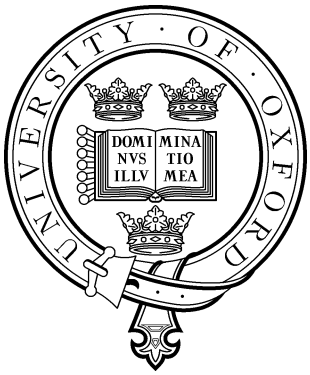
The Residual–Free Bubble Method for Problems with Multiple Scales



Andrea Cangiani
St Hugh's College
University of Oxford

Trinity Term 2004

Thesis submitted for the degree of Doctor of Philosophy at the University of Oxford

Oxford University Computing Laboratory
Numerical Analysis Group
Parks Road
Oxford, OX1 3QD

Abstract

The Residual–Free Bubble Method for Problems with Multiple Scales

Andrea Cangiani
St Hugh's College

Doctor of Philosophy
Trinity Term 2004

This thesis is devoted to the numerical analysis and development of the residual–free bubble finite element method.

We begin with an overview of known results and properties of the method, showing how techniques used on a range of multiscale problems can be cast into the framework of the residual–free bubble method.

Further, we present an *a priori* error analysis of the method applied to convection–dominated diffusion problems on anisotropic meshes. The result has implications for the problem of parameter–tuning in classical stabilised finite element methods (for instance, the streamline–diffusion finite element method). We show how the local SD–parameter should be chosen on meshes with high aspect–ratios.

A new algorithm named RFBe (*enhanced residual–free bubble method*) is proposed for the resolution of boundary layers on coarse meshes. The residual–free bubble finite element space is augmented locally by *ad hoc* bubble functions with support on two elements sharing a particular edge. The idea is presented in a general framework to highlight its applicability to a wide range of multiscale problems.

Finally, we derive an *a posteriori* error estimate for the method and describe an associated adaptive algorithm designed to minimise the computational effort required for reducing the error below a prescribed tolerance. Both norm–error and linear–functional–error bounds are considered. We also propose an automatic procedure for switching off bubble stabilisation locally during the mesh refinement process (*hb*–adaptivity) to further reduce the computational cost.

Acknowledgments

A DPhil thesis is a relatively long project. For some it is just a job, for many a full involvement. Maybe this is the reason why authors feel the need to look back to those years and appreciate the help of both those that have contributed actively and those that where simply there, sharing our lives. Finally realising that the two categories are mixed up.

My supervisor Endre Süli is a perfect example: his contribution to this thesis has been crucial and extremely professional. His patience in dealing with my characteristic latin approach to the job priceless.

Still, the first prize goes to my mother: to her this thesis is dedicated.

A special mention goes to my wife Johanna, for supporting me during the *periodo especial* of writing–up, while it was our honey moon that she deserved.

I thank my father Michele and brother Giovanni for being my greatest supporters, my brother being, as well, a great colleague.

The Computing Laboratory has been a very pleasant place to work, and Oxford a very pleasant place to live. Throughout these years I met many friends and helpful collaborators, like my office mates Max, Manolis and Andris. And had some wonderful teachers in M. Giles, N. Trefethen and A. Wathen.

This thesis has greatly benefited from the collaboration of M. Picasso and G. Sangalli and the inspiring ideas of F. Brezzi.

To my mother Alessandra.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The residual–free bubble (RFB) method is a two–level finite element method originally introduced for the stable and accurate approximation of numerical solutions to convection–dominated diffusion problems. The potential of the method, though, is better appreciated by interpreting it as a general procedure for the solution of multiscale problems.

In the following sections we outline the difficulties encountered in the numerical solution of multiscale problems and describe a class of techniques used to overcome such difficulties. Special attention is devoted to the connections of the techniques described with the RFB method. The importance of convection–diffusion type equations in applied mathematics and numerical analysis, still in the context of the application of RFB, is also discussed.

The chapter concludes with an outline of this thesis.

1.1 Multiscale problems

Many computationally challenging problems that arise in science and engineering exhibit *multiscale behaviour*. The list is long, including turbulent transport in high Reynolds number flows, flow through porous media, structural analysis of composite and foam materials, fine–scale laminates and crystalline microstructures, weather forecasting and large–scale molecular dynamic simulations, to mention just a few.

Sometimes the data of the model is incomplete (for example, the detailed properties of the underlying media may not be known) and *averaging techniques* need to be employed to 'construct' a coarse model that can be solved (see, for example, the review article by Farmer [43] on *upscaling*[1] in porous media).

However, we focus here on multiscale problems admitting a representation through a given differential operator with known coefficients and on their approximate solution using *finite element methods* (FEM).

[1] In *upscaling* a simple model is postulated, as when postulating a model for a new physical situation, and then validated through physical experiments or through a few numerical experiments on the detailed model.

Classical computational algorithms are designed to operate at a certain *preselected scale* fixed by the choice of a *discretisation parameter*. If the scales in the model are very diverse, representing and numerically computing on all physical scales results in excessive algorithmic complexity, both in terms of computation times and memory requirements. Moreover, often one is only interested in quantifying some coarse–scale or *macroscopic* features of the problem.

In the presence of nonlinearities in the model, if the discretisation at a coarse level ignores the fine scale information, then the solution might not be physically meaningful. In other words, for a *coarse scale model* to be successful, the influence of the fine, unresolvable scales on coarse scales must be incorporated.

It is, then, natural to follow a *two–level approach*: first, the effect of the unresolvable (*subgrid*) scales is incorporated into the coarse scale model which is subsequently solved approximately using the preferred analytical or numerical method.

This is the idea behind the classical *theory of homogenization*, *upscaling* and their numerical counterparts. Homogenization (see [13], [34]) refers to a group of techniques aimed at replacing a detailed mathematical model by a simpler one that can reproduce, in some asymptotic limit, some average behaviour of the detailed model. The most common approach is the method of multiple scales, where after appropriate scaling an expansion is derived. In particular, an *effective* operator is obtained which approximates the exact operator by taking the effect of fine scales into account without containing them explicitly. The solution in terms of the effective operator should provide an approximation of the coarse scale behaviour of the solution of the original problem. The main application of this technique is in the study of flow through porous media.

Numerical homogenization or upscaling follows the same pattern, except that the effective operator is computed numerically (see, for example, Babuška *et al.* [8], the multiscale finite element method (MFEM) of Hou *et al.* [59], [60], the heterogeneous multiscale method (HMM) of Engquist and E [37], the wavelet–based numerical homogenization of Engquist and Runborg [40], the upscaling technique of Durlofsky [36] and Arbogast [6] and *local–local* upscaling in general [43]).

Other examples of two–level techniques are given by Franca *et al.* [45], [47] dealing with Helmholtz equations at high wave number, the generalised p–FEM of Matache *et al.* [70], [92], and the two–scale FEM of Matache *et al.* [71] in structural analysis of composite and foam materials and Babuška and coauthors [7] in damage analysis of fiber composites.

Another common characteristic of many of the methods just mentioned is the use of ad–hoc approximation spaces (a common idea is, for example, to define the FEM basis functions through the differential operator at hand). It is understood that to incorporate the effect of the *subgrid* scales some extra resolution properties must be included into the approximation space.

This is certainly a characteristic of the more general variational multiscale method of Hughes *et al.* [65], [64] and the related residual–free bubble method of Brezzi and Russo [29], and Franca and Russo [50] which is the subject of this thesis.

These methods have been introduced to stabilize standard FEMs in convection–dominated diffusion problems with fine scales represented by sharp layers of rapid variation of the solution.

1.2 The residual–free bubble method

Given a bounded polygonal domain Ω in $\mathbb{R}^n$, let $\mathcal{L}(\cdot,\cdot)$ be a bilinear functional defined over some function space, for example $V = H_0^1(\Omega)$. We consider the task of solving numerically the boundary value problem in variational form

$$\begin{cases} \text{find } u \in V \text{ such that} \\ \mathcal{L}(u,v) = (f,v) \qquad \forall v \in V. \end{cases} \tag{1.1}$$

Throughout this thesis $(\cdot,\cdot)$ denotes the inner product in $L_2(\Omega)$.

Assume that a partition $\mathcal{T}_h$ (throughout this thesis, h represents the maximum elemental diameter) of Ω is given which is *conforming*, i.e. any two elements in $\mathcal{T}_h$ either have a common face of dimension $< n$ or they do not intersect at all. Further, let Σ be the skeleton of the partition $\mathcal{T}_h$, i.e. the union of the boundaries of all elements in $\mathcal{T}_h$.

Given a piecewise polynomial finite element space V_h on the partition $\mathcal{T}_h$, the *residual free bubble* (RFB) space V_{RFB} is defined by augmenting V_h with the space of all *bubbles*, i.e. all functions with support in $\Omega \setminus \Sigma$. That is,

$$V_{RFB} = V_h + B_h,$$

where

$$B_h = \bigoplus_{T \in \mathcal{T}_h} H_0^1(T).$$

This choice is made in order to ensure that all the subgrid scales are representable by the space V_{RFB}. The RFB method is the Galerkin formulation of (1.1) on V_{RFB}, namely

$$\begin{cases} \text{find } u_{RFB} \in V_{RFB} \text{ such that} \\ \mathcal{L}(u_{RFB},v) = (f,v) \qquad \forall v \in V_{RFB}. \end{cases} \tag{1.2}$$

As we will see in detail in the next chapter, starting from (1.2) a two–level procedure is obtained by splitting the solution into its polynomial component $u_h \in V_h$ and bubble component $u_b \in B_h$ and *testing* separately in V_h and B_h. At the subgrid level the bubble component of the solution is obtained by solving the *bubble equation*

$$\mathcal{L}(u_b,v) = (f,v) - \mathcal{L}(u_h,v) \qquad \forall v \in B_h.$$

This can be done locally. Formally we write

$$u_b|_T = L_T^{-1}(f - Lu_h)|_T \qquad \forall T \in \mathcal{T}_h,$$

where L and f are the linear differential operator and function associated with $\mathcal{L}$ and F, respectively. The second step consists of the solution in terms of the polynomial component, which has to satisfy

$$\mathcal{L}(u_h, v_h) + \sum_{T \in \mathcal{T}_h} (L_T^{-1}(f - Lu_h), L^* v_h)_T = (f, v_h) \qquad \forall v_h \in V_h; \tag{1.3}$$

here L^* is the differential operator adjoint to L. This is the coarse scale formulation given by the RFB method.

