Convergence analysis of primal-dual algorithms for total variation image restoration

Bingsheng He[†1] and **Xiaoming Yuan**[‡2]

November 4, 2010

[†]Department of Mathematics and National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China

[‡]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

Abstract. Recently, some attractive primal-dual algorithms have been proposed for solving a saddle-point problem, with particular applications in the area of total variation (TV) image restoration. This paper focuses on the convergence analysis of existing primal-dual algorithms and shows that the involved parameters of those primal-dual algorithms (including the step sizes) can be significantly enlarged if some simple correction steps are supplemented. As a result, we present some primal-dual-based contraction methods for solving the saddle-point problem. These contraction methods are in the prediction-correction fashion in the sense that the predictor is generated by a primal-dual method and it is corrected by some simple correction step at each iteration. In addition, based on the context of contraction type methods, we provide a novel theoretical framework for analyzing the convergence of primal-dual algorithms which simplifies existing convergence analysis substantially.

Keywords. Saddle-point problem, total variation, primal-dual, contraction method, variational inequality, proximal point algorithm.

1 Introduction

We consider the saddle-point problem:

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

where $\mathcal{X}\subset\Re^L$, $\mathcal{Y}\subset\Re^N$, $z\in\Re^L$, $A,\ B\ \in\Re^{L\times N}$, $\lambda>0$, $\|\cdot\|$ denotes the Euclidean norm and T denotes the operator of inner product of vectors. An important application of (1.1) is the total variation (TV, see [20]) image restoration problems with blurry and

[1]This author was supported by the NSFC grant 10971095 and Doctoral Fund of Ministry of Education of China 20060284001. Email: hebma@nju.edu.cn

[2]Corresponding author. This author was supported by an Hong Kong General Research Grant. Email: xmyuan@hkbu.edu.hk

noisy image, see e.g. [6, 9, 22]. Note that we can consider the saddle-point problem in more general setting, for example, exactly as [6]:

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\ g(x)-\langle Ax,y\rangle-f^*(y) \tag{1.2}$$

where $\mathcal{X}\subset X$ and $\mathcal{Y}\subset Y$ are closed convex sets; X and Y are two finite-dimensional real vector spaces equipped with an inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|=\langle\cdot,\cdot\rangle^{1/2}$; $g: X\to[0,+\infty)$ and $f^*: Y\to[0,+\infty)$ are proper, convex, lower-semicontinuous (l.s.c.) functions; f^* is itself the convex conjugate of a convex l.s.c. function f; and $A: X\to Y$ is a continuous linear operator with induced norm

$$\|A\|=\max\{\|Ax\|: x\in X \text{ with } \|x\|\le 1\}.$$

For clearer exposure of our analysis and simpler presentation, we focus on (1.1) in the following analysis and we mention the extension to (1.2) at the end.

As analyzed in [6, 9], the saddle-point problem can be regarded as a primal-dual formulation of a nonlinear programming problem. In the literature of TV image restoration, some primal-dual algorithms have been developed and their efficiency has been well illustrated by comprehensive numerical experiments, see e.g. [6, 7, 9, 12, 17, 21, 22]. In particular, the iterative scheme of existing primal-dual algorithms for (1.1) can be summarized as the following procedure.

The primal-dual procedure for (1.1).

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

$$x^{k+1}=\operatorname{Arg}\max_{x\in\mathcal{X}}\big\{\tau\Phi(x,y^k)-\frac{1}{2}\|x-x^k\|^2\big\}, \tag{1.3a}$$

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

$$y^{k+1}=\operatorname{Arg}\min_{y\in\mathcal{Y}}\big\{\sigma\Phi(\bar{x}^k,y)+\frac{1}{2}\|y-y^k\|^2\big\}. \tag{1.3c}$$

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

- When $\theta=0$ in (1.3b), the above primal-dual procedure reduces to the primal-dual hybrid gradient method in [22], which essentially a Arrow-Hurwicz algorithm [1]. In [9], this primal-dual hybrid gradient method was explained as an inexact Uzawa method in [9], and its convergence was analyzed under some conditions which essentially ensure that the step sizes τ and σ are sufficiently 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 [15, 18], the Douglas

Rachford splitting method [8] and the alternating direction method of multipliers [10]. When $\theta = 1$, convergence of the primal-dual algorithm in [6] was proved with the following requirement on step sizes:

$$\tau\sigma < \frac{1}{\|A^T A\|}, \tag{1.4}$$

Thus, the difficulty of choosing very small step sizes in [9] is overcame. Note that under some additional regularity and further convexity assumptions on (1.2), some sophisticated strategies for choosing the parameters (τ, σ and θ) dynamically have also been analyzed in [6] in order to yield accelerated primal-dual algorithms. In this paper, we restrict our discussion into the deterministic scenario where all the parameters (τ, σ and θ) are fixed throughout iterations.

In this paper, we study primal-dual algorithms for (1.1) from the perspective of contraction type methods (see [4] and Section 2.4). We will show that the primal-dual procedure (1.3) is not a contraction method (in terms of the definition of [4]) with respect to the solution set of (1.1) unless $\theta = 1$. However, if the iterates generated by the primal-dual procedure (1.3) are corrected by some simple correction steps, then the resulting iterates constitute a sequence of a contraction method for (1.1). Some primal-dual-based contraction algorithms are thus presented for (1.1) in the prediction-correction fashion, where the primal-dual procedure (1.3) produces a predictor and it is corrected by a simple correction step at each iteration. Our contributions of this paper are:

1). We show that the range of the combination parameter θ can be enlarged to $[-1, 1]$ which is boarder than the result in [6].

