INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 19, Number 2-3, Pages 237–254

REDUCED APPROACH FOR STOCHASTIC OPTIMAL CONTROL PROBLEMS

LIUHONG CHEN, MEIXIN XIONG, AND JU MING*

Abstract. In this paper, we develop and analyze the reduced approach for solving optimal control problems constrained by stochastic partial differential equations (SPDEs). Compared to the classical approaches based on Monte Carlo method to the solution of stochastic optimal control and optimization problems, e.g. Lagrange multiplier method, optimization methods based on sensitivity equations or adjoint equations, our strategy can take best advantage of all sorts of gradient descent algorithms used to deal with the unconstrained optimization problems but with less computational cost. Specifically, we represent the sample solutions for the constrained SPDEs or the state equations by their associated inverse-operators and plug them into the objective functional to explicitly eliminate the constrains, the constrained optimal control problems are then converted into the equivalent unconstrained ones, which implies the computational cost for solving the adjoint equations of the derived Lagrange system is avoided and faster convergent rate is expected. The stochastic Burgers' equation with additive white noise is used to illustrate the performance of our reduced approach. It no doubt has great potential application in stochastic optimization problems.

Key words. SPDEs-constrained optimization problems, Lagrange multiplier method, the reduced approach, Monte Carlo finite element method.

1. Introduction

Over the past decades, the computational community has shown a growing interest in designing fast solution methods for optimal control problems constrained by stochastic partial differential equations [21,23,24,46,53]. In this case, the Monte Carlo methods are typically used in conjunction with the associated Galerkin finite element approximation in space [10,22,28] to overcome the *curse of dimensionality*, i.e., the situation in which the volume of the sample space increases exponentially with the dimension, and obtain a reliable model. However, Monte Carlo simulation typically requires a large number of sample solutions which may lead to formidable computational cost. Therefore, effective algorithms are urgently desired to solve these large-scale SPDE-constrained optimization problems in practice.

In this work, we consider a stochastic optimal control problem with tracking type objective functional. The control goal is to determine a state variable u and a deterministic control variable f to minimize

$$\mathcal{J}(u,f) := \mathbb{E}\left[\frac{1}{2}\int_0^T \|u-U\|^2_{L^2(\mathcal{D})}\,\mathrm{d}t\right] + \frac{\beta}{2}\int_0^T \|f\|^2_{L^2(\mathcal{D})}\,\mathrm{d}t \tag{1}$$

over a convex, bounded and polygonal spatial domain $\mathcal{D} \subset \mathbb{R}^d$ $(d = 2, 3)$, where U is a given expected state and usually assumed to be deterministic, $\mathbb{E}$ denotes an expected value, which is defined as the Lebesque integral in a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ described in section 2, β is a regularization parameter. u is the

Received by the editors September 29, 2021 and, in revised form, March 6, 2022.
2000 *Mathematics Subject Classification.* 35Q30, 65N30, 93E20.
*Corresponding author.

solution to a given SPDE, i.e., the state equation, which can be written in the abstract form

$$\mathcal{A}(u, f) = 0, \tag{2}$$

where the operator $\mathcal{A}$ denotes a SPDE equipped with appropriate boundary and initial conditions. (2) is usually used to model many physical, biological and economic systems subject to the influence of randomness. In brief, the constrained optimization problem we consider is then to find states u and controls f such that the functional given in (1) is minimized subject to (2).

the minimization in (1) is constrained by (2).

In the study of turbulence phenomena, the Burgers' equations can be viewed as a simplified version of the Navier-Stokes equations. Analysis and numerical approximation of optimal control problems constrained by the Burgers' equation are thus important to a variety of more complicated optimization problems in fluid dynamics. Control problems of the deterministic Burgers' equation have been studied by many authors [7, 25, 27, 30, 42, 48, 50–52], and stochastic control problems in [2, 12, 13, 29]. Here we focus on the case of stochastic Burgers' equation with additive white Gaussian noise.

To solve the large-scale optimization problems resulted from the Monte Carlo finite element (MC-FE) discretization, the classical approaches, e.g. Lagrange multiplier method, optimization method based sensitivity equations or adjoint equations [4, 17, 19, 40, 44] require the update of gradient over the samples, thus demanding repeated and costly sample solutions of the state and adjoint equations or sensitivity equations. In practice, they are typically not feasible for large-scale optimization problems due to the unaffordable computational cost for the resulted optimization system .

In this paper, we proposed the reduced approach to the stochastic optimal control problems. The reduced approach has been used to solve PDE-constrained optimization in inverse problems [31], but there are very rare literatures exploring the application of reduced approach to the stochastic optimization problems. For the solvability of the optimal control problems, in literature, there are two different strategies: Discretize-then-Optimize approach [32, 35, 39, 40] and Optimize-then-Discretize approach [32, 36, 37, 41, 45, 46], the former approach is to discretize the continuous problem and then accordingly derive for the optimality conditions, while the latter one refers to optimality condition on the continuous level is formulated first and then discretized. In our reduced approach, the Discretize-then-Optimize strategy will be adopted. Specifically, we first discretize the objective functional and the state equations, then we represent the sample solutions for the constrained SPDEs by their associated inverse-operators $\mathcal{G}$ and plug them into the discrete objective functional to explicitly eliminate the constrains. This elimination leads to a reduced objective functional $\hat{\mathcal{J}}(\mathcal{G}(f), f)$. The derived reduced system no longer has to solve the adjoint equations, but directly obtains the gradient direction through the *chain rule*. From the optimization point of view, the reduced approach can make full use of various gradient descent algorithms for unconstrained optimization problems and has low computational cost. Numerical experiments also show that the new technique works well. Moreover, much of our results and computations can be readily extended to to other optimization control problems. Figure 1 presents the outline of our optimization algorithm.

The remainder of this paper is organized as follows. In Section 2.1, we give a brief overview of some function spaces and notations. And the approximation of Brownian white noise via piecewise constant functions are discussed in Section

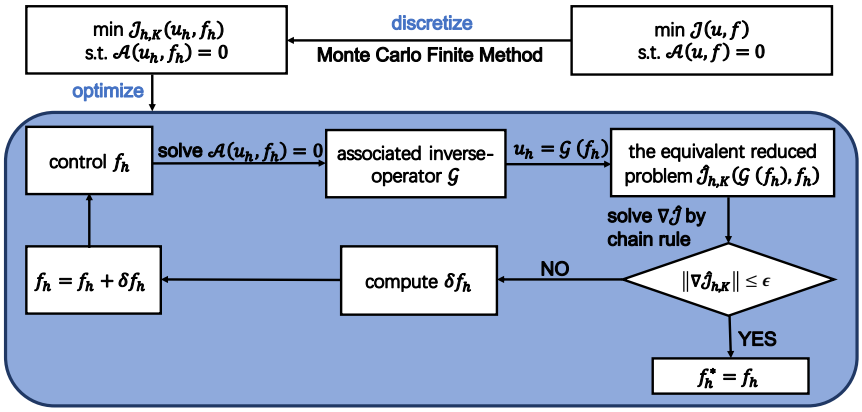


FIGURE 1. The outline of the reduced approach, where $\mathcal{J}$ is the cost functional, $\mathcal{J}_{h,K}$ its MC-FE approximation with mesh spacing h and sample size K, and f_h^* the numerical approximation using a specific optimization algorithm.

2.2. The stochastic control problem constrained by Burgers' equation with additive white noise is introduced in Section 3, as well as its the optimality system is derived by using the Lagrange multiplier method. Next, in Section 4, the reduced approach for solving the large-scale SPDE-constrained optimization problems is illustrated. In Section 4.1, we introduce the MC-FE approximation by sampling white noise for the constrained SPDE and the cost functional. The analyses of the reduced approach including algorithm implementation and complexity estimation are presented in Section 4.2 and 4.3. Numerical experiments are presented to illustrate the efficiency of the reduced approach in Section 5. Finally, in Section 6, we provide some concluding remarks.

2. Preliminaries

2.1. Function spaces and notation. We begin by recalling some function spaces and notations. Throughout this paper, we use the standard notations for Sobolev spaces (see e.g., [1]). Let $L^p(\mathcal{D}), 1 \le p \le \infty$, denote usual Lebesgue space; $\|\cdot\| = \|\cdot\|_{L^2(\mathcal{D})}$ denote the $L^2(\mathcal{D})$-norm induced by the inner product $\langle f, g\rangle = \int_{\mathcal{D}} fg\mathrm{d}\mathcal{D}$, $\forall f, g \in L^2(\mathcal{D})$. $H^r(\mathcal{D})$ is a Sobolev space for all real numbers r with norms $\|y\|_{H^r(\mathcal{D})}$ and semi-norm $|y|_{H^r(\mathcal{D})}$, where

$$\|u\|^2_{H^r(\mathcal{D})} = \sum_{|\boldsymbol{\alpha}|\le r} \left\| \frac{\partial^{|\boldsymbol{\alpha}|} u}{\partial x^{\boldsymbol{\alpha}}} \right\|^2_{L^2(\mathcal{D})}$$

and

$$|u|^2_{H^r(\mathcal{D})} = \sum_{|\boldsymbol{\alpha}|= r} \left\| \frac{\partial^{|\boldsymbol{\alpha}|} u}{\partial x^{\boldsymbol{\alpha}}} \right\|^2_{L^2(\mathcal{D})}.$$

Sobolev spaces

$$H^1(\mathcal{D}) = \left\{ y \in L^2(\mathcal{D}),\ \ \partial_{x_i} y \in L^2(\mathcal{D}),\ \ i = 1, \cdots n \right\}$$

and

$$H^1_0(\mathcal{D}) = \left\{ y \in H^1(\mathcal{D}),\ \ y\,|_{\partial\mathcal{D}} = 0 \right\}.$$

Clearly, $H^1_0(\mathcal{D}) \subset H^1(\mathcal{D})$.

