SIAM J. NUMER. ANAL.
Vol. 50, No. 4, pp. 2012–2038

LOCAL ERROR ANALYSIS OF DISCONTINUOUS GALERKIN METHODS FOR ADVECTION-DOMINATED ELLIPTIC LINEAR-QUADRATIC OPTIMAL CONTROL PROBLEMS*

DMITRIY LEYKEKHMAN† AND MATTHIAS HEINKENSCHLOSS‡

Abstract. This paper analyzes the local properties of the symmetric interior penalty upwind discontinuous Galerkin (SIPG) method for the numerical solution of optimal control problems governed by linear reaction-advection-diffusion equations with distributed controls. The theoretical and numerical results presented in this paper show that for advection-dominated problems the convergence properties of the SIPG discretization can be superior to the convergence properties of stabilized finite element discretizations such as the streamline upwind Petrov Galerkin (SUPG) method. For example, we show that for a small diffusion parameter the SIPG method is optimal in the interior of the domain. This is in sharp contrast to SUPG discretizations, for which it is known that the existence of boundary layers can pollute the numerical solution of optimal control problems everywhere even into domains where the solution is smooth and, as a consequence, in general reduces the convergence rates to only first order. In order to prove the nice convergence properties of the SIPG discretization for optimal control problems, we first improve local error estimates of the SIPG discretization for single advection-dominated equations by showing that the size of the numerical boundary layer is controlled not by the mesh size but rather by the size of the diffusion parameter. As a result, for small diffusion, the boundary layers are too "weak" to pollute the SIPG solution into domains of smoothness in optimal control problems. This favorable property of the SIPG method is due to the weak treatment of boundary conditions, which is natural for discontinuous Galerkin methods, while for SUPG methods strong imposition of boundary conditions is more conventional. The importance of the weak treatment of boundary conditions for the solution of advection dominated optimal control problems with distributed controls is also supported by our numerical results.

Key words. optimal control, advection-diffusion equations, discontinuous Galerkin methods, discretization, local error estimates, distributed control

AMS subject classifications. 49M25, 49K20, 65N30, 65N15, 65J10

DOI. 10.1137/110826953

1. Introduction.

We analyze the convergence behavior of symmetric interior penalty upwind discontinuous Galerkin (SIPG) methods for the numerical solution of linear-quadratic optimal control problems governed by advection dominated elliptic partial differential equations (PDEs) with distributed controls. In particular, we show that for a small diffusion parameter the SIPG method is optimal in the interior of the domain. This is in sharp contrast to stabilized continuous finite element discretizations. For example, we have shown in [17] that underresolved boundary layers in streamline upwind Petrov Galerkin (SUPG) methods can pollute the numerical solution of optimal control problems everywhere even into domains where the solution is smooth. In order to prove the favorable convergence properties of the SIPG discretization for optimal control problems, we also improve local error estimates in [16] for the SIPG discretization for single advection-dominated PDEs. We demonstrate numerically that the favorable convergence properties of the SIPG method is

*Received by the editors March 8, 2011; accepted for publication (in revised form) April 20, 2012; published electronically August 15, 2012.
http://www.siam.org/journals/sinum/50-4/82695.html
†Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009 (leykekhman@math.uconn.edu). This author's work was supported in part by NSF grant DMS-0811167.
‡Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892 (heinken@rice.edu). This author's work was supported in part by NSF grant DMS-0915238 and AFOSR grant FA9550-09-1-0225.

due to the weak treatment of boundary conditions which is natural for discontinuous Galerkin methods, while for SUPG methods strong imposition of boundary conditions is more conventional. Another important aspect of this work is that we estimate the discretization error in local norms. This is crucial for advection dominated problems since the constants in these local estimates depend only on the solution and its derivatives in regions where these are well behaved. Almost all other convergence analyses for advection dominated optimal control problems use global norms which are not very informative, since the constants in these estimates involve the derivatives of the solution in boundary layers and can be huge.

Let Ω be a bounded open, convex domain in $\mathbb{R}^2$ (or in $\mathbb{R}$) and $\Gamma = \partial\Omega$. We consider the model problem

$$\text{minimize } \frac{1}{2}\int_\Omega (y(x)-\widehat{y}(x))^2 dx + \frac{\alpha}{2}\int_\Omega u^2(x)dx \tag{1.1a}$$

subject to

$$-\varepsilon\Delta y(x) + \beta\cdot\nabla y(x) + r(x)y(x) = f(x)+u(x), \qquad x\in\Omega, \tag{1.1b}$$

$$y(x) = d(x), \qquad x\in\Gamma, \tag{1.1c}$$

where $f, r, \widehat{y}, d$ are given functions, the advection $\beta\in\mathbb{R}^2$ is constant, diffusion, and regularization parameters $\varepsilon, \alpha > 0$ are given scalars. We refer to u as the control, to y as the state, and to (1.1b), (1.1c) as the state equation.

Discontinuous Galerkin (DG) methods are attractive alternatives to stabilized continuous finite element methods for the numerical solution of advection-diffusion-reaction PDEs [2, 7, 9, 10, 15, 16, 19, 20, 31] because, e.g., they provide greater flexibility to locally adapt the mesh or the polynomial degree of the basis functions which implies better ability to capture fine scales of the solution. The literature on DG methods for advection diffusion PDEs is already substantial and the research in this area is still active. Surprisingly, there are almost no theoretical or numerical analyses of DG methods for the spatial discretization of optimal control problems such as (1.1). See [8, 30] for some work in this area. Almost all analyses of discretization methods for advection dominated optimal control problems has focused on stabilized finite element methods. See, e.g., [1, 4, 5, 12, 17, 18, 24].

The analysis of discretization schemes for advection dominated optimal control problems is particularly important for reasons which are related to the fact that the numerical solution of such optimal control problems requires the solution of an optimality system which consists of the state equation (1.1b), (1.1c), the adjoint equation

$$-\varepsilon\Delta\lambda(x) - \beta\cdot\nabla\lambda(x) + r(x)\lambda(x) = -(y(x)-\hat{y}(x)), \qquad x\in\Omega, \tag{1.2a}$$

$$\lambda(x) = 0, \qquad x\in\Gamma, \tag{1.2b}$$

and the equation

$$\lambda(x) = \alpha u(x), \qquad x\in\Omega. \tag{1.3}$$

Like the original state equation (1.1b), (1.1c), the adjoint equation (1.2) is also an advection-diffusion equation, but with advection $-\beta$ instead of β. Discretization methods applied to advection dominated optimal control problems can introduce inconsistencies in the discretization of the adjoint PDE which can negatively impact

the convergence behavior. See, e.g., [5, 11, 12, 22]. Additionally, as a result of the transport of information in the optimality system in the direction of the advection in the state PDE as well as in the direction of its negative in the adjoint PDE, the convergence properties of a discretization method applied to the optimal control problem can be substantially different from the convergence properties of the discretization method applied to a single advection-dominated PDE. In [17] we provided a detailed local convergence analysis for SUPG applied to advection dominated optimal control problems. In particular, we have shown in [17] that any boundary layer in either state or adjoint PDE pollutes the numerical solution everywhere in the entire domain, even in subregions where the exact solution is smooth. This reduces the order of convergence to only first order. This is in sharp contrast to the case of a single PDE, where it has been shown analytically that neither layers pollute the numerical solution into domain of smoothness (see, e.g., [29]). As we have mentioned earlier, the goal of this paper is to show that the SIPG methods do not suffer from a deterioration in the order convergence and that the SIPG method is optimal in the interior of the domain.

We estimate the discretization error in local norms, i.e., we derive a priori bounds for the error localized in subdomains $\Omega_0 \subset \Omega$ away from regions where layers occur. The right-hand sides in our error bounds involve derivatives of the solution y, u, λ of (1.1) restricted to $\Omega_0 \subset \Omega$. Since interior or boundary layers of the solution are located outside Ω_0, the right-hand sides of our bounds are independent of ε. Therefore, our local error bounds are much more descriptive than the error bounds in [4, 5, 11, 12, 18, 24, 30], which use global norms. The constants in these global norm estimates involve the derivatives of the solution in boundary layers and can be huge.

We show that interior layers do not pollute the numerical solution obtained using SIPG into subdomains of smoothness. This nice property was also shown in [17] for the SUPG method. However, in the presence of boundary layers the situation is more complicated and the convergence properties of the SIPG and SUPG discretizations differ dramatically. We show that if $\varepsilon \ll h$, the layers are too "weak" to pollute the numerical solution obtained using SIPG and, for example, the convergence rates in the L^2-norm are optimal almost until the error is of order ε. This is in sharp contrast to the SUPG method, where only first order convergence rates in general can be expected. The explanation of this strange at first fact lies in treatment of Dirichlet boundary conditions. The SIPG method naturally enforces the boundary conditions weakly, while for SUPG methods strong imposition of the boundary conditions is more common. For small ε the numerical solution must not only approximate the exact solution, but also the solution to the reduced problem, which can be formally obtained by taking $\varepsilon = 0$. The reduced version of the state equation only requires Dirichlet boundary conditions on the inflow part of the boundary, but no conditions are imposed on the reduced state at the outflow part of the boundary. Thus the weak treatment of Dirichlet boundary seems more advantageous since it does not fix the numerical solution there. There have been several publications advocating weak treatment of boundary conditions for advection-dominating problems even for the SUPG method [3, 14, 26, 27].

If we take for granted that the DG solution well approximates the reduced problem, then the numerical boundary layers are not of order h, which one would naturally expect, but of order ε. For $\varepsilon \ll h$ this is quite remarkable. It means that the numerical layer is deep inside a skin layer of width of just a single element. This paper gives a theoretical justification of this observation. In particular, we improve the local error estimates for a single equation of Guzmán [16]. We show that the boundary layers do not pollute the numerical solution into subdomains which are of order ε distance

away from outflow boundary. In [16] the subdomain Ω_0 had to be of order h distance away form the boundary. This "small" improvement has important consequences for optimal control problems. It implies that the pollution from the numerical boundary layers that propagates into the domain is insignificant for mesh sizes $\varepsilon \ll h$ and, consequently, for $\varepsilon \ll h$ is too weak to affect the convergence rates.

We analyze the SIPG method, which has the property that the two discretization strategies *optimize-then-discretize* and *discretize-then-optimize* lead to the same discrete systems. The same is true for several other DG methods, but not all [22]. For those DG methods for which optimize-then-discretize and discretize-then-optimize lead to the same result, we expect a similar error behavior as established for SIPG in the sense that convergence in regions away from boundary or interior layers is equal to what one can observe for these methods applied to PDEs with small advection. The issue becomes a bit more involved for methods like the nonsymmetric interior penalty discontinuous Galerkin method (NIPG) for which optimize-then-discretize and discretize-then-optimize lead to different systems. Typically, the discretize-then-optimize approach leads to reduced convergence rates. For the optimize-then-discretize approach, we expect that these methods exhibit a similar error behavior as SIPG in the sense specified above. See, e.g., [11, 17, 22].

The rest of this paper is organized as follows. In the next section we state the problem and the standard existence and regularity results. In section 3 we describe the DG method. This section mainly introduces the notation used in this paper and collects some basic results on DG needed in subsequent parts. Section 4 is devoted to the analysis of the SIPG method applied to the state equation. The main result of this section is Theorem 4.5, which improves the result of [16]. The SIPG discretization error for the optimal control problem is analyzed in section 5. The main result in the presence of interior layers is Theorem 5.1. The central result in the presence of boundary layers is Theorem 5.2, which establishes optimal order convergence for $\varepsilon \ll h$. Due to rather technical proofs, we only treat the problems with constant advection field β and in two dimensions. With appropriate changes the analysis can be extended to three dimensions. Using techniques similar to the ones in [2] it seems possible to relax the restriction of constant advection field. However, this would make this paper even more technical and distract from the main points of our analysis. Finally, in section 6 we provide numerical illustrations of our theoretical findings. In addition we demonstrate that if we impose the boundary conditions strongly in DG methods, the numerical layer become of order h, even for $\varepsilon \ll h$, and the pollution of order h spreads across the domain and reduces the convergence rates to the first order even far away from the layers and for high order elements. This is exactly what one observed in [17] for the SUPG method.

2. Optimal control problem. In this section we give the precise statement of the optimal control problem (1.1) and we collect well-known results on the existence, uniqueness, and characterization of solutions. The problem set-up is identical to that in [17]. We repeat the problem specification and some basic results for completeness. The results in this section hold for domains $\Omega \subset \mathbb{R}^n$ and $\beta \in \mathbb{R}^n$.

