Systems Engineering and Computer Science Graduate Program

Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineering

FEDERAL UNIVERSITY OF RIO DE JANEIRO

TECHNICAL REPORT

TR-PESC/Coppe/UFRJ- 791/2024

A Nonquadratic Augmented Lagrangian Algorithm for Nonconvex Programming

Authors:

Lennin Mallma Ramirez[1*], Nelson Maculan [2†], Adilson Elias Xavier3†, Vinicius Layter Xavier4†

[1,3*] Universidade Federal do Rio de Janeiro, PESC/COPPE, Rio de Janeiro, Brazil.

[2]Universidade Federal do Rio de Janeiro, PESC/IM/COPPE, Rio de Janeiro, Brazil.

[4]Universidade do Estado do Rio de Janeiro, Department of Epidemiology, Institute of Social Medicine, Rio de Janeiro, Brazil.

[†]These authors contributed equally to this work

Abstract

In this note, we extend the hyperbolic augmented Lagrangian algorithm (HALA) for solving nonconvex programming problems. That is, we guarantee that the sequence generated by HALA converges under mild assumptions to a Karush—Kuhn-Tucker (KKT) point.

Keywords: Nonlinear programming, First-order optimality, Convergence

Publication Date: 17/09/2024

NOTE
A Nonquadratic Augmented Lagrangian Algorithm for Nonconvex Programming

Lennin Mallma Ramirez[1*], Nelson Maculan [2†], Adilson Elias Xavier[3†], Vinicius Layter Xavier[4†]

[1,3*] Universidade Federal do Rio de Janeiro, PESC/COPPE, Rio de Janeiro, Brazil.
[2]Universidade Federal do Rio de Janeiro, PESC/IM/COPPE, Rio de Janeiro, Brazil.
[4]Universidade do Estado do Rio de Janeiro, Department of Epidemiology, Institute of Social Medicine, Rio de Janeiro, Brazil.

*Corresponding author(s). E-mail(s): lenninmr@cos.ufrj.br ;
Contributing authors: maculan@cos.ufrj.br ; adilson@cos.ufrj.br; viniciuslx@ime.uerj.br;
[†]These authors contributed equally to this work.

Abstract

In this note, we extend the hyperbolic augmented Lagrangian algorithm (HALA) for solving the nonconvex programming problems, that is, we guarantee that the sequence generated by HALA converges under mild assumptions to a Karush-Kuhn-Tucker (KKT) point.

Keywords: Nonlinear programming, First-order optimality, Convergence

1 Introduction

We are interested in the following nonconvex programming problem

$$
\begin{aligned}
(P) \qquad & \min \ f(x) \\
& s.t. \ \ g_i(x) \geq 0, \ \ i = 1, ..., m,
\end{aligned}
$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, ..., m$ are continuously differentiable functions. There is a variety of algorithms that solve the problem (P) when it is nonconvex, such as: [16], [5], [6], [1], [7] and [8]. In particular, the following augmented Lagrangian function is studied in [13] to solve the problem (P)

$$L_\rho(x, \lambda) = f(x) + \frac{\rho}{2} \left\| max \left\{ 0, z(x) + \frac{\lambda}{\rho} \right\} \right\|^2, \tag{1.1}$$

where the function $z(x) = -g(x)$, $\rho > 0$ is the penalty parameter and $\lambda \geq 0$ is the Lagrange multiplier. In [13] the authors study two augmented Lagrangian methods (ALM), they are: Standard ALM and Modified ALM, in both methods, the penalty parameters have an update formula. The main difference of these methods is the updating formula of the Lagrange multipliers. The authors comment that it is convenient to use the Modified ALM, since this method uses the safeguards technique to update the Lagrange multipliers. It is known that the algorithm ALGENCAN uses the safeguards technique, see, [1] and [4]. In [13], the authors present a computational result comparing the Standard ALM and the Modified ALM. On the other hand, HALA solves the problem (P) and its convergence to a Karush-Kuhn-Tucker (KKT) point is guaranteed recently in the work [14] when the problem has convexity assumptions. A characteristic of HALA is that it uses a nonquadratic penalty function. On the other hand, other studies of nonquadratic penalty functions can be seen in [10], [8] and [9]. The first studies on HALA can be seen in the thesis of Xavier [19]. The algorithm HALA belongs to the class of Standard ALM, that is, HALA does not use the safeguards technique. In this work the algorithm HALA has a fixed penalty parameter. The contribution of our work is to guarantee convergence towards a KKT point using HALA to solve problem (P) with nonconvexity assumptions.

