Convergence Analysis of the Hyperbolic Augmented Lagrangian Algorithm

Lennin Mallma Ramirez · Nelson Maculan · Adilson Elias Xavier

May 08, 2021

Lennin Mallma Ramirez, Corresponding author
Federal University of Rio de Janeiro, Systems Engineering and Computer Science Program (PESC), Graduate School of Engineering (COPPE)
Rio de Janeiro, Brazil
lenninmr@gmail.com

Nelson Maculan
Federal University of Rio de Janeiro, Systems Engineering and Computer Science Program (PESC), Graduate School of Engineering (COPPE)
Rio de Janeiro, Brazil
maculan@cos.ufrj.br

Adilson Elias Xavier
Federal University of Rio de Janeiro, Systems Engineering and Computer Science Program (PESC), Graduate School of Engineering (COPPE)
Rio de Janeiro, Brazil
adilson.xavier@gmail.com
Abstract A new approach -Hyperbolic Augmented Lagrangian Algorithm (HALA)- for solving nonlinear programming problem is presented. Under mild assumptions, such as: convexity, Slater’s qualification and differentiability, the convergence of the proposed algorithm is proved.

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

Nonlinear programming draws attention from the scientific community for the large number and variety of problems that are modeled with constraints and solved in different areas of research, see the time-dependent traveling salesman problem in [22], the optimal reactive dispatch problem in [12] and the Euclidean Steiner tree problem in [23]. Classic books in nonlinear programming are [5] and [24].

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

$$\min \left\{ f(x) \mid x \in S \right\}, \quad (1.1)$$
where $S = \{ x \in \mathbb{R}^n \mid g_i(x) \geq 0, i = 1, \ldots, m \}$, $f$ and $g_i$, $i = 1, \ldots, m$ are real-valued functions defined on $\mathbb{R}^n$.

The primal methods solve the problem (1.1), some of them are: Beale’s method, see [2]; cutting-plane method, see [16]; gradient projection method, see [33]; Wolfe’s method, see [35]; the feasible direction methods, see [43] and the convex simplex method, see [42]. To see more algorithms, see [24]. On the other hand, other methods that also solve the problem (1.1), are for example: the barrier methods, these methods use the logarithmic barrier function (LBF) or also the inverse barrier function (IBF), see [11]; the penalty methods, see [6] and mixed interior-exterior penalty method, see [11].

In [4], the advantages of using the multiplier methods (also called the augmented lagrangian method) over the penalty methods are shown. 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 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 [27]; penalty/barrier multiplier method, see [3]; Bregman function for multiplier methods, see [10] 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 [26].

[13] and [29], solve problem (1.1) subject to equality constraints. Later, [32] adapted the Hestenes-Powell formulation for the nonlinear programming problem subject to inequality constraints. This adaptation, defines an augmented Lagrangian function without continuous second derivatives. This new formulation is known as Hestenes-Powell-Rockafellar augmented Lagrangian function. Later this function will be important to construct a new augmented Lagrangian function, which will be continuously differentiable, see [9]. On the other hand, in [17] and [21] proposed a twice differentiable augmented Lagrangian function. Subsequently, a Lagrangian function of class $C^\infty$ is studied in [1] and [28].

In the work of [34], they study the exponential multiplier method, proposed by [19]. They, study two rules for choosing the penalty parameters and guarantee the convergence in the ergodic sense (for the primal sequence). Other works, where convergence is studied in an ergodic sense, are: [14], [15], [18] and [27].

In [36], the hyperbolic penalty algorithm (HPA) is proposed, studied in [37], [38], [40] and applied in different types of problems, see [25] and [41]. The HPA induces a new augmented Lagrangian algorithm, to be called by
HALA-1992, see [39]. The characteristic of HALA-1992 is that it considers the updating of the penalty parameter. With this characteristic of HALA-1992 and under a set of assumptions about the problem (1.1), then in that way the first ideas are given to guarantee the convergence of HALA-1992. Now, on this occasion, unlike HALA-1992, we consider the fixed penalty parameter and we also consider the convexity assumption. This way, 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 existence of solutions and the convergence of HALA. We obtain a nonergodic convergence. In order for us to guarantee the global convergence of the algorithm proposed in this work, we use the following classic assumptions, which are widely used in the literature, such as: Slater constraint qualification, convexity and concavity in the objective function and restrictions, respectively. Some works that consider these assumptions are, [1], [3] and [27].

