Optimization and Engineering (2023) 24:65–91
https://doi.org/10.1007/s11081-022-09716-5

RESEARCH ARTICLE

A refined inertial DC algorithm for DC programming

Yu You[1] · **Yi-Shuai Niu**[2,3]

Received: 29 April 2021 / Revised: 17 February 2022 / Accepted: 24 February 2022 /
Published online: 5 April 2022

Abstract

In this paper we consider the difference-of-convex (DC) programming problems, whose objective function is the difference of two convex functions. The classical DC Algorithm (DCA) is well-known for solving this kind of problems, which generally returns a critical point. Recently, an inertial DC algorithm (InDCA) equipped with heavy-ball inertial-force procedure was proposed in de Oliveira et al. (Set-Valued Variat Anal 27(4):895–919, 2019), which potentially helps to improve both the convergence speed and the solution quality. Based on InDCA, we propose a refined inertial DC algorithm (RInDCA) equipped with enlarged inertial step-size compared with InDCA. Empirically, larger step-size accelerates the convergence. We demonstrate the subsequential convergence of our refined version to a critical point. the Kurdyka-Łojasiewicz (KL) property of the objective function, we establish the sequential convergence of RInDCA. Numerical simulations on checking copositivity of matrices and image denoising problem show the benefit of larger step-size.

Keywords Difference-of-convex programming · Refined inertial DC algorithm · Kurdyka-Łojasiewicz property · Checking copositivity of matrices · Image denoising

Mathematics Subject Classification 90C26 · 90C30 · 68U10

This work is supported by the National Natural Science Foundation of China (Grant 11601327).

✉ Yi-Shuai Niu
yi-shuai.niu@polyu.edu.hk; niuyishuai@sjtu.edu.cn

Yu You
youyu0828@sjtu.edu.cn

[1] School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China

[2] Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[3] School of Mathematical Sciences and SJTU-Paristech, Shanghai Jiao Tong University, Shanghai, China

1 Introduction

Difference-of-convex (DC) programming, referring to the problems of minimizing a function which is the difference of two convex functions, forms an important class of nonconvex programming and has been studied extensively for decades, see e.g., (Hartman 1959; Hiriart-Urruty 1985; Pham and Souad 1986; Pham and Le Thi 1997; Pham and Le Thi 1998; Horst and Thoai 1999; Le Thi and Pham 2005; Aleksandrov 2012; Le Thi et al. 2012; Pham and Le Thi 2014; Souza et al. 2016; Joki et al. 2017; Pang et al. 2017; Le Thi and Pham 2018; de Oliveira 2020).

In the paper, we consider the standard DC program in form of

$$\min\left\{f(\mathbf{x}) := f_1(\mathbf{x}) - f_2(\mathbf{x}) \,:\, \mathbf{x} \in \mathbb{R}^n\right\}, \quad (P)$$

where f_1 and f_2 are proper closed and convex functions. Such a function f is called a DC function, while $f_1 - f_2$ is a DC decomposition of f with f_1 and f_2 being DC components. Formulation (P) also includes convex constrained problems $\min\{\widetilde{f}_1(\mathbf{x}) - f_2(\mathbf{x}) \,:\, \mathbf{x} \in X\}$, where $\widetilde{f}_1 : \mathbb{R}^n \to \mathbb{R}$ is convex and $X \subseteq \mathbb{R}^n$ is nonempty, closed and convex. By introducing the indicator function of X, i.e., $\delta_X(\mathbf{x}) = 0$ if $\mathbf{x} \in X$ and $\delta_X(\mathbf{x}) = \infty$ otherwise, this model recovers (P) as $\min\{(\widetilde{f}_1 + \delta_X)(\mathbf{x}) - f_2(\mathbf{x}) \,:\, \mathbf{x} \in \mathbb{R}^n\}$. Throughout the manuscript, we make the next assumptions for problem (P).

Assumption 1

(a) $\mathrm{dom}(f_1) \subseteq Y \subseteq \mathrm{int}(\mathrm{dom}(f_2))$, where $Y \subseteq \mathbb{R}^n$ is closed.
(b) The function f is bounded below, i.e., there exists a scalar f^* such that $f(\mathbf{x}) \geq f^*$ for all $\mathbf{x} \in \mathbb{R}^n$.

The term (a) of Assumption 1 guarantees the nonemptiness of the subdifferential of f_2 at any point of $\mathrm{dom}(f_1)$. It will be also used in Theorem 4.1.

The DC Algorithm (DCA) (Pham and Le Thi 1997; Pham and Le Thi 1998; Le Thi and Pham 2005, 2018) is renowned for solving (P), which has been introduced by Pham Dinh Tao in 1985 and extensively developed by Le Thi Hoai An and Pham Dinh Tao since 1994. Specifically, at iteration k, DCA obtains the next iterate $\mathbf{x}^{k+1}$ by solving the convex subproblem

$$\mathbf{x}^{k+1} \in \operatorname{argmin}\left\{f_1(\mathbf{x}) - \langle \mathbf{y}^k, \mathbf{x}\rangle \,:\, \mathbf{x} \in \mathbb{R}^n\right\},$$

where $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$. It has been proved in Pham and Le Thi (1997) that any limit point $\bar{\mathbf{x}}$ of the generated sequence $\{\mathbf{x}^k\}$ is a (DC) critical point of problem (P), i.e., the necessary local optimality condition $\partial f_1(\bar{\mathbf{x}}) \cap \partial f_2(\bar{\mathbf{x}}) \neq \emptyset$ is verified.

Recently, several accelerated algorithms for DC programming have been studied. Actacho et al. proposed in Artacho et al. (2018); Artacho and Vuong (2020) the boosted DC algorithms for unconstrained smooth/nonsmooth DC program, where the direction $\mathbf{x}^{k+1} - \mathbf{x}^k$, determined by the consecutive iterates of DCA, is verified as a descent direction of f at $\mathbf{x}^{k+1}$ when f_2 is strongly convex, thus a line-search procedure can be conducted along it to obtain a better candidate with lower objective value. Next, Niu et al. Niu andWang (2019) developed a boosted DCA by incorporating the line-search procedure for DC program (both smooth and nonsmooth) with convex constraints. Later, Actacho et al. Artacho et al. (2019) investigated the line-search idea in DC program with linear constraint, where a sufficient and necessary condition for determining the feasibility of the direction $\mathbf{x}^{k+1} - \mathbf{x}^k$ is proposed. Different from the boosted approaches above, another type of acceleration is the momentum methods. There are two renowned momentum methods: the Polyak's heavy-ball method (Polyak 1987; Zavriev and Kostyuk 1993) and the Nesterov's acceleration technique (Nesterov 1983, 2013). Indeed, Nesterov's acceleration belongs to the heavy-ball family. These two momentum methods have been successfully applied for various nonconvex optimizations problems (see, e.g., Adly et al. 2021; Ochs et al. 2014; Pock and Sabach 2016; Mukkamala et al. 2020). In de Oliveira and Tcheou (2019), de Oliveira et al. proposed an inertial DC algorithm (InDCA) for problem (P) in the premise that f_2 is strongly convex, where the heavy-ball inertial-force procedure is introduced. Phan et al. developed in Phan et al. (2018) an accelerated variant of DCA for (P) by incorporating the Nesterov's acceleration technique. At the same time, Wen et al. (2018) used the same technique for a special class of DC program, where f_1 is the sum of a smooth convex function and a (possibly non-smooth) convex function. Adly et al. Adly et al. (2021) proposed a gradient-based method, using both heavy-ball and Nesterov's inertial-forces, for optimizing a smooth function. They apply the algorithm for (P) by minimizing the DC Moreau envelope ${}^{\mu}f_1 - {}^{\mu}f_2$[1], which can return a critical point of $f_1 - f_2$. However, their methods require computing the proximal operators of f_1 and f_2. The above inertial-type methods often help to improve the performances (compared with their counterparts without inertial-forces) in both the convergence speed and the solution quality. Concerning the optimality condition of the DC program (P), the aforementioned methods only guarantee the critical point, which is weaker than the d-stationary point (see, e.g., Pang et al. 2017; de Oliveira 2020). Next, we introduce some works about obtaining the d-stationary solution of (P). By exploring the structure of f_2, i.e., f_2 is the supremum of finitely many convex smooth functions, Pang et al. (2017) proposed a novel enhanced DC algorithm (EDCA) to obtain a d-stationary solution. Later, Lu et al. (2019) developed an inexact variant of EDCA in order to make EDCA cheaper at each iteration. Furthermore, considering additionally the same structure of f_1 as that in Wen et al. (2018), a proximal version of EDCA was developed, whose cost in each iteration is lower than the inexact version, besides, this algorithm incorporated the Nesterov's extrapolation for possible acceleration.

In this paper, based on the inertial DC algorithm (InDCA) (de Oliveira and Tcheou 2019), we propose a refined version (RInDCA) equipped with larger inertial step-size. Indeed, both InDCA and RInDCA can handle exactness and inexactness in the solution of the convex subproblems. The basic idea of the exact versions of InDCA and RInDCA (denoted by InDCA_e and RInDCA_e) are described as follows.

[1] Let $\mu > 0$ and $i = 0, 1$, ${}^{\mu}f_i$ is the Moreau envelope of f_i with parameter μ, defined as ${}^{\mu}f_i(\mathbf{x}) := \min\{f_i(\mathbf{y}) + \frac{1}{2\mu}\|\mathbf{x} - \mathbf{y}\|^2 : \mathbf{y} \in \mathbb{R}^n\}$.

RInDCA$_e$ obtains at the k-th iteration the next iterate $\mathbf{x}^{k+1}$ by solving the convex subproblem

$$\mathbf{x}^{k+1} \in \operatorname{argmin} \{f_1(\mathbf{x}) - \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}\rangle : \mathbf{x} \in \mathbb{R}^n\},$$

where $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$. Our analysis shows that the inertial step-size $\gamma \in [0, (\sigma_1 + \sigma_2)/2)$ is adequate for the convergence of RInDCA$_e$ compared with $[0, \sigma_2/2)$ in InDCA$_e$, where σ_1 and σ_2 are the strong convex parameters of f_1 and f_2. Thus, the inertial range of RInDCA$_e$ is larger than that of InDCA$_e$ when $\sigma_1, \sigma_2 > 0$. Empirically, larger step-size accelerates the convergence. In some applications, we will encounter the case where f_1 is strongly convex and f_2 is not, then RInDCA$_e$ is applicable but not InDCA$_e$. Later, the details about the exact and inexact versions can be found in Sect. 3. The benefit of larger inertial step-size is discussed in Sect. 6.

Our *contributions* are: (1) propose a refined inertial DC algorithm, which is equipped with larger inertial step-size compared with InDCA (de Oliveira and Tcheou 2019). Besides, the relation between the inertial-type DCA and the classical DCA is pointed out. (2) establish the subsequential convergence of our refined version and the sequential convergence by further assuming the Kurdyka-Łojasiewicz (KL) property of the objective function.

The rest of the paper is organized as follows: Sect. 2 reviews some notations and preliminaries in convex and variational analysis. In Sect. 3, we first introduce our refined inertial DC algorithm (RInDCA), followed by some DC programming applications. Next, the relation between the inertial-type DCA and the successive DCA is discussed. Section 4 focuses on establishing the subsequential convergence of RInDCA. By assuming additionally the KL property of the objective function, we prove in Sect. 5 the sequential convergence of RInDCA. Section 6 summarizes the numerical results on checking copositivity of matrices and image denoising problem, in which the benefit of enlarged inertial step-size is demonstrated. Some concluding remarks are given in the final section.

2 Notations and preliminaries

Let $\mathbb{R}^n$ denote the finite dimensional vector space endowed with the canonical inner product $\langle \cdot, \cdot \rangle$, and the Euclidean norm $\| \cdot \|$, i.e., $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The entry of a vector $\mathbf{x}$ is denoted as $\mathbf{x}_i$, and the entry of a matrix $\mathbf{A}$ is denoted as $\mathbf{A}_{i,j}$.

