

# STOCHASTIC PROXIMAL LINEARIZED ADMM FOR SPARSE DISTRIBUTION CONTROL PROBLEM CONSTRAINED BY RANDOM ELLIPTIC EQUATION\*

Haiming Song and Hao Wang

*School of Mathematics, Jilin University, Changchun 130012, China*  
*Emails: songhaiming@jlu.edu.cn, haowang21@mails.jlu.edu.cn*

Jiageng Wu

*School of Advanced Manufacturing and Robotics, Peking University, Beijing 100871, China*  
*Email: jgwu@pku.edu.cn*

Jinda Yang<sup>1)</sup>

*School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China*  
*Email: yangjinda@lzu.edu.cn*

## Abstract

In this paper, we consider the sparse distributed control problem constrained by a random elliptic equation, which we reformulate as a nonsmooth stochastic optimization problem in Hilbert space. By incorporating the advantages of the stochastic approximation approach and the alternating direction method of multipliers (ADMM), we propose a stochastic ADMM algorithm. This method decouples the stochasticity arising from the random equation constraint from the nonsmoothness of the control objective, allowing them to be tackled separately within the iterations. We introduce stochastic gradients and develop a proximal linearization technique for the stochastic subproblem, allowing each subproblem to admit a closed-form solution. The convergence and a high-probability bound of the proposed method are analyzed for the model problem. Numerical results demonstrate the effectiveness and efficiency of our method.

*Mathematics subject classification:* 90C15, 62L20, 35R60, 60H35.

*Key words:* PDE-constrained optimization, Stochastic optimization, Sparse distributed control, Alternating direction method of multipliers, Proximal linearization.

## 1. Introduction

Optimal control problems (OCPs) modeled by partial differential equations with random coefficients (RPDEs) arise in numerous scientific and engineering applications, and have received significant attention in theoretical and computational aspects. We refer the reader to, e.g., [7, 9, 16, 18, 19, 28] for a few references. Moreover, there has been considerable interest in sparse distributed control for a random elliptic equation, driven by applications to tidal-stream energy farm design [4], economics-constrained tidal turbine array layout [30], and beyond. The term “distributed” herein refers to a control action defined over the spatial domain  $\Omega$ , as opposed to a probability distribution. To promote the robustness and sparsity of the control, the tracking-type objective usually includes a stochastic expectation and  $L^1$  regularization, subject to the

---

\* Received June 22, 2025 / Revised version received October 28, 2025 / Accepted December 1, 2025 /  
Published online April 2, 2026 /

<sup>1)</sup> Corresponding author

state equation with additional constraints on the control variable, reading as follows:

$$\begin{aligned} \min_{y,u} \quad & \frac{1}{2} \int_{\mathcal{E}} \int_{\Omega} |y - y_d|^2 dx d\mathcal{P}(\xi) + \frac{\alpha}{2} \int_{\Omega} |u|^2 dx + \beta \int_{\Omega} |u| dx \\ \text{s.t.} \quad & \begin{cases} -\nabla \cdot (a(x, \xi) \nabla y(x, \xi)) = u(x), & (x, \xi) \in \Omega \times \mathcal{E}, \\ y(x, \xi) = 0, & (x, \xi) \in \partial\Omega \times \mathcal{E}, \\ u \in U^{ad}. \end{cases} \end{aligned} \quad (1.1)$$

Here, we use  $\Omega \subset \mathbb{R}^n$  ( $1 \leq n \leq 3$ ) to denote a bounded, convex, polyhedral, and connected domain with a piecewise Lipschitz-continuous boundary  $\partial\Omega$ . In (1.1), the state variable  $y$  belongs to the Bochner space  $Y := L^2(\mathcal{E}; H_0^1(\Omega))$ , defined by

$$L^2(\mathcal{E}; H_0^1(\Omega)) = \left\{ y : \mathcal{E} \rightarrow H_0^1(\Omega); y \text{ is measurable and } \int_{\mathcal{E}} \|y(\xi, \cdot)\|_{H_0^1(\Omega)}^2 d\mathcal{P}(\xi) < +\infty \right\}$$

with a complete probability space  $(\mathcal{E}, \mathcal{F}, \mathcal{P})$ , where  $\mathcal{E}$  is the sample space,  $\mathcal{F}$  is a  $\sigma$ -algebra of measurable events, and  $\mathcal{P}$  is a probability measure. The measurability of  $H_0^1(\Omega)$ -valued functions is understood with respect to the Borel  $\sigma$ -algebra on  $H_0^1(\Omega)$ , as further discussed in [18]. For the control variable  $u$ , the non-empty admissible set  $U^{ad}$  is defined by

$$U^{ad} = \{u \in L^2(\Omega) : -\infty < u_a \leq u(x) \leq u_b < +\infty \text{ a.e. } x \in \Omega\}$$

with given  $u_a, u_b \in \mathbb{R}$ . The desired state  $y_d$  belongs to  $L^2(\Omega)$  and the function  $a(x, \xi)$  is the random diffusion coefficient. The constants  $\alpha > 0$  and  $\beta > 0$  are parameters for the  $L^2$  and  $L^1$  regularization terms, respectively.

Due to the complex structure of (1.1), these problems are generally challenging from the perspective of algorithmic design. Specifically, the main numerical challenges arise from two key aspects: the non-differentiability of the  $L^1$ -regularization term in a Hilbert space [7, 18, 19], and the high computational cost of evaluating the stochastic expectation. Additionally, control constraints complicate the problem further [3, 8]. The challenge, therefore, lies in tackling these interconnected issues within an efficient algorithmic framework.

For sparse distributed control problems constrained by deterministic elliptic PDEs and control constraints, the semi-smooth Newton (SSN) method is an efficient solver, which utilizes an active-set strategy and ensures superlinear convergence under proper conditions [15, 25]. However, the method's active-set strategy is complex to implement, and when the SSN method is applied to the parabolic PDE-constrained optimization problems, it results in extremely large linear systems and significant computational costs. To address this issue, the alternating direction method of multipliers has been developed and successfully implemented for deterministic PDE-constrained problems with separable structures [11, 12, 26]. Nevertheless, the current applications of ADMM remain confined mainly to deterministic problems and have rarely addressed cases where the objective functional contains stochastic expectations.

For RPDE-constrained problems, one can use the sample average approximation (SAA) method to approximate the original problem with an SAA problem built with enough samples, and then solve it using suitable deterministic algorithms [14, 18, 20]. Clearly, the performance of SAA methods depends on the efficiency of both the sampling and the chosen algorithms. By discretizing the probability space with interpolating points, sparse grid methods can be used for problems involving multiple random dimensions, but the computational cost usually increases

rapidly with dimension [1, 6, 22, 31]. Stochastic approximation (SA) methods, particularly those based on stochastic gradient (SG), can also be considered, employing a stochastic first-order oracle that provides unbiased estimates of the gradient [2, 7, 13]. With appropriate step sizes, the SG methods exhibit the advantages of broad applicability and low per-iteration cost.

In the context of sparse distributed control problems, an attractive idea is to design an algorithm that treats sparse control and distributed control separately, following the concept of SA methods to deal with the stochastic expectation associated with the random elliptic equation constraint. To achieve this, we adopt an ADMM-type framework, which handles the aforementioned tasks through different subproblems derived from a structured splitting formulation. We note that one of the subproblems, due to the presence of the stochastic expectation, lacks a closed-form solution. To address this issue, a proximal linearization of the subproblem has been proposed by incorporating the concept of SA methods. At each iteration, the linearized subproblem admits a closed-form solution with a simple stochastic first-order oracle, which is obtained by solving a sampling elliptic equation and its adjoint equation once per iteration. In addition, we introduce a relaxation scheme to enhance the performance of the algorithm. Numerical experiments demonstrate that our algorithm performs efficiently and robustly on the sparse distributed control problems constrained by random elliptic equations.

The rest of this paper is organized as follows. In Section 2, we present the reformulated model and the proposed algorithmic framework. Section 3 establishes the convergence of the proposed method and discusses its convergence rate and high-probability bounds. Numerical results validating the efficiency of our method are presented in Section 4. Finally, concluding remarks are provided in Section 5.

## 2. Stochastic Proximal Linearized ADMM for the Model Problem (1.1)

### 2.1. Reformulation of the model problem (1.1)

To simplify the presentation, we use  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  for the  $L^2$ -inner product and  $L^2$ -norm in  $L^2(\Omega)$ , respectively, and use  $\|\cdot\|_1$  for the  $L^1$ -norm in  $L^1(\Omega)$ . For a random variable  $\omega : \mathcal{E} \rightarrow \mathbb{R}$  defined on the probability space  $(\mathcal{E}, \mathcal{F}, \mathcal{P})$ , the expectation operator  $\mathbb{E}$  is defined as

$$\mathbb{E}[\omega(\xi)] = \int_{\mathcal{E}} \omega(\xi) d\mathcal{P}(\xi).$$

For the state equation in (1.1), the following lemma gives its well-posedness, which will help us to reformulate the model problem (1.1) using the control-to-state operator.

