Numer Algor (2018) 79:437–455
https://doi.org/10.1007/s11075-017-0445-3

ORIGINAL PAPER

Time-stepping discontinuous Galerkin approximation of optimal control problem governed by time fractional diffusion equation

Zhaojie Zhou[1] **· Chenyang Zhang**[1]

Received: 26 May 2017 / Accepted: 13 November 2017 / Published online: 22 November 2017

Abstract In this paper, a piecewise constant time-stepping discontinuous Galerkin method combined with a piecewise linear finite element method is applied to solve control constrained optimal control problem governed by time fractional diffusion equation. The control variable is approximated by variational discretization approach. The discrete first-order optimality condition is derived based on the *first discretize then optimize* approach. We demonstrate the commutativity of discretization and optimization for the time-stepping discontinuous Galerkin discretization. Since the state variable and the adjoint state variable in general have weak singularity near $t = 0$ and $t = T$, a time adaptive algorithm is developed based on step doubling technique, which can be used to guide the time mesh refinement. Numerical examples are given to illustrate the theoretical findings.

Keywords Time-stepping discontinuous Galerkin method · Optimal control problem · Time fractional diffusion equation · Variational discretization · Time adaptive

Mathematics Subject Classification (2010) 49J20 · 49K20 · 65N15 · 65N30

1 Introduction

In this paper, we mainly focus on developing a time-stepping discontinuous Galerkin finite element approximation of optimal control problem governed by time fractional

✉ Zhaojie Zhou
zhouzhaojie@sdnu.edu.cn

[1] School of Mathematics and Statistics, Shandong Normal University, 250014 Ji'nan, China

diffusion equation. Let Ω be a bounded domain of $R^d (1 \leq d \leq 3)$ with sufficiently smooth boundary $\partial\Omega$. Set $\Omega_T = \Omega \times (0, T)$, $\Gamma_T = \partial\Omega \times (0, T)$. We consider the following optimal control problems governed by time-fractional diffusion equation:

$$\min_{q \in U_{ad}} J(u, q) \tag{1.1}$$

subject to

$$\begin{cases} \dfrac{\partial u}{\partial t} - {}_0^R\partial_t^\beta \triangle u = q(\mathbf{x}, t), & (\mathbf{x}, t) \in \Omega_T, \\ u(\mathbf{x}, t) = 0, & (\mathbf{x}, t) \in \Gamma_T, \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{1.2}$$

Here, ${}_0^R\partial_t^\beta u$ denotes the left Riemann-Liouville fractional derivative of order β $(0 < \beta < 1)$ of the state $u(\mathbf{x}, t)$. The other details will be specified later.

Time fractional diffusion equation plays an important role in many fields, for example, the evolution of probability density function for non-Markovian process can be governed by time fractional diffusion equation. In recent years, lots of literatures are devoted to develop numerical methods and algorithms for this kinds of problem. We refer to [1–5] for finite difference methods, [6–9] for Galerkin finite element methods, [10, 11] for spectral methods, and [12, 13] for fast algorithms.

Compared to the extensive amount of work contributed to developing numerical methods and algorithms for fractional differential equations and optimal control problem governed by integer order differential equations [14–16], the research for optimal control problem governed by fractional differential equation is still immature not only in theoretical analysis but also in numerical methods and algorithms. In [17], the existence, uniqueness, and first-order optimality condition for time fractional optimal control problem with Riemann-Liouville time fractional derivative were studied. In [18], the authors discussed the time fractional optimal control problem with pointwise state constraint. In [19], spectral method using Chebyshev polynomials were used to approximate the optimal control problem governed by fractional ordinary differential equation. An inverse problem of a time fractional diffusion equation was investigated in [20], where a spectral approximation was developed. In [21], finite element method combined with $L1$-scheme was used to approximate the time fractional optimal control problem with Caputo derivative. In [22], the authors developed a fast projection gradient algorithm for space fractional optimal control problem based on finite difference discretization of the state equation.

In the present paper, a piecewise constant time-stepping discontinuous Galerkin method combined with a piecewise linear finite element method is applied to solve control constrained optimal control problem governed by time fractional diffusion equation. The control variable is approximated by variational discretization approach. The discrete first-order optimality condition is derived based on the *first discretize then optimize* approach. We demonstrate the commutativity of discretization and optimization for the time-stepping discontinuous Galerkin discretization scheme. Since the state variable and the adjoint state variable in general have weak singularity near $t = 0$ and $t = T$, a time adaptive algorithm is developed, which can be used to guide the time mesh refinement. Numerical examples are presented to verify the theoretical findings.

The paper is organized as follows. In the next section, we recall some preliminary knowledge to be used in the following sections and derive the first order optimality condition for optimal control problem. In Section 3, we construct a time-stepping discontinuous Galerkin scheme and derive the fully discrete first-order optimality condition based on *first discretize then optimize* approach. A time adaptive algorithm based on double step technique is developed in Section 4. In Section 5, we carry out numerical experiments to confirm our theoretical findings.

2 Optimal control problem

In this section, we briefly recall some knowledge about fractional integrals and fractional derivatives and then derive a continuous first-order optimality condition for the optimal control problem.

The left Caputo and Riemann-Liouville time fractional derivative of order $\beta \in (0, 1)$ are defined by

$$ {}_0^C\partial_t^\beta v = \frac{1}{\Gamma(1-\beta)} \int_0^t \frac{v'(s)}{(t-s)^\beta} ds $$

and

$$ {}_0^R\partial_t^\beta v = \frac{1}{\Gamma(1-\beta)} \frac{d}{dt} \int_0^t \frac{v(s)}{(t-s)^\beta} ds. $$

Similarly, the right Caputo and Riemann-Liouville fractional derivative of order β are given by

$$ {}_t^C\partial_T^\beta v = -\frac{1}{\Gamma(1-\beta)} \int_t^T \frac{v'(s)}{(s-t)^\beta} ds $$

and

$$ {}_t^R\partial_T^\beta v = -\frac{1}{\Gamma(1-\beta)} \frac{d}{dt} \int_t^T \frac{v(s)}{(s-t)^\beta} ds. $$

Following [23] the following relations between Caputo and Riemann-Liouville fractional derivative hold:

$$ {}_0^R\partial_t^\beta v(t) = {}_0^C\partial_t^\beta v + \frac{v(0)t^{-\beta}}{\Gamma(1-\beta)} $$

and

$$ {}_t^R\partial_T^\beta v(t) = {}_t^C\partial_T^\beta v + \frac{v(T)(T-t)^{-\beta}}{\Gamma(1-\beta)}. $$

This implies that the Riemann-Liouville fractional derivative and the Caputo fractional derivative are equal for the homogenous initial condition or terminal condition.

