




On the Rate of Convergence of the Difference-of-Convex Algorithm (DCA)

Hadi Abbaszadehpeivasti¹ · Etienne de Klerk¹  · Moslem Zamani¹

Received: 24 February 2022 / Accepted: 3 March 2023
© The Author(s) 2023

Abstract

In this paper, we study the non-asymptotic convergence rate of the DCA (difference-of-convex algorithm), also known as the convex–concave procedure, with two different termination criteria that are suitable for smooth and non-smooth decompositions, respectively. The DCA is a popular algorithm for difference-of-convex (DC) problems and known to converge to a stationary point of the objective under some assumptions. We derive a worst-case convergence rate of $O(1/\sqrt{N})$ after N iterations of the objective gradient norm for certain classes of DC problems, without assuming strong convexity in the DC decomposition and give an example which shows the convergence rate is exact. We also provide a new convergence rate of $O(1/N)$ for the DCA with the second termination criterion. Moreover, we derive a new linear convergence rate result for the DCA under the assumption of the Polyak–Łojasiewicz inequality. The novel aspect of our analysis is that it employs semidefinite programming performance estimation.

Keywords Convex–concave procedure · Difference-of-convex problems · Performance estimation · Worst-case convergence · Semidefinite programming

1 Introduction

In this paper, we consider the general difference-of-convex (DC) optimization problem,

Communicated by Tibor Illés.

✉ Etienne de Klerk
e.deklerk@tilburguniversity.edu
Hadi Abbaszadehpeivasti
h.abbaszadehpeivasti@tilburguniversity.edu
Moslem Zamani
m.zamani_1@tilburguniversity.edu

¹ Department of Econometrics and Operations Research, Tilburg University, Tilburg, The Netherlands

$$\begin{aligned} \inf f(x) &:= f_1(x) - f_2(x) \\ \text{s.t. } x &\in \mathbb{R}^n, \end{aligned} \quad (1)$$

where f_1, f_2 are extended convex functions on \mathbb{R}^n and f is an extended lower-semicontinuous function on \mathbb{R}^n . Throughout the paper, we assume that the infimum in problem (1) is finite, and denote by f^* a lower bound of f on \mathbb{R}^n .

DC problems appear naturally in many applications, e.g., power allocation in digital communication systems [4], production-transportation planning [22], location planning [13], image processing [31], sparse signal recovering [17], cluster analysis [7, 8], and supervised data classification [6, 29], to name but a few.

This wide range of applications is to be expected, since some important classes of non-convex functions may be represented as DC functions. For instance, twice continuously differentiable functions on any convex subset of \mathbb{R}^n [20] and continuous piece-wise linear functions [34] may be written as DC functions. Furthermore, every continuous function on a compact and convex set can be approximated by a DC function [23, 44]. We refer the interested reader to Hiriart-Urruty [21] and Tuy [44] for more information on DC representable functions.

The celebrated difference-of-convex algorithm (DCA), also known as the convex-concave procedure, has been applied extensively to problem (1); see [28, 30, 40] and the references therein. Algorithm 1 presents the basic form of the DCA.

Algorithm 1 DCA

Pick $x^1 \in \mathbb{R}^n$.

For $k = 1, 2, \dots$ perform the following steps:

1. Choose $g_2^k \in \partial f_2(x^k)$.
2. Choose

$$x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} f_1(x) - f_2(x^k) - \langle g_2^k, x - x^k \rangle. \quad (2)$$

3. If the termination criteria are satisfied, then stop.
-

In the description of the DCA in Algorithm 1, (sub)gradients of f_1 and f_2 are assumed to be available at given points, the so-called black-box formulation. The DCA is sometimes also presented as a primal-dual method, where a dual sub-problem is solved to obtain the required (sub)gradients; see [28, 30] for further discussions of this topic. In recent years, some scholars have also extended the DCA and proposed some new variations; see [19, 32, 33, 36, 39].

The first convergence results for Algorithm 1 were given in [40, Theorem 3(iv)]. The authors showed that, if the sequence of iterates $\{x^k\}$ is bounded, then each accumulation point of this sequence is a critical point of f .

Le Thi et al. [27] established an asymptotic linear convergence rate of $\{x^k\}$ under some conditions, in particular under the assumption that f satisfies the Łojasiewicz gradient inequality at all stationary points. Recall that a differentiable function f is said to satisfy this inequality at a stationary point a ($\nabla f(a) = 0$), if there exist constants

$\theta \in (0, 1)$, $C > 0$ and $\epsilon > 0$ such that

$$|f(x) - f(a)|^\theta \leq C \|\nabla f(x)\| \text{ if } \|x - a\| \leq \epsilon, \tag{3}$$

where the constant θ is called the Łojasiewicz exponent. This inequality is known to hold, for example, for real analytic functions, but has been extended to include classes of non-smooth functions as well by considering general sub-differentials instead of gradients; see [10, 11] and the references therein.

The convergence rates established by Le Thi et al. [27] depend on the value of the Łojasiewicz exponent, as the following theorem shows. The theorem stated here is a special case of Theorems 3.4 and 3.5 in [27], to give a flavor of the convergence results in [27].

Theorem 1.1 (Theorems 3.4 and 3.5 in Le Thi et al. [27]) *Let f_1 and f_2 be proper convex functions and let the domain of f be closed. Also assume that at least one of f_1 and f_2 is strongly convex, and f_1 or f_2 is differentiable with locally Lipschitz gradient in every critical point of the DC problem. Finally, assume the sequence $\{x^k\}$ is bounded, and let x^∞ be a limit point of $\{x^k\}$. Then x^∞ is also a stationary point. Moreover, if f satisfies the Łojasiewicz gradient inequality (3) at all stationary points, then*

1. if $\theta \in (1/2, 1)$, then $\|x^k - x^\infty\| \leq ck^{\frac{1-\theta}{1-2\theta}}$ for some $c > 0$.
2. if $\theta \in (0, 1/2)$, then $\|x^k - x^\infty\| \leq cq^k$ for some $c > 0$ and $q \in (0, 1)$.

In particular, item 2 shows a linear convergence rate when $\theta \in (0, 1/2)$. Yen et al. [45] had already shown linear convergence earlier for a much smaller class of DC functions. We will present a complementary result to this theorem (see Theorem 5.1), for the case $\theta = 1/2$, where we show linear convergence of the objective function values and give explicit expressions for the constants that determine the linear convergence rate. Moreover, we will relax the assumption of a bounded sequence of iterates, and the assumption of strong convexity.

In the absence of conditions like the Łojasiewicz gradient inequality (3), only weaker convergence rates are known for the DCA. In particular, Tao and An [40, Proposition 2] and Le Thi et al. [26, Corollary 1] have shown an $O\left(\frac{1}{\sqrt{N}}\right)$ convergence rate after N iterations under suitable assumptions, as given in the next theorem.

Theorem 1.2 (Corollary 1 in [26], Proposition 2 in [40]) *If x^∞ is a limit point of the iteration sequence generated by the DCA, and at least one of f_1 and f_2 is strongly convex, i.e. for some $\mu_1, \mu_2 \geq 0$ such that $\mu_1 + \mu_2 > 0$,*

$$x \mapsto f_i(x) - \frac{\mu_i}{2} \|x\|^2 \text{ is convex for } i \in \{1, 2\},$$

then the series $\|x^{k+1} - x^k\|$ converges, and, after $N + 1$ iterations,

$$\sum_{k=1}^N \|x^{k+1} - x^k\|^2 \leq \frac{2(f(x^1) - f(x^{N+1}))}{\mu_1 + \mu_2},$$

and, consequently,

$$\min_{1 \leq k \leq N} \|x^{k+1} - x^k\| \leq \sqrt{\frac{2(f(x^1) - f^*)}{(\mu_1 + \mu_2)N}} = O\left(\frac{1}{\sqrt{N}}\right).$$

We will derive some variants on this $O\left(\frac{1}{\sqrt{N}}\right)$ convergence result in Corollary 3.1 and in Sect. 3.2, where we improve the constants in the $O\left(\frac{1}{\sqrt{N}}\right)$ bounds. We also show that we obtain the best possible constants, by demonstrating an example where our bound in Corollary 3.1 is tight.

Outline and Further Contributions of this Paper

The novel aspect of the analysis in this paper is that we will apply performance estimation to derive convergence rates. Drori and Teboulle, in the seminal paper [16], introduced performance estimation as a strong tool for the worst-case analysis of first-order methods. The underlying idea of performance estimation is that the worst-case complexity may be cast as an optimization problem. Furthermore, this optimization problem can often be reformulated as a semidefinite programming problem. It is worth noting that performance estimation has been employed extensively for the analysis of worst-case convergence rates of first-order methods, see, e.g. [1, 14–16, 41, 42], and the references therein.

This paper is organized as follows: In Sect. 2, we review some definitions and notions from convex analysis, which will be used in the following sections. We study the DCA for sufficiently smooth DC decompositions in Sect. 3. By using performance estimation, we give a convergence rate of $O(1/\sqrt{N})$ in Corollary 3.1, without any strong convexity assumption, thus extending and complementing Le Thi et al. [26, Corollary 1]. We construct an example that shows this $O(1/\sqrt{N})$ bound is tight. Since the first termination criterion is not suitable for the analysis of non-smooth DC compositions, we investigate the DCA with another stopping criterion in Sect. 4, and we show a convergence rate of $O(1/N)$. This result is completely new to the best of our knowledge.

In Sect. 5, we study the DCA when the objective function satisfies the Polyak–Łojasiewicz inequality, and we derive a linear convergence rate in Theorem 5.1, thereby refining some linear convergence results in Le Thi et al. [27] as described above.

