Error analysis of the Galerkin FEM in L_2-based norms for problems with layers

On the importance, conception and realization of balancing

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von

Dipl.-Math. Martin Schopf

geboren am 16.01.1984 in Stuttgart

Gutachter : Prof. Dr. rer. nat. habil. Hans-Görg Roos
Prof. Dr. Martin Stynes

Eingereicht am: 16.01.2014

Tag der Disputation: 07.05.2014

Die Dissertation wurde unter der Betreuung von

Professor Hans-Görg Roos

im Institut für Numerische Mathematik angefertigt.

Acknowledgements

A scientific work is never the product of a single individual. I am deeply grateful for all the guidance, support and advice I received from numerous people.

I would like to gratefully and sincerely thank my supervisor Hans-Görg Roos. The time he invested into joint research, his guidance, patience and encouragement is deeply appreciated. I admire his diversified knowledge, not only with respect to numerical analysis but also regarding sports and literature. I am proud to be one of his PhD students. Moreover, he enabled me to visit various conferences and workshops through which I got to know interesting people and beautiful theories. Additionally, I want to thank him and Christian Grossmann for their inspiring lectures on numerical analysis introducing me to this wonderful mathematical field some years ago.

Further, I am grateful to Sebastian Franz, who read large parts of this thesis and gave valuable suggestions for its improvement. He was always willing to participate in discussions on its topics — not afraid of diving into technical details.

I want to thank all the members of the numerical mathematics department for creating an enjoyable, warm and familiar atmosphere — in particular the other PhD students of our group, namely Katharina Höhne, Christian Reibiger and Lars Ludwig, to whom I am also indebted for providing me with some code of his finite element framework (SSOFE) to carry out certain numerical experiments.

During the last years I have had the opportunity to meet outstanding scientists. Out of these I want to thank Natalia Kopteva, Petr Knobloch, Gunnar Matthies and Christos Xenophontos for sharing their expertise with me. I am grateful for all the discussions with Martin Stynes, who has never left any questions unanswered. Also, he introduced me to the problem discussed in this thesis at the well organized 9th annual workshop on numerical methods for problems with layer phenomena, Dublin 2011. The underlaying principle of his mixed method in Section 4.1 is embraced by the continuous interior penalty method of Section 4.2. His willingness to write a report on my thesis somewhat completes it and I want to thank him for that.

Most importantly, I want to thank my family, in particular my mother and my brother, who have always encouraged and supported me. I am grateful for the best parents-in-law one can possibly have, who took care of my little sunshine when stressful weekends had to be coped with. A very special thanks is reserved for the most important person in my life, my wife, who has always believed in me. I will always be grateful for your support, encouragement and quiet patience.

Contents

Notation

Ω	computational domain
$D \subset \Omega$	arbitrary domain
$\operatorname{diam} D$	diameter of D
$\operatorname{meas} D$	measure of the area of D
$\mathcal{L}$	differential operator
∇	gradient
Δ	Laplacian
$\frac{\partial v}{\partial x}$, v_x	partial derivative of v with respect to x
C	generic constant, independent of ε, of any mesh and of the function under consideration
$C_{M,p}$	the constant of Lemma 28 on page 62
$C^k(D)$	space of functions over D with continuous k-th order derivatives
$C^{k,\alpha}(D)$	subspace of $C^k(D)$, k-th order derivatives are Hölder-continuous with exponent α
$L_p(D)$	$p < \infty$: Lebesgue space of p-power integrable functions over D $p = \infty$: Lebesgue space of essentially bounded functions over D
$W_{k,p}(D)$	standard Sobolev space, derivatives up to order k lie in $L_p(D)$
$H^k(D)$	Sobolev space $W_{k,2}(D)$
$H_0^1(D)$	subspace of $H^1(D)$, vanishing boundary traces
$H^{1,2}(\Omega^N)$	broken Sobolev space, see (4.32) on page 92
$Q_p(D)$	space of polynomials of degree p in each variable over D
$P_p(D)$	space of polynomials of absolute degree p over D
ε	perturbation parameter
β	lower bound for convection
c_0	lower bound for reaction
$c^\star$	constant associated with the lower bound for reaction, see page 71

γ	the constant in (2.29), page 21; penalty parameter in (2.76), page 30
σ_e	penalty parameter of the CIP method associated with an edge e, see 92
σ	mesh parameter for S-type meshes
N	number of mesh cells in each coordinate direction
λ	mesh-transition point in Shishkin meshes
$\mathcal{I}$	a certain index set, see 74
$\mathcal{E}(I)$, $\mathcal{E}(II)$, $\mathcal{E}(III)$, $\mathcal{E}(IV)$,	set of interior edges of a certain class, see Definition 6 on page 75
$\sim$	equivalence of two quantities, see page 8
$\mathcal{O}(\cdot)$, $o(\cdot)$	Landau symbols
$\boldsymbol{\alpha}$, $\boldsymbol{\beta}, \boldsymbol{\gamma}$	multi-indices in $\mathbb{N}^2$
$\|\boldsymbol{\alpha}\| \coloneqq \alpha_1 + \alpha_2$	order of the multi-index $\boldsymbol{\alpha}$
$\boldsymbol{x}^{\boldsymbol{\alpha}} \coloneqq x^{\alpha_1} y^{\alpha_2}, \boldsymbol{D}^{\boldsymbol{\alpha}} \coloneqq \frac{\partial^{\alpha_1}}{x^{\alpha_1}} \frac{\partial^{\alpha_2}}{y^{\alpha_1}}$	standard multi-index notation
$\boldsymbol{P}$, $\boldsymbol{Q}$	set of multi-indices
$\boldsymbol{P}(D)$	polynomial function space associated with $\boldsymbol{P}$ over D, see page 45
$\overline{\boldsymbol{P}}$	hull of $\boldsymbol{P}$, see page 46
$H_p^{\boldsymbol{P}}(\Lambda)$	certain Sobolev function space, see page 46
$h_T \coloneqq \operatorname{diam} T$	diameter of an element T
h	$\max_T h_T$, mesh size of the mesh; small step size on a Shishkin mesh
$\boldsymbol{h}_T \coloneqq (h_{T,x}, h_{T,y})$	size vector of the axis-aligned rectangle T
H	large step size on a Shishkin mesh
h_x, h_y	sizes of a mesh rectangle T in the coordinate directions
ρ_T	radius of largest inscribed circle of the element T
$\mathcal{T}$, $\mathcal{T}_h$, Ω^N, Ω_h	triangulations of Ω
$S + E_1 + E_2 + E_3 + E_{12} + E_{23}$	decomposition of solution u, see pages 21, 23
$S + \sum_{i=1}^4 E_i + E_{12} + E_{23} + E_{34} + E_{41}$	decomposition of solution u, see page 71
Γ_1, Γ_2, Γ_3, Γ_4,	the four sides of the unitsquare, see page 21
Ω_f, Ω_f^i, Ω_f^b, Ω_c, Ω_c^i, Ω_c^b	subdomains of Ω, see page 24
Ω_{ij}, Ω_i, Ω_f	subdomains of Ω, see page 72
u^I	piecewise nodal interpolant of u
πu	approximation operator of u, sometimes the L_2-projection
Z_h	Scott-Zhang operator
Π_D	local $L_2(D)$ projection onto a polynomial space

$\hat{\cdot}$	quantity in a reference domain
$\Lambda := [-1,1]^2$	reference domain
$\hat{M}$	reference macro-element
$\mathcal{M}, \mathcal{M}_{\boldsymbol{h}}$	macro-element mesh
S_M	macro-element neighbourhood, see page 57
ω_M	macro-element patch associated with M, see Definition 5 on page 58
$F_i, F_{i,j}, F_{i,j}^{\gamma}$	(associate) functionals
$\sigma_{i,j}$	a macro-element edge on a tensor product mesh with $(x_i, y_j) \subset \sigma_{i,j}$
$\hat{\varphi}_{\pm 1}, \hat{\psi}_{\pm 1}$	Lagrangian $C^1 - P_2$ spline basis functions on $\{[-1,0],[0,1]\}$, see page 42
$\hat{\varphi}_{\pm 1,\pm 1}$, $\hat{\phi}_{\pm 1,\pm 1}$, $\hat{\chi}_{\pm 1,\pm 1}$, $\hat{\psi}_{\pm 1,\pm 1}$	Lagrangian $C^1 - Q_2$ spline basis functions on $\hat{M}$, see page 43
φ_i, ψ_i	Lagrangian $C^1 - P_2$ spline basis functions, see pages 58, 74
$\varphi_{i,j}$, $\phi_{i,j}$, $\chi_{i,j}$, $\psi_{i,j}$	Lagrangian $C^1 - Q_2$ spline basis functions on M, see page 74
ψ_i^d	dual spline basis functions, see pages 59, 74
$\hat{\Pi}$, Π	full $C^1 - Q_2$ interpolation operator, see page 44
$\hat{\Pi}^r$, Π^r	reduced $C^1 - Q_2$ interpolation operator, see page 47
$\tilde{\Pi}$	$C^1 - Q_2$ quasi-interpolation operator, see pages 57, 74
$\hat{\Pi}^x$, Π^x, $\hat{\Pi}^y$, Π^y	anisotropic interpolation operator, see (3.75) on page 67 for $\hat{\Pi}^x$, Π^x the roles of x and y are interchanged
$u^\star$	a certain projection of u on a Shishkin mesh, see pages 74–76
$(\cdot,\cdot)_D$	L_2-scalar product on D
$a(\cdot,\cdot)$, $a_N(\cdot,\cdot)$, $a_{\Gamma_0}(\cdot,\cdot)$, $B(\cdot,\cdot)$, $B^{\pm}(\cdot,\cdot)$	several bilinear forms
σ_e	penalty parameter of the CIP method associated with an edge e, see 92
$\\|\cdot\\|_{0,D}$, $\\|\cdot\\|_{L_p(D)}$	L_2- and L_p-norm on D
$\\|\cdot\\|_{k,D}$, $\|\cdot\|_{k,D}$	standard norm and semi-norm in $H^k(D)$
$\\|\cdot\\|_{\varepsilon}$	energy norm, see 6
$\|\|\|\cdot\|\|\|$	balanced CIP norm defined in (4.38) on page 93
$\\|\cdot\\|_b$	balanced norm defined in (4.5) on page 86
$\\|\cdot\\|_{b,2}$	balanced norm defined in (4.12) on page 87
$\\|\cdot\\|_{X,Y,\theta,p}$	norm in an , see 16
$[\cdot]$, $[\![\cdot]\!]$	jump across an edge

1 Introduction

In a chemical reaction several substances called reactants are transformed into one or more products. The local nature of this process on molecular level has an important implication: The concentration of a product in a certain point will highly depend on the concentration of reactants at that point. Hence, the spacial distribution of the concentrations of the substances has to be considered. Diffusion causes the substances to spread out. Without other transport phenomena this mixing process would yield a uniform concentration distribution eventually. However, there might also be a mass transfer induced by a flow in the chemical reactor. The latter transport process is called convection or advection.

In this work we are mainly interested in problems in which reaction or convection are the dominant processes. The stationary linear reaction-convection-diffusion problem

$$\mathcal{L}u := -\varepsilon u_\varepsilon''(x) + b(x)u_\varepsilon'(x) + c(x)u_\varepsilon(x) = f(x) \quad \text{in } (0,1), \qquad u_\varepsilon(0) = u_\varepsilon(1) = 0, \tag{1.1}$$

can be considered the simplest mathematical model for these and related problems. Note that the second-order derivative that models the diffusion is multiplied by a small coefficient $0 < \varepsilon \ll 1$. Such problems are very common in fluid-flow. Especially they can be interpreted as linearized Navier-Stokes equation with large Reynolds number. Moreover, they appear in mathematical models for semi-conductor devices, see [50, 64, 43] for further applications. If $b \equiv 0$ then (1.1) is of reaction-diffusion type. Generalizations of this equation have widespread applications in the modelling of many biological and chemical processes (see e.g. [36] and the references given in the first paragraph of that paper).

In general one might consider problems involving small parameters as *perturbed* problems of simpler ones also called *reduced problems* in which the small values of the parameters are replaced by zero. In case of a *regular perturbation* it is possible to study the reduced problem instead of the perturbed one because their solutions are somewhat "close" to each other. However, there are cases in which this approach fails:

For constant data (1.1) can be solved analytically. For instance, one obtains

$$u_\varepsilon(x) = x - \frac{\exp\left(-\frac{1-x}{\varepsilon}\right) - \exp\left(-\frac{1}{\varepsilon}\right)}{1 - \exp\left(-\frac{1}{\varepsilon}\right)} \tag{1.2}$$

for $b = f \equiv 1$ and $c \equiv 0$ or

$$u_\varepsilon(x) = 1 - \frac{\exp\left(-\frac{1-x}{\sqrt{\varepsilon}}\right) + \exp\left(-\frac{x}{\sqrt{\varepsilon}}\right)}{1 + \exp\left(-\frac{1}{\sqrt{\varepsilon}}\right)} \tag{1.3}$$

for $c = f \equiv 1$ and $b \equiv 0$.

If we allow the parameter ε to become small then problem (1.1) is singularly perturbed (with respect to the maximum norm): In the limiting case $\varepsilon \to 0$ the order of the differential equation is reduced. However, we still demand u_ε to satisfy both boundary conditions of (1.1). Consequently, the problem is somehow ill-posed. The simple examples (1.2) and (1.3) show that, cf. [64, 43]

$$\lim_{x \to x_b} \lim_{\varepsilon \to 0} u_\varepsilon(x) \neq \lim_{\varepsilon \to 0} \lim_{x \to x_b} u_\varepsilon(x),$$

for some boundary point $x_b \in \{0, 1\}$ (more precisely, for (1.2) we have $x_b = 1$ while for (1.3) one can not interchange the limiting operations at any boundary point). Hence, the function $(x, \varepsilon) \mapsto u_\varepsilon(x)$ has a singularity at $(x_b, 0)$. In [43] one finds a more formal definition:

Definition 1. Let B be a function space with norm $\|\cdot\|_B$. Let $D \subset \mathbb{R}^d$ be a parameter domain. The continuous function $u : D \to B,\ \varepsilon \mapsto u_\varepsilon$ is said to be *regular* for $\varepsilon \to \varepsilon^\star \in \partial D$ if there exists a function $u^\star \in B$ such that:

$$\lim_{\varepsilon \to \varepsilon^\star} \|u_\varepsilon - u^\star\|_B = 0, \tag{1.4}$$

otherwise u is said to be *singular* for $\varepsilon \to \varepsilon^\star$.

Let $(\mathcal{P}_\varepsilon)$ be a problem with solution $u_\varepsilon \in B$ for all $\varepsilon \in D$. We say $(\mathcal{P}_\varepsilon)$ is *singularly perturbed* for $\varepsilon \to \varepsilon^\star \in \partial D$ in the norm $\|\cdot\|_B$ if u is singular for $\varepsilon \to \varepsilon^\star$.

From this formal definition it can be seen that the choice of the norm is very significant. A fact that is also discussed in [24]. For instance, the constant coefficient problems associated with (1.2) and (1.3) are singularly perturbed in the maximum norm because of (1.4). However, both examples are not singularly perturbed in the L_2 norm. In fact if we consider for instance the convection-diffusion problem associated with (1.2) we find that for $u^\star = \mathrm{id} : x \mapsto x$ it holds

$$\|u_\varepsilon - u^\star\|_0 = \mathcal{O}(\varepsilon^{1/2}). \tag{1.5}$$

This function $u^\star$ is deeply connected to the solution (1.2) of the problem itself. A comparison shows that the solution (1.2) can be decomposed into the *smooth* or *regular component* $u^\star$ and a singular one. The regular component is associated with the solution of the reduced problem and the other one is often called a *boundary layer function.* For the solution (1.2) of the convection-diffusion problem considered the boundary layer function is of the form $\exp\big(-(1-x)/\varepsilon\big)$. This function and its lower-order derivatives decay exponentially and can be bounded by ε outside a neighbourhood of the point $x = 1$ with a width proportional to $\varepsilon|\ln\varepsilon|$. Hence, for a small value of ε the solution and its derivatives change rapidly near that point. One refers to these narrow regions of rapid change as *layers.* More precisely, the solution (1.2) features a *boundary layer* of width $\mathcal{O}(\varepsilon|\ln\varepsilon|)$ at $x = 1$.

In view of (1.5) we see that (1.4) is in fact a condition on the underlying boundary layer function and the norm considered: The bound (1.5) shows that the L_2 norm is too weak to capture the boundary layer function. Note that this observation is non-judgmental. An appropriate measurement of accuracy is always user-defined. For certain applications global L_2-error control may be sufficient. Moreover, some methods introduce a natural norm by themselves.

Another L_2-based norm is given by the so-called *energy norm* of problem (1.1), which is an ε-weighted H^1-norm:

$$\|v\|_\varepsilon^2 \coloneqq \varepsilon|v|_1^2 + \|v\|_0^2. \tag{1.6}$$

Generally, we write $H^k(D)$ with a positive integer k for the usual Sobolev space over the domain D with associated semi-norm $|\cdot|_{k,D}$ and norm $\|\cdot\|_{k,D}$. We follow a widespread convention and drop the symbol D from the notation for $D = \Omega$. Let $H_0^1(D)$ denote the space of Sobolev functions in $H^1(D)$ whose traces vanish on ∂D. The inner product in $L_2(D)$, denoted by $(\cdot,\cdot)_D$ and the norm $\|\cdot\|_{0,D}$ with $(v,v)_D = \|v\|_{0,D}^2$ for all $v \in L_2(D)$, are treated similarly. These definitions are extended to vector-valued functions, naturally.

The energy norm (1.6) is a natural choice for the analysis of Galerkin methods and in particular the *Galerkin finite element method* (FEM) for the approximate solution of problem (1.1).

For the boundary layer function $E(x) \coloneqq \exp\big(-(1-x)/\varepsilon\big)$ we observe $\|E\|_\varepsilon^2 = 1/2 + \mathcal{O}(\varepsilon)$. Hence, for this particular problem the energy norm captures the boundary layer function and the problem associated with (1.2) is singularly perturbed in the energy norm. We shall soon discuss the situation for reaction-diffusion problems in more detail.

Before that we turn our attention to the approximate solution of problem (1.1). For problems with layer phenomena the numerical solution is much more challenging than in the standard case. Morton says in [50] that *"Accurate modelling of the interaction between convective and diffusive processes is the most ubiquitous and challenging task in the numerical approximation of partial differential equations"*. For problems in which diffusion is the dominant process, solutions vary gradually and a lot of numerical methods have been developed and analyzed. However, unless the number of degrees of freedom is inversely dependent on (some positive power of) ε, which is impractical, these classical schemes are often inappropriate for $\varepsilon \to 0$.

For instance, these methods fail to resolve the layers. In the case of convection-diffusion the approximate solution is often polluted by huge oscillations due to stability issues rendering it practically useless.

These effects are reflected in classical convergence results for these methods in the following way: Let u_ε^h denote a family of approximate solutions computed by a classical method using equidistant meshes with element diameters h. Then an error estimate of classical type has the form

$$\|u_\varepsilon - u_\varepsilon^h\| \leq Kh^\mu,$$

where $K, \mu > 0$ are positive constants. Estimates of this type are not satisfactory in a singularly perturbed setting because the constant K depends on some norm of certain derivatives of the solution u which in turn depends on some negative power of the perturbation parameter ε. Hence, the constant K blows up for $\varepsilon \to 0$ and a fine mesh does not imply a small error. On the contrary, in certain cases mesh refinement can increase the error, initially. This fact can be observed numerically if we use for a fixed ε a standard scheme on uniform meshes and measure the maximum of the pointwise error in the discrete mesh points. These observations led to the concept of *uniform convergence* (see also [64, Subsection I.2.1.3]).

Definition 2. A discretization method (with parameter h) yielding the family of approximations u_ε^h is called *uniformly convergent* (with respect to ε) of order $\mu > 0$ in the norm $\|\cdot\|$, if there exist positive constants ε_0, C that are independent of ε and of any mesh used, such that for all sufficiently small h (independently of ε), one has

$$\sup_{0<\varepsilon\leq\varepsilon_0} \|u_\varepsilon - u_\varepsilon^h\| \leq Ch^\mu.$$

Remark 1. Throughout this work C denotes a generic positive constant that is independent of ε and of any mesh used. Moreover, for the sake of readability we shall drop the dependence of u_ε on the perturbation parameter ε and write u instead of u_ε.

Methods for which ε-uniform convergence can be proved are called *(parameter) robust*. These methods perform well regardless of how small the perturbation parameter ε is.

In this dissertation we focus on Galerkin methods. There are at least two possibilities to adjust Galerkin methods to obtain good approximations even within the (arbitrarily sharp) layer: one could consider adequate basis functions that reflect the behavior of the solution within the layer or one could choose a fine mesh there. Additionally one may stabilize the numerical method by for instance artificial diffusion (in particular streamline diffusion), local projection or interior penalty, see [64]. Note however, that this will only improve the stability of the method and has no effect on the approximation error which is a lower bound for the error, naturally.

The concept of adapting the basis function is very elaborated in the one-dimensional setting. The construction and application of *exponentially fitted splines* also called *$\overline{\mathcal{L}}$-splines* in a Galerkin method is discussed at length for instance in [64, Section I.2.2.5]. Basically, the idea is to approximate the coefficients of the operator $\mathcal{L}$ by piecewise constants over a one dimensional mesh $\{x_i\}_{i=0}^N$ giving the differential operator $\overline{\mathcal{L}}$. Next *$\overline{\mathcal{L}}$-splines* ℓ_i, $i = 1, \dots, N-1$ are constructed. The function ℓ_i is characterized by lying in the null space of the operator $\overline{\mathcal{L}}$ in every open mesh interval and the Lagrangian relation $\ell_i(i_j) = \delta_{ij}$ with the discrete Kronecker delta. Consequently, it has a support in $[x_{i-1}, x_{i+1}]$.

In the two-dimensional setting the situation is different, see [64, Section II.3.5]. Most research in this direction follows the idea of forming the tensor product of one-dimensional $\overline{\mathcal{L}}$-splines. Hence, these approaches are limited to rectangular tensor product meshes. Sacco, Gatti and Gotusso [65] give a more useful generalization of the $\overline{\mathcal{L}}$-spline concept to two dimensions using on each mesh triangle the local basis functions

$$1, \quad \mathrm{e}^{(\bar{b}_1 x + \bar{b}_2 y)/\varepsilon} \quad \text{and} \quad \bar{b}_1 x - \bar{b}_2 y.$$

Observe that all these functions lie in the null space of the operator $\overline{\mathcal{L}}$ given by

$$\overline{\mathcal{L}}v := -\varepsilon\Delta v + \bar{b}_1 v_x + \bar{b}_2 v_y.$$

The paper [66] also examines this choice of basis functions with piecewise linear test functions in a Petrov-Galerkin method, revealing a remarkable relation to a known but unusual upwinded scheme.

Still, in the two dimensional case all result known so far that use this approach are of low order. For instance, for convection-diffusion problems there seems to be no result in the literature that proves uniform convergence of an order greater than $1/2$ in a norm that is strong enough to capture the exponential layers like the (discrete) L_∞-norm or the energy norm. This is perhaps the most substantial reason why there is a general consensus that the most promising approach to solving any problem with layers (or singularities) is given by mesh adaptation based on *a posteriori* error estimation. In these algorithms a stable numerical method is used to compute an approximate solution on some coarse and standard grid, for instance an equidistant one. Based on the approximate solution some error indicator is obtained revealing where the mesh should be refined (or coarsened) in order to obtain a better-suited mesh. This process is repeated iteratively until some stopping criterion is met. Unfortunately the research in this direction for say convection-diffusion problems in more than one dimension is progressing very slowly, see [64, Section III.3.3.6], [73, Section 10.5] and [58, Section 4]. In this dissertation we shall only consider *a priori* mesh design relying on knowledge of the behaviour of the exact solution.

Let us first fix some notation and definitions. Throughout we shall only consider so-called admissible triangulations. This means that the considered domain Ω is decomposed into non-overlapping (triangular or rectangular) elements and that two distinct elements with non-empty intersection have either a common vertex or an entire common edge. Let $\mathcal{T}_h$ denote a family of triangulations of Ω. For any element $T \in \mathcal{T}_h$ we denote by h_T the diameter of T, by ρ_T the radius of its largest inscribed circle and by $|T|$ its area. Moreover, we set as usual $h := \max_{T\in\mathcal{T}_h} h_T$. We call $\mathcal{T}_h$ *shape regular* if $h_T \le C\rho_T$ for all $T \in \mathcal{T}_h$. Hence, for a shape regular mesh one has

$$h_T \le C\rho_T \le Ch_T \le Ch \quad \text{for all } T \in \mathcal{T}_h \tag{1.7}$$

and a shape regular mesh features a bounded *aspect ratio* h_T/ρ_T. The corresponding element T is called *isotropic*. For triangular elements this is equivalent to Zlámal's *minimal angle condition* [83].

In certain applications (for instance in order to use certain quasi-interpolants) it is important to have a global lower bound in (1.7) as well. A triangulation is called *quasi-uniform* if

$$h \sim h_T \sim \rho_T \quad \text{for all } T \in \mathcal{T}_h.$$

Here the notation $a \sim b$ symbolizes the equivalence of two quantities, i.e. the existence of two positive constants C_0 and C_1 independent of any mesh (and of the function under consideration) such that $C_0 b \le a \le C_1 b$. Obviously, quasi-uniformity of a family of triangulations implies its shape regularity. However, this assumption is very restrictive. For instance, a family of locally refined meshes generated by some adaptive algorithm might not be quasi-uniform. We call a shape regular family of triangulations *locally uniform* if

$$h_T \sim h_{T'} \quad \text{for all } T, T' \in \mathcal{T}_h \text{ such that } T \cap T' \neq \emptyset.$$

As examples for locally uniform families of triangulations used in the approximation of functions with boundary layers let us mention the graded meshes of [27] and [22, 23] or the modifications of standard layer-adapted meshes in [62].

In general a layer-adapted mesh is characterized by condensing grid points within the layer. We shall now present some frequently used layer-adapted meshes and comment on error estimates obtained on these. The interested reader is referred to [43] for details. The presentation is borrowed from [58, 43] and [64].

The idea of adjusting the mesh to the boundary layers for a reaction-diffusion problem was introduced by Bakhvalov [7] in 1969. However, it is easy to extend this idea to other layers of known structure. Let us consider a layer problem over the domain $(0, 1)$. Within the layer, say near $x = 0$, Bakhvalov proposed to use the function $t = q(1 - E(x))$ — a mirrored and scaled boundary layer function E — to map an equidistant t-grid back to the x-axis. More precisely, the grid points x_i are defined by

$$q\big(1 - E(x_i)\big) := q\big(1 - \mathrm{e}^{-\frac{\beta x_i}{\sigma\varepsilon}}\big) = \frac{i}{N} =: t_i \qquad \text{for } i = 0, 1, \ldots, \tau N.$$

The parameter β is intrinsic to the layer problem. It is associated with the coefficients of the differential operator involved. The other parameters are user chosen: $q \in (0,1)$ determines the ratio of the number of mesh points within the layer to N (which is the overall number of mesh intervals). The other parameter $\sigma > 0$ controls the grading. At a transition point τ (in the t-domain) the *Bakhvalov mesh* changes from a graded to a uniform one, i.e. $x_i = \varphi(t_i)$ for $i = 0, 1, \dots, N$ with the *mesh generating function*

$$\varphi(t) = \begin{cases} \chi(t) := -\dfrac{\sigma\varepsilon}{\beta} \ln \dfrac{q-t}{q} & \text{for } t \in [0, \tau], \\ \chi(\tau) + \big(1 - \chi(\tau)\big) \dfrac{t-\tau}{1-\tau} & \text{for } t \in [\tau, 1]. \end{cases}$$

Note that neither the Bakhvalov mesh nor the other following standard layer-adapted meshes are locally uniform. The transition point τ is chosen in such a way that $\varphi \in C^1[0,1]$ with $\varphi(1) = 1$. It can be obtained efficiently by solving the non-linear equation

$$\chi'(\tau) = \frac{1 - \chi(\tau)}{1 - \tau}$$

with a fixed-point iteration (and a start value $\tau = 0$). Note that this equation is not solvable if $\sigma\varepsilon \geq \beta q$. In this case it is common practice to switch to a uniform mesh with N mesh intervals. For the transition point $\chi(\tau)$ with respect to the original x-domain one can show that

$$\chi(\tau) \in \left(\frac{\sigma\varepsilon}{\beta} \ln \Big(\frac{\beta}{\sigma\varepsilon} q \Big), \frac{\sigma\varepsilon}{\beta} \ln \Big(\frac{\beta}{\sigma\varepsilon} \frac{1}{1-q} \Big) \right).$$

Consequently, the explicit definition $\chi_B := \frac{\gamma\varepsilon}{\beta} |\ln \varepsilon|$, which gives $\exp(-\beta x/\varepsilon) \leq \varepsilon^\gamma$ for $x \geq \chi_B$, can be considered a reasonable alternative. Note that for this choice of a transition point it is also possible to bound derivatives of the boundary layer function ε-uniformly in points away from the layer. Meshes using this transition point and an adequate grading within the layer are referred to as *Bakhvalov-type meshes*, i.e. the mesh generating function is given (with $q = 1/2$ for simplicity) by

$$\varphi(t) = \begin{cases} -\dfrac{\sigma\varepsilon}{\beta} \ln \big(1 - 2(1-\varepsilon)t\big) & \text{for } t \in [0, 1/2], \\ 1 - 2(1-\chi_B)(1-t) & \text{for } t \in [1/2, 1]. \end{cases}$$

A *Shishkin-type* mesh is characterized by $\chi_S := \min\{q, \frac{\gamma\varepsilon}{\beta} \ln N\}$. Hence, (for $\chi_S \leq q$ which is practically the only relevant case since otherwise N is exponentially large relative to ε^{-1} and uniform meshes can be used)

$$\exp(-\beta x/\varepsilon) \leq N^{-\gamma} \quad \text{for } x \geq \chi_S,$$

which is sufficient to facilitate ε-uniform error estimates. Consequently, γ is typically chosen equal to the formal order of the method (though sometimes larger values are assumed). The probably most-studied mesh of this class is the *Shishkin mesh* (see e.g. [49]) which is particularly simple because it is piecewise uniform. It is constructed by dividing the intervals $[0, \chi_S]$ and $[\chi_S, 1]$ into qN and $(1-q)N$ equidistant mesh intervals, respectively. Note that a Shishkin mesh does not fully resolve the layer: the derivative of a boundary layer function is still large on part of the first coarse mesh interval.

In two dimensions frequently-studied layer-adapted meshes are tensor products of the one-dimensional meshes considered over rectangular domains under the assumption that only boundary layers and so-called corner layers are present (— for more general layer-adapted meshes see [35], [47] and [64, Remarks III.3.121 and III.3.123]). In these tensor products meshes we observe elements with arbitrarily high aspect ratio h_T/ρ_T. If there is no ε-uniform bound for this ratio of an element T, it is referred to as *anisotropic element*. We call a mesh anisotropic if the underlaying family of meshes contains anisotropic elements.

The performance of the linear or bilinear Galerkin FEM over two dimensional Shishkin meshes of tensor product type (draw diagonals into each mesh rectangle to obtain a Shishkin triangulation into triangles) is well understood for convection-diffusion problems with only exponential layers. For the error of the Galerkin FEM in the energy norm the result

$$\|u - u^N\|_\varepsilon \leq C N^{-1} \ln N \tag{1.8}$$

was established in [74, 16]. See, [59] for similar results over the more general Shishkin-type meshes. The Galerkin FEM over Bakhvalov-type meshes is considered in [60]. It is shown that

$$\|u - u^N\|_\varepsilon \le CN^{-1}Q(\varepsilon, N)$$

with some function Q that depends on ε and N very mildly, i.e. $Q(\varepsilon, N) \le \sqrt{\ln 10}$ for $N \ge 10$ and $\varepsilon \ge 10^{-100}$.

For convection-diffusion problems with characteristic layers Roos [57] proves the validity of (1.8) for the error of the Galerkin FEM on appropriately constructed Shishkin meshes. The generalization to Shishkin-type meshes can be found in [43, Theorem 9.27].

Most results are known for reaction-diffusion problems. For instance, in the energy norm it holds

$$\|u - u^N\|_\varepsilon \le C\bigl(\varepsilon^{1/4}N^{-1}\ln N + N^{-2}\bigr)$$

on appropriately constructed Shishkin meshes, see [16, 64]. It is possible to extend this result to higher order elements (see [2] and [39]) if the solution features sufficient regularity. In this case results proving exponential convergence for the *hp*-FEM [48] over *hp* meshes have been obtained, too.

All the anisotropic layer-adapted meshes mentioned feature small mesh sizes in the direction in which the solution changes rapidly. In the perpendicular direction larger mesh sizes are used. In order to prove that the approximation error benefits from this setting one relies on so-called *anisotropic interpolation error estimates*: Let for instance T be a rectangular element with sides aligned to the x- and y-coordinate axes and of size h_x and h_y, respectively. Then for $v \in H^2(T)$ the bilinear nodal interpolant v^I that interpolates v in the four vertices of T exists and the error of bilinear nodal interpolation satisfies (see [3, 2]):

$$\bigl\|(v - v^I)_x\bigr\|_{0,T} \le C\bigl(h_x\|v_{xx}\|_{0,T} + h_y\|v_{xy}\|_{0,T}\bigr). \tag{1.9}$$

Note that this estimate holds true independently of the aspect ratio of the element T. Consequently, a small mesh size h_x can compensate for a large second derivative of v with respect to x. We shall prove similar but new results in Chapter 3.

In the introduction of [2, page 13] Apel states that *"anisotropic mesh refinement offers a great potential for the construction of efficient numerical procedures, more efficient than it is possible with the restriction to a bounded aspect ratio. So one can expect a broad utilization of such meshes"*.

However, anisotropic meshes have certain drawbacks as well. For instance, the analysis of certain quasi-interpolants is much more elaborated in the case of locally uniform meshes, cp. [62, 2]. Moreover, there is not much known with respect to anisotropic a posteriori mesh refinement which can be challenging even from a practical point of view (for instance for a problem with a curved interior layer). Finally, it is well-known that the usage of anisotropic meshes in a Galerkin FEM leads to very high condition numbers of the system matrices of the resulting linear systems and it is challenging to realize multi-grid or multi-level methods on these usually non-nested families of meshes, see [46].

Let us now come back to the question whether reaction-diffusion problems are singularly perturbed in the energy norm. Let E denote one of the boundary layer functions given by $x \mapsto \exp\bigl(-(1-x)/\sqrt{\varepsilon}\bigr)$ and $x \mapsto \exp\bigl(-x/\sqrt{\varepsilon}\bigr)$. If we measure E in the energy norm a short calculation gives

$$\|E\|_\varepsilon = \mathcal{O}(\varepsilon^{1/4}). \tag{1.10}$$

Hence, the energy norm fails to capture the boundary layer functions for reaction-diffusion problems. In view of Definition 1 reaction-diffusion problems are not singularly perturbed with respect to the energy norm. This observation raises two questions:

1. Does the Galerkin finite element method on standard meshes yield satisfactory approximations for the reaction-diffusion problem with respect to the energy norm?

2. Is it possible to strengthen the energy norm in such a way that the boundary layers are captured and that it can be reconciled with a robust finite element method, i.e. robust with respect to this strong norm?

The first question is interesting not only from a mathematical point of view:

- Any adaptive method that fails to resolve the layers of a reaction-diffusion problem in an initial phase will feature similar error reduction in the energy norm until the mesh is sufficiently layer-adapted.
- In certain applications the energy norm might be a feasible measurement of accuracy. Then our results show that standard methods can be applied. One may use the freedom gained in the choice of the mesh by adapting it to other goals like in the dual weighted residuals approach (in contrast to resolving the layers for stability reasons).
- Whenever resolving the layer in a reaction-diffusion problem is of the essence our results demonstrate that the energy norm is not adequate. This observation questions the quality of certain methods and the significance of the corresponding error estimates in a wide variety of papers.

In Chapter 2 we answer the first question. We show that Galerkin finite element approximation u_h converges ε-uniformly in the energy norm to the solution of the reaction-diffusion problem on standard shape regular meshes with mesh size h. More precisely, we show in Section 2.1 that the error of the linear or bilinear Galerkin FEM satisfies

$$\|u - u_h\|_\varepsilon \le Ch^{1/2}, \tag{1.11}$$

given certain regularity conditions on the data of the problem over a polygonal and convex domain Ω. For instance, (1.11) holds true for $f/c \in H^1(\Omega)$ (see Theorem 5) which has been published in [63]. Extending a technique of Wahlbin and Schatz [67] we prove in Theorem 4 that $\nabla c \in L_\infty(\Omega)$ and $f \in H^{1/2,\infty}$ (which is a certain intermediate Sobolev space) is also sufficient for (1.11). These assumptions allow the presence of interior layers generated by jump discontinuities of f along Lipschitz curves. Numerical experiments (presented in Subsection 2.4) indicate that these estimates are sharp.

These results are completely new in two dimensions. So far a similar estimate has only been known for the local discontinuous Galerkin least-squares method over $\Omega = (0,1)^2$ assuming that $u \in H^3(\Omega)$, see [41].

As already mentioned these results are possible because the energy norm fails to capture the layers. Consequently, one can prove similar ε-uniform error estimates for the Galerkin FEM on standard meshes applied to problems with relatively (with respect to the norm) weak layers. For instance, Leykekhman [37] considers a reaction-diffusion problem with homogeneous Neumann boundary conditions (2.26) which leads to the formation of very weak layers. In that paper he proves almost first order convergence in the global maximum norm.

In Section 2.2 we extend this idea and consider a two dimensional convection-diffusion problem posed in the unit square with characteristic layers and a Neumann outflow condition. Hence, the exponential boundary layer is weakened. We follow [63] and show that the error of the linear or bilinear Galerkin FEM on a shape regular mesh with mesh size h satisfies

$$\|u - u_h\|_\varepsilon \le Ch^{1/3},$$

under reasonable assumptions on the solution u, see Theorem 7.

Finally, Section 2.3 deals with a convection-diffusion problem in $\Omega = (0,1)^2$ with a strong exponential layer and weaker characteristic layers. Assuming a solution decomposition, the Galerkin FEM is proven to be robust on a family of meshes with N^2 elements that is only adapted with respect to the strong exponential layer and neglects the presence of the characteristic layers. This result extends [63, Section 3.2, Theorem 1] by proving that the bilinear Galerkin finite element approximation u^N satisfies

$$\|u - u^N\|_\varepsilon \le CN^{-1/2}. \tag{1.12}$$

In particular, in contrast to the result in [63] no relation of ε and N is required to conclude ε-uniform convergence of order 1/2. Inspecting the proof of (1.12) one finds that imposing the Dirichlet boundary conditions along the characteristic boundary of Ω in a weak sense might be beneficial. We show in Theorem 15 that this is indeed the case if ε is very small in comparison to N^{-1}.

Chapter 3 is concerned with approximation theory. Let V^N be a finite element space of piecewise polynomials over some triangulation Ω^N of Ω and let $\{\varphi_i\}_{i=1}^N$ be a basis of V^N. One can obtain an approximation $\pi v \in V^N$ of a sufficiently smooth function v over Ω by

$$\pi v(x) = \sum_{i=1}^{N} F_i(v)\varphi_i(x),$$

where F_i, $i = 1, \ldots, N$ are linear functionals. If $F_i(v)$ is simply the evaluation of (a certain derivative of) v at some point x_i (or a linear combination of such), then we call this functional *local.* As an example consider nodal interpolation over Lagrange elements, where $F_j(v) := v(x_j)$ and $\varphi_i(x_j) = \delta_{ij}$ for all mesh points x_j of Ω^N. Consequently, $\pi v(x_j) = v(x_j)$.

In contrast, the well-known Clément operator, in which the linear functionals are given by averaging operators, is not local. More precisely if $\{\varphi_i\}_{i=1}^N$ is again the Lagrangian basis of V^N, then $F_j(v)$ is defined by some projection of v over the patch of elements that are adjacent to x_j. The Clément operator is called a *quasi-interpolation operator* because in general one only has $F_j(v) \approx v(x_j)$.

As already mentioned the analysis of numerical methods over anisotropic meshes for layer problems relies critically on so-called anisotropic (quasi-)interpolation error estimates, like (1.9). In Chapter 3 we extend the theory of [2] in the following ways:

- We consider (quasi-)interpolation operators defined on macro-elements. This means that the (quasi-)interpolant of a function v is defined by means of linear functionals F_i on a certain patch of neighbouring elements (— the macro-element). Alternatively, we can point out the main difference by considering (quasi-)interpolation on a single macro-element: Then the value of $F_i(v)$ is based on values of v within the considered macro-element but piecewise polynomial basis functions φ_i have to be used.

- The results of [3] are limited to interpolation operators that are based solely on function values of the function v to be interpolated. In [2] additionally the Scott-Zhang interpolation operator and certain modifications of it are considered. However, there are no results for (quasi-)interpolation operators that process derivatives of v, though the developed theory is capable of dealing with such operators of *Hermite-type.* The problem appears to be that it is hard to determine the so-called *associated functionals*, that are needed within the theoretical framework. We show in Chapter 3 that for rectangular elements that are aligned to the coordinate axes these are given by two dimensional divided differences. This observation makes it possible to obtain anisotropic interpolation error estimates for C^1 elements over tensor product meshes.

Summarizing, a general theory for obtaining anisotropic interpolation error estimates for macro-element interpolation is developed, revealing general construction principles. We apply this theory to interpolation operators on a macro-type of biquadratic C^1 finite elements on rectangular tensor product grids. The resulting macro-element can be viewed as a rectangular version of the C^1 Powell-Sabin element.

This theory also shows how interpolation on the C^1 Bogner-Fox-Schmidt finite element space (or higher order generalizations) can be analyzed in a unified framework. Moreover, we discuss a modification of Scott-Zhang type that (processes derivatives of v and) gives optimal error estimates under the regularity required without imposing quasi-uniformity on the family of macro-element meshes used. Finally, we introduce and analyze an anisotropic macro-element interpolation operator, which is the tensor product of one-dimensional $C^1 - P_2$ macro-interpolation and P_2 Lagrange interpolation.

These results are used to approximate the solution of a singularly perturbed reaction-diffusion problem on a Shishkin mesh that features highly anisotropic elements. Hereby we obtain an approximation whose normal derivative is continuous along certain edges of the mesh, enabling a more sophisticated analysis of a continuous interior penalty (CIP) method in Chapter 4.

We dedicate Chapter 4 to the task of devising finite element methods for reaction-diffusion problems that are robust with respect to a norm that strengthens the energy norm and is able to capture the arising boundary layers. We proceed as follows:

In Section 4.1 the ideas of [42] are sketched. This paper was the first to deal with the particular problem of designing a finite element method for which error estimates in a better suited so-called *balanced* norm could be proven. Also, it coined this expression.

We adapt the main idea of that paper in Section 4.2 to propose and analyze a new C^0 interior penalty method that features improved stability properties in comparison with the Galerkin FEM. For the latter balanced norm results are obtained in Section 4.3. In that section we also examine a supercloseness property of the Galerkin finite element approximation and comment on error estimates in the L_2- and the maximum-norm.

At the end of Chapter 4 we supply numerical experiments, give a brief summary and mention further work in this field of research.

2 Galerkin FEM error estimation in weak norms

2.1 Reaction-diffusion problems

In the introduction we have seen from a 1D example that the reaction-diffusion problem is not singularly perturbed in the energy norm in the sense of Definition 1: The energy norm fails to capture the layers, raising the question whether or not it is possible to prove uniform convergence for the Galerkin finite element method in this weak norm without fitting the mesh to the boundary layers.

In the one-dimensional setting an answer to this question can be found in [6, p. 121]:

Theorem 1. *Let $u \in H_0^1(0,1) \cap H^2(0,1)$ be the unique solution of*

$$\begin{aligned} -\varepsilon u'' + cu &= f \quad \textit{in } (0,1), \\ u(0) &= u(1) = 0, \end{aligned}$$

where $0 < \varepsilon \ll 1$, $c, f \in H^2(0,1)$ and $c \geq c_0 > 0$. Let V^h denote the space of piecewise linear basis functions on a uniform mesh with mesh size h. Then the solution $u^h \in V^h$ of the Galerkin finite element method satisfies the uniform estimate

$$\|u - u^h\|_\varepsilon := \left(\varepsilon|u - u^h|_1^2 + |u - u^h|_0^2\right)^{1/2} \leq Ch^{1/2}.$$

Proof. The proof relies on the quasi-optimality property of the Galerkin error with respect to the energy norm:

$$\|u - u^h\|_\varepsilon \leq C \inf_{v^h \in V^h} \|u - v^h\|_\varepsilon, \tag{2.1}$$

which follows from the fact that the bilinear form associated with the Galerkin method is an inner product on $H_0^1(\Omega)$ which induces a norm that is equivalent to the energy norm. In a standard trick the infimum in the right hand side of (2.1) is estimated by choosing $v^h = u^I$ the Lagrange interpolant of u. Consequently, the following interpolation error estimates and a priori information of u are very helpful:

$$|u - u^I|_1 \leq Ch|u|_2 \leq Ch\varepsilon^{-3/4}, \qquad \|u - u^I\|_0 \leq Ch|u|_1 \leq Ch\varepsilon^{-1/4}, \tag{2.2}$$

$$|u - u^I|_1 \leq C|u|_1 \leq C\varepsilon^{-1/4}, \qquad \|u - u^I\|_0 \leq Ch^{1/2}\|u\|_{L_\infty(\Omega)} \leq Ch^{1/2}. \tag{2.3}$$

For the simple proof consider the cases $h \leq \sqrt{\varepsilon}$ and $h \geq \sqrt{\varepsilon}$ separately and apply (2.2) and (2.3), respectively. □

We see that indeed the Galerkin finite element method converges ε-uniformly on a uniform mesh.

Remark 2. The assumptions $c, f \in H^2(0,1)$ in Theorem 1 were required in [6] to prove the needed a priori estimates on u. Even if one assumes their validity in the two-dimensional case the arguments of the proof do not extend to 2D because the Lagrange interpolant is not H^1 stable in contrast to (2.3) in the one-dimensional case.

In higher dimensions this question appears to be unanswered until recently [63]. We shall shortly give elegant proofs for the two dimensional case that also prove the 1D case reducing the regularity of the data required. Some of the results presented in this chapter have been published in [63].

Let us consider the reaction-diffusion problem

$$-\varepsilon\Delta u + cu = f \quad \text{in } \Omega \subset \mathbb{R}^2, \tag{2.4a}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{2.4b}$$

where Ω is polygonal and convex, $0 < \varepsilon \ll 1$ and $f \in L_2(\Omega)$, $c \in C(\overline{\Omega})$ with $c \geq c_0 > 0$. Under these assumptions problem (2.4) has a unique solution $u \in H_0^1(\Omega) \cap H^2(\Omega)$ which is characterized by sharp boundary layers near $\partial\Omega$.

For the definition of trial and test space for the Galerkin finite element method we introduce a shape-regular mesh $\mathcal{T}_h$ and let $h := \max_{T\in\mathcal{T}_h} \operatorname{diam} T$ denote the maximal mesh size. Let $V^h \subset H_0^1(\Omega)$ denote the space of linear or bilinear elements over the given mesh. The Galerkin finite element approximation $u^h \in V^h$ is then determined as the solution of

$$\varepsilon(\nabla u^h, \nabla v^h) + (cu^h, v^h) = (f, v^h) \quad \text{for all } v^h \in V^h. \tag{2.5}$$

The interesting paper [67] deals mainly with deriving local and global L_∞-error estimates for the Galerkin finite element method (2.5) for problem (2.4). We shall come back to them later. However, Schatz and Wahlbin [67] also prove an uniform error estimate in the $L_2(\Omega)$-norm:

Theorem 2. *Assume $f \in H^{1/2,\infty}(\Omega)$ and $\nabla c \in L_\infty(\Omega)$. Then the error of the linear or bilinear Galerkin finite element method on a shape regular mesh satisfies*

$$\|u - u^h\|_0 \leq Ch^{1/2}. \tag{2.6}$$

In order to supply a deeper understanding of the first assumption and the proof we shall shorty sketch the K-interpolation method for two admissible Banach spaces X, Y (i.e. there is a Hausdorff topological vector space Z, such that there are continuous embeddings $X, Y \hookrightarrow Z$). Note that in this case $X \cap Y$ normed by $\|v\|_{X\cap Y} := \max\{\|v\|_X, \|v\|_Y\}$ and $X + Y$ normed by

$$\|v\|_{X+Y} := \inf_{\substack{v=x+y\\ x\in X,\ y\in Y}} \|x\|_X + \|y\|_Y$$

are Banach spaces and it is possible to define intermediate spaces V such that $X \cap Y \hookrightarrow V \hookrightarrow X + Y$. More precisely for $t > 0$ and $v \in X + Y$ let

$$K(t, v) := \inf_{\substack{v=x+y\\ x\in X,\ y\in Y}} \|x\|_X + t\|y\|_Y.$$

For $\theta \in (0, 1)$, $1 \leq p \leq \infty$ define the interpolation space $(X, Y)_{\theta,p}$ as

$$(X, Y)_{\theta,p} := \{v \in X + Y \ :\ \|v\|_{X,Y,\theta,p} < \infty\},$$

where

$$\|v\|_{X,Y,\theta,p} = \begin{cases} \left(\displaystyle\int_0^\infty \left(t^{-\theta}K(t, v)\right)^p \frac{\mathrm{d}t}{t}\right)^{1/p} & \text{for } 1 \leq p < \infty, \\ \displaystyle\sup_{t>0} t^{-\theta}K(t, v) & \text{for } p = \infty. \end{cases}$$

Remark that for any choice of θ and p the normed space $(X, Y)_{\theta,p}$ is complete. For more information on interpolation spaces we refer the interested reader to [8] or [77].

With this definitions we can specify the space $H^{1/2,\infty}(\Omega) := (L_2(\Omega), H_0^1(\Omega))_{1/2,\infty}$ which is ideally suited for the data of problem (2.4):

- It allows the data to have jump discontinuities along Lipschitz curves in Ω.
- In contrast to the Hilbertian interpolation space $H_{00}^{1/2}(\Omega) := (L_2(\Omega), H_0^1(\Omega))_{1/2,2}$ or H_0^1 the space $H^{1/2,\infty}(\Omega)$ does not demand boundary conditions. This fact can nicely be seen from the result $H^1(\Omega) \hookrightarrow H^{1/2,\infty}(\Omega)$, see [67].

The estimate (2.6) now follows by interpolating the corresponding results for $f \in L_2(\Omega)$ and $f \in H_0^1(\Omega)$. We shall sketch the technique.

Proof of Theorem 2. It is possible to prove

$$\|u - u^h\|_0 \le C\|f\|_0 \tag{2.7a}$$

and if $c \in W_{1,\infty}(\Omega)$ and $f \in H_0^1(\Omega)$

$$\|u - u^h\|_0 \le Ch\|f\|_1, \tag{2.7b}$$

see [67]. The latter first order estimate is not surprising because the last assumption precludes u from having boundary layers due to the fact that the solution f/c of the reduced problem to (2.4) already fulfills the homogeneous boundary conditions and no boundary layer function correcting an incompatibility is needed.

Now let $f \in H^{1/2,\infty}(\Omega)$. Based on the construction of that space $f = f_0 + f_1$ can be decomposed into components $f_0 \in L_2(\Omega)$ and $f_1 \in H_0^1(\Omega)$. By linearity of the involved differential operator this decomposition carries over to the solution of (2.4): $u = u_0 + u_1$, where u_i are the solutions of the problems (2.4) in which f is replaced by f_i. Similarly, such a decomposition can be given for the Galerkin finite element approximation: $u^h = u_0^h + u_1^h$. Hence, by (2.7)

$$\|u - u^h\|_0 \le \|u_0 - u_0^h\|_0 + \|u_1 - u_1^h\|_0 \le C\big(\|f_0\|_0 + h\|f_1\|_1\big).$$

Switching to the infimum of all possible decompositions of $f = f_0 + f_1$ one arrives at

$$\|u - u^h\|_0 \le CK(h,f) = Ch^{1/2}h^{-1/2}K(h,f) \le Ch^{1/2}\|f\|_{L_2(\Omega),H_0^1(\Omega),1/2,\infty}. \qquad \square$$

It turns out that (2.6) holds also true if the L_2-norm of the error is replaced by the stronger energy norm. We shall prove this in Theorem 4. Before that we need the following lemma.

Lemma 3. *The solution u of (2.4) satisfies the following a priori estimates:*

$$\|u\|_\varepsilon \le C\|f\|_0, \tag{2.8}$$

if $f \in H_0^1(\Omega)$ and $c \in W_{1,\infty}(\Omega)$, then

$$\varepsilon^{1/2}|u|_2 + \|u\|_1 \le C\|f\|_1, \tag{2.9}$$

if $f \in H^1(\Omega) \cap L_\infty(\Omega)$ and $c \in W_{1,\infty}(\Omega)$, then

$$\varepsilon^{3/4}|u|_2 + \varepsilon^{1/4}|u|_1 + \|u\|_0 \le C\big(\varepsilon^{1/4}|f|_1 + \|f\|_0 + \|f\|_{L_\infty(\Omega)}\big), \tag{2.10}$$

if $u_0 := f/c \in H^1(\Omega)$, then

$$\varepsilon^{3/4}|u|_2 + \varepsilon^{1/4}|u|_1 + \varepsilon^{-1/4}\|u - u_0\|_0 \le C\big(\varepsilon^{1/4}|u_0|_1 + \|u_0\|_{0,\partial\Omega}\big). \tag{2.11}$$

Proof. The bound (2.8) follows by multiplying (2.4a) with u and applying integration by parts.

Similarly, testing the differential equation with $-\Delta u$ integrating over Ω and using integration by parts as well as $u, f \in H_0^1(\Omega)$ one obtains

$$\varepsilon\|\Delta u\|_0^2 + (c\nabla u, \nabla u) + (u\nabla c, \nabla u) = (\nabla f, \nabla u).$$

Hence,

$$\varepsilon\|\Delta u\|_0^2 + c_0|u|_1^2 \le \|\nabla f - u\nabla c\|_0|u|_1 \le \frac{c_0}{2}|u|_1^2 + \frac{1}{2c_0}\|\nabla f - u\nabla c\|_0^2.$$

Since $|u|_2 \le C\|\Delta u\|_0$ for the considered domain one obtains

$$\varepsilon|u|_2^2 + \frac{c_0}{2}|u|_1^2 \le \frac{1}{2c_0}\big(|f|_1 + \|\nabla c\|_{L_\infty(\Omega)}\|u\|_0\big)^2.$$

By (2.8) the stability estimate (2.9) follows.

The result (2.10) is taken from [42, Section 2] where it is proven rigorously.

For (2.11) we follow [63] and rewrite the differential equation (2.4a) to

$$-\varepsilon\Delta u + c(u-u_0) = 0. \tag{2.12}$$

This implies

$$|u|_2 \le C\varepsilon^{-1}\|u-u_0\|_0. \tag{2.13}$$

Multiplying (2.12) with $u-u_0$ integrating over Ω and applying integration by parts yields

$$\varepsilon\big(\nabla u, \nabla(u-u_0)\big) + \big(c(u-u_0), u-u_0\big) = -\varepsilon\Big(\frac{\partial u}{\partial n}, u_0\Big)_{\partial\Omega}$$

Hence,

$$\varepsilon|u|_1^2 + c_0\|u-u_0\|_0^2 \le -\varepsilon\Big(\frac{\partial u}{\partial n}, u_0\Big)_{\partial\Omega} + \varepsilon(\nabla u, \nabla u_0). \tag{2.14}$$

The fist term on the right hand side is estimated using Young's inequality and a trace inequality:

$$\begin{aligned}\left|\varepsilon\Big(\frac{\partial u}{\partial n}, u_0\Big)_{\partial\Omega}\right| &\le \frac{\alpha\sqrt{\varepsilon}}{2}\|u_0\|_{0,\partial\Omega}^2 + \frac{C\varepsilon^{3/2}}{2\alpha}|u|_1|u|_2\\ &\le \frac{\alpha\sqrt{\varepsilon}}{2}\|u_0\|_{0,\partial\Omega}^2 + \frac{C}{2\alpha}\left(\frac{\beta}{2}\varepsilon|u|_1^2 + \frac{1}{2\beta}\varepsilon^2|u|_2^2\right)\\ &\le \frac{\alpha\sqrt{\varepsilon}}{2}\|u_0\|_{0,\partial\Omega}^2 + \frac{C\beta}{4\alpha}\varepsilon|u|_1^2 + \frac{C}{4\alpha\beta}\|u-u_0\|^2.\end{aligned} \tag{2.15}$$

The last inequality is due to (2.13). Clearly, one can choose $\alpha > 0$ and $\beta > 0$ independently of ε in such a way that the sum of the last two terms can be bounded by $\varepsilon|u|_1^2/4 + c_0\|u-u_0\|_0^2/2$. For the last term of (2.14) we proceed likewise:

$$|\varepsilon(\nabla u, \nabla u_0)| \le \varepsilon|u|_1|u_0|_1 \le \frac{\varepsilon}{4}|u|_1^2 + \varepsilon|u_0|_1^2. \tag{2.16}$$

Collecting (2.14), (2.15) and (2.16) we obtain

$$\varepsilon|u|_1^2 + \|u-u_0\|_0^2 \le C\big(\varepsilon|u_0|_1^2 + \sqrt{\varepsilon}\|u_0\|_{0,\partial\Omega}^2\big) \tag{2.17}$$

Combining (2.13) and (2.17) completes the proof. □

Remark 3. As already mentioned in the proof of Theorem 2 the assumption $f \in H_0^1(\Omega)$ precludes u from having boundary layers. This is why (2.9) is possible.

Theorem 4. *Assume $f \in H^{1/2,\infty}(\Omega)$ and $\nabla c \in L_\infty(\Omega)$. Then the error of the linear or bilinear Galerkin finite element method on a shape regular mesh satisfies*

$$\|u-u^h\|_\varepsilon \le Ch^{1/2}. \tag{2.18}$$

Proof. For the error of the Galerkin finite element method for problem (2.4) with $f \in L_2(\Omega)$ one can improve (2.7a) to

$$\|u-u^h\|_\varepsilon \le C\|f\|_0. \tag{2.19}$$

Let us prove this. We consider the case $h \le \sqrt{\varepsilon}$ first. Let $\pi u \in V^h \subset H_0^1$ denote the L_2 projection which is H^1-stable on the given shape regular triangulation.

$$\begin{aligned}\|u-u^h\|_\varepsilon &\le C\|u-\pi u\|_\varepsilon = C\big(\sqrt{\varepsilon}|u-\pi u|_1 + \|u-\pi u\|_0\big) \le C\big(\sqrt{\varepsilon}\|u\|_1 + h\|u\|_1\big)\\ &\le C\sqrt{\varepsilon}\|u\|_1 \le C\|u\|_\varepsilon \le C\|f\|_0,\end{aligned}$$

where we also used (2.8).

It remains to study the case $\sqrt{\varepsilon} \le h$. With the help of the H^1-stability of π and an inverse estimate we obtain

$$|u-u^h|_1 \le |u-\pi u|_1 + |\pi u - u^h|_1 \le C\big(|u|_1 + h^{-1}\|\pi u - u^h\|_0\big).$$

Hence, using (2.7a), $\sqrt{\varepsilon} \le h$ and the stability estimate $\|u\|_\varepsilon \le C\|f\|_0$

$$\|u-u^h\|_\varepsilon \le C\big(\sqrt{\varepsilon}|u-u^h|_1 + \|u-u^h\|_0\big) \le C\big(\sqrt{\varepsilon}|u|_1 + (1+\sqrt{\varepsilon}h^{-1})\|u-u^h\|_0\big) \le C\|f\|_0.$$

Next we consider problem (2.4) with $f \in H_0^1(\Omega)$: Using the quasi-optimality property and the a priori estimate (2.9) one obtains

$$\begin{aligned}\|u-u^h\|_\varepsilon &\le C\|u-u^I\|_\varepsilon \le C\big(\varepsilon^{1/2}|u-u^I|_1 + \|u-u^I\|_0\big)\\ &\le Ch\big(\varepsilon^{1/2}|u|_2 + |u|_1\big) \le Ch\|f\|_1.\end{aligned} \tag{2.20}$$

The estimate (2.18) now follows by interpolating (2.19) and (2.20) in complete analogy to the proof of Theorem 2. □

As already mentioned Theorem 4 allows the right hand side f to be discontinuous along Lipschitz curves which will cause the presence of interior layers, see Subsection 2.4.2 for a numerical experiment with an example problem. Also the demanded regularity of $f \in H^{1/2,\infty}(\Omega)$ fits nicely with the half convergence order obtained. In fact

$$\|u\|_{L_2(\Omega),H_0^1(\Omega),1/2,\infty} \le C\|f\|_{L_2(\Omega),H_0^1(\Omega),1/2,\infty} \le C,$$

for $c \in W_{1,\infty}(\Omega)$ and choosing $\theta > 1/2$ breaks $\|u\|_{L_2(\Omega),H_0^1(\Omega),\theta,\infty} \le C$ due to the boundary layers regardless of the regularity of f, see [67].

Nevertheless the assumption $c \in W_{1,\infty}(\Omega)$ of Theorem 4 is unsatisfactory. We shall now present a result from [63] that side-steps this assumption allowing a rougher coefficient c.

Theorem 5. *Let the solution $u_0 := f/c$ of the reduced problem to* (2.4) *satisfy the estimate $|u_0|_1 + \|u_0\|_{0,\partial\Omega} \le C$. Then for the Galerkin finite element method on a shape-regular mesh of linear or bilinear elements with mesh size h the ε-uniform error estimate* (2.18) *holds true.*

Proof. Based on (2.1) it remains to estimate the approximation error. We bound the latter by a projection error. To this end, let $\pi u \in V^h$ denote an approximation of u governed by an L_2- and H^1-stable operator π. On a shape regular mesh one might choose the L_2 projection or the Clément quasi-interpolant.

We start off with the error in the L_2-norm:

$$\|u-\pi u\|_0^2 \le C\left(\|u_0-\pi u_0\|_0^2 + \|u-u_0-\pi(u-u_0)\|_0^2\right) \tag{2.21}$$

In contrast to $|u|_1$ the solution u_0 of the reduced problem is bounded uniformly: $|u_0|_1 \le C$. Hence, the first summand on the right hand side of (2.21) can be estimated in a classical manner. For the other summand we use once the same well-known estimate and once the L_2 stability.

$$\|u-\pi u\|_0^2 \le Ch^2|u_0|_1^2 + Ch|u-u_0|_1\|u-u_0\|_0. \tag{2.22}$$

By (2.11) one has $\|u-u_0\|_0 \le C\big(\varepsilon^{1/2}|u_0|_1 + \varepsilon^{1/4}\|u_0\|_{0,\partial\Omega}\big) \le C\varepsilon^{1/4}$ and

$$|u-u_0|_1 \le |u|_1 + |u_0|_1 \le 2|u_0|_1 + \varepsilon^{-1/4}\|u_0\|_{0,\partial\Omega} \le C\varepsilon^{-1/4}.$$

Hence, we can conclude from (2.22) that

$$\|u-\pi u\|_0^2 \le Ch^2 + Ch\varepsilon^{-1/4}\varepsilon^{1/4} \le Ch. \tag{2.23}$$

Similarly, using (2.11) again

$$\begin{aligned}\varepsilon|u-\pi u|_1^2 &= \varepsilon|u-\pi u|_1|u-\pi u|_1\\ &\le C\varepsilon h|u|_2|u|_1 \le C\varepsilon h\varepsilon^{-3/4}\varepsilon^{-1/4} \le Ch.\end{aligned} \tag{2.24}$$

Collect (2.23) and (2.24) to complete the proof. □

Remark 4. By an inverse estimate and (2.11) one can conclude that

$$\sqrt{\varepsilon}|u-u^h|_1 \le C\varepsilon^{1/4} + \sqrt{\varepsilon}|u^h|_1 \le C\varepsilon^{1/4} + C\frac{\sqrt{\varepsilon}}{h}.$$

Hence, for $h \gg \sqrt{\varepsilon}$ the error in the energy norm is dominated by the $L_2(\Omega)$ error. This fact is nicely reflected in Figure 2.2 depicting the results of numerical experiments.

The estimate of Remark 4 as well as others in this section use that the boundary layer is not captured by the energy norm as already mentioned in the introduction, see (1.10). Consequently, the situation is entirely different if stronger norms are considered. In [67], Schatz and Wahlbin consider the maximum norm and prove the quasi-optimality result

$$\|u - u^h\|_{L_\infty(\Omega)} \leq \ln(C + \varepsilon/h) \min_{v^h \in V^h} \|u - v^h\|_{L_\infty(\Omega)} \tag{2.25}$$

for the linear Galerkin finite element method on a quasi-uniform family of triangulations and state that *for certain "higher order" element spaces, the logarithmic factor can be replaced with a constant* C, [67, page 48]. Moreover, they derive a localized version of (2.25) from which uniform almost second order convergence in the maximum norm over certain interior subdomains can be concluded.

