Universidade Federal do Piauí
Centro de Ciências da Natureza
Pós-Graduação em Matemática
Mestrado em Matemática

DC programming applied to image denoising

Maria Cheila Mamédio Cardoso

Teresina - 2025

Maria Cheila Mamédio Cardoso

Dissertação de Mestrado:

DC programming applied to image denoising

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Matemática, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientador:
Prof. Dr. João Carlos de Oliveira Souza

Teresina - 2025

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

DC programming applied to image denoising

Maria Cheila Mamédio Cardoso

Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**, outorgado pela Universidade Federal do Piauí.

A citação de qualquer trecho deste trabalho é permitida, desde que seja feita em conformidade com as normas da ética científica.

Dissertação aprovada em 28 de julho de 2025.

Banca Examinadora:

Documento assinado digitalmente
gov.br
JOAO CARLOS DE OLIVEIRA SOUZA
Data: 19/08/2025 10:48:10-0300
Verifique em https://validar.iti.gov.br

____

Prof. Dr. João Carlos de Oliveira Souza – Orientador

Documento assinado digitalmente
gov.br
RAY VICTOR GUIMARAES SERRA
Data: 15/08/2025 08:28:38-0300
Verifique em https://validar.iti.gov.br

____

Prof. Dr. Ray Victor Guimarães Serra - UFPI

Documento assinado digitalmente
gov.br
ELIANDERSON MENESES SANTOS
Data: 16/08/2025 09:48:19-0300
Verifique em https://validar.iti.gov.br

____

Prof. Dr. Elianderson Meneses Santos - IFMA

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Sistema de Bibliotecas UFPI - SIBi/UFPI
Biblioteca Setorial do CCN

C268d Cardoso, Maria Cheila Mamédio.
DC programming applied to image denoising / Maria Cheila Memédio Cardoso. -- 2025.
89 f.

Dissertação (Mestrado) - Universidade Federal do Piauí, Centro de Ciências da Natureza, Programa de Pós-Graduação em Matemática, Teresina, 2025.
“Orientador: Prof. Dr. João Carlos de Oliveira Sousa”.

1. Função DC. 2. Algoritmo de Diferença Convexa. 3. Processamento de Imagem. I. Sousa, João Carlos de Oliveira. II. Titulo.

CDD 516.362

Bibliotecária: Caryne Maria da Silva Gomes - CRB3/1461

Dedico o presente trabalho à minha mãe, Francisca Das Chagas Mamédio Cardoso.

Agradecimentos

Agradeço a Deus, fonte de toda sabedoria e força, por guiar e fortalecer meu caminho. Nunca me senti sozinha ou desamparada, porque Nele encontro abrigo, acalento e fé.

Agradeço ao meu alicerce: minha família. Aos meus irmãos, Fernando Mamédio e Francisco José Mamédio, e especialmente aos meus pais, Francisca das Chagas e Luís Mendes, por me incentivarem a seguir o caminho do conhecimento, da persistência e da honestidade, e por todo o amor, apoio incondicional e incentivo ao longo dessa jornada. Em especial, agradeço à minha mãe, que por toda a vida foi o mais puro cuidado, paciência, coragem e força. Eu a amo infinitamente.

Agradeço ao meu orientador Professor Doutor João Carlos de Oliveira Souza, pela orientação ao longos desses seis anos - quatro na graduação e dois no mestrado. Sua paciencia e apoio foram fundamentais.

Agradeço também aos meus amigos da House 404 — carinhosamente apelidada por nós. Vocês foram a minha família durante esse percurso: Marcelo Oliveira, Deyvid Andrade, Abel Henrique, Luana Sousa, Gabriel Aires e Natali Silva. Obrigada por cada momento, apoio, risada compartilhada e cada briga também.

Agradeço às minhas amigas de infância Maria de Fátima (prima e grande amiga), Daniele Magalhães e Amanda de Araújo, pelo companheirismo, carinho e consideração ao longo dos anos.

Agradeço a todos os professores do Departamento de Matemática da UFPI pela dedicação, compromisso com a educação e empenho na formação dos discentes. Durante essa caminhada, tive a honra de ser aluna de pessoas marcantes, como os professores João Xavier, Barnabé Lima, Jurandir Lopes, Isaías Pereira, Paulo Alexandre, José Francisco, Ítalo Dowell e Newton Santos.

Agradeço a todos os funcionários técnicos e terceirizados que, com dedicação e cuidado, tornam possível o bom funcionamento do Departamento de Matemática. O trabalho de vocês, muitas vezes nos bastidores, faz toda a diferença na nossa vivência acadêmica e merece ser reconhecido com respeito e gratidão.

Agradeço especialmente aos professores Jackson de Oliveira (meu padrinho), Antônio Marcos (matemático e amigo), Edilson Lopes e Cícero Drum, que me ajudaram a enxergar a beleza na Matemática e a acreditar no meu potencial nos tempos de Olimpíadas.

Agradeço à banca examinadora composta pelos professores doutores Ray Victor Guimarães Serra e Elianderson Meneses Santos, por aceitarem o convite, pelas contribuições e pelo tempo dedicado a este trabalho.

Agradeço a Fapepi pelo apoio financeiro, que foi essencial para que pudesse estudar em tempo integral.

"For my part, I know nothing with any certainty, but the sight of the stars makes me dream".

Vicent Van Gogh.

Resumo

Estudaremos o problema de minimização de uma função possivelmente não convexa, mas que pode ser expressa como a diferença entre duas funções convexas. Para resolver esse problema, utilizaremos três métodos consolidados na literatura: o Difference of Convex Algorithm (DCA), o Boosted Difference of Convex Algorithm (BDCA), que introduz uma direção de descida a partir de cada ponto gerado pelo DCA, acelerando a convergência por meio de uma estratégia de busca de linear; e o Non-Monotone Boosted Difference of Convex Algorithm (nmBDCA), uma variante mais recente que relaxa a exigência de redução monotônica da função objetivo, permitindo uma exploração mais flexível do espaço de busca ao admitir aumentos controlados dessa função, regulados por um parâmetro. Esses métodos serão aplicados ao problema de restauração de imagens degradadas por ruído, com o intuito de comparar sua eficácia e desempenho em termos de qualidade da imagem restaurada e de custo computacional. O objetivo é avaliar se essas estratégias baseadas em decomposição DC proporcionam ganhos significativos em comparação frente ao modelo convexo no contexto do processamento de imagens.

Palavras-chave: Funções DC; Algoritmos DC; nmBDCA; Melhoramento de imagens.

Abstract

We study the problem of minimizing a possibly nonconvex function that can be expressed as the difference of two convex functions. To solve this problem, we employ three well-established methods from the literature: the Difference of Convex Algorithm (DCA), the Boosted Difference of Convex Algorithm (BDCA), which introduces a descent direction from each point generated by DCA and accelerates convergence through a line search strategy; and the Non-Monotone Boosted Difference of Convex Algorithm (nmBDCA), a more recent variant that relaxes the requirement for monotone decrease of the objective function, allowing a more flexible exploration of the search space by admitting controlled increases in the objective, regulated by a parameter. These methods are applied to the problem of restoring images degraded by noise, with the aim of comparing their effectiveness and performance in terms of both image quality and computational cost. The objective is to evaluate whether these DC decomposition-based strategies provide significant gains compared to the convex model in the context of image processing.

Key-words: DC functions; DC Algorithms; nmBDCA; Image denoisy.

List of Figures

List of Tables

Contents

Introduction

In the present work, we aim to solve the following problem: given $\phi : \mathbb{R}^n \to \mathbb{R}$, a possibly non-convex function, but which can be rewritten as the difference between two convex functions $g, h : \mathbb{R}^n \to \mathbb{R}$. Functions with this structure are known as DC functions (difference of convex functions), and our goal is to find a minimizer of ϕ, that is,

$$\min_{x \in \mathbb{R}} \phi(x) = g(x) - h(x). \tag{1}$$

Observe that when the second component h is zero, the problem in (1) becomes convex.

The DC minimization problem has various applications in the literature, such as image processing [19], compressed sensing [31], location problem [9, 21, 12], the minimum sum-of-squares clustering problems [4, 24, 34], the multicast network design problem [32].

Given the problem's formulation and wide range of applications, it is necessary to investigate effective methods for addressing this type, including the subgradient-type [14, 26], proximal subgradient [13, 25, 29, 28], double bundle [46], codifferential [43] and inertial method [44]. In this work we will analyze the following methods: the Difference convex Algorithm (DCA), proposed in [11]; which uses the decomposition of the objective function as the difference of two convex functions to build an iterative sequence of approximations; the Boosted Difference convex Algorithm (BDCA), proposed in [3, 4]. In [3], the convergence of BDCA is proved if both functions g and h are differentiable, and the non-differentiable case is considered in [4], but still thinking that g is differentiable and h may not be.

The main idea behind BDCA is to define a descent direction using the points computed by DCA and a line search to do a longer step than the DCA, thus achieving a greater reduction in the objective value by iteration. In addition, this method speeds up the DCA, as it provides better solutions. The Non-monotone Boosted Difference of Convex Algorithm (nmBDCA), proposed in [5], when we eliminate the hypothesis of g being differentiable, the idea of a descent direction can no longer be applied, so the idea behind the non-monotone line search is that the growth of the function is allowed as long as it is controlled by a parameter.

This work proposes applying the nmBDCA to reconstruct noisy images, as in [7], by minimizing the difference between two convex functions whose components are non-differentiable, to show that nmBDCA outperforms DCA in CPU time, maintaining or outperforming the restored image quality.

This work is divided as follows: In chapter 1, we recall some theoretical results that will be presented and will be important for the development of the present work. In chapter 2, we introduced the DCA, highlighting its motivation and importance, along with convergence results. In chapter 3, the BDCA is presented, with a formal definition, convergence analysis. In chapter 4, we define the nmBDCA, discuss its main properties and convergence analysis. Subsequently, Chapter 5 introduces some definitions of image denoising and an application of the nmBDCA for the noise removal.

Chapter 1

Basic concepts and results

In this chapter we present the main results and basic concepts that will be used as the foundation for the development of this work. The definitions presented here can be easily found in Optimization books, see [2],[8],[15].

1.1 Convex sets and function

Definition 1.1.1. *A set* D *is called convex if for any* $a, b \in D$, *and* $\lambda \in [0, 1]$ *we have*

$$a(1 - \lambda) + b\lambda \in D.$$

Definition 1.1.2. *Let* $D \subset \mathbb{R}^n$ *be a convex set. A function* $f : D \to \mathbb{R}$ *is called convex if for any* $a, b \in D$ *and* $\lambda \in [0, 1]$, *it holds*

$$f(a(1 - \lambda) + b\lambda) \leqslant (1 - \lambda)f(a) + \lambda f(b).$$

Definition 1.1.3. *Let* $D \subset \mathbb{R}^n$ *be a convex set and* $f : D \to \mathbb{R}$ *be a convex function. Moreover, for given* $\sigma > 0$, f *is said to be* **σ*-strongly convex*** *if*

$$f(\lambda x + (1 - \lambda)y) \leqslant \lambda f(x) + (1 - \lambda)f(y) - \frac{\sigma}{2}\lambda(1 - \lambda)\|x - y\|^2,$$

for all $x, y \in \mathbb{R}^n$ *and* $\lambda \in [0, 1]$.

Proposition 1.1.1. *Strongly convex functions are convex, which in turn are continuous.*

Proof. See [15, Page 34]. □

Theorem 1.1.1. *Let* $D \subset \mathbb{R}^n$ *be a convex set and* $f : D \to \mathbb{R}$ *be a strongly convex function with modulus* $\sigma > 0$*. Then* f *has a unique minimizer.*

Proof. See [8, Proposition 3.4.20]. □

Proposition 1.1.2. *Let* $f_1 : \mathbb{R}^n \to \mathbb{R}$ *and* $f_2 : \mathbb{R}^n \to \mathbb{R}$ *be a strong convex and convex function, respectively, with modulus* $\sigma > 0$*. Then* $f_1 + f_2$ *is a strongly convex function with modulus* $\sigma > 0$*.*

Proof. By definition of strong convexity, for all $x, y \in \mathbb{R}^n$,

$$f_1(\lambda x + (1-\lambda)y) \leqslant \lambda f_1(x) + (1-\lambda)f_1(y) - \frac{\sigma}{2}\lambda(1-\lambda)\|x-y\|^2,$$

and

$$f_2(x\lambda + y(1-\lambda) \leqslant \lambda f_2(x) + (1-\lambda)f_2(y).$$

Adding these inequalities gives

$$(f_1+f_2)(x\lambda + (1-\lambda)y) - (f_1+f_2)(x) \leqslant (1-\lambda)(f_1+f_2)(y) - \lambda(1-\lambda)\frac{\sigma}{2}\|x-y\|^2.$$

Furthermore, we conclude that $f_1 + f_2$ is σ-strongly convex. □

Proposition 1.1.3. *Let* $f_1 : \mathbb{R}^n \to \mathbb{R}$ *be a convex function and let* $f_2 : \mathbb{R}^n \to \mathbb{R}$ *be a non-decreasing convex function. Then* $(f_2 \circ f_1)(x)$ *is a convex function.*

Proof. See [22, Proposition 1.39]. □

Definition 1.1.4. *Given a function* $f : \mathbb{R}^n \to \mathbb{R}$*, the conjugate of* f *is defined by*

$$f^*(y) = \sup_{x \in \mathbb{R}^n}\{\langle y, x\rangle - f(x)\}, \quad y \in \mathbb{R}^n.$$

Proposition 1.1.4. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function. For any* $x, y \in \mathbb{R}^n$*, the following assertions are equivalent:*

(i) $y \in \partial f(x)$*;*

(ii) $x \in \partial f^*(y)$*.*

Proof. See [39, Theorem 4.20, pag 104]. □

An important condition to analyze is convex functions in continuity. However, before we analyze this factor; let us define the Lipschitz function.

Definition 1.1.5. *A function* $f : D \subset \mathbb{R}^n \to \mathbb{R}$ *is* ***called Lipschitz continuous*** *if there exists a constant* $K \geqslant 0$*, such that:*

$$|f(x) - f(y)| \leqslant K\|x - y\|.$$

Moreover, we said that f *is* ***locally Lipschitz continuous*** *in* $x^* \in D$ *if there exists* $\epsilon > 0$ *such that* f *is Lipschitz on* $D \cap B(x^*; \epsilon)$*.*

Proposition 1.1.5. *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is a convex function, then* f *is locally Lipschitz. In fact,* f *is continuous in a neighborhood.*

Proof. See [15, page 34] or [39, Theorem 2.21]. □

Theorem 1.1.2. *Let* $D \subset \mathbb{R}^n$ *be a convex set and* $f : D \to \mathbb{R}$ *be a convex function. If* x^* *is a local minimizer of f, then* x^* *is a global minimizer of f.*

Proof. See [8, Theorem 1.2.1]. □

Theorem 1.1.3. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a differentiable function and* $D \subset \mathbb{R}^n$ *be a convex set. The function* f *is convex in* D *if, and only if,*

$$f(y) \geqslant f(x) + \nabla f(x)^T (y - x),$$

for any $x, y \in \mathbb{R}^n$*.*

Proof. See [8, Theorem 3.4.30]. □

1.2 The Subdifferential

Definition 1.2.1. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function. We said that* $y \in \mathbb{R}^n$ *is a subgradient of* f *at point* $x \in \mathbb{R}^n$ *if*

$$f(z) \geqslant f(x) + \langle y, z - x \rangle, \quad \forall z \in \mathbb{R}^n.$$

The set of all subgradients of f *in* x *called the subdifferential of* f *in* x*, and is denoted by* $\partial f(x)$*.*

The subgradient defines a linear approximation of the function f, whose graph is below f and whose value coincides with f at the point x.

Theorem 1.2.1. ***(Directional derivative of a convex function)***
Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function. Then, for all* $x \in \mathbb{R}^n$, f *differentiable in each direction* $d \in \mathbb{R}^n$. *Moreover,*

$$f(x + \alpha d) \geqslant f(x) + \alpha f'(x; d), \quad \forall \alpha \in \mathbb{R}_+,$$

where,

$$f'(x; d) = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha}.$$

Proof. See [8, Theorem 3.4.49]. □

Proposition 1.2.1. ***(The subdifferential of a convex function)***
Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function. Then for all* $x \in \mathbb{R}^n$, *the set* $\partial f(x)$ *is convex, compact, and nonempty. In addition, for all* $d \in \mathbb{R}^n$, *one must*

$$f'(x; d) = \max_{y \in \partial f(x)} \langle y, d \rangle.$$

Proof. See [8, Theorem 3.4.52]. □

Proposition 1.2.2. *A convex function* $f : \mathbb{R}^n \to \mathbb{R}$ *is differentiable at the point* $x \in \mathbb{R}^n$ *if only if the set* $\partial f(x)$ *contains only one element. In this case,* $\partial f(x) = \{\nabla f(x)\}$.

Proof. See [39, Theorem 3.33]. □

Theorem 1.2.2. ***(Optimality conditions for minimization of a convex function in a convex set)***
Let $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function. So* $\overline{x}$ *is a minimizer of* f *in* $\mathbb{R}^n$ *if and only if*

$$0 \in \partial f(\overline{x}).$$

Proof. See [8, Theorem 3.4.54]. □

Proposition 1.2.3. ***(Continuity of the subdifferential)***
Let $f : \mathbb{R}^n \to \mathbb{R}$ *a convex function and let* $\{x^k\} \to x$ $(k \to \infty)$ *and* $y^k \in \partial f(x^k)$ *for all* k. *Then, the sequence* $\{y^k\}$ *is bounded and all the cluster points of* $\{y^k\}_{k \in \mathbb{N}}$ *belong to* $\partial f(x)$.

Proof. See [8, Theorem 3.4.58]. □

Corollary 1.2.1. *If* $\phi : \mathbb{R}^n \to \mathbb{R}$ *is a convex and differentiable function. Then the gradient mapping* $\nabla \phi : \mathbb{R}^n \to \mathbb{R}$ *is continuous.*

Proof. See [27, Corollary 9.20]. □

Theorem 1.2.3. ***(Characterization of strongly differentiable convex functions)*** *Let $\Omega \subset \mathbb{R}^n$ a convex and open set and $f : \Omega \to \mathbb{R}$ a differentiable function in Ω with a continuous derivative in Ω. Then the following properties are equivalent:*

(a) The function f is strongly convex in Ω with modulus $\lambda > 0$.

(b) For each $x \in \Omega$ and all $y \in \Omega$

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle + \lambda \|y - x\|^2.$$

(c) For each $x \in \Omega$ and all $y \in \Omega$

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geqslant 2\lambda \|y - x\|^2.$$

Proof. See [8, Theorem 3.4.39.] □

When the function is not differentiable, the gradient is replaced by a term that belongs to the subdifferential of the function.

Definition 1.2.2. f *is called strongly monotone with modulus $\sigma > 0$ when*

$$\langle f(y) - f(x), y - x \rangle \geqslant \sigma \|y - x\|^2 \quad \forall x, y \in \mathbb{R}^n.$$

Theorem 1.2.4. *The following statements are equivalent:*

(i) $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with modulus $\sigma > 0$;

(ii) $f(y) \geqslant f(x) + \langle v, y - x \rangle + \frac{\sigma}{2}\|y - x\|^2$, for all $x, y \in \mathbb{R}^n$ and $v \in \partial f(x)$.

(iii) The point set operator $\partial f : \mathbb{R}^n \to \mathbb{R}$ is strongly monotone with modulus $\sigma > 0$, i.e.,

$$\langle u - v, x - y \rangle \geqslant \sigma \|x - y\|^2,$$

for all $u \in \partial f(x)$, $v \in \partial f(y)$ and for all $x, y \in \mathbb{R}^n$.

Proof. See [27, Theorem 5.24]. □

Note that, the function $f : \mathbb{R}^n \to \mathbb{R}$ is *strongly convex* with modulus $\sigma > 0$ if only if ∂f is *strongly monotone*.

1.2.1 Clarke's Subdifferential

An important concept to be studied is the Clarke subdifferential, as the generalization of the derivative allows extending the concept of derivative to non-differentiable functions. From studying this concept, it is necessary that all the functions mentioned here be locally Lipschitz continuous.

Definition 1.2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. The Clarke's subdifferential of f at $x \in \mathbb{R}^n$ is given by*

$$\partial_c f(x) = \{v \in \mathbb{R}^n \mid f^\circ(x; d) \geqslant \langle v, d\rangle, \quad \forall d \in \mathbb{R}^n\};$$

or equivalently,

$$\partial_c f(x) = \operatorname{conv}\{\lim_{i \to +\infty} \nabla f(x^i) \mid x^i \to x, \quad \exists \nabla f(x^i)\},$$

where $f^\circ(x; d)$ is the generalized directional derivative of f at x in the direction d given by

$$f^\circ(x; d) = \limsup_{\substack{y \to x \\ \alpha \downarrow 0}} \frac{f(y + \alpha d) - f(y)}{\alpha}.$$

Proposition 1.2.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz function. Then for all $x \in \mathbb{R}^n$ there is*

(i) $\partial_c f(x)$ is a nonempty, convex, compact subset of $\mathbb{R}^n$ and $\|v\| \leqslant K_x$ for all $v \in \partial_c f(x)$, where $K_x > 0$ is a Lipschitz constant of f around x;

(ii) $f^\circ(x; d) = \max_{v \in \partial_c f(x)} \langle v, d\rangle$.

Proof. See [2, Proposition 2.1.2]. □

When evaluated at x, the difference quotient is bounded above by $K|v|$, resulting from the fact that f is Lipschitz. So $f^\circ(x; d)$ is well defined and quantitatively finite. Furthermore, we will see that $f^\circ(x; d)$ is positively homogeneous and subadditive.

Proposition 1.2.5. *If f is convex, then $\partial_c f(x)$ coincides with the subdifferential of $\partial f(x)$ in the sense of convex analysis and $f^\circ(x; d)$ coincides with the usual directional derivative $f'(x; d)$.*

Proof. See [2, Proposition 2.2.7]. □

Proposition 1.2.6. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a locally lipschitz continuous function. Then, for any scalar* S*, one has*

$$\partial_c(Sf)(x) = \partial_c Sf(x).$$

Proof. See [2, Proposition 2.3.1]. □

Proposition 1.2.7. *Let* $f_i : \mathbb{R}^n \to \mathbb{R}$ *with* $i = 1, \cdots, n$*, be locally lipschitz continuous. So*

$$\partial_c(\sum_{i=1}^{n} f_i(x)) \subset \sum_{i=1}^{n} \partial_c f_i(x).$$

Moreover, for any scalar S_i *one has*

$$\partial_c(\sum_{i=1}^{n} (S_i f_i)(x)) \subset \sum_{i=1}^{n} \partial_c S_i f_i(x).$$

Proof. See [2, Proposition 2.3.3 and Corollary 2]. □

Definition 1.2.4. *The vector* $d \in \mathbb{R}^n$ *is called a descent direction for* $f : \mathbb{R}^n \to \mathbb{R}$ *at a point* $x \in \mathbb{R}^n$*, if there exists* $\varepsilon > 0$ *such that:*

$$f(x + td) < f(x), \quad \forall t \in (0, \varepsilon].$$

Proposition 1.2.8. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a locally Lipschitz continuous function at a point* $x \in \mathbb{R}^n$*. So* $d \in \mathbb{R}^n$ *is a descent direction for* f *at* x*, if*

$$f^\circ(x; d) < 0 \quad \text{or} \quad \langle v, d \rangle < 0, \quad \forall v \in \partial_c f(x).$$

Proof. See [1, Theorem 4.5]. □

The concept of generalized directional derivative extends the idea of derivative to a general class of functions, being extended to locally Lipschitz continuous functions.

Proposition 1.2.9. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a locally Lipschitz continuous function at a point* $x \in \mathbb{R}^n$*. If* f *reaches a local minimum at* x*, then* $0 \in \partial_c f(x)$*.*

Proof. See [2, Proposition 2.3.2]. □

Moreover, we said that the point $x^* \in \mathbb{R}^n$ is called Clarke stationary when $0 \in \partial_c f(x^*)$.

1.3 Difference of Convex Functions

In this section, we will present a remarkable concept of a class of functions that can be written as the difference of convex functions. Here, these functions are called DC functions.

A general DC problem is stated as follows:

$$\min_{x \in \mathbb{R}^n} \phi(x) := g(x) - h(x),$$

where g and h are convex functions on $\mathbb{R}^n$. The function ϕ is called a DC function and $g - h$ is a DC decomposition of ϕ with DC components g and h. We will represent the class of continuously differentiable functions in $\mathbb{R}^n$ up to order k, by $C^k(\mathbb{R}^n)$.

