Convergence to the optimal value for barrier methods combined with Hessian Riemannian gradient flows and generalized proximal algorithms*

Felipe Alvarez & Julio López †

Abstract

We consider the problem $\min_{x\in\mathbb{R}^n}\{f(x) \mid Ax = b,\ x \in \overline{C},\ g_j(x) \leq 0,\ j = 1,\ldots,s\}$, where $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$ is a full rank matrix, $\overline{C}$ is the closure of a nonempty, open and convex subset C of $\mathbb{R}^n$, and $g_j(\cdot)$, $j = 1,\ldots,s$, are nonlinear convex functions. Our strategy consists firstly in to introduce a barrier-type penalty for the constraints $g_j(x) \leq 0$, then endowing $\{x \in \mathbb{R}^n \mid Ax = b, x \in C\}$ with the Riemannian structure induced by the Hessian of an essentially smooth convex function h such that $C = \mathrm{int}(\mathrm{dom}\, h)$, and finally considering the flow generated by the Riemannian penalty gradient vector field. Under minimal hypotheses, we investigate the well-posedness of the resulting ODE and we prove that the value of the objective function along the trajectories, which are strictly feasible, converges to the optimal value. Moreover, the value convergence is extended to the sequences generated by an implicit discretization scheme which corresponds to the coupling of an inexact generalized proximal point method with parametric barrier schemes. Specializations and simple illustrations of the general results are given for the positive orthant, the unitary simplex and the second-order cone.

1 Introduction

In this paper we treat a general mathematical programming problem of the type

$$(P) \qquad v(P) \equiv \min\{f(x) \mid Ax = b,\ x \in \overline{C},\ g_j(x) \leq 0,\ j \in I\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_j : \mathbb{R}^m \to \mathbb{R}$ for $j \in I := \{1,\ldots,s\}$ are continuously differentiable convex functions, $A \in \mathbb{R}^{m\times n}$ is a full rank matrix with $m \leq n$, $b \in \mathbb{R}^m$ and $\overline{C}$ is the closure of a nonempty, open and convex set $C \subset \mathbb{R}^n$. The idea here is to distinguish among three kinds of constraints: linear equality constraints of the form $Ax = b$, polyhedral or conic constraints represented by the convex inclusion $x \in \overline{C}$, and additional nonlinear inequality constraints $g_j(x) \leq 0$, $j \in I$.

In order to motivate our approach, let us start with the unconstrained minimization of f. Given a starting point $x^0 \in \mathbb{R}^n$, the gradient algorithm corrects the current iterate by following the steepest descent direction, that is, $x^{k+1} = x^k - \lambda_k \nabla f(x^k)$, where $\lambda_k > 0$ is an appropriate stepsize and ∇f stands for the Euclidean gradient of f. This can be viewed as a discretization of the *continuous-in-time gradient method* given by $\frac{du}{dt}(t) = -\nabla f(u(t)),\ t > 0$, see [19]. An implicit discretization of

*This work was partially supported by CONICYT (Fondecyt 3100131 and Basal grants) and by the Institute on Complex Engineering Systems funded by MIDEPLAN Millennium Scientific Initiative (grant P05-004-F).

†Departamento de Ingeniería Matemática, Centro de Modelamiento Matemático (CNRS UMI 2807), FCFM, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile (falvarez@dim.uchile.cl, jclopez@dim.uchile.cl).

this ODE yields to $x^{k+1} = x^k - \lambda_k \nabla f(x^{k+1})$, which for a convex function is equivalent to an exact iteration of the Euclidean *proximal point algorithm* [28]: $x^{k+1} = \operatorname{argmin}\left\{ f(y) + \frac{1}{2\lambda_k}\|y - x^k\|^2 \right\}$. The latter makes sense even for a nonsmooth objective function. Under several conditions, these methods share some qualitative and asymptotic convergence properties [18, 28].

Suppose now that we are interested in minimizing f on the positive orthant $\mathbb{R}^n_+$. According to the ODE approach, we can consider the following *scaled* version of the gradient method:

$$\frac{du_i}{dt}(t) = -u_i(t)\frac{\partial f}{\partial x_i}(u(t)),\ i = 1, \ldots, n. \tag{1}$$

This equation generates *strictly feasible* trajectories: if $u_i(0) = x_i^0 > 0$ then $u_i(t) > 0$ for all $t > 0$. This is also a descent method as long as $u_i(t) > 0$; in fact $\frac{d}{dt}f(u(t)) = -\sum_{i=1}^n u_i(t)\left|\frac{\partial f}{\partial x_i}(u(t))\right|^2 \leq 0$. The dynamical system given by (1) is a special case of the so called *Hessian Riemannian gradient flows* [2]. More precisely, it turns out that (1) can be expressed as

$$\frac{du}{dt}(t) = -\nabla^2 h(u(t))^{-1}\nabla f(u(t)) \tag{2}$$

where $\nabla^2 h(x)$ stands for the Hessian matrix at x of the function $h(x) = \sum_{i=1}^n x_i \log x_i - x_i$. The vector field on $\mathbb{R}^n_{++}$ defined by $\operatorname{grad}_h f(x) = \nabla^2 h(x)^{-1}\nabla f(x)$ is precisely the gradient of f with respect to the Riemannian metric given by

$$\forall v, w \in \mathbb{R}^n,\ (v, w)_x := \langle \nabla^2 h(x) v, w\rangle. \tag{3}$$

In this special case we obtain $(v, w)_x = \sum_{i=1}^n v_i w_i / x_i$ for $x \in \mathbb{R}^n_{++}$. On the other hand, notice that $\frac{du}{dt} = -\nabla^2 h(u)^{-1}\nabla f(u) \Leftrightarrow \frac{d}{dt}\nabla h(u) + \nabla f(u) = 0$. Thus an implicit discretization of (2) yields $\frac{1}{\lambda_k}[\nabla h(x^{k+1}) - \nabla h(x^k)] + \nabla f(x^{k+1}) = 0$, which is is an exact iteration of the *generalized proximal point algorithm* [9, 11, 12, 23].

The well-posedness, asymptotic behavior as $t \to +\infty$ and other properties of autonomous gradient systems as (2) are investigated in [2]; see also [5, 10]. More precisely, in [2] it is considered the case where the optimization problem is of the form $\min\{f(x) \mid Ax = b,\ x \in \overline{C}\}$, by assuming that there exists an essentially smooth convex function h_C such that in particular $C = \operatorname{int}(\operatorname{dom} h_C)$, whose Hessian matrix $\nabla^2 h_C(x)$ is used to endow the manifold $\{x \in \mathbb{R}^n \mid Ax = b,\ x \in C\}$ with a Hessian Riemannian metric $(\cdot,\cdot)_x$ given by (3) for $h = h_C$. Setting

$$\mathcal{A} = \{x \in \mathbb{R}^n \mid Ax = b\}, \tag{4}$$

the corresponding gradient flow on $\mathcal{A} \cap C$ is then given by $\dfrac{du}{dt}(t) = -\operatorname{grad}_{h_C} f(u(t))$, where $\operatorname{grad}_{h_C} f : \mathcal{A} \cap C \to \ker A$ is the Riemannian gradient with respect to $(\cdot,\cdot)_x$ of f restricted to $\mathcal{A} \cap C$; see Section 2 for more details. The connection of this flow with generalized proximal point algorithms and central paths in linear programming was first investigated in [21].

Let us return to the minimization problem (P). In order to handle the additional nonlinear constraints $g_j(x) \leq 0$ for $j \in J$, we set

$$G := \{x \in \mathbb{R}^n \mid g_j(x) < 0,\ j \in I\}, \tag{5}$$

which is supposed to be nonempty. Next, we introduce a barrier-type penalty function $\theta(\cdot)$ (*e.g.* the *inverse barrier* function $\theta(s) = -1/s$ if $s < 0$ and $+\infty$ otherwise). Under appropriate conditions, following the approach of [2], we may define $h_{C\cap G}(x) := h_C(x) + \sum_{i\in J}\theta(g_j(x))$, and endow $\mathcal{A}\cap C\cap G$, the relative interior of the feasible set of (P), with the Hessian Riemannian metric induced by $h_{C\cap G}$. This yields to the following gradient flow (GF for short) on $\mathcal{A}\cap C\cap G$:

$$(GF) \qquad \frac{du}{dt}(t) = -\mathrm{grad}_{\,h_{C\cap G}} f(u(t))$$

On the other hand, inspired by the coupling of the Euclidean gradient method with penalty schemes [4, 8, 14], we introduce a parametric penalty function as follows: $f_\varepsilon(x) := f(x) + \varepsilon\sum_{j\in I}\theta(g_j(x)/\varepsilon)$, $\varepsilon > 0$, and consider the following hybrid barrier-gradient flow (B-GF) on $\mathcal{A}\cap C\cap G$:

$$(B\text{-}GF) \qquad \frac{du}{dt}(t) = -\mathrm{grad}_{\,h_C} f_{\varepsilon(t)}(u(t)),$$

for a suitable parameterization $\varepsilon : [0,+\infty) \to (0,+\infty)$ such that $\varepsilon(t) \to 0$ as $t \to \infty$.

In this paper we focus on the non-autonomous ODE given by $(B\text{-}GF)$ and some related numerical discretization schemes, providing some comparisons with the alternative approach given by (GF). More precisely, in Section 2 we recall some basic definitions concerning variable metric gradient flows and the coupling with barrier-type penalties, and we identify some hypotheses for our results. In Section 3, we state and prove our main result on the global existence for all $t \geq 0$ of the solution $u(t)$ to $(B\text{-}GF)$, and the convergence of $f(u(t))$ to the optimal value $v(P)$ as $t \to \infty$. In Section 4, we introduce a generalized barrier proximal point algorithm as an implicit discretization in time of $(B\text{-}GF)$, and establish an analogous value convergence result. In Section 5, we discuss some simple specializations of (GF) and $(B\text{-}GF)$ for the positive orthant, the unitary simplex and the second-order cone, and we also compare the behavior of these flows through an explicit Euler's scheme on some toy examples. Finally, in Section 6 we end the paper with some remarks on open questions.