2). When $\theta = -1$, the step size τ and σ can be arbitrarily positive numbers. When $\theta \in (-1, 1]$, the condition on step sizes (1.4) can be relaxed to:

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

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

3). We study the convergence of primal-dual algorithms for (1.1) under the analytic framework of contraction type methods, and thus provide a novel theoretical framework of convergence analysis which differs from the existing methodologies in the literature of primal-dual algorithms for (1.1). This new framework simplifies the proof of primal-dual type methods substantially.

The rest of this paper is organized as follows. In Section 2, we review some preliminaries which are useful for our analysis. In Section 2.4, for the case $\theta \in [-1, 1)$, we present the 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 of primal-dual algorithms are also presented in this section. In Section 4, we present a reduced primal-dual-based contraction method with simpler correction steps. The corresponding requirement on step sizes of the primal-dual procedure is also analyzed. In Section 5, we pay specific attention to the case of $\theta = 1$. We will develop a new primal-dual-based contraction method 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. Finally, some conclusions are made in Section 7.

2 Preliminaries

In this section, we provide some preliminaries which are useful for subsequent analysis. In particular, we review some basic knowledge of variational inequalities (VI), the proximal point algorithm (PPA) and contraction methods, which are cornerstones of our analysis.

2.1 The VI reformulation of (1.1)

We first reformulate (1.1) into a VI. Let (x^*, y^*) be a solution of the saddle-point problem (1.1). Then we have

$$\max_{x\in\mathcal{X}}\Big\{x^T Ay^* + \frac{\lambda}{2}\|By^* - z\|^2\Big\} \le (x^*)^T Ay^* + \frac{\lambda}{2}\|By^* - z\|^2 \le \min_{y\in\mathcal{Y}}\Big\{(x^*)^T Ay + \frac{\lambda}{2}\|By - z\|^2\Big\}.$$

Based on the optimality conditions of the last problems, we can easily verify that the above facts can be characterized by the following variational inequalities:

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

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

$$\mathrm{VI}(\Omega, F): \qquad (u - u^*)^T F(u^*) \ge 0, \quad \forall u \in \Omega, \tag{2.1a}$$

where

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

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

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

Therefore, the VI (2.1a) is monotone and the solution set denoted by Ω^* is not empty. Moreover, let the projection onto Ω under the Euclidean norm be denoted by $P_\Omega(\cdot)$, i.e.,

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

Then, the following lemmas shows that solving the VI (2.1a) amounts to solving a projection equation.

Lemma 2.1. *u^* is a solution of VI(Ω, F) if and only if*

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

Proof. See ([3], p. 267). □

2.2 Proximal point algorithm

Among classical methods for solving the equivalent VI (2.1) is the influential proximal point algorithm (PPA) which was contributed originally in [16] and developed concretely in [19]. More specifically, instead of solving (2.1) directly, PPA solves the following proximal subproblem to generate the new iterate u^{k+1} with the given iterate u^k:

$$\text{(PPA)} \qquad u \in \Omega, \qquad (u' - u^{k+1})^T (F(u^{k+1}) + r(u^{k+1} - u^k)) \geq 0, \qquad \forall\, u' \in \Omega, \quad (2.2)$$

where $r > 0$ is 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

$$u \in \Omega, \qquad (u' - u^{k+1})^T (F(u^{k+1}) + G(u^{k+1} - u^k)) \geq 0, \qquad \forall\, u' \in \Omega. \qquad (2.3)$$

Note that VI (2.3) can be regarded as the PPA with preconditioning proximal terms (see e.g.[9]) or the PPA in the context of G-norm (defined by $\|u\|_G = \sqrt{u^T G u}$) [9, 14]. Based on the convergence analysis of PPA [16, 19], the sequence $\{u^k\}$ generated by (2.2) or (2.3) converges to a solution of (2.1).

2.3 The PPA structure of (1.3)

In this subsection, we show that the primal-dual procedure (1.3) actually takes an analogous structure of the PPA (2.3), but the involved precondition matrix is not symmetric.

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

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

where

$$M = \begin{pmatrix} \frac{1}{\tau} I & A^T \\ \theta A & \frac{1}{\sigma} I \end{pmatrix}_{(L+N)\times(L+N)}. \qquad (2.5)$$

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

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

and

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

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

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

By using the function F defined in (1.3b) and the matrix M defined in (2.5), (2.4) is a compact form of the above VI and the lemma is proved. □

Therefore, unless $\theta = 1$, the matrix M in (2.4) is not symmetric. Thus, (2.4) is not exactly the PPA (2.3) even though it takes the PPA structure. In this sense, the iterative scheme (2.4) can be regarded as a variant of PPA (VPPA) or a PPA with linear proximal term [13].

2.4 Contraction methods

We first state the concept of Fejèr monotone for the VI (Ω, F). For a sequence $\{u^k\}$, if the following property is satisfied:

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

then we say that the sequence $\{u^K\}$ is Fejèr monotone with respect to Ω^*. We refer to [2] for more properties of Fejèr monotonicity.

