Optimization Letters (2022) 16:1675–1700
https://doi.org/10.1007/s11590-021-01832-y

ORIGINAL PAPER

Unified linear convergence of first-order primal-dual algorithms for saddle point problems

Fan Jiang[1] · Zhongming Wu[2] · Xingju Cai[3] · Hongchao Zhang[4]

Received: 22 June 2021 / Accepted: 18 November 2021 / Published online: 13 January 2022

Abstract
In this paper, we study the linear convergence of several well-known first-order primal-dual methods for solving a class of convex-concave saddle point problems. We first unify the convergence analysis of these methods and prove the $O(1/N)$ convergence rates of the primal-dual gap generated by these methods in the ergodic sense, where N counts the number of iterations. Under a mild calmness condition, we further establish the global Q-linear convergence rate of the distances between the iterates generated by these methods and the solution set, and show the R-linear rate of the iterates in the nonergodic sense. Moreover, we demonstrate that the matrix games, fused lasso and constrained TV-ℓ_2 image restoration models as application examples satisfy this calmness condition. Numerical experiments on fused lasso demonstrate the linear rates for these methods.

Keywords First-order primal-dual algorithm · Saddle point problem · Convex optimization · Linear convergence rate

1 Introduction

In this paper, we study the linear convergence rates of several primal-dual algorithms for solving the following convex-concave problem:

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \ \mathcal{L}(x, y) := f(x) + h(x) + \langle Kx, y \rangle - g(y), \tag{1.1}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are two finite-dimensional real Euclidean spaces each equipped with the inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$, $K : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator with the operator norm $L = \|K\|$, $f : \mathcal{X} \to (-\infty, \infty]$ and

This research was partially supported by the National Natural Science Foundation of China under grants 11571178, 11871279, 11971238 and 12001286, and by the USA National Science Foundation under grants 1819161 and 2110722.

Extended author information available on the last page of the article

$g : \mathcal{Y} \to (-\infty, \infty]$ are proper closed convex functions, $h : \mathcal{X} \to \mathcal{R}$ is a smooth convex function with L_h-Lipschitz continuous gradient. It is well known that (1.1) is equivalent to the primal problem

$$\min_{x \in \mathcal{X}} f(x) + h(x) + g^*(Kx) \tag{1.2}$$

and the dual problem

$$\min_{y \in \mathcal{Y}} (f + h)^*(-K^* y) + g(y), \tag{1.3}$$

where g^* is the Fenchel conjugate (see the definition in Sect. 2) of the function g. (1.1) captures a wide spectrum of applications in statistics and machine learning, imaging and data processing [10,35,36,41].

We firstly discuss some algorithms for the special cases of (1.1) with only two functions. When f or h is missing, many splitting methods have been proposed and studied in the literature. One benchmark is the alternating direction method of multipliers [12,13], which can be obtained by applying the Douglas-Rachford splitting method (DRSM) to the dual problem (1.3) [11]. Another group is the primal-dual methods, such as primal-dual hybrid gradient method (PDHG, also known as Chambolle-Pock) [3,16], primal-dual fixed-point algorithm based on the proximal operator (PDFP^{2}O) [5], proximal alternating predictor-corrector (PAPC) [9,22]. Recently, [28] showed the equivalence of DRSM and PDHG. Many variants of PDHG have been well studied and the references are there in [1,2,17,23,30]. Besides, there are also many novel papers studying a more general saddle point problem, such as [24,25].

When comes to the case with three functions, a primal-dual algorithm, which known as Condat-Vu, was independently proposed by Condat [7] and Vu [38]:

$$\text{(I)} \quad \begin{cases} x^{k+1} = \text{prox}_{\tau f}\left(x^k - \tau K^* y^k - \tau \nabla h(x^k)\right), & \text{(1.4a)} \\ y^{k+1} = \text{prox}_{\sigma g}\left(y^k + \sigma K(2x^{k+1} - x^k)\right). & \text{(1.4b)} \end{cases}$$

By casting the scheme (1.4) in the form of forward-backward splitting, the convergence can be proved under the condition $\tau\sigma L^2 + \tau L_h/2 < 1$. Actually, this Condat-Vu scheme is the generation of Chambolle-Pock. As a generalization of PAPC and PDFP^{2}O, the primal-dual fixed-point algorithm (PDFP) was proposed in [6]:

$$\text{(II)} \quad \begin{cases} \bar{x}^{k+1} = \text{prox}_{\tau f}\left(x^k - \tau K^* y^k - \tau \nabla h(x^k)\right), & \text{(1.5a)} \\ y^{k+1} = \text{prox}_{\sigma g}\left(y^k + \sigma K \bar{x}^{k+1}\right), & \text{(1.5b)} \\ x^{k+1} = \text{prox}_{\tau f}\left(x^k - \tau K^* y^{k+1} - \tau \nabla h(x^k)\right). & \text{(1.5c)} \end{cases}$$

The assumptions that $\tau\sigma L^2 < 1$ and $\tau L_h < 2$, which guarantee the convergence of PDFP, are less restrictive than that for Condat-Vu. This advantage comes at the cost of calculating an additional proximal operator of f in each iteration. Another primal-dual algorithm named asymmetric forward-backward-adjoint (AFBA) splitting was proposed in [20]. It is a general operator splitting method and can be applied to solve the primal problem (1.2):

$$\text{(III)}\quad \begin{cases} \bar{x}^{k+1} = \mathrm{prox}_{\tau f}\left(x^k - \tau K^* y^k - \tau \nabla h(x^k)\right), & \text{(1.6a)} \\ y^{k+1} = \mathrm{prox}_{\sigma g}\left(y^k + \sigma K \bar{x}^{k+1}\right), & \text{(1.6b)} \\ x^{k+1} = \bar{x}^{k+1} - \tau K^*(y^{k+1} - y^k). & \text{(1.6c)} \end{cases}$$

The updates of $\bar{x}^{k+1}$ and y^{k+1} are the same in AFBA and PDFP, while (1.6c) is much simpler than (1.5c), which involves an additional proximal operator. However, the condition $\tau\sigma L^2/2 + \sqrt{\tau\sigma L^2}/2 + \tau L_h/2 < 1$, to guarantee the convergence of AFBA (1.6), is more conservative than Condat-Vu and PDFP. More recently, [39] proposed a primal-dual three-operator splitting (PD3O), which also involves two proximal operators in each iteration. It can be seen as a generalization of Chambolle-Pock and PAPC. The relations of these algorithms mentioned above have been well studied in [39].

The convergence rates of these algorithms are also considered under some additional conditions. For the Condat-Vu scheme (1.4), Chambolle and Pock [4] proved the primal-dual gap of the ergodic sequence converges with an $O(1/N)$ convergence rate, where N counts the iteration number. When partial and complete strong convexity is considered, $O(1/N^2)$ convergence rate and linear rate can be achieved. By introducing a combination parameter in (1.4b) and updating τ and σ in every iteration, the convergence rate for primal-dual gap of the ergodic sequence can be accelerated to $O(1/N^2)$ when f or g is strongly convex. When f and g are both strongly convex, linear convergence rate for primal-dual gap of the ergodic sequence can be yielded by replacing $2x^{k+1} - x^k$ by $(1+\theta)x^{k+1} - \theta x^k$ in (1.4b), where $\theta \in (0, 1)$. As to PDFP, Chen et al. [6] showed that the sequence generated by PDFP converges linearly to a saddle point of (1.1) when h and g are both strongly convex. Very recently, Jiang et al. [19] proposed an inexact Chambolle-Pock algorithm and established the global Q-linear convergence rate of the distance between the iterates and the solution set, and the R-linear convergence speed of the nonergodic iterates under a calmness condition. However, most existing linear convergence results are based on strong convexity, which can not be satisfied for many practical problems. Besides, existing results except [19] only establish the R-linear convergence rate [4,5], which is weaker than the Q-linear convergence rate that we will develop for Condat-Vu, PDFP and AFBA in this paper.

In this paper, we aim to establish the linear convergence rates for Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6) under an associated mild condition. More precisely, **(i)** we provide unified analysis to show the global Q-linear convergence rate for the distance between iterates generated by these three methods and the solution set, and the R-linear convergence for the nonergodic iterates under a calmness condition. We show that the fused LASSO and constrained TV-ℓ_2 problems actually satisfy the calmness condition, but either f or g or both functions in these models is not strongly convex; **(ii)** since Chambolle-Pock [3,4], PDFP^{2}O [5] and PAPC [9,22] are special cases of Condat-Vu (1.4) or PDFP (1.5), their linear convergence rates follows directly; **(iii)** the range of acceptable parameters in AFBA can be expanded to the same with PDFP.

The rest of this paper is organized as follows: in Sect. 2, we introduce some basic concepts and summarize some useful results for further analysis. We unify the convergence rates of Condat-Vu, PDFP and AFBA under a mild calmness condition in Sect. 3. In Sect. 4, we provide some practical examples in applications that satisfy the calmness condition. In Sect. 5, numerical experiments on fused lasso model show the superiority of the larger stepsizes for AFBA and the linear convergence of three methods. Finally, we draw some conclusions in Sect. 6.

2 Preliminary

In this section, we summarize some basic concepts that will be useful in the subsequent sections and recall the first-order optimality condition of problem (1.1).

2.1 Basic concepts and optimality conditions

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite-dimensional Euclidean spaces. For a function $f : \mathcal{X} \to (-\infty, \infty]$, the domain of f is defined by $\operatorname{dom} f := \{x \in \mathcal{X} \mid f(x) < \infty\}$. The epigraph of f is defined by $\operatorname{epi} f := \{(x, t) \mid f(x) \le t\}$. f is closed if $\operatorname{epi} f$ is a closed set and f is proper if $\operatorname{dom} f \neq \emptyset$. The conjugate function of f, denoted by f^*, is defined by

$$f^*(v) := \sup_{x \in \mathcal{X}} \{\langle v, x\rangle - f(x)\}.$$

For a proper convex function $f : \mathcal{X} \to (-\infty, \infty]$, the subdifferential of f at $x \in \operatorname{dom} f$ is defined by $\partial f(x) = \{d \mid f(z) \ge f(x) + \langle z - x, d\rangle,\ \forall z \in \mathcal{X}\}$. By convention, $\partial f(x) = \emptyset$ when $x \notin \operatorname{dom} f$. Recall that for a proper, closed and convex function $f : \mathcal{X} \to (-\infty, \infty]$, $y \in \mathcal{X}$ and $\tau > 0$, the proximal operator [29] of τf, denoted by $\operatorname{prox}_{\tau f}$, is given by

$$\operatorname{prox}_{\tau f}(y) = \arg\min_{x \in \mathcal{X}} \left\{ f(x) + \frac{1}{2\tau}\|x - y\|^2 \right\}.$$

If f is the indicator function δ_C of the closed convex set $C \subset \mathcal{X}$, then $\operatorname{prox}_f(\cdot) = \Pi_C(\cdot)$, the projection operator to C.

