

Convergence Analysis of Primal-Dual Algorithms for a Saddle-Point Problem: From Contraction Perspective*

Bingsheng He[†] and Xiaoming Yuan[‡]

Abstract. Recently, some primal-dual algorithms have been proposed for solving a saddle-point problem, with particular applications in the area of total variation image restoration. This paper focuses on the convergence analysis of these primal-dual algorithms and shows that their involved parameters (including step sizes) can be significantly enlarged if some simple correction steps are supplemented. Some new primal-dual-based methods are thus proposed for solving the saddle-point problem. We show that these new methods are of the contraction type: the iterative sequences generated by these new methods are contractive with respect to the solution set of the saddle-point problem. The global convergence of these new methods thus can be obtained within the analytic framework of contraction-type methods. The novel study on these primal-dual algorithms from the perspective of contraction methods substantially simplifies existing convergence analysis. Finally, we show the efficiency of the new methods numerically.

Key words. saddle-point problem, total variation, image restoration, primal-dual method, contraction method, proximal point algorithm

AMS subject classifications. 68U10, 90C25, 65K10, 65J22

DOI. 10.1137/100814494

1. Introduction. We consider the saddle-point problem

$$(1.1) \quad \min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \Phi(x, y) := y^T Ax + \frac{\lambda}{2} \|By - z\|^2,$$

where $\mathcal{X} \subset \mathbb{R}^L$ and $\mathcal{Y} \subset \mathbb{R}^N$ are convex sets, $z \in \mathbb{R}^N$, $A \in \mathbb{R}^{N \times L}$, $B \in \mathbb{R}^{N \times N}$, $\lambda > 0$, $\|\cdot\|$ denotes the Euclidean norm, and T denotes the standard inner product operator. In particular, the model (1.1) captures some image restoration problems involving the total variation (TV) regularization introduced in [23]; see, e.g., [6, 11, 25]. Note that we can consider a saddle-point problem in more general settings such as the model considered in [6],

$$(1.2) \quad \min_{y \in Y} \max_{x \in X} \Phi(x, y) := g(y) + y^T Kx - f^*(x),$$

where X and Y are two finite-dimensional real vector spaces equipped with an inner product and the reduced norm $\|\cdot\|$; $g : Y \rightarrow [0, +\infty)$ and $f^* : X \rightarrow [0, +\infty)$ are proper, convex, lower-

*Received by the editors November 10, 2010; accepted for publication (in revised form) November 1, 2011; published electronically January 24, 2012.

<http://www.siam.org/journals/siims/5-1/81449.html>

[†]Department of Mathematics and National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China (hebma@nju.edu.cn). This author's work was supported by National Natural Science Foundation of China (grants 10971095 and 91130007), the Fund of MOE China (20110091110004), and the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (708044).

[‡]Corresponding author. Department of Mathematics, Hong Kong Baptist University, Hong Kong, China (xmyuan@hkbu.edu.hk). This author's work was supported by the Hong Kong General Research Fund: 203311.

semicontinuous (l.s.c.) functions; f^* is itself the convex conjugate of a convex l.s.c. function f ; and $K : Y \rightarrow X$ is a continuous linear operator with the induced norm

$$\|K\| = \max\{\|Ky\| : y \in Y \text{ with } \|y\| \leq 1\}.$$

For clearer exposition of our analysis and simpler presentation, we focus on (1.1) in the following analysis, and we briefly show the extension to (1.2) in section 6.

As analyzed in [6, 9, 11], the saddle-point problem (1.1) can be regarded as the primal-dual formulation of a nonlinear programming problem, and this fact has inspired some primal-dual algorithms for TV image restoration problems. We refer to, e.g., [6, 7, 9, 11, 15, 20, 24, 25] for their numerical efficiency. More specifically, the iterative schemes of existing primal-dual algorithms for (1.1) can be unified as the following procedure.

The primal-dual procedure for (1.1).

Let $\tau > 0$, $\sigma > 0$, and $\theta \in \mathfrak{R}$. With the given (x^k, y^k) , the new iterate (x^{k+1}, y^{k+1}) is generated by

$$(1.3a) \quad x^{k+1} = \text{Arg max}_{x \in \mathcal{X}} \left\{ \tau \Phi(x, y^k) - \frac{1}{2} \|x - x^k\|^2 \right\},$$

$$(1.3b) \quad \bar{x}^k = x^{k+1} + \theta(x^{k+1} - x^k),$$

$$(1.3c) \quad y^{k+1} = \text{Arg min}_{y \in \mathcal{Y}} \left\{ \sigma \Phi(\bar{x}^k, y) + \frac{1}{2} \|y - y^k\|^2 \right\}.$$

In (1.3), the parameters τ and σ are step sizes of the primal and dual steps, respectively, and θ is called the combination parameter for obvious reasons. With some specific choices of these parameters, some existing primal-dual algorithms for (1.1) are recovered, and their convergence can be guaranteed when certain requirements are imposed on these parameters. Below are some examples.

- When $\theta = 0$ in (1.3b), the primal-dual procedure (1.3) reduces to the Arrow–Hurwicz algorithm [1], which has been highlighted in [25] for TV image restoration problems. In [11], the convergence of the special case of (1.3) with $\theta = 0$ has been studied insightfully by imposing additional restrictions ensuring that the step sizes τ and σ are small.
- When $\theta \in [0, 1]$ in (1.3b), the primal-dual algorithm proposed in [6] is recovered. In [6], it was shown that the primal-dual procedure (1.3) is closely related to many existing methods including the extrapolational gradient method [18, 21], the Douglas–Rachford splitting method [10], and the alternating direction method of multipliers [13]. In particular, when $\theta = 1$, the convergence of (1.3) was proved in [6] with the requirement on step sizes

$$(1.4) \quad \tau\sigma < \frac{1}{\|A^T A\|}.$$

Thus, the difficulty of choosing very small step sizes in [11] is overcome. Note that under some additional regularity and convexity assumptions on (1.2), some sophisticated

strategies for choosing the parameters (τ, σ, θ) dynamically have also been analyzed in [6] in order to yield accelerated primal-dual algorithms. In this paper, we restrict our discussion to the deterministic scenario where all the parameters (τ, σ, θ) are fixed throughout iterations.

In this paper, we study primal-dual algorithms for (1.1) from the perspective of contraction methods (see [5] or section 2.4 for the definition of a contraction method). More specifically, we will show that the primal-dual procedure (1.3) with $\theta = 1$ is essentially an application of the proximal point algorithm (PPA) (see [19]), provided that certain conditions on τ and σ are required, and thus the resulting sequence of (1.3) is contractive with respect to the solution set of (1.1). On the other hand, for the case $\theta \neq 1$, (1.3) still takes an iterative framework analogous to that of PPA although it cannot be interpreted as an application of PPA any longer. However, this PPA-like structure makes it possible to develop some simple steps to correct the output of (1.3) at each iteration. Consequently, the sequence generated by the combination of (1.3) and such a correction step is contractive with respect to the solution set of (1.1). Accordingly, some primal-dual-based contraction algorithms with various correction schemes are presented for (1.1). These new algorithms are in the prediction-correction fashion, where the primal-dual procedure (1.3) produces a predictor and it is corrected by a certain correction step at each iteration. Our contributions in this paper can be summarized as follows:

- (1) We show that the range of the combination parameter θ can be enlarged to $[-1, 1]$ in (1.3), which is broader than the result $\theta \in [0, 1]$ in [6].
- (2) When $\theta = -1$, the step size τ and σ can be arbitrary positive numbers. When $\theta \in (-1, 1]$, the condition (1.4) on step sizes can be relaxed to

$$(1.5) \quad \tau\sigma \frac{(1+\theta)^2}{4} < \frac{1}{\|A^T A\|}.$$

Note that the condition (1.4) is recovered when $\theta = 1$ in (1.5). However, when θ is close to -1 , the step sizes τ and σ can be very large simultaneously.

- (3) We show that the analytic framework of contraction methods is a novel tool for studying the convergence of primal-dual algorithms for (1.1). With this novel analytic framework, the convergence analysis of existing primal-dual algorithms can be simplified substantially.
- (4) We propose some efficient numerical algorithms for (1.1) by blending the ideas of primal-dual and contraction methods.

The rest of this paper is organized as follows. In section 2, we review some preliminaries which are useful for our analysis. In section 3, for the case $\theta \in [-1, 1)$, we present a primal-dual-based contraction method and prove its convergence under the analytic framework of contraction-type methods. The main theoretical results of enlarging the parameters in (1.3) are also presented in this section. In section 4, we present a reduced primal-dual-based contraction method with simpler correction steps for the case $\theta \in [-1, 1)$. The corresponding requirement on the step sizes of (1.3) is also analyzed. In section 5, we pay specific attention to the case of $\theta = 1$. We will develop two new primal-dual-based contraction methods for this special case. Then, we briefly analyze the extension of our convergence analysis for (1.1) to the more general case (1.2) in section 6. In section 7, we illustrate the benefits of enlarging

the involved parameters in (1.3) and the efficiency of the new methods numerically by some experiment results. Finally, some conclusions are drawn in section 8.

2. Preliminaries. In this section, we provide some preliminaries which are useful for our analysis. In particular, we review some basic knowledge of variational inequalities (VIs), the PPA, and contraction methods, which are cornerstones of later analysis.

2.1. The VI reformulation of (1.1). We first reformulate (1.1) into a VI; for more details the reader is referred to [25]. Let (x^*, y^*) be a solution of the saddle-point problem (1.1). Then, we have

$$\max_{x \in \mathcal{X}} \left\{ (y^*)^T Ax + \frac{\lambda}{2} \|By^* - z\|^2 \right\} \leq (y^*)^T Ax^* + \frac{\lambda}{2} \|By^* - z\|^2 \leq \min_{y \in \mathcal{Y}} \left\{ y^T Ax^* + \frac{\lambda}{2} \|By - z\|^2 \right\}.$$

Based on their optimality conditions, we can easily verify that the above problems can be characterized by the following VIs:

$$\begin{cases} x^* \in \mathcal{X}, & (x - x^*)^T (-A^T y^*) \geq 0 & \forall x \in \mathcal{X}, \\ y^* \in \mathcal{Y}, & (y - y^*)^T (Ax^* + \lambda B^T (By^* - z)) \geq 0 & \forall y \in \mathcal{Y}. \end{cases}$$

Therefore, the saddle-point problem (1.1) can be characterized by the following compact VI: Find $u^* \in \Omega$ such that

$$(2.1a) \quad \text{VI}(\Omega, F) : \quad (u - u^*)^T F(u^*) \geq 0 \quad \forall u \in \Omega,$$

where

$$(2.1b) \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(u) = \begin{pmatrix} -A^T y \\ Ax + \lambda B^T (By - z) \end{pmatrix}, \quad \text{and} \quad \Omega := \mathcal{X} \times \mathcal{Y}.$$

It is easy to verify that the mapping $F(u)$ in (2.1b) is monotone with respect to Ω , i.e.,

$$(u - v)^T (F(u) - F(v)) \geq 0 \quad \forall u, v \in \Omega.$$

Therefore, $\text{VI}(\Omega, F)$ is monotone and its solution set denoted by Ω^* is nonempty (see, e.g., [12]).

Let the projection onto Ω under the Euclidean norm be denoted by $P_\Omega(\cdot)$, i.e.,

$$P_\Omega(v) = \text{Argmin}\{\|v - u\|^2 \mid u \in \Omega\}.$$

Then, the following lemma shows that solving $\text{VI}(\Omega, F)$ amounts to solving a projection equation.

