Proceedings of Machine Learning Research vol 278:1–7, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

An Inexact Golden Ratio Primal-Dual Algorithm for a Saddle Point Problem

Jinxiu Liu* LIUJXTM@163.COM
Chongqing University of Posts and Telecommunications, Chongqing, 400000, China

Changjie Fang
Chongqing University of Posts and Telecommunications, Chongqing, 400000, China

Jingtao Qiu
Chongqing University of Posts and Telecommunications, Chongqing, 400000, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Convex optimization problems have wide applications in many fields such as mathematics, finance, industrial engineering, and management science. The primal dual algorithm (PDA), which is a classical approach for tackling a certain class of convex-concave saddle point problems, still has shortcomings such as fixed step size and difficulty in accurately solving certain subproblems. Therefore, designing more efficient inexact algorithms to solve these problems has important practical significance. During this investigation, we introduce an inexact golden ratio primal-dual algorithm based on the absolute error criteria of non-negative summable sequences. We establish the global convergence and the $O(1/N)$ rate of convergence for the proposed inexact algorithm, and the effectiveness of the proposed algorithm is verified by the image restoration experiment.

Keywords: Convex Optimization, Inexact Primal Dual Algorithm, Absolute Error, Golden Ratio, Image Restoration.

1. Introduction

Let X and Y be two real vector spaces of finite dimension, and take into account the following type of problem:

$$\min_{x \in X} \max_{y \in Y} L(x,y) = f(x) + \langle Ax, y \rangle - g(y), \tag{1}$$

f and g are both convex functions, X, Y are two non empty closed convex sets in R^n, $A : X \to Y$ represents a bounded linear operator that is endowed with a norm of $L = \|A\|$.

The pair $(\bar{x}, \bar{y}) \in X \times Y$ is called a saddle point of problem (1) if it satisfies the following inequalities:

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

The above inequality can be rewritten as

$$\begin{cases} P(x) := P_{\bar{x},\bar{y}}(x) = f(x) - f(\bar{x}) + \langle x - \bar{x}, A^* \bar{y} \rangle \geq 0, \forall x \in X, \\ D(y) := D_{\bar{x},\bar{y}}(y) = g(y) - g(\bar{y}) + \langle y - \bar{y}, -A\bar{x} \rangle \geq 0, \forall y \in Y. \end{cases} \tag{2}$$

Let $G(x,y) := L(x, \bar{y}) - L(\bar{x}, y)$ represents the dual gap, then

$$G(x,y) := G_{\bar{x},\bar{y}}(x,y) = P(x) + D(y) \geq 0, \forall (x,y) \in X \times Y.$$

The inequality system (2) can alternatively be restated as the subsequent KKT system.

$$0 \in \partial f(\bar{x}) + A^* \bar{y}, \quad 0 \in \partial g(\bar{y}) - A\bar{x}.$$

It is widely recognized that numerous application problems, including image restoration, magnetic resonance imaging, and computer vision, can be cast into the form of problem (1); for instances, refer to (Lyaqini et al., 2022; Park and Hentenryck, 2022; Kouri and Surowiec, 2022). In order to tackle model (1), the subsequent first - order primal - dual algorithm (PDA) (Chambolle and Pock, 2011) has garnered significant attention:

$$\begin{cases} x^{k+1} = \arg\min_{x \in X} \left\{ L\left(x, y^k\right) + \dfrac{1}{2\tau} \left\| x - x^k \right\|^2 \right\}, \\ \hat{x}^k = x^{k+1} + \theta\left(x^{k+1} - x^k\right), \\ y^{k+1} = \arg\max_{y \in Y} \left\{ L\left(\hat{x}^k, y\right) - \dfrac{1}{2\sigma} \left\| y - y^k \right\|^2 \right\}, \end{cases}$$

where $\tau > 0$, $\sigma > 0$, $\theta \in [0,1]$ is a weight parameter.