For the finite-dimensional real vector spaces $\mathcal{X}$ and $\mathcal{Y}$, the adjoint operator of the linear operator $K : \mathcal{X} \to \mathcal{Y}$ is denoted by K^*. If S is a self-adjoint and positive definite linear operator, we denote $\langle x, Sx\rangle$ by $\|x\|_S^2$. Let C be a nonempty closed convex set of $\mathcal{X}$, the notation of

$$\operatorname{dist}(x, C) := \min_{z \in C}\{\|x - z\|\}$$

denotes the Euclidean distance from any $x \in \mathcal{X}$ to the set C, while

$$\operatorname{dist}_G(x, C) := \min_{z \in C}\{\|x - z\|_G\}$$

denotes the distance in the terms norm $\|\cdot\|_G$ for a self-adjoint and positive definite linear operator G.

A sequence $\{u^k\}$ is said to converge to $\hat{u}$ Q-linearly in terms of G-norm, if there exist a scalar $\xi \in (0, 1)$ such that for sufficiently large k, it has

$$\|u^{k+1} - \hat{u}\|_G \leq \xi \|u^k - \hat{u}\|_G.$$

Moreover, a sequence $\{u^k\}$ is said to converge to $\hat{u}$ R-linearly in terms of G-norm, if there exists a nonnegative scalar sequence $\{w_k\}$ such that

$$\|u^k - \hat{u}\|_G \leq w_k,$$

where $\{w_k\}$ converges to zero Q-linearly.

Throughout the paper, we will assume that the following conditions are satisfied.

Assumption 1 The functions f and g are proper closed convex functions, h is a smooth convex function with L_h-Lipschitz continuous gradient. The solution set of (1.1) is nonempty.

Now, we recall an elementary identity which will be used later.

Lemma 1 *For any vectors a, b, c and d in the finite-dimensional real vector space $\mathcal{X}$, and the self-adjoint and positive definite linear operator $S : \mathcal{X} \to \mathcal{X}$, the following identity holds*

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

In the following, we present the first-order optimality condition of (1.1). Denote $u := (x, y)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, and let $\mathcal{U} := \mathcal{X} \times \mathcal{Y}$. The pair $(\hat{x}, \hat{y})$ defined on $\mathcal{X} \times \mathcal{Y}$ is called a saddle point of (1.1) if it satisfies the inequalities

$$\mathcal{L}(\hat{x}, y) \leq \mathcal{L}(\hat{x}, \hat{y}) \leq \mathcal{L}(x, \hat{y}), \quad \forall\, x \in \mathcal{X}, \ \forall\, y \in \mathcal{Y}.$$

Alternatively, we can rewrite these inequalities as

$$\begin{cases} P_{\hat{x},\hat{y}}(x) := f(x) - f(\hat{x}) + \langle x - \hat{x}, \nabla h(\hat{x}) + K^*\hat{y}\rangle \geq 0, \quad \forall\, x \in \mathcal{X}, \\ D_{\hat{x},\hat{y}}(y) := g(y) - g(\hat{y}) + \langle y - \hat{y}, -K\hat{x}\rangle \geq 0, \quad \forall\, y \in \mathcal{Y}. \end{cases} \tag{2.1}$$

Then we have $\Theta_{\hat{x},\hat{y}}(x, y) := P_{\hat{x},\hat{y}}(x) + D_{\hat{x},\hat{y}}(y) \geq 0$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. When there is no confusion, we will omit the subscript in P, D and Θ, i.e., $\Theta(x, y) := \Theta_{\hat{x},\hat{y}}(x, y)$, $P(x) := P_{\hat{x},\hat{y}}(x)$ and $D(y) := D_{\hat{x},\hat{y}}(y)$. Note that the system (2.1) can be reformulated as the following KKT system

$$\begin{cases} 0 \in \partial f(\hat{x}) + \nabla h(\hat{x}) + K^*\hat{y}, \\ 0 \in \partial g(\hat{y}) - K\hat{x}. \end{cases} \tag{2.2}$$

Denote the solution set to the KKT system (2.2) by $\widehat{\mathcal{U}}$. For analyzing the linear convergence, we introduce the KKT mapping $R : \mathcal{U} \to \mathcal{U}$ as follows

$$R(u) := \begin{pmatrix} x - \text{prox}_f \, (x - (\nabla h(x) + K^* y)) \\ y - \text{prox}_g \, (y + Kx) \end{pmatrix}, \quad \forall\, u = (x, y) \in \mathcal{U}. \tag{2.3}$$

Since the proximal operator of a proper convex function is 1-Lipschitz continuous, the mapping $R(\cdot)$ is continuous on $\mathcal{U}$. Obviously, we have

$$\widehat{\mathcal{U}} = \{u \in \mathcal{U} \,\big|\, R(u) = 0\}.$$

2.2 Locally upper Lipschitz continuity and calmness

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite-dimensional real vector spaces and $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a multivalued mapping. We denote the graph of F by $\text{Gph}\, F$ and the unit ball in $\mathcal{Y}$ by $B_{\mathcal{Y}}$. In the following, we first introduce the definition of locally upper Lipschitz continuity.

Definition 1 [31] The multivalued mapping $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ is locally upper Lipschitz continuous at $x^0 \in \mathcal{X}$ with modulus $\kappa_0 > 0$, if there exists a neighborhood V of x^0 such that

$$F(x) \subseteq F(x^0) + \kappa_0 \|x - x^0\| B_{\mathcal{Y}}, \quad \forall\, x \in V.$$

Note that this definition was first proposed by Robinson [31] to develop an implicit function for generalized variational inequalities. Now, we give the definition of piecewise polyhedral mapping.

Definition 2 [33] The multivalued mapping $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ is said to be piecewise polyhedral, if $\text{Gph}\, F$ is the union of finitely many polyhedral sets.

The following proposition, Robinson [32], established the locally upper Lipschitz continuity of a piecewise polyhedral multivalued mapping.

Proposition 1 [32] *If the multivalued mapping $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ is piecewise polyhedral, then F is locally upper Lipschitz continuous at any $x^0 \in \mathcal{X}$ with modulus κ_0 independent of the choice of x^0.*

The following gives the definition of convex piecewise linear quadratic function, whose subdiffenrential is a piecewise polyhedral multivalued mapping.

Definition 3 A closed proper convex function $f : \mathcal{X} \to (-\infty, \infty]$ is called piecewise linear-quadratic, if $\text{dom}\, f$ is the union of finitely many polyhedral sets and f is an affine or a quadratic function on each of these polyhedral sets.

We now summarize several useful results in the following proposition, whose proof can be found in [33].

Proposition 2 *Let $f : \mathcal{X} \to (-\infty, \infty]$ be a closed proper convex function. Then f is piecewise linear-quadratic if and only if the graph of ∂f is piecewise polyhedral. Moreover, f is piecewise linear-quadratic if and only if f^* is piecewise linear-quadratic. In addition, f is piecewise linear-quadratic function if and only if the proximal mapping of f is piecewise linear.*

Next, we present the definition of calmness for $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ at x^0 for y^0 with $(x^0, y^0) \in \operatorname{Gph} F$, which origins from [8].

Definition 4 Let $(x^0, y^0) \in \operatorname{Gph} F$. The multivalued mapping $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ is calm at x^0 for y^0 with modulus $\kappa_0 \geq 0$, if there exists a neighborhood V of x^0 and a neighborhood W of y^0 such that

$$F(x) \cap W \subseteq F(x^0) + \kappa_0 \|x - x^0\| B_{\mathcal{Y}}, \quad \forall\, x \in V.$$

If $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ is piecewise polyhedral, in particular F is the subdifferential of a convex piecewise linear-quadratic function, it follows from Proposition 1 that F is locally upper Lipschitz continuous at any $x_0 \in \mathcal{X}$ with modulus κ_0 independent of x_0. Consequently, according to Definitions 1 and 4, F is calm at x^0 for y^0 satisfying $(x^0, y^0) \in \operatorname{Gph} F$ with modulus $\kappa_0 > 0$ independent of the choice of (x^0, y^0).

3 Linear convergence of Condat-Vu, PDFP and AFBA

In this section, we first provide unified convergence analysis for Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6). Then, we establish the Q-linear convergence rate of the distance of the iterates to the solution set, which in return leads to the R-linear convergence rate for the iterates generated by these methods.

3.1 Global convergence

For analyzing the global convergence of Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6), we need some of the following assumptions for different algorithms:

(Ia) $\tau\sigma L^2 + \frac{1}{2}\tau L_h < 1$;

(Ib) $\tau\sigma L^2 + \tau L_h < 1$;

(IIa) $\tau\sigma L^2 < 1$ and $\tau L_h < 2$;

(IIb) $\tau\sigma L^2 < 1$ and $\tau L_h < 1$.

We now define the following self-adjoint linear operators:

$$
\begin{aligned}
M_{\mathrm{I}} &: \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y} : (x, y) \mapsto \left(\left(\frac{1}{\tau} - \frac{L_h}{2} \right) x - K^* y, -Kx + \frac{1}{\sigma} y \right), \\
H_{\mathrm{I}} &: \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y} : (x, y) \mapsto \left(\left(\frac{1}{\tau} - L_h \right) x - K^* y, -Kx + \frac{1}{\sigma} y \right), \\
G_{\mathrm{I}} &: \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y} : (x, y) \mapsto \left(\frac{1}{\tau} x - K^* y, -Kx + \frac{1}{\sigma} y \right), \\
M_{\mathrm{II}} &: \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y} : (x, y) \mapsto \left(\left(\frac{1}{\tau} - \frac{L_h}{2} \right) x, \left(\frac{1}{\sigma} I - \tau K K^* \right) y \right), \\
H_{\mathrm{II}} &: \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y} : (x, y) \mapsto \left(\left(\frac{1}{\tau} - L_h \right) x, \left(\frac{1}{\sigma} I - \tau K K^* \right) y \right), \\
G_{\mathrm{II}} &: \mathcal{X} \times \mathcal{Y} \to \mathcal{X} \times \mathcal{Y} : (x, y) \mapsto \left(\frac{1}{\tau} x, \frac{1}{\sigma} y \right).
\end{aligned}
\tag{3.1}
$$

For conceptual clarity, we let $M_{\mathrm{III}} := M_{\mathrm{II}}, G_{\mathrm{III}} := G_{\mathrm{II}}$ and $H_{\mathrm{III}} := H_{\mathrm{II}}$ and let assumptions (IIIa) and (IIIb) be the same as (IIa) and (IIb), respectively. Note that Assumption (ιa) is necessary for establishing convergence of iterates, whereas the stronger Assumption (ιb) is for establishing convergence rate of the gap functions for these methods with $\iota \in \{ \mathrm{I,II,III} \}$ (see Theorem 1).

