A ROBUST NUMERICAL METHOD FOR A CONTROL PROBLEM INVOLVING SINGULARLY PERTURBED EQUATIONS*

ALEJANDRO ALLENDES†, ERWIN HERNÁNDEZ‡, AND ENRIQUE OTÁROLA§

Abstract. We consider an unconstrained linear–quadratic optimal control problem governed by a singularly perturbed convection–reaction–diffusion equation. We discretize the optimality system by using standard piecewise bilinear finite elements on the graded meshes introduced by Durán and Lombardi in [17, 18]. We prove convergence of this scheme. In addition, when the state equation is a singularly perturbed reaction–diffusion equation, we derive quasi–optimal a priori error estimates for the approximation error of the optimal variables on anisotropic meshes. We present several numerical experiments when the state equation is both a reaction–diffusion and a convection-reaction–diffusion equation. These numerical experiments reveal a competitive performance of the proposed solution technique.

Key words. linear–quadratic optimal control problem, convection–reaction–diffusion equation, singularly perturbed equation, finite elements, graded meshes, anisotropic estimates.

AMS subject classifications. 49J20, 49M25, 65N12, 65N30.

1. Introduction.

We are interested in the study of new and efficient solution techniques for an unconstrained linear–quadratic optimal control problem involving a convection–reaction-diffusion equation. To be precise, let $\Omega = (0,1)^2$. Given a desired state $\mathsf{y}_d : \Omega \to \mathbb{R}$, we define the cost functional

$$J(\mathsf{y},\mathsf{u}) = \frac{1}{2}\|\mathsf{y}-\mathsf{y}_d\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|\mathsf{u}\|^2_{L^2(\Omega)}, \tag{1.1}$$

where $\lambda > 0$ denotes the so-called regularization parameter. Let $\mathsf{c}, \mathsf{f} : \Omega \to \mathbb{R}$ be fixed functions and $\mathsf{b} : \Omega \to \mathbb{R}^2$ be a given vector field. We shall be concerned with the following optimal control problem: Find

$$\min J(\mathsf{y},\mathsf{u}) \tag{1.2}$$

subject to the singularly perturbed convection–reaction–diffusion equation

$$-\varepsilon^2\Delta\mathsf{y} + \mathsf{b}\cdot\nabla\mathsf{y} + \mathsf{c}\mathsf{y} = \mathsf{u} + \mathsf{f} \quad \text{in} \quad \Omega, \qquad \mathsf{y} = 0 \quad \text{on} \quad \partial\Omega, \tag{1.3}$$

where ε denotes the *perturbation parameter* and satisfies $0 < \varepsilon \ll 1$. The term u denotes the *control variable*, and y, the solution to the *state equation* (1.3), corresponds to the *state variable*. We will also be interested in the particular scenario where $\mathsf{b} \equiv 0$, which yields a singularly perturbed reaction–diffusion problem as state equation:

$$-\varepsilon^2\Delta\mathsf{y} + \mathsf{c}\mathsf{y} = \mathsf{u} + \mathsf{f} \quad \text{in} \quad \Omega, \qquad \mathsf{y} = 0 \quad \text{on} \quad \partial\Omega. \tag{1.4}$$

*AA has been supported in part by CONICYT through FONDECYT project 11121243 and by the USM project 12.14.2. EH has been supported in part by CONICYT through projects FONDECYT 1140392 and Anillo ACT1106. EO has been supported in part by CONICYT through projects FONDECYT 3160201 and Anillo ACT1106.

†Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. alejandro.allendes@usm.cl

‡Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. erwin.hernandez@usm.cl

§Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. enrique.otarola@usm.cl

Since the optimal control problem (1.2)–(1.3) does not involve control constraints, the associated optimality system corresponds to a coupled one involving two singularly perturbed convection–reaction–diffusion equations: the state and adjoint equations; see [10, 14, 25, 30]. The adjoint equation has a convection component that is the negative of the one appearing in (1.3) [14, 25]. This leads to the computational challenge of how to efficiently resolve the optimality system associated with (1.2)–(1.3), which in turn demands the use of an efficient method to solve the state equation (1.3). When solving the latter, it is known that standard finite element techniques lead to strongly oscillatory solutions unless the mesh–size is sufficiently small with respect to the ratio between ε and $\|\mathsf{b}\|$. In addition, the sharp boundary and interior layers and corner and edge singularities, that usually appear in the solution to these types of problems, must be efficiently resolved [7, 27, 39, 41].

In the context of optimal control problems, to overcome such difficulties, different stabilized finite element techniques have been proposed and analyzed in the literature; see [1, 10, 14, 16, 26, 35, 46, 47]. To the best of our knowledge, the first work that analyzed a stabilized scheme is [14]. This work considers the streamline upwind/Petrov Galerkin (SUPG) method, elaborates on the fact that the optimize–then–discretize and discretize–then–optimize approaches do not coincide and explores their respective advantages. Later, local projection stabilization (LPS) techniques were proposed in [10]. These techniques have the advantage that, due to the symmetry of the proposed stabilization term, optimize–then–discretize and discretize–then–optimize coincide. In addition, the authors derive global a priori error estimates for the approximation error of the optimal variables. However, it is important to notice, as pointed out in [25], that these global error estimates (see also [10, 11, 26, 35, 46]) contain constants that depend on derivatives of the optimal variables and so in the presence of interior and boundary layers such estimates become meaningless. This motivates the local a priori error analysis of [25], where the SUPG method is used to approximate the solution of the state equation (1.3). The authors derive local a priori error estimates in subdomains $\Omega_0 \subset \Omega$ that do not include boundary or interior layers. However, the presence of boundary layers pollute the numerical solution, even in subregions where the solution is smooth. This is due to the fact that the boundary layers are not sufficiently resolved; see [25] for a discussion.

In some particular cases, where some information on the behavior of the solution to the single state equation (1.3) or (1.4) is available, it is possible to design a priori graded meshes to efficiently approximate the boundary layers exhibited by the solution; see [17, 18, 29, 41]. Some well–known approximation techniques of this kind are those based on the so-called Shishkin meshes [9, 34, 41]. In the context of the optimal control theory, some numerical experiments provided in [25, Section 5.2] show promising results when Shishkin meshes are considered. In [40], the one dimensional version of (1.2)–(1.3) is analyzed and discretized using these meshes. The authors derive optimal rates of convergence in the energy–norm, but the presented error estimates for the L^2–norm are not optimal in terms of approximation.

In this work we propose a different approach based on the finite element approximation of the optimality system, associated with the optimal control problem (1.2)–(1.3), on the graded or anisotropic meshes proposed by Durán and Lombardi in [17, 18]. In the case that $\mathsf{b} = 0$, i.e., when (1.4) corresponds to the state equation, we propose a solution technique and derive a priori quasi–optimal error estimates in both the energy and the L^2–norm. The error estimates derived in the energy–norm are ε-independent [19, 20], which is a property that Shishkin meshes do not satisfy.

We also propose an approximation scheme for solving the optimal control problem (1.2)–(1.3), where the state equation (1.3) contains a non–zero constant vector field b. We prove the convergence of the scheme by invoking the theory of Γ–convergence. We design several computational experiments that show a competitive performance of the proposed solution technique when it is compared with adaptive stabilized schemes. In addition, we observe that

- the experimental rates of convergence in both the energy and the L^2–norm are quasi–optimal in terms of approximation.
- the pollution effect discussed in [25] is not observed. This is due to the fact that the boundary layers are appropriately approximated.

We comment that, the numerical analysis of the proposed approximation scheme to solve (1.2)–(1.3), where b is a non–zero and constant vector field, could be derived under strong assumptions on the optimal control variable; see Section 4.2.1 for a discussion. The main disadvantage of our graded-mesh scheme is that it is based on the anisotropic error estimates analyzed in [17] which are valid under a tensor product structure of the domain Ω.

The outline of this paper is as follows. In Section 2 we introduce the functional framework that is suitable for analysing the optimal control problem (1.2)–(1.3) and state existence and uniqueness results in conjunction with optimality conditions. Section 3 is a review of the graded finite element techniques developed in [18] and [17] to solve the state equations (1.3) and (1.4), respectively. In Section 4 we propose a numerical technique to solve the optimal control problem (1.2)–(1.3): a fully discrete scheme that discretizes the optimal control and state with standard piecewise bilinear finite elements on anisotropic meshes. Subsection 4.1 contains a complete quasi–optimal a priori error analysis, in both the energy and the L^2–norm, when (1.4) is considered as the state equation. In Subsection 4.2, we assume that b is a non–zero vector field and prove the convergence of the proposed scheme via the theory of Γ–convergence. Finally, in Section 5 we present several numerical experiments that reveal a competitive performance of the proposed solution technique. We also explore computationally the performance of our method when solving a constrained optimal control problem with (1.3) as the state equation.

Throughout the manuscript, the relation $a \lesssim b$ indicates that $a \leq Cb$, with a constant C that does not depend on either a or b. The value of C might change at each occurrence.

2. The optimal control problem. In this section, we describe the optimal control problem (1.2)–(1.3). Following [10, 25], we present existence and uniqueness results together with necessary and sufficient optimality conditions. Since (1.2)–(1.3) is unconstrained, we reformulate the optimality conditions as a coupled system of partial differential equations (PDE). On the basis of this idea, we propose a numerical technique to solve the optimal control problem (1.2)–(1.3).

We start by considering some suitable functional analysis setting. We denote by $\mathbb{Y}$ the so–called state space. This corresponds to the standard Sobolev space $H_0^1(\Omega)$ equipped with the ε-weighted energy norm, which is defined by

$$\|\mathsf{v}\|_{\varepsilon}^2 = \varepsilon^2 \|\nabla \mathsf{v}\|_{L^2(\Omega)}^2 + \|\mathsf{v}\|_{L^2(\Omega)}^2 \qquad \forall \mathsf{v} \in \mathbb{Y}.$$

Since (1.2)–(1.3) is unconstrained, we define the set of admissible controls as $\mathbb{U}_{\text{ad}} = L^2(\Omega)$. To state existence and uniqueness results we assume that:
(1) $0 < \varepsilon \ll 1$.

(2) $\mathsf{c} \in L^\infty(\Omega)$ and $\mathsf{b} \in [W^{1,\infty}(\Omega)]^2$ satisfy $\mathsf{c} \geq 0$ and

$$\mathsf{c} - \frac{1}{2}\mathrm{div}\,\mathsf{b} \geq \mu > 0.$$

(3) $\mathsf{f}, \mathsf{y}_d \in L^2(\Omega)$.

The standard weak formulation of (1.3) reads as follows: find $\mathsf{y} = \mathsf{y}(\mathsf{u}) \in \mathbb{Y}$ such that

$$\mathcal{B}(\mathsf{y}, \mathsf{v}) = (\mathsf{f} + \mathsf{u}, \mathsf{v})_{L^2(\Omega)} \qquad \forall\, \mathsf{v} \in \mathbb{Y}, \tag{2.1}$$

where $(\cdot,\cdot)_{L^2(\Omega)}$ denotes the standard inner product in $L^2(\Omega)$ and $\mathcal{B}$ is defined by

$$\mathcal{B}(\mathsf{y}, \mathsf{v}) = \int_\Omega \varepsilon^2 \nabla \mathsf{y} \cdot \nabla \mathsf{v} + \mathsf{b} \cdot \nabla \mathsf{y} \mathsf{v} + \mathsf{c} \mathsf{y} \mathsf{v}, \qquad \mathsf{y}, \mathsf{v} \in \mathbb{Y}. \tag{2.2}$$

Under the assumptions (1)–(3), the bilinear form $\mathcal{B}$ is coercive in $\mathbb{Y} \times \mathbb{Y}$, i.e., there exists $\beta > 0$, independent of ε, such that

$$\beta \|\mathsf{v}\|_\varepsilon^2 \leq \mathcal{B}(\mathsf{v}, \mathsf{v}) \qquad \forall \mathsf{v} \in \mathbb{Y}. \tag{2.3}$$

For a given control $\mathsf{u} \in \mathbb{U}_{\rm ad}$ and data satisfying the assumptions (1)–(3), the Lax-Milgram Lemma implies the well–posedness of (2.1). We are then in position to introduce the so–called control–to–state operator.

Definition 2.1 (control–to–state operator). *We define the control to state operator* $\mathbf{S} : \mathbb{U}_{ad} \to \mathbb{Y}$ *such that for a given control* $\mathsf{u} \in \mathbb{U}_{ad}$ *it associates a unique state* $\mathsf{y} = \mathsf{y}(\mathsf{u}) \in \mathbb{Y}$ *via the state equation* (2.1).

It follows that $\mathbf{S}$ is a linear and continuous map from $H^{-1}(\Omega)$ into $\mathbb{Y}$. In view of the continuous embedding $H^{-1}(\Omega) \hookrightarrow L^2(\Omega) \hookrightarrow \mathbb{Y}$, we may also consider $\mathbf{S}$ acting from $L^2(\Omega)$ and with range in $L^2(\Omega)$. With this operator at hand, we define the reduced cost functional

$$f(\mathsf{u}) = \frac{1}{2}\|\mathbf{S}\mathsf{u} - \mathsf{y}_d\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|\mathsf{u}\|^2_{L^2(\Omega)}. \tag{2.4}$$

We now define the optimal state-control pair as follows.

Definition 2.2 (optimal state-control pair). *A state-control pair* $(\bar{\mathsf{y}}(\bar{\mathsf{u}}), \bar{\mathsf{u}}) \in \mathbb{Y} \times \mathbb{U}_{ad}$ *is called optimal for problem* (1.2)–(1.3) *if* $\bar{\mathsf{y}}(\bar{\mathsf{u}}) = \mathbf{S}\bar{\mathsf{u}}$ *and*

$$J(\bar{\mathsf{y}}(\bar{\mathsf{u}}), \bar{\mathsf{u}}) \leq J(\mathsf{y}(\mathsf{u}), \mathsf{u}),$$

for all $(\mathsf{y}(\mathsf{u}), \mathsf{u}) \in \mathbb{Y} \times \mathbb{U}_{ad}$ *such that* $\mathsf{y}(\mathsf{u}) = \mathbf{S}\mathsf{u}$.

