GOLDEN RATIO PRIMAL-DUAL ALGORITHM WITH LINESEARCH*

XIAO-KAI CHANG[†], JUNFENG YANG[‡], AND HONGCHAO ZHANG[§]

Abstract. The golden ratio primal-dual algorithm (GRPDA) is a new variant of the classical Arrow-Hurwicz method for solving structured convex optimization problems, in which the objective function consists of the sum of two closed proper convex functions, one of which involves a composition with a linear transform. The same as the Arrow-Hurwicz method and the popular primal-dual algorithm (PDA) of Chambolle and Pock, GRPDA is full-splitting in the sense that it does not rely on solving any subproblems or linear system of equations iteratively. Compared with PDA, an important feature of GRPDA is that it permits larger primal and dual stepsizes. However, the stepsize condition of GRPDA requires that the spectral norm of the linear transform is known, which can be difficult to obtain in some applications. Furthermore, constant stepsizes are usually overconservative in practice. In this paper, we propose a linesearch strategy for GRPDA, which not only does not require the spectral norm of the linear transform but also allows adaptive and potentially much larger stepsizes. Within each linesearch step, only the dual variable needs to be updated, and it is thus quite cheap and does not require any extra matrix-vector multiplications for many special yet important applications such as a regularized least-squares problem. Global iterate convergence and $\mathcal{O}(1/N)$ ergodic convergence rate results, measured by the function value gap and constraint violations of an equivalent optimization problem, are established, where N denotes the iteration counter. When one of the component functions is strongly convex, faster $O(1/N^2)$ ergodic convergence rate results, quantified by the same measures, are established by adaptively choosing some algorithmic parameters. Moreover, when the subdifferential operators of the component functions are strongly metric subregular, a condition that is much weaker than strong convexity, we show that the iterates converge R-linearly to the unique solution. Numerical experiments on matrix game and LASSO problems illustrate the effectiveness of the proposed linesearch strategy.

Key words. saddle point problem, golden ratio primal-dual algorithm, linesearch, acceleration, ergodic sublinear convergence, linear convergence

MSC codes. 49M29, 65K10, 65Y20, 90C25

DOI. 10.1137/21M1420319

SIAM J. OPTIM. Vol. 32, No. 3, pp. 1584-1613

1. Introduction. Let \mathbb{R}^p and \mathbb{R}^q be finite-dimensional Euclidean spaces, each endowed with an inner product and the induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, respectively. Let $f : \mathbb{R}^p \to (-\infty, +\infty]$ and $g : \mathbb{R}^q \to (-\infty, +\infty]$ be extended real-valued closed proper convex functions, and let $K \in \mathbb{R}^{p \times q}$ be a linear transform from \mathbb{R}^q to \mathbb{R}^p . Denote the Legendre–Fenchel conjugate of f by f^* , i.e., $f^*(y) = \sup_{x \in \mathbb{R}^p} \{\langle y, x \rangle - f(x) \}, y \in \mathbb{R}^p$. In this paper, we focus on the following saddle point

^{*}Received by the editors May 18, 2021; accepted for publication (in revised form) March 14, 2022; published electronically July 18, 2022.

https://doi.org/10.1137/21M1420319

Funding: The first author was supported by the National Natural Science Foundation of China (NSFC-12161053) and the Natural Science Foundation of Gansu Province (21JR7RA253). The second author was supported by the National Natural Science Foundation of China (NSFC-11922111) and Ministry of Science and Technology of China (2020YFA0713800). The third author was supported by the U.S. National Science Foundation under grant DMS-1819161 and DMS-2110722.

[†]School of Science, Lanzhou University of Technology, Lanzhou, Gansu, 730050, People's Republic of China (xkchang@lut.cn).

[‡]Corresponding author. Department of Mathematics, Nanjing University, Nanjing, 210093, People's Republic of China (jfyang@nju.edu.cn, http://maths.nju.edu.cn/~jfyang/).

[§]Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803-4918 USA (hozhang@math.lsu.edu, http://www.math.lsu.edu/~hozhang/).

1585

problem with a bilinear coupling term:

(1)
$$\min_{x} \max_{y} \{ g(x) + \langle Kx, y \rangle - f^{*}(y) \mid x \in \mathbb{R}^{q}, y \in \mathbb{R}^{p} \}.$$

Since the biconjugate of f is itself, i.e., $(f^*)^* = f$ (see [29]), problem (1) reduces to the following primal minimization problem:

(2)
$$\min_{x} \{g(x) + f(Kx) \mid x \in \mathbb{R}^{q}\}.$$

On the other hand, by swapping the "min" and the "max" and using the definition of conjugate function, problem (1) can be transformed to the following dual maximization problem:

3)
$$\max_{y} \{ -f^*(y) - g^*(-K^{\top}y) \mid y \in \mathbb{R}^p \},\$$

where K^{\top} denotes the matrix transpose or adjoint operator of K. Under regularity conditions, e.g., Assumption 2.1 given below, strong duality holds between (2) and (3).

Problems (1)–(3) naturally arise from abundant interesting applications, including signal and image processing, machine learning, statistics, mechanics and economics, and so on; see, e.g., [4, 5, 7, 17, 36] and the references therein. To solve (1)–(3) simultaneously, popular choices include the well-known alternating direction method of multipliers (ADMM) [14, 15], the primal-dual algorithm (PDA) of Chambolle and Pock [7, 19, 28], and their accelerated and generalized variants [24, 26]. The focus of this paper is primal-dual type full-splitting algorithms.¹ We emphasize that the literature on numerical algorithms for solving (1)–(3) has become fairly vast and a thorough overview is not only impossible but also far beyond the focus of this work. Instead, we review only some primal-dual type algorithms that are most closely related to this work. For a thorough treatment of various primal-dual type fullsplitting algorithms, we refer interested readers to the recent monograph [30, Chapter 3]. Before going into details, we define our notation.

1.1. Notation. As already mentioned above, the transpose operation of a matrix or a vector is denoted by superscript " \top ." The spectral norm of K is denoted by L, i.e., $L := ||K|| = \sup\{||Kx|| : ||x|| = 1, x \in \mathbb{R}^q\}$. Let h be any extended real-valued closed proper convex function defined on a finite-dimensional Euclidean space \mathbb{R}^m . The effective domain of h is denoted by $\operatorname{dom}(h) := \{x \in \mathbb{R}^m : h(x) < +\infty\}$, and the subdifferential of h at $x \in \mathbb{R}^m$ is denoted by $\partial h(x) := \{\xi \in \mathbb{R}^m : h(y) \ge h(x) + \langle \xi, y - x \rangle$ for all $y \in \mathbb{R}^m\}$. Furthermore, for $\lambda > 0$, the proximal operator of λh is given by

$$\operatorname{Prox}_{\lambda h}(x) := \arg\min_{y \in \mathbb{R}^m} \left\{ h(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}, \quad x \in \mathbb{R}^m,$$

which is uniquely well defined everywhere. The relative interior of C is denoted by $\operatorname{ri}(C)$. Finally, throughout this paper, we denote the golden ratio by ϕ , i.e., $\phi = \frac{\sqrt{5}+1}{2}$, which is a key parameter in golden ratio type algorithms. Other notation will be specified in the context.

¹An algorithm for solving (1)–(3) is said to be full-splitting if it does not rely on solving any subproblems or linear system of equations iteratively and the main computations per iteration are matrix-vector multiplications and the evaluations of proximal operators.

1586 XIAO-KAI CHANG, JUNFENG YANG, AND HONGCHAO ZHANG

1.2. Related works. A main feature of primal-dual algorithms is that problems (1)-(3) are solved simultaneously by alternatingly updating the primal and the dual variables. Among others, the classical augmented Lagrangian method and its variants such as ADMM [14, 15, 20, 23] are most popular. However, ADMM is not full-splitting since at each iteration it requires solving a subproblem of the form $\min_{x \in \mathbb{R}^q} \frac{1}{2} ||Kx - b_n||^2 + g(x)$, where $b_n \in \mathbb{R}^p$ varies with the iteration counter n. Note that even if the proximal operator of g is easy to evaluate, this subproblem needs to be solved iteratively, unless K is the identity operator. On the other hand, for regularized least-squares problem, ADMM requires solving a linear system of equations at each iteration, which could be prohibitively expensive for large scale applications.

The most classical and simple primal-dual full-splitting algorithm goes back to [33], which is nowadays widely known as the Arrow-Hurwicz method. Started at $x_0 \in \mathbb{R}^q$ and $y_0 \in \mathbb{R}^p$, the Arrow-Hurwicz method iterates for $n \geq 1$ as

(4)
$$\begin{cases} x_n = \operatorname{Prox}_{\tau g}(x_{n-1} - \tau K^\top y_{n-1}), \\ y_n = \operatorname{Prox}_{\sigma f^*}(y_{n-1} + \sigma K x_n), \end{cases}$$

where $\tau, \sigma > 0$ are stepsize parameters. The heuristics of the Arrow-Hurwicz method is to solve the minimax problem (1) by alternatingly minimizing with x, maximizing with y and meanwhile incorporating the proximal technique by taking into account the latest information. Convergence of the Arrow-Hurwicz method with small stepsizes was studied in [13], and a sublinear convergence rate result was obtained in [7, 27] when dom(f^*) is bounded. However, the Arrow-Hurwicz method does not converge, in general. In fact, a divergent example has been constructed in [18]. Nonetheless, this method has been popular in image processing community and is known as the primal-dual hybrid gradient method [7, 13, 37].

To obtain a convergent full-splitting algorithm under a more general setting, Chambolle and Pock [7, 8] modified (4) by adopting an extrapolation step. Specifically, x_n is replaced by the extrapolated point $z_n = x_n + \delta(x_n - x_{n-1})$ in the computation of y_n in (4), where $\delta \in (0, 1]$ is a parameter, resulting in the following iterative scheme:

(5)
$$\begin{cases} x_n = \operatorname{Prox}_{\tau g}(x_{n-1} - \tau K^{\top} y_{n-1}), \\ z_n = x_n + \delta(x_n - x_{n-1}), \\ y_n = \operatorname{Prox}_{\sigma f^*}(y_{n-1} + \sigma K z_n). \end{cases}$$

We refer to (5) as PDA. When $\delta = 1$, convergence of PDA was established in [7] under the condition $\tau \sigma L^2 < 1$. Later, it was shown in [19] that PDA is an application of a weighted proximal point algorithm to solve an equivalent mixed variational inequality (MVI) problem of the optimality condition of (1). Furthermore, PDA is also referred to as the split inexact Uzawa method in [13], where the connection with preconditioned or linearized ADMM has been revealed; see [7, 31]. When g and f^* are piecewise linear-quadratic or their subdifferential operators satisfy certain metric subregularity conditions, linear convergence results were obtained in [21]. Note that, without taking a correction step as in [19], convergence of PDA with $\delta \in (0, 1)$ remains unclear. Overrelaxed, inertial, accelerated, and stochastic variants of PDA were investigated in, e.g., [6, 8].

Recently, Malitsky proposed a golden ratio algorithm and an adaptive variant of it, denoted, respectively, by GRAAL (defined in [25, eq. 10]) and aGRAAL (defined in [25, Algorithm 1]), for solving the MVI problem. It was shown that these algorithms converge under more relaxed conditions determined by the golden ratio, which explains their names. Since the optimality condition of (1) can be represented by the MVI problem, Malitsky's algorithms can be applied readily. Unfortunately, numerical experiments show that straightforward application of GRAAL and aGRAAL to the MVI representation of (1) is much less efficient than primal-dual type methods, e.g., the PDA scheme (5), which are able to take advantage of problem structures thoroughly. While this paper was under review, a first-order primal-dual algorithm with adaptive choice of stepsizes was proposed in [35] to solve the bilinear saddle-point problem (1) by using an idea similar to [25]. However, the local smoothness condition on the function q is required there, which is different from our setting. Moreover, the algorithm still needs to know the spectral norm of K, while we do not in this work. Motivated by [25], we proposed a golden ratio primal-dual algorithm (GRPDA) in [9], which inherits the advantages of both golden ratio and primal-dual type algorithms. Instead of an extrapolation step as taken in PDA, a convex combination of essentially all the previously generated primal iterates are used in the current iteration of GR-PDA. Specifically, given $x_0 \in \mathbb{R}^q$, $y_0 \in \mathbb{R}^p$, and letting $z_0 := x_0$, GRPDA iterates for n > 1 as

(6)
$$\begin{cases} z_n = \frac{\psi - 1}{\psi} x_{n-1} + \frac{1}{\psi} z_{n-1}, \\ x_n = \operatorname{Prox}_{\tau g}(z_n - \tau K^\top y_{n-1}), \\ y_n = \operatorname{Prox}_{\sigma f^*}(y_{n-1} + \sigma K x_n). \end{cases}$$

Global iterative convergence and ergodic convergence rate results are established under the condition $\tau \sigma L^2 < \psi$. Since $\psi \in (1, \phi]$ and $\phi = \frac{\sqrt{5}+1}{2}$, this requirement is much more relaxed than that for PDA, which is $\tau \sigma L^2 < 1$. Hence, GRPDA is not only able to fully exploit problem structures but also permits larger primal and dual stepsizes, which are critical for fast practical convergence.

1.3. Motivations and contributions. In this paper, we incorporate linesearch into GRPDA scheme (6). Our primary motivations have two aspects. First, in many applications, especially when K is large and dense, e.g., CT image reconstruction [2, 32], the exact spectral norm of K can be very expensive to compute or estimate. Second, even if the spectral norm of K can be obtained, the stepsizes governed by the condition $\tau \sigma ||K||^2 < \psi$ are usually too conservative for fast practical convergence. Hence, our goal in this paper is to adapt linesearch into (6) to accelerate the algorithm significantly while theoretically still guaranteeing convergence with desirable convergence rates.

Linesearch strategies are often of fundamentally practical importance for both unconstrained and constrained optimization. They are also commonly used in solving structured nonsmooth convex optimization problems, often combined with certain smoothing and homotopy techniques. For example, an algorithm is presented in [34] for solving the composite model (2) and its variants, and a nonmonotone linesearch algorithm is given in [16] for solving the ℓ_1 -regularized least-squares problem. Moreover, linesearch strategy has been adapted into the PDA in [26]. In general, linesearch requires extra evaluations of the proximal operators and matrix-vector multiplications at every iteration. Interestingly, for many special yet important applications as pointed out in [26], the proximal operator of f^* is extremely simple, and as a consequence the linesearch procedure does not require any extra matrix-vector multiplications.

Our work is an adaptation of linesearch into GRPDA, which deviates from [26] mainly in three aspects. First, the convergence rate results obtained in this paper are different from those in [26]. Specifically, those presented in [26] are measured

by the so-called primal-dual gap function proposed in [7], while ours are measured by the function value gap and constraint violations of a constrained optimization problem equivalent to (2). Although frequently used in the literature to quantify convergence rates of primal-dual type algorithms (see, e.g., [7, 8, 9, 26]), the primaldual gap function has a major flaw, that is, it can vanish at nonstationary points and thus make the existing results measured by this function convey little information. In contrast, the measures adopted in this paper are conventional and very meaningful for constrained optimization. Second, GRPDA is quite different from PDA in the sense that it remains unclear how to analyze GRPDA from the fixed point perspective (see [25] for similar remarks on GRAAL), while PDA fits well into the framework of fixed point iterations of averaged operators; see, e.g., [10, 19]. Finally, our experimental results show that GRPDA in combination with the linesearch given in this paper could significantly improve its numerical performance compared with its PDA counterpart.

Our main contributions of this paper include the design of GRPDAs with linesearch under different settings and their convergence analysis. As will be seen in later sections, we provide very novel approaches to analyze the linesearch behaviors, and our theoretical analysis on the stepsize behaviors is fundamentally different from those presented in [26] or given in all other literature. Moreover, our algorithm combining with linesearch not only does not assume any priori knowledge about the spectral norm of K, but also generates adaptive and potentially much larger stepsizes. When both the component functions f and g are generally convex, we establish global iterate convergence of GRPDA as well as ergodic $\mathcal{O}(1/N)$ sublinear convergence rate in terms of function value gap and constraint violations of an equivalent constrained optimization. When either one of the component functions is strongly convex, GR-PDA with linesearch is shown to converge at the faster $\mathcal{O}(1/N^2)$ ergodic sublinear rate. Furthermore, if the subdifferential operators of both component functions are strongly metric subregular, a notion that is much weaker than strong convexity, nonergodic linear convergence results to the unique solution are established. Hence, even with the stepsize relaxations by the proposed linesearch, the global convergence as well as theoretical convergence rates are still guaranteed to remain consistent with their counterparts without using linesearch. Moreover, numerical experiments show that many practical benefits can be obtained from the proposed linesearch strategies which could be critical in important applications.