**Lemma 2.1 (cf. [20]).** *Let the random diffusion coefficient  $a(\cdot, \xi) : \mathcal{E} \rightarrow C^1(\bar{\Omega})$  be measurable. Suppose there exists constants  $a_{\min}, a_{\max}, \bar{a}_{\max} \in (0, \infty)$  such that*

$$\begin{aligned} a_{\min} &\leq a(x, \xi) \leq a_{\max}, \quad \forall x, \quad \xi \in \Omega \times \mathcal{E}, \\ \|a(\cdot, \xi)\|_{C^1(\bar{\Omega})} &\leq \bar{a}_{\max}, \quad \forall \xi \in \mathcal{E}. \end{aligned}$$

*Then, for every  $\xi \in \mathcal{E}$  and any control variable  $u \in L^2(\Omega)$ , there exists a unique solution  $y(\cdot, \xi) \in H_0^1(\Omega)$  to the variational problem*

$$\int_{\Omega} a(x, \xi) \nabla y \cdot \nabla v \, dx = \int_{\Omega} uv \, dx, \quad \forall v \in H_0^1(\Omega).$$

From Lemma 2.1, it is natural to introduce the linear control-to-state operator  $S_\xi : L^2(\Omega) \rightarrow L^2(\mathcal{E}; H_0^1(\Omega))$ , defined by  $y = S_\xi u$ , where  $y$  is the solution to the state equation for a given control  $u$ . Then, the model problem (1.1) can be written in the following reduced form:

$$\min_{u \in U^{ad}} \mathbb{E} \left[ \frac{1}{2} \|S_\xi(u) - y_d\|^2 \right] + \frac{\alpha}{2} \|u\|^2 + \beta \|u\|_1.$$

By a standard discussion as in [20], the above problem has a unique solution under the assumptions of Lemma 2.1. Additionally, we define the functionals  $f(u)$  and  $g(u)$  as

$$f(u) = \mathbb{E}[F(u, \xi)] = \mathbb{E} \left[ \frac{1}{2} \|S_\xi(u) - y_d\|^2 + \frac{\alpha}{2} \|u\|^2 \right], \quad g(u) = \beta \|u\|_1.$$

To separate the sparse component ( $L^1$ -regularization term) from the objective functional, we introduce an auxiliary variable  $z$ . Then the above problem can be written as the following separable convex optimization problem:

$$\begin{aligned} \min_{u, z} \quad & J(u, z) = f(u) + g(z), \\ \text{s.t.} \quad & u - z = 0, \\ & u \in U^{ad}, \quad z \in L^2(\Omega). \end{aligned} \tag{2.1}$$

It is evident that this reformulation allows us to treat the stochastic expectation part and the  $L^1$ -regularization part separately, a structure that is conducive to employing the SA approach in the specific subproblem. We will present the design concept of our algorithm in the next subsection.

## 2.2. Algorithmic framework of stochastic proximal linearized ADMM

Due to the separable structure of problem (2.1), we can readily obtain the following classic ADMM algorithm as in [5, 10, 27]:

$$\begin{cases} u^{k+1} = \arg \min_{u \in U^{ad}} \left\{ f(u) - \langle \lambda^k, u - z^k \rangle + \frac{\rho}{2} \|u - z^k\|^2 \right\}, \\ z^{k+1} = \arg \min_{z \in L^2(\Omega)} \left\{ g(z) - \langle \lambda^k, u^{k+1} - z \rangle + \frac{\rho}{2} \|u^{k+1} - z\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \rho(u^{k+1} - z^{k+1}), \end{cases}$$

where  $\rho > 0$  is a constant. However, the above iteration scheme can not be used directly because calculating the expected value in  $f(u)$  is numerically expensive or even infeasible. Following the methodology of SA, we can replace  $f(u)$  in the  $u$ -subproblem with its stochastic approximation  $F(u, \xi^k)$ . Due to the box constraint  $u \in U^{ad}$ , the  $u$ -subproblem lacks a closed-form solution. Thus, we linearize  $F(u, \xi^k)$  as

$$F(u, \xi^k) \rightarrow F(u^k, \xi^k) + \langle \nabla F(u^k, \xi^k), u - u^k \rangle.$$

The proximal term  $\eta^k \|u - u^k\|^2/2$  is introduced to enhance algorithmic robustness and guarantee convergence by counteracting the errors induced by the stochastic approximation and linearization. The resulting  $u$ -subproblem becomes

$$u^{k+1} = \arg \min_{u \in U^{ad}} \left\{ \langle \nabla F(u^k, \xi^k), u \rangle - \langle \lambda^k, u - z^k \rangle + \frac{\rho}{2} \|u - z^k\|^2 + \frac{\eta^k}{2} \|u - u^k\|^2 \right\},$$

where the sequence  $\{\eta^k\}$  is a positive and monotonically increasing sequence.

To enhance the numerical performance, we empirically introduce a relaxation in the  $z$ -subproblem by replacing  $u^{k+1}$  with the interpolation  $\varrho u^{k+1} + (1 - \varrho)z^k$  for a given relaxation factor  $\varrho \in (0, 2)$ .

Eventually, we arrive at the following stochastic proximal linearized ADMM scheme:

$$\begin{cases} u^{k+1} = \arg \min_{u \in U^{ad}} \left\{ \langle \nabla F(u^k, \xi^k), u \rangle - \langle \lambda^k, u - z^k \rangle + \frac{\rho}{2} \|u - z^k\|^2 + \frac{\eta^k}{2} \|u - u^k\|^2 \right\}, & (2.2a) \\ z^{k+1} = \arg \min_{z \in L^2(\Omega)} \left\{ g(z) - \langle \lambda^k, u^{k+1} - z \rangle + \frac{\rho}{2} \|\varrho u^{k+1} + (1 - \varrho)z^k - z\|^2 \right\}, & (2.2b) \\ \lambda^{k+1} = \lambda^k - \rho(\varrho u^{k+1} + (1 - \varrho)z^k - z^{k+1}). & (2.2c) \end{cases}$$

We now turn to showing that both decoupled subproblems can be easily solved with computable closed-form solutions. For the  $u$ -subproblem, we first present the calculation of the gradient  $\nabla F(u^k, \xi^k)$ .

**Lemma 2.2** (cf. [18, 19]). *Under the assumptions of Lemma 2.1,*

$$\nabla f(u) = \mathbb{E}[\nabla F(u, \xi)], \quad \nabla F(u, \xi) = \alpha u + p(u, \xi),$$

where  $p(u, \xi)$  is the solution of the following problem almost surely in  $\mathcal{E}$ :

$$\begin{aligned} -\nabla \cdot (a(x, \xi) \nabla p(\cdot, \xi)) &= y(x, \xi) - y_d, & x \in \Omega, \\ p(x, \xi) &= 0, & x \in \partial\Omega. \end{aligned} \quad (2.3)$$

From Lemma 2.2, the variable  $u^{k+1}$  can be solved by

$$\begin{aligned} u^{k+1} &= \arg \min_{u \in U^{ad}} \left\{ \langle u, \nabla F(u^k, \xi^k) - \lambda^k \rangle + \frac{\rho}{2} \|u\|^2 + \frac{\eta^k}{2} \|u\|^2 - \langle u, \rho z^k + \eta^k u^k \rangle \right\} \\ &= \arg \min_{u \in U^{ad}} \left\{ \langle u, \nabla F(u^k, \xi^k) - \lambda^k - \rho z^k - \eta^k u^k \rangle + \frac{\rho + \eta^k}{2} \|u\|^2 \right\} \\ &= \arg \min_{u \in U^{ad}} \left\{ \left\| u - \left( \frac{\lambda^k + \rho z^k + \eta^k u^k - \alpha u^k - p^k}{\rho + \eta^k} \right) \right\|^2 \right\}. \end{aligned}$$

With the definition of  $U^{ad}$ , the solution  $u^{k+1}$  can be computed pointwise as

$$u^{k+1}(x) = \min \left\{ u_b, \max \left\{ u_a, \frac{(\eta^k - \alpha)u^k(x) + \rho z^k(x) - p^k(x) + \lambda^k(x)}{\rho + \eta^k} \right\} \right\}. \quad (2.4)$$

It is easy to check that the  $z$ -subproblem (2.2b) is equivalent to

$$z^{k+1} = \arg \min_{z \in L^2(\Omega)} \left\{ \beta \|z\|_1 + \frac{\rho}{2} \left\| \varrho u^{k+1} + (1 - \varrho)z^k - z - \frac{\lambda^k}{\rho} \right\|^2 \right\}.$$

Let

$$\tilde{z}^{k+1} = \varrho u^{k+1} + (1 - \varrho)z^k - \frac{\lambda^k}{\rho},$$

then the solution  $z^{k+1}$  can be computed as follows:

$$z^{k+1}(x) = \begin{cases} \bar{z}^{k+1}(x) - \frac{\beta}{\rho}, & \text{if } \frac{\beta}{\rho} \leq \bar{z}^{k+1}(x), \\ 0, & \text{if } -\frac{\beta}{\rho} < \bar{z}^{k+1}(x) < \frac{\beta}{\rho}, \\ \bar{z}^{k+1}(x) + \frac{\beta}{\rho}, & \text{if } \bar{z}^{k+1}(x) \leq -\frac{\beta}{\rho}. \end{cases} \quad (2.5)$$

Based on the above observations, we summarize the stochastic proximal linearized ADMM in Algorithm 2.1.