We define the Hilbert space

$$L^2\left(0,T;H^r(\mathcal{D})\right)=\left\{y(t)\in H^r(\mathcal{D}),\int_0^T\|y\|_{H^r(\mathcal{D})}\mathrm{d}t<\infty\right\},$$

equipped with the norm

$$\|y\|_{L^2(0,T;H^r(\mathcal{D}))}=\left(\int_0^T\|y(t)\|^2_{H^r(\mathcal{D})}\,\mathrm{d}t\right)^{\frac{1}{2}}.$$

The mathematical formulation of a probability space denotes by $(\Omega,\mathscr{F},\mathbb{P})$, where Ω, $\mathscr{F}$ and $\mathbb{P}$ are a set of random events, the minimal σ-algebra of subsets of Ω and the probability measure, respectively.

If X is a real random variable in $(\Omega,\mathscr{F},\mathbb{P})$, we denote its expected value by

$$\mathbb{E}(X)=\int_\Omega X(\boldsymbol{\xi})\mathbb{P}(\mathrm{d}\boldsymbol{\xi})=\int_{\mathbb{R}^n}xp(\mathrm{d}x).$$

Here p is the distribution probability measure for X, defined on the Borel set $\mathcal{B}$ of $\mathbb{R}$, given by $p(\mathcal{B})=\mathbb{P}(X^{-1}(\mathcal{B}))$.

Define the stochastic Sobolev space as

$$L^2\left(\Omega;L^2\left(0,T;H^r(\mathcal{D})\right)\right)=\left\{y:\Omega\to L^2\left(0,T;H^r(\mathcal{D})\right),\mathbb{E}\left[\|y\|^2_{L^2(0,T;H^r(\mathcal{D}))}\right]<\infty\right\}$$

with the norm $\|y\|_{L^2(\Omega;L^2(0,T;H^r(\mathcal{D})))}=\mathbb{E}\left[\|y\|^2_{L^2(0,T;H^r(\mathcal{D}))}\right]$. The stochastic Sobolev space $L^2\left(\Omega;L^2\left(0,T;H^r(\mathcal{D})\right)\right)$ is a Hilbert space and is isomorphic to the tensor product space $L^2(\Omega)\otimes L^2\left(0,T;H^r(\mathcal{D})\right)$ [?]. We define the space-time cylinder $Q=(0,T]\times\mathcal{D}$ for given $T>0$. For simplicity, we set

$$\mathcal{H}^r\left(Q\right)=L^2\left(\Omega;L^2\left(0,T;H^r(\mathcal{D})\right)\right),\ \mathcal{H}_0^r\left(Q\right)=L^2\left(\Omega;L^2\left(0,T;H_0^r(\mathcal{D})\right)\right).$$

2.2. The approximation of Brownian white noise. Following [3, 15, 49], we regularize the noise through discretization. For simplicity, we divide the time interval $[0,T]$ into N_t subintervals of duration $\Delta t=T/N_t$, where $t_m=m\Delta t$, $0=t_0<t_1<\cdots<t_{N_t}=T$. For the representation of the white noise $\dot{W}(t)=\mathrm{d}W(t)/\mathrm{d}t$ we employ the piecewise constant approximation [15] defined as

$$\frac{\mathrm{d}\widehat{W}_{N_t}(t;\boldsymbol{\xi})}{\mathrm{d}t}=\frac{1}{\sqrt{\Delta t}}\sum_{n=0}^{N_t-1}\chi_n(t)\eta_n(\boldsymbol{\xi}),\tag{3}$$

where the independent and identically distributed (i.i.d.) η_n satisfies the standard normal distribution $\mathcal{N}(0,1)$ and the characteristic function χ_n is given by

$$\sqrt{\Delta t}\eta_n=\int_{t_n}^{t_{n+1}}\mathrm{d}W(t)\quad\text{and}\quad\chi_n(t)=\begin{cases}1 & \text{if } t\in[t_n,t_{n+1}),\\ 0 & \text{otherwise.}\end{cases}$$

$W(t)$ is the standard one-parameter family Brownian white noise that satisfies $\mathbb{E}\left(\frac{\mathrm{d}W(t)}{\mathrm{d}t}\cdot\frac{\mathrm{d}W(t')}{\mathrm{d}t'}\right)=\delta(t-t')$ where δ denote the usual Dirac δ-function. Then, for the piecewise constant approximation $\frac{\mathrm{d}\widehat{W}_{N_t}}{\mathrm{d}t}$ to is given by

$$\mathbb{E}\left(\frac{\mathrm{d}\widehat{W}_{N_t}(t)}{\mathrm{d}t}\cdot\frac{\mathrm{d}\widehat{W}_{N_t}(t')}{\mathrm{d}t'}\right)=\begin{cases}1/\Delta t & \text{if } t,t'\in[t_n,t_{n+1}),\\ 0 & \text{otherwise.}\end{cases}$$

Hence,

$$\lim_{N_t\to\infty}\mathbb{E}\left(\frac{\mathrm{d}\widehat{W}_{N_t}(t)}{\mathrm{d}t}\cdot\frac{\mathrm{d}\widehat{W}_{N_t}(t')}{\mathrm{d}t'}\right)=\delta(t-t').$$

Therefore, the representation (3) provides a numerical approximation of Brownian white noise.

In addition, for a deterministic function f, we can show (see [6]) that if

$$\frac{|f(t)-f(s)|}{|t-s|} \leq \kappa,$$

then there exists a constant $C(T,\kappa)$ depending on T and κ such that

$$\mathbb{E}\left[\int_0^T f(t)\mathrm{d}W(t) - \int_0^T f(t)\mathrm{d}\widehat{W}_{N_t}(t)\right]^2 \leq C(T,\kappa)\Delta t^2. \tag{4}$$

3. The distributed control problem

In this work, we consider a stochastic optimal control problem where the objective functional is of a velocity tracking type, the stochasticity arises from governing system defined by the stochastic Burgers' equations with additive white noise.

3.1. Formulation of the distributed control problem. The control problem for the viscous Burgers' equations with additive white noise can be stated as follows:

$$\min \mathcal{J}(u,f) := \mathbb{E}\left[\frac{1}{2}\int_0^T \|u-U\|^2 \,\mathrm{d}t\right] + \frac{\beta}{2}\int_0^T \|f\|^2 \,\mathrm{d}t, \tag{5}$$

subject to (s.t.)

$$\begin{cases} u_t - \nu\Delta u + uu_{\mathbf{x}} = f + \phi(t,\mathbf{x})\dot{W}(t;\boldsymbol{\xi}), & (t,\mathbf{x}) \in Q, \boldsymbol{\xi} \in \Omega, \\ u(t,\mathbf{x};\boldsymbol{\xi}) = 0, & (t,\mathbf{x}) \in \Sigma, \boldsymbol{\xi} \in \Omega, \\ u(0,\mathbf{x};\boldsymbol{\xi}) = u_0(\mathbf{x}), & \mathbf{x} \in \mathcal{D}, \boldsymbol{\xi} \in \Omega. \end{cases} \tag{6}$$

Where u and U denote the state variable and the desired state, respectively, ν the constant kinematic viscosity, u_0 a given initial condition. $\mathcal{D} \subset \mathbb{R}^2$ is a bounded spatial domain and $\partial\mathcal{D}$ denotes its boundary. We define $\mathbf{x} = (x,y)$, $Q = (0,T] \times \mathcal{D}$ and $\Sigma = (0,T] \times \partial\mathcal{D}$ for given $T > 0$. For the state variable u, we note that $u_{\mathbf{x}} = u_x + u_y$, Here u_x and u_y denote the derivatives of x and y, respectively. We allow the force term to contain a stochastic perturbation that is modeled by Brownian white noise $\dot{W}(t)$. $\phi(t,\mathbf{x})$ is an amplification factor of Brownian white noise. f is a deterministic control which is assumed measurable with respect to the σ-algebras $\mathcal{B}([0,T]) \times \mathcal{B}(\mathcal{D})$. The nonempty admissible set $\mathcal{S}$ is given by

$$\mathcal{S} = \left\{ f \in L^2(0,T;L^2(D)) : f(t,\mathbf{x}) \geq 0, \forall (t,\mathbf{x}) \in [0,T] \times \mathcal{D} \right\}. \tag{7}$$

The optimization problem we consider is to find an optimal state u and an optimal control f such that the functional $\mathcal{J}(u,f)$ defined in (5) is minimized subject to u, f satisfying the stochastic Burgers' equation (6). In this paper, we assume U is deterministic.

