Instituto Alberto Luiz Coimbra de
Pós-Graduação e Pesquisa de Engenharia

THE HYPERBOLIC AUGMENTED LAGRANGIAN ALGORITHM

Lennin Mallma Ramirez

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientadores: Nelson Maculan Filho
Adilson Elias Xavier

Rio de Janeiro
Março de 2022

THE HYPERBOLIC AUGMENTED LAGRANGIAN ALGORITHM

Lennin Mallma Ramirez

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientadores: Nelson Maculan Filho
Adilson Elias Xavier

Aprovada por: Prof. Nelson Maculan Filho
Prof. Adilson Elias Xavier
Prof. José Mario Martínez Pérez
Prof. Juan Pablo Cajahuanca Luna
Profa. Laura Silvia Bahiense da Silva Leite
Prof. Sérgio Granville
Prof. Vinicius Layter Xavier

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2022

Ramirez, Lennin Mallma

The Hyperbolic Augmented Lagrangian Algorithm/Lennin Mallma Ramirez. – Rio de Janeiro: UFRJ/COPPE, 2022.

XII, 62 p.: il.; 29,7cm.

Orientadores: Nelson Maculan Filho

Adilson Elias Xavier

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2022.

Referências Bibliográficas: p. 55 – 62.

1. Nonlinear programming. 2. Hyperbolic augmented Lagrangian. 3. Constraint qualification. I. Maculan Filho, Nelson *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Dedico esta Tese à minha mãe
Leni, ao meu pai Felipe e aos
meus irmãos.

Agradecimentos

Os meus sinceros agradecimentos para:

Meus pais, Leni e Felipe, por ser minha principal referência e inspiração em minha vida.

Meus irmãos e irmã, por serem sempre meu apoio incondicional em todos esses anos.

Alessandra, Mathias, Rebeca, Romina e meus avós, por serem meus anjos e luz no dia a dia. Também aos meus familiares, pelo seu apoio, apesar da distância.

Aos meus professores e amigos, Nelson Maculan e Adilson Elias Xavier: pela atenção e constante comunicação que recebi de vocês em assuntos pessoais e acadêmicos. Por me fazerem parte da historia de um novo algoritmo que chamamos de HALA ("o algoritmo carioca"), e que agora nasceu no PESC/COPPE/UFRJ/Brasil.

Ao Prof. Vinicius Layter Xavier, pela ajuda nas implementações computacionais do nosso algoritmo HALA.

Ao Prof. Abilio Pereira de Lucena Filho, por sua ajuda quando eu precisei.

Ao Prof. Juan Pablo Luna, por sua atenção e sua vontade de querer ajudar, quando eu procurei.

Aos meus amigos Bao e Charlan, por sua disposição em falar sobre questões de otimização matemática e por toda a sua atenção nestes anos.

Aos professores: Laura Bahiense, Franklin de Lima Marquezino, Luis Felipe Magalhães de Moraes, Luidi Gelabert Simonetti e Glaydston Mattos Ribeiro, pelas aulas recebidas no meu doutorado.

Os professores da banca: Laura Silva Bahiense da Silva Leite, José Mario Martínez Pérez, Juan Pablo Cajahuanca Luna e Sérgio Granville, por aceitarem ser da banca e por fazerem parte da história do algoritmo HALA.

As diferentes pessoas que conheci nesses anos de estudos, pela atenção e pela ajuda nos momentos de dificuldade que recebi delas, levarei várias pessoas no meu coração.

Ao PESC/COPPE/UFRJ, por me proporcionar um ambiente de pesquisa agradável.

Ao Rio de Janeiro e Brasil, por me permitir desfrutar dos diferentes gêneros musicais deste belo país e também quero agradecer às rodas de samba do Rio, pela motivação e inspiração no desenvolvimento desta pesquisa.

Ao *Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq* e para a FAPERJ E-26/201.176/2020(258168), pelo apoio financeiro, para eu fazer minha pesquisa.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

O ALGORITMO LAGRANGIANO HIPERBÓLICO AUMENTADO

Lennin Mallma Ramirez

Março/2022

Orientadores: Nelson Maculan Filho
Adilson Elias Xavier

Programa: Engenharia de Sistemas e Computação

O algoritmo Lagrangiano hiperbólico aumentado (HALA) é introduzido na área de otimização contínua, para a resolução de problema de programação não linear. As hipóteses de convexidade, de diferenciabilidade e da condição de qualificação de Slater são consideradas para demonstrar a convergência do HALA. Estudamos também a teoria da dualidade para o caso da função Lagrangiana hiperbólica aumentada. Finalmente, para ilustrar o algoritmo, apresentamos alguns experimentos computacionais.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

THE HYPERBOLIC AUGMENTED LAGRANGIAN ALGORITHM

Lennin Mallma Ramirez

March/2022

Advisors: Nelson Maculan Filho
Adilson Elias Xavier

Department: Systems Engineering and Computer Science

The hyperbolic augmented Lagrangian algorithm (HALA) is introduced in the area of continuous optimization for solving nonlinear programming problems. Under mild assumptions, such as: convexity, Slater's qualification and differentiability, the convergence of the proposed algorithm is proved. We also study the duality theory for the case of the hyperbolic augmented Lagrangian function. Finally, in order to illustrate the algorithm, we present some computational experiments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

We are interested in the nonlinear programming problem subject to inequality constraints, as follows:

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

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 $I\!R^n$.

There are a wide variety of methods that solve the problem (1.1), some of them are: the gradient projection method (see [69]) and the feasible direction methods (see [92]). For a better idea of these methods, see [53].

On the other hand, methods that also solve the problem (1.1) are for example: the barrier methods, where the logarithmic barrier function (LBF),

$$l(x, r^k) = f(x) - r^k \sum_{i=1}^{m} ln(g_i(x)), \quad r^k > 0, \tag{1.2}$$

or also the inverse barrier function (IBF),

$$I(x, r^k) = f(x) + r^k \sum_{i=1}^{m} \frac{1}{g_i(x)}, \quad r^k > 0, \tag{1.3}$$

is used (see [26]). The penalty methods (see [15]) and mixed interior-exterior penalty method (see [26]) also have an important role to solve the problem (1.1). An interesting work where different penalty functions are studied computationally and the work of Birgin et al [12].

The methodology of the augmented Lagrangians or also called the Lagrange multiplier methods also solve the problem (1.1). The idea of these methods is to convert the constrained problem into a sequence of unconstrained problems. In [11], the advantages of using the multiplier methods over the penalty methods are shown.

When the problem (1.1) has convexity hypothesis, that is, the functions f and $-g_i,\ i=1,...,m$, are convex, there are a variety of augmented Lagrangian methods that solve this problem, some of them are:

(1) The augmented Lagrangian corresponding to the class of function ϕ : Kort and Bertsekas, 1972, see [48]. The subproblem of this augmented Lagrangian algorithm is as follows: Let $\phi : I\!R \to I\!R$,

$$min_{x\in I\!R^n}\left\{f(x)+\sum_{i=1}^{m}\lambda_i^k\phi_\epsilon\left(p_i(x)\right)\right\}, \tag{1.4}$$

the multipliers are updated as follows

$$\lambda_i^{k+1}=\lambda_i^k\phi'_\epsilon\left(p_i(x^k)\right),\quad i=1,...,m, \tag{1.5}$$

where $p_i(x)=-g_i(x),\ i=1,...,m,\ \epsilon>0,\ \phi_\epsilon(t)=\epsilon\phi\left(\frac{t}{\epsilon}\right)$ and ϕ satisfies 5 properties. For more details see [48].

(2) Proximal Point and Augmented Lagrangian Methods: Rockafellar, 1973, see [67] and also see Iusem [41]. The subproblem of this augmented Lagrangian algorithm is as follows:

$$min_{x\in I\!R^n}\left\{f(x)+r^k\sum_{i=1}^{m}\left[\left(max\left\{0,\lambda_i^k+\frac{h_i(x)}{2r^k}\right\}\right)^2-\left(\lambda_i^k\right)^2\right]+r^k\left\|x-x^k\right\|^2\right\}, \tag{1.6}$$

the multipliers are updated as follows

$$\lambda_i^{k+1}=max\left\{0,\lambda_i^k+\frac{h_i(x)}{2r^k}\right\},\quad i=1,...,m, \tag{1.7}$$

where $h_i(x)=-g_i(x),\ i=1,...,m,\ \left\{r^k\right\}\subset[\underline{r},\bar{r}]$ for $\bar{r}\geq\underline{r}>0$.

(3) The augmented Lagrangian corresponding to the class of function $p\in P_I$: Kort and Bertsekas, 1976, see [50]. The subproblem of this augmented Lagrangian algorithm is as follows: Let $p : I\!R^2 \to I\!R$, this function is continuous differentiable on $I\!R\times$

$(0, +\infty)$,

$$min_{x \in I\!R^n} \left\{ f(x) + r^k \sum_{i=1}^{m} p\left(\frac{h_i(x)}{r^k}, \lambda_i \right) \right\}, \tag{1.8}$$

the multipliers are updated as follows

$$\lambda_i^{k+1} = \nabla_1 p\left(\frac{h_i(x^k)}{r^k}, \lambda_i^k \right), \quad i = 1, ..., m, \tag{1.9}$$

where $h_i(x) = -g_i(x),\ i = 1, ..., m,\ r^k > 0$. The function p satisfies 8 properties. For more details see [50].

(4) The Quadratic Augmented Lagrangian Method: Rockafellar, 1973, see [67], Hestenes [37] and Powell [64]. The subproblem of this augmented Lagrangian algorithm is as follows:

$$x^k \in argmin_{x \in I\!R^n} \left\{ f(x) + \frac{1}{2c^k} \sum_{i=1}^{m} max\left\{0, \lambda_i^{k-1} + c^k p_i(x)\right\}^2 \right\}, \tag{1.10}$$

the multipliers are updated as follows

$$\lambda_i^k = max\left\{0, \lambda_i^{k-1} + c^k p_i(x^k)\right\}, \quad i = 1, ..., m. \tag{1.11}$$

where $p_i(x) = -g_i(x) \leq 0,\ c^k > 0$.