For an extended real-valued function $h : \mathbb{R}^n \to (-\infty, \infty]$, the set

$$\operatorname{dom}(h) := \{\mathbf{x} \in \mathbb{R}^n : h(\mathbf{x}) < \infty\}$$

denotes its effective domain. If $\operatorname{dom}(h) \neq \emptyset$, and h does not attain the value $-\infty$, then h is called a proper function. The set

$$\operatorname{epi}(h) := \{(\mathbf{x}, t) : h(\mathbf{x}) \leq t, \mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R}\}$$

denotes the epigraph of h, and h is closed (resp. convex) if $\operatorname{epi}(h)$ is closed (resp. convex). The conjugate function of h is defined as

$$h^*(\mathbf{y}) = \sup\{\langle \mathbf{y}, \mathbf{x}\rangle - h(\mathbf{x}) \,:\, \mathbf{x} \in \mathbb{R}^n\}, \quad \mathbf{y} \in \mathbb{R}^n.$$

Given a proper closed function $h : \mathbb{R}^n \to (-\infty, \infty]$, the Fréchet subdifferential of h at $\mathbf{x} \in \mathrm{dom}(h)$ is given by

$$\partial^F h(\mathbf{x}) := \left\{ \mathbf{y} \in \mathbb{R}^n \,:\, \liminf_{\mathbf{z}\to\mathbf{x}} \frac{h(\mathbf{z}) - h(\mathbf{x}) - \langle \mathbf{y}, \mathbf{z} - \mathbf{x}\rangle}{\|\mathbf{z} - \mathbf{x}\|} \geq 0 \right\},$$

while for $\mathbf{x} \notin \mathrm{dom} h$, $\partial^F h(\mathbf{x}) := \emptyset$. The (limiting) subdifferential of f at $\mathbf{x} \in \mathrm{dom}(h)$ is defined as

$$\partial h(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n : \exists(\mathbf{x}^k \to \mathbf{x}, h(\mathbf{x}^k) \to h(\mathbf{x}), \mathbf{y}^k \in \partial^F h(\mathbf{x}^k)) \text{ such that } \mathbf{y}^k \to \mathbf{y}\},$$

and $\partial h(\mathbf{x}) := \emptyset$ if $\mathbf{x} \notin \mathrm{dom}(h)$. Note that if h is also convex, then the Fréchet subdifferential and the limiting subdifferential will coincide with the convex subdifferential, that is,

$$\partial^F h(\mathbf{x}) = \partial h(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n \,:\, h(\mathbf{z}) \geq h(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x}\rangle \text{ for all } \mathbf{z} \in \mathbb{R}^n\}.$$

Given a nonempty set $\mathcal{C} \subseteq \mathbb{R}^n$, the distance from a point $\mathbf{x} \in \mathbb{R}^n$ to $\mathcal{C}$ is denoted as $\mathrm{dist}(\mathbf{x}, \mathcal{C}) := \inf\{\|\mathbf{x} - \mathbf{z}\| \,:\, \mathbf{z} \in \mathcal{C}\}$. We now recall the Kurdyka-Łojasiewicz (KL) property (Bolte et al. 2007; Attouch and Bolte 2009; Attouch et al. 2013). For $\eta \in (0, \infty]$, we denote by Ξ_η the set of all concave continuous functions $\varphi : [0, \eta) \to [0, \infty)$ that are continuously differentiable over $(0, \eta)$ with positive derivatives and satisfy $\varphi(0) = 0$.

Definition 2.1 (KL property) A proper closed function h is said to satisfy the KL property at $\bar{\mathbf{x}} \in \mathrm{dom}(\partial h) := \{\mathbf{x} \in \mathbb{R}^n \,:\, \partial h(\mathbf{x}) \neq \emptyset\}$ if there exist $\eta \in (0, \infty]$, a neighborhood U of $\bar{\mathbf{x}}$, and a function $\varphi \in \Xi_\eta$ such that for all $\mathbf{x}$ in the intersection

$$U \cap \{\mathbf{x} \in \mathbb{R}^n \,:\, h(\bar{\mathbf{x}}) < h(\mathbf{x}) < h(\bar{\mathbf{x}}) + \eta\},$$

it holds that

$$\varphi'(h(\mathbf{x}) - h(\bar{\mathbf{x}}))\mathrm{dist}(\mathbf{0}, \partial h(\mathbf{x})) \geq 1.$$

If h satisfies the KL property at any point of $\mathrm{dom}(\partial h)$, then h is called a KL function.

The following uniformized KL property plays an important role in our sequential convergence analysis.

Lemma 2.1 (Uniformized KL property, see Bolte et al. (2014)) *Let $\Omega \subseteq \mathbb{R}^n$ be a compact set and let $h : \mathbb{R}^n \to (-\infty, \infty]$ be a proper closed function. If h is constant on Ω and satisfies the KL property at each point of Ω, then there exist $\varepsilon, \eta > 0$ and $\varphi \in \Xi_\eta$ such that*

$$\varphi'(h(\mathbf{x}) - h(\bar{\mathbf{x}}))\mathrm{dist}(\mathbf{0}, \partial h(\mathbf{x})) \geq 1$$

for any $\bar{\mathbf{x}} \in \Omega$ and any $\mathbf{x}$ satisfying $\mathrm{dist}(\mathbf{x}, \Omega) < \varepsilon$ and $h(\bar{\mathbf{x}}) < h(\mathbf{x}) < h(\bar{\mathbf{x}}) + \eta$.

Now, let $h : \mathbb{R}^n \to (-\infty, \infty]$ be a proper closed and convex function, and let $\varepsilon > 0$, the set

$$\partial_\varepsilon h(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^n : h(\mathbf{z}) \geq h(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle - \varepsilon \text{ for all } \mathbf{z} \in \mathbb{R}^n\}$$

denotes the ε-subdifferential of h at $\mathbf{x}$, and any point in $\partial_\varepsilon h(\mathbf{x})$ is called a ε-subgradient of h at $\mathbf{x}$. Clearly, $\partial_\varepsilon h(\mathbf{x}) \neq \emptyset$ implies that $\mathbf{x} \in \mathrm{dom}(h)$. A proper closed function h is called σ-strongly convex (cf. σ-convex) with $\sigma \geq 0$ if for any $\mathbf{x}, \mathbf{z} \in \mathrm{dom}(h)$ and $\lambda \in [0, 1]$, it holds that

$$h(\lambda \mathbf{x} + (1 - \lambda)\mathbf{z}) \leq \lambda h(\mathbf{x}) + (1 - \lambda)h(\mathbf{z}) - \frac{\sigma}{2}\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{z}\|^2.$$

Moreover, let h be a σ-convex function. Then it was known (Beck 2017) that for any $\mathbf{x} \in \mathrm{dom}(\partial h)$, $\mathbf{y} \in \partial h(\mathbf{x})$ and $\mathbf{z} \in \mathrm{dom}(h)$, we have

$$h(\mathbf{z}) \geq h(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle + \frac{\sigma}{2}\|\mathbf{z} - \mathbf{x}\|^2.$$

Finally, we give a result related to strong convexity and ε-subdifferential, which will be used in analyzing our refined inexact DC algorithm.

Lemma 2.2 (Pham et al. (2021)) *Let $h : \mathbb{R}^n \to (-\infty, \infty]$ be a σ-convex function, and let $\varepsilon \geq 0$, $t \in (0, 1]$. Then for any $\mathbf{x} \in \mathrm{dom}(h)$, $\mathbf{z} \in \mathrm{dom}(h)$, and $\mathbf{y} \in \partial_\varepsilon h(\mathbf{x})$,*

$$h(\mathbf{z}) \geq h(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle + \frac{\sigma(1 - t)}{2}\|\mathbf{z} - \mathbf{x}\|^2 - \frac{\varepsilon}{t}. \tag{1}$$

3 A refined inertial DC algorithm

In this section, we first describe a general inertial-type DC algorithm for problem (P), which is taken from de Oliveira and Tcheou (2019) except that the values of λ and γ are not specified. By setting different λ and γ we derive the classical DCA, the inertial DC algorithm (InDCA) proposed in de Oliveira and Tcheou (2019), and our refined version (RInDCA). Next, some practical applications of DC programming problems are introduced, followed by a discussion of the relation between the inertial-type DCA and the successive DCA.

From now on, suppose that the objective function f has a DC decomposition $f = f_1 - f_2$, where f_1 is σ_1-convex and f_2 is σ_2-convex. We describe below in Algorithm 1 a general inertial-type DC algorithm for (P).

Algorithm 1: A general inertial-type DC algorithm for (P)

Input: $\mathbf{x}^0 \in \text{dom}(f_1)$; $\lambda \in [0, 1)$; $\gamma \geq 0$; $\mathbf{x}^{-1} = \mathbf{x}^0$.

for $k = 0, 1, 2, \cdots$ **do**

  find $\mathbf{x}^{k+1} \in \mathbb{R}^n$ such that

$$\partial_{\varepsilon_{k+1}} f_1(\mathbf{x}^{k+1}) \cap (\partial f_2(\mathbf{x}^k) + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1})) \neq \emptyset \tag{2}$$

  with $0 \leq \varepsilon_{k+1} \leq \lambda \frac{\sigma_2}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2$.

end

Note that Algorithm 1 includes exact and inexact cases. The exact case occurs when $\lambda = 0$, then $\varepsilon_{k+1} = 0$ and $\partial_{\varepsilon_{k+1}} f_1(\mathbf{x}^{k+1}) = \partial f_1(\mathbf{x}^{k+1})$, thus the iteration point $\mathbf{x}^{k+1}$ in (2) is obtained by exactly solving the next convex subproblem:

$$\mathbf{x}^{k+1} \in \text{argmin} \{f_1(\mathbf{x}) - \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x} \rangle : \mathbf{x} \in \mathbb{R}^n\}, \tag{3}$$

where $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$. On the other hand, the inexact case occurs when $\lambda > 0$ and $\sigma_2 > 0$, then the iterate $\mathbf{x}^{k+1}$ in (2), which is an approximate minimizer of (3), could be computed by some bundle methods (see de Oliveira and Tcheou 2019). The classical DCA is derived when setting $\lambda = 0$ and $\gamma = 0$ in Algorithm 1. Next, we denote the exact (resp. inexact) versions of InDCA and RInDCA as InDCA_e and RInDCA_e (resp. InDCA_n and RInDCA_n). The following table summarizes the four algorithms which are recovered by specifying some parameters in Algorithm 1 together with the assumptions about σ_1 and σ_2.