Generally, when discussing the stochastic process $u(t,\mathbf{x};\boldsymbol{\xi})$, we will omit the explicit dependence on the probability space, treating each realization as a deterministic PDE. For almost every (a.e.) $\boldsymbol{\xi} \in \Omega$, a given initial condition $u_0 \in H_0^1(\mathcal{D})$ and for the control $f \in L^2(L^2(0,T;L^2(D)))$, the existence and uniqueness conditions for the weak solution of the unsteady Burgers equations have already been studied, see [48] for more details.

3.2. Stochastic optimality system. The constrained problem (5) and (6) may be recast as unconstrained optimization problems through the Lagrange Multiplier method [4, 38, 47]. The first-order necessary conditions result in an *the optimality system* from which optimal states and controls can be determined.

In the following, we will take the state space $Y = \mathcal{H}_0^1(Q)$ and the control space $Z = L^2(0,T;L^2(D))$. By defining two bilinear forms

$$b[u,v] = \mathbb{E}\left[\int_{\mathcal{D}} \nu \nabla u \cdot \nabla v \mathrm{d}\mathcal{D}\right] \quad \text{and} \quad [f,v] = \mathbb{E}\left[\int_{\mathcal{D}} f v \mathrm{d}\mathcal{D}\right]$$

and the trilinear form

$$c[u,v,w] = \mathbb{E}\left[\int_{\mathcal{D}} u \cdot v_{\mathbf{x}} \cdot w \mathrm{d}\mathcal{D}\right],$$

for any $u, v, w \in Y$ and $f \in Z$. We note that $v_{\mathbf{x}} = v_x + v_y$.

Then, a weak formulation of (6) reads: find $u \in Y$ such that

$$[u_t, v] + b[u,v] + c[u,u,v] = [f,v] + \left[\phi \dot{W}, v\right], \forall v \in Y,\ t \in (0,T]. \tag{8}$$

Our problem now can be described, in short, as

$$\text{seek } (u,f) \in \mathcal{U}_{ad} \text{ such that } \mathcal{J}(u,f) \text{ is satisfied s.t. } \mathcal{A}(u,f) = 0, \tag{9}$$

where the admissibility set for the state and control variables is defined by

$$\mathcal{U}_{ad} = \{(u,f) \in Y \otimes \mathcal{S} \text{ such that (8) is satisfied and } \mathcal{J}(u,f) < \infty\}. \tag{10}$$

$\mathcal{J}(u,f)$ and $\mathcal{A}(u,f) = 0$ are given by (5) and (6) respectively.

Lemma 3.1. *It follows from Lions [34] and Fursikov [16] that the distributed optimal control problem (9) has a unique solution pair $(u,f) \in \mathcal{U}_{ad}$. By defining the Lagrangian functional*

$$\mathcal{L}(u,f,\lambda) := \mathcal{J}(u,f) - \int_0^T \left\{[u_t,\lambda] + b[u,\lambda] + c[u,u,\lambda] - [f,\lambda] - \left[\phi \dot{W}, \lambda\right]\right\} \mathrm{d}t$$

for any $(u,f,\lambda) = Y \otimes \mathcal{S} \otimes Y$. A pair (u,f) is the solution of (9) if and only if there is a co-state variable $\lambda \in Y$, such that the triplet (u,λ,f) satisfies the following optimality system:

$$[u_t, v] + b[u,v] + c[u,u,v] = [f,v] + \left[\phi \dot{W}, v\right], \quad \forall v \in Y,\ t \in (0,T]. \tag{11}$$

$$[\lambda_t, \zeta] - b[\lambda,\zeta] + c[u,\lambda,\zeta] = [U-u,\zeta], \quad \forall \zeta \in Y,\ t \in (0,T]. \tag{12}$$

$$\int_0^T [\beta f + \lambda, \vartheta - f]\,\mathrm{d}t \geq 0, \quad \forall \vartheta \in \mathcal{S}. \tag{13}$$

$$u|_{t=0} = u_0; \lambda|_{t=T} = 0. \tag{14}$$

Further, the directional derivative of $\hat{\mathcal{J}}(f) = \mathcal{J}(u(f),f)$ at $f \in \mathcal{S}$ along the direction δf is

$$\hat{\mathcal{J}}'(f)(\delta f) = \int_0^T [\beta f + \lambda, \delta f]\,\mathrm{d}t. \tag{15}$$

The *optimality system* (11)-(14) is a system of *coupled* partial differential equations whose solution yields the optimal control f, the optimal state u and the optimal adjoint λ. For the optimality system (11)-(13), the simplest form of this iteration is to

- guess values for the control f;
- solve the state system (11) for the state u;

• solve the adjoint system (12) for the adjoint λ;
• find new value of the control from the optimality condition (13).

This process is repeated until satisfactory convergence we hope is achieved. It can be shown (see, e.g., [20]) that this simple iterative method is equivalent to a *steepest descent algorithm* with a fixed step size. Clearly, such method requires the repeated solution of state and adjoint equations in the iterative stages. They are typically not feasible for large-scale problems since we cannot afford to solve a large number of state and adjoint equations.

Obviously, it is desirable to explore better ways that can provide the gradient computation with relatively few equations being performed per iteration, so as to achieve the purpose of significantly lessening the computational time.

4. Reduced Approach

In this section, we develop and analyze the reduced approach for solving stochastic optimal control problems constrained by SPDEs. Compared to the popular approaches based on Monte Carlo method to the sample solution of stochastic optimal control problems, e.g. Lagrange multiplier method introduced in section 3.2, our approach makes full use of the advantages of various gradient descent algorithms for unconstrained optimization problems and has less computational cost.

4.1. The Monte Carlo finite element method. In our approach, the strategy we adopt is Discretize-then-Optimize. We begin by discretizing the optimal control problem given in (5) and (6). (4) implies that if the interval Δt is sufficiently small, then $\frac{\mathrm{d}\widehat{W}_{N_t}(t;\boldsymbol{\xi})}{\mathrm{d}t}$ is a good approximation to $\dot{W}(t)$. Therefore, a large time interval subdivision N_t may be required for the accurate representation of the white noise, which would result in a high-dimensional problem. As such, the Monte Carlo finite element method ([22]) is a natural choice for numerical implementation to lessen the curse of dimensionality. We would try to approximate the integral $\mathbb{E}(\cdot)$ numerically by sample averages of realizations corresponding to i.i.d. random inputs, which in our case refer to the additive temporal white noise in equation (6). On the other hand, the standard finite element method is used to effect discretization with respect to (w.r.t.) the spatial variable $\mathbf{x} \in \mathcal{D}$.

Now let us investigate the structure of the algebraic system to the discretized stochastic optimal control problem. We consider a shape regular mesh $\mathcal{T}_h$ of $\mathcal{D}$ with maximum mesh size $h > 0$. u is approximated in the finite element space $V_h \subset H_0^1(\mathcal{D})$. Let $\{\varphi_j(\mathbf{x})\}_{j=1}^{N_s}$ are the basis functions of the space V_h. We approximate the state u and the control f by functions $u_h(t,\mathbf{x};\boldsymbol{\xi})$ and $f_h(t,\mathbf{x})$, respectively. Then we have

$$u_h(t,\mathbf{x};\boldsymbol{\xi}) = \sum_{j=1}^{N_s} u_j(t;\boldsymbol{\xi})\varphi_j(\mathbf{x}) \quad \text{and} \quad f_h(t,\mathbf{x}) = \sum_{j=1}^{N_s} f_j(t)\varphi_j(\mathbf{x}),$$

and (6) may be expressed as

$$\sum_{j=1}^{N_s} \frac{\mathrm{d}u_j}{\mathrm{d}t} \int_{\mathcal{D}} \varphi_j\varphi_i \mathrm{d}\mathcal{D} + \nu \sum_{j=1}^{N_s} u_j \int_{\mathcal{D}} \nabla\varphi_j \cdot \nabla\varphi_i \mathrm{d}\mathcal{D}$$
$$+ \int_{\mathcal{D}} \left(\sum_{k=1}^{N_s} u_k \frac{\partial\varphi_k}{\partial\mathbf{x}}\right)\left(\sum_{j=1}^{N_s} u_j\varphi_j\right)\varphi_i \mathrm{d}\mathcal{D} = \int_{\mathcal{D}} \left(\sum_{j=1}^{N_s} f_j(t) + \phi(t,\mathbf{x})\dot{W}(t)\right)\varphi_i \mathrm{d}\mathcal{D},$$

for $i = 1, \cdots, N_s$. Here we set $\frac{\partial\varphi_k}{\partial\mathbf{x}} = \frac{\partial\varphi_k}{\partial x} + \frac{\partial\varphi_k}{\partial y}$.

