



Convergence rate analysis of an extrapolated proximal difference-of-convex algorithm

Lejia Gao¹ · Bo Wen²

Received: 8 June 2022 / Revised: 17 September 2022 / Accepted: 25 September 2022 /

Published online: 3 October 2022

© The Author(s) under exclusive licence to Korean Society for Informatics and Computational Applied Mathematics 2022

Abstract

This paper considers a class of difference-of-convex (DC) optimization problems, whose objective function is the sum of a convex smooth function and a possibly nonsmooth DC function. We first propose a new extrapolated proximal difference-of-convex algorithm, which incorporates a more general setting of the extrapolation parameters $\{\beta_k\}$. Then we prove the subsequential convergence of the proposed method to a stationary point of the DC problem. Based on the Kurdyka–Łojasiewicz inequality, the global convergence and convergence rate of the whole sequence generated by our method have been established. Finally, some numerical experiments on the DC regularized least squares problems have been performed to demonstrate the efficiency of our proposed method.

Keywords Difference-of-convex optimization · Convergence analysis · Extrapolation parameters · Kurdyka–Łojasiewicz inequality

Mathematics Subject Classification 65K05 · 90C26 · 90C30

This author's work was supported in part by NSFC11801131, Natural Science Foundation of Hebei Province, (Grant No. A2019202229), a scientific grant of Hebei Educational Committee (QN2018101).

✉ Bo Wen
wenboh@163.com

Lejia Gao
ljgaohebut@163.com

¹ School of Science, Hebei University of Technology, Tianjin, People's Republic of China

² Institute of Mathematics, Hebei University of Technology, Tianjin, People's Republic of China

1 Introduction

In recent years, DC optimization problems arise in many application areas, such as machine learning [12, 15] and image processing [7, 9, 21]. Due to their popularity, many numerical algorithms have been proposed for solving these problems [1, 13, 14, 29]. In this paper, we consider a special DC problem, which is in the following form:

$$\min \{ \Psi(x) := f(x) + g(x) \mid x \in \mathbb{R}^n \}, \quad (1.1)$$

where g is a difference-of-convex function:

$$g(x) = g_n(x) - g_s(x), \quad (1.2)$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth convex function with Lipschitz continuous gradient, $g_n : \mathbb{R}^n \rightarrow \mathbb{R}$ is a proper closed convex function and $g_s : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuous convex function. In addition, we assume the objective function Ψ is level-bounded, which means that $\inf \Psi(x) > -\infty$ and the set of global minimizers of (1.1) is nonempty.

Since problem (1.1) is a class of DC problems, the well-known and classical DC algorithm (DCA) [26] can be applied to solving the optimization problem. The main iteration step is:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + g_n(x) - \langle \xi^k, x \rangle \right\}, \quad (1.3)$$

where $\xi^k \in \partial g_s(x^k)$ is a subgradient of g_s at $x^k \in \mathbb{R}^n$. As the scale of this problem becomes larger and larger, the original DCA probably not very efficient. By using the proximal mapping and a specific DC decomposition [27], large-scale DC optimization problems are solved via the proximal DC algorithm (PDCA) [11]. It can be suitably applied to the problem (1.1) for which the new iterate is obtained by solving the following proximal subproblem

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ g_n(x) + \langle \nabla f(x^k) - \xi^k, x \rangle + \frac{L}{2} \|x - x^k\|^2 \right\}, \quad (1.4)$$

where $\xi^k \in \partial g_s(x^k)$ and $L > 0$ is the Lipschitz continuity modulus of ∇f . However, this algorithm may take a large number of iterations.

Therefore, various attempts have been made to accelerate PDCA. By using the extrapolation accelerating techniques, Wen et al. [31] proposed a proximal DC algorithm with extrapolation (pDCAe) to solve the problem (1.1). The key iteration step is as follows:

$$\begin{cases} y^k = x^k + \beta_k (x^k - x^{k-1}), \\ x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ g_n(x) + \langle \nabla f(y^k) - \xi^k, x \rangle + \frac{L}{2} \|x - y^k\|^2 \right\}, \end{cases} \quad (1.5)$$

where $\xi^k \in \partial g_s(x^k)$, $L > 0$ is the Lipschitz continuity modulus of ∇f and $\beta_k (x^k - x^{k-1})$ is an extrapolation term. The choice of extrapolation parameters $\{\beta_k\}$

in [31] can cover many popular choices of extrapolation parameters including those used in fast iterative shrinkage-thresholding algorithm (FISTA) with fixed restart strategy [24] for solving (1.1) in the case $g(x) = g_n(x)$, which is given as follows:

$$\beta_k = \frac{t_{k-1} - 1}{t_k} \text{ with } t_k = \frac{1 + \sqrt{1 + 4t_{k-1}^2}}{2}, \tag{1.6}$$

where $t_{-1} = t_0 = 1$. Recently, based on the difference-of-Moreau-envelopes smoothing technique, Sun et al. [30] developed an inexact gradient descent method for solving the DC programming, which is

$$\begin{cases} x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ g_n(x) + \langle \nabla f(x^k), x \rangle + \frac{1}{2\nu} \|x - z^k\|^2 \right\}, \\ x_{\nu g_s}(z^k) = \arg \min_{x \in \mathbb{R}^n} \left\{ g_s(x) + \frac{1}{2\nu} \|x - z^k\|^2 \right\}, \\ z^{k+1} = z^k + \beta(x^{k+1} - x_{\nu g_s}(z^k)), \end{cases} \tag{1.7}$$

where $\nu \in (0, \frac{1}{L})$, $L > 0$ is the Lipschitz continuity modulus of ∇f and $\beta \in (0, 2)$.

For other accelerated DC algorithms for solving constrained DC programming have been further proposed to improve the quality of solutions and the rate of convergence. Pang et al. [25] proposed an enhanced DC algorithm (EDCA) to solve the considered DC problem, meanwhile speed up the subsequential convergence. Further, based on EDCA, Lu et al. [22, 23] developed the enhanced proximal DC algorithm (EPDCA) to improve the efficiency of solving constrained DC programming with g_s to be the supremum of many convex smooth functions in (1.2), while nonmonotone EPDCA is proposed by utilizing the nonmonotone line search technique. Moreover, Sun et al. [30] proposed algorithms by combining the difference-of-Moreau-envelopes smoothing technique with the augmented Lagrangian function and obtained well convergence results.

In this paper, motivated by the success of pDCAe on accelerating the original PDCA for solving structured DC optimization problems, we aim to investigate a more general extrapolation scheme in order to improve the numerical performance of PDCA. Concretely, inspired by the generalized Nesterov momentum scheme for the accelerated forward-backward algorithm (AFBA) proposed in [16] and pDCAe in [31], we consider a new extrapolated proximal DC algorithm (EPDCA) with a more general setting of the extrapolation parameters $\{\beta_k\}$, where a power parameter $\omega \in (0, 1]$ is introduced in the extrapolation scheme.

We first prove that any accumulation point of the sequence generated by our method is a stationary point under very general conditions. Moreover, according to Kurdyka-Łojasiewicz property, we establish the global convergence and convergence rate of the sequence generated by our method, which does not require the assumption that g_s is continuously differentiable. In the end, we conduct some numerical experiments to show the efficiency of our method.

The main contribution of this work can be summarized as following:

- We propose a novel extrapolation scheme with more general setting of the extrapolation parameters $\{\beta_k\}$ in EPDCA to solve DC optimization problem (1.1) for potential acceleration. Then we establish the convergence results of our method.
- By constructing an auxiliary function, we prove the global convergence results of EPDCA without the assumption that g_s is a locally Lipschitz continuous function, which extended the convergence results of pDCAe developed in [31].
- According to our numerical results, the performance of our method is efficient on ℓ_{1-2} regularized least squares problem [32] and logarithmic regularized least squares problem [6].

The rest of this paper is organized as follows. In Sect. 2, some basic definitions used in the whole paper are given. Section 3 gives the framework of EPDCA and discuss the selection of the extrapolation parameters $\{\beta_k\}$. In Sect. 4, we analyze the convergence behaviors of our method. Some numerical experiments are conducted to show the efficiency of the proposed method in comparison with other existing methods in Sect. 5. Section 6 concludes the whole paper.

2 Preliminaries

In this paper, the Euclidean scalar product of \mathbb{R}^n and its corresponding norm are respectively denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$. For a $m \times n$ matrix A , let its transpose be denoted by A^T . If A is a symmetric matrix, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote the largest and smallest eigenvalues respectively. For a given nonempty closed set $C \subseteq \mathbb{R}^n$, let $\text{dist}(\cdot, C) : \mathbb{R}^n \rightarrow \mathbb{R}$ denotes the distance function to C , i.e.

$$\text{dist}(x, C) := \inf_{y \in C} \|x - y\|, \quad \forall x \in \mathbb{R}^n. \quad (2.1)$$

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

$$\text{dom } f := \{x \in \mathbb{R}^n : f(x) < \infty\} \quad (2.2)$$

denotes its effective domain. f is called proper if it does not take the value $-\infty$ and $\text{dom } f$ is nonempty.

The set

$$\text{epi } f := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\} \quad (2.3)$$

denotes the epigraph of f , and f is closed (or convex) if $\text{epi } f$ is closed (or convex). A proper closed function f is said to be level-bounded if the lower level set $\{x \in \mathbb{R}^n : f(x) \leq r\}$ is bounded for any $r \in \mathbb{R}$.