Consider the following time-fractional optimal control problem

$$ \min_{q \in U_{ad}} J(u, q) := \frac{1}{2} \|u(\mathbf{x}, t) - u_d(\mathbf{x}, t)\|^2_{L^2(\Omega_T)} + \frac{\gamma}{2} \|q(\mathbf{x}, t)\|^2_{L^2(\Omega_T)} \tag{2.1} $$

subject to

$$\begin{cases} \dfrac{\partial u}{\partial t} - {}_0^R\partial_t^\beta \triangle u = f + q, & (\mathbf{x}, t) \in \Omega_T, \\ u(\mathbf{x}, t) = 0, & (\mathbf{x}, t) \in \Gamma_T, \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{2.2}$$

Here, U_{ad} is the admissible set defined by

$$U_{ad} = \{q \in L^2(\Omega_T) : \ a \le q(\mathbf{x}, t) \le b \ \text{ a.e. in } \Omega_T \ \text{ with } \ a, b \in \mathrm{R} \ \text{ and } \ a \le b\}.$$

$f, u_d \in L^\infty(0, T; L^2(\Omega))$ and $u_0(\mathbf{x})$ are given functions. Since the state equation is linear and the objective functional is strictly convex, the existence and uniqueness of the solution of above control problem can be guaranteed by standard theory(see, [24]).

For above optimal control problem, we can derive the following first order optimality conditions.

Theorem 2.1 *Assume that $q \in U_{ad}$ is the solution to optimal control problem* (2.1)–(2.2) *and u is the corresponding state variable given by* (2.2). *Then there exists an adjoint state z such that (u, z, q) satisfies the following optimality conditions:*

$$\begin{cases} \dfrac{\partial u}{\partial t} - {}_0^R\partial_t^\beta \triangle u = f + q, & (\boldsymbol{x}, t) \in \Omega_T, \\ u(\boldsymbol{x}, t) = 0, & (\boldsymbol{x}, t) \in \Gamma_T, \\ u(\boldsymbol{x}, 0) = u_0(\boldsymbol{x}), & \boldsymbol{x} \in \Omega. \end{cases} \tag{2.3}$$

$$\begin{cases} -\dfrac{\partial z}{\partial t} - {}_t^C\partial_T^\beta \triangle z = u - u_d, & (\boldsymbol{x}, t) \in \Omega_T, \\ z(\boldsymbol{x}, t) = 0, & (\boldsymbol{x}, t) \in \Gamma_T, \\ z(\boldsymbol{x}, T) = 0, & \boldsymbol{x} \in \Omega. \end{cases} \tag{2.4}$$

and

$$\int_{\Omega_T} (\gamma q + z)(v - q) \ge 0, \ \forall v \in U_{ad}. \tag{2.5}$$

Proof Let $\hat{J}(q) := J(u(q), q)$ be the reduced functional over U_{ad}. Then the optimal control problem (2.1)–(2.2) can be written as the following optimization problem

$$\min_{q \in U_{ad}} \hat{J}(q). \tag{2.6}$$

For above problem, the first-order optimality condition reads

$$\hat{J}'(q)(v - q) \ge 0, \forall v \in U_{ad}. \tag{2.7}$$

By a simple calculation, we have

$$\hat{J}'(q)(v - q) = \int_{\Omega_T} \gamma q(v - q) + \int_{\Omega_T} u'(q)(v - q)(u - u_d) \ge 0, \ \forall v \in U_{ad}, \tag{2.8}$$

where $u'(q)(v-q)$ denotes the the Frechét derivative in the direction $v-q$. Let $\tilde{u} = u'(q)(v-q)$, which satisfies

$$\begin{cases} \tilde{u}_t - {}_0^R\partial_t^\beta \triangle \tilde{u} = v - q, & (\mathbf{x}, t) \in \Omega_T, \\ \tilde{u}(\mathbf{x}, t) = 0, & (\mathbf{x}, t) \in \Gamma_T, \\ \tilde{u}(\mathbf{x}, 0) = 0, & \mathbf{x} \in \Omega. \end{cases}$$

Then (2.8) reduces to

$$\int_{\Omega_T} \gamma q(v-q) + \int_{\Omega_T} \tilde{u}(u - u_d) \geq 0, \forall v \in U_{ad}. \tag{2.9}$$

To further simplify above condition, we introduce the following adjoint state equation

$$\begin{cases} -\frac{\partial z}{\partial t} - {}_t^C\partial_T^\beta \triangle z = u - u_d, & (\mathbf{x}, t) \in \Omega_T, \\ z(\mathbf{x}, t) = 0, & (\mathbf{x}, t) \in \Gamma_T, \\ z(\mathbf{x}, T) = 0, & \mathbf{x} \in \Omega. \end{cases}$$

It is obvious that for the adjoint state z the Riemann-Liouville derivative is equivalent to the Caputo derivative, since $z(\mathbf{x}, T) = 0$. Multiplying the adjoint state equation by $\tilde{u}$ and integrating over Ω_T yields

$$\int_{\Omega_T} (-\frac{\partial z}{\partial t} - {}_t^C\partial_T^\beta \triangle z)\tilde{u} = \int_{\Omega_T} \tilde{u}(u - u_d).$$

By the integration formula by parts, we deduce

$$-\int_{\Omega_T} \tilde{u} \cdot \frac{\partial z}{\partial t} = \int_{\Omega_T} z \cdot \tilde{u}_t.$$

By the Green formula, the definitions of fractional derivative and the integration formula by parts (see, [23]) we can derive

$$\begin{aligned} -\int_{\Omega_T} {}_t^C\partial_T^\beta \triangle z \cdot \tilde{u} &= -\int_{\Omega_T} \triangle \tilde{u}(\mathbf{x}, t) \cdot {}_t^R\partial_T^\beta z(\mathbf{x}, t) \\ &= \frac{1}{\Gamma(1-\beta)} \int_\Omega \int_0^T \triangle \tilde{u}(\mathbf{x}, t) \frac{d}{dt} \int_t^T \frac{z(\mathbf{x}, s)}{(s-t)^\beta} ds dt \\ &= -\frac{1}{\Gamma(1-\beta)} \int_\Omega \int_0^T \frac{d}{dt} \triangle \tilde{u}(\mathbf{x}, t) \int_t^T \frac{z(\mathbf{x}, s)}{(s-t)^\beta} ds dt \\ &= -\frac{1}{\Gamma(1-\beta)} \int_\Omega \int_0^T z(\mathbf{x}, t) \int_0^t \frac{d}{ds} \triangle \tilde{u}(\mathbf{x}, s) \frac{1}{(t-s)^\beta} ds dt \\ &= -\int_{\Omega_T} z(\mathbf{x}, t) \cdot {}_0^C\partial_t^\beta \triangle \tilde{u}(\mathbf{x}, t). \end{aligned}$$

Therefore, we have

$$\int_{\Omega_T} \tilde{u}(u - u_d) = \int_{\Omega_T} z(v-q). \tag{2.10}$$

Combining (2.9) and (2.10) yields (2.5). □

Let

$$P_{U_{ad}}(q(\mathbf{x},t)) = \max\{a, \min(q(\mathbf{x},t), b)\}$$

denote the pointwise projection onto the admissible set U_{ad}. Then (2.5) is equivalent to

$$q = P_{U_{ad}}(-\frac{1}{\gamma}z).$$

3 Time-stepping discontinuous Galerkin approximation

In this section, we will investigate the time-stepping discontinuous Galerkin approximation of optimal control problem (2.1)–(2.2). For the spatial discretization, we employ the continuous and piecewise linear finite element space, while a piecewise constant, discontinuous Galerkin method is used to approximate the time variable.

