CONVERGENCE ANALYSIS OF PRIMAL-DUAL BASED METHODS FOR TOTAL VARIATION MINIMIZATION WITH FINITE ELEMENT APPROXIMATION

WENYI TIAN* AND XIAOMING YUAN†

Abstract. We consider the total variation minimization model with consistent finite element discretization. It has been shown in the literature that this model can be reformulated as a saddle-point problem and be efficiently solved by the primal-dual method. The convergence for this application of the primal-dual method has also been analyzed. In this paper, we focus on a more general primal-dual scheme with a combination factor for the model and derive its convergence. We also establish the worst-case convergence rate measured by the iteration complexity for this general primal-dual scheme. Furthermore, we propose a prediction-correction scheme based on the general primal-dual scheme, which can significantly relax the step size for the discretization in the time direction. Its convergence and the worst-case convergence rate are established. We present preliminary numerical results to verify the rationale of considering the general primal-dual scheme and the primal-dual-based prediction-correction scheme.

Key words. Total variation minimization, saddle-point problem, finite element method, primal-dual method, convergence rate.

1. Introduction. We consider the total variation (TV) minimization model in [34]:

$$\inf_u E(u) := \|Du\| + \frac{\alpha}{2}\|u-g\|^2_{L^2(\Omega)}, \tag{1.1}$$

where the energy functional $E : BV(\Omega)\cap L^2(\Omega) \to \mathbb{R}$ with a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ with $d = 2$, $g \in L^2(\Omega)$ is a given function, $\alpha > 0$ is a parameter, $BV(\Omega)$ is the bounded variation space consists of all functions $v \in L^1(\Omega)$ satisfying $\|Dv\| < +\infty$, and $\|Dv\|$ denotes the TV norm defined by

$$\|Dv\| := \sup\Big\{ \int_\Omega v \ \mathrm{div}\varphi \ \mathrm{d}x : \varphi \in C^1_c(\Omega;\mathbb{R}^d), \|\varphi\|_\infty \le 1\Big\}. \tag{1.2}$$

In (1.2), $\|\varphi\|_\infty = (\sum_{i=1}^d \sup_{x\in\Omega} |\varphi_i(x)|^2)^{1/2}$, Dv represents the gradient of v in the distributional sense, div denotes the divergence operator, and $C^1_c(\Omega;\mathbb{R}^d)$ is the set of once continuously differentiable $\mathbb{R}^d$-valued functions with compact support in Ω. The $BV(\Omega)$ space endowed with norm $\|v\|_{BV} := \|v\|_{L^1(\Omega)} + \|Dv\|$ is a Banach space. We refer the reader to, e.g., [2, 5, 6, 40] for more details. As well studied in the literature, the model (1.1) has particular effective applications in image processing domain, mainly because of the capability of preserving sharp edges of digital images with piecewise smooth structure. Note that the model (1.1) has a unique solution in $BV(\Omega)$ because of the strict convexity of the quadratic term of the energy functional $E(u)$.

To find a numerical solution for (1.1), we can consider its Euler-Lagrange equation

$$-\mathrm{div}\Big(\frac{\nabla u}{|\nabla u|}\Big) + \alpha(u-g) = 0, \tag{1.3}$$

*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China.

†Department of Mathematics, Hong Kong Baptist University, Hong Kong, China. This author was partially supported by a General Research Fund from Hong Kong Research Grants Council. Email: xmyuan@hkbu.edu.hk

where ∇u is the gradient of u in $L^1(\Omega)$, and $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$. The equation (1.3) indeed characterizes the first-order KKT condition of (1.1). In the literature, there are various efficient numerical schemes that are applicable for solving (1.3) or its regularized equation with homogeneous Neumann boundary condition. For instance, the time marching scheme in [34], the linear semi-implicit fixed-point method in [17, 18, 36], the interior-point primal-dual implicit quadratic methods in [16] and some others in [6, 12, 15, 37]. Furthermore, we can consider the L^2 gradient flow of the model (1.1) and its regularized problem constructed by evolving the Euler-Lagrange equation:

$$\partial_t u - \mathrm{div}\Big(\frac{\nabla u}{|\nabla u|}\Big) + \alpha(u-g) = 0 \tag{1.4}$$

with Neumann boundary condition and initial data from the theoretical and computational aspects. In (1.4), ∂_t means the time partial derivative. We refer to, e.g., [3, 8, 22, 23, 29, 34, 35], for more discussions for (1.4). In particular, in a very recent work [8], the authors studied an algorithm for (1.4) with $\alpha = 0$ which was based on the methods in [7, 14].

Replacing the term $\frac{\nabla u}{|\nabla u|}$ in (1.4) by a new variable p, we can reformulate (1.4) as

$$\partial_t u - \mathrm{div} p + \alpha(u-g) = 0, \quad p \in \partial|\nabla u|, \tag{1.5}$$

where $\partial(|\cdot|)$ denotes the subdifferential of a nonsmooth convex function. Indeed, (1.5) is the L^2 gradient flow of the energy functional $E(u)$ considered in [7]. As indicated in [7], by the equivalence

$$p \in \partial|\nabla u| \;\Leftrightarrow\; \nabla u \in \partial I_B(p),$$

where $B = \{p \in L^1(\Omega;\mathbb{R}^d) : \|p\|_\infty \leq 1\}$ and $I_B(\cdot)$ denotes its indictor function, it motivates us to consider the following system of evolution equations

$$\partial_t u - \mathrm{div} p + \alpha(u-g) = 0, \quad -\sigma\partial_t p + \nabla u \in \partial I_B(p) \tag{1.6}$$

with a parameter $\sigma > 0$ as the scale for ∇u. Note that the system (1.6) can also be regarded as the simultaneous gradient flow for the saddle-point formulation of the model (1.1)

$$\inf_u E(u) = \inf_u \sup_p \mathcal{E}(u,p) := \frac{\alpha}{2}\|u-g\|^2_{L^2(\Omega)} + \int_\Omega \nabla u \cdot p \;\mathrm{d}x - I_B(p). \tag{1.7}$$

As mentioned in [7], the piecewise constant and piecewise affine globally continuous finite element spaces are dense in $BV(\Omega)$ with respect to weak* convergence in $BV(\Omega)$. Thus, the following finite element spaces

$$\begin{cases} \mathcal{S}^1(\mathcal{T}_h) = \{v_h \in C(\bar{\Omega}) : v_h|_T \text{ is affine for each } T \in \mathcal{T}_h\}, \\ \mathcal{L}^0(\mathcal{T}_h) = \{q_h \in L^1(\Omega) : q_h|_T \text{ is constant for each } T \in \mathcal{T}_h\}, \end{cases}$$

are built in [7] to approximate the functions u and p in (1.6), respectively, where $\mathcal{T}_h$ denotes as the regular triangulations of Ω into triangles and $h = \max_{T\in\mathcal{T}_h} \mathrm{diam}(T)$ as the maximal diameters. Furthermore, in [7], the L^2 scalar product to equip $\mathcal{L}^0(\mathcal{T}_h)^d$, i.e., a space of piecewise constant vector fields, is based on the identity

$$\|Du_h\| = \sup_{p_h \in \mathcal{L}^0(\mathcal{T}_h)^d, \|p_h\|_\infty \leq 1} \int_\Omega \nabla u_h \cdot p_h \;\mathrm{d}x.$$

Then, the TV minimization model (1.1) with finite element approximation can be reformulated as the following saddle-point problem:

$$
\begin{aligned}
\inf_{u_h} E(u_h) &= \inf_{u_h \in \mathcal{S}^1(\mathcal{T}_h)} \sup_{p_h \in \mathcal{L}^0(\mathcal{T}_h)^d} \mathcal{E}(u_h, p_h) \\
&= \frac{\alpha}{2}\|u_h - g\|^2_{L^2(\Omega)} + \int_\Omega \nabla u_h \cdot p_h \,\mathrm{d}x - I_B(p_h).
\end{aligned} \tag{1.8}
$$

Note that the discretized saddle-point problem (1.8) has a solution point $(u_h, p_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$, because $\mathcal{E}(u_h, p_h)$ is a closed, proper and convex-concave functional, see, e.g., [7].

For solving a saddle-point problem including the special form (1.8), the primal-dual method has received much attention from different areas, see, e.g., earlier work on the inexact Uzawa method [4, 19] and [10, 19, 24, 25, 33, 41] for saddle-point linear systems resulted from the numerical approximation of elasticity problems or stokes equations, and the quadratic programming with linear constraints. Moreover, in some work such as [14, 20, 38, 39], its particular applications to various image processing problems have been intensively investigated. As analyzed in [20, 38], the primal-dual method is a variant of inexact Uzawa method [4, 19]. In [7], the saddle-point problem (1.8) was solved by an iterative scheme, which is the variational form of the following primal-dual scheme:

$$
u_h^{n+1} = \arg\min \left\{ \mathcal{E}(v_h, p_h^n) + \frac{1}{2\tau}\|v_h - u_h^n\|^2 \mid v_h \in \mathcal{S}^1(\mathcal{T}_h) \right\}, \tag{1.9a}
$$

$$
\tilde{u}_h^{n+1} = 2u_h^{n+1} - u_h^n, \tag{1.9b}
$$

$$
p_h^{n+1} = \arg\max \left\{ \mathcal{E}(\tilde{u}_h^{n+1}, q_h) - \frac{\sigma}{2\tau}\|q_h - p_h^n\|^2 \mid q_h \in \mathcal{L}^0(\mathcal{T}_h)^d \right\}, \tag{1.9c}
$$

where $\tau > 0$ and it can be understood as the step size for implementing gradient-based iterative methods for the minimization and maximization subproblems in (1.9). The scheme (1.9) can be viewed as an semi-implicit difference scheme for (1.6) with finite element discretization in space, and the parameter τ in (1.9) behaves as the step size in the time direction. We refer to [7] for its convergence and numerical efficiency.

Inspired by the work [14, 26], in this paper we consider a more general primal-dual scheme, which includes the scheme (1.9) as a special case, for the model (1.8). More specifically, we propose the following scheme:

$$
u_h^{n+1} = \arg\min \left\{ \mathcal{E}(v_h, p_h^n) + \frac{1}{2\tau}\|v_h - u_h^n\|^2 \mid v_h \in \mathcal{S}^1(\mathcal{T}_h) \right\}, \tag{1.10a}
$$

$$
\tilde{u}_h^{n+1} = u_h^{n+1} + \theta(u_h^{n+1} - u_h^n), \tag{1.10b}
$$

$$
p_h^{n+1} = \arg\max \left\{ \mathcal{E}(\tilde{u}_h^{n+1}, q_h) - \frac{\sigma}{2\tau}\|q_h - p_h^n\|^2 \mid q_h \in \mathcal{L}^0(\mathcal{T}_h)^d \right\}, \tag{1.10c}
$$

where the combination factor $\theta \in [-1, 1]$. Clearly, (1.9) is a special case of (1.10) with $\theta = 1$. Note that it is in [14] that θ was extended to $[0, 1]$ and then in [26] to $[-1, 1]$ in the optimization context. This generalization can accelerate the convergence numerically as shown in [26], and it provides more insights in algorithmic design as shown in [32].

Our contributions can be summarized as follows. 1) We propose the general primal-dual scheme (1.10) and prove its convergence; for which the analytic techniques for contraction type methods (see [9]) are used. 2) We establish the worst-case convergence rate measured by the iteration complexity for the scheme (1.10). 3) We propose a new prediction-correction scheme in which the output of (1.10) needs to be refreshed by a correction step (5.2). This primal-dual-based prediction-correction scheme can significantly relax the restriction on the discretization step size τ from $O(h^2)$ to $O(h)$. 4) We also establish the convergence and the worst-case convergence rate measured by the iteration complexity for the primal-dual-based prediction-correction scheme.

The rest of this paper is organized as follows. In Section 2, some known results in the literature which are useful for further analysis are summarized. In Section 3, we propose the general primal-dual scheme (1.10) in the variational form and give some remarks. Then, the analysis of convergence and convergence rate for the general primal-dual scheme is presented in Section 4. In Section 5, we propose a primal-dual-based prediction-correction scheme and analyze its convergence and convergence rate. Some preliminary numerical results are reported in Section 6 to verify the effectiveness of the general primal-dual scheme and the new primal-dual-based prediction-correction scheme. Finally, some conclusions are made in Section 7.

