Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 2, pp. 497–517, 2022

A NEW PRIMAL-DUAL ALGORITHM FOR STRUCTURED CONVEX OPTIMIZATION INVOLVING A LIPSCHITZIAN TERM*

Danqing Zhou, Xiaokai Chang and Junfeng Yang†

Abstract: We propose and analyze a golden ratio primal-dual algorithm for solving structured optimization problems involving the sum of three convex terms — a smooth function with Lipschitzian gradient and two nonsmooth proximal-friendly functions, one of which is composed with a linear mapping. The proposed algorithm is of primal-dual and full-splitting type as it solves the primal and the dual problems simultaneously and does not rely on solving any subproblems or linear system of equations iteratively, the smooth function is handled by gradient evaluation, and the nonsmooth functions are handled by their proximity operators. Several well-known algorithms are closely related, e.g., the classical Arrow-Hurwicz method and the primal-dual algorithm of Chambolle and Pock. In particular, it extends the golden ratio primal-dual algorithm recently proposed by Chang and Yang by including an extra smooth term with Lipschitzian gradient. The convergence rates $\mathcal{O}(1/N)$ and $\mathcal{O}(1/N^2)$ are established for convex and strongly convex cases, respectively, which differentiate themselves from existing results in terms of the adopted optimality measures. Specifically, to measure optimality, most existing results adopt the primal-dual gap function, a major flaw of which is that it could vanish at nonstationary points. In comparison, we adopt function value residual and feasibility violation as optimality measures, which are conventional for constrained optimization. Finally, preliminary numerical results on image reconstruction and elastic net regularization problems are presented to demonstrate the efficiency of the proposed algorithm.

Key words: *structured convex optimization, primal-dual, full-splitting, saddle point, sublinear convergence rate, golden ratio*

Mathematics Subject Classification: *15A18, 15A69, 65F15, 90C33*

---

1 Introduction

Let $\mathbb{R}^p$ and $\mathbb{R}^q$ be finite-dimensional Euclidean spaces with inner products and induced norms denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$, respectively. Our focus in this paper is the following structured convex optimization problem

$$\min_{x\in\mathbb{R}^q} h(x) + g(x) + f(Kx), \tag{1.1}$$

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

†Corresponding author.

where $f : \mathbb{R}^p \to (-\infty, +\infty]$ and $g : \mathbb{R}^q \to (-\infty, +\infty]$ are extended real-valued closed proper convex functions, $h : \mathbb{R}^q \to \mathbb{R}$ is convex and differentiable with L-Lipschitz continuous gradient, i.e.,

$$\|\nabla h(x) - \nabla h(y)\| \leq L\|x - y\|, \ \ \forall\, x, y \in \mathbb{R}^q,$$

and $K : \mathbb{R}^q \to \mathbb{R}^p$ is a linear operator with induced norm $\|K\| = \sup_x\{\|Kx\| : \|x\| \leq 1\}$. Though we have restricted ourselves in $\mathbb{R}^p$ and $\mathbb{R}^q$, our algorithm and convergence results can be generalized to any finite dimensional real Euclidean space.

Problem (1.1) arises from various applications, including signal and image processing, machine learning, statistics, mechanics and economics, just to name a few, see, e.g., [18, 19, 10] for some ill-posed inverse problems and elastic net regularization problems. The Fenchel dual problem of (1.1) is given by

$$\max_{y \in \mathbb{R}^p} -f^*(y) - (g + h)^*(-K^\top y), \tag{1.2}$$

where $f^*(y) = \sup_{u \in \mathbb{R}^p}\{\langle y, u\rangle - f(u)\}$, $y \in \mathbb{R}^p$, is the Legendre conjugate of f, and $K^\top$ denotes the adjoint operator of K. Furthermore, the saddle-point or primal-dual problem corresponding to (1.1) and (1.2) is given by

$$\min_{x \in \mathbb{R}^q} \max_{y \in \mathbb{R}^p} h(x) + g(x) + \langle Kx, y\rangle - f^*(y). \tag{1.3}$$

On the other hand, by introducing an auxiliary variable $w \in \mathbb{R}^p$, one can reformulate problem (1.1) equivalently as

$$\min_{x \in \mathbb{R}^q, w \in \mathbb{R}^p} \big\{\Phi(x, w) := h(x) + g(x) + f(w) \mid Kx - w = 0\big\}. \tag{1.4}$$

Problems (1.1)-(1.4) are closely related. In fact, the equivalence between (1.1) and (1.4) is apparent. Moreover, under regularity conditions (see Assumption 1.1), the dual solution $\bar{y}$ can be induced from a solution $\bar{x}$ to the primal problem via $\bar{y} \in \arg\max_{y \in \mathbb{R}^p}\langle K\bar{x}, y\rangle - f^*(y)$. Contrarily, the primal solution $\bar{x}$ can be induced from a solution $\bar{y}$ to the dual problem via $\bar{x} \in \arg\min_{x \in \mathbb{R}^q} h(x) + g(x) + \langle Kx, \bar{y}\rangle$. Furthermore, any primal solution $\bar{x}$ and dual solution $\bar{y}$ forms a primal-dual solution $(\bar{x}, \bar{y})$ to the primal-dual problem (1.3).

Let $r : \mathbb{R}^n \to (-\infty, +\infty]$ be an extended real-valued closed, proper convex function. The proximity operator of r is defined by

$$\mathrm{Prox}_r(x) := \arg\min_{y \in \mathbb{R}^n} \big\{r(y) + \frac{1}{2}\|y - x\|^2\big\}, \quad x \in \mathbb{R}^n.$$

Our proposed algorithm relies heavily on the proximity operators of f and g, which are uniquely well-defined everywhere since f and g are closed proper and convex. In this paper, we make the following blanket assumption.

Assumption 1.1. Assume that (i) the proximity operators of f and g either have closed formulas or can be evaluated efficiently, and (ii) there exists $\hat{x} \in \mathrm{ri}(\mathrm{dom}(g))$ such that $K\hat{x} \in \mathrm{ri}(\mathrm{dom}(f))$.

Here $\mathrm{ri}(\cdot)$ and $\mathrm{dom}(\cdot)$ represent, respectively, the relative interior of a set and the effective domain of a function. In many applications, the component functions enforce data fitting and/or regularization and usually preserve simple structures so that item (i) of Assumption 1.1 is fulfilled. Examples of such functions are abundant, see, e.g., [1, Chapter 6]. On the other hand, item (ii) of Assumption 1.1 ensures that problems (1.1)-(1.4) have nonempty solution sets.

The literature about numerical algorithms for solving (1.1)-(1.4) under various settings are tremendously vast. In particular, primal-dual algorithms solve these primal, dual and primal-dual problems simultaneously, the smooth function h is processed via its gradient, the nonsmooth functions f and g by their proximity operators, and furthermore, within each iteration there is no need to solve any subproblems or linear system of equations iteratively. This type of algorithms are commonly referred as full-splitting, see, e.g., [5]. We next recall briefly some primal-dual full-splitting algorithms closely related to this work, most of which focus on the special case $h \equiv 0$, i.e., without the smooth term, and can in principle be extended to the case $h \neq 0$. However, primal-dual algorithms for solving (1.1)-(1.4) with $h \neq 0$ and meanwhile have the full-splitting characteristics are much fewer.

1.1 Related algorithms

A popular primal-dual approach for solving (1.4), and thus (1.1)-(1.3), with $h \equiv 0$, is the alternating direction method of multipliers (ADMM, [7, 8]), a variant of the classical method of multipliers. However, when applied to (1.4), ADMM is not full-splitting as it needs to solve a subproblem of the form $\min_{x \in \mathbb{R}^q} \frac{1}{2}\|Kx - b_n\|^2 + g(x)$, which may not be easily solved unless K is the identity operator. The most classical and heuristically simple primal-dual algorithm (PDA) for solving (1.3) is the Arrow-Hurwicz method [20], which starts at $(x_0, y_0) \in \mathbb{R}^q \times \mathbb{R}^p$ and iterates for $n \geq 1$ as

$$\begin{cases} x_n = \mathrm{Prox}_{\tau g}(x_{n-1} - \tau K^\top y_{n-1}), \\ y_n = \mathrm{Prox}_{\sigma f^*}(y_{n-1} + \sigma K x_n). \end{cases} \tag{1.5}$$

Here $\tau, \sigma > 0$ are step size parameters. This method is also widely known as primal-dual hybrid gradient method in image processing community, see [25, 6]. However, the Arrow-Hurwicz method converges under restrictive conditions [6, 2, 13] and does not converge in general, see [9] for a divergent example. In 2011, Chambolle and Pock [2, 15] modified (1.5) by adopting an extrapolation step, resulting the following iterative scheme

$$\begin{cases} x_n = \mathrm{Prox}_{\tau g}(x_{n-1} - \tau K^\top y_{n-1}), \\ z_n = x_n + \delta(x_n - x_{n-1}), \\ y_n = \mathrm{Prox}_{\sigma f^*}(y_{n-1} + \sigma K z_n), \end{cases} \tag{1.6}$$

where $\delta \in (0, 1]$. For $\delta = 1$, (1.6) reduces to the split inexact Uzawa method studied in [6]. Furthermore, it was shown that (1.6) is a linearized ADMM and a weighted proximal point algorithm applied to the optimality conditions of (1.3), see [2, 9, 17]. Convergence and ergodic convergence rate results of (1.6) with $\delta = 1$ are established in [2] under the condition $\tau\sigma\|K\|^2 < 1$, while convergence for the case $\delta \in (0, 1)$ remains unclear. Recently, by using a convex combination firstly introduced by Malitsky [11] to tackle mixed variational inequality problem, Chang and Yang [3] proposed a golden ratio primal-dual algorithm (GRPDA)

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

Convergence and sublinear convergence rate results are established in [3] under the condition $\tau\sigma\|K\|^2 < \psi$ with $\psi \in (1, \frac{\sqrt{5}+1}{2}]$. Compared with the PDA of Chambolle and Pock, here the convergence condition is significantly relaxed. The numerical results reported in [3] demonstrated the benefits gained by this relaxed step size condition.