2 Basic Definitions and Preliminaries

In this section, we recall some notions and definitions from convex analysis. Throughout the paper, $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the Euclidean norm and the dot product, respectively. $I_{\mathbb{R}_+}$ stands for the indicator function on $\mathbb{R}_+ \cup \{\infty\}$, i.e.,

$$I_{\mathbb{R}_+}(x) = \begin{cases} 1 & x \geq 0 \cup \{\infty\} \\ 0 & x < 0 \cup \{-\infty\}. \end{cases}$$

Let $f : \mathbb{R}^n \rightarrow [-\infty, \infty]$ be an extended convex function. The domain of f is denoted and defined as $\text{dom}(f) := \{x : f(x) < \infty\}$. The function f is called proper if it does not attain the value $-\infty$, and its domain is non-empty. We call f closed if its epi-graph is closed, that is $\{(x, r) : f(x) \leq r\}$ is a closed subset of \mathbb{R}^{n+1} . We denote the convex hull of $X \subseteq \mathbb{R}^n$ by $\text{co}(X)$. We adopt the conventions that, for $a, b, c, d \in \mathbb{R}$ with $c \neq d$ and $a \neq 0$, $\frac{b}{\infty} = 0$, $0 \times \infty = 0$ and $\frac{a\infty+b}{c\infty-d\infty} = \frac{a}{c-d}$. For the function $f : \mathbb{R}^n \rightarrow [-\infty, \infty]$, the conjugate function $f^* : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as $f^*(g) = \max_{x \in \mathbb{R}^n} \langle g, x \rangle - f(x)$. Moreover, we denote the set of subgradients of f at $x \in \text{dom}(f)$ by $\partial f(x)$,

$$\partial f(x) = \{g : f(y) \geq f(x) + \langle g, y - x \rangle, \forall y \in \mathbb{R}^n\}.$$

Let $L \in (0, \infty]$ and $\mu \in (0, \infty)$. We call an extended convex function $f : \mathbb{R}^n \rightarrow [-\infty, \infty]$ L -smooth if for any $x_1, x_2 \in \mathbb{R}^n$,

$$\|g_1 - g_2\| \leq L\|x_1 - x_2\| \quad \forall g_1 \in \partial f(x_1), g_2 \in \partial f(x_2).$$

Note that if $L < \infty$, then f must be differentiable on \mathbb{R}^n . In addition, any extended convex function is ∞ -smooth. Also recall that the function $f : \mathbb{R}^n \rightarrow [-\infty, \infty]$ is called μ -strongly convex function if the function $x \mapsto f(x) - \frac{\mu}{2}\|x\|^2$ is convex. Clearly, any convex function is 0-strongly convex. We denote the set of closed proper convex functions which are L -smooth and μ -strongly convex by $\mathcal{F}_{\mu,L}(\mathbb{R}^n)$.

Let \mathcal{I} be a finite index set and let $\{x^i; g^i; f^i\}_{i \in \mathcal{I}} \subseteq \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$. A set $\{x^i; g^i; f^i\}_{i \in \mathcal{I}}$ is called $\mathcal{F}_{\mu,L}$ -interpolable if there exists $f \in \mathcal{F}_{\mu,L}(\mathbb{R}^n)$ with

$$f(x^i) = f^i, g^i \in \partial f(x^i) \quad i \in \mathcal{I}.$$

The next theorem gives necessary and sufficient conditions for $\mathcal{F}_{\mu,L}$ -interpolability.

Theorem 2.1 [41, Theorem 4] *Let $L \in (0, \infty]$ and $\mu \in [0, \infty)$ and let \mathcal{I} be a finite index set. The set $\{(x^i; g^i; f^i)\}_{i \in \mathcal{I}} \subseteq \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ is $\mathcal{F}_{\mu,L}$ -interpolable if and only if for any $i, j \in \mathcal{I}$, we have*

$$\begin{aligned} & \frac{1}{2(1-\frac{\mu}{L})} \left(\frac{1}{L} \|g^i - g^j\|^2 + \mu \|x^i - x^j\|^2 - \frac{2\mu}{L} \langle g^j - g^i, x^j - x^i \rangle \right) \\ & \leq f^i - f^j - \langle g^j, x^i - x^j \rangle. \end{aligned}$$

In the next lemma, we extend the descent lemma for DCA when L_1 or L_2 is finite.

Lemma 2.1 *Let $f_1 \in \mathcal{F}_{\mu_1,L_1}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}(\mathbb{R}^n)$ and let $f = f_1 - f_2$. If $g_1 \in \partial f_1(x)$ and $g_2 \in \partial f_2(x)$, then*

$$f^* \leq f(x) - \frac{1}{2(L_1-\mu_2)} \|g_1 - g_2\|^2.$$

Proof If $L_1 = \infty$, the proof is immediate. Let $L_1 < \infty$. By L -smoothness and strong convexity, we have

$$\begin{aligned} f_1(y) &\leq f_1(x) + \langle g_1, y - x \rangle + \frac{L_1}{2} \|y - x\|^2, \\ f_2(y) &\geq f_2(x) + \langle g_2, y - x \rangle + \frac{\mu_2}{2} \|y - x\|^2, \end{aligned}$$

for $y \in \mathbb{R}^n$. By the above inequalities, we get

$$f(y) \leq f(x) + \langle g_1 - g_2, y - x \rangle + \frac{L_1 - \mu_2}{2} \|y - x\|^2.$$

Hence, by taking minimum on both sides of the last inequality with respect to y for fixed x , we get

$$f^* \leq f(x) - \frac{1}{2(L_1 - \mu_2)} \|g_1 - g_2\|^2.$$

Since the DC optimization problem (1) may have a non-convex and non-smooth objective function f , we will also need a more general notion of subgradients than in the convex case.

Definition 2.1 Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be lower semi-continuous and let $f(\bar{x})$ be finite.

- The vector g is called regular subgradient of f at \bar{x} , written $g \in \hat{\partial}_l f(\bar{x})$, if for all x in some neighborhood of \bar{x}

$$f(x) \geq f(\bar{x}) + \langle g, x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$

- The vector g is called general subgradient of f at \bar{x} , written $g \in \partial_l f(\bar{x})$, if there exist sequences $\{x^i\}$ and $\{g^i\}$ with $g^i \in \hat{\partial}_l f(x^i)$ such that

$$x^i \rightarrow \bar{x}, f(x^i) \rightarrow f(\bar{x}), g^i \rightarrow g.$$

It is worth mentioning that $\hat{\partial}_l f(\bar{x})$ is a closed convex set. In addition, $\partial_l f(\bar{x})$ is also closed but not necessarily convex. Note that when f is closed proper convex, then $\partial f(x) = \hat{\partial}_l f(x) = \partial_l f(x)$ for $x \in \text{dom}(f)$. We refer the interested reader to Rockafellar and Wets [38] for more discussions on regular and general subdifferentials.

Definition 2.2 Let f_1, f_2 be closed proper convex functions, and let f be lower semi-continuous.

- The point $\bar{x} \in \text{dom}(f)$ is called a critical point of problem (1) if

$$\partial f_1(\bar{x}) \cap \partial f_2(\bar{x}) \neq \emptyset. \quad (4)$$

- The point $\bar{x} \in \text{dom}(f)$ is called a stationary point of problem (1) if

$$0 \in \partial_l f(\bar{x}). \quad (5)$$

Obviously, the stationarity condition is stronger than criticality. We recall that a convex function will be locally Lipschitz around \bar{x} providing it takes finite values in a neighborhood of \bar{x} ; see Theorem 35.1 in [37]. Consequently, if f_1 or f_2 takes finite values around a neighborhood of a stationary point \bar{x} , then \bar{x} is a critical point; see Corollary 10.9 in [38]. However, its converse does not hold in general. For instance, consider $f : \mathbb{R} \rightarrow \mathbb{R}$ given as $f(x) = x$. The function f may be written as $f = f_1 - f_2$ where $f_1(x) = \max(x, 0)$ and $f_2(x) = \max(-x, 0)$. Suppose that $\bar{x} = 0$. It is readily seen that $\partial f_1(\bar{x}) \cap \partial f_2(\bar{x}) \neq \emptyset$, but $\bar{x} = 0$ is not a stationary point of f . It is worth noting that, if f_2 is strictly differentiable at \bar{x} , these definitions are equivalent; see Example 10.10 in [38]. Recall that function f is strictly differentiable at \bar{x} , if

$$\lim_{\substack{(x,x') \rightarrow (\bar{x},\bar{x}) \\ x \neq x'}} \frac{f(x) - f(x') - \langle \nabla f(\bar{x}), x - x' \rangle}{\|x - x'\|} = 0.$$

We refer the interested reader to An and Tao [5], Joki et al. [24] and Pang et al. [36] and references therein for more discussions on optimality conditions for DC problems.

2.1 The DC Problem

In this section, we consider

$$\begin{aligned} \min f(x) &= f_1(x) - f_2(x) \\ \text{s.t. } x &\in \mathbb{R}^n, \end{aligned} \tag{6}$$

where $f_1 \in \mathcal{F}_{\mu_1, L_1}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2, L_2}(\mathbb{R}^n)$. Here, we assume that $L_1, L_2 \in (0, \infty]$ and $\mu_1, \mu_2 \in [0, \infty)$, and consequently, f may be non-differentiable. We may assume without loss of generality that f_1 and f_2 satisfy the following assumptions:

$$L_1 > \mu_2, \quad L_2 > \mu_1. \tag{7}$$

Indeed, if $L_1 \leq \mu_2$, then for $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$, we have

$$\begin{aligned} \lambda f_1(x) + (1 - \lambda) f_1(y) &\leq f_1(\lambda x + (1 - \lambda)y) + \lambda(1 - \lambda) \frac{L_1}{2} \|x - y\|^2 \\ - \lambda f_2(x) - (1 - \lambda) f_2(y) &\leq -f_2(\lambda x + (1 - \lambda)y) - \lambda(1 - \lambda) \frac{\mu_2}{2} \|x - y\|^2; \end{aligned}$$

see Theorem 2.15 and Theorem 2.19 in [35]. By summing the above inequalities, we obtain

$$\lambda f(x) + (1 - \lambda) f(y) \leq f(\lambda x + (1 - \lambda)y) + \lambda(1 - \lambda) \frac{L_1 - \mu_2}{2} \|x - y\|^2,$$

which implies concavity of f on \mathbb{R}^n . In this case, problem (6) will be unbounded from below. This follows from the fact that a concave function on \mathbb{R}^n is unbounded from below unless it is constant. Likewise, one can show that problem (6) will be convex providing $L_2 \leq \mu_1$.