Lemma 2.1. *The point u^* is a solution of $\text{VI}(\Omega, F)$ if and only if*

$$u^* = P_\Omega[u^* - \alpha F(u^*)] \quad \forall \alpha > 0.$$

Proof. See [4, p. 267]. ■

2.2. Proximal point algorithm. Among classical methods applicable for solving $\text{VI}(\Omega, F)$ is the PPA, which was contributed originally in [19] and developed concretely in [22]. More specifically, instead of solving $\text{VI}(\Omega, F)$ directly, PPA generates the sequence $\{u^k\}$ iteratively via solving the proximal subproblem

$$(2.2) \quad (\text{PPA}) \quad u^{k+1} \in \Omega, \quad (u - u^{k+1})^T (F(u^{k+1}) + r(u^{k+1} - u^k)) \geq 0 \quad \forall u \in \Omega,$$

with $r > 0$ being the proximal parameter. Let $G \in \Re^{(L+N) \times (L+N)}$ be a symmetric positive definite matrix. Then, the PPA (2.2) can be generalized into

$$(2.3) \quad u^{k+1} \in \Omega, \quad (u - u^{k+1})^T (F(u^{k+1}) + G(u^{k+1} - u^k)) \geq 0 \quad \forall u \in \Omega,$$

where G is a metric proximal parameter. Note that (2.3) can be regarded as the PPA with a preconditioning proximal term (see, e.g., [11]) or the PPA in the context of the G -norm (defined by $\|u\|_G = \sqrt{u^T G u}$) [11, 17]. Based on the convergence analysis of PPA (see, e.g., [19, 22]), the sequence $\{u^k\}$ generated by (2.2) or (2.3) converges to a solution of $\text{VI}(\Omega, F)$.

2.3. The PPA structure of (1.3). In this subsection, we show that the primal-dual procedure (1.3) actually takes a structure analogous to that of the PPA (2.3), but the involved preconditioning matrix G is not symmetric if $\theta \neq 1$.

Lemma 2.2. *Let Ω and F be defined in (2.1b), and let $u^{k+1} = (x^{k+1}, y^{k+1})$ be generated by the primal-dual procedure (1.3). Then we have*

$$(2.4) \quad u^{k+1} \in \Omega, \quad (u - u^{k+1})^T \{F(u^{k+1}) + M(u^{k+1} - u^k)\} \geq 0 \quad \forall u \in \Omega,$$

where

$$(2.5) \quad M = \begin{pmatrix} \frac{1}{\tau} I & A^T \\ \theta A & \frac{1}{\sigma} I \end{pmatrix}.$$

Proof. It follows from the optimality conditions of (1.3a) and (1.3c) that

$$(2.6) \quad x^{k+1} \in \mathcal{X}, \quad (x - x^{k+1})^T \left\{ (-A^T y^k) + \frac{1}{\tau} (x^{k+1} - x^k) \right\} \geq 0 \quad \forall x \in \mathcal{X}$$

and

$$(2.7) \quad y^{k+1} \in \mathcal{Y}, \quad (y - y^{k+1})^T \left\{ [A\bar{x}^k + \lambda B^T (B y^{k+1} - z)] + \frac{1}{\sigma} (y^{k+1} - y^k) \right\} \geq 0 \quad \forall y \in \mathcal{Y}.$$

Combining (2.6) and (2.7), we get that $(x^{k+1}, y^{k+1}) \in \Omega$ and that

$$\begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} -A^T y^{k+1} \\ A x^{k+1} + \lambda B^T (B y^{k+1} - z) \end{pmatrix} + \left[\begin{pmatrix} \frac{1}{\tau} I & A^T \\ \theta A & \frac{1}{\sigma} I \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right] \right\} \geq 0$$

for any $(x, y) \in \Omega$. By using the function F defined in (2.1b) and the matrix M defined in (2.5), (2.4) is a compact form of the above VI and the lemma is proved. ■

Note that the matrix M defined in (2.5) is not symmetric except for $\theta = 1$. Thus, the primal-dual procedure (1.3) with $\theta \neq 1$ is not an application of the PPA (2.3) even though it takes a structure similar to that of PPA. In this sense, the iterative scheme (2.4) can be regarded as a variant of PPA or the PPA with a linear proximal term, as introduced in [16].

2.4. Contraction methods. We first recall the concept of the Fejér monotonicity for $\text{VI}(\Omega, F)$. For a sequence $\{u^k\}$, if the property

$$\|u^{k+1} - u^*\| \leq \|u^k - u^*\| \quad \forall u^* \in \Omega^*$$

is satisfied, then the sequence $\{u^k\}$ is said to be Fejér monotone with respect to Ω^* . We refer the reader to [2] for more properties of the Fejér monotonicity.

According to [5], if a method generates an iterative sequence $\{u^k\}$ satisfying

$$(2.8) \quad \|u^{k+1} - u^*\| \leq \|u^k - u^*\| - c\|u^k - u^{k+1}\| \quad \forall u^* \in \Omega^*,$$

where $c > 0$ is a constant, then we call this method a contraction method for $\text{VI}(\Omega, F)$. Obviously, the sequence generated by a contraction method for $\text{VI}(\Omega, F)$ is Fejér monotone with respect to Ω^* . Let $G \in \mathfrak{R}^{(L+N) \times (L+N)}$ be a symmetric positive definite matrix. If the G -norm is considered, then the definition of a contraction method can be extended to the case with G -norm

$$(2.9) \quad \|u^{k+1} - u^*\|_G \leq \|u^k - u^*\|_G - c\|u^k - u^{k+1}\|_G \quad \forall u^* \in \Omega^*.$$

According to the above definition, it is easy to verify that the PPA (2.2) or (2.3) is a contraction method. We emphasize that the primal-dual procedure (1.3) with $\theta \neq 1$ cannot be regarded as the PPA due to the lack of symmetry of M . Therefore, the existing primal-dual algorithms with $\theta \neq 1$ for (1.1) do not belong to the category of contraction methods. This fact inspires us to investigate how to develop contraction methods for (1.1) based on the primal-dual procedure (1.3).

In the following, we will show that whenever the matrix M defined in (2.5) is positive definite in spite of the lack of symmetry, we can easily find some simple correction steps to correct the iterates generated by the primal-dual procedure (1.3), and the corrected iterates constitute a contractive sequence with respect to Ω^* . As a result, we can develop primal-dual-based contraction methods for (1.1) in the prediction-correction fashion, where the predictor is generated by (1.3) (or (2.4)) and is corrected by some correction step at each iteration. In fact, it is the purpose of ensuring the positive definiteness of M that enables us to significantly relax the requirements on the step sizes τ and σ for existing primal-dual algorithms. In addition, the global convergence of these primal-dual-based contraction methods can be easily derived under the analytic framework of contraction methods.

In order to present the new methods in the prediction-correction fashion, from now on we denote by $\tilde{u}^k = (\tilde{x}^k, \tilde{y}^k)$ the iterate generated by the primal-dual procedure (1.3). Therefore, the primal-dual procedure (1.3) in the context of VI (i.e., (2.4)) can be redescribed as follows.

The primal-dual procedure for (1.1).

Let $\tau > 0$, $\sigma > 0$, and $\theta \in [-1, 1]$; let F and M be defined in (2.1b) and (2.5), respectively. With the given u^k , generate $\tilde{u}^k \in \Omega$ via solving

$$(2.10) \quad (u - \tilde{u}^k)^T \{F(\tilde{u}^k) + M(\tilde{u}^k - u^k)\} \geq 0 \quad \forall u \in \Omega.$$

In sections 3 and 4, we first restrict our discussion to the case when $\theta \in [-1, 1)$, and then we analyze the case of $\theta = 1$ particularly in section 5.

3. A primal-dual-based contraction method for $\theta \in [-1, 1)$. In this section, we develop a primal-dual-based contraction method for (1.1) when $\theta \in [-1, 1)$ in (1.3), with the particular interest in TV image restoration problems. We first elucidate the details of algorithmic design and then present the algorithm. Finally, we prove its convergence.

3.1. The prediction step. As we have mentioned, the predictor \tilde{u}^k is generated by solving (2.10). Thus, for the case $\theta \in [-1, 1)$, the prediction step is described as follows.

The prediction step at the $(k + 1)$ th iteration.
 Let $\theta \in [-1, 1)$. With the given u^k , generate the predictor \tilde{u}^k via solving (2.10).

In the following, we discuss how to realize the primal-dual procedure (1.3) for some TV- l^2 image restoration problems. For these applications, in (1.1), we have $L = 2N$, \mathcal{X} is the Descartes product of some unit boxes (or balls under the infinity norm) in \mathfrak{R}^2 , \mathcal{Y} is a ball in \mathfrak{R}^N with certain (or infinite) radius or the whole space \mathfrak{R}^N , the matrix A corresponds to the matrix representation of the discrete gradient operator, the matrix B represents a deconvolution or subsampling operator such as the denoising or deblurring operator, and the vector z represents a given image. We refer to [25] for the procedure of reformulating TV- l^2 image restoration problems as the saddle-point problem (1.1).

Recall the definition of $\Phi(x, y)$ in (1.1). By deriving the optimality condition of (1.3a) and using Lemma 2.1, the solution of (1.3a) is given by

$$(3.1) \quad \tilde{x}^k = P_{\mathcal{X}}[x^k + \tau A^T y^k].$$

Due to the simplicity of \mathcal{X} , it is easy to compute the projection on \mathcal{X} in (3.1). Thus, the subproblem (1.3a) is easy for TV- l^2 image restoration problems. Now, we elaborate on how to solve the subproblem (1.3c) for different TV- l^2 image restoration problems; see [25] for more details. In the coming analysis, the ROF model is defined in [23].

- *The constrained ROF model.* The discrete constrained ROF model is

$$(3.2) \quad \min_y \int_D |\nabla y| \quad \text{subject to} \quad \|y - z\|^2 \leq |D|\varsigma^2,$$

where D is the image domain with its area being $|D|$, z is the given observed image, ς^2 is an estimate of the variance of the noise in the image z , and ∇ is the discrete gradient operator (see, e.g., [23, 25]). Note that (3.2) can be reformulated as the special case of (1.1)

$$(3.3) \quad \min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \Phi_1(x, y) := y^T Ax,$$

where $\mathcal{Y} := \{y \in \mathfrak{R}^N \mid \|y - z\|^2 \leq |D|\varsigma^2\}$ and $A = \nabla$. In this case, the solution of (1.3c) is given by

$$(3.4) \quad \tilde{y}^k = P_{\mathcal{Y}}[y^k - \sigma A \tilde{x}^k],$$

and the projection on \mathcal{Y} can be computed easily.

- *The unconstrained ROF model.* The discrete unconstrained ROF model is

$$(3.5) \quad \min_y \int_D |\nabla y| + \frac{\lambda}{2} \|y - z\|^2,$$

where $\lambda > 0$ is a constant balancing the data-fidelity and TV regularization terms, and all other terms are defined as (3.2). As shown in [25], (3.5) can be reformulated as the special case of (1.1)

$$(3.6) \quad \min_{y \in \mathfrak{R}^N} \max_{x \in \mathcal{X}} \Phi_2(x, y) := y^T A x + \frac{\lambda}{2} \|y - z\|^2,$$

where again $A = \nabla$. Since $\mathcal{Y} = \mathfrak{R}^N$, for the given y^k and \bar{x}^k , the solution of (1.3c) is given by

$$\nabla_y \left(\sigma \Phi_2(\bar{x}^k, y) + \frac{1}{2} \|y - y^k\|^2 \right) = 0,$$

which means

$$(3.7) \quad \tilde{y}^k = \frac{1}{1 + \sigma\lambda} y^k + \frac{\sigma}{1 + \sigma\lambda} (\lambda z - A\bar{x}^k).$$

- *TV deblurring model.* The discrete blurry and noisy TV restoration model is

$$(3.8) \quad \min_y \int_D |\nabla y| + \frac{\lambda}{2} \|B y - z\|^2,$$

where B is the matrix representation of a space-invariant blurring operator and all other terms are defined as (3.5). Then, (3.8) can be reformulated as the special case of (1.1)

$$(3.9) \quad \min_{y \in \mathfrak{R}^N} \max_{x \in \mathcal{X}} \Phi_3(x, y) := y^T A x + \frac{\lambda}{2} \|B y - z\|^2.$$

Since $\mathcal{Y} = \mathfrak{R}^N$, for given y^k and \bar{x}^k , the solution of (1.3c) is the solution of

$$\nabla_y \left(\sigma \Phi_3(\bar{x}^k, y) + \frac{1}{2} \|y - y^k\|^2 \right) = 0,$$

from which we have

$$(3.10) \quad (\tilde{y}^k - y^k) + \sigma (\lambda B^T (B \tilde{y}^k - z) + A \bar{x}^k) = 0.$$

For B , the Fourier transform of matrix multiplication by B becomes pointwise multiplication in the frequency domain. Hence, (3.10) can be efficiently solved by

$$(3.11) \quad \tilde{y}^k = \mathcal{F}^{-1} \left[\frac{\mathcal{F}(y^k - \sigma A \bar{x}^k) + \sigma \lambda \mathcal{F}(K)^* \odot \mathcal{F}(z)}{1 + \sigma \lambda \mathcal{F}(K)^* \odot \mathcal{F}(K)} \right],$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}[\cdot]$ are the fast Fourier transform (FFT) and inverse FFT operators, respectively, “*” denotes the complex conjugate, and “ \odot ” is the pointwise multiplication operator. For details, see section 2.4.2 in [25].