We define the state and control space

$$Y = \left\{ y \in H^1(\Omega) \,:\, y = d \text{ on } \Gamma \right\}, \quad U = L^2(\Omega) \tag{2.1}$$

and space of test functions

$$V = \left\{ v \in H^1(\Omega) \,:\, v = 0 \text{ on } \Gamma \right\}. \tag{2.2}$$

The weak form of the state equations (1.1b), (1.1c) is given by

$$a(y,v) + b(u,v) = \langle f, v\rangle \quad \forall v \in V, \tag{2.3}$$

where

$$a(y,v) = \int_\Omega \varepsilon \nabla y(x)\cdot\nabla v(x) + \beta\cdot\nabla y(x)v(x) + r(x)y(x)v(x)dx, \tag{2.4a}$$

$$b(u,v) = -\int_\Omega u(x)v(x)dx, \quad \langle f, v\rangle = \int_\Omega f(x)v(x)dx. \tag{2.4b}$$

We are interested in the solution of the optimal control problem

$$\text{minimize} \quad \frac{1}{2}\|y-\widehat{y}\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}, \tag{2.5a}$$

$$\text{subject to} \quad a(y,v) + b(u,v) = \langle f, v\rangle \quad \forall v \in V, \quad y \in Y,\ u \in U. \tag{2.5b}$$

We assume that

$$f, \widehat{y} \in L^2(\Omega), \quad \beta \in \mathbb{R}^n, \quad r \in L^\infty(\Omega), \quad d \in H^{3/2}(\Gamma), \quad \alpha > 0, \quad \varepsilon > 0, \tag{2.6a}$$

and

$$r(x) \geq r_0 \geq 0 \text{ a.e. in } \Omega. \tag{2.6b}$$

The constant r_0 is introduced to better trace the impact of the reaction term in some estimates (cf. (4.7) and (4.10)).

Under the assumptions (2.6), the bilinear form $a(\cdot,\cdot)$ is continuous on $V\times V$ and V-elliptic. The theory in [23, sect. II.1] guarantees the existence of a unique solution $(y,u) \in Y \times U$ of (2.5) and also provides necessary and, for our model problem, sufficient optimality conditions. These are stated in the following result.

THEOREM 2.1. *If* (2.6) *are satisfied, the optimal control problem* (2.5) *has a unique solution* $(y,u) \in Y\times U$. *The functions* $(y,u)\in Y\times U$ *solve* (2.5) *if and only if there exists an adjoint* $\lambda \in V$ *such that*

$$a(\psi,\lambda) = -\langle y-\hat{y},\psi\rangle \qquad \forall \psi \in V, \tag{2.7a}$$

$$b(w,\lambda) + \alpha\langle u, w\rangle = 0 \qquad \forall w \in U, \tag{2.7b}$$

$$a(y,v) + b(u,v) = \langle f, v\rangle \qquad \forall v \in V \tag{2.7c}$$

holds. The optimality system (2.7) *has a unique solution* $(y,u,\lambda) \in Y\times U\times V$.

The equation (2.7a) is the weak form of the so-called adjoint equation (1.2) It is also an advection-diffusion equation, but advection is now given by $-\beta$. The equation (2.7b) simply means that $\lambda(x) = \alpha u(x)$, $x\in\Omega$ (cf. (1.3)). The state equation (2.7c) is the weak form of (1.1b).

Under our regularity assumptions (2.6a) on the data, the following regularity result for the optimal states and corresponding adjoints is proven in [17].

THEOREM 2.2. *Let* Ω *be a bounded open convex subset of* $\mathbb{R}^n$ *and let the assumptions* (2.6) *be satisfied. There exists a positive constant* C *independent of* ε *such that the unique solution of the optimal control problem* (2.5) *and the associated adjoint satisfy* $(y,\lambda) \in H^2(\Omega)\times H^2(\Omega)$ *and*

$$\varepsilon^{3/2}\|y\|_{H^2(\Omega)} + \varepsilon^{1/2}\|y\|_{H^1(\Omega)} + \|y\|_{L^2(\Omega)} \leq C,$$
$$\varepsilon^{3/2}\|\lambda\|_{H^2(\Omega)} + \varepsilon^{1/2}\|\lambda\|_{H^1(\Omega)} + \|\lambda\|_{L^2(\Omega)} \leq C.$$

3. Discontinuous Galerkin discretization. From now on we restrict the discussion to a bounded domain $\Omega \subset \mathbb{R}^2$. We discretize the system (2.7) by a discontinuous Galerkin (DG) finite element method. More specifically, the diffusion part will be discretized using the symmetric interior penalty Galerkin (SIPG) method and the advection-reaction part will be discretized by the upwind method. This section establishes notation used in the remainder of this section and collects some basic properties of the SIPG method.

Let $T = \{T_h\}_h$ be a family of conforming quasi-uniform triangulations such that $\overline{\Omega} = \cup_{\tau \in T_h} \overline{\tau}$, $\tau_i \cap \tau_j = \emptyset$ for $\tau_i, \tau_j \in T_h$, $i \neq j$. We set $h = \max_{\tau \in T_h} \operatorname{diam}(\tau)$. The assumption that the triangulations are conforming can be relaxed in the formulation of the discontinuous Galerkin discretization.

For an integer l and $\tau \in T_h$ let $\mathbb{P}^l(\tau)$ be the set of all polynomials on τ of degree at most l. We define the discrete state and control spaces to be

$$V_h = Y_h \stackrel{\text{def}}{=} \left\{ y \in L^2(\Omega) \,:\, y\,|_\tau \in \mathbb{P}^k(\tau) \quad \forall \tau \in T_h \right\}, \tag{3.1a}$$

$$U_h \stackrel{\text{def}}{=} \left\{ u \in L^2(\Omega) \,:\, u\,|_\tau \in \mathbb{P}^l(\tau) \quad \forall \tau \in T_h \right\}, \tag{3.1b}$$

respectively. The orders $k, l \in \mathbb{N}$ of the finite element approximation can be different for the states and the controls. Note that since discontinuous Galerkin methods impose boundary conditions weakly, the space Y_h of discrete states and the space of test functions V_h are identical. To emphasize the connection between (2.7) and its discretization by a discontinuous Galerkin method we use both Y_h and V_h.

We split the set of all edges $\mathcal{E}_h$ into the set $\mathcal{E}_h^0$ of interior edges of T_h and the set $\mathcal{E}_h^\partial$ of boundary edges so that $\mathcal{E}_h = \mathcal{E}_h^\partial \cup \mathcal{E}_h^0$. Let $\mathbf{n}$ denote the unit outward normal to Ω. We further decompose the boundary edges into edges $\mathcal{E}_h^-$ that correspond to inflow boundary

$$\Gamma^- \stackrel{\text{def}}{=} \{ x \in \partial\Omega \,:\, \beta \cdot \mathbf{n}(x) < 0 \},$$

and edges $\mathcal{E}_h^+$ that correspond to the outflow boundary $\Gamma^+ \stackrel{\text{def}}{=} \partial\Omega \setminus \Gamma^-$. That is, we decompose $\mathcal{E}_h^\partial = \mathcal{E}_h^+ \cup \mathcal{E}_h^-$, where $\mathcal{E}_h^+ \stackrel{\text{def}}{=} \mathcal{E}_h^\partial \setminus \mathcal{E}_h^-$ and $\mathcal{E}_h^- \stackrel{\text{def}}{=} \left\{ e \in \mathcal{E}_h^\partial \,:\, e \subset \Gamma^- \right\}$.

For $e \in \mathcal{E}_h^0$ we define the averages and jumps of $y \in Y_h$ by

$$[\![y]\!]_e(x) = \lim_{\delta \to 0^+} \left(y(x - \delta \mathbf{n}_e) - y(x + \delta \mathbf{n}_e) \right) \tag{3.2a}$$

$$\{ \nabla_h y \}_e(x) = \frac{1}{2} \lim_{\delta \to 0^+} \left(\mathbf{n}_e \cdot \nabla y(x - \delta \mathbf{n}_e) + \mathbf{n}_e \cdot \nabla y(x + \delta \mathbf{n}_e) \right), \tag{3.2b}$$

where $\mathbf{n}_e$ is one of the normal unit vectors to e. For $e \in \mathcal{E}_h^\partial$

$$[\![y]\!]_e(x) = \lim_{\delta \to 0^+} y(x - \delta \mathbf{n}_e), \quad \{ \nabla_h y \}_e(x) = \lim_{\delta \to 0^+} \mathbf{n}_e \cdot \nabla y(x - \delta \mathbf{n}_e), \tag{3.2c}$$

where $\mathbf{n}_e$ is the outward normal unit vector to the boundary of Ω. Finally, we define $y^\pm(x) = \lim_{\delta \to 0^+} y(x \pm \delta\beta)$.

We use the following inner product and (semi-)norms. Let $D \subset \overline{\Omega}$. For an integer k and a multi-index α we define $(f, g)_D = \int_D fg$, $\|f\|_D^2 = (f, g)_D$, $|f|_{k,D}^2 = \sum_{|\alpha|=k} \int_D |D^\alpha f|^2$, and $\|f\|_{k,D}^2 = \sum_{|\alpha| \leq k} \int_D |D^\alpha f|^2$. If $D = \Omega$, we will drop the subscripts.

3.1. Discontinuous Galerkin discretization of the state equation. In this section we review discontinuous Galerkin discretizations of the state equation (1.1b) for a fixed control u. We follow [19]. As mentioned before, the diffusion part is

discretized using the SIPG method and the advection-reaction part is discretized by the upwind method. By h_e we denote the length of an edge $e \in \mathcal{E}_h$ and σ is a positive parameter to be determined later.

For $y, v \in V_h$, $u \in U_h$, and a constant advection field β we define

$$\begin{aligned} a_h(y,v) =& \varepsilon \sum_{\tau\in T_h} (\nabla y, \nabla v)_\tau \\ &+ \varepsilon \sum_{e\in\mathcal{E}_h} \left(\frac{\sigma}{h_e}([\![y]\!],[\![v]\!])_e - (\{\nabla_h y\},[\![v]\!])_e - ([\![y]\!],\{\nabla_h v\})_e \right) \\ &+ \sum_{\tau\in T_h} (\beta\cdot\nabla y + ry, v)_\tau \\ &+ \sum_{e\in\mathcal{E}_h^0} \left(y^+ - y^-, |\mathbf{n}\cdot\beta| v^+\right)_e + \sum_{e\in\mathcal{E}_h^-} \left(y^+, v^+ |\mathbf{n}\cdot\beta|\right)_e, \end{aligned} \tag{3.3}$$

$$\begin{aligned} b_h(u,v) =& -\sum_{\tau\in T_h} (u,v)_\tau, \\ l_h(v) =& \sum_{\tau\in T_h} (f,v)_\tau + \varepsilon \sum_{e\in\mathcal{E}_h^\partial} \left(\frac{\sigma}{h_e}(d,[\![v]\!])_e - (d,\{\nabla_h v\})_e \right) + \sum_{e\in\mathcal{E}_h^-} \left(d, v^+|\mathbf{n}\cdot\beta|\right)_e. \end{aligned}$$

The DG discretization of the state equation (1.1b) for a fixed control u is now given as follows (cf., (2.3)). Find $y_h \in V_h$ such that

$$a_h(y_h, v) + b_h(u,v) = l_h(v) \qquad \forall v \in V_h. \tag{3.4}$$

Since β is a constant vector, we have $(\beta\cdot\nabla v)v = \frac{1}{2}\beta\cdot\nabla(v^2)$. Using integration by parts one can show (cf. [20])

$$\begin{aligned} a_h(y,y) \geq \|y\|_{DG}^2 \stackrel{\text{def}}{=}& \sum_{\tau\in T_h} \left(\varepsilon\|\nabla y\|_\tau^2 + r_0\|y\|_\tau^2\right) + \sum_{e\in\mathcal{E}_h} \frac{\varepsilon}{h_e}\|[\![y]\!]\|_e^2 \\ &+ \sum_{e\in\mathcal{E}_h^\partial} \frac{1}{2}\|y|\mathbf{n}\cdot\beta|^{1/2}\|_e^2 + \sum_{e\in\mathcal{E}_h^0} \frac{1}{2}\|(y^+ - y^-)|\mathbf{n}\cdot\beta|^{1/2}\|_e^2, \end{aligned} \tag{3.5}$$

provided σ is large enough.

3.2. Discontinuous Galerkin discretization of the optimal control problem. Our DG discretization of the optimal control problem (2.5) is given by

$$\text{minimize} \quad \frac{1}{2}\sum_{\tau\in T_h} \|y_h - \widehat{y}\|_\tau^2 + \frac{\alpha}{2}\sum_{\tau\in T_h}\|u_h\|_\tau^2 \tag{3.6a}$$

$$\begin{aligned} \text{subject to} \quad & a_h(y_h, v) + b_h(u_h, v) = l_h(v) \quad \forall v \in V_h, \\ & (y_h, u_h) \in Y_h \times U_h. \end{aligned} \tag{3.6b}$$

Since the bilinear form $a_h(\cdot,\cdot)$ satisfies (3.5), the same technique used to prove Theorem 2.1 can be applied to establish the following counterpart for the discretized problem (3.6).

THEOREM 3.1. *The discretized optimal control problem* (3.6) *has a unique solution* $y_h \in Y_h$, $u_h \in U_h$. *The functions* $(y_h, u_h) \in Y_h \times U_h$ *solve* (3.6) *if and only if there exists a discrete adjoint* $\lambda_h \in V_h$ *such that*

$$a_h(v, \lambda_h) = -(y - \widehat{y}, v) \qquad \forall v \in V_h, \tag{3.7a}$$

$$(w, \lambda_h)_\tau = \alpha(u_h, w)_\tau \qquad \forall \tau \in T_h, \forall w \in U_h, \tag{3.7b}$$

$$a_h(y_h, v) + b_h(u_h, v) = l_h(v) \qquad \forall v \in V_h \tag{3.7c}$$

holds. The optimality system (3.7) *of the DG discretized optimal control problem* (3.6) *has a unique solution* $(y_h, u_h, \lambda_h) \in Y_h \times U_h \times V_h$.