However, the Galerkin finite element method on a quasi-uniform mesh cannot converge uniformly with respect to ε in the $L_\infty(\Omega)$ norm, because even the best approximation with respect to this norm fails to approximate an arbitrarily sharp layer on these kind of meshes. This conjecture can also be found in [67, page 51].

Remark 5. It is unlikely that the situation improves if the class of triangulations is enlarged by allowing locally uniform triangulations. For instance, simple upwinding on a locally uniform family of meshes can not yield uniform convergence in the discrete maximum norm for the convection-diffusion problem, see [64, Remark I.2.85, page 121]. See also [62] for the difficulties of obtaining uniform error estimates of the Galerkin finite element method on locally uniform meshes for convection-diffusion problems in the energy norm.

Leykekhman [37] considers the case of homogeneous Neumann boundary conditions, more precisely

$$-\varepsilon\Delta u + u = f \quad \text{in } \Omega \subset \mathbb{R}^2, \tag{2.26a}$$

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \partial\Omega, \tag{2.26b}$$

where Ω is a domain in $\mathbb{R}^n$, $n \geq 2$, with smooth boundary, $0 < \varepsilon \leq 1$ and $f \in L_2(\Omega)$. In this problem the boundary layers are of a weaker nature, *less pronounced*, see [67, Page 49]. Using the techniques of [67] he obtains for the linear Galerkin finite element error on quasi-uniform meshes an estimate similarly to (2.25) but in which the best approximation is measured in some weighted $L_\infty(\Omega)$ norm. Like in [67] this result is then localized yielding almost optimal order interior convergence. Moreover, assuming $f \in W_{1,\infty}(\Omega)$ he shows in Corollary 2.3 that

$$\|u - u^h\|_{L_\infty(\Omega)} \leq C(\ln h)^3 \min\{h^2/\varepsilon, h\} \|f\|_{W_{1,\infty}(\Omega)}. \tag{2.27}$$

We see that the Galerkin finite element method for problem (2.26) on a quasi-uniform mesh is almost first order ε-uniformly convergent in the global maximum norm. Hence, we may conclude from this result that the maximum norm is too weak to capture the boundary layers arising in problem (2.26). This is of course due to the very weak nature of the layers. In fact, in one space dimension the solution structure is again well-understood. For $\Omega = (0,1)$ the boundary layer functions behave like $\sqrt{\varepsilon}\exp(-c^\star x/\sqrt{\varepsilon})$ and $\sqrt{\varepsilon}\exp(-c^\star(1-x)/\sqrt{\varepsilon})$. For a simple and elegant proof of this fact, see [51, page 178]. We see that even first-order derivatives remain bounded in the maximum norm for $\varepsilon \to 0$. Consequently, regarding Definition 1 this Neumann problem is not singularly perturbed with respect to the maximum norm.

Note that this observation is in stark contrast to the general statement in [24, Page 7] that the need for a priori information of the solution can be avoided if the maximum norm is used.

2.2 A convection-diffusion problem with weak characteristic layers and a Neumann outflow condition

We have seen at the end of the previous section that Neumann boundary conditions lead to the formation of weaker layers in comparison to Dirichlet boundary conditions. This is theoretically clear based on the theory of matched asymptotic expansion, see e.g. [64, Page 16] concerning a one-dimensional convection-diffusion problem.

In two space dimensions the situation is similar. The paper [52] considers a convection-diffusion problem on the unit square with constant coefficients. The convection is aligned to two edges of the domain leading to the formation of two characteristic layers along these edges. The problem is posed with a Neumann boundary condition along the outflow boundary and Dirichlet conditions on the remaining three sides, where the data is allowed to be incompatible to a certain degree. The paper sheds light on the regularity of the solution. Moreover, a solution decomposition is obtained yielding pointwise bounds on derivatives of the solution and revealing the dependence of the solution on the small diffusion coefficient ε.

In this section we shall consider a similar problem, following [63]:

$$-\varepsilon\Delta u - bu_x + cu = f \quad \text{in } \Omega := (0,1)^2, \tag{2.28a}$$

$$\left.\frac{\partial u}{\partial x}\right|_{x=0} = 0, \quad u|_{x=1} = 0 \quad \text{and} \quad u\big|_{y=0} = u\big|_{y=1} = 0, \tag{2.28b}$$

where $0 < \varepsilon \ll 1$, $b \in W^{1,\infty}(\Omega)$, $b \geq \beta > 0$ with some constant β, $c \in L_\infty(\Omega)$, f is sufficiently smooth and

$$c + \frac{1}{2}b_x \geq \gamma > 0. \tag{2.29}$$

Let us introduce a notation for the sides of Ω:

$$\begin{aligned} \Gamma_1 &= \{(x,0) \,:\, 0 \leq x \leq 1\}, \qquad & \Gamma_2 &= \{(0,y) \,:\, 0 \leq y \leq 1\}, \\ \Gamma_3 &= \{(x,1) \,:\, 0 \leq x \leq 1\}, \qquad & \Gamma_4 &= \{(1,y) \,:\, 0 \leq y \leq 1\}. \end{aligned} \tag{2.30}$$

Under the assumptions made one can easily show that the bilinear form

$$a(w,v) := \varepsilon(\nabla w, \nabla v) + (-bw_x + cw, v) \tag{2.31}$$

is coercive in $V = \{v \in H^1(\Omega) \,:\, v|_{\Gamma_i} = 0,\ i = 1,3,4 \text{ in the sense of traces}\}$ with respect to the energy norm. Hence, by continuity of $a(\cdot,\cdot)$ and the well-known Lax-Milgram Lemma, problem (2.28) is uniquely solvable in V. The solution u possesses characteristic boundary layers of width $\mathcal{O}(\sqrt{\varepsilon}\ln(1/\varepsilon))$ along the *characteristic boundary* $\Gamma_1 \cup \Gamma_3$ that is parallel to the direction of convection. Additionally, the solution possesses a weak exponential layer at the outflow boundary Γ_2 where u is required to satisfy homogeneous Neumann boundary conditions.

In the corners of the domain Ω derivatives of u are unbounded, in general. One refers to the solution components that cause this phenomenon as corner singularities. We shall neglect these, assuming that the right hand side f allows for the following solution decomposition to hold true:

$$u = S + E_1 + E_2 + E_3 + E_{12} + E_{23}, \tag{2.32a}$$

where for all $(x,y) \in \Omega$ and $0 \leq i + j \leq 2$ we have the pointwise estimates

$$\begin{aligned} &\left|\frac{\partial^{i+j}S}{\partial x^i \partial y^j}(x,y)\right| \leq C, & &\left|\frac{\partial^{i+j}E_1}{\partial x^i \partial y^j}(x,y)\right| \leq C\varepsilon^{-j/2}\mathrm{e}^{-y/\sqrt{\varepsilon}}, \\ &\left|\frac{\partial^{i+j}E_2}{\partial x^i \partial y^j}(x,y)\right| \leq C\varepsilon^{1-i}\mathrm{e}^{-\beta x/\varepsilon}, & &\left|\frac{\partial^{i+j}E_{12}}{\partial x^i \partial y^j}(x,y)\right| \leq C\varepsilon^{-(i+j/2-1)}\mathrm{e}^{-\beta x/\varepsilon}\mathrm{e}^{-y/\sqrt{\varepsilon}} \end{aligned} \tag{2.32b}$$

and bounds for E_3 and E_{23} similarly to E_1 and E_{12}, respectively. In (2.32a) the term S corresponds to the well-behaved regular solution component of u. The components E_1 and E_3 contain the characteristic layers along Γ_1 and Γ_3. Similarly, E_2 covers the weak exponential layer near Γ_2. Finally, E_{ij} contains the corner layer that arises at the intersection of Γ_i and Γ_j, $(i,j) \in \{(1,2),(2,3)\}$.

Remark 6. In the case of constant coefficients we could use [52] to specify local compatibility conditions on f at the vertices of Ω such that the bounds (2.32b) can be proven to hold true. Remark that the pointwise bounds assumed here are stronger than needed for the presented error analysis. In fact we only need the estimates (2.33) which bound certain $L_2(\Omega)$-based norms.

The solution decomposition (2.32) implies the following a priori estimates.

Corollary 6. *Assume* (2.32) *and set* $E := u - S$. *Then* u *and the layer components* E *satisfy:*

$$\begin{aligned} &|S|_1 \le C, && |E|_1 \le C\varepsilon^{-1/4}, && |u|_2 \le C\varepsilon^{-3/4}, \\ &\|E\|_0 \le C\varepsilon^{1/4}, && \big\|E_x\big\|_0 \le C\varepsilon^{1/4}, && \big|E_x\big|_1 \le C\varepsilon^{-1/2}. \end{aligned} \tag{2.33}$$

As in the previous Section, let V^h denote the space of linear or bilinear finite elements on a shape-regular mesh. Then the Galerkin approximation $u^h \in V^h$ is governed by solving the linear system

$$a(u^h, v^h) = (f, v^h) \quad \text{for all } v^h \in V^h.$$

Remark that by coercivity of $a(\cdot,\cdot)$ the approximate solution u^h is well defined.

Theorem 7. *Let* (2.33) *hold true. Then the error of the finite element method with linear or bilinear finite elements on a shape-regular mesh satisfies the* ε*-uniform estimate*

$$\|u - u_h\|_\varepsilon \le Ch^{1/3}.$$

Proof. Like in the proof of Theorem 5 let π denote an L_2- and H^1-stable operator giving the approximation $\pi u \in V^h$ of u, for instance the L_2 projection. We can estimate the approximation error as in the proof of Theorem 5:

$$\|u - \pi u\|_0^2 \le C\big(\|S - \pi S\|_0^2 + \|E - \pi E\|_0^2\big) \le C\big(h^2|S|_1^2 + h|E|_1\|E\|_0\big).$$

Hence, by (2.33) the bound (2.23) follows. Similarly, (2.33) yields the estimate (2.24). Consequently, one obtains

$$\|u - \pi u\|_\varepsilon \le Ch^{1/2}. \tag{2.34}$$

It remains to estimate the discrete error component $\pi u - u^h \in V^h$. A standard argument using coercivity and Galerkin orthogonality yields

$$\|\pi u - u^h\|_\varepsilon^2 \le Ca(\pi u - u^h, \pi u - u^h) = Ca(\pi u - u, \pi u - u^h).$$

By the definition of $a(\cdot,\cdot)$ and the Cauchy-Schwarz inequality

$$\begin{aligned} \|\pi u - u^h\|_\varepsilon^2 &\le C\Big(\|\pi u - u\|_\varepsilon \|\pi u - u^h\|_\varepsilon + \big|\big(b(\pi u - u)_x, \pi u - u^h\big)\big|\Big) \\ &\le C\big(\|\pi u - u\|_\varepsilon + \|(\pi u - u)_x\|_0\big)\|\pi u - u^h\|_\varepsilon. \end{aligned} \tag{2.35}$$

Hence, we need to estimate $\|(\pi u - u)_x\|_0$.

$$\begin{aligned} \|(u - \pi u)_x\|_0^3 &\le C\big(\|(S - \pi S)_x\|_0^3 + \|(E - \pi E)_x\|_0^3\big) \le Ch^3 + \|E_x\|_0^2\big\|(E - \pi E)_x\big\|_0 \\ &\le Ch^3 + C\varepsilon^{1/2}h|E_x|_1 \le Ch^3 + Ch. \end{aligned} \tag{2.36}$$

Collect (2.34), (2.35) and (2.36) to complete the proof. □

Remark 7. Numerically we observe ε-uniform convergence of order 0.5. Clearly, Theorem 7 could be improved if one could devise a L_2- and H^1-stable projection πu of u with better approximation properties with respect to $\|(u - \pi u)_x\|_0$.

Remark 8. Note that we only dealt with homogeneous boundary data. The inhomogeneous case is more difficult, in particular Shishkin's obstacle theorem applies (see [72] for a similar result and note that this phenomenon is not caused by the parabolic nature of the problem considered therein). This negative result should be kept in mind whenever a problem whose solution exhibits characteristic boundary layers is discretized on an equidistant mesh. It states that it is impossible to devise a difference scheme (with coefficients drawn from a fixed class of functions) on an equidistant mesh that features uniform convergence in the discrete L_∞-norm inside the characteristic boundary layers without excluding certain classes of (smooth and compatible) boundary data. Remember that (besides the homogeneous boundary data) our result for the Galerkin FEM relies on the weakness of the energy norm; it fails to capture the characteristic layers.

2.3 A mesh that resolves only part of the exponential layer and neglects the weaker characteristic layers

We have seen in Sections 2.1 and 2.2 that the Galerkin finite element solution on standard meshes converges ε-uniformly to the solution u of certain problems in norms that fail to capture the layers of u.

However, in other problems the solution u might contain *several kinds of layers* and the norm considered may fail to capture only *some* of them. With respect to the other results of this chapter it is interesting to know if in such a scenario the Galerkin finite element method is robust on a mesh fitted to the stronger layers only. Besides this theoretical point of view this question is interesting:

- An adaptive algorithm for these singularly perturbed problems based on the Galerkin FEM and a posteriori error estimation may fail to refine the characteristic layers initially due to their weaker nature. Can one still expect a meaningful approximation?
- Problems and costs concerning remeshing of the domain may be avoided. For instance, imagine a weak interior layer moving through the domain in a singularly perturbed parabolic problem.

As a model problem we consider

$$\begin{aligned} -\varepsilon\Delta u - bu_x + cu &= f \quad \text{in } \Omega = (0,1)^2, \\ u|_{\partial\Omega} &= 0, \end{aligned} \tag{2.37}$$

assuming $0 < \varepsilon \ll 1$, $b \in W^{1,\infty}(\Omega)$, $b \geq \beta > 0$ with some constant β, $c \in L_\infty(\Omega)$, f is sufficiently smooth and

$$c + \frac{1}{2}b_x \geq \gamma > 0. \tag{2.38}$$

As in the previous section defined in (2.30), let Γ_i denote the sides of Ω. Problem (2.37) has a unique solution u that is characterized by an exponential boundary layer near Γ_2. Moreover, u possesses two characteristic layers near Γ_1 and Γ_3. We shall shortly see that the energy norm captures the exponential layer but fails to do so for the characteristic layers which must be therefore of a weaker nature.

Again we assume a solution decomposition:

$$u = S + E_1 + E_2 + E_3 + E_{12} + E_{23}, \tag{2.39a}$$

where for all $(x,y) \in \Omega$ and $0 \leq i+j \leq 2$ we have the pointwise estimates

$$\begin{aligned} \left|\frac{\partial^{i+j} S}{\partial x^i \partial y^j}(x,y)\right| &\leq C, & \left|\frac{\partial^{i+j} E_1}{\partial x^i \partial y^j}(x,y)\right| &\leq C\varepsilon^{-j/2}\mathrm{e}^{-y/\sqrt{\varepsilon}}, \\ \left|\frac{\partial^{i+j} E_2}{\partial x^i \partial y^j}(x,y)\right| &\leq C\varepsilon^{-i}\mathrm{e}^{-\beta x/\varepsilon}, & \left|\frac{\partial^{i+j} E_{12}}{\partial x^i \partial y^j}(x,y)\right| &\leq C\varepsilon^{-(i+j/2)}\mathrm{e}^{-\beta x/\varepsilon}\mathrm{e}^{-y/\sqrt{\varepsilon}} \end{aligned} \tag{2.39b}$$

and bounds for E_3 and E_{23} similarly to E_1 and E_{12}, respectively.

Remark 9. From [32, 33] one can deduce compatibility and smoothness conditions on f such that (2.39) holds true in the constant coefficient case.

Measuring the boundary layer component E_2 in the energy norm, we see $\|E\|_\varepsilon = \mathcal{O}(1)$. Hence, problem (2.39) is singularly perturbed in the energy norm in view of Definition 1. However, for the weaker characteristic layer components E_1 and E_3 a short calculation gives $\|E_1\|_\varepsilon + \|E_3\|_\varepsilon = \mathcal{O}(\varepsilon^{1/4})$. Consequently, the energy norm is too weak to capture the characteristic layers.

On a mesh sequence that is fitted to all layers one can achieve uniform first order convergence for the bilinear Galerkin finite element method in the energy norm. For instance, on appropriately constructed Shishkin meshes with N^2 elements the error satisfies

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

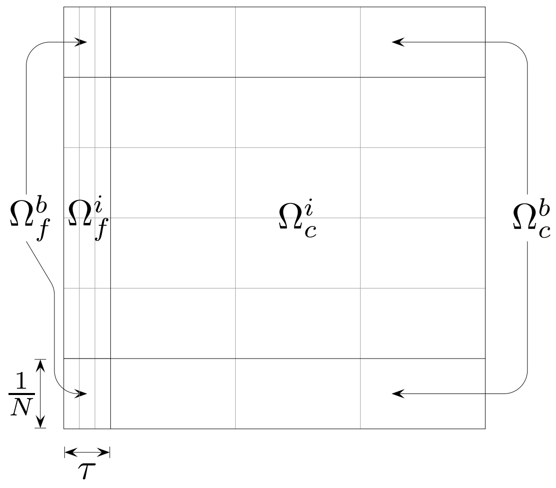

Figure 2.1: Domain decomposition, $N = 6$.

However, the scope of this chapter is different. As motivated in the beginning of this section we introduce a mesh that is fitted to the stronger exponential layer but we ignore the characteristic layers. More precisely, we use a Shishkin mesh. Let N be an even positive integer which will be used to denote the number of mesh intervals in each coordinate direction. Define the transition point

$$\tau := \min\left\{\frac{1}{2}, \frac{2\varepsilon}{\beta}\ln N\right\}.$$

A one-dimensional Shishkin mesh is piecewise equidistant, i.e. over $(0,1)$ it subdivides each of the intervals $(0,\tau)$ and $(\tau,1)$ into $N/2$ intervals giving $N+1$ grid points, i.e. for $\tau < 1/2$:

$$x_i := \begin{cases} ih & \text{for} \quad i = 0,1,\ldots,\dfrac{N}{2} \quad \text{with} \quad h = \dfrac{4\varepsilon}{\beta}N^{-1}\ln N, \\[2ex] \tau + \left(i - \dfrac{N}{2}\right)H & \text{for} \quad i = \dfrac{N}{2}+1,\ldots,N \quad \text{with} \quad H = (1-\tau)\dfrac{2}{N}. \end{cases} \tag{2.40}$$

The mesh is fine near $x = 0$: In $(0,\tau)$ the mesh size is $h = \mathcal{O}(\varepsilon N^{-1}\ln N)$. To the right of τ the size H of the mesh intervals is equivalent to N^{-1}.

We introduce a rectangular mesh Ω^N over Ω by forming the tensor product of the described Shishkin mesh in the x-direction with a uniform mesh (giving the grid points $y_j = jN^{-1}$ for $j = 0,\ldots,N$) in the y-direction.

For our subsequent error analysis we partition Ω as follows, see Figure 2.1:

$$\overline{\Omega} = \overline{\Omega_f^i \cup \Omega_f^b \cup \Omega_c^i \cup \Omega_c^b},$$

where

$$\begin{aligned} \Omega_f &:= (0,\tau)\times(0,1), & \Omega_f^i &:= (0,\tau)\times(y_1,y_{N-1}), & \Omega_f^b &:= (0,\tau)\times\big((0,y_1)\cup(y_{N-1},1)\big), \\ \Omega_c &:= (\tau,1)\times(0,1), & \Omega_c^i &:= (\tau,1)\times(y_1,y_{N-1}), & \Omega_c^b &:= (\tau,1)\times\big((0,y_1)\cup(y_{N-1},1)\big). \end{aligned}$$

Note that Ω_f contains all the *anisotropic elements*, which means that the aspect ratio of these mesh rectangles is unbounded for $\varepsilon \to 0$. In contrast to this the mesh is quasi-uniform in Ω_c. The subscripts f and c refer to the mesh sizes in x-direction which are *fine* in Ω_f and *coarse* elsewhere. Ω_f^b denotes the first and last ply of elements along Γ_1 and Γ_3, i.e. close to the characteristic *boundary*. $\Omega_f^i = \Omega_f \setminus \Omega_f^b$ is the remainder of Ω_f. The symbols Ω_c^b and Ω_c^i have a similar meaning. These subdomains will be useful in the estimation of the characteristic layer component, later on. See Figure 2.1 for some illustration of these definitions and the mesh Ω^6.

In order to define the numerical method let $V^N \subset H_0^1(\Omega)$ denote the space of piecewise bilinears over the mesh Ω_N. The bilinear Galerkin finite element method now reads: Find $u^N \in V^N$ such that the linear system

$$a(u^N, v^N) = (f, v^N) \quad \text{for all } v^N \in V^N \tag{2.41}$$

with $a(\cdot,\cdot)$ from (2.31) is satisfied. Note that u^N is well defined due to the coercivity of the bilinear form in the energy norm with respect to V^N.

If the mesh evolves into a very fine (or even uniform) one, i.e. $N^{-1} \le C\varepsilon$, then we could analyze the Galerkin finite element method in a conventional manner. Note that in this case all the layers — in particular the characteristic layers of width $\mathcal{O}\big(\sqrt{\varepsilon}\ln(1/\varepsilon)\big)$ — are resolved by the mesh. It remains to study $\varepsilon \le CN^{-1}$.

Next, let us consider an intermediate case in which the mesh is layer-adapted, more precisely

$$N^{-1} \le \sqrt{\varepsilon} \le CN^{-1/2}. \tag{2.42}$$

Consequently, we also have

$$\varepsilon^{1/2} \le C(\ln N)^{-2}. \tag{2.43}$$

Lemma 8. *Assuming* (2.42) *the Lagrange interpolation error on a Shishkin mesh with bilinear elements and N intervals in each coordinate direction that is only fitted to the exponential boundary layer satisfies*

$$\|u - u^I\|_\varepsilon \le CN^{-1/2}. \tag{2.44}$$

Proof. We use the solution decomposition (2.39) to estimate the the interpolation error by splitting u into components and applying a triangle inequality. The interpolation error of the regular solution component is easily bounded

$$\|S - S^I\|_0 \le CH^2|S|_2 \le CN^{-2} \qquad \text{and} \qquad |S - S^I|_1 \le CH|S|_2 \le CN^{-1}. \tag{2.45}$$

Since the mesh is fitted to the exponential boundary layer component E_2 one can use standard arguments [64, See e.g. III.3.5.2] and (2.43) to obtain

$$\|E_2 - E_2^I\|_0 \le CN^{-2}, \qquad \varepsilon^{1/2}|E_2 - E_2^I|_1 \le CN^{-1}\ln N,$$

yet,

$$\|E_2 - E_2^I\|_{L_\infty(\Omega)} \le CN^{-2}\ln N^2. \tag{2.46}$$

For the characteristic layer components, for instance for E_1, we use (2.39b) and (2.42) to obtain

$$\|E_1 - E_1^I\|_0 \le CH|E_1|_1 \le C\varepsilon^{-1/4}N^{-1} \le CN^{-1/2}. \tag{2.47}$$

Similarly,

$$\|(E_1 - E_1^I)_x\|_0 \le CH\big(\|(E_1)_{xx}\|_0 + \|(E_1)_{xy}\|_0\big) \le C\varepsilon^{-1/4}N^{-1} \le CN^{-1/2}, \tag{2.48}$$

$$\varepsilon^{1/2}\|(E_1 - E_1^I)_y\|_0 \le C\varepsilon^{1/2}H\big(\|(E_1)_{xy}\|_0 + \|(E_1)_{yy}\|_0\big) \le C\varepsilon^{-1/4}N^{-1} \le CN^{-1/2}., \tag{2.49}$$

Finally, we consider a corner layer, for instance E_{12}. In the subdomain Ω_f the mesh rectangles T of size $\boldsymbol{h} := (h_x, h_y)$ are anisotropic. We use the well-known bilinear anisotropic interpolation error estimates [3, 2]

$$\|v - v^I\|_{0,T} \le C \sum_{|\boldsymbol{\alpha}|=m} \boldsymbol{h}^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}}v\|_{0,T} \quad \text{for } m = 1, 2,$$
$$\|\boldsymbol{D}^{\boldsymbol{\gamma}}(v - v^I)\|_{0,T} \le C \sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{h}^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}}v\|_{0,T} \quad \text{for } |\boldsymbol{\gamma}| = 1.$$

Hence, by (2.43),

$$\|E_{12} - E_{12}^I\|_{0,\Omega_f} \le C\big(h\|(E_{12})_x\|_{0,\Omega_f} + H\|(E_{12})_y\|_{0,\Omega_f}\big) \le C\big(h\varepsilon^{-1/4} + H\varepsilon^{1/4}\big) \le C\varepsilon^{1/4}N^{-1}. \tag{2.50}$$

Similarly, and by (2.42)

$$\varepsilon^{1/2}\|(E_{12}-E_{12}^I)_x\|_{0,\Omega_f} \le C\varepsilon^{1/2}\big(h\|(E_{12})_{xx}\|_{0,\Omega_f}+H\|(E_{12})_{xy}\|_{0,\Omega_f}\big)$$
$$\le C\varepsilon^{1/2}\big(\varepsilon N^{-1}\ln N\varepsilon^{-5/4}+N^{-1}\varepsilon^{-3/4}\big)\le C\varepsilon^{-1/4}N^{-1}\le CN^{-1/2}, \tag{2.51}$$

$$\begin{aligned}\varepsilon^{1/2}\|(E_{12}-E_{12}^I)_y\|_{0,\Omega_f} &\le C\varepsilon^{1/2}\big(h\|(E_{12})_{xy}\|_{0,\Omega_f}+H\|(E_{12})_{yy}\|_{0,\Omega_f}\big)\\ &\le C\varepsilon^{1/2}\big(\varepsilon N^{-1}\ln N\varepsilon^{-3/4}+N^{-1}\varepsilon^{-1/4}\big)\le C\varepsilon^{1/4}N^{-1}.\end{aligned} \tag{2.52}$$

In the remainder of the domain we have $x\ge\tau=2\varepsilon/\beta\ln N$. Consequently, (2.39b) yields pointwise smallness of the corner layer, $\|E_{12}\|_{L_\infty(\Omega_c)}\le CN^{-2}$. By maximum-norm stability of Lagrange interpolation one obtains

$$\|E_{12}-E_{12}^I\|_{0,\Omega_c}\le\|E_{12}-E_{12}^I\|_{L_\infty(\Omega_c)}\le C\|E_{12}\|_{L_\infty(\Omega_c)}\le CN^{-2}. \tag{2.53}$$

Eventually, we use an inverse estimate in Ω_c where the mesh is quasi-uniform with mesh size $H\sim N^{-1}$:

$$\begin{aligned}\varepsilon^{1/2}|E_{12}-E_{12}^I|_{1,\Omega_c} &\le \varepsilon^{1/2}\big(|E_{12}|_{1,\Omega_c}+|E_{12}^I|_{1,\Omega_c}\big)\le C\varepsilon^{1/2}\big(\varepsilon^{-1/4}N^{-2}+H^{-1}\|E_{12}^I\|_{0,\Omega_c}\big)\\ &\le C\big(\varepsilon^{1/4}N^{-2}+N\|E_{12}^I\|_{L_\infty(\Omega_c)}\big)\le CN^{-1}.\end{aligned} \tag{2.54}$$

Summarizing (2.45)–(2.54) we get $\|u-u^I\|_\varepsilon\le C(N^{-1/2}+N^{-1}\ln N)\le CN^{-1/2}$. □

Theorem 9. *Let (2.42) be satisfied. Using the described Shishkin mesh that is only fitted to the exponential boundary layer with N intervals in each coordinate direction the error of the Galerkin finite element method with bilinear elements satisfies*

$$\|u-u^N\|_\varepsilon\le CN^{-1/2}.$$

Proof. Based on Lemma 8 and a triangle inequality it remains to estimate the discrete error component $\xi:=u^I-u^N\in V^N$. Coercivity of the bilinear form with respect to the energy norm and Galerkin orthogonality yield

$$C\|\xi\|_\varepsilon^2\le a(\xi,\xi)\le a(u^I-u,\xi)=\varepsilon\big(\nabla(u^I-u),\nabla\xi\big)-\big(b(u^I-u)_x,\xi\big)+\big(c(u^I-u),\xi\big).$$

Let $E:=E_1+E_3$ denote the sum of the characteristic layer components of u. We introduce the following splitting $u^I-u=E^I-E+(u-E)^I-(u-E)$ and use integration by parts for all terms except for $E-E^I$ to obtain

$$\begin{aligned}C\|\xi\|_\varepsilon^2 &\le \varepsilon\big(\nabla(u^I-u),\nabla\xi\big)-\big(b(E^I-E)_x,\xi\big)+\big(c(E^I-E),\xi\big)\\ &\quad+\Big(b\big((u-E)^I-(u-E)\big),\xi_x\Big)+\Big((b_x+c)\big((u-E)^I-(u-E)\big),\xi\Big).\end{aligned} \tag{2.55}$$

Note that $\xi\in V^N$ implies $\xi\equiv0$ on $x=0$ and $x=1$. Hence, we may choose to apply integration by parts to any component of u separately without introducing terms on the boundary.

All summands but the fourth one of the right hand side in (2.55) are easily estimated using the Cauchy-Schwarz inequality. Collecting (2.44), (2.48) and (2.47) gives

$$\begin{aligned}\|\xi\|_\varepsilon^2 &\le C\big(\|(u-u^I)\|_\varepsilon+\|(E-E^I)_x\|_0+\|E-E^I\|_0\big)\|\xi\|_\varepsilon+C\Big|\Big(b\big((u-E)^I-(u-E)\big),\xi_x\Big)\Big|\\ &\le CN^{-1/2}\|\xi\|_\varepsilon+C\Big|\Big(b\big((u-E)^I-(u-E)\big),\xi_x\Big)\Big|.\end{aligned} \tag{2.56}$$

The last term is split into components and is sometimes estimated in Ω_f and Ω_c separately.

For the regular solution component S we use the Cauchy-Schwarz inequality, (2.45) and (2.42):

$$\big|\big(b(S^I-S),\xi_x\big)\big|\le C\|S-S^I\|_0\|\xi_x\|_0\le CN^{-2}\|\xi_x\|_0\le CN^{-1}\|\xi\|_\varepsilon. \tag{2.57}$$

For the exponential boundary layer component the Hölder inequality and (2.46) yield in Ω_f

$$\begin{aligned}\big|\big(b(E_2^I-E_2),\xi_x\big)_{\Omega_f}\big| &\le C\,\mathrm{meas}(\Omega_f)^{1/2}\|E_2-E_2^I\|_{L_\infty(\Omega_f)}\|\xi_x\|_{0,\Omega_f}\\ &\le \varepsilon^{1/2}N^{-2}(\ln N)^{5/2}\|\xi_x\|_0\le CN^{-1/2}\|\xi\|_\varepsilon.\end{aligned} \tag{2.58}$$

In Ω_f we estimate the corner layer components using the Cauchy-Schwarz inequality with (2.50) and (2.42). For instance, for E_{12} we have

$$\begin{aligned}\left|\left(b(E_{12}^I - E_{12}), \xi_x\right)_{\Omega_f}\right| &\le C\|E_{12} - E_{12}^I\|_{0,\Omega_f}\|\xi_x\|_{0,\Omega_f} \\ &\le C\varepsilon^{1/4}N^{-1}\|\xi_x\|_0 \le CN^{-1/2}\|\xi\|_\varepsilon.\end{aligned} \tag{2.59}$$

Obviously, a similar estimate holds true for E_{23}.

Finally, we consider $\tilde{E} := E_2 + E_{12} + E_{23}$ in Ω_c where the mesh is quasi-uniform and of mesh size $H \sim N^{-1}$. Based on $x \ge \tau = 2\varepsilon/\beta \ln N$ for all points $(x,y) \in \Omega_c$ and (2.39b) we observe that $\tilde{E}$ is pointwise small, i.e. $\|\tilde{E}\|_{L_\infty(\Omega_c)} \le CN^{-2}$. Consequently, an inverse estimate gives

$$\begin{aligned}\left|\left(b(\tilde{E}^I - \tilde{E}), \xi_x\right)_{\Omega_c}\right| &\le C\|\tilde{E} - \tilde{E}^I\|_{0,\Omega_c}\|\xi_x\|_{0,\Omega_c} \\ &\le C\|\tilde{E} - \tilde{E}^I\|_{L_\infty(\Omega_c)}N\|\xi\|_{0,\Omega_c} \\ &\le C\|\tilde{E}\|_{L_\infty(\Omega_c)}N\|\xi\|_\varepsilon \le CN^{-1}\|\xi\|_\varepsilon.\end{aligned} \tag{2.60}$$

Collecting (2.56), (2.57), (2.58), (2.59) and (2.60) completes the proof. □

Finally, we consider the most interesting case in which ε is small in comparison to N^{-1}, i.e.

$$\sqrt{\varepsilon} \le N^{-1}. \tag{2.61}$$

Some techniques used for this case have been published in [63]. Note that the argument presented can be carried out without assuming the solution decomposition (2.39). Instead one may use an asymptotic approximation of Schieweck [68] — as was done in [63].

In view of (2.61) the fact that the characteristic layer terms and the corner layer components are of order $\mathcal{O}(\varepsilon^{1/4})$ in the energy norm gives

$$\|u - (S + E_2)\|_\varepsilon = \|E_1 + E_2 + E_{12} + E_{23}\|_\varepsilon \le C\varepsilon^{1/4} \le CN^{-1/2}. \tag{2.62}$$

This observation motivates the following splitting:

$$\|u - u^N\|_\varepsilon \le \|u - (S + E_2)\|_\varepsilon + \|S + E_2 - \pi(S + E_2)\|_\varepsilon + \|\pi(S + E_2) - u^N\|_\varepsilon \tag{2.63}$$

with some projection $\pi(S + E_2) \in V^N$ of $S + E_2$. By (2.62) we have already an estimate for the first term in (2.63). In order to study the second one we need to specify the projector π. Note that for the sum of regular solution component and exponential boundary layer one has

$$(S + E_2)|_{x=0} = (S + E_2)|_{x=1} = 0 \qquad \text{but in general} \qquad S + E_2 \not\equiv 0 \text{ on } \Gamma_2 \text{ and } \Gamma_4.$$

Hence, we cannot use a standard projector because this would break $\pi(S + E_2) \in V^N$. Instead we modify bilinear interpolation for grid points on $\Gamma_2 \cup \Gamma_4$:

$$(\pi v)(x_i, y_j) := \begin{cases} v(x_i, y_j) & \text{for } 0 \le i \le N,\ 1 \le j \le N-1 \\ 0 & \text{otherwise.}\end{cases}$$

Before we turn our attention to the second term in (2.63) we show a useful stability result for the projector π.

Lemma 10. *Consider the mesh rectangle $T = [x_i, x_{i+1}] \times [y_{N-1}, 1]$ for $i \in \{0, \dots, N-1\}$ with an edge length of H in y-direction. Let $v \in C(T) \cap W_{1,\infty}(T)$. Then*

$$\|(\pi v)_x\|_{0,T} \le C\sqrt{H}\|v_x(\cdot, y_{N-1})\|_{0,(x_i,x_{i+1})}.$$

Proof. One easily checks that for $(x,y) \in T$ and with $h_i := x_{i+1} - x_i$

$$\pi v(x,y) = v(x_i, y_{N-1})\frac{(x_{i+1} - x)(1 - y)}{h_i H} + v(x_{i+1}, y_{N-1})\frac{(x - x_i)(1 - y)}{h_i H}.$$

Hence,

$$(\pi v)_{xy}(x,y) = \frac{1}{h_i H}\left(v(x_i, y_{N-1}) - v(x_{i+1}, y_{N-1})\right) = -\frac{1}{h_i H}\int_{x_i}^{x_{i+1}} v_x(s, y_{N-1})\,\mathrm{d}s.$$

Using this identity and $(\pi v)_x(x,1)=0$ gives

$$(\pi v)_x(x,y) = -\int_y^1 (\pi v)_{xy}(x,t)\,\mathrm{d}t = \frac{1}{h_i H}\int_y^1\int_{x_i}^{x_{i+1}} v_x(s,y_{N-1})\,\mathrm{d}s\mathrm{d}t.$$

Squaring, using $0 \le 1-y \le H$ and applying the Cauchy-Schwarz inequality yields

$$\big((\pi v)_x(x,y)\big)^2 \le \frac{1}{h_i^2H^2}H^2\Big(\int_{x_i}^{x_{i+1}} |v_x(s,y_{N-1})|\,\mathrm{d}s\Big)^2 \le \frac{1}{h_i^2}h_i\int_{x_i}^{x_{i+1}}\big(v_x(s,y_{N-1})\big)^2\,\mathrm{d}s.$$

Integrating this estimate over T and taking the square root the proof is complete. □

Lemma 11. *Assuming* (2.61) *the projection error of* $S+E_2$ *satisfies*

$$\|S+E_2-\pi(S+E_2)\|_\varepsilon \le CN^{-1/2}.$$

Proof. Set $\tilde u = S+E_2$. Since the projection is given by bilinear interpolation in Ω_f^i and Ω_c^i, we can use standard arguments [64, See e.g. III.3.5.2] to obtain

$$\|\tilde u-\pi\tilde u\|_{\varepsilon,\Omega_f^i} \le CN^{-1}\ln N \quad\text{and}\quad \|\tilde u-\pi\tilde u\|_{\varepsilon,\Omega_c^i} \le CN^{-1}.$$

It remains to estimate the projection error in the subdomains Ω_f^b and Ω_c^b adjacent to the characteristic boundary. Here the projection of $\tilde u$ is non-standard because the values along the characteristic boundaries $y=0$ and $y=1$ are ignored and replaced by zero.

In Ω_f^b we use that $\tilde u$ as well as $\pi\tilde u$ are bounded to get

$$\|\tilde u-\pi\tilde u\|_{0,\Omega_f^b} \le \|\tilde u-\pi\tilde u\|_{L_\infty(\Omega_f^b)}(\operatorname{meas}\Omega_f^b)^{1/2} \le C\varepsilon^{1/2}N^{-1/2}(\ln N)^{1/2} \le CN^{-1/2}.$$

Similarly, by an inverse estimate, (2.39b) and (2.61)

$$\begin{aligned}\varepsilon^{1/2}\|\tilde u_y-(\pi\tilde u)_y\|_{0,\Omega_f^b} &\le \varepsilon^{1/2}\big(\|\tilde u_y\|_{0,\Omega_f^b}+\|(\pi\tilde u)_y\|_{0,\Omega_f^b}\big)\\ &\le C\varepsilon^{1/2}\big(\|\tilde u_y\|_{L_\infty(\Omega_f^b)}+H^{-1}\|\pi\tilde u\|_{L_\infty(\Omega_f^b)}\big)(\operatorname{meas}\Omega_f^b)^{1/2}\\ &\le C\varepsilon^{1/2}N^{-1/2}(\ln N)^{1/2} \le CN^{-1/2}.\end{aligned}$$

For the derivative with respect to x we use Lemma 10 and (2.39b):

$$\begin{aligned}\varepsilon^{1/2}\|\tilde u_x-(\pi\tilde u)_x\|_{0,\Omega_f^b} &\le \varepsilon^{1/2}\big(\|\tilde u_x\|_{0,\Omega_f^b}+\|(\pi\tilde u)_x\|_{0,\Omega_f^b}\big)\\ &\le C\varepsilon^{1/2}H^{1/2}\Big(\varepsilon^{-1/2}+\|\tilde u_x(\cdot,y_{N-1})\|_{0,(0,\tau)}+\|\tilde u_x(\cdot,y_1)\|_{0,(0,\tau)}\Big) \le CN^{-1/2}\end{aligned}$$

because $\|\tilde u_x(\cdot,y_{N-1})\|_{0,(0,\tau)}+\|\tilde u_x(\cdot,y_1)\|_{0,(0,\tau)} \le C\varepsilon^{-1/2}$.

Next we consider Ω_c^b and estimate the components S and E_2, separately. For S in Ω_c^b we use the same arguments as for $\tilde u$ in Ω_f^b: By $\|\pi S\|_{L_\infty(\Omega_c^b)} \le \|S\|_{L_\infty(\Omega_c^b)} \le C$ one obtains

$$\|S-\pi S\|_{0,\Omega_c^b} \le \|S-\pi S\|_{L_\infty(\Omega_c^b)}(\operatorname{meas}\Omega_c^b)^{1/2} \le CN^{-1/2}.$$

The derivative with respect to y is bounded using an inverse estimate and (2.61)

$$\begin{aligned}\varepsilon^{1/2}\|S_y-(\pi S)_y\|_{0,\Omega_c^b} &\le \varepsilon^{1/2}\big(\|S_y\|_{0,\Omega_c^b}+\|(\pi S)_y\|_{0,\Omega_c^b}\big)\\ &\le C\varepsilon^{1/2}\big(\|S_y\|_{L_\infty(\Omega_c^b)}+H^{-1}\|\pi S\|_{L_\infty(\Omega_c^b)}\big)(\operatorname{meas}\Omega_c^b)^{1/2}\\ &\le C\varepsilon^{1/2}N^{1/2} \le CN^{-1/2}.\end{aligned}$$

For the x-derivative a triangle inequality and Lemma 10 yields

$$\begin{aligned}\varepsilon^{1/2}\|S_x-(\pi S)_x\|_{0,\Omega_c^b} &\le \varepsilon^{1/2}\big(\|S_x\|_{0,\Omega_c^b}+\|(\pi S)_x\|_{0,\Omega_c^b}\big)\\ &\le C\varepsilon^{1/2}\Big((\operatorname{meas}\Omega_c^b)^{1/2}+\sqrt{H}\Big)\|S_x\|_{L_\infty(\Omega_c^b)} \le C\varepsilon^{1/2}N^{-1/2}.\end{aligned} \tag{2.64}$$

In Ω_c^b we have $x \ge \tau$. Consequently, $\|E_2\|_{L_\infty(\Omega_c^b)} \le CN^{-2}$ and

$$\begin{aligned}\|E_2-\pi E_2\|_{0,\Omega_c^b} &\le C\|E_2\|_{L_\infty(\Omega_c^b)} \le CN^{-2}, \qquad (2.65)\\ \varepsilon^{1/2}|E_2-\pi E_2|_{1,\Omega_c^b} &\le C\varepsilon^{1/2}\big(|E_2|_{1,\Omega_c^b}+H^{-1}\|E_2\|_{L_\infty(\Omega_c^b)}\big) \le C(\varepsilon^{1/2}N^{-1}+N^{-2}),\end{aligned}$$

where we used an inverse estimate again. □

Theorem 12. *Assume that* (2.61) *holds true. The error of the bilinear Galerkin finite element method for problem* (2.37) *on a Shishkin meshes with* N *intervals in each coordinate direction that is only fitted to the exponential boundary layer satisfies*

$$\|u - u^N\|_\varepsilon \le CN^{-1/2}. \tag{2.66}$$

Proof. Based on (2.63), (2.62) and Lemma 11 it remains to estimate the discrete error component $\xi := \pi(S + E_2) - u^N \in V^N$ in (2.63). Using coercivity of $a(\cdot,\cdot)$ and Galerkin orthogonality

$$C\|\xi\|_\varepsilon^2 \le a(\xi,\xi) = a\big(\pi(S+E_2) - u, \xi\big) = a\big(\pi(S+E_2) - (S+E_2), \xi\big) + a(S + E_2 - u, \xi). \tag{2.67}$$

We study the last term first. Integration by parts as in (2.55) but this time for the corner layer terms and the Cauchy-Schwarz inequality gives

$$\begin{aligned}|a(S + E_2 - u, \xi)| &= |a(-E_1 - E_2 - E_{12} - E_{23}, \xi)| \\ &\le C\|E_1 + E_2 + E_{12} + E_{23}\|_\varepsilon \|\xi\|_\varepsilon + \big|\big(b(E_{12} + E_{23}), \xi_x\big)\big| + \big|\big(b(E_1 + E_2)_x, \xi\big)\big| \\ &\le C\big(\varepsilon^{1/4} + \varepsilon^{-1/2}\|E_{12} + E_{23}\|_0 + \|(E_1 + E_2)_x\|_0\big)\|\xi\|_\varepsilon \le C\varepsilon^{1/4}\|\xi\|_\varepsilon.\end{aligned}$$

Hence, by (2.61)

$$a(S + E_2 - u, \xi) \le CN^{-1/2}\|\xi\|_\varepsilon. \tag{2.68}$$

Similarly, to (2.55) we start to estimate the first summand of (2.67), splitting it into

$$\begin{aligned}\big|a\big(\pi(S+E_2) - (S+E_2), \xi\big)\big| &\le \|\pi(S+E_2) - (S+E_2)\|_\varepsilon \|\xi\|_\varepsilon \\ &\quad + \big|\big(b(\pi S - S)_x, \xi\big)\big| + \big|\big(b(\pi E_2 - E_2), \xi_x\big)\big|\end{aligned} \tag{2.69}$$

The first summand in (2.69) is easily bounded by $N^{-1/2}\|\xi\|_\varepsilon$ based on Lemma 11.

In $\Omega^i := \Omega_f^i \cup \Omega_c^i$ where $\pi v = v^I$ coincides with bilinear interpolation one can use standard arguments for the other terms, i.e. (2.58), (2.60) and

$$\big|\big(b(\pi S - S)_x, \xi\big)_{\Omega^i}\big| \le C\|(S^I - S)_x\|_{0,\Omega^i}\|\xi\|_{0,\Omega^i} \le CN^{-1}\|\xi\|_\varepsilon. \tag{2.70}$$

In Ω_c^b we estimate

$$\big|\big(b(\pi S - S)_x, \xi\big)_{\Omega_c^b}\big| \le C\|(\pi S - S)_x\|_{0,\Omega_c^b}\|\xi\|_{0,\Omega_c^b} \le CN^{-1/2}\|\xi\|_\varepsilon \tag{2.71}$$

using the Cauchy-Schwarz inequality and (2.64). Clearly, a similar estimate holds true in Ω_f^b.

By (2.65) the smallness of the exponential boundary layer component E_2 in Ω_c^b and an inverse estimate give

$$\big|\big(b(\pi E_2 - E_2), \xi_x\big)_{\Omega_c^b}\big| \le C\|\pi E_2 - E_2\|_{0,\Omega_c^b}\|\xi_x\|_{0,\Omega_c^b} \le CN^{-2}H^{-1}\|\xi\|_0 \le CN^{-1}\|\xi\|_\varepsilon. \tag{2.72}$$

It remains to estimate $\big|\big(b(\pi E_2 - E_2), \xi_x\big)_{\Omega_f^b}\big|$ to complete the proof. Note that applying Hölder's inequality is too crude by a logarithmic factor:

$$\big|\big(b(\pi E_2 - E_2), \xi_x\big)_{\Omega_f^b}\big| \le C\|\pi E_2 - E_2\|_{L_\infty(\Omega_f^c)}(\operatorname{meas}\Omega_f^c)^{1/2}\|\xi_x\|_0 \le N^{-1/2}(\ln N)^{1/2}\|\xi\|_\varepsilon.$$

We estimate more carefully:

$$\big|\big(b(\pi E_2 - E_2), \xi_x\big)_{\Omega_f^b}\big| \le \big|\big(b(\pi E_2 - E_2^I), \xi_x\big)_{\Omega_f^b}\big| + \big|\big(b(E_2^I - E_2), \xi_x\big)_{\Omega_f^b}\big|$$

The second summand is again bounded by (2.58) in a standard way. Let us estimate the first summand on the lower ply of elements of Ω_f^b denoted by $\omega := [0,\tau] \times [0, y_1]$ as the other upper ply can be treated similarly. By the Cauchy-Schwarz inequality

$$\big|\big(b(\pi E_2 - E_2^I), \xi_x\big)_\omega\big| \le C\|\pi E_2 - E_2^I\|_{0,\omega}\|\xi_x\|_{0,\omega}. \tag{2.73}$$

The term $\pi E_2 - E_2^I$ is discrete (but not in V^N) and the only non-vanishing coefficients in a nodal basis representation of this term correspond to basis functions along the characteristic boundaries Γ_1 and Γ_3. Consequently,

$$\begin{aligned}\left\|\pi E_2 - E_2^I\right\|_{0,\omega}^2 &= \sum_{i=1}^{N/2} \left\|\pi E_2 - E_2^I\right\|_{0,[x_{i-1},x_i]\times[0,y_1]}^2 \le C \sum_{i=1}^{N/2} \int_0^{y_1} \int_{x_{i-1}}^{x_i} \max_{x\in\{x_{i-1},x_i\}} (E_2(x,0))^2 \,\mathrm{d}x\mathrm{d}y \\ &\le C \sum_{i=1}^{N/2} \int_0^{y_1} \int_{x_{i-1}}^{x_i} \mathrm{e}^{-2\beta x_{i-1}/\varepsilon} \,\mathrm{d}x\mathrm{d}y\end{aligned}$$

In order to bound this term we adapt a technique of Stynes and Tobiska, see e.g. [76, Lemma 3.2.]. In fact, $\mathrm{e}^{-2\beta x_{i-1}/\varepsilon} \le \mathrm{e}^{-2\beta x/\varepsilon}$ for $x \in [x_{i-2}, x_{i-1}]$. Hence,

$$\begin{aligned}\left\|\pi E_2 - E_2^I\right\|_{0,\omega}^2 &\le C \sum_{i=1}^{N/2-1} \int_0^{y_1} \int_{x_{i-1}}^{x_i} \mathrm{e}^{-2\beta x/\varepsilon} \,\mathrm{d}x\mathrm{d}y + C \int_0^{y_1} \int_0^{x_1} \mathrm{e}^{-2\beta x_0/\varepsilon} \,\mathrm{d}x\mathrm{d}y \\ &\le CH \int_0^{\tau} \mathrm{e}^{-2\beta x_0/\varepsilon} \,\mathrm{d}x + ChH \le CH(\varepsilon + h) \le C\varepsilon H.\end{aligned} \tag{2.74}$$

Collecting (2.73) and (2.74) gives

$$\left|\left(b(\pi E_2 - E_2^I), \xi_x\right)_\omega\right| \le C\sqrt{\varepsilon H}\|\xi_x\|_{0,\omega} \le CN^{-1/2}\|\xi\|_\varepsilon. \tag{2.75}$$

Similarly, an estimate on the upper ply of elements belonging to Ω_f^b follows. Summarizing the proof is complete. □

Since one can use standard arguments to obtain an error estimate superior to (2.66) in the regime $N^{-1} \le C\varepsilon$ combining Theorems 12 and 9 gives the main result of this section.

Corollary 13. *Assume that the solution u of* (2.37) *can be decomposed according to* (2.39)*. Let Ω^N denote the rectangular Shishkin mesh defined in the beginning of this section which has $N \ge 2$ intervals in each coordinate direction and is fitted to the exponential layer only. Hence, the mesh completely ignores the presence of the characteristic layers of u. Let $V^N \subset C(\overline{\Omega})$ be the space of piecewise bilinears on Ω^N. Then there is a constant C which is independent of ε and N such that the Galerkin finite element approximation u^N defined in* (2.41) *satisfies*

$$\|u - u^N\|_\varepsilon \le CN^{-1/2}.$$

Remark 10. In contrast to [63, Section 3.2 Theorem 1] Corollary 13 states uniform convergence of order 1/2. Unfortunately our proof relies on assuming (2.39) paying the price. Our numerical experiments show that this estimate is sharp.

2.3.1 Weakly imposed characteristic boundary conditions

Inspecting the proof of Theorem 12 we see that most problems are induced by the homogeneous boundary conditions of V^N along the characteristic boundaries Γ_1 and Γ_3. This observation raises the question: *What error estimates can we expect if we impose the boundary conditions weakly?* This would enable us to use standard Lagrange interpolation instead of π.

Note that a characteristic boundary layer occurs in general along the *characteristic boundary* Γ_0 that is parallel to the characteristics of the convective field $\vec{b}$, i.e. on Γ_0 the outer unit normal vector n to Ω is perpendicular to $\vec{b}$. Hence, imposing only the boundary conditions along Γ_0 weakly can be contrasted from the general case, in which artificial terms involving $\vec{b}\cdot n$ are introduced to ensure coercivity of the modified bilinear form, see [64, Part III Section 3.3.2]. In this sense weakly imposed characteristic boundary conditions are more natural than the complete treatment of the entire boundary with this technique.

We study this idea. In our case $\vec{b} = (-b, 0)^T$ and $\Gamma_0 := \Gamma_1 \cup \Gamma_3$. Let us introduce the function space $\tilde{V} := \{v \in H^1(\Omega) : v|_{x=0} = v|_{x=1} = 0 \text{ in the sense of traces}\}$ and denote by $\tilde{V}^N$ the space of piecewise bilinears over the mesh Ω^N. The bilinear Galerkin finite element method for problem (2.37) with weakly imposed characteristic boundary conditions reads: find $\tilde{u}^N \in \tilde{V}^N$ such that

$$a_N(\tilde{u}^N, v^N) = (f, v^N) \quad \text{for each } v^N \in \tilde{V}^N \tag{2.76a}$$

where

$$a_N(w,v) := \varepsilon(\nabla w, \nabla v) + (-bw_x + cw, v) + a_{\Gamma_0}(w,v), \tag{2.76b}$$

$$a_{\Gamma_0}(w,v) := -\varepsilon\Big(\frac{\partial w}{\partial n}, v\Big)_{\Gamma_0} \pm \varepsilon\Big(w, \frac{\partial v}{\partial n}\Big)_{\Gamma_0} + \frac{\gamma\varepsilon}{H}(w,v)_{\Gamma_0}. \tag{2.76c}$$

The positive penalty parameter $\gamma > 0$ will be needed to ensure coercivity if one picks the minus sign in (2.76c). Note that neglecting this term completely also devises a feasible method.

It is easy to see, that these methods are consistent: For a solution $u \in H_0^1 \cap H^2$ one has

$$a_N(u, v^N) = (f, v^N) \quad \text{for each } v^N \in \tilde{V}^N.$$

In order to analyze this method we rely on trace inequalities. We shall prove such a result and state the dependencies of the element-geometry, explicitly. Hence, this result will also be valuable on anisotropic meshes in subsequent chapters.

Lemma 14 (Anisotropic multiplicative trace inequality)**.** *Let T be a rectangle with sides parallel to the coordinate axes and a width in y-direction of h_y. Let ∂T_x denote the union of the two edges parallel to the x-axis. Then for $v \in W_{1,p}(T)$ we have the estimate*

$$\|v\|^p_{L_p(\partial T_x)} \le p\|v\|^{p-1}_{L_p(T)}\|v_y\|_{L_p(T)} + \frac{2}{h_y}\|v\|^p_{L_p(T)} \quad \text{for } p \in [1,\infty), \tag{2.77}$$

$$\|v\|_{L_\infty(\partial T_x)} \le \|v\|_{L_\infty(T)}. \tag{2.78}$$

Proof. The proof follows its isotropic version in [28, Theorem 1.5.1.10] (or [18, Lemma 3.1] in the L_2 setting): Without loss of generality we assume that the origin of the coordinate system is given by the midpoint of the rectangle T. The divergence theorem yields for $v \in C^1(\overline{T})$:

$$\begin{aligned}\int_T \frac{\partial}{\partial y}\left(|v|^p y\right)\mathrm{d}x\mathrm{d}y &= \int_T \nabla\cdot\begin{pmatrix}|v|^p y\\ 0\end{pmatrix}\mathrm{d}x\mathrm{d}y = \int_{\partial T} n\cdot\begin{pmatrix}|v|^p y\\ 0\end{pmatrix}\mathrm{d}s\\ &= \int_{\partial T_x}|v|^p|y|\mathrm{d}s = \frac{h_y}{2}\int_{\partial T_x}|v|^p\mathrm{d}s = \frac{h_y}{2}\|v\|^p_{L_p(\partial T_x)}.\end{aligned} \tag{2.79}$$

Moreover, since $|y| \le h_y/2$ on T an application of the product rule and Hölder's inequality with $\frac{1}{p}+\frac{1}{q}=1$ imply

$$\begin{aligned}\int_T \frac{\partial}{\partial y}\left(|v|^p y\right)\mathrm{d}x\mathrm{d}y &= \int_T \frac{\partial}{\partial y}\left(|v|^p\right) y\,\mathrm{d}x\mathrm{d}y + \int_T |v|^p\mathrm{d}x\mathrm{d}y = p\int_T |v|^{p-2}v\frac{\partial v}{\partial y}y\,\mathrm{d}x\mathrm{d}y + \|v\|^p_{L_p(T)}\\ &\le \frac{ph_y}{2}\int_T |v|^{p-1}\left|\frac{\partial v}{\partial y}\right|\mathrm{d}x\mathrm{d}y + \|v\|^p_{L_p(T)} \le \frac{ph_y}{2}\left(\int_T |v|^p\mathrm{d}x\mathrm{d}y\right)^{1/q}\left\|\frac{\partial v}{\partial y}\right\|_{L_p(T)} + \|v\|^p_{L_p(T)}\\ &\le \frac{ph_y}{2}\|v\|^{p-1}_{L_p(T)}\left\|\frac{\partial v}{\partial y}\right\|_{L_p(T)} + \|v\|^p_{L_p(T)}.\end{aligned}$$

The assertion follows from a standard density argument. The case $p = \infty$ is trivial. □

The trace inequality of Lemma 14 and an inverse estimate for an edge $E \subset \Gamma_0$ of a mesh rectangle T adjacent to Γ_0 yield

$$\varepsilon\left|\Big(\frac{\partial v^N}{\partial n}, v^N\Big)_E\right| \le \varepsilon\|v^N_y\|_{0,E}\|v^N\|_{0,E} \le \frac{C\varepsilon}{\sqrt{H}}\|v^N_y\|_{0,T}\|v^N\|_{0,E} \le \frac{\varepsilon}{2}|v^N|^2_{1,T} + \frac{C\varepsilon}{H}\|v^N\|^2_{0,E}.$$

Hence, the bilinear form $a_N(\cdot,\cdot)$ is coercive in $\tilde{V}^N$ with respect to the norm $|\!|\!|\cdot|\!|\!|$ given by

$$|\!|\!|v^N|\!|\!|^2 := \varepsilon|v^N|^2_1 + \|v^N\|_0 + \frac{\gamma\varepsilon}{H}\|v^N\|^2_{0,\Gamma_0}$$

provided that $\gamma \ge \gamma_0$ is sufficiently large (independently of ε and H). In this case $\tilde{u}^N$ is well defined.

Remark 11. Choosing the plus sign in (2.76c) obviously generates a method with a coercive bilinear form without further restrictions on $\gamma > 0$.

Based on (2.63) and the fact that one can use $(S+E_2)^I$ instead of $\pi(S+E_2)$, standard arguments [64, 43] improve the error estimate to

$$\|u-\tilde{u}^N\|_\varepsilon \le C\big(\varepsilon^{1/4}+N^{-1}\ln N\big)+\|\xi\|_\varepsilon, \tag{2.80}$$

where $\xi = (S+E_2)^I - \tilde{u}^N$.

For the discrete error component ξ we use

$$\|\xi\|_\varepsilon \le |||\xi||| \tag{2.81}$$

and bound ξ in the stronger norm $|||\cdot|||$. By the standard argument that led to (2.67)

$$C|||\xi|||^2 \le \big|a_N\big((S+E_2)^I-(S+E_2),\xi\big)\big| + \big|a_N(S+E_2-u,\xi)\big|.$$

Using again standard arguments [64, 43] for the Galerkin part $a(\cdot,\cdot)$ of the bilinear form $a_N(\cdot,\cdot)$ we arrive at

$$C|||\xi|||^2 \le C\big(\varepsilon^{1/4}+N^{-1}\ln N\big)|||\xi||| + \big|a_{\Gamma_0}\big((S+E_2)^I-(S+E_2),\xi\big)\big| + \big|a_{\Gamma_0}(S+E_2-u,\xi)\big|. \tag{2.82}$$

Let us abbreviate $\tilde{u} := S+E_2$ and estimate the two remaining terms of (2.82) by splitting them into the components of $a_{\Gamma_0}(\cdot,\cdot)$. We start off with the second summand:

$$\big|a_{\Gamma_0}(\tilde{u}^I-\tilde{u},\xi)\big| \le \Big|\varepsilon\Big(\frac{\partial}{\partial n}(\tilde{u}^I-\tilde{u}),\xi\Big)_{\Gamma_0}\Big| + \Big|\varepsilon\Big(\tilde{u}^I-\tilde{u},\frac{\partial\xi}{\partial n}\Big)_{\Gamma_0}\Big| + \Big|\frac{\gamma\varepsilon}{H}(\tilde{u}^I-\tilde{u},\xi)_{\Gamma_0}\Big|. \tag{2.83}$$

We need interpolation error estimates of $\tilde{u}$ on Γ_0. Applying the trace inequality of Lemma 14 yields

$$\|(\tilde{u}^I-\tilde{u})_y\|^2_{0,\Gamma_0} \le C\|(\tilde{u}^I-\tilde{u})_y\|_0\|(\tilde{u}^I-\tilde{u})_{yy}\|_0 + \frac{C}{H}\|(\tilde{u}^I-\tilde{u})_y\|_0^2 \le CN^{-1}, \tag{2.84}$$

$$\|\tilde{u}^I-\tilde{u}\|^2_{0,\Gamma_0} \le C\|\tilde{u}^I-\tilde{u}\|_0\|(\tilde{u}^I-\tilde{u})_y\|_0 + \frac{C}{H}\|\tilde{u}^I-\tilde{u}\|_0^2 \le CN^{-3}. \tag{2.85}$$

By the Cauchy-Schwarz inequality and (2.84)

$$\Big|\varepsilon\Big(\frac{\partial}{\partial n}(\tilde{u}^I-\tilde{u}),\xi\Big)_{\Gamma_0}\Big| \le \frac{\varepsilon^{1/2}H^{1/2}}{\gamma^{1/2}}\|(\tilde{u}^I-\tilde{u})_y\|_{0,\Gamma_0}\frac{\varepsilon^{1/2}\gamma^{1/2}}{H^{1/2}}\|\xi\|_{0,\Gamma_0} \le C\varepsilon^{1/2}N^{-1}|||\xi|||. \tag{2.86}$$

Similarly, a trace inequality and (2.85) give

$$\Big|\varepsilon\Big(\tilde{u}^I-\tilde{u},\frac{\partial\xi}{\partial n}\Big)_{\Gamma_0}\Big| \le \varepsilon^{1/2}\|\tilde{u}^I-\tilde{u}\|_{0,\Gamma_0}\frac{C\varepsilon^{1/2}}{H^{1/2}}\|\xi_y\|_0 \le C\varepsilon^{1/2}N^{-1}|||\xi|||. \tag{2.87}$$

Finally, simply use the Cauchy-Schwarz inequality and (2.85) to obtain

$$\Big|\frac{\gamma\varepsilon}{H}(\tilde{u}^I-\tilde{u},\xi)_{\Gamma_0}\Big| \le \frac{\gamma^{1/2}\varepsilon^{1/2}}{H^{1/2}}\|\tilde{u}^I-\tilde{u}\|_{0,\Gamma_0}\frac{\varepsilon^{1/2}\gamma^{1/2}}{H^{1/2}}\|\xi\|_{0,\Gamma_0} \le C\varepsilon^{1/2}N^{-1}|||\xi|||. \tag{2.88}$$

Next, we consider the last summand of (2.82):

$$\big|a_{\Gamma_0}(\tilde{u}-u,\xi)\big| \le \Big|\varepsilon\Big(\frac{\partial}{\partial n}(\tilde{u}-u),\xi\Big)_{\Gamma_0}\Big| + \Big|\varepsilon\Big(\tilde{u}-u,\frac{\partial\xi}{\partial n}\Big)_{\Gamma_0}\Big| + \Big|\frac{\gamma\varepsilon}{H}(\tilde{u}-u,\xi)_{\Gamma_0}\Big|. \tag{2.89}$$

The characteristic layers and corner layers $\tilde{u}-u$ are pointwise and ε-uniformly bounded by a constant, i.e. $\|\tilde{u}-u\|_{L_\infty(\Omega)} \le C$. Moreover, $\|(\tilde{u}-u)_y\|_{L_\infty(\Omega)} \le C\varepsilon^{-1/2}$. Hence, by a trace inequality

$$\Big|\varepsilon\Big(\frac{\partial}{\partial n}(\tilde{u}-u),\xi\Big)_{\Gamma_0}\Big| \le \varepsilon\|(\tilde{u}-u)_y\|_{0,\Gamma_0}\frac{C}{H^{1/2}}\|\xi\|_0 \le C(\varepsilon N)^{1/2}|||\xi|||. \tag{2.90}$$

Similarly,

$$\left|\varepsilon\Big(\tilde{u}-u,\frac{\partial\xi}{\partial n}\Big)_{\Gamma_0}\right| \le \varepsilon^{1/2}\|\tilde{u}-u\|_{0,\Gamma_0}\frac{C\varepsilon^{1/2}}{H^{1/2}}\|\xi_y\|_0 \le C(\varepsilon N)^{1/2}|\!|\!|\xi|\!|\!| \tag{2.91}$$

and

$$\left|\frac{\gamma\varepsilon}{H}(\tilde{u}-u,\xi)_{\Gamma_0}\right| \le \frac{\gamma^{1/2}\varepsilon^{1/2}}{H^{1/2}}\|\tilde{u}-u\|_{0,\Gamma_0}\frac{\varepsilon^{1/2}\gamma^{1/2}}{H^{1/2}}\|\xi\|_{0,\Gamma_0} \le C(\varepsilon N)^{1/2}|\!|\!|\xi|\!|\!|. \tag{2.92}$$

Collecting (2.80)–(2.83) and (2.86)–(2.92) we obtain the following result.

