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

An Inexact Non-monotone Boosted Difference of Convex Algorithm

Wilkreffy Manoel de Sousa Santos

Teresina - 2024

Wilkreffy Manoel de Sousa Santos

Dissertação de Mestrado:

An Inexact Non-monotone Boosted Difference of Convex Algorithm

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 - 2024

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

An Inexact Non-monotone Boosted Difference of Convex Algorithm

Wilkreffy Manoel de Sousa Santos

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 29 de julho de 2024.

Banca Examinadora:

João Carlos de Oliveira Sz

Prof. Dr. João Carlos de Oliveira Souza – Orientador

JOAO XAVIER DA CRUZ NETO:21810141320
Assinado de forma digital por JOAO XAVIER DA CRUZ NETO:21810141320
Dados: 2024.08.08 14:57:14 -03'00'

Prof. Dr. João Xavier da Cruz Neto - UFPI

Prof. Dr. Gilson do Nascimento Silva - UFPI

Geovani Ngrapiglia

Prof. Dr. Geovani Nunes Grapiglia - UCLouvain

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Divisão de Representação da Informação

S237a Santos, Wilkreffy Manoel de Sousa.
An Inexact Non-monotone Boosted Difference of Convex Algorithm / Wilkreffy Manoel de Sousa Santos. -- 2024.
75 f.

Dissertação (Mestrado) – Universidade Federal do Piauí, Programa de Pós-Graduação em Matemática, Teresina-PI, 2024.
"Orientador: Prof. Dr. João Carlos de Oliveira Souza"

1. Funções DC. 2. Algoritmos DC. 3. nmBDCA inexato. 4. Ilustrações numéricas. I. Souza, João Carlos de Oliveira. II. Título.

CDD 510

Bibliotecária: Francisca das Chagas Dias Leite – CRB3/1004

Dedico este trabalho à minha mãe, Francisca Neide.

Agradecimentos

Agradeço imensamente à minha família por todo apoio e incentivo que me foram dados. Em especial, agradeço à minha mãe por todo seu amor, carinho, incentivos, compreensão, coragem e esforço para superar as dificuldades para que eu pudesse ter a oportunidade de me dedicar aos estudos e trilhar a minha trajetória acadêmica. É com muita felicidade que posso dizer que esta conquista é nossa. Amo você, mãe.

Agradeço a todos os meus amigos que, com todo companheirismo, me incentivaram e apoiaram ao longo desta trajetória. Em particular, deixo expressa toda a minha gratidão a José Vinícius (meu primo e grande amigo), Maria Clara, Grabriele, Nonato Jr., Fabríssio, José Rufino, Luís Estevão, Lucas, João Matheus, Arimatéa Melo Jr, Arimatéa Aguiar Jr, Carlos Eduardo (Cadu) e Igor Kallyo por terem me acompanhado tão de perto. A presença de vocês em minha vida me engrandece enquanto ser humano.

Agradeço ao Professor Dr. João Carlos de Oliveira Souza, meu orientador desde que ingressei na graduação, por todo apoio, amizade e incentivo. Por ser esse profissional dedicado, talentoso, compreensível e sempre disposto a auxiliar e enriquecer o trabalho dos seus orientandos. Muito obrigado por ter aceitado me orientar em mais uma etapa, por acreditar em mim e no meu trabalho. Muito obrigado por me encorajar a seguir os meus sonhos e não desistir diante dos desafios que posso encontrar na ciência, por me mostrar, com seu próprio exemplo, o impacto positivo que um professor pode imprimir na vida de um aluno. Além do saber matemático, com o senhor eu aprendo também a me tornar um ser humano melhor, prestativo, solidário e acolhedor.

Agradeço ao Professor Me. Antônio Oliveira, do IFPI, pelo companheirismo e amizade, por sempre me incentivar a estudar matemática e me encorajar a seguir os meus sonhos. Professor, você tem um papel importante na minha trajetória.

Agradeço cordialmente aos professores Doutores Geovani Nunes Gripaglia, Gilson do

Nascimento Silva e João Xavier da Cruz Neto por aceitarem o convite para participar da Banca Examinadora da defesa de Dissertação e pelas valiosas contribuições. Muito obrigado pela atenção e pelo cuidado nas correções que enriquecem a versão final deste trabalho.

Agradeço a todos os professores do Departamento de Matemática da UFPI, comprometidos com o ensino desta ciência e dedicados a contribuir na trajetória dos seus discentes. Nesse percurso, tive a honra de ser aluno de profissionais incríveis, dentre os quais os professores: Halyson Baltazar, Jurandir Lopes, Ítalo Dowell, Barnabé Lima, Mário Gomes, Antonio Wilson, Paulo Alenxandre e Cícero Aquino.

Agradeço a todos os funcionários técnicos e terceirizados que atuam na manutenção e no bom funcionamento do Departamento de Matemática com empenho e dedicação. Em particular, agradeço a Larisse, secretária da Pós-Graduação em Matemática, por todo o seu empenho e prontidão para ajudar a todos os discentes e docentes.

Agradeço às Instituições de Ensino Superior públicas que possibilitam o acesso a uma educação de qualidade a todos; em particular, agradeço à Universidade Federal do Piauí pelas suas políticas de incentivo e permanência dos estudantes durante sua formação. Agradeço de modo geral a todos professores que em algum momento fizeram parte da minha vida acadêmica como um todo.

Agradeço à CAPES pela concessão do apoio financeiro que possibilitou dedicação integral às atividades do Mestrado.

"*If I have seen further, it is by standing upon the shoulders of giants*".

Isaac Newton.

Resumo

Estudamos o problema de minimização em uma classe de funções possivelmente não convexas e não diferenciáveis, dadas pela diferença de duas funções convexas. Abordamos esse problema por meio de três métodos estabelecidos na literatura: o Difference of Convex Algorithm (DCA), proposto por Tao e Souad [46]; o Boosted Difference of Convex Algorithm (BDCA), formulado por Aragón Artacho e Vuong [4], que considera uma busca monótona em cada iterada a partir da solução encontrada pelo DCA; e o mais recente Non-monotone Boosted Difference of Convex Algorithm, proposto por Fereira, Santos e Souza [21], que considera uma busca não monótona no BDCA, habilitando um possível crescimento na função objetivo controlado por um parâmetro. Além disso, propomos uma abordagem inexata para o nmBDCA e, sob hipóteses razoáveis, recuperamos os resultados de convergência e complexidade da sua versão exata. Realizamos alguns experimentos numéricos para ilustrar os algoritmos.

Palavras-chave: Funções DC; Algoritmos DC; nmBDCA inexato; Ilustrações numéricas.

Abstract

We studied the minimization problem in a class of functions that are possibly non-convex and non-differentiable, given by the difference of two convex functions. We approached this problem through three methods established in the literature: the Difference of Convex Algorithm (DCA), proposed by Tao and Souad [46]; the Boosted Difference of Convex Algorithm (BDCA), formulated by Aragón Artacho and Vuong [4], which considers a monotone line search at each iteration starting from the solution found by DCA; and the latest Non-monotone Boosted Difference of Convex Algorithm, proposed by Ferreira, Santos, and Souza [21], which incorporates a non-monotone line search in BDCA, enabling potential growth in the objective function controlled by a parameter. Additionally, we present an inexact approach for nmBDCA and, under reasonable assumptions, recover the convergence and complexity results of its exact version. We conducted numerical experiments to illustrate the algorithms.

Key-words: DC Function; DC Algorithms; Inexact nmBDCA; Numerical illustrations.

Contents

List of Figures

Introduction

In this work, we consider the problem of minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$ that is possibly non-convex and non-differentiable, which can be expressed as the difference of two convex functions $g, h : \mathbb{R}^n \to \mathbb{R}$. Functions with this representation are called DC (difference of convex) functions, and the DC problem that we approached consists of:

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

Note that when we set h to be identically zero, we retrieve the classical convex minimization problem. This already introduces (1) as an even more general problem.

The DC minimization problem has been studied and developed over the past decades and applied to many relevant problems, such as image processing [34], compressed sensing [49], location problems [2, 11, 13], sparse optimization problems [24], the minimum sum-of-squares clustering problem [4, 17, 40], the bilevel hierarchical clustering problem [38], clusterwise linear regression [8], the multicast network design problem [23], and multidimensional scaling problem [1, 4, 10]. Recently, there have also been works directed towards support vector machines [22], Value-at-Risk Constrained Portfolio Optimization [47], and training deep neural networks [16]. More applications can be found in [48, Part II].

Many methods can be found in the literature to solve the DC problem (1), among which we mention subgradient-type methods [10, 31], proximal subgradient methods [15, 36, 43, 44], proximal bundle methods [18], codifferential methods [7], and inertial methods [20]. However, the Difference of Convex Algorithm (DCA) proposed by Tao and Souad [46] was the first to consider the problem (1) with the particularity of its DC structure. The interpretation of DCA is simple: at each iteration k, the second DC component is replaced by a linear minorant, which reduces the problem to minimizing a convex problem in each iteration. Since then, many variants of DCA have emerged, and their theoretical and practical properties have been studied over the years.

In addition to DCA, we have studied and presented the Boosted Difference of Convex Algorithm (BDCA) proposed by Aragón Artacho and Vuong [4] (see also [3]), whose key idea is to define, at each iteration k, a descent direction for f based on the solution computed by DCA. This subtle modification further decreases the objective function value compared to DCA, making the method in some sense accelerated. The improvement in performance compared to DCA will be demonstrated in some examples distributed throughout the corresponding section. However, achieving this descent property comes at the cost of assuming differentiability of the first DC component. As we will see, removing this hypothesis can result in an ascent direction, making it impossible to perform a search that decreases the objective function value.

To overcome this drawback (differentiability for the first DC component) Ferreira, Santos, and Souza [21] propose a method that allows controlled growth of f using a parameter. This approach extends the applicability of their method to cases where both components may be non-differentiable. In practice, the work by Ferreira, Santos, and Souza [21] has already shown this method to be highly effective in solving several well-established examples in the literature.

Finally, we acknowledge that when implementing an algorithm computationally, we are susceptible to the fact that subproblem solutions are calculated approximately. However, convergence properties of both BDCA and nmBDCA are demonstrated based on the exact solution of the subproblems. Therefore, motivated to ensure convergence of these methods when subproblem solutions are computed inexactly, we propose a new algorithm called Inexact Non-monotone Boosted Difference of Convex which considers inexact computation of the subproblems. Under reasonable assumptions, this algorithm restores convergence properties and complexity similar to its exact counterpart.

This work is divided as follows: in Chapter 1 we compiled useful theoretical results for the development of each of the algorithms, including their convergence analyses. In Chapter 2, we approached the DCA and emphasized its motivation and relevance, in addition to the convergence results and some illustrations of its computational performance. In Chapter 3, the BDCA is presented, establishing its definition, convergence analysis and some graphs of its computational performance. In Chapter 4 we presented the definition and properties of the nmBDCA comparing its computational performance with DCA. Finally, we present new results on an inexact version of nmBDCA.

Chapter 1

Concepts and results of optimization

In this chapter, we will introduce the main theoretical results that will support this entire work. The vast majority of the definitions and results presented here can be found in books on convex analysis; see [12, 35, 26, 28].

1.1 Convex Sets and Functions

Definition 1.1.1. *A set $C \subset \mathbb{R}^n$ is called **convex** if for any $x, y \in C$, and $\lambda \in [0,1]$ it holds that $\lambda x + (1-\lambda)y \in C$.*

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

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y). \tag{1.1}$$

*when the inequality in (1.1) is strict for any $x, y \in C$ and $\lambda \in (0,1)$, then f is called **strictly convex**.*

Definition 1.1.3. *Let $C \subset \mathbb{R}^n$ a convex set. A convex function $f : C \to \mathbb{R}$ is called **strongly convex** with modulus $\rho > 0$ if for all $x, y \in C$ and $\lambda \in [0,1]$ we have*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\rho}{2}\lambda(1-\lambda)||x-y||^2.$$

As we can observe, it is immediate that every strongly convex function with parameter $\rho > 0$ is, in particular, strictly convex and therefore convex. When dealing with convex and strongly convex functions, it is very common to resort to characterizations equivalent to the definitions presented above. We will discuss the most relevant ones for this work in Section 1.2.1, using the definition of subgradient.

Proposition 1.1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strongly convex function with modulus $\rho > 0$. Then f has a unique minimizer.*

Proof. See [9, Theorem 5.25]. □

The Fenchel conjugate of a convex function is a function that plays a fundamental role in convex duality and convex optimization problems. Below, we present its definition and some properties.

Definition 1.1.4. *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a function (not necessarily convex), its Fenchel conjugate $\phi^* : \mathbb{R}^n \to (-\infty, +\infty]$ is*

$$\phi^*(v) := \sup \{\langle v, x\rangle - \phi(x) \mid x \in \mathbb{R}^n\}, \quad v \in \mathbb{R}^n.$$

Proposition 1.1.2. *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ a convex function. Then the following statements are equivalent.*

(i) $x^* \in \partial\phi(x)$;

(ii) $\phi(x) + \phi^*(x^*) = \langle x^*, x\rangle$;

(iii) $x \in \partial\phi^*(x^*)$.

Proof. See [9, Theorem 4.19]. □

An important result concerning convex functions is their continuity within the interior of their domain. Before ensuring this fact, we define what a Lipschitz function is.

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

$$|f(x) - f(y)| \leq L\|x - y\|, \quad \forall x, y \in D.$$

We say that f is ***locally Lipschitz continuous*** *at $\bar{x} \in D$ when there exists $\delta > 0$ such that f is Lipschitz on $D \cap B(\bar{x}; \delta)$. When f is locally Lipschitz for all $x \in D$, we say simply that f is locally Lipschitz.*

Proposition 1.1.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, then f is locally Lipschitz. In particular, f is continuous on C.*

Proof. See [9, Theorem 2.21]. □

Convex functions are not necessarily differentiable. However, we will show that a convex function has directional derivatives. In Section 1.2, we will present Fenchel subdifferential and Clarke subdifferential, which are a kind of generalization of the derivative.

Theorem 1.1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, for all $x \in \mathbb{R}^n$, f is differentiable in each direction $d \in \mathbb{R}^n$. Moreover,*

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

where $f'(x; d) := \lim_{\alpha \downarrow 0} \dfrac{f(x + \alpha d) - f(x)}{\alpha}$ is the directional derivative of f at the point x in the direction d.

In Sections 1.2.1 and 1.2.3, we obtain relations between directional derivative and the Fenchel and Clarke subdifferential of a function.

1.2 Subdifferential calculus

In this section, we present some well-established subdifferentials in the literature that will be used throughout this work. The subdifferentials defined here can be interpreted as set-valued operators. We say that $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a set-valued operator when for each $x \in \mathbb{R}^n$ we associate a subset $F(x) \subset \mathbb{R}^n$.

1.2.1 Fenchel subdifferential

The Fenchel subdifferential is classical in the study of convex functions and arises as a natural generalization of the derivative concept in the case where the function is convex. However, its definition does not require the function to be convex. Through it, we can establish important properties of convex and strongly convex functions, as well as an optimality condition that will be presented in this section.

Definition 1.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ a function and $x \in \mathbb{R}^n$. An element $v \in \mathbb{R}^n$ is called a **subgradient** of f at x if*

$$f(y) \geq f(x) + \langle v, y - x \rangle$$

*for all $y \in \mathbb{R}^n$. The set of all subgradients of f at x is called the **subdifferential** of the function at this point and is denoted by $\partial f(x)$.*

Note that the subgradient of f at x defines a linear approximation of f whose graph lies below that of f and whose value coincides with f at the point x. Next, we present a theorem that establishes a connection between the Fenchel subdifferential and the directional derivative of the function.

Theorem 1.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then for every $x \in \mathbb{R}^n$, the set $\partial f(x)$ is convex, compact, and non-empty. Moreover, for any $d \in \mathbb{R}^n$, we have*

$$f'(x; d) = \max_{y \in \partial f(x)} \langle y, d \rangle. \tag{1.2}$$

Proof. See [35, Proposition 2.47]. □

In the discussion of the well-definition of the algorithms studied in this work, we will see the importance of the previous theorem in ensuring that it is possible to select an element from a subdifferential.

Proposition 1.2.1. *If a convex function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at the point $x \in \mathbb{R}^n$, then the set $\partial f(x)$ contains a single element. In this case, $\partial f(x) = \{\nabla f(x)\}$.*

Proof. See [9, Theorem 3.33]. □

Theorem 1.2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, $x^* \in \mathbb{R}^n$ is a minimizer of f if, and only if*

$$0 \in \partial f(x).$$

Proof. See [9, Theorem 3.63]. □

In other words, we see that the concept of the Fenchel subdifferential is a natural generalization of the derivative for convex functions. Note that when f is differentiable, if x^* is a local minimum, then $\nabla f(x^*) = 0$, which means $0 \in \partial f(x^*)$.

Proposition 1.2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Let $\{x^k\}_{k \in \mathbb{N}}$ be a sequence such that $x^k \to x$, as $k \to +\infty$, and $w^k \in \partial f(x^k)$, for all $k \in \mathbb{N}$, then the sequence $\{w^k\}_{k \in \mathbb{N}}$ is bounded and all cluster points of $\{w^k\}_{k \in \mathbb{N}}$ belong to $\partial f(x)$.*

Proof. See [27, Proposition 6.2.1]. □

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

Proof. See [41, 9.20 Corollary].

□

Below, we present some characterizations of convex and strongly convex functions associated with the Fenchel subdifferential.

Theorem 1.2.3. *The following statements are equivalent:*

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

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

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

$$\langle u - v, x - y \rangle \geq \rho||x - y||^2,$$

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

Proof. See [9, Theorem 5.24]. □

1.2.2 The ε-subdifferential

Definition 1.2.2. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a convex function and* $\varepsilon \geq 0$*. We say that* $y \in \mathbb{R}^n$ *is an* ε*-subgradient of* f *at point* $x \in \mathbb{R}^n$ *if*

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

The set of all ε*-subgradients of* f *at* x*, denoted by* $\partial_\varepsilon f(x)$*, is called the* ε*-subdifferential of* f *at* x*.*

An ε-subgradient of f at x defines a linear function whose value at x is $f(x) - \varepsilon$ and whose graph lies below that of f. As we can see,

$$\partial_{\varepsilon_1} f(x) \subset \partial_{\varepsilon_2} f(x), \quad \forall \varepsilon_2 > \varepsilon_1 \geq 0.$$

In particular,

$$\partial f(x) = \partial_0 f(x) \subset \partial_\varepsilon f(x), \quad \forall x \in \mathbb{R}^n, \quad \forall \varepsilon > 0.$$

Hence, $\varepsilon-$subdifferential is always non-empty. Moreover, it follows immediately from definition than

$$0 \in \partial_\varepsilon f(x) \iff f(x) \leq \inf_{z \in \mathbb{R}^n} f(z) + \varepsilon,$$

which represents an approximate optimality condition. In this case, if $0 \in \partial_\varepsilon f(x^*)$, then we say that x^* is a ε**-critical point**.

