Hamburger Beiträge
zur Angewandten Mathematik

Error estimates for finite element approximations of parabolic equations with measure data

Wei Gong

Nr. 2011-19
October 2011

ERROR ESTIMATES FOR FINITE ELEMENT APPROXIMATIONS OF PARABOLIC EQUATIONS WITH MEASURE DATA

WEI GONG ◇

Abstract: In this paper we study the a priori error estimates for the finite element approximations of parabolic equations with measure data, especially we consider problems with separate measure data in time and space, respectively. The solutions of this kind of problems exhibit low regularities due to the existence of measure data, this introduces some difficulties in both theoretical and numerical analysis. For both cases we use standard piecewise linear and continuous finite elements for the space discretization and derive the a priori error estimates for the semi-discretization problems, while the backward Euler method is then used for time discretization and a priori error estimates for the fully discrete problems are also derived. Numerical results are provided at the end of the paper to confirm our theoretical findings.

Keywords: finite element method, parabolic equation, measure data, semi-discrete error estimates, fully discrete error estimates.

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

1. Introduction

The aim of this paper is to analyze the finite element approximations of parabolic equations with measure data. Let $\Omega_T = \Omega \times (0, T]$, $\Gamma_T = \partial\Omega \times (0, T]$, Ω is an open bounded domain in $\mathbb{R}^d$ ($d = 2$ or 3) with boundary $\Gamma = \partial\Omega$. We consider the following parabolic problems

$$\left\{\begin{aligned} \partial_t y + \mathcal{A}y &= \mu \quad \text{in } \Omega_T,\\ y &= 0 \quad \text{on } \Gamma_T,\\ y(\cdot, 0) &= y_0 \quad \text{in } \Omega, \end{aligned}\right. \tag{1.1}$$

where $\partial_t y = \frac{\partial y}{\partial t}$, the operator $\mathcal{A}$ is assumed to be a second order elliptic partial differential operator, $y_0 \in L^2(\Omega)$ and $T > 0$ are fixed.

Here we consider two kinds of problems with measure data. At first, we consider problem (1.1) with measure data in time, i.e., $\mu = g\omega$, g and ω are given functions such that $g \in \mathcal{C}([0, T]; L^2(\Omega))$ and $\omega \in \mathcal{M}[0, T]$, where $\mathcal{M}[0, T]$ is the space of the real and regular Borel measures in $[0, T]$, which can be defined as the dual space of $\mathcal{C}[0, T]$ with its natural norm

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

One of the most important applications for parabolic equations with measure data in time is related to the first order optimality conditions of parabolic optimal control problems with pointwise state constraints. Optimal control problems governed by parabolic PDE reads as:

$$\min \quad J(y, u) \quad = \tfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega_T)} + \tfrac{\alpha}{2}\|u\|^2_{L^2(\Omega_T)} \tag{1.2}$$

subject to

$$\left\{\begin{aligned} \partial_t y + \mathcal{A}y &= u \quad \text{in } \Omega_T,\\ y &= 0 \quad \text{on } \Gamma_T,\\ y(\cdot, 0) &= y_0 \quad \text{in } \Omega, \end{aligned}\right. \tag{1.3}$$

◇ LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. Email: `wgong@lsec.cc.ac.cn`, and Schwerpunkt Optimierung und Approximation, Universität Hamburg, Bundesstrasse 55, 20146, Hamburg, Germany.

where y denotes the state variable and u denotes the control (see, e.g., [18]). If we impose state constraints pointwise in time, for example (see, e.g., Example 2.3 in [6] and [22])

$$\int_\Omega g(x,t,y(x,t))dx \leqslant b(t) \quad \forall\, t \in [0,T] \tag{1.4}$$

with given function $b(t)$, then the adjoint state p associated to the first order optimality conditions satisfies

$$\begin{cases} -\partial_t p + \mathcal{A}^* p = y - y_d + \mu_{\Omega_T} & \text{in } \Omega_T, \\ p = \mu_{\Sigma_T} & \text{on } \Sigma_T, \\ p(\cdot,T) = \mu_T & \text{in } \Omega \end{cases} \tag{1.5}$$

in the sense of distributions (see, e.g., [6], [10] and [22]). In general, the Lagrange multiplier μ associated to the state constraints for parabolic optimal control problems with pointwise state constraints belongs to $\mathcal{M}(\overline{\Omega}_T)$, where $\mathcal{M}(\overline{\Omega}_T)$ is the space of regular Borel measures on $\overline{\Omega}_T$, $\mu_{\Omega_T} := \mu|_{\Omega_T}$, $\mu_{\Gamma_T} := \mu|_{\Gamma_T}$ and $\mu_T := \mu|_{\overline{\Omega}\times\{T\}}$. However, in this case where only pointwise in time state constraints are imposed, the Lagrange multiplier μ associated to the state constraints (1.4) appears to be a measure only in time, and can be decomposed as $\mu = g\omega$, g and ω are given functions such that $g \in \mathcal{C}([0,T];L^2(\Omega))$ and $\omega \in \mathcal{M}[0,T]$. Thus the associated (to the state) adjoint equation exhibits the similar structure of (1.1).

Then, we consider problems with measure data in space, i.e., $\mu = g\omega$, g and ω are given functions such that $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $\omega \in \mathcal{M}(\Omega)$. Here $\mathcal{M}(\Omega)$ is the space of the real and regular Borel measures on Ω, and can be defined as the dual space of $\mathcal{C}(\overline{\Omega})$ with its natural norm

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

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 μ (see [5]). This kind of problems also arise in other different applications, for instance, modeling of acoustic monopoles, transport equations for effluent discharge in aquatic media (see [1]). In the design and management of waste-water treatment systems, mainly the disposal of sea outfalls discharging polluting effluent from a sewerage system (see [21] for details), the problem can be formulated as an optimal control problem with pointwise state and control constraints and pointwise control, while the governing state equation of which is of form (1.1) with measure data in space

$$\begin{cases} y_t + \mathcal{A}y = \displaystyle\sum_{i=1}^m u_i(t)\delta_{X_i} & \text{in } \Omega_T, \\ y(x,t) = 0 & \text{on } \Gamma_T, \\ y(x,0) = y_0(x) & \text{in } \Omega, \end{cases} \tag{1.6}$$

where δ_{X_i} represents the Dirac measure concentrated at X_i, $i = 1, 2, \cdots, m$. Some other kind of optimal control problems with state equation of form (1.6), for example, problems with pointwise control, can be found in, e.g., [11], [14] and [25].

The existence of solutions for quasi-linear elliptic and parabolic equations involving measure data has been studied by Boccardo and Gallouët in [4], Casas studied linear parabolic problems and improved the results of [4] by exploiting the linearity of the equation in [6]. The finite element method for elliptic equation with measure data has been extensively studied (see, e.g., [2], [5], [28] and [29]). Casas gave an optimal error estimate of order $O(h^{2-\frac{d}{2}})$ in [5], where h is the mesh size of space triangulation and d is the dimension of Ω. Araya et al. obtained a posteriori error estimates for elliptic problems with Dirac delta source terms in [1]. However, there seems to be no such kind of contributions to finite element approximations of parabolic equations with measure data. To the best of our knowledge this paper is among the few contributions on this topic.

In this paper we study the finite element approximations of parabolic equations with measure data, especially we consider problems with separate measure data in time and space, respectively.

We use standard piecewise linear and continuous finite elements for the space discretization and derive the a priori error estimates for the semidiscretization problems, while the backward Euler method is then used for time discretization and a priori error estimates for the fully discrete problems are also derived.

We denote by k the step size in the temporal discretization and by h the maximum element size of the spatial mesh. Then the main results of this paper are as follows. For parabolic equations with measure data in time, we obtain the following estimates of the error between the solution y of the continuous problem and the solution y_h of the semidiscretization one:

$$\|y - y_h\|_{L^2(0,T;L^2(\Omega))} \leqslant Ch\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]} + \|y_0\|_{0,\Omega}\Big)$$

and the estimates of the error between the solution y of the continuous problem and the solution Y_h of the fully discrete one:

$$\|y - Y_h\|_{L^2(0,T;L^2(\Omega))} \leqslant C(h + k^{\frac{1}{2}})\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]} + \|y_0\|_{0,\Omega}\Big).$$

For parabolic equations with measure data in space, we obtain the following estimates of the error between the solution y of the continuous problem and the solution y_h of the semidiscretization one:

$$\|y - y_h\|_{L^2(0,T;L^2(\Omega))} \leqslant Ch^{2-\frac{d}{2}}\Big(\|g\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)} + \|y_0\|_{0,\Omega}\Big)$$

and the estimates of the error between the solution y of the continuous problem and the solution Y_h of the fully discrete one:

$$\|y - Y_h\|_{L^2(0,T;L^2(\Omega))} \leqslant C(h^{2-\frac{d}{2}} + k^{\frac{1}{2}})\Big(\|g\|_{H^{\frac{1}{2}}(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)} + \|y_0\|_{0,\Omega}\Big).$$

Numerical results are provided at the end of the paper to confirm our theoretical findings.

The rest of this paper is organized as follows. In Section 2 we give some notations and present the parabolic equations with measure data in time and space, and analyze the existence results for the unique solution. In Section 3 we establish the continuous time semi-discrete finite element approximation schemes for above two kinds of problems and derive a priori error estimates. Then the fully discrete finite element approximation based on the backward Euler method is introduced and a priori estimate for the discretization error is derived in Section 4. We also carry out some numerical experiments in Section 5 to confirm our theoretical findings. At the end of the paper we give a conclusion and discuss the future work.

2. Parabolic equations with measure data

2.1. **Notation.** Assume that $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3 is a convex polygonal or polyhedral domain, or domain with a $\mathcal{C}^{1,1}$ boundary. We denote by $W^{m,p}(\Omega)$ the usual Sobolev space of order $m \geqslant 0$, $1 \leqslant p < \infty$ with norm $\|\cdot\|_{m,p,\Omega}$ and seminorm $|\cdot|_{m,p,\Omega}$, and the standard modification for $p = \infty$. For $p = 2$ we denote $W^{m,p}(\Omega)$ by $H^m(\Omega)$ and $\|\cdot\|_{m,\Omega} = \|\cdot\|_{m,2,\Omega}$, which is a Hilbert space. Note that $H^0(\Omega) = L^2(\Omega)$ and $H^1_0(\Omega) = \{v \in H^1(\Omega) : \ v = 0 \text{ on } \partial\Omega\}$.

For $p \in [1, \infty)$, the interval $[0, T] \subset \mathbb{R}$ and the Banach space A with norm $\|\cdot\|_A$, we denote by $L^p(0, T; A)$ the set of measurable functions $y : [0, T] \to A$ such that $\int_0^T \|y\|_A^p dt \leqslant \infty$. The norm on $L^p(0, T; A)$ is defined by

$$\|y(t)\|_{L^p(0,T;A)} = \begin{cases} \Big(\displaystyle\int_0^T \|y\|_A^p dt\Big)^{\frac{1}{p}} & 1 \leqslant p < \infty, \\ \operatorname*{ess\ sup}_{t\in[0,T]} \|y(t)\|_A & p = \infty. \end{cases}$$

We denote $\mathcal{D}(\Omega_T)$ the set of $\mathcal{C}^\infty(\Omega_T)$ functions with compact support in Ω_T. Let $H^{s,r}(\Omega_T) = L^2(0, T; H^s(\Omega)) \cap H^r(0, T; L^2(\Omega))$ equipped with the norm

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

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 [19]). We also set

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

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

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

and

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

respectively. 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 $a_0 \in L^\infty(\Omega)$, $a_0(x,t) \geqslant 0$ for all $(x,t) \in \Omega_T$, a_{ij} $(1 \leqslant i,j \leqslant d)$ is Lipschitz continuous on Ω_T and satisfies the following uniform ellipticity condition:

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

Moreover, $\partial_{\boldsymbol{n}_A} = \sum\limits_{i,j=1}^{n} a_{ij}\partial_{x_j} n_i$ and $\boldsymbol{n}$ is the unit outer normal to $\partial\Omega$. We will denote by $\mathcal{A}^*$ the adjoint operator of A:

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

In addition, c and C denote generic positive constants.