2 Preliminaries

2.1 Quick review on Hessian Riemannian gradient flows

Let $Q \subset \mathbb{R}^n$ be a nonempty, open and convex set. We denote by $\mathcal{S}^n_{++}$ the cone of positive definite symmetric $n\times n$ real matrices. Let $h : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a lower semicontinuous and convex function with effective domain $\mathrm{dom} h = \{x \in \mathbb{R}^n \mid h(x) < +\infty\}$. We assume the following conditions:

$$(H_h;Q) \quad \begin{cases} (a) & Q = \mathrm{int}(\mathrm{dom}\, h). \\ (b) & h|_Q \in \mathcal{C}^2(Q;\mathbb{R}) \text{ and } \forall x \in Q, \nabla^2 h(x) \in \mathcal{S}^n_{++}. \\ (c) & \text{The map } x \mapsto \nabla^2 h(x) \text{ is locally Lipschitz continuous on } Q. \\ (d) & \forall \bar{x} \in \partial Q,\ \forall x^k \to \bar{x} \text{ with } x^k \in Q,\ \|\nabla h(x^k)\| \to +\infty. \end{cases}$$

In particular, h is essentially smooth and of Legendre type [27, Chapter 26].

Example 2.1. Examples of functions satisfying $(H_h;Q)$ for different choices of Q are the following:

- $h_1(x) = \frac{1}{2}\|x\|^2 = \frac{1}{2}\sum_{i=1}^n x_i^2$, for $Q = \mathbb{R}^n$.
- $h_2(x) = -\sum_{i=1}^n \sqrt{1-x_i^2}$, for $Q = (-1,1)^n$.
- $h_3(x) = \sum_{i=1}^n x_i \log x_i - x_i$ and $h_4(x) = -\sum_{i=1}^n \log x_i$, both for $Q = \mathbb{R}^n_{++}$.
- Let $Q = \mathcal{L}^n_{++}$ be the interior of the second-order (or Lorentz) cone $\mathcal{L}^n_+$ in $\mathbb{R}^n$, which is given by $\mathcal{L}^n_+ = \{x = (x_1,\bar{x}) \in \mathbb{R}\times\mathbb{R}^{n-1} \mid \|\bar{x}\| \leq x_1\ \}$. For $x = (x_1,\bar{x}) \in \mathbb{R}\times\mathbb{R}^{n-1}$ take the following functions $h_5(x) = -\log\det(x)$ for $x \in \mathcal{L}^n_{++}$ and $h_6(x) = \mathrm{tr}(x\circ\log(x)) - \mathrm{tr}(x)$ for $x \in \mathcal{L}^n_+$, where $\det(x) = x_1^2 - \|\bar{x}\|^2$, $\mathrm{tr}(x) = 2x_1$ and $\circ$ denotes the *Jordan product* for $\mathcal{L}^n_+$ which is defined by $x\circ y = (x^\top y, x_1\bar{y} + y_1\bar{x})$ (see [1] for more details).

Notice that for h_2, h_3 and h_6 we have that $\operatorname{dom} h = \overline{Q}$ (under the usual convention $0 \log 0 = 0$), while for h_4 and h_5 we have $\operatorname{dom} h = Q$.

Next, let us endow Q with the variable metric defined by

$$\forall v, w \in \mathbb{R}^n,\ (v, w)_x := \langle \nabla^2 h(x) v, w \rangle. \tag{6}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function and denote by $df(x) : \mathbb{R}^n \to \mathbb{R}$ the differential of f at $x \in \mathbb{R}^n$. For any $x \in Q$ and $v \in \mathbb{R}^n$ we have that $df(x)v = \langle \nabla f(x), v \rangle = \langle \nabla^2 h(x) \nabla^2 h(x)^{-1} \nabla f(x), v \rangle = (\nabla^2 h(x)^{-1} \nabla f(x), v)_x$. Thus the gradient with respect to the metric $(\cdot, \cdot)_x$ of f restricted to Q is given by

$$\operatorname{grad}_h f(x) = \nabla^2 h(x)^{-1} \nabla f(x), \quad x \in Q. \tag{7}$$

This is a special case of a Riemannian manifold. Indeed, since Q is open, we can take the manifold $M = Q$ with the usual identification for the tangent space $T_x Q \simeq \mathbb{R}^n$ for every $x \in Q$, and the metric defined by (6) endows Q with a Riemannian structure, which is at least locally Lipschitz continuous by virtue of $(H_h; Q)(c)$. In general, if M is a smooth manifold and we denote by $T_x M$ the tangent space to M 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 $T_x M$ such that $(\cdot, \cdot)_x$ depends in a $\mathcal{C}^k$ way on x. The pair $(M, (\cdot, \cdot)_x)$ is called a $\mathcal{C}^k$ Riemannian manifold. This structure permits to define a notion of gradient vector. Indeed, the gradient $\operatorname{grad} f(x)$ of f at $x \in M$ is uniquely determined by the following conditions:

(g_1) tangency condition: $\operatorname{grad} f(x) \in T_x M$,

(g_2) duality condition: for all $v \in T_x M$, $df(x)v = (\operatorname{grad} f(x), v)_x$.

If N is a submanifold of M then $T_x N \subset T_x M$ for all $x \in N$ so that the metric $(\cdot, \cdot)_x$ on M induces a metric on N by restriction. So, for any full rank matrix $A \in \mathbb{R}^{m \times n}$ with $m \leq n$ and $b \in \mathbb{R}^m$, we can take the smooth submanifold of Q defined by $N := \{x \in \mathbb{R}^n \mid Ax = b,\ x \in Q\} = \mathcal{A} \cap Q$, where $\mathcal{A}$ is given by (4), so that $T_x N \simeq \ker A = \{v \in \mathbb{R}^n \mid Av = 0\}$ for each $x \in N$. Definition (6) induces a Riemannian structure on N. Conditions (g_1) and (g_2) imply that the corresponding gradient vector field of f restricted to $\mathcal{A} \cap Q$ is given by

$$\operatorname{grad}_h f(x) = \Pi_x^{\ker A} \nabla^2 h(x)^{-1} \nabla f(x), \tag{8}$$

where $\Pi_x^{\ker A} : \mathbb{R}^n \to \ker A$ is the $(\cdot, \cdot)_x$-orthogonal projection onto the linear subspace $\ker A$. Since A is supposed to have full rank, it is easy to see that

$$\Pi_x^{\ker A} = I - \nabla^2 h(x)^{-1} A^\top (A \nabla^2 h(x)^{-1} A^\top)^{-1} A, \tag{9}$$

and we conclude that for all $x \in \mathcal{A} \cap Q$ we have

$$\operatorname{grad}_h f(x) = \nabla^2 h(x)^{-1} [I - A^\top (A \nabla^2 h(x)^{-1} A^\top)^{-1} A \nabla^2 h(x)^{-1}] \nabla f(x). \tag{10}$$

Notice that in absence of linear equality constraints we can take $A = 0$ and we recover (7) from (10). In any case, given $x \in \mathcal{A} \cap Q$, the vector $-\operatorname{grad}_h f(x)$ can be interpreted as that direction in $\ker A$ such that f decreases the most steeply at x with respect to the metric $(\cdot, \cdot)_x$, which motivates to consider the following dynamical system for the (local) minimization of f on $\mathcal{A} \cap Q$:

$$\frac{du}{dt}(t) = -\operatorname{grad}_h f(u(t)), \tag{11}$$

with initial condition $u(0) = x^0 \in \mathcal{A} \cap Q$. For further developments about this dynamical approach to optimization problems, see [2, 6, 10, 15, 21].

Example 2.2. By taking $A = (1, \dots, 1) \in \mathbb{R}^{1\times n}$, $b = 1$ and $Q = \mathbb{R}^n_{++}$, we get $\mathcal{A} \cap \overline{Q} = \Delta_{n-1} := \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1,\ x \geq 0\}$. Here $N = \mathcal{A} \cap Q = \{x \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 1, x > 0\}$ and $T_x N = \{v \in \mathbb{R}^n \mid \sum_{i=1}^n v_i = 0\}$. Take $h(x) = \sum_{i=1}^n x_i \log x_i - x_i$ so that $\nabla^2 h(x) = \mathrm{diag}(1/x_1, \dots, 1/x_n)$, then (11) amounts to $\frac{du_i}{dt}(t) = -u_i(t) \left[\frac{\partial f}{\partial x_i}(u(t)) - \sum_{j=1}^n u_j(t) \frac{\partial f}{\partial x_j}(u(t))\right]$, $\quad i = 1, \dots, n$. See [17, 19, 22] for applications of this ODE to optimization problems.

2.2 Gradient flows for solving (P)

The feasible solution set $\mathcal{F}$ of (P) is given by $\mathcal{F} = \mathcal{A} \cap \overline{C} \cap \overline{G}$, where $\mathcal{A}$ and $G \neq \emptyset$ are given by (4) and (5), respectively. From now on, the set of optimal solutions of (P) is denoted by $S(P)$, the optimal value of (P) is denoted by $v(P)$, and $\mathcal{F}^0 = \mathcal{A} \cap C \cap G$ stands for the relative interior of $\mathcal{F}$. Throughout this paper, we assume:

$$(H_P) \qquad \begin{cases} (a)\ S(P) \text{ is nonempty and bounded.} \\ (b)\ \mathcal{F}^0 \neq \emptyset \text{ (Slater's condition).} \end{cases}$$

By continuity of all data, we have that $\inf_{x\in\mathcal{F}^0} f(x) = \min_{x\in\mathcal{F}} f(x) = v(P)$.

Remark 2.1. If we replace $Ax = b$ with two sets of inequality constraints, namely $Ax - b \leq 0$ and $b - Ax \leq 0$, these inequalities do not satisfy the Slater condition. Hence, under $(H_P)(b)$ the linear equality constraint $Ax = b$ cannot be integrated into the system of inequalities $g_j(x) \leq 0$, $j \in I$.

