Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 19, No. 1, pp. 21–43, 2023

NEW CONVERGENCE RESULTS OF THE GOLDEN RATIO PRIMAL-DUAL ALGORITHM*

Xiaokai Chang and Junfeng Yang†

Abstract: Recently, we proposed a golden ratio primal-dual algorithm (GRPDA) for solving bilinear saddle point problem. It is full-splitting and can be viewed as a new variant of the classical Arrow-Hurwicz method. Moreover, compared with the famous primal-dual algorithm of Chambolle and Pock, it converges under a much relaxed stepsize condition. However, ergodic convergence rate results have been established for GRPDA only in terms of the so-called "primal-dual gap function", which could vanish at nonstationary points, making existing results less informative. In this work, based on some equivalent reformulations of the bilinear saddle point problem as constrained/unconstrained optimization problems, we establish new convergence rate results measured by the conventional measures of function value residual and constraint violation. We establish in the general convex case $\mathcal{O}(1/N)$ ergodic sublinear convergence rate result, where N denotes the iteration counter. When either one of the component functions is strongly convex, an accelerated GRPDA is constructed, which achieves the faster $\mathcal{O}(1/N^2)$ ergodic convergence rate. These new results enrich the convergence theory of GRPDA. Furthermore, we demonstrate the superior performance of the accelerated GRPDA via preliminary numerical results on the least absolute deviation and the LASSO problems.

Key words: *Bilinear saddle point problem, golden ratio, primal-dual algorithm, augmented Lagrangian, convex combination, convergence rate*

Mathematics Subject Classification: *49M29, 65K10, 65Y20, 90C25*

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. In this paper, we consider the following bilinear saddle point problem

$$\min_{x\in\mathbb{R}^q}\max_{y\in\mathbb{R}^p} g(x) + \langle Kx, y\rangle - f^*(y), \tag{1.1}$$

where $f : \mathbb{R}^p \to (-\infty, +\infty]$ and $g : \mathbb{R}^q \to (-\infty, +\infty]$ are extended real-valued closed proper convex functions [24], f^* denotes the Legendre-Fenchel conjugate of f, i.e., $f^*(\cdot) = \sup_{u\in\mathbb{R}^p}\{\langle\cdot, u\rangle - f(u)\}$, and $K \in \mathbb{R}^{p\times q}$ is a linear operator from $\mathbb{R}^q$ to $\mathbb{R}^p$.

*Supported by the National Natural Science Foundation of China (NSFC-12161053, 11922111), Ministry of Science and Technology of China (2020YFA0713800) and the Fundamental Research Funds for the Central Universities (020314380034).

†Corresponding author.

By using the fact that $(f^*)^* = f$, see, e.g., [24], the bilinear saddle point problem (1.1) can be equivalently reformulated as a primal minimization problem

$$\inf \mathcal{P} := \min_{x \in \mathbb{R}^q} \{\mathcal{P}(x) := g(x) + f(Kx)\}. \tag{1.2}$$

Let $K^\top$ be the adjoint operator of K. The Fenchel dual of (1.2) is given by

$$\sup \mathcal{D} := \max_{y \in \mathbb{R}^p} \{\mathcal{D}(y) := -f^*(y) - g^*(-K^\top y)\}. \tag{1.3}$$

By introducing an auxiliary variable $w \in \mathbb{R}^p$, the primal problem (1.2) can be rewritten as a two-block separable optimization problem with linear equality constraint

$$\min_{x \in \mathbb{R}^q,\, w \in \mathbb{R}^p} \{g(x) + f(w) \ \text{ s.t. } \ Kx - w = 0\}. \tag{1.4}$$

Similarly, the dual problem (1.3) can be equivalently reformulated as

$$\max_{y \in \mathbb{R}^p,\, u \in \mathbb{R}^q} \{-f^*(y) - g^*(u) \ \text{ s.t. } \ K^\top y + u = 0\}. \tag{1.5}$$

It is apparent that the optimal values of (1.4) and (1.5) are equal to $\inf \mathcal{P}$ and $\sup \mathcal{D}$, respectively.

Problem (1.1) and its reformulations (1.2)-(1.5) arise from numerous applications, including signal and image processing, machine learning, statistics, mechanics and economics, to name a few, see, e.g., [5, 4, 30, 16, 3] and the references therein. In many applications, the component functions f and g admit simple structures in the sense that their proximal point mappings (see definition in Section 1.1) can be evaluated efficiently. Examples of such functions are abundant, see [2, Chapter 6]. We thus make the following assumption.

Assumption 1.1. Assume that the proximal point mappings of the component functions f and g either have closed form formulas or can be evaluated efficiently.

In the rest of this section, we define some notation, review some of the most closely related algorithms for solving (1.1)-(1.5), summarize our motivation and contributions and present the organization of this paper.

1.1 Notation

Let h be any extended real-valued closed proper convex function defined on a finite dimensional Euclidean space. The effective domain of h is denoted by $\mathrm{dom}(h) := \{x : h(x) < +\infty\}$, and its subdifferential at x is denoted by $\partial h(x) := \{\xi : h(y) \geq h(x) + \langle \xi, y - x\rangle \text{ for all } y\}$. Furthermore, the proximal point mapping of h is given by

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

Since h is closed proper convex, Prox_h is uniquely well defined everywhere. The relative interior of a set C is denoted by $\mathrm{ri}(C)$.

Throughout this paper, we let $\phi = \frac{\sqrt{5}+1}{2}$ be the golden ratio, which is a key parameter of golden ratio type algorithms, $\mathbb{N} = \{1, 2, 3, \ldots\}$ be the set of positive integers, and $\mathbb{S}^n_+$ (resp., $\mathbb{S}^n_{++}$) be the set of all $n \times n$ symmetric positive semidefinite (resp., positive definite) matrices. The identity operator/matrix is denoted by I, whose domain/order is clear from

the context, the composition of two operators is denoted by "$\circ$", and the operator norm of K is denoted by $L := \|K\| = \sup\{\|Kx\| : \|x\| = 1\}$. Given $H \in \mathbb{S}^n_+$ and $u, v, w \in \mathbb{R}^n$, we let

$$\Delta_H(u, v, w) := \frac{1}{2}\big(\|u - v\|^2_H - \|u - w\|^2_H\big), \tag{1.6}$$

where $\|\cdot\|_H := \sqrt{\langle \cdot, H\cdot\rangle}$. When $H = I$, we omit the subscript H and simply write $\Delta(\cdot) := \Delta_I(\cdot)$. Other notation will be specified later.

1.2 Related algorithms

For large scale applications, traditional optimization approaches such as interior point methods or second-order type methods are generally not suitable because a single iteration of them is just too expensive to be implementable in practice. Moreover, nondifferentiability of optimization problems naturally arises from many applications, especially when regularization technique is adopted. Primal-dual full-splitting algorithms which do not rely on solving any subproblems or linear system of equations iteratively are extremely popular, largely because they are able to take full use of problem structures. The dominant computations at each iteration of such algorithms are several matrix-vector multiplications and evaluations of the proximal point mappings, which are basically the maximum computational burden bearable for large scale problems.

Among others, alternating direction method of multipliers (ADMM, [14, 13, 20]) is a popular primal-dual algorithm for solving (1.1)-(1.5) simultaneously. However, ADMM is not full-splitting because, when applied to (1.4), the x-subproblem can be computationally expensive due to a quadratic term $\|Kx - w\|^2$ appearing in the augmented Lagrangian function. For this reason, ADMM was usually modified in practice, e.g., the variants of proximal/linearized ADMM [10, 17, 12, 9]. One of the most simple primal-dual full-splitting algorithms for solving (1.1)-(1.5) is probably the classical Arrow-Hurwicz method [28], which, started at $(x_0, y_0) \in \mathbb{R}^q \times \mathbb{R}^p$, iterates as

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

for $n \geq 0$, where $\tau, \sigma > 0$ are stepsize parameters. However, existing results indicate that Arrow-Hurwicz method converges under restrictive conditions [11, 5, 23] and does not converge in general, see [18] for a divergent example. By using an extrapolation technique, Chambolle and Pock [5] proposed the following primal-dual algorithm (PDA)

$$\begin{cases} x_{n+1} = \mathrm{Prox}_{\tau g}(x_n - \tau K^\top y_n), \\ \bar{x}_{n+1} = x_{n+1} + \delta(x_{n+1} - x_n), \\ y_{n+1} = \mathrm{Prox}_{\sigma f^*}(y_n + \sigma K \bar{x}_{n+1}), \end{cases} \tag{1.7}$$

where $\delta \in (0, 1]$ is an extrapolation/inertial constant. For the case of $\delta = 1$, the convergence of (1.7) was established in [5, 19] under the condition $\tau\sigma L^2 < 1$, see also [11, 26, 6, 27] for further analysis of (1.7) and its variants.

Recently, by using a convex combination technique originally introduced by Malitsky [21] for solving monotone variational inequality problems, Chang and Yang [7] presented a golden ratio PDA (GRPDA), which iterates as

$$\begin{cases} z_{n+1} = \frac{\psi - 1}{\psi} x_n + \frac{1}{\psi} z_n, \\ x_{n+1} = \mathrm{Prox}_{\tau g}(z_{n+1} - \tau K^\top y_n), \\ y_{n+1} = \mathrm{Prox}_{\sigma f^*}(y_n + \sigma K x_{n+1}). \end{cases} \tag{1.8}$$

Here $\psi \in (1, \phi]$ determines the weight of the convex combination. Compared with (1.7), which converges when $\tau\sigma L^2 < 1$, an advantage of GRPDA is that it converges under the relaxed condition $\tau\sigma L^2 < \psi$. Ergodic convergence rate results similar to those in [5] have been established in [7].

1.3 Motivation and contributions

Let $(\bar{x}, \bar{y})$ be a saddle point of (1.1), which exists under mild conditions (see Assumption 2.1). Then, there hold $-K^\top \bar{y} \in \partial g(\bar{x})$ and $K\bar{x} \in \partial f^*(\bar{y})$, implying that

$$\begin{cases} P(x) := g(x) - g(\bar{x}) + \langle K^\top \bar{y}, x - \bar{x}\rangle \geq 0, & \forall x \in \mathbb{R}^q, \\ D(y) := f^*(y) - f^*(\bar{y}) - \langle K\bar{x}, y - \bar{y}\rangle \geq 0, & \forall y \in \mathbb{R}^p. \end{cases}$$

The so-called "primal-dual gap function" frequently used in the literature, see, e.g., [5, 22, 6, 7] and references therein, is given by $G(x, y) := P(x) + D(y)$ for $(x, y) \in \mathbb{R}^q \times \mathbb{R}^p$. Apparently, $G(x, y)$ is nonnegative for any (x, y). On the other hand, it is easy to show that $G(\tilde{x}, \tilde{y}) = 0$ for any saddle point $(\tilde{x}, \tilde{y})$ of (1.1). Therefore, it seems reasonable to take $G(\cdot)$ as a merit function to quantify the convergence rate of primal-dual algorithms. In particular, we have shown in [7] that, measured by this primal-dual gap function, GRPDA converges in the general convex case at the ergodic $\mathcal{O}(1/N)$ sublinear convergence rate, i.e., the sequence $\{(x_n, y_n)\}$ generated by (1.8) satisfies $G(\bar{x}_N, \bar{y}_N) \leq C/N$ for all $N \geq 1$, where $(\bar{x}_N, \bar{y}_N)$ is a weighted average of $\{(x_n, y_n) : n = 1, \ldots, N\}$ and $C > 0$ is some constant. Moreover, by modifying the algorithm properly, see [7] for details, this rate can be improved to $\mathcal{O}(1/N^2)$ when either g or f^* is strongly convex.

