Commun. Optim. Theory 2025 (2025) 7 https://doi.org/10.23952/cot.2025.7

MATHRES Communications in Optimization Theory
Available online at http://cot.mathres.org

AN INERTIAL BREGMAN PROXIMAL DC ALGORITHM FOR GENERALIZED DC PROGRAMMING WITH APPLICATION TO DATA COMPLETION

CHENJIAN PAN[1], YINGXIN ZHOU[1], HONGJIN HE[1,*], CHEN LING[2]

[1]School of Mathematics and Statistics, Ningbo University, Ningbo, 315211, China
[2]Department of Mathematics, Hangzhou Dianzi University, Hangzhou, 310018, China

Dedicated to Professor Elijah Lucien Polak on the occasion of his 90th birthday

Abstract. In this paper, we consider a class of generalized difference-of-convex functions (DC) programming, whose objective is the difference of two convex (not necessarily smooth) functions plus a decomposable (possibly nonconvex) function with Lipschitz gradient. By employing the Fenchel-Young inequality and Moreau decomposition theorem, we introduce an inertial Bregman proximal DC algorithm to solve the problem under consideration. Our algorithmic framework is able to fully exploit the decomposable structure of the generalized DC programming such that each subproblem of the algorithm is enough easy in many cases. Theoretically, we show that the sequence generated by the proposed algorithm globally converges to a critical point under the Kurdyka-Łojasiewicz condition. A series of numerical results demonstrate that our algorithm runs efficiently on matrix and tensor completion problems.

Keywords. Difference-of-convex programming; Fenchel-Young inequality; Moreau decomposition; Bregman distance; Tensor completion.

2020 Mathematics Subject Classification. 65K10, 90C26.

1. INTRODUCTION

In this paper, we are interested in a class of generalized difference-of-convex functions (DC) programming, which refers to

$$\min_{x\in\mathbb{R}^n} \Phi(x) := f(x) - g(x) + h(x), \tag{1.1}$$

where $f(x)$ and $g(x)$ are proper closed convex functions, and $h(x)$ is a differentiable (possibly nonconvex) function, which is further assumed to be liberally decomposable in the sense of $h(x) = h^+(x) - h^-(x)$ with gradient Lipschitz continuity modulus L_h^+ and L_h^- for $h^+(x)$ and $h^-(x)$, respectively. The generalized DC programming (1.1) has widespread applications in signal/image processing and statistical/machine learning [1, 13, 14, 25, 29], to name just a few.

*Corresponding author.
E-mail address: hehongjin@nbu.edu.cn (H. He).
Received December 5, 2023; Accepted April 26, 2024.

When the last decomposable part $h(x)$ in (1.1) vanishes, a benchmark solver to circumvent the nonconvexity of the resulting problem is the so-named DC algorithm (DCA) originally proposed by Pham Dinh [21] (see also [20]), where $-g(x)$ is replaced by its linear approximation at x^k so that the DCA enjoys a convex subproblem. Naturally, a straightforward application of the DCA to (1.1) produces the following iterative scheme:

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) + h(x) - \langle \xi^{k+1}, x - x^k \rangle \right\}, \tag{1.2}$$

where ξ^{k+1} is a subgradient of $g(x)$ at the iterate x^k. However, the iterative scheme (1.2) is not necessarily implementable in practice due to the simultaneous appearance of $f(x)$ and $h(x)$, especially the possible nonconvexity of $h(x)$. Hence, one more reasonable way is to exploit the differentiability of $h(x)$. In this way, a customized application of the DCA to (1.1) yields the following iterative scheme:

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

where $\nabla h(x^k)$ is the gradient of $h(x)$ at the iterate x^k. Clearly, (1.2) and (1.3) present two different applications of the original DCA. However, (1.3) is more implementable than (1.2) in many cases. To some extent, we can regard the seminal DCA as an algorithmic paradigm. Therefore, how to exploit the DC decomposition in algorithmic design and implementation is of importance for efficiently solving DC programming and general nonconvex optimization problems [20]. Consequently, as studied in [10], we can also reformulate (1.1) as a canonical DC programming:

$$\min_{x\in\mathbb{R}^n} \Phi(x) := \underbrace{\left(f(x) + \frac{L_h}{2}\|x\|^2 \right)}_{\widehat{f}(x)} - \underbrace{\left(g(x) - h(x) + \frac{L_h}{2}\|x\|^2 \right)}_{\widehat{g}(x)}, \tag{1.4}$$

where L_h is the Lipschitz continuity constant of $h(x)$ so that both $\widehat{f}(x)$ and $\widehat{g}(x)$ are convex functions. Then, directly applying the DCA to (1.4) leads to

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{L_h}{2} \left\| x - x^k + \frac{1}{L_h}\left(\nabla h(x^k) - \xi^{k+1} \right) \right\|^2 \right\}, \tag{1.5}$$

which shares the same idea of the proximal DCA introduced in [26]. Moreover, such a scheme immediately reduces to the classical proximal gradient method [4] when $g(x)$ vanishes. Comparatively, (1.5) is more practical than (1.2) and (1.3) for dealing with nonsmooth optimization problems. Although (1.5) enjoys an easier subproblem, it is essentially a gradient-like algorithm which runs slowly in many cases (e.g., see [18] for the discussion on convex optimization). Therefore, many researchers are motivated to develop faster algorithms for DC programming, e.g., see [7, 15, 16, 17, 22, 23, 27, 28] and references therein.

In this paper, we follow the extrapolation idea and the Bregman proximal technique to propose an inertial Bregman Proximal DC Algorithm (iBPDCA) for the generalized DC programming (1.1). It is noteworthy that most DCA-like algorithms must select a subgradient ξ^{k+1}, which is an element of the subdifferential $\partial g(x^k)$ when dealing with nonsmooth DC programming. In this situation, it is not an easy task to determine which one is the most ideal subgradient for approximating $-g(x)$. Therefore, to circumvent this difficulty, we accordingly employ the

well known Fenchel-Young inequality to introduce a surrogate for $-g(x)$. Then, we add a proximal term to enhance the subproblem for updating ξ^{k+1}. One direct benefit is that we will no longer worry about the selection of the subgradient since it will be updated uniquely as long as $g(x)$ is convex. Another remarkable benefit is that we can directly exploit the possibly explicit proximal operator of $g(x)$ by the extended Moreau decomposition theorem [3], which is able to make our algorithm more implementable in practice (see Section 4). Moreover, we shall highlight that our algorithm also possesses an extrapolation step for the algorithmic acceleration, which will be verified in Section 4. Thirdly, we should mention that $h(x)$ is assumed to be liberally decomposable of the form $h(x) = h^+(x) - h^-(x)$. Such a decomposable form allows us to freely reformulate the DC optimization model so that which, together with the Bregman proximal term, can gainfully help us design customized algorithms for solving the problem under consideration. Theoretically, we show that the sequence generated by the proposed iBPDCA converges to a critical point of (1.1). Finally, we apply the iBPDCA to low-rank matrix and tensor data completion. A series of computational results demonstrate that our algorithm works well in solving DC optimization problems.

The structure of this paper is divided into five parts. In Section 2, we summarize some notations and recall some basic definitions and properties. In Section 3, we introduce our iBPDCA algorithm for (1.1) and study its convergence. In Section 4, we conduct the numerical performance of our iBPDCA on data completion. Finally, some concluding remarks are summarized in Section 5.

2. Preliminaries

In this section, we briefly review several basic notions, definitions, and properties of Kurdyka-Łojasiewicz inequality and other related mathematical tools that will be used throughout this paper.

We denote $\mathbb{R}^n$ as the n-dimensional Euclidean space with inner product $\langle\cdot,\cdot\rangle$, and we use $\|\cdot\|_1$, $\|\cdot\|$, $\|\cdot\|_*$ and $\|\cdot\|_F$ to denote the ℓ_1 norm, ℓ_2 norm, nuclear norm and Frobenius norm for vectors or matrices, respectively. For any subset $\mathbb{S}\subset\mathbb{R}^n$ and any point $x\in\mathbb{R}^n$, $\mathbf{dist}(x,\mathbb{S}) = \inf\{\|y-x\| \mid y\in\mathbb{S}\}$ is used to define the distance from x to $\mathbb{S}$. As for a matrix $X\in\mathbb{R}^{m\times n}$, the transpose of X is denoted by $X^\top$.

For an extended-real-value function $f:\mathbb{R}^n\to[-\infty,+\infty]$, we denote the domain and level set of f by $\mathbf{dom}(f) := \{x\in\mathbb{R}^n \mid f(x)<\infty\}$ and $\mathbf{Lev}_f(\alpha_k) = \{x\in\mathbb{R}^n \mid f(x)\le\alpha_k\}$, respectively. It is documented in [24] that a proper closed function f is said to be level-bounded if $\mathbf{Lev}_f(\alpha_k)$ is bounded. Moreover, a proper function is closed if it is lower semicontinuous.

Definition 2.1. Let $f:\mathbb{R}^n\to(-\infty,\infty]$ be a proper and lower semicontinuous function.

(i) For each $x\in\mathbf{dom}(f)$, the Fréchet subdifferential $\widehat{\partial}f(x)$ of f at x is defined by

$$\widehat{\partial}f(x) := \left\{\xi\in\mathbb{R}^n \;\middle|\; \liminf_{\substack{y\neq x\\ y\to x}} \frac{f(y)-f(x)-\langle\xi,y-x\rangle}{\|y-x\|}\ge 0\right\}.$$

In particular, for $x\notin\mathbf{dom}(f)$, we set $\widehat{\partial}f(x)=\emptyset$.

(ii) The limiting subdifferential $\partial f(x)$ of f at $x \in \mathbf{dom}(f)$ is defined by

$$\partial f(x) := \left\{ \xi \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \exists \left(x^k, f(x^k)\right) \to (x, f(x)), \xi^k \in \widehat{\partial} f(x^k) \\ \text{such that } \xi^k \to \xi \text{ as } k \to \infty \end{array} \right\}.$$

If f is continuously differentiable, then $\partial f = \{\nabla f\}$ (e.g., see [24, Exercise 8.8(b)]), where ∇f is denoted as the gradient of f. Moreover, when f is convex, the limiting subdifferential reduces to the classical subdifferential in convex analysis (see [24, Proposition 8.12]). Apart from that, Definition 2.1 implies $\widehat{\partial} f \subset \partial f$ for each $x \in \mathbb{R}^n$. It is worth noting that the set $\widehat{\partial} f$ is both convex and closed while ∂f is closed (e.g., see [24, Theorem 8.6]).

A necessary but not sufficient condition for $x^\star \in \mathbb{R}^n$ to be a local minimizer of f is that $0 \in \partial f(x^\star)$, where $x^\star$ is called a stationary point of f. Throughout this paper, the set of stationary points of f is denoted by $\Xi(f)$. Additionally, as defined in the literature (e.g., [14, 28]), we say that $x^\star \in \mathbb{R}^n$ is a critical point of $\Phi := f - g + h$ given in (1.1) if

$$0 \in \partial f(x^\star) - \partial g(x^\star) + \nabla h(x^\star), \tag{2.1}$$

or equivalently, $\{\partial f(x^\star) + \nabla h(x^\star)\} \cap \partial g(x^\star) \neq \emptyset$. Then, the set of critical points of Φ is denoted by $\mathbf{crit}(\Phi)$. Here, we should emphasize that a stationary point of Φ, i.e., $0 \in \partial(f-g)(x^\star) + \nabla h(x^\star)$, is sharper than a critical point of Φ satisfying (2.1). If g is assumed to be continuously differentiable over $\mathbf{dom}(f+h)$, i.e., $\partial g(x^\star) = \{\nabla g(x^\star)\}$, then a critical point of Φ is also a stationary point of Φ (see [19]). In our paper, we do not assume g being a smooth function and adopt the critical point (2.1) for simplicity.

