Several variants of the primal-dual hybrid gradient algorithm with applications

Jianchao Bai[1,2*], Jicheng Li[2], Zhie Wu[3]

[1]*Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710129, China*
[2]*School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China*
[3]*School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China*

Abstract. By reviewing the primal-dual hybrid gradient algorithm (PDHG) proposed by He, You and Yuan (SIAM J. Imaging Sci. 2014, 7(4): 2526-2537), in this paper we introduce four improved schemes for solving a class of saddle-point problems. Convergence properties of the proposed algorithms are ensured based on weak assumptions, where none of the objective functions are assumed to be strongly convex but the step-sizes in the primal-dual updates are more flexible than the previous. By making use of variational analysis, the global convergence and sublinear convergence rate in the ergodic/nonergodic sense are established, and the numerical efficiency of our algorithms is verified by testing an image deblurring problem compared with several existing algorithms.

1. Introduction

Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}^{m \times n}$ be the space of $m \times n$ dimensional real matrices, and $\mathbb{R}^n$ be the space of n dimensional real column vectors equipped with inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\|z\|_2 = \sqrt{\langle z, z \rangle}$ for any $z \in \mathbb{R}^n$. Consider the following general saddle-point problem

$$\min_{x_i \in \mathscr{X}_i} \max_{y_j \in \mathscr{Y}_j} F(x, y) := \sum_{i=1}^{p} f_i(x_i) - \sum_{i=1}^{p} \sum_{j=1}^{p} \left\langle y_j, A_i x_i \right\rangle - \sum_{j=1}^{p} g_j(y_j), \tag{1.1}$$

where $x = (x_1^\mathsf{T}, x_2^\mathsf{T}, \cdots, x_p^\mathsf{T})^\mathsf{T}$, $y = (y_1^\mathsf{T}, y_2^\mathsf{T}, \cdots, y_p^\mathsf{T})^\mathsf{T}$ are grouped variables; $p \geq 1$ is any positive integer; $\mathscr{X}_i \subset \mathbb{R}^{m_i}$, $\mathscr{Y}_j \subset \mathbb{R}^n$ are structured and closed convex sets; $f_i(x_i) : \mathbb{R}^{m_i} \to \mathbb{R}$, $g_j(y_j) : \mathbb{R}^n \to \mathbb{R}$ are proper closed convex functions but possibly nonsmooth; and all $A_i \in \mathbb{R}^{n \times m_i}$ are given matrices. Throughout the discussions, the solution set of the problem (1.1) is assumed to be nonempty.

Minimization problems in the form of (1.1) arises in many possible applications, such as the 2D image denoising [12] and machine learning [3, Problem (2)]. In the past several years, a number of first-order algorithms had been developed for solving the problem (1.1) with case $p = 1$. For instance, Zhu and Chan [13] firstly proposed the primal-dual hybrid gradient algorithm (PDHG),

---

*Corresponding author. *Email address:* bjc1987@163.com (J. Bai)

whose iteration alternates between the primal and dual formulations, for solving total variation (TV) minimizations with applications in 2D image processing. Later, this PDHG was extended by Esser et al. [6] to solve a broader class of convex optimization models, and the modified version of PDHG was analyzed to have a similarly good empirical convergence rate for TV minimization problems. In 2011, Chambolle and Pock [4] showed an accelerated version of PDHG for non-smooth convex optimization problems with known saddle-point structure. In particular, their algorithm had $\mathcal{O}(1/t)$ convergence rate for non-smooth problems, and $\mathcal{O}(1/t^2)$ convergence rate for problems where either the primal or dual objective is uniformly convex. Here t denotes the iteration number. To better understand how to choose the step-sizes of the primal-dual updates, He et al. [9] revisited convergence of PDHG by an extremely simple example that it is not necessarily convergent when the step-sizes are fixed as tiny constants. The modified PDHG in [9], that is,

$$
\begin{cases}
x^{k+1} = \arg\min_{x \in \mathcal{X}} \left\{ F(x, y^k) + \frac{r}{2} \left\| x - x^k \right\|_2^2 \right\}, \\
y^{k+1} = \arg\max_{y \in \mathcal{Y}} \left\{ F(x^{k+1}, y) - \frac{s}{2} \left\| y - y^k \right\|_2^2 \right\},
\end{cases}
\tag{1.2}
$$

is indeed globally convergent under the following conditions:

- (A1) $f(x)$ is strongly convex with the modulus $\tau > 0$, i.e., there exists a positive constant τ such that for any $\xi \in \partial f(x)$, it holds

$$
f(\widetilde{x}) - f(x) \geq \langle \widetilde{x} - x, \xi \rangle + \frac{\tau}{2} \| \widetilde{x} - x \|_2^2, \quad \forall x, \widetilde{x} \in \mathcal{X};
$$

- (A2) For given matrix A and $\tau > 0$, the parameter s in (1.2) satisfies $s > \frac{\rho(A^T A)}{\tau}$ where $\rho(\cdot)$ denotes the spectral radius of a matrix.

Clearly, (A1-A2) are strong and not always satisfied for some cases in real applications. For example, the TV regularized linear inversion problem, widely used as a model of salt-pepper noisy image deblurring [12], is of the following form

$$
\min_{x \in \mathcal{X}} \left\{ \frac{1}{2} \| Kx - b \|_1 + \lambda \| Ax \|_{2,1} \right\}.
$$

Here $K \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^m$ denote respectively the given structured matrix, reconstructed image and observed image; $\lambda, \| Ax \|_{2,1} = \sum_i \| A_i x \|_2$ are respectively the penalty parameter and discrete form of TV semi-norm; and $\mathcal{X} = \{ x \in \mathbb{R}^n : l_* \leq x_i \leq u_*, i = 1, 2, \cdots, n \}$ is given with bound constants l_* and u_*. Notice that this problem can be transformed into a two-block case of (1.1):

$$
\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \left\{ \frac{1}{2} \| Kx - b \|_1 + \lambda \langle Ax, y \rangle \right\},
$$

where $\mathcal{Y} = \left\{ y \in \mathbb{R}^{2n} | \, \| y_i \|_2 \leq 1, y_i \in \mathbb{R}^2 \right\}$. In such case $f(x) := \frac{1}{2} \| Kx - b \|_1$ is convex but not strongly convex. Hence, the condition (A1) does not hold, which implies divergence of PDHG (1.2) for such image deblurring problem. The counter example considered in [9] can also explain why (1.2) is not convergent (since there exists no constant $\tau > 0$, i.e., the function $f(x) := x$ is not strongly convex). Besides, we can observe that it depends largely on the modulus of $f(x)$ and

$\rho(A^T A)$ for choosing the parameter s, even though $f(x) := \frac{1}{2}\|Kx - b\|_2^2$ is strongly convex in the case that a full column rank matrix K is considered for image deblurring problem [5, 12, 14].

Motivated by the above discussions, we focus on developing several improved PDHG procedures for tackling the general model (1.1), whose assumptions are much weaker than the conditions (A1-A2) but global convergence still holds. In fact, we originally focus on the traditional saddle-point problem ($p = 1$) in order to weaken convergence conditions of the algorithm in [9]. Surprisingly, we find that our obtained convergence conditions and results hold also for the multi-block structured saddle point problem ($p > 1$). For the sake of conciseness, we first present two basic modified algorithms described in Algorithms 1.1-1.2, where $1/r$ and $1/s$ are called the step-sizes of the primal and dual updates, respectively. $\|\mathbf{A}^\mathsf{T}\mathbf{A}\|$ denotes the spectral norm of matrix $\mathbf{A}^\mathsf{T}\mathbf{A}$, i.e., the square root of its largest eigenvalue. The difference between Algorithm 1.1 and Algorithm 1.2 lies in the updating of the y_j-subproblem and the choice of the parameters (r,s).