**Algorithm 2.1:** Stochastic Proximal Linearized ADMM for Problem (2.1).

**Input:** Initial point  $u^0, z^0, \lambda^0$ , parameters  $\rho, \eta^k > 0$  and  $\varrho \in (0, 2)$ .

- 1 **for**  $k = 0, 1, \dots$ , **do**
- 2     Compute  $y^k$  by solving the state equation in (1.1) with sample  $\xi^k$  and control  $u^k$ .
- 3     Compute  $p^k$  by solving the adjoint equation (2.3) with sample  $\xi^k$  and state  $y^k$ .
- 4     Update  $u^{k+1}$  by (2.4).
- 5     Update  $z^{k+1}$  by (2.5).
- 6     Update  $\lambda^{k+1}$  by (2.2c).
- 7 **end**

### 3. Convergence Analysis

#### 3.1. Preliminaries

To present our analysis compactly, we denote variables  $w, \hat{w}^*, w^* \in W := U^{ad} \times L^2(\Omega) \times L^2(\Omega)$ , variables  $v, \hat{v}^* \in V := L^2(\Omega) \times L^2(\Omega)$ , and the operator  $D(w)$  as follows:

$$\begin{aligned} w &:= \begin{pmatrix} u \\ z \\ \lambda \end{pmatrix}, & \hat{w}^* &:= \begin{pmatrix} u^* \\ z^* \\ \lambda^* \end{pmatrix}, & w^* &:= \begin{pmatrix} u^* \\ z^* \\ \lambda^* \end{pmatrix}, \\ v &:= \begin{pmatrix} z \\ \lambda \end{pmatrix}, & \hat{v}^* &:= \begin{pmatrix} z^* \\ \lambda^* \end{pmatrix}, & D(w) &:= \begin{pmatrix} -\lambda \\ \lambda \\ u - z \end{pmatrix}, \end{aligned} \quad (3.1)$$

where  $u^*$  and  $z^*$  are the optimal solutions to problem (2.1), and  $\lambda^*$  is the corresponding optimal Lagrange multiplier. With the above notations, it is easy to see that problem (2.1) can be equivalently reformulated as the variational inequality

$$g(z) - g(z^*) + \langle u - u^*, \nabla f(u^*) \rangle + \langle w - w^*, D(w^*) \rangle \geq 0, \quad \forall w \in W. \quad (3.2)$$

For all  $w_1, w_2 \in W$  and  $\lambda \in L^2(\Omega)$ , it can be verified that

$$\langle D(w_1) - D(w_2), w_1 - w_2 \rangle = 0, \quad (3.3)$$

$$\langle \hat{w}^* - w_1, D(w_1) \rangle = \langle \lambda, u_1 - z_1 \rangle. \quad (3.4)$$

For the matrix operator

$$H = \begin{pmatrix} \frac{\rho}{\varrho} I & \frac{\varrho-1}{\varrho} I \\ \frac{\varrho-1}{\varrho} I & \frac{1}{\rho\varrho} I \end{pmatrix},$$

where  $I : L^2(\Omega) \rightarrow L^2(\Omega)$  represents the identity operator, it can be verified that  $H$  is positive definite when  $\varrho \in (0, 2)$ . Therefore, we define the inner product induced by  $H$  as

$$\langle v_1, H v_2 \rangle = \frac{\rho}{\varrho} \langle z_1, z_2 \rangle + \frac{\varrho-1}{\varrho} \langle z_1, \lambda_2 \rangle + \frac{\varrho-1}{\varrho} \langle z_2, \lambda_1 \rangle + \frac{1}{\rho\varrho} \langle \lambda_1, \lambda_2 \rangle, \quad \forall v_1, v_2 \in V,$$

and the corresponding induced norm is given by

$$\|v\|_H = \sqrt{\langle v, H v \rangle}.$$

In the subsequent discussion, the symbol  $I$  denotes different identity operators depending on the function space.

Furthermore, we introduce two auxiliary sequences  $\{\hat{w}^k\}$  and  $\{\hat{v}^k\}$ ,

$$\hat{w}^k = \begin{pmatrix} \hat{u}^k \\ \hat{z}^k \\ \hat{\lambda}^k \end{pmatrix} = \begin{pmatrix} u^{k+1} \\ z^{k+1} \\ \lambda^k - \rho(u^{k+1} - z^k) \end{pmatrix}, \quad \hat{v}^k = \begin{pmatrix} \hat{z}^k \\ \hat{\lambda}^k \end{pmatrix}. \quad (3.5)$$

From (2.2c) and  $\hat{\lambda}^k = \lambda^k - \rho(u^{k+1} - z^k)$ , we have

$$\lambda^{k+1} - \lambda^k = \varrho(\hat{\lambda}^k - \lambda^k) + \rho(z^{k+1} - z^k), \quad (3.6)$$

and

$$v^{k+1} - v^k = M(\hat{v}^k - v^k),$$

where

$$M = \begin{pmatrix} I & 0 \\ \rho I & \varrho I \end{pmatrix}.$$

**Remark 3.1.** Algorithm 2.1 can be interpreted as a stochastic version of linearized ADMM within a predictor-corrector framework, where the  $\lambda$  subproblem is first computed as  $\hat{\lambda}^k = \lambda^k - \rho(u^{k+1} - z^k)$ , and the correction is then performed by using the matrix  $M$ . Following the scheme of Algorithm 2.1, the correction step is incorporated directly into the  $\lambda$  subproblem, thereby reducing the computational cost of each iteration.

Next, we provide some results that will be useful in the subsequent discussion.

**Lemma 3.1 (cf. [18, 24]).** *Under the assumptions of Lemma 2.1, the functional  $f(u)$  satisfies the strong convexity condition*

$$f(u_1) - f(u_2) \leq \langle \nabla f(u_1), u_1 - u_2 \rangle - \frac{\alpha}{2} \|u_1 - u_2\|^2, \quad \forall u_1, u_2 \in U^{ad}. \quad (3.7)$$

**Lemma 3.2.** *Under the assumptions of Lemma 2.1, for any  $\xi \in \mathcal{E}$  and  $u \in U^{ad}$ , the following gradient bound holds:*

$$\|\nabla F(u, \xi)\| \leq M := \frac{C_p^2}{a_{\min}} \|y_d\| + \left( \frac{C_p^4}{a_{\min}^2} + \alpha \right) \max\{|u_a|, |u_b|\} |\Omega|, \quad (3.8)$$

where

$$C_p = \sup_{y \in Y \setminus \{0\}} \frac{\|y\|}{\|\nabla y\|}$$

is the Poincaré constant and  $|\Omega| = \int_{\Omega} 1 dx$ .

*Proof.* From Lemma 2.2, we have

$$\|\nabla F(u, \xi)\| = \|\alpha u + p(u, \xi)\| \leq \alpha \|u\| + \|p(u, \xi)\|, \quad \forall \xi \in \mathcal{E}, \quad (3.9)$$

where  $p$  satisfies the adjoint equation (2.3). To estimate  $\|p\|$ , we multiply both sides of (2.3) by  $p$ , integrate over  $\Omega$ , and use the uniform boundedness of  $a(x, \xi)$ , which yields

$$\int_{\Omega} (y - y_d) p dx = \int_{\Omega} a \nabla p \cdot \nabla p dx \geq a_{\min} \|\nabla p\|^2.$$

Using the Poincaré inequality and triangle inequality, we get

$$a_{\min} \|p\|^2 \leq C_p^2 \|p\| (\|y\| + \|y_d\|).$$

Following a similar argument applied to the state equation in (1.1), we obtain the estimate

$$a_{\min} \|y\|^2 \leq C_p^2 \|u\| \|y\|.$$

For any  $u \in U^{\text{ad}}$ , we have the pointwise bound  $|u(x)| \leq \max\{|u_a|, |u_b|\}$ , which implies

$$\|u\| \leq \max\{|u_a|, |u_b|\} |\Omega|, \quad \forall u \in U^{\text{ad}}.$$

Given the boundedness of  $u, y$ , and consequently the adjoint state  $p$  is also bounded. Substituting the explicit estimates of  $\|u\|$  and  $\|p\|$  into (3.9) yields the desired bound.  $\square$

**Lemma 3.3 (cf. [29]).** *For the functional  $J(u, z)$  in problem (2.1) and any fixed  $(u, z) \in L^2(\Omega) \times L^2(\Omega)$ , suppose there exists a functional  $\phi$  such that*

$$J(u, z) - J(u^*, z^*) - \langle \lambda, u - z \rangle \leq \phi(\lambda), \quad \forall \lambda \in L^2(\Omega),$$

then, for any  $\gamma > 0$ , it holds that

$$J(u, z) - J(u^*, z^*) + \gamma \|u - z\| \leq \sup_{\|\lambda\| \leq \gamma} \phi(\lambda).$$

Moreover, if  $\epsilon := \sup_{|\lambda| \leq \gamma} \phi(\lambda) \geq 0$  and  $\gamma > \|\lambda^*\|$ , then the following estimates hold:

$$\|u - z\| \leq \frac{\epsilon}{\gamma - \|\lambda^*\|}, \quad -\frac{\|\lambda^*\| \epsilon}{\gamma - \|\lambda^*\|} \leq J(u, z) - J(u^*, z^*) \leq \epsilon.$$

**Lemma 3.4 (cf. [17]).** *Let  $\zeta^k = \zeta^k(\xi^{[k]})$  be deterministic Borel functions of  $\xi^{[k]}$  such that*

$$\mathbb{E}[\zeta^k | \xi^{[k-1]}] = 0, \quad \mathbb{E} \left[ \exp \left\{ \frac{(\zeta^k)^2}{(\nu^k)^2} \right\} \middle| \xi^{[k-1]} \right] \leq \exp \{1\},$$

where  $\nu^k > 0$  are deterministic. Then for all  $\theta > 0$ ,

$$\text{Prob} \left\{ \sum_{k=0}^K \zeta^k > \theta \sqrt{\sum_{k=0}^K (\nu^k)^2} \right\} \leq \exp \left\{ -\frac{\theta^2}{3} \right\},$$

where  $\xi^{[k]} = (\xi^0, \xi^1, \dots, \xi^k)$  is the history of random samples from the 0-th through  $k$ -th iteration.

### 3.2. Convergence property

We begin this subsection with several technical results, and then prove the convergence property of the stochastic proximal linearized ADMM.