3.2. The correction step. Recall that a matrix $M \in \mathfrak{R}^{(L+N) \times (L+N)}$ (not necessarily symmetric) is positive definite if there exists a constant $c > 0$ such that

$$(3.12) \quad (u - \tilde{u})^T M(u - \tilde{u}) \geq c \|u - \tilde{u}\|^2 \quad \forall u \neq \tilde{u} \in \mathfrak{R}^{L+N}.$$

As we have mentioned, the matrix M defined in (2.5) is asymmetric except for $\theta = 1$. Thus, the primal-dual procedure (1.3) with $\theta \in [-1, 1)$ cannot be interpreted as an application of PPA, and the resulting sequence is not necessarily contractive with respect to Ω^* . To construct appropriate correction steps and thus yield primal-dual-based contraction methods for (1.1), our idea can be explained as follows. Because $u^* \in \Omega$, it follows from (2.10) that

$$(\tilde{u}^k - u^*)^T \{-F(\tilde{u}^k) + M(u^k - \tilde{u}^k)\} \geq 0.$$

In addition, since $\tilde{u}^k \in \Omega$ and u^* is a solution of $\text{VI}(\Omega, F)$, we have

$$(\tilde{u}^k - u^*)^T F(u^*) \geq 0.$$

Adding the above two inequalities and using the monotonicity of F , we obtain

$$(\tilde{u}^k - u^*)^T M(u^k - \tilde{u}^k) \geq 0,$$

which implies

$$(3.13) \quad (u^k - u^*)^T M(u^k - \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) \quad \forall u^* \in \Omega^*.$$

Therefore, whenever (3.12) is satisfied, we have that

$$(W(u^k - u^*))^T (-W^{-1}M(u^k - \tilde{u}^k)) \leq -c \|u^k - \tilde{u}^k\|^2,$$

where $W \in \mathfrak{R}^{(L+N) \times (L+N)}$ is an arbitrarily symmetric positive definite matrix. In this case, we conclude that $-W^{-1}M(u^k - \tilde{u}^k)$ is a descent direction of the unknown distance function $\frac{1}{2} \|u - u^*\|_W^2$ at the point $u = u^k$, and this direction is able to yield the contraction of proximity to the set Ω^* if an appropriate step size is chosen. We are thus inspired to propose the correction step as follows, given the positive definiteness of M .

The correction step at the $(k + 1)$ th iteration.

If the matrix M defined in (2.5) is positive definite and the predictor \tilde{u}^k is generated by (2.10), then the new iterate is yielded by correcting \tilde{u}^k via

$$(3.14) \quad u^{k+1} = \tilde{u}^k - \alpha W^{-1}M(u^k - \tilde{u}^k),$$

where $W \in \mathfrak{R}^{(L+N) \times (L+N)}$ is a symmetric positive definite matrix and $\alpha > 0$ is a step size to be specified later.

Therefore, in order to present the complete algorithm, we need to investigate how to ensure the positive definiteness of M in (2.5) and specify the choices of the step size α and the matrix W in (3.14). These are tasks to be addressed in the following two subsections.

3.3. How to ensure the positive definiteness of M . Subsections 3.1 and 3.2 emphasize the importance of the positive definiteness of the matrix M in (2.5) for developing primal-dual-based contraction methods for (1.1). In this subsection, we focus on the determination of the step sizes τ and σ for this purpose.

For the convenience of analysis, we define the block diagonal part of M by

$$(3.15) \quad H = \begin{pmatrix} \frac{1}{\tau}I & 0 \\ 0 & \frac{1}{\sigma}I \end{pmatrix}.$$

Obviously, the matrix H defined by (3.15) is positive definite whenever $\tau > 0$ and $\sigma > 0$. Therefore, we choose $W = H$ in (3.14) for the discussion in sections 3 and 4. We now divide the analysis of this subsection into the cases $\theta = -1$ and $\theta \in (-1, 1)$.

Case 1. $\theta = -1$.

When $\theta = -1$, the matrix M (2.5) becomes

$$M = \begin{pmatrix} \frac{1}{\tau}I & A^T \\ -A & \frac{1}{\sigma}I \end{pmatrix},$$

which is the sum of H (3.15) and a skew-symmetric matrix

$$M = H + \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}.$$

In this case, for any positive τ and σ , we have

$$(3.16) \quad (u - \tilde{u})^T M(u - \tilde{u}) = \|u - \tilde{u}\|_H^2,$$

which indicates the positive definiteness of M . Hence, when $\theta = -1$, the step sizes τ and σ can be any positive numbers in order to ensure the positive definiteness of M .

Case 2. $\theta \in (-1, 1)$.

When $\theta \in (-1, 1)$, the condition on τ and σ to ensure the positive definiteness of M can be summarized by the following lemma.

Lemma 3.1. *Let H be defined in (3.15). For $\theta \in (-1, 1)$, if the step sizes τ and σ in (1.3) satisfy*

$$(3.17) \quad \tau\sigma \frac{(1+\theta)^2}{4} < \frac{1}{\|A^T A\|},$$

then for the matrix M in (2.5), we have

$$(3.18) \quad (u - \tilde{u})^T M(u - \tilde{u}) \geq \frac{\delta}{1+\delta} \|u - \tilde{u}\|_H^2 \quad \forall u \neq \tilde{u} \in \mathfrak{R}^{L+N},$$

where

$$(3.19) \quad \delta = \frac{2}{1+\theta} \sqrt{\frac{1}{\tau\sigma\|A^T A\|}} - 1.$$

Proof. Under the condition (3.17), the scalar δ defined in (3.19) is positive and it holds that

$$(3.20) \quad \tau(1 + \delta)\|A^T A\| \frac{(1 + \theta)^2}{4} = \frac{1}{\sigma(1 + \delta)}.$$

For any $u \neq \tilde{u}$, we have

$$(3.21) \quad (u - \tilde{u})^T M(u - \tilde{u}) = \|u - \tilde{u}\|_H^2 + (1 + \theta)(y - \tilde{y})^T A(x - \tilde{x}).$$

By using the Cauchy-Schwarz inequality and (3.20), we get

$$\begin{aligned} (1 + \theta)(y - \tilde{y})^T A(x - \tilde{x}) &= 2\left(\frac{1 + \theta}{2}\right)(y - \tilde{y})^T A(x - \tilde{x}) \\ &\geq -(\tau(1 + \delta)\|A^T A\|) \frac{(1 + \theta)^2}{4} \|y - \tilde{y}\|^2 - \frac{1}{\tau(1 + \delta)\|A^T A\|} \|A(x - \tilde{x})\|^2 \\ &= -\frac{1}{\sigma(1 + \delta)} \|y - \tilde{y}\|^2 - \frac{1}{\tau(1 + \delta)\|A^T A\|} \|A(x - \tilde{x})\|^2 \\ &\geq -\frac{1}{1 + \delta} \left(\frac{1}{\tau} \|x - \tilde{x}\|^2 + \frac{1}{\sigma} \|y - \tilde{y}\|^2 \right) \\ &= -\frac{1}{1 + \delta} \|u - \tilde{u}\|_H^2. \end{aligned}$$

Substituting this in (3.21), the assertion (3.18) is proved. ■

Now we summarize the requirement on the step sizes τ and σ in (1.3) to ensure the positive definiteness of M .

The step sizes τ and σ of the primal-dual procedure (1.3).

$$(3.22) \quad \begin{cases} \tau \text{ and } \sigma \text{ are any positive numbers} & \text{if } \theta = -1, \\ \tau\sigma \frac{(1+\theta)^2}{4} < \frac{1}{\|A^T A\|} & \text{if } \theta \in (-1, 1). \end{cases}$$

Remark 3.2. Compared to (1.4) analyzed in [6], we now allow the step sizes τ and σ to be chosen according to the rule (3.22). In fact, τ and σ can be arbitrarily large when $\theta = -1$, and they can be arbitrarily large simultaneously if θ is sufficiently close to -1 . Hence, the requirement on the step sizes τ and σ in [6] is significantly relaxed by (3.22).

Remark 3.3. Note that the condition (3.17) is a sufficient condition to ensure the positive definiteness of M . In fact, the positive definiteness of M can be guaranteed if the step sizes τ and σ are chosen to satisfy

$$(3.23) \quad (u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) \geq \frac{\delta}{1 + \delta} \|u^k - \tilde{u}^k\|_H^2 \quad \forall k > 0,$$

where $\delta > 0$ is a constant. In practical computation, we can use the classical Armijo's technique to find a pair of τ and σ to satisfy the condition (3.23) in the absence of the value of $\|A^T A\|$.

3.4. Step size of the correction step. Now we analyze how to determine the step size α for the correction step (3.14) where $W = H$. Moving along the direction $-H^{-1}M(u^k - \tilde{u}^k)$ from the current point u^k by a suitable step size, we can obtain a new iterate which is closer to the solution set Ω^* ; this is the essence of contraction methods. To determine an appropriate value of α , we let

$$u(\alpha) := u^k - \alpha H^{-1}M(u^k - \tilde{u}^k).$$

Further, we denote

$$(3.24) \quad \vartheta_k(\alpha) := \|u^k - u^*\|_H^2 - \|u(\alpha) - u^*\|_H^2,$$

which can measure the progress on the proximity to the solution set Ω^* made by the new iterate $u(\alpha)$. Because $\vartheta_k(\alpha)$ involves the unknown vector u^* , we are not able to maximize it directly. But, with (3.13), we have that

$$\begin{aligned} \vartheta_k(\alpha) &= \|u^k - u^*\|_H^2 - \|u^k - \alpha H^{-1}M(u^k - \tilde{u}^k) - u^*\|_H^2 \\ &= 2\alpha(u^k - u^*)^T M(u^k - \tilde{u}^k) - \alpha^2 \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \\ &\geq 2\alpha(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) - \alpha^2 \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2. \end{aligned}$$

We define

$$(3.25) \quad q_k(\alpha) = 2\alpha(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) - \alpha^2 \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2;$$

then $q_k(\alpha)$ is a lower bound of $\vartheta_k(\alpha)$. Hence, we can choose a value of α such that $q_k(\alpha)$, instead of $\vartheta_k(\alpha)$, is maximized. Note that $q_k(\alpha)$ is a quadratic function of α and reaches its maximum at

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k)}{\|H^{-1}M(u^k - \tilde{u}^k)\|_H^2}.$$

Thus, we choose α_k^* as above for the step size of the correction step (3.14).

3.5. Algorithm. With the specifications of the prediction step, the requirement on the step sizes τ and σ to ensure the positive definiteness of M , and the correction step and its step size, we are now ready to present a primal-dual-based contraction method for (1.1) when $\theta \in [-1, 1)$ in (2.10).

Algorithm 1: A primal-dual-based contraction method for (1.1) with $\theta \in [-1, 1)$ in (2.10).

Step 0. Let $\gamma \in (0, 2)$. Let $\theta \in [-1, 1)$, M be defined in (2.5), and H be defined in (3.15). Take $u^0 \in \mathfrak{R}^{L+N}$. Choose the step sizes τ and σ according to (3.22).

Prediction step: Generate the predictor \tilde{u}^k via solving the primal-dual procedure (1.3), i.e., the VI (2.10).

Correction step: Correct the predictor and generate the new iterate u^{k+1} via

$$(3.26a) \quad u^{k+1} = u^k - \alpha_k H^{-1}M(u^k - \tilde{u}^k),$$

where

$$(3.26b) \quad \alpha_k = \gamma \alpha_k^* \quad \text{and} \quad \alpha_k^* = \frac{(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k)}{\|H^{-1}M(u^k - \tilde{u}^k)\|_H^2}.$$

Remark 3.4. Since we choose the step size α_k^* via minimizing the lower bound $q_k(\alpha)$ in (3.25), rather than the real distance difference $\vartheta_k(\alpha)$ in (3.24), the “optimal” step size α_k^* is usually conservative for the contraction purpose. Thus, we attach a relaxation parameter γ to the step size α_k^* in (3.26b). The reason why we restrict $\gamma \in (0, 2)$ will be clear in the next subsection (see Lemma 3.5).