Algorithm 1.1. (IPDHA1)

Step 1. Initialize the starting points (x^0, y^0) and parameters

$$(r,s) \in \mathcal{K} := \left\{ r > t_i, \ s \geq \max_i \{1/t_i\} \mid \forall t_i \in \mathbb{R}^+, i = 1, 2, \cdots, p \right\}. \tag{1.3}$$

Step 2. For $k = 0, 1, \cdots$ do

$$\begin{cases} \text{For } i = 1, 2, \cdots, p, \\ x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \left\{ F(x_1^k, \cdots, x_i, \cdots, x_p^k, y^k) + \frac{r}{2} \left\| x_i - x_i^k \right\|_2^2 \right\}, \\ \text{For } j = 1, 2, \cdots, p, \\ y_j^{k+1} = \arg\max_{y_j \in \mathcal{Y}_j} \left\{ F(x^{k+1}, y_1^k, \cdots, y_j, \cdots, y_q^k) - \frac{s}{2} \sum_{i=1}^{p} \left\| A_i^T(y_j - y_j^k) \right\|_2^2 \right\}. \end{cases} \tag{1.4}$$

end

Algorithm 1.2. (IPDHA2)

Step 1. Initialize the starting points (x^0, y^0) and parameters (r,s) such that $rs > \|A^\mathsf{T}A\|$.
Step 2. For $k = 0, 1, \cdots$ do

$$\begin{cases} \text{For } i = 1, 2, \cdots, p, \\ x_i^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \left\{ F(x_1^k, \cdots, x_i, \cdots, x_p^k, y^k) + \frac{r}{2} \left\| x_i - x_i^k \right\|_2^2 \right\}, \\ \text{For } j = 1, 2, \cdots, p, \\ y_j^{k+1} = \arg\max_{y_j \in \mathcal{Y}_j} \left\{ F(2x^{k+1} - x^k, y_1^k, \cdots, y_j, \cdots, y_q^k) - \frac{s}{2} \left\| y_j - y_j^k \right\|_2^2 \right\}. \end{cases} \tag{1.5}$$

end

Contributions of this paper are summarized as the following aspects. Firstly, we develop two improved PDHG schemes to tackle the general saddle-point problem (1.1), where the global convergence and the worst-case $\mathcal{O}(1/t)$ convergence rate in the ergodic sense are analyzed. Secondly, based on Algorithms 1.1-1.2 two modified updates with a relaxation step are introduced in Section 3 to accelerate the convergence of our basic algorithms, whose global convergence and worst-case convergence rate in the ergodic/nonergodic sense are also established. Thirdly, the convergence

conditions of four algorithms are much weaker than before in terms of the objective functions, since there is no assumption on strong convexity of $f_i(x_i)$ or $g_j(y_j)$. Especially, Algorithm 1.1 and its relaxation Algorithm 3.1 are more flexible than (1.2) on choosing the parameters (r, s). For instance, the choice of s in Algorithm 1.1 doesn't depend on the spectral radius of any matrix, which is different from the condition (A2), and the parameter s can be arbitrarily small and close to zero. In addition, preliminary numerical examples on solving the TV image deblurring problem are tested to show the efficiency of our proposed algorithms by comparing with several state-of-the-art algorithms.

Notations: Throughout this paper, denoted by $\mathbb{R}^+$ be the set of positive number and $^\mathsf{T}$ be the transpose of a matrix or vector. The symbols $\mathbf{I}, \mathbf{0}$ stand respectively for the identity matrix and zero matrix/vector with proper dimensions. For an extended real-valued function $h : \mathbb{R}^n \to [-\infty, \infty]$, its conjugate function $h^* : \mathbb{R}^n \to [-\infty, \infty]$ is defined by $h^*(y) = \sup_{x \in \mathbb{R}^n} \{\langle x, y \rangle - f(x)\}$. For convenience of analysis, we introduce the following notations

$$x = (x_1^\mathsf{T}, x_2^\mathsf{T}, \cdots, x_p^\mathsf{T})^\mathsf{T}, \ y = (y_1^\mathsf{T}, y_2^\mathsf{T}, \cdots, y_p^\mathsf{T})^\mathsf{T},$$

$$\Omega = \mathscr{X}_1 \times \mathscr{X}_2 \times \cdots \times \mathscr{X}_p \times \mathscr{Y}_1 \times \mathscr{Y}_2 \times \cdots \times \mathscr{Y}_p,$$

$$\phi(w) = \sum_{i=1}^p f_i(x_i) - \sum_{j=1}^p g_j(y_j), \ w = \begin{pmatrix} x \\ y \end{pmatrix}, \ w^k = \begin{pmatrix} x^k \\ y^k \end{pmatrix}. \tag{1.6}$$

For any $\xi_i \in \partial f_i(x_i), \eta_j \in \partial g_j(y_j)$ where $\partial f_i(x_i)$ is the subdifferential of $f_i(x_i)$, let

$$\mathscr{I}(w) = \begin{pmatrix} \xi_1 - A_1^\mathsf{T} \sum\limits_{j=1}^p y_j \\ \vdots \\ \xi_p - A_p^\mathsf{T} \sum\limits_{j=1}^p y_j \\ \eta_1 + \sum\limits_{i=1}^p A_i x_i \\ \vdots \\ \eta_p + \sum\limits_{i=1}^p A_i x_i \end{pmatrix}, \ \mathscr{I}(w^k) = \begin{pmatrix} \xi_1 - A_1^\mathsf{T} \sum\limits_{j=1}^p y_j^k \\ \vdots \\ \xi_p - A_p^\mathsf{T} \sum\limits_{j=1}^p y_j^k \\ \eta_1 + \sum\limits_{i=1}^p A_i x_i^k \\ \vdots \\ \eta_p + \sum\limits_{i=1}^p A_i x_i^k \end{pmatrix}, \tag{1.7}$$

and

$$\mathbf{A} = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \\ A_1 & A_2 & \cdots & A_p \\ \vdots & \vdots & \ddots & \vdots \\ A_1 & A_2 & \cdots & A_p \end{bmatrix} \in \mathbb{R}^{np \times \sum\limits_{i=1}^p m_i}. \tag{1.8}$$

2. Convergence analysis of IPDHA

In this section, our concerns are devoted to analyzing the global convergence of Algorithms 1.1-1.2 and their worst-case convergence rate in the ergodic sense.

For any point $(x, y) \in \Omega$, we have from (1.1) that

$$F(x^*, y) \le F(x^*, y^*) \le F(x, y^*),$$

where $(x^*, y^*) =: w^* \in \mathcal{X} \times \mathcal{Y}$ denotes the solution pair of the problem (1.1). Note that the above saddle-point inequalities can be equivalently expressed as

$$
\begin{cases}
\text{For } i = 1, 2, \cdots, p, \\
\quad x_i^* \in \mathcal{X}_i, \quad f_i(x_i) - f_i(x_i^*) + \left\langle x_i - x_i^*, -A_i^\mathsf{T} \sum_{j=1}^{p} y_j^* \right\rangle \geq 0, \quad \forall x_i \in \mathcal{X}_i, \\
\text{For } j = 1, 2, \cdots, p, \\
\quad y_j^* \in \mathcal{Y}_j, \quad g_j(y_j) - g_j(y_j^*) + \left\langle y_j - y_j^*, \sum_{i=1}^{p} A_i x_i^* \right\rangle \geq 0, \quad \forall y_j \in \mathcal{Y}_j,
\end{cases}
$$

which is further rewritten as a compact variational inequality(VI)

$$
\text{VI}(\phi, M, \Omega): \quad \phi(w) - \phi(w^*) + \langle w - w^*, Mw^* \rangle \geq 0, \quad \forall w \in \Omega. \tag{2.1}
$$

Here ϕ, w are defined in (1.6), and

$$
M = \begin{bmatrix} \mathbf{0} & -\mathbf{A}^\mathsf{T} \\ \mathbf{A} & \mathbf{0} \end{bmatrix}
$$

is a skew-symmetric matrix and has the property (see also [1, Eq. (11)])

$$
\langle w^{k+1} - w, M w^{k+1} \rangle = \langle w^{k+1} - w, Mw \rangle, \quad \forall w, w^{k+1} \in \Omega. \tag{2.2}
$$

We next show convergence of Algorithm 1.1, and an analogous way will be adopted to analyze the second algorithm. We begin to characterize the optimality conditions of the subproblems in (1.4) by the aid of an equivalent variational inequality.

Lemma 2.1. *The sequence $\{w^k\}$ generated by Algorithm 1.1 satisfies*

$$
\phi(w) - \phi(w^{k+1}) + \langle w - w^{k+1}, M w^{k+1} \rangle \geq \langle w - w^{k+1}, Q(w^k - w^{k+1}) \rangle, \quad \forall w \in \Omega \tag{2.3}
$$

where

$$
Q = \begin{bmatrix} rI & A^\mathsf{T} \\ \mathbf{0} & s \sum_{i=1}^{p} A_i A_i^\mathsf{T} \end{bmatrix}. \tag{2.4}
$$

Proof. For $i = 1, 2, \cdots, p$, the x_i-subproblem clearly amounts to

$$
x_i^{k+1} = \arg \min_{x_i \in \mathcal{X}_i} \left\{ f_i(x_i) - \sum_{j=1}^{p} \langle y_j^k, A_i x_i \rangle + \frac{r}{2} \| x_i - x_i^k \|_2^2 \right\},
$$

whose optimality condition is $x_i^{k+1} \in \mathcal{X}_i$ and

$$
f_i(x_i) - f_i(x_i^{k+1}) + \left\langle x_i - x_i^{k+1}, -A_i^\mathsf{T} \sum_{j=1}^{p} y_j^{k+1} + A_i^\mathsf{T} \sum_{j=1}^{p} (y_j^{k+1} - y_j^k) + r(x_i^{k+1} - x_i^k) \right\rangle \geq 0, \ \forall x_i \in \mathcal{X}_i. \tag{2.5}
$$