**Lemma 3.5.** *Let the sequences  $\{u^k\}$ ,  $\{z^k\}$  and  $\{\lambda^k\}$  be generated by Algorithm 2.1. Under the assumptions of Lemma 2.1, for any  $\varrho \in (0, 2)$  and  $\lambda \in L^2(\Omega)$ , we have*

$$\begin{aligned} & \langle \nabla F(u^k, \xi^k), u^{k+1} - u^* \rangle + g(z^{k+1}) - g(z^*) \\ & \quad - \langle \hat{w}^* - \hat{w}^k, D(\hat{w}^k) \rangle + \frac{2-\varrho}{2\rho} \|\lambda^k - \hat{\lambda}^k\|^2 \\ & \leq \frac{1}{2} (\|v^k - \hat{v}^*\|_H^2 - \|v^{k+1} - \hat{v}^*\|_H^2) \\ & \quad + \frac{\eta^k}{2} (\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 - \|u^k - u^{k+1}\|^2). \end{aligned} \quad (3.10)$$

*Proof.* By Algorithm 2.1, using the expression for  $\hat{\lambda}^k$  from (3.5), we derive the first-order optimality conditions for  $u$  subproblem,  $z$  subproblem and  $\lambda$  subproblem, respectively,

$$\langle \nabla F(u^k, \xi^k) - \hat{\lambda}^k + \eta^k(u^{k+1} - u^k), u^{k+1} - u^* \rangle \leq 0, \quad (3.11)$$

$$g(z^{k+1}) - g(z^*) + \langle \hat{\lambda}^k + \rho(z^{k+1} - z^k) + (\varrho - 1)(\hat{\lambda}^k - \lambda^k), z^{k+1} - z^* \rangle \leq 0, \quad (3.12)$$

$$z^{k+1} - u^{k+1} = z^{k+1} - z^k + \frac{1}{\rho}(\hat{\lambda}^k - \lambda^k). \quad (3.13)$$

Using the definition (3.1) of  $D(w)$  and combining (3.6), (3.11)-(3.13), we obtain

$$\begin{aligned} & \langle \nabla F(u^k, \xi^k), u^{k+1} - u^* \rangle + g(z^{k+1}) - g(z^*) - \langle \hat{w}^* - \hat{w}^k, D(\hat{w}^k) \rangle \\ & \leq \eta^k \langle u^{k+1} - u^k, u^* - u^{k+1} \rangle + \langle z^* - z^{k+1}, \rho(z^{k+1} - z^k) + (\varrho - 1)(\hat{\lambda}^k - \lambda^k) \rangle \\ & \quad + \left\langle \lambda - \hat{\lambda}^k, z^{k+1} - z^k + \frac{1}{\rho}(\hat{\lambda}^k - \lambda^k) \right\rangle \\ & = \eta^k \langle u^{k+1} - u^k, u^* - u^{k+1} \rangle + \left\langle z^* - z^{k+1}, \frac{\rho}{\varrho}(z^{k+1} - z^k) + \frac{\varrho - 1}{\varrho}(\lambda^{k+1} - \lambda^k) \right\rangle \\ & \quad + \left\langle \lambda - \hat{\lambda}^k, \frac{\varrho - 1}{\varrho}(z^{k+1} - z^k) + \frac{1}{\rho\varrho}(\lambda^{k+1} - \lambda^k) \right\rangle \\ & = \eta^k \langle u^{k+1} - u^k, u^* - u^{k+1} \rangle + \langle \hat{v}^* - \hat{v}^k, H(v^{k+1} - v^k) \rangle. \end{aligned} \quad (3.14)$$

Because  $H$  is a symmetric matrix, for any  $v_1, v_2, v_3, v_4 \in V$ , it follows directly that

$$\langle v_1 - v_2, H(v_3 - v_4) \rangle = \frac{1}{2} (\|v_1 - v_4\|_H^2 + \|v_2 - v_3\|_H^2 - \|v_1 - v_3\|_H^2 - \|v_2 - v_4\|_H^2).$$

Using this result, we estimate the two terms on the right side of the inequality (3.14) as follows:

$$\begin{aligned} & \langle u^{k+1} - u^k, u^* - u^{k+1} \rangle \\ & = \frac{1}{2} (\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 - \|u^k - u^{k+1}\|^2), \end{aligned} \quad (3.15)$$

$$\begin{aligned} & \langle \hat{v}^* - \hat{v}^k, H(v^{k+1} - v^k) \rangle \\ & = \frac{1}{2} (\|v^k - \hat{v}^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 + \|v^{k+1} - \hat{v}^k\|_H^2 - \|v^k - \hat{v}^k\|_H^2). \end{aligned} \quad (3.16)$$

Since  $\varrho \in (0, 2)$ , we can simplify the last two terms in Eq. (3.16) as

$$\begin{aligned}
& \|v^k - \hat{v}^k\|_H^2 - \|v^{k+1} - \hat{v}^k\|_H^2 \\
&= \|v^k - \hat{v}^k\|_H^2 - \|v^{k+1} - v^k + v^k - \hat{v}^k\|_H^2 \\
&= \|v^k - \hat{v}^k\|_H^2 - \|(I - M)(v^k - \hat{v}^k)\|_H^2 \\
&= \frac{2 - \varrho}{\rho} \|\lambda^k - \hat{\lambda}^k\|^2.
\end{aligned} \tag{3.17}$$

By combining Eqs. (3.14)-(3.17), we complete this proof.  $\square$

**Lemma 3.6.** *Let the sequences  $\{u^k\}$ ,  $\{z^k\}$  and  $\{\lambda^k\}$  be generated by Algorithm 2.1 with  $\eta^k = \alpha(k + 1)$ . Under the assumptions of Lemma 2.1, for any  $\varrho \in (0, 2)$  and  $\lambda \in L^2(\Omega)$ , we have*

$$\begin{aligned}
& \sum_{k=0}^{K-1} \left\{ J(u^{k+1}, z^{k+1}) - J(u^*, z^*) + \langle \hat{w}^k - \hat{w}^*, D(\hat{w}^k) \rangle + \frac{2 - \varrho}{2\rho} \|\lambda^k - \hat{\lambda}^k\|^2 \right\} \\
& \leq \sum_{k=0}^{K-1} \left\{ \langle e(u^k, \xi^k), u^k - u^* \rangle + \frac{1}{\eta^k} (\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2) \right\} + \frac{1}{2} \|v^0 - \hat{v}^*\|_H^2,
\end{aligned} \tag{3.18}$$

where  $e(u^k, \xi^k) = \nabla f(u^k) - \nabla F(u^k, \xi^k)$ .

*Proof.* Based on the strong convexity of  $f(u)$  in Lemma 3.1, as stated in inequality (3.7), we obtain that for any  $u \in U^{ad}$ ,

$$\begin{aligned}
& f(u^{k+1}) - f(u) \\
&= f(u^{k+1}) - f(u^k) + f(u^k) - f(u) \\
&\leq \langle \nabla f(u^{k+1}), u^{k+1} - u^k \rangle + \langle \nabla f(u^k), u^k - u \rangle - \frac{\alpha}{2} \|u^k - u\|^2 \\
&= \langle \nabla f(u^{k+1}), u^{k+1} - u^k \rangle + \langle e(u^k, \xi^k), u^k - u \rangle - \frac{\alpha}{2} \|u^k - u\|^2 \\
&+ \langle \nabla F(u^k, \xi^k), u^k - u^{k+1} \rangle + \langle \nabla F(u^k, \xi^k), u^{k+1} - u \rangle.
\end{aligned} \tag{3.19}$$

According to the definition of  $J(u, z)$  in problem (2.1), choosing  $u = u^*$  in (3.19), and applying inequality (3.10) from Lemma 3.5 with  $\eta^k = \alpha(k + 1)$ , then for any  $\varrho \in (0, 2)$  and  $\lambda \in L^2(\Omega)$ , it holds that

$$\begin{aligned}
& J(u^{k+1}, z^{k+1}) - J(u^*, z^*) - \langle \hat{w}^* - \hat{w}^k, D(\hat{w}^k) \rangle + \frac{2 - \varrho}{2\rho} \|\lambda^k - \hat{\lambda}^k\|^2 \\
&\leq \langle e(u^k, \xi^k), u^k - u^* \rangle + \langle \nabla f(u^{k+1}) - \nabla F(u^k, \xi^k), u^{k+1} - u^k \rangle - \frac{\alpha}{2} \|u^k - u^*\|^2 \\
&+ \langle \nabla F(u^k, \xi^k), u^{k+1} - u^* \rangle + g(z^{k+1}) - g(z^*) - \langle \hat{w}^* - \hat{w}^k, D(\hat{w}^k) \rangle + \frac{2 - \varrho}{2\rho} \|\lambda^k - \hat{\lambda}^k\|^2 \\
&\leq \frac{1}{2} (\|v^k - \hat{v}^*\|_H^2 - \|v^{k+1} - \hat{v}^*\|_H^2) + \frac{\eta^{k-1}}{2} \|u^k - u^*\|^2 - \frac{\eta^k}{2} \|u^{k+1} - u^*\|^2 \\
&+ \langle e(u^k, \xi^k), u^k - u^* \rangle + \langle \nabla f(u^{k+1}) - \nabla F(u^k, \xi^k), u^{k+1} - u^k \rangle - \frac{\eta^k}{2} \|u^k - u^{k+1}\|^2.
\end{aligned} \tag{3.20}$$

By applying Young's inequality and triangle inequality, we obtain that

$$\begin{aligned}
& \langle \nabla f(u^{k+1}) - \nabla F(u^k, \xi^k), u^{k+1} - u^k \rangle - \frac{\eta^k}{2} \|u^k - u^{k+1}\|^2 \\
&\leq \frac{1}{\eta^k} (\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2).
\end{aligned} \tag{3.21}$$

According to (3.21), summing the inequality (3.20) from  $k = 0$  to  $K - 1$  directly yields the desired result of this lemma.  $\square$

**Theorem 3.1.** *Under the assumptions of Lemma 3.6, we have following convergence results:*

$$\lim_{k \rightarrow \infty} \mathbb{E} [\|\bar{u}^k - u^*\| + \|\bar{z}^k - z^*\|] = 0, \quad \lim_{k \rightarrow \infty} \mathbb{E} [\|u^k - u^*\|] = 0,$$

where

$$\bar{u}^k = \frac{1}{k} \sum_{i=1}^k u^i, \quad \bar{z}^k = \frac{1}{k} \sum_{i=1}^k z^i.$$