A solution of the problem (1.1) subject to equality constraints is proposed in [37] and [64]. Later, the Hestenes-Powell formulation was adapted for the nonlinear programming problem subject to inequality constraints (see [67]). 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 [23]). On the other hand, in [50] and [46] a twice differentiable augmented Lagrangian function is proposed. Subsequently, a Lagrangian function of class C^∞ is studied in [2] and [62].

(5) The Exponential Multiplier Method: Tseng and Bertsekas, 1997, see [75]. The subproblem of this augmented Lagrangian algorithm is as follows: Let $\psi : I\!R \to I\!R$

and $\psi(t) = e^t - 1$,

$$x^{k+1} \in argmin_{x\in I\!R^n} \left\{ f(x) + \sum_{i=1}^{m} \frac{\lambda_i^k}{c_i^k} \psi\left(c_i^k h_i(x)\right) \right\}, \tag{1.12}$$

the multipliers are updated as follows

$$\lambda_i^{k+1} = \lambda_i^k e^{c_i^k h_i(x^{k+1})}, \quad i = 1, ..., m, \tag{1.13}$$

where $h_i(x) = -g_i(x) \leq 0$, $c_i^k > 0$. Tseng and Bertsekas study the exponential multiplier method proposed by [48].

The authors Tseng and Bertsekas study two rules for choosing the penalty parameters and guarantentee the convergence in the ergodic sense. Other works, where the convergence is studied in an ergodic sense, are [41], [44], [47] and [60].

(6) Log-Sigmoid Multiplier Method: Polyak, Griva and Sobieszczanski-Sobieski, 1998, see [61] and Polyak, 2001, see [62]. The subproblem of this augmented Lagrangian algorithm is as follows:

$$x^{k+1} = argmin_{x\in I\!R^n} \left\{ f(x) + 2\beta^{-1} \sum_{i=1}^{m} \lambda_i \, ln\left(1 + e^{-\beta g_i(x)}\right) - 2\beta^{-1} \left(\sum_{i=1}^{m} \lambda_i \right) ln\ 2 \right\}, \tag{1.14}$$

the multipliers are updated as follows

$$\lambda_i^{k+1} = \frac{2\lambda_i^k}{1 + e^{\beta g_i(x^{k+1})}}, \quad i = 1, ..., m, \tag{1.15}$$

where $\beta > 0$.

(7) Nonlinear Rescaling Algorithm: Polyak and Teboulle, 1997, see [60]. The subproblem of this augmented Lagrangian algorithm is as follows: Let ψ be a C^2 on the interval $(a, +\infty)$, $-\infty \leq a < 0$, where $\psi(a) = -\infty$ and $\psi'(a) = +\infty$.

$$x^{k+1} \in argmin_{x\in I\!R^n} \left\{ f(x) - u^{-1} \sum_{i=1}^{m} \lambda_i^k \psi\left(u g_i(x)\right) \right\}, \tag{1.16}$$

the multipliers are updated as follows

$$\lambda_i^{k+1} = \lambda_i^k \psi' \left(u g_i(x^{k+1}) \right), \quad i = 1, ..., m, \tag{1.17}$$

where $u > 0$, this function ψ satisfies 6 properties. For more details, see [60].

Roman Polyak ([59]) modifies functions (1.2) and (1.3) as follows: the modified Frisch function, $F(x, \lambda, r) : I\!R^n \times I\!R^m_+ \times I\!R_+ \to I\!R$,

$$F(x, \lambda, r) = f(x) - \frac{1}{r} \sum_{i=1}^{m} \lambda_i ln(r\ g_i(x) + 1), \tag{1.18}$$

and the modified Carroll function, $C(x, \lambda, r) : I\!R^n \times I\!R^m_+ \times I\!R_+ \to I\!R$,

$$C(x, \lambda, r) = f(x) + \frac{1}{r} \sum_{i=1}^{m} \lambda_i \left(\frac{1}{(r\ g_i(x) + 1)} - 1 \right). \tag{1.19}$$

Thus, with the functions (1.18) and (1.19) augmented Lagrangian algorithms are studied, see [59] and [44]. Finally, other algorithms that solve convex problems are: [70], [33], [40], [54] and [57].

Hyperbolic Methodology

This methodology has been developed since previous decades, as follows:

- Xavier, 1982: Master's dissertation ([78]), "*Penalização Hiperbólica: Um Novo Método para Resolução de Problemas de Otimização*", advisor: João Lizardo Rodrigues Hermes de Araujo.

  In [78] the hyperbolic penalty algorithm (HPA) is proposed, it is also studied in [79], [80], [82] and [56].

- Xavier, 1992: Doctoral thesis, *"Penalização Hiperbólica"* ([81]), advisor: Nelson Maculan Filho.

  The HPA induces a new augmented Lagrangian algorithm, called HALA-1992 (see [81]). 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. In other words, convergence was not guaranteed in this work.

- Xavier, 2010: paper ([84]), "The Hyperbolic Smoothing Clustering Method".

  Some applications of this method are the following works: [83], [5], [86], [7], [89] and [85].

- Adilson Elias Xavier and Vinicius Layter Xavier, 2019: United States Patent and Trademark Office (USPTO), "Hyperbolic Smoothing Clustering and Minimum Distance Methods".
- Lennin Mallma-Ramirez, Nelson Maculan, Adilson Elias Xavier and Vinicius Layter Xavier, 2022: preprint, "Convergence Analysis of the Hyperbolic Augmented Lagrangian Algorithm".

"HALA
***E vai entrando na Otimização Matemática*"**

In this occasion, unlike HALA-1992, we consider the fixed penalty parameter and we also consider the convexity assumption. So, we propose in this work a new algorithm, which henceforth we will call HALA.

The main contribution of our work is to have guaranteed a basic existence result and the convergence of HALA. In order for us to guarantee the convergence of the algorithm proposed in this work, we use the following classic assumptions, which are widely used in the literature, such as the Slater constraint qualification and convexity. Some works that consider these assumptions are [2], [9] and [60]. On the other hand, currently different algorithms consider convexity assumptions, see: [28], [29], [72], [51], [21] and [39]. Therefore, after about 30 years we guarantee an existence result and a convergence result for HALA.

The thesis 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 the HALA. We also study some characteristics of this algorithm. We guarantee the convergence of the HALA and computational results are illustrated. In Chapter 4 The duality theory is studied for the case of the augmented hyperbolic Lagrangian function. In Chapter 5 we give some conclusions of our work. In Chapter 6 we propose some future work.

Chapter 2

Preliminaries

Throughout this thesis 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. Thus (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 [26]) 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

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x), \tag{2.1}$$

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.,

$$\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) - \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0, \tag{2.2}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, ..., m, \tag{2.3}$$

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

$$\lambda_i^* \geq 0, \quad i = 1, ..., m. \tag{2.5}$$

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

$$\Phi(\lambda) = \inf_{x \in I\!R^n} L(x, \lambda), \tag{2.6}$$

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)

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

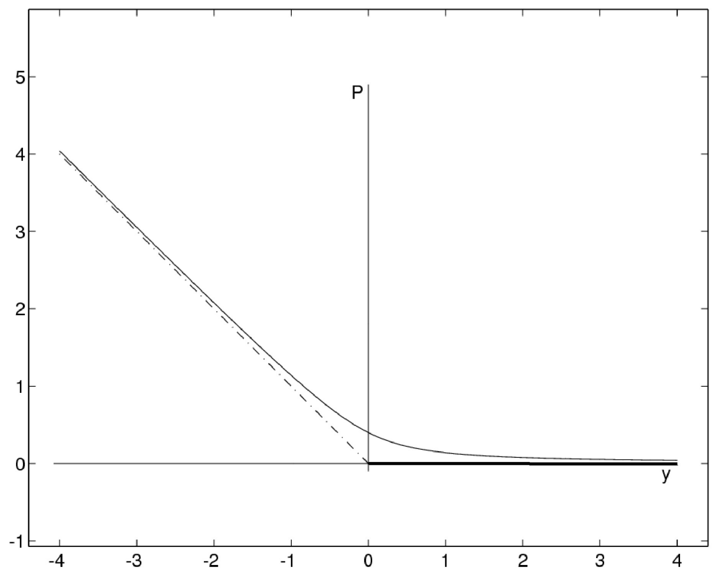

Figure 2.1: Hyperbolic Penalty

where $P : (-\infty, +\infty) \times I\!R_+ \times I\!R_{++} \to I\!R$, see Fig. 2.1

Remark 2.1.1 *Let, $\lambda \geq 0$, $y \geq 0$ and $\tau > 0$. Since we are $0 < \tau^2$, we can obtain the following inequalities,*

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

$$|\lambda y| < \sqrt{(\lambda y)^2 + \tau^2},$$

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

$$-\lambda y - \sqrt{(\lambda y)^2 + \tau^2} < 0 < -\lambda y + \sqrt{(\lambda y)^2 + \tau^2}. \tag{2.8}$$

From (2.8), we get that, $P(y, \lambda, \tau) > 0$.

Remark 2.1.2 *The HPF is originally proposed in [78] and studied in [82]. In these studies, 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* $\quad y \geq 0$.
- $P(y, \lambda, 0) = -2\lambda y, \quad$ *for* $\quad y < 0$.

Due to the properties (a) and (b) the HPF is equivalent to a smoothing of the penalty studied by Zangwill, in [90].

In particular we only use the following properties of the HPF (which are also studied in [78]):

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.,

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

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

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

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

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

P3) $P(0, \lambda, \tau) = \tau$, for $\tau > 0$ and $\lambda \geq 0$.