Similarly, the y_j-subproblem is equivalent to

$$
y_j^{k+1} = \arg \min_{y_j \in \mathcal{Y}_j} \left\{ g_j(y_j) + \sum_{i=1}^{p} \langle y_j, A_i x_i^{k+1} \rangle + \frac{s}{2} \sum_{i=1}^{p} \left\| A_i^\mathsf{T} (y_j - y_j^k) \right\|_2^2 \right\}, \ \forall j = 1, 2, \cdots, p,
$$

whose optimality condition is

$$y_j^{k+1} \in \mathscr{Y}_j, \ g_j(y_j) - g_j(y_j^{k+1}) + \left\langle y_j - y_j^{k+1}, \sum_{i=1}^{p} A_i x_i^{k+1} + s \sum_{i=1}^{p} A_i A_i^\mathsf{T}(y_j^{k+1} - y_j^k) \right\rangle \geq 0, \ \forall y_j \in \mathscr{Y}_j.$$

Rewriting the above optimality conditions as a unified variational inequality, we achieve

$$w^{k+1} \in \Omega, \ \phi(w) - \phi(w^{k+1}) + \left\langle w - w^{k+1}, M w^{k+1} + Q(w^{k+1} - w^k) \right\rangle \geq 0, \ \forall w \in \Omega.$$

This completes the whole proof. ∎

Lemma 2.2. *For any $t_i > 0$ ($i = 1, 2, \cdots, p$), the sequence $\{w^k\}$ generated by Algorithm 1.1 satisfies*

$$\left\langle w^{k+1} - w, G(w^k - w^{k+1}) \right\rangle$$

$$\geq \ \left\langle w^{k+1} - w, \mathscr{J}(w) \right\rangle - \frac{1}{2} \sum_{i=1}^{p} \left\{ t_i \left\| x_i^{k+1} - x_i^k \right\|_2^2 + \frac{1}{t_i} \sum_{j=1}^{p} \left\| A_i^\mathsf{T}(y_j^k - y_j^{k+1}) \right\|_2^2 \right\}, \quad (2.6)$$

where

$$G = \begin{bmatrix} rI & \mathbf{0} \\ \mathbf{0} & s \displaystyle\sum_{i=1}^{p} A_i A_i^\mathsf{T} \end{bmatrix}. \qquad (2.7)$$

Proof. By the convexity of the functions $f_i(x_i)$ and $g_j(y_j)$, we have

$$\begin{cases} f_i(x_i^{k+1}) - f_i(x_i) \geq \left\langle x_i^{k+1} - x_i, \xi_i \right\rangle, & \forall \xi_i \in \partial f_i(x_i), \ i = 1, 2, \cdots, p, \\ g_j(y_j^{k+1}) - g_j(y_j) \geq \left\langle y_j^{k+1} - y_j, \eta_j \right\rangle, & \forall \eta_j \in \partial g_j(y_j), \ j = 1, 2, \cdots, p, \end{cases}$$

which, by summing themselves over $i, j = 1, 2, \cdots, p$, gives

$$\phi(w^{k+1}) - \phi(w) \geq \left\langle w^{k+1} - w, \left(\xi_1^\mathsf{T}, \cdots, \xi_p^\mathsf{T}, \eta_1^\mathsf{T}, \cdots, \eta_p^\mathsf{T} \right)^\mathsf{T} \right\rangle.$$

Adding the equation (2.2) to both sides of the above inequality, we get

$$\phi(w^{k+1}) - \phi(w) + \left\langle w^{k+1} - w, M w^{k+1} \right\rangle \geq \left\langle w^{k+1} - w, \mathscr{J}(w) \right\rangle. \qquad (2.8)$$

Then, substituting (2.8) into (2.3) it holds

$$\left\langle w^{k+1} - w, Q(w^k - w^{k+1}) \right\rangle \geq \left\langle w^{k+1} - w, \mathscr{J}(w) \right\rangle \qquad (2.9)$$

which is equivalent to

$$\left\langle w^{k+1} - w, G(w^k - w^{k+1}) \right\rangle \geq \left\langle w^{k+1} - w, \mathscr{J}(w) \right\rangle - \sum_{i=1}^{p} \left\langle x_i^{k+1} - x_i, A_i^\mathsf{T} \sum_{j=1}^{p}(y_j^k - y_j^{k+1}) \right\rangle. \ (2.10)$$

Now, by the Cauchy-Schwartz inequality, for any $t_i > 0$ ($i = 1, 2, \cdots, p$) it follows that

$$\left\langle x_i^{k+1} - x_i, A_i^\mathsf{T}(y_j^k - y_j^{k+1}) \right\rangle \leq \frac{t_i}{2} \left\| x_i^{k+1} - x_i \right\|_2^2 + \frac{1}{2t_i} \left\| A_i^\mathsf{T}(y_j^k - y_j^{k+1}) \right\|_2^2, \quad \forall x_i \in \mathscr{X}_i. \quad (2.11)$$

Combining the inequalities (2.10) and (2.11), the result is confirmed. ■

Based on Lemmas 2.1-2.2, we next prove the contraction of the sequence $\{w^k - w^*\}$ under a weighted norm, where the solution set Ω^* of VI(ϕ, M, Ω) is convex and could be characterized as

$$\Omega^* = \bigcap_{w \in \Omega} \left\{ \widetilde{w} \mid \langle w - \widetilde{w}, \mathscr{J}(w) \rangle \geq 0 \right\}. \tag{2.12}$$

Note from (2.7) that the matrices $r\mathbf{I}$ and $s\sum_{i=1}^{p} A_i A_i^{\top}$ are respectively strictly positive definite and positive semi-definite for any $(r,s) \in \mathscr{K}$, so we define the weighted G-norm $\|w\|_G = \sqrt{w^{\top} G w}$, and it holds $\|w\|_G > 0$ for a nonzero w. In the following, we also have $\|w\|_P = \sqrt{w^{\top} P w} > 0$ for the similar reasons.

Theorem 2.1. *For any $(r,s) \in \mathscr{K}$, let $\alpha = \max_{i=1}^{p} \{1/t_i\}$. Then, the sequence $\{w^k\}$ generated by Algorithm 1.1 satisfies*

$$\left\| w^{k+1} - w^* \right\|_G^2 \leq \left\| w^k - w^* \right\|_G^2 - \left\| w^k - w^{k+1} \right\|_P^2, \quad \forall w^* \in \Omega^*,$$

where

$$P = \left[\begin{array}{ccc|c} (r - t_1)\mathbf{I} & \cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{0} \\ \mathbf{0} & \cdots & (r - t_p)\mathbf{I} & \\ \hline & \mathbf{0} & & (s - \alpha)\sum_{i=1}^{p} A_i A_i^{\top} \end{array} \right]. \tag{2.13}$$

Proof. Setting $a = w - w^k, b = w - w^{k+1}$ into the identity

$$\langle b, G(b - a) \rangle = \frac{1}{2} \left(\|b\|_G^2 - \|a\|_G^2 + \|a - b\|_G^2 \right),$$

we have

$$\langle w - w^{k+1}, G(w^k - w^{k+1}) \rangle = \frac{1}{2} \left(\left\| w - w^{k+1} \right\|_G^2 - \left\| w - w^k \right\|_G^2 + \left\| w^k - w^{k+1} \right\|_G^2 \right). \tag{2.14}$$

Then, by substituting (2.14) into (2.6) it follows that

$$\frac{1}{2} \left(\left\| w - w^k \right\|_G^2 - \left\| w - w^{k+1} \right\|_G^2 \right) \geq \left\langle w^{k+1} - w, \mathscr{J}(w) \right\rangle + \frac{1}{2} R_{xy}, \tag{2.15}$$

where

$$
\begin{aligned}
R_{xy} &= \left\| w^k - w^{k+1} \right\|_G^2 - \sum_{i=1}^p \left\{ t_i \left\| x_i^{k+1} - x_i^k \right\|_2^2 + \frac{1}{t_i} \sum_{j=1}^p \left\| A_i^\mathsf{T}(y_j^k - y_j^{k+1}) \right\|_2^2 \right\} \\
&= r \sum_{i=1}^p \left\| x_i^{k+1} - x_i^k \right\|_2^2 - \sum_{i=1}^p t_i \left\| x_i^{k+1} - x_i^k \right\|_2^2 \\
&\quad + \sum_{j=1}^p \left(s \sum_{i=1}^p \left\| A_i^\mathsf{T}(y_j^k - y_j^{k+1}) \right\|_2^2 - \sum_{i=1}^p \frac{1}{t_i} \left\| A_i^\mathsf{T}(y_j^k - y_j^{k+1}) \right\|_2^2 \right) \\
&\geq \sum_{i=1}^p (r - t_i) \left\| x_i^{k+1} - x_i^k \right\|_2^2 + (s - \alpha) \sum_{j=1}^p \sum_{i=1}^p \left\| A_i^\mathsf{T}(y_j^k - y_j^{k+1}) \right\|_2^2 \\
&= \left\| w^{k+1} - w^k \right\|_P^2,
\end{aligned}
\tag{2.16}
$$

in which we take $\alpha = \max\limits_{i=1}^p \left\{ \frac{1}{t_i} \right\}$ and P is defined in (2.13). Now, letting $w = w^*$ in the above inequality (2.15) and using (2.12), the proof is completed. ∎

The above Theorem 2.1 implies the contractive property of the sequence $\{ w^k - w^* \}$ under G-norm. In the following, we prove the global convergence of Algorithm 1.1 and its sublinear convergence rate.