We introduce the following bilinear forms associated with $\mathcal{A}$ on Ω and Ω_T:

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

and

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

For $f \in L^2(\Omega_T)$, we assume that ϕ and ψ are the solutions of following forward and backward in time parabolic problems:

$$\begin{cases} \partial_t\phi + \mathcal{A}\phi = f & \text{in } \Omega_T, \\ \phi = 0 & \text{on } \Gamma_T, \\ \phi(0) = 0 & \text{in } \Omega \end{cases} \tag{2.1}$$

and

$$\begin{cases} -\partial_t\psi + \mathcal{A}^*\psi = f & \text{in } \Omega_T, \\ \psi = 0 & \text{on } \Gamma_T, \\ \psi(T) = 0 & \text{in } \Omega. \end{cases} \tag{2.2}$$

Then the following standard stability estimates can be found in, e.g. [19].

Lemma 2.1. *Let ϕ and ψ denote the solutions of problem (2.1) and (2.2), respectively. Then for $v=\phi$ or $v=\psi$ there holds $v\in L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(0,T;L^2(\Omega))\hookrightarrow \mathcal{C}([0,T];H^1(\Omega))$ and satisfies*

$$\|v\|_{L^2(0,T;H^2(\Omega))}+\|v_t\|_{L^2(0,T;L^2(\Omega))}\leqslant C\|f\|_{0,0}, \tag{2.3}$$

and

$$\|\phi(T)\|_{1,\Omega}\leqslant C\|f\|_{L^2(0,T;L^2(\Omega))},\quad \|\psi(0)\|_{1,\Omega}\leqslant C\|f\|_{L^2(0,T;L^2(\Omega))}. \tag{2.4}$$

2.2. **Parabolic equations with measure data in time.** At first, we consider the following parabolic equations with measure data in time:

$$\begin{cases} \partial_t y+\mathcal{A}y=\mu=g\omega & \text{in } \Omega_T,\\ y=0 & \text{on } \Gamma_T,\\ y(\cdot,0)=y_0 & \text{in } \Omega, \end{cases} \tag{2.5}$$

where $g\in\mathcal{C}([0,T];L^2(\Omega))$ and $\omega\in\mathcal{M}[0,T]$. The weak solution of problems (2.5) can be defined by transposition techniques (see Lions and Magenes [19]). In the following theorem we will give the results on the existence and uniqueness as well as regularity of the solution to problem (2.5).

Theorem 2.2. *With the assumption that $\mu=g\omega$, g and ω are given functions such that $g\in\mathcal{C}([0,T];L^2(\Omega))$ and $\omega\in\mathcal{M}[0,T]$, problem (2.5) admits a unique solution $y\in L^2(0,T;H_0^1(\Omega))\cap L^\infty(0,T;L^2(\Omega))$ such that*

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

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

$$\|y\|_{L^2(0,T;H_0^1(\Omega))}+\|y\|_{L^\infty(0,T;L^2(\Omega))}\leqslant C\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}+\|y_0\|_{0,\Omega}\Big). \tag{2.7}$$

Here

$$\langle\mu,v\rangle_{\Omega_T}=\int_{\Omega_T}v\,d\mu=\int_0^T\Big(\int_\Omega g(x,t)v(x,t)dx\Big)d\omega(t),\quad \forall\ v\in\mathcal{C}([0,T];L^2(\Omega)).$$

Proof. The proof follows the idea of [6], here we sketch the proof for completeness. Since the problem is linear, it suffices to consider either $y_0\equiv 0$ or $\mu\equiv 0$.

If $\mu\equiv 0$, $y_0\in L^2(\Omega)$, it is obvious that problem (2.5) admits a unique solution $y\in L^2(0,T;H_0^1(\Omega))\cap L^\infty(0,T;L^2(\Omega))$ satisfying (see Lions and Magenes [19])

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

Now we suppose $y_0\equiv 0$, let $\{\omega_n\}_n\subset\mathcal{C}[0,T]$ be the sequence converging weakly-* to ω in $\mathcal{M}[0,T]$ and satisfy

$$\|\omega_n\|_{L^1[0,T]}\leqslant\|\omega\|_{\mathcal{M}[0,T]}.$$

Let y_n be the solutions of

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

then we have $y_n\in L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(0,T;L^2(\Omega))$. For $f\in\mathcal{D}(\Omega_T)$, let ψ be the solution of problem

$$\begin{cases} -\partial_t\psi+\mathcal{A}^*\psi=f & \text{in } \Omega_T,\\ \psi=0 & \text{on } \Gamma_T,\\ \psi(\cdot,T)=0 & \text{in } \Omega, \end{cases} \tag{2.9}$$

thus we have $\psi \in \mathcal{C}(\overline{\Omega}_T)$ from the regularity theory of parabolic equation. Then from (2.8) we have

$$\begin{aligned}\int_{\Omega_T} f y_n dxdt &= \int_{\Omega_T} (-\partial_t\psi + \mathcal{A}^*\psi) y_n dxdt \\ &= \int_{\Omega_T} g\omega_n \psi dxdt \\ &\leqslant C\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega_n\|_{L^1[0,T]}\|\psi\|_{\mathcal{C}([0,T];L^2(\Omega))} \\ &\leqslant C\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\|\psi\|_{\mathcal{C}([0,T];L^2(\Omega))}.\end{aligned}$$

Standard estimates give (see, e.g., [19])

$$\|\psi\|_{\mathcal{C}([0,T];L^2(\Omega))} \leqslant C\|f\|_{L^1([0,T];L^2(\Omega))} \tag{2.10}$$

and

$$\|\psi\|_{\mathcal{C}([0,T];L^2(\Omega))} \leqslant C\|f\|_{L^2(0,T;H^{-1}(\Omega))}. \tag{2.11}$$

We can conclude from (2.10) that the solution sequence $\{y_n\}_n$ is bounded in the space $L^\infty(0,T;L^2(\Omega))$, while $\{y_n\}_n$ is also bounded in the space $L^2(0,T;H_0^1(\Omega))$ from (2.11). Thus we can take a subsequence such that $y_n \to y$ weakly in $L^2(0,T;H_0^1(\Omega)) \cap L^\infty(0,T;L^2(\Omega))$ and (2.7) is satisfied.

Then we prove (2.6). Let $\psi \in W(0,T)$ and $\psi(\cdot,T) = 0$, multiplying (2.8) by ψ and integrating by parts we have

$$\begin{aligned}&\int_0^T \Big(\int_\Omega g(x,t)\psi(x,t)dx\Big)\omega_n(t)dt \\ &= -\int_{\Omega_T} y_n\partial_t\psi dxdt + \int_{\Omega_T}\Big(\sum_{i,j=1}^d a_{ij}\partial_{x_i}y_n\partial_{x_j}\psi + a_0 y_n\psi\Big)dxdt,\end{aligned} \tag{2.12}$$

passing to the limit in (2.12) we get (2.6).

Finally, we note that uniqueness holds since the only solution for zero data of (2.5) is $y = 0$. □

2.3. **Parabolic equations with measure data in space.** Now we turn to the following parabolic equations with measure data in space:

$$\begin{cases} \partial_t y + \mathcal{A}y = \mu = g\omega & \text{in } \Omega_T, \\ y = 0 & \text{on } \Gamma_T, \\ y(0) = y_0 & \text{in } \Omega, \end{cases} \tag{2.13}$$

where $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $\omega \in \mathcal{M}(\Omega)$. Similarly, the weak solution of problems (2.13) can be defined by transposition techniques. The following theorem gives the results concerning the existence, uniqueness and regularity of the solution to problem (2.13).

Theorem 2.3. *With the assumption that $y_0 \in L^2(\Omega)$, $\mu = g\omega$, g and ω are given functions such that $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $\omega \in \mathcal{M}(\Omega)$, problem (2.13) admits a unique solution $y \in L^2(0,T;L^2(\Omega))$ in the sense that*

$$-(y,\partial_t v)_{\Omega_T} + (y,\mathcal{A}^*v)_{\Omega_T} = \langle\mu,v\rangle_{\Omega_T} + (y_0,v(\cdot,0)) \quad \forall\, v \in X(0,T) \tag{2.14}$$

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

$$\langle\mu,v\rangle_{\Omega_T} = \int_{\Omega_T} v d\mu = \int_\Omega\Big(\int_0^T g(x,t)v(x,t)dt\Big)d\omega(x) \quad \forall\, v \in L^2(0,T;\mathcal{C}(\overline{\Omega})).$$

Besides, there exist a constant C only depending on data, such that

$$\|y\|_{L^2(0,T;L^2(\Omega))} \leqslant C\Big(\|g\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)} + \|y_0\|_{0,\Omega}\Big). \tag{2.15}$$

Moreover, we have $y \in L^1(0,T;W^{1,p}(\Omega)) \cap \mathcal{C}([0,T];W^{1,q}(\Omega)')$ and $\partial_t y \in L^1(0,T;W^{1,q}(\Omega)')$ and

$$\|y\|_{L^1(0,T;W^{1,p}(\Omega))} \leqslant C\Big(\|g\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)} + \|y_0\|_{0,\Omega}\Big), \tag{2.16}$$

where $p \in [1, \frac{d}{d-1})$ and q is the conjugate number of p such that $\frac{1}{p}+\frac{1}{q}=1$.

Proof. Similarly as in the proof of Theorem 2.2, we assume that $y_0 \in L^2(\Omega)$ and $g \in L^2(0,T;\mathcal{C}(\overline{\Omega}))$. For $\mu \equiv 0$ the assertion is obvious. We set $y_0 \equiv 0$ and let $\{\omega_n\}_n \subset \mathcal{C}(\overline{\Omega})$ be the sequence converging weakly-* to ω in $\mathcal{M}(\Omega)$ and satisfy

$$\|\omega_n\|_{L^1(\Omega)} \leqslant \|\omega\|_{\mathcal{M}(\Omega)}.$$

Let y_n be the solutions of (2.8) with righthand side $g\omega_n$, then we have $y_n \in X(0,T)$. For $f \in \mathcal{D}(\Omega_T)$, let ψ be the solution of problem (2.9), thus we have $\psi \in \mathcal{C}(\overline{\Omega}_T)$ from the regularity theory of parabolic equation. We deduce from (2.8) that

$$\begin{aligned}\int_{\Omega_T} f y_n dxdt &= \int_{\Omega_T} (-\partial_t \psi + \mathcal{A}^* \psi) y_n dxdt \\ &= \int_{\Omega_T} g\omega_n \psi dxdt \\ &\leqslant C\|g\|_{L^2(0,T;\mathcal{C}(\overline{\Omega}))}\|\omega_n\|_{L^1(\Omega)}\|\psi\|_{L^2(0,T;\mathcal{C}(\overline{\Omega}))} \\ &\leqslant C\|g\|_{L^2(0,T;\mathcal{C}(\overline{\Omega}))}\|\omega\|_{\mathcal{M}(\Omega)}\|\psi\|_{L^2(0,T;\mathcal{C}(\overline{\Omega}))}.\end{aligned}$$

From embedding theorem we have $L^2(0,T;H^2(\Omega)) \hookrightarrow L^2(0,T;\mathcal{C}(\overline{\Omega}))$. Standard estimates give (see, e.g., [19])

$$\|\psi\|_{L^2(0,T;H^2(\Omega))} \leqslant C\|f\|_{L^2(0,T;L^2(\Omega))}. \tag{2.17}$$

We can conclude from (2.17) that the solution sequence $\{y_n\}_n$ is bounded in the space $L^2(0,T;L^2(\Omega))$. Thus we can take a subsequence such that $y_n \to y$ weakly in $L^2(0,T;L^2(\Omega))$ and (2.15) is satisfied. The rest of the proof is standard.