We define the adjoint state associated to the state $\mathsf{y} = \mathsf{y}(\mathsf{u})$ as follows.

Definition 2.3 (adjoint state). *Given a control* $\mathsf{u} \in \mathbb{U}_{ad}$, *we define the adjoint state* $\mathsf{p} = \mathsf{p}(\mathsf{u}) \in \mathbb{Y}$ *as the solution to*

$$\mathsf{p} \in \mathbb{Y}: \qquad \mathcal{B}(\mathsf{q}, \mathsf{p}) = (\mathsf{y}(\mathsf{u}) - \mathsf{y}_d, \mathsf{q})_{L^2(\Omega)}, \tag{2.5}$$

for all $\mathsf{q} \in \mathbb{Y}$.

We present the following result about existence and uniqueness of the optimal control together with necessary and sufficient optimality conditions for (1.2)–(1.3); see [10, Theorem 1 and Lemma 1] and [25, Theorem 2.1].

Theorem 2.4 (existence, uniqueness and optimality conditions). *The optimal control problem* (1.2)–(1.3) *has a unique optimal solution* $(\bar{\mathsf{y}}(\bar{\mathsf{u}}), \bar{\mathsf{u}}) \in \mathbb{Y} \times \mathbb{U}_{ad}$. *The optimality conditions*

$$\left\{ \begin{array}{lrcll} \bar{\mathsf{y}} \in \mathbb{Y}: & \mathcal{B}(\bar{\mathsf{y}}, \mathsf{v}) & = & (\mathsf{f} + \bar{\mathsf{u}}, \mathsf{v}) & \forall \mathsf{v} \in \mathbb{Y}, \\ \bar{\mathsf{p}} \in \mathbb{Y}: & \mathcal{B}(\mathsf{v}, \bar{\mathsf{p}}) & = & (\bar{\mathsf{y}} - \mathsf{y}_d, \mathsf{v}) & \forall \mathsf{v} \in \mathbb{Y}, \\ & \lambda\bar{\mathsf{u}} + \bar{\mathsf{p}} & = & 0, & \end{array} \right. \tag{2.6}$$

hold. These conditions are necessary and sufficient.

Theorem 2.4, in conjunction with the fact that $\lambda > 0$, yields immediately the following reduced and equivalent optimality system: find $(\bar{\mathsf{y}}, \bar{\mathsf{p}}) \in \mathbb{Y} \times \mathbb{Y}$ such that

$$\left\{ \begin{array}{rcll} \lambda \mathcal{B}(\bar{\mathsf{y}}, \mathsf{v}) + (\bar{\mathsf{p}}, \mathsf{v})_{L^2(\Omega)} & = & \lambda (\mathsf{f}, \mathsf{v})_{L^2(\Omega)} & \forall\, \mathsf{v} \in \mathbb{Y}, \\ \mathcal{B}(\mathsf{q}, \bar{\mathsf{p}}) - (\bar{\mathsf{y}}, \mathsf{q})_{L^2(\Omega)} & = & -(\mathsf{y}_d, \mathsf{q})_{L^2(\Omega)} & \forall\, \mathsf{q} \in \mathbb{Y}. \end{array} \right. \tag{2.7}$$

Inspired in this reformulation of the optimality conditions, a natural and simple technique to solve the optimal control problem is the following: given $\lambda > 0$, $\mathsf{f} \in L^2(\Omega)$ and $\mathsf{y}_d \in L^2(\Omega)$, we solve the coupled PDE system (2.7), thereby obtaining $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$. Setting $\bar{\mathsf{u}} = -\frac{1}{\lambda}\bar{\mathsf{p}}$, we obtain the optimal variables $\bar{\mathsf{u}}$ and $\bar{\mathsf{y}}$ solving (1.2)–(1.3).

We finally define the auxiliary bilinear form

$$\mathcal{A}((\mathsf{y}, \mathsf{p}), (\mathsf{v}, \mathsf{q})) := \lambda \mathcal{B}(\mathsf{y}, \mathsf{v}) + (\mathsf{p}, \mathsf{v})_{L^2(\Omega)} + \mathcal{B}(\mathsf{q}, \mathsf{p}) - (\mathsf{y}, \mathsf{q})_{L^2(\Omega)}, \tag{2.8}$$

for all $(\mathsf{y}, \mathsf{p}), (\mathsf{v}, \mathsf{q}) \in \mathbb{Y} \times \mathbb{Y}$.

Since an efficient technique to solve (2.7) relies on a method to solve the state equations (1.3) and (1.4), in the next section we describe simply and efficient approaches to solve such equations.

3. Discretization of the state equation. In this section we review the graded finite element approximation techniques developed by Durán and Lombardi in [17] and [18] for the state equations (1.3) and (1.4), respectively. To accomplish this task, we start by introducing some terminology and describing the construction of the underlying finite element spaces. We recall that $\Omega = (0,1)^2$. We consider a mesh $\mathscr{T}$ of Ω that is composed of rectangular elements T, with sides parallel to the coordinate axes, such that

$$\bar{\Omega} = \bigcup_{T \in \mathscr{T}} T \quad \text{and} \quad |\Omega| = \sum_{T \in \mathscr{T}} |T|.$$

The mesh $\mathscr{T}$ is assumed to be conforming or compatible. We denote by $\mathrm{N}_{\mathscr{T}}$ the number of degrees of freedom of the partition $\mathscr{T}$, and given $T \in \mathscr{T}$, we define h_T^i as the length of T in the i-th direction, where $i = 1, 2$.

Since we will be working with anisotropic discretizations that do not satisfy the standard shape regularity condition [12, 21], we assume the following *weak shape regularity condition* [17, 36]: there exists a constant $\sigma > 1$ such that, if $T, S \in \mathscr{T}$ are neighboring elements, then

$$h_T^i (h_S^i)^{-1} \leq \sigma, \qquad i = 1, 2. \tag{3.1}$$

We define the finite element space $\mathbb{V}(\mathscr{T})$ by

$$\mathbb{V}(\mathscr{T}) = \left\{ \mathsf{w}_{\mathscr{T}} \in \mathcal{C}^0(\overline{\Omega}) : \mathsf{w}_{\mathscr{T}}|_T \in \mathcal{Q}_1(T)\ \forall\, T \in \mathscr{T},\ \mathsf{w}_{\mathscr{T}}|_{\partial\Omega} = 0 \right\}, \tag{3.2}$$

where, for an element T, the set $\mathcal{Q}_1(T)$ denotes the space of polynomials of degree not larger than one in each variable.

The Galerkin approximation of the solution to (1.3) is given by the unique function $\mathsf{y}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$ that solves the discrete problem

$$\mathcal{B}(\mathsf{y}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}}) = (\mathsf{f} + \mathsf{u}, \mathsf{v}_{\mathscr{T}})_{L^2(\Omega)} \qquad \forall\, \mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \tag{3.3}$$

where $\mathcal{B}$ is defined by (2.2).

The discretization and analysis developed in [17, 18] for the state equation (1.3) differs whether $\mathsf{b} \equiv 0$ or $\mathsf{b} \not\equiv 0$. Therefore, we proceed to review and analyze each approximation scheme separately.

3.1. The reaction-diffusion state equation. In this case the state equation is (1.4) and then the bilinear form $\mathcal{B}$, defined by (2.2), is continuous and coercive on $\mathbb{Y} \times \mathbb{Y}$ with constants that do not depend on ε. For convenience, we will consider $\mathsf{c} \equiv 1$ in (1.4). Then, invoking Galerkin orthogonality [12, 21] we deduce that

$$\|\mathsf{y} - \mathsf{y}_{\mathscr{T}}\|_\varepsilon \leq \|\mathsf{y} - \mathsf{w}_{\mathscr{T}}\|_\varepsilon \quad \forall \mathsf{w}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}). \tag{3.4}$$

Consequently, the numerical analysis of (1.4) reduces to a result in approximation theory: the distance between y and $\mathsf{y}_{\mathscr{T}}$ is bounded by the best approximation error in the finite element space with respect to the ε-norm. To bound such an approximation error it is standard to consider $\mathsf{w}_{\mathscr{T}} = \Pi_{\mathscr{T}} \mathsf{y}$ in (3.4), where $\Pi_{\mathscr{T}}$ denotes a suitable interpolation operator, and then invoke interpolation error estimates, which in turn rely on the fact that y typically exhibits sharp boundary layers of width $\mathcal{O}(\varepsilon|\log\varepsilon|)$ along the boundary $\partial\Omega$. To approximate efficiently such a singular behavior, we consider the graded mesh introduced in [17] as a realization of $\mathscr{T}$. In fact, let $h, \gamma \in (0,1)$ be fixed. We consider a partition $\{\xi_i\}_{i=0}^M$ of the interval $[0, \frac{1}{2}]$ given by

$$\xi_0 = 0, \quad \xi_1 = h^{1/(1-\gamma)}, \quad \xi_{i+1} = \xi_i + h\xi_i^\gamma \text{ for } j = 1, \dots M-2, \quad \xi_M = \tfrac{1}{2}, \tag{3.5}$$

where M is such that $\xi_{M-1} < \frac{1}{2}$, $\xi_{M-1} + h\xi_{M-1}^\gamma \geq \frac{1}{2}$ and $\xi_M = \frac{1}{2}$. If $\frac{1}{2} - \xi_{M-1} < \xi_{M-1} - \xi_{M-2}$ the definition of ξ_{M-1} is modified as follows: $\xi_{M-1} = (\frac{1}{2} + \xi_{M-2})/2$. Invoking symmetry we define a graded partition on $[\frac{1}{2}, 1]$. Collecting this two meshes we obtaining a partition $\{\xi_i\}_{i=0}^{2M}$ of $[0,1]$. We then define a graded mesh $\mathscr{T} = \{T_{ij}\}_{i,j=1}^{2M}$ based on the rectangular elements $T_{ij} = [\xi_{i-1}, \xi_i] \times [\xi_{j-1}, \xi_j]$. Figure 3.1 shows an example of this type of meshes. We remark that the family of meshes $\{\mathscr{T}\}$ satisfies the weak regularity condition (3.1) with a constant $\sigma = 2^\gamma$.

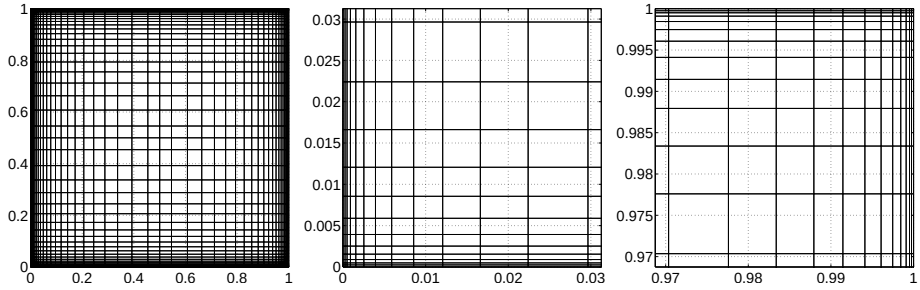

FIG. 3.1. *The left panel shows a graded meshes $\mathscr{T}$ constructed on the basis of* (3.5) *with* $\gamma = 0.75$. *The number of degrees of freedom is* $N_{\mathscr{T}} = 2916$. *The center panel presents a zoom of the graded mesh near to the origin and the right panels shows a zoom near to* $(1,1)$.

REMARK 3.1 (ε-independent mesh). The graded mesh defined by (3.5) is ε-independent. Computationally, this property has the advantage that one graded mesh solves efficiently (3.3) for a whole range of perturbation parameters. This is a property that Shishkin meshes do not have; see [20, Section 5].

We now consider the finite element space $\mathbb{V}(\mathscr{T})$ on these graded meshes. Under this setting, referece [17] provides a quasi-interpolation operator $\Pi_{\mathscr{T}}$ which is built on local averages over stars and has optimal approximation properties on anisotropic meshes; the theory has been recently extended to a Muckenhoupt weighted Sobolev space setting [36]. Then, these results, in view of (3.4), allow us to control $\|\mathsf{y} - \mathsf{y}_{\mathscr{T}}\|_\varepsilon$. To state the precise estimate, we introduce the following notation: we say that ϕ

satisfies (3.6) if

$$\phi \in C^2([0,1]^2) \quad \text{and} \quad \phi(0,0) = \phi(1,0) = \phi(0,1) = \phi(1,1) = 0. \tag{3.6}$$

Then, if u and f satisfy (3.6) and $\frac{1}{2} \leq \gamma < 1$, we have that

$$\|\mathsf{y} - \mathsf{y}_{\mathscr{T}}\|_{\varepsilon} \lesssim |\log \mathrm{N}_{\mathscr{T}}| \mathrm{N}_{\mathscr{T}}^{-1/2}, \tag{3.7}$$

where the hidden constant is independent of ε and $N_{\mathscr{T}}$; see [17, Corollary 4.5]. We remark that (3.7) is called quasi-optimal because of the logarithmic term $|\log \mathrm{N}_{\mathscr{T}}|$. Notice that, up to this logarithmic factor, the estimate (3.7) is the same as the one obtained with uniform refinement for the approximation of a smooth solution. We conclude with the following quasi–optimal estimate in the L^2–norm [20, Corollary 4.9]. If $\frac{3}{4} \leq \gamma < 1$, then

$$\|\mathsf{y} - \mathsf{y}_{\mathscr{T}}\|_{L^2(\Omega)} \lesssim \log\left(\tfrac{1}{\varepsilon}\right)^{1/2} |\log \mathrm{N}_{\mathscr{T}}|^2 \mathrm{N}_{\mathscr{T}}^{-1}, \tag{3.8}$$

where the hidden constant is independent of ε and $\mathrm{N}_{\mathscr{T}}$. We remark that the derivation of (3.8) is non–trivial and involves fine results in analysis such as weighted-Poincaré inequalities and the use of the Hardy–Littlewood maximal function; see [20] for details.