Theorem 15. *Assume that the solution u of* (2.37) *can be decomposed according to* (2.39). *Let Ω^N denote the rectangular Shishkin mesh defined in the beginning of this section which has $N \ge 2$ intervals in each coordinate direction and is fitted to the exponential layer only. Let $V^N \subset C(\overline{\Omega})$ be the space of piecewise bilinears on Ω^N. Then there is a constant C which is independent of ε and N such that the Galerkin finite element approximation $\tilde{u}^N$ with weakly imposed characteristic boundary conditions defined in* (2.76) *satisfies*

$$\|u-\tilde{u}^N\|_\varepsilon \le C\big(\varepsilon^{1/4}+N^{-1}\ln N+(\varepsilon N)^{1/2}\big).$$

Remark 12. Comparing Theorem 15 with the sharp estimate of Corollary 13:

$$\|u-u^N\|_\varepsilon \le CN^{-1/2},$$

we see that it is beneficial to impose the boundary conditions along the characteristic boundary weakly if ε is very small in comparison to N^{-1}. In the extreme case $\varepsilon \le CN^{-4}(\ln N)^4$ (which holds true for instance for the last two columns of Table 2.9) Theorem 15 states convergence of almost first order. This is the best possible result for bilinear elements on a Shishkin mesh.

2.4 Numerical experiments

2.4.1 A reaction-diffusion problem with boundary layers

Consider the test boundary value problem

$$\begin{aligned} -\varepsilon\Delta u + u &= f \quad \text{in } \Omega=(0,1)^2,\\ u &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

where $\varepsilon \in (0,1]$ and f is chosen in such a way that

$$u(x,y)=\hat{u}(x)\hat{u}(y), \quad \hat{u}(t)=-\frac{1-\mathrm{e}^{-1/\sqrt{\varepsilon}}}{1-\mathrm{e}^{-2/\sqrt{\varepsilon}}}\left(\mathrm{e}^{-t/\sqrt{\varepsilon}}+\mathrm{e}^{-(1-t)/\sqrt{\varepsilon}}\right)+1$$

is the exact solution, which exhibits typical boundary layer behavior. Let Ω_N denote the uniform mesh that is generated as the tensor product of two uniform 1D meshes dissecting the interval $(0,1)$ into N elements. Moreover, we define V^N as the FE-space of bilinear functions on Ω_N. We denote by u_N the solution of the Galerkin FE-method. We use some adaptive quadrature algorithm to compute all integrals with a tolerance of 10^{-10}.

In Table 2.1 we see the performance of the method in the setting $\varepsilon = 1$. In agreement with classical analysis one observes convergence of second order in the $L_2(\Omega)$ norm as well as in the $L_\infty(\Omega)$ norm and its discrete version. However, the error of the method measured in the energy norm is $\mathcal{O}(N^{-1})$ because its $H^1(\Omega)$ semi-norm component is only first order convergent and dominates the $L_2(\Omega)$ error (even for small N). Remark that similar results will follow if for the element diameter $h(N)$ it holds $h(N) \le \sqrt{\varepsilon}$, see Table 2.2.

If $\sqrt{\varepsilon}$ is small in comparison to $h(N)$ — as is the case in Table 2.3 — one observes that the rates of converges decline. Furthermore the error in the energy norm is dominated by the $L_2(\Omega)$ norm component for small values of N where the $H^1(\Omega)$ semi-norm component attains convergence rates smaller than 1/2.

Finally, Table 2.4 shows that if the ratio $\sqrt{\varepsilon}/h(N)$ is further decreased one observes no error reduction in the $H^1(\Omega)$ semi-norm or the $L_\infty(\Omega)$ norm. However, the $L_2(\Omega)$ error seems to be

N	$\sqrt{\varepsilon}\lvert u-u_N\rvert_1$		$\lVert u-u_N\rVert_0$		$\lVert u-u_N\rVert_\infty$		$\lVert u^I-u_N\rVert_\infty$	
	error	rate	error	rate	error	rate	error	rate
8	3.92e-3	1.00	1.22e-4		3.57e-3	2.08	1.66e-4	
16	1.96e-3	1.00	3.05e-5	2.00	8.41e-4	2.02	4.12e-5	2.01
32	9.78e-4	1.00	7.63e-6	2.00	2.07e-4	2.01	1.03e-5	2.00
64	4.89e-4	1.00	1.91e-6	2.00	5.16e-5	2.00	2.57e-6	2.00
128	2.45e-4	1.00	4.77e-7	2.00	1.29e-5	2.00	6.42e-7	2.00
256	1.22e-4	1.00	1.19e-7	2.00	3.22e-6	2.00	1.60e-7	2.00
512	6.11e-5		2.98e-8		8.06e-7		4.01e-8	

Table 2.1: $\varepsilon = 1$

N	$\sqrt{\varepsilon}\lvert u-u_N\rvert_1$		$\lVert u-u_N\rVert_0$		$\lVert u-u_N\rVert_\infty$		$\lVert u^I-u_N\rVert_\infty$	
	error	rate	error	rate	error	rate	error	rate
8	1.27e-1	0.94	4.08e-2	1.92	9.54e-2	1.53	4.74e-2	2.24
16	6.65e-2	0.98	1.08e-2	1.98	3.29e-2	1.74	1.00e-2	2.05
32	3.36e-2	1.00	2.73e-3	2.00	9.87e-3	1.86	2.41e-3	2.01
64	1.69e-2	1.00	6.84e-4	2.00	2.72e-3	1.93	5.98e-4	2.00
128	8.44e-3	1.00	1.71e-4	2.00	7.15e-4	1.96	1.49e-4	2.00
256	4.22e-3	1.00	4.28e-5	2.00	1.83e-4	1.98	3.73e-5	2.00
512	2.11e-3		1.07e-5		4.64e-5		9.31e-6	

Table 2.2: $\varepsilon = 10^{-2}$

N	$\sqrt{\varepsilon}\lvert u-u_N\rvert_1$		$\lVert u-u_N\rVert_0$		$\lVert u-u_N\rVert_\infty$		$\lVert u^I-u_N\rVert_\infty$	
	error	rate	error	rate	error	rate	error	rate
8	1.30e-1	0.14	2.96e-1	0.84	8.19e-1	0.42	5.65e-1	0.30
16	1.18e-1	0.37	1.65e-1	1.21	6.13e-1	0.87	4.57e-1	0.88
32	9.17e-2	0.69	7.11e-2	1.63	3.36e-1	1.35	2.48e-1	1.68
64	5.69e-2	0.90	2.30e-2	1.88	1.32e-1	1.46	7.70e-2	2.34
128	3.06e-2	0.97	6.24e-3	1.97	4.81e-2	1.69	1.53e-2	2.01
256	1.56e-2	0.99	1.60e-3	1.99	1.49e-2	1.83	3.78e-3	2.02
512	7.84e-3		4.01e-4		4.20e-3		9.30e-4	

Table 2.3: $\varepsilon = 10^{-4}$

N	$\sqrt{\varepsilon}\lvert u-u_N\rvert_1$		$\lVert u-u_N\rVert_0$		$\lVert u-u_N\rVert_\infty$		$\lVert u^I-u_N\rVert_\infty$	
	error	rate	error	rate	error	rate	error	rate
8	1.41e-2	0.00	3.72e-1	0.49	9.99e-1	0.00	6.08e-1	0.00
16	1.41e-2	0.00	2.65e-1	0.50	9.99e-1	0.00	6.07e-1	0.00
32	1.41e-2	0.00	1.88e-1	0.51	9.97e-1	0.00	6.07e-1	0.00
64	1.41e-2	0.01	1.32e-1	0.52	9.94e-1	0.01	6.07e-1	0.00
128	1.40e-2	0.02	9.17e-2	0.55	9.86e-1	0.04	6.07e-1	0.01
256	1.38e-2	0.03	6.27e-2	0.60	9.60e-1	0.10	6.03e-1	0.03
512	1.35e-2		4.13e-2		8.97e-1		5.90e-1	

Table 2.4: $\varepsilon = 10^{-8}$

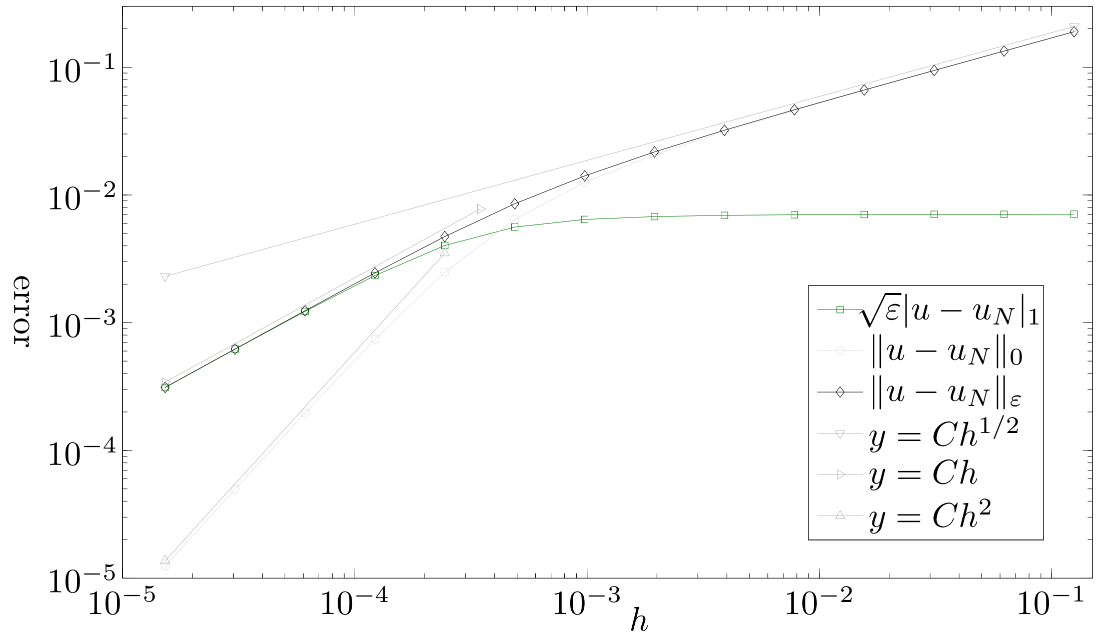

Figure 2.2: 1D reaction-diffusion problem: energy norm errors for $\varepsilon = 10^{-8}$

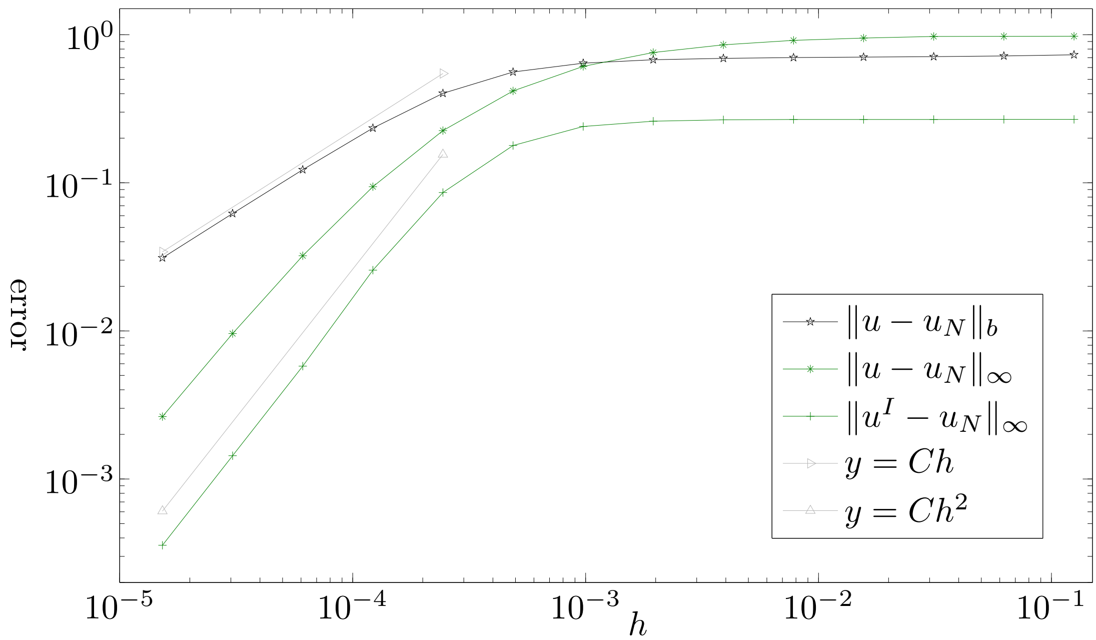

Figure 2.3: 1D reaction-diffusion problem: errors for $\varepsilon = 10^{-8}$

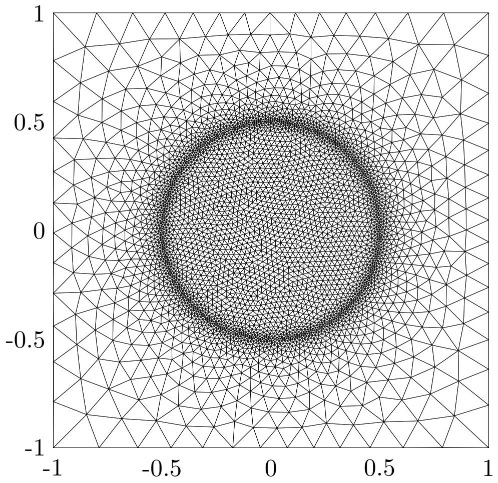

Figure 2.4: Example grid for the reference solution.

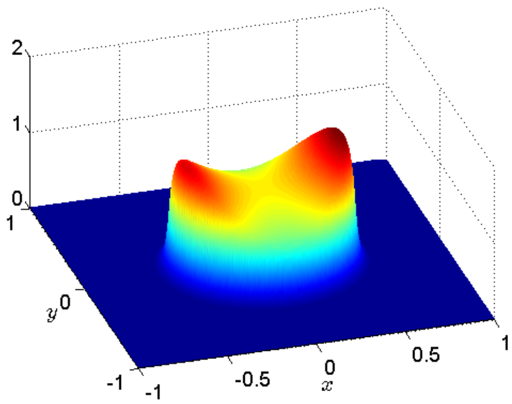

Figure 2.5: Reference solution for $\varepsilon = 10^{-4}$

$\mathcal{O}(N^{-1/2})$. In fact we have $\|u - u_N\|_0 \leq 1.06N^{-1/2}$. Since the $L_2(\Omega)$ error dominates the error in the $H^1(\Omega)$ semi-norm the energy norm error is $\mathcal{O}(N^{-1/2})$, as well. Note that in order for the $L_2(\Omega)$ error to reach the same magnitude as the $H^1(\Omega)$ semi-norm, namely $1.4e-2$, we would need $N > 5000$ elements reducing the critical ratio $\sqrt{\varepsilon}/h(N)$ and resulting in convergence rates similar to those of Table 2.3 (with $N > 32$).

The different phases described above can nicely be seen in Figure 2.2. Here a 1D reaction-diffusion problem similar to the original test problem with $\varepsilon = 10^{-8}$ is considered. Hence, a broader interval for the parameter $h = h(N)$ can be studied. Figure 2.3 shows that the approximations u_N do not converge uniformly in ε to the solution of the problem u neither in the $L_\infty(\Omega)$ norm nor in its discrete counterpart. An initial phase of error stagnation (until $h(N) \leq \sqrt{\varepsilon}$ and the layers are resolved) can also be observed in the stronger balanced norm $\|\cdot\|_b$, i.e. $\|v\|_b^2 := \sqrt{\varepsilon}|v|_1^2 + \|v\|_0$.

2.4.2 A reaction-diffusion problem with an interior layer

In this subsection we want to examine the sharpness of Theorem 4 and consider a problem with an interior layer along a Lipschitz curve:

$$\begin{aligned} -\varepsilon\Delta u + \left(1 + x^2y^2\mathrm{e}^{xy/2}\right)u &= f \quad \text{in } \Omega = (-1,1)^2, \\ u &= 0 \quad \text{on } \partial\Omega \end{aligned}$$

with $\varepsilon \in (0, 1]$ and

$$f(x,y) = \begin{cases} x^3(1+y) + \sin(\pi x^2) + \cos(\pi y/2) & \text{if } x^2 + y^2 \leq 1/4, \\ 0 & \text{else.} \end{cases}$$

Consequently, u will exhibit an interior layer near the circle centered at the origin and with radius 0.5. By the definition of f, the reduced solution f/c vanishes outside that circle. Hence, it satisfies homogeneous boundary conditions and no boundary layers arise. This is for our convenience: we numerically studied the performance of the Galerkin finite element method for reaction-diffusion problems with boundary layers already and want to focus on the interior layers.

Since we can't supply a reasonable solution structure for this problem we use a finely resolved numerical *reference solution* $u^\star$ to calculate the errors of the numerical scheme. By our a priori knowledge of the position of the layers we are able to generate a very fine and graded layer-adapted mesh, similarly to the one presented in Figure 2.4. However, in contrast to the coarse mesh depicted (featuring about 7560 mesh triangles) the high resolution mesh generated accumulates about 1 million triangular elements.We use quadratic finite elements over this mesh to compute the reference solution $u^\star$, which is depicted in Figure 2.5 for the moderate value $\varepsilon = 10^{-4}$. Moreover, we approximate all integrals by a high order Gaussian quadrature rule.

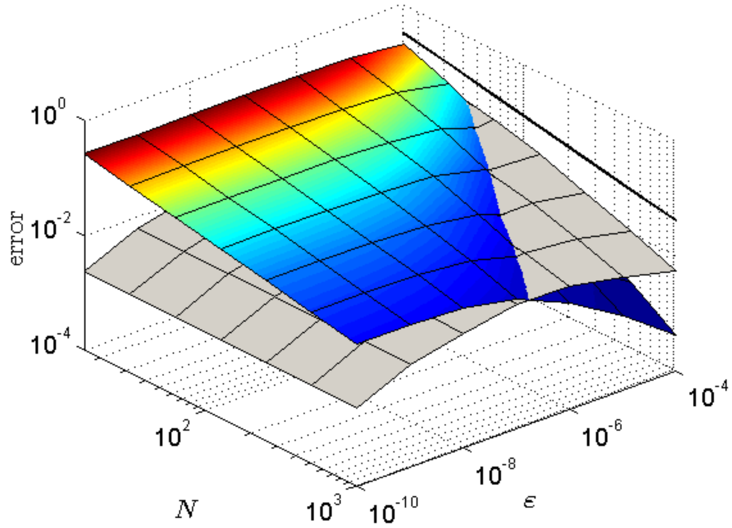

Figure 2.6: L_2-norm error (colored) and H^1 semi-norm error (gray) for various values of ε and N.

The results of this numerical experiment are shown in Table 2.5 and Figure 2.6. In the former errors and convergence orders with respect to the energy norm are displayed in the lower and upper half, respectively. Inspecting that table we find complete agreement with Theorem 4: The Galerkin finite element approximation on uniform meshes appears to converge ε-uniformly of order 1/2 to the (reference) solution in the energy norm. As before we note that the ratio of $1/N$ and $\sqrt{\varepsilon}$ is critical. If these two quantities are comparable in size, then we observe first order convergence of the Galerkin FEM in the energy norm. If however, $\sqrt{\varepsilon} \ll 1/N$ then the convergence order breaks down and numerically computed orders around 0.5 become evident. This also yields sharpness of Theorem 4.

Regarding the distribution of the error to the two components the situation is again similar to the previous numerical experiment. This time we use a three dimensional plot in form of Figure 2.6 to present these results. In that figure errors in both norm components of the energy norm — namely the L_2-norm (colored surface) and the weighted H^1 semi-norm (gray surface) — are plotted as surfaces over pairs (N, ε). Hence, moving further to the front in the parameter domain of Figure 2.6 corresponds to uniform mesh refinement, while for instance moving to the left decreases ε. Note that all three axes of that figure are logarithmic.

The two surfaces intersect along some almost straight line which is apparently of the form $\sqrt{\varepsilon} = C/N$. For $\sqrt{\varepsilon} \ll 1/N$ the error in the energy norm is dominated by the L_2-norm error. This is only due to the small multiplier of the H^1 semi-norm component. In particular the H^1 semi-norm error is almost constant for $\varepsilon = 10^{-10}$ and $N \leq 1024$. Consequently, we would observe error stagnation in the stronger norm $\|\cdot\|_b$ (as the larger multiplier would lift the gray surface). For $\varepsilon < 10^{-7}$ the dominant error component shifts to the H^1- semi-norm component within the observed parameter interval of N. Still the both components can be bounded by $1.3N^{-1/2}$ which is illustrated as black line in the right of Figure 2.6. This bound also holds true for $\varepsilon = 10^{-10}$ which also realizes the maximal error.

N	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-7}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-9}$	$\varepsilon = 10^{-10}$
16	0.76	0.64	0.58	0.55	0.53	0.53	0.53
32	0.83	0.66	0.55	0.52	0.52	0.52	0.52
64	0.91	0.76	0.61	0.54	0.50	0.49	0.49
128	0.94	0.88	0.70	0.58	0.54	0.53	0.54
256	0.96	0.89	0.78	0.61	0.53	0.50	0.49
512	0.98	0.94	0.88	0.72	0.58	0.53	0.52
16	2.28e-01	2.61e-01	2.70e-01	2.71e-01	2.71e-01	2.71e-01	2.71e-01
32	1.35e-01	1.67e-01	1.80e-01	1.85e-01	1.87e-01	1.87e-01	1.87e-01
64	7.60e-02	1.06e-01	1.23e-01	1.29e-01	1.30e-01	1.31e-01	1.31e-01
128	4.05e-02	6.26e-02	8.07e-02	8.87e-02	9.19e-02	9.30e-02	9.33e-02
256	2.10e-02	3.41e-02	4.96e-02	5.93e-02	6.32e-02	6.42e-02	6.43e-02
512	1.08e-02	1.83e-02	2.89e-02	3.89e-02	4.37e-02	4.54e-02	4.57e-02
1024	5.49e-03	9.54e-03	1.57e-02	2.36e-02	2.92e-02	3.15e-02	3.19e-02

Table 2.5: Difference to the reference solution $\|u^\star - u^N\|_\varepsilon$ of the bilinear Galerkin FEM on equidistant meshes for a problem with an interior layer.

2.4.3 A convection-diffusion problem with characteristic layers and a Neumann outflow condition

For the numerical verification of the results of Section 2.2 we consider the problem

$$\begin{gathered} -\varepsilon\Delta u - 2u_x + u = f \quad \text{in} \quad \Omega = (0,1)^2, \\ \left.\frac{\partial u}{\partial x}\right|_{x=0} = 0, \quad u|_{x=1} = 0 \quad \text{and} \quad u|_{y=0} = u|_{y=1} = 0, \end{gathered} \tag{2.93}$$

with $0 < \varepsilon \le 1$ and f chosen in such a way that

$$u(x,y) = \Big(\cos\frac{\pi x}{2} - 2x + 2 - \varepsilon\big(\mathrm{e}^{-2x/\varepsilon} - \mathrm{e}^{-2/\varepsilon}\big)\Big)\Big(1 - \mathrm{e}^{-y/\sqrt{\varepsilon}}\Big)\Big(1 - \mathrm{e}^{-(1-y)/\sqrt{\varepsilon}}\Big)$$

is the exact solution of (2.93). It exhibits typical boundary layer behavior for this kind of problem. We denote by u_N the finite element solution determined by

$$\varepsilon(\nabla u_N, \nabla v_N) + (u_N - 2u_{N,x}, v_N) = (f, v_N) \quad \text{for all } v_N \in V^N.$$

Here V^N is the FE-space of bilinear functions on a uniform mesh with N^2 elements like in the previous subsection.

In Table 2.6 we see the error of the method measured in the $\sqrt{\varepsilon}$-weighted $H^1(\Omega)$ semi-norm. For the last two columns in that table ($\varepsilon = 10^{-8}$ and $\varepsilon = 10^{-12}$) we observe no significant error reduction. In fact the error behaves almost like $\approx 1.8\varepsilon^{1/4}$ independently of N. Since u is of the same quality (according to (2.33): $|u|_1 \le C\varepsilon^{-1/4}$) this indicates that for $\sqrt{\varepsilon} \ll h_N := N^{-1}$ bilinear functions are not able to yield good approximations for a sharp layer function. For a bigger value of ε, namely $\varepsilon = 10^{-4}$ we find that the rates of convergence increase significantly if some hundred elements are considered. Hence, the ratio $\sqrt{\varepsilon}/h_N$ appears to be significant. Remark that a parabolic boundary layer has a width $\mathcal{O}(\sqrt{\varepsilon}\ln(1/\varepsilon))$. If $\sqrt{\varepsilon}/h_N \ge 1$ we observe first order convergence.

The $L_2(\Omega)$ errors and corresponding rates are depicted in Table 2.7. For $\sqrt{\varepsilon}/h_N \ge 1$ second order convergence can be observed. If that ratio is smaller than one the rates start to fall but $1/2$ appears to be a lower bound for all $L_2(\Omega)$ rates, independently of the quotient $\sqrt{\varepsilon}/h_N$. Remark also that for $\sqrt{\varepsilon}/h_N \ll 1$ the $L_2(\Omega)$ error dominates the corresponding $\sqrt{\varepsilon}$-weighted $H^1(\Omega)$ semi-norm error. Hence, in this case the energy norm error is essentially given by its $L_2(\Omega)$ component.

N	$\varepsilon = 1$		$\varepsilon = 10^{-2}$		$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-12}$
	error	rate	error	rate	error	rate	error	error
8	6.98e-2		2.04e-1		1.73e-1		1.86e-2	1.86e-3
16	3.49e-2	1.00	1.06e-1	0.94	1.58e-1	0.13	1.86e-2	1.86e-3
32	1.74e-2	1.00	5.37e-2	0.98	1.25e-1	0.34	1.85e-2	1.86e-3
64	8.72e-3	1.00	2.72e-2	0.98	7.78e-2	0.68	1.85e-2	1.86e-3
128	4.36e-3	1.00	1.37e-2	0.99	4.13e-2	0.91	1.84e-2	1.87e-3
256	2.18e-3	1.00	6.89e-3	1.00	2.09e-2	0.98	1.82e-2	1.87e-3

Table 2.6: $\sqrt{\varepsilon}|u - u_N|_1$ for different values of N and ε.

N	$\varepsilon = 1$		$\varepsilon = 10^{-2}$		$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-8}$		$\varepsilon = 10^{-12}$	
	error	rate	error	rate	error	rate	error	rate	error	rate
8	2.23e-3		5.89e-2		4.00e-1		4.98e-1		4.99e-1	
16	5.56e-4	2.00	1.49e-2	1.99	2.20e-1	0.86	3.52e-1	0.50	3.53e-1	0.50
32	1.39e-4	2.00	3.74e-3	1.99	9.34e-2	1.24	2.48e-1	0.51	2.50e-1	0.50
64	3.47e-5	2.00	9.47e-4	1.98	2.85e-2	1.71	1.74e-1	0.51	1.77e-1	0.50
128	8.68e-6	2.00	2.39e-4	1.98	7.30e-3	1.97	1.21e-1	0.52	1.25e-1	0.50
256	2.17e-6	2.00	6.01e-5	1.99	1.82e-3	2.01	8.25e-2	0.55	8.82e-2	0.50

Table 2.7: $\|u - u_N\|_0$ for different values of N and ε.

2.4.4 A mesh that resolves only part of the exponential layer and neglects the weaker characteristic layers

In the problem of Section 2.3 the energy norm associated with (2.37) is strong enough to capture the exponential layer in contrast to the previous problems. The numerical results reflect this exceptional feature. Let u^N denote the solution of (2.41) on the tensor product mesh composed by a Shishkin mesh in x-direction and a uniform mesh in y-direction introduced in Section 2.3.

N	$\varepsilon = 1$		$\varepsilon = 10^{-2}$		$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-8}$		$\varepsilon = 10^{-12}$	
	error	rate	error	rate	error	rate	error	rate	error	rate
8	2.21e-2		2.59e-1		3.61e-1		3.90e-1		3.90e-1	
16	1.10e-2	1.00	1.70e-1	0.60	2.35e-1	0.62	2.71e-1	0.52	2.72e-1	0.52
32	5.52e-3	1.00	1.06e-1	0.68	1.41e-1	0.74	1.82e-1	0.58	1.82e-1	0.57
64	2.76e-3	1.00	6.35e-2	0.74	8.10e-2	0.80	1.20e-1	0.60	1.21e-1	0.59
128	1.38e-3	1.00	3.70e-2	0.78	4.60e-2	0.81	7.86e-2	0.61	8.01e-2	0.59
256	6.90e-4	1.00	2.11e-2	0.81	2.59e-2	0.83	5.14e-2	0.61	5.36e-2	0.58
512	3.45e-4	1.00	1.18e-2	0.83	1.44e-2	0.84	3.31e-2	0.63	3.64e-2	0.56

Table 2.8: $\sqrt{\varepsilon}|u - u^N|_1$ for different values of N and ε.

N	$\varepsilon = 1$		$\varepsilon = 10^{-2}$		$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-8}$		$\varepsilon = 10^{-12}$	
	error	rate	error	rate	error	rate	error	rate	error	rate
8	6.55e-4		2.57e-2		1.53e-1		1.91e-1		1.91e-1	
16	1.63e-4	2.00	6.12e-3	2.07	8.40e-2	0.87	1.34e-1	0.51	1.34e-1	0.51
32	4.09e-5	2.00	1.53e-3	2.00	3.56e-2	1.24	9.42e-2	0.51	9.50e-2	0.50
64	1.02e-5	2.00	4.05e-4	1.92	1.09e-2	1.71	6.60e-2	0.51	6.72e-2	0.50
128	2.55e-6	2.00	1.11e-4	1.87	2.80e-3	1.96	4.59e-2	0.52	4.75e-2	0.50
256	6.38e-7	2.00	3.11e-5	1.84	6.99e-4	2.00	3.14e-2	0.55	3.36e-2	0.50
512	1.60e-7	2.00	8.85e-6	1.81	1.75e-4	2.00	2.07e-2	0.60	2.37e-2	0.50

Table 2.9: $\|u - u^N\|_0$ for different values of N and ε.

As in the previous subsection we choose $b = 2$ and $c = 1$. Moreover, we determine f in such way that

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

is the exact solution of (2.37).

In Table 2.8 and Table 2.9 the $\sqrt{\varepsilon}$-weighted $H^1(\Omega)$ semi-norm errors and the errors in $L_2(\Omega)$ are presented, respectively. In contrast to previous numerical results both components are almost of the same magnitude if ε is very small. For the $\sqrt{\varepsilon}$-weighted $H^1(\Omega)$ semi-norm error we observe first order convergence in the non-perturbed case and $|u - u^N| \leq CN^{-1}\ln N$ if the singular perturbation is mild ($\varepsilon \geq 10^{-4}$). In the case of a very small perturbation parameter convergence with a rate slightly greater than 1/2 can be observed uniformly in ε. For the $L_2(\Omega)$ error the rates of convergence are close to two for $\varepsilon \geq 10^{-4}$ and $N \geq 64$. If $\sqrt{\varepsilon} \ll N^{-1}$ we observe uniform convergence of order 1/2. Hence, the method is of order 1/2 in the energy norm and our theoretical findings are sharp.

3 Macro-element interpolation on tensor product meshes

3.1 Introduction

There is a high interest in differentiable finite elements and their corresponding interpolation operators as these are used for instance in the construction and analysis of methods for higher order problems like the biharmonic equation. On a triangular mesh the fifth degree Argyris element and its reduced version — the Bell element — are most popular. However, they are rarely used as they introduce a large number of degrees of freedom. In fact, Ženižek [79] showed that on a triangular element with polynomial shape functions at least 18 degrees of freedom are needed to grant the C^1 property. In this respect the Bell element can be considered optimal.

The desire for reducing the number of degrees of freedom used (and therefore the polynomial degree) lead to the construction of macro-elements in 1960s and 1970s. Let us mention the cubic Hsieh-Clough-Tocher macro-element [12] and the quadratic Powell-Sabin macro-element [56]. In the latter, each base triangle is split into six sub-triangles that share an inner point (for instance the center of the inscribed circle) of the base triangle. The inner degrees of freedom are then eliminated by the C^1 property.

While there is a huge amount of literature for triangular macro-elements (see for instance the survey article [53] and the references therein), there appears to be only one publication [31] dealing with rectangular ones. Moreover, to the knowledge of the author, there appears to be no paper dealing with anisotropic interpolation error estimates for macro-element interpolation, i.e. up to now macro-element interpolation has only been considered on quasi-uniform meshes. However, one can certainly improve the approximation quality by allowing elements with an arbitrarily high aspect ratio in certain cases. This benefit becomes obvious if the underlaying domain or the function to be approximated has anisotropic features (like layers).

In Section 3.2 of this chapter we shall briefly introduce the concept of $C^1 - P_2$ macro-interpolation in the 1D case and fix some notation.

The following Section 3.3 starts by showing how the 1D $C^1 - P_2$ macro-element extends to the 2D $C^1 - Q_2$ macro-element on tensor product meshes. Then a general theory for obtaining anisotropic interpolation error estimates for macro-element interpolation is developed and general construction principles are revealed. This theory is then applied in order to analyze the $C^1 - Q_2$ macro-element interpolation operator Π as well as some reduced counterpart.

Thereafter we discuss a modification of Π of Scott-Zhang [71] type in Subsection 3.5.3 giving optimal error estimates under the regularity required. The price to pay is that not all linear functionals that define this modified operator are local, i.e. in order to obtain the value of the quasi-interpolant on a base macro-element M some averaging process of the data on a macro-element edge that does not necessarily belong to M is needed. This causes some difficulties because quasi-interpolation operators of similar type are mostly studied on quasi-uniform meshes.

We summarize our results concerning C^1 (quasi-)interpolation in Subsection 3.5.4 and cite some results of the literature.

In Section 3.6 we introduce and analyze an anisotropic macro-element interpolation operator. Basically, this operator is the tensor product of one-dimensional $C^1 - P_2$ macro-interpolation and P_2 Lagrange interpolation.

We conclude this chapter with Section 3.7 in which we apply the results of the (Sub-)Sections 3.5.3 and 3.6 in order to approximate the solution of a singularly perturbed reaction-diffusion problem on a Shishkin mesh that features anisotropic elements, i.e. elements with an unbounded aspect ratio for $\varepsilon \to 0$. Hereby we obtain an approximation whose normal derivative is continuous along certain edges of the mesh, enabling a more sophisticated analysis of a continuous interior penalty method in the next chapter.

3.2 Univariate $C^1 - P_2$ macro-element interpolation

Consider the 1D Hermite interpolation problem on the interval $[-1, 1]$: Let u be a real function over $[-1, 1]$ such that $u(\pm 1)$, $u'(\pm 1) \in \mathbb{R}$ can be defined. Find $s \in C^1[-1, 1]$, such that

$$s(\pm 1) = u(\pm 1), \qquad s'(\pm 1) = u'(\pm 1). \tag{3.1}$$

In 1983 Schumaker [69] observed that while the Hermite interpolation problem considered is only solvable for a quadratic polynomial $s \in P_2[-1, 1]$ if and only if

$$u'(-1) + u'(1) = u(1) - u(-1),$$

there is always a solution in the space of quadratic splines with one simple knot. We may choose $x = 0$ as this knot and introduce the spline space

$$S^2 := \left\{ v \in C^1[-1, 1] \,:\, v|_T \in P_2(T),\ T \in \{[-1, 0], [0, -1]\} \right\}.$$

Of course other choices for the additional knot are possible. This parameter can be used to grant additional properties of the underlaying interpolation operator, see [69].

A function s that is a quadratic polynomial on each of the intervals $[-1, 0]$ and $[0, 1]$ can be characterized by six parameters of which two are determined by the C^1 property at zero. Hence, the remaining four parameters of a function $s \in S^2$ may be chosen in such a way that (3.1) is fulfilled. In fact, a simple calculation shows that

$$s(x) = \sum_{i=\pm 1} \Big(u(i)\hat{\varphi}_i(x) + u'(i)\hat{\psi}_i(x) \Big), \quad x \in [-1, 1] \tag{3.2}$$

is the unique solution of (3.1) in S^2. Here $\hat{\varphi}_{\pm 1}$ and $\hat{\psi}_{\pm 1} \in S^2$ denote the Lagrangian basis functions

$$\begin{aligned}
\hat{\varphi}_{-1}(x) &= \frac{(x-1)^2}{2} - \begin{cases} x^2, & x \in [-1, 0], \\ 0, & x \in [0, 1], \end{cases} \qquad & \hat{\varphi}_1(x) &= \frac{(x+1)^2}{2} - \begin{cases} 0, & x \in [-1, 0], \\ x^2, & x \in [0, 1], \end{cases} \\
\hat{\psi}_{-1}(x) &= \frac{(x-1)^2}{4} - \begin{cases} x^2, & x \in [-1, 0], \\ 0, & x \in [0, 1], \end{cases} \qquad & \hat{\psi}_1(x) &= -\frac{(x+1)^2}{4} + \begin{cases} 0, & x \in [-1, 0], \\ x^2, & x \in [0, 1], \end{cases}
\end{aligned} \tag{3.3}$$

i.e. these spline functions fulfill the conditions

$$\begin{gathered}
\hat{\varphi}_{-1}(-1) = \hat{\varphi}_1(1) = \hat{\psi}'_{-1}(-1) = \hat{\psi}'_1(1) = 1, \\
\hat{\varphi}_1(-1) = \hat{\psi}_{-1}(-1) = \hat{\psi}_1(-1) = \hat{\varphi}_{-1}(1) = \hat{\psi}_{-1}(1) = \hat{\psi}_1(1) = 0, \\
\hat{\varphi}'_{-1}(-1) = \hat{\varphi}'_1(-1) = \hat{\psi}'_1(-1) = \hat{\varphi}'_{-1}(1) = \hat{\varphi}'_1(1) = \hat{\psi}'_{-1}(1) = 0.
\end{gathered}$$

For a graphical representation of these functions, see Figure 3.1.

Based on the symmetry of the subproblem defining the basis functions we observe

$$\hat{\varphi}_{-1}(x) = \hat{\varphi}_1(-x) \qquad \text{and} \qquad \hat{\psi}_{-1}(x) = -\hat{\psi}_1(-x) \quad \forall x \in [-1, 1].$$

Moreover, $\hat{\varphi}'_{\pm 1}$ are even functions, i.e.

$$\hat{\varphi}'_{\pm 1}(x) = \hat{\varphi}'_{\pm 1}(-x) \quad \forall x \in [-1, 1].$$

From these properties we can deduce that $\hat{\varphi}'_1(x) = -\hat{\varphi}'_{-1}(-x) = \hat{\varphi}'_{-1}(x)$ for all $x \in [-1, 1]$. Hence, similar to a cubic polynomial the derivative s' of a spline $s \in S^2$ is an element of a three dimensional vector space. Since the second derivative of the spline considered is piecewise constant, it belongs to a two dimensional space.

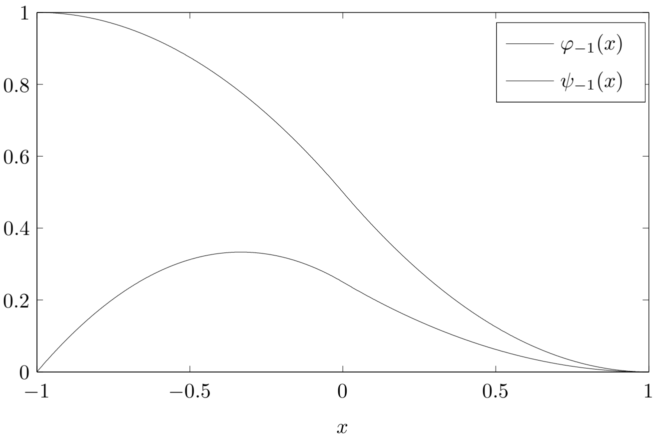

Figure 3.1: Lagrangian basis functions φ_{-1} and ψ_{-1}

This fact can nicely be seen if we switch from the Lagrangian representation (3.2) of the solution of (3.1) to its Newtonian one. Based on $\hat{\psi}_1(\pm 1) = \hat{\psi}_1'(-1) = 0$ we observe, that

$$s(x) = u[-1] + u[-1,-1](x+1) + u[-1,-1,1](x+1)^2 + u[-1,-1,1,1]4\hat{\psi}_1(x). \tag{3.4}$$

Here $u[x_0,\ldots,x_N]$ are the well known divided differences of order N of u with possibly coincident knots $x_0 \le x_1 \le \cdots \le x_N$, recursively defined by

$$u[x_i] := u(x_i) \quad \text{and} \quad u[x_0,\ldots,x_N] := \begin{cases} \dfrac{1}{N!}u^{(N)}(x_0), & \text{if } x_0 = \cdots = x_N, \\ \dfrac{u[x_1,\ldots,x_N] - u[x_0,\ldots,x_{N-1}]}{x_N - x_0}, & \text{else.} \end{cases}$$

A simple calculation shows that

$$\begin{aligned} u[-1] = u(-1), \quad u[-1,-1] = u'(-1), \quad u[-1,-1,1] &= \frac{1}{4}\big(u(1) - u(-1)\big) - \frac{1}{2}u'(-1), \\ u[-1,-1,1,1] &= \frac{1}{4}\big(u(-1) - u(1) + u'(-1) + u'(1)\big). \end{aligned} \tag{3.5}$$

If we substitute the expressions from (3.5) into (3.4) and expand in terms of $u(\pm 1)$ and $u'(\pm 1)$ we re-obtain the Lagrangian representation (3.2) of s. However, the Newtonian form (3.4) of s will prove to be very useful in the derivation of anisotropic interpolation error estimates.

3.3 $C^1 - Q_2$ macro-element interpolation on tensor product meshes

One can easily solve the Hermite interpolation problem (3.1) for a cubic polynomial s. Hence, similar to (3.3) a Lagrangian basis for a cubic C^1 spline can be obtained associated with the values of the function and its first derivative in the endpoints of the interval considered. It is well-known that the tensor product of this basis of the cubic C^1 splines leads to the Bogner-Fox-Schmidt element, which is in fact a C^1 element. Here the 16 degrees of freedom are associated with the values $v(V_i)$, the first derivatives $v_x(V_i)$, $v_y(V_i)$ and the mixed derivative $v_{xy}(V_i)$ of a function $v \in Q_3(T)$ at the four vertices V_i, $i = 1,\ldots,4$ of a rectangle T, see Figure 3.2. Note that the restriction of the generated finite element space to any element T is $Q_3(T)$, where T is a rectangle of the underlaying triangulation with sides aligned to the coordinate axes.

By analogy with the Bogner-Fox-Schmidt element the tensor product of the basis functions (3.3) generates a C^1 macro-element, as well. One obtains 16 basis functions that are piecewise biquadratic:

$$\begin{aligned} \hat{\varphi}_{i,j}(x,y) &:= \hat{\varphi}_i(x)\hat{\varphi}_j(y), & \hat{\phi}_{i,j}(x,y) &:= \hat{\psi}_i(x)\hat{\varphi}_j(y), \\ \hat{\chi}_{i,j}(x,y) &:= \hat{\varphi}_i(x)\hat{\psi}_j(y), & \hat{\psi}_{i,j}(x,y) &:= \hat{\psi}_i(x)\hat{\psi}_j(y), \end{aligned} \quad i,j \in \{-1,1\}. \tag{3.6}$$

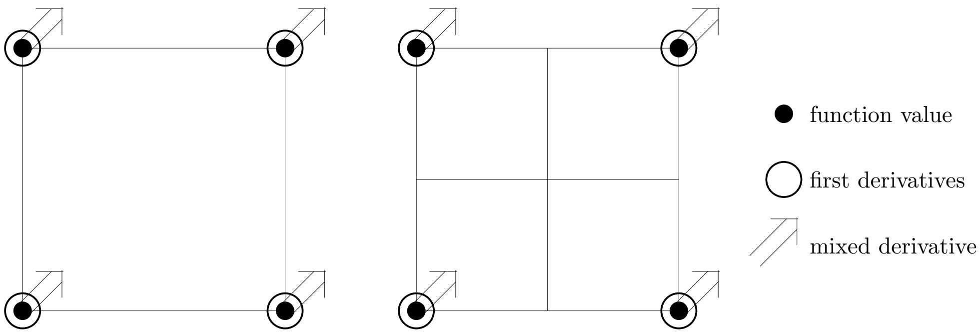

Figure 3.2: The Bogner-Fox-Schmidt Q_3 element (left) and its Q_2 analogue on a macro of four elements (right)

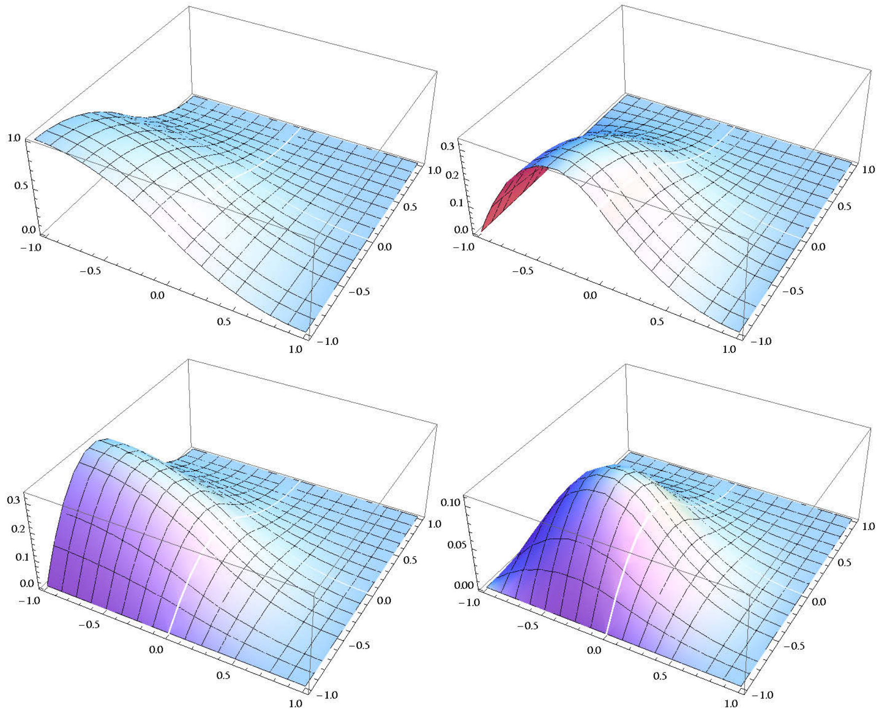

Figure 3.3: The basis functions $\hat{\varphi}_{-1,-1}$, $\hat{\phi}_{-1,-1}$, $\hat{\chi}_{-1,-1}$, $\hat{\psi}_{-1,-1}$ on the reference macro-element.

Whenever definitions are tied to a reference (macro-)element we shall continue to use a hat symbol to emphasize this fact. With the dual functionals

$$\begin{aligned} F_{i,j}^{\hat{\varphi}}(v) &\coloneqq v(i,j), & F_{i,j}^{\hat{\phi}}(v) &\coloneqq v_x(i,j), \\ F_{i,j}^{\hat{\chi}}(v) &\coloneqq v_y(i,j), & F_{i,j}^{\hat{\psi}}(v) &\coloneqq v_{xy}(i,j), \end{aligned} \qquad i,j \in \{-1,1\}.$$

the basis functions obey the Lagrange relation

$$F_{i,j}^{v}(w_{k,\ell}) = \delta_{vw}\delta_{ik}\delta_{j\ell}$$

for $v, w \in \{\hat{\varphi}, \hat{\phi}, \hat{\chi}, \hat{\psi}\}$ and $i, j, k, \ell \in \{-1, 1\}$. We denote by $\hat{M}$ the reference macro-element which is given as the triangulation of the reference domain $\Lambda \coloneqq [-1,1]^2$ induced by the coordinate axes. On $\hat{M}$ the four basis functions for $i = j = -1$ associated with the point $(-1,-1)$ are depicted in Figure 3.3.

In a natural way, a biquadratic interpolant $\hat{\Pi}v \in C^1(\Lambda)$ of a function $v \in C^2(\Lambda)$ is defined by

$$\hat{\Pi}v = \sum_{i,j\in\{-1,1\}} F_{i,j}^{\hat{\varphi}}(v)\hat{\varphi}_{i,j} + F_{i,j}^{\hat{\phi}}(v)\hat{\phi}_{i,j} + F_{i,j}^{\hat{\chi}}(v)\hat{\chi}_{i,j} + F_{i,j}^{\hat{\psi}}(v)\hat{\psi}_{i,j}. \tag{3.7}$$

By affine equivalence, it suffices to define the interpolation operator $\hat{\Pi}$ on the reference macro-element $\hat{M}$. Given a rectangular macro-element mesh $\mathcal{M}$ of tensor product type, the value of the interpolant Πv of a function $v \in C^2(\overline{\Omega})$ in a certain point $(x, y) \in \overline{\Omega}$ of the physical domain can be obtained by identifying a macro-element M such that $(x, y) \in M$ and performing an affine transformation.

After making an independent construction an excessive search of the literature available showed that the $C^1 - Q_2$ macro-element is not new. In fact, it can be traced back to the PhD thesis [5]. In the work [30] the thesis [5] is cited and optimal interpolation error estimates

$$|u - \Pi u|_m \leq C h^{3-m} |u|_3$$

for $m = 0, 1, 2$ are proven for $u \in H^3(\Omega) \cap C^2(\overline{\Omega})$ and a tensor product triangulation which is required to be quasi-uniform.

Strangely, this idea appears to be unpublished until 2011. In [31] the C^1 property of the finite element space V_h introduced by the $C^1 - Q_2$ macro-element on a tensor product triangulation $\mathcal{T}_h$ of a domain Ω is shown. Moreover, it is established that V_h coincides with the full $C^1 - Q_2$ space, i.e.:

$$V_h = \{v_h \in C^1(\Omega) \,:\, v_h|_T \in Q_2(T) \;\forall T \in \mathcal{T}_h\}. \tag{3.8}$$

This appears to be of high interest in certain applications. Finally, optimal interpolation error estimates are derived for an extension of the Girault-Scott operator into the $C^1 - Q_2$ finite element space, i.e. a modification $\tilde{\Pi}$ of the operator Π (defined via an affine transformation as $\hat{\Pi}$ on the reference macro-element $\hat{M}$ in (3.7)) is obtained in such a way that a function $v \in H^2(\Omega)$ can be interpolated and

$$\|v - \tilde{\Pi} v\|_0 + h|v - \tilde{\Pi} v|_1 + h^2 |v - \tilde{\Pi} v|_2 \leq C h^2 |v|_2.$$

However, the analysis in [31] of the interpolation error also requires quasi-uniformity of the triangulation $\mathcal{T}_h$, i.e. it is assumed that there is a positive constant $C > 0$ such that for all axis-aligned mesh rectangles $T \in \mathcal{T}_h$ the edge lengths $h_x(T)$ and $h_y(T)$ in x- and y-direction are equivalent to a global discretization parameter h, i.e.

$$Ch \leq h_x(T), h_y(T) \leq h \quad \forall T \in \mathcal{T}_h. \tag{3.9}$$

On the other hand there are problems that can be treated efficiently if elements with very high aspect ratios are permitted within the triangulation or if edge lengths of neighboring elements are allowed to vary unbounded. As examples, let us mention the approximation of a smooth function over a long and thin domain Ω or solutions of partial differential equations with anisotropic behavior like layers. Wherefore we ask the question: *Is it possible to prove anisotropic interpolation error estimates for the operator* Π *from* (3.7) *or a modification of it?*

It turns out that the wonderful theory of [3, 2] is incapable to handle the analysis of macro-element interpolation. In the following we shall therefore develop a slight modification of it.

3.4 A theory on anisotropic macro-element interpolation

We first introduce some notation, partly adopted from [3].

Let $\hat{M} := \{\hat{T}_i\}_{i=1}^{\ell}$ be our reference macro-element, i.e. a triangulation of some reference domain Λ. For a set of multi-indices $\boldsymbol{P}$ we denote by

$$\boldsymbol{P}(\Lambda) := \operatorname{span}\{\boldsymbol{X} \mapsto \boldsymbol{X}^{\boldsymbol{\alpha}} \,:\, \boldsymbol{\alpha} \in \boldsymbol{P}\} \subset C^\infty(\Lambda) \tag{3.10}$$

the corresponding polynomial function space over Λ that is spanned by the monomials $\boldsymbol{X}^{\boldsymbol{\alpha}}$ ($\boldsymbol{\alpha} \in \boldsymbol{P}$).

Here we used standard multi-index notation:

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2), \quad |\boldsymbol{\alpha}| = \alpha_1 + \alpha_2, \quad \boldsymbol{X}^{\boldsymbol{\alpha}} = x^{\alpha_1} y^{\alpha_2}, \quad \boldsymbol{h}^{\boldsymbol{\alpha}} = h_x^{\alpha_1} h_y^{\alpha_2}, \quad \boldsymbol{D}^{\boldsymbol{\alpha}} = \frac{\partial^{\alpha_1}}{\partial x^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial y^{\alpha_2}}.$$

The hull $\overline{\boldsymbol{P}}$ of $\boldsymbol{P}$ is the set

$$\overline{\boldsymbol{P}} := \boldsymbol{P} \cup \{\boldsymbol{\alpha} + \boldsymbol{e}_i \ : \ \boldsymbol{\alpha} \in \boldsymbol{P},\ i = 1, 2\},$$

where $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ denotes the canonical basis of $\mathbb{R}^2$.

Associated with a set of multi-indices $\boldsymbol{P}$ with $\mathbf{0} := (0,0) \in \boldsymbol{P}$ and $1 \le p \le \infty$ we introduce a norm and a semi-norm on the reference domain Λ:

$$\|v\|^p_{\boldsymbol{P},p} := \sum_{\boldsymbol{\alpha}\in\boldsymbol{P}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} v\|^p_{L_p(\Lambda)}, \qquad |v|^p_{\overline{\boldsymbol{P}},p} := \sum_{\boldsymbol{\alpha}\in\overline{\boldsymbol{P}}\setminus\boldsymbol{P}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} v\|^2_{L_p(\Lambda)},$$

with obvious modifications for $p = \infty$. Furthermore, let $H_p^{\boldsymbol{P}}(\Lambda)$ denote the function space

$$H_p^{\boldsymbol{P}}(\Lambda) := \{v \in L^1(\Lambda) \ : \ \|v\|_{\boldsymbol{P},p} < \infty\} \tag{3.11}$$

and let $S(\hat{M})$ be a spline space such that for $v \in S(\hat{M})$ the restrictions $v|_{\hat{T}_i}$ are polynomials, $i = 1, \ldots, \ell$.

The following two Lemmas are taken from [3].

Lemma 16. *Let $\boldsymbol{P}$ be a set of multi-indices. To each $v \in H_p^{\boldsymbol{P}}(\Lambda)$ there exists a unique $q \in \boldsymbol{P}(\Lambda)$ with*

$$\int_\Lambda \boldsymbol{D}^{\boldsymbol{\alpha}}(v - q)\,\mathrm{d}\boldsymbol{X} = 0 \qquad \forall \boldsymbol{\alpha} \in \boldsymbol{P}.$$

For a short and elegant proof see [3, Lemma 1]. The argument is a slight extension from the well-known Bramble-Hilbert theory.

Lemma 17. *Let $\boldsymbol{P}$ be a set of multi-indices with $\mathbf{0} \in \boldsymbol{P}$. Then there exists a constant C independent of v such that*

$$\|v\|_{\overline{\boldsymbol{P}},p} \le C|v|_{\overline{\boldsymbol{P}},p}$$

for all $v \in H^{\overline{\boldsymbol{P}}}(\Lambda)$ with $\int_\Lambda \boldsymbol{D}^{\boldsymbol{\alpha}} v\,\mathrm{d}\boldsymbol{X} = 0$ for $\boldsymbol{\alpha} \in \boldsymbol{P}$.

An indirect proof can be found in [3, Lemma 2]. It relies on the compactness of a certain embedding, extending a similar result from Bramble and Hilbert.

The next Lemma is an adaptation of [3, Lemma 3] to our patchwise setting.

Lemma 18. *Let $\boldsymbol{\gamma}$ be a multi-index, $I : C^\mu(\Lambda) \to S(\hat{M}) \subset H_p^{\boldsymbol{P}+\boldsymbol{\gamma}}(\Lambda)$, $\mu \in \mathbb{N}$ be a linear operator and let $\boldsymbol{Q}$ be a set of multi-indices with $\mathbf{0} \in \boldsymbol{Q}$ and $\boldsymbol{P} \subset \overline{\boldsymbol{Q}}$. Assume that there are linear functionals $F_i \in \left(H_p^{\overline{\boldsymbol{Q}}}(\Lambda)\right)'$, $i = 1, \ldots, j = \dim \boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$, with the properties*

$$\begin{cases} F_i(\boldsymbol{D}^{\boldsymbol{\gamma}} I u) = F_i(\boldsymbol{D}^{\boldsymbol{\gamma}} u), \qquad i = 1, \ldots, j, \qquad \forall u \in C^\mu(\Lambda) \cap H_p^{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma}}(\Lambda), \\ \big(F_i(\boldsymbol{D}^{\boldsymbol{\gamma}} s) = 0 \quad \text{for } i = 1, \ldots, j\big) \quad \Rightarrow \quad \boldsymbol{D}^{\boldsymbol{\gamma}} s = 0 \qquad \forall s \in S(\hat{M}). \end{cases} \tag{3.12}$$

Then there exists a constant C independent of u such that

$$\|u - Iu\|_{\boldsymbol{P}+\boldsymbol{\gamma},p} \le C\left(|u|_{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma},p} + \|q - Iq\|_{\boldsymbol{P}+\boldsymbol{\gamma},p}\right) \qquad \forall u \in C^\mu(\Lambda) \cap H_p^{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma}}(\Lambda), \tag{3.13}$$

where the polynomial $q \in (\boldsymbol{Q} + \boldsymbol{\gamma})(\Lambda)$ is uniquely determined by

$$\int_\Lambda \boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}}(u - q)\,\mathrm{d}\boldsymbol{X} = 0 \qquad \forall \boldsymbol{\alpha} \in \boldsymbol{Q}. \tag{3.14}$$

Proof. By Lemma 16 the polynomial $q \in (\boldsymbol{Q} + \boldsymbol{\gamma})(\Lambda)$ satisfying (3.14) is indeed unique. The triangle inequality gives

$$\|u - Iu\|_{\boldsymbol{P}+\boldsymbol{\gamma},p} \le \|u - q\|_{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma},p} + \|q - Iq\|_{\boldsymbol{P}+\boldsymbol{\gamma},p} + \|I(q - u)\|_{\boldsymbol{P}+\boldsymbol{\gamma},p}. \tag{3.15}$$

Based on (3.12) we observe that $\sum_{i=1}^{j} |F_i(\cdot)|$ is a norm on $\boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$. Hence, norm equivalence in finite dimensional spaces yields for the last term

$$\begin{aligned}\|I(q-u)\|_{\boldsymbol{P}+\boldsymbol{\gamma},p} = \|\boldsymbol{D}^{\boldsymbol{\gamma}} I(q-u)\|_{\boldsymbol{P},p} &\le C \sum_{i=1}^{j} \left|F_i\big(\boldsymbol{D}^{\boldsymbol{\gamma}} I(q-u)\big)\right| \\ &= C \sum_{i=1}^{j} \left|F_i\big(\boldsymbol{D}^{\boldsymbol{\gamma}} (q-u)\big)\right| \le C\|u-q\|_{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma},p}.\end{aligned} \tag{3.16}$$

With (3.14) an application of Lemma 17 gives

$$\|u-q\|_{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma},p} = \left\|\boldsymbol{D}^{\boldsymbol{\gamma}}(u-q)\right\|_{\overline{\boldsymbol{Q}},p} \le C|\boldsymbol{D}^{\boldsymbol{\gamma}} u|_{\overline{\boldsymbol{Q}},p} = C|u|_{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma},p}. \tag{3.17}$$

Collecting (3.15), (3.16) and (3.17) the result follows. □

Remark 13. The estimate (3.13) shows that a macro-element interpolation operator should be designed in such a way that on the macro-element polynomials with a degree as high as possible are reproduced. Ideally, $q = Iq$ for all $q \in (\boldsymbol{Q}+\boldsymbol{\gamma})(\Lambda)$ which leads to the estimate $\|u - Iu\|_{\boldsymbol{P}+\boldsymbol{\gamma}} \le C|u|_{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma}}$ for all $u \in C^{\mu}(\Lambda) \cap H_p^{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma}}(\Lambda)$. Otherwise an additional error component arises due to the inability to reproduce certain polynomials. This is the only difference in comparison with the theory of [3] caused by a triangle inequality with Iq in (3.15). Such an amendment becomes necessary because in general the polynomial $q \notin S(\hat{M})$ does not lie within the spline space.

Definition 3. Since the interpolation operator is usually defined by linear functionals we follow the nomenclature of [3] and will call the F_i from (3.12) *associated functionals (with respect to $\boldsymbol{D}^{\boldsymbol{\gamma}}$).*

3.5 C^1 macro-interpolation on anisotropic tensor product meshes

Before we turn our attention to a rigorous analysis of Π from (3.7) we want to consider a simpler *reduced* operator. By doing so we demonstrate the developed techniques without getting bogged down in details. Moreover, the insight gained into this reduced interpolation operator will prove to be very useful in the analysis of a quasi-interpolation operator.

3.5.1 A reduced macro-element interpolation operator

Let us consider the reference domain $\Lambda \coloneqq [-1,1]^2$ decomposed into the reference macro-element $\hat{M} \coloneqq \{\hat{T}_i\}_{i=1,\dots,4}$, where $\hat{T}_i$ is the intersection of Λ with the ith quadrant, $i = 1,\dots,4$. With the basis functions from (3.6) we introduce the following *reduced macro-element interpolation operator* $\hat{\Pi}^r : C^1(\Lambda) \to S(\hat{M})$ with $S(\hat{M}) \subset \{v \in C^1(\Lambda) \,:\, v|_{\hat{T}_i} \in Q_2,\ i = 1,\dots,4\}$,

$$(\hat{\Pi}^r v)(x,y) = \sum_{i,j\in\{-1,1\}} v(i,j)\hat{\varphi}_{i,j}(x,y) + v_x(i,j)\hat{\phi}_{i,j}(x,y) + v_y(i,j)\hat{\chi}_{i,j}(x,y), \quad (x,y) \in \Lambda. \tag{3.18}$$

In comparison to Π from (3.7) we discard the basis functions associated with the mixed derivative. Since $\hat{\Pi}$ maps a sufficiently smooth function into $C^1(\Lambda)$, as was shown in [31], we observe for $v \in C^1(\Lambda)$ that

$$\hat{\Pi}^r v = \hat{\Pi}\left(\hat{\Pi}^r v\right) \in C^1(\Lambda).$$

Hence, indeed $S(\hat{M}) \subset \{v \in C^1(\Lambda) \,:\, v|_{\hat{T}_i} \in Q_2,\ i = 1,\dots,4\} \subset W_{2,p}(\Lambda)$. Let us fix $\boldsymbol{\gamma} = (1,0)$. If we seek to apply Lemma 18 to this setting we need to find eight associated functionals F_i, $i = 1,\dots,8$, since

$$\begin{aligned}\boldsymbol{D}^{(1,0)} S(\hat{M}) = \operatorname{span}\{&\hat{\varphi}'_{-1}(x)\hat{\varphi}_{-1}(y),\ \hat{\varphi}'_{-1}(x)\hat{\varphi}_1(y),\ \hat{\psi}'_{-1}(x)\hat{\varphi}_{-1}(y),\ \hat{\psi}'_1(x)\hat{\varphi}_{-1}(y), \\ &\hat{\psi}'_{-1}(x)\hat{\varphi}_1(y),\ \hat{\psi}'_1(x)\hat{\varphi}_1(y),\ \hat{\varphi}'_{-1}(x)\hat{\psi}_{-1}(y),\ \hat{\varphi}'_{-1}(x)\hat{\psi}_1(y)\}\end{aligned} \tag{3.19}$$

is an eight-dimensional space. Setting $\boldsymbol{P} = \boldsymbol{Q} := \{(0,0),(0,1),(1,0)\}$ these functionals must be members of $(W_{2,p}(\Lambda))'$. For $i = 1,\ldots,4$ let V_i denote the four vertices of Λ. Then for $v \in W_{2,p}(\Lambda)$ we find

$$F_i(v) := v(V_i), \quad i = 1,\ldots,4$$

with $|F_i(v)| \leq C\|v\|_{W_{2,p}(\Lambda)}$, $i = 1,\ldots,4$ due to the well known Sobolev embedding $W_{2,p}(\Lambda) \hookrightarrow C(\Lambda)$ in two dimensions. Moreover, for $u \in H_p^{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma}}(\Lambda)$, i.e. $u_x \in W_{2,p}(\Lambda)$ one has

$$F_i(\boldsymbol{D}^{\boldsymbol{\gamma}} u) = u_x(V_i) = F_i(\boldsymbol{D}^{\boldsymbol{\gamma}} \hat{\Pi}^r u), \quad i = 1,\ldots,4.$$

The other four associated functionals are defined on the edges $E_1 := \{(x,-1) : |x| \leq 1\}$ and $E_2 := \{(x,1) : |x| \leq 1\}$ of Λ which are parallel to the x-axis. In fact, they are the mean value and the mean value of the normal derivative:

$$F_{4+i}(v) := \frac{1}{2}\int_{E_i} v(s)\mathrm{d}s, \qquad F_{6+i}(v) := \frac{1}{2}\int_{E_i} v_y(s)\mathrm{d}s, \quad i = 1,2.$$

By well known trace theorems $|F_i(v)| \leq C\|v\|_{W_{2,p}(\Lambda)}$, $i = 5,\ldots,8$ (see e.g. [2] and the references cited in Section 1.3) and

$$F_5(\boldsymbol{D}^{\boldsymbol{\gamma}} u) = \frac{1}{2}\int_{E_1} u_x(s)\mathrm{d}s = \frac{1}{2}\big(u(V_2) - u(V_1)\big) = \frac{1}{2}\int_{E_1} \boldsymbol{D}^{\boldsymbol{\gamma}}\left(\hat{\Pi}^r u\right)(s)\mathrm{d}s = F_5(\boldsymbol{D}^{\boldsymbol{\gamma}}\hat{\Pi}^r u),$$
$$F_7(\boldsymbol{D}^{\boldsymbol{\gamma}} u) = \frac{1}{2}\int_{E_1} u_{xy}(s)\mathrm{d}s = \frac{1}{2}\big(u_y(V_2) - u_y(V_1)\big) = \frac{1}{2}\int_{E_1} \boldsymbol{D}^{\boldsymbol{\gamma}}\left(\hat{\Pi}^r u\right)_y(s)\mathrm{d}s = F_7(\boldsymbol{D}^{\boldsymbol{\gamma}}\hat{\Pi}^r u).$$

Similarly, these identities can be shown to hold true for F_6 and F_8.

