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

Convergence Analysis of the Hyperbolic Augmented Lagrangian Algorithm

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

viniciuslx@ime.uerj.br

Abstract The Hyperbolic Augmented Lagrangian Algorithm (HALA) is a novel algorithm proposed in this work for solving the constrained nonlinear programming problem. Under mild assumptions, such as: convexity, Slater's qualification and differentiability, the convergence of the proposed algorithm is proved. Finally, in order to illustrate the algorithm, we present some computational experiments.

Keywords Hyperbolic augmented Lagrangian · Nonlinear programming · Constrained optimization · Constraint qualification · Hyperbolic penalty · Convergence · Convex problem

Mathematics Subject Classification (2000) 90C30 · 90C25 · 90C46 · 65K05

1 Introduction

We are interested in the nonlinear programming problem subject to inequality constraints

$$min\,\{f(x) \mid x \in S\}\,, \tag{1.1}$$

where $S = \{x \in I\!R^n \mid g_i(x) \geq 0,\ i = 1, ..., m\}$, f and $g_i,\ i = 1, ..., m$ are real-valued functions defined on $I\!R^n$.

The primal methods solve the problem (1.1), some of them are the gradient projection method (see [32]), cutting-plane method (see [16]) and the feasible direction methods (see [50]). For a better idea of these methods, see the book of Minoux [22]. On the other hand, the dual methods also solve the problem (1.1) are for example: the barrier methods, where the logarithmic barrier function (LBF) or also the inverse barrier function (IBF) is used (see [10]); penalty methods (see [5]) and mixed interior-exterior penalty method (see [10] and [22]) also have an important role.

Later the augmented Lagrangian algorithms are studied, some basic references are [12] and [27]. Now, in [3], the advantages of using the multiplier methods (also called the augmented lagrangian method) over the penalty methods are shown, also see Chapter 6 of [22]. The augmented Lagrangian methods are widely used for problems with constraints (1.1). The idea of these methods is to convert the constrained problem into a sequence of the unconstrained problems.

Currently, there are a wide variety of augmented Lagrangian algorithms that solve problem (1.1), for example: entropy-like multiplier methods, see [14]; nonlinear rescaling algorithm, see [25]; penalty/barrier multiplier method, see [2] and multiplier methods based on second order homogeneous kernels, see [1]. The functions LBF and IBF are modified and used in the context of the augmented Lagrangian algorithms, see [24].

A solution of the problem (1.1) subject to equality constraints is proposed in Hestenes [12] and Powell [27]. Later, thus Hestenes-Powell formulation was adapted for the nonlinear programming problem subject to inequality constraints (see [30]). This adaptation defines an augmented Lagrangian function without continuous second derivatives. This new formulation is known as Hestenes-Powell-Rockafellar augmented Lagrangian function. This function had a very important role to construct a new augmented Lagrangian function, which is continuously differentiable, see [8]. Other differentiable augmented Lagrangian functions are proposed, see [21], [17], [1] and [26].

The authors in [36] study the exponential multiplier method, proposed by [19]. They study two rules for choosing the penalty parameters and guarantees that the primal sequence converges in an ergodic sense. Other works, where the convergence of the sequence primal is studied in an ergodic sense, are [14], [15], [18] and [25]. In what follows of this work, we are going to introduce a novel augmented Lagrangian algorithm in continuous optimization.

A Novel Algorithm

In [38] the hyperbolic penalty algorithm (HPA) is proposed, later this algorithm is studied in [39], [40] and [42]. Some works where good computational results of HPA can be observed in [43] and [37]. HPA motivates the development of the hyperbolic smoothing method (HSM), to see details of this connection see [37]. The HSM shows a good computational performance in solving different nondifferentiable problems from mathematical optimization, see: [44], [45], [35], [46], [47] and [48].

The HPA also induces ideas to build a new algorithm of augmented Lagrangian type, called HALA-1992 (see [41]). The characteristic of HALA-1992 is that it considers the updating of the penalty parameter and consider conditions on the dual function, so then the first ideas of convergence are proposed in [41]. Now on this occasion, unlike HALA-1992, we consider the fixed penalty parameter and without the need to consider conditions on the dual function. Therefore we propose in this work a new algorithm, which henceforth we will call HALA.

The main contribution of our work is to have guaranteed the convergence of HALA and we also ensure existence of solutions for the subproblem generated by HALA, considering usual assumptions of the literature. In order for us to guarantee the convergence of the algorithm proposed in this work, we use the classic assumptions, such that the Slater constraint qualification and convexity, some works that consider these assumptions are [1], [2] and [25].

The paper is organized as follows: In Chapter 2 we present some basic results, we also present HPA and some of its properties. In Chapter 3 we present the hyperbolic augmented Lagrangian function and HALA. We study some characteristics of this algorithm. In Chapter 4 we guarantee the convergence of the HALA. In Chapter 5 computational results are illustrated. In Chapter

6 we give some conclusions of our work. In Chapter 7 we propose some future work.

2 Preliminaries

Throughout this paper 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 that $f, \; g_i$ are continuously differentiable. That way (P) is a convex optimization problem. So (P) will be called as the primal problem. We consider the following assumptions.

C1. The optimal set X^* is nonempty, closed, bounded and, consequently, compact.

C2. Slater constraint qualification holds, i.e., there exists $\hat{x} \in S$ which satisfies $g_i(\hat{x}) > 0, \; i = 1, ..., m$.

