SIAM J. NUMER. ANAL.
Vol. 51, No. 5, pp. 2797–2821

OPTIMAL A PRIORI ERROR ESTIMATES OF PARABOLIC OPTIMAL CONTROL PROBLEMS WITH POINTWISE CONTROL*

DMITRIY LEYKEKHMAN† AND BORIS VEXLER‡

Abstract. In this paper we consider a parabolic optimal control problem with a pointwise (Dirac type) control in space, but variable in time, in two space dimensions. To approximate the problem we use the standard continuous piecewise linear approximation in space and the piecewise constant discontinuous Galerkin method in time. Despite low regularity of the state equation, we show almost optimal $h^2 + k$ convergence rate for the control in L^2 norm. This result improves almost twice the previously known estimate in [W. Gong, M. Hinze, and Z. Zhou, *A Priori Error Analysis for Finite Element Approximation of Parabolic Optimal Control Problems with Pointwise Control*, Tech. report, 2011-07, Hamburger Beiträge zur Angewandten Mathematik, Hamburg, Germany, 2011].

Key words. optimal control, pointwise control, parabolic problems, finite elements, discontinuous Galerkin, error estimates, pointwise error estimates

AMS subject classifications. 49M25, 49K20, 65N30, 65N15

DOI. 10.1137/120885772

1. Introduction. In this paper we provide numerical analysis for the following optimal control problem:

$$\min_{q,u} J(q,u) := \frac{1}{2}\int_0^T \|u(t)-\widehat{u}(t)\|^2_{L^2(\Omega)}dt + \frac{\alpha}{2}\int_0^T |q(t)|^2 dt \tag{1.1}$$

subject to the second order parabolic equation

$$u_t(t,x) - \Delta u(t,x) = q(t)\delta_{x_0}(x), \qquad (t,x) \in I\times\Omega, \tag{1.2a}$$

$$u(t,x) = 0, \qquad (t,x)\in I\times\partial\Omega, \tag{1.2b}$$

$$u(0,x) = 0, \qquad x\in\Omega, \tag{1.2c}$$

and subject to pointwise control constraints

$$q_a \le q(t) \le q_b \quad \text{a.e. in } I. \tag{1.3}$$

Here $I = [0,T]$, $\Omega \subset \mathbb{R}^2$ is a convex polygonal domain, $x_0 \in \operatorname{Int}\Omega$ fixed, and δ_{x_0} is the Dirac delta function. The parameter α is assumed to be positive and the desired state $\widehat{u}$ fulfills $\widehat{u} \in L^2(I; L^\infty(\Omega))$. The control bounds $q_a, q_b \in \mathbb{R}\cup\{\pm\infty\}$ fulfill $q_a < q_b$. The precise functional-analytic setting is discussed in the next section.

This setup is a model for problems with pointwise control that can vary in time. For simplicity we consider here the case of only one point source. However, all presented results extend directly to the case of $l \ge 1$ point sources $\sum_{i=1}^{l} q_i(t)\delta_{x_i}(x)$.

There are several applications in the context of optimal control as well as of inverse problems leading to pointwise control. The main mathematical difficulty is

*Received by the editors July 24, 2012; accepted for publication (in revised form) July 2, 2013; published electronically October 16, 2013.
http://www.siam.org/journals/sinum/51-5/88577.html

†Department of Mathematics, University of Connecticut, Storrs, CT 06269 (leykekhman@math.uconn.edu). The first author was partially supported by NSF grant DMS-1115288.

‡Lehrstuhl für Mathematische Optimierung, Technische Universität München, 85748 Garching b. München, Germany (vexler@ma.tum.de). The second author was partially supported by the DFG Priority Program 1253 "Optimization with Partial Differential Equations."

low regularity of the state variable for such problems. We refer to [13, 34] for pointwise control in the context of Burgers type equations and to [9, 16] for pointwise control of parabolic systems. Moreover, a recent approach to sparse control problems utilizes a formulation with control variable from measure spaces; see [7, 8, 10, 33].

For the discretization, we consider the standard continuous piecewise linear finite elements in space and piecewise constant discontinuous Galerkin method in time. This is a special case ($r = 0$, $s = 1$) of so-called dG(r)cG(s) discretization; see, e.g., [19] for analysis of the method for parabolic problems and, e.g., [31, 32] for error estimates in the context of optimal control problems. Throughout, we will denote by h the spatial mesh size and by k the time step; see section 3 for details.

The numerical analysis of the problem under the consideration is challenging due to low regularity of the state equation. On the other hand, the corresponding adjoint (dual) state is more regular, which is exploited in our analysis. In contrast, optimal control problems with state constraints leads to optimality systems with lower regularity of the adjoint state and more regular state; see [14, 30] for a priori error estimates for discretization of state-constrained problems governed by parabolic equations.

Although, numerical analysis for elliptic problems with a rough right-hand side was considered in a number of papers [2, 3, 6, 18, 39], there are few papers that consider parabolic problems with rough sources. We are only aware of the paper [22], where $L^2(I; L^2(\Omega))$ error estimates are considered. Based on the results of this paper, suboptimal error estimates of order $\mathcal{O}(k^{\frac{1}{2}} + h)$ for the optimal control problem under the consideration were derived in [23]. However, the numerical results in the same paper strongly suggest better convergence rates. Examining the error analysis in [23], one can notice that the authors worked with L^2 norm in space for both the state and the adjoint equations. Looking at these equations separately, one can see that only the state equation has a singularity at x_0, the adjoint equation does not. As a result, the solutions to these equations have different regularity. To obtain better order estimates, one must choose the functional spaces for the error analysis more carefully. Roughly speaking, performing an error analysis in $L^1(\Omega)$ norm is space and L^2 norm in time for the state equation as well as an error analysis in L^∞ in space and L^2 norm in time for the adjoint equation, we are able to improve the error estimates for the control to the almost optimal order $\mathcal{O}(k + h^2)$. The main result in the paper is the following.

THEOREM 1.1. *Let $\bar{q}$ be optimal control for the problem* (1.1)–(1.2) *and $\bar{q}_{kh}$ be the optimal dG(0)cG(1) solution. Then there exists a constant C independent of h and k such that*

$$\|\bar{q} - \bar{q}_{kh}\|_{L^2(I)} \leq C\alpha^{-1} d^{-1} |\ln h|^{\frac{7}{2}} \left(k + h^2\right),$$

where d is the radius of the largest ball centered at x_0 that is contained in Ω.

We would also like to point out that, in addition to almost optimal order estimates, our analysis does not require any relationship between the size of the space discretization h and the time steps k. In our opinion any relation between h and k is not natural for the method since the piecewise constant discontinuous Galerkin method is just a variation of backward Euler method and is unconditionally stable.

The main ingredients of our analysis are the global and local pointwise in space error estimates, Theorems 3.1 and 3.5, respectively. In these theorems the discretization error is estimated with respect to the $L^\infty(\Omega; L^2(I))$-norm. These results have an independent interest since the error estimates in such a norm are somewhat nonstandard and are not considered in the finite element literature. We are not aware of any results in this direction. The local estimate in Theorem 3.5 is based on the global result from Theorem 3.1 and uses a localization technique from [36]. This local estimate is essential for our analysis since on the one hand only local error of the adjoint state at point x_0 plays a role (see the proof of Theorem 1.1) and on the other hand the required regularity of the adjoint state can only be expected in the interior of Ω; cf. Proposition 2.3.

Due to substantial technicalities, this paper treats the two dimensional case only. The technique of the proof does not immediately extend to three space dimensions. Moreover, we believe that in three space dimensions, due to stronger singularity, the optimal order estimates cannot hold without special mesh refinement near the singularity. This is a subject of future work.

Throughout this paper we use the usual notation for Lebesgue and Sobolev spaces. We denote by $(\cdot,\cdot)_\Omega$ the inner product in $L^2(\Omega)$ and by $(\cdot,\cdot)_{J\times\Omega}$ with some subinterval $J\subset I$ the inner product in $L^2(J;L^2(\Omega))$.

The rest of this paper is organized as follows. In section 2 we discuss the functional analytic setting of the problem, state the optimality system, and prove regularity results for the state and for the adjoint state. In section 3 we establish important global and local error estimates with respect to the $L^\infty(\Omega;L^2(I))$-norm for the heat equation. In section 4 we prove our main result, and in the last section we provide numerical examples illustrating our error estimates.

2. Optimal control problem and regularity. In order to state the functional analytic setting for the optimal control problem, we first introduce an axillary problem

$$\begin{aligned} v_t(t,x)-\Delta v(t,x) &= f(t,x), && (t,x)\in I\times\Omega,\\ v(t,x) &= 0, && (t,x)\in I\times\partial\Omega,\\ v(0,x) &= 0, && x\in\Omega \end{aligned} \tag{2.1}$$

with a right-hand side $f\in L^2(I;L^p(\Omega))$ for some $1<p<\infty$. This equation possesses a unique solution

$$v\in L^2(I;H^1_0(\Omega))\cap H^1(I;H^{-1}(\Omega)).$$

Due to the convexity of the polygonal domain Ω, the solution v possesses an additional regularity for $p=2$:

$$v\in L^2(I;H^2(\Omega)\cap H^1_0(\Omega))\cap H^1(I;L^2(\Omega)),$$

with the corresponding estimate

$$\|v\|_{L^2(I;H^2(\Omega))}+\|v_t\|_{L^2(I;L^2(\Omega))}\le c\|f\|_{L^2(I;L^2(\Omega))}; \tag{2.2}$$

see, e.g., [20]. Moreover, there holds the following regularity result.

LEMMA 2.1. *If $f\in L^2(I;L^p(\Omega))$ for an arbitrary $p>1$, then $v\in L^2(I;C(\Omega))$ and*

$$\|v\|_{L^2(I;C(\Omega))}\le C_p\|f\|_{L^2(I;L^p(\Omega))},$$

where $C_p\sim\frac{1}{p-1}$, as $p\to 1$.

Proof. This lemma follows from the maximal regularity result [24] that says that if $f\in L^2(I;L^p(\Omega))$ for any $p>1$, then $\Delta v\in L^2(I;L^p(\Omega))$ and $v_t\in L^2(I;L^p(\Omega))$ with the following estimate:

$$\|v_t\|_{L^2(I;L^p(\Omega))}+\|\Delta v\|_{L^2(I;L^p(\Omega))}\le C\|f\|_{L^2(I;L^p(\Omega))}, \tag{2.3}$$

where the constant C does not depend on p. Since by our assumption Ω is polygonal and convex, there exists some $p_\Omega > 2$ (see [25]) such that

$$\|v\|_{L^2(I;W^{2,p}(\Omega))} \le C_p \|\Delta v\|_{L^2(I;L^p(\Omega))}$$

for all $1 < p \le p_\Omega$, where $C_p \sim \frac{1}{p-1}$ as $p \to 1$. The exact form of the constant can be traced, for example, from Theorem 9.9 in [21]. By the embedding $W^{2,1}(\Omega) \hookrightarrow C(\Omega)$ we have $v \in L^2(I; C(\Omega))$ and the desired estimate follows. ∎

We will also need the following local regularity result. Here and in what follows, we will denote an open ball of radius d centered at x_0 by $B_d = B_d(x_0)$.