In practice, the actual computation of the bubble, hidden here in the formal local inversion of L, is carried out numerically by introducing a subgrid. In this way a fully discrete procedure is obtained. The choice of the subgrid dictates which fine scales are incorporated into the coarse scale formulation.

Originally, the residual–free bubble method was defined as a stable FEM for the solution of convection–dominated diffusion problems following the analogy of using bubble functions to stabilise mixed finite element approximations [19].

Only recently, in [20] and [16], it became apparent that the RFB method should be seen as a general technique for solving multiscale problems. Most notably Brezzi and Marini [26] demonstrated that the RFB method recovers the MFEM of Hou *et al.* [59].

The MFEM method is a low–order finite element method for the solution of the classical homogenization problem

$$-\nabla \cdot (d \nabla u) = f, \tag{1.4}$$

subject to appropriate boundary conditions, where the components of the tensor d are rapidly oscillating functions over the computational domain. For instance, this equation can be used to model single–phase flow in porous media and heat or electrical conduction in composite materials. When the symmetric *conductivity tensor* a is oscillatory with periodic oscillations, it is well–known from the theory of homogenization that the *effective* equation corresponding to (1.4) is of the same form as (1.4), with the effective conductivity being symmetric *and* constant (see, e.g. [13], [58] and [34]). Unfortunately, the effective coefficient is, in general, out of reach, and numerical upscaling techniques need to be employed. In particular, the MFEM for upscaling (1.4) consists of using the standard Galerkin formulation by employing ad–hoc basis functions obtained by solving (1.4) element–wise with zero forcing term. As detailed in Section 2.5 of this thesis, the basis functions of the MFEM method are contained in the RFB space V_{RFB} and are the sum of a basis function of the space V_h and a bubble function.

Other problems to which the RFB method has been applied include the Stokes problem [49], the incompressible Navier–Stokes equation [31] and the Helmholtz equation [45]. Moreover, some results concerning the *a posteriori* error analysis of the residual–free bubble method and of the use of the bubble u_b as error estimator can be found in [86] and [84], respectively.

1.2.1 A general framework

An attempt to incorporate the two–scale FEMs described above into a general framework is found in Brezzi and Marini [25].

The idea behind the *general augmented space framework* proposed in [25] is to design the most comprehensive method possible on a fixed partition. The partition may have been fixed to suit the coarse scales that need to be captured or to the computing power at hand.

To illustrate the method, we consider again the abstract problem in variational form (1.1). Let Φ be the space of traces of V on the skeleton of the partition Σ. A general *augmented subspace* is defined by considering all the extensions onto $\bar{\Omega}$ from a finite–dimensional subspace of Φ. That is, given a subspace $\Phi_h \subset \Phi$ of finite dimension, we define the space

$$V_a := \left\{ v \in V \,:\, v_{|\Sigma} \in \Phi_h \right\}.$$

The RFB finite element space is clearly given by the choice $\Phi_h = V_h|_\Sigma$, hence the RFB space is an instance of the augmented subspace.

Since we aim to be as general as possible, the fact that the space of traces Φ_h is fixed may seem arbitrary: though the space of traces may itself be subsequently modified where this is deemed necessary. However, the only limitation is that only approximation spaces *conforming* with the partition can be recovered.

The augmented space formulation is obtained, as for RFB, by restricting (1.1) to V_a:

$$\begin{cases} \text{find } u_a \in V_a \text{ such that} \\ \mathcal{L}(u_a, v_a) = (f, v_a) \qquad \forall v_a \in V_a. \end{cases} \tag{1.5}$$

Formally (1.5) looks like a Galerkin finite element method, although, since V_a is infinite–dimensional, it represents more a framework for devising methods than a numerical method as such.

As we have seen with the RFB method, the idea is that a specific (fully discrete) numerical method may be recovered from (1.5) by discretising V_a appropriately,

Examples of a technique which emerge from the general augmented space formulation that do not fall into the RFB framework are the MFEM method with oscillatory traces [59], [60], the generalised Galerkin methods in [74], the method in [9] and the upscaling technique of [6]; see [25] for details.

1.3 Convection–dominated–diffusion problems

The model problem used throughout this thesis is the steady–state convection–diffusion equation.

From the incipit of K. W. Morton's book [76] entitled '*Numerical Solution of Convection–Diffusion Problems*', we learn 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". Convection–diffusion processes are indeed central to scientific fields as diverse as physiology, oil reservoir simulation, aerodynamics, meteorology and financial modelling. Often, convection dominates diffusion leading to an essentially hyperbolic (i.e. elliptic *singularly perturbed*) set of equations. Techniques which are effective for the numerical solution of elliptic problems with dominant diffusion, such as classical finite element methods, encounter great difficulties when convection prevails. In this case, localized fine scale phenomena in the form of propagating near–shocks and sharp transition layers arise, and their treatment proves to be a challenging computational task.

The evolution of convection–diffusion processes is modelled by the parabolic *convection–diffusion equation* which, in its simplest form, reads

$$u_t - \varepsilon \Delta u + \boldsymbol{a} \cdot \nabla u = f, \tag{1.6}$$

where $\boldsymbol{a}$ is a velocity field, ε the diffusion coefficient and f is a forcing term. The solution u may represent the concentration of a pollutant in a fluid whose motion is described by $\boldsymbol{a}$.

For instance, the finite element treatment of the time–dependent equation (1.6) may be based on a space–time finite element mesh, with basis functions continuous in space but discontinuous in time to permit the use of different spatial meshes as the solution evolves. Due to the singularly perturbed nature of the interaction between convection and diffusion, the choice of the spatial finite element method is a subtle point.

Indeed, the main difficulties are already encountered in the numerical treatment of the steady–state version of equation (1.6), namely

$$-\varepsilon \Delta u + \boldsymbol{a} \cdot \nabla u = f. \tag{1.7}$$

Equation (1.7) will be our model problem throughout this thesis.

The approximation of steady problems is considered to be the core difficulty in the subject of the numerical solution of convection–diffusion problems. For instance, equation (1.7) is a fundamental model problem for the field of computational fluid dynamics since its stable and accurate solution is a crucial step in the treatment of the Oseen problem and, ultimately, of the incompressible Navier–Stokes equations.

For a complete survey of numerical methods for the solution of (1.7) see, for example, Morton [76] and Roos, Stynes and Tobiska [82]. Common techniques are *exponential fitting*, *symmetrization*, *upwinding* and *least squares regularisation*. Ad–hoc meshing, like *graded meshes* [93] and *Shishkin type meshes* [69], and *adaptive mesh refinement* (see, e.g., [41], [11], [4] and [55]) are also, of course, at the center of the subject. Finally, *bubble stabilisation* constitutes an illuminating way of reinterpreting many of the techniques just mentioned. An example is given by the equivalence of the RFB method with the influential *streamline upwind Petrov–Galerkin* (SUPG) method introduced by Hughes and Brooks [63] (see also Johnson and Nävert [66] and Hughes and Brooks [30]). The idea behind the SUPG method is to add to the

standard Galerkin formulation a diffusion term in the direction of convection:

$$\mathcal{L}(u_h, v_h) + \sum_{T \in \mathcal{T}_h} \tau_T (f - Lu_h, -\boldsymbol{a} \cdot \nabla v_h)_T = (f, v_h) \qquad \forall v_h \in V_h, \tag{1.8}$$

(the crucial term is $(\boldsymbol{a} \cdot \nabla u_h, \boldsymbol{a} \cdot \nabla v_h)$, the other extra term is added to preserve consistency), with the aim of suppressing the nonphysical numerical oscillations which would otherwise arise.

By comparing (1.8) with (1.3) we immediately see that the *stabilisation term* introduced by the SUPG method and the RFB method are identical if v_h is linear on every element and the effect of the local inversion L_T^{-1} reduces to multiplication by a constant (which turns out to be the case if $\boldsymbol{a}$ and f are constant inside every element).

In (1.8), τ_T represents a parameter (the SD–parameter) that needs to be tuned by the user. Notice that this has no equivalent in (1.3): we say that the RFB method is *parameter–free*.

1.4 Achievements and outline

The aim of this thesis is to show that the tools and results typical of standard FEMs can also be applied in the RFB framework. We consider, in particular, the *a priori* analysis of the method both on *shape regular* and *anisotropic* partitions, the *a posteriori* analysis of the method and the application of adaptive mesh refinement techniques.

Moreover, we show that it is possible to drastically improve the approximation properties of the RFB method by interpreting it as an augmented subspace method and locally enriching the skeleton space Φ_h with *edge bubbles* defined through an appropriate restriction of the differential operator of the problem.

1.4.1 Outline

The RFB method is rigorously defined in Chapter 2. We then present a result about the convergence properties of the method applied to second–order elliptic p.d.e.'s in divergence form on shape regular partitions, due to Brezzi, Marini and Süli [28]. We also report on the equivalence of the RFB method with the streamline–diffusion or SUPG method and with the MFEM method. The chapter concludes with notes about the implementation of the method and, in particular, about the computation of the bubble part of the RFB solution u_b.

Subsequent chapters contain the original contributions of this thesis. In particular, in Chapter 3 we perform the error analysis of RFB on anisotropic meshes. Moreover, we show that the RFB method delivers the correct choice for the local SD–parameter in classical stabilised methods.

A general methodology for improving the RFB method is proposed in Chapter 4. We show that the introduction of edge bubbles, i.e. functions with support on two elements sharing an edge, improves the stability and accuracy of the method. The edge bubbles are introduced only on those edges which cross the boundary layer. A refined, specialized, pre–asymptotic (precisely, in the regime $\varepsilon \ln(1/\varepsilon) < ch$ for a given constant c) *a priori* error analysis

demonstrates the weaknesses of the RFB method which are contrasted with the superior (pre–asymptotic) approximation properties of the new RFBe method.

Finally, in Chapter 5, we perform the *a posteriori* error analysis of the RFB method in the case of linear functionals of the solution, as well as in the case of norms. The efficiency of an adaptive mesh refinement algorithm, based on the *a posteriori* error bound obtained, is also discussed. In particular, we emphasize the relevance of the information contained in the bubble part of the RFB solution in *a posteriori* error estimation. In the last section of the chapter we propose a new '*hb–adaptive*' algorithm which, while the mesh is refined, automatically decides whether or not the local evaluation of the bubble (equivalently, the introduction of a stabilisation term) is crucial; when this is deemed not to be the case, the bubble part of the solution is locally omitted.

All of the results presented in this thesis are corroborated by numerical experiments.

The thesis concludes with Chapter 6 where we summarize the work presented and we discuss directions for further research.