1.4. Organization. The rest of this paper is organized as follows. In section 2, we make our assumptions and present some useful facts and further notation. Section 3 is devoted to GRPDA with linesearch in the general convex case, followed in section 4 by the case when either g or f^* is strongly convex. The case when both ∂g and ∂f^* are strongly metric subregular is studied in section 5. Numerical results on minimax matrix game and LASSO problems are reported in section 6 to show the benefits gained by adopting the linesearch strategies. Finally, some concluding remarks are drawn in section 7.

2. Assumptions and preliminaries.

2.1. Assumptions and further notation. For convenience of analysis, we introduce an auxiliary variable $w \in \mathbb{R}^p$ and rewrite the primal problem (2), equivalently, as

(7)
$$\min_{x,w} \{g(x) + f(w) \mid Kx - w = 0, x \in \mathbb{R}^q, w \in \mathbb{R}^p\}.$$

Let $y \in \mathbb{R}^p$ be the Lagrange multiplier. The objective and the Lagrangian functions of (7) are denoted, respectively, by

(8)
$$\Phi(x,w) := g(x) + f(w) \text{ and } \mathcal{L}(x,w,y) := \Phi(x,w) + \langle y, Kx - w \rangle.$$

Throughout the paper, we make the following blanket assumptions.

ASSUMPTION 2.1. Assume that the set of solutions of (2), and hence (7), is nonempty and, in addition, there exists $\tilde{x} \in ri(dom(g))$ such that $K\tilde{x} \in ri(dom(f))$.

Under Assumption 2.1, it follows from [29, Corollaries 28.2.2 and 28.3.1] that $(x^*, w^*) \in \mathbb{R}^q \times \mathbb{R}^p$ is a solution of (7) if and only if there exists an optimal solution $y^* \in \mathbb{R}^p$ to the dual problem (3) such that (x^*, w^*, y^*) is a saddle point of $\mathcal{L}(x, w, y)$, i.e.,

$$\mathcal{L}(x^{\star}, w^{\star}, y) \leq \mathcal{L}(x^{\star}, w^{\star}, y^{\star}) \leq \mathcal{L}(x, w, y^{\star}) \text{ for all } (x, w, y) \in \mathbb{R}^{q} \times \mathbb{R}^{p} \times \mathbb{R}^{p}.$$

As such, (x^*, y^*) is a solution of the minimax problem (1) and y^* is a solution of the dual problem (3). We denote the set of saddle points of $\mathcal{L}(x, w, y)$ by Ω , which is nonempty under Assumption 2.1 and is given by

$$\Omega = \{ (x^{\star}, w^{\star}, y^{\star}) \in \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \mid -K^{\top}y^{\star} \in \partial g(x^{\star}), \ y^{\star} \in \partial f(w^{\star}), \ Kx^{\star} = w^{\star} \}.$$

In addition, we make the following assumptions on the proximal operators of f and g, which are widely satisfied in many practical applications; see, e.g., [3, Chapter 6].

ASSUMPTION 2.2. Assume that the proximal operators of the component functions f and g either have closed form formulas or can be evaluated efficiently.

Note that the proximal operators of f^* and g^* are also easily computable under Assumption 2.2 due to the Moreau decomposition theorem [29, Theorem 31.5].

2.2. Facts and identities. The following simple facts and identities are useful in our analysis.

FACT 2.1. Let $h : \mathbb{R}^m \to (-\infty, +\infty]$ be an extended real-valued closed proper and γ -strongly convex function with modulus $\gamma \ge 0$. Then for any $\tau > 0$ and $x \in \mathbb{R}^m$, it holds that $z = \operatorname{Prox}_{\tau h}(x)$ if and only if $h(y) \ge h(z) + \frac{1}{\tau} \langle x - z, y - z \rangle + \frac{\gamma}{2} ||y - z||^2$ for all $y \in \mathbb{R}^m$.

FACT 2.2. Let $\{a_n : n \ge 1\}$ and $\{b_n : n \ge 1\}$ be real and nonnegative sequences. If $a_{n+1} \le a_n - b_n$ for all $n \ge 1$, then $\lim_{n\to\infty} a_n$ exists and $\lim_{n\to\infty} b_n = 0$.

For any $x, y, z \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$, there hold

(9)
$$2\langle x - y, x - z \rangle = \|x - y\|^2 + \|x - z\|^2 - \|y - z\|^2,$$

(10)
$$\|\alpha x + (1-\alpha)y\|^2 = \alpha \|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2.$$

2.3. Metric subregularity. Metric subregularity is a property of set-valued operators. In the case of subdifferential operators of closed proper convex functions, it is equivalent to the quadratic growth condition [1, 11]. Specifically, for a closed proper convex function $h : \mathbb{R}^m \to (-\infty, \infty]$, the subdifferential operator ∂h is metrically subregular at x^* for y^* with $(x^*, y^*) \in \operatorname{gra}(\partial h) := \{(u, v) : v \in \partial h(u)\}$ if and only if there exist a constant c > 0 and a neighborhood U of x^* such that the following growth condition holds:

$$h(x) \ge h(x^*) + \langle y^*, x - x^* \rangle + cd^2(x, (\partial h)^{-1}(y^*)) \text{ for all } x \in U,$$

where d(x, X) denotes certain distance from x to the set X. Furthermore, ∂h is strongly subregular at x^* for y^* with $(x^*, y^*) \in \operatorname{gra}(\partial h)$ if and only if there exist a constant c > 0 and a neighborhood U of x^* such that

(11)
$$h(x) \ge h(x^*) + \langle y^*, x - x^* \rangle + c \|x - x^*\|^2 \text{ for all } x \in U.$$

Note that if h is strongly convex, then it satisfies (11). However, the contrary is not true. In fact, (11) is much weaker than strong convexity as it is a local condition and requires being held at y^* only in a neighborhood of x^* . Moreover, ∂h is globally (strongly) subregular at x^* for y^* if (strong) subregularity holds with $U = \mathbb{R}^m$.

3. General convex case. Recall that $\phi = (\sqrt{5}+1)/2$ denotes the golden ratio. In this section, we introduce a linesearch strategy into GRPDA to choose stepsizes adaptively. Within each linesearch step, only the dual variable needs to be updated. The resulting algorithm, called GRPDA-L, is summarized in Algorithm 3.1.

Algorithm 3.1 (GRPDA-L).

Step 0. Choose $x_0 = z_0 \in \mathbb{R}^q$, $y_0 \in \mathbb{R}^p$, $\psi \in (1, \phi)$, $\sigma \in (0, 1)$, $\beta > 0$, $\mu \in (0, 1)$, and $\tau_0 > 0$. Set $\varphi = \frac{1+\psi}{\psi^2}$ and n = 1.

Step 1. Compute

(

12)
$$z_n = \frac{\psi - 1}{\psi} x_{n-1} + \frac{1}{\psi} z_{n-1},$$

(13)
$$x_n = \operatorname{Prox}_{\tau_{n-1}g}(z_n - \tau_{n-1}K^{\top}y_{n-1}).$$

Step 2. Let $\tau = \varphi \tau_{n-1}$ and compute

(14)
$$y_n = \operatorname{Prox}_{\beta\tau_n f^*}(y_{n-1} + \beta\tau_n K x_n),$$

where $\tau_n = \tau \mu^i$ and *i* is the smallest nonnegative integer such that

15)
$$\sqrt{\beta \tau_n} \| K^\top y_n - K^\top y_{n-1} \| \le \sigma \sqrt{\psi/\tau_{n-1}} \| y_n - y_{n-1} \|.$$

Step 3. Set $n \leftarrow n+1$ and return to Step 1.

Since $\psi \in (1, \phi)$, we have $\varphi = (1 + \psi)/\psi^2 \in (1, 2)$. Thus, the initial trial of τ_n in (14) is strictly greater than τ_{n-1} . By using Moreau's decomposition $y = \operatorname{Prox}_{f/\sigma}(y) + \frac{1}{\sigma}\operatorname{Prox}_{\sigma f^*}(\sigma y)$, for any $\sigma > 0$ and $y \in \mathbb{R}^p$, y_n defined in (14) can be updated as

(16)
$$y_n = y_{n-1} + \beta \tau_n (Kx_n - w_n)$$
 with $w_n = \operatorname{Prox}_{f/(\beta \tau_n)} (y_{n-1}/(\beta \tau_n) + Kx_n)$.

From Step 2 of Algorithm 3.1, the linesearch procedure may require computing $\operatorname{Prox}_{\beta\tau_n f^*}$ and $K^{\top}y_n$ repeatedly to find a proper τ_n at each iteration. However, as pointed out in [25, Remark 2], this procedure becomes extremely simple when $\operatorname{Prox}_{\lambda f^*}$ is linear or affine. Some examples are listed below:

- (a) $\operatorname{Prox}_{\lambda f^*}(u) = u \lambda b$ when $f^*(y) = \langle b, y \rangle$ for some $b \in \mathbb{R}^p$;
- (b) $\operatorname{Prox}_{\lambda f^*}(u) = \frac{1}{1+\lambda}(u+\lambda b)$ when $f^*(y) = \frac{1}{2}||y-b||^2$ for some $b \in \mathbb{R}^p$;
- (c) $\operatorname{Prox}_{\lambda f^*}(u) = u + \frac{b \langle u, a \rangle}{\|a\|^2} a$ when f^* is the indicator function of $H = \{u : \langle a, u \rangle = b\}$ for some $a \in \mathbb{R}^p$ and $b \in \mathbb{R}$.

In all these cases, the evaluation of $\operatorname{Prox}_{\lambda f^*}$ is very simple, and it is unnecessary to compute $K^{\top}y_n$ repeatedly since it can be obtained by combining some already computed quantities. Therefore, in these cases the linesearch step is quite cheap and does not require any additional matrix-vector multiplications. Furthermore, if necessary, one can always exchange the roles of the primal and the dual variables in (1) to take advantage of the above-mentioned structure. Also note that the ratio τ_n/τ_{n-1} is upper bounded by $\varphi \in (1, 2)$. As suggested in [25], one default choice could let $\psi = 1.5$ so that $\varphi = (1 + \psi)/\psi^2 = 10/9$. The parameter $\beta > 0$ in Algorithm 3.1 is introduced to scale the primal and the dual variables so that they will converge in a weighted balance way. Similar settings have also been done in [9] for GRPDA without linesearch.

The following lemma shows that the linesearch step of Algorithm 3.1 is well defined. In addition, it establishes some important properties on $\{\tau_n : n \ge 1\}$ and $\{\delta_n : n \ge 1\}$, with $\delta_n := \tau_n/\tau_{n-1}$, which are essential for establishing the convergence results. Since its proof is rather technical, for fluency of the overall paper, we put the proof in Appendix A.

LEMMA 3.1. Let $\underline{\tau} := \frac{\sigma\sqrt{\psi}}{L\sqrt{\beta\varphi}} > 0$. Then, we have the following properties. (i) The linesearch step of Algorithm 3.1, i.e., Step 2, always terminates. (ii) For any $\rho \in (0,1)$, there exists an infinite subsequence $\{n_k : k \ge 1\} \subseteq \{1,2,\ldots\}$ such that $\tau_{n_k} \ge \underline{\tau}$ and $\delta_{n_k} \ge \rho$. (iii) For any integer N > 0, we have $|\mathcal{K}_N| \ge \hat{c}N$ for some constant $\hat{c} > 0$, where $\mathcal{K}_N := \{1 \le n \le N : \tau_n \ge \underline{\tau} \text{ and } \delta_n \ge 1/\varphi\}$ and $|\mathcal{K}_N|$ is the cardinality of \mathcal{K}_N , which implies $\sum_{n=1}^N \tau_n \ge \underline{c}N$ with $\underline{c} = \hat{c}\underline{\tau}$.

We emphasize that the linesearch procedure adopted by Algorithm 3.1 is motivated but theoretically very different from that of [26, Algorithim 1]. In fact, the sequence $\{\tau_n : n \ge 1\}$ generated by [26, Algorithim 1] is uniformly bounded below by some positive constant; see [26, Lemma 3.3 (ii)]. In contrast, as seen in (iii) of Lemma 3.1, only a subsequence $\{\tau_{n_k} : k \ge 1\}$ is guaranteed to have uniform lower bound $\underline{\tau} > 0$. Similar arguments also apply to Algorithm 4.1 in section 4.

In the rest of this section, without mentioning repeatedly, we always fix arbitrary a primal and dual solution triplet $(x^*, w^*, y^*) \in \Omega$, and let $\{(z_n, x_n, y_n) : n \ge 1\}$ be generated by Algorithm 3.1 and $\{w_n : n \ge 1\}$ be given in (16). Furthermore, for $(x, w, y) \in \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p$, we define

(17)
$$J(x,w,y) := \mathcal{L}(x,w,y) - \mathcal{L}(x^{\star},w^{\star},y) = \Phi(x,w) + \langle y, Kx - w \rangle - \Phi(x^{\star},w^{\star}),$$

where $\Phi(x, w)$, $\mathcal{L}(x, w, y)$, and Ω are defined in (8) and Assumption 2.1. Next, we present two useful lemmas, which play critical roles in the convergence analysis.

LEMMA 3.2. For any $y \in \mathbb{R}^p$, there holds

(

$$\tau_n J(x_n, w_n, y) \le \langle x_{n+1} - z_{n+1}, x^* - x_{n+1} \rangle + \frac{1}{\beta} \langle y_n - y_{n-1}, y - y_n \rangle + \psi \delta_n \langle x_n - z_{n+1}, x_{n+1} - x_n \rangle + \tau_n \langle K^\top(y_n - y_{n-1}), x_n - x_{n+1} \rangle.$$

Proof. It follows from (13), (16), and Fact 2.1 that

(19)

$$\begin{aligned} \tau_n \big(g(x_{n+1}) - g(x^*) \big) &\leq \big\langle x_{n+1} - z_{n+1} + \tau_n K^\top y_n, \ x^* - x_{n+1} \big\rangle \\ &= \big\langle x_{n+1} - z_{n+1} + \tau_n K^\top \big(y_{n-1} + \beta \tau_n (Kx_n - w_n) \big), \\ x^* - x_{n+1} \big\rangle, \end{aligned}$$

(20) $\tau_{n-1}(g(x_n) - g(x_{n+1})) \leq \langle x_n - z_n + \tau_{n-1}K^{\top}y_{n-1}, x_{n+1} - x_n \rangle,$

(21)
$$\tau_n \big(f(w_n) - f(w^\star) \big) \le -\tau_n \big\langle y_{n-1} + \beta \tau_n (Kx_n - w_n), w^\star - w_n \big\rangle.$$

Multiplying (20) by $\delta_n = \tau_n/\tau_{n-1}$ and using $x_n - z_n = \psi(x_n - z_{n+1})$, which follows from (12), we obtain

(22)
$$\tau_n (g(x_n) - g(x_{n+1})) \leq \langle \psi \delta_n (x_n - z_{n+1}) + \tau_n K^\top y_{n-1}, x_{n+1} - x_n \rangle.$$

Recall that $y_n - y_{n-1} = \beta \tau_n (Kx_n - w_n)$ and $Kx^* = w^*$. By taking a sum to (19), (21)–(22), and adding $\tau_n \langle y, Kx_n - w_n \rangle$ to both sides, we obtain

$$\begin{aligned} &\tau_n(\mathcal{L}(x_n, w_n, y) - \mathcal{L}(x^*, w^*, y)) = \tau_n \big(g(x_n) + f(w_n) + \langle y, Kx_n - w_n \rangle - g(x^*) \\ &- f(x^*) \big) \leq \langle x_{n+1} - z_{n+1}, \ x^* - x_{n+1} \rangle + \psi \delta_n \ \langle x_n - z_{n+1}, \ x_{n+1} - x_n \rangle \\ &+ \tau_n \langle Kx_n - w_n, y - y_{n-1} \rangle \\ &+ \beta \tau_n^2 \langle Kx_n - w_n, (Kx^* - Kx_{n+1}) - (w^* - w_n) \rangle \\ &= \langle x_{n+1} - z_{n+1}, \ x^* - x_{n+1} \rangle + \psi \delta_n \ \langle x_n - z_{n+1}, \ x_{n+1} - x_n \rangle + \frac{1}{\beta} \langle y_n - y_{n-1}, y - y_{n-1} \rangle \\ &+ \tau_n \langle y_n - y_{n-1}, w_n - Kx_{n+1} \rangle \\ &= \langle x_{n+1} - z_{n+1}, \ x^* - x_{n+1} \rangle + \psi \delta_n \ \langle x_n - z_{n+1}, \ x_{n+1} - x_n \rangle + \frac{1}{\beta} \langle y_n - y_{n-1}, y - y_n \rangle \\ &+ \tau_n \langle K^\top (y_n - y_{n-1}), x_n - x_{n+1} \rangle, \end{aligned}$$

which, by the definition of $J(\cdot)$ in (17), implies (18) immediately.

