Noname manuscript No.
(will be inserted by the editor)

Duality in Convex Optimization for the Hyperbolic Augmented Lagrangian

Lennin Mallma-Ramirez · Nelson Maculan · Adilson Elias Xavier · Vinicius Layter Xavier

Received: date / Accepted: date

Lennin Mallma-Ramirez, Corresponding author

Federal University of Rio de Janeiro, Systems Engineering and Computer Science Program (COPPE). Rio de Janeiro, Brazil

lenninmr@gmail.com

Nelson Maculan

Federal University of Rio de Janeiro, Systems Engineering and Computer Science Program-Applied Mathematics (COPPE and IM). Rio de Janeiro, Brazil

maculan@cos.ufrj.br

Adilson Elias Xavier

Federal University of Rio de Janeiro, Systems Engineering and Computer Science Program (COPPE). Rio de Janeiro, Brazil

adilson.xavier@gmail.com

Vinicius Layter Xavier

Rio de Janeiro State University, Department of Statistics, Institute of Mathematics and Statistics, Graduate Program in Computational Sciences. Rio de Janeiro, Brazil

viniciuslx@ime.uerj.br

Abstract We guarantee the strong duality and the existence of a saddle point of the hyperbolic augmented Lagrangian function (HALF) in convex optimization. In order to guarantee these results, we assume a set of convexity hypothesis and the Slater condition. Finally we computationally illustrate our theoretical results obtained in this work.

Keywords Hyperbolic augmented Lagrangian · Nonlinear programming · Constrained convex optimization · Hyperbolic penalty function · Saddle point · Duality

Mathematics Subject Classification (2020) 90C30 · 90C25 · 90C46 · 65K05

1 Introduction

Throughout this work, we are interested in studying the following optimization problem

$$(P) \qquad x^* \in X^* = argmin\{f(x) \mid x \in S\},$$

where

$$S = \{x \in I\!R^n \mid g_i(x) \geq 0, \;\; i = 1, ..., m\},$$

is the convex feasible set of the problem (P) and where the function $f : I\!R^n \to I\!R$ is convex, $g_i : I\!R^n \to I\!R,\ i = 1, ..., m$, are concave functions, we assume also that f and g_i are continuously differentiable.

The augmented Lagrangian algorithms solve convex problem (P), see [12] and [13]. Recently it was guaranteed that the hyperbolic augmented Lagrangian algorithm (HALA) also solves problem (P), see [10]. In [10] it is guaranteed that HALA converges towards a Karush-Kuhn-Tucker (KKT) point. The HALA minimizes the HALF . With the solution found in the subproblem, the Lagrange multipliers will be updated, through an update formula. The theory of Duality is a current topic, this theory is studied in more general spaces (vector optimization), see [1]. In this work we are interested in developing the duality theory for HALF in the Euclidean space.

The main result of our work is: we guarantee the strong duality for the HALF, for the convex case. In this way we assure a solution to the primal and dual problems. With these results, we can also note that HALF has properties similar to Log-sigmoid Lagrangian function (LSLF), see [13]; modified Frisch function (MFF) and Modified Carroll function (MCF), these last two functions are studied in [12].

The work is organized as follows: Chapter 2, we present some basic results and we also present the hyperbolic penalty function and some of its properties. Chapter 3, we present the HALF. In Chapter 4, we developed the duality theory for HALF. In Chapter 5, we present a set of computational illustration to verify our theoretical results.

2 Preliminaries

Henceforth we consider the following assumption:

A. The optimal set X^* is nonempty, closed, bounded and, consequently, compact.

B. Slater constraint qualification holds, i.e., there exists $\hat{x} \in S$ which satisfies $g_i(\hat{x}) > 0,\ i = 1, ..., m$.

The Lagrangian function associated with the problem (P) is $L : I\!R^n \times I\!R^m_+ \to I\!R^n$, defined as

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x),$$

where, $\lambda_i \geq 0, i = 1, ..., m$, are called Lagrange multipliers. The dual function $\Phi : I\!R^m_+ \to I\!R$, is defined as

$$\Phi(\lambda) = \inf_{x \in I\!R^n} L(x, \lambda), \tag{2.1}$$

and the dual problem consists of finding

$$(D) \qquad \lambda^* \in \Lambda^* = argmax\{\Phi(\lambda) \mid \lambda \in I\!R^m_+\}.$$

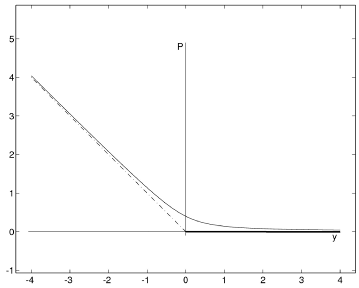

Fig. 1 Hyperbolic Penalty Function

2.1 Hyperbolic Penalty

The hyperbolic penalty method, introduced in [17], is meant to solve the problem (P). The penalty method adopts the hyperbolic penalty function (HPF) as follows

$$P(y, \lambda, \tau) = -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}, \tag{2.2}$$

where $P : (-\infty, +\infty) \times I\!R_+ \times I\!R_{++} \rightarrow I\!R$. The graphic representation of $P(y, \lambda, \tau)$, is as shown in Figure 1. Now in the present work, we consider $\tau > 0$ fixed.

Remark 2.1 *The HPF is originally proposed in [17] and also studied in [18]. In these works, the following properties are important for HPF:*

(a) $P(y, \lambda, \tau)$ *is asymptotically tangent to the straight lines* $r_1(y) = -2\lambda y$ *and* $r_2(y) = 0$ *for* $\tau > 0$.