Given a proper closed function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$, the Fréchet subdifferential of f at $x \in \text{dom } f$ is given by

$$\partial^F f(x) := \left\{ y \in \mathbb{R}^n : \liminf_{z \rightarrow x} \frac{f(z) - f(x) - \langle y, z - x \rangle}{\|z - x\|} \geq 0 \right\}, \quad (2.4)$$

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

$$\partial f(x) := \left\{ y \in \mathbb{R}^n : \exists x^k \rightarrow x, f(x^k) \rightarrow f(x), y^k \in \partial^F f(x^k) \text{ such that } y^k \rightarrow y \right\}, \tag{2.5}$$

while for $x \notin \text{dom } f, \partial f(x) = \emptyset$. If f is convex, the Fréchet subdifferential and the limiting subdifferential will coincide with the convex subdifferential, that is,

$$\partial^F f(x) = \partial f(x) = \left\{ y \in \mathbb{R}^n : f(z) \geq f(x) + \langle y, z - x \rangle \text{ for all } z \in \mathbb{R}^n \right\}. \tag{2.6}$$

A proper closed function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is called σ -strongly convex with $\sigma \geq 0$ if for any $x, z \in \text{dom } f$ and $\lambda \in [0, 1]$, it holds that

$$f(\lambda x + (1 - \lambda)z) \leq \lambda f(x) + (1 - \lambda)f(z) - \frac{\sigma}{2}\lambda(1 - \lambda)\|x - z\|^2. \tag{2.7}$$

Moreover, let f be a σ -strongly convex function, it is known that for any $x \in \text{dom } \partial f, y \in \partial f(x)$ and $z \in \text{dom } f$, we have

$$f(z) \geq f(x) + \langle y, z - x \rangle + \frac{\sigma}{2}\|z - x\|^2. \tag{2.8}$$

For a proper closed convex function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$, let $f^*(u) = \sup_{x \in \mathbb{R}^n} \{ \langle u, x \rangle - f(x) \}$ denotes the conjugate function of $f(x)$. Then f^* is proper closed convex and for any x and y , the Young's inequality holds, that is

$$f(x) + f^*(y) \geq \langle x, y \rangle, \tag{2.9}$$

the equality holds if and only if $y \in \partial f(x)$. Moreover, for any x and y , one has $y \in \partial f(x)$ if and only if $x \in \partial f^*(y)$.

Finally, we present some preliminaries on the Kurdyka–Łojasiewicz property [2–4]. These concepts play a central role in our theoretical and algorithmic developments. For $\eta \in (0, \infty]$, we denote by Ξ_η the set of all concave continuous functions $\phi : [0, \eta) \rightarrow [0, \infty)$ that are continuously differentiable over $(0, \eta)$ with positive derivatives and satisfy $\phi(0) = 0$.

Definition 2.1 (*Kurdyka–Łojasiewicz property*) A proper closed function $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is said to satisfy the Kurdyka–Łojasiewicz property at $\bar{x} \in \text{dom } \partial f$ if there exist $\eta \in (0, \infty]$, a neighborhood W of \bar{x} , and a function $\phi \in \Xi_\eta$ such that for all x in the intersection

$$W \cap \{ x \in \mathbb{R}^n : f(\bar{x}) < f(x) < f(\bar{x}) + \eta \}$$

it holds that

$$\phi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1.$$

If f satisfies the Kurdyka–Łojasiewicz property at any point of $\text{dom } \partial f$, then f is called a Kurdyka–Łojasiewicz function.

The following uniformized Kurdyka–Łojasiewicz property [5] plays an important role in our sequential convergence analysis.

Lemma 2.1 (Uniformized Kurdyka–Łojasiewicz property) *Let $\Omega \subseteq \mathbb{R}^n$ be a compact set and let $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$ be a proper closed function. If f is constant on Ω and satisfies the Kurdyka–Łojasiewicz property at each point of Ω , then there exist $\epsilon, \eta > 0$ and $\phi \in \Xi_\eta$ such that*

$$\phi'(f(x) - f(\bar{x})) \text{dist}(0, \partial f(x)) \geq 1$$

for any $\bar{x} \in \Omega$ and any x satisfying $\text{dist}(x, \Omega) < \epsilon$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \eta$.

3 Extrapolated proximal difference-of-convex algorithm

In this section, we will show the framework of the extrapolated proximal difference-of-convex algorithm (EPDCA) and give the chosen of the extrapolation parameters $\{\beta_k\}$, which are more general than those proposed in (1.6).

Extrapolated proximal difference-of-convex algorithm (EPDCA):

Input: $x^0 \in \text{dom } g_n, \{\beta_k\} \subseteq [0, 1)$. Set $x^{-1} = x^0$.

for $k = 0, 1, 2, \dots$

 Take any $\xi^{k+1} \in \partial g_s(x^k)$ and compute

$$y^k = x^k + \beta_k(x^k - x^{k-1}), \tag{3.1}$$

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ g_n(x) + \langle \nabla f(y^k) - \xi^{k+1}, x \rangle + \frac{L}{2} \|x - y^k\|^2 \right\}.$$

end for

The extrapolation parameters $\{\beta_k\}$ are selected as follows:

$$\beta_k = \frac{t_{k-1} - 1}{t_k} \text{ with } t_{k-1} = a(k-1)^\omega + b, \quad k \geq 1 \tag{3.2}$$

where $\omega \in (0, 1]$, $a \in \mathbb{R}^+$ and $b \in \mathbb{R}$. Set $t_{-1} = t_0 = b$. It is worth noting that in order to avoid the denominator being zero, without further mentioning, we always set $b \in \mathbb{R} \setminus \{-ak^\omega : k \in \mathbb{N}^+\}$.

Remark 3.1 (i) If we choose $\omega = 1, a = \frac{1}{\alpha-1}$ with $\alpha > 3, b = 1$ in (3.2), the extrapolation parameters $\{\beta_k\}$ become as follows:

$$\beta_k = \frac{k-1}{k+\alpha-1}. \tag{3.3}$$

This setting was proposed by Chambolle and Dossal [8], which is a special case of our proposed extrapolation parameters in (3.2).

- (ii) Let the extrapolation parameters $\{\beta_k\}$ in (3.2) with $a = \frac{1}{2}, b = 1$ and $\omega = 1$, we can obtain the extrapolation parameters $\{\beta_k\}$ in (1.6) and (3.2) converge to the same value with the same convergence rate from [17, Proposition 2], which means that the setting (1.6) of the extrapolation parameters $\{\beta_k\}$ are asymptotically consistent to a special case of the parameters we proposed in (3.2).
- (iii) The choice of the power ω depends on the specific problems. We choose $\omega = 0.98$ in the numerical experiments for solving two classes of DC regularized least squares problems, see Sect. 5.
- (iv) We can easily deduce that the sequence $\{\beta_k\}$ is increasing to 1 from (3.2). Therefore, given a fixed positive integer K , we can choose $\beta_k \equiv \beta_K$ whenever $k > K$, which implies the supremum of β_k is strictly less than 1.

More concrete examples of such setting of extrapolation parameters have been extensively used for the Nesterov’s accelerated forward–backward algorithm, see [16] for more details.

4 Convergence analysis

In this section, we are going to analyze the convergence behaviors of EPDCA. We first define the following auxiliary sequence:

$$H(x^k, \xi^k, x^{k-1}) := f(x^k) + g_n(x^k) - \langle x^k, \xi^k \rangle + g_s^*(\xi^k) + \frac{\mu}{2} \|x^k - x^{k-1}\|^2,$$

where $\mu \in [L\bar{\beta}^2, L]$ with $\bar{\beta} = \sup_k \beta_k < 1, \xi^k \in \partial g_s(x^{k-1})$ and $g_s^*(\xi^k)$ is the conjugate function of $g_s(x^k)$. Based on the Young’s inequality, we obtain that

$$H(x^k, \xi^k, x^{k-1}) \geq f(x^k) + g_n(x^k) - g_s(x^k) + \frac{\mu}{2} \|x^k - x^{k-1}\|^2 \geq \Psi(x^k). \tag{4.1}$$

It is worth mentioning that using this auxiliary sequence structure, we can establish the global convergence of the sequence generated by our method to solve problem (1.1) without requiring g_s to be continuously differentiable. The idea of constructing such an auxiliary sequence similarity can be found in [20].

Lemma 4.1 *Let $\{x^k\}$ be a sequence generated by EPDCA and $\mu \in [L\bar{\beta}^2, L]$, it holds that*

$$H(x^{k+1}, \xi^{k+1}, x^k) - H(x^k, \xi^k, x^{k-1}) \leq \frac{\mu - L}{2} \|x^{k+1} - x^k\|^2 + \frac{L\bar{\beta}_k^2 - \mu}{2} \|x^k - x^{k-1}\|^2. \tag{4.2}$$

Furthermore, the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing.