Unfortunately, the above primal-dual gap function could vanish at nonstationary points, which makes existing convergence rate results measured by this gap function less informative. In fact, it has been pointed out in [5] that this "vanishing at nonstationary point" phenomenon can happen even to the following enhanced primal-dual gap function

$$G_{B_1 \times B_2}(x, y) := \max_{y' \in B_2} \{g(x) + \langle y', Kx\rangle - f^*(y')\} - \min_{x' \in B_1} \{g(x') + \langle y, Kx'\rangle - f^*(y)\},$$

where $B_1 \times B_2$ is any set containing any saddle point of (1.1). In fact, it is clear that $G_{B_1 \times B_2}(x, y)$ reduces to $G(x, y)$ if $B_1 = \{\bar{x}\}$ and $B_2 = \{\bar{y}\}$. Furthermore, only when $B_1 = \mathbb{R}^q$ and $B_2 = \mathbb{R}^p$ will $G_{B_1 \times B_2}(x, y)$ reduce to $\mathcal{P}(x) - \mathcal{D}(y)$, i.e., the true primal and dual function value gap, which is a meaningful optimality measure in the sense that $G_{B_1 \times B_2}(x, y) = \mathcal{P}(x) - \mathcal{D}(y) = 0$ if and only if x and y are optimal solutions to (1.2) and (1.3), respectively.

Aiming to fix this defectiveness and resorting to a framework recently proposed by Sabach and Teboulle [25] for analyzing Lagrangian-based methods, we establish in this paper some new convergence rate results of GRPDA based on the equivalent optimization problems (1.2)-(1.5). Our contributions are summarized below.

- In the general convex case, we establish new ergodic $\mathcal{O}(1/N)$ convergence rate results for GRPDA, i.e., the scheme given in (1.8), which are quantified by the conventional measures of function value residual and constraint violation of (1.4).
- When either g or f^* is strongly convex, an accelerated GRPDA is constructed by using variable parameters. We show that the new algorithm converges at the faster $\mathcal{O}(1/N^2)$ ergodic rates, again, quantified by function value residual and constraint violation of (1.4).

- Under Lipschitz continuity assumption on f and g^*, we show that the proposed GRPDA ensures that the function value residuals of the unconstrained optimization problems (1.2) and (1.3), i.e., $\mathcal{P}(x)-\inf\mathcal{P}$ and $\sup\mathcal{D}-\mathcal{D}(y)$, respectively, converge at the ergodic convergence rate $\mathcal{O}(1/N)$ in the general convex case. Furthermore, it is shown that this rate can be improved to $\mathcal{O}(1/N^2)$ when either f or g^* is strongly convex, leading to a convergence rate guarantee of the true primal and dual function value gap, i.e., $\mathcal{P}(x)-\mathcal{D}(y)$.
- We carry out numerical experiments on the least absolute deviation and the LASSO problems, with comparisons to some state-of-the-art algorithms, to demonstrate the favorable performance of the proposed algorithms.

1.4 Organization

The rest of this paper is organized as follows. In Section 2, we summarize some preliminary results and make a technical assumption, which will be used in our analysis. Section 3 is devoted to the accelerated GRPDA and its convergence analysis when g is strongly convex, while the analysis for the general convex case is left to Section 4. Since the analysis for the general convex case is similar and much simpler than the strongly convex case, we only present the intermediate results and the main theorems, while their proofs will be omitted to reduce redundancy. Numerical results are given in Section 5 and, finally, conclusions are drawn in Section 6.

2 Preliminaries

In this section, we define more notation, make a technical assumption and summarize some useful preliminary results, which will be useful in our analysis.

Let $y\in\mathbb{R}^p$ be the Lagrange multiplier and $\sigma>0$ be a penalty parameter. The objective, the Lagrangian and the augmented Lagrangian functions of (1.4) are denoted respectively by

$$\Phi(x,w) := g(x)+f(w), \tag{2.1}$$

$$\mathcal{L}(x,w,y) := \Phi(x,w)+\langle y, Kx-w\rangle, \tag{2.2}$$

$$\mathcal{L}_\sigma(x,w,y) := \mathcal{L}(x,w,y)+\frac{\sigma}{2}\|Kx-w\|^2. \tag{2.3}$$

Throughout this paper, we make the following blanket assumption.

Assumption 2.1. Assume that the set of solutions of (1.2) is nonempty and, in addition, there exists $\tilde{x}\in\mathrm{ri}(\mathrm{dom}(g))$ such that $K\tilde{x}\in\mathrm{ri}(\mathrm{dom}(f))$.

For simplicity, in the rest of this paper, we let $\mathcal{X}:=\mathbb{R}^q\times\mathbb{R}^p\times\mathbb{R}^p$. Under Assumption 2.1, it follows from [24, Corollaries 28.2.2 and 28.3.1] that $(\bar{x},\bar{w})\in\mathbb{R}^q\times\mathbb{R}^p$ is a solution of (1.4) if and only if there exists an optimal solution $\bar{y}\in\mathbb{R}^p$ to the dual problem (1.3) such that $(\bar{x},\bar{w},\bar{y})$ is a saddle point of $\mathcal{L}(x,w,y)$, i.e.,

$$\mathcal{L}(\bar{x},\bar{w},y)\le\mathcal{L}(\bar{x},\bar{w},\bar{y})\le\mathcal{L}(x,w,\bar{y}) \ \text{ for all } \ (x,w,y)\in\mathcal{X}, \tag{2.4}$$

or equivalently, $-K^\top\bar{y}\in\partial g(\bar{x})$, $\bar{y}\in\partial f(\bar{w})$ and $K\bar{x}=\bar{w}$. Furthermore, it holds that

$$\inf\mathcal{P}=\Phi(\bar{x},\bar{w})=\mathcal{L}(\bar{x},\bar{w},\bar{y})=\sup\mathcal{D}.$$

The following lemmas will be used in our analysis. Lemmas 2.1 can be proved via elementary calculus, while Lemma 2.2 is well known as Fenchel's inequality, see [24].

Lemma 2.1. *Let $h : \mathbb{R}^m \to (-\infty, +\infty]$ be an extended real-valued closed proper and γ-strongly convex function with modulus $\gamma \geq 0$, i.e., it holds for all $x, y \in \mathbb{R}^m$ and $u \in \partial h(x)$ that*

$$h(y) \geq h(x) + \langle u, y - x\rangle + \frac{\gamma}{2}\|y - x\|^2.$$

Then for any $\tau > 0$ and $x \in \mathbb{R}^m$, it holds that $z = \mathrm{Prox}_{\tau h}(x)$ if and only if

$$h(y) \geq h(z) + \frac{1}{\tau}\langle x - z, y - z\rangle + \frac{\gamma}{2}\|y - z\|^2, \quad \forall\, y \in \mathbb{R}^m.$$

Lemma 2.2 (Fenchel's inequality)**.** *Let $h : \mathbb{R}^m \to (-\infty, \infty]$ be an extended real-valued closed proper convex function. Then, for any $x, y \in \mathbb{R}^m$, it holds that $h(x) + h^*(y) \geq \langle x, y\rangle$, and equality holds if and only if $x \in \partial h^*(y)$ or $y \in \partial h(x)$.*

3 An accelerated GRPDA and its analysis

In this section, we present an accelerated GRPDA and establish its convergence rate results when either g or f^* is strongly convex. Without loss of generality, we assume that g is γ-strongly convex. We will present the algorithm and analysis for the primal problem (1.4). When f^* is strongly convex, one can apply the proposed method to the dual problem (1.5).

3.1 Accelerated GRPDA

Recall that ϕ represents the golden ratio and $L = \|K\|$ is the operator norm of K. Below, we introduce our accelerated GRPDA.

Algorithm 3.1 (Accelerated GRPDA)**.**

Step 0. Let $\mu \in (0, 1)$, $\psi \in (1, \phi]$, $\sigma_0 = \rho = \frac{(1-\mu)(\psi-1)\gamma}{2L^2}$, $\tau_0 = \frac{(1-\mu)\psi}{\sigma_0 L^2}$. Choose $x_0 \in \mathbb{R}^q$ and $y_0 \in \mathbb{R}^p$. Set $z_0 = x_0$, $t_0 = 1$ and $n = 0$.

Step 1. Compute

$$z_{n+1} = \big(1 - 1/\psi\big)x_n + z_n/\psi, \tag{3.1a}$$
$$x_{n+1} = \mathrm{Prox}_{\tau_n g}(z_{n+1} - \tau_n K^\top y_n), \tag{3.1b}$$
$$w_{n+1} = \mathrm{Prox}_{f/\sigma_n}\big(y_n/\sigma_n + K x_{n+1}\big), \tag{3.1c}$$
$$y_{n+1} = y_n + \sigma_n(K x_{n+1} - w_{n+1}). \tag{3.1d}$$

Step 2. Update parameters by

$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}, \tag{3.2a}$$
$$\sigma_{n+1} = \rho t_{n+1}, \tag{3.2b}$$
$$\tau_{n+1} = \begin{cases} \frac{(1-\mu)\psi}{\sigma_{n+1}L^2 - (1-\mu)\gamma}, & \text{if } \ \sigma_{n+1}L^2 > (1-\mu)\gamma, \\ \frac{(1-\mu)\psi}{\sigma_{n+1}L^2}, & \text{otherwise.} \end{cases} \tag{3.2c}$$

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

We give several remarks on Algorithm 3.1. First, the strong convexity parameter γ is not only used in the analysis but also plays a role in the algorithm itself, which is different from [31], where golden ratio type splitting methods were studied in the presence of an extra smooth Lipschitzian term in (1.1). Second, by the definition of ρ and $\sigma_{n+1} = \rho t_{n+1}$, the condition $\sigma_{n+1}L^2 > (1-\mu)\gamma$ is equivalent to $t_{n+1} > 2/(\psi-1)$, and thus we have $\tau_{n+1} = \frac{(1-\mu)\psi}{\sigma_{n+1}L^2-(1-\mu)\gamma}$ for all $n \geq \lceil 4/(\psi-1) - 3\rceil$ since $t_{n+1} > t_n + 1/2 > (n+3)/2$. Third, Algorithm 3.1 can indeed be viewed as a variant of (1.8) with variable stepsizes τ_n and σ_n determined by (3.2). This can be shown by using the Moreau's decomposition $y = \mathrm{Prox}_{f/\sigma}(y) + \frac{1}{\sigma}\mathrm{Prox}_{\sigma f^*}(\sigma y)$, for any $\sigma > 0$ and $y \in \mathbb{R}^p$. In fact, by combining $w_{n+1} = \mathrm{Prox}_{f/\sigma_n}\big(y_n/\sigma_n + Kx_{n+1}\big)$ and $y_{n+1} = y_n + \sigma_n(Kx_{n+1} - w_{n+1})$ given in (3.1c) and (3.1d), respectively, we obtain $y_{n+1} = \mathrm{Prox}_{\sigma_n f^*}(y_n + \sigma_n K x_{n+1})$. Then, the statement above is clear by comparing (3.1) with (1.8). We will show that due to the strong convexity of g and the adaptive choice of parameters in (3.2), Algorithm 3.1 achieves accelerated $\mathcal{O}(1/N^2)$ convergence rate on some auxiliary sequences, which we define next. In fact, one can always set $\tau_{n+1} = \frac{(1-\mu)\psi}{\sigma_{n+1}L^2}$ for $n \geq 0$ in Algorithm 3.1 and the accelerated $\mathcal{O}(1/N^2)$ convergence results given in Section 3.2 will remain valid. Only slight modifications to the analysis are required. An advantage of this choice of τ_{n+1} is that the strong convexity parameter γ does not need to be known in advance. The adaptive choice of algorithmic parameters in (3.2) is common in accelerating first order methods. Among others, see, e.g., [25]. The value of τ_{n+1} given in (3.2c) is larger when $\sigma_{n+1}L^2 > (1-\mu)\gamma$, which is helpful in practice.