Theorem 2.2. *Let the parameters $(r,s) \in \mathcal{K}$ and the sequence $\{ w^k \}$ be generated by Algorithm 1.1. Then, there exists a $w^\infty \in \Omega^*$ such that $\lim\limits_{k \to \infty} w^{k+1} = w^\infty$. Moreover, for any integer $t > 0$ and $\mathbf{w}_t = \frac{1}{t+1} \sum\limits_{k=0}^t w^{k+1}$, we have*

$$
\langle \mathbf{w}_t - w, \mathscr{J}(w) \rangle \leq \frac{1}{2(t+1)} \| w - w^0 \|_G^2, \quad \forall w \in \Omega.
$$

Proof. We have by Theorem 2.1 that $\{ w^k \}$ is bounded and

$$
\sum_{k=0}^\infty \left\| w^k - w^{k+1} \right\|_P^2 \leq \left\| w^0 - w^* \right\|_G^2 < \infty \implies \lim_{k \to \infty} (w^k - w^{k+1}) = \mathbf{0},
$$

since the fact $\| w \|_P > 0$ for any nonzero $w \in \Omega$ and $\| w \|_P = 0$ if and only if $w = \mathbf{0}$.

Let w^∞ be any accumulation point of $\{ w^k \}$. Then, by taking a subsequence of $\{ w^k \}$ in (2.3) if necessary, we can deduce (since f_i, g_j are proper) that

$$
\phi(w) - \phi(w^\infty) + \langle w - w^\infty, M w^\infty \rangle \geq 0, \qu\forall w \in \Omega,
$$

which implies $w^\infty \in \Omega^*$ comparing to (2.1). Using the result in Theorem 2.1 again, we have

$$
\left\| w^k - w^\infty \right\|_G \leq \left\| w^j - w^\infty \right\|_G \quad \text{for any } k \geq j.
$$

Therefore, the first assertion $\lim\limits_{k \to \infty} w^{k+1} = w^\infty$ holds.

Now, combining (2.16) and (2.15) it holds

$$
\begin{aligned}
\left\langle w^{k+1} - w, \mathscr{J}(w) \right\rangle &\leq \frac{1}{2} \left(\left\| w - w^k \right\|_G^2 - \left\| w - w^{k+1} \right\|_G^2 \right) - \frac{1}{2} \left\| w^{k+1} - w \right\|_P^2 \\
&\leq \frac{1}{2} \left(\left\| w - w^k \right\|_G^2 - \left\| w - w^{k+1} \right\|_G^2 \right),
\end{aligned}
$$

which, by summing up itself over $k = 0, 1, \cdots, t$, gives

$$
\left\langle \sum_{k=0}^{t} w^{k+1} - (1+t)w, \mathscr{J}(w) \right\rangle \leq \frac{1}{2} \left\| w - w^0 \right\|_G^2.
$$

The above inequality can be clearly rewritten as the second assertion. ∎

From the proof of Theorem 2.2 and related work in the literature of the alternating direction method of multipliers, the proximal point algorithm, PDHG and so on (convergence of these first-order algorithms can be analyzed by a unified variational framework analogous to (2.3)), the global convergence and sublinear convergence rate of a similar first-order algorithm can be showed if the extra weighted norm of $w^k - w^{k+1}$ is positive or if its lower bound can be estimated with respect to the iterative error and the constraint error.

Remark 2.1. *The global sublinear convergence of Algorithm 1.2 can be proved in a similar way as that of Algorithm 1.1. Here we give a simple analysis. Regardless of the same x_i-update for $i = 1, 2, \cdots, p$, the optimality condition of y_j-subproblem in (1.5) is*

$$
y_j^{k+1} \in \mathscr{Y}_j, \ g_j(y_j) - g_j(y_j^{k+1}) + \left\langle y_j - y_j^{k+1}, \sum_{i=1}^{p} A_i(2x_i^{k+1} - x_i^k) + s(y_j^{k+1} - y_j^k) \right\rangle \geq 0, \ \forall y_j \in \mathscr{Y}_j.
$$

Combining the above inequality and (2.5), we have a similar VI as (2.3):

$$
\phi(w) - \phi(w^{k+1}) + \left\langle w - w^{k+1}, M w^{k+1} \right\rangle \geq \left\langle w - w^{k+1}, \mathscr{Q}(w^k - w^{k+1}) \right\rangle, \tag{2.17}
$$

where

$$
\mathscr{Q} = \begin{bmatrix} r\boldsymbol{I} & \boldsymbol{A}^\top \\ \boldsymbol{A} & s\boldsymbol{I} \end{bmatrix}.
$$

It easy to deduce from (2.17) by using the similar identity as (2.14) and the replacement of w by w^* that

$$
\left\| w^{k+1} - w^* \right\|_{\mathscr{Q}}^2 \leq \left\| w^k - w^* \right\|_{\mathscr{Q}}^2 - \left\| w^k - w^{k+1} \right\|_{\mathscr{Q}}^2, \quad \forall w^* \in \Omega^*,
$$

which implies the contractive property of the sequence $\{w^k - w^*\}$ under the condition $rs > \|\boldsymbol{A}^\top \boldsymbol{A}\|$. So, Algorithm 1.2 converges globally with a sublinear convergence rate, that is, there exists a $w^\infty \in \Omega^*$ such that the sequence $\{w^k\}$ generated by Algorithm 1.2 satisfies

$$
\lim_{k \to \infty} w^{k+1} = w^\infty \quad \text{and} \quad \langle \boldsymbol{w}_t - w, \mathscr{J}(w) \rangle \leq \frac{1}{2(t+1)} \| w - w^0 \|_{\mathscr{Q}}^2,
$$

whose proof is also analogous to [2, Theorems 1-2] and thus omitted for the sake of conciseness.

3. Acceleration of IPDHA

From now on, the output of IPDHA is denoted by $\bar{w}^k = (\bar{x}^k, \bar{y}^k)$ with given iterate $w^k = (x^k, y^k)$. According to the results in Section 2, we next provide two modified IPDHA schemes with a relaxation step (see Algorithms 3.1-3.2) to accelerate convergence of Algorithms 1.1-1.2. We still show their sublinear convergence rate but in the ergodic and nonergodic sense.

Algorithm 3.1. (RPDHA1)

Step 1. Initialize the starting points (x^0, y^0) and parameters $(r, s) \in \mathcal{K}$ defined in (1.3).
Step 2. For $k = 0, 1, \cdots$ do

$$
\begin{cases}
\text{For } i = 1, 2, \cdots, p, \\
\quad \bar{x}_i^k = \arg\min_{x_i \in \mathcal{X}_i} \left\{ F(x_1^k, \cdots, x_i, \cdots, x_p^k, y^k) + \frac{r}{2} \left\| x_i - x_i^k \right\|_2^2 \right\}, \\
\text{For } j = 1, 2, \cdots, p, \\
\quad \bar{y}_j^k = \arg\max_{y_j \in \mathcal{Y}_j} \left\{ F(\bar{x}^k, y_1^k, \cdots, y_j, \cdots, y_q^k) - \frac{s}{2} \sum_{i=1}^{p} \left\| A_i^T(y_j - y_j^k) \right\|_2^2 \right\}.
\end{cases}
$$

end
Step 3. Relaxation step: determine an inverse matrix D and $\vartheta \in \mathbb{R}^+$, and then update w^{k+1} by

$$w^{k+1} = w^k - \vartheta D(w^k - \bar{w}^k). \tag{3.1}$$

Algorithm 3.2. (RPDHA2)

Step 1. Initialize the starting points (x^0, y^0) and parameters (r, s) such that $rs > \|A^\mathsf{T} A\|$.
Step 2. For $k = 0, 1, \cdots$ do

$$
\begin{cases}
\text{For } i = 1, 2, \cdots, p, \\
\quad \bar{x}_i^k = \arg\min_{x_i \in \mathcal{X}_i} \left\{ F(x_1^k, \cdots, x_i, \cdots, x_p^k, y^k) + \frac{r}{2} \left\| x_i - x_i^k \right\|_2^2 \right\}, \\
\text{For } j = 1, 2, \cdots, p, \\
\quad \bar{y}_j^k = \arg\max_{y_j \in \mathcal{Y}_j} \left\{ F(2\bar{x}^k - x^k, y_1^k, \cdots, y_j, \cdots, y_q^k) - \frac{s}{2} \left\| y_j - y_j^k \right\|_2^2 \right\}.
\end{cases}
$$

end
Step 3. Relaxation step: determine an inverse matrix D and $\vartheta \in \mathbb{R}^+$, and then update w^{k+1} by

$$w^{k+1} = w^k - \vartheta D(w^k - \bar{w}^k).$$

Clearly, the first two steps in Algorithms 3.1-3.2 are the same as that in Algorithms 1.1-1.2 with small modifications, and the final relaxation step follows the same update. However, their convergence conditions are different from IPDHA. In what follows, we would take Algorithm 3.1 for an example to illustrate its convergence properties.