The Toland dual [43] of problem (6) may be written as

$$\begin{aligned} \min \quad & f_2^*(x) - f_1^*(x) \\ \text{s.t.} \quad & x \in \mathbb{R}^n. \end{aligned} \tag{8}$$

It is known that problems (6) and (8) share the same optimal value [43].

In what follows, we investigate the convergence rate of Algorithm 1 with the termination criterion $\|g_1^k - g_2^k\| \leq \epsilon$. As a motivation of this criterion, recall that $\|g_1^k - g_2^k\| = 0$ implies that x^k is a critical point of (1) in the non-smooth case, and a stationary point of f if f_2 is strictly differentiable; see our discussion following Definition 2.2. In Sect. 3, we will derive results for the case that at least one of f_1 or f_2 is differentiable, and we will consider the more general situation in Sect. 4.

For well-definedness of the DCA (Algorithm 1), throughout the paper, we assume that

$$x^k \in \text{dom}(\partial f_1) \cap \text{dom}(\partial f_2) \quad k = 1, 2, \dots,$$

where $\text{dom}(\partial f_1) = \{x : \partial f_1(x) \neq \emptyset\}$. It is worth noting that similar algorithm has been developed for the dual problem in [28] and (2) is equivalent to $x^{k+1} \in \partial f_1^*(g_2^k)$.

3 Performance Analysis of the DCA for Smooth f_1 or f_2

In this subsection, we apply performance estimation for the analysis of Algorithm 1 for the case that at least one of f_1 or f_2 is L -smooth for some finite $L > 0$. The worst-case convergence rate of Algorithm 1 can be obtained by solving the following abstract optimization problem:

$$\begin{aligned} \max \quad & \left(\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\|^2 \right) \\ & g_1^{N+1}, g_2^{N+1}, x^{N+1}, \dots, x^2 \text{ are generated by Algorithm 1 w.r.t. } f_1, f_2, x^1 \\ & f(x) \geq f^* \quad \forall x \in \mathbb{R}^n \\ & f_1 \in \mathcal{F}_{\mu_1, L_1}(\mathbb{R}^n), f_2 \in \mathcal{F}_{\mu_2, L_2}(\mathbb{R}^n) \\ & f_1(x^1) - f_2(x^1) - f^* \leq \Delta \\ & x^1 \in \mathbb{R}^n, \end{aligned} \tag{9}$$

where $\Delta \geq 0$ denote the difference between the optimal value and the value of f at the starting point. Here, f_1, f_2 and x^k, g_1^k and g_2^k ($k \in \{1, \dots, N + 1\}$) are decision variables, and $\Delta, \mu_1, L_1, \mu_2, L_2$ and N are fixed parameters.

Problem (9) is an intractable infinite-dimensional optimization problem with an infinite number of constraints. In what follows, we provide a semidefinite programming relaxation of the problem.

By Theorem 2.1, problem (9) can be written as,

$$\begin{aligned}
 & \max \left(\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\|^2 \right) \\
 & \text{s.t. } \frac{1}{2(1-\frac{\mu_1}{L_1})} \left(\frac{1}{L_1} \|g_1^i - g_1^j\|^2 + \mu_1 \|x^i - x^j\|^2 - \frac{2\mu_1}{L_1} \langle g_1^j - g_1^i, x^j - x^i \rangle \right) \\
 & \quad \leq f_1^i - f_1^j - \langle g_1^j, x^i - x^j \rangle \quad i, j \in \{1, \dots, N+1\} \\
 & \frac{1}{2(1-\frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_2^i - g_2^j\|^2 + \mu_2 \|x^i - x^j\|^2 - \frac{2\mu_2}{L_2} \langle g_2^j - g_2^i, x^j - x^i \rangle \right) \\
 & \quad \leq f_2^i - f_2^j - \langle g_2^j, x^i - x^j \rangle \quad i, j \in \{1, \dots, N+1\} \\
 & g_1^{k+1} = g_2^k \quad k \in \{1, \dots, N\} \\
 & f_1^k - f_2^k - \frac{1}{2(L_1 - \mu_2)} \|g_1^k - g_2^k\|^2 \geq f^* \quad k \in \{1, \dots, N+1\} \\
 & f_1^1 - f_2^1 - f^* \leq \Delta. \tag{10}
 \end{aligned}$$

In problem (10), f^* and $x^k, g_1^k, g_2^k, f_1^k, f_2^k, k \in \{1, \dots, N+1\}$, are decision variables. By virtue of Lemma 2.1, constraints $f(x) \geq f^*$ for each $x \in \mathbb{R}^n$ are replaced by $f_1^k - f_2^k - \frac{1}{2(L_1 - \mu_2)} \|g_1^k - g_2^k\|^2 \geq f^*, k \in \{1, \dots, N+1\}$. Due to the necessary and sufficient optimality conditions for convex problems, $x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} f_1(x) - f_2(x^k) - \langle g_2^k, x - x^k \rangle, k \in \{1, \dots, N\}$ implies $g_1^{k+1} = g_2^k$ for some $g_1^{k+1} \in \partial f(x^{k+1})$; see Theorem 3.63 in [9]. By substituting $g_2^k = g_1^{k+1}, k \in \{1, \dots, N\}$, the above formulation may be written as:

$$\begin{aligned}
 & \max \ell \\
 & \text{s.t. } \|g_1^i - g_1^{i+1}\|^2 \geq \ell \quad i \in \{1, \dots, N\} \\
 & \|g_1^{N+1} - g_2^{N+1}\|^2 \geq \ell \\
 & \frac{1}{2(1-\frac{\mu_1}{L_1})} \left(\frac{1}{L_1} \|g_1^i - g_1^j\|^2 + \mu_1 \|x^i - x^j\|^2 - \frac{2\mu_1}{L_1} \langle g_1^j - g_1^i, x^j - x^i \rangle \right) \\
 & \quad \leq f_1^i - f_1^j - \langle g_1^j, x^i - x^j \rangle \quad i, j \in \{1, \dots, N+1\} \\
 & \frac{1}{2(1-\frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{i+1} - g_1^{j+1}\|^2 + \mu_2 \|x^i - x^j\|^2 - \frac{2\mu_2}{L_2} \langle g_1^{j+1} - g_1^{i+1}, x^j - x^i \rangle \right) \\
 & \quad \leq f_2^i - f_2^j - \langle g_1^{j+1}, x^i - x^j \rangle \quad i, j \in \{1, \dots, N\} \\
 & \frac{1}{2(1-\frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_2^{N+1} - g_1^{j+1}\|^2 + \mu_2 \|x^{N+1} - x^j\|^2 - \frac{2\mu_2}{L_2} \langle g_1^{j+1} - g_2^{N+1}, x^j - x^{N+1} \rangle \right) \\
 & \quad \leq f_2^{N+1} - f_2^j - \langle g_1^{j+1}, x^{N+1} - x^j \rangle \quad j \in \{1, \dots, N\} \\
 & \frac{1}{2(1-\frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{i+1} - g_2^{N+1}\|^2 + \mu_2 \|x^i - x^{N+1}\|^2 - \frac{2\mu_2}{L_2} \langle g_2^{N+1} - g_1^{i+1}, x^{N+1} - x^i \rangle \right) \\
 & \quad \leq f_2^i - f_2^{N+1} - \langle g_2^{N+1}, x^i - x^{N+1} \rangle \quad i \in \{1, \dots, N\}
 \end{aligned}$$

$$\begin{aligned}
 f_1^k - f_2^k - \frac{1}{2(L_1 - \mu_2)} \|g_1^k - g_1^{k+1}\|^2 &\geq f^* \quad k \in \{1, \dots, N\} \\
 f_1^{N+1} - f_2^{N+1} - \frac{1}{2(L_1 - \mu_2)} \|g_1^{N+1} - g_2^{N+1}\|^2 &\geq f^* \\
 f_1^1 - f_2^1 - f^* &\leq \Delta.
 \end{aligned}
 \tag{11}$$

By using this formulation, the next result (Theorem 3.1) provides a convergence rate for Algorithm 1. Since the proof is quite technical, a few remarks are in order. The proof uses the performance estimation technique of Drori and Teboulle [16] that consists of the following steps:

1. Observe that problem (11) may be rewritten as a semidefinite programming (SDP) problem (for sufficiently large N) by replacing all inner products by the entries of an unknown Gram matrix.
2. Use weak duality of SDP to bound the optimal value of (11) by constructing a dual feasible solution.
3. The dual feasible solution is constructed empirically, by first doing numerical experiments with fixed values of the parameters $\Delta, N, \mu_1, L_1, \mu_2, L_2$, and noting the dual multipliers.
4. Subsequently, the analytical expressions of the dual multipliers are guessed, based on the numerical values, and the guess is verified analytically.
5. In the proof of Theorem 3.1, the conjectured dual multipliers are simply stated and then shown to provide the required bound on the optimal value of (11) through the corresponding aggregation of the constraints of (11).