It is of interest whether the optimality system (3.7) of the DG discretized optimal control problem (3.6) is equivalent to the DG discretization of the optimality system (2.7). This is not the case for many stabilized finite element methods and may negatively impact the convergence properties of the method in the optimal control context (see, e.g., [1, 4, 5, 11]). The paper [22] studies a large number of DG methods and identifies whether they have this property or not. In particular, it is show in [22] that for the DG method applied in this paper the DG discretization of the optimality system (2.7) is the optimality system (3.7) of the DG discretized optimal control problem (3.6).

PROPOSITION 3.2. *The optimality system* (2.7) *discretized by the SIPG method is identical to the optimality system* (3.7) *of the DG discretized optimal control problem* (3.6).

3.3. Trace and inverse inequalities. We will frequently use the following trace and inverse inequalities. For $\tau \in \mathcal{T}_h$ and $v \in V$, $v_h \in V_h$ there exist positive constants C_{tr} and C_{inv} independent of τ and v, v_h such that

$$\|v\|_{\partial\tau} \le C_{tr}\big(h^{-1/2}\|v\|_\tau + h^{1/2}\|\nabla v\|_\tau\big), \tag{3.8a}$$

$$\|\nabla v_h\|_\tau \le C_{inv} h^{-1}\|v_h\|_\tau, \tag{3.8b}$$

$$\|v_h\|_{\partial\tau} \le C_{tr}(1 + C_{inv})h^{-1/2}\|v_h\|_\tau. \tag{3.8c}$$

3.4. Coercivity of the bilinear form. In the advection-dominated case we can equip Y_h with a stronger norm than the DG norm defined in (3.5) (cf. [2, sect. 4] or [15, sect. 5]). This norm, which will allow us to provide stronger estimates for the gradient of the error in the direction of the advection β, is given by

$$|||y|||^2 \stackrel{\text{def}}{=} \|y\|_{DG}^2 + \sum_{\tau \in T_h} h_\tau \|\beta \cdot \nabla y\|_\tau^2. \tag{3.9}$$

The following result is proven in [15, Lemma A.1]).

LEMMA 3.3. *There exist constants* C_1 *and* K *such that for all* $y \in Y_h$,

$$C_1|||y|||^2 \le a_h(y, Ky + h\beta \cdot \nabla y).$$

The proof in [15] uses the fact that β is either constant or linear. A more general result can be found in [2, Thm. 4.7].

4. Local error estimates for the state equation. In this section we consider the uncontrolled ($u = 0$) state equation

$$-\varepsilon\Delta y(x) + \beta \cdot \nabla y(x) + r(x)y(x) = f(x), \qquad x \in \Omega, \tag{4.1a}$$

$$y(x) = d(x), \qquad x \in \Gamma. \tag{4.1b}$$

Global error analysis of DG methods for advection-diffusion-reaction equations have been derived in a number of papers; see, e.g., [2, 15, 19]. The estimates for the error

$$e = y - y_h$$

derived in these papers are essentially of the form

$$|||e||| \leq Ch^{k+1/2}\|y\|_{H^{k+1}(\Omega)}.$$

However the presence of layers makes such estimates rather meaningless for advection dominated problems, since $\|y\|_{H^{k+1}(\Omega)}$ depends on ε and usually dominates $h^{k+1/2}$ for $\varepsilon \leq h$. More descriptive local (weighted) error estimates were derived in [16]. Such estimates show that interior or boundary layers do not pollute the numerical solution into subdomains Ω_0, which are sufficiently far away from the layers and the convergence is optimal over such subdomains,

$$|||e|||_{\Omega_0} \leq C|\log h|h^{k+1/2}.$$

In the above estimate the constant C does not depend on ε if the subdomain Ω_0 is $O(h^{1/2}|\log h|)$ away from the interior layers and $O(h|\log h|)$ away from the boundary layers. Although much more precise than global error estimates, the above local error estimate is not sharp for $\varepsilon < h$. Surprisingly we can show that for DG methods the actual and numerical boundary layers almost coincide, i.e., the subdomain Ω_0 needs only to be of $O(\varepsilon|\log\varepsilon|)$ away from the boundary. This seems to be a small improvement, but has important consequences for optimal control problems. We will show later that for $\varepsilon \ll h$ the discretization error in optimal control problems has optimal order of convergence for mesh sizes h almost down to $O(\varepsilon)$. This result should be compared to the corresponding result in [17] for an SUPG discretization of optimal control problems. There, it has been shown that in contrast to a single equation the boundary layers can pollute the optimal control solution everywhere even into subdomain of smoothness and only the first order convergence rates in general are the best possible. This "nice" behavior of the error for DG methods is due to the weak treatment of the boundary conditions, which are natural to DG methods. If in DG methods we impose Dirichlet boundary conditions strongly, then we observe the same deterioration in the order convergence that is known to hold for SUPG (cf. section 6).

An intuitive reason and some analytical justification for the excellent convergence behavior of DG methods has already been provided in [26] in the case of the CIP method with weak treatment of boundary conditions. Roughly speaking, the main idea is that a "good" numerical solution in the case of $\varepsilon \ll h$ does not only have to approximate the exact solution y, but also the solution y_r to a reduced problem

$$\beta \cdot \nabla y_r(x) + r(x)y_r(x) = f(x), \qquad x \in \Omega, \tag{4.2a}$$

$$y_r(x) = d(x), \qquad x \in \Gamma^-, \tag{4.2b}$$

where as before Γ^- denotes the inflow boundary.

Remark 4.1. For smooth domains, the regularity $u \in H^k(\Omega)$ follows from the method of characteristics if $f \in H^k(\Omega)$ and $g \in H^k(\Gamma^-)$ (cf. [13, sect. 3.2]). In the case of a rectangular domain such regularity follows from the differentiability theorem of Rauch [25] under some additional compatibility condition on data at the inflow vertex.

The following result shows that for small ε the error between y and y_r is small on subdomains that are $K\varepsilon$ and $K\sqrt{\varepsilon}$ distances in directions of $\beta = (\beta_1, \beta_2)^T$ and $\beta^\perp = (-\beta_2, \beta_1)^T$, respectively, away from the outflow boundary Γ^+. We define the cross product for two dimensional vectors a and b by $a \times b := a_1 b_2 - a_2 b_1$, which is just a z-component of the cross-product if we think of vectors a and b as three dimensional vectors with z component to be zero.

LEMMA 4.2. *Let $y \in H^1(\Omega)$ solve (4.1) and let the solution y_r of the reduced problem (4.2) satisfy $y_r \in H^2(\Omega)$. Define*

$$\Omega_0 = \{x \in \Omega : \ (x' - x) \cdot \beta \geq K\varepsilon, \ |(x' - x) \times \beta| \geq K\sqrt{\varepsilon} \ \forall x' \in \Gamma^+\}.$$

If K is sufficiently large, then there exists a constant C independent of y and ε such that

$$\varepsilon^{1/2} \|\nabla(y - y_r)\|_{\Omega_0} + \|y - y_r\|_{\Omega_0} \leq C\varepsilon \|\Delta y_r\|_\Omega.$$

Remark 4.3. The conditions on Ω_0 essentially come from the theory of singularly perturbed problems, namely that the typical size of exponential layers is of order ε and a typical size of the parabolic layers is of order $\sqrt{\varepsilon}$. Two examples of Ω_0 for typical advection fields β are given in Figure 4.1.

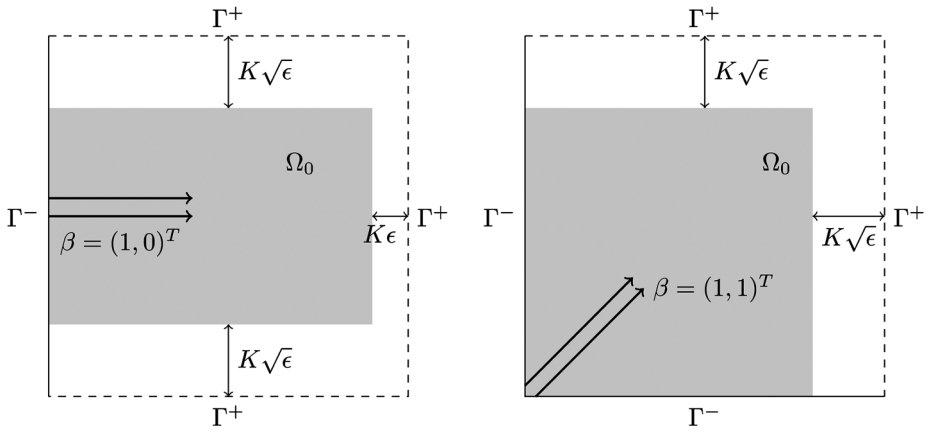

FIG. 4.1. *Sketch of the subdomain Ω_0 for two different advections β.*

Proof. The proof follows the ideas of Lemma 23.2 in [29]. Since we include the case of $r_0 = 0$ the proof is more involved. First, we define two weight functions $\omega \in W^1_\infty(\Omega)$ and $\eta \in C^\infty(\Omega)$ with the following properties:

$$0 \leq \omega(x) \leq 1 \qquad \text{for } x \in \Omega, \tag{4.3a}$$

$$\omega(x) = 1 \qquad \text{for } x \in \Omega_0, \tag{4.3b}$$

$$\omega(x) = 0 \qquad \text{for } x \in \Gamma^+, \tag{4.3c}$$

$$\beta \cdot \nabla\omega(x) \leq 0 \qquad \text{for } x \in \Omega, \tag{4.3d}$$

$$|\beta \cdot \nabla\omega(x)| \leq K^{-1}\varepsilon^{-1}\omega(x) \qquad \text{for } x \in \Omega, \tag{4.3e}$$

$$|\beta^\perp \cdot \nabla\omega(x)| \leq K^{-1}\varepsilon^{-1/2}\omega(x) \qquad \text{for } x \in \Omega. \tag{4.3f}$$

and $\eta = e^{-\gamma\beta\cdot(x-x_0)}$, where $x_0 \in \partial\Omega$ such that $\|\eta\|_{L^\infty(\Omega)} = 1$ and γ is some positive number we specify later. The construction of such a function ω is given in [21, sect. 2] or [16, sect. 2]. If we define $L = diam(\Omega)$, then η has the following properties:

$$e^{-\gamma L|\beta|} \le \eta \le 1, \quad \nabla\eta = -\beta\gamma\eta, \quad \beta\cdot\nabla\eta = -|\beta|^2\gamma\eta. \tag{4.4}$$

We define a bilinear form associated with the advection-diffusion equation by

$$a(y,v) = \varepsilon(\nabla y, \nabla v) + (\beta\cdot\nabla y, v) + (ry, v).$$

Put $e = y - y_r$. Since $\omega^3\eta e \in H_0^1(\Omega)$ ($e = 0$ on Γ^- and $\omega = 0$ on Γ^+),

$$a(e, \omega^3\eta e) = \varepsilon(\Delta y_r, \omega^3\eta e). \tag{4.5}$$

On the other hand

$$a(e,\omega^3\eta e) = \varepsilon(\nabla e, \nabla(\omega^3\eta)e) + \varepsilon(\nabla e, \omega^3\eta\nabla e) + (\beta\cdot\nabla e, \omega^3\eta e) + (re, \omega^3\eta e). \tag{4.6}$$

Applying integration by parts, (4.4), and (4.3d), we find

$$\begin{aligned}(\beta\cdot\nabla e, \omega^3\eta e) &= \frac{1}{2}(\beta\cdot\nabla e, \omega^3\eta e) - \frac{1}{2}(e, \nabla\cdot(\beta\omega^3\eta e))\\ &= -\frac{3}{2}(\beta\cdot(\nabla\omega)e, \omega^2\eta e) + \frac{\gamma|\beta|^2}{2}(e, \omega^3\eta e)\\ &= \frac{3}{2}\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2 + \frac{\gamma|\beta|^2}{2}\|\omega^{3/2}\eta^{1/2}e\|^2.\end{aligned}$$

If we insert the above estimate into (4.6) and use (2.6b), (4.5), we obtain

$$\begin{aligned}&\varepsilon\|\omega^{3/2}\eta^{1/2}\nabla e\|^2 + \frac{3}{2}\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2 + \left(\frac{\gamma|\beta|^2}{2} + r_0\right)\|\omega^{3/2}\eta^{1/2}e\|^2\\ &\qquad = a(e,\omega^3\eta e) - \varepsilon(\nabla e, \nabla(\omega^3\eta)e)\\ &\qquad = \varepsilon(\Delta y_r, \omega^3\eta e) - 3\varepsilon(\omega^2\nabla e, \nabla\omega\eta e) - \varepsilon(\omega^3\nabla e, \nabla\eta e)\\ &\qquad := J_1 + J_2 + J_3.\end{aligned} \tag{4.7}$$

Using the Cauchy–Schwarz inequality, the fact that $\omega, \eta \in [0,1]$, and the arithmetic-geometric mean inequality, we can estimate J_1 as

$$J_1 \le \varepsilon\|\Delta y_r\|\|\omega^{3/2}\eta^{1/2}e\| \le \varepsilon^2\frac{1}{\gamma|\beta|^2}\|\Delta y_r\|^2 + \frac{\gamma|\beta|^2}{4}\|\omega^{3/2}\eta^{1/2}e\|^2. \tag{4.8}$$