Let $\widetilde{w}_0 \in \mathbb{R}^p$ be arbitrarily chosen and set $(\widetilde{x}_0, \widetilde{y}_0, \widehat{y}_0) = (x_0, y_0, y_0)$. For convergence analysis, we define auxiliary sequences $\{(\widetilde{x}_n, \widetilde{w}_n, \widetilde{y}_n, \widehat{y}_n) : n \geq 0\}$ recursively as follows:

$$\begin{pmatrix} \widetilde{x}_{n+1} \\ \widetilde{w}_{n+1} \\ \widehat{y}_{n+1} \end{pmatrix} = (1 - t_n^{-1}) \begin{pmatrix} \widetilde{x}_n \\ \widetilde{w}_n \\ \widehat{y}_n \end{pmatrix} + t_n^{-1} \begin{pmatrix} x_{n+1} \\ w_{n+1} \\ y_{n+1} \end{pmatrix}, \tag{3.3}$$

$$\widetilde{y}_{n+1} = \widetilde{y}_n + \mu\sigma_n(Kx_{n+1} - w_{n+1}). \tag{3.4}$$

We emphasize that the auxiliary sequences $\{(\widetilde{x}_n, \widetilde{w}_n, \widetilde{y}_n, \widehat{y}_n) : n \geq 0\}$ are used only in the convergence rate analysis and need not to be computed in practice. Apparently, their computations only involve some scalar-vector multiplications and vector additions, which are negligible compared to the dominant computations of the algorithm. We next establish a useful relation.

Lemma 3.2. *For all $n \geq 1$, there holds*

$$y_n = \widetilde{y}_n + (1-\mu)\rho t_{n-1}^2(K\widetilde{x}_n - \widetilde{w}_n). \tag{3.5}$$

Proof. It follows from (3.1d), (3.2b) and (3.4) that

$$y_{n+1} = y_n + \widetilde{y}_{n+1} - \widetilde{y}_n + (1-\mu)\rho t_n(Kx_{n+1} - w_{n+1}), \quad \forall\, n \geq 0. \tag{3.6}$$

Recall that $t_n^2 - t_n = t_{n-1}^2$ for $n \geq 1$. Multiplying both sides of (3.3) by t_n^2 and noting the linearity of K, it is easy to deduce

$$t_n(Kx_{n+1} - w_{n+1}) = t_n^2(K\widetilde{x}_{n+1} - \widetilde{w}_{n+1}) - t_{n-1}^2(K\widetilde{x}_n - \widetilde{w}_n), \quad \forall\, n \geq 0.$$

This together with (3.6) implies for all $n \geq 1$ that

$$y_{n+1} - \widetilde{y}_{n+1} - (1-\mu)\rho t_n^2(K\widetilde{x}_{n+1} - \widetilde{w}_{n+1}) = y_n - \widetilde{y}_n - (1-\mu)\rho t_{n-1}^2(K\widetilde{x}_n - \widetilde{w}_n). \tag{3.7}$$

Since $t_0 = 1$, we have from (3.3) that $(\widetilde{x}_1, \widetilde{w}_1) = (x_1, w_1)$. Furthermore, by noting $\sigma_0 = \rho$, $\widetilde{y}_0 = y_0$ and $\widetilde{y}_1 = \widetilde{y}_0 + \mu\sigma_0(Kx_1 - w_1)$, it follows from (3.1d) that

$$y_1 - \widetilde{y}_1 - (1-\mu)\rho t_0^2(K\widetilde{x}_1 - \widetilde{w}_1) = (1-\mu)\sigma_0(Kx_1 - w_1) - (1-\mu)\rho(Kx_1 - w_1) = 0,$$

which, together with (3.7), implies that (3.5) holds for all $n \geq 1$. □

Before carrying out convergence analysis for Algorithm 3.1, we summarize some useful properties of the sequence $\{t_n : n \geq 0\}$ defined in (3.2a).

Lemma 3.3. *Let $t_0 = 1$ and $t_n = \frac{1+\sqrt{1+4t_{n-1}^2}}{2}$ for $n \geq 1$. Then*

(i) *t_n is monotonically increasing, in particular, $t_n \geq t_{n-1} + \frac{1}{2} \geq \frac{n+2}{2}$,*

(ii) *$t_n^2 - t_{n-1}^2 = t_n$,*

(iii) *$t_n^2 \leq t_{n-1}^2 + 2t_{n-1}$,*

(iv) *$t_{n+1} = t_n + \frac{t_{n+1}}{t_n + t_{n+1}}$,*

(v) *$t_n/t_{n-1} \in (1, \frac{1+\sqrt{5}}{2}]$ and $t_n - t_{n-1}$ are monotonically decreasing, and*

(vi) *for any $\kappa > 1$, if $n \geq \lfloor 2\kappa \rfloor$, then $t_{n-1} > \kappa$ and $t_n/t_{n-1} < \frac{1+\sqrt{1+4\kappa^2}}{2\kappa}$.*

Proof. Items (i) to (iv) are easy to verify either by direct calculation or reduction, see also [25]. We omit the details. For (v), direct calculation shows that $t_n/t_{n-1} \in (1, \frac{1+\sqrt{5}}{2}]$, $t_n/t_{n-1} = (1 - 1/t_n)^{-1/2}$ and $t_n - t_{n-1} = \left((t_n/t_{n-1})^{-1} + 1\right)^{-1}$. Since t_n is monotonically increasing, t_n/t_{n-1} is monotonically decreasing, and so is $t_n - t_{n-1}$. Finally, if $n \geq \lfloor 2\kappa \rfloor$, then $t_{n-1} \geq \frac{n+1}{2} > \kappa$. Since $t_n/t_{n-1} > 1$, $t_n/t_{n-1} < \frac{1+\sqrt{1+4\kappa^2}}{2\kappa}$ follows from $\kappa(t_n/t_{n-1} - 1) < t_{n-1}(t_n/t_{n-1} - 1) = \left((t_n/t_{n-1})^{-1} + 1\right)^{-1}$. □

3.2 Convergence results

In this section, we establish convergence rate results for Algorithm 3.1. First, we present some useful lemmas on the sequence $\{(z_n, x_n, w_n, y_n) : n \geq 0\}$ generated by Algorithm 3.1 and the auxiliary sequence $\{(\widetilde{x}_n, \widetilde{w}_n, \widetilde{y}_n, \widehat{y}_n) : n \geq 0\}$ defined in (3.3)-(3.4). In the following, we let $\delta_n := \tau_n/\tau_{n-1}$.

Lemma 3.4. *For any $(x, w, y) \in \mathcal{X}$ and $n \geq 1$, we have*

$$\begin{aligned}
&\mathcal{L}_{\sigma_{n-1}}(x_n, w_n, y_{n-1}) - \mathcal{L}(x, w, y) \\
\leq &\frac{1}{\tau_n}\langle x_{n+1} - z_{n+1},\ x - x_{n+1}\rangle + \frac{\psi\delta_n}{\tau_n}\langle x_n - z_{n+1},\ x_{n+1} - x_n\rangle \\
&- \frac{\gamma}{2}\|x - x_{n+1}\|^2 - \frac{\gamma}{2}\|x_n - x_{n+1}\|^2 + \langle y_n - y, Kx - w\rangle \\
&+ \sigma_{n-1}\langle Kx_n - w_n, w_n - Kx_{n+1}\rangle + \frac{\sigma_{n-1}}{2}\|Kx_n - w_n\|^2. \qquad (3.8)
\end{aligned}$$

Proof. Let $(x,w,y)\in\mathcal{X}$ and $n\geq 1$ be arbitrarily fixed. It follows from (3.1b), Lemma 2.1 and the γ-strong convexity of g that

$$\begin{aligned} g(x_{n+1})-g(x)\leq&\frac{1}{\tau_n}\langle x_{n+1}-z_{n+1}+\tau_n K^\top y_n,\ x-x_{n+1}\rangle-\frac{\gamma}{2}\|x-x_{n+1}\|^2\\ =&\frac{1}{\tau_n}\langle x_{n+1}-z_{n+1},\ x-x_{n+1}\rangle-\frac{\gamma}{2}\|x-x_{n+1}\|^2\\ &+\langle y_{n-1}+\sigma_{n-1}(Kx_n-w_n),\ Kx-Kx_{n+1}\rangle, \end{aligned}\tag{3.9}$$

$$f(w_n)-f(w)\leq-\big\langle y_{n-1}+\sigma_{n-1}(Kx_n-w_n),\ w-w_n\big\rangle.\tag{3.10}$$

Similar to (3.9), we have

$$\begin{aligned} g(x_n)-g(x_{n+1})\leq&\frac{1}{\tau_{n-1}}\langle x_n-z_n+\tau_{n-1}K^\top y_{n-1},\ x_{n+1}-x_n\rangle-\frac{\gamma}{2}\|x_n-x_{n+1}\|^2\\ =&\langle\frac{\psi\delta_n}{\tau_n}(x_n-z_{n+1})+K^\top y_{n-1},\ x_{n+1}-x_n\rangle-\frac{\gamma}{2}\|x_n-x_{n+1}\|^2, \end{aligned}\tag{3.11}$$

where the equality follows from $\delta_n=\tau_n/\tau_{n-1}$ and $x_n-z_n=\psi(x_n-z_{n+1})$ (follows from (3.1a)). By the definitions of $\mathcal{L}_\sigma(\cdot)$ and $\mathcal{L}(\cdot)$ in (2.3) and (2.2), respectively, direct calculations show that the addition of (3.9), (3.10) and (3.11) gives

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

where the last equality used $y_n=y_{n-1}+\sigma_{n-1}(Kx_n-w_n)$. This completes the proof. □

In the following, we let

$$n_0:=\min\{n\in\mathbb{N}:t_{n-1}>2/(\psi-1)\}.\tag{3.12}$$

Since t_n is monotonically increasing, by (3.2c), $\rho=(1-\mu)(\psi-1)\gamma/(2L^2)$ and the definition of n_0 in (3.12), we have $\tau_{n-1}=(1-\mu)\psi/\big(\sigma_{n-1}L^2-(1-\mu)\gamma\big)$ for all $n\geq n_0$.

Lemma 3.5. *For any $(x,w,y)\in\mathcal{X}$ and $n\geq n_0$, we have*

$$\begin{aligned} \mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y)-\mathcal{L}(x,w,y)\leq&\frac{1}{\tau_n}\Delta_{P_n}(x,z_{n+1},z_{n+2})+\frac{1}{\mu\sigma_{n-1}}\Delta(y,\widetilde{y}_{n-1},\widetilde{y}_n)\\ &+\langle y_n-y,Kx-w\rangle-\frac{\gamma}{2}\|x-z_{n+2}\|^2-(1-\mu)\rho t_{n-2}^2\langle K\widetilde{x}_{n-1}-\widetilde{w}_{n-1},Kx_n-w_n\rangle, \end{aligned}\tag{3.13}$$

where the notation $\Delta_{P_n}(\cdot)$ *is defined in* (1.6) *with*

$$P_n = \frac{\psi}{\psi-1}(1+\frac{\gamma\tau_n}{\psi})I \succ 0. \tag{3.14}$$

Proof. Let $(x,w,y)\in\mathcal{X}$ and $n\geq n_0$ be arbitrarily fixed, where n_0 is defined in (3.12). First, it follows from the Cauchy-Schwartz inequality that

$$\begin{aligned}\langle Kx_n-w_n, w_n-Kx_{n+1}\rangle =& -\|Kx_n-w_n\|^2+\langle Kx_n-w_n, Kx_n-Kx_{n+1}\rangle\\ \leq& -\|Kx_n-w_n\|^2+\frac{1}{2(1-\mu)}\|x_{n+1}-x_n\|^2_{K^\top K}\\ &+\frac{1-\mu}{2}\|Kx_n-w_n\|^2\\ =&\frac{1}{2(1-\mu)}\|x_{n+1}-x_n\|^2_{K^\top K}-\frac{1+\mu}{2}\|Kx_n-w_n\|^2.\end{aligned}$$

