An Accelerated Primal-Dual Iterative Scheme for the L^2-TV Regularized Model of Linear Inverse Problems

Wenyi Tian* Xiaoming Yuan†

Abstract

A model with the L^2 and total variational (TV) regularization terms for linear inverse problems is considered. The regularized model is reformulated as a saddle-point problem, and the primal and dual variables are discretized in the piecewise affine and piecewise constant finite element spaces, respectively. An accelerated primal-dual iterative scheme with an $O(\frac{1}{N^2})$ convergence rate is proposed for the discretized problem, where N is the iteration counter. Both the regularization and perturbation errors of the regularized model, and the finite element discretization and iteration errors of the accelerated primal-dual scheme, are estimated. Preliminary numerical results are reported to show the efficiency of the proposed iterative scheme.

Keywords: Linear inverse problem, Primal-dual method, Saddle-point problem, Finite element method, Convergence rate, Error estimate.

1 Introduction

We consider a possibly ill-posed linear inverse problem

$$Au = g, \tag{1.1}$$

where $A : L^2(\Omega) \to L^2(\Omega)$ is a bounded linear operator, $\Omega \subset \mathbb{R}^d$ is a bounded domain with a Lipschitz continuous boundary, $d = 1, 2$, and $g \in L^2(\Omega)$ is a given function. This problem is fundamental in a variety of scientific computing areas, such as astrophysics, signal and image processing, statistical inference, and optics. We refer to, e.g., the monographs [18, 21, 23, 40] and the references therein.

To solve the ill-conditioned problem (1.1), the regularization technique is a standard approach. Among different choices of the regularization term, the total variational regularization has been widely used to recover solutions with discontinuities to inverse problems, especially since the seminal works [22, 34]. As studied in various literature (e.g., [1, 10, 11, 16, 33, 41]), the solution of (1.1) can be approximated by the following minimization model with TV regularization:

$$\inf_u \left\{ \frac{1}{2}\|Au - g\|^2_{L^2(\Omega)} + \alpha\|Du\| \right\}, \tag{1.2}$$

*Center for Applied Mathematics, Tianjin University, Tianjin 300072, China. This author was partially supported by the National Natural Science Foundation of China (No. 11701416). Email: twymath@gmail.com

†Department of Mathematics, The University of Hong Kong, Hong Kong, China. This author was partially supported by the General Research Fund from Hong Kong Research Grants Council: 12300515. Email: xmyuan@hku.hk

where $\alpha > 0$ is a regularization parameter determining the relative weight of the involved two terms in (1.2). The regularization term $\|Du\|$ in (1.2) is the TV norm of u defined by

$$\|Du\| := \sup\left\{ \int_\Omega u \,\mathrm{div}\varphi \,\mathrm{d}x : \varphi \in C_c^1(\Omega;\mathbb{R}^d), \|\varphi\|_\infty \le 1 \right\}, \tag{1.3}$$

where $\|\varphi\|_\infty = \sup_{x\in\Omega}(\sum_{i=1}^d |\varphi_i(x)|^2)^{1/2}$, Du represents the gradient of u in the distributional sense, div denotes the divergence operator, and $C_c^1(\Omega;\mathbb{R}^d)$ is the set of once continuously differentiable $\mathbb{R}^d$-valued functions with compact support in Ω. The $BV(\Omega)$ space endowed with the norm $\|v\|_{BV} := \|v\|_{L^1(\Omega)} + \|Dv\|$ is a Banach space; see, e.g., [2, 4, 5, 45] for more details. Discussions on the existence and uniqueness of the solution of (1.2) can be found in, e.g., [1, 10, 14, 41].

In [14, 26], the following minimization model with both the L^2 and TV regularization terms is studied:

$$\inf_u \left\{ E(u) := \frac{1}{2}\|Au - g\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L^2(\Omega)} + \alpha\|Du\| \right\}, \tag{1.4}$$

with $u \in L^2(\Omega) \cap BV(\Omega)$ and $\beta > 0$. As mentioned in [14], adding the L^2 quadratic regularization term is meaningful. First, in (1.4), the objective function is strongly convex and thus the solution of (1.4) is unique without additional assumptions of the corresponding operator A. Second, the model (1.4) has the possibility of distinguishing the structure of stability results for quadratic regularization terms from that of the non-quadratic BV-term. Third, as stated in [14, Corollary 1.2], $L^2(\Omega) \cap BV(\Omega)$ and $BV(\Omega)$ are equivalent Banach spaces when $d = 1$ or 2; and thus the model (1.4) can keep favorable characters of the TV regularization.

In the models (1.2) and (1.4), the TV term is beneficial for reconstructing discontinuous or piecewise constant solutions of (1.1). Meanwhile, the TV term causes some difficulties, one of which is that this nonsmooth term makes it more challenging to design efficient iterative schemes. It is noted that some authors have suggested replacing the TV term with the smoothing surrogate $\int_\Omega \sqrt{|\nabla u|^2 + \varepsilon}$ with small $\varepsilon > 0$; see, e.g., [1, 41]. Here, we follow the approach in [6, 7, 12] to consider the saddle-point reformulation of the minimization model (1.4):

$$\begin{aligned} \inf_u E(u) &= \inf_u \sup_p \Big\{ \mathcal{E}(u,p) \\ &:= \frac{1}{2}\|Au - g\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L^2(\Omega)} + \alpha \int_\Omega \nabla u \cdot p \,\mathrm{d}x - I_B(p) \Big\}, \end{aligned} \tag{1.5}$$

where $B = \{p \in L^1(\Omega;\mathbb{R}^d) : \|p\|_\infty \le 1\}$ and $I_B(\cdot)$ denotes its indicator function.

We further choose the piecewise constant and piecewise affine globally continuous finite element spaces to discretize the functions u and p in (1.5), respectively. That is, we have

$$\begin{cases} \mathcal{S}^1(\mathcal{T}_h) \;\; := \{v_h \in C(\bar{\Omega}) : v_h|_T \text{ is affine for each } T \in \mathcal{T}_h\}, \\ \mathcal{L}^0(\mathcal{T}_h)^d := \{q_h \in L^1(\Omega;\mathbb{R}^d) : q_h|_T \text{ is constant for each } T \in \mathcal{T}_h\}, \end{cases} \tag{1.6}$$

where $\mathcal{T}_h$ denotes as a regular triangulation of Ω into triangles and $h = \max_{T\in\mathcal{T}_h} \mathrm{diam}(T)$ as the maximal diameter. Such choices are preferred because of some particular reasons as mentioned in [6]. First, both the spaces $\mathcal{S}^1(\mathcal{T}_h)$ and $\mathcal{L}^0(\mathcal{T}_h)$ are dense in $BV(\Omega)$ with respect to weak* convergence in $BV(\Omega)$, which is continuously embedded into $L^r(\Omega)$ for all r satisfying $1 \leq r \leq \frac{d}{d-1}$, see [4, Theorem 10.1.4]. Second, it was demonstrated in [6] that the piecewise constant finite element approximation for u cannot be expected to converge to an exact solution in general. Note that the space $\mathcal{L}^0(\mathcal{T}_h)^d$ consists of piecewise constant vector fields equipped with L^2 scalar product as

$$(p_h, q_h) = \sum_{i=1}^{d} \int_\Omega (p_h)_i (q_h)_i \mathrm{d}x, \tag{1.7}$$

where $(p_h)_i$ stands for the i-th component of the vector-valued function $p_h \in \mathcal{L}^0(\mathcal{T}_h)^d$, and d is the dimension of the domain Ω. The integral (1.7) of piecewise-constant functions on each element $T \in \mathcal{T}_h$ is computed by their values at the centroid of T. Furthermore, the identity

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

is mentioned in [6]. Therefore, the minimization model (1.4) with finite element approximation can be reformulated as the following discretized saddle-point problem

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

This is the model our analysis will be focused on hereafter.

Recently, some concrete progresses have been made for saddle-point problems, especially in convex optimization and image processing literatures. In particular, a number of primal-dual type iterative schemes have been studied by many authors; we refer to, e.g., [8, 12, 19, 24, 32, 36, 37, 44] to mention a few. It is worthy mentioning that these primal-dual type iterative schemes have a deep origin in the so-called inexact Uzawa method [3, 17]. As indicated in [36], primal-dual iterative schemes can be conceptually applicable to discretized saddle-point problems arising from inverse problems; but the resulting subproblems may be highly nontrivial and it deserves to look into the details for these applications. For the case where the regularization term for (1.1) is just the TV term, i.e., the special case of (1.8) with $\beta = 0$, a linearized primal-dual method is proposed in [36]. Its iterative scheme reads as

$$\begin{cases} u_h^{n+1} = \arg\min\limits_{u_h \in \mathcal{S}^1(\mathcal{T}_h)} \Big\{ \mathcal{E}(u_h, p_h^n) + \dfrac{1}{2\tau}\|u_h - u_h^n\|^2_{X, L^2(\Omega)} \Big\}, \\ \tilde{u}_h^n = 2u_h^{n+1} - u_h^n, \\ p_h^{n+1} = \arg\max\limits_{p_h \in \mathcal{L}^0(\mathcal{T}_h)^d} \Big\{ \mathcal{E}(\tilde{u}_h^n, p_h) - \dfrac{\sigma}{2\tau}\|p_h - p_h^n\|^2_{L^2(\Omega)} \Big\}, \end{cases} \tag{1.9}$$

where n is the iteration number; (u_h^n, p_h^n) are the output of the scheme (1.9) at the n-th iteration; $\|\cdot\|^2_{X,L^2(\Omega)} := (X\cdot, \cdot)$; $X := I - \tau A^* A$ in which I is the identity operator and

$\tau > 0$; and $\sigma > 0$. Note that τ and $\frac{\tau}{\sigma}$ can be understood as the step sizes for implementing certain gradient-based iterative methods for computing u_h^{n+1} and p_h^{n+1}, respectively.

In this paper, we study a primal-dual type iterative scheme for the model (1.4). Note that in [36] we proposed a linearized primal-dual iterative scheme for the model (1.2). Compared with (1.2), there is an additional term $\frac{\beta}{2}\|u_h\|_{L^2(\Omega)}^2$ in (1.4) which requires more sophisticated skills in algorithmic design. Meanwhile, the objective functional in (1.4) is strongly convex and thus it is possible to extend some recent works [12, 13] to the saddle-point problem (1.8). This motivates us to consider the following primal-dual iterative scheme:

$$\begin{cases} \tilde{u}_h^n = u_h^n + \theta_n(u_h^n - u_h^{n-1}), \\ p_h^{n+1} = \underset{p_h \in \mathcal{L}^0(\mathcal{T}_h)^d}{\arg\max} \left\{\mathcal{E}(\tilde{u}_h^n, p_h) - \frac{\sigma_n}{2\tau_n}\|p_h - p_h^n\|_{L^2(\Omega)}^2\right\}, \\ u_h^{n+1} = \underset{u_h \in \mathcal{S}^1(\mathcal{T}_h)}{\arg\min} \left\{\mathcal{E}(u_h, p_h^{n+1}) + \frac{1}{2\tau_n}\|u_h - u_h^n\|_{X,L^2(\Omega)}^2\right\}, \end{cases} \tag{1.10}$$

with $X := I - \tau A^*A$ and the parameters τ_n, σ_n and $\theta_n > 0$ satisfy certain conditions which will be specified later (see (3.2)). The main reason of considering the scheme (1.10) with dynamically adjusted parameters is that the convergence rate of this scheme measured by iteration complexity is $O(\frac{1}{N^2})$, while that of the linearized primal-dual scheme (1.9) for the model (1.2) is only $O(\frac{1}{N})$. Here, N is the iteration counter. This acceleration makes sense, particularly because the implementation of these two schemes is of the same level of difficulty (in sense of the difficulty of the resulting subproblems) and they are mainly different in choosing the involved parameters which is nearly computational free. Moreover, as we shall show, the strong convexity of the objective function in (1.4) leads to the error estimate of the finite element solution in Theorem 5.2; and together with the particular choices of the parameters in the iterative scheme (1.10), the strong convexity property enables us to derive the error estimate and convergence elucidated in Theorem 5.3. These properties are not achievable in [36] for the model (1.2) without strong convexity.

The rest of this paper is organized as follows. In Section 2, some known preliminary results are summarized. In Section 3, the accelerated primal-dual iterative scheme (1.10) is specified for the problem (1.8), along with some elaborations. In Section 4, we prove the convergence of the accelerated primal-dual iterative scheme (1.10) and establish its worst-case $O(\frac{1}{N^2})$ convergence rate measured by the iteration complexity. In Section 5, we estimate the regularization and perturbation errors for the model (1.4), and the finite element discretization and iteration errors for the accelerated primal-dual scheme (1.10). Some preliminary numerical results are reported in Section 6 to verify the effectiveness of the proposed method. Finally, some conclusions are made in Section 7.

2 Preliminary

In this section, we recall some known results in the literature for the convenience of further analysis. Throughout, the notation $(\cdot, \cdot)$ stands for the L^2 scalar product.

First, we present the existence and uniqueness results about the solution of the linear inverse problem (1.1) approximated by the L^2-TV regularization model (1.4), which is proved in [14].

Theorem 2.1 ([14])**.** *Let $\alpha > 0$. If one of the following conditions holds:*

(1) $\beta > 0$,

(2) $d = 1$ or $d = 2$, and nonzero constants do not belong to the null space $\ker A$ of A,

then there exists a solution of (1.4). Moreover, if $\beta > 0$ or A is injective, then the solution is unique.

