ADAPTIVE FINITE ELEMENT METHOD FOR PARABOLIC EQUATIONS WITH DIRAC MEASURE

WEI GONG ◇, HUIPO LIU †, AND NINGNING YAN ‡

Abstract: In this paper we study the adaptive finite element method for parabolic equations with Dirac measure. Two kinds of problems with separate measure data in time and measure data in space are considered. It is well known that the solutions of such kind of problems may exhibit lower regularity due to the existence of the Dirac measure, and thus fit to adaptive FEM for space discretization and variable time steps for time discretization. For both cases we use piecewise linear and continuous finite elements for the space discretization and backward Euler scheme, or equivalently piecewise constant discontinuous Galerkin method, for the time discretization, the a posteriori error estimates based on energy and L^2 norms for the fully discrete problems are then derived to guide the adaptive procedure. Numerical results are provided at the end of the paper to support our theoretical findings.

Keywords: parabolic equation, Dirac measure, adaptive finite element method, space-time discretization, a posteriori error estimates.

Subject Classification: 49J20, 49K20, 65N15, 65N30.

1. Introduction

Let $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3 be an open bounded convex polygonal or polyhedral domain with Lipschitz boundary $\Gamma = \partial\Omega$ and $T > 0$ be a real number. The purpose of this paper is to consider the finite element approximations of the following parabolic equations with measure data in space

$$\begin{cases} \partial_t y + \mathcal{A}y = g(x,t)\delta_{\gamma(t)} & \text{in } \Omega_T, \\ y = 0 & \text{on } \Gamma_T, \\ y(\cdot,0) = y_0 & \text{in } \Omega \end{cases} \tag{1.1}$$

and parabolic equations with measure data in time

$$\begin{cases} \partial_t y + \mathcal{A}y = g(x,t)\delta_{t_0} & \text{in } \Omega_T, \\ y = 0 & \text{on } \Gamma_T, \\ y(\cdot,0) = y_0 & \text{in } \Omega, \end{cases} \tag{1.2}$$

where $\Omega_T = \Omega \times (0,T)$ and $\Gamma_T = \partial\Omega \times (0,T)$, $\partial_t y = \frac{\partial y}{\partial t}$, the operator $\mathcal{A}$ is a second order elliptic partial differential operator, $y_0 \in L^2(\Omega)$ is the given initial condition, g is a given function such that $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ for (1.1) and $g \in \mathcal{C}([0,T];L^2(\Omega))$ for (1.2).

In equation (1.1), we assume that $\gamma(t)$ is a lower dimensional time-continuous manifold which is strictly contained in Ω for all $t \in [0,T]$ (see [10, 22] for similar definition). $\delta_{\gamma(t)}$ denotes the Dirac measure in space on $\gamma(t)$. We note that $\gamma(t)$ can be a point, a curve or even a surface if $d = 3$, it can be static and independent of time t or evolves in the time horizon. For simplicity we assume that $\gamma(t)$ is a Lipschitz-continuous m-dimensional manifold in Ω with $0 \leqslant m \leqslant d-1$ for all $t \in [0,T]$, and the distance between $\gamma(t)$ and $\partial\Omega$ is positive for all $t \in [0,T]$. The m-dimensional

Date: August 15, 2017.

◇ NCMIS, LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. Email: wgong@lsec.cc.ac.cn.

† Institute of Applied Physics and Computational Mathematics, Beijing 100094, China. Email: liuhuipo@amss.ac.cn.

‡ NCMIS, LSEC, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. Email: ynn@amss.ac.cn.

Hausdorff measure of $\gamma(t) \subset \Omega$ in $\mathbb{R}^d$ is finite for all $t \in [0,T]$. When $m = 0$, $\gamma(t)$ will reduce to a single point or a finite number of points for each $t \in [0,T]$; when $m = 1$, $\gamma(t)$ is a $\mathcal{C}^2$-curve s.t. $\gamma(t) \subset \partial D$ for some d-dimensional $\mathcal{C}^2$-domain $D \subset\subset \Omega$ for each $t \in [0,T]$; when $m = 2$, $\gamma(t)$ is a $\mathcal{C}^2$-surface s.t. $\gamma(t) \subset \partial D$ for some 3-dimensional $\mathcal{C}^2$-domain $D \subset\subset \Omega$ for each $t \in [0,T]$. In equation (1.2) we assume that δ_{t_0} denotes the Dirac measure in time on given point $t_0 \in (0,T)$.

The problems of form (1.1) with measure data in space can be used to model the potential of an electric field with an electric charge distribution. This kind of problems also arise in other different applications, for instance, modeling of acoustic monopoles, transport equations for effluent discharge in aquatic media, and so on. Also, there are some applications in inverse problems where one attempts to identify the moving pointwise sources in the heat and convection-diffusion equations, see for example [2]. One of the most import applications of parabolic equations with measure data appears in optimal control theory. For instance, problems of form (1.1) with measure data in space can serve as the state equation of some parabolic optimal control problems with pointwise control (see [17], [21] and [24] for more details) and sparse controls (see [8]), or can be used to model air or water pollution control problems (see [31]). Moreover, parabolic optimal control problems with controls acting on a lower dimensional manifold also involve such kind of parabolic equations with Dirac measure in space where the measure data may evolve in the time horizon, we refer to [10] and [22] for more details.

On the other hand, parabolic equations of form (1.2) with measure data in time also appear in the optimality conditions of some optimal control problems with state constraints, for example, pointwise state constraints in time, and serve as the so-called adjoint state equation, we refer to [7] and [32] for more details on this kind of optimal control problems.

There have already appeared some contributions to the theoretical and numerical analysis for partial differential equations with measure data. Boccardo and Gallouët studied the existence of solutions for quasi-linear elliptic and parabolic equations involving measure data in [5], Casas studied in [7] the linear parabolic problems with measure data and improved the results of [5] by exploiting the linearity of the equation. The finite element method for elliptic equation with Dirac measure data has been extensively studied (see, e.g., [6] and the references cited therein). Casas gave an optimal error estimate of order $O(h^{2-\frac{d}{2}})$ in [6], where h is the mesh size of space triangulation and d is the dimension of Ω. Based on the similar duality argument Gong derived a priori error estimates for finite element approximations of parabolic equations with measure data in [23].

It is well known that the solutions of PDEs with measure data exhibit low regularity, thus the well developed adaptive finite element method fits to this kind of problems for the sake of accuracy enhancement. To this end, the a posteriori error estimators should be constructed to guide the adaptive procedure, which is the main purpose of this paper. We refer to [37] for an excellent review for a posteriori error estimates of different types. As for the a posteriori error estimates and adaptive finite element methods for parabolic equations, we mention the earlier work of Eriksson and Johnson in [19] and [20], and followed by Picasso in [34], Chen and Jia ([13]) and Kreuzer et al ([27]). We do not aware of much work on the a posteriori error estimates for PDEs with measure data, among them we should mention the work of Araya et al in [3], where a posteriori error estimates for elliptic problems with Dirac delta source terms are derived. In this work we intend to derive the a posteriori error estimators for parabolic equations with measure data in both space and time. To the best of our knowledge this is the first contribution on this subject in the literature. We emphasize that our problems are different from the case studied in [13], [27] and [34], where the right hand sides of equations (1.1) and (1.2) belong to $L^2(0,T;L^2(\Omega))$ and are much regular than the measure data. At last, the derived a posterior error indicators can be used to guide the space-time adaptive algorithm. In current paper we assume that the domain Ω is convex, we remark that this is not restrictive as the singularity of the solutions of parabolic equation with measure data is not caused by the reentrant corner of the domain but the singular data.

Compared with standard FEM, adaptive finite element methods are more suitable for solving problems with singularity, as it is desirable to solve a series of small scale problems other than solving a large scale problem. For time-dependent problems where the singularity may evolve in the time horizon, adaptive algorithms can provide accurate approximations by producing meshes which can capture the varying singularity through mesh refinement and coarsening, while the standard FEM usually results in a very fine mesh if we want to approximate the solution with similar accuracy. It is worthy pointing out that adaptive finite element methods pose difficulties on implementation than standard FEM, such as the mesh refinement and coarsening algorithm, the interpolation of solutions between different mesh levels and so on. However, a lot of popular finite element packages have already been well developed which make the adaptive algorithms efficient and easy to implement, we refer to, e.g., [11], [35] and [38].

The remainder of this paper is organized as follows. In Section 2 we give some notations and results concerning the parabolic equations with measure data in space and time. In Section 3 and 4 we consider parabolic equations with measure data in space and time separately, and establish the fully discrete finite element approximation based on continuous in space and discontinuous in time method and derive a posteriori error estimates between the corresponding continuous and discrete solutions. The a posteriori error estimates derived below are then used to guide the adaptive procedure in Section 5 with some numerical experiments.

2. Parabolic equations with measure data

In this paper we adopt the standard notation $W^{l,p}(\Omega)$ for Sobolev space on Ω with norm $\|\cdot\|_{l,p,\Omega}$ and seminorm $|\cdot|_{l,p,\Omega}$. We denote $W^{l,2}(\Omega)$ by $H^l(\Omega)$ and set $H_0^1(\Omega) \equiv \{v \in H^1(\Omega) : \ v|_{\partial\Omega} = 0\}$. Let $\mathcal{M}(\overline{\Omega})$ denote the space of the real and regular Borel measures on Ω, which can be defined as the dual space of $\mathcal{C}(\overline{\Omega})$ with its natural norm

$$\|\mu\|_{\mathcal{M}(\overline{\Omega})} = \sup\{\int_{\overline{\Omega}} v d\mu : \ v \in \mathcal{C}(\overline{\Omega}) \text{ and } \|v\|_{\mathcal{C}(\overline{\Omega})} \leqslant 1\},$$

and $\mathcal{M}[0,T]$ be the space of the real and regular Borel measures in $[0,T]$, which can also be defined as the dual space of $\mathcal{C}[0,T]$ with its natural norm

$$\|\mu\|_{\mathcal{M}[0,T]} = \sup\{\int_0^T v d\mu : \ v \in \mathcal{C}[0,T] \text{ and } \|v\|_{\mathcal{C}[0,T]} \leqslant 1\}.$$

Then it is clear that $\delta_{\gamma(t)} \in L^2(0,T;\mathcal{M}(\overline{\Omega}))$ and $\delta_{t_0} \in \mathcal{M}[0,T]$ (see [10, 23]).

We denote by $L^s(0,T;W^{l,p}(\Omega))$ the Banach space of all L^s integrable functions from $(0,T)$ into $W^{l,p}(\Omega)$ with norm $\|v\|_{L^s(0,T;W^{l,p}(\Omega))} = \Big(\int_0^T \|v\|_{l,p,\Omega}^s dt\Big)^{\frac{1}{s}}$ for $s \in [1,\infty)$ and the standard modification for $s = \infty$. Similarly, we define the space $H^r(0,T;W^{l,p}(\Omega))$, the details can be found in [29, Chap.4]. Let $H^{s,r}(\Omega_T) = L^2(0,T;H^s(\Omega) \cap H_0^1(\Omega)) \cap H^r(0,T;L^2(\Omega))$ equipped with the norm

$$\|w\|_{s,r} = \Big(\int_0^T \|w(\cdot,t)\|_{s,\Omega}^2 dt + \int_\Omega \|w(x,\cdot)\|_{r,[0,T]}^2 dx\Big)^{\frac{1}{2}},$$

where $\|\cdot\|_{r,[0,T]}$ denotes the norm on $H^r([0,T])$. We set

$$W(0,T) := L^2(0,T;H_0^1(\Omega)) \cap H^1(0,T;H^{-1}(\Omega)),$$

it is straightforward that $W(0,T) \hookrightarrow \mathcal{C}([0,T];L^2(\Omega))$ (see [29, Chap.1, Sec.3]). We also set

$$X(0,T) := L^2(0,T;H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(0,T;L^2(\Omega)) \hookrightarrow \mathcal{C}([0,T];H_0^1(\Omega)).$$

In addition, c or C denotes a general positive constant independent of mesh size h and time step size k.

The operator $\mathcal{A}$ is assumed to be a second order elliptic partial differential operator of the form

$$\mathcal{A}y = -\sum_{i,j=1}^{d} \partial_{x_j}(a_{ij}\partial_{x_i}y) + a_0 y,$$

where $a_0 \in L^\infty(\Omega)$, $a_0(x) \geqslant 0$ for all $x \in \Omega$, $a_{ij} \in W^{1,\infty}(\Omega)$ $(1 \leqslant i, j \leqslant d)$ and satisfies the following uniform ellipticity condition:

$$\sum_{i,j=1}^{d} a_{ij}\xi_i\xi_j \geqslant c|\xi|^2, \quad c > 0 \ \ \forall\, \xi \in \mathbb{R}^d, \ \ x \in \Omega.$$

We will denote by $\mathcal{A}^*$ the adjoint operator of $\mathcal{A}$:

$$\mathcal{A}^*y = -\sum_{i,j=1}^{d} \partial_{x_j}(a_{ji}\partial_{x_i}y) + a_0 y.$$

Thus we can define the following bilinear forms on Ω and Ω_T:

$$a(v,w) = \sum_{i,j=1}^{d} \int_\Omega (a_{ij}\partial_{x_i}v\partial_{x_j}w + a_0 vw)dx \ \ \forall\, v, w \in H^1(\Omega)$$

and

$$a(v,w)_{\Omega_T} = \sum_{i,j=1}^{d} \int_{\Omega_T} (a_{ij}\partial_{x_i}v\partial_{x_j}w + a_0 vw)dxdt \ \ \forall\, v, w \in L^2(0,T;H^1(\Omega)).$$

We denote the L^2-inner products on $L^2(\Omega)$ and $L^2(\Omega_T)$ by

$$(v,w) = \int_\Omega vwdx \ \ \forall\, v, w \in L^2(\Omega)$$

and

$$(v,w)_{\Omega_T} = \int_{\Omega_T} vwdxdt \ \ \forall\, v, w \in L^2(\Omega_T).$$

Lemma 2.1. *Assume that Ω is a convex domain. For $f \in L^2(\Omega_T)$, let ψ be the solution of following backward in time parabolic problem:*

$$\begin{cases} -\partial_t\psi + \mathcal{A}^*\psi = f & \text{in } \Omega_T, \\ \psi = 0 & \text{on } \Gamma_T, \\ \psi(T) = 0 & \text{in } \Omega. \end{cases} \tag{2.1}$$

Then there holds $\psi \in L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega)) \cap H^1(0,T;L^2(\Omega)) \hookrightarrow \mathcal{C}([0,T];H_0^1(\Omega))$ and satisfies ([29, Ch.4, Sec.6]*)*

$$\|\psi\|_{L^2(0,T;H^2(\Omega))} + \|\psi_t\|_{L^2(0,T;L^2(\Omega))} \leqslant C\|f\|_{L^2(0,T;L^2(\Omega))} \tag{2.2}$$

and

$$\|\psi(\cdot,0)\|_{H^1(\Omega)} \leqslant C\|f\|_{L^2(0,T;L^2(\Omega))}. \tag{2.3}$$

Now we are in the position to study the parabolic equations with measure data. At first we consider the parabolic equations (1.1) with measure data in space. The weak solution of problems (1.1) can be defined by transposition techniques (see Lions and Magenes [29, Ch.4, Sec.9]). The existence of a unique solution of problem (1.1) and it's regularity have been proved in [23].