3.2. Convection–reaction–diffusion equation. In this subsection, we consider the graded finite element approximation developed in [17] for the convection–reaction–diffusion equation (2.1). We start by commenting that, since the bilinear form (2.2) is not uniformly continuous with respect to the parameter ε, the standard finite element theory based on Cea's Theorem cannot be applied. However, [18] hinges on the explicit behavior of the solution to (1.3) and provide optimal interpolation estimates, which are the basis to obtain a priori nearly-optimal error estimates. We review this theory in what follows.

From now on, and following [18, 19], we assume that $\mathsf{b} = (\mathsf{b}_1, \mathsf{b}_2)$ is a constant vector such that $\mathsf{b}_i < -\delta$ for $i = 1, 2$ and $\delta > 0$. Then, the solution to (1.3) presents a boundary layer of width $\mathcal{O}(\varepsilon^2 |\log \varepsilon|)$ at the outflow boundary $\{(x, y) \in \Omega : x = 0 \text{ or } y = 0\}$. In order to design a graded scheme that captures such a singular behavior, we proceed as follows [18]. Given a parameter $h \in (0, 1)$, we define a partition $\{\xi_i\}_{i=0}^{M}$ of the interval $[0, 1]$ based on the mesh–points

$$\xi_0 = 0, \quad \xi_1 = h\varepsilon^2, \quad \xi_{i+1} = \xi_i + h\xi_i \text{ for } j = 1, \dots M-2, \quad \xi_M = 1, \tag{3.9}$$

where M is such that $\xi_{M-1} < 1$ and $\xi_{M-1} + h\xi_{M-1}^{\gamma} \geq 1$. If $1 - \xi_{M-1} < \xi_{M-1} - \xi_{M-2}$ the definition of ξ_{M-1} is modified as follows: $\xi_{M-1} := (1 + \xi_{M-2})/2$. We then define a graded mesh $\mathscr{T} = \{T_{ij}\}_{i,j=1}^{2M}$ where each rectangle T_{ij} is defined by $T_{ij} = [\xi_{i-1}, \xi_i] \times [\xi_{j-1}, \xi_j]$. Figure 3.2 shows an example of this type of meshes.

We consider a family of finite element spaces defined by (3.2) on the graded meshes constructed on the basis of (3.9). We now present the error estimates derived in [18, 19]. To state these estimates appropriately, we introduce the following notation: we say that ϕ satisfies (3.10) if

$$\begin{cases} \phi \in C^4((0,1)^2), \quad \phi(0,0) = \phi(1,0) = \phi(0,1) = \phi(1,1) = 0, \\ \dfrac{\partial^{i+j}\phi}{\partial x_1^i \partial x_2^j}(1,1) = 0, \quad \text{for} \quad 0 \leq i + j \leq 2, \end{cases} \tag{3.10}$$

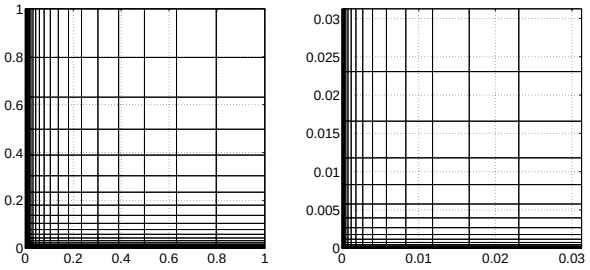

FIG. 3.2. *The left panel shows a graded mesh $\mathscr{T}$ constructed on the basis of* (3.5) *with* $\varepsilon^2 = 10^{-4}$. *The number of degrees of freedom is* $N_{\mathscr{T}} = 1156$. *The right panel presents a zoom of the mesh near to the origin.*

Then, if f and u satisfy (3.10), we have the following quasi-optimal error estimate [18, Corollary 2.3]:

$$\|\mathsf{y} - \mathsf{y}_{\mathscr{T}}\|_{\varepsilon} \lesssim \log^2\left(\tfrac{1}{\varepsilon^2}\right) \mathrm{N}_{\mathscr{T}}^{-1/2},$$

where the hidden constant is independent of ε and $\mathrm{N}_{\mathscr{T}}$. In addition, we present the following quasi-optimal estimate in the L^2-norm [19, Corollary 3.7]:

$$\|\mathsf{y} - \mathsf{y}_{\mathscr{T}}\|_{L^2(\Omega)} \lesssim \log^5\left(\tfrac{1}{\varepsilon^2}\right) \mathrm{N}_{\mathscr{T}}^{-1},$$

where again the hidden constant is independent of ε and $\mathrm{N}_{\mathscr{T}}$.

4. Discretization of the optimal control problem. The goal of this section is to propose and analyze a simple and efficient numerical strategy to solve the optimal control problem (1.2)–(1.3): a fully discrete scheme that discretizes the optimal control and state with standard piecewise bilinear finite elements on the anisotropic meshes described in Sections 3.1 and 3.2, depending on the corresponding singularly perturbed equation. To be concrete, we consider the following fully discrete scheme: Find

$$\min\ J(\mathsf{y}_{\mathscr{T}}, \mathsf{u}_{\mathscr{T}}) \tag{4.1}$$

subject to the discrete singularly perturbed equation

$$\mathsf{y}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}): \quad \mathcal{B}(\mathsf{y}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}}) = (\mathsf{f} + \mathsf{u}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}})_{L^2(\Omega)} \quad \forall\, \mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \tag{4.2}$$

and $\mathsf{u}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$. We recall that $\mathcal{B}$ is defined by (2.2).

We denote by $(\bar{\mathsf{y}}_{\mathscr{T}}, \bar{\mathsf{u}}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T})$ the optimal pair solving (4.1)–(4.2). We define the discrete control–to–state operator $\mathbf{S}_{\mathscr{T}} : \mathbb{V}(\mathscr{T}) \to \mathbb{V}(\mathscr{T})$, which given $\mathsf{u}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$, it associates a unique discrete state $\mathsf{y}_{\mathscr{T}}(\mathsf{u}_{\mathscr{T}}) = \mathbf{S}_{\mathscr{T}}\mathsf{u}_{\mathscr{T}}$ solving (4.2). With this operator at hand, we define the discrete and reduced cost functional by

$$f_{\mathscr{T}}(\mathsf{u}_{\mathscr{T}}) = J(\mathbf{S}_{\mathscr{T}}\mathsf{u}_{\mathscr{T}}, \mathsf{u}_{\mathscr{T}}) = \frac{1}{2}\|\mathbf{S}_{\mathscr{T}}\mathsf{u}_{\mathscr{T}} - \mathsf{y}_d\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|\mathsf{u}_{\mathscr{T}}\|^2_{L^2(\Omega)}. \tag{4.3}$$

We define the discrete adjoint state $\mathsf{p}_{\mathscr{T}} = \mathsf{p}_{\mathscr{T}}(\mathsf{u}_{\mathscr{T}})$ as the unique solution to

$$\mathsf{p}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}): \quad \mathcal{B}(\mathsf{q}_{\mathscr{T}}, \mathsf{p}_{\mathscr{T}}) = (\mathsf{y}_{\mathscr{T}} - \mathsf{y}_d, \mathsf{q}_{\mathscr{T}})_{L^2(\Omega)} \quad \forall \mathsf{q}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}). \tag{4.4}$$

We now state the existence and uniqueness of the discrete optimal control together with the necessary and sufficient optimality conditions for problem (4.1)–(4.2).

THEOREM 4.1 (existence, uniqueness and optimality system). *The fully discrete optimal control problem* (4.1)–(4.2) *has a unique optimal solution* $(\bar{\mathsf{y}}_{\mathscr{T}}, \bar{\mathsf{u}}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times$

$\mathbb{V}(\mathscr{T})$. *The optimality system*

$$\begin{cases} \bar{\mathsf{y}}_{\mathscr{T}}(\bar{\mathsf{u}}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \text{ solution to (4.2)}, \\ \bar{\mathsf{p}}_{\mathscr{T}}(\bar{\mathsf{u}}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \text{ solution to (4.4)}, \\ \lambda \bar{\mathsf{u}}_{\mathscr{T}} + \bar{\mathsf{p}}_{\mathscr{T}} = 0, \end{cases} \tag{4.5}$$

hold. These conditions are necessary and sufficient.

As in the continuous case described in Section 2, we eliminate the discrete control $\mathsf{u}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$ from (4.5) and rewrite the optimality conditions as follows:

$$\left\{ \begin{array}{rcll} \lambda \mathcal{B}(\bar{\mathsf{y}}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}}) + (\bar{\mathsf{p}}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}})_{L^2(\Omega)} & = & \lambda(\mathsf{f}, \mathsf{v}_{\mathscr{T}})_{L^2(\Omega)} & \forall\, \mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \\ \mathcal{B}(\mathsf{q}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}}) - (\bar{\mathsf{y}}_{\mathscr{T}}, \mathsf{q}_{\mathscr{T}})_{L^2(\Omega)} & = & -(\mathsf{y}_d, \mathsf{q}_{\mathscr{T}})_{L^2(\Omega)} & \forall\, \mathsf{q}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}). \end{array} \right. \tag{4.6}$$

We now construct suitable finite element spaces $\mathbb{V}(\mathscr{T})$ that will be essential to design an efficient solution technique to compute the solution to the discrete system (4.6). Since the discretization and analysis for the state equation (1.3) differs of the one for the reaction-diffusion equation (1.4), we analyze each case separately.

4.1. Reaction–diffusion equation. In this subsection we present a numerical technique to solve the optimal control problem composed by (1.2) and the state equation (1.4). In addition, we provide a complete quasi-optimal priori error analysis in both the energy and the L^2–norm. For convenience, we will consider $\mathsf{c} \equiv 1$ in (1.4).

4.1.1. Energy estimates. The a priori error analysis in the energy norm relies on a standard Galerkin approximation technique: the distance between $(\bar{\mathsf{y}}, \bar{\mathsf{p}})$ and $(\bar{\mathsf{y}}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}})$ is bounded by the best approximation error in the finite element space. To estimate such an approximation error we will invoke results in interpolation theory that rely on suitable pointwise estimates for the solution $(\bar{\mathsf{y}}, \bar{\mathsf{p}})$ to (2.7).

If a function ϕ is such that

$$\left| \frac{\partial^k \phi}{\partial x_i^k}(x_1, x_2) \right| \leq C \left(1 + \varepsilon^{-k} \left(e^{-x_i/\varepsilon} + e^{-(1-x_i)/\varepsilon} \right) \right), \tag{4.7}$$

for $i = 1, 2$ and $0 \leq k \leq 4$, we say that ϕ satisfies (4.7). Under the assumption that f and u satisfy (3.6), these pointwise estimates are derived for the solution y to the reaction–diffusion equation (1.4) in [31, Lemmas 2.1–2.5] and [33, Lemma 4.1]. We now proceed to derive the pointwise estimates (4.7) for the solution $(\bar{\mathsf{y}}, \bar{\mathsf{p}})$ to the coupled PDE system (2.7).

THEOREM 4.2 (pointwise estimates). *Let* $(\bar{\mathsf{y}}, \bar{\mathsf{p}})$ *be the unique solution to* (2.7). *If* f *and* y_d *satisfy* (3.6), *then* $\bar{\mathsf{y}}$ *and* $\bar{\mathsf{p}}$ *verify the pointwise estimates* (4.7).

Proof. We start by writing the second equation of (2.7) as follows:

$$\mathcal{B}(\mathsf{q}, \bar{\mathsf{p}}) = (\bar{\mathsf{y}} - \mathsf{y}_d, \mathsf{q})_{L^2(\Omega)} \qquad \forall \mathsf{q} \in \mathbb{Y}.$$

Standard elliptic regularity immediately yields $\bar{\mathsf{p}} \in H^2(\Omega) \cap H_0^1(\Omega)$, which in turn implies that $\bar{\mathsf{p}} \in C^\alpha(\bar{\Omega})$ for $\alpha \in (0, 1)$; see [23, Corollary 7.11]. We now write the first equation of (2.7) as follows:

$$\mathcal{B}(\bar{\mathsf{y}}, \mathsf{v}) = (\mathsf{f} - \bar{\mathsf{p}}/\lambda, \mathsf{v})_{L^2(\Omega)} \qquad \forall v \in \mathbb{Y}.$$

Since $\mathsf{f} - \bar{\mathsf{p}}/\lambda \in C^\alpha(\bar{\Omega})$ for $\alpha \in (0, 1)$, an application of Theorem 1.2 in [41, Part III] allows us to conclude that $\bar{\mathsf{y}} \in C^2(\Omega) \cap C(\bar{\Omega})$. Given the structure of (1.4), the previous

result can be improved by invoking the results of [23, 24, 45]: $\mathsf{y} \in C^{2,\alpha}(\Omega) \cap C^{1,\alpha}(\bar{\Omega})$; see also [41, Part III]. Now, recalling that $\bar{\mathsf{y}} - \mathsf{y}_d = 0$ on the vertices of $\bar{\Omega}$, we conclude, on the basis of [38, Theorem 2.1], that $\mathsf{p} \in C^4(\Omega) \cap C^2(\bar{\Omega})$; see also [24, 32, 33, 45]. Invoking the pointwise estimates derived for the solution to the singularly perturbed reaction-diffusion problem (1.4) in [31, Lemmas 2.1–2.5] and [33, Lemma 4.1], we deduce that $\bar{\mathsf{p}}$ satisfies (4.7). Finally, since $\bar{\mathsf{p}}$ and f satisfy (3.6), the same pointwise estimates hold for $\bar{\mathsf{y}}$. This concludes the proof. ∎

REMARK 4.3 (regularity of the optimal control). The arguments developed in the proof of Theorem 4.2, in conjunction with the relation $\lambda\bar{\mathsf{u}} + \bar{\mathsf{p}} = 0$, immediately yield the following regularity result for the optimal control: $\mathsf{u} \in C^4(\Omega) \cap C^2(\bar{\Omega})$.