3.6. Convergence. In this subsection, we show that the proposed Algorithm 1 is a contraction method according to the definition (2.9), and then we prove its convergence under the analytic framework of contraction methods.

Lemma 3.5. *There exists a constant $c_0 > 0$ such that the sequence $\{u^k\}$ generated by the proposed Algorithm 1 satisfies*

$$(3.27a) \quad (u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) \geq c_0 \|u^k - \tilde{u}^k\|_H^2,$$

$$(3.27b) \quad \|u^{k+1} - u^*\|_H^2 \leq \|u^k - u^*\|_H^2 - \gamma(2 - \gamma)\alpha_k^* c_0 \|u^k - \tilde{u}^k\|_H^2 \quad \forall u^* \in \Omega^*.$$

Proof. First, for the case $\theta = -1$, it follows from (3.16) that the assertion (3.27) holds with $c_0 = 1$, and for the case $\theta \in (-1, 1)$, it follows from (3.18) in Lemma 3.1 that the assertion (3.27) holds with $c_0 = \frac{\delta}{1+\delta}$, where δ is given in (3.19). Therefore, the assertion (3.27a) is proved. Second, using (3.26) and (3.13), we have that

$$\begin{aligned} \|u^{k+1} - u^*\|_H^2 &= \|u^k - u^* - \alpha_k H^{-1} M(u^k - \tilde{u}^k)\|_H^2 \\ &= \|u^k - u^*\|_H^2 - 2\alpha_k (u^k - u^*)^T M(u^k - \tilde{u}^k) + \alpha_k^2 \|H^{-1} M(u^k - \tilde{u}^k)\|_H^2 \\ &\leq \|u^k - u^*\|_H^2 - 2\alpha_k (u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) + \alpha_k^2 \|H^{-1} M(u^k - \tilde{u}^k)\|_H^2 \\ &= \|u^k - u^*\|_H^2 - \gamma(2 - \gamma)\alpha_k^* (u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) \quad \forall u^* \in \Omega^*. \end{aligned}$$

Thus, the assertion (3.27b) is proved by substituting (3.27a) into the above inequality. ■

Based on Lemma 3.5, we can easily show that the proposed Algorithm 1 is a contraction method for (1.1), as stated in the following lemma.

Lemma 3.6. *The proposed Algorithm 1 is a contraction method for (1.1).*

Proof. First, it follows from (3.27a) that the “optimal” step size (see (3.26b)) is bounded below:

$$\alpha_k^* \geq \frac{c_0 \|u^k - \tilde{u}^k\|^2}{\|H^{-1} M(u^k - \tilde{u}^k)\|_H^2} \geq \frac{c_0}{\|M^T H^{-1} M\|}.$$

Consequently, the assertion (3.27b) and the above inequality show

$$(3.28) \quad \|u^{k+1} - u^*\|_H^2 \leq \|u^k - u^*\|_H^2 - \frac{\gamma(2 - \gamma)c_0^2}{\|M^T H^{-1} M\|} \|u^k - \tilde{u}^k\|_H^2 \quad \forall u^* \in \Omega^*.$$

Then, it follows from (3.26a) and (3.28) that (2.9) is satisfied by the sequence $\{u^k\}$ generated by Algorithm 1. That is, Algorithm 1 is a contraction method for (1.1). ■

Now, because of Lemma 3.6, we can easily prove the convergence of the proposed Algorithm 1 by following the standard analytic framework of contraction-type methods; see, e.g., [5].

Theorem 3.7. *The sequence generated by the proposed Algorithm 1 converges to a solution point of (1.1).*

Proof. According to (3.28), the sequence $\{u^k\}$ generated by the proposed Algorithm 1 is bounded. In fact, the sequence is contained by the compact set

$$S := \{u \in \mathfrak{R}^{L+N} \mid \|u - u^*\|_H \leq \|u^0 - u^*\|_H\},$$

where u^* is an arbitrary point in Ω^* and $u^0 \in \mathfrak{R}^{L+N}$. Thus, the sequence $\{u^k\}$ has at least a cluster point, say u^∞ , and we assume that the subsequence $\{u^{k_j}\}$ converges to u^∞ . Note that (3.28) immediately implies that $\|u^{k_j} - \tilde{u}^{k_j}\| \rightarrow 0$ when $k_j \rightarrow \infty$. Thus, taking the limit over $k_j \rightarrow \infty$ in (2.10), we have that u^∞ is a solution of VI (Ω, F) and thus of (1.1). Note that (3.28) also implies that u^∞ is the unique cluster point of the sequence $\{u^k\}$. Thus, $\{u^k\}$ converges to u^∞ , a solution point of (1.1), starting from any u^0 . The global convergence of Algorithm 1 is established. ■

4. A reduced primal-dual-based contraction method for $\theta \in [-1, 1]$. In this section, we show that the procedure of determining the optimal step size α_k^* at the correction step (3.26b) can be reduced. Thus, a reduced primal-dual-based contraction method with an easier correction step for (1.1) is proposed. With an easier correction step, however, we will show that the allowable ranges of the step sizes τ and σ are narrower than those in Algorithm 1.

Instead of choosing the step length α_k judiciously as (3.26b), we can simply take $\alpha_k \equiv 1$ and thus the correction step (3.26a) is modified to

$$(4.1) \quad u^{k+1} = u^k - H^{-1}M(u^k - \tilde{u}^k).$$

The proposed Algorithm 1 but with the easier correction step (4.1), i.e., Algorithm 2 to be proposed, is called a reduced primal-dual-based contraction method. We are now interested in the conditions on τ and σ for ensuring that this new method is a contraction method.

Lemma 4.1. *Assume that the matrix M (2.5) is positive definite. Let H be defined in (3.15). Let $\{u^k\}$ be the sequence generated by the proposed Algorithm 1 but with the easier correction step (4.1). If there exists $c_1 > 0$ such that*

$$(4.2) \quad 2(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) - \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \geq c_1 \|u^k - \tilde{u}^k\|_H^2$$

holds for all $k > 0$, then, the reduced primal-dual-based contraction method with (4.1) is a contraction method in the sense that

$$(4.3) \quad \|u^{k+1} - u^*\|_H^2 \leq \|u^k - u^*\|_H^2 - c_1 \|u^k - \tilde{u}^k\|_H^2 \quad \forall u^* \in \Omega^*.$$

Proof. Using (4.1) and (3.13), we have that

$$\begin{aligned} \|u^{k+1} - u^*\|_H^2 &= \|u^k - u^* - H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \\ &= \|u^k - u^*\|_H^2 - 2(u^k - u^*)^T M(u^k - \tilde{u}^k) + \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \\ &\leq \|u^k - u^*\|_H^2 - 2(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) + \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \\ &\leq \|u^k - u^*\|_H^2 - c_1 \|u^k - \tilde{u}^k\|_H^2 \quad \forall u^* \in \Omega^*, \end{aligned}$$

where the last inequality comes from (4.2). Together with (4.1), the above fact ensures that (2.9) holds for this new algorithm. The assertion of this lemma is proved. ■

Now, we analyze how to determine appropriate values of the step sizes τ and σ in (1.3) to guarantee the validation of (4.2).

Lemma 4.2. *Let M and H be defined as in (2.5) and (3.15), respectively. For $\theta \in [-1, 1)$, if the step sizes τ and σ satisfy*

$$(4.4) \quad \tau\sigma < \frac{1}{\|A^T A\|},$$

then we have

$$(4.5) \quad 2(u - \tilde{u})^T M(u - \tilde{u}) - \|H^{-1}M(u - \tilde{u})\|_H^2 \geq (1 - \tau\sigma\|A^T A\|)\|u - \tilde{u}\|_H^2 \quad \forall u, \tilde{u} \in \mathfrak{R}^{L+N}.$$

Proof. Using the definitions of the matrices M and H , we have that

$$2(u - \tilde{u})^T M(u - \tilde{u}) = \frac{2}{\tau}\|x - \tilde{x}\|^2 + \frac{2}{\sigma}\|y - \tilde{y}\|^2 + 2(1 + \theta)(y - \tilde{y})^T A(x - \tilde{x}).$$

On the other hand, since

$$H^{-1}M(u - \tilde{u}) = \begin{pmatrix} x - \tilde{x} + \tau A^T(y - \tilde{y}) \\ y - \tilde{y} + \sigma\theta A(x - \tilde{x}) \end{pmatrix},$$

we obtain that

$$\begin{aligned} \|H^{-1}M(u - \tilde{u})\|_H^2 &= \frac{1}{\tau}\|(x - \tilde{x}) + \tau A^T(y - \tilde{y})\|^2 + \frac{1}{\sigma}\|(y - \tilde{y}) + \sigma\theta A(x - \tilde{x})\|^2 \\ &= \frac{1}{\tau}\|x - \tilde{x}\|^2 + \frac{1}{\sigma}\|y - \tilde{y}\|^2 + \tau\|A^T(y - \tilde{y})\|^2 + \sigma\theta^2\|A(x - \tilde{x})\|^2 \\ &\quad + 2(1 + \theta)(y - \tilde{y})^T A(x - \tilde{x}). \end{aligned}$$

Therefore, we get

$$(4.6) \quad \begin{aligned} &2(u - \tilde{u})^T M(u - \tilde{u}) - \|H^{-1}M(u - \tilde{u})\|_H^2 \\ &= \left(\frac{1}{\tau}\|x - \tilde{x}\|^2 + \frac{1}{\sigma}\|y - \tilde{y}\|^2\right) - \left(\tau\|A^T(y - \tilde{y})\|^2 + \sigma\theta^2\|A(x - \tilde{x})\|^2\right). \end{aligned}$$

It follows from (4.6) that

$$\begin{aligned} &2(u - \tilde{u})^T M(u - \tilde{u}) - \|H^{-1}M(u - \tilde{u})\|_H^2 \\ &= \left(\frac{1}{\tau}\|x - \tilde{x}\|^2 - \sigma\theta^2\|A(x - \tilde{x})\|^2\right) + \left(\frac{1}{\sigma}\|y - \tilde{y}\|^2 - \tau\|A^T(y - \tilde{y})\|^2\right) \\ &\geq (1 - \tau\sigma\theta^2\|A^T A\|)\frac{1}{\tau}\|x - \tilde{x}\|^2 + (1 - \tau\sigma\|A A^T\|)\frac{1}{\sigma}\|y - \tilde{y}\|^2 \\ &\geq (1 - \tau\sigma\|A^T A\|)\|u - \tilde{u}\|_H^2, \end{aligned}$$

where the last inequality follows from the fact that $\theta^2 \leq 1$. Thus, this lemma is proved. ■

Remark 4.3. According to Lemma 4.2, the condition (4.4) is a sufficient condition to ensure (4.2) with $c_1 = 1 - \tau\sigma\|A^T A\|$. In fact, from the proof of Lemma 4.6, we need only guarantee that the right-hand side of (4.6) is greater than $c_1\|u - \tilde{u}\|_H^2$ with $c_1 > 0$, in order to meet (4.2). In other words, the equality (4.6) implies that (4.2) holds if the step sizes τ and σ are chosen to guarantee that

$$(4.7) \quad \tau\|A^T(y^k - \tilde{y}^k)\|^2 + \sigma\theta^2\|A(x^k - \tilde{x}^k)\|^2 \leq (1 - c_1)\left(\frac{1}{\tau}\|x^k - \tilde{x}^k\|^2 + \frac{1}{\sigma}\|y^k - \tilde{y}^k\|^2\right) \quad \forall k > 0$$

for a given scalar $c_1 > 0$. In practical computation, we can use the classical Armijo technique to find values of τ and σ to satisfy (4.7) in the absence of the value $\|A^T A\|$.