For unifying the convergence analysis, we denote the sequence $\{(x^k, y^k)\}$ generated by Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6) by $\{(x_\iota^k, y_\iota^k)\}$ with $\iota \in \{\mathrm{I, II, III}\}$, respectively. We also denote $v := (\bar{x}, y)$. Note that $\bar{x}^{k+1}$ does not appear in Condat-Vu (1.4) and thus, we simply denote $\bar{x}^{k+1} := x^{k+1}$ for Condat-Vu. For every $\iota \in \{\mathrm{I, II, III}\}$, if ($\iota$a) holds, M_ι and G_ι are positive definite, which implies there exist two positive constants $\beta_\iota \geq \alpha_\iota > 0$ such that for any $u, u' \in \mathcal{U}$,

$$
\alpha_\iota \|u - u'\| \leq \|u - u'\|_{M_\iota} \leq \|u - u'\|_{G_\iota} \leq \beta_\iota \|u - u'\|. \tag{3.2}
$$

To establish global convergence, we start with the following lemma. For simplicity, we omit the subscript index in its proof without causing confusion.

Lemma 2 *Suppose that Assumption 1 holds. Let* $\{(x_\iota^k, y_\iota^k)\}$ *with* $\iota \in \{\mathrm{I, II, III}\}$ *be the sequence generated by Condat-Vu* (1.4)*, PDFP* (1.5) *and AFBA* (1.6)*, respectively. Then, for any* $(x, y) \in \mathcal{X} \times \mathcal{Y}$*, we have*

$$
\Theta_{x,y}(\bar{x}_\iota^{k+1}, y_\iota^{k+1}) \leq \frac{1}{2}\|u_\iota^k - u\|_{G_\iota}^2 - \frac{1}{2}\|u_\iota^{k+1} - u\|_{G_\iota}^2 - \frac{1}{2}\|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}^2, \tag{3.3}
$$

and

$$
\mathcal{L}(\bar{x}_\iota^{k+1}, y) - \mathcal{L}(x, y_\iota^{k+1}) \leq \frac{1}{2}\|u_\iota^k - u\|_{G_\iota}^2 - \frac{1}{2}\|u_\iota^{k+1} - u\|_{G_\iota}^2 - \frac{1}{2}\|u_\iota^k - v_\iota^{k+1}\|_{H_\iota}^2. \tag{3.4}
$$

Proof We consider the cases with $\iota \in \{\mathrm{I, II, III}\}$ respectively.

Case $\iota =$ I. First, it follows from (1.4a) that

$$f(x)-f(x^{k+1})+\langle x-x^{k+1}, K^*y^k+\nabla h(x^k)+\frac{1}{\tau}(x^{k+1}-x^k)\rangle \geq 0, \quad \forall x \in \mathcal{X}. \tag{3.5}$$

By rearranging some terms of (3.5), we have

$$\begin{aligned}\langle x - x^{k+1}, K^*(y^k - y) + \frac{1}{\tau}(x^{k+1} - x^k) + \nabla h(x^k) - \nabla h(x)\rangle \\ \geq f(x^{k+1}) - f(x) + \langle x^{k+1} - x, \nabla h(x) + K^*y\rangle.\end{aligned}$$

Then, using the definition of $P_{x,y}(\cdot)$, we obtain

$$\begin{aligned}\langle x^{k+1} - x, \frac{1}{\tau}(x^k - x^{k+1}) - K^*(y^k - y^{k+1})\rangle + \langle x^{k+1} - x, K^*(y - y^{k+1})\rangle \\ - \langle x^{k+1} - x, \nabla h(x^k) - \nabla h(x)\rangle \geq P_{x,y}(x^{k+1}).\end{aligned} \tag{3.6}$$

Similarly, according to (1.4b), we have

$$g(y)-g(y^{k+1})+\langle y-y^{k+1}, -Kx^{k+1}-K(x^{k+1}-x^k)+\frac{1}{\sigma}(y^{k+1}-y^k)\rangle \geq 0, \quad \forall y \in \mathcal{Y},$$

which by rearranging some terms gives

$$\begin{aligned}\langle y - y^{k+1}, -K(x^{k+1} - x) - K(x^{k+1} - x^k) + \frac{1}{\sigma}(y^{k+1} - y^k)\rangle \\ \geq g(y^{k+1}) - g(y) + \langle y^{k+1} - y, -Kx\rangle.\end{aligned}$$

Then, using the definition of $D_{x,y}(\cdot)$, we obtain

$$\langle y^{k+1}-y, -K(x^k-x^{k+1})+\frac{1}{\sigma}(y^k-y^{k+1})\rangle-\langle y^{k+1}-y, K(x-x^{k+1})\rangle \geq D_{x,y}(y^{k+1}). \tag{3.7}$$

Summing (3.6) and (3.7) and using the definition in (2.1), we can get

$$\begin{aligned}\langle x^{k+1} - x, \frac{1}{\tau}(x^k - x^{k+1}) - K^*(y^k - y^{k+1})\rangle - \langle x^{k+1} - x, \nabla h(x^k) - \nabla h(x)\rangle \\ +\langle y^{k+1} - y, -K(x^k - x^{k+1}) + \frac{1}{\sigma}(y^k - y^{k+1})\rangle \geq \Theta_{x,y}(x^{k+1}, y^{k+1}).\end{aligned} \tag{3.8}$$

By the fact that h is convex and possesses a Lipschitz gradient, we have

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

$$
\begin{aligned}
&= \frac{L_h}{4}\|x^{k+1}-x^k\|^2 - \left\| \frac{\sqrt{L_h}}{2}(x^{k+1}-x^k) + \frac{1}{\sqrt{L_h}}\Big(\nabla h(x^k) - \nabla h(x)\Big) \right\|^2 \\
&\le \frac{L_h}{4}\|x^{k+1}-x^k\|^2, \qquad (3.9)
\end{aligned}
$$

where the first inequality follows from convexity and Lipschitz continuity of ∇h (for example, see [27]). Now, on one hand, according to Lemma 1, we have

$$
\langle x^{k+1}-x, x^k-x^{k+1}\rangle = \frac{1}{2}(\|x^k-x\|^2 - \|x^{k+1}-x\|^2 - \|x^k-x^{k+1}\|^2), \qquad (3.10)
$$

and

$$
\langle y^{k+1}-y, y^k-y^{k+1}\rangle = \frac{1}{2}(\|y^k-y\|^2 - \|y^{k+1}-y\|^2 - \|y^k-y^{k+1}\|^2). \qquad (3.11)
$$

On the other hand, we have

$$
\begin{aligned}
&\langle x-x^{k+1}, K^*(y^k-y^{k+1})\rangle + \langle y-y^{k+1}, K(x^k-x^{k+1})\rangle \\
&\quad = \langle x-x^{k+1}, K^*(y-y^{k+1})\rangle + \langle x-x^{k+1}, K^*(y^k-y)\rangle \\
&\qquad + \langle y-y^{k+1}, K(x^k-x^{k+1})\rangle \\
&\quad = \langle x-x^{k+1}, K^*(y-y^{k+1})\rangle + \langle x-x^k, K^*(y^k-y)\rangle \\
&\qquad + \langle x^k-x^{k+1}, K^*(y^k-y)\rangle + \langle y-y^{k+1}, K(x^k-x^{k+1})\rangle \\
&\quad = \langle x-x^{k+1}, K^*(y-y^{k+1})\rangle + \langle x-x^k, K^*(y^k-y)\rangle \\
&\qquad + \langle x^k-x^{k+1}, K^*(y^k-y^{k+1})\rangle. \qquad (3.12)
\end{aligned}
$$

Substituting (3.10), (3.11) and (3.12) into (3.8), and combining the inequality (3.9), we get (3.3) with $\iota =$ I immediately. Adding $h(x^{k+1}) - h(x)$ to both sides of (3.5) and rearranging some terms, we obtain

$$
\begin{aligned}
&\langle x^{k+1}-x, \frac{1}{\tau}(x^k-x^{k+1}) - K^*(y^k-y^{k+1})\rangle + \langle x^{k+1}-x, K^*(y-y^{k+1})\rangle \\
&\quad - \langle x^{k+1}-x, \nabla h(x^k)\rangle + h(x^{k+1}) - h(x) \\
&\quad \ge f(x^{k+1}) + h(x^{k+1}) - f(x) - h(x) + \langle x^{k+1}-x, K^*y\rangle. \qquad (3.13)
\end{aligned}
$$

In addition, the following inequality holds

$$
\begin{aligned}
&\langle x-x^{k+1}, \nabla h(x^k)\rangle + h(x^{k+1}) - h(x) \\
&\quad \le \langle x-x^{k+1}, \nabla h(x^k)\rangle + h(x^{k+1}) - h(x^k) + \langle x^k-x, \nabla h(x^k)\rangle \\
&\quad = h(x^{k+1}) - h(x^k) - \langle x^{k+1}-x^k, \nabla h(x^k)\rangle \\
&\quad \le \frac{L_h}{2}\|x^k-x^{k+1}\|^2. \qquad (3.14)
\end{aligned}
$$

Combining (3.7), (3.13) and (3.14), the conclusion (3.4) with $\iota =$ I follows directly.