Theorem 2.2. *Assume that $\gamma(t)$ is a Lipschitz-continuous m-dimensional manifold in Ω with $0 \leqslant m \leqslant d-1$ for all $t \in [0,T]$, and the distance between $\gamma(t)$ and $\partial\Omega$ is positive for all $t \in [0,T]$. With the assumption that $y_0 \in L^2(\Omega)$ and $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$, problem (1.1) admits a unique solution $y \in L^2(0,T;L^2(\Omega))$ in the sense that*

$$-(y,\partial_t v)_{\Omega_T} + (y,\mathcal{A}^*v)_{\Omega_T} = \langle g\delta_{\gamma(t)}, v\rangle_{\Omega_T} + (y_0, v(\cdot,0)) \ \ \forall\, v \in X(0,T) \tag{2.4}$$

with $v(\cdot,T)=0$, *here*

$$\langle g\delta_{\gamma(t)},v\rangle_{\Omega_T}=\int_{\overline{\Omega}_T}gvd\delta_{\gamma(t)}=\begin{cases}\int_0^T g(\gamma(t),t)v(\gamma(t),t)dt & \text{if } m=0;\\ \int_0^T\int_{\gamma(t)}g(x,t)v(x,t)dxdt & \text{if } m\geqslant 1\end{cases}$$

for all $v\in L^2(0,T;\mathcal{C}(\overline{\Omega}))$. *Besides, there exist a constant* C *only depending on* Ω *and* $\gamma(t)$, *such that*

$$\|y\|_{L^2(0,T;L^2(\Omega))}\leqslant C(\|g\|_{L^2(0,T;L^\infty(\Omega))}+\|y_0\|_{L^2(\Omega)}).$$

Moreover, we have $y\in L^2(0,T;W_0^{1,p}(\Omega))$ *and* $\partial_t y\in L^2(0,T;W^{-1,p}(\Omega))$ *such that*

$$\langle y_t,v\rangle_I+a(y,v)_{\Omega_T}=\langle g\delta_{\gamma(t)},v\rangle_{\Omega_T}\quad\forall\ v\in L^2(0,T;W^{1,q}(\Omega)),\tag{2.5}$$

where $p\in(1,\frac{d}{d-1})$ *if* $d-m>1$ *and* $p=2$ *if* $d-m=1$, q *is the conjugate number of* p *such that* $\frac{1}{p}+\frac{1}{q}=1$, $\langle\cdot,\cdot\rangle_I$ *denotes the duality pairing between* $L^2(0,T;W^{-1,p}(\Omega))$ *and* $L^2(0,T;W_0^{1,q}(\Omega))$.

Proof. Since $\delta_{\gamma(t)}\in L^2(0,T;\mathcal{M}(\overline{\Omega}))$, the existence of a unique solution in the sense of (2.4) has been proved in [23]. Moreover, in case of $d-m>1$ we have $q>d$ for $p\in(1,\frac{d}{d-1})$ and therefore $W_0^{1,q}(\Omega)$ is embedded into $\mathcal{C}(\overline{\Omega})$, it follows that $g\delta_{\gamma(t)}$ can be identified as an element of $L^2(0,T;W^{-1,p}(\Omega))$. Since $\mathcal{A}:W_0^{1,p}(\Omega)\to W^{-1,p}(\Omega)$ is an isomorphism, using the results from [18] on maximal parabolic regularity we conclude that equation (1.1) admits a unique solution $y\in L^2(0,T;W_0^{1,p}(\Omega))$ and $y_t\in L^2(0,T;W^{-1,p}(\Omega))$ for all $p\in(1,\frac{d}{d-1})$ in the sense of (2.5). Similarly, in case of $d-m=1$ we note that $g\delta_{\gamma(t)}$ can be identified as an element of $L^2(0,T;H^{-1}(\Omega))$ (see [22]) by using the trace theorem, thus we can conclude that $y\in L^2(0,T;H_0^1(\Omega))$ and $y_t\in L^2(0,T;H^{-1}(\Omega))$. This completes the proof. □

In the following, we give several typical examples of parabolic equations with measure data in space appeared in the literature.

Example 2.3.
(1) *In the parabolic optimal control problems with pointwise control (see* [24]*), the measure data is a Dirac measure concentrated on one or a combination of stationary spatial points. In this case the right hand side of (1.1) takes the form* $u(t)\delta_{x_0}$ *with* $u(t)$ *serving as the control, while* $\delta_{\gamma(t)}$ *equals to the Dirac measure* δ_{x_0} *with* x_0 *the spatial point.*
(2) *In the inverse problems of identifying a moving pointwise source (see* [2]*), one usually encounters the parabolic equations with right hand side* $g(x,t)\delta_{x(t)}$, *where* $g(x,t)$ *is the strength of the source and* $\delta_{x(t)}$ *is the Dirac measure concentrated on the moving spatial point* $x(t)$.
(3) *In the parabolic optimal control problems with controls acting on a lower dimensional manifold (see* [10, 22]*), one encounters the parabolic state equation with right hand side* $u(x,t)\delta_{\gamma(t)}$, *here* $u(x,t)$ *is the control and* $\gamma(t)$ *is a lower dimensional continuous manifold which is strictly contained in* Ω *for all* $t\in[0,T]$. $\delta_{\gamma(t)}$ *denotes the Dirac measure on* $\gamma(t)$. *We note that in this case* $\gamma(t)$ *can be a point, a curve if* $d\geqslant 2$ *or even a surface if* $d=3$, *it can be static and independent of time* t *or evolves in the time horizon.*

Then we consider the parabolic equations with measure data in time. Similarly, the weak solution of problems (1.2) can be defined by transposition techniques. In the following theorem we will give the results on the existence and uniqueness as well as regularity of solution of problem (1.2) which was also proved in [23].

Theorem 2.4. [23] *With the assumption that* $y_0\in L^2(\Omega)$ *and* $g\in\mathcal{C}([0,T];L^2(\Omega))$, *problem (1.2) admits a unique solution* $y\in L^2(0,T;H_0^1(\Omega))\cap L^\infty(0,T;L^2(\Omega))$ *such that*

$$-(y,\partial_t v)_{\Omega_T}+a(y,v)_{\Omega_T}=\langle g\delta_{t_0},v\rangle_{\Omega_T}+(y_0,v(\cdot,0))\quad\forall\ v\in W(0,T)\tag{2.6}$$

with $v(\cdot,T)=0$ *and*

$$\|y\|_{L^2(0,T;H_0^1(\Omega))}+\|y\|_{L^\infty(0,T;L^2(\Omega))}\leqslant C(\|g\|_{L^\infty(0,T;L^2(\Omega))}+\|y_0\|_{L^2(\Omega)}).$$

Here

$$\langle g\delta_{t_0}, v\rangle_{\Omega_T} = \int_{\Omega_T} gvd\delta_{t_0} = \int_{\Omega} g(x,t_0)v(x,t_0)dx, \quad \forall v \in \mathcal{C}([0,T];L^2(\Omega)).$$

Remark 2.5. *We remark that although we assume that Ω is convex in current paper, the convexity assumption is not necessary to prove the existence of a unique solution for problems (1.1) and (1.2).*

In the following, we give an example of parabolic equation with measure data in time which appears in the optimal control theory.

Example 2.6. *Consider the following optimal control problems governed by parabolic PDE:*

$$\min \quad J(y,u) \quad = \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega_T)} + \tfrac{\alpha}{2}\|u\|^2_{L^2(\Omega_T)} \tag{2.7}$$

subject to

$$\begin{cases} \partial_t y + \mathcal{A}y = u & \text{in } \Omega_T, \\ y = 0 & \text{on } \Gamma_T, \\ y(\cdot,0) = y_0 & \text{in } \Omega, \end{cases} \tag{2.8}$$

where y denotes the state variable and u denotes the control. If we impose state constraints pointwise in time, i.e. (see, e.g., [32]*)*

$$\int_{\Omega} f(x,t,y(x,t))dx \leqslant b(t) \quad \forall\ t \in [0,T] \tag{2.9}$$

with given functions $b : [0,T] \to \mathbb{R}$ and $f : \Omega \times [0,T] \times L^2(0,T;L^2(\Omega)) \to \mathbb{R}$, then the adjoint state p associated to the first order optimality conditions satisfies

$$\begin{cases} -\partial_t p + \mathcal{A}^* p = y - y_d + \mu_{\Omega_T} & \text{in } \Omega_T, \\ p = \mu_{\Gamma_T} & \text{on } \Gamma_T, \\ p(\cdot,T) = \mu_T & \text{in } \Omega \end{cases} \tag{2.10}$$

in the sense of distributions (see, e.g., [32]*). Here $\mu_{\Omega_T} := \mu|_{\Omega_T}$, $\mu_{\Gamma_T} := \mu|_{\Gamma_T}$ and $\mu_T := \mu|_{\overline{\Omega}\times\{T\}}$. The Lagrange multiplier μ associated to the state constraints (2.9) appears to be a measure only in time, and can be decomposed as $\mu = g\delta_{t_0}$, g and δ_{t_0} are given such that $g \in \mathcal{C}([0,T];L^2(\Omega))$ and $\delta_{t_0} \in \mathcal{M}[0,T]$. Thus the associated (to the state) adjoint equation exhibits the similar structure of (1.2).*

3. Fully discrete approximations of parabolic equations with Dirac measure in space and a posteriori error estimates

Let us consider the finite element approximations of problem (2.4) in this section. To this aim, we consider a family of triangulation $\mathcal{T}^h$ of $\overline{\Omega}$, such that $\overline{\Omega} = \bigcup_{\tau\in\mathcal{T}^h} \overline{\tau}$. We denote h_τ the diameter of each element $\tau \in \mathcal{T}^h$ and set $h := \max_{\tau\in\mathcal{T}^h} h_\tau$. We suppose that $\overline{\Omega}$ is the union of the elements of $\mathcal{T}^h$. This triangulation is supposed to be shape regular in the usual sense ([14, p.124]).

Here we consider only d-simplex elements, as they are among the most widely used ones. Associated with $\mathcal{T}^h$ is a finite dimensional subspace V_h of $\mathcal{C}(\overline{\Omega})$, such that $\chi|_\tau$ are linear polynomials for $\forall \chi \in V_h$ and $\tau \in \mathcal{T}^h$. Wet set $\mathring{V}_h = V_h \cap H^1_0(\Omega)$. Let $\mathcal{E}^h$ be the set consisting of the interelement edges on the interior of the domain. The quantity

$$\Big[\frac{\partial v}{\partial n_{\mathcal{A}}}\Big] = (A\nabla v)_\tau \cdot n_\tau + (A\nabla v)_{\tau'} \cdot n_{\tau'}$$

defined on the edge $l \in \mathcal{E}^h$, $l = \bar{\tau} \cap \bar{\tau'}$, measures the jump of v across the element edge l. Here n_τ denotes the unit outward normal vector to $\partial\tau$ and $A = (a_{ij}(x))_{d\times d}$.

Moreover, we introduce the well known error estimates for the Clément type interpolation (see [15] and [36] for more details), which will be used in the a posteriori error analysis in this paper.

Lemma 3.1. *Let $\hat{\pi}$ be the Clément type interpolation operator defined in* [15]*, then for any $v \in H^1(\Omega)$ and all element τ,*

$$\|v - \hat{\pi}v\|_{L^2(\tau)} + h_\tau \|\nabla(v - \hat{\pi}v)\|_{L^2(\tau)} \leqslant \sum_{\bar{\tau}' \cap \bar{\tau} \neq \emptyset} C h_\tau |\nabla v|_{L^2(\tau')}, \tag{3.1}$$

$$\|v - \hat{\pi}v\|_{L^2(l)} \leqslant \sum_{l \subset \bar{\tau}'} C h_l^{1/2} |\nabla v|_{L^2(\tau')}, \tag{3.2}$$

where l is the edge or face of the element τ.