(b) $\cdot\ P(y, \lambda, 0) = 0, \quad$ *for* $\ y \geq 0.$

· $P(y, \lambda, 0) = -2\lambda y, \quad for \quad y < 0.$

Due to the properties (a) *and* (b) *the HPF perform a smoothing of the penalty studied by Zangwill, see [20].*

For more details of the HPF, see [17] and [18]. In particular we use the following properties of the function P:

P0) $P(y, \lambda, \tau)$ is $k-$times continuously differentiable for any positive integer k for $\tau > 0$.

P1) $P(0, \lambda, \tau) = \tau$, for $\tau > 0$ and $\lambda \geq 0$.

P2) $P(y, \lambda, \tau)$ is strictly decreasing function of y, i.e.,

$$\nabla_y P(y, \lambda, \tau) = -\lambda \left(1 - \frac{\lambda y}{\sqrt{(\lambda y)^2 + \tau^2}} \right) < 0,$$

for $\tau > 0$ and $\lambda > 0$.

Remark 2.2 *For any* $\lambda \geq 0$, $y \geq 0$ *and* $\tau > 0$. *We have* $\tau^2 > 0$, *so we can obtain the following inequalities*

$$(\lambda y)^2 < (\lambda y)^2 + \tau^2,$$

follow,

$$-\lambda y - \sqrt{(\lambda y)^2 + \tau^2} < 0 < -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}, \tag{2.3}$$

considering the definition of the function P *in (2.3), we have*

$$P(y, \lambda, \tau) > 0. \tag{2.4}$$

3 Hyperbolic Augmented Lagrangian Function

We define the HALF of problem (P) by $L_H : I\!R^n \times I\!R^m_{++} \times I\!R_{++} \to I\!R$,

$$L_H(x,\lambda,\tau) = f(x) + \sum_{i=1}^{m} P(g_i(x),\lambda_i,\tau)$$

$$= f(x) + \sum_{i=1}^{m} \left(-\lambda_i g_i(x) + \sqrt{(\lambda_i g_i(x))^2 + \tau^2} \right), \tag{3.5}$$

where $\tau > 0$ is the penalty parameter and we assume a fixed valued. Note that this function belongs to class C^∞ if the involved functions $f(x)$ and $g_i(x),\ i = 1, ..., m$, are too.

Proposition 3.1 *Let us assume that if $f(x)$ and all $g_i(x) \in C^2$ and that $f(x)$ is strictly convex, then $L_H(x,\lambda,\tau)$ is strictly convex in $I\!R^n$ for any fixed $\lambda > 0$ and $\tau > 0$.*

Proof. We only need to prove that the Hessiana of L_H is defined positive. Let are $\lambda = (\lambda_1, ..., \lambda_m) > 0$ and $\tau > 0$ fixed. The Hessian of $L_H(x,\lambda,\tau)$ is

$$\nabla^2_{xx} L_H(x,\lambda,\tau) = \nabla^2_{xx} f(x) - \sum_{i=1}^{m} \lambda_i \nabla^2_{xx} g_i(x)$$

$$+ \sum_{i=1}^{m} \left(\frac{(\lambda_i)^2}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}} - \frac{(\lambda_i)^4 g_i^2(x)}{((\lambda_i g_i(x))^2 + \tau^2)^{\frac{3}{2}}} \right) \nabla_x g_i(x) \nabla_x^T g_i(x)$$

$$+ \sum_{i=1}^{m} \frac{(\lambda_i)^2 g_i(x)}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}} \nabla^2_{xx} g_i(x). \tag{3.6}$$

In (3.6), the $\nabla^2_{xx} g_i(x)$ is factored, so, we can rewrite (3.6), as follows

$$\nabla^2_{xx} L_H(x,\lambda,\tau) = \nabla^2_{xx} f(x) - \sum_{i=1}^{m} \lambda_i \left(1 - \frac{\lambda_i g_i(x)}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}} \right) \nabla^2_{xx} g_i(x)$$

$$+ \sum_{i=1}^{m} \left(\frac{(\lambda_i)^2}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}} - \frac{(\lambda_i)^4 g_i^2(x)}{((\lambda_i g_i(x))^2 + \tau^2)^{\frac{3}{2}}} \right) \nabla_x g_i(x) \nabla_x^T g_i(x). \tag{3.7}$$

On the other hand, since we have $\tau^2 > 0$, we can get

$$(\lambda_i g_i(x))^2 + \tau^2 > (\lambda_i g_i(x))^2 , \tag{3.8}$$

now we multiply by λ_i^2 in (3.8), so it follows that

$$\left((\lambda_i g_i(x))^2 + \tau^2\right) \lambda_i^2 > \lambda_i^4 g_i^2(x),$$

the above inequality, we can rewrite it as

$$\frac{\left((\lambda_i g_i(x))^2 + \tau^2\right)^{\frac{3}{2}}}{\left((\lambda_i g_i(x))^2 + \tau^2\right)^{\frac{1}{2}}} \lambda_i^2 > \lambda_i^4 g_i^2(x),$$

so,

$$\frac{\lambda_i^2}{\left((\lambda_i g_i(x))^2 + \tau^2\right)^{\frac{1}{2}}} > \frac{\lambda_i^4 g_i^2(x)}{\left((\lambda_i g_i(x))^2 + \tau^2\right)^{\frac{3}{2}}},$$

thus, it follows

$$\frac{\lambda_i^2}{\left((\lambda_i g_i(x))^2 + \tau^2\right)^{\frac{1}{2}}} - \frac{\lambda_i^4 g_i^2(x)}{\left((\lambda_i g_i(x))^2 + \tau^2\right)^{\frac{3}{2}}} > 0. \tag{3.9}$$

We replace (3.9) in (3.7) and since $-\lambda_i \left(1 - \frac{\lambda_i g_i(x)}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}}\right) < 0$, in (3.7), we get that, $\nabla^2_{xx} L_H(x, \lambda, \tau) > 0$, for $\lambda > 0$ and $\tau > 0$ fixed. ∎

Recall that strict convexity implies convexity.