Theorem 3.1 *Let $f_1 \in \mathcal{F}_{\mu_1, L_1}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2, L_2}(\mathbb{R}^n)$ and let $f(x^1) - f^* = \Delta$. Suppose that L_1 or L_2 is finite. Then after N iterations of Algorithm 1, one has:*

$$\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\| \leq \sqrt{\frac{\mathcal{A}\Delta}{\mathcal{B}N + \mathcal{C}}},
 \tag{12}$$

where

$$\begin{aligned}
 \mathcal{A} &= 2(L_1 L_2 - \mu_1 L_2 I_{\mathbb{R}_+}(L_1 - L_2) - \mu_2 L_1 I_{\mathbb{R}_+}(L_2 - L_1)), \\
 \mathcal{B} &= L_1 + L_2 + \mu_1 \left(\frac{L_1}{L_2} - 3\right) I_{\mathbb{R}_+}(L_1 - L_2) + \mu_2 \left(\frac{L_2}{L_1} - 3\right) I_{\mathbb{R}_+}(L_2 - L_1),
 \end{aligned}$$

and

$$\mathcal{C} = \frac{L_1 L_2 - \mu_1 L_2 I_{\mathbb{R}_+}(L_1 - L_2) - \mu_2 L_1 I_{\mathbb{R}_+}(L_2 - L_1)}{L_1 - \mu_2}.$$

Proof We investigate two cases $L_1 \geq L_2$ and $L_1 < L_2$. Suppose that U denote the square of the right side of inequality (12) and let $B = \frac{U}{\Delta}$. To prove this bound, we show that U is an upper bound for problem (11). First, we consider $L_1 \geq L_2$. Let

$$\bar{\lambda} = \frac{2(L_1 L_2 - \mu_1(2L_2 - L_1))}{N \left(L_1 + L_2 + \mu_1 \left(\frac{L_1}{L_2} - 3 \right) \right) + \frac{L_2(L_1 - \mu_1)}{L_1 - \mu_2}}$$

$$\begin{aligned} \bar{\eta}_1 &= \frac{L_2 - \mu_1}{\left(L_1 + L_2 + \mu_1\left(\frac{L_1}{L_2} - 3\right)\right)N + \frac{L_2(L_1 - \mu_1)}{L_1 - \mu_2}} \\ \bar{\eta}_k &= \frac{\frac{L_1\mu_1}{L_2} + (L_1 + L_2 - 3\mu_1)}{\left(L_1 + L_2 + \mu_1\left(\frac{L_1}{L_2} - 3\right)\right)N + \frac{L_2(L_1 - \mu_1)}{L_1 - \mu_2}}, \quad k \in \{2, \dots, N\} \\ \bar{\eta}_{N+1} &= 1 - \bar{\eta}_1 - \sum_{k=2}^N \bar{\eta}_k = \frac{\frac{L_1\mu_1}{L_2} + L_1 - 2\mu_1 + \frac{L_2(L_1 - \mu_1)}{L_1 - \mu_2}}{\left(L_1 + L_2 + \mu_1\left(\frac{L_1}{L_2} - 3\right)\right)N + \frac{L_2(L_1 - \mu_1)}{L_1 - \mu_2}}. \end{aligned}$$

By direct calculation, one can verify that

$$\begin{aligned} &\ell - U + \bar{\eta}_1 \left(\|g_1^1 - g_1^2\|^2 - \ell \right) + \sum_{k=2}^N \bar{\eta}_k \left(\|g_1^k - g_1^{k+1}\|^2 - \ell \right) + \bar{\eta}_{N+1} \left(\|g_1^{N+1} - g_2^{N+1}\|^2 - \ell \right) \\ &+ B \left(f^* - f_1^1 + f_2^1 + \Delta \right) + B \left(f_1^{N+1} - f_2^{N+1} - \frac{1}{2(L_1 - \mu_2)} \|g_1^{N+1} - g_2^{N+1}\|^2 - f^* \right) \\ &+ B \sum_{k=1}^N \left(f_1^k - f_1^{k+1} - \langle g_1^{k+1}, x^k - x^{k+1} \rangle - \frac{1}{2(1 - \frac{\mu_1}{L_1})} \left(\frac{1}{L_1} \|g_1^k - g_1^{k+1}\|^2 + \mu_1 \|x^k - x^{k+1}\|^2 \right. \right. \\ &\quad \left. \left. - \frac{2\mu_1}{L_1} \langle g_1^{k+1} - g_1^k, x^{k+1} - x^k \rangle \right) \right) + \bar{\lambda} \sum_{k=1}^{N-1} \left(f_2^{k+1} - f_2^k - \langle g_1^{k+1}, x^{k+1} - x^k \rangle \right. \\ &\quad \left. - \frac{1}{2(1 - \frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{k+1} - g_1^{k+2}\|^2 + \mu_2 \|x^k - x^{k+1}\|^2 - \frac{2\mu_2}{L_2} \langle g_1^{k+2} - g_1^{k+1}, x^{k+1} - x^k \rangle \right) \right) \\ &+ (\bar{\lambda} - B) \sum_{k=1}^{N-1} \left(f_2^k - f_2^{k+1} - \langle g_1^{k+2}, x^k - x^{k+1} \rangle - \frac{1}{2(1 - \frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{k+1} - g_1^{k+2}\|^2 \right. \right. \\ &\quad \left. \left. + \mu_2 \|x^k - x^{k+1}\|^2 - \frac{2\mu_2}{L_2} \langle g_1^{k+2} - g_1^{k+1}, x^{k+1} - x^k \rangle \right) \right) + (\bar{\lambda} - B) \left(f_2^N - f_2^{N+1} - \langle g_2^{N+1}, x^N - x^{N+1} \rangle \right. \\ &\quad \left. - \frac{1}{2(1 - \frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{N+1} - g_2^{N+1}\|^2 + \mu_2 \|x^N - x^{N+1}\|^2 - \frac{2\mu_2}{L_2} \langle g_2^{N+1} - g_1^{N+1}, x^{N+1} - x^N \rangle \right) \right) \\ &+ \bar{\lambda} \left(f_2^{N+1} - f_2^N - \langle g_1^{N+1}, x^{N+1} - x^N \rangle - \frac{1}{2(1 - \frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{N+1} - g_2^{N+1}\|^2 + \mu_2 \|x^N - x^{N+1}\|^2 \right. \right. \\ &\quad \left. \left. - \frac{2\mu_2}{L_2} \langle g_2^{N+1} - g_1^{N+1}, x^{N+1} - x^N \rangle \right) \right) \\ &= -\bar{\beta}_1^{-1} \sum_{i=1}^N \left\| \bar{\beta}_1 g_1^i - \bar{\beta}_1 g_1^{i+1} - \bar{\alpha}_1 x^i + \bar{\alpha}_1 x^{i+1} \right\|^2 - \bar{\alpha}_2^{-1} \sum_{i=1}^{N-1} \left\| \bar{\alpha}_2 x^i - \bar{\alpha}_2 x^{i+1} - \bar{\beta}_2 g_1^{i+1} + \bar{\beta}_2 g_1^{i+2} \right\|^2 \\ &\quad - \bar{\alpha}_2^{-1} \left\| \bar{\alpha}_2 x^N - \bar{\alpha}_2 x^{N+1} - \bar{\beta}_2 g_1^{N+1} + \bar{\beta}_2 g_2^{N+1} \right\|^2 \leq 0, \end{aligned}$$

where

$$\begin{aligned} \bar{\alpha}_1 &= \frac{\mu_1 B}{2(L_1 - \mu_1)}, \quad \bar{\beta}_1 = \frac{\mu_1 B}{2L_2(L_1 - \mu_1)}, \\ \bar{\alpha}_2 &= \frac{(-\mu_1 L_2^2 - 2\mu_1 \mu_2 L_2 + \mu_1 L_1 L_2 + \mu_1 \mu_2 L_1 + \mu_2 L_1 L_2) B}{2(L_1 - \mu_1)(L_2 - \mu_2)}, \\ \bar{\beta}_2 &= \frac{(L_1 L_2 \mu_2 - 2\mu_1 \mu_2 L_2 + \mu_1 \mu_2 L_1 - \mu_1 L_2^2 + \mu_1 L_1 L_2) B}{2L_2(L_1 - \mu_1)(L_2 - \mu_2)}. \end{aligned}$$

It is readily seen that $\bar{\lambda}, \bar{\eta}_k$ ($k \in \{1, \dots, N + 1\}$), $\bar{\lambda} - B, \bar{\beta}_1, \bar{\alpha}_2 \geq 0$. Thus, we have $\ell \leq U$ for any feasible point of problem (11). Now, we consider $L_1 < L_2$. In this case, because bound (12) does not depend on μ_1 , we may assume $\mu_1 = 0$ in problem (11). Let

$$\begin{aligned} \hat{\lambda} &= \frac{2(L_1L_2 - \mu_2(2L_1 - L_2))}{\left(L_1 + L_2 + \mu_2\left(\frac{L_2}{L_1} - 3\right)\right)N + \frac{L_1(L_2 - \mu_2)}{L_1 - \mu_2}} \\ \hat{\eta}_1 &= \frac{\frac{L_2(L_1 + \mu_2)}{L_1} - 2\mu_2}{\left(L_1 + L_2 + \mu_2\left(\frac{L_2}{L_1} - 3\right)\right)N + \frac{L_1(L_2 - \mu_2)}{L_1 - \mu_2}} \\ \hat{\eta}_k &= \frac{\frac{L_2(L_1 + \mu_2)}{L_1} + (L_1 - 3\mu_2)}{\left(L_1 + L_2 + \mu_2\left(\frac{L_2}{L_1} - 3\right)\right)N + \frac{L_1(L_2 - \mu_2)}{L_1 - \mu_2}}, \quad k \in \{2, \dots, N\} \\ \hat{\eta}_{N+1} &= 1 - \hat{\eta}_1 - \sum_{k=2}^N \hat{\eta}_k = \frac{\frac{L_1(L_2 - \mu_2)}{L_1 - \mu_2} + L_1 - \mu_2}{\left(L_1 + L_2 + \mu_2\left(\frac{L_2}{L_1} - 3\right)\right)N + \frac{L_1(L_2 - \mu_2)}{L_1 - \mu_2}}. \end{aligned}$$