Plugging the above inequality into (3.8) and using the three-points identity

$$\langle u-v,u-w\rangle=\frac{1}{2}\|u-v\|^2+\frac{1}{2}\|u-w\|^2-\frac{1}{2}\|v-w\|^2,$$

which holds for any vectors u,v and w of the same lengths, we obtain

$$\begin{aligned}&\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y_{n-1})-\mathcal{L}(x,w,y)\\ \leq&\frac{1}{2\tau_n}\Big(\|z_{n+1}-x\|^2-\|x_{n+1}-z_{n+1}\|^2-\|x_{n+1}-x\|^2\Big)\\ &+\frac{\psi\delta_n}{2\tau_n}\Big(\|z_{n+1}-x_{n+1}\|^2-\|z_{n+1}-x_n\|^2-\|x_{n+1}-x_n\|^2\Big)+\langle y_n-y,Kx-w\rangle\\ &+\frac{\sigma_{n-1}}{2(1-\mu)}\|x_{n+1}-x_n\|^2_{K^\top K}-\frac{\mu\sigma_{n-1}}{2}\|Kx_n-w_n\|^2-\frac{\gamma}{2}\|x-x_{n+1}\|^2-\frac{\gamma}{2}\|x_n-x_{n+1}\|^2\\ \leq&\frac{1}{2\tau_n}\Big(\|z_{n+1}-x\|^2-\|x_{n+1}-z_{n+1}\|^2\Big)-\frac{1+\gamma\tau_n}{2\tau_n}\|x_{n+1}-x\|^2\\ &+\frac{\psi\delta_n}{2\tau_n}\Big(\|z_{n+1}-x_{n+1}\|^2-\|z_{n+1}-x_n\|^2\Big)-\frac{\mu\sigma_{n-1}}{2}\|Kx_n-w_n\|^2+\langle y_n-y,Kx-w\rangle,\end{aligned} \tag{3.15}$$

where the second inequality follows from

$$n\geq n_0\Longrightarrow\tau_{n-1}=\frac{(1-\mu)\psi}{\sigma_{n-1}L^2-(1-\mu)\gamma}\Longrightarrow\Big(\frac{\psi\delta_n}{\tau_n}+\gamma\Big)I-\frac{\sigma_{n-1}}{1-\mu}K^\top K\succeq 0.$$

From (3.1a), we have $x_{n+1}=\frac{\psi}{\psi-1}z_{n+2}-\frac{1}{\psi-1}z_{n+1}$ and $z_{n+2}-z_{n+1}=\frac{\psi-1}{\psi}(x_{n+1}-z_{n+1})$. It then follows from $\|\alpha u+(1-\alpha)v\|^2=\alpha\|u\|^2+(1-\alpha)\|v\|^2-\alpha(1-\alpha)\|u-v\|^2$ for any scalar $\alpha\in\mathbb{R}$ and vectors u and v of the same lengths that

$$\begin{aligned}\|x_{n+1}-x\|^2=&\frac{\psi}{\psi-1}\|z_{n+2}-x\|^2-\frac{1}{\psi-1}\|z_{n+1}-x\|^2+\frac{\psi}{(\psi-1)^2}\|z_{n+2}-z_{n+1}\|^2\\ =&\frac{\psi}{\psi-1}\|z_{n+2}-x\|^2-\frac{1}{\psi-1}\|z_{n+1}-x\|^2+\frac{1}{\psi}\|x_{n+1}-z_{n+1}\|^2.\end{aligned} \tag{3.16}$$

By plugging (3.16) into (3.15) and omitting the non-negative term $\|z_{n+1}-x_n\|^2$, we obtain

$$\begin{aligned}
&\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y_{n-1})-\mathcal{L}(x,w,y)\\
\leq&\frac{1}{2\tau_n}\frac{\psi}{\psi-1}\big((1+\gamma\tau_n/\psi)\|z_{n+1}-x\|^2-(1+\gamma\tau_n)\|z_{n+2}-x\|^2\big)+\langle y_n-y,Kx-w\rangle\\
&-\frac{1}{2\tau_n}\big((1+\gamma\tau_n)/\psi-\psi\delta_n+1\big)\|z_{n+1}-x_{n+1}\|^2-\frac{\mu\sigma_{n-1}}{2}\|Kx_n-w_n\|^2\\
\leq&\frac{1}{\tau_n}\Delta_{P_n}(x,z_{n+1},z_{n+2})-\frac{\mu\sigma_{n-1}}{2}\|Kx_n-w_n\|^2-\frac{\gamma}{2}\|x-z_{n+2}\|^2+\langle y_n-y,Kx-w\rangle,
\end{aligned}\tag{3.17}$$

where the second inequality follows from $1+\gamma\tau_n=(1+\gamma\tau_n/\psi)\big(1+\gamma\tau_n(\psi-1)/(\psi+\gamma\tau_n)\big)$, the definition of $P_n\succ 0$ in (3.14), $(1+\gamma\tau_n)/\psi-\psi\delta_n+1>1/\psi-\psi+1\geq 0$ since for $n\geq n_0$ we have from (3.2c) that $\delta_n=\tau_n/\tau_{n-1}=\big(\sigma_{n-1}L^2-(1-\mu)\gamma\big)/\big(\sigma_nL^2-(1-\mu)\gamma\big)<1$ and $\psi\in(1,\phi]$, and the definition of the notation $\Delta_{P_n}(\cdot)$ in (1.6).

From (3.4) and the three-points identity $\langle u-v,u-w\rangle=\frac{1}{2}\|u-v\|^2+\frac{1}{2}\|u-w\|^2-\frac{1}{2}\|v-w\|^2$ again, we obtain

$$\begin{aligned}
\langle y-\widetilde{y}_{n-1},Kx_n-w_n\rangle&=\frac{1}{\mu\sigma_{n-1}}\langle y-\widetilde{y}_{n-1},\widetilde{y}_n-\widetilde{y}_{n-1}\rangle\\
&=\frac{1}{\mu\sigma_{n-1}}\Delta(y,\widetilde{y}_{n-1},\widetilde{y}_n)+\frac{1}{2\mu\sigma_{n-1}}\|\widetilde{y}_n-\widetilde{y}_{n-1}\|^2\\
&=\frac{1}{\mu\sigma_{n-1}}\Delta(y,\widetilde{y}_{n-1},\widetilde{y}_n)+\frac{\mu\sigma_{n-1}}{2}\|Kx_n-w_n\|^2.
\end{aligned}\tag{3.18}$$

Here $\Delta(\cdot)=\Delta_I(\cdot)$. Considering (3.5), we deduce from (3.18) that

$$\begin{aligned}
\langle y-y_{n-1},Kx_n-w_n\rangle=&\langle y-\widetilde{y}_{n-1},Kx_n-w_n\rangle-(1-\mu)\rho t_{n-2}^2\langle K\widetilde{x}_{n-1}-\widetilde{w}_{n-1},Kx_n-w_n\rangle\\
=&\frac{1}{\mu\sigma_{n-1}}\Delta(y,\widetilde{y}_{n-1},\widetilde{y}_n)+\frac{\mu\sigma_{n-1}}{2}\|Kx_n-w_n\|^2\\
&-(1-\mu)\rho t_{n-2}^2\langle K\widetilde{x}_{n-1}-\widetilde{w}_{n-1},Kx_n-w_n\rangle.
\end{aligned}\tag{3.19}$$

By the definition of $\mathcal{L}_\sigma(\cdot)$ in (2.3), we have

$$\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y_{n-1})+\langle y-y_{n-1},Kx_n-w_n\rangle=\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y).$$

Then, the proof of (3.13) is completed by adding (3.19) to (3.17). □

Lemma 3.6. *For any $(x,w,y)\in\mathcal{X}$ and $n\geq n_0$, we have*

$$\begin{aligned}
t_{n-1}^2S^n_{(1-\mu)\rho t_{n-1}^2}-t_{n-2}^2S^{n-1}_{(1-\mu)\rho t_{n-2}^2}\leq&\frac{t_{n-1}}{\tau_n}\Delta_{P_n}(x,z_{n+1},z_{n+2})+\frac{1}{\mu\rho}\Delta(y,\widetilde{y}_{n-1},\widetilde{y}_n)\\
&-\frac{t_{n-1}\gamma}{2}\|x-z_{n+2}\|^2,
\end{aligned}\tag{3.20}$$

where P_n is defined as in (3.14) *and*

$$S^n_\beta:=S^n_\beta(x,w,y)=\mathcal{L}_\beta(\widetilde{x}_n,\widetilde{w}_n,y)-\mathcal{L}(x,w,\widehat{y}_n).\tag{3.21}$$

Proof. Let $(x,w,y)\in\mathcal{X}$ and $n\geq n_0$ be arbitrarily fixed and recall that $\Phi(\cdot,\cdot)$ is defined in (2.1). From the linearity of K and the convexity of $\Phi(\cdot,\cdot)$, we derive from (3.3) that

$$\langle y, K\widetilde{x}_n-\widetilde{w}_n\rangle=(1-t_{n-1}^{-1})\langle y, K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}\rangle+t_{n-1}^{-1}\langle y, Kx_n-w_n\rangle, \tag{3.22a}$$

$$\Phi(\widetilde{x}_n,\widetilde{w}_n)\leq(1-t_{n-1}^{-1})\Phi(\widetilde{x}_{n-1},\widetilde{w}_{n-1})+t_{n-1}^{-1}\Phi(x_n,w_n). \tag{3.22b}$$

Multiplying both sides of the above relations by t_{n-1}^2 and recalling that $t_{n-1}^2-t_{n-1}=t_{n-2}^2$, we obtain

$$\begin{aligned}&t_{n-1}^2\langle y, K\widetilde{x}_n-\widetilde{w}_n\rangle-t_{n-2}^2\langle y, K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}\rangle=t_{n-1}\langle y, Kx_n-w_n\rangle,\\ &t_{n-1}^2\big(\Phi(\widetilde{x}_n,\widetilde{w}_n)-\Phi(x,w)\big)-t_{n-2}^2\big(\Phi(\widetilde{x}_{n-1},\widetilde{w}_{n-1})-\Phi(x,w)\big)\leq t_{n-1}\big(\Phi(x_n,w_n)-\Phi(x,w)\big).\end{aligned}$$

Adding the above two relations and using the definition of $\mathcal{L}(\cdot)$ in (2.2), we arrive at

$$\begin{aligned}t_{n-1}^2\big(\mathcal{L}(\widetilde{x}_n,\widetilde{w}_n,y)-\mathcal{L}(x,w,y)\big)-t_{n-2}^2\big(\mathcal{L}(\widetilde{x}_{n-1},\widetilde{w}_{n-1},y)-\mathcal{L}(x,w,y)\big)\\ \leq t_{n-1}\big(\mathcal{L}(x_n,w_n,y)-\mathcal{L}(x,w,y)\big).\end{aligned} \tag{3.23}$$

Using (3.3) again to obtain

$$\begin{aligned}\|K\widetilde{x}_n-\widetilde{w}_n\|^2=&(1-t_{n-1}^{-1})^2\|K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}\|^2+t_{n-1}^{-2}\|Kx_n-w_n\|^2\\ &+2t_{n-1}^{-1}(1-t_{n-1}^{-1})\langle K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}, Kx_n-w_n\rangle.\end{aligned} \tag{3.24}$$

Multiplying both sides of the above equality by $\rho t_{n-1}^4/2$ and recalling $t_{n-1}^2-t_{n-1}=t_{n-2}^2$ yield

$$\begin{aligned}&\rho t_{n-1}^4/2\|K\widetilde{x}_n-\widetilde{w}_n\|^2-\rho t_{n-2}^4/2\|K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}\|^2\\ &\quad=\rho t_{n-1}^2/2\|Kx_n-w_n\|^2+\rho t_{n-1}t_{n-2}^2\langle K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}, Kx_n-w_n\rangle.\end{aligned} \tag{3.25}$$

Therefore, by adding (3.23) to (3.25), using the definition of S_β^n in (3.21) and

$$t_{n-1}^2\langle\widehat{y}_n-y, Kx-w\rangle-t_{n-2}^2\langle\widehat{y}_{n-1}-y, Kx-w\rangle=t_{n-1}\langle y_n-y, Kx-w\rangle, \tag{3.26}$$

we deduce that

$$\begin{aligned}t_{n-1}^2S_{\rho t_{n-1}^2}^n-t_{n-2}^2S_{\rho t_{n-2}^2}^{n-1}\leq&t_{n-1}(\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y)-\mathcal{L}(x,w,y))-t_{n-1}\langle y_n-y, Kx-w\rangle\\ &+\rho t_{n-1}t_{n-2}^2\langle K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}, Kx_n-w_n\rangle.\end{aligned}$$