For any
$$y \in \mathbb{R}^p$$
 and $n \ge 1$, define
(23)

$$\begin{cases}
a_n(x^*, y) := \frac{\psi}{\psi - 1} \|z_{n+1} - x^*\|^2 + \frac{1}{\beta} \|y_{n-1} - y\|^2, \\
b_n := \psi \delta_n \|z_{n+1} - x_n\|^2 + (1 - \sigma) (\psi \delta_n \|x_{n+1} - x_n\|^2 + \frac{1}{\beta} \|y_n - y_{n-1}\|^2)
\end{cases}$$

LEMMA 3.3. There holds $a_{n+1}(x^*, y) + 2\tau_n J(x_n, w_n, y) \leq a_n(x^*, y) - b_n$ for any $y \in \mathbb{R}^p$ and $n \geq 1$.

Proof. Fix $n \ge 1$. First, it is easy to verify from $\varphi = (1 + \psi)/\psi^2$ and $\delta_n =$ $\tau_n/\tau_{n-1} \leq \varphi$ that

(24)
$$1 + \frac{1}{\psi} - \psi \delta_n \ge 1 + \frac{1}{\psi} - \psi \varphi = 0.$$

It follows from (15) and the basic inequality $2ab \leq a^2 + b^2$ for any $a, b \in \mathbb{R}$ that

D

$$2\tau_n \|K^\top y_n - K^\top y_{n-1}\| \|x_{n+1} - x_n\| \le \sigma \bigg(\psi \delta_n \|x_{n+1} - x_n\|^2 + \frac{1}{\beta} \|y_n - y_{n-1}\|^2\bigg).$$

Furthermore, by Lemma 3.2, identity (9), and Cauchy–Schwarz inequality, we have

(26)
$$\|x_{n+1} - x^{\star}\|^{2} + \frac{1}{\beta} \|y_{n} - y\|^{2} + 2\tau_{n}J(x_{n}, w_{n}, y)$$

$$\leq \|z_{n+1} - x^{\star}\|^{2} + \frac{1}{\beta} \|y_{n-1} - y\|^{2} + 2\tau_{n} \|K^{\top}y_{n} - K^{\top}y_{n-1}\| \|x_{n+1} - x_{n}\|$$

$$- \psi\delta_{n} \|z_{n+1} - x_{n}\|^{2} - (1 - \psi\delta_{n}) \|x_{n+1} - z_{n+1}\|^{2} - \psi\delta_{n} \|x_{n+1} - x_{n}\|^{2}$$

$$- \frac{1}{\beta} \|y_{n} - y_{n-1}\|^{2}.$$

Since $x_{n+1} = \frac{\psi}{\psi-1}z_{n+2} - \frac{1}{\psi-1}z_{n+1}$, which again follows from (12), we deduce from (10) that

where the second equality is due to $z_{n+2} - z_{n+1} = \frac{\psi - 1}{\psi}(x_{n+1} - z_{n+1})$. By combining (27) with (26), we obtain

$$(28) \quad \frac{\psi}{\psi-1} \|z_{n+2} - x^{\star}\|^{2} + \frac{1}{\beta} \|y_{n} - y\|^{2} + 2\tau_{n}J(x_{n}, w_{n}, y)$$

$$\leq \frac{\psi}{\psi-1} \|z_{n+1} - x^{\star}\|^{2} + \frac{1}{\beta} \|y_{n-1} - y\|^{2} + 2\tau_{n} \|K^{\top}y_{n} - K^{\top}y_{n-1}\| \|x_{n+1} - x_{n}\|$$

$$-\psi\delta_{n} \|z_{n+1} - x_{n}\|^{2} - \left(1 + \frac{1}{\psi} - \psi\delta_{n}\right) \|x_{n+1} - z_{n+1}\|^{2} - \psi\delta_{n} \|x_{n+1} - x_{n}\|^{2}$$

$$-\frac{1}{\beta} \|y_{n} - y_{n-1}\|^{2}$$

$$\leq \frac{\psi}{\psi-1} \|z_{n+1} - x^{\star}\|^{2} + \frac{1}{\beta} \|y_{n-1} - y\|^{2} - \psi\delta_{n} \|z_{n+1} - x_{n}\|^{2}$$

$$-(1 - \sigma)\psi\delta_{n} \|x_{n+1} - x_{n}\|^{2} - \frac{1 - \sigma}{\beta} \|y_{n} - y_{n-1}\|^{2},$$

where the second " \leq " follows from (24) and (25). Finally, the conclusion of Lemma 3.3 follows immediately from (28) and the definitions of $a_n(x^*, y)$ and b_n in (23).

Now, we are ready to establish global iterate convergence and $\mathcal{O}(1/N)$ ergodic sublinear convergence rate of Algorithm 3.1.

THEOREM 3.1 (global convergence). The sequence $\{(x_n, y_n) : n \ge 1\}$ converges to a solution of (1).

Proof. Let $\rho \in (0, 1)$ and $\underline{\tau} > 0$ be defined in Lemma 3.1. By (ii) of Lemma 3.1, there exists an infinite sequence $\{n_k : k \ge 1\}$ such that $\tau_{n_k} \ge \underline{\tau}$ and $\delta_{n_k} \ge \rho$. Since $J(x_n, w_n, y^*) \ge 0$, it follows from Lemma 3.3 that $a_{n+1}(x^*, y^*) \le a_n(x^*, y^*) - b_n$. Hence, by the definition of b_n in (23), we have

(29)
$$\psi \sum_{k=1}^{\infty} \delta_{n_k} \left(\|z_{n_k+1} - x_{n_k}\|^2 + (1-\sigma) \|x_{n_k+1} - x_{n_k}\|^2 \right) \le \sum_{n=1}^{\infty} b_n < \infty.$$

We have that $\sum_{k=1}^{\infty} \delta_{n_k} = \infty$, which together with (29) implies that there exists a subsequence of $\{n_k : k \ge 1\}$, still denoted as $\{n_k : k \ge 1\}$, such that

(30)
$$\lim_{k \to \infty} \|z_{n_k+1} - x_{n_k}\| = 0 \quad \text{and} \quad \lim_{k \to \infty} \|x_{n_k+1} - x_{n_k}\| = 0.$$

Furthermore, Fact 2.2 implies that $\lim_{n\to\infty} a_n(x^*, y^*)$ exists and $\lim_{n\to\infty} b_n = 0$. Hence, both $\{z_n : n \ge 1\}$ and $\{y_n : n \ge 1\}$ are bounded, and $\lim_{n\to\infty} ||y_{n+1} - y_n|| = 0$. Then, (12) implies that $\{x_n : n \ge 1\}$ is also bounded. Therefore, there exist a subsequence of $\{n_k : k \ge 1\}$, still denoted as $\{n_k : k \ge 1\}$, and (x^*, y^*) such that $\lim_{k\to\infty} x_{n_k} = x^*$ and $\lim_{k\to\infty} y_{n_k} = y^*$, which together with (30) implies that $\lim_{k\to\infty} x_{n_k+1} = \lim_{k\to\infty} x_{n_k} = \lim_{k\to\infty} z_{n_k+1} = x^*$. Now, Fact 2.1 and (14) imply

(31)
$$\left\langle \frac{1}{\beta}(y_n - y_{n-1}) - \tau_n K x_n, y - y_n \right\rangle \ge \tau_n \left(f^*(y_n) - f^*(y) \right)$$
 for all $y \in \mathbb{R}^p$.

Similar to (19) and (31), for any (x, y), there hold

(32)
$$\begin{cases} \langle x_{n_k+1} - z_{n_k+1} + \tau_{n_k} K^\top y_{n_k}, \ x - x_{n_k+1} \rangle \ge \tau_{n_k} \big(g(x_{n_k+1}) - g(x) \big), \\ \langle \frac{1}{\beta} (y_{n_k} - y_{n_k-1}) - \tau_{n_k} K x_{n_k}, \ y - y_{n_k} \rangle \ge \tau_{n_k} \big(f^*(y_{n_k}) - f^*(y) \big). \end{cases}$$

Then, dividing $\tau_{n_k} \geq \underline{\tau} > 0$ from both sides of (32), taking into account that both g and f^* are closed (and thus lower semicontinuous) and letting $k \to \infty$, we obtain

(33)
$$\langle K^{\top}y^*, x - x^* \rangle \ge g(x^*) - g(x) \text{ and } - \langle Kx^*, y - y^* \rangle \ge f^*(y^*) - f^*(y).$$

Since (33) holds for any $(x, y) \in \mathbb{R}^q \times \mathbb{R}^p$, we have $-K^{\top}y^* \in \partial g(x^*)$ and $Kx^* \in \partial f^*(y^*)$, i.e., (x^*, y^*) is a solution of (1). Note that Lemma 3.3 holds for any x^* such that $(x^*, w^*, y^*) \in \Omega$ for some (w^*, y^*) . Therefore, one can replace x^* by x^* and meanwhile set $y = y^*$ in the definition of $a_n(x^*, y)$. As such, we have $\lim_{k\to\infty} a_{n_k}(x^*, y^*) = 0$ since $\lim_{k\to\infty} z_{n_k+1} = x^*$ and $\lim_{k\to\infty} y_{n_k-1} = \lim_{k\to\infty} y_{n_k} = y^*$. Since $\{a_n(x^*, y^*) : n \ge 1\}$ is monotonically nonincreasing, it follows that $\lim_{n\to\infty} a_n(x^*, y^*) = 0$. Therefore, $\lim_{n\to\infty} (z_n, y_n) = (x^*, y^*)$. Again by (12), we have $\lim_{n\to\infty} x_n = x^*$. This completes the proof.

THEOREM 3.2 (sublinear convergence rate). For any $N \ge 1$, define

(34)
$$\hat{x}_N = \frac{1}{s_N} \sum_{n=1}^N \tau_n x_n$$
 and $\hat{w}_N = \frac{1}{s_N} \sum_{n=1}^N \tau_n w_n$ with $s_N = \sum_{n=1}^N \tau_n$.

Then, there exists a constant $C_1 > 0$ such that for any $N \ge 1$ we have

$$|\Phi(\hat{x}_N, \hat{w}_N) - \Phi(x^*, w^*)| \le \frac{C_1}{N} \quad and \quad ||K\hat{x}_N - \hat{w}_N|| \le \frac{2C_1}{cN},$$

where c is a constant satisfying $c \geq 2 \|y^{\star}\|$.

Proof. Recall that $a_n(x^*, y)$ and b_n are nonnegative. For any $y \in \mathbb{R}^p$, Lemma 3.3 implies that $2\tau_n J(x_n, w_n, y) \leq a_n(x^*, y) - a_{n+1}(x^*, y)$, a sum of which over $n = 1, \ldots, N$ yields

(35)
$$2\sum_{n=1}^{N} \tau_n J(x_n, w_n, y) \le a_1(x^*, y) - a_{N+1}(x^*, y) \le a_1(x^*, y).$$

Since J(x, w, y) is convex in x and w, it follows from (34) and Jensen's inequality that

(36)
$$J(\hat{x}_N, \hat{w}_N, y) \le \frac{1}{s_N} \sum_{n=1}^N \tau_n J(x_n, w_n, y)$$

Combining (35), (36), and the definition of $J(\cdot)$ in (17), we obtain

(37)
$$\Phi(\hat{x}_N, \hat{w}_N) + \langle y, K\hat{x}_N - \hat{w}_N \rangle - \Phi(x^*, w^*) \le a_1(x^*, y)/(2s_N).$$

By property (iii) in Lemma 3.1, we have $s_N = \sum_{n=1}^N \tau_n \ge \underline{c}N$. By taking the maximum of both sides of (37) over $||y|| \le c$ and defining $C_1 = \sup_y \{a_1(x^*, y) : ||y|| \le c\}/(2\underline{c}) > 0$, we obtain

(38)
$$\Phi(\hat{x}_N, \hat{w}_N) + c \|K\hat{x}_N - \hat{w}_N\| - \Phi(x^\star, w^\star) \le C_1/N,$$

which implies $\Phi(\hat{x}_N, \hat{w}_N) - \Phi(x^*, w^*) \leq C_1/N$. Furthermore, since $\mathcal{L}(x^*, w^*, y^*) \leq \mathcal{L}(\hat{x}_N, \hat{w}_N, y^*), Kx^* = w^*$ and $\|y^*\| \leq c/2$, we have

(39)
$$\Phi(x^*, w^*) - \Phi(\hat{x}_N, \hat{w}_N) \le \langle y^*, K\hat{x}_N - \hat{w}_N \rangle \le (c/2) \|K\hat{x}_N - \hat{w}_N\|,$$

which together with (38) implies

$$c\|K\hat{x}_N - \hat{w}_N\| \le \Phi(x^*, w^*) - \Phi(\hat{x}_N, \hat{w}_N) + C_1/N \le (c/2)\|K\hat{x}_N - \hat{w}_N\| + C_1/N.$$

As a result, we derive $||K\hat{x}_N - \hat{w}_N|| \leq 2C_1/(cN)$. It then follows from (39) that $\Phi(x^*, w^*) - \Phi(\hat{x}_N, \hat{w}_N) \leq C_1/N$, and thus $|\Phi(\hat{x}_N, \hat{w}_N) - \Phi(x^*, w^*)| \leq C_1/N$. The proof is completed.

4. When one component function is strongly convex. Consider the case when either g or f^* is strongly convex. It was shown in [7] that one can choose the stepsizes and some parameters adaptively so that the resulting PDA achieves faster $\mathcal{O}(1/N^2)$ convergence rate. Similar results have been obtained in [26] for PDA with linesearch. In this section, we first propose an adaptive linesearch strategy for GRPDA and then carry out a convergence analysis. By a sign change to the saddle function, the minimax problem (1) is equivalent to

(40)
$$\max_{x} \min_{y} \{ f^{*}(y) + \langle -K^{\top}y, x \rangle - g(x) \mid x \in \mathbb{R}^{q}, y \in \mathbb{R}^{p} \};$$

by swapping " $\max_{x \in \mathbb{R}^q}$ " with " $\min_{y \in \mathbb{R}^p}$ " and (g, K, x, q) with $(f^*, -K^T, y, p)$, problem (40) is reducible to (1). Thus, we need only treat the case when g is strongly convex, while the case when f^* is strongly convex can be treated in the same way, which is thus omitted.

In the rest of this section, we assume that g is strongly convex with modulus $\gamma_g > 0$. The following algorithm is an adaptation of linesearch into GRPDA, which exploits strong convexity of g since the modulus γ_g plays a role in the choices of parameters.

ALGORITHM 4.1 (GRPDA with linesearch when g is γ_g -strongly convex).

Step 0. Let $\psi_0 = 1.3247, \ldots$ be the unique real root of $\psi^3 - \psi - 1 = 0$. Choose $\psi \in (\psi_0, \phi), \ \beta_0 > 0, \ \tau_0 > 0$, and $\mu \in (0, 1)$. Choose $x_0 = z_0 \in \mathbb{R}^q$ and $y_0 \in \mathbb{R}^p$. Set $\varphi = \frac{1+\psi}{\psi^2}$ and n = 1.