It is immediately apparent that each convex function can be expressed as a difference between convex functions; it is easy to assume that the second component is equal to zero. The following results provide a wide range of examples of DC functions, which will be important for a good understanding of the topic and analysis of the methods discussed.

Proposition 1.3.1. *Every function in* $C^2(\mathbb{R}^n)$ *is DC on any compact convex set* $C \subset \mathbb{R}^n$.

Proof. See [17, Proposition 4.2]. □

Proposition 1.3.2. *Every function locally* DC *on* $\mathbb{R}^n$ *is globally* DC *on* $\mathbb{R}^n$.

Proof. See [6, (I) page 707] or [17, Proposition 4.2].

□

Finding a decomposition for a DC function may not be an easy task, but when it is found, it can generate infinitely many others from it. Another important fact is that any difference between convex functions can become a difference between strongly convex functions, that is, $\phi(x) = g_1(x) - h_1(x)$ be a DC function em $\mathbb{R}^n$ and $g, h : \mathbb{R}^n \to \mathbb{R}$ we can rewrite $\phi(x)$ as

$$\phi(x) = g(x) - h(x);$$

where $g_1(x) + \psi(x) = g(x)$ and $h(x) = h_1(x) + \psi(x)$. We can assume that the term $\psi(x)$ is strongly convex with modulus $\sigma > 0$; in this case, the components DC g, h are strongly convex with modulus $\sigma > 0$, by Proposition1.1.2. The parameter $\sigma > 0$ can be easily chosen, for example, if we define $\psi(x) = \frac{\sigma}{2}\|x\|^2$, which is strongly convex with modulus $\sigma > 0$.

Choosing a DC decomposition whose components strongly convex for to facilitate the theoretical development of some calculations, for example in the analysis of convergence.

Proposition 1.3.3. *Every function* $\phi \in DC(\mathbb{R}^n)$ *is locally Lipschitz.*

Proof. Let $\phi(x) := g(x) - h(x)$ where $g, h : \mathbb{R}^n \to \mathbb{R}$ are its DC components. As g, h are convex functions, by Proposition 1.1.5 are locally Lipchitz. Thus, for all $x_0 \in \mathbb{R}^n$ there exists $\delta_1, \delta_2, K_1, K_2$ such that

$$|g(x) - g(y)| \leqslant K_1 \|x - y\| \quad \forall x, y \in B(x_0, \delta_1),$$

and

$$|h(x) - h(y)| \leqslant K_2 \|x - y\| \quad \forall x, y \in B(x_0, \delta_2).$$

Therefore,

$$|\phi(x) - \phi(y)| = |g(x) - h(x) - g(y) + h(y)| \leqslant |g(x) - g(y)| + |h(x) - h(y)|.$$

For all $x, y \in B(x_0, \delta)$, where $\delta = \min\{\delta_1, \delta_2\}$, we obtain

$$|\phi(x) - \phi(y)| \leqslant (K_1 + K_2)\|x - y\|,$$

in other words, ϕ is locally Lipschitz. □

Lemma 1.3.1. *Let* $g : \mathbb{R}^n \to \mathbb{R}$ *be a continuously differentiable function with gradient Lipschitz continuous with constant* $K > 0$*,* $h : \mathbb{R}^n \to \mathbb{R}$ *be a convex function and* $\phi : \mathbb{R}^n \to \mathbb{R}$ *be given by* $\phi(x) = g(x) - h(x)$*. Then, for all* $y^k, d^k \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$*, there holds*

$$\phi(y^k + \lambda d^k) \leqslant \phi(y^k) + \lambda\langle \nabla g(y^k) - w, d^k\rangle + \frac{K}{2}\lambda^2\|d^k\|^2, \quad \forall w \in \partial h(y^k).$$

Furthermore, if h *is strongly convex with modulus* $\sigma > 0$*, then*

$$\phi(y^k + \lambda d^k) \leqslant \phi(y^k) + \lambda\langle \nabla g(y^k) - w, d^k\rangle + \frac{(K - \sigma)}{2}\lambda^2\|d^k\|^2, \quad \forall w \in \partial h(y^k).$$

Proof. See [5, Lemma 10]. □

Next, we present the necessary conditions for local optimality in DC programming. Let us now relate the concept of Clarke's subdifferential with the difference of DC functions. Consider the function $\phi : \mathbb{R}^n \to \mathbb{R}$, defined by

$$\phi(x) = g(x) - h(x),$$

where g and h are convex functions. By Proposition 1.2.6, we have the inclusion

$$\partial_c \phi(x) \subset \partial_c g(x) - \partial_c h(x).$$

Just consider $S_1 = 1$ and $S_2 = -1$. For hypotheses g and h are convex functions, by remark 1.2.5. We obtain

$$\partial_c \phi(x) \subset \partial g(x) - \partial h(x).$$

Next, we will see that criticality is a weaker condition than Clarke Stationary, this is, $0 \in \partial \phi_c(x)$.

Proposition 1.3.4. *Let $g, h : \mathbb{R}^n \to \mathbb{R}$ be convex functions. If x^* is a minimizer point of $\phi : \mathbb{R}^n \to \mathbb{R}$ given by $\phi(x) = g(x) - h(x), \quad \forall x \in \mathbb{R}^n$, then*

$$\partial h(x^*) \subset \partial g(x^*). \tag{1.1}$$

Where ∂h and ∂g are defined as in Definition 1.2.5, the points satisfying (1.1) are called inf-stationary.

Proof. See [10, Theorem 2]. □

Remark 1.3.1. *Such a condition is not easy to verify and hence a relaxed form has been considered in the DC literature.*

Definition 1.3.1. *Given the function $\phi(x) = g(x) - h(x)$, the point $x^* \in \mathbb{R}^n$ is said to be a critical point when*

$$\partial h(x^*) \cap \partial g(x^*) \neq \emptyset. \tag{1.2}$$

If the point x^* satisfies (1.2)*, then is called a critical point.*

There exist some interesting relationships between inf-stationary, Clarke stationary, and critical points. First, inf-stationarity always implies Clarke stationarity. Second, a Clarke stationary point is a critical point; see [16]. However, for the other way around, these implications do not hold without some extra assumptions. These relationships are summarized in the figure below extracted from [16].

In figure 1.1, the DC function $f : \mathbb{R}^n \to \mathbb{R}$, given by $f(x) = f_1(x) - f_2(x)$, is considered. Note that every inf-stationary point is Clarke point, which in turn is a critical point. The reverse implication, however, depends on the differentiability of the components. When

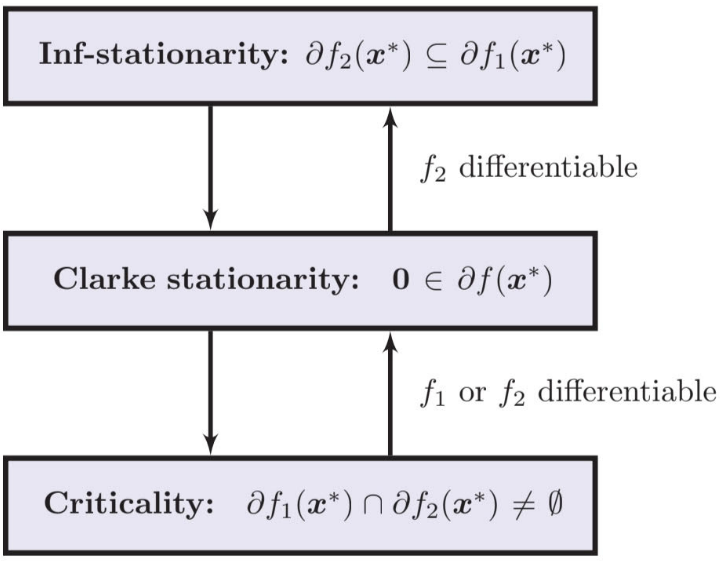


Figure 1.1: Relationship between different stationary concepts [16]

either f_1 or f_2 is differentiable, a critical point implies Clarke stationarity. Moreover, if the point is Clarke and f_2 is differentiable, hence every Clarke point is inf-stationary.

Presented with the formulation of the problem and its range of applications, the need arises to study and develop methods for solving this type, such as the following subgradient-type [14, 26], proximal subgradient [13, 25, 29, 28], double bundle [46], codifferential [43] and inertial method [44]. We will analyze the following methods: the Difference convex Algorithm (DCA), proposed in [11]; the Boosted Difference convex Algorithm (BDCA), proposed in [4]; and the Non-monotone Boosted Difference of Convex Algorithm (nmBDCA), proposed in [5], applied to the enhancement of images with noise, as in [7].

1.4 Floor and Ceiling Functions

Let $\mathbb{Z}$ denote the set of integers and $\mathbb{R}$ the set of real numbers.

Definition 1.4.1. *The **floor function** is the mapping*

$$\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$$

defined by

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} \mid n \leqslant x\}.$$

Definition 1.4.2. *The* ***ceiling function*** *is the mapping*

$$\lceil \cdot \rceil : \mathbb{R} \to \mathbb{Z}$$

defined by

$$\lceil x \rceil = \min\{n \in \mathbb{Z} \mid n \geqslant x\}.$$

Next, we will present some characterizations of inequality involving the concept of floor and ceiling functions.

For all $x \in \mathbb{R}$:

$$\lfloor x \rfloor \leqslant x < \lfloor x \rfloor + 1,$$

$$\lceil x \rceil - 1 < x \leqslant \lceil x \rceil.$$

Moreover:

$$\lceil x \rceil = -\lfloor -x \rfloor.$$

Below we will present some fundamental properties

Proposition 1.4.1. *Regarding the floor and ceiling functions defined above, the following items apply.*

(i) ***Invariance for integers:*** *If $x \in \mathbb{Z}$, then $\lfloor x \rfloor = \lceil x \rceil = x$.*

(ii) ***Integer translation:*** *For all $x \in \mathbb{R}$ and $m \in \mathbb{Z}$:*

$$\lfloor x + m \rfloor = \lfloor x \rfloor + m, \quad \lceil x + m \rceil = \lceil x \rceil + m.$$

(iii) ***Monotonicity:*** *The functions $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ are non-decreasing on $\mathbb{R}$.*

(iv) ***Discontinuity:*** *Both functions are discontinuous exactly at points $x \in \mathbb{Z}$, where they exhibit unit jumps.*

Proof. See [45, Chapter 3]. □

Example 1.4.1.

$$\lfloor 4.7 \rfloor = 4, \quad \lceil 2.7 \rceil = 3,$$

$$\lfloor -2.4 \rfloor = -3, \quad \lceil -2.4 \rceil = -2,$$

$$\lfloor 5 \rfloor = \lceil 5 \rceil = 5.$$

Chapter 2

The Difference of Convex Algorithm

The problem consists in finding a solution to:

$$\min_{x \in \mathbb{R}^n} \phi(x) = g(x) - h(x),$$

where, $g, h : \mathbb{R}^n \to \mathbb{R}$ are a convex functions. For the presentation of the method, it is necessary to consider the following assumptions:

- (AF1) $g, h : \mathbb{R}^n \to \mathbb{R}$ are both strongly convex functions with modulus $\sigma > 0$;
- (AF2) $\phi^* := \inf_{x \in \mathbb{R}^n}\{\phi(x) = g(x) - h(x)\} > -\infty$.

Note that the assumption **(AF1)** is not restrictive. Given two functions g and h, we can add a strongly convex term $\frac{\sigma\|x\|^2}{2}$, we obtain a new decomposition $\phi(x) = (g(x) + \frac{\sigma\|x\|^2}{2}) - (h(x) + \frac{\sigma\|x\|^2}{2})$. Therefore, $\phi(x)$ can be rewritten as the difference of two strongly convex function $\sigma > 0$. The assumption **(AF2)** is common in the context of DC programming; see references [3], [4], [5], [13].

2.1 The Algorithm

The first work related to the difference of convex functions (DC) was addressed by Tao and Souad in [11] to solve non-convex optimization problems. Originally, the DCA was proposed as in **Algorithm 1**, but with a subtle difference: instead of (2.1), in this approach, the subproblem was to find introduced with a subdifferential. The objective was to find $y^k \in \partial g^*(w^k)$, where g^* is the Fenchel conjugate of the function g.

Now, let us present the DCA introduced in [11].

Algorithm 1 Difference of Convex Algorithm (DCA)

Step 1: Choose an initial point $x^0 \in \mathbb{R}^n$ and set $k := 0$;

Step 2: Choose $w^k \in \partial h(x^k)$ and compute y^k the solution of the following convex subproblem

$$y^k = \arg\min_{x \in \mathbb{R}^n} \left\{ g(x) - \langle w^k, x - x^k \rangle \right\}; \tag{2.1}$$

Step 3: If $y^k = x^k$, then stop and return x^k. Otherwise, set $x^{k+1} = y^k$ and go back to Step 2 with $k := k + 1$.

Remark 2.1.1. *Note that finding a solution to the subproblem in* (2.1) *is equivalent to finding* $y^k \in \partial g^*(w^k)$. *By Proposition 1.1.4, we have that* $w^k \in \partial g(y^k)$ *this is* $0 \in \partial g(y^k) - w^k$, *in other words,*

$$y^k = \arg\min_{x \in \mathbb{R}^n} \{g(x) - \langle w^k, x - x^k \rangle\}.$$

Remark 2.1.2. *We will study whether Algorithm 1 is well defined. In fact, since* h *is convex, by Proposition 1.2.1,* $\partial h(x)$ *is non-empty* ($\partial h(x) \neq \emptyset$). *Since* g *is strongly convex, we see that the subproblem in* (2.1) *has a solution by Theorem 1.1.1. Therefore, the construction of Algorithm 1 is well defined.*

The objective of the comment is to explain the motivation and the form of the approximation subfunction used in the DCA. Observe that the problem of the minimizing the difference between the functions $g(x)$ and $\langle w^k, x - x^k \rangle$ from the equation in the *Step 2* can be seen as:

$$w^k \in \partial h(x^k),$$

by definition, $h(x) \geqslant h(x^k) + \langle w^k, x - x^k \rangle$, then

$$-h(x) \leqslant -h(x^k) - \langle w^k, x - x^k \rangle.$$

In other words, $g(x) - h(x) \leqslant g(x) - h(x^k) - \langle w^k, x - x^k \rangle$. Therefore,

$$\phi(x) = g(x) - h(x) \leqslant g(x) - \langle w^k, y - x^k \rangle - h(x^K) \quad \forall x \in \mathbb{R}^n.$$

Ignoring the constant term, the right side of the inequality coincides with the subproblem in (2.1).

2.2 Convergence Analysis DCA

In this section, we will study the convergence of the DCA.

Proposition 2.2.1. *The sequence* $\{x^k\}_{k\in\mathbb{N}}$ *generated by* ***Algorithm 1*** *satisfies one of the following properties.*

(i) The Algorithm 1 stops at a critical point.

(ii) The sequence $\{\phi(x^k)\}_{k\in\mathbb{N}}$ *is decreasing.*

Proof.

(i) If $y^k = x^k$ for any $k \in \mathbb{N}$, then x^k is critical point. By Remark 2.1.1, suppose that y^k is solution for the subproblem, so $w^k \in \partial g(y^k)$. On the other hand, if $w^k \in \partial h(x^k)$ as assumed, and $x^k = y^k$, then

$$w^k \in \partial g(x^k) \cap \partial h(x^k).$$

In other words,

$$\partial g(x^k) \cap \partial h(x^k) \neq \emptyset.$$

Therefore, x^k is a point critical.

(ii) Now, consider $x^{k+1} \neq x^k$. Then by Step 2 e Step 3, there exists $w^k \in \partial g(x^{k+1}) \cap \partial h(x^k)$, for all $k \in \mathbb{N}$. Since that g and h are strongly convex, it follows that:

$$g(x^k) \geqslant g(x^{k+1}) + \langle w^k, x^k - x^{k+1}\rangle + \frac{\sigma}{2}\left\|x^k - x^{k+1}\right\|^2 \tag{2.2}$$

and

$$h(x^{k+1}) \geqslant h(x^k) + \langle w^k, x^{k+1} - x^k\rangle + \frac{\sigma}{2}\left\|x^{k+1} - x^k\right\|^2. \tag{2.3}$$

Adding inequalities (2.2) and (2.3), we obtain:

$$g(x^k) + h(x^{k+1}) \geqslant g(x^{k+1}) + h(x^k) + \langle w^k, x^k - x^{k+1}\rangle - \langle w^k, x^k - x^{k+1}\rangle + \sigma\left\|x^k - x^{k+1}\right\|^2.$$

In other words,

$$g(x^k) - h(x^k) \geqslant g(x^{k+1}) - h(x^{k+1}) + \sigma\left\|x^k - x^{k+1}\right\|^2.$$

As $\left\|x^k - x^{k+1}\right\|^2 > 0$, we obtain

$$\phi(x^k) \geqslant \phi(x^{k+1}) + \sigma\left\|x^k - x^{k+1}\right\|^2 > \phi(x^{k+1}).$$

Therefore, $\phi(x^{k+1}) < \phi(x^k) \quad \forall k \in \mathbb{N}$, which this we conclude that the sequence $\{\phi(x^k)\}_{k\in\mathbb{N}}$ is decreasing. □

Proposition 2.2.2. *Let* $\{x^k\}_{k\in\mathbb{N}}$ *a sequence generated by the Algorithm 1, then* $\{\phi(x^k)\}_{k\in\mathbb{N}}$ *is convergent.*

Proof. As seen in the previous Proposition 2.2.1, ϕ is decreasing. Moreover, $\phi(x^k)$ is bounded below, by hypothesis AF2. Therefore, the sequence $\{\phi(x^k)\}_{k\in\mathbb{N}}$ converges by the Monotone Convergence Theorem. □

Proposition 2.2.3. *Let* $\{x^k\}_{k\in\mathbb{N}}$ *a sequence generated by the* ***Algorithm 1****, then*

$$\sum_{k=0}^{+\infty} \left\|x^k - x^{k+1}\right\|^2 < +\infty.$$

In particular, $\left\|x^k - x^{k+1}\right\| \to 0$ *as* $k \to +\infty$.

Proof. We know that by Proposition 2.2.1, item (ii), one must have:

$$\phi(x^k) \geqslant \phi(x^{k+1}) + \sigma\left\|x^k - x^{k+1}\right\|^2.$$
$$0 \leqslant \sigma\left\|x^k - x^{k+1}\right\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}).$$

We obtain,

$$0 \leqslant \sum_{k=0}^{I} \sigma\left\|x^k - x^{k+1}\right\|^2 \leqslant \sum_{k=0}^{I} \phi(x^k) - \phi(x^{k+1}).$$

Note that

$$\sum_{k=0}^{I} \phi(x^k)-\phi(x^{k+1}) = \phi(x^0)-\phi(x^1)+\phi(x^1)-\phi(x^2)-\cdots+\phi(x^I)-\phi(x^{I+1}) = \phi(x^0)-\phi(x^{I+1}).$$

Then

$$0 \leqslant \sum_{k=0}^{I} \sigma\left\|x^k - x^{k+1}\right\|^2 \leqslant \phi(x^0) - \phi(x^{I+1}).$$

Remember by the AF2 hypothesis, $\phi^* := \inf_{x\in\mathbb{R}^n}\{\phi(x) = g(x) - h(x)\} > -\infty$, it follows that

$$\phi^* \leqslant \phi(x^{I+1}).$$

Hence,

$$-\phi^* \geqslant -\phi(x^{I+1}).$$

Therefore,

$$0 \leqslant \sum_{k=0}^{I} \sigma\left\|x^k - x^{k+1}\right\|^2 \leqslant \phi(x^0) - \phi^*.$$

When $I \to +\infty$, we have to

$$0 \leqslant \sum_{k=0}^{+\infty} \sigma \left\| x^k - x^{k+1} \right\|^2 \leqslant +\infty.$$

In particular,

$$\sum_{k=0}^{+\infty} \left\| x^k - x^{k+1} \right\| \leqslant +\infty,$$

in other words, is convergent. Then $\left\| x^k - x^{k+1} \right\| \to 0$ when $k \to +\infty$. This is,

$$\lim_{k \to +\infty} \left\| x^k - x^{k+1} \right\| = 0.$$

□

Remark 2.2.1. *Note that the entire convergence analysis is demonstrated starting from the hypothesis that the components of $\phi(x)$, namely $g(x)$ and $h(x)$, are strongly convex. Otherwise, we cannot guarantee that the sequence $\{\left\| x^k - x^{k+1} \right\|\}$ converges to 0, see next example.*

Example 2.2.1. *[38, Example 2.2.1] Consider $g, h : \mathbb{R} \to \mathbb{R}$ as convex functions defined as follows: $g(x) = \sup\{-x, 0, x-1\}$ and $h(x) = \sup\{0, -x\}$. The functions g and h are convex, but neither strongly convex nor strictly convex. Starting DCA from the initial $x^0 \in (0, 1)$, for example, $x^0 := 0.1$, note that $w^0 \in \partial h(0.1) = \{0\}$, choose $w^0 = 0$, so $x^1 \in \arg\min\{g(x) - \langle w^0, x - x^0 \rangle\} = [0, 1]$ implies that choosing $x^1 = 0.9$, compute $w^1 \in \partial h(x^1) = 0$. Setting $x^2 = 0.1$, here it goes, a sequence possible for DCA is (0.1, 0.9, 0.1, 0.9, ...), where $\|x^{k+1} - x^k\|$ is a constant sequence (0.8, 0.8, 0.8,...) that obviously does not converge to zero. it does not satisfy Proposition 2.2.3. Then, $\{\phi(x^k)\} = \{g(x^k) - h(x^k)\}$ is a constant equal to zero and cannot be decreasing.*

Theorem 2.2.1. *Any cluster point of the sequence generated by* ***Algorithm 1*** *is a critical point.*

Proof. Let x^* be a cluster point of the sequence $\{x^k\}$. Then there exists a subsequence $\{x^{k_j}\}$ such that $\lim_{j\to\infty} x^{k_j} = x^*$, in other words, $\{x^{k_j}\}$ is bounded. Let us prove that $\lim_{j\to\infty} x^{k_j+1} = x^*$. Being $\{x^k\}$ a sequence, applying Proposition 2.2.3, we obtain $\|x^{k_j+1} - x^{k_j}\| \to 0$. Note that

$$\|x^{k_j+1} - x^*\| = \| - x^{k_j} + x^{k_j+1} + x^{k_j} - x^*\| \leqslant \| - x^{k_j} + x^{k_j+1}\| + \|x^{k_j} - x^*\|, \quad \forall j \in \mathbb{N}.$$

When $j \to \infty$, we obtain $\|x^{k_j+1} - x^*\| \to 0$; then $\lim_{j\to\infty} x^{k_j+1} = x^*$, by Proposition 1.2.3, there exists $w^{k_j} \in \partial h(x^{k_j})$, for all $j \in \mathbb{N}$. The subproblem (2.1) with $k = k_j$, can be rewritten as:

$$w^{k_j} \in \partial g(x^{x_j+1}).$$

Assume without loss of generality that $\{w^{k_j}\}$ is convergent. Thus,

$$w^{k_j} \in \partial g(x^{k_j+1}) \cap \partial h(x^{k_j}), \quad \forall j \in \mathbb{N}.$$

Again, by Proposition 1.2.3 same cluster point of $\{w^{k_j}\} \in \partial g(x^*) \cap \partial h(x^*), \quad \forall j \in \mathbb{N}$

$$\partial g(x^*) \cap \partial h(x^*) \neq \emptyset.$$

Thus, x^* is a critical point. □

Chapter 3

The Boosted Difference of Convex Algorithm

It was initially proposed by Aragón Artacho et al. [3] as a method that accelerated DCA to solve the following problem. The problem consists of finding the solution to the problem:

$$\min_{x \in \mathbb{R}^n} \phi(x) = g(x) - h(x),$$

In the first version of The Boosted Difference of Convex Algorithm (BDCA), both components g and h had to be differentiable and convex. However, in the updated version also proposed by Aragón Artacho and Voung in [4], only the component g is differentiable. This update allows the method to be extended to a class of non-differentiable functions. Therefore, BDCA accelerates the DCA by performing a linear search. For the presentation of the method, it is necessary to consider the following assumptions:

- (AF1) $g, h : \mathbb{R}^n \to \mathbb{R}$ are both strongly convex functions with modulus $\sigma > 0$;
- (AF2) $\phi^* := \inf_{x \in \mathbb{R}^n}\{\phi(x) = g(x) - h(x)\} > -\infty$;
- (AF3) $g : \mathbb{R}^n \to \mathbb{R}$ is differentiable.