The first-order optimality condition for the minimization of the energy functional $E(\cdot)$ in $\mathcal{S}^1(\mathcal{T}_h)$ is stated in the following lemma. The proof is followed by the similar argument for that of Lemma 10.3 in [7].

Lemma 2.1. *The function $u_h^* \in \mathcal{S}^1(\mathcal{T}_h)$ minimizes $E(\cdot)$ 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*

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

We can rewrite the optimality condition (2.1) as the following variational inequality in a compact form: finding $\boldsymbol{\mu}_h^* \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ such that

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

where

$$\boldsymbol{\mu}_h = \begin{pmatrix} u_h \\ p_h \end{pmatrix}, \quad \boldsymbol{\mu}_h^* = \begin{pmatrix} u_h^* \\ p_h^* \end{pmatrix}, \quad F(\boldsymbol{\mu}_h^*) = \begin{pmatrix} \beta u_h^* - \alpha \mathrm{div} p_h^* + A^*(Au_h^* - g) \\ -\alpha \nabla u_h^* \end{pmatrix}, \tag{2.3}$$

and $-\mathrm{div}$ is the conjugate operator of ∇ and $-(\mathrm{div} p_h, u_h) = (p_h, \nabla u_h)$. For any $\boldsymbol{\mu}_h = (u_h; p_h)$ and $\boldsymbol{\nu}_h = (v_h; q_h)$, obviously the mapping $F(\cdot)$ in (2.3) satisfies

$$\big(F(\boldsymbol{\mu}_h) - F(\boldsymbol{\nu}_h), \boldsymbol{\mu}_h - \boldsymbol{\nu}_h\big) = \beta\|u_h - v_h\|^2_{L^2(\Omega)} + \|A(u_h - v_h)\|^2_{L^2(\Omega)}. \tag{2.4}$$

It follows from (2.2), (2.4) and 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.

Remark 2.1. *An inverse estimate in [9] shows that there exists a constant $c > 0$ such that $\|\nabla u_h\|_{L^2(\Omega)} \le c h_{\min}^{-1} \|u_h\|_{L^2(\Omega)}$ for all $u_h \in \mathcal{S}^1(\mathcal{T}_h)$, where $h_{\min} = \min_{T \in \mathcal{T}_h} \mathrm{diam}(T)$. We denote*

$$\|\nabla\| = \sup_{u_h \in \mathcal{S}^1(\mathcal{T}_h)\setminus\{0\}} \frac{\|\nabla u_h\|_{L^2(\Omega)}}{\|u_h\|_{L^2(\Omega)}} \le c h_{\min}^{-1},$$

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

3 An Accelerated Linearized Primal-Dual Scheme

In this section, an accelerated variant of the linearized primal-dual scheme (1.9) is considered. We know that the energy functional in (1.8) is strongly convex with respect to u_h due to the term of its L^2 square norm. Then the linearized primal-dual scheme (1.9) can be accelerated by using the technique in [12, 13]. The accelerated linearized primal-dual scheme (1.10) for the discretized saddle-point problem (1.8) is specified in Algorithm 1.

Algorithm 1: An accelerated linearized primal-dual scheme for (1.8).

Input: Choose an initial iteration $(u_h^0; p_h^0) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$, $u_h^{-1} = u_h^0$, $\tau_0, \sigma_0, \theta_0 > 0$ and $\frac{1-3\|A\|^2\tau_0}{\tau_0} > \|\nabla\|^2\alpha^2\frac{\tau_0}{\sigma_0}$.

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

Generate the new iteration $(u_h^{n+1}; p_h^{n+1})$ satisfying

$$\text{(3.1a)} \quad \tilde{u}_h^n = u_h^n + \theta_n(u_h^n - u_h^{n-1}),$$

$$\text{(3.1b)} \quad p_h^{n+1} = \underset{p_h \in \mathcal{L}^0(\mathcal{T}_h)^d}{\arg\max} \left\{ \alpha \int_\Omega \nabla \tilde{u}_h^n \cdot p_h \,\mathrm{d}x - I_B(p_h) - \frac{\sigma_n}{2\tau_n}\|p_h - p_h^n\|_{L^2(\Omega)}^2 \right\},$$

$$\text{(3.1c)} \quad u_h^{n+1} = \underset{u_h \in \mathcal{S}^1(\mathcal{T}_h)}{\arg\min} \left\{ \begin{array}{c} \big(A^*(Au_h^n - g), u_h\big) + \frac{\beta}{2}\|u_h\|_{L^2(\Omega)}^2 \\ +\alpha \displaystyle\int_\Omega \nabla u_h \cdot p_h^{n+1} \,\mathrm{d}x + \frac{1}{2\tau_n}\|u_h - u_h^n\|_{L^2(\Omega)}^2 \end{array} \right\},$$

where τ_n, σ_n and θ_n are updated by

$$\text{(3.2)} \qquad \theta_{n+1} = \frac{1}{\sqrt{1+2\beta\tau_n}}, \quad \tau_{n+1} = \theta_{n+1}\tau_n, \quad \frac{\sigma_n}{\tau_n} = \frac{\sigma_{n+1}}{\theta_{n+1}\tau_{n+1}}.$$

end

Remark 3.1. *The solution p_h^{n+1} of the subproblem (3.1b) is explicitly given by*

$$p_h^{n+1} = \big(p_h^n + (\alpha\tau_n/\sigma_n)\nabla\tilde{u}_h^n\big) / \max\left\{1, |p_h^n + (\alpha\tau_n/\sigma_n)\nabla\tilde{u}_h^n|\right\},$$

which can be computed element-wise, see [6, 36].

Remark 3.2. *If τ_0 and σ_0 satisfy the condition $\frac{1-3\|A\|^2\tau_0}{\tau_0} > \|\nabla\|^2\alpha^2\frac{\tau_0}{\sigma_0}$, and the parameters $\tau_n, \sigma_n, \theta_n$ are chosen by (3.2), then $\frac{1-3\|A\|^2\tau_n}{\tau_n} > \|\nabla\|^2\alpha^2\frac{\tau_n}{\sigma_n}$ holds for all $n \geq 0$.*

Remark 3.3. *In [38], we applied the techniques in [12] to accelerate the alternating direction method of multipliers (ADMM) with an $O(\frac{1}{N^2})$ convergence rate. As well known in the literature, see, e.g., [12, 19, 35, 38], the linearized version of ADMM can be reduced to the primal-dual hybrid gradient method (PDHG) in [12] if the two parameters of the proximal terms of the PDHG are reciprocal constants. Here, mainly motivated by [13], we consider the linearized version of the PDHG, rather than the ADMM, whose proximal parameters are dynamically chosen in according with the particular rule (3.2), instead of being reciprocal.*

4 Convergence Analysis

In this section, we prove the convergence of Algorithm 1 and estimate its worst-case $O(\frac{1}{N^2})$ convergence rate in the ergodic sense measured by the iteration complexity. It is worthy mentioning that, although we apply the techniques proposed in [12, 13] for choosing the parameters to derive Algorithm 1, the convergence is not proved in [12, 13] for the accelerated algorithms studied therein. Here, we prove the convergence of Algorithm 1.

4.1 Convergence

We first prove that the sequence $\{(u_h^n; p_h^n)\}$ generated by Algorithm 1 converges to a saddle-point of model (1.8) with $n \to 0$, where n is the iteration counter. It is noticed that the sequence $\{(u_h^{n+1}; p_h^{n+1})\}$ generated by Algorithm 1 satisfies the following first-order optimality conditions:

$$(4.1) \qquad \big(-\frac{\sigma_n}{\tau_n}(p_h^{n+1} - p_h^n) + \alpha\nabla\tilde{u}_h^n, p_h - p_h^{n+1}\big) \le 0,\ \forall\, p_h \in \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d),$$

and

$$(4.2) \quad \big(\frac{1}{\tau_n}(u_h^{n+1}-u_h^n)+A^*(Au_h^n-g)+\beta u_h^{n+1}, u_h\big)+\alpha\big(p_h^{n+1}, \nabla u_h\big) = 0, \forall\, u_h \in \mathcal{S}^1(\mathcal{T}_h).$$

Note that (4.1) and (4.2) can be viewed as a numerical discretization of the simultaneous L^2 gradient flow of the model (1.5):

$$(4.3) \qquad -\sigma\frac{\partial p}{\partial t} + \alpha\nabla u \in \partial I_B(p)$$

and

$$(4.4) \qquad \frac{\partial u}{\partial t} - \alpha \mathrm{div} p + A^*(Au - g) + \beta u = 0,$$

respectively, with finite element discretization in space, where $\partial I_B(p)$ denotes the subdifferential of $I_B(p)$.

Theorem 4.1 (Convergence)**.** *Let $\{\boldsymbol{\mu}_h^{n+1} = (u_h^{n+1}; p_h^{n+1})\}$ be the sequence generated by Algorithm 1. Then, the sequence $\{\frac{1}{\tau_{n+1}^2}\|u_h^{n+1} - u_h^*\|^2_{L^2(\Omega)}\}$ is bounded, and the sequence $\{\boldsymbol{\mu}_h^{n+1}\}$ converges to a saddle-point of the problem (1.8) in $\mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$.*

Proof. It follows from (4.1) and (4.2) that the scheme (3.1) can be written as

$$(4.5) \qquad \begin{aligned} &\big(F(\boldsymbol{\mu}_h^{n+1}), \boldsymbol{\mu}_h - \boldsymbol{\mu}_h^{n+1}\big) + \big(H_n(\boldsymbol{\mu}_h^{n+1} - \boldsymbol{\mu}_h^n), \boldsymbol{\mu}_h - \boldsymbol{\mu}_h^{n+1}\big) \\ &\ge \theta_n\alpha\big(\nabla(u_h^n - u_h^{n-1}), p_h - p_h^{n+1}\big) - \alpha\big(\nabla(u_h^{n+1} - u_h^n), p_h - p_h^{n+1}\big) \\ &\quad + \big(A^*A(u_h^{n+1} - u_h^n), u_h - u_h^{n+1}\big),\ \forall\, \boldsymbol{\mu}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \end{aligned}$$

where $F(\cdot)$ is given in (2.3) and

$$(4.6) \qquad H_n = \begin{pmatrix} \frac{1}{\tau_n} I & 0 \\ 0 & \frac{\sigma_n}{\tau_n} I \end{pmatrix}.$$

We denote by $\mu_h^* = (u_h^*; p_h^*) \in \mathcal{S}^1(\mathcal{T}_h)\times\mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ a saddle-point of (1.8). Then, taking $\mu_h = \mu_h^*$ in (4.5), using the relationships $\tau_n = \theta_n\tau_{n-1}$ from (3.2) and $-(\mathrm{div}p_h, u_h) = (p_h, \nabla u_h)$, we have

(4.7)
$$\begin{aligned}&\big(H_n(\boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h^n), \boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^{n+1}\big) - \tau_n\alpha^2\big(\mathrm{div}(p_h^{n+1}-p_h^n), \mathrm{div}(p_h^*-p_h^{n+1})\big)\\ &+\tau_n\Big(\alpha\mathrm{div}(p_h^{n+1}-p_h^n) - \frac{u_h^{n+1}-u_h^n}{\tau_n} + \frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}, \alpha\mathrm{div}(p_h^*-p_h^{n+1}) + \frac{u_h^{n+1}-u_h^n}{\tau_n}\Big)\\ &\geq \big(F(\boldsymbol{\mu}_h^{n+1}) - F(\mu_h^*), \boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h^*\big) + \big(F(\mu_h^*), \boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h^*\big)\\ &+\tau_n\Big(\alpha\mathrm{div}(p_h^{n+1}-p_h^n) - \frac{u_h^{n+1}-u_h^n}{\tau_n} + \frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}, \frac{u_h^{n+1}-u_h^n}{\tau_n}\Big)\\ &+\big(A(u_h^{n+1}-u_h^n), A(u_h^*-u_h^{n+1})\big).\end{aligned}$$

It is easy to check that the following four identities hold:

$$\begin{aligned}&\big(H_n(\boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h^n), \boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^{n+1}\big)\\ &=\frac{1}{2}\Big(\|\boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^n\|_{H_n}^2 - \|\boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^{n+1}\|_{H_n}^2 - \|\boldsymbol{\mu}_h^n-\boldsymbol{\mu}_h^{n+1}\|_{H_n}^2\Big),\end{aligned}\tag{4.8}$$

(4.9)
$$\begin{aligned}&\big(\mathrm{div}(p_h^{n+1}-p_h^n), \mathrm{div}(p_h^*-p_h^{n+1})\big)\\ &=\frac{1}{2}\Big(\|\mathrm{div}(p_h^*-p_h^n)\|_{L^2(\Omega)}^2 - \|\mathrm{div}(p_h^*-p_h^{n+1})\|_{L^2(\Omega)}^2 - \|\mathrm{div}(p_h^n-p_h^{n+1})\|_{L^2(\Omega)}^2\Big),\end{aligned}$$