Theorem 1.2.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then, for all $x \in \mathbb{R}^n$ and $\varepsilon \geq 0$, the set $\partial_\varepsilon f(x)$ is and non-empty, convex and compact.*

Proof. See [26, Theorem 1.14]. □

The following proposition will be used in the convergence analysis of InmBDCA, ensuring that the accumulation points of the sequence generated by the algorithm, if any, are critical in the DC sense.

Proposition 1.2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, $\{x^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^n$ converging to x^*, $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+$ converging to ε^*, and $\{w^k\}_{k\in\mathbb{N}}$ converging to w^*, with $w^k \in \partial_{\varepsilon_k} f(x^k)$ for all $k \in \mathbb{N}$. Then $w^* \in \partial_\varepsilon f(x^*)$.*

Proof. See [26, Proposition 4.1.1]. □

1.2.3 Clarke subdifferential

An important generalization of the concept of subdifferential is the definition of the Clarke subdifferential. We know, by Proposition 1.1.3, that every convex function is locally Lipschitz continuous; however, the converse is not true. Thus, there arises the need for a theory that extends the scope of subdifferential calculus and recovers its main properties when the function under consideration is, in particular, convex.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at $x \in \mathbb{R}^n$, and $d \in \mathbb{R}^n$ be a direction. The *Clarke directional derivative* of f at x in the direction of d, denoted by $f^\circ(x; d)$, is defined as follows:

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

With this, we have the following definition.

Definition 1.2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at $x \in \mathbb{R}^n$, and $d \in \mathbb{R}^n$ be a direction. The **Clarke subdifferential** of f at x, denoted by $\partial_C f(x)$, is defined by*

$$\partial_C f(x) := \{w \in \mathbb{R}^n \mid \langle w, d\rangle \leq f^\circ(x; d), \quad \forall d \in \mathbb{R}^n\},$$

or, equivalently,

$$\partial_C f(x) = conv \left\{ \lim_{i\to+\infty} \nabla f(x^i) \mid x^i \to x \text{ and } \nabla f(x^i) \text{ exists} \right\}.$$

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 hold:*

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

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

Proof. See [12, Proposition 2.1.2]. □

The following proposition demonstrates that the Clarke subdifferential further extends the concept of derivative, being extended to locally Lipschitz continuous functions. In this sense, we will see that when the function is convex, the Clarke subdifferential recovers the Fenchel subdifferential.

Proposition 1.2.5. *When $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then $\partial_C f(x)$ coincides with the Fenchel subdifferential at x in the sense of convex analysis, and $f^\circ(x; v)$ coincides with the directional derivative $f'(x; v)$ for each $v \in \mathbb{R}^n$.*

Proof. See [12, 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*

$$\partial_C(sf)(x) = s\partial_C f(x).$$

Proof. See [12, Proposition 2.3.1]. □

Theorem 1.2.5. *Let $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, N$, be locally Lipschitz continuous functions. Then, for any scalars s_i, one has*

$$\partial_C \left(\sum_{i=1}^{N} s_i f_i \right)(x) \subset \sum_{i=1}^{N} s_i \partial_C f_i(x),$$

and equality holds if all but at most on of f_i are continuously differentiable ($f_i \in C^1(\mathbb{R}^n)$) at x.

Proof. See [12, Corollary 2, p. 39]. □

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

$$f(x + td) < f(x).$$

Proposition 1.2.7. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function at $x \in \mathbb{R}^n$. The direction $d \in \mathbb{R}^n$ is a descent direction for f at x if*

$$\langle s, d\rangle < 0 \quad \textit{for all} \quad s \in \partial_C f(x) \quad \textit{or} \quad f^\circ(x; d) < 0.$$

Proof. See [6, Theorem 4.5]. □

In the following section, we will see that functions that can be expressed as differences of convex functions are locally Lipschitz continuous. As we have seen before, the Clarke subdifferential can be understood as an extension of the derivative concept. In this regard, the following proposition provides a sufficient condition for a point to be a local minimum or maximum of a locally Lipschitz continuous function.

Proposition 1.2.8. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function. If f attains a local minimum or maximum at x^*, then $0 \in \partial_C f(x^*)$.*

Proof. See [12, Proposition 2.3.2]. □

With this, we say that a point $x \in \mathbb{R}^n$ is called **Clarke stationary** for f when $0 \in \partial_C f(x)$.

1.3 Difference of Convex Functions

In this section, we will present a class of functions that can be written as the difference between two convex functions, along with definitions and basic results. These functions, called DC functions, will be considered in the context of minimization problems under certain assumptions.

We will address DC functions defined on $\mathbb{R}^n$. A function $f : \mathbb{R}^n \to \mathbb{R}$ will be called a **difference of convex function**, or simply a DC function, if there exists a pair of convex functions $g, h : \mathbb{R}^n \to \mathbb{R}$, respectively the first and the second components, such that $f(x)$ is the difference

$$f(x) = g(x) - h(x), \quad \forall x \in \mathbb{R}^n.$$

A function is locally DC if each point of its domain has a convex neighbourhood wherein it is DC. The set of DC functions with the usual operations of function addition and scalar multiplication defines a vector space, denoted by $DC(\mathbb{R}^n)$. Moreover, $DC(\mathbb{R}^n)$ is the smallest vector space containing all convex functions (in this case, also continuous) defined on a given set; see [5].

It is immediate that any convex function can be written as a difference of convex functions; one simply considers the second component to be identically zero. The following results provide a wide range of examples of DC functions.

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

Proof. See [48, Proposition 4.2]. □

Theorem 1.3.1. *(Hartman) Every function locally DC on $\mathbb{R}^n$ is globally DC on $\mathbb{R}^n$.*

Proof. See [25, (I)] or [48, Proposition 4.3]. □

In Proposition 1.3.1, considering $\delta > 0$, $x_0 \in \mathbb{R}^n$, and $C = B[x_0; \delta]$, we have that f is DC on $B(x_0; \delta)$. Therefore, by Theorem 1.3.1, f is globally DC, from which we conclude that every function $f \in C^2(\mathbb{R}^n)$ is DC.

Finding a decomposition for a DC function may not be an easy task. However, once one decomposition is found, we can obtain infinitely many others from it. Hence a DC function does note have just one DC decomposition.

Let $f(x) = g_0(x) - h_0(x)$ be a DC function on $\mathbb{R}^n$, with $g, h : \mathbb{R}^n \to \mathbb{R}$ being its DC components. If $\Phi : \mathbb{R}^n \to \mathbb{R}$ is convex, we obtain a new decomposition:

$$f(x) = g(x) - h(x),$$

where $g(x) = g_0(x) + \phi(x)$ and $h(x) = h_0(x) + \phi(x)$. We can further assume $\phi(x)$ to be strongly convex with modulus $\rho > 0$; in this case, $g(x) = g_0(x) + \phi(x)$ and $h(x) = h_0(x) + \phi(x)$ become strongly convex with modulus $\rho > 0$, by Proposition 1.2.3. The parameter $\rho > 0$ can be easily chosen, for example, if we define $\phi(x) = \frac{\rho}{2}\|x\|^2$, which is strongly convex with modulus $\rho > 0$.

Choosing a DC decomposition whose components are strongly convex can facilitate the verification of certain theoretical results. However, we will also analyze the performance of the method when modifying the parameter $\rho > 0$.

We denote the DC minimization problem as follows:

$$\min_{x \in \mathbb{R}^n} f(x) = g(x) - h(x), \tag{1.3}$$

where $f \in DC(\mathbb{R}^n)$, and $g, h : \mathbb{R}^n \to \mathbb{R}$ are its DC components. In this context, it is necessary to introduce the concept of critical point in the DC sense. However, beforehand, we note that DC functions inherit the property of being locally Lipschitz from their convex components.

Proposition 1.3.2. *Every function $f \in DC(\mathbb{R}^n)$ is locally Lipschitz.*

Proof. Consider $f(x) := g(x) - h(x)$, where $g, h : \mathbb{R}^n \to \mathbb{R}$ are its DC components. By Proposition 1.1.3, $g(x)$ and $h(x)$ are locally Lipschitz. Therefore, for any $x_0 \in \mathbb{R}^n$, there exist $\delta_1, \delta_2, L_1, L_2 > 0$ such that

$$|g(x) - g(y)| \leq L_1 \|x - y\| \qquad \forall x, y \in B(x_0, \delta_1)$$

and

$$|h(x) - h(y)| \leq L_2 \|x - y\|, \quad \forall x, y \in B(x_0, \delta_2).$$

Thus,

$$\begin{aligned} |f(x) - f(y)| &= |g(x) - h(x) - g(y) + h(y)| \\ &\leq |g(x) - g(y)| + |h(y) - h(x)| \\ &\leq (L_1 + L_2)\|x - y\|, \end{aligned}$$

for all $x, y \in B(x_0, \delta)$, where $\delta = \min\{\delta_1, \delta_2\}$. Therefore, f is locally Lipschitz. □

Thus, we can consider the Clarke subdifferential of f applied at $x \in \mathbb{R}^n$ as defined in Definition X. Let $f(x) = g(x) - h(x)$ be a DC function on $\mathbb{R}^n$. From Theorem 1.2.5, we have that $\partial_C f(x) \subset \partial_C g(x) - \partial_C h(x)$. By Proposition 1.2.5, the Fenchel subdifferential coincides with the Clarke subdifferential when the function is convex, then we obtain

$$\partial_C f(x) \subset \partial g(x) - \partial h(x).$$

The following proposition presents a necessary condition for a point x^* to be a local minimum (1.3).

Proposition 1.3.3. *Let $g, h : \mathbb{R}^n \to \mathbb{R}$ convex functions. If x^* is a local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = g(x) - h(x)$, for all $x \in \mathbb{R}^n$, then*

$$\partial h(x^*) \subset \partial g(x^*). \tag{1.4}$$

*Points satisfying (1.4) are called **inf-stationary**.*

Proof. See [45, Theorem 2]. □

According to [30], the inf-stationary condition (1.4) is not easy to be verified in practice due to difficulty of calculating the subdifferentials of the DC components g and h. Thus, in numerical algorithms, a relaxed form of condition (1.4) is often used, which requires that

$$\partial g(x^*) \cap \partial h(x^*) \neq \varnothing. \tag{1.5}$$

A point satisfying (1.5) is called a **critical point**. Due to previous proposition, condition (1.5) is also a necessary condition for a local optmality.

There are some interesting relationships between inf-stationary, Clarke-stationary, and critical points mentioned in [30]. These relationships are summarized in the following figure extracted from [29]:

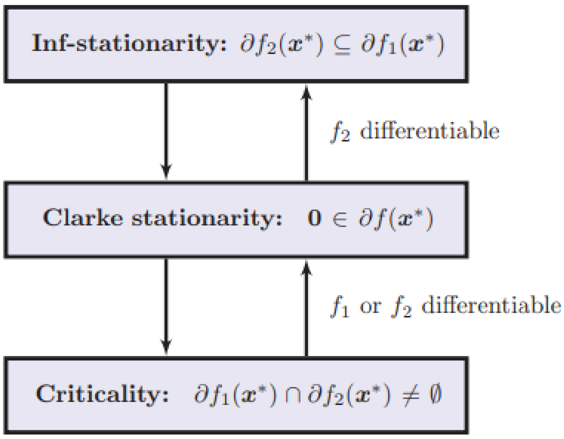


Figure 1.1: Relationship between different stationary concepts [30].

In Figure 1.1, the DC function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(x) = f_1(x) - f_2(x)$ was considered. Note that every inf-stationary point is Clarke stationary. Furthermore, every Clarke stationary point is critical. However, the converse is not generally true. For criticality to imply Clarke stationary, we need to assume that f_1 or f_2 is differentiable. Finally, Clarke stationary implies inf-stationary when the second DC component f_2 is differentiable.

Given the formulation of the DC programming problem and its wide range of applications, there arises the need to develop methods to solve this type of problem, such as subgradient-type [10, 31], proximal subgradient [15, 36, 43, 44], proximal bundle [18], codifferential [7], and inertial methods [20]. However, we will focus on the following methods:

the Difference of Convex Algorithm (DCA), proposed in [46]; the Boosted Difference of Convex Algorithm (BDCA), proposed in [4]; and we propose an inexact version of the Non-monotone Boosted Difference of Convex Algorithm (nmBDCA), proposed in [21].

For further insights into the theory of DC functions, see the following surveys: [5], [19], and [32].

Chapter 2

The Difference of Convex Algorithm

The vector space of DC functions, denoted by $DC(\mathbb{R}^n)$, contains the set of real convex functions defined on $\mathbb{R}^n$ and thus broadens the scope of techniques for solving minimization problems.

We recall the problem DC in (1):

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

where $g, h : \mathbb{R}^n \to \mathbb{R}$ are convex functions.

Throughout this chapter, we make the following assumptions:

(H1) $g, h : \mathbb{R}^n \to \mathbb{R}$ are both strongly convex with modulus $\rho > 0$;

(H2) $f^* := \inf_{x \in \mathbb{R}^n} \{f(x) = g(x) - h(x)\} > -\infty$.

Observe that assumption **(H1)** is not theoretically restrictive. As mentioned in Section 1.3, given two convex functions, g and h, we can add a strongly convex term $\frac{\rho}{2}||x||^2$ to both to achieve a new decomposition, with components strongly convex with modulus $\rho > 0$. The hypothesis **(H2)** is common in the context of DC programming; see [3, 4, 15, 21].

2.1 The algorithm

The Difference of Convex Algorithm (DCA), proposed by Tao and Souad [46], was the first to address problem (2) by exploring its DC structure. Originally, the DCA was proposed as in **Algorithm 1**, but with a subtle difference: instead of (2.1), the subproblem was to find $y^k \in \partial g^*(w^k)$, where g^* is the Fenchel conjugate of the function

g. Proposition 1.1.2 implies that to calculate $y^k \in \partial g^*(w^k)$ is equivalent to $w^k \in \partial g(y^k)$, that it is, $0 \in \partial g(y^k) - w^k$, i.e., $y^k = \arg\min_{x\in\mathbb{R}^n} g(x) - \langle w^k, x - x^k\rangle$, which proves that both subproblems are equivalent.

Next, we present the Difference of Convex Algorithm (DCA) to solve (2).

Algorithm 1 Difference of Convex Algorithm (DCA)[46]

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

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

$$\min_{x\in\mathbb{R}^n} g(x) - \langle w^k, x - x^k\rangle. \tag{2.1}$$

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

Remark 2.1.1. *By Theorem 1.2.1, $\partial h(x^k) \neq \varnothing$ for all $k \in \mathbb{N}$. Furthermore, by item (ii) of Theorem 1.2.3, the function at* (2.1) *is strongly convex with modulus $\rho > 0$. As a consequence of Proposition 1.1.1 the subproblem* (2.1) *has a solution. Thus, the **Algorithm 1** is well-defined.*

Lemma 2.1.1. *If $y^k = x^k$ for some $k \in \mathbb{N}$, then x^k is a critical point of f.*

Proof. Indeed, the expression at (2.1) is equivalent to

$$w^k \in \partial g(y^k).$$

Indeed, suponha que y^k seja solução de (2.1). Then, since the subproblem function is, in particular, convex, we have that

$$0 \in \partial\big(g(\cdot) - \langle w^k, \cdot\rangle + \langle w^k, x^k\rangle\big)(y^k).$$

From Proposition 1.2.1 and Theorem 1.2.2, it follows that $0 \in \partial g(y^k) - w^k$, that is, $w^k \in \partial g(y^k)$. On the other hand, $w^k \in \partial h(x^k)$, due to Step 2. Consequently, if $y^k = x^k$, it follows that $\partial g(x^k) \cap \partial h(x^k) \neq \varnothing$. Thus, x^k is a critical point. □

The subproblem in (2.1) consists of minimizing the difference between $g(x)$ and a linear approximation of $h(x)$. In fact, since $w^k \in \partial h(x^k)$, we have that

$$h(x) \geq h(x^k) + \langle w^k, x - x^k\rangle, \quad \forall x \in \mathbb{R}^n,$$

thus,

$$g(x) - h(x) \leq g(x) - \langle w^k, x - x^k \rangle - h(x^k), \quad \forall x \in \mathbb{R}^n.$$

Ignoring the constant term $h(x^k)$, we obtain exactly the expression in (2.1) on the right-hand side of the last inequality.

2.2 Convergence analysis

The aim of this section is to present convergence analysis results of the DCA.

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

*(i) The **Algorithm 1** terminates at a critical point;*

(ii) The sequence $\{f(x^k)\}_{k\in\mathbb{N}}$ *is decreasing, i.e.,* $f(x^{k+1}) < f(x^k)$*, for all* $k \in \mathbb{N}$.

Proof. Indeed, according Step 3, the **Algorithm 1** terminates when $y^k = x^k$. By Lemma 2.1.1, x^k is a critical point, and this proves the item *(i)*. Suppose that $x^{k+1} \neq x^k, \forall k \in \mathbb{N}$. Since $w^k \in \partial h(x^k) \cap \partial g(x^{k+1})$ for any $k \in \mathbb{N}$, applying the item *(iii)* of the Theorem 1.2.3 we have

$$g(x^k) - g(x^{k+1}) \geq \langle w^k, x^k - x^{k+1} \rangle + \frac{\rho}{2}||x^k - x^{k+1}||^2 \tag{2.2}$$

and

$$h(x^{k+1}) - h(x^k) \geq \langle w^k, x^{k+1} - x^k \rangle + \frac{\rho}{2}||x^{k+1} - x^k||^2 \tag{2.3}$$

By adding the inequalities (2.2) and (2.3) term by term, we have

$$\begin{aligned}
[g(x^k) - h(x^k)] - [g(x^{k+1}) - h(x^{k+1})] &\geq \langle w^k, x^k - x^{k+1} \rangle + \langle w^k, x^{k+1} - x^k \rangle + \rho||x^k - x^{k+1}||^2 \\
&= \langle w^k, x^k - x^{k+1} \rangle - \langle w^k, x^k - x^{k+1} \rangle + \rho||x^k - x^{k+1}||^2 \\
&= \rho||x^k - x^{k+1}||^2.
\end{aligned}$$

Then

$$\rho||x^k - x^{k+1}||^2 \leq f(x^k) - f(x^{k+1}), \quad \forall k \in \mathbb{N}. \tag{2.4}$$

In particular, $x^{k+1} \neq x^k$ implies $||x^k - x^{k+1}||^2 > 0$. Hence,

$$0 < f(x^k) - f(x^{k+1}), \forall k \in \mathbb{N}.$$

Thus, $f(x^{k+1}) < f(x^k)$ for any $k \in \mathbb{N}$ and it proves that $\{f(x^k)\}$ is decreasing. □

Corollary 2.2.1. *If $\{x^k\}_{k\in\mathbb{N}}$ is a sequence generated by **Algorithm 1**, then the sequence $\{f(x^k)\}_{k\in\mathbb{N}}$ is convergent.*

Proof. This follows immediately from the fact that f is lower-bounded, by assumption **(H2)**, and $\{f(x^k)\}_{k\in\mathbb{N}}$ is decreasing. □

Proposition 2.2.2. *If $\{x^k\}_{k\in\mathbb{N}}$ is a sequence generated by **Algorithm 1**, then*

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

and $||x^k - x^{k+1}|| \to 0$ as $k \to +\infty$.