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

\[
(P) \quad x^* \in X^* = \arg \min \{ f(x) \mid x \in S \},
\]

where we denote by

\[
S = \{ x \in \mathbb{R}^n \mid g_i(x) \geq 0, \; i = 1, \ldots, m \},
\]

the convex feasible set of the problem (P) and where the function \( f : \mathbb{R}^n \to \mathbb{R} \) is convex, \( g_i : \mathbb{R}^n \to \mathbb{R}, \; i = 1, \ldots, m, \) are concave functions, we assume that \( f, g_i \) are continuously differentiable. That way (P) is a convex optimization problem. Also (P) will be called as the primal problem. In the following, we consider the assumptions.

**C1.** The optimal set \( X^* \) is nonempty, bounded and, consequently, compact.

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

A consequence of **C1** (see the Theorem 24 and Corollary 20 of [11]) is that the level set \( \{ x \in S \mid f(x) \leq \beta \} \), remains bounded for any value \( \beta \). 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 \( L : \mathbb{R}^n \times \mathbb{R}_+^m \rightarrow \mathbb{R} \), defined as
\[
L(x, \lambda) = f(x) - \lambda^T g(x) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x), \quad \lambda^T = (\lambda_1, ..., \lambda_m) \geq 0, \quad (2.2)
\]
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.,
\[
\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda^*_i \nabla g_i(x^*) = 0, \quad (2.3)
\]
\[
\lambda^*_i g_i(x^*) = 0, \quad i = 1, ..., m, \quad (2.4)
\]
\[
g_i(x^*) \geq 0, \quad i = 1, ..., m, \quad (2.5)
\]
\[
\lambda^*_i \geq 0, \quad i = 1, ..., m. \quad (2.6)
\]
Moreover, the set of optimal Lagrange multipliers \( \lambda^* \) is denoted by
\[
A^* = \left\{ \lambda \in \mathbb{R}_+^m \mid \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0, \ x^* \in X^* \right\},
\]
it is known that \( A^* \) is a bounded set (and hence compact set) due to C2. The dual function \( \Phi : \mathbb{R}_+^m \rightarrow \mathbb{R} \), is defined as follows
\[
\Phi(\lambda) = \inf_{x \in \mathbb{R}^n} L(x, \lambda), \quad (2.7)
\]
and the dual problem consists of finding
\[
(D) \quad \lambda \in A^* = \text{argmax} \{ \Phi(\lambda) \mid \lambda \in \mathbb{R}_+^m \}.
\]
3 Hyperbolic Penalty

The hyperbolic penalty, introduced in [36] and [37], is meant to solve the problem (P). The penalty method adopts the penalty function

\[ P(y, \alpha, \tau) = -\left(\frac{1}{2} \tan \alpha\right) y + \sqrt{\left(\frac{1}{2} \tan \alpha\right)^2 y^2 + \tau^2}, \]  

(3.8)

where \( \alpha \in [0, \frac{\pi}{2}] \) and \( \tau \geq 0 \). The graphic representation of \( P(y, \alpha, \tau) \), as shown in Fig. 1, is a hyperbole with asymptotes forming angles \( (\pi - \alpha) \) and 0 (zero) with the horizontal axis and having \( \tau \) as the intercept with the axis of the ordinates. For more details of this penalty, see [40].

In the hyperbolic penalty method, the solution to problem (P) is obtained by means of solving a sequence of subproblems, \( k = 1, 2, \ldots \), defined by minimization of the modified objective function

\[ F(x, \alpha^k, \tau^k) = f(x) + \sum_{i=1}^{m} P(g_i(x), \alpha^k, \tau^k). \]

The sequence of subproblems is obtained by the controlled variation of the two parameters "\( \alpha \)" and "\( \tau \)" in two different phases of the algorithm. The
geometric idea, which is the foundation of the hyperbolic penalty algorithm, see [36], is as follow:

- as shown in Fig. 2, initially the angle $\alpha$ of the asymptote of the penalty function increases, thus causing a significant increase in penalty outside the feasible region, while at the same time there is a reduction in penalty for points within the feasible region. This process continues until a feasible point is achieved.

- from that point on, $\alpha$ remains constant and the value of $\tau$ decreases sequentially. In this way internal penalty manages to become more and more irrelevant, maintaining the same level of prohibitiveness outside the feasible region. Fig. 3 illustrates the second phase of the process.

[36] offers a description of thirteen properties of function (3.8) that are used in developing the theory of hyperbolic penalty method. In order to continue
with our analyzes, first we will define

\[ \frac{1}{2} \tan \alpha = \lambda. \]  \hspace{1cm} (3.9)

Then we replace (3.9) in (3.8), then we get

\[ P(y, \lambda, \tau) = -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}, \]  \hspace{1cm} (3.10)

where \( P : (-\infty, +\infty) \times \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R} \).

In particular we use the following properties:

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 \( \tau \geq 0 \) and \( \lambda \geq 0 \).