Remark 4.4. As we have mentioned, an important case of the primal-dual procedure (1.3) is the primal-dual hybrid gradient (PDHG) method in [25] with $\theta = 0$ in (1.3b). For this case, we notice that the condition (4.7) for determining values of τ and σ can be simplified to

$$\tau\|A^T(y^k - \tilde{y}^k)\|^2 \leq (1 - c_1)\left(\frac{1}{\tau}\|x^k - \tilde{x}^k\|^2 + \frac{1}{\sigma}\|y^k - \tilde{y}^k\|^2\right) \quad \forall k > 0,$$

for a given scalar $c_1 > 0$.

Now, we are ready to present the reduced primal-dual-based contraction method for (1.1) with $\theta = [-1, 1)$ in (2.10).

Algorithm 2: A reduced primal-dual-based contraction method for (1.1) with $\theta \in [-1, 1)$ in (2.10).

Step 0. Let $\theta \in [-1, 1)$, M be defined in (2.5), and H be defined in (3.15). Take $u^0 \in \mathfrak{R}^{L+N}$. Choose the step sizes τ and σ according to (4.4).

Prediction step: Generate the predictor \tilde{u}^k via solving the primal-dual procedure (1.3), i.e., the VI (2.10).

Correction step: Correct the predictor and generate the new iterate u^{k+1} via

$$(4.8) \quad u^{k+1} = u^k - H^{-1}M(u^k - \tilde{u}^k).$$

Remark 4.5. Since Lemmas 4.1 and 4.2 show that the proposed Algorithm 2 is a contraction method for (1.1) under the condition (4.4), the convergence analysis of Algorithm 2 is similar to that of Algorithm 1 and it is omitted.

To close this subsection, we discuss the relationship of the proposed Algorithm 2 with the PDHG method in [25]. Recall that the PDHG method in [25] is the special case of the primal-dual procedure (1.3) with $\theta = 0$, and its convergence was proved in [11] with very small step sizes τ and σ . When $\theta = 0$, the primal-dual procedure (1.3) reduces to

$$(4.9a) \quad \tilde{x}^k = \text{Arg max}_{x \in \mathcal{X}} \left\{ \tau\Phi(x, y^k) - \frac{1}{2}\|x - x^k\|^2 \right\},$$

$$(4.9b) \quad \tilde{y}^k = \text{Arg min}_{y \in \mathcal{Y}} \left\{ \sigma\Phi(\tilde{x}^k, y) + \frac{1}{2}\|y - y^k\|^2 \right\}.$$

Then, the PDHG method in [25] takes

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \end{pmatrix}$$

as the new iterate. On the other hand, the proposed Algorithm 2 generates the new iterate by

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{x}^k - \tau A^T(y^k - \tilde{y}^k) \\ \tilde{y}^k \end{pmatrix}.$$

Hence, when $\theta = 0$, the proposed Algorithm 2 differs from the PDHG method in [25] in that the variable \tilde{x}^k is further corrected. Recall that the PDHG method in [25] is essentially an Arrow–Hurwicz algorithm in [1]. Since the variable \tilde{x}^k is further corrected and the \tilde{y}^k remains uncorrected, the proposed Algorithm 2 with $\theta = 0$ can be regarded as a semi-implicit Arrow–Hurwicz algorithm.

5. Primal-dual-based contraction methods for $\theta = 1$. In this section, we pay attention to the special case of (1.3) when $\theta = 1$. As we have emphasized, for this case, the matrix M defined in (2.5) becomes

$$(5.1) \quad M = \begin{pmatrix} \frac{1}{\tau}I & A^T \\ A & \frac{1}{\sigma}I \end{pmatrix},$$

and it is symmetric. Thus, the primal-dual procedure (1.3) with $\theta = 1$ is exactly an application of the classical PPA whenever the positive definiteness of M is guaranteed. Then, the iterate scheme (1.3) without any correction step is a contraction method with respect to the solution set Ω^* . Similarly to Lemma 3.1, we can easily show that the positive definiteness of the matrix M in (5.1) is guaranteed if

$$(5.2) \quad \tau\sigma < \frac{1}{\|A^T A\|}.$$

Thus, with the restriction (5.2) on the step sizes, the primal-dual procedure (1.3) with $\theta = 1$ is a contraction-type method, and this is exactly the extrapolational gradient method in [6].

Notice that the condition (5.2) coincides with the condition (3.17) by taking $\theta = 1$ and the condition (4.4). All the analysis conducted in sections 3 and 4 is valid for $\theta = 1$. Thus, we can easily extend the proposed Algorithms 1 and 2 to the case $\theta = 1$. In particular, by extending Algorithm 2, we immediately obtain a reduced primal-dual-based contraction method for (1.1) with $\theta = 1$ in (2.10).

Algorithm 3: A reduced primal-dual-based contraction method for (1.1) with $\theta = 1$ in (2.10).

Step 0. Let $\theta = 1$. Take $u^0 \in \mathfrak{R}^{L+N}$. Choose the step sizes τ and σ according to (5.2).

Prediction step: Generate the predictor \tilde{u}^k via solving the primal-dual procedure (1.3), i.e., the VI (2.10).

Correction step: Correct the predictor and generate the new iterate u^{k+1} via

$$(5.3) \quad u^{k+1} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{x}^k - \tau A^T(y^k - \tilde{y}^k) \\ \tilde{y}^k - \sigma A(x^k - \tilde{x}^k) \end{pmatrix}.$$

Because of the speciality of $\theta = 1$, in this section we show that another primal-dual-based contraction method can be easily derived for (1.1) with $\theta = 1$. The new primal-dual-based contraction method is also in the prediction-correction fashion, and it differs from the

proposed Algorithms 1–3 in that its correction step is much less demanding computationally. Our motivation is that the aforementioned correction steps (3.26a), (4.8), and (5.3) all require multiplications of the matrices A and A^T with some vectors. For TV image restoration problems, these multiplications might be computationally demanding. Thus, it is desired to develop a primal-dual-based contraction method for (1.1) such that its correction step is simple enough to involve no multiplications of matrices and vectors.

Similar to the motivation stated in section 3.2, our idea is that for the case $\theta = 1$, we can easily derive that

$$(5.4) \quad (u^k - u^*)^T M(u^k - \tilde{u}^k) \geq \|u^k - \tilde{u}^k\|^2 \quad \forall u^* \in \Omega^*,$$

which indicates that the direction $-(u^k - \tilde{u}^k)$ is a descent direction of the function $\frac{1}{2}\|u - u^*\|^2$ at the point $u = u^k$. Thus, we can move closer to the solution set Ω^* along the direction $-(u^k - \tilde{u}^k)$. The resulting primal-dual-based contraction method for $\theta = 1$ in (2.10) is presented as follows.

Algorithm 4: A new primal-dual-based contraction method for (1.1) with $\theta = 1$ in (2.10).

Step 0. Let $\theta = 1$ and $\rho \in (0, 2)$. Take $u^0 \in \mathfrak{R}^{L+N}$. Choose the step sizes τ and σ according to (5.2).

Prediction step: Generate the predictor \tilde{u}^k via solving the primal-dual procedure (1.3), i.e., the VI (2.10).

Correction step: Correct the predictor and generate the new iterate u^{k+1} via

$$(5.5) \quad u^{k+1} = u^k - \rho(u^k - \tilde{u}^k).$$

Remark 5.1. Recall that the primal-dual procedure (1.3) is exactly the PPA (see (2.3)) when $\theta = 1$. Thus, Algorithm 4 can also be regarded as a relaxed PPA, or a PPA-based contraction method.

Lemma 5.2. Let M be defined in (5.1), where the condition (5.2) is satisfied. The sequence $\{u^k\}$ generated by the proposed Algorithm 4 satisfies

$$\|u^{k+1} - u^*\|_M^2 \leq \|u^k - u^*\|_M^2 - \rho(2 - \rho)\|u^k - \tilde{u}^k\|_M^2 \quad \forall u^* \in \Omega^*.$$

Proof. First of all, the positive definiteness of M in (5.1) is ensured under the condition (5.2). Then, using (5.5) and (5.4) and by a simple manipulation, we obtain

$$\begin{aligned} \|u^{k+1} - u^*\|_M^2 &= \|u^k - u^* - \rho(u^k - \tilde{u}^k)\|_M^2 \\ &= \|u^k - u^*\|_M^2 - 2\rho(u^k - u^*)^T M(u^k - \tilde{u}^k) + \rho^2\|u^k - \tilde{u}^k\|_M^2 \\ &\leq \|u^k - u^*\|_M^2 - 2\rho\|u^k - \tilde{u}^k\|_M^2 + \rho^2\|u^k - \tilde{u}^k\|_M^2 \\ &= \|u^k - u^*\|_M^2 - \rho(2 - \rho)\|u^k - \tilde{u}^k\|_M^2 \quad \forall u^* \in \Omega^*. \end{aligned}$$

The assertion is proved. ■

Remark 5.3. According to Lemma 5.2, it is clear why the parameter ρ is restricted in the interval $(0, 2)$. In fact, when $\rho \in (0, 2)$, (5.5) and the assertion of Lemma 5.2 immediately indicate that the sequence $\{u^k\}$ generated by Algorithm 4 satisfies (2.9) under the M -norm.

That is, Algorithm 4 is a contraction method with respect to Ω^* . Hence, the convergence of Algorithm 4 can be established easily, and we omit it.

At the end of this section, we analyze the difference between the extrapolational gradient method in [6] and the proposed Algorithm 4. Since $\theta = 1$, the primal-dual procedure (1.3) reduces to

$$(5.6a) \quad \tilde{x}^k = \text{Arg max}_{x \in \mathcal{X}} \left\{ \tau \Phi(x, y^k) - \frac{1}{2} \|x - x^k\|^2 \right\},$$

$$(5.6b) \quad \tilde{y}^k = \text{Arg min}_{y \in \mathcal{Y}} \left\{ \sigma \Phi(2(\tilde{x}^k - x^k), y) + \frac{1}{2} \|y - y^k\|^2 \right\}.$$

Then, the extrapolational gradient method in [6] takes

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \end{pmatrix}$$

as the new iterate, and the convergence is guaranteed under the condition (5.2). Under the same condition (5.2), the proposed Algorithm 4 takes a combination of u^k and \tilde{u}^k to generate a new iterate:

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \rho \begin{pmatrix} x^k - \tilde{x}^k \\ y^k - \tilde{y}^k \end{pmatrix} \quad \text{with } \rho \in (0, 2).$$

Obviously, the extrapolational gradient method in [6] is a special case of the proposed Algorithm 4 with $\rho = 1$.

6. Extension to a general saddle-point problem. As we have mentioned, we can extend our convergence analysis to the saddle-point problem (1.2) which was considered in [6]. The convergence analysis for this general model is completely analogous to the previous analysis for (1.1). Thus we only briefly present the procedure of reformulating (1.2) as a variational inequality and explain how to express existing primal-dual algorithms in [6] by the structure of PPA but with an asymmetric preconditioning matrix.

Similarly, the problem (1.2) can be reformulated as the following monotone variational inequality: find $(x^*, y^*) \in X \times Y$, $F(x^*) \in \partial f^*(x^*)$, and $G(y^*) \in \partial g(y^*)$ such that

$$(6.1) \quad \begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} F(x^*) - K^T y^* \\ G(y^*) + K x^* \end{pmatrix} \geq 0 \quad \forall (x, y) \in X \times Y,$$

where $\partial(\cdot)$ denotes the subdifferential operator of a convex function. By denoting

$$(6.2) \quad u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad U(u) = \begin{pmatrix} F(x) - K^T y \\ G(y) + K x \end{pmatrix}, \quad \text{and } \Omega = X \times Y,$$

(6.1) is a monotone VI denoted by VI (Ω, U) . Note that the monotonicity of the VI (Ω, U) is guaranteed by the convexity of f^* and g .

Recall that the primal-dual algorithm for (1.2) presented in [6] is as follows.