Step 1. Compute

(41)
$$z_n = \frac{\psi - 1}{\psi} x_{n-1} + \frac{1}{\psi} z_{n-1},$$

(42)
$$x_n = \operatorname{Prox}_{\tau_{n-1}g}(z_n - \tau_{n-1}K^{\top}y_{n-1}),$$

(43)
$$\omega_n = \frac{\psi - \varphi}{\psi + \varphi \gamma_g \tau_{n-1}},$$

(44)
$$\beta_n = \beta_{n-1} (1 + \gamma_g \omega_n \tau_{n-1}).$$

Step 2. Let $\tau = \varphi \tau_{n-1}$ and compute

$$y_n = \operatorname{Prox}_{\beta_n \tau_n f^*} (y_{n-1} + \beta_n \tau_n K x_n),$$

where $\tau_n = \tau \mu^i$ and *i* is the smallest nonnegative integer such that

(45)
$$\sqrt{\beta_n \tau_n} \| K^\top y_n - K^\top y_{n-1} \| \le \sqrt{\psi/\tau_{n-1}} \| y_n - y_{n-1} \|.$$

Step 3. Set $n \leftarrow n+1$ and return to Step 1.

Similar to (16), y_n defined in (4.1) can be rewritten as

(46)
$$y_n = y_{n-1} + \beta_n \tau_n (Kx_n - w_n)$$
 with $w_n = \operatorname{Prox}_{f/(\beta_n \tau_n)} (y_{n-1}/(\beta_n \tau_n) + Kx_n)$.

As in [9, Algorithm 4.1], ω_n plays a key role in establishing the accelerated rate. The condition $\psi > \psi_0$, where ψ_0 is the unique real root of $\psi^3 - \psi - 1 = 0$, ensures that $\psi > (1 + \psi)/\psi^2 = \varphi$. Therefore, we have $\omega_n > 0$ and $\beta_n > \beta_{n-1} > 0$ for all $n \ge 1$.

Next, we state a key lemma with respect to Algorithm 4.1, whose detailed proof is left to Appendix B since it similar in spirit to that of Lemma 3.1.

LEMMA 4.1. Let $\{(\tau_n, \beta_n) : n \ge 0\}$ be generated by Algorithm 4.1. Then, we have the following properties. (i) The linesearch step of Algorithm 4.1, i.e., Step 2, always terminates. (ii) There exists a constant c > 0 such that $\beta_n \ge cn^2$ for all $n \ge 1$. (iii) There exists a constant $\tilde{c} > 0$ such that $\sum_{n=1}^N \tau_n \le \tilde{c} \sum_{n \in S_N} \tau_n$ for all $N \ge 1$, where $S_N := \{1 \le n \le N : \sqrt{\beta_n} \tau_n \ge 1/L\}.$

We next establish the promised $O(1/N^2)$ ergodic convergence rate result. Since g is strongly convex, x^* is unique.

THEOREM 4.1 (global convergence and $\mathcal{O}(1/N^2)$ rate). Let $\{(z_n, x_n, y_n, \beta_n, \tau_n) : n \geq 1\}$ be generated by Algorithm 4.1, and let $\{w_n : n \geq 1\}$ be defined in (46). Then, the following holds:

(a) there exist $C_1, C_2 > 0$ such that $||z_{n+1} - x^*|| \le C_1/n$ and $||x_{n+1} - x^*|| \le C_2/n$ for all $n \ge 1$;

(b) the sequence $\{y_n : n \ge 1\}$ is bounded and there exists a subsequence of $\{y_n : n \ge 1\}$ converging to y^* such that (x^*, y^*) is a solution of (1);

(c) there exists a constant $C_3 > 0$ such that for any $N \ge 1$ we have

$$|\Phi(\hat{x}_N, \hat{w}_N) - \Phi(x^\star, w^\star)| \le \frac{C_3}{N^2} \quad and \quad ||K\hat{x}_N - \hat{w}_N|| \le \frac{2C_3}{\breve{c}N^2}$$

where $\breve{c} > 0$ is a constant satisfying $\breve{c} \geq 2 \|y^{\star}\|$ and

(47)
$$\hat{x}_N = \frac{1}{s_N} \sum_{n=1}^N \beta_n \tau_n x_n$$
 and $\hat{w}_N = \frac{1}{s_N} \sum_{n=1}^N \beta_n \tau_n w_n$ with $s_N = \sum_{n=1}^N \beta_n \tau_n x_n$

1597

Proof. Since g is γ_g -strongly convex, it follows from (42) and Fact 2.1 that

(48)
$$\langle x_n - z_n + \tau_{n-1} K^\top y_{n-1}, x - x_n \rangle$$

$$\geq \tau_{n-1} \left(g(x_n) - g(x) + \frac{\gamma_g}{2} \|x_n - x\|^2 \right) \text{ for all } x.$$

By passing n + 1 to n and x^* to x in (48), we obtain

(49)
$$\langle x_{n+1} - z_{n+1} + \tau_n K^\top y_n, x^* - x_{n+1} \rangle$$

$$\geq \tau_n \left(g(x_{n+1}) - g(x^*) + \frac{\gamma_g}{2} \|x_{n+1} - x^*\|^2 \right).$$

Similarly, by passing x_{n+1} to x in (48) and multiplying both sides by $\delta_n = \tau_n/\tau_{n-1}$, we obtain

$$\langle \delta_n(x_n - z_n) + \tau_n K^\top y_{n-1}, \ x_{n+1} - x_n \rangle \ge \tau_n \bigg(g(x_n) - g(x_{n+1}) + \frac{\gamma_g}{2} \|x_{n+1} - x_n\|^2 \bigg).$$

Similar to (21), it follows from (46) that

(51)
$$-\tau_n \langle y_{n-1} + \beta_n \tau_n (Kx_n - w_n), w^* - w_n \rangle \ge \tau_n \big(f(w_n) - f(w^*) \big).$$

By adding (49)–(51), noting $x_n - z_n = \psi(x_n - z_{n+1})$, and using similar arguments as in Lemma 3.2, we obtain for any $y \in \mathbb{R}^p$ that

$$\tau_n J(x_n, w_n, y) \leq \langle x_{n+1} - z_{n+1}, x^* - x_{n+1} \rangle + \frac{1}{\beta_n} \langle y_n - y_{n-1}, y - y_n \rangle z + \psi \delta_n \langle x_n - z_{n+1}, x_{n+1} - x_n \rangle + \tau_n \langle K^\top (y_n - y_{n-1}), x_n - x_{n+1} \rangle (52) \qquad - \frac{\gamma_g \tau_n}{2} (\|x_{n+1} - x^*\|^2 + \|x_{n+1} - x_n\|^2),$$

where $J(\cdot)$ is given in (17). By removing $-\frac{\gamma_g \tau_n}{2} ||x_{n+1} - x_n||^2 \leq 0$ in (52), using (9), (27), and Cauchy–Schwarz inequality, we obtain from direct calculations that

$$(1+\gamma_{g}\tau_{n})\frac{\psi}{\psi-1}\|z_{n+2}-x^{\star}\|^{2}+\frac{1}{\beta_{n}}\|y_{n}-y\|^{2}+2\tau_{n}J(x_{n},w_{n},y)$$

$$\leq \frac{\psi+\gamma_{g}\tau_{n}}{\psi-1}\|z_{n+1}-x^{\star}\|^{2}+\frac{1}{\beta_{n}}\|y_{n-1}-y\|^{2}+\left(\psi\delta_{n}-1-\frac{1+\gamma_{g}\tau_{n}}{\psi}\right)\|x_{n+1}-z_{n+1}\|^{2}$$

$$-\psi\delta_{n}\|z_{n+1}-x_{n}\|^{2}-\psi\delta_{n}\|x_{n+1}-x_{n}\|^{2}-\frac{1}{\beta_{n}}\|y_{n}-y_{n-1}\|^{2}$$

$$+2\tau_{n}\|K^{\top}y_{n}-K^{\top}y_{n-1}\|\|x_{n+1}-x_{n}\|.$$

Recall that $\delta_n = \tau_n / \tau_{n-1}$. Similar to (25), we have

$$2\tau_n \|K^\top y_n - K^\top y_{n-1}\| \|x_{n+1} - x_n\| \le \psi \delta_n \|x_{n+1} - x_n\|^2 + \frac{1}{\beta_n} \|y_n - y_{n-1}\|^2$$

Moreover, $\psi \delta_n - 1 - \frac{1 + \gamma_g \tau_n}{\psi} \le \psi \varphi - 1 - \frac{1}{\psi} - \frac{\gamma_g \tau_n}{\psi} = -\frac{\gamma_g \tau_n}{\psi}$ since $\delta_n \le \varphi$. Thus, (53)

implies

1598

(54)

$$(1 + \gamma_{g}\tau_{n})\frac{\psi}{\psi - 1} \|z_{n+2} - x^{\star}\|^{2} + \frac{1}{\beta_{n}}\|y_{n} - y\|^{2} + 2\tau_{n}J(x_{n}, w_{n}, y)$$

$$\leq \frac{\psi + \gamma_{g}\tau_{n}}{\psi - 1}\|z_{n+1} - x^{\star}\|^{2} + \frac{1}{\beta_{n}}\|y_{n-1} - y\|^{2}$$

$$-\frac{\gamma_{g}\tau_{n}}{\psi}\|x_{n+1} - z_{n+1}\|^{2}.$$

Note that $(1 + \gamma_g \tau_n) \frac{\psi}{\psi - 1} = \frac{\psi(1 + \gamma_g \tau_n)}{\psi + \gamma_g \tau_{n+1}} \frac{\psi + \gamma_g \tau_{n+1}}{\psi - 1}$. It follows from $\tau_{n+1} \leq \varphi \tau_n$ that

$$\frac{\psi(1+\gamma_g\tau_n)}{\psi+\gamma_g\tau_{n+1}} \ge \frac{\psi(1+\gamma_g\tau_n)}{\psi+\gamma_g\varphi\tau_n} = 1 + \frac{\psi-\varphi}{\psi+\gamma_g\varphi\tau_n}\gamma_g\tau_n = 1 + \omega_{n+1}\gamma_g\tau_n,$$

and thus

(55)
$$(1+\gamma_g\tau_n)\frac{\psi}{\psi-1} \ge (1+\omega_{n+1}\gamma_g\tau_n)\frac{\psi+\gamma_g\tau_{n+1}}{\psi-1} = \frac{\beta_{n+1}}{\beta_n}\frac{\psi+\gamma_g\tau_{n+1}}{\psi-1}.$$

Define $A_n(x^*, y) := \frac{\psi + \gamma_g \tau_n}{2(\psi - 1)} \|z_{n+1} - x^*\|^2 + \frac{1}{2\beta_n} \|y_{n-1} - y\|^2$. Combining (54) and (55), we deduce

(56)
$$\beta_{n+1}A_{n+1}(x^{\star}, y) + \beta_n \tau_n J(x_n, w_n, y) \le \beta_n A_n(x^{\star}, y) - \frac{\beta_n \gamma_g \tau_n}{2\psi} \|x_{n+1} - z_{n+1}\|^2.$$

By summing (56) over $n = 1, \ldots, N$, we obtain

$$\beta_{N+1}A_{N+1}(x^{\star},y) + \sum_{n=1}^{N} \beta_n \tau_n J(x_n, w_n, y) + \sum_{n=1}^{N} \frac{\beta_n \gamma_g \tau_n}{2\psi} \|x_{n+1} - z_{n+1}\|^2 \le \beta_1 A_1(x^{\star}, y).$$

Then, from the convexity of $J(\cdot)$ in (x, w), the nonnegativity of $J(x_n, w_n, y^*)$, the definition of $A_n(x^*, y)$, (47) and (57), we obtain

(58)
$$J(\hat{x}_N, \hat{w}_N, y) \le \frac{1}{s_N} \sum_{n=1}^N \beta_n \tau_n J(x_n, w_n, y) \le \frac{\beta_1 A_1(x^*, y)}{s_N}$$
 for all $y \in \mathbb{R}^p$,

(59)
$$||z_{N+2} - x^*||^2 \le \frac{2(\psi - 1)}{\psi + \gamma_g \tau_{N+1}} \frac{\beta_1 A_1(x^*, y^*)}{\beta_{N+1}} \le \frac{2\beta_1 A_1(x^*, y^*)}{\beta_{N+1}}.$$

Then, it is easy to derive from (ii) of Lemma 4.1 and (59) that $||z_{N+1} - x^*|| \leq C_1/N$ with $C_1 := \sqrt{2\beta_1 A_1(x^*, y^*)/c} > 0$ and thus $||x_{N+1} - x^*|| \leq C_2/N$ for some $C_2 > 0$, which follows from (41). Hence, property (a) holds.

Let $S = \{n \in \mathbb{Z}^+ : \sqrt{\beta_n}\tau_n \ge 1/L\}$. Properties (ii) and (iii) in Lemma 4.1 imply that $|S| = \infty$. We next show by contradiction that there exists a subsequence $\{n_k : k \ge 1\} \subseteq S$ such that

(60)
$$\lim_{k \to \infty} \|x_{n_k+1} - z_{n_k+1}\|/\tau_{n_k} = 0.$$

Let $N \ge 1$ be arbitrarily fixed and $\theta = (\psi - \varphi)\gamma_g/\psi > 0$. Lemma 4.1 (ii) and (44) imply

$$c(N+1)^2 \le \beta_{N+1} = \beta_N \left(1 + \frac{(\psi - \varphi)\gamma_g \tau_N}{\psi + \varphi \gamma_g \tau_N} \right) \le \beta_N \left(1 + \theta \tau_N \right) \le \beta_1 \prod_{n=1}^N \left(1 + \theta \tau_n \right).$$

(

1599

Then, by Lemma 4.1 (iii) we deduce

(61)
$$2\ln(N+1) - \ln(\beta_1/c) \le \sum_{n=1}^N \ln(1+\theta\tau_n) \le \theta \sum_{n=1}^N \tau_n \le \theta \tilde{c} \sum_{n \in \mathcal{S}_N} \tau_n.$$

On the other hand, it follows from (57) that

$$\frac{2\psi\beta_1 A_1(x^{\star}, y^{\star})}{\gamma_g} \ge \sum_{n \in \mathcal{S}} \beta_n \tau_n^3 \left(\frac{\|x_{n+1} - z_{n+1}\|}{\tau_n}\right)^2 \ge \frac{1}{L^2} \sum_{n \in \mathcal{S}} \tau_n \left(\frac{\|x_{n+1} - z_{n+1}\|}{\tau_n}\right)^2,$$

from which the existence of a subsequence $\{n_k : k \ge 1\} \subseteq S$ such that (60) holds is guaranteed since (61) indicates that $\sum_{n \in S} \tau_n = \lim_{N \to \infty} \sum_{n \in S_N} \tau_n = \infty$. Now, it follows from (57) that $\{\beta_n A_n : n \ge 1\}$ is bounded, which implies that

Now, it follows from (57) that $\{\beta_n A_n : n \ge 1\}$ is bounded, which implies that $\{y_n : n \ge 1\}$ is also bounded. Let $\{n_k : k \ge 1\} \subseteq S$ be the subsequence satisfying (60), which then has a subsequence, still denoted as $\{n_k : k \ge 1\} \subseteq S$, such that $\lim_{k\to\infty} y_{n_k} = y^*$. Similar to (32), for any $(x, y) \in \mathbb{R}^q \times \mathbb{R}^p$, we have

(62)
$$\begin{cases} \langle x_{n_k+1} - z_{n_k+1} + \tau_{n_k} K^\top y_{n_k}, \ x - x_{n_k+1} \rangle \ge \tau_{n_k} (g(x_{n_k+1}) - g(x)), \\ \langle \frac{1}{\beta_{n_k}} (y_{n_k} - y_{n_k-1}) - \tau_{n_k} K x_{n_k}, \ y - y_{n_k} \rangle \ge \tau_{n_k} (f^*(y_{n_k}) - f^*(y)). \end{cases}$$

Since $n_k \in S$, we have $\sqrt{\beta_{n_k}}\tau_{n_k} \geq 1/L$, which together with $\beta_n \geq cn^2$ for some c > 0 implies $\lim_{k\to\infty} \beta_{n_k}\tau_{n_k} = \infty$. Hence, $\lim_{k\to\infty} ||y_{n_k} - y_{n_k-1}||/(\beta_{n_k}\tau_{n_k}) = 0$ since $\{y_n : n \geq 1\}$ is bounded. Then, dividing τ_{n_k} from both sides of the two inequalities in (62), taking $k \to \infty$, it follows from $\lim_{n\to\infty} x_n = x^*$, $\lim_{k\to\infty} y_{n_k} = y^*$, lower semicontinuous of g and f^* , and (60) that (33) holds, which implies (x^*, y^*) is a saddle point of (1). Hence, property (b) holds.