Case $\iota = \text{II}$. First, it follows from (1.5a) that

$$f(x)-f(\bar{x}^{k+1})+\langle x-\bar{x}^{k+1}, K^*y^k+\nabla h(x^k)+\frac{1}{\tau}(\bar{x}^{k+1}-x^k)\rangle \geq 0, \quad \forall x \in \mathcal{X}. \tag{3.15}$$

Similarly, it follows from (1.5c) that

$$f(x) - f(x^{k+1}) + \langle x - x^{k+1}, K^*y^{k+1} + \nabla h(x^k) + \frac{1}{\tau}(x^{k+1} - x^k)\rangle \geq 0, \quad \forall x \in \mathcal{X}. \tag{3.16}$$

Setting $x = x^{k+1}$ in (3.15) and adding the above two inequalities, we obtain

$$\begin{aligned}
&\langle x - x^{k+1}, K^*y^{k+1}\rangle - \langle x - \bar{x}^{k+1}, K^*y\rangle + \left\langle x^{k+1} - \bar{x}^{k+1}, K^*y^k\right\rangle \\
&\quad + \left\langle x - \bar{x}^{k+1}, \nabla h(x^k) - \nabla h(x)\right\rangle + \frac{1}{\tau}\left\langle x^{k+1} - \bar{x}^{k+1}, \bar{x}^{k+1} - x^k\right\rangle \\
&\quad + \frac{1}{\tau}\langle x - x^{k+1}, x^{k+1} - x^k\rangle \\
&\quad \geq f(\bar{x}^{k+1}) - f(x) + \left\langle \bar{x}^{k+1} - x, K^*y + \nabla h(x)\right\rangle.
\end{aligned}$$

Using the definition of $P_{x,y}(\cdot)$ and Lemma 1, we obtain

$$\begin{aligned}
&\langle x - x^{k+1}, K^*y^{k+1}\rangle - \langle x - \bar{x}^{k+1}, K^*y\rangle + \left\langle x^{k+1} - \bar{x}^{k+1}, K^*y^k\right\rangle \\
&\quad + \left\langle x - \bar{x}^{k+1}, \nabla h(x^k) - \nabla h(x)\right\rangle \\
&\quad + \frac{1}{2\tau}\left(\|x^k - x^{k+1}\|^2 - \|x^k - \bar{x}^{k+1}\|^2 - \|x^{k+1} - \bar{x}^{k+1}\|^2\right) \\
&\quad + \frac{1}{2\tau}\left(\|x - x^k\|^2 - \|x - x^{k+1}\|^2 - \|x^k - x^{k+1}\|^2\right) \geq P_{x,y}(\bar{x}^{k+1}).
\end{aligned} \tag{3.17}$$

According to (1.5b), we obtain

$$g(y) - g(y^{k+1}) + \langle y - y^{k+1}, -K\bar{x}^{k+1} + \frac{1}{\sigma}(y^{k+1} - y^k)\rangle \geq 0, \quad \forall y \in \mathcal{Y},$$

which by simple manipulation gives

$$\begin{aligned}
&\langle y - y^{k+1}, -K(\bar{x}^{k+1} - x) + \frac{1}{\sigma}(y^{k+1} - y^k)\rangle \\
&\qquad \geq g(y^{k+1}) - g(y) + \langle y^{k+1} - y, -Kx\rangle.
\end{aligned}$$

Using the definition of $D_{x,y}(\cdot)$ and Lemma 1, we have

$$\begin{aligned}
&\frac{1}{2\sigma}\left(\|y^k - y\|^2 - \|y^{k+1} - y\|^2 - \|y^k - y^{k+1}\|^2\right) \\
&\qquad + \langle y - y^{k+1}, -K(\bar{x}^{k+1} - x)\rangle \geq D_{x,y}(y^{k+1}).
\end{aligned} \tag{3.18}$$

Note that

$$\begin{aligned}
&\langle x - x^{k+1}, K^* y^{k+1}\rangle - \langle x - \bar{x}^{k+1}, K^* y\rangle \\
&\qquad + \left\langle x^{k+1} - \bar{x}^{k+1}, K^* y^k \right\rangle + \langle y - y^{k+1}, -K(\bar{x}^{k+1} - x)\rangle \\
&\quad = \langle x - x^{k+1}, K^* y^{k+1}\rangle - \left\langle y^{k+1}, K(x - \bar{x}^{k+1})\right\rangle + \left\langle x^{k+1} - \bar{x}^{k+1}, K^* y^k\right\rangle \\
&\quad = \langle \bar{x}^{k+1} - x^{k+1}, K^* y^{k+1}\rangle + \left\langle x^{k+1} - \bar{x}^{k+1}, K^* y^k\right\rangle \\
&\quad = \langle \bar{x}^{k+1} - x^{k+1}, K^*(y^{k+1} - y^k)\rangle.
\end{aligned}$$

Summing (3.17) and (3.18), it follows from the the above equality and (3.9) that

$$\begin{aligned}
&\frac{1}{2\tau}\|x^k - x\|^2 + \frac{1}{2\sigma}\|y^k - y\|^2 - \left(\frac{1}{2\tau}\|x^{k+1} - x\|^2 + \frac{1}{2\sigma}\|y^{k+1} - y\|^2\right) \\
&\quad \geq \left(\frac{1}{2\tau} - \frac{L_h}{4}\right)\|x^k - \bar{x}^{k+1}\|^2 + \langle \bar{x}^{k+1} - x^{k+1}, K^*(y^k - y^{k+1})\rangle \\
&\qquad + \frac{1}{2\tau}\|x^{k+1} - \bar{x}^{k+1}\|^2 + \frac{1}{2\sigma}\|y^k - y^{k+1}\|^2 + \Theta_{x,y}(\bar{x}^{k+1}, y^{k+1}) \\
&\quad \geq \left(\frac{1}{2\tau} - \frac{L_h}{4}\right)\|x^k - \bar{x}^{k+1}\|^2 + \frac{1}{2}\|y^k - y^{k+1}\|^2_{\frac{1}{\sigma}I - \tau K K^*} \\
&\qquad + \Theta_{x,y}(\bar{x}^{k+1}, y^{k+1}).
\end{aligned} \tag{3.19}$$

where the second inequality follows from the Cauchy-Schwarz inequality. Then, we get (3.3) with $\iota = \text{II}$ holds. Setting $x = x^{k+1}$ in (3.15) and summing it with (3.16), we have

$$\begin{aligned}
&\langle x - x^{k+1}, K^* y^{k+1}\rangle - \langle x - \bar{x}^{k+1}, K^* y\rangle + \left\langle x^{k+1} - \bar{x}^{k+1}, K^* y^k\right\rangle \\
&\qquad + \left\langle x - \bar{x}^{k+1}, \nabla h(x^k)\right\rangle + h(x^{k+1}) - h(x) \\
&\qquad + \frac{1}{\tau}\left\langle x^{k+1} - \bar{x}^{k+1}, \bar{x}^{k+1} - x^k\right\rangle + \frac{1}{\tau}\langle x - x^{k+1}, x^{k+1} - x^k\rangle \\
&\quad \geq f(\bar{x}^{k+1}) + h(x^{k+1}) - f(x) - h(x) + \left\langle \bar{x}^{k+1} - x, K^* y\right\rangle.
\end{aligned}$$

Adding this equality to (3.18) and combining it with (3.14), (3.4) with $\iota = \text{II}$ will follow.

Case $\iota = \text{III}$. First, it follows from (1.6a) that

$$f(x) - f(\bar{x}^{k+1}) + \langle x - \bar{x}^{k+1}, K^* y^k + \nabla h(x^k) + \frac{1}{\tau}(\bar{x}^{k+1} - x^k)\rangle \geq 0, \quad \forall x \in \mathcal{X}. \tag{3.20}$$

By rearranging some terms of (3.20), we obtain

$$\begin{aligned}
&\langle x-\bar{x}^{k+1}, K^*(y^k-y)\rangle + \left\langle x-\bar{x}^{k+1}, \nabla h(x^k)-\nabla h(x)\right\rangle \\
&\quad + \frac{1}{\tau}\left\langle x-\bar{x}^{k+1}, \bar{x}^{k+1}-x^k\right\rangle \\
&\quad\geq f(\bar{x}^{k+1})-f(x)+\left\langle \bar{x}^{k+1}-x, K^*y+\nabla h(x)\right\rangle.
\end{aligned}$$

Using the definition of $P(\cdot)$ and Lemma 1, we obtain

$$\begin{aligned}
&\langle x-\bar{x}^{k+1}, K^*(y^k-y)\rangle + \left\langle x-\bar{x}^{k+1}, \nabla h(x^k)-\nabla h(x)\right\rangle \\
&\quad + \frac{1}{2\tau}\left(\|x-x^k\|^2-\|x-\bar{x}^{k+1}\|^2-\|x^k-\bar{x}^{k+1}\|^2\right) \geq P_{x,y}(\bar{x}^{k+1}). \quad (3.21)
\end{aligned}$$

Since (1.6b) is the same with (1.5b), (3.18) still holds for this case. Summing (3.21) and (3.18), we get

$$\begin{aligned}
&\langle x-\bar{x}^{k+1}, K^*(y^k-y^{k+1})\rangle + \langle x-\bar{x}^{k+1}, \nabla h(x^k)-\nabla h(x)\rangle \\
&\quad + \frac{1}{2\tau}\left(\|x-x^k\|^2-\|x-\bar{x}^{k+1}\|^2-\|x^k-\bar{x}^{k+1}\|^2\right) \\
&\quad + \frac{1}{2\sigma}\left(\|y^k-y\|^2-\|y^{k+1}-y\|^2-\|y^k-y^{k+1}\|^2\right) \geq \Theta_{x,y}(\bar{x}^{k+1}, y^{k+1}).
\end{aligned} \quad (3.22)$$

Note that

$$\begin{aligned}
&\|\bar{x}^{k+1}-x\|^2 - 2\tau\langle x-\bar{x}^{k+1}, K^*(y^k-y^{k+1})\rangle \\
&\quad = \|\bar{x}^{k+1}-x-\tau K^*(y^{k+1}-y^k)\|^2 - \tau^2\|K^*(y^k-y^{k+1})\|^2 \\
&\quad = \|x^{k+1}-x\|^2 - \tau^2\|K^*(y^k-y^{k+1})\|^2,
\end{aligned}$$

where the last equality follows from (1.6c). Substituting the above equality and (3.9) into (3.22), (3.3) with $\iota = \text{III}$ will follow. Likewise, instead of using (3.9), we adopt the inequality (3.14), we will get (3.4) with $\iota = \text{III}$. The proof is completed. □

The following theorem shows the ergodic convergence of Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6).