For the matrix Q given in (2.4) and D, ϑ involved in Algorithm 3.1, let

$$\bar{H} := QD^{-1}, \quad \bar{G} := Q^\mathsf{T} + Q - \vartheta D^\mathsf{T} \bar{H} D. \tag{3.2}$$

In fact, these two matrices are given by the proof of Lemma 3.1. We then use the notations $\|w\|_{\bar{H}}^2 = w^\mathsf{T} \bar{H} w, \|w\|_{\bar{G}}^2 = w^\mathsf{T} \bar{G} w$, and the following assumption for the subsequent convergence analysis.

Assumption 3.1. *The matrices $\bar{H}$ and $\bar{G}$ are symmetric positive definite.*

Lemma 3.1. *For the matrices $\bar{H}, \bar{G}$ defined in (3.2), under Assumption 3.1 the sequence $\{w^k\}$ generated by Algorithm 3.1 satisfies*

$$\phi(w) - \phi(\bar{w}^k) + \left\langle w - \bar{w}^k, M\bar{w}^k \right\rangle \geq \frac{1}{2\vartheta}\left(\|w - w^{k+1}\|_{\bar{H}}^2 - \|w - w^k\|_{\bar{H}}^2\right) + \frac{1}{2}\|w^k - \bar{w}^k\|_{\bar{G}}^2. \quad (3.3)$$

Proof. By Lemma 2.1, (3.1)-(3.2), we have

$$\begin{aligned}
&\vartheta\left\{\phi(w) - \phi(\bar{w}^k) + \left\langle w - \bar{w}^k, M\bar{w}^k \right\rangle\right\} \\
&\geq \quad \left\langle w - \bar{w}^k, \bar{H}(w^k - w^{k+1}) \right\rangle \\
&= \quad \frac{1}{2}\left(\|w - w^{k+1}\|_{\bar{H}}^2 - \|w - w^k\|_{\bar{H}}^2\right) + \frac{1}{2}\left(\|w^k - \bar{w}^k\|_{\bar{H}}^2 - \|w^{k+1} - \bar{w}^k\|_{\bar{H}}^2\right), \quad (3.4)
\end{aligned}$$

where the above equality uses the identity

$$\langle a - b, \bar{H}(c - d) \rangle = \frac{1}{2}\left(\|a - d\|_{\bar{H}}^2 - \|a - c\|_{\bar{H}}^2\right) + \frac{1}{2}\left(\|c - b\|_{\bar{H}}^2 - \|d - b\|_{\bar{H}}^2\right)$$

with specifications $a = w, b = \bar{w}^k, c = w^k, d = w^{k+1}$.

Notice that, by using (3.1)-(3.2) the last term of (3.4) can be further simplified as

$$\begin{aligned}
&\|w^k - \bar{w}^k\|_{\bar{H}}^2 - \|\bar{w}^k - w^{k+1}\|_{\bar{H}}^2 \\
&= \quad \|w^k - \bar{w}^k\|_{\bar{H}}^2 - \|w^k - \bar{w}^k - (w^k - w^{k+1})\|_{\bar{H}}^2 \\
&= \quad \|w^k - \bar{w}^k\|_{\bar{H}}^2 - \|w^k - \bar{w}^k - \vartheta D(w^k - \bar{w}^k)\|_{\bar{H}}^2 \\
&= \quad \vartheta(w^k - \bar{w}^k)^\mathsf{T}(\bar{H}D + D^\mathsf{T}\bar{H})(w^k - \bar{w}^k) - \vartheta^2(w^k - \bar{w}^k)^\mathsf{T}D^\mathsf{T}\bar{H}D(w^k - \bar{w}^k) \\
&= \quad \vartheta(w^k - \bar{w}^k)^\mathsf{T}(Q^\mathsf{T} + Q - \vartheta D^\mathsf{T}\bar{H}D)(w^k - \bar{w}^k) \\
&= \quad \vartheta\|w^k - \bar{w}^k\|_{\bar{G}}^2, \quad (3.5)
\end{aligned}$$

where the fourth equality use the symmetric property of the matrix $\bar{H}$. Therefore, substituting (3.5) into (3.4) we complete the whole proof. ∎

Now, by setting $w = w^*$ in (3.3) and using VI(ϕ, M, Ω) in (2.1), the following theorem holds directly.

Theorem 3.1. *For the matrices $\bar{H}, \bar{G}$ defined in (3.2), under Assumption 3.1 the sequence $\{w^k\}$ generated by Algorithm 3.1 satisfies*

$$\left\|w^{k+1} - w^*\right\|_{\bar{H}}^2 \leq \left\|w^k - w^*\right\|_{\bar{H}}^2 - \vartheta\left\|w^k - \bar{w}^k\right\|_{\bar{G}}^2, \quad \forall w^* \in \Omega^*.$$

It is clear that both the terms $\|\cdot\|_{\bar{H}}$ and $\|\cdot\|_{\bar{G}}$ are positive from Assumption 3.1, which shows the contractive property of the sequence $\{w^k - w^*\}$ according to the result in Theorem 3.1. Next, we prove the global convergence of Algorithm 3.1 and its worst-case $\mathcal{O}(1/t)$ convergence rate in the ergodic and nonergodic case, where t denotes the iteration number.

Theorem 3.2. *Let the sequence $\{w^k\}$ be generated by Algorithm 3.1 and Assumption 3.1 hold. Then, there exists a $w^\infty \in \Omega^*$ such that $\lim\limits_{k\to\infty} w^{k+1} = w^\infty$. Moreover, for any integer $t > 0$ and $\bar{w}_t = \frac{1}{t+1}\sum\limits_{k=0}^{t} \bar{w}^k$, we have*

$$\langle \bar{w}_t - w, \mathcal{J}(w) \rangle \leq \frac{1}{2\vartheta(t+1)}\|w - w^0\|_{\bar{H}}^2, \quad \forall w \in \Omega.$$

Proof. The first result of Theorem 3.2 can be proved in a similar way as that of Theorem 2.2 and is omitted here. For the second result, by (2.9) and (3.4)-(3.5), we can get

$$
\begin{aligned}
\langle \bar{w}^k - w, \mathscr{J}(w) \rangle &\leq -\left\langle w - \bar{w}^k, Q(w^k - \bar{w}^k) \right\rangle && (3.6) \\
&= -\frac{1}{\vartheta} \left\langle w - \bar{w}^k, \bar{H}(w^k - w^{k+1}) \right\rangle \\
&= \frac{1}{2\vartheta} \left(\|w - w^k\|_{\bar{H}}^2 - \|w - w^{k+1}\|_{\bar{H}}^2 \right) - \frac{1}{2} \|w^k - \bar{w}^k\|_{\bar{G}}^2 \\
&\leq \frac{1}{2\vartheta} \left(\|w - w^k\|_{\bar{H}}^2 - \|w - w^{k+1}\|_{\bar{H}}^2 \right), && (3.7)
\end{aligned}
$$

where the above last inequality uses the positive definiteness of matrix $\bar{G}$. By summing the above inequality (3.7) over $k = 0, 1, \cdots, t$ and using the notation $\bar{\mathbf{w}}_t$, we deduce

$$
\langle \bar{\mathbf{w}}_t - w, \mathscr{J}(w) \rangle \leq \frac{1}{2\vartheta(t+1)} \|w - w^0\|_{\bar{H}}^2. \quad \blacksquare
$$

Lemma 3.2. *Let the sequence $\{w^k\}$ be generated by Algorithm 3.1 and Assumption 3.1 hold. Then, we have*

$$
\|D(w^{k+1} - \bar{w}^{k+1})\|_{\bar{H}} \leq \|D(w^k - \bar{w}^k)\|_{\bar{H}}, \quad \forall k \geq 0.
$$

Proof. Firstly, by the identity $\|a\|_{\bar{H}}^2 - \|b\|_{\bar{H}}^2 = 2a^{\mathsf{T}}\bar{H}(a - b) - \|a - b\|_{\bar{H}}^2$ with substitutions

$$
a = D(w^k - \bar{w}^k) \quad \text{and} \quad b = D(w^{k+1} - \bar{w}^{k+1}),
$$

it can be achieved that

$$
\begin{aligned}
&\|D(w^k - \bar{w}^k)\|_{\bar{H}}^2 - \|D(w^{k+1} - \bar{w}^{k+1})\|_{\bar{H}}^2 \\
&= 2(w^k - \bar{w}^k)^{\mathsf{T}} D^{\mathsf{T}}\bar{H}D \left\{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\} - \left\| D \left\{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\} \right\|_{\bar{H}}^2 && (3.8)
\end{aligned}
$$