Finally, it follows from (44) and $\omega_{n+1} < 1$ that $\beta_n \tau_n = (\beta_{n+1} - \beta_n)/(\omega_{n+1}\gamma_g) \ge (\beta_{n+1} - \beta_n)/\gamma_g$. Hence, we have from $\beta_n \ge cn^2$ that $s_N = \sum_{n=1}^N \beta_n \tau_n \ge (\beta_{N+1} - \beta_1)/\gamma_g \ge c_1 N^2$ for some $c_1 > 0$. Consequently, for any $y \in \mathbb{R}^p$, it follows from (58) that

$$J(\hat{x}_N, \hat{w}_N, y) = \Phi(\hat{x}_N, \hat{w}_N) + \langle y, K\hat{x}_N - \hat{w}_N \rangle - \Phi(x^*, w^*) \le \beta_1 A_1(x^*, y) / (c_1 N^2).$$

Taking the maximum of both sides over $||y|| \leq \check{c}$, we obtain

$$\Phi(\hat{x}_N, \hat{w}_N) + \breve{c} \| K \hat{x}_N - \hat{w}_N \| - \Phi(x^*, w^*) \le C_3 / N^2,$$

where $C_3 = (\beta_1/c_1) \sup_y \{A_1(x^*, y) : ||y|| \le \check{c}\} > 0$. Hence, by noting $\check{c} \ge 2||y^*||$ and following similar arguments as in Theorem 3.2, we can then show that property (c) holds.

5. Linear convergence under metric subregularity. In this section, we reconsider Algorithm 3.1 with $\sigma \in (0, 1]$ and establish its linear convergence rate under the assumption that the subdifferential operators ∂g and ∂f^* are globally and strongly metric subregular (see section 2.3). In fact, when $\sigma \in (0, 1)$ only locally strongly metric subregularity is needed. Specifically, we make the following assumption.

ASSUMPTION 5.1. Let $(x^*, w^*, y^*) \in \Omega$. Hence, $(x^*, -K^{\top}y^*) \in gra(\partial g)$ and $(y^*, Kx^*) \in gra(\partial f^*)$. Assume that ∂g and ∂f^* are globally and strongly metric subregular, respectively, at x^* for $-K^{\top}y^*$ and at y^* for Kx^* , i.e., there exist constants $\kappa_g > 0$ and $\kappa_f > 0$ such that

$$63) \qquad \begin{cases} g(x) \geq g(x^{\star}) - \langle K^{\top}y^{\star}, x - x^{\star} \rangle + \frac{\kappa_q}{2} \|x - x^{\star}\|^2 \text{ for all } x \in \mathbb{R}^q, \\ f^{\star}(y) \geq f^{\star}(y^{\star}) + \langle Kx^{\star}, y - y^{\star} \rangle + \frac{\kappa_f}{2} \|y - y^{\star}\|^2 \text{ for all } y \in \mathbb{R}^p. \end{cases}$$

The conditions in (63) are known as quadratic growth conditions, which are equivalent to the metric subregularity of ∂g and ∂f^* ; see [1, 11]. Under Assumption 5.1, the primal-dual solution pair (x^*, y^*) is unique. We emphasize that κ_g and κ_f do not need to be known since Algorithm 3.1 does not rely on them, and their existence will be used in the analysis solely. Linear convergence results under metric subregularity conditions have been achieved in [21] for a triangularly preconditioned PDA and in [22] for PDA, which was analyzed within the framework of fixed-point iterations of averaged operators. In contrast, how to analyze golden ratio type algorithms within the framework of fixed-point iterations remains unclear. In the following, we provide more direct analysis for establishing linear convergence which does not rely on the theory of fixed-point iterations.

Let $\{\tau_n : n \geq 0\}$ be the sequence generated by Algorithm 3.1, and let $\delta_n = \tau_n/\tau_{n-1}$. For $n \geq 1$, define $\theta_n := \min\{1/\chi(\tau_n, \delta_n), 1 + \beta \kappa_f \tau_{n-1}\}$, where, for $\tau > 0$ and $0 < \delta \leq \varphi$, $\chi(\tau, \delta)$ is defined by

$$\chi(\tau, \delta) := 1 - (\psi - 1) \min\{k_q, \psi\} \min\{\tau, \delta\} / (2\psi)$$

Note that $\psi > 1$. Hence, without loss of generality, we can assume $\kappa_g > 0$ is sufficiently small such that $\chi(\tau, \delta) \in (0, 1)$ for any $\tau > 0$ and $0 < \delta \leq \varphi$. We have the following key lemma.

LEMMA 5.1. Consider Algorithm 3.1 with $\sigma \in (0, 1]$. Then, we have the following properties. (i) The linesearch step of Algorithm 3.1, i.e., Step 2, always terminates. (ii) Under Assumption 5.1, there exist constants $\theta > 1$ and c > 0 such that $\Gamma_{N+1} := \prod_{n=2}^{N+1} \theta_n \ge c\theta^N$ for any integer $N \ge 1$.

Proof. The conclusion (i) with $\sigma \in (0, 1)$ has already been shown in Lemma 3.1 (i). Similarly, the conclusion (i) with $\sigma = 1$ can be also established. We now prove conclusion (ii). By following almost exactly the same proof, we can show that Lemma 3.1 (iii) holds for any $\sigma \in (0, 1]$, i.e., for any integer $N \ge 1$, we have $|\mathcal{K}_N| \ge \hat{c}N$ for some constant $\hat{c} > 0$, where $\mathcal{K}_N = \{1 \le n \le N : \tau_n \ge \underline{\tau} \text{ and } \delta_n \ge 1/\varphi\}$. Since $\chi(\tau_n, \delta_n) \in (0, 1)$, we have $\theta_n > 1$ for all $n \ge 1$. Furthermore, for any $n \in \mathcal{K}_N$, we have $\tau_n \ge \underline{\tau}$ and $\delta_n \ge 1/\varphi$, and thus $\tau_{n-1} \ge \tau_n/\varphi \ge \underline{\tau}/\varphi$ since $\tau_n \le \varphi \tau_{n-1}$. Note that $\chi(\tau, \delta)$ is nonincreasing with respect to both τ and δ . Hence, for any $n \in \mathcal{K}_N$, we have

$$\theta_n \ge \min\{1/\chi(\underline{\tau}, 1/\varphi), 1 + \beta \kappa_f \underline{\tau}/\varphi\} =: \theta > 1.$$

Therefore, for any $N \ge 1$, it follows from $|\mathcal{K}_N| \ge \hat{c}N$ that

$$\Gamma_{N+1} = \frac{1}{\theta_1} \prod_{n=1}^{N+1} \theta_n \ge \frac{1}{\theta_1} \prod_{n \in \mathcal{K}_N} \theta_n \ge \frac{\tilde{\theta}^{|\mathcal{K}_N|}}{\theta_1} \ge \frac{\tilde{\theta}^{\hat{c}N}}{\theta_1} = c\theta^N,$$

where $\theta := \tilde{\theta}^{\hat{c}} > 1$ and $c := 1/\theta_1 > 0$. This completes the proof.

For $x \in \mathbb{R}^q$ and $y \in \mathbb{R}^p$, define

(64)
$$\begin{cases} P(x) := g(x) - g(x^{\star}) + \langle K^{\top}y^{\star}, x - x^{\star} \rangle, \\ D(y) := f^{\star}(y) - f^{\star}(y^{\star}) - \langle Kx^{\star}, y - y^{\star} \rangle, \\ H(x, y) := P(x) + D(y) - \frac{\kappa_g}{2} ||x - x^{\star}||^2 - \frac{\kappa_f}{2} ||y - y^{\star}||^2. \end{cases}$$

Under Assumption 5.1, we have $H(x, y) \ge 0$ for any $x \in \mathbb{R}^q$ and $y \in \mathbb{R}^p$. Based on Lemma 5.1, we can establish the following nonergodic R-linear convergence results. Some proof details are omitted due to the similarity to the previous arguments in section 3.

THEOREM 5.1 (R-linear convergence). Let $\{(z_n, x_n, y_n) : n \ge 1\}$ be the sequence generated by Algorithm 3.1 with $\sigma \in (0, 1]$. Then, under Assumption 5.1, there exist constants $C_1, C_2 > 0$ and $\theta > 1$ such that $||z_n - x^*|| \le C_1/\theta^n$ and $||y_n - y^*|| \le C_2/\theta^n$ for all $n \ge 1$, which implies that $\{(z_n, y_n) : n \ge 1\}$ converges to (x^*, y^*) R-linearly.

Proof. Following the proof of Lemma 3.2, one can easily show that

$$\begin{aligned} \tau_n H(x_n, y_n) &\leq \langle x_{n+1} - z_{n+1}, x^* - x_{n+1} \rangle + \frac{1}{\beta} \langle y_n - y_{n-1}, y^* - y_n \rangle \\ &+ \psi \delta_n \langle x_n - z_{n+1}, x_{n+1} - x_n \rangle \\ &+ \tau_n \langle K^\top (y_n - y_{n-1}), x_n - x_{n+1} \rangle - \frac{\kappa_g \tau_n}{2} \| x_n - x^* \|^2 - \frac{\kappa_f \tau_n}{2} \| y_n - y^* \|^2, \end{aligned}$$

where $H(\cdot)$ is defined in (64). Note that $H(x_n, y_n) \ge 0$ under Assumption 5.1. Hence, by the same arguments as in (25)–(27), we can derive

$$\kappa_{g}\tau_{n}\|x_{n} - x^{\star}\|^{2} + \psi\delta_{n}\|z_{n+1} - x_{n}\|^{2} + \frac{\psi}{\psi - 1}\|z_{n+2} - x^{\star}\|^{2} + \left(\frac{1}{\beta} + \kappa_{f}\tau_{n}\right)\|y_{n} - y^{\star}\|^{2}$$

$$\leq \frac{\psi}{\psi - 1}\|z_{n+1} - x^{\star}\|^{2} + \frac{1}{\beta}\|y_{n-1} - y^{\star}\|^{2}.$$

It is easy to see that $\kappa_g \tau_n \|x_n - x^\star\|^2 + \psi \delta_n \|z_{n+1} - x_n\|^2 \ge \frac{1}{2} \min(\kappa_g, \psi) \min(\tau_n, \delta_n) \|z_{n+1} - x^\star\|^2$. Furthermore, from the definition of $\chi(\tau, \delta)$, we have $\frac{\psi}{\psi-1} - \frac{1}{2} \min(\kappa_g, \psi) \min(\tau_n, \delta_n) = \chi(\tau_n, \delta_n) \frac{\psi}{\psi-1}$. Therefore, it follows from (65) that

$$\begin{aligned} \frac{\psi}{\psi-1} \|z_{n+2} - x^{\star}\|^2 + \left(\frac{1}{\beta} + \kappa_f \tau_n\right) \|y_n - y^{\star}\|^2 &\leq \chi(\tau_n, \delta_n) \frac{\psi}{\psi-1} \|z_{n+1} - x^{\star}\|^2 \\ &+ \frac{1}{\beta} \|y_{n-1} - y^{\star}\|^2, \end{aligned}$$

from which we deduce $\theta_{n+1}A_{n+1} \leq A_n$ with $\theta_{n+1} = \min\{1/\chi(\tau_{n+1}, \delta_{n+1}), 1 + \beta\kappa_f\tau_n\}$ and

(66)
$$A_n := \chi(\tau_n, \delta_n) \frac{\psi}{\psi - 1} \|z_{n+1} - x^\star\|^2 + \frac{1}{\beta} \|y_{n-1} - y^\star\|^2.$$

Using $\theta_{n+1}A_{n+1} \leq A_n$ inductively for n = 1, ..., N and taking into account Lemma 5.1 (ii), we obtain $A_{N+1} \leq A_1/\Gamma_{N+1} \leq (A_1/c)/\theta^N$. Furthermore, by the definition of $\chi(\tau, \delta), \, \delta_n \leq \varphi$ for all n and $\varphi = (\psi + 1)/\psi^2$, we see that

$$\chi(\tau_{N+1}, \delta_{N+1}) \ge 1 - (\psi - 1)\psi\varphi/(2\psi) = (\psi^2 + 1)/(2\psi^2) > 1/2$$

Then, the claims of this theorem follow immediately from the definition of A_n in (66).

Finally, we note that for $\sigma < 1$ the convergence of $\{(x_n, y_n) : n \ge 1\}$ generated by Algorithm 3.1 to a solution of (1) has already been established in Theorem 3.1. Hence, as mentioned previously, in this case the quadratic growth conditions given in (63) need only be held locally for showing the R-linear convergence of Algorithm 3.1.

6. Numerical results. In this section, we present numerical results to demonstrate the performance of the proposed Algorithms 3.1 and 4.1, which are denoted by GRPDA-L and AGRPDA-L, respectively. We set $\sigma = 0.99$ for GRPDA-L and $\psi = 1.5$ and $\mu = 0.7$ for both algorithms. First, choose y_{-1} arbitrarily in a small

neighborhood of the starting point y_0 such that $K^{\top}(y_{-1} - y_0) \neq 0$ and then compute $\xi = ||y_{-1} - y_0||/||K^{\top}(y_{-1} - y_0)|| \geq 1/L$. Then, set $\tau_0 = \xi \sqrt{\psi/\beta} \geq \sqrt{\psi/\beta}/L$ for GRPDA-L and $\tau_0 = \xi \sqrt{\psi/\beta_0} \geq \sqrt{\psi/\beta_0}/L$ with some $\beta_0 > 0$ for AGRPDA-L. We compare the proposed algorithms with their corresponding counterparts without linesearch, i.e., GRPDA [9, Algorithm 3.1] with $\tau = \sqrt{\psi/\beta}/L$ and $\psi = 1.618$, and the state-of-the-art PDA with linesearch, i.e., PDA-L [26, Algorithm 1], with $\mu = 0.7$, $\delta = 0.99$, and $\tau_0 = ||y_{-1} - y_0||/(\sqrt{\beta}||K^{\top}(y_{-1} - y_0)||)$. Other parameters will be specified in the following.

All experiments were performed within Python 3.8 on an Intel Core i5-4590 CPU 3.30GHz PC with 8GB of RAM running on 64-bit Windows operating system. For reproducible purpose, the codes are provided at https://github.com/cxk9369010/GRPDA-Linesearch. We solve the minimax matrix game problem and the LASSO problem for comparison.

PROBLEM 6.1 (minimax matrix game). Let $\Delta_m = \{x \in \mathbb{R}^m : \sum_i x_i = 1, x \ge 0\}$ be the standard unit simplex in \mathbb{R}^m . The minimax matrix game problem is given by $\min_{x \in \Delta_q} \max_{y \in \Delta_p} \langle Kx, y \rangle$, where $K \in \mathbb{R}^{p \times q}$.

Clearly, Problem 6.1 is a special case of (1) with $g = \iota_{\Delta_q}$ and $f^* = \iota_{\Delta_p}$, where ι_C denotes the indicator function of a set C. Since neither g nor f^* is strongly convex, only the nonaccelerated algorithms GRPDA, GRPDA-L, and PDA-L are relevant here. For a feasible pair $(x, y) \in \Delta_q \times \Delta_p$, it is elementary to verify that the gap between the primal function value of (2) and the dual function value of (3) is given by (see also [8, sect. 7.1])

$$G(x,y) := (g(x) + f(Kx)) - (-f^*(y) - g^*(-K^{\top}y)) = \max_{1 \le i \le p} (Kx)_i - \min_{1 \le j \le q} (K^{\top}y)_j.$$

Initial points for all the algorithms are set to be $x_0 = \frac{1}{q}(1, \ldots, 1)^{\top} \in \mathbb{R}^q$ and $y_0 = \frac{1}{p}(1, \ldots, 1)^{\top} \in \mathbb{R}^p$. The projection onto the unit simplex is computed by the algorithm from [12]. We set $\psi = 1.618$ and $\tau = \sigma = 1/||K||$ for GRPDA, and $\beta = 1$ for PDA-L and GRPDA-L. As in [26], we generated $K \in \mathbb{R}^{p \times q}$ randomly in four different ways with random number generator *seed* = 50:

- (i) All entries of K were generated independently from the uniform distribution in [-1, 1], and (p, q) = (100, 100).
- (ii) All entries of K were generated independently from the normal distribution $\mathcal{N}(0,1)$, and (p,q) = (100, 100).
- (iii) All entries of K were generated independently from the normal distribution $\mathcal{N}(0, 10)$, and (p, q) = (500, 100).
- (iv) The matrix K is sparse with 10% nonzero elements generated independently from the uniform distribution in [0, 1], and (p, q) = (1000, 2000).

