A weighted and balanced FEM for singularly perturbed reaction-diffusion problems

Niall Madden[1] · Martin Stynes[2]

Abstract

A new finite element method is presented for a general class of singularly perturbed reaction-diffusion problems $-\varepsilon^2 \Delta u + bu = f$ posed on bounded domains $\Omega \subset \mathbb{R}^k$ for $k \geq 1$, with the Dirichlet boundary condition $u = 0$ on $\partial\Omega$, where $0 < \varepsilon \ll 1$. The method is shown to be quasioptimal (on arbitrary meshes and for arbitrary conforming finite element spaces) with respect to a weighted norm that is known to be balanced when one has a typical decomposition of the unknown solution into smooth and layer components. A robust (i.e., independent of ε) almost first-order error bound for a particular FEM comprising piecewise bilinears on a Shishkin mesh is proved in detail for the case where Ω is the unit square in $\mathbb{R}^2$. Numerical results illustrate the performance of the method.

Keywords Finite element method · Balanced norm · Quasioptimal

Mathematics Subject Classification 65N30 · 65N12

1 Introduction

Singularly perturbed differential equations of reaction-diffusion type have been extensively studied, as described in [16, 21, 23]. When a standard Galerkin finite element method (FEM) is used to solve these problems, it is straightforward to carry out the usual "energy norm" analysis, but a serious weakness of this measure of the

✉ Martin Stynes
m.stynes@csrc.ac.cn

Niall Madden
niall.madden@nuigalway.ie

1   School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, Ireland

2   Division of Applied and Computational Mathematics, Beijing Computational Science Research Center, Beijing, China

error is that when the singular perturbation parameter (i.e., the diffusion coefficient in the PDE) is very small, the energy norm of the error in the computed solution is essentially no stronger than the L^2 norm of the error; that is, the H^1 component of the energy norm is typically much smaller than its L^2 component. This drawback is discussed at length in [14], where the authors proposed the replacement of the energy norm by a stronger *balanced norm* whose H^1 component is scaled to the correct size, so that both the H^1 and L^2 components of the solution are $O(1)$ when the singular perturbation parameter is small.

[14] also proposed a new bilinear form and FEM that were designed to facilitate analysis in their balanced norm. Subsequently, other authors derived new FEMs and new analyses that yielded convergence in various balanced norms; see in particular [1, 2, 5, 12, 18, 20, 22].

In the present paper, we consider a singularly perturbed reaction-diffusion problem that is posed on an arbitrary bounded domain $\Omega \subset \mathbb{R}^k$ for $k \geq 1$ and present a new and simple way of constructing a FEM that is convergent in a weighted balanced norm. This new norm is stronger than the standard energy norm because of the weighting that it includes.

Our key idea is to modify the standard Galerkin FEM by introducing a weight function that was (essentially) already used in [2], but unlike [2] we do not rewrite the reaction-diffusion problem as a system of equations. Furthermore, the analysis of [2] is for a problem posed on the unit square in $\mathbb{R}^2$, but the analysis in the current paper permits a far more general class of domains.

Our method is simpler than any other FEM that is designed to yield convergence in a balanced norm. It stands on a solid theoretical foundation: Theorem 1 shows that on an arbitrary bounded domain, for arbitrary meshes and an arbitrary conforming FEM space, the computed solution is quasioptimal with respect to our weighted norm. This norm can be shown to be balanced for problems where one has a typical decomposition of the unknown solution into smooth and layer components; see Remarks 3 and 4.

The structure of the paper is as follows. In Sect. 2 we state our reaction-diffusion problem and construct a weighted norm and associated bilinear form for which we derive various fundamental properties. These results are used to construct our weighted FEM in Sect. 3, and quasioptimality of the FEM solution is proved. Then in Sect. 4 we specialise this general theory to the particular case where the domain is the unit square in 2D and the FEM uses bilinears on a Shishkin mesh. Here a detailed error analysis leads to the optimal-order convergence result of Theorem 2, where we show that (up to a log factor) the weighted FEM attains first-order convergence in our weighted norm. Finally, Sect. 5 presents numerical results to show the performance of the weighted FEM.

2 The singularly perturbed reaction–diffusion problem

Let Ω be a bounded domain in $\mathbb{R}^k$, where $k \geq 1$. Write $\bar{\Omega}$ for the closure of Ω and $\partial\Omega$ for its boundary. We shall discuss the elliptic boundary value problem

$$Lu := -\varepsilon^2 \Delta u + bu = f \quad \text{in } \Omega, \tag{1a}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{1b}$$

where the diffusion parameter satisfies $0 < \varepsilon \le 1$, but the challenging and interesting case is when $\varepsilon \ll 1$; then (1) is a singularly perturbed reaction-diffusion problem. Assume that the reaction term satisfies $b \in L_\infty(\Omega)$ with $b_0^2 < b(x) \le b_1^2$ for almost all $x \in \Omega$, where b_0, b_1 are positive constants. This assumption is usual in singularly perturbed problems of this type—see [16, 21, 23].

It is well known that for each $f \in L^2(\Omega)$ the problem (1) has a unique solution $u \in H^2(\Omega) \cap H_0^1(\Omega)$ if Ω is convex or if $\partial\Omega$ is smooth, but throughout Sects. 2 and 3 we assume only that Ω is bounded.

Notation. Let ω be any measurable subset of $\bar{\Omega}$ and let $\partial\Omega$ be its boundary. We use standard notation for the Sobolev spaces $W^{m,p}(\omega)$ and their associated norms $\|\cdot\|_{m,p,\omega}$ and seminorms $|\cdot|_{m,p,\omega}$. If $m = 0$, we simply write $\|\cdot\|_{p,\omega}$; while if $p = 2$ we set $H^m(\omega) = W^{m,2}(\omega)$ and $\|\cdot\|_{m,\omega} = \|\cdot\|_{m,2,\omega}$. As usual, $H_0^1(\omega) = \{v \in H^1(\omega) : v|_{\partial\Omega} = 0\}$, where $v|_{\partial\Omega}$ is the trace of v on $\partial\Omega$, and $L^2(\omega) = H^0(\omega)$.

Throughout this paper, the letter C (with or without subscripts) will denote a generic positive constant that may stand for different values in different places but is independent of the parameter ε and of the mesh diameter.