A consequence of **C1** (see the Theorem 24 and Corollary 20 of [10]) is that the level set $\{x \in S \mid f(x) \leq \beta\}$ remains bounded for any value β. The **C2** assumption guarantees that the interior of S set is nonempty. The condition **C1** also imply the existence of a finite vector x^* and a number f^* such that $f(x^*) = f^* = \inf_S f(x) = \min_S f(x)$.

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

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x), \tag{2.2}$$

where, $\lambda_i \geq 0,\ i = 1, ..., m$, are called dual variables or Lagrange multipliers. Since the problem (P) is convex, we know that due to assumption **C2**, the following results will occur: there exists $\lambda^* = (\lambda_1^*, ..., \lambda_m^*)$, such that, the Karush-Kuhn-Tucker (KKT) conditions hold true, i.e.,

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0, \tag{2.3}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, ..., m, \tag{2.4}$$

$$g_i(x^*) \geq 0, \quad i = 1, ..., m, \tag{2.5}$$

$$\lambda_i^* \geq 0, \quad i = 1, ..., m. \tag{2.6}$$

Moreover, the set of optimal Lagrange multipliers λ^* is denoted by

$$\Lambda^* = \left\{ \lambda \in I\!R^m_+ \mid \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0,\ x^* \in X^* \right\},$$

it is known that Λ^* is a bounded set (and hence compact set) due to **C2**. The dual function $\Phi : I\!R^m_+ \to I\!R$, is defined as follows

$$\Phi(\lambda) = \inf_{x \in I\!R^n} L(x, \lambda), \tag{2.7}$$

and the dual problem consists of finding

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

2.1 **Hyperbolic Penalty**

The hyperbolic penalty is meant to solve the problem (P). The penalty method adopts the hyperbolic penalty function (HPF)

$$P(y, \lambda, \tau) = -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}, \tag{2.8}$$

where $P : (-\infty, +\infty) \times I\!R_+ \times I\!R_{++} \to I\!R$. Notice that, $P(y, \lambda, \tau) > 0$.

Remark 2.1 *The HPF is originally proposed in [38] and studied in [42]. 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)
- $P(y, \lambda, 0) = 0, \quad$ *for* $\; y \geq 0$.
- $P(y, \lambda, 0) = -2\lambda y, \quad$ *for* $\; y < 0$.

Due to the properties (a) *and* (b) *the HPF is equivalent to a smoothing of the penalty studied by Zangwill, in [49].*

Let us note the following properties of the function P (which are also studied in [38]):

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

P1) $P(y, \lambda, \tau)$ is convex function of y, i.e.,

$$\nabla^2_{yy} P(y, \lambda, \tau) = \frac{\lambda^2 \tau^2}{((\lambda y)^2 + \tau^2)^{\frac{3}{2}}} > 0.$$

3 Hyperbolic Augmented Lagrangian

We define the Hyperbolic Augmented Lagrangian Function (HALF) of problem (P) by $L_H : \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}_{++} \to \mathbb{R}$,

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

$$= f(x) + \sum_{i=1}^{m} \left(-\lambda_i g_i(x) + \sqrt{(\lambda_i g_i(x))^2 + \tau^2} \right), \quad (3.9)$$

where $\tau > 0$ is the penalty parameter. Note that this function belongs to class C^∞ if the involved functions $f(x)$ and $g_i(x)$, $i = 1, ..., m$, are too. On the other hand, a variation of (3.9) is proposed and studied in the work of [7] and [28].

By comparing (2.2) and (3.9), we see that the function L_H may be put in the form

$$L_H(x, \lambda, \tau) = L(x, \lambda) + \sum_{i=1}^{m} \sqrt{(\lambda_i g_i(x))^2 + \tau^2}. \quad (3.10)$$

Analysis of expression (3.10) allows us to see that the modified objective function associated with the hyperbolic penalty may be decomposed as the sum of the Lagrangian function along with a summation of terms containing squares of the products between the values of the constraints and their corresponding multipliers (complementary slacks). We are aware that at any optimal point (x^*, λ^*) we must have $\lambda_i^* g_i(x^*) = 0,\ i = 1, ..., m$, and therefore at this point the summation takes on a minimum value equal to $\sum_{i=1}^{m} \tau = m\tau$. From this point of view the summation in expression (3.10) may be interpreted as a penalty for the noncompliance with the condition of complementarity of the slacks which is added to the Lagrangian function. In the composition of the modified objective function, when we attempt to minimize this portion, we will automatically be seeking the optimal solution where equalities $\lambda_i^* g_i(x^*) = 0,\ i = 1, ..., m$ prevail.

C3. For every $\tau > 0$ and $\lambda > 0$, the level set

$$M = \{x \in I\!R^n \ \mid \ L_H(x, \lambda, \tau) \leq \alpha\},$$

is bounded, for every $\alpha < \infty$.

Remark 3.1 *We know that the function P is convex by P1). When f is strongly convex, we get that L_H is strongly convex. Then, we can ensure that the assumption* **C3** *is verified. Other authors who study the assumption of strong convexity are: Auslender et al. ([1]), Kiwiel ([18]), Rockafellar ([31]), Kort and Bertsekas ([21]), Kort and Bertsekas ([20]), Sabach and Teboulle ([33]) and Silva et al. ([34]).*

Now we present the HALA to solve the nonlinear problem (P).