Remark 3.1 *From Lemma 4.1 of [10] and Proposition 3.1 for any $\lambda > 0$ and any $\tau > 0$ there exists a unique minimizer*

$$\breve{x} = \breve{x}(\lambda, \tau) = argmin \{L_H(x, \lambda, \tau) \mid x \in I\!R^n\}$$

for problem (P) with the assumption **A**.

4 Duality

In this section, we adapt the classic results already existing in the literature: Chapter 9 of [11] and Section 7 of [12] for our HALF. The following result is also verified by MFF and MCF, see [12].

Proposition 4.1 *Consider the convex problem (P). Assume the assumption* **B** *it hold. Then $x^* \in S$ is a solution of problem (P) for any $\tau > 0$ if and only if:*

(i) There exists a vector $\lambda^* \geq 0$ such that

$$\lambda_i^* g_i(x^*) = 0,\ i = 1, ..., m \quad \textit{and} \quad L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau), \ \ \forall x \in I\!R^n. \tag{4.10}$$

(ii) The pair (x^*, λ^*) is a saddle point of L_H, that is,

$$L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau) \geq L_H(x^*, \lambda, \tau), \quad \forall x \in I\!R^n, \ \ \forall \lambda \in I\!R^m_+. \tag{4.11}$$

Proof. ($\Rightarrow$) Let any $\tau > 0$ fixed. Assume x^* is a solution for convex problem (P) satisfying the assumption **B**. Then system

$$f(x) - f(x^*) < 0,$$

$$-g_i(x) < 0,\ i = 1, ..., m,$$

has no solution in $I\!R^n$. Hence, by the Proper Separation Theorem (see, Theorem 2.26 (iv) of Dhara and Dutta [4]), there exists a vector $(\tilde{\lambda}, \hat{\lambda}) \neq (0, 0) \in I\!R \times I\!R^m$ such that

$$\tilde{\lambda}\left(f(x) - f(x^*)\right) - \sum_{i=1}^{m} \hat{\lambda}_i g_i(x) \geq 0,$$

for all $x \in I\!R^n$. We rewrite the inequality above as

$$\tilde{\lambda}\left(f(x) - f(x^*)\right) \geq \sum_{i=1}^{m} \hat{\lambda}_i g_i(x), \tag{4.12}$$

for all $x \in I\!R^n$. Now, we follow an analysis similar to Theorem 4.2 of [4], so by **B**, we have that there exists $\lambda_i^* = \frac{\hat{\lambda}_i}{\tilde{\lambda}},\ i = 1, ..., m$, with $\tilde{\lambda} > 0$. Then, by (4.12) we have

$$f(x) - f(x^*) \geq \sum_{i=1}^{m} \lambda_i^* g_i(x), \tag{4.13}$$

for all $x \in I\!R^n$. In particular, (4.13) holds for $x = x^*$. So we get

$$0 \geq \sum_{i=1}^{m} \lambda_i^* g_i(x^*). \tag{4.14}$$

On the other hand, since, $g_i(x^*) \geq 0$ and $\lambda_i^* \geq 0$ for $i = 1, ..., m$, then by (4.14) we obtain

$$\lambda_i^* g_i(x^*) = 0, \ \ i = 1, ..., m, \tag{4.15}$$

holds, so we have the first part of (4.10).

Now, we are interested in proving the second part of (4.10). From (4.15) and (4.13), we have

$$f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) = f(x^*) \leq f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x), \tag{4.16}$$

for all $x \in I\!R^n$. Now, since we have (4.15), also, we can obtain

$$(\lambda_i^* g_i(x^*))^2 + \tau^2 \leq (\lambda_i^* g_i(x))^2 + \tau^2, \ i = 1, ..., m,$$

so, we have the following

$$\sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2} \leq \sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x))^2 + \tau^2}, \tag{4.17}$$

considering (4.16) and (4.17), we have

$$L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau), \;\; \forall x \in I\!R^n, \tag{4.18}$$

in this way, we finish the proof of (4.10).

We are interested in verifying item (ii) now. But, first we will prove that $L_H(x^*, \lambda^*, \tau) = f(x^*) + m\tau$. Indeed, by definition of L_H, we have

$$L_H(x^*, \lambda^*, \tau) = f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}, \tag{4.19}$$

considering (4.15); (4.19) becomes

$$L_H(x^*, \lambda^*, \tau) = f(x^*) + m\tau. \tag{4.20}$$

On the other hand, as x^* is feasible, i.e.,

$$g_i(x^*) \geq 0, \;\; i = 1, ..., m. \tag{4.21}$$

By applying the property P2 of the HPF in (4.21), we obtain

$$P(g_i(x^*), \lambda_i, \tau) \leq P(0, \lambda_i, \tau), \;\; i = 1, ..., m. \tag{4.22}$$

By applying property P1, on the right side of expression (4.22), we will obtain

$$P(g_i(x^*), \lambda_i, \tau) \leq \tau, \quad for \quad \lambda_i \geq 0, \quad i = 1, ..., m. \tag{4.23}$$

By performing the sum of 1 to m in (4.23) it follows that

$$\sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau) \leq \sum_{i=1}^{m} \tau = m\tau.$$

Adding $f(x^*)$ to both sides of the expression, we obtain

$$f(x^*) + \sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau) \leq f(x^*) + m\tau. \tag{4.24}$$

By definition of L_H, (4.24) becomes

$$L_H(x^*, \lambda, \tau) \leq f(x^*) + m\tau. \tag{4.25}$$

Now, by (4.25) and (4.20) we have

$$L_H(x^*, \lambda, \tau) \leq f(x^*) + m\tau = L_H(x^*, \lambda^*, \tau). \tag{4.26}$$

Finally, from (4.18) and (4.26), there is $\lambda^* \geq 0$ such that the primal-dual solution (x^*, λ^*) is a saddle point of L_H, $\forall x \in I\!R^n$ and $\tau > 0$.