The paper is organized as follows: Section 2, we present some basic results. Section 3, we will remember the characteristics of the hyperbolic penalty function and we propose the algorithm HALA and guarantee the convergence of this algorithm. In Section 4, we present a computational illustration to see the performance our proposed algorithm.

2 Preliminaries

In this chapter we consider some basic definitions of nonlinear programming. The Lagrangian function for the problem (P) is $L : \mathbb{R}^n \times \mathbb{R}^m_+ \to \mathbb{R}$,

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

where $\lambda_i \geq 0,\ i = 1, ..., m$, is the vector of Lagrange multipliers. The first-order KKT conditions for the problem (P) hold at the point x^*, if there exists λ^*, called a Lagrange multiplier vector, such that

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

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

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

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

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

$$P(y, \lambda, \tau) = -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}, \tag{2.7}$$

where $P : \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_{++} \to \mathbb{R}$. For more details of this function, see [18]. HPF is equivalent to a smoothing of the penalty studied by Zangwill, see [20].

3 Hyperbolic Augmented Lagrangian Algorithm

The algorithm HALA is studied in [14] for the convex case.

3.1 Hyperbolic Augmented Lagrangian

We define the hyperbolic augmented Lagrangian function of problem (P) by $L_H : \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}_{++} \to \mathbb{R}$,

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

$$= f(x) + \sum_{i=1}^{m} \left(-\lambda_i g_i(x) + \sqrt{(\lambda_i g_i(x))^2 + \tau^2} \right), \tag{3.8}$$

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

The multiplier updating formula (3.10) are derived by noticing the following fact

$$\nabla_x L_H(x^{k+1}, \lambda^k, \tau)$$

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

$$= \nabla f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i(x^{k+1}). \tag{3.11}$$

For any $x \in \mathbb{R}^n$, we define the index sets $I_0(x) = \{i \in \{1, ..., m\} \mid g_i(x) = 0\}$, and $I_+(x) = \{i \in \{1, ..., m\} \mid g_i(x) > 0\}$.

Algorithm 1 (HALA)

Step 1. Let $k := 0$. Let $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m_{++}$ and $\tau > 0$.
Choose a positive sequence $\{\epsilon^k\}_{k \in I\!N} \subset \mathbb{R}_+$ satisfying $\lim_{k\to\infty} \epsilon^k = 0$.

Step 2. If (x^k, λ^k) is a KKT point of the problem (P): Then Stop.

Step 3. Find an approximate minimizer $x^{k+1} \in \mathbb{R}^n$ of $L_H(x, \lambda^k, \tau)$, such that

$$\left\|\nabla_x L_H(x^{k+1}, \lambda^k, \tau)\right\| \leq \epsilon^k. \tag{3.9}$$

Step 4. Updating of Lagrange multipliers:

$$\lambda_i^{k+1} = \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}}\right), \quad i = 1, ..., m. \tag{3.10}$$

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

Remark 3.1. *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, $i \in I_0(x^{k+1})$, then we have at the k-th iteration that $g_i(x^{k+1}) = 0$, then by (3.10), we get, $\lambda_i^{k+1} = \lambda_i^k$.

(c2) If, $i \in I_+(x^{k+1})$, then we have at the k-th iteration that $g_i(x^{k+1}) > 0$, then by (3.10), we get, $\lambda_i^k > \lambda_i^{k+1}$.

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

$$\lambda^k \in \mathbb{R}^m_{++} \quad \textit{then} \quad \lambda^{k+1} \in \mathbb{R}^m_{++}.$$

Proof. See Proposition 3.2.1 of [14]. ■

3.2 Convergence

Theorem 3.1. *Let us consider problem (P). The whole sequences generate by HALA are convergent, i.e.,*

$$\lim_{k\to\infty} \lambda^k = \lambda^* \tag{3.12}$$

and $\lim_{k\to\infty} x^k = x^*$ where x^* is a feasible point, then (x^*, λ^*) is a KKT point of the problem (P).