Chapter 2

The residual–free bubble method

This chapter is devoted to the definition and the mathematical analysis of the *residual–free bubble* (RFB) method.

In particular, we present two definitions of the method which are ultimately equivalent. The first definition is based on the classical framework of *bubble methods* in which the given finite element space V_h is locally *augmented* with a finite–dimensional space of *bubbles*, i.e. functions with support contained in the elements' interior. Based on such definition, the first *a priori* analysis of the method, restricted to linear elements, was presented by Brezzi *et al.* in [24]. Their analysis recovers the bounds already known from the theory of stabilised methods on the linear part of the solution. Moreover, it has the advantage of producing analogous bounds on the full (augmented) RFB solution.

The more general error bound which we present here, due to Brezzi, Marini and Süli [28], is based on the second definition of the method, which was suggested by the equivalence of the RFB method with the variational multiscale method of Hughes [62]. In this case, the RFB finite element space is defined as the space of all admissible extensions from the set of piecewise polynomial traces on the skeleton of the triangulation. This definition also inspired the *generalised augmented space formulation* by Brezzi and Marini [25] which we also present.

Sangalli [85] has subsequently obtained results similar to those in [28] following a different argument (see the next chapter) which allowed him to bound the L_2–norm of the error and to establish local error bounds as well.

In later sections, we introduce the classical stabilised methods for the solution of steady–state convection–diffusion problems and discuss their relation with two–level methods such as RFB. The equivalence of RFB with the MFEM method of Hou *et al.* [59], [60] is also detailed.

The final section of the chapter is dedicated to implementational issues.

2.1 Definition of the method

Let Ω be a bounded polygonal domain in $\mathbb{R}^n$. Given the linear second–order elliptic boundary value problem

$$\begin{cases} Lu = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.1}$$

we consider its variational formulation

$$\begin{cases} \text{find } u \in V \text{ such that} \\ \mathcal{L}(u, v) = (f, v) \qquad \forall v \in V, \end{cases} \tag{2.2}$$

assuming that $V = H_0^1(\Omega)$ and $\mathcal{L}(\cdot, \cdot)$ is a continuous and coercive bilinear functional on $V \times V$.

Let V_h be a finite–dimensional subspace of V, such as a conforming finite element space of piecewise polynomials defined over a partition $\mathcal{T}_h$ of Ω. The *standard Galerkin* method for the solution of (2.2) consists of approximating the variational formulation (2.2) from V_h.

Suppose that we are not satisfied with the approximation properties of V_h, but, at the same time, we are not willing to simply approximate (2.2) on a larger space. For example, the partition $\mathcal{T}_h$ may be the finest we are prepared to work with.

A possible way out is to modify V_h by (temporarily) considering an *augmented space* V_a obtained by adding to V_h some new degrees of freedom internal to the finite elements. We can think of the new elemental functions as being defined through a subgrid or as a set of *bubble* functions. The main point is that we want such new degrees of freedom to be defined element–wise so that they can be successively eliminated by *static condensation*. In this way, the resulting formulation will be in terms of the initial finite element space. As we shall see, adding and eliminating bubbles is equivalent to modifying the bilinear form $\mathcal{L}$; ultimately, we wish to enhance the stability and accuracy of the underlying numerical scheme.

As a first step in this direction, we give the definition of *bubble* proposed by Brezzi *et al.* in [21].

Definition 2.1.1. *A bubble (in V, on $\mathcal{T}_h$) is a function $b \in V$ such that the support* $\text{supp}(b)$ *of b is contained in a single element $T \in \mathcal{T}_h$.*

Let $\widetilde{B}_h \subset V$ be a finite dimensional *bubble space* on $\mathcal{T}_h$; that is,

$$\widetilde{B}_h = \bigoplus_{T \in \mathcal{T}_h} B_T,$$

where, for any $T \in \mathcal{T}_h$, the elements of B_T are characterised by having support in a subset of T, and B_T is finite dimensional.

Remark. In this thesis we often regard bubble functions as functions on a given element. In this case extension by zero elsewhere is assumed.

Notice that the space V_h may already contain bubble functions. For example, if V_h is the space of all piecewise cubic continuous functions, then, for each $T \in \mathcal{T}_h$, V_h contains a one–dimensional subspace of cubic bubbles defined on element T. Since such bubbles can be eliminated through static condensation, it makes sense to think of them as belonging to the bubble space $\widetilde{B}_h$ rather then to V_h. Keeping this in mind, we consider the *augmented space* V_a defined by

$$V_a = V_h \oplus \widetilde{B}_h, \tag{2.3}$$

so that any element of V_a admits a unique decomposition into the sum of an element of V_h and an element of $\widetilde{B}_h$.

We name *bubble approach* or *augmented space approach* the Galerkin formulation of (2.2) on V_a, that is,

$$\begin{cases} \text{find } u_a \in V_a \text{ such that} \\ \mathcal{L}(u_a, v_a) = (f, v_a) \qquad \forall v_a \in V_a. \end{cases} \tag{2.4}$$

Since in (2.3) we have a direct sum, it follows that we have the unique decompositions $u_a = u_h + u_b$ and $v_a = v_h + v_b$ with u_h, $v_h \in V_h$ and u_b, $v_b \in \widetilde{B}_h$. Moreover, for any element $T \in \mathcal{T}$, we can write

$$v_b|_T = v_{b,T} \quad \text{with} \quad v_{b,T} \in B_T,$$

and similarly for u_b. Thus, by testing separately in V_h and then in $\widetilde{B}_h$, (2.4) can be re–stated as follows:

$$\begin{cases} \text{find } u_a = u_h + u_b = u_h + \displaystyle\sum_{T \in \mathcal{T}_h} u_{b,T} \in V_a \text{ such that} \\ \mathcal{L}(u_h, v_h) + \displaystyle\sum_{T \in \mathcal{T}_h} \mathcal{L}_T(u_{b,T}, v_h) = (f, v_h) \qquad \forall v_h \in V_h, \\ \mathcal{L}_T(u_{b,T}, v_{b,T}) + \mathcal{L}_T(u_h, v_{b,T}) = (f, v_{b,T})_T \qquad \forall v_{b,T} \in B_T \quad \text{and} \quad \forall T \in \mathcal{T}_h \end{cases} \tag{2.5}$$

where the notation $\mathcal{L}_T(\cdot,\cdot)$ and $(\cdot,\cdot)_T$ indicates that the integrals involved are restricted to the element T.

Formally, *static condensation* is performed by solving element–wise the second equation in (2.5), or *bubble equation*:

$$\mathcal{L}_T(u_{b,T}, v_{b,T}) = (f, v_{b,T})_T - \mathcal{L}_T(u_h, v_{b,T}) \qquad \forall v_{b,T} \in B_T, \tag{2.6}$$

for $u_{b,T}$ and substituting the resulting expression for $u_{b,T}$ into the first equation in (2.5), see, e.g., [10].

Before presenting the details of the static condensation procedure, let us introduce a particular set of bubbles, namely the *residual–free bubbles*. These were originally proposed by Brezzi and Russo [29] and Franca and Russo [50].

The idea is to fix B_T for $T \in \mathcal{T}_h$ as the set of all solution of (2.6), considered as a problem in $H_0^1(T)$ (instead of B_T), while u_h ranges over V_h. That is, for each $w_h \in V_h$, we define the residual–free bubble $w_{b,T} \in H_0^1(T)$ as the solution of the variational problem

$$\begin{cases} \text{find } w_{b,T} \in H_0^1(T) \text{ such that} \\ \mathcal{L}_T(w_{b,T}, v) = (f, v)_T - \mathcal{L}_T(w_h, v) = (f - Lw_h, v)_T \qquad \forall v \in H_0^1(T). \end{cases} \tag{2.7}$$

Since in (2.7) we are testing in the whole space $H_0^1(T)$, we can also think of $w_{b,T}$ as the weak solution of the problem (2.7) written in terms of the differential operator L, namely

$$\begin{cases} Lw_{b,T} = (f - Lw_h)|_T & \text{in } T \\ w_{b,T} = 0 & \text{on } \partial T. \end{cases} \tag{2.8}$$

Hence, $L(w_h + w_{b,T})|_T = f|_T$, which justifies the name *residual–free bubble*.

If, for example, $\mathcal{L}_T(\cdot,\cdot)$ is a continuous and coercive bilinear form over $H_0^1(T) \times H_0^1(T)$, then, for every $w_h \in V_h$, problem (2.7) has a unique solution in $H_0^1(T)$; thus we can consider the bounded linear operator $L_T^{-1} : H^{-1}(T) \to H_0^1(T)$ defined as the solution operator of (2.8).

Finally, to ensure that the bubble space is linear, we define the *residual–free bubble* space $\widetilde{B}_h$ element–wise as

$$B_T|_T = \{L_T^{-1}(\lambda f - Lw_h)|_T, w_h \in V_h, \lambda \in \mathbb{R}\} \qquad \forall T \in \mathcal{T}_h. \tag{2.9}$$

Notice that $\dim(\widetilde{B}_h|_T) \leq \dim(V_h|_T)$.

With such definition of the bubble space, we can write the bubble part of the solution as

$$u_{b,T} = L_T^{-1}(f - Lu_h)|_T \qquad \forall T \in \mathcal{T}_h. \tag{2.10}$$

Inserting (2.10) into the first equation in (2.5) we obtain:

$$\begin{cases} \text{find } u_h \in V_h \text{ such that} \\ \mathcal{L}(u_h, v_h) + \displaystyle\sum_{T \in \mathcal{T}_h} \mathcal{L}_T(L_T^{-1}(f - Lu_h)_T, v_h) = (f, v_h) \qquad \forall v_h \in V_h, \end{cases} \tag{2.11}$$

which is a *generalised Galerkin* formulation in terms of the original finite element space V_h.

The bubble solution (2.10) together with (2.11) form the *residual–free bubble* (RFB) method.

Remark. The local inversion of L, necessary for the computation of the bubbles, is a task of the same complexity of the solution of the original problem (2.1). We remark that the scope of the introduction and elimination of the bubble is to recover some information from the subgrid scales, rather then resolve them. Hence, it should be expected that a rough approximation of the bubbles may yield satisfactory results. We postpone the discussion of this issue and of the implementation of the RFB method until the end of this chapter.