The conditions **(AF1)** and **(AF2)** are reasonable and were previously addressed in Algorithm 1; **(AF3)** is important because it ensures the existence of descent directions. This is a fundamental requirement for the convergence of the method.

3.1 The Algorithm

The Boosted Difference of Convex Algorithm (BDCA) was proposed by Aragón Artacho and Voung [4]. It was generated from the method proposed by Aragón Artacho in [3]. Now, the BDCA will be presented as **Algorithm 2**.

Algorithm 2 Boosted Difference of Convex Algorithm (BDCA)

Step 1. Fix $\rho > 0$ and $\beta \in (0,1)$. Choose an initial point $x^0 \in \mathbb{R}^n$ and set $k := 0$;

Step 2. Choose $w^k \in \partial h(x^k)$ and compute y^k the solution for the following convex subproblem

$$y^k = \arg\min_{x\in\mathbb{R}^n}\{g(x) - \langle w^k, x - x^k\rangle\}; \tag{3.1}$$

Step 3. Set $d^k = y^k - x^k$. If $d^k = 0$, then STOP and return x^k. **Otherwise**, choose any $\overline{\lambda}_k \geqslant 0$, and set $\lambda_k := \overline{\lambda}_k$. WHILE $\phi(y^k + \lambda_k d^k) > \phi(y^k) - \rho\lambda_k^2\|d^k\|^2$ DO $\lambda_k := \beta\lambda_k$.

Step 4. Set $x^{k+1} := y^k + \lambda_k d^k$; set $k := k+1$ and go to Step 2.

See that when $\overline{\lambda}_k = 0$, the iterations of BDCA coincide with those of DCA. We will show that d^k is a descent direction, because when we take $x^{k+1} = y^k$, the image ϕ is decreasing. If we take x^{k+1} in the descent direction of d^k from y^k, we obtain an even greater decrease. This is the idea behind BDCA. Moreover, this fact improves the performance of DCA in many applications.

Remark 3.1.1. *The well definition of the Step 2 follows similarly to the Remark 2.1.2. To complete the well definition of **Algorithm 2**, let us consider the following proposition.*

Proposition 3.1.1. *For all $k \in \mathbb{N}$, the following holds:*

(i) $\phi(y^k) \leqslant \phi(x^k) - \sigma\|d^k\|^2$;

(ii) $\phi'(y^k; d^k) \leqslant -\sigma\|d^k\|^2$;

(iii) There is some $\delta_k > 0$ such that

$$\phi(y^k + \lambda d^k) \leqslant \phi(y^k) - \alpha\lambda^2\|d^k\|^2, \quad \forall\lambda \in (0, \delta_k].$$

Proof.

(i) Consider y^k as the unique solution for the problem (3.1), then $\nabla(g(x) - \langle w^k, x - x^k\rangle)(y^k) = 0$, this is, $\nabla g(y^k) - w^k = 0$, or equivalently, $\nabla g(y^k) = w^k$, since $w^k \in \partial h(x^k)$. By definition de the strong convex of g and h, we have:

$$g(x^k) - g(y^k) \geqslant \langle w^k, x^k - y^k\rangle + \frac{\sigma}{2}\|x^k - y^k\|^2,$$

and

$$h(y^k) - h(x^k) \geqslant \langle w^k, y^k - x^k \rangle + \frac{\sigma}{2}\left\|y^k - x^k\right\|^2.$$

Adding both inequalities yields the desired result

$$g(x^k) - h(x^k) - [g(y^k) - h(y^k)] \geqslant \sigma\left\|y^k - x^k\right\|^2.$$

This is,

$$\phi(x^k) - \phi(y^k) \geqslant \sigma\left\|y^k - x^k\right\|^2.$$

By definition $d^k = y^k - x^k$, then

$$\phi(x^k) \geqslant \phi(y^k) + \sigma\left\|d^k\right\|^2.$$

Hence,

$$\phi(y^k) \leqslant \phi(x^k) - \|d^k\|^2.$$

(ii) It directly follows from the definition of directional derivative that:

$$\begin{aligned}
\phi'(y^k; d^k) &= \lim_{\lambda\downarrow 0} \frac{\phi(y^k + \lambda d^k) - \phi(y^k)}{\lambda} \\
&= \lim_{\lambda\downarrow 0} \frac{g(y^k + \lambda d^k) - h(y^k + \lambda d^k) - g(y^k) + h(y^k)}{\lambda} \\
&= \lim_{\lambda\downarrow 0} \frac{g(y^k + \lambda d^k) - g(y^k)}{\lambda} - \lim_{\lambda\downarrow 0} \frac{h(y^k + \lambda d^k) - h(y^k)}{\lambda}. \qquad (3.2)
\end{aligned}$$

As g is differentiable and convex and h is convex, consider $v \in \partial h(y^k)$. We have:

$$\lim_{\lambda\downarrow 0} \frac{g(y^k + \lambda d^k) - g(y^k)}{\lambda} = \langle \nabla g(y^k), d^k \rangle,$$

and due to the subdifferential inequality:

$$h(y^k + \lambda d^k) - h(y^k) \geqslant \langle v, \lambda d^k \rangle,$$

what implies:

$$\lim_{\lambda\downarrow 0} \frac{h(y^k + \lambda d^k) - h(y^k)}{\lambda} \geqslant \lim_{\lambda\downarrow 0} \frac{\lambda \langle v, d^k \rangle}{\lambda},$$

or equivalently

$$\lim_{\lambda\downarrow 0} \frac{h(y^k + \lambda d^k) - h(y^k)}{\lambda} \geqslant \langle v, d^k \rangle,$$

or

$$-\lim_{\lambda\downarrow 0} \frac{h(y^k + \lambda d^k) - h(y^k)}{\lambda} \leqslant -\langle v, d^k \rangle.$$

Applying this results in (3.2), we obtain:

$$\begin{aligned}\phi'(y^k; d^k) &\leqslant \langle \nabla g(y^k), d^k\rangle - \langle v, d^k\rangle \\ &= \langle \nabla g(y^k) - v, d^k\rangle \\ &= \langle \nabla g(y^k) - v, y^k - x^k\rangle.\end{aligned}$$

Therefore,

$$\phi'(y^k; d^k) \leqslant \langle \nabla g(y^k) - v, y^k - x^k\rangle.$$

We recall $\nabla g(y^k) = w^k \in \partial h(x^k)$ and the function h is strongly convex, by Theorem 1.2.4, ∂h is strongly monotone with modulus $\sigma > 0$. Since $v \in \partial h(y^k)$, we have:

$$\langle \partial h(y^k) - \partial h(x^k), y^k - x^k\rangle \geqslant \sigma\|y^k - x^k\|^2$$

which implies

$$\langle v - \nabla g(y^k), y^k - x^k\rangle \geqslant \sigma\|y^k - x^k\|^2.$$

Hence

$$\langle \nabla g(y^k) - v, y^k - x^k\rangle \leqslant -\sigma\|y^k - x^k\|^2.$$

That said,

$$\phi'(y^k; d^k) \leqslant \langle \nabla g(y^k) - v, y^k - x^k\rangle \leqslant -\sigma\|y^k - x^k\|^2,$$

in other words,

$$\phi'(y^k; d^k) \leqslant -\sigma\|y^k - x^k\|^2.$$

(iii) When $d^k = 0$ or $\lambda = 0$, then the proof is obvious. Otherwise, we have by definition of directional derivative

$$\phi'(y^k; d^k) = \lim_{\lambda\downarrow 0} \frac{\phi(y^k + td^k) - \phi(y^k)}{\lambda} \leqslant -\sigma\|d^k\|^2 \leqslant -\frac{\sigma}{2}\|d^k\|^2.$$

Hence, there is $\overline{\lambda}_k$ such that

$$\frac{\phi(y^k + \lambda d^k) - \phi(y^k)}{\lambda} \leqslant -\frac{\sigma}{2}\|d^k\|^2 \quad \forall\lambda \in (0, \overline{\lambda}_k].$$

This is

$$\phi(y^k + \lambda d^k) - \phi(y^k) \leqslant -\frac{\sigma\lambda}{2}\|d^k\|^2 \quad \forall\lambda \in (0, \overline{\lambda}_k].$$

Note that,

$$\phi(y^k + \lambda d^k) \leqslant \phi(y^k) - \frac{\sigma\lambda}{2}\|d^k\|^2 \leqslant \phi(y^k) - \alpha\lambda^2\left\|d^k\right\|^2.$$

Then,

$$\frac{\sigma}{2} \geqslant \alpha\lambda \quad \text{or} \quad \lambda \leqslant \frac{\sigma}{2\alpha}.$$

Therefore, setting $\delta_k = \min\{\overline{\lambda}_k, \frac{\sigma}{2\alpha}\}$, we obtain:

$$\phi(y^k + \lambda d^k) \leqslant \phi(y^k) - \alpha\lambda^2\left\|d^k\right\|^2 \quad \forall\lambda \in (0, \delta_k].$$

□

The following example shows the importance of g being differentiable for the convergence of the method.

Example 3.1.1. *([4], Example 3.4) Let* $\phi : \mathbb{R}^2 \to \mathbb{R}$ *be a DC function as DC components* $g(x_1, x_2) = -\frac{5}{2}x_1 + x_1^2 + x_2^2 + |x_1| + |x_2|$ *and* $h(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2)$.
Note that, since h *is differentiable, but* g *is not differentiable. Consider the point* $x^0 = (\frac{1}{2}, 1)$*, so the* $w^0 = \partial h(\frac{1}{2}, 1) = \nabla h(\frac{1}{2}, 1) = (\frac{1}{2}, 1)$*, the next iteration generated by DCA is* $y^0 = (1, 0)$*, because* $w^0 = \partial g(y^0)$*, implies that,* $(\frac{1}{2}, 1) = (-\frac{5}{2} + 2x_1 + 1, 2x_2 + 1)$*. When* $(x_1, x_2) > (0, 0)$*, so* $d^0 = y^0 - x^0 = (\frac{1}{2}, -1)$*. Now, we assert that* d^0 *is not a descent direction from* y^0*. In fact,*

$$\phi'(y^0; d^0) = \lim_{\lambda\downarrow 0} \frac{\phi(y^0 + \lambda d^0) - \phi(y^0)}{\lambda} = \lim_{\lambda\downarrow 0} \frac{3}{4}.$$

Thus, d^0 *is not a descent direction for* ϕ *at* y^0*. In addition,* $\phi(y^0) = -1$ *and* $\phi(y^0+\lambda d^0) = -1 + \dfrac{3}{4}\lambda + \dfrac{5}{8}\lambda^2$*, we conclude that*

$$\phi(y^0 + \lambda d^0) > \phi(y^0), \quad \forall\lambda > 0.$$

Therefore, d^0 *is a ascent direction and a monotone line search cannot be performed.*

3.2 Convergence Analysis

Proposition 3.2.1. *The sequences* $\{x^k\}_{k\in\mathbb{N}}$ *generated by* ***Algorithm 2*** *satisfy one of the following properties:*

(i) The ***Algorithm 2*** *stops at a critical point;*

(ii) The sequence $\{\phi(x^k)\}_{k\in\mathbb{N}}$ *is decreasing.*

Proof.

(i) If $y^k = x^k$ for any $k \in \mathbb{N}$, then x^k is a critical point. Since g is differentiable,by Proposition 1.2.2 and given that y^k is a solution to the expression in (3.1), this is, it can be expressed as:

$$0 = \nabla(g(\cdot) - \langle w^k, \cdot - x^k\rangle)(y^k).$$

This is,

$$\nabla g(y^k) = w^k.$$

By step 2, we have $w^k \in \partial h(x^k)$. Consequently, if $x^k = y^k$ its follows that:

$$\partial h(x^k) \cap \partial g(x^k) = \{\nabla g(x^k)\}.$$

Thus, x^k is a critical point of the subproblem.

(ii) Follows from the *items (i)* and *(iii)* of the Proposition 3.1.1, along with Step 3 of the Algorithm 2:

$$\phi(y^k + \lambda_k d^k) \leqslant \phi(y^k) - \alpha\lambda_k^2 \left\|d^k\right\|^2 \leqslant \phi(x^k) - \sigma\|d^k\|^2 - \alpha\lambda_k^2\|d^k\|^2.$$

In other words,

$$\phi(y^k + \lambda_k d^k) \leqslant \phi(x^k) - \|d^k\|^2(\lambda_k^2\alpha + \sigma), \quad \forall k \in \mathbb{N}.$$

Since that $d^k \neq 0$, and $\sigma, \alpha > 0$ and $\lambda_k \geqslant 0$. Consider, $y^k + \lambda_k d^k = x^{k+1}$, we have:

$$\phi(x^{k+1}) < \phi(x^k), \quad \forall k \in \mathbb{N}.$$

□

Proposition 3.2.2. *Let* $\{x^k\}_{k\in\mathbb{N}}$ *be a sequence generated by* ***Algorithm 2****, then* $\{\phi(x^k)\}_{k\in\mathbb{N}}$ *is convergent.*

Proof. As seen in the previous proposition, ϕ is decreasing. Moreover, $\phi(x^k)$ is bounded below, by hypothesis AF2. Therefore, $\{\phi(x^k)\}_{k\in\mathbb{N}}$ converges. □

Proposition 3.2.3. *If* $\{x^k\}_{k\in\mathbb{N}}$ *the sequence generated by* ***Algorithm 2****. The following statements are valid.*

(i) $\sum_{k=0}^{+\infty} \|d^k\|^2 < +\infty$*, so* $\|y^k - x^k\| \to 0$ *when* $k \to +\infty$*;*

(ii) $\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty$*, so* $\|x^{k+1} - x^k\| \to 0$ *when* $k \to +\infty$*.*

Proof.

(i) By Proposition 3.1.1 and Step 3 of Algorithm 2, we have:

$$\phi(y^k + \lambda_k d^k) \leqslant \phi(x^k) - \|d^k\|^2(\lambda_k^2 \alpha + \sigma), \quad \forall k \in \mathbb{N},$$

thus,

$$\|d^k\|^2(\lambda_k^2 \alpha + \sigma) \leqslant \phi(x^k) - \phi(x^{k+1}), \quad \forall k \in \mathbb{N}.$$

Considered $\alpha \lambda_k \geqslant 0$, then

$$\sigma|d^k\|^2 \leqslant \|d^k\|^2(\lambda_k^2 \alpha + \sigma).$$

Hence,

$$\sigma\|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}).$$

Considered the partial sum

$$\sum_{k=0}^{N} \sigma\|d^k\|^2 \leqslant \sum_{k=0}^{N} \phi(x^k) - \phi(x^{k+1}).$$

This implies that,

$$\sum_{k=0}^{N} \sigma|d^k\|^2 \leqslant \phi(x^0) - \phi(x^{N+1}).$$

By hypotheses AF2, $\phi^* = \inf \phi(x)$, so $\phi^* < \phi(x^{N+1})$ or $-\phi^* > -\phi(x^{N+1})$. Therefore,

$$\sum_{k=0}^{N} \sigma\|d^k\|^2 \leqslant \phi(x^0) - \phi^*, \quad \forall k \in \mathbb{N}.$$

Since $\sigma > 0$ is constant. When $N \to +\infty$, we have:

$$\sum_{k=0}^{+\infty} \sigma\|d^k\|^2 \leqslant +\infty.$$

In particular,

$$\sum_{k=0}^{+\infty} \|d^k\|^2 \leqslant +\infty.$$

Hence, $\|y^k - x^k\| \to 0$ when $k \to +\infty$. This proves *item (i)*.

(ii) Consider $\tilde{\lambda} > 0$ be such that $0 \leqslant \lambda_k \leqslant \tilde{\lambda}, \quad \forall k \in \mathbb{N}$. Now, observe that

$$\begin{aligned}
\|x^{k+1} - x^k\|^2 = \|y^k + \lambda_k d^k - x^k\|^2 &= \|y^k + \lambda_k(y^k - x^k) - x^k\|^2 \\
&= \|y^k(1+\lambda_k) - x^k(1+\lambda_k)\|^2 \\
&= \|(y^k - x^k)(1+\lambda_k)\|^2 \\
&= \|(y^k - x^k)\|^2\|(1+\lambda_k)\|^2 \\
&\leqslant (1+\lambda)^2\|d^k\|^2, \quad \forall k \in \mathbb{N}.
\end{aligned}$$

Its follows by *item (i)* and inequality above that

$$0 < \sum_{k=0}^{+\infty} \|x^{k+1} - x^k\|^2 \leqslant \sum_{k=0}^{+\infty} (1+\lambda)^2 \|d^k\|^2 \leqslant +\infty.$$

Therefore, $\|x^{k+1} - x^k\| \to 0$ when $k \to +\infty$. □

Theorem 3.2.1. *Any cluster point of the sequence generated by **Algorithm 2** is a critical point.*

Proof. Let x^* be a cluster point of the sequence $\{x^k\}$. Then there exists a subsequence $\{x^{k_j}\}$ such that $\lim_{j\to\infty} x^{k_j} = x^*$, that is, $\lim_{j\to\infty} \|x^{k_j} - x^*\| = 0$. Given a $\{y^k\}_{k\in\mathbb{N}}$ sequence generated by Algorithm 2, we have $\lim_{k\to\infty} \|y^k - x^k\| = 0$. We have:

$$\|y^{k_j} - x^*\| = \|y^{k_j} - x^{k_j} + x^{k_j} - x^*\| \leqslant \|y^{k_j} - x^{k_j}\| + \|x^{k_j} - x^*\|, \quad \forall j \in \mathbb{N},$$

which implies that $\lim_{j\to+\infty} \|y^{k_j} - x^*\| = 0$, in other words, $\lim_{j\to+\infty} y^{k_j} = x^*$. It follows from Step 2, y^{k_j} is the solution of the subproblem, then $\nabla g(y^{k_j}) = w^{k_j} \in \partial h(x^{k_j})$, for all $j \in \mathbb{N}$. Suppose that $\lim_{j\to+\infty} w^{k_j} = w$, by Proposition1.2.1 and Corollary 1.2.1, it follows that

$$\lim_{j\to+\infty} \nabla g(y^{k_j}) = \nabla g(x^*) = \lim w^{k_j} = w \in \partial h(x^*).$$

In other words,

$$\nabla g(x^*) \in \partial h(x^*).$$

Therefore, x^* is a critical point. □

Chapter 4

Non-Monotone Boosted Difference of Convex Algorithm

In the previous chapter, we presented the BDCA, where we considered that the first component g was strongly convex and differentiable so that we could ensure that the direction considered by **Algorithm 2** was a descent direction.

Now, let us study the method proposed by Ferreira, Santos, and Souza in [5], in which they proposed a non-monotonic search in the BDCA (nmBDCA) to enable control growth in the objective function values by means of the parameter.

In this chapter, we will assume that the statements have already been seen in previous chapters, namely:

- (AF1) $g, h : \mathbb{R}^n \to \mathbb{R}$ are both strongly convex functions with modulus $\sigma > 0$;
- (AF2) $\phi^* := \inf_{x \in \mathbb{R}^n}\{\phi(x) = g(x) - h(x)\} > -\infty$.

The statements **(AF1)** and **(AF2)** are the same as those adopted in Chapter 2 where they were analyzed and discussed.

4.1 The algorithm

When we nullify the hypothesis that g is differentiable, we cannot guarantee the existence of descent directions that decreases the objective function. So, the main idea of the Non-monotone Boosted Difference of Convex Algorithm (nmBDCA) is that the growth of the objective function is allowed, controlled by a parameter ν_k that will be specified shortly.

Algorithm 3 Non-monotone Boosted Difference of Convex Algorithm (nmBDCA)

Step 1. Fix $\rho > 0$ and $\beta \in (0,1)$. Choose an initial point $x^0 \in \mathbb{R}^n$ and set $k := 0$.

Step 2. Choose $w^k \in \partial h(x^k)$ and compute y^k the solution for the following convex subproblem

$$y^k = \arg\min_{x\in\mathbb{R}^n} g(x) - \langle w^k, x - x^k\rangle. \tag{4.1}$$

Step 3. Set $d^k = y^k - x^k$. If $d^k = 0$, then STOP and return x^k. **Otherwise**, choose any $\overline{\lambda}_k \geqslant 0$, and set $\lambda_k := \overline{\lambda}_k$. WHILE $\phi(y^k + \lambda_k d^k) > \phi(y^k) - \rho\lambda_k^2\|d^k\|^2 + \nu_k$ DO $\lambda_k := \beta\lambda_k$.

Step 4. Set $x^{k+1} := y^k + \lambda_k d^k$; set $k := k+1$ and go to step 2.

Proposition 4.1.1. *For each* $k \in \mathbb{N}$*, the following statements hold:*

(i) If $d^k = 0$*, then* x^k *is a critical point;*

(ii) There holds $\phi(y^k) \leqslant \phi(x^k) - \sigma\|d^k\|^2$.

Proof. The Proof is equivalent to what is presented in Proposition 3.1.1. □

The well-definedness of nmBDCA, we present in the next sections, by assuming that g is possibly non differentiable and then g is continuously differentiable.

4.2 Well-definedness of nmBDCA: g is possibly non-differentiable

Proposition 4.2.1. *Let* $\{(x^k)\}_{k\in\mathbb{N}}$ *be the sequence generated by Algorithm 3. For each* $k \in \mathbb{N}$ *assume that* $d^k \neq 0$ *and* $\nu_k > 0$*. Then, the following statements hold:*

(i) Consider $\hat{\delta}_k := \frac{\nu_k}{g(y^k+d^k)+g(x^k)-2g(y^k)} > 0$*, then*

$$\phi(y^k + \lambda d^k) \leqslant \phi(y^k) - \rho\lambda^2\|d^k\|^2 + \nu_k, \quad \forall\lambda \in (0, \overline{\delta}_k],$$

where $\overline{\delta}_k := \min\{\hat{\delta}_k, 1, \frac{3\sigma}{2\rho}\}$*. Therefore, the line search in Step 3 is well defined.*

(ii) $\phi(y^k + \lambda d^k) \leqslant \phi(x^k) - (\sigma + \rho\lambda_k^2)\|d^k\|^2 + \nu_k$.