($\Leftarrow$) We assume that (x^*, λ^*) is a saddle point of L_H, so (4.11) is hold. Then, for all $x \in I\!R^n$, $\lambda \in I\!R^m_+$ and for any $\tau > 0$ fixed, we have

$$f(x^*) - \sum_{i=1}^{m} \lambda_i g_i(x^*) + \sum_{i=1}^{m} \sqrt{\left(\lambda_i g_i(x^*)\right)^2 + \tau^2} = L_H(x^*, \lambda, \tau)$$

$$\leq L_H(x^*, \lambda^*, \tau) = f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{m} \sqrt{\left(\lambda_i^* g_i(x^*)\right)^2 + \tau^2}. \tag{4.27}$$

From (4.27), we have

$$-\sum_{i=1}^{m} \lambda_i g_i(x^*) + \sum_{i=1}^{m} \sqrt{\left(\lambda_i g_i(x^*)\right)^2 + \tau^2}$$

$$\leq -\sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{m} \sqrt{\left(\lambda_i^* g_i(x^*)\right)^2 + \tau^2}, \tag{4.28}$$

for all $\lambda_i \geq 0,\ i = 1, ..., m$.

This relation (4.28) is possible only if $g_i(x^*) \geq 0$. Since, if this relation is violated (*i.e.*, $g_i(x^*) < 0$) for some index i, we can choose λ_i sufficiently large such that (4.28) becames false. So, x^* is feasible for problem (P).

We will prove the complementarity condition of (4.10). So again, by (4.28), and since that $\lambda_i \geq 0,\ i = 1, ..., m$, in particular taking $\lambda_i = 0,\ i = 1, ..., m$, in (4.28), we obtain

$$\sum_{i=1}^{m} \tau \leq -\sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2},$$

thus, it follows that

$$\sum_{i=1}^{m} \left(\lambda_i^* g_i(x^*) + \tau\right)^2 \leq \sum_{i=1}^{m} \left((\lambda_i^* g_i(x^*))^2 + \tau^2\right),$$

$$\sum_{i=1}^{m} \Big((\lambda_i^* g_i(x^*))^2 + \tau^2 + 2\tau\lambda_i^* g_i(x^*) \Big) \leq \sum_{i=1}^{m} \left((\lambda_i^* g_i(x^*))^2 + \tau^2\right),$$

so,

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leq 0,$$

and since $\lambda_i^* \geq 0$ and $g_i(x^*) \geq 0,\ i = 1, ..., m$, it must be true

$$\lambda_i^* g_i(x^*) = 0,\ i = 1, ..., m. \tag{4.29}$$

By (4.29) and definition of L_H, we obtain

$$L_H(x^*, \lambda^*, \tau) = f(x^*) + m\tau. \tag{4.30}$$

From definition of saddle point, we know that $L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau)$, by (4.30) and by definition of L_H, we can write

$$f(x^*) + m\tau = L_H(x^*, \lambda^*, \tau) \leq L_H(x, \lambda^*, \tau) = f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i^*, \tau). \tag{4.31}$$

On the other hand, once again considering property P2 of HPF, for any feasible point x, i.e., $g_i(x) \geq 0,\ i = 1, ...m$, we will carry out a work similar to that of (4.21)-(4.24), thus, we can obtain

$$f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i^*, \tau) \leq f(x) + m\tau, \tag{4.32}$$

now, we replace (4.32) in (4.31), then follow

$$f(x^*) + \tau m \leq f(x) + \tau m,$$

from this last inequality, we obtain $f(x^*) \leq f(x)$, whenever x is feasible. Therefore, x^* is a global optimal solution of (P). ■

Let's consider the following definitions. Let

$$F_\tau(x) = \sup_{\lambda \geq 0} L_H(x, \lambda, \tau).$$

Then $F_\tau(x) = f(x) + m\tau$, if $g_i(x) \geq 0,\ i = 1, ..., m$ and $F_\tau(x) = \infty$, otherwise. Therefore, we can consider the following problem

$$x^* = argmin \left\{ F_\tau(x) \mid x \in I\!R^n \right\}, \tag{4.33}$$

that is the problem (P) reduces to solving (4.33).

Let

$$\phi_\tau(\lambda) = \inf_{x \in I\!R^n} L_H(x, \lambda, \tau)$$

(possibly $\phi_\tau(\lambda) = -\infty$ for some λ) and consider the following dual problem of (P), that consisting of finding

$$\lambda^* = argmax \left\{ \phi_\tau(\lambda) \mid \lambda \geq 0 \right\}. \tag{4.34}$$

In the following result, we are going to verify the weak duality.

Proposition 4.2 *Let x be a feasible solution to problem (P) and let λ be a feasible solution to problem (4.34). Then*

$$\phi_\tau(\lambda) \leq F_\tau(x) = f(x) + m\tau, \quad \forall x \in S, \ \forall \ \lambda \in I\!R^m_+.$$

Proof. For any feasible x and λ, we then we can get the weak duality. Indeed, by the definition of ϕ_τ, we have

$$\begin{aligned} \phi_\tau(\lambda) = \inf_{w \in I\!R^n} L_H(w, \lambda, \tau) &= \inf_{w \in I\!R^n} \left\{ f(w) + \sum_{i=1}^{m} P(g_i(w), \lambda_i, \tau) \right\} \\ &\leq \inf_{w \in S} \left\{ f(w) + \sum_{i=1}^{m} P(g_i(w), \lambda_i, \tau) \right\} \\ &= f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i, \tau). \end{aligned} \tag{4.35}$$