Next we show that the functionals F_i define a norm in $\boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$. For this purpose let $u \in \boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$ with $F_i(u) = 0$ for $i = 1,\ldots,8$. Based on the relations

$$\hat{\psi}_i'(k)\hat{\varphi}_j(\ell) = \delta_{ik}\delta_{j\ell}, \qquad i,j,k,\ell \in \{-1,1\}$$

and $v(\pm 1, \pm 1) = 0$ for all other basis functions v of $\boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$ in (3.19) we find that

$$u \in \operatorname{span}\{\hat{\varphi}_{-1}'(x)\hat{\varphi}_{-1}(y),\ \hat{\varphi}_{-1}'(x)\hat{\varphi}_1(y),\ \hat{\varphi}_{-1}'(x)\hat{\psi}_{-1}(y),\ \hat{\varphi}_{-1}'(x)\hat{\psi}_1(y)\}.$$

Out of these remaining four basis functions only $\hat{\varphi}_{-1}'(x)\hat{\varphi}_{-1}(y)$ is non-trivial on the edge E_1. Similarly, only $\hat{\varphi}_{-1}'(x)\hat{\varphi}_1(y)$ has values different from zero on E_2. Moreover, these values are all not positive. Since the mean values $F_5(u) = F_6(u) = 0$ of u vanishes on these edges we conclude that

$$u \in \operatorname{span}\{\hat{\varphi}_{-1}'(x)\hat{\psi}_{-1}(y), \hat{\varphi}_{-1}'(x)\hat{\psi}_1(y)\}.$$

The remaining two basis functions are treated in the same way: while the normal derivative of $\hat{\varphi}_{-1}'(x)\hat{\psi}_{-1}(y)$ on the Edge E_1 is non-trivial and not positive we find $\hat{\varphi}_{-1}'(x)\hat{\psi}_1'(y) \equiv 0$ on E_1. On the edge E_2 the relations are exactly the other way round. Hence, $u \equiv 0$.

An application of Lemma 18 yields

$$\|(u - \hat{\Pi}^r u)_x\|_{W_{1,p}(\Lambda)} \leq C\left(|u_x|_{W_{2,p}(\Lambda)} + \|(q - \hat{\Pi}^r q)_x\|_{W_{1,p}(\Lambda)}\right), \tag{3.20}$$

for all $u \in C^1(\Lambda) \cap H_p^{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma}}(\Lambda)$. The latter means that $u_x \in W_{2,p}(\Lambda)$. The polynomial q is determined by (3.14) and we want to estimate the second error component of (3.20) containing it. Obviously, $q \in \big(\boldsymbol{Q} + \boldsymbol{\gamma}\big)(\Lambda)$ has a representation of the form

$$q(x,y) = q_1 x + q_2 xy + q_3 x^2 \qquad (x,y) \in \Lambda.$$

Here the coefficients $q_i \in \mathbb{R}$, $i = 1,2,3$ are determined by u. A direct calculation shows that the function $(x,y) \mapsto x$ is invariant under interpolation:

$$\hat{\Pi}^r x = \big(\hat{\varphi}_{1,-1}(x,y) + \hat{\varphi}_{1,1}(x,y)\big) - \big(\hat{\varphi}_{-1,-1}(x,y) + \hat{\varphi}_{-1,1}(x,y)\big) + \sum_{i,j\in\{-1,1\}} \hat{\phi}_{i,j}(x,y) \tag{3.21}$$
$$= \big(\hat{\varphi}_1(x) - \hat{\varphi}_{-1}(x) + \hat{\psi}_{-1}(x) + \hat{\psi}_1(x)\big)\big(\hat{\varphi}_{-1}(y) + \hat{\varphi}_1(y)\big) = x. \tag{3.22}$$

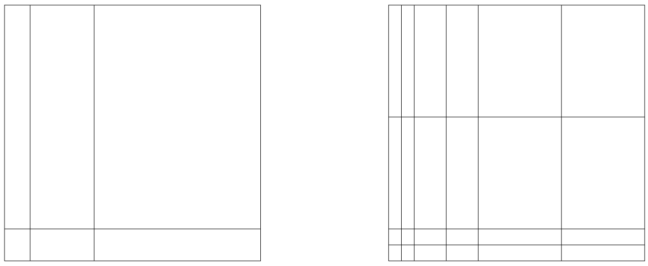

Figure 3.4: macro-element mesh $\mathcal{M}$ (left) and element mesh $\mathcal{T}$ (right).

Similarly, the function $(x, y) \mapsto x^2$ is preserved by the interpolation operator on the macro-element, i.e. $\hat{\Pi}^r(x^2) = x^2$. From (3.14) with $\boldsymbol{\alpha} = (0,1)$ we determine $q_2 = \frac{1}{4}\int_\Lambda u_{xy}(x,y)\,\mathrm{d}x\mathrm{d}y$, hence

$$\|(q - \hat{\Pi}^r q)_x\|_{W_{1,p}(\Lambda)} = |q_2|\,\|(xy - \hat{\Pi}^r xy)_x\|_{W_{1,p}(\Lambda)} \le C\left|\int_\Lambda u_{xy}(x,y)\,\mathrm{d}x\mathrm{d}y\right|. \tag{3.23}$$

Collecting (3.20) and (3.23) we arrive at

$$\|(u - \hat{\Pi}^r u)_x\|_{W_{1,p}(\Lambda)} \le C\left(|u_x|_{W_{2,p}(\Lambda)} + \left|\int_\Lambda u_{xy}(x,y)\,\mathrm{d}x\mathrm{d}y\right|\right), \tag{3.24}$$

for all $\forall u \in C^1(\Lambda) \cap H_p^{\overline{\boldsymbol{Q}}+\boldsymbol{\gamma}}(\Lambda)$.

Remark 14. For $\boldsymbol{\gamma} = (1,1)$ one can choose

$$F_i(v) := \int_{E_i} v\,\mathrm{d}s \quad i = 1,\dots,4 \qquad \text{and} \qquad F_5(v) = \int_\Lambda v\,\mathrm{d}s$$

as associated functionals. Here E_i denotes the ith edge of Λ, $i = 1,\dots,4$. In fact, it is easy to show that $\sum_{i=1}^5 |F_i(\cdot)|$ is a norm on

$$\boldsymbol{D}^{(1,1)}S(\hat{M}) = \operatorname{span}\{\hat{\varphi}'_{-1}(x)\hat{\varphi}'_{-1}(y), \hat{\psi}'_{-1}(x)\hat{\varphi}'_{-1}(y), \hat{\psi}'_1(x)\hat{\varphi}'_{-1}(y), \hat{\varphi}'_{-1}(x)\hat{\psi}'_{-1}(y), \hat{\varphi}'_{-1}(x)\hat{\psi}'_1(y)\}$$

and that $F_i(\boldsymbol{D}^{(1,1)}\hat{\Pi}^r u) = F_i(\boldsymbol{D}^{(1,1)}u)$ for $i = 1,\dots,5$. Moreover, $F_i \in \big(W_{1,p}(\Lambda)\big)'$:

$$\big|F_i(v)\big| = \left|\int_{E_i} v\,\mathrm{d}s\right| \le \|v\|_{L_1(E_i)} \le C\|v\|_{W_{1,1}(\Lambda)} \le C\|v\|_{W_{1,p}(\Lambda)} \quad i = 1,\dots,4,$$
$$\big|F_5(v)\big| = \left|\int_\Lambda v\,\mathrm{d}x\mathrm{d}y\right| \le C\|v\|_{L_p(\Lambda)} \le C\|v\|_{W_{1,p}(\Lambda)},$$

based on Sobolev embeddings and Hölder's inequality. Hence,

$$\|(u - \hat{\Pi}^r u)_{xy}\|_{L_p(\Lambda)} \le C\left(|u_{xy}|_{W_{1,p}(\Lambda)} + \left|\int_\Lambda u_{xy}(x,y)\,\mathrm{d}x\mathrm{d}y\right|\right) \tag{3.25}$$

Now, let $\mathcal{M}$ be a tensor product mesh of Ω. We shall refer to $\mathcal{M}$ as the macro-element mesh and do not require it to be quasi-uniform, i.e. there are no restrictions on the element sizes of the underlaying 1D-triangulations $\mathcal{M}_x$ and $\mathcal{M}_y$ of the macro-element mesh. We obtain the element mesh $\mathcal{T}$ as the tensor product mesh of the two 1D-triangulations that are generated by subdividing every element of $\mathcal{M}_x$ and $\mathcal{M}_y$ uniformly into two elements of equal size. The choice of the midpoint as transition point of a macro-element M is not significant. The theory can handle any subdivision such that the elements within one macro are comparable in size. However, it simplifies the presentation. See Figure 3.4 for a graphical representation of $\mathcal{M}$ and $\mathcal{T}$.

Let $M \in \mathcal{M}$ be the macro-element $M = [x_0 - h_1, x_0 + h_1] \times [y_0 - h_2, y_0 + h_2]$. Note that M consists out of the four elements of $\mathcal{T}$ that share the vertex (x_0, y_0). Introducing the reference mapping F_M from $[-1,1]^2$ to M by

$$x = x_0 + h_1\hat{x}, \qquad y = y_0 + h_2\hat{y}, \tag{3.26}$$

we obtain anisotropic error estimates for the macro-interpolation operator $\Pi^r u := \hat{\Pi}^r\hat{u} \circ F_M^{-1}$ with $\hat{u} := u \circ F_M$ on M.

Theorem 19. *Associated with the shape of the macro-element M let $\boldsymbol{h} := (h_1, h_2)$. For $u \in C^1(M)$ with $u_x \in W_{2,p}(M)$ we have the estimate*

$$\| (u - \Pi^r u)_x \|_{L_p(M)} \leq C \left(\sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{h}^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} u_x\|_{L_p(M)} + h_2 \left| \int_M u_{xy}(x,y)\,\mathrm{d}x\mathrm{d}y \right| \right). \tag{3.27}$$

Proof. The proof uses change of variables, the result (3.24) on the reference macro-element and the relation $\boldsymbol{D}^{\boldsymbol{\alpha}} = \boldsymbol{h}^{-\boldsymbol{\alpha}} \hat{\boldsymbol{D}}^{\boldsymbol{\alpha}}$:

$$\begin{aligned} \|\boldsymbol{D}^{(1,0)} (u - \Pi^r u)\|^p_{L_p(M)} &= h_1^{-p} \left\| \hat{\boldsymbol{D}}^{(1,0)} \left(\hat{u} - \hat{\Pi}^r \hat{u} \right) \right\|^p_{L_p(\Lambda)} h_1 h_2 \\ &\leq C h_1^{-p} \left(\left| \hat{\boldsymbol{D}}^{(1,0)} \hat{u} \right|^p_{W_{2,p}(\Lambda)} + \left| \int_\Lambda \hat{u}_{xy}(\hat{x},\hat{y})\,\mathrm{d}\hat{x}\mathrm{d}\hat{y} \right|^p \right) h_1 h_2 \\ \leq C h_1^{-p} \Bigg(\sum_{|\boldsymbol{\alpha}|=2} h_1^p \boldsymbol{h}^{p\boldsymbol{\alpha}} &\left\| \boldsymbol{D}^{\boldsymbol{\alpha}} \boldsymbol{D}^{(1,0)} u \right\|^p_{L_p(M)} + h_1^p h_2^p \left| \int_M \boldsymbol{D}^{(1,1)} u(x,y)\,\mathrm{d}x\mathrm{d}y \right|^p \Bigg). \end{aligned}$$

Which is the desired estimate. In the case $p = \infty$ some minor modifications are needed. □

Remark 15. The diagonal form of the affine reference mapping F_M according to (3.26) is needed for affine equivalence of the interpolation operator. Note that only in this case Q_k elements are affine equivalent.

Remark 16. While for functions $u \in C^1(M)$ with $u_x \in W_{2,p}(M)$ the reduced interpolation operator Π^r is not of second order in the $W_{1,p}$ semi-norm it is of optimal second order if additionally the mean value of the mixed derivative u_{xy} vanishes on M. Clearly, this reduction in approximation ability corresponds to discarding the basis functions $\hat{\psi}_{\pm1,\pm1}$ in (3.18).

Remark 17. Similarly, one can can deduce from the result in Remark 14 that

$$\|(u - \Pi^r u)_{xy}\|_{L_p(M)} \leq C \left(\sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{h}^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} u_{xy}\|_{L_p(M)} + \left| \int_M u_{xy}(x,y)\,\mathrm{d}x\mathrm{d}y \right| \right).$$

Clearly, this result is in general unsatisfactory. The inability to yield anisotropic interpolation error estimates for second order derivatives of the approximation error is caused by discarding the basis functions corresponding to the mixed derivative.

3.5.2 The full $C^1 - Q_2$ interpolation operator

As a second example we want to consider the interpolation operator $\hat{\Pi}$ of (3.7). We refer to it as *full* not only to contrast it from the reduced operator in the previous subsection but also to underline the property (3.8) of its underlaying macro-element space. Since this operator is closely related to interpolation on the bicubic C^1 Bogner-Fox-Schmidt element, we shall first give a result from the literature. To the knowledge of the author there exists only one paper dealing with anisotropic interpolation error estimates for this element. In [13] the authors derive the result

$$\left\| \hat{\boldsymbol{D}}^{\boldsymbol{\gamma}} \left(\hat{u} - \hat{I}_{12} \hat{u} \right) \right\|_{0,\hat{K}} \leq C \left| \hat{\boldsymbol{D}}^{\boldsymbol{\gamma}} \hat{u} \right|_{4-|\boldsymbol{\gamma}|,\hat{K}}, \tag{3.28}$$

for $|\boldsymbol{\gamma}| \leq 2$ and $\hat{u} \in H^4(\hat{K})$ on the reference element $\hat{K} := [0,1]^2$. Here I_{12} is the analogue of $\hat{\Pi}$ in the space of bicubic polynomials, i.e. the Lagrangian basis functions in (3.7) have to be replaced by bicubic polynomials satisfying the same (duality and Kronecker) relations. Using affine transformation this result can be extended to

$$\|\boldsymbol{D}^{\boldsymbol{\gamma}} (u - I_{12} u)\|_{0,K} \leq C \sum_{|\boldsymbol{\alpha}|=4-|\boldsymbol{\gamma}|} \boldsymbol{h}^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} \boldsymbol{D}^{\boldsymbol{\gamma}} u\|_{0,K}, \tag{3.29}$$

for $|\boldsymbol{\gamma}| \leq 2$ and $u \in H^4(K)$ on a rectangular element K with sides aligned to the coordinate axes and $\boldsymbol{h} = (h_1, h_2)$ with edge lengths h_i, $i = 1, 2$.

However, in [13] the theory of Apel [3, 2] is not used to obtain this result. Instead a new interpolation operator $\hat{L}_1$ is introduced such that $\hat{L}_1(\hat{\boldsymbol{D}}^{\gamma}\hat{u}) = \hat{\boldsymbol{D}}^{\gamma}\hat{I}_{12}\hat{u}$ and standard interpolation theory is applied to obtain a bound for the interpolation error of $\hat{L}_1$. Since we are dealing with only piecewise polynomials this path is blocked for us. A spin-off of our discussion will be how the results (3.28) and (3.29) can be obtained using Apel's theory. The key is to recognize that divided differences can be used as associated functionals. Since we need certain Sobolev embeddings, we focus on the case $p = 2$, which also appears to be the most important one with respect to applications.

Inspired by [13] we generalize the Newtonian representation (3.4) to two dimensions obtaining

$$\begin{aligned}\big(\hat{\Pi}u\big)(x,y) = &\sum_{i=1}^{3}\sum_{j=1}^{3} F_{i,j}(u)(x+1)^{i-1}(y+1)^{j-1} + 4\sum_{j=1}^{3} F_{4,j}(u)\hat{\psi}_1(x)(y+1)^{j-1} \\ &+4\sum_{i=1}^{3} F_{i,4}(u)(x+1)^{i-1}\hat{\psi}_1(y) + 16F_{4,4}(u)\hat{\psi}_1(x)\hat{\psi}_1(y).\end{aligned} \tag{3.30}$$

Here the 16 functionals $F_{i,j}$, $i,j = 1,\dots,4$ are two-dimensional divided differences with multiple knots, see e.g. [55]. If we define a sorted node sequence by

$$n_i = \begin{cases} -1 & \text{for } i = 1, \\ -1,-1 & \text{for } i = 2, \\ -1,-1,1 & \text{for } i = 3, \\ -1,-1,1,1 & \text{for } i = 4, \end{cases}$$

then $F_{i,j}(u) \coloneqq u[n_i; n_j]$ is the divided difference of order $i-1$ to x and order $j-1$ to y:

Definition 4. For a fixed $y \in [-1,1]$ let

$$u_{n_i}(y) \coloneqq u(\cdot, y)[n_i]$$

denote the parametrized one dimensional divided difference (with respect to x and the node sequence n_i). Then the two dimensional divided difference $u[n_i; n_j]$ is defined by

$$u[n_i; n_j] \coloneqq u_{n_i}[n_j].$$

Remark 18. Because of $u[n_i; n_j] = \big(u(x,\cdot)[n_j]\big)[n_i]$ one can start with the evaluation in y, as well.

We find that

$$F_{i,j}(u) = u[\overbrace{-1,\dots,-1}^{i \text{ times}}; \overbrace{-1,\dots,-1}^{j \text{ times}}] = \boldsymbol{D}^{(i-1,j-1)}u(-1,-1), \qquad i,j = 1,2. \tag{3.31}$$

Moreover, using (3.5) for instance

$$\begin{aligned} F_{3,1} &= u[-1,-1,1;-1] = \frac{1}{4}u(1,-1) - \frac{1}{4}u(-1,-1) - \frac{1}{2}u_x(-1,-1), \\ F_{4,1} &= u[-1,-1,1,1;-1] = \frac{1}{4}u(-1,-1) - \frac{1}{4}u(1,-1) + \frac{1}{4}u_x(-1,-1) + \frac{1}{4}u_x(1,-1), \\ F_{3,2} &= u[-1,-1,1;-1,-1] = \frac{1}{4}u_y(1,-1) - \frac{1}{4}u_y(-1,-1) - \frac{1}{2}u_{xy}(-1,-1), \\ F_{4,2} &= u[-1,-1,1,1;-1,-1] = \frac{1}{4}u_y(-1,-1) - \frac{1}{4}u_y(1,-1) + \frac{1}{4}u_{xy}(-1,-1) + \frac{1}{4}u_{xy}(1,-1). \end{aligned}$$

Similarly, the other divided differences can be calculated, e.g.

$$\begin{aligned} F_{3,3} = u[-1,-1,1;-1,-1,1] = &\frac{1}{4}\left(\frac{1}{4}u(1,1) - \frac{1}{4}u(-1,1) - \frac{1}{2}u_x(-1,1)\right) \\ &- \frac{1}{4}\left(\frac{1}{4}u(1,-1) - \frac{1}{4}u(-1,-1) - \frac{1}{2}u_x(-1,-1)\right) \\ &- \frac{1}{2}\left(\frac{1}{4}u_y(1,-1) - \frac{1}{4}u_y(-1,-1) - \frac{1}{2}u_{xy}(-1,-1)\right). \end{aligned}$$

Obviously, all divided differences $F_{i,j}$, $i, j = 1, \ldots, 4$ can be expressed as linear combinations of the interpolation data $\{u(\pm 1, \pm 1), u_x(\pm 1, \pm 1), u_y(\pm 1, \pm 1), u_{xy}(\pm 1, \pm 1)\}$.

In contrast to Subsection 3.5.1 we want to consider an arbitrary multi-index $\boldsymbol{\gamma}$ with $|\boldsymbol{\gamma}| \leq 2$ here. Consequently, certain sets and functionals depend on the specific choice of $\boldsymbol{\gamma}$ and we emphasize this by using the additional subscript or superscript $\boldsymbol{\gamma}$. By applying the differential operator $\boldsymbol{D}^{\boldsymbol{\gamma}}$ to the representation (3.30) we observe that the space $\boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$ can be normed by $\sum_{(i,j) \in J_{\boldsymbol{\gamma}}} |F_{i,j}(\cdot)|$. Here

$$J_{(\gamma_1, \gamma_2)} = \{(i, j) \,:\, i = \gamma_1 + 1, \ldots, 4 \text{ and } j = \gamma_2 + 1, \ldots, 4\}.$$

Note also that by construction $\dim \boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M}) = |J_{\boldsymbol{\gamma}}|$. We want to apply Lemma 18 with $\boldsymbol{Q} \coloneqq \{\boldsymbol{\alpha} \,:\, |\boldsymbol{\alpha}| \leq 3 - |\boldsymbol{\gamma}|\}$. In order to establish that $F_{i,j}$ with $(i, j) \in J_{\boldsymbol{\gamma}}$ are associated functionals according to (3.12) we have to show that the divided differences $F_{i,j}$ can be interpreted as the application of a linear functional $F^{\boldsymbol{\gamma}}_{i,j} \in \left(H^{4-|\boldsymbol{\gamma}|}(\Lambda)\right)'$ on the derivative $\boldsymbol{D}^{\boldsymbol{\gamma}} u$ such that

$$F^{\boldsymbol{\gamma}}_{i,j}(\boldsymbol{D}^{\boldsymbol{\gamma}} \hat{\Pi} u) = F^{\boldsymbol{\gamma}}_{i,j}(\boldsymbol{D}^{\boldsymbol{\gamma}} u), \qquad (i, j) \in J_{\boldsymbol{\gamma}}, \qquad \forall u \in \{v \in C^2(\Lambda) \,:\, \boldsymbol{D}^{\boldsymbol{\gamma}} v \in H^{4-|\boldsymbol{\gamma}|}(\Lambda)\}.$$

Following [13] we reinterpret (3.31) in the form

$$F_{i,j}(u) = \boldsymbol{D}^{(i-1-\gamma_1, j-1-\gamma_2)} \boldsymbol{D}^{\boldsymbol{\gamma}} u(-1, -1) =: F^{\boldsymbol{\gamma}}_{i,j}(\boldsymbol{D}^{\boldsymbol{\gamma}} u), \qquad (i, j) \in J_{\boldsymbol{\gamma}},\ i, j = 1, 2.$$

Since all these $F_{i,j}$ can be expressed as linear combinations of the interpolation data we have

$$F^{\boldsymbol{\gamma}}_{i,j}(\boldsymbol{D}^{\boldsymbol{\gamma}} u) = F_{i,j}(u) = F_{i,j}(\hat{\Pi} u) = F^{\boldsymbol{\gamma}}_{i,j}(\boldsymbol{D}^{\boldsymbol{\gamma}} \hat{\Pi} u) \tag{3.32}$$

for $(i, j) \in J_{\boldsymbol{\gamma}}$, $i, j = 1, 2$. Moreover, from a standard Sobolev embedding $H^{4-|\boldsymbol{\gamma}|} \hookrightarrow C^{2-|\boldsymbol{\gamma}|}$ for $v \in H^{4-|\boldsymbol{\gamma}|}(\Lambda)$

$$\left|F^{\boldsymbol{\gamma}}_{i,j}(v)\right| \leq \left|\boldsymbol{D}^{(i-1-\gamma_1, j-1-\gamma_2)} v(-1, -1)\right| \leq C \|v\|_{4-|\boldsymbol{\gamma}|}, \qquad (i, j) \in J_{\boldsymbol{\gamma}},\ i, j = 1, 2.$$

For the other divided differences we need some kind of Peano form which was developed in [13]. In fact, replacing $u(1)$ and $u'(1)$ in (3.5) by the Taylor expansions

$$u(1) = u(-1) + 2u'(-1) + \int_{-1}^{1} (1 - x) u''(x) \,\mathrm{d}x \qquad \text{and} \qquad u'(1) = u'(-1) + \int_{-1}^{1} u''(x) \,\mathrm{d}x$$

one obtains

$$u[-1, -1, \overbrace{1 \ldots 1}^{i \text{ times}}] = \int_{-1}^{1} s_i(x) u''(x) \mathrm{d}x \qquad \text{with} \qquad s_i(x) = \begin{cases} (1 - x)/4, & \text{for } i = 1, \\ x/4, & \text{for } i = 2. \end{cases} \tag{3.33}$$

With respect to (3.32) it is important to note that this identity does not only hold for $C^2([-1, 1])$ functions but also for the quadratic C^1 splines considered as can be checked by examining all the basis functions $\hat{\varphi}_{\pm 1}$ and $\hat{\psi}_{\pm 1}$, for instance

$$\hat{\varphi}_{-1}[-1, -1, 1] = -\frac{1}{4} = \int_{-1}^{1} \frac{(1 - x)}{4} \hat{\varphi}''_{-1}(x) \mathrm{d}x.$$

Clearly, $|s_i(x)| \leq \frac{1}{2}$ for $x \in [-1, 1]$ and $i = 1, 2$. Hence, one can conclude that

$$\begin{aligned} F_{i,j}(u) = u[n_i; n_j] &= \int_{-1}^{1} s_{i-2}(x) \boldsymbol{D}^{(2, j-1)} u(x, -1) \,\mathrm{d}x \\ &= \int_{-1}^{1} s_{i-2}(x) \boldsymbol{D}^{(2-\gamma_1, j-1-\gamma_2)} \boldsymbol{D}^{\boldsymbol{\gamma}} u(x, -1) \,\mathrm{d}x =: F^{\boldsymbol{\gamma}}_{i,j}(\boldsymbol{D}^{\boldsymbol{\gamma}} u), \end{aligned}$$

for $(i, j) \in J_{\boldsymbol{\gamma}}$, $i = 3, 4$, $j = 1, 2$ and with

$$\left|F^{\boldsymbol{\gamma}}_{i,j}(v)\right| \leq \frac{1}{2} \int_{-1}^{1} \left|\boldsymbol{D}^{(2-\gamma_1, j-1-\gamma_2)} v(x, -1)\right| \,\mathrm{d}x \leq C \|v\|_{4-|\boldsymbol{\gamma}|}, \qquad (i, j) \in J_{\boldsymbol{\gamma}},\ i = 3, 4,\ j = 1, 2.$$

for $v \in H^{4-|\boldsymbol{\gamma}|}(\Lambda)$ by a trace theorem (c.p. [13]). Note that the identities in (3.32) hold true for $(i,j) \in J_{\boldsymbol{\gamma}}$, $i=3,4$, $j=1,2$, as well.

In exactly the same manner we treat the functionals $F_{i,j}$ with $(i,j) \in J_{\boldsymbol{\gamma}}$, $i=1,2$, $j=3,4$:

$$\begin{aligned} F_{i,j}(u) = u[n_i;n_j] &= \int_{-1}^{1} s_{j-2}(y) \boldsymbol{D}^{(i-1,2)} u(-1,y)\,\mathrm{d}x \\ &= \int_{-1}^{1} s_{j-2}(y) \boldsymbol{D}^{(i-1-\gamma_1,2-\gamma_2)} \boldsymbol{D}^{\boldsymbol{\gamma}} u(-1,y)\,\mathrm{d}y =: F_{i,j}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} u). \end{aligned}$$

Using the same argument as before it is easy to obtain $\left|F_{i,j}^{\boldsymbol{\gamma}}(v)\right| \le \|v\|_{4-|\boldsymbol{\gamma}|}$ and (3.32) for $(i,j) \in J_{\boldsymbol{\gamma}}$, $i=1,2$, $j=3,4$.

Finally, we consider $F_{i,j}$ for $i,j=3,4$.

$$\begin{aligned} F_{i,j}(u) = u[n_i;n_j] &= \int_{\Lambda} s_{i-2}(x) s_{j-2}(y) \boldsymbol{D}^{(2,2)} u(x,y)\,\mathrm{d}x\mathrm{d}y \\ &= \int_{\Lambda} s_{i-2}(x) s_{j-2}(y) \boldsymbol{D}^{(2-\gamma_1,2-\gamma_2)} \boldsymbol{D}^{\boldsymbol{\gamma}} u(x,y)\,\mathrm{d}x\mathrm{d}y =: F_{i,j}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} u), \qquad i,j=3,4. \end{aligned}$$

Again the functionals can be shown to be bounded. Using the Cauchy Schwarz inequality

$$\left|F_{i,j}^{\boldsymbol{\gamma}}(v)\right| \le \frac{1}{4} \int_{\Lambda} \left|\boldsymbol{D}^{(2-\gamma_1,2-\gamma_2)} v(x,y)\right| \mathrm{d}x\mathrm{d}y \le C \left\|\boldsymbol{D}^{(2-\gamma_1,2-\gamma_2)} v\right\|_0 \le C\|v\|_{4-|\boldsymbol{\gamma}|}$$

for $i,j=3,4$. Additionally, (3.32) holds true for $i,j=3,4$.

Hence, for a given differential operator $\boldsymbol{D}^{\boldsymbol{\gamma}}$ with $|\boldsymbol{\gamma}| \le 2$ we can use $F_{i,j}^{\boldsymbol{\gamma}}$, $(i,j) \in J_{\boldsymbol{\gamma}}$ as associated functionals and Lemma 18 yields for $u \in C^2(\Lambda)$ with $\boldsymbol{D}^{\boldsymbol{\gamma}} u \in H^{4-|\boldsymbol{\gamma}|}(\Lambda)$ that

$$\left\|\boldsymbol{D}^{\boldsymbol{\gamma}}\left(u - \hat{\Pi} u\right)\right\|_{2-|\boldsymbol{\gamma}|} \le C \left(|\boldsymbol{D}^{\boldsymbol{\gamma}} u|_{4-|\boldsymbol{\gamma}|} + \left\|\boldsymbol{D}^{\boldsymbol{\gamma}}\left(q - \hat{\Pi} q\right)\right\|_{2-|\boldsymbol{\gamma}|}\right). \tag{3.34}$$

Here the polynomial $q \in (\boldsymbol{Q}+\gamma)(\Lambda)$ is defined by (3.14) where $\boldsymbol{Q} := \{\boldsymbol{\alpha} : |\boldsymbol{\alpha}| \le 3 - |\boldsymbol{\gamma}|\}$. Hence, with $\boldsymbol{X} = (x,y)$,

$$q(x,y) = \sum_{\boldsymbol{\alpha} \in \boldsymbol{Q}} q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}} \boldsymbol{X}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} = \sum_{\substack{\boldsymbol{\alpha} \in \boldsymbol{Q} \\ |\boldsymbol{\alpha}+\boldsymbol{\gamma}|<3}} q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}} \boldsymbol{X}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} + \sum_{\substack{\boldsymbol{\alpha} \in \boldsymbol{Q} \\ |\boldsymbol{\alpha}+\boldsymbol{\gamma}|=3}} q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}} \boldsymbol{X}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}},$$

with coefficients $q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}} = q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}}(u) \in \mathbb{R}$ for $\boldsymbol{\alpha} \in \boldsymbol{Q}$. A simple calculation shows that $\hat{\Pi}(x^{\boldsymbol{\beta}}) = x^{\boldsymbol{\beta}}$ on Λ for all $\boldsymbol{\beta}$ satisfying $|\boldsymbol{\beta}| < 3$. Therefore we obtain with a triangle inequality

$$\left\|\boldsymbol{D}^{\boldsymbol{\gamma}}\left(q - \hat{\Pi} q\right)\right\|_{2-|\boldsymbol{\gamma}|} \le \sum_{\substack{\boldsymbol{\alpha} \in \boldsymbol{Q} \\ |\boldsymbol{\alpha}+\boldsymbol{\gamma}|=3}} |q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}}| \left\|\boldsymbol{D}^{\boldsymbol{\gamma}}\left(\boldsymbol{X}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} - \hat{\Pi}(\boldsymbol{X}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}})\right)\right\|_{2-|\boldsymbol{\gamma}|} \le C \sum_{\substack{\boldsymbol{\alpha} \in \boldsymbol{Q} \\ |\boldsymbol{\alpha}+\boldsymbol{\gamma}|=3}} |q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}}|.$$

Note that the last summation is carried out over multi-indices of highest order for which we observe by (3.14)

$$\begin{aligned} \int_{\Lambda} \boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} q(x,y)\,\mathrm{d}x\mathrm{d}y &= 4(\alpha_1+\gamma_1)!(\alpha_2+\gamma_2)!\, q_{\boldsymbol{\alpha}+\boldsymbol{\gamma}} \\ &= \int_{\Lambda} \boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} u(x,y)\,\mathrm{d}x\mathrm{d}y, \qquad \boldsymbol{\alpha} \in \boldsymbol{Q},\ |\boldsymbol{\alpha}+\boldsymbol{\gamma}| = 3. \end{aligned}$$

Hence,

$$\left\|\boldsymbol{D}^{\boldsymbol{\gamma}}\left(q - \hat{\Pi} q\right)\right\|_{2-|\boldsymbol{\gamma}|} \le C \sum_{|\boldsymbol{\alpha}|=3-|\boldsymbol{\gamma}|} \left|\int_{\Lambda} \boldsymbol{D}^{\boldsymbol{\alpha}} \boldsymbol{D}^{\boldsymbol{\gamma}} u(x,y)\,\mathrm{d}x\mathrm{d}y\right|. \tag{3.35}$$

Collecting (3.34) and (3.35) we obtain another main result.

Theorem 20. *Let $\boldsymbol{\gamma}$ be a multi-index with $|\boldsymbol{\gamma}| \le 2$ and let $\hat{\Pi}$ denote the full C^1-Q_2 interpolation operator defined in (3.7) on the reference macro-element $\hat{M}$. For $u \in H^4(\Lambda)$ we have the estimate*

$$\left\|\boldsymbol{D}^{\boldsymbol{\gamma}}(u - \hat{\Pi} u)\right\|_0 \le C \left(|\boldsymbol{D}^{\boldsymbol{\gamma}} u|_{4-|\boldsymbol{\gamma}|} + \sum_{|\boldsymbol{\alpha}|=3-|\boldsymbol{\gamma}|} \left|\int_{\Lambda} \boldsymbol{D}^{\boldsymbol{\alpha}} \boldsymbol{D}^{\boldsymbol{\gamma}} u(x,y)\,\mathrm{d}x\mathrm{d}y\right|\right).$$

Using an affine mapping we can define Π on a macro-element $M \in \mathcal{M}$ and extend the result like in the proof of Theorem 19.

Corollary 21. *Let $M = [x_0 - h_1, x_0 + h_1] \times [y_0 - h_2, y_0 + h_2]$ be the axis-aligned macro-element that contains the four elements sharing the vertex (x_0, y_0). With the reference mapping F_M from $\Lambda := [-1,1]^2$ to M defined in (3.26) one can introduce the full $C^1 - Q_2$ interpolation operator Π on M by $\Pi u := \hat{\Pi}\hat{u} \circ F_M^{-1}$ with $\hat{u} := u \circ F_M$. Let $\boldsymbol{\gamma}$ be a multi-index with $|\boldsymbol{\gamma}| \leq 2$ and $\boldsymbol{h} = (h_1, h_2)$. Then for $u \in H^4(M)$ we have the estimate*

$$\|\boldsymbol{D}^{\boldsymbol{\gamma}}(u - \Pi u)\|_{0,M} \leq C \left(\sum_{|\boldsymbol{\alpha}|=4-|\boldsymbol{\gamma}|} \boldsymbol{h}^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}}\boldsymbol{D}^{\boldsymbol{\gamma}}u\|_{0,M} + \sum_{|\boldsymbol{\alpha}|=3-|\boldsymbol{\gamma}|} \boldsymbol{h}^{\boldsymbol{\alpha}} \left| \int_M \boldsymbol{D}^{\boldsymbol{\alpha}}\boldsymbol{D}^{\boldsymbol{\gamma}}u(x,y)\,\mathrm{d}x\mathrm{d}y \right| \right). \tag{3.36}$$

Corollary 22. *Let I_{12} be the analogue of Π in the space of bicubic polynomials, i.e. the Lagrangian basis functions in (3.7) are replaced by bicubic polynomials satisfying the same duality and Kronecker relations. Let $\boldsymbol{\gamma}$ be a multi-index with $|\boldsymbol{\gamma}| \leq 2$. Then the estimate (3.29) holds true for all $u \in C^2(M)$ with $\boldsymbol{D}^{\boldsymbol{\gamma}}u \in H^{4-\boldsymbol{\gamma}}(M)$.*

Proof. Replace (3.30) by

$$\begin{aligned}\big(I_{12}u\big)(x,y) &= \sum_{i=1}^{3}\sum_{j=1}^{3} F_{i,j}(u)(x+1)^{i-1}(y+1)^{j-1} + \sum_{j=1}^{3} F_{4,j}(u)(x+1)^2(x-1)(y+1)^{j-1} \\ &+ \sum_{i=1}^{3} F_{i,4}(u)(x+1)^{i-1}(y+1)^2(y-1) + F_{4,4}(u)(x+1)^2(x-1)(y+1)^2(y-1).\end{aligned}$$

Now all arguments carry over to I_{12}. Observe that for this interpolation operator on the reference element we find $I_{12}(\boldsymbol{X}^{\boldsymbol{\beta}}) = \boldsymbol{X}^{\boldsymbol{\beta}}$ for all $\boldsymbol{\beta}$ satisfying $|\boldsymbol{\beta}| \leq 3$. Hence, the additional error component containing q vanishes. □

Remark 19. It is possible to extend this result to Hermite interpolation by polynomials of higher degree as was done in [13]. However, in that paper a different technique is used. By identifying possible associate functionals we enable the analysis of these operators using the unified theory of Apel and Dobrowolski, see [3].

Remark 20. Comparing the estimates (3.29) and (3.36) we see that the macro-interpolation attains in general a lower order than the corresponding element interpolation. This is due to the inability of the macro-interpolation operator to reproduce cubic polynomials.

Remark 21. The reduced macro-interpolation operator Π^r is of even lower order compared to Π and it appears doubtful to obtain anisotropic estimates for second order derivatives of the interpolation error of Π^r. However, it does not rely on so much regularity of the function u to be interpolated. Note that the only difference of $\hat{\Pi}^r$ and $\hat{\Pi}$ is the choice of the functional determining the coefficient of the basis functions $\hat{\psi}_{i,j}$, $i,j = -1,1$, see also Table 3.1.

3.5.3 A $C^1 - Q_2$ macro-element quasi-interpolation operator of Scott-Zhang-type on tensor product meshes

Let us start this subsection by recalling the definition of the Scott-Zhang quasi-interpolation operator Z_h. This operator was designed in order to obtain approximations to functions u that are not sufficiently regular for nodal interpolation, see [71]. For instance, one might wish to approximate non-smooth functions. The basic idea is to use local L_2 projections on certain element edges to specify the coefficients of the approximating finite element function $Z_h u$. In contrast to the well-known Clément quasi-interpolant this approach can grant the projection property and the ability to preserve homogeneous boundary conditions.

Since we only want to demonstrate the basic ideas and fix some notation here we shall only consider the function space V_h of continuous piecewise linears induced by a quasi-uniform partition Ω^h of the polygonal domain $\Omega \subset \mathbb{R}^2$ into triangles. For a more extensive presentation we refer the interested reader to [2, Section 3.2].

	Πu	$\Pi^r u$	$\tilde{\Pi} u$
coefficient of the basis-function corresponding to mixed derivative	$u_{xy}(x_i, y_j)$	0	a_{ij}, see (3.44), non-local
required regularity of u	$H^4(\Omega)$	$H^3(\Omega)$	$H^3(\Omega)$
formal order of the first derivative of the approximation error in L_2	2	1	2
best possible order of the first derivative of the approximation error in L_2	3	2	2
anisotropic estimates for the derivatives $\boldsymbol{D}^\gamma$ of the approximation error	$\lvert\gamma\rvert \le 2$	$\lvert\gamma\rvert \le 1$	$\lvert\gamma\rvert \le 1$

Table 3.1: Comparison of the (quasi-)interpolation operators Π^r (Subsection 3.5.1), Π (Subsection 3.5.2) and $\tilde{\Pi}$ (Subsection 3.5.3) on tensor product meshes.

Let φ_i, $i \in I$ denote the nodal basis functions of V_h, i.e. for any grid node $\boldsymbol{X}_j$, $j \in I$ the piecewise linear function $\varphi_i \in V_h$ satisfies

$$\varphi_i(\boldsymbol{X}_j) = \delta_{ij}. \tag{3.37}$$

Next, for each node $\boldsymbol{X}_i$, $i \in I$ of the mesh we pick an edge σ_i of a mesh triangle such that $\boldsymbol{X}_i \in \sigma_i$. If $\boldsymbol{X}_i \in \partial\Omega$ belongs to the boundary then we further restrict the choice of these edges by demanding $\sigma_i \subset \partial\Omega$. This is essential if one wishes to preserve homogeneous boundary conditions. Now the Scott-Zhang operator is defined by

$$Z_h u(x,y) = \sum_{i \in I} \big(\Pi_{\sigma_i} u\big)(\boldsymbol{X}_i)\varphi_i(x,y), \tag{3.38}$$

where $\Pi_{\sigma_i} : L_2(\sigma_i) \to P_1(\sigma_i)$, $i \in I$ is the local L_2-projection operator. It is easy to see that Z_h inherits the property of being a *projector*; actually, $Z_h v_h = v_h$ for all $v_h \in V_h$.

In order to provide an equivalent but more useful definition of the Scott-Zhang quasi-interpolant $Z_h u$ to u let us assume that σ_i is the straight line connecting the nodes $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ for some $j \in I$. On σ_i let $\psi_i^d \in P_1(\sigma_i)$ denote a *dual basis function*, uniquely determined by

$$\int_{\sigma_i} \psi_i^d \varphi_i \,\mathrm{d}s = 1 \qquad \text{and} \qquad \int_{\sigma_i} \psi_i^d \varphi_j \,\mathrm{d}s = 0. \tag{3.39}$$

Obviously $\Pi_{\sigma_i} u \in P_1(\sigma_i)$ can be represented as a linear combination of the restrictions of φ_i and φ_j to σ_i, i.e.

$$\Pi_{\sigma_i} u = b_i \varphi_i\big|_{\sigma_i} + b_j \varphi_j\big|_{\sigma_i}.$$

with real numbers b_i and b_j still to be specified. Hence, by (3.39) and the definition of Π_{σ_i} one finds that

$$b_i = b_i \int_{\sigma_i} \psi_i^d \varphi_i \,\mathrm{d}s = \int_{\sigma_i} (\Pi_{\sigma_i} u)\psi_i^d \,\mathrm{d}s = \int_{\sigma_i} u\psi_i^d \,\mathrm{d}s. \tag{3.40}$$

Finally, by the Kronecker relation (3.37) it is clear that $\big(\Pi_{\sigma_i} u\big)(\boldsymbol{X}_i) = b_i$. Consequently, with (3.40) and (3.38) one obtains

$$Z_h u(x,y) = \sum_{i \in I} \int_{\sigma_i} u\psi_i^d \,\mathrm{d}s\, \varphi_i(x,y). \tag{3.41}$$

From its representation (3.41) it can be seen that the coefficients of the Scott-Zhang interpolant $Z_h u$ to u are *weighted local averages of* u over σ_i. In fact, the dual basis function ψ_i^d can be interpreted as some weighting function since

$$\int_{\sigma_i} \psi_i^d \,\mathrm{d}s = \int_{\sigma_i} \psi_i^d(\varphi_i + \varphi_j)\,\mathrm{d}s = \int_{\sigma_i} \psi_i^d \varphi_i \,\mathrm{d}s = 1,$$

because of (3.39) and the fact that $\{\varphi_i, \varphi_j\}$ is a partition of unity on σ_i. In this light it is clear that stability and error estimates for $Z_h u$ over an element T will be based on the values of derivatives of u on an entire patch ω_T of elements around T. More precisely, a mesh triangle T_j is a subset of ω_T iff T has a vertex $\boldsymbol{X}_i$ such that $\boldsymbol{X}_i \in \sigma_i \subset T_j$.

Moreover, (3.41) extends the domain of definition. Naturally one would demand that for the function u to be approximated it holds $u \in L_2(\sigma_i)$. However, since one has $\psi_i^d \in L_\infty(\sigma_i)$ for the polynomial dual basis functions ψ_i^d, $i \in I$, it is possible to apply Z_h to any function u such that its trace satisfies $u \in L_1(\sigma_i)$.

Under the assumption of a quasi-uniform mesh Ω^h and for $u \in W_{j,p}(\omega_T)$ the stability estimate

$$|Z_h u|_{W_{k,p}(T)} \leq C h^{-k} \sum_{j=0}^{\ell} h^j |u|_{W_{j,p}(\omega_T)}, \qquad \text{for } p \in [1,\infty],\ 0 \leq k \leq \ell \leq 2,\ \ell \geq 1,$$

can be found for instance in [17]. Next, standard arguments can be used to obtain the error estimate

$$|u - Z_h u|_{W_{k,p}(T)} \leq C h^{\ell-k} |u|_{W_{\ell,p}(\omega_T)}, \qquad \text{for } p \in [1,\infty],\ 0 \leq k \leq \ell \leq 2,\ \ell \geq 1.$$

In [2] the Scott-Zhang operator is studied over anisotropic meshes of tensor product type. It is shown in Theorem 3.1 of that book that for $p \in [1,\infty]$ and some rectangular axis-aligned element T this operator grants a stability estimate and an anisotropic quasi-interpolation error estimate for $\|Z_h u\|_{L_p(T)}$ and $\|u - Z_h u\|_{L_p(T)}$, respectively. Moreover, in [2] one finds a counterexample showing that in general the original Scott-Zhang operator does not provide such an estimate for derivatives of the approximation error. Therefore the original operator is modified in several ways in the Sections 3.3, 3.4 and 3.5 of [2] and anisotropic quasi-interpolation error estimates for the resulting operators are obtained. However, in the entire third chapter of that book it is assumed that there is no abrupt change in the element sizes. This means that while elements are allowed to have an arbitrary aspect ratio h_x/h_y the edge length h_x and h_y have to vary gradually when moving from one element to a neighboring one, see [2, (3.4) on page 100]. Clearly, this assumption is quite restrictive. For instance, the frequently used Shishkin-type meshes do not meet this requirement.

The paper [4] deals with the possibility of applying the Scott-Zhang operator on Shishkin meshes Ω^N of tensor product type. The authors suggest to choose the element edges σ_i for every mesh node $\boldsymbol{X}_i \in \sigma_i$, $i = 1, \ldots, N^2$ in a special way:

- Certain edges σ_i on the boundary may be chosen arbitrarily but the rest has to be parallel to one coordinate axis, say the x-axis.

- The ratio of the size of the patch ω_T to the size of the element T must have an ε-uniform upper bound in both coordinate directions. Consequently, for instance an element T with a small side in the x-direction must be associated with a patch ω_T with the same property.

This modified Scott-Zhang operator Q_N can be applied on a Shishkin mesh. Unfortunately the authors needed more regularity of the regular solution component $S \in W_{2,\infty}(\Omega)$ of a convection-diffusion problem to prove optimal quasi-interpolation error estimates. Still, this result shows that the Scott-Zhang operator is quite flexible and that it can be tailored to suit an application on meshes with abrupt changes in the mesh sizes.

Note that the original Scott-Zhang operator and its modifications sketched so far were introduced for elements of Lagrange-type, i.e. the linear functionals associated with the element are function evaluations in certain points. The $C^1 - Q_2$ macro-element however features also the point evaluation of derivatives. We want to apply the basic ideas of the Scott-Zhang operator to the components of the $C^1 - Q_2$ macro-element space that are associated with the evaluation of the mixed second derivative. We do so with the aim of reducing the regularity required to prove anisotropic quasi-interpolation error estimates. In view of Remark 21 we study the question, whether it is possible to define a new C^1 interpolation operator $\tilde{\Pi}$ by introducing the right functional corresponding to the mixed derivative in such a way that estimates like (3.36) are possible assuming only some $W_{3,p}$ regularity of u.

The quasi-interpolant $\tilde{\Pi}u$ to u over some macro-element M will be governed on a macro-element neighbourhood or macro-element patch around M. More precisely, the coefficients of the basis functions that correspond to the mixed derivative are calculated by some weighted averaging process of the mixed derivative of u over macro-element edges that do not necessarily belong to M. Because of this non-local character of $\tilde{\Pi}$ we have to be very careful when a reference mapping to some reference domain is used to prevent imposing very restrictive conditions on the geometry of the macro-element patch. Instead we shall use some ideas of [2] and estimate directly on the world domain.

Let $\boldsymbol{X}_{ij} := (x_i, y_j)$, $(i,j) \in I$ denote the nodes of a rectangular tensor product mesh $\mathcal{M}_{\boldsymbol{h}}$, generated by the two arbitrary one-dimensional triangulations $\{x_i\}_{i=0}^n$ and $\{y_j\}_{j=0}^m$. We shall refer to $\mathcal{M}_{\boldsymbol{h}}$ as *macro-element mesh.* We use

$$h_i := \frac{1}{2}(x_i - x_{i-1}), \quad i = 1, \ldots, n \qquad \text{and} \qquad k_j := \frac{1}{2}(y_j - y_{j-1}), \quad j = 1, \ldots, m,$$

to denote the local step sizes in x- and y-direction. Each macro-element $M \in \mathcal{M}_{\boldsymbol{h}}$ is subdivided into four congruent elements introducing new mesh nodes with subscript $\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$. The generated *element mesh* is denoted by $\mathcal{T}_{\boldsymbol{h}}$, see Figure 3.4. Note that one may chose a different refinement of the macro-element mesh such that the elements within one macro-element remain comparable in size. We choose the presented uniform one in order to simplify the presentation. Now each macro-element $M_{ij} := [x_{i+1/2} - h_i, x_{i+1/2} + h_i] \times [y_{j+1/2} - k_j, y_{j+1/2} + k_j] \in \mathcal{M}_{\boldsymbol{h}}$ is centered around $(x_{i+1/2}, y_{j+1/2})$ and consists of four elements of size $\boldsymbol{h}_{ij} := (h_i, k_j)$. Moreover, we denote by $I_{ij} := I_{M_{ij}} := \{(i,j), (i,j+1), (i+1,j), (i+1,j+1)\}$ the set of the four node indices that are vertices of M_{ij}.

Let $V_{\boldsymbol{h}}$ denote the space of $C^1 - Q_2$ finite element functions over the tensor product mesh $\mathcal{T}_{\boldsymbol{h}}$. Using the reference mapping $F_{ij} : [-1,1]^2 \to M_{ij} \in \mathcal{M}_{\boldsymbol{h}}$ with

$$x = x_{i+1/2} + h_i \hat{x}, \qquad \text{and} \qquad y = y_{j+1/2} + k_j \hat{y}$$

we can specify basis functions of $V_{\boldsymbol{h}}$ in the world domain using (3.6). Consider for instance the lower right vertex $\boldsymbol{X}_{i+1,j}$ of the macro-element M_{ij}. Then the basis function $\psi_{i+1,j}$ associated with the mixed derivative in $\boldsymbol{X}_{i+1,j}$ admits the representation

$$\psi_{i+1,j}\big|_{M_{ij}} = h_i k_j \hat{\psi}_{1,-1} \circ F_{ij}^{-1},$$

where $\hat{\psi}_{1,-1}$ was defined in (3.6). Similarly,

$$\psi_{i,j}\big|_{M_{ij}} = h_i k_j \hat{\psi}_{-1,-1} \circ F_{ij}^{-1}, \quad \psi_{i,j+1}\big|_{M_{ij}} = h_i k_j \hat{\psi}_{-1,1} \circ F_{ij}^{-1}, \quad \psi_{i+1,j+1}\big|_{M_{ij}} = h_i k_j \hat{\psi}_{1,1} \circ F_{ij}^{-1}.$$

Let us now define a (quasi-)interpolation operator $\tilde{\Pi}$ by

$$\tilde{\Pi}u\big|_M := \Pi^r \left(u\big|_M\right) + \sum_{(k,\ell)\in I_M} a_{k,\ell}\, \psi_{k,\ell}, \tag{3.42}$$

with the reduced interpolation operator Π^r from Subsection 3.5.1 and real numbers $a_{k,\ell}$ still to be determined. Note that the choice of $a_{k,\ell}$ does not alter the ability of $\tilde{\Pi}$ to reproduce inhomogeneous Dirichlet boundary conditions g (if $g \in V_h|_{\partial\Omega}$, *i.e.* $g \in C^1(\partial\Omega)$ and piecewise quadratic), because $\psi_{k,\ell}$, $(k,\ell) \in I_{ij}$ vanishes on the boundary of M_{ij}, see Figure 3.3.

The local choices $a_{k,\ell} = 0$ and $a_{k,\ell} = u_{xy}(\boldsymbol{X}_{k\ell})$ correspond to $\tilde{\Pi} = \Pi^r$ and $\tilde{\Pi} = \Pi$, respectively. Next, we want to follow the approach of Scott and Zhang [71] and define the coefficients $a_{k,\ell}$ using certain mean values of u_{xy} along macro-element edges $\sigma_{k,\ell}$. Hence, as already mentioned, the interpolation operator is of non-local character and the theory developed in Section 3.4 can not be applied to $\tilde{\Pi}$. However, the definition of $\tilde{\Pi}u$ on a macro-element M is not global but shall be based on the values of u_{xy} on the *macro-element neighbourhood* S_M of M:

$$S_M = \bigcup \{M' \,:\, M' \in \mathcal{M}_{\boldsymbol{h}},\, M' \cap M \neq \emptyset\}. \tag{3.43}$$

More precisely, we associate every node $\boldsymbol{X}_{ij}$ of the macro-element mesh $\mathcal{M}_{\boldsymbol{h}}$ with a macro-element edge $\sigma_{i,j} \subset S_M$ such that $\boldsymbol{X}_{ij} \in \sigma_{i,j}$, see Figure 3.5 for some illustration.

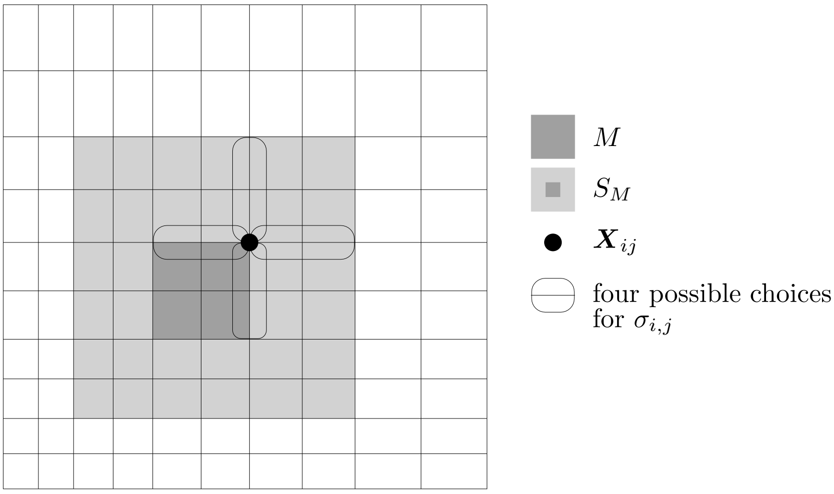

Figure 3.5: Definitions on the macro-element mesh.

Once the edges $\sigma_{i,j} \in S_M$ for $(i,j) \in I_M$ are chosen we can define the *associated macro-element patch* ω_M around M. Another patch neighbourhood of M is needed because the value of our quasi-interpolation operator will be based on values of its interpolant on M and $\sigma_{i,j}$. Hence, if the approximation error is estimated norms of the interpolant over a patch of macro-elements will appear on the right hand side of the estimate. On the other hand estimates that use the full neighbourhood S_M might be too crude.

Definition 5. The smallest (in area) rectangular patch of macro-elements that contains the convex hull of $\{\sigma_{i,j} : (i,j) \in I_M\}$ is called the *associated macro-element patch* ω_M around M.

Note that $M \subset \omega_M \subset S_M$. If for instance at each node $\boldsymbol{X}_{ij}$ of the tensor product macro-element mesh the set $\sigma_{i,j}$ is chosen to be the edge to the left of that point, then the associated macro-element patch ω_M around M is defined as the union of M and its left macro-element neighbour.

We plan to set

$$a_{i,j} = \frac{\partial^2}{\partial x \partial y} \left(\Pi_{\sigma_{i,j}} u\right)(\boldsymbol{X}_{ij}), \tag{3.44}$$

with a suitable projector $\Pi_{\sigma_{i,j}}$. Assuming that $\sigma_{i,j}$ is the horizontal macro-element edge $(x_i, x_{i+1}) \times \{y_j\}$ that connects the macro-element vertices $\boldsymbol{X}_{ij}$ and $\boldsymbol{X}_{i+1,j}$ we set

$$\Pi_{\sigma_{i,j}} u = b_{i,j}\psi_{i,j} + b_{i+1,j}\psi_{i+1,j}. \tag{3.45}$$

We determine the real coefficients $b_{i,j}$ and $b_{i+1,j}$ by

$$\int_{\sigma_{i,j}} \frac{\partial^2}{\partial x \partial y} \left(\Pi_{\sigma_{i,j}} u\right)(x, y_j) v(x) \,\mathrm{d}x = \int_{\sigma_{i,j}} \frac{\partial^2 u(x, y_j)}{\partial x \partial y} v(x) \,\mathrm{d}x \qquad \text{for all } v \in \mathbb{V}_{i+1/2}, \tag{3.46}$$

with $\mathbb{V}_{i+1/2} = \operatorname{span}\{\psi_i + \psi_{i+1}, \theta_{i+1/2}\}$ and $\theta_{i+1/2}(x) = \left(\frac{x - x_{i+1}}{h_i}\right)^2 - \frac{1}{6}$. Here ψ_i and ψ_{i+1} are the one dimensional spline basis functions from (3.3) scaled to $\sigma_{i,j}$, i.e.

$$\begin{aligned}
\psi_i(x) &= \frac{h_i}{4} - \frac{x - x_{i+1/2}}{2} \begin{cases} -\dfrac{3(x - x_{i+1/2})^2}{4h_i}, & x_i \le x \le x_{i+1/2}, \\ +\dfrac{(x - x_{i+1/2})^2}{4h_i}, & x_{i+1/2} \le x \le x_{i+1}, \end{cases} \\
\psi_{i+1}(x) &= -\frac{h_i}{4} - \frac{x - x_{i+1/2}}{2} \begin{cases} -\dfrac{(x - x_{i+1/2})^2}{4h_i}, & x_i \le x \le x_{i+1/2}, \\ +\dfrac{3(x - x_{i+1/2})^2}{4h_i}, & x_{i+1/2} \le x \le x_{i+1} \end{cases}
\end{aligned} \tag{3.47}$$

and $h_i = x_{i+1} - x_{i+1/2} = x_{i+1/2} - x_i$. Note that these functions have $\mathcal{O}(h_i)$ scalings while their first derivatives have $\mathcal{O}(1)$ scalings on (x_i, x_{i+1}). Moreover, we would like to recall $\psi_{i,j}(x,y) = \psi_i(x)\psi_j(y)$ and point out that the ansatz (3.45) is justified by the fact that $\frac{\partial^2}{\partial x \partial y}\psi_{k,\ell}$ vanishes on $\sigma_{i,j}$ if $\boldsymbol{X}_{k,\ell} \notin \sigma_{i,j}$, i.e. only adjacent basis functions contribute to the integral over $\sigma_{i,j}$ on the left hand side of (3.46). The choice of $\mathbb{V}_{i+1/2}$ will become clear in the next Lemma. Basically, we need that this space is L_2-orthogonal to certain functions to prove that discrete functions are left invariant, see Lemma 23.

Next, we want to find a more suitable representation of $a_{i,j}$ according to (3.44). For this purpose let us define the dual basis function $\psi_i^d \in \mathbb{V}_{i+1/2}$ by

$$\int_{\sigma_{i,j}} \frac{\partial^2 \psi_{k,j}(x, y_j)}{\partial x \partial y} \psi_i^d(x)\,\mathrm{d}x = \delta_{k,i}. \tag{3.48}$$

This system yields with (3.45)

$$b_{i,j} = \sum_{k=i}^{i+1} b_{k,j} \int_{\sigma_{i,j}} \frac{\partial^2 \psi_{k,j}(x,y_j)}{\partial x \partial y} \psi_i^d(x)\,\mathrm{d}x = \int_{\sigma_{i,j}} \frac{\partial^2}{\partial x \partial y} \left(\Pi_{\sigma_{i,j}} u\right)(x, y_j)\psi_i^d(x)\,\mathrm{d}x.$$

An application of (3.46) then gives

$$b_{i,j} = \int_{\sigma_{i,j}} \frac{\partial^2 u(x,y_j)}{\partial x \partial y} \psi_i^d(x)\,\mathrm{d}x.$$

Finally, we use the Lagrange relation $\frac{\partial^2 \psi_{k,\ell}(x_i,y_j)}{\partial x \partial y} = \delta_{k,i}\delta_{\ell,j}$ to obtain

$$a_{i,j} = \frac{\partial^2}{\partial x \partial y}\left(\Pi_{\sigma_{i,j}} u\right)(\boldsymbol{X}_{ij}) = b_{i,j} = \int_{\sigma_{i,j}} \frac{\partial^2 u(x,y_j)}{\partial x \partial y} \psi_i^d(x)\,\mathrm{d}x. \tag{3.49}$$

Hence, $a_{i,j}$ is indeed a weighted mean value of u_{xy} on the macro-element edge $\sigma_{i,j}$. For the weighting function we solve (3.48) to find

$$\begin{aligned}
\psi_i^d(x) &= -\frac{h_i^2 + 12h_i(x - x_{i+1/2})}{2h_i^3} \begin{cases} -\dfrac{3(x - x_{i+1/2})^2}{h_i^3}, & x_i \le x \le x_{i+1/2}, \\ +\dfrac{9(x - x_{i+1/2})^2}{h_i^3}, & x_{i+1/2} \le x \le x_{i+1}, \end{cases} \\
\psi_{i+1}^d(x) &= -\frac{h_i^2 - 12h_i(x - x_{i+1/2})}{2h_i^3} \begin{cases} +\dfrac{9(x - x_{i+1/2})^2}{h_i^3}, & x_i \le x \le x_{i+1/2}, \\ -\dfrac{3(x - x_{i+1/2})^2}{h_i^3}, & x_{i+1/2} \le x \le x_{i+1}. \end{cases}
\end{aligned} \tag{3.50}$$

Here $\psi_{i+1}^d \in \mathbb{V}_{i+1/2}$ is the other dual basis function on $\sigma_{i,j}$, satisfying (3.48) with i replaced by $i+1$. Note that $\psi_i^d, \psi_{i+1}^d \in C^1\big(x_i, x_{i+1}\big)$ and that $\|\psi_i^d\|_{L_\infty(\sigma_{i,j})} \le Ch_i^{-1}$ with a similar bound for $\|\psi_{i+1}^d\|_{L_\infty(\sigma_{i,j})}$. A simple calculation shows the important property

$$\int_{\sigma_{i,j}} \psi_i^d(x)\,\mathrm{d}x = 1, \tag{3.51}$$

which again underlines the role of ψ_i^d as a weighting function.