Proof Using the strong convexity of the objective in the subproblem (3.1), we can compare the objective values of this strongly convex function at x^{k+1} and x^k that

$$\begin{aligned} &g_n(x^{k+1}) + \langle \nabla f(y^k) - \xi^{k+1}, x^{k+1} \rangle + \frac{L}{2} \|x^{k+1} - y^k\|^2 \\ &\leq g_n(x^k) + \langle \nabla f(y^k) - \xi^{k+1}, x^k \rangle + \frac{L}{2} \|x^k - y^k\|^2 - \frac{L}{2} \|x^{k+1} - x^k\|^2. \end{aligned}$$

Rearranging terms, we obtain that

$$\begin{aligned} g_n(x^{k+1}) &\leq g_n(x^k) + \langle -\nabla f(y^k) + \xi^{k+1}, x^{k+1} - x^k \rangle + \frac{L}{2} \|x^k - y^k\|^2 \\ &\quad - \frac{L}{2} \|x^{k+1} - y^k\|^2 - \frac{L}{2} \|x^{k+1} - x^k\|^2. \end{aligned}$$

On the other hand, using the fact that ∇f is Lipschitz continuous with a Lipschitz constant of $L > 0$, we have

$$f(x^{k+1}) \leq f(y^k) + \langle \nabla f(y^k), x^{k+1} - y^k \rangle + \frac{L}{2} \|x^{k+1} - y^k\|^2.$$

Using the above two inequalities, we obtain further that

$$\begin{aligned} &f(x^{k+1}) + g_n(x^{k+1}) \\ &\leq f(y^k) + \langle \nabla f(y^k), x^k - y^k \rangle + \frac{L}{2} \|x^k - y^k\|^2 + g_n(x^k) + \langle \xi^{k+1}, x^{k+1} - x^k \rangle \\ &\quad - \frac{L}{2} \|x^{k+1} - x^k\|^2 \\ &\leq f(x^k) + \frac{L}{2} \|x^k - y^k\|^2 + g_n(x^k) + \langle \xi^{k+1}, x^{k+1} - x^k \rangle - \frac{L}{2} \|x^{k+1} - x^k\|^2 \\ &\leq f(x^k) + \frac{L}{2} \beta_k^2 \|x^k - x^{k-1}\|^2 + g_n(x^k) + \langle \xi^{k+1}, x^{k+1} - x^k \rangle - \frac{L}{2} \|x^{k+1} - x^k\|^2, \end{aligned} \tag{4.3}$$

where the second inequality follows from the convexity of f , and the last inequality holds due to the relation $x^k - y^k = x^k - (x^k + \beta_k(x^k - x^{k-1})) = -\beta_k(x^k - x^{k-1})$. From the definition of the auxiliary sequence $\{H(x^k, \xi^k, x^{k-1})\}$ and the conjugate function g_s^* , we have

$$\begin{aligned} &H(x^{k+1}, \xi^{k+1}, x^k) = f(x^{k+1}) + g_n(x^{k+1}) - \langle x^{k+1}, \xi^{k+1} \rangle + g_s^*(\xi^{k+1}) + \frac{\mu}{2} \|x^{k+1} - x^k\|^2 \\ &\leq f(x^k) + \frac{L}{2} \beta_k^2 \|x^k - x^{k-1}\|^2 + g_n(x^k) + \langle \xi^{k+1}, x^{k+1} - x^k \rangle - \frac{L}{2} \|x^{k+1} - x^k\|^2 \\ &\quad - \langle x^{k+1}, \xi^{k+1} \rangle + g_s^*(\xi^{k+1}) + \frac{\mu}{2} \|x^{k+1} - x^k\|^2 \\ &= f(x^k) + \frac{L}{2} \beta_k^2 \|x^k - x^{k-1}\|^2 + g_n(x^k) + \langle \xi^{k+1}, x^{k+1} - x^k \rangle - \frac{L}{2} \|x^{k+1} - x^k\|^2 \end{aligned}$$

$$\begin{aligned}
 & + \langle x^k - x^{k+1}, \xi^{k+1} \rangle - g_s(x^k) + \frac{\mu}{2} \|x^{k+1} - x^k\|^2 \\
 & = f(x^k) + \frac{L}{2} \beta_k^2 \|x^k - x^{k-1}\|^2 + g_n(x^k) + \left(\frac{\mu}{2} - \frac{L}{2}\right) \|x^{k+1} - x^k\|^2 - g_s(x^k), \quad (4.4)
 \end{aligned}$$

where the first inequality follows from (4.3), and the second equality follows from the fact that $\xi^{k+1} \in \partial g_s(x^k)$. Combing the Young’s inequality on (4.4), we obtain that

$$\begin{aligned}
 & H(x^{k+1}, \xi^{k+1}, x^k) \\
 & \leq f(x^k) + \frac{L}{2} \beta_k^2 \|x^k - x^{k-1}\|^2 + g_n(x^k) + \left(\frac{\mu}{2} - \frac{L}{2}\right) \|x^{k+1} - x^k\|^2 \\
 & \quad - \langle x^k, \xi^k \rangle + g_s^*(\xi^k) \\
 & = f(x^k) + g_n(x^k) - \langle x^k, \xi^k \rangle + g_s^*(\xi^k) + \frac{\mu}{2} \|x^k - x^{k-1}\|^2 \\
 & \quad + \left(\frac{\mu}{2} - \frac{L}{2}\right) \|x^{k+1} - x^k\|^2 + \left(\frac{L}{2} \beta_k^2 - \frac{\mu}{2}\right) \|x^k - x^{k-1}\|^2 \\
 & = H(x^k, \xi^k, x^{k-1}) + \left(\frac{\mu}{2} - \frac{L}{2}\right) \|x^{k+1} - x^k\|^2 + \left(\frac{L}{2} \beta_k^2 - \frac{\mu}{2}\right) \|x^k - x^{k-1}\|^2.
 \end{aligned}$$

where the second equality follows from the definition of the sequence $\{H(x^k, \xi^k, x^{k-1})\}$. Therefore,

$$\begin{aligned}
 & H(x^{k+1}, \xi^{k+1}, x^k) - H(x^k, \xi^k, x^{k-1}) \\
 & \leq \left(\frac{\mu}{2} - \frac{L}{2}\right) \|x^{k+1} - x^k\|^2 + \left(\frac{L}{2} \beta_k^2 - \frac{\mu}{2}\right) \|x^k - x^{k-1}\|^2.
 \end{aligned}$$

Recall that $L\bar{\beta}^2 \leq \mu \leq L$ and $\bar{\beta} = \sup_k \beta_k < 1$, we obtain that for any $k > 0$,

$$\frac{\mu}{2} - \frac{L}{2} \leq 0, \text{ and } \frac{L}{2} \beta_k^2 - \frac{\mu}{2} \leq \frac{L}{2} \bar{\beta}^2 - \frac{\mu}{2} \leq 0.$$

Consequently, we see immediately that

$$H(x^{k+1}, \xi^{k+1}, x^k) - H(x^k, \xi^k, x^{k-1}) \leq 0,$$

which means that the auxiliary sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing. This completes the proof. □

Before analyzing the convergence behaviors, on the basis of Fermat’s rule [28, Theorem 10.1], we give the definition of stationary points of the objective function Ψ as follows.

Definition 4.1 Let Ψ be given in (1.1). We say that \tilde{x} is a stationary point of Ψ if

$$0 \in \partial \nabla f(\tilde{x}) + \partial g_n(\tilde{x}) - \partial g_s(\tilde{x}).$$

The set of all stationary points of Ψ is denoted by \mathcal{X} .

We next give the global subsequential convergence of the sequence $\{x^k\}$ generated by EPDCA to solve problem (1.1).

Theorem 4.1 (Global subsequential convergence) *Let $\{x^k\}$ be the sequence generated by EPDCA and $\mu \in [L\beta^2, L]$. Then the following statements hold:*

- (i) *The sequences $\{x^k\}$ and $\{\xi^k\}$ are bounded.*
- (ii) *$\|x^k - x^{k-1}\| \rightarrow 0$ as $k \rightarrow \infty$.*
- (iii) *Any accumulation point of the sequence $\{x^k\}$ is a stationary point of Ψ .*

Proof (i) Since the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing from Lemma 4.1, together with the inequality (4.1), we obtain that

$$\Psi(x^k) \leq H(x^k, \xi^k, x^{k-1}) \leq H(x^0, \xi^0, x^{-1}) < \infty.$$

Noting that by assumption Ψ is level-bounded, thus we have the sequence $\{x^k\}$ is bounded. Moreover, due to the boundedness of $\{x^k\}$, the continuity of g_s , and the fact that $\xi^k \in \partial g_s(x^{k-1})$, we immediately have the sequence $\{\xi^k\}$ is also bounded. This proves (i).

(ii) From Lemma 4.1 and according to the fact that $\mu \leq L$, we have

$$\begin{aligned} & H(x^{k+1}, \xi^{k+1}, x^k) - H(x^k, \xi^k, x^{k-1}) \\ & \leq \frac{\mu - L}{2} \|x^{k+1} - x^k\|^2 + \frac{L\beta_k^2 - \mu}{2} \|x^k - x^{k-1}\|^2 \\ & \leq \frac{L\beta_k^2 - \mu}{2} \|x^k - x^{k-1}\|^2. \end{aligned}$$

By rearranging the terms above, we obtain that

$$\frac{\mu - L\beta_k^2}{2} \|x^k - x^{k-1}\|^2 \leq H(x^k, \xi^k, x^{k-1}) - H(x^{k+1}, \xi^{k+1}, x^k). \tag{4.5}$$

Furthermore, summing up both sides of the above inequality with k ranging from 1 to m , we have

$$\begin{aligned} 0 & \leq \sum_{k=1}^m \left[\frac{\mu - L\beta_k^2}{2} \|x^k - x^{k-1}\|^2 \right] \leq \sum_{k=1}^m \left(H(x^k, \xi^k, x^{k-1}) - H(x^{k+1}, \xi^{k+1}, x^k) \right) \\ & = H(x^1, \xi^1, x^0) - H(x^{m+1}, \xi^{m+1}, x^m). \end{aligned} \tag{4.6}$$

Letting m to the limit in (4.6), we conclude that

$$\frac{\mu - L\beta_k^2}{2} \sum_{k=1}^{\infty} \|x^k - x^{k-1}\|^2 < \infty.$$

Using the fact that $\mu \geq L\bar{\beta}^2 \geq L\beta_k^2$, we have $\sum_{k=1}^{\infty} \|x^k - x^{k-1}\|^2 < \infty$. Therefore, we immediately obtain that $\|x^k - x^{k-1}\| \rightarrow 0$ as $k \rightarrow \infty$. This proves (ii).