We next consider the fully discrete approximations by using the piecewise constant discontinuous Galerkin scheme (dG(0) for short) in time, or equivalently backward Euler scheme. We consider a partitioning of the time interval $\bar{I} = [0, T]$ as

$$\bar{I} = \{0\} \cup I_1 \cup I_2 \cup \cdots \cup I_N$$

with subintervals $I_n = (t_{n-1}, t_n]$ of size $k_n := t_n - t_{n-1} > 0$ and time points

$$0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T.$$

We define the discretization parameter k as a piecewise constant function by setting $k|_{I_n} = k_n$ for $n = 1, 2, \cdots, N$. For $n = 0, 1, 2, \cdots, N$, we construct the finite element spaces $\mathring{V}_h^n \in H_0^1(\Omega)$ (similar to $\mathring{V}_h$) with the mesh $\mathcal{T}_n^h$ (similar to $\mathcal{T}^h$). Similar to $\mathcal{E}^h$ we also denote $\mathcal{E}_n^h$ the union of interelement edges of $\mathcal{T}_n^h$. On each I_n we define the L^2-projection operator π_n onto the piecewise constant space $P_0(I_n)$ as $\pi_n v := \frac{1}{k_n} \int_{I_n} v dt$. Let P_h^n be the $L^2(\Omega)$-projection operator defined from $L^2(\Omega)$ to $\mathring{V}_h^n$:

$$(P_h^n y, v_h) = (y, v_h) \ \ \forall\ v_h \in \mathring{V}_h^n. \tag{3.3}$$

At first we consider the fully discrete finite element approximations of parabolic equations (1.1) with Dirac measure in space, which is to find $Y_h^n \in \mathring{V}_h^n,\ n = 1, 2, \cdots, N$, such that

$$\begin{cases} \Big(\dfrac{Y_h^n - Y_h^{n-1}}{k_n}, w_h\Big) + a(Y_h^n, w_h) = \langle g\delta_{\gamma(t)}, w_h\rangle_{I_n}, \quad \forall\ w_h \in \mathring{V}_h^n, \\ Y_h^0(x) = y_0^h(x), \qquad x \in \Omega, \end{cases} \tag{3.4}$$

where $y_0^h := P_h^0 y_0 \in \mathring{V}_h^0$ is an approximation of y_0. Here

$$\langle g\delta_{\gamma(t)}, v_h\rangle_{I_n} = \frac{1}{k_n} \int_{\overline{\Omega} \times I_n} g v_h d\delta_{\gamma(t)} = \begin{cases} \frac{1}{k_n} \int_{I_n} g(\gamma(t), t) v_h(\gamma(t)) dt & \text{if } m = 0,\ \forall\ v_h \in \mathring{V}_h^n; \\ \frac{1}{k_n} \int_{I_n} \int_{\gamma(t)} g(x, t) v_h(x) dx dt & \text{if } m \geqslant 1,\ \forall\ v_h \in \mathring{V}_h^n. \end{cases}$$

Obviously the above fully discrete finite element scheme admits a unique solution, the stability result of above scheme was also proved in [23]. In the following we denote Y_h the fully discrete finite element approximation of y, which is piecewise constant in time and piecewise linear in space on each time interval.

On each time interval I_n we denote $\mathcal{M}_n^h$ the set consisting of the elements where the measure data $\delta_{\gamma(t)}$ concentrates on. Note that we can only expect that $y \in \mathcal{C}([0, T]; W^{-\epsilon,p}(\Omega))$ for any $\epsilon > 0$ in the case $d - m > 1$. Therefore, we are not able to derive a posteriori error estimates under $L^\infty(0, T; L^2(\Omega))$-norm as in [13] and [27].

Since the solution y of problem (2.4) belongs to $L^2(0, T; W^{1,p}(\Omega))$ for all $p \in (1, \frac{d}{d-1})$, we have $y \in L^2(0, T; L^s(\Omega))$ for all $s < \frac{dp}{d-p}$, which means $s < +\infty$ when $d = 2$ and $s < 3$ when $d = 3$. In the following we will derive $L^2(0, T; L^2(\Omega))$-norm a posteriori error estimates. To improve the readability of the paper we postpone all the proofs to an appendix.

Theorem 3.2. *Assume that $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$, $y_0 \in L^2(\Omega)$ and $\gamma(t)$ is a m-dimensional manifold with $d-m>1$. Let $y \in L^2(0,T;L^2(\Omega))$ be the solution of problem (2.4), and $Y_h^n \in \mathring{V}_h^n, n = 1,2,\cdots,N$ be the solution of problem (3.4). Then there holds*

$$\sum_{n=1}^{N}\int_{I_n}\|y-Y_h^n\|^2_{L^2(\Omega)}dt \leqslant C\Big(\sum_{n=1}^{N}k_n(\xi^n_{res}+\xi^n_{rhs}+\xi^n_{time})+\xi_{ini}\Big), \tag{3.5}$$

where

$$\begin{aligned}
\xi^n_{res}: &= \sum_{\tau\in\mathcal{T}_n^h} h_\tau^4\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|^2_{L^2(\tau)}+\sum_{l\in\mathcal{E}_n^h}h_l^3\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|^2_{L^2(l)};\\
\xi_{ini} &:= \|y_0-Y_h^0\|^2_{L^2(\Omega)};\\
\xi^n_{rhs} &:= \begin{cases}\frac{1}{k_n}\int_{I_n}\sum\limits_{\tau\in\mathcal{M}_n^h}(\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^\infty(\tau)}+h_\tau^{4-d}\|g\|^2_{L^\infty(\tau)})dt & \text{if } m=0;\\ \frac{1}{k_n}\int_{I_n}(\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^2(\gamma(t))}+\sum\limits_{\tau\in\mathcal{M}_n^h}h_\tau^{4-d}\|g\|^2_{L^2(\gamma(t)\cap\tau)})dt & \text{if } m>0;\end{cases}\\
\xi^n_{time} &:= \|Y_h^n-Y_h^{n-1}\|^2_{L^2(\Omega)},
\end{aligned}$$

C is a positive constant depending only on Ω, $\gamma(t)$ and the shape regularity of the triangulation.

In the following we will derive a posteriori error estimate under $L^2(0,T;W^{1,p}(\Omega))$-norm ($p\in(1,\frac{d}{d-1})$) between the solutions of problem (2.4) and (3.4).

Theorem 3.3. *Assume that $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $y_0 \in L^2(\Omega)$, $\gamma(t)$ is a m-dimensional manifold with $d-m>1$. Let $y \in L^2(0,T;W_0^{1,p}(\Omega))$ be the solution of problem (2.4), and $Y_h^n \in \mathring{V}_h^n, n = 1,2,\cdots,N$ be the solution of problem (3.4). Then it holds that*

$$\sum_{n=1}^{N}\int_{I_n}\|y-Y_h^n\|^2_{W^{1,p}(\Omega)}dt \leqslant C\Big(\sum_{n=1}^{N}k_n(\eta^n_{res}+\eta^n_{rhs}+\eta^n_{time})+\eta_{ini}\Big), \tag{3.6}$$

where p is defined in Theorem 2.2 and

$$\begin{aligned}
\eta^n_{res}: &= \Big(\sum_{\tau\in\mathcal{T}_n^h} h_\tau^p\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|^p_{L^p(\tau)}+\sum_{l\in\mathcal{E}_n^h}h_l\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|^p_{L^p(l)}\Big)^{\frac{2}{p}};\\
\eta_{ini} &:= \|y_0-Y_h^0\|^2_{L^p(\Omega)};\\
\eta^n_{rhs} &:= \begin{cases}\frac{1}{k_n}\int_{I_n}(\sum\limits_{\tau\in\mathcal{M}_n^h}\|g-\pi_n g(\gamma(t_n),t)\|^p_{L^\infty(\tau)}+h_\tau^{d-(d-1)p}\|g\|^p_{L^\infty(\tau)})^{\frac{2}{p}} & \text{if } m=0;\\ \frac{1}{k_n}\int_{I_n}(\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^2(\gamma(t))}+(\sum\limits_{\tau\in\mathcal{M}_n^h}h_\tau^{d-(d-1)p}\|g\|^p_{L^2(\gamma(t)\cap\tau)})^{\frac{2}{p}}) & \text{if } m>0;\end{cases}\\
\eta^n_{time} &:= \|Y_h^n-Y_h^{n-1}\|^2_{W^{1,p}(\Omega)},
\end{aligned}$$

C is a positive constant depending only on Ω, $\gamma(t)$ and the shape regularity of the triangulation.

Remark 3.4. *If the measure data $\delta_{\gamma(t)}$ is of Dirac type δ_{x_c} with stationary point x_c, then the error estimator ξ^n_{rhs} reduces to*

$$\frac{1}{k_n}\int_{I_n}((g-\pi_n g)^2(x_c,t)+h_{\tau_{x_c}}^{4-d}g^2(x_c,t))dt,$$

where x_c is the point where Dirac measure δ_{x_c} concentrates on and τ_{x_c} is the element containing x_c. Furthermore, if x_c is the node of the triangulation $\mathcal{T}_n^h$ ($n=1,2,\cdots,N$), then the second term of the above error estimator vanishes, since in the error estimate of E_4 the term $(\psi-\pi_h^n\psi)(x_c,t)=0$ for all triangulation nodes x_c. The same situations occur for the error estimator η^n_{rhs}.

Note that if $d-m=1$, we can conclude from Theorem 2.2 that $y\in L^2(0,T;H^1(\Omega))\cap \mathcal{C}([0,T];L^2(\Omega))$. Then we are able to derive the following standard a posteriori error estimates following the approaches of [9] and [19].

Theorem 3.5. *Assume that $g\in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $y_0\in L^2(\Omega)$, $\gamma(t)$ is a m-dimensional manifold with $d-m=1$. Let $y\in L^2(0,T;H_0^1(\Omega))$ be the solution of problem (2.4), and $Y_h^n\in \mathring{V}_h^n, n=1,2,\cdots,N$ be the solution of problem (3.4). Then for any $1\leqslant J\leqslant N$ it holds that*

$$\|y(\cdot,t_J)-Y_h^J\|^2_{L^2(\Omega)}\leqslant C_J\max_{1\leqslant n\leqslant J}(\eta^n_{res}+\hat{\eta}^n_{rhs}+\eta^n_{time})+\eta_{ini}, \tag{3.7}$$

where η^n_{res}, η_{ini} and η^n_{time} are defined in Theorem 3.3 in the case $p=2$, and

$$\hat{\eta}^n_{rhs}:=\int_{I_n}(\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^2(\gamma(t))}+\sum_{\tau\in\mathcal{M}_n^h}h_\tau\|g\|^2_{L^2(\gamma(t)\cap\tau)}),$$

C_J ($J=1,\cdots,N$) are positive constants depending only on Ω, $\gamma(t)$ and the shape regularity of the triangulation.

Proof. The proof is very similar to [9] and [19] and we omit it here. □

Remark 3.6. *The a posteriori error estimators presented in Theorem 3.2, 3.3 and 3.5 consist of three parts: error estimator ξ_{time} (or η_{time}) due to time discretization, error estimator ξ_{space} (or η_{space}) and ξ_{rhs} (or η_{rhs}) due to space discretization and the part ξ_{ini} (or η_{ini}) from initial value approximation. In our adaptive finite element procedure, the estimator ξ_{time} (or η_{time}) is used to guide time step adaptivity and the estimator ξ_{space} (or η_{space}) and ξ_{rhs} (or η_{rhs}) are used as indicator for space mesh adaptivity. ξ_{ini} (or η_{ini}) is used to generate a mesh on which an accurate initial value approximation can be obtained.*

4. Fully discrete approximations of parabolic equations with Dirac measure in time and a posteriori error estimate

Now we are in a position to define the fully discrete approximations to parabolic equations with measure data in time. The fully discrete approximation scheme of (2.6) is to find $Y_h^n\in\mathring{V}_h^n$, $n=1,2,\cdots,N$, such that

$$\begin{cases}\Big(\dfrac{Y_h^n-Y_h^{n-1}}{k_n},w_h\Big)+a(Y_h^n,w_h)=\langle g\delta_{t_0},w_h\rangle_{I_n},\ \forall w_h\in\mathring{V}_h^n,\\ Y_h^0(x)=y_0^h(x),\qquad x\in\Omega.\end{cases} \tag{4.1}$$

Here

$$\langle g\delta_{t_0},v_h\rangle_{I_n}=\frac{1}{k_n}\int_{\Omega\times I_n}v_h g d\delta_{t_0}=\frac{1}{k_n}\int_\Omega g(x,t_0)v_h(x)dx,\quad\forall v_h\in\mathring{V}_h^n,$$

where we required $t_0\in(t_n,t_{n+1})$ for some $n\in\mathbb{N}$, and y_0^h is the L^2-projection of y_0 in $\mathring{V}_h^0$.

We denote $\mathcal{I}$ the set of indices for time partitions where the measure data δ_{t_0} concentrates on. Now we are ready to estimate the error between the solution y of continuous problem (2.6) and the solution Y_h^n ($n=1,2,\cdots,N$) of the fully discrete problem (4.1). Similar to Theorem 3.2 we are able to derive $L^2(0,T;L^2(\Omega))$-norm a posteriori error estimates.

Theorem 4.1. *Assume that $g\in\mathcal{C}([0,T];L^2(\Omega))$ and $y_0\in L^2(\Omega)$. Let $y\in L^2(0,T;H_0^1(\Omega))\cap L^\infty(0,T;L^2(\Omega))$ be the solution of problem (2.6), and $Y_h^n\in\mathring{V}_h^n$ ($n=1,2,\cdots,N$) be the solution of problem (4.1). Then we have*

$$\sum_{n=1}^N\int_{I_n}\|y-Y_h^n\|^2_{L^2(\Omega)}dt\leqslant C\Big(\sum_{n=1}^N k_n(\xi^n_{res}+\xi^n_{time})+\widetilde{\xi}_{rhs}+\xi_{ini}\Big), \tag{4.2}$$

where ξ^n_{res}, ξ_{ini} *and* ξ^n_{time} *are defined in Theorem 3.2, and*

$$\widetilde{\xi}_{rhs} := \sum_{\tau\in\mathcal{T}_{\mathcal{I}}^h} (k_{\mathcal{I}}\|g(\cdot,t_0)\|^2_{L^2(\tau)} + h_\tau^2\|g(\cdot,t_0)\|^2_{L^2(\tau)}),$$

C is a positive constant depending only on Ω, t_0 and the shape regularity of the triangulation.

