] for a quadratic objective function as a continuous model of the

Shasha.eps

Figure 1: A trajectory of (35).

interior point algorithm introduced by him in [28]. Equation (34) is studied by Faybusovich in [18, 19, 20] when (P) is a linear program, establishing connections with completely integrable Hamiltonian systems and exponential convergence rate, and by Herzel et al. in [23], who prove quadratic convergence for an explicit discretization.

Take now the log barrier kernel θ_1 and $h(x) = -\sum_{i=1}^n \ln x_i$. Since $\nabla^2 h(x) = X^{-2}$ with X defined as above, the associated differential equation is

$$\dot{x} + [I - X^2 A^T (A X^2 A^T)^{-1} A] X^2 \nabla f(x) = 0. \tag{36}$$

This equation was considered by Bayer and Lagarias in [5] for a linear program. In the particular case $f(x) = \langle c, x\rangle$ and without linear equality constraints, (36) amounts to $\dot{x} + X^2 c = 0$, or $\dot{y} + c = 0$ for $y = \nabla h(x) = -X^{-1}e$ with $e = (1, \cdots, 1) \in \mathbb{R}^n$, which gives $x(t) = (1/(1/x_1^0 + c_1 t), ..., 1/(1/x_n^0 + c_n t))$, $T_m \le t \le T_M$, with $T_m = \max\{-1/x_i^0 c_i \mid c_i > 0\}$ and $T_M = \min\{-1/x_i^0 c_i \mid c_i < 0\}$ (see [5, pag. 515]). Denote by Π_{A_0} the Euclidean orthogonal projection onto $\mathcal{A}_0$. To study the associated trajectories for a general linear program, it is introduced in [5] the *Legendre transform coordinates* $y = \Pi_{\mathcal{A}_0} \nabla h(x) = [I - A^T (A A^T)^{-1} A] X^{-1} e$, which still linearizes (36) when f is linear (see section 5 for an extension of this result), and permits to establish some remarkable analytic and geometric properties of the trajectories. A similar system was considered in [21, 34] as a continuous log-barrier method for nonlinear inequality constraints and with $\mathcal{A}_0 = \mathbb{R}^n$.

New systems may be derived by choosing other kernels. For instance, taking $h(x) = -1/\gamma \sum_{i=1}^n x_i^\gamma$ with $\gamma \in (0,1)$, $A = (1, \ldots, 1) \in \mathbb{R}^{1\times n}$ and $b = 1$, we obtain

$$\dot{x}_i + \frac{x_i^{2-\gamma}}{1-\gamma}\left(\frac{\partial f}{\partial x_i} - \sum_{j=1}^n \frac{x_j^{2-\gamma}}{\sum_{k=1}^n x_k^{2-\gamma}} \frac{\partial f}{\partial x_j}\right) = 0, \quad i = 1, \ldots, n. \tag{37}$$

4.5 Convergence results for linear programming

Let us consider the specific case of a linear program

$$(LP) \qquad \min_{x\in\mathbb{R}^n} \{\langle c, x\rangle \mid Bx \ge d,\ Ax = b\},$$

where A and b are as in section 2.1, $c \in \mathbb{R}^n$, B is a $p \times n$ full rank real matrix with $p \ge n$ and $d \in \mathbb{R}^p$. We assume that the optimal set satisfies

$$S(LP) \text{ is nonempty and bounded,} \tag{38}$$

and there exists a Slater point $x^0 \in \mathbb{R}^n$, $Bx^0 > d$ and $Ax^0 = b$. Take the Legendre function

$$h(x) = \sum_{i=1}^n \theta(g_i(x)), \quad g_i(x) = \langle B_i, x\rangle - d_i, \tag{39}$$

where $B_i \in \mathbb{R}^n$ is the ith-row of B and the Legendre kernel θ satisfies (H_1). By (38), (WP_1) holds and therefore $(H\text{-}SD)$ is well-posed due to Theorem 4.1. Moreover, $x(t)$ is bounded and all its cluster points belong to $S(LP)$ by Proposition 4.1. The variational property (20) ensures the convergence of $x(t)$ and gives a variational characterization of the limit as well. Indeed, we have the following result:

Proposition 4.4. *Let h be given by* (39) *with θ satisfying* (H_1). *Under* (38), (H-SD) *is well-posed and $x(t)$ converges as $t \to +\infty$ to the unique solution x^* of*

$$\min_{x \in S(LP)} \sum_{i \notin I_0} D_\theta(g_i(x), g_i(x^0)), \tag{40}$$

where $I_0 = \{i \in I \mid g_i(x) = 0$ for all $x \in S(LP)\}$.

Proof. Assume that $S(LP)$ is not a singleton, otherwise there is nothing to prove. The relative interior $\mathrm{ri}\, S(LP)$ is nonempty and moreover $\mathrm{ri}\, S(LP) = \{x \in \mathbb{R}^n \mid g_i(x) = 0$ for $i \in I_0,\ g_i(x) > 0$ for $i \notin I_0,\ Ax = b\}$. By compactness of $S(LP)$ and strict convexity of $\theta \circ g_i$, there exists a unique solution x^* of (40). Indeed, it is easy to see that $x^* \in \mathrm{ri}\,(LP)$. Let $\bar{x} \in S(LP)$ and $t_j \to +\infty$ be such that $x(t_j) \to \bar{x}$. It suffices to prove that $\bar{x} = x^*$. When $\theta(0) < +\infty$, the latter follows by the same arguments as in Corollary 4.1. When $\theta(0) = +\infty$, the proof of [4, Theorem 3.1] can be adapted to our setting (see also [27, Theorem 2]). Set $x^*(t) = x(t) - \bar{x} + x^*$. Since $Ax^*(t) = b$ and $D_h(x, x^0) = \sum_{i=1}^m D_\theta(g_i(x), g_i(x^0))$, (20) gives

$$\langle c, x(t)\rangle + \frac{1}{t}\sum_{i=1}^m D_\theta(g_i(x(t)), g_i(x^0)) \le \langle c, x^*(t)\rangle + \frac{1}{t}\sum_{i=1}^m D_\theta(g_i(x^*(t)), g_i(x^0)). \tag{41}$$

But $\langle c, x(t)\rangle = \langle c, x^*(t)\rangle$ and $\forall i \in I_0,\ g_i(x^*(t)) = g_i(x(t)) > 0$. Since $x^* \in \mathrm{ri}\, S(LP)$, for all $i \notin I_0$ and j large enough, $g_i(x^*(t_j)) > 0$. Thus, the right-hand side of (41) is finite at t_j, and it follows that $\sum_{i \notin I_0} D_\theta(g_i(\bar{x}), g_i(x^0)) \le \sum_{i \notin I_0} D_\theta(g_i(x^*), g_i(x^0))$. Hence, $\bar{x} = x^*$. □

Rate of convergence. We turn now to the case where there is no equality constraint so that the linear program is

$$\min_{x \in \mathbb{R}^n} \{\langle c, x\rangle \mid Bx \ge d\}. \tag{42}$$

We assume that (42) admits a unique solution a and we study the rate of convergence when θ is a Bregman function with zone $\mathbb{R}_+$. To apply Proposition 4.2, we need:

Lemma 4.4. *Set $C = \{x \in \mathbb{R}^n | Bx > d\}$. If* (42) *admits a unique solution $a \in \mathbb{R}^n$ then $\exists k_0 > 0$, $\forall y \in \overline{C}$, $\langle c, y - a\rangle \ge k_0 \mathcal{N}(y - a)$, where $\mathcal{N}(x) = \sum_{i \in I} |\langle B_i, x\rangle|$ is a norm on $\mathbb{R}^n$.*

Proof. Set $I_0 = \{i \in I \mid \langle B_i, a\rangle = d_i\}$. The optimality conditions for a imply the existence of a *multiplier* vector $\lambda \in \mathbb{R}^p_+$ such that $\lambda_i[d_i - \langle B_i, a\rangle] = 0$, $\forall i \in I$, and $c = \sum_{i \in I} \lambda_i B_i$. Let $y \in \overline{C}$. We deduce that $\langle c, y - a\rangle = N(y - a)$ where $N(x) = \sum_{i \in I_0} \lambda_i |\langle B_i, x\rangle|$. By uniqueness of the optimal solution, it is easy to see that $\mathrm{span}\{B_i \mid i \in I_0\} = \mathbb{R}^n$, hence N is a norm on $\mathbb{R}^n$. Since $\mathcal{N}(x) = \sum_{i \in I} |\langle B_i, x\rangle|$ is also a norm on $\mathbb{R}^n$ (recall that B is a full rank matrix), we deduce that $\exists k_0$ such that $N(x) \ge k_0 \mathcal{N}(x)$. □

The following lemma is a sharper version of Proposition 4.2 in the linear context.