Furthermore, the second part of this theorem has been proved in [6], see also [21]. Actually, Theorem 6.3 in [6] implies the existence of a unique solution $y \in L^1(0,T;W^{1,p}(\Omega))$ for all $p \in [1, \frac{d}{d-1})$ and $\partial_t y \in L^1(0,T;W^{1,q}(\Omega)')$ in the sense of (2.14), such that (2.16) is satisfied, hence we have $y \in \mathcal{C}([0,T];W^{1,q}(\Omega)')$ after a modification on a set of zero measure. □

Remark 2.4. *If $d=1$, the function $\mu = g(x,t)\omega(x)$ belongs to $L^2(0,T;H^{-1}(\Omega))$, this property implies in turn that (see [14])*

$$y \in L^2(0,T;H^1(\Omega)) \cap \mathcal{C}([0,T];L^2(\Omega)), \quad \frac{\partial y}{\partial t} \in L^2(0,T;H^{-1}(\Omega)).$$

However, we don't proceed with this issue and consider only cases when $d=2$ or 3.

3. Error estimates for the continuous time Galerkin approximations

Let us consider the continuous time finite element approximations of the problems (2.6) and (2.14). To this aim, we consider a family of triangulation $\mathscr{T}^h$ of $\overline{\Omega}$, such that $\overline{\Omega} = \bigcup_{\tau\in\mathscr{T}^h}\overline{\tau}$. We suppose that $\overline{\Omega}$ is the union of the elements of $\mathscr{T}^h$ so that element edges lying on the boundary may be curved. This triangulation is supposed to be regular in the usual sense. For each element $\tau \in \mathscr{T}^h$ we associate two parameters $\rho(\tau)$ and $\sigma(\tau)$, where $\rho(\tau)$ denotes the diameter of the element τ and $\sigma(\tau)$ is the supremum of the diameters of all circles contained in τ. Define the size of the mesh by $h = \max_{\tau\in\mathscr{T}^h} \rho(\tau)$. We suppose that the following regularity assumptions are satisfied: There exists a positive constant C such that

$$\frac{\rho(\tau)}{\sigma(\tau)} \leqslant C, \quad \frac{h}{\rho(\tau)} \leqslant C \tag{3.1}$$

hold for all $\tau \in \mathscr{T}^h$ and all $h>0$.

Here we consider only n-simplex elements, as they are among the most widely used ones. Associated with $\mathscr{T}^h$ is a finite dimensional subspace V^h of $\mathcal{C}(\overline{\Omega})$, such that $\chi|_\tau$ are polynomials of order m ($m \geqslant 1$) for $\forall\, \chi \in V^h$ and $\tau \in \mathscr{T}^h$. Here we only consider piecewise linear elements, i.e., $m = 1$. We also set $V_0^h = V^h \cap H_0^1(\Omega)$.

Note that the regular assumption (3.1) guarantees the following inverse properties for $v_h \in V^h$:

$$\|v_h\|_{s,\Omega} \leqslant Ch^{l-s}\|v_h\|_{l,\Omega} \quad 0 \leqslant l \leqslant s \leqslant 1 \tag{3.2}$$

and

$$\|v_h\|_{0,\infty,\Omega} \leqslant Ch^{-\frac{d}{2}}\|v_h\|_{0,\Omega}. \tag{3.3}$$

Let $\Pi_h : \mathcal{C}(\overline{\Omega}) \to V^h$ denote the standard Lagrange interpolation operator, then interpolation error estimate implies that for $y \in H^2(\Omega)$ (see, e.g., [8])

$$\|y - \Pi_h y\|_{0,\Omega} + h\|y - \Pi_h y\|_{1,\Omega} \leqslant Ch^2\|y\|_{2,\Omega}, \tag{3.4}$$

and

$$\|y - \Pi_h y\|_{0,\infty,\Omega} \leqslant Ch^{2-\frac{d}{2}}\|y\|_{2,\Omega}. \tag{3.5}$$

Let $\mathcal{P}_h$ be the $L^2(\Omega)$-projection operator defined from $L^2(\Omega)$ to V^h:

$$(\mathcal{P}_h y, v_h) = (y, v_h) \quad \forall\, v_h \in V^h \tag{3.6}$$

and $\mathcal{R}_h : H_0^1(\Omega) \to V_0^h$ denote the Ritz projection operator defined as

$$a(\mathcal{R}_h y, v_h) = a(y, v_h) \quad \forall\, v_h \in V_0^h. \tag{3.7}$$

Then we have the following error estimates (see, e.g., [8] and [26])

Lemma 3.1. *Let $\mathcal{P}_h$ and $\mathcal{R}_h$ be the L^2-projection operator and Ritz projection operator defined above. Then there holds:*

$$\|y - \mathcal{P}_h y\|_{-1,\Omega} + h\|y - \mathcal{P}_h y\|_{0,\Omega} \leqslant Ch^2\|y\|_{1,\Omega}, \tag{3.8}$$

$$\|y - \mathcal{R}_h y\|_{0,\Omega} + h\|y - \mathcal{R}_h y\|_{1,\Omega} \leqslant Ch^2\|y\|_{2,\Omega}. \tag{3.9}$$

Moreover, we have

$$\|y - \mathcal{R}_h y\|_{0,\infty,\Omega} \leqslant Ch^{2-\frac{d}{2}}\|y\|_{2,\Omega}. \tag{3.10}$$

Proof. Here we only prove $\|y - \mathcal{P}_h y\|_{-1,\Omega}$. From the definition of L^2-projection we have

$$\begin{aligned}
\|y - \mathcal{P}_h y\|_{-1,\Omega} &= \sup_{v \in H^1(\Omega)} \frac{(y - \mathcal{P}_h y, v)}{\|v\|_{1,\Omega}} \\
&= \sup_{v \in H^1(\Omega)} \frac{(y - \mathcal{P}_h y, v - \mathcal{P}_h v)}{\|v\|_{1,\Omega}} \\
&\leqslant \sup_{v \in H^1(\Omega)} \frac{Ch^2\|y\|_{1,\Omega}\|v\|_{1,\Omega}}{\|v\|_{1,\Omega}} \\
&\leqslant Ch^2\|y\|_{1,\Omega}.
\end{aligned}$$

□

Since the solutions of problems (2.6) and (2.14) have low regularities, it seems to be natural to estimate the error between the solutions of continuous problem and semidiscretization problem under the norm $L^2(0,T;L^2(\Omega))$. To achieve this we need to use duality argument. Thus, we introduce the semi-discrete finite element approximation of the backward parabolic problem (2.2):

$$\begin{cases} -(\partial_t \psi_h, v_h)_{\Omega_T} + a(\psi_h, v_h)_{\Omega_T} = (f, v_h)_{\Omega_T} \quad \forall\, v_h \in V_0^h, \\ (\psi_h(T), w_h) = 0 \quad \forall\, w_h \in V_0^h, \end{cases} \tag{3.11}$$

where $\psi_h(t) \in H^1(0,T;V_0^h)$.

At first, we need to derive the error estimates for the solutions of the backward parabolic problem (2.2) and its semidiscretization approximation (3.11) under the norms $L^2(0,T;L^\infty(\Omega))$ and $L^\infty(0,T;L^2(\Omega))$, which will play a crucial role in the derivation of our main results.

Lemma 3.2. *Let $\psi \in X(0,T) \hookrightarrow \mathcal{C}([0,T];H^1(\Omega))$ and $\psi_h \in H^1(0,T;V_0^h)$ be the solutions of problem (2.2) and (3.11), respectively. Then we have the following uniformly in time and space error estimate:*

$$\|\psi - \psi_h\|_{L^\infty(0,T;L^2(\Omega))} \leqslant Ch\Big(\|\psi\|_{L^2(0,T;H^2(\Omega))} + \|\psi_t\|_{L^2(0,T;L^2(\Omega))}\Big) \tag{3.12}$$

and

$$\|\psi - \psi_h\|_{L^2(0,T;L^\infty(\Omega))} \leqslant Ch^{2-\frac{d}{2}}\Big(\|\psi\|_{L^2(0,T;H^2(\Omega))} + \|\psi_t\|_{L^2(0,T;L^2(\Omega))}\Big). \tag{3.13}$$

Proof. Actually, we need to prove the error estimates under the regularity conditions stated in this lemma, and the techniques are different from the standard proof for the semidiscrete error estimates which requires higher regularity, see, e.g., [30]. Using the similar arguments as in [7] it is not difficult to prove following a priori error estimates for the backward parabolic equations

$$\begin{aligned} &\|\psi(t) - \psi_h(t)\|_{0,\Omega} + \|\psi - \psi_h\|_{L^2(0,T;H^1(\Omega))} \\ \leqslant\ & Ch\Big(\|\psi\|_{L^2(0,T;H^2(\Omega))} + \|\psi_t\|_{L^2(0,T;L^2(\Omega))}\Big) \end{aligned} \tag{3.14}$$

and

$$\|\psi - \psi_h\|_{L^2(0,T;L^2(\Omega))} \leqslant Ch^2\Big(\|\psi\|_{L^2(0,T;H^2(\Omega))} + \|\psi_t\|_{L^2(0,T;L^2(\Omega))}\Big). \tag{3.15}$$

The key ingredient of the proof is the introduction of L^2-projection instead of the Ritz-projection used in the other literatures.

Error estimate (3.12) is a direct consequence of (3.14). To prove (3.13), let $\Pi_h\psi$ be the piecewise linear interpolation of ψ defined above, then from (3.5) we have

$$\begin{aligned} \|\psi - \psi_h\|_{L^2(0,T;L^\infty(\Omega))} &\leqslant \|\psi - \Pi_h\psi\|_{L^2(0,T;L^\infty(\Omega))} + \|\Pi_h\psi - \psi_h\|_{L^2(0,T;L^\infty(\Omega))} \\ &\leqslant Ch^{2-\frac{d}{2}}\|\psi\|_{L^2(0,T;H^2(\Omega))} + Ch^{-\frac{d}{2}}\|\Pi_h\psi - \psi_h\|_{L^2(0,T;L^2(\Omega))} \\ &\leqslant Ch^{2-\frac{d}{2}}\|\psi\|_{L^2(0,T;H^2(\Omega))} + Ch^{-\frac{d}{2}}\|\psi - \psi_h\|_{L^2(0,T;L^2(\Omega))}, \end{aligned} \tag{3.16}$$

where we have used the standard interpolation error estimate and inverse estimate (3.3). This together with (3.15) implies (3.13).

□

3.1. **Finite element approximations to parabolic equations with measure data in time.** We now turn to defining the continuous time finite element approximation scheme for problems (2.6). Based on the weak form stated in Theorem 2.2, we can define the following semi-discrete finite element approximation of (2.6):

$$-(y_h, \partial_t v_h)_{\Omega_T} + a(y_h, v_h)_{\Omega_T} = \langle \mu, v_h\rangle_{\Omega_T} + (y_0^h, v_h(\cdot,0)) \ \ \forall\ v_h \in H^1(0,T;V_0^h) \tag{3.17}$$

with $v_h(\cdot,T) = 0$, where $y_h(t) \in L^2(0,T;V_0^h)$, and $y_0^h \in V_0^h$ is an approximation of y_0. We set $y_0^h = \mathcal{P}_h y_0$ be the L^2-projection of y_0. Here

$$\langle \mu, v_h\rangle_{\Omega_T} = \int_{\overline{\Omega}_T} v_h d\mu = \int_0^T (\int_\Omega g(x,t)v_h(x)dx)d\omega(t) \ \ \forall\ v_h \in V^h.$$

Now we are in a position to state our main result of this subsection, i.e., the estimates of the error between the solution y of the continuous problem (2.6) and the solution y_h of the semidiscretization one (3.17).

