Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 2, pp. 415–437, 2022

DOUBLE-INERTIAL PROXIMAL GRADIENT ALGORITHM FOR DIFFERENCE-OF-CONVEX PROGRAMMING*

Tanxing Wang, Xingju Cai, Yongzhong Song and Xue Gao†

Abstract: In this paper we study a class of difference-of-convex programming whose objective function is the sum of a smooth convex function with Lipschitz gradient, a proper closed convex function and a proper closed concave function composited with a linear operator. First, we consider the primal-dual reformulation of difference-of-convex programming. Then, adopting the framework of the double-proximal gradient algorithm (DPGA) and the inertial technique for accelerating the first-order algorithms, we propose a double-inertial proximal gradient algorithm (DiPGA) which includes some classical algorithms as its special cases. Under the assumption that the underlying function satisfies the Kurdyka-Łojasiewicz (KL) property and some suitable conditions on the parameters, we prove that each bounded sequence generated by DiPGA globally converges to a critical point of the objective function. Finally, we apply the algorithm to image processing model and compare it with DPGA to show its efficiency.

Key words: *difference-of-convex programming, double-inertial, proximal gradient algorithm, Kurdyka-Łojasiewicz property*

Mathematics Subject Classification: *90C26, 49M29, 65K05*

1 Introduction

In this paper, we consider the difference-of-convex (DC) programming whose objective function can be written as the difference of a proper closed convex function and a continuous convex function. It arises in various applications such as digital communication system [1], assignment and power allocation [26], optimal transport [11], sparse signal recovering [14, 32] and so on.

A classical algorithm for solving DC programming is so-called DC algorithm (DCA), which was proposed by Tao and An [28]. In each iteration, this algorithm replaces the concave part of the objective function by a linear majorization and solves the resulting convex optimization problem. In 2003, Sun et al. [27] introduced a proximal point algorithm for DC programming. This algorithm not only approximates the concave part in the objective function by a linear majorization in each iteration, but also approximates the convex part by a quadratic majorization. In [2, 16], the authors extended the algorithm in [27] to solve the objective function with a further convex smooth part. In a recent research on this topic [3], an accelerated variant of DCA was proposed under the additional assumption that both the convex and concave parts are continuously differentiable.

*This work was supported by the National Natural Science Foundation of China (11871279).

†Corresponding author

However, there are some drawbacks in the DCA and its variants for solving DC programming. First, the subdifferential may be a non-singleton set. In particular it may be empty or may consist of several distinct elements. Second, even if the subdifferential is singleton in each step, it might be highly discontinuous. So small deviations might lead to very different behaviors of the iteration [21]. To overcome the above drawbacks, Banert and Boţ [6] considered a primal-dual reformulation of the difference-of-convex programming. Furthermore, they proposed a general double-proximal gradient algorithm (DPGA) and proved every cluster point is a solution of the optimization problem. Actually, DPGA can fall in the framework of proximal alternating linearized minimization (PALM) algorithm [8], which is a popular algorithm for solving nonconvex and nonsmooth optimization. Under the assumption that the objective function satisfies the Kurdyka-Łojasiewicz property and some suitable conditions, [8] proved that each bounded sequence generated by PALM globally converges to a critical point.

As DPGA and PALM are the first-order algorithms, acceleration techniques are of great practical interests to improve the performance of these algorithms. In general, there are two classes of acceleration techniques, including Nesterov's acceleration [18] and inertial technique [19, 20]. In this paper, we focus on the inertial technique which was first proposed by Polyak [23]. Recently, there are increasing interests in studying inertial type algorithms, such as inertial forward-backward splitting methods for separable optimization problems under the nonconvex setting [20] and strongly convex setting [19], inertial versions of the Douglas-Rachford operator splitting method [10], inertial forward-backward-forward method [9] based on Tseng's approach [31] and general inertial proximal point method for the mixed variational inequality problem [12]. Specially, Pock and Sabach [22] proposed the inertial version of PALM (iPALM for short), and they proved that the generated bounded sequence globally converges to critical point of the objective function, assuming that the objective function possesses the Kurdyka-Łojasiewicz property. Further, [13] considered a Gauss-Seidel type inertial proximal alternating linearized minimization (GiPALM) scheme where the inertial step is performed whenever the x- or y- subproblem is updated.

The contributions of this paper can be summarized as follows. First, we propose the double-inertial proximal gradient algorithm (DiPGA), which combines the Gauss-Seidel type inertial techniques with double-proximal gradient algorithm, for solving DC programming. In this case, DiPGA may not only have better acceleration effect, but also be more universality. Furthermore, we give the global convergence analysis of DiPGA under the assumption that the objective function satisfies the Kurdyka-Łojasiewicz (KL) property. Finally, we present the numerical experimental results of the proposed algorithm applied to the image processing model and show its efficiency.

The rest of this paper is organized as follows. In Section 2, we introduce some notations and preliminary. In Section 3, we present the problem to be solved and the primal-dual reformulation of the difference-of-convex programming. We show that the primal-dual minimization problem and the primal optimization problem have a same optimal value. Meanwhile, any critical point of primal problem and critical point of its Toland dual problem forms the critical point of the primal-dual problem. In Section 4, we state the DiPGA in detail and provide its convergence analysis. Section 5 illustrates the numerical results of applying DiPGA to image processing model and compares it with DPGA.

2 Preliminaries

In this section, we summarize some notations and elementary facts for further analysis.

The theory of convex analysis in finite-dimensional spaces can refer to [24]. We shall consider functions taking values in the extended real line $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$. Let $\mathbb{H}$ be a real finite-dimensional Hilbert space, $\langle \cdot, \cdot \rangle$ denotes the inner product and the induced norm by $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The function $f : \mathbb{H} \to \overline{\mathbb{R}}$ is convex if

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda) f(x) + \lambda f(y),$$

for all $x, y \in \mathbb{H}$ and $0 \leq \lambda \leq 1$. The conjugate function $f^* : \mathbb{H} \to \overline{\mathbb{R}}$ of f is defined by

$$f^*(x^*) = \sup \{\langle x^*, x \rangle - f(x) \mid x \in \mathbb{H}\}.$$

If f is proper convex lower semicontinuous, then $f^{**} = (f^*)^* = f$ by the Fenchel-Moreau theorem [7]. If $F : \mathbb{H} \to \mathbb{G}$ is a point-to-set mapping, its graph is defined by

$$\text{Graph } F := \{(x, y) \in \mathbb{H} \times \mathbb{G} : \ y \in F(x)\}.$$

Similarly the graph of a real-extended-valued function $f : \mathbb{H} \to \overline{\mathbb{R}}$ is defined by

$$\text{Graph } f := \left\{(x, s) \in \mathbb{H} \times \overline{\mathbb{R}} : \ s = f(x)\right\}.$$

For any subset $S \subseteq \mathbb{H}$ and any point $x \in \mathbb{H}$, the distance from x to S, denoted by $\text{dist}(x, S)$, is defined as

$$\text{dist}(x, S) := \inf_{y \in S} \|y - x\|.$$

When $S = \emptyset$, we set $\text{dist}(x, S) = +\infty$ for all $x \in \mathbb{H}$.

Let us recall a few definitions concerning subdifferential calculus [4]. Recall that for $f : \ \mathbb{H} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous function, the domain of f is defined through

$$\text{dom } f := \{x \in \mathbb{H} : f(x) < +\infty\}.$$

Definition 2.1. Let $f : \ \mathbb{H} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous function.

(i) For each $x \in \text{dom } f$, the Fréchet subdifferential of f at x, written as $\hat{\partial} f(x)$, is the set of vectors $u \in \mathbb{H}$ which satisfy

$$\liminf_{y \neq x \ y \to x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0.$$

If $x \notin \text{dom } f$, we set $\hat{\partial} f(x) = \emptyset$.

(ii) The limiting-subdifferential, or simply the subdifferential for short, of f at $x \in \text{dom } f$, written as $\partial f(x)$, is defined as

$$\partial f(x) := \left\{ u \in \mathbb{H} : \exists \ x_n \to x, \ f(x_n) \to f(x), \ u_n \in \hat{\partial} f(x_n) \to u \right\}.$$

Remark 2.2. From Definition 2.1, we can find that

(i) The above definition implies $\hat{\partial} f(x) \subseteq \partial f(x)$ for each $x \in \mathbb{H}$, where the first set is closed and convex while the second one is only closed.

(ii) Let $(x_n, u_n) \in \text{Graph } \partial f$ be a sequence that converges to (x, u). By the definition of ∂f, if $f(x_n)$ converges to $f(x)$ as $n \to +\infty$, then $(x, u) \in \text{Graph } \partial f$.

(iii) A necessary condition for $x \in \mathbb{H}$ to be a local minimizer of f is that x is a critical point, that is,

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

(iv) If $f : \mathbb{H} \to \overline{\mathbb{R}}$ is proper lower semicontinuous and $h : \mathbb{H} \to \mathbb{R}$ is continuously differentiable, then for any $x \in \mathrm{dom}\ f$, $\partial(f+h)(x) = \partial f(x) + \nabla h(x)$.

Definition 2.3 (Kurdyka-Łojasiewicz property [5])**.** Let $f : \mathbb{H} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \leq +\infty$, set

$$[\eta_1 < f < \eta_2] := \{x \in \mathbb{H} : \eta_1 < f(x) < \eta_2\}.$$

We say that function f has the Kurdyka-Łojasiewicz (KL) property at $\bar{x} \in \mathrm{dom}\ \partial f$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of $\bar{x}$ and a continuous concave function $\vartheta : [0, \eta) \to \mathbb{R}_+$ such that

(i) $\vartheta(0) = 0$;

(ii) ϑ is C^1 on $(0, \eta)$;

(iii) for all $s \in (0, \eta)$, $\vartheta'(s) > 0$;

(iv) for all x in $U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, the Kurdyka-Łojasiewicz inequality holds, i.e.,

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

Definition 2.4. (Kurdyka-Łojasiewicz function [5]) Denote θ_η be the set of functions which satisfy the above definitions (i), (ii) and (iii). If f satisfies the KL property at each point of dom ∂f, then f is called a KL function.