Proof. Considering the partial sum at inequality (2.4):

$$0 \leq \sum_{k=0}^{N} \rho||x^k - x^{k+1}||^2 \leq \sum_{k=0}^{N} [f(x^k) - f(x^{k+1})]$$

we obtain

$$0 \leq \sum_{k=0}^{N} \rho||x^k - x^{k+1}||^2 \leq f(x^0) - f(x^{N+1})$$

By assumption **(H2)**, $f^* \leq f(x^{N+1})$, for all $N \in \mathbb{N}$, which implies that $-f(x^{N+1}) \leq -f^*$, for all $N \in \mathbb{N}$. Therefore,

$$0 \leq \sum_{k=0}^{N} \rho||x^k - x^{k+1}||^2 \leq f(x^0) - f^*, \forall k \in \mathbb{N}. \tag{2.5}$$

Taking the limit as $N \to +\infty$ at (2.5) we have

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

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

Remark 2.2.1. *Note that, by taking the limit in (2.4) as $k \to +\infty$ and using Corollary 2.2.1, we can already obtain $\lim_{k\to+\infty} ||x^k - x^{k+1}|| = 0$.*

Remark 2.2.2. *When we consider a decomposition of $f(x)$ where the components $g(x)$ and $h(x)$ are not strongly convex, then the sequence $\{||x^k - x^{k+1}||\}_{k\in\mathbb{N}}$ may not converge to 0. See the following example taken from [39].*

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 $h(x) = \sup\{-x, 0\}$. The functions g and h are piecewise linear* *and convex, but neither strongly convex nor strictly convex. Starting DCA from the initial point* $x^0 \in (0,1)$*, for example,* $x^0 := 0.1$*, we obtain* $w^0 \in \partial h(x^0) = \{0\}$*, and* $x^1 \in \arg\min\{g(x) - \langle w^0, x - x^0\rangle\} = [0,1]$ *(subproblem (2.1)). Choosing* $x^1 = 0.9$*, we then compute* $w^1 \in \partial h(x^1) = \{0\}$*. Thus,* $x^2 \in \arg\min\{g(x) - \langle w^1, x - x^1\rangle\} = [0,1]$ *(subproblem (2.1)). Setting* $x^2 = 0.1$ *and proceeding in this manner, DCA could generate a sequence* $\{x^k\}_{k\in\mathbb{N}} \subset (0,1)$ *such as* $(0.1, 0.9, 0.1, 0.9, \ldots)$*. Hence,* $\{\|x^k - x^{k+1}\|\}_{k\in\mathbb{N}}$ *is a constant sequence* $(0.8, 0.8, 0.8, \ldots)$ *whose limit is nonzero. In this case,* $\{\|x^k - x^{k+1}\|\}_{k\in\mathbb{N}}$ *does not satisfy Proposition 2.2.2. Note that the sequence* $\{f(x^k)\}$ *is the constant zero-sequence, which is convergent but without verifying item (ii) of Proposition 2.2.1.*

Theorem 2.2.1. *Every cluster point of the sequence* $\left\{x^k\right\}_{k\in\mathbb{N}}$ *generated by* ***Algorithm 1****, if any, is a critical point.*

Proof. Let x^* be a critical point of $\left\{x^k\right\}_{k\in\mathbb{N}}$. Then there exists a subsequence $\left\{x^{k_j}\right\}_{j\in\mathbb{N}}$, such that $\lim\limits_{j\to\infty} x^{k_j} = x^*$. In particular, $\left\{x^{k_j}\right\}_{j\in\mathbb{N}}$ is bounded.
We affirm that $\lim\limits_{j\to\infty} x^{k_j+1} = x^*$. Indeed, by triangular inequality, we have that

$$\begin{aligned} ||x^{k_j+1} - x^*|| &= ||(x^{k_j+1} - x^{k_j}) + (x^{k_j} - x^*)|| \\ &\leq ||x^{k_j+1} - x^{k_j}|| + ||x^{k_j} - x^*||. \end{aligned}$$

Then,

$$||x^{k_j+1} - x^*|| \leq ||x^{k_j+1} - x^{k_j}|| + ||x^{k_j} - x^*||, \forall j \in \mathbb{N}. \tag{2.6}$$

Taking the limit as $j \to \infty$ at (2.6) and applying the *Proposition 2.2.2*, we have that $\lim\limits_{j\to+\infty} ||x^{k_j+1} - x^*|| = 0$, hence $\lim\limits_{j\to+\infty} x^{k_j+1} = x^*$.
By the boundedness of $\left\{x^{k_j}\right\}_{j\in\mathbb{N}}$ and applying the *Proposition 1.2.2*, we obtain the boundedness of $\left\{w^{k_j}\right\}_{j\in\mathbb{N}}$, where $w^{k_j} \in \partial h(x^{k_j})$, for all $j \in \mathbb{N}$.
The first-order optimality condition for (2.1) at $k = k_j$ implies

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

Without loss of generality, we suppose $\left\{w^{k_j}\right\}_{j\in\mathbb{N}}$ convergent. Therefore,

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

By *Proposition 1.2.2*, we have that

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

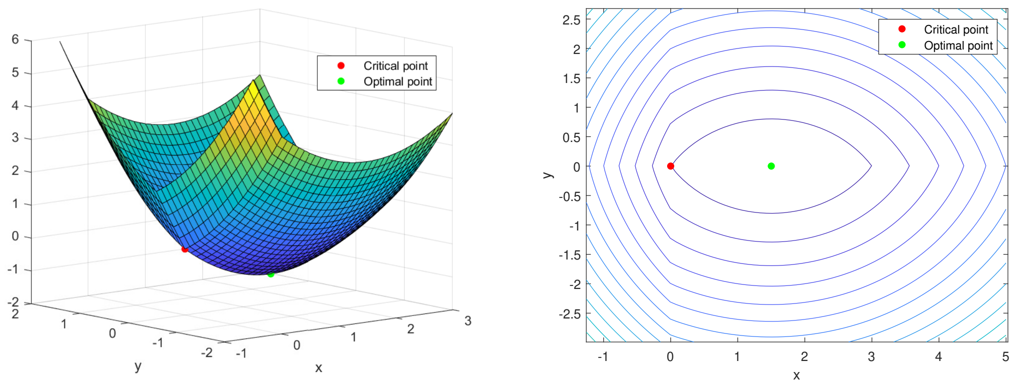
Therefore, x^* is a critical point. □

2.3 Numerical illustration

The numerical illustrations in this section were conducted using MATLAB software. The initial points were randomly chosen within the box $[-10, 10] \times [-10, 10]$. To solve the subproblems, we used the `fminsearch` toolbox with the inner stop rule: `optimset('TolX',1e-7,'TolFun',1e-7)`. The stopping criterion for the algorithm was $||x^{k+1} - x^k|| < 10^{-5}$.

Example 2.3.1. *Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *given by* $f(x, y) = \frac{1}{2}(x^2 + y^2) + |x| + |y| - \frac{5}{2}x$. *We can obtain a DC decomposition of* f *as follows:* $f(x, y) = g(x, y) - h(x, y)$, *where* $g(x, y) = x^2 + y^2 + |x| + |y| - \frac{5}{2}x$ *and* $h(x, y) = \frac{1}{2}(x^2 + y^2)$. *The minimum point of* f *is* $x_{opt} = (1.5, 0)$ *and the optimum value is* $f_{opt} = -1.125$.

To illustrate **Algorithm 1** applied to Example 2.3.1, we consider the point $x^0 = (-0.9259, 1.9735)$ chosen randomly, see Figure 2.2. In this case, the algorithm found the global solution of the problem, $x^* = (1.5, 0)$. Since $f(1.5, 0) = -1.125$, we have $f^* = -1.125$, which is the optimum value.

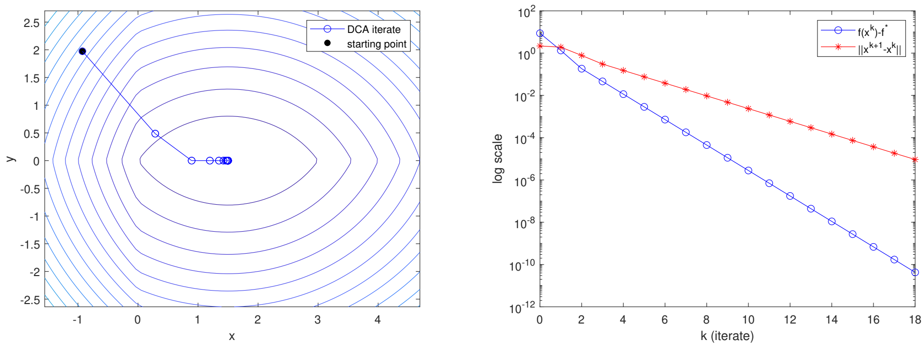


(a) Graph of the function f. (b) Contour lines.

Figure 2.1: Function f from Example 2.3.1.

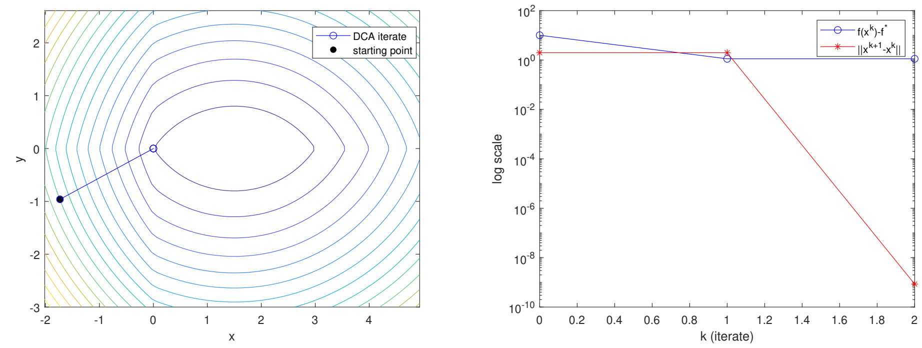
However, the function f does not have only one critical point. When the point $x^0 = (-1.7312, -0.9610)$ was randomly chosen, **Algorithm 1** found the critical point $x^* = (0.1613, 0.0763) \cdot 1.0e - 7$, which is very close to $(0, 0)$, another critical point. However, $(0, 0)$ is not the global minimum of f.

Figure 2.3 illustrates the convergence of the sequence generated by **Algorithm 1** approaching the critical point $(0, 0)$. In Table 2.1, we gather the data from **Algorithm**



(a) $x^0 = (-0.9259, 1.9735)$ (b) Values of $\|x^{k+1} - x^k\|$.

Figure 2.2: **Algorithm 1** for Example 2.3.1 with $x^0 = (-0.9259, 1.9735)$.



(a) $x^0 = (-1.7312, -0.9610)$ (b) Values of $\|x^{k+1} - x^k\|$.

Figure 2.3: **Algorithm 1** for Example 2.3.1 with $x^0 = (-1.7312, -0.9610)$.

1 being run 100 times, starting from random points taken within the box $[-10, 10] \times [-10, 10]$. In Table 2.1, min_k represents the minimum number of iterations taken to find a critical point; max_k denotes the maximum number of iterations required to find the critical point; med_k indicates the average number of iterations computed for the algorithm to find a critical point. The columns min_t, max_t, and med_t denote, respectively, the minimum, maximum, and average time for the algorithm to halt at a critical point. The last column min_{global} presents the percentage of times the algorithm found the optimal solution.

Table 2.1: 100 times with random initial points in $[-10, 10] \times [-10, 10]$.

Function f	min_k	max_k	med_k	min_t (s)	max_t (s)	med_t (s)	min_{global}
Example 2.3.1	2	20	19	0.0028469	0.20579	0.022441	77%
Example 2.3.2	11	14	13	0.01433	0.14426	0.020827	25%

Example 2.3.2. *Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *given by* $f(x, y) = x^2 + y^2 + x + y - |x| - |y|$*. We can obtain a DC decomposition of* f *as follows:* $f(x, y) = g(x, y) - h(x, y)$*, where* $g(x, y) = \frac{3}{2}(x^2 + y^2) + x + y$ *and* $h(x, y) = \frac{1}{2}(x^2 + y^2) + |x| + |y|$*.*

The function f from Example 2.3.2 has four critical points represented in its graph in Figure 2.4. However, it has a unique global minimizer, which is the point $x_{opt} = (-1, -1)$. To illustrate **Algorithm 1** in Example 2.3.2, we consider the initial point $x^0 = (-4.3119, -1.8040)$ chosen randomly. In Table 2.1, we can observe a trend of **Algorithm 1** finding the critical points of f.

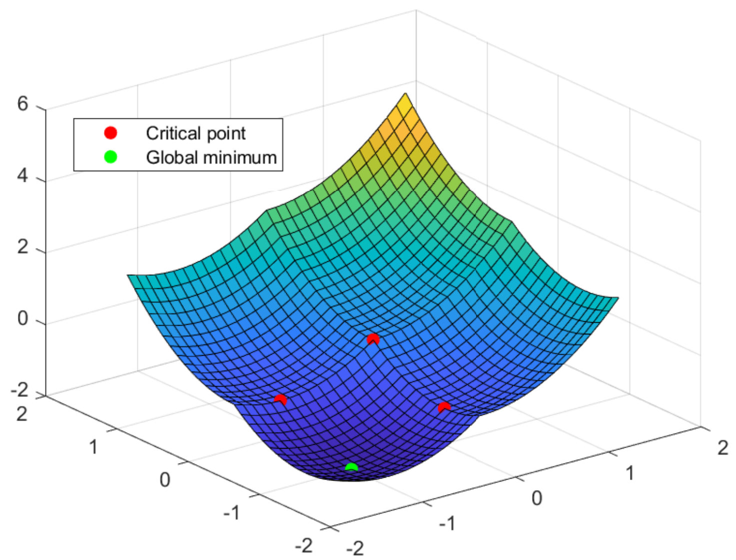
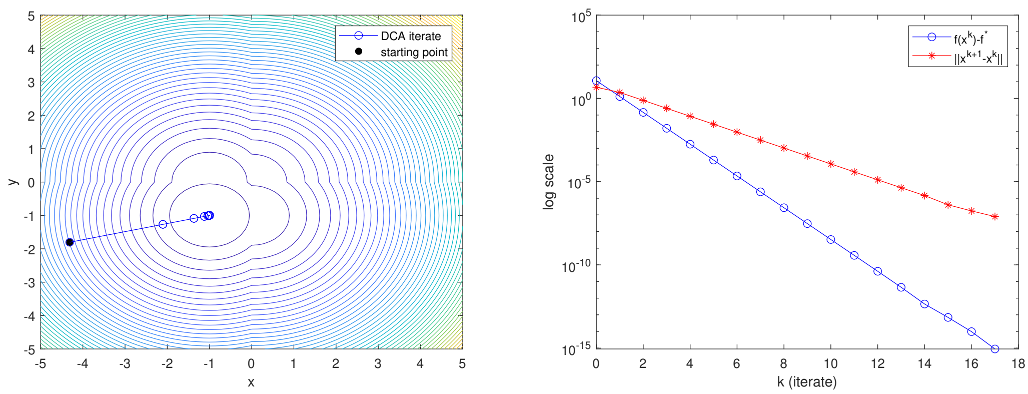


Figure 2.4: Graph of the function f from Example 2.3.2.



(a) $x^0 = (-4.3119, -1.8040)$.

(b) Values of $\|x^{k+1} - x^k\|$.

Figure 2.5: **Algorithm 1** for Example 2.3.2 with $x^0 = (-4.3119, -1.8040)$.

Chapter 3

The Boosted Difference of Convex Algorithm

In the work by Aragón Artacho et al. [3], a method was proposed to solve the DC problem in (1):

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

supposing that the DC components $g, h : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable strongly convex functions with modulus $\rho > 0$ and $\inf_{x \in \mathbb{R}^n} f(x) > -\infty$, named the Boosted Difference of Convex Algorithm (BDCA). However, we focus on the work of Aragón Artacho and Vuong [4], which relaxes the assumptions on the DC components by assuming only $g(x)$ is differentiable, consequently expanding the method's applicability to this specific class of non-differentiable functions. Thus, DBCA accelerates the convergence of DCA thanks to a line search step.

Throughout this chapter, consider the DC programming problem (3) considering the following assumptions:

(H1) $g, h : \mathbb{R}^n \to \mathbb{R}$ are both strongly convex with modulus $\rho > 0$;

(H2) $f^* := \inf_{x \in \mathbb{R}^n} f(x) > -\infty$.

(H3) $g : \mathbb{R}^n \to \mathbb{R}$ is differentiable.

The step size search structure considered in the algorithm requires assumption **(H3)**; otherwise, the obtained direction may be ascending (see Example 3.1.1). Assumptions **(H1)** and **(H2)** are reasonable and have also been considered and discussed in Chapter 2.

3.1 The algorithm

The Boosted Difference of Convex Algorithm (BDCA) was proposed by Aragón Artacho and Vuong [4] as a generalization of the method proposed by Aragón Artacho et al. [3]. The BDCA will be introduced in **Algorithm 2**.

Algorithm 2 Boosted Difference of Convex Algorithm (BDCA)[4].

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

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

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

3: Set $d^k := y^k - x^k$. **If** $d^k = 0$ **then** STOP and return x^k. **Otherwise**, choose any $\bar{\lambda}_k \geq 0$, and set $\lambda_k := \bar{\lambda}_k$. **While** $f(y^k + \lambda_k d^k) > f(y^k) - \alpha{\lambda_k}^2\|d^k\|^2$ DO $\lambda_k := \beta\lambda_k$.

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

Note that if $\bar{\lambda}_k = 0$, then the iterates of the BDCA coincide with the iterates of the DCA. In this case, convergence results apply, particularly to the DCA. We saw in Proposition 2.2.1 that, by setting $x^{k+1} := y^k$, the image of f undergoes a decrease, i.e., $f(y^k) \leq f(x^k)$. Furthermore, we will show that by taking x^{k+1} in the direction of d^k from y^k, we achieve an even greater decrease. This fact is the main idea behind the BDCA and enhances the performance of the DCA in many applications.

Remark 3.1.1. *Proceeding analogously to Remark 2.1.1, we verify that Step 2 is executable. To complete the well-definition of Algorithm 2, consider the following proposition:*

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

(i) $f(y^k) \leq f(x^k) - \rho\|d^k\|^2$;

(ii) $f'(y^k; d^k) \leq -\rho\|d^k\|^2$;

(iii) For all $k \in \mathbb{N}$, there exists some $\delta_k > 0$ such that

$$f(y^k + \lambda d^k) \leq f(y^k) - \alpha\lambda^2\|d^k\|, \quad \forall\lambda \in [0, \delta_k],$$

and hence, the line search in Step 3 of BDCA terminates finitely.

Proof. Consider $k \in \mathbb{N}$.