Lemma 4.5. *Under the assumptions of Proposition 4.4, assume in addition that θ is a Bregman function with zone $\mathbb{R}_-$ and that there exist $\alpha > 0$, $\beta \geq 1$ and $\varepsilon > 0$ such that*

$$\forall s \in (0, \varepsilon),\ \alpha D_\theta(0, s)^\beta \leq s. \tag{43}$$

Then there exists positive constants K, L, M such that for all $t > 0$ the trajectory of $(H\text{-}SD)$ satisfies $D_h(a, x(t)) \leq Ke^{-Lt}$ if $\beta = 1$, and $D_h(a, x(t)) \leq M/t^{\frac{1}{\beta-1}}$ if $\beta > 1$.

Proof. By Lemma 4.4, there exists k_0 such that for all $t > 0$,

$$\langle c, x(t) - a\rangle \geq \sum_{i\in I} k_0 |\langle B_i, x(t)\rangle - \langle B_i, a\rangle|. \tag{44}$$

Now, if we prove that $\exists \lambda > 0$ such that

$$|\langle B_i, x(t)\rangle - \langle B_i, a\rangle| \geq \lambda D_\theta(\langle B_i, a\rangle - d_i, \langle B_i, x(t)\rangle - d_i) \tag{45}$$

for all $i \in I$ and for t large enough, then from (44) it follows that $f(\cdot) = \langle c, \cdot\rangle$ satisfies the assumptions of Proposition 4.2 and the conclusion follows easily. Since $x(t) \to a$, to prove (45) it suffices to show that $\forall r_0 \geq 0$, $\exists \eta, \mu > 0$ such that $\forall s$, $|s - r_0| < \eta$, $\mu D_\theta(r_0, s)^\beta \leq |r_0 - s|$. The case where $r_0 = 0$ is a direct consequence of (43). Let $r_0 > 0$. An easy computation yields $\frac{d^2}{ds^2} D_\theta(r_0, s)_{|s=r_0} = \theta''(r_0)$, and by Taylor's expansion formula

$$D_\theta(r_0, s) = \frac{\theta''(r_0)}{2}(s - r_0)^2 + o(s - r_0)^2 \tag{46}$$

with $\theta''(r_0) > 0$ due to (H_1)(iii). Let η be such that $\forall s$, $|s - r_0| < \eta$, $s > 0$, $D_\theta(r_0, s) \leq \theta''(r_0)(s - r_0)^2$ and $D_\theta(r_0, s) \leq 1$; since $\beta \geq 1$, $D_\theta(r_0, s)^\beta \leq D_\theta(r_0, s) \leq \theta''(r_0)|s - r_0|$. □

To obtain Euclidean estimates, the functions $s \mapsto D_\theta(r_0, s)$, $r_0 \in \mathbb{R}_+$ have to be locally compared to $s \mapsto |r_0 - s|$. By (46) and the fact that $\theta'' > 0$, for each $r_0 > 0$ there exists $K, \eta > 0$ such that $|r_0 - s| \leq K\sqrt{D_\theta(r_0, s)}$, $\forall s$, $|r_0 - s| < \eta$. This shows that, in practice, the Euclidean estimate depends only on a property of the type (43). Examples:

- The Boltzmann-Shannon entropy $\theta_3(s) = s\ln(s) - s$ and $\theta_6(s) = s\ln s$ satisfy $D_{\theta_i}(0, s) = s$, $s > 0$; hence for some $K, L > 0$, $|x(t) - a| \leq Ke^{-Lt}$, $\forall t \geq 0$.
- With either $\theta_4(s) = -s^\gamma/\gamma$ or $\theta_5(s) = (\gamma s - s^\gamma)/(1 - \gamma)$, $\gamma \in (0, 1)$, we have $D_{\theta_i}(0, s) = (1 + 1/\gamma)s^\gamma$, $s > 0$; hence $|x(t) - a| \leq K/t^{\frac{\gamma}{2-2\gamma}}$, $\forall t > 0$.

4.6 Dual convergence

In this section we focus on the case $C = \mathbb{R}^n_{++}$, so that the minimization problem is

$$(P) \qquad \min\{f(x) \mid x \geq 0,\ Ax = b\}.$$

We assume

$$f \text{ is convex and } S(P) \neq \emptyset, \tag{47}$$

together with the Slater condition

$$\exists x^0 \in \mathbb{R}^n,\ x^0 > 0,\ Ax^0 = b. \tag{48}$$

In convex optimization theory, it is usual to associate with (P) the *dual* problem given by

$$(D) \qquad \min\{p(\lambda) \mid \lambda \geq 0\},$$

where $p(\lambda) = \sup\{\langle \lambda, x\rangle - f(x) \mid Ax = b\}$. For many applications, dual solutions are as important as primal ones. In the particular case of a linear program where $f(x) = \langle c, x\rangle$ for some $c \in \mathbb{R}^n$, writing $\lambda = c + A^T y$ with $y \in \mathbb{R}^m$ the linear dual problem may equivalently be expressed as $\min\{\langle b, y\rangle \mid A^T y + c \geq 0\}$. Thus, λ is interpreted as a vector of *slack* variables for the dual inequality constraints. In the general case, $S(D)$ is nonempty and bounded under (47) and (48), and moreover $S(D) = \{\lambda \in \mathbb{R}^n \mid \lambda \geq 0,\ \lambda \in \nabla f(x^*) + \operatorname{Im} A^T,\ \langle \lambda, x^*\rangle = 0\}$, where x^* is any solution of (P); see for instance [24, Theorems VII.2.3.2 and VII.4.5.1].
Let us introduce a Legendre kernel θ satisfying (H_1) and define

$$h(x) = \sum_{i=1}^{n} \theta(x_i). \tag{49}$$

Suppose that $(H\text{-}SD)$ is well-posed. Integrating the differential inclusion (17), we obtain

$$\lambda(t) \in c(t) + \operatorname{Im} A^T, \tag{50}$$

where $c(t) = \frac{1}{t}\int_0^t \nabla f(x(\tau))d\tau$ and $\lambda(t)$ is the *dual trajectory* defined by

$$\lambda(t) = \frac{1}{t}[\nabla h(x^0) - \nabla h(x(t))]. \tag{51}$$

Assume that $x(t)$ is bounded. From (47), it follows that ∇f is constant on $S(P)$, and then it is easy to see that $\nabla f(x(t)) \to \nabla f(x^*)$ as $t \to +\infty$ for any $x^* \in S(P)$. Consequently, $c(t) \to \nabla f(x^*)$. By (51) together with [36, Theorem 26.5], we have $x(t) = \nabla h^*(\nabla h(x^0) - t\lambda(t))$, where the Fenchel conjugate h^* is given by $h^*(\lambda) = \sum_{i=1}^n \theta^*(\lambda_i)$. Take any solution $\widetilde{x}$ of $A\widetilde{x} = b$. Since $Ax(t) = b$, we have $\widetilde{x} - \nabla h^*(\nabla h(x^0) - t\lambda(t)) \in \operatorname{Ker} A$. On account of (50), $\lambda(t)$ is the unique optimal solution of

$$\lambda(t) \in \operatorname{Argmin}\left\{ \langle \widetilde{x}, \lambda\rangle + \frac{1}{t}\sum_{i=1}^{n} \theta^*(\theta'(x_i^0) - t\lambda_i) \mid \lambda \in c(t) + \operatorname{Im} A^T \right\}. \tag{52}$$