For a given $\epsilon > 0$, we terminate the algorithms when $G(x_n, y_n) < \epsilon$ or $n = n_{\max}$, where n_{\max} is the maximum number of iterations allowed. In this section, we set $n_{\max} = 3 \times 10^5$ and examine how the values of the primal-dual function value gap $G(x_n, y_n)$ decrease against CPU time. Table 1 presents the total CPU time (Time, in seconds), the number of iterations (Iter), and the number of extra linesearch trial steps (#LS) of PDA-L and GRPDA-L as compared with their counterparts without linesearch. We emphasize that for each trial of linesearch a projection onto the unit simplex is required for this example. The decreasing behavior of the primal-dual function value gap (abbreviated as PD gap) versus CPU time is shown in Figure 1 for the compared algorithms with $\epsilon = 10^{-10}$. The details of linesearch are illustrated in Figure 2. TABLE 1

Results on Problem 6.1. In the table, "—" represents that the algorithm reached the maximum number of iterations without satisfying the stopping condition.

ε	Test	GRPDA			PDA-L		GRPDA-L			
		Iter	Time	Iter	#LS	Time	Iter	#LS	Time	
10 ⁻⁷	(i)	25688	3.7	18816	18582	4.2	12944	3824	2.1	
	(ii)	103788	14.6	41486	41014	9.2	31631	9345	6.8	
	(iii)			80040	79708	31.4	63815	18853	21.4	
	(iv)			56119	54803	115.7	30345	8958	60.5	
10 ⁻¹⁰	(i)	151134	20.2	58879	58161	15.3	47564	14050	9.2	
	(ii)	245612	32.5	90415	89387	22.4	74378	21971	14.2	
	(iii)			161868	161182	67.3	144064	42558	48.2	
	(iv)		—		—	—	—	—	—	



FIG. 1. Comparison results of GRPDA, GRPDA-L, PDA-L, and aGRAAL on Problem 6.1: primal-dual function value gap (vertical axes) versus CPU time (horizontal axes).

It can be seen from Table 1 that PDA-L requires approximately one extra linesearch trial step per outer iteration, while GRPDA-L requires roughly one extra linesearch trial step per three outer iterations. As a result, GRPDA-L consumed less CPU time than PDA-L. From Figure 1, it is clear that GRPDA-L performs the best for all four tests, followed by PDA-L, and both are faster than GRPDA. Test (iv) is a difficult case, for which all the compared algorithms fail to reduce the primal-dual function value gap to be less than 10^{-10} within the prescribed maximum number of iterations. Figure 2 (a) shows that PDA-L takes 0–4 linesearch trials per iteration for case (iii), while GRPDA-L takes 0–5 per iteration and on average GRPDA-L takes much less extra linesearch steps than PDA-L (see the cumulative results in Figure 2 (b)). Note that here for illustration purposes, we only present results of the first



FIG. 2. Details of linesearch steps for case (iii) of Problem 6.1. Left (a): comparison of extra linesearch trial steps taken by PDA-L and GRPDA-L. Right (b): cumulative results for both algorithms.

200 iterations of case (iii). For other cases, similar remarks can be said, e.g., for case (i) PDA-L takes 0–2 linesearch trials per iteration, while GRPDA-L takes 0–3 per iteration, and the total number of linesearch steps taken by GRPDA-L is again much less than that of PDA-L.

PROBLEM 6.2 (LASSO). Let $K \in \mathbb{R}^{p \times q}$ be a sensing matrix, and let $b \in \mathbb{R}^p$ be an observation vector. One form of the LASSO problem is to recover a sparse signal via solving

(67)
$$\min_{x} F(x) := \eta \|x\|_{1} + \frac{1}{2} \|Kx - b\|^{2},$$

where $\eta > 0$ is a regularization parameter.

It is easy to verify that the LASSO problem (67) can be represented as the saddle point problem (1) with $g(x) = \eta \|x\|_1$ and $f^*(y) = \frac{1}{2} \|y\|^2 + \langle b, y \rangle$. Thus, the proximal operator $\operatorname{Prox}_{\tau f^*}(\cdot)$ is linear, and there is no extra matrix-vector multiplications needed within a linesearch step for GRPDA-L as $K^{\top}y_n$ can always be obtained via a convex combination of $K^{\top}y_{n-1}$ and $K^{\top}(Kx_n - b)$. On the other hand, problem (1) is equivalent to (40). Then, by swapping $\max_{x \in \mathbb{R}^q}$ with $\min_{y \in \mathbb{R}^p}$ and (g, K, x, q) with $(f^*, -K^T, y, p)$, the strong convexity of $\frac{1}{2} \|y\|^2 + \langle b, y \rangle$ (previously f^*) can be transferred to g, which enables the application of the accelerated version, i.e., AGRPDA-L. Therefore, the algorithms to compare in this experiment are GRPDA-L, AGRPDA-L, and PDA-L. GRPDA without linesearch will not be compared in this case since it is less efficient.

We set seed = 100 and generate a random vector $x^* \in \mathbb{R}^q$ for which s random coordinates are drawn from $\mathcal{N}(0, 1)$ and the rest are set to be zero. Then, we generate $\omega \in \mathbb{R}^p$ with entries drawn from $\mathcal{N}(0, 0.1)$ and set $b = Kx^* + \omega$. The matrix $K \in \mathbb{R}^{p \times q}$ is constructed in the following ways:

- (i) All entries of K are generated independently from $\mathcal{N}(0,1)$. The s entries of x^* are drawn from the uniform distribution in [-10, 10].
- (ii) First, we generate a matrix $A \in \mathbb{R}^{p \times q}$, whose entries are independently drawn from $\mathcal{N}(0, 1)$. Then, for a scalar $v \in (0, 1)$ we construct the matrix K column by column as follows: $K_1 = A_1/\sqrt{1-v^2}$ and $K_j = vK_{j-1} + A_j$, $j = 2, \ldots, q$. Here K_j and A_j represent the *j*th column of K and A, respectively. As $v \in (0, 1)$ becomes larger, K becomes more ill-conditioned. In this experiment we take v = 0.5 and v = 0.9, respectively. The sparse vector x^* is generated

in the same way as in case (i).

In both cases, the regularization parameter η was set to be 0.1. Similar to [26], we set $\beta = 400$ for PDA-L and GRPDA-L. For AGRPDA-L, we set $\gamma = 0.01$ and $\beta_0 = 1$ as in [9, 26]. The initial points for all algorithms are $x_0 = (0, \ldots, 0)^{\top}$ and $y_0 = Kx_0 - b$.

In this experiment, we first ran all the algorithms by a sufficiently large number of iterations and then chose the minimum attainable function value as an approximation of the optimal value F^* of (67). Again, for a given $\epsilon > 0$, we terminate the algorithms when $F(x_n) - F^* < \epsilon$ or $n = n_{\text{max}}$. In this experiment, we set $\epsilon = 10^{-12}$ and $n_{\text{max}} = 8 \times 10^4$ to examine their convergence behavior. The comparison results on the number of iterations, the number of extra linesearch steps, and CPU time are given in Table 2. The evolution of function value residuals $F(x_n) - F^*$ versus CPU time is given in Figure 3.

TABLE 2 Comparison results on the LASSO problem (67).

ϵ	Test	PDA-L			GRPDA-L			AGRPDA-L		
		Iter	#LS	Time	Iter	#LS	Time	Iter	#LS	Time
10^{-8}	(i)	4693	4619	13.7	4292	1251	11.5	2422	718	9.2
	(ii) $v = 0.5$	6208	6161	20.3	5017	1465	16.5	1754	518	7.1
	(ii) $v = 0.9$	27915	27902	92.2	26679	7869	83.3	7465	2208	32.5
10^{-12}	(i)	11177	11020	32.8	9734	2859	26.1	3512	1038	13.2
	(ii) $v = 0.5$	14621	14527	48.1	12145	3573	38.1	3124	925	12.9
	(ii) $v = 0.9$	62928	62917	211.5	56059	16547	180.1	12209	3609	53.2

It can be seen from the results in Table 2 that a similar conclusion can be drawn, i.e., PDA-L requires approximately one extra linesearch trial step per outer iteration, while GRPDA-L and AGRPDA-L require roughly one extra linesearch trial step per three outer iterations. Since the proximal operator $\operatorname{Prox}_{\tau f^*}(\cdot)$ is linear and does not incur extra computations, GRPDA-L and PDA-L perform similarly, although GRPDA-L performs slightly better, in terms of outer iteration and CPU time. In comparison, AGRPDA-L, which takes advantage of strong convexity, is much faster than both GRPDA-L and PDA-L. The detailed convergence behavior of the three algorithms is given in Figure 3. From these results, we see that strong convexity of the component functions, if properly explored, could significantly improve the performance of primal-dual type algorithms.

7. Conclusions. In this paper, we have incorporated linesearch strategy into the GRPDA recently proposed in [9]. Global convergence and $\mathcal{O}(1/N)$ ergodic convergence rate measured by the function value gap and constraint violations are established in the general convex case. When one of the component functions is strongly convex, accelerated GRPDA with linesearch is proposed, which achieves faster $\mathcal{O}(1/N^2)$ ergodic rate of convergence quantified by the same measures. Furthermore, when the subdifferential operators of both component functions are strongly metric subregular, R-linear convergence results are established. The proposed linesearch strategy does not require evaluating the spectral norm of K and adopts potentially much larger stepsizes. In many practical cases, such as the regularized leastsquares problem, the proposed linesearch strategy only requires minimal extra computational cost and thus is particularly useful. Our numerical experiments on minimax matrix game and LASSO problems demonstrate the benefits gained by incorporating our proposed linesearch and taking advantage of strong convexity of component



FIG. 3. Experimental results on the LASSO problem. Decreasing behavior of function value versus CPU time. From left to right: case (i) with (p,q,s) = (1000, 2000, 100), case (ii) with (p,q,s,v) = (1000, 2000, 10, 0.5), and case (ii) with (p,q,s,v) = (1000, 2000, 10, 0.9).

functions in the objective function. Experimentally, the extra linesearch trial steps used by golden ratio type primal-dual algorithms are about one-third of those proposed by Malitsky and Pock [26] and larger stepsizes can be accepted, which could be significant when the evaluations of proximal operators are highly nontrivial.

Appendix A. Proof of Lemma 3.1.

Proof. (i) Since (15) is fulfilled whenever τ_n satisfies $\sqrt{\beta\tau_n}L \leq \sigma\sqrt{\psi/\tau_{n-1}}$, we have from $\tau_n = \varphi\tau_{n-1}\mu^i$ that (15) is fulfilled whenever $\sqrt{\mu^i} \leq \underline{\tau}/\tau_{n-1}$. Hence, (15) will be fulfilled by the linesearch procedure in Step 2 of Algorithm 3.1 since $\mu \in (0, 1)$.

(ii) We consider two cases.

Case 1. There exists a \overline{k} such that $\tau_n \geq \underline{\tau}$ for all $n \geq \overline{k}$. If there is no infinite subsequence $\{n_k : k \geq 1\} \subseteq \{1, 2, \ldots\}$ such that $\delta_{n_k} \geq \rho$, we have $\delta_n < \rho < 1$ for all n sufficiently large. Then, we will have from $\tau_n = \tau_0 \prod_{i=1}^n \delta_i$ and $\rho < 1$ that $\lim_{n\to\infty} \tau_n = 0$, which contradicts with $\tau_n \geq \underline{\tau}$ for all $n \geq \overline{k}$. Hence, in this case, property (ii) holds.

Case 2. There exists an infinite subsequence $\{n_i : i \geq 1\}$ such that $\tau_{n_i} < \underline{\tau}$. By the linesearch procedure in Step 2, for any $\tau_n \leq \underline{\tau}$, the initial trial $\tau_{n+1} = \varphi \tau_n$ will satisfy (15) and be accepted by the linesearch. Hence, defining $\ell(t) = \lfloor \log_{\varphi}(\underline{\tau}/t) \rfloor$ where $t < \underline{\tau}$, we have the following property:

If
$$\tau_n < \underline{\tau}$$
, then with $k' := n + \ell(\tau_n)$ we have
(68) $\tau_k < \underline{\tau}$ and $\tau_{k+1} = \varphi \tau_k$ for all $k = n, n+1, \dots, k'$, and $\underline{\tau} \le \tau_{k'+1} < \varphi \underline{\tau}$

Here, $\lfloor t \rfloor$ is the largest integer less than or equal to t. So, for any $\tau_{n_i} < \underline{\tau}$, we have $\tau_{k'+1} \geq \underline{\tau}$, where $k' = n_i + \ell(\tau_{n_i})$. In addition, we have $\delta_{k'+1} = \tau_{k'+1}/\tau_{k'} = \varphi > \rho$. Hence, in this case property (ii) also holds.

(iii) First, if $\tau_0 < \underline{\tau}$, by property (68), we have $\tau_s = \tau_0 \varphi^s \ge \underline{\tau}$, where $s = \ell(\tau_0) + 1$. Hence, without loss of generality, to show property (iii), we can simply assume $\tau_1 \ge \underline{\tau}$. Now, we show the following property:

(69) For any
$$\tau_{n-1} \ge \underline{\tau}$$
 and $\tau_n < \underline{\tau}$, we have that (70) and (71) hold.

Since $\tau_n < \tau_{n-1}$, by the linesearch procedure in Step 2, we have $\tau_n = \varphi \tau_{n-1} \mu^j$ with $j \ge 1$ and $\sqrt{\beta \varphi \tau_{n-1} \mu^{j-1}} L > \sigma \sqrt{\psi/\tau_{n-1}}$, which is equivalent to

On the other hand, by (70) and $j \ge 1$, we have

(71)
$$1 + \ell(\tau_n) = 1 + \lfloor \log_{\varphi}(\underline{\tau}/\tau_n) \rfloor = 1 + \lfloor \log_{\varphi}(\underline{\tau}/(\varphi\tau_{n-1}\mu^j)) \rfloor$$
$$= \lfloor \log_{\varphi}(\underline{\tau}/(\tau_{n-1}\mu^j)) \rfloor \leq \lfloor \frac{j+1}{2} \log_{\varphi}(1/\mu) \rfloor \leq j \log_{\varphi}(1/\mu).$$

Let $\mathcal{Z}^+ := \{1, 2, 3, \ldots\}$ be the set of positive integers. Given two integers $i_1 \leq i_2$, let interval $[i_1, i_2] := \{i \in \mathcal{Z}^+ : i_1 \leq i \leq i_2\}$ and interval $[i_1, \infty) := \{i \in \mathcal{Z}^+ : i_1 \leq i < \infty\}$. Then, based on properties (68), (69) and the assumption $\tau_1 \geq \underline{\tau}$, there exist a set of positive integers $\mathcal{K} := \bigcup_{i=1}^{|\mathcal{K}|} \{k_i\} \subseteq \mathcal{Z}^+$ and an associated integer set $\mathcal{M} := \bigcup_{i=1}^{|\mathcal{K}|-1} \{m_i\} \subseteq \mathcal{Z}^+$, where $|\mathcal{K}| \geq 1$ denotes the cardinality of \mathcal{K} that is either a finite number or infinity, such that they partition \mathcal{Z}^+ , i.e., $\mathcal{Z}^+ = \bigcup_{i=1}^{\infty} [k_i, k_{i+1} - 1]$ if $|\mathcal{K}| = \infty$ or $\mathcal{Z}^+ = \bigcup_{i=1}^{|\mathcal{K}|-1} [k_i, k_{i+1} - 1] \cup [k_{|\mathcal{K}|}, \infty)$ if $|\mathcal{K}| < \infty$, and the following properties hold:

- (a) $k_1 = 1$ and $k_i < m_i < k_{i+1}$ for all *i*.
- (b) $\tau_k \geq \underline{\tau}$ for all $k \in [k_i, m_i 1]$ and $\tau_k < \underline{\tau}$ for all $k \in [m_i, k_{i+1} 1]$; see the diagram below

$$\dots, \quad k_i - 1, \quad \overbrace{k_i, \dots, m_i - 1}^{\tau_k \ge \underline{\tau}, \ (m_i - k_i) \text{ times}}, \quad \overbrace{m_i, \dots, k_{i+1} - 1}^{\tau_k < \underline{\tau}, \ (k_{i+1} - m_i) \text{ times}}, \quad k_{i+1}, \dots$$

- (c) If $|\mathcal{K}| < \infty$, $\tau_k \geq \underline{\tau}$ for all $k \geq k_{|\mathcal{K}|}$; otherwise, $|\mathcal{K}| = \infty$ and $\mathcal{Z}^+ = \bigcup_{i=1}^{\infty} [k_i, k_{i+1} 1]$.
- (d) $\tau_{k_i} < \varphi \underline{\tau}$ by property (68) for all $k_i \in \mathcal{K} \setminus \{k_1\}$.
- (e) $\tau_{m_i-1} > \underline{\tau}\mu^{-(j-1)/2}$ by (70) and $k_{i+1} m_i = \ell(\tau_{m_i}) + 1 \le j \log_{\varphi}(1/\mu)$ by (71) for all $m_i \in \mathcal{M}$ and some $j \ge 1$ depending on m_i .