For $h \neq 0$, the above mentioned methods can conceptually be applied by replacing g with $h + g$. However, it is apparent that the proximity operator of the sum $h + g$ could be much more difficult to evaluate than that of g. Instead, a popular and efficient modification is to replace h by its linear approximation. For example, Condat [5] extended (1.6) to the case $h \neq 0$ and established convergence results under the condition $\frac{1}{\tau} - \sigma\|K\|^2 > \frac{L}{2}$. A similar splitting algorithm was proposed in [21] in the setting of monotone operator inclusion problems. See also [4, 22] for primal-dual full-splitting algorithm studied from the perspective of fixed point iteration. This work can be viewed as an extension of GRPDA (1.7) to solve (1.1)-(1.4), which includes an extra smooth term h with Lipschitzian gradient. Our convergence rate results will be established based on the conventional optimality measures, i.e., function value residual and feasibility violation, rather than the primal-dual gap function as used in, e.g., [2, 12]. As pointed out in [2], a major flaw of the primal-dual gap function is that it could vanish at nonstationary points. Our analysis is motivated by the recent work of Sabach and Teboulle [16], which proposed some principles for analyzing Lagrangian type methods.

1.2 Notation and preliminaries

As used in (1.4), we let $\Phi(x, w) := h(x) + g(x) + f(w)$ for $x \in \mathbb{R}^q$ and $w \in \mathbb{R}^p$. The Lagrangian function $\mathcal{L} : \mathbb{R}^p \times \mathbb{R}^q \to (-\infty, +\infty]$ associated with (1.4) is defined by

$$\mathcal{L}(x, w, y) := \Phi(x, w) + \langle y, Kx - w\rangle, \tag{1.8}$$

where $y \in \mathbb{R}^p$ is the Lagrange multiplier associated with the linear constraint $Kx = w$. The augmented Lagrangian function is defined by

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

where $\sigma > 0$ is a penalty parameter. Denote the set of saddle points of (1.4) by

$$\mathcal{S} = \{(\bar{x}, \bar{w}, \bar{y}) \in \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p : \ 0 \in K^\top \bar{y} + \nabla h(\bar{x}) + \partial g(\bar{x}),\ K\bar{x} = \bar{w},\ \bar{y} \in \partial f(\bar{w})\}, \tag{1.9}$$

which is nonempty under (ii) of Assumption 1.1. The identity matrix of appropriate order is denoted by I. Denote the set of n-by-n symmetric positive semidefinite (resp., positive definite) matrices by $\mathbb{S}^n_+$ (resp., $\mathbb{S}^n_{++}$). For $P \in \mathbb{S}^n_+$, we let $\|u\|_P := \sqrt{\langle u, Pu\rangle}$ be the semi-norm of $u \in \mathbb{R}^n$. For any $u, v, w \in \mathbb{R}^n$, we denote $\Delta_P(u, v, w) := \frac{1}{2}(\|u - v\|_P^2 - \|u - w\|_P^2)$. Specially, when $P = I$, we let $\Delta(u, v, w) := \Delta_I(u, v, w) = \frac{1}{2}(\|u - v\|^2 - \|u - w\|^2)$ for simplicity, where $\|\cdot\|$, as used above, denotes the Euclidean norm. The subdifferential of a closed proper convex function $r : \mathbb{R}^n \to (-\infty, \infty]$ at a given $x \in \mathbb{R}^n$ is denoted by $\partial r(x) := \{v \in \mathbb{R}^n : r(y) \geq r(x) + \langle v, y - x\rangle \ \text{ for all } \ y \in \mathbb{R}^n\}$. For a nonempty set $C \subset \mathbb{R}^n$, we let ι_C be the indicator function of C, i.e., $\iota_C(x) = 0$ if $x \in C$ and ∞ if otherwise. Finally, throughout this paper, we denote the golden ratio by ϕ, i.e., $\phi = \frac{\sqrt{5}+1}{2}$.

We next present some useful identities and fact. Given any matrix $P \in \mathbb{S}^n_+$, for any $u, v, w \in \mathbb{R}^n$, the Pythagoras three-points identity has the form

$$2\langle u - w, P(w - v)\rangle = \|u - v\|_P^2 - \|u - w\|_P^2 - \|v - w\|_P^2. \tag{1.10}$$

Moreover, for any $u, v \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, we have the following identity

$$\|\alpha u + (1 - \alpha)v\|^2 = \alpha\|u\|^2 + (1 - \alpha)\|v\|^2 - \alpha(1 - \alpha)\|u - v\|^2. \tag{1.11}$$

The following lemma, whose proof is elementary and is thus omitted, is very useful and will be used repeatedly in our analysis.

Lemma 1.1. *Let $r : \mathbb{R}^n \to (-\infty, +\infty]$ be an extended real-valued closed proper and γ-strongly convex function with modulus $\gamma \geq 0$. Then, for any $\tau > 0$ and $x \in \mathbb{R}^n$, it holds that $z = \mathrm{Prox}_{\tau r}(x)$ if and only if $r(y) \geq r(z) + \frac{1}{\tau}\langle x - z, y - z\rangle + \frac{\gamma}{2}\|y - z\|^2$ for all $y \in \mathbb{R}^n$.*

1.3 Organization

The rest of this paper is organized as follows. In Section 2, we present our algorithm in the general convex case and compare it with some existing primal-dual full-splitting algorithms. Sublinear $\mathcal{O}(1/N)$ convergence rate results measured by function value residual and feasibility violation are established in Section 3. An accelerated method is proposed when either g or h is strongly convex in Section 4. The accelerated method achieves faster $\mathcal{O}(1/N^2)$ convergence rate measured by the same criteria as for the convex case. Some numerical results are given in Section 5 to demonstrate the efficiency of the proposed algorithms. Finally, we give some concluding remarks in Section 6.

2 Extended GRPDA

The algorithm to be proposed in this section is an extension of (1.7). We therefore refer to it as extended GRPDA or E-GRPDA. Let $\tau, \sigma > 0$, $\psi \in (1, \phi]$, $z_0 = x_0 \in \mathbb{R}^q$ and $y_0 \in \mathbb{R}^p$. Our proposed E-GRPDA iterates for $n \geq 0$ as follows

$$\begin{cases} z_{n+1} = \frac{\psi-1}{\psi}x_n + \frac{1}{\psi}z_n, \\ x_{n+1} = \mathrm{Prox}_{\tau g}(z_{n+1} - \tau K^\top y_n - \tau\nabla h(x_n)), \\ y_{n+1} = \mathrm{Prox}_{\sigma f^*}(y_n + \sigma K x_{n+1}). \end{cases} \tag{2.1}$$

By using the famous Moreau decomposition theorem, i.e., $y = \mathrm{Prox}_{\frac{1}{\sigma}f}(y) + \frac{1}{\sigma}\mathrm{Prox}_{\sigma f^*}(\sigma y)$ for any $y \in \mathbb{R}^q$, the scheme (2.1) can be rewritten as

$$\begin{cases} z_{n+1} = \frac{\psi-1}{\psi}x_n + \frac{1}{\psi}z_n, \\ x_{n+1} = \mathrm{Prox}_{\tau g}(z_{n+1} - \tau K^\top y_n - \tau\nabla h(x_n)), \\ w_{n+1} = \mathrm{Prox}_{\frac{1}{\sigma}f}(\frac{1}{\sigma}y_n + K x_{n+1}), \\ y_{n+1} = y_n + \sigma(K x_{n+1} - w_{n+1}). \end{cases} \tag{2.2}$$

We will show in Section 3 that sublinear convergence rate of E-GRPDA (2.1) or (2.2) is guaranteed under the condition $\tau(\frac{\sigma}{1-\mu}\|K\|^2 + 2L) \leq \psi \in (1, \phi]$ for any $\mu \in (0, 1)$. For clear comparison between E-GRPDA and some existing full-splitting algorithms, we next present the iterative formulas as well as convergence conditions of Condat and Vu [5, 21], primal-dual fixed point algorithm [4] and the primal-dual three operator splitting algorithm [22] for solving (1.1).

Condat-Vu [5, 21]: The Condat-Vu's algorithm is a generalization of the Chambolle-Pock's PDA (1.6) with $\delta = 1$ and iterates as follows

$$\begin{cases} x_{n+1} = \mathrm{Prox}_{\tau g}(x_n - \tau K^\top y_n - \tau\nabla h(x_n)), \\ z_{n+1} = 2x_{n+1} - x_n, \\ y_{n+1} = \mathrm{Prox}_{\sigma f^*}(y_n + \sigma K z_{n+1}), \end{cases}$$

whose global convergence was established under the stepsize condition $\tau(\sigma\|K\|^2 + L/2) < 1$.

PDFP [4]: The primal-dual fixed point algorithm iterates as

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

whose convergence was established under the condition $\tau\sigma\|K\|^2 < 1$ and $\tau L < 2$.