Similar to the last section, in the following we derive an error estimate under $L^2(0,T;H^1(\Omega))$-norm.

Theorem 4.2. *Assume that $g \in \mathcal{C}([0,T];L^2(\Omega))$ and $y_0 \in L^2(\Omega)$. Let $y \in L^2(0,T;H^1_0(\Omega)) \cap L^\infty(0,T;L^2(\Omega))$ be the solution of problem (2.6), and $Y^n_h \in \mathring{V}^n_h$ $(n = 1,2,\cdots,N)$ be the solution of problem (4.1). Then we have*

$$\sum_{n=1}^N \int_{I_n} \|y - Y^n_h\|^2_{H^1(\Omega)} dt \leqslant C\Big(\sum_{n=1}^N k_n(\eta^n_{res} + \eta^n_{time}) + \widetilde{\eta}_{rhs} + \eta_{ini}\Big), \tag{4.3}$$

where η^n_{res}, η_{ini} and η^n_{time} are defined in Theorem 3.3 in the case $p = 2$, and

$$\widetilde{\eta}_{rhs} := \sum_{\tau\in\mathcal{T}_{\mathcal{I}}^h} \|g(\cdot,t_0)\|^2_{L^2(\tau)},$$

C is a positive constant depending only on Ω, t_0 and the shape regularity of the triangulation.

5. Adaptive algorithm and numerical experiments

Adaptive finite element procedures consist of the following steps

$$\text{Solve} \to \text{Estimate} \to \text{Refine/Coarsen}.$$

The step Solve is beyond the scope of this paper, while the step Estimate has been done in the above sections. This section is devoted to the constructing of time-step and mesh refinement and coarsening algorithms based on the error equi-distribution strategy. For a detailed research on algorithm design and convergence analysis for linear parabolic equation we refer to [13] and [27].

We can not directly use the results of Theorem 3.2, 3.3, 4.1 and 4.2 for the adaptive strategy because the time stock. As in [9] and [33] we may use $\zeta^n_{space} := \xi^n_{res} + \xi^n_{rhs}$, $\zeta^n_{time} := \xi^n_{time}$, $\zeta_{init} := \xi_{init}$ in Theorem 3.2 and $\zeta^n_{space} := \eta^n_{res} + \eta^n_{rhs}$, $\zeta^n_{time} := \eta^n_{time}$, $\zeta_{init} := \eta_{init}$ in Theorem 3.3 as error indicators on each time step for parabolic equation with measure data in space and the similar adaptation of Theorem 4.1 and 4.2 for parabolic equation with measure data in time. In the following numerical examples we only consider the adaptivity by using the error estimates derived under $L^2(0,T;L^2(\Omega))$-norm.

Let TOL_{time} be the total tolerance allowed for the part of a posteriori error estimate related to the time discretization, that is

$$\Big(\sum_{n=1}^N k_n(\zeta^n_{time})^2\Big)^{\frac{1}{2}} \leqslant \mathrm{TOL}_{time}. \tag{5.1}$$

Our goal is to achieve the following error equi-distributions:

$$\zeta^n_{time} \leqslant \frac{\mathrm{TOL}_{time}}{T}. \tag{5.2}$$

If (5.2) is violated in current step, the step size should be refined, otherwise, if

$$\zeta^n_{time} \leqslant \theta_{time}\frac{\mathrm{TOL}_{time}}{T}, \tag{5.3}$$

then the current step size should be enlarged to reduce computational cost.

For the space mesh refinement we follow the similar idea. Let TOL_{space} denote the tolerance for the part of a posteriori error estimates related to space triangulation, our aim is to achieve

$$\Big(\sum_{n=1}^{N} k_n(\zeta_{space}^n)^2\Big)^{\frac{1}{2}} \leqslant \mathrm{TOL}_{space}. \tag{5.4}$$

We do space adaptivity on each time step until that

$$\zeta_{space}^n \leqslant \frac{\mathrm{TOL}_{space}}{T}. \tag{5.5}$$

On each time step we use the fully adapted space mesh from last time step as the initial mesh. Since the location of singularity of the solution may vary with respect to time evolution, we need the mesh coarsening to save the computation cost, here we following the ideas of the adaptive finite element library AFEM ([12]), where the mesh refinement and coarsening algorithms are implemented by using the bisection algorithm. On each time step we refine the minimum set of elements $S_n \subset \mathcal{T}_n^h$ verifying

$$\zeta_{space}^n(S_n) > \theta\zeta_{space}^n(\mathcal{T}_n^h),$$

and we mark the minimum set of elements $C_n \subset \mathcal{T}_n^h$ to coarsen verifying

$$\zeta_{space}^n(C_n) \leqslant \theta_c\zeta_{space}^n(\mathcal{T}_n^h),$$

where $\eta_{space}^n(S_n)$ denotes the sum of elementwise error estimators in S_h and $\eta_{space}^n(\mathcal{T}_n^h)$ the sum of elementwise error estimators in $\mathcal{T}_n^h$. Now we give the space-time adaptive algorithm as follows.

Algorithm 5.1. *Space-time adaptive algorithm*

Given tolerances TOL_{time}, TOL_{space}, parameter $\delta_1 \in (0,1)$, $\delta_2 > 1$ and $\theta_{time} \in (0,1)$. Start with initial time step k_0, initial mesh $\mathcal{T}_0^h$, and initial solution Y_h^0. Set $n = 1$ and $t_0 = 0$.

(I) Compute the initial error indicator ζ_{init}, refine $\mathcal{T}_0^h$ to get a mesh such that $\zeta_{init} < TOL_{space}$.

(II) While $t_n < T$, do

Given Y_h^{n-1} from the previous time step at time t_{n-1}, with the mesh $\mathcal{T}_{n-1}^h$ and time step size k_{n-1}.

(1) *Set $\mathcal{T}_n^h := \mathcal{T}_{n-1}^h$, $k_n := k_{n-1}$ and $t_n = t_{n-1} + k_n$.*
   (a) *Solve the discrete equation for Y_h^n on mesh $\mathcal{T}_n^h$ with time step size k_n and data Y_h^{n-1}.*
   (b) *Compute the a posteriori error estimates on $\mathcal{T}_n^h$.*

(2) *while (5.2) is not satisfied*
   (a) *Set $k_n := \delta_1 k_{n-1}$ and $t_n := t_{n-1} + k_n$.*
   (b) *Solve the discrete equation for Y_h^n on mesh $\mathcal{T}_n^h$ with time step size k_n and data Y_h^{n-1}.*
   (c) *Compute the a posteriori error estimates on $\mathcal{T}_n^h$.*

   end while.

(3) *while (5.5) is not satisfied*
   (a) *Refine and coarsen mesh $\mathcal{T}_n^h$ to generate a new mesh $\mathcal{T}_n^h$.*
   (b) *Solve the discrete equation for Y_h^n on mesh $\mathcal{T}_n^h$ with time step size k_n and data Y_h^{n-1}.*
   (c) *Compute the a posteriori error estimates on $\mathcal{T}_n^h$.*
   (d) *while (5.2) is not satisfied do*
      (i) *Set $k_n := \delta_1 k_{n-1}$ and $t_n := t_{n-1} + k_n$.*
      (ii) *Solve the discrete equation for Y_h^n on mesh $\mathcal{T}_n^h$ with time step size k_n and data Y_h^{n-1}.*
      (iii) *Compute the a posteriori error estimates on $\mathcal{T}_n^h$.*

   end while

   end while

(4) *If (5.3) is satisfied then*
   $k_n := \delta_2 k_n$.

   end if

end while

In the following of this section we will carry out some numerical experiments to support our theoretical findings. We set the parameters appear in Algorithm 5.1 as $\delta_1 = 0.5$, $\delta_2 = 2$, $\theta_{time} = 0.5$, $\theta = 0.2$ and $\theta_c = 0.05$.

5.1. **Parabolic equations with measure data in space.** At first, we consider the following parabolic equation with Dirac source term in space:

$$\begin{cases} y_t - \Delta y = f + g(t)\delta_\gamma & \text{in } \Omega_T, \\ y = y_d & \text{on } \Gamma_T, \\ y(0) = 0 & \text{in } \Omega. \end{cases}$$

For ease of constructing examples we may admit some additional regular parts to appear in the right hand side.

Example 5.2. *The first example is taken from* [22] *with modifications. Let* $\Omega_T = (-1,1)^2 \times (0,1)$. *We take the exact solution as*

$$y(x,t) = -\frac{1}{2\pi} \log|x - \gamma(t)| \cdot g(t),$$

where $g(t)$ *is defined as follows*

$$g(t) = 1 - e^{-500\times(t-0.5)^2}.$$

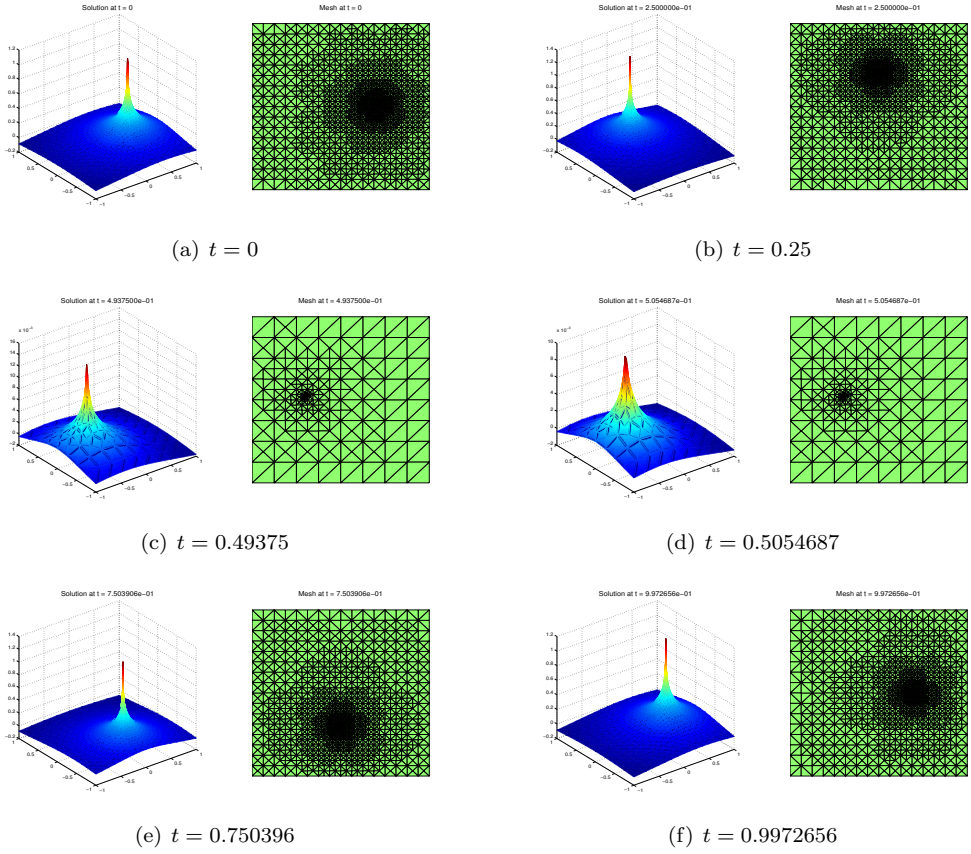
After simple calculation we see δ_γ *is the Dirac measure at spatial point* $\gamma(t) = (0.4\cos(2\pi t), 0.4\sin(2\pi t))$.

In this example we set $\mathrm{TOL}_{time} = \mathrm{TOL}_{space} = 0.005$. In Figure 1 we show the profiles of the solutions and their corresponding adaptively refined meshes at different time nodes. Since the Dirac source is moving around, the mesh refinement and coarsening on each time step are crucial to reduce the computational cost. From Figure 1 we can observe that the singularity can be captured by the adaptive mesh satisfactorily on each time node. In Figure 2 we report the number of degrees of freedom (DOF) and time step size, and find that the adaptive refined mesh and time step are adapted to the exact solution very well. Especially, when the solution changes very fast in time around time level 0.5 the time step size becomes very small. We also report the L^2-norm errors of the solutions at different time nodes in Figure 3 together with the error indicators.

To give some quantitative assessments of derived a posteriori error estimators and the performance of the adaptive algorithm, we list in Table 1 the number of time steps N, the average number of nodes of the meshes, the total estimated error η, the total L^2-norm error $\mathcal{E}$ and the effectiveness index defined by $\mathrm{eff} := \frac{\eta}{\mathcal{E}}$, for different values of TOL for Example 5.2. We observe that the effectiveness index is stable with respect to the tolerance, which is very similar to that of [13] and confirms the efficiency of the error estimator derived in this paper.

TABLE 1. The number of time steps N, the average number of nodes of the meshes 'Node', the total estimated error η, the total L^2-norm error $\mathcal{E}$ and the effectiveness index 'eff', for different values of 'TOL' for Example 5.2.

TOL	N	Node	$\mathcal{E}$	η	eff
0.06	1694	224	9.273e-2	6.207e-1	6.694
0.04	2177	280	7.470e-2	5.510e-1	7.376
0.02	3552	507	5.151e-2	3.976e-1	7.719
0.01	4266	991	3.616e-2	2.523e-1	6.977
0.005	4856	2018	1.808e-2	1.548e-1	8.562



(a) $t = 0$ (b) $t = 0.25$

(c) $t = 0.49375$ (d) $t = 0.5054687$

(e) $t = 0.750396$ (f) $t = 0.9972656$

FIGURE 1. The profiles of the solutions Y_h (left) and their corresponding adaptively refined meshes (right) at different time nodes $t = 0, 0.25, 0.49375, 0.5054687, 0.750396, 0.9972656$ for Example 5.2.