Chambolle and Pock's primal-dual algorithm for (1.2).

$$(6.3a) \quad x^{k+1} = \operatorname{Arg} \max_{x \in X} \left\{ \tau ((y^k)^T K x - f^*(x)) - \frac{1}{2} \|x - x^k\|^2 \right\},$$

$$(6.3b) \quad \bar{x}^k = x^{k+1} + \theta(x^{k+1} - x^k), \quad \theta \in [0, 1],$$

$$(6.3c) \quad y^{k+1} = \operatorname{Arg} \min_{y \in Y} \left\{ \sigma(g(y) + y^T K \bar{x}^k) + \frac{1}{2} \|y - y^k\|^2 \right\}.$$

With analogous reasoning, we can easily verify that the iterate (x^{k+1}, y^{k+1}) generated by (6.3) can be characterized as follows: $(x^{k+1}, y^{k+1}) \in X \times Y$, $F(x^{k+1}) \in \partial f^*(x^{k+1})$, and $G(y^{k+1}) \in \partial g(y^{k+1})$ such that

$$\begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} F(x^{k+1}) - K^T y^{k+1} \\ G(y^{k+1}) + K x^{k+1} \end{pmatrix} + \begin{bmatrix} \frac{1}{\tau} I & K^T \\ \theta K & \frac{1}{\sigma} I \end{bmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right\} \geq 0$$

for any $(x, y) \in X \times Y$. Using the notation in (6.2), we have the compact form

$$(u - u^{k+1})^T \{U(u^{k+1}) + M(u^{k+1} - u^k)\} \geq 0 \quad \forall u \in \Omega,$$

where M is defined the same as (2.5) if we regard K as A . Therefore, we obtain the same assertion in Lemma 2.2 for the case (1.2). Consequently, all the convergence analysis conducted in sections 3–5 can be implemented analogously for the model (1.2).

7. Numerical experiments. In this section, we apply the proposed algorithms to solve some TV image deblurring and inpainting problems, and report the numerical results. For these applications, the linear operator A in (1.1) is thus the matrix representation of the discrete gradient operator, and the matrix B in (1.1) is a deconvolution or subsampling operator. We also compare these algorithms numerically with some existing efficient methods. More specifically, this section consists of the following three parts.

- Subsection 7.1 illustrates the sensitivity of the involved parameters of the proposed algorithms by a TV image deblurring problem. The sensitivity of the step sizes τ and σ , the combination parameter θ , and the parameter ρ of the proposed algorithms will be investigated in this subsection.
- Subsection 7.2 compares the proposed algorithms numerically with the PDGH method in [25] for solving some TV image deblurring problems.
- Subsection 7.3 compares the proposed algorithms numerically with the PDHG method in [25] and the first-order primal-dual algorithm in [6] for some TV image inpainting problems.

To report the numerical results, “It.” and “CPU” in the following tables represent the iteration numbers and computing time in seconds, respectively. The quality of restored images is measured by the value of the signal-to-noise ratio (SNR) given by

$$\text{SNR} := 20 \log_{10} \frac{\|y^*\|}{\|\bar{y} - y^*\|},$$

where \bar{y} is the image restored by a certain algorithm and y^* is the original one.

All the proposed algorithms were coded by MATLAB 7.9, and all the numerical experiments were conducted on a Lenovo laptop with Intel Core CPU 2.30 GHZ and 8G memory.

7.1. Sensitivity of parameters. One of the contributions in this paper is that we enlarge the allowable range of the combination parameter θ from $[0, 1]$ to $[-1, 1]$. Recall that Algorithms 1 and 2 are proposed for the case $\theta \in [-1, 1)$, while Algorithms 3 and 4 are for the case $\theta = 1$. In addition, the requirements on the step sizes τ and σ are different when $\theta = -1$ and $\theta \in (-1, 1)$; see (3.22) and (4.4), respectively. Thus, we divide the discussion of this subsection into three cases, $\theta \in (-1, 1)$, $\theta = -1$, and $\theta = 1$, and report the respective sensitivity results in three different subsections. In the following, we summarize the main task of each case and thus clarify how we will test the sensitivity of parameters for the proposed algorithms.

- Let $\theta \in (-1, 1)$. Due to the symmetric role of the step sizes τ and σ , we take τ as a fixed value throughout this case and test various values of σ . More specifically, for a given value of σ , we test the variation of Algorithm 1's numerical performance with respect to various choices of $\theta \in (-1, 1)$. The sensitivity of θ and σ are thus illustrated in some senses for Algorithm 1. Since Algorithm 2 is the reduced version of Algorithm 1 and our numerical experiments indicate the similar numerical performance between these two algorithms, we omit the sensitivity results of Algorithm 2 for the sake of succinctness. Another reason for excluding the sensitivity results of Algorithm 2 is its similarity to Algorithm 3, whose sensitivity results will be tested for the case $\theta = 1$.
- Let $\theta = -1$. We test the variation of Algorithm 1's numerical performance with respect to various choices of τ or σ when one of them is fixed. For the same reason as before, we omit the details of the sensitivity of Algorithm 2 for this case.
- Let $\theta = 1$. We first test the variation of Algorithm 3's numerical performance with respect to various choices of τ or σ when one of them is fixed. Note that we omit the sensitivity results of τ and σ for Algorithm 4, as they are analogous to those of Algorithm 3. Then, since Algorithm 4 involves the parameter ρ in the correction step, we also test the sensitivity of ρ for Algorithm 4 when the step sizes τ and σ are fixed.

We illustrate the sensitivity of parameters by deblurring the image `Cameraman.png` (256×256). The original image is degraded by the `gaussian` blur with `hsize` = 21 and `sigma` = 5 (which can be realized by the script `fspecial` in the MATLAB Image Processing Toolbox). Moreover, the additive zero-mean Gaussian noise with the standard deviation 10^{-3} is added to the blurred image. The original and degraded images are shown in Figure 1. To deblur this corrupted image, we take $\lambda = 1000$ in (1.1). Note that there is a relaxation parameter γ for Algorithm 1. For the simpler exposition of the main results, we simply take $\gamma = 1.6$ throughout when we implement Algorithm 1.

7.1.1. $\theta \in (-1, 1)$. For this case, recall that the requirement on the step sizes τ and σ is (3.22). Thus, when τ is fixed, the allowable step size σ can be expressed as

$$\sigma = \frac{\mu}{\tau(1 + \theta)^2 \|A^T A\|} \quad \text{with } \mu \in (0, 4).$$

In our numerical experiment, we fixed $\tau = 0.1$ when we implemented Algorithm 1. In Figure 2, we plot the evolution of SNR values with respect to 200 iterations, for various values of θ and μ .



Figure 1. Left: original *Camerman.png* (256×256). Right: degraded image.

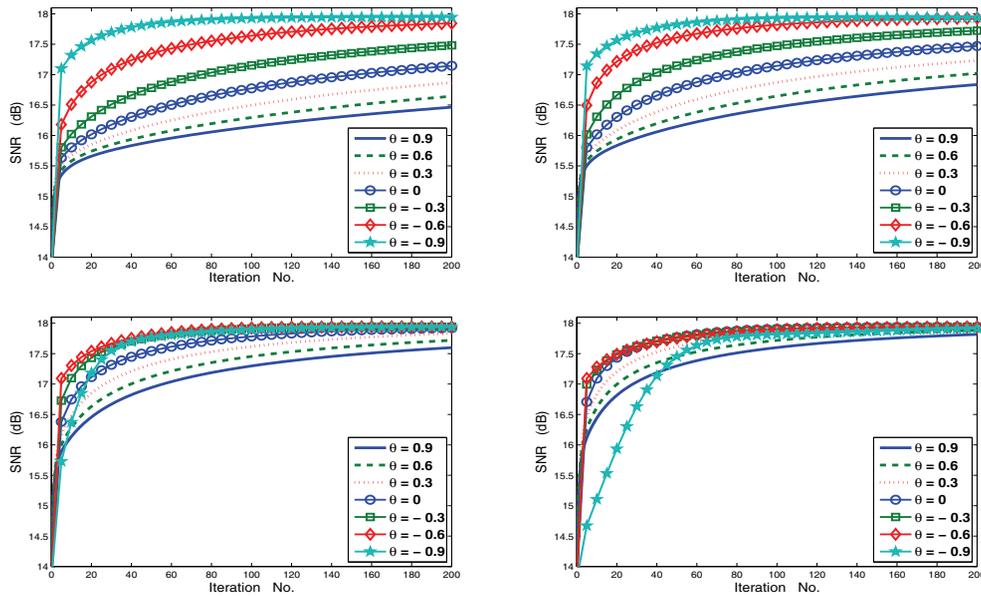


Figure 2. Evolutions of SNRs with respect to iterations for $\theta \in (-1, 1)$. From top to bottom and left to right: $\mu = 0.05$, $\mu = 0.1$, $\mu = 0.5$, $\mu = 1$.

According to the curves in Figure 2, the numerical performance of Algorithm 1 could be improved by taking negative values of θ when other parameters are fixed. This verifies empirically the necessity of enlarging the allowable range of θ from $[0, 1]$ to $[-1, 1]$.

7.1.2. $\theta = -1$. For this case, recall that the step sizes τ and σ can be arbitrarily large positive numbers. In Figure 3, we show the variation of Algorithm 1's numerical performance with respect to various choices of σ when τ is fixed as 0.01, and that of τ when σ is fixed as 0.1.

The curves in Figure 3 indicate that Algorithm 1 could be accelerated when relatively larger step sizes are chosen appropriately. For instance, when $\tau = 0.01$, Algorithm 1 with $\sigma = 100$ performs much faster than $\sigma = 0.1$ and 0.001. Thus, the enlargement of the allowable ranges of τ and σ provides more freedom for tuning these parameters, and it makes it possible to accelerate Algorithm 1 considerably.

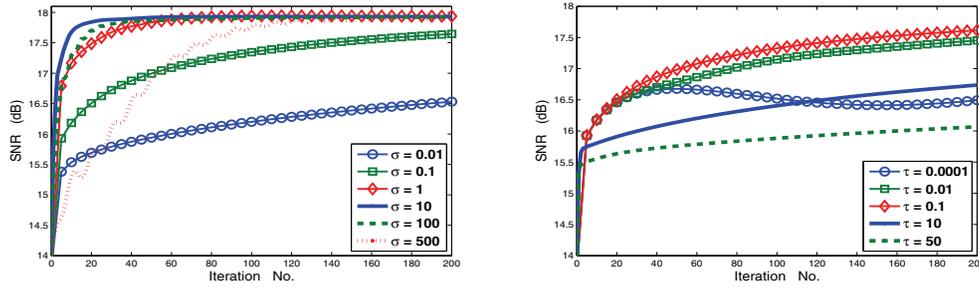


Figure 3. Evolutions of SNRs with respect to iterations for $\theta = -1$. Left: $\tau = 0.01$. Right: $\sigma = 0.1$.

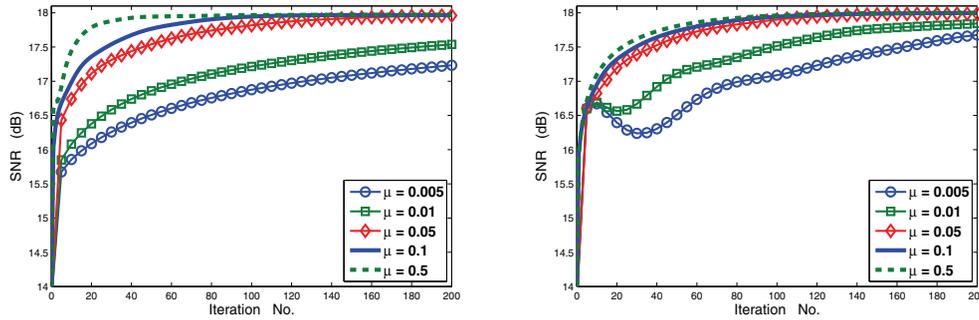


Figure 4. Evolutions of SNRs with respect to iterations for $\theta = 1$. Left: $\tau = 0.01$. Right: $\sigma = 0.1$.

7.1.3. $\theta = 1$. For this case, recall that the requirement on the step sizes τ and σ is (4.4). Thus, when τ is fixed, the allowable step size σ can be expressed as

$$\sigma = \frac{\mu}{\tau \|A^T A\|} \text{ with } \mu \in (0, 1),$$

and when σ is fixed, the allowable step size τ can be expressed as

$$\tau = \frac{\mu}{\sigma \|A^T A\|} \text{ with } \mu \in (0, 1).$$

In Figure 4, we show the variation of Algorithm 3’s numerical performance with respect to various choices of μ when τ is fixed as 0.01 and when σ is fixed as 0.1. Figure 4 shows clearly that larger step sizes are possible to improve the numerical performance of Algorithm 3.