2.2 Alternative definition of the method

An interesting development (cf. [20]) is the proof of the equivalence of the residual–free bubble method with the *variational multiscale method* (see [62], [64], [65] and [52]). The latter consists of considering the given problem as decomposed into a problem on a coarse scale, the original finite element mesh, and a problem on a fine scale (the subgrid scale). The fine scale solution is determined element–wise and the result is then substituted into the coarse scale problem. This method can be interpreted as a bubble method in which the local space of bubbles is the whole of $H_0^1(T)$. In fact, it turns out that, the bubble method with $B_T = H_0^1(T)$ is equivalent to the residual–free bubble method defined in the previous section. This is quite obvious since we can still write the bubble part of the solution as in (2.10) and so the variational multiscale approach recovers (2.11).

To see this in more detail, let us first re–define the RFB method as the method obtained through the variational multiscale approach. We introduce the *residual–free bubble* space

$$V_{RFB} = V_h + B_h, \tag{2.12}$$

where the set of bubbles is given by

$$B_h = \bigoplus_{T \in \mathcal{T}_h} H_0^1(T). \tag{2.13}$$

Notice that the space V_{RFB} admits the representation

$$V_{RFB} \equiv \{v \in V : v|_{\partial T} = v_h|_{\partial T} \text{ for some } v_h \in V_h, \forall T \in \mathcal{T}_h\}. \tag{2.14}$$

In other words, V_{RFB} is obtained as follows. To start with, a space of traces is defined as the restriction of V_h onto the skeleton of the triangulation. Successively, V_{RFB} is defined by considering all extensions in V from such space of traces.

We then define the infinite–dimensional version of the RFB method as:

$$\begin{cases} \text{find } u_{RFB} \in V_{RFB} \text{ such that} \\ \mathcal{L}(u_{RFB}, v) = (f, v) \qquad \forall v \in V_{RFB}. \end{cases} \tag{2.15}$$

First off all, the formulation (2.15) is still locally residual–free. Indeed, the solution u_{RFB} satisfies weakly

$$L_T u_{RFB} = f|_T \qquad \forall T \in \mathcal{T}_h. \tag{2.16}$$

To show this, we only need to test in (2.15) with the function $v \in V_{RFB}$ defined as $v_{|_T} = \phi \in C_0^\infty(T)$ and $v = 0$ outside T. This gives $(L_T u_{RFB}, \phi)_T = (f, \phi)_T$ and we thus conclude (2.16) since ϕ is arbitrarily chosen in $C_0^\infty(T)$.

Let $u_{RFB}|_T = u_h + u_b$ be a possible decomposition of u_{RFB} according to (2.12). Recalling (2.16), we have that u_b is given by

$$u_b|_T = L_T^{-1}(f - Lu_h)|_T \qquad \text{in each } T \in \mathcal{T}_h. \tag{2.17}$$

In other words, u_b belongs to the finite–dimensional set of bubbles $\widetilde{B}_h$ defined in the previous section. We have just proved that u_{RFB} belongs to

$$\widetilde{V}_{RFB} = V_h \oplus \widetilde{B}_h.$$

Further, u_{RFB} coincides with the solution of the finite–dimensional problem

$$\begin{cases} \text{find } \tilde{u}_{RFB} \in \widetilde{V}_{RFB} \text{ such that} \\ \mathcal{L}(\tilde{u}_{RFB}, v) = (f, v) \qquad \forall v \in \widetilde{V}_{RFB}, \end{cases} \tag{2.18}$$

which is nothing else but (2.4).

The new interpretation of the method as an infinite–dimensional Galerkin formulation is rather appealing since it clarifies its approximation properties.

Before focusing on the numerical analysis of the method, let us introduce the following generalisation which will become useful later on.

2.2.1 The general augmented space framework

Let Σ be the skeleton of our partition $\mathcal{T}_h$, i.e. the union of the boundaries of all elements in $\mathcal{T}_h$, and let Φ be the space of traces of V on Σ.

A general *augmented subspace* is defined by considering all the extensions onto $\bar{\Omega}$ from a finite–dimensional subspace of Φ. That is, given a subspace $\Phi_h \subset \Phi$ of finite dimension, we define the space

$$V_a := \left\{ v \in V \,:\, v_{|\Sigma} \in \Phi_h \right\}.$$

The augmented space formulation is obtained by restricting (2.2) to V_a:

$$\begin{cases} \text{find } u_a \in V_a \text{ such that} \\ \mathcal{L}(u_a, v_a) = (f, v_a) \qquad \forall v_a \in V_a. \end{cases} \tag{2.19}$$

Existence and uniqueness are ensured by the coercivity of the bilinear functional $\mathcal{L}$.

By definition, the augmented space V_a always contains as a subspace the residual–free bubble space B_h as defined in (2.13). Moreover, the RFB space is a special case of V_a obtained from the above general formulation simply by choosing Φ_h as the space spanned by the traces of V_h.

We can also identify a second subspace V_l, which depends on the bilinear form $\mathcal{L}$:

$$V_l := \{ v_l \in V_a \,:\, \mathcal{L}(v_l, v_b) = 0 \quad \forall v_b \in B_h \}, \tag{2.20}$$

and observe that we have the splitting

$$V_a = V_l \oplus B_h. \tag{2.21}$$

In the special case in which V_a is equal to V_{RFB}, we also have, by definition,

$$V_a = V_{RFB} = V_h + B_h.$$

Since V_h and V_l are not equal, the two characterisations of the RFB space are different. Thus, the augmented space formulation naturally gives a new interpretation to the RFB method. Indeed, the solution u_a of (2.19) can be characterised as follows (see [25]).

Theorem 2.2.1 *Let u_a be the unique solution of (2.19). Then, its decomposition according to (2.21) is given by $u_a = u_l + u_b^f$ where u_b^f is the unique solution in B_h of*

$$\mathcal{L}(u_b^f, v_b) = (f, v_b) \qquad \forall v_b \in B_h, \tag{2.22}$$

and u_l can be characterised as the unique solution in V_l of

$$\mathcal{L}(u_l, v_l) + \mathcal{L}(u_b^f, v_l) = (f, v_l) \qquad \forall v_l \in V_l. \tag{2.23}$$

Proof. The existence of unique solutions to problems (2.22) and (2.23) is ensured by the coercivity of $\mathcal{L}$. Let $u_a = u_l + u_b$ be the unique decomposition of u_a according to (2.21). Testing in (2.19) with $v_a = v_b \in B_h$ and using the definition (2.20) we have

$$(f, v_b) = \mathcal{L}(u_a, v_b) = \mathcal{L}(u_l, v_b) + \mathcal{L}(u_b, v_b) = \mathcal{L}(u_b, v_b),$$

that is, $u_b = u_b^f$. Finally, testing in (2.19) with $v_a = v_l \in V_l$, we see that u_l satisfies (2.23). □

Remark. The *bubble equation* (2.22), i.e. the equation obtained from the augmented space formulation by testing in the bubble space, has the appealing property of being independent of u_l. Moreover the two equations (2.22) and (2.23) decouple if $\mathcal{L}$ is a symmetric (bilinear) functional, since in this case $\mathcal{L}(u_b^f, v_l) = \mathcal{L}(v_l, u_b^f) = 0$ by (2.20).

2.3 Convergence results

The *a priori* error analysis of the RFB method (2.15) for the solution of a non–self–adjoint version of (2.1) was performed by Brezzi *et al.* [28]. We report here the main result of that paper.

We assume that the second–order linear elliptic operator L is of the form

$$L = \varepsilon D + A \tag{2.24}$$

where

$$Dw \equiv -\sum_{i,j=1}^{n} \frac{\partial}{\partial x_j}\left(d_{ij}(x)\frac{\partial w}{\partial x_i}\right), \qquad Aw \equiv \sum_{i=1}^{n} a_i(x)\frac{\partial w}{\partial x_i}.$$

Moreover we assume that:

- D is symmetric and normalised to one, i.e. for almost every x in Ω, the $n \times n$ matrix $(d_{ij}(x))$ is symmetric and positive definite, with smallest eigenvalue $\geq \delta > 0$ and largest eigenvalue ≤ 1. Hence, with $V = H_0^1(\Omega)$, the associated bilinear functional a satisfies:

$$\begin{aligned} &d(v,v) \geq \delta |v|^2_{H^1(\Omega)} && \forall v \in V, \\ &|d(w,v)| \leq |w|_{H^1(\Omega)} |v|_{H^1(\Omega)} && \forall w, v \in V; \end{aligned}$$

- for almost every $x \in \Omega$, the n-component vector $(a_i(x))$ has Euclidean norm $\leq \gamma$, with $\gamma > 0$ independent of x. We finally assume that the associated bilinear functional is skew–symmetric, hence:

$$\begin{aligned} &|a(w,v)| \leq \gamma |w|_{H^1(\Omega)} \|v\|_{L_2(\Omega)} && \forall w \in H^1(\Omega), \forall v \in L_2(\Omega), \\ &a(w,v) = -a(v,w) && \forall w, v \in V. \end{aligned}$$

Skew–symmetry is assured, for instance, by requiring that the vector field $\mathbf{a} = (a_1, \ldots, a_n)$ is divergence–free on Ω in the sense of distributions.

In this way, it follows that the bilinear form associated with the operator L, namely

$$\mathcal{L}(w,v) = \varepsilon d(w,v) + a(w,v) \qquad \forall w, v \in V,$$

satisfies

$$\begin{aligned} &\mathcal{L}(v,v) \geq \varepsilon \delta |v|^2_{H^1(\Omega)} && \forall v \in V, \\ &|\mathcal{L}(w,v)| \leq C |w|_{H^1(\Omega)} |v|_{H^1(\Omega)} && \forall w, v \in V. \end{aligned}$$

Thus, by the Lax-Milgram lemma, the variational problem associated with (2.1),

$$\begin{cases} \text{find } u \in V \text{ such that} \\ \mathcal{L}(u,v) = (f,v) \qquad \forall v \in V, \end{cases} \tag{2.25}$$

has a unique solution in V.

We then consider the RFB approximation of (2.25) on affine–equivalent finite element methods (the authors of [28] dealt with triangular elements, but the extension of their analysis to, for example, parallelepipedal elements is only a notational matter).

Consider a *conforming* and *shape regular* family of partitions $\{\mathcal{T}_h\}_h$ of $\Omega \subset \mathbb{R}^n$, i.e. with the property that

1. *Conformity*: Any two elements in $\mathcal{T}_h$ either have a common face of dimension $< n$ or they do not intersect at all;

2. *Shape regularity*: For any element, the ratio of the largest circumscribed sphere to that of the smallest inscribed sphere is bounded above by a constant μ which does not depend on the element and on the mesh parameter h.