2.1 Weighted norm

To solve (1), we shall apply a FEM that uses standard trial and test functions on general meshes; its distinguishing feature is that it incorporates a special weight in the bilinear form and its associated norm, where the weight is chosen in such a way that the norm is balanced (i.e., the $H^1(\Omega)$ and $L^2(\Omega)$ weighted components of our norm are commensurable for typical solutions of (1)); see Remarks 3 and 4.

For each $x \in \Omega$, set

$$d(x) = \min\{|x - z| : z \in \partial\Omega\},$$

where $|x - z|$ denotes the Euclidean distance from x to z. That is, $d(x)$ is the Euclidean distance from x to the boundary of Ω.

For any $y \in \Omega$, choose $z \in \partial\Omega$ such that $d(y) = |y - z|$. Then $d(x) \le |x - z| \le |x - y| + d(y)$, by a triangle inequality. Similarly $d(y) \le |y - x| + d(x)$. Hence

$$|d(x) - d(y)| \le |x - y| \quad \text{for all } x, y \in \Omega. \tag{2}$$

(This proof of (2) comes from [10, p.354].) The inequality (2) says that $d(\cdot)$ is uniformly Lipschitz continuous on $\bar{\Omega}$. Then Rademacher's Theorem [7, p.296] guarantees that the function $d(\cdot)$ is differentiable almost everywhere in Ω, and *a fortiori* (2) clearly implies that

$$|\nabla d(x)| \leq 1 \text{ almost everywhere.} \tag{3}$$

Define our weight function β by

$$\beta(x) = 1 + \frac{1}{\varepsilon} \exp\left(-\frac{\gamma d(x)}{\varepsilon}\right) \text{ for } x \in \Omega, \tag{4}$$

where the positive constant γ is (for the moment) arbitrary. Then

$$\nabla\beta(x) = -\frac{\gamma \nabla d(x)}{\varepsilon^2} \exp\left(-\frac{\gamma d(x)}{\varepsilon}\right) \text{ almost everywhere in } \Omega.$$

Recalling (3), we get

$$|\nabla\beta(x)| \leq \frac{\gamma}{\varepsilon^2} \exp\left(-\frac{\gamma d(x)}{\varepsilon}\right) \leq \frac{\gamma}{\varepsilon} \beta(x) \text{ almost everywhere in } \Omega. \tag{5}$$

Let $(\cdot, \cdot)$ denote the scalar and vector inner products in $L^2(\Omega)$. Write $\|\cdot\|$ for the norm associated with this inner product. For each $v \in L^2(\Omega)$ and $w \in H^1(\Omega)$, define

$$\|v\|_\beta := (\beta v, v)^{1/2} \text{ and } |||w|||_\beta := \left[\varepsilon^2 \|\nabla w\|_\beta^2 + \|w\|_\beta^2\right]^{1/2}.$$

This norm is balanced (i.e., its components $\varepsilon^2(\beta\nabla u, \nabla u)$ and $\|u\|_\beta^2$ are both $O(1)$) for a typical solution u of (1); see Remarks 3 and 4.

Remark 1 (Comparison with other balanced norms used to solve (1)) In [2] the balanced norm used comprises the above norm $|||w|||_\beta$ with some extra terms added. The balanced norm used in [5, 18, 20, 22] is $\|w\|_\varepsilon := \left[\varepsilon\|\nabla w\|^2 + \|w\|^2\right]^{1/2}$. In [1, 12, 14] extra terms are added to $\|w\|_\varepsilon$ to form a balanced norm. Of course, all these norms are stronger than the standard "energy norm" $\left[\varepsilon^2\|\nabla w\|^2 + \|w\|^2\right]^{1/2}$ that is often associated with (1).

Remark 2 (Motivation and applicability) The weight function β defined in (4) is a generalisation of the weight function used in [2] for a problem posed on the unit square in $\mathbb{R}^2$. It is motivated by the desire to have the simplest weight with the following two properties: that $\beta \approx 1$ away from layers so the method is like the standard Galerkin method, and that $\beta \approx 1/\varepsilon$ inside layers so that the component $\varepsilon^2\|\nabla w\|^2$ in the standard energy norm is essentially replaced by $\varepsilon\|\nabla w\|^2$ in regions where we want this term to be $O(1)$.

Our weight function β should work satisfactorily for any layer that has the same exponential decay nature as the boundary layers described in Lemma 3 below, provided that β is suitably modified: one replaces "distance to $\partial\Omega$" by "distance to the layer location". For example, the layers appearing in certain reaction-diffusion problems with interior layers are of this type; see [4, 8].

Remark 3 (Heuristic justification that $|||\cdot|||_\beta$ is balanced) Given sufficient information about the data of (1), a standard technique due to Shishkin enables us to decompose its solution as $u = v + w$, where v satisfies $Lv = f$ on Ω with all low-order derivatives of v bounded independently of ε, while $Lw = 0$ on Ω with $w = -v$ on $\partial\Omega$, so w is the boundary layer component of u. This is done for example in [6], where $\Omega \subset \mathbb{R}^2$ is the unit square. Using the properties of v and w, one finds that typically $|||v|||_\beta \sim \|v\|_\beta = O(1)$ and $|||w|||_\beta \sim \varepsilon\|\nabla w\|_\beta = O(1)$, so each component of $|||u|||_\beta$ is $O(1)$, demonstrating that $|||\cdot|||_\beta$ is balanced.

2.2 Weighted bilinear form

For all $v, w \in H_0^1(\Omega)$, define the bilinear form

$$B_\beta(v, w) := (\varepsilon^2 \nabla v, \nabla(\beta w)) + (bv, \beta w),$$

We prove two fundamental properties of $B_\beta(\cdot, \cdot)$.

Lemma 1 (coercivity of $B_\beta(\cdot, \cdot)$) *Assume that $0 < \gamma \le b_0$. Then for the weight function β specified in* (4), *one has*

$$B_\beta(v, v) > C_1 \, |||v|||_\beta^2 \text{ for all } v \in H_0^1(\Omega),$$

where $C_1 := \min\{1/2, b_0^2/2\}$.