By (H_1)(iii), θ' is increasing in $\mathbb{R}_{++}$. Set $\eta = \lim_{s\to+\infty} \theta'(s) \in (-\infty, +\infty]$. Since θ^* is a Legendre type function, $\operatorname{int}\operatorname{dom}\theta^* = \operatorname{dom}\partial\theta^* = \operatorname{Im}\partial\theta = (-\infty, \eta)$. From $(\theta^*)' = (\theta')^{-1}$, it follows that $\lim_{u\to-\infty}(\theta^*)'(u) = 0$ and $\lim_{u\to\eta^-}(\theta^*)'(u) = +\infty$. Consequently, (52) can be interpreted as a *penalty approximation scheme* of the dual problem (D), where the dual positivity constraints are penalized by a separable strictly convex function. Similar schemes have been treated in [4, 14, 26]. Consider the additional condition

$$\text{Either } \theta(0) < \infty, \text{ or } S(P) \text{ is bounded, or } f \text{ is linear.} \tag{53}$$

As a direct consequence of [26, Propositions 10 and 11], we obtain that under (47), (48), (53) and (H_1), $\{\lambda(t) \mid t \to +\infty\}$ is bounded and its cluster points belong to $S(D)$. The convergence of $\lambda(t)$ is more difficult to establish. In fact, under some additional conditions on θ^* (see [14, Conditions (H_0)-(H_1)] or [26, Conditions (A7) and (A8)]) it is possible to show that $\lambda(t)$ converges to a particular element of the dual optimal set (the "θ^*-center" in the sense of [14, Definition 5.1] or the $D_h(\cdot, x^0)$-center as defined in [26, pag. 616]), which is characterized as the unique solution of a *nested hierarchy* of optimization problems on the dual optimal set. We will not develop this point here. Let us only mention that for all the examples of section 4.4, θ_i^* satisfies such additional conditions and consequently:

Proposition 4.5. *Under* (47), (48) *and* (53), *for each of the explicit Legendre kernels given in section* 4.4, $\lambda(t)$ *given by* (51) *converges to a particular dual solution.*

5 Legendre transform coordinates

5.1 Legendre functions on affine subspaces

The first objective of this section is to slightly generalize the notion of Legendre type function to the case of functions whose domains are contained in an affine subspace of $\mathbb{R}^n$. We begin by noticing that the Legendre type property does not depend on canonical coordinates.

Lemma 5.1. *Let* $g \in \Gamma_0(\mathbb{R}^r)$, $r \geq 1$, *and* $T : \mathbb{R}^r \to \mathbb{R}^r$ *an affine invertible mapping. Then* g *is of Legendre type iff* $g \circ T$ *is of Legendre type.*

Proof. The proof is elementary and is left to the reader. □

From now on, $\mathcal{A}$ is the affine subspace defined by (1), whose dimension is $r = n - m$.

Definition 5.1. *A function* $g \in \Gamma_0(\mathcal{A})$ *is said to be of* Legendre type *if there exists an affine invertible mapping* $T : \mathcal{A} \to \mathbb{R}^r$ *such that* $g \circ T^{-1}$ *is a Legendre type function in* $\Gamma_0(\mathbb{R}^r)$.

By Lemma 5.1, the previous definition is consistent.

Proposition 5.1. *Let* $h \in \Gamma_0(\mathbb{R}^n)$ *be a function of Legendre type with* $C = \operatorname{int} \operatorname{dom} h$. *If* $\mathcal{F} = C \cap \mathcal{A} \neq \emptyset$ *then the restriction* $h_{|\mathcal{A}}$ *of* h *to* $\mathcal{A}$ *is of Legendre type and moreover* $\operatorname{int}_{\mathcal{A}}\operatorname{dom} h_{|\mathcal{A}} = \mathcal{F}$ *(*$\operatorname{int}_{\mathcal{A}}B$ *stands for the interior of* B *in* $\mathcal{A}$ *as a topological subspace of* $\mathbb{R}^n$*).*

Proof. From the inclusions $\mathcal{F} \subset \operatorname{dom} h_{|\mathcal{A}} \subset \overline{\mathcal{F}} = \overline{C} \cap \mathcal{A}$ and since $\operatorname{ri} \overline{\mathcal{F}} = \mathcal{F}$, we conclude that $\operatorname{int}_{\mathcal{A}}\operatorname{dom} h_{|\mathcal{A}} = \mathcal{F} \neq \emptyset$. Let $T : \mathbb{R}^r \to \mathcal{A}$ be an invertible transformation with $Tz = Lz + x^0$ for all $z \in \mathbb{R}^r$, where $x^0 \in \mathcal{A}$ and $L : \mathbb{R}^r \to \mathcal{A}_0$ is a nonsingular linear mapping. Define $k = h_{|\mathcal{A}} \circ T$. Clearly, $k \in \Gamma_0(\mathbb{R}^r)$. Let us prove that k is essentially smooth. We have $\operatorname{dom} k = T^{-1}\operatorname{dom} h_{|\mathcal{A}}$ and therefore $\operatorname{int} \operatorname{dom} k = T^{-1}\mathcal{F}$. Since h is differentiable on C, we conclude that k is differentiable on $\operatorname{int} \operatorname{dom} k$. Now, let $(z^j) \in \operatorname{int} \operatorname{dom} k$ be a sequence that converges to a boundary point $z \in \operatorname{bd} \operatorname{dom} k$. Then, $Tz^j \in \operatorname{int}_{\mathcal{A}}\operatorname{dom} h_{|\mathcal{A}}$ and $Tz^j \to Tz \in \operatorname{bd}_{\mathcal{A}}\operatorname{dom} h_{|\mathcal{A}} \subset \operatorname{bddom} h$. Since h is essentially smooth, $|\nabla h(Tz^j)| \to +\infty$. Thus, to prove that

$|\nabla k(z^j)| \to +\infty$ it suffices to show that there exists $\lambda > 0$ such that $|\nabla k(z^j)| \geq \lambda |\nabla h(Tz^j)|$ for all j large enough. Note that $\nabla k(z^j) = \nabla[h_{|\mathcal{A}} \circ T](z^j) = L^*\nabla h_{|\mathcal{A}}(Tz^j) = L^*\Pi_{\mathcal{A}_0}\nabla h(Tz^j)$, where $L^* : \mathcal{A}_0 \to \mathbb{R}^r$ is defined by $\langle z, L^*x\rangle = \langle Lz, x\rangle$, $\forall (z,x) \in \mathbb{R}^r \times \mathcal{A}_0$. Of course, L^* is linear with Ker $L^* = \{0\}$. Therefore $\frac{\nabla k(z^j)}{|\nabla h(Tz^j)|} = L^*\Pi_{\mathcal{A}_0}\frac{\nabla h(Tz^j)}{|\nabla h(Tz^j)|}$. Let ω denote the nonempty and compact set of cluster points of the normalized sequence $\nabla h(Tz^j)/|\nabla h(Tz^j)|$, $j \in \mathbb{N}$. By Lemma 4.1, we have that $\omega \subset \{\nu \in N_{\overline{C}}(Tz) \mid |\nu| = 1\}$, and consequently Lemma 4.2 yields $\Pi_{\mathcal{A}_0}\omega \cap \{0\} = \emptyset$. By compactness of ω, we obtain $\liminf_{j\to+\infty} |\Pi_{\mathcal{A}_0}\nabla h(Tz^j)|/|\nabla h(Tz^j)| > 0$, which proves our claim. Finally, the strict convexity of k on $\operatorname{dom}\partial k = \operatorname{int}\operatorname{dom} k = T^{-1}\mathcal{F}$ is a direct consequence of the strict convexity of h in $\mathcal{F}$. □