Equivalently, the semi-discrete system resulting from a finite element discretization of the Burgers' equations constitutes a nonlinear system of algebraic equations:

$$\mathbb{G}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(t;\boldsymbol{\xi})+\nu\mathbb{K}\boldsymbol{u}(t;\boldsymbol{\xi})+\mathbb{N}(\boldsymbol{u}(t;\boldsymbol{\xi}))=\mathbb{G}\boldsymbol{f}(t)+\mathbb{W},\quad t\in(0,T], \tag{16}$$

where the Gram matrix $\mathbb{G}\in\mathbb{R}^{N_s\times N_s}$, stiffness matrix $\mathbb{K}\in\mathbb{R}^{N_s\times N_s}$ and nonlinear vector function $\mathbb{N}(\boldsymbol{u}(t;\boldsymbol{\xi}))\in\mathbb{R}^{N_s}$ and $\mathbb{W}\in\mathbb{R}^{N_s}$ are respectively given by

$$\begin{aligned}
&\mathbb{G}_{ij}=\int_{\mathcal{D}}\varphi_j\varphi_i\mathrm{d}\mathcal{D},\quad \mathbb{K}_{ij}=\int_{\mathcal{D}}\nabla\varphi_j\cdot\nabla\varphi_i\mathrm{d}\mathcal{D},\\
&\mathbb{N}_i=\int_{\mathcal{D}}\frac{\partial\varphi_j}{\partial\mathbf{x}}\varphi_k\varphi_i\mathrm{d}\mathcal{D}u_k(t)u_j(t),\quad \mathbb{W}_i=\int_{\mathcal{D}}(\phi\dot{W}(t))\varphi_i\mathrm{d}\mathcal{D},\\
&\boldsymbol{u}(t;\boldsymbol{\xi})=(u_1(t;\boldsymbol{\xi}),\cdots,u_{N_s}(t;\boldsymbol{\xi}))^T\quad\text{and}\quad\boldsymbol{f}(t)=(f_1(t),\cdots,f_{N_s}(t))^T.
\end{aligned}$$

The objective functional can be also discretized by the piecewise quadratic finite elements as:

$$\mathcal{J}_h(\boldsymbol{u},\boldsymbol{f})=\mathbb{E}\left[\int_0^T\frac{1}{2}\left(\boldsymbol{u}(t;\boldsymbol{\xi})-\boldsymbol{U}(t)\right)^T\mathbb{G}\left(\boldsymbol{u}(t;\boldsymbol{\xi})-\boldsymbol{U}(t)\right)\mathrm{d}t\right]+\int_0^T\frac{\beta}{2}\boldsymbol{f}^T\mathbb{G}\boldsymbol{f}\mathrm{d}t,$$

where $(\boldsymbol{U}(t))_i=\int_{\mathcal{D}}U(\mathbf{x},t)\varphi_i(\mathbf{x})\mathrm{d}\mathcal{D}$.

Given the K i.i.d. sample realizations of the white noise function $\frac{\mathrm{d}\widehat{W}_{N_t}(t;\boldsymbol{\xi}_k)}{\mathrm{d}t}$, $k=1,\cdots,K$, we denote $\boldsymbol{u}_m^k=\boldsymbol{u}(t_m;\boldsymbol{\xi}_k)$, $\boldsymbol{U}_m=\boldsymbol{U}(t_m)$ and $\boldsymbol{f}_m=\boldsymbol{f}(t_m)$ at the different time-step $m=0,1,\cdots,N_t$. The nonlinear convective term may be linearized by *Newton's method* [18] at each time iteration step of the evolution process to handle the nonlinearity. This scheme is a semi-implicit method if we perform only one Newton iteration at each time step (see [18], Chapter 10). Stability and convergence of the semi-implicit scheme for the Burgers' equation in [5]. Moreover, using the *rectangle rule* in time leads to the following discrete objective functional

$$\mathcal{J}_{h,K}\left(\vec{\boldsymbol{u}},\vec{\boldsymbol{f}}\right)=\frac{1}{K}\sum_{k=1}^{K}\sum_{m=1}^{N_t}\frac{\Delta t}{2}\left(\boldsymbol{u}_m^k-\boldsymbol{U}_m\right)^T\mathbb{G}\left(\boldsymbol{u}_m^k-\boldsymbol{U}_m\right)+\sum_{m=1}^{N_t}\frac{\beta\Delta t}{2}\boldsymbol{f}_m^T\mathbb{G}\boldsymbol{f}_m. \tag{17}$$

We set $\vec{\boldsymbol{f}}=(\boldsymbol{f}_1,\cdots,\boldsymbol{f}_{N_t})$ and $\vec{\boldsymbol{u}}=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_{N_t})$. $\left(\boldsymbol{u}_1^k,\cdots,\boldsymbol{u}_{N_t}^k\right)^T$ are the solution of the backward Euler difference discretized approximation form for the Burgers' equations:

$$\begin{cases}
\left(\mathbb{G}+\nu\Delta t\mathbb{K}+\Delta t\mathbb{B}(\boldsymbol{u}_m^k)+\Delta t\mathbb{C}(\boldsymbol{u}_m^k)\right)\boldsymbol{u}_{m+1}^k-\left(\mathbb{G}+\Delta t\mathbb{C}(\boldsymbol{u}_m^k)\right)\boldsymbol{u}_m^k\\
-\Delta t\widehat{\mathbb{W}}_{m+1}=\Delta t\mathbb{G}\boldsymbol{f}_{m+1},\\
\boldsymbol{u}(0;\boldsymbol{\xi}_k)=\boldsymbol{u}_0,
\end{cases} \tag{18}$$

for $m=0,1,\cdots,N_t, k=1,\cdots,K$, $\boldsymbol{u}_0=(u_0(\mathbf{x}_1),\cdots,u_0(\mathbf{x}_{N_s}))^T$ and

$$\widehat{\mathbb{W}}_m=\int_{\mathcal{D}}(\phi\frac{\mathrm{d}\widehat{W}_{N_t}(t_m)}{\mathrm{d}t})\varphi_i\mathrm{d}\mathcal{D},(\mathbb{B})_{ij}=\int_{\mathcal{D}}u\frac{\partial\varphi_i}{\partial\mathbf{x}}\varphi_j\mathrm{d}\mathcal{D},(\mathbb{C})_{ij}=\int_{\mathcal{D}}u_{\mathbf{x}}\varphi_i\varphi_j\mathrm{d}\mathcal{D}.$$

The discretized optimal control problem is then given as follows: minimize $\mathcal{J}_{h,K}$ subject to the K realizations (18) of the constrained equations. After having derived the discreted system corresponding to the optimal control problem, we need to discuss how to solve this system efficiently.

4.2. Reduced approach. To solve the coupled optimality system by the Lagrange Multiplier method requires solving a large number of constrained SPDEs or state equations and adjoint equations at each update, which forces us to expect an effective method with lower computational cost and faster convergence speed. Most widely applied optimization algorithms use the gradient of the objective functional w.r.t. the controls to determine the descent direction. A typical optimization algorithm proceeds as follows.

Algorithm 1 `Gradient-based algorithm`

Input: initial guess $\vec{\boldsymbol{f}}^{(0)}$ for the control variable
Output: optimal control $\vec{\boldsymbol{f}}^{(*)}$
1: For $\ell = 0, 1, 2, ...,$ until satisfactory convergence is achieved;
2: Solve (18) to obtain the corresponding state $\vec{\boldsymbol{u}}^{k,(\ell)} = \boldsymbol{u}(\vec{\boldsymbol{f}}^{(\ell)})$ for $k = 1, \cdots, K$;
3: Compute the gradient of the functional $\nabla \mathcal{J}_{h,K}^{(\ell)}$;
4: Use the results of step 3 to compute a step $\delta \vec{\boldsymbol{f}}^{(\ell)}$;
5: Set $\vec{\boldsymbol{f}}^{(\ell+1)} = \vec{\boldsymbol{f}}^{(\ell)} + \delta \vec{\boldsymbol{f}}^{(\ell)}$;
6: **return** $\vec{\boldsymbol{f}}^{(\ell)}$

Here, we focus on Step 3. It will be important to investigate the possibilities of computing the derivative of the reduced objective function $\hat{\mathcal{J}}_{h,K}$. After having derived the discrete system (17) and (18), the so-called reduced approach we propose is based on a elimination of the constraints to formulate an unconstrained optimization problem. We represent the sample solutions for the constrained SPDEs or the state equations by their associated inverse-operators. For mth time step of the kth realization, we rewrite the discrete system (18) as

$$\begin{aligned} \boldsymbol{u}_{m+1}^k := & \left[\mathbb{G} + \nu\Delta t\mathbb{K} + \Delta t\mathbb{B}(\boldsymbol{u}_m^k) + \Delta t\mathbb{C}(\boldsymbol{u}_m^k)\right]^{-1} \\ & \left[\left(\mathbb{G} + \Delta t\mathbb{C}(\boldsymbol{u}_m^k)\right)\boldsymbol{u}_m^k + \Delta t\mathbb{G}\boldsymbol{f}_{m+1} + \Delta t\widehat{\mathbb{W}}_{m+1}\right]. \end{aligned} \tag{19}$$