Proof Let $v \in H_0^1(\Omega)$. Then

$$\begin{aligned}
B_\beta(v, v) &= (\varepsilon^2 \nabla v, \nabla(\beta v)) + (bv, \beta v) \\
&\ge \varepsilon^2(\nabla v, \beta \nabla v + v \nabla \beta) + (b_0 v, \beta v) \\
&= \varepsilon^2 \|\nabla v\|_\beta^2 + (\varepsilon^2 \nabla v, v \nabla \beta) + b_0^2 \|v\|_\beta^2.
\end{aligned} \tag{6}$$

Recalling (5), we have

$$\begin{aligned}
\left| (\varepsilon^2 \nabla v, v \nabla \beta) \right| &\le (\varepsilon|\nabla v|, |v|\gamma\beta) \\
&\le \varepsilon\|\nabla v\|_\beta \, \gamma\|v\|_\beta \\
&\le \frac{\varepsilon^2}{2}\|\nabla v\|_\beta^2 + \frac{\gamma^2}{2}\|v\|_\beta^2,
\end{aligned}$$

by the Cauchy-Schwarz and Young inequalities. Inserting this bound into (6), one obtains

$$B_\beta(v, v) \ge \frac{\varepsilon^2}{2}\|\nabla v\|_\beta^2 + \left(b_0^2 - \frac{\gamma^2}{2}\right)\|v\|_\beta^2.$$

The lemma follows, since $\gamma \le b_0$. $\qquad\square$

Lemma 2 (*boundedness of $B_\beta(\cdot, \cdot)$*) *For the weight function β specified in* (4), *one has*

$$|B_\beta(v, w)| \le C_2 |||v|||_\beta \, |||w|||_\beta \text{ for all } v, w \in H_0^1(\Omega),$$

where $C_2 := \max\{1, \gamma, b_1^2\}$.

Proof Let $v, w \in H_0^1(\Omega)$. Then

$$B_\beta(v, w) = (\varepsilon^2 \nabla v, \nabla(\beta w)) + (bv, \beta w) = (\varepsilon^2 \beta \nabla v, \nabla w) + (\varepsilon^2 \nabla v, w \nabla \beta) + (bv, \beta w). \tag{7}$$

Now (5) yields

$$|(\varepsilon^2 \nabla v, w \nabla \beta)| \le \gamma \varepsilon(|\nabla v|, |w|\beta) \le \gamma \varepsilon \|\nabla v\|_\beta \|w\|_\beta$$

by a Cauchy-Schwarz inequality. Applying two more Cauchy-Schwarz inequalities to (7), we get

$$\left| B_\beta(v, w) \right| \le \varepsilon^2 \|\nabla v\|_\beta \|\nabla w\|_\beta + \gamma \varepsilon \|\nabla v\|_\beta \|w\|_\beta + b_1^2 \|v\|_\beta \|w\|_\beta.$$

The desired result follows. $\square$

From now on, we assume that $f \in H^{-1}(\Omega) := \left(H_0^1(\Omega)\right)'$. One can then define a weak solution $u \in H_0^1(\Omega)$ of (1) by requiring it to satisfy

$$B_\beta(u, v) = (f, \beta v) \text{ for all } v \in H_0^1(\Omega), \tag{8}$$

where β is defined in (4) and the constant γ satisfies $0 < \gamma \le b_0$. By the Lax-Milgram theorem, invoking Lemmas 1 and 2, the problem (8) has a unique solution $u \in H_0^1(\Omega)$. Any strong solution $u \in H^2(\bar{\Omega}) \cap H_0^1(\Omega)$ of the original problem (1) is also a weak solution of (8), as can be seen by multiplying (1a) by βv then integrating by parts.

3 The weighted finite element method

Let Ω_h be an arbitrary mesh on $\bar{\Omega}$. Let $V_h \subset H_0^1(\Omega)$ be a conforming finite element space defined on this mesh. Suppose that $0 < \gamma \le b_0$. From the Lax-Milgram theorem and Lemmas 1 and 2, there is a unique $u_h \in V_h$ such that

$$B_\beta(u_h, v_h) = (f, \beta v_h) \text{ for all } v_h \in V_h. \tag{9}$$

Combining (8) and (9) gives the Galerkin orthogonality property

$$B_\beta(u - u_h, v_h) = 0 \text{ for all } v_h \in V_h. \tag{10}$$

Theorem 1 (quasioptimal FEM error bound) *Let Ω be a bounded subset of $\mathbb{R}^k$ for some $k \geq 1$. Let Ω_h be an arbitrary mesh on $\bar{\Omega}$, and $V_h \subset H_0^1(\Omega)$ a finite element space defined on this mesh. Choose $\gamma \in \mathbb{R}$ to satisfy $0 < \gamma \leq b_0$. Let u be the weak solution of (8) and u_h the solution of (9). Then one has*

$$|||u - u_h|||_\beta \leq \frac{C_2}{C_1} \inf_{w_h \in V_h} |||u - w_h|||_\beta.$$

Proof Let $w_h \in V_h$ be arbitrary. Invoking Lemma 1, then equation (10), then Lemma 2, we get

$$\begin{aligned} C_1 |||u - u_h|||_\beta^2 &\leq B_\beta(u - u_h, u - u_h) \\ &= B_\beta(u - u_h, u - w_h) \\ &\leq C_2 |||u - u_h|||_\beta |||u - w_h|||_\beta. \end{aligned}$$

Hence $|||u - u_h|||_\beta \leq (C_2/C_1)|||u - w_h|||_\beta$. As $w_h \in V_h$ was arbitrary, we are done.
$\square$

The analysis up to this point is for a general bounded domain, a general mesh, and a general conforming finite element space V_h. This generality cannot however yield a bound in Theorem 1 that will ensure that $|||u - u_h|||_\beta$ is small; to get this desirable outcome, one must tailor V_h or the mesh to the singularly perturbed nature of the problem. In Sect. 4 we show how this is done for a specific domain by a suitable choice of mesh.

4 Reaction-diffusion problem on the unit square in $\mathbb{R}^2$

We now specialise the theory of Sect. 2 to a 2D reaction-diffusion problem that has been considered in many balanced-norm papers, including [1, 12, 14].

During Sect. 4 we take $\Omega = (0, 1)^2$, the unit square in $\mathbb{R}^2$. Assume that $f, b \in C^{0,\alpha}(\bar{\Omega})$, where we use the standard notation for Hölder spaces. Assume the corner compatibility conditions

$$f(0, 0) = f(1, 0) = f(0, 1) = f(1, 1) = 0.$$

Then (1) has a unique solution $u \in C^{2,\alpha}(\bar{\Omega})$; see, e.g., [11]. Furthermore, this solution has typically an exponential boundary layer in a neighbourhood of $\partial\Omega$ of width $O(\varepsilon |\ln \varepsilon|)$; see Lemma 3 below for more details.