Since we know that x is feasible, we have $g_i(x) \geq 0, \ i = 1, ..., m$, immediately then, for the property P2 of the HPF, we get the following expressions

$$P(g_i(x), \lambda_i, \tau) \leq P(0, \lambda_i, \tau), \ \ i = 1, ..., m,$$

we rewrite the expression above, as follows

$$\sum_{i=1}^{m} P(g_i(x), \lambda_i, \tau) \leq \sum_{i=1}^{m} P(0, \lambda_i, \tau),$$

now, we apply property P1, on the right side of the previous inequality

$$\sum_{i=1}^{m} P(g_i(x), \lambda_i, \tau) \leq \sum_{i=1}^{m} \tau = m\tau,$$

we add $f(x)$, to both sides of the inequality above

$$f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i, \tau) \leq f(x) + m\tau, \tag{4.36}$$

we replace (4.36) in (4.35), so

$$\phi_\tau(\lambda) \leq f(x) + m\tau, \quad \forall x \in S, \;\; \forall \lambda \in I\!R^m_+. \tag{4.37}$$

■

If $\hat{x}$ and $\hat{\lambda}$ are feasible solutions of the primal and dual problems and $F_\tau(\hat{x}) = \phi_\tau(\hat{\lambda})$, then $\hat{x} = x^*$ and $\hat{\lambda} = \lambda^*$. From Remark 3.1, with the smoothness of $f(x)$ and $g_i(x),\ i = 1, ..., m$, we ensure the smoothness for the dual function $\phi_\tau(\lambda)$.

Theorem 4.1 *The problem (P) is considered. The assumption* **B** *is verified. Then the existence of a solution of problem (P) implies that the problem (4.34) has a solution and*

$$\phi_\tau(\lambda^*) = f(x^*) + m\tau, \;\; \textit{for any} \;\; \tau > 0. \tag{4.38}$$

Proof. Let x^* be a solution of problem (P). By **B**, we get $\lambda^* \geq 0$, such that (4.10) is verified. So we have

$$\phi_\tau(\lambda^*) = \min_{x \in I\!R^n} L_H(x, \lambda^*, \tau) = L_H(x^*, \lambda^*, \tau)$$

$$\geq L_H(x^*, \lambda, \tau) \geq \min_{x \in I\!R^n} L_H(x, \lambda, \tau) = \phi_\tau(\lambda), \;\; \forall \lambda \geq 0.$$

Therefore $\phi_\tau(\lambda^*) = \max\left\{\phi_\tau(\lambda) \mid \lambda \in I\!R^m_+\right\}$, in this way $\lambda^* \in I\!R^m_+$ is a solution of the dual problem and since we have $L_H(x^*, \lambda^*, \tau) = f(x^*) + m\tau$, so (4.38) hold. ■

Proposition 4.3 *Suppose that (4.38) holds, for the viable points x^* and λ^*, then x^* is a solution of the problem (P) and λ^* is a solution of the dual problem (4.34).*

Proof. Let $g_i(x^*) \geq 0,\ i = 1, ..., m$, with $x^* \in S$, $\lambda_i^* \geq 0,\ i = 1, ..., m$ and (4.38) with $\tau > 0$ fixed. Then for (4.37) where x and λ are viable, we can obtain the following

$$f(x) + m\tau \geq \phi_\tau(\lambda^*) = f(x^*) + m\tau \geq \phi_\tau(\lambda),$$

that is, x^* is solution of the problem (P) and λ^* is solution of (4.34), which corresponds the validity of the strong duality. ■

5 Computational Illustration

We use HALA, proposed in [10] to guarantee the theory proposed in this work. The program were compiled by the GNU Fortran compiler version 4:7.4.0-1ubuntu2.3. The numerical Experiments are conducted on a Notebook with operating system Ubuntu 18.04.5, CPU i7-3632QM and 8GB RAM. The unconstrained minimization tasks were carried out by means of a Quasi-Newton algorithm employing the BFGS updating formula, with the function

VA13 from HSL library [7]. The algorithm stop when the solution is viable (feasible) an the absolute value of the difference of the two consecutive solutions $\|x^k - x^{k-1}\|$ is less than $1.D-5$.

We are going to take advantage of this section to make some comparisons of our algorithm HALA (see Table 14) with respect to the following algorithms:

Alg1=[5] which is an truncated Newton method;

Alg2=[6] which is a primal-dual interior point method;

Alg3=[15] which is an interior-point algorithm;

Alg4= [14] which is a QP-free method;

Alg5=[3] which is a primal-dual feasible interior-point method;

Alg6=[19] which is a feasible sequential linear equation algorithm;

Alg7=[16] which is an inexact first-order method;

Alg8=[8] which is a feasible direction interior-point technique;

Alg9=[2] which is an interior point algorithm.

For completeness reasons, we are going to present HALA:

5.1 **Algorithm HALA**

Step 1. Let $k := 0$ (initialization).

Take initial values $\lambda^0 = (\lambda_1^0, ..., \lambda_m^0) \in \mathbb{R}_{++}^m$, $\tau \in \mathbb{R}_{++}$.

Step 2. Solve the unconstrained minimization problem (primal update):

$$x^{k+1} \in argmin_{x\in I\!R^n}\ L_H(x,\lambda^k,\tau)$$
$$= argmin_{x\in I\!R^n}\left\{ f(x) + \sum_{i=1}^{m}\left(-\lambda_i^k g_i(x) + \sqrt{\left(\lambda_i^k g_i(x)\right)^2+\tau^2}\right)\right\}.$$

Step 3. Updating of Lagrange multipliers (dual update):

$$\lambda_i^{k+1} = \lambda_i^k\left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{(\lambda_i^k g_i(x^{k+1}))^2+\tau^2}}\right),\quad i=1,...,m. \tag{5.39}$$

Step 4. If the pair (x^{k+1},λ^{k+1}) satisfies the stopping criteria: Then Stop.

Step 5. $k := k+1$. Go to Step 2.