According to [4], if a method generates a iterative sequence $\{u^k\}$ satisfying the following property:

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

where $c > 0$ is a constant. Then, we call this method a contraction method for VI (Ω, F). Obviously, the sequence generated by a contraction method for VI (Ω, F) is Fejèr monotone with respect to Ω^*. Let $G \in \Re^{(L+N)\times(L+N)}$ be a symmetric positive definite matrix. If the G-norm is considered, then the inequality (2.8) in the definition of contraction methods can be altered to

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

Then, 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 sequence generated by the primal-dual procedure (1.3) (with $\theta \neq 1$) is not Fejèr monotone with respect to Ω^* because of the lack of symmetry of M. Therefore, the existing primal-dual algorithms for (1.1) (except for $\theta = 1$) do not belong to the category of contraction methods. This observation inspires us to seek a contraction method for (1.1) based on the primal-dual procedure (1.3), and the principle advantage for doing so is that the requirement on parameters can be significantly relaxed.

3 A primal-dual-based contraction method for $\theta \in [-1, 1)$

As we have observed, due to the lack of the symmetry of M, the primal-dual procedure (1.3) (unless $\theta = 1$) is not in the nature of contraction methods even though it takes an analogous structure of PPA (2.4). In the following, we will show that whenever M is positive definite, we can easily develop some simple correction steps to correct the iterates generated by the primal-dual procedure (1.3), and the corrected iterates constitute a sequence of a contraction method for (1.1). As a result, we can develop a primal-dual-based contraction method for (1.1) in the prediction-correction fashion where the predictor is generated by the primal-dual procedure (1.3) (or (2.4)) and it is corrected by some correction step at each iteration. We will show that the convergence of the resulting primal-dual contraction method can be easily derived under the analytic framework of contraction type methods. We will also show that the purpose of ensuring the positive definiteness of M entails the requirement on the step sizes τ and σ, and makes it possible to relax the requirements on τ and σ in the existing literature [6, 22].

We first restrict our discussion into the case that $\theta \in [-1, 1)$ in Sections 3 and 4; and then we will analyze the case of $\theta = 1$ separately in Section 5.

3.1 The prediction step

For the convenience of presenting 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 process (1.3). Therefore, the primal-dual-procedure (1.3) (i.e. (2.4)) can be re-described as follows.

The prediction step at the k-th iteration.

Let $\theta \in [-1, 1)$; F be defined in (2.1b) and M be defined in (2.5). With given u^k, apply the primal-dual procedure (1.3) to generate the predictor $\tilde{u}^k \in \Omega$:

$$(u - \tilde{u}^k)^T \{F(\tilde{u}^k) + M(\tilde{u}^k - u^k)\} \geq 0, \quad \forall\, u \in \Omega. \tag{3.1}$$

In the following, we discuss how to realize the primal-dual procedure (1.3) for some TV image restoration models. For these applications, note that in (1.1), $\mathcal{X}$ is the Descartes product of some unit balls in $\Re^2$, $\mathcal{Y}$ is a ball in $\Re^N$ with certain (or infinite) radius, A is corresponding to the discrete gradient operator of the total variation term, B is the matrix representing denoising or deblurring operators, and z is a the vector of a given image. We refer to [20] for the TV image restoration models and [6, 22] for how to reformulate these TV image restoration models into 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

$$\tilde{x}^k = P_{\mathcal{X}}(x^k + \tau A^T y^k). \tag{3.2}$$

Since $\mathcal{X}$ is the Descartes product of some unit balls under the Euclidean-norm in $\Re^2$, the projection on $\mathcal{X}$ is easy to be carried out. Thus, the subproblem (1.3a) is easy to be realized for TV image restoration problems.

Now, we elaborate on how to realize the subproblem (1.3c) for different TV image restoration models.

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

$$\min_u \int_D |\nabla u| \quad \text{s.t.} \quad \|u - f\|_2^2 \leq |D|\sigma^2, \tag{3.3}$$

  where D is the image domain with its area being $|D|$, f is the given observed image and σ^2 is an estimation of the variance of the noise in the image f. Note that (3.3) can be reformulated into the saddle-point problem:

$$\min_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} \Phi_1(x, y) := y^T A x, \tag{3.4}$$

  where $\mathcal{Y}$ is a ball in $\Re^N$ with certain radius. In this case, the solution of (1.3c) is given by

$$\tilde{y}^k = P_{\mathcal{Y}}\big(y^k - \sigma A \bar{x}^k\big), \tag{3.5}$$

  and the projection on $\mathcal{Y}$ can be carried out easily.

- `The unconstrained ROF model`. The discrete unconstrained ROF model is:

$$\min_u \int_D |\nabla u| + \frac{\lambda}{2}\|u - f\|_2^2, \tag{3.6}$$

and it can be reformulated into the saddle-point problem (1.1):

$$\min_{y\in\Re^N} \max_{x\in\mathcal{X}} \Phi_2(x,y) := y^T Ax + \frac{\lambda}{2}\|y - z\|^2. \tag{3.7}$$

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

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

Thus, $\tilde{y}^k$ is given explicitly by:

$$\tilde{y}^k = \frac{1}{1+\sigma\lambda} y^k + \frac{\sigma}{1+\sigma\lambda}\big(\lambda z - A\bar{x}^k\big). \tag{3.8}$$

- `TV deblurring model`. The discrete blurry and noisy image model is:

$$\min_u \int_D |\nabla u| + \frac{\lambda}{2}\|Ku - f\|_2^2, \tag{3.9}$$

where K is a given linear blurring operator and every other term is defined the same as before. Then, (3.9) can be reformulated into the saddle-point problem (1.1):

$$\min_{y\in\Re^N} \max_{x\in\mathcal{X}} \Phi_3(x,y) := y^T Ax + \frac{\lambda}{2}\|By - z\|^2. \tag{3.10}$$