(4.10)
$$\begin{aligned}&\Big(\alpha\mathrm{div}(p_h^{n+1}-p_h^n) - \frac{u_h^{n+1}-u_h^n}{\tau_n} + \frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}, \alpha\mathrm{div}(p_h^*-p_h^{n+1}) + \frac{u_h^{n+1}-u_h^n}{\tau_n}\Big)\\ &=\frac{1}{2}\|\alpha\mathrm{div}(p_h^*-p_h^n) + \frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}\|_{L^2(\Omega)}^2 - \frac{1}{2}\|\alpha\mathrm{div}(p_h^*-p_h^{n+1}) + \frac{u_h^{n+1}-u_h^n}{\tau_n}\|_{L^2(\Omega)}^2\\ &\quad-\frac{1}{2}\|\alpha\mathrm{div}(p_h^{n+1}-p_h^n) - \frac{u_h^{n+1}-u_h^n}{\tau_n} + \frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}\|_{L^2(\Omega)}^2,\end{aligned}$$

and

(4.11)
$$\begin{aligned}&\big(A(u_h^{n+1}-u_h^n), A(u_h^*-u_h^{n+1})\big)\\ &=\frac{1}{2}\|A(u_h^*-u_h^n)\|_{L^2(\Omega)}^2 - \frac{1}{2}\|A(u_h^*-u_h^{n+1})\|_{L^2(\Omega)}^2 - \frac{1}{2}\|A(u_h^n-u_h^{n+1})\|_{L^2(\Omega)}^2.\end{aligned}$$

Since the first-order optimality condition (4.2) holds for the n-th iterate u_h^n, we have

(4.12)
$$\big(\frac{1}{\tau_{n-1}}(u_h^n-u_h^{n-1}) + A^*(Au_h^{n-1}-g) + \beta u_h^n, u_h\big) + \alpha\big(p_h^n, \nabla u_h\big) = 0, \forall\, u_h \in \mathcal{S}^1(\mathcal{T}_h).$$

Then, taking $u_h = u_h^{n+1}-u_h^n$ in (4.2) and (4.12) yields

$$\begin{aligned}&\tau_n\Big(\alpha\mathrm{div}(p_h^{n+1}-p_h^n) - \frac{u_h^{n+1}-u_h^n}{\tau_n} + \frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}, \frac{u_h^{n+1}-u_h^n}{\tau_n}\Big)\\ &\quad= \beta\|u_h^{n+1}-u_h^n\|_{L^2(\Omega)}^2 + \big(A(u_h^n-u_h^{n-1}), A(u_h^{n+1}-u_h^n)\big).\end{aligned}\tag{4.13}$$

From (3.2), it holds that $\frac{1}{2\tau_n^2}+\frac{\beta}{\tau_n}=\frac{1}{2\tau_{n+1}^2}$ and $\frac{\sigma_n}{\tau_n^2}=\frac{\sigma_{n+1}}{\tau_{n+1}^2}$. Thus, we have

$$\frac{1}{2\tau_n}\|\boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^{n+1}\|_{H_n}^2+\frac{\beta}{\tau_n}\|u_h^*-u_h^{n+1}\|_{L^2(\Omega)}^2=\frac{1}{2\tau_{n+1}}\|\boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^{n+1}\|_{H_{n+1}}^2. \tag{4.14}$$

It also follows from (2.2) that $\big(F(\mu_h^*),\boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h^*\big)\geq 0$. Therefore, substituting (4.8)-(4.11), (4.13), (4.14) and the property (2.4) into (4.7), we obtain the following inequality

$$\begin{aligned}
&\Big(\frac{1}{2\tau_n}\|\boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^n\|_{H_n}^2-\frac{1}{2}\|\alpha\mathrm{div}(p_h^*-p_h^n)\|_{L^2(\Omega)}^2-\frac{1}{2\tau_n}\|A(u_h^*-u_h^n)\|_{L^2(\Omega)}^2\\
&\quad+\frac{1}{2}\|\alpha\mathrm{div}(p_h^*-p_h^n)+\frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}\|_{L^2(\Omega)}^2\Big)\\
&-\Big(\frac{1}{2\tau_{n+1}}\|\boldsymbol{\mu}_h^n-\boldsymbol{\mu}_h^{n+1}\|_{H_{n+1}}^2-\frac{1}{2}\|\alpha\mathrm{div}(p_h^n-p_h^{n+1})\|_{L^2(\Omega)}^2\\
&\quad-\frac{1}{\tau_{n+1}}\|A(u_h^n-u_h^{n+1})\|_{L^2(\Omega)}^2-\frac{1}{2\tau_n}\|A(u_h^{n-1}-u_h^n)\|_{L^2(\Omega)}^2\\
&\quad+\frac{1}{2}\|\alpha\mathrm{div}(p_h^{n+1}-p_h^n)-\frac{u_h^{n+1}-u_h^n}{\tau_n}+\frac{u_h^n-u_h^{n-1}}{\tau_{n-1}}\|_{L^2(\Omega)}^2\Big)\\
&\geq\Big(\frac{1}{2\tau_{n+1}}\|\boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^{n+1}\|_{H_{n+1}}^2-\frac{1}{2}\|\alpha\mathrm{div}(p_h^*-p_h^{n+1})\|_{L^2(\Omega)}^2\\
&\quad-\frac{1}{2\tau_{n+1}}\|A(u_h^*-u_h^{n+1})\|_{L^2(\Omega)}^2+\frac{1}{2}\|\alpha\mathrm{div}(p_h^*-p_h^{n+1})+\frac{u_h^{n+1}-u_h^n}{\tau_n}\|_{L^2(\Omega)}^2\Big).
\end{aligned}\tag{4.15}$$

As it is assumed in Algorithm 1 that $1-3\|A\|^2\tau_0>0$ and $\tau_{n+1}\leq\tau_n$. Therefore, it holds that $1-3\|A\|^2\tau_n>0$ for any $n>0$ and we have

$$\begin{aligned}
&\sum_{n=0}^{N}\Big(\frac{1}{2\tau_{n+1}^2}\|u_h^n-u_h^{n+1}\|_{L^2(\Omega)}^2-\frac{1}{\tau_{n+1}}\|A(u_h^n-u_h^{n+1})\|_{L^2(\Omega)}^2-\frac{1}{2\tau_n}\|A(u_h^{n-1}-u_h^n)\|_{L^2(\Omega)}^2\Big)\\
&=\frac{1}{2\tau_{N+1}^2}\|u_h^N-u_h^{N+1}\|_{L^2(\Omega)}^2-\frac{1}{\tau_{N+1}}\|A(u_h^N-u_h^{N+1})\|_{L^2(\Omega)}^2\\
&\quad+\sum_{n=0}^{N-1}\Big(\frac{1}{2\tau_{n+1}^2}\|u_h^n-u_h^{n+1}\|_{L^2(\Omega)}^2-\frac{3}{2\tau_{n+1}}\|A(u_h^n-u_h^{n+1})\|_{L^2(\Omega)}^2\Big)\geq 0,
\end{aligned}\tag{4.16}$$

where $u_h^{-1}=u_h^0$ (as set in Algorithm 1). Taking the summation of (4.15) over n from 0 to N, we conclude that the following sequence

$$\Big\{\frac{1}{2\tau_{n+1}}\|\boldsymbol{\mu}_h^*-\boldsymbol{\mu}_h^{n+1}\|_{H_{n+1}}^2-\frac{1}{2}\|\alpha\mathrm{div}(p_h^*-p_h^{n+1})\|_{L^2(\Omega)}^2-\frac{1}{2\tau_{n+1}}\|A(u_h^*-u_h^{n+1})\|_{L^2(\Omega)}^2\Big\}$$

is bounded. With the definition of H_n by (4.6), and the condition

$$\frac{1-3\|A\|^2\tau_n}{\tau_n}>\|\nabla\|^2\alpha^2\frac{\tau_n}{\sigma_n}$$

as indicated in Remark 3.2, we know that both the sequences $\{\frac{1}{\tau_{n+1}^2}\|u_h^*-u_h^{n+1}\|_{L^2(\Omega)}^2\}$ and $\{\|p_h^*-p_h^{n+1}\|_{L^2(\Omega)}^2\}$ are bounded. From Remark 4.1, it holds that $\tau_n\sim O(n^{-1})$. Then, it is clear that $\lim_{n\to\infty}u_h^{n+1}=u_h^*$ and the sequence $\{p_h^{n+1}\}$ is also bounded. Thus, the sequence $\{p_h^{n+1}\}$ has a cluster point; let us denote it by q_h^*. In addition, summing (4.15) over $n = 0, \cdots, N$, we obtain that $\lim\limits_{n\to\infty} \|p_h^n - p_h^{n+1}\| = 0$ and thus $\lim\limits_{n\to\infty} p_h^{n+1} = q_h^*$. We conclude from (4.5) that $(u_h^*; q_h^*)$ satisfies (2.2), and thus $(u_h^*; q_h^*)$ is a saddle-point of (1.8). □

4.2 Convergence Rate

In this subsection, we estimate a worst-case $O(\frac{1}{N^2})$ convergence rate in the ergodic sense for Algorithm 1, where N denotes the iteration counter. Note that we follow [29, 30] and many others, to call a worst-case $O(\frac{1}{N^2})$ convergence rate by meaning that the accuracy to a solution under certain criteria is of the order $O(\frac{1}{N^2})$ after N iterations of an iterative scheme; or equivalently, it requires at most $O(\frac{1}{\sqrt{\epsilon}})$ iterations to achieve an approximate solution with an accuracy of ϵ.

A criterion to measure the accuracy of an approximation of the variational inequality (2.2) is introduced in the following theorem.

Theorem 4.2. *The solution set of variational inequality* (2.2) *is convex and can be characterized as*

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

Proof. A similar proof is referred to that of Theorem 2.3.5 in [20] or Theorem 2.1 in [25]. □

Then, it follows from the criterion in [31] and Theorem 4.2 that $\tilde{\boldsymbol{\mu}}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ is an approximate solution of the variational inequality (2.2) with an accuracy of ϵ if the following inequality holds:

$$\big(F(\boldsymbol{\mu}_h), \tilde{\boldsymbol{\mu}}_h - \boldsymbol{\mu}_h\big) \leq \epsilon, \quad \forall\, \boldsymbol{\mu}_h \in \mathcal{D}(\tilde{\boldsymbol{\mu}}_h), \tag{4.17}$$

where $\mathcal{D}(\tilde{\boldsymbol{\mu}}_h) := \{\boldsymbol{\mu}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d) : \|\boldsymbol{\mu}_h - \tilde{\boldsymbol{\mu}}_h\|_{H_0} \leq 1\}$.

To estimate the convergence rate of Algorithm 1, we first give a lemma.

Lemma 4.1. *Let $\{\boldsymbol{\mu}_h^{n+1} = (u_h^{n+1}; p_h^{n+1})\}$ be the sequence generated by Algorithm 1, we have*

$$\big(F(\boldsymbol{\mu}_h), \boldsymbol{\mu}_h^{n+1} - \boldsymbol{\mu}_h\big) + \frac{1}{\theta_{n+1}} S_{n+1}(\boldsymbol{\mu}_h) \leq S_n(\boldsymbol{\mu}_h),\ \forall\, \boldsymbol{\mu}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d), \tag{4.18}$$

where H_n is given by (4.6) *and*

$$S_n(\boldsymbol{\mu}_h) = \frac{1}{2}\|\boldsymbol{\mu}_h - \boldsymbol{\mu}_h^n\|_{H_n}^2 - \theta_n \alpha\big(\nabla(u_h^n - u_h^{n-1}), p_h - p_h^n\big) + \theta_n^2 \|\nabla\|^2 \alpha^2 \tau_n \frac{\|u_h^n - u_h^{n-1}\|_{L^2(\Omega)}^2}{2\sigma_n}.$$

Proof. First, it is easy to check that the following identity holds:

$$
\begin{aligned}
&\big(H_n(\boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h^n),\boldsymbol{\mu}_h-\boldsymbol{\mu}_h^{n+1}\big)\\
&=\frac{1}{2}\Big(\|\boldsymbol{\mu}_h-\boldsymbol{\mu}_h^n\|_{H_n}^2-\|\boldsymbol{\mu}_h-\boldsymbol{\mu}_h^{n+1}\|_{H_n}^2-\|\boldsymbol{\mu}_h^n-\boldsymbol{\mu}_h^{n+1}\|_{H_n}^2\Big).
\end{aligned}
\tag{4.19}
$$

We also have that

$$
\begin{aligned}
&\theta_n\alpha\big(\nabla(u_h^n-u_h^{n-1}),p_h-p_h^{n+1}\big)-\theta_n\alpha\big(\nabla(u_h^n-u_h^{n-1}),p_h-p_h^n\big)\\
&=\theta_n\alpha\big(\nabla(u_h^n-u_h^{n-1}),p_h^n-p_h^{n+1}\big)\\
&\geq-\theta_n^2\|\nabla\|^2\alpha^2\tau_n\frac{\|u_h^{n-1}-u_h^n\|_{L^2(\Omega)}^2}{2\sigma_n}-\sigma_n\frac{\|p_h^n-p_h^{n+1}\|_{L^2(\Omega)}^2}{2\tau_n}.
\end{aligned}
\tag{4.20}
$$

Adding $\big(F(\boldsymbol{\mu}_h),\boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h\big)$ to (4.5), together with (4.19), (4.20), the property of $F(\cdot)$ in (2.4) and the definition of H_n in (4.6), then we have

(4.21)

$$
\begin{aligned}
&\frac{1}{2}\|\boldsymbol{\mu}_h-\boldsymbol{\mu}_h^n\|_{H_n}^2-\theta_n\alpha\big(\nabla(u_h^n-u_h^{n-1}),p_h-p_h^n\big)+\theta_n^2\|\nabla\|^2\alpha^2\tau_n\frac{\|u_h^{n-1}-u_h^n\|_{L^2(\Omega)}^2}{2\sigma_n}\\
&\geq\big(F(\boldsymbol{\mu}_h),\boldsymbol{\mu}_h^{n+1}-\boldsymbol{\mu}_h\big)+\frac{1+2\beta\tau_n}{2\tau_n}\|u_h-u_h^{n+1}\|_{L^2(\Omega)}^2+\frac{\sigma_n}{2\tau_n}\|p_h-p_h^{n+1}\|_{L^2(\Omega)}^2\\
&\quad-\alpha\big(\nabla(u_h^{n+1}-u_h^n),p_h-p_h^{n+1}\big)+\frac{2-\|A\|^2\tau_n}{4\tau_n}\|u_h^n-u_h^{n+1}\|_{L^2(\Omega)}^2.
\end{aligned}
$$

Because the parameters τ_n,σ_n and θ_n satisfy (3.2) for $n\geq 0$, it holds that $\frac{1+2\beta\tau_n}{\tau_n}=\frac{1}{\theta_{n+1}\tau_{n+1}}$. Together with Remark 3.2 and (4.21), we obtain (4.18). ☐

From the result in Lemma 4.1, we obtain the following theorem which essentially implies the estimate of the convergence rate of Algorithm 1.