Before stating the Kurdyka-Łojasiewicz (KŁ) property, we first use Υ_ζ to denote the set of all concave continuous functions $\phi : [0, \zeta) \to \mathbb{R}_+$ that are continuously differentiable on $(0, \zeta)$ with positive derivatives and satisfy $\phi(0) = 0$. Below, we recall the widely used KŁ property (e.g., see [2, Definition 3.1]) that plays an instrumental role in convergence analysis of many nonconvex optimization methods.

Definition 2.2 (KŁ property and KŁ function)**.** Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper and lower semicontinuous function.

(i) We say that the function f has the KŁ property at $\bar{x} \in \mathbf{dom}(\partial f)$ if there exist $\zeta \in (0, \infty]$, a neighborhood $\mathcal{N}$ of $\bar{x}$ and a continuous concave function $\phi \in \Upsilon_\zeta$ such that for all $x \in \mathcal{N}$ with $f(\bar{x}) < f(x) < f(\bar{x}) + \zeta$, the following KŁ inequality holds, i.e.,

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

(ii) If f satisfies the above KŁ property at each point in $\mathbf{dom}(\partial f)$, then f is called a KŁ function.

Hereafter, we recall the uniformized KŁ property introduced in [5, Lemma 6].

Lemma 2.3 (uniformized KŁ property)**.** *Let Γ be a compact set and let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper and lower semicontinuous function. Assume that f is constant on Γ and satisfies the KŁ property at each point of Γ. Then, there exist $\varsigma > 0$, $\zeta > 0$ and $\phi \in \Upsilon_\zeta$ such that for each $\bar{x} \in \Gamma$ the following KŁ inequality*

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

holds for any $x \in \{x \in \mathbb{R}^n \mid \mathbf{dist}(x, \Gamma) < \varsigma\} \cap \{x \in \mathbb{R}^n \mid f(\bar{x}) < f(x) < f(\bar{x}) + \zeta\}$.

Note that we assume $h(x)$ being differentiable. Here, we recall the famous descent lemma for the so-called L-smooth optimization problems (see more details in [3, Lemma 5.7]).

Lemma 2.4. *Let $f:\mathbb{R}^n\to(-\infty,\infty]$ be an L-smooth function ($L\geq 0$) over a given convex set $\mathbb{S}$, i.e., f is differentiable over $\mathbb{S}$ and satisfies*

$$\|\nabla f(x)-\nabla f(y)\|\leq L\|x-y\|,\quad \forall x,y\in\mathbb{S}.$$

Then, for any $x,y\in\mathbb{S}$, we have

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

Now, we present the definition of conjugate function, which plays a significant role in finding a surrogate of $-g(x)$ for our algorithmic design.

Definition 2.5. Let $f:\mathbb{R}^n\to[-\infty,\infty]$ be an extended real-valued function. The conjugate function of f is defined by

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

From [3, Chapter 4], f^* is proper, lower semicontinuous and convex provided f is proper, lower semicontinuous and convex. Furthermore, Definition 2.5 implies

$$f(x)+f^*(y)\geq\langle x,y\rangle \tag{2.2}$$

holds. Especially, the equality holds if and only if $y\in\partial f(x)$. Besides, if f is proper, closed and convex, then for any x and y, one has $y\in\partial f(x)$ if and only if $x\in\partial f^*(y)$ (e.g., see [3, Theorem 4.20]). In the literature, (2.2) is often referred to as the Fenchel-Young inequality, which will be vital for our algorithmic design.

Definition 2.6 (proximal mapping)**.** Given a function $\vartheta:\mathbb{R}^n\to(-\infty,\infty]$ and a scalar $t>0$, the proximal mapping of $t\vartheta$ is the operator given by

$$\mathbf{Prox}_{t\vartheta}(\boldsymbol{a})=\arg\min_{x\in\mathbb{R}^n}\left\{t\vartheta(x)+\frac{1}{2}\|x-\boldsymbol{a}\|^2\right\}\quad\text{for any }\boldsymbol{a}\in\mathbb{R}^n. \tag{2.3}$$

Theorem 2.7 (extended Moreau decomposition theorem)**.** *Let $f:\mathbb{R}^n\to(-\infty,\infty]$ be proper closed and convex, and let the scalar $t>0$. Then for any $x\in\mathbb{R}^n$, we have*

$$\mathbf{Prox}_{tf}(x)+t\mathbf{Prox}_{t^{-1}f^*}\left(\frac{x}{t}\right)=x.$$

This theorem connects the proximal operator of proper closed convex functions and their conjugates.

We conclude this section by introducing the definition of Bregman distance and its associated properties (see [6] and [3, Chapter 9]). These concepts will serve as valuable tools for our algorithmic design and convergence analysis.

Definition 2.8 (Bregman distance)**.** Let $\psi:\mathbb{R}^n\to(-\infty,+\infty]$ be a strongly convex and continuously differentiable function over $\mathbf{dom}\,(\partial\psi)$. The Bregman distance associated with the kernel ψ is the function $\mathscr{B}_\psi:\mathbf{dom}\,(\psi)\times\mathbf{dom}\,(\partial\psi)\to\mathbb{R}$ given by

$$\mathscr{B}_\psi(x,y)=\psi(x)-\psi(y)-\langle\nabla\psi(y),x-y\rangle,\quad\forall x\in\mathbf{dom}\,(\psi),y\in\mathbf{int}\,(\mathbf{dom}\,(\psi)).$$

Lemma 2.9. *Suppose that $\mathbb{S} \subseteq \mathbb{R}^n$ is nonempty, closed and convex, and the function ψ is proper closed convex and differentiable over* $\mathbf{dom}(\partial\psi)$. *If $\mathbb{S} \subseteq \mathbf{dom}(\psi)$ and $\psi + \mathscr{I}_{\mathbb{S}}$ with $\mathscr{I}_{\mathbb{S}}$ being an indicator function associated with $\mathbb{S}$ is ρ-strongly convex ($\rho > 0$), then the Bregman distance $\mathscr{B}_\psi$ associated with ψ has the following properties:*

(i) $\mathscr{B}_\psi(x,y) \geq \frac{\rho}{2}\|x-y\|^2$ *for all $x \in \mathbb{S}$ and $y \in \mathbb{S} \cap \mathbf{dom}(\partial\psi)$;*

(ii) *Let $x \in \mathbb{S}$ and $y \in \mathbb{S} \cap \mathbf{dom}(\partial\psi)$. Then $\mathscr{B}_\psi(x,y) \geq 0$, and in particular, the equality holds if and only if $x = y$;*

(iii) *If $\nabla\psi$ is Lipschitz continuous with modulus L_ψ, it holds that $\mathscr{B}_\psi(x,y) \leq \frac{L_\psi}{2}\|x-y\|^2$ for all $x \in \mathbb{S}$ and $y \in \mathbb{S} \cap \mathbf{dom}(\partial\psi)$.*

3. Algorithm and convergence analysis

In this section, we aim to introduce a new algorithm for (1.1) and establish the convergence result for the proposed algorithm.

3.1. Algorithm: iBPDCA.

The inertial technique has been widely used for algorithmic acceleration. Therefore, we first follow the inertial idea to generate an intermediate point via

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

where $\alpha_k \in (0,1]$ is an extrapolation parameter. Then, we employ the Fenchel-Young inequality (2.2) to iteratively construct a majorized surrogate of $-g(x)$ at $\widehat{x}^k$:

$$g^*(\xi) - \langle \xi, \widehat{x}^k \rangle \geq -g(\widehat{x}^k), \quad \forall \xi \in \mathbb{R}^n.$$

In this situation, we can construct a ξ-subproblem to update ξ^{k+1} uniquely by attaching a proximal term $\frac{\beta}{2}\|\xi - \xi^k\|^2$, i.e.,

$$\xi^{k+1} = \arg\min_{\xi \in \mathbb{R}^n} \left\{ g^*(\xi) - \langle \widehat{x}^k, \xi \rangle + \frac{\beta}{2}\|\xi - \xi^k\|^2 \right\}, \tag{3.2}$$

where $\beta > 0$ is a regularization parameter. Finally, we invoke the decomposable form of $h(x) = h^+(x) - h^-(x)$, and reformulate (1.1) as

$$\min_{x \in \mathbb{R}^n} \Phi(x) = \big(f(x) + h^+(x)\big) - \big(g(x) + h^-(x)\big).$$

Consequently, we follow the DCA spirit and attach a Bregman proximal term to the x-subproblem. Specifically, we update x^{k+1} via

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + h^+(x) - \langle x - \widehat{x}^k, u^k \rangle + \tau \mathscr{B}_\psi(x, \widehat{x}^k) \right\}, \tag{3.3}$$

where $u^k = \xi^{k+1} + \nabla h^-(\widehat{x}^k)$ and $\tau > 0$ is also a proximal regularization parameter. Concretely, we summarize the details for solving (1.1) in Algorithm 1.

Remark 3.1. In the literature, most existing DC-type algorithms usually randomly select a subgradient of $g(x^k)$ when g is not differentiable. Comparatively, we can obtain a unique ξ^{k+1} via (3.2) as long as $g(x)$ is a convex function. Actually, updating ξ^{k+1} amounts to evaluating

Algorithm 1 Inertial Bregman Proximal DC Algorithm (iBPDCA) for solving (1.1).

Choose a starting point x^0, set $x^{-1}=x^0, \{\alpha_k\}\subseteq[\alpha_{\min},1], 0<\alpha_{\min}<1, \beta>1/2$ and $\tau>0$.
for $k=0,1,2,\cdots,$ **do**
  Calculate $\widehat{x}^k$ via (3.1).
  Compute ξ^{k+1}via (3.2).
  Update x^{k+1} via (3.3).
end for

$\mathbf{Prox}_{\frac{1}{\beta}g^*}\left(\xi^k+\frac{1}{\beta}\widehat{x}^k\right)$, which, with the help of the extended Moreau Decomposition Theorem 2.7, can be easily obtained via

$$\mathbf{Prox}_{\frac{1}{\beta}g^*}\left(\xi^k+\frac{1}{\beta}\widehat{x}^k\right)=\xi^k+\frac{1}{\beta}\widehat{x}^k-\frac{1}{\beta}\mathbf{Prox}_{\beta g}\left(\beta\xi^k+\widehat{x}^k\right).$$

Therefore, we do not care about the explicit form of $g^*(x)$, and our algorithm enjoys a simpler iterative scheme.

3.2. **Convergence analysis.** In this subsection, we are concerned with the convergence properties of Algorithm 1. To begin with, we state some standard assumptions on (1.1), which will be used in convergence analysis.