Theorem 1 *Suppose that Assumptions 1 and (ιa) hold for $\iota \in \{\text{I}, \text{II}, \text{III}\}$. Let $\{(x_\iota^k, y_\iota^k)\}$ with $\iota \in \{\text{I}, \text{II}, \text{III}\}$ be the sequence generated by Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6), respectively. Then $\{(x_\iota^k, y_\iota^k)\}$ converges to a saddle point $(x_\iota^\infty, y_\iota^\infty)$ of (1.1). If, in addition, Assumption (ιb) holds, for the ergodic sequence $\{(X_\iota^N, Y_\iota^N)\}$ defined by*

$$X_\iota^N = \frac{1}{N}\sum_{k=1}^{N}\bar{x}_\iota^k \quad \textit{and} \quad Y_\iota^N = \frac{1}{N}\sum_{k=1}^{N} y_\iota^k, \quad (3.23)$$

we have

$$\mathcal{L}(X_\iota^N, y) - \mathcal{L}(x, Y_\iota^N) \le \frac{1}{2N}\|u_\iota^0 - u\|_{G_\iota}^2, \quad \forall\,(x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{3.24}$$

Proof Suppose that (ιa) holds. Let $(\hat{x}, \hat{y})$ be an arbitrary saddle point of (1.1). Summing the inequality (3.3) over $k = 0, 1, \ldots, N-1$ and setting $(x, y) = (\hat{x}, \hat{y})$ yields that for any $N \ge 0$,

$$\begin{aligned}&\frac{1}{2}\|u_\iota^N - \hat{u}\|_{G_\iota}^2 + \frac{1}{2}\sum_{k=0}^{N-1}\|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}^2 + \sum_{k=0}^{N-1}\Theta(\bar{x}_\iota^{k+1}, y_\iota^{k+1})\\ &\quad\le \frac{1}{2}\|u_\iota^0 - \hat{u}\|_{G_\iota}^2.\end{aligned} \tag{3.25}$$

Combining (3.25), $\Theta(\bar{x}_\iota^{k+1}, y_\iota^{k+1}) \ge 0$ and the positive definiteness of M_ι and H_ι, we conclude that $\{(x_\iota^k, y_\iota^k)\}$ is bounded and

$$\sum_{k=0}^{\infty}\|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}^2 < \infty.$$

Consequently, $(x_\iota^k, y_\iota^k) - (\bar{x}_\iota^{k+1}, y_\iota^{k+1}) \to 0$ as $k \to \infty$. Since $\{(x_\iota^k, y_\iota^k)\}$ has at least one limit point $(x_\iota^\infty, y_\iota^\infty)$ by its boundedness, there exists a subsequence $\{(x_\iota^{k_j}, y_\iota^{k_j})\}$ such that $(x_\iota^{k_j}, y_\iota^{k_j}) \to (x_\iota^\infty, y_\iota^\infty)$ as $j \to \infty$. For any $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$, we have

$$\begin{aligned}&f(x) - f(\bar{x}_\iota^{k_j+1}) + \langle x - \bar{x}_\iota^{k_j+1}, \nabla h(x_\iota^{k_j}) + K^* y_\iota^{k_j}\rangle\\ &\quad+\langle x - \bar{x}_\iota^{k_j+1}, \frac{1}{\tau}(\bar{x}_\iota^{k_j+1} - x_\iota^{k_j})\rangle \ge 0, \quad \forall\, x \in \mathcal{X}.\end{aligned}$$

For $\iota = \mathrm{I}$, it follows from (1.4b) that

$$\begin{aligned}&g(y) - g(y_\iota^{k_j+1}) + \langle y - y_\iota^{k_j+1}, -K(2\bar{x}_\iota^{k_j+1} - x_\iota^{k_j})\rangle\\ &\quad+\langle y - y_\iota^{k_j+1}, \frac{1}{\sigma}(y_\iota^{k_j+1} - y_\iota^{k_j})\rangle \ge 0, \quad \forall\, y \in \mathcal{Y}.\end{aligned}$$

For $\iota \in \{\mathrm{II}, \mathrm{III}\}$, according to (1.5b) and (1.6b), we have

$$g(y) - g(y_\iota^{k_j+1}) + \left\langle y - y_\iota^{k_j+1}, -K\bar{x}_\iota^{k_j+1} + \frac{1}{\sigma}(y_\iota^{k_j+1} - y_\iota^{k_j})\right\rangle \ge 0, \quad \forall\, y \in \mathcal{Y}.$$

Passing $j \to \infty$ in the above three inequalities, it follows that $(x_\iota^\infty, y_\iota^\infty)$ is a solution of (1.1). Note that (3.3) holds for any solution of (1.1), hence

$$\|u_\iota^{k+1} - u_\iota^\infty\|_{G_\iota}^2 \le \|u_\iota^k - u_\iota^\infty\|_{G_\iota}^2, \quad \forall\, k \ge 0. \tag{3.26}$$

Since $(x_\iota^{k_j}, y_\iota^{k_j}) \to (x_\iota^\infty, y_\iota^\infty)$ and $(x_\iota^k, y_\iota^k) - (\bar{x}_\iota^{k+1}, y_\iota^{k+1}) \to 0$, for any given $\epsilon > 0$ and $\iota \in \{\text{I, II, III}\}$, there exists a positive integer l such that

$$\|v_\iota^{k_l+1} - u_\iota^\infty\|_{G_\iota}^2 < \frac{\epsilon}{2}, \quad \text{and} \quad \|v_\iota^{k_l+1} - u_\iota^{k_l}\|_{G_\iota}^2 < \frac{\epsilon}{2}. \tag{3.27}$$

Therefore, for any $k \geq k_l$, it follows from (3.26) and (3.27) that

$$\begin{aligned}\|u_\iota^k - u_\iota^\infty\|_{G_\iota}^2 &\leq \|u_\iota^{k_l} - u_\iota^\infty\|_{G_\iota}^2 \\ &\leq \|v_\iota^{k_l+1} - u_\iota^\infty\|_{G_\iota}^2 + \|u^{k_l} - v_\iota^{k_l+1}\|_{G_\iota}^2 < \epsilon.\end{aligned}$$

This shows that the sequence $\{(x_\iota^k, y_\iota^k)\}$ converges to $(x_\iota^\infty, y_\iota^\infty)$.

Now, suppose that (ιb) holds, then the operator H_ι defined in (3.1) are positive definite. So, summing the inequality (3.4) over $k = 0, 1, \ldots, N-1$, we obtain

$$\sum_{k=0}^{N-1} \left(\mathcal{L}(\bar{x}_\iota^{k+1}, y) - \mathcal{L}(x, y_\iota^{k+1})\right) \leq \frac{1}{2}(\|u_\iota^0 - u\|_{G_\iota}^2 - \|u_\iota^N - u\|_{G_\iota}^2).$$

It follows from the convexity of $f,\ g,\ h$ and the definition of (X_ι^N, Y_ι^N) in (3.23) that

$$N(\mathcal{L}(X_\iota^N, y) - \mathcal{L}(x, Y_\iota^N)) \leq \frac{1}{2}\|u_\iota^0 - u\|_{G_\iota}^2.$$

Then, the assertion (3.24) follows directly. □

Note that the condition (ιa) with $\iota \in \{\text{I, II, III}\}$ is enough to guarantee the convergence of Condat-Vu, PDFP, and AFBA, respectively. However, if one would like to establish the convergence rate for primal-dual gap, the stronger condition (ιb) needs to be assumed. For Condat-Vu, the conclusions of Lemma 2 and Theorem 1 had been extablished in [4,38]. As for PDFP, [6] proved that the sequences $\{x^k\}$ and $\{\bar{x}^k\}$ converge to a solution of the primal problem (1.2) by fixed point theory; however, no convergence rate of the primal-dual gap is analyzed. Actually, by Theorem 1, we see that the primal-dual gap of the ergodic sequence converges to zero with an $O(1/N)$ rate. As for AFBA, [20] proved that the sequence $\{(x^k, y^k)\}$ converge to a saddle point of the min-max problem (1.1) under the condition $\tau\sigma L^2 + \sqrt{\tau\sigma L^2} + \tau L_h < 2$. This condition is much more conservative than the conditions posed in the Theorem 1, where the $O(1/N)$ convergence rate of the primal-dual gap of the ergodic sequence of AFBA is also established. Moreover, although the update of x^{k+1} in AFBA (1.6c) differs from that one in PDFP (1.5c), we can see from our analysis in Lemma 2 and Theorem 1 that they actually lead to the same contractive property.

3.2 Linear convergence

In this subsection, we study the Q-linear convergence and R-linear convergence properties of Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6).

Firstly, we derive an useful upper bound of $\|R(\cdot)\|$ at the iterates $\{(\bar{x}_\iota^{k+1}, y_\iota^{k+1})\}$ generated by Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6).

Lemma 3 *Suppose that Assumptions 1 and (ιa) hold for $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$. Let $\{(x_\iota^k, y_\iota^k)\}$ with $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$ be the sequence generated by Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6), respectively. Then for any $k \geq 0$, there exists a constant $\kappa_\iota > 0$ such that*

$$\|R(v_\iota^{k+1})\|^2 \leq \kappa_\iota \|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}^2, \tag{3.28}$$

where

$$\kappa_\iota = \frac{1}{\alpha_\iota^2} \max\left\{3L_h^2 + \frac{3}{\tau^2} + 2L^2, 3L^2 + \frac{2}{\sigma^2}\right\} \tag{3.29}$$

and $\alpha_\iota > 0$ is a constant given in (3.2).

Proof Since we have denoted $\bar{x}_{\mathrm{I}}^{k+1} = x_{\mathrm{I}}^{k+1}$ for Condat-Vu, by (1.4a), (1.5a) and (1.6a), for $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$ it has

$$0 \in \partial f(\bar{x}_\iota^{k+1}) + \nabla h(x_\iota^k) + \frac{1}{\tau}(\bar{x}_\iota^{k+1} - x_\iota^k) + K^* y_\iota^k,$$

which implies that

$$\bar{x}_\iota^{k+1} = \operatorname{prox}_f\left(\bar{x}_\iota^{k+1} - (\nabla h(x_\iota^k) + \frac{1}{\tau}(\bar{x}_\iota^{k+1} - x_\iota^k) + K^* y_\iota^k)\right).$$

Therefore,

$$\begin{aligned}
&\|\bar{x}_\iota^{k+1} - \operatorname{prox}_f\left(\bar{x}_\iota^{k+1} - (\nabla h(\bar{x}_\iota^{k+1}) + K^* y_\iota^{k+1})\right)\|^2 \\
&\leq \|\nabla h(\bar{x}_\iota^{k+1}) - \nabla h(x_\iota^k) - \frac{1}{\tau}(\bar{x}_\iota^{k+1} - x_\iota^k) + K^*(y_\iota^{k+1} - y_\iota^k)\|^2 \\
&\leq 3L_h^2 \|x_\iota^k - \bar{x}_\iota^{k+1}\|^2 + \frac{3}{\tau^2}\|x_\iota^k - \bar{x}_\iota^{k+1}\|^2 + 3L^2\|y_\iota^k - y_\iota^{k+1}\|^2.
\end{aligned} \tag{3.30}$$