To estimate J_2 we recall that $\beta = (\beta_1,\beta_2)^T$, $\beta^\perp = (-\beta_2,\beta_1)^T$, and we use

$$\nabla\omega = \frac{1}{|\beta|^2}\left(\beta\ \beta^\perp\right)\left(\beta\ \beta^\perp\right)^T\nabla\omega = (\omega_{x_1}, \omega_{x_2})^T,$$

where

$$\omega_{x_1} = \frac{1}{|\beta|^2}\left(\beta_1(\beta\cdot\nabla\omega) - \beta_2(\beta^\perp\cdot\nabla\omega)\right), \quad \omega_{x_2} = \frac{1}{|\beta|^2}\left(\beta_2(\beta\cdot\nabla\omega) + \beta_1(\beta^\perp\cdot\nabla\omega)\right).$$

Then,

$$\begin{aligned}J_2 =& -3\varepsilon(\omega^2\nabla e, \nabla\omega\eta e) = -3\varepsilon\Big[(\omega^2 e_{x_1}, \omega_{x_1}\eta e) + (\omega^2 e_{x_2}, \omega_{x_2}\eta e)\Big]\\ =& -\frac{3\varepsilon}{|\beta|^2}\Big[(\omega^2 e_{x_1}, (\beta_1(\beta\cdot\nabla\omega) - \beta_2(\beta^\perp\cdot\nabla\omega))\eta e)\\ &\qquad + (\omega^2 e_{x_2}, (\beta_2(\beta\cdot\nabla\omega) + \beta_1(\beta^\perp\cdot\nabla\omega))\eta e)\Big].\end{aligned}$$

Using (4.3d), the Cauchy–Schwarz and the arithmetic-geometric mean inequalities, and the property (4.3e) of ω, we have

$$\begin{aligned}\frac{3\varepsilon\beta_1}{|\beta|^2}(\omega^2 e_{x_1}, \beta\cdot\nabla\omega\eta e) &= -\frac{3\varepsilon\beta_1}{|\beta|^2}(\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e_{x_1}, \omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e)\\ &\le \varepsilon\frac{3|\beta_1|}{|\beta|^2}\,\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e_{x_1}\|\,\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|\\ &\le \varepsilon^2\frac{18\beta_1^2}{|\beta|^4}\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e_{x_1}\|^2 + \frac{1}{8}\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2\\ &\le \frac{C\varepsilon}{K}\|\omega^{3/2}\eta^{1/2}\nabla e\|^2 + \frac{1}{8}\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2\end{aligned}$$

for some constant C. Similarly, using the Cauchy–Schwarz and the arithmetic-geometric mean inequalities and the property of ω (4.3f), we have

$$\begin{aligned}\frac{3\varepsilon\beta_2}{|\beta|^2}(\omega^2\eta^{1/2}e_{x_1}, \beta^\perp\cdot\nabla\omega\eta^{1/2}e) &\le \frac{C\varepsilon}{K\sqrt{\varepsilon}}\|\omega^{3/2}\eta^{1/2}e_{x_1}\|\|\omega^{3/2}\eta^{1/2}e\|\\ &\le \frac{\varepsilon}{4}\|\omega^{3/2}\eta^{1/2}\nabla e\|^2 + \frac{C}{K^2}\|\omega^{3/2}\eta^{1/2}e\|^2\end{aligned}$$

for some constant C. Similarly we can estimate the other terms in J_2 and arrive at

$$J_2 \le \varepsilon\left(\frac{1}{2}+\frac{2C}{K}\right)\|\omega^{3/2}\eta^{1/2}\nabla e\|^2 + \frac{1}{4}\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2 + \frac{2C}{K^2}\|\omega^{3/2}\eta^{1/2}e\|^2.$$

Finally,

$$\begin{aligned}J_3 = -\varepsilon\gamma(\omega^3\nabla e, \beta\eta e) &\le \varepsilon|\beta|\gamma\|\omega^{3/2}\eta^{1/2}\nabla e\|\|\omega^{3/2}\eta^{1/2}e\|\\ &\le \frac{\varepsilon}{4}\|\omega^{3/2}\eta^{1/2}\nabla e\|^2 + \varepsilon|\beta|^2\gamma^2\|\omega^{3/2}\eta^{1/2}e\|^2.\end{aligned}$$

In summary, we have the bound

$$\begin{aligned}J_1+J_2+J_3 \le{}& \varepsilon^2\frac{1}{\gamma|\beta|^2}\|\Delta y_r\|^2 + \varepsilon\left(\frac{3}{4}+\frac{2C}{K}\right)\|\omega^{3/2}\eta^{1/2}\nabla e\|^2\\ &+\frac{1}{4}\|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2\\ &+\left(\frac{\gamma|\beta|^2}{4}+\frac{2C}{K^2}+\varepsilon|\beta|^2\gamma^2\right)\|\omega^{3/2}\eta^{1/2}e\|^2.\end{aligned}\tag{4.9}$$

Hence if we choose γ and K such that

$$\frac{3}{4}+\frac{2C}{K}<1,\quad \frac{\gamma|\beta|^2}{4}+\frac{2C}{K^2}+\varepsilon|\beta|^2\gamma^2 < \frac{\gamma|\beta|^2}{2}+r_0,\tag{4.10}$$

the estimates (4.7) and (4.9) imply the existence of a constant $c>0$ such that

$$\varepsilon\|\omega^{3/2}\eta^{1/2}\nabla e\|^2 + \|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2 + \|\omega^{3/2}\eta^{1/2}e\|^2 \le c\varepsilon^2\|\Delta y_r\|^2.$$

Using $e^{-\gamma L|\beta|}\le\eta$ (see (4.4)) and the properties (4.3a) and (4.3b) we obtain

$$\begin{aligned}\varepsilon\|\nabla e\|_{\Omega_0}^2 + \|e\|_{\Omega_0}^2 &\le e^{\gamma L|\beta|}\left(\varepsilon\|\omega^{3/2}\eta^{1/2}\nabla e\|^2 + \|\omega|\beta\cdot\nabla\omega|^{1/2}\eta^{1/2}e\|^2 + \|\omega^{3/2}\eta^{1/2}e\|^2\right)\\ &\le ce^{\gamma L|\beta|}\varepsilon^2\|\Delta y_r\|^2,\end{aligned}$$

which implies the desired inequality. ∎

Next we show that the DG approximation y_h well approximates y_r globally on Ω for $\varepsilon \ll h$.

PROPOSITION 4.4. *Let Ω be a bounded open convex subset of $\mathbb{R}^n$. Assume the solution to the reduced problem (4.2) satisfies $y_r \in H^{k+1}(\Omega)$, $k \geq 1$. If y_h is the DG approximation to y obtained by solving (3.4) using polynomials of degree $k \geq 1$, then there exists a constant C independent of ε and h such that*

$$|||y_r - y_h||| \leq C(\varepsilon h^{-3/2} + h^{k+1/2})\|y_r\|_{k+1} \quad \text{for } \varepsilon < h.$$

Proof. In the proof we adapt the technique of [26]. Let $I_h : H^{k+1}(\Omega) \to V_h$ be the standard interpolation operator. Put $v = I_h y_r - y_h$ and $w = v + h\beta \cdot \nabla v$. Note that $v, w \in V_h$. By Lemma 3.3 and Galerkin orthogonality ($a(y - y_h, w) = 0$),

$$C_1 |||v|||^2 \leq a_h(v, w) = a_h(I_h y_r - y_r, w) + a_h(y_r - y, w) = I_1 + I_2. \tag{4.11}$$

From $(A.8)$ and $(A.9)$ in [15] we have the following estimate for I_1:

$$I_1 \leq C(\varepsilon^{1/2} h^k + h^{k+1/2})\|y_r\|_{k+1}|||v||| \leq C h^{k+1/2}\|y_r\|_{k+1}|||v||| \tag{4.12}$$

for $\varepsilon < h$. Assume for now we also have the estimate

$$I_2 \leq C\varepsilon h^{-3/2}\|y_r\|_{k+1}|||v|||. \tag{4.13}$$

(The proof of (4.13) is lengthy and will be given in Lemma 4.6). Combining (4.11), (4.12), and (4.13), we have

$$|||v||| \leq C(\varepsilon h^{-3/2} + h^{k+1/2})\|y_r\|_{k+1}$$

for $\varepsilon < h$. By the trace inequalities and the approximation properties of the interpolant one can show

$$|||I_h y_r - y_r||| \leq C(\varepsilon^{1/2} h^k + h^{k+1/2})\|y_r\|_{k+1} \leq C h^{k+1/2}\|y_r\|_{k+1}$$

for $\varepsilon < h$. In fact, by the approximation properties of the interpolant [6, sect. 4.4] we get

$$\varepsilon \sum_{\tau \in T_h} \|I_h y_r - y_r\|_\tau^2 \leq C\varepsilon h^{2k+2} \sum_{\tau \in T_h} \|y_r\|_{k+1,\tau}^2 \leq C\varepsilon h^{2k+2}\|y_r\|_{k+1}^2.$$

Moreover, by the trace inequality (3.8a) and the approximation properties of the interpolant [6, sect. 4.4],

$$\begin{aligned}\sum_{e \in \mathcal{E}_h} \frac{\varepsilon}{h_e}\|[\![I_h y_r - y_r]\!]\|_e^2 &\leq C\varepsilon h^{-1} \sum_{\tau \in T_h} \left(h^{-1}\|I_h y_r - y_r\|_\tau^2 + h\|I_h y_r - y_r\|_{1,\tau}^2\right) \\ &\leq C\varepsilon h^{-1} \sum_{\tau \in T_h} h^{2k+1}\|y_r\|_{k+1,\tau}^2 \leq \varepsilon h^{2k}\|y_r\|_{k+1}^2.\end{aligned}$$

The estimates of other terms are similar. Thus, by the triangle inequality we have

$$|||y_r - y_h||| \leq |||I_h y_r - y_r||| + |||I_h y_r - y_h||| \leq C(\varepsilon h^{-3/2} + h^{k+1/2})\|y_r\|_{k+1}. \tag{4.14}$$ ∎

The above result shows that for $\varepsilon \ll h$, the DG solution y_h approximates the solution y_r of the reduced problem with optimal order on the whole domain Ω. Combining this result with Lemma 4.2 we immediately obtain the following result.

THEOREM 4.5. *Let Ω be a bounded open convex subset of $\mathbb{R}^n$. Assume that the solution to the reduced problem* (4.2) *satisfies $y_r \in H^{k+1}(\Omega)$, $k \geq 1$. Furthermore, let y_h be the DG approximation to y obtained by solving* (3.4) *using polynomials of degree $k \geq 1$. If the subdomain Ω_0 is given as in Lemma* 4.2*, then there exists a constant C independent of ε and h such that*

$$\varepsilon^{1/2}\|\nabla(y-y_h)\|_{\Omega_0} + \|y-y_h\|_{\Omega_0} \leq C(\varepsilon h^{-3/2} + h^{k+1/2})\|y_r\|_{k+1} \quad \text{for } \varepsilon < h.$$

Proof. For any subdomain $\Omega_0 \subset \Omega$, Proposition 4.4 implies that

$$|||y_r - y_h|||_{\Omega_0} \leq C(\varepsilon h^{-3/2} + h^{k+1/2})\|y_r\|_{k+1} \quad \text{for } \varepsilon < h.$$

Furthermore, since $k \geq 1$, we have $\|\Delta y_r\| \leq Ch^{-3/2}\|y_r\|_{k+1}$ for all $h \leq \operatorname{diam}(\Omega)$. Using these estimates, the triangle inequality, and Lemma 4.2, we immediately conclude

$$\begin{aligned}
&\varepsilon^{1/2}\|\nabla(y-y_h)\|_{\Omega_0} + \|y-y_h\|_{\Omega_0} \\
&\leq \varepsilon^{1/2}\|\nabla(y_r-y_h)\|_{\Omega_0} + \|y_r-y_h\|_{\Omega_0} + \varepsilon^{1/2}\|\nabla(y_r-y)\|_{\Omega_0} + \|y_r-y\|_{\Omega_0} \\
&\leq C(\varepsilon h^{-3/2} + h^{k+1/2})\|y_r\|_{k+1} + C\varepsilon\|\Delta y_r\| \leq C(\varepsilon h^{-3/2} + h^{k+1/2})\|y_r\|_{k+1}
\end{aligned}$$

for $\varepsilon < h$. ∎

To complete the proof of Proposition 4.4, we still have to show (4.13), which we state as a separate lemma.