5.2 Legendre transform coordinates

The prominent fact of Legendre functions theory is that $h \in \Gamma_0(\mathbb{R}^n)$ is of Legendre type iff its Fenchel conjugate h^* is of Legendre type [36, Theorem26.5], and $\nabla h : \operatorname{int}\operatorname{dom} h \to \operatorname{int}\operatorname{dom} h^*$ is onto with $(\nabla h)^{-1} = \nabla h^*$. In the case of Legendre functions on affine subspaces, we have the following generalization:

Proposition 5.2. *If $g \in \Gamma_0(\mathcal{A})$ is of Legendre type in the sense of Definition 5.1, then $\nabla g(\operatorname{int}_{\mathcal{A}}\operatorname{dom} g)$ is a nonempty, open and convex subset of $\mathcal{A}_0$. In addition, ∇g is a one-to-one continuous mapping from $\operatorname{int}_{\mathcal{A}}\operatorname{dom} g$ onto its image.*

Proof. Let $Tx = Lx + z_0$ with $L : \mathcal{A}_0 \to \mathbb{R}^r$ being a linear invertible mapping and $z_0 \in \mathbb{R}^p$. Set $k = g\circ T^{-1} \in \Gamma_0(\mathbb{R}^r)$, which is of Legendre type. We have $\operatorname{dom} k = T\operatorname{dom} g$. Define $L^* : \mathbb{R}^r \to \mathcal{A}_0$ by $\langle L^*z, x\rangle = \langle z, Lx\rangle$, $\forall (z,x) \in \mathbb{R}^r \times \mathcal{A}_0$. We have that $\nabla g(x) = \nabla[k\circ T](x) = L^*\nabla k(Tx)$ for all $x \in \operatorname{int}_{\mathcal{A}}\operatorname{dom} g$. Therefore $\nabla g(\operatorname{int}_{\mathcal{A}}\operatorname{dom} g) = L^*\nabla k(T\operatorname{int}_{\mathcal{A}}\operatorname{dom} g) = L^*\nabla k(\operatorname{int}_{\mathbb{R}^r}\operatorname{dom} k) = L^*\operatorname{int}_{\mathbb{R}^r}\operatorname{dom} k^*$. Since $\operatorname{int}_{\mathbb{R}^r}\operatorname{dom} k^*$ is a nonempty, open and convex subset of $\mathbb{R}^r$ and L^* is an invertible linear mapping, then $L^*\operatorname{int}_{\mathbb{R}^r}\operatorname{dom} k^*$ is an open and nonempty subset of $\mathcal{A}_0$. Moreover, by [36, Theorem 6.6], we have $L^*\operatorname{int}_{\mathbb{R}^r}\operatorname{dom} k^* = \operatorname{ri} L^*\operatorname{dom} k^*$. Consequently, $\nabla g(\operatorname{int}_{\mathcal{A}}\operatorname{dom} g) = \operatorname{ri} L^*\operatorname{dom} k^* = \operatorname{int}_{\mathcal{A}_0} L^*\operatorname{dom} k^* \neq \emptyset$. Finally, since $\nabla k : \operatorname{int}_{\mathbb{R}^r}\operatorname{dom} k \to \operatorname{int}_{\mathbb{R}^r}\operatorname{dom} k^*$ is one-to-one and continuous, the same result holds for $\nabla g = L^* \circ \nabla k \circ T$ on $\operatorname{int}_{\mathcal{A}}\operatorname{dom} g$. □

In the sequel, we assume that h satisfies the basic condition (H_0) and $\mathcal{F} = C \cap \mathcal{A} \neq \emptyset$. The *Legendre transform coordinates mapping* on $\mathcal{F}$ associated with h is defined by

$$\begin{array}{rccl} \phi_h : & \mathcal{F} & \to & \mathcal{F}^* = \phi_h(\mathcal{F}) \\ & x & \mapsto & \phi_h(x) = \nabla(h_{|\mathcal{A}}) = \Pi_{\mathcal{A}_0}\nabla h(x). \end{array} \tag{54}$$

This definition retrieves the Legendre transform coordinates introduced by Bayer and Lagarias in [5] for the particular case of the log-barrier on a polyhedral set.

Theorem 5.1. *Under the above definitions and assumptions, $\mathcal{F}^*$ is a convex, (relatively) open and nonempty subset of $\mathcal{A}_0$, ϕ_h is a $\mathcal{C}^1$ diffeomorphism from $\mathcal{F}$ to $\mathcal{F}^*$, and for all $x \in \mathcal{F}$, $d\phi_h(x) = \Pi_{\mathcal{A}_0}H(x)$ and $d\phi_h(x)^{-1} = \sqrt{H(x)^{-1}}\Pi_{\sqrt{H(x)}\mathcal{A}_0}\sqrt{H(x)^{-1}}$, where $H(x) = \nabla^2 h(x)$.*

Proof. By Propositions 5.1 and 5.2, $\mathcal{F}^*$ is a convex, open and nonempty subset of $\mathcal{A}_0$ and ϕ_h is a continuous bijection. By (H_0)(ii), ϕ_h is of class $\mathcal{C}^1$ on $\mathcal{F}$ and we have for all $x \in \mathcal{F}$, $d\phi_h(x) = \Pi_{\mathcal{A}_0}\nabla^2 h(x) = \Pi_{\mathcal{A}_0} H(x)$. Let $v \in \mathcal{A}_0$ be such that $d\phi_h(x)v = 0$. It follows that $H(x)v \in \mathcal{A}_0^\perp$ and in particular $\langle H(x)v, v\rangle = 0$. Hence, $v = 0$ thanks to (H_0)(iii). The implicit function theorem implies then that ϕ_h is a $\mathcal{C}^1$ diffeomorphism. The formula concerning $d\phi_h(x)^{-1}$ is a direct consequence of the next lemma.

Lemma 5.2. *Define the linear operators* $L_i : \mathbb{R}^n \to \mathbb{R}^n$ *by* $L_1 = \Pi_{\mathcal{A}_0} H(x)$ *and* $L_2 = \sqrt{H(x)^{-1}}\Pi_{\sqrt{H(x)}\mathcal{A}_0}\sqrt{H(x)^{-1}}$. *Then* $L_2 L_1 v = v$ *for all* $v \in \mathcal{A}_0$.

This follows by the same method as in [5], pag. 545; we leave the proof to the reader. □

Similarly to the classical Legendre type functions theory, the inverse of ϕ_h can be expressed in terms of Fenchel conjugates. For that purpose, we notice that inverting ϕ_h is a minimization problem. Indeed, given $y \in \mathcal{A}_0$, the problem of finding $x \in \mathcal{F}$ such that $y = \Pi_{\mathcal{A}_0}\nabla h(x)$ is equivalent to $x = \mathrm{Argmin}\{h(z) - \langle y, z\rangle | z \in \mathcal{A}\}$, or equivalently

$$x = \mathrm{Argmin}\{(h + \delta_{\mathcal{A}})(z) - \langle y, z\rangle\}, \tag{55}$$

where $\delta_{\mathcal{A}}$ is the *indicator* of $\mathcal{A}$, i.e. $\delta_{\mathcal{A}}(z) = 0$ if $z \in \mathcal{A}$ and $+\infty$ otherwise. Let us recall the definition of *epigraphical sum* of two functions $g_1, g_2 \in \Gamma_0(\mathbb{R}^n)$, which is given by $(g_1 \Box g_2)(y) = \inf\{g_1(u) + g_2(v) | u + v = y\}$, $\forall y \in \mathbb{R}^n$. We have $g_1 \Box g_2 \in \Gamma_0(\mathbb{R}^n)$ and if g_1 and g_2 satisfy $\mathrm{ri\,dom}\, g_1 \cap \mathrm{ri\,dom}\, g_2 \neq \emptyset$ then $(g_1 + g_2)^* = g_1^* \Box g_2^*$ (see [36]).