Moreover, we assume that any $T \in \mathcal{T}_h$ is affine–equivalent to the n–dimensional unit simplex or unit hypercube through an affine map

$$F_T(\hat{x}) = M\hat{x} + \boldsymbol{t},$$

where M is an invertible $n \times n$ matrix and $\boldsymbol{t}$ is a vector in $\mathbb{R}^n$.

From now on, we denote by $\mathcal{P}_k$ the space of algebraic polynomials of degree $\leq k$, i.e.

$$\mathcal{P}_k = \left\{ p : \mathbb{R}^n \to \mathbb{R} \ : \ p = \sum_{|\alpha| \leq k} c_\alpha \boldsymbol{x}^\alpha \right\}, \tag{2.26}$$

and by $\mathcal{Q}_k$ the space of algebraic polynomials of degree $\leq k$ with respect to each variable, i.e.

$$\mathcal{Q}_k = \left\{ q : \mathbb{R}^n \to \mathbb{R} \ : \ q = \sum_{0 \leq \alpha_i \leq k} c_\alpha \boldsymbol{x}^\alpha \right\}, \tag{2.27}$$

where $\alpha = \{\alpha_1, \ldots, \alpha_n\} \in \mathbb{N}^n$ represents a multi–index and $|\alpha| = \sum_{i=1}^n \alpha_i$. We set

$$V_{RFB} = \Big\{ v \in V : \big(v_{|_T} \circ F_T\big)_{|_{\hat{e}}} \in \mathcal{P}_{k|_{\hat{e}}} \\ \text{for each } (n-1)\text{–dimensional face } \hat{e} \text{ of } \widehat{T} \text{ and any element } T \in \mathcal{T}_h \Big\}. \tag{2.28}$$

As we have seen in the previous sections, $V_{RFB} = V_h + B_h$ where B_h is given by (2.13) and V_h is the standard conforming finite element space

$$V_h = \left\{ v_h \in H_0^1(\Omega) : \begin{cases} w_{h|_T} \in \mathcal{P}_k, & \text{if } T \text{ is a simplex} \\ w_{h|_T} \circ F_T \in \mathcal{Q}_k, & \text{if } T \text{ is a parallelepiped} \end{cases} \right\}.$$

The RFB finite element discretisation of (2.25) is given, as before, by (2.15).

The following is the main result in [28].

Theorem 2.3.1 *Let u and u_{RFB} be the solutions of (2.25) and (2.15), respectively. Assuming that $u \in H^{k+1}(\Omega) \cap H_0^1(\Omega)$, there exists a positive constant $\beta^* = \beta^*(\mu, k)$, independent of h, δ, γ and ε, such that*

$$\varepsilon^{1/2} |u - u_{RFB}|_{H^1(\Omega)} \leq \frac{\beta^*}{\delta} \left(\sum_{T \in \mathcal{T}_h} \left(\varepsilon h_T^{2r} + \gamma_T h_T^{2r+1} \right) |u|^2_{H^{r+1}(\omega(T))} \right)^{1/2}, \tag{2.29}$$

where $0 \leq r \leq k$. Here $\omega(T)$ denotes the patch of elements consisting of T and its immediate neighbours.

In [28], in order to clarify the strength of the norm $\varepsilon^{1/2}|u - u_{RFB}|_{H^1(\Omega)}$, the following result is shown.

Theorem 2.3.2 *Let u and u_{RFB} be the solutions of (2.25) and (2.15), respectively. Then,*

$$\|A(u-u_{RFB})\|_*^2 \leq 4\varepsilon |u-u_{RFB}|^2_{H^1(\Omega)}, \tag{2.30}$$

where

$$\|v\|_*^2 = \sum_T \frac{1}{\gamma_T}\|v\|^2_{(X(\infty,T))'},$$

and $X(\infty,T)$ is the fractional–order Besov space $\overset{\circ}{B}{}^{1/2}_{2,\infty}(T) = \big(L_2(T), H^1_0(T)\big)_{1/2,\infty}$.

Remark. Besov spaces are fractional order Banach spaces obtained through the real method of interpolation applied to the Sobolev spaces $W^{k,p}$, see Adams and Fournier [3].

Remark. The norm $\|\cdot\|_*$ can be thought of as a broken norm on a space with negative–order 1/2. Indeed, there exist two positive constants $\beta_1 = \beta_1(\mu,k)$, and $\beta_2 = \beta_2(\mu,k)$ independent of T, such that

$$\beta_1\|v\|_* \leq \left(\sum_T \frac{h_T}{\gamma_T}\|v\|^2_{L_2(T)}\right)^{1/2} \leq \beta_2\|v\|_* \tag{2.31}$$

for each v in V_h. Thus, the right–hand side of (2.29) bounds also the left–hand side of (2.30). In particular, a control on the L_2–norm of the streamline derivative is implied by the theorem.

Actually, as mentioned in [24], $\varepsilon^{1/2}|u|_{1,\Omega}$, where u is solution of a general problem with smooth data, is the strongest norm of u which can be expected to remain bounded independent of ε in the limit $\varepsilon \to 0$.

2.4 Steady–state convection–diffusion

The residual–free bubble method was first introduced for the solution of b.v.p.'s for the linear steady–state convection–diffusion equation

$$Lu \equiv -\varepsilon\Delta u + \boldsymbol{a}\cdot\nabla u = f \qquad \text{in } \Omega, \tag{2.32}$$

where ε is a positive parameter and $\boldsymbol{a}$ is a continuously differentiable field. This is also the problem on which this work is mainly focused on.

When ε is small w.r.t. $\boldsymbol{a}$ (*convection–dominated regime*), the solution of (2.32) may exhibit sharp layers, i.e. narrow regions where the solution and its derivatives change very rapidly (see Appendix A.1 for the definition of boundary and internal layer thicknesses). Classical numerical techniques, like Galerkin finite element methods as well as central and upwind finite difference methods, are globally unstable or inaccurate when the layer thickness is smaller then the scale of the discretisation, that is, when the mesh used does not resolve the *small scales* of

the problem (in this case, the boundary and internal layers). This concept is made precise by the definition of the *mesh Péclet number*

$$Pe_T = \frac{\|\boldsymbol{a}\|_{\infty,T} h_T}{2\varepsilon},$$

which measures if the numerical approximation is convection–dominated. Typically, the Galerkin finite element solution presents unphysical, maximum–principle–violating oscillations when $Pe_T > 1$ for elements T contained in the subdomains of Ω which surround the layers.

2.4.1 Classical stabilised methods

Classical *stabilised* finite element methods achieve the double goal of stability and accuracy in the solution of (2.32) in both the convection–dominated and diffusion–dominated regimes, by modifying the bilinear form $\mathcal{L}$. The Galerkin formulation, which reads,

$$\text{find } u_h \in V_h: \quad \mathcal{L}(u_h, v_h) = (f, v_h) \qquad \forall v_h \in V_h,$$

is modified into a *generalised Galerkin* formulation by introducing a *stabilisation term* designed to improve the stability of the method without destroying consistency. In particular, this can be achieved by introducing a discrete diffusion operator associated with the streamline direction (streamline–diffusion).

The three streamline–diffusion (SD) stabilised methods that have been most frequently studied in the literature add the following term to the left–hand side of the standard Galerkin formulation:

- *Streamline Upwind/Petrov-Galerkin* (SUPG) stabilisation:

$$S = \sum_{T \in \mathcal{T}_h} \tau_T \int_T \left(-\varepsilon \Delta u_h + \boldsymbol{a} \cdot \nabla u_h - f\right) \left(\boldsymbol{a} \cdot \nabla v_h\right) \, d\boldsymbol{x} \tag{2.33}$$

- *Galerkin/Least-Squares* (GALS) stabilisation:

$$S = \sum_{T \in \mathcal{T}_h} \tau_T \int_T \left(-\varepsilon \Delta u_h + \boldsymbol{a} \cdot \nabla u_h - f\right) \left(-\varepsilon \Delta v_h + \boldsymbol{a} \cdot \nabla v_h\right) \, d\boldsymbol{x} \tag{2.34}$$

- *Douglas-Wang/Galerkin* (DWG) stabilisation:

$$S = -\sum_{T \in \mathcal{T}_h} \tau_T \int_T \left(-\varepsilon \Delta u_h + \boldsymbol{a} \cdot \nabla u_h - f\right) \left(-\varepsilon \Delta v_h - \boldsymbol{a} \cdot \nabla v_h\right) \, d\boldsymbol{x}, \tag{2.35}$$

where τ_T is a user–chosen parameter, called the SD–parameter.

All these methods can be interpreted as Petrov–Galerkin methods. Notice that, for linear or bilinear finite elements $\Delta u_h|_T = 0$ and $\Delta v_h|_T = 0$ for every $T \in \mathcal{T}_h$, so the three choices give

rise to the same method which is referred to as the *streamline diffusion finite element method* (SDFEM).

In general the SD–parameter depends both on the mesh size and on the mesh Péclet number. A typical result for the SUPG method, taken from [82], is the following.

Under appropriate assumptions, the SUPG method with

$$\tau_T = \begin{cases} \tau_0 h_T & \text{if } Pe_T > 1 \text{ (convection-dominated case)} \\ \tau_1 h_T^2/\varepsilon & \text{if } Pe_T \leq 1 \text{ (diffusion-dominated case),} \end{cases} \tag{2.36}$$

satisfies the global error estimate

$$|||u - u_h|||_{SD} \leq C \left(\varepsilon^{1/2} + h^{1/2}\right) h^k |u|_{k+1}, \tag{2.37}$$

where the SD–norm is defined as

$$|||v|||_{SD} := \left(\varepsilon |v|_1^2 + \sum_{T \in \mathcal{T}_h} \tau_T \|\boldsymbol{a} \cdot \nabla v\|_{0,T}^2\right)^{1/2}.$$

The error bound (2.37) is optimal. A bound on the L_2–norm error of the SUPG method can be added in the presence of a reaction term, although this turns out to be 1/2–order suboptimal: a price has to be paid for improving the stability of the method. This deficiency is not due to an incorrect choice of the SD–parameter. Indeed, it was shown by Zhou [96] through the *a priori* error analysis of the method that on certain meshes the error bound (2.37) and the L_2–norm error bound are sharp. The same is true of the RFB method which has the same order of convergence as SDFEM, cf. the error bound (2.29).