Secondly, we prove

$$
(w^k - \bar{w}^k)^{\mathsf{T}} D^{\mathsf{T}}\bar{H}D \left\{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\} \geq \frac{1}{2\vartheta} \left\| (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\|_{Q^{\mathsf{T}}+Q}^2. \quad (3.9)
$$

Note that (3.6) with $k := k+1$ and $w := \bar{w}^k$ gives

$$
(\bar{w}^{k+1} - \bar{w}^k)^{\mathsf{T}} \mathscr{J}(\bar{w}^k) \leq (\bar{w}^{k+1} - \bar{w}^k)^{\mathsf{T}} Q(w^{k+1} - \bar{w}^{k+1}), \quad (3.10)
$$

and (3.6) itself with $w := \bar{w}^{k+1}$ implies

$$
(\bar{w}^k - \bar{w}^{k+1})^{\mathsf{T}} \mathscr{J}(\bar{w}^{k+1}) \leq (\bar{w}^k - \bar{w}^{k+1})^{\mathsf{T}} Q(w^k - \bar{w}^k). \quad (3.11)
$$

Adding (3.10) to (3.11) together with (2.12), we obtain

$$
(\bar{w}^k - \bar{w}^{k+1})^{\mathsf{T}} Q \left\{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\} \geq 0. \quad (3.12)
$$

Then, by adding the term

$$
\left\{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\}^{\mathsf{T}} Q \left\{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\} = \frac{1}{2} \|(w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1})\|_{Q^{\mathsf{T}}+Q}^2
$$

to both sides of (3.12), we have

$$(w^k - w^{k+1})^\mathsf{T} Q \left\{ (w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1}) \right\} \geq \frac{1}{2} \|(w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1})\|^2_{Q^\mathsf{T}+Q},$$

which immediately confirms (3.9) by combining (3.1) and $Q = \bar{H}D$.

Finally, substituting (3.9) into (3.8) and using Assumption 3.1, we get

$$\|D(w^k - \bar{w}^k)\|^2_{\bar{H}} - \|D(w^{k+1} - \bar{w}^{k+1})\|^2_{\bar{H}} \geq \frac{1}{\vartheta} \|(w^k - \bar{w}^k) - (w^{k+1} - \bar{w}^{k+1})\|^2_{\bar{G}} \geq 0. \quad \blacksquare$$

Theorem 3.3. *Let the sequence $\{w^k\}$ be generated by Algorithm 3.1 and Assumption 3.1 hold. Then, for any integer $t > 0$, there exists a constant $c_0 > 0$ such that*

$$\|D(w^t - \bar{w}^t)\|^2_{\bar{H}} \leq \frac{1}{(t+1)\vartheta c_0} \|w^0 - \bar{w}^*\|^2_{\bar{H}}.$$

Proof. By Theorem 3.1 and Assumption 3.1, there exists a constant $c_0 > 0$ such that

$$\left\| w^{k+1} - w^* \right\|^2_{\bar{H}} \leq \left\| w^k - w^* \right\|^2_{\bar{H}} - \vartheta c_0 \left\| D(w^k - \bar{w}^k) \right\|^2_{\bar{H}}, \quad \forall w^* \in \Omega^*,$$

which deduces to

$$\vartheta c_0 \sum_{k=0}^{t} \left\| D(w^k - \bar{w}^k) \right\|^2_{\bar{H}} \leq \left\| w^0 - w^* \right\|^2_{\bar{H}}, \tag{3.13}$$

Besides, it follows from Lemma 3.2 that

$$(t+1) \left\| D(w^t - \bar{w}^t) \right\|^2_{\bar{H}} \leq \sum_{k=0}^{t} \left\| D(w^k - \bar{w}^k) \right\|^2_{\bar{H}}. \tag{3.14}$$

Combining the above inequalities (3.13) and (3.14), the proof is completed. $\blacksquare$

Remark 3.1. *By Theorem 3.3, we know that Algorithm 3.1 with a relaxation step is globally convergent with the worst-case $\mathcal{O}(1/t)$ convergence rate in the nonergodic sense. Analogous to the whole analysis in Section 3, the global convergence and the same convergence rate in the ergodic/nonergodic sense of Algorithm 3.2 can be proved.*

Remark 3.2. *Note that Theorems 3.1-3.2 are very similar to Theorems 2.1-2.2. For the choice of ϑ, it is not easy to directly obtain its region for any inverse matrix D (we can treat it as a preconditioned matrix from the prospect of numerical algebra) in our certain knowledge, but we can list two special cases to determine it. For instance, in Algorithm 3.2,*

- *when taking $D = \varrho I$ with $\varrho > 0$, we have $\bar{G} = \varrho(2 - \vartheta\varrho)\bar{H}$ and*

$$\|w\|^2_{\bar{G}} = \varrho(2 - \vartheta\varrho)\|w\|^2_{\bar{H}} = \frac{2 - \varrho\vartheta}{\varrho} \|Dw\|^2_{\bar{H}},$$

which implies $c_0 = \frac{2-\varrho\vartheta}{\varrho}$ and $\vartheta \in (0, \frac{2}{\varrho})$. In such case, by Theorems 3.2-3.3, we conjecture that Algorithm 3.2 may converge faster than Algorithm 1.2 when ϑ closes to $2/\varrho$. What's more, if we take $D = I$ and $\vartheta \in (0, 2)$, then the relaxation step is similar to relaxed proximal point methods (see e.g. [2, 7, 8]), which shows that our algorithm is more general.

- *when taking $D = U$ where U is orthogonal matrix, we deduce $\bar{G} = (2I - \vartheta U^{\mathsf{T}})\bar{H}U$ and*

$$
\begin{aligned}
\|w\|_{\bar{G}}^2 &= \left\langle (2I - \vartheta U^{\mathsf{T}})^{\mathsf{T}}w, \bar{H}Uw \right\rangle \\
&= \left\langle (2I - \vartheta U^{\mathsf{T}})^{\mathsf{T}}U^{\mathsf{T}}Uw, \bar{H}Uw \right\rangle \\
&= \left\langle (2U^{\mathsf{T}} - \vartheta I)Dw, \bar{H}Dw \right\rangle \\
&= c_0 \left\langle Dw, \bar{H}Dw \right\rangle = c_0\|Dw\|_{\bar{H}}^2,
\end{aligned}
$$

which together with $c_0 > 0$ implies that ϑ should be smaller than the eigenvalues of matrix $2U^{\mathsf{T}}$.

4. Numerical experiments

This section investigates the performance of our proposed algorithms by solving the standard image deblurring problem about *man*, *salt-pepper* and *earth* images shown in Figure 1, and numerical results are reported and analyzed by comparing to several existing algorithms. All experiments are conducted in MATLAB 7.10(R2010a) on a PC with Intel Core i5 processor(3.3GHz) with 4 GB memory.
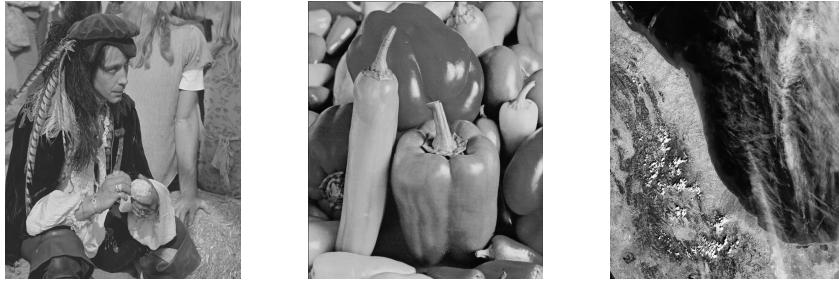

Figure 1: The left is a 512×512 original *man* image, the middle is a 512×512 original *salt-pepper* image, and the right is a 1024×1024 original *earth* image.

4.1. Test problem

The popular Gaussian noise image deblurring problem is of the following form

$$
\min_{x \in \mathbb{R}^{mn}} \left\{ \frac{\lambda}{2}\|\mathscr{P}x - b\|_2^2 + \|\mathscr{A}x\|_{2,1} \right\}, \tag{4.1}
$$

where $\lambda > 0$ denotes a weighting parameter; $\mathscr{P} \in \mathbb{R}^{mn \times mn}$ is convolution operator (possibly large and ill-conditioned); $\mathscr{A} \in \mathbb{R}^{2mn \times mn}$ is total difference operator consisting of vertical and horizontal difference operator; $b \in \mathbb{R}^{mn}$ is given vector vectorized from an observed image; the pixel values of the image x belong to $[0, 255]$ and can be scaled to $[0, 1]$. Notice that the second term in the objective function of (4.1) can be equivalently transformed into

$$
\|\mathscr{A}x\|_{2,1} := \sum_{i=1}^{mn} \|\mathscr{A}_i x\|_2 = \max_{v \in \mathscr{V}} \langle \mathscr{A}x, v \rangle = \max_{v \in \mathbb{R}^{2mn}} \left\{ \langle \mathscr{A}x, v \rangle - \delta_{\mathscr{V}}(v) \right\}, \tag{4.2}
$$

in which

$$
\mathscr{V} = \{v \in \mathbb{R}^{2mn} \mid \|v_i\|_2 \le 1, v_i \in \mathbb{R}^2, i = 1, 2, \cdots, mn\}
$$

is a convex set whose indicator function is denoted by $\delta_{\mathscr{V}}(v)$. By (4.2) the problem (4.1) is rewritten as

$$\min_{x \in \mathbb{R}^{mn}} \max_{v \in \mathscr{V}} \left\{ \frac{\lambda}{2} \|\mathscr{P}x - b\|_2^2 + \langle \mathscr{A}x, v \rangle \right\} \tag{4.3}$$

or expressed as the following form [12]:

$$\min_{x \in \mathbb{R}^{mn}} \max_{v \in \mathbb{R}^{2mn}} \left\{ \frac{\lambda}{2} \|\mathscr{P}x - b\|_2^2 + \langle \mathscr{A}x, v \rangle - \delta_{\mathscr{V}}(v) + \delta_{[0,1]}(x) \right\}, \tag{4.4}$$

where $\delta_{[0,1]}(x)$ is the indicator function of $[0,1]$ and can be converted to a dual expression

$$\delta_{[0,1]}(x) = \max_{w \in \mathbb{R}^{mn}} \left\{ \langle x, w \rangle - \delta_{[0,1]}^*(w) \right\}. \tag{4.5}$$