P2) \( P(y, \lambda, \tau) \):

- is a convex decreasing function of \( y \) for \( \tau > 0 \) and \( \lambda \geq 0 \),
- is a convex nonincreasing function of \( y \) for \( \tau = 0 \) and \( \lambda \geq 0 \),
- is a convex function equal to \( \tau \) for \( \lambda = 0 \).
P3) $\lim_{y \to +\infty} P(y, \lambda, \tau) = 0$, for $\tau > 0$ and $\lambda > 0$.

P4) $\lim_{y \to -\infty} P(y, \lambda, \tau) = +\infty$, for $\tau > 0$ and $\lambda > 0$.

**Proposition 3.1** $\lim_{y \to +\infty} \{P(-y, \lambda, \tau) + P(y, \lambda, \tau)\} = +\infty$.

**Proof.** Considering P3 and P4, the proof is direct. □

4 Hyperbolic Augmented Lagrangian

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

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

$$= f(x) + \sum_{i=1}^{m} \left(-\lambda_i g_i(x) + \sqrt{(\lambda_i g_i(x))^2 + \tau^2}\right), \quad (4.11)$$

where $\tau > 0$ is the penalty parameter.

Note that this function belongs to class $C^\infty$ if the involved functions $f(x)$ and $g_i(x), \quad i = 1, ..., m,$ are too. On the other hand, a variation of (4.11) is proposed and studied in the work of [8] and [30].

By comparing (2.2) and (4.11), we see that the function $L_H$ may be put in the form

$$L_H(x, \lambda, \tau) = L(x, \lambda) + \sum_{i=1}^{m} \sqrt{(\lambda_i g_i(x))^2 + \tau^2}. \quad (4.12)$$

Analysis of expression (4.12) 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 \((x^*, \lambda^*)\) 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 \(\sum_{i=1}^{m} \tau\). From this point of view the summation in expression (4.12) 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.

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

4.1 Algorithm

**Step 1.** Let \(k := 0\) (initialization).

Take initial values \(\lambda^0 = (\lambda^0_1, ..., \lambda^0_m) \in \mathbb{R}^m_+, \ \tau \in \mathbb{R}_+\).

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

\[
x^{k+1} \in \text{argmin}_{x \in \mathbb{R}^n} L_H(x, \lambda^k, \tau) = \text{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^{m} \left( -\lambda^k_i g_i(x) + \sqrt{(\lambda^k_i g_i(x))^2 + \tau^2} \right) \right\}.
\]

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

\[
\lambda^{k+1}_i = \lambda^k_i \left( 1 - \frac{\lambda^k_i g_i(x^{k+1})}{\sqrt{(\lambda^k_i g_i(x^{k+1}))^2 + \tau^2}} \right), \ \ i = 1, ..., m.
\]

**Step 4.** If the pair \((x^{k+1}, \lambda^{k+1})\) 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 [32]. 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 [7]:

\[-\min_{i=1,...,m} g_i(x^k) < \beta \quad \text{and} \quad \frac{|f(x^k) - f(x^{k-1})|}{1 + |f(x^{k-1})|} < 10^{-2}\beta,\]

or

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

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, this can also be seen in [3].

### 4.2 Existence Result

Now, we are going to prove that the minimizer $x^{k+1}$ exists in the **Step 2** for all $k \geq 0$. Other existence results can be seen in [21] and [27]. Our next result follows a reasoning similar to Lemma 12 of [11] and [3].
Lemma 4.1  Suppose assumption C1 and C2 hold and the HPF, satisfies the properties P0-P4. Then for every $\tau > 0$ and $\lambda > 0$, the solution set of the unconstrained problem $\min_{x \in \mathbb{R}^n} L_H(x, \lambda, \tau)$, is nonempty.

Proof. Let $\tau > 0$ and $\lambda > 0$ be fixed. We are going to demonstrate that for any $l < \infty$, the level set

$$M = \{x \in \mathbb{R}^n \mid L_H(x, \lambda, \tau) \leq l\},$$

is bounded. For contradiction, there is a sequence $x_k \in M$ with $\lim_{k \to \infty} \|x_k\| = \infty$. Let $x^0 \in \text{int}(S) = \{x \in \mathbb{R}^n \mid g_i(x) > 0, i = 1, \ldots, m\} \neq \emptyset$, by C2. For some $\epsilon > 0$ define the set

$$D = \{x \in S \mid f(x) \leq f(x^0) + \epsilon\}.$$

From C1, we know that set $\{x \in S \mid f(x) \leq v^* = \inf_{x \in S} f(x)\} \neq \emptyset$ is compact. By Corollary 20, by [11], we obtain that the set $D$ is bounded (i.e., compact set). Therefore for $k$ large enough $x_k \notin D$ while $x^0 \in \text{int}(D)$. Let $w_k$ be a point on the line segment $[x^0, x_k]$ which is on $B$ ($B$ denote the boundary set of $D$ and note that $B$ is compact also). Now consider