Remark 2.5. One can easily check the Kurdyka-Łojasiewicz property is automatically satisfied at any noncritical point $\bar{x} \in \mathrm{dom}\ f$ [5].

Lemma 2.6 (Uniformized KL property [8])**.** *Let Ω be a compact set and let $f : \mathbb{H} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous function. Assume that f is constant on Ω and satisfies the KL property at each point of Ω. Then, there exist $\varepsilon > 0$, $\eta > 0$ and $\vartheta \in \theta_\eta$ such that for any $\bar{x} \in \Omega$ and all x in the following intersection:*

$$\{x \in \mathbb{H} :\ \mathrm{dist}(x, \Omega) < \varepsilon\} \cap [f(\bar{x}) < f(x) < f(\bar{x}) + \eta],$$

one has,

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

The Moreau envelope function and proximal operator are fundamental for introducing DiPGA and conducting its convergence analysis. Thus, we recall their definitions and summarize their basic properties. Let $f : \mathbb{H} \to \overline{\mathbb{R}}$ be a proper convex lower semicontinuous function. Given $v \in \mathbb{H}$ and $\lambda > 0$, the Moreau envelope function e_λ^f and the proximal operator prox_λ^f are defined respectively as

$$e_\lambda^f(v) := \inf \left\{ f(x) + \frac{1}{2\lambda} \|x - v\|^2 : x \in \mathbb{H} \right\}$$

and

$$\mathrm{prox}_\lambda^f(v) := \arg\min \ \left\{ f(x) + \frac{1}{2\lambda} \|x - v\|^2 : x \in \mathbb{H} \right\}. \tag{2.1}$$

Proposition 2.7 (Proximal behavior [25])**.** *Let $f : \mathbb{H} \to \overline{\mathbb{R}}$ be a proper lower semicontinuous function with $\inf_{\mathbb{H}} f > -\infty$. Then, for every $\lambda \in (0, +\infty)$, the set $prox_\lambda^f(v)$ is nonempty and compact. In addition, $e_\lambda^f(v)$ is finite and depends continuously on (v, λ).*

Remark 2.8. Let $\lambda > 0$ and $f : \mathbb{H} \to \overline{\mathbb{R}}$ be proper convex lower semicontinuous. The set $\text{prox}_\lambda^f(v)$ is a singleton and the proximal point is characterised by the following inequality [7],

$$f(x) \geq f(\text{prox}_\lambda^f(v)) + \frac{1}{\lambda}\langle x - \text{prox}_\lambda^f(v), v - \text{prox}_\lambda^f(v)\rangle.$$

The following lemma for smooth functions is very useful for the convergence analysis [17].

Lemma 2.9. *Let $h : \mathbb{G} \to \mathbb{R}$ be a continuous differentiable function and gradient ∇h is Lipschitz continuous with the modulus $L > 0$, then for any $x, y \in \mathbb{G}$,*

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

3 Problem Statement

Let $\mathbb{G}$ and $\mathbb{H}$ be real finite-dimensional Hilbert spaces. Let $g : \mathbb{H} \to \overline{\mathbb{R}}$ and $h : \mathbb{G} \to \overline{\mathbb{R}}$ be proper convex lower semicontinuous functions. Let $\varphi : \mathbb{H} \to \mathbb{R}$ be a convex differentiable function with L_1 Lipschitz continuous gradient for some $L_1 > 0$, and let $K : \mathbb{H} \to \mathbb{G}$ be a linear mapping and $K^* : \mathbb{G} \to \mathbb{H}$ be its adjoint. In this paper, we consider the following optimization problem:

$$\min_{x \in \mathbb{H}} \quad P(x) := g(x) + \varphi(x) - h(Kx). \tag{3.1}$$

Its Toland dual problem [29, 30] can be described as:

$$\min_{y \in \mathbb{G}} \quad D(y) := h^*(y) - (g + \varphi)^*(K^*y). \tag{3.2}$$

Furthermore, the primal-dual problem corresponding to (3.1) and (3.2) is given by

$$\min_{x \in \mathbb{H}, y \in \mathbb{G}} \quad \Phi(x, y) := g(x) + \varphi(x) + h^*(y) - \langle y, Kx\rangle, \tag{3.3}$$

where $\Phi : \mathbb{H} \times \mathbb{G} \to \overline{\mathbb{R}}$ is proper lower semicontinuous.

Let us give some relations between the problem (3.1), (3.2) and (3.3).

Proposition 3.1. (i) *(3.1), (3.2) and (3.3) have a same optimal value.*

(ii) *For all $x \in \mathbb{H}$ and $y \in \mathbb{G}$, then*

$$\Phi(x, y) \geq P(x) \quad \textit{and} \quad \Phi(x, y) \geq D(y).$$

(iii) *Let $\bar{x} \in \mathbb{H}$ be a solution of (3.1). Then $0 \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}) - \partial(h \circ K)(\bar{x})$.*

(iv) *Let $\bar{y} \in \mathbb{G}$ be a solution of (3.2). Then $0 \in \partial h^*(\bar{y}) - \partial\left((g + \varphi)^* \circ K^*\right)(\bar{y})$.*

(v) Let $(\bar{x},\bar{y}) \in \mathbb{H}\times\mathbb{G}$ be a solution of (3.3). Then $\bar{x}$ is a solution of (3.1), and $\bar{y}$ is a solution of (3.2). Furthermore, the following inclusion relations hold:

$$K^*\bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}), \tag{3.4}$$

$$K\bar{x} \in \partial h^*(\bar{y}). \tag{3.5}$$

Proof. (i) Applying the Fenchel-Moreau theorem [7] to h, we have

$$\begin{aligned}
&\inf\{g(x)+\varphi(x)-h(Kx) \mid x\in\mathbb{H}\}\\
&=\inf\{g(x)+\varphi(x)-h^{**}(Kx) \mid x\in\mathbb{H}\}\\
&=\inf\{g(x)+\varphi(x)-\sup\{\langle y,Kx\rangle-h^*(y) \mid y\in\mathbb{G}\} \mid x\in\mathbb{H}\}\\
&=\inf\{g(x)+\varphi(x)+h^*(y)-\langle y,Kx\rangle \mid x\in\mathbb{H}, y\in\mathbb{G}\}\\
&=\inf\{h^*(y)-\sup\{\langle x,K^*y\rangle-(g+\varphi)(x) \mid x\in\mathbb{H}\} \mid y\in\mathbb{G}\}\\
&=\inf\{h^*(y)-(g+\varphi)^*(K^*y) \mid y\in\mathbb{G}\}.
\end{aligned}$$

(ii) From the definition of the conjugate function, we know that for any $x\in\mathbb{H}$ and $y\in\mathbb{G}$,

$$\begin{aligned}
g(x)+\varphi(x)-h(Kx) &= g(x)+\varphi(x)-h^{**}(Kx)\\
&= g(x)+\varphi(x)-\sup\{\langle Kx,\tilde{y}\rangle-h^*(\tilde{y}) \mid \tilde{y}\in\mathbb{G}\}\\
&\le g(x)+\varphi(x)-\langle Kx,y\rangle+h^*(y).
\end{aligned}$$

The other inequality is verified by an analogous calculation.
(iii) By definition, it is trivial [27].
(iv) The proof of this statement is analogous with (iii).
(v) Let $(\bar{x},\bar{y})$ be a solution of (3.3). It is worth noting that the optimal values of (3.1), (3.2) and (3.3) must be finite if such a solution exists. Fixed $\bar{y}\in\mathbb{H}$, the function $\Phi(x,\bar{y})$ is convex and takes a minimum at $\bar{x}$. Thus

$$0\in\partial g(\bar{x})+\nabla\varphi(\bar{x})-K^*\bar{y}.$$

(3.4) follows directly. The same argument works for the function $\Phi(\bar{x},y)$ and implies

$$0\in\partial h^*(\bar{y})-K\bar{x}.$$

According to Young-Fenchel inequality [7], we know the following equality from (3.4) and (3.5),

$$h^*(\bar{y})+h(K\bar{x})=\langle\bar{y},K\bar{x}\rangle,$$

$$(g+\varphi)^*(K^*\bar{y})+(g+\varphi)(\bar{x})=\langle\bar{x},K^*\bar{y}\rangle.$$

Therefore, by subtracting these equalities,

$$\begin{aligned}
(g+\varphi)(\bar{x})-h(K\bar{x}) &= h^*(\bar{y})-(g+\varphi)^*(K^*\bar{y})\\
&= h^*(\bar{y})-\sup\{\langle x,K^*\bar{y}\rangle-g(x)-\varphi(x) \mid x\in\mathbb{H}\}\\
&\le h^*(\bar{y})+g(\bar{x})+\varphi(\bar{x})-\langle\bar{x},K^*\bar{y}\rangle.
\end{aligned}$$

Since $(\bar{x},\bar{y})$ is a solution of (3.3), by (i), we have $P(\bar{x})=\inf P(x)=Q(\bar{y})$. □

Definition 3.2. (Critical points [28]) We denote the set of critical points $(\bar{x}, \bar{y}) \in \mathbb{H} \times \mathbb{G}$ of the function Φ, which satisfy the inclusions (3.4) and (3.5), by crit Φ. The set of critical points of the objective function P is defined by

$$\text{crit } P := \{x \in \mathbb{H} : \partial(h \circ K)(x) \cap (\partial g(x) + \nabla \varphi(x)) \neq \emptyset\},$$

and similarly the set of critical points of the objective function D is denoted by

$$\text{crit } D := \{y \in \mathbb{G} : \partial\left((g + \varphi)^* \circ K^*\right)(y) \cap \partial h^*(y) \neq \emptyset\}.$$

Remark 3.3. From Definition 3.2, we know the following results.

(i) If $(\bar{x}, \bar{y}) \in \mathbb{H} \times \mathbb{G}$ is a critical point of Φ, then

$$K^*\bar{y} \in K^*\partial h(K\bar{x}) \cap (\partial g(\bar{x}) + \nabla \varphi(\bar{x})), \tag{3.6}$$

$$K\bar{x} \in K\partial(g + \varphi)^*(K^*\bar{y}) \cap \partial h^*(\bar{y}). \tag{3.7}$$

(ii) Since $K^*\partial h(Kx) \subseteq \partial(h \circ K)(x)$ and $K\partial(g + \varphi)^*(K^*y) \subseteq \partial\left((g + \varphi) \circ K^*\right)(y)$, we know that if $(\bar{x}, \bar{y}) \in \mathbb{H} \times \mathbb{G}$ is a critical point of the objective function Φ, then $\bar{x}$ is a critical point of P and $\bar{y}$ is a critical point of D.