Proposition 5.3. *We have that* $\phi_h^{-1} : \mathcal{F}^* \to \mathcal{F}$ *is given by* $\phi_h^{-1}(y) = \nabla[h^* \Box (\delta_{\mathcal{A}_0^\perp} + \langle \cdot, \widetilde{x}\rangle)](y)$, *for any* $\widetilde{x} \in \mathcal{A}$, *and moreover* $\mathcal{F}^* = \Pi_{\mathcal{A}_0}\mathrm{int\,dom}\, h^*$.

Proof. The optimality condition for (55) yields $y \in \partial(h + \delta_{\mathcal{A}})(x)$. Thus, $x \in \partial(h + \delta_{\mathcal{A}})^*(y)$. From $\mathcal{F} \neq \emptyset$, we conclude that the function $g \in \Gamma_0(\mathbb{R}^n)$ defined by $g = (h + \delta_{\mathcal{A}})^*$ satisfies $g = h^* \Box \delta_{\mathcal{A}}^* = h^* \Box (\delta_{\mathcal{A}_0^\perp} + \langle \cdot, \widetilde{x}\rangle)$ with $\widetilde{x} \in \mathcal{A}$. Moreover, by [36, Corollary 26.3.2], g is essentially smooth and we deduce that indeed $x = \nabla g(y)$. Since g is essentially smooth, $\mathrm{dom}\, \partial g = \mathrm{int\,dom}\, g$. By definition of epigraphical sum, $g(y) = \inf\{h^*(u) + \delta_{\mathcal{A}_0^\perp}(v) + \langle v, \widetilde{x}\rangle | u + v = y\}$, and consequently we have that $y \in \mathrm{dom}\, g$ iff $y \in \mathrm{dom}\, h^* + \mathcal{A}_0^\perp$. Hence, $\mathrm{int\,dom}\, g = \mathrm{int\,dom}\, h^* + \mathcal{A}_0^\perp$ (see for instance [36, Corollary 6.6.2]). Recalling that $\mathcal{F}^*$ is a relatively open subset of $\mathcal{A}_0$, we deduce that $\mathcal{F}^* = \Pi_{\mathcal{A}_0}\mathrm{dom}\, \partial g = \Pi_{\mathcal{A}_0}\mathrm{int\,dom}\, h^*$. □

5.3 Linear problems in Legendre transform coordinates

5.3.1 Polyhedral sets in Legendre transform coordinates

One of the first interest of Legendre transform coordinates is to transform linear constraints into positive cones.

Proposition 5.4. *Assume that* $C = \{x \in \mathbb{R}^n | Bx > d\}$, *where* B *is a* $p \times n$ *full rank matrix, with* $p \geq n$. *Suppose also that* h *is of the form (39) with* θ *satisfying* (H_1), *and let* $\eta = \lim_{s \to +\infty} \theta'(s) \in (-\infty, +\infty]$. *If* $\eta < +\infty$ *then* $\overline{\mathrm{dom}\, h^*} = \{y \in \mathbb{R}^n \mid y + B^T\lambda = 0,\ \lambda_i \geq -\eta\}$, *and* $\mathrm{dom}\, h^* = \mathbb{R}^n$ *when* $\eta = +\infty$.

Proof. By [37, Theorem 11.5], $\overline{\mathrm{dom}\, h^*} = \{y \in \mathbb{R}^n \mid \langle y, d\rangle \leq h^\infty(d) \text{ for all } d \in \mathbb{R}^n\}$, where h^∞ is the *recession* function, also known as *horizon* function, of h. The recession function is defined by $h^\infty(d) = \lim_{t\to+\infty} \frac{1}{t}[h(\bar{x} + td) - h(\bar{x})]$, $d \in \mathbb{R}^n$, where $\bar{x} \in \mathrm{dom}\, h$; this limit does not depend of $\bar{x}$ and eventually $h^\infty(d) = +\infty$ (see also [36]). In this case, it is easy to verify that $h^\infty(d) = \sum_{i=1}^p \theta^\infty(\langle B_i, d\rangle)$. Clearly, $\theta^\infty(-1) = +\infty$ and $\theta^\infty(1) = \lim_{s\to+\infty} \theta'(s) = \eta$. In particular, if $\eta = +\infty$ then $\mathrm{dom}\, h^* = \mathbb{R}^n$. If $\eta < +\infty$ then $y \in \overline{\mathrm{dom}\, h^*}$ iff for all $d \in \mathbb{R}^n$ such that $Bd \geq 0$, $\langle y, d\rangle \leq h^\infty(d) = \sum_{i=1}^p \eta\langle B_i, d\rangle$, that is $\langle y - \eta B^T e, d\rangle \leq 0$ with $e = (1, \cdots, 1)$. Thus, by the Farkas lemma, $y \in \overline{\mathrm{dom}\, h^*}$ iff $\exists \mu \geq 0$, $y - \eta B^T e + B^T \mu = 0$. □

As a direct consequence of Propositions 5.3 and 5.4:

Corollary 5.1. *Under the assumptions of Proposition 5.4, if $\eta = 0$ then $\mathcal{F}^*$ is a positive convex cone and if $\eta = +\infty$ then $\mathcal{F}^* = \mathcal{A}_0$.*

5.3.2 $(H\text{-}SD)$-trajectories in Legendre transform coordinates

In the sequel, we assume that $f(x) = \langle c, x\rangle$ for some $c \in \mathbb{R}^n$. As another striking application of Legendre transform coordinates, we prove now that the trajectories of $(H\text{-}SD)$ may be seen as straight lines in $\mathcal{F}^*$. Recall that the *push forward* vector field of $\nabla_H f_{|\mathcal{F}}$ by ϕ_h is defined for every $y \in \mathcal{F}^*$ by $[(\phi_h)_* \nabla_H f_{|\mathcal{F}}](y) = d\phi_h(\phi_h^{-1}(y))\nabla_H f_{|\mathcal{F}}(\phi_h^{-1}(y))$.

Proposition 5.5. *For all $y \in \mathcal{F}^*$, $[(\phi_h)_* \nabla_H f_{|\mathcal{F}}](y) = \Pi_{\mathcal{A}_0} c$.*

Proof. Let $y \in \mathcal{F}^*$. Setting $x = \phi_h^{-1}(y)$, by Theorem 5.1 we get $[(\phi_h)_* \nabla_H f_{|\mathcal{F}}](y) = d\phi_h(x)\nabla_H f_{|\mathcal{F}}(x) = \Pi_{\mathcal{A}_0} H(x)H(x)^{-1}[I - A^T(AH(x)^{-1}A^T)^{-1}AH(x)^{-1}]c = \Pi_{\mathcal{A}_0} c - \Pi_{\mathcal{A}_0} A^T z$, where $z = [(AH(x)^{-1}A^T)^{-1}AH(x)^{-1}]c$. Since $\mathrm{Im}\, A^T = \mathcal{A}_0^\perp$, the conclusion follows. □

Next, we give two optimality characterizations of the orbits of $(H\text{-}SD)$, extending thus to the general case the results of [5] for the log-metric.