Proof. We recall that $d^k = y^k - x^k$. Consider $d^k \neq 0$ and take $w^k \in \partial h(x^k)$. Since h is strongly convex with modulus $\sigma > 0$, by Theorem 1.2.3. We obtain:

$$h(y^k + \lambda d^k) \geqslant h(y^k) + \langle s, \lambda d^k \rangle + \lambda^2 \frac{\sigma}{2} \|d^k\|^2, \quad \forall s \in \partial h(y^k). \tag{4.2}$$

Now, note that $w^k \in \partial h(x^k)$. Applying the Theorem 1.2.4, we obtain:

$$\langle s - w^k, y^k - x^k \rangle \geqslant \sigma \|y^k - x^k\|^2.$$

Thus,

$$\langle s, d^k \rangle - \langle w^k, d^k \rangle \geqslant \sigma \|d^k\|^2,$$

and hence,

$$\langle s, d^k \rangle \geqslant \langle w^k, d^k \rangle + \sigma \|d^k\|^2.$$

Replacing in (4.2) we have,

$$h(y^k + \lambda d^k) \geqslant h(y^k) + \lambda \langle w^k, d^k \rangle + \lambda \sigma \|d^k\|^2 + \lambda^2 \frac{\sigma}{2} \|d^k\|^2.$$

Considering y^k solution for the subproblem of the Algorithm 3, we have:

$$g(y^k) - \langle w^k, y^k - x^k \rangle \leqslant g(x^k) - \langle w^k, x^k - x^k \rangle.$$

This implies that $-\langle w^k, y^k - x^k \rangle \leqslant g(x^k) - g(y^k)$, which replacing in (4.2):

$$-h(y^k + \lambda d^k) \leqslant -h(y^k) + \lambda \left(g(x^k) - g(y^k) \right) - \lambda \sigma \|d^k\|^2 - \lambda^2 \frac{\sigma}{2} \|d^k\|^2. \tag{4.3}$$

On the other hand, by using the strong convexity of g with modulus $\sigma > 0$, we have:

$$\begin{aligned} g(y^k + \lambda d^k) - g(y^k) &= g(\lambda(y^k + d^k) + (1-\lambda) y^k) - g(y^k) \\ &\leqslant \lambda g(y^k + d^k) + (1-\lambda) g(y^k) - \lambda(1-\lambda) \frac{\sigma}{2} \|d^k\|^2 - g(y^k) \\ &= \lambda \left(g(y^k + d^k) - g(y^k) \right) - \lambda(1-\lambda) \frac{\sigma}{2} \|d^k\|^2, \end{aligned} \tag{4.4}$$

$\forall \lambda \in [0, 1]$. By definition of ϕ with (4.4) and (4.3), we obtain

$$\begin{aligned} \phi(y^k + \lambda d^k) - \phi(y^k) &= g(y^k + \lambda d^k) - h(y^k + \lambda d^k) - g(y^k) + h(y^k) \\ &= g(y^k + \lambda d^k) - g(y^k) - \left(h(y^k + \lambda d^k) - h(y^k) \right) \\ &\leqslant \lambda \left(g(y^k + d^k) - g(y^k) \right) - \lambda(1-\lambda) \frac{\sigma}{2} \|d^k\|^2 \\ &\quad + \lambda \left(g(x^k) - g(y^k) \right) - \sigma \lambda \|d^k\|^2 - \lambda^2 \frac{\sigma}{2} \|d^k\|^2 \\ &\leqslant -\frac{3\sigma}{2} \lambda \|d^k\|^2 + \lambda \left(g(y^k + d^k) + g(x^k) - 2 g(y^k) \right). \end{aligned} \tag{4.5}$$

Moreover, it follows from Theorem 1.2.3 that for all $w \in \partial g(y^k)$ and any $x^k, y^k + d^k \in \mathbb{R}^n$, we have:

$$\begin{aligned} g(y^k + d^k) &\geqslant g(y^k) + \langle w, (y^k + d^k) - y^k \rangle + \frac{\sigma}{2}\|(y^k + d^k) - y^k\|^2 \\ &\geqslant g(y^k) + \langle w, d^k \rangle + \frac{\sigma}{2}\|d^k\|^2 \end{aligned}$$

and

$$\begin{aligned} g(x^k) &\geqslant g(y^k) + \langle w, x^k - y^k \rangle + \frac{\sigma}{2}\|x^k - y^k\|^2 \\ &\geqslant g(y^k) - \langle w, y^k - x^k \rangle + \frac{\sigma}{2}\|d^k\|^2 \\ &\geqslant g(y^k) - \langle w, d^k \rangle + \frac{\sigma}{2}\|d^k\|^2. \end{aligned}$$

This implies that

$$g(y^k + d^k) + g(x^k) \geqslant 2g(y^k) + \sigma\|d^k\|^2.$$

In other words

$$g(y^k + d^k) + g(x^k) - 2g(y^k) \geqslant \sigma\|d^k\|^2 > 0.$$

As $\nu_k > 0$, we have $0 < \hat{\delta}_k := \frac{\nu_k}{g(y^k+d^k)+g(x^k)-2g(y^k)}$, this proves the first statement of item (i). Furthermore,

$$0 < \lambda \left(g(y^k + d^k) + g(x^k) - 2g(y^k) \right) \leqslant \nu_k, \quad \forall \lambda \in (0, \hat{\delta}_k].$$

Note that, $-\frac{3\sigma}{2}\lambda\|d^k\|^2 \leqslant -\rho\lambda^2\|d^k\|^2$ implies that, $\frac{3\sigma}{2}\lambda\|d^k\|^2 \geqslant \rho\lambda^2\|d^k\|^2$. Hence,

$$\frac{3\sigma}{2\rho} \geqslant \lambda.$$

Define $\overline{\delta}_k := \min\{\hat{\delta}_k, 1, \frac{3\sigma}{2\rho}\}$. Therefore, from inequality (4.5). We have:

$$\phi(y^k + \lambda d^k) - \phi(y^k) \leqslant -\rho\lambda^2\|d^k\|^2 + \nu_k, \quad \forall \lambda \in (0, \overline{\delta}_k].$$

Thus, it concludes the second statement of item (i). As $\beta \in (0, 1)$, so $\lim_{j\to\infty} \beta^j \overline{\lambda}_k = 0$ the line search in step 3 is well defined. To prove item (ii), observe that item (i) implies that step 4 is well defined for $\nu_k > 0$. Thus, since that $\lim_{j\to\infty} \beta^j \overline{\lambda}_k = 0$, there exists $j \in \mathbb{N}$ such that $\lambda_k = \beta^j \overline{\lambda}_k$ satisfies

$$\phi(y^k + \lambda d^k) - \phi(y^k) \leqslant -\rho\lambda^2\|d^k\|^2 + \nu_k. \tag{4.6}$$

From Proposition 4.1.1, we have

$$\phi(y^k) \leqslant \phi(x^k) - \sigma\|d^k\|^2. \tag{4.7}$$

Adding up the equations in (4.6) and (4.7), so

$$\phi(y^k + \lambda d^k) \leqslant \phi(x^k) - (\rho\lambda^2 + \sigma)\|d^k\|^2 + \nu_k.$$

Therefore, the proof of the proposition is completed. □

Remark 4.2.1. *When g is not differentiable and convex, the direction $d^k \neq 0$ generated by step 3 of Algorithm 3 is generally not a descent direction of ϕ at y^k. For this reason, in step 3 of Algorithm 3, we assume that $\nu_k > 0$; otherwise, it would not be possible to compute the linear search in the algorithm. When g is differentiable, we assume that $\nu_k \geqslant 0$.*

4.3 Well-definedness of nmBDCA: g is continuously differentiable

In the section, we assume that the function g satisfies the statement follow. In addition to the statements already proposed, **(AF1)** and **(AF2)**.

- (AF3)' g is continuously differentiable.

Then, the parameter $\nu_k \geqslant 0$.

Proposition 4.3.1. *Suppose that $g : \mathbb{R}^n \to \mathbb{R}$ satisfies **(AF3)'**. For each $k \in \mathbb{N}$, assume that $d^k \neq 0$ and $\nu_k \geqslant 0$.Then, the following statements are valid.*

(i) $\phi'(y^k; d^k) = -\sigma\|d^k\|^2 < 0$ *and there exists a constant* $\delta_k > 0$ *such that* $\phi(y^k+\lambda d^k) \leqslant \phi(y^k)-\rho\lambda^2\|d^k\|^2+\nu_k, \forall\lambda \in (0, \delta_k]$. *Consequently, the line search in Step 3 is well defined.*

(ii) $\phi)(x^{k+1}) \leqslant \phi(x^k) - (\sigma + \rho\lambda^2)\|d^k\|^2 + \nu_k$.

Proof. The proof of item (i) follows from [Proposition 3.1.1 (ii)-(iii)] together with the factor of $\nu_k \geqslant 0$. Now, item (ii) follows from Proposition 4.1.1 and item (i). □

Remark 4.3.1. *When $\nu_k = 0$ for all $k \in \mathbb{N}$, then the non-monotone line search of Algorithm 3 coincides with the monotone search of Algorithm 2, that is to say, the nmBDCA is a natural extension of the BDCA introduced in [3] and [4]. Moreover, if $\nu_k > 0$, for all $k \in \mathbb{N}$, so the nmBDCA can be with a less stringent of the BDCA.*

4.4 Strategies to Choose ν_k

Next, we discuss some strategies to choose the sequence of parameters $\{\nu_k\}_{k\in\mathbb{N}}$, suggested by Ferreira, Santos, and Souza in [5]:

- **(S1)** Given $\Delta_{min} \in [0,1)$, the sequence $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is defined as follows: $\nu_0 \geqslant 0$ and ν_{k+1}, for each $\Delta_{k+1} \in [\Delta_{min}, 1]$, satisfies the condition below.

$$0 \leqslant \nu_{k+1} \leqslant (1-\Delta_{k+1})(\phi(x^k) - \phi(x^{k+1}) + \nu_k), \quad \forall k \in \mathbb{N}. \tag{4.8}$$

- **(S2)** $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is such that $\sum_{k=0}^{+\infty} \nu_k < +\infty$;

- **(S3)** $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is such that for every $\delta > 0$, there exists $k_0 \in \mathbb{N}$ with the following property $\nu_k \leqslant \delta\|d^k\|^2$, for all $k \geqslant k_0$.

Lemma 4.4.1. *Since $\Delta_{k+1} \geqslant \Delta_{min} > 0$, if $\{\nu_k\}_{k\in\mathbb{N}}$ satisfies **(S1)**, then satisfies **(S2)**.*

Proof. Observe that, by equation (4.8), we obtain:

$$0 \leqslant \Delta_{k+1}(\phi(x^k) - \phi(x^{k+1}) + \nu_k) \leqslant \phi(x^k) + \nu_k - (\phi(x^{k+1}) + \nu_{k+1}), \quad \forall k \in \mathbb{N}.$$

By hypothesis **(S1)**, we know that $\Delta_{k+1} \geqslant \Delta_{min} > 0$, and by Proposition 4.3.1, item(ii)

$$0 < \sigma\|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}) + \nu_k.$$

Thus, implies that

$$\Delta_{min}(\phi(x^k) - \phi(x^{k+1}) + \nu_k) \leqslant \phi(x^k) + \nu_k - (\phi(x^{k+1}) + \nu_{k+1}). \tag{4.9}$$

Considering the sum partial of the last inequality, we obtain

$$\begin{aligned}\sum_{k=0}^{N} \Delta_{min}(\phi(x^k) - \phi(x^{k+1}) + \nu_k) &\leqslant \sum_{k=0}^{N} \phi(x^k) + \nu_k - (\phi(x^{k+1}) + \nu_{k+1}) \\ &= \phi(x^0) - \phi(x^N) + \nu_0 - \nu_{N+1}.\end{aligned}$$

Since, $\phi^* = \inf_{x\in\mathbb{R}^n} \phi(x)$, for all $k \in \mathbb{N}$, implies that $-\phi^* \geqslant -\phi(x^{N+1})$. Then $\nu_k \geqslant 0$, in particular, $-\nu_{N+1} \leqslant 0$. Then,

$$\sum_{k=0}^{N} \Delta_{min}(\phi(x^k) - \phi(x^{k+1}) + \nu_k) \leqslant \phi(x^0) - \phi^* + \nu_0. \tag{4.10}$$

When $N \to +\infty$ in (4.10) to concluded that

$$\Delta_{min} \sum_{k=0}^{+\infty} (\phi(x^k) - \phi(x^{k+1}) + \nu_k) < +\infty. \tag{4.11}$$

By rearranging equation and (4.9) and applying (4.11), we arrive at

$$\nu_{k+1} \leqslant (1 - \Delta_{min})(\phi(x^k) - \phi(x^{k+1}) + \nu_k).$$

In other words,

$$\sum_{k=0}^{+\infty} \nu_k < +\infty.$$

Therefore, the sequence $\{\nu_k\}_{k\in\mathbb{N}}$ satisfies **(S2)**. □

Remark 4.4.1. *We can consider the following strategy:*

- ***(S'3)*** *Fix any* $\overline{\delta} \in (0, \sigma)$. *There exists* $k_0 \in \mathbb{N}$ *such that* $\nu_k \leqslant \overline{\delta}\|d^k\|^2$, *for all* $k_0 \geqslant k$.

Now, we will present some examples according to the strategies discussed.

Example 4.4.1. *Take any* $\nu_0 > 0$, *and define* Δ_{k+1} *and* ν_k *as follows*

$$0 < \Delta_{min} \leqslant \Delta_{k+1} < 1, \quad 0 < \nu_{k+1} := (1 - \Delta_{k+1})(\sigma + \rho\lambda^2)\|d^k\|^2, \quad \forall k \in \mathbb{N}. \tag{4.12}$$

Follows the Proposition 4.1.1 that $(\sigma + \rho\lambda^2)\|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}) - \nu_k$. *Then, for every* $d^k \neq 0$, *we have* $0 < \nu_{k+1} \leqslant (1 - \Delta_{k+1})(\phi(x^k) - \phi(x^{k+1}) - \nu_k)$. *When* $\{\nu_k\}_{k\in\mathbb{N}}$ *defined in* (4.12) *satisfies **(S1)**. As a result, considering* $0 < \Delta_{min}$, *by Lemma 4.4.1 we concluded that* $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *also satisfies **(S2)**.*

Example 4.4.2. *Let* $\omega > 0$ *be a constant. Consider a sequence* $\{\nu_k\}_{k\in\mathbb{N}}$ *defined by*

$$\nu_k := \frac{\omega\|d^k\|^2}{k+1}, \quad \forall k \in \mathbb{N}.$$

*This sequence satisfies **(S3)**. In fact, note that* $\lim_{k\to+\infty} \frac{\omega}{k+1} = 0$, *then for each* $\delta > 0$ *there exists* $k_0 \in \mathbb{N}$ *such that* $k_0 \geqslant k$ *implies that* $|\frac{\omega}{k+1}| \leqslant \delta$, *in other words,* $\frac{\omega}{k+1} \leqslant \delta$. *Hence, we obtain* $\nu_k \leqslant \delta\|d^k\|^2$. *A similar situation arises when we define*

$$\nu_k := \frac{\omega\|d^k\|^2}{\ln(k+2)}, \quad \forall k \in \mathbb{N}.$$

*Also satisfies **(S3)**.*

4.5 Convergence Analysis: g is possibly non-differentiable

The objective of this section is to show the convergence results and the analysis of iteration complexity of the nmBDCA for the case when g is non differentiable. Assume $\nu_k > 0$.

4.5.1 Asymptotic converge analysis

The main result of this section is the relationship between the cluster point of the sequence $\{x^k\}_{k\in\mathbb{N}}$, if any, and the critical point of ϕ.

Theorem 4.5.1. *If* $\lim_{k\to+\infty}\|d^k\| = 0$*, then every cluster point of the sequence* $\{x^k\}_{k\in\mathbb{N}}$*, if any, is a critical point of* ϕ.

Proof. Let x^* a cluster point of $\{x^k\}_{k\in\mathbb{N}}$, then exists a subsequence $\{x^{k_l}\}_{l\in\mathbb{N}}$ of $\{x^k\}_{k\in\mathbb{N}}$ such that $\lim_{l\to+\infty} x^{k_l} = x^*$. Consider the sequences $\{w^{k_l}\}$ and $\{y^{k_l}\}$ generated by Algorithm 3, this is, $w^{k_l} \in \partial h(x^{k_l})$. Since, y^{k_l} is solution of the subproblem (4.1) we have $w^{k_l} \in \partial g(y^{k_l})$. As $\lim_{k\to+\infty}\|d^k\| = 0$ and $\lim_{l\to+\infty} x^{k_l} = x^*$ follows that

$$\|y^{k_l} - x^*\| = \|y^{k_l} - x^{k_l} + x^{k_l} - x^*\| \leqslant \|y^{k_l} - x^{k_l}\| + \|x^{k_l} - x^*\|$$

$\lim_{l\to+\infty} y^{k_l} = x^*$. Now, considering that $\lim_{l\to+\infty} w^{k_l} = w^*$ and $w^{k_l} \in \partial h(x^{k_l}) \cap \partial g(y^{k_l})$, we can apply Proposition 1.2.3 obtain $w^* \in \partial h(x^*) \cap \partial g(x^*)$, finishing the proof. □

Theorems that follow prove asymptotic convergence of Algorithm 3 when $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ and satisfies **(S2)**, **(S3)** or **(S1)**.

Theorem 4.5.2. *If the sequence* $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *chosen according to the strategy* ***(S2)****, then every cluster point, if any, is a critical point of* ϕ.

Proof. By Proposition 4.2.1, item (ii), it follows

$$0 \leqslant \sigma\|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}) + \nu_k, \quad \forall k \in \mathbb{N}.$$

Compute the finite sum of k, where k varies from 0 to N. Thus:

$$\begin{aligned} 0 \leqslant \sum_{k=0}^{N} \sigma\|d^k\|^2 &\leqslant \sum_{k=0}^{N} (\phi(x^k) - \phi(x^{k+1}) + \nu_k \\ &\leqslant \frac{1}{\sigma}(\phi(x^0) - \phi(x^{N+1}) + \sum_{k=0}^{N} \nu_k, \quad \forall k \in \mathbb{N}. \end{aligned}$$

Being, $\phi^* \leqslant \phi(x^{N+1})$ e doing $N \to +\infty$, we obtain

$$0 \leqslant \sum_{k=0}^{N} \|d^k\|^2 \leqslant \frac{1}{\sigma}(\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty} \nu_k) < +\infty.$$

Therefore, $\lim_{k\to+\infty} \|d^k\| = 0$, by Theorem 4.5.1, follows that every cluster point of the sequence $\{x^k\}_{k\in\mathbb{N}}$ is a critical point of ϕ. □

Theorem 4.5.3. *If the sequence $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ chosen according to the strategy **(S3)**, then every cluster point, if any, is a critical point of ϕ.*

Proof. Using the **(S3)** strategy, there exists $k_0 \in \mathbb{N}$ such that $0 \leqslant \nu_k \leqslant \frac{\sigma}{2}\|d^k\|^2, \quad \forall k \geqslant k_0$. It follows that

$$0 \leqslant \frac{\sigma}{2}\|d^k\|^2 \leqslant \sigma\|d^k\|^2 - \nu_k, \quad \forall k \geqslant k_0.$$

By the Proposition 4.2.1, item (ii) and **(AF2)** we have

$$0 \leqslant \frac{\sigma}{2}\|d^k\|^2 \leqslant \sigma\|d^k\|^2 - \nu_k \leqslant \phi(x^k) - \phi(x^{k+1}), \quad \forall k \geqslant k_0,$$

which implies that

$$\sum_{k=0}^{+\infty} \|d^k\|^2 \leqslant \frac{2}{\sigma}(\phi(x^0) - \phi^*) < +\infty.$$

Therefore, $\lim_{k\to+\infty} \|d^k\| = 0$. By Theorem 4.5.1, the result follows. □

Theorem 4.5.4. *If the sequence $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ chosen according to the strategy **(S1)**, then the following statements hold:*

(i) The sequence $\{\phi(x^k) + \nu_k\}_{k\in\mathbb{N}}$ is non-increasing and convergent;

(ii) If $\lim_{k\to\infty} \nu_k = 0$, then every cluster point of $\{x^k\}_{k\in\mathbb{N}}$, if any, is a critical point of ϕ;

(iii) If $\Delta_{min} > 0$, then every cluster point of $\{x^k\}_{k\in\mathbb{N}}$, if any, is a critical point of ϕ.

Proof.

(i) If **(S1)** holds, then

$$0 \leqslant \Delta_{k+1}(\phi(x^k) - \phi(x^{k+1}) + \nu_k) \leqslant \phi(x^k) + \nu_k - (\phi(x^{k+1}) + \nu_{k+1}).$$

Hence, $\phi(x^k) + \nu_k \geqslant (\phi(x^{k+1}) + \nu_{k+1})$ is non-increasing. Therefore, using that $\phi^* = \inf_{x\in\mathbb{R}^n}\{\phi(x)\} \leqslant -\infty$ and using the hypothesis **(S1)** that ν_k is bounded below, we have that $\{\phi(x^k) + \nu_k\}_{k\in\mathbb{N}}$ is lower bound. Moreover, $\{\phi(x^k) + \nu_k\}_{k\in\mathbb{N}}$ is convergent.

(ii) Since $\lim_{k\to+\infty}\nu_k = 0$ for item (i), it follows that $\{\phi(x^k)\}_{k\in\mathbb{N}}$ is convergent. On the other hand, from Proposition 4.2.1 item (ii), we have

$$0 \leqslant \sigma\|d^k\|^2 \leqslant \phi(x^k) + \nu_k - \phi(x^{k+1}), \quad \forall k \in \mathbb{N},$$

$$\begin{aligned} 0 \leqslant \lim_{k\to\infty} \sigma\|d^k\|^2 &\leqslant \lim_{k\to\infty} (\phi(x^k) + \nu_k - \phi(x^{k+1})) \\ &\leqslant \phi^* - \phi^* \\ &= 0, \quad \forall k \in \mathbb{N}. \end{aligned}$$

Thus, $\lim_{k\to\infty}\|d^k\| = 0$, from the Theorem 4.5.1, we complete the proof.

(iii) If $\Delta_{min} > 0$ and $\{\nu_k\}_{k\in\mathbb{N}}$ satisfies **S1**, then, by Lemma 4.4.1, we have $\sum_{k=0}^{+\infty}\nu_k < \infty$. This condition corresponds exactly to the (S2) strategy. Therefore, by the Theorem 4.5.2, the desired result follows. □

4.6 Iteration-complexity analysis

Theoretically assessing the computational resources needed to run an algorithm in relation to the volume of input data is known as complexity analysis. Execution time and memory usage are the main metrics used to quantify such resources. Regardless of particular implementations or hardware capacity, this method makes it possible to compare the effectiveness of various algorithms.

In this section, we will present the iteration complexity limit for $\{x^k\}_{k\in\mathbb{N}}$ generated by the presented algorithm. For the case where $\{\nu_k\}_{k\in\mathbb{N}}$ is chosen according to each strategy.

Theorem 4.6.1. *Suppose that the sequence* $\{\nu_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *is chosen according to strategy **(S2)**. For each* $N \in \mathbb{N}$*, we have*

$$\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{\sqrt{\phi(x^0) - \phi^* + \sum_{k=0}^{\infty}\nu_k}}{\sqrt{\sigma}}\frac{1}{\sqrt{N}}.$$

As a result, for a given accuracy $\epsilon > 0$*, if* $N \geqslant \frac{\phi(x^0)-\phi^*+\sum_{k=0}^{\infty}\nu_k}{\sigma\epsilon^2}$*, then the following inequality holds* $\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \epsilon$.

Proof. Since, $\inf_{x\in\mathbb{R}^n}\phi(x) \leqslant \phi(x^k), \quad \forall k \in \mathbb{N}$, follows Proposition 4.2.1, item (ii) that

$$\sum_{k=0}^{N-1}\|d^k\|^2 \leqslant \frac{1}{\sigma}(\phi(x^0) - \phi(x^N) + \sum_{k=0}^{N-1}\nu_k) \leqslant \frac{1}{\sigma}(\phi(x^0) - \phi^* + \sum_{k=0}^{N-1}\nu_k).$$

Then, $\sum_{k=0}^{N-1} \|d^k\|^2 \leqslant \frac{1}{\sigma}(\phi(x^0) - \phi^* + \sum_{k=0}^{N-1} \nu_k)$. Consider,

$$N \cdot \min\Big\{\|d^k\|^2 : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{1}{\sigma}(\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty} \nu_k).$$

In other words,

$$\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{\sqrt{(\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty} \nu_k)}}{\sqrt{\sigma}} \frac{1}{\sqrt{N}}. \tag{4.13}$$

Given $\epsilon > 0$, if $N \geqslant \frac{(\phi(x^0) - \phi^* + \sum_{k=0}^{\infty} \nu_k)}{\sigma \epsilon^2}$. Hence,

$$\frac{1}{\sqrt{N}} \leqslant \frac{\sqrt{\sigma \epsilon^2}}{\sqrt{(\phi(x^0) - \phi^* + \sum_{k=0}^{\infty} \nu_k)}}. \tag{4.14}$$

We show that (4.13) holds, combining with (4.14) we get

$$\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \epsilon.$$

□

Remark 4.6.1. *The Theorem above provides performance guarantees, which is essential for practical applications where computational cost is important.*

Theorem 4.6.2. *Suppose that the sequence $\{\nu_k\}_{k \in \mathbb{N}} \subset \mathbb{R}_{++}$ is chosen according to the strategy **(S3)**. Consider $0 < \beta < 1$ and $k_0 \in \mathbb{N}$ such that $\nu_k \leqslant \beta \sigma \|d^k\|^2$, for all $k \geqslant k_0$.*

$$\min\Big\{\|d^k\| : K = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{\sqrt{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1} \nu_k}}{\sqrt{(1-\beta)\sigma}} \frac{1}{\sqrt{N}}.$$

As a result, for a given accuracy $\epsilon > 0$ and $k_0 \in \mathbb{N}$ such that $\nu_k \leqslant \beta \sigma \|d^k\|^2$ for all $k \geqslant k_0$, if $N \geqslant \max\{k_0, \frac{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1} \nu_k}{\sigma(1-\beta)\epsilon^2}\}$, then the following inequality holds $\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \epsilon$.