2. Preliminary. In this section, we summarize some known results in the literature for the convenience of further analysis. Most of the results can be found in [7]. The notation $(\cdot,\cdot)$ stands for the L^2 scalar product throughout this paper.

First, the first-order optimality condition for the minimization of the energy functional $E(u)$ in $\mathcal{S}^1(\mathcal{T}_h)$ is stated in the following lemma.

LEMMA 2.1 ([7]). *The function $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ minimizes the energy functional $E(u)$ in $\mathcal{S}^1(\mathcal{T}_h)$ if and only if there exists $p_h \in \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d) := \{q_h \in \mathcal{L}^0(\mathcal{T}_h)^d : \|q_h\|_\infty \le 1\}$ such that*

$$\begin{aligned} \big(p_h, \nabla v_h\big) + \alpha\big(u_h - g, v_h\big) = 0, &\quad \forall\, v_h \in \mathcal{S}^1(\mathcal{T}_h), \\ \big(\nabla u_h, q_h - p_h\big) \le 0, &\quad \forall\, q_h \in \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \end{aligned} \tag{2.1}$$

Notice that the optimality condition (2.1) can be rewritten as the following variational inequality (VI) in a compact form: Finding $\mu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ such that

$$\big(F(\mu_h), \nu_h - \mu_h\big) \ge 0, \quad \forall\, \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \tag{2.2}$$

where

$$\mu_h = \begin{pmatrix} u_h \\ p_h \end{pmatrix}, \quad \nu_h = \begin{pmatrix} v_h \\ q_h \end{pmatrix}, \quad F(\mu_h) = \begin{pmatrix} -\mathrm{div} p_h + \alpha(u_h - g) \\ -\nabla u_h \end{pmatrix}, \tag{2.3}$$

and $-\mathrm{div}$ is the conjugate operator of ∇ and $-(\mathrm{div} q_h, v_h) = (q_h, \nabla v_h)$.

Obviously, it is easy to see that the mapping $F(\cdot)$ in (2.3) satisfies

$$\big(F(\mu_h) - F(\nu_h), \mu_h - \nu_h\big) = \alpha \|u_h - v_h\|^2_{L^2(\Omega)}. \tag{2.4}$$

It follows from Lemma 2.1 that the first component u_h of a solution pair of (1.8) is unique. But the second one p_h is not unique in general.

The error of the finite element approximation of function u in (1.8) is given by the following theorem.

THEOREM 2.2 ([7]). *We have $u_h \to u$ in $L^2(\Omega)$ as $h \to 0$, and if $d = 2$, $\Omega = (0,1)^2$ and $u \in Lip(\beta, L^2(\Omega))$ for some $0 < \beta < 1$, then*

$$\|u - u_h\|^2_{L^2(\Omega)} \le c h^{\frac{\beta}{1+\beta}},$$

where $u \in Lip(\beta, L^2(\Omega))$ *if*

$$\sup_{t>0} t^{-\beta} \sup_{|y|\le t} \Big(\int_\Omega |u(x+y)-u(x)|^2 \mathrm{d}x \Big)^{1/2} < +\infty.$$

3. A General Primal-Dual Scheme. In this section, we specify the general primal-dual scheme (1.10) in the variational form for solving (1.8) and present the algorithm based on it. Let us use the notation

$$d_t v^{n+1} = \frac{v^{n+1} - v^n}{\tau}$$

for a sequence $\{v^n\}_{n\in\mathbb{N}}$, where $\tau > 0$ is the discretization step size.

Algorithm 1: A general primal-dual scheme for (1.8)

Input: Choose an initial iteration $(u_h^0, p_h^0) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$. Choose constants $\tau, \sigma > 0$ and $\theta \in [-1, 1]$.

for $n = 0, 1, 2, \cdots,$ **do**

Step 1 Update u_h^{n+1} by solving

$$\big(d_t u_h^{n+1} + \alpha(u_h^{n+1} - g), v_h\big) + \big(p_h^n, \nabla v_h\big) = 0, \quad \forall\, v_h \in \mathcal{S}^1(\mathcal{T}_h); \tag{3.1}$$

Step 2 Set

$$\tilde{u}_h^{n+1} = u_h^{n+1} + \theta\tau d_t u_h^{n+1}; \tag{3.2}$$

Step 3 Update p_h^{n+1} satisfying

$$\big(-\sigma d_t p_h^{n+1} + \nabla \tilde{u}_h^{n+1}, q_h - p_h^{n+1}\big) \le 0, \quad \forall\, q_h \in \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{3.3}$$

end

REMARK 3.1. *As mentioned, Algorithm 1 with* $\theta = 1$ *reduces to the primal-dual scheme in [7]. In [7], it is noticed that* p_h^{n+1} *satisfying* (3.3) *is the unique minimizer of*

$$q_h \mapsto \frac{\sigma}{2\tau}\|q_h - p_h^n\|^2 - \big(q_h, \nabla \tilde{u}_h^{n+1}\big) + I_B(q_h);$$

and it is given by

$$p_h^{n+1} = \big(p_h^n + (\tau/\sigma)\nabla \tilde{u}_h^{n+1}\big) / \max\big\{1, |p_h^n + (\tau/\sigma)\nabla\tilde{u}_h^{n+1}|\big\}.$$

REMARK 3.2. *The first and third steps in Algorithm 1 can be respectively viewed as the discretization of the systems*

$$\begin{cases} \partial_t u_h = -\partial_v \mathcal{E}(u_h, p_h), \\ \sigma \partial_t p_h \in \partial_q \mathcal{E}(u_h, p_h), \end{cases} \tag{3.4}$$

which are the finite element discretizations of the evolution systems (1.6)*. Moreover, the parameter* τ *in Algorithm 1 behaves as the step size of the discretization in the time direction.*

REMARK 3.3. *An inverse estimate in [11] shows that there exists* $c > 0$ *such that* $\|\nabla v_h\|_{L^2(\Omega)} \le c h_{\min}^{-1} \|v_h\|_{L^2(\Omega)}$ *for all* $v_h \in \mathcal{S}^1(\mathcal{T}_h)$*, where* $h_{\min} = \min_{T\in\mathcal{T}_h} \mathrm{diam}(T)$.

We denote

$$\|\nabla\| = \sup_{v_h \in \mathcal{S}^1(\mathcal{T}_h)\backslash\{0\}} \frac{\|\nabla v_h\|_{L^2(\Omega)}}{\|v_h\|_{L^2(\Omega)}} \leq ch_{\min}^{-1},$$

which will emerge in the theoretical analysis in the sequel. For a regular mesh $\mathcal{T}_h$, it yields from the above estimate that $\|\nabla\| \leq ch^{-1}$.

4. Convergence Analysis for Algorithm 1. In this section, we prove the convergence for Algorithm 1 and establish its worst-case convergence rate measured by the iteration complexity. As [26], our analysis follows the framework for contraction type methods in [9].

4.1. Convergence. First, for the iteration $\mu_h^{n+1} = (u_h^{n+1}; p_h^{n+1}) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ generated by Algorithm 1, it is easy to see that it satisfies the VI:

$$\big(F(\mu_h^{n+1}) + M(\mu_h^{n+1} - \mu_h^n), \nu_h - \mu_h^{n+1}\big) \geq 0, \ \ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d) \tag{4.1}$$

with

$$M = \begin{pmatrix} \frac{1}{\tau} I & \mathrm{div} \\ -\theta\nabla & \frac{\sigma}{\tau} I \end{pmatrix}. \tag{4.2}$$

We prove an inequality in the following lemma; it is useful for further analysis.

LEMMA 4.1. *Let the sequence $\{\mu_h^{n+1} = (u_h^{n+1}; p_h^{n+1})\}$ be generated by Algorithm 1 with $\theta \in [-1, 1]$. Then, we have*

$$\begin{aligned} \big(G(\mu_h^{n+1} - \mu_h^n), \nu_h - \mu_h^{n+1}\big) \geq{}& \alpha\|v_h - u_h^{n+1}\|_{L^2(\Omega)}^2 + \big(F(\nu_h), \mu_h^{n+1} - \nu_h\big) \\ & - (1-\theta)\big(\nabla(v_h - u_h^{n+1}), p_h^n - p_h^{n+1}\big), \\ & \forall\ \nu_h = (v_h; q_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \end{aligned} \tag{4.3}$$

where

$$G = \begin{pmatrix} \frac{1}{\tau} I & \theta\mathrm{div} \\ -\theta\nabla & \frac{\sigma}{\tau} I \end{pmatrix}. \tag{4.4}$$

Proof. We can rewrite (4.1) as follows

$$\begin{aligned} &\big(F(\mu_h^{n+1}) + G(\mu_h^{n+1} - \mu_h^n), \nu_h - \mu_h^{n+1}\big) - (1-\theta)\big(\nabla(v_h - u_h^{n+1}), p_h^{n+1} - p_h^n\big) \\ &\geq 0, \ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \end{aligned} \tag{4.5}$$

Then, adding $\big(F(\nu_h), \mu_h^{n+1} - \nu_h\big)$ to both sides of (4.5) yields

$$\begin{aligned} \big(G(\mu_h^{n+1} - \mu_h^n), \nu_h - \mu_h^{n+1}\big) \geq{}& \big(F(\nu_h) - F(\mu_h^{n+1}), \nu_h - \mu_h^{n+1}\big) \\ & + \big(F(\nu_h), \mu_h^{n+1} - \nu_h\big) \\ & - (1-\theta)\big(\nabla(v_h - u_h^{n+1}), p_h^n - p_h^{n+1}\big), \\ & \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \end{aligned} \tag{4.6}$$

which completes the proof by using the property (2.4) of F. ∎

Using the inequality proved in the above lemma, we can show that the sequence generated by Algorithm 1 is strictly contractive with respect to the solution set of (1.8) under some conditions. We summarize the assertion in the following theorem.

THEOREM 4.2 (Contraction). *Let* μ_h *be the solution point of* (1.8) *and* $\{\mu_h^{n+1}\}$ *be the sequence generated by Algorithm 1 with* $\theta \in [-1, 1]$. *Under the condition*

$$\Big(\theta^2 + \frac{(1-\theta)^2}{2\alpha\tau}\Big)\frac{\tau^2\|\nabla\|^2}{\sigma} < 1, \tag{4.7}$$

we have

$$\|\mu_h^{n+1} - \mu_h\|_G^2 \le \|\mu_h^n - \mu_h\|_G^2 - \|\mu_h^n - \mu_h^{n+1}\|_Q^2, \tag{4.8}$$

where G *is given by* (4.4) *and*

$$Q = \begin{pmatrix} \frac{1}{\tau}I & \theta\mathrm{div} \\ -\theta\nabla & \Big(\frac{\sigma}{\tau} - \frac{(1-\theta)^2\|\nabla\|^2}{2\alpha}\Big)I \end{pmatrix}. \tag{4.9}$$

Proof. Using Cauchy-Schwarz inequality and it follows from (4.3) that

(4.10)
$$\begin{aligned}\big(G(\mu_h^{n+1} - \mu_h^n), \nu_h - \mu_h^{n+1}\big) \ge{}& \alpha\|v_h - u_h^{n+1}\|_{L^2(\Omega)}^2 + \big(F(\nu_h), \mu_h^{n+1} - \nu_h\big) \\ &- \frac{\alpha}{\|\nabla\|^2}\|\nabla(v_h - u_h^{n+1})\|_{L^2(\Omega)}^2 \\ &- \frac{(1-\theta)^2\|\nabla\|^2}{4\alpha}\|p_h^n - p_h^{n+1}\|_{L^2(\Omega)}^2 \\ \ge{}& \big(F(\nu_h), \mu_h^{n+1} - \nu_h\big) - \frac{(1-\theta)^2\|\nabla\|^2}{4\alpha}\|p_h^n - p_h^{n+1}\|_{L^2(\Omega)}^2, \\ & \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).\end{aligned}$$

Applying the identity

$$\big(G(b-a), b\big) = \frac{1}{2}\big(\|b\|_G^2 - \|a\|_G^2 + \|a-b\|_G^2\big) \tag{4.11}$$

to the term on the left-hand side of (4.10) with $b = \mu_h^{n+1} - \nu_h$ and $a = \mu_h^n - \nu_h$, we derive

(4.12)
$$\begin{aligned}2\big(F(\nu_h), \mu_h^{n+1} - \nu_h\big) \le{}& \|\mu_h^n - \nu_h\|_G^2 - \|\mu_h^{n+1} - \nu_h\|_G^2 \\ &- \Big(\|\mu_h^n - \mu_h^{n+1}\|_G^2 - \frac{(1-\theta)^2\|\nabla\|^2}{2\alpha}\|p_h^n - p_h^{n+1}\|_{L^2(\Omega)}^2\Big), \\ & \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).\end{aligned}$$

Thus, it yields

$$\begin{aligned}2\big(F(\nu_h), \mu_h^{n+1} - \nu_h\big) \le \|\mu_h^n - \nu_h\|_G^2 - \|\mu_h^{n+1} - \nu_h\|_G^2 - \|\mu_h^n - \mu_h^{n+1}\|_Q^2, \\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).\end{aligned} \tag{4.13}$$

Setting $\nu_h = \mu_h$ in (4.13) and using the optimality condition (2.2), then we get the result (4.8). ∎

We emphasize that the condition (4.7) is essential for ensuring the positive definiteness of the matrix form operators G and Q given by (4.4) and (4.9), respectively. The strict contraction of the sequence generated by Algorithm 1, which is implied by the assertion (4.8), essentially means that the sequence $\{\mu_h^{n+1}\}$ converges to the solution set of (1.8). We summarize the convergence result in the following theorem.