(iii) Let x^* be an accumulation point of $\{x^k\}$. Then there exists a subsequence $\{x^{k_j}\}$ convergent to x^* . From the first-order optimality condition of the subproblem (3.1), we obtain that

$$0 \in \partial g_n(x^{k_j+1}) + \nabla f(y^{k_j}) - \xi^{k_j+1} + L(x^{k_j+1} - y^{k_j}), \tag{4.7}$$

where $\xi^{k_j+1} \in \partial g_s(x^{k_j})$. On the other hand, due to the fact that $\|x^{k_j} - x^{k_j-1}\| \rightarrow 0$ from (ii), we have

$$\begin{aligned} x^{k_j+1} - y^{k_j} &= x^{k_j+1} - (x^{k_j} + \beta_{k_j}(x^{k_j} - x^{k_j-1})) \\ &= (x^{k_j+1} - x^{k_j}) - \beta_{k_j}(x^{k_j} - x^{k_j-1}) \rightarrow 0. \end{aligned} \tag{4.8}$$

In addition, recall that the sequence $\{\xi^{k_j+1}\}$ is bounded from (i), so without loss of generality we can assume that the limit of the sequence $\{\xi^{k_j+1}\}$ exists, and it belongs to $\partial g_s(x^*)$ due to the closedness of ∂g_s . Noting the closedness of ∂g_n and the continuity of ∇f , then taking j to the limit in (4.7), we obtain that

$$0 \in \partial g_n(x^*) + \nabla f(x^*) - \partial g_s(x^*), \tag{4.9}$$

which means that x^* is a stationary point of Ψ . This completes the proof. □

Before analyzing the global convergence results of EPDCA, we will give some properties of the auxiliary sequence $\{H(x^k, \xi^k, x^{k-1})\}$.

Proposition 4.1 *Let $\{x^k\}$ be the sequence generated by EPDCA for solving (1.1) and $\mu \in [L\bar{\beta}^2, L]$. Then the following statements hold:*

- (i) $\lim_{k \rightarrow \infty} \text{dist}((0, 0, 0), \partial H(x^k, \xi^k, x^{k-1})) = 0$.
- (ii) $\zeta := \lim_{k \rightarrow \infty} H(x^k, \xi^k, x^{k-1})$ exists.
- (iii) The set of accumulation points of $\{(x^k, \xi^k, x^{k-1})\}$, denoted by Υ , is a nonempty compact set. Moreover, $H \equiv \zeta$ on Υ .

Proof (i) Taking the subdifferential of the auxiliary sequence $\{H(x^k, \xi^k, x^{k-1})\}$, we obtain that for $k \geq 1$,

$$\partial H(x^k, \xi^k, x^{k-1}) = \begin{bmatrix} \nabla f(x^k) + \partial g_n(x^k) - \xi^k + \mu(x^k - x^{k-1}) \\ -x^k + \partial g_s^*(\xi^k) \\ -\mu(x^k - x^{k-1}) \end{bmatrix}.$$

Using the first-order optimality of the subproblem (3.1), we have

$$-\nabla f(y^{k-1}) + \xi^k - L(x^k - y^{k-1}) \in \partial g_n(x^k).$$

From the above relation and the fact that $x^{k-1} \in \partial g_s^*(\xi^k)$, we obtain further that

$$\left[\begin{array}{c} \nabla f(x^k) - \nabla f(y^{k-1}) - L(x^k - y^{k-1}) + \mu(x^k - x^{k-1}) \\ x^{k-1} - x^k \\ -\mu(x^k - x^{k-1}) \end{array} \right] \in \partial H(x^k, \xi^k, x^{k-1}),$$

for all $k \geq 1$. This together with the relation that $x^k - y^{k-1} = x^k - x^{k-1} - \beta_{k-1}(x^{k-1} - x^{k-2})$ and the Lipschitz continuity of ∇f , we conclude that there exists $A > 0$ such that

$$\text{dist}((0, 0, 0), \partial H(x^k, \xi^k, x^{k-1})) \leq A \left(\|x^k - x^{k-1}\| + \|x^{k-1} - x^{k-2}\| \right). \tag{4.10}$$

Since the facts that $\|x^k - x^{k-1}\| \rightarrow 0$ and $\|x^{k-1} - x^{k-2}\| \rightarrow 0$ from Theorem 4.1 (ii), we deduce that $\lim_{k \rightarrow \infty} \text{dist}((0, 0, 0), \partial H(x^k, \xi^k, x^{k-1})) = 0$. This proves (i).

(ii) Noting that the inequality $H(x^k, \xi^k, x^{k-1}) \geq \Psi(x^k)$ holds in (4.1) and due to the level-boundedness of Ψ , we obtain that the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is bounded below. In addition, the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing from Lemma 4.1. Therefore, we conclude that $\zeta := \lim_{k \rightarrow \infty} H(x^k, \xi^k, x^{k-1})$ exists. This proves (ii).

(iii) Using the facts $\{x^k\}$ and $\{\xi^k\}$ are bounded from Theorem 4.1 (i), we know immediately that the set of accumulation points of $\{(x^k, \xi^k, x^{k-1})\}$ is a nonempty compact set Υ .

We now show that $H \equiv \zeta$ on Υ . To this end, take any $(\hat{x}, \hat{\xi}, \hat{v}) \in \Upsilon$, then there exists a convergent subsequence $\{(x^{k_j}, \xi^{k_j}, x^{k_j-1})\}$ such that $\lim_{j \rightarrow \infty} (x^{k_j}, \xi^{k_j}, x^{k_j-1}) = (\hat{x}, \hat{\xi}, \hat{v})$. From the lower semicontinuity of the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ and the definition of ζ , we have

$$H(\hat{x}, \hat{\xi}, \hat{v}) \leq \liminf_{j \rightarrow \infty} H(x^{k_j}, \xi^{k_j}, x^{k_j-1}) = \zeta. \tag{4.11}$$

On the other hand, since x^{k_j} is the minimizer of the subproblem (3.1), we have

$$\begin{aligned} g_n(x^{k_j}) + \langle \nabla f(y^{k_j-1}) - \xi^{k_j}, x^{k_j} \rangle + \frac{L}{2} \|x^{k_j} - y^{k_j-1}\|^2 \\ \leq g_n(\hat{x}) + \langle \nabla f(y^{k_j-1}) - \xi^{k_j}, \hat{x} \rangle + \frac{L}{2} \|\hat{x} - y^{k_j-1}\|^2. \end{aligned} \tag{4.12}$$

By rearranging the terms above, we obtain that

$$\begin{aligned}
 &g_n(x^{k_j}) + \langle \nabla f(y^{k_j-1}) - \xi^{k_j}, x^{k_j} - \hat{x} \rangle + \frac{L}{2} \|x^{k_j} - y^{k_j-1}\|^2 \\
 &\leq g_n(\hat{x}) + \frac{L}{2} \|\hat{x} - y^{k_j-1}\|^2.
 \end{aligned}
 \tag{4.13}$$

Using the facts that $y^{k_j-1} = x^{k_j-1} + \beta_{k_j-1}(x^{k_j-1} - x^{k_j-2})$, $\|x^k - x^{k-1}\| \rightarrow 0$ from Theorem 4.1(ii), and $\lim_{j \rightarrow \infty} x^{k_j} = \hat{x}$, we have the following relations:

$$\|x^{k_j} - y^{k_j-1}\| \leq \|x^{k_j} - x^{k_j-1}\| + \bar{\beta} \|x^{k_j-1} - x^{k_j-2}\| \rightarrow 0,
 \tag{4.14}$$

$$\|\hat{x} - y^{k_j-1}\| = \|\hat{x} - x^{k_j} + x^{k_j} - y^{k_j-1}\| \leq \|\hat{x} - x^{k_j}\| + \|x^{k_j} - y^{k_j-1}\| \rightarrow 0.
 \tag{4.15}$$

In addition, using the continuous differentiability of f , $\lim_{j \rightarrow \infty} x^{k_j} = \hat{x}$, and the boundedness of $\{\xi^{k_j}\}$ and $\{y^{k_j}\}$, we have

$$\lim_{j \rightarrow \infty} \langle \nabla f(y^{k_j-1}) - \xi^{k_j}, x^{k_j} - \hat{x} \rangle = 0.
 \tag{4.16}$$

Therefore, from the definition of the auxiliary sequence $\{H(x^k, \xi^k, x^{k-1})\}$, we obtain that

$$\begin{aligned}
 \zeta &= \lim_{j \rightarrow \infty} H(x^{k_j}, \xi^{k_j}, x^{k_j-1}) \\
 &= \lim_{j \rightarrow \infty} f(x^{k_j}) + g_n(x^{k_j}) - \langle x^{k_j}, \xi^{k_j} \rangle + g_s^*(\xi^{k_j}) + \frac{L}{2} \|x^{k_j} - x^{k_j-1}\|^2 \\
 &= \lim_{j \rightarrow \infty} f(x^{k_j}) + g_n(x^{k_j}) - \langle x^{k_j}, \xi^{k_j} \rangle + g_s^*(\xi^{k_j}) + \frac{L}{2} \|x^{k_j} - y^{k_j-1}\|^2 \\
 &= \lim_{j \rightarrow \infty} f(x^{k_j}) + \langle \nabla f(y^{k_j-1}) - \xi^{k_j}, x^{k_j} - \hat{x} \rangle + \frac{L}{2} \|x^{k_j} - y^{k_j-1}\|^2 \\
 &\quad + g_n(x^{k_j}) - \langle x^{k_j}, \xi^{k_j} \rangle + g_s^*(\xi^{k_j}) \\
 &\leq \limsup_{j \rightarrow \infty} f(x^{k_j}) + \frac{L}{2} \|\hat{x} - y^{k_j-1}\|^2 + g_n(\hat{x}) - \langle x^{k_j}, \xi^{k_j} \rangle + g_s^*(\xi^{k_j}) \\
 &= \limsup_{j \rightarrow \infty} f(x^{k_j}) + g_n(\hat{x}) - \langle x^{k_j} - x^{k_j-1}, \xi^{k_j} \rangle - g_s(x^{k_j-1}) \\
 &= f(\hat{x}) + g_n(\hat{x}) - g_s(\hat{x}) = \Psi(\hat{x}) \leq H(\hat{x}, \hat{\xi}, \hat{v}),
 \end{aligned}
 \tag{4.17}$$

where the third equality follows from the facts that $\|x^{k_j} - x^{k_j-1}\| \rightarrow 0$ and (4.14), the fourth equality follows from (4.16), the first inequality holds due to (4.13), the fifth equality follows from (4.15) and the fact that $\xi^{k_j} \in \partial g_s(x^{k_j-1})$, the sixth equality

holds due to $\|x^{k_j} - x^{k_j-1}\| \rightarrow 0$ from Theorem 4.1 (ii), the boundedness of $\{\xi^k\}$ and the continuity of f and g_s , while the last inequality follows from the relation (4.1).