5.2. **Parabolic equations with measure data in time.** Then we consider the following parabolic equation with Dirac right hand side in time:

$$\begin{cases} \partial_t y - \Delta y = f + g(x,t)\delta_{t_0} & \text{in } \Omega_T, \\ y = y_d & \text{on } \Gamma_T, \\ y(\cdot, 0) = y_0 & \text{in } \Omega, \end{cases}$$

where $g(x,t) \in \mathcal{C}([0,T]; L^2(\Omega))$ and δ_{t_0} is the Dirac measure concentrated on the time node $t_0 \in (0,T)$. For ease of constructing examples we may also admit some additional regular parts to appear in the right hand side.

Example 5.3. *The second example is a parabolic equation with Dirac measure in time (see* [23]*). Let* $\Omega_T = (-1,1)^2 \times (0,1)$*. We take the exact solution as*

$$y(x,t) = 0.1\exp(-25|x-(t-0.5)|^2) \cdot \begin{cases} t^2, & t < \dfrac{1}{2}; \\ t^2 + 2t, & t \geqslant \dfrac{1}{2}. \end{cases}$$

After simple calculation we have

$$g(x,t)\delta_{t_0} = 0.1\exp(-25|x-(t-0.5)|^2)\delta_{\frac{1}{2}}(t) \tag{5.6}$$

with $t_0 = 0.5$ *and additional right hand side* f*.*

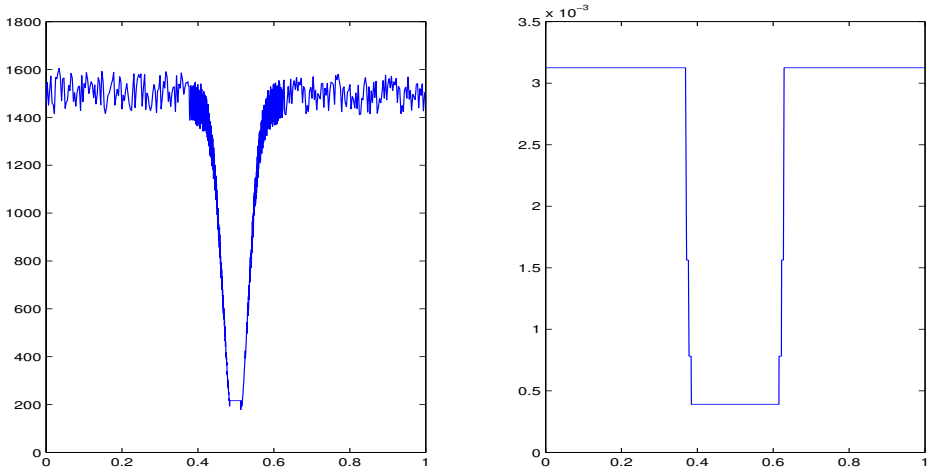


FIGURE 2. Numbers of DOF for spatial triangulation (left) on different time nodes and the corresponding time step sizes (right) for time discretization for Example 5.2.

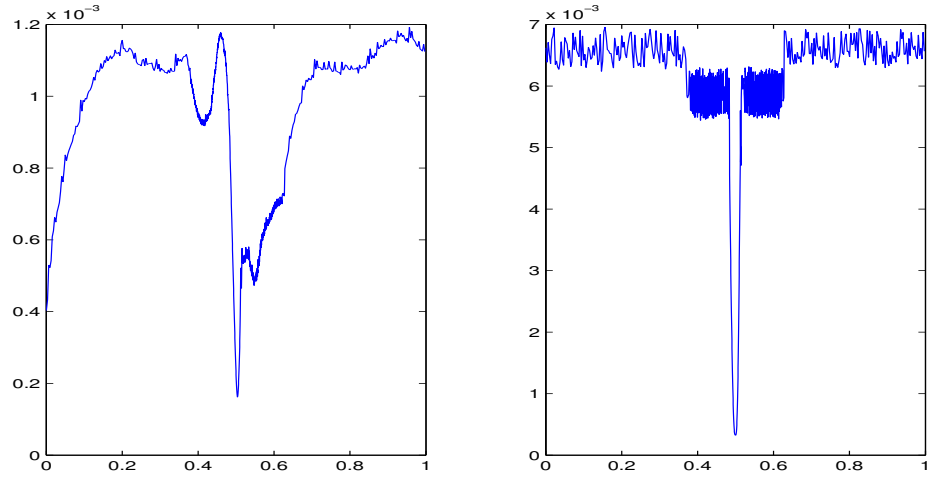


FIGURE 3. The L^2-norm errors (left) of the discrete solutions Y_h and the corresponding error estimators (right) on different time nodes for Example 5.2.

Note that the solution of this example is not continuous and exhibits a jump at $t = 0.5$. Therefore, with the choice of $\zeta^n_{space} := \xi^n_{res} + \xi^n_{rhs}$ and $\zeta^n_{time} =: \xi^n_{time}$ the time step size may be refined never-ending around $t = 0.5$, caused by the jump of the data and the solution. We plot in Figure 4 the numbers of DOF for spatial triangulation on different time nodes and the corresponding time step sizes for time discretization, where we set a minimum time step size and choose $\mathrm{TOL}_{time} = 0.0025, \mathrm{TOL}_{space} = 0.025$. We observe that the meshes should be refined fine enough to control the bound of the error indicators. In Figure 5 we show the L^2-norm errors of the solution at different time nodes and the error indicators. It is clear that the error indicator is very big around the time node $t = 0.5$ in this case.

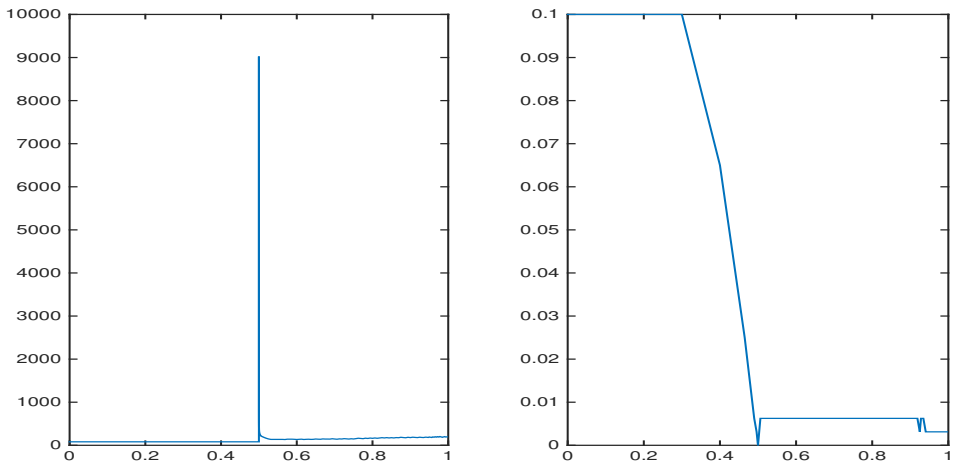


FIGURE 4. Numbers of DOF for spatial triangulation (left) on different time nodes and the corresponding time step sizes (right) for time discretization for Example 5.3.

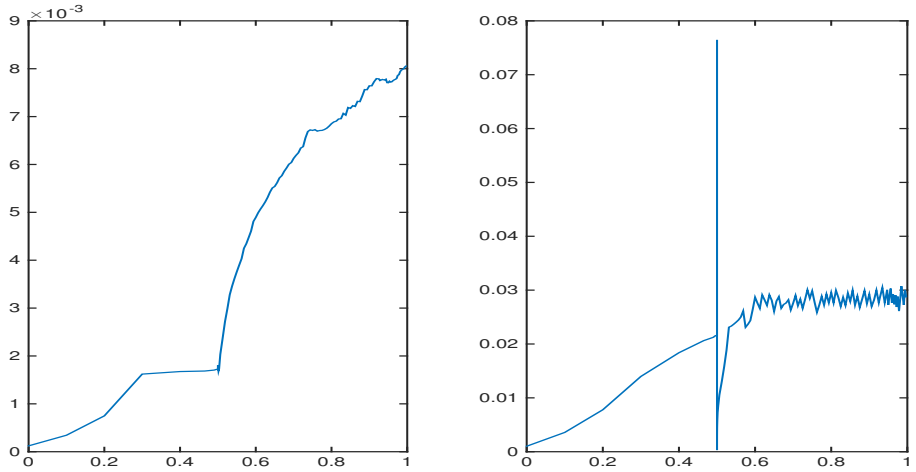
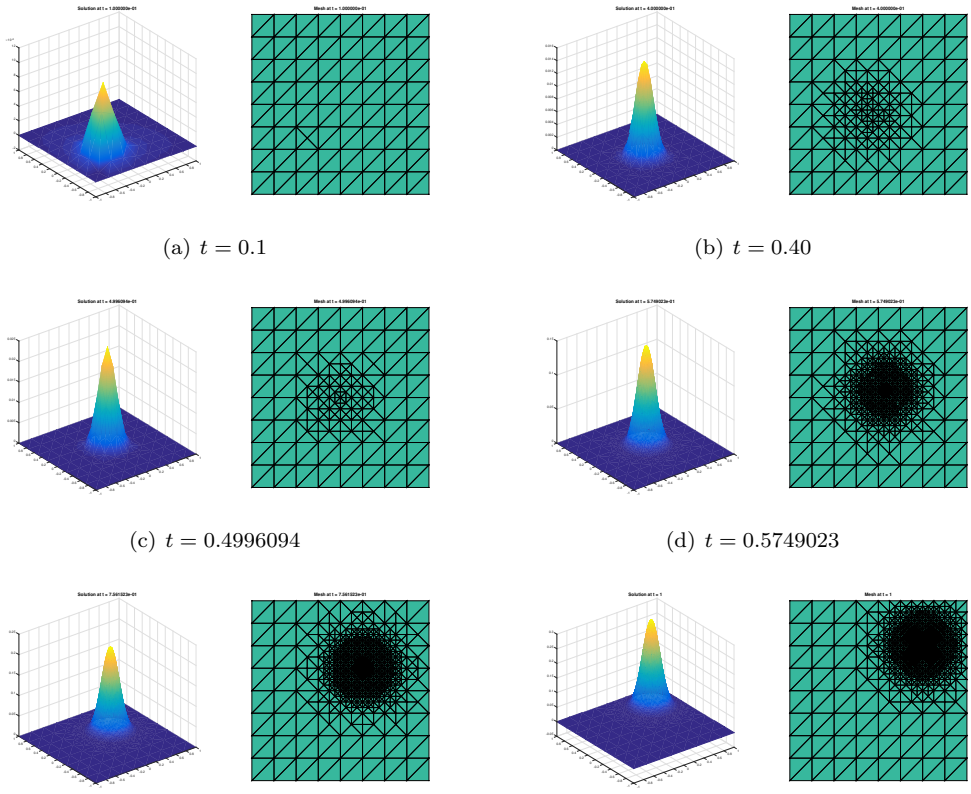


FIGURE 5. The L^2-norm errors (left) of the discrete solutions Y_h and the corresponding error estimators (right) on different time nodes for Example 5.3.

As a remedy we use $(\zeta_{space}^n)^2 := k_n((\xi_{res}^n)^2 + (\xi_{rhs}^n)^2)$ and $(\zeta_{time}^n)^2 := k_n(\xi_{time}^n)^2$ as error indicators for parabolic equation with Dirac measure in time, compared to Example 5.2. In this case we set $\mathrm{TOL}_{time} = \mathrm{TOL}_{space} = 2.5e-4$. As in Example 5.2 we show in Figure 6 the profiles of the solutions and their corresponding adaptively refined meshes at different time nodes. From Figure 6 we can observe that the mesh nodes are concentrated on the place where the solution varies largely. In Figure 7 we report the number of degrees of freedom (DOF) and the time step size, and find that the adaptive refined mesh and time step are adapted to the exact solution very well. Especially, when the solution has singularity in time around time node 0.5 the time step size becomes very small. Moreover, to be consistent with the error indicators we compute the discrete L^2-norm errors of the solution at different time nodes, i.e., $(k_n\|Y_h^n - y(\cdot,t_n)\|^2_{L^2(\Omega)})^{\frac{1}{2}}$, which are

illustrated in Figure 8 together with the error indicators. Figure 6, 7 and 8 show that the solution is very small at the beginning and grows slowly and shows a jump at $t = 0.5$. We remark that the last step size at time $t = 1$ is very small because we set $t = 1.0$ as the stopping rule of the code, and it is not caused by the singularity of the solution.

Similarly, we list in Table 2 the number of time steps N, the average number of nodes of the meshes, the total estimated error η, the total L^2-norm error $\mathcal{E}$ and the effectiveness index, for different values of TOL for Example 5.3. We can observe the similar phenomena as in Example 5.2.



(a) $t = 0.1$ (b) $t = 0.40$

(c) $t = 0.4996094$ (d) $t = 0.5749023$

(e) $t = 0.7561523$ (f) $t = 1.0$

FIGURE 6. The profiles of the solutions Y_h (left) and their corresponding adaptively refined meshes (right) at different time nodes $t = 0.1, 0.4, 0.4996094, 0.5749023, 0.7561523, 1.0$ for Example 5.3.

ACKNOWLEDGMENTS

This work was partially supported by the National Natural Science Foundation of China under Grants 11001027, 11571356, 11671391, 91530204, and the National Basic Research Program under the Grant 2011CB309705 and 2012CB821204. The authors would like to thank Prof. Chensong Zhang for help on the implementation of the space-time adaptive algorithm and two anonymous referees for their helpful suggestions.