3.1 Time-stepping discontinuous Galerkin discrete scheme for the state equation

Let V_h be the finite element space consisting of continuous piecewise linear functions over the triangulation T_h:

$$V_h = \{v_h \in H_0^1(\Omega) \cap C(\Omega);\ v_h \text{ is a linear function over } K, \forall K \in T_h\}.$$

To define the fully discrete scheme, we introduce a time partition. Let $\Delta_\tau : 0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$ be a time grid with $\tau_n = t_n - t_{n-1},\ n = 1, 2, \cdots, N$ and $\tau = \max\limits_n \tau_n$. Let $I_n = (t_{n-1}, t_n]$ be a half-open interval. Define the fully discrete finite element space

$$V_{hk} = \{\phi : \Omega \times [0,T] \to R;\ \phi(\mathbf{x},\cdot) \in V_h,\ \phi(\cdot,t)|_{I_n} \in P_0, n = 1, 2, \cdots, N\}.$$

This implies that ϕ is a piecewise constant with respect to time. Thus for $\phi \in V_{hk}$, we have

$$\phi(\cdot,t) = \phi^n \in V_h,\quad t \in I_n.$$

In particular $\phi(t_n) = \phi^n$. We write $\phi_+^n = \phi(t_n^+) = \lim\limits_{t \to t_n^+} \phi(t)$, and $[\phi]^n = \phi_+^n - \phi^n$. Since ϕ is a constant on I_{n+1}, we have $[\phi]^n = \phi^{n+1} - \phi^n$.

Then the time stepping discontinuous Galerkin discrete scheme for the state equation reads:

$$\int_0^T (U', X) + \sum_{n=1}^{N-1} ([U^n], X_+^n) + \int_0^T ({}_0^R\partial_t^\beta \nabla U, \nabla X) + (U_+^0, X_+^0)$$
$$= \int_0^T (f + q, X) + (\tilde{u}_0, X_+^0), \forall X \in V_{hk}. \tag{3.1}$$

Here, $\tilde{u}_0$ is an approximation of $u_0(x)$. Note that U is a piecewise constant in time. Then we have $U' \equiv 0$ and $U(t) = U^n = U_+^{n-1}$ in I_n. Then by choosing X to vanish outside I_n above discrete scheme reduces to

$$(U^n, v_h)+\int_{I_n}({}_0^R\partial_t^\beta \nabla U, \nabla v_h) = (U^{n-1}, v_h)+\int_{I_n}(f+q, v_h), \forall v_h \in V_h, n = 1, 2, \cdots, N. \quad (3.2)$$

By the definition of fractional derivative, we have

$$\begin{aligned}\int_{I_n} {}_0^R\partial_t^\beta U(t)dt &= \frac{1}{\Gamma(1-\beta)}\int_0^{t_n}\frac{U(s)}{(t_n-s)^\beta} - \frac{1}{\Gamma(1-\beta)}\int_0^{t_{n-1}}\frac{U(s)}{(t_{n-1}-s)^\beta} \\ &:= ({}_0I_t^\beta U)(t_n) - ({}_0I_t^\beta U)(t_{n-1}).\end{aligned}$$

Since U is a piecewise constant in time, then we can derive

$$\begin{aligned}({}_0I_t^\beta U)(t_n) &= \frac{1}{\Gamma(1-\beta)}\sum_{k=1}^{n}\int_{t_{k-1}}^{t_k}\frac{U(s)}{(t_n-s)^\beta} \\ &= \frac{1}{\Gamma(1-\beta)}\Big(U^1\cdot\int_{t_0}^{t_1}\frac{1}{(t_n-s)^\beta} + U^2\cdot\int_{t_1}^{t_2}\frac{1}{(t_n-s)^\beta} \\ &\quad +\cdots+U^n\cdot\int_{t_{n-1}}^{t_n}\frac{1}{(t_n-s)^\beta}\Big).\end{aligned}$$

Thus we have

$$\begin{aligned}({}_0I_t^\beta U)(t_n)-({}_0I_t^\beta U)(t_{n-1}) &= \frac{1}{\Gamma(1-\beta)}\sum_{k=1}^{n-1}\Big(\int_{t_{k-1}}^{t_k}\frac{1}{(t_n-s)^\beta}-\int_{t_{k-1}}^{t_k}\frac{1}{(t_{n-1}-s)^\beta}\Big)U^k \\ &\quad +\frac{U^n}{\Gamma(1-\beta)}\cdot\int_{t_{n-1}}^{t_n}\frac{1}{(t_n-s)^\beta}.\end{aligned}$$

For simplicity, we define

$$C_{n,k} = \begin{cases}\dfrac{1}{\Gamma(1-\beta)}\displaystyle\int_{t_{k-1}}^{t_k}\Big(\frac{1}{(t_n-s)^\beta}-\frac{1}{(t_{n-1}-s)^\beta}\Big), & k = 1, 2, \cdots, n-1, \\ \dfrac{1}{\Gamma(1-\beta)}\displaystyle\int_{t_{n-1}}^{t_n}\frac{1}{(t_n-s)^\beta}, & k = n.\end{cases}$$

By some simple calculations, we further derive

$$C_{n,k} = \begin{cases}\dfrac{1}{\Gamma(2-\beta)}\Big((t_n-t_{k-1})^{1-\beta}-(t_n-t_k)^{1-\beta}-(t_{n-1}-t_{k-1})^{1-\beta} \\ \qquad +(t_{n-1}-t_k)^{1-\beta}\Big), \quad k = 1, 2, \cdots, n-1, \\ \dfrac{\tau_n^{1-\beta}}{\Gamma(2-\beta)}, \quad k = n.\end{cases}$$

Then we have

$$({}_0I_t^\beta U)(t_n) - ({}_0I_t^\beta U)(t_{n-1}) := \sum_{k=1}^{n} C_{n,k}U^k.$$

Using above notations, the time stepping discontinuous discrete scheme for the state equation can be characterized by

$$(U^n, v_h)+\sum_{k=1}^{n} C_{n,k}(\nabla U^k, \nabla v_h)=(U^{n-1}, v_h)+\int_{I_n}(f+q, v_h), \forall v_h \in V_h, n=1,2,\cdots,N. \quad (3.3)$$

Further, we can rewrite above discrete scheme as a modified backward Euler scheme:

$$\left(\frac{U^n-U^{n-1}}{\tau_n}, v_h\right)+\frac{1}{\tau_n}\sum_{k=1}^{n} C_{n,k}(\nabla U^k,\nabla v_h)=\frac{1}{\tau_n}\int_{I_n}(f+q, v_h), \ \forall v_h\in V_h. \quad (3.4)$$

3.2 Time-stepping discontinuous Galerkin scheme for the control problem

In this section, we will discuss the time-stepping discontinuous Galerkin approximation of optimal control problem (2.1)–(2.2).

Firstly, we discretize the cost functional by the right rectangular rule:

$$J_{h,\tau}(U^n, Q^n) := \frac{1}{2}\sum_{n=1}^{N}\tau_n\Big(\| U^n-\bar{u}_d^n \|^2_{L^2(\Omega)} +\gamma \| Q^n \|^2_{L^2(\Omega)} \Big). \quad (3.5)$$

Here, $\bar{u}_d^n = \frac{1}{\tau_n}\int_{I_n} u_d dt$.