PD3O [22]: The primal-dual three operator splitting algorithm is also a generalization of Chambolle-Pock's PDA (1.6) with $\delta = 1$, and the convergence condition is the same as that of PDFP, i.e., $\tau\sigma\|K\|^2 < 1$ and $\tau L < 2$. The iterative scheme of PD3O is given by

$$\begin{cases} x_{n+1} = \text{Prox}_{\tau g}(x_n - \tau K^\top y_n - \tau \nabla h(x_n)), \\ z_{n+1} = 2x_{n+1} - x_n + \tau \nabla h(x_n) - \tau \nabla h(x_{n+1}), \\ y_{n+1} = \text{Prox}_{\sigma f^*}(y_n + \sigma K z_{n+1}). \end{cases}$$

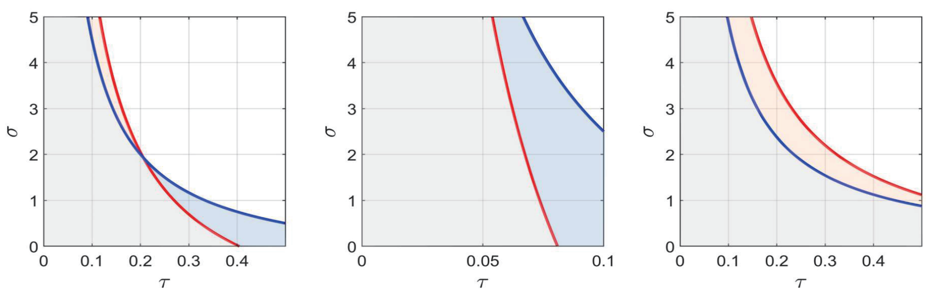

Figure 1: The sets of feasible parameters of Condat-Vu (with the blue line as boundary) and E-GRPDA (with the red line as boundary). The three plots from left to right corresponds to $\|K\| = \sqrt{2}$ and $L = 1, 10, 0.5$, respectively, with $\mu = 0$ and $\psi = \frac{\sqrt{5}+1}{2}$.

It is easy to show that the parameter conditions guaranteeing global convergence of the above methods do not have simple inclusion relations. Roughly speaking, if L is large, the region of parameters allowed by Condat-Vu, PDFP and PD3O is wider than that of E-GRPDA. On the other hand, if L is small, i.e., when h is approximately linear, E-GRPDA has a wider parameter region guaranteeing global convergence. To show this clearly, we have plotted for several scenarios the boundaries of the sets of feasible parameters of Condat-Vu and E-GRPDA in Figure 1.

It can be seen from Figure 1 that when $\|K\|$ is fixed, the region guaranteeing convergence of E-GRPDA is narrower than that of Condat-Vu in the case of $L = 10$ (the plot in the middle), while the opposite is true in the case of $L = 0.5$ (the plot on the right-hand-side). For $L = 1$, the convergence regions of the two algorithms do not have simple inclusion relation (the plot on the left-hand-side).

3 Analysis of E-GRPDA

The E-GRPDA (2.1) has been proposed and analyzed in [24] in the general convex case, where global iterate convergence of $\{(x_n, y_n) : n \geq 1\}$ to a primal-dual optimal solution of (1.3) has been established under the condition $\frac{\psi}{\tau} - \sigma\|K\|^2 > 2L$, where $\psi \in (1, \phi]$.

Furthermore, it has been shown in [24] that the primal-dual gap function converges at the $\mathcal{O}(1/N)$ sublinear rate. As pointed above, see also [2], a major flaw of the primal-dual gap function is that it could vanish at nonstationary points, in which case the sublinear convergence rate is less informative. In this section, we focus on the general convex case and establish $\mathcal{O}(1/N)$ convergence rate results for E-GRPDA with fixed step sizes. Instead of the primal-dual gap function, we adopt conventional optimality measures for the constrained optimization problem (1.4), i.e., function value residual and feasibility violation.

In the rest of this section, we let $t_n = n+1$ for $n \geq -1$. In particular, $t_{-1} = 0$. Our analysis is motivated by the recent work of Sabach and Teboulle [16], which proposed some principles for analyzing Lagrangian type methods. To take advantage of the analytic techniques proposed in [16], we need to represent the algorithm E-GRPDA. To begin with, we need the following lemma.

Lemma 3.1. *Choose $x_0 \in \mathbb{R}^q$, $w_0 \in \mathbb{R}^p$ and $y_0 \in \mathbb{R}^p$, and set $(\widetilde{x}_0, \widetilde{w}_0, \widetilde{y}_0) = (x_0, w_0, y_0)$ and*

$$\widetilde{y}_{n+1} = \widetilde{y}_n + \mu\sigma(Kx_{n+1} - w_{n+1}), \tag{3.1}$$

$$\begin{pmatrix} \widetilde{x}_{n+1} \\ \widetilde{w}_{n+1} \end{pmatrix} = (1 - t_n^{-1}) \begin{pmatrix} \widetilde{x}_n \\ \widetilde{w}_n \end{pmatrix} + t_n^{-1} \begin{pmatrix} x_{n+1} \\ w_{n+1} \end{pmatrix}, \tag{3.2}$$

for $n \geq 0$. Then, we have $y_n = \widetilde{y}_n + (1-\mu)\sigma t_{n-1}(K\widetilde{x}_n - \widetilde{w}_n)$ for all $n \geq 0$.

Proof. Let $n \geq 0$ be arbitrarily fixed. From (2.2) and (3.1), we obtain

$$y_{n+1} = y_n + \widetilde{y}_{n+1} - \widetilde{y}_n + (1-\mu)\sigma(Kx_{n+1} - w_{n+1}). \tag{3.3}$$

Noting $t_n - t_{n-1} = 1$ and the linearity of K, we deduce from (3.2) that

$$(Kx_{n+1} - w_{n+1}) = t_n(K\widetilde{x}_{n+1} - \widetilde{w}_{n+1}) - t_{n-1}(K\widetilde{x}_n - \widetilde{w}_n).$$

This together with (3.3) implies for all $n \geq 0$ that

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

Since $\widetilde{y}_0 = y_0$ and $t_{-1} = 0$, the right hand side is 0 for $n = 0$. This completes the proof. □

We emphasize that the auxiliary sequences $\{(\widetilde{x}_n, \widetilde{w}_n, \widetilde{y}_n) : n \geq 0\}$ 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. By using Lemma 3.1, we can represent E-GRPDA (2.1) or (2.2) formally as follows.

Algorithm 3.2 (E-GRPDA)**.**

Step 0. Let $\tau, \sigma > 0$, $\mu \in (0,1)$ and $\psi \in (1, \phi]$ with $\tau(2L + \sigma\|K\|^2/(1-\mu)) \leq \psi$. Choose $(x_0, y_0) \in \mathbb{R}^q \times \mathbb{R}^p$ and $\widetilde{w}_0 \in \mathbb{R}^p$. Set $(\widetilde{x}_0, \widetilde{y}_0) = (x_0, y_0)$, $z_0 = x_0$, $t_0 = 1$ and $n = 0$.

Step 1. Compute

$$\begin{aligned}
z_{n+1} &= \Big(1-\frac{1}{\psi}\Big)x_n + \frac{1}{\psi}z_n, && (3.4a)\\
x_{n+1} &= \text{Prox}_{\tau g}\big(z_{n+1}-\tau K^\top y_n - \tau\nabla h(x_n)\big), && (3.4b)\\
w_{n+1} &= \text{Prox}_{\frac{1}{\sigma}f}(y_n/\sigma + Kx_{n+1}), && (3.4c)\\
\widetilde{y}_{n+1} &= \widetilde{y}_n + \mu\sigma(Kx_{n+1}-w_{n+1}), && (3.4d)\\
\begin{pmatrix}\widetilde{x}_{n+1}\\ \widetilde{w}_{n+1}\end{pmatrix} &= (1-t_n^{-1})\begin{pmatrix}\widetilde{x}_n\\ \widetilde{w}_n\end{pmatrix} + t_n^{-1}\begin{pmatrix}x_{n+1}\\ w_{n+1}\end{pmatrix}, && (3.4e)\\
y_{n+1} &= \widetilde{y}_{n+1} + (1-\mu)\sigma t_n(K\widetilde{x}_{n+1}-\widetilde{w}_{n+1}). && (3.4f)
\end{aligned}$$

Step 2. Update $t_{n+1} = t_n + 1$. Set $n \leftarrow n+1$ and return to Step 1.

We next establish some lemmas, which are useful in our convergence rate analysis.

Lemma 3.3. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 3.2. Then, for any $(\bar{x}, \bar{w}, \bar{y}) \in \mathcal{S}$ and $n \geq 1$, we have*

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

Proof. It follows from (3.4b) and Lemma 1.1 with $\gamma = 0$ that

$$\begin{aligned}
g(x_{n+1}) - g(\bar{x}) &\leq \frac{1}{\tau}\langle x_{n+1}-z_{n+1}+\tau K^\top y_n + \tau\nabla h(x_n),\ \bar{x}-x_{n+1}\rangle\\
&= \frac{1}{\tau}\langle x_{n+1}-z_{n+1},\ \bar{x}-x_{n+1}\rangle + \langle y_{n-1}+\sigma(Kx_n-w_n),\ K\bar{x}-Kx_{n+1}\rangle\\
&\quad +\langle\nabla h(x_n),\ \bar{x}-x_{n+1}\rangle, && (3.5)\\
g(x_n) - g(x_{n+1}) &\leq \frac{1}{\tau}\langle x_n - z_n + \tau K^\top y_{n-1} + \tau\nabla h(x_{n-1}),\ x_{n+1}-x_n\rangle\\
&= \langle\frac{\psi}{\tau}(x_n - z_{n+1}) + K^\top y_{n-1} + \nabla h(x_{n-1}),\ x_{n+1}-x_n\rangle, && (3.6)
\end{aligned}$$

where the equality in (3.6) follows from $x_n - z_n = \psi(x_n - z_{n+1})$. Similarly, it holds that

$$f(w_n) - f(\bar{w}) \leq -\big\langle y_{n-1} + \sigma(Kx_n - w_n),\ \bar{w} - w_n\big\rangle. \tag{3.7}$$

Moreover, from the convexity of h it is clear that

$$h(x_n) - h(\bar{x}) \leq \langle\nabla h(x_n), x_n - \bar{x}\rangle. \tag{3.8}$$