Accordingly, we denote the inverse-operator

$$\mathcal{G}_m^k := \left[\mathbb{G} + \nu\Delta t\mathbb{K} + \Delta t\mathbb{B}(\boldsymbol{u}_m^k) + \Delta t\mathbb{C}(\boldsymbol{u}_m^k)\right]^{-1}. \tag{20}$$

In order to emphasize the dependence of $\boldsymbol{u}_{m+1}^k$ and $\boldsymbol{f}_{m+1}$, we will omit the other terms in (19). Let $\boldsymbol{u}_{m+1}^k := \mathcal{G}_m^k(\boldsymbol{f}_{m+1})$ denote the dependence. In this case there exists a unique vector $\boldsymbol{u}_{m+1}^k$ for any $\boldsymbol{f}_{m+1}$. We plug it into the objective functional to explicitly eliminate the constrains, the constrained optimal control problem is then converted into the equivalent unconstrained problem. The discretized, reduced optimal control problem is as follows:

$$\begin{aligned} \hat{\mathcal{J}}_{h,K}\left(\vec{\boldsymbol{f}}\right) = & \frac{1}{K}\sum_{k=1}^{K}\sum_{m=1}^{N_t}\frac{\Delta t}{2}\left(\mathcal{G}_{m-1}^k(\boldsymbol{f}_m) - \boldsymbol{U}_m\right)^T\mathbb{G}\left(\mathcal{G}_{m-1}^k(\boldsymbol{f}_m) - \boldsymbol{U}_m\right) \\ & + \sum_{m=1}^{N_t}\frac{\beta\Delta t}{2}\boldsymbol{f}_m^T\mathbb{G}\boldsymbol{f}_m. \end{aligned} \tag{21}$$

Assume that when we are going to update the control $\boldsymbol{f}_m$ of the current instant t_m, the control at the previous $m-1$ instant has been updated, that is, the update of the control variable at t_m is independent of the previous $m-1$ instant. Based

on the chain rule, we can find the gradient directly of the unconstrained reduced objective functional at each time step $t_m, m = 0, 1, \cdots, N_t$

$$\nabla \hat{\mathcal{J}}_{h,K}(\boldsymbol{f}_m) = \frac{D\hat{\mathcal{J}}_{h,K}}{D\boldsymbol{f}_m} = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{\mathrm{d}\boldsymbol{u}_m^k}{\mathrm{d}\boldsymbol{f}_m}\right)^T \frac{\partial \hat{\mathcal{J}}_{h,K}}{\partial \boldsymbol{u}_m^k} + \frac{\partial \hat{\mathcal{J}}_{h,K}}{\partial \boldsymbol{f}_m}. \tag{22}$$

The terms $\partial\hat{\mathcal{J}}_{h,K}/\partial\boldsymbol{u}_m^k$ and $\partial\hat{\mathcal{J}}_{h,K}/\partial\boldsymbol{f}_m$ are usually easy to determine. The terms $\partial\hat{\mathcal{J}}_{h,K}/\partial\boldsymbol{u}_m^k$ and $\partial\hat{\mathcal{J}}_{h,K}/\partial\boldsymbol{f}_m$ of the unconstrained optimization problem (21) are:

$$\frac{\partial \hat{\mathcal{J}}_{h,K}}{\partial \boldsymbol{u}_m^k} = \Delta t\mathbb{G}\left(\mathcal{G}_{m-1}^k(\boldsymbol{f}_m) - \boldsymbol{U}_m\right), \quad \frac{\partial \hat{\mathcal{J}}_{h,K}}{\partial \boldsymbol{f}_m} = \beta\Delta t\mathbb{G}\boldsymbol{f}_m. \tag{23}$$

By direct differentiation of the state discrete algebraic systems (19) w.r.t the control $\boldsymbol{f}_m$, the *sensitivities* $\mathrm{d}\boldsymbol{u}_m^k/\mathrm{d}\boldsymbol{f}_m$ can be directly written as

$$\frac{\mathrm{d}\boldsymbol{u}_m^k}{\mathrm{d}\boldsymbol{f}_m} = \left[\left(\mathbb{G} + \nu\Delta t\mathbb{K} + \Delta t\mathbb{B}(\boldsymbol{u}_{m-1}^k) + \Delta t\mathbb{C}(\boldsymbol{u}_{m-1}^k)\right)\right]^{-1}\Delta t\mathbb{G}. \tag{24}$$

Substitute (23) and (24) into (22), the forms of the gradient of $\hat{\mathcal{J}}_{h,K}$ are given by

$$\begin{aligned}\nabla \hat{\mathcal{J}}_{h,K}(\boldsymbol{f}_m) =& \frac{1}{K}\sum_{k=1}^{K}\left\{\left[\left(\mathbb{G} + \nu\Delta t\mathbb{K} + \Delta t\mathbb{B}(\boldsymbol{u}_{m-1}^k) + \Delta t\mathbb{C}(\boldsymbol{u}_{m-1}^k)\right)\right]^{-1}\Delta t\mathbb{G}\right\}^T \\ & \Delta t\mathbb{G}\left(\boldsymbol{u}_m^k - \boldsymbol{U}_m\right) + \beta\Delta t\mathbb{G}\boldsymbol{f}_m,\end{aligned} \tag{25}$$

for the $m = 1, 2, \cdots, N_t$. Moreover, it should be noted that gradient computation depends on the selection of time discrete scheme for time-dependent optimization problems.

Formulation (25) implies that via the reduced approach, the gradient of the reduced objective functional could be obtained directly and the computational cost for solving the adjoint equations of the derived Lagrange system is saved. Moreover, the reduced objective functional satisfies the constraint condition in each iteration in the reduced approach.

A basis *Steepest descent method* algorithm for minimizing $\hat{\mathcal{J}}_{h,K}$ via the reduced approach is given in Algorithm 2. If the algorithm terminates successfully, the final iterates additionally satisfy $\frac{\left|\hat{\mathcal{J}}_{h,K}^{(\ell)} - \hat{\mathcal{J}}_{h,K}^{(\ell+1)}\right|}{\left|\hat{\mathcal{J}}_{h,K}^{(\ell)}\right|} \leq \epsilon$ or $\left\|\nabla\hat{\mathcal{J}}_{h,K}^{(\ell)}\right\| \leq \epsilon$.

For Step 9 of Algorithm 2, one can use one's favorite *optimization method* [43] for $\delta\boldsymbol{f}_m^{(\ell)}$, e.g., Newton method [11, 33], quasi-Newton method [8, 9], trust region method [14, 26], etc. Since it is not the focus in our work, we will not go into details here. For the numerical experiment in Section 5, we will give the optimization results of few optimization method.

4.3. Computational complexity estimates. From Section (4.2), we know that the computational cost for solving the adjoint equations of the derived Lagrange system is avoided for the reduced approach, but directly solves the gradient of the reduced objective functional. A summary of the computational costs per iteration of Lagrange Multiplier method (LMM) and the reduced approach (RA) is given in Table 1. The symmetric, positive definite coefficient matrix $\mathbb{G}, \mathbb{M}$ does not change from time step to time step, so that it needs to be factored only once for the numerical implementation. In addition, it is also noted that both methods adopt the steepest gradient descent method, so the number of optimization iterations of reduced approach is not more than that of Lagrange multiplier method. We can

Algorithm 2 Steepest descent method algorithm for minimizing $\hat{\mathcal{J}}_{h,K}$ via the reduced approach.

Input: A tessellation $\mathcal{T}_h$ of $\mathcal{D}$, time step Δt, interval size N_t.
Output: Numerical optimal control $\vec{\boldsymbol{f}}^{(*)}$.
I. Initialization.
1: Construct $V_h \subset H_0^1(\mathcal{D})$, initial control guess $f^{(0)}$, step size γ, tolerance ϵ, $\ell = 0$.
2: Sample the piecewise constant white noise $\frac{\mathrm{d}\widehat{W}_{N_t}(t,\boldsymbol{\xi}_k)}{\mathrm{d}t}, k = 1, \cdots, K$.
3: Solve $\boldsymbol{u}_m^{k,(\ell)} := \mathcal{G}_{m-1}^k\left(\boldsymbol{f}_m^{(\ell)}\right)$ for $k = 1, \cdots, K$; $m = 1, \cdots, N_t$.
II. Optimization loop.
4: **while** $\frac{\left|\hat{\mathcal{J}}_{h,K}^{(\ell)} - \hat{\mathcal{J}}_{h,K}^{(\ell+1)}\right|}{\left|\hat{\mathcal{J}}_{h,K}^{(\ell)}\right|} \leq \epsilon$ or $\left\|\boldsymbol{g}^{k,(\ell)}\right\| \leq \epsilon$ **do**
5: **for** $k = 1, \cdots, K$ **do**
6: Compute the reduced gradient $\boldsymbol{g}_m^{k,(\ell)}$ by (25) of $\hat{\mathcal{J}}_{h,K}^{(\ell)}$

$$\boldsymbol{g}_m^{k,(\ell)} = \left\{\left[\left(\mathbb{G} + \nu\Delta t\mathbb{K} + \Delta t\mathbb{B}(\boldsymbol{u}_{m-1}^{k,(\ell)}) + \Delta t\mathbb{C}(\boldsymbol{u}_{m-1}^{k,(\ell)})\right)\right]^{-1}\mathbb{G}\right\}^T$$
$$\Delta t\mathbb{G}\left(\boldsymbol{u}_m^{k,(\ell)} - \boldsymbol{U}_m\right) + \beta\Delta t\mathbb{G}\boldsymbol{f}_m^{(\ell)}, \text{ for } m = 1, \cdots, N_t.$$