3.1 **Algorithm**

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):

$$x^{k+1} \in argmin_{x\in\mathbb{R}^n}\ L_H(x, \lambda^k, \tau)$$

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

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

$$\lambda_i^{k+1} = \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{(\lambda_i^k g_i(x^{k+1}))^2 + \tau^2}} \right), \quad i = 1, ..., m. \tag{3.11}$$

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.

The HALA considers an initial vector $\lambda^0 > 0$ and $\tau > 0$ which is fixed. Considering a fixed penalty parameter can also be observed in the works of [15], [19], [20] and [30]. With this information, the HALA generate the primal sequence in **Step 2** and the multiplier estimates in **Step 3**. In **Step 4**, we can consider different stopping criteria. For example, we can consider some of the following criteria studied in [6]:

$$-\min_{i=1,\dots,m} g_i(x^k) < \beta \quad \textit{and} \quad \frac{\left|f(x^k) - f(x^{k-1})\right|}{1 + |f(x^{k-1})|} < 10^{-2}\beta,$$

or

$$\max \left\{ -\min_{i=1,\dots,m} g_i(x^k), \frac{\sum_{i=1}^m \lambda_i^k \left|g_i(x^k)\right|}{1 + \|x^k\|_2}, \frac{\left\|\nabla f(x^k) - \sum_{i=1}^m \lambda_i^k \nabla g_i(x^k)\right\|_\infty}{1 + \|x^k\|_2} \right\} < \beta,$$

where $\beta > 0$.

Notice that HALA is based in the exact unconstrained minimization of the HALF. In [21] an exact unconstrained minimization of the augmented Lagrangian is also discussed, also see [2].

3.2 **Study of the HALA**

By **C3** hence there exists $x^{k+1} \in I\!R^n$ such that

$$L_H(x^{k+1}, \lambda^k, \tau) = min_{x \in I\!R^n} L_H(x, \lambda^k, \tau),$$

thus $\nabla_x L_H(x^{k+1}, \lambda^k, \tau) = 0$ holds, i.e.,

$$\nabla f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{(\lambda_i^k g_i(x^{k+1}))^2 + \tau^2}} \right) \nabla g_i(x^{k+1}) = 0, \tag{3.12}$$

substituting (3.11) in (3.12), we have

$$\nabla_x L_H(x^{k+1}, \lambda^k, \tau) = \nabla f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i(x^{k+1}) = \nabla_x L(x^{k+1}, \lambda^{k+1}) = 0, \tag{3.13}$$

for any $\tau > 0$. We observe that x^{k+1} and λ^{k+1} satisfy $\nabla_x L(x^{k+1}, \lambda^{k+1}) = 0$, shows that x^{k+1} is the minimizer of $L(x, \lambda^{k+1})$ (i.e., x^{k+1} attains the minimum in (2.7)), i.e.,

$$\Phi(\lambda^{k+1}) = L(x^{k+1}, \lambda^{k+1}) = \min_{x \in I\!R^n} L(x, \lambda^{k+1}) \quad and \quad \lambda^{k+1} \in I\!R^m_{++},$$

thus, it follows that

$$\Phi(\lambda^{k+1}) = f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} g_i(x^{k+1}). \tag{3.14}$$

From (3.14) we obtain

$$-g(x^{k+1}) = \left(-g_1(x^{k+1}), \cdots, -g_m(x^{k+1})\right)^T \in \partial\Phi(\lambda^{k+1}),$$

where $\partial\Phi(\lambda^{k+1})$ is the subdifferential of $\Phi(\lambda)$ at $\lambda = \lambda^{k+1}$. In the following remark, we analyze what happens with Lagrange multipliers (iteration (3.11)) depending on the type of restriction we have. First, for $x \in I\!R^n$, we define the following sets of indices

$$I_0 = \{i \in \{1, ..., m\} \mid g_i(x) = 0\},$$

$$I_- = \{i \in \{1, ..., m\} \mid g_i(x) < 0\},$$

$$I_+ = \{i \in \{1, ..., m\} \mid g_i(x) > 0\},$$

such that $I_0 \cap I_+ = \emptyset$, $I_0 \cap I_- = \emptyset$, $I_+ \cap I_- = \emptyset$ and $I_0 \cup I_+ \cup I_- = \{1, ..., m\}$.

Remark 3.2 *Let $\{\lambda^k\}$ be a sequence generated by HALA such that $\lambda_i^k > 0$, $i = 1, ..., m$ and let $\tau > 0$ fixed. Let us consider the following cases:*

(c1) If $i \in I_0$, then we have at the k-th iteration $g_i(x^{k+1}) = 0$, then by (3.11), we get, $\lambda_i^{k+1} = \lambda_i^k$. We also obtain:

$$\left(\lambda_i^k - \lambda_i^{k+1}\right) g_i(x^{k+1}) = 0, \ \forall i \in I_0.$$

(c2) If $i \in I_+$, then we have at the k-th iteration $g_i(x^{k+1}) > 0$, then by (3.11), we get, $\lambda_i^k > \lambda_i^{k+1}$. We also obtain:

$$\left(\lambda_i^k - \lambda_i^{k+1}\right) g_i(x^{k+1}) > 0, \ \forall i \in I_+.$$

(c3) If $i \in I_-$, then we have at the k-th iteration $g_i(x^{k+1}) < 0$, then by (3.11), we get, $\lambda_i^k < \lambda_i^{k+1}$. We also obtain:

$$\left(\lambda_i^k - \lambda_i^{k+1}\right) g_i(x^{k+1}) > 0, \ \ \forall i \in I_-.$$

Of the three previous cases, we can note that we have the following

$$\left(\lambda_i^k - \lambda_i^{k+1}\right) g_i(x^{k+1}) \geq 0, \quad i = 1, ..., m.$$

Now, we will demonstrate the positivity of the updated Lagrange multipliers.