4 DiPGA and its Convergence Analysis

We propose the following double-inertial proximal gradient algorithm (DiPGA) iterative scheme for difference-of-convex programming.

Algorithm 1: Double-inertial proximal gradient algorithm.

Step 1. Choose starting point $(x_0, \bar{x}_0, y_0, \bar{y}_0) \in \mathbb{H} \times \mathbb{H} \times \mathbb{G} \times \mathbb{G}$, $\bar{x}_0 = x_0,\ \bar{y}_0 = y_0$. Take $\alpha_1, \beta_1, \alpha_2, \beta_2 \in [0, 1)$ and $\gamma > 0, \mu > 0$.
Step 2. For each $n = 0, 1, \cdots$, $\{(x_n, y_n)\}_{n \in \mathbb{N}}$ is generated as follows:
Compute

$$x_{n+1} = \text{prox}_{\gamma g}\left(x_n + \gamma K^*\bar{y}_n - \gamma \nabla \varphi(\bar{x}_n) + \beta_1\left(x_n - \bar{x}_{n-1}\right)\right), \tag{4.1}$$

$$\bar{x}_{n+1} = x_{n+1} + \alpha_1(x_{n+1} - \bar{x}_n). \tag{4.2}$$

Compute

$$y_{n+1} = \text{prox}_{\mu h^*}\left(y_n + \mu K\bar{x}_{n+1} + \beta_2\left(y_n - \bar{y}_{n-1}\right)\right), \tag{4.3}$$

$$\bar{y}_{n+1} = y_{n+1} + \alpha_2(y_{n+1} - \bar{y}_n). \tag{4.4}$$

Remark 4.1. (i) If $\alpha_1 = \alpha_2 = 0$, then DiPGA reduces to inertial proximal algorithm [20]. It should be noted that the inertial proximal algorithm is also a special case of iPALM [22]. Specially, we take the Gauss-Seidel type inertial scheme in DiPGA, i.e., the inertial step is performed whenever $x-$ or $y-$ subproblem is updated. Correspondingly, iPALM takes the Jacobian type inertial scheme and the inertial step can be iterated in parallel.

(ii) If $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$, then DiPGA reduces to a Gauss-Seidel type inertial proximal alternating linearized minimization algorithm [13].

4.1 Some preparations

First, we introduce some useful notations in the whole process as follows:

$$z_n := (x_n, y_n),\ \ \bar{z}_n := (\bar{x}_n, \bar{y}_n) \text{ and } \hat{z}_n := (x_n, y_n, \bar{x}_n, \bar{y}_n).$$

Lemma 4.2. *Let $\{\hat{z}_n\}_{n\in\mathbb{N}}$ be the sequence generated by DiPGA. Then we have*

$$\begin{aligned}&\Phi(x_{n+1}, y_{n+1}) + M_1 \|x_{n+1} - \bar{x}_{n+1}\|^2 + M_2 \|y_{n+1} - \bar{y}_{n+1}\|^2\\ &\le \Phi(x_n, y_n) + M_3 \|x_n - \bar{x}_n\|^2 + M_4 \|y_n - \bar{y}_n\|^2,\end{aligned} \tag{4.5}$$

with

$$\begin{aligned}&M_1 = \frac{1}{2\gamma\alpha_1^2} - \frac{L_1}{2\alpha_1^2} - \frac{1+\alpha_2}{2\alpha_1^2} - \frac{1}{2\alpha_1^2}, \qquad M_2 = \frac{1}{2\mu\alpha_2^2} - \frac{\alpha_1^2\|K\|^2}{2\alpha_2^2},\\ &M_3 = \frac{1}{2\gamma} + \frac{(\alpha_1-\beta_1)^2}{2\gamma\alpha_1^2} + \frac{1}{2\alpha_1},\\ &M_4 = \frac{1}{2\mu} + \frac{(1+\alpha_1)\|K\|^2}{2} + \frac{(\alpha_2-\beta_2)^2}{2\mu\alpha_2^2} + \frac{\alpha_1^2\|K\|^2}{2\alpha_2},\end{aligned}$$

where L_1 is the Lipschitz constant of $\nabla\varphi$.

Proof. By the definition of the proximal operator given in (2.1) and the iterative step (4.1), we know

$$\begin{aligned}x_{n+1} = \underset{x\in\mathbb{H}}{\operatorname{argmin}}\ \{g(x) + \frac{1}{2\gamma}\|x - \bar{x}_n - \gamma K^*\bar{y}_n\\ + \gamma\nabla\varphi(\bar{x}_n) + (\alpha_1-\beta_1)(x_n - \bar{x}_{n-1})\|^2\}.\end{aligned} \tag{4.6}$$

The optimality condition of (4.6) is, for any $x \in \mathbb{H}$,

$$g(x) - g(x_{n+1}) + \frac{1}{\gamma}\langle x - x_{n+1}, x_{n+1} - \bar{x}_n - \gamma K^*\bar{y}_n + \gamma\nabla\varphi(\bar{x}_n) + (\alpha_1-\beta_1)(x_n - \bar{x}_{n-1})\rangle \ge 0.$$

Particularly, taking $x = x_n$,

$$\begin{aligned}g(x_{n+1}) - g(x_n) &\le \langle x_n - x_{n+1}, \nabla\varphi(\bar{x}_n) - K^*\bar{y}_n\rangle + \frac{\alpha_1-\beta_1}{\gamma}\langle x_n - x_{n+1}, x_n - \bar{x}_{n-1}\rangle\\ &\quad + \frac{1}{\gamma}\langle x_n - x_{n+1}, x_{n+1} - \bar{x}_n\rangle\\ &\le \langle x_n - x_{n+1}, \nabla\varphi(\bar{x}_n) - K^*\bar{y}_n\rangle + \frac{\alpha_1-\beta_1}{\gamma}\langle x_n - x_{n+1}, x_n - \bar{x}_{n-1}\rangle\\ &\quad + \frac{1}{2\gamma}\|x_n - \bar{x}_n\|^2 - \frac{1}{2\gamma}\|x_{n+1} - \bar{x}_n\|^2 - \frac{1}{2\gamma}\|x_{n+1} - x_n\|^2,\end{aligned} \tag{4.7}$$

where the second inequality dues to the relation

$$\langle a-b, c-d\rangle = \frac{1}{2}(\|a-d\|^2 - \|a-c\|^2) + \frac{1}{2}(\|b-c\|^2 - \|b-d\|^2).$$

Similarly, using (2.1) and the iterative step (4.3),

$$y_{n+1} = \underset{y\in\mathbb{G}}{\operatorname{argmin}}\ \{h^*(y) + \frac{1}{2\mu}\|y - \bar{y}_n - \mu K\bar{x}_{n+1} + (\alpha_2-\beta_2)(y_n - \bar{y}_{n-1})\|^2\}.$$

Particularly, taking $y = y_n$,

$$\begin{aligned} h^*(y_{n+1}) - h^*(y_n) \leq & \langle y_n - y_{n+1}, -K\bar{x}_{n+1}\rangle + \frac{\alpha_2 - \beta_2}{\mu}\langle y_n - y_{n+1}, y_n - \bar{y}_{n-1}\rangle \\ & + \frac{1}{2\mu}\|y_n - \bar{y}_n\|^2 - \frac{1}{2\mu}\|y_{n+1} - \bar{y}_n\|^2 - \frac{1}{2\mu}\|y_{n+1} - y_n\|^2. \end{aligned} \tag{4.8}$$

On the other hand, using the smoothness and convexity of φ, we have

$$\varphi(x_{n+1}) - \varphi(\bar{x}_n) \leq \langle \nabla\varphi(\bar{x}_n), x_{n+1} - \bar{x}_n\rangle + \frac{L_1}{2}\|x_{n+1} - \bar{x}_n\|^2, \tag{4.9}$$

$$\varphi(\bar{x}_n) - \varphi(x_n) \leq \langle \nabla\varphi(\bar{x}_n), \bar{x}_n - x_n\rangle. \tag{4.10}$$

Adding (4.7), (4.8), (4.9) and (4.10) to get

$$\begin{aligned} & g(x_{n+1}) + \varphi(x_{n+1}) + h^*(y_{n+1}) - g(x_n) - \varphi(x_n) - h^*(y_n) \\ & \leq \langle x_n - x_{n+1}, \nabla\varphi(\bar{x}_n) - K^*\bar{y}_n\rangle + \frac{\alpha_1 - \beta_1}{\gamma}\langle x_n - x_{n+1}, x_n - \bar{x}_{n-1}\rangle + \frac{L_1}{2}\|x_{n+1} - \bar{x}_n\|^2 \\ & + \frac{1}{2\gamma}\|x_n - \bar{x}_n\|^2 - \frac{1}{2\gamma}\|x_{n+1} - \bar{x}_n\|^2 - \frac{1}{2\gamma}\|x_{n+1} - x_n\|^2 + \langle \nabla\varphi(\bar{x}_n), x_{n+1} - x_n\rangle \\ & + \langle y_n - y_{n+1}, -K\bar{x}_{n+1}\rangle + \frac{\alpha_2 - \beta_2}{\mu}\langle y_n - y_{n+1}, y_n - \bar{y}_{n-1}\rangle \\ & + \frac{1}{2\mu}\|y_n - \bar{y}_n\|^2 - \frac{1}{2\mu}\|y_{n+1} - \bar{y}_n\|^2 - \frac{1}{2\mu}\|y_{n+1} - y_n\|^2. \end{aligned}$$