THEOREM 4.3 (Convergence). *Let the sequence $\{\mu_h^{n+1} = (u_h^{n+1}; p_h^{n+1})\}$ be generated by Algorithm 1 with $\theta \in [-1, 1]$. Under the condition* (4.7)*, the sequence $\{u_h^{n+1}\}$ converges to the unique minimizer of the problem* (1.1) *in $\mathcal{S}^1(\mathcal{T}_h)$.*

Proof. According to (4.8), for any integer $N > 0$, we have

$$\sum_{n=0}^{N} \|\mu_h^n - \mu_h^{n+1}\|_Q^2 \le \|\mu_h - \mu_h^0\|_G^2.$$

Thus, we conclude

$$\lim_{n\to\infty} \|\mu_h^n - \mu_h^{n+1}\|_Q^2 = 0.$$

As Q is positive definite under the condition (4.7), then $\lim_{n\to\infty}(\mu_h^n - \mu_h^{n+1}) = 0$. Substituting it into (4.1), we obtain that

$$\big(F(\mu_h^{n+1}), \nu_h - \mu_h^{n+1}\big) \ge 0, \quad \forall\, \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d),$$

which means $\lim_{n\to\infty} \mu_h^{n+1}$ is a solution point of (2.2). Thus, the sequence $\{u_h^{n+1}\}$ converges to the unique minimizer of energy functional E in $\mathcal{S}^1(\mathcal{T}_h)$ by Lemma 2.1. □

4.2. Convergence Rate. In this subsection, we estimate a worst-case $O(\frac{1}{N})$ convergence rate measured by the iteration complexity for Algorithm 1 with $\theta \in [-1, 1]$ under the condition (4.7), where N denotes the iteration counter. Note that we follow [30, 31] and many others, a worst-case $O(\frac{1}{N})$ convergence rate means the accuracy to a solution under certain criteria is of the order $O(\frac{1}{N})$ after N iterations of an iterative scheme; or equivalently, it requires at most $O(\frac{1}{\epsilon})$ iterations to achieve an approximate solution with an accuracy of ϵ.

First, we introduce a criterion to measure the accuracy of an approximation of the VI (2.2).

THEOREM 4.4. *The solution set of VI* (2.2) *is convex and can be characterized as*

$$\Theta = \bigcap_{\nu_h} \big\{\tilde{\mu}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d) : \big(F(\nu_h), \nu_h - \tilde{\mu}_h\big) \ge 0\big\}.$$

Proof. The proof can refer to Theorem 2.3.5 in [21] or Theorem 2.1 in [27]. □

Theorem 4.4 implies that $\tilde{\mu}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ is an approximate solution of VI (2.2) with an accuracy of ϵ if

$$\big(F(\nu_h), \tilde{\mu}_h - \nu_h\big) \le \epsilon, \quad \forall\, \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{4.14}$$

The result in the following theorem shows that we can find $\tilde{\mu}_N$ such that (4.14) is satisfied with $\epsilon = O(\frac{1}{N})$ after N iterations of Algorithm 1. Therefore, a worst-case $O(\frac{1}{N})$ convergence rate is established for Algorithm 1.

THEOREM 4.5 (Convergence rate in the ergodic sense). *Let the sequence $\{\mu_h^{n+1}\}$ be generated by Algorithm 1 with $\theta \in [-1, 1]$ under the condition* (4.7)*. For any integer*

$N > 0$, *let*

$$\tilde{\mu}_N = \frac{1}{N+1}\sum_{n=0}^{N} \mu_h^{n+1}.$$

Then, we have

$$\big(F(\nu_h), \tilde{\mu}_N - \nu_h\big) \le \frac{1}{2(N+1)}\|\nu_h - \mu_h^0\|_G^2,\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{4.15}$$

Proof. It follows from (4.13) that

$$\big(F(\nu_h), \mu_h^{n+1} - \nu_h\big) \le \frac{1}{2}\big(\|\nu_h - \mu_h^n\|_G^2 - \|\nu_h - \mu_h^{n+1}\|_G^2\big),\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{4.16}$$

Summing (4.16) with $n = 0, 1, \cdots, N$, we have

$$\big(F(\nu_h), \sum_{n=0}^{N} \mu_h^{n+1} - (N+1)\nu_h\big) \le \frac{1}{2}\big(\|\nu_h - \mu_h^0\|_G^2 - \|\nu_h - \mu_h^{N+1}\|_G^2\big), \tag{4.17}$$

which yields the result (4.15). □

This theorem shows a worst-case $O(\frac{1}{N})$ convergence rate in the ergodic sense for Algorithm 1. The ergodic sense is because of the fact that the approximate solution with an accuracy of $O(\frac{1}{N})$ is the average of all the N iterations generated by Algorithm 1. For the special case $\theta = 1$ of Algorithm 1, i.e., the primal-dual scheme in [7], we can obtain a stronger worst-case $O(\frac{1}{N})$ convergence rate in a nonergodic sense. But it is not clear if this convergence rate in a nonergodic sense can be extended to the general case of Algorithm 1 with $\theta \in [-1, 1)$. The main reason is that the matrix form operator M defined in (4.2) is not symmetric if $\theta \ne 1$. Hence, it becomes difficult to define a norm with these matrix form operators to measure the progress of proximity to the solution set. We summarize the stronger worst-case $O(\frac{1}{N})$ convergence rate in a nonergodic sense for Algorithm 1 in the following theorem. This is a by-produce of this paper.

THEOREM 4.6 (Convergence rate for $\theta = 1$ in a nonergodic sense). *Let μ_h be the solution of* (1.8) *and the sequence $\{\mu_h^{n+1}\}$ be generated by Algorithm 1 with $\theta = 1$ under the condition* (4.7). *Then for any integer $N > 0$, it exists*

$$\|\mu_h^N - \mu_h^{N+1}\|_G^2 \le \frac{1}{(N+1)}\|\mu_h - \mu_h^0\|_G^2. \tag{4.18}$$

Proof. First, it follows from (4.1) when $\theta = 1$ that

$$\big(F(\mu_h^{n+1}) + G(\mu_h^{n+1} - \mu_h^n), \nu_h - \mu_h^{n+1}\big) \ge 0,\ \ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{4.19}$$

And it also holds

$$\big(F(\mu_h^{n+2}) + G(\mu_h^{n+2} - \mu_h^{n+1}), \nu_h - \mu_h^{n+2}\big) \ge 0,\ \ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{4.20}$$

Setting $\nu_h = \mu_h^{n+2}$ in (4.19) and $\nu_h = \mu_h^{n+1}$ in (4.20), and then combining them together, we obtain

$$\begin{aligned} &\big(F(\mu_h^{n+1}) - F(\mu_h^{n+2}), \mu_h^{n+2} - \mu_h^{n+1}\big) \\ &- \big(G((\mu_h^{n+2} - \mu_h^{n+1}) - (\mu_h^{n+1} - \mu_h^n)), \mu_h^{n+2} - \mu_h^{n+1}\big) \ge 0. \end{aligned} \tag{4.21}$$

Applying the equalities (4.11) and (2.4) to (4.21) yields

$$
\begin{aligned}
\|\mu_h^{n+1}-\mu_h^{n+2}\|_G^2 \le \|\mu_h^{n}-\mu_h^{n+1}\|_G^2 - 2\alpha\|u_h^{n+1}-u_h^{n+2}\|_{L^2(\Omega)}^2 \\
- \|(\mu_h^{n+2}-\mu_h^{n+1})-(\mu_h^{n+1}-\mu_h^{n})\|_G^2 \\
\le \|\mu_h^{n}-\mu_h^{n+1}\|_G^2.
\end{aligned} \tag{4.22}
$$

With $\theta = 1$ and the definition of Q in (4.9), we have from the result of Theorem 4.2 that

$$
\|\mu_h-\mu_h^{n+1}\|_G^2 \le \|\mu_h-\mu_h^{n}\|_G^2 - \|\mu_h^{n}-\mu_h^{n+1}\|_G^2. \tag{4.23}
$$

Summing (4.23) over $n = 0, 1, \cdots, N$ yields that

$$
\sum_{n=0}^{N} \|\mu_h^{n}-\mu_h^{n+1}\|_G^2 \le \|\mu_h-\mu_h^{0}\|_G^2. \tag{4.24}
$$

The estimate (4.22) reveals that $\|\mu_h^{n}-\mu_h^{n+1}\|_G^2$ is monotonically non-increasing, then we obtain

$$
(N+1)\|\mu_h^{N}-\mu_h^{N+1}\|_G^2 \le \|\mu_h-\mu_h^{0}\|_G^2, \tag{4.25}
$$

which yields the result (4.18). □

It is follows from (4.19) that μ_h^{N+1} belongs to the solution set of VI (2.2) if $\|\mu_h^{N}-\mu_h^{N+1}\|_G^2 = 0$ since G is positive definite under the condition (4.7) with $\theta = 1$. In other words, if $\|\mu_h^{N}-\mu_h^{N+1}\|_G^2 = 0$, we have

$$
\big(F(\mu_h^{N+1}), \nu_h - \mu_h^{N+1}\big) \ge 0, \ \ \forall\, \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d),
$$

which implies μ_h^{N+1} is a solution of (1.8) characterized by VI (2.2). Then the quantity $\|\mu_h^{N}-\mu_h^{N+1}\|_G^2$ can be used to measure the accuracy of an approximate solution of (1.8). Thus, the assertion in Theorem 4.6 shows a worst-case $O(\frac{1}{N})$ convergence rate measured by the iteration complexity in a nonergodic sense for Algorithm 1 with $\theta = 1$.

5. A Primal-Dual Based Prediction-Correction Scheme. In Section 3, we propose Algorithm 1 which is more general than the primal-dual scheme in [7]. We will show in Section 6.1 that this general scheme with $\theta \neq 1$ can accelerate the convergence numerically; it thus makes sense to consider the generalization for $\theta \in [-1, 1]$. Meanwhile, we have analyzed that the convergence of Algorithm 1 can be guaranteed under the condition (4.7). As indicated in Remark 3.3, we have $\|\nabla\|^2 \le ch^{-2}$ with the regular mesh $\mathcal{T}_h$. When $\theta \neq 1$ for Algorithm 1, the requirement $\tau \le ch^2$ is necessary to make the condition (4.7) satisfied. Here τ stands for the discretization step size in the time direction if we regard Algorithm 1 as the discretizations of (3.4) (which is the evolution systems (1.6) with finite element approximation in space). In this sense, we may wish to relax the requirement on τ from the order of $O(h^2)$ to $O(h)$ if possible. This is the main motivation we consider the new primal-dual based prediction-correction scheme in this section.