Theorem 3.3. *Assume that $\mu = g\omega$, g and ω are given functions such that $g \in \mathcal{C}([0,T];L^2(\Omega))$ and $\omega \in \mathcal{M}[0,T]$. Let $y \in L^2(0,T;H_0^1(\Omega)) \cap L^\infty(0,T;L^2(\Omega))$ and $y_h \in L^2(0,T;V_0^h)$ be the solutions of problem (2.6) and (3.17), respectively. Then we have the following error estimate:*

$$\|y - y_h\|_{L^2(0,T;L^2(\Omega))} \leqslant Ch\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]} + \|y_0\|_{0,\Omega}\Big). \tag{3.18}$$

Proof. Let ψ be the solution of problem (2.2) with $f \in L^2(0,T;L^2(\Omega))$. Then from (2.6), (3.17) and orthogonality property we have

$$\begin{aligned}
\int_{\Omega_T}(y-y_h)f dxdt &= \int_0^T\int_\Omega (y-y_h)(-\partial_t\psi+\mathcal{A}^*\psi)dxdt \\
&= (y,-\partial_t\psi)_{\Omega_T}+a(y,\psi)_{\Omega_T}+(y_h,\partial_t\psi)_{\Omega_T}-a(y_h,\psi)_{\Omega_T} \\
&= \langle\mu,\psi\rangle_{\Omega_T}+(y_0,\psi(0))+(y_h,\partial_t\psi_h)_{\Omega_T}-a(y_h,\psi_h)_{\Omega_T} \\
&= \langle\mu,\psi\rangle_{\Omega_T}+(y_0,\psi(0))-\langle\mu,\psi_h\rangle_{\Omega_T}-(\mathcal{P}_hy_0,\psi_h(0)) \\
&= \int_{\overline{\Omega}_T}(\psi-\psi_h)d\mu+(y_0,\psi(0)-\psi_h(0)) \\
&= \int_0^T\Big(\int_\Omega g(x,t)(\psi-\psi_h)dx\Big)d\omega(t)+(y_0,\psi(0)-\psi_h(0)) \\
&\leqslant C\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}+\|y_0\|_{0,\Omega}\Big)\|\psi-\psi_h\|_{L^\infty(0,T;L^2(\Omega))}.
\end{aligned}$$

Lemma 2.1 and Lemma 3.2 yield

$$\begin{aligned}
\int_{\Omega_T}(y-y_h)f dxdt &\leqslant C\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}+\|y_0\|_{0,\Omega}\Big)\|\psi-\psi_h\|_{L^\infty(0,T;L^2(\Omega))} \\
&\leqslant Ch\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}+\|y_0\|_{0,\Omega}\Big)\Big(\|\psi\|_{L^2(0,T;H^2(\Omega))} \\
&\quad +\|\partial_t\psi\|_{L^2(0,T;L^2(\Omega))}\Big) \\
&\leqslant Ch\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}+\|y_0\|_{0,\Omega}\Big)\|f\|_{L^2(0,T;L^2(\Omega))},
\end{aligned}$$

then from the definition of $L^2(\Omega_T)$ norm we have

$$\begin{aligned}
\|y-y_h\|_{L^2(0,T;L^2(\Omega))} &= \sup_{f\in L^2(0,T;L^2(\Omega)),f\neq 0}\frac{(f,y-y_h)_{\Omega_T}}{\|f\|_{L^2(0,T;L^2(\Omega))}} \\
&\leqslant Ch\Big(\|y_0\|_{0,\Omega}+\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\Big),
\end{aligned}$$

which completes the proof. □

3.2. **Finite element approximations to parabolic equations with measure data in space.** Similarly, based on the results of Theorem 2.3 we can define the semi-discrete finite element approximation of (2.14) as follows:

$$-(y_h,\partial_t v_h)_{\Omega_T}+a(y_h,v_h)_{\Omega_T}=\langle\mu,v_h\rangle_{\Omega_T}+(y_0^h,v_h(\cdot,0))\quad\forall\, v_h\in H^1(0,T;V_0^h) \tag{3.19}$$

with $v_h(\cdot,T)=0$, where $y_h(t)\in L^2(0,T;V_0^h)$, and $y_0^h\in V_0^h$ is an approximation of y_0. We set $y_0^h=P_hy_0$ be the L^2-projection of y_0. Here

$$\langle\mu,v_h\rangle_{\Omega_T}=\int_{\overline{\Omega}_T}v_hd\mu=\int_0^T(\int_\Omega g(x,t)v_h(x)d\omega(x))dt\quad\forall\, v_h\in V^h.$$

With above preparations now we can state our main result in the follwing theorem.

Theorem 3.4. *Assume that g and ω are given functions such that $g\in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $\omega\in\mathcal{M}(\Omega)$, $y_0\in L^2(\Omega)$. Let $y\in L^2(0,T;L^2(\Omega))$ and $y_h\in L^2(0,T;V_0^h)$ be the solutions of problem (2.13) and (3.19), respectively. Then we have the following error estimate:*

$$\|y-y_h\|_{L^2(0,T;L^2(\Omega))}\leqslant Ch^{2-\frac{d}{2}}\Big(\|g\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)}+\|y_0\|_{0,\Omega}\Big). \tag{3.20}$$

Proof. Let ψ be the solution of problem (2.2) with $f \in L^2(0,T;L^2(\Omega))$. Then from (2.14) , (3.19), orthogonality property and proceeding as in the proof of Theorem 3.3 we have

$$\begin{aligned}
\int_{\Omega_T} (y-y_h) f dxdt &= \int_0^T \int_\Omega (y-y_h)(-\partial_t \psi + \mathcal{A}^* \psi) dxdt \\
&= \int_{\overline{\Omega}_T} (\psi - \psi_h) d\mu + (y_0, \psi(\cdot,0) - \psi_h(\cdot,0)) \\
&= \int_0^T \Big(\int_\Omega g(x,t)(\psi - \psi_h) d\omega(x) \Big) dt + (y_0, \psi(\cdot,0) - \psi_h(\cdot,0)) \\
&\leqslant C\Big(\|g\|_{L^2(0,T;L^\infty(\Omega))} \|\omega\|_{\mathcal{M}(\Omega)} \|\psi - \psi_h\|_{L^2(0,T;L^\infty(\Omega))} \\
&\quad + \|y_0\|_{0,\Omega} \|\psi - \psi_h\|_{\mathcal{C}([0,T];L^2(\Omega))} \Big).
\end{aligned}$$

It now follows from Lemma 2.1 and Lemma 3.2 that

$$\begin{aligned}
\int_{\Omega_T} (y-y_h) f dxdt &\leqslant C\Big(h^{2-\frac{d}{2}} \|g\|_{L^2(0,T;L^\infty(\Omega))} \|\omega\|_{\mathcal{M}(\Omega)} + Ch \|y_0\|_{0,\Omega} \Big) \\
&\quad \Big(\|\psi\|_{L^2(0,T;H^2(\Omega))} + \|\partial_t \psi\|_{L^2(0,T;L^2(\Omega))} \Big) \\
&\leqslant Ch^{2-\frac{d}{2}} \Big(\|g\|_{L^2(0,T;L^\infty(\Omega))} \|\omega\|_{\mathcal{M}(\Omega)} + \|y_0\|_{0,\Omega} \Big) \|f\|_{L^2(0,T;L^2(\Omega))},
\end{aligned}$$

which proves (3.20) from the definition of $L^2(\Omega_T)$ norm. □

Remark 3.5. *The a priori error estimate we obtained in Theorem 3.4, which is of order $O(h^{2-\frac{d}{2}})$, seems to be optimal compared with the results presented in* [5], *where finite element approximation for elliptic equations with measure data is studied and a priori error estimate of order $O(h^{2-\frac{d}{2}})$ is derived.*

4. Error estimates for fully discrete finite element approximations

We next consider the fully discrete approximations for above semidiscrete problems by using the backward Euler scheme in time. We consider a partitioning of the time interval $\overline{I} = [0,T]$ as

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

with subintervals $I_i = (t_{i-1}, t_i]$ of size k_i 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 $k|_{I_i} = k_i$ for $i = 1, 2, \cdots, N$. For $i = 1, 2, \cdots, N$, construct the finite element spaces $V_i^h \in H^1(\Omega)$ (similar to V^h) with the mesh $\mathscr{T}_i^h$. For simplicity we consider equal partition in time, i.e., $k_i \equiv k$, where k denotes the time step size, and the same finite element space on each time step. For our error analysis in the following we set $k = O(h^d)$ throughout the paper.

4.1. **Fully discrete approximations of parabolic equations with measure data in time.** 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 (3.17) is to find $Y_h^i \in V_0^h$, $i = 1, 2, \cdots, N$, such that

$$\left\{ \begin{aligned} &(\frac{Y_h^i - Y_h^{i-1}}{k}, w_h) + a(Y_h^i, w_h) = \langle \mu, w_h \rangle_{I_i}, \ \forall\, w_h \in V_0^h, \ \ i = 1, \cdots, N, \\ &\qquad\qquad\qquad Y_h^0(x) = y_0^h(x), \quad x \in \Omega. \end{aligned} \right. \tag{4.1}$$

Here

$$\langle \mu, v_h \rangle_{I_i} = \frac{1}{k} \int_{\Omega \times (t_{i-1}, t_i]} v_h d\mu = \frac{1}{k} \int_{t_{i-1}}^{t_i} (\int_\Omega g(x,t) v_h(x) dx) d\omega(t), \quad \forall\, v_h \in V^h.$$

In the following we denote Y_h the fully discrete finite element approximation of y, it is obvious that Y_h is piecewise constant in time and piecewise linear in space on each time interval.

Then we can derive following stability estimate for numerical scheme (4.1) of problem (2.5).

Lemma 4.1. *Let $Y_h^i \in V_0^h,\ i = 1, 2, \cdots, N$ be the solutions of fully discrete scheme (4.1), $y_0^h = \mathcal{P}_h y_0$ and assume that $k \leqslant Ch^d$, then there exists a constant C independent of h, k and the data (g, ω, y_0) such that*

$$\sum_{i=1}^{N} \|Y_h^i - Y_h^{i-1}\|_{0,\Omega}^2 + k\|Y_h^N\|_{1,\Omega}^2 \leqslant C\Big(\|y_0\|_{0,\Omega}^2 + \|g\|_{L^\infty(0,T;L^2(\Omega))}^2 \|\omega\|_{\mathcal{M}[0,T]}^2\Big) \tag{4.2}$$

and

$$\|Y_h^N\|_{0,\Omega}^2 + \sum_{i=1}^{N} k\|Y_h^i\|_{1,\Omega}^2 \leqslant C\Big(\|y_0\|_{0,\Omega}^2 + \|g\|_{L^\infty(0,T;L^2(\Omega))}^2 \|\omega\|_{\mathcal{M}[0,T]}^2\Big). \tag{4.3}$$

Proof. Let $w_h = k(Y_h^i - Y_h^{i-1})$ in (4.1) we get

$$(Y_h^i - Y_h^{i-1}, Y_h^i - Y_h^{i-1}) + ka(Y_h^i, Y_h^i - Y_h^{i-1}) \quad = \quad k\langle \mu, Y_h^i - Y_h^{i-1}\rangle_{I_i},$$

thus we have

$$\begin{aligned}
&\|Y_h^i - Y_h^{i-1}\|_{0,\Omega}^2 + k\|Y_h^i\|_{1,\Omega}^2 \\
\leqslant\ & ka(Y_h^i, Y_h^{i-1}) + \int_{t_{i-1}}^{t_i} (g(t), Y_h^i - Y_h^{i-1}) d\omega(t) \\
\leqslant\ & \frac{1}{2}k\|Y_h^i\|_{1,\Omega}^2 + \frac{1}{2}k\|Y_h^{i-1}\|_{1,\Omega}^2 + \|Y_h^i - Y_h^{i-1}\|_{0,\Omega} \int_{t_{i-1}}^{t_i} \|g(t)\|_{0,\Omega} d\omega(t) \\
\leqslant\ & \frac{1}{2}k\|Y_h^i\|_{1,\Omega}^2 + \frac{1}{2}k\|Y_h^{i-1}\|_{1,\Omega}^2 + C\Big(\int_{t_{i-1}}^{t_i} \|g(t)\|_{0,\Omega} d\omega(t)\Big)^2 + \frac{1}{2}\|Y_h^i - Y_h^{i-1}\|_{0,\Omega}^2. \qquad (4.4)
\end{aligned}$$

Summing the above equations over i from 1 to N we obtain

$$\begin{aligned}
\sum_{i=1}^{N} \|Y_h^i - Y_h^{i-1}\|_{0,\Omega}^2 + k\|Y_h^N\|_{1,\Omega}^2 \quad &\leqslant \quad k\|\mathcal{P}_h y_0\|_{1,\Omega}^2 + C\sum_{i=1}^{N}\Big(\int_{t_{i-1}}^{t_i} \|g(t)\|_{0,\Omega} d\omega(t)\Big)^2 \\
&\leqslant \quad k\|\mathcal{P}_h y_0\|_{1,\Omega}^2 + C\Big(\int_0^T \|g(t)\|_{0,\Omega} d\omega(t)\Big)^2 \\
&\leqslant \quad C\|y_0\|_{0,\Omega}^2 + C\|g\|_{L^\infty(0,T;L^2(\Omega))}^2 \|\omega\|_{\mathcal{M}[0,T]}^2, \qquad (4.5)
\end{aligned}$$

where we have used the inverse estimate $k\|\mathcal{P}_h y_0\|_{1,\Omega}^2 \leqslant kh^{-2}\|\mathcal{P}_h y_0\|_{0,\Omega}^2 \leqslant C\|y_0\|_{0,\Omega}^2$. This proves (4.2). Similarly, by setting $v_h = kY_h^i$ in (4.1) we can prove (4.3). $\square$

With the above preparations we are ready to estimate the error between the solution y of continuous problem (2.6) and the solution Y_h of the fully discrete problem (4.1), which is the main result of this paper. Instead of the standard approaches based on Ritz-projection, see, e.g, [30], we use duality argument to carry out the error analysis, and the stability results stated in Lemma 4.1 for numerical scheme (4.1) play an important role.