Finally, multiplying both sides of the above inequality by $(1-\mu)$ and adding to (3.13) multiplied by t_{n-1} yield

$$\begin{aligned}&(1-\mu)\big(t_{n-1}^2S_{\rho t_{n-1}^2}^n-t_{n-2}^2S_{\rho t_{n-2}^2}^{n-1}\big)\\ \leq&t_{n-1}(\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y)-\mathcal{L}(x,w,y))-(1-\mu)t_{n-1}\langle y_n-y, Kx-w\rangle\\ &+(1-\mu)\rho t_{n-1}t_{n-2}^2\langle K\widetilde{x}_{n-1}-\widetilde{w}_{n-1}, Kx_n-w_n\rangle-\mu t_{n-1}(\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y)-\mathcal{L}(x,w,y))\\ \leq&\frac{t_{n-1}}{\tau_n}\Delta_{P_n}(x,z_{n+1},z_{n+2})+\frac{t_{n-1}}{\mu\sigma_{n-1}}\Delta(y,\widetilde{y}_{n-1},\widetilde{y}_n)-\frac{t_{n-1}\gamma}{2}\|x-z_{n+2}\|^2\\ &-\mu t_{n-1}\big(\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y)-\mathcal{L}(x,w,y)-\langle y_n-y, Kx-w\rangle\big).\end{aligned} \tag{3.27}$$

Note that

$$\begin{aligned}
&t_{n-1}\big(\mathcal{L}_{\sigma_{n-1}}(x_n,w_n,y)-\mathcal{L}(x,w,y)-\langle y_n-y,Kx-w\rangle\big)\\
\geq&t_{n-1}\big(\mathcal{L}(x_n,w_n,y)-\mathcal{L}(x,w,y)-\langle y_n-y,Kx-w\rangle\big)\\
\geq&t_{n-1}^2\big(\mathcal{L}(\widetilde{x}_n,\widetilde{w}_n,y)-\mathcal{L}(x,w,\widehat{y}_n)\big)-t_{n-2}^2\big(\mathcal{L}(\widetilde{x}_{n-1},\widetilde{w}_{n-1},y)-\mathcal{L}(x,w,\widehat{y}_{n-1})\big),
\end{aligned}$$

where the last inequality follows from (3.23) and (3.26). Combining the above inequality with (3.27) and considering $\sigma_{n-1}=\rho t_{n-1}$ and the definition of S_β^n in (3.21), we obtain (3.20). Note that compared to (3.27), the subindex of S^n in (3.20) has been multiplied by $(1-\mu)$. This completes the proof. □

Now, we are ready to establish the main convergence result of the accelerated GRPDA.

Theorem 3.7. *Let $(\bar{x},\bar{w},\bar{y})$ be a saddle point of $\mathcal{L}(\cdot)$ satisfying (2.4) and $c>0$ be a constant such that $c\geq 2\|\bar{y}\|$. Then, there exist constants $C_1,C_2>0$ such that for any $N\geq n_0$ there hold*

$$|\Phi(\widetilde{x}_N,\widetilde{w}_N)-\Phi(\bar{x},\bar{w})|\leq\frac{C_1}{(N+1)^2},\quad \|K\widetilde{x}_N-\widetilde{w}_N\|\leq\frac{2C_1}{c(N+1)^2}$$
$$\text{and}\quad \|x_{N+1}-\bar{x}\|\leq\frac{C_2}{N+2}.$$

Proof. Let $n\geq n_0$. Then, we have $\tau_i=\frac{(1-\mu)\psi}{\sigma_i L^2-(1-\mu)\gamma}$ for $i=n,n+1$. Define

$$l_n:=\frac{1}{2\tau_n}\frac{\psi+\gamma\tau_n}{\psi-1}\geq 0\ \text{ and }\ h_n:=\frac{\psi+\psi\gamma\tau_n}{\psi+\gamma\tau_{n+1}}\frac{\tau_{n+1}}{\tau_n}=\frac{\rho L^2t_n+(1-\mu)(\psi-1)\gamma}{\rho L^2t_{n+1}}.\quad(3.28)$$

Then, it is easy to verify from the notation $\Delta_{P_n}(\cdot)$ defined in (1.6) and P_n defined in (3.14) that

$$\frac{1}{\tau_n}\Delta_{P_n}(\bar{x},z_{n+1},z_{n+2})-\frac{\gamma}{2}\|\bar{x}-z_{n+2}\|^2=l_n\|\bar{x}-z_{n+1}\|^2-l_{n+1}h_n\|\bar{x}-z_{n+2}\|^2.\quad(3.29)$$

Let $\pi(n)=\frac{t_n(t_{n+1}-t_{n-1})}{t_{n-1}}$ for all $n\geq 1$. It follows from (v) of Lemma 3.3 that $\pi(n)$ is nonincreasing. Thus, $\pi(n)\leq\pi(1)=t_1(t_2-1)\approx 1.9312<2$, and by further considering the definition of ρ, we deduce

$$\frac{\rho L^2}{1-\mu}\frac{t_n(t_{n+1}-t_{n-1})}{(\psi-1)t_{n-1}}<\gamma.\quad(3.30)$$

Then, it is elementary to show from (3.30) and the definition of h_n in (3.28) that $h_n\geq t_n/t_{n-1}$. As a result, it follows directly from (3.20) and (3.29) that

$$\begin{aligned}
&t_{n-1}^2S^n_{(1-\mu)\rho t_{n-1}^2}-t_{n-2}^2S^{n-1}_{(1-\mu)\rho t_{n-2}^2}\\
\leq&t_{n-1}l_n\|\bar{x}-z_{n+1}\|^2-t_nl_{n+1}\|\bar{x}-z_{n+2}\|^2+\frac{1}{2\mu\rho}\left(\|y-\widetilde{y}_{n-1}\|^2-\|y-\widetilde{y}_n\|^2\right).
\end{aligned}\quad(3.31)$$

By the monotonicity of $\{t_n\}$ and the definition of n_0 in (3.12), we can verify via computer that $n_0\geq 5$. Summing (3.31) for all $n=n_0,\ldots,N$ and dropping some negative terms, we obtain

$$\begin{aligned}
&t_{N-1}^2S^N_{(1-\mu)\rho t_{N-1}^2}-t_{n_0-2}^2S^{n_0-1}_{(1-\mu)\rho t_{n_0-2}^2}\\
&\qquad\leq t_{n_0-1}l_{n_0}\|\bar{x}-z_{n_0+1}\|^2-t_Nl_{N+1}\|\bar{x}-z_{N+2}\|^2+\frac{1}{2\mu\rho}\|y-\widetilde{y}_{n_0-1}\|^2.
\end{aligned}\quad(3.32)$$

Set $(x,w)=(\bar{x},\bar{w})$ in $S^n_{(1-\mu)\rho t^2_{n-1}}$ defined as in (3.21). Then, by noting $K\bar{x}=\bar{w}$ we obtain

$$\begin{aligned} S^N_{(1-\mu)\rho t^2_{N-1}} = S^N_{(1-\mu)\rho t^2_{N-1}}(\bar{x},\bar{w},y) &= \mathcal{L}_{(1-\mu)\rho t^2_{N-1}}(\widetilde{x}_N,\widetilde{w}_N,y) - \mathcal{L}(\bar{x},\bar{w},\widehat{y}_N) \\ &\geq \Phi(\widetilde{x}_N,\widetilde{w}_N) + \langle y, K\widetilde{x}_N - \widetilde{w}_N\rangle - \Phi(\bar{x},\bar{w}). \end{aligned}$$

Recall that $\Phi(\cdot,\cdot)$ is defined in (2.1). Therefore, (3.32) implies

$$\begin{aligned} & t^2_{N-1}\big(\Phi(\widetilde{x}_N,\widetilde{w}_N) + \langle y, K\widetilde{x}_N - \widetilde{w}_N\rangle - \Phi(\bar{x},\bar{w})\big) \\ \leq & t_{n_0-1}l_{n_0}\|\bar{x}-z_{n_0+1}\|^2 - t_N l_{N+1}\|\bar{x}-z_{N+2}\|^2 + \frac{1}{2\mu\rho}\|y-\widetilde{y}_{n_0-1}\|^2 + t^2_{n_0-2}S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}} \qquad (3.33) \\ \leq & t_{n_0-1}l_{n_0}\|\bar{x}-z_{n_0+1}\|^2 + \frac{1}{2\mu\rho}\|y-\widetilde{y}_{n_0-1}\|^2 + t^2_{n_0-2}S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}}. \qquad (3.34) \end{aligned}$$

Note that $S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}} = S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}}(\bar{x},\bar{w},y)$ also depends on y and is given by

$$S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}} = \mathcal{L}_{(1-\mu)\rho t^2_{n_0-2}}(\widetilde{x}_{n_0-1},\widetilde{w}_{n_0-1},y) - \mathcal{L}(\bar{x},\bar{w},\widehat{y}_n) = \langle y, K\widetilde{x}_{n_0-1} - \widetilde{w}_{n_0-1}\rangle + C,$$

with $C := \Phi(\widetilde{x}_{n_0-1},\widetilde{w}_{n_0-1}) + \frac{(1-\mu)\rho t^2_{n_0-2}}{2}\|K\widetilde{x}_{n_0-1}-\widetilde{w}_{n_0-1}\|^2 - \Phi(\bar{x},\bar{w})$ is a constant. By taking the maximum of both sides of (3.34) over $\|y\|\leq c$ and noting that $t^2_{N-1} > (N+1)^2/4$ (see (i) of Lemma 3.3), we obtain

$$\Phi(\widetilde{x}_N,\widetilde{w}_N) - \Phi(\bar{x},\bar{w}) + c\|K\widetilde{x}_N - \widetilde{w}_N\| \leq C_1/(N+1)^2, \qquad (3.35)$$

where $C_1 := 4C'$ with $C'>0$ given by

$$C' := t_{n_0-1}l_{n_0}\|\bar{x}-z_{n_0+1}\|^2 + \frac{1}{2\mu\rho}(c+\|\widetilde{y}_{n_0-1}\|)^2 + t^2_{n_0-2}\big(c\|K\widetilde{x}_{n_0-1}-\widetilde{w}_{n_0-1}\| + |C|\big). \qquad (3.36)$$

Then we obviously have $\Phi(\widetilde{x}_N,\widetilde{w}_N) - \Phi(\bar{x},\bar{w}) \leq C_1/(N+1)^2$. Furthermore, since $(\bar{x},\bar{w},\bar{y})$ is a saddle point of $\mathcal{L}(\cdot)$ and $\|\bar{y}\|\leq c/2$, we have

$$\Phi(\bar{x},\bar{w}) - \Phi(\widetilde{x}_N,\widetilde{w}_N) \leq \langle \bar{y}, K\widetilde{x}_N - \widetilde{w}_N\rangle \leq \frac{c}{2}\|K\widetilde{x}_N - \widetilde{w}_N\|, \qquad (3.37)$$

which together with (3.35) implies

$$c\|K\widetilde{x}_N - \widetilde{w}_N\| \leq \Phi(\bar{x},\bar{w}) - \Phi(\widetilde{x}_N,\widetilde{w}_N) + \frac{C_1}{(N+1)^2} \leq \frac{c}{2}\|K\widetilde{x}_N - \widetilde{w}_N\| + \frac{C_1}{(N+1)^2}.$$

As a result, we obtain $\|K\widetilde{x}_N - \widetilde{w}_N\| \leq \frac{2C_1}{c(N+1)^2}$. It then follows from (3.37) that $\Phi(\bar{x},\bar{w}) - \Phi(\widetilde{x}_N,\widetilde{w}_N) \leq C_1/(N+1)^2$, and thus $|\Phi(\widetilde{x}_N,\widetilde{w}_N) - \Phi(\bar{x},\bar{w})| \leq C_1/(N+1)^2$. Finally, setting $y=\bar{y}$ in (3.33) and using the first inequality in (3.37) to obtain $t_N l_{N+1}\|\bar{x}-z_{N+2}\|^2 \leq C'$, where C' is given by (3.36). Furthermore, it is elementary to verify that

$$2t_N l_{N+1} = \frac{t_N}{\tau_{N+1}}\frac{\psi+\gamma\tau_{N+1}}{\psi-1} \geq \frac{t_N}{\tau_{N+1}}\frac{\psi}{\psi-1} \geq C'' t_N t_{N+1} \geq \frac{C''(N+2)^2}{4},$$

where $C'' := \frac{\rho L^2 t_{n_0+1} - (1-\mu)\gamma}{(1-\mu)(\psi-1)t_{n_0+1}} > 0$. Then, we have $\|z_{N+1}-\bar{x}\| \leq C_2/(N+2)$ with $C_2 := \sqrt{8C'/C''}$, and therefore, by noting (3.1a), $\|x_{N+1}-\bar{x}\| \leq C_2/(N+2)$. This completes the proof. □

We next show that the function value residuals for the unconstrained optimization problems (1.2) and (1.3) also converge at the order $\mathcal{O}(1/n^2)$ under additional assumptions that f and g^* are Lipschitz continuous. Similar results have been achieved in [32].