The 4 sides of Ω will be denoted by

$$\begin{aligned} \Gamma_1 &:= \{(x, 0)|0 \leq x \leq 1\}, \quad \Gamma_2 := \{(0, y)|0 \leq y \leq 1\}, \\ \Gamma_3 &:= \{(x, 1)|0 \leq x \leq 1\}, \quad \Gamma_4 := \{(1, y)|0 \leq y \leq 1\}. \end{aligned}$$

The corners of this domain are $z_1 := (0, 0), z_2 := (1, 0), z_3 := (1, 1), z_4 := (0, 1)$.

In [6, Theorem 2.2] and [17, Lemma 1.2] the solution u of (1) with $\Omega = (0, 1)^2$ is decomposed into smooth and layer components, and bounds are derived on

certain derivatives of these components. The analysis in these papers makes stronger assumptions than ours, but an inspection of their arguments shows however that, under our assumptions on the data, one obtains the following result, where for brevity the derivative $\partial^{m+n} g / \partial^m x \partial^n y$ of any function g is written as $D_x^m D_y^n g$.

Lemma 3 *The solution u of (1) can be decomposed as*

$$u = v + \sum_{k=1}^{4} w_k + \sum_{k=1}^{4} z_k, \tag{11}$$

where each w_k is a layer associated with the edge Γ_k and each z_k is a layer associated with the corner c_k. There exists a constant C such that for all $(x, y) \in \bar{\Omega}$ and $0 \le m + n \le 2$ one has

$$\|D_x^m D_y^n v\|_{0,\infty,\bar{\Omega}} \le C, \tag{12a}$$

$$\left| D_x^m D_y^n w_1(x, y) \right| \le C \varepsilon^{-n} e^{-b_0 y / \varepsilon}, \tag{12b}$$

$$\left| D_x^m D_y^n z_1(x, y) \right| \le C \varepsilon^{-m-n} e^{-b_0 (x+y) / \varepsilon}. \tag{12c}$$

Bounds for w_2, w_3 and w_4 that are analogous to (12b) and bounds for z_2, z_3 and z_4 that are analogous to (12c) also hold.

Remark 4 For a solution u that enjoys the properties described in Lemma 3, one can verify that for any constant $\gamma > 0$, one has $\varepsilon \|\nabla u\|_\beta = O(1)$ and $\|u\|_\beta = O(1)$. That is, the norm $||| \cdot |||_\beta$ is balanced for this problem.

4.1 The Shishkin mesh

To solve the problem numerically, we shall use a piecewise-uniform Shishkin mesh. These meshes are a popular tool in the numerical solution of problems such as (1). For an introduction to their properties and usage, see [16, 21, 23].

Let N be an even positive integer. The mesh will use N mesh intervals in each coordinate direction. The *mesh transition parameter* λ will specify where the mesh changes from coarse to fine; we define it by

$$\lambda = \min \left\{ \frac{1}{4}, \frac{\sigma \varepsilon \ln N}{b_0} \right\}. \tag{13}$$

In this formula, the constant σ will be chosen later to facilitate our numerical analysis and is then used in the implementation of the finite element method.

Then, without loss of generality, one can assume that N is so large that (13) simplifies to

$$\lambda = \sigma \varepsilon b_0^{-1} \ln N. \tag{14}$$

Partition Ω as follows (see Fig. 1): $\bar{\Omega} = \Omega_{11} \cup \Omega_{21} \cup \Omega_{12} \cup \Omega_{22}$, where

$$\Omega_{11} = [\lambda, 1 - \lambda] \times [\lambda, 1 - \lambda], \qquad \Omega_{21} = ([0, \lambda] \cup [1 - \lambda, 1]) \times [\lambda, 1 - \lambda],$$
$$\Omega_{12} = [\lambda, 1 - \lambda] \times ([0, \lambda] \cup [1 - \lambda, 1]),$$
$$\Omega_{22} = ([0, \lambda] \times ([0, \lambda] \cup [1 - \lambda, 1])) \cup ([1 - \lambda, 1] \times ([0, \lambda] \cup [1 - \lambda, 1])).$$

Divide each of the x-intervals $[0, \lambda]$ and $[1 - \lambda, 1]$ into $N/4$ equidistant subintervals and divide $[\lambda, 1 - \lambda]$ into $N/2$ equidistant subintervals. This gives a coarse mesh on $[\lambda, 1 - \lambda]$ and a fine mesh on $[0, \lambda] \cup [1 - \lambda, 1]$. Divide the y-interval $[0,1]$ in the same way. Then the final 2-dimensional mesh is a tensor product of these 1-dimensional Shishkin meshes; see Fig. 1, where $N = 8$.

An explicit description of the mesh follows: one has $0 = x_0 < x_1 < \cdots < x_N = 1$ and $0 = y_0 < y_1 < \cdots < y_N = 1$, with mesh sizes $h_i := x_i - x_{i-1}$ and $k_j := y_j - y_{j-1}$ that are defined by

$$h_i = \begin{cases} h := 4\lambda N^{-1} & \text{for } i = 0, \ldots, N/4 \text{ and } i = 3N/4 + 1, \ldots, N, \\ H := 2(1 - 2\lambda)N^{-1} & \text{for } i = N/4 + 1, \ldots, 3N/4, \end{cases} \tag{15a}$$

and

$$k_j = \begin{cases} h & \text{for } j = 0, \ldots, N/4 \text{ and } j = 3N/4 + 1, \ldots, N, \\ H & \text{for } j = N/4 + 1, \ldots, 3N/4. \end{cases} \tag{15b}$$