With some calculation, one can establish that

$$\begin{aligned} &\ell - U + \hat{\eta}_1 \left(\|g_1^1 - g_1^2\|^2 - \ell \right) + \sum_{k=2}^N \hat{\eta}_k \left(\|g_1^k - g_1^{k+1}\|^2 - \ell \right) + \hat{\eta}_{N+1} \left(\|g_1^{N+1} - g_2^{N+1}\|^2 - \ell \right) \\ &+ B \left(f^* - f_1^1 + f_2^1 + \Delta \right) + B \left(f_1^{N+1} - f_2^{N+1} - \frac{1}{2(L_1 - \mu_2)} \|g_1^{N+1} - g_2^{N+1}\|^2 - f^* \right) \\ &+ (\hat{\lambda} - B) \sum_{k=1}^N \left(f_1^{k+1} - f_1^k - \langle g_1^k, x^{k+1} - x^k \rangle - \frac{1}{2L_1} \|g_1^{k+1} - g_1^k\|^2 \right) \\ &+ \hat{\lambda} \sum_{k=1}^N \left(f_1^k - f_1^{k+1} - \langle g_1^{k+1}, x^k - x^{k+1} \rangle - \frac{1}{2L_1} \|g_1^k - g_1^{k+1}\|^2 \right) \\ &+ B \sum_{k=1}^{N-1} \left(f_2^{k+1} - f_2^k - \langle g_1^{k+1}, x^{k+1} - x^k \rangle \right) \\ &- \frac{1}{2(1 - \frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{k+1} - g_1^{k+2}\|^2 + \mu_2 \|x^k - x^{k+1}\|^2 - \frac{2\mu_2}{L_2} \langle g_1^{k+2} - g_1^{k+1}, x^{k+1} - x^k \rangle \right) \\ &+ B \left(f_2^{N+1} - f_2^N - \langle g_1^{N+1}, x^{N+1} - x^N \rangle \right) \\ &- \frac{1}{2(1 - \frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{N+1} - g_2^{N+1}\|^2 + \mu_2 \|x^N - x^{N+1}\|^2 - \frac{2\mu_2}{L_2} \langle g_2^{N+1} - g_1^{N+1}, x^{N+1} - x^N \rangle \right) \\ &= -\hat{\beta}_1^{-1} \sum_{i=1}^N \left\| \hat{\beta}_1 g_1^i - \hat{\beta}_1 g_1^{i+1} - \hat{\alpha}_1 x_i^i + \hat{\alpha}_1 x^{i+1} \right\|^2 - \hat{\alpha}_2^{-1} \sum_{i=1}^{N-1} \left\| \hat{\alpha}_2 x^i - \hat{\alpha}_2 x^{i+1} - \hat{\beta}_2 g_1^{i+1} + \hat{\beta}_2 g_1^{i+2} \right\|^2 \\ &- \hat{\alpha}_2^{-1} \left\| \hat{\alpha}_2 x^N - \hat{\alpha}_2 x^{N+1} - \hat{\beta}_2 g_1^{N+1} + \hat{\beta}_2 g_2^{N+1} \right\|^2 \leq 0, \end{aligned}$$

where

$$\hat{\alpha}_1 = \frac{\mu_2 B(1 - \frac{L_1}{L_2})}{2L_1(1 - \frac{\mu_2}{L_2})}, \quad \hat{\alpha}_2 = \frac{\mu_2 L_1 B}{2(L_2 - \mu_2)}, \quad \hat{\beta}_1 = \frac{\mu_2 B(1 - \frac{L_1}{L_2})}{2L_1^2(1 - \frac{\mu_2}{L_2})}, \quad \hat{\beta}_2 = \frac{\mu_2 B}{2(L_2 - \mu_2)}.$$

It is readily seen that $\hat{\lambda}, \hat{\eta}_k$ ($k \in \{1, \dots, N + 1\}$), $\hat{\lambda} - B, \hat{\beta}_1, \hat{\alpha}_2 \geq 0$. The rest of proof is similar to that of the former case, and the proof is complete.

The theorem implies that Algorithm 1 is convergent when at least one of the Lipschitz constants is finite. In the following corollary, we simplify the inequality (12) for some special cases of L_1, L_2, μ_1 , and μ_2 .

Corollary 3.1 *Suppose that $f_1 \in \mathcal{F}_{\mu_1, L_1}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2, L_2}(\mathbb{R}^n)$. Then, after N iterations of Algorithm 1, one has:*

(i) *If $L_1 = \infty, L_2 < \infty$, then*

$$\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\| \leq \sqrt{\frac{2L_2^2 (f(x^1) - f^*)}{N(L_2 + \mu_1)}}.$$

(ii) *If $L_2 = \infty, L_1 < \infty$, then*

$$\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\| \leq \sqrt{\frac{2L_1^2 (L_1 - \mu_2) (f(x^1) - f^*)}{(L_1^2 - \mu_2^2) N + L_1^2}}. \tag{13}$$

(iii) *If $L_1, L_2 < \infty$, and $\mu_1 = \mu_2 = 0$ then*

$$\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\| \leq \sqrt{\frac{2L_1 L_2 (f(x^1) - f^*)}{(L_1 + L_2) N + L_2}}.$$

One can compare the results in Corollary 3.1 to that of Le Thi et al. [26] as reviewed earlier in Theorem 1.2. First of all, Corollary 3.1 part *iii*) does not assume strict convexity of f_1 or f_2 , and in this sense it is more general than the result in Theorem 1.2. If we do assume $\mu_1 + \mu_2 > 0$, then, for example, if $L_1 < \infty$, Theorem 1.2 implies,

$$\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\| \leq L_1 \sqrt{\frac{2 (f(x^1) - f^*)}{(\mu_1 + \mu_2) N}},$$

which is weaker than our bound (13) since $\mu_1 \leq L_1$, although the $O(1/\sqrt{N})$ dependence on N is the same. We will do a further, more direct, comparison of Theorem 1.2 and Corollary 3.1 in Sect. 3.2, where we consider the convergence rate of the sequence $\|x^{k+1} - x^k\|$.

3.1 An Example to Prove Tightness

In what follows, we give a class of functions for which the bound in Corollary 3.1, part *ii*), is attained, implying that the $O(1/\sqrt{N})$ convergence rate is tight. This result is new to the best of our knowledge.

Example 3.1 Let $L_1 \in (0, \infty)$. Suppose that N is selected such that $U := \sqrt{\frac{2}{L_1(N+1)}} < 1$. Let $f_1 : \mathbb{R} \rightarrow \mathbb{R}$ be given as follows,

$$f_1(x) = \begin{cases} \frac{L_1}{2}(x - i(1 - U))^2 + \frac{L_1 U i(i-1)(1-U)}{2} & x \in [\alpha_i, \beta_{i+1}) \\ L_1 U \beta_i(x - \beta_i) + \frac{\beta_i L_1 U^2}{2} + \frac{\beta_i(\beta_i-1)L_1 U}{2} & x \in [\beta_i, \alpha_i) \\ \frac{L_1}{2}x^2 & x \in (-\infty, 0), \end{cases}$$

where for $i \in \{1, \dots, N + 1\}$, $\alpha_i = i - U$, $\beta_i = i - 1$, and $\beta_{N+2} = \infty$. Note that $f_1 \in \mathcal{F}_{0,L_1}(\mathbb{R})$. Suppose that $f_2 : \mathbb{R} \rightarrow \mathbb{R}$ is given by

$$f_2(x) = \max_{1 \leq i \leq N+1} \left\{ L_1 U(i - 1)(x - i) + \frac{i(i-1)L_1 U}{2} \right\}.$$

An easy computation shows that

$$\begin{cases} \partial f_2(i) = [L_1 U(i - 1), L_1 U i] & i \in \{1, \dots, N, \} \\ \partial f_2(N + 1) = L_1 U N. \end{cases}$$

Note that $f_2 \in \mathcal{F}_{0,\infty}(\mathbb{R})$. One can check that, at $x^1 = N + 1$, one has $f_1(x^1) - f_2(x^1) = 1$, $\min_{x \in \mathbb{R}} f_1(x) - f_2(x) = 0$ and $\operatorname{argmin}_{x \in \mathbb{R}} f_1(x) - f_2(x) = [0, 1 - U]$. By taking x^1 as a starting point, Algorithm 1 can generate the following iterates:

$$x^k = N + 2 - k, \quad k \in \{1, \dots, N + 1\}.$$

Here at iteration, $k \in \{1, \dots, N + 1\}$, we set $g_2^k = L_1 U(N + 1 - k)$. It follows that $|\nabla f_1(x_k) - g_2^k| = \sqrt{\frac{2L_1}{N+1}}$, $k \in \{1, \dots, N + 1\}$. Hence,

$$\min_{1 \leq k \leq N+1} \|g_1^k - g_2^k\| = \sqrt{\frac{2L_1}{N+1}},$$

which shows bound (13) in Corollary 3.1 is exact for this example.

3.2 Convergence Rates for the Iterates

In this section, we investigate the implications of our results so far on convergence rates of the iterates $\{x^k\}$.