Using the definition of Φ, we obtain

$$\begin{aligned} \Phi(x_{n+1}, y_{n+1}) \leq & \Phi(x_n, y_n) + \langle x_n - x_{n+1}, -K^*\bar{y}_n\rangle + \langle y_n - y_{n+1}, -K\bar{x}_{n+1}\rangle \\ & - \langle y_{n+1}, Kx_{n+1}\rangle + \langle y_n, Kx_n\rangle + \frac{L_1}{2}\|x_{n+1} - \bar{x}_n\|^2 \\ & + \frac{1}{2\gamma}\|x_n - \bar{x}_n\|^2 - \frac{1}{2\gamma}\|x_{n+1} - \bar{x}_n\|^2 - \frac{1}{2\gamma}\|x_{n+1} - x_n\|^2 \\ & + \frac{1}{2\mu}\|y_n - \bar{y}_n\|^2 - \frac{1}{2\mu}\|y_{n+1} - \bar{y}_n\|^2 - \frac{1}{2\mu}\|y_{n+1} - y_n\|^2 \\ & + \frac{|\alpha_1 - \beta_1|}{2s_1\gamma}\|x_n - \bar{x}_{n-1}\|^2 + \frac{|\alpha_1 - \beta_1|s_1}{2\gamma}\|x_n - x_{n+1}\|^2 \\ & + \frac{|\alpha_2 - \beta_2|}{2s_2\mu}\|y_n - \bar{y}_{n-1}\|^2 + \frac{|\alpha_2 - \beta_2|s_2}{2\mu}\|y_n - y_{n+1}\|^2, \end{aligned} \tag{4.11}$$

where $s_1 > 0, s_2 > 0$. From (4.2), we obtain

$$\bar{x}_{n+1} - x_{n+1} = \alpha_1(x_{n+1} - \bar{x}_n) = \alpha_1(x_{n+1} - x_n) + \alpha_1(x_n - \bar{x}_n).$$

Thus,

$$\|x_{n+1} - \bar{x}_n\|^2 = \frac{1}{\alpha_1^2}\|\bar{x}_{n+1} - x_{n+1}\|^2,$$

and

$$\begin{aligned} \|x_{n+1} - x_n\|^2 = & \frac{1}{\alpha_1^2}\|(\bar{x}_{n+1} - x_{n+1}) - \alpha_1(x_n - \bar{x}_n)\|^2 \\ = & \frac{1}{\alpha_1^2}\|\bar{x}_{n+1} - x_{n+1}\|^2 + \|x_n - \bar{x}_n\|^2 - \frac{2}{\alpha_1}\langle \bar{x}_{n+1} - x_{n+1}, x_n - \bar{x}_n\rangle \\ \leq & \frac{1 + \alpha_1}{\alpha_1^2}\|\bar{x}_{n+1} - x_{n+1}\|^2 + \frac{1 + \alpha_1}{\alpha_1}\|x_n - \bar{x}_n\|^2, \end{aligned} \tag{4.12}$$

where the last inequality follows from Cauchy-Schwarz and Young inequalities.

Similarly, using (4.4) we have

$$\|y_{n+1}-\bar{y}_n\|^2 = \frac{1}{\alpha_2^2}\|\bar{y}_{n+1}-y_{n+1}\|^2,$$

and

$$\|y_{n+1}-y_n\|^2 \le \frac{1+\alpha_2}{\alpha_2^2}\|\bar{y}_{n+1}-y_{n+1}\|^2 + \frac{1+\alpha_2}{\alpha_2}\|y_n-\bar{y}_n\|^2. \tag{4.13}$$

On the other hand, for any $s_3>0$ and $s_4>0$, we have

$$\begin{aligned}
&\langle Kx_n - Kx_{n+1}, -\bar{y}_n\rangle + \langle y_n - y_{n+1}, -K\bar{x}_{n+1}\rangle - \langle y_{n+1}, Kx_{n+1}\rangle + \langle y_n, Kx_n\rangle \\
&= \langle K(\bar{x}_{n+1}-x_{n+1}), y_{n+1}-y_n\rangle + \langle K(x_n - x_{n+1}), y_n-\bar{y}_n\rangle \\
&\le \frac{\|K\|^2}{2s_3}\|\bar{x}_{n+1}-x_{n+1}\|^2 + \frac{s_3}{2}\|y_{n+1}-y_n\|^2 + \frac{\|K\|^2 s_4}{2}\|x_{n+1}-x_n\|^2 + \frac{1}{2s_4}\|y_n-\bar{y}_n\|^2.
\end{aligned} \tag{4.14}$$

Combining (4.11), (4.12), (4.13) with (4.14) yields

$$\begin{aligned}
&\Phi(x_{n+1},y_{n+1}) \le \Phi(x_n,y_n) \\
&+ \left(\frac{1}{2\gamma} + \frac{(1+\alpha_1)(|\alpha_1-\beta_1|s_1-1)}{2\alpha_1\gamma} + \frac{|\alpha_1-\beta_1|}{2s_1\gamma\alpha_1^2} + \frac{(1+\alpha_1)\|K\|^2 s_4}{2\alpha_1}\right)\|x_n-\bar{x}_n\|^2 \\
&+ \left(\frac{1}{2\mu} + \frac{(1+\alpha_2)(|\alpha_2-\beta_2|s_2-1)}{2\alpha_2\mu} + \frac{1}{2s_4} + \frac{|\alpha_2-\beta_2|}{2s_2\mu\alpha_2^2} + \frac{(1+\alpha_2)s_3}{2\alpha_2}\right)\|y_n-\bar{y}_n\|^2 \\
&- \left(\frac{1-(1+\alpha_1)(|\alpha_1-\beta_1|s_1-1)}{2\gamma\alpha_1^2} - \frac{L_1}{2\alpha_1^2} - \frac{\|K\|^2}{2s_3} - \frac{(1+\alpha_1)\|K\|^2 s_4}{2\alpha_1^2}\right)\|x_{n+1}-\bar{x}_{n+1}\|^2 \\
&- \left(\frac{1}{2\mu\alpha_2^2} - \frac{(1+\alpha_2)(|\alpha_2-\beta_2|s_2-1)}{2\mu\alpha_2^2} - \frac{(1+\alpha_2)s_3}{2\alpha_2^2}\right)\|y_{n+1}-\bar{y}_{n+1}\|^2.
\end{aligned} \tag{4.15}$$

Take $s_1 = \frac{1}{|\alpha_1-\beta_1|}$, $s_2 = \frac{1}{|\alpha_2-\beta_2|}$, $s_3 = \frac{\alpha_1^2\|K\|^2}{1+\alpha_2}$ and $s_4 = \frac{1}{(1+\alpha_1)\|K\|^2}$, we know

$$\begin{aligned}
\Phi(x_{n+1},y_{n+1}) \le \Phi(x_n,y_n) &+ \left(\frac{1}{2\gamma} + \frac{(\alpha_1-\beta_1)^2}{2\gamma\alpha_1^2} + \frac{1}{2\alpha_1}\right)\|x_n-\bar{x}_n\|^2 \\
&- \left(\frac{1}{2\gamma\alpha_1^2} - \frac{L_1}{2\alpha_1^2} - \frac{1+\alpha_2}{2\alpha_1^2} - \frac{1}{2\alpha_1^2}\right)\|x_{n+1}-\bar{x}_{n+1}\|^2 \\
&+ \left(\frac{1}{2\mu} + \frac{(1+\alpha_1)\|K\|^2}{2} + \frac{(\alpha_2-\beta_2)^2}{2\mu\alpha_2^2} + \frac{\alpha_1^2\|K\|^2}{2\alpha_2}\right)\|y_n-\bar{y}_n\|^2 \\
&- \left(\frac{1}{2\mu\alpha_2^2} - \frac{\alpha_1^2\|K\|^2}{2\alpha_2^2}\right)\|y_{n+1}-\bar{y}_{n+1}\|^2.
\end{aligned} \tag{4.16}$$

It can be seen that (4.5) follows immediately from (4.16). □

Remark 4.3. Let $\epsilon > 0$ be a real number for which

$$s = (1-\epsilon) - (\alpha_1^2 + (\alpha_1-\beta_1)^2)(1+\epsilon) > 0, \quad t = (1-\epsilon) - (\alpha_2^2 + (\alpha_2-\beta_2)^2)(1+\epsilon) > 0.$$

(i) If

$$\delta_1 = \frac{(L_1 + 2 + \alpha_2)(\alpha_1^2 + (\alpha_1 - \beta_1)^2)}{2\alpha_1^2 s} + \frac{1}{2\alpha_1 s},$$

$$\frac{1}{\gamma} = 2 + L_1 + \alpha_2 + 2(1+\epsilon)\alpha_1^2 \delta_1, \tag{4.17}$$

then $M_1 = (1+\epsilon)\delta_1$ and $M_3 = (1-\epsilon)\delta_1$.

(ii) If

$$\delta_2 = \frac{\alpha_1^2 \|K\|^2 (\alpha_2^2 + (\alpha_2 - \beta_2)^2)}{2\alpha_2^2 t} + \frac{(1+\alpha_1)\|K\|^2}{2t} + \frac{\alpha_1^2 \|K\|^2}{2\alpha_2 t},$$

$$\frac{1}{\mu} = \alpha_1^2 \|K\|^2 + 2(1+\epsilon)\alpha_2^2 \delta_2, \tag{4.18}$$

then $M_2 = (1+\epsilon)\delta_2$ and $M_4 = (1-\epsilon)\delta_2$.

Equipped with the above results, we construct a function $S : \mathbb{H} \times \mathbb{G} \times \mathbb{H} \times \mathbb{G} \to \overline{\mathbb{R}}$,

$$S(x, y, \bar{x}, \bar{y}) = \Phi(x, y) + \delta_1 \|x - \bar{x}\|^2 + \delta_2 \|y - \bar{y}\|^2 .$$

Lemma 4.4. *Take the parameters γ, μ according to* (4.17) *and* (4.18) *respectively as shown in Remark 4.3. Suppose Φ is assumed to be bounded below. Let $\{\hat{z}_n\}_{n\in\mathbb{N}}$ be the sequence generated by DiPGA. Then the following statements hold.*

(i) *The sequence $\{S(\hat{z}_n)\}_{n\in\mathbb{N}}$ is nonincreasing and in particular,*

$$\rho_1 \left(\|z_n - \bar{z}_n\|^2 + \|z_{n+1} - \bar{z}_{n+1}\|^2 \right) \leq S(\hat{z}_n) - S(\hat{z}_{n+1}), \tag{4.19}$$

where $\rho_1 = \epsilon \min\{\delta_1, \delta_2\}$.