Assumption 1. *$f(x)$ is a proper lower semicontinuous function and $g(x)$ is a continuous and convex function.*

Assumption 2. $\inf_x\Phi(x):=\inf_x\{f(x)-g(x)+h(x)\}>-\infty$.

Assumption 3. *$h(x)$ is decomposable of the form $h(x):=h^+(x)-h^-(x)$ with gradient Lipschitz continuity modulus L_h^+ and L_h^- for $h^+(x)$ and $h^-(x)$, respectively.*

In the coming analysis, for notational simplicity, we denote

$$\Theta(x,\xi)=f(x)+g^*(\xi)-\langle\xi,x\rangle+h^+(x)-h^-(x)$$

as a surrogate objective function to approximate $\Phi(x)$. From the Fenchel-Young inequality, it is easy to get $\Theta(x,\xi)\geq\Phi(x)$.

Lemma 3.2. *Suppose that Assumptions 1, 2 and 3 hold. Then, there exist two scalars H and P such that*

$$\Theta(x^{k+1},\xi^{k+1})\leq\Theta(x^k,\xi^k)+H\|\widehat{x}^k-x^k\|^2+P\|\widehat{x}^{k+1}-x^{k+1}\|^2+\left(\frac{1}{2}-\beta\right)\|\xi^{k+1}-\xi^k\|^2, \quad (3.4)$$

where

$$H=\frac{\tau L_\psi+1+2(L_h^-)^2-2\tau\rho}{2} \quad \text{and} \quad P=\frac{2(L_h^-)^2+L_h^--2\tau\rho+1}{2\alpha_{\min}^2}. \quad (3.5)$$

Proof. From the first-order optimality condition of (3.2), we easily obtain

$$\widehat{x}^k-\beta(\xi^{k+1}-\xi^k)\in\partial g^*(\xi^{k+1}).$$

Combining the convexity of g^*, we get

$$g^*(\xi^{k+1})\leq g^*(\xi^k)+\langle\xi^{k+1}-\xi^k,\widehat{x}^k-\beta(\xi^{k+1}-\xi^k)\rangle.$$

According to the update scheme (3.3), it yields

$$\begin{aligned} &f(x^{k+1})+h^+(x^{k+1})-\langle x^{k+1}-\widehat{x}^k,u^k\rangle+\tau\mathscr{B}_\psi(x^{k+1},\widehat{x}^k)\\ &\leq f(x^k)+h^+(x^k)-\langle x^k-\widehat{x}^k,u^k\rangle+\tau\mathscr{B}_\psi(x^k,\widehat{x}^k). \end{aligned}$$

Adding the above two inequalities, and then reorganizing the resulting inequality, we have

$$\begin{aligned} &f(x^{k+1})+g^*(\xi^{k+1})+h^+(x^{k+1})-h^+(x^k)+\langle\widehat{x}^k,\xi^k-\xi^{k+1}\rangle+\langle x^k-x^{k+1},\xi^{k+1}+\nabla h^-(\widehat{x}^k)\rangle\\ &\leq f(x^k)+g^*(\xi^k)-\beta\|\xi^k-\xi^{k+1}\|^2+\tau\mathscr{B}_\psi(x^k,\widehat{x}^k)-\tau\mathscr{B}_\psi(x^{k+1},\widehat{x}^k). \end{aligned}$$

Subtracting $\langle x^k,\xi^k\rangle$ both sides simultaneously and arranging each term, we arrive at

$$\begin{aligned} &f(x^{k+1})+g^*(\xi^{k+1})-\langle x^{k+1},\xi^{k+1}\rangle+h^+(x^{k+1})-h^+(x^k)\\ &\leq f(x^k)+g^*(\xi^k)-\langle x^k,\xi^k\rangle-\langle x^k-x^{k+1},\nabla h^-(\widehat{x}^k)\rangle-\beta\|\xi^k-\xi^{k+1}\|^2\\ &\quad+\tau\mathscr{B}_\psi(x^k,\widehat{x}^k)-\tau\mathscr{B}_\psi(x^{k+1},\widehat{x}^k)+\langle x^k-\widehat{x}^k,\xi^k-\xi^{k+1}\rangle. \end{aligned}\tag{3.6}$$

Recalling the strong convexity of ψ with strongly convex modulus ρ, and the Lipschitz gradient continuity property with constant L_ψ, we have

$$\tau\mathscr{B}_\psi(x^k,\widehat{x}^k)-\tau\mathscr{B}_\psi(x^{k+1},\widehat{x}^k)\leq\frac{\tau L_\psi}{2}\|x^k-\widehat{x}^k\|^2-\frac{\tau\rho}{2}\|x^{k+1}-\widehat{x}^k\|^2.$$

From Cauchy-Schwarz inequality, there holds

$$\langle x^k-\widehat{x}^k,\xi^k-\xi^{k+1}\rangle\leq\frac{1}{2}\|\xi^{k+1}-\xi^k\|^2+\frac{1}{2}\|\widehat{x}^k-x^k\|^2.$$

Subsequently, we get

$$\begin{aligned} &f(x^{k+1})+g^*(\xi^{k+1})-\langle x^{k+1},\xi^{k+1}\rangle+h^+(x^{k+1})-h^-(x^{k+1})\\ &-(h^+(x^k)-h^-(x^k))\\ &\leq f(x^k)+g^*(\xi^k)-\langle x^k,\xi^k\rangle-\langle x^k-x^{k+1},\nabla h^-(\widehat{x}^k)\rangle\\ &\quad-h^-(x^{k+1})+h^-(x^k)-\beta\|\xi^k-\xi^{k+1}\|^2\\ &\quad+\frac{\tau L_\psi}{2}\|x^k-\widehat{x}^k\|^2-\frac{\tau\rho}{2}\|x^{k+1}-\widehat{x}^k\|^2+\frac{1}{2}\|\widehat{x}^k-x^k\|^2+\frac{1}{2}\|\xi^{k+1}-\xi^k\|^2. \end{aligned}\tag{3.7}$$

Based on the Assumption 1, it follows from Lemma 2.4 that

$$-h^-(x^{k+1})+h^-(x^k)\leq\langle-\nabla h^-(x^{k+1}),x^{k+1}-x^k\rangle+\frac{L_h^-}{2}\|x^{k+1}-x^k\|^2,$$

which further implies that

$$\begin{aligned} &-\langle x^k-x^{k+1},\nabla h^-(\widehat{x}^k)\rangle-h^-(x^{k+1})+h^-(x^k)\\ &\leq\langle x^k-x^{k+1},\nabla h^-(x^{k+1})-\nabla h^-(\widehat{x}^k)\rangle+\frac{L_h^-}{2}\|x^{k+1}-x^k\|^2\\ &\leq\frac{(L_h^-)^2}{2}\|\widehat{x}^k-x^{k+1}\|^2+\frac{L_h^-+1}{2}\|x^{k+1}-x^k\|^2, \end{aligned}\tag{3.8}$$

where the last inequality holds from the Cauchy-Schwarz inequality. Based on the fact $\|x^{k+1}-x^k\|^2=\frac{1}{\alpha_{k+1}^2}\|\widehat{x}^{k+1}-x^{k+1}\|^2$, by substituting (3.8) into (3.7), we have

$$\begin{aligned}
&f(x^{k+1})+g^*(\xi^{k+1})-\langle x^{k+1},\xi^{k+1}\rangle+h^+(x^{k+1})-h^-(x^{k+1})\\
&\le f(x^k)+g^*(\xi^k)-\langle x^k,\xi^k\rangle+h^+(x^k)-h^-(x^k)+\left(\frac{1}{2}-\beta\right)\|\xi^k-\xi^{k+1}\|^2\\
&\quad+\frac{\tau L_\psi+1}{2}\|x^k-\widehat{x}^k\|^2+\frac{(L_h^-)^2-\tau\rho}{2}\|\widehat{x}^k-x^{k+1}\|^2+\frac{L_h^-+1}{2\alpha_{k+1}^2}\|\widehat{x}^{k+1}-x^{k+1}\|^2.
\end{aligned}$$

Since

$$\|x^{k+1}-\widehat{x}^k\|^2=\left\|\frac{1}{\alpha_k}(\widehat{x}^{k+1}-x^{k+1})+(x^k-\widehat{x}^k)\right\|^2\le\frac{2}{\alpha_{k+1}^2}\|\widehat{x}^{k+1}-x^{k+1}\|^2+2\|x^k-\widehat{x}^k\|^2,$$

and for any $k\in\mathbb{N}$, $0<\alpha_{\min}\le\alpha_k$, it yields

$$\begin{aligned}
&f(x^{k+1})+g^*(\xi^{k+1})-\langle x^{k+1},\xi^{k+1}\rangle+h^+(x^{k+1})-h^-(x^{k+1})\\
&\le f(x^k)+g^*(\xi^k)-\langle x^k,\xi^k\rangle+h^+(x^k)-h^-(x^k)+(\frac{1}{2}-\beta)\|\xi^k-\xi^{k+1}\|^2\\
&\quad+\frac{\tau L_\psi+1+2(L_h^-)^2-2\tau\rho}{2}\|x^k-\widehat{x}^k\|^2+\frac{2(L_h^-)^2+L_h^--2\tau\rho+1}{2\alpha_{\min}^2}\|\widehat{x}^{k+1}-x^{k+1}\|^2,
\end{aligned}$$

which, together with the notations H and P in (3.5) implies the assertion of this lemma. □

Now, we construct a merit function $\widehat{\Theta}:\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^n\to\bar{\mathbb{R}}$:

$$\widehat{\Theta}(x,\xi,\widehat{x},\zeta)=\Theta(x,\xi)+\delta\|x-\widehat{x}\|^2+\eta\|\xi-\zeta\|^2,\tag{3.9}$$

For all $k\ge1$, letting $\left\{(x^k,\xi^k,\widehat{x}^k,\xi^{k-1})\right\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1, and setting $x=x^k,\xi=\xi^k,\widehat{x}=\widehat{x}^k$ and $\zeta=\xi^{k-1}$ in function (3.9), we immediately define a new sequence as follows:

$$\left\{\widehat{\Theta}(x^k,\xi^k,\widehat{x}^k,\xi^{k-1})\right\}_{k\in\mathbb{N}}=\left\{\Theta(x^k,\xi^k)+\delta\|x^k-\widehat{x}^k\|^2+\eta\|\xi^k-\xi^{k-1}\|^2\right\}_{k\in\mathbb{N}},\tag{3.10}$$

where

$$\delta=\frac{\tau L_\psi-L_h^-}{2[(1-\varepsilon)+(1+\varepsilon)\alpha_{\min}^2]},\quad\tau=\frac{2(L_h^-)^2+L_h^-+1+2(1+\varepsilon)\delta}{2\rho},\quad\eta=\frac{2\beta-1}{2(1+\varepsilon)}.$$

Then, let $\beta>\frac{1}{2}$, $L_h^-<\tau L_\psi$, and $\varepsilon>0$ be an arbitrary real number satisfying $1-\varepsilon>0$. We below show that $\left\{\widehat{\Theta}(x^k,\xi^k,\widehat{x}^k,\xi^{k-1})\right\}_{k\in\mathbb{N}}$ defined by (3.10) satisfies following decreasing property. In what follows, we denote $\omega:=(x,\xi)$ and $\widehat{\omega}=:(x,\xi,\widehat{x},\zeta)$ for simplicity.