We would reiterate that the convergence analysis for Algorithm 1 in Section 4 mainly follows the analytic framework for contraction type methods. For its analysis, a key technique is that the condition (4.7) can ensure the positive definiteness of the matrix form operators G in (4.4) and Q in (4.9). With their positive definiteness, we can measure the progress of proximity between two consecutive iterations and eventually establish the strict contraction property for the sequence obtained by Algorithm 1 which essentially implies the convergence of Algorithm 1. It is seen from (4.4) that the off-diagonal entries of G is not zero operator, which means that it renders two non-square terms in the expansion of any quadratic term associated with the G-norm. This fact essentially raises the reason of considering the condition (4.7) to sufficiently ensure the positive definiteness of G. For the purpose of relaxing the restriction on τ, we may consider only keeping the diagonal entries of G as the matrix form operator for defining the norm when measuring the progress of proximity between two consecutive iterations. That is,

$$G = \begin{pmatrix} \frac{1}{\tau}I & \theta\mathrm{div} \\ -\theta\nabla & \frac{\sigma}{\tau}I \end{pmatrix} \to H := \begin{pmatrix} \frac{1}{\tau}I & 0 \\ 0 & \frac{\sigma}{\tau}I \end{pmatrix}. \tag{5.1}$$

Based on these analysis, we will follow the idea in [26] and modify Algorithm 1 as a new prediction-correction scheme whose each iteration consists of the primal-dual step (3.1)-(3.3) and a correction step (5.2). With the additional correction step, the requirement on τ can be relaxed from $O(h^2)$ to $O(h)$. Moreover, the worst-case convergence rate in both the ergodic and nonergodic senses can be established for the new primal-dual-based prediction-correction scheme.

5.1. Algorithm. We summarize the new primal-dual-based prediction correction scheme as follows.

Algorithm 2: A primal-dual-based prediction-correction scheme for (1.8)

Input: Choose an initial iteration $(u_h^0, p_h^0) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$. Choose constants $\tau, \sigma > 0$, $\gamma \in (0,1]$ and $\theta \in [-1,1]$

for $n = 0, 1, 2, \cdots$, **do**

Prediction step Obtain the predictor $\bar{\mu}_h^n$ by Algorithm 1 with input μ_h^n, i.e. (3.1)-(3.3);

Correction step Generate the new iteration μ_h^{n+1} by solving

$$\begin{aligned} &\big((u_h^{n+1} - u_h^n) + \gamma(u_h^n - \bar{u}_h^n), v_h\big) - \tau\gamma\big(\nabla v_h, p_h^n - \bar{p}_h^n\big) = 0, \qquad \forall\ v_h \in \mathcal{S}^1(\mathcal{T}_h), \\ &\big((p_h^{n+1} - p_h^n) + \gamma(p_h^n - \bar{p}_h^n), q_h\big) - \gamma\theta\frac{\tau}{\sigma}\big(\nabla(u_h^n - \bar{u}_h^n), q_h\big) = 0,\ \forall\ q_h \in \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \end{aligned} \tag{5.2}$$

end

REMARK 5.1. *For the correction step* (5.2)*, it can be rewritten as the compact form*

$$\big(H(\mu_h^{n+1} - \mu_h^n), \nu_h\big) + \gamma\big(M(\mu_h^n - \bar{\mu}_h^n), \nu_h\big) = 0, \quad \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \tag{5.3}$$

where M and H are defined in (4.2) *and* (5.1)*, respectively. It is noticed that the correction step* (5.2) *in Algorithm 2 is not difficult to compute because it is essentially a system of linear algebra equations with a symmetric and positive definite mass matrix.*

REMARK 5.2. *The parameter $\gamma \in (0,1]$ in* (5.2) *is a relaxation factor which can potentially accelerate numerical performance. Instead, we can simply take $\gamma \equiv 1$ if the number of parameters is a concern for implementation.*

5.2. Convergence. In this subsection, we prove the convergence for Algorithm 2. First, we prove an inequality which is important for the convergence analysis. Let us recall (4.1). Thus, the predictor $\bar{\mu}_h^n$ generated by Algorithm 2 satisfies

$$\big(F(\bar{\mu}_h^n) + M(\bar{\mu}_h^n - \mu_h^n), \nu_h - \bar{\mu}_h^n\big) \geq 0, \ \ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{5.4}$$

Lemma 5.1. *Let the sequence $\{\mu_h^{n+1}\}$ be generated by Algorithm 2 with $\theta \in [-1, 1]$, $\gamma \in (0, 1]$ and*

$$\frac{\tau^2\|\nabla\|^2}{\sigma} < 1. \tag{5.5}$$

Then we have

$$\begin{aligned}2\big(H(\mu_h^n - \mu_h^{n+1}), \mu_h^n - \nu_h\big) - \|\mu_h^n - \mu_h^{n+1}\|_H^2 \geq \frac{1}{4}\Big(1 - \frac{\tau^2\|\nabla\|^2}{\sigma}\Big)\|\mu_h^n - \mu_h^{n+1}\|_H^2 \\ +2\gamma\big(F(\nu_h), \bar{\mu}_h^n - \nu_h\big),\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d),\end{aligned} \tag{5.6}$$

where H is defined in (5.1).

Proof. Adding $\big(F(\nu_h), \bar{\mu}_h^n - \nu_h\big)$ to both sides of (5.4), we have

$$\begin{aligned}\big(F(\bar{\mu}_h^n) - F(\nu_h) + M(\bar{\mu}_h^n - \mu_h^n), \nu_h - \bar{\mu}_h^n\big) \geq \big(F(\nu_h), \bar{\mu}_h^n - \nu_h\big), \\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).\end{aligned} \tag{5.7}$$

Then, we derive

$$\begin{aligned}\big(M(\mu_h^n - \bar{\mu}_h^n), \mu_h^n - \nu_h\big) \geq \big(M(\mu_h^n - \bar{\mu}_h^n), \mu_h^n - \bar{\mu}_h^n\big) + \big(F(\nu_h), \bar{\mu}_h^n - \nu_h\big) \\ + \alpha\|v_h - \bar{u}_h^n\|_{L^2(\Omega)}^2,\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).\end{aligned} \tag{5.8}$$

With (5.3), we obtain

$$\begin{aligned}\big(H(\mu_h^n - \mu_h^{n+1}), \mu_h^n - \nu_h\big) &= \gamma\big(M(\mu_h^n - \bar{\mu}_h^n), \mu_h^n - \nu_h\big) \\ &\geq \gamma\big(M(\mu_h^n - \bar{\mu}_h^n), \mu_h^n - \bar{\mu}_h^n\big) + \gamma\big(F(\nu_h), \bar{\mu}_h^n - \nu_h\big), \\ &\quad \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).\end{aligned} \tag{5.9}$$

Using the definition of M in (4.2) and H in (5.1), we can expand the term on the right-hand side of (5.9) as

$$\big(M(\mu_h^n - \bar{\mu}_h^n), \mu_h^n - \bar{\mu}_h^n\big) = \|\mu_h^n - \bar{\mu}_h^n\|_H^2 - (1+\theta)\big(\nabla(u_h^n - \bar{u}_h^n), p_h^n - \bar{p}_h^n\big). \tag{5.10}$$

With (5.3), we also have

$$\begin{aligned}\|\mu_h^n - \mu_h^{n+1}\|_H^2 &= \gamma\big(M(\mu_h^n - \bar{\mu}_h^n), \mu_h^n - \mu_h^{n+1}\big) \\ &= \gamma\big(H(\mu_h^n - \mu_h^{n+1}), H^{-1}M(\mu_h^n - \bar{\mu}_h^n)\big) \\ &= \gamma^2\big(M(\mu_h^n - \bar{\mu}_h^n), H^{-1}M(\mu_h^n - \bar{\mu}_h^n)\big),\end{aligned} \tag{5.11}$$

whose last term can be evaluated by the definitions of H and M as

$$\begin{aligned}\big(M(\mu_h^n - \bar{\mu}_h^n), H^{-1}M(\mu_h^n - \bar{\mu}_h^n)\big) = \|\mu_h^n - \bar{\mu}_h^n\|_H^2 + \tau\|\mathrm{div}(p_h^n - \bar{p}_h^n)\|^2 \\ + \theta^2\frac{\tau}{\sigma}\|\nabla(u_h^n - \bar{u}_h^n)\|^2 \\ - 2(1+\theta)\big(\nabla(u_h^n - \bar{u}_h^n), p_h^n - \bar{p}_h^n\big).\end{aligned} \tag{5.12}$$

Then, together with (5.9)-(5.12), we get

$$
\begin{aligned}
&2\big(H(\mu_h^n-\mu_h^{n+1}),\mu_h^n-\nu_h\big)-\|\mu_h^n-\mu_h^{n+1}\|_H^2\\
&\geq 2\gamma\big(M(\mu_h^n-\bar\mu_h^n),\mu_h^n-\bar\mu_h^n\big)-\gamma^2\big(M(\mu_h^n-\bar\mu_h^n),H^{-1}M(\mu_h^n-\bar\mu_h^n)\big)\\
&\quad+2\gamma\big(F(\nu_h),\bar\mu_h^n-\nu_h\big)\\
&=(2\gamma-\gamma^2)\|\mu_h^n-\bar\mu_h^n\|_H^2-2(1+\theta)(\gamma-\gamma^2)\big(\nabla(u_h^n-\bar u_h^n),p_h^n-\bar p_h^n\big)\\
&\quad-\gamma^2\Big(\tau\|\mathrm{div}(p_h^n-\bar p_h^n)\|^2+\theta^2\frac{\tau}{\sigma}\|\nabla(u_h^n-\bar u_h^n)\|^2\Big)+2\gamma\big(F(\nu_h),\bar\mu_h^n-\nu_h\big)\\
&\geq\Big((2\gamma-\gamma^2)-\big((1+\theta)(\gamma-\gamma^2)+\theta^2\gamma^2\big)\frac{\tau^2\|\nabla\|^2}{\sigma}\Big)\frac{1}{\tau}\|u_h^n-\bar u_h^n\|_{L^2(\Omega)}^2\\
&\quad+\Big(\big((2\gamma-\gamma^2)-(1+\theta)(\gamma-\gamma^2)\big)-\gamma^2\frac{\tau^2\|\nabla\|^2}{\sigma}\Big)\frac{\sigma}{\tau}\|p_h^n-\bar p_h^n\|_{L^2(\Omega)}^2\\
&\quad+2\gamma\big(F(\nu_h),\bar\mu_h^n-\nu_h\big)\\
&\geq\gamma^2\Big(1-\frac{\tau^2\|\nabla\|^2}{\sigma}\Big)\|\mu_h^n-\bar\mu_h^n\|_H^2+2\gamma\big(F(\nu_h),\bar\mu_h^n-\nu_h\big),\\
&\forall\ \nu_h\in\mathcal{S}^1(\mathcal{T}_h)\times\mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).
\end{aligned}\tag{5.13}
$$

Therefore, because of (5.11) and (5.12), using Cauchy-Schwarz inequality and the condition (5.5), we can derive

$$
\begin{aligned}
\|\mu_h^n-\mu_h^{n+1}\|_H^2&=\gamma^2\Big(\|\mu_h^n-\bar\mu_h^n\|_H^2+\tau\|\mathrm{div}(p_h^n-\bar p_h^n)\|^2+\theta^2\frac{\tau}{\sigma}\|\nabla(u_h^n-\bar u_h^n)\|^2\\
&\quad-2(1+\theta)\big(\nabla(u_h^n-\bar u_h^n),p_h^n-\bar p_h^n\big)\Big)\\
&\leq\gamma^2\Big(1+\big(\theta^2+1+\theta\big)\frac{\tau^2\|\nabla\|^2}{\sigma}\Big)\frac{1}{\tau}\|u_h^n-\bar u_h^n\|_{L^2(\Omega)}^2\\
&\quad+\gamma^2\Big(1+\frac{\tau^2\|\nabla\|^2}{\sigma}+(1+\theta)\Big)\frac{\sigma}{\tau}\|p_h^n-\bar p_h^n\|_{L^2(\Omega)}^2\\
&\leq 4\gamma^2\|\mu_h^n-\bar\mu_h^n\|_H^2.
\end{aligned}\tag{5.14}
$$

Then the result is obtained from (5.13) and (5.14). ◻

Using the result in the above lemma, we can easily derive that the sequence generated by Algorithm 2 is strictly contractive with respect to the solution set of VI (2.2). We summarize it in the following theorem.