(ii) *We have*

$$\sum_{n=0}^{+\infty} \|z_n - \bar{z}_n\|^2 < +\infty,$$

which means that $\lim_{n\to+\infty} \|z_n - \bar{z}_n\| = 0$, *and hence* $\lim_{n\to+\infty} \|z_{n+1} - z_n\| = 0$.

Proof. (i) Take $\rho_1 = \epsilon \min\{\delta_1, \delta_2\}$, then the conclusion follows immediately from Lemma 4.2 and Remark 4.3.

(ii) Since Φ is assumed to be bounded below, S is also bounded below. From (i), we know $\{S(\hat{z}_n)\}_{n\in\mathbb{N}}$ is nonincreasing, hence it converges to some real number $\bar{S}$.

Let N be a positive integer. Summing (4.19) from $n = 0$ to $N - 1$ we get

$$\sum_{n=0}^{N-1} \left(\|z_n - \bar{z}_n\|^2 + \|z_{n+1} - \bar{z}_{n+1}\|^2 \right) \leq \frac{1}{\rho_1} (S(\hat{z}_0) - S(\hat{z}_N)).$$

According to Remark 4.3, it is easy to check that ρ_1 is inf-bounded. Taking limit as $N \to +\infty$, we obtain

$$\sum_{n=0}^{+\infty}\left(\|z_n-\bar{z}_n\|^2+\|z_{n+1}-\bar{z}_{n+1}\|^2\right)\leq \frac{1}{\rho_1}(S(\hat{z}_0)-\bar{S})<+\infty.$$

Hence, $\sum_{n=0}^{+\infty}\|z_n-\bar{z}_n\|^2<+\infty$, which deduces that $\lim_{n\to+\infty}\|z_n-\bar{z}_n\|=0$.

From (4.12) and (4.13), we have

$$\sum_{n=0}^{+\infty}\|x_{n+1}-x_n\|^2<+\infty,$$

and

$$\sum_{n=0}^{+\infty}\|y_{n+1}-y_n\|^2<+\infty.$$

So we can easily get

$$\sum_{n=0}^{+\infty}\|z_{n+1}-z_n\|^2=\sum_{n=0}^{+\infty}\left(\|x_{n+1}-x_n\|^2+\|y_{n+1}-y_n\|^2\right)<+\infty,$$

which deduces that $\lim_{n\to+\infty}\|z_{n+1}-z_n\|=0$. □

Lemma 4.5. *Take the parameters γ, μ according to (4.17) and (4.18) respectively as shown in Remark 4.3. Let $\{\hat{z}_n\}_{n\in\mathbb{N}}$ be the sequence generated by DiPGA. For all $n\geq 0$, define*

$$A^{n+1}=(A_x^{n+1},A_y^{n+1},A_{\bar{x}}^{n+1},A_{\bar{y}}^{n+1}),$$

where

$$\begin{cases} A_x^{n+1}=\dfrac{\bar{x}_n-x_{n+1}}{\gamma}+\nabla\varphi(x_{n+1})-\nabla\varphi(\bar{x}_n)-K^*y_{n+1}+K^*\bar{y}_n \\ \qquad\quad +2\delta_1(x_{n+1}-\bar{x}_{n+1})+\dfrac{\beta_1-\alpha_1}{\gamma}(x_n-\bar{x}_{n-1}), \\ A_y^{n+1}=\dfrac{\bar{y}_n-y_{n+1}}{\mu}+K\bar{x}_{n+1}-Kx_{n+1}+2\delta_2(y_{n+1}-\bar{y}_{n+1})+\dfrac{\beta_2-\alpha_2}{\mu}(y_n-\bar{y}_{n-1}), \\ A_{\bar{x}}^{n+1}=2\delta_1(\bar{x}_{n+1}-x_{n+1}), \\ A_{\bar{y}}^{n+1}=2\delta_2(\bar{y}_{n+1}-y_{n+1}). \end{cases}$$

Then $A^{n+1}\in\partial S(\hat{z}_{n+1})$ and

$$\|A^{n+1}\|\leq\rho_2(\|\bar{z}_{n+1}-z_{n+1}\|+\|\bar{z}_n-z_n\|),$$

where $\rho_2=\sqrt{2}\max\left\{\dfrac{1+L_1\gamma}{\alpha_1\gamma}+\|K\|+4\delta_1,\dfrac{1+\mu\|K^*\|}{\alpha_2\mu}+4\delta_2,\dfrac{|\beta_1-\alpha_1|}{\alpha_1\gamma},\dfrac{|\beta_2-\alpha_2|}{\alpha_2\mu}\right\}.$

Proof. The first order optimality conditions of (4.1) and (4.3) yield

$$0 \in \partial g(x_{n+1}) + \frac{1}{\gamma}\left(x_{n+1} - \bar{x}_n - \gamma K^* \bar{y}_n + \gamma \nabla\varphi(\bar{x}_n) + (\alpha_1 - \beta_1)(x_n - \bar{x}_{n-1})\right),$$

$$0 \in \partial h^*(y_{n+1}) + \frac{1}{\mu}(y_{n+1} - \bar{y}_n - \mu K \bar{x}_{n+1} + (\alpha_2 - \beta_2)(y_n - \bar{y}_{n-1})).$$

Consider the function $S(x, y, \bar{x}, \bar{y}) = \Phi(x, y) + \delta_1\|x - \bar{x}\|^2 + \delta_2\|y - \bar{y}\|^2$. By the general calculation of the subdifferential [25] and Remark 2.2 (iii), we obtain

$$\begin{cases} \partial_x S(x, y, \bar{x}, \bar{y}) = \partial g(x) + \nabla\varphi(x) - K^* y + 2\delta_1(x - \bar{x}), \\ \partial_y S(x, y, \bar{x}, \bar{y}) = \partial h^*(y) - Kx + 2\delta_2(y - \bar{y}), \\ \partial_{\bar{x}} S(x, y, \bar{x}, \bar{y}) = 2\delta_1(\bar{x} - x), \\ \partial_{\bar{y}} S(x, y, \bar{x}, \bar{y}) = 2\delta_2(\bar{y} - y). \end{cases}$$

Then we get that $A^{n+1} \in \partial S(\hat{z}_{n+1})$. Moreover,

$$\begin{aligned} \left\|A_x^{n+1}\right\| \le{}& \frac{1}{\alpha_1\gamma}\|x_{n+1} - \bar{x}_{n+1}\| + \frac{L_1}{\alpha_1}\|x_{n+1} - \bar{x}_{n+1}\| \\ &+ \|K^*\|\,\|y_{n+1} - \bar{y}_n\| + 2\delta_1\|x_{n+1} - \bar{x}_{n+1}\| + \frac{|\beta_1 - \alpha_1|}{\alpha_1\gamma}\|x_n - \bar{x}_n\| \\ &= (\frac{1 + L_1\gamma}{\alpha_1\gamma} + 2\delta_1)\|x_{n+1} - \bar{x}_{n+1}\| + \frac{\|K^*\|}{\alpha_2}\|y_{n+1} - \bar{y}_{n+1}\| + \frac{|\beta_1 - \alpha_1|}{\alpha_1\gamma}\|x_n - \bar{x}_n\|, \end{aligned}$$

and

$$\begin{aligned} \left\|A_y^{n+1}\right\| \le{}& \frac{1}{\alpha_2\mu}\|y_{n+1} - \bar{y}_{n+1}\| + \|K\|\,\|x_{n+1} - \bar{x}_{n+1}\| + 2\delta_2\|y_{n+1} - \bar{y}_{n+1}\| \\ &+ \frac{|\beta_2 - \alpha_2|}{\alpha_2\mu}\|y_n - \bar{y}_n\| \\ &= (\frac{1}{\alpha_2\mu} + 2\delta_2)\|y_{n+1} - \bar{y}_{n+1}\| + \|K\|\,\|x_{n+1} - \bar{x}_{n+1}\| + \frac{|\beta_2 - \alpha_2|}{\alpha_2\mu}\|y_n - \bar{y}_n\|. \end{aligned}$$

Thus we can get

$$\begin{aligned} \left\|A^{n+1}\right\| \le{}& \left\|A_x^{n+1}\right\| + \left\|A_y^{n+1}\right\| + \left\|A_{\bar{x}}^{n+1}\right\| + \left\|A_{\bar{y}}^{n+1}\right\| \\ &\le (\frac{1 + L_1\gamma}{\alpha_1\gamma} + \|K\| + 4\delta_1)\|x_{n+1} - \bar{x}_{n+1}\| + (\frac{1 + \mu\|K^*\|}{\alpha_2\mu} + 4\delta_2)\|y_{n+1} - \bar{y}_{n+1}\| \\ &+ \frac{|\beta_1 - \alpha_1|}{\alpha_1\gamma}\|x_n - \bar{x}_n\| + \frac{|\beta_2 - \alpha_2|}{\alpha_2\mu}\|y_n - \bar{y}_n\| \\ &\le \rho_2(\|\bar{z}_{n+1} - z_{n+1}\| + \|\bar{z}_n - z_n\|). \end{aligned}$$

This completes the proof. □

4.2 Convergence analysis of DiPGA

The following result summarizes several properties of the cluster point set. Let $\hat{z}^* := (x^*, y^*, x^*, y^*)$ be a cluster point of the sequence $\{\hat{z}_n\}$. The cluster point set of $\{\hat{z}_n\}$ is denoted by Ω^*.