Collecting (3.4) and (3.5) gives the time-stepping discontinuous Galerkin discrete scheme for the control problem (2.1)–(2.2): finding $(U^n, Q^n) \in V_h \times U_{ad}$ such that

$$\min_{U^n\in V_h, Q^n\in U_{ad}} J_{h,\tau}(U^n, Q^n) \quad (3.6)$$

subject to

$$\begin{cases} (\frac{U^n-U^{n-1}}{\tau_n}, v_h)+\frac{1}{\tau_n}\sum_{k=1}^{n} C_{n,k}(\nabla U^k, \nabla v_h) = (\bar{f}^n+Q^n, v_h), \\ U^0=\tilde{u}_0. \end{cases} \quad (3.7)$$

Here, $\bar{f}^n = \frac{1}{\tau_n}\int_{I_n} f dt$ and the control variable was implicitly discretized by variational discretization concept. In general Q^n is not a finite element function.

To obtain the discrete first-order optimality conditions of optimal control problem (3.6)–(3.7), we define a Lagrange functional as

$$\begin{aligned}\mathcal{L}(\vec{U}, \vec{Z}, \vec{Q}) := &\sum_{n=1}^{N}\tau_n\Bigg((\bar{f}^n+Q^n, Z^{n-1})-\frac{1}{\tau_n}\sum_{k=1}^{n}C_{n,k}(\nabla U^k, \nabla Z^{n-1}) \\ &- \left(\frac{U^n-U^{n-1}}{\tau_n}, Z^{n-1}\right)\Bigg)+(U^0, Z^0)+J_{h,\tau}(U^n, Q^n).\end{aligned}$$

Here, $\vec{U} = (U^1, U^2, \cdots, U^N)^T$, $\vec{Z} = (Z^0, Z^1, \cdots, Z^{N-1})^T$ and $\vec{Q} = (Q^1, Q^2, \cdots, Q^N)^T$. The discrete first order optimality condition can be deduced by computing

$$\frac{\partial \mathcal{L}(\vec{U}, \vec{Z}, \vec{Q})}{\partial Z^{n-1}}(v_h) = 0, \quad \frac{\partial \mathcal{L}(\vec{U}, \vec{Z}, \vec{Q})}{\partial U^n}(v_h) = 0, \quad \frac{\partial \mathcal{L}(\vec{U}, \vec{Z}, \vec{Q})}{\partial Q^n}(w_h) \geq 0, \ n = 1, 2, \cdots, N.$$

Now, we derive the discrete scheme for the adjoint state equation. Firstly, by rearranging the terms, we can derive

$$\begin{aligned}
&\sum_{n=1}^{N} \tau_n \left(\frac{U^n - U^{n-1}}{\tau_n}, Z^{n-1} \right) + (U^0, Z^0) \\
&= (U^0, Z^0) + (U^1 - U^0, Z^0) + (U^2 - U^1, Z^1) + \cdots + (U^N - U^{N-1}, Z^{N-1}) \\
&= (Z^0 - Z^1, U^1) + (Z^1 - Z^2, U^2) + \cdots + (Z^{N-1} - Z^N, U^N) + (U^N, Z^N).
\end{aligned}$$

By setting $Z^N = 0$, we have

$$\sum_{n=1}^{N} \tau_n \left(\frac{U^n - U^{n-1}}{\tau_n}, Z^{n-1} \right) + (U^0, Z^0) = \sum_{n=1}^{N} \tau_n \left(\frac{Z^{n-1} - Z^n}{\tau_n}, U^n \right).$$

Secondly, note that

$$\begin{aligned}
&\sum_{n=1}^{N} \sum_{k=1}^{n} C_{n,k} (\nabla U^k, \nabla Z^{n-1}) \\
&= (C_{1,1} \nabla U^1, \nabla Z^0) \\
&\quad + (C_{2,1} \nabla U^1, \nabla Z^1) + (C_{2,2} \nabla U^2, \nabla Z^1) \\
&\quad + (C_{3,1} \nabla U^1, \nabla Z^2) + (C_{3,2} \nabla U^2, \nabla Z^2) + (C_{3,3} \nabla U^3, \nabla Z^2) \\
&\quad \cdots \\
&\quad + (C_{N-1,1} \nabla U^1, \nabla Z^{N-2}) + (C_{N-1,2} \nabla U^2, \nabla Z^{N-2}) + \cdots + (C_{N-1,N-1} \nabla U^{N-1}, \nabla Z^{N-2}) \\
&\quad + (C_{N,1} \nabla U^1, \nabla Z^{N-1}) + (C_{N,2} \nabla U^2, \nabla Z^{N-1}) + \cdots + (C_{N,N} \nabla U^N, \nabla Z^{N-1}) \\
&= (C_{1,1} \nabla Z^0, \nabla U^1) + (C_{2,1} \nabla Z^1, \nabla U^1) + \cdots + (C_{N-1,1} \nabla Z^{N-2}, \nabla U^1) + (C_{N,1} \nabla Z^{N-1}, \nabla U^1) \\
&\quad + (C_{2,2} \nabla Z^1, \nabla U^2) + (C_{3,2} \nabla Z^2, \nabla U^2) + \cdots + (C_{N,2} \nabla Z^{N-1}, \nabla U^2) \\
&\quad \cdots \\
&\quad + (C_{N,N} \nabla Z^{N-1}, \nabla U^N).
\end{aligned}$$

Then we have

$$\sum_{n=1}^{N} \sum_{k=1}^{n} C_{n,k} (\nabla U^k, \nabla Z^{n-1}) = \sum_{n=1}^{N} \sum_{k=n}^{N} C_{k,n} (\nabla Z^{k-1}, \nabla U^n).$$

From

$$\frac{\partial \mathcal{L}(\vec{U}, \vec{Z}, \vec{Q})}{\partial U^n}(v_h) = 0,$$

we obtain the discrete scheme for the adjoint state:

$$\begin{cases} \left(\dfrac{Z^{n-1}-Z^n}{\tau_n}, v_h\right) + \dfrac{1}{\tau_n}\displaystyle\sum_{k=n}^{N} C_{k,n}(\nabla Z^{k-1}, \nabla v_h) = (U^n - \bar{u}_d^n, v_h), \\ Z^N = 0. \end{cases} \tag{3.8}$$

By

$$\frac{\partial \mathcal{L}(\vec{U}, \vec{Z}, \vec{Q})}{\partial Q^n}(w_h) \geq 0$$

we obtain

$$(\gamma Q^n + Z^{n-1}, w_h - Q^n) \geq 0, \quad \forall w_h \in U_{ad}.$$

Therefore, the discrete optimality conditions are given by:

$$\begin{cases} \left(\dfrac{U^n - U^{n-1}}{\tau_n}, v_h\right) + \dfrac{1}{\tau_n}\displaystyle\sum_{k=1}^{n} C_{n,k}(\nabla U^k, \nabla v_h) = (\bar{f}^n + Q^n, v_h), \ n = 1, 2, \cdots, N, \\ \left(\dfrac{Z^{n-1}-Z^n}{\tau_n}, v_h\right) + \dfrac{1}{\tau_n}\displaystyle\sum_{k=n}^{N} C_{k,n}(\nabla Z^{k-1}, \nabla v_h) = (U^n - \bar{u}_d^n, v_h), \ n = N, N-1, \cdots, 1, \\ (\gamma Q^n + Z^{n-1}, w_h - Q^n) \geq 0, \quad \forall w_h \in U_{ad}, \ n = 1, 2, \cdots, N, \\ U^0 = \tilde{u}_0, \ \ Z^N = 0. \end{cases} \tag{3.9}$$

By the projection operator $P_{U_{ad}}$, the discrete control variable Q^n can be expressed as follows

$$Q^n = P_{U_{ad}}\left(-\frac{1}{\gamma} Z^{n-1}\right).$$

Above discrete first-order optimality condition is derived based on the *first discretize then optimize* approach. In the following, we shall show that discretization and optimization are commutative for the time-stepping discontinuous Galerkin method. This means that the *first discretize then optimize* approach and *first optimize then discretize* approach lead to the same discrete optimality system.