Next, let us take a function h_C satisfying $(H_h; C)$, and a *barrier-type* function $\theta : \mathbb{R} \to (0, +\infty]$ with $\mathrm{dom}\,\theta = (-\infty, 0)$ such that:

$$(H_\theta) \qquad \begin{cases} (a)\ \theta : (-\infty, 0) \to \mathbb{R} \text{ is smooth and convex.} \\ (b)\ \theta(s) > 0, \text{ for all } s \in (-\infty, 0), \text{ with } \lim_{s\to 0^-} \theta(s) = +\infty. \\ (c)\ \theta'(s) > 0 \text{ with } \lim\limits_{s\to -\infty} \theta'(s) = 0 \text{ and } \lim\limits_{s\to 0^-} \theta'(s) = +\infty. \end{cases}$$

An example of such a function is the inverse barrier $\theta(s) = -1/s$ if $s < 0$ and $+\infty$ otherwise. We have two alternatives to derive θ-based gradient flows on $\mathcal{F}^0$:
(A1) Riemannian gradients flow using the Hessian of the extended function given by

$$h_{C\cap G}(x) := h_C(x) + \sum_{i\in J} \theta(g_j(x)), \tag{12}$$

under second-order regularity conditions on θ and all g_j, $j \in J$ (at least $\mathcal{C}^2$).
(A2) Hybrid barrier-gradient flows using h_C and replacing the original objective function with the penalty approximate defined by

$$f_\varepsilon(x) = f(x) + \varepsilon \sum_{j\in I} \theta(g_j(x)/\varepsilon), \tag{13}$$

where $\varepsilon > 0$ is a scalar parameter which will ultimately go to 0, and all data is assumed to be at least continuously differentiable.

In the first alternative (A1), we notice that the extended function $h_{C\cap G} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by (12) satisfies $(H_h; C \cap G)$, hence $\mathcal{F}^0$ can be endowed with the Riemannian structure induced by the Hessian of $h_{C\cap G}$, that is

$$\nabla^2 h_{C\cap G}(x) = \nabla^2 h_C(x) + \sum_{j\in I} \theta''(g_j(x)) \nabla g_j(x) \nabla g_j(x)^\top + \sum_{j\in I} \theta'(g_j(x)) \nabla^2 g_j(x). \tag{14}$$

The gradient flow corresponding to (11) is given in this case by

$$(GF;u^0)\qquad \begin{cases} \dfrac{du}{dt}(t) = -\mathrm{grad}_{\,h_{C\cap G}} f(u(t)), \\ u(0) = u^0 \in \mathcal{F}^0. \end{cases}$$

Here $\mathrm{grad}_{\,h_{C\cap Q}} f : \mathcal{F}^0 \to \mathbb{R}^n$ stands for the Riemannian gradient vector field of f restricted to $\mathcal{F}^0$, with respect to the Hessian metric given by (6) for $Q = C \cap G$ and $h = h_{C\cap G}$. Therefore $\mathrm{grad}_{\,h_{C\cap G}} f(x) = \Pi_x^{\ker A} \nabla^2 h_{C\cap G}(x)^{-1} \nabla f(x)$, and the projection mapping $\Pi_x^{\ker A} : \mathbb{R}^n \to \ker A$ is given by (9) for $h = h_{C\cap G}$.

On the other hand, the second alternative (A2) is inspired by previous work on the coupling of the Euclidean steepest descent method with penalty schemes [4, 8, 14]. In this case, we consider the non-autonomous Cauchy problem

$$(B\text{-}GF;u^0)\qquad \begin{cases} \dfrac{du}{dt}(t) = -\mathrm{grad}_{\,h_C} f_{\varepsilon(t)}(u(t)), \\ u(0) = u^0 \in \mathcal{F}^0, \end{cases}$$

where the vector field $\mathrm{grad}_{\,h} f_\varepsilon : \mathcal{F}^0 \to \mathbb{R}^n$ stands for the gradient of f_ε restricted to $\mathcal{F}^0$ with respect to the Hessian Riemannian metric given by (6). Here $\varepsilon : [0, +\infty) \to (0, +\infty)$ is a continuously differentiable parameterization in time of the penalty scheme such that

$$(H_\varepsilon)\qquad \varepsilon(t) > 0, \quad \dot\varepsilon(t) \le 0 \quad \text{and} \quad \lim_{t\to+\infty} \varepsilon(t) = 0.$$

3 Global existence and convergence to the optimal value

Under (H_P), $(H_{h_C}; C)$ and (H_θ), and provided that all data is sufficiently regular, it follows from [2, Theorem 4.1] that the Cauchy problem $(GF;u^0)$ is well-posed in the sense that the solution trajectory $u(t)$ is well-defined for all $t > 0$. Moreover, as f is convex, it follows from [2, Proposition 4.4] that the objective function along the trajectories $f(u(t))$ converges to the optimal value $v(P)$ as $t \to \infty$. The next result establishes that the same holds for $(B\text{-}GF;u^0)$ under additional condition (H_ε).

Theorem 3.1. *Under (H_P), $(H_h; C)$, (H_θ) and (H_ε), the following statements hold:*

(i) The Cauchy problem $(B\text{-}GF;u^0)$ admits a unique $\mathcal{C}^1$ solution $u : [0, +\infty) \to \mathcal{F}^0$.

(ii) The mapping $t \mapsto f_{\varepsilon(t)}(u(t))$ is nonincreasing, the trajectory $\{u(t) \mid t \in [0, +\infty)\}$ is bounded, and $(\dot u, \dot u)_u \in L^1([0, +\infty); \mathbb{R})$.

(iii) For all $a \in \mathcal{F}^0$ and for all $t > 0$

$$f_{\varepsilon(t)}(u(t)) \le f(a) + \frac{1}{t}\left[D_h(a, x^0) - D_h(a, u(t)) + \sum_{j\in I} \theta(g_j(a)/\varepsilon_0) \int_0^t \varepsilon(s)ds \right], \tag{15}$$

where

$$D_h(y, x) = h(y) - h(x) - \langle \nabla h(x), y - x\rangle \ge 0. \tag{16}$$

Hence,

$$\lim_{t\to+\infty} f_{\varepsilon(t)}(u(t)) = \lim_{t\to+\infty} f(u(t)) = \min_{\mathcal{F}} f = v(P),$$

and every cluster point of $\{u(t) \mid t \to +\infty\}$ belongs to $S(P)$.

Proof. Under the conditions (H_P), $(H_h;C)$, (H_θ) and (H_ε) the mapping $(u,t)\mapsto \operatorname{grad}_h f_{\varepsilon(t)}|_{\mathcal{A}\cap C}$ is locally Lipschitz continuous on $\mathcal{A}\cap C$ and satisfies the hypotheses of the classical Cauchy-Lipschitz Theorem, which gives the existence and uniqueness of a local solution of $(B\text{-}GF;u^0)$. The global existence of $u(t)$ for all $t\in[0,+\infty)$ is not immediate because of the singular behavior of the vector field near the boundary of $\mathcal{A}\cap C$. Then, we first define

$$T_{max}=\sup\{T>0\mid \exists!\text{ solution } u \text{ of } (B\text{-}GF;u^0) \text{ on } [0,T) \text{ such that } u([0,T))\subset\mathcal{F}^0\}. \tag{17}$$

It follows that $T_{max}>0$. Let $u:[0,T_{max})\to\mathcal{F}^0$ be a maximal solution of $(B\text{-}GF;u^0)$. The definition of the projection $\Pi_x^{\ker A}$ (see (9)) implies that, for all $t\in[0,T_{max})$,

$$\dot u+\nabla^2h(u)^{-1}\nabla f_\varepsilon(u)=\nabla^2h(u)^{-1}A^\top(A\nabla^2h(u)^{-1}A^\top)^{-1}A\nabla^2h(u)^{-1}\nabla f_\varepsilon(u).$$

Then, for all $y\in\ker A$, we get $(\dot u+\nabla^2h(u)^{-1}\nabla f_\varepsilon(u),y+\dot u)_u=0$, that is,

$$\langle\nabla^2h(u)\dot u+\nabla f_\varepsilon(u),y+\dot u\rangle=0. \tag{18}$$

Letting $y=0$ in (18) yields

$$\langle\nabla f_\varepsilon(u),\dot u\rangle+\langle\nabla^2h(u)\dot u,\dot u\rangle=0, \tag{19}$$

and thus by $(H_h;C)(b)$ one has that $\langle\nabla f_\varepsilon(u),\dot u\rangle\le 0$. On the other hand, we note that

$$\frac{d}{dt}f_\varepsilon(u)=\langle\nabla f_\varepsilon(u),\dot u\rangle+\frac{\partial}{\partial\varepsilon}f_\varepsilon(u)\dot\varepsilon$$

with

$$\frac{\partial}{\partial\varepsilon}f_\varepsilon(u)=\sum_{j\in I}[\theta(g_j(u)/\varepsilon)-\dot\theta(g_j(u)/\varepsilon)\frac{g_j(u)}{\varepsilon}]\ \text{ if } g_j(u)<0.$$

From $(H_\theta)(b)$,(c), we get $\frac{\partial}{\partial\varepsilon}f_\varepsilon(u)\ge 0$, and as ε satisfies (H_ε), we have $\frac{\partial}{\partial\varepsilon}f_\varepsilon(u)\dot\varepsilon\le 0$.

Thus, the mapping $t\mapsto f_{\varepsilon(t)}(u(t))$ is nonincreasing. Moreover, as $S(P)\neq\emptyset$ and $(H_\theta)(b)$ holds, we have that $f_{\varepsilon(\cdot)}(\cdot)$ is bounded from below by $v(P)$ and consequently $f_{\varepsilon(t)}(u(t))$ is convergent as $t\to T_{max}$. In particular, $f_{\varepsilon(t)}(u(t))$ is bounded. Moreover, as $f_{\varepsilon(t)}(u(t))$ is nonincreasing, we have that $u(t)\in L_f=\{y\in\mathcal{F}\mid f(y)\le f_{\varepsilon_0}(x^0)\}$, for all $t\in[0,T_{max})$ because $f_\varepsilon(u)\ge f(u)$. By $(H_P)(a)$ together with convexity, it follows that the level set L_f is bounded, hence $\{u(t)\mid t\in[0,T_{max})\}$ is bounded.