LEMMA 4.6 (estimate (4.13)). *If the assumptions of Proposition* 4.4 *are valid, then there exists a constant C such that* (4.13) *holds.*

Proof. Recall that $y \in H^2(\Omega)$ and $y_r \in H^{k+1}(\Omega)$, $k \geq 1$. Using the definition (4.2) of y_r, $y_r \in C^0(\Omega)$, and integration by parts we have

$$\begin{aligned}
a_h(y_r - y, w) =& \varepsilon \sum_{\tau \in T_h} (\nabla y_r, \nabla w)_\tau \\
&+ \varepsilon \sum_{e \in \mathcal{E}_h} \left(\frac{\sigma}{h_e}([\![y_r]\!], [\![w]\!])_e - (\{\nabla y_r\}, [\![w]\!])_e - ([\![y_r]\!], \{\nabla w\})_e \right) \\
=& \varepsilon \sum_{\tau \in T_h} (-\Delta y_r, w)_\tau + \varepsilon \sum_{e \in \mathcal{E}_h} \left(\frac{\sigma}{h_e}([\![y_r]\!], [\![w]\!])_e - ([\![y_r]\!], \{\nabla w\})_e \right). \qquad (4.15)
\end{aligned}$$

If we recall the definition of $w = v + h\beta \cdot \nabla v$ and apply the local inverse inequality (3.8b), we find that

$$\sum_{\tau \in T_h} \|w\|_\tau^2 \leq C \sum_{\tau \in T_h} \|v\|_\tau^2 = C\|v\|_\Omega^2. \qquad (4.16)$$

Using the inequality

$$\sum_i |a_i b_i| \leq \left(\sum_i |a_i|^2 \right)^{1/2} \left(\sum_i |b_i|^2 \right)^{1/2} \quad \text{for } a_i, b_i \in \mathbb{R}, \qquad (4.17)$$

and (4.16) the first term on the right-hand side of (4.15), we can estimate

$$\varepsilon \sum_{\tau \in T_h} (-\Delta y_r, w)_\tau \leq \varepsilon \left(\sum_{\tau \in T_h} \|\Delta y_r\|_\tau^2 \right)^{1/2} \left(\sum_{\tau \in T_h} \|w\|_\tau^2 \right)^{1/2} \leq C\varepsilon\|y_r\|_{2,\Omega}\|v\|_\Omega. \qquad (4.18)$$

To estimate the second sum in (4.15) we notice that since $y_r \in H^2$ we have $[\![y_r]\!]_e = 0$ on all interior edges e. Thus we need only estimate

$$\varepsilon \sum_{e\in\mathcal{E}_h^\partial} \left(\frac{\sigma}{h_e}([\![y_r]\!],[\![w]\!])_e - ([\![y_r]\!],\{\nabla w\})_e \right).$$

Using (4.17) and the local inverse inequality (3.8c), we have

$$\begin{aligned}
\varepsilon \sum_{e\in\mathcal{E}_h^\partial} \left(\frac{\sigma}{h_e}([\![y_r]\!],[\![w]\!])_e \right) &\le C\varepsilon h^{-1} \left(\sum_{e\in\mathcal{E}_h^\partial} \|[\![y_r]\!]\|_e^2 \right)^{1/2} \left(\sum_{e\in\mathcal{E}_h^\partial} \|[\![w]\!]\|_e^2 \right)^{1/2} \\
&\le \frac{C\varepsilon}{h} \left(\sum_{e\in\mathcal{E}_h^\partial} \|[\![y_r]\!]\|_e^2 \right)^{1/2} \left(\sum_{\tau\in T_h} h^{-1}\|v\|_\tau^2 + h\|\beta\cdot\nabla v\|_\tau^2 \right)^{1/2} \qquad (4.19) \\
&\le C\varepsilon h^{-3/2} \left(\sum_{e\in\mathcal{E}_h^\partial} \|[\![y_r]\!]\|_e^2 \right)^{1/2} |||v|||.
\end{aligned}$$

Using the continuous embedding $H^2(\Omega) \hookrightarrow C(\overline{\Omega})$ and that the number of boundary edges is of order h^{-1}, we have

$$\sum_{e\in\mathcal{E}_h^\partial} \|[\![y_r]\!]\|_e^2 \le Ch\|y_r\|_{L^\infty(\Omega)}^2 \sum_{e\in\mathcal{E}_h^\partial} 1 \le C\|y_r\|_{H^2(\Omega)}^2. \qquad (4.20)$$

Similarly, using the trace inequality (3.8c) applied to ∇w followed by the inverse inequality (3.8b) applied to ∇w and (4.20), we find

$$\begin{aligned}
\varepsilon \sum_{e\in\mathcal{E}_h^\partial} ([\![y_r]\!],\{\nabla w\})_e &\le \varepsilon \left(\sum_{e\in\mathcal{E}_h^\partial} \|[\![y_r]\!]\|_e^2 \right)^{1/2} \left(\sum_{e\in\mathcal{E}_h^\partial} \|\{\nabla_h w\}\|_e^2 \right)^{1/2} \\
&\le C\varepsilon\|y_r\|_{H^2(\Omega)} \left(\sum_{\tau\in T_h} h^{-3}\|w\|_\tau^2 \right)^{1/2} \qquad (4.21) \\
&\le C\varepsilon\|y_r\|_{H^2(\Omega)} \left(\sum_{\tau\in T_h} h^{-3}\|v\|_\tau^2 + h^{-1}\|\beta\cdot\nabla v\|_\tau^2 \right)^{1/2} \\
&\le C\varepsilon h^{-3/2}\|y_r\|_{H^2(\Omega)}|||v|||.
\end{aligned}$$

Equations (4.19) and (4.21) imply

$$\begin{aligned}
\varepsilon \sum_{e\in\mathcal{E}_h} \left(\frac{\sigma}{h_e}([\![y_r]\!],[\![w]\!])_e - ([\![y_r]\!],\{\nabla w\})_e \right) &= \varepsilon \sum_{e\in\mathcal{E}_h^\partial} \left(\frac{\sigma}{h_e}([\![y_r]\!],[\![w]\!])_e - ([\![y_r]\!],\{\nabla_h w\})_e \right) \\
&\le C\varepsilon h^{-3/2}\|y_r\|_{H^2(\Omega)}|||v|||.
\end{aligned}$$

This inequality combined with (4.15) and (4.18) implies (4.13). ∎

5. Local error estimates for the optimal control problem. We now turn to optimal control problems. We assume

- $\varepsilon \leq h$, i.e., we consider only the advection-dominating case, and
- $Y_h = U_h = V_h$.

Consider the optimality systems (2.7) and (3.7). From (2.7b) we can conclude that $\lambda = \alpha u$, which leads to the reduced optimality system

$$\alpha a(\varphi, u) + (y, \varphi) = \langle \hat{y}, \varphi \rangle \qquad \forall \varphi \in V, \tag{5.1a}$$

$$a(y, \phi) - (u, \phi) = \langle f, v \rangle \qquad \forall \phi \in V. \tag{5.1b}$$

Similarly, from (3.7b) we obtain $\lambda_h = \alpha u_h$, which leads to the reduced discrete optimality system

$$\alpha a_h(\varphi_h, u_h) = -\langle y_h - \hat{y}, \varphi_h \rangle \qquad \forall \varphi_h \in V_h, \tag{5.2a}$$

$$a_h(y_h, \phi_h) - (u_h, \phi_h) = l_h(\phi_h) \qquad \forall \phi_h \in V_h. \tag{5.2b}$$

The system (5.2) motivates the definition of the reduced bilinear form $\mathcal{A}^{DG}(\cdot, \cdot)$ on $(V_h \times V_h) \times (V_h \times V_h)$ given by

$$\mathcal{A}^{DG}((y_h, u_h), (\phi_h, \varphi_h)) = a_h(y_h, \phi_h) - (u_h, \phi_h) + \alpha a_h(\varphi_h, u_h) + (y_h, \varphi_h). \tag{5.3}$$

The reduced discrete optimality system (5.2) can be written as

$$\mathcal{A}^{DG}((y_h, u_h), (\phi_h, \varphi_h)) = l_h(\phi_h) + \langle \hat{y}, \varphi_h \rangle \quad \forall (\phi_h, \varphi_h) \in V_h \times V_h. \tag{5.4}$$

Notice that the discontinuous Galerkin method is consistent, i.e., provided that the exact solution is regular enough (e.g., y and u, λ in H^2), then

$$\mathcal{A}^{DG}((y, u), (\phi_h, \varphi_h)) = l_h(\phi_h) + \langle \hat{y}, \varphi_h \rangle \quad \forall (\phi_h, \varphi_h) \in V_h \times V_h. \tag{5.5}$$

In particular, (5.4) and (5.5) imply the Galerkin orthogonality condition

$$\mathcal{A}^{DG}\big((y - y_h, u - u_h), (\phi_h, \varphi_h)\big) = 0 \quad \forall (\phi_h, \varphi_h) \in V_h \times V_h. \tag{5.6}$$

5.1. The case of interior layers. First we state an estimate for the local error between the solution of the infinite dimensional optimal control problem (2.5) and the solution of the discretized problem (3.6) in the presence of interior layers. We will show that the interior layers do not pollute the numerical solution to the coupled optimality system (3.7) obtained using SIPG. For the SIPG discretization of a single equation, such a result is shown in [16], and for numerical solution of optimal control problems using the SUPG method the same behavior was proven in [17, sect. 3].

The results in, e.g., [28, p. 473] or [29, L. 23.1] describe what parts of the forcing term f influence the exact solution of a single advection dominated PDE at any fixed point $x_0 \in \Omega$: The force term in the entire upstream direction of x_0 influences the exact solution at x_0, but only the force term from within an $\varepsilon|\log\varepsilon|$-neighborhood in the streamline (downwind) direction and within a $\sqrt{\varepsilon}|\log(\varepsilon)|$-neighborhood in the crosswind direction influence exact solution at x_0. The same behavior can be observed from the properties of the corresponding Green's function. In the presence of interior layers only, the exact solution may vary strongly in the crosswind direction, but not in the streamline direction. Since the adjoint equation has similar properties, the same behavior of the solution can be expected from the coupled optimality system.

Recall that for $a, b \in \mathbb{R}^2$ the cross product is defined by $a \times b := a_1 b_2 - a_2 b_1$. Let $A_1 < A_2$, let $K > 0$ be a sufficiently large constant, and let $s > 0$. We define the strips

$$\Omega_0 = \{x \in \Omega : A_1 \le (x \times \beta) \le A_2\},$$
$$\Omega_s^+ = \{x \in \Omega : A_1 - sK\sqrt{h}|\log h| \le (x \times \beta) \le A_2 + sK\sqrt{h}|\log h|\}$$

along β of width $|A_2 - A_1|$ and $|A_2 - A_1 + 2sK\sqrt{h}|\log h|$, respectively.

THEOREM 5.1. *Let Ω be a bounded open convex subset of $\mathbb{R}^n$ and let (y, u) and (y_h, u_h) satisfy (5.6). If $h \le C_2\alpha$ for some constant C_2 and $\varepsilon \le h$, then there exists a constant C independent of y, u, ε, and h such that for any $s > 0$ and mesh sizes $\varepsilon \le h$,*

$$\begin{aligned}\|y - y_h\|_{\Omega_0} + \alpha\|u - u_h\|_{\Omega_0} \le & C\left(h^{3/2}\|y\|_{2,\Omega_s^+} + h^{s+3/2}\|y\|_{2,\Omega}\right) \\ & + C\alpha\left(h^{3/2}\|u\|_{2,\Omega_s^+} + h^{s+3/2}\|u\|_{2,\Omega}\right).\end{aligned}$$

The proof of Theorem 5.1 uses weighted error estimates, where the purpose of the weighting function is to isolate the domains of smoothness from the layers. The main ideas of the proof are already contained in [17, sect. 3], where the same result is proven for the SUPG discretization of the optimal control problem (1.1). Since the proof is rather long, we omit it here.[1]

The interpretation of Theorem 5.1 is essentially the same as that given in [17, p. 4615] and we adapt it here for completeness. The right-hand side in the error estimate of Theorem 5.1 depends on local and global norms of the state and the adjoint. The local norms associated with $h^{3/2}$ are independent of ε if Ω_s^+ does not contain interior layers. The global norms depend on $\|y\|_{2,\Omega}$ and $\|u\|_{2,\Omega}$ and because of the regularity result and in view of Theorem 2.2 may depend on negative powers of ε. However, they are associated with the higher order terms $h^{s+3/2}$. Thus negative powers of ε can be compensated by h^s for sufficiently large s, provided that for these values of s the subdomain Ω_s^+ does not contain interior layers.

5.2. The case of boundary layers. In this section we extend the main result of section 4 to optimal control problems. Because the optimality system consists of coupled advection-diffusion-reaction equations the analysis is more involved.

The reduced optimality system corresponding to (1.1b), (1.1c), (1.2), (1.3) is given by

$$\beta \cdot \nabla y_r(x) + r(x)y_r(x) = f(x) + u_r(x), \quad x \in \Omega, \tag{5.7a}$$
$$y_r(x) = d(x), \qquad x \in \{x \in \partial\Omega \,:\, \beta \cdot \mathbf{n}(x) < 0\}, \tag{5.7b}$$
$$-\alpha\beta \cdot \nabla u_r(x) + \alpha r(x)u_r(x) = -(y_r(x) - \widehat{y}(x)), \quad x \in \Omega, \tag{5.7c}$$
$$u_r(x) = 0, \qquad x \in \{x \in \partial\Omega \,:\, \beta \cdot \mathbf{n}(x) > 0\}. \tag{5.7d}$$

THEOREM 5.2. *Assume that the solution y_r, u_r to reduced problem (5.7) satisfies $y_r, u_r \in H^2(\Omega)$ and that there exists Ω_0 such that*

$$\begin{aligned}&\varepsilon^{1/2}\|\nabla(y - y_r)\|_{\Omega_0} + \|y - y_r\|_{\Omega_0} + \varepsilon^{1/2}\|\nabla(u - u_r)\|_{\Omega_0} + \|u - u_r\|_{\Omega_0} \\ &\le C\varepsilon\left(\|\Delta y_r\|_\Omega + \|\Delta u_r\|_\Omega\right).\end{aligned} \tag{5.8}$$

[1] An "appendix" with the proof of Theorem 5.1 is available at http://www.caam.rice.edu/∼heinken/papers/DLeykekhman_MHeinkenschloss_2010a.html.