Finally, we investigate the sensitivity of the parameter ρ for Algorithm 4. In Figure 5, for Algorithm 4, we plot the evolutions of SNR values with respect to 200 iterations for different values of ρ , when the step sizes are fixed. In this figure, $\rho \in (1, 2)$ means that its value is randomly generated in the interval $(1, 2)$. These curves in Figure 5 indicate that $\rho \in (1.5, 1.9)$ is preferable when we implement Algorithm 4 in practice.

7.2. TV image deblurring problem. In this subsection, we apply the proposed algorithms to solve some TV image deblurring problems and compare them numerically with the PDHG

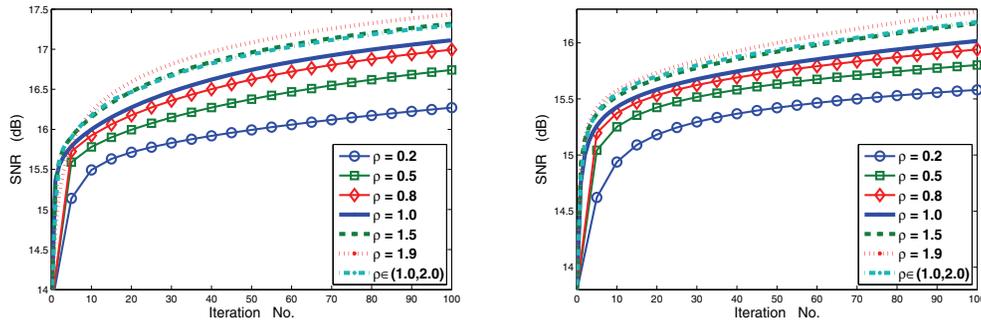


Figure 5. Evolutions of SNRs with respect to iterations for Algorithm 4 with different ρ . Left: ($\tau = 0.01, \sigma = 0.1$). Right: ($\tau = 0.1, \sigma = 0.01$)



Figure 6. Original images: *Barbara.png* (512×512) and *Man.png* (1024×1024).

Table 1

Tested scenarios in subsection 7.2.

| | motion blur | gaussian blur |
|-----------------|-------------------|--------------------|
| Medium scenario | theta=135, len=21 | hsize=21, sigma=5 |
| Severe scenario | theta=135, len=91 | hsize=41, sigma=10 |

method in [25]. More specifically, we test the images *Barbara.png* (512×512) and *Man.png* (1024×1024), as shown in Figure 6. These images are then degraded by convolutions and the zero-mean Gaussian noise with the standard deviation 10^{-3} . The blur operator and the additive noise are generated by the respective scripts `fspecial` and `imnoise` in the MATLAB Image Processing Toolbox, and the specific scenarios (i.e., the input of the script `fspecial`) we test are listed in Table 1. In Figures 7 and 8, we display the blurred images to be tested.

To deblur these corrupted images, in (1.1) we take $\lambda = 250$ for the motion blur cases and $\lambda = 1000$ for the gaussian blur cases. All the algorithms including Algorithms 1–4 and the PDHG use the stopping criterion

$$(7.1) \quad \frac{\|y^{k+1} - y^k\|}{\|y^{k+1}\|} < \text{Tol},$$

where $\{y^k\}$ is the sequence generated by one of the tested algorithms and Tol is the error tolerance. In (7.1), we take Tol = 10^{-4} for the motion blur cases and Tol = 5×10^{-5} for the gaussian blur cases. All the algorithms start their iterations with the degraded images. The



Figure 7. Degraded images. From left to right: Barbara with medium motion blur, Man with medium motion blur, Barbara with severe motion blur, Man with severe motion blur.



Figure 8. Degraded images. From left to right: Barbara with medium gaussian blur, Man with medium gaussian blur, Barbara with severe gaussian blur, Man with severe gaussian blur.

Table 2

Toned values of parameters for Algorithms 1–4 for image deblurring.

| | |
|-------------|--|
| Algorithm 1 | $\tau = 0.03, \sigma = 5.0, \theta = -0.2, \gamma = 1.6$ |
| Algorithm 2 | $\tau = 0.03, \sigma = 5.0, \theta = -0.2$ |
| Algorithm 3 | $\tau = 0.03, \sigma = 4.0, \theta = 1$ |
| Algorithm 4 | $\tau = 0.03, \sigma = 4.0, \theta = 1, \rho = 1.8$ |

code of PDHG with toned parameters was downloaded from <http://pages.cs.wisc.edu/~swright/GPUreconstruction/>. When the proposed Algorithms 1–4 are implemented, their respective parameters are given in Table 2.

We report the numerical results in Tables 3 and 4 for the motion blur cases and gaussian blur cases, respectively. These tables show that all the proposed algorithms can restore images with better quality (i.e., higher SNR values) than that restored by PDHG. For the cases with medium blurs, the proposed algorithms are all faster than PDHG. For the cases with severe blurs, the proposed Algorithm 4 is faster than PDHG, and Algorithms 1–3 are also very competitive with PDHG in terms of the restoration speed. According to Tables 3 and 4, it is easy to find that the computational cost per iteration of Algorithms 1–3 is larger than that of Algorithm 4. The reason is that Algorithms 1–3 require multiplications of matrices and vectors (which are computationally demanding for image deblurring problems) at their correction steps, while Algorithm 4 does not. In addition, we can see that Algorithm 1 requires fewer iterations than Algorithm 2 for achieving the same stopping criterion, with the same values of parameters and starting iterate. This is due to the fact that Algorithm 2 is the reduced version of Algorithm 1 without choosing appropriate step sizes at correction

Table 3*Numerical results of image deblurring (motion blur).*

| | Medium motion blur | | | | | | Severe motion blur | | | | | |
|-------|--------------------|-----|-------|-----|------|-------|--------------------|------|-------|-----|------|-------|
| | Barbara | | | Man | | | Barbara | | | Man | | |
| | It. | CPU | SNR | It. | CPU | SNR | It. | CPU | SNR | It. | CPU | SNR |
| Algo1 | 26 | 3.8 | 24.48 | 34 | 21.7 | 25.54 | 72 | 11.4 | 18.95 | 63 | 41.3 | 22.37 |
| Algo2 | 32 | 4.7 | 24.48 | 42 | 27.1 | 25.54 | 98 | 14.6 | 18.95 | 85 | 53.4 | 22.37 |
| Algo3 | 37 | 5.1 | 24.48 | 48 | 28.8 | 25.53 | 113 | 15.8 | 18.94 | 97 | 57.3 | 22.36 |
| Algo4 | 44 | 3.8 | 24.48 | 47 | 15.7 | 25.54 | 75 | 7.3 | 18.95 | 65 | 28.8 | 22.37 |
| PDHG | 94 | 8.1 | 21.77 | 100 | 40.3 | 23.34 | 150 | 13.7 | 16.59 | 171 | 67.2 | 17.49 |

Table 4*Numerical results of image deblurring (gaussian blur).*

| | Medium gaussian blur | | | | | | Severe gaussian blur | | | | | |
|-------|----------------------|------|-------|-----|-------|-------|----------------------|------|-------|-----|-------|-------|
| | Barbara | | | Man | | | Barbara | | | Man | | |
| | It. | CPU | SNR | It. | CPU | SNR | It. | CPU | SNR | It. | CPU | SNR |
| Algo1 | 82 | 12.2 | 17.26 | 85 | 55.1 | 18.85 | 136 | 18.8 | 16.54 | 139 | 94.5 | 17.12 |
| Algo2 | 100 | 14.6 | 17.26 | 106 | 69.2 | 18.85 | 172 | 21.4 | 16.54 | 177 | 116.9 | 17.12 |
| Algo3 | 109 | 15.2 | 17.26 | 115 | 69.6 | 18.86 | 187 | 22.9 | 16.54 | 194 | 120.0 | 17.12 |
| Algo4 | 84 | 8.5 | 17.26 | 89 | 41.5 | 18.85 | 146 | 12.7 | 16.54 | 146 | 70.5 | 17.12 |
| PDHG | 262 | 21.8 | 17.18 | 313 | 129.3 | 16.12 | 219 | 19.6 | 15.81 | 263 | 125.9 | 15.88 |

steps. Therefore, we can expect that the new iterate generated by Algorithm 1 is closer to the solution set than that generated by Algorithm 2. On the other hand, because of the alleviation of determining step size, the computational cost per iteration of Algorithm 2 is less than that of Algorithm 1.

To further visualize the numerical comparison between the proposed algorithms and the PDHG, in Figure 9 we plot their respective evolutions of SNR values with respect to iterations and computing time for the image Barbara.png. In Figures 10 and 11, we list the images restored by Algorithm 1 and the PDHG. Since the SNR values of the images restored by Algorithms 1–4 are almost the same, we list only the restored images by Algorithm 1 for the sake of succinctness. To see the difference in restoration clearly, we zoom some parts of the images (Barbara.png with the severe motion blur and Man.png with the severe gaussian blur) restored by PDHG and Algorithm 1 in Figure 12.

7.3. TV image inpainting problem. In this section, we apply the proposed algorithms to solve some TV image inpainting problems, and we compare them numerically with the PDHG method in [25] and the first-order primal-dual algorithm (denoted by CP) in [6]. Since Algorithm 4 outperforms other proposed algorithms based on Tables 3 and 4, among the proposed algorithms, we focus on the implementation of Algorithm 4 in this subsection.

We first review the image inpainting problem briefly. Image inpainting refers to filling in missing or damaged regions in images either in the pixel domain or in a transformed domain, and it plays pivotal roles in many image precessing tasks; see, e.g., [3, 8]. The image inpainting problem with TV regularization fits in the model (1.1) with $B \in \mathfrak{R}^{N \times N}$ as the mask operator, i.e., a diagonal matrix whose zero entries denote missed information, and identity entries indicate observed information.

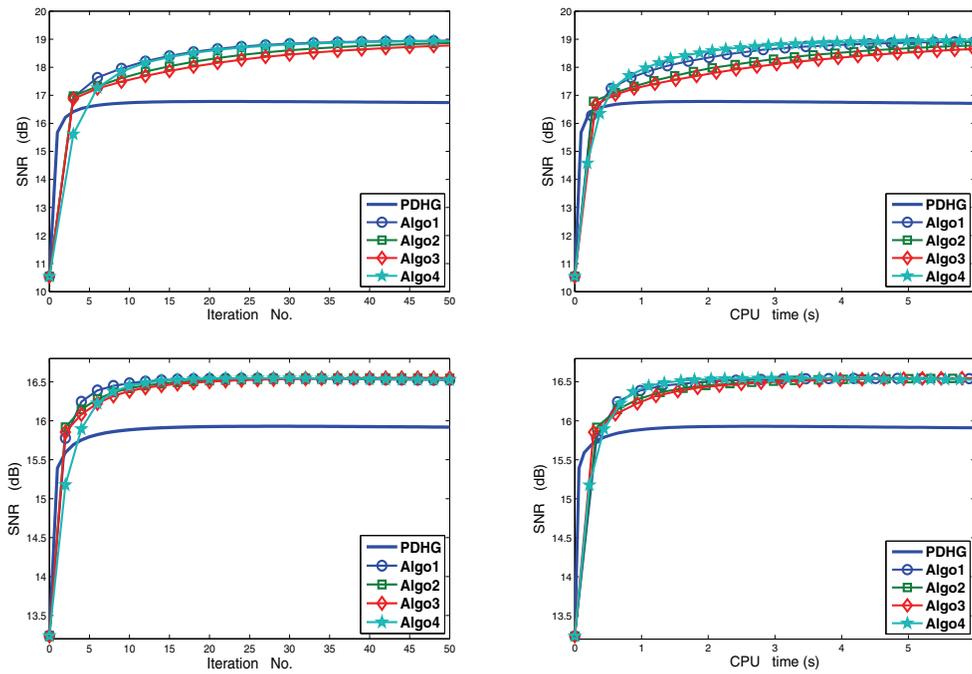


Figure 9. Evolutions of SNRs with respect to iterations and computing times for *Barbara.png*. First row: the scenario of severe motion blur. Second row: the scenario of severe gaussian blur.