Theorem 4.2. *Assume that $\mu = g\omega$, g and ω are given functions such that $g \in \mathcal{C}([0,T]; L^2(\Omega))$ and $\omega \in \mathcal{M}[0,T]$. 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 be the solution of problem (4.1), then we have*

$$\|y - Y_h\|_{L^2(0,T;L^2(\Omega))} \leqslant C(h + k^{\frac{1}{2}})\Big(\|g\|_{L^\infty(0,T;L^2(\Omega))} \|\omega\|_{\mathcal{M}[0,T]} + \|y_0\|_{0,\Omega}\Big). \tag{4.6}$$

Proof. Let ψ be the solution of problem (2.2) with $f \in L^2(0,T;L^2(\Omega))$. It follows from (2.6) that

$$\begin{aligned}
\int_{\Omega_T}(y-Y_h)f dxdt &= \int_0^T\int_\Omega (y-Y_h)(-\partial_t\psi+\mathcal{A}^*\psi)dxdt \\
&= -(y,\partial_t\psi)_{\Omega_T}+a(y,\psi)_{\Omega_T}+\sum_{n=1}^N\int_{I_n}((Y_h^n,\partial_t\psi)-a(Y_h^n,\psi))dt \\
&= \langle\mu,\psi\rangle_{\Omega_T}+(y_0,\psi(0))+\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n,\psi^n-\psi^{n-1})-a(Y_h^n,\psi))dt \\
&= \langle\mu,\psi\rangle_{\Omega_T}+(y_0,\psi(0))-\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1})+a(Y_h^n,\psi))dt \\
&\quad +(Y_h^N,\psi^N)-(Y_h^0,\psi(0)) \\
&= -\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1})+a(Y_h^n,\psi))dt \\
&\quad +\langle\mu,\psi\rangle_{\Omega_T}+(y_0-Y_h^0,\psi(0))
\end{aligned}$$

with $\psi^n := \psi(:,t_n)$. Note that from (4.1) we have

$$\sum_{n=1}^N(k^{-1}(Y_h^n-Y_h^{n-1},\overline{R_h}\psi)+a(Y_h^n,\overline{R_h}\psi))=\sum_{n=1}^N\langle\mu,\overline{R_h}\psi\rangle_{I_n},$$

where $\overline{R_h}\psi \in V^h$ is defined on I_n as follows:

$$\overline{R_h}\psi=\overline{R_h}\psi^n=\frac{1}{k}\int_{I_n}R_h\psi(\cdot,t)dt, \quad n>0, \tag{4.7}$$

and $\overline{R_h}\psi^N = R_h\psi(T)$. Here and in what follows we denote $\overline{\psi}$ the average of ψ in I_n as defined in (4.7) for all $\psi \in L^1(I_n)$. It is straightforward to see that

$$\int_{I_n}(\psi-\overline{\psi})dt=0. \tag{4.8}$$

Therefore we have

$$\begin{aligned}
\int_{\Omega_T}(y-Y_h)f dxdt &= \langle\mu,\psi\rangle_{\Omega_T}-\sum_{n=1}^N\int_{I_n}\langle\mu,\overline{R_h}\psi\rangle_{I_n} \\
&\quad -\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1})+a(Y_h^n,\overline{\psi}))dt \\
&\quad +(y_0-Y_h^0,\psi(0))+\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\overline{R_h}\psi)+a(Y_h^n,\overline{R_h}\psi))dt \\
&= -\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1}-\overline{R_h}\psi)+a(Y_h^n,\overline{\psi}-\overline{R_h}\psi))dt \\
&\quad +\langle\mu,\psi\rangle_{\Omega_T}-\sum_{n=1}^N\int_{I_n}\langle\mu,\overline{R_h}\psi\rangle_{I_n}+(y_0-Y_h^0,\psi(0)) \\
&= E_1+E_2+E_3. \qquad (4.9)
\end{aligned}$$

Note that we have

$$
\begin{aligned}
|E_2| &= \Big|\langle \mu, \psi\rangle_{\Omega_T} - \sum_{n=1}^{N} \int_{I_n} \langle \mu, \overline{R_h}\psi\rangle_{I_n}\Big| \\
&= \Big|\sum_{n=1}^{N} \int_{t_{n-1}}^{t_n} \Big(\int_\Omega g(x,t)(\psi - \overline{R_h}\psi)(x)dx\Big) d\omega(t)\Big| \\
&\leqslant C\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\|\psi - \overline{R_h}\psi\|_{L^\infty(0,T;L^2(\Omega))}. \qquad (4.10)
\end{aligned}
$$

Note that

$$
\begin{aligned}
\|\psi - \overline{R_h}\psi\|_{L^\infty(0,T;L^2(\Omega))} &\leqslant \|\psi - \overline{\psi}\|_{L^\infty(0,T;L^2(\Omega))} + \|\overline{\psi} - \overline{R_h}\psi\|_{L^\infty(0,T;L^2(\Omega))} \\
&\leqslant Ck^{\frac{1}{2}}\|\psi\|_{H^1(0,T;L^2(\Omega))} + Ch\|\overline{\psi}\|_{L^\infty(0,T;H^1(\Omega))} \\
&\leqslant C(k^{\frac{1}{2}} + h)\|\psi\|_{2,1}, \qquad (4.11)
\end{aligned}
$$

where standard error estimates were used in above equation. Thus

$$
\begin{aligned}
|E_2| &\leqslant C(k^{\frac{1}{2}} + h)\|\psi\|_{2,1}\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]} \\
&\leqslant C(k^{\frac{1}{2}} + h)\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\|f\|_{L^2(0,T;L^2(\Omega))}. \qquad (4.12)
\end{aligned}
$$

We also have

$$
\begin{aligned}
|E_3| &= |(y_0 - Y_h^0, \psi(0))| \leqslant \|y_0 - \mathcal{P}_h y_0\|_{-1,\Omega}\|\psi(0)\|_{1,\Omega} \\
&\leqslant Ch\|y_0\|_{0,\Omega}\|f\|_{L^2(0,T;L^2(\Omega))}. \qquad (4.13)
\end{aligned}
$$

Then it remains to estimate E_1. Application of the Cauchy-Schwarz inequality gives

$$
|E_1| \leqslant F_1 \cdot F_2, \qquad (4.14)
$$

where

$$
F_1 = \Big(\sum_{n=1}^{N}(\|Y_h^n - Y_h^{n-1}\|_{0,\Omega}^2 + ka(Y_h^n, Y_h^n))\Big)^{\frac{1}{2}}
$$

and

$$
F_2 = \Big(\sum_{n=1}^{N}(\|\psi^{n-1} - \overline{R_h}\psi\|_{0,\Omega}^2 + ka(\overline{\psi} - \overline{R_h}\psi, \overline{\psi} - \overline{R_h}\psi))\Big)^{\frac{1}{2}}.
$$

Lemma 4.1 gives

$$
F_1 \leqslant C\Big(\|y_0\|_{0,\Omega} + \|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\Big). \qquad (4.15)
$$

Standard error estimates yield

$$
\begin{aligned}
\|\psi^{n-1} - \overline{R_h}\psi\|_{0,\Omega} &\leqslant \|\psi^{n-1} - \overline{\psi}\|_{0,\Omega} + \|\overline{\psi} - \overline{R_h}\psi\|_{0,\Omega} \\
&\leqslant \|\psi^{n-1} - \overline{\psi}\|_{0,\Omega} + Ch^2\|\overline{\psi}\|_{2,\Omega} \qquad (4.16)
\end{aligned}
$$

and

$$
\|\psi^{n-1} - \overline{\psi}\|_{0,\Omega} \leqslant k^{\frac{1}{2}}\|\psi_t\|_{L^2(t_{n-1},t_n;L^2(\Omega))}. \qquad (4.17)
$$

From Lemma 3.1 we also have that

$$
a(\overline{\psi} - \overline{R_h}\psi, \overline{\psi} - \overline{R_h}\psi) \leqslant Ch^2\|\overline{\psi}\|_{2,\Omega}. \qquad (4.18)
$$

It is straightforward to show that

$$
\|\overline{\psi}\|_{2,\Omega} \leqslant k^{-\frac{1}{2}}\|\psi\|_{L^2(t_{n-1},t_n;H^2(\Omega))}. \qquad (4.19)
$$

Then from (4.16)-(4.19) we can conclude that

$$
\begin{aligned}
|F_2| &\leqslant C\Big(\sum_{n=1}^{N}((h^4+kh^2)\|\overline{\psi}\|_{2,\Omega}^2+k\|\psi_t\|^2_{L^2(t_{n-1},t_n;L^2(\Omega))})\Big)^{\frac{1}{2}}\\
&\leqslant C\Big(\sum_{n=1}^{N}((h^4+kh^2)k^{-1}\|\overline{\psi}\|^2_{L^2(t_{n-1},t_n;H^2(\Omega))}+k\|\psi_t\|^2_{L^2(t_{n-1},t_n;L^2(\Omega))})\Big)^{\frac{1}{2}}\\
&\leqslant C(h+k^{\frac{1}{2}})\|\psi\|_{2,1}, \qquad (4.20)
\end{aligned}
$$

thus

$$
\begin{aligned}
|E_1| &\leqslant C(h+k^{\frac{1}{2}})\|\psi\|_{2,1}\Big(\|y_0\|_{0,\Omega}+\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\Big)\\
&\leqslant C(h+k^{\frac{1}{2}})\|f\|_{L^2(0,T;L^2(\Omega))}\Big(\|y_0\|_{0,\Omega}+\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\Big). \qquad (4.21)
\end{aligned}
$$

It follows from (4.9), (4.12), (4.13) and (4.21) that

$$
\begin{aligned}
\|y-Y_h\|_{L^2(0,T;L^2(\Omega))} &= \sup_{f\in L^2(0,T;L^2(\Omega)),f\neq 0}\frac{(f,y-Y_h)_{\Omega_T}}{\|f\|_{L^2(0,T;L^2(\Omega))}}\\
&\leqslant C(h+k^{\frac{1}{2}})\Big(\|y_0\|_{0,\Omega}+\|g\|_{L^\infty(0,T;L^2(\Omega))}\|\omega\|_{\mathcal{M}[0,T]}\Big),
\end{aligned}
$$

which completes the proof. □

Remark 4.3. *In Theorem 3.3 we obtain a priori error estimates of order $O(h)$ for the semidiscretization finite element approximation of parabolic equations with measure data in time, and the same result with respect to space discretization is derived for fully discrete approximation in Theorem 4.2. The order $O(h)$ seems to be optimal in view of the regularity of solution y, which belongs to $L^2(0,T;H^1(\Omega))$ as presented in Theorem 2.2. The convergence order with respect to time discretization is $O(k^{\frac{1}{2}})$, which should also be optimal.*

4.2. **Fully discrete approximations of parabolic equations with measure data in space.** This subsection is devoted to the fully discrete approximations of parabolic equations with measure data in space. The fully discrete approximation scheme of (3.19) is to find $Y_h^i\in V_0^h,\ i=1,2,\cdots,N$, such that

$$
\begin{cases}
(\dfrac{Y_h^i-Y_h^{i-1}}{k},w_h)+a(Y_h^i,w_h)=\langle\mu,w_h\rangle_{I_i}, \quad \forall\ w_h\in V_0^h,\ \ i=1,\cdots,N,\\
\qquad Y_h^0(x)=y_0^h(x)\ \ x\in\Omega.
\end{cases} \qquad (4.22)
$$

Here

$$
\langle\mu,v_h\rangle_{I_i}=\frac{1}{k}\int_{\Omega\times(t_{i-1},t_i]}v_h d\mu=\frac{1}{k}\int_{t_{i-1}}^{t_i}(\int_\Omega g(x,t)v_h(x)d\omega(x))dt\ \ \forall\ v_h\in V^h.
$$

Also we denote Y_h the fully discrete finite element approximation of y.