With the following examples proposed in the book [9], we are going to verify the strong duality. On the other hand, in each example, the value of m means the total number of restrictions. Also, in all the examples starting points are considered, so that assumption **B** is verified.

Example 5.1 *Problem 1 (HS1).*

$$\min_{x\in I\!R^2} f(x) = 100(x_2 - x_1^2)^2 + (1-x_1)^2$$
$$s.t.\ \ g_1(x) = x_2 + 1.5 \geq 0.$$

Starting with $x^0 = (-2,1)$ (feasible), $f(x^0) = 909$ and $m = 1$. The minimum value is $f(x^*) = 0$ at the optimal solution $x^* = (1,1)$.

Example 5.2 *Problem 30 (HS30).*

$$\min_{x\in I\!R^3} f(x) = x_1^2 + x_2^2 + x_3^2$$
$$s.t.\ \ g_1(x) = x_1^2 + x_2^2 - 1 \geq 0,$$
$$g_2(x) = x_1 - 1 \geq 0,$$

$$g_3(x) = 10 - x_1 \geq 0,$$

$$g_4(x) = x_2 + 10 \geq 0,$$

$$g_5(x) = 10 - x_2 \geq 0,$$

$$g_6(x) = x_3 + 10 \geq 0,$$

$$g_7(x) = 10 - x_3 \geq 0.$$

Starting with $x^0 = (1,1,1)$ (feasible), $f(x^0) = 3$ and $m = 7$. The minimum value is $f(x^*) = 1$ at the optimal solution $x^* = (1,0,0)$.

Example 5.3 *Problem 66 (HS66).*

$$\min_{x \in I\!R^3} f(x) = 0.2x_3 - 0.8x_1$$

$$s.t.\;\; g_1(x) = x_2 - e^{x_1} \geq 0,$$

$$g_2(x) = x_3 - e^{x_2} \geq 0,$$

$$g_3(x) = x_1 \geq 0,$$

$$g_4(x) = x_2 \geq 0,$$

$$g_5(x) = x_3 \geq 0,$$

$$g_6(x) = 100 - x_1 \geq 0,$$

$$g_7(x) = 100 - x_2 \geq 0,$$

$$g_8(x) = 10 - x_3 \geq 0.$$

Starting with $x^0 = (0, 1.05, 2.9)$ (feasible), $f(x^0) = 0.58$ and $m = 8$. The minimum value is $f(x^*) = 0.5181632741$ at the optimal solution

$x^* = (0.1841264879, 1.202167873, 3.327322322)$.

Example 5.4 *Problem 76 (HS76).*

$$\min_{x\in I\!R^4} f(x) = x_1^2 + 0.5x_2^2 + x_3^2 + 0.5x_4^2 - x_1x_3 + x_3x_4 - x_1 - 3x_2 + x_3 - x_4$$

$$s.t.\;\; g_1(x) = 5 - x_1 - 2x_2 - x_3 - x_4 \geq 0,$$

$$g_2(x) = 4 - 3x_1 - x_2 - 2x_3 + x_4 \geq 0,$$

$$g_3(x) = x_2 + 4x_3 - 1.5 \geq 0,$$

$$g_4(x) = x_1 \geq 0,$$

$$g_5(x) = x_2 \geq 0,$$

$$g_6(x) = x_3 \geq 0,$$

$$g_7(x) = x_4 \geq 0.$$

Starting with $x^0 = (0.5, 0.5, 0.5, 0.5)$ (feasible), $f(x^0) = -1.25$ and $m = 7$. The minimum value is $f(x^*) = -4.681818181$ at the optimal solution $x^* = (0.2727273, 2.090909, -0.26E-10, 0.5454545)$.

Example 5.5 *Problem 100 (HS100).*

$$\min_{x\in I\!R^7} f(x) = (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6$$
$$+7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7$$

$$s.t.\;\; g_1(x) = 127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0,$$

$$g_2(x) = 282 - 7x_1 - 3x_2 - 10x_3^2 - x_4 + x_5 \geq 0,$$

$$g_3(x) = 196 - 23x_1 - x_2^2 - 6x_6^2 + 8x_7 \geq 0,$$

$$g_4(x) = -4x_1^2 - x_2^2 + 3x_1x_2 - 2x_3^2 - 5x_6 + 11x_7 \geq 0.$$

Starting with $x^0 = (1, 2, 0, 4, 0, 1, 1)$ (feasible), $f(x^0) = 714$ and $m = 4$. The minimum value is $f(x^*) = 680.6300573$ at the optimal solution $x^* = (2.330499, 1.951372, -0.4775414, 4.365726, -0.6244870, 1.038131, 1.594227)$.

5.2 Results

For each table, the letter N indicates the name of the problem, λ is the multiplier Lagrange, x is the primal variable, $f(x)$ is the value of the objective function, $g_i(x)$ are the constraints of each problem, $L_H(\cdot, \cdot, \cdot)$ is the value of the HALF and $via = viable = feasible$ where, in each iteration, the obtained point can be viable, then its value is "$0 = yes''$ or the point can be inviable, then the value is "$1 = not''$ and τ is the penalty parameter. In all of our examples, we will use $\tau = 0.10E - 04$. We are going to analyze the Examples.

- Example 5.1: The HALA solves this example even though function f is nonconvex, see Tables 1 and 2.

- Example 5.2: the function f is strictly convex. From Table 3, we can see that in iteration 2 the Theorem 4.1 can be verified, that is, we have the following

$$f(x^*) + m\tau = 1.00000000 + (7)(0.00001) = 1.00007$$

and

$$\phi_\tau(\lambda^*) = L_H(x^*, \lambda^*, \tau) = 1.00007,$$

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. So, $x^* = (0.100000000E+01, 0.100000000E+01)$ is the solution of the primal problem and from Table 4 and Table 5, we can see the λ^* is the solution of the dual problem in the iteration 2.