For this purpose, we just need to check the discretization of fractional derivative term in the adjoint state equation. By the definition of the right fractional integral and derivative we derive

$$\begin{aligned} \int_{I_n} {}_t^R\partial_T^\beta Z(t)dt \ &= \ -\frac{1}{\Gamma(1-\beta)} \int_{t_n}^{T} \frac{Z(s)}{(s-t_n)^\beta} + \frac{1}{\Gamma(1-\beta)} \int_{t_{n-1}}^{T} \frac{Z(s)}{(s-t_{n-1})^\beta} \\ &:= \ ({}_tI_T^\beta Z)(t_n) - ({}_tI_T^\beta Z)(t_{n-1}). \end{aligned}$$

Similar to the state variable, we deduce

$$
\begin{aligned}
&({}_tI_T^{\beta}Z)(t_n)\\
&= -\frac{1}{\Gamma(1-\beta)}\sum_{k=n+1}^{N}\int_{t_{k-1}}^{t_k}\frac{Z(s)}{(s-t_n)^{\beta}}\\
&= -\frac{1}{\Gamma(1-\beta)}\left(Z^n\cdot\int_{t_n}^{t_{n+1}}\frac{1}{(s-t_n)^{\beta}}+Z^{n+1}\cdot\int_{t_{n+1}}^{t_{n+2}}\frac{1}{(s-t_n)^{\beta}}\right.\\
&\left.+\ \cdots+Z^{N-1}\cdot\int_{t_{N-1}}^{t_N}\frac{1}{(s-t_n)^{\beta}}\right)
\end{aligned}
$$

Then we have

$$
\begin{aligned}
({}_tI_T^{\beta}Z)(t_n)-({}_tI_T^{\beta}Z)(t_{n-1}) &= \frac{1}{\Gamma(1-\beta)}\sum_{k=n+1}^{N}\left(\int_{t_{k-1}}^{t_k}\frac{1}{(s-t_{n-1})^{\beta}}-\int_{t_{k-1}}^{t_k}\frac{1}{(s-t_n)^{\beta}}\right)Z^{k-1}\\
&\quad+\frac{Z^{n-1}}{\Gamma(1-\beta)}\cdot\int_{t_{n-1}}^{t_n}\frac{1}{(s-t_{n-1})^{\beta}}.
\end{aligned}
$$

For simplicity, we set

$$
\widetilde{C}_{n,k}=\begin{cases}\dfrac{1}{\Gamma(1-\beta)}\displaystyle\int_{t_{k-1}}^{t_k}\left(\frac{1}{(s-t_{n-1})^{\beta}}-\frac{1}{(s-t_n)^{\beta}}\right), & k=n+1,n+2,\cdots,N,\\[2ex] \dfrac{1}{\Gamma(1-\beta)}\displaystyle\int_{t_{n-1}}^{t_n}\frac{1}{(s-t_{n-1})^{\beta}}=\frac{\tau_n^{1-\beta}}{\Gamma(2-\beta)}, & k=n.\end{cases}
$$

Then we obtain

$$
({}_tI_T^{\beta}Z)(t_n)-({}_tI_T^{\beta}Z)(t_{n-1}):=\sum_{k=n}^{N}\widetilde{C}_{n,k}Z^{k-1}. \tag{3.10}
$$

Now it remains to check the relationship of $\widetilde{C}_{n,k}$ and $C_{n,k}$. Indeed, using integration by substitution, we can easily prove that $\widetilde{C}_{n,k}=C_{k,n}$ for $k=n+1,n+2,\cdots,N$. This property guarantees the commutativity of optimization and discretization.

Finally, we shall derive the stability estimate for above discrete scheme.

Lemma 3.1 *Let (U^n, Z^{n-1}, Q^n) be the solution of (3.9). Then the following stability estimates hold*

$$
\|U^n\|\le\|\tilde{u}_0\|+\sum_{n=1}^{N}\tau_n\|\bar{f}^n+Q^n\|
$$

and

$$
\|Z^{n-1}\|\le\sum_{n=1}^{N}\tau_n\|U^n-\bar{u}_d^n\|.
$$

Proof Taking $v_h = U^n$ in the discrete state equation gives

$$(U^n - U^{n-1}, U^n) + \sum_{k=1}^{n} C_{n,k}(\nabla U^k, \nabla U^n) = \tau_n(\bar{f}^n + Q^n, U^n).$$

It is easy to check that

$$(U^n - U^{n-1}, U^n) \geq \frac{1}{2}\|U^n\|^2 - \frac{1}{2}\|U^{n-1}\|^2$$

Summing from $n = 1$ to m leads to

$$\frac{1}{2}\|U^m\|^2 - \frac{1}{2}\|U^0\|^2 + \sum_{n=1}^{m}\sum_{k=1}^{n} C_{n,k}(\nabla U^k, \nabla U^n) \leq \sum_{n=1}^{m} \tau_n\|\bar{f}^n + Q^n\| \cdot \|U^n\|.$$

According to [8, 12], we have

$$\begin{aligned}\sum_{n=1}^{m}\sum_{k=1}^{n} C_{n,k}(\nabla U^k, \nabla U^n) &= \sum_{n=1}^{m}(\int_{I_n} {}_0^R\partial_t^\beta \nabla U(t)dt, \nabla U^n)\\ &= \int_0^{t_m} ({}_0^R\partial_t^\beta \nabla U(t), \nabla U(t))dt \geq 0.\end{aligned}$$

Then we obtain

$$\|U^m\|^2 \leq \|U^0\|^2 + \sum_{n=1}^{m} \tau_n\|\bar{f}^n + Q^n\| \cdot \|U^n\|.$$

Let $\|U^*\| = \max_{0\leq n\leq N} \|U^n\|$. We further derive

$$\|U^*\|^2 \leq (\|U^0\| + \sum_{n=1}^{N} \tau_n\|\bar{f}^n + Q^n\|) \cdot \|U^*\|. \tag{3.11}$$

The stability estimate for the adjoint state can proved in an analogue way. Here, we just skip it. □

4 Time apaptive algorithm

As is shown in many references(see, for example, [25]), the solution of time fractional diffusion equation in general has weak singularity near the initial value $t = 0$, which reduces the convergence rate of numerical methods on uniform mesh. There are some references developing numerical methods on graded mesh or nonuniform mesh to improve the behavior of numerical solutions, for example, [4, 8, 9]. In the optimal control problem governed by time fractional diffusion equation, the state variable and adjoint state variable have weak singularity near $t = 0$ and $t = T$, respectively, which make the problem become more complicated and more difficult to construct graded mesh and carry out corresponding numerical analysis.