5.3.3 Geodesic curves

First, we claim that the orbits of $(H\text{-}SD)$ can be regarded as geodesics curves with respect to some appropriate metric on $\mathcal{F}$. To this end, we endow $\mathcal{F}^* = \phi_h(\mathcal{F})$ with the Euclidean metric, which allows us to define on $\mathcal{F}$ the metric

$$(\cdot, \cdot)^{H^2} = (\phi_h)^* \langle \cdot, \cdot \rangle, \tag{56}$$

that is, $\forall (x, u, v) \in \mathcal{F} \times \mathbb{R}^n \times \mathbb{R}^n$, $(u, v)_x^{H^2} = \langle d\phi_h(x)u, d\phi_h(x)v\rangle = \langle \Pi_{\mathcal{A}_0} H(x)u, \Pi_{\mathcal{A}_0} H(x)v\rangle$. For each initial condition $x^0 \in \mathcal{F}$, and for every $c \in \mathbb{R}^n$ we set

$$v = d\phi_h(x^0)^{-1}\Pi_{\mathcal{A}_0} c = \sqrt{H(x^0)^{-1}}\Pi_{\sqrt{H(x^0)}\mathcal{A}_0}\sqrt{H(x^0)^{-1}}\Pi_{\mathcal{A}_0} c. \tag{57}$$

Theorem 5.2. *Let $(x^0, c) \in \mathcal{F} \times \mathbb{R}^n$, set $f(x) = \langle c, x\rangle$, $\forall x \in C$ and define v as in (57). If $\mathcal{F}$ is endowed with the metric $(\cdot, \cdot)^{H^2}$ given by (56), then the solution $x(t)$ of $(H\text{-}SD)$ is the unique geodesic passing through x^0 with velocity v.*

Proof. Since $\mathcal{F}$, $(\cdot,\cdot)^{H^2}$ is isometric to the Euclidean Riemannian manifold $\mathcal{F}^*$, the geodesic joining two points of $\mathcal{F}$ exists and is unique. Let us denote by $\gamma : J \subset \mathbb{R} \mapsto \mathcal{F}$ the geodesic passing through x^0 with velocity v. By definition of $(\cdot,\cdot)^{H^2}$, $\phi_h(\gamma)$ is a geodesic in $\mathcal{F}^*$. Whence $\phi_h(\gamma(t)) = \phi_h(x^0) + td\phi_h(x^0)v$, where $t \in J$. In view of (57), this can be rewritten $\phi_h(\gamma(t)) = \phi_h(x^0) + t\Pi_{\mathcal{A}_0}c$. By Proposition 5.5 we know that $(\phi_h)_* \nabla_H f_{|\mathcal{F}} = \Pi_{\mathcal{A}_0}c$, and therefore $\phi_h^{-1}(\phi_h(\gamma)) = \gamma$ is exactly the solution of $(H\text{-}SD)$. □

Remark 5.1. A Riemannian manifold is called *geodesically complete* if the maximal interval of definition of every geodesic is $\mathbb{R}$. When $\Pi_{\mathcal{A}_0}c \neq 0$ and $\mathcal{F}^*$ is not an affine subspace of $\mathbb{R}^n$, the Riemannian manifold $\mathcal{F}$, $(\cdot,\cdot)^{H^2}$ is not complete in this sense.

5.3.4 Lagrange equations

Following the ideas of [5], we describe the orbits of $(H\text{-}SD)$ as orthogonal projections on $\mathcal{A}$ of $\dot{q}-$trajectories of a specific *Lagrangian system.* Recall that given a real-valued mapping $\mathcal{L}(q,\dot{q})$ called the Lagrangian, where $q = (q_1,\dots,q_n)$ and $\dot{q} = (\dot{q}_1,\dots,\dot{q}_n)$, the associated Lagrange equations of motion are the following

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{q}_i} = \frac{\partial \mathcal{L}}{\partial q_i}, \quad \frac{d}{dt}q_i = \dot{q}_i, \quad \forall i = 1\dots n. \tag{58}$$

Their solutions are C^1-piecewise paths $\gamma : t \longmapsto (q(t),\dot{q}(t))$, defined for $t \in J \subset \mathbb{R}$, that satisfy (58), and appear as extremals of the functional $\widehat{\mathcal{L}}(\gamma) = \int_J \mathcal{L}(q(t),\dot{q}(t))dt$. Notice that in general, the solutions are not unique, in the sense that they do not only depend on the initial condition $\gamma(0)$. Let us introduce the Lagrangian $\mathcal{L} : \mathbb{R}^n \times C \to \mathbb{R}$ defined by

$$\mathcal{L}(q,\dot{q}) = \langle \Pi_{\mathcal{A}_0}c, q\rangle - h(\Pi_{\mathcal{A}}\dot{q}), \tag{59}$$

where $\Pi_{\mathcal{A}}$ is the orthogonal projection onto $\mathcal{A}$, i.e. $\Pi_{\mathcal{A}}x = \widetilde{x} + \Pi_{\mathcal{A}_0}(x - \widetilde{x})$ for any $\widetilde{x} \in \mathcal{A}$.

Theorem 5.3. *For any solution $\gamma(t) = (q(t),\dot{q}(t))$ of the Lagrangian dynamical system* (58) *with Lagrangian given by* (59)*, the projection $x(t) = \Pi_{\mathcal{A}}\dot{q}(t)$ is the solution of (H-SD) with initial condition $x^0 = \Pi_{\mathcal{A}}\dot{q}(0)$.*

Proof. It is easy to verify that $\nabla(h \circ \Pi_{\mathcal{A}})(x) = \Pi_{\mathcal{A}_0}\nabla h(\Pi_{\mathcal{A}}x)$ for any $x \in \mathbb{R}^n$. Given a solution $\gamma(t) = (q(t),\dot{q}(t))$ of (59) defined on J, we set $p(t) = (p_1(t),\dots,p_n(t)) = \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_1}(\gamma(t)),\dots,\frac{\partial \mathcal{L}}{\partial \dot{q}_n}(\gamma(t))\right)$. We have $p(t) = \nabla(h\circ\Pi_{\mathcal{A}})(\dot{q}(t)) = \Pi_{\mathcal{A}_0}\nabla h(\Pi_{\mathcal{A}}\dot{q}(t)) = \phi_h(\Pi_{\mathcal{A}}\dot{q}(t))$. Equations of motion become $\frac{d}{dt}p(t) = \Pi_{\mathcal{A}_0}c$, that is, $\frac{d}{dt}\phi_h(\Pi_{\mathcal{A}}\dot{q}(t)) = \Pi_{\mathcal{A}_0}c$. Since $\phi_h : \mathcal{F} \to \mathcal{F}^*$ is a diffeomorphism, the latter means, according to Proposition 5.5, that $\Pi_{\mathcal{A}}\dot{q}(t)$ is a trajectory for the vector field $\nabla_H f_{|\mathcal{F}}$. Notice that C being convex, as soon as $\dot{q}(0) \in C$, $\Pi_{\mathcal{A}}\dot{q}(0) \in C \cap \mathcal{A} = \mathcal{F}$, and what precedes forces $\Pi_{\mathcal{A}}\dot{q}(t)$ to stay in $\mathcal{F}$ for any $t \in J$. □

5.3.5 Completely integrable Hamiltonian systems

In the sequel, all mappings are supposed to be at least of class $\mathcal{C}^2$. Let us first recall the notion of Hamiltonian system. Given an integer $r \geq 1$ and a real-valued mapping $\mathcal{H}(q,p)$ on $\mathbb{R}^{2r}$ with coordinates $(q,p) = (q_1,\ldots,q_r,p_1,\ldots,p_r)$, the *Hamiltonian vector field* $X_{\mathcal{H}}$ *associated with* $\mathcal{H}$ is defined by $X_{\mathcal{H}} = \sum_{i=1}^{r} \frac{\partial \mathcal{H}}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial \mathcal{H}}{\partial q_i}\frac{\partial}{\partial p_i}$. The trajectories of the dynamical system induced by $X_{\mathcal{H}}$ are the solutions to

$$\left\{ \begin{array}{l} \dot{p}_i(t) = -\frac{\partial}{\partial q_i}\mathcal{H}(q(t),p(t)),\ i=1,\ldots,r, \\ \dot{q}_i(t) = \frac{\partial}{\partial p_i}\mathcal{H}(q(t),p(t)),\ i=1,\ldots,r. \end{array} \right. \tag{60}$$

Following a standard procedure, Lagrangian functions $\mathcal{L}(q,\dot{q})$ are associated with Hamiltonian systems by means of the so-called Legendre transform

$$\Phi : \left\{ \begin{array}{ccc} \mathbb{R}^{2r} & \longrightarrow & \mathbb{R}^{2r} \\ (q,\dot{q}) & \longmapsto & (q, \frac{\partial \mathcal{L}}{\partial \dot{q}}(q,\dot{q})) \end{array} \right.$$

In fact, when Φ is a diffeomorphism, the Hamiltonian function $\mathcal{H}$ associated with the Lagrangian $\mathcal{L}$ is defined on $\Phi(\mathbb{R}^{2r})$ by $\mathcal{H}(p,q) = \sum_{i=1}^{r} p_i\dot{q}_i - \mathcal{L}(q,\dot{q}) = \langle p, \psi^{-1}(q,p)\rangle - \mathcal{L}(q,\psi^{-1}(q,p))$, where $(q,\psi^{-1}(q,p)) := \Phi^{-1}(q,p)$. With these definitions, Φ sends the trajectories of the corresponding Lagrangian system on the trajectories of the Hamiltonian system (60).