(i) Since y^k is the unique solution of (3.1), we have that $\nabla g(y^k) - w^k = 0$, i.e., $\nabla g(y^k) = w^k$. By the strong convexity of g and h, as $w^k \in \partial h(x^k)$ we have

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

and

$$h(y^k) - h(x^k) \geq \langle w^k, y^k - x^k \rangle + \frac{\rho}{2}\|y^k - x^k\|^2.$$

By adding both the inequalities above term by term, we have

$$\begin{aligned}
[g(x^k) - h(x^k)] - [g(y^k) - h(y^k)] &\geq \langle w^k, x^k - y^k \rangle + \langle w^k, y^k - x^k \rangle + \rho||x^k - x^{k+1}||^2 \\
&= \langle w^k, x^k - y^k \rangle - \langle w^k, x^k - y^k \rangle + \rho||x^k - y^k||^2 \\
&= \rho||x^k - y^k||^2.
\end{aligned}$$

We recall that $d^k = y^k - x^k$ and $f(x) = g(x) - h(x)$. Thus,

$$f(y^k) \leq f(x^k) - \rho\|d^k\|^2.$$

(ii) By definition of one-side directional derivative, we have

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

Since g is differentiable, we obtain

$$f'(y^k; d^k) = \langle \nabla g(y^k), d^k \rangle - \lim_{\lambda \downarrow 0} \frac{h(y^k + \lambda d^k) - h(y^k)}{\lambda}. \tag{3.2}$$

On the other hand, by convexity of h we can to choose $u \in \partial h(y^k)$, and then

$$h(y^k + \lambda d^k) - h(y^k) \geq \langle u, \lambda d^k \rangle,$$

which implies

$$\frac{h(y^k + \lambda d^k) - h(y^k)}{\lambda} \geq \langle u, d^k \rangle, \quad \forall \lambda > 0,$$

and hence,

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

Applying this fact in (3.2), we obtain

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

Therefore

$$f'(y^k; d^k) \leq \langle \nabla g(y^k) - u, y^k - x^k \rangle. \tag{3.3}$$

We recall that $\nabla g(y^k) = w^k \in \partial h(x^k)$. The function h is strongly convex with modulus ρ, then, by Proposition 1.2.3, ∂h is strongly monotone with modulus ρ. Thus, since $u \in \partial h(y^k)$, it holds that

$$\langle \nabla g(y^k) - u, x^k - y^k \rangle \geq \rho \|x^k - y^k\|^2.$$

Hence

$$\langle \nabla g(y^k) - u, y^k - x^k \rangle \leq -\rho \|x^k - y^k\|^2.$$

The proof follows by combining the last inequality with (3.3).

(iii) If $d^k = 0$ or $\lambda = 0$, there is nothing to prove. Otherwise, we have

$$\lim_{\lambda \downarrow 0} \frac{f(y^k + \lambda d^k) - f(y^k)}{\lambda} = f'(y^k; d^k) \leq -\rho \|x^k - y^k\|^2 < -\frac{\rho}{2} \|d^k\|^2.$$

Thus there is some $\bar{\delta}_k > 0$ such that

$$\frac{f(y^k + \lambda d^k) - f(y^k)}{\lambda} < -\frac{\rho}{2} \|d^k\|^2, \quad \forall \lambda \in \left(0, \bar{\delta}_k\right],$$

i.e.,

$$f(y^k + \lambda d^k) < f(y^k) - \frac{\rho \lambda}{2} \|d^k\|^2, \quad \forall \lambda \in \left(0, \bar{\delta}_k\right].$$

Pick $\lambda \in \left(0, \bar{\delta}_k\right]$. Hence

$$f(y^k) - \frac{\rho \lambda}{2} \|d^k\|^2 \leq f(y^k) - \alpha \lambda^2 \|d^k\|^2,$$

which is equivalent to

$$\lambda \leq \frac{\rho}{2\alpha}.$$

Therefore, setting $\delta_k := \min \left\{ \bar{\delta}_k, \dfrac{\rho}{2\alpha} \right\}$, we obtain

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

$\square$

Remark 3.1.2. *Since $\beta \in (0,1)$, then $\lim_{j\to\infty} \beta^j \bar{\lambda}_k = 0$ and, therefore, by Proposition 3.1.1(iii), the line search in Step 3 is well-defined. Combining this fact with Remark 3.1.1 we conclude that* ***Algorithm 2*** *is well-defined.*

The following example illustrates a case where the DC function f has a non-differentiable first component. We will show that, in this case, the direction computed by BDCA is ascending from the found solution.

Example 3.1.1. *([4, Example 3.4]) Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a DC function with DC components*

$$g(x) = -\frac{5}{2}x_1 + x_1^2 + x_2^2 + |x_1| + |x_2| \quad \textit{and} \quad h(x) = \frac{1}{2}(x_1^2 + x_2^2).$$

The minimum point of ϕ is $x^* = (1.5, 0)^\top$ and the optimum value is $\phi^* = -1.125$.

Clearly, the second component h is differentiable, but g is not. Following the definition of **Algorithm 2**, in Step 1 we choose $x^0 = (\frac{1}{2}, 1)$. In Step 2, we compute $w^0 \in \partial h\left(\frac{1}{2}, 1\right)$. Since h is differentiable,we apply Proposition 1.2.1 and obtain $w^0 = \nabla h\left(\frac{1}{2}, 1\right) = \left(\frac{1}{2}, 1\right)$. Continuing, we solve the subproblem (3.1), whose solution is $y^0 = (1, 0)$ and thus $d^0 = \left(\frac{1}{2}, -1\right)$. Now, we assert that d^0 is not a descent direction from y^0. In fact, we can verify that

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

descent direction of f from y^0. In addiction,

$$f(y^0 + \lambda d) - f(y^0) = \left(-1 + \frac{3}{4}\lambda + \frac{5}{8}\lambda^2\right) - 1 = \frac{3}{4}\lambda + \frac{5}{8}\lambda^2 > 0, \quad \forall \lambda > 0,$$

which shows that d^0 is, in fact, an ascending direction of f from y^0.

Lemma 3.1.1. *If $y^k = x^k$ for some $k \in \mathbb{N}$, then x^k is a critical point of f.*

Proof. Indeed, the expression at (3.1) is equivalent to

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

On the other hand, $w^k \in \partial h(x^k)$, due to Step 2. Consequently, if $y^k = x^k$, it follows that $\partial g(x^k) \cap \partial h(x^k) = \{\nabla g(x^k)\}$. Thus, x^k is a critical point of f. $\square$

3.2 Convergence analysis

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

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

(ii) The sequence $\{f(x^k)\}_{k\in\mathbb{N}}$ *is decreasing, i.e.,* $f(x^{k+1}) \leq f(x^k)$*, for all* $k \in \mathbb{N}$*.*

Proof. Conforming to Step 3, the **Algorithm 2** terminates when $y^k = x^k$. By Lemma 3.1.1, x^k is a critical point, and this proves item *(i)*. By combining the items *(i)* and *(iii)* of the Proposition 3.1.1 with Step 4, we obtain

$$\begin{aligned} f(y^k + \lambda_k d^k) &\leq f(y^k) - \alpha\lambda_k^2 \|d^k\| \\ &\leq f(x^k) - \rho\|d^k\|^2 - \alpha\lambda_k^2\|d^k\| \\ &= f(x^k) - (\rho + \alpha\lambda_k^2)\|d^k\|^2, \quad \forall k \in \mathbb{N}, \end{aligned} \tag{3.4}$$

where $\rho > 0$ is the constant of strong convexity for the functions g and h. Thus, since $d^k \neq 0$ and $x^{k+1} = y^k + \lambda_k d^k$, we have

$$f(x^{k+1}) = f(y^k + \lambda_k d^k) < f(x^k), \quad \forall k \in \mathbb{N},$$

which completes the proof. □

Remark 3.2.1. *When* $d^k \neq 0$*, the inequalities at* (3.4) *shows that The BDCA method provides a decrease in the objective function* f*, at each iteration, larger than the DCA. Due to this, it is expected that the BDCA converges faster than the DCA.*

Corollary 3.2.1. *If* $\{x^k\}_{k\in\mathbb{N}}$ *is a sequence generated by* ***Algorithm 2****, then the sequence* $\{f(x^k)\}_{k\in\mathbb{N}}$ *is convergent.*

Proof. This follows immediately from the fact that f is lower-bounded, by assumption **(H3)**, and $\{f(x^k)\}_{k\in\mathbb{N}}$ is decreasing. □

Proposition 3.2.2. *If* $\{x^k\}_{k\in\mathbb{N}}$ *is a sequence generated by* ***Algorithm 2****, then the following statements holds:*

(i) $\sum_{k=0}^{+\infty} \|d^k\|^2 < +\infty$*. In particular,* $\|y^k - x^k\| \to 0$ *as* $k \to +\infty$*.*

(ii) If $\{\lambda_k\}_{k\in\mathbb{N}}$ *is bounded, then* $\sum_{k=0}^{+\infty}\|x^{k+1}-x^k\|^2<+\infty$*. In particular,* $\|x^{k+1}-x^k\|\to 0$ *as* $k\to+\infty$*.*

Proof.

(i) By Proposition 3.1.1 and Step 4 of **Algorithm 2**, we have

$$f(x^{k+1})\leq f(x^k)-\left(\alpha\lambda_k^2+\rho\right)\|d^k\|^2,\quad \forall k\in\mathbb{N}.$$

Thus,

$$\left(\alpha\lambda_k^2+\rho\right)\|d^k\|^2\leq f(x^k)-f(x^{k+1}),\quad \forall k\in\mathbb{N}.$$

Since $\alpha\lambda_k^2\geq 0$, it holds that $\rho\|d^k\|^2\leq(\alpha\lambda_k^2+\rho)\|d^k\|^2$. Hence,

$$\rho||d^k||^2\leq f(x^k)-f(x^{k+1}),\quad \forall k\in\mathbb{N}. \tag{3.5}$$

Considering the partial sum at inequality (3.5):

$$0\leq\sum_{k=0}^{N}\rho||d^k||^2\leq\sum_{k=0}^{N}\left[f(x^k)-f(x^{k+1})\right]$$

we obtain

$$0\leq\sum_{k=0}^{N}\rho||d^k||^2\leq f(x^0)-f(x^{N+1})$$

Since $f^*=\inf_{x\in\mathbb{R}^n}f(x)$, by **(H2)**, then $f^*\leq f(x^{N+1})$ implies $-f^*\geq -f(x^{N+1})$, for all $N\in\mathbb{N}$. Therefore,

$$0\leq\sum_{k=0}^{N}\rho||d^k||^2\leq f(x^0)-f^*,\forall k\in\mathbb{N}. \tag{3.6}$$

Since $\rho>0$ is constant, taking the limit as $N\to+\infty$ at (3.6) we have

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

(ii) Let $\bar{\lambda}>0$ be such that $0\leq\lambda_k\leq\bar{\lambda}$, for all $k\in\mathbb{N}$. Firstly, note that

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

Because of this, and by using item *(i)*, it holds that

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

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

□

Theorem 3.2.1. *Every cluster point of $\left\{x^k\right\}_{k\in\mathbb{N}}$ generated by **Algorithm 2**, if any, is a critical point of f.*

Proof. Since x^* is a critical point of $\left\{x^k\right\}_{k\in\mathbb{N}}$, there is a subsequence $\left\{x^{k_j}\right\}_{j\in\mathbb{N}}$ such that $\lim_{j\to+\infty} x^{k_j} = x^*$, i.e., $\lim_{j\to+\infty} ||x^{k_j} - x^*|| = 0$. By the last proposition, if $\left\{y^k\right\}_{k\in\mathbb{N}}$ is the sequence generated by Step 2 of **Algorithm 2**, then $\lim_{k\to+\infty} ||y^k - x^k|| = 0$. Hence,

$$\begin{aligned} \|y^{k_j} - x^*\| &= \|y^{k_j} - x^{k_j} + x^{k_j} - x^*\| \\ &\leq \|y^{k_j} - x^{k_j}\| + \|x^{k_j} - x^*\|, \quad \forall j \in \mathbb{N}, \end{aligned}$$

which implies that $\lim_{j\to+\infty} \|y^{k_j} - x^*\| = 0$, i.e., $\lim_{j\to+\infty} y^{k_j} = x^*$. Step 2 give us $\nabla g(y^{k_j}) \in \partial h(x^{k_j})$, for all $j \in \mathbb{N}$. By Proposition 1.2.2 with Corollary 1.2.1, it follows that

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

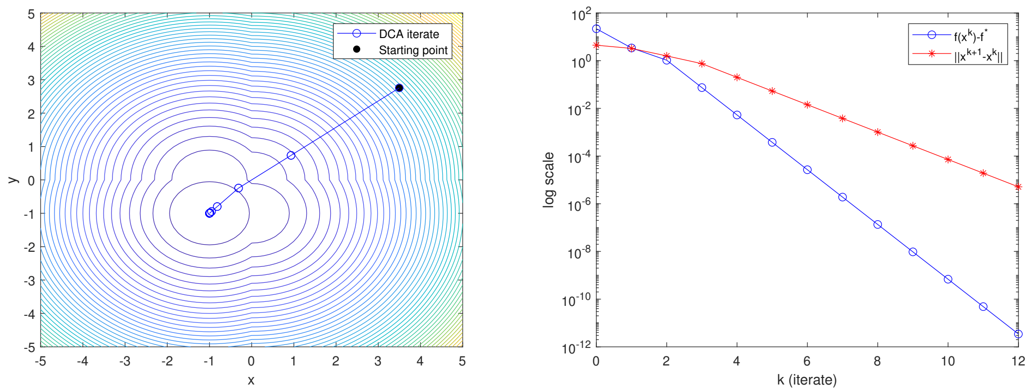
Thus, x^* is a critical point of f. □

3.3 Numerical illustration

The numerical illustrations in this section were conducted using MATLAB software. The initial points were randomly chosen within the box $[-10, 10] \times [-10, 10]$. To solve the subproblems, we used the `fminsearch` toolbox with the inner stop rule: `optimset('TolX',1e-7,'TolFun',1e-7)`. The stopping criterion for the algorithm was $||x^{k+1} - x^k|| < 10^{-5}$. In the example below, the constants in the definition of Algorithm 2 were set as $\alpha = 0.6$, $\beta = 0.1$ and $\bar{\lambda} = 1$.

Example 3.3.1. *(Example 2.3.2 revisited) Let $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2+y^2+x+y-|x|-|y|$. We can obtain a DC decomposition of f as follows: $f(x, y) = g(x, y)-h(x, y)$, where $g(x, y) = \frac{3}{2}(x^2 + y^2) + x + y$ and $h(x, y) = \frac{1}{2}(x^2 + y^2) + |x| + |y|$. The minimum point of f is $x_{opt} = (-1, -1)$ and the optimum value is $f_{opt} = -2$.*

To illustrate **Algorithm 2**, we consider Example 3.3.1 and take the initial point $x^0 = (3.4975, 2.7560)$ randomly chosen within the box $[-5, 5] \times [-5, 5]$. In this case, the algorithm finds the critical point $x^* = (-1.0000, -1.0000)$, which is exactly the global minimum of f. The convergence of the method is illustrated in Figure 3.1. Note in Figure 3.1(b) that toel $= 1e - 5$ is being satisfied.



(a) $x^0 = (3.4975, 2.7560)$. (b) Values of $\|x^{k+1} - x^k\|$.

Figure 3.1: **Algorithm 2** for Example 3.3.1.

To illustrate a comparison between **Algorithm 1** and **Algorithm 2**, we selected 100 random points within the box $[-10, 10] \times [-10, 10]$ and compiled the data in Table 3.3, where the column min_k represents the minimum number of iterations taken to find a critical point, max_k denotes the maximum number of iterations required to find the critical point, med_k indicates the average number of iterations computed for the algorithm to find a critical point. The columns min_t, max_t, and med_t denote, respectively, the minimum, maximum, and average time for the algorithm to halt at a critical point. The last column min_{global} presents the percentage of times the algorithm found the optimal solution. The data collected for the column min_{global} were obtained separately so that storing the images would not affect the time.

Table 3.1: Comparing **Algorithm 1** and **Algorithm 2** 100 times with random initial points in $[-10, 10] \times [-10, 10]$.

Function	min_k	max_k	med_k	min_t (s)	max_t (s)	med_t (s)	min_{global}
Algorithm 1 (DCA)	11	14	13	0.010056	0.033005	0.015545	28%
Algorithm 2 (BDCA)	9	16	12	0.0097658	0.1503	0.014328	100%

Chapter 4

Non-monotone Boosted Difference of Convex Algorithm

In Chapter 3, we approached the BDCA, which strongly considers the differentiability of the first DC component in problem (3) to guarantee that the direction defined in the algorithm is descent. As observed, this process enhances the DCA by further decreasing the function f along the sequence generated by the BDCA, thus preserving the monotonicity of the search.

The method proposed by Ferreira, Santos, and Souza [21] employs a non-monotone line search in BDCA to enable a possible growth in the objective function values controlled by a parameter. Before to introduce the method, we recall the DC problem presented at (1):

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

where $g, h : \mathbb{R}^n \to \mathbb{R}$ are convex functions.

Throughout this chapter, we will use the same assumptions as in the previous chapter, except for the differentiability of the function g. In other words, we will assume:

(H1) $g, h : \mathbb{R}^n \to \mathbb{R}$ are both strongly convex with modulus $\rho > 0$;

(H2) $f^* := \inf_{x \in \mathbb{R}^n} \{f(x) = g(x) - h(x)\} > -\infty$.

The assumptions **(H1)** and **(H2)** are the same as those adopted in Chapter 2, where they were analyzed and discussed.

4.1 Exact computation of the subproblems

The main idea of the Non-monotone Boosted Difference of Convex Algorithm (nmB-DCA) is to allow potential growth in the objective function values controlled by a parameter, enabling the removal of the differentiability assumption in the first DC component considered in [3] and [4].

Next, we present the structure of the nmBDCA followed by the results that can be found in [21]. In this section, proofs will be omitted as they are recovered by the inexact method when choosing parameters appropriately.

Algorithm 3 Non-monotone Boosted Difference of Convex Algorithm (nmBDCA)[21].

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

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

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

3: Set $d^k := y^k - x^k$. **If** $d^k = 0$ **then** STOP and return x^k. **Otherwise**, choose $v_k \geq 0$ (to be specified later), any $\bar{\lambda}_k \geq 0$, and set $\lambda_k := \bar{\lambda}_k$. **While** $f(y^k + \lambda_k d^k) > f(y^k) - \alpha {\lambda_k}^2 \|d^k\|^2 + v_k$ DO $\lambda_k := \beta \lambda_k$.

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

Note that we have a term v_k added to the line search described in Step 3. Some examples of sequences $\{v_k\}_{k \in \mathbb{N}}$ will be mentioned in the inexact method section (see section 4.1.3).

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 of f.

(ii) There holds $f(y^k) \leq f(x^k) - \rho||d^k||^2$.

Proof. See [3, Proposition 3]. □

To ensure the well-definition of **Algorithm 3**, we consider two cases: first, g is possibly non-differentiable; second, g is continuously differentiable.

4.1.1 Well-definedness of nmBDCA: g possibly non-differentiable

In this section, we consider the function g is possibly non-differentiable. However, we also need to assume that $v_k > 0$. The next proposition guarantees that **Algorithm 3** is well-defined.