Proof. We know that $g_i(x^*) \geq 0$, $i = 1, ..., m$ by hypothesis. By Proposition 3.1 and (3.12) we get

$$\lim_{k\to\infty} \lambda_i^k = \lambda_i^* \geq 0, \; i = 1, ..., m. \tag{3.13}$$

Now, let us see the following cases:

(i) If, $i \in I_0(x^*)$, we have $g_i(x^*) = 0$, and by (3.13) then we can get

$$\lambda_i^* g_i(x^*) = 0, \ \forall i \in I_0(x^*).$$

(ii) If, $i \in I_+(x^*)$, we have $g_i(x^*) > 0$, then there exists $\bar{k}$ such that for all $k \geq \bar{k}$, we have $g_i(x^{k+1}) > 0$, then by $(c2)$ of the Remark 3.1, Proposition 3.1 and (3.10), we get

$$0 < \lambda_i^{k+1} = \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}} \right) < \lambda_i^k, \ \forall k \geq \bar{k}, \ \forall i \in I_+(x^*). \quad (3.14)$$

By the squeeze theorem and (3.12), we obtain

$$\lim_{k \geq \bar{k}} \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}} \right) = 1, \quad \forall i \in I_+(x^*),$$

it immediately follows that

$$\lambda_i^* g_i(x^*) = 0, \quad \forall i \in I_+(x^*).$$

Therefore, from (i) and (ii), we obtain

$$\lambda_i^* g_i(x^*) = 0, \ i = 1, ..., m, \quad (3.15)$$

then, the condition of complementarity is assured. Note that from (3.10) and (3.15) we can get

$$\lambda_i^* \left(1 - \frac{\lambda_i^* g_i(x^*)}{\sqrt{\left(\lambda_i^* g_i(x^*)\right)^2 + \tau^2}} \right) = \lambda_i^*, \quad i = 1, ..., m. \quad (3.16)$$

On the other hand, from (3.9), (3.16), (3.11) and since we have $\lim_{k\to\infty} \epsilon^k = 0$, we have

$$\lim_{k\to\infty} \epsilon^k \geq \lim_{k\to\infty} \left\| \nabla_x L_H(x^{k+1}, \lambda^k, \tau) \right\|$$

$$= \lim_{k\to\infty} \left\| \nabla f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i(x^{k+1}) \right\| = \lim_{k\to\infty} \left\| \nabla L(x^{k+1}, \lambda^{k+1}) \right\| = 0,$$

i.e., we obtain $\nabla L(x^*, \lambda^*) = 0$. In this way we ensure that the sequence generated by HALA converges to a KKT point. ■

Remark 3.2. *In Theorem 3.1 it is assumed that the primal sequence converges to a feasible point, this assumption is also considered in the Theorem 5.1 of [2] and Theorem 2.3 of [13]. Also see [3], [12] and [15].*

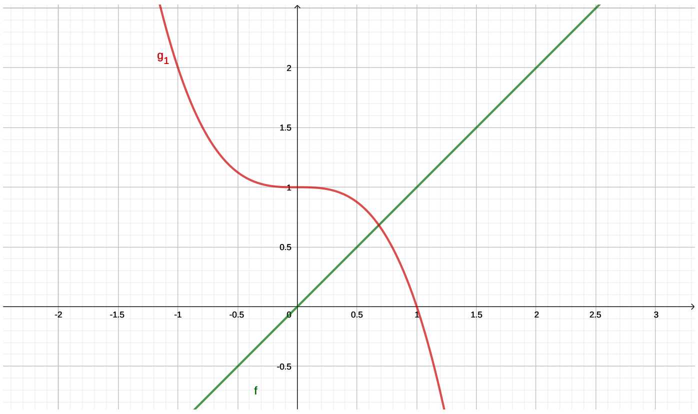

Fig. 1 Example 4.1

4 A Computer Illustration

The computational illustration presented below were obtained with a preliminary Fortran implementation for the 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 [11]. Let us now consider the following example which is also studied by Kanzow and Steck [13].

Example 4.1. *See Section 3 of [13].*

$$\min_{x\in\mathbb{R}} f(x) = x$$

$$s.t. \quad g_1(x) = 1 - x^3 \leq 0.$$

Starting from $x^0 = -1$. *The unique solution is* $x^* = 1$. *Moreover* $(x^*, \lambda^*) = (1, \frac{1}{3})$, *see Fig. 1.*

Our algorithm HALA will stop when the point x^{k+1} is feasible and the stationarity condition is satisfied, as follows

$$\left\|\nabla f(x^{k+1}) - \lambda^{k+1}\nabla g(x^{k+1})\right\| \leq 10^{-4}. \tag{4.17}$$