The mesh divides Ω into a set $T^{N,N}$ of mesh rectangles R whose sides are parallel to the axes—see Fig. 1. The mesh is coarse on Ω_{11}, coarse/fine on $\Omega_{21} \cup \Omega_{12}$, and fine on Ω_{22}. The mesh is quasiuniform on Ω_{11} and its diameter d there satisfies $\sqrt{2}/N \leq d \leq 2\sqrt{2}/N$; on $\Omega_{12} \cup \Omega_{21}$, each mesh rectangle has dimensions $O(N^{-1})$ by $O(\varepsilon N^{-1} \ln N)$; and on Ω_{22} each rectangle is $O(\varepsilon N^{-1} \ln N)$ by $O(\varepsilon N^{-1} \ln N)$. These dimensions will be used in the error analysis of our finite element method.
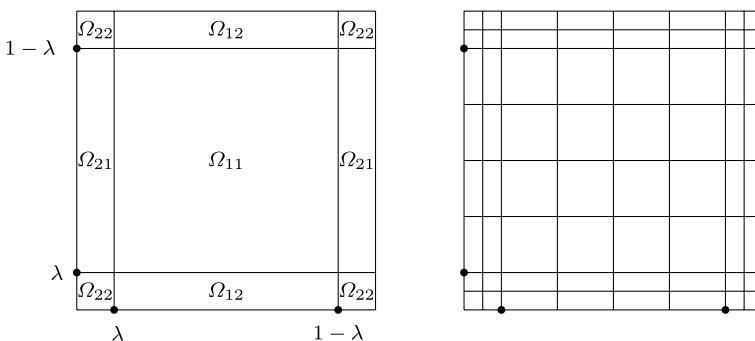

Fig. 1 Shishkin mesh for reaction-diffusion

4.2 Bilinear FEM on Shishkin mesh

We assume henceforth that the user-chosen constant γ satisfies $0 < \gamma \le b_0$, so Theorem 1 is valid.

For our finite element method, choose the finite space $V_h \subset C(\bar{\Omega}) \cap H_0^1(\Omega)$ to comprise piecewise bilinears on the Shishkin mesh of Sect. 4.1. Given any function $g \in C(\bar{\Omega}) \cap H_0^1(\Omega)$, we write g^I for the nodal interpolant of g from V_h.

The following interpolation error bound is suitable for the highly anisotropic Shishkin mesh.

Lemma 4 [3, *Theorem* 2.7] *Let R be any Shishkin mesh rectangle with dimensions $h_x \times k_y$. Let $\phi \in H^2(R)$. Then its bilinear nodal interpolant ϕ^I satisfies the bounds*

$$\|\phi - \phi^I\|_{\infty,R} \le C\left(h_x^2 \|\phi_{xx}\|_{\infty,R} + k_y^2 \|\phi_{yy}\|_{\infty,R} \right),$$

$$\|(\phi - \phi^I)_x\|_{\infty,R} \le C\left(h_x \|\phi_{xx}\|_{\infty,R} + k_y \|\phi_{xy}\|_{\infty,R} \right),$$

$$\|(\phi - \phi^I)_y\|_{\infty,R} \le C\left(h_x \|\phi_{xy}\|_{\infty,R} + k_y \|\phi_{yy}\|_{\infty,R} \right),$$

where the constant C is independent of ϕ, h_x and k_y.

Theorem 1 gives us

$$|||u - u_h|||_\beta \le \frac{C_2}{C_1} |||u - u^I|||_\beta. \tag{16}$$

We shall bound the right-hand side of (16) by using the decomposition of u from Lemma 3, comparing each component there with its interpolant from V_h.

Lemma 5 *For the component v in Lemma 3, one has*

$$|||v - v^I|||_\beta \le C\varepsilon N^{-1} + N^{-2}.$$

Proof Lemmas 3 and 4 yield $\|v - v^I\|_{\infty,\Omega} \le CN^{-2} |v|_{2,\infty,\Omega} \le CN^{-2}$. Hence

$$\|v - v^I\|_\beta = \left[\int_\Omega \beta(v - v^I)^2 \right]^{1/2} \le CN^{-2} \left[\int_\Omega \beta \right]^{1/2} \le CN^{-2}.$$

By a similar argument, again using Lemmas 3 and 4, one has $\|\nabla(v - v^I)\|_{\infty,\Omega} \le CN^{-1} |v|_{2,\infty,\Omega} \le CN^{-1}$ and hence $\|\nabla(v - v^I)\|_\beta \le CN^{-1}$. To finish the proof, recall the definition of $||| \cdot |||_\beta$. $\square$

For any measurable $\omega \subset \Omega$, and each $v \in L^2(\omega)$ and $w \in H^1(\omega)$, define

$$\|v\|_{\beta,\omega} := \left[\int_\omega \beta v^2 \right]^{1/2} \quad \text{and} \quad |||w|||_{\beta,\omega} := \left[\left(\varepsilon^2 \int_\omega \beta |\nabla w|^2 \right) + \|w\|_{\beta,\omega}^2 \right]^{1/2}.$$

Lemma 6 *For the components w_1, w_2, w_3, w_4 of Lemma 3, one has*

$$|||w_j - w_j^I|||_\beta \leq C(N^{-1} \ln N + N^{-\sigma}) \text{ for } j = 1, 2, 3, 4.$$

Proof We give the proof only for w_1, as the other w_j are similar.

Set $\Omega_y = \{(x, y) \in \Omega : y \geq \lambda\}$. By Lemma 3, one has

$$\left| D_x^m D_y^n w_1(x, y) \right| \leq C\varepsilon^{-n} N^{-\sigma} \quad \text{for } 0 \leq m + n \leq 2 \text{ and } (x, y) \in \Omega_y.$$

Hence,

$$\|w_1 - w_1^I\|_{\infty, \Omega_y} \leq \|w_1\|_{\infty, \Omega_y} + \|w_1^I\|_{\infty, \Omega_y} = 2\|w_1\|_{\infty, \Omega_y} \leq CN^{-\sigma}$$

and

$$\|\nabla(w_1 - w_1^I)\|_{\infty, \Omega_y} \leq \|\nabla w_1\|_{\infty, \Omega_y} + \|\nabla w_1^I\|_{\infty, \Omega_y} \leq 2\|\nabla w_1\|_{\infty, \Omega_y} \leq C\varepsilon^{-1} N^{-\sigma}.$$