Proof. Let $\beta \in (0, 1)$ and $k_0 \in \mathbb{N}$ such that $\nu_k \leqslant \beta \sigma \|d^k\|^2, \quad \forall k \geqslant k_0$ it follows that Proposition (4.2.1), item (ii)

$$0 \leqslant \sigma \|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}) + \nu_k, \quad \forall k = 0, 1, 2, \dots, N-1.$$

Summing the inequality from 1 to $N-1$

$$0 \leqslant \sum_{k=0}^{N-1} \sigma \|d^k\|^2 \leqslant \phi(x^0) - \phi^* + \sum_{k=0}^{N-1} \nu_k.$$

Considered that $\nu_k \leqslant \beta\sigma\|d^k\|^2, \quad \forall k \geqslant k_0$, then

$$\sum_{k=0}^{N-1} \sigma\|d^k\|^2 \leqslant \phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1} \nu_k + \sum_{k=0}^{N-1} \beta\sigma\|d^k\|^2.$$

This implies that $\sum_{k=0}^{N-1}(1-\beta)\sigma\|d^k\|^2 \leqslant \phi(x^0) - \phi * + \sum_{k=0}^{k_0-1} \nu_k$, in other words,

$$\sum_{k=0}^{N-1} \|d^k\|^2 \leqslant \frac{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1} \nu_k}{(1-\beta)\sigma}.$$

Analogous to the Theorem 4.6.1,

$$\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{\sqrt{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1} \nu_k}}{\sqrt{(1-\beta)\sigma}} \frac{1}{\sqrt{N}}.$$

Now, given $\epsilon > 0$ and if $N \geqslant \max\{k_0, \frac{(\phi(x^0)-\phi^*+\sum_{k=0}^{k_0-1})}{\sigma(1-\beta)\epsilon^2}\}$. Then, if $N \geqslant \frac{(\phi(x^0)-\phi^*+\sum_{k=0}^{k_0-1})}{\sigma(1-\beta)\epsilon^2}$ it follows that

$$\min\Big\{\|d^k\|^2 : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \epsilon.$$

Otherwise, $N \geqslant k_0 \geqslant \frac{(\phi(x^0)-\phi^*+\sum_{k=0}^{k_0-1})\nu_k}{\sigma(1-\beta)\epsilon^2}$ it is a similar case. □

4.7 Convergence Analysis: ∇g is Lipschitz continuous

In this section, we will present convergence analysis and complexity analysis of iteration when g is continuously differentiable and ∇g is globally Lipschitz continuous. Consider

$$\inf_{k\in\mathbb{N}} \bar{\lambda}_k \geqslant \lambda_{-1} > 0. \tag{4.15}$$

The steps $\bar{\lambda}_k$ are as large as one wants, but limited below by λ_{-1}. We need to assume that $\nu_k \geqslant 0$ and an important property for error control.

- **(AF4)** ∇g is Lipschitz continuous with constant $K > 0$.

Let $\lambda_{-1} > 0$ defined as (4.15) before stating the next result, we need to define the following constant

$$\lambda_{min} := \min\Big\{\lambda_{-1}, \frac{2\sigma\beta}{(K+2\rho)}\Big\}. \tag{4.16}$$

It follows from Step 3 that we can simplify the notation by defining

$$\lambda_k := \beta^{j_k}\bar{\lambda}_k, \quad j_k := \min\{j \in \mathbb{N} : \phi(y^k + \beta^j\bar{\lambda}_k d^k) \leqslant \phi(y^k) - \rho(\beta^j\bar{\lambda}_k)^2\|d^k\|^2 + \nu_k\}, \tag{4.17}$$

for each $k \in \mathbb{N}$.

Lemma 4.7.1. *If* g *satisfies* ***(AF4)****, then* $\lambda_k \geqslant \lambda_{min}$*, for all* $k \in \mathbb{N}$*.*

Proof. If $j_k = 0$, then $\lambda_k = \bar{\lambda}_k$ by (4.15) and (4.16), follows $\lambda_k \geqslant \lambda_{min}$. Otherwise, assume that $j_k > 0$. Since $\lambda_k = \beta^{j_k}\bar{\lambda}_k$ we obtain from (4.17) that

$$\phi(y^k + \frac{\lambda_k}{\beta}d^k) > \phi(y^k) - \frac{\rho\lambda_k^2}{\beta^2}\|d^k\|^2 + \nu_k. \tag{4.18}$$

On the other hand, by Lemma 1.3.1. Where, consider $\lambda = \frac{\lambda_k}{\beta}$, we obtain

$$\phi(y^k + \frac{\lambda_k}{\beta}d^k) - \phi(y^k) \leqslant \frac{\lambda_k}{\beta}\langle \nabla g(y^k)) - s^k, d^k\rangle + \frac{K}{2}\frac{\lambda_k^2}{\beta^2}\|d^k\|^2, \quad \forall s^k \in \partial h(y^k). \tag{4.19}$$

Let y^k be a solution to the subproblem, so $\nabla g(y^k) = w^k \in \partial h(x^k)$. Hence, since that h is strongly convex, then ∂h is strongly monotone, this is

$$\begin{aligned}\langle w^k - s^k, x^k - y^k\rangle &= -\langle w^k - s^k, y^k - x^k\rangle \\ &= -\langle \nabla g(y^k) - s^k, d^k\rangle \\ &\leqslant -\sigma\|d^k\|^2.\end{aligned}$$

Therefore, since that $\nu_k \geqslant 0$, we obtain from the equation (4.19)

$$\phi(y^k + \frac{\lambda_k}{\beta}d^k) - \phi(y^k) \leqslant -\frac{\lambda_k\sigma}{\beta}\|d^k\|^2 + \frac{K}{2}\frac{\lambda_k^2}{\beta^2}\|d^k\|^2 + \nu_k. \tag{4.20}$$

Combining (4.18) and (4.20) we have

$$\begin{aligned}-\frac{\rho\lambda_k^2}{\beta^2}\|d^k\|^2 &< -\frac{\lambda_k\sigma}{\beta}\|d^k\|^2 + \frac{K}{2}\frac{\lambda_k^2}{\beta^2}\|d^k\|^2 \\ &< -\frac{\lambda_k\|d^k\|^2}{\beta}(\sigma - \frac{K\lambda_k}{2\beta}) \\ &< -\frac{\lambda_k\|d^k\|^2}{2\beta^2}(2\beta\sigma - K\lambda_k).\end{aligned}$$

Therefore,

$$2(\rho\lambda_k) + K\lambda_k > 2\beta\sigma,$$

or

$$\lambda_k > \frac{2\beta\sigma}{K + 2\rho}. \tag{4.21}$$

Considering that $\rho > 0, K > 0$ and $d^k \neq 0$, by equation (4.21) and (4.16), we found

$$\lambda_k > \frac{2\beta\sigma}{(K + 2\rho)} \geqslant \lambda_{min}.$$

□

Corollary 4.7.1. *If* g *satisfies **(AF4)**, then* $(\sigma+\rho\lambda_{min}^2)\|d^k\|^2 \leqslant \phi(x^k)-\phi(x^{k+1})+\nu_k$, *for all* $k \in \mathbb{N}$.

Proof. By Proposition 4.3.1, item (ii) it follows that

$$(\sigma + \rho\lambda_k^2)\|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}) + \nu_k,$$

and by Lemma 4.16 $\lambda_k \geqslant \lambda_{min}$, for all $k \in \mathbb{N}$, then

$$(\sigma + \rho\lambda_k^2)\|d^k\|^2 \geqslant (\sigma + \rho\lambda_{min}^2)\|d^k\|^2, \forall k \in \mathbb{N}.$$

Therefore,

$$(\sigma + \rho\lambda_{min}^2)\|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}) + \nu_k, \forall k \in \mathbb{N},$$

where, $\sigma > 0$ and $\rho > 0$. □

4.7.1 Iteration complexity bounds

The objective of this section is to present some iteration limits for the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by the Algorithm 3, in the case where ∇g is Lipschitz continuous. In fact, assume:

$$\sup_{k\in\mathbb{N}} \bar{\lambda}_k \leqslant \lambda_{sup} \leqslant +\infty. \tag{4.22}$$

Lemma 4.7.2. *Suppose that* g *satisfies **(AF4)**. Let* $j_k \in \mathbb{N}$ *be a integer defined in* (4.17), *and let* J_k *be the number of functions evaluations* ϕ *in the linear search in Algorithm 3 after* $k \geqslant 1$ *iterations. Then,*

$$j_k \leqslant \left\lfloor \frac{(\log \lambda_{min} - \log \lambda_{sup}}{\log \beta} \right\rfloor + 1,$$

and

$$J_k \leqslant (k+1)\left(\frac{\log \lambda_{min} - \log \lambda_{sup}}{\log \beta} + 2 \right),$$

where $\lambda_{sup} := \sup_{k\in\mathbb{N}} \bar{\lambda}_k$.

Proof. Using by (4.22), we have that

$$\sup \bar{\lambda}_k \leqslant \lambda_{sup} < +\infty.$$

It follows from Lemma 4.16, $\lambda_k \geqslant \lambda_{min}, \quad \forall k \geqslant k \in \mathbb{N}$. From (4.17), we obtain:

$$\lambda_k := \beta^{j_k}\bar{\lambda}_k.$$

Hence,

$$0 < \lambda_{min} \leqslant \lambda_k := \beta^{j_k}\overline{\lambda}_k \leqslant \lambda_{sup}, \quad \forall k \in \mathbb{N}. \tag{4.23}$$

By applying the logarithm function to this inequality (4.23), we can deduce that

$$0 < \log\lambda_{min} \leqslant \log\lambda_k = j_k(\log\beta) + \log\overline{\lambda}_k \leqslant \log\lambda_{sup}, \quad \forall k \in \mathbb{N}.$$

Then, taking $\beta \in (0,1)$ and $0 < \overline{\lambda}_k < \lambda_{sup}$ and consider that
$\lambda_{min} \leqslant \overline{\lambda}_k \leqslant \lambda_{sup}$, implies that, $-\log\lambda_{min} \geqslant -\log\overline{\lambda}_k \geqslant -\log\lambda_{sup}$. Therefore, $\frac{\log\lambda_{min}-\log\overline{\lambda}_k}{\log\beta} \leqslant \frac{\log\lambda_{min}-\log\lambda_{sup}}{\log\beta} \leqslant \left\lfloor\frac{\log\lambda_{min}-\log\lambda_{sup}}{\log\beta}\right\rfloor + 1$, we obtain

$$\begin{aligned} j_k = \frac{\log\lambda_k - \log\overline{\lambda}_k}{\log\beta} &\leqslant \frac{\log\lambda_{min}}{\log\beta} \\ &\leqslant \frac{\log\lambda_{min} - \log\overline{\lambda}_k}{\log\beta} \\ &\leqslant \frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta}, \quad \forall k \in \mathbb{N}. \end{aligned}$$

This prove the first part of the inequality. To prove the second one we sun the first inequality from $l = 0$ to k and we get

$$\begin{aligned} \sum_{l=0}^{k} j_l \leqslant \sum_{l=0}^{k} \frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta} \\ = (k+1)\frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta}. \end{aligned} \tag{4.24}$$

On the hand, the definition of J_k and (4.24) implies that

$$\begin{aligned} J_k = \sum_{l=0}^{k}(j_l + 2) = 2(k+1) + \sum_{l=0}^{k} j_l \\ \leqslant (k+1)\left(\frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta} + 2\right). \end{aligned}$$

□

Theorem 4.7.1. *Suppose that $\{\nu^k\}_{k\in\mathbb{N}}$ is chosen according to the strategy **(S2)** and g satisfies **(AF4)**. For each $N \in \mathbb{N}$, we obtain*

$$\min\left\{\|d^k\| : k = 0,1,2,3,\dots,N-1\right\} \leqslant \frac{\sqrt{\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty}\nu_k}}{\sqrt{\sigma + \rho\lambda_{min}^2}}\frac{1}{\sqrt{N}}.$$

Accordingly, for a given $\epsilon > 0$, if $N \geqslant \frac{\phi(x^0)-\phi^*+\sum_{k=0}^{+\infty}\nu_k}{(\sigma+\rho\lambda_{min}^2)\epsilon^2}$, then the following inequality holds

$$\min\left\{\|d^k\| : k = 0,1,\dots,N-1\right\} \leqslant \epsilon.$$

Proof. Given $\phi^* := \inf_{x\in\mathbb{R}} \phi(x) \leqslant \phi(x^k), \quad \forall k \in \mathbb{N}$, by Corollary 4.7.1, we have

$$(\sigma + \rho\lambda_{min}^2)\|d^k\|^2 \leqslant \phi(x^0) - \phi(x^{k+1}) + \nu_k.$$

Taking the sum with k varying from 0 to $N-1$ and considering the statements about ϕ and ν_k, it follows that

$$(\sigma + \rho\lambda_{min}^2)\sum_{k=0}^{N-1}\|d^k\|^2 \leqslant \phi(x^0) - \phi^* + \sum_{k=0}^{+\infty}\nu_k.$$

Then,

$$N \cdot \min\Big\{\|d^k\|^2 : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{1}{(\sigma + \rho\lambda_{min}^2)}(\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty}\nu_k).$$

Finally,

$$\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{\sqrt{\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty}\nu_k}}{\sqrt{(\sigma + \rho\lambda_{min}^2)}}\frac{1}{\sqrt{N}}. \tag{4.25}$$

If $N > \frac{\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty}\nu_k}{(\sigma+\rho\lambda_{min}^2)\epsilon^2}$ and given $\epsilon > 0$, by equation (4.25). Here is the second part of the demonstration. □

Theorem 4.7.2. *Suppose that $\{\nu^k\}_{k\in\mathbb{N}}$ is chosen according to the strategy **(S2)** and g satisfies **(AF4)**. Given $\epsilon > 0$, the number of functions evaluations ϕ in the Algorithm 3 for computer d^k such that $\|d^k\| \leqslant \epsilon$ is at most*

$$(k+1)\left(\frac{\phi(x^0) - \phi^* + \sum_{k=0}^{+\infty}\nu_k}{(\sigma + \rho\lambda_{min}^2)\epsilon^2} + 1\right)\left(\frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta} + 2\right).$$

Proof. By Lemma 4.7.2, we have:

$$J_k \leqslant (k+1)\left(\frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta} + 2\right).$$

Now, by Theorem 4.7.1 if $N \geqslant \frac{\phi(x^0) - \phi^* + \sum_{k=0}^{\infty}\nu_k}{(\sigma+\rho\lambda_{min}^2)}$, then $\|d^k\| \leqslant \epsilon$. Since we need an integer, consider

$$\overline{N} = \left\lceil \frac{\phi(x^0) - \phi^* + \sum_{k=0}^{\infty}\nu_k}{(\sigma + \rho\lambda_{min}^2)\epsilon^2} \right\rceil \leqslant \left(\frac{\phi(x^0) - \phi^* + \sum_{k=0}^{\infty}\nu_k}{(\sigma + \rho\lambda_{min}^2)\epsilon^2} + 1\right).$$

Each iteration uses at most J_k iterations of ϕ. The total number of function evaluations after $\overline{N}$ iterations is limited by

$$\overline{N} \cdot J_k \leqslant (k+1)\left(\frac{\phi(x^0) - \phi^* + \sum_{k=0}^{\infty}}{(\sigma + \rho\lambda_{min}^2)\epsilon^2} + 1\right)\left(\frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta} + 2\right).$$

□

Theorem 4.7.3. *Suppose that* $\{\nu^k\}_{k\in\mathbb{N}}$ *is chosen according to the strategy* ***(S3)*** *and* g *satisfies* ***(AF4)****. Let* $0 < \beta < 1$ *and* $k_0 \in \mathbb{N}$ *such that* $\nu_k \leqslant \beta(\sigma + \rho\lambda^2_{min})\|d^k\|^2$*, for all* $k_0 \geqslant k$*. Then, for each* $N \in \mathbb{N}$ *such that* $N \geqslant k_0$*, there holds*

$$\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{\sqrt{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1}\nu_k}}{\sqrt{(1-\beta)(\sigma + \rho\lambda^2_{min})}}\frac{1}{\sqrt{N}}.$$

Accordingly, for a given $\epsilon > 0$ *and* $k_0 \in \mathbb{N}$ *such that* $\nu_k \leqslant \beta(\sigma + \rho\lambda^2_{min})\|d^k\|^2$ *for all* $k \geqslant k_0$*, if* $N \geqslant \max\Big\{k_0, \frac{\phi(x^0)-\phi^*+\sum_{k=0}^{k_0-1}\nu_k}{(1-\beta)(\sigma+\rho\lambda^2_{min})\epsilon^2}\Big\}$*, then the following inequality holds*

$$\min\Big\{\|d^k\| : k = 0, 1, \dots, N-1\Big\} \leqslant \epsilon.$$

Proof. Let $\beta \in (0,1)$ and $k_0 \in \mathbb{N}$ such that $\nu_k \leqslant \beta(\sigma + \rho\lambda_{min})\|d^k\|^2, \quad \forall k \geqslant k_0$. It follows from Corollary 4.7.1 that

$$0 \leqslant (\sigma + \rho\lambda_{min})\|d^k\|^2 \leqslant \phi(x^k) - \phi(x^{k+1}) + \nu_k, \quad \forall k = 0, 1, 2, \dots, N-1.$$

Summing up the inequality from 1 to $N-1$ and considered that $\nu_k \leqslant \beta(\sigma + \lambda^2_{min})\|d^k\|^2, \quad \forall k \geqslant k_0$, then

$$\begin{aligned} 0 \leqslant \sum_{k=0}^{N-1}(\sigma + \rho\lambda_{min})\|d^k\|^2 &\leqslant \phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1}\nu_k + \sum_{k=0}^{N-1}\nu_k \\ &\leqslant \phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1}\nu_k + \beta(\sigma + \rho\lambda^2_{min})\sum_{k=k_0}^{N-1}\|d^k\|^2 \\ &\leqslant \phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1}\nu_k + \beta(\sigma + \rho\lambda^2_{min})\sum_{k=0}^{N-1}\|d^k\|^2. \end{aligned}$$

This implies that $\sum_{k=0}^{N-1}(1-\beta)(\sigma + \rho\lambda^2_{min})\|d^k\|^2 \leqslant \phi(x^0) - \phi * + \sum_{k=0}^{k_0-1}\nu_k$, in other words,

$$\sum_{k=0}^{N-1}\|d^k\|^2 \leqslant \frac{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1}\nu_k}{(1-\beta)(\sigma + \lambda^2_{min})}.$$

Analogous to the Theorem 4.7.1,

$$\min\Big\{\|d^k\| : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \frac{\sqrt{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1}\nu_k}}{\sqrt{(1-\beta)(\sigma + \rho\lambda^2_{min})}}\frac{1}{\sqrt{N}}.$$

Now, given $\epsilon > 0$ and if $N \geqslant \max\{k_0, \frac{(\phi(x^0)-\phi^*+\sum_{k=0}^{k_0-1})}{(\sigma+\rho\lambda^2_{min})(1-\beta)\epsilon^2}\}$ then, if $N \geqslant \frac{(\phi(x^0)-\phi^*+\sum_{k=0}^{k_0-1})}{(\sigma+\rho\lambda^2_{min})(1-\beta)\epsilon^2}$ it follows that

$$\min\Big\{\|d^k\|^2 : k = 0, 1, 2, 3, \dots, N-1\Big\} \leqslant \epsilon.$$

Otherwise, $N \geqslant k_0 \geqslant \frac{(\phi(x^0)-\phi^*+\sum_{k=0}^{k_0-1})\nu_k}{(\sigma+\rho\lambda^2_{min})(1-\beta)\epsilon^2}$ it is a similar case. □

Theorem 4.7.4. *Suppose that* $\{\nu^k\}_{k\in\mathbb{N}}$ *is chosen according to the strategy* ***(S3)*** *and* g *satisfies* ***(AF4)****. Let* $0 < \beta < 1$ *and* $k_0 \in \mathbb{N}$ *such that* $\nu_k \leqslant \beta(\sigma + \rho\lambda_{min}^2)\|d^k\|^2$*, for all* $k_0 \geqslant k$*. Then, the number of functions evaluations in Algorithm for computer* d^k *such that* $\|d^k\| \leqslant \epsilon$ *is at most*

$$(k+1)\left(\max\left\{k_0, \frac{\phi(x^0) - \phi^* + \sum_{k=0}^{k_0-1}\nu_k}{(1-\beta)(\sigma+\rho\lambda_{min}^2)\epsilon^2}\right\}+1\right)\left(\frac{\log\lambda_{min} - \log\lambda_{sup}}{\log\beta} + 2\right). \quad (4.26)$$

Proof. The proof follows similarly to the Theorem 4.7.2; just combine the Lemma 4.7.2 with the Theorem 4.7.3. Then, obtain the equation (4.26). □

4.7.2 Full convergence under the Kuryka–Łojasiewicz property

Our goal in this section is to show that the sequence generated by the algorithm converges completely, provided that we consider that ϕ satisfies the Kurdyka–Łojasiewicz property (K-Ł property); see for instance [20],[18]. In practice, this property guarantees that the closer you are to the critical point, the more the subgradient points towards it.

Definition 4.7.1. *Let* $C^1[(0,+\infty)]$ *be the set of all continuously differentiable functions defined in* $(0,+\infty)$*,* $\phi : \mathbb{R}^n \to \mathbb{R}$ *be a locally Lipschitz function and* $\partial_c\phi(\cdot)$ *be a Clarke's subdifferential of* ϕ*. The function* ϕ *is said to have the Kurdyka–Lojasiewicz property at* $x^* \in dom\partial\phi$ *if here exist* $\eta \in (0,+\infty]$*, a neighborhood* U *of* x^* *and a continuous concave function* $\gamma : [0,\eta) \to \mathbb{R}_+$ *(called desingularizing function) such that:*

(i) $\gamma(0) = 0$

(ii) γ *is* C^1 *on* $[(0,+\infty)]$,

(iii) For all $s \in (0,\eta)$ $\gamma'(s) > 0$,

(iv) And for all $x \in U \cap \{x \in \mathbb{R}^n \mid \phi(x^*) < \phi < \phi(x^*) + \eta\}$*, the Kurdyka-Lojasiewicz inequality holds*

$$\gamma'(\phi(x) - \phi(x^*))\,dist(0, \partial_c\phi(x)) \geqslant 1,$$

where $dist(0, \partial_c\phi(x)) := \inf\{\|s\| : s \in \partial_c\phi(x)\}$.

Theorem 4.7.5. *Suppose that* $\{\nu_k\}_{k\in\mathbb{N}}$ *is chosen according to strategy* ***(S3)****. Assume that* $\{x^k\}_{k\in\mathbb{N}}$ *has a cluster point* x^**,* ∇g *is locally Lipschitz continuous around* x^**, since that* ϕ *satisfies K-L property at* x^**. Then* $\{x^k\}_{k\in\mathbb{N}}$ *converges to* x^**, which is a critical point of* ϕ.