The pointwise estimates (4.7) satisfied by $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ allow us to obtain the weighted a priori estimates derived in [17, Lemmas 4.1–4.3], which together with the anisotropic interpolation estimates developed in [17, Section 3] yield a quasi-optimal a priori error estimate in the energy norm. Before elaborating on this, let us comment about the family of meshes yielding quasi-optimality of our graded solution technique.

REMARK 4.4 (graded–mesh for the optimality system). The functions $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ solve a singularly perturbed reaction–diffusion equation. Consequently, they will typically exhibit the same singular behavior: sharp boundary layers of width $\mathcal{O}(\varepsilon|\log\varepsilon|)$ along the boundary $\partial\Omega$. Therefore, the graded mesh based on (3.1) is the appropriate refinement law that yields the quasi-optimality of our graded solution technique approximating the system (2.7); see Theorem 4.5.

THEOREM 4.5 (quasi-optimal a priori error estimate in energy norm). *Let* $(\bar{\mathsf{y}}, \bar{\mathsf{p}}) \in \mathbb{Y} \times \mathbb{Y}$ *and* $(\bar{\mathsf{y}}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T})$ *be the unique solutions to* (2.7) *and* (4.6)*, respectively. Then, we have the following a priori error estimate*

$$\lambda\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}}\|_\varepsilon + \|\bar{\mathsf{p}} - \bar{\mathsf{p}}_{\mathscr{T}}\|_\varepsilon \lesssim |\log \mathrm{N}_{\mathscr{T}}|\mathrm{N}_{\mathscr{T}}^{-\frac{1}{2}}, \tag{4.8}$$

where the hidden constant is independent of ε *and* $\mathrm{N}_{\mathscr{T}}$.

Proof. Define $e_{\bar{\mathsf{y}}} = \bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}}$ and $e_{\bar{\mathsf{p}}} = \bar{\mathsf{p}} - \bar{\mathsf{p}}_{\mathscr{T}}$. We write the systems (2.7) and (4.6) as single equations. Subtracting these derived equations and invoking the definition of $\mathcal{A}$, given by (2.8), we arrive at the following Galerkin orthogonality property:

$$\mathcal{A}\left((e_{\bar{y}}, e_{\bar{p}}), (\mathsf{v}_{\mathscr{T}}, \mathsf{q}_{\mathscr{T}})\right) = 0 \qquad \forall\, (\mathsf{v}_{\mathscr{T}}, \mathsf{q}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T}). \tag{4.9}$$

Invoking again the definition of the bilinear form $\mathcal{A}$ and exploiting (4.9), we obtain

$$\begin{aligned}
\lambda\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}}\|_\varepsilon^2 + \|\bar{\mathsf{p}} - \bar{\mathsf{p}}_{\mathscr{T}}\|_\varepsilon^2 &= \lambda\mathcal{B}(e_{\bar{\mathsf{y}}}, e_{\bar{\mathsf{y}}}) + \mathcal{B}(e_{\bar{\mathsf{p}}}, e_{\bar{\mathsf{p}}}) = \mathcal{A}\left((e_{\bar{\mathsf{y}}}, e_{\bar{\mathsf{y}}}), (e_{\bar{\mathsf{p}}}, e_{\bar{\mathsf{p}}})\right) \\
&= \mathcal{A}\left((e_{\bar{\mathsf{y}}}, \bar{\mathsf{y}} - \mathsf{v}_{\mathscr{T}}), (e_{\bar{\mathsf{p}}}, \bar{\mathsf{p}} - \mathsf{q}_{\mathscr{T}})\right) \\
&\lesssim \|e_{\bar{\mathsf{y}}}\|_\varepsilon\|\bar{\mathsf{y}} - \mathsf{v}_{\mathscr{T}}\|_\varepsilon + \|e_{\bar{\mathsf{p}}}\|_\varepsilon\|\bar{\mathsf{p}} - \mathsf{q}_{\mathscr{T}}\|_\varepsilon,
\end{aligned}$$

which, together with Young's inequality yields

$$\lambda\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}}\|_\varepsilon + \|\bar{\mathsf{p}} - \bar{\mathsf{p}}_{\mathscr{T}}\|_\varepsilon \lesssim \inf_{\mathsf{v}_{\mathscr{T}} \in \mathbb{V}_{\mathscr{T}}} \lambda\|\bar{\mathsf{y}} - \mathsf{v}_{\mathscr{T}}\|_\varepsilon + \inf_{\mathsf{q}_{\mathscr{T}} \in \mathbb{V}_{\mathscr{T}}} \|\bar{\mathsf{p}} - \mathsf{q}_{\mathscr{T}}\|_\varepsilon.$$

Let us estimate $\inf_{\mathsf{v}_{\mathscr{T}} \in V_{\mathscr{T}}} \lambda\|\bar{\mathsf{y}} - \mathsf{v}_{\mathscr{T}}\|_\varepsilon$, the estimate of $\inf_{\mathsf{q}_{\mathscr{T}} \in V_{\mathscr{T}}} \|\bar{\mathsf{p}} - \mathsf{q}_{\mathscr{T}}\|_\varepsilon$ being similar. We bound this best approximation error by using the $\mathcal{Q}_1$-quasi interpolation operator analyzed in [17, 36]: $\inf_{\mathsf{v}_{\mathscr{T}} \in V_{\mathscr{T}}} \|\bar{\mathsf{y}} - \mathsf{v}_{\mathscr{T}}\|_\varepsilon \leq \|\bar{\mathsf{y}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}\|_\varepsilon$. We now invoke the pointwise estimates for $\bar{\mathsf{y}}$ derived in Theorem 4.2 to obtain the weighted estimates of [17, Lemmas 4.1–4.3]. These estimates, in conjunction with the anisotropic estimates of [17, Theorem 4.4] on the family of graded meshes $\{\mathscr{T}\}$, defined in Section 3.1, yield

$$\|\bar{\mathsf{y}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}\|_\varepsilon \lesssim |\log \mathrm{N}_{\mathscr{T}}|\mathrm{N}_{\mathscr{T}}^{-1/2}.$$

The hidden constant is independent of ε and $\mathrm{N}_{\mathscr{T}}$. This concludes the proof. ∎

4.1.2. L^2 estimates. In this subsection, we follow [20] and derive a quasi–optimal a priori error analysis in the L^2–norm. We start with the following estimates.

LEMMA 4.6 (auxiliary estimates). *Let* $(\bar{\mathsf{y}}, \bar{\mathsf{p}})$ *be the solution to* (2.7) *with data* f *and* y_d *satisfying* (3.6). *If* γ, *defining the graded mesh* (3.5), *satisfies* $\gamma \in [\frac{3}{4}, 1]$, *then*

$$\left|\varepsilon^2 \int_\Omega \nabla(\bar{\mathsf{y}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}) \cdot \nabla \mathsf{v}_{\mathscr{T}}\right| + \left|\varepsilon^2 \int_\Omega \nabla(\bar{\mathsf{p}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}) \cdot \nabla \mathsf{v}_{\mathscr{T}}\right| \lesssim \log\left(\tfrac{1}{\varepsilon}\right)^{\frac{1}{2}} |\log \mathrm{N}_{\mathscr{T}}|^2 \mathrm{N}_{\mathscr{T}}^{-1} \|\mathsf{v}_{\mathscr{T}}\|_\varepsilon, \quad (4.10)$$

and

$$\left|\int_\Omega (\bar{\mathsf{y}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}})\mathsf{v}_{\mathscr{T}}\right| + \left|\int_\Omega (\bar{\mathsf{p}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}})\mathsf{v}_{\mathscr{T}}\right| \lesssim \log\left(\tfrac{1}{\varepsilon}\right)^{\frac{1}{2}} |\log \mathrm{N}_{\mathscr{T}}|^2 \mathrm{N}_{\mathscr{T}}^{-1} \|\mathsf{v}_{\mathscr{T}}\|_\varepsilon, \quad (4.11)$$

for all $\mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$. *In these estimates* $\Pi_{\mathscr{T}}$ *denotes the Lagrange interpolation operator, and the hidden constants are independent of* ε *and* $\mathrm{N}_{\mathscr{T}}$.

Proof. Since $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ satisfy the pointwise estimates (4.7), we apply [20, Lemma 3.1] to deduce some suitable weighted estimates for $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$. These estimates are the key ingredients in the derivation of the desired estimates (4.10) and (4.11), which follow from an application of [20, Lemmas 4.5 and 4.6]. ∎

With the help of estimates (4.10) and (4.11), we are able to derive a quasi-optimal a priori error analysis in the L^2–norm. To present it, we prove the following auxiliary error estimate that is instrumental in the analysis.

LEMMA 4.7 (auxiliary error estimate). *Let* $(\bar{\mathsf{y}}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}})$ *be the unique solution to* (4.6). *If* f *and* y_d *satisfy* (3.6), *then we have the following error estimate*

$$\lambda\|\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}\|_\varepsilon + \|\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}\|_\varepsilon \lesssim \log\left(\tfrac{1}{\varepsilon}\right)^{\frac{1}{2}} |\log \mathrm{N}_{\mathscr{T}}|^2 \mathrm{N}_{\mathscr{T}}^{-1}, \quad (4.12)$$

for all $\mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T})$. *In these estimates* $\Pi_{\mathscr{T}}$ *denotes the Lagrange interpolation operator, and the hidden constants do not depend on* ε *and* $\mathrm{N}_{\mathscr{T}}$.

Proof. We start by using the definition of the bilinear forms $\mathcal{A}$ and $\mathcal{B}$, given by (2.8) and (2.2), respectively, to derive

$$\begin{aligned}
&\lambda\|\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}\|_\varepsilon^2 + \|\bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}\|_\varepsilon^2 \\
&= \lambda\mathcal{B}(\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}, \bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}) + \mathcal{B}(\bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}, \bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}) \\
&= \mathcal{A}((\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}, \bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}), (\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}, \bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}})) \\
&= \mathcal{A}((\bar{\mathsf{y}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}, \bar{\mathsf{p}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}), (\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}, \bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}})).
\end{aligned}$$

where in the last inequality we have used the Galerkin orthogonality property (4.9): $\mathcal{A}((e_{\bar{\mathsf{y}}}, e_{\bar{\mathsf{p}}}), (\bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}, \bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}})) = 0$. Therefore, the estimate (4.12) follows by using (4.10) and (4.11) with $\mathsf{v}_{\mathscr{T}} = \bar{\mathsf{y}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}$ and $\mathsf{v}_{\mathscr{T}} = \bar{\mathsf{p}}_{\mathscr{T}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}$ accordingly. This concludes the proof. ∎

THEOREM 4.8 (quasi-optimal interpolation error estimate in the L^2–norm). *Let* $(\bar{\mathsf{y}}, \bar{\mathsf{p}})$ *be the solution to* (2.7) *and* $\Pi_{\mathscr{T}}$ *be the Lagrange interpolation operator. If the grading parameter* γ *in* (3.5) *satisfies* $\gamma \in [\frac{3}{4}, 1]$ *and* f *and* y_d *verify* (3.6), *then*

$$\|\bar{\mathsf{y}} - \Pi_{\mathscr{T}}\bar{\mathsf{y}}\|_{L^2(\Omega)} + \|\bar{\mathsf{p}} - \Pi_{\mathscr{T}}\bar{\mathsf{p}}\|_{L^2(\Omega)} \lesssim \log(\tfrac{1}{\varepsilon})^{\frac{1}{2}} |\log \mathrm{N}_{\mathscr{T}}|^2 \mathrm{N}_{\mathscr{T}}^{-1}, \quad (4.13)$$

where the hidden constant is independent of ε *and* $\mathrm{N}_{\mathscr{T}}$.

Proof. The key pointwise estimates derived in Theorem 4.2 yields the weighted estimates derived in [20, Lemma 3.1]. These estimates in conjunction with [20, Lemma 4.1], which only requires to have a function in $H^2((0,1)^2)$, are the main ingredients to derive the quasi-optimal interpolation error estimate (4.13) via an application of [20, Theorem 4.4.]. ∎

We conclude with the following quasi-optimal a priori error estimate in the L^2–norm for the optimal control variables $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$.

THEOREM 4.9 (quasi-optimal a priori error estimate in the L^2–norm). *Let* $(\bar{\mathsf{y}}, \bar{\mathsf{p}}) \in \mathbb{Y} \times \mathbb{Y}$ *and* $(\bar{\mathsf{y}}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T})$ *be the unique solutions to* (2.7) *and* (4.6)*, respectively. Then, we have the following error estimate*

$$\lambda \|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}}\|_{L^2(\Omega)} + \|\bar{\mathsf{p}} - \bar{\mathsf{p}}_{\mathscr{T}}\|_{L^2(\Omega)} \lesssim \log(\tfrac{1}{\varepsilon})^{\frac{1}{2}} |\log \mathrm{N}_{\mathscr{T}}|^2 \mathrm{N}_{\mathscr{T}}^{-1}, \tag{4.14}$$

where the hidden constant does not depend on ε *and* $\mathrm{N}_{\mathscr{T}}$.

Proof. The estimate (4.14) follows as a trivial application of the triangle inequality and Lemma 4.7 and Theorem 4.8. ∎

REMARK 4.10 (quasi optimal error estimate: advantages and disadvantages). Theorems 4.5 and 4.9 show quasi-optimal error estimates in terms of approximation. Since the optimal variables $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ have the same singular behavior on $\partial\Omega$, the graded mesh described in Section 3.1 for a single singularly perturbed equation (1.4) dictates the appropriate refinement technique to obtain the optimality of the proposed graded scheme (4.5). The main disadvantage of the latter is that is valid under a tensor product structure of the domain Ω and that relies on strong regularity assumptions on f and y_d; see condition (3.6).

4.2. Convection–reaction–diffusion equation. In this subsection we design an efficient solution technique to approximate the solution to the optimal control problem (1.2)–(1.3). To do this, we recall that in Section 3.2 we have assumed that $\mathsf{b} = (\mathsf{b}_1, \mathsf{b}_2)$ with $\mathsf{b}_i < -\delta$ for $i = 1, 2$ and $\delta > 0$.