It now follows, as in the proof of Lemma 5, that

$$|||w_1 - w_1^I|||_{\beta, \Omega_y} \leq CN^{-\sigma}. \tag{17}$$

On $\Omega \setminus \Omega_y$ invoke Lemma 4 and (15) to get

$$\begin{aligned}
\|w_1 - w_1^I\|_{\infty, \Omega \setminus \Omega_y} &\leq C\Big[H^2 \|(w_1)_{xx}\|_{\infty, \Omega \setminus \Omega_y} + h^2 \|(w_1)_{yy}\|_{\infty, \Omega \setminus \Omega_y}\Big] \\
&\leq C\big[N^{-2} + (\varepsilon N^{-1} \ln N)^2 \varepsilon^{-2}\big] \\
&= C(N^{-1} \ln N)^2
\end{aligned}$$

and

$$\begin{aligned}
\|\nabla(w_1 - w_1^I)\|_{\infty, \Omega \setminus \Omega_y} &\leq C\Big[H\|(w_1)_{xx}\|_{\infty, \Omega \setminus \Omega_y} + (h + H)\|(w_1)_{xy}\|_{\infty, \Omega \setminus \Omega_y} \\
&\quad + h\|(w_1)_{yy}\|_{\infty, \Omega \setminus \Omega_y}\Big] \\
&\leq C\big[N^{-1}(1 + \varepsilon^{-1}) + (\varepsilon N^{-1} \ln N)(\varepsilon^{-1} + \varepsilon^{-2})\big] \\
&= C\varepsilon^{-1}(N^{-1} \ln N),
\end{aligned}$$

where we used Lemma 3 to bound the derivatives of w_1. From these estimates it follows, as in the proof of Lemma 5, that

$$|||w_1 - w_1^I|||_{\beta, \Omega \setminus \Omega_y} \leq CN^{-1} \ln N. \tag{18}$$

Putting together (17) and (18) completes the proof. □

Next we prove the corresponding result for the components z_j of u; the proof resembles the proof of Lemma 6 but there are some differences.

Lemma 7 *For the components z_1, z_2, z_3, z_4 in Lemma 3, one has*

$$|||z_j - z_j^I|||_\beta \le C(N^{-1} \ln N + N^{-\sigma}) \quad \text{for } j = 1, 2, 3, 4.$$

Proof We give the proof only for z_1, as the other z_j are similar.

Set $\Omega_{xy} = \{(x, y) \in \Omega : x \ge \lambda \text{ or } y \ge \lambda\}$. By Lemma 3, one has

$$\left| D_x^m D_y^n w_1(x, y) \right| \le C\varepsilon^{-m-n} N^{-\sigma} \quad \text{for } 0 \le m + n \le 2 \text{ and } (x, y) \in \Omega_{xy}.$$

Hence,

$$\|z_1 - z_1^I\|_{\infty,\Omega_y} \le \|z_1\|_{\infty,\Omega_y} + \|z_1^I\|_{\infty,\Omega_y} = 2\|z_1\|_{\infty,\Omega_y} \le CN^{-\sigma}$$

$$\text{and}$$

$$\|\nabla(z_1 - z_1^I)\|_{\infty,\Omega_y} \le \|\nabla z_1\|_{\infty,\Omega_y} + \|\nabla z_1^I\|_{\infty,\Omega_y} \le 2\|\nabla z_1\|_{\infty,\Omega_y} \le C\varepsilon^{-1} N^{-\sigma}.$$

It now follows, as in the proof of Lemma 5, that

$$|||z_1 - z_1^I|||_{\beta,\Omega_y} \le CN^{-\sigma}. \tag{19}$$

On $\Omega \setminus \Omega_{xy}$ invoke Lemma 4 and (15) to get

$$\begin{aligned}
\|z_1 - z_1^I\|_{\infty,\Omega\setminus\Omega_y} &\le C\Big[h^2\|(z_1)_{xx}\|_{\infty,\Omega\setminus\Omega_y} + h^2\|(z_1)_{yy}\|_{\infty,\Omega\setminus\Omega_y}\Big] \\
&\le C(\varepsilon N^{-1} \ln N)^2 \varepsilon^{-2} \\
&= C(N^{-1} \ln N)^2 \\
&\text{and}
\end{aligned}$$

$$\begin{aligned}
\|\nabla(z_1 - z_1^I)\|_{\infty,\Omega\setminus\Omega_y} &\le C\Big[h\|(z_1)_{xx}\|_{\infty,\Omega\setminus\Omega_y} + h\|(z_1)_{xy}\|_{\infty,\Omega\setminus\Omega_y} \\
&\qquad + h\|(z_1)_{yy}\|_{\infty,\Omega\setminus\Omega_y}\Big] \\
&\le C(\varepsilon N^{-1} \ln N)(1 + \varepsilon^{-1} + \varepsilon^{-2}) \\
&= C\varepsilon^{-1}(N^{-1} \ln N),
\end{aligned}$$

where we used Lemma 3 to bound the derivatives of z_1. From these estimates it follows, as in the proof of Lemma 5, that

$$|||z_1 - z_1^I|||_{\beta,\Omega\setminus\Omega_y} \le CN^{-1} \ln N. \tag{20}$$

Combine (19) and (20) to finish the argument. $\qquad\square$

We can now give a precise error estimate for our finite element method.

Theorem 2 *Let u be the solution of* (1) *and u_h the solution of* (9), *where the finite element method uses piecewise bilinears on the Shishkin mesh. Then there exists a constant C such that*

$$|||u - u_h|||_\beta \le C(N^{-1} \ln N + N^{-\sigma}).$$

With the choice $\sigma = 1$ in the definition (14) *of the Shishkin mesh transition parameter*, *this bound becomes*

$$|||u - u_h|||_\beta \leq CN^{-1} \ln N.$$

Proof Lemmas 5, 6 and 7 and the decomposition of Lemma 3 yield

$$|||u - u^I|||_\beta \leq C(N^{-1} \ln N + N^{-\sigma}).$$

The result then follows from Lemma 1.   □

Remark 5 One could instead use a FEM space that comprises polynomials of higher degree and carry out a similar error analysis to obtain a higher order of convergence in Theorem 2, but this necessitates imposing more conditions on the data of (1) so that more derivatives of u are bounded in Lemma 3, as in [6, 17].