*Proof.* By applying Jensen's inequality, from (3.3) and (3.18), and the identity  $\lambda^k - \hat{\lambda}^k = \rho(u^{k+1} - z^k)$ , we obtain

$$\begin{aligned} & J(\bar{u}^K, \bar{z}^K) - J(u^*, z^*) + \langle \bar{w}^K - \hat{w}^*, D(\hat{w}^*) \rangle + \frac{\rho(2-\varrho)}{2} \left\| \bar{u}^K - \frac{1}{K} \sum_{k=0}^{K-1} z^k \right\|^2 \\ & \leq \frac{1}{K} \sum_{k=0}^{K-1} \left\{ J(u^{k+1}, z^{k+1}) - J(u^*, z^*) + \langle \hat{w}^k - \hat{w}^*, D(\hat{w}^*) \rangle + \frac{\rho(2-\varrho)}{2} \|u^{k+1} - z^k\|^2 \right\} \\ & = \frac{1}{K} \sum_{k=0}^{K-1} \left\{ J(u^{k+1}, z^{k+1}) - J(u^*, z^*) + \langle \hat{w}^k - \hat{w}^*, D(\hat{w}^k) \rangle + \frac{2-\varrho}{2\rho} \|\lambda^k - \hat{\lambda}^k\|^2 \right\} \\ & \leq \frac{1}{K} \sum_{k=0}^{K-1} \left\{ \langle e(u^k, \xi^k), u^k - u^* \rangle + \frac{1}{\eta^k} (\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2) \right\} \\ & \quad + \frac{1}{2K} \|v^0 - \hat{v}^*\|_H^2, \end{aligned} \tag{3.22}$$

where

$$\bar{w}^K = (\bar{u}^K, \bar{z}^K, \bar{\lambda}^K)^\top, \quad \bar{\lambda}^K = \frac{1}{K} \sum_{k=0}^{K-1} \hat{\lambda}^k.$$

From the strong convexity of  $f(u)$ , the variational inequality characterization of the optimality conditions in (3.2), and the definition of the objective functional  $J(u, z)$  in problem (2.1), we obtain the following inequality:

$$J(\bar{u}^K, \bar{z}^K) - J(u^*, z^*) + \langle \bar{w}^K - w^*, D(w^*) \rangle \geq \frac{\alpha}{2} \|\bar{u}^K - u^*\|^2. \tag{3.23}$$

Substituting  $\lambda = \lambda^*$  into inequality (3.22), together with (3.23), it follows that

$$\begin{aligned} & \frac{\alpha}{2} \|\bar{u}^K - u^*\|^2 + \frac{\rho(2-\varrho)}{4} \left\| \bar{u}^K - \frac{K-1}{K} \bar{z}^{K-1} \right\|^2 \\ & \leq \frac{\alpha}{2} \|\bar{u}^K - u^*\|^2 + \frac{\rho(2-\varrho)}{2} \left\| \bar{u}^K - \frac{K-1}{K} \bar{z}^{K-1} - \frac{z^0}{K} \right\|^2 + \frac{\rho(2-\varrho)}{2K^2} \|z^0\|^2 \\ & \leq \frac{1}{K} \sum_{k=0}^{K-1} \left\{ \langle e(u^k, \xi^k), u^k - u^* \rangle + \frac{1}{\eta^k} (\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2) \right\} \\ & \quad + \frac{1}{2K} \|v^0 - v^*\|_H^2 + \frac{\rho(2-\varrho)}{2K^2} \|z^0\|^2. \end{aligned} \tag{3.24}$$

Based on the fact that  $u^k$  depends only on  $\xi^{[k-1]}$  and is independent of  $\xi^k$ , taking the expectation of the first term on the right-hand side of (3.24), we obtain

$$\mathbb{E}[\langle e(u^k, \xi^k), u^k - u^* \rangle] = \mathbb{E}[\langle \nabla f(u^k) - \mathbb{E}[\nabla F(u^k, \xi^k) | \xi^{[k-1]}], u^k - u^* \rangle] = 0. \quad (3.25)$$

Using the estimate (3.8) in Lemma 3.2 and the choice  $\eta^k = \alpha(k+1)$ , we deduce that

$$\sum_{k=0}^{K-1} \frac{1}{\eta^k} \mathbb{E} [\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2] \leq \frac{2M^2(\ln K + 1)}{\alpha}. \quad (3.26)$$

From the inequalities (3.24)-(3.26), we obtain that

$$\lim_{K \rightarrow \infty} \mathbb{E} \|\bar{u}^K - u^*\| = 0, \quad \lim_{K \rightarrow \infty} \mathbb{E} \left\| \bar{u}^K - \frac{K-1}{K} \bar{z}^{K-1} \right\| = 0. \quad (3.27)$$

Based on the convergence result in (3.27) and the following inequality:

$$\begin{aligned} \|\bar{z}^{K-1} - z^*\| &\leq \left\| \frac{K}{K-1} \left( \frac{K-1}{K} \bar{z}^{K-1} - z^* \right) + \frac{1}{K-1} z^* \right\| \\ &\leq \frac{K}{K-1} \left\| \frac{K-1}{K} \bar{z}^{K-1} - z^* \right\| + \frac{1}{K-1} \|z^*\| \\ &\leq \frac{K}{K-1} \left( \|\bar{u}^K - u^*\| + \left\| \bar{u}^K - \frac{K-1}{K} \bar{z}^{K-1} \right\| \right) + \frac{1}{K-1} \|z^*\|, \end{aligned}$$

the first conclusion is proved.

Substituting  $\lambda = \lambda^*$  into (3.20), and combining the gradient estimate (3.21) with the first-order optimality condition (3.2) for problem (2.1), we obtain

$$\begin{aligned} &\frac{\eta^k}{2} \|u^{k+1} - u^*\|^2 - \frac{\eta^{k-1}}{2} \|u^k - u^*\|^2 + \frac{2-\rho}{2\rho} \|\lambda^k - \hat{\lambda}^k\|^2 \\ &\leq \frac{1}{2} (\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2) + J(u^*, z^*) - J(u^{k+1}, z^{k+1}) + \langle w^* - \hat{w}^k, D(\hat{w}^k) \rangle \\ &\quad + \langle e(u^k, \xi^k), u^k - u^* \rangle + \frac{1}{\eta^k} (\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2) \\ &\leq \frac{1}{2} (\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2) + \langle e(u^k, \xi^k), u^k - u^* \rangle \\ &\quad + \frac{1}{\eta^k} (\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2). \end{aligned} \quad (3.28)$$

Taking expectation on both sides of inequality (3.28) and summing over  $k = 0$  to  $K-1$ , then applying conditions (3.25) and (3.26), we derive

$$\frac{\alpha K}{2} \mathbb{E} [\|u^K - u^*\|^2] \leq \frac{1}{2} \|v^0 - v^*\|_H^2 + \frac{2M^2}{\alpha} \sum_{k=1}^K \frac{1}{k} \leq \frac{1}{2} \|v^0 - v^*\|_H^2 + \frac{2M^2(\ln K + 1)}{\alpha},$$

which establishes the second result.  $\square$

### 3.3. Convergence rate and high probability bounds

Now, we are in a position to establish the convergence rate for the stochastic proximal linearized ADMM and provide its high-probability bounds.

**Theorem 3.2.** *Under the assumptions of Theorem 3.1, for any  $\gamma > \|\lambda^*\|$ , the following convergence rate estimates hold:*

$$\begin{aligned} \mathbb{E} [J(\bar{u}^K, \bar{z}^K) - J(u^*, z^*)] &\leq \frac{C_1 \ln K}{K} + \frac{C_2}{K}, \\ \mathbb{E} [\|\bar{u}^K - \bar{z}^K\|] &\leq \frac{1}{\gamma - \|\lambda^*\|} \left( \frac{C_1 \ln K}{K} + \frac{C_2}{K} \right), \end{aligned} \quad (3.29)$$

where

$$\begin{aligned} C_1 &= \frac{2M^2}{\alpha}, \quad C_2 = C_1 + \frac{\mu}{2} (\|z^0 - z^*\|^2 + 2\|\lambda^0\|^2 + 2\gamma^2), \\ \mu &= \frac{\rho^2 + 1 + \sqrt{(\rho^2 + 1)^2 + 4\rho^2\varrho(\varrho - 2)}}{2\rho\varrho}. \end{aligned}$$

*Proof.* Taking  $w_1 = \bar{w}^K$  and  $w_2 = \hat{w}^*$  in inequalities (3.3) and (3.4), we obtain

$$-\langle \lambda, \bar{u}^K - \bar{z}^K \rangle = \langle \bar{w}^K - \hat{w}^*, D(\hat{w}^*) \rangle. \quad (3.30)$$

Substituting identity (3.30) into the left-hand side of inequality (3.22), and applying Lemma 3.3, we obtain the following estimate:

$$\begin{aligned} &J(\bar{u}^K, \bar{z}^K) - J(u^*, z^*) + \gamma \|\bar{u}^K - \bar{z}^K\| \\ &\leq \frac{1}{K} \sum_{k=0}^{K-1} \left\{ \langle e(u^k, \xi^k), u^k - u^* \rangle + \frac{1}{\eta^k} (\|\nabla f(u^{k+1})\|^2 + \|\nabla F(u^k, \xi^k)\|^2) \right\} \\ &\quad + \frac{1}{2K} \max_{\|\lambda\| \leq \gamma} \{\|v^0 - \hat{v}^*\|_H^2\}. \end{aligned} \quad (3.31)$$

By taking the expectation of (3.31) and applying the estimates from (3.25) and (3.26), we derive the following bound:

$$\begin{aligned} &\mathbb{E} [J(\bar{u}^K, \bar{z}^K) - J(u^*, z^*) + \gamma \|\bar{u}^K - \bar{z}^K\|] \\ &\leq \frac{1}{2K} \max_{\|\lambda\| \leq \gamma} \{\|v^0 - \hat{v}^*\|_H^2\} + \frac{2M^2(\ln K + 1)}{\alpha K} \\ &\leq \frac{\mu}{2K} \max_{\|\lambda\| \leq \gamma} \{\|v^0 - \hat{v}^*\|^2\} + \frac{2M^2(\ln K + 1)}{\alpha K} \\ &= \frac{\mu}{2K} \left( \|z^0 - z^*\|^2 + \max_{\|\lambda\| \leq \gamma} \{\|\lambda^0 - \lambda\|^2\} \right) + \frac{2M^2(\ln K + 1)}{\alpha K} \\ &\leq \frac{\mu}{2K} (\|z^0 - z^*\|^2 + 2\|\lambda^0\|^2 + 2\gamma^2) + \frac{2M^2(\ln K + 1)}{\alpha K}. \end{aligned} \quad (3.32)$$

The conclusion (3.29) follows by applying Lemma 3.3 in conjunction with the inequality (3.32). The proof is complete.  $\square$

Next, we will complement our discussion with the high-probability bound for our method. From Lemma 3.2, one can show that for all  $u \in U^{ad}$  it holds that

$$\mathbb{E} \left[ \exp \left\{ \frac{\|e(u, \xi)\|^2}{4M^2} \right\} \right] \leq \mathbb{E} \left[ \exp \left\{ \frac{2\|\nabla f(u)\|^2 + 2\|\nabla F(u, \xi)\|^2}{4M^2} \right\} \right] \leq \exp\{1\}. \quad (3.33)$$

With the above observation, we can derive high-probability bounds of the objective functional and constraint violation, as shown in Theorem 3.3.