Lemma 3.3. *Suppose Assumptions 1, 2 and 3 hold, and the parameters are set as in (3.10). Let $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}:=\left\{(x^k,\xi^k,\widehat{x}^k,\xi^{k-1})\right\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1. Assume that $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$ is bounded. Then, the following statements hold.*

(i) *$\{\widehat{\Theta}(\widehat{\omega}^k)\}_{k\in\mathbb{N}}$ is monotonically nonincreasing and there exists a $\sigma=\varepsilon\min\{\delta,\eta\}$, such that*

$$\sigma\left(\|x^k-\widehat{x}^k\|^2+\|x^{k+1}-\widehat{x}^{k+1}\|^2+\|\xi^k-\xi^{k+1}\|^2\right)\le\widehat{\Theta}(\widehat{\omega}^k)-\widehat{\Theta}(\widehat{\omega}^{k+1}).\tag{3.11}$$

(ii)

$$\sum_{k=0}^{\infty}\|x^{k+1}-\widehat{x}^{k+1}\|^2<\infty \quad \textit{and} \quad \sum_{k=0}^{\infty}\|\xi^{k}-\xi^{k+1}\|^2<\infty,$$

In particular, $\lim_{k\to\infty}\|\xi^{k+1}-\xi^{k}\|=0$, *and* $\lim_{k\to\infty}\|x^{k+1}-\widehat{x}^{k+1}\|=0$, *which implies* $\lim_{k\to\infty}\|x^{k}-x^{k+1}\|=0$.

Proof. (i) The conclusion can be obtained immediately from Lemma 3.2 and the definition of (3.10) with the specific parameter settings.

(ii) From Assumption 2 and the definition of Θ, we know that the function Θ is bounded below. Hence the function $\widehat{\Theta}$ is also bounded below. Consequently, it follows from (i) that $\{\widehat{\Theta}(\widehat{\omega}^k)\}$ is monotonically nonincreasing. Therefore, $\{\widehat{\Theta}(\widehat{\omega}^k)\}$ is convergent. Let N be a positive integer. Summing up (3.11) from $k=0$ to $N-1$ leads to

$$\sigma\sum_{k=0}^{N-1}\left(\|x^{k}-\widehat{x}^{k}\|^2+\|x^{k+1}-\widehat{x}^{k+1}\|^2+\|\xi^{k}-\xi^{k+1}\|^2\right)\le\widehat{\Theta}(\widehat{\omega}^0)-\widehat{\Theta}(\widehat{\omega}^N).$$

From the definition of σ and taking the limit as $N\to\infty$, immediately yields

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

Hence $\sum_{k=0}^{\infty}\|x^{k+1}-\widehat{x}^{k+1}\|^2<\infty$. We can further obtain $\lim_{k\to\infty}\|x^{k+1}-\widehat{x}^{k+1}\|^2=0$. Based on the fact $\|x^{k+1}-x^{k}\|^2=\frac{1}{\alpha_{k+1}^2}\|\widehat{x}^{k+1}-x^{k+1}\|^2$, it immediately conclude that $\lim_{k\to\infty}\|x^{k+1}-x^{k}\|^2=0$. Likewise, we have $\sum_{k=0}^{\infty}\|\xi^{k}-\xi^{k+1}\|^2<\infty$, which means $\lim_{k\to\infty}\|\xi^{k}-\xi^{k+1}\|^2=0$.

□

Next, we will show that the subgradient of auxiliary function $\widehat{\Theta}$ is controlled by iteration gap.

Lemma 3.4. *Suppose Assumptions 1, 2 and 3 hold, and the parameters are set as in* (3.10). *Let* $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$ *be a sequence generated by Algorithm 1. Assume that* $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$ *is bounded. We define* $\upsilon^{k+1}\in\partial\widehat{\Theta}(\widehat{\omega}^{k+1})$, *where*

$$\upsilon^{k+1}=(\upsilon_x^{k+1},\upsilon_\xi^{k+1},\upsilon_{\widehat{x}}^{k+1},\upsilon_\zeta^{k+1}).$$

It holds

$$\begin{cases}\upsilon_x^{k+1}=\nabla h^-(\widehat{x}^k)-\nabla h^-(x^{k+1})+2\delta(x^{k+1}-\widehat{x}^{k+1})-\tau(\nabla\psi(x^{k+1})-\nabla\psi(\widehat{x}^k)),\\ \upsilon_\xi^{k+1}=\widehat{x}^k-x^k+\dfrac{1}{\alpha_{k+1}}(x^{k+1}-\widehat{x}^{k+1})+(2\eta-\beta)(\xi^{k+1}-\xi^k),\\ \upsilon_{\widehat{x}}^{k+1}=-2\delta(x^{k+1}-\widehat{x}^{k+1}),\\ \upsilon_\zeta^{k+1}=2\eta(\xi^k-\xi^{k+1}).\end{cases}\tag{3.12}$$

Then, there exists $\theta>0$ *such that*

$$\mathbf{dist}(0,\partial\widehat{\Theta}(\widehat{\omega}^{k+1}))\le\ \theta\left(\|x^{k}-\widehat{x}^{k}\|+\|x^{k+1}-\widehat{x}^{k+1}\|+\|\xi^{k+1}-\xi^{k}\|\right).\tag{3.13}$$

Proof. It is clear that

$$\begin{cases} \partial_x \widehat{\Theta}(\widehat{\omega}^{k+1}) = \partial f(x^{k+1}) - \xi^{k+1} + \nabla h^+(x^{k+1}) - \nabla h^-(x^{k+1}) + 2\delta(x^{k+1} - \widehat{x}^{k+1}), \\ \partial_\xi \widehat{\Theta}(\widehat{\omega}^{k+1}) = \partial g^*(\xi^{k+1}) - x^{k+1} + 2\eta(\xi^{k+1} - \xi^k), \\ \partial_{\widehat{x}} \widehat{\Theta}(\widehat{\omega}^{k+1}) = -2\delta(x^{k+1} - \widehat{x}^{k+1}), \\ \partial_\zeta \widehat{\Theta}(\widehat{\omega}^{k+1}) = 2\eta(\xi^k - \xi^{k+1}). \end{cases}$$

From the first order optimality conditions of (3.2)-(3.3), we have

$$\widehat{x}^k - \beta(\xi^{k+1} - \xi^k) \in \partial g^*(\xi^{k+1}),$$

and

$$\nabla h^-(\widehat{x}^k) - \nabla h^+(x^{k+1}) - \tau(\nabla \psi(x^{k+1}) - \nabla \psi(\widehat{x}^k)) + \xi^{k+1} \in \partial f(x^{k+1}).$$

Then, we can immediately get the formula (3.12). On the other hand, from the Lipschitz gradient continuity property of h^- and ψ, we further obtain

$$\|v_x^{k+1}\| \le \left(L_h^- + \tau L_\psi\right) \|x^k - \widehat{x}^k\| + \left(\frac{L_h^-}{\alpha_{k+1}} + 2\delta + \frac{\tau L_\psi}{\alpha_{k+1}}\right) \|x^{k+1} - \widehat{x}^{k+1}\|$$

and it is evident that

$$\begin{aligned} &\|v_\xi^{k+1}\| \le \|\widehat{x}^k - x^k\| + \frac{1}{\alpha_k}\|x^{k+1} - \widehat{x}^{k+1}\| + (2\eta - \beta)\|\xi^{k+1} - \xi^k\|, \\ &\|v_{\widehat{x}}^{k+1}\| \le 2\delta\|x^{k+1} - \widehat{x}^{k+1}\| \quad \text{and} \quad \|v_\zeta^{k+1}\| \le 2\eta\|\xi^{k+1} - \xi^k\|. \end{aligned}$$

Thus, there exists $\theta > 0$ such that

$$\begin{aligned} \mathbf{dist}(0, \partial\widehat{\Theta}(\widehat{\omega}^{k+1})) \le \|v^{k+1}\| &= \sqrt{\|v_x^{k+1}\|^2 + \|v_\xi^{k+1}\|^2 + \|v_{\widehat{x}}^{k+1}\|^2 + \|v_\zeta^{k+1}\|^2} \\ &\le \|v_x^{k+1}\| + \|v_\xi^{k+1}\| + \|v_{\widehat{x}}^{k+1}\| + \|v_\zeta^{k+1}\| \\ &\le \theta\left(\|x^k - \widehat{x}^k\| + \|x^{k+1} - \widehat{x}^{k+1}\| + \|\xi^{k+1} - \xi^k\|\right), \end{aligned}$$

which completes the proof. □

For notational simplicity, we introduce some new symbols. For sequence $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$, we denote $\widehat{\omega}^\star := (x^\star, \xi^\star, x^\star, \xi^\star)$ as its cluster point, and the set composed of these cluster points is defined as $\widehat{\Omega}^\star$. Similarly, for $\{\omega^k\}_{k\in\mathbb{N}}$, let $\omega^\star =: (x^\star, \xi^\star)$ be its cluster point, and we use $\Omega^\star$ to represent the cluster point set of $\{\omega^k\}_{k\in\mathbb{N}}$.

Lemma 3.5. *Suppose Assumptions 1, 2 and 3 hold, and the parameters are set as in (3.10). Let $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1. Assume that $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$ is bounded. Then the following statements hold.*

(i) *$\Omega^\star$ is a nonempty compact set, and $\lim_{k\to\infty} \mathbf{dist}(\widehat{\omega}^k, \widehat{\Omega}^\star) = \lim_{k\to\infty} \mathbf{dist}(\omega^k, \Omega^\star) = 0$.*
(ii) *$\widehat{\Theta}(\cdot)$ is a constant on $\widehat{\Omega}^\star$.*
(iii) *$\widehat{\Omega}^\star \subseteq \Xi(\widehat{\Theta})$.*

Proof. We prove the assertions of this lemma one by one.

(i) From the definition of $\widehat{\Omega}^\star$ and $\Omega^\star$, it is trivial to get Item (i).