To approximate the optimal control problem (1.2)–(1.3), we rewrite the optimality conditions as the system of PDE (2.7). To design an efficient technique to approximate such a system, it is essential to realize the structure of the adjoint equation:

$$-\Delta \mathsf{p} - \mathsf{b} \cdot \nabla \mathsf{p} + c\mathsf{p} = \mathsf{y} - \mathsf{y}_d,$$

i.e., it is a linear convection–reaction-diffusion equation with a convective term that is the negative of the one appearing in the state equation (1.3). As a consequence of the assumptions on the vector field b, the solution p to the adjoint equation exhibit a boundary layer of width $\mathcal{O}(\varepsilon^2 |\log \varepsilon|)$ at $\{(x, y) \in \Omega : x = 1 \text{ or } y = 1\}$. Inspired in [18], we now design a graded mesh to capture the singular behavior of the optimal variables $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ solving (2.7). Given a parameter $h \in (0, 1)$, we construct a partition $\{\xi_i\}_{i=0}^M$ of the interval $[0, \frac{1}{2}]$ with mesh–points:

$$\xi_0 = 0, \quad \xi_1 = h\varepsilon^2, \quad \xi_{i+1} = \xi_i + h\xi_i \text{ for } j = 1, \dots M-2, \quad \xi_M = 1/2, \tag{4.15}$$

where M is such that $\xi_{M-1} < \frac{1}{2}$ and $\xi_{M-1} + h\xi_{M-1} \geq \frac{1}{2}$. If $\frac{1}{2} - \xi_{M-1} < \xi_{M-1} - \xi_{M-2}$ we consider $\xi_{M-1} = (\frac{1}{2} + \xi_{M-2})/2$. We now invoke symmetry to define a partition on $[\frac{1}{2}, 1]$, thereby obtaining a partition $\{\xi_i\}_{i=0}^{2M}$ of $[0, 1]$. With this setting, we construct a graded mesh $\mathscr{T} = \{T_{ij}\}_{i,j=1}^{2M}$ where each rectangular element $T_{ij} = [\xi_{i-1}, \xi_i] \times [\xi_{j-1}, \xi_j]$. Figure 4.1 shows an example of this type of meshes.

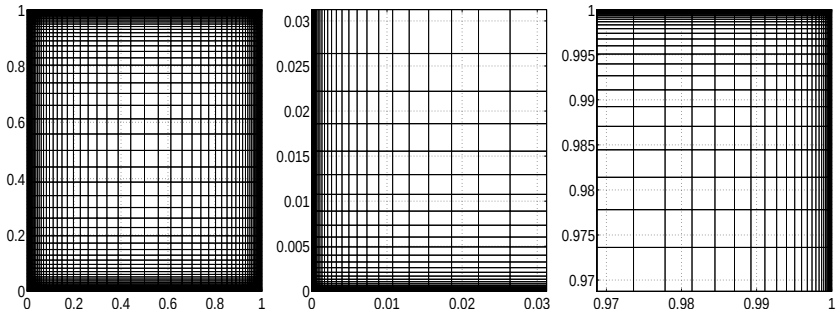

FIG. 4.1. *The left panel shows a graded mesh $\mathscr{T}$ constructed on the basis of* (3.9) *with* $\varepsilon = 10^{-4}$. *The number of degrees of freedom is* $N_{\mathscr{T}} = 18496$. *The center panel presents a zoom of the graded mesh near to the origin and the right panel shows a zoom near to* $(1,1)$.

4.2.1. Convergence. To derive an a priori error analysis in both the energy and the L^2–norm we could follow similar arguments to the ones developed in Section 4.1 and [18, 19]. In fact, following the analysis presented in [18] for the state equation (1.3), the first step is the derivation of suitable pointwise estimates for $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ (see [19, inequality (1.3)]). However, the derivation of these estimates requires strong regularity assumptions on $\bar{\mathsf{y}}$ and $\bar{\mathsf{u}}$ which, in principle, could be unrealistic. In what follows, we show convergence of our graded-scheme, without rates, by invoking Γ–convergence techniques [8, 15].

To avoid irrelevant technicalities let us here assume that $\mathsf{f} \equiv 0$. In this setting $\mathbf{S}$ is linear. Before starting with the convergence analysis, we remark that the family of discrete control–to–state operators $\{\mathbf{S}_{\mathscr{T}}\}$ is uniformly bounded in $\mathcal{B}(L^2(\Omega))$. In addition, since the error estimates are valid for right–hand sides that range over a dense subset of $L^2(\Omega)$, we obtain the pointwise convergence of these operators to $\mathbf{S}$. Similar conclusions hold for the family $\{\mathbf{S}^*_{\mathscr{T}}\}$.

We now show the convergence of the fully discrete scheme (4.6).

THEOREM 4.11 (convergence). *The sequence* $\{\bar{\mathsf{u}}_{\mathscr{T}}\}$ *is uniformly bounded. In addition,* $\{\bar{\mathsf{u}}_{\mathscr{T}}\}$ *contains a subsequence that converges strongly to* $\bar{\mathsf{u}}$ *in* $L^2(\Omega)$.

Proof. Since $\bar{\mathsf{u}}_{\mathscr{T}}$ minimizes the reduced cost functional $f_{\mathscr{T}}$, defined by (4.3), we deduce that $\{\bar{\mathsf{u}}_{\mathscr{T}}\}$ is uniformly bounded. Now, if $\mathsf{u}_0 \in L^2(\Omega)$ and $\Pi_{\mathscr{T}}$ denotes the quasi-interpolation operator of [17, 36], the uniform boundedness of both families $\{\Pi_{\mathscr{T}}\}$ and $\{\mathbf{S}_{\mathscr{T}}\}$ yield $f_{\mathscr{T}}(\bar{\mathsf{u}}_{\mathscr{T}}) \leq f_{\mathscr{T}}(\Pi_{\mathscr{T}}\mathsf{u}_0) \lesssim \|\mathsf{u}_0\|_{L^2(\Omega)} + \|\mathsf{y}_d\|_{L^2(\Omega)}$. This implies the existence of a weakly convergent subsequence. We proceed to invoke the theory of Γ–convergence and prove that this subsequence converges to $\bar{\mathsf{u}}$. To acomplish this task, we follow [15] and verify some suitable assumptions.

(a.1) Let us assume that $\mathsf{u}_{\mathscr{T}} \rightharpoonup \mathsf{u}$ in $L^2(\Omega)$ and observe that

$$(\mathbf{S}_{\mathscr{T}}\mathsf{u}_{\mathscr{T}} - \mathbf{S}\mathsf{u}, \mathsf{v})_{L^2(\Omega)} = (\mathbf{S}_{\mathscr{T}}\mathsf{u} - \mathbf{S}\mathsf{u}, \mathsf{v})_{L^2(\Omega)} + (\mathbf{S}_{\mathscr{T}}(\mathsf{u}_{\mathscr{T}} - \mathsf{u}), \mathsf{v})_{L^2(\Omega)}$$

for $\mathsf{v} \in L^2(\Omega)$. The pointwise convergence of $\{\mathbf{S}_{\mathscr{T}}\}$ to $\mathbf{S}$ implies that $(\mathbf{S}_{\mathscr{T}}\mathsf{u} - \mathbf{S}\mathsf{u}, \mathsf{v})_{L^2(\Omega)} \to 0$, while the pointwise convergence of the sequence $\{\mathbf{S}^*_{\mathscr{T}}\}$ shows that $(\mathbf{S}_{\mathscr{T}}(\mathsf{u}_{\mathscr{T}} - \mathsf{u}), \mathsf{v})_{L^2(\Omega)} = (\mathsf{u}_{\mathscr{T}} - \mathsf{u}, \mathbf{S}^*_{\mathscr{T}}\mathsf{v})_{L^2(\Omega)} \to 0$. Consequently $\mathbf{S}_{\mathscr{T}}\mathsf{u}_{\mathscr{T}} \rightharpoonup \mathbf{S}\mathsf{u}$. Invoking the lower semicontinuity property of the norms, we derive that

$$f(\mathsf{u}) \leq \liminf f_{\mathscr{T}}(\mathsf{u}_{\mathscr{T}}),$$

where f denotes the reduced cost functional defined in (2.4).

(a.2) Let $\mathsf{u} \in L^2(\Omega)$, then $\Pi_{\mathscr{T}}\mathsf{u} \to \mathsf{u}$ in $L^2(\Omega)$. This implies that $\mathbf{S}_{\mathscr{T}}\Pi_{\mathscr{T}}\mathsf{u} \to \mathbf{S}\mathsf{u}$ in $L^2(\Omega)$ as well. Then, the continuity of $f_{\mathscr{T}}$ yields

$$f(\mathsf{u}) \geq \limsup f_{\mathscr{T}}(\Pi_{\mathscr{T}}\mathsf{u}).$$

(a.3) Given that $f_{\mathscr{T}}(\mathsf{u}_{\mathscr{T}}) \geq \frac{\lambda}{2}\|\mathsf{u}_{\mathscr{T}}\|^2_{L^2(\Omega)}$, [15, Proposition 7.7] implies that the family $\{f_{\mathscr{T}}\}$ is equicoercive.

Properties (a.1) and (a.2) imply the Γ–convergence of $f_{\mathscr{T}}$ to f, which implies that minimizers of $f_{\mathscr{T}}$, if they converge, must do so to a minimizer of f [15, Corollary 7.20]. Property (a.3) and the uniqueness of the minimizer of the reduced cost f are the conditions for the Lemma of Γ–convergence [15, Corollary 7.24]: $\{\bar{\mathsf{u}}_{\mathscr{T}}\}$ converges weakly to $\bar{\mathsf{u}}$. The strong convergence follow similar arguments to the ones elaborated in the proof of [8, Theorem 37]. ∎

5. Numerical examples. Here we explore computationally the performance of the proposed solution scheme to solve the optimal control (1.2)–(1.3). With the graded schemes introduced in Section 4, we solve (4.6) and explore several test cases. In addition, we compare our solution technique with a combined method that involves adaptivity and stabilized schemes. Computationally, we observe a competitive performance of our method in terms of accuracy. We also observe optimal experimental rates of convergence in both the energy and the $L^2(\Omega)$–norm. The pollution effect discussed in [25] is not observed. This is due to the fact that the boundary layers are appropriately approximated. Let us finally remark that the computational implementation of the proposed scheme is simpler than implementing adaptive stabilized procedures, but it relies on the a priori knowledge of the boundary layers.

We recall the finite element discretization of the coupled system (2.7) introduced in Section 4: Find $(\bar{\mathsf{y}}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}}) \in \mathbb{V}(\mathscr{T}) \times \mathbb{V}(\mathscr{T})$ such that

$$\begin{cases} \lambda\mathcal{B}(\bar{\mathsf{y}}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}}) + (\bar{\mathsf{p}}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}})_{L^2(\Omega)} &=& \lambda(\mathsf{f}, \mathsf{v}_{\mathscr{T}})_{L^2(\Omega)} & \forall\, \mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \\ \mathcal{B}(\mathsf{q}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}}) - (\bar{\mathsf{y}}_{\mathscr{T}}, \mathsf{q})_{L^2(\Omega)} &=& -(\mathsf{y}_d, \mathsf{q}_{\mathscr{T}})_{L^2(\Omega)} & \forall\, \mathsf{q}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \end{cases}$$

where the data y_d and f satisfy (3.6) in the case that $\mathsf{b} \equiv 0$. We design the graded meshes on the basis of (3.5) and (4.15) depending on the considered state equation. This solution technique is implemented with the help of a code that we implemented using `C++`. In all our numerical examples the domain is $\Omega = (0,1)^2$, and the regularized parameter is $\lambda = 1$. The stiffness matrix of this discrete system is assembled exactly. The forcing terms and the approximation errors are computed by a quadrature formula which is exact for polynomials of degree 10. The resulting linear system is solved by using the multifrontal massively parallel sparse direct solver (MUMPS) [5, 6].

We now proceed to explore and examine several numerical examples.

5.1. Reaction–diffusion equation: double boundary layer test. The purpose of this numerical example is to show how a standard $\mathcal{Q}_1$–finite element technique based on the graded meshes (3.5) solves efficiently the coupled system (2.7) when (1.4) is considered as state equation. We set $\gamma = 0.75$ in (3.5) and $\mathsf{c} = 1$ in (1.4). We design the numerical experiment such that the optimal variables $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ exhibit boundary layers on $\partial\Omega$. To do that, we consider the data y_d and f such that:

$$\bar{\mathsf{y}} = x_2(1-x_2)\left(1 - x_1 - \frac{e^{-\frac{x_1}{\varepsilon^2}} - e^{-\frac{1}{\varepsilon^2}}}{1 - e^{-\frac{1}{\varepsilon^2}}}\right), \quad \bar{\mathsf{p}} = x_1(1-x_1)\left(1 - x_2 - \frac{e^{-\frac{x_2}{\varepsilon^2}} - e^{-\frac{1}{\varepsilon^2}}}{1 - e^{-\frac{1}{\varepsilon^2}}}\right).$$

The optimal state variable $\bar{\mathsf{y}}$ presents a singular behavior at $\{(x_1, x_2) \in \partial\Omega : \ x_1 = 0\}$. Meanwhile, the optimal adjoint state $\bar{\mathsf{p}}$ exhibits a boundary layer at $\{(x_1, x_2) \in$

$\partial\Omega : \ x_2 = 0\}$. The asymptotic relations $\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}}\|_{\varepsilon} \approx \mathrm{N}_{\mathscr{T}}^{-1/2}$ and $\|\bar{\mathsf{p}} - \bar{\mathsf{p}}_{\mathscr{T}}\|_{\varepsilon} \approx \mathrm{N}_{\mathscr{T}}^{-1/2}$ are shown in Figure 5.1. This illustrates the quasi-optimal decay rate, in the energy norm, of our graded solution technique for all choices of the parameter ε considered. We also report the asymptotic relation $\|\bar{\mathsf{y}} - \bar{\mathsf{y}}_{\mathscr{T}}\|_{L^2(\Omega)} \approx \mathrm{N}_{\mathscr{T}}^{-1}$ and $\|\bar{\mathsf{p}} - \bar{\mathsf{p}}_{\mathscr{T}}\|_{L^2(\Omega)} \approx \mathrm{N}_{\mathscr{T}}^{-1}$, which are optimal in terms of approximation. We conclude computational robustness of our technique with respect to the parameter ε.