Here $\delta_{[0,1]}^*(w)$ stands for the conjugate function of $\delta_{[0,1]}(x)$. Substituting (4.5) into (4.4), we have

$$\min_{x \in \mathbb{R}^{mn}} \max_{v \in \mathbb{R}^{2mn}, w \in \mathbb{R}^{mn}} \left\{ \frac{\lambda}{2} \|\mathscr{P}x - b\|_2^2 + \langle \mathscr{A}x, v \rangle - \delta_{\mathscr{V}}(v) + \langle x, w \rangle - \delta_{[0,1]}^*(w) \right\}, \tag{4.6}$$

which is a special case of the problem (1.1) with $p = q = 1$ and specifications

$$f(x) = \frac{\lambda}{2} \|\mathscr{P}x - b\|_2^2, \ g(y) = \delta_{\mathscr{V}}(v) + \delta_{[0,1]}^*(w), \ \bar{A} = -\left(\begin{array}{c} \mathscr{A} \\ \mathbf{I} \end{array} \right), \ y = \left(\begin{array}{c} v \\ w \end{array} \right).$$

Next, we use Algorithms 1.2 and 3.2 to solve the transformed problem (4.6), while we apply Algorithms 1.1 and 3.1 to solve the problem (4.3). For the sake of conciseness, we take an example to illustrate how to solve the involved subproblems of Algorithm 1.2. It is easy to know that the x-update is

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^{mn}} \left\{ \frac{\lambda}{2} \|\mathscr{P}x - b\|_2^2 + \langle \mathscr{A}x, v \rangle + \langle x, w \rangle + \frac{r}{2} \|x - x^k\|_2^2 \right\},$$

whose first-order optimality condition is $(\lambda \mathscr{P}^\top \mathscr{P} + r\mathbf{I})x = \lambda \mathscr{P}^\top b - \mathscr{A}^\top v - w + r x^k =: \bar{b}$. By making use of the convolution theorem of Fourier transforms as shown in [11], the closed form solution for the above linear equation is derived as

$$x^{k+1} = \mathscr{F}^{-1} \left(\frac{\mathscr{F}(\bar{b})}{\lambda \mathscr{F}(\mathscr{P})^* \circ \mathscr{F}(\mathscr{P}) + r} \right), \tag{4.7}$$

where $\mathscr{F}, \mathscr{F}^{-1}$ denote fast Fourier transform and its inverse transform, respectively; the symbols $*$ and $\circ$ stand respectively for complex conjugacy and componentwise multiplication in which the division is computed component-wisely. Besides, the y-update is equivalent to

$$y^{k+1} = \arg \max_{y=(v,w)} \left\{ \langle \mathscr{A}\bar{x}^k, v \rangle - \delta_{\mathscr{V}}(v) + \langle \bar{x}^k, w \rangle - \delta_{[0,1]}^*(w) - \frac{s}{2} \left(\|v - v^k\|_2^2 + \|w - w^k\|_2^2 \right) \right\} \tag{4.8}$$

with $\bar{x}^k = 2x^{k+1} - x^k$. Note that the above y-subproblem is separable. Therefore, we have from (4.2) and (4.8) that

$$\begin{aligned} v^{k+1} &= \arg \max_{\|v_i\|_2 \leq 1} \left\{ \langle \mathscr{A}\bar{x}^k, v \rangle - \frac{s}{2} \|v - v^k\|_2^2 \right\} \\ &= \arg \min_{\|v_i\|_2 \leq 1} \frac{s}{2} \|v - v^k - \mathscr{A}\bar{x}^k/s\|_2^2 \\ &= \mathscr{M}_{\|v_i\|_2 \leq 1} \left(v^k + \mathscr{A}\bar{x}^k/s \right), \end{aligned} \tag{4.9}$$

where $\mathscr{M}_\Omega(\cdot)$ denotes the projection operator onto the set Ω, and

$$
\begin{aligned}
w^{k+1} &= \arg\max_{w\in\mathbb{R}^{mn}} \left\{ \langle \bar{x}^k, w \rangle - \delta^*_{[0,1]}(w) - \frac{s}{2}\|w - w^k\|_2^2 \right\} \\
&= \arg\min_{w\in\mathbb{R}^{mn}} \left\{ \frac{s}{2}\|w - w^k - \bar{x}^k/s\|_2^2 + \delta^*_{[0,1]}(w) \right\} \\
&= w^k + \frac{1}{s}\bar{x}^k - \frac{1}{s}\mathscr{M}_{[0,1]}\left(sw^k + \bar{x}^k \right),
\end{aligned}
\tag{4.10}
$$

where the above last equality comes from the Moreau proximal identity and definition of proximal operator (see e.g. [10, Eq.(2.5)] for more details).

4.2. Numerical results

In this subsection, we test the above mentioned example by seven well-established algorithms with tuned values of parameters: **Algorithms 1.1-3.2** with $D = \mathbf{I}, \vartheta = 0.6$ which are set by Remark 3.2; **CP** (i.e., Algorithm 1 in [4]) with its parameters satisfying $\sigma\tau L^2 = 1$ and $\theta = 0.6$ that has the same meaning as ϑ; **PD** (i.e., the scheme (17) in [14]) and **PDHG** (i.e., PDHG-D in [13]), where Algorithms 1.2 and 3.2 are used to solve the transformed problem (4.6) and Algorithms 1.1 and 3.1 are used to solve (4.3). All codes are written based on PDHG (the x-subproblem is solved by fast Fourier transform for all algorithms) that is downloaded from http://pages.cs.wisc.edu/~swright/ GPUreconstruction/, and the primal and dual steps are respectively fixed as unified values 0.25 and 0.0018 for the sake of comparison experiments (note that these settings can be satisfied for our proposed algorithms when taking $t_i \in (0.0018, 4)$). The testing images are degraded by convolutions and zero-mean Gaussian noise with variance 10^{-6}, where the blur operator and the additive noise are generated by scripts `fspecial()` and `imnoise()` in MATLAB Image Processing Toolbox. Specifically, we set the medium Gaussian blur with `hsize=12` and `sigma = 5`. In order to deblur corrupted images, the problem parameter in (4.1) is set as $\lambda = 5500$ for the Gaussian blur cases, and the following stopping criterion is used to terminate all algorithms:

$$
\text{IT\_res}(k+1) := \frac{\|x^{k+1} - x^k\|_2}{\|x^{k+1}\|_2} \leq 1.0 \times 10^{-4}
$$

under the maximal iteration 1000. We also use the following relative error to evaluate the quality of the restored image:

$$
\text{RE}(k+1) := \frac{\|x^{k+1} - \hat{x}\|_2}{\|\hat{x}\|_2},
\tag{4.11}
$$

where x^{k+1} and $\hat{x}$ denote the image restored by a certain algorithm and the original image, respectively.

Figure 2: From left to right and up to down: *man* image with medium Gaussian blur, restored images by algorithms PD, CP, PDHG, IPDHA1, IPDHA2, RPDHA2 and RPDHA1, respectively.