(ii) Let $\widehat{\omega}^\star \in \widehat{\Omega}^\star$ then there exists a subsequence $\{\widehat{\omega}^{k_q}\}_{q\in\mathbb{N}}$ of $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$ converging to $\widehat{\omega}^\star$. From the second conclusion of Lemma 3.3, we have already known

$$\lim_{q\to\infty} x^{k_q+1} = \lim_{q\to\infty} \widehat{x}^{k_q+1} = x^\star \ , \ \lim_{q\to\infty} \xi^{k_q+1} = \xi^\star.$$

Since f is lower semicontinuous, $\widehat{\Theta}$ is also lower semicontinuous, which implies

$$\widehat{\Theta}(\widehat{\omega}^\star) \le \lim_{q\to\infty} \inf \widehat{\Theta}(\widehat{\omega}^{k_q+1}). \tag{3.14}$$

From (3.2) and (3.3), letting $k = k_q$, it yields

$$g^*(\xi^{k_q+1}) - \langle \widehat{x}^{k_q}, \xi^{k_q+1}\rangle + \frac{\beta}{2}\|\xi^{k_q+1} - \xi^{k_q}\|^2 \le g^*(\xi^\star) - \langle \widehat{x}^{k_q}, \xi^\star\rangle + \frac{\beta}{2}\|\xi^\star - \xi^{k_q}\|^2$$

and

$$\begin{aligned}
&f(x^{k_q+1}) + h^+(x^{k_q+1}) - \langle x^{k_q+1} - \widehat{x}^{k_q}, \xi^{k_q+1} + \nabla h^-(\widehat{x}^{k_q})\rangle + \tau\mathscr{B}_\psi(x^{k_q+1}, \widehat{x}^{k_q}) \\
&\le f(x^\star) + h^+(x^\star) - \langle x^\star - \widehat{x}^{k_q}, \xi^{k_q+1} + \nabla h^-(x^{k_q})\rangle + \tau\mathscr{B}_\psi(x^\star, \widehat{x}^{k_q}).
\end{aligned}$$

Combining the above two inequalities, adding $-h^-(x^{k_q+1})$ to the resulting inequality and reorganize it, we have

$$\begin{aligned}
&f(x^{k_q+1}) + g^*(x^{k_q+1}) - \langle x^{k_q+1}, \xi^{k_q+1}\rangle + h^+(x^{k_q+1}) - h^-(x^{k_q+1}) \\
&\le f(x^\star) + g^*(\xi^\star) - \langle \widehat{x}^\star, \xi^{k_q+1}\rangle + h^+(x^\star) - h^-(x^{k_q+1}) - \langle \widehat{x}^{k_q}, \xi^\star - \xi^{k_q+1}\rangle \\
&\quad - \langle x^\star - \widehat{x}^{k_q+1}, \nabla h^-(x^\star)\rangle + \frac{\beta}{2}\|\xi^\star - \xi^{k_q}\|^2 - \frac{\beta}{2}\|\xi^{k_q+1} - \xi^{k_q}\|^2 \\
&\quad + \tau\mathscr{B}_\psi(x^\star, \widehat{x}^{k_q}) - \tau\mathscr{B}_\psi(x^{k_q+1}, \widehat{x}^{k_q}).
\end{aligned}$$

Taking limit in the above inequality, we arrive at

$$\begin{aligned}
&\lim_{q\to\infty} \sup \widehat{\Theta}(\widehat{\omega}^{k_q+1}) \\
&= \lim_{q\to\infty} \sup \left\{ \Theta(\omega^{k_q+1}) + \delta\|x^{k_q+1} - \widehat{x}^{k_q+1}\|^2 + \eta\|\xi^{k_q+1} - \xi^{k_q}\|^2 \right\} \\
&\le \lim_{q\to\infty} \sup \Big\{ f(x^\star) + g^*(\xi^\star) - \langle \widehat{x}^\star, \xi^{k_q+1}\rangle + h^+(x^\star) - h^-(x^{k_q+1}) - \langle \widehat{x}^{k_q}, \xi^\star - \xi^{k_q+1}\rangle \\
&\quad - \langle x^\star - \widehat{x}^{k_q+1}, \nabla h^-(x^\star)\rangle + \frac{\beta}{2}\|\xi^\star - \xi^{k_q}\|^2 - \frac{\beta}{2}\|\xi^{k_q+1} - \xi^{k_q}\|^2 \\
&\quad + \tau\mathscr{B}_\psi(x^\star, \widehat{x}^{k_q}) - \tau\mathscr{B}_\psi(x^{k_q+1}, \widehat{x}^{k_q}) + \delta\|x^{k_q+1} - \widehat{x}^{k_q+1}\|^2 + \eta\|\xi^{k_q+1} - \xi^{k_q}\|^2 \Big\} \\
&= \widehat{\Theta}(\widehat{\omega}^\star) = \Theta(\omega^\star),
\end{aligned}$$

where the second equality holds from the continuity of h and Item (ii) of Lemma 3.3. Obviously, this inequality means

$$\lim_{q\to\infty} \sup \widehat{\Theta}(\widehat{\omega}^{k_q+1}) \le \widehat{\Theta}(\widehat{\omega}^\star),$$

which, together with the fact (3.14), yields

$$\lim_{q\to\infty} \widehat{\Theta}(\widehat{\omega}^{k_q+1}) = \widehat{\Theta}(\widehat{\omega}^\star) = \Theta(\omega^\star). \tag{3.15}$$

Hence $\widehat{\Theta}(\widehat{\omega})$ is constant on $\widehat{\Omega}^\star$.

(iii) By invoking Lemmas 3.3 and 3.4, we have $v^{k+1} \in \partial\widehat{\Theta}(\widehat{\omega}^{k+1})$. Since $\{\widehat{\Theta}(\widehat{\omega}^k)\}$ converges uniformly on $\widehat{\Omega}^\star$, the order of subgradient and limit can be exchanged. Letting $k \to \infty$, it yields $v^{k+1} \to 0$, which immediately implies $0 \in \partial\widehat{\Theta}(\widehat{\omega}^\star)$ by the closedness of $\partial\widehat{\Theta}$. Hence $\widehat{\omega}^\star$ is a stationary point of $\widehat{\Theta}$.

The proof is complete. □

To end this subsection, we now show that $\{x^k\}_{k\in\mathbb{N}}$ converges to a stationary point of (1.1).

Theorem 3.6. *Suppose $\widehat{\Theta}(\widehat{\omega})$ is a KŁ function and Assumptions 1-3 hold. The parameters are set in (3.10). Let $\{\omega^k\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 1. Assume $\{\omega^k\}_{k\in\mathbb{N}}$ is bounded. Then, the following statements hold.*

(i) *The sequence $\{\omega^k\}_{k\in\mathbb{N}}$ has finite length, i.e.,*

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

(ii) *The sequence $\{x^k\}_{k\in\mathbb{N}}$ converges to a critical point $x^\star$ of $\Phi(x)$.*

Proof. The proof is divided into two parts.

(i) To prove Item (i), we further consider two cases based on (3.15).

(a) If there exists a positive integer k_1 such that, for any $k > k_1$, $\widehat{\Theta}(\widehat{\omega}^k) = \widehat{\Theta}(\widehat{\omega}^\star)$. Then from Item (i) of Lemma 3.3, it is clear that

$$\begin{aligned}\sigma\left(\|x^k - \hat{x}^k\|^2 + \|x^{k+1} - \hat{x}^{k+1}\|^2 + \|\xi^k - \xi^{k+1}\|^2\right)\\ \le \widehat{\Theta}(\widehat{\omega}^k) - \widehat{\Theta}(\widehat{\omega}^{k+1}) \le \widehat{\Theta}(\widehat{\omega}^\star) - \widehat{\Theta}(\widehat{\omega}^\star) = 0.\end{aligned}$$

We can immediately get $x^{k+1} = \hat{x}^{k+1}$ for all $k > k_1$, it follows $x^{k+1} = x^k$. Then,

$$\sum_{k=0}^{\infty} \|\omega^{k+1} - \omega^k\| \le \sum_{k=0}^{\infty} \left\{\|x^k - x^{k+1}\| + \|\xi^k - \xi^{k+1}\|\right\} < \infty \tag{3.16}$$

holds for all k. Hence the conclusion is proved.

(b) Now, we assume $\widehat{\Theta}(\widehat{\omega}^k) > \widehat{\Theta}(\widehat{\omega}^\star)$ for all $k > 0$. Since (3.15) implies that for any $\hat{\zeta} > 0$, there exists a positive integer k_2 such that $\widehat{\Theta}(\widehat{\omega}^{k+1}) < \widehat{\Theta}(\widehat{\omega}^\star) + \hat{\zeta}$ for all $k > k_2$. It follows from Item (i) of Lemma 3.5 that, for any $\varsigma > 0$, there exists a positive integer k_3 such that $\mathbf{dist}(\widehat{\omega}^{k+1}, \widehat{\Omega}^\star) < \varsigma$ for all $k > k_3$. Now we choose $d = \max\{k_2, k_3\}$, then for any $k > d$, we have

$$\mathbf{dist}(\widehat{\omega}^{k+1}, \widehat{\Omega}^\star) < \varsigma \quad \text{and} \quad \widehat{\Theta}(\widehat{\omega}^\star) < \widehat{\Theta}(\widehat{\omega}^{k+1}) < \widehat{\Theta}(\widehat{\omega}^\star) + \hat{\zeta}.$$

In addition, $\widehat{\Omega}^\star$ is a nonempty compact set and $\widehat{\Theta}$ is constant over such a set. Then, the condition of uniformized KŁ property (Lemma 2.3) is satisfied. Consequently, we deduce that for any $k > d$,

$$\phi'(\widehat{\Theta}(\widehat{\omega}^{k+1}) - \widehat{\Theta}(\widehat{\omega}^\star))\mathbf{dist}(0, \partial\widehat{\Theta}(\widehat{\omega}^{k+1})) \ge 1. \tag{3.17}$$

According to Lemma 3.4, we have

$$\mathbf{dist}(0, \partial\widehat{\Theta}(\widehat{\omega}^{k+1})) \le \theta\left(\|x^k - \hat{x}^k\| + \|x^{k+1} - \hat{x}^{k+1}\| + \|\xi^{k+1} - \xi^k\|\right). \tag{3.18}$$

It follows from the concavity of ϕ that

$$\begin{aligned}&\phi\big(\widehat{\Theta}(\widehat{\omega}^k)-\widehat{\Theta}(\widehat{\omega}^\star)\big)-\phi\big(\widehat{\Theta}(\widehat{\omega}^{k+1})-\widehat{\Theta}(\widehat{\omega}^\star)\big)\\&\geq\phi'\big(\widehat{\Theta}(\widehat{\omega}^k)-\widehat{\Theta}(\widehat{\omega}^\star)\big)\big(\widehat{\Theta}(\widehat{\omega}^k)-\widehat{\Theta}(\widehat{\omega}^{k+1})\big).\end{aligned}\tag{3.19}$$

We denote $\triangle_k=\phi\big(\widehat{\Theta}(\widehat{\omega}^k)-\widehat{\Theta}(\widehat{\omega}^\star)\big)-\phi\big(\widehat{\Theta}(\widehat{\omega}^{k+1})-\widehat{\Theta}(\widehat{\omega}^\star)\big)$ for simplicity. Associating (3.17), (3.18) with (3.19), it follows that

$$\widehat{\Theta}(\widehat{\omega}^k)-\widehat{\Theta}(\widehat{\omega}^{k+1})\leq\theta\triangle_k\big(\|x^k-\widehat{x}^k\|+\|x^{k+1}-\widehat{x}^{k+1}\|+\|\xi^{k+1}-\xi^k\|\big).$$

From (3.11), we can further obtain

$$\begin{aligned}&\Big(\|x^k-\widehat{x}^k\|^2+\|x^{k+1}-\widehat{x}^{k+1}\|^2+\|\xi^k-\xi^{k+1}\|^2\Big)\\&\leq\frac{\theta}{\sigma}\triangle_k\big(\|x^k-\widehat{x}^k\|+\|x^{k+1}-\widehat{x}^{k+1}\|+\|\xi^{k+1}-\xi^k\|\big).\end{aligned}\tag{3.20}$$