LEMMA 2.2. *If $\overline{B}_{2d} \subset \Omega$ and $f \in L^2(I; L^2(\Omega)) \cap L^2(I; L^p(B_{2d}))$ for some $2 \le p < \infty$, then $v \in L^2(I; W^{2,p}(B_d)) \cap H^1(I; L^p(B_d))$ and there exists a constant C independent of p and d such that*

$$\|v_t\|_{L^2(I;L^p(B_d))} + \|v\|_{L^2(I;W^{2,p}(B_d))} \le Cp(\|f\|_{L^2(I;L^p(B_{2d}))} + d^{-1}\|f\|_{L^2(I;L^2(\Omega))}).$$

Proof. To obtain the local estimate we introduce a smooth cut-off function ω with properties that

$$\omega(x) \equiv 1, \quad x \in B_d(x_0), \tag{2.4a}$$

$$\omega(x) \equiv 0, \quad x \in \Omega \backslash B_{2d}(x_0), \tag{2.4b}$$

$$|\nabla \omega| \le C d^{-1}, \quad |\nabla^2 \omega| \le C d^{-2}. \tag{2.4c}$$

Define

$$\bar{v}(t) = \frac{1}{|B_{2d}|} \int_{B_{2d}} v(t,x)dx.$$

By the Cauchy–Schwarz inequality we have

$$\bar{v}_t \le \frac{1}{|B_{2d}|} |B_{2d}|^{1/2} \|v_t\|_{L^2(B_{2d})} \le C d^{-1} \|v_t\|_{L^2(B_{2d})}. \tag{2.5}$$

We set $\tilde{v} = (v - \bar{v})\omega$. There holds

$$\Delta \tilde{v} = \omega \Delta v + \nabla v \cdot \nabla \omega + (v - \bar{v}) \Delta \omega.$$

Therefore, $\tilde{v}$ satisfies the following equation:

$$\tilde{v}_t - \Delta \tilde{v} = g, \quad v(0,x) = 0,$$

on B_{2d} with homogeneous Dirichlet boundary conditions, where

$$\begin{aligned} g &= (v_t - \Delta v)\omega - \nabla v \cdot \nabla \omega - (v - \bar{v})\Delta \omega - \bar{v}_t \omega \\ &= f\omega - \nabla v \cdot \nabla \omega - (v - \bar{v})\Delta \omega - \bar{v}_t \omega. \end{aligned}$$

We have

$$\begin{aligned} \|g\|_{L^2(I;L^p(B_{2d}))} \le C\Big(&\|f\|_{L^2(I;L^p(B_{2d}))} + d^{-1}\|\nabla v\|_{L^2(I;L^p(B_{2d}))} \\ &+ d^{-2}\|v - \bar{v}\|_{L^2(I;L^p(B_{2d}))} + \|\bar{v}_t\|_{L^2(I;L^p(B_{2d}))}\Big). \end{aligned}$$

Using the Sobolev embedding theorem and (2.2), we have

$$\|\nabla v\|_{L^2(I;L^p(B_{2d}))} \le C\|v\|_{L^2(I;H^2(B_{2d}))} \le C\|f\|_{L^2(I;L^2(\Omega))}.$$

Similarly, using the Poincare inequality first, we obtain

$$\|v-\bar{v}\|_{L^2(I;L^p(B_{2d}))} \le Cd\|\nabla v\|_{L^2(I;L^p(B_{2d}))} \le Cd\|f\|_{L^2(I;L^2(\Omega))}.$$

Also, by (2.5) we have

$$\|\bar{v}_t\|_{L^2(I;L^p(B_{2d}))} \le Cd^{\frac{2}{p}-1}\|v_t\|_{L^2(I;L^2(B_{2d}))}. \tag{2.6}$$

By the maximum regularity estimate [24] we obtain

$$\begin{aligned}\|\tilde{v}_t\|_{L^2(I;L^p(B_{2d}))} + \|\Delta\tilde{v}\|_{L^2(I;L^p(B_{2d}))} &\le C\|g\|_{L^2(I;L^p(B_{2d}))}\\ &\le C\left(d^{-1}\|f\|_{L^2(I;L^2(\Omega))} + \|f\|_{L^2(I;L^p(B_{2d}))}\right),\end{aligned}$$

and due to the fact that B_{2d} has a smooth boundary, we also have

$$\|\tilde{v}\|_{L^2(I;W^{2,p}(B_{2d}))} \le Cp\|\Delta\tilde{v}\|_{L^2(I;L^p(B_{2d}))}$$

for any $2 \le p < \infty$. Observing that $\nabla^2 v = \nabla^2\tilde{v}$ on B_d we obtain the desired estimate for $\|v\|_{L^2(I;W^{2,p}(B_d))}$. The estimate for $\|v_t\|_{L^2(I;L^p(B_d))}$ follows by the fact that $v_t = \tilde{v}_t + \bar{v}_t$ on B_d, estimate (2.6), and by the triangle inequality. This completes the proof. ◻

To introduce a weak solution of the state equation (1.2) we use the method of transposition; cf. [29]. For a given control $q \in Q = L^2(I)$ we denote by $u = u(q) \in L^2(I;L^2(\Omega))$ a weak solution of (1.2), if for all $\varphi \in L^2(I;L^2(\Omega))$ there holds

$$(u,\varphi)_{I\times\Omega} = \int_I w(t,x_0)q(t)\,dt,$$

where $w \in L^2(I;H^2(\Omega)\cap H^1_0(\Omega)) \cap H^1(I;L^2(\Omega))$ is the weak solution of the adjoint equation

$$\begin{aligned}-w_t(t,x) - \Delta w(t,x) &= \varphi(t,x), \quad && (t,x) \in I\times\Omega,\\ w(t,x) &= 0, && (t,x) \in I\times\partial\Omega,\\ w(T,x) &= 0, && x \in \Omega.\end{aligned} \tag{2.7}$$

The existence of this weak solution $u = u(q)$ follows by the Riesz representation theorem using the embedding $L^2(I;H^2(\Omega)) \hookrightarrow L^2(I;C(\Omega))$. Using Lemma 2.1 we can prove additional regularity for the state variable $u = u(q)$.

PROPOSITION 2.1. *Let $q \in Q = L^2(I)$ be given and let $u = u(q)$ be the solution of the state equation* (1.2)*. Then $u \in L^2(I;L^p(\Omega))$ for any $p < \infty$ and the following estimate holds for $p \to \infty$ with a constant C independent of p:*

$$\|u\|_{L^2(I;L^p(\Omega))} \le Cp\|q\|_{L^2(I)}.$$

Proof. To establish the result we use a duality argument. There holds

$$\|u\|_{L^2(I;L^p(\Omega))} = \sup_{\|\varphi\|_{L^2(I;L^s(\Omega))}=1} (u,\varphi)_{I\times\Omega}, \quad \text{where} \quad \frac{1}{p}+\frac{1}{s} = 1.$$

Let w be the solution to (2.7) for $\varphi \in L^2(I; L^s(\Omega))$ with $\|\varphi\|_{L^2(I;L^s(\Omega))} = 1$. From Lemma 2.1, $w \in L^2(I; C(\Omega))$ and the following estimate holds:

$$\|w\|_{L^2(I;C(\Omega))} \le \frac{C}{s-1}\|\varphi\|_{L^2(I;L^s(\Omega))} = \frac{C}{s-1} \le Cp \text{ as } p \to \infty.$$

Thus

$$\begin{aligned}\|u\|_{L^2(I;L^p(\Omega))} &= \sup_{\|\varphi\|_{L^2(I;L^s(\Omega))}=1} (u,\varphi)_{I\times\Omega}\\ &= \int_I w(t,x_0)q(t)\,dt \le \|q\|_{L^2(I)}\|w\|_{L^2(I;C(\Omega))} \le Cp\|q\|_{L^2(I)}. \qquad \square\end{aligned}$$

A further regularity result for the state equation follows from [17].

PROPOSITION 2.2. *Let $q \in Q = L^2(I)$ be given and let $u = u(q)$ be the solution of the state equation* (1.2)*. Then for each $\frac{3}{2} < s < 2$ and $\varepsilon > 0$, there holds*

$$u \in L^2(I; W_0^{1,s}(\Omega)), \quad u_t \in L^2(I; W^{-1,s}(\Omega)), \quad \text{and} \quad u \in C(\bar{I}; W^{-\varepsilon,s}(\Omega))$$

for any $\varepsilon > 0$. Moreover, the state u fulfills the following weak formulation:

$$\langle u_t, \varphi\rangle + (\nabla u, \nabla\varphi) = \int_I \varphi(t,x_0)q(t)\,dt \quad \text{for all } \varphi \in L^2(I; W^{1,s'}(\Omega)),$$

where $\frac{1}{s'} + \frac{1}{s} = 1$ and $\langle\cdot,\cdot\rangle$ is the duality product between $L^2(I; W^{-1,s}(\Omega))$ and $L^2(I; W_0^{1,s'}(\Omega))$.

Proof. For $s < 2$ we have $s' > 2$ and, therefore, $W_0^{1,s'}(\Omega)$ is embedded into $C(\bar{\Omega})$. Therefore, the right-hand side $q(t)\delta_{x_0}$ of the state equation can be identified with an element in $L^2(I; W^{-1,s}(\Omega))$. Using the result from [17, Theorem 5.1] on maximal parabolic regularity and exploiting the fact that $-\Delta\colon W_0^{1,s}(\Omega) \to W^{-1,s}(\Omega)$ is an isomorphism (see [27]), we obtain

$$u \in L^2(I; W_0^{1,s}(\Omega)) \quad \text{and} \quad u_t \in L^2(I; W^{-1,s}(\Omega)).$$

The assertion $u \in C(\bar{I}; W^{-\varepsilon,s}(\Omega))$ then follows by embedding and interpolation; see [1, Chap. III, Theorem 4.10.2]. Given the above regularity the corresponding weak formulation is fulfilled by a standard density argument. $\square$

As the next step we introduce the reduced cost functional $j\colon Q \to \mathbb{R}$ on the control space $Q = L^2(I)$ by

$$j(q) = J(q, u(q)),$$

where J is the cost function in (1.1) and $u(q)$ is the weak solution of the state equation (1.2) as defined above. The optimal control problem can then be equivalently reformulated as

$$\min\, j(q), \quad q \in Q_{\mathrm{ad}}, \tag{2.8}$$

where the set of admissible controls is defined according to (1.3) by

$$Q_{\mathrm{ad}} = \{\, q \in Q \mid q_a \le q(t) \le q_b \text{ almost everywhere (a.e.) in } I \,\}. \tag{2.9}$$

By standard arguments this optimization problem possesses a unique solution $\bar{q} \in Q = L^2(I)$ with the corresponding state $\bar{u} = u(\bar{q}) \in L^2(I; L^p(\Omega))$; see Proposition 2.1

for the regularity of $\bar{u}$. Due to the fact that this optimal control problem is convex, the solution $\bar{q}$ is equivalently characterized by the optimality condition

$$j'(\bar{q})(\delta q - \bar{q}) \geq 0 \quad \text{for all } \delta q \in Q_{\text{ad}}. \tag{2.10}$$

The (directional) derivative $j'(q)(\delta q)$ for given $q, \delta q \in Q$ can be expressed as

$$j'(q)(\delta q) = \int_I (\alpha q(t) + z(t, x_0))\, \delta q(t)\, dt,$$

where $z = z(q)$ is the solution of the adjoint equation

$$-z_t(t,x) - \Delta z(t,x) = u(t,x) - \widehat{u}(t,x), \qquad (t,x) \in I \times \Omega, \tag{2.11a}$$

$$z(t,x) = 0, \qquad (t,x) \in I \times \partial\Omega, \tag{2.11b}$$

$$z(T,x) = 0, \qquad x \in \Omega, \tag{2.11c}$$

and $u = u(q)$ on the right-hand side of (2.11a) is the solution of the state equation (1.2). The adjoint solution, which corresponds to the optimal control $\bar{q}$, is denoted by $\bar{z} = z(\bar{q})$.

The optimality condition (2.10) is a variational inequality, which can be equivalently formulated using the pointwise projection

$$P_{Q_{\text{ad}}} \colon Q \to Q_{\text{ad}}, \quad P_{Q_{\text{ad}}}(q)(t) = \min\big(q_b, \max(q_a, q(t))\big).$$

The resulting condition reads

$$\bar{q} = P_{Q_{\text{ad}}}\left(-\frac{1}{\alpha}\bar{z}(\cdot, x_0)\right). \tag{2.12}$$

In the next proposition we provide an important regularity result for the solution of the adjoint equation.