Theorem 4.3. *Let* $\{\boldsymbol{\mu}_h^{n+1}=(u_h^{n+1};p_h^{n+1})\}$ *be the sequence generated by Algorithm 1. For any integer* $N>0$, *let* $T_N=\sum_{n=0}^N\frac{\varrho_n}{\varrho_0}$, $\varrho_n=\frac{\tau_n}{\sigma_n}$ *and*

$$
\tilde{\boldsymbol{\mu}}_N=\frac{1}{T_N}\sum_{n=0}^N\frac{\varrho_n}{\varrho_0}\boldsymbol{\mu}_h^{n+1}.
\tag{4.22}
$$

Then, we have

$$
\big(F(\boldsymbol{\mu}_h),\tilde{\boldsymbol{\mu}}_N-\boldsymbol{\mu}_h\big)+\frac{1}{\varrho_0\sigma_{N+1}}\frac{1}{2T_N}\|u_h-u_h^{N+1}\|_{L^2(\Omega)}^2\leq\frac{1}{2T_N}\|\boldsymbol{\mu}_h-\boldsymbol{\mu}_h^0\|_{H_0}^2,
\tag{4.23}
$$

for any $\boldsymbol{\mu}_h=(u_h;p_h)\in\mathcal{S}^1(\mathcal{T}_h)\times\mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$.

Proof. Multiplying (4.18) by $\frac{\varrho_n}{\varrho_0}$, summing it from $n=0$ to N, and using the convexity of $F(\cdot)$ and (3.2), we obtain

$$
T_N\big(F(\boldsymbol{\mu}_h),\tilde{\boldsymbol{\mu}}_N-\boldsymbol{\mu}_h\big)+\frac{1}{\varrho_0}\frac{\tau_{N+1}}{\sigma_{N+1}}S_{N+1}(\boldsymbol{\mu}_h)\leq S_0(\boldsymbol{\mu}_h).
\tag{4.24}
$$

Thus it holds that

$$S_{N+1}(\boldsymbol{\mu}_h) \geq \frac{1}{2\tau_{N+1}} \|u_h - u_h^{N+1}\|_{L^2(\Omega)}^2. \tag{4.25}$$

Recall that $u_h^{-1} = u_h^0$ is assumed in Algorithm 1. The result (4.23) is followed from (4.24) and (4.25). □

Remark 4.1. *Since the parameters $\tau_n, \sigma_n, \theta_n$ in Algorithm 1 are chosen by (3.2), as proved in [12], it holds that $\tau_n \sim O(\frac{1}{\beta n})$. We also have $\frac{\sigma_{n+1}}{\tau_{n+1}^2} = \frac{\sigma_n}{\tau_n^2} = \frac{\sigma_0}{\tau_0^2}$ because of (3.2) and thus $\sigma_n \sim O(\frac{1}{\beta^2 n^2})$. In addition, with (3.2) and the definitions of ϱ_n and T_N in Theorem 4.3, it holds that $\tau_{n+1}\varrho_{n+1} = \tau_n \varrho_n = \tau_0 \varrho_0$. Then, we have $\varrho_n \sim O(\beta n)$ and $T_N \sim O(\beta N^2)$.*

In the following theorem, we show that an iterate $\tilde{\boldsymbol{\mu}}_N$ can be found such that (4.17) is satisfied with $\epsilon = O(\frac{1}{N^2})$ after N iterations of Algorithm 1. Therefore, a worst-case $O(\frac{1}{N^2})$ convergence rate is established for Algorithm 1.

Theorem 4.4 (Convergence rate in the ergodic sense)**.** *Let $\{\boldsymbol{\mu}_h^{n+1} = (u_h^{n+1}; p_h^{n+1})\}$ be the sequence generated by Algorithm 1. For any integer $N > 0$, let $T_N = \sum_{n=0}^{N} \frac{\varrho_n}{\varrho_0}$, $\varrho_n = \frac{\tau_n}{\sigma_n}$ and $\tilde{\boldsymbol{\mu}}_N$ be given by (4.22). Then, we have*

$$\big(F(\boldsymbol{\mu}_h), \tilde{\boldsymbol{\mu}}_N - \boldsymbol{\mu}_h\big) \leq \frac{1}{2T_N} \|\boldsymbol{\mu}_h - \boldsymbol{\mu}_h^0\|_{H_0}^2 \leq \frac{c}{\beta N^2}, \ \forall\, \boldsymbol{\mu}_h \in \mathcal{D}(\tilde{\boldsymbol{\mu}}_N). \tag{4.26}$$

Proof. The proof follows directly from (4.23) and the fact $T_N \sim O(\beta N^2)$ in Remark 4.1. □

5 Error Analysis

In this section, we conduct error analysis for the model (1.4) and Algorithm 1. More specifically, we estimate the regularization and perturbation errors for the model (1.4) by mainly using the methodology in [14]; and then analyze the finite element discretization error $\|u_h^* - u^*\|_{L^2(\Omega)}^2$ and the iteration error $\|u_h^{N+1} - u_h^*\|_{L^2(\Omega)}^2$ for the sequence generated by Algorithm 1.

5.1 Regularization and Perturbation Errors of (1.4)

We first analyze the regularization and perturbation errors of the model (1.4). Our main purpose is generalizing the result in [14, Theorem 4.5] for the special case $\alpha = \beta$ to the general case of (1.4) where $\alpha \neq \beta$. We first recall some notations used in [14].

Let u_0^* be the "minimal" solution of the un-regularized least-squares problem:

$$\min_u \frac{1}{2}\|Au - g_0\|_{L^2(\Omega)}^2, \quad u \in \mathcal{X} := L^2(\Omega) \cap BV(\Omega), \tag{5.1}$$

in which g_0 represents the unperturbed data of g, and $\|g - g_0\|_{L^2(\Omega)} \leq \delta$ with $\delta > 0$ the perturbation parameter. For the case $\alpha \neq \beta$, the "minimal" solution to (5.1) is of the following sense:

$$\min_u \Big\{\frac{1}{2}\|u\|_{L^2(\Omega)}^2 + \frac{\alpha}{\beta}\|Du\|\Big\}, \quad u \in S := \big\{u \in \mathcal{X} : Au = \hat{g}_0\big\}, \tag{5.2}$$

where S is the solution set of (5.1), and $\hat{g}_0$ is the projection of g_0 on the range $A(\mathcal{X})$. Let $\mathcal{M}^d(\Omega)$ denote the vectorial space of real Borel measures endowed with the norm

$$\|\mu\| = \sup\Big\{ \int_\Omega \varphi \, \mathrm{d}\mu : \varphi \in C_c(\Omega;\mathbb{R}^d), \|\varphi\|_\infty \leq 1 \Big\}.$$

Then, by [14, Theorem 2.2], there exists Λ^* in the dual space of $\mathcal{M}^d(\Omega)$ such that

$$\begin{cases} \beta(u_0^*, u - u_0^*) - \alpha(\operatorname{div} \Lambda^*, u - u_0^*)_{BV^*,BV} \geq 0, & \forall\, u \in S, \\ (\Lambda^*, \mu - \nabla u_0^*)_{\mathcal{M}^{d,*},\mathcal{M}^d} + \|Du_0^*\| \leq \|\mu\|, & \forall\, \mu \in \mathcal{M}^d(\Omega), \end{cases} \tag{5.3}$$

where "$-$div" stands for the conjugate of $\nabla : BV(\Omega) \to \mathcal{M}^d(\Omega)$. That is

$$(-\operatorname{div} \Lambda^*, u)_{BV^*,BV} = (\Lambda^*, \nabla u)_{\mathcal{M}^{d,*},\mathcal{M}^d}, \ \forall\, u \in BV(\Omega).$$

Moreover, as in [14], for a convex set C in $\mathcal{X}$, the tangent cone to C at $u \in C$ is defined as

$$T(C,u) = \big\{ v \in \mathcal{X} : \exists\, v_n \in C \text{ and } \lambda_n > 0 \text{ with } \lim_{n\to\infty} \lambda_n(v_n - u) = v \big\},$$

in which the limit is taken in $\mathcal{X}$, and the negative polar cone $T(C,u)^- \subset \mathcal{X}^*$ (the dual space of $\mathcal{X}$) at $u \in C$ is given by

$$T(C,u)^- = \big\{ v \in \mathcal{X}^* : (v, \bar{u})_{\mathcal{X}^*,\mathcal{X}} \leq 0 \text{ for all } \bar{u} \in T(C,u) \big\}.$$

Now, we analyze the regularization and perturbation errors for the generic case of the model (1.4) without the restriction $\alpha = \beta$. The proof is motivated by [14]. Recall that $\delta > 0$ is the perturbation parameter of the perturbed data g.

Theorem 5.1. *Let u^* be the solution point of (1.4), u_0^* be the "minimal" solution point of (5.1) with $\alpha, \beta > 0$, and $Rg(A^*)$ be the range of A^*. If $-u_0^* \in -\frac{\alpha}{\beta}\operatorname{div} \Lambda^* + T(\mathcal{X}, u_0^*)^- + Rg(A^*)$ holds, then we have*

$$\|u^* - u_0^*\|_{L^2(\Omega)}^2 \leq c(\delta + \beta)^2/\beta, \tag{5.4}$$

in which $c = \max\{1, \|w\|_{L^2(\Omega)}\}$ for some $w \in L^2(\Omega)$.

Proof. As u^* is the solution point of the model (1.4), we have

$$\begin{aligned} &\frac{1}{2}\|Au^* - g\|_{L^2(\Omega)}^2 + \frac{\beta}{2}\|u^*\|_{L^2(\Omega)}^2 + \alpha\|Du^*\| \\ &\leq \frac{1}{2}\|Au_0^* - g\|_{L^2(\Omega)}^2 + \frac{\beta}{2}\|u_0^*\|_{L^2(\Omega)}^2 + \alpha\|Du_0^*\|. \end{aligned} \tag{5.5}$$

Adding $\frac{1}{2}\|Au^* - \hat{g}_0\|_{L^2(\Omega)}^2 + \frac{\beta}{2}\|u^* - u_0^*\|_{L^2(\Omega)}^2$ to both sides of (5.5) gives us

$$
\begin{aligned}
&\frac{1}{2}\|Au^* - \hat{g}_0\|_{L^2(\Omega)}^2 + \frac{\beta}{2}\|u^* - u_0^*\|_{L^2(\Omega)}^2 \\
&\leq \frac{1}{2}\|Au^* - \hat{g}_0\|_{L^2(\Omega)}^2 + \frac{1}{2}\|Au_0^* - g\|_{L^2(\Omega)}^2 - \frac{1}{2}\|Au^* - g\|_{L^2(\Omega)}^2 \\
&\quad + \frac{\beta}{2}\|u^* - u_0^*\|_{L^2(\Omega)}^2 + \frac{\beta}{2}\|u_0^*\|_{L^2(\Omega)}^2 - \frac{\beta}{2}\|u^*\|_{L^2(\Omega)}^2 + \alpha(\|Du_0^*\| - \|Du^*\|) \\
&= (Au^* - \hat{g}_0, g - \hat{g}_0) + \beta(u_0^*, u_0^* - u^*) + \alpha(\|Du_0^*\| - \|Du^*\|) \\
&= (Au^* - \hat{g}_0, g - g_0) + (Au^* - \hat{g}_0, g_0 - \hat{g}_0) \\
&\quad + \beta(u_0^*, u_0^* - u^*) + \alpha(\|Du_0^*\| - \|Du^*\|) \\
&= (Au^* - \hat{g}_0, g - g_0) + \beta(u_0^*, u_0^* - u^*) + \alpha(\|Du_0^*\| - \|Du^*\|),
\end{aligned} \tag{5.6}
$$

where $(Au^* - \hat{g}_0, g_0 - \hat{g}_0) = 0$ as $\hat{g}_0$ is the projection of g_0 on range $A(\mathcal{X})$ and $\hat{g}_0 = Au_0^*$. By the condition $-u_0^* \in -\frac{\alpha}{\beta}\text{div}\,\Lambda^* + T(\mathcal{X}, u_0^*)^- + Rg(A^*)$, there exists $w \in L^2(\Omega)$ and $\eta \in T(\mathcal{X}, u_0^*)^-$ such that

$$
u_0^* = \frac{\alpha}{\beta}\text{div}\,\Lambda^* - \eta + A^* w. \tag{5.7}
$$

Then, it yields from (5.6) and (5.7) that

$$
\begin{aligned}
&\frac{1}{2}\|Au^* - \hat{g}_0\|_{L^2(\Omega)}^2 + \frac{\beta}{2}\|u^* - u_0^*\|_{L^2(\Omega)}^2 \\
&\leq (Au^* - \hat{g}_0, g - g_0) + \beta(A^*w, u_0^* - u^*)_{\mathcal{X}^*,\mathcal{X}} + \beta(\eta, u^* - u_0^*)_{\mathcal{X}^*,\mathcal{X}} \\
&\quad + \alpha(\Lambda^*, \nabla u^* - \nabla u_0^*)_{\mathcal{M}^{d,*},\mathcal{M}^d} + \alpha(\|Du_0^*\| - \|Du^*\|) \\
&\leq (\delta + \beta\|w\|_{L^2(\Omega)})\|Au^* - \hat{g}_0\|_{L^2(\Omega)},
\end{aligned} \tag{5.8}
$$

where $(\eta, u^* - u_0^*) \leq 0$ as $(u^* - u_0^*) \in T(\mathcal{X}, u_0^*)$, and $(\Lambda^*, \nabla u^* - \nabla u_0^*)_{\mathcal{M}^{d,*},\mathcal{M}^d} + \|Du_0^*\| - \|Du^*\| \leq 0$ by (5.3). Thus, the result (5.4) is followed by (5.8). □

5.2 Discretization and Iteration Errors of Algorithm 1

Next, we estimate the finite element discretization and iteration errors for Algorithm 1.