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

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

In other words, $\tilde{y}^k$ is the unique vector which satisfies

$$(\tilde{y}^k - y^k) + \sigma\big(\lambda B^T(B\tilde{y}^k - z) + A\bar{x}^k\big) = 0. \tag{3.11}$$

Since B is the matrix representation of a convolution operator K, the Fourier transform of matrix multiplication by B becomes point-wise multiplication in the frequency domain. Hence (3.11) can be efficiently solved by FFT and inverse FFT:

$$\tilde{y}^k = \mathcal{F}^{-1}\Big[\frac{\mathcal{F}(y^k - \sigma A\bar{x}^k) + \sigma\lambda\mathcal{F}(K)^* \odot \mathcal{F}(z)}{1 + \sigma\lambda\mathcal{F}(K)^* \odot \mathcal{F}(K)}\Big], \tag{3.12}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}$ are FFT and inverse FFT operator, respectively, * denotes the complex conjugate and $\odot$ is the point-wise multiplication operator. For details, see § 2.4.2 in [22].

To end this subsection, we point out that the primal-dual procedure (1.3) is also easily implementable for some TV-L^1 models. For given y, since

$$\|A^T y\|_2 = \max\{x^T A^T y : \|x\|_2 \le 1\}$$

and

$$\|A^T y\|_1 = \max\{x^T A^T y : \|x\|_\infty \le 1\},$$

the set $\mathcal{X}$ is changed to the Descartes product of some unit boxes (ball in max-norm) in $\Re^2$. The projection on such $\mathcal{X}$ is also very easy to be carried out.

3.2 The correction step

Recall that the matrix $M \in \Re^{(L+N)\times(L+N)}$ (not necessarily symmetric) is positive definite if and only if there exists a constant $c_0 > 0$, such that

$$(u - \tilde{u})^T M(u - \tilde{u}) \ge c_0\|u - \tilde{u}\|^2, \quad \forall\, u \ne \tilde{u}. \tag{3.13}$$

As we have mentioned, the matrix M (2.5) is asymmetric unless $\theta = 1$. Thus, the primal-dual procedure (1.3) with $\theta \in [-1, 1)$ is not a contraction method. Our idea of constructing appropriate correction steps and thus yielding primal-dual-based contraction methods for (1.1) can be explained as follows. Because $u^* \in \Omega$, it follows from (3.1) that

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

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

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

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

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

From the last inequality follows that

$$(u^k - u^*)^T M(u^k - \tilde{u}^k) \ge (u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k), \quad \forall\, u^* \in \Omega^*. \tag{3.14}$$

Therefore, whenever M is positive definite, i.e., (3.13) is satisfied, we have that

$$\langle H(u^k - u^*), -H^{-1}M(u^k - \tilde{u}^k)\rangle \le -c_0\|u^k - \tilde{u}^k\|^2,$$

where $H \in \Re^{(L+N)\times(L+N)}$ is an arbitrary positive definite matrix. In this case, we conclude that $-H^{-1}M(u^k - \tilde{u}^k)$ is a descent direction of the unknown distance function $\frac{1}{2}\|u - u^*\|_H^2$ at the point u^k. Hence, we can correct the predictor $\tilde{u}^k$ by the following descent step and thus generate the new iterate u^{k+1}:

The correction step at the k-th iteration.

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

where $\alpha > 0$ is a step size along the descent direction to be specified, $H \in \Re^{(L+N)\times(L+N)}$ is a positive definite matrix, and $\tilde{u}^k$ is generated by the prediction step (3.1).

3.3 How to ensure the positive definiteness of M

Subsections 3.1 and 3.2 delineate how to construct a primal-dual-based contraction method for the case $\theta \in [-1, 1)$ under the condition that the matrix M (2.5) is positive definite. As we can see, the positive definiteness of M relies on values of the step sizes τ and σ of the prediction step when the combination parameter θ is given. In fact, the purpose of ensuring the positive definiteness of M entails the requirement on the step sizes τ and σ.

In this subsection, we focus on how to determine the step sizes τ and σ to ensure the positive definiteness of M when $\theta \in [-1, 1)$. For analysis convenience, we define the block diagonal part of M by

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

We divide the analysis of this subsection into the case $\theta = -1$ and $\theta = (-1, 1)$.

3.3.1 The case $\theta = -1$

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

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

In other words, M is the sum of H (3.15) and a skew-symmetric matrix. In this case, for any $\tau, \sigma > 0$, we have

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

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

3.3.2 The case $\theta \in (-1, 1)$

When $\theta \in (-1, 1)$, the condition on τ and σ to ensure positive definiteness of M can be summarized as 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 of the primal-dual procedure* (1.3) *satisfy:*

$$\tau\sigma \frac{(1+\theta)^2}{4} \|A^T A\| < 1. \tag{3.16}$$

Then, for the matrix M in (2.5), *we have*

$$(u - \tilde{u})^T M (u - \tilde{u}) \geq \frac{\delta}{1+\delta} \|u - \tilde{u}\|_H^2, \quad \forall u \neq \tilde{u}, \tag{3.17}$$

where

$$\delta = \frac{2}{1+\theta} \sqrt{\frac{1}{\tau\sigma\|A^T A\|}} - 1. \tag{3.18}$$

Proof. Under the condition (3.16), the scalar δ defined in (3.18) is positive and it holds that

$$\tau(1+\delta)\|A^TA\|\frac{(1+\theta)^2}{4} = \frac{1}{\sigma(1+\delta)}. \tag{3.19}$$

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

$$(u-\tilde{u})^T M(u-\tilde{u}) = \|u-\tilde{u}\|_H^2 + (1+\theta)(y-\tilde{y})^T A(x-\tilde{x}). \tag{3.20}$$

By using the Cauchy-Schwarz Inequality and (3.19), we get

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

Substituting it in (3.20), the assertion (3.17) is proved. □

We summarize the requirement on the step sizes τ and σ of the primal-dual procedure (1.3) as follows.