Proposition 3.1 *Let* $\left\{\lambda^k = (\lambda_1^k, ..., \lambda_m^k) \mid k = 1, 2, ...\right\} \subset I\!R^m$. *If*

$$\lambda^k \in I\!R_{++}^m \qquad then \qquad \lambda^{k+1} \in I\!R_{++}^m, \ \ i = 1, ..., m.$$

Proof. Let $\tau > 0$ be fixed. Since we have $0 < \tau^2$, we can obtain the following

$$\left(\lambda_i^k g_i(x^{k+1})\right)^2 < \left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2, \ \ i = 1, ..., m,$$

from this, we can get

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

from the latter it follows that

$$0 < \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}}\right) < 2\lambda_i^k, \ \ i = 1, ..., m, \tag{3.15}$$

then from the expression above and by (3.11), we get that, $\lambda_i^{k+1} > 0, \ i = 1, ..., m$. ∎

Remark 3.3 *From inequality (3.15), we can see that iteration (3.11) has the following characteristic*

$$0 < \lambda_i^{k+1} < 2\lambda_i^k, \ \ i = 1, ..., m. \tag{3.16}$$

Remark 3.4 *From* **C3** *and Proposition 3.1 we obtain that HALA is well defined.*

Theorem 3.1 *Let* $\{\lambda^k\}$ *be a sequence generated by HALA. The sequence* $\{\Phi(\lambda^k)\}$ *is monotone nondecreasing for all* $k \in I\!N$.

Proof. From the concavity of $\Phi(\cdot)$ and since $-g(x^{k+1}) \in \partial\Phi(\lambda^{k+1})$, we obtain

$$\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \sum_{i=1}^{m} \left(g_i(x^{k+1})\right)\left(\lambda_i^k - \lambda_i^{k+1}\right). \tag{3.17}$$

On the other hand, we can rewrite (3.11), as follows,

$$\lambda_i^k - \lambda_i^{k+1} = \frac{\left(\lambda_i^k\right)^2 g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}}, \quad i = 1, ..., m, \tag{3.18}$$

this expression (3.18), is replaced on the right side of inequality (3.17), we get

$$\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \sum_{i=1}^{m} \left(\frac{\left(\lambda_i^k g_i(x^{k+1})\right)^2}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}} \right) \geq 0, \tag{3.19}$$

so, we have, $\Phi(\lambda^{k+1}) \geq \Phi(\lambda^k)$. ■

Proposition 3.2 *The sequence of dual objective function values* $\{\Phi(\lambda^k)\}$ *is bounded and monotone nondecreasing, hence it converges.*

Proof. By Theorem 3.1 we obtain $\Phi(\lambda^{k+1}) \geq \Phi(\lambda^k)$, then $\{\Phi(\lambda^k)\}$ is nondecreasing sequence for all $k \in I\!N$ and considering the weak duality theorem, we obtain $\Phi(\lambda^k) \leq \Phi(\lambda^{k+1}) \leq f^*$, $\forall k$, i.e., $\{\Phi(\lambda^k)\}$ is bounded from above by the optimal value. Then $\{\Phi(\lambda^k)\}$ is convergent. ■

Proposition 3.3 *The sequence* $\{\lambda^k\}$ *generated by the HALA is bounded.*

Proof. From **C2** we know that Λ^* is nonempty and compact. So, one level set of $\Phi(\cdot)$ is compact. Then, all of these level sets are compact, see Corollary 8.7.1 of [29]. By Proposition 3.2 we obtain in particular $\lambda^k \in \Gamma = \{\lambda \in$

$\mathbb{R}^m_+ \mid \Phi(\lambda^0) \leq \Phi(\lambda)\}$ for all $k \in \mathbb{N}$ and hence $\{\lambda^k\}$ is a bounded sequence. ■

We present a preliminary result which will be used to guarantee the complementarity condition in our algorithm.

Lemma 3.1 *Let $d > 0$ and a sequence $\{a^k\} \subset \mathbb{R}_+$. If*

$$\lim_{k\to\infty} \left(a^k/\sqrt{a^k+d}\right) = 0 \qquad \textit{then} \qquad \lim_{k\to\infty} a^k = 0.$$

Proof. Let us fix $\epsilon \in (0,1)$. By hypothesis, there exists $k_0 \in \mathbb{N}$ such that

$$\frac{a^k}{2\sqrt{a^k+d}} < \epsilon, \ \ \forall k \geq k_0. \tag{3.20}$$

On the other hand, we know that $\left(\sqrt{a^k+d}-1\right)^2 \geq 0$, then, $a^k+d+1 \geq 2\sqrt{a^k+d}$, from (3.20) and from the previous inequality, we obtain

$$\frac{a^k}{a^k+d+1} \leq \frac{a^k}{2\sqrt{a^k+d}} < \epsilon, \ \forall k \geq k_0, \tag{3.21}$$

then of (3.21), we get $a^k \leq (\epsilon(1+d)/(1-\epsilon))$, $\forall k \geq k_0$, which implies, $\lim_{k\to\infty} a^k = 0$. ■

Theorem 3.2 *Let the sequences $\{x^k\}$ and $\{\lambda^k\}$ be generated by HALA. Then*

$$\lim_{k\to\infty} \left(\lambda_i^k g_i(x^k)\right) = 0, \quad i = 1, ..., m.$$

Proof. Let be $\tau > 0$ fixed. Since $\Phi(\cdot)$ is concave we have the expression (3.19).