Chapter 3

Hyperbolic Augmented Lagrangian

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

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

$$= f(x) + \sum_{i=1}^{m} \left(-\lambda_i g_i(x) + \sqrt{(\lambda_i g_i(x))^2 + \tau^2} \right), \tag{3.1}$$

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 (3.1) is proposed and studied in the work of [17] and [65].

By comparing (2.1) and (3.1), we see that the function L_H may be put in the form

$$L_H(x, \lambda, \tau) = L(x, \lambda) + \sum_{i=1}^{m} \sqrt{(\lambda_i g_i(x))^2 + \tau^2}. \tag{3.2}$$

Analysis of expression (3.2) 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 (3.2) 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, let us consider the following assumption:

C3. For every $\tau > 0$ and $\lambda > 0$. Also for every $l < \infty$, the level set

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

is bounded.

Remark 3.0.1 *We know that the function P is convex by P1). Now the assumption* **C3** *is verified if in particular the function f is strongly convex in x. That way L_H will also be strongly convex in x. Other works that also use the strong convexity assumptions are: [2], [47], [10], [73], [68], [45], [71], [73] and [87]. See also [30].*

We present the HALA to solve the nonlinear problem (P).

3.1 Algorithm HALA

Step 1. Let $k := 0$ (initialization).
Take initial values $\lambda^0 = (\lambda_1^0, ..., \lambda_m^0) \in \mathbb{R}^m_{++}$, $\tau \in \mathbb{R}_{++}$.

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

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

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

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

$$\lambda_i^{k+1} = \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{(\lambda_i^k g_i(x^{k+1}))^2 + \tau^2}} \right), \quad i = 1, ..., m. \tag{3.3}$$

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.

HALA considers an initial vector $\lambda^0 > 0$ and $\tau > 0$. Note that HALA assumes a fixed value for the penalty parameter τ. Considering a fixed penalty parameter can also be observed in the following studies [44], [48], [49] and [67]. 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 [16]:

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

or

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

where $\beta > 0$.

Notice that HALA is based in the exact unconstrained minimization of the HALF. In [50] an exact unconstrained minimization of the augmented Lagrangian is also discussed, also see [9].

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.,

$$\nabla f(x^{k+1}) - \sum_{i=1}^m \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{(\lambda_i^k g_i(x^{k+1}))^2 + \tau^2}}\right) \nabla g_i(x^{k+1}) = 0, \tag{3.4}$$

substituting (3.3) in (3.4), we have

$$\nabla_x L_H(x^{k+1}, \lambda^k, \tau) = \nabla f(x^{k+1}) - \sum_{i=1}^m \lambda_i^{k+1} \nabla g_i(x^{k+1}) = \nabla_x L(x^{k+1}, \lambda^{k+1}) = 0, \tag{3.5}$$

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 (2.6)), i.e.,

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

thus, it follows that

$$\Phi(\lambda^{k+1}) = f(x^{k+1}) - \sum_{i=1}^{m} \lambda_i^{k+1} g_i(x^{k+1}). \tag{3.6}$$

From (3.6) 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 (3.3)) depending on the type of restriction we have. First, for $x \in I\!R^n$, we define the following sets of indices

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

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

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 3.2.1 *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 $i \in I_0$, then we have at the k-th iteration $g_i(x^{k+1}) = 0$, then by (3.3), 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 $g_i(x^{k+1}) > 0$.

So we can get,

$$\lambda_i^k g_i(x^{k+1}) > 0,$$

$$\frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}} > 0,$$

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

$$\lambda_i^k > \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}}\right), \tag{3.7}$$

then by (3.3) in (3.7), 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* $g_i(x^{k+1}) < 0$.

So we can get,

$$\lambda_i^k g_i(x^{k+1}) < 0,$$

$$\frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}} < 0,$$

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

$$\lambda_i^k < \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}}\right), \tag{3.8}$$

then by (3.3) in (3.8), 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

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

In the following result, we will demonstrate the positivity of the updated Lagrange multipliers.

Proposition 3.2.1 *Let* $\{\lambda^k = (\lambda_1^k, ..., \lambda_m^k) \mid k = 1, 2, ...\} \subset I\!R^m$. *If*

$$\lambda^k \in I\!R^m_{++} \quad \textit{then} \quad \lambda^{k+1} \in I\!R^m_{++}, \ \ i = 1, ..., m.$$

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

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

so,

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

thus,

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

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

$$0 < \lambda_i^k \left(1 - \frac{\lambda_i^k g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}}\right) < 2\lambda_i^k, \ i = 1, ..., m, \tag{3.9}$$

then from the expression above and by (3.3), we get that, $\lambda_i^{k+1} > 0, \ i = 1, ..., m.$ ■

Remark 3.2.2 *From inequality (3.9), we can see that iteration (3.3) has the following characteristic*

$$0 < \lambda_i^{k+1} < 2\lambda_i^k, \ \ i = 1, ..., m. \tag{3.10}$$

Remark 3.2.3 *From* **C3** *and Proposition 3.2.1, we obtain that HALA is well defined.*

Theorem 3.2.1 *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

$$\Phi(\lambda) - \Phi(\lambda^{k+1}) \leq \left(-g(x^{k+1})\right)\left(\lambda - \lambda^{k+1}\right), \tag{3.11}$$

now by rewriting (3.11) and considering $\lambda = \lambda^k$, we obtain

$$\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \left(g(x^{k+1})\right)\left(\lambda^k - \lambda^{k+1}\right), \tag{3.12}$$

writing again (3.12), so we have the expression

$$\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \sum_{i=1}^{m}\left(g_i(x^{k+1})\right)\left(\lambda_i^k - \lambda_i^{k+1}\right). \tag{3.13}$$

On the other hand, we can rewrite (3.3), as follows,

$$\lambda_i^k - \lambda_i^{k+1} = \frac{\left(\lambda_i^k\right)^2 g_i(x^{k+1})}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2 + \tau^2}}, \quad i = 1, ..., m, \tag{3.14}$$

this expression (3.14) is replaced on the right side of inequality (3.13), we get

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

so, we have,

$$\Phi(\lambda^{k+1}) \geq \Phi(\lambda^k).$$

■

Proposition 3.2.2 *The sequence of dual objective function values $\{\Phi(\lambda^k)\}$ is bounded and monotone nondecreasing, hence it converges.*

Proof. By Theorem 3.2.1 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 3.2.3 *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 [66]. By Proposition 3.2.2 we obtain in particular $\lambda^k \in \Gamma = \{\lambda \in I\!R_+^m \mid \Phi(\lambda^0) \leq \Phi(\lambda)\}$ for all $k \in I\!N$ 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 3.2.1 *Let $d > 0$ and a sequence $\{a^k\} \subset I\!R_+$. 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. By the hypothesis, for $\epsilon \in (0, 1)$ fixed, there exists $k_0 \in I\!N$, such that

$$\left|\frac{a^k}{\sqrt{a^k + d}}\right| < \epsilon, \ \ \forall k \geq k_0,$$

$$-\epsilon < \frac{a^k}{\sqrt{a^k + d}} < \epsilon, \ \ \forall k \geq k_0,$$

$$0 < \frac{a^k}{\sqrt{a^k + d}} + \epsilon < 2\epsilon, \ \ \forall k \geq k_0. \tag{3.16}$$

On the other hand, we know that $0 < \epsilon$, then

$$\frac{a^k}{\sqrt{a^k + d}} < \frac{a^k}{\sqrt{a^k + d}} + \epsilon, \tag{3.17}$$

now, we replace (3.17) in (3.16), then we get

$$\frac{a^k}{\sqrt{a^k + d}} < \frac{a^k}{\sqrt{a^k + d}} + \epsilon < 2\epsilon,$$

so, from the inequality above we have

$$\frac{a^k}{2\sqrt{a^k + d}} < \epsilon, \ \ \forall k \geq k_0. \tag{3.18}$$

Also, on the other hand, we know that $\left(\sqrt{a^k + d} - 1\right)^2 \geq 0$, then

$$a^k + d + 1 \geq 2\sqrt{a^k + d}, \ \ \forall k \geq k_0,$$

so,

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

from (3.18) and from the previous inequality, we obtain the following

$$\frac{a^k}{a^k + d + 1} \leq \frac{a^k}{2\sqrt{a^k + d}} < \epsilon, \ \ \forall k \geq k_0, \tag{3.19}$$

then of (3.19), we get

$$\frac{a^k}{a^k+d+1} \leq \epsilon, \quad \forall k \geq k_0,$$

$$a^k \leq a^k\epsilon + \epsilon\,(d+1)\,, \quad \forall k \geq k_0,$$

$$a^k - a^k\epsilon \leq \epsilon\,(d+1)\,, \quad \forall k \geq k_0,$$

$$a^k\,(1-\epsilon) \leq \epsilon\,(d+1)\,, \quad \forall k \geq k_0,$$

thus, $a^k \leq \left(\frac{\epsilon(d+1)}{1-\epsilon}\right), \quad \forall k \geq k_0$ which implies that $\lim_{k\to\infty} a^k = 0$. ■

Theorem 3.2.2 *Let the sequences* $\{x^k\}$ *and* $\{\lambda^k\}$ *be generated by HALA. Then*

$$\lim_{k\to\infty} \left(\lambda_i^k g_i(x^k)\right) = 0, \quad i = 1, ..., m. \tag{3.20}$$

Proof. Let be $\tau > 0$ fixed. Since $\Phi(\cdot)$ is concave we have the expression (3.15).