Then we can derive following stability estimate for numerical scheme (4.22) of problem (2.13).

Lemma 4.4. *Assume that g and ω are given functions such that $g\in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $\omega\in\mathcal{M}(\Omega)$, $y_0\in L^2(\Omega)$. Let $Y_h^i\in V_0^h,\ i=1,2,\cdots,N$ be the solutions of fully discrete scheme (4.22), $y_0^h=\mathcal{P}_h y_0$ and assume that $k\leqslant Ch^d$, then there exists a constant C independent of h,k and the data (g,ω,y_0) such that*

$$
\sum_{i=1}^{N}\|Y_h^i-Y_h^{i-1}\|^2_{0,\Omega}+k\|Y_h^N\|^2_{1,\Omega}\leqslant C\Big(\|y_0\|^2_{0,\Omega}+\|g(t)\|^2_{L^2(0,T;L^\infty(\Omega))}\|\omega\|^2_{\mathcal{M}(\Omega)}\Big) \qquad (4.23)
$$

and

$$
\|Y_h^N\|^2_{0,\Omega}+\sum_{i=1}^{N}k\|Y_h^i\|^2_{1,\Omega}\leqslant C\Big(\|y_0\|^2_{0,\Omega}+\|g(t)\|^2_{L^2(0,T;L^\infty(\Omega))}\|\omega\|^2_{\mathcal{M}(\Omega)}\Big). \qquad (4.24)
$$

Proof. The proof is similar to Lemma 4.1. Let $w_h = k(Y_h^i - Y_h^{i-1})$ in (4.22) we get

$$(Y_h^i - Y_h^{i-1}, Y_h^i - Y_h^{i-1}) + ka(Y_h^i, Y_h^i - Y_h^{i-1}) \quad = \quad k\langle \mu, Y_h^i - Y_h^{i-1}\rangle_{I_i},$$

thus we have

$$\begin{aligned}
& \|Y_h^i - Y_h^{i-1}\|_{0,\Omega}^2 + k\|Y_h^i\|_{1,\Omega}^2 \\
\leqslant\ & ka(Y_h^i, Y_h^{i-1}) + \int_{t_{i-1}}^{t_i} (\int_\Omega g(x,t)(Y_h^i - Y_h^{i-1}) d\omega(x)) dt \\
\leqslant\ & \frac{1}{2}k\|Y_h^i\|_{1,\Omega}^2 + \frac{1}{2}k\|Y_h^{i-1}\|_{1,\Omega}^2 + \int_{t_{i-1}}^{t_i} \|Y_h^i - Y_h^{i-1}\|_{0,\infty,\Omega}\|g(t)\|_{0,\infty,\Omega}\|\omega\|_{\mathcal{M}(\Omega)} dt \\
\leqslant\ & \frac{1}{2}k\|Y_h^i\|_{1,\Omega}^2 + \frac{1}{2}k\|Y_h^{i-1}\|_{1,\Omega}^2 + C\|g(t)\|_{L^2(t_{i-1},t_i;L^\infty(\Omega))}^2\|\omega\|_{\mathcal{M}(\Omega)}^2 \\
& + \frac{1}{2}\|Y_h^i - Y_h^{i-1}\|_{0,\Omega}^2, \qquad (4.25)
\end{aligned}$$

where we have used the following inverse estimate:

$$\begin{aligned}
\sqrt{k}\|Y_h^i - Y_h^{i-1}\|_{0,\infty,\Omega} \quad & \leqslant \quad C\sqrt{k}h^{-\frac{d}{2}}\|Y_h^i - Y_h^{i-1}\|_{0,\Omega} \\
& \leqslant \quad C\|Y_h^i - Y_h^{i-1}\|_{0,\Omega}.
\end{aligned}$$

Summing the above equations over i from 1 to N and using inverse estimate we get

$$\begin{aligned}
\sum_{i=1}^N \|Y_h^i - Y_h^{i-1}\|_{0,\Omega}^2 + k\|Y_h^N\|_{1,\Omega}^2 \quad & \leqslant \quad k\|Y_h^0\|_{1,\Omega}^2 + C\sum_{i=1}^N \|g(t)\|_{L^2(t_{i-1},t_i;L^\infty(\Omega))}^2\|\omega\|_{\mathcal{M}(\Omega)}^2 \\
& \leqslant \quad k\|\mathcal{P}_h y_0\|_{1,\Omega}^2 + C\|g(t)\|_{L^2(0,T;L^\infty(\Omega))}^2\|\omega\|_{\mathcal{M}(\Omega)}^2 \\
& \leqslant \quad C\|y_0\|_{0,\Omega}^2 + C\|g(t)\|_{L^2(0,T;L^\infty(\Omega))}^2\|\omega\|_{\mathcal{M}(\Omega)}^2, \qquad (4.26)
\end{aligned}$$

which proves (4.23). Similarly, by setting $v_h = kY_h^i$ in (4.22) we can prove (4.24).

□

Now we are in a position to estimate the error between the solutions of problem (2.14) and (4.22), which is one of the main results of this paper.

Theorem 4.5. *Assume that g and ω are given functions such that $g \in L^2(0,T;\mathcal{C}(\overline{\Omega})) \cap H^{\frac{1}{2}}(0,T;L^\infty(\Omega))$ and $\omega \in \mathcal{M}(\Omega)$, $y_0 \in L^2(\Omega)$. Let $y \in L^2(0,T;L^2(\Omega))$ be the solution of problem (2.14), and Y_h be the solution of problem (4.22), then we have*

$$\|y - Y_h\|_{L^2(0,T;L^2(\Omega))} \leqslant C(h^{2-\frac{d}{2}} + k^{\frac{1}{2}})\Big(\|g\|_{H^{\frac{1}{2}}(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)} + \|y_0\|_{0,\Omega}\Big). \qquad (4.27)$$

Proof. Again, we use duality argument to prove this theorem. Let ψ be the solution of problem (2.2) with $f \in L^2(0,T;L^2(\Omega))$. Note that $\psi = 0$ on $\partial\Omega$, $\psi^N = \psi(T) = 0$, it follows from (2.14) that

$$
\begin{aligned}
\int_{\Omega_T}(y-Y_h)f dxdt &= \int_0^T\int_\Omega (y-Y_h)(-\partial_t\psi+\mathcal{A}^*\psi)dxdt \\
&= -(y,\partial_t\psi)_{\Omega_T}+(y,\mathcal{A}^*\psi)_{\Omega_T}+\sum_{n=1}^N\int_{I_n}((Y_h^n,\partial_t\psi)-a(Y_h^n,\psi))dt \\
&= \langle\mu,\psi\rangle_{\Omega_T}+(y_0,\psi(\cdot,0))+\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n,\psi^n-\psi^{n-1})-a(Y_h^n,\psi))dt \\
&= \langle\mu,\psi\rangle_{\Omega_T}+(y_0,\psi(\cdot,0))-\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1})+a(Y_h^n,\psi))dt \\
&\quad +(Y_h^N,\psi^N)-(Y_h^0,\psi(\cdot,0)) \\
&= -\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1})+a(Y_h^n,\psi))dt \\
&\quad +\langle\mu,\psi\rangle_{\Omega_T}+(y_0-Y_h^0,\psi(\cdot,0)).
\end{aligned}
$$

Note that from (4.22) we have

$$
\sum_{n=1}^N(k^{-1}(Y_h^n-Y_h^{n-1},\overline{R_h}\psi)+a(Y_h^n,\overline{R_h}\psi))=\sum_{n=1}^N\langle\mu,\overline{R_h}\psi\rangle_{I_n},
$$

where $\overline{R_h}\psi\in V_0^h$ is defined in (4.7). Thus

$$
\begin{aligned}
\int_{\Omega_T}(y-Y_h)f dxdt &= \langle\mu,\psi\rangle_{\Omega_T}-\sum_{n=1}^N\int_{I_n}\langle\mu,\overline{R_h}\psi\rangle_{I_n} \\
&\quad +\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\overline{R_h}\psi)+a(Y_h^n,\overline{R_h}\psi))dt \\
&\quad -\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1})+a(Y_h^n,\overline{\psi}))dt+(y_0-Y_h^0,\psi(0)) \\
&= -\sum_{n=1}^N\int_{I_n}(k^{-1}(Y_h^n-Y_h^{n-1},\psi^{n-1}-\overline{R_h}\psi)+a(Y_h^n,\overline{\psi}-\overline{R_h}\psi))dt \\
&\quad +\langle\mu,\psi\rangle_{\Omega_T}-\sum_{n=1}^N\int_{I_n}\langle\mu,\overline{R_h}\psi\rangle_{I_n}+(y_0-Y_h^0,\psi(0)) \\
&= \widetilde{E}_1+\widetilde{E}_2+\widetilde{E}_3. \qquad (4.28)
\end{aligned}
$$

From (4.8) we deduce

$$
\begin{aligned}
|\widetilde{E}_2| &= \Big|\langle\mu,\psi\rangle_{\Omega_T}-\sum_{n=1}^{N}\int_{I_n}\langle\mu,\overline{R_h}\psi\rangle_{I_n}\Big|\\
&= \Big|\sum_{n=1}^{N}\int_{\Omega}\Big(\int_{t_{n-1}}^{t_n} g(x,t)(\psi-\overline{R_h}\psi)(x,t)dt\Big)d\omega(x)\Big|\\
&= \Big|\sum_{n=1}^{N}\int_{\Omega}\Big(\int_{t_{n-1}}^{t_n} (g(x,t)\psi(x,t)-\overline{g}(x,t)R_h\psi(x,t))dt\Big)d\omega(x)\Big|\\
&\leqslant \Big|\sum_{n=1}^{N}\int_{\Omega}\Big(\int_{t_{n-1}}^{t_n} \psi(x,t)(g(x,t)-\overline{g}(x,t))dt\Big)d\omega(x)\Big|\\
&\quad +\Big|\sum_{n=1}^{N}\int_{\Omega}\Big(\int_{t_{n-1}}^{t_n} \overline{g}(x,t)(\psi-R_h\psi)(x,t)dt\Big)d\omega(x)\Big|\\
&\leqslant C\|g-\overline{g}\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)}\|\psi\|_{L^2(0,T;L^\infty(\Omega))}\\
&\quad +C\|\overline{g}\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)}\|\psi-R_h\psi\|_{L^2(0,T;L^\infty(\Omega))}.
\end{aligned}
\tag{4.29}
$$

Standard error estimates yield

$$\|g-\overline{g}\|_{L^2(0,T;L^\infty(\Omega))}\leqslant Ck^{\frac12}\|g\|_{H^{\frac12}(0,T;L^\infty(\Omega))} \tag{4.30}$$

and

$$\|\psi-R_h\psi\|_{L^2(0,T;L^\infty(\Omega))} \leqslant Ch^{2-\frac{d}{2}}\|\psi\|_{L^2(0,T;H^2(\Omega))}. \tag{4.31}$$