Algorithm	λ	γ	t	Assumption
InDCA_e	0	$[0, \sigma_2/2)$	–	$\sigma_2 > 0$
InDCA_n	(0,1)	$[0, (1-\lambda)\sigma_2/2)$	–	$\sigma_2 > 0$
RInDCA_e	0	$[0, (\sigma_1 + \sigma_2)/2)$	–	$\sigma_1 + \sigma_2 > 0$
RInDCA_n	(0,1)	$[0, (\sigma_1(1-t) + \sigma_2 - \lambda\sigma_2/t)/2)$	(0,1]	$\sigma_2 > 0$

Comparing InDCA_e with RInDCA_e, the inertial range of the latter is larger than the former if $\sigma_1 > 0$. Moreover, RInDCA_e is applicable when $\sigma_2 = 0$ and $\sigma_1 > 0$, while this is not the case for InDCA_e. Next, for comparing InDCA_n and RInDCA_n, we need to optimize t in RInDCA_n such that the supremum of the inertial step-size γ is maximized. To this end, fixing $\lambda \in (0, 1)$ and let $\sigma(\lambda, t) := \sigma_1(1-t) + \sigma_2 - \lambda\sigma_2/t$, we can maximize $\sigma(\lambda, t)$ with respect to $t \in [0, 1)$. It is easy to verify that

$$\text{argmax} \{\sigma(\lambda, t) : t \in (0, 1]\} = \begin{cases} 1 & \text{if } \sigma_1 = 0, \\ 1 & \text{if } \sigma_1 \neq 0, \sqrt{\lambda\sigma_2/\sigma_1} \geq 1, \\ \sqrt{\lambda\sigma_2/\sigma_1} & \text{if } \sigma_1 \neq 0, \sqrt{\lambda\sigma_2/\sigma_1} < 1. \end{cases}$$

For the first two cases, $\sigma(\lambda, 1) = (1-\lambda)\sigma_2$, this implies the same inertial range between InDCA_n and RInDCA_n. Next, for the last case, we have $\sigma(\lambda, 1) < \sigma(\lambda, \sqrt{\lambda\sigma_2/\sigma_1})$, thus the inertial range of RInDCA_n is larger than that in

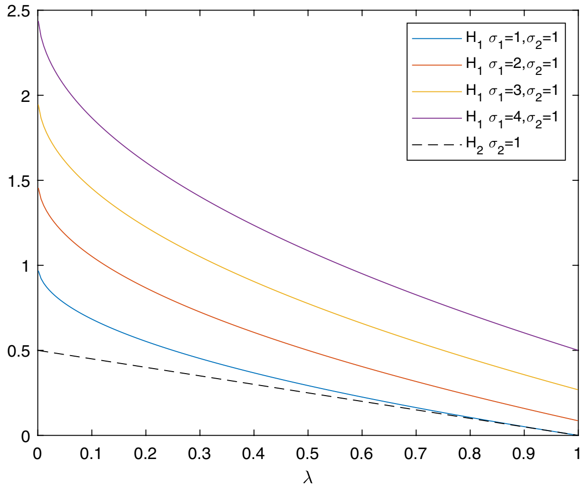

Fig. 1 The values of H_1 with $(\sigma_1, \sigma_2) \in \{(1,1),(2,1),(3,1),(4,1)\}$ and H_2 with respect to $\lambda \in (0,1)$

InDCA$_\text{n}$. Now, let $H_1(\lambda) := \sigma(\lambda, \sqrt{\lambda\sigma_2/\sigma_1})/2$ and $H_2(\lambda) := (1-\lambda)\sigma_2/2$, we visualize in Fig. 1 the values of H_1 for $(\sigma_1, \sigma_2) \in \{(1,1),(2,1),(3,1),(4,1)\}$ as well as the values of H_2 for $\sigma_1 = 1$ with respect to $\lambda \in (0,1)$ respectively.

It can be observed that when fixing $\lambda \in (0,1)$ and $\sigma_2 = 1$, the value of $H_1(\lambda)$ is getting larger as σ_1 increases, moreover, all of these values are larger than $H_2(\lambda)$. This observation implies that a larger strong convexity parameter of f_1 will result in a larger inertial range.

Now, we introduce some DC programming problems, in which the first DC components are strongly convex.

Example 3.1 (Convex constraint quadratic problem) Consider the quadratic problem with convex constraint

$$\min\left\{f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top\mathbf{A}\mathbf{x} + \delta_Y(\mathbf{x}) \,:\, \mathbf{x} \in \mathbb{R}^n\right\},$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric and Y is a closed convex set. This model has a DC decomposition $f = f_1 - f_2$ with $f_1(\mathbf{x}) = \frac{L}{2}\|\mathbf{x}\|^2 + \delta_Y(\mathbf{x})$ and $f_2(\mathbf{x}) = \frac{L}{2}\|\mathbf{x}\|^2 - \frac{1}{2}\mathbf{x}^\top\mathbf{A}\mathbf{x}$, where $L \geq \|\mathbf{A}\|$[2]. When Y is the nonnegative orthant $\mathbb{R}^n_+$, DCA applied for the above DC decomposition has been used for checking copositivity of matrices (Dür and Hiriart-Urruty 2013). Note that for this DC decomposition, the first DC component is L-convex.

Example 3.2 (1D signals denoising) Consider the nonconvex sparse signal recovery problem for 1D signals:

[2] $\|\mathbf{A}\|$ is the spectral norm of $\mathbf{A}$.

$$\min\left\{\frac{\mu}{2}\|\mathbf{x}-\mathbf{b}\|^2+\sum_{i=1}^{n-1}\phi(|\mathbf{x}_{i+1}-\mathbf{x}_i|) : \mathbf{x}\in\mathbb{R}^n\right\},$$

where $\mu>0$ is a trade-off constant, $\mathbf{b}$ is the noise signal in $\mathbb{R}^n$, and ϕ is a concave function for inducing sparsity, e.g., $\phi(r):=\log(1+2r)/r$. In this model, taking $\phi(r)=\log(1+2r)/r$, we obtain a DC decomposition as f_1-f_2 with $f_1(\mathbf{x})=\frac{\mu}{2}\|\mathbf{x}-\mathbf{b}\|^2+\sum_{i=1}^{n-1}|\mathbf{x}_{i+1}-\mathbf{x}_i|$ and $f_2=\sum_{i=1}^{n-1}|\mathbf{x}_{i+1}-\mathbf{x}_i|-\sum_{i=1}^{n-1}\phi(|\mathbf{x}_{i+1}-\mathbf{x}_i|)$. Note that f_1 is strongly convex, while f_2 is not. Thus InDCA can not be directly applied. In order to apply InDCA, a strong convexity term $\psi : \mathbf{x}\mapsto \rho\|\mathbf{x}\|^2/2$ is added into both f_1 and f_2 in de Oliveira and Tcheou (2019), yielding the new DC decomposition $(f_1+\psi)-(f_2+\psi)$, where the DC components are both strongly convex.

Remark 3.1 As shown in Examples 3.1 and 3.2, the first DC components are strongly convex. However, InDCA only takes into account the strong convexity of the second DC component. This often causes the inertial range smaller than that of our refined version RInDCA, which involves the strong convexity of both DC components.

Before ending this section, we mention out that the inertial-type DCA (both InDCA_e and RInDCA_e) is related to the successive DCA, namely SDCA (see Le Thi and Pham 2018). Let us take at the k-th iteration of DCA the successive DC decomposition $f=(f_1+\varphi^k)-(f_2+\varphi^k)$ with $\varphi^k(\mathbf{x})=\frac{\gamma}{2}\|\mathbf{x}-\mathbf{x}^{k-1}\|^2$, then DCA applied to this special DC decomposition yields

$$\mathbf{x}^{k+1}\in \operatorname{argmin}\left\{f_1(\mathbf{x})+\varphi^k(\mathbf{x})-\langle\nabla\varphi^k(\mathbf{x}^k)+\mathbf{y}^k,\mathbf{x}\rangle : \mathbf{x}\in\mathbb{R}^n\right\}, \tag{4}$$

where $\mathbf{y}^k\in\partial f_2(\mathbf{x}^k)$. Then (4) reads as

$$\mathbf{x}^{k+1}\in \operatorname{argmin}\left\{f_1(\mathbf{x})+\varphi^k(\mathbf{x})-\langle\mathbf{y}^k+\gamma(\mathbf{x}^k-\mathbf{x}^{k-1}),\mathbf{x}\rangle : \mathbf{x}\in\mathbb{R}^n\right\}. \tag{5}$$

Comparing SDCA described in (5) with RInDCA_e (or InDCA_e) in (2), we observe that the only difference is the presence of the proximal term $\varphi^k(\mathbf{x})$ in (5). This term plays a role of regularizer which can be interpreted as finding a point $\mathbf{x}^{k+1}$ close to $\mathbf{x}^{k-1}$ and a minimizer of the function $\mathbf{x}\mapsto f_1(\mathbf{x})-\langle\mathbf{y}^k+\gamma(\mathbf{x}^k-\mathbf{x}^{k-1}),\mathbf{x}\rangle$. This is the basic idea of proximal point methods, thus this SDCA is often called proximal DCA as well. The absence of φ^k in RInDCA_e means that the iteration point $\mathbf{x}^{k+1}$ is not supposed to be close to $\mathbf{x}^{k-1}$, which will lead to potential acceleration. In conclusion, the term φ^k plays a key role to make RInDCA_e differ from SDCA, and possibly yields acceleration in RInDCA_e.

Remark 3.2 Indeed, the convex problem of (5) is irrelevant to $\mathbf{x}^{k-1}$ since

$$(5) = \operatorname{argmin}\left\{f_1(\mathbf{x})+\frac{\gamma}{2}\|\mathbf{x}\|^2-\langle\mathbf{y}^k+\gamma\mathbf{x}^k,\mathbf{x}\rangle : \mathbf{x}\in\mathbb{R}^n\right\},$$

which is exactly the subproblem of DCA for the fixed DC decomposition

$$f = \left(f_1 + \frac{\gamma}{2}\|\cdot\|^2\right) - \left(f_2 + \frac{\gamma}{2}\|\cdot\|^2\right).$$

4 Subsequential convergence analysis

This section focuses on showing the subsequential convergence of RInDCA (both RInDCA_e and RInDCA_n), some results for further developing the sequential convergence of RInDCA in the next section are also proved.

4.1 Subsequential convergence of RInDCA_e

The following assumption is made in this subsection.

Assumption 2 Suppose that the DC components f_1 and f_2 are respectively σ_1-convex and σ_2-convex with $\sigma_1 + \sigma_2 > 0$.

Now, we start to establish the subsequential convergence of RInDCA_e.

Lemma 4.1 *Let* $\{\mathbf{x}^k\}$ *be the sequence generated by* RInDCA_e *and* $\gamma \in [0, (\sigma_1 + \sigma_2)/2)$, *then the sequence* $\{f(\mathbf{x}^k) + \frac{\sigma_1+\sigma_2-\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2\}$ *is nonincreasing and for all* $k \geq 0$,

$$f(\mathbf{x}^{k+1}) + \frac{\sigma_1 + \sigma_2 - \gamma}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \leq f(\mathbf{x}^k) + \frac{\sigma_1 + \sigma_2 - \gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 - \frac{\sigma_1 + \sigma_2 - 2\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2. \tag{6}$$

Proof Because $\mathbf{x}^{k+1}$ satisfies (3), thus $\mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}) \in \partial f_1(\mathbf{x}^{k+1})$, then by σ_1-convexity of f_1, we have

$$f_1(\mathbf{x}^k) \geq f_1(\mathbf{x}^{k+1}) + \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle + \frac{\sigma_1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \tag{7}$$

On the other hand, it follows from $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$ and σ_2-convexity of f_2 that

$$f_2(\mathbf{x}^{k+1}) \geq f_2(\mathbf{x}^k) + \langle \mathbf{y}^k, \mathbf{x}^{k+1} - \mathbf{x}^k\rangle + \frac{\sigma_2}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \tag{8}$$

Summing (7) to (8) and reshuffling the terms, we derive that

$$f(\mathbf{x}^k) \geq f(\mathbf{x}^{k+1}) + \gamma\langle \mathbf{x}^k - \mathbf{x}^{k-1}, \mathbf{x}^k - \mathbf{x}^{k+1}\rangle + \frac{\sigma_1 + \sigma_2}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \tag{9}$$

By applying $\langle \mathbf{x}^k - \mathbf{x}^{k-1}, \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \geq -(\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 + \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2)/2$ to (9), we obtain

$$f(\mathbf{x}^{k+1}) + \frac{\sigma_1+\sigma_2-\gamma}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2 \le f(\mathbf{x}^k) + \frac{\sigma_1+\sigma_2-\gamma}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 - \frac{\sigma_1+\sigma_2-2\gamma}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2.$$

Moreover, we get from $\gamma < (\sigma_1+\sigma_2)/2$ that $(\sigma_1+\sigma_2-2\gamma)/2 > 0$, thus the sequence $\{f(\mathbf{x}^k) + \frac{\sigma_1+\sigma_2-\gamma}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2\}$ is nonincreasing. ☐

Theorem 4.1 *Let* $\{\mathbf{x}^k\}$ *be the sequence generated by RInDCA*$_\mathrm{e}$ *and* $\gamma \in [0, (\sigma_1+\sigma_2)/2)$. *Then the following statements hold*:

(i) $\lim\limits_{k\to\infty} \|\mathbf{x}^k - \mathbf{x}^{k-1}\| = 0$;
(ii) *If* $\{\mathbf{x}^k\}$ *is bounded, then any limit point of* $\{\mathbf{x}^k\}$ *is a critical point of* (P).