Remark 22. In Section 4 of [31] a similar macro-element edge based approach is used to reduce the regularity demanded of the function to be interpolated. There, the Girault-Scott operator is extended to the $C^1 - Q_2$ macro-element. In [31] integration by parts is applied to an identity similar to (3.49) which results in a different system defining the dual basis functions. However, this approach appears to be not suitable for anisotropic quasi-interpolation error estimates. Another difference to that paper is that here we mix local and non-local functionals for the definition of our quasi-interpolation operator which is reflected in the sophisticated choice of $\mathbb{V}_{i+1/2}$.

Lemma 23. $\tilde{\Pi}$ *preserves* $V_{\boldsymbol{h}}$ *functions, i.e.*

$$\tilde{\Pi} v_{\boldsymbol{h}} = v_{\boldsymbol{h}} \qquad \text{for all } v_{\boldsymbol{h}} \in V_{\boldsymbol{h}}. \tag{3.52}$$

Proof. Since every function $v_{\boldsymbol{h}} \in V_{\boldsymbol{h}}$ is uniquely determined by the nodal values

$$v_{\boldsymbol{h}}(\boldsymbol{X}_{ij}), \qquad \frac{\partial v_{\boldsymbol{h}}}{\partial x}(\boldsymbol{X}_{ij}), \qquad \frac{\partial v_{\boldsymbol{h}}}{\partial y}(\boldsymbol{X}_{ij}), \qquad \frac{\partial^2 v_{\boldsymbol{h}}}{\partial x \partial y}(\boldsymbol{X}_{ij}),$$

in the macro-element vertices $\boldsymbol{X}_{ij}$ with $(i,j) \in I$, it remains to prove that these functionals are invariant to the application of the quasi-interpolation operator $\tilde{\Pi}$. Let us prove the identity of the last functional involving the mixed derivative as the other ones are trivial. We observe with (3.49) that

$$\frac{\partial^2 \tilde{\Pi} v_{\boldsymbol{h}}}{\partial x \partial y}(\boldsymbol{X}_{ij}) = a_{i,j} = \int_{\sigma_{i,j}} \frac{\partial^2 v_{\boldsymbol{h}}(x, y_j)}{\partial x \partial y} \psi_i^d(x) \, \mathrm{d}x. \tag{3.53}$$

Since $v_{\boldsymbol{h}} \in V_{\boldsymbol{h}}$ it can be expanded on the macro-element M considered in terms of the basis functions $\varphi_{i,j}$, $\phi_{i,j}$, $\chi_{i,j}$ and $\psi_{i,j}$, $(i,j) \in I_M$ according to (3.6). For the mixed derivative on $\sigma_{i,j}$ we find

$$\begin{aligned} \left.\frac{\partial^2 v_{\boldsymbol{h}}}{\partial x \partial y}\right|_{\sigma_{i,j}} &= \sum_{\ell=i}^{i+1} \frac{\partial v_{\boldsymbol{h}}}{\partial y}(\boldsymbol{X}_{\ell j}) \frac{\partial^2}{\partial x \partial y} \chi_{\ell,j} + \frac{\partial^2 v_{\boldsymbol{h}}}{\partial x \partial y}(\boldsymbol{X}_{\ell j}) \frac{\partial^2}{\partial x \partial y} \psi_{\ell,j} \\ &= \sum_{\ell=i}^{i+1} \frac{\partial v_{\boldsymbol{h}}}{\partial y}(\boldsymbol{X}_{\ell j}) \varphi_\ell'(x) \underbrace{\psi_j'(y_j)}_{=1} + \frac{\partial^2 v_{\boldsymbol{h}}}{\partial x \partial y}(\boldsymbol{X}_{\ell j}) \frac{\partial^2}{\partial x \partial y} \psi_{\ell,j}, \end{aligned}$$

since the mixed derivative of the other basis functions vanishes on $\sigma_{i,j}$. The functions φ_ℓ', $\ell = i, i+1$ are continuous, piecewise linear and vanish in the endpoints of the interval (x_i, x_{i+1}). Hence, the odd function $\psi_i + \psi_{i+1}$ is $L_2(\sigma_{i,j})$ orthogonal to them. A direct calculation shows the same orthogonality relation for θ_{i+1}, i.e. $\mathbb{V}_{i+1/2} \perp_{L_2(\sigma_{i,j})} \varphi_\ell'$, $\ell = i, i+1$. Using this orthogonality and (3.48) in (3.53) we see that

$$\frac{\partial^2 \tilde{\Pi} v_{\boldsymbol{h}}}{\partial x \partial y}(\boldsymbol{X}_{ij}) = \int_{\sigma_{i,j}} \frac{\partial^2 v_{\boldsymbol{h}}(x, y_j)}{\partial x \partial y} \psi_i^d(x) \, \mathrm{d}x = \frac{\partial^2 v_{\boldsymbol{h}}}{\partial x \partial y}(\boldsymbol{X}_{ij}).$$

From which the assertion follows. ☐

Remark 23. With (3.49) the quasi-interpolation operator $\tilde{\Pi}$ from (3.42) is a projector due to (3.52).

Lemma 24. *For some macro-element* $M \in \mathcal{M}_{\boldsymbol{h}}$ *let* $v \in Q_2(\omega_M)$ *i.e.* v *is biquadratic on the associated macro-element patch* ω_M *around the macro-element* M*, then*

$$\left.\tilde{\Pi} v\right|_M = v|_M \,. \tag{3.54}$$

Proof. We use the Q_2-preservation of the interpolation operator Π and Lemma 23:

$$v|_M = (\Pi v)|_M = \left.\left(\tilde{\Pi}(\Pi v)\right)\right|_M = \left.(\tilde{\Pi} v)\right|_M .$$

In the second identity we applied Lemma 23 and need $v \in Q_2(\omega_M)$ because of the non-local character of $\tilde{\Pi}$. ☐

The following lemma is taken from [15, Theorem 1.1].

Lemma 25. *Let* $\Omega \subset \mathbb{R}^n$ *be convex with diameter* d *and let* $g \in W_{\nu,p}(\Omega)$*,* $\nu \in \mathbb{N}$*,* $p \in [1, \infty]$*. Then there exists a polynomial* $p_\nu^g \in P_{\nu-1}$ *for which*

$$|g - p_\nu^g|_{W_{k,p}(\Omega)} \le C(n,\nu) d^{\nu-k} |g|_{W_{\nu,p}(\Omega)}, \quad k = 0, 1, \dots, \nu.$$

Here the polynomial

$$p_\nu^g(x) = Q^\nu\big(g(A\cdot)\big)\big(A^{-1}x\big)$$

is constructed using the averaged Taylor polynomial Q^ν over the ball $B(0,1) \subset \mathbb{R}^n$ and A is John's optimal affine transform with respect to Ω, cp. [15]. The basic idea of this paper is the usage of ellipsoids in contrast to balls which is more suitable for anisotropic elements to which we want to apply this result. Yet, we shall first give a small modification of it.

Lemma 26. *Let $\Omega \subset \mathbb{R}^n$ be convex with diameter d, $\boldsymbol{\gamma}$ be a multi-index with $|\boldsymbol{\gamma}| = m \in \mathbb{N}$, and let $v \in W_{\ell,p}(\Omega)$, $m \le \ell \in \mathbb{N}$, $p \in [1,\infty]$. Then there exists a polynomial $p_\ell^v \in P_{\ell-1}$ for which*

$$|\boldsymbol{D}^{\boldsymbol{\gamma}}(v - p_\ell^v)|_{W_{k,p}(\Omega)} \le C(n, \ell - m) d^{\ell-m-k} |\boldsymbol{D}^{\boldsymbol{\gamma}} v|_{W_{\ell-m,p}(\Omega)}, \quad k = 0, 1, \ldots, \ell - m.$$

Proof. First assume that $v \in C^\ell(\Omega)$. Applying Lemma 25 with $\nu = \ell - m$ and $g = \boldsymbol{D}^{\boldsymbol{\gamma}} v$ yields the existence of a polynomial $p_{\ell-m}^{\boldsymbol{D}^{\boldsymbol{\gamma}} v} \in P_{\ell-m-1}$ such that

$$|\boldsymbol{D}^{\boldsymbol{\gamma}} v - p_{\ell-m}^{\boldsymbol{D}^{\boldsymbol{\gamma}} v}|_{k,p} \le C(n, \ell - m) d^{\ell-m-k} |\boldsymbol{D}^{\boldsymbol{\gamma}} v|_{\ell-m,p}, \quad k = 0, 1, \ldots, \ell - m.$$

Next one finds that for $1 \le m \le \ell - 1$

$$p_{\ell-m}^{\boldsymbol{D}^{\boldsymbol{\gamma}} v}(x) = Q^{\ell-m}\big((\boldsymbol{D}^{\boldsymbol{\gamma}} v)(A\cdot)\big)\big(A^{-1}x\big) = \boldsymbol{D}^{\boldsymbol{\gamma}}\Big(Q^\ell\big(v(A\cdot)\big)\big(A^{-1}x\big)\Big) = \boldsymbol{D}^{\boldsymbol{\gamma}} p_\ell^v,$$

i.e. the averaged Taylor polynomial and differentiation commute in some sense [15, Corollary 3.4]. Now the case $v \in W_{\ell,p}(\Omega)$ follows by standard arguments based on the density of $C^\infty(\Omega)$ in $W_{\ell,p}(\Omega)$. For $m = 0$ the assertion of the Lemma is given by Lemma 25 and for $m = \ell$ the assertion is trivial. □

Remark 24. A slightly more general result is given in [2, Lemma 2.1]. However, there the dependencies of the constant of geometrical properties of the domain considered is not stated explicitly.

Assumption 1. Let for each node $\boldsymbol{X}_{ij}$ of a macro-element M the macro-element edges $\sigma_{i,j}$ be chosen in such a way that with the associated macro-element patch ω_M around M it holds

$$h_k(\omega_M) \le C h_k(M) \quad k = 1, 2. \tag{3.55}$$

Here and in the following $h_k(T)$ denotes the size of an axis-aligned rectangle T in x_k-direction, $k = 1, 2$. Moreover, we set $\boldsymbol{h}_M = (h_1(M), h_2(M))$ for any macro-element M.

Lemma 27. *Based on Assumption 1 for any $u \in W_{\ell,p}(\omega_M)$ there is a polynomial $q \in P_{\ell-1}(\omega_M)$ with*

$$\sum_{|\boldsymbol{\alpha}| \le \ell - m} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \, |\boldsymbol{D}^{\boldsymbol{\alpha}}(u - q)|_{W_{m,p}(\omega_M)} \le C \sum_{|\boldsymbol{\alpha}| = \ell - m} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \, |\boldsymbol{D}^{\boldsymbol{\alpha}} u|_{W_{m,p}(\omega_M)},$$

for all $m = 0, \ldots, \ell$.

Proof. Using an affine transformation we can map the macro-element M to the reference macro-element $[-1,1]^2$. This transformation maps ω_M to $\hat{\omega}_M$. Based on (3.55) we see that the diameter of the rectangle $\hat{\omega}_M$ can be bounded by a constant. Hence, we can apply Lemma 26 in the transformed domain. Scaling back to ω_M we obtain due to $\boldsymbol{h}_M^{\boldsymbol{\alpha}} \boldsymbol{D}^{\boldsymbol{\alpha}} = \hat{\boldsymbol{D}}^{\boldsymbol{\alpha}}$ that

$$\sum_{|\boldsymbol{\alpha}| \le \ell - m} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}}(u - q)\|_{L_p(\omega_M)} \le C \sum_{|\boldsymbol{\alpha}| = \ell - m} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} u\|_{L_p(\omega_M)},$$

for a multi-index $\boldsymbol{\gamma}$ with $|\boldsymbol{\gamma}| = m$. The assertion follows by summing up over all of these multi-indices. □

Remark 25. A similar lemma is given in [2, Lemma 3.1]. However, there the mesh is required to have no abrupt changes in the element sizes. Clearly, Assumption 1 can be dropped then. Note that (3.55) can also be found in the paper [4].

Lemma 28 (Stability of $\tilde{\Pi}$)**.** *Under Assumption 1 the quasi-interpolation operator $\tilde{\Pi}$ satisfies the stability estimate*

$$\left|\tilde{\Pi}u\right|_{W_{1,p}(M)} \le C C_{M,p} \sum_{|\boldsymbol{\alpha}|\le 2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} |\boldsymbol{D}^{\boldsymbol{\alpha}} u|_{W_{1,p}(\omega_M)}$$

with

$$C_{M,p} := \left(\frac{\operatorname{meas} M}{\min_{M'\in\mathcal{M}_{\boldsymbol{h}},\, M'\subset\omega_M} \operatorname{meas} M'} \right)^{1/p} \ge 1,$$

provided that $u \in W_{3,p}(\omega_M) \cap C^1(M)$ with $p \in [1,\infty]$.

Proof. Let $M \in \mathcal{M}_{\boldsymbol{h}}$ be a macro-element and set $\boldsymbol{h}_M = (h_1(M), h_2(M))$. We consider a first derivative in x-direction. Using the definition of $\tilde{\Pi}$ and a triangle inequality we find that

$$\left\|\left(\tilde{\Pi}u\right)_x\right\|_{L_p(M)} \le \|(\Pi^r u)_x\|_{L_p(M)} + \left\| \sum_{(i,j)\in I_M} a_{i,j} \frac{\partial \psi_{i,j}}{\partial x} \right\|_{L_p(M)}, \tag{3.56}$$

with coefficients $a_{i,j}$ depending on the direction of $\sigma_{i,j}$ given by

$$a_{i,j} = \begin{cases} \displaystyle\int_{\sigma_{i,j}} \frac{\partial^2 u(x,y_j)}{\partial x \partial y} \psi_i^d(x)\,\mathrm{d}x & \text{if } \sigma_{i,j} \text{ is horizontal,} \\ \displaystyle\int_{\sigma_{i,j}} \frac{\partial^2 u(x_i,y)}{\partial x \partial y} \psi_j^d(y)\,\mathrm{d}y & \text{if } \sigma_{i,j} \text{ is vertical.} \end{cases} \tag{3.57}$$

We estimate the first term on the right hand side of (3.56) using Theorem 19

$$\begin{aligned} \left\|\left(\Pi^r u\right)_x\right\|_{L_p(M)} &\le \|u_x\|_{L_p(M)} + \|(u-\Pi^r u)_x\|_{L_p(M)} \\ &\le C \sum_{|\boldsymbol{\alpha}|\le 2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} u_x\|_{L_p(M)}. \end{aligned} \tag{3.58}$$

For the other term we use $\left\|\frac{\partial}{\partial x}\psi_{i,j}\right\|_{L_\infty(M)} \le C h_2(M)$ which yields

$$\left\| \sum_{(i,j)\in I_M} a_{i,j} \frac{\partial \psi_{i,j}}{\partial x} \right\|_{L_p(M)} \le C (\operatorname{meas} M)^{1/p} h_2(M) \max_{(i,j)\in I_M} |a_{i,j}|. \tag{3.59}$$

Next we use $\|\psi_k^d\|_{\infty,\sigma_{i,j}} \le C \operatorname{meas}(\sigma_{i,j})^{-1}$ for $k=i,j$ and obtain with a Hölder inequality

$$|a_{i,j}| \le C \operatorname{meas}(\sigma_{i,j})^{-1} \left\| \frac{\partial^2 u}{\partial x \partial y} \right\|_{L_1(\sigma_{i,j})} \quad \text{for } (i,j) \in I_M. \tag{3.60}$$

Set

$$M' := \operatorname*{arg\,min}_{\substack{\tilde{M}\in\mathcal{M}_{\boldsymbol{h}} \\ \tilde{M}\subset\omega_M}} (\operatorname{meas} \tilde{M}),$$

i.e. the macro-element $M' \in \mathcal{M}_{\boldsymbol{h}}$ belongs to the associated macro-element patch ω_M around M and realizes the smallest surface measure. Using the embeddings $W_{1,p}(\hat{\omega}_M) \hookrightarrow W_{1,p}(\hat{M}') \hookrightarrow L_1(\hat{\sigma}_{i,j})$ in a transformed domain $\hat{\omega}_M$ and scaling back to the original one, we see that

$$\|v\|_{L_1(\sigma_{i,j})} \le \operatorname{meas}(\sigma_{i,j}) \operatorname{meas}(M')^{-1/p} \sum_{|\boldsymbol{\alpha}|\le 1} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} v\|_{L_p(\omega_M)} \tag{3.61}$$

for $v \in W_{1,p}(\omega_M)$. Here we also used Assumption 1. Collecting (3.59), (3.60) and (3.61) with $v = \frac{\partial^2 u}{\partial x \partial y}$ we obtain

$$\left\| \sum_{(i,j)\in I_M} a_{i,j} \frac{\partial \psi_{i,j}}{\partial x} \right\|_{L_p(M)} \le C \frac{(\operatorname{meas} M)^{1/p}}{(\operatorname{meas} M')^{1/p}} \sum_{|\boldsymbol{\alpha}|\le 1} \boldsymbol{h}_M^{\boldsymbol{\alpha}+(0,1)} \left\| \boldsymbol{D}^{\boldsymbol{\alpha}+(0,1)} u_x \right\|_{L_p(\omega_M)}. \tag{3.62}$$

Together with (3.56) and (3.58) the assertion of the lemma is proven since the first derivative in y-direction can be estimated analogously. ☐

Theorem 29. *Based on Assumption 1 for the quasi-interpolation operator $\tilde{\Pi}$ the approximation error estimate*

$$|u - \tilde{\Pi}u|_{W_{1,p}(M)} \leq CC_{M,p} \sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} |\boldsymbol{D}^{\boldsymbol{\alpha}} u|_{W_{1,p}(\omega_M)} \tag{3.63}$$

holds true provided that $u \in W_{3,p}(\omega_M) \cap C^1(M)$ with $p \in [1, \infty]$. Here $C_{M,p}$ is the constant from Lemma 28.

Proof. Let $q \in P_2(\omega_M)$ denote the polynomial of Lemma 27 with $\ell = 3$. A triangle inequality gives

$$|u - \tilde{\Pi}u|_{W_{1,p}(M)} \leq |u - q|_{W_{1,p}(M)} + |q - \tilde{\Pi}u|_{W_{1,p}(M)}. \tag{3.64}$$

As a polynomial $q \in P_2(\omega_M)$ is preserved by $\tilde{\Pi}$ on the macro-element M considered, see Lemma 24. Hence, we can use the stability of $\tilde{\Pi}$ shown in Lemma 28 to get a bound for the second summand

$$|q - \tilde{\Pi}u|_{W_{1,p}(M)} = |\tilde{\Pi}(q - u)|_{W_{1,p}(M)} \leq CC_{M,p} \sum_{|\boldsymbol{\alpha}|\leq 2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} |\boldsymbol{D}^{\boldsymbol{\alpha}}(q - u)|_{W_{1,p}(\omega_M)}. \tag{3.65}$$

The first summand is estimated as follows:

$$|u - q|_{W_{1,p}(M)} \leq C \sum_{|\boldsymbol{\alpha}|\leq 2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} |\boldsymbol{D}^{\boldsymbol{\alpha}}(u - q)|_{W_{1,p}(M)}, \tag{3.66}$$

which can be proven to hold true by setting $v := \boldsymbol{D}^{\boldsymbol{\gamma}}(u - q)$ with $|\boldsymbol{\gamma}| = 1$ in

$$\|v\|_{L_p(M)} \leq C \sum_{|\boldsymbol{\alpha}|\leq 2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} v\|_{L_p(M)}.$$

This is in turn the embedding $W_{2,p}(M) \hookrightarrow L_p(M)$ on the reference macro-element and appropriate scaling. Collecting (3.64), (3.65) and (3.66) we arrive at

$$|u - \tilde{\Pi}u|_{W_{1,p}(M)} \leq CC_{M,p} \sum_{|\boldsymbol{\alpha}|\leq 2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} |\boldsymbol{D}^{\boldsymbol{\alpha}}(u - q)|_{W_{1,p}(\omega_M)} \leq CC_{M,p} \sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} |\boldsymbol{D}^{\boldsymbol{\alpha}} u|_{W_{1,p}(\omega_M)},$$

due to the special choice of q and Lemma 27. $\square$

Remark 26. The absence of abrupt changes in the mesh sizes leads not only to Assumption 1 always being satisfied but also to $C_{M,p} \leq C$ in (3.63), similar to the results in [2]. If on the contrary there are abrupt changes in the mesh sizes of arbitrary magnitude then (3.63) can become useless for $p < \infty$ — an observation that was made in [4], as well.

Remark 27. Inspecting the proofs of Lemma 28 and Theorem 29 one sees that under the same assumptions the approximation error estimate

$$\|u - \tilde{\Pi}u\|_{L_p(M)} \leq CC_{M,p} \sum_{|\boldsymbol{\alpha}|=3} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} u\|_{L_p(\omega_M)} \tag{3.67}$$

holds true for $p > 1$. In fact, the stability estimate

$$\|\Pi^r u\|_{L_p(M)} \leq CC_{M,p} \sum_{|\boldsymbol{\alpha}|\leq 3} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} u\|_{L_p(M)}$$

can be established based on the embedding $W_{3,p}(\Lambda) \hookrightarrow C^1(\Lambda)$ (which holds true for $p \geq 2$ in two dimensions) on the reference macro-element and a scaling argument. Moreover, one can make use of $\|\psi_{i,j}\|_{L_\infty(M)} \leq C\boldsymbol{h}_M^{(1,1)}$ for $(i,j) \in I_M$. If one only has $u \in W_{2,\infty}(\omega_M)$ one can still obtain

$$\|u - \tilde{\Pi}u\|_{L_\infty(M)} \leq C \sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} u\|_{L_\infty(\omega_M)}$$

by estimating (3.57) directly.

Remark 28. Similarly to the situation in which the interpolation operator is defined by local functionals it is again important that *polynomials are reproduced on larger entities.* While we demanded this property for macro-elements in the local setting we need it now on patches of macro-elements. This seems to be an underlaying principle.

We now turn our attention to second order derivatives. Inspecting the arguments in Theorem 29 for the possibility to prove L_p-bounds for second order derivatives of the approximation error, we see that stability of $\tilde{\Pi}$ is crucial.

It is possible to prove

$$\left\| \left(\tilde{\Pi}u\right)_{xy} \right\|_{L_p(M)} \leq C C_{M,p} \sum_{|\boldsymbol{\alpha}|\leq 1} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \left\| \boldsymbol{D}^{\boldsymbol{\alpha}} u_{xy} \right\|_{L_p(\omega_M)}.$$

However, it is unclear how to obtain a similar estimate for the other second order derivatives. We therefore restrict the subsequent study to the case of an *isotropic macro-element patch* ω_M. These results will be useful in Section 3.7. There we want to apply $\tilde{\Pi}$ in the fine regions of a Shishkin mesh close to the corners of the domain where the mesh is uniform.

Assumption 2. Let $M \in \mathcal{M}_{\boldsymbol{h}}$ denote a macro-element such that the restriction of $\mathcal{M}_{\boldsymbol{h}}$ to the associated macro-element patch ω_M is locally uniform with mesh size h_M.

Theorem 30. *Based on Assumption 2 the quasi-interpolation operator* $\tilde{\Pi}$ *satisfies the approximation error estimate*

$$\left|u - \tilde{\Pi}u\right|_{W_{k,p}(M)} \leq C h_M^{3-k} |u|_{W_{3,p}(\omega_M)}, \tag{3.68}$$

for $u \in W_{3,p}(\omega_M) \cap C^1(M)$ *with* $p \in [1, \infty]$ *and* $k \leq 2$.

Proof. Under Assumption 2 the estimates (3.63) and (3.67) simplify to (3.68) for $k \leq 1$ and it remains to validate this estimate for $k = 2$.

Let $v \in C^1(M)$ with $v_{xy}|_{\sigma_{ij}} \in L_1(\sigma_{i,j})$ for all $(i,j) \in I_M$ so that Πv is well defined. By Assumption 2 and the fact that Πv is piecewise biquadratic an inverse estimate yields

$$\|\tilde{\Pi}v\|_{W_{2,p}(M)} \leq C h_M^{-1} \|\tilde{\Pi}v\|_{W_{1,p}(M)}. \tag{3.69}$$

We proceed as in Theorem 29. By Lemma 27 with $\ell = 3$ there exits a unique polynomial $q \in P_2(\omega_M)$ such that

$$\sum_{k=0}^{3} h_M^k \, |u - q|_{W_{k,p}(\omega_M)} \leq C h_M^3 |u|_{W_{3,p}(\omega_M)}, \tag{3.70a}$$

$$\sum_{k=0}^{2} h_M^k \, |u - q|_{W_{k+1,p}(\omega_M)} \leq C h_M^2 |u|_{W_{3,p}(\omega_M)}. \tag{3.70b}$$

A triangle inequality implies

$$\left|\tilde{\Pi}u - u\right|_{W_{2,p}(M)} \leq |u - q|_{W_{2,p}(M)} + \left|\tilde{\Pi}(q - u)\right|_{W_{2,p}(M)}. \tag{3.71}$$

The first summand is easily bounded by (3.70a). For the other one we use the inverse estimate (3.69), the stability estimates for low order derivatives of $\tilde{\Pi}$, see Lemma 28 and Remark 27, and (3.70):

$$\begin{aligned}
\left|\tilde{\Pi}(q-u)\right|_{W_{2,p}(M)} &\leq C h_M^{-1} \left\|\tilde{\Pi}(q-u)\right\|_{W_{1,p}(M)} \leq C h_M^{-1} \left(\left|\tilde{\Pi}(q-u)\right|_{W_{1,p}(M)} + \left\|\tilde{\Pi}(q-u)\right\|_{L_p(M)} \right) \\
&\leq C h_M^{-1} \left(\sum_{k=0}^{2} h_M^k |q-u|_{W_{k+1,p}(\omega_M)} + \sum_{k=0}^{3} h_M^k |q-u|_{W_{k,p}(\omega_M)} \right) \\
&\leq C h_M |u|_{W_{3,p}(\omega_M)}.
\end{aligned} \tag{3.72}$$

Collecting (3.71), (3.70a) and (3.72) the result follows. $\square$

Remark 29. For $p < \infty$ the constant $C_{M,p}$ in the estimates (3.63) and (3.67) renders them useless on meshes of Shishkin type or any other mesh with abrupt changes in the mesh sizes. In this case L_∞ estimates are desirable. For second order derivatives we were able to prove a result of classical type with Theorem 30. In order to prove anisotropic error estimates it might be necessary to specify additional rules for the choice of the macro-element edges $\sigma_{i,j}$ associated with the macro-element vertices $\boldsymbol{X}_{ij}$, $(i,j) \in I_M$. Moreover, Theorem 29 shows two things:

- Firstly, it is possible to design useful quasi-interpolation operators that are defined by a mix of local and non-local functionals. This is particularly true if the element considered is not of Lagrange type. Extending this idea one might use different entities $\sigma_{i,j}$ for every component of a quasi-interpolation operator.

- Secondly, by using non-local functionals only for the coefficients of basis functions associated with higher order derivatives the resulting quasi-interpolation operators of Scott-Zhang type seem to be very flexible with respect to the choice of the entities $\sigma_{i,j}$. Note that in [2] derivatives of adaptations of the Scott-Zhang operator were only proven to obey anisotropic interpolation error estimates if the entities $\sigma_{i,j}$ were chosen all parallel.

3.5.4 Summary: anisotropic C^1 (quasi-)interpolation error estimates

In this Section we want to summarize our results and those of [13]. To the knowledge of the author these are the only sources of anisotropic (quasi-)interpolation error estimates for C^1 Hermite(-type) interpolation. All estimates are valid on rectangular tensor product meshes such that the edges of an element K are aligned with the coordinate axes. In all estimates C is a generic constant that does not depend on u or the mesh.

The work [13] addresses for $N \geq 1$ two C^{N-1} Hermite interpolation operators I_{12} and I_{22} into the piecewise Q_{2N-1} and Q_{2N} functions, respectively. Its main results are the anisotropic error estimates

$$
\begin{aligned}
|u - I_{12}u|_{N,K} &\leq C \sum_{|\boldsymbol{\beta}|=N} \boldsymbol{h}_K^{\boldsymbol{\beta}} |\boldsymbol{D}^{\boldsymbol{\beta}} u|_{N,K}, \\
|u - I_{22}u|_{N,K} &\leq C \sum_{|\boldsymbol{\beta}|=N+1} \boldsymbol{h}_K^{\boldsymbol{\beta}} |\boldsymbol{D}^{\boldsymbol{\beta}} u|_{N,K},
\end{aligned}
$$

for $u \in H^{2N}(K)$ and with $\boldsymbol{h}_K = (h_{1,K}, h_{2,K})$ where $h_{i,K}$ is the size of K in x_i-direction.

Inspecting their proofs for $N = 2$ we see that there is a C^1 Hermite interpolation operator I_{12} into the piecewise bicubic functions (more precisely the Bogner-Fox-Schmidt element space) such that

$$
\left\| \boldsymbol{D}^{\boldsymbol{\gamma}} \left(u - I_{12}u\right) \right\|_{0,K} \leq C \sum_{|\boldsymbol{\alpha}|=4-|\boldsymbol{\gamma}|} \boldsymbol{h}_K^{\boldsymbol{\alpha}} \left\| \boldsymbol{D}^{\boldsymbol{\alpha}} \boldsymbol{D}^{\boldsymbol{\gamma}} u \right\|_{0,K},
$$

for $|\boldsymbol{\gamma}| \leq 2$ and $u \in H^4(K)$. We want to emphasize that this result originally obtained by [13] can alternatively be proven using Apel's theory and our key observation that two dimensional divided differences may be used as associated functionals (cf. Corollary 22).

We refer to [13] for a note on the three dimensional case.

In the case of piecewise biquadratic functions we extended the results of [31] to the anisotropic case using new results on macro-interpolation. If the mesh can be generated as a uniform refinement of a macro-element mesh $\mathcal{M}_{\boldsymbol{h}}$, then there is a C^1 Hermite interpolation operator Π into the piecewise biquadratic functions such that (cf. Corollary 21)

$$
\left\| \boldsymbol{D}^{\boldsymbol{\gamma}} \left(u - \Pi u\right) \right\|_{0,M} \leq C \left(\sum_{|\boldsymbol{\alpha}|=4-|\boldsymbol{\gamma}|} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \left| \boldsymbol{D}^{\boldsymbol{\alpha}} \boldsymbol{D}^{\boldsymbol{\gamma}} u \right|_{0,M} + \sum_{|\boldsymbol{\alpha}|=3-|\boldsymbol{\gamma}|} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \left| \int_M \boldsymbol{D}^{\boldsymbol{\alpha}} \boldsymbol{D}^{\boldsymbol{\gamma}} u(x,y) \,\mathrm{d}x\mathrm{d}y \right| \right)
$$

on a macro-element $M \in \mathcal{M}_{\boldsymbol{h}}$ for a multi-index $\boldsymbol{\gamma}$ with $|\boldsymbol{\gamma}| \leq 2$ and $u \in C^2(M)$ such that $\boldsymbol{D}^{\boldsymbol{\gamma}} u \in H^{4-|\boldsymbol{\gamma}|}(M)$.

In order to reduce the regularity required we use non-local information of the interpolant in order to define the coefficient of the basis function associated with the mixed second derivative, creating the quasi-interpolation operator $\tilde{\Pi}$. For its analysis we need Assumption 1 to be satisfied. Collecting the results of Theorem 29, Remark 27 we summarize that for $p \in [2,\infty]$ and $m = 0, 1$ the error estimate

$$|u - \tilde{\Pi}u|_{W_{m,p}(M)} \le C C_{M,p} \sum_{|\boldsymbol{\alpha}|=3-m} \boldsymbol{h}_M^{\boldsymbol{\alpha}} |\boldsymbol{D}^{\boldsymbol{\alpha}} u|_{W_{m,p}(\omega_M)} \tag{3.73}$$

holds true, provided $u \in W_{3,p}(\omega_M)$. Here

$$C_{M,p} := \left(\frac{\operatorname{meas} M}{\min_{T \in \mathcal{M},\, T \subset \omega_M} \operatorname{meas} T} \right)^{1/p} \ge 1$$

and ω_M is the associated macro-element patch ω_M around M, cf. Definition 5.

In the case of a more regular mesh (more precisely: under Assumption 2) the operator $\tilde{\Pi}$ satisfies error estimates of classical type even for second order derivatives, see Theorem 30. Note that the absence of abrupt changes in the mesh sizes implies the validity of Assumption 1 and a simplification of the estimates (3.73) due to $C_{M,p} \le C$, cf. Remark 26.

It would be very interesting to check numerically if there is hope for the Girault-Scott operator of [31, Section 4] to allow anisotropic interpolation error estimates given only some $W_{2,p}$ regularity of the function to be approximated. However, certain details in that paper are unclear — especially the scaling of the true dual basis functions (given only as a brief note) is questionable.

3.6 An anisotropic macro-element of tensor product type

In Section 3.2 we have seen 1D Hermite interpolation in the space of quadratic C^1 splines. The tensor product of this 1D macro-element with itself created a 2D macro-element and the induced interpolation operator Π for which we were able to prove certain anisotropic interpolation error estimates. However, the usage of this operator on for instance a Shishkin mesh (where the direction of anisotropy and mesh sizes changes abruptly) does not lead to optimal results. The main reason for this failure is that the C^1 operators Π or $\tilde{\Pi}$ do not satisfy certain L_∞-stability estimates. Based on the usage of derivatives one has for instance on some macro-element $M \in \mathcal{M}_{\boldsymbol{h}}$ with sizes $\boldsymbol{h}_M$ that

$$\|\Pi v\|_{L_\infty(M)} \le C \left(\sum_{|\boldsymbol{\alpha}| \le 1} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}} v\|_{L_\infty(M)} + \boldsymbol{h}_M^{(1,1)} \|\boldsymbol{D}^{(1,1)} v\|_{L_\infty(M)} \right),$$

holds true, i.e. L_∞ norms of derivatives appear on the right hand side. Hence, if one wants to bound the error in the interior with large elements one can no longer use that the interpolant is small there but has to demand that also derivatives of the interpolant are small. This is however not true on a Shishkin mesh as already mentioned in the introduction. In order to remedy this problem we consider the following anisotropic macro-element.

We form a macro of two rectangles and use as degrees of freedom the function value and the value of a certain first derivate in six points along the boundary of the macro (cf. Figure 3.6). Note that this macro-element can be considered as the tensor product of one dimensional $C^1 - P_2$ macro-interpolation and P_2 Lagrange interpolation. Hence, we *leave the realm of C^1 macro-elements* but preserve the property of a continuous normal derivative across some macro-element edges. This will be vital in the next section.

More precisely, assuming that, as illustrated in Figure 3.6, the reference macro-element $\hat{M} := \{[-1,1] \times [-1,0], [-1,1] \times [0,1]\}$ over the reference domain $\Lambda = [-1,1]^2$ is mapped to an anisotropic one for which the aspect ratio h_x/h_y is very large we use quadratic C^1 splines in y direction (small side) and P_2 in x direction (large side). This space $S(\hat{M})$ is 12 dimensional and from (3.4) and

$$p(x) = p[-1] + p[-1,0](x+1) + p[-1,0,1](x+1)x \quad \forall p \in P_2([-1,1]),$$

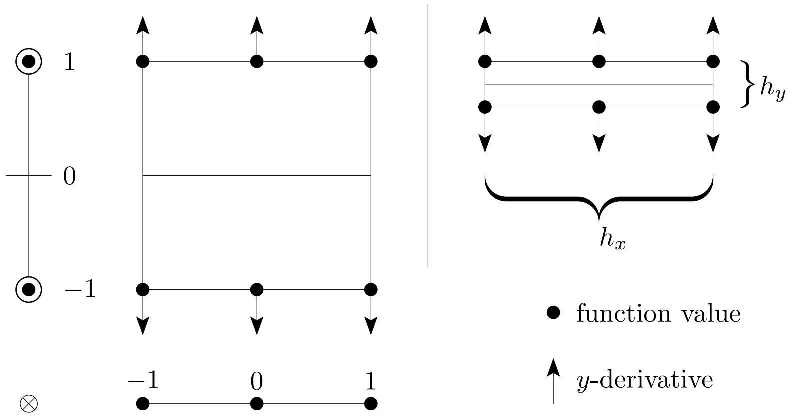

Figure 3.6: Degrees of freedom of the anisotropic macro-element on the reference macro-element $\hat{M}$ (left) and on some anisotropic macro in the world domain (right).

we can obtain the representation

$$\begin{aligned} s(x,y) = \sum_{j=1}^{3} \big(&F_{1j}(s)(y+1)^{j-1} + F_{2j}(s)(x+1)(y+1)^{j-1} + F_{3j}(s)(x+1)x(y+1)^{j-1}\big) \\ &+ 4\big(F_{14}(s) + F_{24}(s)(x+1) + F_{34}(s)(x+1)x\big)\hat{\psi}_1(y) \quad \forall s \in S(\hat{M}). \end{aligned} \tag{3.74}$$

By Π^x we denote the macro-element interpolation operator such that the roles of the sizes h_x and h_y of a macro-element M are interchanged, i.e. $h_x \gg h_y$.

The functionals F_{ij} are again defined as two dimensional divided differences:

$$F_{ij}(s) := s[m_i; n_j] \quad \text{with} \quad m_i = \begin{cases} -1 & \text{for } i = 1, \\ -1, 0 & \text{for } i = 2, \\ -1, 0, 1 & \text{for } i = 3, \end{cases} \quad \text{and} \quad n_j = \begin{cases} -1 & \text{for } j = 1, \\ -1, -1 & \text{for } j = 2, \\ -1, -1, 1 & \text{for } j = 3, \\ -1, -1, 1, 1 & \text{for } j = 4. \end{cases}$$

It is easy to establish the H^1-conformity of this macro-element. Moreover, we find that the y-derivative along the edge $y = \pm 1$ of Λ can be expressed by

$$\begin{aligned} \frac{\partial s}{\partial y}(x, \pm 1) = \frac{\partial s}{\partial y}(0, \pm 1) &+ \frac{1}{2}\left(\frac{\partial s}{\partial y}(1, \pm 1) - \frac{\partial s}{\partial y}(-1, \pm 1)\right) x \\ &+ \frac{1}{2}\left(\frac{\partial s}{\partial y}(-1, \pm 1) - 2\frac{\partial s}{\partial y}(0, \pm 1) + \frac{\partial s}{\partial y}(1, \pm 1)\right) x^2. \end{aligned}$$

Hence, if two such macro-elements are combined in y-direction the normal derivative along the common edge parallel to the x-axis (long side) is continuous. Clearly, this macro-element induces another interpolation operator $\hat{\Pi}^y : C^1(\Lambda) \to S(\hat{M})$:

$$\hat{\Pi}^y u(x,y) := \sum_{i \in \{-1,0,1\}} \sum_{j \in \{-1,1\}} \left(u(i,j)\hat{\ell}_i(x)\hat{\varphi}_j(y) + \frac{\partial u}{\partial y}(i,j)\hat{\ell}_i(x)\hat{\psi}_j(y) \right). \tag{3.75}$$

Here $\hat{\ell}_i \in P_2[-1,1]$ denotes the quadratic Lagrange basis function that corresponds to the node $i \in \{-1, 0, 1\}$, i.e.

$$\hat{\ell}_{-1} := x(x-1)/2, \quad \hat{\ell}_0 := -(x+1)(x-1), \quad \hat{\ell}_{+1} := (x+1)x/2.$$

Let $M = [x_0 - h_x/2, x_0 + h_x/2] \times [y_0 - h_y/2, y_0 + h_y/2$ denote a macro-element. From the representation (3.75) and the affine reference mapping $F_M : [-1,1] \to M$:

$$x = x_0 + h_x \hat{x}, \qquad y = y_0 + h_y \hat{y}, \tag{3.76}$$

it is easy to deduce for the interpolation operator $\Pi^y u := \hat{\Pi}^y \hat{u} \circ F_M^{-1}$ with $\hat{u} := u \circ F_M$ on the macro-element M in the world domain the stability property

$$\|\Pi^y u\|_{L_\infty(M)} \le C \left(\|u\|_{L_\infty(M)} + h_y \left\| \frac{\partial u}{\partial y} \right\|_{L_\infty(M)} \right). \tag{3.77}$$

Remark 30. Note that by construction h_y is the length of the small side of the macro-element M in the world domain. Hence, the first derivative in (3.77) is combined with a small multiplier.

Next we study the approximation properties of this interpolation operator.

Theorem 31. *For $u \in H^3(\Lambda)$ and a multi-index $\boldsymbol{\gamma}$ with $|\boldsymbol{\gamma}| \le 2$ we have the estimates*

$$\|\boldsymbol{D}^{\boldsymbol{\gamma}}(u - \Pi^y u)\|_0 \le C\, |\boldsymbol{D}^{\boldsymbol{\gamma}} u|_{3-|\boldsymbol{\gamma}|} \qquad \text{for } \boldsymbol{\gamma} \ne (2,0) \tag{3.78a}$$

$$\|(u - \Pi^y u)_{xx}\|_0 \le C\big(|u_{xx}|_1 + |u_x|_2\big) \tag{3.78b}$$

Proof. We shall apply Lemma 18 in order to prove (3.78a). Thus, we set $\boldsymbol{P} := \boldsymbol{Q} := P_{2-|\boldsymbol{\gamma}|}$. By a direct calculation similarly to (3.21) we observe that the additional error component involving the polynomial $q \in P_2(\Lambda)$ vanishes since $\Pi^y v = v$ holds true for any function $v \in Q_2(\Lambda) \supset P_2(\Lambda)$. It remains to specify the associate functionals $F_{ij}^{\boldsymbol{\gamma}}$ according to (3.12) for a given differential operator $\boldsymbol{D}^{\boldsymbol{\gamma}}$ with $|\boldsymbol{\gamma}| \le 2$. We use the same techniques as in Theorem 20. Firstly, it can be seen by applying the differential operator $\boldsymbol{D}^{\boldsymbol{\gamma}}$ to the representation (3.74) of an element $s \in S(\hat{M})$ that $\boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$ can be normed by

$$\sum_{(i,j)\in J_{\boldsymbol{\gamma}}} |F_{ij}(\cdot)| \qquad \text{with} \quad J_{\boldsymbol{\gamma}} := \{(i,j) \,:\, i = \gamma_1 + 1, \ldots, 3,\ j = \gamma_2 + 1, \ldots, 4\}.$$

Clearly, $F_{ij}(u) = F_{ij}(\Pi^y u)$ for all $i \in \{1,2,3\}$ and $j \in \{1,2,3,4\}$ because the divided differences are linear combinations of the interpolation data $\{u(k,\ell), u_y(k,\ell)\}_{k\in\{-1,0,1\},\ell\in\{-1,1\}}$. The associated functionals $F_{ij}^{\boldsymbol{\gamma}}$ for $(i,j) \in J_{\boldsymbol{\gamma}}$ are listed in Table 3.2. Using Sobolev embeddings like in the proof of Theorem 20 it is easy to check that $F_{ij}^{\boldsymbol{\gamma}} \in \big(H_{3-|\boldsymbol{\gamma}|}(\Lambda)\big)'$. Moreover,

$$F_{ij}(u) = F_{ij}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} u) \quad \text{and} \quad F_{ij}(\Pi^y u) = F_{ij}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} \Pi^y u)$$

for $(i,j) \in J_{\boldsymbol{\gamma}}$. The first identity follows from the techniques in the proof of Theorem 20, especially (3.33). A simple computation for each basis function in $S(\hat{M})$ shows the second identity, due to the linearity of F_{ij} and $F_{ij}^{\boldsymbol{\gamma}}$. Hence, indeed $F_{ij}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} \Pi^y u) = F_{ij}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} u)$. We shall demonstrate this procedure for F_{33}. A calculation gives

$$\begin{aligned} F_{33}(u) = u[-1,0,1;-1,-1,1] &= \frac{1}{2}\big(u(-1,\cdot)[-1,-1,1] - 2u(0,\cdot)[-1,-1,1] + u(1,\cdot)[-1,-1,1]\big) \\ &= \frac{1}{2}\int_{-1}^{1} s_1(y)\big(u_{yy}(-1,y) - 2u_{yy}(0,y) + u_{yy}(1,y)\big)\,\mathrm{d}y, \end{aligned} \tag{3.79}$$

where we used (3.33) with $s_1(y) = (1-y)/4$ and from which $F_{33}^{(0,1)}$ and $F_{33}^{(0,2)}$ can be deduced. Moreover, we may rewrite this identity to obtain

$$F_{33}(u) = \frac{1}{2}\int_{-1}^{1} s_1(y) \left(\int_0^1 u_{xyy}(x,y)\,\mathrm{d}x - \int_{-1}^{0} u_{xyy}(x,y)\,\mathrm{d}x \right) \mathrm{d}y.$$

A reinterpretation of this equation according to $F_{33}(u) = F_{33}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} u)$ gives $F_{33}^{(1,0)}$ and $F_{33}^{(1,1)}$. A computation shows $F_{33}(s) = F_{33}^{\boldsymbol{\gamma}}(\boldsymbol{D}^{\boldsymbol{\gamma}} s)$ for all $s \in S(\hat{M})$ and $|\boldsymbol{\gamma}| \le 2$, $\boldsymbol{\gamma} \ne (2,0)$. Hence, the estimate (3.78a) is proven.

For $\boldsymbol{\gamma} = (2,0)$ it appears impossible to provide the associated functionals by the above technique. Consider for instance the divided difference F_{33}. Using Taylor expansion it is possible to rewrite the equation (3.79) to

$$F_{33}(u) = \frac{1}{2}\int_{-1}^{1} s_1(y) \left(\int_{-1}^{0} (1+x) u_{xxyy}(x,y)\,\mathrm{d}x + \int_0^1 (1-x) u_{xxyy}(x,y)\,\mathrm{d}x \right) \mathrm{d}y.$$

This however comes at the price of demanding higher regularity. Clearly, we have to approach this problem differently. Let $\boldsymbol{P} := \{(2,0),(1,1),(1,0)\}$ and $q \in \boldsymbol{P}(\Lambda)$ denote the polynomial with

$$\int_\Lambda \boldsymbol{D}^{\boldsymbol{\alpha}}(u-q)\,\mathrm{d}\boldsymbol{x} = 0 \qquad \forall \boldsymbol{\alpha} \in \boldsymbol{P}.$$

By Lemma 16 the polynomial q exits and is unique. From Lemma 17 we can deduce by setting $v := (u-q)_{xx}$ that

$$\|v\|_1 = \|(u-q)_{xx}\|_1 \le C|(u-q)_{xx}|_1 = C|u_{xx}|_1, \tag{3.80}$$

since $\int_\Lambda v\,\mathrm{d}\boldsymbol{x} = \int_\Lambda (u-q)_{xx}\,\mathrm{d}\boldsymbol{x} = 0$. Similarly, Lemma 17 implies that for $v := (u-q)_x$ we find

$$\|v\|_2 = \|(u-q)_x\|_2 \le C|(u-q)_x|_2 = C|u_x|_2, \tag{3.81}$$

based on

$$\int_\Lambda \boldsymbol{D}^{\boldsymbol{\alpha}} v\,\mathrm{d}\boldsymbol{x} = 0 \qquad \forall\,|\boldsymbol{\alpha}| \le 1 \quad \Leftrightarrow \quad \int_\Lambda \boldsymbol{D}^{\boldsymbol{\alpha}}(u-q)_x\,\mathrm{d}x = 0 \qquad \forall\,|\boldsymbol{\alpha}| \le 1.$$

Next from (3.78a) for $\boldsymbol{\gamma} = (1,0)$ we obtain the following stability estimate

$$\left\|\left(\Pi^y v\right)_x\right\|_0 \le \|v_x\|_0 + \left\|\left(v - \Pi^y v\right)_x\right\|_0 \le C\|v_x\|_2. \tag{3.82}$$

A triangle inequality implies due to $q = \Pi^y q$ that

$$\left\|\left(u - \Pi^y u\right)_{xx}\right\|_0 \le \|(u-q)_{xx}\|_0 + \left\|\left(\Pi^y(q-u)\right)_{xx}\right\|_0. \tag{3.83}$$

The first summand on the right hand side of (3.83) is estimated using (3.80), while for the other one we use the inverse estimate

$$\|s_{xx}\|_0 \le C\|s_x\|_0 \quad \forall s \in S,$$

which is easily verified in the four dimensional space $\boldsymbol{D}^{(2,0)}S(\hat{M})$ over the reference macro-element. In fact, the optimal constant in this estimate is given by $C = \sqrt{3}$. Hence, by (3.83),

$$\left\|\left(u - \Pi^y u\right)_{xx}\right\|_0 \le C\left(|u_{xx}|_1 + \left\|\left(\Pi^y(q-u)\right)_x\right\|_0\right).$$

We finish the proof of (3.78b) by using (3.82) for $v = q - u$ and (3.81). □

Using affine equivalence (cf. (3.76) and the proof of Theorem 19) we obtain on a macro-element M in the world domain the following result.

Corollary 32. *For $u \in H^3(M)$ and a multi-index $\boldsymbol{\gamma}$ with $|\boldsymbol{\gamma}| \le 2$ we have the estimates*

$$\|\boldsymbol{D}^{\boldsymbol{\gamma}}(u - \Pi^y u)\|_{0,M} \le C \sum_{|\boldsymbol{\alpha}|=3-|\boldsymbol{\gamma}|} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \left\|\boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} u\right\|_{0,M} \qquad \textit{for } \boldsymbol{\gamma} \ne (2,0) \tag{3.84a}$$

$$\left\|\left(u - \Pi^y u\right)_{xx}\right\|_{0,M} \le C\left(\sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}} u_{xx}\|_{0,M} + \sum_{|\boldsymbol{\alpha}|=2} \frac{\boldsymbol{h}_M^{\boldsymbol{\alpha}}}{h_x}\|\boldsymbol{D}^{\boldsymbol{\alpha}} u_x\|_{0,M}\right) \tag{3.84b}$$

Remark 31. By construction h_x denotes the length of the long side of M. Hence, the estimate (3.84b) is useful even in the anisotropic case.

Before we end this section we prove a suboptimal but useful error estimate for $\boldsymbol{\gamma} = (0,0)$.

Lemma 33. *Let $u \in H^3(M)$ then*

$$\|u - \Pi^y u\|_{0,M} \le C \sum_{|\boldsymbol{\alpha}|=2} \left(\boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}} u\|_{0,M} + \boldsymbol{h}_M^{\boldsymbol{\alpha}} h_y \|\boldsymbol{D}^{\boldsymbol{\alpha}} u_y\|_{0,M}\right).$$

$\boldsymbol{\gamma}$	$\dim \boldsymbol{D}^{\boldsymbol{\gamma}} S(\hat{M})$	associate functionals
$(0,0)$	12	$F_{ij}^{(0,0)} \coloneqq F_{ij} \quad i \in \{1,2,3\}$ and $j \in \{1,2,3,4\}$
$(1,0)$	8	$F_{2\ell}^{(1,0)}(v) \coloneqq \int_{-1}^{0} \frac{\partial^{\ell-1} v}{\partial y^{\ell-1}}(x,-1)\,\mathrm{d}x, \quad \ell \in \{1,2\}$
		$F_{2k}^{(1,0)}(v) \coloneqq \int_{-1}^{1}\int_{-1}^{0} s_{k-2}(y) v_{yy}(x,y)\,\mathrm{d}x\mathrm{d}y, \quad k \in \{3,4\}$
		$F_{3\ell}^{(1,0)}(v) \coloneqq \frac{1}{2}\int_{0}^{1} \frac{\partial^{\ell-1} v}{\partial y^{\ell-1}}(x,-1)\,\mathrm{d}x - \frac{1}{2}\int_{-1}^{0} \frac{\partial^{\ell-1} v}{\partial y^{\ell-1}}(x,-1)\,\mathrm{d}x$
		$F_{3k}^{(1,0)}(v) \coloneqq \frac{1}{2}\int_{-1}^{1} s_{k-2}(y)\left(\int_{0}^{1} v_{yy}(x,y)\,\mathrm{d}x - \int_{-1}^{0} v_{yy}(x,y)\,\mathrm{d}x\right)\mathrm{d}y$
$(0,1)$	9	$F_{12}^{(0,1)}(v) \coloneqq v(-1,-1)$
		$F_{1k}^{(0,1)}(v) \coloneqq \int_{-1}^{1} s_{k-2}(y) v_y(-1,y)\,\mathrm{d}y, \quad k \in \{3,4\}$
		$F_{22}^{(0,1)}(v) \coloneqq v(0,-1) - v(-1,-1)$
		$F_{2k}^{(0,1)}(v) \coloneqq \int_{-1}^{1}\int_{-1}^{0} s_{k-2}(y) v_{xy}(x,y)\,\mathrm{d}x\mathrm{d}y, \quad k \in \{3,4\}$
		$F_{32}^{(0,1)}(v) \coloneqq \frac{1}{2}\big(v(-1,-1) - 2v(0,-1) + v(1,-1)\big)$
		$F_{3k}^{(0,1)}(v) \coloneqq \frac{1}{2}\int_{-1}^{1} s_{k-2}(y)\big(v_y(-1,y) - 2v_y(0,y) + v_y(1,y)\big)\,\mathrm{d}y$
$(1,1)$	6	$F_{22}^{(1,1)}(v) \coloneqq \int_{-1}^{0} v(x,-1)\,\mathrm{d}x$
		$F_{2k}^{(1,1)}(v) \coloneqq \int_{-1}^{1} s_{k-2}(y)\int_{-1}^{0} v_y(x,y)\,\mathrm{d}x\mathrm{d}y, \quad k \in \{3,4\}$
		$F_{32}^{(1,1)}(v) \coloneqq \frac{1}{2}\int_{0}^{1} v(x,-1)\,\mathrm{d}x - \frac{1}{2}\int_{-1}^{0} v(x,-1)\,\mathrm{d}x$
		$F_{3k}^{(1,1)}(v) \coloneqq \frac{1}{2}\int_{-1}^{1} s_{k-2}(y)\left(\int_{0}^{1} v_y(x,y)\,\mathrm{d}x - \int_{-1}^{0} v_y(x,y)\,\mathrm{d}x\right)\mathrm{d}y$
$(0,2)$	6	$F_{1k}^{(0,2)}(v) \coloneqq \int_{-1}^{1} s_{k-2}(y) v(-1,y)\,\mathrm{d}y, \quad k \in \{3,4\}$
		$F_{2k}^{(0,2)}(v) \coloneqq \int_{-1}^{1} s_{k-2}(y)\int_{-1}^{0} v_x(x,y)\,\mathrm{d}x\mathrm{d}y, \quad k \in \{3,4\}$
		$F_{3k}^{(0,2)}(v) \coloneqq \frac{1}{2}\int_{-1}^{1} s_{k-2}(y)\big(v(-1,y) - 2v(0,y) + v(1,y)\big)\mathrm{d}y$

Table 3.2: Associated functionals $F_{i,j}^{\boldsymbol{\gamma}}$ for the operator Π^y over $\hat{M}$ with respect to $\boldsymbol{D}^{\boldsymbol{\gamma}}$ and $s_1(y) = (1-y)/4$, $s_2(y) = y/4$.

Proof. Let $q \in \boldsymbol{P}_1(\Lambda)$ denote the linear polynomial such that

$$\int_\Lambda \boldsymbol{D}^{\boldsymbol{\alpha}}(u-q)\mathrm{d}\boldsymbol{x} = 0 \quad \forall \boldsymbol{\alpha} \in \boldsymbol{P}_1 := \{(0,0),(1,0),(0,1)\}.$$

Then from Lemma 17 it follows that $\|u-q\|_2 \leq C|u-q|_2 = C|u|_2$. Using this and $\Pi^y q = q$ we see that

$$\begin{aligned}
\|u - \Pi^y u\|_0 &\leq \|u-q\|_0 + \|\Pi^y(q-u)\|_0 \\
&\leq \|u-q\|_2 + C \sum_{i\in\{-1,0,1\}} \sum_{j\in\{-1,1\}} \left(|\Pi^y(q-u)(i,j)| + \left| \frac{\partial \Pi^y(q-u)}{\partial y}(i,j) \right| \right) \\
&\leq |u|_2 + C \sum_{i\in\{-1,0,1\}} \sum_{j\in\{-1,1\}} \left(|(q-u)(i,j)| + \left| \frac{\partial (q-u)}{\partial y}(i,j) \right| \right) \\
&\leq |u|_2 + C \left(\|u-q\|_2 + \left\| \frac{\partial (q-u)}{\partial y} \right\|_2 \right).
\end{aligned}$$

From

$$\|(q-u)_y\|_2 \leq \|q-u\|_2 + |(q-u)_y|_2 \leq C|u|_2 + |u_y|_2,$$

the estimate follows on the reference macro $\hat{M}$. The assertion of the lemma is again easily obtained by affine transformation. □

3.7 Application of macro-element interpolation on a tensor product Shishkin mesh

As an application of the anisotropic quasi-interpolation error estimates obtained we want to examine the approximation error of the solution of a reaction-diffusion problem on an anisotropic mesh. Let u denote the solution of the singularly perturbed linear reaction-diffusion problem

$$-\varepsilon\Delta u + cu = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \tag{3.85}$$

where $0 < \varepsilon \ll 1$, $0 < 2(c^\star)^2 \leq c$ and c and f are smooth functions on some bounded two dimensional domain Ω with Lipschitz-continuous boundary $\partial\Omega$. We consider the unit square $\Omega := (0,1)^2$ with the four edges

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

In the corners of the domain Ω derivatives of u are unbounded, in general. One refers to the solution components that cause this phenomenon as corner singularities. If we however assume the corner compatibility conditions

$$f(0,0) = f(1,0) = f(0,1) = f(1,1) = 0, \tag{3.86}$$

then third derivatives of u are smooth up to the boundary, $u \in C^3(\overline{\Omega})$, see, e.g. [29].

The following solution decomposition is taken from [45, Lemma 1.1 and Lemma 1.2]

Lemma 34. *The solution $u \in C^3(\overline{\Omega})$ of* (3.85) *can be decomposed as*

$$u = S + \sum_{i=1}^{4} E_i + E_{12} + E_{23} + E_{34} + E_{41}. \tag{3.87a}$$

Here E_i is a boundary layer associated with the edge Γ_i. Similarly, E_{ij} is a corner layer associated with the corner that is formed by the edges Γ_i and Γ_j. Moreover, there are positive

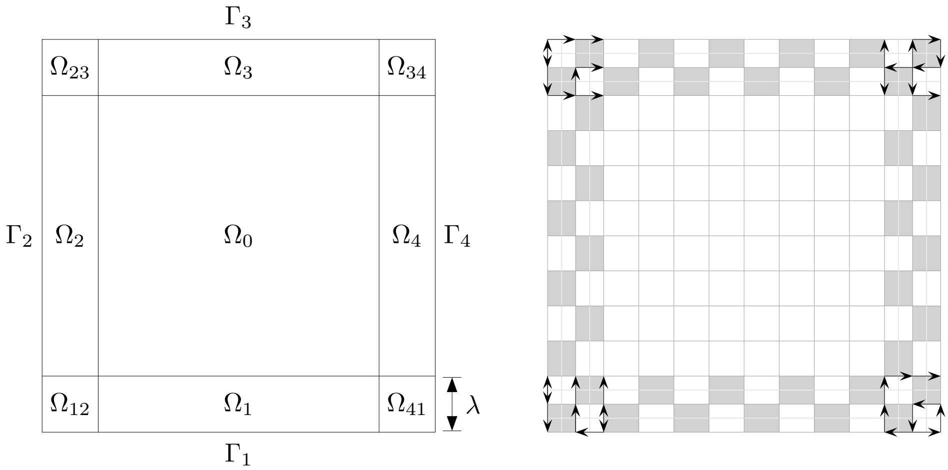

Figure 3.7: Domain decomposition (left) and anisotropic mesh Ω^N (right) for $N = 16$, corresponding macro-element triangulation $\mathcal{M}^{16}$ of $\Omega \setminus \Omega_0$ as checkerboard and possible choice for $\sigma_{i,j}$ symbolized by black arrows pointing to the corresponding mesh node $\boldsymbol{X}_{ij}$.

constants $C > 0$ such that for all $(x, y) \in \overline{\Omega}$ and $0 \leq i + j \leq 3$ we have

$$\left| \frac{\partial^{i+j} S(x,y)}{\partial x^i \partial y^j} \right| \leq C\big(1 + \varepsilon^{1-(i+j)/2}\big) \tag{3.87b}$$

$$\left| \frac{\partial^{i+j} E_1(x,y)}{\partial x^i \partial y^j} \right| \leq C\big(1 + \varepsilon^{1-i/2}\big)\varepsilon^{-j/2} \mathrm{e}^{-c^\star y/\sqrt{\varepsilon}} \tag{3.87c}$$

$$\left| \frac{\partial^{i+j} E_{12}(x,y)}{\partial x^i \partial y^j} \right| \leq C\varepsilon^{-(i+j)/2} \mathrm{e}^{-c^\star (x+y)/\sqrt{\varepsilon}} \tag{3.87d}$$

and analogous bounds for the other boundary and corner layers.

Next we introduce a standard domain decomposition. Let N denote a multiple of eight — N will later denote the number of mesh intervals in each coordinate direction — and define the transition point

$$\lambda \coloneqq \min\left\{ \frac{1}{4}, \frac{\lambda_0 \sqrt{\varepsilon}}{c^\star} \ln N \right\} \qquad \text{with } \lambda_0 \geq 3. \tag{3.88}$$

For our subsequent error analysis we shall make the practical and standard assumption

$$\sqrt{\varepsilon} \leq CN^{-1},$$

from which $\lambda < 1/4$ follows.

For our approximation error analysis we use a standard approach and split the domain into several subdomains

$$\begin{aligned}
\Omega_0 &\coloneqq (\lambda, 1-\lambda)^2, & \Omega_{12} &\coloneqq (0,\lambda)^2, \\
\Omega_1 &\coloneqq (\lambda, 1-\lambda) \times (0,\lambda), & \Omega_{23} &\coloneqq (0,\lambda) \times (1-\lambda, 1), \\
\Omega_2 &\coloneqq (0,\lambda) \times (\lambda, 1-\lambda), & \Omega_{34} &\coloneqq (1-\lambda, \lambda)^2, \\
\Omega_3 &\coloneqq (\lambda, 1-\lambda) \times (1-\lambda, 1), & \Omega_{41} &\coloneqq (1-\lambda, \lambda) \times (0,\lambda), \\
\Omega_4 &\coloneqq (1-\lambda, 1) \times (\lambda, 1-\lambda), & \Omega_f &\coloneqq \Omega_{12} \cup \Omega_{23} \cup \Omega_{34} \cup \Omega_{41},
\end{aligned}$$

as shown in the left of Figure 3.7.

We use λ to construct a 1D Shishkin mesh as follows: subdivide each of the intervals $[0, \lambda]$, $[1-\lambda, 1]$ into $N/4$ subintervals, equidistantly. Giving the small grid size $h = \lambda/(N/4-2)$. Next, divide the third subinterval $[\lambda, 1-\lambda]$ into $N/2$ subintervals of same size H. Hence, the mesh is uniform in each of the subintervals $[0, \lambda]$, $[\lambda, 1-\lambda]$ and $[1-\lambda, 1]$ but it changes from fine to coarse at the transition points λ and $1-\lambda$. Remark that since N is a multiple of eight the number of subintervals within $[0,\lambda]$, $[1-\lambda,1]$ and $[\lambda,1-\lambda]$ is even. Finally, form the tensor product of this one-dimensional mesh with itself to obtain our anisotropic Shishkin mesh Ω^N with the mesh nodes $\{(x_i,y_j)\}_{i,j=0,\ldots,N}$.

Note that by the definition of λ in the inner subdomain Ω_0 all the layers have declined such that they can be bounded pointwise by a constant times $N^{-\lambda_0}$. This is however not true for their derivatives. Consequently, it is very challenging to define a C^1 (quasi-)interpolant of $u \in C^3(\overline{\Omega})$ in the function space of piecewise biquadratics over Ω^N featuring anisotropic error estimates. We relax this too ambitious objective by defining a quasi-interpolant $u^\star$ of u, such that the normal derivative of $u^\star$ is continuous only across certain edges of Ω^N. For this purpose we shall use the results of the previous sections on macro-element quasi-interpolation.

In Ω_f, i.e. close to corners of the domain, we combine four neighbouring elements of equal shape to form a macro-element $M = [x_{i-1}, x_{i+1}] \times [y_{j-1}, y_{j+1}]$ and in $\Omega_1 \cup \Omega_3$ we combine two neighbouring elements to get $M = [x_i, x_{i+1}] \times [y_{j-1}, y_{j+1}]$ as shown in the right of Figure 3.7. In $\Omega_2 \cup \Omega_4$ we proceed likewise. We denote the obtained macro-element triangulation by $\mathcal{M}^N$. Note that the mesh Ω^N can also be understood as the result of a refinement routine of the macro-element mesh $\mathcal{M}^N$.