In 2020, Malitsky (2020) proposed a golden ratio algorithm with fully adaptive step size for variational inequalities. Inspired by Malitsky's convex combination technique with golden ratio, Chang and Yang (2021) proposed the following GRPDA algorithm in 2021:

$$\begin{cases} z^{k+1} = \dfrac{\phi - 1}{\phi} x^k + \dfrac{1}{\phi} z^k, \\ x^{k+1} = \text{Prox}_{\tau f}(z^{k+1} - \tau A^* y^k), \\ y^{k+1} = \text{Prox}_{\sigma g}(y^k + \sigma A x^{k+1}). \end{cases}$$

When the condition $\tau\sigma\|A\|^2 < \phi, \phi \in \left(1, \frac{1+\sqrt{5}}{2}\right]$ is met, the point sequence convergence of the algorithm can be obtained. It can be observed that this step size condition is much looser than the $\tau\sigma\|A\|^2 < 1$ in the PDA algorithm. Therefore, GRPDA allows for larger original and dual step sizes, which is crucial for fast practical convergence.

In 2020, Rasch and Chambolle (2020) introduced four types of approximation criteria to measure the effectiveness of solving sub problems with near non precision. They proposed an inexact primal dual algorithm, whose general steps are as follows:

$$\begin{cases} \hat{y} \approx_2^{\delta} \text{prox}_{\sigma g}(\bar{y} + \sigma K \tilde{x}), \\ \hat{x} \approx_i^{\varepsilon} \text{prox}_{\tau f}\left(\bar{x} - \tau\left(K^* \hat{y} + \nabla f(\bar{x}) + e\right)\right), \end{cases}$$

where $i = 1, 2$. In 2021, Jiang et al. (2021, 2020) carried out research on two kinds of imprecise primal - dual algorithms, adopting absolute error and relative error criteria separately. In 2024, Hien et al. (2023) proposed an inexact primal dual algorithm with absolute error criteria for solving non bilinear saddle point problems, but this algorithm has constraints in scenarios where the objective function exhibits strong convexity properties.

In order to further ease the conditions regarding the primal-dual step size and deal with the challenge of accurately resolving problems, spurred by the research in (Chang and Yang, 2021;

Jiang et al., 2020), we introduce an inexact golden ratio primal-dual algorithm based on the absolute error standard to solve Problem (1) in this research.

2. Preliminaries

Here, we recap key basic notions that are relevant for the upcoming sections.

Definition 2.1 Let h be a convex function on $X \subseteq R^n$, and $D \in R^{n\times n}$ be a given positive definite matrix. For any $D \succ 0$ and given $y \in R^n$, then

$$J_y(x) := h(x) + \frac{1}{2\tau}\|x-y\|_D^2, \quad \forall x \in R^n.$$

The proximal operator of h is

$$\mathrm{Prox}_{\tau h}^{D}(y) = \operatorname*{argmin}_{x\in X}\left\{h(x) + \frac{1}{2\tau}\|x-y\|_D^2\right\},$$

where $\|x\|_D^2 = \langle x, Dx\rangle$.

Assumption 2.1 (Jiang et al., 2020) There exists a mapping $\mathcal{F} : Y\times X\times\mathbb{R}^+\times X\times N \to X\times X$ such that for any $l \in N$, if $(x^l, d^l) = \mathcal{F}(y, \bar{x}, \tau, \tilde{x}, l)$, then

$$\lim_{l\to\infty} d_1^l = 0, \quad d_1^l \in \partial_x\left[f(x) + \langle Ax, y\rangle + \frac{1}{2\tau}\|x-\bar{x}\|^2\right]_{x=x^l}.$$

Assumption 2.2 (Jiang et al., 2020) There exists a mapping $\mathcal{G} : Y\times X\times\mathbb{R}^+\times Y\times N \to Y\times Y$ such that for any $l \in N$, if $(y^l, d^l) = \mathcal{G}(x, \bar{y}, \tau, \tilde{y}, l)$, then

$$\lim_{l\to\infty} d_2^l = 0, \quad d_2^l \in \partial_y\left[g(y) - \langle Ax, y\rangle + \frac{1}{2\tau}\|y-\bar{y}\|^2\right]_{y=y^l}.$$

Lemma 2.1 If assumption 2.1 holds, then sequence $\{x^l\}$ generated by $(x^l, d^l) = \mathcal{F}(y, \bar{x}, \tau, \tilde{x}, l)$ always converges to the unique solution of $f(x) + \langle Ax, y\rangle + \frac{1}{2\tau}\|x-\bar{x}\|^2$ in X.