PROPOSITION 2.3. *Let $q \in Q$ be given, let $u = u(q)$ be the corresponding state fulfilling* (1.2), *and let $z = z(q)$ be the corresponding adjoint state fulfilling* (2.11). *Then,*

(a) $z \in L^2(I; H^2(\Omega) \cap H^1_0(\Omega)) \cap H^1(I; L^2(\Omega))$ *and the following estimate holds:*

$$\|\nabla^2 z\|_{L^2(I;L^2(\Omega))} + \|z_t\|_{L^2(I;L^2(\Omega))} \leq c(\|q\|_{L^2(I)} + \|\hat{u}\|_{L^2(I;L^2(\Omega))}).$$

(b) *If $\overline{B}_{2d} \subset \Omega$, then $z \in L^2(I; W^{2,p}(B_d)) \cap H^1(I; L^p(B_d))$ for all $2 \leq p < \infty$ and the following estimate holds:*

$$\|\nabla^2 z\|_{L^2(I;L^p(B_d))} + \|z_t\|_{L^2(I;L^p(B_d))} \leq cp^2 d^{-1}(\|q\|_{L^2(I)} + \|\hat{u}\|_{L^2(I;L^\infty(\Omega))}).$$

Proof.

(a) The right-hand side of the adjoint equation fulfills $u - \widehat{u} \in L^2(I; L^p(\Omega))$ for all $1 < p < \infty$; see Proposition 2.1. Due to the convexity of the domain Ω we directly obtain $z \in L^2(I; H^2(\Omega) \cap H^1_0(\Omega)) \cap H^1(I; L^2(\Omega))$ and the estimate

$$\|\nabla^2 z\|_{L^2(I;L^2(\Omega))} + \|z_t\|_{L^2(I;L^2(\Omega))} \leq c\|u - \widehat{u}\|_{L^2(I;L^2(\Omega))}.$$

The result from Proposition 2.1 leads directly to the first estimate.

(b) From Lemma 2.2 for $p \geq 2$ we have

$$\|\nabla^2 z\|_{L^2(I;L^p(B_d))} + \|z_t\|_{L^2(I;L^p(B_d))} \leq Cpd^{-1}\|u - \hat{u}\|_{L^2(I;L^p(\Omega))}.$$

Hence, by the triangle inequality and Proposition 2.1 we obtain

$$\|u - \hat{u}\|_{L^2(I;L^p(\Omega))} \leq C\left(p\|q\|_{L^2(I)} + \|\hat{u}\|_{L^2(I;L^\infty(\Omega))}\right).$$

This completes the proof. ∎

Remark 2.1. From Proposition 2.3 one concludes that $z \in H^{1-\varepsilon}(I; C(B_d))$ for all $\varepsilon > 0$ using an embedding result from [12, Chap. XVIII, Theorem 6, p. 494]. Hence, there holds $z(\cdot, x_0) \in H^{1-\varepsilon}(I)$. Using the pointwise representation (2.12) of the optimal control $\bar{q}$ and the fact that this projection operator preserves H^s-regularity for $0 \leq s \leq 1$ (see [28, Lemma 3.3]), we obtain $\bar{q} \in H^{1-\varepsilon}(I)$. We do not need this regularity for the proof of our error estimates, but the order of convergence in Theorem 1.1 is consistent with this regularity result.

3. Discretization and the best approximation results for parabolic problem.

3.1. Space-time discretization and notation. For the discretization of the problem under the consideration we introduce a partition of $I = [0, T]$ into subintervals $I_m = (t_{m-1}, t_m]$ of length $k_m = t_m - t_{m-1}$, where $0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = T$. The maximal time step is denoted by $k = \max_m k_m$. The semidiscrete space X_k^0 of piecewise constant functions in time is defined by

$$X_k^0 = \{v_k \in L^2(I; H_0^1(\Omega)) : \ v_k|_{I_m} \in \mathcal{P}_0(H_0^1(\Omega)), \ m = 1, 2, \ldots, M\},$$

where $\mathcal{P}_0(V)$ is the space of constant functions in time with values in V. We will employ the following notation for functions in X_k^0:

$$v_m^+ = \lim_{\varepsilon \to 0^+} v(t_m + \varepsilon) := v_{m+1}, \ v_m^- = \lim_{\varepsilon \to 0^+} v(t_m - \varepsilon) = v(t_m) := v_m, \ [v]_m = v_m^+ - v_m^-. \tag{3.1}$$

Let $\mathcal{T}$ denote a quasi-uniform triangulation of Ω with a mesh size h, i.e., $\mathcal{T} = \{\tau\}$ is a partition of Ω into triangles τ of diameter h_τ such that for $h = \max_\tau h_\tau$,

$$\operatorname{diam}(\tau) \leq h \leq C|\tau|^{\frac{1}{2}} \quad \forall \tau \in \mathcal{T}$$

hold. Let V_h be the set of all functions in $H_0^1(\Omega)$ that are linear on each τ, i.e., V_h is the usual space of linear finite elements. We will use the usual nodewise interpolation $\pi_h \colon C_0(\Omega) \to V_h$, the Clement intepolation $\pi_h \colon L^1(\Omega) \to V_h$, and the L^2-projection $P_h \colon L^2(\Omega) \to V_h$ defined by

$$(P_h v, \chi)_\Omega = (v, \chi)_\Omega \quad \forall \chi \in V_h. \tag{3.2}$$

To obtain the fully discrete approximation, we consider the space-time finite element space

$$X_{k,h}^{0,1} = \{v_{kh} \in X_k^0 : \ v_{kh}|_{I_m} \in \mathcal{P}_0(V_h), \ m = 1, 2, \ldots, M\}. \tag{3.3}$$

We will also need the following semidiscrete projection $\pi_k \colon C(\bar{I}; H_0^1(\Omega)) \to X_k^0$ defined by

$$\pi_k v|_{I_m} = v(t_m), \quad m = 1, 2, \ldots, M.$$

To introduce the dG(0)cG(1) discretization we define the following bilinear form:

$$B(v,\varphi)=\sum_{m=1}^{M}\langle v_t,\varphi\rangle_{I_m\times\Omega}+(\nabla v,\nabla\varphi)_{I\times\Omega}+\sum_{m=2}^{M}([v]_{m-1},\varphi_{m-1}^{+})_\Omega+(v_0^{+},\varphi_0^{+})_\Omega, \tag{3.4}$$

where $\langle\cdot,\cdot\rangle_{I_m\times\Omega}$ is the duality product between $L^2(I_m;W^{-1,s}(\Omega))$ and $L^2(I_m;W_0^{1,s'}(\Omega))$. We note that the first sum vanishes for $v\in X_k^0$. Rearranging the terms we obtain an equivalent (dual) expression of B:

$$B(v,\varphi)=-\sum_{m=1}^{M}\langle v,\varphi_t\rangle_{I_m\times\Omega}+(\nabla v,\nabla\varphi)_{I\times\Omega}-\sum_{m=1}^{M-1}(v_m^{-},[\varphi_k]_m)_\Omega+(v_M^{-},\varphi_M^{-})_\Omega. \tag{3.5}$$

In the two following subsections we establish global and local pointwise in space best approximation type results for the error between the solution v of the axillary equation (2.1) and its dG(0)cG(1) approximation $v_{kh}\in X_{k,h}^{0,1}$ defined as

$$B(v_{kh},\varphi_{kh})=(f,\varphi_{kh})_{I\times\Omega}+(v_0,\varphi_{kh,0}^{+})_\Omega \quad \text{for all}\ \ \varphi_{kh}\in X_{k,h}^{0,1} \tag{3.6}$$

and $v_0=0$. Since dG(0)cG(1) method is a consistent discretization we have the following Galerkin orthogonality relation:

$$B(v-v_{kh},\varphi_{kh})=0 \quad \text{for all}\ \ \varphi_{kh}\in X_{k,h}^{0,1}.$$

3.2. Global pointwise in space error estimate. In this section we prove the following global approximation result with respect to the $L^\infty(\Omega;L^2(I))$-norm.

THEOREM 3.1 (global best approximation). *Assume v and v_{kh} satisfy* (2.1) *and* (3.6)*, respectively. Then there exists a constant C independent of k and h such that for any $1\le p\le\infty$,*

$$\sup_{y\in\overline{\Omega}}\int_0^T|(v-v_{kh})(t,y)|^2dt \le C|\ln h|^2\inf_{\chi\in X_{k,h}^{0,1}}\left(\|v-\chi\|^2_{L^2(I;L^\infty(\Omega))}+h^{-\frac{4}{p}}\|\pi_k v-\chi\|^2_{L^2(I;L^p(\Omega))}\right).$$

Proof. To establish the result we use a duality argument. Let $y\in\overline{\Omega}$ be fixed, but arbitrary. First, we introduce a smoothed Delta function [38, Appendix], which we will denote by $\tilde\delta=\tilde\delta_y=\tilde\delta_y^h$. This function is supported in one cell, denoted by τ_y, and satisfies

$$(\chi,\tilde\delta)_{\tau_y}=\chi(y)\quad\forall\chi\in\mathbb{P}^1(\tau_y).$$

In addition we also have

$$\|\tilde\delta\|_{W_p^s(\Omega)}\le Ch^{-s-2(1-\frac{1}{p})},\quad 1\le p\le\infty,\quad s=0,1. \tag{3.7}$$

Thus, in particular, $\|\tilde\delta\|_{L^1(\Omega)}\le C$, $\|\tilde\delta\|_{L^2(\Omega)}\le Ch^{-1}$, and $\|\tilde\delta\|_{L^\infty(\Omega)}\le Ch^{-2}$.

We define g to be a solution to the following backward parabolic problem:

$$\begin{aligned} -g_t(t,x)-\Delta g(t,x)&=v_{kh}(t,y)\tilde\delta_y(x), && (t,x)\in I\times\Omega,\\ g(t,x)&=0, && (t,x)\in I\times\partial\Omega,\\ g(T,x)&=0, && x\in\Omega. \end{aligned} \tag{3.8}$$

Let the $g_{kh} \in X^{0,1}_{k,h}$ be dG(0)cG(1) solution defined by

$$B(\varphi_{kh}, g_{kh}) = (v_{kh}(t,y)\tilde{\delta}_y, \varphi_{kh})_{I\times\Omega} \quad \forall \varphi_{kh} \in X^{0,1}_{k,h}. \tag{3.9}$$

Then using that dG(0)cG(1) method is consistent, we have

$$\begin{aligned} \int_0^T |v_{kh}(t,y)|^2 dt &= B(v_{kh}, g_{kh}) = B(v, g_{kh}) \\ &= (\nabla v, \nabla g_{kh})_{I\times\Omega} - \sum_{m=1}^{M} (v_m, [g_{kh}]_m)_\Omega, \end{aligned} \tag{3.10}$$

where we have used the dual expression for the bilinear form B (3.5) and the fact that the last term in (3.5) can be included in the sum by setting $g_{kh,M+1} = 0$ and defining, consequently, $[g_{kh}]_M = -g_{kh,M}$. The first sum in (3.5) vanishes due to $g_{kh} \in X^{0,1}_{k,h}$. For each t, integrating by parts elementwise and using that g_{kh} is linear in the spacial variable, by the Hölder's inequality we have

$$(\nabla v, \nabla g_{kh})_\Omega = \frac{1}{2}\sum_\tau (v, [\![\partial_n g_{kh}]\!])_{\partial\tau} \le C\|v\|_{L^\infty(\Omega)} \sum_\tau \|[\![\partial_n g_{kh}]\!]\|_{L^1(\partial\tau)}, \tag{3.11}$$

where $[\![\partial_n g_{kh}]\!]$ denotes the jumps of the normal derivatives across the element faces. Next, we introduce a weight function

$$\sigma(x) = \sqrt{|x-y|^2 + h^2}. \tag{3.12}$$

One can easily check that σ satisfies the following properties:

$$\|\sigma^{-1}\|_{L^2(\Omega)} \le C|\ln h|^{\frac{1}{2}}, \tag{3.13a}$$

$$|\nabla\sigma| \le C, \tag{3.13b}$$

$$|\nabla^2\sigma| \le C|\sigma^{-1}|. \tag{3.13c}$$

From Lemma 2.4 in [35] we have

$$\sum_\tau \|[\![\partial_n g_{kh}]\!]\|_{L^1(\partial\tau)} \le C|\ln h|^{\frac{1}{2}} \left(\|\sigma\Delta_h g_{kh}\|_{L^2(\Omega)} + \|\nabla g_{kh}\|_{L^2(\Omega)}\right).$$