7: **end for**
8: $\boldsymbol{g}_m^{(\ell)} = \frac{1}{K}\sum_{k=1}^K \boldsymbol{g}_m^{k,(\ell)}$, use $\boldsymbol{g}_m^{(\ell)}$ to compute a step $\delta\boldsymbol{f}_m^{(\ell)}$ for $m = 1, \cdots, N_t$.
9: Choose $\delta\boldsymbol{f}_m^{(\ell)} = -\gamma\boldsymbol{g}_m^{(\ell)}$ with step size γ,

$$\boldsymbol{f}_m^{(\ell+1)} = \boldsymbol{f}_m^{(\ell)} + \delta\boldsymbol{f}_m^{(\ell)} \text{ for } m = 1, \cdots, N_t.$$

10: Solve $\boldsymbol{u}_m^{k,(\ell+1)} := \mathcal{G}_{m-1}^k\left(\boldsymbol{f}_m^{(\ell+1)}\right)$ for $k = 1, \cdots, K$; $m = 1, \cdots, N_t$.
11: $\ell = \ell + 1$.
12: **end while**
13: **return** $\boldsymbol{f}_m^{(*)} := \boldsymbol{f}_m^{(\ell)}, m = 1, \cdots, N_t$.

TABLE 1. Computational and storage costs per iteration of different methods; K denotes the number of realizations and N_s denotes the spatial dimension and N_t denotes the number of parameters.

Method	# PDE's	# Storage
Lagrange Multiplier method	$2K$	$(2KN_s + 1)N_t$
Reduced approach	K	$(KN_s + 1)N_t$

also view that the number of iterations of the reduced approach (T_{RA}) is much smaller than that of the Lagrangian multiplier method (T_{LMM}) from the numerical experiment results in Section 5. We have

$$\mathcal{O}\left(KT_{\mathrm{RA}}\right)) \ll \mathcal{O}\left(2KT_{\mathrm{LMM}}\right). \tag{26}$$

5. Numerical experiments

In this section, numerical experiments on the stochastic control problem (9) are presented to validate the effectiveness of our proposed reduced approach. We carry out simulations by using MATLAB R2017a software on an Intel Xeon W machine

with 256GB of memory. The examples demonstrate that the reduced approach does a very good job of stochastic optimal control problems.

The numerical examples are tested for the above algorithm with following setting. Our model problem has the desired state

$$U(\mathbf{x},t) = (1+t^2)\sin(2\pi x)\sin(2\pi y).$$

The Figure 2 shows the desired state at $t = 0.2, 0.5, 0.8$.

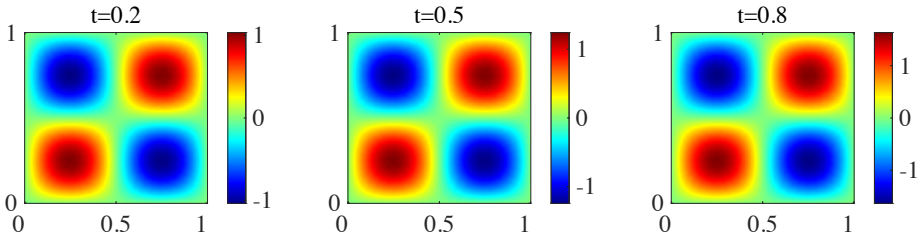


FIGURE 2. Desired state at $t = 0.2, 0.5, 0.8$.

We choose the domain $D = [0,1] \times [0,1]$, $\mathbf{x} = (x, y)$. The initial control guess can be chosen arbitrarily as long as the boundary conditions on u are not violated, here we let $f = 0$. The initial state is chosen as

$$u_0(\mathbf{x}) = \sin(\pi x)\sin(\pi y).$$

We choose the viscosity constant $\nu = 0.01$, regularization parameter $\beta = 1 \times 10^{-4}$ and the time step $\Delta t = 0.005$ for the computation. The Taylor-Hood elements are utilized to effect discretization w.r.t. the spatial variable $\mathbf{x}$. In addition, the tolerance parameter $\epsilon = 1 \times 10^{-3}$.

For $t \in [0, T]$, the temporal white noise $\mathrm{d}W(t;\boldsymbol{\xi})/\mathrm{d}t$ can be approximated by

$$\frac{\mathrm{d}\widehat{W}_{N_t}(t;\boldsymbol{\xi})}{\mathrm{d}t} = \frac{\sigma}{\sqrt{\Delta t}} \sum_{i=0}^{N_t-1} \chi_i(t)\eta_i(\boldsymbol{\xi}).$$

Figure 3 shows the average values of $\mathrm{d}W(t;\boldsymbol{\xi})/\mathrm{d}t$ by taking $K = 100$ realizations, where we set $\sigma = 0.1$ and $\phi(\boldsymbol{x}) = 10\cos(x_1)\sin(x_2)$. Specifically, we plot the approximation

$$\mathbb{E}_K\left(\mathrm{d}W(t;\boldsymbol{\xi})/\mathrm{d}t\right) = \frac{1}{K}\sum_{k=1}^{K} \frac{\mathrm{d}\widehat{W}_{N_t}(t;\boldsymbol{\xi}_k)}{\mathrm{d}t}$$

of the expect value $\mathbb{E}\left(\mathrm{d}W(t;\boldsymbol{\xi})/\mathrm{d}t\right)$ of $\mathrm{d}W(t;\boldsymbol{\xi})/\mathrm{d}t$.

The computational results obtained by implementing Algorithm 2 with the above data are presented in graphical form. The effect of the above optimization algorithm is presented in Figure 4 and 5. In Figure 6 we show the optimal control obtained when the stopping criterion was met.

We present the comparison results of the Lagrangian multiplier method and the reduced approach. In order to demonstrate and isolate the stochastic effects, we first consider the two methods with constant step sizes. The step size is related to the derived gradient scheme (25). The gradient scheme contains Δt^2. In our numerical experiment,Δt taken as $1/200$, which leads to the small value. In addition, this is also different from the gradient scheme on the whole time integral derived by Lagrange multiplier method. In order to eliminate the stochastic effects, we set that the increment in the control variable $\delta\vec{\boldsymbol{f}}$ of the two methods are on the same

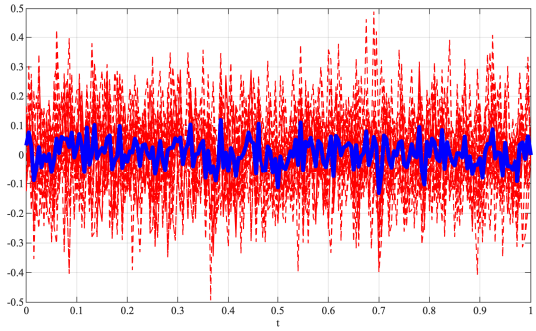


FIGURE 3. The red dotted lines present the realizations of the white noise $\mathrm{d}W(t;\boldsymbol{\xi})/\mathrm{d}t$. The blue line presents the approximation of expected value $\mathbb{E}_K\left(\mathrm{d}W(t;\boldsymbol{\xi})/\mathrm{d}t\right)$.

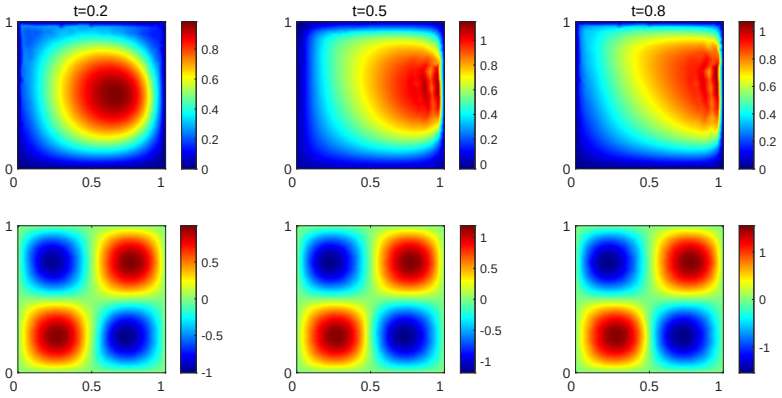


FIGURE 4. Row 1 is the sample of the initial state $u^{(0)}$ without control at t=0.2, 0.5, 0.8, respectively. Row 2 is the sample of the controlled stochastic state at t=0.2, 0.5, 0.8, respectively.

order of magnitude at the initial iteration $\ell = 0$. Hence, the step size of the reduced approach at the initial iteration $\ell = 0$ reached the order of 10^7 in our experiment. Both of them choose the steepest descent direction of the move to the next iterate. The control results including step size, iterations, CPU times, time per iteration, the initial objective functional $\hat{\mathcal{J}}_{h,K}^{(0)}$, the final objective functional $\hat{\mathcal{J}}_{h,K}^{(*)}$ and the ratio $\hat{\mathcal{J}}_{h,K}^{(*)}/\hat{\mathcal{J}}_{h,K}^{(0)}$ are shown in Table 2.