THEOREM 5.2 (Contraction). *Let μ_h be the solution of* (1.8) *and the sequence $\{\mu_h^{n+1}\}$ be generated by Algorithm 2 with $\theta\in[-1,1]$ under the condition* (5.5). *Then we have*

$$
\|\mu_h^{n+1}-\mu_h\|_H^2\leq\|\mu_h^n-\mu_h\|_H^2-\frac{1}{4}\Big(1-\frac{\tau^2\|\nabla\|^2}{\sigma}\Big)\|\mu_h^n-\mu_h^{n+1}\|_H^2.\tag{5.15}
$$

Proof. Obviously, we have

$$
\begin{aligned}
\|\mu_h^{n+1}-\nu_h\|_H^2&=\|\mu_h^n-\nu_h-(\mu_h^n-\mu_h^{n+1})\|_H^2\\
&=\|\mu_h^n-\nu_h\|_H^2-2\big(H(\mu_h^n-\mu_h^{n+1}),\mu_h^n-\nu_h\big)+\|\mu_h^n-\mu_h^{n+1}\|_H^2.
\end{aligned}\tag{5.16}
$$

Applying the result (5.6) in Lemma 5.1 to (5.16), setting $\nu_h=\mu_h$ and noticing $\big(F(\mu_h),\bar\mu_h^n-\mu_h\big)\geq 0$, we obtain (5.15). ◻

With the strict contraction property established in the last theorem, it becomes easy to prove the convergence for Algorithm 2. The convergence of Algorithm 2 is summarized in the following theorem.

THEOREM 5.3 (Convergence). *Let the sequence* $\{\mu_h^{n+1} = (u_h^{n+1}, p_h^{n+1})\}$ *be generated by Algorithm 2 with* $\theta \in [-1, 1]$ *and* $\gamma \in (0, 1]$ *under the condition* (5.5)*. Then, the sequence* $\{u_h^{n+1}\}$ *converges to the unique minimizer of energy functional* E *in* $\mathcal{S}^1(\mathcal{T}_h)$*.*

Proof. From (5.13) and (5.16), we obtain

$$\|\mu_h^{n+1} - \mu_h\|_H^2 \le \|\mu_h^n - \mu_h\|_H^2 - \gamma^2\Big(1 - \frac{\tau^2\|\nabla\|^2}{\sigma}\Big)\|\mu_h^n - \bar{\mu}_h^n\|_H^2. \tag{5.17}$$

Then, the above inequality and (5.15) imply that

$$\lim_{n\to\infty}(\mu_h^n - \mu_h^{n+1}) = \lim_{n\to\infty}(\mu_h^n - \bar{\mu}_h^n) = 0.$$

Thus, we have $\lim_{n\to\infty}\mu_h^{n+1} = \lim_{n\to\infty}\bar{\mu}_h^n$. With (5.4), we derive that $\lim_{n\to\infty}\mu_h^{n+1}$ satisfies the VI (2.2). So, the sequence $\{u_h^{n+1}\}$ converges to the unique minimizer of energy functional E in $\mathcal{S}^1(\mathcal{T}_h)$ by Lemma 2.1. ∎

Note that Algorithm 2 requires an additional correction step compared with Algorithm 1. But the condition (5.5) which guarantees the convergence for Algorithm 2 can be satisfied if $\tau \le ch$ for some $c > 0$ as $\|\nabla\| \le ch^{-1}$. This is an significantly relaxed condition compared with the requirement $\tau \le ch^2$ for Algorithm 1. This is the main advantage of Algorithm 2. We will numerically verify its superiority in Section 6.2.

5.3. Convergence Rate. In this subsection, we establish the worst-case $O(\frac{1}{N})$ convergence rate in both the ergodic and nonergodic senses for Algorithm 2 with $\theta \in [-1, 1]$. Recall the lack of worst-case convergence rate in a nonergodic sense of Algorithm 1 with $\theta \in [-1, 1)$. Thus, the provable worst-case convergence rate in a nonergodic sense is another theoretical advantage of Algorithm 2.

5.3.1. Convergence Rate in the Ergodic Sense. We first establish the worst-case $O(\frac{1}{N})$ convergence rate in the ergodic sense for Algorithm 2 in the following theorem. The proof is analogous to that of Theorem 4.5.

THEOREM 5.4 (Convergence rate in the ergodic sense). *Let the sequence* $\{\mu_h^{n+1}\}$ *be generated by Algorithm 2 with* $\theta \in [-1, 1]$ *and* $\gamma \in (0, 1]$ *under the condition* (5.5). *For any integer* N*, let* $\bar{\mu}_N$ *be defined as*

$$\bar{\mu}_N = \frac{1}{N+1}\sum_{n=0}^{N}\bar{\mu}_h^n.$$

Then we have

$$\big(F(\nu_h), \bar{\mu}_N - \nu_h\big) \le \frac{1}{2\gamma(N+1)}\|\nu_h - \mu_h^0\|_H^2, \ \ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{5.18}$$

Proof. It follows from (5.16) and (5.6) that

$$\begin{aligned}\big(F(\nu_h), \bar{\mu}_h^n - \nu_h\big) \le \frac{1}{2\gamma}\big(\|\nu_h - \mu_h^n\|_H^2 - \|\nu_h - \mu_h^{n+1}\|_H^2\big),& \\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d).&\end{aligned} \tag{5.19}$$

Summarizing the inequalities (5.19) for the cases of $n = 0, 1, \cdots, N$, we obtain

$$\big(F(\nu_h), \sum_{n=0}^{N} \bar{\mu}_h^n - (N+1)\nu_h\big) \le \frac{1}{2\gamma}\big(\|\nu_h - \mu_h^0\|_H^2 - \|\nu_h - \mu_h^{N+1}\|_H^2\big), \quad \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \tag{5.20}$$

which leads to the result (5.18). ▯

5.3.2. Convergence Rate in a Nonergodic Sense. Now we establish the worst-case $O(\frac{1}{N})$ convergence rate in a nonergodic sense for Algorithm 2. The analysis is based on the strict contraction property (5.15) and the monotonicity of the sequence $\{\|\mu_h^n - \mu_h^{n+1}\|_H^2\}$. We thus first prove this property in the following lemma.

LEMMA 5.5. *Let the sequence $\{\mu_h^{n+1}\}$ be generated by Algorithm 2 with $\theta \in [-1, 1]$ and $\gamma \in (0, 1]$. Then, we have*

$$\|\mu_h^{n+1} - \mu_h^{n+2}\|_H^2 \le \|\mu_h^n - \mu_h^{n+1}\|_H^2. \tag{5.21}$$

Proof. Because of the optimality condition (5.4) of the prediction step, we have

$$\big(F(\bar{\mu}_h^n) + M(\bar{\mu}_h^n - \mu_h^n), \nu_h - \bar{\mu}_h^n\big) \ge 0,\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \tag{5.22}$$

$$\big(F(\bar{\mu}_h^{n+1}) + M(\bar{\mu}_h^{n+1} - \mu_h^{n+1}), \nu_h - \bar{\mu}_h^{n+1}\big) \ge 0,\ \forall\ \nu_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d). \tag{5.23}$$

Taking $\nu_h = \bar{\mu}_h^{n+1}$ in (5.22) and $\nu_h = \bar{\mu}_h^n$ in (5.23), and adding them together, we have

$$\big(M\big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big), \bar{\mu}_h^n - \bar{\mu}_h^{n+1}\big) \ge 0. \tag{5.24}$$

Adding the term

$$\big(M\big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big), \big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big)\big)$$

to both sides of (5.24), it yields

$$\begin{aligned}&\big(M\big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big), \mu_h^n - \mu_h^{n+1}\big)\\ &\ge \big(M\big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big), \big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big)\big),\end{aligned} \tag{5.25}$$

together with which, it follows from (5.3) that

$$\begin{aligned}&\big(H\big((\mu_h^n - \mu_h^{n+1}) - (\mu_h^{n+1} - \mu_h^{n+2})\big), \mu_h^n - \mu_h^{n+1}\big)\\ &\ge \gamma\big(M\big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big), \big((\mu_h^n - \bar{\mu}_h^n) - (\mu_h^{n+1} - \bar{\mu}_h^{n+1})\big)\big).\end{aligned} \tag{5.26}$$

Now we use the identity

$$\|a\|_H^2 - \|b\|_H^2 = 2(H(a - b), a) - \|a - b\|_H^2$$

with $a = \mu_h^n - \mu_h^{n+1}$ and $b = \mu_h^{n+1} - \mu_h^{n+2}$ and thus get

$$\begin{aligned}&\|\mu_h^n - \mu_h^{n+1}\|_H^2 - \|\mu_h^{n+1} - \mu_h^{n+2}\|_H^2\\ &= 2\big(H\big((\mu_h^n - \mu_h^{n+1}) - (\mu_h^{n+1} - \mu_h^{n+2})\big), \mu_h^n - \mu_h^{n+1}\big)\\ &\quad - \|(\mu_h^n - \mu_h^{n+1}) - (\mu_h^{n+1} - \mu_h^{n+2})\|_H^2.\end{aligned} \tag{5.27}$$

Then, it follows from (5.26), (5.27) and (5.3) that

(5.28)
$$\begin{aligned}
&\|\mu_h^n-\mu_h^{n+1}\|_H^2-\|\mu_h^{n+1}-\mu_h^{n+2}\|_H^2\\
&\geq 2\gamma\big(M\big((\mu_h^n-\bar{\mu}_h^n)-(\mu_h^{n+1}-\bar{\mu}_h^{n+1})\big),\big((\mu_h^n-\bar{\mu}_h^n)-(\mu_h^{n+1}-\bar{\mu}_h^{n+1})\big)\big)\\
&\quad-\|(\mu_h^n-\mu_h^{n+1})-(\mu_h^{n+1}-\mu_h^{n+2})\|_H^2\\
&=2\gamma\big(M\big((\mu_h^n-\bar{\mu}_h^n)-(\mu_h^{n+1}-\bar{\mu}_h^{n+1})\big),\big((\mu_h^n-\bar{\mu}_h^n)-(\mu_h^{n+1}-\bar{\mu}_h^{n+1})\big)\big)\\
&\quad-\gamma^2\big(M\big((\mu_h^n-\bar{\mu}_h^n)-(\mu_h^{n+1}-\bar{\mu}_h^{n+1})\big),H^{-1}M\big((\mu_h^n-\bar{\mu}_h^n)-(\mu_h^{n+1}-\bar{\mu}_h^{n+1})\big)\big).
\end{aligned}$$

We can show that the right-hand side term of (5.28) is nonnegative, just as the same approach in (5.13). The result (5.21) is thus proved. ◻

Next we establish the worst-case $O(\frac{1}{N})$ convergence rate in a nonergodic sense for Algorithm 2. We summarize the result in the following theorem.

THEOREM 5.6 (Convergence rate in a nonergodic sense). *Let μ_h be the solution of* (1.8) *and the sequence $\{\mu_h^{n+1}\}$ be generated by Algorithm 2 with $\theta\in[-1,1]$ and $\gamma\in(0,1]$ under the condition* (5.5). *Then for any integer $N>0$, it holds*

$$\|\mu_h^N-\mu_h^{N+1}\|_H^2\leq\frac{1}{r(N+1)}\|\mu_h-\mu_h^0\|_H^2, \tag{5.29}$$

where r is

$$r=\frac{1}{4}\Big(1-\frac{\tau^2\|\nabla\|^2}{\sigma}\Big)>0.$$

Proof. It follows from (5.15) that

$$r\|\mu_h^n-\mu_h^{n+1}\|_H^2\leq\|\mu_h-\mu_h^n\|_H^2-\|\mu_h-\mu_h^{n+1}\|_H^2. \tag{5.30}$$

Summarizing the inequalities (5.30) for the cases $n=0,\cdots,N$, we have

$$r\sum_{n=0}^{N}\|\mu_h^n-\mu_h^{n+1}\|_H^2\leq\|\mu_h-\mu_h^0\|_H^2-\|\mu_h-\mu_h^{N+1}\|_H^2. \tag{5.31}$$

From the result (5.21) of Lemma 5.5, we know that $\|\mu_h^n-\mu_h^{n+1}\|_H^2$ is monotonically non-increasing. Therefore, it yields

$$(N+1)\|\mu_h^N-\mu_h^{N+1}\|_H^2\leq\sum_{n=0}^{N}\|\mu_h^n-\mu_h^{n+1}\|_H^2. \tag{5.32}$$

Then the assertion (5.29) is obtained with (5.31) and (5.32). ◻

6. Numerical Examples. In this section, we report some preliminary numerical results to show the efficiency of the proposed algorithms. The rationale of considering the general primal-dual scheme (1.10) and the new primal-dual-based prediction-correction scheme is thus verified. Our main purpose is to illustrate: 1) the combination factor $\theta\neq1$ sometimes can accelerate the convergence of Algorithm 1 with $\theta=1$; and 2) Algorithm 2 with a relaxed requirement on τ could be numerically faster than Algorithm 1.