To estimate the term involving the jumps in (3.10), we first use the Hölder's inequality and the inverse estimate to obtain

$$\sum_{m=1}^{M} (v_m, [g_{kh}]_m)_\Omega \le c\sum_{m=1}^{M} k_m^{\frac{1}{2}}\|v_m\|_{L^p(\Omega)} k_m^{-\frac{1}{2}} h^{-\frac{2}{p}} \|[g_{kh}]_m\|_{L^1(\Omega)}. \tag{3.14}$$

Now we use the fact that (3.9) can be rewritten on the each time level as

$$(\nabla\varphi_{kh}, \nabla g_{kh})_{I_m\times\Omega} - (\varphi_{kh,m}, [g_{kh}]_m)_\Omega = (v_{kh}(t,y)\tilde{\delta}_y, \varphi_{kh})_{I_m\times\Omega},$$

or, equivalently, as

$$-k_m\Delta_h g_{kh,m} - [g_{kh}]_m = k_m v_{kh,m}(y) P_h\tilde{\delta}_y, \tag{3.15}$$

where $P_h\colon L^2(\Omega) \to V_h$ is the L^2-projection (see (3.2)) and $\Delta_h\colon V_h \to V_h$ is the discrete Laplace operator. We test (3.15) with $\varphi = -\mathrm{sgn}([g_{kh}]_m)$ and obtain

$$\|[g_{kh}]_m\|_{L^1(\Omega)} \le k_m\|\Delta_h g_{kh,m}\|_{L^1(\Omega)} + k_m\|P_h\tilde{\delta}\|_{L^1(\Omega)}|v_{kh,m}(y)|.$$

Using that the L^2-projection is stable in L^1-norm (cf. [11]), we have

$$\|P_h\tilde{\delta}\|_{L^1(\Omega)} \le C\|\tilde{\delta}\|_{L^1(\Omega)} \le C.$$

Inserting the above estimate into (3.14), we obtain

$$\begin{aligned}\sum_{m=1}^M (v_m, [g_{kh}]_m)_\Omega &\le Ch^{-\frac{2}{p}} \sum_{m=1}^M k_m^{\frac12}\|v_m\|_{L^p(\Omega)} k_m^{\frac12}\left(\|\Delta_h g_{kh,m}\|_{L^1(\Omega)} + |v_{kh,m}(y)|\right)\\ &\le Ch^{-\frac{2}{p}}\left(\sum_{m=1}^M k_m\|v_m\|^2_{L^p(\Omega)}\right)^{\frac12}\left(\sum_{m=1}^M k_m\|\Delta_h g_{kh,m}\|^2_{L^1(\Omega)} + k_m|v_{kh,m}(y)|^2\right)^{\frac12}\\ &\le Ch^{-\frac{2}{p}}\|\pi_k v\|_{L^2(I;L^p(\Omega))}\left(\int_0^T |\ln h|\|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} + |v_{kh}(t,y)|^2 dt\right)^{\frac12}.\end{aligned}$$

Combining (3.10) with the above estimate we have

$$\begin{aligned}\int_0^T |v_{kh}(t,y)|^2 dt &\le C|\ln h|^{\frac12}\left(\|v\|_{L^2(I;L^\infty(\Omega))} + h^{-\frac{2}{p}}\|\pi_k v\|_{L^2(I;L^p(\Omega))}\right)\\ &\quad\times\left(\int_0^T \|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} + \|\nabla g_{kh}\|^2_{L^2(\Omega)} + |v_{kh}(t,y)|^2 dt\right)^{\frac12}.\end{aligned} \tag{3.16}$$

To complete the proof of the theorem we need to show that

$$\int_0^T \left(\|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} + \|\nabla g_{kh}\|^2_{L^2(\Omega)}\right) dt \le C|\ln h|\int_0^T |v_{kh}(t,y)|^2 dt. \tag{3.17}$$

The above result will follow from the series of lemmas. The first lemma treats the term $\|\sigma\Delta_h g_{kh}\|^2_{L^2(I;L^2(\Omega))}$.

LEMMA 3.2. *For any $\varepsilon > 0$ there exists C_ε such that*

$$\begin{aligned}\int_0^T \|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} dt &\le C_\varepsilon\int_0^T \left(|v_{kh}(t,y)|^2 + \|\nabla g_{kh}\|^2_{L^2(\Omega)}\right) dt\\ &\quad + \varepsilon\sum_{m=1}^M k_m^{-1}\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)}.\end{aligned}$$

Proof. Equation (3.9) for each time interval I_m can be rewritten as (3.15). Testing (3.15) with $\varphi = -\sigma^2\Delta_h g_{kh}$ we have

$$\begin{aligned}\int_{I_m}\|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} dt &= -([g_{kh}]_m, \sigma^2\Delta_h g_{kh,m})_\Omega - (v_{kh}(t,y)P_h\tilde{\delta}_y, \sigma^2\Delta_h g_{kh})_{I_m\times\Omega}\\ &= -([\sigma^2 g_{kh}]_m, \Delta_h g_{kh,m})_\Omega - (v_{kh}(t,y)P_h\tilde{\delta}_y, \sigma^2\Delta_h g_{kh})_{I_m\times\Omega}\\ &= ([\nabla(\sigma^2 g_{kh})]_m, \nabla g_{kh,m})_\Omega + ([\nabla(P_h - I)\sigma^2 g_{kh}]_m, \nabla g_{kh,m})_\Omega\\ &\quad - (v_{kh}(t,y)P_h\tilde{\delta}_y, \sigma^2\Delta_h g_{kh})_{I_m\times\Omega} = J_1 + J_2 + J_3.\end{aligned}$$

We have

$$J_1 = 2(\sigma\nabla\sigma[g_{kh}]_m, \nabla g_{kh,m})_\Omega + (\sigma[\nabla g_{kh}]_m, \sigma\nabla g_{kh,m})_\Omega = J_{11} + J_{12}.$$

By the Cauchy–Schwarz inequality and using (3.13b) we get

$$J_{11} \le C\|\sigma[g_{kh}]_m\|_{L^2(\Omega)}\|\nabla g_{kh,m}\|_{L^2(\Omega)}.$$

Using the identity

$$([w_{kh}]_m, w_{kh,m})_\Omega = \frac{1}{2}\|w_{kh,m+1}\|^2_{L^2(\Omega)} - \frac{1}{2}\|w_{kh,m}\|^2_{L^2(\Omega)} - \frac{1}{2}\|[w_{kh}]_m\|^2_{L^2(\Omega)}, \tag{3.18}$$

we have

$$J_{12} = \frac{1}{2}\|\sigma\nabla g_{kh,m+1}\|^2_{L^2(\Omega)} - \frac{1}{2}\|\sigma\nabla g_{kh,m}\|^2_{L^2(\Omega)} - \frac{1}{2}\|\sigma[\nabla g_{kh}]_m\|^2_{L^2(\Omega)}.$$

Using the generalized geometric-arithmetic mean inequality for J_{11} and neglecting $-\frac{1}{2}\|\sigma[\nabla g_{kh}]_m\|^2_{L^2(\Omega)}$ in J_{12} we obtain

$$\begin{aligned} J_1 \le \frac{1}{2}\|\sigma\nabla g_{kh,m+1}\|^2_{L^2(\Omega)} - \frac{1}{2}\|\sigma\nabla g_{kh,m}\|^2_{L^2(\Omega)} + C_\varepsilon k_m\|\nabla g_{kh,m}\|^2_{L^2(\Omega)} \\ + \frac{\varepsilon}{k_m}\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)}. \end{aligned} \tag{3.19}$$

To estimate J_2, first by the Cauchy–Schwarz inequality and the approximation theory, we have

$$\begin{aligned} J_2 &= \sum_\tau([\nabla(P_h - I)\sigma^2 g_{kh}]_m, \nabla g_{kh,m})_\tau \\ &\le Ch\sum_\tau \|[\nabla^2(\sigma^2 g_{kh})]_m\|_{L^2(\tau)}\|\nabla g_{kh,m}\|_{L^2(\tau)}. \end{aligned}$$

Using that g_{kh} is piecewise linear we have

$$\nabla^2(\sigma^2 g_{kh}) = \nabla^2(\sigma^2)g_{kh} + \nabla(\sigma^2)\cdot\nabla g_{kh} \quad \text{on} \quad \tau.$$

There holds $\partial_{ij}(\sigma^2) = (\partial_i\sigma)(\partial_j\sigma) + \sigma\partial_{ij}\sigma$ and $\nabla(\sigma^2) = 2\sigma\nabla\sigma$. Thus by the properties of σ (3.13b) and (3.13c), we have

$$|\nabla^2(\sigma^2)| \le c \quad \text{and} \quad |\nabla(\sigma^2)| \le c\,\sigma.$$

Using these estimates, the fact that $h \le \sigma$, and the inverse inequality, we obtain

$$J_2 \le C\|\sigma[g_{kh}]_m\|_{L^2(\Omega)}\|\nabla g_{kh,m}\|_{L^2(\Omega)} \le C_\varepsilon k_m\|\nabla g_{kh,m}\|^2_{L^2(\Omega)} + \frac{\varepsilon}{k_m}\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)}. \tag{3.20}$$

To estimate J_3 we first show that

$$\|\sigma P_h\tilde{\delta}\|_{L^2(\Omega)} \le C. \tag{3.21}$$

By the triangle inequality we get

$$\|\sigma P_h\tilde{\delta}\|_{L^2(\Omega)} \le \|\sigma\tilde{\delta}\|_{L^2(\Omega)} + \|\sigma(P_h - I)\tilde{\delta}\|_{L^2(\Omega)}.$$

Using that the support of $\tilde{\delta}_y$ is in a single element τ_y and using (3.7), we have

$$\|\sigma\tilde{\delta}\|^2_{L^2(\Omega)} = \int_{\tau_y} |\sigma\tilde{\delta}|^2 dx \le \|\tilde{\delta}\|^2_{L^\infty(\Omega)} \int_{\tau_y} (|x-y|^2 + h^2) dx \le Ch^{-4}h^2|\tau_y| \le C.$$

Similarly, using that $\|\sigma(P_h - I)\tilde{\delta}\|_{L^2(\Omega)} \le Ch\|\sigma\nabla\tilde{\delta}\|_{L^2(\Omega)}$ and (3.7), we have

$$\|\sigma\nabla\tilde{\delta}\|^2_{L^2(\Omega)} = \int_{\tau_y} |\sigma\nabla\tilde{\delta}|^2 dx \le \|\nabla\tilde{\delta}\|^2_{L^\infty(\Omega)} \int_{\tau_y} (|x-y|^2 + h^2) dx \le Ch^{-6}h^2|\tau_y| \le Ch^{-2}.$$

This establishes (3.21). By the Cauchy–Schwarz inequality, (3.21), and the arithmetic-geometric mean inequality we obtain

$$J_3 \le C \int_{I_m} |v_{kh}(t,y)|^2 dt + \frac{1}{2} \int_{I_m} \|\sigma\Delta_h g_{kh,m}\|^2_{L^2(\Omega)} dt. \tag{3.22}$$

Using the estimates (3.19), (3.20), and (3.22) we have

$$\begin{aligned}\int_{I_m} \|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} dt \le C_\varepsilon \int_{I_m} \left(|v_{kh}(t,y)|^2 + \|\nabla g_{kh}\|^2_{L^2(\Omega)}\right) dt \\ + \frac{\varepsilon}{k_m}\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)} + \frac{1}{2}\|\sigma\nabla g_{kh,m+1}\|^2_{L^2(\Omega)} - \frac{1}{2}\|\sigma\nabla g_{kh,m}\|^2_{L^2(\Omega)}.\end{aligned}$$

Summing over m and using that $g_{kh,M+1} = 0$ we obtain the lemma. ∎

The second lemma treats the term involving jumps.