The elements of our Shishkin mesh Ω^N are axis-parallel rectangles with side lengths

$$h := \frac{4\lambda}{N-8} = \mathcal{O}(\sqrt{\varepsilon} N^{-1} \ln N) \qquad \text{or} \qquad H := \frac{2(1-2\lambda)}{N} \sim N^{-1}. \tag{3.89}$$

The sizes of a macro-element are equivalent to the sizes of the containing mesh elements.

Close to the corners of the domain, i.e. in Ω_f we want to approximate u by the quasi-interpolant $\tilde{\Pi}u$, see Subsection 3.5.3. Hence, we have to specify how the macro-element edges $\sigma_{i,j}$ associated with the macro-element vertices $\boldsymbol{X}_{ij} \in \overline{\Omega_f}$ are chosen. If we want to satisfy Assumption 1 on our anisotropic mesh we have to choose carefully whenever $\boldsymbol{X}_{ij}$ lies on one of the lines $x = x_{N/4} = \lambda$, $x = x_{3N/4} = 1-\lambda$ or $y = y_{N/4} = \lambda$, $y = y_{3N/4} = 1-\lambda$ where the mesh sizes change abruptly. Restricted to Ω_f our Shishkin mesh Ω^N is (quasi-)uniform, hence any choice that satisfies

$$\sigma_{i,j} \subset \overline{\Omega_f} \tag{3.90}$$

is possible. One may fulfill (3.90) as demonstrated in the right of Figure 3.7. In that Figure a macro-element edge $\sigma_{i,j}$ is symbolized by an arrow pointing to $\boldsymbol{X}_{ij}$.

Let us recall the functions $\varphi_i, \psi_i \in C^1[0,1]$, supported within $[x_{1-2}, x_{i+2}]$, defined by

$$\varphi_i(x) := \begin{cases} \dfrac{1}{2} + \dfrac{x-x_{i-1}}{h_{i-1}} + \dfrac{(x-x_{i-1})^2}{2h_{i-1}^2} & \text{in } [x_{i-2}, x_{i-1}], \\[2mm] \dfrac{1}{2} + \dfrac{x-x_{i-1}}{h_{i-1}} - \dfrac{(x-x_{i-1})^2}{2h_{i-1}^2} & \text{in } [x_{i-1}, x_{i}], \\[2mm] \dfrac{1}{2} - \dfrac{x-x_{i+1}}{h_{i+1}} - \dfrac{(x-x_{i+1})^2}{2h_{i+1}^2} & \text{in } [x_{i}, x_{i+1}], \\[2mm] \dfrac{1}{2} - \dfrac{x-x_{i+1}}{h_{i+1}} + \dfrac{(x-x_{i+1})^2}{2h_{i+1}^2} & \text{in } [x_{i+1}, x_{i+2}], \end{cases}$$

x	$\varphi_i(x)$	$\varphi_i'(x)$
x_{i-2}	0	0
x_{i-1}	0.5	>0
x_i	1	0
x_{i+1}	0.5	<0
x_i	0	0

$$\psi_i(x) := \begin{cases} -\dfrac{h_{i-1}}{4} - \dfrac{x-x_{i-1}}{2} - \dfrac{(x-x_{i-1})^2}{4h_{i-1}} & \text{in } [x_{i-2}, x_{i-1}], \\[2mm] -\dfrac{h_{i-1}}{4} - \dfrac{x-x_{i-1}}{2} + \dfrac{3(x-x_{i-1})^2}{4h_{i-1}} & \text{in } [x_{i-1}, x_{i}], \\[2mm] \dfrac{h_{i+1}}{4} - \dfrac{x-x_{i+1}}{2} - \dfrac{3(x-x_{i+1})^2}{4h_{i+1}} & \text{in } [x_{i}, x_{i+1}], \\[2mm] \dfrac{h_{i+1}}{4} - \dfrac{x-x_{i+1}}{2} + \dfrac{(x-x_{i+1})^2}{4h_{i+1}} & \text{in } [x_{i+1}, x_{i+2}], \end{cases}$$

x	$\psi_i(x)$	$\psi_i'(x)$
x_{i-2}	0	0
x_{i-1}	<0	-0.5
x_i	0	1
x_{i+1}	>0	-0.5
x_i	0	0

with $h_{i-1} := x_{i-1} - x_{i-2} = x_i - x_{i-1}$ and $h_{i+1} := x_{i+1} - x_i = x_{i+2} - x_{i+1}$, i.e. $h_i = h$ for $i < N/4$ or $i > 3N/4$ and $h_i = H$ else. Based on these one-dimensional functions one can define the global basis functions in the world domain

$$\begin{aligned} \varphi_{i,j}(x,y) &:= \varphi_i(x)\varphi_j(y), & \phi_{i,j}(x,y) &:= \psi_i(x)\varphi_j(y), \\ \chi_{i,j}(x,y) &:= \varphi_i(x)\psi_j(y), & \psi_{i,j}(x,y) &:= \psi_i(x)\psi_j(y), \end{aligned} \qquad i,j = 0,\dots,N. \tag{3.91}$$

Now we are able to define our quasi-interpolation operator into the finite element space

$$V^N := \{v \in H^1(\Omega) : v|_T \in Q_2(T) \ \ \forall T \in \Omega^N\}. \tag{3.92}$$

As already mentioned, for $M = [x_{i-1}, x_{i+1}] \times [y_{j-1}, y_{j+1}] \subset \overline{\Omega_f}$, $M \in \mathcal{M}^N$ close to the corners of the domain we use the quasi-interpolation operator $\tilde{\Pi}$ from Subsection 3.5.3, i.e.

$$u^\star\big|_M = (\tilde{\Pi}u)|_M = \sum_{\substack{k=i-1,i+1\\ \ell=j-1,j+1}} u(x_k,y_\ell)\varphi_{k,\ell} + u_x(x_k,y_\ell)\phi_{k,\ell} + u_y(x_k,y_\ell)\chi_{k,\ell} + a_{k,\ell}\psi_{k,\ell}.$$

The coefficients $a_{k,\ell}$ depend on the direction of $\sigma_{k,\ell}$ given by (3.57):

$$a_{k,\ell} = \begin{cases} \displaystyle\int_{\sigma_{k,\ell}} \frac{\partial^2 u(x,y_\ell)}{\partial x \partial y}\psi_k^d(x)\,\mathrm{d}x & \text{if } \sigma_{k,\ell} \text{ is horizontal,} \\ \displaystyle\int_{\sigma_{k,\ell}} \frac{\partial^2 u(x_k,y)}{\partial x \partial y}\psi_\ell^d(y)\,\mathrm{d}y & \text{if } \sigma_{k,\ell} \text{ is vertical,} \end{cases}$$

with the dual basis functions ψ_k^d obtained in (3.50):

$$\psi_k^d(x) := \begin{cases} -\dfrac{h_{k-1}^2 + 12h_{k-1}(x - x_{k-1})}{2h_{k-1}^3} \begin{cases} -\dfrac{3(x-x_{k-1})^2}{h_{k-1}^3}, & x_{k-2} \le x \le x_{k-1}, \\ +\dfrac{9(x-x_{k-1})^2}{h_{k-1}^3}, & x_{k-1} \le x \le x_k, \end{cases} \\ -\dfrac{h_{k+1}^2 - 12h_{k+1}(x - x_{k+1})}{2h_{k+1}^3} \begin{cases} +\dfrac{9(x-x_{k+1})^2}{h_{k+1}^3}, & x_k \le x \le x_{k+1}, \\ -\dfrac{3(x-x_{k+1})^2}{h_{k+1}^3}, & x_{k+1} \le x \le x_{k+2}. \end{cases} \end{cases}$$

In Ω_0 we use on the element level the standard biquadratic nodal interpolant u^I of u. Set $\mathcal{I} := \{\frac{N}{4}, \frac{N}{4}+\frac{1}{2}, \frac{N}{4}+1, \frac{N}{4}+\frac{3}{2}, \dots, \frac{3}{4}N\}$. Let ℓ_i denote the *1D quadratic Lagrange basis functions*, $i \in \mathcal{I}$ with

$$\ell_i(x) = \begin{cases} \dfrac{2}{h_i^2}(x - x_{i-1})(x - x_{i-1/2}), & x_{i-1} \le x \le x_i \\ \dfrac{2}{h_{i+1}^2}(x_{i+1} - x)(x_{i+1/2} - x), & x_i \le x \le x_{i+1} \end{cases} \qquad \text{for } i \in \mathcal{I} \cap \mathbb{N},$$

$$\ell_{i+1/2}(x) = \frac{4}{h_{i+1}^2}(x_{i+1} - x)(x - x_i), \quad \text{for } i \in \mathcal{I} \cap \mathbb{N}, i \neq N,$$

where $x_{i+1/2} := (x_i + x_{i+1})/2$, $i \in \mathcal{I} \cap \mathbb{N}$ with $i \neq N$, denotes the midpoint of the interval $[x_i, x_{i+1}]$. Now for $T \subset \overline{\Omega_0}$ we set

$$u^\star|_T(x,y) := u^I|_T(x,y) = \sum_{i,j \in \mathcal{I}} u(x_i,y_j)\ell_i(x)\ell_j(y), \quad (x,y) \in T.$$

Finally, we need some modified anisotropic macro-interpolation operator in $\bigcup_{i=1}^4 \Omega_i$ to glue these interpolants together. Let us consider a macro-element $M = [x_i, x_{i+1}] \times [y_{j-1}, y_{j+1}] \subset \overline{\Omega_1}$. The two elements contained in this macro-element have a long side of length H in x-direction and a short one in y-direction (with length h). On all of these macro-elements $M \subset \overline{\Omega_1}$ that are not adjacent to $\partial\Omega_0$ we use the anisotropic macro-interpolation Π^y as introduced and analyzed in Section 3.6, c.p. (3.75):

$$u^\star|_M(x,y) = \Pi^y u(x,y) := \sum_{\substack{k\in\{i,i+1/2,i+1\}\\ m\in\{j-1,j+1\}}} \left(u(k,m)\ell_k(x)\varphi_m(y) + \frac{\partial u}{\partial y}(k,m)\ell_k(x)\psi_m(y)\right).$$

We use the same interpolation operator for $M \subset \overline{\Omega_3}$. In $\Omega_2 \cup \Omega_4$ we use Π^x instead. Hence, the roles of x and y are interchanged, there.

On macro-elements that are adjacent to $\partial\Omega_0$ we modify the anisotropic macro-interpolation operator in order to archive continuity of the normal derivative $\partial_n u^\star$ across $\partial\Omega_0$. Let for instance $M = [x_i, x_{i+1}] \times [y_{N/4-2}, y_{N/4}] \subset \overline{\Omega_1}$ denote such a macro-element. Then on M the interpolant $u^\star$ is of the form:

$$\begin{aligned} u^\star|_M(x,y) = \sum_{k\in\{i,i+1/2,i+1\}} \Bigg(&\sum_{j\in\{N/4-2,N/4\}} u(x_k,y_j)\ell_k(x)\varphi_j(y) \\ &+ \frac{\partial u}{\partial y}(x_k, y_{N/4-2})\ell_k(x)\psi_{N/4-2}(y) + \frac{\partial(u^I|_{\Omega_0})}{\partial y}(x_k,y_{N/4})\ell_k(x)\psi_{N/4}(y)\Bigg). \end{aligned}$$

In the other subdomains we proceed likewise. Since $\left.\frac{\partial(u^I|_{\Omega_0})}{\partial y}\right|_{M\cap\Omega_0}$ and $\left.\frac{\partial(u^\star|_M)}{\partial y}\right|_{M\cap\Omega_0}$ are quadratic polynomials they are indeed uniquely determined by the values in three distinct points along the edge where they coincide. Note further that $\frac{\partial(u^I|_{\Omega_0})}{\partial y}(x_k, y_{N/4})$ is simply a linear combination of the nodal values $u(x_k, y_{N/4})$, $u(x_k, y_{N/4+1/2})$ and $u(x_k, y_{N/4+1})$. Hence, this coefficient is well defined along element interfaces due to the continuity of u^I.

Summarizing,

$$u^\star(x,y) = \begin{cases} (\tilde\Pi u)|_M & (x,y)\in M\subset\overline{\Omega_f}, \\ (\Pi^y u)|_M + \displaystyle\sum_{\substack{i=N/2\\ j\in\{N/4,3N/4\}}}^{3N/2} \left.\frac{\partial(u^I-u)}{\partial y}\right|_{\Omega_0}(x_{i/2},y_j)\,\ell_{i/2}(x)\psi_j(y) & (x,y)\in M\subset\overline{\Omega_1}\cup\overline{\Omega_3}, \\ (\Pi^x u)|_M + \displaystyle\sum_{\substack{j=N/2\\ i\in\{N/4,3N/4\}}}^{3N/2} \left.\frac{\partial(u^I-u)}{\partial x}\right|_{\Omega_0}(x_i,y_{j/2})\,\psi_i(x)\ell_{j/2}(y) & (x,y)\in M\subset\overline{\Omega_2}\cup\overline{\Omega_4}, \\ u^I|_T & (x,y)\in T\subset\overline{\Omega_0}. \end{cases}$$

By construction the normal derivative of $u^\star$ is only discontinuous along short edges of anisotropic elements (type-III edges) and interior edges of Ω_0 (type I edges). For some illustration see Figure 3.8.

Before we analyze $u^\star$ on the Shishkin mesh ω^N let us assign a type to each element edge as shown in the left of Figure 3.8:

Definition 6. A type-I edge $e \subset \overline{\Omega_0}$ is a long edge given as the intersection of two isotropic elements. An edge that belongs to at least one anisotropic element is of type II if it is a long one. Otherwise it is short and of type III. A remaining type-IV edge $e \subset \overline{\Omega_f}$ belongs to two small and square shaped elements and is close to a corner of Ω. Let $\mathcal{E}(I)$ be the set of interior edges of type I and introduce similar symbols for $\mathcal{E}(II)$, $\mathcal{E}(III)$ and $\mathcal{E}(IV)$.

First we show that the modification is small in various L_2-based norms. By the solution decomposition (3.87), standard interpolation error estimates and the choice of λ wee find that

$$\begin{aligned} |u-u^I|_{W_{1,\infty}(\Omega_0)} &\le |S-S^I|_{W_{1,\infty}(\Omega_0)} + |(u-S)-(u-S)^I|_{W_{1,\infty}(\Omega_0)} \\ &\le C\big(H^2|S|_{W_{3,\infty}} + |u-S|_{W_{1,\infty}(\Omega_0)} + |(u-S)^I|_{W_{1,\infty}(\Omega_0)}\big) \\ &\le C\big(H^2\varepsilon^{-1/2} + \varepsilon^{-1/2}N^{-\lambda_0} + H^{-1}N^{-\lambda_0}\big) \le C\varepsilon^{-1/2}N^{-2}. \end{aligned}$$

Here we also used an inverse estimate. Let ω_1 denote the strip of macro-elements in Ω_1 that are adjacent to Ω_0 then for $|\boldsymbol{\alpha}| \le 2$ it holds

$$\begin{aligned} &\left\|\boldsymbol{D}^{\boldsymbol{\alpha}} \sum_{i=N/2}^{3N/2} \left.\frac{\partial(u^I-u)}{\partial y}\right|_{\Omega_0}(x_{i/2},y_{N/4})\,\ell_{i/2}\psi_{N/4}\right\|_{0,\omega_1} \le |u-u^I|_{W_{1,\infty}(\Omega_0)} \left\|\sum_{i=N/2}^{3N/2} \ell_{i/2}^{(\alpha_1)}\psi_{N/4}^{(\alpha_2)}\right\|_{0,\omega_1} \\ &\quad\le C\varepsilon^{-1/2}N^{-2}\,\mathrm{meas}(\Omega_1)^{1/2} \left\|\sum_{i=N/2}^{3N/2}\ell_{i/2}^{(\alpha_1)}\right\|_{L_\infty([x_{N/4},x_{3N/4}])} \|\psi_{N/4}^{(\alpha_2)}\|_{L_\infty([y_{N/4-2},y_{N/4}])} \qquad (3.93) \\ &\quad\le C\varepsilon^{-1/4}N^{-5/2}(\ln N)^{1/2}H^{-\alpha_1}h^{1-\alpha_2} \le C\varepsilon^{1/4-\alpha_2/2}N^{-7/2+\alpha_1+\alpha_2}(\ln N)^{1/2-\alpha_2}. \end{aligned}$$

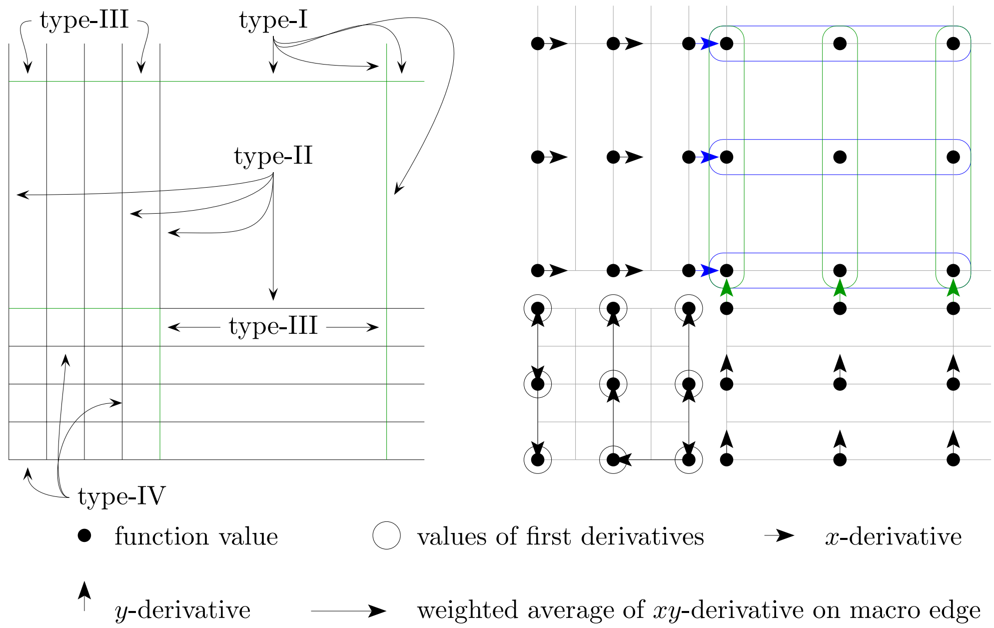

Figure 3.8: The normal derivative of $u^\star$ is discontinuous along the edges of type I and III highlighted in green (left) and linear functionals of u that enter in the definition of $u^\star$ in the various subdomains (right).

For $|\boldsymbol{\alpha}| = 2$ the L_2 norms have to be read as norms in the broken Sobolev space over $\mathcal{M}^N$. Bounds for the other three strips ω_i in Ω_i for $i = 2, 3, 4$ that are adjacent to Ω_0 follow similarly.

Since the Shishkin mesh is (quasi-)uniform in Ω_f and by the choice of the macro-element edges according to (3.90) the interpolation error estimates for $\tilde{\Pi}$ simplify to (c.p. Theorem 30)

$$\left|v - \tilde{\Pi} v\right|_{k,M} \le C h^{3-k} |v|_{3,\omega_M} \quad \text{for } v \in H^3(\omega_M) \text{ and } k \le 2. \tag{3.94}$$

Next we estimate the approximation error of $u - u^\star$:

Lemma 35. *There exists a constant $C > 0$ such that*

$$\|u - u^\star\|_0 \le C\big(N^{-2} + \varepsilon^{1/4} N^{-2} (\ln N)^2\big), \tag{3.95a}$$

$$\varepsilon^{1/4} |u - u^\star|_1 \le C\big(\varepsilon^{1/4} N^{-1} + N^{-2} (\ln N)^2\big), \tag{3.95b}$$

$$\varepsilon^{3/4} \Big(\sum_{M \in \mathcal{M}^N} |u - u^\star|^2_{2,M} \Big)^{1/2} \le C\big(\varepsilon N^{-1} (\ln N)^2 + N^{-1} \ln N\big). \tag{3.95c}$$

If $\varepsilon^{1/4} \le (\ln N)^{-2}$, then

$$\|u - u^\star\|_0 \le C N^{-2}, \tag{3.95d}$$

$$\varepsilon^{3/4} \Big(\sum_{M \in \mathcal{M}^N} |u - u^\star|^2_{2,M} \Big)^{1/2} \le C N^{-1} \ln N. \tag{3.95e}$$

If $|S|_3 \le C \varepsilon^{-1/4}$, then

$$\varepsilon^{1/4} |u - u^\star|_1 \le C N^{-2} (\ln N)^2. \tag{3.95f}$$

Suppose $\varepsilon^{1/4} \le (\ln N)^{-3}$ and $|S|_3 + \sum_{i \in 1,3} \|\boldsymbol{D}^{(3,0)} E_i\|_{0,\Omega_i} + \sum_{j \in 2,4} \|\boldsymbol{D}^{(0,3)} E_j\|_{0,\Omega_j} \le C$, then

$$\|u - u^\star\|_0 \le C N^{-3} (\ln N)^3. \tag{3.95g}$$

Proof. We use the solution decomposition (3.87) several times without mentioning it explicitly and different techniques in each subdomain.

In Ω_f the approximation error is small because the mesh is very fine. We use (3.94):

$$\begin{aligned} |u - u^\star|_{k,\Omega_f} = \left|u - \tilde{\Pi} u\right|_{k,\Omega_f} \le C h^{3-k} |u|_{3,\Omega_f} &\le C h^{3-k} \varepsilon^{-3/2} \operatorname{meas}(\Omega_f)^{1/2} \\ &= C \varepsilon^{(1-k)/2} N^{k-3} (\ln N)^{2-k}. \end{aligned} \tag{3.96}$$

In Ω_0 the Shishkin mesh is coarse but all layer components have declined sufficiently. With the L_∞- stability of the nodal interpolant we get

$$\|(u-S)^I\|_{L_\infty(\Omega_0)} \le C\|u-S\|_{L_\infty(\Omega_0)} \le CN^{-\lambda_0} \le CN^{-3}.$$

Hence, we obtain for the layer components of u and $k \le 2$ with an inverse estimate

$$\begin{aligned} |(u-S)-(u-S)^\star|_{k,\Omega_0} &\le |(u-S)-(u-S)^I|_{k,\Omega_0} \le |(u-S)|_{k,\Omega_0} + |(u-S)^I|_{k,\Omega_0} \\ &\le C\big(\varepsilon^{1/4-k/2}N^{-\lambda_0} + H^{-k}\|(u-S)^I\|_{0,\Omega_0}\big) \le C\big(\varepsilon^{1/4-k/2}N^{-3} + N^{k-3}\big). \end{aligned} \tag{3.97}$$

For the smooth solution component S we estimate

$$|S-S^\star|_{k,\Omega_0} = |S-S^I|_{k,\Omega_0} \le CH^{2-k}|S|_{2,\Omega_0} \le CN^{k-2} \quad \text{for } k=0,1 \tag{3.98a}$$

and

$$|S-S^\star|_{2,\Omega_0} = |S-S^I|_{2,\Omega_0} \le CH|S|_{3,\Omega_0} \le C\varepsilon^{-1/2}N^{-1}. \tag{3.98b}$$

Obviously these bounds can be improved to $|S-S^\star|_{k,\Omega_0} \le CN^{k-3}$ if $|S|_3 < C$.

In the remainder of the domain the elements of the Shishkin mesh are anisotropic. For the smooth part S we use Lemma 33, for instance in Ω_1:

$$\begin{aligned} \|S-\Pi^y S\|_{0,\Omega_1} &\le C \sum_{|\boldsymbol{\alpha}|=2} \big(\boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}}S\|_{0,\Omega_1} + \boldsymbol{h}_M^{\boldsymbol{\alpha}+(0,1)}\|\boldsymbol{D}^{\boldsymbol{\alpha}}S_y\|_{0,\Omega_1}\big) \\ &\le C(H^2 + \varepsilon^{-1/2}H^2h) \le CN^{-2}. \end{aligned} \tag{3.99}$$

If $|S|_{3,\Omega_1} < C$ we could improve the estimate to $\|S-\Pi^y S\|_{0,\Omega_1} \le CN^{-3}$ using (3.84a). In the other subdomains Ω_i for $i=2,3,4$ the smooth part is estimated similarly. For the layer term E_1 Lemma 33 yields

$$\begin{aligned} \|E_1-\Pi^y E_1\|_{0,\Omega_1} &\le C \sum_{|\boldsymbol{\alpha}|=2} \big(\boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}}E_1\|_{0,\Omega_1} + \boldsymbol{h}_M^{\boldsymbol{\alpha}+(0,1)}\|\boldsymbol{D}^{\boldsymbol{\alpha}+(0,1)}E_1\|_{0,\Omega_1}\big) \\ &\le H^2\|\boldsymbol{D}^{(2,0)}E_1\|_{0,\Omega_1} + Hh\|\boldsymbol{D}^{(1,1)}E_1\|_{0,\Omega_1} + h^2\|\boldsymbol{D}^{(0,2)}E_1\|_{0,\Omega_1} \\ &\quad + H^2h\|\boldsymbol{D}^{(2,1)}E_1\|_{0,\Omega_1} + Hh^2\|\boldsymbol{D}^{(1,2)}E_1\|_{0,\Omega_1} + h^3\|\boldsymbol{D}^{(0,3)}E_1\|_{0,\Omega_1} \\ &\le C\big(N^{-2}\varepsilon^{1/4} + \varepsilon^{1/2}N^{-2}\ln N\varepsilon^{-1/4} + \varepsilon N^{-2}(\ln N)^2\varepsilon^{-3/4} \\ &\quad + \varepsilon^{1/2}N^{-3}\ln N\varepsilon^{-1/4} + \varepsilon N^{-3}(\ln N)^2\varepsilon^{-3/4} + \varepsilon^{3/2}N^{-3}(\ln N)^3\varepsilon^{-5/4} \\ &\le C\varepsilon^{1/4}N^{-2}(\ln N)^2. \end{aligned} \tag{3.100}$$

If $\|\boldsymbol{D}^{(3,0)}E_1\|_{0,\Omega_1} \le C$ this bound can be improved to $\|E_1-\Pi^y E_1\|_{0,\Omega_1} \le CN^{-3}(\ln N)^3$ with (3.84a). With the same technique one can estimate the layer component E_i on Ω_i, $i=2,3,4$. The other layer components are small on Ω_1, for instance for the corner layer E_{12} it holds

$$\begin{aligned} \|E_{12}-\Pi^y E_{12}\|_{0,\Omega_1} &\le \|E_{12}\|_{0,\Omega_1} + (\operatorname{meas}\Omega_1)^{1/2}\|\Pi^y E_{12}\|_{L_\infty(\Omega_1)} \\ &\le C(\operatorname{meas}\Omega_1)^{1/2}\big(\|E_{12}\|_{L_\infty(\Omega_1)} + h\|\boldsymbol{D}^{(0,1)}E_{12}\|_{L_\infty(\Omega_1)}\big) \\ &\le C\varepsilon^{1/4}(\ln N)^{1/2}\big(N^{-\lambda_0} + \varepsilon^{1/2}N^{-1}\ln N\varepsilon^{-1/2}N^{-\lambda_0}\big) \\ &\le C\varepsilon^{1/4}N^{-\lambda_0}(\ln N)^{1/2}. \end{aligned} \tag{3.101}$$

Here we used the stability estimate (3.77). Proceed similarly for all layer components E_j on Ω_i with $j \in \{1,\dots,4,12,23,34,41\}$ and $i=1,\dots,4$ with $i \ne j$. Now collect (3.93) for $\boldsymbol{\alpha}=(0,0)$, (3.96), (3.97), (3.98) with $k=0$, (3.99), (3.100) and (3.101) to obtain (3.95a).

Next if we want to estimate $\varepsilon^{1/4}|u-u^\star|_1$ it remains to estimate the error on the anisotropic elements, for instance on Ω_1. There the smooth solution component can be bounded with (3.84a). Let $|\boldsymbol{\gamma}|=1$, then

$$\begin{aligned} \|\boldsymbol{D}^{\boldsymbol{\gamma}}(S-\Pi^y S)\|_{0,\Omega_1} &\le C \sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}}S\|_{0,\Omega_1} \le CN^{-2}\|S\|_{W_{3,\infty}(\Omega_1)}(\operatorname{meas}\Omega_1)^{1/2} \\ &\le CN^{-2}\varepsilon^{-1/2}\varepsilon^{1/4}(\ln N)^{1/2} \le C\varepsilon^{-1/4}N^{-2}(\ln N)^{1/2}. \end{aligned} \tag{3.102}$$

The other domains Ω_i, $i = 2, 3, 4$ are treated similarly. From (3.84a) we deduce for the boundary layer component E_1 that

$$\begin{aligned} \|\boldsymbol{D}^{(0,1)}(E_1 - \Pi^y E_1)\|_{0,\Omega_1} &\le C \sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}+(0,1)} E_1\|_{0,\Omega_1} \\ &\le C\big(H^2\|\boldsymbol{D}^{(2,1)}E_1\|_{0,\Omega_1} + Hh\|\boldsymbol{D}^{(1,2)}E_1\|_{0,\Omega_1} + h^2\|\boldsymbol{D}^{(0,3)}E_1\|_{0,\Omega_1}\big) \\ \le C\big(N^{-2}\varepsilon^{-1/4} + \varepsilon^{1/2}N^{-2}\ln N\varepsilon^{-3/4} &+ \varepsilon N^{-2}(\ln N)^2\varepsilon^{-5/4} \le C\varepsilon^{-1/4}N^{-2}(\ln N)^2. \end{aligned} \tag{3.103}$$

The derivative with respect to x is better behaved and the same bound holds true:

$$\begin{aligned} \|\boldsymbol{D}^{(1,0)}(E_1 - \Pi^y E_1)\|_{0,\Omega_1} &\le C \sum_{|\boldsymbol{\alpha}|=2} \boldsymbol{h}_M^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}+(1,0)} E_1\|_{0,\Omega_1} \\ &\le C\big(H^2\|\boldsymbol{D}^{(3,0)}E_1\|_{0,\Omega_1} + Hh\|\boldsymbol{D}^{(2,1)}E_1\|_{0,\Omega_1} + h^2\|\boldsymbol{D}^{(1,2)}E_1\|_{0,\Omega_1}\big) \\ \le C\big(N^{-2}\varepsilon^{-1/4} + \varepsilon^{1/2}N^{-2}\ln N\varepsilon^{-1/4} &+ \varepsilon N^{-2}(\ln N)^2\varepsilon^{-3/4} \le C\varepsilon^{-1/4}N^{-2}(\ln N)^2. \end{aligned} \tag{3.104}$$

Obviously this bound holds also on Ω_3 where the anisotropy of the elements is in the same direction compared to Ω_1. In Ω_2 (or Ω_4) we use inverse estimates and the stability of Π^x:

$$\begin{aligned} |E_1 - \Pi^x E_1|_{1,\Omega_2} &\le |E_1|_{1,\Omega_2} + Ch^{-1}\|\Pi^x E_1\|_{0,\Omega_2} \\ &\le C(\operatorname{meas}\Omega_2)^{1/2}\big(|E_1|_{W_{1,\infty}(\Omega_2)} + h^{-1}(\|E_1\|_{L_\infty(\Omega_2)} + h\|\boldsymbol{D}^{(1,0)}E_1\|_{L_\infty(\Omega_2)})\big) \\ &\le C\varepsilon^{1/4}(\ln N)^{1/2}\big(\varepsilon^{-1/2}N^{-\lambda_0} + \varepsilon^{-1/2}N(\ln N)^{-1}N^{-\lambda_0} + N^{-\lambda_0}\big) \le C\varepsilon^{-1/4}N^{-2}(\ln N)^{-1/2}. \end{aligned} \tag{3.105}$$

Clearly, this technique can also be applied to estimate E_i, $i = 2, 3, 4$. The corner layer components are bounded in exactly the same way. Consider for instance E_{12} on Ω_1:

$$\begin{aligned} |E_{12} - \Pi^y E_{12}|_{1,\Omega_1} &\le |E_{12}|_{1,\Omega_1} + Ch^{-1}\|\Pi^y E_{12}\|_{0,\Omega_1} \\ &\le C(\operatorname{meas}\Omega_1)^{1/2}\big(|E_{12}|_{W_{1,\infty}(\Omega_1)} + h^{-1}(\|E_{12}\|_{L_\infty(\Omega_1)} + h\|\boldsymbol{D}^{(0,1)}E_{12}\|_{L_\infty(\Omega_1)})\big) \\ &\le C\varepsilon^{1/4}(\ln N)^{1/2}\big(\varepsilon^{-1/2}N^{-\lambda_0} + \varepsilon^{-1/2}N(\ln N)^{-1}N^{-\lambda_0}\big) \le C\varepsilon^{-1/4}N^{-2}(\ln N)^{-1/2}. \end{aligned} \tag{3.106}$$

Collecting (3.93) for $|\boldsymbol{\alpha}| = 1$, (3.96), (3.97), (3.98) with $k = 1$, (3.102), (3.103), (3.104), (3.105) and (3.106) yields (3.95b).

Finally, we consider second order derivatives. Unfortunately $u^\star \notin H^2(\Omega)$. However, $u^\star \in H^2(T)$ for all $T \in \Omega^N$ and even $u^\star \in H^2(M)$ for all $M \in \mathcal{M}^N$. Hence, we introduce the abbreviation $\|v\|_{0,\mathcal{M}(V)} := (\sum_{M\in\mathcal{M}, M\subset V}\|v\|_{0,M}^2)^{1/2}$. Now let $|\boldsymbol{\gamma}| = 2$, then by (3.84a) and (3.84b) we find for instance in Ω_1 that

$$\begin{aligned} \|\boldsymbol{D}^{\boldsymbol{\gamma}}(S - \Pi^y S)\|_{0,\mathcal{M}(\Omega_1)} &\le C\Big(\sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}}S\|_{0,\Omega_1} + \sum_{|\boldsymbol{\alpha}|=2} \frac{\boldsymbol{h}_M^{\boldsymbol{\alpha}}}{H}\|\boldsymbol{D}^{\boldsymbol{\alpha}}\boldsymbol{D}^{(1,0)}S\|_{0,\Omega_1}\Big) \\ &\le CN^{-1}\|S\|_{W_{3,\infty}(\Omega_1)}(\operatorname{meas}\Omega_1)^{1/2} \\ &\le CN^{-1}\varepsilon^{-1/2}\varepsilon^{1/4}(\ln N)^{1/2} \le C\varepsilon^{-1/4}N^{-1}(\ln N)^{1/2}. \end{aligned} \tag{3.107}$$

Similar bounds hold on Ω_i for $i = 2, 3, 4$. In oder to obtain bounds for the layer components E_i on Ω_i $(i = 1, \ldots, 4)$ we use (3.84a) and (3.84b) more careful.

$$\begin{aligned} \|\boldsymbol{D}^{(0,2)}(E_1 - \Pi^y E_1)\|_{0,\mathcal{M}(\Omega_1)} &\le C \sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}+(0,2)}E_1\|_{0,\Omega_1} \\ &\le C\big(H\|\boldsymbol{D}^{(1,2)}E_1\|_{0,\Omega_1} + h\|\boldsymbol{D}^{(0,3)}E_1\|_{0,\Omega_1}\big) \\ &\le C\big(N^{-1}\varepsilon^{-3/4} + \varepsilon^{1/2}N^{-1}\ln N\varepsilon^{-5/4}\big) \le \varepsilon^{-3/4}N^{-1}\ln N, \end{aligned} \tag{3.108}$$

$$\begin{aligned} \|\boldsymbol{D}^{(1,1)}(E_1 - \Pi^y E_1)\|_{0,\mathcal{M}(\Omega_1)} &\le C \sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}+(1,1)}E_1\|_{0,\Omega_1} \\ &\le C\big(H\|\boldsymbol{D}^{(2,1)}E_1\|_{0,\Omega_1} + h\|\boldsymbol{D}^{(1,2)}E_1\|_{0,\Omega_1}\big) \\ &\le C\big(N^{-1}\varepsilon^{-1/4} + \varepsilon^{1/2}N^{-1}\ln N\varepsilon^{-3/4}\big) \le \varepsilon^{-1/4}N^{-1}\ln N, \end{aligned} \tag{3.109}$$

$$
\begin{aligned}
\|\boldsymbol{D}^{(2,0)}(E_1-\Pi^y E_1)\|_{0,\mathcal{M}(\Omega_1)} &\le C\Big(\sum_{|\boldsymbol{\alpha}|=1}\boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}+(2,0)}E_1\|_{0,\Omega_1}+\sum_{|\boldsymbol{\alpha}|=2}\frac{\boldsymbol{h}_M^{\boldsymbol{\alpha}}}{H}\|\boldsymbol{D}^{\boldsymbol{\alpha}}\boldsymbol{D}^{(1,0)}E_1\|_{0,\Omega_1}\Big)\\
&\le C\big(H\|\boldsymbol{D}^{(3,0)}E_1\|_{0,\Omega_1}+h\|\boldsymbol{D}^{(2,1)}E_1\|_{0,\Omega_1}+h^2H^{-1}\|\boldsymbol{D}^{(1,2)}E_1\|_{0,\Omega_1}\big)\\
&\le C\big(N^{-1}\varepsilon^{-1/4}+\varepsilon^{1/2}N^{-1}\ln N\varepsilon^{-1/4}+\varepsilon N^{-1}(\ln N)^2\varepsilon^{-3/4}\big) \qquad (3.110)\\
&\le C\big(\varepsilon^{-1/4}N^{-1}+\varepsilon^{1/4}N^{-1}(\ln N)^2\big).
\end{aligned}
$$

The same technique can be used to bound the error of E_i on the anisotropic part of the Shishkin mesh along the opposite edge. In Ω_2 (or Ω_4) inverse estimates and the stability of Π^x yield again:

$$
\begin{aligned}
|E_1-\Pi^x E_1|_{2,\mathcal{M}(\Omega_1)} &\le |E_1|_{2,\Omega_2}+Ch^{-2}\|\Pi^x E_1\|_{0,\Omega_2}\\
&\le C(\operatorname{meas}\Omega_2)^{1/2}\big(|E_1|_{W_{2,\infty}(\Omega_2)}+h^{-2}(\|E_1\|_{L_\infty(\Omega_2)}+h\|\boldsymbol{D}^{(1,0)}E_1\|_{L_\infty(\Omega_2)})\big)\\
&\le C\varepsilon^{1/4}(\ln N)^{1/2}\big(\varepsilon^{-1}N^{-\lambda_0}+\varepsilon^{-1}N^2(\ln N)^{-2}N^{-\lambda_0}+\varepsilon^{-1/2}N(\ln N)^{-1}N^{-\lambda_0}\big)\\
&\le C\varepsilon^{-3/4}N^{-1}(\ln N)^{-3/2}.
\end{aligned} \qquad (3.111)
$$

The corner layers are handled similarly, for instance E_{12} on Ω_1:

$$
\begin{aligned}
|E_{12}-\Pi^y E_{12}|_{2,\mathcal{M}(\Omega_1)} &\le |E_{12}|_{2,\Omega_1}+Ch^{-2}\|\Pi^y E_{12}\|_{0,\Omega_1}\\
&\le C(\operatorname{meas}\Omega_1)^{1/2}\big(|E_{12}|_{W_{2,\infty}(\Omega_1)}+h^{-2}(\|E_{12}\|_{L_\infty(\Omega_1)}+h\|\boldsymbol{D}^{(0,1)}E_{12}\|_{L_\infty(\Omega_1)})\big)\\
&\le C\varepsilon^{1/4}(\ln N)^{1/2}\big(\varepsilon^{-1}N^{-\lambda_0}+\varepsilon^{-1}N^2(\ln N)^{-2}N^{-\lambda_0}+\varepsilon^{-1/2}N(\ln N)^{-1}\varepsilon^{-1/2}N^{-\lambda_0}\big)\\
&\le C\varepsilon^{-3/4}N^{-2}(\ln N)^{-1/2}. \qquad (3.112)
\end{aligned}
$$

Collect (3.93) for $|\boldsymbol{\alpha}|=2$, (3.96), (3.97), (3.98) with $k=2$, (3.107), (3.108), (3.109), (3.110), (3.111) and (3.112) to obtain (3.95c). The other assertions of the Lemma follow easily. □

After quantifying the approximation properties of $u^\star$ we want to study certain traces of $u-u^\star$ along interior edges.

Since Ω^N is an admissible triangulation two elements $T_1,T_2\in\Omega^N$ define traces of a function $v\in H^1(T_1\cup T_2)\cap H^2(T_1)\cap H^2(T_2)$ along an interior edge e. We associate a unit normal vector n with each edge. If $e\subset\partial\Omega$ is an edge along the boundary we define n as the unit outer normal to $\partial\Omega$. In a similar manner there are two traces of the normal derivative $\frac{\partial v}{\partial n}\in L_2(e)$. Assuming n is oriented from T_1 to T_2 we obtain jumps $[\![\frac{\partial v}{\partial n}]\!]$ of these traces as follows:

$$
\Big[\!\!\Big[\frac{\partial v}{\partial n}\Big]\!\!\Big] := \frac{\partial v}{\partial n}\Big|_{T_1}-\frac{\partial v}{\partial n}\Big|_{T_2}\in L_2(e).
$$

Lemma 36. *Suppose $\varepsilon^{1/4}\le(\ln N)^{-2}$. Then there is a positive constant C such that*

$$
\|u-u^\star\|_{0,e}^2\le CN^{-5}\qquad \text{on a long edge } e\text{, i.e. of type I or II,} \qquad (3.113)
$$

$$
\sum_{e\in\mathcal{E}(III)}\|u-u^\star\|_{0,e}^2\le C\varepsilon^{-1/2}N^{-5}(\ln N)^2, \qquad (3.114)
$$

$$
\sum_{e\in\mathcal{E}(IV)}\|u-u^\star\|_{0,e}^2\le C\varepsilon^{1/2}N^{-5}(\ln N)^3. \qquad (3.115)
$$

Proof. Let $e\subset\overline{\Omega}\setminus\Omega_f$ denote a long type-I or type-II edge of a possibly anisotropic element. For instance, on a long edge $e\subset\Omega_1$ the interpolant $\Pi^y v$ of v is a quadratic polynomial which is uniquely described by its values in the endpoints and the midpoint of e. Hence, on long edges Π^y coincides with the 1D Lagrange interpolation and we find that

$$
\begin{aligned}
\|(S+E_1)-\Pi^y(S+E_1)\|_{0,e}^2 &\le \operatorname{meas}(e)\|(S+E_1)-\Pi^y(S+E_1)\|_{L_\infty(e)}^2\\
&\le CH\,H^{-4}\|(S+E_1)_{yy}\|_{L_\infty(e)}^2\le CN^{-5}.
\end{aligned} \qquad (3.116)
$$

Any other layer component $E:=u-S-E_1$ is estimated using a stability argument of the interpolation operator involved on a macro-element M that is adjacent to $e\subset M$:

$$
\begin{aligned}
\|E-\Pi^y E\|_{0,e}^2 &\le \operatorname{meas}(e)\|E-\Pi^y E\|_{L_\infty(e)}^2\le CH\big(\|E\|_{L_\infty(M)}^2+\|\Pi^y E\|_{L_\infty(M)}^2\big)\\
&\le CH\big(\|E\|_{L_\infty(M)}^2+h\|E_y\|_{L_\infty(M)}^2\big)\le CN^{-\lambda_0-1}\le CN^{-7}.
\end{aligned} \qquad (3.117)
$$

Similarly to (3.116), we estimate the smooth part S on any edge $e \subset \overline{\Omega}_0$ in the interior subdomain:

$$\|S - S^I\|_{0,e}^2 \le \operatorname{meas}(e)\|S - S^I\|_{L_\infty(e)}^2 \le CH\,H^{-4}|S|_{W_{2,\infty}(e)}^2 \le CN^{-5}. \tag{3.118}$$

Next we use that all the layer components $E := u - S$ have declined sufficiently. Let $T \subset \overline{\Omega_0}$ denote an element that has the edge e, then

$$\begin{aligned}\|E - E^I\|_{0,e}^2 &\le \operatorname{meas}(e)\|E - E^I\|_{L_\infty(e)}^2 \le CH\big(\|E\|_{L_\infty(T)}^2 + \|E^I\|_{L_\infty(T)}^2\big)\\ &\le CH\|E\|_{L_\infty(T)}^2 \le CN^{-\lambda_0-1} \le CN^{-7}.\end{aligned} \tag{3.119}$$

Collecting (3.116), (3.117), (3.118) and (3.119) gives (3.113).

Now we consider the short type-III edge e of an anisotropic element T for instance in Ω_1. We use the trace Lemma 37 and (3.84a):

$$\begin{aligned}\|S - \Pi^y S\|_{0,e}^2 &\le C\|S - \Pi^y S\|_{0,T}\|(S - \Pi^y S)_x\|_{0,T} + \frac{1}{H}\|S - \Pi^y S\|_{0,T}^2 \le CH^5|S|_{3,M}^2\\ &\le C\operatorname{meas}(M)H^5|S|_{W_{3,\infty}(M)}^2 \le C\varepsilon^{1/2}N^{-2}\ln N\,N^{-5}\varepsilon^{-1} \le C\varepsilon^{-1/2}N^{-7}\ln N\end{aligned} \tag{3.120}$$

Here M denotes the macro-element such that $T \subset M$. Similarly, we obtain for the layer E_1

$$\|E_1 - \Pi^y E_1\|_{0,e}^2 \le C\|E_1 - \Pi^y E_1\|_{0,T}\|(E_1 - \Pi^y E_1)_x\|_{0,T} + \frac{1}{H}\|E_1 - \Pi^y E_1\|_{0,T}^2.$$

Hence, a summation over all type-III edges gives with Young's inequality

$$\sum_{e\in\mathcal{E}(III)} \|E_1 - \Pi^y E_1\|_{0,e}^2 \le C\varepsilon^{-1/2}N^{-5}(\ln N)^2, \tag{3.121}$$

due to (3.104) and a similar estimate with (3.84a) and $\varepsilon^{1/4} \le (\ln N)^{-2}$ for $\|E_1 - \Pi^y E_1\|_{0,\Omega_1}$, namely

$$\begin{aligned}\|E_1 - \Pi^y E_1\|_{0,\Omega_1} &\le C\sum_{|\boldsymbol{\alpha}|=3} \boldsymbol{h}_M^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}}E_1\|_{0,\Omega_1} \le C\big(H^3\|\boldsymbol{D}^{(3,0)}E_1\|_{0,\Omega_1}\\ &\quad + H^2h\|\boldsymbol{D}^{(2,1)}E_1\|_{0,\Omega_1} + Hh^2\|\boldsymbol{D}^{(1,2)}E_1\|_{0,\Omega_1} + h^3\|\boldsymbol{D}^{(0,3)}E_1\|_{0,\Omega_1}\big)\\ &\le C\big(\varepsilon^{-1/4}N^{-3} + \varepsilon^{1/4}N^{-3}(\ln N)^3\big).\end{aligned}$$

The other layer components can be estimated like in (3.117). With (3.120) and (3.121) we arrive at (3.114).

For the short type-IV edges of Ω_f close to the corners of the domain we again use the a trace Lemma and (3.96) to obtain

$$\sum_{e\in\mathcal{E}(IV)} \|u - \tilde{\Pi}u\|_{0,e}^2 \le C\varepsilon^{1/2}N^{-5}(\ln N)^3,$$

which is (3.115). □

Lemma 37. *Let T be a rectangle with sides parallel to the coordinate axes and a width in x-direction of h_x. Let ∂T_y denote the union of the two edges parallel to the y-axis having length h_y. Denote by $v^I \in Q_2(T)$ the nodal interpolant of $v \in C(\bar{T})$. Then for $v \in H^3(T)$ it holds*

$$\big\|\big(v - v^I\big)_x\big\|_{0,\partial T_y} \le C\big(h_x^{3/2}\|v_{xxx}\|_{0,T} + \sqrt{h_x}h_y\|v_{xxy}\|_{0,T} + \frac{h_y^2}{\sqrt{h_x}}\|v_{xyy}\|_{0,T}\big). \tag{3.122}$$

Proof. Lemma 14 and Young's inequality yield

$$\big\|\big(v - v^I\big)_x\big\|_{0,\partial T_y}^2 \le C\big(\frac{1}{h_x}\big\|\big(v - v^I\big)_x\big\|_{0,T}^2 + h_x\big\|\big(v - v^I\big)_{xx}\big\|_{0,T}^2\big).$$

With the well known anisotropic nodal interpolation error estimates for $v \in H^3(T)$:

$$\begin{aligned}\big\|\big(v - v^I\big)_x\big\|_{0,T} &\le C\big(h_x^2\|v_{xxx}\|_{0,T} + h_xh_y\|v_{xxy}\|_{0,T} + h_y^2\|v_{xyy}\|_{0,T}\big),\\ \big\|\big(v - v^I\big)_{xx}\big\|_{0,T} &\le C\big(h_x\|v_{xxx}\|_{0,T} + h_y\|v_{xxy}\|_{0,T}\big),\end{aligned}$$

we complete the proof. □

Lemma 38. *Assume $|S|_3 \le C$ and $\varepsilon^{1/2} \le (\ln N)^{-2}$ then there is a positive constant C such that*

$$\sum_{e\in\mathcal{E}(I)} \left\| \left[\!\left[\frac{\partial(u-u^\star)}{\partial n} \right]\!\right] \right\|_{0,e}^2 \le CN^{-3}, \tag{3.123}$$

$$\sum_{e\in\mathcal{E}(III)} \left\| \left[\!\left[\frac{\partial(u-u^\star)}{\partial n} \right]\!\right] \right\|_{0,e}^2 \le C\varepsilon^{-1/2}N^{-3}(\ln N)^4. \tag{3.124}$$

Proof. Recall that by construction the normal derivative of $u^\star$ is continuous across type-II and type-IV edges, i.e. across long edges of anisotropic elements and within the subdomains close to the four corners of Ω. Let $e \subset \overline{\Omega_0}$ be a type-I edge. Since $u^\star$ is defined by nodal interpolation on the the two elements T_1 and T_2 that share the edge e we find with Lemma 37 that

$$\left\| \left[\!\left[\frac{\partial(S-S^I)}{\partial n} \right]\!\right] \right\|_{0,e} = \left\| \frac{\partial(S-S^I)}{\partial n}\Big|_{T_1} \right\|_{0,e} + \left\| \frac{\partial(S-S^I)}{\partial n}\Big|_{T_2} \right\|_{0,e} \le CH^{3/2}\big(|S|_{3,T_1} + |S|_{3,T_2}\big).$$

Hence,

$$\sum_{e\in\mathcal{E}^N_{int}(\Omega_0)} \left\| \left[\!\left[\frac{\partial(S-S^I)}{\partial n} \right]\!\right] \right\|_{0,e}^2 \le CH^3|S|^2_{3,\Omega_0} \le CN^{-3}. \tag{3.125}$$

Next we abbreviate $E = u - S$. In Ω_0 the layer components E are pointwise small and smooth, hence on a type-I edge $e \in \mathcal{E}(I)$ we use inverse estimates to obtain

$$\begin{aligned}\left\| \left[\!\left[\frac{\partial(E-E^I)}{\partial n} \right]\!\right] \right\|_{0,e} &= \left\| \left[\!\left[\frac{\partial(E^I)}{\partial n} \right]\!\right] \right\|_{0,e} = \left\| \frac{\partial(E^I)}{\partial n}\Big|_{T_1} \right\|_{0,e} + \left\| \frac{\partial(E^I)}{\partial n}\Big|_{T_2} \right\|_{0,e} \\ &\le CH^{-1/2}\left(\left\| \frac{\partial(E^I)}{\partial n} \right\|_{0,T_1} + \left\| \frac{\partial(E^I)}{\partial n} \right\|_{0,T_2} \right) \le CH^{-3/2}\big(\|E^I\|_{0,T_1} + \|E^I\|_{0,T_2}\big).\end{aligned}$$

A summation over all type-I edges then yields

$$\sum_{e\in\mathcal{E}(I)} \left\| \left[\!\left[\frac{\partial(E-E^I)}{\partial n} \right]\!\right] \right\|_{0,e}^2 \le CH^{-3}\|E^I\|^2_{0,\Omega_0} \le CH^{-3}\|E\|^2_{\infty,\Omega_0} \le CH^{-3}N^{-2\lambda_0} \le CN^{-3}. \tag{3.126}$$

Combining (3.125) and (3.126) we arrive at (3.123).

It remains to estimate the jump of the normal derivative across short edges of anisotropic elements which are of type III. Let $e = T_1 \cap T_2 \subset \overline{\Omega_1}$ denote such an edge. We shall first deal with the case that T_1 and T_2 are anisotropic elements. Again, we split u into smooth and layer components and estimate

$$\big\|[\![\boldsymbol{D}^{(1,0)}(S-\Pi^y S)]\!]\big\|_{0,e} \le \big\|\boldsymbol{D}^{(1,0)}(S-\Pi^y S)|_{T_1}\big\|_{0,e} + \big\|\boldsymbol{D}^{(1,0)}(S-\Pi^y S)|_{T_2}\big\|_{0,e}.$$

Lemma 14 gives for the smooth part

$$\begin{aligned}\big\|\boldsymbol{D}^{(1,0)}(S-\Pi^y S)|_T\big\|^2_{0,e} &\le C\Big(\big\|\boldsymbol{D}^{(1,0)}(S-\Pi^y S)\big\|_{0,T}\big\|\boldsymbol{D}^{(2,0)}(S-\Pi^y S)\big\|_{0,T} \\ &\qquad + \frac{1}{H}\big\|\boldsymbol{D}^{(1,0)}(S-\Pi^y S)\big\|^2_{0,T}\Big) \\ &\le C(H^2H + H^{-1}H^4)|S|^2_{3,T} \le CN^{-3}|S|^2_{3,T}.\end{aligned}$$

A summation of all type-III edges then yields

$$\sum_{e\in\mathcal{E}(III)} \left\| \left[\!\left[\frac{\partial}{\partial n}(S-S^\star) \right]\!\right] \right\|_{0,e}^2 \le CN^{-3}|S|^2_{3,\bigcup_{i=1}^4\Omega_i} \le CN^{-3}. \tag{3.127}$$

With the layer component E_1 we proceed in a similar manner

$$\|\boldsymbol{D}^{(1,0)}(E_1 - \Pi^y E_1)|_T\|_{0,e}^2 \le C\Big(\|\boldsymbol{D}^{(1,0)}(E_1 - \Pi^y E_1)\|_{0,T}\|\boldsymbol{D}^{(2,0)}(E_1 - \Pi^y E_1)\|_{0,T} + \frac{1}{H}\|\boldsymbol{D}^{(1,0)}(E_1 - \Pi^y E_1)\|_{0,T}^2\Big).$$

A summation gives with (3.104) and (3.110)

$$\sum_{e\in\mathcal{E}(III)} \Big\|\Big[\!\!\Big[\frac{\partial}{\partial n}(E_1 - E_1^\star)\Big]\!\!\Big]\Big\|_{0,e}^2 \le C\varepsilon^{-1/2}N^{-3}(\ln N)^4. \tag{3.128}$$

Any other layer component $E \neq E_1$ can handled similarly as in the interior subdomain Ω_0:

$$\|[\boldsymbol{D}^{(1,0)}(E - \Pi^y E)]\|_{0,e} \le CH^{-3/2}\big(\|\Pi^y E\|_{0,T_1} + \|\Pi^y E\|_{0,T_2}\big).$$

Hence,

$$\sum_{e\in\mathcal{E}(III)} \Big\|\Big[\!\!\Big[\frac{\partial}{\partial n}(E - E^\star)\Big]\!\!\Big]\Big\|_{0,e}^2 \le CH^{-3}\big(\|\Pi^y E\|_{0,\Omega_1\cup\Omega_3}^2 + \|\Pi^x E\|_{0,\Omega_2\cup\Omega_4}^2\big) \le C\varepsilon^{1/2}N^{-3}\ln N, \tag{3.129}$$

as shown in (3.101). In order to estimate the jump of the normal derivative of $u - u^\star$ across short interior edges of for instance Ω_1 it remains to estimate the jump of the x-derivative of the term

$$\sum_{\substack{i=N/2\\ j\in\{N/4,3N/4\}}}^{3N/2} \frac{\partial(u^I - u)}{\partial y}\Big|_{\Omega_0}(x_{i/2}, y_j)\,\ell_{i/2}(x)\psi_j(y)$$

across these edges. With Lemma 14 and (3.93) one easily sees that this term is better behaved than $[\boldsymbol{D}^{(1,0)}(u - \Pi^y u)]$.

Finally, we consider type-III edges that are shared by an anisotropic element and a small square shaped one in the subdomains close to the corners of Ω. The common edge is then a subset of $\partial\Omega_f \setminus \partial\Omega$. Let for instance $T_1 \in \overline{\Omega_1}$ and $T_2 \in \overline{\Omega_{12}}$ denote such elements. Then the normal derivative of $u^\star$ jumps across the common edge at $x = \lambda$. Since

$$\Big\|\Big[\!\!\Big[\frac{\partial(u - u^\star)}{\partial n}\Big]\!\!\Big]\Big\|_{0,e} = \Big\|\frac{\partial(u - u^\star)}{\partial n}\Big|_{T_1}\Big\|_{0,e} + \Big\|\frac{\partial(u - \tilde\Pi u)}{\partial n}\Big|_{T_2}\Big\|_{0,e},$$

we can estimate the first summand like before and it remains to estimate the second one. We start off with a trace inequality

$$\|\boldsymbol{D}^{(1,0)}(u - \tilde\Pi u)|_{T_2}\|_{0,e}^2 \le C\big(\frac{1}{h}|u - \tilde\Pi u|_{1,T_2}^2 + h|u - \tilde\Pi u|_{2,T}^2\big).$$

Hence, with (3.94):

$$\begin{aligned}\|\boldsymbol{D}^{(1,0)}(u - \tilde\Pi u)|_{\Omega_{12}}\|_{0,x=\lambda}^2 &\le C\big(\frac{1}{h}|u - \tilde\Pi u|_{1,\Omega_{12}}^2 + h|u - \tilde\Pi u|_{2,\Omega_{12}}^2\big)\\ &\le Ch^3|u|_{3,\Omega_{12}}^2 \le Ch^3\,\mathrm{meas}(\Omega_{12})|u|_{W_{3,\infty}(\Omega_{12})}^2\\ &\le C\varepsilon^{3/2}N^{-3}(\ln N)^3\varepsilon\ln N\varepsilon^{-3} = C\varepsilon^{-1/2}N^{-3}(\ln N)^4.\end{aligned} \tag{3.130}$$

Collecting (3.123), (3.127), (3.128), (3.129) and (3.130) we arrive at (3.124) and finish the proof. □

Remark 32. Under additional compatibility conditions on the right hand side f it should be possible to remove the dependency of the third-order derivatives of the smooth part S on ε in (3.87b), giving $\|S\|_3 \le C$. However, assuming $|S|_3 \le C$ is of course weaker than requiring that all third-order derivatives of u are pointwise bounded uniformly with respect to ε.

Remark 33. Let e denote a horizontal long edge of an anisotropic macro-element. The interpolation operator Π^y features a stability of the form

$$\|(\Pi^y v)_y\|_{\infty,e} \leq C\|v_y\|_{\infty,e}.$$

However, this seems to lead only to the estimate $\|(\Pi^y E_1)_y\|_{0,e}^2 = \mathcal{O}(\varepsilon^{-1})$ which is not good enough for our purposes. That is why we use a modification of $\tilde{\Pi}$ in the definition of $u^\star$ in order to match the normal derivatives on both sides of $\partial\Omega_0$.

4 Balanced norm results for 2D reaction-diffusion problems

In the present chapter we seek to obtain numerical approximations to the solution of the singularly perturbed linear elliptic boundary value problem

$$-\varepsilon\Delta u + cu = f \qquad \text{in } \Omega, \tag{4.1a}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{4.1b}$$

where Ω is a bounded two dimensional domain with Lipschitz-continuous boundary $\partial\Omega$, $0 < \varepsilon \ll 1$ is a small positive parameter and c is a smooth function that satisfies $0 < 2(c^\star)^2 \le c$.

A standard weak formulation associated with problem (4.1) reads: find $u \in V$, such that

$$a(u, v) \coloneqq \varepsilon(\nabla u, \nabla v) + (c\, u, v) = (f, v) \quad \forall v \in V. \tag{4.2}$$

If $f \in L_2(\Omega)$, then problem (4.2) has a unique solution $u \in V \coloneqq H_0^1(\Omega)$ which is characterized by the presence of exponential boundary layers of width $\mathcal{O}(\varepsilon^{1/2}\ln(1/\varepsilon))$ along the entire boundary $\partial\Omega$. Additionally, internal layers and corner singularities may be present.

We shall consider finite element methods for the approximation of u. It is thus natural to use L_2 based norms to measure their performances. For instance, the Galerkin finite element method in which V in (4.2) is replaced with a finite dimensional subspace $V^h \subset V$ is easily analyzed in the energy norm

$$\|v\|_\varepsilon \coloneqq \varepsilon^{1/2}|v|_1 + \|v\|_0 \quad \text{for all } v \in H_0^1(\Omega). \tag{4.3}$$

In fact, the bilinear form in (4.2) has some nice properties. In particular $a(\cdot,\cdot)$ is coercive in V (and V^h) with respect to the energy norm. Hence, the Galerkin approximation u^h is easily shown to be quasi-optimal in this norm:

$$\|u - u^h\|_\varepsilon \le C \inf_{v^h \in V^h} \|u - v^h\|_\varepsilon. \tag{4.4}$$

Finally, if u has sufficient regularity, the approximation error in (4.4) can be bounded uniformly with respect to ε by replacing v^h with some projection of u onto a spline space on a layer adapted mesh.

However, we have seen in Chapter 2 that if one aims for information of the solution within the layer not every norm yields a meaningful result in the case of singular perturbation. In fact, in the following paragraph we will explain that the energy norm is too weak to capture the boundary layers.

The behaviour of the solution u of (4.1) is well understood if f is smooth and nicely reflected in solution decompositions splitting u into a sum of several components. In this introduction we shall focus on only two of these. While for the smooth solution component S, lower order derivatives can be bounded pointwise uniformly with respect to ε, a typical boundary layer component E_1 behaves like the function $\exp(-\tilde{c}\,\mathrm{dist}(x, \partial\Omega)/\sqrt{\varepsilon})$. Here $\tilde{c}$ is a constant and $\mathrm{dist}(x, \partial\Omega)$ is the distance of the point $x \in \mathbb{R}^2$ to the boundary of Ω. Consequently, measuring the sum of these components in the energy norm gives

$$\|S + E_1\|_\varepsilon \le \varepsilon\big(|S|_1 + |E_1|_1\big) + \|S + E_1\|_0 = \mathcal{O}(\varepsilon^{1/4}) + \|S\|_0,$$

because the boundary layer function is of order $\mathcal{O}(\varepsilon^{1/4})$ in $\|\cdot\|_\varepsilon$. Hence, the predominant feature of u — the boundary layer — is neglected for $\varepsilon \to 0$, making the energy norm essentially no stronger than the L_2 norm. This failure of the energy norm has been known for some time but an appropriate treatment for this problem was unknown until recently.

Remark 34. For the convection-diffusion equation there is no such problem if only exponential boundary layers are present which are of the form $\exp(-\tilde{c}\,\mathrm{dist}(x,\partial\Omega)/\varepsilon)$. If however, the convection is aligned to certain edges of the domain and characteristic layers arise, then the corresponding solution components are again suppressed by a too small multiplier. Consequently, characteristic layers are not well represented in the energy norm. The interested reader is referred to [25]. There, the authors propose a differently weighted and better suited norm in which the error of the bilinear streamline diffusion finite element method is analyzed on a Shishkin mesh.

For instance, if the Galerkin finite element method is applied to (4.1) the situation is delicate: reducing the multiplier $\varepsilon^{1/2}$ of the H^1 semi-norm component of $\|\cdot\|_\varepsilon$ to $\varepsilon^{1/4}$ yields the so-called *balanced* norm $\|\cdot\|_b$ with

$$\|v\|_b := \varepsilon^{1/4}|v|_1 + \|v\|_0 \quad \text{for all } v \in H^1(\Omega), \tag{4.5}$$

whose components are correctly scaled by powers of ε. However, the standard error analysis of the Galerkin finite element method can no longer be applied due to the lack of coercivity with respect to this norm. We will see how to circumvent this problem in Section 4.3.

In Section 4.1 we want to sketch the ideas of [42]. This paper was the first to deal with the particular problem of designing a finite element method for which error estimates in a better suited norm could be proven. Also, it coined the expression *balanced norm.* We adapt the main idea in Section 4.2 to propose a new C^0 interior penalty method that features improved stability properties in comparison with the Galerkin Finite Element Method. At the end of this chapter we supply numerical experiments, give a brief summary and mention further work in this field of research.