**Theorem 3.3.** *Under the assumptions of Theorem 3.1, for any  $\gamma > \|\lambda^*\|$  and  $\theta \geq 0$ , we have the following probability bound:*

$$\begin{aligned} \text{Prob} \left( J(\bar{u}^K, \bar{z}^K) - J(u^*, z^*) < \frac{\theta C_3 \sqrt{\ln K}}{\sqrt{K}} \right) &\geq 1 - K^{-\frac{\theta^2}{3}}, \\ \text{Prob} \left( \|\bar{u}^K - \bar{z}^K\| < \frac{\theta C_3 \sqrt{\ln K}}{(\gamma - \|\lambda^*\|)\sqrt{K}} \right) &\geq 1 - K^{-\frac{\theta^2}{3}}, \end{aligned} \quad (3.34)$$

where

$$C_3 = 2M|\Omega|(u_b - u_a) + \frac{4M^2}{\alpha} + \frac{\mu}{2} (\|z^0 - z^*\|^2 + 2\|\lambda^0\|^2 + 2\gamma^2).$$

*Proof.* According to the inequality (3.31) and  $\eta^k = \alpha(k+1)$ , the following inequality holds:

$$\begin{aligned} &J(\bar{u}^K, \bar{z}^K) - J(u^*, z^*) + \gamma \|\bar{u}^K - \bar{z}^K\| \\ &\leq \frac{1}{K} \sum_{k=0}^{K-1} \langle e(u^k, \xi^k), u^k - u^* \rangle + \frac{2M^2(\ln K + 1)}{\alpha K} \\ &\quad + \frac{1}{2K} \max_{\|\lambda\| \leq \gamma} \{\|v^0 - \hat{v}^*\|_H^2\}, \quad \forall \gamma > 0. \end{aligned} \quad (3.35)$$

Using a similar derivation as in (3.32), we have

$$\max_{\|\lambda\| \leq \gamma} \left\{ \frac{1}{2K} \|v^0 - \hat{v}^*\|_H^2 \right\} \leq \frac{\mu}{2K} (\|z^0 - z^*\|^2 + 2\|\lambda^0\|^2 + 2\gamma^2). \quad (3.36)$$

By setting  $\zeta^k = \langle e(u^k, \xi^k), u^k - u^* \rangle / K$  and using the Cauchy-Schwarz inequality, we obtain

$$\zeta^k \leq \frac{\|e(u^k, \xi^k)\| \|u^k - u^*\|}{K}. \quad (3.37)$$

Based on the estimate (3.33) and (3.37), we derive

$$\mathbb{E} \left[ \exp \left\{ \frac{(\zeta^k)^2}{(2M|\Omega|(u_b - u_a)/K)^2} \right\} \right] \leq \exp\{1\}.$$

Then, using Lemma 3.4 and setting  $\nu^k = 2M|\Omega|(u_b - u_a)/K$ , for any  $\Theta > 0$ , we have

$$\text{Prob} \left( \sum_{k=0}^{K-1} \zeta^k > \frac{2M|\Omega|(u_b - u_a)\Theta}{\sqrt{K}} \right) \leq \exp \left\{ -\frac{\Theta^2}{3} \right\}. \quad (3.38)$$

Applying Lemma 3.3 to estimate (3.35), combining inequalities (3.36) and (3.38) with  $\Theta = \theta\sqrt{\ln K}$ , we obtain the conclusion of this theorem.  $\square$

## 4. Numerical Experiment

In this section, we evaluate the performance of the proposed method with two examples. We report a series of numerical results to demonstrate the effectiveness of our method in the first example, and compare it with some existing methods in the second example.

The initial values of all algorithms are uniformly set to the midpoint of the box constraint, and  $\lambda^0$  is consistent with  $u^0$ . To address the stochasticity introduced by the random diffusion coefficient and to ensure statistical reliability, each experiment is independently repeated 50 times. The final numerical results for the control variable  $u$  and the auxiliary variable  $z$  are obtained by averaging over these repetitions. For computational implementation, the linear finite element method is employed to discretize problem (1.1), where  $h$  denotes the mesh size of the triangulation. All experiments are conducted using MATLAB (version 2020b) on a system equipped with an AMD Ryzen 7 5800H CPU and Radeon Graphics running at 3.20 GHz.

#### 4.1. Example 1

We consider the model problem (1.1) with the domain  $\Omega = [0, 1]^2$  and the additional control constraints  $u_a = -30, u_b = 30$ . The desired state  $y_d$  is specified by

$$y_d = \frac{1}{6} \sin(2\pi x_1) \sin(2\pi x_2) \exp(2x_1).$$

The random diffusion coefficient appearing in the constraint equation in (1.1) is given by  $a(x, \xi) = \xi$ , where  $\xi$  is a uniformly distributed random variable on the interval  $[0.5, 1.5]$ . This setting represents a stochastic extension of the deterministic model introduced in [15], for which a more accurate benchmark solution to the problem (1.1) can be computed.

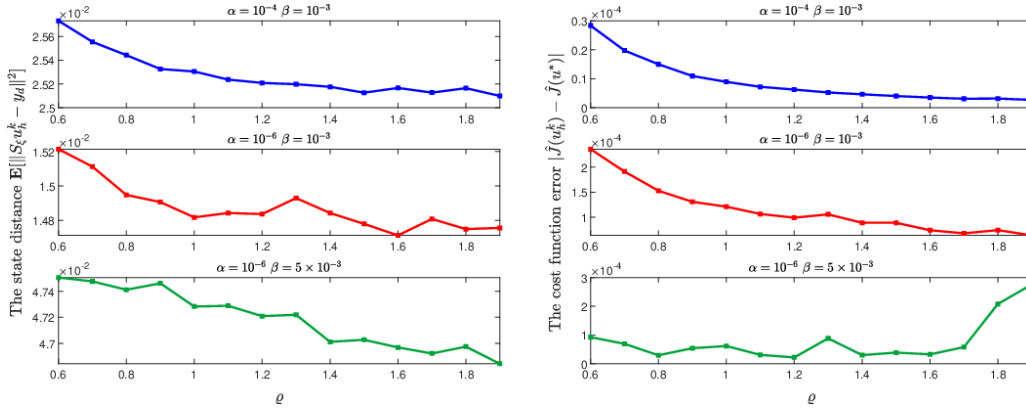
Since the solution operator of the state equation in (1.1) can be given by  $S_\xi = S/\xi$ , where  $S$  denotes the solution operator for the deterministic case with constant diffusion coefficient  $a(x, \xi) = 1$ , then model problem (1.1) reduces to the following deterministic optimization problem:

$$\begin{aligned} \min_{u \in U^{ad}} \bar{J}(u) &= \frac{1}{2} \mathbb{E} [\|Su/\xi - y_d\|^2] + \frac{\alpha}{2} \|u\|^2 + \beta \|u\|_1 \\ &= \frac{1}{2} \mathbb{E} [1/\xi^2] \|Su\|^2 + \frac{1}{2} \|y_d\|^2 - \mathbb{E}[1/\xi] \langle Su, y_d \rangle + \frac{\alpha}{2} \|u\|^2 + \beta \|u\|_1 \\ &= \frac{2}{3} \|Su\|^2 + \frac{1}{2} \|y_d\|^2 - \ln 3 \langle Su, y_d \rangle + \frac{\alpha}{2} \|u\|^2 + \beta \|u\|_1. \end{aligned}$$

In view of this, we use the numerical solution of the above deterministic model for the benchmark solution  $u^*$ , which is computed by the proximal gradient descent method with the mesh size  $h = 2^{-7}$ .

To evaluate the impact of the relaxation factor  $\varrho$  on the performance of Algorithm 2.1, we fix the mesh size at  $h = 2^{-5}$  and set the number of iterations to  $K = 500$ . We then examine the state distance and the functional error as  $\varrho$  varies under different regularization parameters. The parameters in Algorithm 2.1 are set to  $\rho = 10^{-2}$  and  $\eta^k = \alpha(k + 1)$ . Fig. 4.1 plots the results for the different relaxation factor  $\varrho$ .

It can be observed that using a relaxation factor  $\varrho > 1$  generally enhances convergence efficiency. However, in the third test case, the error in the objective functional deteriorates as  $\varrho$  approaches 2. This behavior is rooted in the properties of the relaxation factor. While values of  $\varrho > 1$  typically accelerate convergence by promoting more aggressive updates, those nearing the theoretical upper bound (here,  $\varrho = 2$ ) can undermine the positive definiteness of the matrix operator  $H$ . As  $\varrho \rightarrow 2$ , the positive definiteness of  $H$  is compromised, directly amplifying numerical errors and ultimately leading to instability. This effect is particularly pronounced in sensitivity tests, as in the third case. Overall, we recommend selecting  $\varrho$  within the range of  $[1.3, 1.6]$  to maintain a robust and efficient convergence.