Lemma 2.2 If assumption 2.2 holds, then sequence $\{y^l\}$generated by $(y^l, d^l) = \mathcal{G}(x, \bar{y}, \tau, \bar{y}, l)$ always converges to the unique solution of $g(y) - \langle Ax, y\rangle + \frac{1}{2\sigma}\|y-\bar{y}\|^2$ in Y.

Lemma 2.3 For any x, y, z, and given symmetric positive definite matrix D, we have

$$\langle D(x-y), x-z\rangle = \frac{1}{2}\left[\|x-y\|_D^2 + \|x-z\|_D^2 - \|y-z\|_D^2\right],$$
$$\|\alpha x + (1-\alpha)y\|_D^2 = \alpha\|x\|_D^2 + (1-\alpha)\|y\|_D^2 - \alpha(1-\alpha)\|x-y\|_D^2.$$

Lemma 2.4 Assuming that λ and Λ are the minimum and maximum eigenvalues of a symmetric positive definite matrix D, we have

$$\sqrt{\lambda}\|x\| \le \|x\|_D \le \sqrt{\Lambda}\|x\|.$$

3. Algorithm and convergence properties

In the following section, we introduce our inexact GRPDA incorporating absolute error criteria. Subsequently, we demonstrate its global convergence property and derive the $O(1/N)$ convergence rate analysis for the proposed algorithm.

3.1. Inexact GRPDA with absolute error criteria

Algorithm 1 Inexact GRPDA with absolute error criteria (Ia-GRPDA)

Step 1. Let $\varphi = \frac{\sqrt{5}+1}{2}$ be the golden ratio, that is $\varphi^2 = 1 + \varphi$. Choose $x^0 = z^0 \in \mathbb{R}^n$, $y^0 \in \mathbb{R}^m$, $\phi \in (1, \varphi)$, $\eta \in (0,1)$, $\mu \in (0,1)$, $\tau > 0$, $\beta > 0$. $S \in \mathbb{R}^{n\times n}$ and $T \in \mathbb{R}^{m\times m}$ are given symmetric positive definite matrix. Let $\psi = \frac{1+\phi}{\phi^2}$, $k = 0$.

Step 2. Let $l = 0$ and compute $z^{k+1} = \frac{\phi-1}{\phi}x^k + \frac{1}{\phi}z^k$ and

$$x^{k+1} \approx \underset{x\in X}{\operatorname{argmin}} \left\{ L\left(x, y^k\right) + \frac{1}{2\tau}\left\|x - z^{k+1}\right\|_S^2 \right\}.$$

2.a. Compute $(x^{k,l}, d_1^{k,l}) = \mathcal{F}(y^k, x^k, \tau, x^k, l)$.

2.b. If $\left\|d_1^{k,l}\right\| \le \frac{\epsilon_{k+1}}{\max\left\{\beta_1, \left\|x^{k,l}\right\|\right\}}$ then $x^{k+1} = x^{k,l}$, $d_1^k = d_1^{k,l}$. Otherwise, let $l = l+1$ and return 2.a.

Step 3. Let $l = 0$ and compute $y^{k+1} \approx \underset{y\in Y}{\operatorname{argmax}} \left\{ L\left(x^{k+1}, y\right) + \frac{1}{2\beta\tau}\left\|y - y^k\right\|_T^2 \right\}$.

3.a. Compute $\left(y^{k,l}, d_2^{k,l}\right) = \mathcal{G}\left(x^{k+1}, y^k, \beta\tau, y^k, l\right)$.