Then, it follows from (1.4b) that for $\iota = \mathrm{I}$,

$$0 \in \partial g(y_\iota^{k+1}) - K(2\bar{x}_\iota^{k+1} - x_\iota^k) + \frac{1}{\sigma}(y_\iota^{k+1} - y_\iota^k),$$

which implies that

$$y_\iota^{k+1} = \operatorname{prox}_g\left(y_\iota^{k+1} - (-K(2\bar{x}_\iota^{k+1} - x_\iota^k) + \frac{1}{\sigma}(y_\iota^{k+1} - y_\iota^k))\right).$$

Therefore, we have

$$
\begin{aligned}
&\|y_\iota^{k+1} - \mathrm{prox}_g\left(y_\iota^{k+1} + K\bar{x}_\iota^{k+1}\right)\|^2 \\
&\le \|K(\bar{x}_\iota^{k+1} - x_\iota^k) + \frac{1}{\sigma}(y_\iota^{k+1} - y_\iota^k)\|^2 \\
&\le 2L^2\|x_\iota^k - \bar{x}_\iota^{k+1}\|^2 + \frac{2}{\sigma^2}\|y_\iota^k - y_\iota^{k+1}\|^2, \qquad (3.31)
\end{aligned}
$$

where the first inequality follows from the 1-Lipschitz continuity of $\mathrm{prox}_g(\cdot)$, and the second inequality follows from the Cauchy-Schwarz inequality. Similarly, according to (1.5b) and (1.6b), we have for $\iota \in \{\mathrm{II}, \mathrm{III}\}$,

$$
y_\iota^{k+1} = \mathrm{prox}_g\left(y_\iota^{k+1} - (-K\bar{x}_\iota^{k+1} + \frac{1}{\sigma}(y_\iota^{k+1} - y_\iota^k))\right),
$$

which gives

$$
\|y_\iota^{k+1} - \mathrm{prox}_g\left(y_\iota^{k+1} + K\bar{x}_\iota^{k+1}\right)\|^2 \le \frac{1}{\sigma^2}\|y_\iota^k - y_\iota^{k+1}\|^2. \qquad (3.32)
$$

Consequently, it follows (3.31) and (3.32) that for any $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$,

$$
\|y_\iota^{k+1} - \mathrm{prox}_g\left(y_\iota^{k+1} + K\bar{x}_\iota^{k+1}\right)\|^2 \le 2L^2\|x_\iota^k - \bar{x}_\iota^{k+1}\|^2 + \frac{2}{\sigma^2}\|y_\iota^k - y_\iota^{k+1}\|^2. \qquad (3.33)
$$

Then, by combining (3.30) and (3.33), we obtain from the definition of $R(\cdot)$ in (2.3) that

$$
\begin{aligned}
\|R(v_\iota^{k+1})\|^2 &= \|\bar{x}_\iota^{k+1} - \mathrm{prox}_f\left(\bar{x}_\iota^{k+1} - (\nabla h(\bar{x}_\iota^{k+1}) + K^* y_\iota^{k+1})\right)\|^2 \\
&\quad + \|y_\iota^{k+1} - \mathrm{prox}_g\left(y_\iota^{k+1} + K\bar{x}_\iota^{k+1}\right)\|^2 \\
&\le (3L_h^2 + \frac{3}{\tau^2} + 2L^2)\|x_\iota^k - \bar{x}_\iota^{k+1}\|^2 + (3L^2 + \frac{2}{\sigma^2})\|y_\iota^k - y_\iota^{k+1}\|^2 \\
&\le \kappa_\iota \|u_\iota^k - v_\iota^{k+1}\|^2_{M_\iota},
\end{aligned}
$$

where the last inequality follows from (3.2) and the definition of κ_ι in (3.29). □

Now, we are ready to establish the linear convergence rate of Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6) under a calmness condition. If $R^{-1} : \mathcal{U} \to \mathcal{U}$ is calm at the origin for a point u^∞ with modulus $\theta > 0$, it follows from Definition 4 that there exist a neighborhood $B(0, s)$ of the origin and a neighborhood $B(u^\infty, r)$ of u^∞ with $r, s > 0$ such that

$$
R^{-1}(x) \cap B(u^\infty, r) \subseteq R^{-1}(0) + \theta\|x\| B_{\mathcal{U}}, \quad \forall x \in B(0, s), \qquad (3.34)
$$

where $B_{\mathcal{U}}$ is the unit ball in $\mathcal{U}$. By the Lipschitz continuity of KKT mapping R, we can choose $r > 0$ sufficiently small in (3.34) such that $R(u) \in B(0, s)$ for any

$u \in B(u^\infty, r)$. Then, noticing that $R^{-1}(0) = \widehat{\mathcal{U}}$ and considering $x = R(u) \in B(0, s)$, the relation (3.34) would imply

$$\operatorname{dist}(u, \widehat{\mathcal{U}}) \leq \theta \|R(u)\|, \quad u \in \{u \in \mathcal{U} \,\big|\, \|u - u^\infty\| \leq r\}.$$

Note that this calmness condition is also proposed for establishing linear convergence of the alternating direction method of multipliers [15] and the first-order inexact primal-dual algorithm [19].

Theorem 2 *Suppose that Assumptions 1 and (ιa) hold for $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$. Let $\{(x_\iota^k, y_\iota^k)\}$ with $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$ be the sequence generated by Condat-Vu (1.4), PDFP (1.5) and AFBA (1.6), respectively. Then we have the following properties.*
(1) *There exists a saddle point $u_\iota^\infty = (x_\iota^\infty, y_\iota^\infty)$ of (1.1) such that the $\{(x_\iota^k, y_\iota^k)\}$ converges to u_ι^∞.*
(2) *If R^{-1} is calm at the origin for u_ι^∞ with modulus $\theta_\iota > 0$, i.e.,*

$$dist(u, \widehat{\mathcal{U}}) \leq \theta_\iota \|R(u)\|, \quad \forall\, u \in \{u \in \mathcal{U} \,\big|\, \|u - u_\iota^\infty\| \leq r_\iota\}, \tag{3.35}$$

for some $r_\iota > 0$, there exists a positive number $0 < \xi_\iota < 1$ given by

$$\xi_\iota = \sqrt{1 - \frac{\alpha_\iota^2}{\beta_\iota^2(\theta_\iota \beta_\iota \sqrt{\kappa_\iota} + 1)^2}}$$

such that

$$dist_{G_\iota}(u_\iota^{k+1}, \widehat{\mathcal{U}}) \leq \xi_\iota dist_{G_\iota}(u_\iota^k, \widehat{\mathcal{U}}), \tag{3.36}$$

for all $k \geq 0$, where α_ι, β_ι and κ_ι are given in (3.2) and (3.29), respectively. Moreover, $\{u_\iota^k\} := \{(x_\iota^k, y_\iota^k)\}$ converges to u_ι^∞ R-linearly.

Proof Property (1) has already been established in Theorem 1. Hence, there exists $\bar{k} \geq 0$ such that for all $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$ it has

$$\|v_\iota^{k+1} - u_\iota^\infty\| \leq r_\iota, \quad \forall\, k \geq \bar{k}.$$

Thus, by Lemma 3 and (3.35), we know that for all $k \geq \bar{k}$,

$$\operatorname{dist}(v_\iota^{k+1}, \widehat{\mathcal{U}}) \leq \theta_\iota \|R(v_\iota^{k+1})\| \leq \theta_\iota \sqrt{\kappa}_\iota \|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}, \tag{3.37}$$

where κ_ι is given in Lemma 3. Next, we have

$$\operatorname{dist}(v_\iota^{k+1}, \widehat{\mathcal{U}}) \geq \frac{1}{\beta_\iota} \operatorname{dist}_{M_\iota}(v_\iota^{k+1}, \widehat{\mathcal{U}}) \geq \frac{1}{\beta_\iota} \left(\operatorname{dist}_{M_\iota}(u_\iota^k, \widehat{\mathcal{U}}) - \|u_\iota^k - v_\iota^{k+1}\|_{M_\iota} \right). \tag{3.38}$$

By combining (3.37) with (3.38), we obtain for $k \geq \bar{k}$

$$\operatorname{dist}_{G_\iota}(u_\iota^k, \widehat{\mathcal{U}}) \leq \frac{\beta_\iota}{\alpha_\iota} \operatorname{dist}_{M_\iota}(u_\iota^k, \widehat{\mathcal{U}}) \leq \frac{\beta_\iota(\theta_\iota \beta_\iota \sqrt{\kappa_\iota} + 1)}{\alpha_\iota} \|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}. \tag{3.39}$$

Note that for any $(\hat{x}, \hat{y}) \in \hat{\mathcal{U}}$, it follows from (3.3) and $\Theta(\cdot, \cdot) \geq 0$ that

$$\|u_\iota^k - \hat{u}\|_{G_\iota}^2 - \|u_\iota^{k+1} - \hat{u}\|_{G_\iota}^2 \geq \|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}^2. \tag{3.40}$$

Since $\widehat{\mathcal{U}}$ is a nonempty closed convex set, the above inequality (3.40) implies for $k \geq \bar{k}$,

$$\begin{aligned} \mathrm{dist}_{G_\iota}^2(u_\iota^k, \widehat{\mathcal{U}}) - \mathrm{dist}_{G_\iota}^2(u_\iota^{k+1}, \widehat{\mathcal{U}}) &\geq \|u_\iota^k - v_\iota^{k+1}\|_{M_\iota}^2 \\ &\geq \frac{\alpha_\iota^2}{\beta_\iota^2(\theta_\iota \beta_\iota \sqrt{\kappa_\iota} + 1)^2} \mathrm{dist}_{G_\iota}^2(u_\iota^k, \widehat{\mathcal{U}}). \end{aligned} \tag{3.41}$$

where the second inequality follows from (3.39). Then, (3.40), (3.41), and Lemma 2 imply that the property (2) holds, i.e., (3.36) holds for all $k \geq 0$.