In general, the Lagrangian (59) does not lead to an invertible Φ on $\mathbb{R}^{2n}$. However, we are only interested in the projections $\Pi_{\mathcal{A}}\dot{q}$ of the trajectories, which, according to Theorem 5.3, take their values in $\mathcal{F}$. Moreover, notice that for any differentiable path $t \mapsto q^{\perp}(t)$ lying in $\mathcal{A}_0^{\perp}$, $t \mapsto (q(t),\dot{q}(t))$ is a solution of (58) iff $t \mapsto (q(t)+q^{\perp}(t), \dot{q}(t)+\dot{q}^{\perp}(t))$ is. This legitimates the idea of restricting $\mathcal{L}$ to $\mathcal{A}_0 \times \Pi_{\mathcal{A}_0}\mathcal{F}$. Hence and from now on, $\mathcal{L}$ denotes the function:

$$\mathcal{L} : \left\{ \begin{array}{ccc} \mathcal{A}_0 \times \Pi_{\mathcal{A}_0}\mathcal{F} & \longrightarrow & \mathbb{R} \\ (q,\dot{q}) & \longmapsto & \mathcal{L}(q,\dot{q}). \end{array} \right. \tag{61}$$

Taking $(q_1,\ldots,q_r)$, with $r = n-m$, a linear system of coordinates induced by an Euclidean orthonormal basis for $\mathcal{A}_0$, we easily see that this "new" Lagrangian has trajectories $(q(t),\dot{q}(t))$ lying in $\mathcal{A}_0 \times \Pi_{\mathcal{A}_0}\mathcal{F}$, whose projections $\Pi_{\mathcal{A}}\dot{q}(t)$ are exactly the (*H-SD*) trajectories. Moreover, an easy computation yields

$$\frac{\partial \mathcal{L}}{\partial \dot{q}}(q,\dot{q}) = \Pi_{\mathcal{A}_0}\nabla h(\Pi_{\mathcal{A}_0}\dot{q}) = [\phi_h \circ \Pi_{\mathcal{A}}](\dot{q}),$$

which is a diffeomorphism by Proposition 5.1. The Legendre transform is then given by

$$\Phi : \left\{ \begin{array}{ccc} \mathcal{A}_0 \times \Pi_{\mathcal{A}_0}\mathcal{F} & \longrightarrow & \mathcal{A}_0 \times \mathcal{F}^* \\ (q,\dot{q}) & \longmapsto & (q, [\phi_h \circ \Pi_{\mathcal{A}}](\dot{q})), \end{array} \right.$$

and therefore, $\mathcal{L}$ is converted into the Hamiltonian system associated with

$$\mathcal{H} : \left\{ \begin{array}{ccc} \mathcal{A}_0 \times \mathcal{F}^* & \longrightarrow & \mathbb{R} \\ (q,p) & \longmapsto & \langle p, [\phi_h \circ \Pi_{\mathcal{A}}]^{-1}(p)\rangle - \mathcal{L}(q, [\phi_h \circ \Pi_{\mathcal{A}}]^{-1}(p)). \end{array} \right. \tag{62}$$

Let us now introduce the concept of completely integrable Hamiltonian system. The Poisson bracket of two real valued functions f_1, f_2 on $\mathbb{R}^{2r}$ is given by $\{f_1, f_2\} = \sum_{i=1}^{r} \frac{\partial f_1}{\partial p_i}\frac{\partial f_2}{\partial q_i} - \frac{\partial f_1}{\partial q_i}\frac{\partial f_2}{\partial p_i}$. Notice that, from the definitions, we have $\{f_1, f_2\} = X_{f_1}(f_2)$ and $X_{\{f_1,f_2\}} = [X_{f_1}, X_{f_2}]$, where $[\cdot,\cdot]$ is the standard *bracket product* of vector fields [33]. Now, the system (60) is called *completely integrable* if there exist r functions $f_1, \ldots, f_r$ with $f_1 = \mathcal{H}$, satisfying

$$\begin{cases} \{f_i, f_j\} = 0, \quad \forall i, j = 1, \ldots, r. \\ df_1(x), \ldots, df_r(x) \text{ are linearly independent at any } x \in \mathbb{R}^{2r}. \end{cases}$$

As a motivation for completely integrable systems, we will just point out the following: the functions f_i are called *integrals of motions* because $X_{\mathcal{H}}(f_i) = \{h, f_i\} = 0$, which means that any trajectory of $X_{\mathcal{H}}$ lies on the level sets of each f_i (the same holds for all X_{f_j}). Also, the trajectory passing through (q_0, p_0) lies in the set $\bigcap_{i=1\ldots r} f_i^{-1}(\{f_i(q_o, p_0)\})$. Besides, $[X_{f_i}, X_{f_j}] = 0$ implies that we can find, at least locally, coordinates $(x_1, \ldots, x_r)$ on this set such that $X_{\mathcal{H}} = \frac{\partial}{\partial x_1}, X_{f_2} = \frac{\partial}{\partial x_2}, \ldots, X_{f_r} = \frac{\partial}{\partial x_r}$, that is, in these coordinates, the trajectories of X_{f_i} are straight lines.

Theorem 5.4. *Suppose $\Pi_{\mathcal{A}_0} c \neq 0$. The Lagrangian system on $\mathcal{A}_0 \times \Pi_{\mathcal{A}_0}\mathcal{F}$ associated with* (59), (61) *gives rise, by the Legendre transform, to a completely integrable Hamiltonian system on $\mathcal{A}_0 \times \mathcal{F}^*$ with Hamiltonian given by* (62).