Remark 6 [9] extend the reaction-diffusion analysis of [20] to the convection-diffusion problem

$$-\varepsilon \Delta v + a v_x + b v = f \text{ on } \Omega = (0, 1)^2, \quad v = 0 \text{ on } \partial\Omega, \tag{21}$$

where the functions a and b are positive. (Here, as is customary, we have replaced the ε^2 coefficient of (1a) by ε since this is a convection–diffusion problem.) Typical solutions of this problem exhibit an exponential layer along the side $x = 1$ of Ω, and parabolic layers along the sides $y = 0$ and $y = 1$. The associated energy norm $\left[\varepsilon|\nabla v|^2 + \|v\|_0^2\right]^{1/2}$ is correctly balanced for the exponential layer, but is unbalanced for the weaker parabolic layers, for which it reduces essentially to the $L^2(\Omega)$ norm when $\varepsilon \ll 1$. In [9] an FEM comprising streamline-diffusion (SDFEM) in the x-direction and standard Galerkin in the y-direction is used and is shown to converge in the balanced norm $\left[\varepsilon|v_x|^2 + \varepsilon^{1/2}|v_y|^2 + \|v\|_0^2\right]^{1/2}$ on a Shishkin mesh that is appropriate for this problem.

One can carry out an analogous construction and analysis in our setting: working with piecewise bilinears on a Shishkin mesh like that of [9], construct an FEM that uses SDFEM in the x-direction (note that we can use the standard SDFEM, unlike the special variant of SDFEM that is used in [9]) and replaces (4) by the weight function

$$\hat{\beta}(x, y) := 1 + \frac{1}{\varepsilon^{1/2}} \exp\left(-\frac{\gamma[1/2 - |y - 1/2|]}{\varepsilon^{1/2}}\right) \text{ for } (x, y) \in \Omega.$$

This weight is a function of y only and is large inside the parabolic layers. One can prove convergence on the Shishkin mesh in the weighted balanced norm

$$\left[\varepsilon|v_x|^2 + \varepsilon\hat{\beta}|v_y|^2 + \|v\|_0^2\right]^{1/2}. \tag{22}$$

We do not give details here as they are lengthy and require no new idea beyond a synthesis of our earlier analysis and that of [9].

5 Numerical results

Our test problem is

$$-\varepsilon^2 \Delta u + u = f \text{ in } \Omega = (0,1)^2, \qquad u = 0 \text{ on } \partial\Omega, \tag{23}$$

where f chosen so that

$$u(x,y) = \left(\cos\left(\frac{\pi}{2}x\right) - \frac{e^{-x/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}}\right)\left(1 - y - \frac{e^{-y/\varepsilon} - e^{-1/\varepsilon}}{1 - e^{-1/\varepsilon}}\right). \tag{24}$$

This problem is taken from [13] and is widely used in the literature (see, e.g., [1]). Its solution exhibits only two layers, which are near $x = 0$ and $y = 0$. Nonetheless, the bounds presented in Lemma 3 hold and are sharp for v, w_1, w_2, and z_1. Thus, the example is sufficiently typical to verify the results of Theorem 2.

Since this problem has layers along only two edges, we modify the Shishkin mesh of Sect. 4.1 so that it is a tensor product of two one-dimensional meshes with $N/2$ equidistant subintervals on each of $[0, \lambda]$ and $[\lambda, 1]$. In our experiments we have taken $\gamma = 0.98$ in (4), and $b_0 = 0.99$ and $\sigma = 1$ in (13).

Our results are computed using Firedrake; see [19]. All results presented here are for bilinear elements; consistent results were obtained using biquadratic elements (see Remark 5).

In Table 1 we present the errors in the solutions computed using our proposed method. Observe that, for sufficiently small ε, the errors are independent of ε, and converge at a rate that is $\mathcal{O}(N^{-1} \ln N)$, verifying Theorem 2. This contrasts with the observed results for the classical Galerkin method (i.e., $\beta(\cdot) \equiv 1$) for this problem on this Shishkin mesh, where the computed errors in the standard energy norm scale like $\varepsilon^{1/2}N^{-1} \ln N$; see, e.g., [17, Table 1].

Although a discrete maximum norm error analysis is beyond the scope of this paper, in Table 2 we show the maximum error observed at the mesh points. One sees again that, for sufficiently small ε, the error is independent of ε and converges at a rate that is $\mathcal{O}(N^{-1})$. We remark that choosing $\sigma = 2$ in (13) improves the experimental convergence rate to $\mathcal{O}(N^{-2} \ln^2 N)$, though we do not show these results here.

Table 1 $|||u - u_h|||_\beta$ where u is as in (24) and u_h the solution of (9) on a Shishkin mesh

ε	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = 512$
1	4.303e−03	2.151e−03	1.076e−03	5.378e−04	2.689e−04
10^{-1}	2.975e−02	1.787e−02	1.043e−02	5.320e−03	2.660e−03
10^{-2}	3.322e−02	1.985e−02	1.157e−02	6.610e−03	3.718e−03
10^{-3}	3.360e−02	2.007e−02	1.169e−02	6.675e−03	3.754e−03
10^{-4}	3.358e−02	2.008e−02	1.170e−02	6.682e−03	3.758e−03
10^{-5}	3.366e−02	2.009e−02	1.170e−02	6.682e−03	3.758e−03
10^{-6}	3.366e−02	2.010e−02	1.170e−02	6.683e−03	3.758e−03
10^{-7}	3.366e−02	2.010e−02	1.170e−02	6.683e−03	3.758e−03
10^{-8}	3.366e−02	2.010e−02	1.170e−02	6.683e−03	3.758e−03

Table 2 $|||u - u_h|||_{\infty, T^{N,N}}$ for the weighted method on a Shishkin mesh

ε	$N = 32$	$N = 64$	$N = 128$	$N = 256$	$N = 512$
1	3.901e−05	9.673e−06	2.372e−06	5.699e−07	1.313e−07
10^{-1}	2.678e−03	9.603e−04	3.283e−04	8.664e−05	2.231e−05
10^{-2}	1.580e−02	4.857e−03	8.965e−04	1.549e−04	4.249e−05
10^{-3}	3.180e−02	1.445e−02	5.990e−03	2.090e−03	7.605e−04
10^{-4}	3.406e−02	1.670e−02	8.142e−03	3.912e−03	1.818e−03
10^{-5}	2.553e−02	1.398e−02	7.951e−03	4.132e−03	2.041e−03
10^{-6}	2.526e−02	1.306e−02	6.694e−03	3.413e−03	1.753e−03
10^{-7}	2.526e−02	1.306e−02	6.697e−03	3.412e−03	1.729e−03
10^{-8}	2.526e−02	1.306e−02	6.697e−e−03	3.412e−03	1.729e−03

Finally, we mention that when a solution $\tilde{u}_h$ is computed using the classical Galerkin FEM, our numerical results reveal that $|||u - \tilde{u}_h|||_\beta$ closely matches the values of $|||u - u_h|||_\beta$ stated in Table 1. This is curious, and worthy of further investigation, particularly since our experiments also suggest that when errors are measured in the discrete maximum norm, then our weighted method is more accurate than the classical Galerkin FEM by more than a factor of two.