Proposition 4.1.2. *Let $\{x^k\}_{k\in\mathbb{N}}$ be the sequence generated by **Algorithm 3**. Assume that $d^k \neq 0$ and $v_k > 0$ for each $k \in \mathbb{N}$. Then, the following statements hold:*

(i) There holds $\bar{\delta}_k := v_k/(g(y^k + d^k) + g(x^k) - 2g(y^k)) > 0$, and

$$f(y^k + \lambda d^k) \leq f(y^k) - \alpha\lambda^2\|d^k\|^2 + v_k, \quad \forall\lambda \in (0, \delta_k]\,,$$

where $\delta_k := \min\left\{\bar{\delta}_k, 1, \frac{3\rho}{2\alpha}\right\}$. Consequently, the line search in Step 3 is well-defined.

(ii) $f(x^{k+1}) \leq f(x^k) - (\rho + \alpha\lambda_k^2)||d^k||^2 + v_k$, for all $k \in \mathbb{N}$.

Proof. See [21, Proposition 13]. □

Remark 4.1.1. *Since $x^{k+1} := y^k + \lambda_k d^k$, then $x^{k+1} = x^k$ implies that $x^k = y^k + \lambda_k d^k$, and hence $(1 + \lambda_k)d^k = 0$. As $\lambda_k > 0$, we obtain that $d^k = 0$. By Proposition 4.1.1, x^k is a critical point of f. Thus, we assume the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated by **Algorithm 3** is infinite.*

As mentioned by Ferreira, Santos, and Sousa [21], when g is convex and nondifferentiable, the direction $d^k \neq 0$ generated by Step 3 in **Algorithm 3** may not be a descent direction of f from y^k; this can be verified in Example 4.3.2. Because of this, we need to assume that $v^k > 0$; otherwise, we could not compute $\lambda_k > 0$ satisfying the line search in Step 3.

Next, we will see that considering g convex and differentiable, we can assume $v_k \geq 0$ to compute $\lambda_k > 0$ satisfying the line search in **Algorithm 3**.

4.1.2 Well-definedness of nmBDCA: g differentiable

Considering the function g to be continuously differentiable, we can assume that $v_k \geq 0$. Thus, in addition to hypotheses **(H1)** and **(H2)**, we adopt the following assumption:
(H3) g is differentiable.

Throughout this work, **(H3)** will be used only when explicitly mentioned.

Proposition 4.1.3. *Suppose that $g : \mathbb{R}^n \to \mathbb{R}$ satisfies **(H3)**. For each $k \in \mathbb{N}$, assume that $d^k \neq 0$ and $v_k \geq 0$. Then, the following statements hold:*

(i) $f'(y^k; d^k) \leq \rho||d^k||^2 < 0$ *and there exists a constant δ_k such that* $f(y^k + \lambda d^k) \leq f(y^k) - \rho\lambda^2||d^k||^2 + v_k$*, for all $\lambda \in (0, \delta_k]$. Consequently, the line search in Step 4 is well-defined.*

(ii) $f(x^{k+1}) \leq f(x^k) - (\rho + \alpha\lambda_k^2)\,||d^k||^2 + v_k$.

Proof. See [21, Proposition 14]. □

Note that when $v_k = 0$, the non-monotone line search of **Algorithm 3** retrieves the monotone line search of **Algorithm 2**. Thus, as pointed out by Ferreira, Santos, and Sousa [21], the nmBDCA is a natural extension of the BDCA established in [4] and [3]. Moreover, if $v_k > 0$, then the nmBDCA can be seen as a relaxed version of the BDCA.

4.1.3 Strategies to choose v_k

In this section, we will introduce some of the sequence selection strategies $\{v_k\}_{k\in\mathbb{N}}$ suggested by Ferreira, Santos, and Souza [21].

(E1) Given $\Delta_{min} \in [0, 1)$, the sequence $\{v_k\}_{k\in\mathbb{N}}$ is defined as follows: $v_0 \geq 0$ and v_{k+1}, for each $\Delta_{k+1} \in [\Delta_{min}, 1]$, satisfies the following condition:

$$0 \leq v_{k+1} \leq (1 - \Delta_{k+1})\left(f(x^k) - f(x^{k+1}) + v_k\right), \quad \forall k \in \mathbb{N};$$

(E2) $\{v_k\}_{k\in\mathbb{N}}$ is such that $\sum_{k=0}^{+\infty} v_k < +\infty$;

(E3) $\{v_k\}_{k\in\mathbb{N}}$ satisfies: for every, $\delta > 0$, there exists $k_0 \in \mathbb{N}$ such that $v_k \leq \delta||d^k||^2$, for all $k \geq k_0$.

Remark 4.1.2. *In Lemma 4.2.1, as particular case, we will prove that if $\Delta_{min} > 0$, then **(E1)** implies **(E2)**.*

As we will see in Section 4.2.1, strategies **(E1)**, **(E2)**, and **(E3)** will be employed in the proposal of an inexact method that is a relaxed version of **Algorithm 2**. With a subtle modification, we revive the described strategies, and for this reason, the comments and examples will be maintained in Section 4.2.1.

4.1.4 Convergence analysis: g possibly non-differentiable

In this section, convergence results will be presented. Here, results from Section 4.1.1 will be used; hence, we also need to assume that $v_k > 0$, for all $k \in \mathbb{N}$.

The results to be presented were established by Ferreira, Santos, and Souza [21]. In Section 4.2.2, we will see that, under reasonable assumptions, the following theorems are particular cases of their inexact version, proposed in Section 4.2. Therefore, their proofs will be omitted.

Theorem 4.1.1. *If* $\lim_{k\to\infty} \|d^k\| = 0$*, then every cluster point of* $\{x^k\}_{k\in\mathbb{N}}$*, if any, is a critical point of* f*.*

Proof. See [21, Theorem 15]. □

Theorem 4.1.2. *If the sequence* $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *is chosen either according to strategy* ***(E2)****, or to strategy* ***(E3)****, then every cluster point of* $\{x^k\}_{k\in\mathbb{N}}$*, if any, is a critical point of* f*.*

Proof. See [21, Theorem 16]. □

Theorem 4.1.3. *If the sequence* $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *is chosen according to strategy* ***(E1)****, then the following statements hold:*

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

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

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

Proof. See [21, Theorem 17]. □

4.1.5 Iteration-complexity analysis

The following theorems established by Ferreira, Santos, and Souza [21] present some complexity limitations for $\{x^k\}_{k\in\mathbb{N}}$ generated by **Algorithm 2**. The following results address the cases where the sequence $\{v_k\}_{k\in\mathbb{N}}$ was chosen according to the strategies **(E2)** and **(E3)**.

Theorem 4.1.4. *Suppose that the sequence $(v_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is chosen according to strategy **(E2)**. For each $N \in \mathbb{N}$, we have*

$$\min\left\{\|d^k\| : \ k = 0, 1, \cdots, N-1\right\} \leq \frac{\sqrt{f(x^0) - f^* + \sum_{k=0}^{+\infty} v_k}}{\sqrt{\rho}} \frac{1}{\sqrt{N}}.$$

Consequently, for a given accuracy $\varepsilon > 0$, if $N \geq \left(f(x^0) - f^* + \sum_{k=0}^{\infty} v_k\right)/(\rho\epsilon^2)$, then the following inequality holds $\min\{\|d^k\| : \ k = 0, 1, \cdots, N-1\} \leq \varepsilon$.

Proof. See [21, Theorem 18]. □

Theorem 4.1.5. *Suppose that the sequence $(v_k)_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is chosen according to strategy **(S3)**. Let $0 < \xi < 1$ and $k_0 \in \mathbb{N}$ such that $v_k \leq \xi\rho\|d^k\|^2$, for all $k \geq k_0$. Then, for each $N \in \mathbb{N}$ such that $N > k_0$, one has*

$$\min\{\|d^k\| : k = 0, 1, \cdots, N-1\} \leq \frac{\sqrt{f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k}}{\sqrt{(1-\xi)\rho}} \frac{1}{\sqrt{N}}.$$

Consequently, for a given $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ such that $v_k \leq \xi\rho\|d^k\|^2$ for all $k \geq k_0$, if $N \geq$ $\max\{k_0, (f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k)/(\rho(1-\xi)\varepsilon^2)\}$, *then* $\min\left\{\|d^k\| : \ k = 0, 1, \cdots, N-1\right\} \leq \varepsilon$.

Proof. See [21, Theorem 19]. □

4.2 Inexact computation of the subproblems

We propose an inexact approach (InmBDCA) for the Non-monotone Boosted Difference of Convex Algorithm (nmBDCA), enabling a relaxed version that, under suitable assumptions, naturally recovers the exact version.

Consider the exact model described in **Algorithm 3**. From a computational standpoint, it can be difficult to compute an element in the subdifferential of the function h or to solve the subproblem (4.1). Often, the subproblems are solved approximately by adjusting parameters that make the inexact solution appropriate for each situation.

Proposition 4.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a DC function, given by $f(x) = g(x) - h(x)$, where $g, h : \mathbb{R}^n \to \mathbb{R}$ are strongly convex with modulus $\rho > 0$ and g is differentiable. Let $x \in \mathbb{R}^n$ and $w \in \partial h(x)$ be given. Then, for every $y \in \mathbb{R}^n$ such that $y \neq x$ and $||\nabla g(y) - w|| < \rho||y - x||$, we have that $f'(y; y - x) < 0$.*

Proof. Given $y \in \mathbb{R}^n$, we have

$$f'(y; y - x) = g'(y; y - x) - h'(y; y - x). \tag{4.2}$$

In particular, since g is differentiable, $g'(y; y - x) = \langle \nabla g(y), y - x \rangle$ and, by Theorem 1.2.1, $h'(y; y - x) = \max_{s \in \partial h(y)} \langle s, y - x \rangle$. Thus, (4.2) becomes

$$\begin{aligned} f'(y; y - x) &\leq \langle \nabla g(y), y - x \rangle - \langle v, y - x \rangle \\ &= \langle \nabla g(y) - v, y - x \rangle, \quad v \in \partial h(y). \end{aligned} \tag{4.3}$$

On the other hand,

$$\begin{aligned} \langle \nabla g(y) - v, y - x \rangle &= \langle \nabla g(y) - w + w - v, y - x \rangle \\ &= \langle \nabla g(y) - w, y - x \rangle + \langle w - v, y - x \rangle. \end{aligned} \tag{4.4}$$

Since $w \in \partial h(x)$ and ∂h is strongly monotone with modulus $\rho > 0$ (Theorem 1.2.3), we obtain

$$\langle w - v, y - x \rangle \leq -\rho ||y - x||^2. \tag{4.5}$$

Moreover, by Cauchy-Schwarz inequality,

$$\langle \nabla g(y) - w, y - x \rangle \leq ||\nabla g(y) - w|| \cdot ||y - x||. \tag{4.6}$$

Substituting (4.6) and (4.5) into (4.4), we obtain

$$\langle \nabla g(y) - v, y - x \rangle \leq ||\nabla g(y) - w|| \cdot ||y - x|| - \rho ||y - x||^2. \tag{4.7}$$

By combining (4.7) with (4.3) and using the fact that $||\nabla g(y) - w|| < \rho ||y - x||$, we have

$$\begin{aligned} f'(y; y - x) &\leq ||\nabla g(y) - w|| \cdot ||y - x|| - \rho ||y - x||^2 \\ &< \rho ||y - x|| \cdot ||y - x|| - \rho ||y - x||^2 \\ &= 0. \end{aligned}$$

□

Remark 4.2.1. *Note that if y^k is the exact solution of subproblem* (3.1) *in Step 2 of BDCA and $d^k = y^k - x^k \neq 0$, then $\nabla g(y^k) = w^k \in \partial h(x^k)$, and hence,*

$$0 = ||\nabla g(y^k) - \nabla g(y^k)|| = ||\nabla g(y^k) - w^k|| < \rho ||y^k - x^k||.$$

In other words, for every $k \in \mathbb{N}$, the exact solution y^k of of subproblem (3.1) *in Step 2 of BDCA satisfies $||\nabla g(y^k) - w^k|| < \rho ||y^k - x^k||$.*

Therefore, a natural question arises: besides the exact solution of subproblem (3.1), are there other points $y \in \mathbb{R}^n$ such that $d = y - x^k$ is a descent direction of f at y? The following corollary provides an answer to this question.

Corollary 4.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a DC function, given by $f(x) = g(x) - h(x)$, where $g, h : \mathbb{R}^n \to \mathbb{R}$ are strongly convex with modulus $\rho > 0$ and g is differentiable. Let $x \in \mathbb{R}^n$ and $w \in \partial h(x)$ be given. If $\hat{y} \in \mathbb{R}^n$ is such that $\hat{y} \neq x$ and $||\nabla g(\hat{y}) - w|| < \rho||\hat{y} - x||$, then there exists $r > 0$ such that, for all $y \in B(\hat{y}; r)$,*

$$||\nabla g(y) - w|| < \rho||y - x||. \tag{4.8}$$

Proof. Take any $\hat{y} \in \mathbb{R}^n$ satisfying (4.8). Let us suppose that, for each $l \in \mathbb{N}$, there exists $y^l \in \mathbb{R}^n$ such that $\|\hat{y} - y^l\| < \frac{1}{l}$ and (4.8) does not hold, i.e.,

$$\rho||y^l - x|| \leq ||\nabla g(y^l) - w||. \tag{4.9}$$

Taking the limit as $l \to +\infty$ on (4.9), by Proposition 1.2.1, ∇g is continuous, and hence,

$$\rho||\hat{y} - x|| \leq ||\nabla g(\hat{y}) - w||,$$

which is a contradiction. □

To ensure mathematical efficiency of the nmBDCA even when handling the problem in an approximate manner, motivated by Proposition 4.2.1 and Corollary 4.2.1, we propose an Inexact Non-monotone Boosted Difference of Convex Algorithm (InmBDCA) defined in **Algorithm 4**. In particular, our proposal is also an inexact method for BDCA. To the best our knowledge, inexact versions of BDCA or nmBDCA have not been considerd in the literature. On the other hand, many works have considered addressing various problems in an inexact context, see, for instance, [14, 20, 33, 37, 42, 43, 44, 50].

Remark 4.2.2. *Setting $\theta = 0$ and $\varepsilon_k = 0$ for all $k \in \mathbb{N}$, we have $w^k \in \partial h(x^k)$, and* (4.11) *implies $w^k = \xi^k$. Thus, the subproblem in Step 2 becomes to compute y^k such that $w^k \in \partial g(y^k)$, i.e.,*

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

which corresponds to (4.1)*. As a result,* ***Algorithm 4*** *is an inexact version of* ***Algorithm 3****.*

Algorithm 4 Inexact Non-monotone Boosted Difference of Convex Algorithm

1: Fix $\alpha > 0$, $\theta \in \left[0, \frac{\rho}{2}\right)$, $\beta \in (0,1)$, and a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_+$. Choose any initial point $x^0 \in \mathbb{R}^n$ and set $k := 0$.

2: Choose $w^k \in \partial_{\varepsilon_k} h(x^k)$ and compute $\left(y^k, \xi^k\right)$ a solution of the following subproblem

$$\xi^k \in \partial g(y^k), \tag{4.10}$$

$$||w^k - \xi^k|| \leq \theta ||y^k - x^k||. \tag{4.11}$$

3: Set $d^k := y^k - x^k$. **If** $d^k = 0$ **then** STOP and return x^k. **Otherwise**, choose $v_k \geq 0$ (to be specified later), any $\bar{\lambda}_k \geq 0$, and set $\lambda_k := \bar{\lambda}_k$. **While** $f(y^k + \lambda_k d^k) > f(y^k) - \alpha {\lambda_k}^2 \|d^k\|^2 + v_k$ DO $\lambda_k := \beta\lambda_k$.

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

Remark 4.2.3. *Note that, in the exact case, the convex minimization subproblem in* ***Step 2*** *consists of finding*

$$y^k = \underset{x\in\mathbb{R}^n}{\arg\min}\, g(x) - \langle w^k, x - x^k\rangle, \tag{4.12}$$

which is generally solved approximately. Thus, one way to verify (4.10) *and* (4.11) *is to check if the solution y^k obtained from the machine in* (4.12) *satisfies these conditions. For example, when the DC component $g : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then $\xi^k = \nabla g(y^k)$, and in this case,* (4.11) *can be easily verified, provided that we know the expression for the gradient of g. If g is not differentiable, the verification may become more difficult, as we would need to know the subdifferential of g to take $\xi^k \in \partial g(y^k)$ and then verify* (4.11)*. In this way, in both cases, the value of θ could be adjusted so that* (4.11) *is satisfied.*

Remark 4.2.4. *By using the Cauchy-Schwarz Inequality, we have that $\langle \xi^k - w^k, y^k - x^k\rangle \leq ||\xi^k - w^k|| \cdot ||y^k - x^k||$, which, when combined with equation* (4.11)*, yields*

$$\langle \xi^k - w^k, y^k - x^k\rangle \leq \theta||y^k - x^k||^2 \leq \frac{\rho}{2}||y^k - x^k||^2,$$

then

$$\langle \xi^k, x^k - y^k\rangle + \frac{\rho}{2}||y^k - x^k||^2 \geq -\langle w^k, y^k - x^k\rangle. \tag{4.13}$$

On the other hand, since $\xi^k \in \partial g(y^k)$ and g is strongly convex with modulus $\rho > 0$, it follows that

$$g(x^k) \geq g(y^k) + \langle \xi^k, x^k - y^k\rangle + \frac{\rho}{2}||x^k - y^k||^2,$$

which combing with (4.13)*, becomes*

$$g(x^k) \geq g(y^k) - \langle w^k, y^k - x^k \rangle. \tag{4.14}$$

Proposition 4.2.2. *For each $k \in \mathbb{N}$ the following statements hold:*

(i) If $d^k = 0$, then x^k is a ε_k-critical point of f;

(ii) There holds $f(y^k) \leq f(x^k) - \left(\frac{\rho}{2} - \theta\right) ||d^k||^2 + \varepsilon_k$.

Proof. Proof of item (i): fix any $k \in \mathbb{N}$ and take $w^k \in \partial_{\varepsilon_k} h(x^k)$. Since $d^k = y^k - x^k$ and $||w^k - \xi^k|| \leq \theta ||y^k - x^k||$, by (4.11), if $d^k = 0$, then $y^k = x^k$ and $w^k = \xi^k \in \partial g(y^k) = \partial g(x^k) \subset \partial_{\varepsilon_k} g(x^k)$. Therefore, $\partial_{\varepsilon_k} g(x^k) \cap \partial_{\varepsilon_k} h(x^k) \neq \varnothing$, which proves this item.