4.1 The balanced finite element method of Lin and Stynes

It is very easy to show that the solution u of (4.1) satisfies the stability estimate

$$\|u\|_\varepsilon = \varepsilon^{1/2}|u|_1 + \|u\|_0 \le C\|f\|_0 \tag{4.6}$$

in the energy norm. In [42] the authors present a nice example which shows that such an estimate can not hold if the energy norm is replaced by $\|\cdot\|_b$. In fact if one only has $f \in L_2(\Omega)$ then the exponent of ε can not be reduced without breaking the stability property.

However, if $f \in C(\overline{\Omega}) \cap H^1(\Omega)$ and Ω is convex then it is possible to obtain a stability estimate for the solution $u \in H^2(\Omega)$ of (4.1) in a balanced norm (see [42]): By testing the differential equation (4.1a) with $-\Delta u$, integration by parts and using (4.1b), (4.6), Young's inequality as well as a maximum principle one gets

$$\varepsilon^{3/4}\|\Delta u\|_0 + \|u\|_b \le C\big(\|f\|_b + \|f\|_{L_\infty(\Omega)}\big). \tag{4.7}$$

Observe that for a typical boundary layer component E_1 it holds $\|\Delta E_1\|_0 = \mathcal{O}(\varepsilon^{-3/4})$. Hence, this norm is in fact balanced and stronger than the energy norm.

Following the derivation of (4.7) it is possible to provide a new variational formulation of problem (4.1). The idea of testing the differential equation (4.1a) with the Laplacian of a test function can be traced back to H^1-Galerkin methods [21] and has, according to Douglas, significant practical advantages over other standard approaches, particularly for non-linear problems [20, 19]. Yet this concept appears to be new in the singularly perturbed case. However, a direct multiplication of (4.1a) by $-\varepsilon\Delta v$ with $v \in H^2(\Omega)$ being a test function would require C^1-finite elements after a conforming discretization.

In [42] the authors circumvent this difficulty by a mixed finite element framework and we follow their argumentation: Rewriting (4.1a) as a first-order system yields

$$\begin{aligned} \boldsymbol{p} - \nabla u &= \boldsymbol{0}, \\ -\varepsilon\nabla\cdot\boldsymbol{p} + cu &= f. \end{aligned} \tag{4.8}$$

The first equation of (4.8) is rescaled by a multiplication with $\varepsilon^{1/4}$ resulting in the first-order system $A\boldsymbol{u} = \boldsymbol{f}$ in Ω with

$$\boldsymbol{u} = \begin{pmatrix} \boldsymbol{p} \\ u \end{pmatrix}, \qquad A\boldsymbol{u} = \begin{pmatrix} \varepsilon^{1/4}(\boldsymbol{p} - \nabla u) \\ -\varepsilon\nabla\cdot\boldsymbol{p} + cu \end{pmatrix} \quad \text{and} \quad \boldsymbol{f} = \begin{pmatrix} \boldsymbol{0} \\ f \end{pmatrix}, \tag{4.9}$$

which is equivalent to (4.1a). In order to supply a weak formulation the space $\boldsymbol{H}(\Omega)$ is introduced by

$$\boldsymbol{H}(\Omega) \coloneqq H(\mathrm{div};\Omega) \times H_0^1(\Omega),$$

where $H(\mathrm{div};\Omega)$ denotes the standard space of vector-valued functions which are in $L_2(\Omega)$ in both components and whose weak divergence is in $L_2(\Omega)$, as well. A weak formulation then reads: find $\boldsymbol{u} = (\boldsymbol{p}\ u)^T \in \boldsymbol{H}(\Omega)$, such that

$$B(\boldsymbol{u},\boldsymbol{v}) := (A\boldsymbol{u},\tilde{A}\boldsymbol{v}) = (\boldsymbol{f},\tilde{A}\boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in \boldsymbol{H}(\Omega), \tag{4.10}$$

where

$$\boldsymbol{v} = \begin{pmatrix} \boldsymbol{q} \\ v \end{pmatrix} \quad \text{and} \quad \tilde{A}\boldsymbol{v} = \begin{pmatrix} \varepsilon^{1/4}(\boldsymbol{q} - \nabla v) \\ \frac{1}{c}(-\varepsilon^{1/2}\nabla \cdot \boldsymbol{q} + cv) \end{pmatrix}. \tag{4.11}$$

Similarly to (4.9) one may think of q as corresponding to the gradient of v.

Lin and Stynes then introduce a balanced norm $\|\cdot\|_{b,2}$ for $\boldsymbol{v} \in \boldsymbol{H}(\Omega)$ of the form

$$\|\boldsymbol{v}\|_{b,2} = \left(\varepsilon^{3/2}\|\nabla \cdot \boldsymbol{q}\|_0^2 + \varepsilon^{1/2}(\|\boldsymbol{q}\|_0^2 + \|\nabla v\|_0^2) + \|v\|_0^2\right)^{1/2}. \tag{4.12}$$

Note that if $\boldsymbol{u} = (\nabla u\ u)^T$ is measured each component of this norm has indeed the same order of magnitude and that this norm can be viewed as a translation of the weighted norm on the left-hand side of (4.7) into the system context. With this norm at hand the analysis of a finite element method based on (4.10) follows the standard pattern sketched in the introduction.

Theorem 39. *The bilinear form $B(\cdot,\cdot)$ is coercive and bounded with respect to the balanced norm $\|\cdot\|_{b,2}$, i.e.*

$$C\|\boldsymbol{v}\|_{b,2}^2 \le B(\boldsymbol{v},\boldsymbol{v}) \quad \textit{for all } \boldsymbol{v} \in \boldsymbol{H}(\Omega), \tag{4.13}$$

$$|B(\boldsymbol{v},\boldsymbol{w})| \le C\|\boldsymbol{v}\|_{b,2}\|\boldsymbol{w}\|_{b,2} \quad \textit{for all } \boldsymbol{v},\boldsymbol{w} \in \boldsymbol{H}(\Omega). \tag{4.14}$$

The elegant proof in [42] uses only simple ingredients. Let $\boldsymbol{v},\boldsymbol{w} \in \boldsymbol{H}(\Omega)$ be arbitrary with $\boldsymbol{v} = (\boldsymbol{q}\ v)^T$ and $\boldsymbol{w} = (\boldsymbol{r}\ w)^T$. In order to prove (4.14) apply integration by parts to all the $L_2(\Omega)$ inner products of a divergence of $\boldsymbol{q}$ or $\boldsymbol{r}$ with v or w. To derive (4.13) additionally the binomial identities

$$\|\boldsymbol{q} \pm \nabla v\|_0^2 = \|\boldsymbol{q}\|_0^2 \pm 2(\boldsymbol{q},\nabla v) + \|\nabla v\|_0^2$$

are used.

Next (4.10) is discretized by replacing $\boldsymbol{H}(\Omega)$ by some finite element subspace $\boldsymbol{V}_h$. The solution $\boldsymbol{u}_h = (\boldsymbol{p}_h\ u_h)^T \in \boldsymbol{V}_h$ of the discrete problem

$$B(\boldsymbol{u}_h,\boldsymbol{v}_h) = (\boldsymbol{f},\tilde{A}\boldsymbol{v}_h) \quad \text{for all } \boldsymbol{v}_h \in \boldsymbol{V}_h \tag{4.15}$$

is unique due to the well-known lemma of Lax-Milgram. Moreover, it is quasi-optimal.

Theorem 40. *Let u and $\boldsymbol{u}_h$ be the solutions of (4.1) and (4.15), respectively and set $\boldsymbol{u} = (\nabla u\ u)^T$. Then there is a constant C, which is independent of ε, such that*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{b,2} \le C \inf_{\boldsymbol{v}_h \in \boldsymbol{V}_h} \|\boldsymbol{u} - \boldsymbol{v}_h\|_{b,2}.$$

Proof. Let $\boldsymbol{v}_h \in \boldsymbol{V}_h$ be arbitrary. By $A\boldsymbol{u} = \boldsymbol{f}$, the conformity of the discretization (i.e. $\boldsymbol{V}_h \subset \boldsymbol{H}(\Omega)$) and the definitions of $B(\cdot,\cdot)$ and $\boldsymbol{u}_h$ it is easy to establish the Galerkin orthogonality property

$$B(\boldsymbol{u} - \boldsymbol{u}_h,\boldsymbol{v}_h) = 0 \quad \text{for all } \boldsymbol{v}_h \in \boldsymbol{V}_h.$$

This and Theorem 39 are used in a standard argument to get

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{b,2}^2 \le CB(\boldsymbol{u} - \boldsymbol{u}_h,\boldsymbol{u} - \boldsymbol{u}_h) = CB(\boldsymbol{u} - \boldsymbol{u}_h,\boldsymbol{u} - \boldsymbol{v}_h) \le C\|\boldsymbol{u} - \boldsymbol{u}_h\|_{b,2}\|\boldsymbol{u} - \boldsymbol{v}_h\|_{b,2},$$

from which the result follows. ☐

By this Theorem error estimates can be obtained by studying the approximation error of a conforming finite element space $\boldsymbol{V}_h \subset \boldsymbol{H}(\Omega)$. Let $\{\mathcal{T}_h\}$ denote a family of regular triangulations of Ω into triangles or rectangles with a maximal element diameter h.

A sufficient condition for $H(\mathrm{div};\Omega)$-conformity is given by continuity of the normal derivative across inter-element boundaries, which can easily be seen using integration by parts. In fact, let $\boldsymbol{q}_h$ denote a vector-valued function such that each component is polynomial on every element of $\mathcal{T}_h$. Then

$$\begin{aligned}(\boldsymbol{q}_h,\nabla\varphi) &= \sum_{T\in\mathcal{T}_h}(\boldsymbol{q}_h,\nabla\varphi)_T = -\sum_{T\in\mathcal{T}_h}(\operatorname{div}\boldsymbol{q}_h,\varphi) + \sum_{T\in\mathcal{T}_h}(n\cdot\boldsymbol{q}_h,\varphi)_{\partial T}\\ &= -\sum_{T\in\mathcal{T}_h}(\operatorname{div}\boldsymbol{q}_h,\varphi) + \sum_{e\in\mathcal{E}_h}([\![n_e\cdot\boldsymbol{q}_h]\!],\varphi)_e \quad \text{for all } \varphi\in C_0^\infty(\Omega).\end{aligned}$$

Here $\mathcal{E}_h$ is the set of inner edges of $\{\mathcal{T}_h\}$, n is the unit outer normal to ∂T and $[\![n_e\cdot\boldsymbol{q}_h]\!]_e$ denotes the jump of $n_e\cdot\boldsymbol{q}_h$ across $e = T_1\cap T_2\in\mathcal{E}_h$, i.e.

$$[\![v]\!]_e = v|_{T_1\cap e} - v|_{T_2\cap e}$$

with n_e being the normal vector associated with the edge e pointing from T_1 to T_2.

For instance, the Raviart-Thomas elements (see e.g. [10]) of index $k\geq 0$ fulfill this condition and their corresponding finite element space $RT_{k,h}$ is therefore a subspace of $H(\mathrm{div};\Omega)$.

Remark 35. The continuity of the normal components is not necessary to yield $H(\mathrm{div};\Omega)$-conformity. For instance, in [9] a technique based on Lagrange multipliers is used to enforce the continuity weakly, relaxing this condition. Since the vector-valued finite element space is needed to approximate the gradient of the solution u this remark points into the direction that any method that follows the approach of Lin and Stynes will have to feature some control over the jumps of the normal derivative along inter-element edges, cf. the continuous interior penalty method of Section 4.2.

Let V_h^k denote the $H^1(\Omega)$-conforming finite element space generated by Q_k elements if the mesh cells of $\{\mathcal{T}_h\}$ are rectangles. Otherwise use P_k elements. Set $V_{h,0}^k := V_h^k\cap H_0^1(\Omega)$. In [42] the authors use well-known approximation error estimates [10] for $\boldsymbol{u}$ in the finite element space

$$\boldsymbol{V}_h := RT_{k,h}\times V_{h,0}^k \subset \boldsymbol{H}(\Omega),$$

to obtain the result

$$\|\boldsymbol{u}-\boldsymbol{u}_h\|_{b,2} \leq C\big(h^{k+1}|u|_{k+1} + \varepsilon^{1/4}h^k|u|_{k+1} + \varepsilon^{3/4}h^k|\Delta u|_k\big), \tag{4.16}$$

provided u has the regularity required. Moreover, they use a duality argument of Aubin-Nitsche type to derive a higher-order bound for the $L_2(\Omega)$ error $\|u-u_h\|_0$ under the additional assumption that the differential operator of (4.1) enjoys full elliptic regularity:

$$\|u-u_h\|_0 \leq Ch\varepsilon^{-1/4}\big(h^{k+1}|u|_{k+1} + \varepsilon^{1/4}h^k|u|_{k+1} + \varepsilon^{3/4}h^k|\Delta u|_k\big). \tag{4.17}$$

The estimates (4.16) and (4.17) are of classical type and not satisfactory in the case of singular perturbation because the right hand side blows up for $\varepsilon\to 0$. In fact, typically derivatives of the solution of (4.1) are bounded sharply by $|u|_s = \mathcal{O}(\varepsilon^{-s/2+1/4})$ for $s\geq 1$. Nevertheless these estimates show that the method performs well in the regime $h<\varepsilon^{1/2}$, i.e. when the mesh is very fine globally and there are mesh points inside the layer region. In this case (4.17) proves a higher rate of convergence in the $L_2(\Omega)$-norm.

In order to provide meaningful bounds for the error of the finite element method in the interesting regime $\varepsilon^{1/2}<h$ Theorem 40 shows that the approximation error is the key. If the approximation error is small by using a layer-adapted mesh so will be the error of the finite element method. This will enable one to include coarse elements into the mesh in parts of the domain avoiding the unrealistic assumption $h<\varepsilon^{1/2}$ (which would in general imply a gigantic number of degrees of freedom).

In Section 5 of [42] the authors analyze their finite element method on a Shishkin mesh of the unit square $\Omega=(0,1)^2$ in the regime $\varepsilon\leq CN^{-1}$. They assume that the data c and f of (4.1) lie in the Hölder space $C^{4,\alpha}(\overline{\Omega})$ for some $\alpha\in(0,1]$ and the corner compatibility conditions (3.86) which guarantee that third derivatives of u are bounded even in the corners of Ω. In this respect the authors remark the possibility to extend their analysis to higher order elements by assuming additional corner compatibility conditions. Without these the authors prove convergence for lowest-order Raviart-Thomas and bilinear finite elements, i.e. $\boldsymbol{V}_h := RT_{0,h} \times V_{h,0}^1$.

Next, Lin and Stynes describe the solution decomposition and a priori estimates of [45, Lemma 1.1 and Lemma 1.2] as stated in Lemma 34. Moreover, they construct a Shishkin mesh similar to the one introduced in Section 3.7 with $\lambda_0 := 2$ and estimate the approximation error.

Lemma 41. *Let u^I denote the piecewise bilinear nodal interpolant of u on a Shishkin mesh with N intervals in each coordinate direction described in Section 3.7. Then there is a constant C such that*

$$\|u - u^I\|_b \leq CN^{-1} \ln N$$

Proof. Again, the argument follows [42]. An estimate for the $L_2(\Omega)$ error can be found in [45, Lemma 2.3]:

$$\|u - u^I\|_0 \leq CN^{-2}.$$

In order to estimate $\varepsilon^{1/4}|u - u^I|_1$ the solution is split into $u = S + \sum_{i=1}^4 E_i + E_{12} + E_{23} + E_{34} + E_{41}$ according to Lemma 34. For the smooth component S one obtains by a well-known interpolation error estimate and (3.87b) that

$$\varepsilon^{1/4}|S - S^I|_1 \leq C\varepsilon^{1/4}N^{-1}|S|_2 \leq C\varepsilon^{1/4}N^{-1}. \tag{4.18}$$

Next the authors use the well-known anisotropic interpolation error estimate [2]

$$\|\boldsymbol{D}^{\boldsymbol{\gamma}}(v - v^I)\|_{0,T} \leq C \sum_{|\boldsymbol{\alpha}|=1} \boldsymbol{h}_T^{\boldsymbol{\alpha}} \|\boldsymbol{D}^{\boldsymbol{\alpha}+\boldsymbol{\gamma}} v\|_{0,T} \tag{4.19}$$

for a rectangular axis-parallel element $T \in \mathcal{T}_h$, a multi-index $\boldsymbol{\gamma}$ with $|\boldsymbol{\gamma}| = 1$ and $v \in H^2(T)$. Consider for instance the boundary layer E_1, by (4.19) and (3.87c),

$$\begin{aligned}\varepsilon^{1/4}\left\|\frac{\partial}{\partial y}(E_1 - E_1^I)\right\|_{0,\Omega_1\cup\Omega_3} &\leq C\varepsilon^{1/4}\left(N^{-1}\left\|\frac{\partial^2}{\partial x \partial y}E_1\right\|_{0,\Omega_1\cup\Omega_3} + \varepsilon^{1/2}N^{-1}\ln N\left\|\frac{\partial^2}{\partial y^2}E_1\right\|_{0,\Omega_1\cup\Omega_3}\right)\\ &\leq C\varepsilon^{1/4}\left(N^{-1}\varepsilon^{-1/4} + \varepsilon^{1/2}N^{-1}\ln N\varepsilon^{-3/4}\right) \leq CN^{-1}\ln N.\end{aligned} \tag{4.20}$$

Proceed similarly in Ω_h where the mesh is very fine. In the remainder of the domain, $\Omega_0 \cup \Omega_2 \cup \Omega_4$ the mesh is coarse in y-direction and the layer E_1 is pointwise small, i.e. $|E_1| \leq CN^{-2}$. An inverse estimate yields for instance in Ω_0

$$\begin{aligned}\varepsilon^{1/4}\left\|\frac{\partial}{\partial y}(E_1 - E_1^I)\right\|_{0,\Omega_0} &\leq \varepsilon^{1/4}\left(\left\|\frac{\partial}{\partial y}E_1\right\|_{0,\Omega_0} + \left\|\frac{\partial}{\partial y}E_1^I\right\|_{0,\Omega_0}\right)\\ &\leq C\varepsilon^{1/4}\left(\varepsilon^{-1/4}N^{-2} + N\|E_1^I\|_{0,\Omega_0}\right)\\ &\leq C(N^{-2} + \varepsilon^{1/4}N^{-1}).\end{aligned} \tag{4.21}$$

A derivative with respect to x is easier to handle. The estimate (4.19) and (3.87c) yield

$$\begin{aligned}\varepsilon^{1/4}\left\|\frac{\partial}{\partial x}(E_1 - E_1^I)\right\|_0 &\leq C\varepsilon^{1/4}N^{-1}\left(\left\|\frac{\partial^2}{\partial x^2}E_1\right\|_0 + \left\|\frac{\partial^2}{\partial x \partial y}E_1\right\|_0\right)\\ &\leq C\varepsilon^{1/4}N^{-1}(1 + \varepsilon^{-1/4}) \leq CN^{-1}.\end{aligned} \tag{4.22}$$

For the corner layer components we use the same ideas. Consider for instance E_{12}. In Ω_h one obtains

$$\begin{aligned}\varepsilon^{1/4}\left|E_{12} - E_{12}^I\right|_{1,\Omega_h} &\leq C\varepsilon^{1/4}h|E_{12}|_{2,\Omega_h} \leq C\varepsilon^{1/4}\varepsilon^{1/2}N^{-1}\ln N\varepsilon^{-3/4}\\ &\leq CN^{-1}\ln N.\end{aligned} \tag{4.23}$$

In the remainder of the domain this layer component is small again. In Ω_0 we proceed similar to the estimation of E_1 in Ω_0, while in $\Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \Omega_4$ we additionally use the smallness of the domain, for instance in Ω_1:

$$\begin{aligned}
\varepsilon^{1/4}|E_{12} - E_{12}^I|_{1,\Omega_1} &\leq \varepsilon^{1/4}\left(|E_{12}|_{1,\Omega_1} + \left|E_{12}^I\right|_{1,\Omega_1}\right) \\
&\leq C\varepsilon^{1/4}\left(\varepsilon^{-1/4}N^{-2} + \varepsilon^{-1/2}N(\ln N)^{-1}\|E_1^I\|_{0,\Omega_1}\right) \\
&\leq C\big(N^{-2} + \varepsilon^{-1/4}N(\ln N)^{-1}\operatorname{meas}(\Omega_1)^{1/2}\|E_1\|_{L_\infty(\Omega_1)}\big) \\
&\leq CN^{-1}(\ln N)^{-1/2}.
\end{aligned} \tag{4.24}$$

Collecting (4.18), (4.20), (4.21), (4.22), (4.23) and (4.24) we finish the proof. □

It remains to establish approximation error estimates for the vector-valued component in $\boldsymbol{V}_h$ for the gradient of the solution. The lowest-order Raviart-Thomas element induces a projection $\Pi_{0,h}\boldsymbol{q}$ of $\boldsymbol{q} \in \left(H^1(\Omega)\right)^2$ into the finite element space

$$RT_{0,h} = \{\boldsymbol{r} \in H(\operatorname{div};\Omega):\ (\boldsymbol{r}|_T)_i \in \operatorname{span}\{1, x_i\} \quad \text{for } i = 1,2 \text{ and } T \in \mathcal{T}_h\}.$$

More precisely the projection operator $\Pi_{0,h}$ is constructed locally on each rectangular element $T \in \mathcal{T}_h$ by

$$\int_e \left(\Pi_{0,h}\boldsymbol{q}|_T - \boldsymbol{q}\right) \cdot n_e \mathrm{d}s = 0$$

for every side e of T.

Lemma 42. *Let $\Pi_{0,h}\nabla u$ denote the projection of ∇u into the finite element space $RT_{0,h}$ on a Shishkin mesh with N intervals in each coordinate direction as described in Section 3.7. Then there is a constant C such that*

$$\varepsilon^{3/4}\|\nabla \cdot (\nabla u - \Pi_{0,h}\nabla u)\|_0 + \varepsilon^{1/4}\|\nabla u - \Pi_{0,h}\nabla u\|_0 \leq CN^{-1}\ln N.$$

Proof. We follow [42]. Based on the projection error estimates [1, Remark 4.1]

$$\|(\boldsymbol{q} - \Pi_{0,h}\boldsymbol{q})_i\|_{0,T} \leq C\sum_{|\boldsymbol{\alpha}|=1}\boldsymbol{h}_T^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}}\boldsymbol{q}_i\|_{0,T} \quad \text{for i=1,2,}$$

for $\boldsymbol{q} = (\boldsymbol{q}_1, \boldsymbol{q}_2) \in \left(H^1(T)\right)^2$, inverse inequalities in $RT_{0,h}$ and stability estimates for each component of $\Pi_{0,h}$ the bound of $\varepsilon^{1/4}\|\nabla u - \Pi_{0,h}\nabla u\|_0$ is obtained similarly to the proof of Lemma 41.

To bound the other term the "commuting diagram" property [10] is used, i.e.

$$\nabla \cdot (\Pi_{0,h}\boldsymbol{q}) = P_h(\nabla \cdot \boldsymbol{q}) \quad \text{for all } \boldsymbol{q} \in \left(H^1(\Omega)\right)^2, \tag{4.25}$$

where $P_h : L_2(K) \to P_0(K)$ is the local L_2-projection onto a constant for which

$$\|v - P_h v\|_{0,T} \leq C\sum_{|\boldsymbol{\alpha}|=1}\boldsymbol{h}_T^{\boldsymbol{\alpha}}\|\boldsymbol{D}^{\boldsymbol{\alpha}}v\|_{0,T} \tag{4.26}$$

for $v \in H^1(T)$ holds [40, Lemma 2.1 (i)]. Setting $\boldsymbol{q} = \nabla u$ in (4.25) one sees that

$$\varepsilon^{3/4}\|\nabla \cdot (\nabla u - \Pi_{0,h}\nabla u)\|_0 = \varepsilon^{3/4}\|\Delta u - P_h(\Delta u)\|_0. \tag{4.27}$$

The right hand side of (4.27) is bounded using the solution decomposition of Lemma 34 and (4.26). First, by (3.87b),

$$\varepsilon^{3/4}\|\Delta S - P_h(\Delta S)\|_0 \leq C\varepsilon^{3/4}N^{-1}|\Delta S|_1 \leq C\varepsilon^{1/4}N^{-1}. \tag{4.28}$$

For the boundary layer component E_1 the estimate (4.26) and (3.87c) yield

$$\begin{aligned}
\varepsilon^{3/4}\|\Delta E_1 - P_h(\Delta E_1)\|_{0,\Omega_1} &\leq C\varepsilon^{3/4}\left(N^{-1}\left\|\frac{\partial(\Delta E_1)}{\partial x}\right\|_{0,\Omega_1} + \varepsilon^{1/2}N^{-1}\ln N\left\|\frac{\partial(\Delta E_1)}{\partial y}\right\|_{0,\Omega_1}\right) \\
&\leq CN^{-1}\ln N.
\end{aligned} \tag{4.29}$$

Proceed similarly in Ω_3 and Ω_h where the mesh is very fine. In the remainder of the domain the layer component E_1 is small:

$$\varepsilon^{3/4}\|\Delta E_1\|_{0,\Omega\setminus(\Omega_1\cup\Omega_3\cup\Omega_h)} \le CN^{-2},$$

due to (3.87c) and $\lambda_0 = 2$. Next, let $T \subset \Omega\setminus(\Omega_1\cup\Omega_3\cup\Omega_h)$ denote a mesh rectangle with sides of length h_x and $h_y = CN^{-1}$ then

$$\left|P_h(\Delta E_1)|_T\right| = \frac{1}{\operatorname{meas}(T)}\left|\int_T \Delta E_1 \,\mathrm{d}x\mathrm{d}y\right| \le \frac{C}{\operatorname{meas}(T)} h_x \varepsilon^{-1/2}N^{-2},$$

where we used (3.87c) and the choice of λ again. Consequently,

$$\left\|P_h(\Delta E_1)\right\|_{0,T} = \left|P_h(\Delta E_1)|_T\right| \operatorname{meas}(T)^{1/2} \le C\varepsilon^{-1/2}N^{-2}.$$

Squaring and summing up over all remaining $\mathcal{O}(N^2)$ rectangles a triangle inequality yields

$$\varepsilon^{3/4}\|\Delta E_1 - P_h(\Delta E_1)\|_{0,\Omega\setminus(\Omega_1\cup\Omega_3\cup\Omega_h)} \le C(N^{-2} + \varepsilon^{1/4}N^{-1}). \tag{4.30}$$

Collecting (4.28), (4.29) and (4.30) gives

$$\varepsilon^{3/4}\|\Delta E_1 - P_h(\Delta E_1)\|_0 \le CN^{-1}\ln N.$$

For the corner layer E_{12} similar arguments give

$$\varepsilon^{3/4}\|\Delta E_{12} - P_h(\Delta E_{12})\|_0 \le \varepsilon^{1/4}N^{-1}\ln N. \tag{4.31}$$

In Ω_h the mesh is uniform and very fine. Hence, obtaining (4.31) is easy there. For $\Omega\setminus\Omega_h$ it is useful to apply the argument that led to (4.30) only on the $\mathcal{O}(N)$ elements that are adjacent to Ω_h while on the remainder of the domain the layer is very small. □

Lemma 43. *Let $\boldsymbol{u}^I := (\Pi_{0,h}\nabla u, u^I) \in \boldsymbol{V_h} = RT_{0,h}\times V^1_{h,0}$ denote the interpolant based on the projection into the lowest-order Raviart Thomas space and bilinear elements on a Shishkin mesh with N intervals in each coordinate direction. Then there is a constant C such that*

$$\|\boldsymbol{u} - \boldsymbol{u}^I\|_{b,2} \le CN^{-1}\ln N.$$

Proof. Combine Lemmas 41 and 42. □

Based on this result and the quasi-optimality of the method convergence follows.

Theorem 44. *Consider the solution u of (4.1) with $\Omega = (0,1)^2$ and smooth data, satisfying the corner compatibility conditions (3.86) and set $\boldsymbol{u} := (\nabla u, u)$. Let $\boldsymbol{V_h} = RT_{0,h}\times V^1_{h,0}$ denote the product of the $H(\operatorname{div};\Omega)$-conforming lowest-order Raviart-Thomas space and the space of piecewise bilinears on a rectangular Shishkin mesh with N intervals in each coordinate direction. Then the finite element solution $\boldsymbol{u}_h$ from (4.15) satisfies*

$$\|\boldsymbol{u} - \boldsymbol{u}_h\|_{b,2} \le CN^{-1}\ln N.$$

Proof. This main result of [42] follows from Theorem 40 and Lemma 43. □

4.2 A C^0 interior penalty method

Despite the beauty of the theory of the mixed method of [42] shown in the previous section, it has some minor drawbacks: It involves the usage of conforming Raviart-Thomas elements and introduces additional unknowns that bind a great number of degrees of freedom. Moreover, as a consequence of the correct weighing in (4.10) the symmetry of the problem is sacrificed. The method presented in this Section allows to use standard finite elements. Additionally, the method can be modified to conserve the symmetry of the problem (in the constant coefficient case no modification is needed to achieve this). Still the method yields robust error estimates in a balanced norm which is quite similar to $\|\cdot\|_{b,2}$ in (4.12).

The basic idea is to imitate the technique of Lin and Stynes in a broken Sobolev space and to supply sufficient control over the jumps of the normal derivative of the approximate solution along inter-element edges, cf. Remark 35. The presentation follows [61].

Let Ω^N denote an admissible triangulation of Ω into rectangles associated with some discretization parameter N and define the broken Sobolev space

$$H^{1,2}(\Omega^N) := \{v \in H^1(\Omega) : v|_T \in H^2(T) \ \ \forall T \in \Omega^N,\ v = 0 \text{ on } \partial\Omega\}. \tag{4.32}$$

Let $\mathcal{E}^N$ be the set of interior edges. Since Ω^N is an admissible triangulation two elements $T_1, T_2 \in \Omega^N$ define traces of a function $v \in H^{1,2}(\Omega^N)$ along an interior edge $e \in \mathcal{E}^N$. We associate a unit normal vector n with each edge. If $e \subset \partial\Omega$ is an edge along the boundary we define n as the unit outer normal to $\partial\Omega$. In a similar manner there are two traces of the normal derivative $\frac{\partial v}{\partial n} \in L_2(e)$. Assuming n is oriented from T_1 to T_2 we obtain jumps $[\![\frac{\partial v}{\partial n}]\!]$ of these traces as follows:

$$\left[\!\!\left[\frac{\partial v}{\partial n}\right]\!\!\right] := \left.\frac{\partial v}{\partial n}\right|_{T_1} - \left.\frac{\partial v}{\partial n}\right|_{T_2} \in L_2(e), \quad e \in \mathcal{E}^N.$$

Testing the partial differential equation (4.1a) with $-\varepsilon^{1/2}\Delta v$, integrating over each element T of the triangulation and summing up over all $T \in \Omega^N$ we obtain for $v \in H^{1,2}(\Omega^N)$ the variational equation

$$\varepsilon^{3/2} \sum_{T\in\Omega^N} (\Delta u, \Delta v)_T - \varepsilon^{1/2} \sum_{T\in\Omega^N} (c\,u, \Delta v)_T = -\varepsilon^{1/2} \sum_{T\in\Omega^N} (f, \Delta v)_T. \tag{4.33}$$

Next an application of Green's theorem to the reaction part gives

$$\begin{aligned} \varepsilon^{3/2} \sum_{T\in\Omega^N} (\Delta u, \Delta v)_T + \varepsilon^{1/2}(c\,\nabla u + \nabla c\, u, \nabla v) - \varepsilon^{1/2} \sum_{e\in\mathcal{E}^N} \left(c \left[\!\!\left[\frac{\partial v}{\partial n}\right]\!\!\right], u\right)_e \\ = -\varepsilon^{1/2} \sum_{T\in\Omega^N} (f, \Delta v)_T. \end{aligned} \tag{4.34}$$

Adding (4.2) and (4.34) we arrive at a new weak formulation for problem (4.1). It can be viewed as the scalar version of (4.10) without imposing continuity of the normal derivatives along inter-element edges. Find $u \in H^{1,2}(\Omega^N)$, such that

$$B^{\pm}(u, v) = L(v) \quad \forall v \in H^{1,2}(\Omega^N), \tag{4.35}$$

where the bilinear forms $B^{\pm} : H^{1,2}(\Omega^N) \times H^{1,2}(\Omega^N) \to \mathbb{R}$ and the linear functional $L : H^{1,2}(\Omega^N) \to \mathbb{R}$ are defined as

$$\begin{aligned} B^{\pm}(w, v) := \varepsilon^{3/2} \sum_{T\in\Omega^N} (\Delta w, \Delta v)_T + \varepsilon(\nabla w, \nabla v) + \varepsilon^{1/2}(c\,\nabla w + \nabla c\, w, \nabla v) + (c\,w, v) \\ + \sum_{e\in\mathcal{E}^N} \left(-\varepsilon^{1/2} \left(c\,w, \left[\!\!\left[\frac{\partial v}{\partial n}\right]\!\!\right]\right)_e \pm \varepsilon^{1/2} \left(c \left[\!\!\left[\frac{\partial w}{\partial n}\right]\!\!\right], v\right)_e + \left(\sigma_e \left[\!\!\left[\frac{\partial w}{\partial n}\right]\!\!\right], \left[\!\!\left[\frac{\partial v}{\partial n}\right]\!\!\right]\right)_e \right), \end{aligned} \tag{4.36}$$

$$L(v) := (f, v) - \varepsilon^{1/2} \sum_{T\in\Omega^N} (f, \Delta v)_T, \tag{4.37}$$

for $v, w \in H^{1,2}(\Omega^N)$ and $\sigma_e \geq 0$ for all $e \in \mathcal{E}^N$. Note that for $w \in H^2(\Omega)$ the last two terms of the right hand side of (4.36) vanish since $[\![\partial w/\partial n]\!] \equiv 0$ on $e \in \mathcal{E}^N$ by a well-known Sobolev embedding. The introduction of these artificial terms is motivated by symmetrization and coercivity of the bilinear form. After the discretization the last term penalizes jumps of the normal derivative across interior edges. It plays an important role with respect to the stability of the method and renders it a continuous interior penalty method.

Remark 36. In the case of a constant coefficient c the bilinear form B^- is symmetric whereas B^+ is asymmetric. If the conservation of the symmetry of the problem is desired in the variable coefficient case, it might be advantageous to use a piecewise constant approximation c_N instead of c. By doing so the term $(\nabla c\,w, \nabla v)$ in (4.36) vanishes and c_N moves inside the jump brackets.

In order to carry out the error analysis of the method (4.35) we will introduce the following adequate norm:

$$|||v|||^2 = \varepsilon^{3/2} \sum_{T\in\Omega^N} \|\Delta v\|_{0,T}^2 + \varepsilon^{1/2}|v|_1^2 + \|v\|_0^2 + \sum_{e\in\mathcal{E}^N} \Big(\sigma_e \Big[\!\Big[\frac{\partial v}{\partial n}\Big]\!\Big], \Big[\!\Big[\frac{\partial v}{\partial n}\Big]\!\Big]\Big)_e. \tag{4.38}$$

Note that this norm corresponds to the left hand side of (4.7) in the context of our broken Sobolev space. Consequently, similarly to the norm (4.12) the norm $|||\cdot|||$ is balanced and stronger than the energy norm or $\|\cdot\|_b$ from (4.5).

For our discretization we introduce the FE space

$$V^N := \{v \in H^{1,2}(\Omega^N) : v|_T \in Q_2(T) \;\; \forall T \in \Omega^N\}. \tag{4.39}$$

The use of bilinear elements would result in $\Delta v^N = 0$ and (4.33) would not evolve the method. It should be possible to extend the results to Q_k elements with $k > 2$. However, this requires even more compatibility conditions of the data.

Lemma 45 (Coercivity). *Assume that the following mild condition holds:*

$$c - \frac{\varepsilon^{1/2}}{2}\Delta c \geq c_0 > 0. \tag{4.40}$$

Then there exists a positive constant C such that

$$B^+(v,v) \geq C|||v|||^2 \quad \forall v \in H^{1,2}(\Omega^N). \tag{4.41}$$

Additionally, let the penalty parameter satisfy

$$\sigma_e \geq C_\star \frac{\varepsilon}{h_e^\perp} \quad \text{for all } e \in \mathcal{E}^N, \tag{4.42}$$

where $C_\star$ is a constant and $h_e^\perp$ denotes the minimal length of all edges orthogonal to e. Then the symmetric bilinear form B^- is coercive in the discrete space V^N, i.e.

$$B^-(v^N, v^N) \geq C|||v^N|||^2 \quad \forall v^N \in V^N. \tag{4.43}$$

Proof. A straight-forward calculation yields the validity of (4.41) with $C = \min\{1, c_0\}$. Here we used the identity $(\nabla cv, \nabla v) = (\Delta c, v^2)/2$ which follows from integration by parts.

To prove (4.43) we start off with $v^N \in V^N$ and the identity

$$B^-(v^N, v^N) = B^+(v^N, v^N) - 2\varepsilon^{1/2} \sum_{e\in\mathcal{E}^N} \Big(c\, v^N, \Big[\!\Big[\frac{\partial v^N}{\partial n}\Big]\!\Big]\Big)_e. \tag{4.44}$$

We estimate the modulus of the last summand of (4.44). The triangle and the Cauchy-Schwarz inequality give

$$\Big|2\varepsilon^{1/2} \sum_{e\in\mathcal{E}^N} \Big(c\, v^N, \Big[\!\Big[\frac{\partial v^N}{\partial n}\Big]\!\Big]\Big)_e\Big| \leq 2\varepsilon^{1/2}\|c\|_{L_\infty(\Omega)} \sum_{e\in\mathcal{E}^N} \|v^N\|_{0,e} \Big\|\Big[\!\Big[\frac{\partial v^N}{\partial n}\Big]\!\Big]\Big\|_{0,e}.$$

Next the Young inequality yields

$$\Big|2\varepsilon^{1/2} \sum_{e\in\mathcal{E}^N} \Big(c\, v^N, \Big[\!\Big[\frac{\partial v^N}{\partial n}\Big]\!\Big]\Big)_e\Big| \leq \varepsilon^{1/2}\|c\|_{L_\infty(\Omega)} \sum_{e\in\mathcal{E}^N} \Big(\mu_e \|v^N\|_{0,e}^2 + \frac{1}{\mu_e}\Big\|\Big[\!\Big[\frac{\partial v^N}{\partial n}\Big]\!\Big]\Big\|_{0,e}^2\Big), \tag{4.45}$$

where $\mu_e > 0$ are positive constants on each edge $e \in \mathcal{E}^N$ that will be specified soon. The local inverse estimate

$$\|v^N\|_{0,e}^2 \leq \frac{75}{h_e^\perp}\|v^N\|_{0,T(e)}^2 \tag{4.46}$$

is well-known. It holds for biquadratic elements on rectangles $T(e)$ with edge e, see e.g. [11].

We want to estimate the first summand of the right hand side of (4.45) against $\frac{c_0}{2}\|v^N\|_0^2$. If we substitute (4.46) into (4.45) we achieve this by determining μ_e in such a way that the coefficient of $\|v^N\|^2_{0,T(e)}$ is bounded by $c_0/8$. Note that each element can only be referred to by at most four edges. Consequently, we set

$$\mu_e := \frac{1}{600}\frac{c_0}{\|c\|_{L_\infty(\Omega)}}\frac{h_e^\perp}{\varepsilon^{1/2}}.$$

With this choice the coefficient of the last summand of the right hand side of (4.45) amounts to

$$\frac{\varepsilon^{1/2}\|c\|_{L_\infty(\Omega)}}{\mu_e} = 600\,\frac{\|c\|^2_{L_\infty(\Omega)}}{c_0}\frac{\varepsilon}{h_e^\perp}.$$

Hence, if we set σ_e according to (4.42) with $C_\star := 1200\frac{\|c\|^2_{L_\infty(\Omega)}}{c_0}$ we obtain

$$\left|2\varepsilon^{1/2}\sum_{e\in\mathcal{E}^N}\left(c\,v^N, \left[\!\left[\frac{\partial v^N}{\partial n}\right]\!\right]\right)_e\right| \le \frac{c_0}{2}\|v^N\|_0^2 + \frac{1}{2}\sum_{e\in\mathcal{E}^N}\left(\sigma_e\left[\!\left[\frac{\partial v^N}{\partial n}\right]\!\right], \left[\!\left[\frac{\partial v^N}{\partial n}\right]\!\right]\right)_e.$$

From this together with (4.44) the assertion (4.43) follows. □

The asymmetric continuous interior penalty method proposed now reads: find $u^{N,+} \in V^N$ such that

$$B^+(u^{N,+}, v^N) = L(v^N) \quad \forall v^N \in V^N. \tag{4.47}$$

If the bilinear form in (4.15) is replaced with $B^-(\cdot,\cdot)$ we denote the obtained approximate solution by $u^{N,-}$. Similarly, one might obtain an approximate solution $u^{N,0}$ by using the bilinear form that is created by discarding the first artificial term in (4.36). We shall refer to this method as *incomplete continuous interior penalty method*. By the coercivity property of Lemma 45 the function $u^{N,+}$ is well defined. Similarly, $u^{N,-}$ and $u^{N,0}$ are well defined if σ_e is chosen appropriately. Subsequently, we shall focus on the asymmetric method (4.47) and specify variations of the analysis or in the results of the other methods. Note that $u^{N,+}$ is well defined due to the coercivity property (4.41).

Since our method (4.15) is consistent and $V^N \subset H^{1,2}(\Omega^N)$ the following identity known as Galerkin orthogonality holds:

$$B^+(u - u^N, v^N) = 0 \quad \forall v^N \in V^N. \tag{4.48}$$

In every finite element analysis the approximation error comes into play, eventually. In Section 3.7 we considered this problem for a Shishkin mesh of the unit square $\Omega = (0,1)^2$.

Under the assumptions that the smooth part S of the solution decomposition (3.87) satisfies

$$|S|_3 \le C, \tag{4.49}$$

cf. Remark 32, that the corner compatibility conditions (3.86) as well as the mild condition

$$\varepsilon^{1/2} \le (\ln N)^{-2} \tag{4.50}$$

hold true we proved the following:

There is a projection $u^\star \in V^N$ of u by a quasi-interpolation operator such that

$$\|u - u^\star\|_0 \le CN^{-2}, \quad \text{(3.95d)} \qquad \sum_{e\in\mathcal{E}(I)\cup\mathcal{E}(II)} \|u - u^\star\|^2_{0,e} \le CN^{-3}, \quad \text{(3.113)}$$

$$\varepsilon^{1/4}|u - u^\star|_1 \le CN^{-2}(\ln N)^2, \quad \text{(3.95f)} \qquad \sum_{e\in\mathcal{E}(III)} \|u - u^\star\|^2_{0,e} \le C\varepsilon^{-1/2}N^{-5}(\ln N)^2, \quad \text{(3.114)}$$

$$\varepsilon^{3/2}\sum_{T\in\Omega^N} |u - u^\star|^2_{2,T} \le CN^{-2}(\ln N)^2, \quad \text{(3.95e)} \qquad \sum_{e\in\mathcal{E}(IV)} \|u - u^\star\|^2_{0,e} \le C\varepsilon^{1/2}N^{-5}(\ln N)^3, \quad \text{(3.115)}$$

$$\sum_{e\in\mathcal{E}(I)} \left\| \left[\!\!\left[\frac{\partial(u-u^\star)}{\partial n} \right]\!\!\right] \right\|_{0,e}^2 \le CN^{-3}, \tag{3.123}$$

$$\sum_{e\in\mathcal{E}(III)} \left\| \left[\!\!\left[\frac{\partial(u-u^\star)}{\partial n} \right]\!\!\right] \right\|_{0,e}^2 \le C\varepsilon^{-1/2}N^{-3}(\ln N)^4. \tag{3.124}$$

Here we summarized the results of the Lemmas 35, 36 and 38 from Chapter 3. For the definition of the type of an edge and $u^\star$ we refer to Definition 6 and the beginning of Section 3.7, respectively.

Remark 37. The need to introduce a new projection arises since for our reformulation of the problem the standard nodal interpolant $u^I \in V^N$ to u appears to be incapable to yield balanced error estimates that are uniform with respect to ε. The problem is caused by the jump of the normal derivative of $u - u^I$ on long edges of anisotropic elements of Ω^N.

Lemma 46. *Suppose that* (3.86), (4.49) *and* (4.50) *holds true. Then* $|||u-u^\star||| \le CN^{-1}\ln N$, *if the parameter* $\sigma_e \ge 0$ *satisfies*

$$\sigma_e \le C \begin{cases} N(\ln N)^2 & \text{if } e \text{ is of type } I, \\ \varepsilon^{1/2}N(\ln N)^{-2} & \text{if } e \text{ is of type } III. \end{cases} \tag{4.51}$$

Proof. Combine (3.95d), (3.95f) and (3.95e) with (3.123) and (3.124). $\square$

Note that the approximation error is not adversely affected regardless of the choice of σ_e on type-II or type-IV edges.

Theorem 47. *Consider* (4.1) *and assume that the corner compatibility conditions* (3.86) *as well as the Assumptions* (4.49), (4.50) *and* (4.40) *hold true. Moreover, let the penalty parameter* σ_e *be chosen by*

$$\sigma_e := \sigma_0 \begin{cases} \varepsilon & \text{if } e \text{ is of type } I \text{ or type } II, \\ \varepsilon^{1/2} & \text{if } e \text{ is of type } III, \\ \varepsilon^{3/2} & \text{if } e \text{ is of type } IV, \end{cases}$$

with a sufficiently large constant $\sigma_0 > 0$. *Then the CIP-method approximation* u^N, *i.e. the piecewise biquadratic FE-function on the rectangular Shishkin mesh defined in Section 3.7 with* N *intervals in each coordinate direction satisfies the robust error estimate*

$$|||u-u^N||| \le CN^{-1}(\ln N)^{3/2} \tag{4.52}$$

in the balanced norm $|||\cdot|||$.

Proof. Splitting the error into an approximation error $\eta := u - u^\star$ and a discrete component $\xi := u^\star - u^N \in V^N$ we obtain by Lemma 46:

$$|||u-u^N||| \le |||\eta||| + |||\xi||| \le CN^{-1}\ln N + |||\xi|||.$$

It remains to establish a bound for $|||\xi|||$.

A standard argument using coercivity of B^+ and (4.48) gives:

$$|||\xi|||^2 \le -CB^+(\eta,\xi) \le C|B^+(\eta,\xi)|.$$

After a triangle inequality we apply the Cauchy-Schwarz inequality to all symmetric terms of $B^+(\eta,\xi)$ and obtain with Lemma 46:

$$|B^+(\eta,\xi)| \le CN^{-1}\ln N|||\xi||| + \left|\varepsilon^{1/2}(\nabla c\,\eta,\nabla\xi)\right| + \varepsilon^{1/2}\sum_{e\in\mathcal{E}^N}\left(\left| \left(c\,\eta, \left[\!\!\left[\frac{\partial\xi}{\partial n}\right]\!\!\right]\right)_e \right| + \left| \left(c\left[\!\!\left[\frac{\partial\eta}{\partial n}\right]\!\!\right],\xi\right)_e \right| \right). \tag{4.53}$$

An inverse estimate and (3.95d) give

$$\begin{aligned} \left|\varepsilon^{1/2}(\nabla c\,\eta,\nabla\xi)\right| &\le \|\nabla c\|_{L_\infty(\Omega)}\varepsilon^{1/2}\|\eta\|_0\|\nabla\xi\|_0 \le C\frac{\varepsilon^{1/2}}{\varepsilon^{1/2}N^{-1}\ln N}\|\eta\|_0\|\xi\|_0 \\ &\le CN^{-1}(\ln N)^{-1}|||\xi|||. \end{aligned} \tag{4.54}$$

The last term is estimated against $\|\xi\|_0 \le |||\xi|||$ as follows:

$$\varepsilon^{1/2}\sum_{e\in\mathcal{E}^N}\left|\left(c\left[\!\left[\frac{\partial\eta}{\partial n}\right]\!\right],\xi\right)_e\right| \le \varepsilon^{1/2}\,\|c\|_{L_\infty(\Omega)}\sum_{e\in\mathcal{E}^N}\left\|\left[\!\left[\frac{\partial\eta}{\partial n}\right]\!\right]\right\|_{0,e}\|\xi\|_{0,e}.$$

Since every inner edge $e\in\mathcal{E}^N$ belongs to two rectangular elements $T_1(e), T_2(e)\in\Omega^N$ with sides perpendicular to e of length $h_{1,\perp}, h_{2,\perp}$ an inverse estimate yields

$$\|\xi\|_{0,e} \le C\min\{h_{1,\perp}, h_{2,\perp}\}^{-1/2}\|\xi\|_{0,T_1(e)\cup T_2(e)}.$$

Across type-II and type-IV edges the jump of the normal derivative of η vanishes. Thus, the Cauchy-Schwarz inequality gives with (3.123) and (3.124):

$$\begin{aligned}\varepsilon^{1/2}\sum_{e\in\mathcal{E}^N}\left|\left(c\left[\!\left[\frac{\partial\eta}{\partial n}\right]\!\right],\xi\right)_e\right| &\le \varepsilon^{1/2}\,\|c\|_{L_\infty(\Omega)}\sum_{e\in\mathcal{E}(I)\cup\mathcal{E}(III)}\left\|\left[\!\left[\frac{\partial\eta}{\partial n}\right]\!\right]\right\|_{0,e}\|\xi\|_{0,e}\\ &\le C\varepsilon^{1/2}h^{-1/2}\Bigg(\sum_{e\in\mathcal{E}(I)\cup\mathcal{E}(III)}\left\|\left[\!\left[\frac{\partial\eta}{\partial n}\right]\!\right]\right\|_{0,e}^2\Bigg)^{1/2}\|\xi\|_0 \qquad (4.55)\\ &\le C\varepsilon^{1/2}\varepsilon^{-1/4}N^{1/2}(\ln N)^{-1/2}\varepsilon^{-1/4}N^{-3/2}(\ln N)^2\|\xi\|_0\\ &\le CN^{-1}(\ln N)^{3/2}|||\xi|||.\end{aligned}$$

Finally, we consider

$$\begin{aligned}\varepsilon^{1/2}\sum_{e\in\mathcal{E}^N}\left|\left(c\,\eta,\left[\!\left[\frac{\partial\xi}{\partial n}\right]\!\right]\right)_e\right| &\le \varepsilon^{1/2}\,\|c\|_{L_\infty(\Omega)}\sum_{e\in\mathcal{E}^N}\|\eta\|_{0,e}\left\|\left[\!\left[\frac{\partial\xi}{\partial n}\right]\!\right]\right\|_{0,e}\\ &\le C\varepsilon^{1/2}\left(\sum_{e\in\mathcal{E}^N}\frac{1}{\sigma_e}\|\eta\|_{0,e}^2\right)^{1/2}\left(\sum_{e\in\mathcal{E}^N}\sigma_e\left\|\left[\!\left[\frac{\partial\xi}{\partial n}\right]\!\right]\right\|_{0,e}^2\right)^{1/2} \qquad (4.56)\\ &\le C\left(\sum_{e\in\mathcal{E}^N}\frac{\varepsilon}{\sigma_e}\|\eta\|_{0,e}^2\right)^{1/2}|||\xi||| \le CN^{-3/2}|||\xi|||.\end{aligned}$$

Here the last inequality is due to our choice for σ_e, (3.113), (3.114) and (3.115). The result follows from (4.53) with (4.54), (4.55) and (4.56). □

Remark 38. Inspecting the proof of Theorem 47 one sees that a similar estimate holds for the CIP-method based on the bilinear form B^- (cf. Lemma 45) if the penalty parameter is chosen according to

$$\sigma_e := \sigma_0\begin{cases}\varepsilon N & \text{if } e \text{ is of type I},\\ \varepsilon^{1/2}N(\ln N)^{-1} & \text{else.}\end{cases}$$

With the same choice for the parameter σ_e the incomplete CIP-method (in which the non-symmetric artificial term of (4.36) is removed) gives an optimal approximation, i.e. $|||u-u^N||| \le CN^{-1}\ln N$.

Remark 39. The estimate (4.55) is crude on most edges: The edge perpendicular to a type-I edge is always long. Hence, one might use $H^{-1/2}$ instead of $h^{-1/2}$ in this case. In fact, only a type-III edge $e\in\overline{\Omega_f}$ is estimated sharply. This observation fits perfectly to our numerical experiments in which we observed oscillations in the fine mesh part over Ω_f. These are damped efficiently if we simplify our choice for the penalty parameter to

$$\sigma_e := \sigma_0\begin{cases}\varepsilon & \text{if } e \text{ is a long (type-I or type-II) edge},\\ \varepsilon^{1/2} & \text{if } e \text{ is a short (type-III or type-IV) edge.}\end{cases}$$

Note that even without this adjustment we numerically observe first order convergence for the CIP-method in agreement with Theorem 47. However, the initial rates are slightly smaller then which accounts for a larger error.

4.3 Galerkin finite element method

In this section we shall examine the performance of the Galerkin finite element method

$$\varepsilon(\nabla u^N, \nabla v^N) + (c\, u^N, v^N) = (f, v^N) \quad \forall v^N \in V^N$$

for problem (4.1) on the unit square $\Omega = (0,1)^2$. Here $V^N \subset H_0^1(\Omega)$ denotes the space of bilinear finite elements on the Shishkin mesh Ω^N described in Section 3.7 with $\lambda_0 = 2$ in (3.88). Alternatively we may consider linear finite elements over a triangulation $\tilde{\Omega}^N$ obtained form Ω^N by drawing diagonals in each mesh rectangle. The presentation follows [61].

As already mentioned in the introduction the analysis of the Galerkin FEM is fairly easy in the energy norm $\|\cdot\|_\varepsilon$ defined in (4.3). Assuming that f satisfies the corner compatibility conditions (3.86) then for the Lagrange interpolant $u^I \in V^N$ of u on Shishkin meshes it holds

$$\|u - u^I\|_\varepsilon \le C\big(\varepsilon^{1/4} N^{-1} \ln N + N^{-2}\big). \tag{4.57}$$

Actually (4.57) is a mere consequence of Lemma 41 which proved for bilinears that

$$\varepsilon^{1/4} |u - u^I|_1 \le C N^{-1} \ln N \qquad \text{and} \qquad \|u - u^I\|_0 \le C N^{-2}. \tag{4.58}$$

Similarly, one may obtain this estimate for linear elements and for the pointwise interpolation error (see e.g. [64, 43]):

$$\|u - u^I\|_{L_\infty(\Omega_0)} \le C N^{-2} \qquad \text{and} \qquad \|u - u^I\|_{L_\infty(\Omega \setminus \Omega_0)} \le C(N^{-1} \ln N)^2. \tag{4.59}$$

Based on the quasi-optimality (4.4) the bound (4.57) implies the error estimate

$$\|u - u^N\|_\varepsilon \le C \|u - u^I\|_\varepsilon \le C\big(\varepsilon^{1/4} N^{-1} \ln N + N^{-2}\big). \tag{4.60}$$

However, we are interested in more meaningful results as the energy norm is too weak. Unfortunately it is not clear how to obtain a quasi-optimality result for the balanced norm $\|\cdot\|_b$ defined in (4.38) because in the latter the multiplier of the H^1 semi-norm component is reduced to $\varepsilon^{1/4}$. This implies that the bilinear form $a(\cdot,\cdot)$ form is not coercive in V (or even V^N) with respect to $\|\cdot\|_b$.

The key idea to sidestep the coercivity loss is to stick with the energy norm but to use instead of the Lagrange interpolant a projection that enables one to separate the two components of the bilinear form. We shall shorty see that the weighted global L_2 projection $\pi u \in V^N$ of u defined by

$$\big(c(\pi u - u), v^N\big) = 0 \quad \text{for all } v^N \in V^N \tag{4.61}$$

does this trick. Therefore we first discuss the approximation properties of this projection.

Lemma 48. *Let* (4.58) *and* (4.59) *be satisfied. Then for the weighted global L_2 projection on a Shishkin mesh it holds*

$$\|u - \pi u\|_{L_\infty(\Omega)} \le C \|u - u^I\|_{L_\infty(\Omega)} \qquad \text{and} \qquad \varepsilon^{1/4} |u - \pi u|_1 \le C N^{-1} (\ln N)^{3/2}. \tag{4.62}$$

Proof. For the maximum-norm bound we use the L_∞-stability of the global L_2 projection which holds true on standard layer adapted meshes [54]:

$$\|u - \pi u\|_{L_\infty(\Omega)} \le \|u - u^I\|_{L_\infty(\Omega)} + \|\pi(u - u^I)\|_{L_\infty(\Omega)} \le C \|u - u^I\|_{L_\infty(\Omega)}. \tag{4.63}$$

Consequently, by (4.59),

$$\|u - \pi u\|_{0,\Omega_0} \le C \|u - \pi u\|_{L_\infty(\Omega_0)} \le C \|u - u^I\|_{L_\infty(\Omega_0)} \le C N^{-2}. \tag{4.64}$$

For the other estimate we start off in Ω_0. Another triangle inequality yields:

$$|u - \pi u|_{1,\Omega_0} \le |u - u^I|_{1,\Omega_0} + |u^I - \pi u|_{1,\Omega_0}. \tag{4.65}$$

In Ω_0 the mesh is (quasi-)uniform with a mesh size $H = \mathcal{O}(N^{-1})$. Applying an inverse estimate to the discrete component gives with (4.64) and (4.58) that

$$|u^I - \pi u|_{1,\Omega_0} \le C N \|u^I - \pi u\|_{0,\Omega_0} \le C N \big(\|u^I - u\|_{0,\Omega_0} + \|u - \pi u\|_{0,\Omega_0}\big) \le C N^{-1}.$$

Collecting (4.58), (4.65) and (4.3) we arrive at

$$|u-\pi u|_{1,\Omega_0} \le C\varepsilon^{-1/4}N^{-1}\ln N.$$

In $\Omega\setminus\Omega_0$ a similar argument gives

$$\begin{aligned}
|u-\pi u|_{1,\Omega\setminus\Omega_0} &\le |u-u^I|_{1,\Omega\setminus\Omega_0} + |u^I-\pi u|_{1,\Omega\setminus\Omega_0}\\
&\le |u-u^I|_1 + \frac{C}{h}\|u^I-\pi u\|_{0,\Omega\setminus\Omega_0}\\
&\le C\Big(\varepsilon^{-1/4}N^{-1}\ln N + \frac{\operatorname{meas}(\Omega\setminus\Omega_0)^{1/2}}{\varepsilon^{1/2}N^{-1}\ln N}\|u^I-\pi u\|_{L_\infty(\Omega\setminus\Omega_0)}\Big).
\end{aligned}$$

Note that $\operatorname{meas}(\Omega\setminus\Omega_0)^{1/2} = \varepsilon^{1/4}(\ln N)^{1/2}$ and furthermore from (4.63) and (4.59) it follows that

$$\|u^I-\pi u\|_{L_\infty(\Omega\setminus\Omega_0)} \le CN^{-2}(\ln N)^2.$$

Hence, the proof of Lemma 48 is complete. □

Remark 40. In [54] it is established that the L_2-orthogonal projection onto linear spline spaces is bounded as an operator in L_∞ provided two geometric conditions on the underlaying family of triangulations are met: The valence of any node and the so called depth of local area growth with ratio r are bounded uniformly. These conditions are fulfilled on a Shishkin-mesh because the maximal valence is eight due to the construction of the mesh. Moreover, it can be broken down into nine uniform sub-triangulations (c.f. the domain decomposition in the left of Figure 3.7) and consequently in each sub-triangulation two triangles have an area ratio of one. This result holds also for other two-dimensional meshes of Shishkin- or Bakhvalov-type. For the proof the scaled mass matrix is studied which has some nice properties if linear elements in 2D are considered. Therefore the proof in [54] is limited to this setting.

For bilinear finite elements the L_∞-stability of the global L_2 projection follows immediately from the univariate case [14] without any restrictions on the family of underlaying tensor product partitions. Moreover, the stability constant $C = 9$ can sharply be specified [54].

Since the integral kernel c of (4.61) is bounded by $0 < 2(c^\star)^2 \le c \le \|c\|_{L_\infty(\Omega)}$ these results carry over to the weighted L_2 projection.

Now we are able to prove the main result of this section: uniform convergence of the Galerkin finite element solution in the balanced norm $\|\cdot\|_b$.

Theorem 49. *Assume that the Lagrange interpolation error estimates* (4.58) *and* (4.59) *are satisfied. Then there exists a constant C such that*

$$\|u-u^N\|_b \le CN^{-1}(\ln N)^{3/2} \tag{4.66}$$

where u^N is the Galerkin finite element solution obtained on a Shishkin mesh with $\mathcal{O}(N^2)$ linear or bilinear elements.

Proof. A triangle inequality and Lemma 48 yield

$$|u-u^N|_1 \le |u-\pi u|_1 + |\pi u-u^N|_1 \le \varepsilon^{-1/4}N^{-1}(\ln N)^{3/2} + |\pi u-u^N|_1. \tag{4.67}$$

For the discrete component $\xi := \pi u - u^N$ we use coercivity with respect to the energy norm and Galerkin orthogonality

$$\varepsilon|\xi|_1^2 \le \|\xi\|_\varepsilon \le C\big(\varepsilon|\xi|_1^2 + (c\xi,\xi)\big) = C\varepsilon\big(\nabla(\pi u-u),\nabla\xi\big) + \big(c(\pi u-u),\xi\big). \tag{4.68}$$

Based on (4.61) this yields

$$|\xi|_1 = |\pi u-u^N|_1 \le C|\pi u-u|_1. \tag{4.69}$$

Collecting (4.62), (4.69) and (4.67) we arrive at

$$\varepsilon^{1/4}|u-u^N|_1 \le N^{-1}(\ln N)^{3/2}.$$

For the L_2-norm error we use $\|u-u^N\|_0 \le \|u-u^N\|_\varepsilon$ and (4.60) which proves first order convergence uniformly in ε. Hence, the proof is complete. □

Remark 41. Given certain additional compatibility conditions one may prove

$$\|u - u^N\|_b \le CN^{-k}(\ln N)^{k+1/2}$$

for the Galerkin finite element solution u^N on a Shishkin mesh formed by Q_k elements with $\lambda_0 = k + 1$ and $k > 1$. Note that for a tensor product mesh the L_∞-stability of the weighted L_2 projection again follows from the corresponding one-dimensional result, c.f. Remark 40. Moreover, one may use cut-off functions to conclude such a result in certain regions of the domain even if the compatibility conditions are violated.

4.3.1 L_2-norm error bounds and supercloseness

There are in principle several ways to derive an estimate for $\|u - u^N\|_0$. We may use the estimate for the energy norm. In the interesting regime $\sqrt{\varepsilon} \le CN^{-1}$ the bound (4.60) gives

$$\|u - u^N\|_0 \le CN^{-3/2} \ln N. \tag{4.70}$$

In order to derive a better estimate one can try to apply some duality argument, also known as Aubin-Nitsche trick. This was done in [42] for the method of Section 4.1. However, it is unclear how to bound the energy norm of the solution w of $a(w, v) = (u - u^N, v)$ against $\|u - u^N\|_0$ in such a way that robust second order convergence can be concluded in the singularly perturbed case.

An approach similar to the derivation of (4.69) involves the H_0^1-projection $\pi^\star v \in V^N$ of $v \in H^1(\Omega)$ defined by

$$\big(\nabla(v - \pi^\star v), \nabla v^N\big) = 0 \quad \text{for all } v^N \in V^N.$$

For $\pi^\star u - u^N$ we obtain similarly to (4.68) that

$$\|\pi^\star u - u^N\|_0 \le C\|\pi^\star u - u\|_0.$$

Hence,

$$\|u - u^N\|_0 \le \|\pi^\star u - u\|_0 + \|\pi^\star u - u^N\|_0 \le C\|\pi^\star u - u\|_0.$$

Unfortunately it is unclear how to estimate the last term. Let us assume that the H_0^1-projection onto the space of linear or bilinear elements over a family of Shishkin meshes is $\big(L_\infty(\Omega), L_2(\Omega)\big)$-stable, i.e.

$$\|\pi^\star v\|_0 \le C\|v\|_\infty. \tag{4.71}$$

Then we could prove that

$$\|\pi^\star u - u\|_0 \le \|\pi^\star(u - u^I)\|_0 + \|u - u^I\|_0 \le C\|u - u^I\|_{L_\infty(\Omega)} \le N^{-2}(\ln N)^2.$$

Note that in 1D (4.71) holds true for linear elements, because the H_0^1-projection $\pi^\star v$ of v coincides with the nodal interpolant v^I of v. This fact can easily be checked using integration by parts:

$$\big((u - u^I)', (v^N)'\big)_{(x_{i-1}, x_i)} = (u - u^I)(v^N)'\big|_{x_{i-1}}^{x_i} - \big(u - u^I, (v^N)''\big) = 0.$$

With respect to the two-dimensional case the situation is unclear. The work [26] may be considered a starting point in which bounds for the inverse of the stiffness matrix (the flexibility matrix) over non-uniform meshes are obtained.