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

$$\begin{cases} \tau, \sigma \text{ are any positive numbers,} & \text{if } \theta = -1; \\ \tau\sigma\frac{(1+\theta)^2}{4}\|A^TA\| < 1, & \text{if } \theta \in (-1,1). \end{cases} \tag{3.21}$$

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 (3.21). In fact, τ and σ can be arbitrarily large when $\theta = -1$; and they can be arbitrarily large simultaneously if we take θ is very close to -1. Hence, the requirement on the step sizes τ and σ in [6] is significantly relaxed by our proposed primal-dual-based contraction method.

Remark 3.3. Note that the condition (3.16) 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:

$$(u^k-\tilde{u}^k)^T M(u^k-\tilde{u}^k) \geq \frac{\delta}{1+\delta}\|u^k-\tilde{u}^k\|_H^2, \quad \forall\, k>0, \tag{3.22}$$

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 (3.22) in the absence of the value of $\|A^TA\|$.

3.4 Step size of the correction step

Now, we analyze how to determine the step size α of the correction step. 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^*. For discussing how to determine a reasonable step length α, we let

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

Further, we denote

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

which can measure the progress of the proximity to the solution set Ω^* made by the k-th iteration. Because $\vartheta_k(\alpha)$ involves the unknown vector u^*, we cannot maximize it directly. However, with (3.14), we have that

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

By defining

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

the last inequality tells us that $q_k(\alpha)$ is a lower bound of $\vartheta_k(\alpha)$. Note that $q_k(\alpha)$ is a quadratic function of α, it 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 can choose α_k^* defined above as the step size of the correction step.

3.5 Algorithm

With the specifications of the prediction and correction steps, and their step sizes, we are now ready to present the primal-dual-based contraction method for (1.1) as follows.

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

Step 0. Let $\gamma \in (0, 2)$. Let $\theta \in [-1, 1)$ and H be defined in (3.15). Take $u^0 \in \Re^l$. Choose the step sizes τ and σ according to (3.21).

Prediction Step: Generate the predictor $\tilde{u}^k$ via solving (3.1).

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

$$u^{k+1} = u^k - \alpha_k H^{-1}M(u^k - \tilde{u}^k), \tag{3.23a}$$

where

$$\alpha_k = \gamma\alpha_k^*, \qquad \alpha_k^* = \frac{(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k)}{\|H^{-1}M(u^k - \tilde{u}^k)\|_H^2}. \tag{3.23b}$$

Remark 3.4. Since some inequalities are used in the analysis, the 'optimal' step size α_k is usually conservative for contraction methods. Thus, we attach a relaxation parameter γ to the step size α_k^*. The reason why we restrict $\gamma \in (0, 2)$ will be clear in the next subsection (see Lemma 3.6).

Remark 3.5. Compare to the primal-dual procedure (1.3), the correction step can be realized via much little computation. Thus, the price of relaxing the step sizes τ and σ, and the combination parameter θ, is cheap when the proposed primal-dual-based contraction method is implemented.

3.6 Convergence

In this subsection, we show that the proposed primal-dual-based contraction method is a contraction method in terms of the definition (2.9) and then we prove its convergence under the analytic framework of contraction type methods.

Lemma 3.6. *Let the matrix M be defined in (2.5) and H be defined in (3.15). Then, the sequence $\{u^k\}$ generated by the proposed Algorithm 1 satisfies:*

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

Proof. First, we notice that (3.13) is satisfied. Then, using (3.23) and (3.14), we have that

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

Substituting (3.13) in the last inequality, the assertion is proved. □

Based on Lemma 3.6, we can easily show that the proposed Algorithm 1 is a contraction method for (1.1), as stated in the following lemma.

Lemma 3.7. *The proposed Algorithm 1 is a contraction method for (1.1).*

Proof. First, under the condition (3.13), the 'optimal' step size (see (3.23b)) 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, from Lemma 3.6, we get

$$\|u^{k+1} - u^*\|_H^2 \leq \|u^k - u^*\|_H^2 - \frac{\gamma(2-\gamma)c_0^2}{\|M^T H^{-1} M\|} \|u^k - \tilde{u}^k\|^2, \quad \forall\, u^* \in \Omega^*. \tag{3.24}$$

Therefore, according to the definition (2.9), the proposed Algorithm 1 is a contraction method for (1.1) under the H-norm whenever $\gamma \in (0, 2)$. □

Now, because of Lemma 3.7, we can easily prove the convergence of the proposed Algorithm 1 under the analytic framework of contraction type methods in the following theorem.