- Example 5.3: the function f is linear. From Table 6, we can see that in iteration 3 the Theorem 4.1 can be verified, that is, we have the following

$$f(x^*) + m\tau = 0.518163274 + (8)(0.00001) = 0.518243274$$

and

$$\phi_\tau(\lambda^*) = L_H(x^*, \lambda^*, \tau) = 0.518243274,$$

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. So, x^* is the solution of the primal problem and from Table 7 and Table 8, we can see the λ^* is the solution of the dual problem in the iteration 3.

- Example 5.4: the function f is strictly convex. From Table 9, we can see that in iteration 2, the Theorem 4.1 can be verified, that is, we have the following

$$f(x^*) + m\tau = -4.68181818 + (7)(0.00001) = -4.68174818$$

and

$$\phi_\tau(\lambda^*) = L_H(x^*, \lambda^*, \tau) = -4.68174818,$$

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. So, x^* is the solution of the primal problem and from Table 10 and Table 11, we can see the λ^* is the solution of the dual problem in the iteration 2.

- Example 5.5: the function f is convex. From Table 5.2, we can see that in iteration 2, the Theorem 4.1 can be verified, that is, we have the following

$$f(x^*) + m\tau = 680.630057 + (4)(0.00001) = 680.630097$$

and

$$\phi_\tau(\lambda^*) = L_H(x^*, \lambda^*, \tau) = 680.630097,$$

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. The optimal value x^* is reported in the Table 5.2 and the optimal value λ^* is reported in the Table 12.

In Table 14: we can see that HALA is more efficient in the sense that it uses fewer iterations with respect to the other algorithms. We can observe in the computational results that the HALA remains in the viable region in all the examples. On the other hand, despite being the theory developed in this work on convexity hypothesis, our algorithm shows in the Example 5.1 that it can also solve non-convex problems.

Acknowledgements This work was partially supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq*; FAPERJ E-26/201.176/2020 (258168) and COPPETEC Foundation, Brazil.

Table 1 Example 5.1

k	x_1	x_2	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	-0.200000000E+01	0.100000000E+01	0.909000000E+03	0.909000000E+03	0
1	0.100000000E+01	0.100000000E+01	0.976202768E-23	0.200017780E-11	0
2	0.100000000E+01	0.100000000E+01	0.976202768E-23	0.999999800E-05	0

Table 2 Example 5.1

	$g_1(x)$	
k	via	λ_1
0	0	0.100000000E+02
1	0	0.800470801E-12
2	0	0.800470641E-12

Table 3 Example 5.2

k	x_1	x_2	x_3	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	0.100000000E+01	0.100000000E+01	0.100000000E+01	0.300000000E+01	0.300001000E+01	0
1	0.100000179E+01	-0.773363648E-10	-0.477975865E-10	0.100000359E+01	0.100000755E+01	0
2	0.100000000E+01	-0.198472639E-09	0.162237191E-09	0.100000000E+01	0.100007000E+01	0

Table 4 Example 5.2

	$g_1(x)$		$g_2(x)$		$g_3(x)$	
k	via	λ_1	via	λ_2	via	λ_3
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	0	0.367214575E+00	0	0.126557312E+01	0	0.610622664E-13
2	0	0.367214283E+00	0	0.126557139E+01	0	0.610622630E-13

References

1. Adán M, Novo V (2005) Duality and saddle-points for convex-like vector optimization problems on real linear spaces. Top 13(2):343-357

Table 5 Continuation of Table 4

	$g_4(x)$		$g_5(x)$		$g_6(x)$		$g_7(x)$	
k	via	λ_4	via	λ_5	via	λ_6	via	λ_7
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	0	0.499600361E-13	0	0.499600361E-13	0	0.499600361E-13	0	0.499600361E-13
2	0	0.499600336E-13	0	0.499600336E-13	0	0.499600336E-13	0	0.499600336E-13

Table 6 Example 5.3

k	x_1	x_2	x_3	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	0.000000000E+00	0.105000000E+01	0.290000000E+01	0.580000000E+00	0.580010000E+00	0
1	0.184125306E+00	0.120216905E+01	0.332733118E+01	0.518165991E+00	0.518168851E+00	0
2	0.184126486E+00	0.120216787E+01	0.332732231E+01	0.518163274E+00	0.518243274E+00	0
3	0.184126486E+00	0.120216787E+01	0.332732231E+01	0.518163274E+00	0.518243274E+00	0

Table 7 Example 5.3

	$g_1(x)$		$g_2(x)$		$g_3(x)$		$g_4(x)$	
k	via	λ_1	via	λ_2	via	λ_3	via	λ_4
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	0	0.665464690E+00	0	0.200000037E+00	0	0.147483137E-09	0	0.345945494E-11
2	0	0.665464311E+00	0	0.199999981E+00	0	0.147482736E-09	0	0.345945351E-11
3	0	0.665463933E+00	0	0.199999924E+00	0	0.147482336E-09	0	0.345945207E-11

Table 8 Continuation of Table 7

	$g_5(x)$		$g_6(x)$		$g_7(x)$		$g_8(x)$	
k	via	λ_5	via	λ_6	via	λ_7	via	λ_8
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	0	0.450750548E-12	0	0.111022302E-14	0	0.000000000E+00	0	0.113242749E-12
2	0	0.450750480E-12	0	0.111022301E-14	0	0.000000000E+00	0	0.113242740E-12
3	0	0.450750413E-12	0	0.111022300E-14	0	0.000000000E+00	0	0.113242731E-12