$$w_k = \alpha_k x_k + (1 - \alpha_k) x^0, \quad 0 < \alpha_k < 1, \quad (4.14)$$

then we can obtain that

$$\|x_k\| - \|x^0\| \leq \|x_k - x^0\| = \frac{\|w_k - x^0\|}{\alpha_k},$$

so,

$$\|x_k\| \leq \frac{\|w_k - x^0\|}{\alpha_k} + \|x^0\|,.$$
since \( w^k \in B \) and \( B \) is bounded, the fact that \( \lim_{k \to \infty} \| x^k \| = \infty \) implies that

\[
\lim_{k \to \infty} \alpha^k = 0. \quad (4.15)
\]

By concavity of the \( g_i, i = 1, \ldots, m \), and from (4.14) it follows that

\[
g_i(w^k) \geq \alpha^k g_i(x^k) + (1 - \alpha^k)g_i(x^0), \quad i = 1, \ldots, m. \quad (4.16)
\]

Now, we claim that \( \{ g_i(x^k) \} \) is bounded in (4.16). Indeed (by contradiction), i.e., for some index \( i \), the function \( g_i \) is unbounded, that is, if \( g_i(x^k) \to -\infty \), then by P4, we obtain that \( P(g_i(x^k), \lambda_i, \tau) \to +\infty \). If for some \( j \neq i \), \( g_j(x^k) \to +\infty \), then by P3, we obtain that \( P(g_j(x^k), \lambda_j, \tau) \to 0 \), then the sum of the two terms, implies that

\[
L_{H}(x^k, \lambda, \tau) = f(x^k) + P(g_i(x^k), \lambda, \tau) + P(g_j(x^k), \lambda, \tau) \to +\infty,
\]

i.e., by Proposition 3.1, we get

\[
L_{H}(x^k, \lambda, \tau) \to +\infty,
\]

this contradicting the fact that \( x^k \in M \). Now, since we have (4.15), \( g_i(x^k) \) is bounded and \( g_i(x^0) > 0 \), then from (4.16), we have, for \( k \) large enough

\[
g_i(w^k) > 0, \quad i = 1, \ldots, m,
\]

hence \( w^k \in \text{int}(S) \), but since \( w^k \in B = \partial D \), it follows that

\[
f(w^k) = f(x^0) + \epsilon. \quad (4.17)
\]

By convexity of \( f \), (4.14) and (4.17), it follows that

\[
\alpha^k f(x^k) \geq f(w^k) - (1 - \alpha^k)f(x^0),
\]
i.e.,

\[ f(x^k) \geq f(x^0) + \frac{\epsilon}{\alpha^k}, \quad (4.18) \]

notice that since \( \epsilon > 0 \), we obtain \( f(x^k) \rightarrow +\infty \) as \( \alpha^k \rightarrow 0, k \rightarrow \infty \). Rewriting (4.16), it follows that

\[
\alpha^k g_i(x^k) \leq g_i(w^k) - (1 - \alpha^k)g_i(x^0),
\]

\[
g_i(x^k) \leq \frac{g_i(w^k) - (1 - \alpha^k)g_i(x^0)}{\alpha^k} \leq \max_{x \in B} \left\{ g_i(x^0) + \frac{g_i(x) - g_i(x^0)}{\alpha^k} \right\},
\]

so,

\[
g_i(x^k) \leq g_i(x^0) + \frac{d^0_i}{\alpha^k}, \quad i = 1, ..., m, \quad (4.19)
\]

where \( d^0_i = \max_{x \in B} \{ g_i(x) - g_i(x^0) \}, \quad i = 1, ..., m \) (this exists because \( B \) is compact), let \( d^0 = (d^0_1, ..., d^0_m) \). On the other hand, we now apply the property P2 in (4.19)

\[
P \left( g_i(x^0) + \frac{d^0_i}{\alpha^k}, \lambda, \tau \right) \leq P(g_i(x^k), \lambda, \tau), \quad i = 1, ..., m, \quad (4.20)
\]

we add the expressions (4.18) and (4.20), then

\[
f(x^k) + P(g_i(x^k), \lambda, \tau) \geq f(x^0) + \frac{\epsilon}{\alpha^k} + P \left( g_i(x^0) + \frac{d^0_i}{\alpha^k}, \lambda, \tau \right), \quad i = 1, ..., m,
\]

we rewrite the expression above,

\[
L_H(x^k, \lambda, \tau) \geq f(x^0) + \frac{\epsilon}{\alpha^k} + \sum_{i=1}^{m} P \left( g_i(x^0) + \frac{d^0_i}{\alpha^k}, \lambda, \tau \right), \quad (4.21)
\]