Therefore, we conclude that $H(\hat{x}, \hat{\xi}, \hat{\nu}) = \lim_{j \rightarrow \infty} H(x^{k_j}, \xi^{k_j}, x^{k_j-1})$ from (4.11) and (4.17). Due to the arbitrariness of $(\hat{x}, \hat{\xi}, \hat{\nu})$, we have $H \equiv \zeta$ on Υ . This completes the proof. □

Next, we are going to analyze the global convergence and the convergence rate of the sequence $\{x^k\}$ generated by EPDCA to solve (1.1).

Theorem 4.2 (Global convergence) *Let $\{x^k\}$ be the sequence generated by EPDCA for solving (1.1) and $\mu \in [L\bar{\beta}^2, L]$. Assume the auxiliary function $H(x, \xi, \nu)$ is a Kurdyka–Łojasiewicz function. Then the sequence $\{x^k\}$ converges to a stationary point of Ψ ; moreover, $\sum_{k=1}^\infty \|x^k - x^{k-1}\| < \infty$.*

Proof From Theorem 4.1 (iii), it is sufficient to prove the sequence $\{x^k\}$ is convergent. We first consider the case that there exists an integer $k > 0$, such that $H(x^k, \xi^k, x^{k-1}) = \zeta$. Due to the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing and convergent to ζ , we obtain that for any $\bar{j} \geq 0$, $H(x^{k+\bar{j}}, \xi^{k+\bar{j}}, x^{k+\bar{j}-1}) = \zeta$ is satisfied. Therefore, from (4.5) we conclude that for all $k > \bar{j} \geq 0$, it holds that $x^k = x^{k+\bar{j}}$, which means that the sequence $\{x^k\}$ converges finitely.

In the following, we consider the case that for any $k > 0$, $H(x^k, \xi^k, x^{k-1}) > \zeta$ holds. Recall from Proposition 4.1 (iii) that Υ is a compact subset and $H \equiv \zeta$ on Υ , where Υ is the set of the accumulation points of $\{(x^k, \xi^k, x^{k-1})\}$. In addition, noting that $H(x, \xi, \nu)$ is a Kurdyka–Łojasiewicz function, according to Lemma 2.1, there exists a positive scalar ϵ and a continuous concave function $\phi \in \Xi_\eta$ with $\eta > 0$ such that

$$\phi'(H(x, \xi, \nu) - \zeta) \cdot \text{dist}((0, 0, 0), \partial H(x, \xi, \nu)) \geq 1$$

for all $(x, \xi, \nu) \in W$, where

$$W = \{(x, \xi, \nu) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \text{dist}((x, \xi, \nu), \Upsilon) < \epsilon\} \\ \cap \{(x, \xi, \nu) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \zeta < H(x, \xi, \nu) < \zeta + \eta\}.$$

Since $\{x^k\}$ and $\{\xi^k\}$ are bounded from Theorem 4.1 (i), and Υ is the set of accumulation points of $\{(x^k, \xi^k, x^{k-1})\}$ from Proposition 4.1 (iii), we obtain that

$$\lim_{k \rightarrow \infty} \text{dist} \left((x^k, \xi^k, x^{k-1}), \Upsilon \right) = 0.$$

Hence, there exists $\bar{k}_1 > 0$ such that $\text{dist}((x^k, \xi^k, x^{k-1}), \Upsilon) < \epsilon$ whenever $k \geq \bar{k}_1$. Similarly, from the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing and convergent to ζ , there exists $\bar{k}_2 > 0$ such that $\zeta < H(x^k, \xi^k, x^{k-1}) < \zeta + \eta$ for all $k \geq \bar{k}_2$. Taking

$\bar{k} = \max \{ \bar{k}_1, \bar{k}_2 \}$, then the sequence $\{ (x^k, \xi^k, x^{k-1}) \}$ belongs to W whenever $k \geq \bar{k}$. Hence we have

$$\phi' \left(H \left(x^k, \xi^k, x^{k-1} \right) - \zeta \right) \cdot \text{dist} \left((0, 0, 0), \partial H \left(x^k, \xi^k, x^{k-1} \right) \right) \geq 1, \quad \forall k \geq \bar{k}. \tag{4.18}$$

Since ϕ is a concave function, we see further that

$$\begin{aligned} & \left[\phi \left(H \left(x^k, \xi^k, x^{k-1} \right) - \zeta \right) - \phi \left(H \left(x^{k+1}, \xi^{k+1}, x^k \right) - \zeta \right) \right] \\ & \quad \cdot \text{dist} \left((0, 0, 0), \partial H \left(x^k, \xi^k, x^{k-1} \right) \right) \\ & \geq \phi' \left(H \left(x^k, \xi^k, x^{k-1} \right) - \zeta \right) \cdot \text{dist} \left((0, 0, 0), \partial H \left(x^k, \xi^k, x^{k-1} \right) \right) \\ & \quad \cdot \left(H \left(x^k, \xi^k, x^{k-1} \right) - H \left(x^{k+1}, \xi^{k+1}, x^k \right) \right) \\ & \geq H \left(x^k, \xi^k, x^{k-1} \right) - H \left(x^{k+1}, \xi^{k+1}, x^k \right) \geq T \left\| x^k - x^{k-1} \right\|^2, \end{aligned}$$

where the second inequality holds due to (4.18) and the fact that $\{ H(x^k, \xi^k, x^{k-1}) \}$ is nonincreasing, the third inequality follows from (4.5) and $T = \frac{\mu - L\beta_k^2}{2}$. From the above inequality and (4.10), we obtain that

$$\begin{aligned} \left\| x^k - x^{k-1} \right\|^2 & \leq \frac{A}{T} \left(\phi \left(H \left(x^k, \xi^k, x^{k-1} \right) - \zeta \right) - \phi \left(H \left(x^{k+1}, \xi^{k+1}, x^k \right) - \zeta \right) \right) \\ & \quad \cdot \left(\left\| x^{k-1} - x^{k-2} \right\| + \left\| x^k - x^{k-1} \right\| \right). \end{aligned} \tag{4.19}$$

Taking the square root on both sides of (4.19) and using the inequality of arithmetic and geometric means, we have

$$\begin{aligned} \left\| x^k - x^{k-1} \right\| & \leq \sqrt{\frac{2A}{T} \left(\phi \left(H \left(x^k, \xi^k, x^{k-1} \right) - \zeta \right) - \phi \left(H \left(x^{k+1}, \xi^{k+1}, x^k \right) - \zeta \right) \right)} \\ & \quad \times \sqrt{\frac{\left\| x^{k-1} - x^{k-2} \right\| + \left\| x^k - x^{k-1} \right\|}{2}} \\ & \leq \frac{A}{T} \left(\phi \left(H \left(x^k, \xi^k, x^{k-1} \right) - \zeta \right) - \phi \left(H \left(x^{k+1}, \xi^{k+1}, x^k \right) - \zeta \right) \right) \\ & \quad + \frac{1}{4} \left\| x^{k-1} - x^{k-2} \right\| + \frac{1}{4} \left\| x^k - x^{k-1} \right\|, \end{aligned}$$

which implies that

$$\begin{aligned} \frac{1}{2} \left\| x^k - x^{k-1} \right\| & \leq \frac{A}{T} \left(\phi \left(H \left(x^k, \xi^k, x^{k-1} \right) - \zeta \right) - \phi \left(H \left(x^{k+1}, \xi^{k+1}, x^k \right) - \zeta \right) \right) \\ & \quad + \frac{1}{4} \left(\left\| x^{k-1} - x^{k-2} \right\| - \left\| x^k - x^{k-1} \right\| \right). \end{aligned} \tag{4.20}$$

Summing the above relation from $k = \bar{k}$ to ∞ , we obtain that

$$\sum_{k=\bar{k}}^{\infty} \|x^k - x^{k-1}\| \leq \frac{2A}{T} \phi \left(H \left(x^{\bar{k}}, \xi^{\bar{k}}, x^{\bar{k}-1} \right) - \zeta \right) + \frac{1}{2} \|x^{\bar{k}-1} - x^{\bar{k}-2}\| < \infty,$$

which implies that $\sum_{k=1}^{\infty} \|x^k - x^{k-1}\| \leq \infty$, i.e., the sequence $\{x^k\}$ is a Cauchy sequence. Thus, the sequence $\{x^k\}$ converges to a stationary point of Ψ from Theorem 4.1 (iii). This completes the proof. \square

Remark 4.1 It is worth mentioning that the idea of Theorem 4.2 is similar to Theorem 2.9 proposed in [4], but the details are slightly different. In order to make the proof easier to read and understand, we write out the concrete proof.

In the following, we consider the convergence rate of the sequence $\{x^k\}$ under the assumption that the auxiliary function $H(x, \xi, v)$ is a KL function whose $\phi \in \Xi_\eta$ takes the form $\phi(s) = cs^{1-\theta}$ for some $\theta \in [0, 1)$.