Table 9 Example 5.4

k	x_1	x_2	x_3	x_4	$f(x)$	$L_H(x,\lambda,\tau)$	via
0	0.500000000E+00	0.500000000E+00	0.500000000E+00	0.500000000E+00	-0.125000000E+01	-0.125000000E+01	0
1	0.272727650E+00	0.209090766E+01	0.147253122E-05	0.545452356E+00	-0.468181418E+01	-0.468180958E+01	0
2	0.272727273E+00	0.209090909E+01	0.413220517E-10	0.545454545E+00	-0.468181818E+01	-0.468174818E+01	0

Table 10 Example 5.4

	$g_1(x)$		$g_2(x)$		$g_3(x)$		$g_4(x)$	
k	via	λ_1	via	λ_2	via	λ_3	via	λ_4
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	0	0.454545522E+00	0	0.186739513E-11	0	0.143196566E-10	0	0.672217837E-10
2	0	0.454545455E+00	0	0.186739456E-11	0	0.143196445E-10	0	0.672216605E-10

Table 11 Continuation of Table 10

	$g_5(x)$		$g_6(x)$		$g_7(x)$	
k	via	λ_5	via	λ_6	via	λ_7
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	0	0.114352972E-11	0	0.172728506E+01	0	0.168043357E-10
2	0	0.114352972E-11	0	0.172728506E+01	0	0.168043357E-10

Table 12 Example 5.5

	$g_1(x)$		$g_2(x)$		$g_3(x)$		$g_4(x)$	
k	via	λ_1	via	λ_2	via	λ_3	via	λ_4
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	0	0.113971988E+01	0	0.000000000E+00	0	0.000000000E+00	0	0.368614695E+00
2	0	0.113971989E+01	0	0.000000000E+00	0	0.000000000E+00	0	0.368614517E+00

2. Akrotirianakis I, Rustem B (2005) Globally convergent interior-point algorithm for nonlinear programming. Journal of Optimization Theory and Applications 125(3):497-521
3. Bakhtiari S, Tits AL (2003) A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent. Computational Optimization and Applications 25:17-38

Table 13 Example 5.5

k	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$f(x)$	$L_H(x,\lambda,\tau)$	via
0	0.100000000E+01	0.200000000E+01	0.000000000E+00	0.400000000E+01	0.000000000E+00	0.100000000E+01	0.100000000E+01	0.714000000E+03	0.714000000E+03	0
1	0.233049916E+01	0.195137237E+01	-0.477541322E+00	0.436572623E+01	-0.624486971E+00	0.103813098E+01	0.159422676E+01	0.680630061E+03	0.680630065E+03	0
2	0.233049937E+01	0.195137237E+01	-0.477541393E+00	0.436572623E+01	-0.624486971E+00	0.103813102E+01	0.159422671E+01	0.680630057E+03	0.680630097E+03	0

Table 14 Iterations

N	HALA	Alg1	Alg2	Alg3	Alg4	Alg5	Alg6	Alg7	Alg8	Alg9
HS1	2	18	34	32	40	24	36	260	36	27
HS30	2	3	8	11	7	7	10	7	11	10
HS66	3	12	12	13		11	11	23	5	20
HS76	2	28	9	11	10		9	23	7	12
HS100	2	18	10	11	15	9	14	99	13	14

4. Dhara A, Dutta J (2012) Optimality Conditions in Convex Optimization: A finite-Dimensional View. CRC Press, Boca Raton

5. Di Pillo G, Lucidi S, Palagi L (2000) A truncated Newton method for constrained optimization. In G. Di Pillo and F. Giannessi, editors, Nonlinear Optimization and Related Topics. Kluwer Academic, Dordrecht

6. Griva I, Shanno DF, Vanderbei RJ, Benson HY (2008) Global convergence of a primal-dual interior point method for nonlinear programming. Algorithmic Operations Research 3:12-29

7. Harwell Subroutine Library (2002) A collection of Fortran codes for large scale scientific computation, $url = https://www.hsl.rl.ac.uk/$

8. Herskovits J (1998) Feasible direction interior-point technique for nonlinear optimization. Journal of Optimization Theory and Applications 99:121-146

9. Hock W, Schittkowski K (1981) Test Examples for Nonlinear Programming Codes. Lecture Notes in Economics and Mathematical Systems 187, Springer-Verlag, Berlin, Germany

10. Mallma-Ramirez L, Maculan N, Xavier AE (2021) Convergence analysis of the hyperbolic augmented Lagrangian algorithm. Technical Report, PESC-COPPE, Federal University of Rio de Janeiro. ES-2991/21 (https://www.cos.ufrj.br/index.php/pt-BR/publicacoes-pesquisa/details/15/2991)

11. Polyak BT (1987) Introduction to Optimization. Optimization Software, Inc., New York

12. Polyak RA (1992) Modified barrier functions: theory and methods. Mathematical Programming 54:177-222

13. Polyak RA (2001) Log-sigmoid multipliers method in constrained optimization. Annals of Operations Research 101:427-460

14. Qi HD, Qi L (2000) A new qp-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. SIAM J OPTIM 11:113-132

15. Vanderbei RJ, Shanno DF (1999) An interior-point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications 13:231-252

16. Wang H, Zhang F, Wang J, Rong Y. An inexact first-order method for constrained nonlinear optimization. Optimization Methods and Software. https://doi.org/10.1080/10556788.2020.1712601 (online)

17. Xavier AE (1982) *Penalização Hiperbólica- Um Novo Método para Resolução de Problemas de Otimização* (master's thesis). Federal University of Rio de Janeiro \ COPPE, Rio de Janeiro, Brazil

18. Xavier AE (2001) Hyperbolic penalty: a new method for nonlinear programming with inequalities. Intl Trans in Op Res 8:659-671

19. Yang YF, Li DH, Qi L (2003) A feasible sequential linear equation method for inequality constrained optimization. SIAM J OPTIM 13(4):1222-1244

20. Zangwill WI (1967) Non-linear programming via penalty function. Management Science 13(5):344-358