now in expression (4.21), let’s notice the following sets

\[
L_H(x^k, \lambda, \tau) \geq f(x^0) + \frac{\epsilon}{\alpha^k} + \sum_{i \in \{i | d^0_i > 0\}} P \left( g_i(x^0) + \frac{d^0_i}{\alpha^k}, \lambda, \tau \right)
\]
rewriting the inequality above, it follows
\[ L_H(x^k, \lambda, \tau) \geq f(x^0) + \sum_{i : d_0^i > 0} P \left( g_i(x^0) + \frac{d_0^i}{\alpha^k}, \lambda, \tau \right) \]
\[ + \frac{1}{\alpha^k} \left( \epsilon + \sum_{i : d_0^i < 0} \alpha^k P \left( g_i(x^0) + \frac{d_0^i}{\alpha^k}, \lambda, \tau \right) \right) \]
\[ + \frac{1}{\alpha^k} \left( \sum_{i : d_0^i = 0} \alpha^k P \left( g_i(x^0) + \frac{d_0^i}{\alpha^k}, \lambda, \tau \right) \right), \quad (4.22) \]
we will analyze the expression above with (4.15), as follows:

(b1) If \( i \in I_1 \), it follows that \( d_0^i > 0 \). From P3, there exists \( r > 0 \) such that
\[ P \left( y, \lambda, \tau \right) < 1, \quad \forall y > r. \quad (4.23) \]
Now, since \( \alpha^k > 0 \) and \( \lim_{k \to +\infty} \alpha^k = 0 \), there exists \( k_0 > 0 \), thus it follows
\[ \frac{d_0^i}{\alpha^k} > 0, \]
\[ g_i(x^0) + \frac{d_0^i}{\alpha^k} > g_i(x^0), \quad \forall k > k_0, \]
we consider (4.23) in the inequality above, this implies
\[ P \left( g_i(x^0) + \frac{d_0^i}{\alpha^k}, \lambda, \tau \right) \leq 1, \]
multiplying on both sides of the inequality by \( \alpha^k \), we get
\[ \alpha^k P \left( g_i(x^0) + \frac{d_0^i}{\alpha^k}, \lambda, \tau \right) \leq \alpha^k, \]
in the inequality above we apply the comparison test, so
\[ \lim_{k \to +\infty} \sum_{i : d_0^i > 0} \alpha^k P \left( g_i(x^0) + \frac{d_0^i}{\alpha^k}, \lambda, \tau \right) = 0. \]
(b2) If $i \in I_2$, it follows $d_i^0 < 0$. From P4, we obtain that $\forall \gamma > 0$, there exists $\beta$ such that

$$P(y, \lambda, \tau) > \gamma, \quad \forall y < \beta.$$  

(4.24)

Now, since $\lim_{k \to +\infty} \alpha^k = 0$ then there exists $k_0 > 0$, thus it follows

$$\frac{d_i^0}{\alpha^k} < 0,$$

$$g_i(x^0) + \frac{d_i^0}{\alpha^k} < g_i(x^0), \quad \forall k > k_0,$$

we consider (4.24) in the inequality above, so

$$P\left(g_i(x^0) + \frac{d_i^0}{\alpha^k}, \lambda, \tau\right) > \gamma,$$

from the above inequality, we can get

$$\lim_{k \to +\infty} \alpha^k \sum_{I_2 = \{i | d_i^0 < 0\}} P\left(g_i(x^0) + \frac{d_i^0}{\alpha^k}, \lambda, \tau\right) = +\infty.$$

(b3) If $i \in I_3$, it follows that $d_i^0 = 0$. We know that $\lim_{k \to +\infty} \alpha^k = 0$. Now

$$\lim_{k \to +\infty} \alpha^k P\left(g_i(x^0) + \frac{d_i^0}{\alpha^k}, \lambda, \tau\right) = P\left(g_i(x^0), \lambda, \tau\right) \lim_{k \to +\infty} \alpha^k = 0,$$

so,

$$\lim_{k \to +\infty} \alpha^k P\left(g_i(x^0) + \frac{d_i^0}{\alpha^k}, \lambda, \tau\right) = 0.$$

Finally, considering (b1), (b2) and (b3), in the expression (4.22), we obtain that

$$L_H(x^k, \lambda, \tau) \to +\infty,$$

a contradiction to $x^k \in M$. Thus, we obtain what is desired.
4.3 Study of the HALA

From Lemma 4.1 hence, there exists \( x^{k+1} \in \mathbb{R}^n \) such that

\[
L_H(x^{k+1}, \lambda^k, \tau) = \min_{x \in \mathbb{R}^n} L_H(x, \lambda^k, \tau),
\]