Consider $[k_i, l]$, where $l \in [k_i, m_i - 1]$. Since $\tau_l = \tau_{k_i - 1} \prod_{j=k_i}^l \delta_j$, $\tau_{k_i - 1} < \underline{\tau}$, and $\tau_l \geq \underline{\tau}$, we have $\prod_{j=k_i}^l \delta_j \geq 1$. Then, it follows from $\delta_j \leq \varphi$ for all j that $|\{k_i \leq j \leq l : \delta_j \geq 1/\varphi\}| \geq \lfloor \frac{l-k_i+1}{2} \rfloor + 1$, and thus

(72)
$$\left|\left\{k_i \leq j \leq l : \tau_j \geq \underline{\tau} \text{ and } \delta_j \geq 1/\varphi\right\}\right| \geq \left\lfloor \frac{l-k_i+1}{2} \right\rfloor + 1 \text{ for all } l \in [k_i, m_i-1].$$

Now, we consider any interval $[k_i, k_{i+1} - 1] = [k_i, m_i - 1] \cup [m_i, k_{i+1} - 1]$. Let $j \ge 1$ be the integer associated with m_i such that property (e) holds. Since $\tau_{k+1} \le \varphi \tau_k$ for all k, we have $\tau_{m_i-1} < \tau_{k_i} \varphi^{m_i-1-k_i}$. Then, by properties (d) and (e), for $k_i \ne k_1 = 1$, we have $\tau_{k_i} < \varphi \underline{\tau}$ and $\tau_{m_i-1} \ge \underline{\tau} \mu^{-(j-1)/2}$, which together with the above inequality gives $\underline{\tau} \mu^{-(j-1)/2} \le \tau_{m_i-1} < \tau_{k_i} \varphi^{m_i-1-k_i} < \underline{\tau} \varphi^{m_i-k_i}$, which is equivalent to $(j-1)\log_{\varphi}(1/\sqrt{\mu}) < m_i - k_i$. For $k_i = k_1 = 1$, we have $\underline{\tau} \mu^{-(j-1)/2} < \tau_1 \varphi^{m_i-2}$, or $(j-1)\log_{\varphi}(1/\sqrt{\mu}) \le \log_{\varphi}(\tau_1/\underline{\tau}) + m_i - 2$. So, there exists an integer constant $\overline{j} \ge 1$, which does not depend on either k_i or m_i , such that

(73)
$$m_i - k_i \ge \frac{j}{4} \log_{\varphi}(1/\mu)$$
 whenever $j \ge \overline{j}$.

Next, we show that there exists a constant $\bar{c} > 0$, which does not depend on m_i or $[k_i, k_{i+1} - 1]$, such that

(74)
$$m_i - k_i \ge \bar{c} (k_{i+1} - m_i).$$

Noticing that by property (e), we have $k_{i+1} - m_i \leq j \log_{\varphi}(1/\mu)$. Hence, it follows from (73) that if $j \geq \overline{j}$, we have

$$m_i - k_i \ge \frac{\log_{\varphi}(1/\mu)}{4\log_{\varphi}(1/\mu)}(k_{i+1} - m_i) = \frac{1}{4}(k_{i+1} - m_i).$$

On the other hand, if $1 \le j < \overline{j}$, we have from $k_{i+1} - m_i \le j \log_{\varphi}(1/\mu)$ that

$$m_i - k_i \ge 1 > j/\overline{j} \ge \frac{k_{i+1} - m_i}{\overline{j} \log_{\varphi}(1/\mu)}.$$

Hence, (74) holds with $\bar{c} = \min\{1/(j \log_{\varphi}(1/\mu)), 1/4\} > 0.$

It follows from (74) that $m_i - k_i \ge \tilde{c} (k_{i+1} - k_i)$ with $\tilde{c} := \bar{c}/(1 + \bar{c}) \in (0, 1)$. If $|\mathcal{K}| = \infty$, given any $N \ge 1$, it follows from property (c) that $N \in [k_i, k_{i+1} - 1]$ for certain $i \ge 1$. Hence, by the definition of $\mathcal{K}_N = \{1 \le n \le N : \tau_n \ge \underline{\tau}, \ \delta_n \ge 1/\varphi\}$, we have $\mathcal{K}_N = \bigcup_{j=1}^i (\mathcal{K}_N \cap [k_j, k_{j+1} - 1])$. Then, it follows from (72), (74), and property (b) that

$$\begin{aligned} |\mathcal{K}_N| &= \sum_{j=1}^{i} |\mathcal{K}_N \cap [k_j, k_{j+1} - 1]| \ge \sum_{j=1}^{i} |\mathcal{K}_N \cap [k_j, m_j - 1]| \\ &\ge \left\lfloor \frac{\min\{N, m_i - 1\} - k_i + 1}{2} \right\rfloor + 1 + \sum_{j=1}^{i-1} \left(\left\lfloor \frac{m_j - k_j}{2} \right\rfloor + 1 \right) \\ &\ge \frac{1}{2} \left(\min\{N, m_i - 1\} - k_i + 1 + \sum_{j=1}^{i-1} (m_j - k_j) \right) \\ &\ge \frac{\tilde{c}}{2} \left(N - k_i + 1 + \sum_{j=1}^{i-1} (k_{j+1} - k_j) \right) = \check{c}N, \end{aligned}$$

where $\check{c} = \tilde{c}/2 \in (0, 1)$ and the last " \geq " is because $m_i - k_i \geq \tilde{c} (k_{i+1} - k_i), k_{i+1} \geq N+1$, and $\tilde{c} < 1$. If $|\mathcal{K}| < \infty$, by property (c), for all $n \geq k_{|\mathcal{K}|}$ we have $\tau_n \geq \underline{\tau}$, and by following the same arguments as for (72) we have $|\{k_{|\mathcal{K}|} \leq j \leq n : \tau_j \geq \underline{\tau} \text{ and } \delta_j \geq 1\}$ $1/\varphi\} \ge \lfloor \frac{n-k_{|\mathcal{K}|}+1}{2} \rfloor + 1$. This property, together with (75), implies that for all $N \ge 1$ we have $|\mathcal{K}_N| \ge \hat{c}N$ for some $\hat{c} > 0$.

Appendix B. Proof of Lemma 4.1.

Proof. (i) This conclusion follows almost from an identical proof of conclusion (i)

in Lemma 3.1 except by replacing β by β_n and setting $\sigma = 1$. (ii) Let $h(\tau) := 1 + \frac{(\psi - \varphi)\gamma_g \tau}{\psi + \varphi \gamma_g \tau}$. Since $h(\tau)$ is strictly increasing with respect to $\tau > 0$, we have that $1 < 1 + \gamma_g \omega_n \tau_{n-1} = h(\tau_{n-1}) < \varsigma := \psi/\varphi$. So, by (44), $\beta_{n-1} < \beta_n = \psi/\varphi$. $\beta_{n-1}h(\tau_{n-1}) < \beta_{n-1}$. Therefore, for any $\sqrt{\beta_{n-1}}\tau_{n-1} \leq 1/L$, we have $\sqrt{\beta_n}\tau_{n-1} \leq 1/L$ $\sqrt{\zeta\beta_{n-1}}\tau_{n-1} \leq \sqrt{\zeta}/L = \frac{1}{L}\sqrt{\frac{\psi}{\varphi}}$, which by the linesearch procedure in Step 2 implies that the initial trial $\tau_n = \varphi \tau_{n-1}$ will satisfy (45) and be accepted by the linesearch.

(75)

Hence, analogous to property (68), we have the following property.

- If $\sqrt{\beta_n}\tau_n < \frac{1}{L}$, then there exists an integer $\ell_n \ge 0$ such that $k' := n + \ell_n$ satisfies
- (76) $\sqrt{\beta_k}\tau_k < \frac{1}{L}$ and $\tau_{k+1} = \varphi \tau_k$ for all $k = n, n+1, \dots, k'$, and $\sqrt{\beta_{k'+1}}\tau_{k'+1} \ge \frac{1}{L}$,

which, by $\beta_n < \beta_{n+1} < \varsigma \beta_n$ and $\tau_{n+1} \leq \varphi \tau_n$ for all $n \geq 1$, also implies

(77)
$$\ell_n \leq \vartheta \left(\sqrt{\beta_n} \tau_n \right) \text{ and } \sqrt{\beta_{k'+1}} \tau_{k'+1} < \sqrt{\varsigma \beta_{k'}} \varphi \tau_{k'} = \sqrt{\psi \varphi} \sqrt{\beta_{k'}} \tau_{k'} < \theta,$$

where $\vartheta(t) := \lfloor \log_{\varphi}(1/(Lt)) \rfloor$ for t < 1/L and $\theta := \sqrt{\psi\varphi}/L$. Analogous to property (69), we show the following property:

(78) For any $\sqrt{\beta_{n-1}}\tau_{n-1} \ge \frac{1}{L}$ and $\sqrt{\beta_n}\tau_n < \frac{1}{L}$, we have that (79) and (80) hold.

Since $\beta_n > \beta_{n-1}$ and $\sqrt{\beta_n}\tau_n < \sqrt{\beta_{n-1}}\tau_{n-1}$, by the linesearch procedure in Step 2, we have $\tau_n = \varphi \tau_{n-1} \mu^j$ with $j \ge 1$ and $\sqrt{\beta_n \varphi \tau_{n-1} \mu^{j-1}} L > \sqrt{\psi/\tau_{n-1}}$, which gives

(79)
$$\sqrt{\beta_n}\tau_{n-1} \ge \sqrt{\frac{\psi}{\varphi}} \frac{\mu^{-\frac{j-1}{2}}}{L} \quad \text{and} \quad \sqrt{\beta_{n-1}}\tau_{n-1} > \sqrt{\beta_n/\varsigma}\tau_{n-1} \ge \frac{\mu^{-\frac{j-1}{2}}}{L}.$$

It follows from (77), $\psi > \varphi > 1$, $j \ge 1$, and (79) that

$$1 + \ell_n \leq 1 + \vartheta(\sqrt{\beta_n}\tau_n) = 1 + \left\lfloor \log_{\varphi} \left(1/(L\sqrt{\beta_n}\tau_n) \right) \right\rfloor$$
$$= 1 + \left\lfloor \log_{\varphi} \left(1/(L\sqrt{\beta_n}\varphi\tau_{n-1}\mu^j) \right) \right\rfloor = \left\lfloor \log_{\varphi} (1/(L\sqrt{\beta_n}\tau_{n-1}\mu^j)) \right\rfloor$$
$$\leq \left\lfloor \frac{j+1}{2} \log_{\varphi} (1/\mu) + \frac{1}{2} \log_{\varphi} (\varphi/\psi) \right\rfloor \leq j \log_{\varphi} (1/\mu).$$

Let $\mathcal{Z}^+ := \{1, 2, \ldots\}$ be the set of positive integers. Given two integers $i_1 \leq i_2$, let $[i_1, i_2] := \{i \in \mathcal{Z}^+ : i_1 \leq i \leq i_2\}$ and $[i_1, \infty) := \{i \in \mathcal{Z}^+ : i_1 \leq i < \infty\}$. To show this lemma, without loss of generality, by property (76), we can simply assume $\sqrt{\beta_1}\tau_1 \geq 1/L$. Then, based on properties (76) and (78), there exist a set of positive integers $\mathcal{K} := \bigcup_{i=1}^{|\mathcal{K}|} \{k_i\} \subseteq \mathcal{Z}^+$ and an associated integer set $\mathcal{M} := \bigcup_{i=1}^{|\mathcal{K}|-1} \{m_i\} \subseteq \mathcal{Z}^+$, where $|\mathcal{K}| \geq 1$ denotes the cardinality of \mathcal{K} that is either a finite number or infinity, such that they partition \mathcal{Z}^+ , i.e., $\mathcal{Z}^+ = \bigcup_{i=1}^{\infty} [k_i, k_{i+1} - 1]$ if $|\mathcal{K}| = \infty$ or $\mathcal{Z}^+ = \bigcup_{i=1}^{|\mathcal{K}|-1} [k_i, k_{i+1} - 1] \cup [k_{|\mathcal{K}|}, \infty)$ if $|\mathcal{K}| < \infty$, and the following properties hold:

- (a) $k_1 = 1$ and $k_i < m_i < k_{i+1}$ for all *i*.
- (b) $\sqrt{\beta_k}\tau_k \ge 1/L$ for all $k \in [k_i, m_i 1]$ and $\sqrt{\beta_k}\tau_k < 1/L$ for all $k \in [m_i, k_{i+1} 1]$.
- (c) If $|\mathcal{K}| < \infty$, $\sqrt{\beta_k} \tau_k \ge 1/L$ for all $k \ge k_{|\mathcal{K}|}$; otherwise, $|\mathcal{K}| = \infty$ and $\mathcal{Z}^+ = \bigcup_{i=1}^{\infty} [k_i, k_{i+1} 1].$
- (d) $\sqrt{\beta_{k_i}} \tau_{k_i} < \theta$ by (77) for all $k_i \in \mathcal{K} \setminus \{k_1\}$, where $\theta = \sqrt{\psi \varphi}/L$.
- (e) $\sqrt{\beta_{m_i-1}}\tau_{m_i-1} > \mu^{-(j-1)/2}/L$ by (79) and $k_{i+1} m_i = \ell_{m_i} + 1 \le j \log_{\varphi}(1/\mu)$ by (80) for all $m_i \in \mathcal{M}$ and some $j \ge 1$ depending on m_i , where ℓ_{m_i} is defined in property (76) associated with $\sqrt{\beta_{m_i}}\tau_{m_i}$.

Now, we consider any interval $[k_i, k_{i+1} - 1] = [k_i, m_i - 1] \cup [m_i, k_{i+1} - 1]$. Let $j \ge 1$ be the integer associated with m_i such that property (e) holds. Since $\sqrt{\beta_{k+1}\tau_{k+1}} < \sqrt{\varsigma\beta_k}\varphi\tau_k = \sqrt{\psi\varphi}\sqrt{\beta_k}\tau_k$ for all k, we have

$$\sqrt{\beta_{m_i-1}}\tau_{m_i-1} < \sqrt{\beta_{k_i}}\tau_{k_i}\rho^{(m_i-1-k_i)/2},$$

where $\rho := \psi \varphi > 1$. By properties (d) and (e), for $k_i \neq k_1 = 1$, we have $\sqrt{\beta_{k_i} \tau_{k_i}} < \theta$ and $\sqrt{\beta_{m_i-1}} \tau_{m_i-1} \ge \mu^{-(j-1)/2}/L$, which together with the above inequality gives

$$\mu^{-(j-1)/2}/L < \theta \rho^{(m_i - 1 - k_i)/2} \iff j \log_{\rho}(1/\mu) < \log_{\rho}\left((\theta L)^2/(\mu\rho)\right) + m_i - k_i$$

= $\log_{\rho}\left(1/\mu\right) + m_i - k_i.$

For $k_i = k_1 = 1$, we have $j \log_{\rho}(1/\mu) \leq \log_{\rho}((\theta_1 L)^2/(\mu\rho)) + m_i - 1$, where $\theta_1 := \sqrt{\beta_1}\tau_1$. So, there exists an integer constant $\overline{j} \geq 1$, which does not depend on either k_i or m_i , such that

(81)
$$m_i - k_i \ge \frac{j}{2} \log_{\rho}(1/\mu)$$
 whenever $j \ge \overline{j}$.