In order to improve the quality of numerical solution near $t = 0$ and $t = T$, we resort to the time adaptive algorithm, where the mesh size of time partition is chosen according to the behavior of the solution.

As we know the key problem of adaptive algorithm is the error estimator. Based on step-doubling technique [13, 26], we define the error indicators for the state and adjoint state variables:

$$\mathcal{E}_u^n := \Big(\int_\Omega (U^n - \widehat{U}^n)^2 dx \Big)^{\frac{1}{2}}$$

and

$$\mathcal{E}_z^n := \Big(\int_\Omega (Z^{n-1} - \widehat{Z}^{n-1})^2 dx \Big)^{\frac{1}{2}}.$$

Here, $U^n(Z^{n-1})$ and $\widehat{U}^n(\widehat{Z}^{n-1})$ are the discrete state and adjoint state solved on (3.9) with a full step τ_n and a half step $\tau_n/2$, respectively.

We define the error estimator $\mathcal{E}_{all}^n = \mathcal{E}_u^n + \mathcal{E}_z^n$. For a given tolerance $\mathbb{O}_{time}$, we refine the time mesh until the estimator $\mathcal{E}_{all}^n$ is smaller than the tolerance $\mathbb{O}_{time}$. Then we use the last time step as the appropriate one. In the time adaptive algorithm, we use the Dörfler's (see, [27]) strategy as the marking strategy. Our time adaptive algorithm for the time fractional optimal control problem is given below.

Algorithm 1 Time step adaptive algorithm for optimal control problem

1. Choose an initial spatial triangulation $\mathscr{T}^h$ and time partition $\{I_n\}_0$, and set $m = 0$. Given a tolerance $\mathbb{O}_{time} > 0$.
2. Loop
   - (a) Solve the discrete optimal system (3.9) on the time mesh $\{I_n\}_m$ and $\{I_{\frac{n}{2}}\}_m$ to get (U^n, Z^{n-1}, Q^n) and $(\widehat{U}^n, \widehat{Z}^{n-1}, \widehat{Q}^n)$.
   - (b) Evaluate the indicator $\mathcal{E}_u^n$ and $\mathcal{E}_z^n$.
     If $\mathcal{E}_{all}^n \geq \mathbb{O}_{time}$,
     Mark the time intervals with Dörfler's strategy and refine them to generate a new time mesh $\{I_n\}_{m+1}$.
     end.
   - (c) Update m and mesh information.
3. End loop, if $\mathcal{E}_{all}^n < \mathbb{O}_{time}$.

5 Numerical experiments

In this section, numerical experiments will be carried out to illustrate the numerical scheme and algorithm presented in Sections 3 and 4. In the following example for smooth function the mean term, $\bar{f}^n$ and $\bar{u}_d^n$ are approximated by $f(t_n)$ and $u_d(t_n)$, respectively. For the function with limited smoothness, we use numerical integral formula to approximate them, for example, Gauss-Legendre formula with three points.

The discrete norm $l^2(0; T; L^2(\Omega))$ is used to measure the error $u - U$, $z - Z$ and $q - Q$, for example,

$$\|u - U\|_{l^2(0;T;L^2(\Omega))} = \Big(\sum_{n=1}^{N} \tau_n \|u(t_n) - U^n\|^2_{L^2(\Omega)} \Big)^{\frac{1}{2}}.$$

Example 5.1 In this example, we consider problem (1.1)–(1.2) with $\Omega = [0, 1]$, $\gamma = 1$ and $T = 1.0$. The exact solutions are given by

$$\begin{aligned} u &= t^{\nu} sin(2\pi x), \\ z &= (1-t)^{\nu} cos(2\pi x), \\ q &= \max\Big(-0.5, \min(-z, -0.1)\Big). \end{aligned}$$

The right-hand term f and the desired state u_d can be calculated by the exact solutions and governing equations. $\nu > 0$ is a parameter. In this example, the state variable and adjoint state variable admit typical layers at $t = 0$ and $t = T$ for small ν, respectively.

In this example, the state variable and the adjoint state variable are approximated by a time-stepping discontinuous Galerkin method combined with piecewise linear polynomial spaces, while the control variable is discretized by a variational discretization concept. This means the control variable is not explicitly approximated by finite element function. For more details about variational discretization concept, one can refer to [28].

Firstly, we consider the case $\nu = 2$. It is easy to see that the solution is smooth with respect to time variable and space variable. The purpose of this case is to show the accuracy of the time-stepping discontinuous Galerkin scheme. Figures 1 and 2 present the surface and space-time surface of the state variable, adjoint state variable

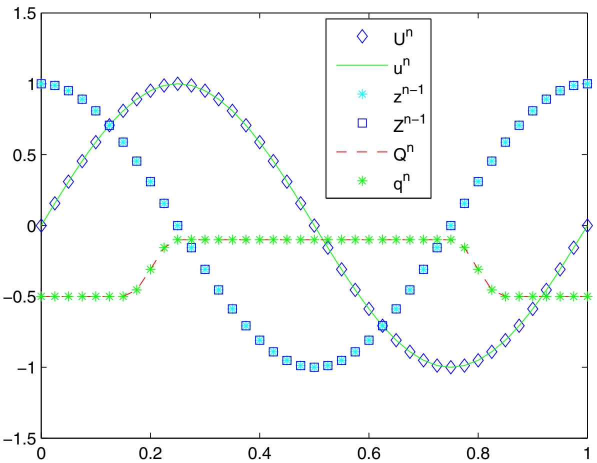

Fig. 1 The surfaces of state (at $t = T$), adjoint state(at $t = 0$) and control (at $t = T$) with $\nu = 2$

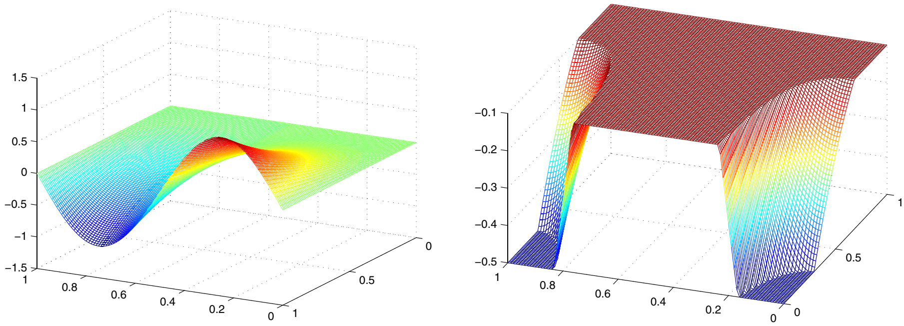

Fig. 2 The space-time surface of discrete state U and control Q with $\nu = 2$

and control variable, respectively. In order to investigate the spatial and temporal convergence rates, we need to couple τ and h as $\tau = h^2$ and $h = \tau^2$, respectively. The spatial convergence rate of the state variable and the control variable in L^2 norm are displayed in Fig. 3 for different β. We can observe that the spatial convergence rate is second order, which is optimal for piecewise linear polynomial approximation. Figure 4 presents the temporal convergence rate for different β, which implies the temporal convergence rate for the control variable is first order, while the temporal convergence rates for the state is a little better than one.