Thus we have

$$|\widetilde{E}_2|\leqslant C(k^{\frac12}+h^{2-\frac{d}{2}})\|\omega\|_{\mathcal{M}(\Omega)}\|g\|_{H^{\frac12}(0,T;L^\infty(\Omega))}\|f\|_{L^2(0,T;L^2(\Omega))}. \tag{4.32}$$

Similar to (4.13) we also have

$$|\widetilde{E}_3| \leqslant Ch\|y_0\|_{0,\Omega}\|f\|_{L^2(0,T;L^2(\Omega))}. \tag{4.33}$$

Then it remains to estimate $\widetilde{E}_1$. Cauchy-Schwarz inequality gives

$$|\widetilde{E}_1|\leqslant \widetilde{F}_1\cdot\widetilde{F}_2, \tag{4.34}$$

where

$$\widetilde{F}_1=\Big(\sum_{n=1}^{N}(\|Y_h^n-Y_h^{n-1}\|_{0,\Omega}^2+ka(Y_h^n,Y_h^n))\Big)^{\frac12}$$

and

$$\widetilde{F}_2=\Big(\sum_{n=1}^{N}(\|\psi^{n-1}-\overline{R_h}\psi\|_{0,\Omega}^2+ka(\overline{\psi}-\overline{R_h}\psi,\overline{\psi}-\overline{R_h}\psi))\Big)^{\frac12}.$$

From Lemma 4.4 we have

$$\widetilde{F}_1\leqslant C\Big(\|y_0\|_{0,\Omega}+\|g\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)}\Big). \tag{4.35}$$

Similar to (4.20) we can conclude from (4.16)-(4.19) that

$$|\widetilde{F}_2| \leqslant C(h^{2-\frac{d}{2}}+k^{\frac12})\|\psi\|_{2,1}, \tag{4.36}$$

thus

$$
\begin{aligned}
|\widetilde{E}_1| &\leqslant C(h^{2-\frac{d}{2}}+k^{\frac12})\|\psi\|_{2,1}\Big(\|y_0\|_{0,\Omega}+\|g\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)}\Big)\\
&\leqslant C(h^{2-\frac{d}{2}}+k^{\frac12})\|f\|_{L^2(0,T;L^2(\Omega))}\Big(\|y_0\|_{0,\Omega}+\|g\|_{L^2(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)}\Big).
\end{aligned}
\tag{4.37}
$$

Then from (4.28), (4.32), (4.33) and (4.37) we have

$$\begin{aligned}\|y-Y_h\|_{L^2(0,T;L^2(\Omega))} &= \sup_{f\in L^2(0,T;L^2(\Omega)),f\neq 0}\frac{(f,y-Y_h)_{\Omega_T}}{\|f\|_{L^2(0,T;L^2(\Omega))}}\\ &\leqslant C(h^{2-\frac{d}{2}}+k^{\frac{1}{2}})\Big(\|y_0\|_{0,\Omega}+\|g\|_{H^{\frac{1}{2}}(0,T;L^\infty(\Omega))}\|\omega\|_{\mathcal{M}(\Omega)}\Big),\end{aligned}$$

which completes the proof.

□

5. Numerical Examples

In this section we will carry out some numerical experiments to confirm our theoretical findings. For the computation the software package AFEPack ([17]) has been used. To validate the estimates developed in the previous section, we show the convergence order by separating the discretization errors. At first we consider the behavior of the error for a sequence of discretizations with different mesh sizes and a fixed time steps. Then we show the behavior of the error for different time steps but a fixed spatial triangulation.

In the following numerical examples, we define an error functional to show the experimental order of convergence by

$$\text{rate}=\frac{\log E(h_1)-\log E(h_2)}{\log h_1-\log h_2},$$

where $E(h)$ denotes the error on triangulation with mesh size h or time step k. Then it is easy to see that "rate $=\gamma$" means that "error $=O(h^\gamma)$".

5.1. **Parabolic equations with measure data in time.** At first we consider the following parabolic equation with Dirac righthand side in time:

$$\begin{cases}\partial_t y-\Delta y=g\omega & \text{in } \Omega_T,\\ y=0 & \text{on } \Gamma_T,\\ y(\cdot,0)=y_0 & \text{in } \Omega,\end{cases}$$

where $g(x,t)\in\mathcal{C}([0,T];L^2(\Omega))$ and $\omega(t)\in\mathcal{M}[0,T]$. For ease of constructing examples we may admit some additional regular parts to appear in the righthand side.

Example 5.1. *The first example is a modification from the example presented in* [22]*. Let* $\Omega_T=B(0,1)\times[0,1]$*, where* $B(0,1)$ *is the unit circle centered at zero with radius 1,* $\gamma\in(0,1)$ *and* $\lambda\in\mathbb{R}$*. Let*

$$\epsilon(t)=(e^{-\lambda t}-e^{-\frac{\lambda}{2}}),$$

we take the exact solution as

$$y(x,t) = \sin(\pi|x|^2)\frac{e^{\lambda t}}{\lambda(1-\gamma)}\cdot\begin{cases}\epsilon(0)^{1-\gamma}, & t\geqslant\frac{1}{2};\\ \epsilon(0)^{1-\gamma}-\epsilon(t)^{1-\gamma}, & t<\frac{1}{2}.\end{cases}$$

After simple calculation we have

$$\begin{aligned}\mu(x,t) &= \sin(\pi|x|^2)\delta(t)+(\sin(\pi|x|^2)\cdot\frac{e^{\lambda t}}{1-\gamma}+(-4\pi\cos(\pi|x|^2)+\\ &\quad 4\pi^2|x|^2\sin(\pi|x|^2))\frac{e^{\lambda t}}{\lambda(1-\gamma)})\cdot\begin{cases}\epsilon(0)^{1-\gamma}, & t\geqslant\frac{1}{2};\\ \epsilon(0)^{1-\gamma}-\epsilon(t)^{1-\gamma}, & t<\frac{1}{2},\end{cases}\end{aligned}$$

Table 1. Error of y for Example 5.1 with $\lambda = 1$.

Dofs	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate	N	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate
25	1.805765645861	\	3	0.229100247225	\
81	0.479593557806	1.9127	9	0.197363792883	0.1357
289	0.119994521144	1.9988	27	0.131680471328	0.3683
1089	0.031089301244	1.9485	81	0.080625762381	0.4465
4225	0.010045681219	1.6298	243	0.048726635543	0.4584
16641	0.006520318504	0.6236	729	0.029977569575	0.4422

Table 2. Error of y for Example 5.1 with $\lambda = 2$.

Dofs	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate	N	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate
25	2.405100524743	\	3	0.283293492668	\
81	0.636075956291	1.9188	9	0.157102357581	0.5367
289	0.158595488032	2.0038	27	0.106189907672	0.3565
1089	0.041037524463	1.9503	81	0.067208827352	0.4164
4225	0.013412438652	1.6134	243	0.041682864989	0.4348
16641	0.008934235708	0.5862	729	0.026138785274	0.4248

Table 3. Error of y for Example 5.1 with $\lambda = 10$.

Dofs	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate	N	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate
25	788.214049388618	\	3	2421.450238197359	\
81	200.611180708993	1.9742	9	614.302084069153	1.2485
289	48.311558519598	2.0540	27	162.023290177949	1.2131
1089	11.935206551992	2.0171	81	48.670234674881	1.0947
4225	2.934940200289	2.0238	243	14.900101550467	1.0774
16641	0.885011153476	1.7296	729	5.036596568855	0.9873

where

$$\delta(t) = \begin{cases} 0, & t \geqslant \dfrac{1}{2}; \\ \epsilon(t)^{-\gamma}, & t < \dfrac{1}{2}. \end{cases}$$

To confirm our theoretical results we test the convergence order with respect to space discretization and time discretization, respectively. To investigate the convergence order with respect to the space discretization we fixed the time discretization with $N = 33333$ for $\lambda = 1, 2$ and $N = 3333$ for $\lambda = 10$, while the space discretization is fixed with 16641 Dofs to investigate the convergence order with respect to the time discretization, the results for $\lambda = 1$, $\lambda = 2$ and $\lambda = 10$ are listed in Table 1, 2 and 3, respectively.

We can see from Table 1-3 that the convergence order with respect to the spatial discretization is almost 2, which is optimal and better than our predicted order $O(h)$. While the convergence order with respect to the time discretization is almost $O(k^{\frac{1}{2}})$ for the cases $\lambda = 1$ and $\lambda = 2$, which is consistent with our theoretical results. The convergence order for the time discretization is $O(k)$ for larger λ, as presented in Table 3.

Example 5.2. *The second example is constructed inspired by Example 4.2 of* [9]*. Let* $\Omega_T = B(0,1) \times [0,1]$*, and we take the exact solution as*

$$y(x,t) = \sin(\pi|x|^2) \cdot \begin{cases} t^2, & t < 0.5; \\ t^2 + 2t, & t \geqslant 0.5. \end{cases}$$

Table 4. Error of y for Example 5.2.

Dofs	$\|y-Y_h\|_{L^2(\Omega_T)}$	rate	N	$\|y-Y_h\|_{L^2(\Omega_T)}$	rate
25	1.003408828629	\	2	0.067209558965	\
81	0.265418710871	1.9186	4	0.033600480166	1.0002
289	0.065958680971	2.0086	8	0.016860913555	0.9948
1089	0.016567151789	1.9932	16	0.008466148990	0.9939
4225	0.004146042659	1.9985	32	0.004286813653	0.9818
16641	0.001034571264	2.0027	64	0.002276249156	0.9132

We know that $\partial_t y = \sin(\pi|x|^2)\cdot\delta_t(\frac{1}{2}) + \sin(\pi|x|^2)\cdot\gamma(t)$, where $\delta_t(z)$ denotes the Dirac measure with respect to the variable t concentrated at $t=z$, and

$$\gamma(t) = \begin{cases} 2t, & t < 0.5; \\ 2t+2, & t \geqslant 0.5. \end{cases}$$

Thus, after simple calculation we have

$$\begin{aligned} \mu(x,t) &= \sin(\pi|x|^2)\cdot\delta_t(\frac{1}{2}) + \sin(\pi|x|^2)\cdot\gamma(t) \\ &\quad +(-4\pi\cos(\pi|x|^2) + 4\pi^2|x|^2\sin(\pi|x|^2))\cdot\begin{cases} t^2, & t<0.5; \\ t^2+2t, & t\geqslant 0.5. \end{cases} \end{aligned}$$

At first we fixed the time discretization with 4096 time steps to investigate the behavior of error with respect to the spatial discretization, then the space discretization is fixed with 16641 Dofs to investigate the convergence order with respect to the time discretization. The results are listed in Table 4. From Table 4 we found that the convergence orders with respect to the spatial and time discretization are $O(h^2)$ and $O(k)$, respectively, both of them are higher than our predicted results which are $O(h)$ and $O(k^{\frac{1}{2}})$, respectively.

5.2. **Parabolic equations with measure data in space.** Since our theoretical results are also valid for Neumann boundary conditions, we consider in this subsection the following parabolic equation with Dirac source term in space and Neumann boundary condition:

$$\begin{cases} y_t - \Delta y + y = \mu & \text{in } \Omega_T, \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \Gamma_T, \\ y(0) = y_0 & \text{in } \Omega, \end{cases}$$

where $\mu = g(x,t)\omega(x)$, $g(x,t)\in L^2(0,T;\mathcal{C}(\overline{\Omega}))$ and $\omega(x)\in\mathcal{M}(\Omega)$. For ease of constructing examples we also admit some additional regular parts to appear in the righthand side.