Lemma 4.6. *Take the parameters γ, μ according to* (4.17) *and* (4.18) *respectively as shown in Remark 4.3. The sequence generated by DiPGA is denoted as $\{\hat{z}_n\}_{n\in\mathbb{N}}$, which is assumed to be bounded. Then the following assertions hold.*

(i) *Ω^* is a nonempty compact set, and*

$$\lim_{n\to+\infty} \operatorname{dist}(\hat{z}_n, \Omega^*) = 0. \tag{4.20}$$

(ii) *$\Omega^* \subseteq crit\ S$.*

(iii) *The function S is constant on Ω^*.*

Proof. (i) First, it is clear that the set of cluster points of a bounded sequence is nonempty. Set $\Omega = \Omega^*$. Observe that Ω can be viewed as an intersection of compact sets

$$\Omega = \bigcap_{q\in N} \overline{\bigcup_{n\geq q} \{\hat{z}_n\}},$$

so it is also compact. Moreover, assume (4.20) does not hold. In this case, there exist an $\varepsilon > 0$ and a subsequence $\{\hat{z}_{n_k}\}_{k\in\mathbb{N}}$ of $\{\hat{z}_n\}_{n\in\mathbb{N}}$ with $\operatorname{dist}(\hat{z}_{n_k}, \Omega^*) > \varepsilon$ for all $k \geq 0$. However, the cluster point of this subsequence is also an element of Ω^*. This contradicts the assumption that $\operatorname{dist}(\hat{z}_{n_k}, \Omega^*) > \varepsilon$ for all $k \geq 0$.
(ii) Let $\hat{z}^* \in \Omega^*$, then there exists a subsequence $\{\hat{z}_{n_k}\}_{k\in\mathbb{N}}$ of $\{\hat{z}_n\}_{n\in\mathbb{N}}$ converging to $\hat{z}^*$. So we have $\bar{x}_{n_k-1} \to x^*$ as $k \to +\infty$. Note that Lemma 4.4 implies

$$\lim_{k\to+\infty} \|x_{n_k} - \bar{x}_{n_k-1}\| = 0.$$

From the iterative step (4.1), we have

$$\begin{aligned} x_{n+1} \in \operatorname*{argmin}_{x\in\mathbb{H}} \Big\{ & g(x) + \frac{1}{\gamma} \langle \gamma\nabla\varphi(\bar{x}_n) - \gamma K^*\bar{y}_n + (\alpha_1 - \beta_1)(x_n - \bar{x}_{n-1}), x - \bar{x}_n \rangle \\ & + \frac{1}{2\gamma} \|x - \bar{x}_n\|^2 \Big\}. \end{aligned} \tag{4.21}$$

Take $x = x^*$ in (4.21), we get

$$\begin{aligned} & g(x_{n+1}) + \langle x_{n+1} - \bar{x}_n, \nabla\varphi(\bar{x}_n) - K^*\bar{y}_n - \frac{\beta_1 - \alpha_1}{\gamma}(x_n - \bar{x}_{n-1}) \rangle + \frac{1}{2\gamma} \|x_{n+1} - \bar{x}_n\|^2 \\ & \leq g(x^*) + \langle x^* - \bar{x}_n, \nabla\varphi(\bar{x}_n) - K^*\bar{y}_n - \frac{\beta_1 - \alpha_1}{\gamma}(x_n - \bar{x}_{n-1}) \rangle + \frac{1}{2\gamma} \|x^* - \bar{x}_n\|^2 . \end{aligned}$$

Choosing $n = n_k - 1$ leads to

$$\begin{aligned} & g(x_{n_k}) + \langle x_{n_k} - \bar{x}_{n_k-1}, \nabla\varphi(\bar{x}_{n_k-1}) - K^*\bar{y}_{n_k-1} \rangle \\ & + \langle x_{n_k} - \bar{x}_{n_k-1}, -\frac{\beta_1 - \alpha_1}{\gamma}(x_{n_k-1} - \bar{x}_{n_k-2}) \rangle + \frac{1}{2\gamma} \|x_{n_k} - \bar{x}_{n_k-1}\|^2 \\ & \leq g(x^*) + \langle x^* - \bar{x}_{n_k-1}, \nabla\varphi(\bar{x}_{n_k-1}) - K^*\bar{y}_{n_k-1} \rangle \\ & + \langle x^* - \bar{x}_{n_k-1}, -\frac{\beta_1 - \alpha_1}{\gamma}(x_{n_k-1} - \bar{x}_{n_k-2}) \rangle + \frac{1}{2\gamma} \|x^* - \bar{x}_{n_k-1}\|^2 . \end{aligned}$$

By taking the limit superior above as $k \to +\infty$, we obtain

$$\begin{aligned}\limsup_{k\to+\infty} g(x_{n_k}) \leq & g(x^*) + \limsup_{k\to+\infty}\Big(\langle x^* - \bar{x}_{n_k-1}, -\frac{\beta_1-\alpha_1}{\gamma}(x_{n_k-1} - \bar{x}_{n_k-2})\rangle \\ & + \langle x^* - \bar{x}_{n_k-1}, \nabla\varphi(\bar{x}_{n_k-1}) - K^*\bar{y}_{n_k-1}\rangle + \frac{1}{2\gamma}\|x^* - \bar{x}_{n_k-1}\|^2\Big),\end{aligned} \tag{4.22}$$

where using the facts that the boundedness of the sequence $\{z_n\}_{n\in\mathbb{N}}$ and γ defined as (4.17), the continuity of $\nabla\varphi$ and $\lim_{k\to+\infty}\|x_{n_k} - \bar{x}_{n_k-1}\| = 0$. Hence (4.22) reduces to

$$\limsup_{k\to+\infty} g(x_{n_k}) \leq g(x^*). \tag{4.23}$$

Similarly, we have

$$\limsup_{k\to+\infty} h^*(y_{n_k}) \leq h^*(y^*). \tag{4.24}$$

On the other hand, by the lower semicontinuity of g and h^*, we get

$$g(x^*) \leq \liminf_{k\to+\infty} g(x_{n_k}), \quad h^*(y^*) \leq \liminf_{k\to+\infty} h^*(y_{n_k}). \tag{4.25}$$

In view of (4.23), (4.24) and (4.25), we can obtain that $g(x_{n_k}) \to g(x^*)$ and $h^*(y_{n_k}) \to h^*(y^*)$ as $k \to +\infty$. Hence,

$$\begin{aligned}\lim_{k\to+\infty} S(\hat{z}_{n_k}) &= \lim_{k\to+\infty}\{g(x_{n_k}) + h^*(y_{n_k}) + \varphi(x_{n_k}) - \langle y_{n_k}, Kx_{n_k}\rangle \\ &\quad +\delta_1\|x_{n_k} - \bar{x}_{n_k}\|^2 + \delta_2\|y_{n_k} - \bar{y}_{n_k}\|^2\Big\} \\ &= g(x^*) + h^*(y^*) + \varphi(x^*) - \langle y^*, Kx^*\rangle \\ &= S(\hat{z}^*).\end{aligned}$$

Then following from Lemma 4.4 (ii) and Lemma 4.5, we know that $A^n \in \partial S(\hat{z}_n)$ and $A^n \to 0$ as $n \to +\infty$. The closeness of ∂S implies that $0 \in \partial S(\hat{z}^*)$. Thus $\hat{z}^*$ is a critical point of S.
(iii) Take an arbitrary point $\hat{z}^* \in \Omega^*$, then there exists a subsequence $\{\hat{z}_{n_k}\}_{k\in\mathbb{N}}$ converging to $\hat{z}^*$ as $k \to +\infty$. Besides, the sequence $S(\hat{z}_{n_k})$ converges to $S(\hat{z}^*)$ and $\{S(\hat{z}_n)\}_{n\in\mathbb{N}}$ is nonincreasing. Thus we have $\lim_{n\to+\infty} S(\hat{z}_n) = S(\hat{z}^*)$ independent of $\hat{z}^*$. Hence S is constant on Ω^*. □

Remark 4.7. If one of the following statements is true,

(i) The objective function Φ is coercive.

(ii) The lower level sets of Φ are bounded.

then the sequence $\{\hat{z}_n\}_{n\in\mathbb{N}}$ generated by DiPGA is bounded.

Remark 4.8. Take the parameters γ, μ according to (4.17) and (4.18) respectively as shown in Remark 4.3. The sequence generated by DiPGA is denoted as $\{\hat{z}_n\}_{n\in\mathbb{N}}$, which is assumed to be bounded. Then similar to Lemma 4.6, one can deduce that every cluster point of $\{z_n\}_{n\in\mathbb{N}}$ is also a critical point of Φ.

We are now ready for proving the main result of the paper.

Theorem 4.9. *Suppose that Φ is a KL function and take the parameters $\gamma,\ \mu$ according to (4.17) and (4.18) respectively as shown in Remark 4.3. The sequence generated by DiPGA is denoted as $\{\hat{z}_n\}_{n\in\mathbb{N}}$, which is assumed to be bounded. Then*

$$\sum_{n=0}^{+\infty} \|z_{n+1} - z_n\| < +\infty,$$

and $\{z_n\}_{n\in\mathbb{N}}$ converges to a critical point of Φ.

Proof. As Lemma 4.6 (iii),

$$\lim_{n\to+\infty} S(\hat{z}_n) = S(\hat{z}^*).$$

We consider the following two cases.

(i) If there exists an integer n' for which $S(\hat{z}_{n'}) = S(\hat{z}^*)$. Rearranging terms of (4.19), for any $n > n'$, we have

$$\rho_1 \|z_{n+1} - \bar{z}_{n+1}\|^2 \le S(\hat{z}_n) - S(\hat{z}_{n+1}) \le S(\hat{z}_{n'}) - S(\hat{z}^*) = 0.$$

So, we have $z_{n+1} = \bar{z}_{n+1}$ for any $n > n'$. Associated with (4.12) and (4.13), it follows that $z_{n+1} = z_n$ for any $n > n'$ and the assertion holds.