Proof of item (ii): Since $w^k \in \partial_{\varepsilon_k} h(x^k)$, by Definition 1.2.2, it follows that

$$h(y^k) \geq h(x^k) + \langle w^k, y^k - x^k \rangle - \varepsilon_k \tag{4.15}$$

The function g is strongly convex with modulus $\rho > 0$. Then, by (4.10) and Theorem 1.2.3, it holds

$$g(x^k) \geq g(y^k) + \langle \xi^k, x^k - y^k \rangle + \frac{\rho}{2} \|x^k - y^k\|^2 \tag{4.16}$$

By adding both the inequalities (4.15) and (4.16) term by term, we obtain

$$\begin{aligned} h(y^k) + g(x^k) &\geq h(x^k) + g(y^k) + \langle w^k, y^k - x^k \rangle + \langle -\xi^k, y^k - x^k \rangle + \frac{\rho}{2} ||x^k - y^k||^2 - \varepsilon_k \\ &= h(x^k) + g(y^k) + \langle w^k - \xi^k, y^k - x^k \rangle + \frac{\rho}{2} ||y^k - x^k||^2 - \varepsilon_k, \end{aligned}$$

which implies that

$$-h(y^k) - g(x^k) \leq -h(x^k) - g(y^k) + \langle \xi^k - w^k, y^k - x^k \rangle - \frac{\rho}{2} ||y^k - x^k||^2 + \varepsilon_k$$

and hence,

$$g(y^k) - h(y^k) \leq g(x^k) - h(x^k) + \langle \xi^k - w^k, y^k - x^k \rangle - \frac{\rho}{2} ||y^k - x^k||^2 + \varepsilon_k. \tag{4.17}$$

The Cauchy-Schwarz inequality yields

$$\langle \xi^k - w^k, y^k - x^k \rangle \leq ||\xi^k - w^k|| \cdot ||y^k - x^k||.$$

Since $||w^k - \xi^k|| \leq \theta ||y^k - x^k||$, it follows that

$$\langle \xi^k - w^k, y^k - x^k \rangle \leq \theta ||y^k - x^k||^2. \tag{4.18}$$

By combining (4.17) with (4.18) and by using that $f(x) = g(x) - h(x)$, for all $x \in \mathbb{R}^n$, we obtain

$$\begin{aligned} f(y^k) &\leq f(x^k) + \theta||y^k - x^k||^2 - \frac{\rho}{2}||y^k - x^k||^2 + \varepsilon_k \\ &= f(x^k) - \left(\frac{\rho}{2} - \theta\right)||y^k - x^k||^2 + \varepsilon_k. \end{aligned}$$

We recall that $d^k = y^k - x^k$. Therefore, it holds that $f(y^k) \leq f(x^k) - \left(\frac{\rho}{2} - \theta\right)||d^k||^2 + \varepsilon_k$. $\square$

Note that when $\varepsilon_k = 0$ for each $k \in \mathbb{N}$, then Proposition 4.2.2 recovers Proposition 4.1.1.

Proposition 4.2.3. *Let $\{x^k\}_{k\in\mathbb{N}}$ be the sequence generated by **Algorithm 4**. Assume that $d^k \neq 0$ and $v_k > 0$ for each $k \in \mathbb{N}$. Then,*

$$\hat{\tau}_k := v_k / \big(g(y^k + d^k) + g(x^k) - 2g(y^k) + \varepsilon_k\big) > 0,$$

and

$$f(y^k + \lambda d^k) \leq f(y^k) - \alpha\lambda^2\|d^k\|^2 + v_k, \quad \forall \lambda \in (0, \tau_k],$$

where $\tau_k := \min\left\{1, \hat{\tau}_k, \frac{\rho}{\alpha}\right\}$. Consequently, the line search in Step 3 is well-defined.

Proof. Before starting the proof, we recall that $d^k = y^k - x^k$. Fix $s \in \partial h(y^k)$. Since h is strongly convex with modulus $\rho > 0$, then by Theorem 1.2.3, we have

$$h(y^k + \lambda d^k) \geq h(y^k) + \lambda\langle s, d^k\rangle + \frac{\rho}{2}\lambda^2||d^k||^2, \quad \forall \lambda \in \mathbb{R}. \tag{4.19}$$

In particular, for $\lambda = -1$, we obtain

$$h(x^k) = h(y^k - y^k + x^k) \geq h(y^k) + \langle -s, d^k\rangle + \frac{\rho}{2}||d^k||^2. \tag{4.20}$$

Moreover, since $w^k \in \partial_{\varepsilon_k} h(x^k)$, it follows that

$$h(y^k) \geq h(x^k) + \langle w^k, y^k - x^k\rangle - \varepsilon_k. \tag{4.21}$$

By adding both the inequalities (4.20) and (4.21) term by term, it holds

$$h(x^k) + h(y^k) \geq h(y^k) + h(x^k) + \langle -s, d^k\rangle + \frac{\rho}{2}||d^k||^2 + \langle w^k, y^k - x^k\rangle - \varepsilon_k,$$

which implies that $\langle s, d^k\rangle \geq \langle w^k, y^k - x^k\rangle + \frac{\rho}{2}||d^k||^2 - \varepsilon_k$. Then,

$$\lambda\langle s, d^k\rangle \geq \lambda\langle w^k, y^k - x^k\rangle + \frac{\rho}{2}\lambda||d^k||^2 - \lambda\varepsilon_k, \quad \forall \lambda \geq 0. \tag{4.22}$$

Combing the expressions at (4.19) and (4.22), we obtain

$$h(y^k + \lambda d^k) \geq h(y^k) + \lambda \langle w^k, y^k - x^k \rangle + \frac{\rho}{2}\lambda ||d^k||^2 - \lambda \varepsilon_k + \frac{\rho}{2}\lambda^2 ||d^k||^2, \quad \forall \lambda \geq 0. \quad (4.23)$$

Given that y^k satifies (4.14), i.e., $\langle w^k, y^k - x^k \rangle \geq g(y^k) - g(x^k)$, then (4.23) becomes

$$\begin{aligned} h(y^k + \lambda d^k) &\geq h(y^k) + \lambda \left(g(y^k) - g(x^k) \right) + \frac{\rho}{2}\lambda(1+\lambda)||d^k||^2 - \lambda \varepsilon_k \\ &= h(y^k) + \lambda \left(g(y^k) - g(x^k) - \varepsilon_k \right) + \frac{\rho}{2}\lambda(1+\lambda)||d^k||^2, \quad \forall \lambda \geq 0, \end{aligned}$$

which gives

$$-\left(h(y^k + \lambda d^k) - h(y^k) \right) \leq \lambda \left(g(x^k) - g(y^k) + \varepsilon_k \right) - \frac{\rho}{2}\lambda(1+\lambda)||d^k||^2, \quad (4.24)$$

for all $\lambda \geq 0$.

On the other hand, since g is strongly convex with modulus $\rho > 0$, using Definition 1.1.3, we have

$$\begin{aligned} g(y^k + \lambda d^k) - g(y^k) &= g\left(\lambda(y^k + d^k) + (1-\lambda)y^k \right) - g(y^k) \\ &\leq \lambda g(y^k + d^k) + (1-\lambda)g(y^k) - \frac{\rho}{2}\lambda(1-\lambda)||d^k||^2 - g(y^k) \\ &= \lambda \left(g(y^k + d^k) - g(y^k) \right) - \frac{\rho}{2}\lambda(1-\lambda)||d^k||^2, \quad (4.25) \end{aligned}$$

for all $\lambda \in [0,1]$. Note that

$$\begin{aligned} f(y^k + \lambda d^k) - f(y^k) &= \left(g(y^k + \lambda d^k) - h(y^k + \lambda^k) \right) - \left(g(y^k) - h(y^k) \right) \\ &= \left(g(y^k + \lambda d^k) - g(y^k) \right) - \left(h(y^k + \lambda d^k) - h(y^k) \right). \end{aligned}$$

Using this fact with (4.24) and (4.25), we obtain

$$\begin{aligned} f(y^k + \lambda d^k) - f(y^k) &\leq \lambda \left(g(y^k + d^k) + g(x^k) - 2g(y^k) + \varepsilon_k \right) - \frac{\rho}{2}\lambda(1+\lambda)||d^k||^2 \\ &\quad - \frac{\rho}{2}\lambda(1-\lambda)||d^k||^2 \\ &= \lambda \left(g(y^k + d^k) + g(x^k) - 2g(y^k) + \varepsilon_k \right) - \rho\lambda||d^k||^2, \quad (4.26) \end{aligned}$$

for all $\lambda \in [0,1]$.

Moreover, it follows from Theorem 1.2.3 that

$$g(y^k + d^k) \geq g(y^k) + \langle u, d^k \rangle + \frac{\rho}{2}||d^k||^2, \quad \text{and} \quad g(x^k) \geq g(y^k) + \langle u, -d^k \rangle + \frac{\rho}{2}||d^k||^2,$$

for all $u \in \partial g(y^k)$, where $d^k = y^k - x^k$. Adding either the inequalities above, it holds that $g(y^k+d^k)+g(x^k)-2g(y^k) \geq \rho||d^k||^2 > 0$, because $d^k \neq 0$. Hence, since $v_k > 0$ and $\varepsilon_k \geq 0$, we have that $\hat{\tau}_k := v_k/\big(g(y^k+d^k)+g(x^k)-2g(y^k)+\varepsilon_k\big) > 0$, as stated. Furthermore,

$$0 < \lambda \left(g(y^k+d^k)+g(x^k)-2g(y^k)+\varepsilon_k\right) \leq v_k, \quad \forall \lambda \in (0, \hat{\tau}_k]\,. \tag{4.27}$$

Also note that, considering $\lambda > 0$, we have $-\rho\lambda||d^k||^2 \leq -\alpha\lambda^2||d^k||^2$ if, only if, $\lambda \leq \dfrac{\rho}{\alpha}$. Thus, by combining this fact with (4.27) in (4.26), it follows that setting $\tau_k := \min\left\{1, \hat{\tau}_k, \frac{\rho}{\alpha}\right\}$, we have

$$f(y^k+\lambda d^k) \leq f(y^k) - \alpha\lambda^2||d^k||^2 + v_k, \quad \forall \lambda \in (0, \tau_k]\,.$$

Finally, given $\bar{\lambda} \geq 0$, due to $\beta \in (0,1)$, it follows that $\lim\limits_{j\to+\infty} \beta^j\bar{\lambda}_k = 0$. Hence, there is a sufficient large $j \in \mathbb{N}$ such that $\lambda_k := \beta^j\bar{\lambda}_k$ satisfies

$$f(y^k+\lambda_k d^k) \leq f(y^k) - \alpha\lambda_k^2||d^k||^2 + v_k,$$

which proves that the line search in Step 3 is well-defined. Moreover, setting $x^{k+1} := y^k + \lambda_k d^k$, we obtain the well-definition of Step 4. $\square$

Proposition 4.2.4. *Let $\left\{x^k\right\}_{k\in\mathbb{N}}$ be the sequence generated by **Algorithm 4**. Assume that $d^k \neq 0$ and $v_k > 0$ for each $k \in \mathbb{N}$. Then,*

$$f(x^{k+1}) \leq f(x^k) - \left(\frac{\rho}{2} - \theta + \alpha\lambda_k^2\right)||d^k||^2 + v_k + \varepsilon_k,$$

for each $k \in \mathbb{N}$. In particular,

$$\left(\frac{\rho}{2} - \theta\right)||d^k||^2 \leq f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k, \tag{4.28}$$

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

Proof. For each $k \in \mathbb{N}$, Proposition 4.2.2 (ii) guarantees that $f(y^k) \leq f(x^k) - \left(\frac{\rho}{2} - \theta\right)||d^k||^2 + \varepsilon_k$. Moreover, by the previous proposition, there is $\lambda_k > 0$ such that $f(y^k+\lambda_k d^k) \leq f(y^k) - \alpha\lambda_k^2\|d^k\|^2 + v_k$. Therefore, setting $x^{k+1} := y^k + \lambda_k d^k$, we obtain

$$\begin{aligned} f(x^{k+1}) &\leq f(y^k) - \alpha\lambda_k^2\|d^k\|^2 + v_k \\ &\leq f(x^k) - \left(\frac{\rho}{2} - \theta\right)||d^k||^2 - \alpha\lambda_k^2\|d^k\|^2 + v_k + \varepsilon_k \\ &= f(x^k) - \left(\frac{\rho}{2} - \theta + \alpha\lambda_k^2\right)||d^k||^2 + v_k + \varepsilon_k, \end{aligned}$$

for all $k \in \mathbb{N}$, which implies that

$$\left(\frac{\rho}{2} - \theta + \alpha\lambda_k^2\right) ||d^k||^2 \leq f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k, \quad \forall k \in \mathbb{N}.$$

Since $\alpha\lambda_k^2 \geq 0$ for all $k \in \mathbb{N}$, we have that $\left(\frac{\rho}{2} - \theta\right) ||d^k||^2 \leq \left(\frac{\rho}{2} - \theta + \alpha\lambda_k^2\right)$ for all $k \in \mathbb{N}$ and hence

$$\left(\frac{\rho}{2} - \theta\right) ||d^k||^2 \leq f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k, \quad \forall k \in \mathbb{N}.$$

□

4.2.1 Strategies to choose v_k

In this section, some strategies to choose the terms of the sequence $\{v_k\}_{k\in\mathbb{N}}$ will be introduced. The following strategies were taken from [21] with one adaptation in **(S1)**.

(S1) Given $\Delta_{min} \in [0, 1)$, the sequence $\{v_k\}_{k\in\mathbb{N}}$ is defined as follows: $v_0 \geq 0$ and v_{k+1}, for each $\Delta_{k+1} \in [\Delta_{min}, 1]$, satisfies the following condition:

$$0 \leq v_{k+1} \leq (1 - \Delta_{k+1}) \left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right), \quad \forall k \in \mathbb{N}; \tag{4.29}$$

(S2) $\{v_k\}_{k\in\mathbb{N}}$ is such that $\sum_{k=0}^{+\infty} v_k < +\infty$;

(S3) $\{v_k\}_{k\in\mathbb{N}}$ satisfies: for every $\delta > 0$, there exists $k_0 \in \mathbb{N}$ such that $v_k \leq \delta||d^k||^2$, for all $k \geq k_0$.

The strategy **(S3)** may seem strong at first glance. However, Example 4.2.2 provides other examples that satisfy **(S3)**. On the other hand, following [21], we also consider the following alternative form:

(S3') Fix any $\bar{\delta} \in (0, \frac{\rho}{2} - \theta)$. There exists $k_0 \in \mathbb{N}$ such that $v_k \leq \bar{\delta}\|d^k\|^2$, for all $k \geq k_0$.

Remark 4.2.5. *By Proposition* (4.2.4)*, we have* $0 \leq \left(\frac{\rho}{2} - \theta\right) ||d^k||^2 \leq f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k$*, for all* $k \in \mathbb{N}$*. Then, we can take* $v_{k+1} \geq 0$ *satisfying* (4.29)*. In particular, when* $\varepsilon_k = 0$*, for all* $k \in \mathbb{N}$*,* (4.29) *recovers the strategy defined in [21].*

Lemma 4.2.1. *If* $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$ *and* $\{v_k\}_{k\in\mathbb{N}}$ *satisfies* ***(S1)*** *with* $\Delta_{min} > 0$*, then* $\{v_k\}_{k\in\mathbb{N}}$ *satisfies* ***(S2)****.*

Proof. Note that (4.29) provides

$$0 \leq \Delta_{k+1}\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right) \leq \left(f(x^k) + v_k\right) - \left(f(x^{k+1}) + v_{k+1}\right) + \varepsilon_k, \quad \forall k \in \mathbb{N}.$$

Hence, $0 < \Delta_{min} \leq \Delta_{k+1}$ implies that

$$0 \leq \Delta_{min}\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right) \leq \left(f(x^k) + v_k\right) - \left(f(x^{k+1}) + v_{k+1}\right) + \varepsilon_k. \quad (4.30)$$

Considering the partial sum in the expressions of the last inequalities, it holds that

$$\begin{aligned} \Delta_{min}\sum_{k=0}^{N}\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right) &\leq \sum_{k=0}^{N}\left[\left(f(x^k) + v_k\right) - \left(f(x^{k+1}) + v_{k+1}\right) + \varepsilon_k\right] \\ &= f(x^0) + v_0 - f(x^{N+1}) - v_{N+1} + \sum_{k=0}^{N}\varepsilon_k \end{aligned}$$

Since $f^* = \inf_{x\in\mathbb{R}^n} f(x)$ and $v_k \geq 0$ for all $k \in \mathbb{N}$, then $f^* \leq f(x^{N+1})$ implies $-f^* \geq -f(x^{N+1})$, for all $N \in \mathbb{N}$, and $-v_{N+1} \leq 0$. Therefore,

$$\Delta_{min}\sum_{k=0}^{N}\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right) \leq f(x^0) + v_0 - f^* + \sum_{k=0}^{N}\varepsilon_k$$

Thus, since $\sum_{k=0}^{+\infty}\varepsilon_k < +\infty$ and $\Delta_{min} > 0$, we take the limit as $N \to +\infty$ to conclude that $\sum_{k=0}^{+\infty}\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right) < +\infty$.

On the other hand, (4.30) implies that

$$0 \leq v_{k+1} \leq (1 - \Delta_{min})\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right)$$

Since $\sum_{k=0}^{+\infty}\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right) < +\infty$, we conclude that

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

Therefore, $\{v_k\}_{k\in\mathbb{N}}$ satisfies **(S2)** and the claim is proved. □

Example 4.2.1. *Take any $v_0 > 0$, and define Δ_{k+1} and v_k as follows:*

$$0 < \Delta_{min} \leq \Delta_{k+1} < 1, \quad 0 < v_k := (1 - \Delta_{k+1})\left(\frac{\rho}{2} - \theta + \alpha\lambda_k^2\right)||d^k||^2, \quad \forall k \in \mathbb{N}. \quad (4.31)$$

Then, Proposition 4.2.4 yields $\left(\frac{\rho}{2} - \theta + \alpha\lambda_k^2\right)||d^k||^2 \leq f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k$. Thus, whenever $d^k \neq 0$, we have

$$0 < v_{k+1} \leq (1 - \Delta_{k+1})\left(f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k\right).$$

*Hence, $\{v_k\}_{k\in\mathbb{N}}$ defined in (4.31) satisfies **(S1)**. Therefore, considering that $\Delta_{min} > 0$, and $\sum_{k=0}^{+\infty}\varepsilon_k < +\infty$, we conclude from Lemma 4.2.1 that $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ also satisfies **(S2)**.*

Example 4.2.2. *[21, Example 4] Let* $\omega > 0$ *be a constant. Then, the sequence* $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *defined by*

$$v_k := \frac{\omega}{k+1}||d^k||^2,$$

for each $k \in \mathbb{N}$*, satisfies* ***(S3)****. Indeed, given* $\delta > 0$*, since* $\lim_{k\to+\infty} \frac{\omega}{k+1} = 0$*, then there exists* $k_0 \in \mathbb{N}$ *such that*

$$\frac{\omega}{k+1} \leq \delta,$$

for all $k \geq k_0$*, which implies that* $v_k \leq \delta||d^k||^2$*, for all* $k \geq k_0$*.*

More generally, if $\{u_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *is such that* $\lim_{k\to+\infty} u_k = +\infty$*, then* $v_k := \frac{\omega}{u_k}||d^k||^2$ *satisfies* ***(S3)****. Indeed, given* $\delta > 0$*, there exists* $k_0 \in \mathbb{N}$ *such that* $0 < \frac{\omega}{\delta} \leq u^k$*, for all* $k \geq k_0$*, i.e.,* $\frac{\omega}{u^k} \leq \delta$*, for all* $k \geq k_0$*, which implies that*

$$v_k = \frac{\omega}{u_k}||d^k||^2 \leq \delta||d^k||^2,$$

for all $k \geq k_0$*.*

4.2.2 Convergence Analysis

Note that in the convergence analysis of **Algorithm 1** and **Algorithm 2**, we used the fact that $\lim_{k\to+\infty} ||d^k|| = 0$, which arises naturally from the fact that $\left\{f(x^k)\right\}_{k\in\mathbb{N}}$ is convergent. The following proposition provides a sufficient condition to guarantee that each cluster point, if any, is a critical point.