Theorem 3.8. *The sequence generated by the proposed Algorithm 1 converges to a solution point of (1.1).*

Proof. According to (3.24), the sequence $\{u^k\}$ generated by the proposed Algorithm 1 is bounded. In fact, the sequence is contained in the compact set

$$S := \{u \in \Re^l \mid \|u - u^*\|_H \le \|u^0 - u^*\|_H\},$$

Thus, there exists a cluster point of $\{u^k\}$, denoted by u^∞. We assume that the subsequence $\{u^{k_j}\}$ converges to u^∞. Then, based on standard techniques of Fejèr monotonicity (see e.g.[2]), it is easy to show that u^∞ is a solution of (1.1) and u^∞ is the unique cluster point of the sequence $\{u^k\}$. Thus, this theorem is proved. □

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 (3.23b) can be reduced. Thus, a reduced primal-dual-based contraction method for (1.1) is proposed. With the easier treatment of correction steps, however, we will show that the ranges of the step sizes τ and σ of the prediction step are narrower.

Instead of choosing the step length α_k in (3.23a) judiciously, we can simply take $\alpha_k \equiv 1$ and thus the correction step (3.23a) is modified to:

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

The resulting primal-dual-based contraction method with the correction step (4.1) is called a reduced primal-dual-based contraction method. The following theorem indicates that the reduced primal-dual-based contraction method is also a contraction method if some mild assumptions are assumed.

Lemma 4.1. *Assume that the matrix M (2.5) is positive definite. Let H be defined in (3.15). If the following condition*

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

is hold for all $k > 0$ where c_1 is a positive number. Then, the reduced primal-dual-based contraction method with (4.1) is a contraction method in the sense that:

$$\|u^{k+1} - u^*\|_H^2 \le \|u^k - u^*\|_H^2 - c_1\|u^k - \tilde{u}^k\|_H^2, \quad \forall\, u^* \in \Omega^*. \tag{4.3}$$

Proof. Indeed, the condition (4.2) guarantees that the 'optimal' step length α_k^* in (3.23b) is greater than $\frac{1}{2}$. Using (4.1) and (3.14), we have that

$$\begin{aligned}
\|u^{k+1} - u^*\|_H^2 &= \|(u^k - u^*) - H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \\
&= \|u^k - u^*\|_H^2 - 2(u^k - u^*)^T M(u^k - \tilde{u}^k) + \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \\
&\le \|u^k - u^*\|_H^2 - 2(u^k - \tilde{u}^k)^T M(u^k - \tilde{u}^k) + \|H^{-1}M(u^k - \tilde{u}^k)\|_H^2 \\
&\le \|u^k - u^*\|_H^2 - c_1\|u^k - \tilde{u}^k\|_H^2.
\end{aligned}$$

The assertion of this lemma is proved. $\square$

Now, we analyze how to determine values of the step sizes τ and σ of the primal-dual procedure (1.3) to guarantee the validation of (4.2).

Lemma 4.2. *Let H be as defined in* (3.15)*. For $\theta \in [-1, 1)$, let the step sizes τ and σ of the primal-dual procedure* (1.3) *satisfy:*

$$\tau\sigma\|A^T A\| < 1. \tag{4.4}$$

Then, for the matrix M defined in (2.5)*, we have*

$$2(u - \tilde{u})^T M(u - \tilde{u}) - \|H^{-1}M(u - \tilde{u})\|_H^2 \geq \delta\|u - \tilde{u}\|_H^2, \tag{4.5}$$

where

$$\delta = 1 - \tau\sigma\|A^T A\|.$$

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

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

and

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

Therefore, by a manipulation, we obtain

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

and

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

It follows from (4.6) that

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

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

Remark 4.3. According to Lemma 4.2, the condition (4.4) is a sufficient condition to ensure (4.5). In fact, from the proof of Lemma 4.6, we can see that we only need to guarantee that the right-hand side of (4.6) is greater than

$$\big(1 - \tau\sigma\|A^T A\|\big)\|u - \tilde{u}\|_H^2,$$

in order to ensure (4.5). In other words, the equality (4.6) implies that (4.5) holds if the step sizes τ and σ are chosen to guarantee

$$\begin{aligned} &\tau\|A^T(y^k - \tilde{y}^k)\|^2 + \sigma\theta^2\|A(x^k - \tilde{x}^k)\|^2 \\ \leq\ & (1-\delta)\Big(\frac{1}{\tau}\|x^k - \tilde{x}^k\|^2 + \frac{1}{\sigma}\|y^k - \tilde{y}^k\|^2\Big),\ \forall\, k > 0, \end{aligned} \tag{4.7}$$

for a scalar $\delta > 0$. In practical computation, we can use the classical Armijo's 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 method in [22] 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:

$$\tau\|A^T(y^k - \tilde{y}^k)\|^2 \leq (1-\delta)\Big(\frac{1}{\tau}\|x^k - \tilde{x}^k\|^2 + \frac{1}{\sigma}\|y^k - \tilde{y}^k\|^2\Big),\ \ \forall\, k > 0.$$

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

Algorithm 2: A reduced primal-dual-based contraction method for (1.1) with $\theta \in [-1, 1)$.
Step 0. Let $\theta \in [-1, 1)$ and H be defined in (3.15). Take $u^0 \in \Re^l$. Choose the step sizes τ and σ according to (4.4).
Prediction Step: Generate the predictor $\tilde{u}^k$ via solving (3.1).
Correction Step: Correct the predictor and generate the new iterate u^{k+1} via:

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

Remark 4.5. Since Lemma 4.1 shows that the proposed Algorithm 2 is a contraction method for (1.1), the convergence analysis of Algorithm 2 is similar to that of Algorithm 1. Thus, it is omitted.