3.b. If $\left\|d_2^{k,l}\right\| \le \frac{\varsigma_{k+1}}{\max\left\{\beta_2, \left\|y^{k,l}\right\|\right\}}$ then $y^{k+1} = y^{k,l}$, $d_2^k = d_2^{k,l}$. Otherwise, let $l = l+1$ and return 3.a.

3.2. Convergence properties

Lemma 3.1 (Jiang et al., 2020) Under the validity of Assumptions 2.1 and 2.2, the inner iteration of Algorithm 3.1 will cease after a limited number of steps.

Lemma 3.2 Let $\lambda_1, \lambda_2 > v + 1$ are the minimum eigenvalues of matrices S and T, respectively. If assumptions 2.1 and 2.2 hold, $\left\{\left(z^{k+1}, x^{k+1}, \mathrm{y}^{k+1}\right) : \mathrm{k} \ge 0\right\}$ is a sequence generated by Algorithm 3.1, $(\bar{x}, \bar{y})$ is any saddle point of problem (1), then

$$\begin{aligned}
&\frac{\phi}{\phi-1}\left\|z^{k+3} - \bar{x}\right\|_S^2 + \frac{1}{\beta}\left\|y^{k+1} - \bar{y}\right\|_T^2 + 2\tau G\left(x^{k+1}, y^{k+1}\right) \\
&\le \frac{\phi}{\phi-1}\left\|z^{k+2} - \bar{x}\right\|_S^2 + \frac{1}{\beta}\left\|y^{k} - \bar{y}\right\|_T^2 + 2\tau\left(\frac{\tau\epsilon_{k+1}^2}{2\beta_1^2} + \epsilon_{k+2} + \|\bar{x}\|\frac{\epsilon_{k+2}}{\beta_1} + \varsigma_{k+1} + \|\bar{y}\|\frac{\zeta_{k+1}}{\beta_2}\right).
\end{aligned}$$

Theorem 3.1 If assumptions 2.1 and 2.2 hold, $\left\{\left(z^{k+1}, x^{k+1}, \mathrm{y}^{k+1}\right) : \mathrm{k} \ge 0\right\}$ is a sequence generated by Algorithm 3.1. So, $\left\{\left(x^{k+1}, \mathrm{y}^{k+1}\right) : \mathrm{k} \ge 0\right\}$ is bounded, with each of its limit points being a solution to Problem (1).

Theorem 3.2 For the following traversal sequence $\left\{X^N, Y^N\right\}$, where

$$X^N = \frac{1}{N}\sum_{k=0}^{N-1} x^k, Y^N = \frac{1}{N}\sum_{k=0}^{N-1} y^k$$

There are

$$G\left(X^N, Y^N\right) \leq \frac{c_0}{2N}, \quad c_0 = \frac{1}{\tau}\left(\frac{\phi}{\phi-1}\left\|z^2-\bar{x}\right\|_S^2 + \frac{1}{\beta}\left\|y^0-\bar{y}\right\|_T^2\right) + \frac{\tau}{\beta_1^2}\sum_{k=0}^{N-1}\epsilon_{k+1}^2 + \left(2+\frac{2\|\bar{x}\|}{\beta_1}\right)\sum_{k=0}^{N-1}\epsilon_{k+2} + \left(2+\frac{2\|\bar{y}\|}{\beta_2}\right)\sum_{k=0}^{N-1}\varsigma_{k+1}.$$

4. Numerical experiments

In this section, we focus on the following image restoration problem with total variation constraints:

$$\min_{y\in\mathcal{M}}\left\{\||By\||_1 + \frac{\varrho}{2}\|My-a\|^2\right\},$$

Where B is a discrete gradient operator, a is the observed image, M is the matrix representation of the blur operator.