LEMMA 3.3. *There exists a constant C such that*

$$\sum_{m=1}^{M} k_m^{-1}\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)} \le C \int_0^T \left(\|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} + |v_{kh}(t,y)|^2\right) dt.$$

Proof. We test (3.15) with $\varphi = \sigma^2[g_{kh}]_m$ and obtain

$$\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)} = -(\Delta_h g_{kh}, \sigma^2[g_{kh}]_m)_{I_m\times\Omega} - (v_{kh}(t,y)P_h\tilde{\delta}, \sigma^2[g_{kh}]_m)_{I_m\times\Omega}. \tag{3.23}$$

The first term on the right-hand side of (3.23) using the geometric-arithmetic mean inequality can be easily estimated as

$$(\Delta_h g_{kh}, \sigma^2[g_{kh}]_m)_{I_m\times\Omega} \le Ck_m \int_{I_m} \|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} dt + \frac{1}{4}\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)}.$$

The last term on the right-hand side of (3.23) can easily be estimated using (3.21) as

$$(v_{kh}(t,y)P_h\tilde{\delta}, \sigma^2[g_{kh}]_m)_{I_m\times\Omega} \le Ck_m \int_{I_m} |v_{kh}(t,y)|^2 dt + \frac{1}{4}\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)}.$$

Combining the above two estimates we obtain

$$\|\sigma[g_{kh}]_m\|^2_{L^2(\Omega)} \le Ck_m \int_{I_m} \left(\|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} + |v_{kh}(t,y)|^2\right) dt.$$

Summing over m we obtain the lemma. ∎

LEMMA 3.4. *There exists a constant C such that*

$$\|\nabla g_{kh}\|^2_{L^2(I;L^2(\Omega))} \le C|\ln h| \int_0^T |v_{kh}(t,y)|^2 dt.$$

Proof. Adding the primal (3.4) and the dual (3.5) representation of the bilinear form $B(\cdot,\cdot)$, one immediately arrives at

$$\|\nabla v\|^2_{I\times\Omega} \le B(v,v) \quad \text{for all } \ v \in X_k^0;$$

see, e.g., [31]. Applying this inequality together with the discrete Sobolev inequality (see [5, Lemma 4.9.2]) results in

$$\begin{aligned}\|\nabla g_{kh}\|^2_{I\times\Omega} \le B(g_{kh},g_{kh}) = (v_{kh}(t,y)\tilde{\delta}_y, g_{kh})_{I\times\Omega} &= \int_0^T v_{kh}(t,y) g_{kh}(t,y)\,dt \\ &\le \left(\int_0^T |v_{kh}(t,y)|^2\,dt\right)^{\frac{1}{2}} \|g_{kh}\|_{L^2(I;L^\infty(\Omega))} \\ &\le c|\ln h|^{\frac{1}{2}} \left(\int_0^T |v_{kh}(t,y)|^2\,dt\right)^{\frac{1}{2}} \|\nabla g_{kh}\|_{I\times\Omega}.\end{aligned}$$

This gives the desired estimate. ∎

We proceed with the proof of Theorem 3.1. From Lemmas 3.2–3.4, it follows that

$$\begin{aligned}\int_0^T \left(\|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} + \|\nabla g_{kh}\|^2_{L^2(\Omega)}\right) dt \le C_\varepsilon |\ln h| \int_0^T |v_{kh}(t,y)|^2 dt \\ + C\varepsilon \int_0^T \|\sigma\Delta_h g_{kh}\|^2_{L^2(\Omega)} dt.\end{aligned}$$

Taking ε sufficiently small we have (3.17). From (3.16) we can conclude that

$$\int_0^T |v_{kh}(t,y)|^2 dt \le C|\ln h|^2 \left(\|v\|^2_{L^2(I;L^\infty(\Omega))} + h^{-\frac{4}{p}} \|\pi_k v\|^2_{L^2(I;L^p(\Omega))}\right)$$

for some constant C independent of h, k, and y. Using the dG(0)cG(1) method is invariant on $X^{0,1}_{k,h}$, by replacing v and v_{kh} with $v-\chi$ and $v_{kh}-\chi$ for any $\chi \in X^{0,1}_{k,h}$, by taking the supremum over y, using the triangle inequality, and using $\int_0^T |(v-\chi)(t,y)|^2 dt \le \|v-\chi\|^2_{L^2(I;L^\infty(\Omega))}$, we obtain Theorem 3.1. ∎

3.3. Local error estimate. For the error at point x_0 we are able to obtain a sharper result. For elliptic problems a similar result was obtained in [37]. As before, we denote by $B_d = B_d(x_0)$ the ball of radius d centered at x_0, and $\pi_k v = v(t_m)$.

THEOREM 3.5 (local approximation). *Assume v and v_{kh} satisfy* (2.1) *and* (3.6), *respectively, and let $d > 4h$. Then there exists a constant C independent of h, k, and d such that for any $1 \le p \le \infty$*

$$\begin{aligned}&\int_0^T |(v - v_{kh})(t,x_0)|^2 dt \\ &\quad\le C|\ln h|^3 \inf_{\chi\in X^{0,1}_{k,h}} \int_0^T \|v-\chi\|^2_{L^\infty(B_d(x_0))} + h^{-\frac{4}{p}} \|\pi_k v - \chi\|^2_{L^p(B_d(x_0))} dt \\ &\qquad + Cd^{-2}|\ln h| \int_0^T \|v - v_{kh}\|^2_{L^2(\Omega)} dt.\end{aligned} \tag{3.24}$$

Proof. As in the proof of Proposition (2.3) let $\omega(x)$ be a smooth cut-off function with the properties (2.4). Define

$$\widetilde{v}(t,x) = \omega(x)v(t,x). \tag{3.25}$$

Let $\widetilde{v}_{kh}$ be dG(0)cG(1) approximation of $\widetilde{v}$ defined by

$$B(\widetilde{v} - \widetilde{v}_{kh}, \varphi_{kh}) = 0 \quad \forall \varphi_{kh} \in X^{0,1}_{k,h}.$$

Adding and subtracting $\widetilde{v}_{kh}$, we have

$$(v - v_{kh})(t,x_0) = (\widetilde{v} - v_{kh})(t,x_0) = (\widetilde{v} - \widetilde{v}_{kh})(t,x_0) + (\widetilde{v}_{kh} - v_{kh})(t,x_0).$$

By the global best approximation result Theorem 3.1 with $\chi \equiv 0$ we have

(3.26)
$$\begin{aligned}\int_0^T |(\widetilde{v} - \widetilde{v}_{kh})(t,x_0)|^2 dt &\le C|\ln h|^2 \int_0^T \|\widetilde{v}\|^2_{L^\infty(B_{2d}(x_0))} + h^{-\frac{4}{p}} \|\pi_k \widetilde{v}\|^2_{L^p(B_{2d}(x_0))} dt \\ &\le C|\ln h|^2 \int_0^T \|v\|^2_{L^\infty(B_{2d}(x_0))} + h^{-\frac{4}{p}} \|\pi_k v\|^2_{L^p(B_{2d}(x_0))} dt.\end{aligned}$$

The discrete function

$$\psi_{kh} := \widetilde{v}_{kh} - v_{kh}$$

satisfies

$$B(\psi_{kh}, \varphi_{kh}) = 0 \quad \forall \varphi_{kh} \in X^{0,1}_{k,h}(B_d(x_0)), \tag{3.27}$$

where $X^{0,1}_{k,h}(B_d(x_0))$ is the subspace of $X^{0,1}_{k,h}$ functions that vanish outside of $B_d(x_0)$. We will need the following discrete version of the Sobolev type inequality.

LEMMA 3.6. *For any $\chi \in V_h$ and $h \le d$, there exists a constant C independent of h such that*

$$\chi(x_0) \le C|\ln h|^{\frac{1}{2}} \left(\|\nabla\chi\|_{L^2(B_{2d}(x_0))} + d^{-1}\|\chi\|_{L^2(B_{2d}(x_0))} \right).$$

Proof. The proof goes along the lines of [36, Lemma 1.1]. Let $\omega(x)$ be a smooth cut-off function as in (2.4) and let $\Gamma_{x_0}(x)$ denote Green's function for the Laplacian on $B_{2d}(x_0)$ with homogeneous Dirichlet boundary conditions. Then

$$\begin{aligned}\chi(x_0) = (\omega\chi)(x_0) &= \int_{B_{2d}(x_0)} \nabla_x \Gamma_{x_0}(x) \cdot \nabla(\omega\chi)(x) dx \\ &\le \int_{B_h(x_0)} \nabla_x \Gamma_{x_0}(x) \cdot \nabla\chi(x) dx + \int_{B_{2d}(x_0)\setminus B_h(x_0)} \nabla_x \Gamma_{x_0}(x) \cdot \nabla(\omega\chi)(x) dx \\ &:= J_1 + J_2.\end{aligned}$$

Using the estimate $|\nabla_x \Gamma_{x_0}(x)| \le \frac{C}{|x-x_0|}$ and the inverse inequality we have

$$J_1 \le C\|\nabla\chi\|_{L^\infty(B_h(x_0))} \int_{B_h(x_0)} \frac{dx}{|x-x_0|} \le Ch^{-1}\|\nabla\chi\|_{L^2(B_h(x_0))} h \le C\|\nabla\chi\|_{L^2(B_{2d}(x_0))}.$$

Similarly we have

$$\begin{aligned}J_2 &\le \|\nabla\Gamma_{x_0}\|_{L^2(B_{2d}(x_0)\setminus B_h(x_0))} \left(|\omega| \|\nabla\chi\|_{L^2(B_{2d}(x_0))} + |\nabla\omega| \|\chi\|_{L^2(B_{2d}(x_0))} \right) \\ &\le C|\ln h|^{\frac{1}{2}} \left(\|\nabla\chi\|_{L^2(B_{2d}(x_0))} + d^{-1}\|\chi\|_{L^2(B_{2d}(x_0))} \right).\end{aligned}$$

This completes the proof. ∎

Applying the above lemma with $d/4$ in the place of d, we have

(3.28)
$$\int_0^T |\psi_{kh}(t,x_0)|^2\,dt \le C|\ln h| \int_0^T \left(\|\nabla\psi_{kh}\|^2_{L^2(B_{d/2}(x_0))} + d^{-2}\|\psi_{kh}\|^2_{L^2(B_{d/2}(x_0))} \right) dt.$$

To treat $\|\nabla\psi_{kh}\|_{L^2(I;L^2(B_{d/2}(x_0)))}$ we need the following lemma.

