An Inexact Double Primal-Dual Algorithm for Saddle Point Problems

Changjie Fang\textsuperscript{1*}, Jingyu Chen\textsuperscript{2}

\textsuperscript{1} College of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.

\textsuperscript{2} College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.

*Corresponding author’s e-mail: fangcj@cqupt.edu.cn

Abstract. In this paper, we study a saddle point problem and propose an inexact double primal-dual algorithm for solving the saddle point problem. The sequence generated by the proposed algorithm is proved to globally converge to the saddle point of the saddle point problem in the ergodic sense, and the convergence rate of the algorithm is also obtained. For the model of total variation image processing, the performance of the algorithm is demonstrated. Simulation results show that our method is more efficient in image denoising compared with existing methods.

1. Introduction
Image restoration is an important problem in the field of communication, medical image analysis and other image processing. A blurred image brings great difficulties to image feature extraction and edge detection. Therefore, effective image processing has always been the focus of attention. In recent years, there are many algorithms for image restoration, such as Primal-Dual Hybrid Gradient (PDHG)\cite{9}, Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)\cite{1}, First-Order Primal-Dual Algorithm (FOPDA)\cite{3}, forward-backward splitting algorithm\cite{8,12}. These algorithms need to be solved accurately. However, in some cases, these algorithms can only be performed with a certain accuracy. For example, TV regularization inverse problem\cite{1,13} and matrix completion problem\cite{4,14}. In order to solve this problem, the proximal point methods\cite{6,15} and the proximal Forward-Backward splitting algorithms\cite{7} have been proposed, that is to say, the solution obtained is regarded as the solution of the problem under the condition of certain error. In this paper, we consider the following saddle point problem:

$$\min_{x \in X} \max_{y \in Y} L(x, y) := \theta_1(x) - \langle y, Ax \rangle - \theta_2(y)$$

(1)

where $A \in \mathbb{R}^n$ is a given matrix. $X$ and $Y$ are two nonempty, closed and convex subsets. $\theta_1(\cdot): \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, $\theta_2(\cdot): \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, lower semicontinuous convex functions.

Recently, Chambolle and Pock\cite{3} proposed a first-order primal-dual algorithm for solving (1), in which an inertial step for the primal (or dual) variable was treated. Further, Rasch and Chambolle\cite{5} reported the inexact first-order primal-dual algorithms for saddle point problems and investigated the convergence rates for the proposed methods. As both primal and dual subproblems usually have their
own good properties, Wang and He\cite{2} introduced a double inertial primal-dual algorithm for solving (1), in which the inertial effects were incorporate into both primal and dual variables.

Motivated by the research works mentioned above, in this paper, we present an inexact double primal-dual algorithm for solving (1), in which the type-2 and type-1 approximations of the proximal point for primal and dual variables are executed, respectively. At the same time, the inertial steps for both primal and dual variables are considered. Moreover, we obtain the convergence result for the proposed method in the ergodic sense. In addition, we also analyze the convergence rate of our method. Finally, the numerical results illustrate the efficiency of our method.

Our method is described as follows:

**Algorithm 1** An Inexact Double Primal-Dual Algorithm

Step 1: Select $r > 0$, $s > 0$ and $r > \|A\|^2 + 1, s > \|A\|^2 + 1$.

Step 2: Choose any $x^0, y^0, x^{-1}$ and $y^{-1}$.

Step 3: Updates $(x^{k+1}, y^{k+1})$ via

\[
\begin{align*}
    x^{k+1} &= \frac{1}{2} \arg \min_{x \in \mathbb{X}} \left\{ L(x, y^k) + \frac{r}{2} \|x - x^k\|^2 \right\} \quad (1) \\
    \tilde{x}^{k+1} &= 2x^k - x^{k-1} \quad (2) \\
    y^{k+1} &= \frac{1}{2} \arg \max_{y \in \mathbb{Y}} \left\{ L(\tilde{x}^{k+1}, y) - \frac{s}{2} \|y - y^k\|^2 \right\} \quad (3) \\
    \tilde{y}^{k+1} &= 2y^{k+1} - y^k \quad (4)
\end{align*}
\]

until meeting stopping criterion.

2. Preliminaries

Suppose $h(x): \mathbb{X} \rightarrow [0, \infty]$ mapping from a Hilbert space $\mathbb{X}$ to the extended real line $\mathbb{R} = \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function. And $y \in \mathbb{X}$ the proximal point, denote

\[
G_y(x) := \arg \min_{x \in \mathbb{X}} \left\{ h(x) + \frac{1}{2\tau} \|x - y\|^2 \right\}
\]

and define

\[
\text{prox}_{\tau h}(y) := \arg \min_{x \in \mathbb{X}} G_y(x)
\]

The following definitions can be found in [5].