Proof (*i*) Let us denote $\eta_1 := (\sigma_1+\sigma_2-\gamma)/2$ and $\eta_2 := (\sigma_1+\sigma_2-2\gamma)/2$. Clearly, $\eta_1 > 0$, $\eta_2 > 0$, and the inequality (6) reads as

$$f(\mathbf{x}^{k+1}) + \eta_1\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2 \le f(\mathbf{x}^k) + \eta_1\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 - \eta_2\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2. \quad (10)$$

By summing (10) from $k=0$ to n, we obtain that

$$\eta_2 \sum_{k=0}^{n} \|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 \le f(\mathbf{x}^0) - \left(f(\mathbf{x}^{n+1}) + \eta_1\|\mathbf{x}^{n+1}-\mathbf{x}^n\|^2\right). \quad (11)$$

Recalling that f is bounded below by a scalar f^* (term (b) of Assumption 1), then we have

$$f(\mathbf{x}^{n+1}) + \eta_1\|\mathbf{x}^{n+1}-\mathbf{x}^n\|^2 \ge f^*,$$

thus we obtain

$$\sum_{k=0}^{n} \|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 \le \frac{f(\mathbf{x}^0)-f^*}{\eta_2}, \quad \forall n \in \mathbb{N}.$$

Thus, $\sum_{k=0}^{\infty}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2 < \infty$ which implies $\lim\limits_{k\to\infty}\|\mathbf{x}^k-\mathbf{x}^{k-1}\| = 0$. (*ii*) The boundedness of $\{\mathbf{x}^k\}$ together with term (a) of Assumption 1 implies the boundedness of $\{\mathbf{y}^k\}$ (see Theorem 3.16 in Beck (2017)). Let $\bar{\mathbf{x}}$ be any limit point of $\{\mathbf{x}^k\}$, then there exists a convergent subsequence such that $\lim\limits_{i\to\infty}\mathbf{x}^{k_i} = \bar{\mathbf{x}}$. Moreover, since $\{\mathbf{y}^{k_i}\}$ is bounded in $\mathbb{R}^n$, there exists a convergent subsequence of $\{\mathbf{y}^{k_i}\}$. Without loss of generality, we can assume that the sequence $\{\mathbf{y}^{k_i}\}$ is convergent and $\lim\limits_{i\to\infty}\mathbf{y}^{k_i} = \bar{\mathbf{y}}$. Then taking into account that $\mathbf{y}^{k_i} \in \partial f_2(\mathbf{x}^{k_i})$ for all i, $\mathbf{y}^{k_i} \to \bar{\mathbf{y}}$, $\mathbf{x}^{k_i} \to \bar{\mathbf{x}}$, and the closedness of the graph ∂f_2 (see Theorem 24.4 in Rockafellar (1970)), we obtain

$$\bar{\mathbf{y}} \in \partial f_2(\bar{\mathbf{x}}). \quad (12)$$

On the other hand, the next relation holds for all i

$$\mathbf{y}^{k_i} + \gamma(\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}) \in \partial f_1(\mathbf{x}^{k_i+1}).$$

By the fact that $\lim_{i\to\infty} \mathbf{x}^{k_i+1} = \bar{\mathbf{x}}$ since $\lim_{i\to\infty} \|\mathbf{x}^{k_i+1} - \mathbf{x}^{k_i}\| = 0$, then combining $\mathbf{y}^{k_i} \to \bar{\mathbf{y}}$, $\gamma(\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}) \to 0$, $\mathbf{x}^{k_i+1} \to \bar{\mathbf{x}}$, and the closedness of the graph ∂f_1, we have

$$\bar{\mathbf{y}} \in \partial f_1(\bar{\mathbf{x}}). \tag{13}$$

We conclude from (12) and (13) that $\bar{\mathbf{y}} \in \partial f_1(\bar{\mathbf{x}}) \cap \partial f_2(\bar{\mathbf{x}}) \neq \emptyset$, thus $\bar{\mathbf{x}}$ is a critical point of problem (P). □

4.2 Subsequential convergence of RInDCA$_\text{n}$

The following assumption is made in this part.

Assumption 3 Suppose that the second DC component f_2 is σ_2-convex with $\sigma_2 > 0$.

Now, we start to establish the subsequential convergence of RInDCA$_\text{n}$.

Lemma 4.2 *Let* $\{\mathbf{x}^k\}$ *be the sequence generated by RInDCA*$_\text{n}$ *and* $\gamma \in [0, \bar{\sigma}/2)$, *where* $\bar{\sigma} = \sigma_1(1-t) + \sigma_2 - \lambda\sigma_2/t$ *with* $\lambda \in (0,1)$, $t \in (0,1]$. *Then the sequence* $\{f(\mathbf{x}^k) + \frac{\bar{\sigma}-\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2\}$ *is nonincreasing and for all* $k \geq 0$, *we have*

$$\begin{aligned} f(\mathbf{x}^{k+1}) + \frac{\bar{\sigma}-\gamma}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \leq & f(\mathbf{x}^k) + \frac{\bar{\sigma}-\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \\ & - \frac{\bar{\sigma}-2\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2. \end{aligned} \tag{14}$$

Proof Because $\mathbf{x}^{k+1}$ satisfies (2), let $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$ such that $\mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}) \in \partial_{\varepsilon^{k+1}} f_1(\mathbf{x}^{k+1})$, we have

$$\begin{aligned} f_1(\mathbf{x}^k) \overset{(1)}{\geq} & f_1(\mathbf{x}^{k+1}) + \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \\ & + \frac{\sigma_1(1-t)}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - \frac{\varepsilon^{k+1}}{t} \\ \geq & f_1(\mathbf{x}^{k+1}) + \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \\ & + \frac{\sigma_1(1-t) - \lambda\sigma_2/t}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2, \end{aligned} \tag{15}$$

where the second inequality is implied by the fact $\varepsilon^{k+1} \leq \lambda\frac{\sigma_2}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2$. On the other hand, $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$ and σ_2-convexity of f_2 imply that

$$f_2(\mathbf{x}^{k+1}) \geq f_2(\mathbf{x}^k) + \langle \mathbf{y}^k, \mathbf{x}^{k+1} - \mathbf{x}^k\rangle + \frac{\sigma_2}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \tag{16}$$

Summing (15) to (16) and reshuffling the terms, we obtain that

$$f(\mathbf{x}^k) \geq f(\mathbf{x}^{k+1}) + \gamma\langle \mathbf{x}^k - \mathbf{x}^{k-1}, \mathbf{x}^k - \mathbf{x}^{k+1}\rangle + \frac{\bar{\sigma}}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \tag{17}$$

By applying $\langle \mathbf{x}^k - \mathbf{x}^{k-1}, \mathbf{x}^k - \mathbf{x}^{k+1}\rangle \geq -(\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 + \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2)/2$ to (17), we obtain (14). Moreover, $\gamma \in [0, \bar{\sigma}/2)$ implies $(\bar{\sigma} - 2\gamma)/2 > 0$, thus the sequence $\{f(\mathbf{x}^k) + \frac{\bar{\sigma}-\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2\}$ is nonincreasing. □

The proof of the next theorem is omitted since it follows from the similar arguments as in Theorem 4.1.

Theorem 4.2 *Let $\{\mathbf{x}^k\}$ be the sequence generated by RInDCA$_\text{n}$ and $\gamma \in [0, \bar{\sigma}/2)$, where $\bar{\sigma} = \sigma_1(1-t) + \sigma_2 - \lambda\sigma_2/t$ with $\lambda \in (0,1)$, $t \in (0,1]$. Then the following statements hold*:

(i) $\lim_{k\to\infty} \|\mathbf{x}^k - \mathbf{x}^{k-1}\| = 0$;
(ii) *If $\{\mathbf{x}^k\}$ is bounded, then any limit point of $\{\mathbf{x}^k\}$ is a critical point of (P).*

4.3 Some results on the sequences obtained by RInDCA

In this subsection, we provide some important results on the sequences obtained by RInDCA$_\text{e}$ and RInDCA$_\text{n}$.

Proposition 4.1 *Suppose that the assumptions in Theorem 4.1 hold. Let $\{\mathbf{x}^k\}$ be the sequence generated by RInDCA$_\text{e}$ and let Ω_1 be the set of the limit points of $\{\mathbf{x}^k\}$. If $\{\mathbf{x}^k\}$ is bounded, then we have*

(i) $\lim_{k\to\infty} f(\mathbf{x}^k) := \zeta$ *exists*;
(ii) $f \equiv \zeta$ *on* Ω_1.

Proof (i) Lemma 4.1 implies that the sequence $\{f(\mathbf{x}^k) + \frac{\sigma_1+\sigma_2-\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2\}$ is nonincreasing, combining the fact that this sequence is bounded below, thus it converges. On the other hand, (ii) of Theorem 4.1 implies that $\lim_{k\to\infty} \|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 = 0$. Therefore, $\lim_{k\to\infty} f(\mathbf{x}^k) := \zeta$ exists. (ii) Because Ω_1 is nonempty, thus given any $\bar{\mathbf{x}} \in \Omega_1$, there exists a subsequence $\{\mathbf{x}^{k_i}\}$ such that $\lim_{i\to\infty} \mathbf{x}^{k_i} = \bar{\mathbf{x}}$. Then, we get from $\mathbf{y}^{k_i} \in \partial f_2(\mathbf{x}^{k_i})$ and σ_2-convexity of f_2 that

$$f_2(\mathbf{x}^{k_i+1}) \geq f_2(\mathbf{x}^{k_i}) + \langle \mathbf{y}^{k_i}, \mathbf{x}^{k_i+1} - \mathbf{x}^{k_i}\rangle + \frac{\sigma_2}{2}\|\mathbf{x}^{k_i+1} - \mathbf{x}^{k_i}\|^2. \tag{18}$$

Similarly, we obtain as well from $\mathbf{y}^{k_i} + \gamma(\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}) \in \partial f_1(\mathbf{x}^{k_i+1})$ and σ_1-convexity of f_1 that

$$f_1(\bar{\mathbf{x}}) \geq f_1(\mathbf{x}^{k_i+1}) + \langle \mathbf{y}^{k_i} + \gamma(\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}), \bar{\mathbf{x}} - \mathbf{x}^{k_i+1} \rangle + \frac{\sigma_1}{2}\|\bar{\mathbf{x}} - \mathbf{x}^{k_i+1}\|^2. \quad (19)$$

Then we have

$$\begin{aligned}
\zeta &= \lim_{k\to\infty} f(\mathbf{x}^k) \\
&= \lim_{i\to\infty} f_1(\mathbf{x}^{k_i+1}) - f_2(\mathbf{x}^{k_i+1}) \\
&\overset{(18)}{\leq} \limsup_{i\to\infty} f_1(\mathbf{x}^{k_i+1}) - \{f_2(\mathbf{x}^{k_i}) + \langle \mathbf{y}^{k_i}, \mathbf{x}^{k_i+1} - \mathbf{x}^{k_i} \rangle + \frac{\sigma_2}{2}\|\mathbf{x}^{k_i+1} - \mathbf{x}^{k_i}\|^2\} \\
&\overset{(19)}{\leq} \limsup_{i\to\infty} f_1(\bar{\mathbf{x}}) - \langle \mathbf{y}^{k_i} + \gamma(\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}), \bar{\mathbf{x}} - \mathbf{x}^{k_i+1} \rangle - \frac{\sigma_1}{2}\|\mathbf{x}^{k_i+1} - \bar{\mathbf{x}}\|^2 \\
&\quad - f_2(\mathbf{x}^{k_i}) - \langle \mathbf{y}^{k_i}, \mathbf{x}^{k_i+1} - \mathbf{x}^{k_i} \rangle - \frac{\sigma_2}{2}\|\mathbf{x}^{k_i+1} - \mathbf{x}^{k_i}\|^2 \\
&= f(\bar{\mathbf{x}}),
\end{aligned}$$

where the last equality follows that $\lim_{i\to\infty}\|\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}\| = 0$, f_2 is continuous over $\mathrm{dom} f_1$, and $\{\mathbf{y}^{k_i}\}$ is bounded. On the other hand, because f is a closed function, then we have

$$\begin{aligned}
f(\bar{\mathbf{x}}) &= f_1(\bar{\mathbf{x}}) - f_2(\bar{\mathbf{x}}) \\
&\leq \liminf_{i\to\infty} f_1(\mathbf{x}^{k_i}) - f_2(\mathbf{x}^{k_i}) \\
&= \liminf_{i\to\infty} f(\mathbf{x}^{k_i}) = \zeta.
\end{aligned}$$

Thus, we obtain that $f(\bar{\mathbf{x}}) = \zeta$ and conclude that $f \equiv \zeta$ on Ω_1. □

We can obtain a similar proposition for $\mathrm{RInDCA_n}$ described as follows whose proof is omitted.