Using the fact $\frac{1}{\sqrt{n}}(a_1+\cdots+a_n)\leq(a_1^2+\cdots+a_n^2)^{\frac{1}{2}}$ and $\sqrt{ab}\leq\frac{1}{2}(\gamma a+\frac{b}{\gamma})$ for $a,b,\gamma>0$, we then have

$$\begin{aligned}&\left(\frac{\theta}{\sigma}\triangle_k\big(\|x^k-\widehat{x}^k\|+\|x^{k+1}-\widehat{x}^{k+1}\|+\|\xi^{k+1}-\xi^k\|\big)\right)^{\frac{1}{2}}\\&\leq\frac{1}{2}\left(\frac{\theta\gamma}{\sigma}\triangle_k+\frac{1}{\gamma}\big(\|x^k-\widehat{x}^k\|+\|x^{k+1}-\widehat{x}^{k+1}\|+\|\xi^{k+1}-\xi^k\|\big)\right),\end{aligned}\tag{3.21}$$

and

$$\begin{aligned}&\frac{1}{\sqrt{3}}\Big(\|x^k-\widehat{x}^k\|+\|x^{k+1}-\widehat{x}^{k+1}\|+\|\xi^k-\xi^{k+1}\|\Big)\\&\leq\Big(\|x^k-\widehat{x}^k\|^2+\|x^{k+1}-\widehat{x}^{k+1}\|^2+\|\xi^k-\xi^{k+1}\|^2\Big)^{\frac{1}{2}}.\end{aligned}\tag{3.22}$$

Now, by taking the square root of both sides of inequality (3.20), and substituting (3.21) and (3.22) into it, we obtain

$$\left(1-\frac{\sqrt{3}}{2\gamma}\right)\Big(\|x^k-\widehat{x}^k\|+\|\xi^k-\xi^{k+1}\|+\|x^{k+1}-\widehat{x}^{k+1}\|\Big)\leq\frac{\sqrt{3}}{2}\frac{\theta\gamma}{\sigma}\triangle_k.\tag{3.23}$$

Summing (3.23) from $k=N$ to ∞ immediately yields

$$\begin{aligned}&\left(1-\frac{\sqrt{3}}{2\gamma}\right)\sum_{k=N}^{\infty}\Big(\|x^k-\widehat{x}^k\|+\|\xi^k-\xi^{k+1}\|+\|x^{k+1}-\widehat{x}^{k+1}\|\Big)\\&\leq\frac{\sqrt{3}}{2}\frac{\theta\gamma}{\sigma}\phi\big(\widehat{\Theta}(\widehat{\omega}^N)-\widehat{\Theta}(\widehat{\omega}^\star)\big).\end{aligned}\tag{3.24}$$

Due to the arbitrariness of γ, we just require $\gamma>\frac{\sqrt{3}}{2}$, which immediately ensures $\left(1-\frac{\sqrt{3}}{2\gamma}\right)>0$. As a consequence, the boundedness of the right hand side of (3.24) leads to

$$\left(1-\frac{\sqrt{3}}{2\gamma}\right)\sum_{k=N}^{\infty}\big(\|x^k-\widehat{x}^k\|+\|\xi^k-\xi^{k+1}\|+\|x^{k+1}-\widehat{x}^{k+1}\|\big)<\infty,$$

which means $\sum_{k=N}^{\infty}\|\xi^k-\xi^{k+1}\|<\infty$ and $\sum_{k=N}^{\infty}\|x^{k+1}-\widehat{x}^{k+1}\|<\infty$. Invoking the updating scheme of $\widehat{x}^k$ in (3.1), we have

$$\widehat{x}^{k+1}-x^{k+1}=\alpha_{k+1}(x^{k+1}-x^k),\quad \alpha_{k+1}\in(0,1].$$

Therefore, we can easily obtain $\sum_{k=N}^{\infty}\|x^{k+1}-x^k\|<\infty$. From (3.16), we finally draw the conclusion.

(ii) Now, we aim to show $x^\star\in\mathbf{crit}(\Phi)$. Since sequence $\{\omega^k\}_{k\in\mathbb{N}}$ is convergent, which implies $\{\widehat{\omega}^k\}_{k\in\mathbb{N}}$ and $\{x^k\}_{k\in\mathbb{N}}$ are also convergent. In view of Item (iii) of Lemma 3.5, from $0\in\partial\widehat{\Theta}(\widehat{\omega}^\star)$, we have

$$\begin{cases}x^\star\in\partial g^*(\xi^\star),\\ \xi^\star\in\partial f(x^\star)+\nabla h^+(x^\star)-\nabla h^-(x^\star).\end{cases}\tag{3.25}$$

Since $x^\star\in\partial g^*(\xi^\star)$, by the properties of conjugate functions and the continuity of g, we have $\xi^\star\in\partial g(x^\star)$, which together with (3.25) implies that

$$\{\partial f(x^\star)+\nabla h(x^\star)\}\cap\partial g(x^\star)=\xi^\star\neq\emptyset,$$

that is,

$$0\in\partial f(x^\star)-\partial g(x^\star)+\nabla h^+(x^\star)-\nabla h^-(x^\star),$$

which means that $x^\star$ is a critical point of (1.1).

The proof is complete. □

4. Numerical experiments

In this section, we apply Algorithm 1 (iBPDCA for short) to low-rank matrix and tensor completion problems, and compare it with the classical DCA (i.e., iterative scheme (1.2) [21]) and the BPDCA (i.e., iBPDCA without the extrapolation (3.1)). All codes were written by Matlab 2021b and all experiments were conducted on a laptop computer with Intel (R) core (TM) i7-7500 CPU @ 2.70GHz and 8GB memory.

4.1. Low-rank matrix completion.

In this subsection, we consider the following low-rank matrix completion problem:

$$\min_{X\in\mathbb{R}^{m\times n}}\ \lambda(\|X\|_*-\|X\|_F)+\frac{1}{2}\|\mathscr{P}_\Omega(X-M)\|_F^2,\tag{4.1}$$

where M is an incompleted matrix, Ω is a subset of the index set of entries $\{1,\ldots,m\}\times\{1,\ldots,n\}$, $\mathscr{P}_\Omega(\cdot):\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}$ is the orthogonal projection so that the ij-th entry of $\mathscr{P}_\Omega(M)$ is M_{ij} if $(i,j)\in\Omega$ and zero otherwise.

We can easily reformulate model (4.1) as a special case of (1.1) by setting $x=X$,

$$f(X)=\lambda\|X\|_*,\ g(X)=\lambda\|X\|_F,\ h^+(X)=\frac{1}{2}\|\mathscr{P}_\Omega(X-M)\|_F^2,\ \text{and}\ h^-(X)=0.$$

Due to the appearance of $\mathscr{P}_\Omega$, we take the Bregman kernel function as $\psi_k(\cdot)=\frac{1}{2}\|\cdot\|_{M_k}^2$, where $M_k=\mu_k I-\mathscr{P}_\Omega^\top\mathscr{P}_\Omega$ with some appropriate $\mu_k>0$ such that M_k is positive definite. Throughout this subsection, we set $\mu_k=1.1,\lambda=0.5,\beta=1$ and $\tau=1$. At each iteration, we update $\alpha_k=\frac{t^{k-1}-1}{t^k}$, where $t^k=\frac{1}{2}(1+\sqrt{1+4(t^{k-1})^2})$ and $t^0=1$. When implementing the classical DCA

and BPDCA, the underlying $h(X)$ is directly set as $h(X) = \frac{1}{2}\|\mathscr{P}_\Omega(X-M)\|_F^2$. Then, applying the DCA (1.2) to (4.1) yields

$$X^{k+1} = \arg\min_X \left\{ \lambda\|X\|_* - \langle \xi^{k+1}, X - X^k\rangle + \frac{1}{2}\|\mathscr{P}_\Omega(X-M)\|_F^2 \right\} \tag{4.2}$$

where $\xi^{k+1} \in \partial(\lambda\|X^k\|_F)$. However, we notice that the subproblem (4.2) has no closed-form solution, which is not practically feasible. Therefore, we approximately obtain X^{k+1} via the powerful alternating direction method of multipliers (ADMM [8, 9]), where $h(X)$ is further transformed into $h(Y)$ via an auxiliary variable Y so that $\|X\|_*$ and $h(Y)$ are separable. In our experiments, we set the penalty parameter as 1.1^j in the $\{k,j\}$-th outer-inner iteration, and take $\|X^{k,j} - Y^j\|_F \le 10^{-3}$ as the stopping criterion for ADMM, where k and j count the outer and inner iterations, respectively. For simplicity, we denote this algorithm by ADMM-DCA for short. The iterative scheme of BPDCA for (4.1) reads as

$$\begin{aligned} X^{k+1} &= \arg\min_X \left\{ \lambda\|X\|_* + \langle \nabla h(X^k) - \xi^{k+1}, X - X^k\rangle + \frac{\mu_k}{2}\|X - X^k\|_F^2 \right\} \\ &= \mathscr{D}_{\frac{\lambda}{\mu_k}}\left(X^k - \frac{1}{\mu_k}\left(\nabla h(X^k) - \xi^{k+1}\right)\right), \end{aligned}$$

where ξ^{k+1} is updated by scheme (3.2) with $\widehat{X}^k = X^k$ and $\mathscr{D}_\kappa(\cdot)$ is the well-known singular value shrinkage operator, which is defined by $\mathscr{D}_\kappa(A) = U\mathscr{D}_\kappa(\Sigma)V^\top$ with $A = U\Sigma V^\top$ being the SVD of a matrix $A \in \mathbb{R}^{m\times n}$ of rank r in the reduced form, i.e., $\Sigma = \mathrm{diag}(\{\sigma_i\}_{1\le i\le r})$, U and V are orthogonal matrices, and $\mathscr{D}_\kappa(\Sigma) = \mathrm{diag}(\max\{\sigma_i - \kappa, 0\})$ for $\kappa \ge 0$.

We first consider some synthetic datasets to evaluate the feasibility and reliability of iBPDCA. More concretely, we generate a low-rank matrix $X_{\rm true}$ as a ground truth matrix, which follows $X_{\rm true} = UV + 0.01\cdot \mathrm{randn}(m,n)$, where $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{r\times n}$ with the entries being random samples drawn from a uniform distribution, and r is much less than $\min\{m,n\}$. In this way, we construct a low-rank matrix whose rank is less than r. In our experiments, we take $r = 10$, and set $m = n = 100, 500$ and 1000, respectively. Moreover, we apply Matlab script $\Omega = \mathrm{rand}(m,n) <$ sample ratio (sr), and then set the observed matrix $M = \mathscr{P}_\Omega(X_{\rm true})$. Throughout this subsection, we set sr $= 0.2$ and 0.5. We define

$$\mathrm{RSE} = \frac{\|X^* - X_{\rm true}\|}{\|X_{\rm true}\|},$$

to measure the quality of the recovered matrix. Moreover, we report the computing time in seconds (Time(s)), number of iteration (Iter) and the rank of the recovered matrix (rank) in Table 1. We see that three algorithms can achieve the almost same RSE and rank, but our iBPDCA takes the least time in all cases.