**Definition 2.1** The notion of an $\varepsilon$ - subdifferential of $h(\cdot): \mathbb{X} \rightarrow \mathbb{R}$ at $z \in \mathbb{X}$:

\[
\partial_\varepsilon h(z) = \left\{ p \in \mathbb{X} \mid h(x) \geq h(z) + \langle p, x - z \rangle - \varepsilon, \forall x \in \mathbb{X} \right\}
\]

**Definition 2.2** Let $\varepsilon \geq 0$ and $z \in \mathbb{X}$. For minimization problem

\[
z \approx^\varepsilon_1 \text{prox}_{\tau h}(y) \iff 0 \in \partial_\varepsilon G_y(z)
\]

**Definition 2.3** Let $\varepsilon \geq 0$ and $z \in \mathbb{X}$. For minimization problem

\[
z \approx^\varepsilon_2 \text{prox}_{\tau h}(y) \iff \frac{1}{\tau}(y - z) \in \partial_\varepsilon h(z)
\]
Lemma 2.1 [5] Suppose \( z \in \mathbb{X} \) and \( z \approx \arg \min_{x \in \mathbb{X}} \{ h(x) + \frac{1}{2\tau} \| x - y \|^2 \} \), then there exists \( q \in \mathbb{X} \) with \( \| q \| \leq \sqrt{2\tau}r \), such that
\[
\frac{y - z - q}{\tau} \in \partial_{x}h(z)
\]
(10)

3. Convergence Analysis

Lemma 3.1 Let \( L = \| A \|, \sigma > 0 \) and \( \beta > 0 \). Suppose \( (x^{k+1}, y^{k+1}) \) is generated by Algorithm 1. For each \( (x, y) \in \mathbb{X} \times \mathbb{Y} \), we have

\[
L(x^{k+1}, y) - L(x, y^{k+1}) \leq \frac{r}{2} \| x - x^k \|^2 - \frac{r}{2} \| x - x^{k+1} \|^2 - \frac{r - \beta}{2} \| x^{k+1} - x^k \|^2 + \frac{L^2}{2\tau} \| x - x^{k+1} \|^2
\]
\[
+ \frac{s}{2} \| y - y^k \|^2 - \frac{s}{2} \| y - y^{k+1} \|^2 - \frac{s - \sigma}{2} \| y^{k+1} - y \|^2 + \frac{L^2}{2\beta} \| y^{k+1} - y \|^2 + \langle y^{k+1} - y, A(x - x^{k+1}) \rangle
\]
\[
- \langle y^{k+1} - y^k, (x - x^{k+1}) \rangle + \langle y^{k+1} - y^k, A(x^{k+1} - x^k) \rangle - \langle y^k - y, A(x^{k+1} - x^k) \rangle
\]
\[
+ \sqrt{2s\epsilon_k} \| y - y^{k+1} \| + \epsilon_{k+1} + \delta_{k+1}
\]
(11)

Proof. According to (1), we have
\[
x^{k+1} = \delta_{k+1} \arg \min_{x \in \mathbb{X}} \left\{ \theta_1(x) + \frac{r}{2} \| x - x^k - \frac{1}{r} A^T y^k \|^2 \right\}
\]
(12)

From (7), the definition of \( \varepsilon \)-subdifferential, we have
\[
\theta_1(x^{k+1}) - \theta_1(x) - \langle y^{k+1} - y, A(x - x^{k+1}) \rangle \leq \frac{r}{2} \| x - x^k \|^2 - \frac{r}{2} \| x^{k+1} - x^k \|^2 - \frac{r - \beta}{2} \| x^{k+1} - x^k \|^2
\]
\[
+ \langle y^{k+1} - y, (x^{k+1} - x^k) \rangle + \langle y^{k+1} - y^k, A(x^{k+1} - x^k) \rangle + \delta_{k+1}
\]
(13)

According to (3), we have
\[
y^{k+1} = \epsilon_{k+1} \arg \min_{y \in \mathbb{Y}} \left\{ \theta_2(y) + \frac{s}{2} \| y - y^k - \frac{1}{s} A^T x^{k+1} \|^2 \right\}
\]
(14)

From (7), the definition of \( \varepsilon \)-subdifferential, we have
\[
\theta_2(y^{k+1}) - \theta_2(y) - \langle y^{k+1} - y, A(x^{k+1} - x^k) \rangle \leq \frac{s}{2} \| y - y^k \|^2 - \frac{s}{2} \| y^{k+1} - y^k \|^2 - \frac{s}{2} \| y - y^{k+1} \|^2
\]
\[
+ \langle y^{k+1} - y, (x^{k+1} - x^k) \rangle - \langle y^k - y, A(x^k - x^{k+1}) \rangle - \langle y^k - y^{k+1}, A(x^{k+1} - x^k) \rangle
\]
\[
+ \sqrt{2s\epsilon_k} \| y - y^{k+1} \| + \epsilon_{k+1}
\]
(15)