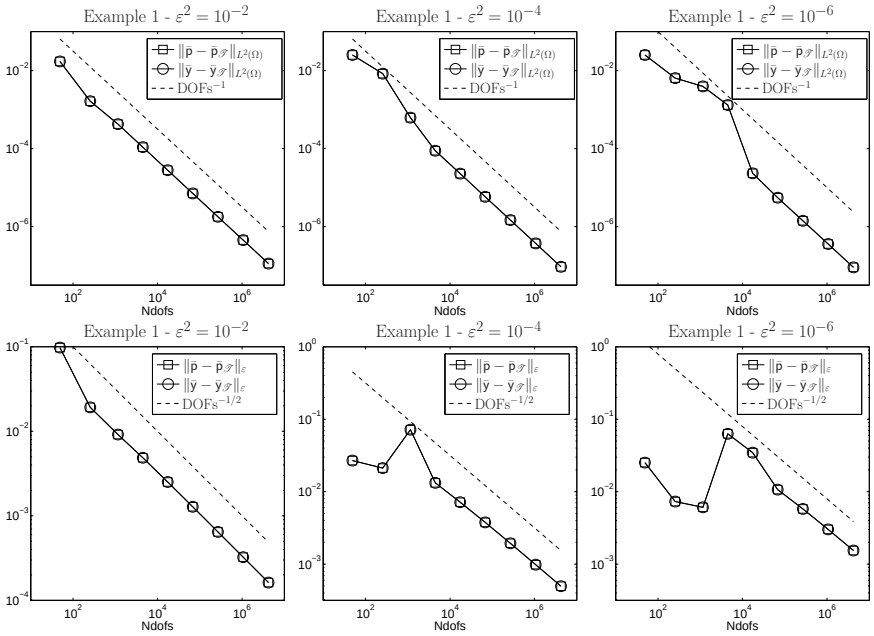

FIG. 5.1. *Example 1: Computational rates of convergence for our graded solution technique in the L^2–norm (top) and in the energy–norm (bottom) with respect to the number of degrees of freedom (DOFs). The values of the parameter ε^2 considered are: 10^{-2}, 10^{-4} and 10^{-6}. The graded meshes are based on* (3.5) *with $\gamma = 0.75$. In all cases we recover optimal rates of convergence.*

In Figure 5.2, we present the numerical approximation of both the optimal state $\bar{\mathsf{y}}$ and the optimal adjoint state $\bar{\mathsf{p}}$. To obtain them, we have considered a standard $\mathcal{Q}_1$ finite element method on the graded meshes based on (3.5) with $\mathrm{N}_{\mathscr{T}} = 192721$ and $\varepsilon^2 = 10^{-6}$.

5.2. Convection–reaction–diffusion equation: double boundary layer test. In this second numerical example we explore the use of the graded meshes based on (4.15) to approximate the solution to the optimal control problem (1.2)–(1.3). We consider a convection–reaction–diffusion equation as a state equation: we set $\mathsf{b} = (-1, -1)$ and $\mathsf{c} = 1$ in (1.3). In addition, we consider a suitable data y_d and f such that the optimal variables $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ both have boundary layers. In fact,

$$\bar{\mathsf{y}} = \left(1 - x_1 - \frac{e^{-\frac{x_1}{\varepsilon^2}} - e^{-\frac{1}{\varepsilon^2}}}{1 - e^{-\frac{1}{\varepsilon^2}}}\right)\left(1 - x_2 - \frac{e^{-\frac{x_2}{\varepsilon^2}} - e^{-\frac{1}{\varepsilon^2}}}{1 - e^{-\frac{1}{\varepsilon^2}}}\right),$$

$$\bar{\mathsf{p}} = \left(x_1 - \frac{e^{-\frac{(1-x_1)}{\varepsilon^2}} - e^{-\frac{1}{\varepsilon^2}}}{1 - e^{-\frac{1}{\varepsilon^2}}}\right)\left(x_2 - \frac{e^{-\frac{(1-x_2)}{\varepsilon^2}} - e^{-\frac{1}{\varepsilon^2}}}{1 - e^{-\frac{1}{\varepsilon^2}}}\right).$$

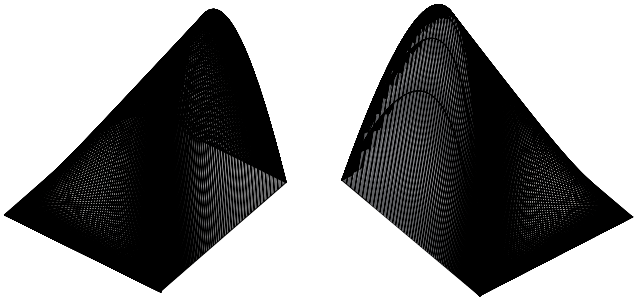

FIG. 5.2. *Example 1: We present the discrete optimal adjoint state* $\bar{\mathsf{p}}_{\mathscr{T}}$ *(left) and the discrete optimal state* $\bar{\mathsf{y}}_{\mathscr{T}}$ *(right) approximating the optimal variables that solve the optimal control problem* (1.2)-(1.3). *We consider* $\mathsf{b} = 0$, $\mathsf{c} = 1$ *and* $\varepsilon^2 = 10^{-6}$. *The number of degrees of freedom is* $\mathrm{N}_{\mathscr{T}} = 192721$.

The optimal state $\bar{\mathsf{y}}$ presents a boundary layer at $\{(x_1, x_2) \in \partial\Omega : \ x_1 = 0 \text{ and } x_2 = 0\}$. Meanwhile the optimal adjoint state $\bar{\mathsf{p}}$ at $\{(x_1, x_2) \in \partial\Omega : \ x_1 = 1 \text{ and } x_2 = 1\}$. We remark that that this situation represents a more complicated scenario to the one considered in Section 5.1: the region where one optimal variable is smooth corresponds to a sector where the other optimal variable exhibit a singular behavior. This leads to the challenge of capturing efficiently and robustly both solutions. Our graded solution technique does not present the pollution effect discussed in [25], which is due to the fact that the boundary layers are appropriately approximated.

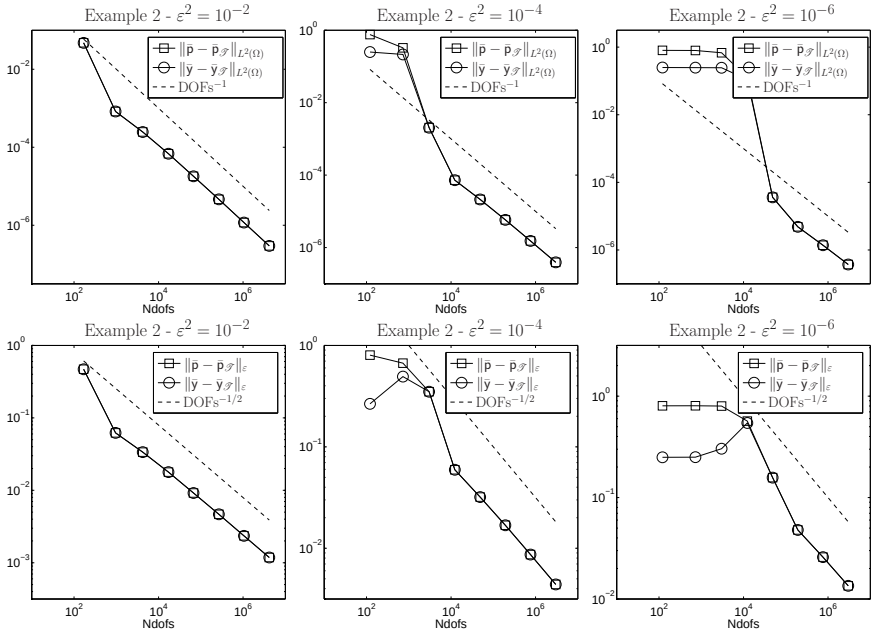

FIG. 5.3. *Example 2: Computational rates of convergence for our graded solution technique in the* L^2*–norm (top) and in the energy–norm (bottom) with respect to the number of degrees of freedom (DOFs). We consider* $\mathsf{b} = (-1, -1)$ *and* $\mathsf{c} = 1$ *in* (1.3). *The values of the parameter* ε^2 *considered are:* 10^{-2}, 10^{-4} *and* 10^{-6}. *The graded meshes are based on* (4.15). *In all cases we recover optimal rates of convergence.*

In Figure 5.3, we present the computational rates of convergence for the error approximating the optimal variables in the energy and in the L^2–norm. We consider different values of the parameter ε^2: 10^{-2}, 10^{-4} and 10^{-6}. For both norms, our method delivers optimal rates of convergence. We also conclude robustness of our graded technique with respect to the parameter ε^2.

In Figure 5.4, we show the numerical approximation of both optimal variables $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$ obtained by using our graded technique with $N_{\mathscr{T}} = 192721$ and $\varepsilon^2 = 10^{-6}$.

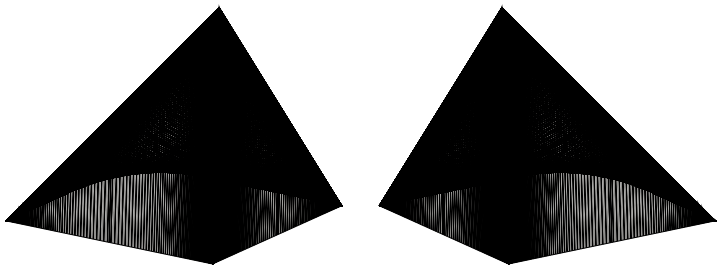

FIG. 5.4. *Example 2: We show the discrete optimal adjoint state $\bar{\mathsf{p}}_{\mathscr{T}}$ (left), and the discrete optimal state $\bar{\mathsf{y}}_{\mathscr{T}}$ (right) approximating the optimal variables solving the optimal control problem* (1.2)-(1.3) *with* $\mathsf{b} = (-1,-1)$, $\mathsf{c} = 1$ *and* $\varepsilon^2 = 10^{-6}$. *The number of degrees of freedom is* $\mathrm{N}_{\mathscr{T}} = 44100$.

5.3. Convection–reaction–diffusion equation: single interior layer test. Here, we explore the performance of our method by solving the optimal control problem (1.2)–(1.3) with $\mathsf{b} = (1,0)$ and $\mathsf{c} = 1$ in (1.3) . We remark that under this setting our theoretical results can not be applied: the optimal adjoint state $\bar{\mathsf{p}}$ exhibits an interior layer. We consider a smooth solution as optimal state variable $\bar{\mathsf{y}}$, and study computationally if any pollution effect occurs. To do this, we set

$$\bar{\mathsf{y}} = x_1x_2(1-x_1)(1-x_2), \quad \bar{\mathsf{p}} = x_1x_2(1-x_1)(1-x_2)\mathrm{tg}^{-1}\left(\frac{x_1-\frac{1}{2}}{\varepsilon^2}\right).$$

as the exact solution to (2.7). The adjoint variable $\bar{\mathsf{p}}$ present an interior layer on the line $x_1 = 0.5$.

The asymptotic relations $\|\bar{\mathsf{y}}-\bar{\mathsf{y}}_{\mathscr{T}}\|_{\varepsilon} \approx \mathrm{N}_{\mathscr{T}}^{-1/2}$, $\|\bar{\mathsf{p}}-\bar{\mathsf{p}}_{\mathscr{T}}\|_{\varepsilon} \approx \mathrm{N}_{\mathscr{T}}^{-1/2}$, $\|\bar{\mathsf{y}}-\bar{\mathsf{y}}_{\mathscr{T}}\|_{L^2(\Omega)} \approx \mathrm{N}_{\mathscr{T}}^{-1}$ and $\|\bar{\mathsf{p}}-\bar{\mathsf{p}}_{\mathscr{T}}\|_{L^2(\Omega)} \approx \mathrm{N}_{\mathscr{T}}^{-1}$ are shown in Figure 5.5. We conclude that in this case, which is not is not covered by our theoretical results, we again recover optimality of our method. This also exhibits robustness with respect to the perturbation parameter ε. We remark that the pollution effect discussed in [25] is not observed. In Figure 5.6, we present the numerical approximations $\bar{\mathsf{y}}_{\mathscr{T}}$ and $\bar{\mathsf{p}}_{\mathscr{T}}$ of both optimal variables $\bar{\mathsf{y}}$ and $\bar{\mathsf{p}}$, respectively. They were obtained using our graded solution technique based on the graded mesh (4.15) with $\varepsilon^2 = 10^{-6}$. The number of degrees of freedom is $\mathrm{N}_{\mathscr{T}} = 192721$.