APPENDIX

Proof of Theorem 3.2: To derive the a posteriori error estimates we use the duality argument. Let ψ be the solution of problem (2.1) with $f \in L^2(0, T; L^2(\Omega))$. Note that $\psi = 0$ on $\partial\Omega$,

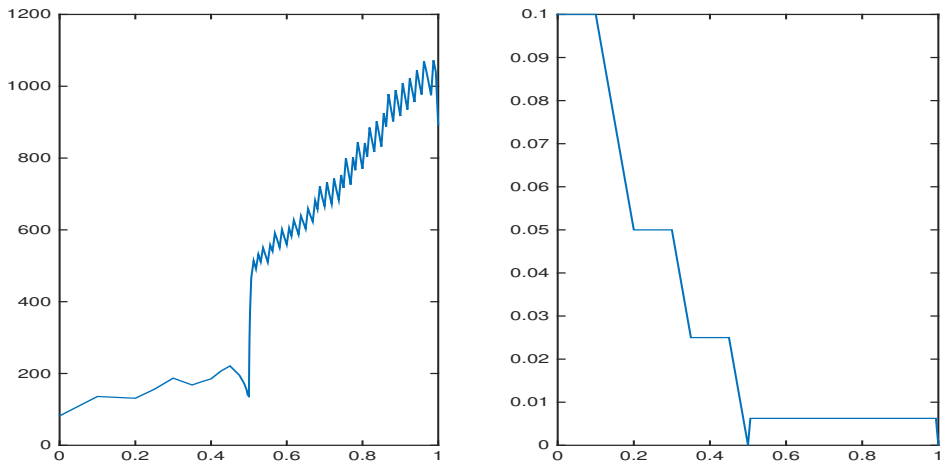


FIGURE 7. Numbers of DOF for spatial triangulation (left) on different time nodes and the corresponding time step sizes (right) for time discretization for Example 5.3.

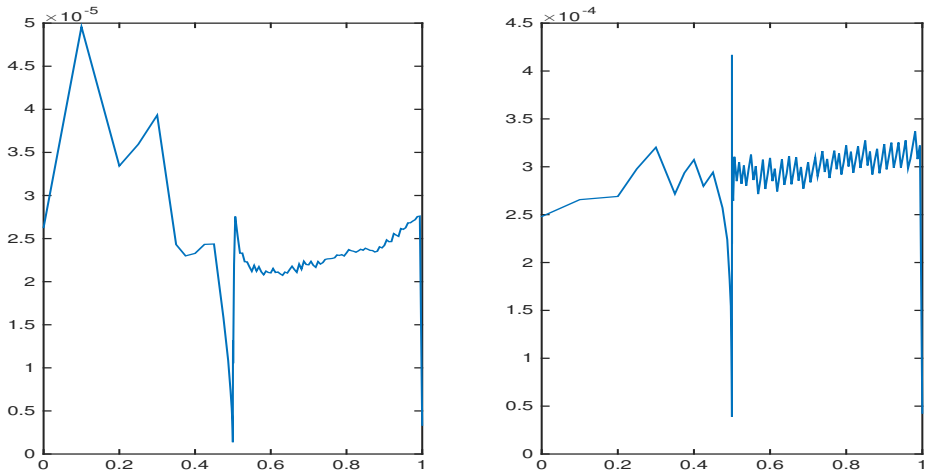


FIGURE 8. The L^2-norm errors (left) of the discrete solutions Y_h and the corresponding error estimators (right) on different time nodes for Example 5.3.

$\psi^N = \psi(T) = 0$, it follows from (2.4) and integrating by parts that

$$\int_{\Omega_T} (y - Y_h) f dx dt = \int_0^T \int_\Omega (y - Y_h)(-\partial_t \psi + \mathcal{A}^* \psi) dx dt$$

$$= -(y, \partial_t \psi)_{\Omega_T} + (y, \mathcal{A}^* \psi)_{\Omega_T} + \sum_{n=1}^N \int_{I_n} ((Y_h, \partial_t \psi) - a(Y_h, \psi)) dt$$

$$= -\sum_{n=1}^N \int_{I_n} (k_n^{-1}(Y_h^n - Y_h^{n-1}, \psi^{n-1}) + a(Y_h^n, \psi)) dt + \langle g \delta_{\gamma(t)}, \psi \rangle_{\Omega_T} + (y_0 - Y_h^0, \psi(\cdot, 0)).$$

TABLE 2. The number of time steps N, the average number of nodes of the meshes 'Node', the total estimated error η, the total L^2-norm error $\mathcal{E}$ and the effectiveness index 'eff', for different values of 'TOL' for Example 5.3.

TOL	N	Node	$\mathcal{E}$	η	eff
4e-3	30	200	4.076e-3	6.267e-2	15.38
2e-3	46	284	3.321e-3	4.931e-2	14.85
1e-3	68	421	2.743e-3	3.931e-2	14.33
5e-4	105	624	2.275e-3	2.991e-2	13.15
2.5e-4	171	949	1.817e-3	2.454e-2	13.51

We define ψ_I such that $\psi_I|_{I_n} := \pi_h^n\pi_n\psi \in \mathring{V}_h^n$ on each time interval I_n with π_h^n the standard Lagrange interpolation operator onto $\mathring{V}_h^n$. Note that $\psi \in L^2(0,T;H^2(\Omega)) \hookrightarrow L^2(0,T;\mathcal{C}(\overline{\Omega}))$, so the Lagrange interpolation is well-defined. It follows from (3.4) that

$$\sum_{n=1}^{N} k_n^{-1}(Y_h^n - Y_h^{n-1},\psi_I) + a(Y_h^n,\psi_I) = \sum_{n=1}^{N}\langle g\delta_{\gamma(t)},\psi_I\rangle_{I_n}. \tag{5.7}$$

Then we have

$$\begin{aligned}
&\int_{\Omega_T}(y-Y_h)f\,dxdt\\
=\;& -\sum_{n=1}^{N}\int_{I_n}\Big(k_n^{-1}(Y_h^n - Y_h^{n-1},\psi^{n-1}-\psi_I) + a(Y_h^n,\psi-\psi_I)\Big)dt\\
&+(y_0 - Y_h^0,\psi(\cdot,0)) + \langle g\delta_{\gamma(t)},\psi-\psi_I\rangle_{\Omega_T}\\
:=\;& E_1+E_2+E_3. \qquad (5.8)
\end{aligned}$$

Now we estimate the above terms one by one. Integrating by parts, using the definition of π_n and the estimates of the Lagrange interpolation ([14, Sec.3.1]) we are led to

$$\begin{aligned}
E_1 =\;& -\sum_{n=1}^{N}\int_{I_n}\Big(k_n^{-1}(Y_h^n - Y_h^{n-1},\psi^{n-1}-\pi_n\psi) + \sum_{l\in\mathcal{E}_n^h}\int_l\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big](\psi-\pi_h^n\psi)ds\\
&+\sum_{\tau\in\mathcal{T}_n^h}\int_\tau\big(k_n^{-1}(Y_h^n - Y_h^{n-1}) + \mathcal{A}Y_h^n\big)(\psi-\pi_h^n\psi)dx\Big)dt\\
\leqslant\;& C\Big[\sum_{n=1}^{N}k_n\Big(\|Y_h^n - Y_h^{n-1}\|_{L^2(\Omega)}^2 + \sum_{\tau\in\mathcal{T}_n^h}h_\tau^4\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n} + \mathcal{A}Y_h^n\Big\|_{L^2(\tau)}^2\\
&+\sum_{l\in\mathcal{E}_n^h}h_l^3\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|_{L^2(l)}^2\Big)\Big]^{\frac12}(\|\partial_t\psi\|_{L^2(0,T;L^2(\Omega))} + \|\psi\|_{L^2(0,T;H^2(\Omega))}). \qquad (5.9)
\end{aligned}$$

For E_2 we have

$$E_2 = (y_0 - Y_h^0,\psi(\cdot,0)) \leqslant C\|y_0 - Y_h^0\|_{L^2(\Omega)}\|\psi(\cdot,0)\|_{L^2(\Omega)}. \tag{5.10}$$

In case that $m=0$, from the structure of E_3 and L^∞-error estimate for the Lagrange interpolation we can deduce

$$\begin{aligned}
E_3 &= \langle g\delta_{\gamma(t)}, \psi-\psi_I\rangle_{\Omega_T} = \sum_{n=1}^{N}\int_{I_n} g(\gamma(t),t)(\psi-\psi_I)(\gamma(t),t)dt \\
&= \sum_{n=1}^{N}\Big(\int_{I_n} g(\gamma(t),t)(\psi-\pi_n\psi)(\gamma(t),t)dt + \int_{I_n} g(\gamma(t),t)\pi_n(\psi-\pi_h^n\psi)(\gamma(t),t)dt\Big) \\
&\leqslant C\sum_{n=1}^{N}\int_{I_n}\sum_{\tau\in\mathcal{M}_n^h}\Big(\|g-\pi_n g(\gamma(t_n),t)\|_{L^\infty(\tau)}\|\psi-\pi_n\psi\|_{L^\infty(\tau)} \\
&\quad +\|g\|_{L^\infty(\tau)}\|\pi_n(\psi-\pi_h^n\psi)\|_{L^\infty(\tau)}\Big) \\
&\leqslant C\Big(\sum_{n=1}^{N}\int_{I_n}\sum_{\tau\in\mathcal{M}_n^h}(\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^\infty(\tau)} + h_\tau^{4-d}\|g\|^2_{L^\infty(\tau)})dt\Big)^{\frac{1}{2}}\|\psi\|_{L^2(0,T;H^2(\Omega))}. \qquad (5.11)
\end{aligned}$$

Similarly, in case that $m>0$ we can derive

$$\begin{aligned}
E_3 &= \langle g\delta_{\gamma(t)}, \psi-\psi_I\rangle_{\Omega_T} = \sum_{n=1}^{N}\int_{I_n}\int_{\gamma(t)} g(x,t)(\psi-\psi_I)(x,t)dxdt \\
&= \sum_{n=1}^{N}\Big(\int_{I_n}\int_{\gamma(t)} g(x,t)(\psi-\pi_n\psi)(x,t)dxdt + \int_{I_n}\int_{\gamma(t)} g(x,t)\pi_n(\psi-\pi_h^n\psi)(x,t)dxdt\Big) \\
&\leqslant C\sum_{n=1}^{N}\int_{I_n}\Big(\|g-\pi_n g(\gamma(t_n),t)\|_{L^2(\gamma(t))}\|\psi-\pi_n\psi\|_{L^2(\gamma(t))} \\
&\quad +\|g\|_{L^2(\gamma(t))}\|\pi_n(\psi-\pi_h^n\psi)\|_{L^2(\gamma(t))}\Big)dt \\
&\leqslant C\Big(\sum_{n=1}^{N}\int_{I_n}(\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^2(\gamma(t))} + \sum_{\tau\in\mathcal{M}_n^h} h_\tau^{4-d}\|g\|^2_{L^2(\gamma(t)\cap\tau)})dt\Big)^{\frac{1}{2}}\|\psi\|_{L^2(0,T;H^2(\Omega))}. \qquad (5.12)
\end{aligned}$$

Combining (5.8)-(5.12) and recalling (2.2)-(2.3) we obtain

$$\begin{aligned}
&\|y-Y_h\|_{L^2(0,T;L^2(\Omega))} \\
&\leqslant C\Big(\sum_{n=1}^{N}k_n\Big(\sum_{\tau\in\mathcal{T}_n^h}h_\tau^4\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|^2_{L^2(\tau)} + \sum_{l\in\mathcal{E}_n^h}h_l^3\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|^2_{L^2(l)}\Big)\Big)^{\frac{1}{2}} \\
&\quad +C\|y_0-Y_h^0\|_{L^2(\Omega)} + C\Big(\sum_{n=1}^{N}\int_{I_n}\sum_{\tau\in\mathcal{M}_n^h}(\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^\infty(\tau)} + h_\tau^{4-d}\|g\|^2_{L^\infty(\tau)})dt\Big)^{\frac{1}{2}} \\
&\quad +C\Big(\sum_{n=1}^{N}k_n\|Y_h^n-Y_h^{n-1}\|^2_{L^2(\Omega)}\Big)^{\frac{1}{2}} \qquad (5.13)
\end{aligned}$$

in the case $m=0$. We can prove the similar result for the case $m>0$, this completes the proof. □

Proof of Theorem 3.3: Let q be the conjugate number of p such that $\frac{1}{p}+\frac{1}{q}=1$, then we have $q>d\geqslant 2$. Given $\vec{\Psi}\in L^2(0,T;L^q(\Omega)^d)$, consider now the following problem:

$$-\langle\psi_t,w\rangle_I + a(w,\psi)_{\Omega_T} = \int_0^T\int_\Omega \vec{\Psi}\cdot\nabla w dxdt \quad \forall w\in L^2(0,T;W_0^{1,p}(\Omega)) \qquad (5.14)$$

with $\psi(\cdot,T)=0$. Note that $\vec{\Psi}$ can be identified as an element of $L^2(0,T;W^{-1,q}(\Omega))$, it follows that the above problem admits a unique solution $\psi\in L^2(0,T;W_0^{1,q}(\Omega))\cap H^1(0,T;W^{-1,q}(\Omega))$ (see [1] and [16]). From the embedding theorem in [1] (see, e.g., P. 180, Theorem 4.10.2) we conclude that $\psi\in\mathcal{C}([0,T];(W^{-1,q}(\Omega),W^{1,q}(\Omega))_{\frac{1}{2},2})$. Interpolation theory for Sobolev space and embedding theorem for Besov space ([28, Corollary A.28]) imply that $\psi\in\mathcal{C}([0,T];L^q(\Omega))$. Moreover, there holds

$$\|\psi\|_{L^2(0,T;W_0^{1,q}(\Omega))}+\|\partial_t\psi\|_{L^2(0,T;W^{-1,q}(\Omega))}+\|\psi\|_{\mathcal{C}([0,T];L^q(\Omega))}\leqslant C\|\vec{\Psi}\|_{L^2(0,T;L^q(\Omega)^d)}. \tag{5.15}$$

It follows from (2.5) and (5.14) that

$$\begin{aligned}
&\int_0^T\int_\Omega\nabla(y-Y_h)\cdot\vec{\Psi}dxdt=-\langle\psi_t,y-Y_h\rangle_I+\int_0^T a(y-Y_h,\psi)dt\\
=\;&\langle y_t,\psi\rangle_I+\langle\psi_t,Y_h\rangle_I+\int_0^T(a(y,\psi)-a(Y_h,\psi))dt+(\psi(\cdot,0),y_0)\\
=\;&\langle g\delta_{\gamma(t)},\psi\rangle_{\Omega_T}+(\psi(\cdot,0),y_0-Y_h^0)-\sum_{n=1}^N\int_{I_n}(k_n^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1})+a(Y_h^n,\psi))dt. \qquad (5.16)
\end{aligned}$$