Add (13) and (15), we have
L(x^{k+1}, y) - L(x, y^{k+1}) \leq \frac{r}{2} \left\| x - x^0 \right\|^2 - \frac{r-\beta}{2} \left\| x^{k+1} - x^k \right\|^2 - \frac{r-\beta}{2} \left\| x^k - x^1 \right\|^2 + \frac{L^2}{2\sigma} \left\| x^k - x^{k-1} \right\|^2
+ \frac{s}{2} \left( y - y^0 \right)^2 - \frac{s-\sigma}{2} \left( y^{k+1} - y^k \right)^2 + \frac{L^2}{2\beta} \left( y^k - y^{k-1} \right)^2 + \frac{r(s-\sigma)}{2(s-\sigma)} \left\| x - x^N \right\|^2
- \frac{r-\beta}{2} \left\| x^{k+1} - x^k \right\|^2 - \frac{r(s-\sigma)}{2(s-\sigma)} \sum_{k=1}^{N-1} \left\| x^k - x^{k-1} \right\|^2 - \frac{r-\beta}{2} \sum_{k=1}^{N-1} \left\| y^k - y^{k-1} \right\|^2
+ \sqrt{2s}\varepsilon + \delta + \sum_{k=1}^{N-1} (\varepsilon + \delta).

\text{Lemma 3.2} \text{ Let } L = \left\| A \right\|, \sigma > 0, s > L^2 + 1 \text{ and } r > \frac{L^2}{2}. \text{ For any } N \geq 1, \text{ there is a saddle point } (x^*, y^*) \in \mathbb{X} \times \mathbb{Y}, \text{ with}

L(X^N, y^*) - L(x^*, Y^N) \leq \frac{r}{2N} \left( \left\| x^0 - x^* \right\| + \sqrt{s} \left\| y^0 - y^* \right\| + \frac{2\beta}{\sqrt{tr}} A_N + \sqrt{2s}\varepsilon \right)^2.

\text{Proof.} \text{ Let } x^0 = x^1 \text{ and } y^0 = y^1, \text{ sum (11) from } k = 0, 1, \ldots, N-1 \text{ and using the Young's inequality to obtain}

\sum_{k=1}^{N} L(x^{k+1}, y) - L(x, y^{k+1}) \leq \frac{r}{2} \left\| x - x^0 \right\|^2 + \frac{s}{2} \left\| y - y^0 \right\|^2 - \frac{r(s-\sigma)}{2(s-\sigma)} \left\| x - x^N \right\|^2
- \frac{r-\beta}{2} \left\| x^{k+1} - x^k \right\|^2 - \frac{s-\sigma}{2} \sum_{k=1}^{N-1} \left\| x^k - x^{k-1} \right\|^2 - \frac{r-\beta}{2} \sum_{k=1}^{N-1} \left\| y^k - y^{k-1} \right\|^2
+ \sqrt{2s}\varepsilon + \delta + \sum_{k=1}^{N-1} (\varepsilon + \delta).

In (17), we let \sigma = s - L^2 \text{ and } \beta = r - L^2, \text{ we have}

\sum_{k=1}^{N} L(x^{k+1}, y) - L(x, y^{k+1}) \leq \frac{r}{2} \left\| x - x^0 \right\|^2 + \frac{s}{2} \left\| y - y^0 \right\|^2 - \frac{r-1}{2} \left\| x - x^N \right\|^2
- \frac{s-1}{2} \left\| y - y^0 \right\|^2 - \frac{L^2}{2(s-L^2)} \sum_{k=1}^{N-1} \left\| x^k - x^{k-1} \right\|^2 - \frac{L^2}{2(s-L^2)} \sum_{k=1}^{N-1} \left\| y^k - y^{k-1} \right\|^2
+ \sqrt{2s}\varepsilon + \delta + \sum_{k=1}^{N-1} (\varepsilon + \delta).

In (18), we let \( x, y = (x^*, y^*) \) and similar to the proof way in [5], we have

L(X^N, y^*) - L(x^*, Y^N) \leq \frac{r}{2N} \left( \left\| x^0 - x^* \right\| + \sqrt{s} \left\| y^0 - y^* \right\| + \frac{2\beta}{\sqrt{tr}} A_N + \sqrt{2s}\varepsilon \right)^2.
**Theorem 3.1** Suppose that sequence \((x^n, y^n)\) is generated by Algorithm 1, \((X^N, Y^N)\) is defined by Theorem 3.2 with \(A_N\) and \(B_N\) are additive. Then every weak accumulation point \((\hat{x}^*, \hat{y}^*)\) of \((X^N, Y^N)\) is the saddle point of problem (1). If \(X\) and \(Y\) are finite sets, then there is a saddle point \(\hat{x}^*, \hat{y}^* \in X \times Y\) such that \(x^n \rightarrow \hat{x}^*, \ y^n \rightarrow \hat{y}^*\).