Theorem 4.3 (Rate of convergence) *Let $\{x^k\}$ be the sequence generated by EPDCA for solving (1.1) and $\mu \in [L\beta^2, L]$. Suppose that $\{x^k\}$ converges to some $x^* \in \mathcal{X}$, the auxiliary function $H(x, \xi, v)$ is a KL function with ϕ in the KL inequality taking the form $\phi(s) = cs^{1-\theta}$ for some $\theta \in [0, 1)$ and $c > 0$. Then the following statements hold:*

- (i) *If $\theta = 0$, then there exists $k_0 > 0$ so that x^k is constant for $k > k_0$.*
- (ii) *If $\theta \in (0, \frac{1}{2}]$, then there exist $c_1 > 0, k_1 > 0$, and $\eta \in (0, 1)$ such that $\|x^k - x^*\| < c_1 \eta^k$ for $k > k_1$.*
- (iii) *If $\theta \in (\frac{1}{2}, 1)$, then there exist $c_2 > 0$ and $k_2 > 0$ such that $\|x^k - x^*\| < c_2 k^{-\frac{1-\theta}{2\theta-1}}$ for $k > k_2$.*

Proof (i) If $\theta = 0$, we will prove that there must exist an integer $k_0 > 0$ such that $H(x^{k_0}, \xi^{k_0}, x^{k_0-1}) = \zeta$. We assume to the contrary that for all $k > 0$, it holds that $H(x^k, \xi^k, x^{k-1}) > \zeta$. Since $\lim_{k \rightarrow \infty} x^k = x^*$ and the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing and convergent to ζ , from $\phi(s) = cs$ and the inequality (4.18), we have

$$\text{dist} \left((0, 0, 0), \partial H \left(x^k, \xi^k, x^{k-1} \right) \right) \geq \frac{1}{c},$$

for all sufficiently large k , which contradicts Proposition 4.1 (i). Therefore, there exists $k_0 > 0$ so that $H(x^{k_0}, \xi^{k_0}, x^{k_0-1}) = \zeta$. Due to the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing and convergent to ζ , it must holds that $H(x^{k_0+\bar{j}}, \xi^{k_0+\bar{j}}, x^{k_0+\bar{j}-1}) = \zeta$ for any $\bar{j} \geq 0$. Thus, we conclude from (4.5) that there exists $k_0 > 0$ so that x^k is constant for $k > k_0$. This proves (i).

(ii–iii) We next consider the case that $\theta \in (0, 1)$. If there exists $k_0 > 0$ such that $H(x^{k_0}, \xi^{k_0}, x^{k_0-1}) = \zeta$, then we can show that the sequence $\{x^k\}$ is finitely convergent as above, and the desired conclusions hold trivially. Therefore, for $\theta \in (0, 1)$, we only need to consider the case when $H(x^k, \xi^k, x^{k-1}) > \zeta$ for all $k > 0$.

Define $R_k = H(x^k, \xi^k, x^{k-1}) - \zeta$ and $S_k = \sum_{j=k}^{\infty} \|x^j - x^{j-1}\|$, where S_k is well-defined due to Theorem 4.2. Using (4.20), for any $k \geq \bar{k}$, where \bar{k} is defined as in (4.18), we have

$$\begin{aligned} S_k &= 2 \sum_{j=k}^{\infty} \frac{1}{2} \|x^j - x^{j-1}\| \\ &\leq 2 \sum_{j=k}^{\infty} \left[\frac{A}{T} \left(\phi \left(H(x^j, \xi^j, x^{j-1}) - \zeta \right) - \phi \left(H(x^{j+1}, \xi^{j+1}, x^j) - \zeta \right) \right) \right. \\ &\quad \left. + \frac{1}{4} \left(\|x^{j-1} - x^{j-2}\| - \|x^j - x^{j-1}\| \right) \right] \\ &\leq \frac{2A}{T} \phi \left(H(x^k, \xi^k, x^{k-1}) - \zeta \right) + \frac{1}{2} \|x^{k-1} - x^{k-2}\| \\ &= \frac{2A}{T} \phi(R_k) + \frac{1}{2} (S_{k-1} - S_k). \end{aligned}$$

Using the above inequality and the fact that the sequence $\{S_k\}$ is nonincreasing, we obtain that for all $k \geq \bar{k}$,

$$S_k \leq \frac{2A}{T} \phi(R_k) + \frac{1}{2} (S_{k-2} - S_k). \tag{4.21}$$

On the other hand, since $\lim_{k \rightarrow \infty} x^k = x^*$ and the sequence $\{H(x^k, \xi^k, x^{k-1})\}$ is nonincreasing and convergent to ζ , from (4.18) with $\phi(s) = cs^{1-\theta}$, we have

$$c(1 - \theta) (R_k)^{-\theta} \text{dist} \left((0, 0, 0), \partial H(x^k, \xi^k, x^{k-1}) \right) \geq 1, \tag{4.22}$$

for all sufficiently large k . In addition, from (4.10) and the definition of S_k , we obtain that

$$\begin{aligned} \text{dist} \left((0, 0, 0), \partial H(x^k, \xi^k, x^{k-1}) \right) &\leq A \left(\|x^{k-1} - x^{k-2}\| + \|x^k - x^{k-1}\| \right) \\ &= A (S_{k-1} - S_{k+1}) \leq A (S_{k-2} - S_k). \end{aligned} \tag{4.23}$$

Combining the relations (4.22) and (4.23), we further have

$$(R_k)^\theta \leq A \cdot c(1 - \theta) (S_{k-2} - S_k).$$

Raising to a power of $\frac{1-\theta}{\theta}$ and multiplying by c to both sides of the above inequality, we obtain that

$$c (R_k)^{1-\theta} \leq c (A \cdot c(1 - \theta) \cdot (S_{k-2} - S_k))^{\frac{1-\theta}{\theta}}.$$

Combining this with (4.21) and according to the fact that $\phi(R_k) = c(R_k)^{1-\theta}$, we see that for all sufficiently large k ,

$$S_k \leq A_1 (S_{k-2} - S_k)^{\frac{1-\theta}{\theta}} + \frac{1}{2} (S_{k-2} - S_k) \leq A_1 (S_{k-2} - S_k)^{\frac{1-\theta}{\theta}} + S_{k-2} - S_k, \tag{4.24}$$

where $A_1 = \frac{2A}{T} c(A \cdot c(1 - \theta))^{\frac{1-\theta}{\theta}}$.

(ii) When $\theta \in (0, \frac{1}{2}]$, the inequality $\frac{1-\theta}{\theta} \geq 1$ holds. According to the fact that $\|x^k - x^{k-1}\| \rightarrow 0$ from Theorem 4.1 (ii) and the definition of S_k , we obtain that $S_{k-2} - S_k \rightarrow 0$. Combining these and (4.24), we conclude that there exists $k_1 > 0$ so that for all $k \geq k_1$, it holds that $S_k \leq (A_1 + 1)(S_{k-2} - S_k)$, which implies that $S_k \leq \frac{A_1+1}{A_1+2} S_{k-2}$. Therefore, we obtain that for any $k \geq k_1$,

$$\|x^k - x^*\| \leq \sum_{j=k}^{\infty} \|x^j - x^{j-1}\| = S_k \leq S_{k_1-2} \left(\sqrt{\frac{A_1 + 1}{A_1 + 2}} \right)^{k-k_1+1}.$$

This proves (ii).

(iii) When $\theta \in (\frac{1}{2}, 1)$, we deduce that $\frac{1-\theta}{\theta} < 1$. From this with (4.24) and the fact $S_{k-2} - S_k \rightarrow 0$, we see that there exists $k_2 > 0$ such that for all $k \geq k_2$,

$$S_k \leq A_1 (S_{k-2} - S_k)^{\frac{1-\theta}{\theta}} + (S_{k-2} - S_k)^{\frac{1-\theta}{\theta}} = (A_1 + 1) (S_{k-2} - S_k)^{\frac{1-\theta}{\theta}}.$$

Raising the above inequality to a power of $\frac{\theta}{1-\theta}$ to both sides, we have for any $k \geq k_2$,

$$S_k^{\frac{\theta}{1-\theta}} \leq A_2 (S_{k-2} - S_k),$$

where $A_2 = (A_1 + 1)^{\frac{\theta}{1-\theta}}$. From [2, Theorem 2], we obtain that there exists $M > 0$ and all sufficiently large k , the inequality $S_k \leq M k^{-\frac{1-\theta}{2\theta-1}}$ holds. This completes the proof. \square

5 Numerical experiments

In this section, we perform numerical experiments to illustrate the efficiency of EPDCA for solving the optimization problem (1.1). We compare our algorithm with pDCAe proposed in [31], which has different choices of the extrapolation parameters $\{\beta_k\}$ given in (1.6), and the general iterative shrinkage and thresholding algorithm (GIST) proposed in [10] on a DC regularized least squares problem as follows:

$$\min \left\{ \Psi(x) := \frac{1}{2} \|Ax - b\|^2 + g_n(x) - g_s(x) \mid x \in \mathbb{R}^n \right\}, \tag{5.1}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $g_n : \mathbb{R}^n \rightarrow \mathbb{R}$ is a proper closed convex function and $g_s : \mathbb{R}^n \rightarrow \mathbb{R}$ is a continuous convex function. More concrete examples can be found in [31].

In the following we consider two different classes of regularizers. One is the ℓ_{1-2} regularized least squares problem [32]:

$$\min \left\{ \Psi_{\ell_{1-2}}(x) = \frac{1}{2} \|Ax - b\|^2 + \lambda \|x\|_1 - \lambda \|x\| \mid x \in \mathbb{R}^n \right\}, \tag{5.2}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\lambda > 0$ is the regularization parameter. It is easy to see that (5.2) is in the form of (5.1) with $g_n(x) = \lambda \|x\|_1$ and $g_s(x) = \lambda \|x\|$. We assume that A does not have zero columns so that $\Psi_{\ell_{1-2}}$ is level-bounded [18]. So by Theorem 4.1, the sequence $\{x^k\}$ generated by our method to solve problem (5.2) is bounded, and any accumulation point of $\{x^k\}$ is a stationary point. In addition, noting from [31, Example 4.1] that if chosen $\lambda < \frac{1}{2} \|A^T b\|_\infty$, we obtain that the sequence $\{x^k\}$ generated by our method for solving problem (5.2) is globally convergent.