Proposition 3.1 Let $f_1 \in \mathcal{F}_{\mu_1,L_1}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2,L_2}(\mathbb{R}^n)$ and let $f(x^1) - f^* \leq \Delta$. If μ_1 or μ_2 is strictly positive, then after N iterations of Algorithm 1, one has:

$$\min_{1 \leq k \leq N} \|x^{k+1} - x^k\| \leq \left(\frac{\mathcal{A}}{\mathcal{B}N + \mathcal{C}} \cdot \Delta \right)^{\frac{1}{2}},$$

where

$$\begin{aligned}
 A &= 2 \left(\mu_2^{-1} \mu_1^{-1} - L_2^{-1} \mu_1^{-1} I_{\mathbb{R}_+} (\mu_2^{-1} - \mu_1^{-1}) - L_1^{-1} \mu_2^{-1} I_{\mathbb{R}_+} (\mu_1^{-1} - \mu_2^{-1}) \right), \\
 B &= \mu_2^{-1} + \mu_1^{-1} + L_2^{-1} \left(\frac{\mu_1}{\mu_2} - 3 \right) I_{\mathbb{R}_+} (\mu_2^{-1} - \mu_1^{-1}) + L_1^{-1} \left(\frac{\mu_2}{\mu_1} - 3 \right) I_{\mathbb{R}_+} (\mu_1^{-1} - \mu_2^{-1}), \\
 \text{and} \\
 C &= \frac{\mu_2^{-1} \mu_1^{-1} - L_2^{-1} \mu_1^{-1} I_{\mathbb{R}_+} (\mu_2^{-1} - \mu_1^{-1}) - L_1^{-1} \mu_2^{-1} I_{\mathbb{R}_+} (\mu_1^{-1} - \mu_2^{-1})}{\mu_2^{-1} - L_1^{-1}}.
 \end{aligned}$$

Proof The proof is based on the computation of the worst-case convergence rate of DCA for problem (8) by applying Theorem 3.1. By Toland duality, f^* is also a lower bound of problem (8). By virtue of conjugate function properties, it follows that $f_2^*(g_2^1) - f_1^*(g_1^1) - f^* \leq \Delta$ and $f_2^* \in \mathcal{F}_{L_2^{-1}, \mu_2^{-1}}(\mathbb{R}^n)$ and $f_1^* \in \mathcal{F}_{L_1^{-1}, \mu_1^{-1}}(\mathbb{R}^n)$. In addition, $x^{k+1} \in \partial f_1^*(g_1^k)$ and $x^k \in \partial f_2^*(g_2^k)$ for $k \in \{1, \dots, N\}$. Hence, all assumptions of Theorem 3.1 hold, and subsequently the bound follows from Theorem 3.1.

Recall the known result from Theorem 1.2:

$$\min_{1 \leq k \leq N} \|x^{k+1} - x^k\| \leq \left(\frac{2(f(x^1) - f^*)}{N(\mu_1 + \mu_2)} \right)^{\frac{1}{2}}. \tag{14}$$

By employing Theorem 3.1, we get

$$\min_{1 \leq k \leq N} \|x^{k+1} - x^k\| \leq \left(\frac{2(f(x^1) - f^*)}{N(\mu_1 + \mu_2) + \mu_1} \right)^{\frac{1}{2}},$$

which is tighter than the bound (14). Moreover, the bound given in Proposition 3.1 provides more information concerning the worst-case convergence rate of the DCA when $L_1 < \infty$ or $L_2 < \infty$.

4 Performance Estimation using a Convergence Criterion for Critical Points in the Non-smooth Case

Theorem 3.1 addresses the case that f_1 or f_2 is L -smooth with $L < \infty$. In what follows, we investigate the case that f_1 and f_2 are proper convex functions and where both may be non-smooth. For this general case, we need to adopt a different termination criterion to obtain results, since the termination criterion $\|g_1^k - g_2^k\| \leq \epsilon$ may be of no use in this case. For example, suppose that a DC function $f : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ is given by

$$f(x) = \begin{cases} f_1(x) - f_2(x) & x \geq 0 \\ \infty & x < 0, \end{cases}$$

where

$$f_1(x) = \max_{n \in \mathbb{N} \cup \{0\}} \{-n(x - 2^{-n}) + 2 - 2^{1-n} - n2^{-n}\},$$

$$f_2(x) = \max_{n \in \mathbb{N} \cup \{0\}} \{-(n + 1)(x - 2^{-n}) + 2 - 3(2^{-n}) - n2^{-n}\}.$$

With $x^1 = 1$ and the given DC decomposition, Algorithm 1 may generate

$$x^k = 2^{-k}, \quad g_1^k = -(k - 1), \quad g_2^k = -k, \quad k \in \{1, 2, \dots\}.$$

As $|g_1^k - g_2^k| = 1$, Algorithm 1 never stops by employing the given termination criterion while it is convergent to global minimum $\bar{x} = 0$. We therefore will use the termination criterion of the following value being sufficiently small:

$$T(x^{k+1}) := f_1(x^k) - f_2(x^k) - \min_{x \in \mathbb{R}^n} \left(f_1(x) - f_2(x^k) - \langle g_2^k, x - x^k \rangle \right)$$

$$= f_1(x^k) - f_1(x^{k+1}) - \langle g_2^k, x^k - x^{k+1} \rangle. \tag{15}$$

Note that $T(x^{k+1}) \geq 0$. It follows that if $T(x^{k+1}) = 0$ then $f(x^k) = f(x^{k+1})$, and $x^k \in \operatorname{argmin}_{x \in \mathbb{R}^n} f_1(x) - f_2(x^k) - \langle g_2^k, x - x^k \rangle$. Indeed, by the optimality conditions for convex problems, we have $\partial f_1(x^k) \cap \partial f_2(x^k) \neq \emptyset$. Consequently, $T(x^{k+1}) = 0$ implies that x^k is a critical point of problem (6). The aforementioned stopping criterion has also been employed for the analysis of the Frank–Wolfe method for non-convex problems; see Eq. (2.6) in [18].

In what follows, we investigate Algorithm 1 with the termination criterion $T(x^{k+1}) < \epsilon$ for the given accuracy $\epsilon > 0$. The performance estimation problem with termination criterion (15) may be written as follows,

$$\begin{aligned} & \max \ell \\ & \text{s.t. } f_1(x^k) - f_1(x^{k+1}) - \langle g_1^{k+1}, x^k - x^{k+1} \rangle \geq \ell \quad i \in \{1, \dots, N\} \\ & \frac{1}{2(1-\frac{\mu_1}{L_1})} \left(\frac{1}{L_1} \|g_1^i - g_1^j\|^2 + \mu_1 \|x^i - x^j\|^2 - \frac{2\mu_1}{L_1} \langle g_1^j - g_1^i, x^j - x^i \rangle \right) \\ & \leq f_1^i - f_1^j - \langle g_1^j, x^i - x^j \rangle \quad i, j \in \{1, \dots, N + 1\} \\ & \frac{1}{2(1-\frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{i+1} - g_1^{j+1}\|^2 + \mu_2 \|x^i - x^j\|^2 - \frac{2\mu_2}{L_2} \langle g_1^{j+1} - g_1^{i+1}, x^j - x^i \rangle \right) \\ & \leq f_2^i - f_2^j - \langle g_1^{j+1}, x^i - x^j \rangle \quad i, j \in \{1, \dots, N\} \\ & \frac{1}{2(1-\frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_2^{N+1} - g_1^{j+1}\|^2 + \mu_2 \|x^{N+1} - x^j\|^2 - \frac{2\mu_2}{L_2} \langle g_1^{j+1} - g_2^{N+1}, x^j - x^{N+1} \rangle \right) \\ & \leq f_2^{N+1} - f_2^j - \langle g_1^{j+1}, x^{N+1} - x^j \rangle \quad j \in \{1, \dots, N\} \\ & \frac{1}{2(1-\frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_1^{i+1} - g_2^{N+1}\|^2 + \mu_2 \|x^i - x^{N+1}\|^2 - \frac{2\mu_2}{L_2} \langle g_2^{N+1} - g_1^{i+1}, x^{N+1} - x^i \rangle \right) \end{aligned}$$

$$\begin{aligned} &\leq f_2^i - f_2^{N+1} - \langle g_2^{N+1}, x^i - x^j \rangle \quad i \in \{1, \dots, N\} \\ f_1^k - f_2^k &\geq f^* \quad k \in \{1, \dots, N + 1\} \\ f_1^1 - f_2^1 - f^* &\leq \Delta. \end{aligned} \tag{16}$$

Note that we do not employ Lemma 2.1 in this formulation because we consider a general DC problem. Using the performance estimation procedure as described before the proof of Theorem 3.1 once more, we obtain the following result.

Theorem 4.1 *Let $f_1 \in \mathcal{F}_{\mu_1, L_1}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2, L_2}(\mathbb{R}^n)$. Then, after N iterations of Algorithm 1, one has*

$$\begin{aligned} &\min_{1 \leq k \leq N} f_1(x^k) - f_1(x^{k+1}) - \langle g_2^k, x^k - x^{k+1} \rangle \\ &\leq \min \left\{ \frac{L_1}{N(L_1 + \mu_2)}, \frac{L_2}{N(L_2 + \mu_1) - \mu_1} \right\} (f(x^1) - f^*). \end{aligned} \tag{17}$$

Proof We show separately that $\frac{L_1(f(x^1) - f^*)}{N(L_1 + \mu_2)}$ and $\frac{L_2(f(x^1) - f^*)}{N(L_2 + \mu_1) - \mu_1}$ are upper bounds for problem (16). The proof is analogous to that of Theorem 3.1. First, consider the bound $\frac{L_1(f(x^1) - f^*)}{N(L_1 + \mu_2)}$. Since the given bound does not depend on μ_1 and L_2 , we may assume without loss of generality that $L_2 = \infty$ and $\mu_1 = 0$. Suppose that $B_1 = \frac{L_1}{N(L_1 + \mu_2)}$. With some algebra, one can show that

$$\begin{aligned} &\ell - B_1 \Delta + \frac{1}{N} \sum_{k=1}^N (f_1^k - f_1^{k+1} - \langle g_1^{k+1}, x^k - x^{k+1} \rangle - \ell) + B_1 (f_1^{N+1} - f_2^{N+1} - f^*) \\ &\quad + B_1 (f^* - f_1^1 + f_2^1 + \Delta) + (\frac{1}{N} - B_1) \sum_{k=1}^N (f_1^{k+1} - f_1^k - \langle g_1^k, x^{k+1} - x^k \rangle - \frac{1}{2L_1} \|g_1^{k+1} - g_1^k\|^2) \\ &\quad + B_1 \sum_{k=1}^N (f_2^{k+1} - f_2^k - \langle g_1^{k+1}, x^{k+1} - x^k \rangle - \frac{\mu_2}{2} \|x^{k+1} - x^k\|^2) \\ &= -\frac{B_1 \mu_2}{2} \sum_{k=1}^N \|x^k - x^{k+1} - \frac{1}{L_1} (g_1^k - g_1^{k+1})\|^2 \leq 0. \end{aligned}$$