Let y_h and u_h be the SIPG approximation to y and u using polynomials of degree $k \geq 1$ and satisfy (5.6). *Assume $\varepsilon \leq h$. Then for h sufficiently small there exists a constant C independent of ε, h, y, and u such that*

$$|||y - y_h|||_{\Omega_0} + |||u - u_h|||_{\Omega_0} \leq C(\varepsilon h^{-3/2} + h^{k+1/2}) \left(\|y_r\|_{k+1} + \|u_r\|_{k+1}\right).$$

Remark 5.3. From Lemma 4.2 for a single equation, we expect (5.8) to hold for

$$\Omega_0 = \{x \in \Omega : \ |(x' - x) \cdot \beta| \geq K\varepsilon, \ |(x' - x) \times \beta| \geq K\sqrt{\varepsilon} \ \forall x' \in \partial\Omega\}.$$

Before we provide the proof of the theorem, let us first collect some result we will use in the proof.

5.2.1. Preliminary results. The first result we will need is a simplified version of Lemma 4.1 in [2]. To state the result we need the function $\eta = e^{-\beta\cdot\gamma(x-x_0)}$ from section 4 with the properties (4.4). We will also need an exponential function $\eta^* = e^{\beta\cdot\gamma(x-x_1)}$, where $x_1 \in \partial\Omega$ such that $\|\eta^*\|_{L^\infty(\Omega)} = 1$. Then η^* has the following properties:

$$e^{-\gamma L|\beta|} \leq \eta^* \leq 1, \quad \nabla\eta^* = \beta\gamma\eta^*, \quad \beta \cdot \nabla\eta^* = |\beta|^2\gamma\eta^*. \tag{5.9}$$

LEMMA 5.4. *There exists a constant C_2 such that for all $y \in Y_h$,*

$$C_2(\|y\|_{DG}^2 + \gamma\|y\|^2) \leq a(y, \eta y).$$

Proof. The proof of this result is straightforward. ∎

For $v \in H^{k+1}(\Omega)$, we let $\widetilde{v}$ denote the local L^2-projection of v onto V_h defined by

$$(v - \widetilde{v}, \chi)_\tau = 0 \quad \forall \chi \in \mathbb{P}^k(\tau) \quad \tau \in T_h.$$

Recall the standard estimates

$$\|v - \widetilde{v}\|_{s,\tau} \leq Ch_\tau^{k+1-s}|v|_{k+1,\tau}, \quad s = 0, 1, \tag{5.10a}$$

$$\|v - \widetilde{v}\|_{\partial\tau} \leq Ch_\tau^{k+1/2}|v|_{k+1,\tau}. \tag{5.10b}$$

The following superapproximation result shows that functions of special form can be approximated very well by an L^2-projection.

LEMMA 5.5 (superapproximation). *Let η be from above. Then for any $v \in V_h$ there exists a constant C independent of h such that for $h \leq \gamma$,*

$$\|\eta v - \widetilde{\eta v}\|_\tau + h^{1/2}\|\eta v - \widetilde{\eta v}\|_{\partial\tau} + h|\eta v - \widetilde{\eta v}|_{1,\tau} \leq Ch\gamma\|v\|_\tau.$$

Proof. The proof is standard. One needs to use the approximation properties of the L^2-projection (5.10), $|v|_{H^{k+1}(\tau)} = 0$, $\|\eta\|_{W_\infty^l} \leq C\gamma^l$, and the inverse inequality $\|v\|_{H^l(\tau)} \leq Ch^{-l}\|v\|_\tau$. ∎

LEMMA 5.6. *For any $v \in V_h$ and any constant δ there exists a constant C_δ independent of h and ε such that*

$$a_h(v, \eta v - \widetilde{\eta v}) \leq \delta|||v|||^2 + C_\delta\gamma^2(\varepsilon + h)\|v\|^2.$$

Proof. The proof uses the superapproximation result, Lemma 5.5, and the Cauchy–Schwarz and the arithmetic-geometric mean inequalities. We give some illustration. By the Cauchy–Schwarz inequality, superapproximation result Lemma 5.5, and the arithmetic-geometric mean inequality,

$$\begin{aligned}\varepsilon(\nabla v, \nabla(\eta v - \widetilde{\eta v}))_\tau &\leq \varepsilon\|\nabla v\|_\tau\|\nabla(\eta v - \widetilde{\eta v})\|_\tau \\ &\leq \varepsilon\|\nabla v\|_\tau C\gamma\|v\|_\tau \leq \delta\varepsilon\|\nabla v\|_\tau^2 + C_\delta\varepsilon\gamma^2\|v\|_\tau^2.\end{aligned}$$

Similarly, employing in addition the inverse inequality,

$$\begin{aligned}\varepsilon \sum_{e\in\mathcal{E}_h} (\{\nabla_h v\}, [\![\eta v - \widetilde{\eta v}]\!])_e &\le \varepsilon \sum_{\tau\in T_h} h^{-1/2}\|\nabla v\|_\tau C h^{1/2}\gamma\|v\|_\tau \\ &\le \sum_{\tau\in T_h} \delta\varepsilon\|\nabla v\|_\tau^2 + C_\delta \varepsilon\gamma^2\|v\|_\tau^2\end{aligned}$$

and

$$\varepsilon \sum_{e\in\mathcal{E}_h} \frac{\sigma}{h}([\![v]\!], [\![\eta v - \widetilde{\eta v}]\!])_e \le \varepsilon\delta \sum_{e\in\mathcal{E}_h} \frac{1}{h}\|[\![v]\!]\|_e^2 + C_\delta\varepsilon\gamma^2 \sum_{\tau\in T_h}\|v\|_\tau^2.$$

Finally we can estimate the advection terms by

$$(\beta\cdot\nabla v, \eta v - \widetilde{\eta v})_\tau \le \|\beta\cdot\nabla v\|_\tau C h\gamma\|v\|_\tau \le \delta h\|\beta\cdot\nabla v\|_\tau + C_\delta h\gamma^2\|v\|_\tau^2$$

and

$$\begin{aligned}\sum_{e\in\mathcal{E}_h} ((v^+ - v^-)|\beta\cdot\mathbf{n}|, (\eta v - \widetilde{\eta v})^+)_e &\le \sum_{e\in\mathcal{E}_h} \|(v^+ - v^-)|\beta\cdot\mathbf{n}|\|_e C h^{1/2}\gamma\|v\|_{S_e} \\ &\le \delta \sum_{e\in\mathcal{E}_h} \|(v^+ - v^-)|\beta\cdot\mathbf{n}|^{1/2}\|_e^2 + C_\delta\varepsilon\gamma^2 \sum_{\tau\in T_h}\|v\|_\tau^2. \qquad \square\end{aligned}$$

5.2.2. Proof of Theorem 5.2. Put $v = I_h y_r - y_h$ and $w = I_h u_r - u_h$, then by Lemma 5.6,

$$\begin{aligned}C_1(|||v|||^2 + \gamma\|v\|^2 + |||w|||^2 + \gamma\|w\|^2) &- (w, Kv + h\beta\cdot\nabla v + \eta v) \\ &+ (v, Kw - h\beta\cdot\nabla w + \eta^* w) \\ &\le \mathcal{A}^{DG}((v,w),(Kv + h\beta\cdot\nabla v + \eta v, Kw - h\beta\cdot\nabla w + \eta^* w)).\end{aligned}$$

By the Cauchy–Schwarz, the arithmetic-geometric, and the inverse inequalities, we have

$$-(w, Kv + h\beta\cdot\nabla v + \eta v) + (v, Kw - h\beta\cdot\nabla w + \eta^* w) \le (K + C_{inv} + 1)(\|v\|^2 + \|w\|^2).$$

Adding and subtracting $\widetilde{\eta y}$ and $\widetilde{\eta^* u}$ we have

$$\begin{aligned}&\mathcal{A}^{DG}((v,w),(Kv + h\beta\cdot\nabla v + \eta v, Kw - h\beta\cdot\nabla w + \eta^* w)) \\ &= \mathcal{A}^{DG}((v,w),(Kv + h\beta\cdot\nabla v + \widetilde{\eta v}, Kw - h\beta\cdot\nabla w + \widetilde{\eta^* w})) \\ &\quad + \mathcal{A}^{DG}((v,w),(\eta v - \widetilde{\eta v}, \eta^* w - \widetilde{\eta^* w})) \\ &\stackrel{\text{def}}{=} I_1 + I_2.\end{aligned}$$

To treat I_1 we add and subtract y_r and u_r and use the Galerkin orthogonality. Notice that $Kv + h\beta\cdot\nabla v + \widetilde{\eta v} \in V_h$. Thus,

$$\begin{aligned}I_1 &= \mathcal{A}^{DG}((I_h y_r - y_r, I_h u_r - u_r),(Kv + h\beta\cdot\nabla v + \widetilde{\eta v}, Kw - h\beta\cdot\nabla w + \widetilde{\eta^* w})) \\ &\quad + \mathcal{A}^{DG}((y_r - y, u_r - u),(Kv + h\beta\cdot\nabla v + \widetilde{\eta v}, Kw - h\beta\cdot\nabla w + \widetilde{\eta^* w})) \\ &\stackrel{\text{def}}{=} J_1 + J_2.\end{aligned}$$

Similarly to Theorem 4.5, from $(A.8)$ and $(A.9)$ in [15] we have the following estimate for J_1:

$$J_1 \le C(\varepsilon^{1/2}h^k + h^{k+1/2})(\|y_r\|_{k+1}|||v||| + \|u_r\|_{k+1}|||w|||). \tag{5.11}$$

Along the lines of Lemma 4.6 we can obtain an estimate for J_2,

$$J_2 \le C(\varepsilon h^{-3/2} + h^{k+1/2})(\|y_r\|_{k+1}|||v||| + \|u_r\|_{k+1}|||w|||).$$

Next we will treat I_2. To obtain the desirable estimate we use Lemma 5.6 with $\delta = C_1/2$ and observe that the coupling terms do not pose problems and, for example, can be estimated as

$$(w, \eta v - \widetilde{\eta v})_\tau \le \|w\|_\tau Ch\gamma\|v\|_\tau \le Ch\gamma(\|v\|_\tau^2 + \|w\|_\tau^2).$$

Thus, we have

$$I_2 \le \frac{C_1}{2}(|||v|||^2 + |||w|||^2) + Ch\gamma(\|v\|^2 + \|w\|^2).$$

Thus provided $\varepsilon \le h$ and h is small enough, by combining estimates for I_1, I_2, J_1, and J_2 and choosing $\gamma \ge K + C_{inv} + 1 + Ch\gamma^2$, we have

$$|||v|||^2 + |||w|||^2 \le C(\varepsilon h^{-3/2} + h^{k+1/2})(\|y_r\|_{k+1} + \|u_r\|_{k+1}).$$

The above inequality implies that for any subdomain $\Omega_0 \subset \Omega$,

$$|||v|||_{\Omega_0}^2 + |||w|||_{\Omega_0}^2 \le C(\varepsilon h^{-3/2} + h^{k+1/2})(\|y_r\|_{k+1} + \|u_r\|_{k+1}).$$

Let Ω_0 be as in the statement of the theorem, then by the triangle inequality we finally can conclude

$$|||y - y_h|||_{\Omega_0}^2 + |||u - u_h|||_{\Omega_0}^2 \le C(\varepsilon h^{-3/2} + h^{k+1/2})(\|y_r\|_{k+1} + \|u_r\|_{k+1}). \qquad \square$$

6. Numerical results. We illustrate our theoretical findings of the previous sections with a few simple examples.

6.1. Example 1. To support the theoretical result of section 4, we consider

$$-\varepsilon y''(x) + y'(x) = f(x) \quad \text{on } (0,1), \quad y(0) = y(1) = 0 \tag{6.1a}$$

with $\varepsilon = 10^{-9}$. The right-hand side function f is such that the exact solution is

$$y(x) = x^4 - \frac{e^{\frac{x-1}{\varepsilon}} - e^{-\frac{1}{\varepsilon}}}{1 - e^{-\frac{1}{\varepsilon}}}. \tag{6.1b}$$

The solution has a boundary layer at $x = 1$ of width $O(\varepsilon|\log\varepsilon|)$. We compute the L^2 and H^1 norm errors between the computed solution and the exact solution over the subdomain $\Omega_0 = (0, 1 - 6\varepsilon|\log\varepsilon|)$.