(ii) Now we assume $S(\hat{z}_n) > S(\hat{z}^*)$ for all $n \ge 0$. Since $\lim_{n\to+\infty} S(\hat{z}_n) = S(\hat{z}^*)$, it follows that for any $\eta > 0$, there exists a nonnegative integer n_0 such that $S(\hat{z}_n) < S(\hat{z}^*) + \eta$ for all $n \ge n_0$. From (4.20), we know that $\lim_{n\to+\infty} \operatorname{dist}(\hat{z}_n, \Omega^*) = 0$. This means that for any $\varepsilon > 0$, there exists a positive integer n_1 such that $\operatorname{dist}(\hat{z}_n, \Omega^*) < \varepsilon$ for all $n \ge n_1$. Consequently, for all $n > l := \max\{n_0, n_1\}$,

$$\operatorname{dist}(\hat{z}_n, \Omega^*) < \varepsilon,\ \ S(\hat{z}^*) < S(\hat{z}_n) < S(\hat{z}^*) + \eta.$$

Since Ω^* is nonempty compact set and S is constant on Ω^*, applying Lemma 2.6, we deduce that for any $n > l$,

$$\vartheta'(S(\hat{z}_n) - S(\hat{z}^*))\operatorname{dist}(0, \partial S(\hat{z}_n)) \ge 1.$$

From Lemma 4.5, we get that

$$\vartheta'(S(\hat{z}_n) - S(\hat{z}^*)) \ge \frac{1}{\rho_2(\|\bar{z}_n - z_n\| + \|\bar{z}_{n-1} - z_{n-1}\|)} = \frac{1}{\rho_2(G_n + G_{n-1})}, \tag{4.26}$$

where $G_n = \|\bar{z}_n - z_n\|$. On the other hand, from the concavity of ϑ, we get that

$$\begin{aligned}\vartheta(S(\hat{z}_n) - S(\hat{z}^*)) - \vartheta(S(\hat{z}_{n+1}) - S(\hat{z}^*)) &\ge \vartheta'(S(\hat{z}_n) - S(\hat{z}^*))(S(\hat{z}_n) - S(\hat{z}_{n+1})) \\ &\ge \frac{\rho_1 G_n^2}{\rho_2(G_n + G_{n-1})}.\end{aligned} \tag{4.27}$$

For convenience, we define

$$\Delta_n = \vartheta(S(\hat{z}_n) - S(\hat{z}^*)) - \vartheta(S(\hat{z}_{n+1}) - S(\hat{z}^*)).$$

From (4.27), we obtain

$$G_n^2 \le \frac{2\rho_2}{\rho_1}\Delta_n \times \frac{1}{2}(G_n + G_{n-1}).$$

Using the fact that $2\sqrt{\xi\eta} \le \xi + \eta$ for all $\xi, \eta \ge 0$, we infer

$$2G_n \le 2\sqrt{\frac{2\rho_2}{\rho_1}\Delta_n \times \frac{1}{2}(G_n + G_{n-1})} \le \frac{2\rho_2}{\rho_1}\Delta_n + \frac{1}{2}(G_n + G_{n-1}). \tag{4.28}$$

So using inequality (4.28), we can obtain the fact that for all $j \geq 1$,

$$\sum_{i=1}^{j} G_i \leq \frac{1}{2}(G_0 - G_j) + \frac{2\rho_2}{\rho_1}[\vartheta(S(\hat{z}_1) - S(\hat{z}^*)) - \vartheta(S(\hat{z}_{j+1}) - S(\hat{z}^*))]. \tag{4.29}$$

Then from (4.29), we can easily get

$$\sum_{i=1}^{j} G_i \leq \frac{1}{2}G_0 + \frac{2\rho_2}{\rho_1}\vartheta(S(\hat{z}_1) - S(\hat{z}^*)) < +\infty.$$

Hence we have

$$\sum_{n=0}^{+\infty} \|\bar{z}_n - z_n\| < +\infty.$$

From (4.12) and (4.13), $\sum_{n=1}^{+\infty} \|x_{n+1} - x_n\| < +\infty$ and $\sum_{n=1}^{+\infty} \|y_{n+1} - y_n\| < +\infty$ hold. It is obvious that

$$\sum_{n=0}^{+\infty} \|z_{n+1} - z_n\| \ < +\infty, \tag{4.30}$$

and the sequence $\{z_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence and hence is convergent. Now the result follows immediately from Remark 4.8, that is, $\{z_n\}_{n\in\mathbb{N}}$ converges to a critical point of Φ. □

5 Application to Image Processing

Image processing problems [6, 33] can be conveniently described as follows. Given a matrix $L : \mathbb{H} \to \mathbb{H}$ denoting space-invariant blurring linear operator, a noise ν, and $b \in \mathbb{R}^m_+$ the vector of measurement. The goal is to reconstruct the original gray-scale picture of the size $m \times n$ pixels $x \in \mathbb{H} := \mathbb{R}^{mn}$ with entries in [0,1], where 0 represents pure black and 1 represents pure white, from the noisy measurement b. Often these problems can be represented as a minimize problem as follows,

$$\min_{x\in\mathbb{H}} \ \frac{\mu}{2}\|Lx - b\|^2 + J(Dx), \tag{5.1}$$

where $\mu > 0$ is a regularization parameter, $D : \mathbb{R}^{mn} \to \mathbb{R}^{2mn}$ is the discrete gradient operator and $J : \mathbb{H} \to \mathbb{G}$ is a regularization function. There are several choices of the function J proposed by [14, 15, 33], all of which are commonly to characterize the inherent properties of Dx, such as sparsity. In this paper, we choose Zhang penalty[33].

The Zhang penalty is defined by

$$\text{Zhang}_\alpha(z) = \sum_{j=1}^{2mn} g_\alpha(z_j),$$

where $\alpha > 0$ and

$$\begin{aligned} g_\alpha(z_j) &= \begin{cases} \frac{1}{\alpha}|z_j| & \text{if} \quad |z_j| < \alpha, \\ 1 & \text{if} \quad |z_j| \geq \alpha. \end{cases} \\ &= \frac{1}{\alpha}|z_j| - \begin{cases} 0 & \text{if} \quad |z_j| < \alpha, \\ \frac{1}{\alpha}(|z_j| - \alpha) & \text{if} \quad |z_j| \geq \alpha. \end{cases} \end{aligned}$$

Denoting the part after the curly brace as $h_\alpha(z_j)$ and $h_\alpha(z) := \sum_{j=1}^{2mn} h_\alpha(z_j)$.

Now the model (5.1) can be written as a DC optimization problem as follows:

$$\min \quad \frac{\mu}{2}\|Lx-b\|^2+\frac{1}{\alpha}\|Dx\|_1-h_\alpha(Dx). \tag{5.2}$$

It can be seen that (5.2) is consistent with (3.1) when $g(x)=\frac{1}{\alpha}\|Dx\|_1, \varphi(x)=\frac{\mu}{2}\|Lx-b\|^2$ and $h(Kx)=h_\alpha(Dx)$. Hence, DiPGA can be applied to the model (5.2).

5.1 The concrete solution method of subproblem

In order to apply DiPGA to the problem (5.2), we have to solve the following optimization subproblem

$$\inf\ \{\frac{1}{2\gamma}\|x-b\|^2+\|Dx\|_1 \mid x\in\mathbb{H}\}, \tag{5.3}$$

where $\gamma>0$, $b\in\mathbb{H}$ and $\|\cdot\|_1$ denotes (as usual) the sum of the absolute values. It can be seen that (5.3) is equivalent to the following problem,

$$\begin{aligned} \min \quad & \|y\|_1+\frac{1}{2\gamma}\|x-b\|^2 \\ \text{s.t.} \quad & Dx=y. \end{aligned} \tag{5.4}$$

We can solve (5.4) by alternating direction method of multipliers (ADMM). Its augmented Lagrange function is

$$L_\beta(x,y,\lambda)=\|y\|_1+\frac{1}{2\gamma}\|x-b\|^2-\lambda^T(Dx-y)+\frac{\beta}{2}\|Dx-y\|^2,$$

where λ is Lagrange multiplier and β is the penalty parameter. Apply ADMM to (5.4) and generate the iterates via

$$\begin{cases} x^{k+1}=\underset{x\in\mathbb{H}}{\operatorname{argmin}}\ \{\frac{1}{2\gamma}\|x-b\|^2-(\lambda^k)^T(Dx-y^k)+\frac{\beta}{2}\|Dx-y^k\|^2\}, \\ y^{k+1}=\underset{y\in\mathbb{H}}{\operatorname{argmin}}\ \{\|y\|_1-(\lambda^k)^T(Dx^{k+1}-y)+\frac{\beta}{2}\|Dx^{k+1}-y\|^2\}, \\ \lambda^{k+1}=\lambda^k-\beta(Dx^{k+1}-y^{k+1}). \end{cases}$$

5.2 Numerical results

Numerical experiment has been implemented in MATLAB on a PC with Inter Core i5 8265U and 8GB of RAM.

As follows, we test three images. Their names are Lean, Cameraman and Chart, and their sizes are 256×256, 512×512 and 1024×1024 respectively. We add the Gaussian white noise with mean 0 and standard deviation 0.01 for the three images. We solve the model (5.2) by double-proximal gradient algorithm (DPGA) and DiPGA. We choose initial value $x_0=b$, pick $v_0\in\partial h(Kx_0)$ and compute the parameters μ and γ by invoking Remark 4.3. The original clean images, the input noisy images and the recovered images by DPGA and DiPGA are shown in Figure 1.