All codes were written in C++ based on the finite element library `AFEPack` [28] and all experiments were run on a Linux workstation with a ten core Intel 3.0GHz dual Processors and 128GB Memory. The stopping criterion for implementing Algorithm 1 and 2 is throughout chosen as

$$\frac{\|u_h^{n+1} - u_h^n\|_{L^2(\Omega)}}{\|u_h^{n+1}\|_{L^2(\Omega)}} \leq \texttt{Tol},$$

with the tolerance $\texttt{Tol} > 0$.

6.1. Numerical Results for Algorithm 1. We first test Example 6.1, which is an example similar as the one in [7], to show the numerical efficiency of Algorithm 1 with $\theta \neq 1$.

EXAMPLE 6.1. *Let* $\Omega = (-0.5, 0.5)^2$, $\alpha = 200$, *and* $g = g_0 + \delta\xi_h$ *with* $g_0 = \iota_{B(0,r)}$, *which is the indicator function of* $B(0,r) = \{x \in \Omega : \ |x| \leq r\}$, $r = 0.2$, *and* ξ_h *is a mesh-dependent perturbation function.*

To see the effectiveness of the combination factor θ, we fix the other parameters as $\sigma = 1.0$ and $\delta = 0.1$; and the tolerance in the stopping criterion is set as $\texttt{Tol} = 1.0 \times 10^{-4}$ for the experiments for Example 6.1. The domain Ω is partitioned by the triangulation mesh $\mathcal{T}_h$ with 1110 nodes and 2090 elements, as shown in the left part of Figure 1. The right part of Figure 1 shows the plot of g_0 over the mesh $\mathcal{T}_h$. The discretized function $g_h \in \mathcal{L}^0(\mathcal{T}_h)$ is defined by $g_h|_T = g_0(x_T) + \delta\xi_h|_T$ for each $T \in \mathcal{T}_h$, where the perturbation function $\xi_h \in \mathcal{L}^0(\mathcal{T}_h)$ is a normally distributed random value in each element $T \in \mathcal{T}_h$. The initial guess u_h^0 is taken as the projection of function g_h onto the finite element space $\mathcal{S}^1(\mathcal{T}_h)$, and p_h^0 is initialized as zero function.

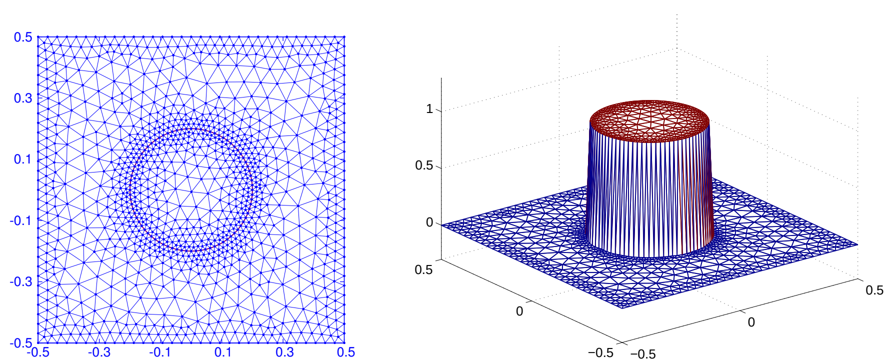

FIG. 1. *Triangular mesh over* Ω *(left) and the function* g_0 *(right) for Example 6.1.*

To implement Algorithm 1, obeying the condition (4.7), we choose the step size τ as

$$\tau = 2\frac{\sigma}{\|\nabla\|^2} \Big/ \left(\frac{(1-\theta)^2}{2\alpha} + \sqrt{\frac{(1-\theta)^4}{4\alpha^2} + 4\theta^2 \frac{\sigma}{\|\nabla\|^2}} \right)$$

and $1/\|\nabla\|^2$ is about 1.0×10^{-5} to ensure the convergence of all $\theta \in [-1, 1]$. We test the cases of $\theta \in [-1, 1]$ with an equal distance of 0.1 and plot the iteration numbers and the values of the step size τ in Figure 2. From this figure, we see that some cases of $\theta \in [-0.1, 1)$ require less iterations. Thus the numerical efficiency of Algorithm 1 with $\theta \neq 1$ is demonstrated. We can also observe in Figure 2 that the value of τ plays a key role for the total iteration number of Algorithm 1 under the same stopping criterion, a larger value of τ satisfying the convergence condition turns out to be with less iteration number. It should be noticed that the largest step size τ to guarantee the convergence for $\theta \in [-1, 1)$ can be larger than that of $\theta = 1.0$ when the mesh size h is not small enough, although $\tau = O(h)$ for $\theta = 1$ and $\tau = O(h^2)$ for $\theta \in [-1, 1)$ are required to guarantee the convergence, just as it is showed in Figure 2 that the step size τ with $\theta \in [-0.1, 1)$ is larger than that with $\theta = 1.0$.

In Figure 3, the energy $E(u_h^n)$ at the iterations u_h^n for the cases of $\theta = 0.3$ and 1.0 with their corresponding step size τ for Example 6.1 are plotted. This figure shows that the energy for the case with $\theta = 0.3$ decreases more quickly than that with $\theta = 1.0$. It is thus again verified to consider the general primal-dual scheme: Algorithm 1 with $\theta \neq 1$. We show the iterations u_h^n for the cases of $\theta = 0.3$ and 1.0 in Figure 4; the cases with $n = 0, 20, 40, n^{stop}$ are listed from top to bottom.

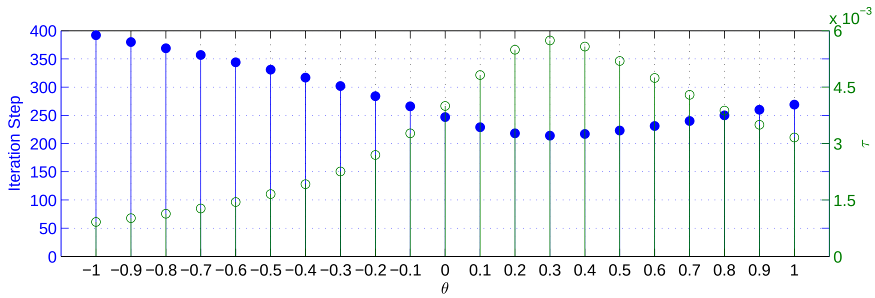

FIG. 2. *Iteration step number ('•') and step size τ ('∘') with different θ for Example 6.1.*

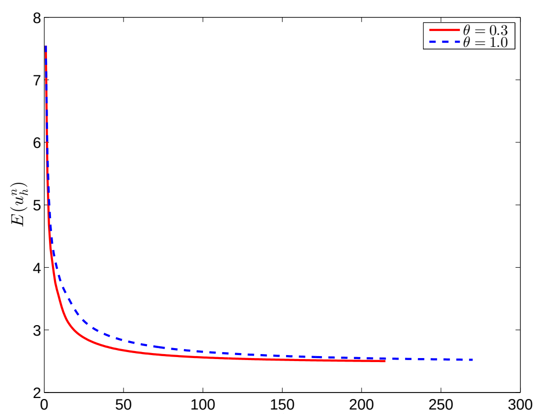

FIG. 3. *Energy $E(u_h^n)$ of iterations u_h^n for $\theta = 0.3$ and 1.0 with their corresponding τ for Example 6.1.*

6.2. Numerical Results for Algorithm 2. Then, we consider Example 6.2 to verify the efficiency of Algorithm 2. Its efficiency is demonstrated by comparison with Algorithm 1.

EXAMPLE 6.2. *Let $\alpha = 400$, $g = g_0 + \delta\xi_h$ and ξ_h be a mesh-dependent perturbation function, where g_0 is the solution at $t = 1.0$ of the 2D Allen-Cahn equation [1] over*

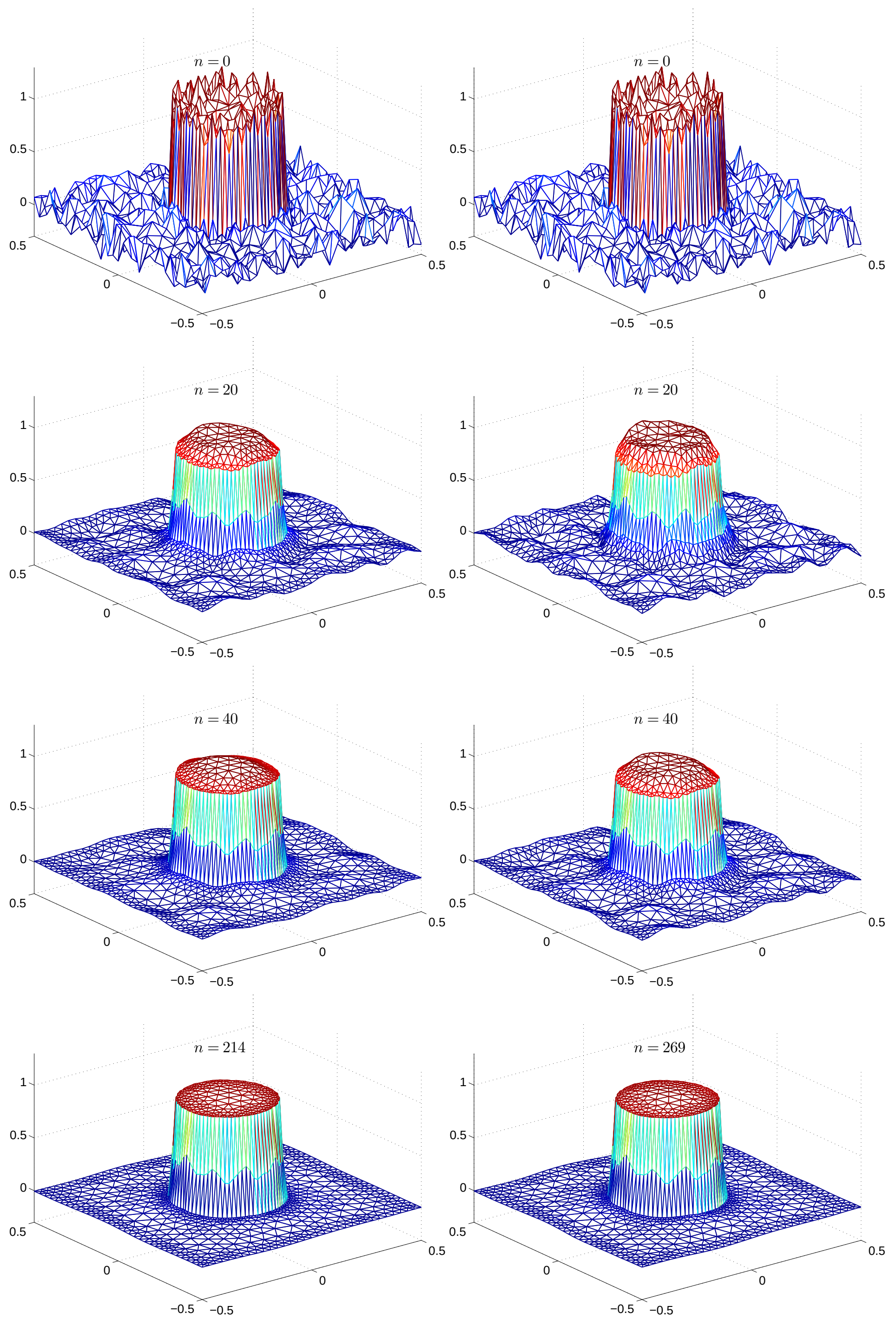

FIG. 4. *Iterations* u_h^n *for* $\theta = 0.3$ *(left) and* 1.0 *(right) for Example 6.1 with* $n = 0, 20, 40, n^{stop}$ *from top to bottom.*

$\Omega = (0,1)^2$ *subjected to periodic boundary condition,*

$$\partial_t u = D\big(\epsilon \Delta u - \frac{1}{\epsilon} F'(u)\big), \tag{6.1}$$

where $F(u) = \frac{1}{4}(u^2-1)^2$*, the initial value is taken as the following random initial value*

$$u(x,0) = 0.05(2\,\textit{rand} - 1).$$

We obtain g_0 by solving the Allen-Cahn equation (6.1), and the image of g_0 over the mesh $\mathcal{T}_h$ with $10,201$ nodes and $20,000$ elements is shown in Figure 5. The perturbation function $\xi_h \in \mathcal{S}^1(\mathcal{T}_h)$ evaluated at each node of mesh $\mathcal{T}_h$ is a random value sampled from the normally distribution, the noise level δ is 0.2, the initial guess u_h^0 is set by function g_h and p_h^0 is chosen as 0.0.