Theorem 3.8. *Let $\{(x_n, w_n, y_n, z_n)\}$ be generated by Algorithm 3.1, $\{(\widetilde{x}_n, \widetilde{w}_n, \widetilde{y}_n, \widehat{y}_n)\}$ be given by (3.3) and (3.4). Assume that f is M_f-Lipschitz continuous, i.e., $|f(u) - f(v)| \leq M_f\|u - v\|$ for all $u, v \in \mathrm{dom}(f)$. Then, there exists $C_3 > 0$ such that $\mathcal{P}(\widetilde{x}_n) - \inf \mathcal{P} \leq C_3/(n+1)^2$ for all $n \geq n_0$. If in addition g^* is M_{g^*}-Lipschitz continuous and $\mathrm{dom}(g^*)$ is open, then there exists $C_4 > 0$ such that $\sup \mathcal{D} - \mathcal{D}(\widehat{y}_n) \leq C_4/(n+1)^2$, and hence $\mathcal{P}(\widetilde{x}_n) - \mathcal{D}(\widehat{y}_n) \leq (C_3 + C_4)/(n+1)^2$, for all $n \geq n_0$.*

Proof. For any $(x, w, y) \in \mathcal{X}$, we define $A_n(x, w, y) := \mathcal{L}_{(1-\mu)\rho t_{n-1}^2}(\widetilde{x}_n, \widetilde{w}_n, y) - \mathcal{L}(x, w, \widehat{y}_n)$ and

$$a_n(x, w, y) := t_{n-1}^2 A_n(x, w, y) + t_n l_{n+1}\|x - z_{n+2}\|^2 + \frac{1}{2\mu\rho}\|y - \widetilde{y}_n\|^2. \tag{3.38}$$

Then, the key result (3.31) can be rewritten as $a_n(x, w, y) \leq a_{n-1}(x, w, y)$, which holds for any $n \geq n_0$. As a result, we have

$$t_{n-1}^2(\mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, y) - \mathcal{L}(x, w, \widehat{y}_n)) \leq t_{n-1}^2 A_n(x, w, y) \leq a_n(x, w, y) \leq \ldots \leq a_{n_0}(x, w, y). \tag{3.39}$$

Here the first inequality holds because $\mathcal{L}_\sigma(\cdot) \geq \mathcal{L}(\cdot)$, see (2.2) and (2.3). On the other hand, the M_f-Lipschitz continuity of f implies that

$$f(K\widetilde{x}_n) \leq f(\widetilde{w}_n) + M_f\|K\widetilde{x}_n - \widetilde{w}_n\| = f(\widetilde{w}_n) + \langle \breve{y}_n, K\widetilde{x}_n - \widetilde{w}_n\rangle,$$

where $\breve{y}_n := M_f(K\widetilde{x}_n - \widetilde{w}_n)/\|K\widetilde{x}_n - \widetilde{w}_n\|$ if $K\widetilde{x}_n \neq \widetilde{w}_n$ and $\breve{y}_n := 0$ otherwise. Then, we have

$$\begin{aligned}\mathcal{P}(\widetilde{x}_n) - \inf \mathcal{P} \leq& g(\widetilde{x}_n) + f(\widetilde{w}_n) + \langle \breve{y}_n, K\widetilde{x}_n - \widetilde{w}_n\rangle - \Phi(\bar{x}, \bar{w})\\ =& \mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, \breve{y}_n) - \mathcal{L}(\bar{x}, \bar{w}, \widehat{y}_n) \leq a_{n_0}(\bar{x}, \bar{w}, \breve{y}_n)/t_{n-1}^2,\end{aligned}$$

where the second inequality follows from (3.39). Since $\|\breve{y}_n\| \leq M_f$, it is clear that

$$\begin{aligned}A_{n_0}(\bar{x}, \bar{w}, \breve{y}_n) &= \mathcal{L}_{(1-\mu)\rho t_{n_0-1}^2}(\widetilde{x}_{n_0}, \widetilde{w}_{n_0}, \breve{y}_n) - \mathcal{L}(\bar{x}, \bar{w}, \widehat{y}_{n_0})\\ &\leq C' := |\Phi(\widetilde{x}_{n_0}, \widetilde{w}_{n_0}) - \Phi(\bar{x}, \bar{w})| + M_f\|K\widetilde{x}_{n_0} - \widetilde{w}_{n_0}\| + \frac{(1-\mu)\rho t_{n_0-1}^2}{2}\|K\widetilde{x}_{n_0} - \widetilde{w}_{n_0}\|^2.\end{aligned}$$

It then follows from the definition of $a_n(\bar{x}, \bar{w}, \breve{y}_n)$ in (3.38) that

$$a_{n_0}(\bar{x}, \bar{w}, \breve{y}_n) = t_{n_0-1}^2 A_{n_0}(\bar{x}, \bar{w}, \breve{y}_n) + t_{n_0} l_{n_0+1}\|\bar{x} - z_{n_0+2}\|^2 + \frac{1}{2\mu\rho}\|\breve{y}_n - \widetilde{y}_{n_0}\|^2 \leq C'',$$

where $C'' := t_{n_0-1}^2 C' + t_{n_0} l_{n_0+1}\|\bar{x} - z_{n_0+2}\|^2 + \frac{1}{2\mu\rho}(M_f + \|\widetilde{y}_{n_0}\|)^2 > 0$. Then, by setting $C_3 = 4C''$ and noting $t_{n-1} \geq (n+1)/2$, we obtain $\mathcal{P}(\widetilde{x}_n) - \inf \mathcal{P} \leq C_3/(n+1)^2$ for all $n \geq n_0$.

Now, assume in addition that g^* is Lipschitz continuous with constant M_{g^*} and $\mathrm{dom}(g^*)$ is open. Let $\breve{x}_n \in \partial g^*(-K^\top \widehat{y}_n)$ and $\breve{w}_n \in \partial f^*(\widehat{y}_n)$. Then, Lemma 2.2 and (1.3) imply that

$$\mathcal{D}(\widehat{y}_n) = g(\breve{x}_n) + \langle K\breve{x}_n, \widehat{y}_n\rangle - f^*(\widehat{y}_n) = \mathcal{L}(\breve{x}_n, \breve{w}_n, \widehat{y}_n).$$

It follows from (2.4) that $\sup \mathcal{D} = \inf \mathcal{P} = \Phi(\bar{x}, \bar{w}) = \mathcal{L}(\bar{x}, \bar{w}, \bar{y}) \leq \mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, \bar{y})$, where $(\bar{x}, \bar{w}, \bar{y})$ is any saddle point of $\mathcal{L}(\cdot)$. Thus, it follows from (3.39) that

$$\sup \mathcal{D} - \mathcal{D}(\widehat{y}_n) \leq \mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, \bar{y}) - \mathcal{L}(\breve{x}_n, \breve{w}_n, \widehat{y}_n) \leq a_{n_0}(\breve{x}_n, \breve{w}_n, \bar{y})/t_{n-1}^2.$$

Since $t_{n-1} \geq (n+1)/2$, it remains to show that $a_{n_0}(\breve{x}_n, \breve{w}_n, \bar{y})$ is uniformly bounded above.

Since $\breve{x}_n \in \partial g^*(-K^\top \widehat{y}_n)$ and g^* is M_{g^*}-Lipschitz continuous, we have from [2, Theorem 3.61] that $\|\breve{x}_n\| \leq M_{g^*}$ for all n, which together with the fact that g is lower semicontinuous implies that $g(\breve{x}_n) \geq \widetilde{C}$ for all n and some constant $\widetilde{C}$, see [1, Corollary 9.20]. Since $y_n \in \mathrm{dom}(f^*)$, it follows from (3.3) that $\widehat{y}_n \in \mathrm{dom}(f^*)$ for all n. Then, Lemma 2.2, $\|K\breve{x}_n\| \leq LM_{g^*}$ and $f^*(\widehat{y}_{n_0}) < \infty$ imply that

$$\mathcal{L}(\breve{x}_n, \breve{w}_n, \widehat{y}_{n_0}) \geq g(\breve{x}_n) + \langle K\breve{x}_n, \widehat{y}_{n_0}\rangle - f^*(\widehat{y}_{n_0}) \geq \widetilde{C} - LM_{g^*}\|\widehat{y}_{n_0}\| - f^*(\widehat{y}_{n_0}) > -\infty.$$

Further considering the definition of $A_n(x, w, y)$, we have

$$A_{n_0}(\breve{x}_n, \breve{w}_n, \bar{y}) \leq \left|\mathcal{L}_{(1-\mu)\rho t_{n_0-1}^2}(\widetilde{x}_{n_0}, \widetilde{w}_{n_0}, \bar{y})\right| + \left|\widetilde{C} - LM_{g^*}\|\widehat{y}_{n_0}\| - f^*(\widehat{y}_{n_0})\right| =: \widehat{C} < \infty.$$

Substituting the above inequality into the definition of $a_{n_0}(\breve{x}_n, \breve{w}_n, \bar{y})$ yields

$$a_{n_0}(\breve{x}_n, \breve{w}_n, \bar{y}) \leq t_{n_0-1}^2 A_{n_0}(\breve{x}_n, \breve{w}_n, \bar{y}) + t_{n_0} l_{n_0+1}\|\breve{x}_n - z_{n_0+2}\|^2 + \frac{1}{2\mu\rho}\|\bar{y} - \widetilde{y}_{n_0}\|^2 \tag{3.40}$$

$$\leq t_{n_0-1}^2 \widehat{C} + t_{n_0} l_{n_0+1}(M_{g^*} + \|z_{n_0+2}\|)^2 + \frac{1}{2\mu\rho}\|\bar{y} - \widetilde{y}_{n_0}\|^2. \tag{3.41}$$

Therefore, $a_{n_0}(\breve{x}_n, \breve{w}_n, \bar{y})$ is indeed bounded above uniformly with respect to n. As a result, there exists a constant $C_4 > 0$ such that $\sup \mathcal{D} - \mathcal{D}(\widehat{y}_n) \leq C_4/(n+1)^2$, and, since $\inf \mathcal{P} = \sup \mathcal{D}$, $\mathcal{P}(\widetilde{x}_n) - \mathcal{D}(\widehat{y}_n) \leq (C_3 + C_4)/(n+1)^2$ for all $n \geq n_0$. This completes the proof. □

The analysis of Theorem 3.7 is largely motivated by the unified analytic framework given in [25] for analyzing Lagrangian type splitting methods, while that of Theorem 3.8 is inspired by the recent work [32]. It might be interesting to explore the possibility of a unified analysis. Finally, we note that linear convergence results can generally be established for primal-dual type splitting methods when both g and f^* are strongly convex, see, e.g., [5] and also our recent analysis of GRPDA with line search [8]. If only one of the objective function is strongly convex, linear convergence results are also attainable in the presence of some error bound or smoothness conditions, see, e.g., [29] for details.