Signal-to-noise ratio (SNR), which is typically considered as a measurement of denoising quality, defined as

$$\text{SNR}=20\log_{10}\frac{\|u^*\|_2}{\|u-u^*\|_2},$$

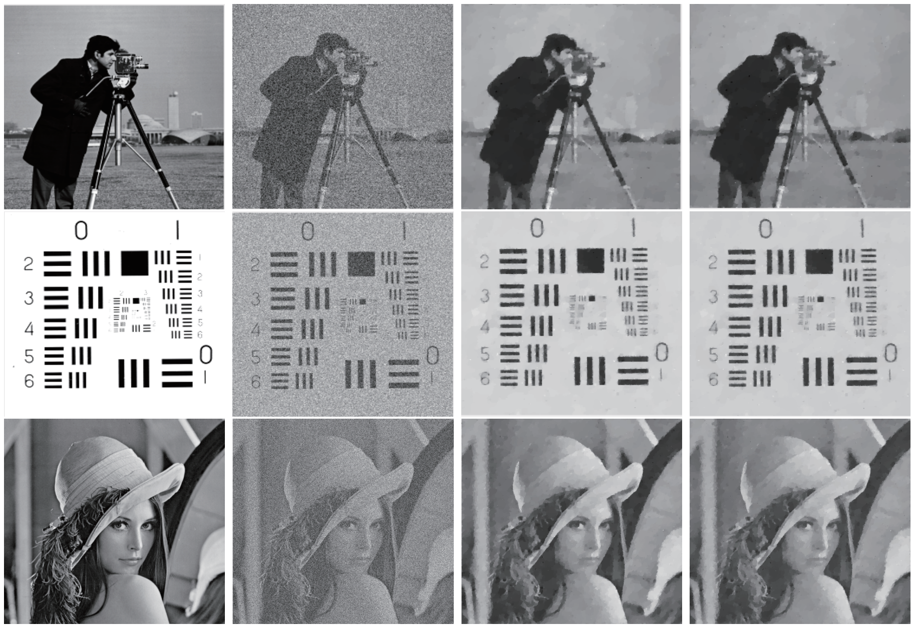

Figure 1: From left to right: original clean image, noisy image with Gaussian noise, image denoising using DPGA, image denoising using DiPGA. Test problem Cameraman (the first row), 512×512; test problem Chart (the second row), 1024×1024; test problem Lean (the third row), 256×256.

where u^* and u represent the original and the recovered images respectively. Table 1 reports the number of iteration, CPU time and the SNR values of DPGA and DiPGA with respect to the three test images. Obviously, both DPGA and DiPGA are effective for solving the above model. With respect to the number of iteration, CPU time and the value of SNR, DiPGA is the winner. Hence, we know that DiPGA is feasible significant.

Figure 2 shows the development of the SNR values when choosing different parameters in DiPGA. It shows the efficiency of DiPGA perform better when the values of parameters are larger. It is worth noting that DiPGA reduces to GiPALM when $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$. It is obvious that algorithm DiPGA has an advantage over algorithm GiPALM in terms of iteration.

Figure 3 shows the evolution of SNR of the test problem Cameraman and problem Lean. We run the test problem for 50 iterations and plot the evolution of SNR using DPGA and DiPGA.

6 Conclusions

In this paper, we proposed the improved optimization algorithm called the double-inertial proximal gradient algorithm (DiPGA). This algorithm combined the double-proximal gradient algorithm with the inertial scheme. It was applicable to solve the difference-of-convex programming. Suppose that the underlying function has the Kurdyka-Łojasiewicz property and the parameters satisfy certain conditions, the global convergence results of DiPGA can be established. In addition, we applied the proposed algorithm to image processing model

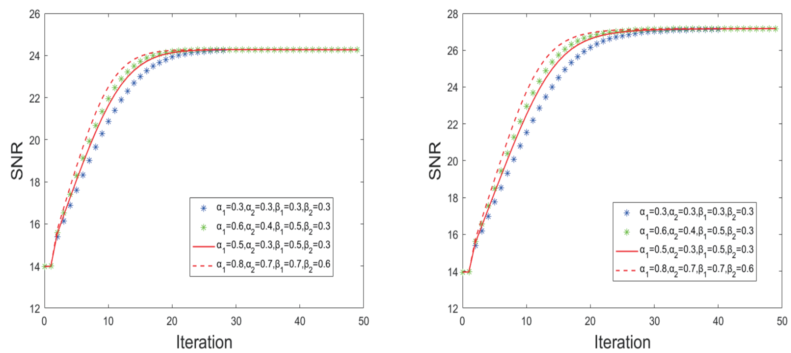

Figure 2: Development in signal-to-noise ration versus iterations. Left: problem Cameraman. Right: problem Lean.

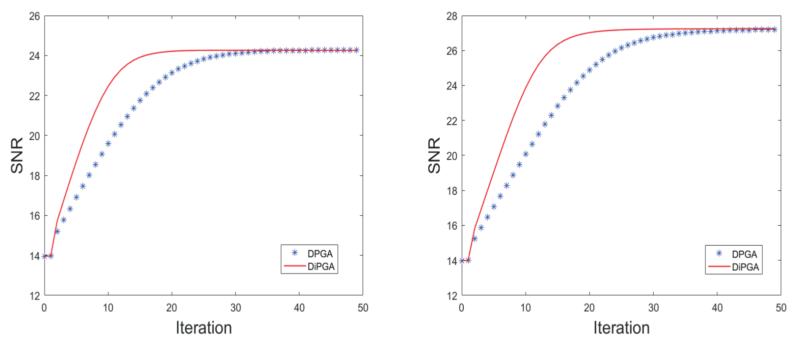

Figure 3: Plot of SNR versus iterations. Left: problem Cameraman. Right: problem Lean.

Table 1: Comparison between DPGA and DiPGA for test problems

Algs. \ Images	Cameraman			Chart			Lean		
	Iter.	CPU(s)	SNR	Iter.	CPU(s)	SNR	Iter.	CPU(s)	SNR
DPGA	28	3.6719	23.9586	41	6.4844	27.0971	38	4.3281	26.9921
DiPGA	17	2.9531	24.0006	25	5.9219	27.1044	22	3.7500	27.0052

to show its efficiency.

References

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

[2] N.T. An and N.M. Nam, Convergence analysis of a proximal point algorithm for minimizing differences of functions, *Optimization* 66 (2017) 129–147.

[3] F.J.A. Artacho, R.M.T. Fleming and P.T. Vuong, Accelerating the DC algorithm for smooth functions, *Math. Program.* 169 (2018) 95–118.

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

[5] H. Attouch, J. Bolte and B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program.* 137 (2013) 91–129.

[6] S. Banert and R.I. Boţ, A general double-proximal gradient algorithm for d.c. programming, *Math. Program.* 178 (2019) 301–326.

[7] H.H. Bauschke and P.L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2011.

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

[9] R.I. Boţ and E.R. Csetnek, An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems, *J. Optim. Theory Appl.* 171 (2016) 600–616.

[10] R.I. Boţ, E.R. Csetnek and C. Hendrich, Inertial Douglas-Rachford splitting for monotone inclusion problems, *Appl. Math. Comput.* 256 (2015) 472–487.

[11] G. Carlier, Remarks on Toland's duality, convexity constraint and optimal transport, *Pacific J. Optim.* 4 (2008) 423–432.

[12] C.H. Chen, S.Q. Ma and J.F. Yang, A general inertial proximal point algorithm for mixed variational inequality problem, *SIAM J. Optim.* 25 (2015) 2120–2142.

[13] X. Gao, X.J. Cai and D.R. Han, A Gauss-Seidel type inertial proximal alternating linearized minimization for a class of nonconvex optimization problems, *J. Global Optim.* 76 (2020) 863–887.

[14] G. Gasso, A. Rakotomamonjy and S. Canu, Recovering sparse signals with a certain family of nonconvex penalties and DC programming, *IEEE Trans. Signal Process.* 57 (2009) 4686–4698.

[15] Y. Lou, T. Zeng, S. Osher and J. Xin, A weighted difference of anisotropic and isotropic total variation model for image processing, *SIAM J. Imaging Sci.* 8 (2015) 1798–1823.

[16] P.E. Maingé and A. Moudafi, Convergence of new inertial proximal methods for DC programming, *SIAM J. Optim.* 19 (2008) 397–413.

[17] Y. Nesterov, *Lectures on convex optimization*, Springer Optimization and Its Applications, Springer Verlag, 2018.

[18] Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, *Dokl. Akad. Nauk* 269 (1983) 543–547.

[19] P. Ochs, T. Brox and T. Pock, iPiasco: Inertial proximal algorithm for strongly convex optimization, *J. Math. Imaging. Vision.* 53 (2015) 171–181.

[20] P. Ochs, Y.J. Chen, T. Brox and T. Pock, iPiano: Inertial proximal algorithm for nonconvex optimization, *SIAM J. Imaging Sci.* 7 (2014) 1388–1419.

[21] J.-S. Pang, M. Razaviyayn and A. Alvarado, Computing B-stationary points of nonsmooth DC programs, *Math. Oper. Res.* 42 (2017) 95–118.

[22] T. Pock and S. Sabach, Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems, *SIAM J. Imaging. Sci.* 9 (2016) 1756–1787.

[23] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, *Comput. Math. Math. Phys.* 4 (1964) 1–17.

[24] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[25] R.T. Rockafellar and R.J.B. Wets, *Variational Analysis*, Springer, 2009.

[26] M. Sanjabi, M. Razaviyayn and Z.-Q. Luo, Optimal joint base station assignment and beamforming for heterogeneous networks, *IEEE Trans. Signal Process.* 62 (2014) 1950–1961.

[27] W. Sun, R.J.B. Sampaio and M.A.B. Candido, Proximal point algorithm for minimization of DC function, *J. Comput. Math.* 21 (2003) 451–462.

[28] P.D. Tao and L.T.H. An, Convex analysis approach to D.C. programming: theory, algorithms and applications, *Acta Math. Vietnam.* 22 (1997) 289–355.

[29] J.F. Toland, Duality in nonconvex optimization, *J. Math. Anal. Appl.* 66 (1978) 399–415.

[30] J.F. Toland, A duality principle for non-convex optimisation and the calculus of variations, *Arch. Ration. Mech. Anal.* 71 (1979) 41–61.

[31] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM J. Control Optim.* 38 (2000) 431–446.

[32] H. Tuy, *Convex analysis and global optimization*, Springer Optimization and Its Applications, Springer Verlag, 2016.

[33] T. Zhang, Some sharp performance bounds for least squares regression with l_1 regularization, *Ann. Statist.* 37 (2009) 2109–2144.

Manuscript received 22 June 2021
revised 30 September 2021
accepted for publication 8 October 2021

Tanxing Wang
School of Mathematical Sciences
Jiangsu Key Lab for NSLSCS
Nanjing Normal University, Nanjing 210023, China
E-mail address: wangtanxing2021@163.com

Xingju Cai
School of Mathematical Sciences
Jiangsu Key Lab for NSLSCS
Nanjing Normal University, Nanjing 210023, China
E-mail address: caixingju@njnu.edu.cn

Yongzhong Song
School of Mathematical Sciences
Jiangsu Key Lab for NSLSCS
Nanjing Normal University, Nanjing 210023, China
E-mail address: yzsong@njnu.edu.cn

Xue Gao
Institute of Mathematics, Hebei University of Technology
Tianjin, 300401, China
E-mail address: xgao@hebut.edu.cn