Recall the definition of the Lipschitz space $\text{Lip}(\gamma, L^2(\Omega))$ with $0 < \gamma \leq 1$ in [15]. It consists of all functions $v \in L^2(\Omega)$ such that

$$
|v|_{\text{Lip}(\gamma, L^2(\Omega))} := \sup_{t>0}\{t^{-\gamma}\omega(v,t)\} < \infty,
$$

where $\omega(v,t) = \sup_{|y|\leq t}\big(\int_\Omega |v(x+y) - v(x)|^2 \mathrm{d}x\big)^{1/2}$ is called the first order modulus of smoothness of $v \in L^2(\Omega)$. In the following theorem, the error of the finite element solution of the model (1.4) in the two-dimensional case is estimated. Similar techniques can be found in the proof of Theorem 3.2 in [36].

Theorem 5.2. *Assume $\Omega \subset \mathbb{R}^2$ and $\alpha, \beta > 0$. Let $u^* \in L^2(\Omega) \cap BV(\Omega) \cap \text{Lip}(\gamma, L^2(\Omega))$ for some $0 < \gamma \leq 1$ and $u_h^* \in \mathcal{S}^1(\mathcal{T}_h)$ be the minimizers of (1.4) in $L^2(\Omega) \cap BV(\Omega)$ and $\mathcal{S}^1(\mathcal{T}_h)$, respectively. Then we have*

$$
\frac{\beta}{2}\|u_h^* - u^*\|_{L^2(\Omega)}^2 \leq E(u_h^*) - E(u^*) \leq c(\alpha + \beta + \|A\|)h^{\gamma/(\gamma+1)}. \tag{5.9}
$$

Proof. First, it follows from [42] that

$$\|u_\varepsilon^* - u^*\|_{L^2(\Omega)} \le c|u^*|_{\mathrm{Lip}(\gamma, L^2(\Omega))}\varepsilon^\gamma,$$
$$\|\mathcal{I}_h u_\varepsilon^* - u_\varepsilon^*\|_{L^2(\Omega)} \le c|u^*|_{\mathrm{Lip}(\gamma, L^2(\Omega))}h^\gamma,$$

where u_ε^* represents the mollification of u^* and $\mathcal{I}_h u_\varepsilon^* \in \mathcal{S}^1(\mathcal{T}_h)$ is the nodal interpolation of u_ε^*. Then it yields the following estimate

$$\begin{aligned}\|\mathcal{I}_h u_\varepsilon^* - u^*\|_{L^2(\Omega)} &\le \|\mathcal{I}_h u_\varepsilon^* - u_\varepsilon^*\|_{L^2(\Omega)} + \|u_\varepsilon^* - u^*\|_{L^2(\Omega)} \\ &\le c(h^\gamma + \varepsilon^\gamma).\end{aligned} \tag{5.10}$$

Further, because of [7, Lemma 10.1], we have

$$\|\nabla \mathcal{I}_h u_\varepsilon^*\|_{L^1(\Omega)} \le (1 + ch/\varepsilon + c\varepsilon)\|Du^*\|. \tag{5.11}$$

By the bounded property of the operator A in (1.1) and the fact that $\|Du\|$ and $\|\mathcal{I}_h u_\varepsilon\|_{L^2(\Omega)}$ are bounded provided that $h \le c\varepsilon$, together with (5.10) and (5.11), we have

(5.12)

$$\begin{aligned}E(\mathcal{I}_h u_\varepsilon^*) - E(u^*) &= \alpha\|\nabla \mathcal{I}_h u_\varepsilon^*\|_{L^1(\Omega)} + \frac{\beta}{2}\|\mathcal{I}_h u_\varepsilon^*\|_{L^2(\Omega)}^2 + \frac{1}{2}\|A(\mathcal{I}_h u_\varepsilon^*) - g\|_{L^2(\Omega)}^2 \\ &\quad - \alpha\|Du^*\| - \frac{\beta}{2}\|u^*\|_{L^2(\Omega)}^2 - \frac{1}{2}\|Au^* - g\|_{L^2(\Omega)}^2 \\ &= \alpha\|\nabla \mathcal{I}_h u_\varepsilon^*\|_{L^1(\Omega)} - \alpha\|Du^*\| + \frac{\beta}{2}\big(\mathcal{I}_h u_\varepsilon^* + u^*, \mathcal{I}_h u_\varepsilon^* - u^*\big) \\ &\quad + \frac{1}{2}\big(A(\mathcal{I}_h u_\varepsilon^* + u^*) - 2g, A(\mathcal{I}_h u_\varepsilon^* - u^*)\big) \\ &\le c\alpha(h\varepsilon^{-1} + \varepsilon)\|Du^*\| + \frac{\beta}{2}\|\mathcal{I}_h u_\varepsilon^* + u^*\|_{L^2(\Omega)}\|\mathcal{I}_h u_\varepsilon^* - u^*\|_{L^2(\Omega)} \\ &\quad + \frac{\|A\|}{2}\|A(\mathcal{I}_h u_\varepsilon^* + u^*) - 2g\|_{L^2(\Omega)}\|\mathcal{I}_h u_\varepsilon^* - u^*\|_{L^2(\Omega)} \\ &\le c\big(\alpha(h\varepsilon^{-1} + \varepsilon) + (\beta + \|A\|)(h^\gamma + \varepsilon^\gamma)\big).\end{aligned}$$

Because u_h^* is the minimizer of $E(\cdot)$ in $\mathcal{S}^1(\mathcal{T}_h)$ and the energy functional $E(\cdot)$ in (1.4) is β-strongly convex, it follows from (5.12) that

$$\begin{aligned}\frac{\beta}{2}\|u_h^* - u^*\|_{L^2(\Omega)}^2 &\le E(u_h^*) - E(u^*) \le E(\mathcal{I}_h u_\varepsilon^*) - E(u^*) \\ &\le c\big(\alpha(h\varepsilon^{-1} + \varepsilon) + (\beta + \|A\|)(h^\gamma + \varepsilon^\gamma)\big).\end{aligned} \tag{5.13}$$

Therefore, the result (5.9) is derived by setting $\varepsilon = h^{1/(\gamma+1)}$ in (5.13). □

Remark 5.1. *Together with Theorems 5.1 and 5.2, we have*

$$\|u_h^* - u_0^*\|_{L^2(\Omega)}^2 \le c\Big(\frac{(\delta + \beta)^2}{\beta} + \frac{(\alpha + \beta + \|A\|)}{\beta}h^{\gamma/(\gamma+1)}\Big), \tag{5.14}$$

where u_0^* is the minimal solution point of (5.1) in $L^2(\Omega) \cap BV(\Omega)$ and u_h^* is the solution point of (1.4) in the finite element space $\mathcal{S}^1(\mathcal{T}_h)$. Hence, in (5.14), the regularization, perturbation and discretization errors are estimated all together in terms of the regularization parameters α and β, perturbation parameter δ and mesh size h.

Based on Remark 4.1, Theorem 4.1 or Theorem 4.3, we obtain that the sequence $\{u_h^{n+1}\}$ generated by Algorithm 1 converges to the minimizer of the problem (1.4) in $\mathcal{S}^1(\mathcal{T}_h)$ with a rate of $O(\frac{1}{N^2})$. The following theorem summarizes the convergence and the error estimate simultaneously.

Theorem 5.3 (Convergence and Error Analysis)**.** *Let $\boldsymbol{\mu}_h^* = (u_h^*; p_h^*)$ be a saddle-point of (1.8) with $\beta > 0$ and the sequence $\{\boldsymbol{\mu}_h^{N+1} = (u_h^{N+1}; p_h^{N+1})\}$ be generated by Algorithm 1. Then the sequence $\{u_h^{N+1}\}$ converges to the minimizer of the problem (1.4) in $\mathcal{S}^1(\mathcal{T}_h)$, and we have*

$$\|u_h^{N+1} - u_h^*\|_{L^2(\Omega)}^2 \leq \frac{\tau_{N+1}^2}{\tau_0}\|\boldsymbol{\mu}_h^* - \boldsymbol{\mu}_h^0\|_{H_0}^2 \leq \frac{c}{\beta^2 N^2}. \tag{5.15}$$

Proof. According to Theorem 4.1, the sequence $\{\frac{1}{\tau_{n+1}^2}\|u_h^* - u_h^{n+1}\|_{L^2(\Omega)}^2\}$ is bounded. Then, together with the relationship $\tau_n \sim O(\frac{1}{\beta n})$ from Remark 4.1, the result (5.15) is proved. Alternatively, the estimate (5.15) can be obtained from Theorem 4.3. Indeed, choosing $\boldsymbol{\mu}_h = \boldsymbol{\mu}_h^*$ in (4.24), together with (2.2) and (4.25), we immediately obtain

$$\|u_h^{N+1} - u_h^*\|_{L^2(\Omega)}^2 \leq \varrho_0 \sigma_{N+1}\|\boldsymbol{\mu}_h^* - \boldsymbol{\mu}_h^0\|_{H_0}^2,$$

which implies (5.15) because of $\frac{\sigma_{n+1}}{\tau_{n+1}^2} = \frac{\sigma_n}{\tau_n^2} = \frac{\sigma_0}{\tau_0^2}$ and $\tau_n \sim O(\frac{1}{\beta n})$ as delineated in Remark 4.1. □

Now, with Theorems 5.2 and 5.3, we are ready to estimate the discretization and iteration errors for the sequence generated by Algorithm 1. The result is summarized in the following theorem.

Theorem 5.4. *Assume $\Omega \subset \mathbb{R}^2$, $\alpha, \beta > 0$ and $0 < \gamma \leq 1$. Let $u^* \in L^2(\Omega) \cap BV(\Omega) \cap \mathrm{Lip}(\gamma, L^2(\Omega))$ be the minimizer of the energy functional $E(\cdot)$ in (1.4), and $u_h^{N+1} \in \mathcal{S}^1(\mathcal{T}_h)$ be generated by Algorithm 1. Then we have*

$$\|u_h^{N+1} - u^*\|_{L^2(\Omega)}^2 \leq c\Big(\frac{1}{\beta^2 N^2} + \frac{(\alpha + \beta + \|A\|)}{\beta} h^{\gamma/(\gamma+1)}\Big). \tag{5.16}$$

Proof. The assertion (5.16) is an immediate conclusion of Theorems 5.2 and 5.3, and the following inequality:

$$\|u_h^{N+1} - u^*\|_{L^2(\Omega)}^2 \leq 2(\|u_h^{N+1} - u_h^*\|_{L^2(\Omega)}^2 + \|u_h^* - u^*\|_{L^2(\Omega)}^2).$$

□

Remark 5.2. *As mentioned, the schemes (4.1)-(4.2) are the finite difference in time and finite element discretization in space of the L^2 gradient flow (4.3)-(4.4) of (1.5). Therefore, the estimate in Theorem 5.4 can also be viewed as an error estimate on u in both the time and space for the evolution systems (4.3)-(4.4). Hence, for Algorithm 1, we estimate both the finite element discretization error and the iteration error of the accelerated primal-dual method in Theorem 5.4. This is something new and it seems not achievable for the methods in [12, 36].*

5.3 Further Analysis

Our analysis above can be further consolidated to analyze the dependence of h for the coefficients of the estimates (5.15) and (5.16). The results are summarized in the following theorem. Recall the condition on the initial parameters τ_0 and σ_0 in Algorithm 1.