We are going to verify that the series in (3.19) is convergent; (3.19) gives by summation

$$0 \leq \sum_{k=1}^{\infty}\sum_{i=1}^{m}\left(\frac{\left(\lambda_i^k g_i(x^{k+1})\right)^2}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2+\tau^2}}\right) \leq \sum_{k=1}^{\infty}\left(\Phi(\lambda^{k+1})-\Phi(\lambda^k)\right),$$

we notice that $\sum_{k=1}^{\infty}\left(\Phi(\lambda^{k+1})-\Phi(\lambda^k)\right)$ is a convergent series (i.e., the partial sum is bounded above), it follows

$$\sum_{k=1}^{\infty}\sum_{i=1}^{m}\left(\frac{\left(\lambda_i^k g_i(x^{k+1})\right)^2}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2+\tau^2}}\right)\leq \lim_{k\to\infty}\left(\Phi(\lambda^k)-\Phi(\lambda^1)\right)\leq f^*-\Phi(\lambda^1)<\infty,$$

therefore, for the test of comparison, we obtain

$$\lim_{k\to\infty}\sum_{i=1}^{m}\left(\frac{\left(\lambda_i^k g_i(x^{k+1})\right)^2}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2+\tau^2}}\right)=0. \tag{3.22}$$

We note each term in the summation of (3.22) is nonnegative, thus

$$\lim_{k\to\infty}\left(\frac{\left(\lambda_i^k g_i(x^{k+1})\right)^2}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2+\tau^2}}\right)=0,\ i=1,...,m, \tag{3.23}$$

in (3.23), we can apply the Lemma 3.1 with $a^k=\left(\lambda_i^k g_i(x^{k+1})\right)^2$ and $d=\tau^2$ and thus we obtain $\lim_{k\to\infty}\left(\lambda_i^k g_i(x^{k+1})\right)^2=0,\ i=1,...,m$, so,

$$\lim_{k\to\infty}\left(\lambda_i^k g_i(x^{k+1})\right)=0,\ i=1,...,m. \tag{3.24}$$

Because $\Phi(\cdot)$ is a concave function and by Remark 3.2 we get

$$\Phi(\lambda^{k+1})-\Phi(\lambda^k)\geq\sum_{i=1}^{m}\left(g_i(x^{k+1})\right)\left(\lambda_i^k-\lambda_i^{k+1}\right)\geq 0, \tag{3.25}$$

and by Proposition 3.2 we know that $\{\Phi(\lambda^k)\}$ is convergent, so, it follows $\lim_{k\to\infty}\left\{\Phi(\lambda^{k+1})-\Phi(\lambda^k)\right\}=0$, and so from (3.25) we obtain

$$\lim_{k\to\infty}\sum_{i=1}^{m}\left(g_i(x^{k+1})\right)\left(\lambda_i^k-\lambda_i^{k+1}\right)=0, \tag{3.26}$$

now since $\left(g_i(x^{k+1})\right)\left(\lambda_i^k-\lambda_i^{k+1}\right)\geq 0$, of (3.26) and (3.24), it follows that

$$\lim_{k\to\infty}\left(\lambda_i^{k+1} g_i(x^{k+1})\right)=0,\ i=1,...,m. \tag{3.27}$$

■

4 Convergence Result

In this section, we are going to consider the following assumption.

C4. The whole sequence $\left\{x^k\right\}$ is convergent to $\bar{x}$, where $\bar{x}$ is assumed a feasible point, i.e., $g_i(\bar{x}) \geq 0, \ i = 1, ..., m$.

Similar to assumption **C4** can also be seen in [11], [23], [4] and [9]. Finally, we ensure that the subsequence generated by the algorithm HALA converges to a KKT point.

Theorem 4.1 *The convex problem (P) satisfies* **C1***,* **C2***,* **C3** *and* **C4***. Let sequences* $\{x^k\}$ *and* $\{\lambda^k\}$ *generated by HALA. Then any limit point of a subsequence* $\{x^k\}$ *and* $\{\lambda^k\}$ *are an optimal solution-Lagrange multiplier pair for the problem (P).*

Proof. Let be $\tau > 0$ fixed. By **C3** follows the boundedness of the sequence $\{x^k\}$, and also we know that the sequence $\{\lambda^k\}$ of the Lagrange multipliers generated by the HALA is bounded, see Proposition 3.3. So, there are limit points $\bar{x}$ and $\bar{\lambda}$. Henceforth, we can consider the following convergent subsequences $\lim_{k\to\infty} x^k = \bar{x}$ and $\lim_{k\to\infty} \lambda^k = \bar{\lambda}$ with $k \in K_1 \subset I\!N$.