Secondly, we carry out numerical experiments for small ν, for example $\nu = 0.2, 0.6$. In this case, the state and adjoint state variable only have limited regularity with respect to time variable and admit typical boundary layer near $t = 0$ and $t = T$, respectively, which leads to degenerated time convergence rate in the error of time stepping discontinuous Galerkin approximation on uniform time grids. This is confirmed by the numerical results shown in Fig. 5, where we set $h = \tau^2$. It is easy to see that the convergence rate for the state and adjoint state are less than one, while the convergence rate for the control variable seems to be better than those of state and adjoint state.

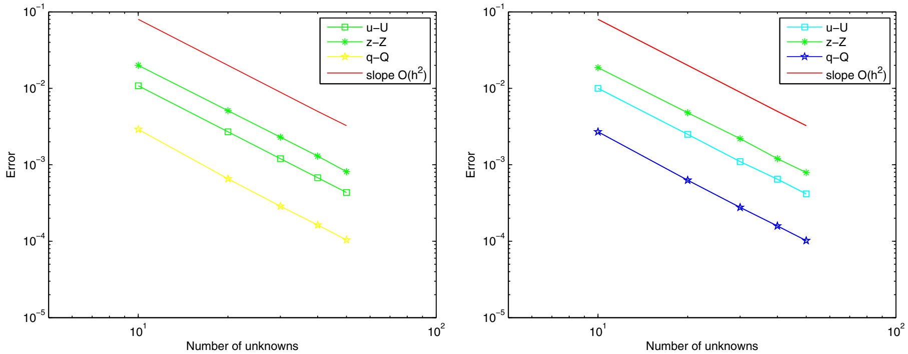

Fig. 3 The space convergence rate for state and control with $\nu = 2$. Left for $\beta = 0.4$, right for $\beta = 0.8$

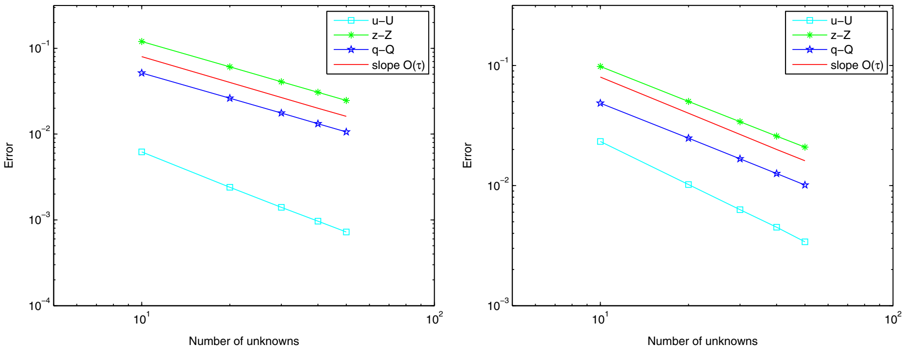

Fig. 4 The temporal convergence rate for state and control with $\nu = 2$. Left for $\beta = 0.4$, right for $\beta = 0.8$

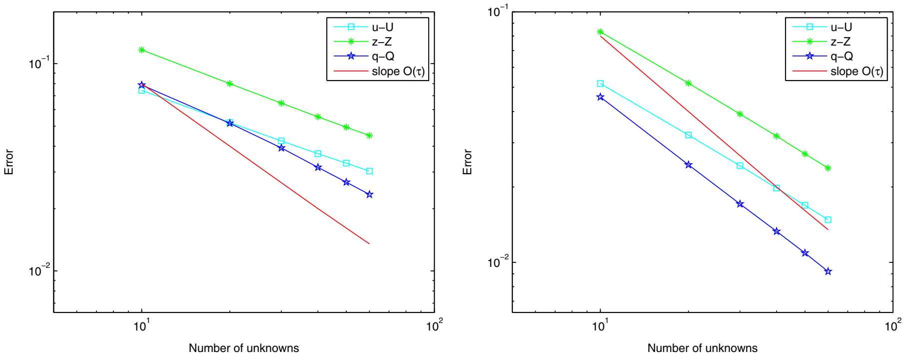

Fig. 5 The temporal convergence rate for state and control. Left for $\nu = 0.2$, $\beta = 0.4$, right for $\nu = 0.6$, $\beta = 0.8$

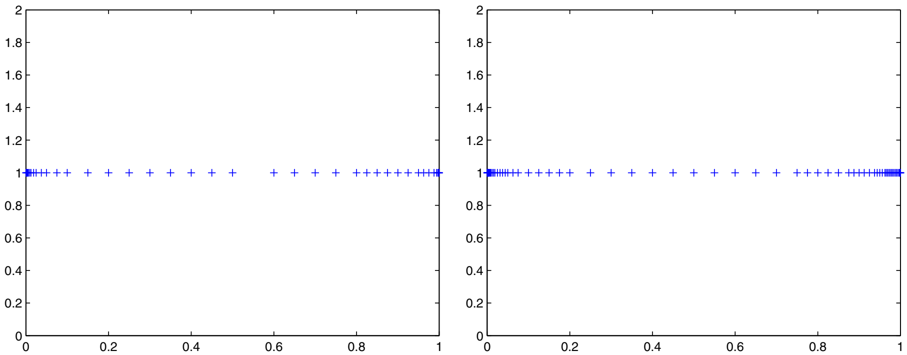

Fig. 6 The adaptive time grids for $\nu = 0.2$. Left for $N = 42$, right for $N = 153$

Table 1 The error of state variable, adjoint state variable, and control variable on adaptive time grids and uniform time grids with space dofs $M = 100$ and $\nu = 0.2$ and $\beta = 0.4$

Grids		$\|u - U\|_{l^2(0,T;L^2(\Omega))}$	$\|z - Z\|_{l^2(0,T;L^2(\Omega))}$	$\|q - Q\|_{l^2(0,T;L^2(\Omega))}$
Uniform grids	$N = 50$	0.0332	0.0494	0.0268
	$N = 100$	0.0241	0.0350	0.0159
	$N = 300$	0.0153	0.0213	0.0069
Adaptive grids	$N = 30$	0.0320	0.0526	0.0171
	$N = 56$	0.0210	0.0327	0.0102
	$N = 108$	0.0137	0.0206	0.0060

Finally, to improve the quality of numerical solution and reduce computational cost we use the time adaptive algorithm to guide the time mesh refinement. In Fig. 6, we plot the adaptive time grid for $\nu = 0.2$. The adaptive refinement near the time boundary $t = 0$ and $t = T$ demonstrates that our indicator can efficiently capture the singularities well. The errors of state variable, adjoint state variable and control variable on uniform time grids and adaptive time grids are presented in Table 1. We can observe that using adaptive time grids can greatly reduce the error and save computational cost. The convergence rates of state, adjoint state and control variable on adaptive time grids are presented in Fig. 7.