Theorem 5.5. *Assume $\Omega \subset \mathbb{R}^2$, $\alpha, \beta > 0$ and $0 < \gamma \leq 1$. Let $u^* \in L^2(\Omega) \cap BV(\Omega) \cap \mathrm{Lip}(\gamma, L^2(\Omega))$ and $u_h^* \in \mathcal{S}^1(\mathcal{T}_h)$ be the minimizers of (1.4) in $L^2(\Omega) \cap BV(\Omega)$ and $\mathcal{S}^1(\mathcal{T}_h)$, respectively; and $u_h^{N+1} \in \mathcal{S}^1(\mathcal{T}_h)$ be generated by Algorithm 1. For Algorithm 1, if the condition of the initial parameters τ_0 and σ_0 is slightly changed to $r \cdot \frac{1-3\|A\|^2\tau_0}{\tau_0} = \|\nabla\|^2\alpha^2\frac{\tau_0}{\sigma_0}$ for some $r \in (0,1)$, then it holds that*

$$\|u_h^{N+1}-u_h^*\|_{L^2(\Omega)}^2 \leq \Big(\frac{1}{\tau_0^2}\|u_h^*-u_h^0\|_{L^2(\Omega)}^2+\frac{\sigma_0}{\tau_0^2}\|p_h^*-p_h^0\|_{L^2(\Omega)}^2\Big)\tau_{N+1}^2 \leq c\Big(1+\frac{\alpha^2}{h^2}\Big)\frac{1}{\beta^2N^2}.$$

Further, with the estimate in Theorem 5.2, we have

$$\|u_h^{N+1}-u^*\|_{L^2(\Omega)}^2 \leq c\Big(\big(1+\frac{\alpha^2}{h^2}\big)\frac{1}{\beta^2N^2}+\frac{(\alpha+\beta+\|A\|)}{\beta}h^{\gamma/(\gamma+1)}\Big).$$

Proof. The first assertion follows from Remark 2.1 and (5.15) immediately. Moreover, with the estimate (5.9) in Theorem 5.2, we have

$$\begin{aligned}\|u_h^{N+1}-u^*\|_{L^2(\Omega)}^2 &\leq 2(\|u_h^{N+1}-u_h^*\|_{L^2(\Omega)}^2+\|u_h^*-u^*\|_{L^2(\Omega)}^2)\\ &\leq c\Big(\big(1+\frac{\alpha^2}{h^2}\big)\frac{1}{\beta^2N^2}+\frac{(\alpha+\beta+\|A\|)}{\beta}h^{\gamma/(\gamma+1)}\Big),\end{aligned}$$

which implies the second assertion. □

Finally, with Theorems 5.1 and 5.5, we easily have the following all-in-one estimate in terms of both the regularization and perturbation errors of the model (1.4), and the discretization and iteration errors of Algorithm 1.

Theorem 5.6. *Assume $\Omega \subset \mathbb{R}^2$, $\alpha, \beta > 0$ and $0 < \gamma \leq 1$. Let u_0^* be the minimal solution point of (5.1), and $u_h^{N+1} \in \mathcal{S}^1(\mathcal{T}_h)$ be generated by Algorithm 1. For Algorithm 1, if the condition of the initial parameters τ_0 and σ_0 is slightly changed to $r \cdot \frac{1-3\|A\|^2\tau_0}{\tau_0} = \|\nabla\|^2\alpha^2\frac{\tau_0}{\sigma_0}$ for some $r \in (0,1)$, then we have*

$$\|u_h^{N+1}-u_0^*\|_{L^2(\Omega)}^2 \leq c\Big(\big(1+\frac{\alpha^2}{h^2}\big)\frac{1}{\beta^2N^2}+\frac{(\alpha+\beta+\|A\|)}{\beta}h^{\gamma/(\gamma+1)}+\frac{(\delta+\beta)^2}{\beta}\Big). \tag{5.17}$$

Proof. It follows from Theorems 5.1 and 5.5 that

$$\begin{aligned}\|u_h^{N+1}-u_0^*\|_{L^2(\Omega)}^2 &\leq 2(\|u_h^{N+1}-u^*\|_{L^2(\Omega)}^2+\|u^*-u_0^*\|_{L^2(\Omega)}^2)\\ &\leq c\Big(\big(1+\frac{\alpha^2}{h^2}\big)\frac{1}{\beta^2N^2}+\frac{(\alpha+\beta+\|A\|)}{\beta}h^{\gamma/(\gamma+1)}+\frac{(\delta+\beta)^2}{\beta}\Big),\end{aligned}$$

which is the assertion (5.17). □

Remark 5.3. *For the special case where $\alpha = \beta$, it is easy to see that (5.17) reduces to*

$$\|u_h^{N+1} - u_0^*\|_{L^2(\Omega)}^2 \le c\Big(\big(1 + \frac{\alpha^2}{h^2}\big)\frac{1}{\alpha^2 N^2} + \frac{(2\alpha + \|A\|)}{\alpha} h^{\gamma/(\gamma+1)} + \frac{(\delta+\alpha)^2}{\alpha}\Big),$$

which can be regarded as a generalized estimate of Theorem 4.5 in [14] with additional considerations of the discretization error of the model (1.4) and the iteration error of Algorithm 1.

6 Numerical Results

In this section, we test some examples and numerically show the efficiency of Algorithm 1, by comparison with some existing methods of the same primal-dual kind. We test the Fredholm integral equations of the first kind in one and two dimensions. Such an integral equation arises from a variety of practical applications in, e.g., remote sensing, indirect measurement, identification of distributed parameters. We refer to [21, 28, 39, 41, 43] for more discussions.

All our codes were written in C++ based on the finite element library `AFEPack`[1] and the numerical experiments were run on a Linux desktop with i5-4570s Intel 2.9GHz processor and 8GB memory. The stopping criterion for implementing all the algorithms is

$$\frac{\|u_h^{n+1} - u_h^n\|_{L^2(\Omega)}}{\|u_h^{n+1}\|_{L^2(\Omega)}} \le 10^{-4}, \tag{6.1}$$

and the numerical accuracy is measured by the L^2 error defined as

$$\text{L2err} := \|u_h^N - u_h^*\|_{L^2(\Omega)}. \tag{6.2}$$

6.1 Algorithms to be Compared

To compare with Algorithm 1, we select several methods of the same primal-dual kind that have been well studied in the literature. The following three primal-dual type methods are tested.

- **Linearized Primal-Dual Method (LPDM).** The scheme (1.9) with $X = I - \tau A^*A$ was proposed in [36] for (1.2). Its application to (1.8) is equivalent to finding the iterate $\boldsymbol{\mu}_h^{n+1} = (u_h^{n+1}; p_h^{n+1})$ in $\mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ such that

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

  where $F(\cdot)$ and $\boldsymbol{\mu}_h$ are defined in (2.3), and

  $$M = \begin{pmatrix} \frac{1}{\tau}I - A^*A & \alpha\text{div} \\ -\alpha\nabla & \frac{\sigma}{\tau}I \end{pmatrix}. \tag{6.3}$$

  Note that the positive definiteness of M in (6.3) is ensured by the following condition

  $$\tau < \tau^* := \frac{\sqrt{\sigma^2\|A\|^4 + 4\sigma\alpha^2\|\nabla\|^2} - \sigma\|A\|^2}{2\alpha^2\|\nabla\|^2}. \tag{6.4}$$

[1] `http://dsec.pku.edu.cn/~rli/source_code/AFEPack.tar.gz`

- **Primal-Dual Method Without Linearization (PDCG and PDGMRES).** Instead of linearizing the term $\frac{1}{2}\|Au_h - g\|^2_{L^2(\Omega)}$ in the u_h-subproblem, one can directly apply the primal-dual method in [12] to solving (1.8): finding $\boldsymbol{\mu}_h^{n+1} = (u_h^{n+1}; p_h^{n+1})$ in $\mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d)$ such that

$$\big(F(\boldsymbol{\mu}_h^{n+1}) + G(\boldsymbol{\mu}_h^{n+1} - \boldsymbol{\mu}_h^n), \boldsymbol{\mu}_h - \boldsymbol{\mu}_h^{n+1}\big) \geq 0, \quad \forall\, \boldsymbol{\mu}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d),$$

where $F(\cdot)$ and $\boldsymbol{\mu}_h$ are defined in (2.3); and

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

The parameters τ and σ are restricted as

$$\tau < \tau^{\#} := \sqrt{\sigma}/(\alpha\|\nabla\|) \tag{6.5}$$

to ensure the positive definiteness of G and thus the convergence of the scheme. Also, for this method, the u_h-subproblem requires solving

$$\big(\frac{1}{\tau}(u_h^{n+1} - u_h^n) + A^*(Au_h^{n+1} - g) + \beta u_h^{n+1}, u_h\big) + \alpha\big(p_h^{n+1}, \nabla u_h\big) = 0, \tag{6.6}$$

for any $u_h \in \mathcal{S}^1(\mathcal{T}_h)$. We apply the standard conjugate gradient (CG) and the generalized minimum residual (GMRES) in [27] to this subproblem; and denote them by "PDCG" and "PDGMRES", respectively.

- **Primal-Dual-Dual Method (PDDM).** To avoid solving the linear algebra equation generated by the u_h-subproblem in (6.6), an alternative approach proposed in [12, Section 6.3.1] is to additionally dualize the data fidelity term in (1.8). The resulting saddle-point reformulation of (1.4) is

$$\inf_{u_h \in \mathcal{S}^1(\mathcal{T}_h)} \sup_{\substack{p_h \in \mathcal{L}^0(\mathcal{T}_h)^d \\ q_h \in \mathcal{S}^1(\mathcal{T}_h)}} \left\{ \begin{array}{c} \big(Au_h - g, q_h\big) - \frac{1}{2}\|q_h\|^2_{L^2(\Omega)} + \frac{\beta}{2}\|u_h\|^2_{L^2(\Omega)} \\ +\alpha \displaystyle\int_\Omega \nabla u_h \cdot p_h \,\mathrm{d}x - I_B(p_h) \end{array} \right\}, \tag{6.7}$$

where u_h is the primal variable and both p_h and q_h are dual variables (hence, the name "primal-dual-dual" is used). Then solving (6.7) amounts to finding $\boldsymbol{\mu}_h^{n+1} = (u_h^{n+1}; p_h^{n+1}; q_h^{n+1})$ in $\mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d) \times \mathcal{S}^1(\mathcal{T}_h)$ such that

$$\big(F(\boldsymbol{\mu}_h^{n+1}) + Q(\boldsymbol{\mu}_h^{n+1} - \boldsymbol{\mu}_h^n), \boldsymbol{\mu}_h - \boldsymbol{\mu}_h^{n+1}\big) \geq 0, \quad \forall\, \boldsymbol{\mu}_h \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{B}_1(\mathcal{L}^0(\mathcal{T}_h)^d) \times \mathcal{S}^1(\mathcal{T}_h),$$

where

$$\boldsymbol{\mu}_h = \begin{pmatrix} u_h \\ p_h \\ q_h \end{pmatrix}, \quad F(\boldsymbol{\mu}_h) = \begin{pmatrix} \beta u_h - \alpha \mathrm{div} p_h + A^* q_h \\ -\alpha \nabla u_h \\ q_h - Au_h + g \end{pmatrix}, \quad Q = \begin{pmatrix} \frac{1}{\tau} I & \alpha\mathrm{div} & -A^* \\ -\alpha\nabla & \frac{\sigma}{\tau} I & 0 \\ -A & 0 & \frac{\sigma}{\tau} I \end{pmatrix}.$$

Thus, the primal-dual scheme in [12] can be applied. Note that if the matrix form operator Q defined above is guaranteed to be symmetric and positive definite, the convergence of the scheme is ensured. This requires the condition

$$\tau < \tau^{\ddagger} := \sqrt{\frac{\sigma}{2(\alpha^2\|\nabla\|^2 + \|A\|^2)}}. \tag{6.8}$$

6.2 Numerical Comparisons

Below we list the values of parameters for the algorithms to be tested.

- Algorithm 1: parameters $\tau_n, \sigma_n, \theta_n$ are updated by (3.2);
- LPDM: $\tau = 0.95\tau^*$, where τ^* is given by (6.4);
- PDCG and PDGMRES: $\tau = 0.95\tau^\#$, where $\tau^\#$ is given by (6.5), with an accuracy of 1.0×10^{-6} for the CG and GMRES procedures in the u_h-subproblem of (6.6);
- PDDM: $\tau = 0.95\tau^\ddagger$, where $\tau^\ddagger$ is given by (6.8).

The accuracy for the internal iterations of PDCG and PDGMRES is chosen as 10^{-6}; because it is observed in our experiments that lower accuracy may lead to divergence or very slow convergence, while higher accuracy can hardly improve the performance. In the numerical implementations of the above mentioned algorithms, we take the initial iterate as $u_h^0 = g_h$, where g_h is the finite element discretization of function $g(x)$ with noise; p_h^0 and q_h^0 are zero functions.

For the first example, the linear operator A and functional g in (1.4) are given as below.

Example 6.1. *Consider the numerical solution to the following Fredholm integral equation of the first kind in one dimension:*

$$Au(x) := \frac{1}{\sqrt{2\pi}\eta} \int_0^1 \mathrm{e}^{-\frac{|x-\xi|^2}{2\eta^2}} u(\xi) \mathrm{d}\xi = g(x), \tag{6.9}$$

where $\eta = 0.05$ *and* g *is chosen as* $g := Au^*$ *with*

$$u^*(x) = \begin{cases} 1, & x \in [0.2, 0.4] \cup [0.6, 0.8], \\ 0, & \textit{otherwise}. \end{cases}$$

The problem (6.9) is ill-posed, and we utilize the minimization model (1.4) with L^2-TV regularization terms to approximate its solution. The function $g(x)$ on the right-hand-side of (6.9) is added by some random noise $\delta\|g\|_{L^2(\Omega)}\mathrm{randn}(x)$, where the values of $\mathrm{randn}(x)$ is sampled from the standard normal distribution, and $\delta > 0$ is the noise level parameter. The finite element mesh size over $[0, 1]$ is taken to be 0.01, over which the condition number of the corresponding discretized matrix of A^*A is 1.1883×10^9. With our choice of the kernel function in (6.9) and the finite element discretization mesh, we have $\|A\|^2 = 0.0098$ and $1/\|\nabla\| = 3.0 \times 10^{-3}$. For Algorithm 1, the initial parameters $\tau_0 = 2.0$ and $\sigma_0 = 0.1$ are taken, τ_n, σ_n and θ_n are updated by (3.2); for LPDM, PDCG/PDGMRES and PDDM, we fix $\sigma = 0.03$, and thus the values of τ are $0.9822544, 0.987269, 0.6002106$, respectively.