We are going to verify that the series in (3.15) is convergent; (3.15) 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

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

therefore, for the test of comparison, we obtain

$$\lim_{k\to\infty}\sum_{i=1}^{m}\left(\frac{\left(\lambda_i^k g_i(x^{k+1})\right)^2}{\sqrt{\left(\lambda_i^k g_i(x^{k+1})\right)^2+\tau^2}}\right) = 0. \tag{3.21}$$

We note each term in the summation of (3.21) is nonnegative, thus

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

in (3.22), we can apply the Lemma 3.2.1 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,

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

Because $\Phi(\cdot)$ is a concave function and by Remark 3.2.1 we get

$$\Phi(\lambda^{k+1}) - \Phi(\lambda^k) \geq \sum_{i=1}^{m}\left(g_i(x^{k+1})\right)\left(\lambda_i^k - \lambda_i^{k+1}\right) \geq 0, \tag{3.24}$$

and by Proposition 3.2.2 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 (3.24) we obtain

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

now since $\left(g_i(x^{k+1})\right)\left(\lambda_i^k - \lambda_i^{k+1}\right) \geq 0$ (by Remark 3.2.1), of (3.25) and (3.23), it follows that

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

3.3 Convergence Result

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

C4. The whole sequence $\{x^k\}$ 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 Hartman [34], Nguyen and Strodiot [55], [14], [27] and [24]. Finally, we ensure that the subsequence generated by the algorithm HALA converges to a KKT point, but not necessarily the entire sequence.

Theorem 3.3.1 *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 3.2.3. 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 3.2.1 we obtain,

$$\lim_{k\to\infty} \lambda_i^k = \bar{\lambda}_i \geq 0, \quad i = 1, ..., m. \tag{3.27}$$

Passsing the limit in (3.20), we have

$$\lim_{k\to\infty} \left(\lambda_i^k g_i(x^k)\right) = \bar{\lambda}_i g_i(\bar{x}) = 0, \quad \forall i = 1, ..., m. \tag{3.28}$$

Moreover, passing the limit in (3.5), 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 $(2.2) - (2.5)$ 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. ■

3.4 Computational Illustration

The computationally illustrate 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 [35]. 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^{-7}.

3.4.1 Test Problems

In this section we show five test problems.

Example 3.4.1

$$\begin{aligned} &\min_{x\in \mathbb{R}^3} f(x) = 2x_1^2 + 3x_2^2 + x_3^2 \\ s.t. \quad & g_1(x) = x_1^2 + x_2^2 + x_3^2 - 8 \leq 0, \\ & g_2(x) = 4x_1^2 - 32x_1 + 36x_2^2 - 144x_2 + 9x_3^2 - 18x_3 + 181 \leq 0. \end{aligned}$$

Starting with $x^0 = (1.5, 1.5, 1)$ and $f(x^0) = 12.2500000000$. The minimum value is $f(x^*) = 11.3792836271$ at the optimal solution
$x^* = (1.3958680939, 1.5256305407, 0.7069246891)$.

Example 3.4.2 *Problem 43 (Rosen-Suzuki) of [38].*

$$\begin{aligned} \min_{x \in I\!R^4} \; & f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\ s.t. \;\; & g_1(x) = 8 - x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 \geq 0, \\ & g_2(x) = 10 - x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 \geq 0, \\ & g_3(x) = 5 - 2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 \geq 0. \end{aligned}$$

Starting with $x^0 = (0, 0, 0, 0)$ and $f(x^0) = 0$. The minimum value is $f(x^*) = -44$ at the optimal solution $x^* = (0, 1, 2, -1)$.

Example 3.4.3 *Problem 11 of [38].*

$$\begin{aligned} \min_{x \in I\!R^2} \; & f(x) = (x_1 - 5)^2 + x_2^2 - 25 \\ s.t. \;\; & g_1(x) = -x_1^2 + x_2 \geq 0. \end{aligned}$$

Starting with $x^0 = (4.9, 0.1)$ (not feasible) and $f(x^0) = -24.98$. The minimum value is $f(x^*) = -8.498464223$.

Example 3.4.4 *See, Remark 4.2, of [62].*

$$\begin{aligned} \min_{x \in I\!R} \; & f(x) = 3x \\ s.t. \;\; & g_1(x) = x \geq 0. \end{aligned}$$

Starting with $\lambda^0 = 1$. The optimal solution is $X^* = \{0\}$ and $\Lambda^* = \{3\}$.

Example 3.4.5

$$
\begin{aligned}
\min_{x\in \mathbb{R}^4} f(x) = -6x_1 - 3x_2 - 2x_3 - x_4 \\
s.t. \qquad g_1(x) = x_1 \geq 0, \\
g_2(x) = x_2 \geq 0, \\
g_3(x) = x_3 \geq 0, \\
g_4(x) = x_4 \geq 0, \\
g_5(x) = -x_1 + 1 \geq 0, \\
g_6(x) = -x_2 + 1 \geq 0, \\
g_7(x) = -x_3 + 1 \geq 0, \\
g_8(x) = -x_4 + 1 \geq 0, \\
g_9(x) = -3x_1 - 2x_2 - 3x_3 - 3x_4 + 6 \geq 0.
\end{aligned}
$$

Starting with $x^0 = (0.1, 0.1, 0.5, 0.5)$ and $f(x^0) = -0.240000000E + 01$. The minimum value is $f(x^*) = -0.966666667E + 01$ at the optimal solution $x^* = (0.100000000E + 01, 0.100000000E + 01, 0.333333332E + 00, 0.621067930E - 09)$.

Example 3.4.6 *See, pag. 84 of [25].*

$$
\begin{aligned}
\min_{x\in \mathbb{R}} f(x) = 0x \\
s.t. \qquad g_1(x) = x \geq 0.
\end{aligned}
$$

Example 3.4.7 *See, pag. 30 of [42].*

$$
\begin{aligned}
\min_{x\in \mathbb{R}} f(x) = 1 \\
s.t. \qquad g_1(x) = 1 - e^x \geq 0.
\end{aligned}
$$

3.4.2 Results

Tables 3.1-3.11 summarize the computational results for these five problems. For each test problem we present two tables. The first table contains information about the primal sequence and the second table contains information about the dual sequence. Examples 3.4.1 and 3.4.5 are proposed by us. In Tables 3.1-3.11, we can see the feasible starting points considered.

For each table, k is the number of iterations, τ is the penalty parameter, λ is the Lagrange multiplier, x is the primal variable, $f(x)$ is the objective value, $g_i(x)$ are the constraints of each example, $L_H(x, \lambda, \tau)$ 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 Tables 3.1, 3.3, 3.5, 3.7 and 3.9, we reports the optimal solutions, the value of the objective function and the value of the HALF found by our proposed algorithm. For Problems 3.4.1, 3.4.2, 3.4.3, 3.4.4 and 3.4.5 our algorithm converges to the exact solution within the precision of the computer.

In Tables 3.2, 3.4, 3.6, 3.8, 3.10 and 3.11 we reports the behavior of the multipliers, this issue is studied in Subsection 3.2 of this work. In particular, in Table 3.2 and Table 3.4 we clearly observe that $\lambda_i^{k+1} = 2\lambda_i^k$, this happens when the solution is not viable and the value of τ is small. In these tables, we can also see the active constraints, for each proposed example. In Table 3.4 our active constraints are $g_1(x)$ and $g_3(x)$ as suggested in different papers, in particular see [50].

In the Example 3.4.3 when we consider the infeasible starting point $(4.9, 0.1)$ suggested by [38] our algorithm converges to the optimal solution in 7 iterations. When we consider the initial viable point $(1, 2)$ suggested by us, our algorithm converges in 5 iterations, see Table 3.5. In the Example 3.4.4 when we used the initial value $\lambda^0 = 1$ used by [62] with $\tau = 1$ we observe that our algorithm does not converge. But our algorithm converge when $\lambda^0 = 10$ and $\tau = 1$, then the results obtained by HALA are presented in Tables 3.7-3.8. In Problem 3.4.5, we can notice that the primal and dual sequence always remains in the viable region, see Tables 3.9-3.11.

In the Example 3.4.6, we consider the following initial parameters: $x^0 = 1$, $\lambda^0 = 1$ and $\tau = 1$. In this way our algorithm HALA converges to the primal solution and a Lagrange multiplier, see Tables 3.12 -3.13. On the other hand, the augmented Lagrangian algorithm studied by Eckstein also studies the same problem, but to guarantee the convergence of the Lagrange multipliers, he needs a condition on the gradient of the augmented Lagrangian function, for more details of this example see [25].

Table 3.1: Example 3.4.1 with $\tau = 0.10E - 04$

k	x_1	x_2	x_3	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	1.5000000000	1.5000000000	1.0000000000	12.2500000000	12.2500000	0
1	0.1538461538	0.3870967742	0.1525423729	0.5201381914	2.9858922	1
2	0.2962962963	0.6486486486	0.2647058823	1.5078874026	5.0869781	1
3	0.5517241379	0.9795918367	0.4186046511	3.6628294023	7.8299588	1
4	0.9696969696	1.3150684931	0.5901639344	7.4171333201	10.4781607	1
5	1.3958664938	1.5256298991	0.7069243209	11.3792683000	11.3793010	1
6	1.3958680939	1.5256305407	0.7069246891	11.3792836271	11.3793036	0
7	1.3958680939	1.5256305407	0.7069246891	11.3792836271	11.3793036	0

Table 3.2: Example 3.4.1 with $\tau = 0.10E - 04$

	$g_1(x)$		$g_2(x)$	
k	via	λ_1	via	λ_2
0	0	0.0100000	0	0.0100000
1	0	0.0000000	1	0.0200000
2	0	0.0000000	1	0.0400000
3	0	0.0000000	1	0.0800000
4	0	0.0000000	1	0.1600000
5	0	0.0000000	1	0.2680100
6	0	0.0000000	0	0.2680103
7	0	0.0000000	0	0.2680107

In the Example 3.4.7, we consider the following initial parameters: $x^0 = -2$, $\lambda^0 = 1$ and $\tau = 1$. In this way our algorithm HALA converges to the primal solution and a Lagrange multiplier, see Tables 3.14-3.15. On the other hand, using the generalized augmented Lagrangian method (GALB), convergence is not guaranteed for this example. Now, considering the generalized doubly augmented Lagrangian method (is based on proximal point methods), convergence is guaranteed for this example. For more details of this example see [42].