Proof. Assume that $\{\nu_k\}_{k\in\mathbb{N}}$ satisfies **(S3)**, there exists $k_0 \in \mathbb{N}$ such that $\nu_k \leqslant (\frac{\sigma}{2})\|d^k\|^2$, for all $k \geqslant k_0$,

$$0 < (\frac{\sigma}{2})\|d^k\|^2 = \sigma\|d^k\|^2 - (\frac{\sigma}{2})\|d^k\|^2 \leqslant \sigma\|d^k\|^2 - \nu_k, \quad \forall k \geqslant k_0. \tag{4.27}$$

Combining (4.27) with the Proposition 4.3.1, item (ii). We have,

$$0 < (\frac{\sigma}{2} + \rho\lambda_k^2)\|d^k\|^2 \leqslant (\sigma + \rho\lambda_k^2)\|d^k\|^2 - \nu_k \leqslant \phi(x^k) - \phi(x^{k+1}), \quad \forall k \geqslant k_0. \tag{4.28}$$

Consider x^* as a cluster point of $\{x^k\}_{k\in\mathbb{N}}$, then there exists a subsequence $\{x^{kl}\}_{l\in\mathbb{N}}$ such that $\lim_{l\to+\infty} x^{kl} = x^*$, which combined with (4.28) implies $\lim_{k\to+\infty}\phi(x) = \phi(x^*)$, since $\lim_{l\to\infty} = \phi(x^{kl}) = \phi(x^*)$ and ϕ is decreasing and bounded below. If there exists an integer $k \geqslant k_0$ with $\phi(x) = \phi(x^*)$, it follows that $d^k = 0$. In this case, Algorithm 3 stops after a finite number of steps and the proof it is done. Currently, assume $\phi(x) > \phi(x^*)$ for all $k \geqslant k_0$. Since ∇g is locally Lipschitz around x^*, there exists $\hat{\delta} > 0$ and $K > 0$ such that

$$\|\nabla g(x) - \nabla g(y)\| \leqslant K\|x - y\|, \quad \forall x, y \in B(x^*, \hat{\delta}). \tag{4.29}$$

By hypothesis, ϕ satisfies the K-L inequality at x^*, there exist $\eta \in (0, +\infty]$, a neighborhood U of x^* and a continuous concave function $\gamma : [0, \eta) \to \mathbb{R}_+$ such that for every $x \in U$ with $\phi(x^*) < \phi(x) < \phi(x^*) + \eta$, we obtain

$$\gamma'(\phi(x) - \phi(x^*))\text{dist}(0, \partial_c\phi(x)) \geqslant 1. \tag{4.30}$$

Take $\overline{\delta} > 0$ such that $B(x^*, \hat{\delta}) \subset U$ and set $\delta := \min \frac{1}{2}\{\overline{\delta}, \hat{\delta}\} > 0$. Considered $\lim_{k\to\infty}\phi(x^k) = \phi(x^*)$, it follows from (4.28) that $\lim_{k\to\infty} d^k = 0$. Then, there exists $k_1 \in \mathbb{N}$ such that

$$\|y^k - x^k\| = \|d^k\| \leqslant \delta, \quad \forall k \geqslant k_1.$$

Moreover, for all $k \geqslant k_1$ with $x^k \in B(x^*, \delta)$ we have that

$$\|y^k - x^*\| \leqslant \|y^k - x^k\| + \|x^k - x^*\| \leqslant 2\delta \leqslant \hat{\delta}.$$

Note that, for all $k \geqslant k_1$ such that $x^k \in B(x^*, \delta)$. We obtain $x^k, y^k \in B(x^*, \delta)$, using (4.29) we conclude that

$$\|\nabla g(y^k) - \nabla g(x^k)\| \leqslant K\|y^k - x^k\|.$$

Remember that $\nabla g(y^k) = w^k$, $\nabla g(x^k) - w^k \in \partial_c \phi(x^k) \subseteq \nabla g(x^k) - \partial h(x^k) = \{\nabla g(x^k) - v : v \in \partial h(x^k)\}$ and take $x^{k+1} - x^k = (1+\lambda_k)(y^k - x^k)$, we have:

$$\begin{aligned} \operatorname{dist}(0, \partial_c \phi(x^k)) &\leqslant \| \nabla g(y^k) - w^k\| \\ &= \| \nabla g(y^k) - \nabla g(x^k)\| \\ &\leqslant \frac{K}{1+\lambda_k}\|x^{k+1} - x^k\|, \end{aligned} \tag{4.31}$$

for all $k \geqslant k_1$ with $x^k \in B(x^*, \delta)$

$$L := \max_{\lambda \geqslant 0} \frac{K(1+\lambda)}{\frac{\sigma}{2} + \rho\lambda^2} > 0. \tag{4.32}$$

Observe that $\lim_{l\to\infty} x^{k_l} = x^*$, $\lim_{k\to\infty} x^k = \phi(x^*)$ and $\phi(x^k) > \phi(x^*)$, for all $k \geqslant k_0$ and ϕ is continuous, we can take $N \geqslant \max\{k_0, k_1\}$ satisfies

$$x^N \in B(x^*, \delta) \subset U, \quad \phi(x^*) < \phi(x^N) < \phi(x^*) + \eta. \tag{4.33}$$

In addition, due to $\gamma(0) = 0$, we can also assume that $N \geqslant \max\{k_0, k_1\}$ such that

$$\|x^N - x^*\| + L\gamma(\phi(x^N) - \phi(x^*)) < \delta. \tag{4.34}$$

However, for all such that $x^k \in B(x^*, \delta) \subset U$, (4.30) and (4.31) it follows that

$$\gamma'(\phi(x^k) - \phi(x^*)) \geqslant \frac{1}{\operatorname{dist}(0, \partial_c \phi(x^k))} \geqslant \frac{1+\lambda_k}{K\|x^k - x^{k+1}\|}. \tag{4.35}$$

Being γ a concave function, combining the last inequality (4.35) with (4.28) we have

$$\begin{aligned} \gamma(\phi(x^k) - \phi(x^*)) - \gamma(\phi(x^{k+1}) - \phi(x^*)) &\geqslant \gamma'(\phi(x^k) - \phi(x^*))(\phi(x^k) - \phi(x^{k+1})) \\ &\geqslant \frac{1+\lambda_k}{K\|x^k - x^{k+1}\|} \left(\frac{\sigma}{2} + \rho\lambda_k^2\right) \|d^k\|^2. \end{aligned}$$

Moreover, using $x^{k+1} - x^k = (1+\lambda_k)d^k$, $0 < \lambda_k \leqslant \lambda_{-1}$ and (4.32) we have

$$\begin{aligned} \left(\gamma(\phi(x^k) - \phi(x^*)) - \gamma(\phi(x^{k+1}) - \phi(x^*))\right) &\geqslant \frac{\|x^k - x^{k+1}\|(1+\lambda_k)K}{K^2\|x^k - x^{k+1}\|^2} \left(\frac{\sigma}{2} + \rho\lambda_k^2\right) \|d^k\|^2 \\ &= \frac{\|x^k - x^{k+1}\|}{K(1+\lambda_k)} \left(\frac{\sigma}{2} + \rho\lambda_k^2\right) \\ &= \|x^k - x^{k+1}\| \frac{\left(\frac{\sigma}{2} + \rho\lambda_k^2\right)}{K(1+\lambda_k)} \\ &= \frac{\|x^k - x^{k+1}\|}{L}, \end{aligned}$$

for all $k \geqslant N$ such that $x^k \in B(x^*, \delta)$. Let's prove next that $x^k \in B(x^*, \delta)$ for all $k \geqslant N$. When $k = N$, the statement is true follows from the equation in (4.33). Otherwise, suppose that $x^k \in B(x^*, \delta)$ for all $k = N+1, \cdots, N+p-1$ for some $p \geqslant 2$. Since $\phi(x^k) > \phi(x^*)$ for all $k \geqslant k_0$, consider (4.28), (4.33) and $N > \max\{k_0, k_1\}$, we obtain:

$$\phi(x^*) < \phi(x^{k+1}) < \phi(x^k) < \phi(x^k) + \eta.$$

For all $k = N+1, \cdots, N+p-1$. Now, we need to show that $x^{N+p} \in B(x^*, \delta)$. Then, by using the induction hypothesis, triangular inequality and (4.33), we concluded:

$$\begin{aligned} \|x^{N+P} - x^*\| &\leqslant \|x^N - x^*\| + \sum_{i=1}^{p} \|x^{N+i} - x^{N+i-1}\| \\ &\leqslant \|x^N - x^*\| + L \sum_{i=1}^{p} \left[\gamma(\phi(x^{N+i-1}) - \phi(x^*)) - \gamma\big(\phi(x^{N+i}) - \phi(x^*)\big)\right]. \end{aligned} \tag{4.36}$$

Summing up the inequality (4.36) and taking into account that $\gamma(\phi(x^{N+i}) - \phi(x^*)) \geqslant 0$ and (4.34), it follows

$$\begin{aligned} \|x^{N+P} - x^*\| =& \|x^N - x^*\| + L\gamma(\phi(x^N) - \phi(x^*)) - L\gamma(\phi(x^{N+p}) - \phi(x^*)) \\ &\leqslant \|x^N - x^*\| + L\gamma(\phi(x^N) - \phi(x^*)), \end{aligned}$$

and the induction completed. At long last, considering that $x^k \in B(x^*, \delta)$ for all $k \geqslant N$, use the same argument above along with (4.33) and (4.34), it follows:

$$\begin{aligned} \sum_{k=N}^{N+p} \|x^{k+1} - x^k\| &\leqslant \sum_{k=N}^{N+p} L\left(\gamma(\phi(x^k) - \phi(x^*)) - \gamma\ (\phi(x^{k+1}) - \phi(x^*)\right) \\ &= L\gamma\ (\phi(x^N) - \phi(x^*)) - K\gamma(\phi(x^{N+p} - \phi(x^*)) \\ &\leqslant L\gamma(\phi(x^N) - \phi(x^*)) < \delta. \end{aligned} \tag{4.37}$$

Taking the limit in (4.37) a p goes ∞ we obtain

$$\sum_{k=N}^{N+p} \|x^{k+1} - x^k\| < \infty.$$

Hence, $(x^k)_{k\in\mathbb{N}}$ is a Cauchy sequence. For other hand, due to x^* being a cluster point of $(x^k)_{k\in\mathbb{N}}$, then the whole sequence $(x^k)_{k\in\mathbb{N}}$ converge to x^*. Therefore, it is enough to consider the Theorem 4.5.3 and the proof is complete. □

Chapter 5

Image Representation Models

5.1 Image Representation Models

Image problems can be seen as representations of linear systems as follows:

$$Ax = b + \epsilon,$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are known, $\epsilon \in \mathbb{R}^m$ it is a noise vector, unknown, and x the value to be estimated. The reconstruction problem is solved by minimizing an energy function E, such that:

$$\hat{x} = \arg\min_x E(x).$$

According to [23], $\hat{x}$ can be set by,

$$\hat{x} = \arg\min_x \frac{\mu}{2}\|x - b\|^2 + TV_f(x), \quad (5.1)$$

where μ is the regularization parameter and $\|x-b\|^2$ is a fidelity term that denotes the difference between the original image and the noisy images and $TV_f(x) = \sum_{i=1}^{n} f(\|(\nabla x)_i\|; a)$. The function $f : \mathbb{R} \to \mathbb{R}$ is the regularization (or penalty) function parameterized by variable a selected to guarantee convexity of the energy function, and $\|(\nabla x)_i\| = \sqrt{(D_h x)_i^2 + (D_v x)_i^2}$, with the linear operators $D_h, D_v \in \mathbb{R}^{m\times n}$ representing finite difference approximations of first-order horizontal and vertical partial derivatives, respectively. If $\phi(x) = x$, then $TV_f(x) = TV(x)$. For more information on the problem in (5.1), they can be found in [23]. When $\mu = 0$, there is no regularization, the result is identical to the noisy image. For $\mu \to +\infty$, regularization dominates, resulting in very smoothed images, which can lead to loss of details. Hence, the value of μ should be adjusted according to the noise level:

the more noise, the higher it should be μ, but too much noise can blur details. We can observe that the choice of the penalty function and the regularization parameter define the balance between noise removal and preservation of details. This model is fundamental for the application of the algorithm discussed in this work.

5.2 Noise removal in non-convex problems

In the first chapters, we provide an improved variant of the DCA, the BDCA Algorithm, was proposed by Aragon Artacho in [4] to solve problems where only the first component is differentiable. When this concept is questioned, Ferreira in [5] introduce the nmBDCA, a variant that employs a non-monotonic search and allows the objective function to increase, provided a certain parameter is used. Now, we present a non-convex TV model formulated as a DC programming problem.

The application of the (convex) TV model for image noise removal allows image reconstruction by various methods. However, studies such as those by Lanza et al. [30] and Nikolova et al. [35], have demonstrated that non-convex regularization techniques can outperform convex models in specific scenarios—particularly in preserving sharp transitions and fine edges in piecewise constant images. In the formulation presented in equation (5.1), the function f defines the regularization behavior and directly influences the convexity of the overall objective. Depending on its properties, f can yield either a convex or a non-convex (and possibly non-differentiable) optimization problem. Several effective penalty functions for image denoising based on this formulation are discussed in Selesnick and Bayram [33].

Under certain conditions, namely when $f : \mathbb{R}^+ \to \mathbb{R}$ is concave and non-decreasing, the model becomes a DC (Difference of Convex functions) problem. As shown by De Oliveira and Tcheou [36], the functional

$$\mathsf{TV}_f(x) = \sum_{i=1}^{n} f\left(\|(\nabla x)_i\|\right)$$

can be decomposed as a difference between two convex functions:

$$\mathsf{TV}_f(x) = \tau\mathsf{TV}(x) - \left[\tau\mathsf{TV}(x) - \mathsf{TV}_f(x)\right],$$

where the inner term is convex provided that $\tau \geqslant f'_+(0) \geqslant 0$. This decomposition enables the use of DC programming techniques, offering a flexible and efficient framework for solving non-convex image denoising problems. The problem considered is

$$\arg\min_{x\in\mathbb{R}^n} \frac{\mu}{2}\|x-b\|^2 + \mathsf{TV}_f(x) = g(x) - h(x), \tag{5.2}$$

where g and h are a convex functions, this is (5.2) is a DC problem to be solved.

Proposition 5.2.1. *Let $\mu > 0$ and $b \in \mathbb{R}^n$ be fixed, and $\sigma > 0$ a strongly convex parameter. Assume that $f : \mathbb{R}_+ \to \mathbb{R}$ is a concave and non-decreasing function . Then, $\phi(x) = \frac{\mu}{2}\|x-b\|^2 + \mathsf{TV}_f(x)$ is a DC function with strongly convex DC components*

$$g(x) = \frac{\mu+\sigma}{2}\|x-b\|^2 + \tau\mathsf{TV}(x); \tag{5.3}$$

and

$$h(x) = \frac{\sigma}{2}\|x-b\|^2 + \tau\mathsf{TV}(x) - \mathsf{TV}_f(x), \tag{5.4}$$

for every $\tau \geqslant f'_+(0)$.

Proof. Since f is concave and non-decreasing, then f' is non-increasing in an interval $I \subset \mathbb{R}_+$. Hence, for any $a, b \in I$, we have

$$f'_+(a) \geqslant f'_+(b), \quad \forall 0 \leqslant a \leqslant b. \tag{5.5}$$

Consider $\phi(x) = g(x) - h(x)$, with

$$g(x) = \frac{\mu+\sigma}{2}\|x-b\|^2 + \tau\mathsf{TV}(x)$$

and

$$h(x) = \frac{\sigma}{2}\|x-b\|^2 + \tau\mathsf{TV}(x) - \mathsf{TV}_f(x),$$

for a given parameter $\sigma > 0$ and $\tau \geqslant f'_+(0) > 0$, because f is concave. Now, we have to show that g and h are strongly convex functions with modulus $\sigma > 0$. In fact, g is strongly convex because it is the sum of the convex function $g_1(x) = \tau\mathsf{TV}(x)$ and the strongly convex function $g_2(x) = \frac{\mu+\sigma}{2}\|x-b\|^2$, by Proposition 1.1.2. For h observe that $h_1(x) = \tau\mathsf{TV} - \mathsf{TV}_f(x)$ is convex, we will show next, and $h_2(x) = \frac{\sigma}{2}\|x-b\|^2$ is strongly convex, by Proposition 1.1.2 h is strongly convex. It follows the definition of the function TV_f in (5.1) that

$$\tau\mathsf{TV}(x) - \mathsf{TV}_f(x) = \psi(\mathsf{TV}(x)),$$

where $\psi(x) = \tau x - f(x)$. Since $\psi''(x) = -f''(x)$ and $f(x)$ is concave, $f''(x) < 0$, then $-f''(x) > 0$, we have that $\psi(x)$ is convex. In addition, consider $\tau \geqslant f'_+(0)$ and (5.5). Hence, $\psi'_+(x) = \tau - f'_+(x) \geqslant f'_+(0) - f'_+(x) \geqslant 0$, for all $x \geqslant 0$. Consequently, we find that $\psi(x)$ is non-decreasing on $x \geqslant 0$ and a convex function. Since $\mathsf{TV}(x)$ is a convex function, by Proposition 1.1.3, we obtain that $\psi(\mathsf{TV}(x))$ is a convex function. Therefore, the demonstration is complete. □

In the previous proposition, we showed that the non-convex TV model can be formulated as a DC problem, with g and h being non-differentiable. Thus, we can apply a DC algorithm, in particular, the nmBDCA to enhance images using the non-convex TV model.

In next section, nmBDCA will be presented for solving the problem in (5.2). The following algorithm is an implementation of nmBDCA for the context of noise removal, which begins with obtaining and vectorizing the image to be analyzed. The configuration parameters, which can vary based on the problem, are then defined. These parameters define the base parameter ν for generating the nmBDCA sequence, the fidelity parameter μ of the image model (5.2), the strongly convex function $\sigma > 0$, and the maximum number of reconstruction iterations (`max_iter`), and the Gaussian noise variance (ϵ) inserted into the image (b).

Table 5.1: Penalty functions for $a > 0$

	$f_{\log}$	f_{atan}	$f_{\exp}$
$f_a(r)$	$\frac{\log(1+ar)}{a}$	$\frac{\frac{\arctan(1+ar)}{\sqrt{3}} - \frac{\pi}{6}}{a\sqrt{\frac{3}{2}}}$	$\frac{1-e^{-ar}}{a}$
$f_a(r)$	$\frac{1}{1+ar}$	$\frac{1}{1+ar+a^2r^2}$	$\frac{1}{e^{ar}}$

5.3 Solving the Subproblems

The Fast Iterative Shrinkage Thresholding Algorithm (FISTA) is a faster version of the Iterative Shrinkage Thresholding (ISTA) developed by Daubechies et al in [37]. Beck and Teboulle in [40] proposed ISTA to tackle the following problem.

$$\min\{F(x) = f(x) + g(x) : x \in \mathbb{R}^n\}, \tag{5.6}$$

where:

- $g : \mathbb{R} \to \mathbb{R}^n$ is a convex function possibly non-differentiable;
- $f : \mathbb{R} \to \mathbb{R}^n$ is a differentiable convex function, continuously differentiable with Lipschitz gradient, this is:
  $\|\nabla f(x) - \nabla f(y)\| \leqslant K\|x - y\|$, for all $x, y \in \mathbb{R}^n$ and with $K > 0$.

The problem (5.6) has a solution; for more information, see [7].

The Fast Projected Gradient (FGP) was considered by Bech and Teboulle in [41]. It is an improved version of FISTA, where FISTA techniques are used with the gradient projection method by Nesterov [42]. For more information on this method, see [7].

Using FGP with nmBDCA requires solving a discrete penalized version of the TV-based model at each iteration k

$$\min_{x \in \mathbb{R}^n} \|x - s^k\|^2 + 2\lambda \mathrm{TV}(x), \tag{5.7}$$

where λ is a any parameter and s^k is a corrupted image. The following proposition guarantees the possibility of applying (FGP) to the resolution of subproblems of nmBDCA.

Proposition 5.3.1. *Let* g *and* h *be the DC components described according to* (5.3) *and* (5.4) *of the non-convex model in* (5.2). *Then, the subproblem* (4.1) *can be written as* (5.7).

Proof. See [7, Appendix B]. □

5.3.1 Convergence guarantee to nmBDCA

It follows from Proposition 4.2.1 that the non-monotone line search in Algorithm 4 is well-defined, this is, it ends in a finite number of iterations. For this consider $\phi(x) = g(x) - h(x)$ as in (5.2). Moreover, by Theorem 4.5.1, we have that the cluster points of the nmBDCA, if any, will be a critical point of the (5.2).

5.4 Numerical experiments

This section presents the results obtained from image reconstructions in MATLAB. For the experiments with Test images in 5.2, we used the same machine is used is all implementations: a laptop with the Windows 11 Home Single Language Version 21H2 operating system, an Intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz, 1.19 GHz, 8 GB of

RAM. Now, for the experiments with Medical Images in 5.6, we performed on a desktop computer running Windows 10 Home Single Language Version 21H2 operating system, equipped with an Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz, 2.30 GHz, 16 GB of RAM. Reconstructions remove Gaussian noise with different variances added to the original image.

Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) metrics evaluate the reconstructions and check image contrast, resolution, noise, and luminance. These metrics are widely used in digital image processing. The objective of the experiments is to analyze the performance of the FISTA, DCA, and nmBDCA techniques. The regularization parameter μ was modified for each noise variance to determine the μ for the best SSIM.

Table 5.2: Images used in Section 5.5.

Image	Description	Characteristics	Pixels
1	Sun	black and white	1440×1440
2	Qr code	black and white	736×736
3	Eiffel	black and white	675×1200
4	CM	black and white	1600×1200

Table 5.3: Images used in Section 5.6.

Image	Description	Characteristics	Pixels
1	IM 1	black and white	550×406
2	IM 2	black and white	498×403
3	IM 3	black and white	600×423

5.5 Test images

Now, we will reconstruct the images of the table corrupted by Gaussian noise with variance 0.01, 0.10 and 0.20 are performed.

Table 5.5 presents the values of μ that achieve the best SSIM for each technique, along with the parameters used to configure the implemented methods. It is worth noting that changing the noise level in each image affects the regularization parameter μ, and in the case of nmBDCA, the parameter ν is particularly sensitive as it influences the algorithm's convergence time. The nmBDCA relies on specific parameters (β, λ_0, τ, and ν) which are tuned for each noise level during the Armijo line search. Adjusting these parameters directly impacts the processing time, since they affect how the search is performed.

The parameter λ_k, starting with λ_0 in the first iteration, determines the amount of regularization applied at the initial point, which can influence the time needed to complete the search. The parameter τ serves to adjust and balance the effect of λ_k during the search. The sequence generated by the algorithm depends on ν, which controls when the objective function value is allowed to increase during the process, ensuring this only happens when necessary. This sequence can be configured to give more or less weight to the influence of τ and λ_k. Finally, the parameter β controls the reduction of the regularization factor throughout the search. It can significantly impact nmBDCA's performance, as it defines the step size the algorithm takes to find the point where the objective function decreases most rapidly. Therefore, when we analyze the table, we can see that the methods using DC functions outperform FISTA by almost 100%, which shows the performance difference between the TV model used by FISTA and the non-convex TV model used by the DC methods. The following figures and tables present a comprehensive evaluation of the performance of each method, highlighting both the quantitative metrics and the visual reconstruction quality results achieved by each method at different noise levels. Despite the insertion of Gaussian noise into the original images, the techniques were able to produce acceptable reconstructions. However, the presence of gray pixels can be observed in some cases, particularly in the images obtained using FISTA (Example: Figures 5.2 (c), 5.2 (h),5.2 (m) and 5.2 (r)). This effect becomes more pronounced as the noise level increases and visually illustrates the qualitative difference between the convex and non-convex models, emphasizing the greater robustness of the non-convex approach and reinforcing its suitability for image reconstruction tasks.

Table 5.4: Evaluation metrics (FISTA's tolerance is 5×10^{-10} and DC techniques tolerance is 5×10^{-4}).