We first fix the regularization parameters in (1.4) as $\alpha = \beta = 5.0 \times 10^{-4}$, and implement Algorithm 1 with the aforementioned settings. The experimental results are reported in Table 1, in which the iteration numbers ("N"), CPU times in seconds ("CPU(s)") and

Table 1: Comparison of Algorithm 1, LPDM, PDCG, PDGMRES and PDDM for Example 6.1.

	$\delta = 20\%$			$\delta = 10\%$			$\delta = 5\%$			$\delta = 1\%$		
	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err
Algorithm 1	383	0.44	0.0577	354	0.41	0.0329	362	0.42	0.0199	350	0.41	0.0127
LPDM	610	0.72	0.0571	555	0.65	0.0335	587	0.68	0.0216	618	0.72	0.0143
PDCG	613	0.74	0.0571	556	0.67	0.0335	588	0.72	0.0216	617	0.74	0.0144
PDGMRES	612	0.75	0.0571	556	0.68	0.0335	588	0.72	0.0216	617	0.75	0.0143
PDDM	631	0.83	0.0603	670	0.88	0.0340	710	0.93	0.0247	804	1.07	0.0209

Table 2: Iteration numbers N, computing time in seconds and L^2 errors by Algorithm 1 for Example 6.1 with different parameters.

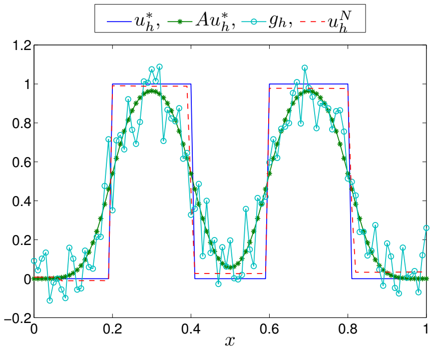
	$\delta = 20\%$			$\delta = 10\%$			$\delta = 5\%$			$\delta = 1\%$		
(α, β)	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err
$(5, 5) \times 10^{-4}$	383	0.45	0.0577	354	0.41	0.0329	362	0.42	0.0199	350	0.40	0.0127
$(5, 4) \times 10^{-4}$	360	0.43	0.0580	328	0.39	0.0324	356	0.43	0.0196	330	0.39	0.0120
$(5, 2) \times 10^{-4}$	322	0.37	0.0582	324	0.37	0.0323	336	0.39	0.0187	308	0.36	0.0114
$(5, 1) \times 10^{-4}$	312	0.36	0.0581	330	0.39	0.0325	363	0.42	0.0201	301	0.35	0.0112
$(5, 0.5) \times 10^{-4}$	312	0.36	0.0581	328	0.38	0.0325	362	0.42	0.0201	300	0.35	0.0112
$(5, 0.2) \times 10^{-4}$	311	0.36	0.0581	327	0.38	0.0325	361	0.42	0.0202	301	0.35	0.0112
$(5, 0.1) \times 10^{-4}$	311	0.36	0.0581	327	0.38	0.0325	361	0.43	0.0202	301	0.35	0.0112

the L^2 errors ("L2err") defined in (6.2) are compared for Example 6.1 with noise levels as $\delta = 20\%, 10\%, 5\%$ and 1%, respectively. According to this table, we observe that Algorithm 1 outperforms other algorithms of the same kind significantly in terms of both iteration number and CUP time, for achieving almost the same accuracy. For all results, the L^2 errors of the numerical solutions decrease with respect to noise levels. In Figure 1, we plot the graphs of u_h^*, Au_h^*, g_h and the numerical solution u_h^N generated by Algorithm 1, where u_h^* and g_h are the finite element approximation of functions u^* and g in Example 6.1, respectively. We see that the exact solution of Example 6.1 is accurately recovered; and thus the effectiveness of the model (1.4) and the efficiency of Algorithm 1 are well verified. We further demonstrate the convergence rate of Algorithm 1 and LPDM in terms of $\|u_h^N - u_h^*\|^2_{L^2(\Omega)}$ with respect to N in Figure 2.

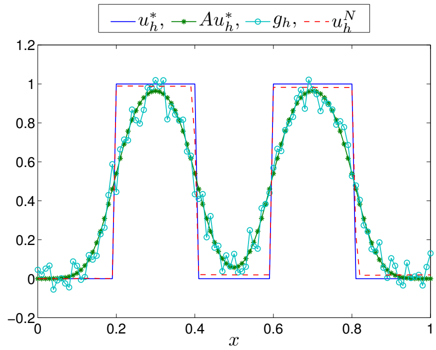
Finally, we also test the sensitivity of Algorithm 1's numerical performance with respect to the parameters α and β by Example 6.1. For succinctness, we fix α as 5.0×10^{-4} and choose varying values from 0.1×10^{-4} to 5.0×10^{-4} for β, and list the numerical results in Table 2. It is observed from Table 2 that Algorithm 1 is generally robust for various choices of α and β, and values of the same scale are especially preferred.

For the second example, the linear operator A and functional g in (1.4) are given as below.

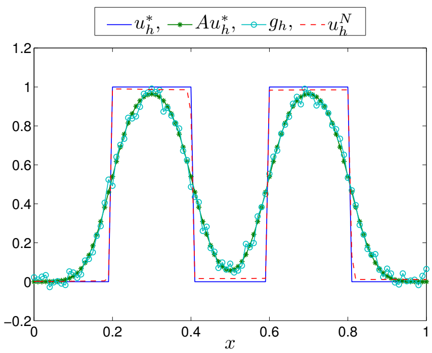
Example 6.2. *Let* $\Omega = (0, 1)^2$ *and* $\mathbb{B}(x_0) = \{x \in \Omega : |x - x_0| \leq 0.2,\ x_0 = (0.5, 0.5)\}$,



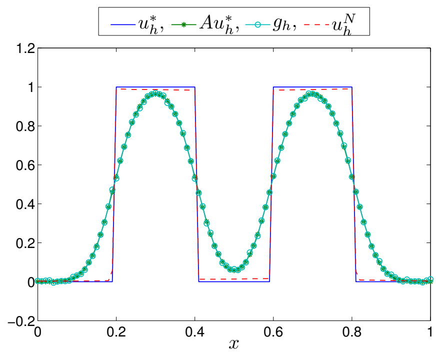
(a) noise level $\delta = 20\%$



(b) noise level $\delta = 10\%$

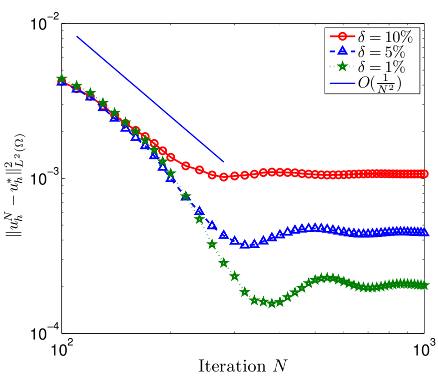


(c) noise level $\delta = 5\%$

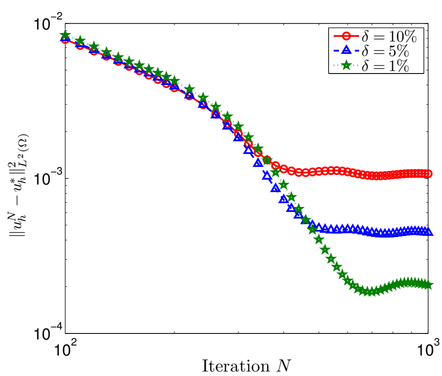


(d) noise level $\delta = 1\%$

Figure 1: Graphs of u_h^*, Au_h^*, g_h and the numerical solution u_h^N of Example 6.1 for different noise levels by Algorithm 1. (u_h^*: exact solution; g_h: data with noise)



(a) By Algorithm 1



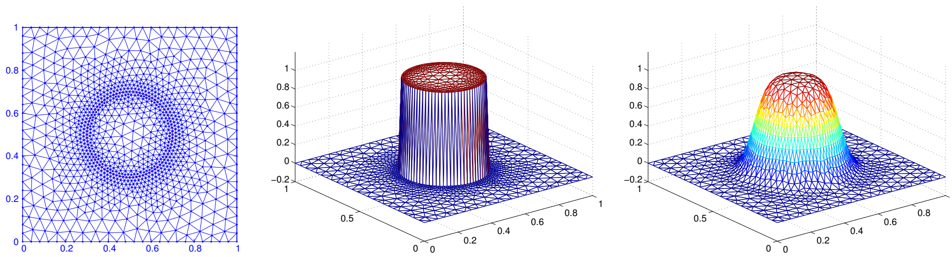
(b) By LPDM

Figure 2: Convergence rate measured by $\|u_h^N - u_h^*\|_{L^2(\Omega)}^2$ of Algorithm 1 and LPDM for Example 6.1.

$u^* = \iota_{\mathbb{B}(x_0)}$ *be the characteristic function of* $\mathbb{B}(x_0)$ *and* g *be chosen as* $g := Au^*$ *with*

$$Au(x) := \frac{1}{2\pi\eta^2}\int_\Omega \mathrm{e}^{-\frac{|x-\xi|^2}{2\eta^2}} u(\xi)\mathrm{d}\xi, \tag{6.10}$$

where $\eta = 0.05$.



(a) mesh (b) u_h^* (c) Au_h^*

Figure 3: Triangular mesh over Ω and the graphs of u_h^* and Au_h^* for Example 6.2.

The domain $\Omega = (0,1)^2$ is partitioned by `EasyMesh`[2] as the regular triangular mesh with 1046 nodes and 1990 elements shown in Figure 3(a). The graphs of u_h^* and Au_h^* are plotted in Figures 3(b) and 3(c), respectively, where u_h^* is the finite element approximation of the function u^* in Example 6.2. Accordingly, the condition number of the discretized matrix of A^*A in (6.10) is 5.4031×10^{10}. The function $g(x)$ in (6.10) is disturbed by some random noise $\delta\|g\|_{L^2(\Omega)}\mathrm{randn}(x)$, where $\mathrm{randn}(x)$ is the standard normal distribution over the finite element mesh and δ is the noise level parameter. Note that for the kernel function and finite element mesh for Example 6.2, we have $\|A\|^2 = 0.002$ and $1/\|\nabla\| = 3.5 \times 10^{-3}$. For Algorithm 1, the initial parameters $\tau_0 = 1.7$ and $\sigma_0 = 0.05$ are taken, τ_n, σ_n and θ_n are updated by (3.2); for LPDM, PDCG/PDGMRES and PDDM, we fix $\sigma = 0.025$ and thus the values of τ are 1.05029, 1.05146, 0.709538, respectively.

Table 3: Comparison of Algorithm 1, LPDM, PDCG, PDGMRES and PDDM for Example 6.2.

	$\delta = 20\%$			$\delta = 10\%$			$\delta = 5\%$			$\delta = 1\%$		
	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err
Algorithm 1	290	5.90	0.0291	266	5.49	0.0153	254	5.17	0.0080	221	4.60	0.0065
LPDM	360	7.66	0.0304	383	7.93	0.0172	385	8.02	0.0093	375	7.69	0.0057
PDCG	370	20.77	0.0307	394	21.69	0.0175	396	21.83	0.0095	382	21.15	0.0057
PDGMRES	374	21.93	0.0297	394	22.83	0.0163	385	22.25	0.0088	355	20.34	0.0058
PDDM	474	11.64	0.0313	489	12.02	0.0191	525	12.83	0.0106	557	14.20	0.0058

Again, we first fix the regularization parameters in (1.4) as $\alpha = \beta = 5.0 \times 10^{-4}$, and implement Algorithm 1 with the aforementioned settings. The experimental results are reported in Table 3. These results show again that Algorithm 1 outperforms other algorithms

[2] http://web.mit.edu/easymesh_v1.4/www/easymesh.html

Table 4: Iteration numbers N, computing time in seconds and L^2 errors by Algorithm 1 for Example 6.2 with different parameters.

	$\delta = 20\%$			$\delta = 10\%$			$\delta = 5\%$			$\delta = 1\%$		
(α, β)	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err	N	CPU(s)	L2err
$(5, 5) \times 10^{-4}$	290	6.10	0.0291	266	5.64	0.0153	254	5.24	0.0080	221	4.56	0.0065
$(5, 4) \times 10^{-4}$	291	5.93	0.0286	263	5.47	0.0146	244	5.00	0.0078	215	4.43	0.0066
$(5, 2) \times 10^{-4}$	297	6.08	0.0277	248	5.06	0.0138	224	4.67	0.0077	208	4.31	0.0067
$(5, 1) \times 10^{-4}$	287	5.87	0.0274	242	4.95	0.0134	231	4.78	0.0077	210	4.46	0.0066
$(5, 0.5) \times 10^{-4}$	287	5.85	0.0272	240	4.89	0.0132	227	4.79	0.0077	209	4.35	0.0067
$(5, 0.2) \times 10^{-4}$	286	5.88	0.0271	239	4.88	0.0131	225	4.66	0.0077	207	4.27	0.0067
$(5, 0.1) \times 10^{-4}$	285	5.93	0.0271	238	4.86	0.0131	224	4.69	0.0077	207	4.40	0.0067

for Example 6.2 as well. In Figure 4, the graphs of g_h (the finite element approximation of g) and the corresponding numerical solutions u_h^N generated by Algorithm 1 are plotted for Example 6.2 with different noise levels. We see that the exact solution of Example 6.2 can be approximated accurately. Hence, the effectiveness of the model (1.4) and the efficiency of Algorithm 1 are verified again. In Figure 5, the convergence rate measured by $\|u_h^N - u_h^*\|_{L^2(\Omega)}^2$ with respect to N is plotted for Algorithm 1 and LPDM. Last, we test the sensitivity of Algorithm 1's numerical performance with respect to the parameters α and β by Example 6.2, and report the numerical results in Table 4. Similar conclusions to Table 2 are observed.