The other is the logarithmic regularized least square problem [6]:

$$\min \left\{ \Psi_{log}(x) = \frac{1}{2} \|Ax - b\|^2 + \sum_{i=1}^n [\lambda \log(|x_i| + \epsilon) - \lambda \log \epsilon] \mid x \in \mathbb{R}^n \right\}, \tag{5.3}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\epsilon > 0$ is a constant and $\lambda > 0$ is the regularization parameter. We observe that (5.3) is in the form of (5.1) with $g_n(x) = \frac{\lambda}{\epsilon} \|x\|_1$ and $g_s(x) = \sum_{i=1}^n \lambda \left[\frac{|x_i|}{\epsilon} - \log(|x_i| + \epsilon) + \log \epsilon \right]$. It is easy to show that Ψ_{log} is level-bounded. This together with Theorem 4.2, we see that the sequence $\{x^k\}$ generated by EPDCA to solve problem (5.3) is globally convergent.

In the following, we perform several numerical tests to compare the behaviors of our method with pDCAe and GIST for solving (5.2) and (5.3), respectively. All the numerical experiments have been performed in Matlab 2018b on a 64-bit PC with an Intel(R) Core(TM) i7-4790 CPU (3.60GHz) and 32GB of RAM.

In our numerical experiments, for given (m, n, s) , we first randomly generate a matrix $A \in \mathbb{R}^{m \times n}$ with i.i.d. standard Gaussian entries, and each column of A is normalized. Then a s -sparsity vector $\hat{y} \in \mathbb{R}^n$ is uniformly generated at random and let $y = \text{sign}(\hat{y})$. Finally, we generate $b = Ay + 0.01 \cdot \hat{n}$, where $\hat{n} \in \mathbb{R}^m$ is a random vector with i.i.d. standard Gaussian entries.

For each triple $(m, n, s) = (720i, 2560i, 80i)$, $i = 1, 2, \dots, 8$, we generate 50 instances randomly as described above and compare all three methods in terms of the average performance over the instances. We choose the Lipschitz constant L of ∇f as $\lambda_{\max}(A^T A)$, initialize these algorithms at the origin, and terminate these algorithms when the iterations reach the maximum number of 5000 times or $\|x^k - x^{k-1}\| / \max\{1, \|x^k\|\} \leq 10^{-5}$.

With regard to parameters, for EPDCA, we choose the extrapolation parameters $\{\beta_k\}$ as in (3.2) and set $a = 0.8, b = 5, \omega = 0.98$. In addition, we adopt the restart strategy and set $K = 500$ in Remark 3.1 (iv) on our method to guarantee $\bar{\beta} = \sup_k \beta_k <$

Table 1 Solving (5.2) on 50 random instances, $\lambda = 1 \times 10^{-3}$

Problem size m	n	s	$t_{y,max}$	iter		CPU time			fval	
				GIST	pDCA _e	GIST	EPDCA	pDCA _e	GIST	pDCA _e
720	2560	80	0.1	929	600	402	0.5	0.7	6.0973e-02	6.0965e-02
1440	5120	160	0.9	923	602	438	2.4	3.2	1.2174e-01	1.2173e-01
2160	7680	240	0.8	932	602	449	5.6	7.5	1.8538e-01	1.8536e-01
2880	10,240	320	2.1	929	602	439	10.5	14.6	2.4268e-01	2.4265e-01
3600	12,800	400	3.6	905	602	425	32.5	23.8	3.1455e-01	3.1451e-01
4320	15,360	480	5.8	991	602	472	24.9	28.9	3.8343e-01	3.8337e-01
5040	17,920	560	8.7	946	602	485	31.6	39.4	4.4599e-01	4.4594e-01
5760	20,480	640	11.8	927	602	484	39.3	48.8	4.9912e-01	4.9906e-01

Table 2 Solving (5.2) on 50 random instances, $\lambda = 5 \times 10^{-4}$

Problem size m	n	s	$t_{y,\max}$	iter		CPU time			fval			
				GIST	pDCA _e	EPDCA	GIST	pDCA _e	EPDCA	GIST	pDCA _e	EPDCA
720	2560	80	0.1	1733	902	603	2.1	0.8	0.5	3.0561e-02	3.0551e-02	3.0551e-02
1440	5120	160	0.8	1744	902	603	10.9	4.5	3.0	6.1028e-02	6.1003e-02	6.1002e-02
2160	7680	240	1.0	1765	1002	574	29.0	11.0	6.0	9.2934e-02	9.2884e-02	9.2889e-02
2880	10,240	320	2.1	1689	903	602	45.5	17.4	11.3	1.2166e-01	1.2161e-01	1.2160e-01
3600	12,800	400	3.3	1712	903	596	72.2	29.4	19.1	1.5768e-01	1.5761e-01	1.5761e-01
4320	15,360	480	5.9	1764	1002	587	106.7	47.0	27.3	1.9219e-01	1.9209e-01	1.9210e-01
5040	17,920	560	10.0	1791	903	603	154.1	64.5	41.7	2.2354e-01	2.2347e-01	2.2346e-01
5760	20,480	640	15.0	1728	902	603	202.5	73.6	49.2	2.5017e-01	2.5008e-01	2.5008e-01

Table 3 Solving (5.3) on 50 random instances, $\lambda = 1 \times 10^{-3}$

Problem size		$t_{y,max}$	iter		CPU time			fval		
m	n		s	GIST	pDCA _e	EPDCA	GIST	pDCA _e	EPDCA	
720	2560	80	478	343	308	1.0	0.5	0.4	7.6369e-02	7.6368e-02
1440	5120	160	467	396	304	3.6	2.3	1.7	1.4746e-01	1.4745e-01
2160	7680	240	507	402	303	9.1	5.2	3.9	2.3269e-01	2.3268e-01
2880	10,240	320	430	402	281	13.6	9.3	6.6	2.9620e-01	2.9619e-01
3600	12,800	400	471	402	307	25.2	15.1	11.3	3.6935e-01	3.6934e-01
4320	15,360	480	480	402	303	30.2	18.6	14.2	4.5260e-01	4.5259e-01
5040	17,920	560	473	402	288	40.4	25.5	18.3	5.2674e-01	5.2673e-01
5760	20,480	640	464	402	292	54.5	34.9	25.5	5.9371e-01	5.9368e-01

Table 4 Solving (5.3) on 50 random instances, $\lambda = 5 \times 10^{-4}$

Problem size m	n	s	$t_{y,\max}$	iter		CPU time			fval		
				GIST	pDCA _e	GIST	pDCA _e	EPDCA	GIST	pDCA _e	EPDCA
720	2560	80	0.1	844	602	1.2	0.5	0.4	3.8445e-02	3.8438e-02	3.8440e-02
1440	5120	160	0.8	860	602	5.8	3.0	2.1	7.4229e-02	7.4217e-02	7.4217e-02
2160	7680	240	1.0	924	602	15.3	6.9	5.5	1.1710e-01	1.1708e-01	1.1709e-01
2880	10,240	320	2.0	848	602	23.7	12.1	9.0	1.4909e-01	1.4906e-01	1.4906e-01
3600	12,800	400	3.8	871	602	38.7	19.7	16.3	1.8593e-01	1.8591e-01	1.8591e-01
4320	15,360	480	6.6	874	602	56.2	27.8	21.9	2.2783e-01	2.2780e-01	2.2781e-01
5040	17,920	560	9.7	886	602	80.5	38.2	31.3	2.6511e-01	2.6506e-01	2.6507e-01
5760	20,480	640	14.4	860	602	94.7	49.2	35.7	2.9887e-01	2.9881e-01	2.9881e-01

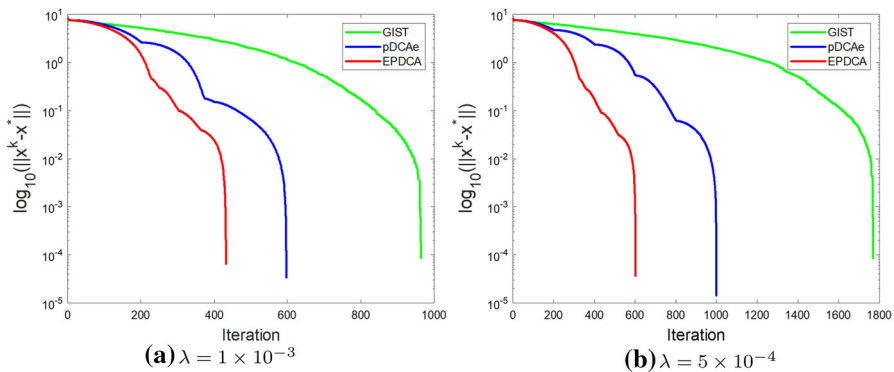


Fig. 1 Comparison of GIST, pDCAe and EPDCA with the problem sizes $m = 720$, $n = 2560$ and $s = 80$ on solving ℓ_{1-2} regularized least squares problem (5.2)

1. For pDCAe, the extrapolation parameters are chosen as the extrapolation coefficients used in [31]. For GIST, the algorithmic framework and the setting of parameters can be found in [19, Appendix A].