Variance 0.01										
Image	**Technique**	μ	β	λ_0	τ	ν	PSNR	SSIM	Iter.	CPU Time (s)
Sun	FISTA	0.2	-	-	-	-	23.6147	0.7121	10000.00	908.8376
	DCA	5	-	-	-	-	**24.7986**	**0.7557**	21.00	451.8472
	nmBDCA	5	0.05	0.9	0.15	50	24.7964	0.7555	**17.00**	**369.9986**
Qr code	FISTA	0.15	-	-	-	-	26.7116	0.7543	10000.00	200.4989
	DCA	5	-	-	-	-	27.0117	0.7576	9.00	15.2581
	nmBDCA	4	0.05	0.9	0.15	50	**27.2084**	**0.7586**	**5.00**	**10.6773**
Eiffel	FISTA	0.15	-	-	-	-	24.7135	0.8668	10000.00	975.9296
	DCA	9	-	-	-	-	**26.6410**	0.9099	14.00	**156.9336**
	nmBDCA	9	0.05	0.9	0.15	50	26.6401	**0.9278**	**12.00**	224.9976
CM	FISTA	0.09	-	-	-		31.3728	**0.9519**	10000.00	1565.9015
	DCA	11	-	-	-	-	32.6322	0.9278	13.00	224.9976
	nmBDCA	10	0.05	0.9	0.15	60	**32.8828**	0.9435	**9.00**	**222.4932**

The comparative analysis of the FISTA, DCA, and nmBDCA methods for image reconstruction under Gaussian noise with variances $\sigma^2 = 0.01$, 0.10, and 0.20 reveals significant differences in reconstruction quality, computational cost, and robustness to noise. At $\sigma^2 = 0.01$, the nmBDCA algorithm already outperforms the others in several cases. For instance, in the *Qr code* image, nmBDCA achieved 27.2084 dB PSNR and 0.7586 SSIM, surpassing DCA (27.0117 dB, SSIM 0.7576) and FISTA (26.7116 dB, SSIM 0.7543), with only 5 iterations and a CPU time of 10.8773 s, whereas FISTA required 10,000 iterations and 200.4989 s. For the *CM* image, nmBDCA reached 32.8828 PSNR and 0.9458 SSIM with only 9 iterations and 222.4932 s, again outperforming both FISTA and DCA. At $\sigma^2 = 0.10$, noise effects became more evident. The FISTA method experienced a performance drop, achieving only 14.7574 dB PSNR and 0.1577 SSIM in the *Sun* image, compared to nmBDCA (19.3479 dB, 0.5859) and DCA (19.3467 dB, 0.5858), both converging to a number close to the iterations. In the *Eiffel* image, FISTA yielded 15.1455 dB and 0.2042 SSIM, while nmBDCA reached 17.1184 dB and 0.6668 SSIM in 419.1017 s, outperforming FISTA, which required 62.4168 s and 1983 iterations. At the highest noise level, $\sigma^2 = 0.20$, the methods behaviors diverge further. FISTA showed a significant quality drop, in the *Sun* image with 11.5786 dB PSNR and 0.0889 SSIM, while both DCA and nmBDCA achieved 18.1905 dB and SSIM near 0.5734 in fewer than

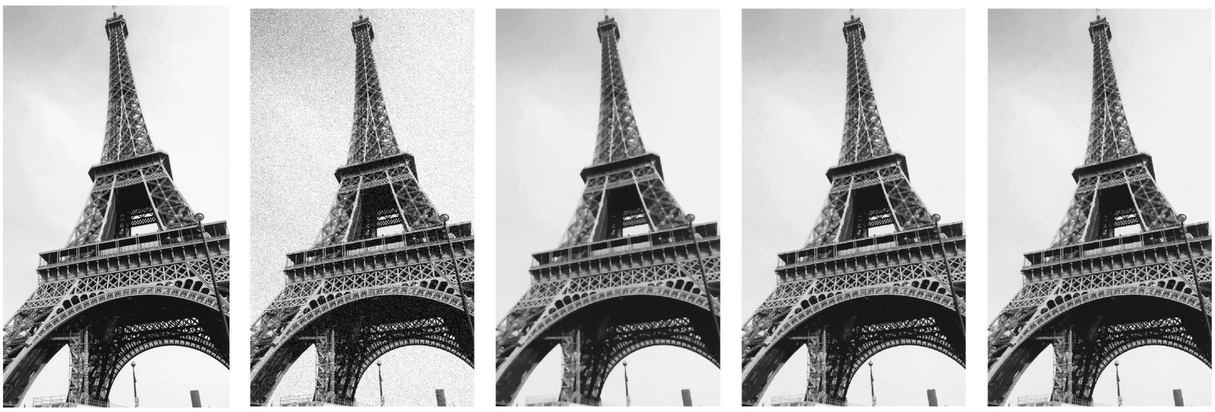
Variance 0.10										
Image	**Technique**	μ	β	λ_0	τ	ν	PSNR	SSIM	Iter.	CPU Time (s)
Sun	FISTA	0.05	-	-	-	-	14.7574	0.1577	1535.00	**134.7577**
	DCA	0.75	-	-	-	-	19.3467	0.5858	22.00	1023.6578
	nmBDCA	0.75	0.05	0.6	0.10	150	**19.3479**	**0.5859**	**21.00**	1022.3306
Qr code	FISTA	0.10	-	-	-	-	16.7203	0.3010	7293.00	133.9924
	DCA	1.2	-	-	-	-	17.3563	0.6746	49.00	99.3265
	nmBDCA	1.0	0.05	0.3	0.5	100	**17.4162**	**0.7058**	**32.00**	**83.1411**
Eiffel	FISTA	0.05	-	-	-	-	15.1455	0.2042	1983.00	**62.4166**
	DCA	0.98	-	-	-	-	17.1168	**0.6668**	41.00	488.2389
	nmBDCA	0.98	0.05	0.3	0.10	250	**17.1184**	**0.6668**	**36.00**	419.1017
CM	FISTA	0.08	-	-	-	-	17.2035	0.1281	3695.00	**276.2876**
	DCA	0.58	-	-	-	-	19.9242	0.7837	33.00	1605.0559
	nmBDCA	0.58	0.05	0.6	0.20	350	**19.9300**	**0.7838**	**25.00**	1226.7555

Variance 0.20										
Image	**Technique**	μ	β	λ_0	τ	ν	PSNR	SSIM	Iter.	CPU Time (s)
Sun	FISTA	0.55	-	-	-	-	11.5788	0.0889	1264.00	**94.3366**
	DCA	1.2	-	-	-	-	**18.1905**	0.5732	51.00	1729.7572
	nmBDCA	1.2	0.05	0.5	0.20	400	**18.1905**	**0.5734**	**43.00**	1559.8656
Qr code	FISTA	0.52	-	-	-	-	14.1127	0.6785	10000.00	154.7320
	DCA	0.64	-	-	-	-	14.5462	0.6807	35.00	70.2442
	nmBDCA	0.64	0.05	0.9	0.10	75	**14.5554**	**0.6813**	**26.00**	**66.6629**
Eiffel	FISTA	0.98	-	-	-	-	14.9618	0.6276	10000.00	273.9737
	DCA	2.0	-	-	-	-	14.3324	0.2344	140.00	865.2360
	nmBDCA	1.0	0.05	0.9	0.15	500	**15.4091**	**0.6374**	**23.00**	**248.0373**
CM	FISTA	0.94	-	-	-	-	16.8812	0.7176	10000.00	**685.8192**
	DCA	1.0	-	-	-	-	16.9617	0.7153	48.00	1964.1206
	nmBDCA	1.0	0.05	0.9	0.15	350	**16.9685**	**0.7155**	**33.00**	1390.3263

43 iterations. For the *CM* image, nmBDCA again performed best, reaching 16.9685 dB PSNR and 0.7155 SSIM with 33 iterations and 1390.3283 s of CPU time, whereas FISTA, with $10,000$ iterations, achieved 16.8812 dB and 0.7176 SSIM but at a much higher cost (685.8192 s). Overall, FISTA proved to be limited, especially under heavier noise, requiring a high number of iterations and resulting in lower reconstruction quality. DCA was more efficient and produced good results but was consistently outperformed by nmBDCA. The latter not only achieved the best results in PSNR and SSIM, but also exhibited fast convergence, fewer iterations, and competitive computational performance, making it the most robust and efficient among the three methods evaluated

(a) Sun (b) Noise (c) FISTA (d) DCA (e) nmBDCA

(f) Qr code (g) Noise (h) FISTA (i) DCA (j) nmBDCA

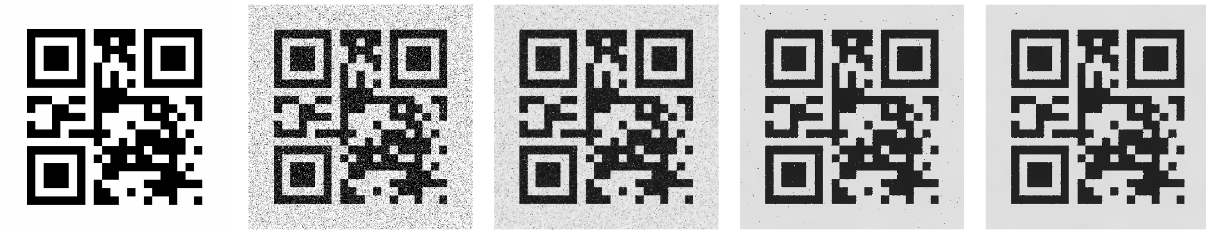


(k) Eiffel (l) Noise (m) FISTA (n) DCA (o) nmBDCA

(p) CM (q) Noise (r) FISTA (s) DCA (t) nmBDCA

Figure 5.1: Results with a variance of 0.01

(a) Sun (b) Noise (c) FISTA (d) DCA (e) nmBDCA

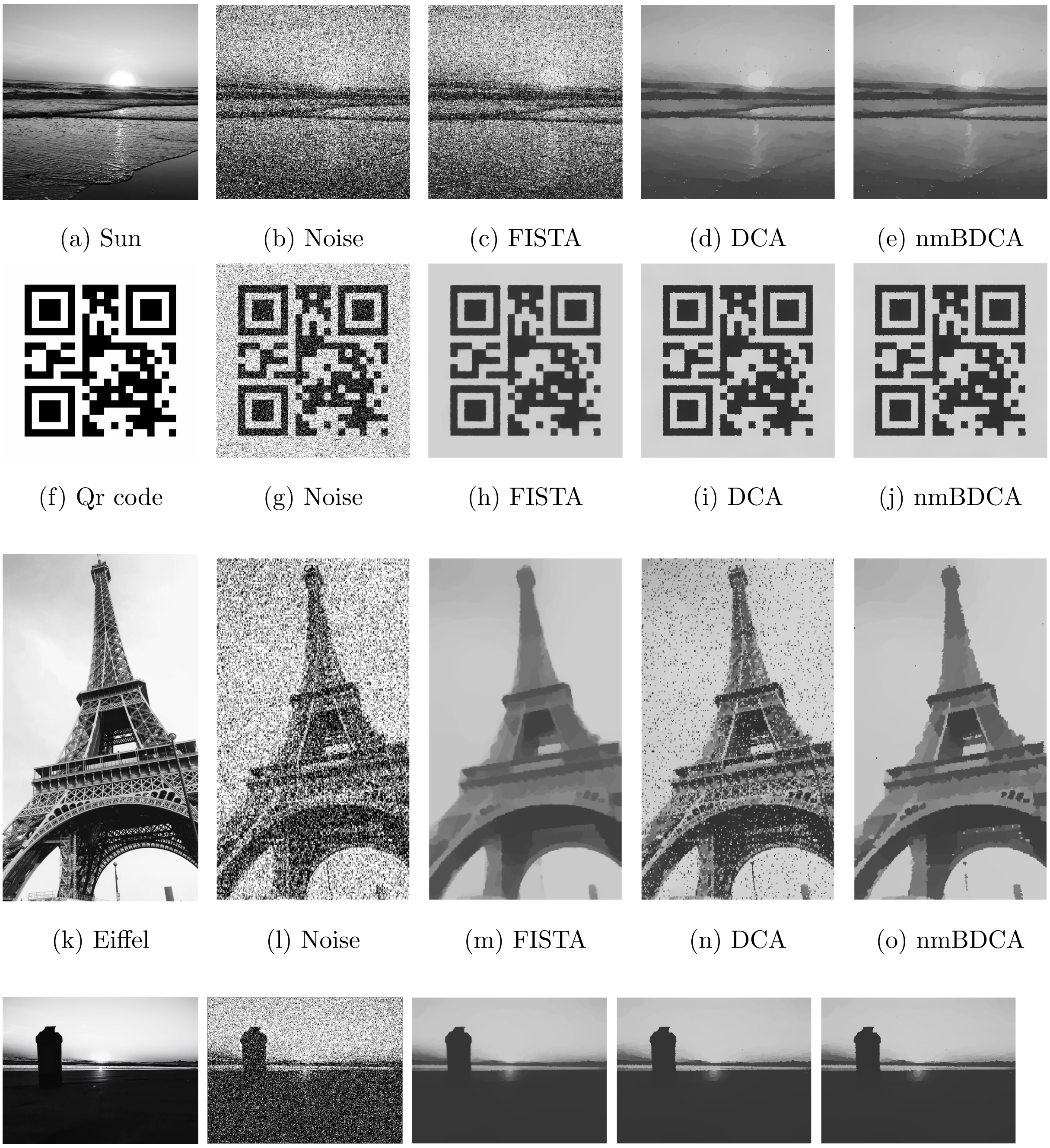


(f) Qr code (g) Noise (h) FISTA (i) DCA (j) nmBDCA

(k) Eiffel (l) Noise (m) FISTA (n) DCA (o) nmBDCA

(p) CM (q) Noise (r) FISTA (s) DCA (t) nmBDCA

Figure 5.2: Results with a variance of 0.10



(a) Sun (b) Noise (c) FISTA (d) DCA (e) nmBDCA

(f) Qr code (g) Noise (h) FISTA (i) DCA (j) nmBDCA

(k) Eiffel (l) Noise (m) FISTA (n) DCA (o) nmBDCA

(p) CM (q) Noise (r) FISTA (s) DCA (t) nmBDCA

Figure 5.3: Results with a variance of 0.20

5.6 Medical images

Similar a section 5.5, this section reconstructs images from Table 5.3 corrupted by Gaussian noise with variances of 0.005, 0.0075 and 0.01. The Figs. 5.7, 5.8 and 5.9, show the results obtained by scanning the parameter μ (constructed by varying the values of μ from 0.01 to 0.99 and from 1 to 12), and allow estimating the best values to obtain of PSNR and SSIM for each technique. It is observed that as the noise variance increases, the suitable range of values for the parameter μ becomes smaller. In this scenario, FISTA shows a narrower effective range compared to DCA and nmBDCA. This suggests that, within this range, small variations in μ significantly affect the performance of the method.

Lower values of μ tend to enhance noise removal but also slow down the algorithm. However, if μ is too small, it may cause distortion in the image pixels. On the other hand, DCA and nmBDCA operate efficiently over similar ranges of μ and are able to reconstruct images with comparable PSNR and SSIM values, while requiring less computation time.

The optimal SSIM measure for FISTA is $\mu < 1$. In most experiments, DCA and nmBDCA have μ greater than 1, showing that the approach has a specific value for each problem. Because the scans were only performed once for each variable, the average of the executions may not accurately reflect the results. The techniques all used the same image. DCA and nmBDCA use non-convex models for image reconstruction, with the following stopping criteria. The algorithm has a limit of 200 iterations, a stop tolerance of 5×10^{-4}, a penalty function $f = f_{atan}$, and a strongly convex function parameter $\sigma = 1$.

The parameters were defined to ensure the proper functioning of the applied techniques. However, for problems different from those analyzed in this study, it is essential to reassess these parameters to ensure the convergence of the methods under the new conditions. With the original stopping criterion, FISTA demonstrated high speed, although it might not always return a solution of optimal quality. Therefore, a stricter stopping tolerance of 5×10^{-10} was adopted, along with a maximum of 10.000 iterations, in order to obtain the best possible result from the method. The final performance metrics were calculated as the average over 100 independent runs with different random seeds, to prevent any potential bias in the results. The parameters used for each technique are presented in Table 5.6. It is observed that the parameter μ varies significantly in magnitude between the convex model ($\mu < 1$) and the non-convex model ($\mu > 1$), highlighting the importance of thoroughly investigating this parameter in the reconstruction process.

The search configuration for the nmBDCA algorithm was defined with fixed parameters: $\beta = 0,05$, $\lambda_0 = 0,9$, and $\tau = 0,15$. The value of ν parameter was changed in some tests to guarantee the lowest CPU time for nmBDCA (this can be noticed in both iterations and CPU time) and demonstrates that each problem requires studying the parameters applied and how laborious this can be. It is possible to observe that the number of iterations and the processing time (CPU) do not always increase proportionally to the level of noise introduced. This can be attributed to the influence of the parameters defined in each technique — in this case, to the parameter ν of the nmBDCA.

The reconstructed CT images are displayed side by side in Figures 5.4, 5.5, and 5.6. Despite the addition of Gaussian noise, the applied techniques were able to reconstruct the images effectively, preserving most of the original details across all tested noise levels. However, even though the visual appearance of the reconstructions is similar, it is noticeable that DCA and nmBDCA, which are based on the non-convex TV model, produce results with less distortion. This becomes evident when comparing the textures of the original and reconstructed images. In the images obtained by FISTA, one can observe the presence of more homogeneous regions — as shown in Figures 5.4 (h), 5.5 (h), and 5.6 (m) and a less effective noise removal, particularly noticeable in Figure 5.5 (c).

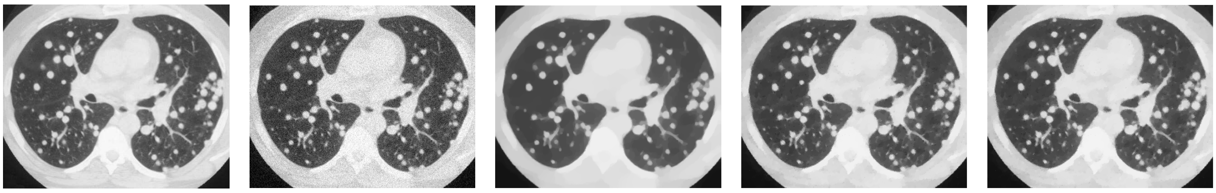
Table 5.5: Evaluation metrics (FISTA's tolerance is 5×10^{-10} and DC techniques tolerance is 5×10^{-4}).

Variance 0.005										
Image	**Technique**	μ	β	λ_0	τ	ν	PSNR	SSIM	Iter.	CPU Time (s)
IM 1	FISTA	0.08	-	-	-	-	28.9367	0.7573	10000.00	61.2180
	DCA	15	-	-	-	-	29.4337	0.7858	9.00	**5.7477**
	nmBDCA	15	0.05	0.9	0.15	60	**29.4363**	**0.7859**	**7.00**	5.9916
IM 2	FISTA	0.17	-	-	-	-	**30.2332**	0.8577	10000.00	62.9855
	DCA	15	-	-	-	-	30.1256	0.8703	9.00	5.3588
	nmBDCA	15	0.05	0.9	0.15	70	30.1342	**0.8709**	**6.00**	**4.7205**
IM 3	FISTA	0.12	-	-	-	-	29.1517	0.7150	10000.00	71.0679
	DCA	7.9	-	-	-	-	29.5499	**0.7166**	14.00	**18.7410**
	nmBDCA	7.9	0.05	0.9	0.15	80	**29.5514**	**0.7166**	**12.00**	19.3372

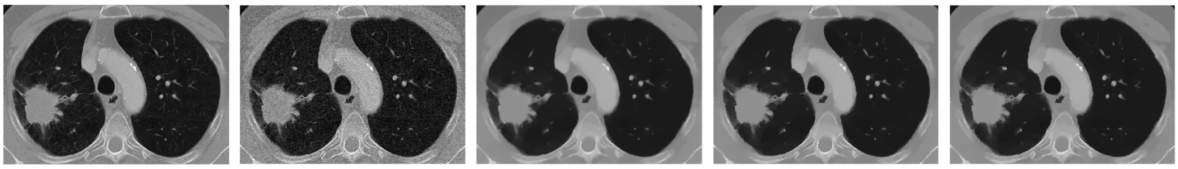
Variance 0.0075										
Image	**Technique**	μ	β	λ_0	τ	ν	PSNR	SSIM	Iter.	CPU Time (s)
IM 1	FISTA	0.09	-	-	-	-	28.3040	0.7433	10000.00	64.6295
	DCA	10.9	-	-	-	-	28.3022	0.7500	12.00	8.8936
	nmBDCA	10.9	0.05	0.9	0.15	80	**28.3065**	**0.7502**	**9.00**	**8.5203**
IM 2	FISTA	0.17	-	-	-	-	**29.9194**	0.8541	10000.00	60.1766
	DCA	10.8	-	-	-	-	29.3092	0.8559	12.00	8.1711
	nmBDCA	10.8	0.05	0.9	0.15	70	29.3178	**0.8563**	**8.00**	**6.8443**
IM 3	FISTA	0.11	-	-	-	-	29.2029	0.7189	10000.00	75.9062
	DCA	8.8	-	-	-	-	29.4938	0.7203	16.00	19.7299
	nmBDCA	8.8	0.05	0.9	0.15	70	**29.4965**	**0.7204**	**13.00**	**19.5268**

Variance 0.01										
Image	**Technique**	μ	β	λ_0	τ	ν	PSNR	SSIM	Iter.	CPU Time (s)
IM 1	FISTA	0.13	-	-	-	-	27.0120	0.6980	10000.00	62.6606
	DCA	10	-	-	-	-	27.5600	0.7283	16.00	11.7963
	nmBDCA	10	0.05	0.9	0.15	80	**27.5637**	**0.7285**	**12.00**	**11.5122**
IM 2	FISTA	0.3	-	-	-	-	28.7559	0.8259	10000.00	57.8191
	DCA	10	-	-	-	-	28.4954	0.8314	14.00	9.0755
	nmBDCA	9	0.05	0.9	0.15	80	**28.5902**	**0.8419**	**11.00**	**7.9845**
IM 3	FISTA	0.12	-	-	-	-	28.7606	0.7099	10000.00	78.9507
	DCA	10.4	-	-	-	-	28.8550	0.7100	18.00	20.3676
	nmBDCA	10.4	0.05	0.9	0.15	80	**28.8602**	**0.7102**	**15.00**	**18.5047**

(a) IM 1 (b) Noise (c) FISTA (d) DCA (e) nmBDCA

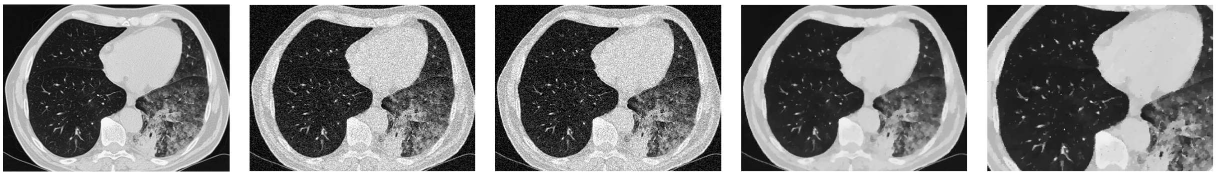


(f) IM 2 (g) Noise (h) FISTA (i) DCA (j) nmBDCA

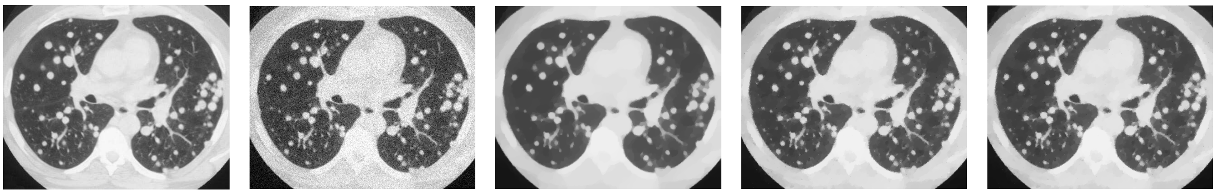


(k) IM 3 (l) Noise (m) FISTA (n) DCA (o) nmBDCA

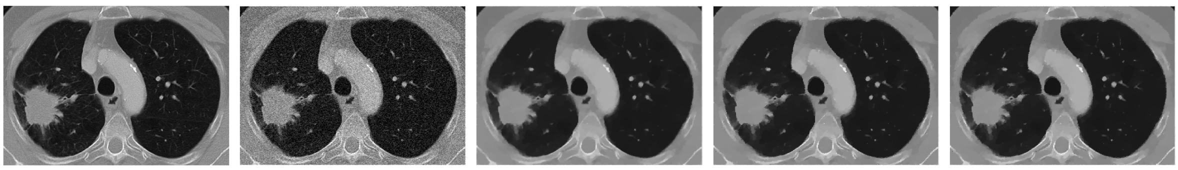
Figure 5.4: Results with a variance of 0.005