The definition of τ_T given by (2.36) states that, where the problem is convection-dominated, the SD–parameter must be proportional to the element size h_T. Still the problem of identifying an optimal value of τ_T, in the sense of providing a sharp, non–oscillatory approximation in the layers, is left open. Indeed, the value for τ_0 and τ_1 is not implied by the *a priori* analysis.

The difficulty encountered with the problem of parameter identification may be seen as a consequence of the lack of physical justification for the stabilised method based on the notion of relevant scales.

Indeed, one of the reasons for the success of two–level, or subgrid scale, methods such as the variational multiscale method (or local Green's function approach, see [62]) and the residual–free bubble method introduced in [29], is that they can provide the required theoretical foundation to classical stabilised techniques.

For each of the methods (2.33), (2.34) and (2.35) one can find in the literature an equivalence with some subgrid scale model, at least for low–order finite element methods.

The formal equivalence between bubble methods and the SUPG, DWG and GALS is discussed by Brezzi *et al.* in [18] and [29] and by Baiocchi *et al.* in [10]. Notably, Baiocchi *et al.*, in their article on 'Virtual Bubbles' [10], show that with a suitable choice of the space B_T, it is possible to choose the bilinear functional corresponding to L_T from a vast class of bilinear

functionals and, in particular, to recreate the stabilisation terms of all the known stabilisation methods. Another notable example is given by Hughes in [62] where a theoretically optimal value for the SD–parameter of the DWG method is obtained through the use of the variational multiscale approach.

2.4.2 A special case

Let us focus, for instance, on the equivalence between the SDFEM and the RFB method proved by Brezzi and Russo [29]. This equivalence is limited to linear finite elements under the assumption of piecewise constant coefficients w.r.t. the partition $\mathcal{T}_h$.

We consider the definition of the RFB method as a generalised Galerkin formulation (2.11). By applying Green's formula to the stabilisation term, this can be written

$$\begin{cases} \text{find } u_h \in V_h \text{ such that} \\ \mathcal{L}(u_h, v_h) + \displaystyle\sum_{T\in\mathcal{T}_h} (L_T^{-1}(f - Lu_h)|_T, L^* v_h)_T = (f, v_h) \qquad \forall v_h \in V_h, \end{cases} \tag{2.38}$$

where L^* is the formal adjoint of L.

Let L be the convection–diffusion operator (2.32) and assume that $\boldsymbol{a}$ and f are piecewise constant over $\mathcal{T}_h$ and that V_h is the space of piecewise linear functions in $H_0^1(\Omega)$, defined over elements T of $\mathcal{T}_h$. In this case, on every element $T \in \mathcal{T}$ we have that

$$f - Lu_h = f - (-\varepsilon\Delta u_h + \boldsymbol{a}\cdot\nabla u_h) = f - \boldsymbol{a}\cdot\nabla u_h,$$

is constant. Hence, by the definition (2.9) the local bubble space B_T is one-dimensional and we may define as basis for B_T the function $b_T \in H_0^1(T)$ which solves the problem

$$\begin{cases} -\varepsilon\Delta b_T + \boldsymbol{a}\cdot\nabla b_T = 1 & \text{in } T, \\ b_T = 0 & \text{on } \partial T. \end{cases} \tag{2.39}$$

In this way we have that $L_T b_T = 1$, $b_T = L_T^{-1}1$, and hence

$$L_T^{-1}(f - Lu_h)|_T = (f - \boldsymbol{a}\cdot\nabla u_h)|_T b_T,$$

so the stabilisation term in (2.38) becomes

$$\begin{aligned} \left(L_T^{-1}(f - Lu_h)|_T, L^* v_h\right)_T &= -(f - \boldsymbol{a}\cdot\nabla u_h)|_T (\boldsymbol{a}\cdot\nabla v_h)|_T \int_T b_T \, d\boldsymbol{x} \\ &= \frac{\int_T b_T \, d\boldsymbol{x}}{|T|} \int_T (\boldsymbol{a}\cdot\nabla u_h - f)(\boldsymbol{a}\cdot\nabla v_h) \, d\boldsymbol{x} \end{aligned}$$

This is the stabilisation term of the SDFEM with a particular choice of the stabilisation parameter τ_T, hence the two methods are equivalent. The advantage of the RFB approach is

that the stabilisation parameter is produced by the method rather than by ad hoc tuning. In this case,

$$\tau_T = \frac{\int_T b_T \, d\boldsymbol{x}}{|T|}. \tag{2.40}$$

The equivalence between SDFEM and RFB is limited to the case of linear elements. For instance it does not hold for bilinear elements. Indeed in this case, although the stability properties of the two methods are comparable (see [51]), they do not coincide. This fact was proved by Brezzi *et al.* [22] through a standard stability argument. By considering the limit of vanishing viscosity and testing in the RFB stabilisation term with the bilinear component of the solution u_h, we get (cf. [22]):

$$\frac{1}{2|\boldsymbol{a}|} \int_{\partial T_+} [u_h(\boldsymbol{x}_Q) - u_h(\boldsymbol{x}_P)]^2 \, \boldsymbol{a} \cdot \boldsymbol{n} \, d\Gamma,$$

where $\boldsymbol{x}_P$ is the point on the inflow part ∂T_- of the boundary of T which is aligned with the point $\boldsymbol{x}_Q$ on the outflow part ∂T_+ of ∂T in the streamline direction. From this expression we see that the effect of the residual–free bubbles on the formulation is to add stability thorugh the 'jump' of the bilinear solution in the streamline direction.

2.5 RFB and upscaling

We survey the use of the RFB method for the solution of

$$-\nabla \cdot (d(\boldsymbol{x}) \nabla u) = f \qquad \text{in } \Omega, \tag{2.41}$$

together with homogeneous Dirichlet boundary conditions. The *conductivity tensor* $d(\boldsymbol{x}) = (d_{ij}(\boldsymbol{x}))_{i,j}$ is assumed to be symmetric and positive definite.

The interest and difficulty in the solution of (2.41) arises when the tensor d, which models the properties of the underlying medium, is characterised by a multiple–scale structure. This is often the case of interest when (2.41) is used to model physical phenomena like single phase steady flow in porous media and steady state heat or electrical conduction in composite materials.

A successful direct numerical solution typically requires that all the scales embedded in the problem are resolved, i.e. that the size of the numerical mesh is comparable with the finest scale of the problem.

Often, though, the quantities of interest are only some *macroscopic* features (e.g., the global average) of the solution. This fact leads to the development of methods which are able to capture the fine scale effects on the coarse, macroscopic scale without necessarily resolving all the fine scales.

A classical example is given by the *homogenization theory*, which deals with problems with oscillatory coefficients (the reference book is by Bensoussan, Lions and Papanicolau [13] and,

more recently, Cioranescu and Donato [34]). It is assumed that $d = d(\boldsymbol{x}/\varepsilon)$, where ε is a small parameter (small with respect to the dimensions of Ω) and that d is periodic with respect to the *fast variable* $\boldsymbol{y} = \boldsymbol{x}/\varepsilon$. Under appropriate regularity assumptions on d, it is proven that the solution u to (2.41) admits an asymptotic expansion of the form

$$u(\boldsymbol{x}, \boldsymbol{y}) = u_0(\boldsymbol{x}) + \varepsilon u_1(\boldsymbol{x}, \boldsymbol{y}) + O(\varepsilon^2),$$

where u_0 is the solution of the *homogenized equation*

$$-\nabla \cdot (d^* \nabla u_0) = f \qquad \text{in } \Omega,$$

and d^* is the symmetric and positive definite (constant) *effective coefficient* determined by solving a problem similar to (2.41). Moreover, u converges to u_0 as $\varepsilon \to 0$ weakly in $H_0^1(\Omega)$ (and point–wise under stronger smoothness conditions).

Obtaining the effective coefficient d^* is not an easy task in general, an exception being given by the one–dimensional problem discussed below.

Numerical upscaling methods seek to approximate d^* to successively obtain u_0 as an approximation of u. We say that such methods are two–level procedures: at the fine level the effects of the fine scales on the macroscopic scales is incorporated in d^*, while at the coarse level the macroscopic features of the solution are approximated by solving for u_0.

The RFB method is also a two level procedure (static condensation of the bubble followed by solution on the coarse grid) and, as it turns out, it is an upscaling method. As we show below, this can be seen very clearly when the RFB method is applied to a one–dimensional problem. For the interpretation of RFB as an upscaling technique in many spacial dimensions see, instead, Sangalli [87].

We refer to Section 2.1 for the definition of the RFB method. As usual we have the characterisation of the RFB space

$$V_{RFB} = V_h \oplus B_h,$$

(let us just assume, here, that the two subspaces are in direct sum).

Given the bilinear form

$$\mathcal{L}(u, v) = \int_\Omega d_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} \, d\boldsymbol{x},$$

associated to (2.41), we have a two level procedure by testing in V_h and B_h separately:

$$\begin{cases} \text{find } u_{RFB} = u_h + u_b \in V_h \oplus B_h \text{ such that} \\ \mathcal{L}(u_h, v_h) + \mathcal{L}(u_b, v_h) = (f, v_h) & \forall v_h \in V_h, \\ \mathcal{L}(u_b, v_b) + \mathcal{L}(u_h, v_b) = (f, v_b) & \forall v_b \in V_b. \end{cases} \tag{2.42}$$

We then (formally) perform static condensation of the bubble by solving the bubble equation in (2.42) locally:

$$u_b|_T = L_T^{-1}(f - Lu_h)|_T,$$

where L represents the linear differential operator associated with $\mathcal{L}$.

Finally, after substitution of u_b into the first equation in (2.42), we obtain the 'coarse grid' RFB formulation: find $u_h \in V_h$ such that

$$\mathcal{L}(u_h, v_h) - \sum_{T\in\mathcal{T}_h} \left(L_T^{-1}(Lu_h), Lv_h\right)_T = (f, v_h) - \sum_{T\in\mathcal{T}_h} \left(L_T^{-1} f, Lv_h\right)_T \qquad \forall v_h \in V_h, \tag{2.43}$$

since L is self–adjoint.