In this case, We can clearly see that the advantages of our method are further magnified in Table 2. The CPU time per iteration has been reduced by more than 60 seconds! In addition, the number of iterations of the reduced approach satisfying the same termination condition is much less than that of the Lagrange multiplier method. We know that with the increase in the number of iterations, the computer performance will decline, resulting in a longer execution time. The CPU time of the LMM simulation is 7.9×10^4 seconds, and the CPU time of RA is 2.1×10^4 seconds. The RA reduces the computational cost by about 73%. Since the optimality system derived by the Lagrangian multiplier method needs to solve the state equations and

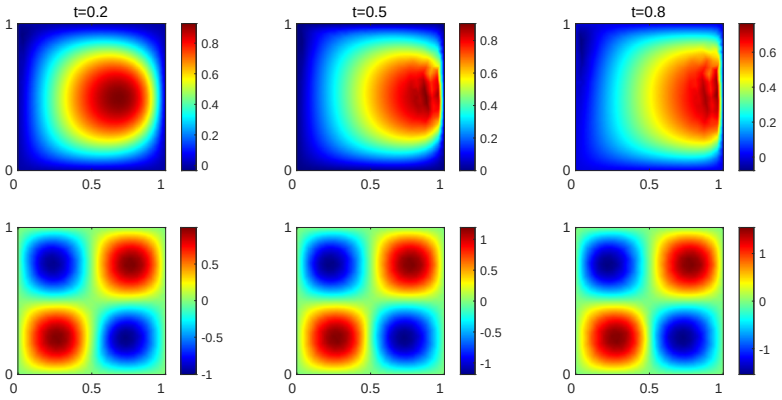


FIGURE 5. Row 1 is the mean of the initial state $u^{(0)}$ without control at t=0.2, 0.5, 0.8, respectively. Row 2 is the mean of the controlled stochastic state at t=0.2, 0.5, 0.8, respectively.

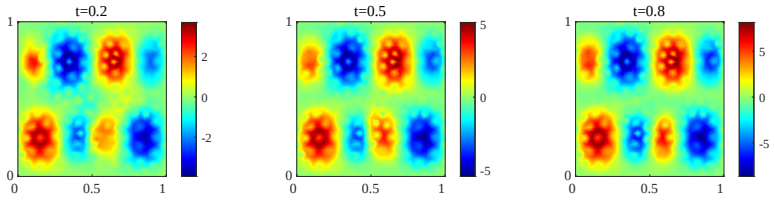


FIGURE 6. Optimized control f at t=0.2, 0.5, 0.8, respectively.

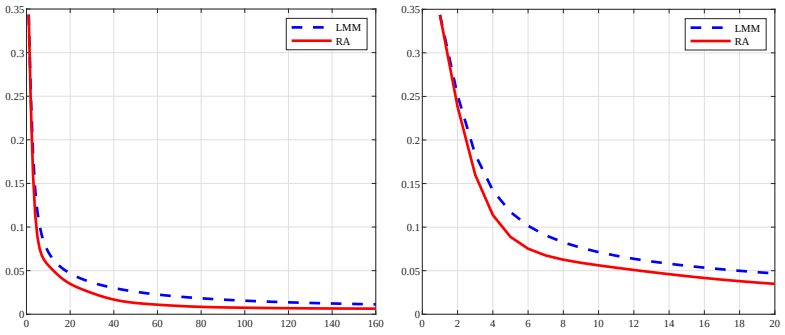
adjoint equations of K samples at each iteration, the high computational cost is unaffordable for large-scale problem!

TABLE 2. Simulation results of Lagrange multiplier method and the reduced approach.

	step size	iteration	CPU times	time per iteration	$\hat{\mathcal{J}}_{h,K}^{(0)}$	$\hat{\mathcal{J}}_{h,K}^{(*)}$	$\hat{\mathcal{J}}_{h,K}^{(*)}/\hat{\mathcal{J}}_{h,K}^{(0)}$
LMM	1	413	79335s	192.1s	0.3436	0.0070	0.0203
RA	10^7	164	21328s	130.0s	0.3436	0.0064	0.0186

The remarkable thing is that the reduced approach does not need to calculate the adjoint equations, which is equivalent to decoupling the adjoint variables and the state variables. In the implementation of the algorithm, we can pre-store the state variables of N_t instants (possibly increase the storage), and then compute the gradient of each time step in parallel. Such processing may further speed up the algorithm's execution time. This makes the approach more attractive for large-scale problems from a computational point-of-view.

We present the convergence history of the online iteration with objective functional. As observed from the Figure 7, the reduced approach converge faster than the Lagrange Multiplier method.



(a) (b)

FIGURE 7. (a) The convergence history of the objective functional during the iteration. (b) The objective functional for the first 20 steps of iteration.

Clearly, our approach can directly apply all types of gradient descent algorithms designed for the unconstrained optimization problems. We now use several widely used optimization algorithms to illustrate the advantages of the our method. Newton and quasi-newton methods take the constant step sizes. For the line search methods, we employ the *Weak Wolf condition*. As can be observed from Table 3, the optimization algorithm based on the reduced approach do a good job for solving the stochastic control problem proposed in this work. Obviously, the numerical results fully prove the effectiveness and advantages of our proposed algorithm.

TABLE 3. Simulation results for other unconstrained optimization algorithms with the reduced approach.

	step size	iteration	CPU times	time per iteration	$\hat{\mathcal{J}}_{h,K}^{(0)}$	$\hat{\mathcal{J}}_{h,K}^{(*)}$	$\hat{\mathcal{J}}_{h,K}^{(*)}/\hat{\mathcal{J}}_{h,K}^{(0)}$
Newton	10^{-2}	69	22774s	330.1s	0.3436	0.0063	0.0183
Line search	10^{7}	61	32386s	530.9s	0.3436	0.0059	0.0172
BFGS	10^{-2}	89	17818s	200.2s	0.3436	0.0067	0.0195

6. Conclusions

In this paper, we show how the reduced approach is used to efficiently obtain accurate approximations of the solutions of optimal control problems for stochastic Burgers' equation with additive white noise. The method is based on the so-called Discretize-then-Optimize strategy. For numerical discretization, the MC-FE method is adopted to overcome the curse of dimensionality. For the solution of such a large-scale SPDEs-constrained optimization problem, we present the reduced approach. Unlike the classical approaches, Lagrange multiplier method, optimization method based sensitivity equations or adjoint equations, the proposed reduced approach directly plug the sample solutions of the constraints (18) into the discretize objective functional (17) to explicitly eliminate the constrains. This reformulation results in an equivalent unconstrained optimization problem. The reduced approach

does not require the computational cost of solving the adjoint equations of the Lagrange system, and it can also take great advantage of all types of gradient descent algorithms. This makes the approach more attractive for large-scale problems from a computational point-of-view. Numerical results on optimal control problems for stochastic Burgers' equation with additive white noise validate the feasibility and the effectiveness of the proposed methods.

However, there are still many open questions to be answered, for examples, although the reduced approach shows significant improvement, theoretical analysis is still absent. In addition, although the reduced approach has been proposed a long time ago, there may be two reasons why it has not been widely studied and applied. One is that for a given control f, when the state equation $\mathcal{A}(u, f) = 0$ does not have a solution or it has infinitely many solutions, it is not always clear that it can be used. For example, the Navier-Stokes equation is only guaranteed to have a unique solution if μ is large (Reynolds number is small) relative to control f. The other is that the reduced approach leads to perform large-scale matrix operations. For higher-dimensional problems, this may increase the computational cost. The effectiveness of the reduced approach needs to be studied. We hope that in the near future, we will be able to find answers to these questions.