5.4. Graded scheme and an adaptive stabilized scheme. The purpose of this section is to compare our proposed numerical technique with an adaptive stabilized scheme. Stabilized finite element methods have being proposed to improve the stability properties of the pure Galerkin method and to reduce its oscillatory behavior at solving a convection–reaction–diffusion equation [27, 39, 41]. These techniques have

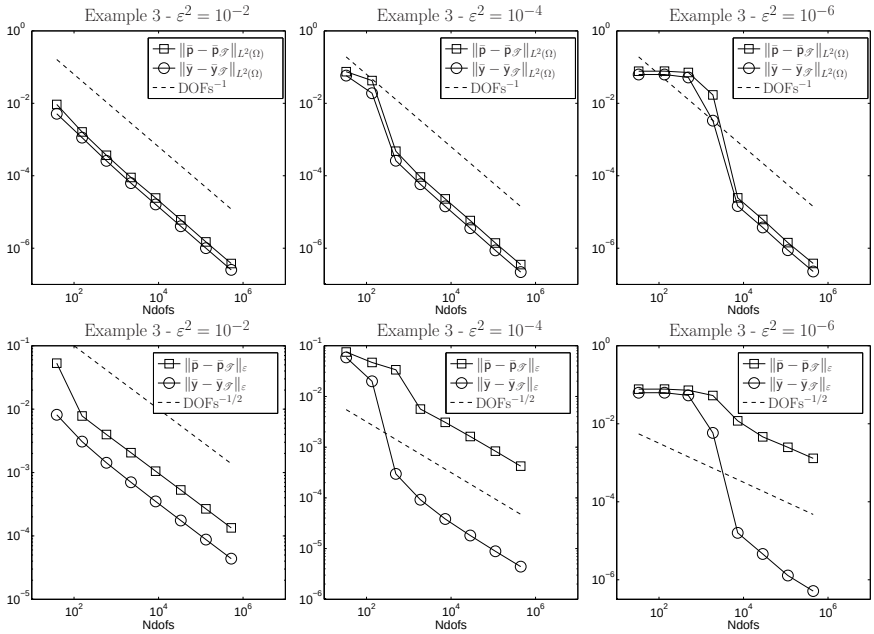

FIG. 5.5. *Example 3: Computational rates of convergence for our graded solution technique in the L^2–norm (top) and in the energy–norm (bottom) with respect to the number of degrees of freedom (DOFs). We consider $\mathsf{b} = (1, 0)$ and $c = 1$ in (1.3). The values of the parameter ε^2 considered are: 10^{-2}, 10^{-4} and 10^{-6}. The graded meshes are based on (4.15). In all cases we recover optimal rates of convergence.*

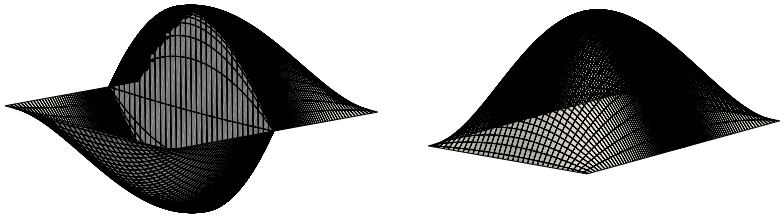

FIG. 5.6. *Example 3: We present the discrete optimal adjoint state $\bar{\mathsf{p}}_{\mathscr{T}}$ (left), and the discrete optimal state $\bar{\mathsf{y}}_{\mathscr{T}}$ (right) approximating the optimal variables solving (2.7) with $\mathsf{b} = (1, 0)$, $\mathsf{c} = 1$ and $\varepsilon^2 = 10^{-6}$. The number of degrees of freedom is $\mathrm{N}_{\mathscr{T}} = 192721$.*

been extended to solve the optimal control problem (1.2)–(1.3) [10, 25]. We review briefly an scheme based on SUPG, which leads to the following system of equations

$$\left\{ \begin{array}{rcll} \lambda \mathcal{B}_S(\bar{\mathsf{y}}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}}) + (\bar{\mathsf{p}}_{\mathscr{T}}, \mathsf{v}_{\mathscr{T}})_S & = & \lambda(\mathsf{f}, \mathsf{v}_{\mathscr{T}})_S & \forall\, \mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \\ \mathcal{B}_S(\mathsf{q}_{\mathscr{T}}, \bar{\mathsf{p}}_{\mathscr{T}}) - (\bar{\mathsf{y}}_{\mathscr{T}}, \mathsf{q}_{\mathscr{T}})_S & = & -(\mathsf{y}_d, \mathsf{q}_{\mathscr{T}})_S & \forall\, \mathsf{q}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \end{array} \right.$$

where each stabilized term $\mathcal{B}_S$ and $(\cdot, \cdot,)_S$ add, to the usual formulation (2.7), element residuals; see [25, equations (2.13a)–(2-13b)]. The finite element space is given by

$$\mathbb{V}(\mathscr{T}) = \left\{ W \in \mathcal{C}^0\left(\overline{\Omega}\right) :\ W|_T \in \mathbb{P}_1(T)\ \forall\ T \in \mathscr{T},\ W|_{\partial\Omega} = 0 \right\},$$

where $\mathbb{P}_1$ is the set of polynomials of degree one and $\{\mathscr{T}\}$ is a shape regular family of triangulation of Ω. As it is stated in [25], the properties of the stabilized SUPG method when is applied to the optimal control problem (1.2)–(1.3) are substantially different from the ones observed at solving the single equation (1.3). In [25], the authors observe optimal error estimates in the H^1–norm; the discrete space $\mathbb{V}(\mathscr{T})$ is based on piecewise linear functions. However, the observed estimates in the L^2–norm are not optimal: $\mathcal{O}(h)$. This is due to the fact that the boundary layers are not properly resolved, and then they pollute the numerical solution in the entire domain. On the contrary, our proposed graded method delivers optimal computational rates of convergence in both norms and no pollution effects occur.

We explore a SUPG technique combined with adaptivity. This has as a key ingredient an a posteriori error estimator and leads to selectively refinements where the solution is poorly resolved; see [3, 37, 44]. Recently, and in the context of optimal control problems, reference [28] provides a general framework to obtain a posteriori error estimators. Following Section 4 in [28] and Section 4 in [43], the following a posteriori error bound can be derived

$$\lambda\|\mathsf{y}-\bar{\mathsf{y}}_{\mathscr{T}}\|_{\varepsilon}+\|\mathsf{p}-\bar{\mathsf{p}}_{\mathscr{T}}\|_{\varepsilon}\leq C\left(\sum_{T\in\mathscr{T}}\eta_T^2\right)^{1/2}\quad\text{with}\quad\eta_T^2:=\eta_{T,\bar{\mathsf{y}}}^2+\eta_{T,\bar{\mathsf{p}}}^2,$$

where

$$\begin{aligned}\eta_{T,\bar{\mathsf{y}}}^2=\min\left\{\frac{h_T}{\varepsilon},1\right\}^2&\Big(\left\|\Pi_T(f)+\lambda\left(\varepsilon^2\Delta\bar{\mathsf{y}}_{\mathscr{T}}-\Pi_T(\mathsf{b}\cdot\nabla\bar{\mathsf{y}}_{\mathscr{T}})-\mathsf{c}\bar{\mathsf{y}}_{\mathscr{T}}\right)-\bar{\mathsf{p}}_{\mathscr{T}}\right\|_{L^2(T)}^2\\&+\|\mathsf{f}-\Pi_T(\mathsf{f})-\lambda(\mathsf{b}\cdot\nabla\bar{\mathsf{y}}_{\mathscr{T}}-\Pi_T(\mathsf{b}\cdot\nabla\bar{\mathsf{y}}_{\mathscr{T}}))\|_{L^2(T)}^2\Big)\\&+\sum_{E\in\partial T}\min\left\{\frac{h_E}{\varepsilon^2},\frac{1}{\varepsilon}\right\}\left\|\lambda\varepsilon^2[\![\partial_n\bar{\mathsf{y}}_{\mathscr{T}}]\!]\right\|_{L^2(E)}^2,\end{aligned}$$

and

$$\begin{aligned}\eta_{T,\bar{\mathsf{p}}}^2=\min\left\{\frac{h_T}{\varepsilon^2},1\right\}^2&\Big(\left\|-\Pi_T(\mathsf{y}_d)+\varepsilon^2\Delta\bar{\mathsf{p}}_{\mathscr{T}}+\Pi_T(\mathsf{b}\cdot\nabla\bar{\mathsf{p}}_{\mathscr{T}})-\mathsf{c}\bar{\mathsf{p}}_{\mathscr{T}}+\bar{\mathsf{y}}_{\mathscr{T}}\right\|_{L^2(T)}^2\\&+\|\mathsf{y}_d-\Pi_T(\mathsf{y}_d)+\mathsf{b}\cdot\nabla\bar{\mathsf{p}}_{\mathscr{T}}-\Pi_T(\mathsf{b}\cdot\nabla\bar{\mathsf{p}}_{\mathscr{T}}))\|_{L^2(T)}^2\Big)\\&+\sum_{E\in\partial T}\min\left\{\frac{h_E}{\varepsilon^2},\frac{1}{\varepsilon}\right\}\left\|\varepsilon^2[\![\partial_n\bar{P}_{\mathscr{T}}]\!]\right\|_{L^2(E)}^2,\end{aligned}$$

where Π_T denotes the L^2–orthogonal projection over $\mathbb{P}_1(T)$. We construct an adaptive procedure, where refinement take place using the so–call *maximum strategy*: the element $\hat{T}$ is marked for refinement if

$$\eta_{\hat{T}}\geq\frac{1}{2}\max_{T\in\mathscr{T}}\{\eta_T\}.$$

The refinement is made by using bisection [37]. Local efficiency can be established by using standard bubble function arguments [3, 44] on the error equation. We remark that other techniques can be applied to obtain different error estimators, which deliver fully computable bounds and in some cases robustness [2, 4, 13, 22].

In Figure 5.7, we present the computational rates of convergence obtained by using the proposed adaptive stabilized scheme to solve examples 2 and 3, for both:

the energy and the L^2–norm. To be precise, we consider (1.3) as state equation with $\mathsf{b} = (-1,-1)$, $\mathsf{c} = 1$ and $\varepsilon = 10^{-4}$, and $\mathsf{b} = (1,0)$, $\mathsf{c} = 1$ and $\varepsilon^2 = 10^{-4}$, respectively. We observe that the proposed adaptive stabilized technique recover optimal rates of convergence, but at the expense of adding residual terms into the formulation, calculating error indicators and conducting a marking/refinement procedure. In Figure 5.7, we also report the computational rates of convergence for our graded scheme, which shows a better performance than the one using an adaptive stabilized method, both in terms of accuracy and computational efficiency.

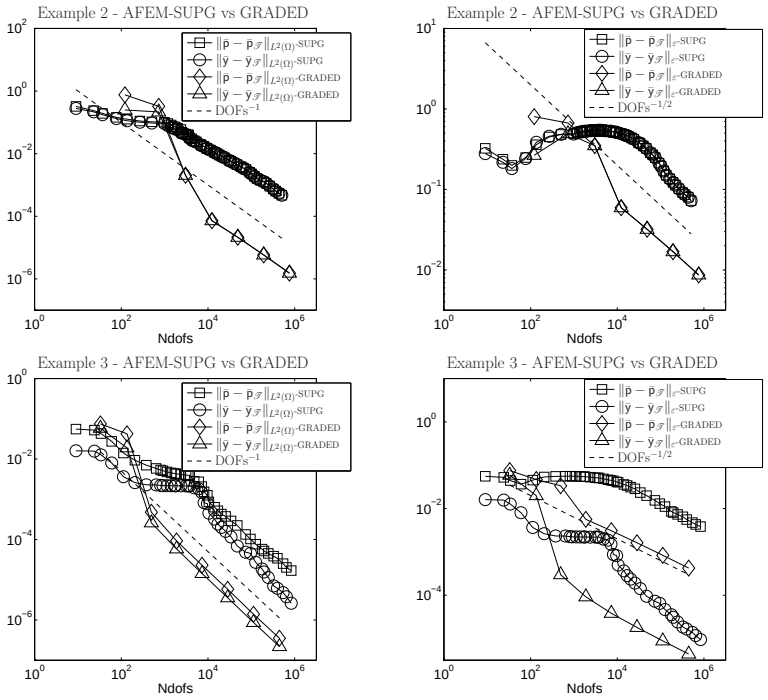

FIG. 5.7. *Example 4: Computational rates of convergence in the energy (right) and the L^2–norm (left) for our graded scheme and the proposed adaptive SUPG stabilized scheme, with respect to the number of degrees of freedom (DOFs). We consider two examples. First, we set* $\mathsf{b} = (-1,-1)$, $\mathsf{c} = 1$ *and* $\varepsilon = 10^{-4}$. *Second, we set* $\mathsf{b} = (1,0)$, $\mathsf{c} = 1$ *and* $\varepsilon^2 = 10^{-4}$. *In all cases we recover optimal rates of convergence. A better performance of our graded scheme is observed.*

5.5. A control constrained problem. To explore the versatility of the proposed graded scheme, in this section we consider a control–constrained optimal control problem that minimizes the functional (1.1) subject to the convection–reaction-diffusion equation (1.3) and the following control constraints:

$$\mathsf{u} \in \mathbb{U}_{\rm ad} := \left\{ \mathsf{v} \in L^2(\Omega) : \ \mathsf{u}_a \leq \mathsf{v} \leq \mathsf{u}_b \text{ a.e. in } \Omega \right\},$$

with $\mathsf{u}_a < \mathsf{u}_b$ and $\mathsf{u}_a, \mathsf{u}_b \in \mathbb{R}$. The existence and uniqueness follows from [42, Theorem 2.14]. In addition, the necessary and sufficient first–order optimality condition is equivalent to the projection formula

$$\bar{\mathsf{u}} = \mathbb{P}_{[\mathsf{u}_a,\mathsf{u}_b]}\left\{ -\frac{1}{\lambda}\bar{\mathsf{p}} \right\} = \max\left\{ \mathsf{u}_a, \min\left\{ \mathsf{u}_b, -\frac{1}{\lambda}\bar{\mathsf{p}} \right\} \right\} \quad \text{a.e. in } \Omega;$$

[42, Theorem 2.28]. This projection formula tells us that, in general, we can only expect $\bar{\mathsf{u}} \in W^{1,\infty}(\Omega)$. Consequently our a priori error analysis can not be developed due to the restricted regularity properties of the optimal control $\bar{\mathsf{u}}$.