Fig. 4.1. The results of Algorithm 2.1 for different  $\rho$  in Example 1.

Furthermore, we set  $\rho = 1.5$  and test the proposed method with various regularization parameters  $\alpha$  and  $\beta$ , as well as finite element mesh sizes  $h$ . The results of the objective functional error are reported in Table 4.1. We can see that the objective functional value of the numerical solution is close to that of the benchmark solution as the mesh is refined and the iteration number  $k$  increases, demonstrating the robustness of our algorithm for different mesh resolutions and regularization parameters.

Next, we present the numerical solution  $u_h^k$  for different regularization parameters with  $h = 2^{-7}$ ,  $k = 10^3$  and  $\rho = 1.5$ . Fig. 4.2 illustrates the impact of the parameters  $\alpha$  and  $\beta$  on the shape and sparsity pattern of the obtained optimal control. Specifically, the comparison between the latter two cases demonstrates that a smaller  $\beta$  leads to reduced sparsity in the control. Meanwhile, the comparison between the first two cases reveals that a smaller  $\alpha$  enlarges the size of the active set, that is, the regions where the control constraints are active. Intuitively, a smaller  $\alpha$  permits a larger  $L^2$  norm of the control. Understanding such sparsity patterns provides a priori insights into the most effective regions for the control process [25].

Finally, we investigate the shock zone and the average performance of our algorithm under mesh size  $h = 2^{-5}$  with  $\rho = 1.5$  over 50 independent iterations, considering various regularization parameter settings. The experimental means of the objective functional  $J(u_h^k, z_h^k)$  and the

Table 4.1: The results of Algorithm 2.1 for  $|\bar{J}(u_h^k) - \bar{J}(u^*)|$  in Example 1.

Regularization parameters	Iteration number	$h = 2^{-5}$	$h = 2^{-6}$	$h = 2^{-7}$
$\alpha = 1 \times 10^{-4}, \beta = 1 \times 10^{-3}$	$k = 1 \times 10^2$	$3.09 \times 10^{-4}$	$1.75 \times 10^{-4}$	$1.34 \times 10^{-4}$
	$k = 5 \times 10^2$	$1.71 \times 10^{-4}$	$3.80 \times 10^{-5}$	$4.09 \times 10^{-6}$
	$k = 1 \times 10^3$	$1.68 \times 10^{-4}$	$3.47 \times 10^{-5}$	$1.15 \times 10^{-6}$
$\alpha = 1 \times 10^{-6}, \beta = 1 \times 10^{-3}$	$k = 1 \times 10^2$	$7.40 \times 10^{-4}$	$6.40 \times 10^{-4}$	$6.04 \times 10^{-4}$
	$k = 5 \times 10^2$	$2.21 \times 10^{-4}$	$1.07 \times 10^{-4}$	$8.39 \times 10^{-5}$
	$k = 1 \times 10^3$	$1.68 \times 10^{-4}$	$7.74 \times 10^{-5}$	$5.44 \times 10^{-5}$
$\alpha = 1 \times 10^{-6}, \beta = 5 \times 10^{-3}$	$k = 1 \times 10^2$	$3.65 \times 10^{-4}$	$2.63 \times 10^{-4}$	$2.40 \times 10^{-4}$
	$k = 5 \times 10^2$	$1.82 \times 10^{-4}$	$9.19 \times 10^{-5}$	$4.15 \times 10^{-5}$
	$k = 1 \times 10^3$	$1.45 \times 10^{-4}$	$5.94 \times 10^{-5}$	$4.02 \times 10^{-5}$

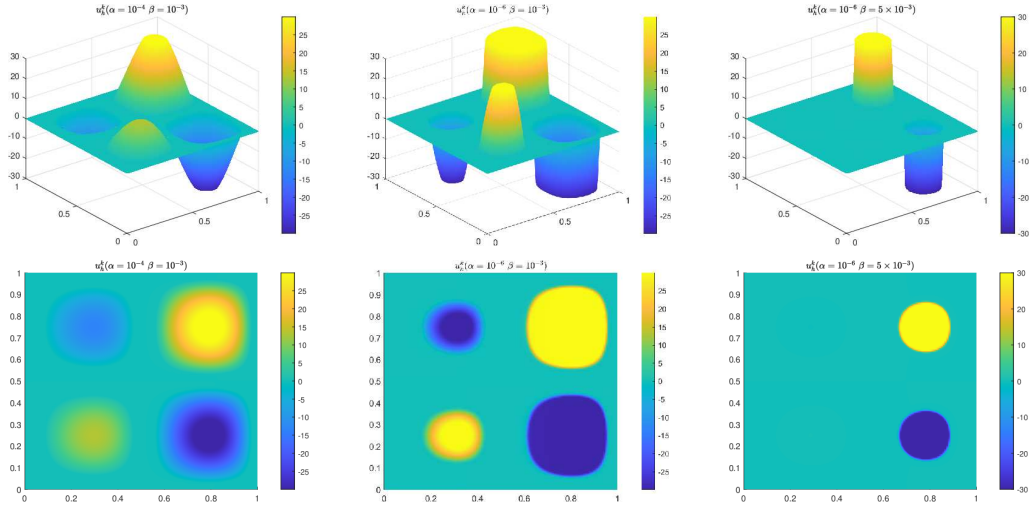


Fig. 4.2. The results of Algorithm 2.1 for numerical solutions in Example 1.

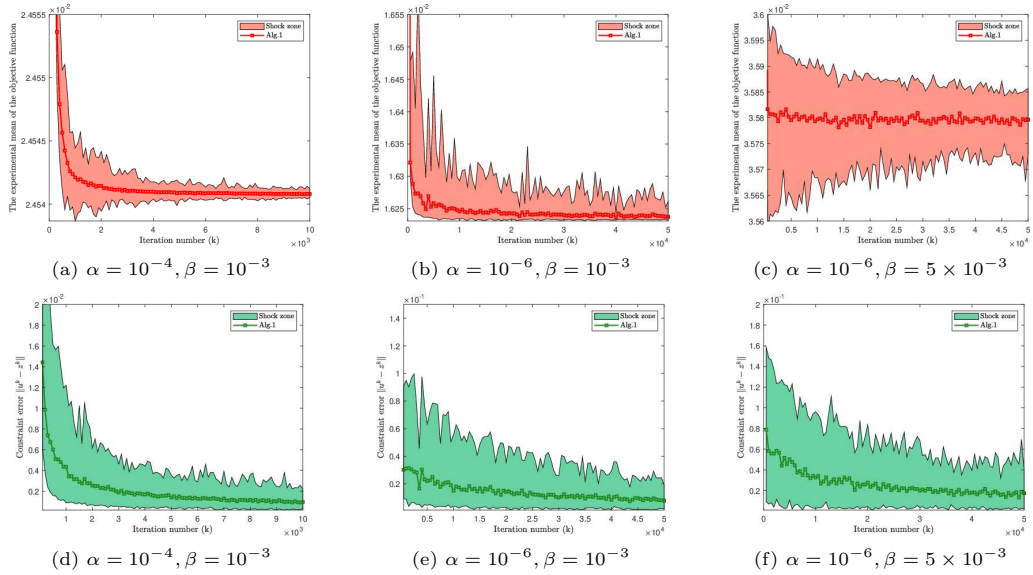


Fig. 4.3. The results of Algorithm 2.1 for high probability bounded in Example 1.

constraint error  $\|u_h^k - z_h^k\|$  are presented in the first and second rows of Fig. 4.3, respectively. While upper bounds cannot be precisely quantified, they indicate that the algorithm converges with high probability.

### 4.2. Example 2

In this example, we consider the model problem tested in [20], where the control domain  $\Omega = (0, 1)^2$  and the desired state  $y_d$  is given by

$$y_d = \begin{cases} -1, & x \in (1/4, 3/4)^2, \\ 1, & \text{otherwise,} \end{cases}$$

representing a piecewise constant target. The random diffusion coefficient of the state equation in (1.1) is set to

$$a(x, \xi) = \exp(\xi_1 \cos(1.1\pi x_1) + \xi_2 \cos(1.2\pi x_1) + \xi_3 \sin(1.3\pi x_2) + \xi_4 \sin(1.4\pi x_2)),$$

where  $\xi_1, \dots, \xi_4$  are independent random variables uniformly distributed on  $[-1, 1]$ . The control constraints are set to  $u_a = -6$  and  $u_b = 6$ .

We compared our method with the projected stochastic gradient (PSG) method [9], the stochastic gradient descent (SGD) method [23], and the stochastic subgradient projection (SSP) method [21] in this example. Referring to [9, 21, 23], the step size  $r^k$  of the PSG methods is set as

$$r^k = \frac{\theta_{psg}}{k + \nu_{psg}}, \quad \theta_{psg} = \frac{1}{2\alpha} + 1, \quad \nu_{psg} = \frac{2\theta_{psg} \times 10^{-4}}{2\alpha\theta_{psg} - 1} - 1,$$

and the step sizes  $r^k$  of the SGD and SSP methods are set to  $1/(\alpha k)$  and  $\min\{1/\alpha, 8/(\alpha(k+1))\}$ , respectively. For our method, the parameters are set as  $\rho = 10^{-2}$ ,  $\varrho = 1.5$  and  $\eta^k = \alpha(k+1)$ . Due to the complexity of the random diffusion coefficient, a benchmark solution is unavailable. Therefore, performance is evaluated by plotting the empirical average of the objective functional against running time. As shown in Fig. 4.4, our proposed method, along with SGD and SSP, consistently outperforms PSG, with our method achieving the lowest objective value.