**Proof.** Similar to the proof of Theorem 3.8 in [5].

**Lemma 3.3** If
\[
\varepsilon_n = O\left(\frac{1}{n^{\frac{\alpha+1}{2}}}\right), \quad \delta_n = O\left(\frac{1}{n^{\frac{\alpha+1}{2}}}\right),
\]
then
\[
L(X^N, y^*) - L(x^*, Y^N) = O\left(\frac{\ln^2(N)}{N}\right) \quad \alpha = \frac{1}{2}
\]
\[
O\left(\frac{1}{N}\right) \quad \alpha > \frac{1}{2}
\]
\[
O\left(N^{-2\alpha}\right) \quad 0 < \alpha < \frac{1}{2}
\]

**Proof.** Similar to the proof of Corollary 3.6 in [5].

4. **Numerical experiment**

In this section, we will study the performance of Algorithm 1 in image processing, so we will compare four algorithms, FOPDA[3], GPDHG[10], DEPDA[2] and ADMM[11], respectively.

We consider the following TV-L1 model
\[
\min_{u \in X} F(u) = \|Au - f\|_1 + \lambda \|Du\|_1
\]

In the experiments, we add the zero-mean Gaussian noise with the standard deviation 0.001 and severe motion blur operator to the image by MATLAB scripts fspecial and imnoise.

Figure 1. From left to right: the original images, the degraded images (noise standard deviation 0.001) and the restored images by Algorithm 1.

From Figure 1, we can see that the quality of the image restored by our algorithm is relatively good. Our algorithm compares the number of iterations and the SNR with the other four algorithms. The comparison results are shown in Table 1.

| Image  | Method | Tol=10^{-3} | Tol=10^{-4} | Tol=10^{-5} |
|--------|--------|-------------|-------------|-------------|
| Hill   | FOPDA  | 31/22.24    | 101/22.88   | 183/22.90   |

Table 1. The results of image restoration with different Tols.
5. Conclusion
In this paper, we propose an inexact double primal-dual algorithm. We prove the convergence of our algorithm. From the experiment, the image restored by our algorithm is relatively good.

Acknowledgments
This work is partially supported by Basic and Advanced Research Project of CQ CSTC(Nos.cstc2020jcyj-msxmX0738 and cstc2018jcyjAX0605).

References
[1] Beck A., Teboulle M. (2009) Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems. IEEE Transactions on Image Processing, 18(11):2419-2434.
[2] Wang K., He H. (2020) A Double Extrapolation Primal-Dual Algorithm for Saddle Point Problems. Journal of Scientific Computing, 85(2):1-30.
[3] Chambolle A., Pock T. (2011) A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. Journal of Mathematical Imaging and Vision, 40(1):120-145.
[4] Ma S., Goldfarb D., Chen L. (2011) Fixed point and Bregman iterative methods for matrix rank minimization. Mathematical Programming, 128(1-2):321-353.
[5] Rasch J., Chambolle A. (2020) Inexact First-Order Primal-Dual Algorithms. Computational Optimization and Applications, 76:381–430.
[6] Rockafellar R.T. (1976) Monotone operators and the proximal point algorithm. SIAM Journal on Control and Optimization, 14(5):877-898.
[7] Combettes P.L., Wajs, Valérie R. (2006) Signal recovery by proximal forward-backward splitting. Multiscale Model Simul., 4(4):1168-1200.
[8] P.L. Lions, B. Mercier. (1979) Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis, 16(6):964–979.
[9] M. Zhu, T. Chan. (2008) An efficient primal-dual hybrid gradient algorithm for total variation image restoration, CAM Reports 08-34, UCLA, Center for Applied Math.
[10] He B., Ma F., Yuan X. (2017) An Algorithmic Framework of Generalized Primal-Dual Hybrid Gradient Methods for Saddle Point Problems. Journal of Mathematical Imaging and Vision, 58(2):279–293.
[11] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein. (2010) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations & Trends in Machine Learning, 3:1-122.
[12] Combettes P.L., Pesquet J.C. (2011) Proximal Splitting Methods in Signal Processing. Heinz H Bauschke, 49:185-212.
[13] Peyré, Gabriel, Fadili J. (2011) Total variation projection with first order schemes. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 20(3):657-69.
[14] Cai J, Candes E J, Shen Z. (2010) A Singular Value Thresholding Algorithm for Matrix Completion. SIAM Journal on Optimization, 20(4):1956–1982.
[15] Güler, Osman. (1992) New Proximal Point Algorithms for Convex Minimization. SIAM Journal on Optimization, 2(4):649-664.