By combining (3.5), (3.6), (3.7) as well as (3.8), and noting that $K\bar{x} - \bar{w} = 0$, we derive

$$
\begin{aligned}
& \mathcal{L}_\sigma(x_n, w_n, y_{n-1}) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y_{n-1}) \\
= \ & g(x_n) + f(w_n) + h(x_n) + \langle y_{n-1}, Kx_n - w_n\rangle + \frac{\sigma}{2}\|Kx_n - w_n\|^2 - (g(\bar{x}) + f(\bar{w}) + h(\bar{x})) \\
\le \ & \frac{1}{\tau}\langle x_{n+1} - z_{n+1},\ \bar{x} - x_{n+1}\rangle + \frac{\psi}{\tau}\langle x_n - z_{n+1},\ x_{n+1} - x_n\rangle \\
& + \langle y_{n-1}, (K\bar{x} - Kx_{n+1}) - (\bar{w} - w_n) + (Kx_{n+1} - Kx_n) + (Kx_n - w_n)\rangle \\
& + \sigma\langle Kx_n - w_n, (K\bar{x} - Kx_{n+1}) - (\bar{w} - w_n)\rangle + \frac{\sigma}{2}\|Kx_n - w_n\|^2 \\
& + \langle \nabla h(x_n) - \nabla h(x_{n-1}),\ x_n - x_{n+1}\rangle \\
= \ & \frac{1}{\tau}\langle x_{n+1} - z_{n+1},\ \bar{x} - x_{n+1}\rangle + \frac{\psi}{\tau}\langle x_n - z_{n+1},\ x_{n+1} - x_n\rangle \\
& + \sigma\langle Kx_n - w_n, w_n - Kx_{n+1}\rangle \\
& + \frac{\sigma}{2}\|Kx_n - w_n\|^2 + \langle \nabla h(x_n) - \nabla h(x_{n-1}),\ x_n - x_{n+1}\rangle,
\end{aligned}
$$

which completes the proof. □

Lemma 3.4. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 3.2. Then, for any $(\bar{x}, \bar{w}, \bar{y}) \in \mathcal{S}$, $y \in \mathbb{R}^p$ and $n \ge 1$, we have*

$$
\begin{aligned}
\mathcal{L}_\sigma(x_n, w_n, y) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y) \le \ & \frac{1}{\tau}\Delta_P(\bar{x}, z_{n+1}, z_{n+2}) + \frac{1}{\mu\sigma}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n) \\
& - (1-\mu)\sigma t_{n-2}\langle K\widetilde{x}_{n-1} - \widetilde{w}_{n-1}, Kx_n - w_n\rangle \\
& - \frac{1}{2}\left(\frac{\psi}{\tau} - \frac{\sigma}{1-\mu}\|K\|^2 - L\right)\|x_{n+1} - x_n\|^2 \\
& + \frac{L}{2}\|x_n - x_{n-1}\|^2,
\end{aligned}
$$

where $P = \frac{\psi}{\psi-1}I$.

Proof. By using the Cauchy-Schwarz inequality, we obtain

$$
\begin{aligned}
\langle Kx_n - w_n, w_n - Kx_{n+1}\rangle = \ & -\|Kx_n - w_n\|^2 + \langle Kx_n - w_n, Kx_n - Kx_{n+1}\rangle \\
\le \ & -\|Kx_n - w_n\|^2 \\
& + \frac{1}{2(1-\mu)}\|x_{n+1} - x_n\|^2_{K^\top K} + \frac{1-\mu}{2}\|Kx_n - w_n\|^2 \\
= \ & \frac{1}{2(1-\mu)}\|x_{n+1} - x_n\|^2_{K^\top K} - \frac{1+\mu}{2}\|Kx_n - w_n\|^2. \quad (3.9)
\end{aligned}
$$

Combining the result in Lemma 3.3, Pythagoras three-points identity (1.10) and (3.9), we have

$$\begin{aligned}\mathcal{L}_\sigma(x_n, w_n, y_{n-1}) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y_{n-1}) \leq & \frac{1}{2\tau}\big(\|z_{n+1} - \bar{x}\|^2 - \|x_{n+1} - z_{n+1}\|^2 - \|x_{n+1} - \bar{x}\|^2\big) \\ & + \frac{\psi}{2\tau}\big(\|z_{n+1} - x_{n+1}\|^2 - \|z_{n+1} - x_n\|^2 - \|x_{n+1} - x_n\|^2\big) \\ & + \frac{\sigma}{2(1-\mu)}\|x_{n+1} - x_n\|^2_{K^\top K} - \frac{\mu\sigma}{2}\|Kx_n - w_n\|^2 \\ & + \langle \nabla h(x_n) - \nabla h(x_{n-1}),\ x_n - x_{n+1}\rangle \\ = & \frac{1}{2\tau}\big(\|z_{n+1} - \bar{x}\|^2 - \|x_{n+1} - z_{n+1}\|^2\big) - \frac{1}{2\tau}\|x_{n+1} - \bar{x}\|^2 \\ & + \frac{\psi}{2\tau}\big(\|z_{n+1} - x_{n+1}\|^2 - \|z_{n+1} - x_n\|^2\big) - \frac{\mu\sigma}{2}\|Kx_n - w_n\|^2 \\ & - \frac{1}{2\tau}\|x_{n+1} - x_n\|^2_{\psi I - \frac{\sigma\tau}{1-\mu}K^\top K} + \langle \nabla h(x_n) - \nabla h(x_{n-1}),\ x_n - x_{n+1}\rangle.\end{aligned} \tag{3.10}$$

Since $x_{n+1} = \frac{\psi}{\psi-1}z_{n+2} - \frac{1}{\psi-1}z_{n+1}$, it follows from (1.11) that

$$\begin{aligned}\|x_{n+1} - \bar{x}\|^2 &= \frac{\psi}{\psi-1}\|z_{n+2} - \bar{x}\|^2 - \frac{1}{\psi-1}\|z_{n+1} - \bar{x}\|^2 + \frac{\psi}{(\psi-1)^2}\|z_{n+2} - z_{n+1}\|^2 \\ &= \frac{\psi}{\psi-1}\|z_{n+2} - \bar{x}\|^2 - \frac{1}{\psi-1}\|z_{n+1} - \bar{x}\|^2 + \frac{1}{\psi}\|x_{n+1} - z_{n+1}\|^2,\end{aligned} \tag{3.11}$$

where the second equality is due to $z_{n+2} - z_{n+1} = \frac{\psi-1}{\psi}(x_{n+1} - z_{n+1})$. By plugging (3.11) into (3.10) and omitting the term $\|z_{n+1} - x_n\|^2$, we obtain

$$\begin{aligned}\mathcal{L}_\sigma(x_n, w_n, y_{n-1}) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y_{n-1}) &\leq \frac{\psi}{2\tau(\psi-1)}\Big(\|z_{n+1} - \bar{x}\|^2 - \|z_{n+2} - \bar{x}\|^2\Big) \\ & - \frac{\mu\sigma}{2}\|Kx_n - w_n\|^2 - \frac{1}{2\tau}\Big(\frac{1}{\psi} - \psi + 1\Big)\|z_{n+1} - x_{n+1}\|^2 \\ & - \frac{1}{2\tau}\|x_{n+1} - x_n\|^2_{\psi I - \frac{\sigma\tau}{1-\mu}K^\top K} + L\|x_n - x_{n-1}\|\|x_n - x_{n+1}\| \\ \leq\ & \frac{1}{\tau}\Delta_P(\bar{x}, z_{n+1}, z_{n+2}) - \frac{\mu\sigma}{2}\|Kx_n - w_n\|^2 \\ & - \frac{1}{2\tau}\|x_{n+1} - x_n\|^2_{\psi I - \frac{\sigma\tau}{1-\mu}K^\top K} + \frac{L}{2}\|x_n - x_{n-1}\|^2 + \frac{L}{2}\|x_n - x_{n+1}\|^2 \\ \leq\ & \frac{1}{\tau}\Delta_P(\bar{x}, z_{n+1}, z_{n+2}) - \frac{\mu\sigma}{2}\|Kx_n - w_n\|^2 \\ & - \frac{1}{2}\left(\frac{\psi}{\tau} - \frac{\sigma}{1-\mu}\|K\|^2 - L\right)\|x_{n+1} - x_n\|^2 + \frac{L}{2}\|x_n - x_{n-1}\|^2,\end{aligned}$$

where the second inequality follows from $\frac{1}{\psi} - \psi + 1 \geq 0$ as $\psi \leq \phi$. Using Lemma 3.1, the notation Δ and (1.10), we obtain

$$\begin{aligned}\langle y - \widetilde{y}_{n-1}, Kx_n - w_n\rangle &= \frac{1}{\mu\sigma}\langle y - \widetilde{y}_{n-1}, \widetilde{y}_n - \widetilde{y}_{n-1}\rangle \\ &= \frac{1}{\mu\sigma}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n) + \frac{\mu\sigma}{2}\|Kx_n - w_n\|^2.\end{aligned} \tag{3.12}$$

Now considering $y_{n-1} = \widetilde{y}_{n-1} + (1-\mu)\sigma t_{n-2}(K\widetilde{x}_{n-1} - \widetilde{w}_{n-1})$, we deduce from (3.12) that

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

Note that $\mathcal{L}_\sigma(\bar{x}, \bar{w}, y_{n-1}) = \mathcal{L}_\sigma(\bar{x}, \bar{w}, y)$. The proof of Lemma 3.4 is completed by adding the above equality to (3.12). □