On the other hand, integrating (19) from 0 to t, with $t\in[0,T_{max})$, we obtain

$$\int_0^t\langle\nabla^2h(u)\dot u,\dot u\rangle ds=\int_0^t\left[\frac{\partial}{\partial\varepsilon}f_\varepsilon(u)\dot\varepsilon-\frac{d}{ds}f_\varepsilon(u)\right]ds\le f_{\varepsilon_0}(x^0)-f_{\varepsilon(t)}(u(t)).$$

Then, we get that

$$\langle\nabla^2h(u(\cdot))\dot u(\cdot),\dot u(\cdot)\rangle\in L^1([0,T_{max});\mathbb{R}). \tag{20}$$

Note that if $T_{\max}=+\infty$, (ii) follows. Let us argue by contradiction and assume that $T_{max}<+\infty$. Let ω be the set of limit points of u, which is nonempty, and set $K=u([0,T_{\max}))\cup\omega$. Note that K is compact. If $K\subset C\cap G$, where $G=\{x\in\mathbb{R}^n\mid g_j(x)<0,\ j\in I\}$ as in (5), then the compactness of K implies that $u(t)$ can be extended for t beyond T_{max}, which contradicts the maximality of T_{max}.

Let us prove that $K\subset C\cap G$. It suffices to prove that $\omega\subset C\cap G$ for which we will argue by contradiction. Assume that $u(t_j)\to u^*$ with $t_j\in(0,T_{max})$, $t_j\to T_{max}$ as $j\to+\infty$ and

$u^* \in \partial C \cap \partial G$. First, if $g_{i_0}(u^*) = 0$ for some $i_0 \in I$, then from $(H_\theta)(b)$ and the fact that $\varepsilon(t_j) \to \varepsilon(T_{max}) > 0$ we must have that

$$\theta(g_{i_0}(u(t_j))/\varepsilon(t_j)) \to \theta(0^-) = +\infty \text{ as } j \to +\infty,$$

so $f_{\varepsilon(t_j)}(u(t_j)) = f(u(t_j)) + \varepsilon(t_j) \sum_{j \in I} \theta(g_j(u(t_j))/\varepsilon(t_j)) \to +\infty$, which contradicts the boundedness of $f_{\varepsilon(t)}(u(t))$ for all $t \in [0, T_{max})$. Hence $u^* \in G$. Since this is valid for any cluster point, we have $\omega \subset G$.

Now, assume that $u^* \in \partial C$. To obtain a contradiction in this case, we can apply the arguments of [2] based on the following auxiliary lemmas:

Lemma 3.2 ([2, Lemma 4.2])**.** *Let h be a convex function with $\operatorname{int}(\operatorname{dom} h) = C$ and $\{x^j\} \subset C$ a sequence such that $x^j \to x^* \in \partial C$ and $\frac{\nabla h(x^j)}{\|\nabla h(x^j)\|} \to \xi \in \mathbb{R}$, as $j \to +\infty$. Then, ξ belongs to $\mathcal{N}_{\overline{C}}(x^*)$, the normal cone to $\overline{C}$ at x^*.*

Lemma 3.3 ([2, Lemma 4.3])**.** *Let C be a nonempty convex subset of $\mathbb{R}$, and $\mathcal{A}$ an affine space of $\mathbb{R}$ such that $\mathcal{A} \cap C \neq \emptyset$. If $x^* \in \partial\mathcal{A} \cap C$, then $\mathcal{N}_{\overline{C}}(x^*) \cap \mathcal{A}_0^\perp = \{0\}$, with $\mathcal{A}_0 = \mathcal{A} - \mathcal{A}$.*

Since h satisfies $(H_h; C)(b)$, as we are assuming that $u(t_j) \to u^* \in \partial C$, we must have $\|\nabla h(u(t_j))\| \to +\infty$, and we may assume that $\nabla h(u(t_j))/\|\nabla h(u(t_j))\| \to \xi \in \mathbb{R}^n$ with $\|\xi\| = 1$. From Lemma 3.2 it follows that $\xi \in \mathcal{N}_{\overline{C}}(u^*)$. Let $\xi_0 = \operatorname{Proj}_{\ker A}\xi$ be the Euclidean orthogonal projection of ξ onto $\ker A$, and take $y = \xi_0$ in (18). Using (19), the expression of ∇f_ε, and integrating from 0 to t_j gives

$$\begin{aligned} \langle \nabla h(u(t_j)), \xi_0 \rangle &= \left\langle \nabla h(u^0) - \int_0^{t_j} \nabla f(u(s)) ds, \xi_0 \right\rangle \\ &\quad - \left\langle \sum_{j \in I} \int_0^{t_j} \theta'(g_j(u(s))/\varepsilon(s)) \nabla g_j(u(s)) ds, \xi_0 \right\rangle \end{aligned} \tag{21}$$

By the boundedness property of $u([0, T_{max}))$ and as $\omega \subset D$, the right-hand side of (21) is bounded under the assumption $T_{max} < +\infty$. Therefore, to draw a contradiction from the latter we just have to prove that the left-hand side of (21) tends towards $+\infty$ as $j \to +\infty$. Indeed, since $\langle \nabla h(u(t_j))/\|\nabla h(u(t_j))\|, \xi_0 \rangle \to \|\xi_0\|^2$, the proof of the result is complete if we check that $\xi_0 \neq 0$. We observe that $\xi_0 = 0$ if and only if $\xi \in (\ker A)^\perp$. But if $\xi \in (\ker A)^\perp$ and as $\xi \in \mathcal{N}_{\overline{C}}(u^*)$, it follows from Lemma 3.3 that $\xi = 0$, which is a contradiction because $\|\xi\| = 1$. These contradictions arise from the assumption $u^* \in \partial C$. Hence $u^* \in C$ and consequently $\omega \subset C$.

Therefore, $T_{max} = +\infty$, which completes the proof of (i), (ii) and (iii).

To conclude the proof of the Theorem, fix $a \in \mathcal{F}^0$. For each $t \geq 0$, take $y = u(t) - a$ in (18) to obtain

$$\langle \nabla f_{\varepsilon(t)}(u(t)) + \frac{d}{dt} \nabla h(u(t)), u(t) - a + \dot{u}(t) \rangle = 0.$$

By using (19), we get

$$\langle \nabla f_{\varepsilon(t)}(u(t)), u(t) - a \rangle + \langle \frac{d}{dt} \nabla h(u(t)), u(t) - a \rangle = 0.$$

From (16) we have that the solution $u(t)$ of $(B\text{-}GF; u^0)$ satisfies

$$\frac{d}{dt} D_h(a, u(t)) + \langle \nabla f_{\varepsilon(t)}(u(t)), u(t) - a \rangle = 0, \quad \forall t \geq 0. \tag{22}$$

Then, the convexity of f_ε implies $\frac{d}{dt} D_h(a, u(t)) + f_{\varepsilon(t)}(u(t)) \leq f_{\varepsilon(t)}(a)$. Integrating from 0 to t yields

$$D_h(a, u(t)) + \int_0^t [f_{\varepsilon(s)}(u(s)) - f_{\varepsilon(s)}(a)] ds \leq D_h(a, x^0).$$

Since $s \mapsto f_{\varepsilon(s)}(u(s))$ is nonincreasing, it follows that

$$t f_{\varepsilon(t)}(u(t)) - \int_0^t f_{\varepsilon(s)}(a) ds \le D_h(a, u_0) - D_h(a, u(t)). \tag{23}$$

On the other hand, since $\varepsilon(\cdot) > 0$ is decreasing, $g_j(a) \le 0$ and θ is nondecreasing, one has

$$f_{\varepsilon(s)}(a) \le f(a) + \varepsilon(s) \sum_{j \in I} \theta(g_j(a)/\varepsilon_0). \tag{24}$$

Hence, the estimate is obtained from (24) and (23). Moreover, letting $t \to +\infty$ and as $D_h(a, u(t)) \ge 0$ by convexity of h, it follows that $\limsup_{t\to+\infty} f_{\varepsilon(t)}(u(t)) \le f(a)$. Thus $\limsup_{t\to+\infty} f_{\varepsilon(t)}(u(t)) \le \inf_{\mathcal{F}^0} f = v(P)$ by virtue of the continuity of all data. On the other hand, from $(H_\theta)(b)$ and (H_ε), we deduce $f_{\varepsilon(t)}(u(t)) \ge f(u(t)) \ge v(P)$. Passing to the limit as $t \to +\infty$ we get $\liminf_{t\to+\infty} f_{\varepsilon(t)}(u(t)) \ge v(P)$. Thus, $\lim_{t\to+\infty} f_{\varepsilon(t)}(u(t)) = v(P)$ and therefore $\lim_{t\to+\infty} f(u(t)) = v(P)$, which concludes the proof.

□

4 Generalized barrier proximal point algorithm

The purpose of this section is to provide one discrete version of the Theorem 3.1. Let us begin by noticing that from (9), it follows easily that the solution $u(t)$ of $(B\text{-}GF; u^0)$ satisfies

$$\begin{cases} \frac{d}{dt}\nabla h(u(t)) + \nabla f_{\varepsilon(t)}(u(t)) & \in & (\ker A)^\perp = \operatorname{Im} A^\top, \\ u(t) & \in & \mathcal{F}^0, \\ u(0) & = & x^0. \end{cases} \tag{25}$$

An implicit discretization of (25) yields the following iterative scheme:

$$\nabla h(x^k) - \nabla h(x^{k-1}) + \lambda_k \nabla f_{\varepsilon_k}(x^k) \in \operatorname{Im} A^\top,\ Ax^k = b,$$

with x^0 being a given starting point, $\varepsilon_k > 0$ a sequence of penalty parameters decreasing to 0, and $\lambda_k > 0$ a sequence of stepsizes. By convexity, this stationary condition is equivalent to the generalized barrier proximal point iteration

$$x^k \in \operatorname{Argmin}\{f_{\varepsilon_k}(u) + \lambda_k^{-1} D_h(u, x^{k-1}) \mid Au = b\}, \tag{26}$$

where D_h is given by (16). The constraints $u \in \overline{C}$ and $g_j(u) \le 0$, $j \in I$, are implicit in the definitions of $D_h(u, x^{k-1})$ and $f_{\varepsilon_k}(u)$, respectively. In fact, due to the behavior of h and θ, we have that $x^k \in C$ and $g(x^k) < 0$, $j \in I$. In this sense, (26) is a strictly feasible algorithm.