Proposition 4.2 *Suppose that the assumptions in Theorem 4.2 hold. Let $\{\mathbf{x}^k\}$ be the sequence generated by $\mathrm{RInDCA_n}$ and let Ω_2 be the set of the limit points of $\{\mathbf{x}^k\}$. If $\{\mathbf{x}^k\}$ is bounded, then we have*

(i) $\lim_{k\to\infty} f(\mathbf{x}^k) := \widetilde{\zeta}$ *exists*;
(ii) $f \equiv \widetilde{\zeta}$ *on* Ω_2.

5 Sequential convergence analysis

In this section, we will establish the sequential convergence of RInDCA. Let us consider the next auxiliary functions:

$$E(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f_1(\mathbf{x}) - \langle \mathbf{x}, \mathbf{y} \rangle + f_2^*(\mathbf{y}) + \frac{\sigma_1 - \gamma}{2} \|\mathbf{x} - \mathbf{z}\|^2, \tag{20}$$

$$\widetilde{E}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f_1(\mathbf{x}) - \langle \mathbf{x}, \mathbf{y} \rangle + f_2^*(\mathbf{y}) + \frac{\bar{\sigma} - \sigma_2 - \gamma}{2} \|\mathbf{x} - \mathbf{z}\|^2, \tag{21}$$

where f_1 and f_2 are defined in problem (P) with the strong convexity parameters σ_1 and σ_2, f_2^* is the conjugate function of f_2, and $\bar{\sigma}$ is defined in Lemma 4.2. Note that our construction of E and $\widetilde{E}$ is motivated by Liu et al. (2019), and our sequential convergence analysis follows similar arguments as in Attouch et al. (2013); Liu et al. (2019).

First, the following Proposition 5.1 establishes the non-increasing property of the sequence $\{E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})\}_{k \geq 1}$.

Proposition 5.1 *Suppose that the assumptions in Theorem 4.1 hold. Let E be defined as in (20), $\{\mathbf{x}^k\}$ be the sequence generated by RInDCA$_\mathrm{e}$, and $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$ for all $k \geq 0$. Then we have*

(i) *For any $k \geq 1$,*

$$E(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{x}^k) \leq E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}) - \frac{\sigma_1 + \sigma_2 - 2\gamma}{2} \|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2; \tag{22}$$

(ii) *If $\{\mathbf{x}^k\}$ is bounded, let Υ be the set of all limit points of the sequence $\{(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})\}$. Then Υ is a nonempty compact set,*

$$\lim_{k \to \infty} E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}) := \zeta$$

exists, and $E \equiv \zeta$ on Υ.

Proof (*i*) For any $k \geq 1$, $\mathbf{y}^{k-1} \in \partial f_2(\mathbf{x}^{k-1})$ together with σ_2-convexity of f_2 implies that

$$f_2(\mathbf{x}^k) \geq f_2(\mathbf{x}^{k-1}) + \langle \mathbf{y}^{k-1}, \mathbf{x}^k - \mathbf{x}^{k-1} \rangle + \frac{\sigma_2}{2} \|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2. \tag{23}$$

Thus we have

$$\begin{aligned} \langle \mathbf{y}^{k-1}, \mathbf{x}^k \rangle - f_2(\mathbf{x}^k) + \frac{\sigma_2}{2} \|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 &\overset{(23)}{\leq} \langle \mathbf{y}^{k-1}, \mathbf{x}^{k-1} \rangle - f_2(\mathbf{x}^{k-1}) \\ &= f_2^*(\mathbf{y}^{k-1}), \end{aligned} \tag{24}$$

where the equality is implied by $\mathbf{y}^{k-1} \in \partial f_2(\mathbf{x}^{k-1})$, and thus $\langle \mathbf{y}^{k-1}, \mathbf{x}^{k-1} \rangle = f_2(\mathbf{x}^{k-1}) + f_2^*(\mathbf{y}^{k-1})$ (see Theorem 4.20 in Beck (2017)). Moreover, $\mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}) \in \partial f_1(\mathbf{x}^{k+1})$ and σ_1-convexity of f_1 imply that

$$f_1(\mathbf{x}^k) \geq f_1(\mathbf{x}^{k+1}) + \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle + \frac{\sigma_1}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2. \quad (25)$$

Next, we have

$$\begin{aligned}
& E(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{x}^k) \\
& = f_1(\mathbf{x}^{k+1}) - \langle \mathbf{x}^{k+1}, \mathbf{y}^k \rangle + f_2^*(\mathbf{y}^k) + \frac{\sigma_1 - \gamma}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& \overset{(25)}{\leq} f_1(\mathbf{x}^k) - \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle - \frac{\sigma_1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& \quad - \langle \mathbf{x}^{k+1}, \mathbf{y}^k \rangle + f_2^*(\mathbf{y}^k) + \frac{\sigma_1 - \gamma}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& = f_1(\mathbf{x}^k) - \langle \mathbf{y}^k + \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle - \frac{\sigma_1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& \quad - \langle \mathbf{x}^{k+1}, \mathbf{y}^k \rangle + \langle \mathbf{y}^k, \mathbf{x}^k \rangle - f_2(\mathbf{x}^k) + \frac{\sigma_1 - \gamma}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& = f_1(\mathbf{x}^k) - f_2(\mathbf{x}^k) - \langle \gamma(\mathbf{x}^k - \mathbf{x}^{k-1}), \mathbf{x}^k - \mathbf{x}^{k+1} \rangle - \frac{\gamma}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\
& \leq f_1(\mathbf{x}^k) - f_2(\mathbf{x}^k) + \frac{\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \\
& \overset{(24)}{\leq} f_1(\mathbf{x}^k) - \langle \mathbf{x}^k, \mathbf{y}^{k-1} \rangle + f_2^*(\mathbf{y}^{k-1}) + \frac{\gamma - \sigma_2}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \\
& = E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}) - \frac{\sigma_1 + \sigma_2 - 2\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2,
\end{aligned}$$

where the second equality is implied by $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$, and thus $f_2(\mathbf{x}^k) + f_2^*(\mathbf{y}^k) = \langle \mathbf{x}^k, \mathbf{y}^k \rangle$; the second inequality follows from $\langle \mathbf{x}^k - \mathbf{x}^{k-1}, \mathbf{x}^k - \mathbf{x}^{k+1} \rangle \geq -(\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 + \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2)/2$.

(*ii*) The boundedness of $\{\mathbf{x}^k\}$ implies the boundedness of $\{\mathbf{y}^{k-1}\}$. Thus, the sequence $\{(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})\}$ is bounded, yielding that Υ is a nonempty compact set. Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Upsilon$, it is easy to see that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{z}}) \in \Upsilon$ implies $\bar{\mathbf{x}} = \bar{\mathbf{z}} \in \Omega_1$. For any $k \geq 1$, we have

$$\begin{aligned}
E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}) & = f_1(\mathbf{x}^k) - \langle \mathbf{x}^k, \mathbf{y}^{k-1} \rangle + f_2^*(\mathbf{y}^{k-1}) + \frac{\sigma_1 - \gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \\
& \geq f_1(\mathbf{x}^k) - f_2(\mathbf{x}^k) + \frac{\sigma_1 - \gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \\
& \geq f^* + \frac{\sigma_1 - \gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2,
\end{aligned}$$

where the first inequality follows from $f_2(\mathbf{x}^k) + f_2^*(\mathbf{y}^{k-1}) \geq \langle \mathbf{x}^k, \mathbf{y}^{k-1} \rangle$ and f^* is the lower bound of f. Thus, $\liminf_{k\to\infty} E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}) \geq f^*$. Then, combining the fact that the sequence $\{E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})\}$ is nonincreasing (statement (i)), we obtain the existence of $\lim_{k\to\infty} E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})$. Next, we prove the last part of (ii). Given any $(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}) \in \Upsilon$, there exists a subsequence $(\mathbf{x}^{k_i}, \mathbf{y}^{k_i-1}, \mathbf{x}^{k_i-1})$ such that

$$\lim_{i\to\infty} \|(\mathbf{x}^{k_i}, \mathbf{y}^{k_i-1}, \mathbf{x}^{k_i-1}) - (\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{x}})\| = 0.$$

Then we have

$$\begin{aligned}
&\lim_{k\to\infty} E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1}) \\
&\quad = \lim_{i\to\infty} E(\mathbf{x}^{k_i+1}, \mathbf{y}^{k_i}, \mathbf{x}^{k_i}) \\
&\quad = \lim_{i\to\infty} f_1(\mathbf{x}^{k_i+1}) - \langle \mathbf{x}^{k_i+1}, \mathbf{y}^{k_i}\rangle + f_2^*(\mathbf{y}^{k_i}) + \frac{\sigma_1-\gamma}{2}\|\mathbf{x}^{k_i+1}-\mathbf{x}^{k_i}\|^2 \\
&\quad = \lim_{i\to\infty} f_1(\mathbf{x}^{k_i+1}) - \langle \mathbf{x}^{k_i+1}, \mathbf{y}^{k_i}\rangle - f_2(\mathbf{x}^{k_i}) + \langle \mathbf{y}^{k_i}, \mathbf{x}^{k_i}\rangle + \frac{\sigma_1-\gamma}{2}\|\mathbf{x}^{k_i+1}-\mathbf{x}^{k_i}\|^2 \\
&\quad \le \limsup_{i\to\infty} f_1(\mathbf{x}^{k_i}) - \langle \mathbf{y}^{k_i} + \gamma(\mathbf{x}^{k_i}-\mathbf{x}^{k_i-1}), \mathbf{x}^{k_i}-\mathbf{x}^{k_i+1}\rangle - \frac{\sigma_1}{2}\|\mathbf{x}^{k_i+1}-\mathbf{x}^{k_i}\|^2 \\
&\qquad - \langle \mathbf{x}^{k_i+1}, \mathbf{y}^{k_i}\rangle - f_2(\mathbf{x}^{k_i}) + \langle \mathbf{y}^{k_i}, \mathbf{x}^{k_i}\rangle + \frac{\sigma_1-\gamma}{2}\|\mathbf{x}^{k_i+1}-\mathbf{x}^{k_i}\|^2 \\
&\quad = \limsup_{i\to\infty} f(\mathbf{x}^{k_i}) - \gamma\langle \mathbf{x}^{k_i}-\mathbf{x}^{k_i-1}, \mathbf{x}^{k_i}-\mathbf{x}^{k_i+1}\rangle - \frac{\gamma}{2}\|\mathbf{x}^{k_i+1}-\mathbf{x}^{k_i}\|^2 \\
&\quad = \limsup_{i\to\infty} f(\mathbf{x}^{k_i}) \\
&\quad = \zeta \\
&\quad = f(\bar{\mathbf{x}}) \le E(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}),
\end{aligned}$$

where the third equality follows from $\mathbf{y}^{k_i} \in \partial f_2(\mathbf{x}^{k_i})$ and thus $f_2(\mathbf{x}^{k_i}) + f_2^*(\mathbf{y}^{k_i}) = \langle \mathbf{y}^{k_i}, \mathbf{x}^{k_i}\rangle$; the first inequality is implied by $\mathbf{y}^{k_i} + \gamma(\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}) \in \partial f_1(\mathbf{x}^{k_i+1})$ and σ_1-convexity of f_1; the fifth equality follows from $\lim_{i\to\infty} \|\mathbf{x}^{k_i} - \mathbf{x}^{k_i-1}\| = 0$. On the other hand, the closedness of E implies that $E(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}) \le \zeta$. Thus, we obtain that $E(\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}) = \zeta$ and conclude that $E \equiv \zeta$ on Υ. □

Next, we start to create the sequential convergence of RInDCA_e based on the KL property of E and Proposition 5.1. Note that the KL property of E can be implied by the KL property of f (see Liu et al. 2019).