The left plot in Figure 6.1 shows the exact solution (6.1b) and the solution computed using the SIPG method with piecewise quadratic elements on a uniform mesh with mesh size $h = 1/10$. Without any special mesh design, the SIPG method fails to resolve the boundary layer for meshes with $h > \varepsilon$. However, in contrast to other stabilized methods, where boundary conditions are imposed strongly, such as in the SUPG method, the numerical layer in the SIPG method is only of order $O(\varepsilon|\log\varepsilon|)$ and not $O(h|\log h|)$, as one would expect. The right plot in Figure 6.1 shows the L^2- and H^1-errors between the exact and computed solution on the subdomain Ω_0, where the computed solution is obtained using the SIPG method with piecewise linear and piecewise quadratic elements. The numerical results confirm our

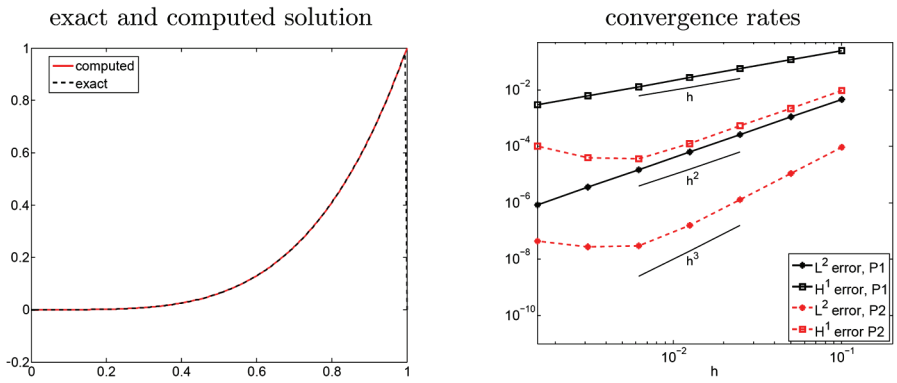

FIG. 6.1. *Results for Example* 1. *The left plot shows the exact solution* (6.1b) *and the computed solution using the SIPG method with piecewise quadratic* ($P2$) *elements on a uniform mesh with mesh size* $h = 1/10$. *The right plot shows the* L^2- *and* H^1-*errors between the exact and computed solutions of the PDE* (6.1a) *on the subdomain* $\Omega_0 = (0, 1 - 6\varepsilon|\log\varepsilon|)$, *when the computed solution is obtained using the SIPG method with piecewise linear* ($P1$) *and piecewise quadratic* ($P2$) *elements.*

theoretical findings. For example, for piecewise quadratic elements, $k = 2$, the estimate in Theorem 4.5 reads $\varepsilon^{1/2}\|\nabla(y - y_h)\|_{\Omega_0} + \|y - y_h\|_{\Omega_0} \le C(\varepsilon h^{-3/2} + h^{5/2})$ for $\varepsilon < h$. In particular for $h \ge \varepsilon^{1/4} = 10^{-2.25}$, we have $\varepsilon h^{-3/2} + h^{5/2} \le 2h^{5/2}$. Hence, $\|y - y_h\|_{\Omega_0} \le C(\varepsilon h^{-3/2} + h^{5/2}) \le Ch^{5/2}$. The right plot in Figure 6.1 shows even cubic convergence for $h \ge \varepsilon^{1/4} = 10^{-2.25}$.

6.2. Example 2. We apply the SIPG method to the optimal control problem (1.1) on $\Omega = (0, 1)$. The right-hand side f and the desired solution $\widehat{y}$ are selected such that the optimal state y, control u, and adjoint λ are given by

$$y(x) = x^4 - \frac{e^{\frac{x-1}{\varepsilon}} - e^{-\frac{1}{\varepsilon}}}{1 - e^{-\frac{1}{\varepsilon}}}, \quad \alpha u(x) = \lambda(x) = (1 - x)^4 - \frac{e^{-\frac{x}{\varepsilon}} - e^{-\frac{1}{\varepsilon}}}{1 - e^{-\frac{1}{\varepsilon}}}. \tag{6.2}$$

We set the diffusion and regularization parameters to $\varepsilon = 10^{-9}$ and $\alpha = 10^{-1}$. Note that the solution is constructed such that the optimal state y has a boundary layer at $x = 1$, and the optimal control u has a boundary layer at $x = 0$ (cf. Figure 6.2).

The convergence behavior of the SIPG method for the optimal control problem is a direct consequence of the behavior of the SIPG method for a single equation, as illustrated in Example 1. Since the numerical boundary layer is of order $O(\varepsilon|\log\varepsilon|)$ we expect the convergence to be optimal all the way down to order ε and then deteriorate because of the pollution effect. Figure 6.3 confirms this prediction.

On the other hand, if we impose boundary conditions strongly, we can see (cf. Figure 6.4) that since the boundary layers are of order $O(h|\log h|)$ they now pollute the numerical solution everywhere. As a consequence, the convergence rates are reduced to only first order in both $L^2(\Omega_0)$ or $H^1(\Omega_0)$ norms for both piecewise linear and piecewise quadratic elements (cf. Figure 6.5). A similar pollution effect was already observed for the SUPG method in [17].

6.3. Example 3. In the previous example, we selected the optimal state and control and constructed the other problem data from the optimality conditions. Now we specify the right-hand side f and desired state $\widehat{y}$ rather than the solution of the

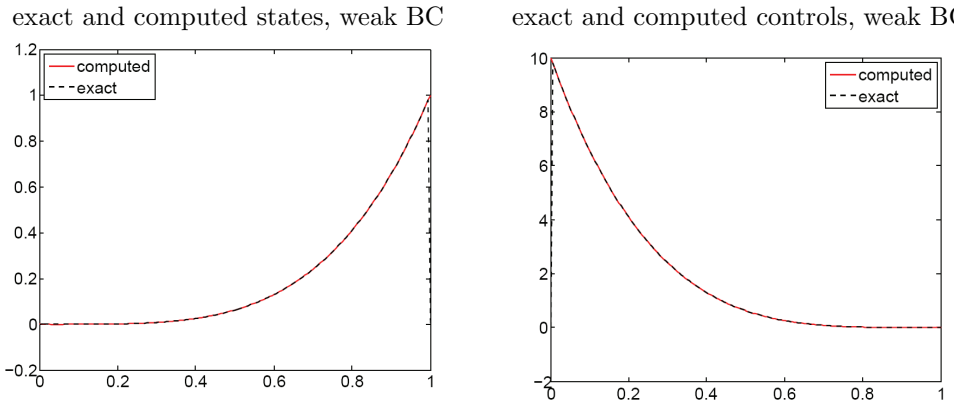

FIG. 6.2. *Exact and computed states (left plot) and controls (right plot) for Example* 2. *The computed state and control are obtained using the SIPG method with piecewise quadratic (P*2*) elements on a uniform mesh with mesh size* $h = 1/10$.

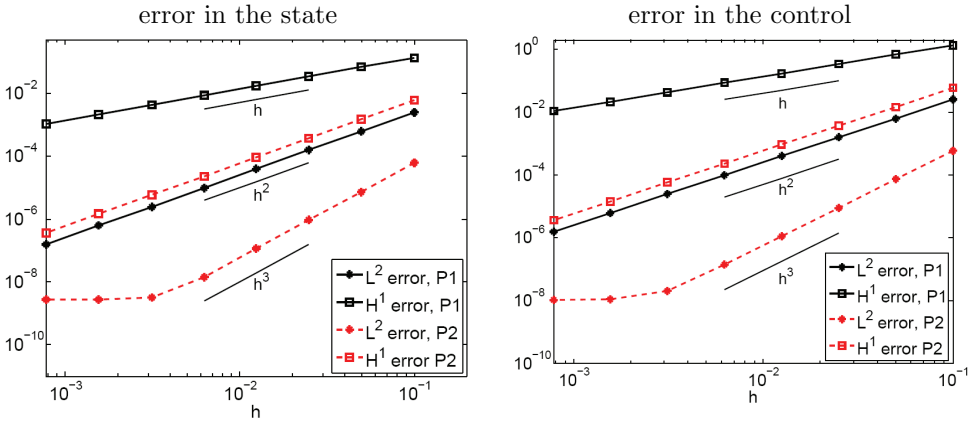

FIG. 6.3. *The left (right) plot shows the* L^2*- and* H^1*-errors between the exact and computed state (control) for Example* 2 *on the subdomain* $\Omega_0 = (0.1, 0.9)$, *when the computed solution is obtained using SIPG with piecewise linear (P*1*) and piecewise quadratic (P*2*) elements.*

optimal control problem. Let $\Omega = (0, 1)$ and

$$f \equiv 1, \quad \widehat{y} \equiv 1, \quad \varepsilon = 10^{-9}, \quad \alpha = 10^{-1}.$$

The optimal state, control, and adjoint for this problem are not known analytically. Instead we compute the solution of the optimal control problem using the SIPG method on a fine grid with mesh size $h = 1/(5 * 2^{10})$. We refer to this solution as the "exact" solution. We compare this "exact" solution with the computed solution on meshes with mesh sizes $h = 1/5$ to $h = 1/(5 * 2^8)$. Figure 6.6 shows the "exact" and the approximate states and controls.

Figure 6.7 shows the L^2- and H^1-errors between the exact and the computed states and controls on the subinterval $\Omega_0 = (0.1, 0.9)$ for various mesh sizes. The errors behave optimally even down to $o(\varepsilon)$ for both the L^2- and H^1-norms and for both piecewise linear and piecewise quadratic elements.

If we impose Dirichlet boundary conditions strongly, then the convergence rate of the SIPG method deteriorate to first order as already observed in the previous

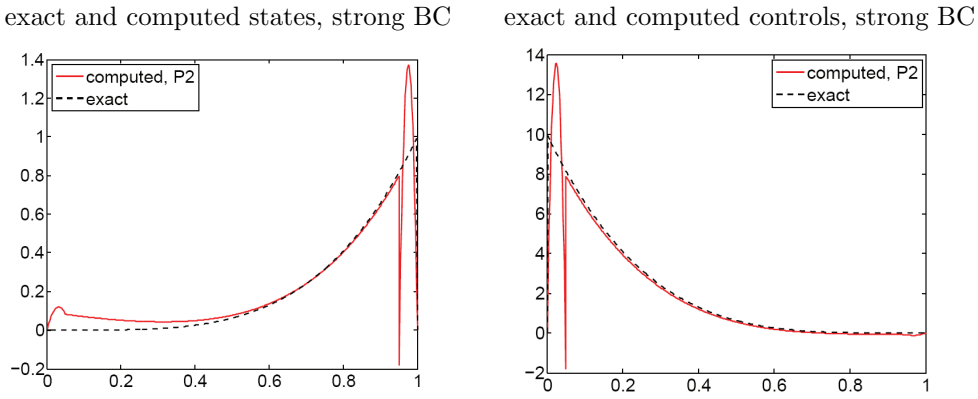

FIG. 6.4. *Exact and computed states (left plot) and controls (right plot) for Example* 2. *The computed state and control are obtained using the SIPG method with strong implementation of boundary conditions with piecewise quadratic (P*2*) elements on a uniform mesh with mesh size* $h = 1/20$.

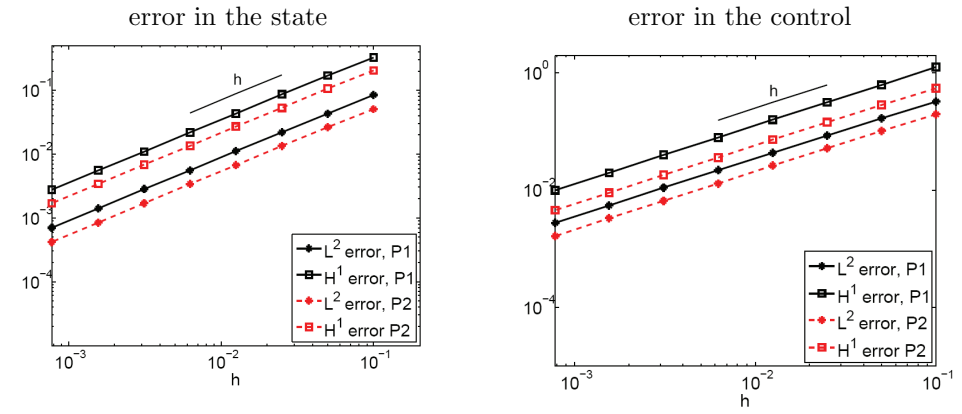

FIG. 6.5. *The left (right) plot shows the* L^2*- and* H^1*-errors between the exact and computed state (control) for Example* 2 *on the subdomain* $\Omega_0 = (0.1, 0.9)$, *when the computed solution is obtained using the SIPG method with strong implementation of boundary conditions with piecewise quadratic (P*2*) elements.*

example.

6.4. Example 4. Theorem 5.1, among other things, shows that interior layers do not pollute the solution. To illustrate this statement numerically we consider the system (5.1) with $\Omega = (0,1)^2$, $\varepsilon = 10^{-5}$, $\alpha = 10^{-1}$, and $\beta = (1,0)^T$. The functions f and $\widehat{y}$ are computed such that the exact solution is

$$y(x_1, x_2) = (1 - x_1)^3 \tan^{-1}\left(\frac{x_2 - 0.5}{\varepsilon}\right), \quad u(x_1, x_2) = x_1(1 - x_1)x_2(1 - x_2).$$

Figure 6.8 shows the exact state and control.