The associated optimal control problem can be written as follows:

$$\begin{aligned} -\varepsilon^2\Delta\bar{\mathsf{y}} + \mathsf{b}\cdot\nabla\bar{\mathsf{y}} + \mathsf{c}\bar{\mathsf{y}} &= \mathsf{f} + \bar{\mathsf{u}} \quad \text{in } \Omega, \qquad \bar{\mathsf{y}} = 0 \quad \text{on } \partial\Omega \\ -\varepsilon^2\Delta\bar{\mathsf{p}} - \mathsf{b}\cdot\nabla\bar{\mathsf{p}} + \mathsf{c}\bar{\mathsf{p}} &= \bar{\mathsf{y}} - \mathsf{y}_d \quad \text{in } \Omega, \qquad \bar{\mathsf{p}} = 0 \quad \text{on } \partial\Omega \end{aligned}$$

and

$$\bar{\mathsf{u}} = \begin{cases} \mathsf{u}_a & \text{if } \lambda\bar{\mathsf{u}}(x) + \bar{\mathsf{p}}(x) < u_a, \\ -\frac{\bar{\mathsf{p}}}{\lambda} & \text{if } \lambda\bar{\mathsf{u}}(x) + \bar{\mathsf{p}}(x) \in [u_a, u_b], \\ \mathsf{u}_b & \text{if } \lambda\bar{\mathsf{u}}(x) + \bar{\mathsf{p}}(x) > u_b. \end{cases} \tag{5.1}$$

To solve this constrained problem, we use a primal-dual active set strategy [42, §2.12.4], which relies on solving the following system: Find $(\mathsf{y}, \mathsf{p}, \mathsf{u})$ such that

$$\begin{aligned} -\varepsilon^2\Delta\mathsf{y} + \mathsf{b}\cdot\nabla\mathsf{y} + \mathsf{c}\mathsf{y} - \mathsf{u} &= \mathsf{f}, \quad \mathsf{y} = 0 \text{ on } \partial\Omega, \\ -\varepsilon^2\Delta\mathsf{p} - \mathsf{b}\cdot\nabla\mathsf{p} + \mathsf{c}\mathsf{p} - \mathsf{y} &= -\mathsf{y}_d, \quad \mathsf{p} = 0 \text{ on } \partial\Omega, \\ (1 - \chi_a - \chi_b)\frac{\mathsf{p}}{\lambda} + \mathsf{u} &= \chi_a\mathsf{u}_a + \chi_b\mathsf{u}_b, \end{aligned}$$

where χ_a and χ_b are the characteristic functions of the active and inactive sets based on (5.1). We now construct an exact solution assuming that $\mathsf{b} = 0$ and $\mathsf{c} = 1$. We set the optimal adjoint state as

$$\bar{\mathsf{p}}(x_1, x_2) = -\lambda x_2(1 - x_2)\mathcal{E}(x_1, \varepsilon^2) = -\lambda\mathsf{u}_f(x_1, x_2),$$

and the optimal control

$$\bar{\mathsf{u}}(x_1, x_2) = \begin{cases} \mathsf{u}_a & \text{if } \mathsf{u}_f(x_1, x_2) < \mathsf{u}_a, \\ \mathsf{u}_f(x_1, x_2) & \text{if } \mathsf{u}_f(x_1, x_2) \in [\mathsf{u}_a, \mathsf{u}_b], \\ \mathsf{u}_b & \text{if } \mathsf{u}_f(x_1, x_2) > \mathsf{u}_b, \end{cases}$$

where $\mathsf{u}_f(x_1, x_2) = x_2(1 - x_2)\mathcal{E}(x_1, \varepsilon^2)$. In addition, we set $\bar{\mathsf{y}}(x_1, x_2) = x_1x_2(1 - x_1)(1 - x_2)$ and the data $\mathsf{y}_d = \varepsilon^2\Delta\bar{\mathsf{p}} + \mathsf{b}\cdot\nabla\bar{\mathsf{p}} - c\bar{\mathsf{p}} + \bar{\mathsf{y}}$ and $\mathsf{f} = -\varepsilon^2\Delta\bar{\mathsf{y}} + \mathsf{b}\cdot\nabla\bar{\mathsf{y}} + \bar{\mathsf{y}} - \bar{\mathsf{u}}$. To generate a singular behavior, we consider

$$\mathcal{E}(x_1, \varepsilon) = x_1 - \frac{e^{-(1-x_1)/\varepsilon^2} - e^{-1/\varepsilon^2}}{1 - e^{-1/\varepsilon^2}}.$$

Figure 5.8 shows the computational rates of convergence for our graded method, which solves the state and adjoint equations using the finite element space (3.2) on the graded meshes (3.5). The optimal control is approximated within the following finite element space

$$\tilde{\mathbb{V}}(\mathscr{T}) = \left\{\mathsf{u}_{\mathscr{T}} \in L^2(\Omega) : \mathsf{u}_{\mathscr{T}}|_T \in \mathbb{R}\,\forall T \in \mathscr{T},\ \mathsf{u}_a \le \mathsf{u}_{\mathscr{T}} \le \mathsf{u}_b\right\},$$

that is based on the graded meshes (3.5). The proposed graded technique delivers optimal rates of convergence; see Figure 5.8.

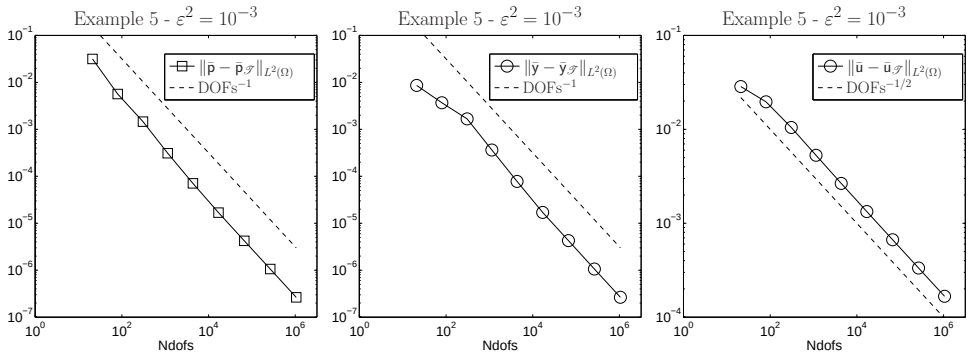

FIG. 5.8. *Example 5: Computational rates of convergence for our graded solution technique in the L^2–norm, with respect to the number of degrees of freedom (DOFs), when solving a constrained optimal control problem. We consider $\mathsf{b} = (0, 0)$ and $\mathsf{c} = 1$ in (1.3). The value of the parameter ε^2 considered is 10^{-3}. The graded meshes are based on (3.5) with $\gamma = 0.75$. In all cases we recover optimal rates of convergence.*

[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(3):229 – 251, 2004.

[2] M. Ainsworth, A. Allendes, G. R. Barrenechea, and R. Rankin. Fully computable a posteriori error bounds for stabilised FEM approximations of convection-reaction-diffusion problems in three dimensions. *Internat. J. Numer. Methods Fluids*, 73(9):765–790, 2013.

[3] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience, New York, 2000.

[4] M. Ainsworth and T. Vejchodský. Fully computable robust a posteriori error bounds for singularly perturbed reaction-diffusion problems. *Numer. Math.*, 119(2):219–243, 2011.

[5] P.R. Amestoy, I.S. Duff, and J.-Y. L'Excellent. Multifrontal parallel distributed symmetric and unsymmetric solvers. *Computer Methods in Applied Mechanics and Engineering*, 184(24):501 – 520, 2000.

[6] P.R. Amestoy, I.S. Duff, J.-Y. L'Excellent, and J. Koster. A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM J. Matrix Anal. Appl.*, 23(1):15–41 (electronic), 2001.

[7] V.B. Andreev and N. Kopteva. Pointwise approximation of corner singularities for a singularly perturbed reaction-diffusion equation in an L-shaped domain. *Math. Comp.*, 77(264):2125–2139, 2008.

[8] H. Antil, E. Otárola, and A.J. Salgado. A Space-Time Fractional Optimal Control Problem: Analysis and Discretization. *SIAM J. Control Optim.*, 54(3):1295–1328, 2016.

[9] T. Apel. *Anisotropic finite elements: local estimates and applications*. Advances in Numerical Mathematics. B. G. Teubner, Stuttgart, 1999.

[10] R. Becker and B. Vexler. Optimal control of the convection-diffusion equation using stabilized finite element methods. *Numer. Math.*, 106(3):349–367, 2007.

[11] M. Braack. Optimal control in fluid mechanics by finite elements with symmetric stabilization. *SIAM J. Control Optim.*, 48(2):672–687, 2009.

[12] S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[13] I. Cheddadi, R. Fučík, M. I. Prieto, and M. Vohralík. Guaranteed and robust a posteriori error estimates for singularly perturbed reaction-diffusion problems. *M2AN Math. Model. Numer. Anal.*, 43(5):867–888, 2009.

[14] S. Collis and M. Heinkenschloss. Analysis of the streamline upwind/Petrov Galerkin method applied to the solution of optimal control problems. *Technical report TR0201, Department of Computational and Applied Mathematics, Rice University, Houston, TX*, 2002.

[15] G. Dal Maso. *An introduction to Γ-convergence*. Progress in Nonlinear Differential Equations and their Applications, 8. Birkhäuser Boston, Inc., Boston, MA, 1993.

[16] L. Dede' and A. Quarteroni. Optimal control and numerical adaptivity for advection-diffusion equations. *M2AN Math. Model. Numer. Anal.*, 39(5):1019–1040, 2005.

[17] R. Durán and A. Lombardi. Error estimates on anisotropic Q_1 elements for functions in

weighted Sobolev spaces. *Math. Comp.*, 74(252):1679–1706 (electronic), 2005.

[18] R. Durán and A. Lombardi. Finite element approximation of convection diffusion problems using graded meshes. *Appl. Numer. Math.*, 56(10-11):1314–1325, 2006.

[19] R. Durán, A. Lombardi, and M.I. Prieto. Superconvergence for finite element approximation of a convection-diffusion equation using graded meshes. *IMA J. Numer. Anal.*, 32(2):511–533, 2012.

[20] R. Durán, A. Lombardi, and M.I. Prieto. Supercloseness on graded meshes for Q_1 finite element approximation of a reaction-diffusion equation. *J. Comput. Appl. Math.*, 242:232–247, 2013.

[21] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[22] A. Ern, A. F. Stephansen, and M. Vohralík. Guaranteed and robust discontinuous Galerkin a posteriori error estimates for convection-diffusion-reaction problems. *J. Comput. Appl. Math.*, 234(1):114–130, 2010.

[23] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[24] H. Han and R. B. Kellogg. Differentiability properties of solutions of the equation $-\epsilon^2\Delta u+ru = f(x, y)$ in a square. *SIAM J. Math. Anal.*, 21(2):394–408, 1990.

[25] 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(6):4607–4638, 2010.

[26] M. Hinze, N. Yan, and Z. Zhou. Variational discretization for optimal control governed by convection dominated diffusion equations. *J. Comput. Math.*, 27(2-3):237–253, 2009.

[27] C. Johnson. *Numerical solution of partial differential equations by the finite element method*. Studentlitteratur, Lund, 1987.

[28] K. Kohls, A. Rösch, and K. G. Siebert. A posteriori error analysis of optimal control problems with control constraints. *SIAM J. Control Optim.*, 52(3):1832–1861, 2014.

[29] N. Kopteva and E. O'Riordan. Shishkin meshes in the numerical solution of singularly perturbed differential equations. *Int. J. Numer. Anal. Model.*, 7(3):393–415, 2010.

[30] D. Leykekhman and M. Heinkenschloss. Local error analysis of discontinuous Galerkin methods for advection-dominated elliptic linear-quadratic optimal control problems. *SIAM J. Numer. Anal.*, 50(4):2012–2038, 2012.

[31] J. Li. Quasioptimal uniformly convergent finite element methods for the elliptic boundary layer problem. *Comput. Math. Appl.*, 33(10):11–22, 1997.

[32] J. Li and I. M. Navon. Uniformly convergent finite element methods for singularly perturbed elliptic boundary value problems. I. Reaction-diffusion type. *Comput. Math. Appl.*, 35(3):57–70, 1998.

[33] J. Li and M. F. Wheeler. Uniform convergence and superconvergence of mixed finite element methods on anisotropically refined grids. *SIAM J. Numer. Anal.*, 38(3):770–798, 2000.

[34] T. Linß. Uniform superconvergence of a Galerkin finite element method on Shishkin-type meshes. *Numer. Methods Partial Differential Equations*, 16(5):426–440, 2000.

[35] G. Lube and B. Tews. Optimal control of singularly perturbed advection-diffusion-reaction problems. *Math. Models Methods Appl. Sci.*, 20(3):375–395, 2010.

[36] R. H. Nochetto, E. Otárola, and A.J. Salgado. Piecewise polynomial interpolation in Muckenhoupt weighted Sobolev spaces and applications. *Numer. Math.*, 132(1):85–130, 2016.

[37] R.H. Nochetto and A. Vesser. Primer of adaptive finite element methods. In *Multiscale and Adaptivity: Modeling, Numerics and Applications, CIME Lectures*. Springer, 2011.

[38] E. O'Riordan and M. Stynes. A globally uniformly convergent finite element method for a singularly perturbed elliptic problem in two dimensions. *Math. Comp.*, 57(195):47–62, 1991.

[39] A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations*, volume 23 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1994.

[40] H. Roos and C. Reibiger. Numerical analysis of a system of singularly perturbed convection-diffusion equations related to optimal control. *Numer. Math. Theory Methods Appl.*, 4(4):562–575, 2011.

[41] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust numerical methods for singularly perturbed differential equations*, volume 24 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2008. Convection-diffusion-reaction and flow problems.

[42] F. Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.

[43] R. Verfürth. A posteriori error estimators for convection-diffusion equations. *Numer. Math.*, 80(4):641–663, 1998.

[44] R. Verfürth. *A posteriori error estimation techniques for finite element methods*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2013.

[45] E. A. Volkov. Differential properties of solutions of boundary value problems for the Laplace equation on polygons. *Trudy Mat. Inst. Steklov*, 77:113–142, 1965.

[46] N. Yan and Z. Zhou. A priori and a posteriori error estimates of streamline diffusion finite element method for optimal control problem governed by convection dominated diffusion equation. *Numer. Math. Theory Methods Appl.*, 1(3):297–320, 2008.

[47] Z. Zhou, X. Yu, and N. Yan. Local discontinuous Galerkin approximation of convection-dominated diffusion optimal control problems with control constraints. *Numer. Methods Partial Differential Equations*, 30(1):339–360, 2014.