LEMMA 3.7. *Let ψ_{kh} satisfy* (3.27)*, then there exists a constant C such that*

$$\int_0^T \|\nabla\psi_{kh}\|^2_{L^2(B_d(x_0))}dt \le Cd^{-2}\int_0^T \|\psi_{kh}\|^2_{L^2(B_{2d}(x_0))}dt.$$

Proof. Let ω be as in (2.4). Thus we have

$$\int_0^T \|\nabla\psi_{kh}\|^2_{L^2(B_d(x_0))}\,dt \le \int_0^T \|\omega\nabla\psi_{kh}\|^2_{L^2(\Omega)}\,dt.$$

We can rewrite (3.27) on each time level I_m as

$$(-\Delta_h\psi_{kh},\varphi)_{I_m\times\Omega} + ([\psi_{kh}]_{m-1},\varphi_m)_\Omega = 0 \quad \forall\varphi\in H^1_0(B_d(x_0)) \text{ and } \varphi\,|_{\Omega\setminus B_d(x_0)} = 0.$$

In other words

$$-k_m\Delta_h\psi_{kh,m} + [\psi_{kh}]_{m-1} = 0,$$

inside the ball $B_d(x_0)$. Multiplying the above equation by $\omega^2\psi_{kh,m}$ we have

$$(-\Delta_h\psi_{kh},\omega^2\psi_{kh})_{I_m\times\Omega} + ([\psi_{kh}]_{m-1},\omega^2\psi_{kh,m})_\Omega = 0.$$

Using the identity

(3.29)
$$([w_{kh}]_{m-1},w_{kh,m})_\Omega = \frac{1}{2}\|w_{kh,m}\|^2_{L^2(\Omega)} - \frac{1}{2}\|w_{kh,m-1}\|^2_{L^2(\Omega)} + \frac{1}{2}\|[w_{kh}]_{m-1}\|^2_{L^2(\Omega)},$$

the last term can be rewritten as

$$([\omega\psi_{kh}]_{m-1},\omega\psi_{kh,m})_\Omega = \frac{1}{2}\|\omega\psi_{kh,m}\|^2_{L^2(\Omega)} - \frac{1}{2}\|\omega\psi_{kh,m-1}\|^2_{L^2(\Omega)} + \frac{1}{2}\|[\omega\psi_{kh}]_m\|^2_{L^2(\Omega)}.$$

For the first term we have

$$\begin{aligned}
-(\Delta_h\psi_{kh},\omega^2\psi_{kh})_{I_m\times\Omega} &= -k_m(\Delta_h\psi_{kh,m},P_h(\omega^2\psi_{kh,m}))_\Omega \\
&= k_m(\nabla\psi_{kh,m},\nabla P_h(\omega^2\psi_{kh,m}))_\Omega \\
&= k_m(\nabla\psi_{kh,m},\nabla(\omega^2\psi_{kh,m}))_\Omega + k_m(\nabla\psi_{kh,m},\nabla(P_h(\omega^2\psi_{kh,m})-\omega^2\psi_{kh,m}))_\Omega \\
&= k_m\|\omega\nabla\psi_{kh,m}\|^2_{L^2(\Omega)} + k_m(\omega\nabla\psi_{kh,m},2\nabla\omega\psi_{kh,m})_\Omega \\
&\quad + k_m(\nabla\psi_{kh,m},\nabla(P_h(\omega^2\psi_{kh,m})-\omega^2\psi_{kh,m}))_\Omega \\
&:= \|\omega\nabla\psi_{kh,m}\|^2_{L^2(I_m;L^2(\Omega))} + J_1 + J_2.
\end{aligned}$$

Using the Cauchy–Schwarz, (3.13c), and the geometric-arithmetic mean inequalities, we have

$$\begin{aligned}
J_1 &\le Cd^{-1}\|\omega\nabla\psi_{kh}\|_{L^2(I_m;L^2(\Omega))}\|\psi_{kh}\|_{L^2(I_m;L^2(\Omega))} \\
&\le \frac{1}{4}\|\omega\nabla\psi_{kh}\|^2_{L^2(I_m;L^2(\Omega))} + Cd^{-2}\|\psi_{kh}\|^2_{L^2(I_m;L^2(\Omega))}.
\end{aligned} \tag{3.30}$$

To estimate J_2 we need the following superapproximation result which essentially follows from [15].

LEMMA 3.8 (superapproximation). *For any $\chi \in V_h$ and $\omega(x)$ as in (2.4), there exists a constant C independent of h and d such that*

$$\|\nabla(P_h(\omega^2\chi) - \omega^2\chi)\|_{L^2(\Omega)} \le Ch\left(d^{-1}\|\omega\nabla\chi\|_{L^2(\Omega)} + d^{-2}\|\chi\|_{L^2(B_{2d})}\right), \tag{3.31a}$$

$$\|P_h(\omega^2\chi) - \omega^2\chi\|_{L^2(\Omega)} \le Ch^2\left(d^{-1}\|\omega\nabla\chi\|_{L^2(\Omega)} + d^{-2}\|\chi\|_{L^2(B_{2d})}\right). \tag{3.31b}$$

By the Cauchy–Schwarz inequality, the superapproximation (3.31a), and the inverse inequality, we have

$$\begin{aligned}
J_2 &\le k_m\|\nabla\psi_{kh,m}\|_{L^2(B_{2d})}Chd^{-1}(\|\omega\nabla\psi_{kh,m}\|_{L^2(\Omega)} + d^{-1}\|\psi_{kh,m}\|_{L^2(B_{2d})})\\
&\le Ck_m\|\psi_{kh,m}\|_{L^2(B_{2d})}(d^{-1}\|\omega\nabla\psi_{kh,m}\|_{L^2(\Omega)} + d^{-2}\|\psi_{kh,m}\|_{L^2(B_{2d})})\\
&\le \frac{1}{8}\|\omega\nabla\psi_{kh}\|^2_{L^2(I_m;L^2(\Omega))} + Cd^{-2}\|\psi_{kh}\|^2_{L^2(I_m;L^2(B_{2d}))}.
\end{aligned} \tag{3.32}$$

Combining (3.30) and (3.32), we have

$$\int_{I_m}\|\omega\nabla\psi_{kh}\|^2_{L^2(\Omega)}dt + \|\omega\psi_{kh,m}\|^2_{L^2(\Omega)} - \|\omega\psi_{kh,m-1}\|^2_{L^2(\Omega)}dt \le Cd^{-2}\int_{I_m}\|\psi_{kh}\|^2_{L^2(B_{2d})}dt.$$

Summing over m we obtain Lemma 3.7. $\square$

3.4. Proof of Theorem 3.5. Applying Lemma 3.7 to (3.28) with $d/2$ instead of d, we have

$$\int_0^T |\psi_{kh}(x_0)|^2\,dt \le C|\ln h|d^{-2}\|\psi_{kh}\|^2_{L^2(I;L^2(B_d(x_0)))}.$$

Since on $B_d(x_0)$ we have $\widetilde{v} = v$, by the triangle inequality

$$\|\psi_{kh}\|_{L^2(I;L^2(B_d(x_0)))} \le \|\widetilde{v} - \widetilde{v}_{kh}\|_{L^2(I;L^2(B_d(x_0)))} + \|v - v_{kh}\|_{L^2(I;L^2(B_d(x_0)))}.$$

Using that $|B_d| \le Cd^2$, we have

$$\|\widetilde{v} - \widetilde{v}_{kh}\|_{L^2(I;L^2(B_d(x_0)))} \le Cd\,\|\widetilde{v} - \widetilde{v}_{kh}\|_{L^2(I;L^\infty(B_d(x_0)))}.$$

Applying Theorem 3.1, similarly to (3.26) we have

(3.33)

$$\begin{aligned}
d^{-2}\int_0^T \|\widetilde{v} - \widetilde{v}_{kh}\|^2_{L^2(B_d(x_0))}dt &= d^{-2}\int_0^T\int_{B_d(x_0)} |(\widetilde{v} - \widetilde{v}_{kh})(t,x)|^2dxdt\\
&= d^{-2}\int_{B_d(x_0)}\int_0^T |(\widetilde{v} - \widetilde{v}_{kh})(t,x)|^2dtdx\\
&\le C\sup_{x\in B_d(x_0)}\int_0^T |(\widetilde{v} - \widetilde{v}_{kh})(t,x)|^2dt\\
&\le C|\ln h|^2\int_0^T \|v\|^2_{L^\infty(B_{2d}(x_0))} + h^{-\frac{4}{p}}\|\pi_k v\|^2_{L^p(B_{2d}(x_0))}dt.
\end{aligned}$$

Combining (3.26) and (3.33) we have

$$\begin{aligned}
\int_0^T |(v - v_{kh})(t,x_0)|^2dt \le C|\ln h|^3\int_0^T &\left(\|v\|^2_{L^\infty(B_{2d}(x_0))} + h^{-\frac{4}{p}}\|\pi_k v\|^2_{L^p(B_{2d}(x_0))}\right)dt\\
&+ Cd^{-2}|\ln h|\int_0^T \|v - v_{kh}\|^2_{L^2(\Omega)}dt.
\end{aligned}$$

Again using that dG(0)cG(1) method is invariant on $X^{0,1}_{k,h}$, by replacing v and v_{kh} with $v-\chi$ and $v_{kh}-\chi$ for any $\chi \in X^{0,1}_{k,h}$ we obtain Theorem 3.5 with an inessential difference of having $2d$ in the place of d. ∎

4. Discretization of the optimal control problem. In this section we describe the discretization of the optimal control problem (1.1)–(1.2) and prove our main result, Theorem 1.1. We start with discretization of the state equation. For a given control $q \in Q$ we define the corresponding discrete state $u_{kh} = u_{kh}(q) \in X^{0,1}_{k,h}$ by

$$B(u_{kh}, \varphi_{kh}) = \int_0^T q(t)\varphi_{kh}(t, x_0)\, dt \quad \text{for all } \ \varphi_{kh} \in X^{0,1}_{k,h}. \tag{4.1}$$

Using the weak formulation for $u = u(q)$ from Proposition 2.2 we obtain that this discretization is consistent, i.e., the Galerkin orthogonality holds

$$B(u - u_{kh}, \varphi_{kh}) = 0 \quad \text{for all } \ \varphi_{kh} \in X^{0,1}_{k,h}.$$

Note that the jump terms involving u vanish due to the fact that $u \in C(I; W^{-\varepsilon,s}(\Omega))$ and $\varphi_{kh,m} \in W^{1,\infty}(\Omega)$.

As on the continuous level, we define the discrete reduced cost functional $j_{kh} \colon Q \to \mathbb{R}$ by

$$j_{kh}(q) = J(q, u_{kh}(q)),$$

where J is the cost function in (1.1). The discretized optimal control problem is then given as

$$\min\, j_{kh}(q), \quad q \in Q_{\mathrm{ad}}, \tag{4.2}$$

where Q_{ad} is the set of admissible controls (2.9). We note that the control variable q is not explicitly discretized; cf. [26]. With standard arguments one proves the existence of a unique solution $\bar{q}_{kh} \in Q_{\mathrm{ad}}$ of (4.2). Due to convexity of the problem, the following condition is necessary and sufficient for the optimality:

$$j'_{kh}(\bar{q}_{kh})(\delta q - \bar{q}_{kh}) \ge 0 \quad \text{for all } \ \delta q \in Q_{\mathrm{ad}}. \tag{4.3}$$

As on the continuous level, the directional derivative $j'_{kh}(q)(\delta q)$ for given $q, \delta q \in Q$ can be expressed as

$$j'_{kh}(q)(\delta q) = \int_I (\alpha q(t) + z_{kh}(t, x_0))\, \delta q(t)\, dt,$$

where $z_{kh} = z_{kh}(q)$ is the solution of the discrete adjoint equation

$$B(\varphi_{kh}, z_{kh}) = (u_{kh}(q) - \widehat{u}, \varphi_{kh}) \quad \text{for all } \ \varphi_{kh} \in X^{0,1}_{k,h}. \tag{4.4}$$

The discrete adjoint state, which corresponds to the discrete optimal control $\bar{q}_{kh}$, is denoted by $\bar{z}_{kh} = z(\bar{q}_{kh})$. The variational inequality (4.3) is equivalent to the following pointwise projection formula (cf. (2.12)):

$$\bar{q}_{kh} = P_{Q_{\mathrm{ad}}}\left(-\frac{1}{\alpha}\bar{z}_{kh}(\cdot, x_0)\right).$$

Due to the fact that $\bar{z}_{kh} \in X^{0,1}_{k,h}$, we have that $\bar{z}_{kh}(\cdot, x_0)$ is piecewise constant and, therefore, by the projection formula $\bar{q}_{kh}$ is also piecewise constant.

To prove Theorem 1.1 we first need estimates for the error in the state and in the adjoint variables for a given (fixed) control q. Due to the structure of the optimality conditions, we will have to estimate the error $\|z(\cdot, x_0) - z_{kh}(\cdot, x_0)\|_I$, where $z = z(q)$ and $z_{kh} = z_{kh}(q)$. Note that z_{kh} is not the Galerkin projection of z due to the fact that the right-hand side of the adjoint equation (2.11) involves $u = u(q)$ and the right-hand side of the discrete adjoint equation (4.4) involves $u_{kh} = u_{kh}(q)$. To obtain an estimate of optimal order, we will first estimate the error $u - u_{kh}$ with respect to the $L^2(I; L^1(\Omega))$ norm. Note that an L^2 estimate would not lead to an optimal result.