Images	Algorithm	Iter(k)	CPU	RE	IT_res
	CP	307	24.6482	0.0122	9.9931e-5
	PD	491	37.8302	0.0122	9.9933e-5
	PDHG	491	38.1875	0.0122	9.9931e-5
Man	IPDHA1	453	61.5720	0.0120	9.9899e-5
	IPDHA2	491	36.2686	0.0122	9.9939e-5
	RPDHA1	282	36.7803	0.0122	9.9773e-5
	RPDHA2	295	23.5014	0.0122	9.9893e-5
	CP	250	21.9213	0.0118	9.9958e-5
	PD	400	34.3046	0.0118	9.9940e-5
	PDHG	400	30.1706	0.0118	9.9938e-5
Salt-pepper	IPDHA1	369	48.8739	0.0115	9.9860e-5
	IPDHA2	402	29.8726	0.0118	9.9907e-5
	RPDHA1	243	33.9582	0.0118	9.9956e-5
	RPDHA2	258	20.6498	0.0118	9.9738e-5
	CP	369	238.8214	0.0149	9.9948e-5
	PD	590	282.4554	0.0149	9.9963e-5
	PDHG	590	244.3288	0.0149	9.9961e-5
Earth	IPDHA1	559	438.0651	0.0148	9.9938e-5
	IPDHA2	604	261.9419	0.0145	9.9963e-5
	RPDHA1	360	308.6289	0.0151	9.9847e-5
	RPDHA2	410	217.8890	0.0149	9.9931e-5

Table 1: Comparative results for deblurring corrupted images by different algorithms.

Table 1 reports some numerical comparison results for deblurring *man*, *salt-pepper* and *earth* images by different algorithms, and Figures 2-4 show the corresponding blurred-noisy images(**BNI**) and deblurred-recovery images(**DRI**), respectively. In Figure 5, we also plot the curves of the iterative residual error IT_res with respect to iteration numbers by different algorithms. We can observe from Table 1 that

- No matter which image is deblurred, the algorithm IPDHA1 (i.e., Algorithm 1.1) could obtain a better quality image than others since the value RE defined in (4.11) is of the smallest, although it costs more CPU time than some of comparison algorithms.

- Although the residual error RE obtained by RPDHA1-RPDHA2 is slightly larger than that of IPDHA1-IPDHA2, the former two algorithms can significantly improve the convergence speed than the latter in terms of the iteration number and the CPU time costed, which verifies the theoretical analysis in Section 3.

- If the iteration number is considered preferentially, then the algorithm RPDHA1 performs the best and CP underperforms RPDHA1. If the CPU time is considered preferentially, then the algorithm RPDHA2 outperforms the others.

Reported results in Table 1 together with the recovery images shown in Figures 2-4 and convergence curves depicted in Figure 5 support our theoretical improvements and verify the numerical performance of our modified algorithms to some extent.

Figure 3: From left to right and up to down: *salt-pepper* image with medium Gaussian blur, restored images by algorithms PD, CP, PDHG, IPDHA1, IPDHA2, RPDHA2 and RPDHA1, respectively.
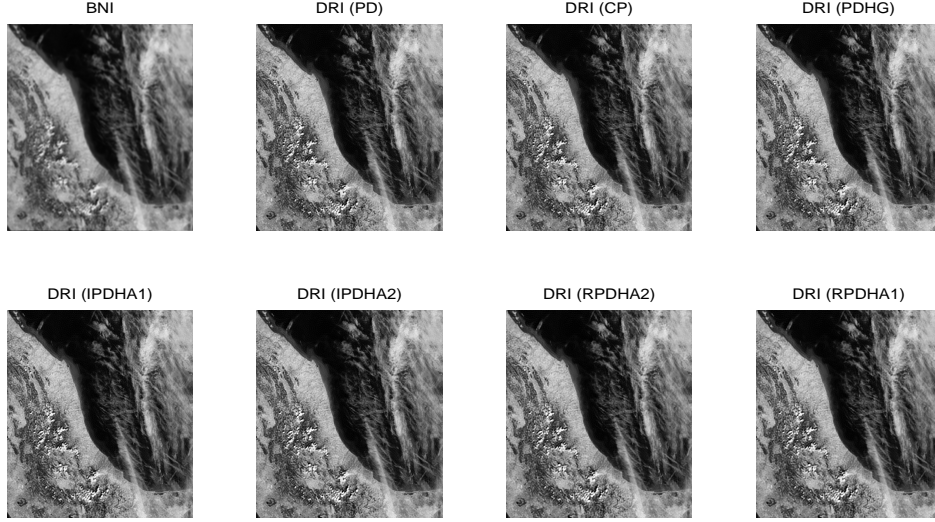
Figure 4: From left to right and up to down: *earth* image with medium Gaussian blur, restored images by algorithms PD, CP, PDHG, IPDHA1, IPDHA2, RPDHA2 and RPDHA1, respectively.
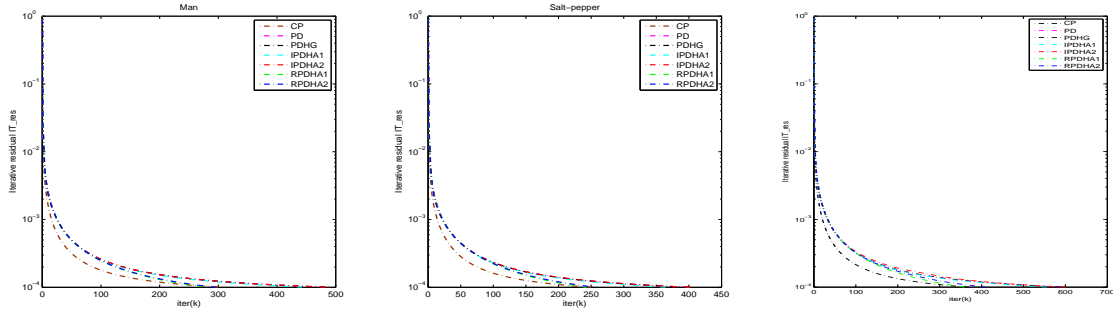

Figure 5: Evaluation of the iterative residual IT_res with respect to iterations for *man*(left), *salt-pepper*(middle) and *earth* (right) images with medium Gaussian blur.

5. Conclusion

This paper presents four variants of the primal-dual hybrid gradient algorithm for solving a family of saddle-point problems, where our algorithms are designed with some altered metrics in the dual step and with relaxation steps to accelerate the convergence. By the aid of a unified framework of variational inequality, we characterize the solution point of the problem and analyze the global convergence and sublinear convergence rate of the proposed algorithms in the ergodic/nonergodic sense. Numerical results about the image deblurring problem on testing *man*, *salt-pepper* and *earth* images of different dimensions illustrate the efficiency of the proposed algorithms to some extent.

Note that in Algorithms 3.1-3.2 we only explain the general matrix D should be nonsingular (it is not necessary to calculate its inverse in the updates of the algorithms) and give a remark to illustrate two easily used cases, so a better choice of the matrix D and parameter ϑ is worth studying in the further work. Besides, the convergence of these two algorithms are analyzed under Assumption 3.1. Whether they are convergent or not in the case that the matrices $\bar{H}, \bar{G}$ are not always positive

definite needs further investigations, since convergence of the generalized symmetric ADMM [1] was showed for multi-block separable convex programming but without using positive definiteness assumption on the involved matrices.

Acknowledgments

References

[1] J.C. BAI, J.C. LI, F.M. XU AND H.C. ZHANG, *Generalized symmetric ADMM for separable convex optimization*, Comput. Optim. Appl., 70 (2018), pp. 129–170.

[2] J.C. BAI, H.C. ZHANG AND J.C. LI, *A parameterized proximal point algorithm for separable convex optimization*, Optim. Lett., 12 (2018), pp. 1589–1608.

[3] A. CHAMBOLLE, M.J. EHRHARDT, P. RICHTARIK AND C.B. SCHONLIEB, *Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications*, SIAM J. Optim., 28 (2018), pp. 2783–2808.

[4] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145.

[5] Y.M. CHEN, G.H. LAN AND Y.Y. OUYANG, *Optimal primal-dual methods for a class of saddle point problems*, SIAM J. Optim., 24 (2014), pp. 1779–1814.

[6] E. ESSER, X.Q. ZHANG AND T.F. CHAN, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci., 4(2010), pp. 1015–1046.

[7] G.Y. GU, B.S. HE AND X.M. YUAN, *Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach*, Comput. Optim. Appl., 59 (2014), pp. 135–161.

[8] B.S. HE, X.M. YUAN AND W.X. ZHANG, *A customized proximal point algorithm for convex minimization with linear constraints*, Comput. Optim. Appl., 56 (2013), pp. 559–572.

[9] B.S. HE, Y.F. YOU AND X.M. YUAN, *On the convergence of primal-dual hybrid gradient algorithm*, SIAM J. Imaging Sci., 7 (2014), pp. 2526–2537.

[10] R. SHEFI AND M. TEBOULLE, *Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization*, SIAM J. Optim., 24 (2014), pp. 269–297.

[11] Y.L. WANG, J.F. YANG, W.T. YIN AND Y. ZHANG, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1 (2008), pp. 248–272.

[12] Z.P. XIE, *A primal-dual method with linear mapping for a saddle point problem in image deblurring*, J. Vis. Commun. Image R., 42 (2017), pp. 112–120.

[13] M. ZHU AND T.F. CHAN, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, CAM Report 08-34, UCLA, Los Angeles, CA. 2008.

[14] B.X. ZHANG, Z.B. ZHU AND S. WANG, *A simple primal-dual method for total variation image restoration*, J. Vis. Commun. Image R., 38 (2016), pp. 814-823.