For the ℓ_{1-2} regularized least squares problem, the numerical results are presented in Tables 1 and 2, which correspond with $\lambda = 1 \times 10^{-3}$ and $\lambda = 5 \times 10^{-4}$ respectively. While for the logarithmic regularized least square problem, the numerical results are presented in Tables 3 and 4, which also correspond with $\lambda = 1 \times 10^{-3}$ and $\lambda = 5 \times 10^{-4}$ respectively. In all the tables, we give the computation of $L = \lambda_{\max}(A^T A)$ ($t_{\lambda_{\max}}$), the number of iterations, CPU times in seconds which does not include the time for computing L , and the function values at termination, averaged over the 50 random instances.

In Fig. 1, We plot $\|x^k - x^*\|$ against the number of iterations to compare with the convergence performance of each algorithm on solving ℓ_{1-2} regularized least squares problem (5.2), where x^* denotes the approximate solution obtained at termination of the respective method. We choose $\lambda = 1 \times 10^{-3}$ in Fig. 1a and $\lambda = 5 \times 10^{-4}$ in Fig. 1b, which correspond with Tables 1 and 2, respectively. In Fig. 2, We plot $\|x^k - x^*\|$ against the number of iterations to compare with the convergence performance of each algorithm on solving logarithmic regularized least square problem (5.3). We choose $\lambda = 1 \times 10^{-3}$ in Fig. 2a and $\lambda = 5 \times 10^{-4}$ in Fig. 2b, which correspond with Tables 3 and 4, respectively. The green, blue and red line correspond to the three methods, GIST, pDCAe and EPDCA, respectively.

We conclude from these tables and figures that our method takes fewer iterations and less CPU time than pDCAe and GIST to get a good approximation of the minimum value of the two test regularized least squares problems.

6 Conclusion

In this paper, based on the generalized Nesterov momentum scheme for the accelerated forward–backward algorithm (AFBA) proposed in [16] and pDCAe [31] for DC

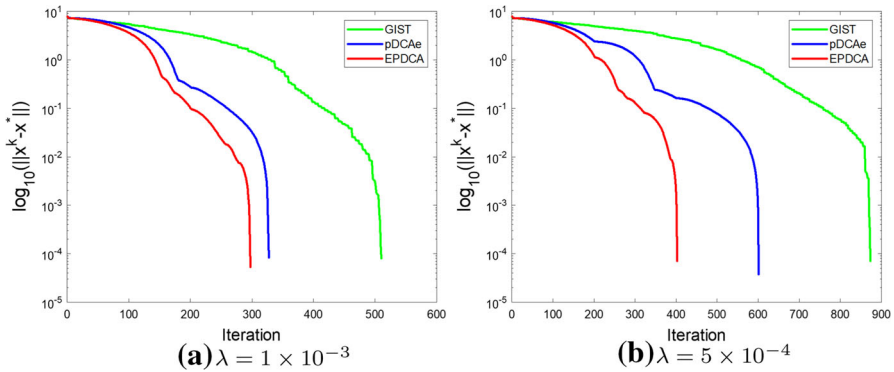


Fig. 2 Comparison of GIST, pDCAe and EPDCA with the problem sizes $m = 720$, $n = 2560$ and $s = 80$ on solving logarithmic regularized least square problem (5.3)

programming, we propose a new extrapolated proximal difference-of-convex algorithm (EPDCA), which equips a more general setting of the extrapolation parameters $\{\beta_k\}$ for solving the DC optimization problem (1.1). We establish global subsequential convergence of the sequence generated by our method. In addition, by assuming the Kurdyka–Łojasiewicz property of the auxiliary function, we establish global convergence of the sequence generated by EPDCA and analyze its convergence rate, which can allow g_s to be a nonsmooth function. To evaluate the performance of the proposed method, we consider two classes of DC regularized least squares problems including the ℓ_{1-2} regularized least squares problem and logarithmic regularized least square problem. Numerical results illustrate the efficiency of our method and its superiority over other well-known methods.

In future research, we can utilize the advantage of nonmonotone line search technique in [23] to enlarge the step size and meanwhile speed up the convergence of the proposed algorithm. Furthermore, the difference-of-Moreau-envelopes smoothing technique has been developed to solve DC programming in [30]. These variational methods may inspire further improvements and extensions.

Acknowledgements The authors would like to thank the editors and anonymous reviewers for their insight and helpful comments and suggestions which improve the quality of the paper. This work is also supported in part by NSFC11801131, Natural Science Foundation of Hebei Province (Grant No. A2019202229), a scientific grant of Hebei Educational Committee (QN2018101).

Declarations

Conflict of interest The authors declare that there are no conflicts of interest with regards to the publication of this paper.

References

1. Alvarado, A., Scutari, G., Pang, J.S.: A new decomposition method for multiuser DC programming and its applications. *IEEE Trans. Signal Process.* **62**, 2984–2998 (2014)

2. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program.* **116**, 5–16 (2009)
3. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. *Math. Oper. Res.* **3**, 438–457 (2010)
4. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. *Math. Program.* **137**, 91–129 (2013)
5. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.* **146**, 459–494 (2014)
6. Candès, E.J., Wakin, M., Boyd, S.: Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.* **14**, 877–905 (2008)
7. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**, 89–97 (2004)
8. Chambolle, A., Dossal, C.: On the convergence of the iterates of “FISTA”. *J. Optim. Theory Appl.* **166**, 25 (2015)
9. Gaso, G., Rakotomamonjy, A., Canu, S.: Recovering sparse signals with a certain family of nonconvex penalties and DC programming. *IEEE Trans. Signal Process.* **57**, 4686–4698 (2009)
10. Gong, P., Zhang, C., Lu, Z., Huang, J.Z., Ye, J.: A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In: *ICML* (2013)
11. Gotoh, J., Takeda, A., Tono, K.: DC formulations and algorithms for sparse optimization problems. Preprint, METR 2015-27, Department of Mathematical Informatics, University of Tokyo. <http://www.keisu.t.u-tokyo.ac.jp/research/techrep/index.html>
12. Le Thi, H.A., Nguyen, M.C.: DCA based algorithms for feature selection in multi-class support vector machine. *Ann. Oper. Res.* **249**, 273–300 (2017)
13. Le Thi, H.A., Pham, D.T.: The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Ann. Oper. Res.* **133**, 23–46 (2005)
14. Le Thi, H.A., Pham, D.T.: DC programming and DCA: thirty years of developments. *Math. Program.* **169**, 5–68 (2018)
15. Le Thi, H.A., Pham, D.T., Le, H.M., Vo, X.Y.: DC approximation approaches for sparse optimization. *Eur. J. Oper. Res.* **244**, 26–46 (2015)
16. Lin, Y., Li, S., Zhang, Y.Z.: Convergence rate analysis of accelerated forward–backward algorithm with generalized Nesterov momentum scheme. [arXiv: 2112.05873](https://arxiv.org/abs/2112.05873)
17. Lin, Y., Schmidlein, C.R., Li, Q., Li, S., Xu, Y.: A Krasnoselskii–Mann algorithm with an improved EM preconditioner for PET image reconstruction. *IEEE Trans. Med. Imaging* **38**, 2114–2126 (2019)
18. Liu, T., Pong, T.K.: Further properties of the forward–backward envelope with applications to difference-of-convex programming. *Comput. Optim. Appl.* **67**, 489–520 (2017)
19. Liu, T., Pong, T.K., Takeda, A.: A successive difference-of-convex approximation method for a class of nonconvex nonsmooth optimization problems. *Math. Program.* **176**, 339–367 (2018)
20. Liu, T., Pong, T.K., Takeda, A.: A refined convergence analysis of pDCA_e with applications to simultaneous sparse recovery and outlier detection. *Math. Program.* **176**, 339–367 (2019)
21. Lou, Y., Zeng, T., Osher, S., Xin, J.: A weighted difference of anisotropic and isotropic total variation model for image processing. *SIAM J. Imaging Sci.* **8**, 1798–1823 (2015)
22. Lu, Z., Zhou, Z., Sun, Z.: Enhanced proximal DC algorithms with extrapolation for a class of structured nonsmooth DC minimization. *Math. Program.* **176**, 369–401 (2019)
23. Lu, Z., Zhou, Z.: Nonmonotone enhanced proximal DC algorithms for a class of structured nonsmooth DC programming. *SIAM J. Optim.* **29**, 2725–2752 (2019)
24. O’Donoghue, B., Candès, E.J.: Adaptive restart for accelerated gradient schemes. *Found. Comput. Math.* **15**, 715–732 (2015)
25. Pang, J.-S., Razaviyayn, M., Alvarado, A.: Computing B-stationary points of nonsmooth DC programs. *Math. Oper. Res.* **42**, 95–118 (2017)
26. Pham, D.T., Le Thi, H.A.: Convex analysis approach to D.C. programming: theory, algorithms and applications. *Acta Math. Vietnam.* **22**, 289–355 (1997)
27. Pham, D.T., Le Thi, H.A.: A D.C. optimization algorithm for solving the trust-region subproblem. *SIAM J. Optim.* **8**, 476–505 (1998)
28. Rockafellar, R.T., Wets, R.: *Variational Analysis*. Grundlehren der Mathematischen Wissenschaften. Springer, Berlin (1998)

29. Sanjabi, M., Razaviyayn, M., Luo, Z.-Q.: Optimal joint base station assignment and beamforming for heterogeneous networks. *IEEE Trans. Signal Process.* **62**, 1950–1961 (2014)
30. Sun, K., Sun, X.A.: Algorithms for difference-of-convex (DC) programs based on difference-of-Moreau-envelopes smoothing. [arXiv: 2104.01470](https://arxiv.org/abs/2104.01470)
31. Wen, B., Chen, X., Pong, T.K.: A proximal difference-of-convex algorithm with extrapolation. *Comput. Optim. Appl.* **62**, 297–324 (2018)
32. Yin, P., Lou, Y., He, Q., Xin, J.: Minimization of ℓ_{1-2} for compressed sensing. *SIAM J. Sci. Comput.* **37**, A536–A563 (2015)

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.