Table 3.3: Example 3.4.2 with $\tau = 0.10E-05$

k	x_1	x_2	x_3	x_4	$f(x)$	$L_H(x,\lambda,\tau)$	via
0	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0
1	2.3074726	2.3204238	5.0921913	-3.2752419	-79.7053622	-77.0243381	1
2	2.2312597	2.2670664	4.9403997	-3.0011307	-79.3079453	-74.4427220	1
3	1.9829192	2.0720262	4.6755395	-2.7161503	-78.1500359	-69.9582510	1
4	1.6153987	1.7745712	4.2248525	-2.1925613	-74.7549829	-63.0318244	1
5	1.0327825	1.2993558	3.5507600	-1.4755845	-66.4076345	-54.3478952	1
6	0.5301862	0.8883313	2.5995884	-0.8980092	-52.5776378	-47.4647475	1
7	0.1198347	1.0617978	2.1622807	-0.9736842	-46.6910951	-44.2047663	1
8	0.0000002	1.0000001	2.0000001	-0.9999997	-44.0000009	-43.9999972	1
9	0.0000000	1.0000000	2.0000000	-1.0000000	-44.0000000	-43.9999970	0
10	0.0000000	1.0000000	2.0000000	-1.0000000	-44.0000000	-43.9999970	0

Table 3.4: Example 3.4.2 with $\tau = 0.10E-05$

	$g_1(x)$		$g_2(x)$		$g_3(x)$	
k	via	λ_1	via	λ_2	via	λ_3
0	0	0.0100000	0	0.0100000	0	0.0100000
1	1	0.0200000	1	0.0200000	1	0.0200000
2	1	0.0400000	1	0.0400000	1	0.0400000
3	1	0.0800000	1	0.0800000	1	0.0800000
4	1	0.1600000	1	0.1600000	1	0.1600000
5	1	0.3200000	1	0.3200000	1	0.3200000
6	1	0.6400000	1	0.6400000	1	0.6400000
7	1	1.2800000	0	0.0000000	1	1.2800000
8	0	1.0000009	0	0.0000000	1	1.9999986
9	0	1.0000000	0	0.0000000	0	2.0000000
10	0	0.9999991	0	0.0000000	0	2.0000014

Table 3.5: Example 3.4.3 with $\tau = 0.10E-01$

k	x_1	x_2	$f(x)$	$L_H(x,\lambda,\tau)$	via
0	0.100000000E+01	0.200000000E+01	-0.500000000E+01	-0.499995000E+01	0
1	0.166667546E+01	0.999992090E+00	-0.128889633E+02	-0.933330521E+01	1
2	0.123552039E+01	0.152343892E+01	-0.850782714E+01	-0.848994743E+01	1
3	0.123477347E+01	0.152466288E+01	-0.849847226E+01	-0.848846423E+01	1
4	0.123477247E+01	0.152466328E+01	-0.849846350E+01	-0.848846422E+01	0
5	0.123477247E+01	0.152466328E+01	-0.849846350E+01	-0.848846422E+01	0

Table 3.6: Example 3.4.3 with $\tau = 0.10E - 01$

$g_1(x)$		
k	via	λ_1
0	0	0.100000000E+01
1	1	0.199998418E+01
2	1	0.304687788E+01
3	1	0.304932575E+01
4	0	0.304910476E+01
5	0	0.304888381E+01

Table 3.7: Example 3.4.4 with $\tau = 0.10E + 01$

k	x	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	0.100000000E+01	0.300000000E+01	0.304987562E+01	0
1	0.980196015E-01	0.294058804E+00	0.714142843E+00	0
2	0.177847483E-07	0.533542449E-07	0.100000000E+01	0
3	0.177847483E-07	0.533542449E-07	0.100000000E+01	0

Table 3.8: Example 3.4.4 with $\tau = 0.10E + 01$

$g_1(x)$		
k	via	λ_1
0	0	0.100000000E+02
1	0	0.300000016E+01
2	0	0.300000000E+01
3	0	0.299999984E+01

Table 3.9: Example 3.4.5 with $\tau = 0.10E + 00$

k	x_1	x_2	x_3	x_4	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	0.100000000E+00	0.100000000E+00	0.500000000E+00	0.500000000E+00	-0.240000000E+01	-0.239984689E+01	0
1	0.998900380E+00	0.998287244E+00	0.332446822E+00	0.221619670E-02	-0.965537385E+01	-0.964379954E+01	0
2	0.999350555E+00	0.998842976E+00	0.333095079E+00	0.158967189E-02	-0.966041209E+01	-0.876677236E+01	0
3	0.100000000E+01	0.100000000E+01	0.333333332E+00	0.621067930E-09	-0.966666667E+01	-0.876670427E+01	0
4	0.100000000E+01	0.100000000E+01	0.333333332E+00	0.621067930E-09	-0.966666667E+01	-0.876670427E+01	0

Table 3.10: Example 3.4.5 with $\tau = 0.10E + 00$

	$g_1(x)$		$g_2(x)$		$g_3(x)$		$g_4(x)$		$g_5(x)$	
k	via	λ_1	via	λ_2	via	λ_3	via	λ_4	via	λ_5
0	0	0.100000000E+04	0	0.100000000E+04	0	0.100000000E+04	0	0.100000000E+04	0	0.100000000E+04
1	0	0.501101427E-05	0	0.501717157E-05	0	0.452403132E-04	0	0.101646128E+01	0	0.410961635E+01
2	0	0.501076333E-05	0	0.501692014E-05	0	0.452334958E-04	0	0.100003904E+01	0	0.399997100E+01
3	0	0.501051225E-05	0	0.501666844E-05	0	0.452266755E-04	0	0.100003903E+01	0	0.399997100E+01
4	0	0.501026120E-05	0	0.501641677E-05	0	0.452198573E-04	0	0.100003902E+01	0	0.399997101E+01

Table 3.11: Example 3.4.5, continuation of Table 3.10 with $\tau = 0.10E+00$

	$g_6(x)$		$g_7(x)$		$g_8(x)$		$g_9(x)$	
k	via	λ_6	via	λ_7	via	λ_8	via	λ_9
0	0	0.100000000E+04	0	0.100000000E+04	0	0.100000000E+04	0	0.100000000E+04
1	0	0.170008383E+01	0	0.112201396E-04	0	0.502223585E-05	0	0.667605632E+00
2	0	0.166664901E+01	0	0.112193000E-04	0	0.502198402E-05	0	0.666678005E+00
3	0	0.166664901E+01	0	0.112184609E-04	0	0.502173182E-05	0	0.666678003E+00
4	0	0.166664901E+01	0	0.112176218E-04	0	0.502147964E-05	0	0.666678001E+00

Table 3.12: Example 3.4.6 with $\tau = 0.10E + 01$

k	x	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	0.100000000E+01	0.000000000E+00	0.414213562E+00	0
1	0.454823513E+08	0.000000000E+00	0.745058060E-08	0
2	0.536830181E+24	0.000000000E+00	0.745058060E-08	0
3	0.536830181E+24	0.000000000E+00	0.999999993E+00	0

Table 3.13: Example 3.4.6 with $\tau = 0.10E + 01$

	$g_1(x)$	
k	via	λ_1
0	0	0.100000000E+01
1	0	0.111022302E-15
2	0	0.123259516E-31
3	0	0.123259516E-31

Table 3.14: Example 3.4.7 with $\tau = 0.10E + 01$

k	x	$f(x)$	$L_H(x, \lambda, \tau)$	via
0	-0.200000000E+01	0.100000000E+01	0.145732056E+01	0
1	-0.374299419E+02	0.100000000E+01	0.141421356E+01	0
2	-0.374299419E+02	0.100000000E+01	0.174911755E+01	0

Table 3.15: Example 3.4.7 with $\tau = 0.10E + 01$

	$g_1(x)$	
k	via	λ_1
0	0	0.100000000E+01
1	0	0.292893219E+00
2	0	0.210565435E+00

Chapter 4

Duality Theory for the Hyperbolic Augmented Lagrangian

In this section we are interested in developing the duality theory for HALF in the Euclidean space.

The main result of this section is guarantee the strong duality for HALF for the convex case. In this way we assure a solution to the primal and dual problems. With these results, we can also note that HALF has properties similar to Log-sigmoid Lagrangian function (LSLF), see [62]; modified Frisch function (MFF) and Modified Carroll function (MCF), these last two functions are studied in [59].

Proposition 4.0.1 *Let us assume that if $f(x)$ and all $g_i(x) \in C^2$ and that $f(x)$ is strictly convex, then $L_H(x, \lambda, \tau)$ is strictly convex in $I\!R^n$ for any fixed $\lambda > 0$ and $\tau > 0$.*

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

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

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

$$+ \sum_{i=1}^{m} \frac{(\lambda_i)^2 g_i(x)}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}} \nabla^2_{xx} g_i(x). \tag{4.1}$$

In (4.1), the $\nabla^2_{xx} g_i(x)$ is factored, so, we can rewrite (4.1), as follows

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

$$+ \sum_{i=1}^{m} \left(\frac{(\lambda_i)^2}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}} - \frac{(\lambda_i)^4 g_i^2(x)}{((\lambda_i g_i(x))^2 + \tau^2)^{\frac{3}{2}}} \right) \nabla_x g_i(x) \nabla_x g_i^T(x). \tag{4.2}$$

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

$$\left(\lambda_i g_i(x)\right)^2 + \tau^2 > \left(\lambda_i g_i(x)\right)^2, \tag{4.3}$$

now we multiply by λ_i^2 in (4.3), so it follows that

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

the above inequality, we can rewrite it as

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

so,

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

thus, it follows

$$\frac{\lambda_i^2}{\left(\left(\lambda_i g_i(x)\right)^2 + \tau^2\right)^{\frac{1}{2}}} - \frac{\lambda_i^4 g_i^2(x)}{\left(\left(\lambda_i g_i(x)\right)^2 + \tau^2\right)^{\frac{3}{2}}} > 0. \tag{4.4}$$

We replace (4.4) in (4.2) and since $-\lambda_i \left(1 - \frac{\lambda_i g_i(x)}{\sqrt{(\lambda_i g_i(x))^2 + \tau^2}}\right) < 0$, (by P2) in (4.2), we get that, $\nabla^2_{xx} L_H(x,\lambda,\tau) > 0$, for $\lambda > 0$ and $\tau > 0$ fixed. ∎

Recall that strict convexity implies convexity.