Hereafter, we are concerned with the numerical performance of our algorithm on real-world datasets. In our experiments, we choose five widely used color images ('fruits', 'tulips', 'butterfly', 'lena' and 'barbara') with size of $256\times 256\times 3$. Here, we convert the original image $\mathscr{X}_{\rm true} \in \mathbb{R}^{256\times 256\times 3}$ into a matrix $X_{\rm true} \in \mathbb{R}^{256\times 768}$. To report the results, we further employ the Peak Signal-to-Noise Ratio (PSNR) defined by

$$\mathrm{PSNR} = 10\log_{10}\frac{(X_{\rm true}^{\max})^2(\#\Omega^C)}{\|X^* - X_{\rm true}\|_F^2},$$

TABLE 1. Computational results of solving problem (4.1) on synthetic data

sr	size	BPDCA				iBPDCA				ADMM-DCA			
		RSE	Time(s)	Iter	rank	RSE	Time(s)	Iter	rank	RSE	Time(s)	Iter	rank
0.5	(100,100)	1.63e-02	0.26	121	10	1.62e-02	0.21	87	10	1.73e-02	0.60	15	10
	(500,500)	2.40e-03	13.49	192	10	2.41e-03	5.44	76	10	2.45e-03	39.71	20	10
	(1000,1000)	1.19e-03	185.00	262	10	1.19e-03	59.49	84	10	1.21e-03	540.16	27	10
0.2	(100,100)	6.29e-02	0.93	387	10	6.28e-02	0.56	207	10	6.48e-02	1.87	50	10
	(500,500)	7.39e-03	33.38	432	10	7.37e-03	10.48	124	10	8.47e-03	100.53	48	10
	(1000,1000)	2.35e-02	316.92	500	10	3.21e-03	99.21	147	10	3.52e-03	1023.96	62	10

TABLE 2. Computational results of solving problem (4.1) on color images

sr	image	BPDCA				iBPDCA				ADMM-DCA			
		PSNR	Time(s)	Iter	rank	PSNR	Time(s)	Iter	rank	PSNR	Time(s)	Iter	rank
0.5	fruits	27.78	2.94	103	120	27.79	1.81	63	120	27.67	8.53	15	126
	tulips	25.93	3.63	127	143	25.93	2.11	74	143	25.81	10.80	19	152
	buttergly	24.90	2.63	137	119	24.95	1.58	80	119	24.87	8.99	20	126
	lena	28.72	2.41	85	113	28.73	1.85	65	113	28.60	7.26	13	120
	barbara	28.18	2.76	91	120	28.19	2.02	70	120	28.07	7.90	14	128
0.2	fruits	21.66	6.85	242	79	21.67	2.91	103	79	21.60	19.90	34	100
	tulips	21.75	8.37	238	94	21.75	2.50	86	94	21.69	20.64	33	118
	butterfly	19.12	8.39	295	76	19.12	4.20	115	76	19.09	28.85	43	96
	lena	22.70	6.14	197	73	22.70	3.01	102	73	22.62	17.82	30	92
	barbara	22.13	8.05	233	78	22.14	3.65	125	78	22.05	19.11	32	99

to measure the quality of the recovered image, where $X_{\text{true}}^{\max}$ is the maximum element of X_{true}, where $\#\Omega^C$ denotes the number of elements in the complementary set of Ω. All results are summarized in Table 2 and Figure 1. It can be seen from Figure 1 that the recovered images by three algorithms have almost the same quality. However, we can see from Table 2 that BPDCA and iBPDCA perform better than ADMM-DCA in terms of computing time and PSNR values.

4.2. **Low-tubal-rank tensor completion.** The objective variable in model (4.1) is a matrix. In this subsection, we consider a more general case where the variable is a tensor in the objective function. Before our experiments, we briefly recall some basic concepts of the tensor nuclear norm (see [11, 12] for more details). We represent the tensor $\bar{\mathscr{X}}$ in terms of $\mathscr{X}$ after performing the Fast Fourier Transform (FFT) on each tube, i.e., $\bar{\mathscr{X}} = \text{fft}(\mathscr{X}, [], 3)$ and $\mathscr{X} = \text{ifft}(\bar{\mathscr{X}}, [], 3)$ in MATLAB. We use X_k (resp. $\bar{X}_k$) to denote the k-th frontal slice of a third-order tensor $\mathscr{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ (resp. $\bar{\mathscr{X}} \in \mathbb{R}^{n_1\times n_2\times n_3}$). With the above preparations, we consider the following low-tubal-rank tensor completion problem:

$$\min_{\mathscr{X}\in\mathbb{R}^{n_1\times n_2\times n_3}} \lambda(\|\mathscr{X}\|_{\text{TNN}} - \|\mathscr{X}\|_F) + \frac{1}{2}\|\mathscr{P}_\Omega(\mathscr{X} - \mathscr{M})\|_F^2, \tag{4.3}$$

where $\|\mathscr{X}\|_{\text{TNN}} := \frac{1}{n_3}\sum_{k=1}^{n_3}\sum_{i=1}^{\min\{n_1,n_2\}} \sigma_i(\bar{X}_k)$, and $\sigma_i(\bar{X}_k)$ is the i-th singular value of $\bar{X}_k$. Clearly, model (4.3) can also be regarded as a special case of (1.1) by setting $x = \mathscr{X}$,

$$f(\mathscr{X}) = \lambda\|\mathscr{X}\|_*,\ g(\mathscr{X}) = \lambda\|\mathscr{X}\|_F,\ h^+(\mathscr{X}) = 0,\ \text{and}\ h^-(\mathscr{X}) = -\frac{1}{2}\|\mathscr{P}_\Omega(\mathscr{X} - \mathscr{M})\|_F^2.$$

$\mathscr{X}(:,:,k)$ (resp. $\bar{\mathscr{X}}(:,:,k)$) in Matlab

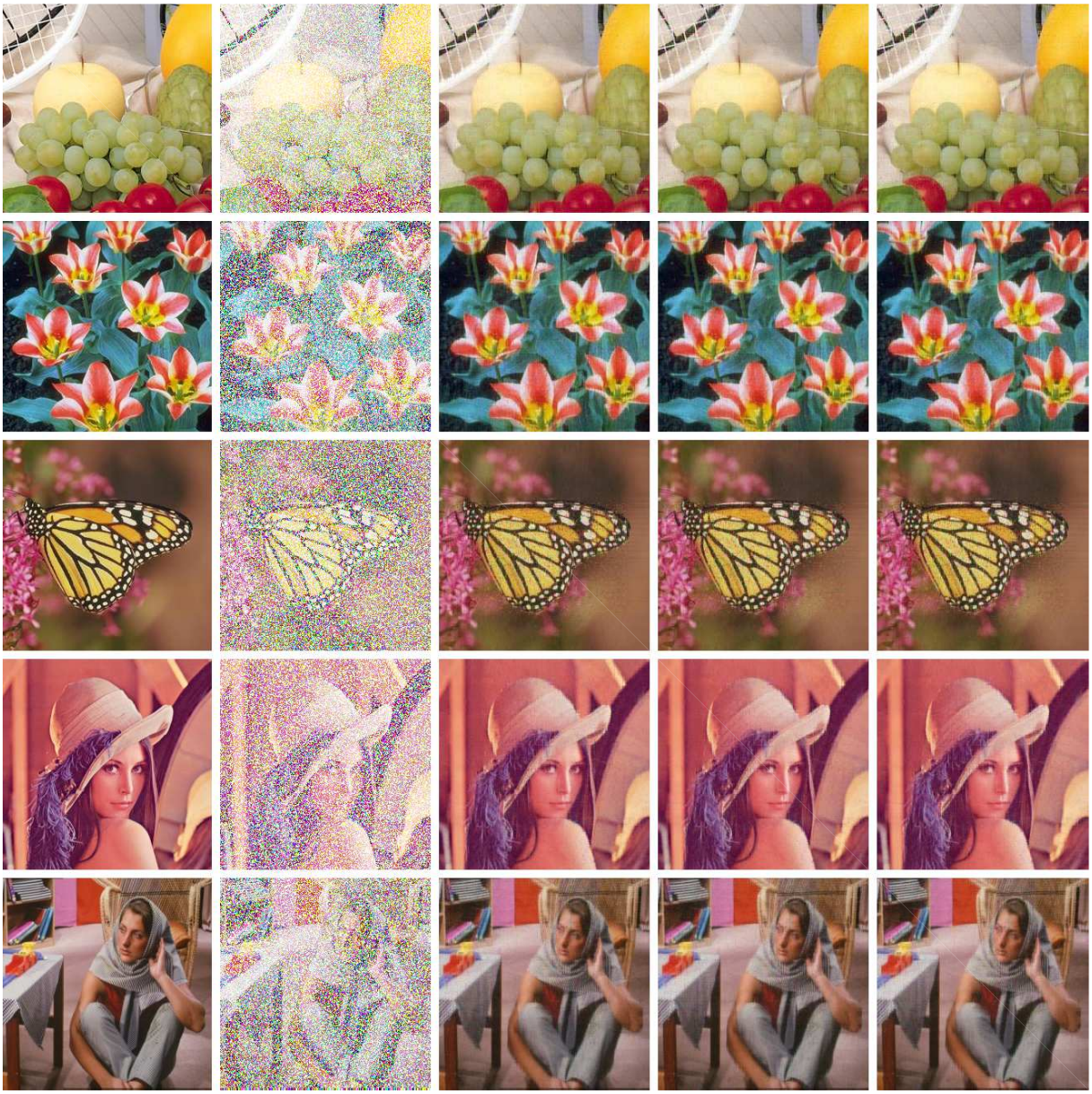

FIGURE 1. Images recovered by BPDCA, iBPDCA and ADMM-DCA. From top to bottom: fruits, tulips, butterfly, lena, and barbara. From left to right: original images, downsampled images with sr= 0.5, image recoverd by BPDCA, iBPDCA and ADMM-DCA, respectively.