THEOREM 4.1. *Let $q \in Q$ be given, let $u = u(q)$ be the solution of the state equation* (1.2)*, and let $u_{kh} = u_{kh}(q) \in X^{0,1}_{k,h}$ be the solution of the discrete state equation* (4.1)*. Then there holds the following estimate:*

$$\|u - u_{kh}\|_{L^2(I;L^1(\Omega))} \le cd^{-1}|\ln h|^{\frac{5}{2}}(k + h^2)\|q\|_I,$$

where d is the radius of the largest ball centered at x_0 that is contained in Ω.

Proof. We denote by $e = u - u_{kh}$ the error and consider the following auxiliary dual problem:

$$\begin{aligned} -w_t(t,x) - \Delta w(t,x) &= g(t,x), && (t,x) \in I \times \Omega, \\ w(t,x) &= 0, && (t,x) \in I \times \partial\Omega, \\ w(T,x) &= 0, && x \in \Omega, \end{aligned}$$

where $g(t,x) = \operatorname{sgn}(e(t,x))\|e(t,\cdot)\|_{L^1(\Omega)}$ and the corresponding discrete solution $w_{kh} \in X^{0,1}_{k,h}$ defined by

$$B(\varphi_{kh}, w - w_{kh}) = 0 \quad \forall \varphi_{kh} \in X^{0,1}_{k,h}.$$

Using the Galerkin orthogonality for $u - u_{kh}$ and $w - w_{kh}$, we obtain

$$\begin{aligned} \int_0^T \|e(t,\cdot)\|^2_{L^1(\Omega)}\,dt &= (e, \operatorname{sgn}(e)\|e(t,\cdot)\|_{L^1(\Omega)})_{I\times\Omega} = (e, g)_{I\times\Omega} \\ &= B(u - u_{kh}, w) = B(u - u_{kh}, w - w_{kh}) \\ &= B(u, w - w_{kh}) \\ &= \int_0^T q(t)(w - w_{kh})(t, x_0)dt \\ &\le \|q\|_I \left(\int_0^T |(w - w_{kh})(t,x_0)|^2 dt \right)^{\frac{1}{2}}. \end{aligned} \tag{4.5}$$

Using the local estimate from Theorem 3.5 we obtain

$$\begin{aligned} \int_0^T |(w - w_{kh})(t,x_0)|^2 dt \le C|\ln h|^3 \int_0^T & \|w - \chi\|^2_{L^\infty(B_d(x_0))} \\ &+ h^{-\frac{4}{p}}\|\pi_k w - \chi\|^2_{L^p(B_d(x_0))} dt \\ &+ Cd^{-2}|\ln h| \int_0^T \|w - w_{kh}\|^2_{L^2(\Omega)} dt := J_1 + J_2 + J_3. \end{aligned}$$

Taking $\chi = \pi_h \pi_k w$, where π_h is the Clement interpolation by the triangle inequality and the inverse estimate, we have

$$J_1 \le C|\ln h|^3 \int_0^T \|w - \pi_h w\|^2_{L^\infty(B_d(x_0))} + \|\pi_h(w - \pi_k w)\|^2_{L^\infty(B_d(x_0))} dt$$
$$\le C|\ln h|^3 \int_0^T \|w - \pi_h w\|^2_{L^\infty(B_d(x_0))} + h^{-\frac{4}{p}} \|\pi_h(w - \pi_k w)\|^2_{L^p(B_d(x_0))} dt.$$

Using the fact that the Clement interpolation is stable with respect to any L^p-norm and the corresponding interpolation estimates (see, e.g., [4]), we obtain

$$J_1 \le C|\ln h|^3 \int_0^T h^{4-\frac{4}{p}} \|w\|^2_{W^{2,p}(B_{2d}(x_0))} + h^{-\frac{4}{p}} \|w - \pi_k w\|^2_{L^p(B_{2d}(x_0))} dt$$
$$\le Ch^{-\frac{4}{p}} |\ln h|^3 (h^4 + k^2) \int_0^T \|w\|^2_{W^{2,p}(B_{2d}(x_0))} + \|w_t\|^2_{L^p(B_{2d}(x_0))} dt.$$

J_2 can be estimated similarly since for $\chi = \pi_h \pi_k w$ by the triangle inequality we have

$$\|\pi_k w - P_h \pi_k w\|_{L^p(B_d(x_0))} \le \|\pi_k w - w\|_{L^p(B_d(x_0))} + \|w - \pi_h w\|_{L^p(B_d(x_0))} + \|\pi_h(w - \pi_k w)\|_{L^p(B_d(x_0))}.$$

This results in

$$J_1 + J_2 \le Ch^{-\frac{4}{p}} |\ln h|^3 (h^4 + k^2) \int_0^T \|w\|^2_{W^{2,p}(B_{2d}(x_0))} + \|w_t\|^2_{L^p(B_{2d}(x_0))} dt.$$

Using Lemma 2.2 we obtain

$$\int_0^T \|w\|^2_{W^{2,p}(B_{2d}(x_0))} + \|w_t\|^2_{L^p(B_{2d}(x_0))} dt \le cd^{-2} p^2 \|g\|^2_{L^2(I;L^p(\Omega))} \le cd^{-2} p^2 \|e\|^2_{L^2(I;L^1(\Omega))}.$$

For the term J_3 we obtain using an L^2-estimate from [31],

$$J_3 \le cd^{-2} |\ln h| (h^4 + k^2) \left(\|\nabla^2 w\|^2_{L^2(I;L^2(\Omega))} + \|w_t\|^2_{L^2(I;L^2(\Omega))} \right)$$
$$\le cd^{-2} |\ln h| (h^4 + k^2) \|g\|^2_{L^2(I;L^2(\Omega))}$$
$$\le cd^{-2} |\ln h| (h^4 + k^2) \|e\|^2_{L^2(I;L^1(\Omega))}.$$

Combining the estimate for J_1, J_2, and J_3 and inserting them into (4.5) we obtain

$$\|e\|_{L^2(I;L^1(\Omega))} \le c|\ln h|^{\frac{3}{2}} d^{-1} (ph^{-\frac{2}{p}} + 1)(h^2 + k).$$

Setting $p = |\ln h|$ completes the proof. ∎

In the following theorem we provide an estimate of the error in the adjoint state for fixed control q.

THEOREM 4.2. *Let $q \in Q$ be given, let $z = z(q)$ be the solution of the adjoint equation* (2.11)*, and let $z_{kh} = z_{kh}(q) \in X^{0,1}_{k,h}$ be the solution of the discrete adjoint equation* (4.4)*. Then there holds the estimate*

$$\left(\int_0^T |z(t, x_0) - z_{kh}(t, x_0)|^2 \, dt \right)^{\frac{1}{2}} \le cd^{-1} |\ln h|^{\frac{7}{2}} (k + h^2) \left(\|q\|_I + \|\widehat{u}\|_{L^2(I;L^\infty(\Omega))} \right),$$

where d is the radius of the largest ball centered at x_0 that is contained in Ω.

Proof. We introduce an intermediate adjoint state $\widetilde{z}_{kh} \in X^{0,1}_{k,h}$ defined by

$$B(\varphi_{kh}, \widetilde{z}_{kh}) = (u - \widehat{u}, \varphi_{kh}) \quad \text{for all} \;\; \varphi_{kh} \in X^{0,1}_{k,h},$$

where $u = u(q)$ and, therefore, $\widetilde{z}_{kh}$ is the Galerkin projection of z. By the local best approximation result of Theorem 3.5 for any $\chi \in X^{0,1}_{k,h}$ we have

$$\int_0^T |(\widetilde{z}_{kh} - z)(t, x_0)|^2 \, dt \leq C|\ln h|^3 \int_0^T \|z - \chi\|^2_{L^\infty(B_d(x_0))} + h^{-\frac{4}{p}} \|\pi_k z - \chi\|^2_{L^p(B_d(x_0))} dt$$
$$+ Cd^{-2}|\ln h| \int_0^T \|\widetilde{z}_{kh} - z\|^2_{L^2(\Omega)} dt := J_1 + J_2 + J_3.$$

The terms J_1, J_2, and J_3 are estimated in the same way as in the proof of Theorem 4.1 using the regularity result for the adjoint state z from Proposition 2.3. This results in

$$\left(\int_0^T |(\widetilde{z}_{kh} - z)(t, x_0)|^2 \, dt \right)^{\frac{1}{2}}$$
$$\leq c|\ln h|^{\frac{3}{2}} d^{-2}(p^2 h^{-\frac{2}{p}} + 1)(h^2 + k) \left(\|q\|_{L^2(I)} + \|\hat{u}\|_{L^2(I;L^\infty(\Omega))} \right).$$

Setting $p = |\ln h|$ we obtain

$$\left(\int_0^T |(\widetilde{z}_{kh} - z)(t, x_0)|^2 \, dt \right)^{\frac{1}{2}} \leq c|\ln h|^{\frac{7}{2}}(h^2 + k) \left(\|q\|_{L^2(I)} + \|\hat{u}\|_{L^2(I;L^\infty(\Omega))} \right). \tag{4.6}$$

It remains to estimate the corresponding error between $\widetilde{z}_{kh}$ and z_{kh}. We denote $e_{kh} = \widetilde{z}_{kh} - z_{kh} \in X^{0,1}_{k,h}$. Then we have

$$B(\varphi_{kh}, e_{kh}) = (u - u_{kh}, \varphi_{kh}) \quad \text{for all} \;\; \varphi \in X^{0,1}_{k,h}.$$

As in the proof of Lemma 3.4, we use the fact that

$$\|\nabla v\|^2_{I\times\Omega} \leq B(v, v).$$

Applying this inequality together with the discrete Sobolev inequality (see [5]) results in

$$\begin{aligned} \|\nabla e_{kh}\|^2_{I\times\Omega} &\leq B(e_{kh}, e_{kh}) = (u - u_{kh}, e_{kh}) \\ &\leq \|u - u_{kh}\|_{L^2(I;L^1(\Omega))} \|e_{kh}\|_{L^2(I;L^\infty(\Omega))} \\ &\leq c|\ln h|^{\frac{1}{2}} \|u - u_{kh}\|_{L^2(I;L^1(\Omega))} \|\nabla e_{kh}\|_{I\times\Omega}. \end{aligned}$$

Therefore, we have

$$\|\nabla e_{kh}\|_{I\times\Omega} \leq c|\ln h|^{\frac{1}{2}} \|u - u_{kh}\|_{L^2(I;L^1(\Omega))}$$

and, consequently (again by the discrete Sobolev inequality),

$$\|e_{kh}\|_{L^2(I;L^\infty(\Omega))} \leq c|\ln h| \|u - u_{kh}\|_{L^2(I;L^1(\Omega))}.$$

Using Theorem 4.1 and

$$\left(\int_0^T |e_{kh}(t,x_0)|^2 dt\right)^{1/2} \le \|e_{kh}\|_{L^2(I;L^\infty(\Omega))},$$

we obtain

$$\left(\int_0^T |e_{kh}(t,x_0)|^2 dt\right)^{1/2} \le c d^{-1} |\ln h|^{\frac{7}{2}} (k+h^2)\|q\|_I.$$

Combining this estimate with (4.6) we complete the proof. ∎

Using the result of Theorem 4.2 we proceed with the proof of Theorem 1.1.