(a) IM 1 (b) Noise (c) FISTA (d) DCA (e) nmBDCA

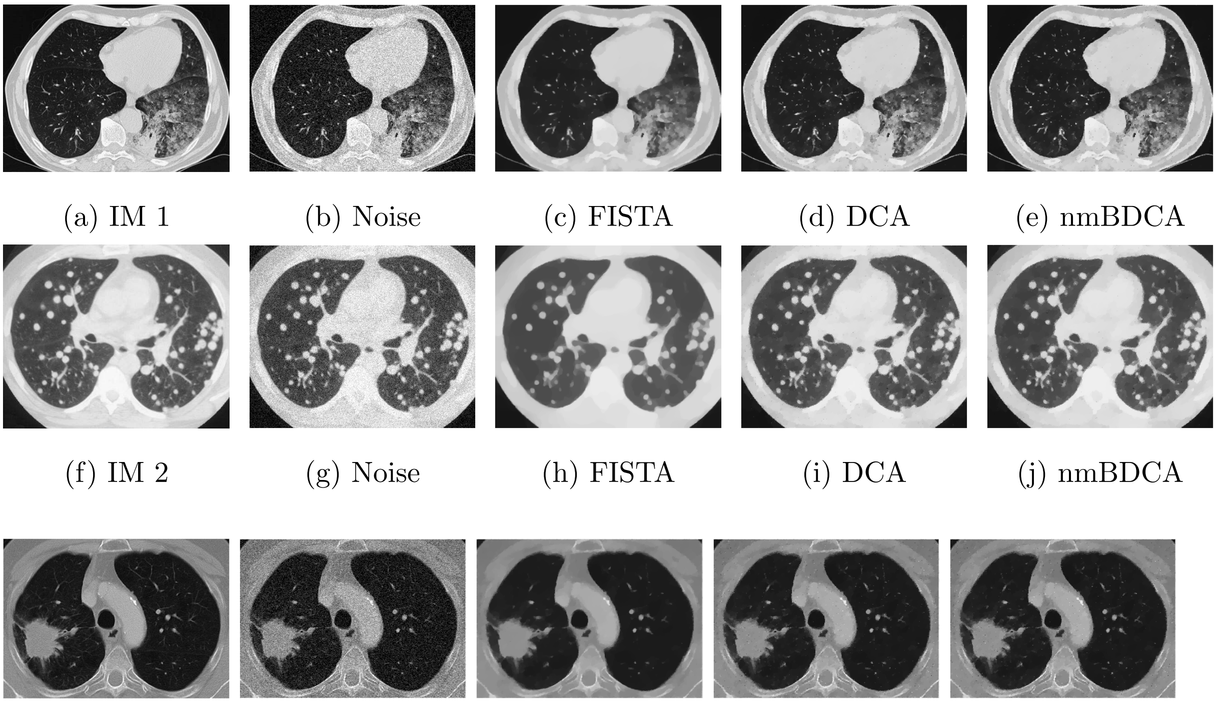


(f) IM 2 (g) Noise (h) FISTA (i) DCA (j) nmBDCA



(k) IM 3 (l) Noise (m) FISTA (n) DCA (o) nmBDCA

Figure 5.5: Results with a variance of 0.0075

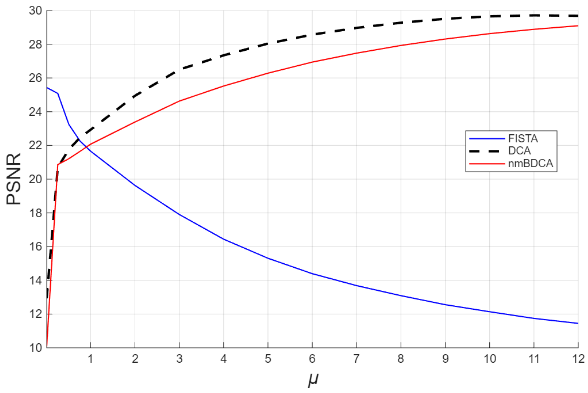


(a) IM 1 (b) Noise (c) FISTA (d) DCA (e) nmBDCA

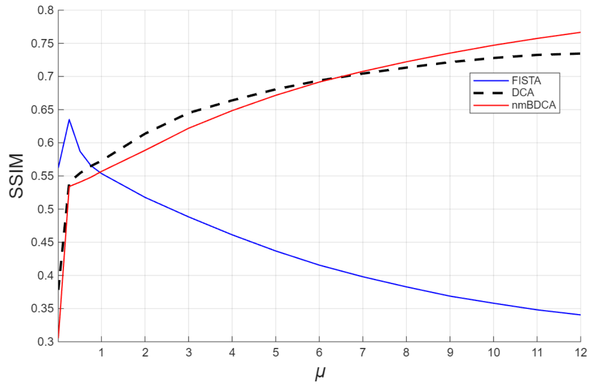
(f) IM 2 (g) Noise (h) FISTA (i) DCA (j) nmBDCA

(k) IM 3 (l) Noise (m) FISTA (n) DCA (o) nmBDCA

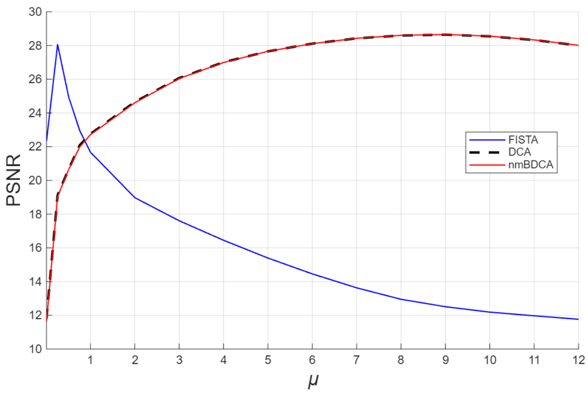
Figure 5.6: Results with a variance of 0.01



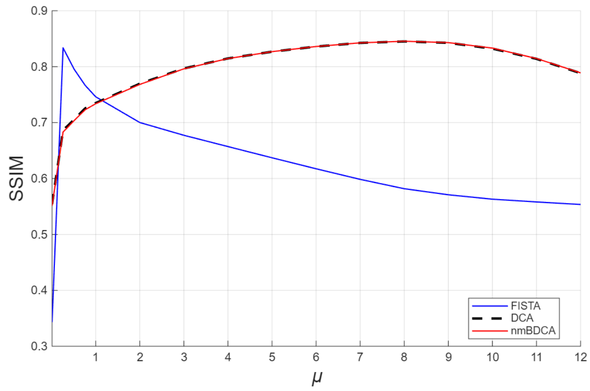
(a) IM 1-PSRN



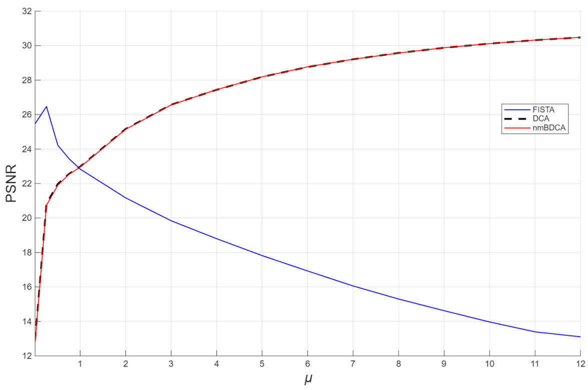
(b) IM 1-SSIM



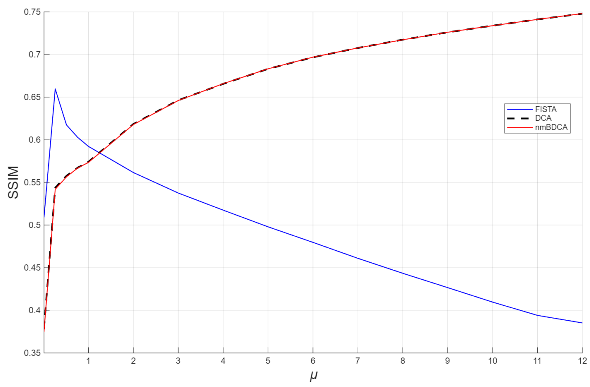
(c) IM 2-PSRN



(d) IM 2-SSIM

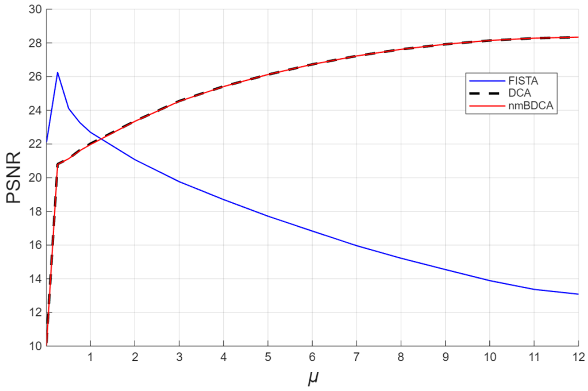


(e) IM 3-PSRN

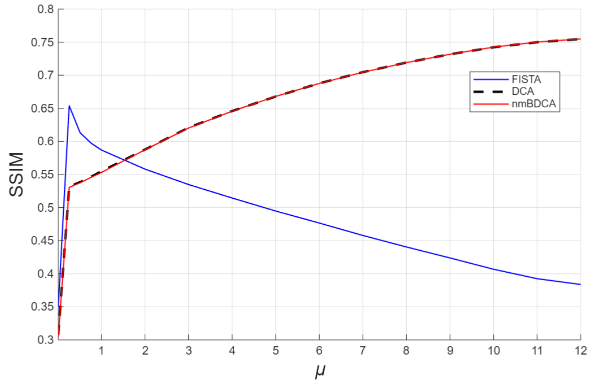


(f) IM 3-SSIM

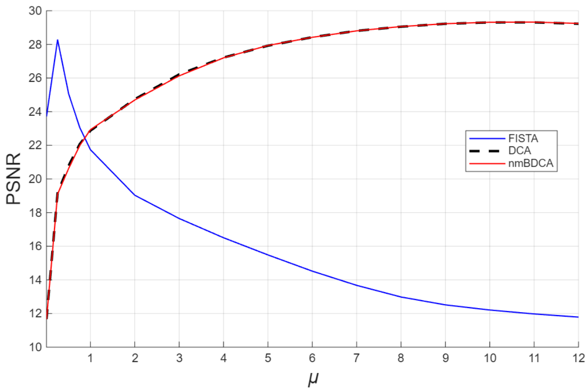
Figure 5.7: Results with a variance of 0.005



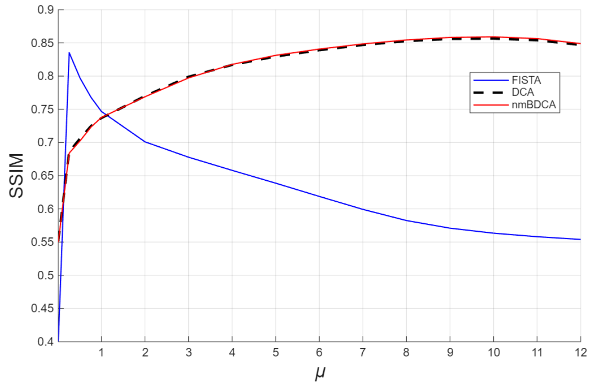
(a) IM 1-PSRN



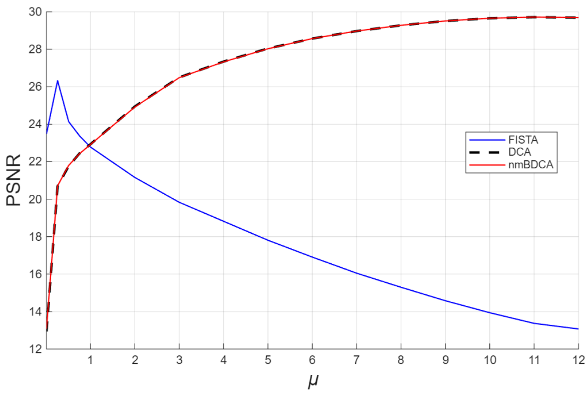
(b) IM 1-SSIM



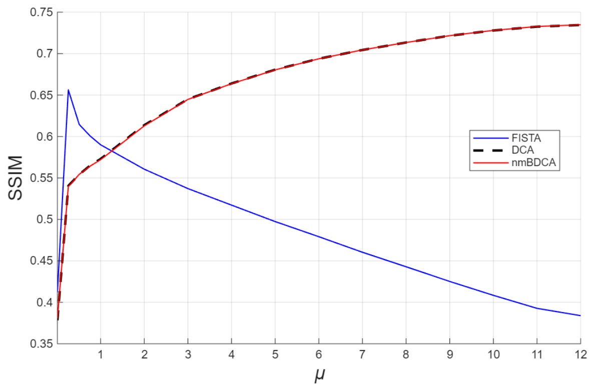
(c) IM 2-PSRN



(d) IM 2-SSIM

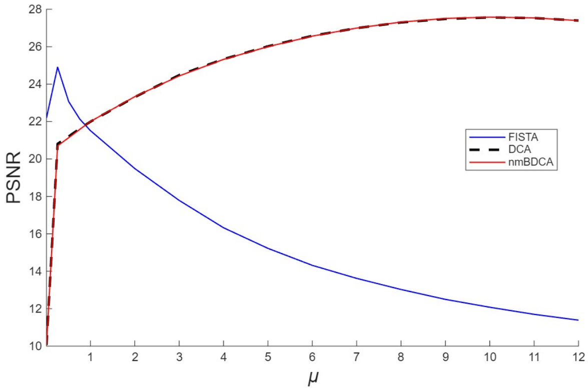


(e) IM 3-PSRN

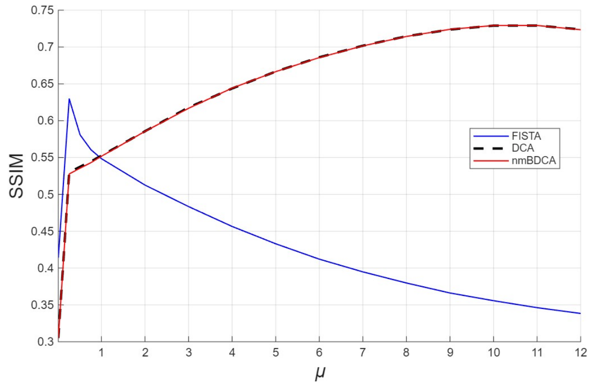


(f) IM 3-SSIM

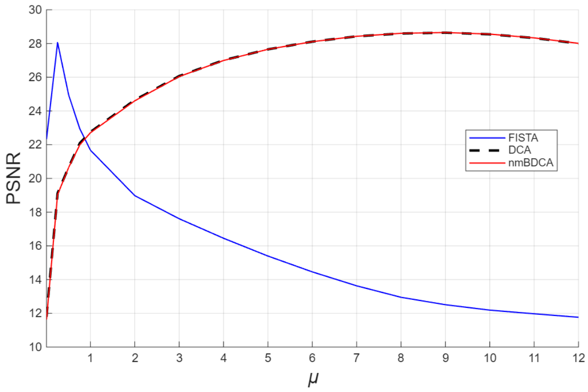
Figure 5.8: Results with a variance of 0.0075



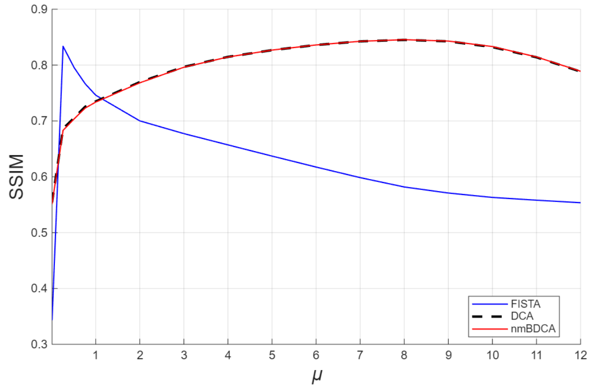
(a) IM 1-PSRN



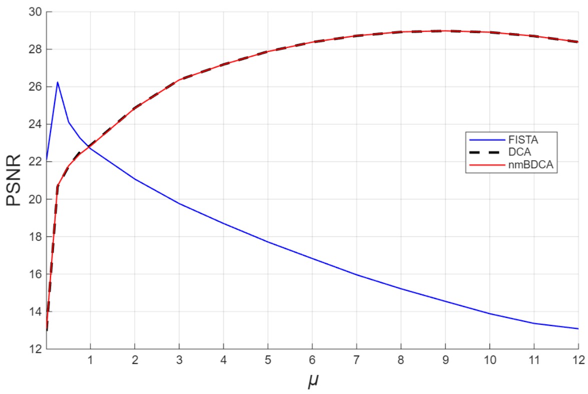
(b) IM 1-SSIM



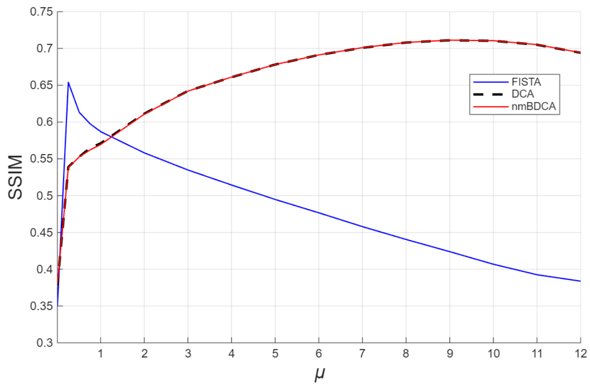
(c) IM 2-PSRN



(d) IM 2-SSIM



(e) IM 3-PSRN



(f) IM 3-SSIM

Figure 5.9: Results with a variance of 0.01

Conclusion

This study used the DCA, BDCA, and nmBDCA algorithms to solve non-convex optimization problems expressed as the difference of convex (DC) functions. Under the right circumstances, theoretical analysis guaranteed these methods' validity and convergence.

These DC-based methods, in particular nmBDCA, showed better performance in terms of convergence speed and reconstruction quality when applied to the problem of image reconstruction from noisy data. This demonstrates how well and reliably DC programming techniques can handle crucial image restoration and enhancement tasks, which are crucial in a number of domains like signal processing, security, and medical imaging. Because DC algorithms can handle non-convex objective functions, regularization can be more precisely controlled, leading to more accurate reconstructions, particularly when noise levels are high.

By examining the nmBDCA algorithm's performance in larger problem classes and under various functional frameworks, future research seeks to increase the algorithm's applicability. This will demonstrate even more the adaptability and strength of DC-based techniques in handling challenging non-convex optimization problems outside of image processing.

Bibliography

[1] Bagirov, A., Karmitsa, N., and Mäkelä, M. M. *Introduction to Nonsmooth Optimization: Theory, Practice and Software*. Springer. Vol. 12, (2014).

[2] Clarke, F. H. *Optimization and Nonsmooth Analysis. SIAM.*(1990).

[3] Aragón Artacho, F. J., Fleming, R. M., and Vuong, P. T. *Accelerating the DC algorithm for smooth functions. Mathematical Programming*. Vol 169, 95–118, (2018).

[4] Aragón Artacho, F. J., and Vuong, P. T. *The boosted difference of convex functions algorithm for nonsmooth functions. SIAM Journal on Optimization*. Vol 30, 980–1006, (2020).

[5] Ferreira, O. P., Santos, E. M., and Souza, J. C. O. *A boosted DC algorithm for non-differentiable DC components with non-monotone line search. Computational Optimization and Applications*. Vol 88, 783–818, (2024).

[6] Hartman, P. *On functions representable as a difference of convex functions. Pacific Journal of Mathematics*.Vol 9, 707–713, (1959).

[7] Ferreira, O. P., Rabelo, R. A. L., Ribeiro, P. H. A., Santos, E. M., and Souza, J. C. O. *Image denoising with a non-monotone boosted DCA for non-convex models. Computers and Electrical Engineering*. Vol 117, 109306, (2024).

[8] Izmailov, A., and Solodov, M. *Otimização, Volume 1: Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. IMPA, (2005).

[9] An, N. T., Nam, N. M., and Yen, N. D. *Algorithm via convex analysis approach for solving a location problem involving sets. Journal of Convex Analysis*. Vol 23, 77–101, (2016).

[10] Tao, P. D., and An, L. H. *Convex analysis approach to DC programming: theory, algorithms and applications. Acta Mathematica Vietnamica.* Vol 22, 289–355, (1997).

[11] Tao, P. D., and Souad, E. B. *Algorithms for solving a class of nonconvex optimization problems: methods of subgradients. North-Holland Mathematics Studies.* Vol 129, 249–271, (1986).

[12] Cruz Neto, J. X., Lopes, J. O., Santos, P. S. M., and Souza, J. C. O. *An interior proximal linearized method for DC programming based on Bregman distance or second-order homogeneous kernels. Optimization.* Vol 68, 1305–1319, (2019).

[13] Cruz Neto, J. X., Oliveira, P. R., Soubeyran, A., and Souza, J. C. O. *A generalized proximal linearized algorithm for DC functions with application to the optimal size of the firm problem. Annals of Operations Research.* Vol 289, 313–339, (2020).

[14] Khamaru, K., and Wainwright, M. J. *Convergence guarantees for a class of non-convex and non-smooth optimization problems. Journal of Machine Learning Research.* Vol 20, 1–52, (2019).

[15] Clarke, F. H. *Optimization and Nonsmooth Analysis. Canadian Mathematical Society Series of Monographs and Advanced Texts.* Wiley, (1983).

[16] Joki, K., Bagirov, A. M., Karmitsa, N., Mäkelä, M. M., and Taheri, S. *Double bundle method for finding Clarke stationary points in nonsmooth DC programming. SIAM Journal on Optimization.* Vol 2, 1892–1919, (2018).

[17] Tuy, H. *Convex Analysis and Global Optimization. Springer.* Vol 110, (2016).

[18] Attouch, H., Bolte, J., Redont, P., and Soubeyran, A. *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality. Mathematics of Operations Research.* Vol 35, 438–457, (2010).

[19] Lou, Y., Zeng, T., Osher, S., and Xin, J. *A weighted difference of anisotropic and isotropic total variation model for image processing. SIAM Journal on Imaging Sciences.* Vol 8, 179–1823, (2015).

[20] Attouch, H., Bolte, J., and Svaiter, B. F. *Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods. Mathematical Programming.* Vol 137, 91–129, (2013).

[21] Brimberg, J. *The Fermat–Weber location problem revisited. Mathematical Programming.* Vol 71, 71–76, (1995).

[22] Mordukhovich, B. S., and Nam, N. M. *An Easy Path to Convex Analysis and Applications. Springer*, (2014).

[23] Li, X., Hu, Y., Gao, X., Tao, D., and Ning, B. A multi-frame image super-resolution method. *Signal Processing.* Vol 90, 405–414, (2010).

[24] Cuong, T. H., Yao, J.-C., and Yen, N. D. *Qualitative properties of the minimum sum-of-squares clustering problem. Optimization.* Vol 69, 2131–2154, (2020).

[25] Souza, J. C. O., Oliveira, P. R., and Soubeyran, A. *Global convergence of a proximal linearized algorithm for difference of convex functions. Optimization Letters.* Vol 10, 1529–1539, (2016).

[26] Beck, A., and Hallak, N. On the convergence to stationary points of deterministic and randomized feasible descent directions methods. *SIAM Journal on Optimization*, Vol 30, 56-79,(2020).

[27] Rockafellar, R. T., and Wets, R. J.-B. *Variational Analysis*, Vol. 317. Springer, (2009).

[28] Sun, W.-Y., Sampaio, R. J., and Candido, M. Proximal point algorithm for minimization of DC function. *Journal of Computational Mathematics* 451–462, (2003).

[29] Moudafi, A., and Maingé, P.-E. On the convergence of an approximate proximal method for DC functions. *Journal of Computational Mathematics* 475–480, (2016).

[30] Lanza, A., Morigi, S., and Sgallari, F. Convex image denoising via non-convex regularization with parameter selection. *Journal of Mathematical Imaging and Vision.* Vol 56, 195–220, (2016).

[31] Yin, P., Lou, Y., He, Q., and Xin, J. Minimization of ℓ_{1-2} for compressed sensing. *SIAM Journal on Scientific Computing*. Vol 37, A536–A563, (2015).

[32] Geremew, W., Nam, N. M., Semenov, A., Boginski, V., and Pasiliao, E. *A DC programming approach for solving multicast network design problems via the Nesterov smoothing technique. Journal of Global Optimization*. Vol 72, 705–729, (2018).

[33] Selesnick, I. W., and Bayram, I. *Sparse signal estimation by maximally sparse convex optimization. IEEE Transactions on Signal Processing*. Vol 62, 1078–1092, (2014).

[34] Ordin, B., and Bagirov, A. M. *A heuristic algorithm for solving the minimum sum-of-squares clustering problems. Journal of Global Optimization*. Vol 61, 341–361, (2015).

[35] Nikolova, M., Ng, M. K., and Tam, C. P. *Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. IEEE Transactions on Image Processing*. Vol 19, 3073–3088, (2010).

[36] De Oliveira, W., and Tcheou, M. P. *An inertial algorithm for DC programming. Set-Valued and Variational Analysis*. Vol 27, 895–919, (2019).

[37] Daubechies, I., Defrise, M., and De Mol, C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics*. Vol 57, 1413–1457, (2004).

[38] Santos, W.M.S. *An Inexact Non-monotone Boosted Difference of Convex Algorithm*. Dissertação de Mestrado, Universidade Federal do Piauí, Programa de Pós-Graduação em Matemática, (2024).

[39] Beck, A. *First-Order Methods in Optimization. SIAM*, (2017). `https://epubs.siam.org/doi/10.1137/1.9781611974997`

[40] Beck, A., and Teboulle, M. *A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences*. Vol 2, 183–202, (2009).

[41] Beck, A., and Teboulle, M. *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Transactions on Image Processing*. Vol 18, 2419–2434, (2009).

[42] Nesterov, Y. E. *A method for solving the convex programming problem with convergence rate $\mathcal{O}(\frac{1}{k^2})$. In: Dokl. Akad. Nauk SSSR.* 543–547, (1983). https://www.mathnet.ru/eng/dan/v269/i3/p543

[43] Bagirov, A.M. and Ugon, J. *Codifferential method for minimizing nonsmooth DC functions. Journal of Global Optimization.* Vol 50, 3-22, (2021).

[44] de Oliveira, W. and Tcheou, M.P. *An inertial algorithm for DC programming. Set-Valued and Variational Analysis.* Vol 27 , 895–919, (2019).

[45] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics: A Foundation for Computer Science.* 2nd ed., Addison-Wesley, Reading, MA,(1994).

[46] Joki, K., Bagirov, A.M., Karmitsa, N., Mäkelä, M.M., and Taheri, S. *Double bundle method for finding Clarke stationary points in nonsmooth DC programming. SIAM Journal on Optimization.* Vol 28, 1892–1919, (2018).