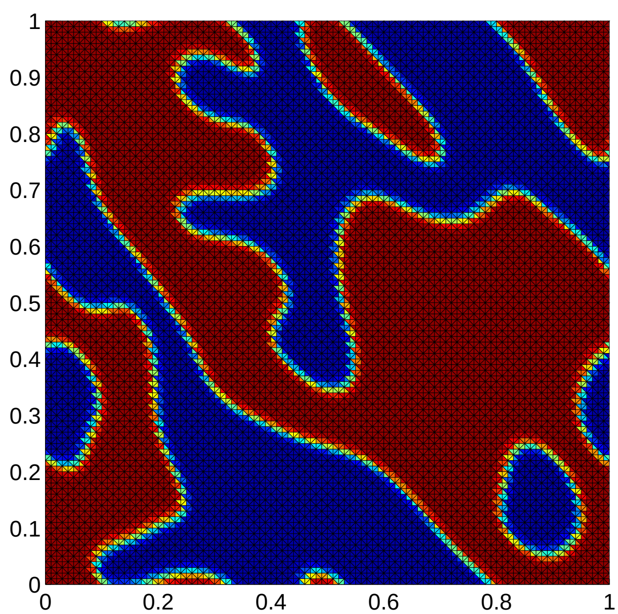

FIG. 5. *The function g_0 over the uniformly triangular mesh with square edge length* $1/100$.

We fix $\sigma = 0.1$ and $\gamma = 1.0$ for Algorithm 1 and 2.

In Table 1, the iteration numbers and CPU time in seconds when Algorithms 1 and 2 are applied to Example 6.2 are listed for the cases of $\theta \in [-0.9, -0.2]$ with an equal distance of 0.1, where the tolerance in the stopping criterion is set as $\texttt{Tol} = 1.0 \times 10^{-4}$. According to the table, it seems that Algorithm 1 is more sensitive to the step size τ near the critical one certifying the convergence. We can also observe that the largest step size τ for guaranteeing the convergence of Algorithm 1 is much smaller than that of Algorithm 2. This coincides with the necessary requirements of the step size with $\tau \leq ch^2$ for Algorithm 1 and $\tau \leq ch$ for Algorithm 2, respectively satisfying our theoretical conditions (4.7) and (5.5). The comparison of iteration numbers and CPU time in seconds is displayed in Figure 6. Moreover, we test more cases with $\theta = -0.2, -0.4, -0.6, -0.8$ with more values of the step size τ; and plot the results in Figure 7. This figure shows that the iteration numbers and the CPU time of Algorithm 1 are first decreasing and then increasing once it is convergent; while those of Algorithm 2 are monotonically increasing with small values of the step size τ. In Figure 8, we plot the input u_h^0 and outputs (first row: input u_h^0; second row: output by Algorithm 1; third row: output by Algorithm 2) when $\theta = -0.4$ (left) and -0.8 (right) for Example 6.2 with $\texttt{Tol} = 1.0 \times 10^{-5}$.

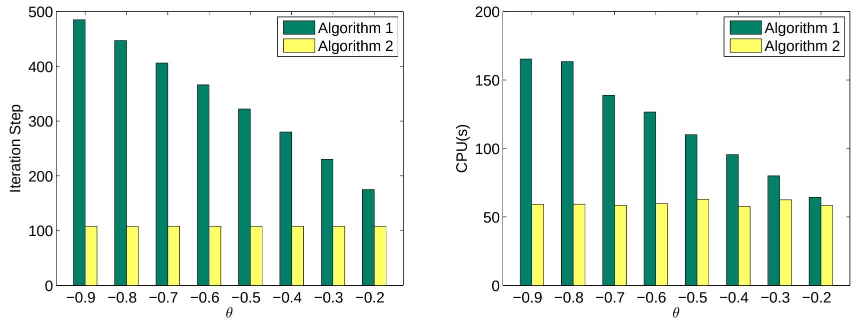

Fig. 6. *The iteration step (left) and CPU time (right) in seconds for Example 6.2 by Algorithm 1 and Algorithm 2 with $\theta = -0.9, \cdots, -0.2$ and their corresponding step size τ marked in Table 1.*

7. Conclusions. In this paper, we focus on the application of primal dual schemes to the saddle-point reformulation of the total variation minimization model with consistent finite element discretization and study its convergence. We first generalize the primal-dual scheme in [7], and then prove its convergence and establish its worst-case convergence rate measured by the iteration complexity. Then, we propose a new primal-dual scheme in the prediction-correction framework, whose necessary requirement of the step size τ satisfying convergence condition can be significantly relaxed. This new primal-dual-based prediction-correction scheme enjoys the same convergence turns out to have both stronger convergence rate and better numerical performance than the generalized primal-dual scheme. We report some preliminary numerical results to verify the theoretical assertions. It is interesting to extend our analysis to some more complicated total variational minimization models (such as [13]). We leave it as our future work.

REFERENCES

[1] S. M. Allen and J. W. Cahn, *A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening*, Acta Metall., 27 (1979), pp. 1085–1095.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, The Clarendon Press, Oxford University Press, New York, 2000.

[3] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón, *Minimizing total variation flow*, Differential Integral Equations, 14 (2001), pp. 321–360.

[4] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*, Stanford University Press, Stanford, 1958.

[5] H. Attouch, G. Buttazzo, and G. Michaille, *Variational analysis in Sobolev and BV spaces: Applications to PDEs and Optimization*, SIAM/MPS, Philadelphia, PA, 2006.

[6] G. Aubert and P. Kornprobst, *Mathematical problems in image processing: Partial differential equations and the calculus of variations*, Springer, New York, 2nd ed., 2006.

[7] S. Bartels, *Total variation minimization with finite elements: convergence and iterative solution*, SIAM J. Numer. Anal., 50 (2012), pp. 1162–1180.

TABLE 1

Iteration step numbers and CPU time in seconds obtained by Algorithm 1 and 2 for Example 6.2 with $\theta = -0.2, \cdots, -0.9$ and different step size τ ('-' means not convergent).

$\theta = -0.2$					$\theta = -0.3$					$\theta = -0.4$					$\theta = -0.5$				
	Alg 1		Alg 2			Alg 1		Alg 2			Alg 1		Alg 2			Alg 1		Alg 2	
τ	step	CPU(s)	step	CPU(s)	τ	step	CPU(s)	step	CPU(s)	τ	step	CPU(s)	step	CPU(s)	τ	step	CPU(s)	step	CPU(s)
1.0/1200.0	-	-	-	-	1.0/1200.0	-	-	-	-	1.0/1200.0	-	-	-	-	1.0/1200.0	-	-	-	-
1.0/1600.0	-	-	**108**	**58.29**	1.0/1600.0	-	-	**108**	**62.46**	1.0/1600.0	-	-	**108**	**57.73**	1.0/1600.0	-	-	**108**	**62.87**
1.0/2000.0	393	135.52	118	64.95	1.0/2000.0	-	-	118	63.71	1.0/2400.0	-	-	126	68.69	1.0/2400.0	-	-	126	69.27
1.0/2400.0	218	75.40	126	69.68	1.0/2400.0	-	-	126	70.24	1.0/3200.0	-	-	141	77.87	1.0/3200.0	-	-	141	81.47
1.0/2800.0	186	63.42	133	72.48	1.0/3200.0	362	124.14	141	77.32	1.0/4000.0	583	207.01	154	88.96	1.0/4800.0	947	349.12	166	92.03
1.0/3200.0	178	65.38	141	79.84	1.0/3600.0	281	96.70	147	84.97	1.0/4800.0	345	124.00	166	92.28	1.0/6000.0	400	139.61	183	100.44
1.0/3600.0	**175**	**64.26**	147	81.15	1.0/4000.0	250	85.74	154	83.90	1.0/5600.0	289	105.39	177	95.84	1.0/7000.0	336	115.73	197	106.40
1.0/4000.0	179	62.54	154	84.56	1.0/4400.0	235	81.73	160	87.09	1.0/6400.0	283	96.38	189	104.57	1.0/8000.0	326	113.74	212	122.96
1.0/4400.0	183	63.84	160	86.94	1.0/4800.0	231	79.62	166	92.00	1.0/7200.0	**280**	**95.51**	200	109.72	1.0/9000.0	**322**	**110.06**	227	126.07
1.0/4800.0	189	65.47	166	89.86	1.0/5200.0	231	80.65	171	92.94	1.0/8000.0	284	98.12	212	117.38	1.0/10000.0	333	114.65	241	132.97
1.0/5200.0	195	66.34	171	92.44	1.0/5600.0	**230**	**79.93**	177	96.94	1.0/8800.0	293	102.50	224	123.43	1.0/11000.0	338	115.54	256	141.70
1.0/5600.0	201	70.24	177	97.78	1.0/6000.0	234	85.66	183	98.37	1.0/9600.0	302	105.02	235	127.96	1.0/12000.0	350	121.81	270	149.74
1.0/6000.0	207	71.01	183	98.99	1.0/6400.0	238	82.21	189	101.05	1.0/10400.0	311	110.49	247	136.77	1.0/13000.0	363	125.27	285	158.67
1.0/6400.0	215	73.98	189	103.02	1.0/6800.0	242	83.84	194	105.82	1.0/11200.0	320	110.52	258	142.58	1.0/14000.0	376	136.61	299	159.98
1.0/6800.0	220	76.74	194	112.95	1.0/7200.0	245	85.54	200	109.57	1.0/12000.0	332	115.13	270	148.44	1.0/15000.0	386	131.68	313	175.99

$\theta = -0.6$					$\theta = -0.7$					$\theta = -0.8$					$\theta = -0.9$				
	Alg 1		Alg 2			Alg 1		Alg 2			Alg 1		Alg 2			Alg 1		Alg 2	
τ	step	CPU(s)	step	CPU(s)	τ	step	CPU(s)	step	CPU(s)	τ	step	CPU(s)	step	CPU(s)	τ	step	CPU(s)	step	CPU(s)
1.0/1200.0	-	-	-	-	1.0/1200.0	-	-	-	-	1.0/1200.0	-	-	-	-	1.0/1200.0	-	-	-	-
1.0/1600.0	-	-	**108**	**59.78**	1.0/1600.0	-	-	**108**	**58.43**	1.0/1600.0	-	-	**108**	**59.31**	1.0/1600.0	-	-	**109**	**59.26**
1.0/3800.0	-	-	151	82.93	1.0/4400.0	-	-	160	93.16	1.0/4800.0	-	-	166	91.34	1.0/5400.0	-	-	174	95.75
1.0/6000.0	741	255.81	183	100.99	1.0/7200.0	724	252.39	200	109.95	1.0/8000.0	899	329.49	212	115.44	1.0/9200.0	875	299.85	229	126.58
1.0/7200.0	453	155.69	200	112.47	1.0/8400.0	500	171.25	218	122.08	1.0/9200.0	592	216.13	229	133.29	1.0/10400.0	621	213.72	247	135.74
1.0/8400.0	385	134.80	218	126.61	1.0/9600.0	426	151.14	235	129.04	1.0/10400.0	493	182.97	247	139.84	1.0/11600.0	532	185.68	264	151.20
1.0/9600.0	372	136.19	235	131.26	1.0/10800.0	418	147.46	253	137.61	1.0/11600.0	454	161.17	264	154.12	1.0/12800.0	498	169.71	282	153.63
1.0/10800.0	**366**	**126.64**	253	136.80	1.0/12000.0	**406**	**138.78**	270	149.38	1.0/12800.0	451	152.97	282	155.04	1.0/14000.0	493	170.16	299	166.52
1.0/12000.0	378	131.03	270	148.80	1.0/13200.0	416	142.28	288	157.94	1.0/14000.0	**447**	**163.49**	299	165.23	1.0/15200.0	**485**	**165.28**	316	175.05
1.0/13200.0	386	133.08	288	158.34	1.0/14400.0	422	144.99	305	169.28	1.0/15200.0	455	155.25	316	173.73	1.0/16400.0	490	166.73	333	181.41
1.0/14400.0	399	144.95	305	169.65	1.0/15600.0	432	149.07	322	176.74	1.0/16400.0	462	159.97	333	180.33	1.0/17600.0	495	169.43	350	191.56
1.0/15600.0	412	141.22	322	179.41	1.0/16800.0	444	154.93	339	184.93	1.0/17600.0	473	164.95	350	198.36	1.0/18800.0	505	171.61	367	201.19
1.0/16800.0	426	144.89	339	196.98	1.0/18000.0	456	157.07	355	196.11	1.0/18800.0	481	163.79	367	198.40	1.0/20000.0	511	173.94	384	211.20
1.0/18000.0	442	156.65	355	193.74	1.0/19200.0	471	160.70	372	209.25	1.0/20000.0	497	174.61	384	212.97	1.0/21200.0	527	183.21	400	232.90
1.0/19200.0	455	166.20	372	205.89	1.0/20400.0	483	165.71	389	215.05	1.0/21200.0	506	174.21	400	232.42	1.0/22400.0	534	183.35	416	222.16