Lemma 3.5. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 3.2. If*

$$\frac{\psi}{\tau} - \frac{\sigma}{1-\mu}\|K\|^2 \geq 2L \quad \textit{with} \quad \mu \in (0,1), \tag{3.13}$$

then, for any $(\bar{x}, \bar{w}, \bar{y}) \in \mathcal{S}$, $y \in \mathbb{R}^p$ and $n \geq 1$, we have

$$\begin{aligned}t_{n-1}S^n_{(1-\mu)\sigma t_{n-1}} - t_{n-2}S^{n-1}_{(1-\mu)\sigma t_{n-2}} \leq\ & \frac{1}{\tau}\Delta_P(\bar{x}, z_{n+1}, z_{n+2}) + \frac{1}{\mu\sigma}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n) \\ & -\frac{L}{2}\|x_{n+1} - x_n\|^2 + \frac{L}{2}\|x_n - x_{n-1}\|^2,\end{aligned} \tag{3.14}$$

where $P = \frac{\psi}{\psi-1}I$ and S^n_β is defined as

$$S^n_\beta := \mathcal{L}_\beta(\widetilde{x}_n, \widetilde{w}_n, y) - \mathcal{L}_\beta(\bar{x}, \bar{w}, y) = \mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, y) - \mathcal{L}(\bar{x}, \bar{w}, y) + \frac{\beta}{2}\|K\widetilde{x}_n - \widetilde{w}_n\|^2. \tag{3.15}$$

Proof. Let $y \in \mathbb{R}^p$ be arbitrarily fixed. Then, it is clear from (3.4e), the linearity of K and the convexity of $\Phi(\cdot,\cdot)$ that

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

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

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

Adding the above two relations and using the definition of $\mathcal{L}(\cdot)$ in (1.8), we arrive at

$$\begin{aligned}t_{n-1}\big(\mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, y) - \mathcal{L}(\bar{x}, \bar{w}, y)\big) - t_{n-2}\big(\mathcal{L}(\widetilde{x}_{n-1}, \widetilde{w}_{n-1}, y) - \mathcal{L}(\bar{x}, \bar{w}, y)\big) \\ \leq \mathcal{L}(x_n, w_n, y) - \mathcal{L}(\bar{x}, \bar{w}, y).\end{aligned} \tag{3.16}$$

In addition, taking into account (3.4e) and (1.11), we derive

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

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

$$\begin{aligned}\sigma t_{n-1}^2/2\|K\widetilde{x}_n - \widetilde{w}_n\|^2 - \sigma t_{n-2}^2/2\|K\widetilde{x}_{n-1} - \widetilde{w}_{n-1}\|^2 \\ = \sigma/2\|Kx_n - w_n\|^2 + \sigma t_{n-2}\langle K\widetilde{x}_{n-1} - \widetilde{w}_{n-1}, Kx_n - w_n\rangle.\end{aligned} \tag{3.17}$$

Therefore, by adding (3.16) to (3.17) and using the definition of S^n_β in (3.15), we deduce

$$t_{n-1}S^n_{\sigma t_{n-1}} - t_{n-2}S^{n-1}_{\sigma t_{n-2}} \le (\mathcal{L}_\sigma(x_n, w_n, y) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y)) + \sigma t_{n-2}\langle K\widetilde{x}_{n-1} - \widetilde{w}_{n-1}, Kx_n - w_n\rangle.$$

Finally, multiplying both sides of the above inequality by $(1-\mu)$ to obtain

$$\begin{aligned}
& (1-\mu)(t_{n-1}S^n_{\sigma t_{n-1}} - t_{n-2}S^{n-1}_{\sigma t_{n-2}}) \\
\le\ & (1-\mu)(\mathcal{L}_\sigma(x_n, w_n, y) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y)) + (1-\mu)\sigma t_{n-2}\langle K\widetilde{x}_{n-1} - \widetilde{w}_{n-1}, Kx_n - w_n\rangle \\
\le\ & -\mu(\mathcal{L}_\sigma(x_n, w_n, y) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y)) + \frac{1}{\tau}\Delta_P(\bar{x}, z_{n+1}, z_{n+2}) + \frac{1}{\mu\sigma}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n) \\
& -\frac{L}{2}\|x_{n+1} - x_n\|^2 + \frac{L}{2}\|x_n - x_{n-1}\|^2,
\end{aligned} \tag{3.18}$$

where the second inequality follows from Lemma 3.4 and (3.13). Furthermore, it follows from (3.16) and the definitions of $\mathcal{L}_\sigma$ and $\mathcal{L}$ that

$$\begin{aligned}
\mathcal{L}_\sigma(x_n, w_n, y) - \mathcal{L}_\sigma(\bar{x}, \bar{w}, y) &= \mathcal{L}(x_n, w_n, y) - \mathcal{L}(\bar{x}, \bar{w}, y) + \frac{\sigma}{2}\|Kx_n - w_n\|^2 \\
&\ge t_{n-1}\big(\mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, y) - \mathcal{L}(\bar{x}, \bar{w}, y)\big) - t_{n-2}\big(\mathcal{L}(\widetilde{x}_{n-1}, \widetilde{w}_{n-1}, y) - \mathcal{L}(\bar{x}, \bar{w}, y)\big).
\end{aligned} \tag{3.19}$$

Combining (3.18) and (3.19), we obtain

$$\begin{aligned}
& (1-\mu)(t_{n-1}S^n_{\sigma t_{n-1}} - t_{n-2}S^{n-1}_{\sigma t_{n-2}}) \\
\le\ & -\mu t_{n-1}\big(\mathcal{L}(\widetilde{x}_n, \widetilde{w}_n, y) - \mathcal{L}(\bar{x}, \bar{w}, y)\big) + \mu t_{n-2}\big(\mathcal{L}(\widetilde{x}_{n-1}, \widetilde{w}_{n-1}, y) - \mathcal{L}(\bar{x}, \bar{w}, y)\big) \\
& + \frac{1}{\tau}\Delta_P(\bar{x}, z_{n+1}, z_{n+2}) + \frac{1}{\mu\sigma}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n) - \frac{L}{2}\|x_{n+1} - x_n\|^2 + \frac{L}{2}\|x_n - x_{n-1}\|^2.
\end{aligned}$$

Finally, (3.14) follows from the above inequality and the definition of S^n_β in (3.15). □

Now, we are ready to establish the convergence rate of E-GRPDA.

Theorem 3.6. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 3.2, $(\bar{x}, \bar{w}, \bar{y})$ be any saddle point, i.e., an element in $\mathcal{S}$, and $\nu > 0$ be a constant such that $\nu \ge 2\|\bar{y}\|$. Then, there exists constant $C > 0$ such that for any $N \ge 1$ there hold*

$$|\Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) - \Phi(\bar{x}, \bar{w})| \le \frac{C}{N} \quad \text{and} \quad \|K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\| \le \frac{2C}{\nu N}. \tag{3.20}$$

Proof. Recall that $\Delta_P(u, v, w) := \frac{1}{2}(\|u-v\|_P^2 - \|u-w\|_P^2)$, $\Delta(u, v, w) = \frac{1}{2}(\|u-v\|^2 - \|u-w\|^2)$ and $t_{-1} = 0$. The sum of (3.14) for $n = 1, \ldots, N$ yields

$$t_{N-1}(\Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) - \Phi(\bar{x}, \bar{w}) + \langle y, K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\rangle) \le C(y),$$

where $C(y) := \frac{1}{2\tau}\|\bar{x} - z_2\|_P^2 + \frac{1}{2\mu\sigma}\|y - \widetilde{y}_0\|^2 + \frac{L}{2}\|x_1 - x_0\|^2$, and we have dropped on the left-hand-side the quadratic term in $S^N_{(1-\mu)\sigma t_{N-1}}$. By taking the maximum of both sides over $\|y\| \le \nu$ and noting $t_{N-1} = N$, we obtain

$$\Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) - \Phi(\bar{x}, \bar{w}) + \nu\|K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\| \le \frac{C}{N}, \tag{3.21}$$

where $C = \frac{1}{2\tau}\|\bar{x} - z_2\|_P^2 + \frac{1}{2\mu\sigma}(\nu + \|\widetilde{y}_0\|)^2 + \frac{L}{2}\|x_0 - x_1\|^2$ is an upper bound of $C(y)$ over $\|y\| \leq \nu$. Then, it is clear that $\Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) - \Phi(\bar{x}, \bar{w}) \leq \frac{C}{N}$. Besides, since $(\bar{x}, \bar{w}, \bar{y})$ is a saddle point of $\mathcal{L}(x, w, y)$, then by $\mathcal{L}(\bar{x}, \bar{w}, \bar{y}) \leq \mathcal{L}(\tilde{x}_{N-1}, \tilde{w}_{N-1}, \bar{y})$ and $\|\bar{y}\| \leq \nu/2$ we have

$$\Phi(\bar{x}, \bar{w}) - \Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) \leq \langle \bar{y}, K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\rangle \leq \frac{\nu}{2}\|K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\|. \tag{3.22}$$

Finally, the conclusion (3.20) follows from the combination of (3.21) and (3.22). □

Recall that $P = \frac{\psi}{\psi-1}I$ and thus $\|\cdot\|_P$, which is used repeatedly on the above, is not a generic weighted norm. We keep using the notation $\|\cdot\|_P$ and $\Delta_P(\cdot)$ because this way the coefficient $\frac{\psi}{\psi-1}$ can be absorbed, resulting to simpler formulas, and the recent work [16], which motivated this work, adopted this same notation.

4 Analysis of an accelerated E-GRPDA

As shown in the literature [2, 14, 3], faster $\mathcal{O}(1/N^2)$ convergence rate can be achieved if one of the component functions in (1.3) is strongly convex. In this section, we show that Algorithm 3.2 can be modified so that when g or h is strongly convex the modified algorithm achieves the same accelerated convergence rate measured by the same criteria, i.e., function value residual and feasibility violation. We shall only treat the case when g is strongly convex, while the case when h is strongly convex is completely analogues, simply, for h, due to the definition of strongly convex, the same extra added term $-\frac{\gamma}{2}\|\bar{x} - x_{n+1}\|^2$ can be obtained in Lemma 4.3 compared with Lemma 3.3, and then satisfies all the consequences thereafter.