To end this subsection, we discuss the relationship of the proposed Algorithm 2 with the primal-dual hybrid gradient algorithm in [22]. Recall that the primal-dual hybrid gradient algorithm in [22] is the special case of the primal-dual procedure (1.3) with $\theta = 0$ and its convergence was proved in [9] with very small step sizes τ and σ. For the case $\theta = 0$, the primal-dual procedure (1.3) reduces to:

$$\tilde{x}^k = \text{Arg}\max_{x\in\mathcal{X}} \big\{\tau\Phi(x, y^k) - \frac{1}{2}\|x - x^k\|^2\big\}, \tag{4.9a}$$

$$\tilde{y}^k = \text{Arg}\min_{y\in\mathcal{Y}} \big\{\sigma\Phi(\tilde{x}^k, y) + \frac{1}{2}\|y - y^k\|^2\big\}. \tag{4.9b}$$

Then, the primal-dual hybrid gradient algorithm in [22] 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 primal-dual hybrid gradient algorithm in [22] in that the variable $\tilde{x}^k$ is further corrected. Recall that the primal-dual hybrid gradient algorithm in [22] is essentially the 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 case that $\theta = 1$ in the primal-dual procedure (1.3). As we have emphasized, for this case, the matrix M (2.5) becomes:

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

which is symmetric. Thus, the primal-dual procedure (1.3) with the matrix M in (5.1) 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 steps is a contraction method with respect to the solution set Ω^*. We first investigate the condition to ensure the positive definiteness of the matrix M in (5.1). Similar as Lemma 3.1, we can easily to show that the positive definiteness of the matrix M in (5.1) is guaranteed if

$$\tau\sigma\|A^T A\| < 1. \tag{5.2}$$

Thus, with the restriction (5.2) on 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 the condition (3.16) (with $\theta = 1$) and the condition (4.4). All the analysis carried out in Sections 3 and 4 is valid for $\theta = 1$. Thus, we can easily extend the proposed Algorithms 1 and 2 to the case with $\theta = 1$. In particular, a resulting primal-dual-based contraction method for the case $\theta = 1$ thus can be described as follows.

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

Step 0. Let $\theta = 1$. Take $u^0 \in \Re^l$. Choose the step sizes τ and σ according to (5.2).
Prediction Step: Generate the predictor $\tilde{u}^k$ via solving (3.1).
Correction Step: Correct the predictor and generate the new iterate u^{k+1} via:

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

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 $\theta = 1$, in addition to Algorithm 3. The new primal-dual-based contraction method is also in the prediction-correction fashion where the predictor is generated by the primal-dual procedure (1.3) and it is corrected by a very simple combination step. Similar as the motivation stated in Section 3.2, our idea is that for the case $\theta = 1$, we can easily derive that

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

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

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

Step 0. Let $\theta = 1$ and $\rho \in (0, 2)$. Take $u^0 \in \Re^l$. Choose the step sizes τ and σ according to (5.2).
Prediction Step: Generate the predictor $\tilde{u}^k$ via solving (3.1).
Correction Step: Correct the predictor and generate the new iterate u^{k+1} via:

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

Remark 5.1. Since the primal-dual procedure (1.3) is exactly the PPA (see (2.3)) when $\theta = 1$, the above algorithm can also be regarded as an extended PPA, or a PPA-based descent method.

Remark 5.2. It should be mentioned that the above algorithm differs from the previous algorithms in that we have to estimate the quantity of $\|A^T A\|$ in order to determine the values of the step sizes τ and σ and so as to ensure the positive definiteness of M, while this estimation can be avoided by Armijo's rule for the previous algorithms. Fortunately, for many applications such as those arising in TV image restoration problems, it is easy to complete this estimation. For example, when A the corresponding discrete gradient operator of TV terms, according to the Gershgorin's theorem [11], we have that $\|A^T A\| \leq 8$.

The following lemma shows that the proposed Algorithm 4 is a contraction method for (1.1), based on which we can prove the convergence of Algorithm 4 easily.

Lemma 5.3. *The sequence $\{u^k\}$ generated by the proposed Algorithm 4 satisifies:*

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

Proof. Using (5.5) and (5.4) and by a simple manipulation, we obtain

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

The assertion is proved. $\square$

Remark 5.4. According to Lemma 5.3, it is clear why the parameter ρ is restricted into the interval $(0,2)$. When $\rho \in (0,2)$, Lemma 5.3 indicates that the sequence $\{u^k\}$ generated by the proposed Algorithm 4 is Fejèr monotone with respect to the solution set Ω^* under the M-norm. Hence, the convergence of Algorithm 4 is trivial. 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

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

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

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 Extensions to generic saddle-point problems

As we have mentioned at the beginning of this paper, we can extend our convergence analysis to the generic saddle-point problem (1.2) which was considered in [6] (where we change the letters G, F and K in [6] to g, f and $-A$, respectively). The methodology is completely the same as the proposed analysis. Thus we only briefly describe the procedure of reformulating (1.2) into an variational inequality reformulation and expressing the existing primal-dual algorithms by a variant of PPA with asymmetric precondition matrix.

Similarly, the problem (1.2) can be reformulated intto the following monotone variational inequality: Find $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$, such that

$$\begin{pmatrix} x - x^* \\ y - y^* \end{pmatrix}^T \begin{pmatrix} \partial g(x^*) - A^T y^* \\ Ax^* + \partial f^*(y^*) \end{pmatrix} \geq 0, \quad \forall (x,y) \in \mathcal{X} \times \mathcal{Y}, \tag{6.1}$$

where ∂g and ∂f^* are sub-gradient of g and f^*, respectively. Then, (6.1) is of the form of the VI (Ω, F) (2.1a) with

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad F(u) = \begin{pmatrix} \partial g(x) - A^T y \\ Ax + \partial f^*(y) \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{X} \times \mathcal{Y}. \tag{6.2}$$

Because f and g are convex, the mapping F in (6.2) is monotone.