Now by **C4**, we have $\lim_{k\to\infty} g_i(x^k) = g_i(\bar{x}) \geq 0, \ i = 1, ..., m$. From Proposition 3.1 we obtain,

$$\lim_{k\to\infty} \lambda_i^k = \bar{\lambda}_i \geq 0, \ \ i = 1, ..., m. \tag{4.28}$$

Passsing the limit in (3.27), we have

$$\lim_{k\to\infty}\left(\lambda_i^k g_i(x^k)\right) = \bar{\lambda}_i g_i(\bar{x}) = 0, \quad \forall i = 1, ..., m. \tag{4.29}$$

Moreover, passing the limit in (3.13), we obtain

$$\nabla_x L(\bar{x}, \bar{\lambda}) = \nabla f(\bar{x}) - \sum_{i=1}^{m} \bar{\lambda}_i \nabla g_i(\bar{x}) = 0.$$

Thus $(\bar{x}, \bar{\lambda})$ satisfies (2.3) – (2.6) for all $i = 1, ..., m$, hence $(\bar{x}, \bar{\lambda})$ is a KKT point. Thus $\bar{x}$ is optimal for the problem (P) and $\bar{\lambda}$ is a Lagrange multiplier.

■

5 Computational Illustration

The computer illustrations presented below were obtained with a preliminary Fortran implementation for HALA. 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 [13]. The algorithm stop when the solution is viable (feasible) an the absolute value of the difference of the two consecutives solutions $\left|x^k - x^{k-1}\right|$ is less than 10^{-5}.

5.1 **Test Problem**

Problem 5.1

$$\min_{x \in I\!R^n} f(x) = x^T A x + x^T b$$
$$s.t.\ \ 10 \le x_i \le 100,$$

where $x,\ b \in I\!R^n$, *A is a symmetric matrix $n \times n$ and b is a vector with all its variables equal to 10.*

The construction rule for matrix A is as follows: $a_{i,i} = 1 + \sqrt{i}$ and $a_{i,j} = \frac{a_{i,i}+a_{j,j}}{n(i+j)}$. In particular we will only show details for case $n = 2$, where matrix A is symmetric and positive definite.

Problem 5.2

$$\min_{x \in I\!R^2} f(x) = x^T \begin{pmatrix} 2 & \frac{3+\sqrt{2}}{6} \\ \frac{3+\sqrt{2}}{6} & 1+\sqrt{2} \end{pmatrix} x + x^T \begin{pmatrix} 10 \\ 10 \end{pmatrix}$$
$$s.t.\ \ 10 \le x_1 \le 100,$$
$$10 \le x_2 \le 100.$$

We are going to rewrite the constraints above as:

$$g_1(x) = 100 - x_1 \ge 0,$$
$$g_2(x) = x_1 - 10 \ge 0,$$
$$g_3(x) = 100 - x_2 \ge 0,$$
$$g_4(x) = x_2 - 10 \ge 0.$$

5.2 **Results**

The Tables 1-2 summarize the computational results for the problem 5.2. For each table, k is the number of iterations, it is the total number of iterations, λ is the multiplier Lagrange, x is the primal variable, $f(x)$ is the objective value, $g_i(x)$ are the constraints of each problem, $L_H(x, \lambda, \tau)$ 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 "$1 = not$".

In the Table 1, we reports the optimal solutions, the value of the objective function and the value of HALF found by our proposed algorithm. In the Table 2, we reports the behavior of the multipliers, this issue is studied in Subsection 3.2 of this work. For the Problem 5.2 our algorithm converges to the exact solution within the precision of the computer.

Finally in Table 3, we solve problem 5.1 for cases $n = 2, 50, 100, 150, 200$. In this table we show the time used to solve each case and the number of iterations. In all cases we use the value of $\tau = 0.10E-02$. We only assure that for the case $n = 2$ the matrix A is positive definite.

Table 1 Problem 5.2 with $\tau = 0.10E - 02$

k	x_1	x_2	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	0.500000000E+02	0.500000000E+02	0.157140452E+05	0.157140452E+05	0
1	0.195759857E+01	0.147451342E+01	0.514816799E+02	0.382839440E+03	1
2	0.587279571E+01	0.442354027E+01	0.257408400E+03	0.645556960E+03	1
3	0.999998038E+01	0.999996351E+01	0.788557875E+03	0.788565160E+03	1
4	0.100000000E+02	0.100000000E+02	0.788561812E+03	0.788565808E+03	0
5	0.999999994E+01	0.999999997E+01	0.788561802E+03	0.788565808E+03	0
6	0.100000000E+02	0.100000000E+02	0.788561814E+03	0.788565808E+03	0
7	0.999999998E+01	0.999999994E+01	0.788561803E+03	0.788565808E+03	0
8	0.100000000E+02	0.100000000E+02	0.788561812E+03	0.788565808E+03	0
9	0.100000000E+02	0.999999994E+01	0.788561804E+03	0.788565808E+03	0
10	0.100000000E+02	0.100000000E+02	0.788561812E+03	0.788565808E+03	0
11	0.100000000E+02	0.100000000E+02	0.788561812E+03	0.788565808E+03	0

6 Conclusions

- In this work, we mainly introduce new algorithms of the augmented Lagrangian type in the area of continuous optimization.
- The results presented in this work provide the necessary theoretical framework for the construction of a new algorithm to which we give the name Hyperbolic Augmented Lagrangian Algorithm. The convergence of the algorithm proposed was also demonstrated.
- The HPF belongs to class C^∞. Hence, $L_H(x, \lambda, \tau)$ will be class C^∞ if the involved functions $f(x)$ and $g_i(x)$, $i = 1, ..., m$, are too. This is an outstanding property from the computational point of view.