On each time interval I_n we also define $\psi_I|_{I_n}:=\pi_h^n\pi_n\psi\in\mathring{V}_h^n$ with π_h^n the standard Lagrange interpolation operator onto $\mathring{V}_h^n$, which is well-defined due to the fact that $\psi\in L^2(0,T;W^{1,q}(\Omega))\hookrightarrow L^2(0,T;\mathcal{C}(\overline{\Omega}))$. Again, using (5.7) we arrive at

$$\begin{aligned}
\int_0^T\int_\Omega\nabla(y-Y_h)\cdot\vec{\Psi}dxdt&=-\sum_{n=1}^N\int_{I_n}(k_n^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1}-\psi_I)+a(Y_h^n,\psi-\psi_I))dt\\
&\quad+(y_0-Y_h^0,\psi(\cdot,0))+\langle g\delta_{\gamma(t)},\psi-\psi_I\rangle_{\Omega_T}\\
&:=F_1+F_2+F_3. \qquad (5.17)
\end{aligned}$$

Now it remains to estimate the above terms one by one. Integrating by parts, using the definition of π_n and the estimates of the Lagrange interpolation ([14, Sec.3.1]) result in

$$\begin{aligned}
F_1&=-\sum_{n=1}^N\int_{I_n}\Big(k_n^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1}-\pi_n\psi)+\sum_{l\in\mathcal{E}_n^h}\int_l\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big](\psi-\pi_h^n\psi)ds\\
&\quad+\sum_{\tau\in\mathcal{T}_n^h}\int_\tau(k_n^{-1}(Y_h^n-Y_h^{n-1}+\mathcal{A}Y_h^n)(\psi-\pi_h^n\psi)dx)\Big)dt\\
&\leqslant C\sum_{n=1}^N\int_{I_n}\Big(\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}\Big\|_{W^{1,p}(\Omega)}\|\psi^{n-1}-\pi_n\psi\|_{W^{-1,q}(\Omega)}\\
&\quad+\sum_{\tau\in\mathcal{T}_n^h}\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|_{L^p(\tau)}\|\psi-\pi_h^n\psi\|_{L^q(\tau)}\\
&\quad+\sum_{l\in\mathcal{E}_n^h}\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|_{L^p(l)}\|\psi-\pi_h^n\psi\|_{L^q(l)}\Big)dt\\
&\leqslant C\Big[\sum_{n=1}^N k_n\Big(\sum_{\tau\in\mathcal{T}_n^h}\|Y_h^n-Y_h^{n-1}\|_{W^{1,p}(\tau)}^p+h_\tau^p\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|_{L^p(\tau)}^p\\
&\quad+\sum_{l\in\mathcal{E}_n^h}h_l\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|_{L^p(l)}^p\Big)^{\frac{2}{p}}\Big]^{\frac{1}{2}}(\|\partial_t\psi\|_{L^2(0,T;W^{-1,q}(\Omega))}+\|\psi\|_{L^2(0,T;W^{1,q}(\Omega))}). \qquad (5.18)
\end{aligned}$$

For F_2 we have

$$F_2=(y_0-Y_h^0,\psi(\cdot,0))\leqslant C\|y_0-Y_h^0\|_{L^p(\Omega)}\|\psi(\cdot,0)\|_{L^q(\Omega)}. \tag{5.19}$$

In the case of $m = 0$, from the structure of F_3 and L^∞-norm error estimate of the Lagrange interpolation we can deduce

$$
\begin{aligned}
F_3 &= \langle g\delta_{\gamma(t)}, \psi - \psi_I\rangle_{\Omega_T} = \sum_{n=1}^{N}\int_{I_n} g(\gamma(t),t)(\psi-\psi_I)(\gamma(t),t)dt \\
&= \sum_{n=1}^{N}\Big(\int_{I_n} g(\gamma(t),t)(\psi-\pi_n\psi)(\gamma(t),t)dt + \int_{I_n} g(\gamma(t),t)\pi_n(\psi-\pi_h^n\psi)(\gamma(t),t)dt\Big) \\
&\leqslant C\sum_{n=1}^{N}\int_{I_n}\sum_{\tau\in\mathcal{M}_n^h}\Big(\|g-\pi_n g(\gamma(t_n),t)\|_{L^\infty(\tau)}\|\psi-\pi_n\psi\|_{L^\infty(\tau)} \\
&\quad +\|g\|_{L^\infty(\tau)}\|\pi_n(\psi-\pi_h^n\psi)\|_{L^\infty(\tau)}\Big) \\
&\leqslant C\Big(\sum_{n=1}^{N}\int_{I_n}\big(\sum_{\tau\in\mathcal{M}_n^h}\|g-\pi_n g(\gamma(t_n),t)\|_{L^\infty(\tau)}^p + h_\tau^{d-(d-1)p}\|g\|_{L^\infty(\tau)}^p\big)^{\frac{2}{p}}dt\Big)^{\frac{1}{2}}\|\psi\|_{L^2(0,T;W^{1,q}(\Omega))}. \qquad (5.20)
\end{aligned}
$$

Combining (5.18)-(5.20) and recalling (5.15) we obtain

$$
\begin{aligned}
&\int_0^T\int_\Omega \nabla(y-Y_h)\cdot\vec{\Psi}dxdt \\
\leqslant\ & C\Big(\Big[\sum_{n=1}^{N}k_n\Big(\sum_{\tau\in\mathcal{T}_n^h}h_\tau^p\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|_{L^p(\tau)}^p+\sum_{l\in\mathcal{E}_n^h}h_l\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|_{L^p(l)}^p\Big)^{\frac{2}{p}}\Big]^{\frac{1}{2}} \\
&+\|y_0-Y_h^0\|_{L^p(\Omega)}+(\sum_{n=1}^{N}\int_{I_n}(\sum_{\tau\in\mathcal{M}_n^h}\|g-\pi_n g(\gamma(t_n),t)\|_{L^\infty(\tau)}^p+h_\tau^{d-(d-1)p}\|g\|_{L^\infty(\tau)}^p)^{\frac{2}{p}}dt)^{\frac{1}{2}} \\
&+(\sum_{n=1}^{N}k_n\|Y_h^n-Y_h^{n-1}\|_{W^{1,p}(\Omega)}^2)^{\frac{1}{2}}\Big)\|\vec{\Psi}\|_{L^2(0,T;L^q(\Omega)^d)}. \qquad (5.21)
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\|y-Y_h\|_{L^2(0,T;W^{1,p}(\Omega))} = \sup_{\vec{\Psi}\in L^2(0,T;L^q(\Omega)^d)\setminus\{0\}}\frac{\int_0^T\int_\Omega\nabla(y-Y_h)\cdot\vec{\Psi}dxdt}{\|\vec{\Psi}\|_{L^2(0,T;L^q(\Omega)^d)}} \\
\leqslant\ & C\Big(\Big[\sum_{n=1}^{N}k_n\Big(\sum_{\tau\in\mathcal{T}_n^h}h_\tau^p\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|_{L^p(\tau)}^p+\sum_{l\in\mathcal{E}_n^h}h_l\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|_{L^p(l)}^p\Big)^{\frac{2}{p}}\Big]^{\frac{1}{2}} \\
&+\|y_0-Y_h^0\|_{L^p(\Omega)}+\big(\sum_{n=1}^{N}\int_{I_n}(\sum_{\tau\in\mathcal{M}_n^h}\|g-\pi_n g(\gamma(t_n),t)\|_{L^\infty(\tau)}^p+h_\tau^{d-(d-1)p}\|g\|_{L^\infty(\tau)}^p)^{\frac{2}{p}}dt\big)^{\frac{1}{2}} \\
&+\big(\sum_{n=1}^{N}k_n\|Y_h^n-Y_h^{n-1}\|_{W^{1,p}(\Omega)}^2\big)^{\frac{1}{2}}\Big), \qquad (5.22)
\end{aligned}
$$

this completes the proof of the case $m = 0$. Similarly, in case that $m > 0$ we can derive

$$\begin{aligned}
F_3 &= \langle g\delta_{\gamma(t)}, \psi - \psi_I\rangle_{\Omega_T} = \sum_{n=1}^{N}\int_{I_n}\int_{\gamma(t)} g(x,t)(\psi-\psi_I)(x,t)dxdt \\
&= \sum_{n=1}^{N}\Big(\int_{I_n}\int_{\gamma(t)} g(x,t)(\psi-\pi_n\psi)(x,t)dxdt + \int_{I_n}\int_{\gamma(t)} g(x,t)\pi_n(\psi-\pi_h^n\psi)(x,t)dxdt\Big) \\
&\leqslant C\sum_{n=1}^{N}\int_{I_n}\Big(\|g-\pi_n g(\gamma(t_n),t)\|_{L^2(\gamma(t))}\|\psi-\pi_n\psi\|_{L^2(\gamma(t))} + \|g\|_{L^2(\gamma(t))}\|\pi_n(\psi-\pi_h^n\psi)\|_{L^2(\gamma(t))}\Big)dt \\
&\leqslant C\Big(\sum_{n=1}^{N}\int_{I_n}\|g-\pi_n g(\gamma(t_n),t)\|^2_{L^2(\gamma(t))} + \big(\sum_{\tau\in\mathcal{M}_n^h} h_\tau^{d-(d-1)p}\|g\|^p_{L^2(\gamma(t)\cap\tau)}\big)^{\frac{2}{p}}\Big)^{\frac{1}{2}}\|\psi\|_{L^2(0,T;W^{1,q}(\Omega))}.
\end{aligned}$$

Proceeding as above we can derive the desired results. □

Proof of Theorem 4.1: Let ψ be the solution of problem (2.1) with $f \in L^2(0,T;L^2(\Omega))$ and $\psi_I|_{I_n} := \pi_h^n\pi_n\psi$ be the corresponding space-time interpolation. Proceeding as in the proof of Theorem 3.2 we have

$$\begin{aligned}
\int_{\Omega_T}(y-Y_h)f dxdt &= -\sum_{n=1}^{N}\int_{I_n}(k_n^{-1}(Y_h^n - Y_h^{n-1}, \psi^{n-1}-\psi_I) + a(Y_h^n, \psi-\psi_I))dt \\
&\quad +(y_0 - Y_h^0, \psi(\cdot,0)) + \langle g\delta_{t_0}, \psi-\psi_I\rangle_{\Omega_T} \\
&:= \widetilde{E}_1 + \widetilde{E}_2 + \widetilde{E}_3. \qquad (5.23)
\end{aligned}$$

The estimates of $\widetilde{E}_1$ and $\widetilde{E}_2$ are similar to E_1 and E_2 as in the proof of Theorem 3.2. For $\widetilde{E}_3$ we have

$$\begin{aligned}
\widetilde{E}_3 &= \langle g\delta_{t_0}, \psi-\psi_I\rangle_{\Omega_T} = \sum_{n=1}^{N}\int_{I_n}\int_{\Omega} g(x,t)(\psi-\psi_I)(x,t)dxd\delta_{t_0} \\
&= \int_{\Omega} g(x,t_0)(\psi-\pi_n\psi)(x,t_0)dx + \int_{\Omega} g(x,t_0)\pi_n(\psi-\pi_h^n\psi)(x,t_0)dx \\
&\leqslant C\|g(\cdot,t_0)\|_{L^2(\Omega)}\|\psi-\pi_n\psi\|_{L^\infty(\mathcal{I};L^2(\Omega))} + C\|g(\cdot,t_0)\|_{L^2(\Omega)}\|\pi_n(\psi-\pi_h^n\psi)\|_{L^\infty(\mathcal{I};L^2(\Omega))} \\
&\leqslant C\Big(\sum_{\tau\in\mathcal{T}_{\mathcal{I}}^h}(k_{\mathcal{I}}\|g(\cdot,t_0)\|^2_{L^2(\tau)} + h_\tau^2\|g(\cdot,t_0)\|^2_{L^2(\tau)})\Big)^{\frac{1}{2}}(\|\psi\|_{L^\infty(0,T;H^1(\Omega))} + \|\psi\|_{H^1(0,T;L^2(\Omega))}), \qquad (5.24)
\end{aligned}$$

where we used the error estimates

$$\|\psi-\pi_n\psi\|_{L^\infty(\mathcal{I};L^2(\Omega))} \leqslant Ck_{\mathcal{I}}^{\frac{1}{2}}\|\psi\|_{H^1(0,T;L^2(\Omega))}, \quad \|\psi-\pi_h^n\psi\|_{L^\infty(I_n;L^2(\tau))} \leqslant Ch_\tau\|\psi\|_{L^\infty(I_n;H^1(\tau))}.$$

Combining (5.9), (5.10), (5.23) and (5.24) we proved

$$\begin{aligned}
&\|y-Y_h\|^2_{L^2(0,T;L^2(\Omega))} \\
\leqslant\ & C\sum_{n=1}^{N}k_n\Big(\sum_{\tau\in\mathcal{T}_n^h}h_\tau^4\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n} + \mathcal{A}Y_h^n\Big\|^2_{L^2(\tau)} + \sum_{l\in\mathcal{E}_n^h}h_l^3\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|^2_{L^2(l)}\Big) \\
&+C\|y_0-Y_h^0\|^2_{L^2(\Omega)} + C\sum_{\tau\in\mathcal{T}_{\mathcal{I}}^h}(k_{\mathcal{I}}\|g(\cdot,t_0)\|^2_{L^2(\tau)} + h_\tau^2\|g(\cdot,t_0)\|^2_{L^2(\tau)}) \\
&+C\sum_{n=1}^{N}k_n\|Y_h^n - Y_h^{n-1}\|^2_{L^2(\Omega)}, \qquad (5.25)
\end{aligned}$$

which completes the proof. □

Proof of Theorem 4.2: Similar to the proof of Theorem 3.3, given $\vec{\Psi} \in L^2(0,T;L^2(\Omega)^d)$, consider now the following problem: Find $\psi \in L^2(0,T;H_0^1(\Omega)) \cap H^1(0,T;H^{-1}(\Omega))$ such that

$$-(\psi_t, w)_{\Omega_T} + a(w,\psi)_{\Omega_T} = (\vec{\Psi}, \nabla w)_{\Omega_T} \quad \forall w \in L^2(0,T;H_0^1(\Omega)). \tag{5.26}$$