Assume that f and g are "simple" in the sense that their resolvent operator defined by

$$x = (I + \tau\partial g)^{-1}(\xi) = \operatorname{Arg}\min_{x\in\mathcal{X}}\big(\tau g(x) + \frac{1}{2}\|x-\xi\|^2\big)$$

has a closed-form solution or can be efficiently solved up to a high precision, for any given $\xi \in X$ and $\tau > 0$. Then the following primal-dual algorithm for (1.2) was presented in [6].

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

$$x^{k+1} = \operatorname{Arg}\min_{x\in\mathcal{X}} \big\{\tau\big(g(x) - \langle Ax, y^k\rangle\big) + \frac{1}{2}\|x - x^k\|^2\big\}, \tag{6.3a}$$

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

$$y^{k+1} = \operatorname{Arg}\min_{y\in\mathcal{Y}} \big\{\sigma\big(f^*(y) + \langle A\bar{x}^k, y\rangle\big) + \frac{1}{2}\|y - y^k\|^2\big\}. \tag{6.3c}$$

With analogous reasoning, we can easily verify that the iterate (x^{k+1}, y^{k+1}) can be characterized by the following variational inequality:

$$\begin{pmatrix} x - x^{k+1} \\ y - y^{k+1} \end{pmatrix}^T \left\{ \begin{pmatrix} \partial g(x^{k+1}) - A^T y^{k+1} \\ Ax^{k+1} + \partial f^*(y^{k+1}) \end{pmatrix} + \left[\begin{pmatrix} \frac{1}{\tau}I & A^T \\ \theta A & \frac{1}{\sigma}I \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{pmatrix} \right] \right\} \geq 0, \ \ \forall (x,y) \in \Omega.$$

Using the definition of $F(u)$ in (6.2), the last variational inequality has the following compact form:

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

where M is defined exactly the same as (2.5). Therefore, we obtain the same assertion in Lemma 2.2. Consequently, all the convergence analysis carried out in Sections 3-5 can be implemented analogously to (6.1).

7 Conclusions

In this paper, from the point of view of contraction methods, we analyze the convergence of primal-dual algorithms for a saddle-point problem which has particular applications in the area of total variation image restoration. We propose some primal-dual-based contraction methods in the fashion of prediction-correction where the primal-dual algorithms generate predictors and they are corrected by some simple correction steps. Under the framework of contraction type methods, the conditions on the involved parameters of existing primal-dual algorithms are relaxed significantly and the convergence analysis of existing primal-dual algorithms are simplified substantially.

References

[1] K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in linear and non-linear programming*. With contributions by H. B. Chenery, S. M. Johnson, S. Karlin, T. Marschak, R. M. Solow. Stanford Mathematical Studies in the Social Science, Vol. II. Stanford University Press, Stanford, Calif., 1958.

[2] H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejèr-monotone methods in Hilbert spaces, Math. Oper. Res., 26(2) (2001), pp. 248-264.

[3] D. P. Bertsekas and J. N. Tsitsiklis, Parallel and Distributed Computation, Numerical Methods. Prentice-Hall, Englewood Cliffs, NJ, (1989).

[4] E. Blum and W. Oettli, Mathematische Optimierung, Econometrics and Operations Research XX, Springer Verlag, 1975.

[5] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imag. Vis., 20 (2004), pp. 89-97.

[6] A. Chambolle and T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, manuscript, 2010.

[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(6) (1999), pp. 1964-1977.

[8] J. Douglas and H. H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, Trans. American Math. Soc., 82 (1956), pp. 421-439.

[9] E. Esser, X. Zhang and T. Chan, A general framework for a class of first order primal-dual algorithms for TV minimization. CAM Reports 09-31, UCLA, Center for Applied Math., 2009.

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

[11] S. Gerschgorin, *Ü*ber die Abgrenzung der Eigenwerte einer Matrix." Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk, 7(1931), pp. 749-754.

[12] 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.

[13] B. S. He, X. L. Fu and Z. K. Jiang, Proximal point algorithm using a linear proximal term, J. Optimi. Theory Appli., 141 (2009), pp. 209-239.

[14] B. S. He, L. Z. Liao, D. R. Han and H. Yang, A new inexact alternating directions method for monontone variational inequalities, Math. Program., 92 (2002), pp. 103-118.

[15] G. M. Korpelevič, An extragradient method for finding saddle points and for other problems, Èkonom. i Mat. Metody, 12(4) (1976), pp. 747-756.

[16] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, Rev. Francaise d'Inform. Recherche Oper., 4 (1970), pp. 154-159.

[17] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, An algorithm for minimizing the Mumford-Shah functional. In ICCV Proceedings, LNCS. Springer, 2009.

[18] L. D. Popov., A modification of the Arrow-Hurwitz method of search for saddle points, Mat. Zametki, 28(5) (1980), pp. 777-784.

[19] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM, J. Control Optim. 14 (1976), pp. 877-898.

[20] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), pp. 227-238.

[21] X. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, J. Sci. Comput., to appear.

[22] M. Zhu and T. F. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Reports 08-34, UCLA, Center for Applied Math., 2008.