Next, we show that there exists a constant $\bar{c} > 0$, which does not depend on m_i , such that

(82)
$$m_i - k_i \ge \bar{c} (k_{i+1} - m_i + 1).$$

Notice that by property (e), we have

(83)
$$k_{i+1} - m_i + 1 \le j \log_{\varphi}(1/\mu) + 1 \le j (\log_{\varphi}(1/\mu) + 1).$$

Hence, it follows from (81) and (83) that whenever $j \ge \overline{j}$ we have

$$m_i - k_i \ge \frac{\log_{\rho}(1/\mu)}{2(\log_{\varphi}(1/\mu) + 1)}(k_{i+1} - m_i + 1).$$

On the other hand, if $1 \le j < \overline{j}$, then by (83) we have

$$m_i - k_i \ge 1 > j/\bar{j} \ge \frac{k_{i+1} - m_i + 1}{\bar{j}(\log_{\varphi}(1/\mu) + 1)}.$$

Hence, (82) holds with $\bar{c} = \min\{1/(\bar{j}(\log_{\varphi}(1/\mu)+1)), \log_{\rho}(1/\mu)/(2(\log_{\varphi}(1/\mu)+1))\} > 0.$

Now, by (44) we have $\beta_{n+1} = \beta_n h(\tau_n) = \beta_n (1+t_n)$, where $t_n = \frac{(\psi - \varphi)\gamma_g \tau_n}{\psi + \varphi \gamma_g \tau_n}$. Then, we have

(84)
$$\sqrt{\beta_{n+1}} - \sqrt{\beta_n} = \frac{\beta_{n+1} - \beta_n}{\sqrt{\beta_{n+1}} + \sqrt{\beta_n}} = \frac{t_n \beta_n}{\sqrt{\beta_{n+1}} + \sqrt{\beta_n}}$$
$$\geq \frac{t_n \beta_n}{(\sqrt{\varsigma} + 1)\sqrt{\beta_n}} \geq \tilde{c} \min\{\tau_n, 1\}\sqrt{\beta_n},$$

where $\tilde{c} > 0$ is some constant. Consider any $k \in [k_i, k_{i+1}-1] = [k_i, m_i-1] \cup [m_i, k_{i+1}-1]$. 1]. If $k \in [m_i, k_{i+1}-1]$, by (82) we have

(85)
$$\frac{m_i - k_i}{k_{i+1} - k_i + 1} = \frac{m_i - k_i}{(m_i - k_i) + (k_{i+1} - m_i + 1)} \ge \frac{1}{1 + 1/\bar{c}} = \frac{\bar{c}}{1 + \bar{c}},$$

and it thus follows from (84), property (b), $\beta_n \ge \beta_0 > 0$, and (85) that

$$\sqrt{\beta_{k}} - \sqrt{\beta_{k_{i}}} = \sum_{n=k_{i}}^{k-1} \left(\sqrt{\beta_{n+1}} - \sqrt{\beta_{n}} \right) \ge \tilde{c} \sum_{n=k_{i}}^{k-1} \min\{\tau_{n}, 1\} \sqrt{\beta_{n}}$$
$$\ge \tilde{c} \sum_{n=k_{i}}^{m_{i}-1} \min\{\tau_{n}, 1\} \sqrt{\beta_{n}} \ge \tilde{c} \sum_{n=k_{i}}^{m_{i}-1} \min\{1/L, \sqrt{\beta_{0}}\}$$
$$= \tilde{c}(m_{i} - k_{i}) \min\{1/L, \sqrt{\beta_{0}}\} \ge \hat{c}(k_{i+1} - k_{i} + 1) \ge \hat{c} (k - k_{i} + 1),$$

where $\hat{c} := \tilde{c} \, \bar{c} \min\{1/L, \sqrt{\beta_0}\}/(1+\bar{c}) > 0$, and \bar{c} and \tilde{c} are constants given in (82) and (84), respectively. On the other hand, if $k \in [k_i, m_i - 1]$, similar to (86) we can show

(87)
$$\sqrt{\beta_k} - \sqrt{\beta_{k_i}} \ge \tilde{c} \min\{1/L, \sqrt{\beta_0}\}(k-k_i) \ge \hat{c} (k-k_i).$$

If $|\mathcal{K}| = \infty$, given any $n \ge 1$, it follows from property (c) that $n \in [k_i, k_{i+1} - 1]$ for certain $i \ge 1$. Hence, by (86), (87), and $0 < \beta_j < \beta_{j+1}$ for any $j \ge 1$, we have

$$\sqrt{\beta_n} = \sqrt{\beta_n} - \sqrt{\beta_{k_i}} + \sum_{j=1}^{i-1} \left(\sqrt{\beta_{k_{j+1}}} - \sqrt{\beta_{k_j}} \right) + \sqrt{\beta_{k_1}} \\
= \sqrt{\beta_n} - \sqrt{\beta_{k_i}} + \sum_{j=1}^{i-1} \left[\left(\sqrt{\beta_{k_{j+1}-1}} - \sqrt{\beta_{k_j}} \right) + \left(\sqrt{\beta_{k_{j+1}}} - \sqrt{\beta_{k_{j+1}-1}} \right) \right] + \sqrt{\beta_{k_1}} \\
\ge \sqrt{\beta_n} - \sqrt{\beta_{k_i}} + \sum_{j=1}^{i-1} \left(\sqrt{\beta_{k_{j+1}-1}} - \sqrt{\beta_{k_j}} \right) \\
(88) \ge \hat{c} \left[(n-k_i) + \sum_{j=1}^{i-1} (k_{j+1} - k_j) \right] = \hat{c} (n-1).$$

If $|\mathcal{K}| < \infty$, by property (c), for any $k \ge k_{|\mathcal{K}|}$ we have $\sqrt{\beta_k}\tau_k \ge 1/L$ and thus $\sqrt{\beta_k} - \sqrt{\beta_{k_{|\mathcal{K}|}}} \ge \hat{c}(k - k_{|\mathcal{K}|})$. This together with (86) and property (c) also implies (88) holds for $n \ge 1$, and thus conclusion (ii) follows.

(iii). Note that for any $\sqrt{\beta_p}\tau_p \ge \sqrt{\beta_q}\tau_q$ with $p \le q$, we have from $\beta_p \le \beta_q$ that $\tau_p \ge \tau_q$. So, based on property (b) in (ii) and (82), we have from (82) the property

$$\begin{array}{ll} (\mathbf{b}') \quad \tau_p \geq \tau_q z \quad \text{for all } p \in [k_i, m_i - 1], q \in [m_i, k_{i+1} - 1] \\ \text{and therefore } \sum_{n=k_i}^{m_i - 1} \tau_n \geq \bar{c} \sum_{n=m_i}^{k_{i+1} - 1} \tau_n. \end{array}$$

Let $\mathcal{K} \subset \mathcal{Z}^+$ be the set given in the proof of (ii). If $|\mathcal{K}| = \infty$, then for any $N \ge 1$ it follows from property (c) that $N \in [k_j, k_{j+1} - 1]$ for certain $j \ge 1 = k_1$. Hence, we have from property (b') that

$$\sum_{n=1}^{N} \tau_n = \sum_{n=k_j}^{N} \tau_n + \sum_{s=1}^{j-1} \sum_{n=k_s}^{k_{s+1}-1} \tau_n = \sum_{n=k_j}^{N} \tau_n + \sum_{s=1}^{j-1} \left(\sum_{n=k_s}^{m_s-1} \tau_n + \sum_{n=m_s}^{k_{s+1}-1} \tau_n \right)$$
$$\leq (1+1/\bar{c}) \sum_{n=k_j}^{\min\{N,m_j-1\}} \tau_n + \sum_{s=1}^{j-1} \left(\sum_{n=k_s}^{m_s-1} \tau_n + 1/\bar{c} \sum_{n=k_s}^{m_s-1} \tau_n \right)$$
$$= (1+1/\bar{c}) \left(\sum_{n=k_j}^{\min\{N,m_j-1\}} \tau_n + \sum_{s=1}^{j-1} \sum_{n=k_s}^{m_s-1} \tau_n \right) = (1+1/\bar{c}) \sum_{n\in\mathcal{S}_N} \tau_n.$$

When $|\mathcal{K}| < \infty$, we can also similarly prove that $\sum_{n=1}^{N} \tau_n \leq (1+1/\bar{c}) \sum_{n \in S_N} \tau_n$ for any $N \geq 1$, because $\sqrt{\beta_n} \tau_n \geq 1/L$ for all $n \geq k_{|\mathcal{K}|}$. The proof is completed by letting $\tilde{c} = 1 + 1/\bar{c}$.

Acknowledgments. We appreciate the associate editor and two anonymous referees for their constructive comments and suggestions which helped to improve this paper significantly.

REFERENCES

- F. A. ARTACHO AND M. GEOFFROY, Characterization of metric regularity of subdifferentials, J. Convex Anal., 15 (2008), pp. 365–380.
- [2] H. H. BARRETT AND K. J. MYERS, Foundations of Image Science, John Wiley & Sons, Hoboken, NJ, 2004.
- [3] A. BECK, First-Order Methods in Optimization, MOS-SIAM Ser. Optim. 25, SIAM, Philadelphia, 2017, https://doi.org/10.1137/1.9781611974997.
- [4] D. P. BERTSEKAS AND E. M. GAFNI, Projection methods for variational inequalities with application to the traffic assignment problem, Math. Programming Stud., No. 17 (1982), pp. 139–159, https://doi.org/10.1007/bfb0120965.
- [5] T. BOUWMANS, N. S. AYBAT, AND E. H. ZAHZAH, EDS., Handbook of Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing, Chapman and Hall/CRC, Boca Raton, FL, 2016.
- [6] A. CHAMBOLLE, M. J. EHRHARDT, P. RICHTARIK, AND C.-B. SCHONLIEB, Stochastic primaldual hybrid gradient algorithm with arbitrary sampling and imaging application, SIAM J. Optim., 28 (2018), pp. 2783–2808, https://doi.org/10.1137/17M1134834.
- [7] A. CHAMBOLLE AND T. POCK, A first-order primal-dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
- [8] A. CHAMBOLLE AND T. POCK, On the ergodic convergence rates of a first-order primaldual algorithm, Math. Program., 159 (2016), pp. 253–287, https://doi.org/10.1007/ s10107-015-0957-3.
- X. CHANG AND J. F. YANG, A golden ratio primal-dual algorithm for structured convex optimization, J. Sci. Comput., 87 (2021), 47.
- [10] L. CONDAT, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, J. Optim. Theory Appl., 158 (2013), pp. 460–479.

Downloaded 05/29/23 to 130.39.169.8 . Redistribution subject to SIAM license or copyright; see https://epubs.siam.org/terms-privacy

- [11] D. DRUSVYATSKIY AND A. S. LEWIS, Error bounds, quadratic growth, and linear convergence of proximal methods, Math. Oper. Res., 43 (2018), pp. 919–948.
- [12] J. DUCHI, S. SHALEV-SHWARTZ, Y. SINGER, AND T. CHANDRA, Efficient projections onto the l₁-ball for learning in high dimensions, in Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 272–279.
- [13] E. ESSER, X. ZHANG, AND T. F. CHAN, A general framework for a class of first order primaldual algorithms for convex optimization in imaging science, SIAM J. Imaging Sci., 3 (2010), pp. 1015–1046, https://doi.org/10.1137/09076934X.
- [14] D. GABAY AND B. MERCIER, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, Comput. Math. Appl., 2 (1976), pp. 17–40.
- [15] R. GLOWINSKI AND A. MARROCCO, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires, Rev. Française Automat. Informat. Recherche Opfationnelle Sér. Rouge Anal. Numér., 9 (1975), pp. 41–76.
- [16] E. T. HALE, W. YIN, AND Y. ZHANG, Fixed-point continuation applied to compressed sensing: Implementation and numerical experiments, J. Comput. Math., 28 (2010), pp. 170–194, https://doi.org/10.4208/jcm.2009.10-m1007.
- [17] S. HAYDEN AND O. STANLEY, A low patch-rank interpretation of texture, SIAM J. Imaging Sci., 6 (2013), pp. 226–262, https://doi.org/10.1137/110854989.
- [18] B. HE, Y. YOU, AND X. YUAN, On the convergence of primal-dual hybrid gradient algorithm, SIAM J. Imaging Sci., 7 (2014), pp. 2526–2537, https://doi.org/10.1137/140963467.
- [19] B. HE AND X. YUAN, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, SIAM J. Imaging Sci., 5 (2012), pp. 119–149, https://doi.org/10.1137/100814494.
- [20] E. JONATHAN AND P. DIMITRI, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Programming, 55 (1992), pp. 293– 318.
- [21] P. LATAFAT, N. M. FRERIS, AND P. PATRINOS, A new randomized block-coordinate primal-dual proximal algorithm for distributed optimization, IEEE Trans. Automat. Control, 64 (2019), pp. 4050–4065.
- [22] J. LIANG, J. FADILI, AND G. PEYR, Convergence rates with inexact nonexpansive operators,

Math. Program., 159 (2016), pp. 403–434.

- [23] P. L. LIONS AND B. MERCIER, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal., 16 (1979), pp. 964–979, https://doi.org/10.1137/0716071.
- [24] Y. LIU, Y. XU, AND W. YIN, Acceleration of primal-dual methods by preconditioning and simple subproblem procedures, J. Sci. Comput., 86 (2021), 21.
- Y. MALITSKY, Golden ratio algorithms for variational inequalities, Math. Program., 184 (2020), pp. 383-410.
- [26] Y. MALITSKY AND T. POCK, A first-order primal-dual algorithm with linesearch, SIAM J. Optim., 28 (2018), pp. 411–432, https://doi.org/10.1137/16M1092015.
- [27] A. NEDIC AND A. OZDAGLAR, Subgradient methods for saddle-point problems, J. Optim. Theory Appl., 142 (2009), pp. 205–228, https://doi.org/10.1007/s10957-009-9522-7.
- [28] T. POCK AND A. CHAMBOLLE, Diagonal preconditioning for first order primal-dual algorithms in convex optimization, in Proceedings of the 2011 IEEE International Conference on Computer Vision, 2011, pp. 1762–1769.
- [29] R. T. ROCKAFELLAR, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
- [30] E. K. RYU AND W. YIN, Large-Scale Convex Optimization: Algorithm Analysis via Monotone Operators, Cambridge University Press, Cambridge, UK, to appear.
- [31] 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, https://doi.org/10.1137/130910774.
- [32] E. Y. SIDKY, J. H. JORGENSEN, AND X. C. PAN, Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm, Phys. Med. Biol., 57 (2012), pp. 3065–3091.
- [33] H. UZAWA, Iterative methods for concave programming, in Studies in Linear and Nonlinear Programming, K. J. Arrow, L. Hurwicz, and H. Uzawa, eds., Stanford University Press, Stanford, CA, 1958, pp. 154–165.
- [34] Q. VAN NGUYEN, O. FERCOQ, AND V. CEVHER, Smoothing Technique for Nonsmooth Composite Minimization with Linear Operator, preprint, https://arxiv.org/abs/1706.05837, 2017.
- [35] M.-L. VLADAREAN, Y. MALITSKY, AND V. CEVHER, A First-Order Primal-Dual Method with Adaptivity to Local Smoothness, preprint, https://arxiv.org/abs/2110.15148, 2021.
- [36] J. YANG AND Y. ZHANG, Alternating direction algorithms for ℓ₁-problems in compressive sensing, SIAM J. Sci. Comput., 33 (2011), pp. 250–278, https://doi.org/10.1137/090777761.
- [37] 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.