The iteration (26) makes sense even for nonsmooth data. More generally, for each $k = 1, 2, \ldots$, let generate a sequence $\{x^k\} \in \mathcal{F}^0$ satisfying

$$\frac{\nabla h(x^k) - \nabla h(x^{k-1}) + A^\top w^k}{\lambda_k} \in -\partial_{\zeta_k} f_{\varepsilon_k}(x^k);\ Ax^k = b, \tag{27}$$

for some $w^k \in \mathbb{R}^m$ and $\zeta_k \ge 0$ is a tolerance for the computation of approximate subgradients: $\partial_\zeta f_\varepsilon(x) = \left\{s \in \mathbb{R}^n \mid \forall y \in \mathbb{R}^n,\ f_\varepsilon(x) + s^\top(y - x) - \zeta \le f_\varepsilon(y)\right\}$. In the Euclidean case $C = \mathbb{R}^n$, $\nabla h(x) = x$ and $A = 0$, this kind of method is studied, for instance, in [14].

The following result, which is a discrete version of Theorem 3.1, generalizes the estimate derived in [12, Lemma 3.3(iii)] and extends the value convergence result established in [12, Theorem 3.4] for the exact version of (27) without penalty parameters, i.e, $\varepsilon_k = \zeta_k = 0$, $\forall k$ and also with $A = 0$.

Theorem 4.1. *Let $\{x^k\} \subset \mathcal{F}^0$ be the sequence generated by the generalized barrier proximal point algorithm (27) with $\{\varepsilon_k\}$ being decreasing to 0. Set $\sigma_n = \sum_{k=1}^n \lambda_k$. If $\sum_{k=1}^\infty \zeta_k < \infty$, then the following statements hold:*

(*i*) *The real sequence $\{f_{\varepsilon_n}(x^n)\}$ is convergent, the sequence $\{x^n\}$ is bounded and we have that $\sum_{k=1}^\infty \lambda_k^{-1}\langle \nabla h(x^k) - \nabla h(x^{k-1}), x^k - x^{k-1}\rangle < +\infty$.*

(*ii*) *For all $a \in \mathcal{F}^0$ and for all $n \geq 1$ we have*

$$\begin{aligned}\sigma_n(f_{\varepsilon_n}(x^n) - f(a)) \quad &\leq \quad \sum_{j\in I}\theta(g_j(a)/\varepsilon_0)\sum_{k=1}^n \lambda_k\varepsilon_k + D_h(a,x^0) - D_h(a,x^n) \\ &\qquad - \sum_{k=1}^n \sigma_k\lambda_k^{-1}D_h(x^k,x^{k-1}) + \sum_{k=1}^n \sigma_k\zeta_k. \end{aligned} \tag{28}$$

(*iii*) *If $\sigma_n \to +\infty$, then the sequence $\{f_{\varepsilon_n}(x^n)\}$, converges to $v(P)$, hence $\{f(x^n)\}$ does so and every cluster point of $\{x^n\}$ belongs to $S(P)$.*

Proof. (*i*) Using the definition of the ζ-subdifferential, for all $a \in \mathcal{F}^0$, we have

$$\begin{aligned} f_{\varepsilon_k}(a) \quad &\geq \quad f_{\varepsilon_k}(x^k) + \lambda_k^{-1}\langle \nabla h(x^{k-1}) - \nabla h(x^k) - A^\top w^k, a - x^k\rangle - \zeta_k \\ &= \quad f_{\varepsilon_k}(x^k) + \lambda_k^{-1}\langle \nabla h(x^{k-1}) - \nabla h(x^k), a - x^k\rangle - \zeta_k. \end{aligned} \tag{29}$$

Taking in particular $a = x^{k-1}$, from the convexity of h it follows that $f_{\varepsilon_k}(x^k) \leq f_{\varepsilon_k}(x^{k-1}) + \zeta_k$. Since ε_k is decreasing, $g_j(x^{k-1}) < 0$ and θ is nondecreasing, one has that

$$f_{\varepsilon_k}(x^k) \leq f_{\varepsilon_{k-1}}(x^{k-1}) + \zeta_k. \tag{30}$$

In particular $0 \leq f_{\varepsilon_k}(x^k) - v(P) \leq f_{\varepsilon_{k-1}}(x^{k-1}) - v(P) + \zeta_k$, from which we conclude that the sequence $\{f_{\varepsilon_k}(x^k)\}$ converges by virtue of part (a) of the following technical lemma [24, 26].

Lemma 4.2. (a) *Let $\{v_k\}$ and $\{\alpha_k\}$ be nonnegative real sequences satisfying $v_k \leq v_{k-1} + \alpha_k$ for $\sum \alpha_k < \infty$. Then the sequence $\{v_k\}$ converges.*
(b) *Let $\{\sigma_k\}$ be an increasing sequence of positive numbers and $\{a_k\}$ a real sequence. If $\sigma_n \to +\infty$ as $n \to +\infty$ and $\sum a_k < +\infty$, then $\sigma_n^{-1}\sum_{k=1}^n \sigma_k a_k \to 0$ as $n \to +\infty$.*

Next, summing (30) over $k = 1, \ldots, n$, one has $f_{\varepsilon_n}(x^n) \leq f_{\varepsilon_0}(x^0) + \sum_{k=1}^n \zeta_k \leq f_{\varepsilon_0}(x^0) + \bar{\zeta}$, where $\bar{\zeta} = \sum_{k=1}^\infty \zeta_k < +\infty$. But $f_{\varepsilon_n}(x^n) \geq f(x^n)$. As a consequence, $x^n \in L_f = \{y \in \mathcal{F} \mid f(y) \leq f_{\varepsilon_0}(x^0) + \bar{\zeta}\}$ for all $n \geq 0$. By $(H_P)(a)$ it follows that $\{x^n\}$ is a bounded sequence. On the other hand, taking $a = x^{k-1}$ in (29) and using the fact that $f_{\varepsilon_k}(x^{k-1}) \leq f_{\varepsilon_{k-1}}(x^{k-1})$, we get

$$\lambda_k^{-1}\langle \nabla h(x^k) - \nabla h(x^{k-1}), x^k - x^{k-1}\rangle \leq f_{\varepsilon_{k-1}}(x^{k-1}) - f_{\varepsilon_k}(x^k) + \zeta_k.$$

Summing over $k = 1, \ldots, n$, we obtain

$$\sum_{k=1}^n \lambda_k^{-1}\langle \nabla h(x^k) - \nabla h(x^{k-1}), x^k - x^{k-1}\rangle \leq f_{\varepsilon_0}(x^0) - f_{\varepsilon_n}(x^n) + \sum_{k=1}^n \zeta_k \leq f_{\varepsilon_0}(x^0) - v(P) + \sum_{k=1}^n \zeta_k.$$

Letting $n \to +\infty$ one has $\sum_{k=1}^\infty \lambda_k^{-1}\langle \nabla h(x^k) - \nabla h(x^{k-1}), x^k - x^{k-1}\rangle < +\infty$.
(*ii*) Applying to (29) the three points identity [12, Lemma 3.1], which relies only on the very definition of the pseudo-metric (16), we obtain

$$\lambda_k(f_{\varepsilon_k}(x^k) - f_{\varepsilon_k}(a)) \leq D_h(a,x^{k-1}) - D_h(a,x^k) - D_h(x^k,x^{k-1}) + \lambda_k\zeta_k. \tag{31}$$

As $\varepsilon_k > 0$ is decreasing, $g_j(a) < 0$ and θ is nondecreasing, we have that $f_{\varepsilon_k}(a) = f(a) + \varepsilon_k \sum_{j\in I} \theta(g_j(a)/\varepsilon_k) \le f(a) + \varepsilon_k \sum_{j\in I} \theta(g_j(a)/\varepsilon_0)$. Therefore, in (31) one has that

$$\lambda_k(f_{\varepsilon_k}(x^k) - f(a) - \varepsilon_k \sum_{j\in I} \theta(g_j(a)/\varepsilon_0)) \le D_h(a, x^{k-1}) - D_h(a, x^k) - D_h(x^k, x^{k-1}) + \lambda_k \zeta_k$$

Summing over $k = 1, \ldots, n$ we get

$$\begin{aligned}\sum_{k=1}^n \lambda_k f_{\varepsilon_k}(x^k) - \sigma_n f(a) \quad \le \quad & \sum_{k=1}^n \lambda_k \varepsilon_k \sum_{j\in I} \theta(g_j(a)/\varepsilon_0)) + D_h(a, x^0) - D_h(a, x^n) \\ & - \sum_{k=1}^n D_h(x^k, x^{k-1}) + \sum_{k=1}^n \lambda_k \zeta_k. \end{aligned} \tag{32}$$

Now, setting $a = x^{k-1}$ in (31) yields $f_{\varepsilon_k}(x^k) - f_{\varepsilon_k}(x^{k-1}) \le -\lambda_k^{-1} D_h(x^k, x^{k-1}) + \zeta_k$. Multiplying by σ_{k-1} and using the fact that $\sigma_k = \lambda_k + \sigma_{k-1}$ (with $\sigma_0 = 0$), one has $\sigma_k f_{\varepsilon_k}(x^k) - \lambda_k f_{\varepsilon_k}(x^k) - \sigma_{k-1} f_{\varepsilon_k}(x^{k-1}) \le -\sigma_{k-1}\lambda_k^{-1} D_h(x^k, x^{k-1}) + \sigma_{k-1}\zeta_k$, whence $\sigma_k f_{\varepsilon_k}(x^k) - \sigma_{k-1} f_{\varepsilon_{k-1}}(x^{k-1}) - \lambda_k f_{\varepsilon_k}(x^k) \le -\sigma_{k-1}\lambda_k^{-1} D_h(x^k, x^{k-1}) + \sigma_{k-1}\zeta_k$, because ε_k is decreasing. Summing over $k = 1, \ldots, n$ we get $-\sum_{k=1}^n \lambda_k f_{\varepsilon_k}(x^k) + \sigma_n f_{\varepsilon_n}(x^n) \le -\sum_{k=1}^n \sigma_{k-1}\lambda_k^{-1} D_h(x^k, x^{k-1}) + \sum_{k=1}^n \sigma_{k-1}\zeta_k$. Adding this inequality to (32) and recalling that $\lambda_k + \sigma_{k-1} = \sigma_k$, we obtain (28).