Proposition 4.2.5. *If* $\lim_{k\to+\infty} ||d^k|| = 0$ *and* $\lim_{k\to+\infty} \varepsilon_k = 0$*, then each cluster point of* $\left\{x^k\right\}_{k\in\mathbb{N}}$*, if any, is critical.*

Proof. Let x^* be a cluster point of $\left\{x^k\right\}_{k\in\mathbb{N}}$. Then there exists a subsequence $\left\{x^{k_j}\right\}$ such that $\lim_{j\to+\infty} x^{k_j} = x^*$. As $d^k = y^k - x^k$ and $\lim_{k\to+\infty} ||d^k|| = 0$, due to

$$\begin{aligned} ||y^{k_j} - x^*|| &= ||y^{k_j} - x^{k_j} + x^{k_j} - x^*|| \\ &\leq ||y^{k_j} - x^{k_j}|| + ||x^{k_j} - x^*||, \end{aligned}$$

we have that $\lim_{j\to+\infty} ||y^{k_j} - x^*|| = 0$, i.e., $\lim_{j\to+\infty} y^{k_j} = x^*$.

Moreover, due to $\xi^{k_j} \in \partial g(y^{k_j})$, for all $j \in \mathbb{N}$, and Proposition 1.2.2, the sequence $\left\{\xi^{k_j}\right\}_{j\in\mathbb{N}}$ is bounded. Without loss of generality, we suppose it convergent and consider

$\xi^* = \lim_{j\to+\infty} \xi^{k_j}$. Note that $||\xi^k - w^k|| \leq \theta||y^k - x^k||$ implies $\lim_{j\to+\infty} ||\xi^{k_j} - w^{k_j}|| = 0$. As

$$\begin{aligned} ||w^{k_j} - \xi^*|| &= ||w^{k_j} - \xi^{k_j} + \xi^{k_j} - \xi^*|| \\ &\leq ||w^{k_j} - \xi^{k_j}|| + ||\xi^{k_j} - \xi^*||, \end{aligned}$$

we have that $\lim_{j\to+\infty} ||w^{k_j} - \xi^*|| = 0$, i.e., $\lim_{j\to+\infty} w^{k_j} = \xi^*$.

Fix an arbitrary $y \in \mathbb{R}^n$. Since $w^{k_j} \in \partial_{\varepsilon_{k_j}} h(x^{k_j})$, we have

$$h(y) \geq h(x^{k_j}) + \langle w^{k_j}, y - x^{k_j}\rangle - \varepsilon_{k_j}.$$

Taking the limit as $j \to +\infty$, and considering the continuity of h, it holds that $h(y) \geq h(x^*) + \langle \xi^*, y - x^*\rangle$. Due to the arbitrariness of $y \in \mathbb{R}^n$, we conclude that $\xi^* \in \partial h(x^*)$. Furthermore, as $\xi^{k_j} \in \partial g(y^{k_j})$, we obtain that $\xi^* \in \partial g(x^*)$. Therefore, x^* is a critical point. □

Setting $\varepsilon_k = 0$ for all $k \in \mathbb{N}$, the last proposition recovers Proposition 4.1.1.

Theorem 4.2.1. *Suppose that $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ satisfies **(S2)**, and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ satisfies $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$. Then, each cluster point of $\{x_k\}_{k\in\mathbb{N}}$, if any, is critical.*

Proof. Proposition 4.2.4 provides $f(x^{k+1}) \leq f(x^k) - \left(\frac{\rho}{2} - \theta + \alpha\lambda_k^2\right) ||d^k||^2 + v_k + \varepsilon_k$, for all $k \in \mathbb{N}$. Thus, since $\theta \in \left[0, \frac{\rho}{2}\right)$, we have that $\frac{\rho}{2} - \theta > 0$. Hence,

$$\begin{aligned} \left(\frac{\rho}{2} - \theta\right) ||d^k||^2 &\leq \left(\frac{\rho}{2} - \theta + \alpha{\lambda_k}^2\right) ||d^k||^2 \\ &\leq f(x^k) - f(x^{k+1}) + v^k + \varepsilon_k. \end{aligned}$$

Taking the N-th partial sum in the above inequality, we obtain

$$\begin{aligned} \sum_{k=0}^{N} \left(\frac{\rho}{2} - \theta\right) ||d^k||^2 &\leq \sum_{k=0}^{N} \left[f(x^k) - f(x^{k+1})\right] + \sum_{k=0}^{N} v_k + \sum_{k=0}^{N} \varepsilon_k \\ &= f(x^0) - f(x^{N+1}) + \sum_{k=0}^{N} v_k + \sum_{k=0}^{N} \varepsilon_k \\ &\leq f(x^0) - f^* + \sum_{k=0}^{N} v_k + \sum_{k=0}^{N} \varepsilon_k, \end{aligned}$$

where $f^* := \inf_{x\in\mathbb{R}^n} f(x) > -\infty$. Taking the limit as $N \to +\infty$, we have

$$\sum_{k=0}^{+\infty} \left(\frac{\rho}{2} - \theta\right) ||d^k||^2 \leq f(x^0) - f^* + \sum_{k=0}^{+\infty} v_k + \sum_{k=0}^{+\infty} \varepsilon_k < +\infty,$$

which proves that $\sum_{k=0}^{+\infty} ||d^k||^2 < +\infty$. In particular, $\lim_{k\to+\infty} ||d^k||^2 = 0$. Thus, by the previous proposition, the result follows. □

Theorem 4.2.2. *Suppose that* $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *satisfies* ***(S3)****, and* $\{\varepsilon_k\}_{k\in\mathbb{N}}$ *satisfies* $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$*. Then, each cluster point of* $\{x_k\}_{k\in\mathbb{N}}$*, if any, is critical.*

Proof. Set $\delta := \frac{1}{2}\left(\frac{\rho}{2} - \theta\right) > 0$. Then, for all $k \geq k_0$,

$$v_k \leq \delta||d^k||^2 = 2\delta||d^k||^2 - \delta||d^k||^2,$$

which implies that $\delta||d^k||^2 \leq 2\delta||d^k||^2 - v_k$. Therefore, by Proposition 4.2.4, we have that

$$\begin{aligned}\delta||d^k||^2 &\leq 2\delta||d^k||^2 - v_k \\ &= \left(\frac{\rho}{2} - \theta\right)||d^k||^2 - v_k \\ &\leq f(x^k) - f(x^{k+1}) + \varepsilon_k, \quad \forall k \geq k_0. \end{aligned} \tag{4.32}$$

Taking the partial sum in the expression above and considering $f^* := \inf_{x\in\mathbb{R}^n} f(x)$, we have that

$$\begin{aligned}\delta\sum_{k=0}^{N}||d^k||^2 &\leq \sum_{k=0}^{N}\left[f(x^k) - f(x^{k+1}) + \varepsilon_k\right] \\ &= f(x^0) - f(x^{N+1}) + \sum_{k=0}^{N}\varepsilon_k \\ &\leq f(x^0) - f^* + \sum_{k=0}^{N}\varepsilon_k.\end{aligned}$$

Taking the limit as $N \to +\infty$, since $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$, we obtain $\sum_{k=0}^{+\infty} ||d^k||^2 < +\infty$. In particular, $\lim_{k\to+\infty} ||d^k||^2 = 0$. □

Remark 4.2.6. *Note that Theorem 4.2.1 and Theorem 4.2.2 recover Theorem 4.1.2, when we consider* $\varepsilon_k = 0$ *for all* $k \in \mathbb{N}$*. Moreover, if there exists* $k_0 \in \mathbb{N}$ *such that* $\varepsilon_k \leq \delta||d^k||^2$ *for all* $k \geq k_0$ *in Theorem 4.2.2, then* (4.32) *implies*

$$f(x^{k+1}) \leq f(x^k),$$

for all $k \geq k_0$*. In this case, it is sufficient suppose that* $\lim_{k\to+\infty} \varepsilon_k = 0$ *to guarantee that* $\lim_{k\to+\infty} ||d^k|| = 0$.

Theorem 4.2.3. *If the sequence* $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *is chosen according to strategy* ***(S1)****, with* $\Delta_{min} > 0$*, and* $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$*, then every cluster point of* $\{x_k\}_{k\in\mathbb{N}}$*, if any, is a critical point of* f.

Proof. Since $\Delta_{min} > 0$ and $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$, Lemma 4.2.1 implies that $\{v_k\}_{k\in\mathbb{N}}$ satisfies **(S2)**. Then, by using Theorem 4.2.1, we obtain that every cluster point of $\{x_k\}_{k\in\mathbb{N}}$, if any, is a critical point of f. □

The last theorem recover item (iii) of Theorem 4.1.3, when we set $\varepsilon_k = 0$ for all $k \in \mathbb{N}$, and the next theorem recovers Theorem 4.1.3.

Theorem 4.2.4. *Suppose $\varepsilon_k = 0$ for all $k \in \mathbb{N}$. If the sequence $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is chosen according strategy **(S1)**, the following statements hold:*

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

(ii) If $\lim\limits_{k\to+\infty} v_k = 0$, then every cluster point of $\left\{x^k\right\}_{k\in\mathbb{N}}$, if any, is a critical point of f.

Proof. Proof of item (i): Setting $\varepsilon_k = 0$ for all $k \in \mathbb{N}$, then (4.30) in Lemma 4.2.1 becomes

$$0 \leq \Delta_{min}\left(f(x^k) - f(x^{k+1}) + v_k\right) \leq \left(f(x^k) + v_k\right) - \left(f(x^{k+1}) + v_{k+1}\right),$$

which implies that $f(x^k) + v_k \leq f(x^{k+1}) + v_{k+1}$ for all $k \in \mathbb{N}$, i.e., $\left\{f(x^k) + v_k\right\}_{k\in\mathbb{N}}$ is non-increasing. By using assumption**(H2)** and $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$, we obtain that either $\left\{f(x^k)\right\}$ and $\{v_k\}$ are lower bounded. Thus $\left\{f(x^k) + v_k\right\}_{k\in\mathbb{N}}$ is bounded and non-increasing, therefore it is convergent. Proof of item (ii): Since $\lim\limits_{k\in\mathbb{N}} v_k = 0$, from item (i) we have that $\left\{f(x^k)\right\}$ is convergent. On the other hand, since $\varepsilon_k = 0$ for all $k \in \mathbb{N}$, inequality (4.28) in Proposition 4.2.4 becomes

$$\left(\frac{\rho}{2} - \theta\right) \|d^k\|^2 \leq f(x^k) - f(x^{k+1}) + v_k,$$

for all $k \in \mathbb{N}$. Thus, taking the limit on the last inequality, we have that $\lim\limits_{k\to+\infty} \|d^k\|^2 = 0$. Therefore, by using Proposition 4.2.5 we complete the proof.

□

4.2.3 Iteration-complexity analysis

In this section, we present our results of iteration-complexity bounds for $\left\{x^k\right\}_{k\in\mathbb{N}}$ by **Algorithm 4**, which recover Theorem 4.1.4 and Theorem 4.1.5. We consider the cases whose the sequence $\{v_k\}_{k\in\mathbb{N}}$ is choosing according to **(S2)** and **(S3)**. The following results are based on Proposition 4.2.4 which implies, in particular,

$$\left(\frac{\rho}{2} - \theta\right) ||d^k||^2 \leq f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k, \tag{4.33}$$

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

Theorem 4.2.5. *Suppose that the sequence $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is chosen according to strategy **(S2)** and $\sum_{k=0}^{+\infty} \varepsilon < +\infty$. For each $N \in \mathbb{N}$, we have*

$$\min\left\{\|d^k\|:\ k=0,1,\cdots,N-1\right\} \leq \frac{\sqrt{f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k+\sum_{k=0}^{+\infty}\varepsilon_k}}{\sqrt{\frac{\rho}{2}-\theta}}\frac{1}{\sqrt{N}}.$$

Consequently, for a given accuracy $\varepsilon > 0$, if

$$N \geq \left(f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k+\sum_{k=0}^{+\infty}\varepsilon_k\right) / \left[\left(\frac{\rho}{2}-\theta\right)\varepsilon^2\right],$$

then the following inequality holds $\min\{\|d^k\|:\ k=0,1,\cdots,N-1\} \leq \varepsilon$.

Proof. By assumption **(H2)**, $f^* = \inf_{x\in\mathbb{R}^n} f(x) \leq f(x^k)$, for all $k \in \mathbb{N}$, from (4.33) we obtain that

$$\sum_{k=0}^{N-1}\|d^k\|^2 \leq \frac{1}{\frac{\rho}{2}-\theta}\Big(f(x^0)-f(x^N)+\sum_{k=0}^{N-1} v_k++\sum_{k=0}^{N-1}\varepsilon_k\Big) \leq \frac{1}{\frac{\rho}{2}-\theta}\Big(f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k++\sum_{k=0}^{+\infty}\varepsilon_k\Big).$$

and hence

$$N\cdot\min\{\|d^k\|^2:\ k=0,1,\cdots N-1\} \leq \frac{f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k++\sum_{k=0}^{+\infty}\varepsilon_k}{\frac{\rho}{2}-\theta}.$$

Thus,

$$\min\left\{\|d^k\|:\ k=0,1,\cdots,N-1\right\} \leq \frac{\sqrt{f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k+\sum_{k=0}^{+\infty}\varepsilon_k}}{\sqrt{\frac{\rho}{2}-\theta}}\frac{1}{\sqrt{N}}. \tag{4.34}$$

Moreover, given $\varepsilon > 0$, if

$$N \geq \frac{f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k+\sum_{k=0}^{+\infty}\varepsilon_k}{\left(\frac{\rho}{2}-\theta\right)\varepsilon^2},$$

then

$$f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k+\sum_{k=0}^{+\infty}\varepsilon_k \leq N\Big(\frac{\rho}{2}-\theta\Big)\varepsilon^2,$$

which combining with (4.34) provides

$$\begin{aligned}\min\left\{\|d^k\|:\ k=0,1,\cdots,N-1\right\} &\leq \frac{\sqrt{f(x^0)-f^*+\sum_{k=0}^{+\infty} v_k+\sum_{k=0}^{+\infty}\varepsilon_k}}{\sqrt{\frac{\rho}{2}-\theta}}\frac{1}{\sqrt{N}}\\ &\leq \frac{\sqrt{N\varepsilon^2\left(\frac{\rho}{2}-\theta\right)}}{\sqrt{\left(\frac{\rho}{2}-\theta\right)}}\frac{1}{\sqrt{N}}\\ &= \varepsilon,\end{aligned}$$

which concludes the proof. ☐

Theorem 4.2.6. *Suppose that the sequence $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ is chosen according to strategy* ***(S3)*** *and $\{\varepsilon_k\}_{k\in\mathbb{N}}$ is such that $\varepsilon_k \leq \xi\Big(\frac{\rho}{2}-\theta\Big)||d^k||^2$, for all $k \in \mathbb{N}$. Let $0 < \xi < 1/2$ and $k_0 \in \mathbb{N}$ such that $v_k \leq \xi\Big(\frac{\rho}{2}-\theta\Big)\|d^k\|^2$, for all $k \geq k_0$. Then, for each $N \in \mathbb{N}$ such that $N > k_0$, one has*

$$\min\{\|d^k\| : k = 0, 1, \cdots, N-1\} \leq \frac{\sqrt{f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k}}{\sqrt{(1-\xi)\Big(\frac{\rho}{2}-\theta\Big)}} \frac{1}{\sqrt{N}}.$$

Consequently, for a given $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ such that $v_k \leq \xi\Big(\frac{\rho}{2}-\theta\Big)\|d^k\|^2$ for all $k \geq k_0$, if

$$N \geq \max\left\{k_0, \frac{f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k}{\Big(\frac{\rho}{2}-\theta\Big)(1-\xi)\varepsilon^2}\right\},$$

then $\min\Big\{\|d^k\| : \ k = 0, 1, \cdots, N-1\Big\} \leq \varepsilon$.

Proof. Let $\xi \in (0, 1/2)$ and $k_0 \in \mathbb{N}$ such that $v_k \leq \xi\Big(\frac{\rho}{2}-\theta\Big)\|d^k\|^2$, for all $k \geq k_0$. It follows from (4.33) that $\Big(\frac{\rho}{2}-\theta\Big)\|d^k\|^2 \leq f(x^k) - f(x^{k+1}) + v_k + \varepsilon_k$, for all $k = 0, \cdots, N-1$. Summing up the last inequality from $k = 0$ to $K = N-1$ and using assumption **(H2)** we have that

$$\Big(\frac{\rho}{2}-\theta\Big)\sum_{k=0}^{N-1}\|d^k\|^2 \leq f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=k_0}^{N-1} v_k + \sum_{k=0}^{k_0-1}\varepsilon_k + \sum_{k=k_0}^{N-1}\varepsilon_k$$

Since $\varepsilon_k \leq \xi\Big(\frac{\rho}{2}-\theta\Big)||d^k||^2$ and $v_k \leq \xi\Big(\frac{\rho}{2}-\theta\Big)||d^k||^2$,for all $k \in \mathbb{N}$, the a=last inequality becomes

$$\begin{aligned}\sum_{k=0}^{N-1}\Big(\frac{\rho}{2}-\theta\Big)\|d^k\|^2 \leq & f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=k_0}^{N-1}\xi\Big(\frac{\rho}{2}-\theta\Big)||d^k||^2 + \sum_{k=0}^{k_0-1}\varepsilon_k \\ & + \sum_{k=k_0}^{N-1}\xi\Big(\frac{\rho}{2}-\theta\Big)||d^k||^2 \\ = & f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1}\varepsilon_k + 2\sum_{k=k_0}^{N-1}\xi\Big(\frac{\rho}{2}-\theta\Big)||d^k||^2,\end{aligned}$$

which implies that

$$\sum_{k=0}^{N-1}\Big(\frac{\rho}{2}-\theta\Big)\|d^k\|^2 \leq f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1}\varepsilon_k + 2\sum_{k=k_0}^{N-1}\xi\Big(\frac{\rho}{2}-\theta\Big)||d^k||^2,$$

And then

$$\sum_{k=0}^{N-1}(1-2\xi)||d^k||^2 \leq f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k.$$

Therefore, we have

$$N \cdot \min\left\{||d^k||^2 : \ k = 0, \dots, N-1\right\} \leq \frac{f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k}{(1-2\xi)\left(\frac{\rho}{2} - \theta\right)},$$

which implies that

$$\min\left\{||d^k|| : \ k = 0, \cdots, N-1\right\} \leq \frac{\sqrt{f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k}}{\sqrt{(1-2\xi)\left(\frac{\rho}{2} - \theta\right)}} \frac{1}{\sqrt{N}}, \tag{4.35}$$

and it proves the first inequality. Moreover, if

$$\max\left\{k_0, \frac{f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k}{\left(\frac{\rho}{2} - \theta\right)(1-2\xi)\varepsilon^2}\right\} \leq N,$$

then, in particular,

$$f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k \leq N\left(\frac{\rho}{2} - \theta\right)(1-2\xi)\varepsilon^2,$$

which combined with (4.35) provides

$$\begin{aligned}
\min\left\{||d^k|| : \ k = 0, \cdots, N-1\right\} &\leq \frac{\sqrt{f(x^0) - f^* + \sum_{k=0}^{k_0-1} v_k + \sum_{k=0}^{k_0-1} \varepsilon_k}}{\sqrt{(1-2\xi)\left(\frac{\rho}{2} - \theta\right)}} \frac{1}{\sqrt{N}} \\
&\leq \frac{\sqrt{N\left(\frac{\rho}{2} - \theta\right)(1-2\xi)\varepsilon^2}}{\sqrt{(1-2\xi)\left(\frac{\rho}{2} - \theta\right)N}} \\
&= \varepsilon,
\end{aligned}$$

thus we finish the demonstration □

Theorem 4.2.7. *Suppose that the sequences* $\{v_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *and* $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset \mathbb{R}_{++}$ *satisfy* $\lim\limits_{N\to\infty} \frac{\sum_{k=0}^{N-1} v_k}{N} = 0$ *and* $\lim\limits_{N\to\infty} \frac{\sum_{k=0}^{N-1} \varepsilon_k}{N} = 0$. *Then,* $\liminf\limits_{k\to+\infty} \|d^k\| = 0$.