In practice the static condensation is carried out as follows. Let $\{\varphi_j\}_{j=1}^{N_T}$ be the set of local basis functions for V_h in T, and write $u_h|_T = \sum_{j=1}^{N_T} U_j\varphi_j$. We have

$$L_T^{-1}(Lu_h)|_T = L_T^{-1}\left(\sum_{j=1}^{N_T} U_j L\varphi_j|_T\right) = \sum_{j=1}^{N_T} U_j L_T^{-1}(L\varphi_j|_T) = \sum_{j=1}^{N_T} U_j b_j, \tag{2.44}$$

where, as suggested in the last equality, b_j is defined element-wise as the solution of the boundary value problem

$$\begin{cases} -\nabla\cdot(d\nabla b_j) = -\nabla\cdot(d\nabla\varphi_j) & \text{in } T, \\ b_j = 0 & \text{on } \partial T. \end{cases} \tag{2.45}$$

We can proceed in a similar fashion for the bubble term on the right–hand side of (2.43) realizing that to carry out the static condensation, we also need to solve on every element T the problem

$$\begin{cases} -\nabla\cdot(d\nabla b_f) = f & \text{in } T, \\ b_f = 0 & \text{on } \partial T. \end{cases} \tag{2.46}$$

Over all, the set of functions $\{b_1, \ldots, b_{N_T}, b_f\}$ form a local basis for a finite–dimensional subspace of the space of bubbles B_h and we have $u_b|_T = b_f - \sum_{j=1}^{N_T} U_j b_j$.

As in the case of convection–diffusion problems, we are left with a generalised Galerkin formulation for the FEM space V_h. Only the interpretation of the bubble terms in (2.43) changes: this time they are thought as representing the effect of the fine scales onto the macroscopic scale. In practice this difference is crucial: the more accurately the bubble terms are computed, the more details will be included and, as a consequence, the more accurate u_h will be as an approximation of the coarse scale solution. Interpreting the RFB method as an upscaling procedure, we say that the accuracy put into the computation of the bubbles determines the effective coefficient of the method.

2.5.1 A one–dimensional example

For relatively simple problems, like one–dimensional problems or when $a(\boldsymbol{x})$ is separable in space, we can easily obtain analytically the subgrid functions needed by the two–level procedure

(2.42). This allows us to analyse the RFB method without worrying about the resolution at the subgrid level.

To start with, let us consider the analogue of problem (2.41) on the one–dimensional domain $\Omega = (0,1)$, that is

$$\begin{cases} -(d_\varepsilon(x)u'_\varepsilon(x))' = f(x) & \text{in } (0,1), \\ u(0) = u(1) = 0, \end{cases} \tag{2.47}$$

where $d_\varepsilon(x)$ is an abbreviation for $d(x/\varepsilon)$ and $d(x)$ is a continuous 1-periodic function, ε is a small parameter and the forcing term $f \in L_2(0,1)$.

In this case,

$$\mathcal{L}(u,v) = \int_0^1 d_\varepsilon(x)u'(x)v'(x)dx,$$

and we assume that $d_\varepsilon(x) > 0$ to ensure that $\mathcal{L}$ is elliptic.

The homogenized problem associated to (2.47) is simply given by (see, again, [13])

$$-(1/m(1/d))u''_0 = f, \tag{2.48}$$

where

$$m(1/d) = \int_0^1 \frac{dt}{d(t)}.$$

We consider linear finite element discretisations of (2.47) over partitions $\{x_j\}_{j=1}^N$ of $(0,1)$ into non–overlapping sub–intervals $T_j = [x_j, x_{j+1}]$.

The solutions of the bubble problems (2.45) are in this case trivially given by the two local bubbles

$$b_j(x) = \frac{1}{x_{j+1}-x_j}\left(-(x-x_j) + (x_{j+1}-x_j)\frac{\int_{x_j}^x \frac{1}{d_\varepsilon(t)}\,dt}{\int_{x_j}^{x_{j+1}} \frac{1}{d_\varepsilon(t)}dt}\right) \quad \text{and} \quad b_{j+1}(x) = 1 - b_{j|_{T_j}}(x).$$

Having solved the subgrid problems, we are ready to compute the stiffness matrix associated with (2.42). The RFB method (2.42) takes the following algebraic form:

$$(D+B)U = F_D + F_B.$$

Assume that the partition is uniform and that every sub–interval is of length h. Then, since we are considering linear finite elements, the standard Galerkin term is simply given by

$$D = \frac{1}{h^2}\text{tridiag}\left(-\int_{T_{j-1}} d_\varepsilon(x)\,dx, \int_{T_{j-1}} d_\varepsilon(x)\,dx + \int_{T_j} d_\varepsilon(x)\,dx, -\int_{T_j} d_\varepsilon(x)\,dx\right),$$

while the bubble term becomes

$$B = \frac{1}{h}\text{tridiag}\left(\int_{T_{j-1}}(d_\varepsilon\, b'_{j-1})(x)\,dx, \int_{T_{j-1}}(d_\varepsilon\, b'_j)(x)\,dx - \int_{T_j}(d_\varepsilon\, b'_j)(x)\,dx, \int_{T_j}(d_\varepsilon\, b'_{j+1})(x)\,dx\right).$$

We notice that the entries of D depend on $1/h$ times the average of the coefficient d over an element. In particular, if h is a multiple of ε, we have that D becomes the discrete Laplace operator w.r.t. the constant diffusion $m(d) := \int_0^1 d(x)dx$. In other words, the standard Galerkin method picks-up the wrong homogenized operator

$$-m(d)\frac{d^2}{dx^2}.$$

On the other hand, the entries of B are of the form

$$\begin{aligned}\int_{T_j}(d_\varepsilon\, b'_j)(x)\,dx &= \int_{T_j}\left(-\frac{1}{h} + \frac{1}{d_\varepsilon(x)\int_{T_j}\frac{dt}{d_\varepsilon(t)}}\right)d_\varepsilon(x)\,dx \\ &= -\frac{\int_{T_j} d_\varepsilon(x)\,dx}{h} + \frac{h}{\int_{T_j}\frac{dx}{d_\varepsilon(x)}}.\end{aligned}$$

Thus, summing up the contributions of D and B we get

$$D + B = \text{tridiag}\left(\frac{-1}{\int_{T_{j-1}}\frac{dx}{d_\varepsilon(x)}}, \frac{1}{\int_{T_{j-1}}\frac{dx}{d_\varepsilon(x)}} + \frac{1}{\int_{T_j}\frac{dx}{d_\varepsilon(x)}}, \frac{-1}{\int_{T_j}\frac{dx}{d_\varepsilon(x)}}\right).$$

This time, when h is a multiple of ε, we see that the RFB method coincides with the Galerkin linear FEM applied to the correct homogenized operator

$$-(1/m(1/d))\frac{d^2}{dx^2}.$$

(Notice the analogy with convection–diffusion b.v.p's for which static condensation of the bubble leads to a centered difference approximation of the regularized operator including the numerical viscosity term.)

Hence, in one–dimension, the RFB method is a numerical upscaling method in which the effective coefficient is computed exactly when the mesh size is a multiple of ε (when this is not the case, it is easy to see that the difference between the effective coefficient and the coefficient given by RFB is of order $\mathcal{O}(\varepsilon/h)$). For the interpretation of RFB as an upscaling technique in many spatial dimensions see Sangalli [87].

The effect of upscaling can be seen in the following example. We solve (2.47) in $\Omega = (0,1)$ with $d = 1/(2 + 1.8\sin(2\pi x/\varepsilon))$, $\varepsilon = 0.04$ and $f = -1$. The exact solution and a standard Galerkin solution can be seen in the top–left plot in Figure 2.1: although the mesh used is almost fine enough to resolve the oscillatory behaviour, the Galerkin approximation fails to capture the coarse scale behaviour of the exact solution. On the other hand the exact solution is very well approximated by the RFB method. Indeed, the RFB solution corresponds to the interpolant of u from V_{RFB}, hence the RFB method is nodally exact in one–dimension.

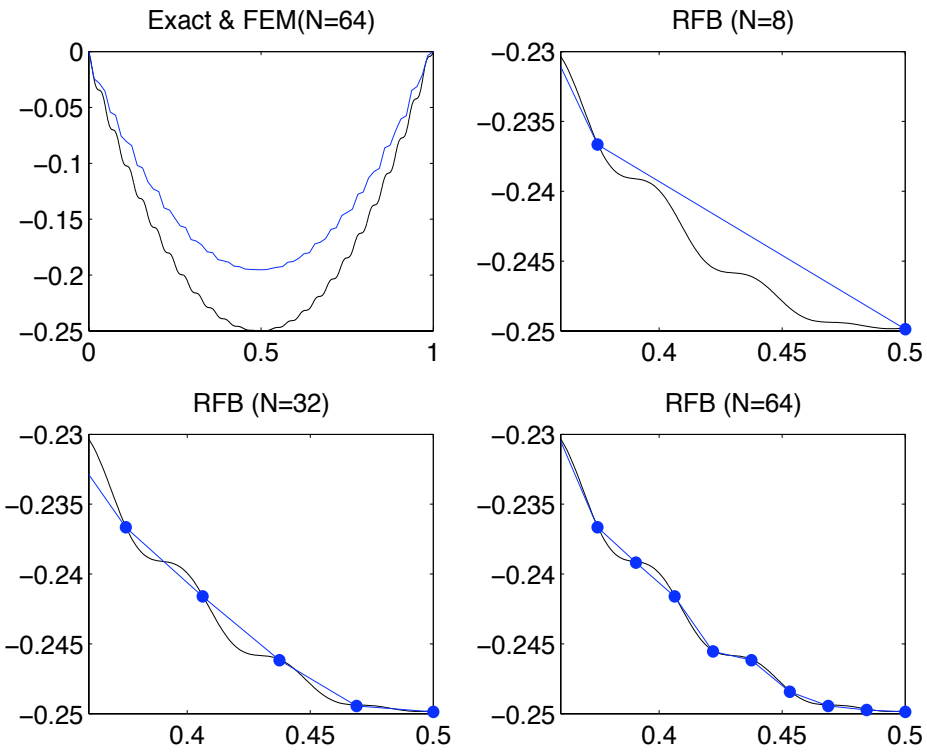


Figure 2.1: Solution of (2.41) in $\Omega = (0,1)$ with $d = 1/(2 + 1.8\sin(2\pi x/\varepsilon))$, $\varepsilon = 0.04$ and $f = -1$.

2.5.2 The MFEM method

A typical difficulty in numerical upscaling is the 'resonance' effect between the mesh size and the scales of the continuous problem [59].

We can analyse the effect of resonance in the RFB method by exploiting the equivalence with the MFEM (*multiscale finite element method*) formulated by Hou and Wu in [59] and analysed by Hou, Wu and Cai in [60].