(iii) Dividing (28) by σ_n, passing to the limit as $n \to +\infty$, using the fact that $\sum_{k=1}^\infty \zeta_k < \infty$, that $\varepsilon_k \to 0$ and invoking Lemma 4.2(b), we have that $\limsup_{n\to+\infty} f_{\varepsilon_n}(x^n) \le f(a)$ for all $a \in \mathcal{F}^0$, whence $\limsup_{n\to+\infty} f_{\varepsilon_n}(x^n) \le \inf_{\mathcal{F}^0} f = v(P)$. By $H_\theta(b)$ we deduce $f_{\varepsilon_n}(x^n) \ge f(x^n) \ge v(P)$. Passing to the limit as $n \to +\infty$ we get $\liminf_{n\to+\infty} f_{\varepsilon_n}(x^n) \ge v(P)$. □

5 Some simple specializations and numerical illustrations

In this section we will present some specific instances of the Hessian Riemannian gradient flow $(GF; u^0)$ and the hybrid barrier-gradient flow $(B\text{-}GF; u^0)$. Moreover, we will illustrate them through some very simple computational examples; codes were all written in MATLAB 7.7, Release 2008b.

In fact, we will consider a *separable* function h_C that can be expressed as

$$h_C(x) = \sum_{i=1}^n \psi(x_i) \tag{33}$$

for some suitable scalar function $\psi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ satisfying $(H_h; (0, +\infty))$, so that in particular

$$\nabla^2 h_C(x) = \operatorname{diag}(\psi''(x_1), \ldots, \psi''(x_n)), \quad x \in C.$$

From now on, we take the inverse barrier function $\theta(s) = -1/s$ if $s < 0$ and $+\infty$ otherwise. This choice in (12) leads to $h_{C\cap G}(x) = \sum_{i=1}^n \psi(x_i) - \sum_{i\in J} \frac{1}{g_j(x)}$, whose Hessian is given by

$$\nabla^2 h_{C\cap G}(x) = \operatorname{diag}(\psi''(x_1), \ldots, \psi''(x_n)) - \sum_{j\in I} \frac{2}{g_j(x)^3} \nabla g_j(x) \nabla g_j(x)^\top + \sum_{j\in I} \frac{1}{g_j(x)^2} \nabla^2 g_j(x). \tag{34}$$

On the other hand, we may consider the following reparameterization of the penalty function (13):

$$f(x, r) = f(x) - r \sum_{j\in I} \frac{1}{g_j(x)}, \tag{35}$$

where $r = \varepsilon^2 > 0$. Notice that the Euclidean gradient $\nabla_x f(x, r)$ is given componentwise by

$$\frac{\partial f}{\partial x_i}(x, r) = \frac{\partial f}{\partial x_i}(x) + r \sum_{j\in I} \frac{1}{g_j(x)^2} \frac{\partial g_j}{\partial x_i}(x), \quad i = 1, \ldots, n. \tag{36}$$

5.1 Positive orthant

First, suppose that the problem we want to solve is the following

$$(P_1) \qquad \min\{f(x) \mid x \geq 0,\ g_j(x) \leq 0,\ j \in I\},$$

so that there is no linear equality constraint. In this case, for a separable function (33) satisfying $(H_h; \mathbb{R}^n_{++})$, the associated ODE in $(GF; u^0)$ has the form: $\frac{du}{dt}(t) = -\nabla^2 h_{C\cap G}(u(t))^{-1}\nabla f(u(t))$, for $\nabla^2 h_{C\cap G}$ given by (34). On the other hand, the system in $(B\text{-}GF; u^0)$ can be written componentwise as $\frac{du_i}{dt}(t) = -\psi''(u_i(t))^{-1}\frac{\partial f}{\partial x_i}(u(t), r(t))$, $\quad i = 1, \ldots, n$, for the reparametrization $r(t) = \varepsilon(t)^2$ with $\varepsilon(\cdot)$ satisfying (H_ε) and $\frac{\partial f}{\partial x_i}(x, r)$ given by (36). In order to illustrate the behavior of the solution trajectories of these equations we will consider a very simple explicit discretization scheme, namely Euler's method with constant stepsize $\Delta t > 0$. For $(GF; u^0)$, this takes the form

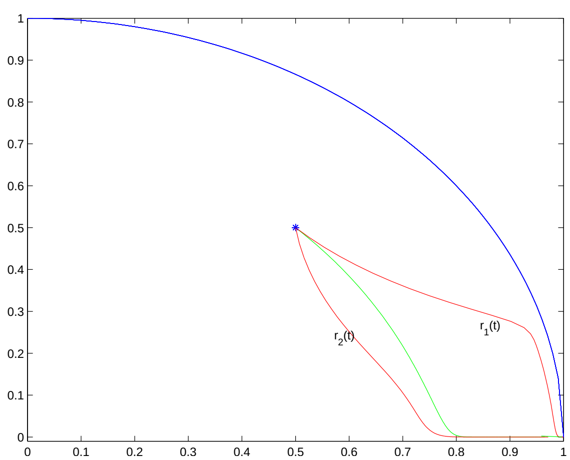
$$\begin{array}{rcl} \nabla^2 h_{C\cap G}(u^k)d^k &=& -\nabla f(u^k), \\ u^{k+1} &=& u^k + \Delta t\, d^k, \end{array} \tag{37}$$

while for $(B\text{-}GF; u^0)$ we get

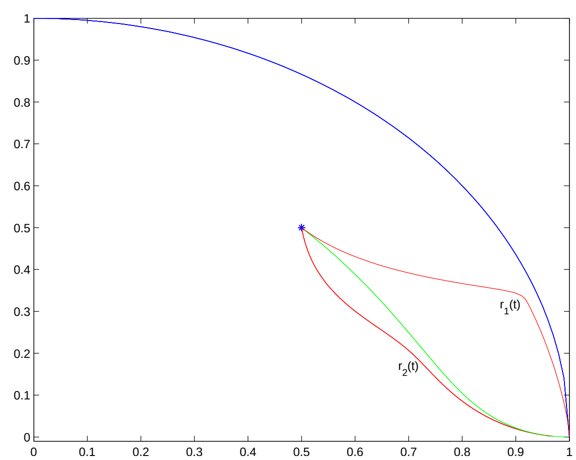
$$u_i^{k+1} = u_i^k - \Delta t\, \psi''(u_i^k)^{-1}\frac{\partial f}{\partial x_i}(u^k, r(t_k)),\ i = 1, \ldots, n, \tag{38}$$

with $t_k = k\Delta t$ for $k = 0, 1, \ldots$

Example 5.1. Consider the following problem $\min\{c^\top x \mid x \geq 0, \sum_{i=1}^n x_i^2 \leq 1\}$. First, take $c = (-\frac{1}{2}, \frac{1}{2})$ for which the unique minimizer is $(1, 0)$. In Figure 1 we illustrate the trajectories obtained through piecewise-linear interpolation of the explicit Euler schemes (37) and (38) with starting point $u^0 = (\frac{1}{2}, \frac{1}{2})$ and stepsize $\Delta t = 0.1$, for two different choices of ψ and $r(t)$.



(a) Taking $\psi(\lambda) = \lambda \log \lambda - \lambda$



(b) Taking $\psi(\lambda) = -\log \lambda$

Figure 1: In green the trajectory of $(GF; (\frac{1}{2}, \frac{1}{2}))$ for Example 5.1 with $c = (-\frac{1}{2}, \frac{1}{2})$. In red, two trajectories of $(B\text{-}GF; (\frac{1}{2}, \frac{1}{2}))$ for the same problem with $r_1(t) = \varepsilon_1(t)^2 = \frac{1}{(t+10)^3}$ and $r_2(t) = \varepsilon_2(t)^2 = \frac{1}{t+10}$.

Now, let us take $n = 10, 100, 1000$ with $c = (-\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ and starting point $u_i^0 = \frac{1}{\sqrt{2n}}$, $i = 1, \ldots, n$. Tables 1 and 2 provide comparisons between the two iterative schemes (37) and (38) in terms of total number of iterations as well as CPU time required to satisfy the stopping rule $f(u^k) < v(P_1) + Tol$ with $Tol = 10^{-2}$, which makes sense in this case because we know that the optimal value is $v(P_1) = -\frac{1}{n}$. We take $\psi(\lambda) = -\log\lambda$, $r_1(t) = \frac{1}{(t+10)^3}$ and $r_2(t) = \frac{1}{t+10}$. We consider two choices for the the constant step-size, either $\Delta t = \sqrt{n/2}$ or $\Delta t = n/2$. All numerical tests were performed on a Toshiba Satellite laptop with an Intel Pentium Core 2 Duo CPU 2.20GHz processor and 4GB of RAM, running Microsoft Windows XP operating system.