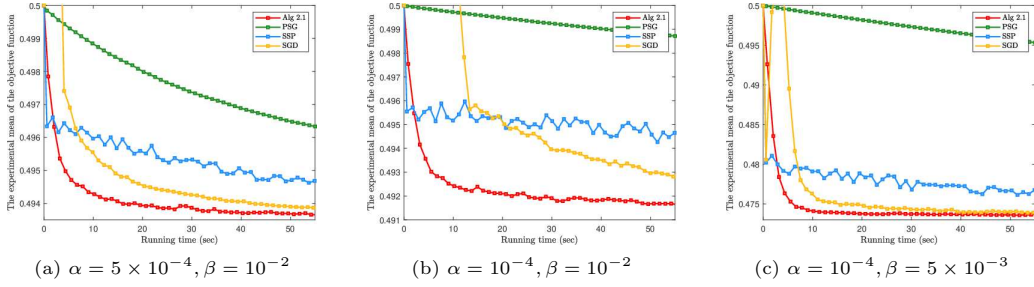


Fig. 4.4. The results of the empirical average values of objective functional in Example 2.

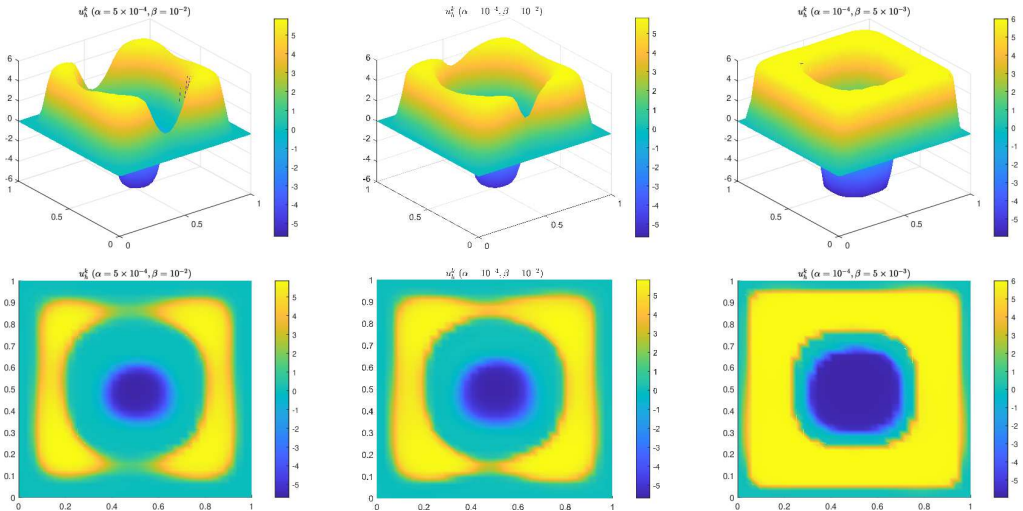


Fig. 4.5. The results of Algorithm 2.1 for numerical solutions in Example 2.

Fig. 4.5 plots the results of the numerical solution given by our method for solving the model problem with different regularization parameters, from which we can draw an analogous conclusion as in Example 1 regarding the influence of regularization parameters  $\alpha$  and  $\beta$  on the active region and sparsity of the control variable  $u$ .

## 5. Conclusion

We have developed a stochastic ADMM iteration for a class of nonsmooth stochastic optimization problems arising from the sparse distributed control of the random elliptic equations. For the separable structure of the problem, we untied the nonsmooth part from the stochastic part and proposed a proximal linearization of the stochastic subproblem by stochastic gradient. At each iteration, all subproblems can be solved explicitly. We have presented the analyses of the convergence and high probability bound of our method. Numerical results demonstrate the efficacy and competitiveness of the proposed method.

**Acknowledgements.** The work was supported by the Deep Earth Probe and Mineral Resources Exploration National Science and Technology Major Project (Grants 2025ZD1008500 and 2025ZD1008503). The work was also supported by the Supercomputing Center of Lanzhou University.

## References

- [1] I. Babuška, R. Tempone, and G.E. Zouraris, Galerkin finite element approximations of stochastic elliptic partial differential equations, *SIAM J. Numer. Anal.*, **42** (2004), 800–825.
- [2] Y.Z. Cao, S. Das, and H.W. Wyk, Adaptive stochastic gradient descent for optimal control of parabolic equations with random parameters, *Numer. Methods Partial Differential Equations*, **38** (2022), 2104–2122.
- [3] M. Eisenmann, T. Stillfjord, and M. Williamson, Sub-linear convergence of a stochastic proximal iteration method in Hilbert space, *Comput. Optim. Appl.*, **83** (2022), 181–210.
- [4] S.W. Funke, S.C. Kramer, and M.D. Piggott, Design optimisation and resource assessment for tidal-stream renewable energy farms using a new continuous turbine approach, *Renew. Energy*, **99** (2016), 1046–1061.
- [5] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, *Comput. Math. Appl.*, **2** (1976), 17–40.
- [6] L. Ge, T.J. Sun, W.F. Shen, and W.B. Liu, Adaptive stochastic meshfree methods for optimal control problem governed by random elliptic equations, *J. Comput. Math.*, **43** (2025), 813–839.
- [7] C. Geiersbach and G.C. Pflug, Projected stochastic gradients for convex constrained problems in Hilbert spaces, *SIAM J. Optim.*, **29** (2019), 2079–2099.
- [8] C. Geiersbach and T. Scarinici, Stochastic proximal gradient methods for nonconvex problems in Hilbert spaces, *Comput. Optim. Appl.*, **78** (2021), 705–740.
- [9] C. Geiersbach and W. Wollner, A stochastic gradient method with mesh refinement for PDE-constrained optimization under uncertainty, *SIAM J. Sci. Comput.*, **42** (2020), A2750–A2772.
- [10] R. Glowinski and A. Marrocco, Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité, d’une classe de problèmes de Dirichlet non linéaires, *Rev. Fr. Autom. Inform. Rech. Opér., Anal. Numér.*, **9** (1975), 41–76.
- [11] R. Glowinski, Y.C. Song, and X.M. Yuan, An ADMM numerical approach to linear parabolic state constrained optimal control problems, *Numer. Math.*, **144** (2020), 931–966.

- [12] R. Glowinski, Y.C. Song, X.M. Yuan and H.R. Yue, Application of the alternating direction method of multipliers to control constrained parabolic optimal control problems and beyond, *Ann. Appl. Math.*, **38** (2022), 115–158.
- [13] B. Gong, W.B. Liu, T. Tang, W.D. Zhao, and T. Zhou, An efficient gradient projection method for stochastic optimal control problems, *SIAM J. Numer. Anal.*, **55** (2017), 2982–3005.
- [14] P.A. Guth, V. Kaarnioja, F.Y. Kuo, C. Schillings, and I.H. Sloan, Parabolic PDE-constrained optimal control under uncertainty with entropic risk measure using quasi-Monte Carlo integration, *Numer. Math.*, **156** (2024), 565–608.
- [15] Y.H. He and J. Liu, Smoothing analysis of two robust multigrid methods for elliptic optimal control problems, *SIAM J. Matrix Anal. Appl.*, **44** (2023), 1–26.
- [16] T.L. Hou, C.M. Liu, and Y. Yang, Error estimates and superconvergence of a mixed finite element method for elliptic optimal control problems, *Comput. Math. Appl.*, **74** (2017), 714–726.
- [17] G.H. Lan, First-order and Stochastic Optimization Methods for Machine Learning, in: *Springer Series in the Data Sciences*, Springer, 2020.
- [18] M. Martin, S. Krumscheid, and F. Nobile, Complexity analysis of stochastic gradient methods for PDE-constrained optimal control problems with uncertain parameters, *ESAIM Math. Model. Numer. Anal.*, **55** (2021), 1599–1633.
- [19] M. Martin, F. Nobile and P. Tsilifis, A multilevel stochastic gradient method for PDE-constrained optimal control problems with uncertain parameters, *arXiv:1912.11900*, 2019.
- [20] J. Milz, Reliable error estimates for optimal control of linear elliptic PDEs with random inputs, *SIAM/ASA J. Uncertain. Quantification*, **11** (2023), 1139–1163.
- [21] I. Necoara and N.K. Singh, Stochastic subgradient for composite convex optimization with functional constraints, *J. Mach. Learn. Res.*, **23** (2022), 1–35.
- [22] C. Schwab and C.J. Gittelsohn, Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs, *Acta Numer.*, **20** (2011), 291–467.
- [23] O. Shamir and T. Zhang, Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes, in: *Proceedings of the 30th International Conference on Machine Learning*, **28** (2013), 71–79.
- [24] W. Shen, T. Sun, B. Gong, and W.B. Liu, Stochastic Galerkin method for constrained optimal control problem governed by an elliptic integro-differential PDE with stochastic coefficients, *Int. J. Numer. Anal. Model.*, **12** (2015), 593–616.
- [25] G. Stadler, Elliptic optimal control problems with  $L^1$ -control cost and applications for the placement of control devices, *Comput. Optim. Appl.*, **44** (2009), 159–181.
- [26] H.L. Wang, C.J. Yu, and Y.C. Song, An augmented Lagrangian method for state constrained linear parabolic optimal control problems, *J. Optim. Theory Appl.*, **203** (2024), 196–226.
- [27] M.Y. Wu, W.B. Ai, J.H. Yuan, and H. Tian, A symmetric inertial alternating direction method of multipliers for elliptic equation constrained optimization problem, *Adv. Appl. Math. Mech.*, **14** (2022), 596–621.
- [28] S.L. Wu and T. Zhou, Diagonalization-based parallel-in-time algorithms for parabolic PDE-constrained optimization problems, *ESAIM Control Optim. Calc. Var.*, **26** (2020), 88.
- [29] Y.Y. Xu, Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming, *SIAM J. Optim.*, **27** (2017), 1459–1484.
- [30] C. Zhang, X.M. Cheng, A. Angeloudis, S.C. Kramer, C.S. Wu, Y. Chen, and M.D. Piggott, Economics-constrained tidal turbine array layout optimisation at the Putuoshan-Hulu island waterway, *Ocean Eng.*, **314** (2024), 119618.
- [31] T. Zhou and T. Tang, Note on coefficient matrices from stochastic Galerkin methods for random diffusion equations, *J. Comput. Phys.*, **229** (2010), 8225–8230.