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

\[
\nabla f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i \left( 1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{(\lambda_i^k g_i(x^{k+1}))^2 + \tau^2}} \right) \nabla g_i(x^{k+1}) = 0, \quad (4.25)
\]

now, substituting (4.13) in (4.25), we have

\[
\nabla_x L_H(x^{k+1}, \lambda^k, \tau) = \nabla f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i(x^{k+1}) = \nabla_x L(x^{k+1}, \lambda^{k+1}) = 0, \quad (4.26)
\]

for any \( \tau > 0 \). That way, then \( x^{k+1} \) and \( \lambda^{k+1} \) also satisfies, \( \nabla_x L(x^{k+1}, \lambda^{k+1}) = 0 \), this showing 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.,

\[
\Phi(\lambda^{k+1}) = L(x^{k+1}, \lambda^{k+1}) = \min_{x \in \mathbb{R}^n} L(x, \lambda^{k+1}) \quad \text{and} \quad \lambda^{k+1} \in \mathbb{R}^m_{++},
\]

thus, it follows that

\[
\Phi(\lambda^{k+1}) = f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} g_i(x^{k+1}). \quad (4.27)
\]

From (4.27), we obtain that \( -g(x^{k+1}) = (-g_1(x^{k+1}), \ldots, -g_m(x^{k+1}))^T \in \partial \Phi(\lambda^{k+1}) \), where \( \partial \Phi(\lambda^{k+1}) \) is the subdifferential of \( \Phi(\lambda) \) at \( \lambda = \lambda^{k+1} \). On the other hand, since that the set of optimal Lagrangian multiplers is compact, the subdifferential of \( g \) is nonempty, see [31].
In the following remark, we analyze what happens with Lagrange multipliers (iteration (4.13)) depending on the type of restriction we have. First, for $x \in \mathbb{R}^n$, we define the following sets of indices

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

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

and

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

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 4.1** Let $\{\lambda^k\}$ be a sequence generated by HALA such that $\lambda^k_i > 0$, $i = 1, ..., m$ 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 that $g_i(x^{k+1}) = 0$, then by (4.13), we get, $\lambda^{k+1}_i = \lambda^k_i$. We can also obtain that, $(\lambda^k_i - \lambda^{k+1}_i) g_i(x^{k+1}) = 0, \forall i \in I_0$.

(c2) If $i \in I_+$, then we have at the $k$-th iteration that $g_i(x^{k+1}) > 0$, then by (4.13), we get, $\lambda^k_i > \lambda^{k+1}_i$. We can also obtain that, $(\lambda^k_i - \lambda^{k+1}_i) g_i(x^{k+1}) > 0, \forall i \in I_+$.

(c3) If $i \in I_-$, then we have at the $k$-th iteration that $g_i(x^{k+1}) < 0$, then by (4.13), we get, $\lambda^k_i < \lambda^{k+1}_i$. We can also obtain that, $(\lambda^k_i - \lambda^{k+1}_i) g_i(x^{k+1}) > 0, \forall i \in I_-$.

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

$$(\lambda^k_i - \lambda^{k+1}_i) g_i(x^{k+1}) \geq 0, \quad i = 1, ..., m.$$
In the following result, we will demonstrate the positivity of the updated Lagrange multipliers.

**Proposition 4.1** Let \( \{ \lambda^k = (\lambda^k_1, ..., \lambda^k_m) \mid k = 1, 2, ... \} \subset \mathbb{R}^m \). If

\[
\lambda^k \in \mathbb{R}^m_{++} \quad \text{then} \quad \lambda^{k+1} \in \mathbb{R}^m_{++}, \quad i = 1, ..., m.
\]

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

\[
(\lambda^k_i g_i(x^{k+1}))^2 < (\lambda^k_i g_i(x^{k+1}))^2 + \tau^2, \quad i = 1, ..., m,
\]

i.e.,

\[
-1 < -1 \frac{\lambda^k_i g_i(x^{k+1})}{\sqrt{(\lambda^k_i g_i(x^{k+1}))^2 + \tau^2}} < 1, \quad i = 1, ..., m,
\]

from the latter it follows that

\[
0 < \lambda^k_i \left(1 - \frac{\lambda^k_i g_i(x^{k+1})}{\sqrt{(\lambda^k_i g_i(x^{k+1}))^2 + \tau^2}}\right) < 2\lambda^k_i, \quad i = 1, ..., m,
\]

then from the expression above and by (4.13), we get that, \( \lambda^{k+1}_i > 0, \quad i = 1, ..., m \).

**Remark 4.2** From Lemma 4.1 and Proposition 4.1, we obtain that the HALA is well defined.

In the following results of this section, we follow a similar reasoning to the work of [27].