Table 1: Computational results with $c = (-\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$, $\psi(\lambda) = -\log\lambda$ and $\Delta t = \sqrt{n/2}$.

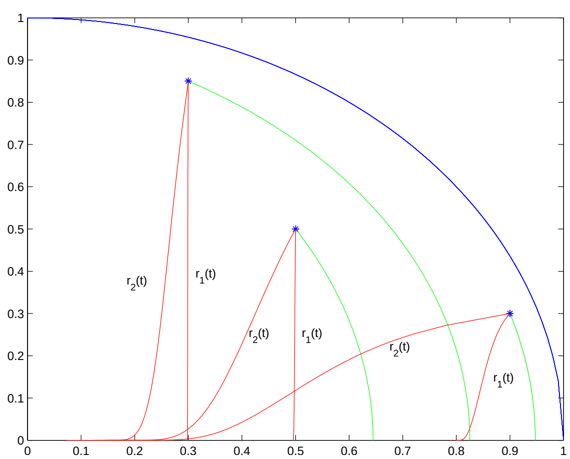
	# Iterations			CPU time		
dimension n	(37)	(38)		(37)	(38)	
		$r_1(t)$	$r_2(t)$		$r_1(t)$	$r_2(t)$
10	825	16	190	00'00".0871	00'00".0006	00'00".0046
100	1322	188	1277	00'01".4927	00'00".0131	00'00".0896
1000	2577	1959	2483	06'24".0242	00'00".8109	00'01".0762

Table 2: Computational results with $c = (-\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$, $\psi(\lambda) = -\log\lambda$ and $\Delta t = n/2$.

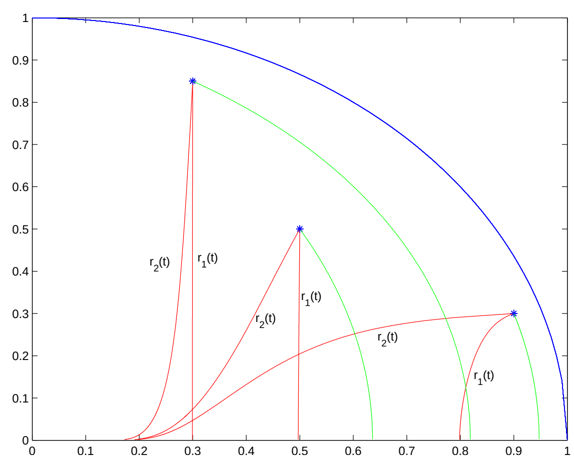
	# Iterations			CPU time		
dimension n	(37)	(38)		(37)	(38)	
		$r_1(t)$	$r_2(t)$		$r_1(t)$	$r_2(t)$
10	367	8	60	00'00".0417	00'00".0005	00'00".0018
100	186	28	50	00'00".2642	00'00".0023	00'00".0036
1000	115	91	93	00'15".9470	00'00".0462	00'00".0471

Notice that the results depend strongly on the choice of the step-size. In any case, (38) is much faster for $r_1(t) = \frac{1}{(t+10)^3}$ than for the slow parameterization $r_2(t) = \frac{1}{t+10}$, while (37) has always the worst performance. In fact, even when (37) is similar to (38)-$r_2(t)$ in terms of total number of iterations, the latter is much faster than the former in terms of CPU time because (37) solves a linear system at each iteration for obtaining the descent direction.

Next, consider $c = (0, \frac{1}{2})$ so that the optimal set is $[0,1] \times \{0\}$. Figure 2 shows the trajectories corresponding to three different starting points u^0: $(0.5, 0.5)$, $(0.3, 0.85)$ and $(0.9, 0.3)$.



(a) Taking $\psi(\lambda) = \lambda\log\lambda - \lambda$



(b) Taking $\psi(\lambda) = -\log\lambda$

Figure 2: For three different starting points, in green the trajectories of $(GF; u^0)$ for Example 5.1 with $c = (0, \frac{1}{2})$; in red, the trajectories of $(B\text{-}GF; u^0)$ for the same problem with $r_1(t)$ and $r_2(t)$ as in Figure 1.

Even for nonunique optimal solutions, the trajectories seems to converge. Notice that the limit point of the trajectories of $(GF; u^0)$ depends on the starting point. The same behavior holds for $(B\text{-}GF; u^0)$ under the fast parameterization $r_1(t)$. But the trajectories of $(B\text{-}GF; u^0)$ with the slow parameterization $r_2(t)$ appear to approach the origin in order to minimize the penalty term in (35) which is given in this case by $\frac{1}{1-\sum_{i=1}^n x_i^2}$.

5.2 Unitary simplex

Suppose now that the problem we want to solve is the following

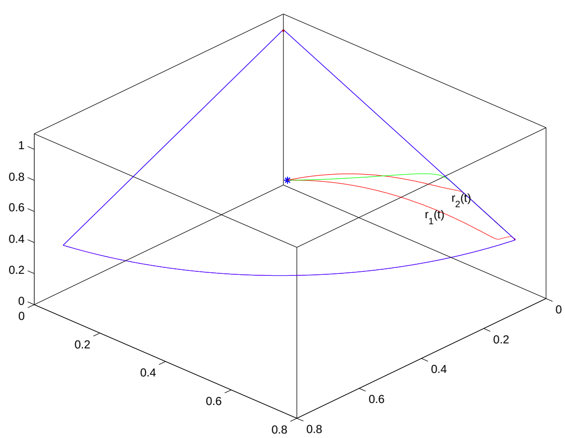
$$(P_2) \qquad \min\{f(x) \mid \sum_{i=1}^{n} x_i = 1,\, x \geq 0,\, g_j(x) \leq 0,\ j \in I\}.$$

In this case, $A = [1 \dots 1] \in \mathbb{R}^{1\times n}$, $b = 1$ and $C = \mathbb{R}^n_{++}$. Again we take a separable h as (33). According to (10), a simple computation shows that the ODE in $(B\text{-}GF; u^0)$ can be written as

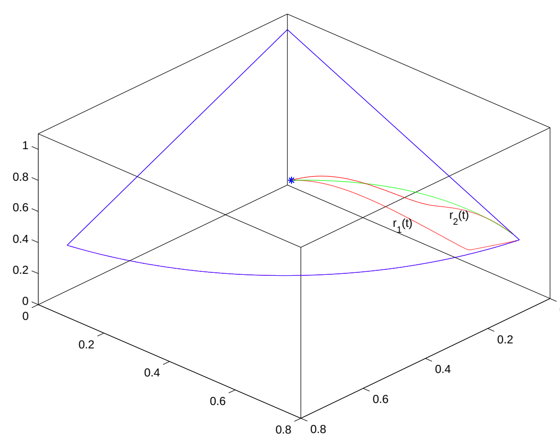
$$\frac{du_i}{dt}(t) = -\psi''(u_i(t))^{-1}\frac{\partial f}{\partial x_i}(u(t), r(t)) + \frac{\psi''(u_i(t))^{-1}}{\sum_{\ell=1}^n \psi''(u_\ell(t))^{-1}} \sum_{\ell=1}^{n} \psi''(u_\ell(t))^{-1}\frac{\partial f}{\partial x_\ell}(u(t), r(t)), \quad (39)$$

for each $i = 1, \dots, n$ with $\frac{\partial f}{\partial x_i}(x, r)$ being given by (36).

Example 5.2. Consider the problem $\min\{c^\top x \mid \sum_{i=1}^n x_i = 1,\, x \geq 0,\, \sum_{i=1}^{n-1} x_i^2 + (x_n - 1)^2 \leq 1\}$. Figure 3 illustrates the case $n = 3$, $c = (1, 3, 2)$, $u^0 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and $\Delta t = 0.1$ in an explicit Euler's scheme for the numerical integration of $(GF; u^0)$ and (39) analogous to (37) and (38), respectively.



(a) Taking $\psi(\lambda) = \lambda \log \lambda - \lambda$



(b) Taking $\psi(\lambda) = -\log \lambda$

Figure 3: Trajectories for $(B\text{-}GF; u^0)$ of Example 5.2 with $r_1(t) = \frac{1}{(t+10)^3}$ and $r_2(t) = \frac{1}{t+10}$ in red, and comparison with the trajectory of $(GF; u^0)$ in green.

5.3 Second-order cone

Suppose now that the problem we want to solve is the following

$$(P_3) \qquad \min\{f(x) \mid x \in \mathcal{L}^n_+,\, g_j(x) \leq 0,\ j \in I\},$$

where $\mathcal{L}^n_+ = \{x = (x_1, \bar{x}) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|\bar{x}\| \le x_1 \}$ (see Example 2.1). We will consider the so called *spectrally defined functions* h satisfying $(H_h; C)$ for $C = \mathcal{L}^n_{++}$ which are given by

$$h(x) = \psi(\lambda_1(x)) + \psi(\lambda_2(x)) = \psi(x_1 - \|\bar{x}\|) + \psi(x_1 + \|\bar{x}\|)$$

As before, some natural choices for ψ are the following: $\psi_1(\lambda) = \lambda \log \lambda - \lambda$ and $\psi_2(\lambda) = -\log(\lambda)$. It is well known (see, for instance, [1]) that the Hessian of h_1 corresponding to ψ_1 is given by

$$\nabla^2 h_1(x) = \frac{2}{\det(x)} \begin{pmatrix} x_1 & -\bar{x}^\top \\ -\bar{x} & \frac{\det(x)}{2\|\bar{x}\|} \log(\frac{x_1+\|\bar{x}\|}{x_1-\|\bar{x}\|}) I_{n-1} + (x_1 - \frac{\det(x)}{2\|\bar{x}\|} \log(\frac{x_1+\|\bar{x}\|}{x_1-\|\bar{x}\|})) \frac{\bar{x}\bar{x}^\top}{\|\bar{x}\|^2} \end{pmatrix}, \tag{40}$$

where $\det(x) = \lambda_1(x)\lambda_2(x) = x_1^2 - \|\bar{x}\|^2$, while the Hessian of h_2 corresponding to ψ_2 is given by

$$\nabla^2 h_2(x) = 2(\mathcal{Q}_x)^{-1}, \tag{41}$$

where

$$\mathcal{Q}_x = \begin{pmatrix} \|x\|^2 & 2x_1\bar{x}^\top \\ 2x_1\bar{x} & \det(x) I_{n-1} + 2\bar{x}\bar{x}^\top \end{pmatrix}.$$