The main difference between problem (4.3) and (4.1) is that the objective variable is changed from a matrix X to a tensor $\mathscr{X}$, so the corresponding tensor nuclear norm $\|\cdot\|_{\text{TNN}}$ is considered. All parameters for iBPDCA for solving (4.3) are same to the matrix case. Similar to model (4.1), we also employ the BPDCA and ADMM-DCA to solve problem (4.3). The iterative scheme of

TABLE 3. Computational results of solving problem (4.3) on synthetic data

sr	size	BPDCA				iBPDCA				A		DMM-DCA	
		RSE	Time(s)	Iter	Trank	RSE	Time(s)	Iter	Trank	RSE	Time(s)	Iter	Trank
0.5	(20,20,10)	1.66-02	0.89	654	10	1.64e-02	0.40	226	10	1.66e-02	1.51	82	11
	(50,50,10)	3.35-03	2.53	443	6	3.33e-03	0.76	146	5	3.58e-03	5.03	48	7
	(100,100,10)	1.56e-03	11.77	600	5	1.55e-03	3.07	161	5	1.61e-03	27.65	62	5
0.2	(20,20,10)	4.75e-02	1.27	867	6	4.79e-02	0.37	215	6	4.75e-02	2.32	116	8
	(50,50,10)	3.25e-02	5.58	1090	14	3.32e-02	1.43	281	14	3.31e-02	12.47	141	18
	(100,100,10)	7.39e-03	57.49	2945	22	5.36e-03	7.55	375	19	6.01e-03	134.52	354	26

BPDCA for solving (4.3) can be written as follow:

$$\begin{aligned}\mathscr{X}^{k+1} &= \arg\min_{\mathscr{X}}\left\{\lambda\|\mathscr{X}\|_{\mathrm{TNN}} + \langle\nabla h(\mathscr{X}^k) - \xi^{k+1}, \mathscr{X} - \mathscr{X}^k\rangle + \frac{\mu_k}{2}\|\mathscr{X} - \mathscr{X}^k\|_F^2\right\}\\ &= \mathscr{U} * \mathrm{ifft}\left(\widehat{\mathscr{D}}_{\frac{\lambda}{\mu_k}}(\Gamma), [], 3\right) * \mathscr{V}^{\hat{\top}},\end{aligned}$$

where ξ^{k+1} is updated by scheme (3.2) with $\widehat{\mathscr{X}^k} = \mathscr{X}^k, \mathscr{X}^k - \frac{1}{\mu_k}\left(\nabla h(\mathscr{X}^k) - \xi^{k+1}\right) = \mathscr{U} * \Gamma * \mathscr{V}^{\hat{\top}}$, which is the tensor singular value decomposition (see [12] for more details) with tubal rank r, where $\mathscr{U} \in \mathbb{R}^{n_1\times r\times n_3}$ and $\mathscr{V} \in \mathbb{R}^{r\times n_2\times n_3}$ are othogonal tensors and $\Gamma \in \mathbb{R}^{r\times r\times n_3}$ is an f-diagonal tensor, and $\widehat{\mathscr{D}}_\kappa(\Gamma) \in \mathbb{R}^{r\times r\times n_3}$ is defined by

$$\widehat{\mathscr{D}}_\kappa(\Gamma)_{iik} = \max\{\bar{\Gamma}_{iik} - \kappa, 0\},\ i = 1,\ldots,r;\ k = 1,\ldots,n_3,$$

where $\kappa \geq 0$. When applying the DCA (1.2) to solve (4.3), we have

$$\mathscr{X}^{k+1} = \arg\min_X\left\{\lambda\|\mathscr{X}\|_{\mathrm{TNN}} - \langle\xi^{k+1}, \mathscr{X} - \mathscr{X}^k\rangle + \frac{1}{2}\|\mathscr{P}_\Omega(\mathscr{X} - \mathscr{M})\|_F^2\right\}$$

where $\xi^{k+1} \in \partial(\lambda\|\mathscr{X}^k\|_F)$. Like the matrix case, we also introduce an auxiliary variable $\mathscr{Y} = \mathscr{X}$ to reformulate $h(\mathscr{X})$ as $h(\mathscr{Y})$. Then, we obtain $\mathscr{X}^{k+1}$ by the ADMM with same settings used in (4.2).

We start to apply our iBPDCA to solving problem (4.3) with synthetic datasets. Here, we randomly generate low-tubal-rank $\mathscr{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and Ω by the following way. First, we obtain $\mathscr{U} = \mathrm{randn}(n_1, r, n_3)$ and $\mathscr{V} = \mathrm{randn}(r, n_2, n_3)$ by Matlab. Then, we let $\mathscr{X} = \mathscr{U} * \mathscr{V} + 0.01 \cdot \mathrm{randn}(n_1, n_2, n_3)$. For Ω, we apply Matlab script $\mathrm{rand}(n_1, n_2, n_3) < \mathrm{sr}$. In this part, we set $r = 5$, $n_1 = n_2 = \{20, 50, 100\}$ and $n_3 = 10$. In Table 3, we report RSE, Time(s), Iter and the Tubal rank (Trank) of using three algorithms to solve model (4.3). It is easy to see that iBPDCA is faster than the other two algorithms in terms of computing time. In addition, iBPDCA can achieve lower tubal rank than ADMM-DCA, which, to some extent, means that iBPDCA can recover better solutions.

Finally, we are interested in recovering grayscale videos from incomplete observations. Here, we choose four grayscale videos including 'Airport', 'Lobby', 'Akiyo' and 'Hall'. Some preliminary numerical results are summarized in Table 4. We can see that iBPDCA always outperforms BPDCA and ADMM-DCA which further verifies the efficiency of our algorithm. In

$\mathscr{X}^{\hat{\top}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is the conjugate transpose for a tensor $\mathscr{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, obtained by conjugate transposing each of the frontal slices of $\mathscr{X}$ and then reversing the order of transposed frontal slices 2 through n_3.

http://trace.eas.asu.edu/yuv/.

TABLE 4. Computational results of solving problem (4.3) on video datasets

sr	image	BPDCA				iBPDCA				A		DMM-DCA	
		PSNR	Time(s)	Iter	Trank	PSNR	Time(s)	Iter	Trank	PSNR	Time(s)	Iter	Trank
0.5	airport	29.54	22.42	148	116	29.63	11.51	71	116	29.63	81.68	23	117
	lobby	40.06	12.18	98	107	40.10	7.12	59	107	40.04	36.19	13	109
	akiyo	41.13	14.32	96	101	41.24	8.92	60	101	41.11	46.59	13	102
	hall	40.31	13.42	92	122	40.44	8.45	57	122	40.36	47.26	13	122
0.2	airport	23.85	45.92	308	84	23.96	16.75	104	84	23.96	174.49	53	89
	lobby	35.42	29.92	254	92	35.60	10.79	65	92	35.32	90.68	33	93
	akiyo	34.61	32.87	220	76	34.84	20.44	122	76	34.67	107.37	30	78
	hall	34.38	30.75	202	99	34.66	18.13	110	99	34.48	102.40	29	100

FIGURE 2. Some frames recovered by BPDCA, iBPDCA and ADMM-DCA. From top to bottom: airport, lobby, akiyo, and hall. From left to right: original frames, downsampled frames with sr= 0.5, frames recovered by BPDCA, iBPDCA, and ADMM-DCA, respectively.

addition, some recovered frames by three algorithms are shown in Figure 2, which show that all algorithms can achieve satisfied results.

5. CONCLUSION

In this paper, we introduce a new efficient DC algorithm for solving a class of generalized DC programming. With the help of the Fenchel-Young inequality, we constructed a well-defined subproblem in the sense a unique solution can be determined iteratively, thereby circumventing

the difficulty of selecting an ideal subgradient for algorithmic implementation. Moreover, we can gainfully exploit the possibly explicit form of the proximal operator of $g(x)$ by the extended Moreau decomposition theorem. The imposed Bregman proximal term in the x-subproblem also makes our algorithm versatile for designing customized algorithms. Some computational results on matrix/tensor completion problems demonstrate that our algorithm performs well in practice.

Acknowledgments

The authors would like to thank the referees for their valuable comments, which helped us improve the presentation of this paper. H. He was supported in part by Zhejiang Provincial Natural Science Foundation of China (No. LZ24A010001), Ningbo Natural Science Foundation (Project ID: 2023J014) and National Natural Science Foundation of China (No. 12371303) and C. Ling was supported in part by National Natural Science Foundation of China (No. 11971138).

REFERENCES

[1] M. Aharon, M. Elad, A. Bruckstein, K-svd: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (2006) 4311-4322.

[2] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality, Math. Oper. Res. 35 (2010) 438-457.

[3] A. Beck, First-order Methods in Optimization, SIAM, Philadelphia (2017)

[4] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (2009) 183-202.

[5] J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Math. Program. 146 (2014), 459-494.

[6] L. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, U.S.S.R. Computational Math. Math. Phys. 7 (1967) 200-217.

[7] C.S. Chuang, H. He, L. Zhang, A unified Douglas–Rachford algorithm for generalized DC programming, J. Global Optim. 82 (2022) 331-349.

[8] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximations, Comput. Math. Appl. 2 (1976) 16-40.

[9] R. Glowinski, A. Marrocco, Approximation par éléments finis d'ordre un et résolution par pénalisation-dualité d'une classe de problèmes non linéaires, R.A.I.R.O. R2 (1975) 41-76.

[10] J. Gotoh, A. Takeda, K. Tono, DC formulations and algorithms for sparse optimization problems, Math. Program. 169 (2017) 141-176.

[11] H. He, C. Ling, W. Xie, Tensor completion via a generalized transformed tensor t-product decomposition without t-SVD, J. Sci. Comput. 93 (2022) 47.

[12] M. Kilmer, C. Martin, Factorization strategies for third-order tensors, Linear Algebra Appl. 435 (2011) 641-658.

[13] H. Le Thi, T. Pham Dinh, Feature selection in machine learning: an exact penalty approach using a difference of convex function algorithm, Mach. Learn. 101 (2015) 163-186.

[14] H. Le Thi, T. Pham Dinh, DC programming and DCA: Thirty years of developments, Math. Program. 169 (2018) 5-68.

[15] T. Liu, T. Pong, A. Takeda, A refined convergence analysis of pDCA$_e$ with applications to simultaneous sparse recovery and outlier detection, Comput. Optim. Appl. 73 (2019), 69-100.

[16] Z. Lu, Z. Zhou, Nonmonotone enhanced proximal DC algorithms for structured nonsmooth DC programming, SIAM J. Optim. 29 (2019) 2725-2752.

[17] Z. Lu, Z. Zhou, Z. Sun, Enhanced proximal DC algorithms with extrapolation for a class of structured nonsmooth DC minimization, Math. Program. 176 (2019) 369-401.
[18] B. O'Donoghue, E. Candés, Adaptive restart for accelerated gradient schemes, Found. Comput. Math. 15 (2015) 715-732.
[19] W. de Oliveira, The ABC of DC programming, Set-Valued Var. Anal. 28 (2020) 679-706.
[20] T. Pham Dinh, H. Le Thi, Convex analysis approach to DC programming: Theory, algorithms and applications. Acta Math. Vietnamica 22 (1997) 289-355.
[21] T. Pham Dinh, E.B. Souad, Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients, In: J.B. Hiriart-Urruty (ed.) Fermat Days 85: Mathematics for Optimization, North-Holland Mathematics Studies, vol. 129, pp. 249 – 271. North-Holland (1986)
[22] D. Nhat Phan, H. Minh Le, H. Le Thi, Accelerated difference of convex functions algorithm and its application to sparse binary logistic regression, IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 1369-1375, (2018)
[23] D. Nhat Phan, H. Le Thi, Difference-of-convex algorithm with extrapolation for nonconvex, nonsmooth optimization problems, Math. Oper. Res. 49 (2024) 1303-2047.
[24] R.T. Rockafellar, R. Wets, Variational Analysis, Springer-Verlag, (1998)
[25] Y. Shan, D. Hu, Z. Wang, T. Jia, Multi-channel nuclear norm minus frobenius norm minimization for color image denoising, Signal Process 207 (2023) 108-959.
[26] W. Sun, R. Sampaio, M. Candido, Proximal point algorithm for minimization of DC functions, J. Comput. Math. 21 (2003) 451-462.
[27] S. Takahashi, M. Fukuda, M. Tanaka, New Bregman proximal type algorithm for solving DC optimization problems, Comput. Optim. Appl. 83 (2002) 893-931.
[28] B. Wen, X. Chen, T.K., Pong, A proximal difference-of-convex algorithm with extrapolation, Comput. Optim. Appl. 69 (2018) 297-324.
[29] P. Yin, Y. Lou, Q. He, J. Xin, Minimization of ℓ_{1-2} for compressed sensing, SIAM J. Sci. Comput. 37 (2015) A536–A563.