Remark 7 For the problem of Remark 6, we have verified experimentally the accuracy of the numerical method described in that remark, using the standard SDFEM with its stabilisation parameter chosen as in [15]. Here we chose f in (21) such that the solution has parabolic layers along $y = 0$ and $y = 1$, but no exponential layer at the outflow boundary $x = 1$, so that the error in the parabolic layer dominates. Our results in the weighted norm (22) show almost first-order convergence, independently of ε.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Adler, J., MacLachlan, S., Madden, N.: A first-order system Petrov–Galerkin discretization for a reaction–diffusion problem on a fitted mesh. IMA J. Numer. Anal. **36**(3), 1281–1309 (2016). https://doi.org/10.1093/imanum/drv045
2. Adler, J.H, MacLachlan, S., Madden, N.: First-order system least squares finite-elements for singularly perturbed reaction–diffusion equations. In: Lirkov, I., Margenov, S. (eds.) Large-Scale Scientific Computing, pp. 3–14. Springer International Publishing, Cham (2020)

3. Apel, T.: Anisotropic Finite Elements: Local Estimates and Applications. Advances in Numerical Mathematics, B. G. Teubner, Stuttgart (1999)
4. Brayanov, I.A.: Numerical solution of a two-dimensional singularly perturbed reaction–diffusion problem with discontinuous coefficients. Appl. Math. Comput. **182**(1), 631–643 (2006). https://doi.org/10.1016/j.amc.2006.04.027
5. Cai, Z., Ku, J.: A dual finite element method for a singularly perturbed reaction–diffusion problem. SIAM J. Numer. Anal. **58**(3), 1654–1673 (2020). https://doi.org/10.1137/19M1264229
6. Clavero, C., Gracia, J.L., O'Riordan, E.: A parameter robust numerical method for a two dimensional reaction–diffusion problem. Math. Comput. **74**(252), 1743–1758 (2005). https://doi.org/10.1090/S0025-5718-05-01762-X
7. Evans, L.C.: Partial Differential Equations, Graduate Studies in Mathematics, vol. 19, 2nd edn. American Mathematical Society, Providence, RI (2010). https://doi.org/10.1090/gsm/019
8. de Falco, C., O'Riordan, E.: Interior layers in a reaction-diffusion equation with a discontinuous diffusion coefficient. Int. J. Numer. Anal. Model **7**(3), 444–461 (2010)
9. Franz, S., Roos, H.G.: Error estimation in a balanced norm for a convection–diffusion problem with two different boundary layers. Calcolo **51**(3), 423–440 (2014). https://doi.org/10.1007/s10092-013-0093-5
10. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order. Classics in Mathematics, Springer, Berlin, reprint of the 1998 edn. (2001)
11. Han, H., Kellogg, R.: Differentiability properties of solutions of the equation $-\epsilon^2 \Delta u + ru = f(x, y)$ in a square. SIAM J. Math. Anal. **21**, 394–408 (1990)
12. Heuer, N., Karkulik, M.: A robust DPG method for singularly perturbed reaction–diffusion problems. SIAM J. Numer. Anal. **55**(3), 1218–1242 (2017). https://doi.org/10.1137/15M1041304
13. Kopteva, N.: Maximum norm a posteriori error estimate for a 2D singularly perturbed semilinear reaction–diffusion problem. SIAM J. Numer. Anal. **46**(3), 1602–1618 (2008). https://doi.org/10.1137/060677616
14. Lin, R., Stynes, M.: A balanced finite element method for singularly perturbed reaction–diffusion problems. SIAM J. Numer. Anal. **50**(5), 2729–2743 (2012). https://doi.org/10.1137/110837784
15. Linß, T.: Anisotropic meshes and streamline-diffusion stabilization for convection-diffusion problems. Commun. Nume,r Methods Eng. **21**(10), 515–525 (2005). https://doi.org/10.1002/cnm.764
16. Linß, T.: Layer-Adapted Meshes for Reaction–Convection–Diffusion Problems. Lecture Notes in Mathematics, vol. 1985. Springer, Berlin (2010)
17. Liu, F., Madden, N., Stynes, M., Zhou, A.: A two-scale sparse grid method for a singularly perturbed reaction–diffusion problem in two dimensions. IMA J. Numer. Anal. **29**(4), 986–1007 (2009). https://doi.org/10.1093/imanum/drn048
18. Melenk, J.M., Xenophontos, C.: Robust exponential convergence of *hp*-FEM in balanced norms for singularly perturbed reaction–diffusion equations. Calcolo **53**(1), 105–132 (2016). https://doi.org/10.1007/s10092-015-0139-y
19. Rathgeber, F., Ham, D.A., Mitchell, L., Lange, M., Luporini, F., McRae, A.T.T., Bercea, G.T., Markall, G.R., Kelly, P.H.J.: Firedrake: automating the finite element method by composing abstractions. ACM Trans. Math. Softw. 43(3):24:1–24:27 (2016) https://doi.org/10.1145/2998441. http://arxiv.org/abs/1501.01809
20. Roos, H.G., Schopf, M.: Convergence and stability in balanced norms of finite element methods on Shishkin meshes for reaction–diffusion problems. Z. Angew. Math. Mech. **95**(6), 551–565 (2015). https://doi.org/10.1002/zamm.201300226
21. Roos, HG., Stynes, M., Tobiska, L.: Robust Numerical Methods for Singularly Perturbed Differential Equations, Springer Series in Computational Mathematics, Convection–Diffusion–Reaction and Flow Problems vol 24, 2nd edn. Springer, Berlin (2008)
22. Russell, S., Stynes, M.: Balanced-norm error estimates for sparse grid finite element methods applied to singularly perturbed reaction-diffusion problems. J. Numer. Math. **27**(1), 37–55 (2019). https://doi.org/10.1515/jnma-2017-0079
23. Stynes, M., Stynes, D.: Convection-diffusion problems, Graduate Studies in Mathematics, vol 196. American Mathematical Society, Providence, RI; Atlantic Association for Research in the Mathematical Sciences (AARMS), Halifax, NS, An Introduction to Their Analysis and Numerical Solution (2018)