Proof. Due to the quadratic structure of discrete reduced functional j_{kh}, the second derivative $j''_{kh}(q)(p,p)$ is independent of q and there holds

$$j''_{kh}(q)(p,p) \ge \alpha\|p\|_I^2 \quad \text{for all} \quad p \in Q. \tag{4.7}$$

Using optimality conditions (2.10) for $\bar{q}$ and (4.3) for $\bar{q}_{kh}$ and the fact that $\bar{q}, \bar{q}_{kh} \in Q_{\text{ad}}$, we obtain

$$-j'_{kh}(\bar{q}_{kh})(\bar{q}-\bar{q}_{kh}) \le 0 \le -j'(\bar{q})(\bar{q}-\bar{q}_{kh}).$$

Using coercivity (4.7), we get

$$\begin{aligned}\alpha\|\bar{q}-\bar{q}_{kh}\|_I^2 &\le j''_{kh}(\bar{q})(\bar{q}-\bar{q}_{kh},\bar{q}-\bar{q}_{kh}) = j'_{kh}(\bar{q})(\bar{q}-\bar{q}_{kh}) - j'_{kh}(\bar{q}_{kh})(\bar{q}-\bar{q}_{kh})\\ &\le j'_{kh}(\bar{q})(\bar{q}-\bar{q}_{kh}) - j'(\bar{q})(\bar{q}-\bar{q}_{kh}) = (z(\bar{q})(t,x_0)-z_{kh}(\bar{q})(t,x_0), \bar{q}-\bar{q}_{kh})_I\\ &\le \left(\int_0^T |z(\bar{q})(t,x_0)-z_{kh}(\bar{q})(t,x_0)|^2\, dt\right)^{\frac{1}{2}} \|\bar{q}-\bar{q}_{kh}\|_I.\end{aligned}$$

Applying Theorem 4.2 completes the proof. ∎

5. Numerical illustration. In this section we illustrate our main results. First we demonstrate the estimate from Theorem 4.1 for the discretization error in the state equation with respect to the $L^2(I;L^1(\Omega))$ norm. To this end we take a fixed control to be $q(t) = \frac{-1}{\sqrt{t} * \ln t}$, which is barely in $L^2(0,T)$. We choose the domain $\Omega = B_1(0)$ and $T = 0.1$. In Tables 5.1 and 5.2 for the comparison we provide the errors $u(q) - u_{kh}(q)$ in $L^2(I;L^2(\Omega))$ and $L^2(I;L^1(\Omega))$ norms for the temporal (Table 5.1) and the spatial (Table 5.2) refinement. As predicted by the theory, the errors in the $L^2(I;L^1(\Omega))$

TABLE 5.1
Errors for the state variable with 20609 *nodes in space,* 2^{k+1} *time steps.*

Time level	$L^2(I;L^2(\Omega))$ error	Rate	$L^2(I;L^1(\Omega))$ error	Rate
1	$6.541 \cdot 10^{-3}$	–	$6.353 \cdot 10^{-3}$	–
2	$4.135 \cdot 10^{-3}$	0.66	$3.408 \cdot 10^{-3}$	0.90
3	$2.619 \cdot 10^{-3}$	0.66	$1.796 \cdot 10^{-3}$	0.92
4	$1.667 \cdot 10^{-3}$	0.65	$9.356 \cdot 10^{-4}$	0.94
5	$1.067 \cdot 10^{-3}$	0.64	$4.830 \cdot 10^{-4}$	0.95
6	$6.851 \cdot 10^{-4}$	0.64	$2.474 \cdot 10^{-4}$	0.97
7	$4.395 \cdot 10^{-4}$	0.64	$1.257 \cdot 10^{-4}$	0.98

TABLE 5.2
Errors for the state variable with 4096 time steps and five levels of mesh refinement.

Mesh level	$L^2(I; L^2(\Omega))$ error	Rate	$L^2(I; L^1(\Omega))$ error	Rate
1	$1.077 \cdot 10^{-2}$	–	$8.325 \cdot 10^{-3}$	–
2	$5.435 \cdot 10^{-3}$	0.99	$2.676 \cdot 10^{-3}$	1.64
3	$2.748 \cdot 10^{-3}$	0.98	$8.147 \cdot 10^{-4}$	1.72
4	$1.393 \cdot 10^{-3}$	0.98	$2.256 \cdot 10^{-4}$	1.85
5	$6.405 \cdot 10^{-4}$	1.12	$5.093 \cdot 10^{-5}$	2.15

norm are optimal both in space and time and the errors in the $L^2(I; L^2(\Omega))$ norm are not. The mesh levels of refinements in Table 5.2 correspond to 25, 89, 337, 1313, and 5185 degrees of freedom.

The second example is taken from [23]. We consider the optimal control problem with $\Omega = B_1(0)$, $T = 1$. The control is unconstrained, i.e., $Q_{ad} = L^2(0, T)$, and the exact solution is given as

$$\bar{u}(x,t) = -\frac{1}{2\pi} t(1-t) \ln |x|, \quad \bar{q}(t) = t(1-t),$$

and

$$\bar{z}(x,t) = -t(1-t) \cos\left(\frac{\pi}{2}|x|^2\right).$$

We report the convergence rates for $\|\bar{q} - \bar{q}_{kh}\|_{L^2(I)}$ separately in time and space; see Table 5.3.

TABLE 5.3
Errors with 20609 nodes in space for time refinement (left) and error with 4096 time steps for space refinement (right).

Time level	$\|\bar{q} - \bar{q}_{kh}\|_{L^2(I)}$	Rate
1	$8.241 \cdot 10^{-2}$	–
2	$4.809 \cdot 10^{-2}$	0.78
3	$2.661 \cdot 10^{-2}$	0.85
4	$1.421 \cdot 10^{-2}$	0.91
5	$7.380 \cdot 10^{-3}$	0.95
6	$3.766 \cdot 10^{-3}$	0.97
7	$1.903 \cdot 10^{-3}$	0.99

Mesh level	$\|\bar{q} - \bar{q}_{kh}\|_{L^2(I)}$	Rate
1	$2.783 \cdot 10^{-2}$	–
2	$6.768 \cdot 10^{-3}$	2.04
3	$1.677 \cdot 10^{-3}$	2.01
4	$4.188 \cdot 10^{-4}$	2.00
5	$1.160 \cdot 10^{-4}$	1.85

Acknowledgment. We would like to thank Kontsantin Pieper for his help by creating numerical results.

REFERENCES

[1] H. AMANN, *Linear and Quasilinear Parabolic Problems. Vol.* I, in Abstract Linear Theory, in Monogr. Math. 89, Birkhäuser Boston, Boston, 1995.
[2] T. APEL, O. BENEDIX, D. SIRCH, AND B. VEXLER, *A priori mesh grading for an elliptic problem with Dirac right-hand side*, SIAM J. Numer. Anal., 49 (2011), pp. 992–1005.
[3] I. BABUŠKA, *Error-bounds for finite element method*, Numer. Math., 16 (1970/1971), pp. 322–333.
[4] C. BERNARDI AND V. GIRAULT, *A local regularization operator for triangular and quadrilateral finite elements*, SIAM J. Numer. Anal., 35 (1998), pp. 1893–1916.

[5] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite element Methods*, 3rd ed, Texts Appl. Math. 15, Springer, New York., 2008.
[6] E. Casas, *L^2 estimates for the finite element method for the Dirichlet problem with singular data*, Numer. Math., 47 (1985), pp. 627–632.
[7] E. Casas, C. Clason, and K. Kunisch, *Approximation of elliptic control problems in measure spaces with sparse solutions*, SIAM J. Control Optim., 50 (2012), pp. 1735–1752.
[8] E. Casas, C. Clason, and K. Kunisch, *Parabolic control problems in measure spaces with sparse solutions*, SIAM J. Control Optim., 51 (2013), pp. 28–63.
[9] I. Chryssoverghi, *Approximate methods for optimal pointwise control of parabolic systems*, Systems Control Lett., 1 (1981/82), pp. 216–219.
[10] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in non-reflexive Banach spaces*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 243–266.
[11] M. Crouzeix and V. Thomée, *The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces*, Math. Comp., 48 (1987), pp. 521–532.
[12] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology: Evolution Problems* I, Vol. 5, Springer-Verlag, Berlin, 1992.
[13] E. J. Dean and P. Gubernatis, *Pointwise control of Burgers' equation—a numerical approach*, Comput. Math. Appl., 22 (1991), pp. 93–100.
[14] K. Deckelnick and M. Hinze, *Variational discretization of parabolic control problems in the presence of pointwise state constraints*, J. Comput. Math., 29 (2011), pp. 1–15.
[15] A. Demlow, J. Guzmán, and A. H. Schatz, *Local energy estimates for the finite element method on sharply varying grids*, Math. Comp., 80 (2011), pp. 1–9.
[16] J. Droniou and J.-P. Raymond, *Optimal pointwise control of semilinear parabolic equations*, Nonlinear Anal., 39 (2000), pp. 135–156.
[17] J. Elschner, J. Rehberg, and G. Schmidt, *Optimal regularity for elliptic transmission problems including C^1 interfaces*, Interfaces Free Bound., 9 (2007), pp. 233–252.
[18] K. Eriksson, *Finite element methods of optimal order for problems with singular data*, Math. Comp., 44 (1985), pp. 345–360.
[19] K. Eriksson, C. Johnson, and V. Thomée, *Time discretization of parabolic problems by the discontinuous Galerkin method*, RAIRO Modél. Math. Anal. Numér., 19 (1985), pp. 611–643.
[20] L. C. Evans, *Partial Differential Equations*, Grad. Stud. Math. 19, AMS, Providence, RI, 2002.
[21] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Classics Math., Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.
[22] W. Gong, *Error estimates for finite element approximations of parabolic equations with measure data*, Math. Comp., 82 (2013), pp. 69–98.
[23] W. Gong, M. Hinze, and Z. Zhou, *A Priori Error Analysis for Finite Element Approximation of Parabolic Optimal Control Problems with Pointwise Control*, Tech. report, 2011-07, Hamburger Beiträge zur Angewandten Mathematik, Hamburg, Germany, 2011.
[24] J. A. Griepentrog, H.-C. Kaiser, and J. Rehberg, *Heat kernel and resolvent properties for second order elliptic differential operators with general boundary conditions on L^p*, Adv. Math. Sci. Appl., 11 (2001), pp. 87–112.
[25] P. Grisvard, *Singularities in Boundary Value Problems*, Springer-Verlag, Masson, Paris, Berlin, 1992.
[26] M. Hinze, *A variational discretization concept in control constrained optimization: The linear-quadratic case*, Comput. Optim. Appl., 30 (2005), pp. 45–61.
[27] D. Jerison and C. E. Kenig, *The inhomogeneous Dirichlet problem in Lipschitz domains*, J. Funct. Anal., 130 (1995), pp. 161–219.
[28] K. Kunisch and B. Vexler, *Constrained Dirichlet boundary control in L^2 for a class of evolution equations*, SIAM J. Control Optim., 46 (2007), pp. 1726–1753.
[29] J.-L. Lions and E. Magenes, *Nonhomogeneous Boundary Value Problems and Applications Vol.* II, Springer-Verlag, New York, Heidelberg, 1972.
[30] 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 (2011), pp. 1961–1997.
[31] D. Meidner and B. Vexler, *A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part* I: *Problems without control constraints*, SIAM J. Control Optim., 47 (2008), pp. 1150–1177.
[32] D. Meidner and B. Vexler, *A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part* II: *Problems with control constraints*, SIAM J. Control Optim., 47 (2008), pp. 1301–1329.
[33] K. Pieper and B. Vexler, *A priori error analysis for discretization of sparse elliptic optimal control problems in measure space*, SIAM J. Control Optim., 51 (2013), pp. 2788–2808.

[34] 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.

[35] R. Rannacher, *L^∞-stability estimates and asymptotic error expansion for parabolic finite element equations*, in Extrapolation and Defect Correction, Bonner Math. Schriften 228, University of Bonn, Bonn, 1991, pp. 74–94.

[36] A. H. Schatz, V. C. Thomée, and L. B. Wahlbin, *Maximum norm stability and error estimates in parabolic finite element equations*, Comm. Pure Appl. Math., 33 (1980), pp. 265–304.

[37] A. H. Schatz and L. B. Wahlbin, *Interior maximum norm estimates for finite element methods*, Math. Comp., 31 (1977), pp. 414–442.

[38] A. H. Schatz and L. B. Wahlbin, *Interior maximum-norm estimates for finite element methods.* II, Math. Comp., 64 (1995), pp. 907–928.

[39] R. Scott, *Finite element convergence for singular data*, Numer. Math., 21 (1973/74), pp. 317–327.