Theorem 5.1 *Suppose that the assumptions in Proposition 5.1 hold. Let E defined in (20) be a KL function, $\{\mathbf{x}^k\}$ be the sequence generated by RInDCA_e, and $\mathbf{y}^k \in \partial f_2(\mathbf{x}^k)$ for all $k \ge 0$. Then we have*

(i) *There exists $D > 0$ such that $\forall k \ge 1$,*

$$\operatorname{dist}(\mathbf{0}, \partial E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})) \le D(\|\mathbf{x}^k - \mathbf{x}^{k-1}\| + \|\mathbf{x}^{k-1} - \mathbf{x}^{k-2}\|). \quad (26)$$

(ii) *If $\{\mathbf{x}^k\}$ is bounded, then $\{\mathbf{x}^k\}$ converges to a critical point of problem (P).*

Proof (i) Note that the subdifferential of E at $\mathbf{w}^k := (\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})$ is:

$$\partial E(\mathbf{w}^k) = \begin{bmatrix} \partial f_1(\mathbf{x}^k) - \mathbf{y}^{k-1} + \frac{\gamma-\sigma_2}{2}(\mathbf{x}^k - \mathbf{x}^{k-1}) \\ -\mathbf{x}^k + \partial f_2^*(\mathbf{y}^{k-1}) \\ -\frac{\gamma-\sigma_2}{2}(\mathbf{x}^k - \mathbf{x}^{k-1}) \end{bmatrix}.$$

Because $\mathbf{y}^{k-1} \in \partial f_2(\mathbf{x}^{k-1})$, thus $\mathbf{x}^{k-1} \in \partial f_2^*(\mathbf{y}^{k-1})$; besides, we have $\mathbf{y}^{k-1} + \gamma(\mathbf{x}^{k-1} - \mathbf{x}^{k-2}) \in \partial f_1(\mathbf{x}^k)$. Combining these relations, we have

$$\begin{bmatrix} \gamma(\mathbf{x}^{k-1} - \mathbf{x}^{k-2}) + \frac{\gamma-\sigma_2}{2}(\mathbf{x}^k - \mathbf{x}^{k-1}) \\ -\mathbf{x}^k + \mathbf{x}^{k-1} \\ -\frac{\gamma-\sigma_2}{2}(\mathbf{x}^k - \mathbf{x}^{k-1}) \end{bmatrix} \in \partial E(\mathbf{w}^k).$$

Thus, it is easy to see that there exists $D > 0$ such that $\forall k \geq 1$,

$$\text{dist}(\mathbf{0}, \partial E(\mathbf{x}^k, \mathbf{y}^{k-1}, \mathbf{x}^{k-1})) \leq D(\|\mathbf{x}^k - \mathbf{x}^{k-1}\| + \|\mathbf{x}^{k-1} - \mathbf{x}^{k-2}\|).$$

(ii) It is sufficient to prove that $\sum_{k=1}^{\infty} \|\mathbf{x}^k - \mathbf{x}^{k-1}\| < \infty$. If there exists $N_0 \geq 1$ such that $E(\mathbf{w}^{N_0}) = \zeta_1$, then the inequality (22) implies that $\mathbf{x}^k = \mathbf{x}^{N_0}, \forall k \geq N_0$; naturally, the sequence $\{\mathbf{x}^k\}$ is convergent. Thus, we can assume that

$$E(\mathbf{w}^k) > \zeta, \forall k \geq 1.$$

Then, it follows from Υ is a compact set, $\Upsilon \subseteq \text{dom}\partial E$ and $E \equiv \zeta$ on Υ that there exists $\varepsilon > 0$, $\eta > 0$ and $\varphi \in \Xi_\eta$ such that

$$\varphi'(E(\mathbf{x}, \mathbf{y}, \mathbf{z}) - \zeta)\text{dist}(\mathbf{0}, \partial E(\mathbf{x}, \mathbf{y}, \mathbf{z})) \geq 1$$

for all $(\mathbf{x}, \mathbf{y}, \mathbf{z}) \in U$ with

$$U = \{(\mathbf{x}, \mathbf{y}, \mathbf{z}) : \text{dist}((\mathbf{x}, \mathbf{y}, \mathbf{z}), \Upsilon) < \varepsilon\} \cap \{(\mathbf{x}, \mathbf{y}, \mathbf{z}) : \zeta < E(\mathbf{x}, \mathbf{y}, \mathbf{z}) < \zeta + \eta\}.$$

Moreover, it is easy to derive from $\lim_{k\to\infty} \text{dist}(\mathbf{w}^k, \Upsilon) = 0$ and $\lim_{k\to\infty} E(\mathbf{w}^k) = \zeta$ that there exists $N \geq 1$ such that $\mathbf{w}^k \in U, \forall k \geq N$. Thus,

$$\varphi'(E(\mathbf{w}^k) - \zeta)\text{dist}(\mathbf{0}, \partial E(\mathbf{w}^k) \geq 1, \ \forall k \geq N.$$

Using the concavity of φ, we see that $\forall k \geq N$,

$$\begin{aligned} &[\varphi(E(\mathbf{w}^k) - \zeta) - \varphi(E(\mathbf{w}^{k+1}) - \zeta)]\text{dist}(\mathbf{0}, \partial E(\mathbf{w}^k)) \\ &\geq \varphi'(E(\mathbf{w}^k) - \zeta)\text{dist}(\mathbf{0}, \partial E(\mathbf{w}^k))(E(\mathbf{w}^k) - E(\mathbf{w}^{k+1})) \\ &\geq E(\mathbf{w}^k) - E(\mathbf{w}^{k+1}). \end{aligned} \tag{27}$$

Let $\Delta_k := \varphi(E(\mathbf{w}^k) - \zeta) - \varphi(E(\mathbf{w}^{k+1}) - \zeta)$ and $C := \frac{\sigma_1+\sigma_2-2\gamma}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2$. Then combining (27), (26) and (i) of Proposition 4.2, we have

$$\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2 \leq \frac{D}{C}\Delta_k(\|\mathbf{x}^k - \mathbf{x}^{k-1}\| + \|\mathbf{x}^{k-1} + \mathbf{x}^{k-2}\|).$$

Thus, taking square roots on both sides, we obtain

$$\|\mathbf{x}^k - \mathbf{x}^{k-1}\| \leq \sqrt{\frac{2D}{C}\Delta_k}\sqrt{\frac{\|\mathbf{x}^k - \mathbf{x}^{k-1}\| + \|\mathbf{x}^{k-1} - \mathbf{x}^{k-2}\|}{2}}$$
$$\leq \frac{D}{C}\Delta_k + \frac{1}{4}\|\mathbf{x}^k - \mathbf{x}^{k-1}\| + \frac{1}{4}\|\mathbf{x}^{k-1} - \mathbf{x}^{k-2}\|,$$

this yields

$$\frac{1}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\| \leq \frac{D}{C}\Delta_k + \frac{1}{4}\|\mathbf{x}^{k-1} - \mathbf{x}^{k-2}\| - \frac{1}{4}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|.$$

Considering also that $\sum_{k=1}^{\infty}\Delta_k < \infty$, thus the above inequality implies that $\sum_{k=1}^{\infty}\|\mathbf{x}^k - \mathbf{x}^{k-1}\| < \infty$, which indicates that $\{\mathbf{x}^k\}$ is a Cauchy sequence, and thus convergent. The proof is completed. □

The sequential convergence of $\mathrm{RInDCA_n}$ can be similarly developed based on the KL property of $\widetilde{E}$, Proposition 4.2, and Proposition 5.1 for $\widetilde{E}$, thus the proof is omitted.

6 Numerical results

In this part, we conduct numerical simulations on testing copositivity of matrices (constrained case) and image denoising problem (unconstrained case). All experiments are implemented in Matlab 2019a on a 64-bit PC with an Intel(R) Core(TM) i5-6200U CPU (2.30GHz) and 8GB of RAM.

6.1 Checking copositivity of matrices

Recall that a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is called copositive if $\mathbf{x}^\top\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \geq 0$, and called non-copositive otherwise. Consider the next formulation

$$\begin{aligned} \min \quad & \frac{1}{2}\mathbf{x}^\top\mathbf{A}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \in \mathbb{R}^n_+. \end{aligned} \tag{COPO}$$

Clearly, $\mathbf{A}$ is copositive if and only if the optimal value of problem (COPO) is nonnegative, and problem (COPO) is equivalent to the next DC program

$$f(\mathbf{x}) = \underbrace{\frac{L}{2}\|\mathbf{x}\|^2 + \delta_{\mathbb{R}^n_+}(\mathbf{x})}_{f_1(\mathbf{x})} - \underbrace{(\frac{L}{2}\|\mathbf{x}\|^2 - \frac{1}{2}\mathbf{x}^\top\mathbf{A}\mathbf{x})}_{f_2(\mathbf{x})}, \tag{DC_{copo}}$$

where $L \geq \|\mathbf{A}\|$. In Dür and Hiriart-Urruty (2013), DCA applied for the above DC decomposition is used to check the copositivity of $\mathbf{A}$. For our experiments, the same instances of $\mathbf{A}$ as in Dür and Hiriart-Urruty (2013) will be used. A brief introduction about the instances is given as follows. Let $\mathbf{E}_n \in \mathbb{R}^{n\times n}$ be the matrix with all entries being one. The matrix $\mathbf{A}_{\mathrm{cycle}} = (a_{ij}) \in \mathbb{R}^{n\times n}$ is given component-wise by

Table 1 Performances of DCA, InDCA$_e$, and RInDCA$_e$ for checking the copositivity of $\mathbf{H}_n$ with $n \in \{500, 1000, 1500, 2000\}$

Size (n)	Algorithm	It.	Time	Fval
500	DCA	1963	0.0799	4.0483e–14
	InDCA$_e$	1965	0.0817	4.0470e–14
	RInDCA$_e$	**1020**	**0.0399**	**9.6531e–15**
1000	DCA	2915	1.1634	1.6403e–13
	InDCA$_e$	2916	1.1692	1.6449e–13
	RInDCA$_e$	**1562**	**0.6276**	**3.9847e–14**
1500	DCA	4772	4.9034	3.7082e–13
	InDCA$_e$	4773	4.9557	3.7170e–13
	RInDCA$_e$	**2542**	**2.5938**	**9.1359e–14**
2000	DCA	5829	10.8607	6.6003e–13
	InDCA$_e$	5830	11.1934	6.6094e–13
	RInDCA$_e$	**3129**	**6.0101**	**1.6295e–13**

$$a_{ij} := \begin{cases} 1 & \text{if } |i-j| \in \{1, n-1\}, \\ 0 & \text{otherwise.} \end{cases}$$

It was known that the following matrix

$$\mathbf{Q}_n^{\mu} := \mu(\mathbf{E}_n - \mathbf{A}_{\text{cycle}}) - \mathbf{E}_n \in \mathbb{R}^{n\times n}$$

is copositive for $\mu \geq 2$ and non-copositive for $\mu < 2$. When $\mu = 2$, $\mathbf{Q}_n^{\mu}$ is called Horn matrix, denoted as $\mathbf{H}_n$.

In order to apply DCA, InDCA$_e$, and RInDCA$_e$ for (DC_{copo}), we set $L = \|\mathbf{A}\|$ for DCA and $L = \|\mathbf{A}\| + 1$ for InDCA$_e$ and RInDCA$_e$. The inertial ranges of InDCA$_e$ and RInDCA$_e$ are $[0, 1/2)$ and $[0, (\|\mathbf{A}\| + 2)/2)$, respectively. For InDCA$_e$, $\gamma = 0.499$ is set, while for RInDCA$_e$ we choose $\gamma = 0.499(\|\mathbf{A}\| + 2)$. Moreover, the common initial point for all involved algorithms is randomly generated as follows: we first use the standard normal distribution to generate a point $\mathbf{y} \in \mathbb{R}^n$, then we obtain the initial point $\mathbf{x}^0 = (\mathbf{x}_j^0)$ by setting

$$\mathbf{x}_j^0 = \frac{e^{\mathbf{y}_j}}{\sum_{j=1}^{n} e^{\mathbf{y}_j}}$$

for $j = 1, \dots, n$. We report the number of iterations, the computational time, and the objective functions values. The bold values in Tables 1 and 2 highlight the best numerical results. The computational time presented is the average execution time (in seconds) over 10 runs.