We tested two images, "man. png (256×256)" and "peppersrgb. png (512×512)", respectively (see the first and third images in Figure 1). Each of the two images undergoes contamination by a blur operator of size 25×25, followed by degradation through zero-mean Gaussian noise with a standard deviation equivalent to 0.002. The contaminated images are listed in the second and fourth images of Fig.1, respectively. Then we compare Ia-GRPDA with PDA in (Chambolle and Pock, 2011), GRPDA in (Chang and Yang, 2021) and ICPA in (Rasch and Chambolle, 2020).

Figure 1: The original and degraded images.

Table 1: Numerical results of tested algorithms for the two images

	man.png				peppersrgb.png			
	Out-iter	In-iter	Time(s)	SNR	Out-iter	In-iter	Time(s)	SNR
PDA	56	1320	2.13	19.45	58	1173	10.06	17.32
GRPDA	43	845	1.76	19.56	51	822	8.73	17.32
ICPA	47	469	1.42	19.73	55	549	6.16	17.32
Ia-GRPDA	50	366	1.34	19.91	46	381	4.21	17.32

In Table 1, we list the required number of outer iterations (Out-iter), inner iterations (In-iter), computation time (Time (s)), and signal-to-noise ratio (SNR) for each testing algorithm in the restoration of two different images. Experimental data shows that the difference in the number of outer iterations between non exact algorithms and exact algorithms is relatively small, but Ia GRPDA has significantly fewer inner iterations and computation time than other testing algorithms, and the SNR is higher, indicating that the restored image quality is better than other algorithms. Through the comparison of the restored images presented in Figure 2, the advantages of the algorithm can also be more intuitively observed.

Figure 2: The restored images by PDA, GRPDA, ICPA, Ia-GRPDA from left to right.

5. Conclusion

This article mainly studies the inexact golden ratio primal dual algorithm based on the absolute error criterion of non negative summation sequences, introducing an extended proximal operator with matrix norm to approximate the subproblem. Further analysis was conducted on the convergence and convergence rate of the proposed algorithm, and the effectiveness of the inexact golden ratio primal dual algorithm was verified through numerical experiments on image restoration.

References

Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. doi: 10.1007/s10851-010-0251-1.

Xiaokai Chang and Junfeng Yang. A golden ratio primal–dual algorithm for structured convex optimization. *J. Sci. Comput.*, 87(2), May 2021. doi: 10.1007/s10915-021-01452-9.

Le Thi Khanh Hien, Renbo Zhao, and William B. Haskell. An inexact primal-dual smoothing framework for large-scale non-bilinear saddle point problems. *J. Optim. Theory Appl.*, 200(1): 34–67, December 2023. doi: 10.1007/s10957-023-02351-9.

Fan Jiang, Xingju Cai, Zhongming Wu, and Deren Han. Approximate first-order primal-dual algorithms for saddle point problems. *Mathematics of Computation*, page 1, 2020.

Fan Jiang, Zhongming Wu, Xingju Cai, and Hongchao Zhang. A first-order inexact primal-dual algorithm for a class of convex-concave saddle point problems. *Numerical Algorithms*, 88(3): 1109–1136, 2021. doi: 10.1007/s11075-021-01069-x.

Drew P. Kouri and Thomas M. Surowiec. A primal–dual algorithm for risk minimization. *Mathematical Programming*, 193(1):337–363, 2022. doi: 10.1007/s10107-020-01608-9.

S. Lyaqini, M. Nachaoui, and A. Hadri. An efficient primal-dual method for solving non-smooth machine learning problem. *Chaos, Solitons & Fractals*, 155:111754, 2022. doi: https://doi.org/10.1016/j.chaos.2021.111754.

Yura Malitsky. Golden ratio algorithms for variational inequalities. *Mathematical Programming*, 184(1):383–410, 2020. doi: 10.1007/s10107-019-01416-w.

Seonho Park and Pascal Van Hentenryck. Self-supervised primal-dual learning for constrained optimization, 2022.

Julian Rasch and Antonin Chambolle. Inexact first-order primal–dual algorithms. *Computational Optimization and Applications*, 76(2):381–430, 2020. doi: 10.1007/s10589-020-00186-y.