Then we have $\psi \in \mathcal{C}([0,T];L^2(\Omega))$ and

$$\|\psi\|_{L^2(0,T;H_0^1(\Omega))} + \|\partial_t \psi\|_{L^2(0,T;H^{-1}(\Omega))} + \|\psi\|_{\mathcal{C}([0,T];L^2(\Omega))} \leqslant C\|\vec{\Psi}\|_{L^2(0,T;L^2(\Omega)^d)}. \tag{5.27}$$

It follows from (2.6) and (5.26) and proceeding as in (5.17) we arrive at

$$\begin{aligned}
\int_0^T (\nabla(y - Y_h), \vec{\Psi}) dt &= -\sum_{n=1}^N \int_{I_n} (k_n^{-1}(Y_h^n - Y_h^{n-1}, \psi^{n-1} - \psi_I) + a(Y_h^n, \psi - \psi_I)) dt \\
&\quad + (y_0 - Y_h^0, \psi(\cdot,0)) + \langle \mu, \psi - \psi_I \rangle_{\Omega_T} \\
&:= \widetilde{F}_1 + \widetilde{F}_2 + \widetilde{F}_3, \qquad (5.28)
\end{aligned}$$

where in this case $\psi_I|_{I_n} := \pi_n \hat{\pi}_h^n \psi$ with $\hat{\pi}_h^n$ the Clément-type interpolation operator defined in Lemma 3.1. Now it remains to estimate the above terms one by one. Similar to the proof of (5.18) and (5.19), by setting $p = 2$ we can obtain

$$\begin{aligned}
\tilde{F}_1 &= -\sum_{n=1}^N \int_{I_n} \Big(k_n^{-1}(Y_h^n - Y_h^{n-1}, \psi^{n-1} - \pi_n \psi) + \sum_{l \in \mathcal{E}_n^h} \int_l \Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}} \Big] (\psi - \hat{\pi}_h^n \psi) ds \\
&\quad + \sum_{\tau \in \mathcal{T}_n^h} \int_\tau (k_n^{-1}(Y_h^n - Y_h^{n-1}) + \mathcal{A} Y_h^n)(\psi - \hat{\pi}_h^n \psi) dx \Big) dt \\
&\leqslant C \Big[\sum_{n=1}^N k_n \Big(\sum_{\tau \in \mathcal{T}_n^h} \|Y_h^n - Y_h^{n-1}\|_{H^1(\tau)}^2 + h_\tau^2 \Big\| \frac{Y_h^n - Y_h^{n-1}}{k_n} + \mathcal{A} Y_h^n \Big\|_{L^2(\tau)}^2 \\
&\quad + \sum_{l \in \mathcal{E}_n^h} h_l \Big\| \Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}} \Big] \Big\|_{L^2(l)}^2 \Big) \Big]^{\frac{1}{2}} \\
&\quad (\|\partial_t \psi\|_{L^2(0,T;H^{-1}(\Omega))} + \|\psi\|_{L^2(0,T;H^1(\Omega))}) \qquad (5.29)
\end{aligned}$$

as well as

$$\widetilde{F}_2 = (y_0 - Y_h^0, \psi(\cdot,0)) \leqslant C\|y_0 - Y_h^0\|_{L^2(\Omega)} \|\psi(\cdot,0)\|_{L^2(\Omega)}. \tag{5.30}$$

From the structure of $\widetilde{F}_3$ and the stability of Clément type interpolation we can deduce

$$\begin{aligned}
\widetilde{F}_3 &= \langle g\delta_{t_0}, \psi - \psi_I \rangle_{\Omega_T} = \sum_{n=1}^N \int_{I_n} \int_\Omega g(x,t)(\psi - \psi_I)(x,t) dx dt \delta_{t_0} \\
&= \int_\Omega g(x,t_0)(\psi - \pi_n \psi)(x,t_0) dx + \int_\Omega g(x,t_0) \pi_n (\psi - \hat{\pi}_h^n \psi)(x,t_0) dx \\
&\leqslant C\|g(\cdot,t_0)\|_{L^2(\Omega)} \|\psi - \pi_n \psi\|_{L^\infty(\mathcal{I};L^2(\Omega))} + C\|g(\cdot,t_0)\|_{L^2(\Omega)} \|\pi_n(\psi - \hat{\pi}_h^n \psi)\|_{L^\infty(\mathcal{I};L^2(\Omega))} \\
&\leqslant C\Big(\sum_{\tau \in \mathcal{T}_{\mathcal{I}}^h} \|g(\cdot,t_0)\|_{L^2(\tau)}^2 \Big)^{\frac{1}{2}} \|\psi\|_{L^\infty(0,T;L^2(\Omega))}. \qquad (5.31)
\end{aligned}$$

Combining (5.29)-(5.31) and recalling (5.27) we obtain

$$\begin{aligned}
&\int_0^T (\nabla(y-Y_h),\vec{\Psi})dt\\
\leqslant\ & C\Big(\Big[\sum_{n=1}^N k_n\Big(\sum_{\tau\in\mathcal{T}_n^h} h_\tau^2\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|^2_{L^2(\tau)}+\sum_{l\in\mathcal{E}_n^h} h_l\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|^2_{L^2(l)}\Big)\Big]^{\frac12}\\
&+C\|y_0-Y_h^0\|_{L^2(\Omega)}+C(\sum_{\tau\in\mathcal{T}_{\mathcal{I}}^h}\|g(\cdot,t_0)\|^2_{L^2(\tau)})^{\frac12}\\
&+C(\sum_{n=1}^N k_n\|Y_h^n-Y_h^{n-1}\|^2_{H^1(\Omega)})^{\frac12}\Big)\|\vec{\Psi}\|_{L^2(0,T;L^2(\Omega)^d)}. \qquad (5.32)
\end{aligned}$$

Therefore

$$\begin{aligned}
&\|y-Y_h\|_{L^2(0,T;H^1(\Omega))}=\sup_{\vec{\Psi}\in L^2(0,T;L^2(\Omega)^d)\setminus\{0\}}\frac{\int_0^T(\nabla(y-Y_h),\vec{\Psi})dt}{\|\vec{\Psi}\|_{L^2(0,T;L^2(\Omega)^d)}}\\
\leqslant\ & C\Big(\Big[\sum_{n=1}^N k_n\Big(\sum_{\tau\in\mathcal{T}_n^h} h_\tau^2\Big\|\frac{Y_h^n-Y_h^{n-1}}{k_n}+\mathcal{A}Y_h^n\Big\|^2_{L^2(\tau)}+\sum_{l\in\mathcal{E}_n^h} h_l\Big\|\Big[\frac{\partial Y_h^n}{\partial n_{\mathcal{A}}}\Big]\Big\|^2_{L^2(l)}\Big)\Big]^{\frac12}\\
&+\|y_0-Y_h^0\|_{L^2(\Omega)}+(\sum_{\tau\in\mathcal{T}_{\mathcal{I}}^h}\|g(\cdot,t_0)\|^2_{L^2(\tau)})^{\frac12}+(\sum_{n=1}^N k_n\|Y_h^n-Y_h^{n-1}\|^2_{H^1(\Omega)})^{\frac12}\Big), \qquad (5.33)
\end{aligned}$$

this completes the proof. □

References

[1] H. Amann. *Linear and Quasilinear Parabolic Problems. Volume I: abstract linear theory.* Birkhäuser Verlag, Basel, 1995.

[2] M. Andrle, F. Ben Belgacem, and A. El Badia. Identification of moving pointwise sources in an advenction-dispersion-reaction equation. *Inverse Problems*, 27:025007, 2011.

[3] R. Araya, E. Behrens, and R. Rodríguez. A posteriori error estimates for elliptic problems with Dirac delta source terms. *Numer. Math.*, 105:193–216, 2006.

[4] L. Boccardo, A. Dall'Aglio, T. Gallouët, and L. Orsina. Nonlinear parabolic equations with measure data. *J. Funct. Anal.*, 147:237–258, 1997.

[5] L. Boccardo and T. Gallouët. Non-linear elliptic and parabolic equations involving measure data. *J. Funct. Anal.*, 87:149–169, 1989.

[6] E. Casas. L^2 estimates for the finite element method for the Dirichlet problem with singular data. *Numer. Math.*, 47:627–632, 1985.

[7] E. Casas. Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM J. Control Optim.*, 35:1297–1327, 1997.

[8] E. Casas, C. Clason, and K. Kunisch. Parabolic control problems in measure spaces with sparse solutions. *SIAM J. Control Optim.*, 51:28–63, 2013.

[9] J. M. Cascon, L. Ferragut and M. I. Asensio. Space-time adaptive algorithm for the mixed parabolic problem. *Numer. Math.*, 103:367–392, 2006.

[10] C. Castro and E. Zuazua. Unique continuation and control for the heat equation from an oscillating lower dimensional manifold. *SIAM J. Control Optim.*, 43:1400–1434, 2004.

[11] L. Chen and C. S. Zhang. AFEM@matlab: a Matlab package of adaptive finite element methods. *Technique report*, 2006.

[12] L. Chen and C. S. Zhang. A coarsening algorithm on adaptive grids by newest vertex bisection and its applications. *J. Comput. Math.*, 28(6):767–789, 2010.

[13] Z. M. Chen and F. Jia. An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems. *Math. Comput.*, 73:1167–1193, 2004.

[14] P. G. Ciarlet. *The Finite Element Methods for Elliptic Problems.* North-Holland, 1978.

[15] Ph. Clément. Approximation by finite element functions using local regularization. *RAIRO Anal. Numer.*, 9: 77–84, 1975.

[16] M. Dauge. Neumann and mixed problems on curvilinear polyhedra. *Integral Equations Oper. Theory*, 15:227–261, 1992.

[17] J. Droniou and J. P. Raymond. Optimal pointwise control of semilinear parabolic equations. *Nonlinear Analysis*, 39:43–77, 2000.
[18] J. Elschner, J. Rehberg and G. Schmidt. Optimal regularity for elliptic transmission problems including $\mathcal{C}^1$ interfaces. *Interfaces Free Bound.*, 9:233–252, 2007.
[19] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems I: A linear model problem. *SIAM J. Numer. Anal.*, 28:43–77, 1991.
[20] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems II: Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. *SIAM J. Numer. Anal.*, 32:706–740, 1995.
[21] R. Glowinski and J. L. Lions. Exact and approximate controllability for distributed parameter systems. *Acta Numerica*, 159–333, 1996.
[22] W. Gong, and N. N. Yan. Finite element approximations of parabolic optimal control problems with controls acting on a lower dimensional manifold. *SIAM J. Numer. Anal.*, 54(2):1229–1262, 2016.
[23] W. Gong. Error estimates for finite element approximations of parabolic equations with measure data. *Math. Comput.*, 82:69–98, 2013.
[24] W. Gong, M. Hinze, and Z. J. Zhou. A priori error analysis for the finite element approximations of parabolic optimal control problems with pointwise control. *SIAM J. Control Optim.*, 52:97–119, 2014.
[25] P. Grisvard. *Singularities in Boundary Value Problems*. Springer-Verlag, Masson, Paris, Berlin, 1992.
[26] P. Grisvard. Singular behavior of elliptic problems in non-Hilbertian Sobolev spaces. *J. Math. Pures Appl.*, 74(9):3–33, 1995.
[27] C. Kreuzer, C. Möller, A. Schmidt, and K. G. Siebert. Design and convergence analysis for an adaptive discretisation of the heat equation. *IMA J. Numer. Anal.*, 32:1375–1403, 2012.
[28] K. Krumbiegel, and J. Rehberg. Second order sufficient optimality conditions for parabolic optimal control problems with pointwise state constraints. *SIAM J. Control Optim.*, 51:304–331, 2013.
[29] J. L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications*. Springer-Verlag, Berlin, 1972.
[30] W. B. Liu, H. P. Ma, T. Tang and N. N. Yan. A posteriori error estimates for discontinuous Galerkin time-stepping method for optimal control problems governed by parabolic equations. *SIAM J. Numer. Anal.*, 42:1032–1061, 2004.
[31] A. Martínez, C. Rodríguez, and M. E. Vázquez-Méndez. Theoretical and numerical analysis of an optimal control problem related to wastewater treatment. *SIAM J. Control Optim.*, 38:1534–1553, 2000.
[32] D. Meidner, R. Rannacher, and B. Vexler. A priori error estimates for finite element discretizations of parabolic optimization problems with pointwise state constraints in time. *SIAM J. Control Optim.*, 49(5):1961–1997, 2011.
[33] R. H. Nochetto, A. Schmidt and C. Verdi. A posteriori error estimation and adaptivity for degenerate parabolic problems. *Math. Comp.*, 69:1–24, 2000.
[34] M. Picasso. Adaptive finite elements for a linear parabolic problem. *Comput. Methods Appl. Mech. Engrg.*, 167:223–237, 1998.
[35] A. Schmidt, K. G. Siebert, C. J. Heine, D. Köster and O. Kriessl. ALBERTA: An adaptive hierarchical finite element toolbox. Available at: http://www.alberta-fem.de/. Versions 1.2 and 2.0, 2005.
[36] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comput.*, 54:483–493, 1990.
[37] R. Verfürth. *A Review of a Posteriori Error Estimation and Adaptive Mesh Refinement*. Wiley-Teubner, London, UK, 1996.
[38] PHG, http://lsec.cc.ac.cn/phg/ .