4 Analysis of GRPDA — Convex case

In this section, we study the convergence of GRPDA for solving the general convex case of the bilinear saddle point problem (1.1) and its equivalent problems. In this case, the algorithm appears to be much simpler than Algorithm 3.1 since the stepsize parameters τ_n and σ_n keep constants. We choose to present algorithm and its detailed analysis for the strongly convex case first, followed by the much simpler general convex case, because this way the proofs of results for the later case can be omitted since they are completely analogous to those of the former.

In the following, we first restate the complete algorithm for clearness, and then present the intermediate results and the main theorems, while their proofs will be omitted for simplicity.

Algorithm 4.1 (GRPDA)**.**

Step 0. Let $\tau, \sigma > 0$, $\mu \in (0,1)$ and $\psi \in (1, \phi]$ be such that $\tau\sigma L^2 = (1-\mu)\psi$. Choose $x_0 \in \mathbb{R}^q$ and $y_0 \in \mathbb{R}^p$. Set $z_0 = x_0$ and $n = 0$.

Step 1. Compute

$$\begin{aligned} z_{n+1} &= \big(1 - 1/\psi\big)x_n + z_n/\psi, \\ x_{n+1} &= \mathrm{Prox}_{\tau g}(z_{n+1} - \tau K^\top y_n), \\ w_{n+1} &= \mathrm{Prox}_{f/\sigma}(y_n/\sigma + K x_{n+1}), \\ y_{n+1} &= y_n + \sigma(K x_{n+1} - w_{n+1}). \end{aligned}$$

Step 2. Set $n \leftarrow n+1$ and return to Step 1.

In the rest of this section, we let $t_{-1} = 0$ and $t_{n+1} = t_n + 1$ for $n \geq 0$, i.e., $t_n = n+1$ for $n \geq -1$. Let $\widetilde{w}_0 \in \mathbb{R}^p$ be arbitrarily chosen and set $(\widetilde{x}_0, \widetilde{y}_0, \widehat{y}_0) = (x_0, y_0, y_0)$. Exactly the same as for the strongly convex case, we define the auxiliary sequence $\{(\widetilde{x}_n, \widetilde{w}_n, \widehat{y}_n) : n \geq 0\}$ recursively as in (3.3) with $t_n = n+1$. Similar to (3.4), we define

$$\widetilde{y}_{n+1} = \widetilde{y}_n + \mu\sigma(K x_{n+1} - w_{n+1}) \quad \text{for} \quad n \geq 0. \tag{4.1}$$

Again, the auxiliary sequences $\{(\widetilde{x}_n, \widetilde{w}_n, \widetilde{y}_n, \widehat{y}_n) : n \geq 0\}$ are used only in the convergence rate analysis and need not to be computed in practice. In the following, without repeatedly mentioning, we let $\{(z_n, x_n, y_n, w_n) : n \geq 0\}$ be the sequence generated by Algorithm 4.1 and $\{(\widetilde{x}_n, \widetilde{w}_n, \widehat{y}_n, \widetilde{y}_n) : n \geq 0\}$ be defined in (3.3) and (4.1). Similar to Lemmas 3.2, 3.4-3.6, we have the following results.

Lemma 4.2. *For all $n \geq 1$, there holds $y_n = \widetilde{y}_n + (1-\mu)\rho t_{n-1}(K\widetilde{x}_n - \widetilde{w}_n)$.*

Lemma 4.3. *For any $(x, w, y) \in \mathcal{X}$ and $n \geq 1$, we have*

$$\begin{aligned} \mathcal{L}_\sigma(x_n, w_n, y_{n-1}) - \mathcal{L}(x, w, y) \leq{}& \frac{1}{\tau}\langle x_{n+1} - z_{n+1},\ x - x_{n+1}\rangle + \frac{\psi}{\tau}\langle x_n - z_{n+1},\ x_{n+1} - x_n\rangle \\ &+ \sigma\langle K x_n - w_n, w_n - K x_{n+1}\rangle + \frac{\sigma}{2}\|K x_n - w_n\|^2 + \langle y_n - y, Kx - w\rangle. \end{aligned}$$

Lemma 4.4. *Let $P = \frac{\psi}{\psi - 1} I \succeq 0$. For any $(x, w, y) \in \mathcal{X}$ and $n \geq 1$, we have*

$$\begin{aligned} \mathcal{L}_\sigma(x_n, w_n, y) - \mathcal{L}(x, w, y) \leq{}& \frac{1}{\tau}\Delta_P(x, z_{n+1}, z_{n+2}) + \frac{1}{\mu\sigma}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n) + \langle y_n - y, Kx - w\rangle \\ &- (1-\mu)\sigma t_{n-2}\langle K\widetilde{x}_{n-1} - \widetilde{w}_{n-1}, K x_n - w_n\rangle. \end{aligned}$$

Lemma 4.5. *For any $(x, w, y) \in \mathcal{X}$ and $n \geq 1$, we have*

$$t_{n-1} S^n_{(1-\mu)\sigma t_{n-1}} - t_{n-2} S^{n-1}_{(1-\mu)\sigma t_{n-2}} \leq \frac{1}{\tau}\Delta_P(x, z_{n+1}, z_{n+2}) + \frac{1}{\mu\sigma}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n), \tag{4.2}$$

where $P = \frac{\psi}{\psi - 1} I \succeq 0$ and S^n_β is identically defined as in (3.21)

The proofs of Lemmas 4.2-4.5 are completely analogous to those of Lemmas 3.2 and 3.4-3.6, respectively. The only differences are to replace the relation $t_{n-1}^2 - t_{n-1} = t_{n-2}^2$ by $t_{n-1} - 1 = t_{n-2}$ and to multiply (3.22) and (3.24) by t_{n-1} and $\sigma t_{n-1}^2/2$, respectively. To reduce redundancy, we omit the details. Based on the lemmas above, we can now establish the convergence rate of GRPDA.

Theorem 4.6. *Let $(\bar{x}, \bar{w}, \bar{y})$ be a saddle point of $\mathcal{L}(\cdot)$ satisfying (2.4) and $c > 0$ be a constant such that $c \geq 2\|\bar{y}\|$. Then, there exists a constant $C_5 > 0$ such that for any $N \geq 1$ we have*

$$|\Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) - \Phi(\bar{x}, \bar{w})| \leq \frac{C_5}{N} \quad \textit{and} \quad \|K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\| \leq \frac{2C_5}{cN}.$$

Theorem 4.6 can be proved by taking the sum of (4.2) over $n = 1, \ldots, N$ and following the same line of proofs as Theorem 3.7. Again, we omit the details for succinctness.

Similarly to Theorem 3.8, we can quantify the convergence rate of the auxiliary sequence $\{(\widetilde{x}_n, \widehat{y}_n) : n \geq 1\}$ via the function value residual of the unconstrained problems (1.2) and (1.3) under the Lipschitz continuity conditions on f and g^*, whose proofs are completely analogous to Theorem 3.8 and, again, are omitted.

Theorem 4.7. *Assume that f is M_f-Lipschitz continuous. Then, there exists $C_6 > 0$ such that $\mathcal{P}(\widetilde{x}_n) - \inf \mathcal{P} \leq C_6/n$ for all $n \geq 1$. If in addition g^* is M_{g^*}-Lipschitz continuous and $\mathrm{dom}(g^*)$ is open, then there exists $C_7 > 0$ such that $\sup \mathcal{D} - \mathcal{D}(\widehat{y}_n) \leq C_7/n$, and hence $\mathcal{P}(\widetilde{x}_n) - \mathcal{D}(\widehat{y}_n) \leq (C_6 + C_7)/n$, for all $n \geq 1$.*

5 Numerical Experiments

In this section, we demonstrate the performance of the proposed Algorithm 4.1 (GRPDA) and its accelerated variant Algorithm 3.1 (A-GRPDA) via preliminary numerical results on the least absolute deviation (LAD) and the LASSO problems, both of which are popular in recovering sparse signals. Comparison results with Chambolle and Pock's and Tran-Dinh and Zhu's PDAs [5, 6, 27], denoted respectively by CP and TDZ, and their accelerated counterparts, denoted respectively by A-CP and A-TDZ, will be given. All the algorithms were implemented in Matlab (R2019b), running on a Laptop with an Intel(R) Core(TM) i5-4590 CPU@3.30 GHz and 8GB of RAM within Microsoft Windows. All the experimental results presented in this section are reproducible by specifying the random number generator `seed` in our code.

5.1 LAD problem

Let $K \in \mathbb{R}^{p\times q}$ be a sensing matrix, $x^\flat \in \mathbb{R}^q$ be an s-sparse signal and $b = \mathbf{N}(Kx^\flat) \in \mathbb{R}^p$ be an observation of $x^\flat$, where $\mathbf{N}(\cdot)$ denotes an impulsive noise corruption procedure. To recover $x^\flat$ from b, we consider the following regularized LAD problem

$$\min_x \mathcal{P}(x) := \|Kx - b\|_1 + g(x),$$

where g serves as a regularizer. We consider the following two cases.

General convex case. For this case, we set $g(x) = \eta\|x\|_1$. The entries of K were generated from $\mathcal{N}(0, 1)$, the normal distribution with mean 0 and standard deviation 1. $x^\flat$ is an s-sparse vector with components randomly generated via the Matlab built-in function `randn`, and the impulsive noise corruption procedure was simulated by replacing 5% randomly chosen entries of $Kx^\flat$ with $\|Kx^\flat\|_\infty$ or $-\|Kx^\flat\|_\infty$, both with probability 0.5. In this experiment, we set $(p, q, s) = (2000, 640, 200)$ and $\eta = 0.05$.

Strongly convex case. For this case, we set $g(x) := \eta\|x\|_1 + \frac{\gamma_g}{2}\|x\|^2$, which is also known as the elastic net regularization. In this experiment, we first generate $\{\varrho_j \in \mathbb{R}^p : j =$

Codes available at `https://github.com/quoctd/PrimalDualCvxOpt`.

$1, \ldots, q\}$ independently from the Gaussian distribution with mean 0 and covariance I, and then define the j-th column of K, denoted by K_j, recursively as follows

$$K_1 = \varrho_1 \sqrt{(1-\upsilon)^2/(1-\upsilon^2)} \quad \text{and} \quad K_j = \upsilon K_{j-1} + (1-\upsilon)\varrho_j,\ j = 2, \ldots, q,$$

where $\upsilon = 0.5$. This way of generating K has been frequently tested in the literature, see, e.g., [27]. In fact, K becomes more ill-conditioned as υ increases from 0 to 1. Other data was generated in the same way as for the general convex case. For this experiment, we set $(p, q, s) = (4000, 1280, 400)$, $\eta = 0.05$ and $\gamma_g = 0.1$.

In both cases, we set $f(\cdot) = \|\cdot - b\|_1$, and thus the proximal point mappings of both f and g can be represented by the soft-thresholding operator. We omit the details since it is well known. In the general convex case, the algorithms to be compared are GRPDA, CP and TDZ, while in the strongly convex case their accelerated counterparts, i.e., A-GRPDA, A-CP and A-TDZ, will be compared. All the algorithms were initialized at the origin. For A-GRPDA, we set $\tau_0 = \frac{(1-\mu)\psi}{\sigma_0\|K\|^2}$ with $\sigma_0 = \frac{(1-\mu)(\psi-1)\gamma_g}{2\|K\|^2}$, $\psi = 1.6$ and $\mu = 0.01$. For A-CP, we set $\tau_0 = 1/\|K\|$, and for A-TDZ we chose $c = 4$, $\gamma = 0.75$, $\Gamma = 2 - 1/\gamma$ and $\rho_0^2 := \frac{c(c-1)\Gamma\gamma_g}{(2c-1)\|K\|^2}$ as in [27].