For small ε the exact state has an interior layer along the line $x_2 = 0.5$. The SIPG method without special treatment does not resolve the interior layer even in the case of a single equation. Actually, since the mesh is aligned with the layer, the SIPG method just ignores the layer. Because of the coupling the computed control is

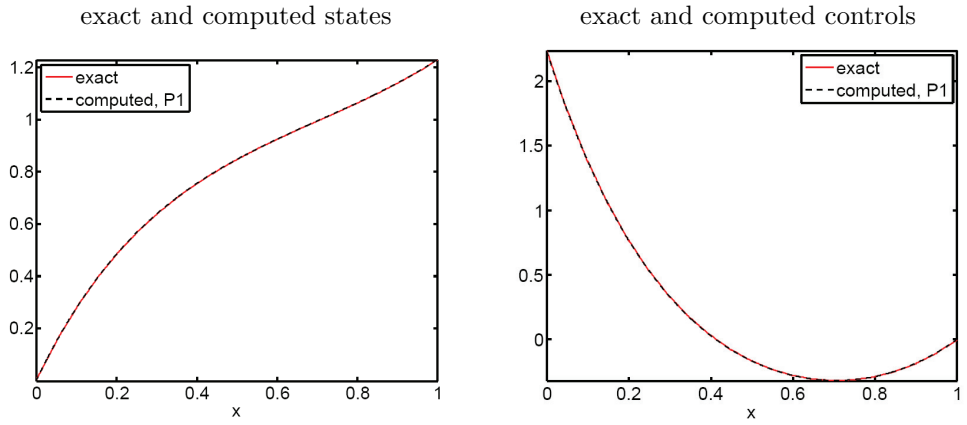

FIG. 6.6. *Exact and computed states (left plot) and controls (right plot) for Example* 3. *The computed states and adjoint are obtained using the SIPG method with piecewise linear* ($P1$) *elements on a uniform mesh with mesh size* $h = 1/20$.

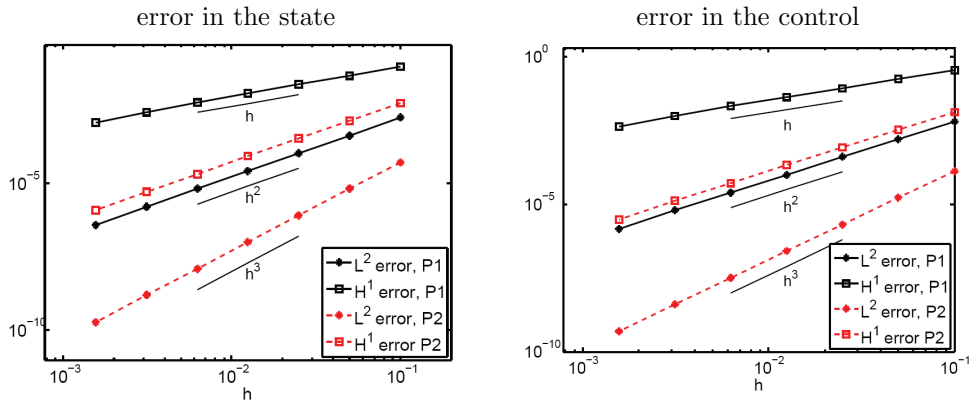

FIG. 6.7. *The left (right) plot shows the* L^2- *and* H^1-*errors between the exact and computed state (control) for Example* 3 *on the subdomain* $\Omega_0 = (0.1, 0.9)$, *when the computed solution is obtained using the SIPG method with piecewise linear* ($P1$) *and piecewise quadratic* ($P2$) *elements.*

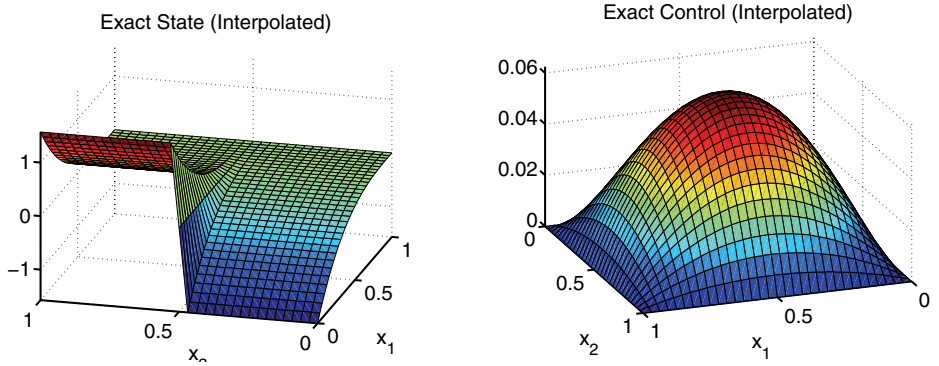

FIG. 6.8. *Exact state and adjoint for Example* 4.

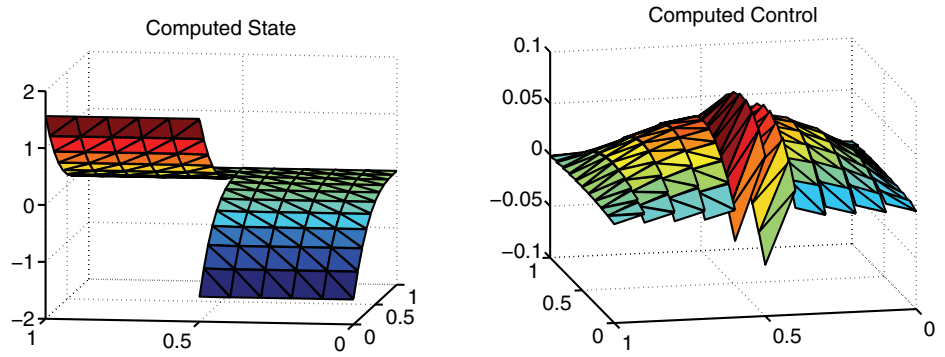

FIG. 6.9. *Computed state and control for Example* 4 *using SIPG with piecewise linear elements on a uniform mesh with mesh size* $h = 1/10$.

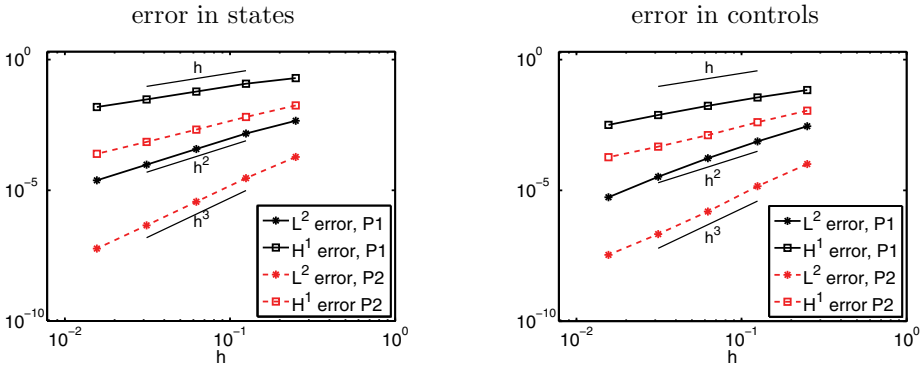

FIG. 6.10. *The left (right) plot shows the* L^2*- and* H^1*-errors between the exact and computed state (control) for Example* 4 *on the subdomain* $\Omega_0 = [0,\ 1] \times [0.6,\ 1]$*, when the computed solution is obtained using SIPG with piecewise linear (*$P1$*) and piecewise quadratic (*$P2$*) elements.*

not resolved along the location of the interior layer, the line $x_2 = 0.5$, despite the fact that the exact control is smooth (cf. Figure 6.9). On the other hand, Theorem 5.1 says that the interior layers do not pollute the SIPG solutions into the subdomains of smoothness. This fact we observe numerically in Figure 6.10.

7. Conclusions. We have provided a careful local error analysis of the SIPG discretization of distributed optimal control problems governed by advection-dominated elliptic PDEs. We have proven that in the presence of boundary layers the convergence rate is optimal in the interior of the domain. This is in sharp contrast to the convergence behavior of SUPG discretizations of the same optimal control problem [17]. Numerical examples indicate that this favorable behavior of the SIPG discretization is due to the weak imposition of Dirichlet boundary conditions. In addition we have proven that in the presence of interior layers the convergence rate for the SIPG discretized solution is optimal in regions away from the interior layer. The same convergence behavior was proven in an earlier paper [17] for the SUPG discretization.

Acknowledgment. The authors thank the referees for their constructive comments which led to improvements on the presentation.

REFERENCES

[1] F. Abraham, M. Behr, and M. Heinkenschloss, *The effect of stabilization in finite element methods for the optimal boundary control of the Oseen equations*, Finite Elem. Anal. Des., 41 (2004), pp. 229–251.

[2] B. Ayuso and L. D. Marini, *Discontinuous Galerkin methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 47 (2009), pp. 1391–1420.

[3] Y. Bazilevs and T. J. R. Hughes, *Weak imposition of Dirichlet boundary conditions in fluid mechanics*, Comput. & Fluids, 36 (2007), pp. 12–26.

[4] R. Becker and B. Vexler, *Optimal control of the convection-diffusion equation using stabilized finite element methods*, Numer. Math., 106 (2007), pp. 349–367.

[5] M. Braack, *Optimal control in fluid mechanics by finite elements with symmetric stabilization*, SIAM J. Control Optim., 48 (2009), pp. 672–687.

[6] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 2nd ed., Springer Verlag, New York, 2002.

[7] F. Brezzi, B. Cockburn, L. D. Marini, and E. Süli, *Stabilization mechanisms in discontinuous Galerkin finite element methods*, Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 3293–3310.

[8] G. Chen and S. S. Collis, *Optimal control for Burgers flow using the discontinuous Galerkin method*, in AIAA Region IV Student Paper Conference, 2003.

[9] B. Cockburn, *Discontinuous Galerkin methods*, ZAMM Z. Angew. Math. Mech., 83 (2003), pp. 731–754.

[10] B. Cockburn, B. Dong, and J. Guzmán, *Optimal convergence of the original DG method for the transport-reaction equation on special meshes*, SIAM J. Numer. Anal., 46 (2008), pp. 1250–1265.

[11] S. S. Collis and M. Heinkenschloss, *Analysis of the Streamline Upwind/Petrov Galerkin Method Applied to the Solution of Optimal Control Problems*, Technical Report TR02–01, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005–1892, 2002; available online at http://www.caam.rice.edu/∼heinken.

[12] L. Dedé and A. Quarteroni, *Optimal control and numerical adaptivity for advection-diffusion equations*, M2AN Math. Modell. Numer. Anal., 39 (2005), pp. 1019–1040.

[13] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, RI, 1998.

[14] J. Freund and R. Stenberg, *On weakly imposed boundary conditions for second order problems*, in Proceedings of the Ninth International Conference Finite Elements in Fluids, M. M. C. et al., eds., Venice, 1995, pp. 327–336.

[15] J. Gopalakrishnan and G. Kanschat, *A multilevel discontinuous Galerkin method*, Numer. Math., 95 (2003), pp. 527–550.

[16] J. Guzmán, *Local analysis of discontinuous Galerkin methods applied to singularly perturbed problems*, J. Numer. Math., 14 (2006), pp. 41–56.

[17] M. Heinkenschloss and D. Leykekhman, *Local error estimates for SUPG solutions of advection-dominated elliptic linear-quadratic optimal control problems*, SIAM J. Numer. Anal., 47 (2010), pp. 4607–4638.

[18] M. Hinze, N. Yan, and Z. Zhou, *Variational discretization for optimal control governed by convection dominated diffusion equations*, J. Comput. Math., 27 (2009), pp. 237–253.

[19] P. Houston, C. Schwab, and E. Süli, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 39 (2002), pp. 2133–2163.

[20] C. Johnson and J. Pitkäranta, *An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation*, Math. Comp., 46 (1986), pp. 1–26.

[21] C. Johnson, A. H. Schatz, and L. B. Wahlbin, *Crosswind smear and pointwise errors in streamline diffusion finite element methods*, Math. Comp., 49 (1987), pp. 25–38.

[22] D. Leykekhman, *Investigation of commutative properties of discontinuous Galerkin methods in PDE constrained optimal control problems*, J. Sci. Comput., submitted.

[23] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer Verlag, New York, Berlin, 1971.

[24] G. Lube and B. Tews, *Distributed and boundary control of singularly perturbed advection-diffusion-reaction problems*, in BAIL 2008—Boundary and interior layers, Lect. Notes Comput. Sci. Eng. 69, Springer, Berlin, 2009, pp. 205–215.

[25] J. Rauch, *$\mathcal{L}_2$ is a continuable initial condition for Kreiss' mixed problems*, Comm. Pure Appl. Math., 25 (1972), pp. 265–285.

[26] F. Schieweck, *On the role of boundary conditions for CIP stabilization of higher order finite elements*, Electron. Trans. Numer. Anal., 32 (2008), pp. 1–16.

[27] R. Stenberg, *On some techniques for approximating boundary conditions in the finite element method*, J. Comput. Appl. Math., 63 (1995), pp. 139–148.
[28] M. Stynes, *Steady-state convection-diffusion problems*, Acta Numer., 14 (2005), pp. 445–508.
[29] L. B. Wahlbin, *Local behavior in finite element methods*, in Handbook of Numerical Analysis, Vol. II, Handb. Numer. Anal., II, North-Holland, Amsterdam, 1991, pp. 353–522.
[30] Z. Zhou and N. Yan, *The local discontinuous Galerkin method for optimal control problem governed by convection diffusion equations*, Int. J. Numer. Anal. Model., 7 (2010), pp. 681–699.
[31] P. Zunino, *Discontinuous Galerkin methods based on weighted interior penalties for second order PDEs with non-smooth coefficients*, J. Sci. Comput., 38 (2009), pp. 99–126.