Example 5.3. *The third example is a modification from Example 4.2 of* [9]*. Let $\Omega_T = [0,1]^2\times[0,1]$, we take the exact solution as*

$$y(x,t) = (e^t+1)\cdot\begin{cases} 0.5 - x_1^2, & x_1 < 0.5, \\ 0.25, & x_1 \geqslant 0.5, \end{cases}$$

since y does not depend on the spatial variable x_2, we find that $\Delta y = \Delta_{xx} y = (\delta_{x_1}(\frac{1}{2}) - \psi(x_1))\cdot(e^t+1)$, where

$$\psi(x_1) = \begin{cases} 2, & x_1 < 0.5, \\ 0, & x_1 \geqslant 0.5, \end{cases}$$

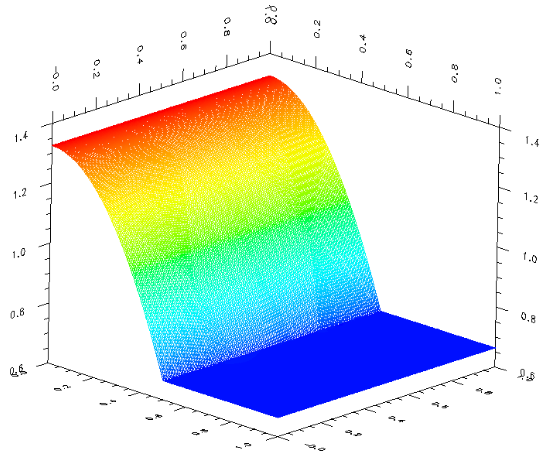

FIGURE 1. The discrete solution Y_h of Example 5.3 at time $t = 0.5$ with 22785 Dofs.

TABLE 5. Error of y for Example 5.3 with respect to space and time.

Dofs	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate	N	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate
31	0.033467303580	\	2	0.069700966171	\
105	0.007254281003	2.5066	4	0.030875705044	1.1747
385	0.002842650160	1.4421	8	0.014389164719	1.1015
1473	0.001350586630	1.1092	16	0.006927239146	1.0546
5761	0.000551335197	1.3139	32	0.003398471622	1.0274
22785	0.000194881287	1.5127	64	0.001688698129	1.0090

and $\delta_{x_i}(z)$ denotes the Dirac measure with respect to the variable x_i concentrated at $x_i = z$. Then after simple calculation we have

$$\mu(x,t) = y(x,t) - (e^t + 1) \cdot (\delta_{x_1}(\frac{1}{2}) - \psi(x_1)) + e^t \cdot \begin{cases} 0.5 - x_1^2, & x_1 < 0.5, \\ 0.25, & x_1 \geqslant 0.5. \end{cases}$$

To investigate the convergence order with respect to the space discretization we fix the time discretization with $N = 2048$, while the spatial discretization is fixed with 22785 Dofs to investigate the convergence order with respect to the time discretization, the results are listed in Table 5 and Figure 1 presents the numerical result at time $t = 0.5$ for a gird with 22785 Dofs. We can see from Table 5 that the convergence order w.r.t the space discretization is almost 1, which is consistent with our theoretical results. While the convergence order w.r.t the time discretization is 1, which is better than our predicted result of order $k^{\frac{1}{2}}$.

Example 5.4. *The fourth example is a modification from Example 4.1 of* [9]*, see also* [1]*. Let $\Omega_T = B(0,1) \times [0,1]$, we take the exact solution as*

$$y(x,t) = -\frac{1}{2\pi} \log|x| \cdot (e^t + 1),$$

then after simple calculation we have

$$\mu(x,t) = (e^t + 1)\delta_0 + y(x,t) - \frac{1}{2\pi} \log|x| \cdot e^t,$$

where δ_0 is the Dirac function at $x = (0,0)$.

To investigate the convergence order with respect to the space discretization we fixed the time discretization with $N = 2048$, while the space discretization is fixed with 66049 Dofs to investigate the convergence order with respect to the time discretization, the results are listed in Table 6 and

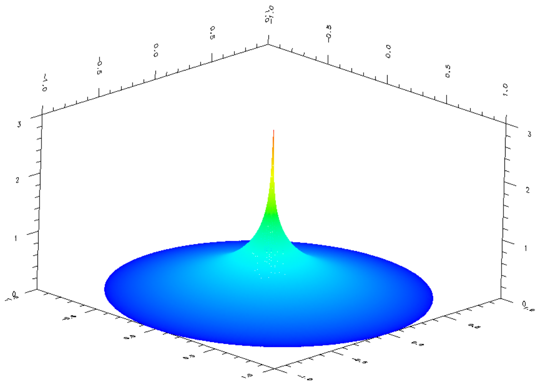

FIGURE 2. The discrete solution Y_h of Example 5.4 at time $t = 0.5$ with 66049 Dofs.

TABLE 6. Error of y for Example 5.4 with respect to space and time.

Dofs	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate	N	$\|y - Y_h\|_{L^2(\Omega_T)}$	rate
25	0.051725167740	\	2	0.029719956266	\
81	0.026461768120	0.9670	4	0.013208571995	1.1700
289	0.013361104038	0.9859	8	0.006210402376	1.0887
1089	0.006716533360	0.9923	16	0.003085031620	1.0094
4225	0.003366984764	0.9963	32	0.001685293247	0.8723
16641	0.001685479572	0.9983	64	0.001112777391	0.5988

Figure 2 presents the numerical result at time $t = 0.5$ for a gird with 66049 Dofs. We can see from Table 6 that the convergence order w.r.t the space discretization is 1, which is consistent with our theoretical results. While the convergence order w.r.t the time discretization is also 1, which is better than our predicted result of order $k^{\frac{1}{2}}$.

6. Conclusion and future work

In this paper we study the a priori error estimates for the finite element approximations of parabolic equations with measure data separately in time and space, respectively. The space discretization is done using piecewise linear and continuous finite elements, whereas the time discretization is based on the backward Euler method. We derive the a priori error estimates for the semidiscretization problems and the fully discrete problems, respectively. Numerical results are provided at the end of the paper to confirm our theoretical findings.

To the best of author's knowledge, this paper is among the fewer contributions to finite element method for partial differential equations involving measure data. Especially, the results obtained in this paper constitute the crucial ingredients to derive the error estimates for some kind of parabolic optimal control problems with state constraints. The traditional approach of error analysis for parabolic optimal control problems with state constraints is to avoid the error estimates for adjoint state equation, which is caused by the lack of error estimates for parabolic equations involving measure data (see, e.g., [10], [15] and [22]). We believe that the results in this paper provide a shortcut for the error analysis of such kind of problems.

Moreover, the results obtained in this paper can be viewed as the first step, but crucial step, for the error analysis of the finite element approximation to parabolic optimal control problems with pointwise control, where control acts on finitely many points of the domain. The state equation has the form of (2.13) with righthand side $\mu = u(t)\omega(x)$ (see (1.6)), where $u(t)$ denotes the control variable, the details can be found in [11], [14] and [25]. The finite element approximation of such kind of problems and corresponding error analysis will be addressed in our future work. In addition, as pointed out in the introduction, parabolic equations with measure data find many applications in optimal control theory. But generally the Lagrange multiplier μ associated to the state constraints for parabolic optimal control problems with pointwise state constraints belongs to $\mathcal{M}(\overline{\Omega}_T)$ (see (1.5)), so it is also very interesting to study the finite element approximation of parabolic problems with measure data in both space and time (see [4] for the analytical setting), the approaches developed in the current paper seems to be unapplicable for such extreme case.

On the other hand, since the solutions of parabolic problems involving measure data have lower regularities, only reduced convergence order can be expected by standard finite element approximation. Thus the a posteriori error estimate and adaptive finite element method for such kind of problems are necessary and deserved further study. Araya et al. ([1]) have studied a posteriori error estimates for elliptic problems with Dirac delta source terms, the applications of their approaches to our setting will be postponed to our future work.

Acknowledgements

The author is grateful to the anonymous referees for their careful reviews and many valuable suggestions that led to a considerably improved paper, he also acknowledges professor Ningning Yan from AMSS, CAS for a careful reading of the preliminary version of this article. The author would like to thank Alexander von Humboldt Foundation for support during his stay in University of Hamburg, Germany. He is very grateful to the Department of Mathematics, University of Hamburg for the hospitality and support.

References

[1] R. Araya, E. Behrens and R. Rodríguez, A posteriori error estimates for elliptic problems with Dirac delta source terms, Numer. Math., 105 (2006), pp. 193-216.
[2] I. Babuska, Error bounds for the finite element method, Numer. Math., 16 (1971), pp. 322-333.
[3] L. Boccardo, A. Dall'Aglio, T. Gallouët and L. Orsina, Nonlinear parabolic equations with measure data, J. Funct. Anal., 147 (1997), pp. 237-258.
[4] L. Boccardo and T. Gallouët, Non-linear elliptic and parabolic equations involving measure data, J. Funct. Anal., 87 (1989), pp. 149-169.
[5] E. Casas, L^2 estimates for the finite element method for the Dirichlet problem with singular data, Numer. Math., 47 (1985), pp. 627-632.
[6] E. Casas, Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations, SIAM J. Control Optim., 35 (1997), pp. 1297-1327.
[7] K. Chrysafinos and L. S. Hou, Error estimates for semidiscrete finite element approximations of linear and semilinear parabolic equations under minimal regularity assumptions, SIAM J. Numer. Anal., 40 (2002), pp. 282-306.
[8] P. G. Ciarlet, The finite element methods for elliptic problems, North-Holland, Amsterdam, 1978.
[9] K. Deckelnick and M. Hinze, Convergence of a fnite element approximation to a state constrained elliptic control problem, SIAM J. Numer. Anal., 45 (2007), pp. 1937–1953.
[10] K. Deckelnick and M. Hinze, Variational discretization of parabolic control problems in the presence of pointwise state constraints, J. Comput. Math., 29 (2011), pp. 1-16.
[11] J. Droniou and J. P. Raymond, Optimal pointwise control of semilinear parabolic equations, Nonlinear Analysis, 39 (2000), pp. 135-156.
[12] 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 (1995), pp. 706-740.
[13] D. A. French and J. T. King, Analysis of a robust finite element approximation for a parabolic equation with rough boundary data, Math. Comput., 60 (1993), pp. 79-104.
[14] R. Glowinski and J. L. Lions, Exact and approximate controllability for distributed parameter systems, Acta Numerica, 1996, pp. 159-333.
[15] W. Gong and M. Hinze, Error estimates for parabolic optimal control problems with control and state constraints, preprint, Hamburger Beiträge zur Angewandten Mathematik, 2010-13.
[16] W. Gong and N. N. Yan, A mixed finite element scheme for optimal control problems with pointwise state constraints, J. Sci. Comput., 46 (2011), pp. 182-203.
[17] R. Li and W. B. Liu, AFEPack available online at *http://circus.math.pku.edu.cn/AFEPack*.
[18] J. L. Lions, Optimal control of systems governed by partial differential equations, Springer-Verlag, Berlin, 1971.
[19] J. L. Lions and E. Magenes, Non-Homogeneous Boundary Value Problems and Applications, Springer-Verlag, Berlin, 1972.

[20] M. Luskin and R. Rannacher, On the smoothing property of the Galerkin method for parabolic equations, SIAM J. Numer. Anal., 19 (1982), pp. 93-113.
[21] 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 (2000), pp. 1534-1553.
[22] 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, preprint-Nr.: SPP1253-098, to appear in SIAM J. Control Optim., 2011.
[23] R. H. Nochetto and C. Verdi, Convergence past singularities for a fully discrete approximation of curvature-driven interfaces, SIAM J. Numer. Anal., 34 (1997), pp. 490-512.
[24] A. Quarteroni and A. Valli, Numerical approximation of partial differential equations, volume 23 of Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 1994.
[25] A. M. Ramos, R. Glowinski and J. Periaux, Pointwise control of the Burgers equation and related Nash equilibrium problems: Computational approach, J. Optim. Theory Appl., 112 (2002), pp. 499-516.
[26] R. Rannacher and R. Scott, Some optimal error estimates for piecewise linear finite element approximations, Math. Comput., 38 (1982), pp. 437-445.
[27] A. H. Schatz, Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids. I: Global estimates, Math. Comput., 67 (1998), pp. 877-899.
[28] R. Scott, Finite element convergence for singular data, Numer. Math., 21 (1973), pp. 317-327.
[29] R. Scott, Optimal L^∞ estimates for the finite element method on irregular meshes, Math. Comput., 30 (1976), pp. 681-697.
[30] V. Thomée, Galerkin finite element methods for parabolic problems, Springer-Verlag, Berlin, 2006.