Now, we show that $\{u_\iota^k\}$ converges to u_ι^∞ R-linearly. Let $\hat{u}_\iota^k \in \widehat{\mathcal{U}}$ such that $\mathrm{dist}_{G_\iota}(u_\iota^k, \widehat{\mathcal{U}}) = \|u_\iota^k - \hat{u}_\iota^k\|_{G_\iota}$. By (3.3), we have

$$\begin{aligned} \|u_\iota^k - u_\iota^{k+1}\|_{G_\iota} &\leq \|u_\iota^k - \hat{u}_\iota^k\|_{G_\iota} + \|u_\iota^{k+1} - \hat{u}_\iota^k\|_{G_\iota} \\ &\leq 2\|u_\iota^k - \hat{u}_\iota^k\|_{G_\iota} = 2\mathrm{dist}_{G_\iota}(u_\iota^k, \widehat{\mathcal{U}}) \\ &\leq 2\xi_\iota^k \mathrm{dist}_{G_\iota}(u_\iota^0, \widehat{\mathcal{U}}). \end{aligned}$$

Consequently,

$$\begin{aligned} \|u_\iota^k - u_\iota^\infty\|_{G_\iota} &= \|\sum_{j=k}^{\infty}(u_\iota^k - u_\iota^{k+1})\|_{G_\iota} \\ &\leq \sum_{j=k}^{\infty} \|u_\iota^k - u_\iota^{k+1}\|_{G_\iota} \leq 2\mathrm{dist}_{G_\iota}(u_\iota^0, \widehat{\mathcal{U}}) \sum_{j=k}^{\infty} \xi_\iota^j \\ &= \frac{2\xi_\iota^k}{1 - \xi_\iota} \mathrm{dist}_{G_\iota}(u_\iota^0, \widehat{\mathcal{U}}), \end{aligned}$$

which shows $\{u_\iota^k\}$ converges to u_ι^∞ R-linearly. □

Under proper calmness condition (3.35), Theorem 2 shows the Q-linear convergence rate of $\mathrm{dist}_{G_\iota}(u_\iota^k, \widehat{\mathcal{U}})$ and the nonergodic R-linear convergence rate of $\{u_\iota^k\}$. Since Chambolle-Pock is a special case of Condat-Vu, the linear convergence results also holds for Chambolle-Pock. In addition, since PDFP includes PAPC [9,22] and PDFP^{2}O [5] as special cases, the linear convergence results for PAPC and PDFP^{2}O also follows directly. However, since the constant θ_ι in the calmness condition (3.35) is not known explicitly, the exact linear convergence ratio ξ_ι is not explicitly known either.

Corollary 1 *Suppose that Assumptions 1 and (ιa) hold for $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$. Let $\{(x_\iota^k, y_\iota^k)\}$ with $\iota \in \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$ be the sequence generated by Condat-Vu (1.4), PDFP (1.5) and*

AFBA (1.6)*, respectively. If the mapping* $R : \mathcal{U} \to \mathcal{U}$ *is piecewise polyhedral, then the following properties hold.*
(1) There exists a constant $\hat{\theta} > 0$ *such that for all* $k \geq 0$ *we have*

$$dist(u_\iota^k, \widehat{\mathcal{U}}) \leq \hat{\theta} \| R(u_\iota^k) \|. \tag{3.42}$$

(2) For all $k \geq 0$ *we have*

$$dist_{G_\iota}(u_\iota^{k+1}, \widehat{\mathcal{U}}) \leq \hat{\xi}_\iota \; dist_{G_\iota}(u_\iota^k, \widehat{\mathcal{U}}), \tag{3.43}$$

where

$$\hat{\xi}_\iota = \sqrt{1 - \frac{\alpha_\iota^2}{\beta_\iota^2 (\hat{\theta} \beta_\iota \sqrt{\kappa_\iota} + 1)^2}} < 1.$$

Moreover, $\{u_\iota^k\} := \{(x_\iota^k, y_\iota^k)\}$ *converges to* u_ι^∞ *R-linearly.*

Proof Since R^{-1} is piecewise polyhedral if and only if R is piecewise polyhedral [15], it follows from the discussion at the end of Sect. 2 that there exist two constants $\theta > 0$ and $s > 0$ such that

$$\mathrm{dist}(u, \widehat{\mathcal{U}}) \leq \theta \| R(u) \|, \quad \forall u \in \{u \in \mathcal{U} \,\big|\, \|R(u)\| \leq s\}. \tag{3.44}$$

By Theorem 2, we know $\{u_\iota^k\}$ converges to $u_\iota^\infty \in \widehat{\mathcal{U}}$. Hence, there exists a constant $r > 0$ such that $\|u_\iota^k - u_\iota^\infty\| \leq r$ for all $k \geq 0$. Note that when $\|R(u_\iota^k)\| > s$, we have

$$\mathrm{dist}(u_\iota^k, \widehat{\mathcal{U}}) \leq \|u_\iota^k - u_\iota^\infty\| \leq r < \frac{r}{s} \|R(u_\iota^k)\|. \tag{3.45}$$

Combining (3.44) and (3.45), we have (3.42) holds with $\hat{\theta} := \max\{\theta, \frac{r}{s}\}$. Using (3.42), the property (2) can be similarly proved as the proof in Theorem 2. □

4 Applications to some convex models

In this section, we show some practical examples, where the linear convergence results in the previous section will apply. As one can see in Theorem 2, the calmness condition is the key assumption for linear convergence. So, to show the linear convergence of Condat-Vu (1.4), PDFP (1.5) and (1.6) for solving the example problems, we need to verify that the KKT mapping (2.3) of these problems satisfies the calmness condition (3.35). From Corollary 1, it is sufficient to show the KKT mapping defined in (2.3) is piecewise polyhedral.

Note that the following examples do not satisfy the strongly convex condition required in [4,6]. Hence, the theoretical results given in [4,6] do not imply the linear convergence rate of Condat-Vu, PDFP or AFBA for solving the following example problems. However, from our analysis, these models satisfy the calmness condition and the linear convergence rate can be obtained immediately.

4.1 Matrix games

Consider the following min-max matrix game [4,30]:

$$\min_{x\in\mathcal{R}^n}\max_{y\in\mathcal{R}^m} \delta_{\Delta_n}(x)+\langle Kx,y\rangle-\delta_{\Delta_m}(y), \tag{4.1}$$

where $K\in\mathcal{R}^{m\times n}$, Δ_n and Δ_m denote the standard unit simplices in $\mathcal{R}^n$ and $\mathcal{R}^m$, respectively. [19] showed that the KKT mapping of (4.1) is piecewise polyhedral.

4.2 Fused lasso

The fused lasso problem, which was proposed for group variable selection [21,37,40], can be written as:

$$\min_{x\in\mathcal{R}^n}\frac{1}{2}\|Ax-b\|^2+\mu_1\|x\|_1+\mu_2\|Kx\|_1, \tag{4.2}$$

where $A\in\mathcal{R}^{m\times n}$, $b\in\mathcal{R}^m$, and $K\in\mathcal{R}^{(n-1)\times n}$ is given by

$$K=\begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \cdots & \cdots & \\ & & & -1 & 1 \end{pmatrix}$$

The model (4.2) can be reformulated as

$$\min_{x\in\mathcal{R}^n}\max_{y\in\mathcal{R}^m} f(x)+h(x)+\langle Kx,y\rangle-g(y), \tag{4.3}$$

where $f(x)=\mu_1\|x\|_1$, $h(x)=\frac{1}{2}\|Ax-b\|^2$, $g(y)=\delta_{\mathcal{B}_\infty}(\frac{y}{\mu_2})$ and $\delta_{\mathcal{B}_\infty}(\cdot)$ is the indicator function of the ball $\mathcal{B}_\infty=\{y:\|y\|_\infty\le 1\}$. Then the KKT mapping for this model (4.3) is

$$R(u):=\begin{pmatrix} x-\mathrm{prox}_f(x-Ax+b-K^*y) \\ y-\Pi_{\mu_2\mathcal{B}_\infty}(y-Kx)\end{pmatrix},\quad \forall u\in\mathcal{U}.$$

Since f is piecewise linear, according to Proposition 2, $\mathrm{prox}_f(\cdot)$ is piecewise linear. By recalling that $\mathcal{B}_\infty$ is a polyhedral set, Proposition 2 implies that $\Pi_{\mathcal{B}_\infty}(\cdot)$ is piecewise linear. Therefore, R is piecewise polyhedral and so is R^{-1} [15].

4.3 TV-ℓ_2 image restoration

Many image processing problems involve two regularized terms such as tomography reconstruction, where nonnegative constraint and total variation regularization appear. Consider the following constrained TV-ℓ_2 image restoration problem [6,14,18,26]

$$\min_{x\in\mathcal{B}} \||Kx|\|_1 + \frac{\rho}{2}\|Ax-b\|^2, \tag{4.4}$$

where $b \in \mathcal{R}^n$ is the observed image, A is a blur operator, K is the discrete gradient operator [34] in order to promote sparsity, $(|z|)_i := \sqrt{(z_1)_i^2 + (z_2)_i^2},\ i = 1, 2, \ldots, n$ where $z = (z_1, z_2) \in \mathcal{R}^n \times \mathcal{R}^n, \mathcal{B} = [0, 1]^n$, and ρ is a positive parameter for balancing the data-fidelity and TV regularization. $n = n_1 \times n_2$ is the total number of pixels, where n_1 and n_2 are the numbers of pixels in the horizontal and vertical directions, respectively. Note that the model (4.4) can be reformulated as the following saddle point problem

$$\min_{x\in\mathcal{R}^n} \max_{y\in\mathcal{R}^n\times\mathcal{R}^n} \left\{\delta_{\mathcal{B}}(x) + \frac{\rho}{2}\|Ax-b\|^2 + \langle Kx, y\rangle - \delta_{\mathcal{B}_\infty}(y)\right\}. \tag{4.5}$$

Clearly, (4.5) is the special case of (1.1) with $f(x) = \delta_{\mathcal{B}}(x),\ g(y) = \delta_{\mathcal{B}_\infty}(y)$, and $h(x) = \frac{\rho}{2}\|Ax-b\|^2$. Then the KKT mapping for this model (4.4) is

$$R(u) := \begin{pmatrix} x - \Pi_{\mathcal{B}}(x - Ax + b - K^*y) \\ y - \Pi_{\mathcal{B}_\infty}(y - Kx) \end{pmatrix}, \quad \forall u \in \mathcal{U}.$$

Similarity, we can conclude that both R and R^{-1} are piecewise polyhedral.

5 Numerical experiments

In this section, we firstly show that the larger stepsizes presented in assumption (IIIa) for AFBA result in better performance on fused lasso model. Then we show the linear convergence rate results of Condat-Vu, PDFP and AFBA. All codes were written by MATLAB R2016a and all the numerical experiments were conducted on a personal computer with a 2.20GHz i7 processor and a 16GB memory.

Consider the fused lasso model (4.2). In this simulation, we use the same setting as [39]. The entries of A are generated by the standard Gaussian distribution $\mathcal{N}(0, 1)$ and b is obtained by $b = Ax + \rho e$, where e is a standard distributed Gaussian noise and $\rho = 0.01$. The parameters are set as $\mu_1 = 20$ and $\mu_2 = 200$. Each entry of the initial point (x^0, y^0) is independently generated from $\mathcal{N}(0, 1)$. Recall that the $n-1$ eigenvalues of KK^* can be analytically computed as $2 - 2\cos(i\pi/n), i = 1, 2, \cdots, n-1$.