Remark 4.0.1 *From* **C3** *and Proposition 4.0.1 for any* $\lambda > 0$ *and any* $\tau > 0$ *there exists a unique minimizer*

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

for problem (P) with the assumption **C1**.

4.1 Duality

In this section, we adapt the classic results already existing in the literature: Chapter 9 of [58] and Section 7 of [59] for our HALF. The following result is also verified by MFF and MCF, see [59].

Proposition 4.1.1 *Consider the convex problem (P). Assume the assumption* **C2** *it hold. Then $x^* \in S$ is a solution of problem (P) for any $\tau > 0$ if and only if:*

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

$$\lambda_i^* g_i(x^*) = 0,\ i = 1, ..., m \quad \text{and} \quad L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau),\ \ \forall x \in I\!R^n. \quad (4.5)$$

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

$$L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau) \geq L_H(x^*, \lambda, \tau), \quad \forall x \in I\!R^n,\ \ \forall \lambda \in I\!R^m_+. \quad (4.6)$$

Proof. ($\Rightarrow$) Let any $\tau > 0$ fixed. Assume x^* is a solution for convex problem (P) satisfying the assumption **C2**. Then system

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

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

has no solution in $I\!R^n$. Hence, by the Proper Separation Theorem (see, Theorem 2.26 (iv) of Dhara and Dutta [20]), there exists a vector $(\tilde{\lambda}, \hat{\lambda}) \neq (0, 0) \in I\!R \times I\!R^m$ such that

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

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

$$\tilde{\lambda}\,(f(x) - f(x^*)) \geq \sum_{i=1}^{m} \hat{\lambda}_i g_i(x), \quad (4.7)$$

for all $x \in I\!R^n$. Now, we follow an analysis similar to Theorem 4.2 of [20], so by **C2**, we have that there exists $\lambda_i^* = \frac{\hat{\lambda}_i}{\tilde{\lambda}}$, $i = 1, ..., m$, with $\tilde{\lambda} > 0$. Then, by (4.7) we have

$$f(x) - f(x^*) \geq \sum_{i=1}^{m} \lambda_i^* g_i(x), \quad (4.8)$$

for all $x \in I\!R^n$. In particular, (4.8) holds for $x = x^*$. So we get

$$0 \geq \sum_{i=1}^{m} \lambda_i^* g_i(x^*). \tag{4.9}$$

On the other hand, since, $g_i(x^*) \geq 0$ and $\lambda_i^* \geq 0$ for $i = 1, ..., m$, then by (4.9) we obtain

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, ..., m, \tag{4.10}$$

so we have the first part of (4.5).

Now, we are interested in proving the second part of (4.5). From (4.10) and (4.8), we have

$$f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) = f(x^*) \leq f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x), \tag{4.11}$$

for all $x \in I\!R^n$. Now, since we have (4.10), also, we can obtain

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

so, we have the following

$$\sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2} \leq \sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x))^2 + \tau^2}, \tag{4.12}$$

adding the expressions (4.11) and (4.12), we get

$$L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau), \quad \forall x \in I\!R^n, \tag{4.13}$$

in this way, we finish the proof of (4.5).

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

$$L_H(x^*, \lambda^*, \tau) = f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}, \tag{4.14}$$

considering (4.10); (4.14) becomes

$$L_H(x^*, \lambda^*, \tau) = f(x^*) + m\tau. \tag{4.15}$$

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

$$g_i(x^*) \geq 0, \quad i = 1, ..., m. \tag{4.16}$$

By applying the property P2 of HPF in (4.16), we obtain

$$P(g_i(x^*), \lambda_i, \tau) \leq P(0, \lambda_i, \tau), \quad i = 1, ..., m. \tag{4.17}$$

By applying property P3, on the right side of expression (4.17), we will obtain

$$P(g_i(x^*), \lambda_i, \tau) \leq \tau, \quad for \quad \lambda_i \geq 0, \quad i = 1, ..., m. \tag{4.18}$$

By performing the sum of 1 to m in (4.18) it follows

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

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

$$f(x^*) + \sum_{i=1}^{m} P(g_i(x^*), \lambda_i, \tau) \leq f(x^*) + m\tau. \tag{4.19}$$

By definition of L_H, (4.19) becomes

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

Now, by (4.20) and (4.15) we have

$$L_H(x^*, \lambda, \tau) \leq f(x^*) + m\tau = L_H(x^*, \lambda^*, \tau). \tag{4.21}$$

Finally, from (4.13) and (4.21), there is $\lambda^* \geq 0$ such that the primal-dual solution (x^*, λ^*) is a saddle point of L_H, $\forall x \in I\!R^n$ and $\tau > 0$.

($\Leftarrow$) We assume that (x^*, λ^*) is a saddle point of L_H, so (4.6) is hold. Then, for all $x \in I\!R^n$, $\lambda \in I\!R^m_+$ and for any $\tau > 0$ fixed, we have

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

$$\leq L_H(x^*, \lambda^*, \tau) = f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}. \tag{4.22}$$

From (4.22), we obtain

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

$$\leq -\sum_{i=1}^{m} \lambda_i^* g_i(x^*) + \sum_{i=1}^{m} \sqrt{(\lambda_i^* g_i(x^*))^2 + \tau^2}, \tag{4.23}$$

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

This relation (4.23) is possible only if $g_i(x^*) \geq 0$. Since, if this relation is violated ($i.e.,\ g_i(x^*) < 0$) for some index i, we can choose λ_i sufficiently large such that (4.23) becames false. So, x^* is feasible for problem (P).

We will prove the complementarity condition of (4.5). So again, by (4.23), and since that $\lambda_i \geq 0,\ i = 1, ..., m$, in particular taking $\lambda_i = 0,\ i = 1, ..., m$, in (4.23), we obtain

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

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

thus, it follows that

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

follows that

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

so,

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

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

$$\lambda_i^* g_i(x^*) = 0,\ i = 1, ..., m. \tag{4.24}$$

By (4.24) and definition of L_H, we obtain

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

From definition of saddle point, we know that $L_H(x, \lambda^*, \tau) \geq L_H(x^*, \lambda^*, \tau)$, by (4.25) and by definition of L_H, we can write

$$f(x^*) + m\tau = L_H(x^*, \lambda^*, \tau) \leq L_H(x, \lambda^*, \tau) = f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i^*, \tau). \tag{4.26}$$

On the other hand, once again considering property P2 of HPF, for any feasible point x, i.e., $g_i(x) \geq 0,\ i = 1, ...m$, we will carry out a work similar to that of (4.16)-(4.19), thus, we can obtain

$$f(x) + \sum_{i=1}^{m} P(g_i(x), \lambda_i^*, \tau) \leq f(x) + m\tau, \tag{4.27}$$

now, we replace (4.27) in (4.26), then follow

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

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

Let's consider the following definitions. Let

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

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

$$x^* = argmin \left\{ F_\tau(x) \mid x \in \mathbb{R}^n \right\}, \tag{4.28}$$

that is the problem (P) reduces to solving (4.28).

Let

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

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

$$\lambda^* = argmax \left\{ \phi_\tau(\lambda) \mid \lambda \geq 0 \right\}. \tag{4.29}$$

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

Proposition 4.1.2 *Let x be a feasible solution to problem (P) and let λ be a feasible solution to problem (4.29). Then*

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

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

$$\begin{aligned} \phi_\tau(\lambda) = \inf_{w \in I\!R^n} L_H(w, \lambda, \tau) &= \inf_{w \in I\!R^n} \left\{ f(w) + \sum_{i=1}^m P(g_i(w), \lambda_i, \tau) \right\} \\ &\leq \inf_{w \in S} \left\{ f(w) + \sum_{i=1}^m P(g_i(w), \lambda_i, \tau) \right\} \\ &= f(x) + \sum_{i=1}^m P(g_i(x), \lambda_i, \tau). \end{aligned} \tag{4.30}$$

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

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

we rewrite the expression above, as follows

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

now, we apply property P3, on the right side of the previous inequality

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

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

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

we replace (4.31) in (4.30), so

$$\phi_\tau(\lambda) \leq f(x)+m\tau, \quad \forall x \in S, \; \forall \lambda \in I\!R^m_+. \tag{4.32}$$

■

If $\hat{x}$ and $\hat{\lambda}$ are feasible solutions of the primal and dual problems and $F_\tau(\hat{x}) = \phi_\tau(\hat{\lambda})$, then $\hat{x} = x^*$ and $\hat{\lambda} = \lambda^*$. From Remark 4.0.1, with the smoothness of $f(x)$ and $g_i(x),\ i = 1,...,m$, we ensure the smoothness for the dual function $\phi_\tau(\lambda)$.