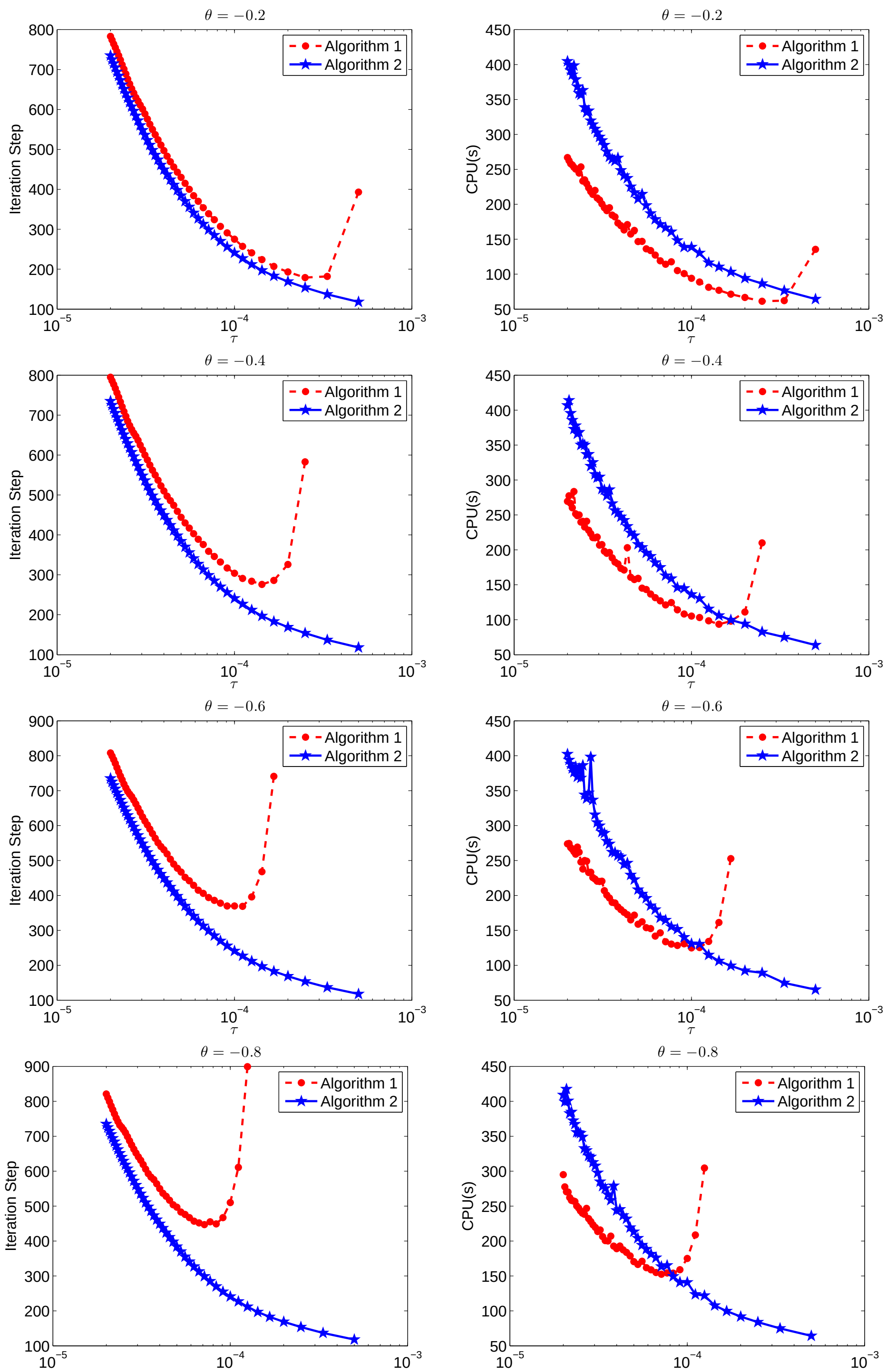

FIG. 7. *The iteration step (left) and CPU time (right) in seconds for Example 6.2 by Algorithm 1 and Algorithm 2 with* $\theta = -0.2, -0.4, -0.6, -0.8$ *and* $\tau =$ `1./[1000:1000:50000]`.

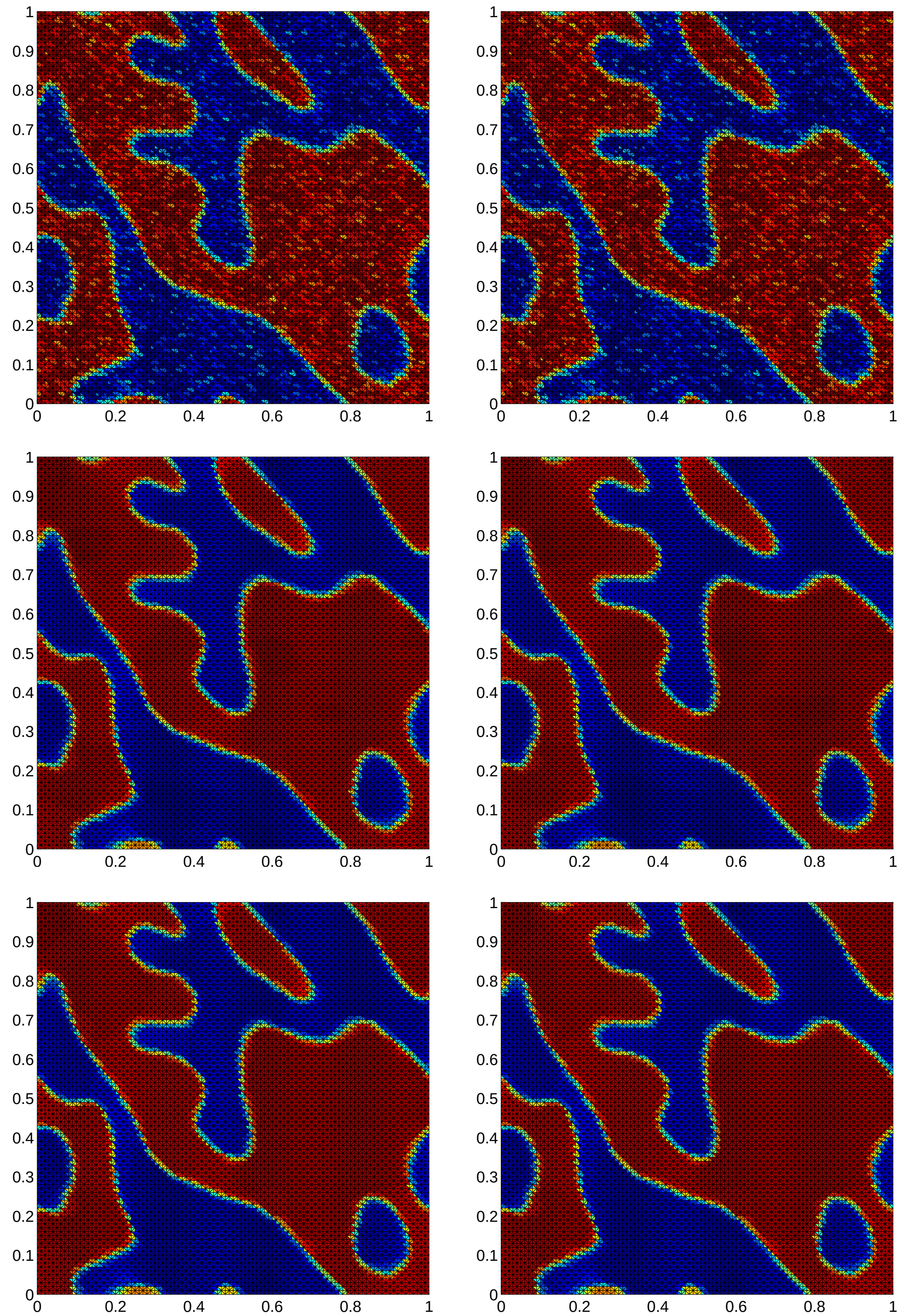

FIG. 8. *Input u_h^0 and Outputs for $\theta = -0.4$ (left) and -0.8 (right) for Example 6.2 with noise level $\delta = 0.2$ and* `Tol` *$= 1.0 \times 10^{-5}$. (First row: input u_h^0; Second row: Output by Algorithm 1; Third row: Output by Algorithm 2)*

[8] S. Bartels, R. H. Nochetto, and A. J. Salgado, *Discrete total variation flows without regularization*, SIAM J. Numer. Anal., 52 (2014), pp. 363–385.
[9] E. Blum and W. Oettli, *Mathematische Optimierung*, Econometrics and Operations Research XX, Springer-Verlag, 1975.
[10] J. H. Bramble, J. E. Pasciak, and A. T. Vassilev, *Analysis of the inexact Uzawa algorithm for saddle point problems*, SIAM J. Numer. Anal., 34 (1997), pp. 1072–1092.
[11] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, Springer, New York, 3rd ed., 2008.
[12] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.
[13] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
[14] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
[15] R. H. Chan and K. Chen, *A multilevel algorithm for simultaneously denoising and deblurring images*, SIAM J. Sci. Comput., 32 (2010), pp. 1043–1063.
[16] 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.
[17] T. F. Chan and P. Mulet, *On the convergence of the lagged diffusivity fixed point method in total variation image restoration*, SIAM J. Numer. Anal., 36 (1999), pp. 354–367.
[18] D. C. Dobson and C. R. Vogel, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal., 34 (1997), pp. 1779–1791.
[19] H. C. Elman and G. H. Golub, *Inexact and preconditioned Uzawa algorithms for saddle point problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1645–1661.
[20] 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.
[21] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Vol. I*, Springer-Verlag, New York, 2003.
[22] X. Feng and A. Prohl, *Analysis of total variation flow and its finite element approximations*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 533–556.
[23] X. Feng, M. von Oehsen, and A. Prohl, *Rate of convergence of regularization procedures and finite element approximations for the total variation flow*, Numer. Math., 100 (2005), pp. 441–456.
[24] M. Fortin and R. Glowinski, *Augmented Lagrangian methods: Applications to the numerical solution of boundary value problems*, North-Holland Publishing Co., Amsterdam, 1983.
[25] R. Glowinski, *Numerical methods for nonlinear variational problems*, Springer-Verlag, New York, 1984.
[26] B. S. He and X. M. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci., 5 (2012), pp. 119–149.
[27] ———, *On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.
[28] R. Li, *AFEPack*. `http://dsec.pku.edu.cn/~rli/`.
[29] A. Marquina and S. Osher, *Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal*, SIAM J. Sci. Comput., 22 (2000), pp. 387–405.
[30] A. S. Nemirovsky and D. B. Yudin, *Problem complexity and method efficiency in optimization*, John Wiley & Sons, Inc., New York, 1983. Wiley-Interscience Series in Discrete Mathematics.
[31] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.
[32] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in Computer Vision (ICCV), 2011 IEEE International Conference on, Nov 2011, pp. 1762–1769.
[33] W. Queck, *The convergence factor of preconditioned algorithms of the Arrow-Hurwicz type*, SIAM J. Numer. Anal., 26 (1989), pp. 1016–1030.
[34] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[35] G. Sapiro and V. Caselles, *Histogram modification via differential equations*, J. Differential Equations, 135 (1997), pp. 238–268.
[36] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.

[37] J. Zhang, K. Chen, and B. Yu, *An iterative Lagrange multiplier method for constrained total-variation-based image denoising*, SIAM J. Numer. Anal., 50 (2012), pp. 983–1003.
[38] X. Zhang, M. Burger, and S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46.
[39] 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.
[40] W. P. Ziemer, *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*, Springer-Verlag, New York, 1989.
[41] W. Zulehner, *Analysis of iterative methods for saddle point problems: a unified approach*, Math. Comp., 71 (2002), pp. 479–505.