We firstly show the advantage of larger range of acceptable parameters in AFBA. For AFBA with $\tau\sigma L^2 + \sqrt{\tau\sigma L^2} + \tau L_h/2 < 1$, we set $\lambda := \tau\sigma$ as $1/16$ and $\tau = 2(0.99 - \tau\sigma L^2 - \sqrt{\tau\sigma L^2})/L_h$, whereas we set $\lambda = 1/4$ and $\tau = 1.9/L_h$ for AFBA with larger stepsizes. The stopping criterion is set as $\|u^k - u^{k+1}\|/\|u^k\| \le 10^{-5}$. The optimal solution $\hat{u}$ are obtained by running these methods for 10,000 iterations. Define Error $:= \|u^k - \hat{u}\|^2_{G_\iota}, \iota \in \{\text{I, II, III}\}$ and Error converges Q-linearly according to Theorem 2. We run 5 groups of problems and report the average performances including the computing time in seconds (Time), iteration number (Iter.) and Error in Table 1.

Table 1 Comparisons for different stepsizes in AFBA on fused lasso

r	n	$\tau\sigma L^2 + \sqrt{\tau\sigma L^2} + \tau L_h/2 < 1$			$\tau\sigma L^2 < 1, \tau L_h < 2$		
		Time	Iter.	Error	Time	Iter.	Error
25	500	1.36	695	1.98e−04	0.08	218	1.60e−05
50	1000	4.92	930	2.15e−04	0.60	301	1.05e−04
75	1500	39.52	2063	4.06e−04	5.60	722	2.51e−04
100	2000	121.48	3111	1.22e−03	20	1160	2.23e−04
200	4000	280.58	3416	4.08e−04	37.63	1011	9.61e−05

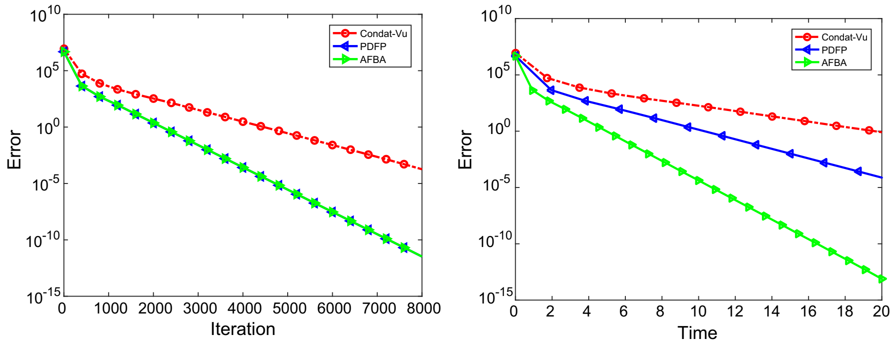

Fig. 1 Numerical experiments for fused lasso (4.2)

As to the linear convergence of these three methods, we set $r = 100$ and $n = 2000$. For Condat-Vu (1.4), the product λ is set as $1/8$ and $\tau = 2(0.99 - \lambda L^2)/L_h$. As to PDFP (1.5), we set $\lambda = 1/4$ and $\tau = 1.9/L_h$. For AFBA (1.6), we set $\lambda = 1/4$ and $\tau = 1.9/L_h$ for better performance. We plot the performance of Error with respect to iteration number and computing time in Fig. 1.

According to Table 1, the larger range of acceptable parameters results in much better performance. Figure 1 (left) shows that 'Error' of these methods converges to zero at a linear rate. Moreover, after running same iteration number, AFBA has the same performance with PDFP, which is better than Condat-Vu. From Fig. 1 (right), after executing same computing time, we can observe that AFBA outperforms PDFP, which performs better than Condat-Vu.

6 Conclusions

In this paper, we provide unified convergence analysis to establish global convergence and the linear convergence rate of a class of primal-dual algorithms such as Condat-Vu, PDFP and AFBA for solving saddle point problems. With a mild calmness condition of the KKT mapping, which naturally holds for many convex models in practical applications, we have established the Q-linear convergence of the distance between the current iterate and the solution set, and the R-linear convergence of the nonergodic iterates for Condat-Vu, PDFP and AFBA. Since Chambolle-Pock, PDFP^{2}O and PAPC are special cases of these algorithms, the global convergence and linear rates of Chambolle-Pock, PDFP^{2}O and PAPC will also follow immediately.

References

1. Cai, X., Han, D., Xu, L.: An improved first-order primal-dual algorithm with a new correction step. J. Global Optim. **57**, 1419–1428 (2013)
2. Chambolle, A., Ehrhardt, M.J., Richtárik, P., Schonlieb, C.-B.: Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications. SIAM J. Optim. **28**, 2783–2808 (2018)
3. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120–145 (2011)
4. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. Math. Program. **159**, 253–287 (2016)
5. Chen, P., Huang, J., Zhang, X.: A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration. Inverse Prob. **29**, 025011 (2013)
6. Chen, P., Huang, J., Zhang, X.: A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions. Fixed Point Theory Appl. **2016**, 1–18 (2016)
7. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**, 460–479 (2013)
8. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings. Springer Monographs in Mathematics, vol. 208. Springer, Berlin (2009)
9. Drori, Y., Sabach, S., Teboulle, M.: A simple algorithm for a class of nonsmooth convex-concave saddle-point problems. Oper. Res. Lett. **43**, 209–214 (2015)
10. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imag. Sci. **3**, 1015–1046 (2010)
11. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, pp. 299–331. North-Holland, Amsterdam (1983)
12. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**, 17–40 (1976)
13. Glowinski, R., Marroco, A.: *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires*, ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 9 (1975), pp. 41–76
14. Han, D., He, H., Yang, H., Yuan, X.: A customized Douglas–Rachford splitting algorithm for separable convex minimization with linear constraints. Numer. Math. **127**, 167–200 (2014)
15. Han, D., Sun, D., Zhang, L.: Linear rate convergence of the alternating direction method of multipliers for convex composite programming. Math. Oper. Res. **43**, 622–637 (2017)
16. He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imag. Sci. **5**, 119–149 (2012)
17. He, H., Desai, J., Wang, K.: A primal-dual prediction-correction algorithm for saddle point optimization. J. Global Optim. **66**, 573–583 (2016)
18. Jiang, F., Cai, X., Wu, Z., Han, D.: Approximate first-order primal-dual algorithms for saddle point problems. Math. Comput. **90**, 1227–1262 (2021)
19. Jiang, F., Wu, Z., Cai, X., Zhang, H.: A first-order inexact primal-dual algorithm for a class of convex-concave saddle point problems. Numer. Algorithms **88**, 1109–1136 (2021)
20. Latafat, P., Patrinos, P.: Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators. Comput. Optim. Appl. **68**, 57–93 (2017)
21. Liu, J., Yuan, L., Ye, J.: An efficient algorithm for a class of fused lasso problems. In: Proceedings of the 16th ACM SIGKDD International Donference on Knowledge Discovery and Data Mining, ACM, pp. 323–332 (2010)
22. Loris, I., Verhoeven, C.: On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty. Inverse Prob. **27**, 125007 (2011)

23. Malitsky, Y., Pock, T.: A first-order primal-dual algorithm with linesearch. SIAM J. Optim. **28**, 411–432 (2018)
24. Mokhtari, A., Ozdaglar, A., Pattathil, S.: A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. In: International Conference on Artificial Intelligence and Statistics, PMLR, pp. 1497–1507 (2020)
25. Mokhtari, A., Ozdaglar, A.E., Pattathil, S.: Convergence rate of O(1/k) for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. SIAM J. Optim. **30**, 3230–3251 (2020)
26. Morini, B., Porcelli, M., Chan, R.H.: A reduced Newton method for constrained linear least-squares problems. J. Comput. Appl. Math. **233**, 2200–2212 (2010)
27. Nesterov, Y.: Lectures on Convex Optimization, vol. 137. Springer, Berlin (2018)
28. O'Connor, D., Vandenberghe, L.: On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting. Math. Program. **179**, 85–108 (2020)
29. Parikh, N., Boyd, S., et al.: Proximal algorithms, Foundations and Trends®. Optimization **1**, 127–239 (2014)
30. Rasch, J., Chambolle, A.: Inexact first-order primal-dual algorithms. Comput. Optim. Appl. **2**, 381–430 (2020)
31. Robinson, S.M.: An implicit-function theorem for generalized variational inequalities. tech. rep., Wisconsin Univ Madison Mathmatics Research Center, (1976)
32. Robinson, S.M.: Some continuity properties of polyhedral multifunctions. In: König, H., Korte, B., Ritter, K. (eds.) Mathematical Programming at Oberwolfach, pp. 206–214. Springer, Berlin (1981)
33. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis, vol. 317. Springer Science & Business Media, Berlin (2009)
34. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**, 259–268 (1992)
35. Shi, F., Cheng, J., Wang, L., Yap, P.-T., Shen, D.: Low-rank total variation for image super-resolution. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 155–162 (2013)
36. Shi, F., Cheng, J., Wang, L., Yap, P.-T., Shen, D.: LRTV: MR image super-resolution with low-rank and total variation regularizations. IEEE Trans. Med. Imaging **34**, 2459–2466 (2015)
37. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and smoothness via the fused lasso. J. Royal Stat. Soc. Ser. B (Stat. Methodol.) **67**, 91–108 (2005)
38. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**, 667–681 (2013)
39. Yan, M.: A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. J. Sci. Comput. **76**, 1698–1717 (2018)
40. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. Royal Stat. Soc. Ser. B (Statistical Methodology) **68**, 49–67 (2006)
41. Zhu, M., Chan, T.: *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, UCLA CAM Report, 34 (2008)

Authors and Affiliations

Fan Jiang[1] · Zhongming Wu[2] · Xingju Cai[3] · Hongchao Zhang[4]

✉ Fan Jiang
15905154902@163.com

Zhongming Wu
wuzm@nuist.edu.cn

Xingju Cai
caixingju@njnu.edu.cn

Hongchao Zhang
hozhang@math.lsu.edu

[1] Department of Information and Computing Science, School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2] School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

[3] School of Mathematical Sciences, Jiangsu Key Laboratory for NSLSCS, Nanjing Normal University, Nanjing 210023, China

[4] Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918, USA