Theorem 4.1.1 *The problem (P) is considered. The assumption* **C2** *is verified. Then the existence of a solution of problem (P) implies that the problem (4.29) has a solution and*

$$\phi_\tau(\lambda^*) = f(x^*)+m\tau, \quad for\ any\ \ \tau > 0. \tag{4.33}$$

Proof. Let x^* be a solution of problem (P). By **C2**, we get $\lambda^* \geq 0$, such that (4.5) is verified. So we have

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

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

Therefore $\phi_\tau(\lambda^*) = \max\left\{\phi_\tau(\lambda) \mid \lambda \in I\!R^m_+\right\}$, in this way $\lambda^* \in I\!R^m_+$ is a solution of the dual problem and since we have $L_H(x^*,\lambda^*,\tau) = f(x^*)+m\tau$, so (4.33) hold. ■

Proposition 4.1.3 *Suppose that (4.33) holds, for the viable points x^* and λ^*, then x^* is a solution of the problem (P) and λ^* is a solution of the dual problem (4.29).*

Proof. Let $g_i(x^*) \geq 0,\ i = 1,...,m$, with $x^* \in S$, $\lambda_i^* \geq 0,\ i = 1,...,m$ and (4.33) with $\tau > 0$ fixed. Then for (4.32) where x and λ are viable, we can obtain the following

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

that is, x^* is solution of the problem (P) and λ^* is solution of (4.29), which corresponds the validity of the strong duality. ∎

4.2 Computational Illustration

We use HALA to guarantee the theory proposed in this work. The program were compiled by the GNU Fortran compiler version 4:7.4.0-1ubuntu2.3. The numerical Experiments are conducted on a Notebook with operating system Ubuntu 18.04.5, CPU i7-3632QM and 8GB RAM. The unconstrained minimization tasks were carried out by means of a Quasi-Newton algorithm employing the BFGS updating formula, with the function VA13 from HSL library [35]. The algorithm stop when the solution is viable (feasible) an the absolute value of the difference of the two consecutive solutions $\left|x^k - x^{k-1}\right|$ is less than $1.D - 5$.

We are going to take advantage of this section to make some comparisons of our algorithm HALA (see Table 4.14) with respect to the following algorithms:

Alg1=[22] which is an truncated Newton method;

Alg2=[32] which is a primal-dual interior point method;

Alg3=[76] which is an interior-point algorithm;

Alg4= [63] which is a QP-free method;

Alg5=[6] which is a primal-dual feasible interior-point method;

Alg6=[88] which is a feasible sequential linear equation algorithm;

Alg7=[77] which is an inexact first-order method;

Alg8=[36] which is a feasible direction interior-point technique;

Alg9=[1] which is an interior point algorithm.

4.2.1 Test Problems

With the following examples proposed in the book [38], we are going to verify the strong duality. On the other hand, in each example, the value of m means the total number of restrictions. Also, in all the examples starting points are considered, so that assumption **C2** is verified.

Example 4.2.1 *Problem 1 (HS1).*

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

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

Example 4.2.2 *Problem 30 (HS30).*

$$\begin{aligned} \min_{x \in \mathbb{R}^3} \; & f(x) = x_1^2 + x_2^2 + x_3^2 \\ s.t. \;\; & g_1(x) = x_1^2 + x_2^2 - 1 \geq 0, \\ & g_2(x) = x_1 - 1 \geq 0, \\ & g_3(x) = 10 - x_1 \geq 0, \\ & g_4(x) = x_2 + 10 \geq 0, \\ & g_5(x) = 10 - x_2 \geq 0, \\ & g_6(x) = x_3 + 10 \geq 0, \\ & g_7(x) = 10 - x_3 \geq 0. \end{aligned}$$

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

Example 4.2.3 *Problem 66 (HS66).*

$$
\begin{aligned}
\min_{x \in I\!R^3} \; & f(x) = 0.2x_3 - 0.8x_1 \\
s.t. \;\; & g_1(x) = x_2 - e^{x_1} \geq 0, \\
& g_2(x) = x_3 - e^{x_2} \geq 0, \\
& g_3(x) = x_1 \geq 0, \\
& g_4(x) = x_2 \geq 0, \\
& g_5(x) = x_3 \geq 0, \\
& g_6(x) = 100 - x_1 \geq 0, \\
& g_7(x) = 100 - x_2 \geq 0, \\
& g_8(x) = 10 - x_3 \geq 0.
\end{aligned}
$$

Starting with $x^0 = (0, 1.05, 2.9)$ (feasible), $f(x^0) = 0.58$ and $m = 8$. The minimum value is $f(x^*) = 0.5181632741$ at the optimal solution
$x^* = (0.1841264879, 1.202167873, 3.327322322)$.

Example 4.2.4 *Problem 76 (HS76).*

$$
\begin{aligned}
\min_{x \in I\!R^4} \; & f(x) = x_1^2 + 0.5x_2^2 + x_3^2 + 0.5x_4^2 - x_1x_3 + x_3x_4 - x_1 - 3x_2 + x_3 - x_4 \\
s.t. \;\; & g_1(x) = 5 - x_1 - 2x_2 - x_3 - x_4 \geq 0, \\
& g_2(x) = 4 - 3x_1 - x_2 - 2x_3 + x_4 \geq 0, \\
& g_3(x) = x_2 + 4x_3 - 1.5 \geq 0, \\
& g_4(x) = x_1 \geq 0, \\
& g_5(x) = x_2 \geq 0, \\
& g_6(x) = x_3 \geq 0, \\
& g_7(x) = x_4 \geq 0.
\end{aligned}
$$

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

Example 4.2.5 *Problem 100 (HS100).*

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

$$\begin{aligned} s.t. \ \ g_1(x) &= 127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0, \\ g_2(x) &= 282 - 7x_1 - 3x_2 - 10x_3^2 - x_4 + x_5 \geq 0, \\ g_3(x) &= 196 - 23x_1 - x_2^2 - 6x_6^2 + 8x_7 \geq 0, \\ g_4(x) &= -4x_1^2 - x_2^2 + 3x_1x_2 - 2x_3^2 - 5x_6 + 11x_7 \geq 0. \end{aligned}$$

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

4.2.2 Results

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

- Example 4.2.1: The HALA solves this example even though function f is nonconvex, see Tables 4.1 and 4.2.

- Example 4.2.2: the function f is strictly convex. From Table 4.3, we can see that in iteration 2 the Theorem 4.1.1 can be verified, that is, we have the following

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

  and

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

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. So, $x^* = (0.100000000E + 01, 0.100000000E + 01)$ is the solution of the primal problem and from Table 4.4 and Table 4.5, we can see the λ^* is the solution of the dual problem in the iteration 2.

- Example 4.2.3: the function f is linear. From Table 4.6, we can see that in iteration 3 the Theorem 4.1.1 can be verified, that is, we have the following

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

and

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

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. So, x^* is the solution of the primal problem and from Table 4.7 and Table 4.8, we can see the λ^* is the solution of the dual problem in the iteration 3.

- Example 4.2.4: the function f is strictly convex. From Table 4.9, we can see that in iteration 2, the Theorem 4.1.1 can be verified, that is, we have the following

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

and

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

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. So, x^* is the solution of the primal problem and from Table 4.10 and Table 4.11, we can see the λ^* is the solution of the dual problem in the iteration 2.

- Example 4.2.5: the function f is convex. From Table 4.2.2, we can see that in iteration 2, the Theorem 4.1.1 can be verified, that is, we have the following

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

and

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

then, $\phi_\tau(\lambda^*) = f(x^*) + m\tau$. The optimal value x^* is reported in the Table 4.2.2 and

Table 4.1: Example 4.2.1

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

Table 4.2: Example 4.2.1

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

the optimal value λ^* is reported in the Table 4.12.

In Table 4.14: we can see that HALA is more efficient in the sense that it uses fewer iterations with respect to the other algorithms. We can observe in the computational results that the HALA remains in the viable region in all the examples. On the other hand, despite being the theory developed in this work on convexity hypothesis, our algorithm shows in the Example 4.2.1 that it can also solve non-convex problems.

Table 4.3: Example 4.2.2

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

Table 4.4: Example 4.2.2

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

Table 4.5: Continuation of Table 4.4

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

Table 4.6: Example 4.2.3

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

Table 4.7: Example 4.2.3

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

Table 4.8: Continuation of Table 4.7

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

Table 4.9: Example 4.2.4

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

Table 4.10: Example 4.2.4

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

Table 4.11: Continuation of Table 4.10

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

Table 4.12: Example 4.2.5

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

Table 4.13: Example 4.2.5

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

Table 4.14: Iterations

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

Chapter 5

Conclusions

- 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. In this way, we introduce a new algorithm in the area of mathematical optimization.
- 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.
- The smooth behavior of the modified objective function offers the possibility to use the best unconstrained minimization techniques, which use second-order derivatives.

Chapter 6

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.
- Considering this first work that contains results of existence and convergence with strong assumptions, we consider it a future and natural work to carry out research considering more relaxed assumptions, to obtain more general results.
- Extend the convergence result of the HALA for the nonconvex problem:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ s.t.\quad & h(x) = 0, \\ & l \le x \le u \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable functions, l and u are vectors in $\mathbb{R}^n$ corresponding to lower and upper bounds in the variable, respectively, see [31]. This model is also studied in [3], [52], [19] and [74].

We are also interested in solving the following problem

$$\begin{aligned} &\min f(x) \\ s.t.\quad & g_i(x) = 0, \quad i = 1, ..., m < n \\ & h_i(x) \le 0, \quad j = 1, ..., r \\ & x^{min} \le x \le x^{max}, \end{aligned}$$

where $x \in I\!R^n$, g e h are continuously differentiable, see [8]. This model is also studied in [13] and [4].

- Solve the nonlinear programmin problem with equality constraints considering the hyperbolic proximal algorithm

$$x^{k+1} \in argmin \left\{ L_H(x, \lambda^k, \tau) + \frac{1}{2c^k} \left\| x - x^k \right\|^2 \right\},$$

$c^k > 0$ and update λ^{k+1} as in (3.3). A similar idea can be seen at [42], [91] and [43].