Finally, one may use a supercloseness result to deduce a bound for the L_2-error. Supercloseness is the terminology for the phenomenon that the difference between the approximate solution u^N and the nodal interpolant u^I of the exact solution u is of higher order than the error itself. Supercloseness results for the reaction-diffusion problem have been obtained in [38, 82, 81, 23]. However, in [38] the transition point λ of the piecewise uniform anisotropic mesh is chosen proportional to $\varepsilon|\ln \varepsilon|$ and the other papers consider graded meshes. A supercloseness result can easily be obtained by applying the techniques of [44, 80]. These papers deal with convection-diffusion problems.

Theorem 50. *Let the corner compatibility conditions* (3.86) *be satisfied so that Lemma 34 holds true. Assume* $\varepsilon^{1/2} \le (\ln N)^{-2}$ *and that for the constant in* (3.88) *it holds* $\lambda_0 \ge 2$. *Then the bilinear Galerkin FEM approximation* u^N *on the Shishkin mesh of Section 3.7 has the supercloseness property*

$$\|u^I - u^N\|_\varepsilon \le CN^{-2}.$$

Proof. Using the coercivity of the bilinear form $a(\cdot,\cdot)$ with respect to the energy norm and Galerkin orthogonality one obtains for $\xi := u^I - u^N$:

$$\|\xi\|_\varepsilon^2 \le |a(\xi,\xi)| = |a(u-u^I,\xi)| \le \varepsilon\big|\big(\nabla(u-u^I),\nabla\xi\big)\big| + \big|\big(c(u-u^I),\xi\big)\big|. \tag{4.72}$$

While the reaction term can be bounded by means of the Cauchy-Schwarz inequality and (4.58):

$$\big|\big(c(u-u^I),\xi\big)\big| \le \|c\|_{L_\infty(\Omega)}\|u-u^I\|_0\|\xi\|_0 \le CN^{-2}\|\xi\|_\varepsilon, \tag{4.73}$$

the diffusion term needs a more sophisticated approach. From the so-called Lin identities (see e.g. [44, 80, 43, 64]) one can deduce for bilinear elements that

$$\big|\big((v-v^I)_x,\xi_x\big)_T\big| \le Ch_{T,y}^2\|\xi_x\|_{0,T}\|v_{xyy}\|_{0,T},$$

where $v \in H^3(T)$, T is an axis-parallel rectangular element and $h_{T,y}$ denotes its size in y-direction. We use this estimate for u in $\Omega_f \cup \Omega_1 \cup \Omega_3$, i.e. whenever $h_{T,y} = h$ is small:

$$\begin{aligned}\varepsilon\big|\big((u-u^I)_x,\xi_x\big)_{\Omega_f\cup\Omega_1\cup\Omega_3}\big| &\le C\varepsilon^2N^{-2}(\ln N)^2\|\xi_x\|_{0,\Omega_f\cup\Omega_1\cup\Omega_3}\|u_{xyy}\|_{0,\Omega_f\cup\Omega_1\cup\Omega_3}\\ &\le C\varepsilon^{1/2}N^{-2}(\ln N)^2\|\xi_x\|_{\varepsilon,\Omega_f\cup\Omega_1\cup\Omega_3}.\end{aligned} \tag{4.74}$$

Here we also used (3.87). In the remainder of the domain $\Omega_0 \cup \Omega_2 \cup \Omega_4$ the involved grid size is large. We estimate the components of (3.87a) separately. Yet we obtain for S based on (3.87b) that

$$\begin{aligned}\varepsilon\big|\big((S-S^I)_x,\xi_x\big)_{\Omega_0\cup\Omega_2\cup\Omega_4}\big| &\le C\varepsilon N^{-2}\|\xi_x\|_{0,\Omega_0\cup\Omega_2\cup\Omega_4}\|S_{xyy}\|_{0,\Omega_0\cup\Omega_2\cup\Omega_4}\\ &\le CN^{-2}\|\xi_x\|_{\varepsilon,\Omega_f\cup\Omega_1\cup\Omega_3}.\end{aligned} \tag{4.75}$$

Proceed in a similar manner with $E := E_2 + E_4$. Note that $\|E_{xyy}\|_{0,\Omega_0\cup\Omega_2\cup\Omega_4} \le C\varepsilon^{-1/4}$ in contrast to $\|S_{xyy}\|_{0,\Omega_0\cup\Omega_2\cup\Omega_4} \le C\varepsilon^{-1/2}$. Hence, E_{xyy} is better behaved than S_{xyy} in this subregion.

Finally, set $E := E_1 + E_3 + E_{12} + E_{23} + E_{34} + E_{41}$. It remains to obtain an estimate for E in $\Omega_0 \cup \Omega_2 \cup \Omega_4$. We use the smallness of the domain $\operatorname{meas}(\Omega_2 \cup \Omega_4) = \varepsilon^{1/4}\ln N$ and obtain with (3.87) that

$$\begin{aligned}\varepsilon\big|\big((E-E^I)_x,\xi_x\big)_{\Omega_2\cup\Omega_4}\big| &\le \varepsilon\|(E-E^I)_x\|_{L_\infty(\Omega_2\cup\Omega_4)}\operatorname{meas}(\Omega_2\cup\Omega_4)^{1/2}\|\xi_x\|_{0,\Omega_2\cup\Omega_4}\\ &\le C\varepsilon^{1/4}N^{-\lambda_0}(\ln N)^{1/2}\|\xi_x\|_{\varepsilon,\Omega_2\cup\Omega_4},\end{aligned} \tag{4.76}$$

because of a stability result for bilinear nodal interpolation:

$$\begin{aligned}\|(E-E^I)_x\|_{L_\infty(\Omega_2\cup\Omega_4)} &\le \|E_x\|_{L_\infty(\Omega_2\cup\Omega_4)} + \|(E^I)_x\|_{L_\infty(\Omega_2\cup\Omega_4)}\\ &\le C\|\nabla E\|_{L_\infty(\Omega_2\cup\Omega_4)} \le \varepsilon^{-1/2}N^{-\lambda_0}.\end{aligned}$$

In Ω_0 the mesh is uniform with mesh size $H = \mathcal{O}(N^{-1})$. Hence, by the Cauchy-Schwarz inequality and an inverse estimate

$$\begin{aligned}\varepsilon\big|\big((E-E^I)_x,\xi_x\big)_{\Omega_0}\big| &\le \varepsilon\|(E-E^I)_x\|_{0,\Omega_0}\|\xi_x\|_{0,\Omega_0}\\ &\le C\varepsilon\|(E-E^I)_x\|_{0,\Omega_0}N\|\xi\|_{0,\Omega_0}\\ &\le \varepsilon^{1/2}N^{-\lambda_0}\|\xi\|_{\varepsilon,\Omega_0}.\end{aligned} \tag{4.77}$$

In the last inequality we used the sharp interpolation error estimate (4.19) and (3.87):

$$\begin{aligned}\|(E-E^I)_x\|_{0,\Omega_0} &\le H(\|E_{xx}\|_{0,\Omega_0} + \|E_{xy}\|_{0,\Omega_0})\\ &\le CN^{-1}\big(\varepsilon^{-1/4}N^{-\lambda_0} + \varepsilon^{-1/2}N^{-2\lambda_0}\big)\end{aligned}$$

Collect (4.72), (4.73), (4.74), (4.75), (4.76) and (4.77) to obtain the estimate for the x-derivative. The other derivative with respect to y is bounded similarly. □

Remark 42. Inspecting the proof of Theorem 50 we see that given $\|\nabla S_{xy}\|_{0,\Omega\setminus\Omega_f} \le C\varepsilon^{-1/4}$ one can prove for a bilinear function v^N that

$$\varepsilon\big|\big(\nabla(u-u^I),\nabla v^N\big)\big| \le C\varepsilon^{1/4}N^{-2}(\ln N)^2\big(\varepsilon^{1/2}|v^N|_1+\varepsilon^{1/4}\|v^N\|_0\big). \tag{4.78}$$

Here we sharpened (4.75) and also used the extra multiplier $\varepsilon^{1/4}$ in (4.77). Note that this is the only term estimated against an L_2-Norm of ξ. The additional positive powers in ε in the right hand side of (4.78) are again evidence, that the energy norm is not correctly scaled for this problem. In fact (4.78) shows that

$$\varepsilon^{1/2}\big|\big(\nabla(u-u^I),\nabla v^N\big)\big| \le CN^{-2}(\ln N)^2\|v^N\|_b. \tag{4.79}$$

The estimate (4.79) raises hope that the Galerkin finite element solution provides some kind of supercloseness result even in the balanced norm. However, it is not clear how to proceed.

From Theorem 50 we deduce an optimal L_2-error estimate.

Lemma 51. *Let the assumptions of Theorem 50 be satisfied. Then for the bilinear Galerkin FEM approximation u^N on the Shishkin mesh of Section 3.7 it holds*

$$\|u-u^N\|_0 \le CN^{-2}.$$

Proof. Combine Theorem 50 and (4.58). Note that if the energy norm was adequate for this problem we would have obtained a result with the additional factor $(\ln N)^2$. Hence, we used the weakness of the energy norm to our advantage. □

4.3.2 Maximum-norm error bounds

In contrast to the energy norm the right balancing of the norm makes it possible to obtain L_∞-error bounds in subregions of the boundary. We study this question in this section.

Firstly, we want to derive a bound for the interior subdomain Ω_0. We start off with a triangle inequality and (4.59):

$$\|u-u^N\|_{L_\infty(\Omega_0)} \le \|u-u^I\|_{L_\infty(\Omega_0)} + \|u^I-u^N\|_{L_\infty(\Omega_0)} \le N^{-2} + \|u^I-u^N\|_{L_\infty(\Omega_0)}.$$

In the interior subdomain Ω_0 the mesh is (quasi-)uniform with the large mesh size $H=\mathcal{O}(N^{-1})$. Consequently, an inverse estimate gives

$$\|u^I-u^N\|_{L_\infty(\Omega_0)} \le CN\|u^I-u^N\|_{0,\Omega_0}. \tag{4.80}$$

For bilinear elements we obtain from Theorem 50:

$$\|u-u^N\|_{L_\infty(\Omega_0)} \le CN^{-1}.$$

If linear elements are considered we use (4.70). This yields for $\sqrt{\varepsilon}\le CN^{-1}$ that

$$\|u-u^N\|_{L_\infty(\Omega_0)} \le CN^{-1/2}\ln N.$$

Next we turn our attention to the boundary region Ω_1. We use a technique presented in [43, page 279]. There the author refers to [75, pp. 11, 12].

Let (x_i,y_j) be any mesh point in $\overline{\Omega_1}$. Using $(\pi u-u^N)(x_i,0)=0$ one finds that

$$\big|(\pi u-u^N)(x_i,y_j)\big| = \left|\int_0^{y_j}(\pi u-u^N)_y(x_j,t)\,\mathrm{d}t\right|.$$

Next, applying an inverse estimate for the modulus of the integrand with respect to x yields

$$\big|(\pi u-u^N)(x_i,y_j)\big| \le CN\int_0^{y_j}\int_{x_{i-1}}^{x_i}\big|(\pi u-u^N)_y(s,t)\big|\,\mathrm{d}s\mathrm{d}t.$$

If $x_i=\lambda$ we integrate over (x_i,x_{i+1}) instead of (x_{i-1},x_i). Since $y_i\le\lambda\le C\varepsilon^{1/2}\ln N$ a Cauchy-Schwarz inequality gives

$$\begin{aligned}\big|(\pi u-u^N)(x_i,y_j)\big| &\le CN\int_0^{\lambda}\int_{x_{i-1}}^{x_i}\big|(\pi u-u^N)_y(s,t)\big|\,\mathrm{d}s\mathrm{d}t\\ &\le CN\varepsilon^{1/4}N^{-1/2}(\ln N)^{1/2}|\pi u-u^N|_1.\end{aligned} \tag{4.81}$$

Collecting (4.69), (4.62) and (4.81) one arrives at

$$\left|(\pi u-u^N)(x_i,y_j)\right| \le C\varepsilon^{1/4}N^{1/2}(\ln N)^{1/2}|\pi u-u|_1 \le CN^{-1/2}(\ln N)^2. \tag{4.82}$$

Hence, by (4.82), (4.62) and (4.59)

$$\|u-u^N\|_{L_\infty(\Omega_1)} \le \|u-\pi u\|_{L_\infty(\Omega_1)} + \|\pi u-u^N\|_{L_\infty(\Omega_1)} \le CN^{-1/2}(\ln N)^2. \tag{4.83}$$

Clearly, an analogous result holds true in Ω_i, $i\in\{2,3,4\}$.

It might be possible to obtain a similar result in Ω_h using certain finite difference arguments similarly to [74, 75].

Finally, we want to mention that for linear elements Kopteva [34] provided numerical evidence and a theoretical justification that for the Galerkin finite element method applied to reaction-diffusion problems one can not expect an error bound better than

$$\|u-u^N\|_{L_\infty(\Omega)} \le CN^{-1}\ln N$$

uniformly in ε without imposing restrictions on the bisection pattern that generated $\tilde\Omega^N$ from the anisotropic mesh Ω^N. Note that in her example $u\in C^\infty(\overline{\Omega})$ and $\|u-u^I\|_{L_\infty(\Omega)} \le CN^{-2}(\ln N)^2$. More precisely, if the directions of the diagonals of $\tilde\Omega^N$ change in certain regions close to the boundary a considerable reduction in convergence speed can be observed. Hence, the use of anisotropic triangulations in general breaks the quasi-optimality property (2.25) in the maximum norm.

4.4 Numerical verification

In this Section we want to verify our theoretical findings. Consider the test problem

Example 1.

$$-\varepsilon\Delta u + \left(1+x^2y^2\mathrm{e}^{xy/2}\right)u = f \qquad \text{in } \Omega=(0,1)^2, \tag{4.84a}$$

$$u = u_d \qquad \text{on } \partial\Omega, \tag{4.84b}$$

where the data u_d and f are chosen in such a way that the exact solution of (4.84) is given by

$$\begin{aligned} u(x,y) &:= x^3(1+y^2)+\sin(\pi x^2)+\cos(\pi y/2)\\ &\quad + (x+y)\left(\mathrm{e}^{-2x/\sqrt{\varepsilon}}+\mathrm{e}^{-2(1-x)/\sqrt{\varepsilon}}+\mathrm{e}^{-3y/\sqrt{\varepsilon}}+\mathrm{e}^{-3(1-y)/\sqrt{\varepsilon}}\right).\end{aligned}$$

This problem was also studied in [45] and [42] numerically. It fulfills the estimates of Lemma 34. Consequently, (4.58) and (4.59) hold true as well.

In order to permit a comparison with the numerical experiments of [42] we use the same Shishkin mesh, i.e. we set the constants of (3.88) in Section 3.7 to $c^\star=\sqrt{1/2}$ and $\lambda_0=2$. In fact, this gives slightly better results than the choice $\lambda_0=3$ which is required by our theory for the CIP-method. A similar behaviour is observed for the biquadratic Galerkin FEM and other test problems considered. We use a standard sixteen-point Gauss-Legendre quadrature rule in the assembly routines of the stiffness matrix and the load vector and in order to calculate the L_2 based errors on each mesh rectangle. Similarly, we use the corresponding four-point rule on edges of the triangulation. We adopt the table layout of [42] and present in each table the errors (lower half) and rates (upper half) for a range of values of ε and N.

First, we examine the accuracy of the Galerkin FEM according to (4.2) and bilinear elements. In agreement with (4.66) of Theorem 49 Table 4.1 shows uniform and balanced convergence with respect to the perturbation parameter ε. For very small ε, i.e. $\varepsilon\le 10^{-8}$, the error $e_N := \|u-u^N\|_b = \left(\varepsilon^{1/2}|u-u^N|_1^2+\|u-u^N\|_0^2\right)^{1/2}$ attains same values rounded to two decimal places. With respect to the discretization parameter N linear convergence up to a logarithmic factor is observed. The rate of convergence p is calculated using two consecutive discretization levels (N_1,e_{N_1}) and (N_2,e_{N_2}):

$$p=\frac{\ln e_{N_1}-\ln e_{N_2}}{\ln(N_1^{-1}\ln N_1)-\ln(N_2^{-1}\ln N_2)}.$$

N	$\varepsilon = 1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-12}$	$\varepsilon = 10^{-16}$	max $\varepsilon = 10^{-2m}$ $0 \le m \le 8$
16	1.47	1.27	0.39	0.39	0.39	0.39	0.39
32	1.35	1.30	0.61	0.61	0.61	0.61	0.61
64	1.29	1.27	0.81	0.81	0.81	0.81	0.81
128	1.24	1.24	0.92	0.92	0.92	0.92	0.92
256	1.20	1.20	0.97	0.97	0.97	0.97	0.97
512	1.18	1.18	0.99	0.99	0.99	0.99	0.99
16	2.02e-1	1.10e-0	2.01e-0	2.01e-0	2.01e-0	2.01e-0	2.01e-0
32	1.01e-1	6.06e-1	1.68e-0	1.67e-0	1.67e-0	1.67e-0	1.68e-0
64	5.07e-2	3.11e-1	1.23e-0	1.22e-0	1.22e-0	1.22e-0	1.23e-0
128	2.53e-2	1.57e-1	7.93e-1	7.92e-1	7.92e-1	7.92e-1	7.93e-1
256	1.27e-2	7.85e-2	4.73e-1	4.72e-1	4.72e-1	4.72e-1	4.73e-1
512	6.33e-3	3.93e-2	2.70e-1	2.70e-1	2.70e-1	2.70e-1	2.70e-1
1024	3.17e-3	1.96e-2	1.51e-1	1.51e-1	1.51e-1	1.51e-1	1.51e-1

Table 4.1: Error $\|u - u^N\|_b$ of the bilinear Galerkin FEM on a sequence of Shishkin-meshes.

N	$\varepsilon = 1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-12}$	$\varepsilon = 10^{-16}$	max $\varepsilon = 10^{-2m}$ $0 \le m \le 8$
16	2.93	2.67	1.07	1.07	1.07	1.07	1.07
32	2.71	2.64	1.44	1.44	1.44	1.44	1.44
64	2.57	2.55	1.73	1.73	1.73	1.73	1.73
128	2.48	2.47	1.89	1.89	1.89	1.89	1.89
256	2.41	2.41	1.96	1.96	1.96	1.96	1.96
16	1.17e-2	2.42e-1	1.05e-0	1.05e-0	1.05e-0	1.05e-0	1.05e-0
32	2.95e-3	6.89e-2	6.37e-1	6.35e-1	6.35e-1	6.35e-1	6.37e-1
64	7.39e-4	1.79e-2	3.06e-1	3.05e-1	3.05e-1	3.05e-1	3.06e-1
128	1.85e-4	4.51e-3	1.20e-1	1.20e-1	1.20e-1	1.20e-1	1.20e-1
256	4.62e-5	1.13e-3	4.17e-2	4.16e-2	4.16e-2	4.16e-2	4.17e-2
512	1.15e-5	2.83e-4	1.35e-2	1.34e-2	1.34e-2	1.34e-2	1.35e-2

Table 4.2: Error $\|u - u^N\|_b$ of the biquadratic Galerkin FEM on a sequence of Shishkin meshes.

Next, we study the performance of the Galerkin FEM with Q_2 elements. Again the error is balanced in the norm $\|\cdot\|_b$ and robust convergence of order two can be observed, cf. Table 4.2: For all test problems the error behaves like

$$\|u - u^N\|_b \le CN^{-2}(\ln N)^2$$

which corresponds to the results of §2 up to a root of a logarithmic factor.

Now we turn our attention to the accuracy of the results of the CIP-method introduced in Section 4.2. Here we shall consider Q_2 elements only. Note that inhomogeneous Dirichlet boundary conditions alter the CIP-method slightly. More precisely, they give rise to the additional term $\sqrt{\varepsilon}(c\frac{\partial v}{\partial n}, u_d)_{\partial\Omega}$ on the right hand side of (4.37).

According to Remark 39 we set

$$\sigma_e := \begin{cases} \varepsilon & \text{if } e \text{ is a long (type-I or type-II) edge,} \\ \varepsilon^{1/2} & \text{if } e \text{ is a short (type-III or type-IV) edge.} \end{cases}$$

In case that the parameters ε and N are such that the mesh used is not of Shishkin type but a uniform one we use $\sigma_e := \varepsilon$ on all edges.

One more slight modification is needed in order to facilitate a comparison with the results of [42]: Lin and Stynes scale the dominant error component that corresponds to the L_2 norm of the Laplacian of the error by the additional constant $c_1 := 1 + \mathrm{e}$ which is the maximum of c,

N	$\varepsilon = 1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-12}$	$\varepsilon = 10^{-16}$	max $\varepsilon = 10^{-2m}$ $0 \leq m \leq 8$
16	1.30	1.27	0.47	0.47	0.47	0.47	0.47
32	1.28	1.29	0.68	0.68	0.68	0.68	0.68
64	1.26	1.26	0.84	0.84	0.84	0.84	0.84
128	1.23	1.22	0.94	0.94	0.94	0.94	0.94
256	1.21	1.20	0.98	0.98	0.98	0.98	0.98
16	1.31e-0	1.86e-0	3.24e-0	3.24e-0	3.24e-0	3.24e-0	3.24e-0
32	7.08e-1	1.02e-0	2.60e-0	2.60e-0	2.60e-0	2.60e-0	2.60e-0
64	3.68e-1	5.30e-1	1.84e-0	1.84e-0	1.84e-0	1.84e-0	1.84e-0
128	1.87e-1	2.69e-1	1.17e-0	1.17e-0	1.17e-0	1.17e-0	1.17e-0
256	9.39e-2	1.36e-1	6.93e-1	6.89e-1	6.89e-1	6.89e-1	6.93e-1
512	4.69e-2	6.81e-2	3.95e-1	3.93e-1	3.93e-1	3.93e-1	3.95e-1

Table 4.3: Error $|||u - u^N|||$ of the CIP-method on a sequence of Shishkin meshes with Q_2 elements.

N	$\varepsilon = 1$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-8}$	$\varepsilon = 10^{-12}$	$\varepsilon = 10^{-16}$	max $\varepsilon = 10^{-2m}$ $0 \leq m \leq 8$
16	1.46	1.33	0.27	0.27	0.27	0.27	0.27
32	1.35	1.32	0.78	0.78	0.78	0.78	0.78
64	1.29	1.28	1.20	1.20	1.20	1.20	1.20
128	1.24	1.24	1.31	1.32	1.32	1.32	1.32
256	1.20	1.20	1.18	1.19	1.19	1.19	1.19
16	7.54e-1	1.70e-0	5.53e-0	5.60e-0	5.61e-0	5.61e-0	5.61e-0
32	3.79e-1	9.07e-1	4.86e-0	4.93e-0	4.94e-0	4.94e-0	4.94e-0
64	1.90e-1	4.62e-1	3.26e-0	3.31e-0	3.31e-0	3.31e-0	3.31e-0
128	9.50e-2	2.32e-1	1.71e-0	1.73e-0	1.73e-0	1.73e-0	1.73e-0
256	4.75e-2	1.16e-1	8.20e-1	8.25e-1	8.25e-1	8.25e-1	8.25e-1
512	2.38e-2	5.82e-2	4.15e-1	4.15e-1	4.15e-1	4.15e-1	4.15e-1

Table 4.4: Error $|||u - u^N|||$ of the Galerkin FEM on a sequence of Shishkin meshes with Q_2 elements.

see [42, (3.5)]. Consequently, we modify the norm (4.38) to

$$|||v|||^2 := \frac{\varepsilon^{3/2}}{c_1} \sum_{T \in \Omega^N} \|\Delta v\|_{0,T}^2 + \varepsilon^{1/2} |v|_1^2 + \|v\|_0^2 + \sum_{e \in \mathcal{E}^N} \left(\sigma_e \left[\!\left[\frac{\partial v}{\partial n} \right]\!\right], \left[\!\left[\frac{\partial v}{\partial n} \right]\!\right] \right)_e. \tag{4.85}$$

The CIP-method presented appears to be robust in the balanced norm $|||\cdot|||$ as shown by Table 4.3: A uniform first order of convergence can be observed. We also measured the performance of the biquadratic Galerkin FEM in the norm $|||\cdot|||$. These results are shown in Table 4.4. Rates and errors of both methods are quite similar. The error in the norm $|||\cdot|||$ appears to be proportional to $N^{-1} \ln N$ uniformly with respect to ε. However, it is not clear how to explain this for the Galerkin FEM, theoretically. If ε is very small (i.e. $\varepsilon \leq 10^{-8}$) the errors of the CIP-method are slightly smaller. In this regime the error component introduced by the jump of the normal derivatives along inter-element edges is significantly smaller for the CIP-method, in fact for $N = 256$ this error component is 23 times larger for the Galerkin FEM. For $N = 512$ this quotient rises over 33.

The mixed method of [42] yields an approximate solution of similar quality which can be seen by comparing the Tables 4.3 and 4.4 with Table 6.1 of that paper. If $\sqrt{\varepsilon}$ is small compared to $1/N$ the CIP-method performs slightly better for this test problem. In the other case the mixed method yields slightly smaller errors which are in turn improved by the Galerkin FEM.

Numerically we observe second order convergence for the lower order components of $|||\cdot|||$: If we measure the error of the CIP-method in the norm $\|\cdot\|_b$ we obtain results similar to Table 4.2. However, the errors are slightly smaller for the Galerkin FEM: In the interesting regime $\varepsilon \leq 10^{-4}$ the quotient of both errors is bounded by 1.7 for the model problem considered.

N	Q_1 elements						Q_2 elements	
	$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-6}$		$\varepsilon = 10^{-2k}$, $k \in \{4,\dots,8\}$		$\varepsilon = 10^{-2k}$, $k \in \{2,\dots,8\}$	
	error	rate	error	rate	error	rate	error	rate
12	4.57e-1		4.60e-1		4.60e-1		1.51e-1	
24	3.04e-1	0.92	3.05e-1	0.92	3.06e-1	0.92	6.88e-2	1.76
48	1.88e-1	0.97	1.89e-1	0.97	1.89e-1	0.97	2.68e-2	1.90
96	1.12e-1	0.99	1.12e-1	0.99	1.12e-1	0.99	9.52e-3	1.96
192	6.45e-2	1.00	6.48e-2	1.00	6.49e-2	1.00	3.18e-3	1.99
384	3.65e-2	1.00	3.67e-2	1.00	3.67e-2	1.00		

Table 4.5: Approximate error $\|\tilde{u}^{4N} - u^N\|_b$ of the Galerkin FEM on a sequence of Shishkin meshes with Q_1 elements for problem (4.86).

N	$\|\cdot\|_b$		$\|\!\|\!\|\cdot\|\!\|\!\|$					
	$\varepsilon = 10^{-2k}$, $k \in \{2,\dots,8\}$		$\varepsilon = 10^{-4}$		$\varepsilon = 10^{-6}$		$\varepsilon = 10^{-2k}$, $k \in \{4,\dots,8\}$	
	error	rate	error	rate	error	rate	error	rate
12	1.76e-1		5.85e-1		5.83e-1		5.83e-1	
24	7.91e-2	1.78	4.02e-1	0.84	4.00e-1	0.84	4.00e-1	0.85
48	3.07e-2	1.91	2.54e-1	0.93	2.52e-1	0.93	2.51e-1	0.93
96	1.09e-2	1.97	1.52e-1	0.97	1.50e-1	0.97	1.50e-1	0.98
192	3.62e-3	1.99	8.83e-2	0.99	8.71e-2	0.99	8.69e-2	0.99

Table 4.6: Approximate errors $\|\tilde{u}^{4N} - u^N\|_b$ and $\|\!|\tilde{u}^{4N} - u^N|\!\|$ of the CIP-method on a sequence of Shishkin meshes with Q_2 elements for problem (4.86).

Similar convergence behavior was observed in numerical experiments for problems with unknown solutions even if the compatibility conditions (3.86) were violated. In these experiments the error was approximated following a double mesh principle: For the approximate solution obtained on a grid associated with the parameter N a reference solution $\tilde{u}^{4N}$ is computed using a refined mesh by performing two uniform refinements of the original grid. Hence, both grids have the same transition points (namely λ from (3.88)) but while the original grid has N^2 elements the grid for the reference solution consists of 16 times that number (this choice of refinement lead to a reliable approximation of the error in those cases in which the exact solution was known).

Moreover, for constant coefficient problems we also used an adaptive quadrature algorithm to compute all integrals up to a tolerance of 10^{-10}. A comparison of the results showed that the initial low convergence orders (see e.g. Table 4.1) — a phenomenon that is encountered in most numerical experiments for layer problems — are due to inexact assembly of the linear system. This problem gains significance if in the computation of the approximate solution meshes are used that are not sufficiently layer-adapted.

Example 2.

$$-\varepsilon\Delta u + u = 1 \qquad \text{in } \Omega = (0,1)^2 \tag{4.86a}$$
$$u = 0 \qquad \text{on } \partial\Omega. \tag{4.86b}$$

The solution of this problem will contain corner singularities that reduce its regularity. However, we still observe the same convergence behaviour for both methods. Table 4.5 shows that the error is balanced in the norm $\|\cdot\|_b$. Linear and quadratic rates of convergence are observed for Q_1 and Q_2 elements, respectively.

Table (4.6) shows the quality of the approximate solutions of the CIP-method for test problem (4.86). Again, the error is balanced in both norms: $\|\cdot\|_b$ as well as $\|\!|\cdot|\!\|$. Second order convergence is observed if the error is measured in the balanced version of the $H^1(\Omega)$ norm $\|\cdot\|_b$ uniformly with respect to the perturbation parameter ε. Comparing the last error column of Table 4.5 with the first one of Table (4.6) we see that if Q_2 elements are used for the Galerkin FEM as well as for the CIP-method then the errors are of comparable magnitude. In the stronger norm $\|\!|\cdot|\!\|$ the CIP-method appears to be first order convergent uniformly with respect to ε.

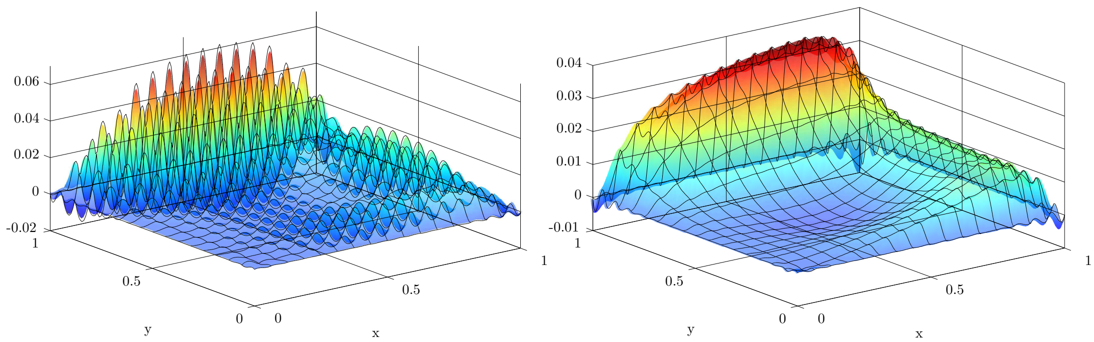

Figure 4.1: Error $u^N - u$ of the CIP-method on a uniform mesh with $N = 16$, $\varepsilon = 0.05$, $\sigma_e \equiv 0$ (left) and $\sigma_e \equiv 1$ (right)

Finally, we want to address the question how the enhanced stability of the CIP-method manifests itself in the obtained approximate solution. Therefore we consider the model problem (4.84) on a uniform mesh.

In Figure 4.1 we see that the approximate solution of the CIP-method without penalty can tend to large oscillations (left) and is unstable while for $\sigma_e \equiv 1$ one finds a considerable damping and smoothing of the error. Numerically we observe that the distance of the approximate solutions of the Galerkin FEM and the CIP-method without penalty tends to zero for $\varepsilon \to 0$: For $N = 16$ and $\varepsilon = 10^{-16}$ one has $\|u^N_{Gal} - u^N_{CIP}\|_\infty \leq 2 \cdot 10^{-5}$. Hence, in the interesting regime $\varepsilon \to 0$ the fate of the Galerkin FEM appears to be bound to the one of an unstabilized CIP method.

Studying the stability on a Shishkin mesh we change to a different and simpler test problem.

Example 3.

$$-\varepsilon\Delta u + u = f \qquad \text{in } \Omega = (0,1) \times (0,1), \tag{4.87a}$$
$$u = u_d \qquad \text{on } \partial\Omega, \tag{4.87b}$$

where the Dirichlet boundary conditions u_d and f are chosen in such a way that the exact solution is given by

$$u(x,y) := \left(\cos(\pi x) - \frac{\mathrm{e}^{-x/\sqrt{\varepsilon}} - \mathrm{e}^{-1/\sqrt{\varepsilon}}}{1 - \mathrm{e}^{-1/\sqrt{\varepsilon}}}\right)\left(1 - y - \frac{\mathrm{e}^{-y/\sqrt{\varepsilon}} - \mathrm{e}^{-1/\sqrt{\varepsilon}}}{1 - \mathrm{e}^{-1/\sqrt{\varepsilon}}}\right).$$

Thus, u is characterized by typical exponential boundary layers along the edges $x = 0$, $y = 0$ and a corner layer near the origin.

The need to introduce another test problem is caused by the fact that we do not observe strong oscillations of the errors of both methods close to the layers for Example 1. Note that the boundary conditions of Example 1 can not be satisfied by piecewise polynomials. The approximation of the boundary conditions gives rise to an additional error component that seems to cover the smaller oscillations of the Galerkin FEM along the boundary. Hence, one can expect the oscillations to emerge once the boundary condition is approximated sufficiently well.

Figure 4.2 shows the error function $u^N - u$ for $\varepsilon = 10^{-8}$ using Q_2 elements on a Shishkin mesh with $N = 32$ intervals in each coordinate direction. The error is characterized by oscillations within the first rows of anisotropic elements along the critical boundary indicating a stability problem.

The error $u^N - u$ of the CIP-method with Q_2 elements, $N = 32$, $\varepsilon = 10^{-8}$ is shown in Figure 4.3. In comparison to the corresponding Galerkin error (Figure 4.2) the oscillations in the layer lose their high-frequency character — the error is smoother but a dominant pointwise error is dragged from the boundary further into the domain. All these effects can be explained by the fact that the jump of the normal derivative along inter-element edges is being penalized: The ability of the approximate solution to fold over these edges is reduced. This also explains why the pointwise error is increased.

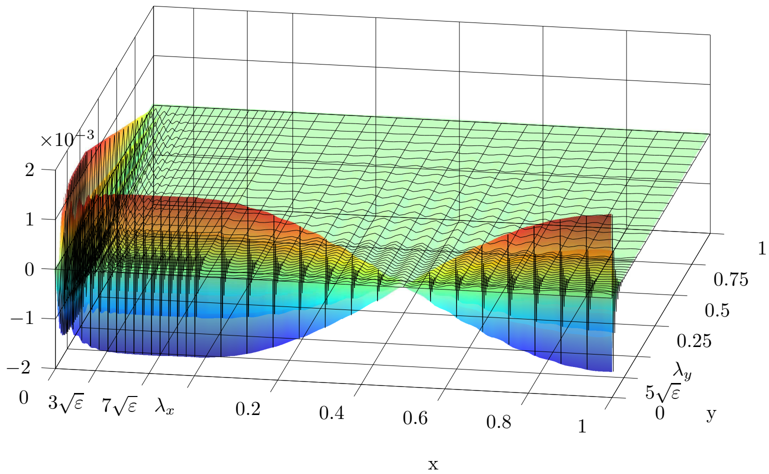

Figure 4.2: Error $u^N - u$ of the Q_2-Galerkin FEM, $N = 32$ and $\varepsilon = 10^{-8}$. The layer region is plotted with exaggerated width.

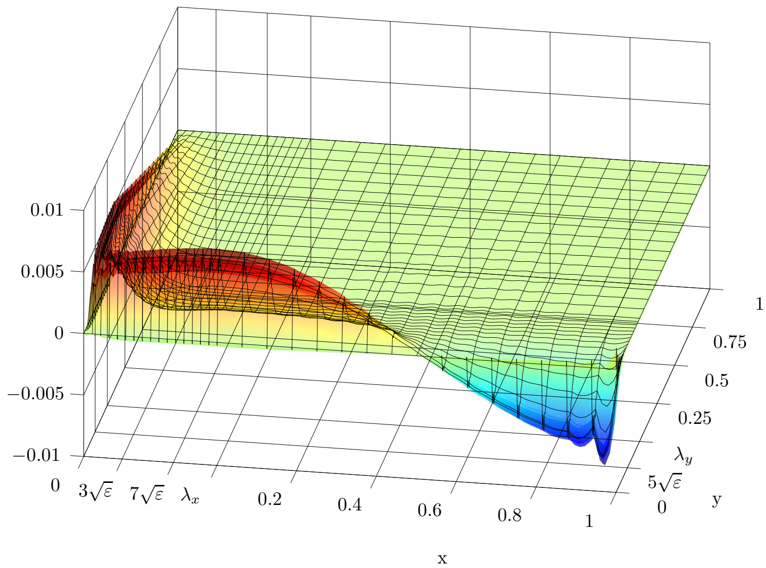

Figure 4.3: Error $u^N - u$ of the Q_2-CIP-method, $N = 32$, $\varepsilon = 10^{-8}$. The layer region is plotted with exaggerated width.

In this respect we remark that it is well known that any stabilized method on a layer-adapted mesh cannot have better convergence properties than the underlaying Galerkin method. Nevertheless practically one prefers stabilized versions: the discrete problems are easier to solve and oscillations on the discrete solution are damped.

4.5 Further developments and summary

In this chapter we have seen that the energy norm is to weak to yield meaningful results for singularly perturbed reaction-diffusion problems whenever the behavior of the solution within the layer is of critical interest. The energy norm is unable to capture the boundary layers which are the predominant feature of the solutions to these problems. With the balanced method of Lin and Stynes [42] (see also Section 4.1) and the CIP method of Section 4.2 two methods have been developed for which error estimates in a suitable balanced norm can be proven. Note that the bilinear forms of both methods are coercive with respect to the balanced norm in which even second order derivatives are measured — weighted by appropriate powers of ε.

In contrast to this the bilinear form associated with the Galerkin finite element method is coercive with respect to the energy norm and it is unclear how to prove coercivity with respect to a balanced norm (and a subspace of $H_0^1(\Omega)$). Nevertheless it is possible to prove a convergence result in the balanced norm $\|\cdot\|_b$, see Section 4.3 and [61]. The key idea is the usage of a projection associated with the reduced problem. This way the components of the bilinear form decouple and one can use standard energy arguments to estimate the resulting error terms separately.

This trick in its general form has also been used in the analysis of a bilinear streamline-diffusion stabilized FEM for a convection-diffusion problem with characteristic layers in a balanced norm [25]. As mentioned in Remark 34 the characteristic layers are not well represented in the energy norm. The argument is essentially the same. However, the analysis is more difficult due to the complicate structure of the projection needed for the trick to work.

Melenk and Xenophontos try to generalize this trick into a different direction. We follow [78] and study the one-dimensional case. Set $\Omega := (0,1)$. For the construction of a rp-version of the Galerkin FEM for problem (4.1) a so-called spectral boundary layer mesh is introduced:

$$\Omega^S := \begin{cases} \{0,1\} & \text{for } \sqrt{\varepsilon}\kappa p \geq 1/2, \\ \{0, \sqrt{\varepsilon}\kappa p, 1-\sqrt{\varepsilon}\kappa p, 1\} & \text{for } \sqrt{\varepsilon}\kappa p < 1/2. \end{cases}$$

Note that for all values of the parameters $\kappa > 0$, $p \in \mathbb{N}$ and $0 < \varepsilon \leq 1$ the triangulation is composed of at most three elements. For instance, an increase in p — which will later denote the polynomial degree over all elements — leads to the two inner mesh points being moved but no further mesh point is inserted. In contrast to the parameter λ of a Shishkin mesh the parameter κ has to be sufficiently small, i.e. $\kappa \in (0,\kappa_0)$, where κ_0 depends only on the data of problem (4.1).

Next introduce the finite element space $V^p := \{v \in H_0^1(\Omega) \,:\, v|_T \in P_p(T) \text{ for all } T \in \Omega^S\}$ and the Galerkin approximation $u^p \in V^p$ determined by

$$a(u^p, v^p) = (f, v^p) \quad \text{for all } v^p \in V^p.$$

Moreover, let $\pi : L_2(\Omega) \to V^p$ denote the projector

$$\big(c(u-\pi u), v^p\big) = 0 \quad \text{for each } v^p \in V^p.$$

Similarly to the proof of Theorem 49 one shows that

$$|u-u^p|_1 \leq C|u-\pi u|_1. \tag{4.88}$$

With the application of inverse estimates in mind the Space $V^p = V_1^p + V_\varepsilon^p$ is decomposed. In the interesting case in which there are small $\mathcal{O}(\sqrt{\varepsilon})$ elements in the underlaying mesh Ω^S one sets

$$V_\varepsilon^p := \{v^p \in V^p \,:\, \operatorname{supp} v_p \subset \overline{\Omega}_\varepsilon\},$$

with $\Omega_\varepsilon := (0, \kappa p\sqrt{\varepsilon}) \cup (1-\kappa p\sqrt{\varepsilon}, 1)$ denoting the boundary region formed by the two small mesh cells. Note that in this case the decomposition $V^p = V_1^p + V_\varepsilon^p$ is a direct sum (in the other case simply set $V_1^p := V^p$). Hence, one can decompose $\pi v \in V^p$ into components:

$$\pi v = (\pi v)_1 + (\pi v)_\varepsilon, \quad \text{with unique } (\pi v)_1 \in V_1^p \text{ and } (\pi v)_\varepsilon \in V_\varepsilon^p.$$

On the other hand one may introduce the operator $\pi_\varepsilon : H^1(\Omega) \to V_\varepsilon^p$ by

$$\big(c(u - \pi_\varepsilon u), v^p\big) = 0 \quad \text{for each } v^p \in V_\varepsilon^p.$$

Now the inverse estimates improve to

$$|v^p|_1 \le Cp^2 \|v^p\|_0 \quad \text{for each } v^p \in V_1^p, \tag{4.89a}$$

$$|v^p|_1 \le C\frac{p^2}{\kappa p\sqrt{\varepsilon}} \|v^p\|_{0,\Omega_\varepsilon} \quad \text{for each } v^p \in V_\varepsilon^p. \tag{4.89b}$$

Assuming

$$p\varepsilon^{1/4}\sqrt{\kappa p} \le C \tag{4.90}$$

Melenk and Xenophontos prove the following stability results:

$$\|(\pi v)_1\|_0 \le C\|v\|_0, \qquad \|(\pi v)_\varepsilon\|_0 \le C\big(\|\pi_\varepsilon v\|_0 + (p\varepsilon^{1/4}\sqrt{\kappa p})^3\|v\|_0\big), \tag{4.91a}$$

$$\|\pi_\varepsilon v\|_0 \le C\|v\|_0, \qquad \|\pi_\varepsilon v\|_{0,\Omega_\varepsilon} \le C\kappa p\sqrt{\varepsilon}|v|_{1,\Omega_\varepsilon} \quad \text{for each } v \in H_0^1(\Omega). \tag{4.91b}$$

Remark 43. If the Assumption (4.90) is violated one can use energy norm results [70, 48] to conclude

$$\varepsilon^{1/4}|u - u^p|_1 \le \varepsilon^{-1/4}\|u - u^p\|_\varepsilon \le C\sqrt{\kappa} p^{3/2} \mathrm{e}^{-\beta p},$$

where $C, \beta > 0$ are constants independent of ε and p.

Using the inverse estimates (4.89) and the stability results (4.91) it is possible to estimate the projection error $|u - \pi u|_1$.

Lemma 52. *Let assumption* (4.90) *be satisfied. Then there are constants $C, \beta > 0$ independent of ε and p such that*

$$\varepsilon^{1/4}|u - \pi u|_1 \le C\mathrm{e}^{-\beta p}.$$

Proof. Similarly to the proof of Lemma 48 the basic idea is to use the stability of the operator π and a function with known approximation properties. In contrast to nodal interpolation Melenk and Xenophontos propose to use an approximation $u_p \in V^p$ of u defined in [70, Section 5] which achieves robust exponential convergence, i.e. (see [70, Theorem 5.1]):

$$|u - u_p|_1 \le C\varepsilon^{-1/4}\mathrm{e}^{-\beta p} \qquad \text{and} \qquad \|u - u_p\|_0 \le C\varepsilon^{1/4}\mathrm{e}^{-\beta p}. \tag{4.92}$$

Next a triangle inequality yields

$$|u - \pi u|_1 \le |u - u_p|_1 + |u_p - \pi u|_1. \tag{4.93}$$

The first summand of the right hand side of (4.93) is bounded using (4.92) and it remains to estimate the second one. By the projection property $\pi u_p = u_p \in V^p$ we have based on the decomposition of V^p that

$$u_p - \pi u = \pi(u_p - u) = \big(\pi(u_p - u)\big)_1 + \big(\pi(u_p - u)\big)_\varepsilon. \tag{4.94}$$

These terms are estimated separately. For the first one (4.89a) and (4.91a) yield

$$\Big|\big(\pi(u_p - u)\big)_1\Big|_1 \le Cp^2\Big\|\big(\pi(u_p - u)\big)_1\Big\|_0 \le Cp^2\Big\|u_p - u\Big\|_0 \le C\varepsilon^{-1/4}\mathrm{e}^{-\beta p}. \tag{4.95}$$

In the last step we used (4.92) and (4.90). For the other term in (4.94) one proceeds in a similar manner. The inverse estimate (4.89b) and the stability estimates (4.91a), (4.91b) give

$$\begin{aligned}\left|\big(\pi(u_p-u)\big)_\varepsilon\right|_1 &\le C\frac{p^2}{\kappa p\sqrt{\varepsilon}}\left\|\big(\pi(u_p-u)\big)_\varepsilon\right\|_0\\ &\le C\frac{p^2}{\kappa p\sqrt{\varepsilon}}\Big(\|\pi_\varepsilon(u_p-u)\|_{0,\Omega_\varepsilon}+(p\varepsilon^{1/4}\sqrt{\kappa p})^3\|u_p-u\|_0\Big)\\ &\le C\Big(p^2|u_p-u|_{1,\Omega_\varepsilon}+\varepsilon^{1/4}p^3\sqrt{\kappa p}\|u_p-u\|_0\Big)\le C\varepsilon^{-1/4}\mathrm{e}^{-\beta p}.\end{aligned} \tag{4.96}$$

The last inequality is due to $|u_p-u|_{1,\Omega_\varepsilon}\le C\varepsilon^{-\beta p}$ (cf. [70, (5.8)]), (4.92) and (4.90). Collect (4.93), (4.92), (4.94), (4.95) and (4.96) to complete the proof. □

Theorem 53. *Let κ be sufficiently small (independent of ε and p) and assume that* (4.90) *holds true. Then the error of the rp Galerkin FEM on a spectral boundary layer mesh satisfies*

$$\|u-u^p\|_b\le C\mathrm{e}^{-\beta p}$$

where the constants $\beta, C>0$ are independent of ε and p.

Proof. For the L_2-norm error the result is known, see [70]:

$$\|u-u^p\|_0\le\|u-u^p\|_\varepsilon\le C\mathrm{e}^{-\beta p}.$$

For the term $\varepsilon^{1/4}|u-u^p|_1$ combine (4.88) and Lemma 52. □

Corollary 54. *Under the assumptions of Theorem 53 the error of the rp Galerkin FEM on a spectral boundary layer mesh satisfies*

$$\|u-u^p\|_{L_\infty(\Omega)}\le C\sqrt{\kappa p}\mathrm{e}^{-\beta p}.$$

Proof. The proof is similar to the arguments of Subsection 4.3.2. From (4.92) one obtains (see [70])

$$\|u-u_p\|_{L_\infty(\Omega)}\le 2\|u-u_p\|_0^{1/2}|u-u_p|_1^{1/2}\le C\mathrm{e}^{-\beta p}.$$

Hence, the result follows in the interior by an inverse estimate for $\|u_p-u^p\|_{L_\infty(\Omega\setminus\Omega_\varepsilon)}$. Finally, let $x\in\Omega_\varepsilon$, for instance $x\in(0,\sqrt{\varepsilon}\kappa p)$. Then

$$|(u-u^p)(x)|=\left|\int_0^x(u-u^p)'(t)\,\mathrm{d}t\right|\le\varepsilon^{1/4}\sqrt{\kappa p}|u-u^p|_1\le C\varepsilon^{1/4}\sqrt{\kappa p}\varepsilon^{-1/4}\mathrm{e}^{-\beta p}\le C\sqrt{\kappa p}\mathrm{e}^{-\beta p}.$$

Obviously, the same argument works if $x\in(1-\sqrt{\varepsilon}\kappa p,1)$. □

Affirmation

1. Hereby I affirm, that I wrote the present thesis without any inadmissible help by a third party and without using any other means than indicated. Thoughts that were taken directly or indirectly from other sources are indicated as such. This thesis has not been presented to any other examination board in this or a similar form, neither in this nor in any other country.

2. The present thesis has been developed at the Institut für Numerische Mathematik, Departement of Mathematics, Faculty of Science, TU Dresden under supervision of Prof. H.-G. Roos.

3. There have been no prior attempts to obtain a PhD degree at any university.

4. I accept the requirements for obtaining a PhD (Promotionsordnung) of the Faculty of Science of the TU-Dresden, issued February 23, 2011 with changes in effect since June 15, 2011.

Versicherung

1. Hiermit versichere ich, dass ich die vorliegene Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

2. Die vorliegende Arbeit wurde am Institut für Numerische Mathematik, Fachbereich Mathematik, Fakultät Mathematik und Naturwissenschaften, TU Dresden unter Betreuung von Prof. H.-G. Roos angefertigt.

3. Es wurden zuvor keine Promotionsvorhaben unternommen.

4. Ich erkenne die Promotionsordnung der Fakultät Mathematik und Naturwissenschaften der TU-Dresden vom 23. Februar 2011, in der geänderten Fassung mit Gültigkeit vom 15. Juni 2011 an.

Dresden, den 14. Januar 2014.

References

[1] G. Acosta and R. G. Durán, *The Maximum Angle Condition for Mixed and Nonconforming Elements: Application to the Stokes Equations*, SIAM J. Numer. Anal. 37(1) (1999), pp. 18–36.

[2] T. Apel, *Anisotropic finite elements: local estimates and applications*, Teubner, 1999.

[3] T. Apel and M. Dobrowolski, *Anisotropic Interpolation with Applications to the Finite Element Method*, Computing, 47 (1992), pp. 277–293.

[4] T. Apel and H.-G. Roos, *Remarks on the analysis of finite element methods on a Shishkin mesh: are Scott-Zhang interpolants applicable?*, Preprint MATH-NM-06-2008, TU Dresden, 2008

[5] S. Asaturyan, *Shape preserving surface interpolation schemes*, PhD thesis, The University of Dundee, 1989.

[6] B. M. Bagaev and V. V. Shaidurov, *Numerical Methods for the Solution of Problems with a Boundary Layer*, Nauka, Novosibirsk (in Russian).

[7] A. S. Bakhvalov, *On the optimization of methods for solving boundary value problems with boundary layers* (in Russian), Zh. Vychisl. Mat. i Mat. Fis., 9 (1969) pp. 841–859.

[8] J. Bergh and J. Löfström, *Interpolation Spaces: An Introduction*, Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[9] F. Brezzi, J. Douglas, Jr., and L. D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numer. Math., 47 (1985), pp. 217–235.

[10] F. Brezzi and Michel Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Ser. Comput. Math. 15, Springer-Verlag, New York, 1991.

[11] E. Burman and A. Ern, *Continuous interior penalty hp-finite element methods for transport operators*, In: A.B. de Castro et al. (ed.) Numerical mathematics and advanced applications. Proceedings of ENUMATH 2005, the 6th European conference on numerical mathematics and advanced applications, Santiago de Compostela, Spain, July 18–22, 2005, pp. 504–511. Springer, Berlin (2006).

[12] P. Ciarlet, *The finite element method for elliptic problems*, North-Holland, Amsterdam, 1978.

[13] S. Chen, Y. Yang and S. Mao, *Anisotropic conforming rectangular elements for elliptic problems of any order*, J. Appl. Numer. Math. 59 (2009), pp. 1137–1148.

[14] M. Crouzeix and V. Thomée, *The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces*, Math. Comput. 48 (1987), pp. 521–532.

[15] S. Dekel and D. Leviatan, *The Bramble–Hilbert Lemma for Convex Domains* SIAM J. Math. Anal., 35(5), pp. 1203–1212.

[16] M. Dobrowolski and H.-G. Roos, *A priori estimates for the solution of convection- diffusion problems and interpolation on Shishkin meshes*, Z. Anal. Anwendungen, 16(4) (1997) pp. 1001–1012.

[17] M. Dobrowolski, *Finite Elemente*, lecture notes, University of Würzburg, available under http://www.mathematik.uni-wuerzburg.de/~dobro/pub/fem.pdf, (in German).

[18] V. Dolejší, M. Feistauer and C. Schwab, *A finite volume discontinuous Galerkin scheme for nonlinear convection-diffusion problems*, Calcolo 39 (2002), pp. 1–40.

[19] J. Douglas, Jr., *H^1-Galerkin Methods for a Nonlinear Dirichlet Problem*, Mathematical Aspects of Finite Element Methods, Lecture Notes in Mathematics, 606, Springer-Verlag, 1977, pp. 64–86

[20] J. Douglas, Jr., T. Dupont, P. Percell and R. Scott, *A family of $C1$ finite elements with optimal approximation properties for various Galerkin methods for 2nd and 4th order problems*, RAIRO Modél. Math. Anal. Numér. 13 (1979), pp. 227–255.

[21] J. Douglas, Jr., T. Dupont and M. F. Wheeler, *H^1-Galerkin Methods for the Laplace and Heat Equations*, Mathematical Aspects of Finite Elements in Partial Differential Equations, ed. C. De Boor. Academic Press, 1974, pp. 383–416

[22] R. G. Durán, A. L. Lombardi, *Finite element approximation of convection diffusion problems using graded meshes*, Appl. Numer. Math. 56 (2006), pp. 1314–1325.

[23] R. G. Durán, A. L. 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 (2013), pp. 232–247.

[24] P. A. Farrell, A. F. Hegarty, J. J. H. Miller, E. O?Riordan and G. I. Shishkin, *Robust Computational Techniques for Boundary Layers*, Chapman-Hall/CRC, 2000.

[25] S. Franz and H.-G. Roos, *Error estimation in a balanced norm for a convection-diffusion problem with characteristic boundary layers*, Calcolo (2013), doi:10.1007/s10092-013-0093-5.

[26] I. Fried, *Bounds on the spectral and maximum norms of the finite element stiffness, flexibility and mass matrices*, Int. J. Solids Structures 9 (1973), pp. 1013–1034.

[27] E. C. Gartland, *Graded-mesh difference schemes for singularly perturbed two-point boundary value problems*, Math. Comp. 51 (1988), pp. 631–657.

[28] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman Advanced Publishing Program, Boston, 1985.

[29] 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. Numer. Anal. 21 (1990), pp. 394–408.

[30] M. Hassan, *C^1-quadratische Splines und ihr Einsatz bei der Mehrgitter-Finite-Elemente-Approximation*, PhD thesis, Technical University of Dresden, 1997.

[31] J. Hu, Y. Huang and S. Zhang, *The Lowest Order Differentiable Finite Element on Rectangular Grids*, SIAM J. Numer. Anal. 49 (2011), pp. 1350–1368.

[32] R. B. Kellogg and M. Stynes, *Corner singularities and boundary layers in a simple convection?diffusion problem*, J. Differential Equations 213 (2005), pp. 81–120.

[33] R. B. Kellogg and M. Stynes, *Sharpened bounds for corner singularities and boundary layers in a simple convection?diffusion problem*, Appl. Math. Lett. 20 (2007), pp. 539–544.

[34] N. Kopteva, *Linear finite elements may be only first-order pointwise accurate on anisotropic triangulations*, Math. Comp. (2013), accepted for publication.

[35] N. Kopteva, *Maximum norm error analysis for a 2d singularly perturbed semilinear reaction-diffusion problem*, Math. Comput. 76(258) (2007), pp. 631–646.

[36] N. Kopteva and M. Stynes, *Stabilised approximation of interior-layer solutions of a singularly perturbed semilinear reaction-diffusion problem*, Numer. Math. 119(4) (2011), pp. 787–810.

[37] D. Leykekhman, *Uniform error estimates in the finite element methods for a singularly perturbed reaction-diffusion problem*, Math. Comp. 77 (2008), pp. 21–38.

[38] J. Li, *Convergence and superconvergence analysis of finite element methods on highly nonuniform anisotropic meshes for singularly perturbed reaction?diffusion problems*, Appl. Numer. Math. 36 (2001), pp. 129–154.

[39] J. Li, *Uniform error estimates in the finite element method for a singularly perturbed reaction-diffusion problem*, Appl. Numer. Math. 36 (2001), pp. 129–154.

[40] 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) (2000), pp. 770–798.

[41] R. Lin, *Discontinuous discretization for least-squares formulation of singularly perturbed reaction-diffusion problems in one and two dimensions*, SIAM J. Numer. Anal. 47(1) (2008/09), pp. 89–108.

[42] R. Lin and M. Stynes, *A balanced finite element method for singularly perturbed reaction-diffusion problems*, SIAM J. Numer. Anal. 50 (2012), pp. 2729–2743.

[43] T. Linss, *Layer-adapted meshes for reaction-convection-diffusion problems*, Lecture Notes in Mathematics, vol 1985, Springer, 2010.

[44] T. Linss, *Uniform superconvergence of a Galerkin finite element method on Shishkin-type meshes*, Numer. Methods Partial Differ. Equations 16 (2000), pp. 426–440.

[45] F. Liu, N. Madden, M. Stynes and A. Zhou, *A two-scale sparse grid method for a singularly perturbed reaction-diffusion problem in two dimensions*, IMA J. Numer. Anal. 39(4) (2009), pp. 986–1007.

[46] S. MacLachlan and N. Madden, *Robust Solution of Singularly Perturbed Problems Using Multigrid Methods*, SIAM J. Sci. Comput., 35(5) (2013), A2225–A2254, (30 pages).

[47] N. Madden and M. Stynes, *Efficient generation of oriented meshes for solving convection-diffusion problems*, Int. J. Numer. Methods Engrg. 40 (1997), pp. 565–576.

[48] J. M. Melenk, *hp-finite element methods for singular perturbations*, Springer, 2002.

[49] J. J. H. Miller, E. O'Riordan and G. I. Shishkin, Fitted numerical methods for singular perturbation problems. Error estimates in the maximum norm for linear problems in one and two dimensions, World Scientific, Singapore, 1996.

[50] K. W. Morton, *Numerical Solution of Convection-Diffusion Problems*, Chapman & Hall, London, 1996.

[51] S. Natesan and N. Ramanujam, *Booster method for singular perturbed one-dimensional reaction-diffusion Neumann problems*, J. Optim. Theory and Appl. 104 (2000), pp. 175–194.

[52] A. Naughton, R. B. Kellogg and M. Stynes, *Regularity and derivative bounds for a convection-diffusion problem with a Neumann outflow condition*, J. Differential Equations 247(9) (2009), pp. 2495–2516.

[53] G. Nürnberger and F. Zeilfelder, *Developments in bivariate spline interpolation*, J. Comp. Appl. Math. 121 (2000), 125–152.

[54] P. Oswald, L_∞-bounds for the L_2-projection onto linear spline spaces, in: Recent advances in harmonic analysis and applications. In honor of Konstantin Oskolkov on the occasion of his 65th birthday. Based on the AMS sectional meeting, Georgia Southern University, Statesboro, GA, USA, March 11–13, 2011, Springer, 2013, pp. 303–316.

[55] O. T. Pop and D. Barbosu, *Two dimensional divided differences with multiple knots*, An. St. Univ. Ovidius Constanta 17 (2009), pp. 181–190.

[56] M. J. D. Powell and M. A. Sabin, *Piecewise quadratic approximations on triangles*, ACM Trans. Math. Software, 3–4 (1977), pp. 316–325.

[57] H.-G. Roos, *Optimal convergence of basic schemes for elliptic boundary value problems with strong parabolic layers*, J. Math. Anal. Appl. 267 (2002), pp. 194–208.

[58] H.-G. Roos, *Robust numerical methods for singularly perturbed differential equations: a survey covering 2008-2012*, ISRN Applied Mathematics, vol. 2012 (2012), Article ID 379547, 30 pages, doi:10.5402/2012/379547.

[59] H.-G. Roos and T. Linss, *Sufficient conditions for uniform convergence on layer-adapted grids*, Computing, 63(1) (1999), pp. 27–45.

[60] H.-G. Roos and M. Schopf, *Analysis of finite element methods on Bakhvalov-type meshes for linear convection-diffusion problems in 2D*, Appl. Math., 57(2) (2012), pp. 97–108.

[61] H.-G. Roos and M. Schopf, *Convergence and stability in balanced norms of finite element methods on Shishkin meshes for reaction-diffusion problems*, to appear in ZAMM, Z. Angew. Math. Mech.

[62] H.-G. Roos and M. Schopf, *Finite elements on locally uniform meshes for convection-diffusion problems with boundary layers*, Computing, 92 (2011), pp. 285–296.

[63] H.-G. Roos and M. Schopf, *Error estimation in energy norms: Is it necessary to fit the mesh to boundary layers?*, Dimov, Farago, Vulkov (Eds.): NAA 2012, LNCS 8236 (2013), pp. 95–109.

[64] H.-G. Roos, M. Stynes and L. Tobiska, *Robust numerical methods for singularly perturbed differential equations*, Springer, 2008.

[65] R. Sacco, E. Gatti and L. Gotusso, *A nonconforming exponentially fitted finite element method for two-dimensional drift-diffusion models in semiconductors*, Numer. Methods Partial Differential Equations 15 (1999), p. 133–150.

[66] R. Sacco and M. Stynes, *Finite element methods for convection-diffusion problems using exponential splines on triangles*, Computers Math. Applic. 35 (1998), pp. 35–45.

[67] A. H. Schatz and L. B. Wahlbin, *On the finite element method for singularly perturbed reaction-diffusion problems in two and one dimensions*, Math. Comp. 40(161) (1983), pp. 47–89.

[68] F. Schieweck, Eine asymptotisch angepaßte Finite-Element-Methode für singulär gestörte elliptische Randwertaufgaben, PhD thesis, Technical University Magdeburg, 1986.

[69] L. L. Schumaker, *On shape-preserving quadratic spline interpolation*, SIAM J. Numer. Anal. 20 (1983), pp. 854–864.

[70] C. Schwab and M. Suri, *The p and hp versions of the finite element method for problems with boundary layers*, Math. Comp. 65 (1996), pp. 1403–1429.

[71] L. R. Scott and S. Zhang, *Finite element interpolation of non-smooth functions satisfying boundary conditions*, Math. Comp. 54 (1990), pp. 483–493.

[72] G. I. Shishkin, *Approximation of solutions of singularly perturbed boundary value problems with a parabolic boundary layer*, USSR Comput. Maths. Math. Phys. 29(4) (1989), pp. 1–10.

[73] M. Stynes *Steady-state convection-diffusion problems*, Acta Numerica, 14 (2005), pp. 445–508 doi:10.1017/S096249290400.

[74] M. Stynes and E. O'Riordan, *A uniformly convergent Galerkin method on a Shishkin mesh for a convection-diffusion problem*, J. Math. Anal. Appl. 214 (1997), pp. 36–54.

[75] M. Stynes and L. Tobiska, *Analysis of the streamline-diffusion finite element method on a Shishkin mesh for a convection-diffusion problem with exponential layers*, Fakultät für Mathematik, Otto-von-Guericke Universität Magdeburg, Preprint 17, 1998.

[76] M. Stynes and L.Tobiska, *The SDFEM for a convection-diffusion problem with a boundary layer: optimal error analysis and enhancement of accuracy*, SIAM J. Numer. Anal. 41 (2003), pp. 1620–1642.

[77] H. Triebel, *Interpolation theory, function spaces, differential operators*, Johann Ambrosius Barth, Heidelberg, second ed., 1995.

[78] C. Xenophontos, *private communication.*

[79] A. Ženižek, *Interpolation polynomials on the triangle*, Numer. Math. 15 (1970), pp. 238–296.

[80] Z. Zhang, *Finite element superconvergence on Shishkin mesh for 2-d convection-diffusion problems*, Math. Comp. 72 (2003), pp. 1147–1177.

[81] G. Zhu and S. Chen, *Convergence and superconvergence analysis of an anisotropic nonconforming finite element methods for singularly perturbed reaction?diffusion problems*, J. Comput. Appl. Math. 234 (2010), pp. 3048–3063.

[82] G. Zhu and S. Chen, *Convergence and superconvergence analysis of finite element methods on graded meshes for singularly and semisingularly perturbed reaction-diffusion problems*, J. Comput. Appl. Math. 220 (2008), pp. 373–393.

[83] M. Zlámal, *On the finite element method*, Numer. Math. 12 (1968), pp. 394–409.