We are going to consider the following initial conditions:

$$x^0 = -1, \quad \lambda^0 = 10 \quad and \quad \tau = 0.10E - 09.$$

With the initial conditions considered, we report the following obtained values:

$$\left\|\nabla f(x^{k+1}) - \lambda^{k+1}\nabla g(x^{k+1})\right\| = 0.1711528E - 04,$$

$$x^* = 0.100000000E + 01 \quad and \quad \lambda^* = 0.333339038E + 00,$$

see, Table 1 and Table 2. From these tables, we can also observe that our algorithm HALA converges in two iterations. The Modified ALM solves this example in 42 iterations, see Kanzow and Steck [13]. We can also observe that our algorithm HALA converges in fewer iterations despite considering a fixed penalty parameter.

Table 1 Example 4.1

iteration	*viavel*	λ_1
0	0	0.100000000E+02
1	0	0.200000000E+02
2	1	0.333339038E+00

Table 2 Example 4.1

iteration	x	$f(x)$	$L_H(x, \lambda, \tau)$	*viavel*
0	-0.100000000E+01	-0.100000000E+01	0.390000000E+02	0
1	-0.129099445E+00	-0.129099445E+00	0.199139337E+02	0
2	0.100000000E+01	0.100000000E+01	0.100000000E+01	1

Declarations

- Funding: This work has been partially supported by FAPERJ (grant E-26/205.684/2022), COPPETEC Foundation, and CNPq (grants 302435/2019-0).

References

[1] Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. *SIAM J. OPTIM.* **18***(4)*, 1286–1309 (2007)

[2] Andreani, R., Birgin, E.G., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM J. OPTIM.* **20***(6)*, 3533–3554 (2010)

[3] Birgin, E.G., Fernández, D., Martínez, F.: The boundednesss of penalty parameters in an augmented Lagrangian method with constrained subproblems. *Optimization Methods and software.* **27**, 1001–1024 (2012)

[4] Birgin, E.G., Martínez, J.M.:*Practical Augmented Lagrangian Methods for Constrained Optimization.* Vol. 10. SIAM, Philadelphia (2014)

[5] Conn, A.R., Gould, N., Toint, PhL.: A globally convergent Lagrangian barrier algorithm for optimization with general inequality constraints and simple bounds. *Mathematics of Computation.* **66**, 261–288 (1997)

[6] Di Pillo, G., Lucidi, S.: An augmented Lagrangian function with improved exactness properties, *SIAM J. Optim.* **12***(2)*, 376–406 (2001)

[7] Dutta. J., Deb, K., Tulshyan, R., Arora, R.: Approximate kkt points and a proximity measure for termination. *J. Glob. Optim.* **56**, 1463–1499 (2013)

[8] Dussault, J.P.: Augmented non-quadratic penalty algorithms. *Math. Program.* **99**, 467–486 (2004)

[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] Goldfarb, D., Polyak, R., Scheinberg, K., Yuzefovich, I.: A mofified barrier-augmented Lagrangian method for constrained minimization. *Computational Optimization and Applications.* **14**, 55–74 (1999)

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

[12] Hartman, J.K.: Iterative determination of parameters for an exact penalty function. *Journal of Optimization Theory and Applications.* **16**, 49–66 (1975)

[13] Kanzow, C., Steck, D.: An example comparing the standard and safeguarded augmented Lagrangian methods. *Operations Research Letters.* **45**, 598–603 (2017)

[14] Mallma-Ramirez Lennin: *The hyperbolic augmented Lagrangian algorithm* (Ph.D thesis). Federal University of Rio de Janeiro/COPPE, Rio de Janeiro, Brazil (2022)

[15] Nguyen, V.H., Strodiot, J.J.: On the convergence rate for a penalty function method of exponential type. *Journal of Optimization Theory and Applications.* **27**, 495–508 (1979)

[16] Rockafellar, R.T.: Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control.* **12**, 268–285 (1974)

[17] 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 (1982)

[18] Xavier, A.E.: Hyperbolic penalty: a new method for nonlinear programming with inequalities. *Intl. Trans. in Op. Res.* **8**: 659–671 (2001)

[19] Xavier, A.E.: *Penalização Hiperbólica* (Ph.D. thesis). Federal University of Rio de Janeiro/COPPE, Rio de Janeiro, Brazil (1992)

[20] Zangwill, W.I.: Non-linear programming via penalty functions. *Management Sciences.* **13**, 344–358 (1967)