References

[1] R. A. Adams. Sobolev spaces. Academic Press, New York, 1978.
[2] R. Agarwal, S.-P. Shue, and Y. Tupe. Optimal and robust control of Burgers equation and heat equation with random forcing, pages 1–13. 1996.
[3] E. J. Allen, S. J. Novosel, and Z. Zhang. Finite element and difference approximation of some linear stochastic partial differential equations. Stochastics and stochastics reports, 64(1-2):117–142, 1998.
[4] D. P. Bertsekas. Constrained optimization and lagrange multiplier methods. 1982.
[5] N. Bressan and A. Quarteroni. An implicit/explicit spectral method for burgers' equation. Calcolo, 23(3):265–284, 1986.
[6] J. Burkardt, M. D. Gunzburger, and C. Webster. Reduced order modeling of some nonlinear stochastic partial differential equations. International journal of numerical analysis and modeling, 4(3-4):368–391, 2007.
[7] J. A. Burns and S. Kang. A control problem for burgers' equation with bounded input/output. Nonlinear dynamics, 2(4):235–262, 1991.
[8] R. H. Byrd, S. L. Hansen, J. Nocedal, and Y. Singer. A stochastic quasi-newton method for large-scale optimization. SIAM journal on optimization, 26(2):1008–1031, 2016.
[9] R. H. Byrd and J. Nocedal. A tool for the analysis of quasi-newton methods with application to unconstrained minimization. SIAM journal on numerical analysis, 26(3):727–739, 1989.
[10] Y. Cao, M. Y. Hussaini, and T. A. Zang. An efficient monte carlo method for optimal control problems with uncertainty. Computational optimization and applications, 26(3):219–230, 2003.
[11] C. Cartis, N. I. M. Gould, and P. L. Toint. On the complexity of steepest descent, newton's and regularized newton's methods for nonconvex unconstrained optimization problems. SIAM journal on optimization, 20(6):2833–2852, 2010.
[12] H. Choi, R. Temam, P. Moin, and J. Kim. Feedback control for unsteady flow and its application to the stochastic burgers equation. Journal of fluid mechanics, 253(1):509–543, 1993.
[13] G. Da Prato and A. Debussche. Control of the stochastic burgers model of turbulence. SIAM journal on control and optimization, 37(4):1123–1149, 1999.
[14] J. E. Dennis, M. Heinkenschloss, and L. N. Vicente. Trust-region interior-point sqp algorithms for a class of nonlinear programming problems. SIAM journal on scientific computing, 36(5):1750–1794, 1998.
[15] Q. Du and T. Zhang. Numerical approximation of some linear stochastic partial differential equations driven by special additive noises. SIAM journal on numerical analysis, 40(4):1421–1445, 2002.

[16] A. V. Fursikov. Optimal control of distributed systems; theory and applications. Scitech Book News, 2000.
[17] M. J. Grote, J. Huber, D. Kourounis, and O. Schenk. Inexact interior-point method for pde-constrained nonlinear optimization. SIAM Journal on scientific computing, 2015.
[18] M. D. Gunzburger. Finite element methods for viscous incompressible flows: A guide to theory, practice, and algorithms. Academic Press, Boston, 1989.
[19] M. D. Gunzburger and L. S. Hou. Finite-dimensional approximation of a class of constrained nonlinear optimal control problems. SIAM journal on control and optimization, 34(3):1001–1043, 1996.
[20] M. D. Gunzburger and H. C. Lee. Analysis, approximation, and computation of a coupled solid fluid temperature control problem. Computer methods in applied mechanics and engineering, 118(1-2):133–152, 1994.
[21] M. D. Gunzburger and J. Ming. Optimal control of stochastic flow over a backward-facing step using reduced-order modeling. SIAM journal on scientific computing, 33(5):2641–2663, 2011.
[22] M. D. Gunzburger, C. G. Webster, and G. Zhang. Stochastic finite element methods for partial differential equations with random input data. Acta numerica, 26:521–650, 2014.
[23] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. Optimization with PDE constraints, mathematical modeling: Theory and applications. Springer-Verlag, New York, 2009.
[24] L. Hou and Y. Yan. Dynamics and approximations of a velocity tracking problem for the navier-stokes flows with piecewise distributed controls. SIAM journal on control and optimization, 35(6):1847–1885, 1997.
[25] A. A. Jeanne, T. B. Jeffrey, and B. K. Belinda. Reduced order controllers for burgers' equation with a nonlinear observer. International journal of applied mathematics and computer science, 11:1311–1330, 2001.
[26] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. Van Bloemen Waanders. A trust-region algorithm with adaptive stochastic collocation for pde optimization under uncertainty. SIAM journal on scientific computing, 35(4):A1847–A1879, 2013.
[27] K. Kunisch and S. Volkwein. Control of the burgers equation by a reduced-order approach using proper orthogonal decomposition. Journal of optimization theory and applications, 102(2):345–371, 1999.
[28] H.-C. Lee and J. Lee. A stochastic galerkin method for stochastic control problems. Communications in computational physics, 14(1):77–106, 2013.
[29] H.-C. Lee and Y. Nam. Distributed control of the stochastic burgers equation with random input data. East Asian journal on applied mathematics, 6(1):89–108, 2016.
[30] H.-C. Lee and G.-R. Piao. Boundary feedback control of the burgers equations by a reduced-order approach using centroidal voronoi tessellations. Journal of scientific computing, 43(3):369–387, 2010.
[31] T. v. Leeuwen and F. J. Herrmann. A penalty method for pde-constrained optimization in inverse problems. Inverse problems, 32(1):15007, 2015.
[32] D. Leykekhman. Investigation of commutative properties of discontinuous galerkin methods in pde constrained optimal control problems. Journal of scientific computation, 53(3):483–511, 2012.
[33] C.-J. Lin and J. J. Moré. Newton's method for large bound-constrained optimization problems. SIAM journal on optimization, 9(4):1100–1127, 1999.
[34] J. L. Lions and S. K. Mitter. Optimal control of systems governed by partial differential equations. Springer-Verlag, New York;Berlin;, 1971.
[35] J. Liu and Z. Wang. Non-commutative discretize-then-optimize algorithms for elliptic pde-constrained optimal control problems. SIAM journal on scientific computing, 362:596–613, 2019.
[36] X. Luo, Y. Chen, and Y. Huang. Some error estimates of finite volume element approximation for elliptic optimal control problems. International journal of numerical analysis and modeling, 10(3):697–711, 2013.
[37] X. Luo, Y. Chen, Y. Huang, and T. Hou. Some error estimates of finite volume element method for parabolic optimal control problems. Optimal control applications & methods, 35(2):145–165, 2014.
[38] J. Martínez-Frutos and F. Periago Esparza. Optimal control of PDEs under uncertainty: An introduction with application to optimal shape design of structures. Springer International Publishing AG, Cham, 2018.

[39] T. P. Mathew, M. Sarkis, and C. E. Schaerer. Analysis of block matrix preconditioners for elliptic optimal control problems. Numerical linear algebra with applications, 14(4):257–279, 2007.
[40] T. P. Mathew, M. Sarkis, and C. E. Schaerer. Analysis of block parareal preconditioners for parabolic optimal control problems. SIAM journal on scientific computing, 32(3):1180–1200, 2010.
[41] I. Neitzel, U. Prüfert, and T. Slawig. Strategies for time-dependent pde control with inequality constraints using an integrated modeling and simulation environment. Numerical algorithms, 50(3):241–269, 2009.
[42] A. Noack and A. Walther. Adjoint concepts for the optimal control of burgers equation. Computational optimization and applications, 36(1):109–133, 2007.
[43] J. Nocedal and S. J. Wright. Numerical optimization. Springer-Verlag, New York, 1998.
[44] T. Rees, H. S. Dollar, and A. J. Wathen. Optimal solvers for pde-constrained optimization. SIAM journal on scientific computing, 32(1):271–298, 2010.
[45] R. Sandilya and S. Kumar. On discontinuous finite volume approximations for semilinear parabolic optimal control problems. International journal of numerical analysis and modeling, 13(4):545–568, 2016.
[46] H. Tiesler, R. M. Kirby, D. Xiu, and T. Preusser. Stochastic collocation for optimal control problems with stochastic pde constraints. SIAM journal on control and optimization, 50(5):2659–2682, 2012.
[47] F. Troltzsch and J. Sprekels. Optimal control of partial differential equations; theory, methods, and applications. SciTech Book News, 34(3):1001–1043, 2010.
[48] S. Volkwein. Distributed control problems for the burgers equation. Computational optimization and applications, 18(2):115–140, 2001.
[49] Y. Yan. Semidiscrete galerkin approximation for a linear stochastic parabolic partial differential equation driven by an additive noise. BIT numerical mathematics, 44(4):829–847, 2004.
[50] F. Yilmaz and B. Karasozen. An all-at-once approach for the optimal control of the unsteady burgers equation. Journal of computational and applied mathematics, 259:771–779, 2014.
[51] F. Yilmaz and B. Karasözen. Solving optimal control problems for the unsteady burgers equation in comsol multiphysics. Journal of computational and applied mathematics, 235(16):4839–4850, 2011.
[52] M. Zeng and G. Zhang. Preconditioning optimal control of the unsteady burgers equations with h_1 regularized term. Applied mathematics and computation, 254:133–147, 2015.
[53] G. Zheng and J. Yin. Numerical approximation for a time optimal control problems governed by semi-linear heat equations. Advances in difference equations, 2014(1):1–7, 2014.

Department of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China 430073

E-mail: liuhong_c@hust.edu.cn and xmx2018@hust.edu.cn and jming@hust.edu.cn

URL: https://mymathorg.wordpress.com