First, we consider to check the copositivity of the Horn matrices. All algorithms are terminated when $\|\mathbf{x}^k - \mathbf{x}^{k-1}\| < 10^{-9}$. Table 1 describes the performances of DCA, InDCA$_e$, and RInDCA$_e$ for checking the copositivity of $\mathbf{H}_n$ with $n \in \{500, 1000, 1500, 2000\}$. It is observed that DCA and InDCA$_e$ perform almost the same since the number of iterations of InDCA$_e$ is very close to that of DCA. We also observe that RInDCA$_e$ performs the best among the three algorithms,

Table 2 Performances of DCA, InDCA_e, and RInDCA_e for checking the copositivity of $\mathbf{Q}_n^\mu$ ($\mu = 1.9$) with $n \in \{500, 1000, 1500, 2000\}$

Size (n)	Algorithm	It.	Time	Fval
500	DCA	430	0.0330	– 9.9882e–07
	InDCA_e	430	0.0359	– 9.9139e–07
	RInDCA_e	**209**	**0.0169**	– 9.8189e–07
1000	DCA	824	0.5786	– 9.9785e–07
	InDCA_e	825	0.6014	– 9.9907e–07
	RInDCA_e	**405**	**0.2991**	– 9.9934e–07
1500	DCA	2036	4.6078	– 9.9965e–07
	InDCA_e	2037	4.6695	– 9.9975e–07
	RInDCA_e	**1021**	**2.3593**	– 9.9985e–07
2000	DCA	5094	14.8749	– 9.9993e–07
	InDCA_e	5095	15.1039	– 9.9987e–07
	RInDCA_e	**2559**	**7.7138**	– 9.9977e–07

terminating with almost half of the number of iterations of DCA or InDCA_e, and with the smallest objective function values. The reason why InDCA_e loses the inertial-force effect may be that 0.499 is much smaller than $0.499(\|\mathbf{A}\| + 2)$ for the instance $\mathbf{H}_n$ when n is large.

Next, we check the non-copositivity of $\mathbf{Q}_n^\mu$ ($\mu = 1.9$). Here, the stopping criterion for the involved algorithms DCA, InDCA_e, and RInDCA_e is that $f(\mathbf{x}^k) \leq -10^{-6}$. Once this condition is verified, then the non-copositivity of the involved matrix is determined. Table 2 shows the performances of DCA, InDCA_e, and RInDCA_e for checking the copositivity of $\mathbf{Q}_n^\mu$ ($\mu = 1.9$) with $n \in \{500, 1000, 1500, 2000\}$.

It is also observed that among the involved algorithms RInDCA_e performs the best, hitting the stopping criterion $f(\mathbf{x}^k) \leq -10^{-6}$ with about half of the number of iterations as well as the time of DCA or InDCA_e. The behavior of the algorithms for checking the noncopositivity almost coincides with the previous experiment.

The above experiments show that the effect of heavy-ball inertial-force may be lost if the strong convexity of the first DC component is not considered.

6.2 Image denoising problem

Consider a gray-scale image of $m \times n$ pixels and with entries in [0, 1], where 0 represents pure black and 1 represents pure white. The original image $\mathbf{X} \in \mathbb{R}^{m\times n}$ is assumed to be corrupted with a noise $\mathbf{B}$. Then, we obtain the observed image $\widetilde{\mathbf{X}} = \mathbf{X} + \mathbf{B}$, which is known to us. We aim to recover the original image $\mathbf{X}$ by solving the next problem

$$\min \left\{ \frac{\mu}{2} \|\mathbf{X} - \widetilde{\mathbf{X}}\|_F^2 + \text{TV}_\phi(\mathbf{X}) \, : \, \mathbf{X} \in \mathbb{R}^{m\times n} \right\}, \tag{28}$$

where $\mu > 0$ is a trade-off constant; $\widetilde{\mathbf{X}}$ is the observed image; $\|\cdot\|_F$ represents the Frobenius norm; $\phi : \mathbb{R}_+ \to \mathbb{R}$ is a concave function and TV_ϕ is defined as:

Table 3 Some examples of concave function $\phi_a : \mathbb{R}_+ \to \mathbb{R}$ parameterized by $a > 0$

	$\phi_{\log}$	ϕ_{rat}	ϕ_{atan}	$\phi_{\exp}$
$\phi_a(r)$	$\frac{\log(1+ar)}{a}$	$\frac{r}{1+ar/2}$	$\frac{\text{atan}((1+ar)/\sqrt{3})-\pi/6}{a\sqrt{3}/4}$	$\frac{1-\exp(-ar)}{a}$
$\phi_a'(r)$	$\frac{1}{1+ar}$	$\frac{1}{(1+ar/2)^2}$	$\frac{4}{a^2r^2+2ar+4}$	$\frac{1}{\exp(ar)}$

$$\text{TV}_\phi(\mathbf{X}) = \sum_{i,j} \phi(\|\nabla \mathbf{X}_{i,j}\|)$$

with

$$\|\nabla \mathbf{X}_{i,j}\| = \begin{cases} |\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}| & i = m, j \neq n, \\ |\mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}| & i \neq m, j = n, \\ 0 & i = m, j = n, \\ \sqrt{(\mathbf{X}_{i,j} - \mathbf{X}_{i+1,j})^2 + (\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1})^2} & \text{otherwise.} \end{cases}$$

When $\phi \equiv 1$, we denote TV as TV_ϕ. The model (28) with $\phi \equiv 1$ is convex. In (de Oliveira and Tcheou 2019), it has shown that the convex model ($\phi \equiv 1$) is not effective to preserve edges in the restoration process of piecewise-constant images than the model (28) with ϕ being concave, such as the ϕ described in Table 3. Moreover it has also shown how to derive the DC decomposition for TV_ϕ. Concretely, $\text{TV}_\phi(\mathbf{X})$ has the following DC decomposition

$$\text{TV}_\phi(\mathbf{X}) = \text{TV}(\mathbf{X}) - (\text{TV}(\mathbf{X}) - \text{TV}_\phi(\mathbf{X})).$$

Then, we get from (28) the next two equivalent DC programs

$$\min\{\underbrace{\frac{\mu}{2}\|\mathbf{X} - \widetilde{\mathbf{X}}\|_F^2 + \text{TV}(\mathbf{X})}_{f_1(\mathbf{X})} - \underbrace{(\text{TV}(\mathbf{X}) - \text{TV}_\phi(\mathbf{X}))}_{f_2(\mathbf{X})} : \mathbf{X} \in \mathbb{R}^{m \times n}\}, \tag{29}$$

and

$$\min\{\underbrace{\frac{\mu+\rho}{2}\|\mathbf{X} - \widetilde{\mathbf{X}}\|_F^2 + \text{TV}(\mathbf{X})}_{\widetilde{f}_1(\mathbf{X})} - \underbrace{(\frac{\rho}{2}\|\mathbf{X} - \widetilde{\mathbf{X}}\|_F^2 + \text{TV}(\mathbf{X}) - \text{TV}_\phi(\mathbf{X}))}_{\widetilde{f}_2(\mathbf{X})}\}, \tag{30}$$

where $\rho > 0$.

We will compare DCA, InDCA_e, RInDCA_e, and ADCA (Phan et al. 2018) (a DCA based algorithm using the Nesterov's extrapolation). Note that DCA and ADCA are applied for (29), while InDCA_e and RInDCA_e are applied for (30). All of the involved algorithms require solving the convex subproblems in form of

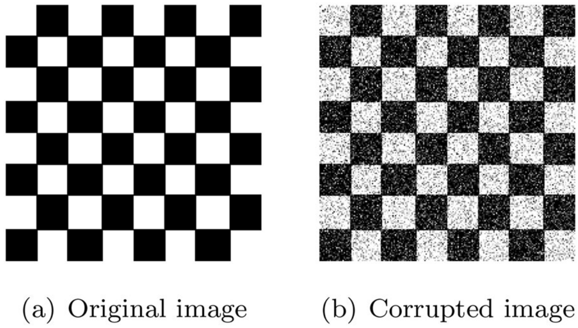

Fig. 2 Original and corrupted images

$$\min\left\{\frac{\tilde{t}}{2}\|\mathbf{X}-\mathbf{Z}\|_F^2+\mathrm{TV}(\mathbf{X})\,:\,\mathbf{X}\in\mathbb{R}^{m\times n}\right\},$$

where $\tilde{t}>0$ and $\mathbf{Z}\in\mathbb{R}^{m\times n}$, which can be efficiently solved by the fast gradient projection (FGP) method proposed in Beck and Teboulle (2009).

We use the gray-scale image checkerboard.png (200×200 pixels) for our experiments. White noise with variance 0.1 is added to the image to produce the observed image $\tilde{\mathbf{X}}$ in both (29) and (30). Here, we use the same method as described in de Oliveira and Tcheou (2019) to generate the corrupted image. The original image and the corrupted image are shown in Fig. 2. To measure the quality of reconstruction images, we use the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index (see (Monga 2017)). In general, the larger PSNR means the better restoration, while the closer to 1 of SSIM indicates the better reconstruction quality. For all the involved algorithms, the observed images $\tilde{\mathbf{X}}$ are set as the common initial point, whose convex subproblems are all solved by FGP with the stopping condition that the distance between two consecutive iterates is smaller than 10^{-4}. It is also worth noting that for problems (29) and (30), another DCA based method with Nesterov's extrapolation developed in Wen et al. (2018) reduces to DCA.

We test DCA and ADCA for (29), $\mathrm{InDCA_e}$ and $\mathrm{RInDCA_e}$ for (30). The parameter μ for both (29) and (30) ranges from 0.8 to 1.6, and we set $\phi=\phi_{\mathrm{rat}}$ with $a=6$, moreover, the parameter ρ in (30) is fixed as 1. The inertial step-size γ in $\mathrm{InDCA_e}$ is set as $0.5\times\rho\times 99\%=0.495$, and in $\mathrm{RInDCA_e}$ as $0.5\times(\mu+2\rho)\times 99\%=0.495\mu+0.99$. Besides, the parameter q in ADCA is set as 3. In Table 4, we present in detail the best objective values f_{best} within 100 iterations and the corresponding values of PSNR and SSIM. The trends of the objective function values together with the values of PSNR and SSIM when $\mu=1.2$ with respect to running time (average on 10 runs) are plotted in Fig 3. Moreover, the recovered images within 100 iterations are demonstrated in Fig 4. Here, we observes that $\mathrm{RInDCA_e}$ always obtains the lowest objective function values, and almost always the best PSNR and SSIM. The benefit of enlarged inertial step-size is indicated by comparing $\mathrm{InDCA_e}$ and $\mathrm{RInDCA_e}$. Moreover, we also observe that ADCA performs better than DCA and $\mathrm{InDCA_e}$. Finally, by comparing DCA and $\mathrm{InDCA_e}$, we observe that adding a strong convexity term to the original DC program (29) degrades the effectiveness of DCA-based algorithm, which can not be offset by the benefit of inertial-force.