References

[1] Akrotirianakis, I., Rustem B.: Globally convergent interior-point algorithm for nonlinear programming. JOTA. 125, 497-521 (2005)

[2] 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)

[3] Andreani, R., Birgin, E. G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program. 111, 5-32 (2008)

[4] Andreani, R., Birgin, E. G., Martínez, J.M., Schuverdt, M.L.: Second-order negative-curvature methods for box-constrained and general constrained optimization. Comput Optim Appl. 45, 209-236 (2010)

[5] Bagirov, A.M., Al Nuaimat, A., Sultanova, N.: Hyperbolic smoothing function method for minimax problems. Optimization. 62, 759-782 (2013)

[6] Bakhtiari, S., Tits, A.L.: A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent. Computational Optimization and Applications. 25, 17-38 (2003)

[7] Bagirov, A.M., Ordin, B., Ozturk, G., Xavier, A. E.: An incremental clustering algorithm based on hyperbolic smoothing. Computational Optimization and Applications. 61, 219-241 (2015)

[8] Baptista, E.C., Belati, E. A., Costa, G.R.M.: Logarithmic barrier-augmented Lagrangian function to the optimal power flow problem. Electrical Power and energy Systems. 27, 528-532 (2005)

[9] Ben-Tal, A., Zibulevsky, M.: Penalty/Barrier multiplier methods for convex programming problems. SIAM J. OPTIM. 7, 347-366 (1997)

[10] Bertsekas, D.: On the Method of Multipliers for Convex Programming. IEEE Transactions on Automatic Control. 385-388 (1975)

[11] Bertsekas, D.P.: Multiplier methods: a survey. Automatica. 12, 133-145 (1976)

[12] Birgin, E.G., Castillo, R., Martínez, J.M.: Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems. Computational Optimization and Applications. 31, 31-56 (2005)

[13] Birgin, E.G., Martínez, J.M.: Improving ultimate convergence of an augmented Lagrangian method. Optimization Methods and Software. 23, 177-195 (2008)

[14] 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)

[15] Boukari, D., Fiacco, A.V.: Survey of penalty, exact-penalty and multiplier methods from 1968 to 1993. Optimization. 32, 301-334 (1995)

[16] Breitfeld, M.G., Shanno, D. F.: Computational experience with penalty-barrier methods for nonlinear programming. Annals of Operations Research. 62, 439-463 (1996)

[17] 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)

[18] Cominetti, R., Dussault, J.P.: A stable exponential penalty algorithm with superlinear convergence. J.O.T.A. 83, 285-309 (1994)

[19] Cristofari, A., Di Pillo, G., Liuzzi, G., Lucidi, S.: An Augmented Lagrangian Method Exploiting an Active-Set Strategy and Second-Order Information. Journal of Optimization Theory and Applications. https://doi.org/10.1007/s10957-022-02003-4 (2022)

[20] Dhara, A., Dutta, J.: Optimality conditions in Convex Optimization. A finite-dimensional View. CRC Press, Boca Raton (2012)

[21] Deng, W., Yin, W.: On the global and linear convergence of the generalized alternating direction method of multipliers. J Sci Comput. 66, 889-916 (2016)

[22] Di Pillo, G., Lucidi, S., Palagi, L.: A truncated Newton method for constrained optimization. In G. Di Pillo and F. Giannessi, editors, Nonlinear Optimization and Related Topics. Kluwer Academic, Dordrecht

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

[24] 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)

[25] Eckstein, J.: A practical general approximation criterion for methods of multipliers based on Bregman distances. Math. Program. 96, 61-86 (2003)

[26] Fiacco, A. V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. SIAM Classic in Applied Mathematics, Philadelphia (1990)

[27] Fernández, D., Pilotta, E.A., Torres, G.A.: An inexact restoration strategy for the globalization of the sSQP method. Comput Optim Appl. 54, 595-617 (2013)

[28] Goldfarb, D., Ma, S., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. Math. Program. 141, 349-382 (2013)

[29] Goldstein, T., O'Donoghue, B., Setzer, S., Baraniuk, R. : Fast alternating direction optimization methods. SIAM J. Imag. Sci. 7, 1588-1623 (2014)

[30] Goldfarb, D., Polyak, R., Scheinberg, K., Yuzefovich, I.: A modified barrier-augmented Lagrangian method for constrained minimization. Comput. Optim. Appl. 14, 55-74 (1999)

[31] Granville, S.: Optimal reactive dispatch through interior point methods. IEEE Trans. Power Systems. 9, 136-146 (1994)

[32] Griva, I., Shanno, D.F., Vanderbei R.J., Benson, H.Y: Global convergence of a primal-dual interior point method for nonlinear programming. Algorithmic Operations Research. 3, 12-29 (2008)

[33] Gonzaga, C.C., Castillo, R.A.: A nonlinear programming algorithm based on non-coercive penalty functions. Math. Program. 96, 87-101 (2003)

[34] Hartman, J.K.: Iterative determination of parameters for an exact penalty function. J. Optim. Theory Appl. 16, 49-66 (1975)

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

[36] Herskovits, J.: Feasible direction interior-point technique for nonlinear optimization. JOTA. 99, 121-146 (1998)

[37] Hestenes, M.R.: Multiplier and gradient methods. JOTA. 4, 303-320 (1969)

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

[39] Hong, M., Luo, Z.Q.: On the linear convergence of the alternating direction method of multipliers. Math. Program. 162, 165-199 (2017)

[40] Humes Jr, C., Silva, P.J.S., Svaiter, B.F.: Some inexact hybrid proximal augmented Lagrangian algorithm. numerical Algorithms. 35, 175-184 (2004)

[41] Iusem, A.N., Svaiter, B.F., Teboulle, M.: Entropy-like proximal methods in convex programming. Mathematics of Operations Research. 19, 790-814 (1994)

[42] Iusem, A.N.: Augmented Lagrangian methods and proximal point methods for convex optimization. Investigación Operativa. 8, 11-49 (1999)

[43] Izmailov, A.F., Solodov, M.V.: Perturbed Augmented LagrangianMethod Framework with Applications to Proximal and Smoothed Variants. Journal of Optimization Theory and Applications. https://doi.org/10.1007/s10957-021-01914-y (2021)

[44] 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)

[45] Kang M., Kang, M., Jung, M.: Inexact accelerated augmented Lagrangian methods. Comput Optim Appl. 62, 373-404 (2015).

[46] Kiwiel, K.C.: On the twice differentiable cubic augmented Lagrangian. Journal of Optimization Theory and Applications. 88, 233-236 (1996).

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

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

[49] 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)

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

[51] Lan, G., Monteiro, R.D.C.: Iteration-complexity of first-order augmented Lagrangian methods for convex programming. Math. Program. 155, 511-547 (2016)

[52] Leyffer, S., Vanaret, C.: An augmented Lagrangian filter method. Mathematical Methods of Operations Research. 92, 343-376 (2020)

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

[54] Necoara, I., Nesterov, Y., Glineur, F.: Linear convergence of first order methods for non-strongly convex optimization. Math. Program. 175, 69-107 (2019)

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

[56] Pereira, A.I.P.N., Fernandes, E.M.G.P.: A hyperbolic penalty filter method for semi-infinite programming. Numerical Analysis and Applied Mathematics. T.E. Simos, G. Psihoyios, and Ch. Tsitouras, eds., AIP Conference Proceedings, Springer-Verlag, Kos, Greece. 1048, 269-273 (2008)

[57] Pinheiro, R.B.N.M., Nepomuceno, L., Balbo, A.R.: Solving large-scale reactive optimal power flow problems by a primal-dual M^2BF approach. Optimization and Engineering. 21, 485-515 (2020)

[58] Polyak, B.T.: Introduction to Optimization. Optimization Software, Inc., New York (1987)

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

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

[61] Polyak, R., Griva, I., Sobieszczanski-Sobieski, J.: The Newton log-sigmoid method in constrained optimization. In Proceedings of the AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization. 2193–2201 (1998)

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

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

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

[65] 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).

[66] Rockafellar, R.T.: Convex analysis. Princeton University Press, Princeton, NJ. (1970)

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

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

[69] 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)

[70] Sagastizábal, C., Solodov, M.: An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter. SIAM J. Optim. 16, 146-169 (2005)

[71] Sabach, S., Teboulle, M.: Faster Lagrangian-based methods in convex optimization. arXiv:2010.14314v3 (2020)

[72] Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. SIAM J. Optim. 24, 269-297 (2014)

[73] 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)

[74] Torrealba, E.M.R., Silva, J.G., Matioli, L.C., Kolossoski, O., Santos, P.S.M.: Augmented Lagrangian algorithms for solving the continuous nonlinear resource allocation problem. European Journal of Operational Research. 299, 46-59 (2022)

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

[76] Vanderbei, R.J., Shanno, D. F.: Interior-point algorithm for nonconvex nonlinear programming. Computational Optimization and Applications. 13, 231-252 (1999)

[77] Wang, H., Zhang, F., Wang, J., Rong, Y.: An inexact first-order methodd for constrained nonlinear optimization. Optimization Method and Software. https://doi.org/10.1080/10556788.2020.1712601 (2020)

[78] 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)

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

[80] 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)

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

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

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

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

[85] 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)

[86] 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)

[87] Xu, Y.: Iteration complexity of inexact augmented lagrangian methods for constrained convex programming. Mathematical Programming. 185, 199-244 (2021)

[88] Yang, Y.F., Li, D.H., Qi, L.: A feasible sequential linear equation method for inequality constrained optimization. SIAM J OPTIM. 13, 1222-1244 (2003)

[89] Yilmaz, N., Sahiner, A.: Generalization of hyperbolic smoothing approach for non-smooth and non-Lipschitz functions. Journal of Industrial and Management Optimization. $doi : 10.3934/jimo.2021170$

[90] Zangwill, W.I.: Non-linear programming via penalty function. Management Science. 13, 344-358 (1967)

[91] Zhang, Y., Wu, J., Zhang, L.: The rate of convergence of proximal method of multipliers for equality constrained optimization problems. Optimization Letters. 14, 1599-1613 (2020)

[92] 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)