Table 2 Problem 5.2, with $\tau = 0.10E - 02$

	$g_1(x)$		$g_2(x)$		$g_3(x)$		$g_4(x)$	
k	via	λ_6	via	λ_7	via	λ_8	via	λ_9
0	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02	0	0.100000000E+02
1	1	0.200000000E+02	1	0.200000000E+02	0	0.520139487E-11	0	0.515143483E-11
2	1	0.400000000E+02	1	0.400000000E+02	0	0.520139232E-11	0	0.515143230E-11
3	1	0.646951555E+02	1	0.729970805E+02	0	0.520138989E-11	0	0.515142991E-11
4	0	0.645434539E+02	0	0.729199045E+02	0	0.520138745E-11	0	0.515142752E-11
5	0	0.648090459E+02	0	0.730998835E+02	0	0.520138502E-11	0	0.515142513E-11
6	0	0.646568096E+02	0	0.728647417E+02	0	0.520138258E-11	0	0.515142274E-11
7	0	0.647429498E+02	0	0.731620337E+02	0	0.520138015E-11	0	0.515142036E-11
8	0	0.647250580E+02	0	0.729264919E+02	0	0.520137771E-11	0	0.515141797E-11
9	0	0.647165850E+02	0	0.732242880E+02	0	0.520137528E-11	0	0.515141558E-11
10	0	0.647128348E+02	0	0.729883452E+02	0	0.520137284E-11	0	0.515141319E-11
11	0	0.647090851E+02	0	0.727539205E+02	0	0.520137041E-11	0	0.515141080E-11

- The smooth behavior of the modified objective function offers the possibility to use the best unconstrained minimization techniques, which use second-order derivatives.

Table 3 Problem 5.1

n	time	it
2	0.001159	11
50	0.122072	14
100	0.514586	15
150	1.350671	12
200	2.379733	6

7 Future Work

⋆ Although important theoretical points have been developed, we are far from having exhausted our studies. In fact, the connections between hyperbolic penalty and the Lagrangian function extend even further the horizons of new theoretical lines and practical experimentation to be researched.

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.

References

1. Auslender, A., Teboulle, M., Ben-Tiba, S.: Interior proximal and multiplier methods based on second order homogeneous kernels. Mathematics of Operations Research. 24, 645-668 (1999)
2. Ben-Tal, A., Zibulevsky, M.: Penalty/Barrier multiplier methods for convex programming problems. SIAM J. OPTIM. 7, 347-366 (1997)
3. Bertsekas, D.P.: Multiplier methods: a survey. Automatica. 12, 133-145 (1976)

4. Birgin, E.G., Fernández, D., Martínez, J.M.: The boundedness of penalty parameters in an augmented Lagrangian method with constrained subproblems. Optimization Methods and Software. 27, 1001-1024 (2012)
5. Boukari, D., Fiacco, A.V.: Survey of penalty, exact-penalty and multiplier methods from 1968 to 1993. Optimization. 32, 301-334 (1995)
6. Breitfeld, M.G., Shanno, D.F.: Computational experience with penalty-barrier methods for nonlinear programming. Annals of Operations Research. 62, 439-463 (1996)
7. Costa, M.F.P., Rocha, A.M.A.C., Fernandes, E.M.G.P.: An artificial fish swarm algorithm based hyperbolic augmented Lagrangian method. Journal of Computational and Applied Mathematics. 259, 868-876 (2014)
8. Di Pillo, G., Lucidi, S.: An augmented Lagrangian function with improved exactness properties. SIAM J. Optim. 12, 376-406 (2001)
9. Echebest, N., Sánchez, M.D., Schuverdt, M.L.: Convergence results of an augmented Lagrangian method using the exponential penalty function. J. Optim. Theory Appl. 168, 92-108 (2016)
10. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. SIAM Classic in Applied Mathematics, Philadelphia (1990)
11. Hartman, J.K.: Iterative determination of parameters for an exact penalty function. J. Optim. Theory Appl. 16, 49-66 (1975)
12. Hestenes, M.R.: Multiplier and gradient methods. JOTA. 4, 303-320 (1969)
13. Harwell Subroutine Library: A collection of Fortran Codes for Large Scale Scientific Computation, $url = https://www.hsl.rl.ac.uk/$ (2002)
14. Iusem, A.N., Svaiter, B.F., Teboulle, M.: Entropy-like proximal methods in convex programming. Mathematics of Operations Research. 19, 790-814 (1994)
15. Jensen, D.L., Polyak, R.A.: The convergence of a modified barrier method for convex programming. IBM Journal of Research and Development. 38, 307-321 (1994)
16. Kelley, J.E.: The cutting-plane method for solving convex programs. J. Soc. Indust. Appl. Math. 8, 703-712 (1960).
17. Kiwiel, K.C.: On the twice differentiable cubic augmented Lagrangian. Journal of Optimization Theory and Applications. 88, 233-236 (1996).