Assumption 4.1. Assume that g is γ-strongly convex, which means

$$g(y) \geq g(x) + \langle u, y - x\rangle + \frac{\gamma}{2}\|y - x\|^2, \quad \forall u \in \partial g(x).$$

For this case, we propose the following accelerated variant of Algorithm 3.2.

Algorithm 4.1 (Accelerated E-GRPDA)**.**

Step 0. Let $\rho > 0$, $\sigma_0 = \rho$, $\mu \in (0, 1)$, $\psi \in (1, \phi]$ and $\tau_0 = \psi/\big(\frac{L(\psi+1)}{2} + \frac{\sigma_0\|K\|^2}{1-\mu}\big)$. Choose $(x_0, y_0) \in \mathbb{R}^q \times \mathbb{R}^p$ and $\widetilde{w}_0 \in \mathbb{R}^p$. Set $(\widetilde{x}_0, \widetilde{y}_0) = (x_0, y_0)$, $z_0 = x_0$, $t_0 = 1$ and $n = 0$.

Step 1. Compute

$$\begin{aligned}
z_{n+1} &= \Big(1 - \frac{1}{\psi}\Big)x_n + \frac{1}{\psi}z_n, \\
x_{n+1} &= \mathrm{Prox}_{\tau_n g}(z_{n+1} - \tau_n K^\top y_n - \tau_n \nabla h(x_n)), \\
w_{n+1} &= \mathrm{Prox}_{f/\sigma_n}(y_n/\sigma_n + Kx_{n+1}), \\
\widetilde{y}_{n+1} &= \widetilde{y}_n + \mu\sigma_n(Kx_{n+1} - w_{n+1}), \\
\begin{pmatrix}\widetilde{x}_{n+1} \\ \widetilde{w}_{n+1}\end{pmatrix} &= (1 - t_n^{-1})\begin{pmatrix}\widetilde{x}_n \\ \widetilde{w}_n\end{pmatrix} + t_n^{-1}\begin{pmatrix}x_{n+1} \\ w_{n+1}\end{pmatrix}, \\
y_{n+1} &= \widetilde{y}_{n+1} + (1 - \mu)\rho t_n^2(K\widetilde{x}_{n+1} - \widetilde{w}_{n+1}).
\end{aligned}$$

Step 2. Update the parameters

$$t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2},\ \sigma_{n+1} = \rho t_{n+1},\ \tau_{n+1} = \psi/\Big(\frac{L(\psi+1)}{2} + \frac{\sigma_n\|K\|^2}{1-\mu}\Big).$$

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

We emphasize that Step 1 of Algorithm 4.1 is a representation of the three-line scheme (2.1) with $\tau = \tau_n$ and $\sigma = \sigma_n$ by using a result similar to Lemma 3.1. Again, this representation is only convenient for analysis and is not necessary in implementation.

We first present some useful properties of the sequence $\{t_n : n \geq 0\}$ defined in Algorithm 4.1. Since these properties are easy to verify, we omit their proofs.

Lemma 4.2. *Let $t_0 = 1$ and $t_{n+1} = \big(1+\sqrt{1+4t_n^2}\big)/2$ for $n \geq 0$. We have*

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

(ii) *$t_n^2 = t_n + t_{n-1}^2$ and $t_n^2 \leq 2t_{n-1} + t_{n-1}^2$;*

(iii) *$\frac{t_n}{t_{n-1}} \in (1, \phi]$ and $t_n - t_{n-1}$ are monotonically decreasing;*

(iv) *for any $\kappa > 1$, if $n \geq \lfloor 2\kappa \rfloor$, where $\lfloor a \rfloor$ denotes the largest integer no greater than a, then $t_{n-1} > \kappa$ and $\frac{t_n}{t_{n-1}} < \frac{1+\sqrt{1+4\kappa^2}}{2\kappa}$.*

We next present some useful results in Lemmas 4.3, 4.4 and 4.5, which are completely analogues to Lemmas 3.3, 3.4 and 3.5, respectively. The key difference is that some relevant inequalities can be enhanced using the γ-strong convexity of g, as presented in Lemma 1.1 with $\gamma > 0$. Due to the high similarity of their respective proofs, we omit the details for simplicity.

Lemma 4.3. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 4.1. Then, for any $(\bar{x}, \bar{w}, \bar{y}) \in \mathcal{S}$ and $n \geq 1$, we have*

$$\begin{aligned}\mathcal{L}_{\sigma_{n-1}}(x_n, w_n, y_{n-1}) - \mathcal{L}_{\sigma_{n-1}}(\bar{x}, \bar{w}, y_{n-1}) \leq{}& \frac{1}{\tau_n}\langle x_{n+1} - z_{n+1},\ \bar{x} - x_{n+1}\rangle \\ &+ \frac{\psi}{\tau_{n-1}}\langle x_n - z_{n+1},\ x_{n+1} - x_n\rangle + \sigma_{n-1}\langle Kx_n - w_n, w_n - Kx_{n+1}\rangle \\ &+ \frac{\sigma_{n-1}}{2}\|Kx_n - w_n\|^2 + \langle \nabla h(x_n) - \nabla h(x_{n-1}),\ x_n - x_{n+1}\rangle \\ &- \frac{\gamma}{2}\|\bar{x} - x_{n+1}\|^2.\end{aligned}$$

Lemma 4.4. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 4.1. Then, for any $(\bar{x}, \bar{w}, \bar{y}) \in \mathcal{S}$, $y \in \mathbb{R}^p$ and $n \geq 1$, we have*

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

where

$$P_n = \frac{\psi}{\psi - 1}\big(1 + \frac{\gamma\tau_n}{\psi}\big)I. \tag{4.2}$$

Lemma 4.5. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 4.1. Then, for any $(\bar{x}, \bar{w}, \bar{y}) \in \mathcal{S}$, $y \in \mathbb{R}^p$ and $n \geq 1$, we have*

$$t_{n-1}^2 S^n_{(1-\mu)\rho t_{n-1}^2} - t_{n-2}^2 S^{n-1}_{(1-\mu)\rho t_{n-2}^2} \leq \frac{t_{n-1}}{\tau_n}\Delta_{P_n}(\bar{x}, z_{n+1}, z_{n+2}) + \frac{1}{\mu\rho}\Delta(y, \widetilde{y}_{n-1}, \widetilde{y}_n) + \frac{t_{n-1}L}{2}\big(\|x_n - x_{n-1}\|^2 - \psi\|x_n - x_{n+1}\|^2\big) - \frac{\gamma t_{n-1}}{2}\|\bar{x} - z_{n+2}\|^2,$$

where P_n is defined in (4.2) and S^n_β in (3.15).

Now, based on the results of Lemmas 4.3, 4.4 and 4.5, we are ready to establish the accelerated convergence results of Algorithm 4.1 when g is γ-strongly convex. Note that the strong convexity module $\gamma > 0$ is only used in the analysis but not in the algorithm itself.

Theorem 4.6. *Let $\{(z_n, x_n, y_n, w_n, \widetilde{x}_n, \widetilde{y}_n, \widetilde{w}_n)\}$ be the sequence generated by Algorithm 4.1. Define*

$$n_0 = \lfloor 2a/(a^2 - 1)\rfloor + 1 \quad \text{with} \quad a = \frac{L(\psi+1) + \psi\gamma}{L(\psi+1) + \gamma}, \tag{4.3}$$

$$\pi(n) = \frac{(t_{n+1} - t_{n-1})t_n}{\psi t_{n-1} - t_n} \quad \text{and} \quad q(n) = \frac{t_n - t_{n-1}}{\psi t_{n-1} - t_n}. \tag{4.4}$$

Let $\rho \in (0, \frac{\gamma(1-\mu)}{2\|K\|^2\pi(n_0)}]$, $(\bar{x}, \bar{w}, \bar{y}) \in \mathcal{S}$ be a saddle point,and $\nu > 0$ be a constant such that $\nu \geq 2\|\bar{y}\|$. Then, there exist constants $C_1, C_2 > 0$ such that for any $N \geq 1$ we have

$$|\Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) - \Phi(\bar{x}, \bar{w})| \leq \frac{C_1}{(N+1)^2}\ , \ \|K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\| \leq \frac{2C_1}{\nu(N+1)^2}, \tag{4.5}$$

as well as $\|\bar{x} - z_{N+2}\| \leq \sqrt{2C_1/C_2}/(N+2)$.