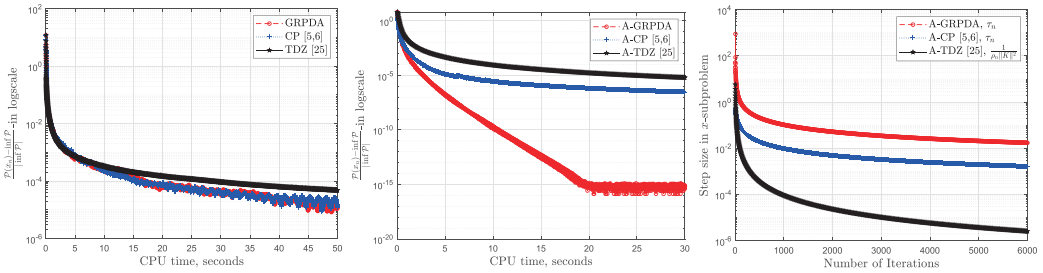

Figure 5.1: Comparison results of GRPDA, CP, TZD and their accelerated counterparts on the regularized LAD problem. Left: general convex case; Middle: strongly convex case; Right: stepsize parameters generated by the three accelerated algorithms throughout the execution.

The relative function value residuals, which decreased as the CPU time proceeded, were plotted in Figure 5.1 for both the general convex case (plot on the left-hand-side) and the strongly convex case (plot in the middle). In both cases, the true optimal function value $\inf \mathcal{P}$ was computed by CVX [15]. In the general convex case, all the compared algorithms achieve the $\mathcal{O}(1/n)$ sublinear rate. It can be seen from Figure 5.1 (left) that GRPDA and CP perform closely and TDZ falls behind slightly. It is also worth noting that all the three nonaccelerated algorithms only achieved relatively lower accuracy compared to $\inf \mathcal{P}$ computed by CVX. On the other hand, all the three accelerated algorithms use different ways to update stepsizes and achieve the faster $\mathcal{O}(1/n^2)$ sublinear rate. It can be seen from Figure 5.1 (middle) that A-GRPDA performs the best, followed by A-CP and A-TDZ. In fact, A-GRPDA converges much faster than the other two algorithms, and A-CP is only slightly faster than A-TDZ. A possible reason for the slower convergence of A-TDZ might be that it needs to compute x-subproblems twice at each iteration. Furthermore, our observation shows that the stepsizes generated by A-TDZ shrunk to 0 the fastest, followed by A-CP, and A-GRPDA adopted the largest stepsize parameters and shrunk to 0 at the slowest speed among the three algorithms, see Figure 5.1 (plot on the right). The larger stepsize of A-GRPDA is due to the way of computing τ_n in (3.2c). Since larger stepsizes

usually results in better performance, this might be a possible explanation of the faster convergence of A-GRPDA.

5.2 LASSO problem

In the case of Gaussian noise, it is desirable to recover the sparse signal $x^\flat$ via solving the LASSO problem

$$\min_x \mathcal{P}(x) := \eta\|x\|_1 + \frac{1}{2}\|Kx - b\|^2, \tag{5.1}$$

where K and b are the same as in LAD except that $b = Kx^\flat + \nu$ and ν is a Gaussian noise. It is easy to verify that the LASSO problem (5.1) fits into (1.1) with $g(x) = \eta\|x\|_1$ and $f^*(y) = \frac{1}{2}\|y\|^2 + \langle b, y\rangle$. 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 applications of the accelerated algorithms A-GRPDA, A-CP and A-TDZ.

In this experiment, the values of the nonzero components of $x^\flat$ were drawn from the uniform distribution in $[-10, 10]$, while their positions were determined uniformly at random. The additive noise $\nu \in \mathbb{R}^p$ was generated from $\mathcal{N}(0, 0.1I)$. Furthermore, two types of K were tested, i.e., partial discrete cosine transform (DCT) and partial discrete Fourier transform (DFT), where the rows of the DCT and the DFT matrices were selected uniformly at random. Note that partial fast transforms were frequently used in compressive sensing, see e.g., [30]. We tested partial DCT with $(p, q) = (4000, 1280)$ and partial DFT with $(p, q) = (8000, 2560)$. In both cases, we set $s = p/10$ and $\eta = 0.1$. For these two sensing matrices, we have $KK^\top = I$ and thus $L = \|K\| = 1$. In this experiment, we computed $\inf \mathcal{P}$ by running A-GRPDA to a sufficiently high accuracy, instead of calling CVX as it cannot be used since K is not explicitly saved in memory. The same as before, all the compared algorithms were initialized at the origin and adopted the same set of parameters as specified in the strongly convex case of the LAD problem.

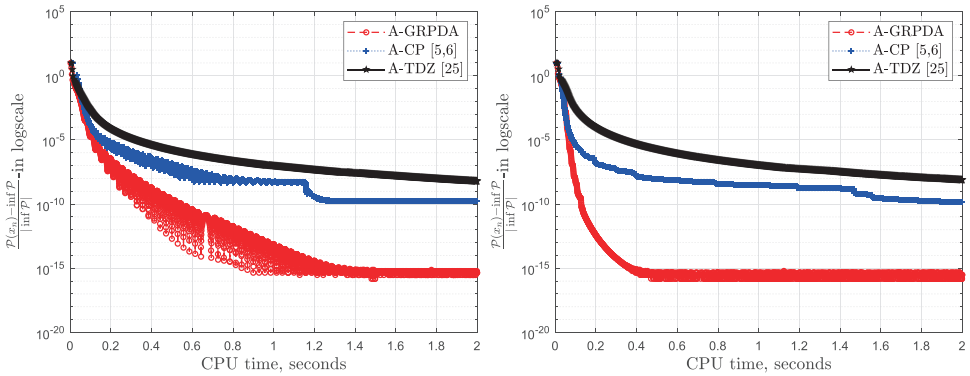

Figure 5.2: Comparison results of A-GRPDA, A-CP and A-TDZ on the LASSO problem. Left: partial DCT with $(p, q, s) = (4000, 1280, 400)$. Right: partial DFT with $(p, q, s) = (8000, 2560, 800)$.

Similar comparison results of A-GRPDA, A-CP and A-TDZ as in the case for the LAD problem are given in Figure 5.2, from which roughly the same conclusion can be drawn, i.e., A-GRPDA performs the best and much better than the other two algorithms, and A-CP

follows next, which is slightly faster than A-TDZ. We attribute the faster convergence of A-GRPDA to its larger stepsizes, as shown in Figure 5.1 (right). Again, A-TDZ appears to be the slowest due to the fact that it adopted the smallest stepsizes and solved x-subproblems twice per iteration.

6 Conclusions

Golden ratio primal-dual algorithm (GRPDA) is an efficient new variant of the classical Arrow-Hurwicz method for solving the bilinear saddle point problem. At present, existing convergence rate results of GRPDA are defective because they are based on the so-called "primal-dual gap function", which could vanish at nonstationary points. In this paper, based on equivalent reformulations as optimization problems, we have established some new convergence rate results for GRPDA and an accelerated variant of it. These new results are based on function value residual and constraint violation, which are conventional optimality measures for constrained optimization problems. Specifically, in the general convex case, some auxiliary sequences generated by GRPDA enjoy $\mathcal{O}(1/N)$ sublinear convergence rate, while in the strongly convex case we have constructed an accelerated GRPDA, which achieves the faster $\mathcal{O}(1/N^2)$ sublinear rate. Moreover, we have shown that the same sublinear rates measured by function value residual for the unconstrained optimization problems (1.2) and (1.3) can be achieved in both the convex and the strongly convex cases if f and g^* are Lipschitz continuous, leading to a convergence rate guarantee on the true primal and dual function value gap. These new convergence rate results have definitely enriched the convergence theory of GRPDAs. Our preliminary numerical results on the least absolute deviation and the LASSO problems also show the superior performance of the accelerated GRPDA. Finally, it is believed that the analysis presented in this paper can be extended to other augmented Lagrangian related splitting algorithms, which are interesting to explore further.

References

[1] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, second edition, Springer, Cham, 2011.

[2] A. Beck, *First-Order Methods in Optimization*, MOS-SIAM Series on Optimization. SIAM-Society for Industrial and Applied Mathematics, 2017.

[3] D.P. Bertsekas and E.M. Gafni, Projection methods for variational inequalities with application to the traffic assignment problem, *Math. Program. Stud.* 17 (1982) 139–159.

[4] T. Bouwmans, N.S. Aybat, and E.H. Zahzah, *Handbook of "Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing"*, volume 45. 2016.

[5] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* 40 (2011) 120–145.

[6] A. Chambolle and T. Pock, On the ergodic convergence rates of a first-order primal-dual algorithm, *Math. Program.*, 159 (2016) 253–287.

[7] X. Chang and J. Yang, A golden ratio primal-dual algorithm for structured convex optimization, *J. Sci. Comput.* 87 (2021): Paper No. 47.

[8] X. Chang, J. Yang and H. Zhang, Golden ratio primal-dual algorithm with linesearch, *Manscript*, 2021.

[9] C. Chen, R.H. Chan, S. Ma and J. Yang, Inertial proximal ADMM for linearly constrained separable convex optimization, *SIAM J. Imaging Sci.* 8 (2015) 2239–2267.

[10] J. Eckstein, Some saddle-function splitting methods for convex programming, *Optim. Method Softw.* 4 (1994) 75–83.

[11] E. Esser, X. Zhang and T.F. Chan, A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science, *SIAM J. Imaging Sci.* 3(2010) 1015–1046.

[12] M. Fazel, T.K. Pong, D. Sun and P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, *SIAM J. Matrix Anal. Appl.* 34 (2013) 946–977.

[13] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, *Comput. Math. Appl.* 2 (1976) 17–40.

[14] 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, *R.A.I.R.O., R2*, 9 (1975) 41–76.

[15] M. Grant, S. Boyd and Y. Ye. CVX: Matlab software for disciplined convex programming, 2008.

[16] S. Hayden and O. Stanley, A low patch-rank interpretation of texture, *SIAM J. Imaging Sci.* 6 (2013) 226–262.

[17] B. He, L.-Z. Liao, D. Han and H. Yang, A new inexact alternating directions method for monotone variational inequalities, *Math. Program.* 92 (2002) 103–118.

[18] B. He, Y. You, and X. Yuan, On the convergence of primal-dual hybrid gradient algorithm, *SIAM J. Imaging Sci.* 7 (2014) 2526–2537.

[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) 119–149.

[20] E. Jonathan and P. Dimitri, On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* 55 (1992) 293–318.

[21] Y. Malitsky, Golden ratio algorithms for variational inequalities, *Math. Program.* 184 (2020) 383–410.

[22] Y. Malitsky and T. Pock, A first-order primal-dual algorithm with linesearch, *SIAM J. Optim.* 28 (2018) 411–432.

[23] A. Nedic and A. Ozdaglar, Subgradient methods for saddle-point problems, *J. Optim. Theory Appl.* 142 (2009) 205–228.

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

[25] S. Sabach and M. Teboulle, Faster lagrangian-based methods in convex optimization, *arXiv 2010.14314*, 2020.

[26] 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) 269–297.

[27] Q. Tran-Dinh and Y. Zhu, Non-stationary first-order primal-dual algorithms with faster convergence rates, *SIAM J. Optim.* 30 (2020) 2866–2896.

[28] 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.

[29] K. Wang and H. He, A double extrapolation primal-dual algorithm for saddle point problems, *J. Sci. Comput.* 85 (2020): Paper No. 30.

[30] J. Yang and Y. Zhang, Alternating direction algorithms for ℓ_1-problems in compressive sensing, *SIAM J. Sci. Comput.* 33 (2011), 250–278.

[31] D. Zhou, X. Chang and J. Yang, A new primal-dual algorithm for structured convex optimization involving a lipschitzian term, *Pac. J. Optim., to appear.*

[32] Y. Zhu, D. Liu and Q. Tran-Dinh, A new primal-dual algorithm for a class of nonlinear compositional convex optimization problems. *arXiv:2006.09263v2*, 2021.

Manuscript received 28 August 2021
revised 15 October 2021
accepted for publication 31 October 2021

Xiaokai Chang
School of Science, Lanzhou University of Technology
Lanzhou, Gansu, P.R. China
E-mail address: xkchang@lut.cn

Junfeng Yang
Department of Mathematics, Nanjing University
Nanjing, P.R. China
E-mail address: jfyang@nju.edu.cn