Proof. Taking the partial sum in (4.33) and using assumption **(H2)**, we obtain

$$\begin{aligned}
\sum_{k=0}^{N-1}\left(\frac{\rho}{2}-\theta\right)||d^k||^2 &\leq \sum_{k=0}^{N-1}\left[f(x^k)-f(x^{k+1})\right]+\sum_{k=0}^{N-1}v_k+\sum_{k=0}^{N-1}\varepsilon_k \\
&= f(x^0)-f(x^N)+\sum_{k=0}^{N-1}v_k+\sum_{k=0}^{N-1}\varepsilon_k \\
&\leq f(x^0)-f^*+\sum_{k=0}^{N-1}v_k+\sum_{k=0}^{N-1}\varepsilon_k.
\end{aligned}$$

Therefore,

$$N\cdot\min\left\{||d^k||^2 :\ k=0,\dots,N-1\right\} \leq \frac{1}{\frac{\rho}{2}-\theta}\left(f(x^0)-f^*+\sum_{k=0}^{N-1}v_k+\sum_{k=0}^{N-1}\varepsilon_k\right),$$

which implies

$$\min\left\{||d^k||^2 :\ k=0,\dots,N-1\right\} \leq \frac{1}{\frac{\rho}{2}-\theta}\left(\frac{f(x^0)-f^*}{N}+\frac{\sum_{k=0}^{N-1}v_k}{N}+\frac{\sum_{k=0}^{N-1}\varepsilon_k}{N}\right),$$

and then

$$\min\left\{||d^k|| :\ k=0,\dots,N-1\right\} \leq \sqrt{\frac{1}{\frac{\rho}{2}-\theta}\left(\frac{f(x^0)-f^*}{N}+\frac{\sum_{k=0}^{N-1}v_k}{N}+\frac{\sum_{k=0}^{N-1}\varepsilon_k}{N}\right)}. \tag{4.36}$$

Taking the limit as $N\to+\infty$ in (4.36), we obtain that $\lim_{N\to\infty}\min\left\{||d^k|| :\ k=0,\dots,N-1\right\}$ $=0$. Thus, there exists a subsequence of $\{\|d^k\|\}_{k\in\mathbb{N}}$ that converges to 0 as $k\to\infty$. Therefore, since $||d^k||>0$ for all $k\in\mathbb{N}$, it follows that $\liminf_{k\to+\infty}\|d^k\|=0$. □

4.3 Numerical illustration

The numerical illustrations in this section were conducted using MATLAB software. The initial points were randomly chosen within the box $[-10,10]\times[-10,10]$. To solve the subproblems, we used the `fminsearch` toolbox with the inner stop rule: `optimset('TolX',1e-7,'TolFun',1e-7)`. The stopping criterion for the algorithm was $||x^{k+1}-x^k||<10^{-5}$. In the Example 4.3.1, the constants in the definition of **Algorithm 4** were set as $\alpha=0.6$, $\beta=0.1$, $\bar{\lambda}=1$ and $\theta=0.2$.

MATLAB solves Example 4.3.1 (see Example 3.3.1) inaccurately when computing the subproblem using the **fminsearch** toolbox. In this section, we verify computationally that the solution found by MATLAB satisfies the inequalities (4.10) and (4.11) in Figures 4.1 and 4.2.

Example 4.3.1. *(Example 3.3.1 revisited) Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *given by* $f(x, y) = x^2+y^2+x+y-|x|-|y|$*. We can obtain a DC decomposition of* f *as follows:* $f(x, y) = g(x, y)-h(x, y)$*, where* $g(x, y) = \frac{3}{2}(x^2 + y^2) + x + y$ *and* $h(x, y) = \frac{1}{2}(x^2 + y^2) + |x| + |y|$*.*

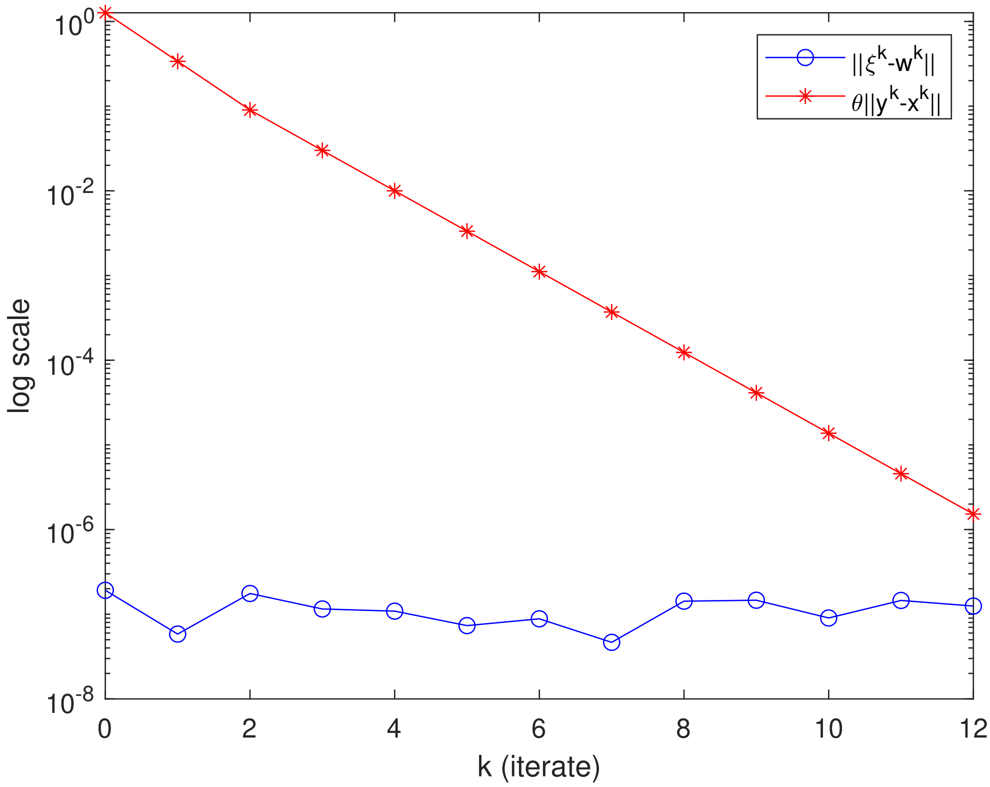


Figure 4.1: Example 4.3.1 starting from $x^0 = (6.2945, 8.1158)$.

In the Example 4.3.2, the constants in the definition of **Algorithm 4** were set as $\alpha = 0.6$, $\beta = 0.1$, $\bar{\lambda} = 1$ and $\theta = 0.2$. The sequence of parameters $\{v_k\}_{k\in\mathbb{N}}$ were chosen as $v_k = 0.01\frac{||d^k||^2}{k+1}$, for all $k \in \mathbb{N}$.

Example 4.3.2. *(Example 2.3.1 revisited) Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *given by* $f(x, y) = \frac{1}{2}(x^2+y^2)+|x|+|y|-\frac{5}{2}x$*. We can obtain a DC decomposition of* f *as follows:* $f(x, y) = g(x, y)-h(x, y)$*, where* $g(x, y) = x^2 + y^2 + |x| + |y| - \frac{5}{2}x$ *and* $h(x, y) = \frac{1}{2}(x^2 + y^2)$*. The minimum point of* f *is* $x_{opt} = (1.5, 0)$ *and the optimum value is* $f_{opt} = -1.125$*.*

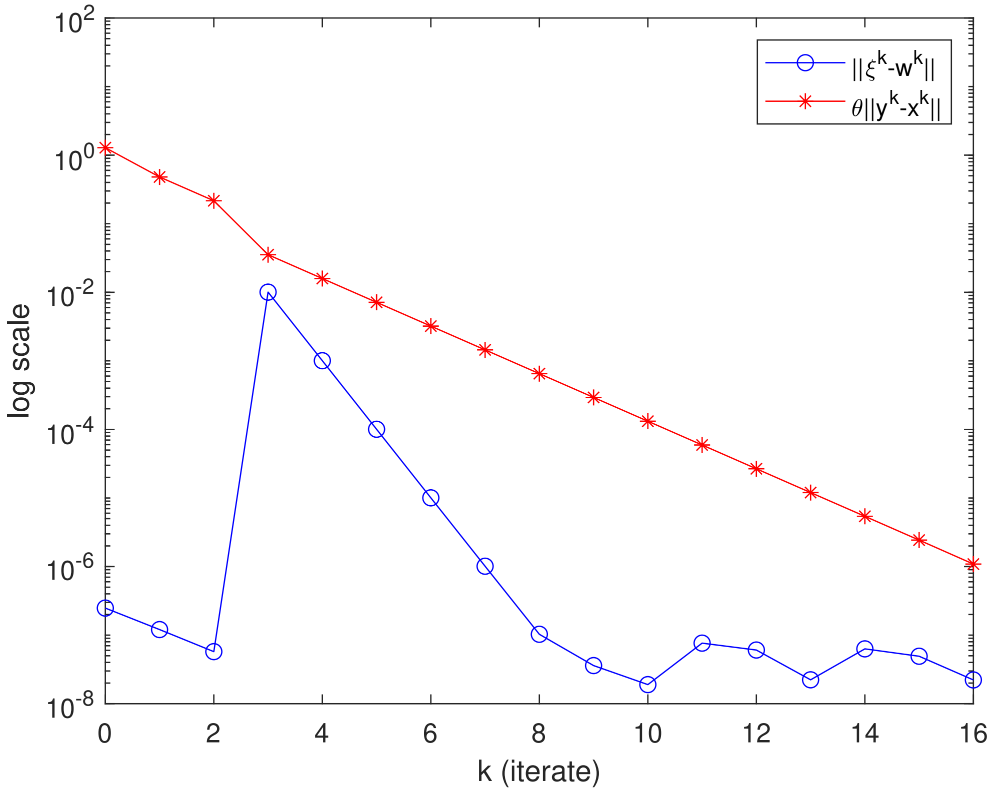


Figure 4.2: Example 4.3.2 starting from $x^0 = (-4.4615, -9.0766)$.

Conclusion

We have proposed an inexact version of BDCA and more general nmBDCA, where both the subgradient of the first component and the subproblem are computed inexactely. We have proposed a sufficient condition for any inexact direction to be a inexact direction as in BDCA. To this end we have studied the convergence analysis of DCA, BDCA and nmBDCA. We have shown that our inexact version have the same properties then its exact version. All the methods have been illustrated computationally.

Bibliography

[1] An, L. T. H., and Tao, P. D. Dc programming approach to the multidimensional scaling problem. *From Local to Global Optimization* (2001), 231–276.

[2] An, N. T., Nam, N. M., and Yen, N. D. A dc algorithm via convex analysis approach for solving a location problem involving sets. *arXiv preprint arXiv:1404.5113* (2014).

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

[4] Aragón Artacho, F. J., and Vuong, P. T. The boosted difference of convex functions algorithm for nonsmooth functions. *SIAM Journal on Optimization 30*, 1 (2020), 980–1006.

[5] Bačák, M., and Borwein, J. M. On difference convexity of locally lipschitz functions. *Optimization 60*, 8-9 (2011), 961–978.

[6] Bagirov, A., Karmitsa, N., and Mäkelä, M. M. *Introduction to Nonsmooth Optimization: theory, practice and software*, vol. 12. Springer, 2014.

[7] Bagirov, A. M., and Ugon, J. Codifferential method for minimizing nonsmooth dc functions. *Journal of Global Optimization 50* (2011), 3–22.

[8] Bagirov, A. M., and Ugon, J. Nonsmooth dc programming approach to clusterwise linear regression: optimality conditions and algorithms. *Optimization methods and software 33*, 1 (2018), 194–219.

[9] Beck, A. *First-order methods in optimization.* SIAM, 2017.

[10] Beck, A., and Hallak, N. On the convergence to stationary points of deterministic and randomized feasible descent directions methods. *SIAM Journal on Optimization 30*, 1 (2020), 56–79.

[11] Brimberg, J. The fermat—weber location problem revisited. *Mathematical programming 71* (1995), 71–76.

[12] Clarke, F. H. *Optimization and nonsmooth analysis.* SIAM, 1990.

[13] 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 68*, 7 (2019), 1305–1319.

[14] Cruz Neto, J. X., Lopes, J. O., Soubeyran, A., and Souza, J. C. O. Abstract regularized equilibria: application to becker's household behavior theory. *Ann Oper Res 316* (2022), 1279–1300.

[15] 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 289* (2020), 313–339.

[16] Cui, Y., He, Z., and Pang, J.-S. Multicomposite nonconvex optimization for training deep neural networks. *SIAM Journal on Optimization 30*, 2 (2020), 1693–1723.

[17] Cuong, T. H., Yao, J.-C., and Yen, N. D. Qualitative properties of the minimum sum-of-squares clustering problem. *Optimization 69*, 9 (2020), 2131–2154.

[18] de Oliveira, W. Proximal bundle methods for nonsmooth dc programming. *Journal of Global Optimization 75*, 2 (2019), 523–563.

[19] de Oliveira, W. The abc of dc programming. *Set-Valued and Variational Analysis 28* (2020), 679–706.

[20] De Oliveira, W., and Tcheou, M. P. An inertial algorithm for dc programming. *Set-Valued and Variational Analysis 27*, 4 (2019), 895–919.

[21] 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 88* (2024), 783–818.

[22] Gaudioso, M., Giallombardo, G., and Miglionico, G. Sparse optimization via vector k-norm and dc programming with an application to feature selection for support vector machines. *Computational Optimization and Applications 86*, 2 (2023), 745–766.

[23] 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 72*, 4 (2018), 705–729.

[24] Gotoh, J.-y., Takeda, A., and Tono, K. Dc formulations and algorithms for sparse optimization problems. *Mathematical Programming 169* (2018), 141–176.

[25] Hartman, P. On functions representable as a difference of convex functions.

[26] Hiriart-Urruty, J., and Lemarechal, C. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods.* Springer Berlin Heidelberg, 1993.

[27] Hiriart-Urruty, J.-B., and Lemaréchal, C. *Convex analysis and minimization algorithms I: Fundamentals*, vol. 305. Springer science & business media, 1993.

[28] Izmailov, A., and Solodov, M. *Otimização, volume 1: condições de otimalidade, elementos de análise convexa e de dualidade.* Impa, 2005.

[29] Joki, K., Bagirov, A. M., Karmitsa, N., and Mäkelä, M. M. A proximal bundle method for nonsmooth dc optimization utilizing nonconvex cutting planes. *Journal of Global Optimization 68*, 3 (2017), 501–535.

[30] Joki, K., Bagirov, A. M., Karmitsa, N., Makela, M. M., and Taheri, S. Double bundle method for finding clarke stationary points in nonsmooth dc programming. *SIAM Journal on Optimization 28*, 2 (2018), 1892–1919.

[31] Khamaru, K., and Wainwright, M. J. Convergence guarantees for a class of non-convex and non-smooth optimization problems. *Journal of Machine Learning Research 20*, 154 (2019), 1–52.

[32] Le Thi, H. A., and Pham Dinh, T. Dc programming and dca: thirty years of developments. *Mathematical Programming 169*, 1 (2018), 5–68.

[33] Liu, T., and Takeda, A. An inexact successive quadratic approximation method for a class of difference-of-convex optimization problems. *Computational Optimization and Applications 82*, 1 (2022), 141–173.

[34] 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 8*, 3 (2015), 1798–1823.

[35] Mordukhovich, B. S., and Nam, N. M. *An easy path to convex analysis and applications*. Springer, 2014.

[36] Moudafi, A., and Maingé, P.-E. On the convergence of an approximate proximal method for dc functions. *Journal of computational Mathematics* (2006), 475–480.

[37] Nakayama, S., Narushima, Y., and Yabe, H. Inexact proximal dc newton-type method for nonconvex composite functions. *Computational Optimization and Applications 87*, 2 (2024), 611–640.

[38] Nam, N. M., Geremew, W., Reynolds, S., and Tran, T. Nesterov's smoothing technique and minimizing differences of convex functions for hierarchical clustering. *Optimization Letters 12* (2018), 455–473.

[39] Niu, Y.-S. On the convergence analysis of dca. *arXiv preprint arXiv:2211.10942* (2022).

[40] Ordin, B., and Bagirov, A. M. A heuristic algorithm for solving the minimum sum-of-squares clustering problems. *Journal of Global Optimization 61*, 2 (2015), 341–361.

[41] Rockafellar, R. T., and Wets, R. J.-B. *Variational analysis*, vol. 317. Springer Science & Business Media, 2009.

[42] Souza, J. C. O., and Oliveira, P. R. A proximal point algorithm for dc fuctions on hadamard manifolds. *Journal of Global Optimization 63* (2015), 797–810.

[43] Souza, J. C. O., Oliveira, P. R., and Soubeyran, A. Global convergence of a proximal linearized algorithm for difference of convex functions. *Optimization Letters 10*, 7 (2016), 1529–1539.

[44] Sun, W.-y., Sampaio, R. J., and Candido, M. Proximal point algorithm for minimization of dc function. *Journal of computational Mathematics* (2003), 451–462.

[45] Tao, P. D., and An, L. H. Convex analysis approach to dc programming: theory, algorithms and applications. *Acta mathematica vietnamica 22*, 1 (1997), 289–355.

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

[47] Thormann, M.-L., Vuong, P. T., and Zemkoho, A. B. The boosted difference of convex functions algorithm for value-at-risk constrained portfolio optimization. *arXiv preprint arXiv:2402.09194* (2024).

[48] Tuy, H. *Convex Analysis and Global Optimization*, vol. 110. 01 2016.

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

[50] Zhang, Y., and Yamada, I. An inexact proximal linearized dc algorithm with provably terminating inner loop. *Optimization* (2024), 1–33.