Proof. Recall that P_n is defined in (4.2). Then, it is elementary to verify that

$$\frac{1}{\tau_n}\Delta_{P_n}(\bar{x}, z_{n+1}, z_{n+2}) - \frac{\gamma}{2}\|\bar{x} - z_{n+2}\|^2 = a_n\|\bar{x} - z_{n+1}\|^2 - a_{n+1}b_n\|\bar{x} - z_{n+2}\|^2,$$

where

$$a_n := \frac{1}{2\tau_n}\frac{\psi + \gamma\tau_n}{\psi - 1} \quad \text{and} \quad b_n := \frac{\psi + \psi\gamma\tau_n}{\psi + \gamma\tau_{n+1}}\frac{\tau_{n+1}}{\tau_n} = \frac{\frac{L(\psi+1)}{2} + \frac{\rho t_n}{1-\mu}\|K\|^2 + \gamma\psi}{\frac{L(\psi+1)}{2} + \frac{\rho t_{n+1}}{1-\mu}\|K\|^2 + \gamma}.$$

It follows from Lemma 4.2 and the definitions of $\pi(n)$ and $q(n)$ in (4.4) that $\pi(n)$ is non-increasing and $q(n)$ decreases monotonically with $q(n) \to 0$ as $n \to \infty$. Moreover, it follows from (iv) of Lemma 4.2 and (4.3) that $t_n/t_{n-1} < \frac{L(\psi+1)+\psi\gamma}{L(\psi+1)+\gamma} < \psi$. Since $\rho \in (0, \frac{\gamma(1-\mu)}{2\|K\|^2\pi(n_0)}]$, we obtain

$$\frac{\rho\|K\|^2}{1-\mu}\pi(n) + \frac{L(\psi+1)}{2}q(n) \leq \frac{\gamma}{2} + \frac{\gamma}{2} \leq \gamma, \quad \forall n \geq n_0,$$

which implies $b_n \geq \frac{t_n}{t_{n-1}}$ for any $n \geq n_0$. Then it follows from Lemma 4.5 that

$$\begin{aligned} t_{n-1}^2 S^n_{(1-\mu)\rho t_{n-1}^2} - t_{n-2}^2 S^{n-1}_{(1-\mu)\rho t_{n-2}^2} &\leq t_{n-1}a_n\|\bar{x} - z_{n+1}\|^2 - t_n a_{n+1}\|\bar{x} - z_{n+2}\|^2 \\ &\quad + \frac{1}{2\mu\rho}(\|y - \widetilde{y}_{n-1}\|^2 - \|y - \widetilde{y}_n\|^2) \\ &\quad + \frac{L}{2}(t_{n-1}\|x_n - x_{n-1}\|^2 - t_n\|x_{n+1} - x_n\|^2). \end{aligned}$$

Sum this inequality for $n = n_0, \ldots, N$ and drop the quadratic term in $S^N_{(1-\mu)\rho t^2_{N-1}}$, we deduce

$$\begin{aligned} t^2_{N-1}(\Phi(\widetilde{x}_{N-1}, \widetilde{w}_{N-1}) - \Phi(\bar{x}, \bar{w}) + \langle y, K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\rangle) \\ \leq C' - t_N a_{N+1}\|\bar{x} - z_{n+2}\|^2 + \frac{1}{2\mu\rho}\|y - \widetilde{y}_{n_0-1}\|^2 + t^2_{n_0-2} S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}} \\ \leq C' + \frac{1}{2\mu\rho}\|y - \widetilde{y}_{n_0-1}\|^2 + t^2_{n_0-2} S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}}, \end{aligned} \tag{4.6}$$

where $C' := t_{n_0-1}a_{n_0}\|\bar{x} - z_{n_0+1}\|^2 + \frac{Lt_{n_0-1}}{2}\|x_{n_0} - x_{n_0-1}\|^2$. Note that $S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}}$ depends on y and is given by $S^{n_0-1}_{(1-\mu)\rho t^2_{n_0-2}} = C'' + \langle y, K\widetilde{x}_{n_0-1} - \widetilde{w}_{n_0-1}\rangle$, where

$$C'' := \Phi(\widetilde{x}_{n_0-1}, \widetilde{w}_{n_0-1}) - \Phi(\bar{x}, \bar{w}) + \frac{(1-\mu)\rho t^2_{n_0-2}}{2}\|K\widetilde{x}_{n_0-1} - \widetilde{w}_{n_0-1}\|^2.$$

By taking the maximum of both sides of (4.6) over $\|y\| \leq \nu$ with the fact $t^2_{N-1} > (N+1)^2/4$, we deduce

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

where $C_1 := 4C' + \frac{2}{\mu\rho}(\nu^2 + \|\widetilde{y}_{n_0-1}\|^2) + 2t^2_{n_0-2}\big(|C''| + \nu\|K\widetilde{x}_{N-1} - \widetilde{w}_{N-1}\|\big)$. Then, similar to Theorem 3.6, it is easy to derive (4.5). Furthermore, it is easy to verify that

$$2t_N a_{N+1} = \frac{t_N}{\tau_{N+1}}\frac{\psi + \gamma\tau_{N+1}}{\psi - 1} \geq \frac{t_N}{\tau_{N+1}}\frac{\psi}{\psi - 1} \geq C_2 t_{N+1} t_N \geq \frac{C_2(N+2)^2}{4}, \tag{4.7}$$

where $C_2 := \frac{\rho\|K\|^2}{(\psi-1)(1-\mu)}$. Finally, setting $y = \bar{y}$ in (4.6) and considering (3.22), we obtain

$$t_N a_{N+1}\|\bar{x} - z_{N+2}\|^2 \leq C_1/4,$$

which together with (4.7) implies that $\|\bar{x} - z_{N+2}\| \leq \sqrt{2C_1/C_2}/(N+2)$. □

5 Numerical Experiments

In this section, we conduct preliminary numerical experiments on computed tomography (CT) image reconstruction problem and elastic net regularization problem to validate the performances of the proposed algorithm (E-GRPDA). Based on our preliminary computational experience, under the choices of the same set of parameters, PDFP [4] and PD3O [22] perform very closely to Condat-Vu's algorithm [5, 21] in the following examples. For simplicity, we only present comparison results between E-GRPDA and Condat-Vu's PDA. Since the per-iteration cost of E-GRPDA is approximately identical to that of Condat-Vu's PDA, we do not present the comparison results of CPU time. Instead, we demonstrate how the (relative or absolute) errors measured by function value residual and deviation from the optimal solution are decreased as the algorithms proceed with iterations. Note that we actually implemented (2.1) with $\psi = 1.618$, which is equivalent to Algorithm 3.2. The algorithms were implemented within MATLAB R2018a on a desktop with Intel 64 Model 61 Stepping 4 Genuine Intel 1801 MHz and 8GB memory. We next describe the two tested problems.

Problem 5.1 (CT image reconstruction problem)**.** Consider CT image reconstruction problem of the following form

$$\min_{x\in\mathbb{R}^q} \frac{1}{2}\|Ax-b\|_2^2 + \iota_C(x) + \mu\|Kx\|_1,$$

where $A \in \mathbb{R}^{p\times q}$ is a Radon transform matrix with size 9250×16384, $K \in \mathbb{R}^{2q\times q}$ contains the global (horizontal and vertical) finite difference operators (with periodic boundary conditions) so that $\|Kx\|_1$ represents the anisotropic total variation regularization, C is the nonnegative orthant $\mathbb{R}^q_+$, and $b \in \mathbb{R}^p$ is the measurement vector contaminated by independent and normally distributed noise with mean 0 and variance 1. Note that the Radon transform matrix A maps an input image to the sinogram data. The data set we used here is the same as that in [23], where the true image is the 128×128 Shepp-Logan phantom image and the sinogram data is measured from 50 uniformly oriented projections. Moreover, the regularization parameter μ is set to be 0.05. For this example, we have $h(x) = \frac{1}{2}\|Ax-b\|_2^2$, $g(x) = \iota_C(x)$ and $f(w) = \mu\|w\|_1$. The experimental results are given in Figures 2 and 3.

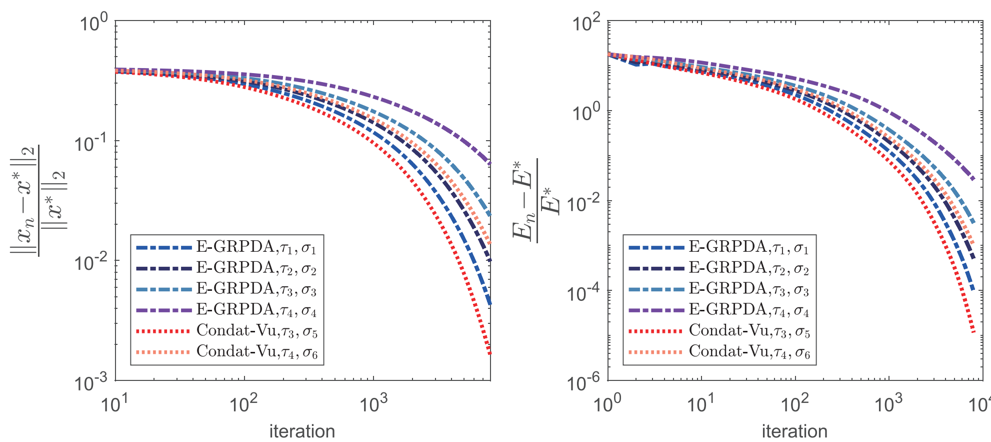

Figure 2: Evolution of the relative errors $\|x_n - x^*\|_2/\|x^*\|_2$ and $(E_n - E^*)/E^*$ for CT image reconstruction problem with different σ and τ, where $\tau_i = (5-i)/5L$ for $i = 1,2,3,4$, $\sigma_1 = \frac{0.09L}{4\|K\|^2}$, $\sigma_2 = \frac{2.09L}{3\|K\|^2}$, $\sigma_3 = \frac{4.09L}{2\|K\|^2}$, $\sigma_4 = \frac{6.09L}{\|K\|^2}$, $\sigma_5 = \frac{2.00L}{\|K\|^2}$ and $\sigma_6 = \frac{4.50L}{\|K\|^2}$. Here x_n denotes the n-th iterate, E_n represents the objective value at x_n, x^* and E^* denote, respectively, the true solution and the minimum function value, which can be computed approximately in advance by running, e.g., E-GRPDA, with sufficiently many iterations.

Problem 5.2 (Elastic net regularization)**.** The elastic net regularization combines ℓ_1 and ℓ_2 penalties of the lasso and ridge regressions linearly, which helps to overcome the limitations of each individual penalty and has been widely applied in, e.g., supported vector machine, metric learning, portfolio optimization and so on. The problem we consider here has the following form

$$\min_x \mu_1\|x\|_2^2 + \mu_2\|x\|_1 + l(Kx, b)$$

where l represents a loss function and can be chosen as zero-one loss, mean squared error, mean absolute error and so on. We set $l(Kx,b) = \|Kx-b\|^2$ to be the (mean) squared error function. The (i,j)-th element K_{ij} of K is randomly generated from $\mathcal{N}(0, 0.01)$, $x^* \sim \mathcal{N}(0,1)$ and $b = Kx^* + \epsilon$ with $\epsilon \sim \mathcal{N}(0, 0.04I)$. We tested $p = 1000$, $q = 300$, $\mu_1 = 0.005$ and $\mu_2 = 0.01$. For this example, we have $h(x) = \mu_1\|x\|_2^2$, $g(x) = \mu_2\|x\|_1$ and $f(w) = \|w-b\|^2$. The experimental results are given in Figure 4.