**Theorem 4.1** Let \( \{ \lambda^k \} \) be a sequence generated by HALA. The sequence \( \{ \Phi(\lambda^k) \} \) is monotone nondecreasing for all \( k \in \mathbb{N} \).
Proof. From the concavity of $\Phi(\cdot)$ and since $-g(x^{k+1}) \in \partial \Phi(\lambda^{k+1})$, we obtain

$$\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \sum_{i=1}^{m} \left( g_i(x^{k+1}) \right) \left( \lambda_i^k - \lambda_i^{k+1} \right). \tag{4.28}$$

On the other hand, we can rewrite (4.13), as follows,

$$\lambda_i^k - \lambda_i^{k+1} = \frac{\left( \lambda_i^k \right)^2 g_i(x^{k+1})}{\sqrt{\left( \lambda_i^k g_i(x^{k+1}) \right)^2 + \tau^2}}, \tag{4.29}$$

this expression (4.29), is replaced on the right side of inequality (4.28), we get that

$$\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \sum_{i=1}^{m} \left( \frac{\left( \lambda_i^k g_i(x^{k+1}) \right)^2}{\sqrt{\left( \lambda_i^k g_i(x^{k+1}) \right)^2 + \tau^2}} \right) \geq 0, \tag{4.30}$$

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

Proposition 4.2 The sequence of dual objective function values $\{\Phi(\lambda^k)\}$ is bounded and monotone nondecreasing, hence it converges.

Proof. By Theorem 4.1, we obtain $\Phi(\lambda^{k+1}) \geq \Phi(\lambda^k)$, then $\{\Phi(\lambda^k)\}$ is nondecreasing sequence for all $k \in \mathbb{N}$ 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 4.3 The sequence $\{\lambda^k\}$ generated by the HALA is bounded.

Proof. Recall that for $C2$ we obtain that $A^*$ is nonempty and compact, i.e., one level set of $\Phi(\cdot)$ is compact. Now, by Proposition 4.2, 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.
Now, we present a preliminary result, which will be used to guarantee the complementarity condition in our algorithm.

**Lemma 4.2** Let \( d > 0 \) and a sequence \( \{a^k\} \subset \mathbb{R}_+ \). If

\[
\lim_{k \to \infty} \left( \frac{a^k}{\sqrt{a^k + d}} \right) = 0 \quad \text{then} \quad \lim_{k \to \infty} a^k = 0.
\]

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

\[
\frac{a^k}{2\sqrt{a^k + d}} < \epsilon, \quad \forall k \geq k_0.
\]

On the other hand, we know that \( (\sqrt{a^k + d} - 1)^2 \geq 0 \), then, \( a^k + d + 1 \geq 2\sqrt{a^k + d} \), from (4.31) and from the previous inequality, we obtain

\[
\frac{a^k}{a^k + d + 1} \leq \frac{a^k}{2\sqrt{a^k + d}} < \epsilon, \quad \forall k \geq k_0,
\]

then of (4.32), we get \( a^k \leq (1 + d)/(1 - \epsilon) \), \( \forall k \geq k_0 \), which implies,

\[
\lim_{k \to \infty} a^k = 0.
\]

**Theorem 4.2** Let the sequences \( \{x^k\} \) and \( \{\lambda^k\} \) be generated by HALA. Then

\[
\lim_{k \to \infty} (\lambda^k_i g_i(x^k)) = 0, \quad i = 1, ..., m.
\]

**Proof.** Let be \( \tau > 0 \) fixed. Since \( \Phi(\cdot) \) is concave, we have the expression (4.30).

Now, we are going to verify that the series in (4.30) is convergent, i.e.,

\[
0 \leq \sum_{k=1}^{\infty} \sum_{i=1}^{m} \frac{(\lambda^k_i g_i(x^{k+1}))^2}{(\lambda^k_i g_i(x^{k+1}))^2 + \tau^2} \leq \sum_{k=1}^{\infty} (\Phi(\lambda^{k+1}) - \Phi(\lambda^k)),
\]

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

\[
0 \leq \sum_{k=1}^{\infty} \sum_{i=1}^{m} \frac{(\lambda^k_i g_i(x^{k+1}))^2}{(\lambda^k_i g_i(x^{k+1}))^2 + \tau^2} \leq \lim_{k \to \infty} (\Phi(\lambda^k) - \Phi(\lambda^1)) \leq f^* - \Phi(\lambda^1) < \infty,
\]
therefore, for the test of comparison, we have
\[
\lim_{k \to \infty} \sum_{i=1}^{m} \left( \frac{\left( \lambda_i^k g_i(x^{k+1}) \right)^2}{\sqrt{\left( \lambda_i^k g_i(x^{k+1}) \right)^2 + \tau^2}} \right) = 0. \tag{4.33}
\]