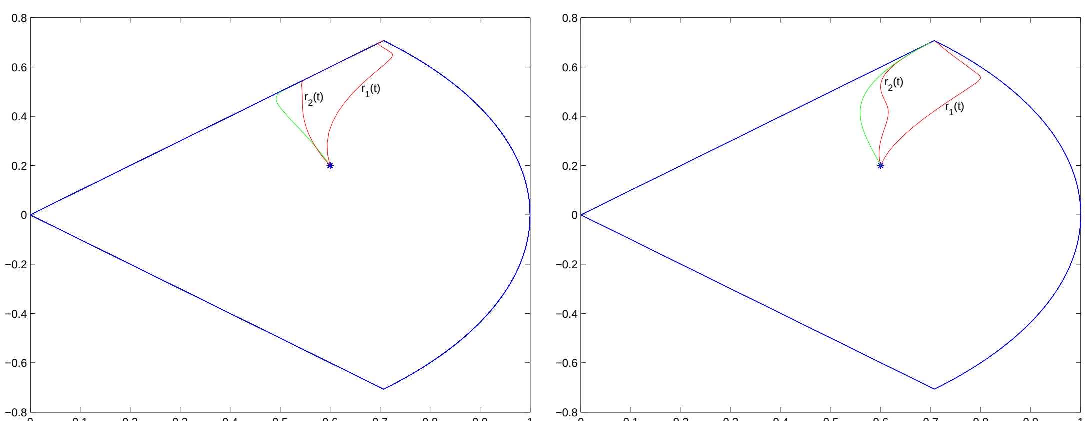
In particular, as there is no linear equality constraint, for the function h_2 the corresponding ODE is the following

$$\begin{aligned} \frac{du}{dt}(t) &= -\frac{1}{2} \mathcal{Q}_{u(t)} \nabla f(u(t), r(t)) \\ &= -\frac{1}{2} \begin{pmatrix} \|u(t)\|^2 (\nabla f)_1 + 2u_1(t)\bar{u}(t)^\top \overline{\nabla f} \\ 2u_1(t)(\nabla f)_1 \bar{u}(t) + \det(u(t))\overline{\nabla f} + 2(\bar{u}(t)^\top \overline{\nabla f})\bar{u}(t) \end{pmatrix}, \end{aligned} \tag{42}$$

where $\nabla f(u(t), r(t)) = ((\nabla f)_1, \overline{\nabla f})$ and $u(t) = (u_1(t), \bar{u}(t)) \in \mathbb{R} \times \mathbb{R}^{n-1}$.
On the other hand, for the second approach $h_{C \cap Q}(x) = \psi(\lambda_1(x)) + \psi(\lambda_2(x)) - \sum_{i \in J} \frac{1}{g_j(x)}$. In the case of the function ψ_2, the Hessian of the corresponding function $h_{C\cap Q}$ is given by

$$\nabla^2 h_{C\cap Q}(x) = 2\mathcal{Q}_x^{-1} - \sum_{j \in I} \frac{2}{g_j(x)^3} \nabla g_j(x) \nabla g_j(x)^\top + \sum_{j\in I} \frac{1}{g_j(x)^2} \nabla^2 g_j(x).$$

Example 5.3. Consider the problem $\min\{c^\top x \mid x \in \mathcal{L}^n_+,\ \sum_{i=1}^n x_i^2 \le 1\}$. The inverse barrier penalty function is the same as in the Example 5.1. In Figure 4 we consider the case $n = 2$, $c = (1, -2)$, $u^0 = (0.6, 0.2)$ and $\Delta t = 0.1$ for the corresponding Euler explicit schemes.



(a) Taking $\psi(\lambda) = \lambda \log \lambda - \lambda$

(b) Taking $\psi(\lambda) = -\log \lambda$

Figure 4: Trajectories in Example 5.3 for $r_1(t) = \frac{1}{(t+10)^3}$ and $r_2(t) = \frac{1}{t+10}$, and comparison with the trajectory $(GF; u^0)$.

6 Concluding remarks

It would be interesting to investigate the full convergence of the trajectories (resp. the sequences) generated by $(B\text{-}GF; u^0)$ (resp. (27)) in the degenerate case where the optimal set $S(P)$ is not a singleton. By virtue of Theorem 3.1 (resp. Theorem 4.1), when h is supposed to be a *Bregman function with zone* C (see [9, 23]), a standard argument [2, 12] shows that the full convergence result amounts to showing that $D_h(\bar{x}, u(t))$ (resp. $D_h(\bar{x}, x^k)$) converges for any $\bar{x} \in S(P)$, where $D_h(y, x)$ is given by (16) even for $y \in \partial C$. In absence of the penalty function, such a convergence property follows easily from the monotonicity of the (sub)gradient of the convex objective function; see, for instance, [2]. When combined with parametric barrier-penalty schemes, in general one cannot expect to have monotonic convergence. However, motivated by the full convergence results known for the Euclidean case where $h(x) = \frac{1}{2}\|x\|^2$ (see, for instance, [3, 4, 8, 13, 14]), it seems natural to try to find similar results for more general Bregman functions h, possibly under additional conditions on the parameterization $\varepsilon(t)$ (resp. $\{\varepsilon_k\}$).

On the other hand, different explicit discretization schemes for the system $(B\text{-}GF; u^0)$ can be viewed as numerical optimization algorithms to solve (P). It would be interesting to obtain convergence and rate of convergence results for some of them, following similar results in the autonomous case [19, 20, 25]. In this context, one may expect that a variable step-size Δt_k according to some suitable rule may improve the algorithms' performance. This should be supplemented with a complete numerical investigation, as the results presented here do not allow us to infer general conclusions.

Acknowledgments The authors would like to thank an anonymous referee for some useful suggestions that allowed them to improve the paper presentation.

References

[1] F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Mathematical Programming, vol. 95 (2003), no. 1, Ser. B, pp. 3-51.

[2] F. Alvarez, J. Bolte and O. Brahic, *Hessian Riemannian gradient flows in convex programming*, SIAM J. Control Optim., vol. 43 (2004), no 2, pp. 477-501.

[3] F. Alvarez and R. Cominetti, *Primal and dual convergence of a proximal point exponential penalty method for linear programming*, Math. Prog., vol. 93 (2002) no. 1, Ser. A, pp. 87-96.

[4] H. Attouch and R. Cominetti, *A dynamical approach to convex minimization coupling approximation with the steepest descent method*, J. Differential Equations, vol. 128 (1996), pp. 519-540.

[5] H. Attouch and M. Teboulle, *A regularized Lotka Volterra dynamical system as a continuous proximal-like method in optimization*, J. Optim. Theory Appl., vol. 121 (2004), no. 3, pp. 541-570.

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

[7] A. Auslender, *Penalty and barrier methods: a unified framework*, SIAM J. Optim., vol. 10 (1999), no. 1, pp. 211-230.

[8] J.-B. Baillon and R. Cominetti, *A convergence result for non-autonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming*, J. Funct. Anal., vol. 187 (2001), pp. 263–273.

[9] H. Bauschke, J.M. Borwein and P. Combettes, *Bregman monotone optimization algorithms*, SIAM J. Control Optim., vol. 42 (2003), no. 2, pp. 596-636.

[10] J. Bolte and M. Teboulle, *Barrier operators and associated gradient-like dynamical systems for constrained minimization problems*, SIAM J. Control Optim., vol. 42 (2003), no. 4, pp. 1266-1292.

[11] Y. Censor and S.A. Zenios, *The proximal minimization algorithm with D-function*, J. Optim. Theory and Appl., vol. 73 (1992), no. 3, pp. 451-464.

[12] G. Chen and M. Teboulle, *Convergence analysis of proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., vol. 3 (1993), no. 3, pp. 538-543.

[13] R. Cominetti, *Coupling the proximal point algorithm with approximation methods*, J. Optim. Theory Appl., vol. 95 (1995), pp. 581-600.

[14] R. Cominetti and M. Courdurier, *Coupling general penalty schemes for convex programming with the steepest descent and the proximal point algorithm*, SIAM J. Optim., vol. 13, (2002), no 3, pp. 745-765.

[15] J.X. da Cruz Neto, O.P. Ferreira, P.R. Oliveira and R.C.M. Silva, *Central paths in semidefinite programming, generalized proximal-point method and Cauchy trajectories in Riemannian manifolds*, J. Optim. Theory Appl., vol. 139 (2008), no. 2, pp. 227-242.

[16] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res., vol. 18 (1993), pp. 202-226.

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

[18] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., vol. 29 (1991), no 2, pp. 403-419.

[19] U. Helmke and J.B. Moore, "Optimization and dynamical systems", Communications and Control Engineering Series, Springer-Verlag London, Ltd., London, 1994.

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

[21] A.N. Iusem, B.F. Svaiter and J.X. da Cruz Neto, *Central paths, generalized proximal point methods and Cauchy trajectories in Riemannian manifolds*, SIAM J. Control Optim., vol. 37 (1999), no 2, pp. 566-588.

[22] N. Karmarkar, *Riemannian geometry underlying interior point methods for linear programming*, Contemporary Mathematics, vol. 114 (1990), pp. 51-75.

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

[24] K. Knopp, "Theory and Application of Infinite Series", Blackie & Son, Glasgow, 1951.

[25] G. Pacelli and M.C. Recchioni, *Monotone variable-metric algorithm for linearly constrained nonlinear programming*, J. Optim. Theory Appl., vol. 104 (2000), no. 2, pp. 255-279.

[26] B. Polyak, "Introduction to Optimization", Optimization Software, New York, 1987.

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

[28] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optimization 14 (1976), no. 5, pp. 877-898.