(a) Input Image (b) Reconstructed Image with E-GRPDA,τ_1, σ_1

Figure 3: Left: original input image (size 128×128); Right: reconstructed using E-GRPDA with τ_1 and σ_1.

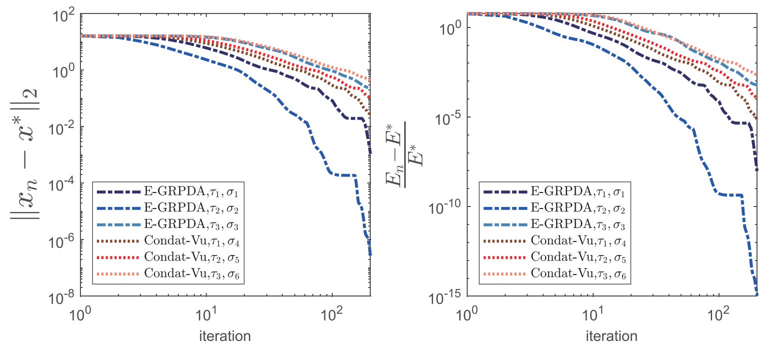

Figure 4: Evolution of the absolute error $\|x_n - x^*\|_2$ and the relative error $(E_n - E^*)/E^*$ for the elastic net regularization problem with different σ and τ, where $\tau_i = (4-i)/5L$ for $i = 1, 2, 3$, $\sigma_1 = \frac{2.09L}{3\|K\|^2}$, $\sigma_2 = \frac{4.09L}{2\|K\|^2}$, $\sigma_3 = \frac{6.09L}{\|K\|^2}$, $\sigma_4 = \frac{3.50L}{3\|K\|^2}$, $\sigma_5 = \frac{2.00L}{\|K\|^2}$ and $\sigma_6 = \frac{4.50L}{\|K\|^2}$. Here x_n, x^*, E_n and E^* have the same meaning as noted in Figure 2.

Note that for the CT image reconstruction problem, evaluating the proximity operator of $g + h$ is equivalent to solving a nonnegativity constrained generic least-squares problem, which apparently does not have closed form solution. In this case, the proposed full-splitting algorithm E-GRPDA, as well as PDFP [4], PD3O [22] and Condat-Vu [5, 21], is quite useful. On the other hand, for the elastic net regularization problem, $g + h$ is equal to a weighted sum of the ℓ_1-norm and the ℓ_2-norm squared, thus the proximity operator of $g + h$ indeed has closed form solution given by the so-called soft-thresholding operator. In this case, the problem can be solved by the original non-extended PDAs such as [2] and [3]. Nonetheless, our purpose here is to demonstrate the viability of the proposed algorithm.

It can be seen from Figure 2 that both E-GRPDA and Condat-Vu's PDA improve the quality of iterates steadily as the algorithms proceed and E-GRPDA can be competitive with Condat-Vu's PDA if the parameters are properly chosen. In particular, Condat-Vu's PDA with $\tau = \tau_3$ and $\sigma = \sigma_5$ performs the best, followed by E-GRPDA with $\tau = \tau_1$ and $\sigma = \sigma_1$, which is only slightly slower. A plausible explanation of the slightly inferior performance of E-GRPDA compared to Condat-Vu's PDA might be that in this example $L = \|A^\top A\|$ is quite large, which results in a relatively narrower convergence region and thus more restrictive choice of parameters, as shown in the middle plot in Figure 1. On the other hand, for the elastic net regularization problem, as shown in Figure 4, E-GRPDA with appropriate choice of step size parameters can be much more efficient than Condat-Vu's PDA. In this example, we have $h(x) = \mu_1\|x\|_2^2$ and thus $L = \sup_x \|\nabla^2 h(x)\| = \mu_1$, which is set to be 0.005 in this test. For small L, E-GRPDA allows more relaxed choice of parameters as shown in the right-hand-side plot in Figure 1. The reason for the above inconsistent relative performance remains unclear. We have also compared E-GRPDA, as well as Condat-Vu's PDA, with a version of the ADMM [7, 8]. Since ADMM needs to solve a linear system of equations at each iteration, it usually consumed more CPU time than E-GRPDA and Condat-Vu's PDA, which are full-splitting. For this reason, we have not included the comparison results with ADMM. Nonetheless, our experimental results given above definitely demonstrate the feasibility and the competitive performance of the proposed E-GRPDA.

6 Conclusions

In this paper, we have proposed and analyzed an extended golden ratio primal-dual algorithm for solving $\min_x h(x) + g(x) + f(Kx)$. The proposed algorithm is an extension of the golden ratio primal-dual algorithm recently proposed by Chang and Yang [3] by introducing an extra smooth term h whose gradient is Lipschitz continuous. In [3, 24], sublinear convergence rate results are established, which adopted the so-called primal-dual gap function as a measure of optimality. A major flaw of the primal-dual gap function is that it could vanish at nonstationary points, see, e.g., [2]. Thereby, convergence results based on the primal-dual gap function are not quite informative. In this paper, motivated by the recent work [16], we have carried out an analysis for the equivalent reformulated problem (1.4). In particular, we have shown in the general convex case that E-GRPDA converges at a sublinear rate $\mathcal{O}(1/N)$, where the optimality is measured by the function value residual and feasibility violation. If g or h is strongly convex, we have shown that E-GRPDA can be accelerated to achieve faster $\mathcal{O}(1/N^2)$ convergence rate. These results extend and complement those derived in [3, 24]. Our numerical results on CT image reconstruction problem and elastic net regularization problem also demonstrate the competitive performance of E-GRPDA. In particular, E-GRPDA performs favorably when compared to Condat-Vu's PDA.

Our proposed algorithm requires the spectral norm of K and the Lipschitz constant L, which could be difficult to evaluate or estimate in practice. In this case, E-GRPDA is hard to implement. A possible remedy is to incorporate certain adaptive technology, e.g., line search, into the algorithmic framework so that $\|K\|$ and L can be avoided and the whole algorithm can be largely accelerated. We leave this as future work.

References

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

[2] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *Journal of Mathematical Imaging and Vision* 40 (2011) 120–145.

[3] X. Chang and J. Yang, A golden ratio primal-dual algorithm for structured convex optimization, *Journal of Scientific Computing* 87 (2021) 1–26.

[4] P. Chen, J. Huang and X. Zhang, A primal-dual fixed-point algorithm for minimization of the sum of three convex separable functions. *Fixed Point Theory and Applications* 2016 (2016) 1–18.

[5] L. Condat, A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms, *Journal of Optimization Theory & Applications* 158 (2013) 460–479.

[6] 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 Journal on Imaging Sciences* 3 (2010) 1015–1046.

[7] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, *Computers and Mathematics with Applications* 2 (1976) 17–40.

[8] R. Glowinski and A. Marrocco, Sur l'approximation, par *é*l*é*ments finis d'ordre un, et la r*é*solution, par p*é*nalisation-dualit*é*, d'une classe de probl*è*mes de Dirichlet non lin*é*aires, *R.A.I.R.O.* 2 (1975) 41–76.

[9] B. He and X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective, *SIAM Journal on Imaging Sciences* 5 (2012) 119–149.

[10] Z. Hui and T. Hastie, Erratum: Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 67 (2015) 301–320.

[11] Y. Malitsky, Golden ratio algorithms for variational inequalities, *Mathematical Programming* 184 (2020) 383–410.

[12] Y. Malitsky and T. Pock, A first-order primal-dual algorithm with linesearch, *SIAM Journal on Optimization* 28 (2018) 411–432.

[13] A. Nedic and A. Ozdaglar, Subgradient methods for saddle-point problems, *Journal of Optimization Theory & Applications* 142 (2009) 205–228.

[14] Y. Nesterov, Smooth minimization of non-smooth functions, *Mathematical Programming* 103 (2005) 127–152.

[15] T. Pock and A. Chambolle, Diagonal preconditioning for first order primal-dual algorithms in convex optimization, in: *IEEE International Conference on Computer Vision*, 2011, pp.1762–1769.

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

[17] R. Shefi and M. Teboulle, Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization, *SIAM Journal on Optimization* 24 (2014) 269–297.

[18] E. Y. Sidky, J. H. Jrgensen and X. Pan, Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle-pock algorithm, *Physics in Medicine & Biology* 57 (2011) 3065–3091.

[19] R. Tibshirani, M. Saunders, S. Rosset, Z. Ji and K. Knight. Sparsity and smoothness via the fused lasso, *Journal of the Royal Statistical Society* 67 (2010) 91–108.

[20] H. Uzawa, Iterative methods for concave programming, in: *Studies in Linear* & *Nonlinear Programming*, 1959.

[21] B.C. Vu, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Advances in Computational Mathematics* 38 (2013) 667–681.

[22] M. Yan, A new primal-dual algorithm for minimizing the sum of three functions with a linear operator, *Journal of Scientific Computing* 76 (2016) 1698–1717.

[23] X. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on bregman iteration, *Journal of Scientific Computing* 46 (2011) 20–46.

[24] D. Zhou, X. Chang and J. Yang. A new golden ratio primal-dual algorithm, *Numerical Mathematics, A Journal of Chinese Universities*, To appear, 2021 (in Chinese).

[25] 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.

Manuscript received 23 June 2021
revised 2 August 2021
accepted for publication 27 September 2021

Danqing Zhou
Department of Mathematics, Nanjing University
Nanjing, P.R. China
E-mail address: zhoudanqing@smail.nju.edu.cn

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