We note each term in the summation of (4.33) is nonnegative, then we obtain
\[
\lim_{k \to \infty} \left( \frac{\left( \lambda_i^k g_i(x^{k+1}) \right)^2}{\sqrt{\left( \lambda_i^k g_i(x^{k+1}) \right)^2 + \tau^2}} \right) = 0, \quad i = 1, \ldots, m, \tag{4.34}
\]

in (4.34), we can apply Lemma 4.2, with \( a^k = \left( \lambda_i^k g_i(x^{k+1}) \right)^2 \) and \( d = \tau^2 \) and thus we obtain that \( \lim_{k \to \infty} \left( \lambda_i^k g_i(x^{k+1}) \right)^2 = 0, \quad i = 1, \ldots, m, \) i.e., we obtain,
\[
\lim_{k \to \infty} \left( \lambda_i^k g_i(x^{k+1}) \right) = 0, \quad i = 1, \ldots, m. \tag{4.35}
\]

Because \( \Phi(\cdot) \) is a concave function and by Remark 4.1, we get that
\[
\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \sum_{i=1}^{m} (g_i(x^{k+1})) \left( \lambda_i^k - \lambda_i^{k+1} \right) \geq 0, \tag{4.36}
\]

and by Proposition 4.2, we know that \( \{ \Phi(\lambda^k) \} \) is convergent, so, it follows that
\[
\lim_{k \to \infty} \{ \Phi(\lambda^{k+1}) - \Phi(\lambda^k) \} = 0. \] Now, from (4.36) we obtain
\[
\lim_{k \to \infty} \sum_{i=1}^{m} (g_i(x^{k+1})) \left( \lambda_i^k - \lambda_i^{k+1} \right) = 0, \tag{4.37}
\]

since, \( (g_i(x^{k+1})) \left( \lambda_i^k - \lambda_i^{k+1} \right) \geq 0, \) of (4.37) and (4.35), it follows that
\[
\lim_{k \to \infty} \left( \lambda_i^{k+1} g_i(x^{k+1}) \right) = 0, \quad i = 1, \ldots, m. \tag{4.38}
\]
5 Convergence Result

Finally we guarantee that the sequence generated by HALA will converge towards a point of KKT.

**Theorem 5.1** The convex problem (P) satisfies C1-C2. Let sequences \( \{x^k\} \) and \( \{\lambda^k\} \) generated by HALA. Then any limit point of a sequence \( \{x^k\} \) and \( \{\lambda^k\} \) is an optimal solution-Lagrange multiplier pair for the problem (P).

**Proof.** Let be \( \tau > 0 \) fixed. From the Lemma 4.1 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 4.3. 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 \( k \in K_1 \subset \mathbb{N} \).

Now, we first show that the point \( \bar{x} \) is feasible (by contradiction). Assumed there exists at least some index \( i \) such that \( g_i(\bar{x}) < 0 \). From the continuity of \( g_i \) and the Remark 4.1, (c3), we can derive that the corresponding multiplier is increasing (i.e., will tend to infinity), this contradicts to the boundedness of multipliers. Hence \( \bar{x} \) is feasible. From Proposition 4.1, we obtain that, \( \lim_{k \to \infty} \lambda^k_i = \bar{\lambda}_i \geq 0, \ i = 1, \ldots, m \).

On the other hand, if we take the limit in (4.38), we obtain

\[
\lim_{k \to \infty} (\lambda^k_i + g_i(x^{k+1})) = \bar{\lambda}_i g_i(\bar{x}) = 0, \ i = 1, \ldots, m.
\]
Moreover, passing the limit in (4.26), we obtain
\[
\nabla_x L(\tilde{x}, \tilde{\lambda}) = \nabla f(\tilde{x}) - \sum_{i=1}^{m} \tilde{\lambda}_i \nabla g_i(\tilde{x}) = 0,
\]
thus \((\tilde{x}, \tilde{\lambda})\) satisfies (2.3) – (2.6) for all \(i = 1, ..., m\), hence \((\tilde{x}, \tilde{\lambda})\) is a KKT point. Thus \(\tilde{x}\) is optimal for the problem (P) and \(\tilde{\lambda}\) is a Lagrange multiplier.

6 Conclusions

- These results provide the necessary theoretical framework for the construction of a new algorithm to which we give the name Hyperbolic Augmented Lagrangian Algorithm. The completion of the development demonstrate the convergence of the proposed algorithm.
- The HPF belongs to class \(C^\infty\). Hence, \(L_H(x, \lambda, \tau)\) will be class \(C^\infty\) 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.
- The smooth behavior of the modified objective function offers the possibility of the best unconstrained minimization techniques, which use second-order derivatives.

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 supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, Brazil and FAPERJ E-26/201.176/2020 (258168), Brazil.

**References**