18. Kiwiel, K.C.: Proximal minimization methods with generalized bregman function. SIAM J. Control Optim. 35, 1142-1168 (1997)

19. Kort, B.W., Bertsekas, D.P.: A new penalty function method for constrained minimization. Proceedings 1972 IEEE Conf. Decision and Control. 162-166 (1972)

20. Kort, B.W., Bertsekas, D.P.: Multiplier method for convex programming. IEEE Conference on Decision and Control Including the 12th Symposium on Adaptive Processes. 428-432 (1973)

21. Kort, B.W., Bertsekas, D.P.: Combined primal-dual and penalty methods for convex programming. SIAM J. Control and Optimization. 14, 268-294 (1976)

22. Minoux, M.: Mathematical Programming: Theory and Algorithms. Chichester: Wile, New York (1986)

23. Nguyen, V.H., Strodiot, J.J.: On the convergence rate for a penalty function method of exponential type. JOTA. 27, 495-508 (1979)

24. Polyak, R.A.: Modified barrier functions: theory and methods. Mathematical Programming. 54, 177-222 (1992)

25. Polyak, R., Teboulle, M.: Nonlinear rescaling and proximal-like methods in convex optimization. Mathematical Programming. 76, 265–284 (1997)

26. Polyak, R.A.: Log-sigmoid multipliers method in constrained optimization. Annals of Operations Research. 101, 427-460 (2001)

27. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. Optimization, *Ed. R. Fletcher*, 283–298 (1969)

28. Rocha, A.M.A.C., Costa, M.F.P., Fernandes, E.M.G.P.: A shifted hyperbolic augmented Lagrangian-based artificial fish two-swarm algorithm with guaranteed convergence for constrained global optimization. Engineering Optimization. 48, 2114–2140 (2016).

29. Rockafellar, R.T.: Convex analysis. Princeton University Press, Princeton, NJ. (1970)

30. Rockafellar, R.T.: The multiplier method of Hestenes and Powell applied to convex programming. JOTA. 12, 555-562 (1973)

31. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research. 1, 97-116 (1976)

32. Rosen, J.B.: The Gradient Projection Method for Nonlinear Programming. Part I. Linear Constraints. Journal of the Society for Industrial and Applied Mathematics. 8, 181-217 (1960)

33. Sabach, S., Teboulle, M.: Faster Lagrangian-based methods in convex optimization. arXiv:2010.14314 (2020)

34. Silva, P.J.S., Eckstein, J., Humes, C.: Rescaling and stepsize selection in proximal methods using separable generalized distances. SIAM J. Optim. 12, 238-261 (2001)

35. Souza, M., Xavier, A.E., Lavor, C., Maculan, N.: Hyperbolic smoothing and penalty techniques applied to molecular structure determination. Operations Research Letters. 39, 461-465 (2011)

36. Tseng, P., Bertsekas, D.P.: On the convergence of the exponential multiplier method for convex programming. Mathematical Programming. 60, 1-19 (1993)

37. Venceslau, H.M., Lubke, D.C., Xavier, A.E: Optimal covering of solid bodies by spheres via the hyperbolic smoothing technique. Optim. Methods Softw. 30, 391-403 (2015)

38. Xavier, A.E.: *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 (1982a)

39. Xavier, A.E.: *Penalização Hiperbólica.* Anais do I Congresso Latino Americano de Pesquisa Operacional e Engenharia de Sistemas. Rio de Janeiro. 468-482 (1982b)

40. Xavier, A.E., Maculan, N.: *Extrapolação em Penalização Hiperbólica*, II Congreso Latino Americano de Investigación Operativa e Ingenieria de Sistemas (trabalhos seleccionados). Buenos Aires. 24-38 (1984)

41. Xavier, A.E.: *Penalização Hiperbólica* (Ph.D. thesis). Federal University of Rio de Janeiro \ COPPE, Rio de Janeiro, Brazil (1992)

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

43. Xavier, A. E., Oliveira, A.A.F.D.: Optimal covering of plane domains by circles via hyperbolic smoothing. Journal of Global Optimization. 31, 493-504 (2005)

44. Xavier, A. E.: The hyperbolic smoothing clustering method. Pattern Recognition. 43, 731-737 (2010)

45. Xavier, A.E., Xavier, V.L.: Solving the minimum sum-of-squares clustering problem by hyperbolic smoothing and partition into boundary and gravitational regions. Pattern Recognition. 44, 70-77 (2011)
46. Xavier, V.L., França, F.M.G., Xavier, A.E., Lima, P.M.V.: A hyperbolic smoothing approach to the Multisource Weber problem. Journal of Global Optimization. 60, 49-58 (2014)
47. Xavier, A.E., Gesteira, C.M., Xavier, V.L.: Solving the continuous multiple allocation p-hub median problem by the hyperbolic smoothing approach. Optimization. 64, 2631-2647 (2015)
48. Yilmaz, N., Sahiner, A.: Generalization of hyperbolic smoothing approach for non-smooth and non-Lipschitz functions. Journal of Industrial and Management Optimization. doi: 10.3934/jimo.2021170 (2020)
49. Zangwill, W.I.: Non-linear programming via penalty function. Management Science. 13, 344-358 (1967)
50. Zoutendijk, G.: Some algorithms based on the principle of feasible directions. Nonlinear Programming, Eds. J.B. Rosen, O.L Mangasarian and K. Ritter (Academic Press, New York - London, 1970. 93-122 (1970)