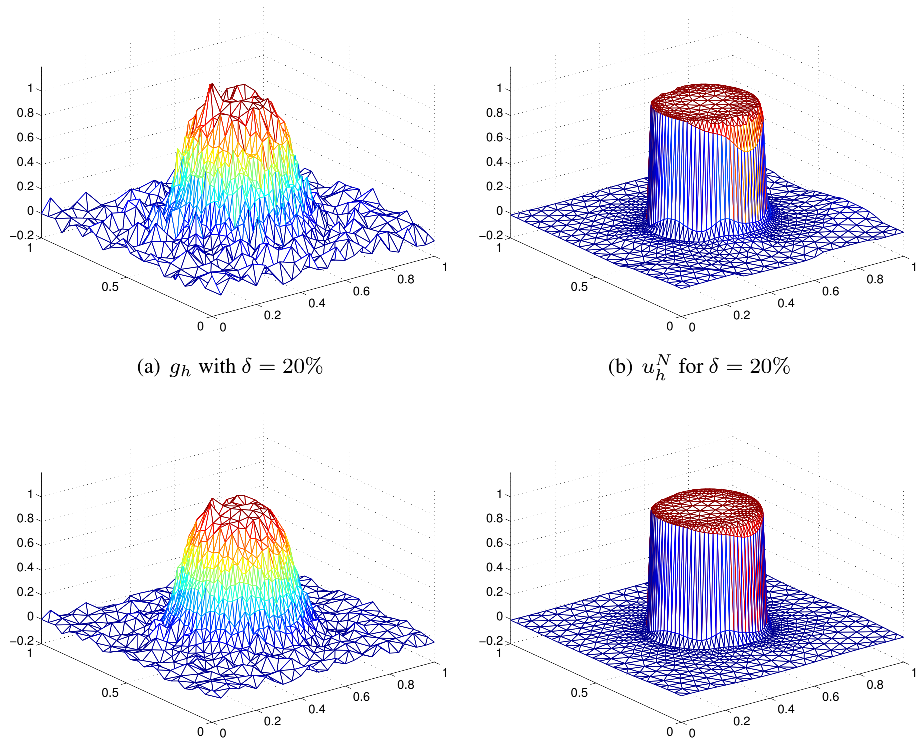
7 Conclusions

We considered a model with the L^2 and total variational regularization terms for linear inverse problems; and further reformulated the regularized model as a saddle-point problem. Applying the finite element discretization to the primal and dual variables of the reformulated saddle-point problem, we proposed a primal-dual iterative scheme to solve the resulting discretized saddle-point problem. The strong convexity of the discretized saddle-point problem enabled us to employ some rules proposed by Chambolle and Pock for dynamically adjusting the involved parameters so as to derive a worst-case $O(\frac{1}{N^2})$ convergence rate measured by iteration complexity for the proposed iterative scheme. Preliminary numerical results show that the proposed primal-dual iterative scheme is very efficient. We also estimated the regularization and perturbation errors for the regularized model, and the finite element discretization and iteration errors for the proposed primal-dual scheme.

References

[1] Acar, R., Vogel, C.R.: Analysis of bounded variation penalty methods for ill-posed problems. Inverse Problems **10**(6), 1217–1229 (1994)

[2] Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. Clarendon Press, Oxford (2000)



(a) g_h with $\delta = 20\%$ (b) u_h^N for $\delta = 20\%$

(c) g_h with $\delta = 10\%$ (d) u_h^N for $\delta = 10\%$

Figure 4: Graphs of g_h with noise levels $\delta = 20\%, 10\%$ and the corresponding numerical solution u_h^N by Algorithm 1 of Example 6.2.

[3] Arrow, K.J., Hurwicz, L., Uzawa, H.: Studies in linear and non-linear programming. Stanford University Press, Stanford (1958)

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

[5] Aubert, G., Kornprobst, P.: Mathematical problems in image processing: Partial differential equations and the calculus of variations, second edn. Springer, New York (2006)

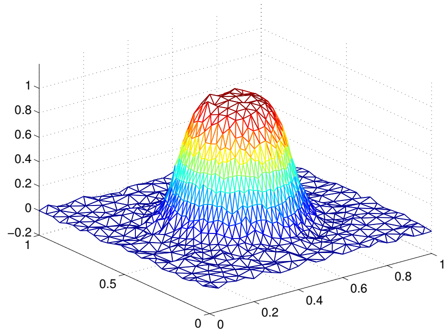
[6] Bartels, S.: Total variation minimization with finite elements: convergence and iterative solution. SIAM J. Numer. Anal. **50**(3), 1162–1180 (2012)

[7] Bartels, S.: Numerical Methods for Nonlinear Partial Differential Equations. Springer International Publishing, AG Switzerland (2015)

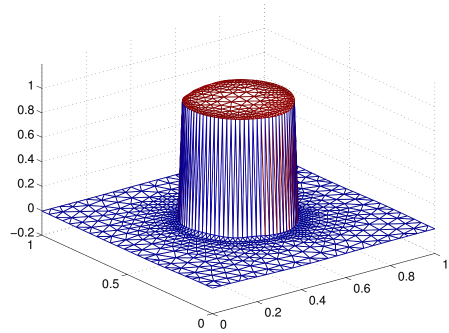
[8] Bredies, K., Sun, H.: Preconditioned Douglas-Rachford splitting methods for convex-concave saddle-point problems. SIAM J. Numer. Anal. **53**(1), 421–444 (2015)

[9] Brenner, S.C., Scott, L.R.: The mathematical theory of finite element methods, third edn. Springer, New York (2008)

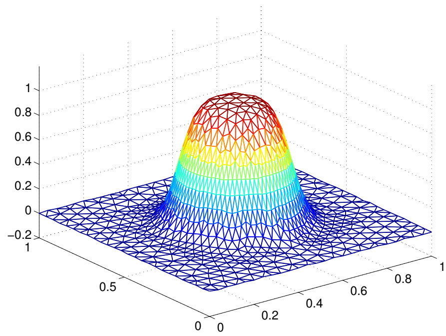
[10] Casas, E., Kunisch, K., Pola, C.: Regularization by functions of bounded variation and applications to image enhancement. Appl. Math. Optim. **40**(2), 229–257 (1999)



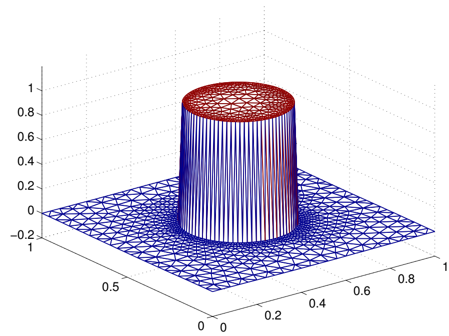
(a) g_h with $\delta = 5\%$



(b) u_h^N for $\delta = 5\%$

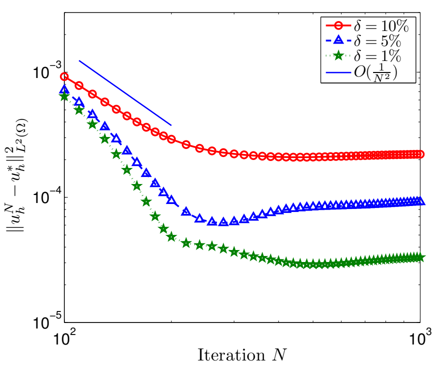


(c) g_h with $\delta = 1\%$

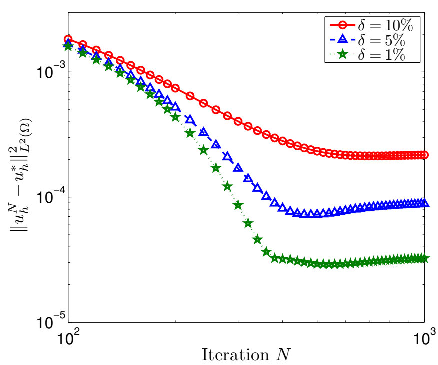


(d) u_h^N for $\delta = 1\%$

Figure 4: (con't) Graphs of g_h with noise levels $\delta = 5\%, 1\%$ and the corresponding numerical solution u_h^N by Algorithm 1 of Example 6.2.



(a) By Algorithm 1



(b) By LPDM

Figure 5: The convergence rate measured by $\|u_h^N - u_h^*\|^2_{L^2(\Omega)}$ of Algorithm 1 and LPDM for Example 6.2 with different noise levels.

[11] Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**(2), 167–188 (1997)

[12] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vision **40**(1), 120–145 (2011)

[13] Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. Math. Program. **159**(1), 253–287 (2016)

[14] Chavent, G., Kunisch, K.: Regularization of linear least squares problems by total bounded variation. ESAIM Control Optim. Calc. Var. **2**, 359–376 (1997)

[15] DeVore, R.A., Lorentz, G.G.: Constructive approximation. Springer-Verlag, Berlin (1993)

[16] Dobson, D.C., Santosa, F.: Recovery of blocky images from noisy and blurred data. SIAM J. Appl. Math. **56**(4), 1181–1198 (1996)

[17] Elman, H.C., Golub, G.H.: Inexact and preconditioned Uzawa algorithms for saddle point problems. SIAM J. Numer. Anal. **31**(6), 1645–1661 (1994)

[18] Engl, H.W., Hanke, M., Neubauer, A.: Regularization of inverse problems. Kluwer Academic Publishers Group, Dordrecht (1996)

[19] Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**(4), 1015–1046 (2010)

[20] Facchinei, F., Pang, J.S.: Finite-dimensional variational inequalities and complementarity problems. Vol. I. Springer-Verlag, New York (2003)

[21] Groetsch, C.W.: The theory of Tikhonov regularization for Fredholm equations of the first kind. Pitman (Advanced Publishing Program), Boston, MA (1984)

[22] Gutman, S.: Identification of discontinuous parameters in flow equations. SIAM J. Control Optim. **28**(5), 1049–1060 (1990)

[23] Gzyl, H., Velásquez, Y.: Linear Inverse Problems: The Maximum Entropy Connection. World Scientific Publishing Co. Pte. Ltd., Singapore (2011)

[24] He, B.S., Yuan, X.M.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. **5**(1), 119–149 (2012)

[25] He, B.S., Yuan, X.M.: On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J. Numer. Anal. **50**(2), 700–709 (2012)

[26] Hintermüller, M., Kunisch, K.: Total bounded variation regularization as a bilaterally constrained optimization problem. SIAM J. Appl. Math. **64**(4), 1311–1333 (2004)

[27] Kelley, C.T.: Iterative Methods for Linear and Nonlinear Equations. SIAM, Philadelphia (1995)

[28] Kress, R.: Linear integral equations, third edn. Springer, New York (2014)

[29] Nemirovsky, A.S., Yudin, D.B.: Problem complexity and method efficiency in optimization. John Wiley & Sons, Inc., New York (1983)

[30] Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSR **269**(3), 543–547 (1983), (In Russian. Translated in Soviet Math. Dokl., 27 (1983), pp. 372-376.)

[31] Nesterov, Y.: Gradient methods for minimizing composite functions. Math. Program. **140**(1), 125-161 (2013)

[32] Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1762–1769 (2011)

[33] Ring, W.: Structural properties of solutions to total variation regularization problems. ESAIM Math. Model. Numer. Anal. **34**(4), 799–810 (2000)

[34] Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**, 259–268 (1992)

[35] Shefi, R.: Rate of convergence analysis for convex nonsmooth optimization algorithms. Doctoral dissertation, Tel Aviv University, Israel, 2015.

[36] Tian, W.Y., Yuan, X.M.: Linearized primal-dual methods for linear inverse problems with total variation regularization and finite element discretization. Inverse Problems **32**(11), 115011 (2016)

[37] Tian, W.Y., Yuan, X.M.: Convergence analysis of primal-dual based methods for total variation minimization with finite element approximation. J. Sci. Comput. **76**(1), 243–274 (2018)

[38] Tian, W.Y., Yuan, X.M.: An alternating direction method of multipliers with a worst-case $O(1/n^2)$ convergence rate. Math. Comp. (2018) doi:10.1090/mcom/3388

[39] Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-Posed Problems. Wiley, New York (1977), (Translated from the Russian)

[40] Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., Yagola, A.G.: Numerical Methods for the Solution of Ill-Posed Problems. Kluwer, Dordrecht (1995), (Translated from the 1990 Russian original by R. A. M. Hoksbergen and revised by the authors)

[41] Vogel, C.R.: Computational methods for inverse problems. SIAM, Philadelphia, PA (2002)

[42] Wang, J., Lucier, B.J.: Error bounds for finite-difference methods for Rudin-Osher-Fatemi image smoothing. SIAM J. Numer. Anal. **49**(2), 845–868 (2011)

[43] Wing, G.M.: A primer on integral equations of the first kind: The problem of deconvolution and unfolding. SIAM, Philadelphia, PA (1991)

[44] Zhu, M., Chan, T.F.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, UCLA, Los Angeles, CA (2008)

[45] Ziemer, W.P.: Weakly differentiable functions: Sobolev spaces and functions of bounded variation. Springer-Verlag, New York (1989)