The rest of proof is similar to that of Theorem 3.1. Now, we consider the bound $\frac{L_2(f(x^1) - f^*)}{N(L_2 + \mu_1) - \mu_1}$. Without loss generality, we may assume that $L_1 = \infty$ and $\mu_2 = 0$. By doing some calculus, one can show that

$$\begin{aligned} &\ell - B_2 \Delta + B_2 (f_1^1 - f_1^2 - \langle g_1^2, x^1 - x^2 \rangle - \ell) + B_2 (f_1^{N+1} - f_2^{N+1} - f^*) \\ &\quad + B_2 (f^* - f_1^1 + f_2^1 + \Delta) + \frac{1 - B_2}{N - 1} \sum_{k=2}^N (f_1^k - f_1^{k+1} - \langle g_1^{k+1}, x^k - x^{k+1} \rangle - \ell) \\ &\quad + \alpha \sum_{k=2}^N (f_1^{k+1} - f_1^k - \langle g_1^k, x^{k+1} - x^k \rangle - \frac{\mu_1}{2} \|x^{k+1} - x^k\|^2) \\ &\quad + B_2 \sum_{k=1}^N (f_2^{k+1} - f_2^k - \langle g_1^{k+1}, x^{k+1} - x^k \rangle - \frac{1}{2L_2} \|g_1^{k+2} - g_1^{k+1}\|^2) \end{aligned}$$

$$\begin{aligned}
 &+ B_2 \left(f_2^{N+1} - f_2^N - \langle g_1^{N+1}, x^{N+1} - x^N \rangle - \frac{1}{2L_2} \|g_2^{N+1} - g_1^{N+1}\|^2 \right) \\
 &= -\frac{B_2}{2L_2} \|g_2^{N+1} - g_1^{N+1}\|^2 - \frac{B_2}{2L_2} \sum_{k=2}^N \|g_1^k - g_1^{k+1} - \frac{\alpha L_2}{B_2} (x^k - x^{k+1})\|^2 \leq 0,
 \end{aligned}$$

where $B_2 = \frac{L_2}{N(L_2 + \mu_1) - \mu_1}$ and $\alpha = \frac{1 - B_2}{N - 1} - B_2$. Since we assume $L_2 > \mu_1$, we have $B_2, \alpha \geq 0$. The rest of the proof runs as before. \square

The important point is that the last result provides a rate of convergence even if neither L_1 nor L_2 is finite, and we therefore state it as a corollary.

Corollary 4.1 *Let $f_1 \in \mathcal{F}_{\mu_1, \infty}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2, \infty}(\mathbb{R}^n)$, i.e. consider any DC decomposition in problem (1). Then, after N iterations of Algorithm 1, one has*

$$\min_{1 \leq k \leq N} f_1(x^k) - f_1(x^{k+1}) - \langle g_2^k, x^k - x^{k+1} \rangle \leq \frac{1}{N} (f(x^1) - f^*).$$

This result is new to the best of our knowledge.

5 Linear Convergence of the DCA under the Polyak–Łojasiewicz Inequality

In the section, we provide some sufficient conditions under which the DCA is linearly convergent. Similar to the former sections, we employ the performance estimation for obtaining convergence rate.

In recent years, the linear convergence of some optimization methods for non-convex problems has been investigated under the Polyak–Łojasiewicz (PL) inequality; see [2, 12, 25] and the reference therein. We say that f satisfies PL inequality on X if there exists $\eta > 0$ such that

$$f(x) - f^* \leq \frac{1}{2\eta} \|\xi\|^2, \quad \forall x \in X, \forall \xi \in \text{co}(\partial_l f(x)). \tag{18}$$

Note that when f is differentiable inequality (18) is a special case of (3) with $\theta = \frac{1}{2}$ and different ground set. If f_1 or f_2 is strictly differentiable, we have $\text{co}(\partial_l f) = \partial f_1 - \partial f_2$; see Example 10.10 in [38]. Hence, the performance estimation problem with the PL inequality may be formulated as follows:

$$\begin{aligned}
 &\max \frac{(f_1^2 - f_2^2) - f^*}{(f_1^1 - f_2^1) - f^*} \\
 &\text{s.t. } \frac{1}{2(1 - \frac{\mu_1}{L_1})} \left(\frac{1}{L_1} \|g_1^i - g_1^j\|^2 + \mu_1 \|x^i - x^j\|^2 - \frac{2\mu_1}{L_1} \langle g_1^j - g_1^i, x^j - x^i \rangle \right) \\
 &\quad \leq f_1^i - f_1^j - \langle g_1^j, x^i - x^j \rangle \quad i, j \in \{1, 2\} \\
 &\quad \frac{1}{2(1 - \frac{\mu_2}{L_2})} \left(\frac{1}{L_2} \|g_2^i - g_2^j\|^2 + \mu_2 \|x^i - x^j\|^2 - \frac{2\mu_2}{L_2} \langle g_2^j - g_2^i, x^j - x^i \rangle \right)
 \end{aligned}$$

$$\begin{aligned}
 &\leq f_2^i - f_2^j - \left\langle g_2^j, x^i - x^j \right\rangle \quad i, j \in \{1, 2\} \\
 f_1^k - f_2^k &\geq f^* \quad k \in \{1, 2\} \\
 g_2^1 &= g_1^2 \\
 (f_1^k - f_2^k) - f^* &\leq \frac{1}{2\eta} \|g_1^k - g_2^k\|^2, \quad k \in \{1, 2\}.
 \end{aligned} \tag{19}$$

By doing constraint aggregation in problem (19) as before (i.e. demonstrating a dual feasible solution and using weak duality), we obtain the following linear convergence rate for the DCA under the PL inequality.

Theorem 5.1 *Let $f_1 \in \mathcal{F}_{\mu_1, L_1}(\mathbb{R}^n)$ and $f_2 \in \mathcal{F}_{\mu_2, L_2}(\mathbb{R}^n)$. If L_1 or L_2 is finite and if f satisfies PL inequality on $X = \{x : f(x) \leq f(x^1)\}$, then for x^2 from Algorithm 1, we have*

$$\frac{f(x^2) - f^*}{f(x^1) - f^*} \leq \left(\frac{1 - \frac{\eta}{L_1}}{1 + \frac{\eta}{L_2}} \right). \tag{20}$$

Proof Since the given bound is independent of μ_1 and μ_2 , without loss of generality, we assume that $\mu_1 = \mu_2 = 0$. In addition, we assume that $f^* = 0$. Direct calculation shows that

$$\begin{aligned}
 &(f_1^2 - f_2^2) - f^* - \left(\frac{1 - \frac{\eta}{L_1}}{1 + \frac{\eta}{L_2}} \right) \left((f_1^1 - f_2^1) - f^* \right) + \left(\frac{1}{1 + \frac{\eta}{L_2}} \right) \\
 &\quad \times \left(f_1^1 - f_2^1 - \left\langle g_1^1, x^1 - x^2 \right\rangle - \frac{1}{2L_1} \|g_1^1 - g_2^1\|^2 \right) \\
 &\quad + \left(\frac{1}{1 + \frac{\eta}{L_2}} \right) \left(f_2^2 - f_2^1 - \left\langle g_1^1, x^2 - x^1 \right\rangle - \frac{1}{2L_2} \|g_1^1 - g_2^1\|^2 \right) + \left(\frac{\frac{\eta}{L_1}}{1 + \frac{\eta}{L_2}} \right) \\
 &\quad \times \left(\frac{1}{2\eta} \|g_1^1 - g_2^1\|^2 - f_1^1 + f_2^1 \right) + \left(\frac{\frac{\eta}{L_2}}{1 + \frac{\eta}{L_2}} \right) \left(\frac{1}{2\eta} \|g_1^2 - g_2^2\|^2 - f_1^2 + f_2^2 \right) = 0.
 \end{aligned}$$

As all the multipliers in the last expression are non-negative, for any feasible solution of problem (11), we have

$$f(x^2) - f^* - \left(\frac{1 - \frac{\eta}{L_1}}{1 + \frac{\eta}{L_2}} \right) (f(x^1) - f^*) \leq 0,$$

completing the proof.

Note that Theorem 1.1 by Le Thi et al. [27] does not imply Theorem 5.1 if inequality (3) holds on $\{x : f(x) \leq f(x^1)\}$ with $\theta = \frac{1}{2}$, since we assume neither strong convexity of f_1 or f_2 , nor boundedness of the sequence of iterates. Moreover, we give explicit expressions for the constants that determine the linear convergence rate of the sequence of objective values.

6 Conclusion

We have shown that the performance estimation framework of Drori and Teboulle [16] yields new insights into the convergence behavior of the difference-of-convex algorithm (DCA). As future work, one may also consider the convergence of the DCA on more restricted classes of DC problems, e.g. where f_1 and f_2 are convex polynomials, as studied in [3]. For constrained problems, even the case where f_1 and f_2 are quadratic polynomials is of interest, e.g. in the study of (extended) trust region problems.