Figure 10. Restored images by PDHG (first row) and Algorithm 1 (second row). From left to right: *Barbara* with medium motion blur, *Man* with medium motion blur, *Barbara* with severe motion blur, *Man* with severe motion blur.

We test the images *House.png* (256×256) and *Peppers.png* (512×512). These images are degraded as follows. For *House.png*, the operator B is a character mask where about 15% of pixels are missed, and for *Peppers.png*, the operator B is the mask where we retain the pixels at rows 1, 9, 17, ..., 503 (i.e., we retain only one row for every eight rows) and about 87% of

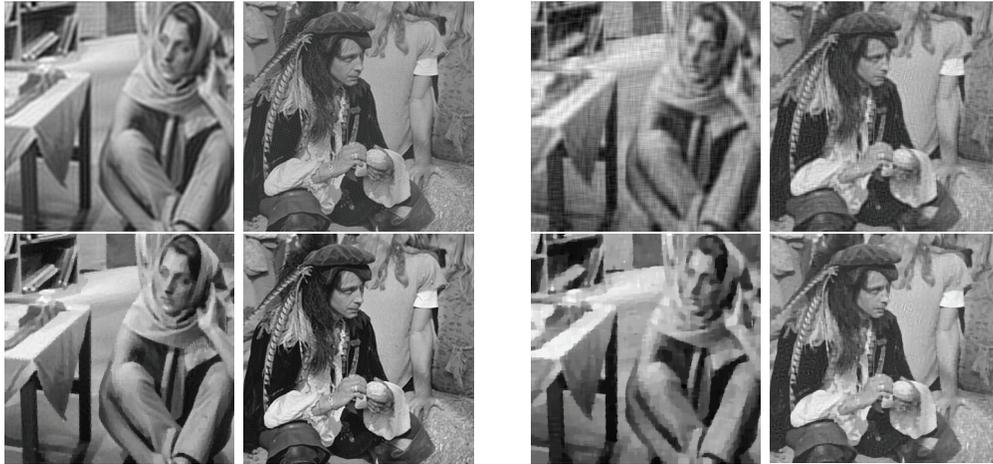


Figure 11. Restored images by PDHG (first row) and Algorithm 1 (second row). From left to right: Barbara with medium gaussian blur, Man with medium gaussian blur, Barbara with severe gaussian blur, Man with severe gaussian blur.



Figure 12. From left to right: original images, degraded images, zooming of the restored images by PDHG and by Algorithm 1.

pixels are missed. For both images, we add the zero-mean Gaussian noise with the standard deviation 0.02. The original and degraded images are shown in Figure 13.

To inpaint these images, we take $\lambda = 50$ in the model (1.1) for all methods. Since the efficient strategies of determining the step sizes τ and σ were proposed particularly for image deblurring problems in [25] and we found they are not efficient for the TV image inpainting problem under consideration, we choose invariant values for τ and σ when we apply PDHG to solve the TV image inpainting problem. More specifically, we tuned the values τ and σ via a number of numerical experiments, and we found that PDHG performs well for the TV image inpainting problem when $\tau\theta \approx 0.08$. For this reason, we choose $(\tau, \theta) = (8, 0.01)$ for PDHG. Similarly, we manually tuned the values of τ and σ for the CP method (i.e., Algorithm 1 in [6]),

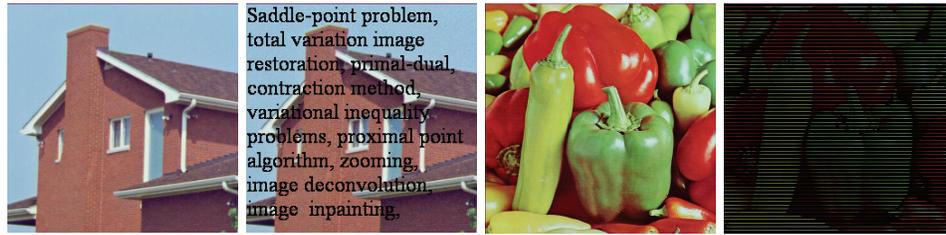


Figure 13. From left to right: original *House.png* (256×256), degraded *House*, original *Peppers.png* (512×512), degraded *Peppers*.

Table 5

Numerical results of image inpainting.

| | House | | | Peppers | | |
|-------------|-------|------|-------|---------|-------|-------|
| | It. | CPU | SNR | It. | CPU | SNR |
| Algorithm 4 | 48 | 1.44 | 28.80 | 114 | 13.29 | 16.56 |
| PDHG | 106 | 2.83 | 28.37 | 313 | 29.98 | 15.44 |
| CP | 57 | 1.51 | 28.52 | 169 | 17.32 | 15.73 |

and we found that it performs well with $(\tau, \sigma) = (0.02, \frac{1}{\tau \|A^T A\|})$ for the TV image inpainting problem under consideration. To implement Algorithm 4, we choose $\tau = 3.0$, $\sigma = 0.04$, and $\rho = 1.8$. All the algorithms start their iterations with the degraded images, and the stopping criterion is (7.1) with $\text{Tol} = 10^{-3}$.

In Table 5, we report the numerical performance of these three algorithms for the TV image inpainting problems. According to this table, we find that these three algorithms can restore the image of House with almost the same quality, and Algorithm 4 outperforms all the others for the image of Peppers (with the highest SNR value). In terms of restoration speed, Algorithm 4 and CP are much faster than PDHG. In Figure 14, we show the images restored by these three algorithms. To see the difference in restoration clearly, in Figure 15, we zoom some parts of the Peppers images restored by these three algorithms.

8. Conclusions. In this paper, we highlight the novelty of applying the analytic framework of contraction methods to analyze the convergence of primal-dual algorithms for a saddle-point problem which has particular applications in the area of TV image restoration. With the analytic framework of contraction-type methods, the conditions on the involved parameters of existing primal-dual algorithms are relaxed significantly, and the convergence analysis is simplified substantially. We also propose some efficient primal-dual-based contraction methods in the prediction-correction fashion and verify their efficiency numerically.

It should be mentioned that the proposed primal-dual-based contraction methods require the estimate of $\|A^T A\|$ to determine the involved step sizes when Armijo’s rule is too expensive to implement (e.g., TV image restoration problems). Fortunately, this estimate is easy for many applications in TV image restoration problems. For example, when A is the matrix representation of the discrete gradient operator, according to Gerschgorin’s theorem [14], we have that $\|A^T A\| \leq 8$.

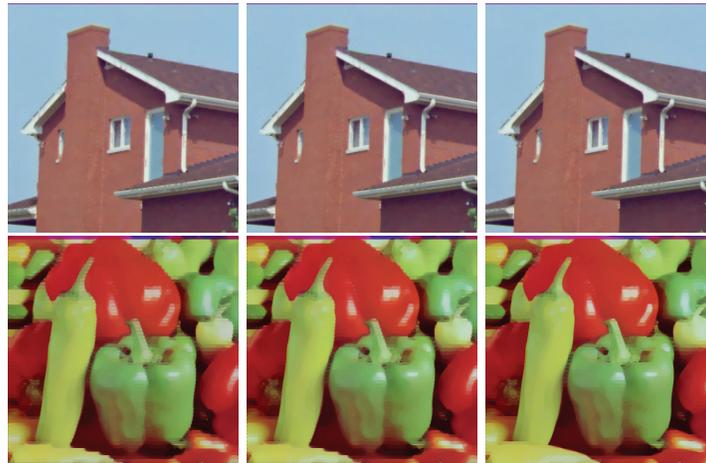


Figure 14. Restored images by PDHG (left column), CP (middle column), and Algorithm 4 (right column).

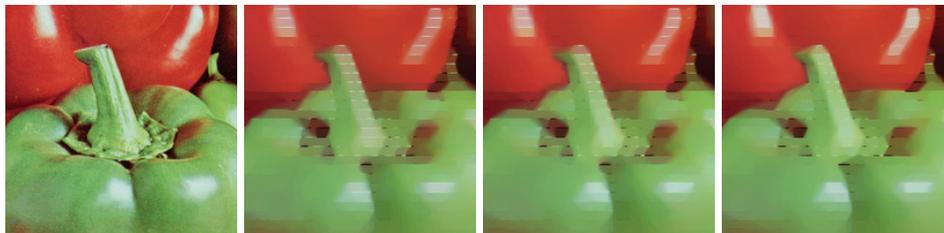


Figure 15. From left to right: zooming of the original image, the restored images by PDHG, CP, and Algorithm 4.

Acknowledgments. The authors are grateful to three anonymous referees for their valuable comments which have helped us substantially improve the presentation of this paper, and they would like to thank Wenxing Zhang for his help on the numerical experiments.

REFERENCES

- [1] K. J. ARROW, L. HURWICZ, AND H. UZAWA, *Studies in Linear and Non-Linear Programming*, Stanford Mathematical Studies in the Social Sciences, II, Stanford University Press, Stanford, CA, 1958.
- [2] H. H. BAUSCHKE AND P. L. COMBETTES, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces*, Math. Oper. Res., 26 (2001), pp. 248–264.
- [3] M. BERTALMIO, G. SAPIRO, V. CASELLES, AND C. BALLESTER, *Image inpainting*, in Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, ACM Press/Addison-Wesley, New York, 2000, pp. 417–424.
- [4] D. P. BERTSEKAS AND J. N. TSITSIKLIS, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
- [5] E. BLUM AND W. OETTLI, *Mathematische Optimierung*, Econometrics and Operations Research XX, Springer-Verlag, Berlin, New York, 1975.
- [6] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vison, 40 (2011), pp. 120–145.
- [7] T. F. CHAN, G. H. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.

- [8] T. F. CHAN AND J. SHEN, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, 2005.
- [9] P. L. COMBETTES, D. DÜNG, AND B. C. VŨ, *Dualization of signal recovery problems*, *Set-Valued Anal.*, 18 (2010), pp. 373–404.
- [10] J. DOUGLAS AND H. H. RACHFORD, *On the numerical solution of the heat conduction problem in 2 and 3 space variables*, *Trans. Amer. Math. Soc.*, 82 (1956), pp. 421–439.
- [11] E. ESSER, X. ZHANG, AND T. F. CHAN, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, *SIAM J. Imaging Sci.*, 3 (2010), pp. 1015–1046.
- [12] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.
- [13] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems via finite-element approximations*, *Comput. Math. Appl.*, 2 (1976), pp. 17–40.
- [14] S. GERSCHGORIN, *Über die Abgrenzung der Eigenwerte einer Matrix*, *Izv. Akad. Nauk. USSR, Ser. Fiz.-Mat. Nauk*, 7 (1931), pp. 749–754.
- [15] D. GOLDFARB AND W. YIN, *Second-order cone programming methods for total variation-based image restoration*, *SIAM J. Sci. Comput.*, 27 (2005), pp. 622–645.
- [16] B. S. HE, X. L. FU, AND Z. K. JIANG, *Proximal point algorithm using a linear proximal term*, *J. Optim. Theory Appl.*, 141 (2009), pp. 209–239.
- [17] B. S. HE, L. Z. LIAO, D. R. HAN, AND H. YANG, *A new inexact alternating directions method for monotone variational inequalities*, *Math. Program.*, 92 (2002), pp. 103–118.
- [18] G. M. KORPELEVIČ, *An extragradient method for finding saddle points and for other problems*, *Èkonom. i Mat. Metody*, 12 (1976), pp. 747–756.
- [19] B. MARTINET, *Régularisation d'inéquations variationnelles par approximations successives*, *Rev. Française Informat. Recherche Operationelle*, 4 (1970), pp. 154–158.
- [20] T. POCK, D. CREMERS, H. BISCHOF, AND A. CHAMBOLLE, *An algorithm for minimizing the Mumford-Shah functional*, in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2009.
- [21] L. D. POPOV, *A modification of the Arrow-Hurwitz method of search for saddle points*, *Mat. Zametki*, 28 (1980), pp. 777–784.
- [22] R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm*, *SIAM J. Control Optim.*, 14 (1976), pp. 877–898.
- [23] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, *Phys. D*, 60 (1992), pp. 227–238.
- [24] X. ZHANG, M. BURGER, AND S. OSHER, *A unified primal-dual algorithm framework based on Bregman iteration*, *J. Sci. Comput.*, 46 (2010), pp. 20–46.
- [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.