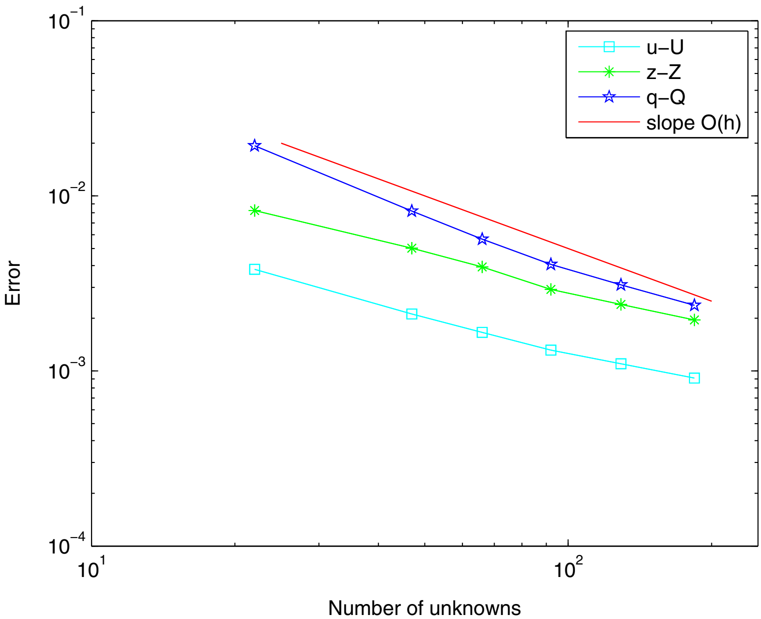

Fig. 7 The convergence rate for state, adjoint state and control variable on adaptive time grids with $M = 200$, $\nu = 0.2$ and $\beta = 0.4$

Acknowledgements The author would like to thank the support of National Natural Science Foundation of China (No. 11301311 and 11471196) and Natural Science Foundation of Shandong Province (No. ZR2016JL004).

References

1. Zeng, F., Li, C., Liu, F., Turner, I.: Numerical algorithms for time-fractional subdiffusion equation with second-order accuracy. SIAM J. Sci. Comput. **37**(1), 55–78 (2015)
2. Liu, F., Zhuang, P., Anh, V., Turner, I., Burrage, K.: Stability and convergence of the difference methods for the space-time fractional advection-diffusion equation. Appl. Math. Comput. **191**, 12–20 (2007)
3. Zhang, Y.N., Sun, Z.Z., Liao, H.L.: Finite difference methods for the time fractional diffusion equation on non-uniform meshes. J. Comput. Phys. **265**(15), 195–210 (2014)
4. Stynesy, M., O'riordan, E., Luis Gracia, J.: Error analysis of a finite difference method on graded meshes for a time fractional diffusion equation. SIAM J. Numer. Anal. **55**(2), 1057-1079 (2017)
5. Zeng, F., Li, C., Liu, F., Turner, I.: The use of finite difference/element approximations for solving the time-fractional subdiffusion equation. SIAM J. Sci. Comput. **35**(6), 2976–3000 (2013)
6. Jin, B.T., Lazarov, R., Pasciak, J., Zhou, Z.: Error analysis of semidiscrete finite element methods for inhomogeneous time-fractional diffusion. IMA J. Numer. Anal. **35**(2), 561–582 (2015)
7. Jin, B.T., Lazarov, R., Zhou, Z.: Error estimates for a semidiscrete finite element method for fractional-order parabolic equations. SIAM J. Numer. Anal. **51**(1), 445–466 (2013)
8. Mustapha, K.A., Abdallah, B.S., Furati, K.M., Nour, M.: A discontinuous Galerkin method for time fractional diffusion equations with variable coefficients. Numerical Algorithms. Accepted (2016)
9. McLean, W., Mustapha, K.: Convergence analysis of a discontinuous Galerkin method for a sub-diffusion equation. Numer. Algorithms **52**, 69–88 (2009)
10. Lin, Y., Xu, C.: Finite difference/spectral approximations for the time-fractional diffusion equation. J. Comput. Phys. **225**(2), 1533–1552 (2007)
11. Lin, Y.M., Li, X.J., Xu, C.J.: Finite difference/spectral approximations for the fractional cable equation. Math. Comp. **80**(275), 1369–1396 (2011)
12. Mclean, W.: Fast summation by interval clustering for an evolution equation with memory. SIAM J. Sci. Comput. **34**, A3039–A3056 (2012)
13. Yuste, S.antos.B., Quintana-Murillo, J.: Fast, accurate and robust adaptive finite difference methods for fractional diffusion equations. Numerical Algorithms **71**(1), 207–228 (2016)
14. Liu, W.B., Yan, N.N.: Adaptive Finite Element Methods for Optimal Control Governed by PDEs. Science Press, Beijing (2008)
15. Zhou, Z.J., Yu, X.M., Yan, N.N.: Local discontinuous galerkin approximation of convection-dominated diffusion optimal control problem with control constraints. Numer. Meth. Part. D. E. **30**, 338–360 (2014)
16. Zhou, Z.J.: A posteriori error estimates for discontinuous Galerkin approximation of nonstationary convection diffusion optimal control problems. Int. J. Comput. Math. **93**, 2106–2123 (2016)
17. Mophou, G.: Optimal control of fractional diffusion equation. Comput. Math. Appl. **61**, 68–78 (2011)
18. Mophou, G., N'Guérékata, G.M.: Optimal control of fractional diffusion equation with state constraints. Comput. Math. Appl. **62**, 1413–1426 (2011)
19. Sweilam, N.H., Al-Ajami, T.M., Hoppe, R.H.W.: Numerical solution of some types of fractional optimal control problems. Sci. World J. **2013**, 9 (2013). Article ID 306237
20. Ye, X.Y., Xu, C.J.: Spectral optimization methods for the time fractional diffusion inverse problem. Numer. Math. Theory Methods Appl. **6**(3), 499–519 (2013)
21. Zhou, Z.J., Gong, W.: Finite element approximation of optimal control problems governed by time fractional diffusion equations. Comput. Math. Appl. **71**, 301–308 (2016)
22. Du, N., Wang, H., Liu, W.B.: A fast gradient projection method for a constrained fractional optimal control. J. Sci. Comput. **68**, 1–20 (2016)
23. Podlubny, I.: Fractional Differential Equations. Academy Press, New York (1999)
24. Lions, J.L.: Optimal Control of Systems Governed by Partial Differential Equations. Springer, Berlin (1971)

25. McLean, W.: Regularity of solutions to a time-fractional diffusion equation. ANZIAM Journal **52**, 123–138 (2010)
26. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes: the Art of Scientific Computing, 3rd edn. Cambridge University Press, New York (2007)
27. Döfler, W.: A convergent adaptive algorithm for Poisson's equation. SIAM J. Numer. Anal. **33**, 1106–1124 (1996)
28. Hinze, M.: A variational discretization concept in control constrained optimization: the linear-quadratic case. Comput. Optim. Appl. **30**(1), 45–61 (2005)