Acknowledgements This work was supported by the Dutch Scientific Council (NWO) Grant *Optimization for and with Machine Learning*, OCENW.GROOT.2019.015.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abbaszadehpeivasti, H., De Klerk, E., Zamani, M.: The exact worst-case convergence rate of the gradient method with fixed step lengths for L-smooth functions. *Optim. Lett.* **16**(6), 1649–1661 (2022). <https://doi.org/10.1007/s11590-021-01821-1>
2. Abbaszadehpeivasti, H., De Klerk, E., Zamani, M.: Conditions for linear convergence of the gradient method for non-convex optimization. *Optim. Lett.* (2023). <https://doi.org/10.1007/s11590-023-01981-2>
3. Ahmadi, A.A., Hall, G.: DC decomposition of nonconvex polynomials with algebraic techniques. *Math. Program.* **169**(1), 69–94 (2018). <https://doi.org/10.1007/s10107-017-1144-5>
4. Alvarado, A., Scutari, G., Pang, J.S.: A new decomposition method for multiuser DC-programming and its applications. *IEEE Trans. Signal Process.* **62**(11), 2984–2998 (2014). <https://doi.org/10.1109/TSP.2014.2315167>
5. An, L.T.H., Tao, P.D.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Ann. Oper. Res.* **133**(1–4), 23–46 (2005). <https://doi.org/10.1007/s10479-004-5022-1>
6. Astorino, A., Fuduli, A., Gaudioso, M.: Margin maximization in spherical separation. *Comput. Optim. Appl.* **53**(2), 301–322 (2012). <https://doi.org/10.1007/s10589-012-9486-7>
7. Bagirov, A.M., Ugon, J.: Nonsmooth DC programming approach to clusterwise linear regression: optimality conditions and algorithms. *Optim. Methods Softw.* **33**(1), 194–219 (2018). <https://doi.org/10.1080/10556788.2017.1371717>
8. Bagirov, A.M., Taheri, S., Ugon, J.: Nonsmooth DC programming approach to the minimum sum-of-squares clustering problems. *Pattern Recogn.* **53**, 12–24 (2016). <https://doi.org/10.1016/j.patcog.2015.11.011>
9. Beck, A.: *First-order Methods in Optimization*. SIAM, Philadelphia (2017)
10. Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM J. Optim.* **17**(4), 1205–1223 (2006). <https://doi.org/10.1137/050644641>
11. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **146**(1), 459–494 (2014). <https://doi.org/10.1007/s10107-013-0701-9>

12. Bolte, J., Nguyen, T.P., Peypouquet, J., Suter, B.W.: From error bounds to the complexity of first-order descent methods for convex functions. *Math. Program.* **165**, 471–507 (2017). <https://doi.org/10.1007/s10107-016-1091-6>
13. Chen, P.C., Hansen, P., Jaumard, B., Tuy, H.: Solution of the multisource Weber and conditional Weber problems by D.C. programming. *Oper. Res.* **46**(4), 548–562 (1998). <https://doi.org/10.1287/opre.46.4.548>
14. De Klerk, E., Glineur, F., Taylor, A.B.: Worst-case convergence analysis of inexact gradient and Newton methods through semidefinite programming performance estimation. *SIAM J. Optim.* **30**(3), 2053–2082 (2020). <https://doi.org/10.1137/19M1281368>
15. De Klerk, E., Glineur, F., Taylor, A.B.: On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions. *Optim. Lett.* **11**(7), 1185–1199 (2017). <https://doi.org/10.1007/s11590-016-1087-4>
16. Drori, Y., Teboulle, M.: Performance of first-order methods for smooth convex minimization: a novel approach. *Math. Program.* **145**(1), 451–482 (2014). <https://doi.org/10.1007/s10107-013-0653-0>
17. Gasso, G., Rakotomamonjy, A., Canu, S.: Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *IEEE Trans. Signal Process.* **57**(12), 4686–4698 (2009). <https://doi.org/10.1109/TSP.2009.2026004>
18. Ghadimi, S.: Conditional gradient type methods for composite nonlinear and stochastic optimization. *Math. Program.* **173**(1), 431–464 (2019). <https://doi.org/10.1007/s10107-017-1225-5>
19. Jy, Gotoh, Takeda, A., Tono, K.: DC formulations and algorithms for sparse optimization problems. *Math. Program.* **169**(1), 141–176 (2018). <https://doi.org/10.1007/s10107-017-1181-0>
20. Hartman, P.: On functions representable as a difference of convex functions. *Pac. J. Math.* **9**(3), 707–713 (1959)
21. Hiriart-Urruty, J.B.: Generalized differentiability/duality and optimization for problems dealing with differences of convex functions. In: Ponstein, J. (ed.) *Convexity and Duality in Optimization*, vol. 256. Springer, Berlin, Heidelberg (1985). https://doi.org/10.1007/978-3-642-45610-7_3
22. Holmberg, K., Tuy, H.: A production-transportation problem with stochastic demand and concave production costs. *Math. Program.* **85**(1), 157–179 (1999). <https://doi.org/10.1007/s101070050050>
23. Horst, R., Thoai, N.V.: DC programming: overview. *J. Optim. Theory Appl.* **103**(1), 1–43 (1999). <https://doi.org/10.1023/A:1021765131316>
24. Joki, K., Bagirov, A.M., Karmitsa, N., Mäkelä, M.M., Taheri, S.: Double bundle method for finding Clarke stationary points in nonsmooth DC programming. *SIAM J. Optim.* **28**(2), 1892–1919 (2018). <https://doi.org/10.1137/16M1115733>
25. Karimi, H., Nutini, J., Schmidt, M.: Linear convergence of gradient and proximal-gradient methods under the Polyak–Lojasiewicz condition. In: Frasconi, P., Landwehr, N., Manco, G., Vreeken, J. (eds.) *Machine Learning and Knowledge Discovery in Databases*, vol. 9851. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46128-1_50
26. Le Thi, H.A., Phan, D.N., Dinh, T.P.: DCA based approaches for bi-level variable selection and application for estimate multiple sparse covariance matrices. *Neurocomputing* **466**, 162–177 (2021). <https://doi.org/10.1016/j.neucom.2021.09.039>
27. Le Thi, H.A., Dinh, T.P., Pham, D.T.: Convergence analysis of difference-of-convex algorithm with subanalytic data. *J. Optim. Theory Appl.* **179**(1), 103–126 (2018). <https://doi.org/10.1007/s10957-018-1345-y>
28. Le Thi, H.A., Dinh, T.P.: DC programming and DCA: thirty years of developments. *Math. Program.* **169**(1), 5–68 (2018). <https://doi.org/10.1007/s10107-018-1235-y>
29. Le Thi, H.A., Nguyen, M.C.: DCA based algorithms for feature selection in multi-class support vector machine. *Ann. Oper. Res.* **249**(1–2), 273–300 (2017). <https://doi.org/10.1007/s10479-016-2333-y>
30. Lipp, T., Boyd, S.: Variations and extension of the convex-concave procedure. *Optim. Eng.* **17**(2), 263–287 (2016). <https://doi.org/10.1007/s11081-015-9294-x>
31. Lou, Y., Zeng, T., Osher, S., Xin, J.: A weighted difference of anisotropic and isotropic total variation model for image processing. *SIAM J. Imag. Sci.* **8**(3), 1798–1823 (2015). <https://doi.org/10.1137/14098435X>
32. Lu, Z., Zhou, Z.: Nonmonotone enhanced proximal DC algorithms for a class of structured nonsmooth DC programming. *SIAM J. Optim.* **29**(4), 2725–2752 (2019). <https://doi.org/10.1137/18M1214342>
33. Lu, Z., Zhou, Z., Sun, Z.: Enhanced proximal DC algorithms with extrapolation for a class of structured nonsmooth DC minimization. *Math. Program.* **176**(1), 369–401 (2019). <https://doi.org/10.1007/s10107-018-1318-9>

34. Melzer, D.: On the expressibility of piecewise-linear continuous functions as the difference of two piecewise-linear convex functions. In: Demyanov, V.F., Dixon, L.C.W. (eds.) Quasidifferential. Calculus Mathematical Programming Studies, vol. 29. Springer, Berlin, Heidelberg (1986). <https://doi.org/10.1007/BFb0121142>
35. Nesterov, Y.: Lectures on Convex Optimization. Springer, Cham (2018)
36. Pang, J.S., Razaviyayn, M., Alvarado, A.: Computing B-stationary points of nonsmooth DC programs. *Math. Oper. Res.* **42**(1), 95–118 (2017). <https://doi.org/10.1287/moor.2016.0795>
37. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
38. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, New York (2009)
39. Sun, K., Sun, X.A.: Algorithms for difference-of-convex programs based on difference-of-Moreau-envelopes smoothing. *INFORMS J. Optim.* (2022). <https://doi.org/10.1287/ijoo.2022.0087>
40. Tao, P.D., An, L.T.H.: Convex analysis approach to DC programming: theory, algorithms and applications. *Acta Math. Vietnam* **22**(1), 289–355 (1997)
41. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Math. Program.* **161**(1–2), 307–345 (2017). <https://doi.org/10.1007/s10107-016-1009-3>
42. Taylor, A.B., Hendrickx, J.M., Glineur, F.: Exact worst-case performance of first-order methods for composite convex optimization. *SIAM J. Optim.* **27**(3), 1283–1313 (2017). <https://doi.org/10.1137/16M108104X>
43. Toland, J.F.: A duality principle for non-convex optimisation and the calculus of variations. *Arch. Ration. Mech. Anal.* **71**(1), 41–61 (1979). <https://doi.org/10.1007/BF00250669>
44. Tuy, H.: Convex Analysis and Global Optimization. Springer, Dordrecht (1998)
45. Yen, I.E., Peng, N., Wang, P.W., Lin, S.D. (2012). On convergence rate of concave-convex procedure. In: Sra, S., Agarwal, A. (eds.) Proceedings of the NIPS 2012 Optimization Workshop, pp. 31–35

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.