Proof. There only remains to prove the complete integrability of the system. To this end, we adapt the proof of [5, Theorem II.12.2] to our abstract framework. Take the integrals of motion to be $f_1 = \mathcal{H}$, $f_i(q,p) = \langle c_i, p\rangle$, $i = 2, \ldots, r$ where $r = n - m$ and $\{\Pi_{\mathcal{A}_0} c,\ c_2, \ldots, c_r\}$ is chosen as to be an orthonormal basis of $\mathcal{A}_0$. For any $i, j \in \{2, \ldots, r\}$, $\{f_i, f_j\}$ is zero since f_i and f_j only depend on p. Let $\phi_{h,l}^{-1}(q,p)$ (resp. $(\Pi_{\mathcal{A}_0} c)_l$) stand for the l-th component of $\phi_h^{-1}(q,p)$ (resp. the l-th component of $\Pi_{\mathcal{A}_0} c$) and take some $k \in \{1, ..., r\}$. Since

$$\begin{aligned} \frac{\partial \mathcal{H}}{\partial q_k}(q,p) &= \frac{\partial(\sum_{l=1}^{r} p_l \phi_{h,l}^{-1})}{\partial q_k}(q,p) - \frac{\partial(\mathcal{L} \circ \Phi^{-1})}{\partial q_k}(q,p) \\ &= \sum_{l=1}^{r} p_l \frac{\partial \phi_{h,l}^{-1}}{\partial q_k}(p,q) - \frac{\partial \mathcal{L}}{\partial q_k}(q, \phi_h^{-1}(q,p)) - \sum_{l=1}^{r} \frac{\partial \mathcal{L}}{\partial \dot{q}_l}(q, \phi_h^{-1}(q,p)) \frac{\partial \phi_{h,l}}{\partial q_k}(q,p) \\ &= -(\Pi_{\mathcal{A}_0} c)_k \end{aligned}$$

we deduce that for all $i \in \{2, ..., r\}$, $\{\mathcal{H}, f_i\} = \sum_{k=1}^{r} -\frac{\partial f_i}{\partial p_k}\frac{\partial \mathcal{H}}{\partial q_k} = \langle \Pi_{\mathcal{A}_0} c, c_i\rangle = 0$. The second condition for complete integrability is satisfied too, as the $r \times 2r$ matrix

$$\left([\frac{\partial f_i}{\partial q_1}, \ldots, \frac{\partial f_i}{\partial q_r}, \frac{\partial f_i}{\partial p_1}, \ldots, \frac{\partial f_i}{\partial p_r}] \right)_{i=1,\ldots,r} = \begin{pmatrix} \Pi_{\mathcal{A}_0} c^T & \star \\ & c_1^T \\ 0 & \ldots \\ & c_r^T \end{pmatrix}$$

is full rank. $\square$

References

[1] E. Akin, "The geometry of population genetics", Lecture Notes in Biomathematics 31, Springer-Verlag, Berlin, 1979.

[2] H. Attouch, *Viscosity solutions of minimization problems*, SIAM J. Optim., 6 (1996), No. 3, pp. 769-806.

[3] H. Attouch and M. Teboulle, *A regularized Lotka Volterra dynamical system as a continuous proximal-like method in optimization*, December 2001. Submitted.

[4] A. Auslender, R. Cominetti and M. Haddou, *Asymptotic analysis for penalty and barrier methods in convex and linear programming*, Math. Oper. Res., 22 (1997), pp. 43-62.

[5] D.A. Bayer and J.C. Lagarias, *The nonlinear geometry of linear programming I. Affine and projective scaling trajectories; II. Legendre transform coordinates and central trajectories* , Trans. Amer. Math. Soc., 314 (1989), No. 2, pp. 499-526 and 527-581.

[6] J. Bolte and M. Teboulle, *Barrier operators and associated gradient-like dynamical systems for constrained minimization problems*, submitted (June 2002).

[7] L.M. Bregman, *The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming*, Zh. Vychisl. Mat. i Mat. Fiz., 7 (1967), pp. 620-631 (in Russian). English transl. in U.S.S.R. Comput. Math. and Math. Phys., 7 (1967), pp. 200-217.

[8] R.W. Brockett, *Dynamical systems that sort lists and solve linear programming problems*, Proc. IEEE Conf. Decision and Control, Austin, Texas, 1988, pp. 779-803.

[9] R.W. Brockett, *Dynamical systems that sort lists, diagonalize matrices and solve linear programming problems*, Linear Alg. Appl., 146 (1991), pp. 79-91.

[10] R.E. Bruck, *Asymptotic convergence of non linear contraction semi-groups in Hilbert space*, J. Func. Anal., 18 (1974), pp 15-26.

[11] Y. Censor and A. Lent, *An iterative row action method for interval convex programming*, J. Optim. Theory Appl., 34 (1981), pp. 321-353.

[12] Y. Censor and S.A. Zenios, *Proximal minimization algorithm with D-functions*, J. Optim. Theory Appl., 73 (1992), pp. 451-464.

[13] G. Chen and M. Teboulle, *Convergence analysis of a proximal-like optimization algorithm using Bregman functions*, SIAM J. Optim., 3 (1993), pp. 538-543.

[14] R. Cominetti, *Nonlinear average and convergence of penalty trajectories in convex programming*, in "Ill-posed variational problems and regularization techniques (Trier, 1998)", Lecture Notes in Econom. and Math. Systems 477, Springer, Berlin, 1999, pp. 65-78.

[15] R. Cominetti and J. San Martín, *Asymptotic Analysis of the Exponential Penalty Trajectory in Linear Programming*, Math. Programming, 67 (1994), pp. 169-187.

[16] M.P. do Carmo, "Riemannian Geometry (Mathematics, Theory and Applications)", Birkhäuser, Boston, 1992.

[17] J.J. Duistermaat, *On Hessian Riemannian structures*, Asian J. Math., 5 (2001), No. 1, pp. 79-91.

[18] L.E. Faybusovich, *Dynamical systems which solve optimization problems with linear constraints*, IMA J. Math. Control and Inf., 8 (1991), pp. 135-149.

[19] L.E. Faybusovich, *Hamiltonian structure of dynamical systems which solve linear programming problems*, Phys. D, 53 (1991), pp. 217-232.

[20] L.E. Faybusovich, *Interior point methods and entropy*, Proc. IEEE Conf. Decision and Control, Tucson, Arizona, 1992, pp. 1626-1631.

[21] A.V. Fiacco, *Perturbed variations of penalty function methods. Example: Projective SUMT*, Annals of Oper. Res., 27 (1990), pp. 371-380.

[22] U. Helmke and J.B. Moore, "Optimization and Dynamical Systems", Springer-Verlag, London, 1994.

[23] S. Herzel, M.C. Recchini and F. Zirilli, *A quadratically convergent method for linear programming*, Linear Alg. Appl., 151 (1991), pp. 255-290.

[24] J.B. Hiriart-Urruty and C. Lemaréchal, "Convex Analysis and Minimization Algorithms II", Springer-Verlag, Berlin, 1996.

[25] J. Hofbauer and K. Sigmund, "Evolutionary Games and Population Dynamics", Cambridge University Press, 1998.

[26] A.N. Iusem and R.D.C. Monteiro, *On dual convergence of the generalized proximal point method with Bregman distances*, Math. Oper. Res., 25 (2000), No. 4, pp. 606-624.

[27] 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 (1999), No. 2, pp. 566-588.

[28] N. Karmarkar, *A new polynomial time algorithm for linear programming*, Combinatorica 4 (1984), pp. 373-395.

[29] N. Karmarkar, *Riemannian geometry underlying interior point methods for linear programming*, in "Mathematical Developments Arising from Linear Programming", Contemporary Mathematics 114, J.C. Lagarias and M.J. Todd (eds.), AMS, Providence, RI, 1990, pp. 51-76.

[30] N. Kenmochi, and I. Pawlow, *A class of doubly nonlinear elliptic-parabolic equations with time dependent constraints*, Nonlinear Analysis, 10 (1986),pp 1181-1202

[31] K.C. Kiwiel, *Free-steering relaxation methods for problems with strictly convex costs*, Math. Oper. Res., 22 (1997), No. 2, pp. 326-349.

[32] K.C. Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142-1168.

[33] S. Lang, "Differential and Riemannian Manifolds", Springer-Verlag, New York, 1995.

[34] G.P. McCormick, *The continuous Projective SUMT method for convex programming*, Math. Oper. Res. 14, (1989), No. 2, pp. 203-223.

[35] J. Palis and W. De Melo, "Geometric theory of dynamical systems", Springer, 1982.

[36] R.T. Rockafellar, "Convex Analysis", Princeton University Press, Princeton, NJ, 1970.

[37] R.T. Rockafellar and J-B. R. Wets, "Variational Analysis", Grundlehren der mathematischen Wissenschaften 317, Springer-Verlag, Berlin (1998).

[38] M. Teboulle, *Entropic proximal mappings with applications to nonlinear programming*, Math. Oper. Res., 17 (1992), pp. 670-690.