Table 4 Restoration of the corrupted checkerboard image

Algorithm	μ	f_{best}	PSNR	SSIM
DCA		2.2920e+03	24.2319	0.8268
$InDCA_e$	0.8	2.2937e+03	24.2902	0.8269
$RInDCA_e$		**2.2896e+03**	24.4612	**0.8324**
ADCA		2.2909e+03	**24.4654**	0.8272
DCA		2.4923e+03	24.1899	0.8324
$InDCA_e$	0.9	2.4937e+03	24.1308	0.8307
$RInDCA_e$		**2.4876e+03**	**24.3191**	**0.8363**
ADCA		2.4907e+03	24.2289	0.8344
DCA		2.6944e+03	23.8179	0.8265
$InDCA_e$	1.0	2.6974e+03	23.7593	0.8262
$RInDCA_e$		**2.6854e+03**	**24.1334**	**0.8390**
ADCA		2.6914e+03	23.9806	0.8308
DCA		2.8987e+03	23.3509	0.8202
$InDCA_e$	1.1	2.9017e+03	23.2949	0.8171
$RInDCA_e$		**2.8839e+03**	**23.7590**	**0.8433**
ADCA		2.8937e+03	23.4928	0.8272
DCA		3.0961e+03	23.0492	0.8114
$InDCA_e$	1.2	3.0971e+03	22.9270	0.8114
$RInDCA_e$		**3.0811e+03**	**23.3767**	**0.8353**
ADCA		3.0927e+03	23.1191	0.8166
DCA		3.2964e+03	22.4314	0.7942
$InDCA_e$	1.3	3.2997e+03	22.4193	0.7933
$RInDCA_e$		**3.2745e+03**	**23.0360**	**0.8279**
ADCA		3.2908e+03	22.5914	0.8030
DCA		3.4935e+03	21.9890	0.7723
$InDCA_e$	1.4	3.4967e+03	21.9640	0.7700
$RInDCA_e$		**3.4688e+03**	**22.5424**	**0.8127**
ADCA		3.4876e+03	22.0525	0.7797
DCA		3.6881e+03	21.5175	0.7517
$InDCA_e$	1.5	3.6907e+03	21.5080	0.7487
$RInDCA_e$		**3.6584e+03**	**22.1562**	**0.8004**
ADCA		3.6817e+03	21.5928	0.7569
DCA		3.8768e+03	21.1589	0.7272
$InDCA_e$	1.6	3.8791e+03	21.1393	0.7240
$RInDCA_e$		**3.8478e+03**	**21.6137**	**0.7752**
ADCA		3.8704e+03	21.2309	0.7357

(Bold values for the best numerical results)

As a conclusion, for the image denoising problem, our method $RInDCA_e$ is a promising approach, which outperforms the other tested algorithms.

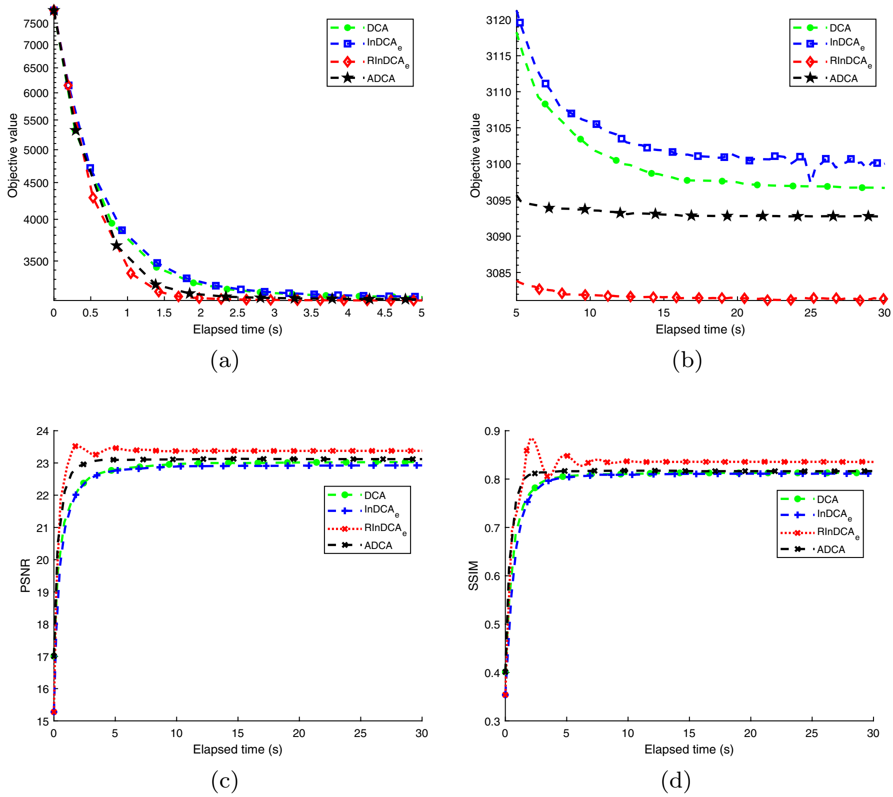

Fig. 3 The trends of the objective function values together with PSNR and SSIM

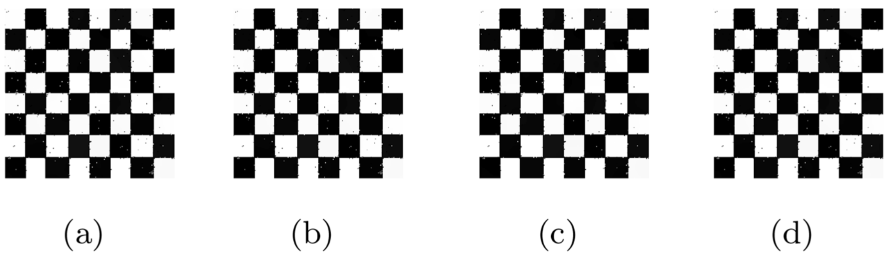

Fig. 4 Image reconstructions where **a, b, c, d** is respectively the resulting image by DCA, InDCA$_e$, RInDCA$_e$, and ADCA

7 Conclusion and perspective

In this paper, based on the inertial DC algorithm (de Oliveira and Tcheou 2019) for DC programming, we propose a refined version with larger inertial step-size, for which we prove the subsequential convergence and the sequential convergence by further assuming the KL property. Numerical simulations on checking copositivity of matrices and image denoising problem show the good performance of our proposed methods and the benefit of larger step-size.

About the future works, the numerical test on RInDCA_n will be investigated. We may also introduce non-heavy-ball and non-Nesterov acceleration to DC programming. Furthermore, the inertial-force procedure will be extended to the partial DC programming (Pham et al. 2021) in which the objective function f depends on two variables $\mathbf{x}$ and $\mathbf{y}$ where $f(\mathbf{x}, .)$ and $f(., \mathbf{y})$ are both DC functions.

Acknowledgements This work is supported by the National Natural Science Foundation of China (Grant 11601327). We thank the anonymous referees for their valuable remarks.

References

Adly S, Attouch H, Le HM (2021) First order inertial optimization algorithms with threshold effects associated with dry friction. hal:03284220
Aleksandrov AD (2012) On the surfaces representable as difference of convex functions. Sib Elektron Mat Izvest 9:360–376
Artacho FJA, Campoy R, Vuong PT (2019) The boosted dc algorithm for linearly constrained dc programming. arXiv:1908.01138
Artacho FJA, Fleming RM, Vuong PT (2018) Accelerating the dc algorithm for smooth functions. Math Programm 169(1):95–118
Artacho FJA, Vuong PT (2020) The boosted dc algorithm for nonsmooth functions. SIAM J Optim 30(1):980–1006
Attouch H, Bolte J (2009) On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math Programm 116(1):5–16
Attouch H, Bolte J, Svaiter BF (2013) Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss-seidel methods. Math Programm 137(1):91–129
Beck A (2017) First-order methods in optimization. SIAM
Beck A, Teboulle M (2009) Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans Image Process 18(11):2419–2434
Bolte J, Daniilidis A, Lewis A (2007) The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J Optim 17(4):1205–1223
Bolte J, Sabach S, Teboulle M (2014) Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math Programm 146(1):459–494
de Oliveira W (2020) The abc of dc programming. Set-Valued Variat Anal 28(4):679–706
de Oliveira W, Tcheou MP (2019) An inertial algorithm for dc programming. Set-Valued Variat Anal 27(4):895–919
Dür M, Hiriart-Urruty JB (2013) Testing copositivity with the help of difference-of-convex optimization. Math Program 140(1 Ser B):31–43
Hartman P (1959) On functions representable as a difference of convex functions. Pac J Math 9(3):707–713
Hiriart-Urruty JB (1985) Generalized differentiability/duality and optimization for problems dealing with differences of convex functions. In: Ponstein J (ed) Convexity and duality in optimization. Springer, Berlin, pp 37–70
Horst R, Thoai NV (1999) Dc programming: overview. J Optim Theory Appl 103:1–43
Joki K, Bagirov AM, Karmitsa N, Mäkelä MM (2017) A proximal bundle method for nonsmooth dc optimization utilizing nonconvex cutting planes. J Glob Optim 68(3):501–535
Le Thi HA, Pham DT (2005) 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
Le Thi HA, Pham DT (2018) Dc programming and dca: thirty years of developments. Math Programm 169(1):5–68

Le Thi HA, Pham DT, Ngai HV (2012) Exact penalty and error bounds in dc programming. J Glob Optim 52:509–535
Liu T, Pong TK, Takeda A (2019) A refined convergence analysis of pDCA$_e$ ith applications to simultaneous sparse recovery and outlier detection. Comput Optim Appl 73(1):69–100
Lu Z, Zhou Z, Sun Z (2019) Enhanced proximal dc algorithms with extrapolation for a class of structured nonsmooth dc minimization. Math Programm 176(1):369–401
Monga V (2017) Handbook of convex optimization methods in imaging science, vol 1. Springer, Berlin
Mukkamala MC, Ochs P, Pock T, Sabach S (2020) Convex-concave backtracking for inertial Bregman proximal gradient algorithms in nonconvex optimization. SIAM J Math Data Sci 2(3):658–682
Nesterov YE (1983) A method for solving the convex programming problem with convergence rate $O(1/k^2)$. In: Dokl. akad. nauk Sssr, vol. 269, pp. 543–547
Nesterov YE (2013) Gradient methods for minimizing composite functions. Math Programm 140(1):125–161
Niu YS, Wang YJ (2019) Higher-order moment portfolio optimization via the difference-of-convex programming and sums-of-squares. arXiv:1906.01509
Ochs P, Chen Y, Brox T, Pock T (2014) ipiano: inertial proximal algorithm for nonconvex optimization. SIAM J Imaging Sci 7(2):1388–1419
Pang JS, Razaviyayn M, Alvarado A (2017) Computing b-stationary points of nonsmooth dc programs. Math Oper Res 42(1):95–118
Pham DT, Huynh VN, Le Thi HA, Ho VT (2021) Alternating dc algorithm for partial dc programming problems. J Glob Optim. https://doi.org/10.1007/s10898-021-01043-w
Pham DT, Le Thi HA (1997) Convex analysis approach to d.c. programming: theory, algorithms and applications. Acta Math Vietnam 22(1):289–355
Pham DT, Le Thi HA (1998) A dc optimization algorithm for solving the trust-region subproblem. SIAM J Optim 8(2):476–505
Pham DT, Le Thi HA (2014) Recent advances in DC programming and DCA. Transactions on Computational Intelligence XIII, pp 1–37
Pham DT, Souad EB (1986) Algorithms for solving a class of nonconvex optimization problems methods of subgradients. In: Hiriart-Urruty J-B (ed) Fermat days 85: mathematics for optimization, vol 129. Elsevier, New York, pp 249–271
Phan D.N, Le H.M, Le Thi H.A (2018) Accelerated difference of convex functions algorithm and its application to sparse binary logistic regression. In: IJCAI, pp. 1369–1375
Pock T, Sabach S (2016) Inertial proximal alternating linearized minimization (ipalm) for nonconvex and nonsmooth problems. SIAM J Imaging Sci 9(4):1756–1787
Polyak BT (1987) Introduction to optimization, vol 1. Optimization Software Inc., New York
Rockafellar RT (1970) Convex analysis, vol 36. Princeton University Press, Princeton
Souza JCO, Oliveira PR, Soubeyran A (2016) Global convergence of a proximal linearized algorithm for difference of convex functions. Optim Lett 10(7):1529–1539
Wen B, Chen X, Pong TK (2018) A proximal difference-of-convex algorithm with extrapolation. Comput Optim Appl 69(2):297–324
Zavriev S, Kostyuk F (1993) Heavy-ball method in nonconvex optimization problems. Comput Math Model 4(4):336–341