The equivalence between the two methods was shown by Brezzi and Marini [26] and is easily explained by using the general augmented space formulation [25] which we have described in Section 2.2.1. We have seen that the RFB finite element space V_{RFB} admits the alternative splitting

$$V_{RFB} = V_l \oplus B_h,$$

where

$$V_l := \{v_l \in V_{RFB} \,:\, \mathcal{L}(v_l, v_b) = 0 \quad \forall v_b \in B_h\}. \tag{2.49}$$

Moreover, with such splitting, the RFB solution is given by $u_{RFB} = u_l + u_b^f$ where (cf. Theorem 2.2.1) u_b^f is the unique solution of

$$\mathcal{L}(u_b^f, v_b) = (f, v_b) \qquad \forall v_b \in B_h, \tag{2.50}$$

and u_l is characterised as the unique solution of

$$\mathcal{L}(u_l, v_l) = (f, v_l) \qquad \forall v_l \in V_l, \tag{2.51}$$

Hence, the two problems defining u_b^f and u_l decouple (as we remarked in Section 2.2.1, this happens only when $\mathcal{L}$ is symmetric).

Notice that we could drop u_b^f and simply consider the method given by the numerical solution of (2.51). Such a 'reduced' RFB method coincides with the MFEM of Hou *et al.* [59], [60].

The philosophy of the MFEM is to use the usual FEM formulation but employing *ad hoc* basis functions designed to capture the small-scale information within each element. The basis functions are defined as local solutions of the original p.d.e. but with zero forcing term, hence they must be evaluated numerically using a subgrid.

The method works on axiparallel polygonal domains and is based on a linear or bilinear FEM formulation.

Let $x_j, j = 1, \ldots, N$ be the nodes of the triangulation. The basis function associated with x_j is defined as the solution of the boundary value problem

$$\begin{cases} -\nabla \cdot (d \nabla \phi_j) = 0 & \text{in} T, \quad \forall T \in \mathcal{T}_h, \\ \phi_j(x_i) = \delta_{i,j} & \text{for any node } x_i \text{ that is a vertex of } T, \\ \phi_j \text{ linear on every edge of } T. \end{cases} \tag{2.52}$$

The standard form of the method is:

$$\begin{cases} \text{find } u_{MFEM} = \displaystyle\sum_{j=1}^{N} U_j \phi_j \text{ such that} \\ \mathcal{L}(u_{MFEM}, \phi_i) = (f, \phi_i) \qquad \forall i = 1, \ldots, N. \end{cases} \tag{2.53}$$

By definition, the set $\{\phi_j\}_{j=1}^N$ is a basis for the finite element space V_l; hence (2.53) is equivalent to (2.51). The equivalence can also be seen by comparing the basis functions of the RFB and MFEM methods. Indeed, the homogeneous problem with non–homogeneous boundary conditions (2.52), which gives the basis of the MFEM space, corresponds, through the bilinear basis φ_j, to the non–homogeneous problem with homogeneous boundary conditions (2.45), which gives the basis for the bubble space. That is, one can easily verify that $\phi_j = -b_j + \psi_j$.

Under the hypothesis that $d(\boldsymbol{x})$ is periodic with period ε, the following convergence results are proven in [60] using a standard FEM analysis:

Theorem 2.5.1 *Let u and u_{MFEM} be the weak solutions of (2.41) and (2.53), respectively. Then there exist constants C_0, C_1, independent of h and ε, such that*

$$\|u - u_{MFEM}\|_{0,\Omega} \leq C_0 \left(h/\varepsilon\right)^2 \|f\|_{0,\Omega}, \tag{2.54}$$

and

$$\|u - u_{MFEM}\|_{1,\Omega} \leq C_1 (h/\varepsilon) \|f\|_{0,\Omega}. \tag{2.55}$$

The above bounds are not sharp when $h > \varepsilon$, which is the regime of practical interest. To understand the properties of the method when $h > \varepsilon$, a second asymptotic analysis is carried out in [59] for the limit as $\varepsilon \to 0$. This involves the use of the homogenisation theory of equation (2.41) and is carried out in order to prove that, in the limit as $\varepsilon \to 0$, the MFEM solution converges to the solution of the homogenised problem. This property is not shared by the standard Galerkin FEM, as we have verified with the one–dimensional example above. The result proved in [59] is the following.

Theorem 2.5.2 *Let u and u_h be the weak solutions of (2.41) and (2.53), respectively. Then there exist constants C_1 and C_2, independent of h and ε, such that*

$$\|u - u_{MFEM}\|_{1,\Omega} \leq C_1 h \|f\|_{0,\Omega} + C_2 (\varepsilon/h)^{\frac{1}{2}}. \tag{2.56}$$

Moreover, there exist constants C_1, C_2 and C_3 independent of h and ε, such that

$$\|u - u_{MFEM}\|_{0,\Omega} \leq C_1 h^2 \|f\|_{0,\Omega} + C_2 \varepsilon + C_3 \left(\varepsilon/h\right). \tag{2.57}$$

Similar bounds are proven for the full RFB formulation by Sangalli [87].

The effect of resonance is expressed by the last term on the right–hand side of the error bounds (2.56) and (2.57) and can be appreciated in the loglog plots in Figure 2.2.

It is clear that, if the problem contains many scales, resonance can completely spoil convergence.

For this reason in [59] a nonconforming version of the MFEM method based on an *over–sampling* technique is also developed. This departs from the general augmented space formulation, and hence the analogy with the RFB method is also lost. Sangalli [87] has instead obtained an improved RFB method by considering *macro*–bubbles, i.e. bubbles with support on patches of elements. The advantage of the new method is that the resonance is in terms of the ratio ε/H, where H is the diameter of the patch.

2.6 Computation of the bubbles

The implementation of the RFB method requires the solution of the local problems (2.8) (otherwise the implementation is straight–forward by considering the method in its generalised

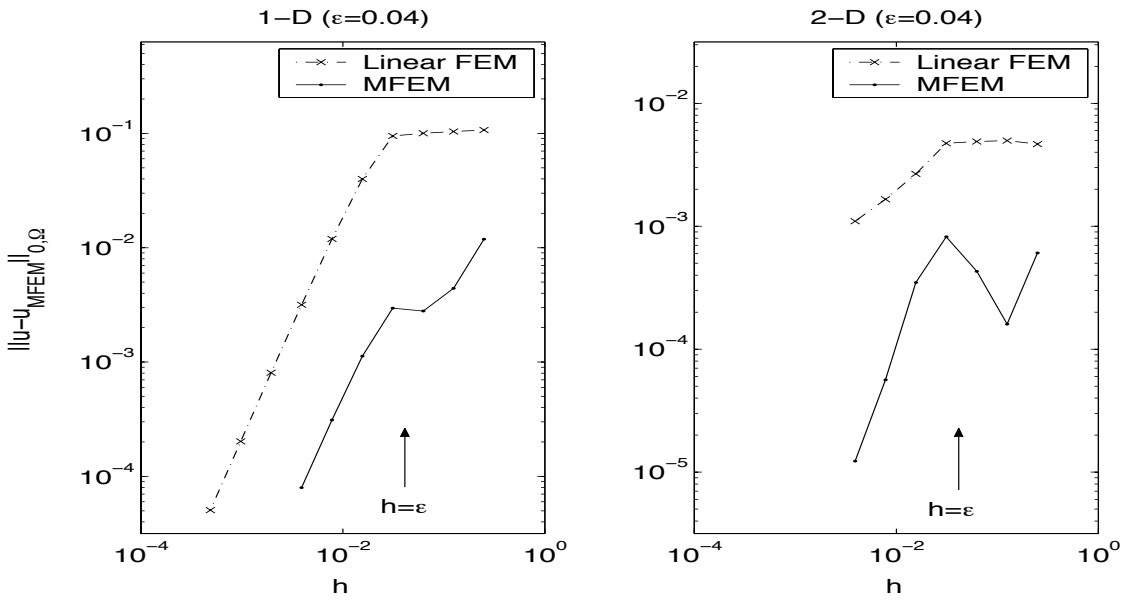


Figure 2.2: L_2–norm error of the MFEM method. The one–dimensional problem (left plot) is the one specified in Figure 2.1. As for the two–dimensional problem (right plot), $\Omega = (0,1)^2$, $d = 1/(2+1.8\sin(2\pi(x-y)/\varepsilon))$, $\varepsilon = 0.04$ and $f = -1/2((6x^2-1)(y^4-y^2)+(6y^2-1)(x^4-x^2))$. In order to ensure that the subgrid discretisation error is of higher order, a 32×32 sub–grid has been used for the two–dimensional computations.

Galerkin formulation (2.11)). More precisely, we need to calculate, for each $T \in \mathcal{T}_h$, a basis of the local bubble space B_T given by (2.9).

As we have seen in Section 2.5, a basis for B_T can be picked as follows. Let $\{\varphi_j\}_{j=1}^{N_T}$ be the local shape functions of the finite element space V_h associated with a generic element T. We define, for $j = 1, \ldots, N_T$, the bubble $b_j \in H_0^1(T)$ as the solution of the problem

$$\begin{cases} Lb_j = -L\varphi_j & \text{in } T, \\ b_j = 0 & \text{on } \partial T, \end{cases} \tag{2.58}$$

and define $b_f \in H_0^1(T)$ as the solution of the problem

$$\begin{cases} Lb_f = f & \text{in } T, \\ b_f = 0 & \text{on } \partial T. \end{cases} \tag{2.59}$$

The local space of residual–free bubbles B_T is then given by

$$B_T = \text{span } \{b_1, \ldots, b_{N_T}, b_f\}.$$

In some cases, the number of bubbles that need to be computed can be reduced since $b_1, \ldots, b_{N_T}$ and b_f need not be linearly independent. For instance, as we have seen, B_T is one–dimensional

for piecewise linear elements and piecewise constant coefficients problems (in the case of bilinear elements the dimension of B_T turns out to be equal to 2).

At any rate the bubbles cannot be computed exactly, and we have no choice but to employ approximate solutions. These may be obtained, for instance, by introducing a subgrid on each $T \in \mathcal{T}_h$, and the question then arises as to how accurate the subgrid computations need to be in order to maintain the accuracy of the method.

An interesting result in this direction has been obtained for convection–diffusion problems by Brezzi and Marini [25], considering the case of linear FEs. Under certain assumptions on the approximability properties of the local problems, it is shown that the error bound (2.29) still holds as long as some nodes of the subgrid used to solve (2.39) are located inside the boundary layer of the local problem.

This suggests the use of very crude Shishkin–type meshes (see Appendix A.2); in fact, this is the method of choice for the computations included in the present thesis. In general any computationally cheap and stable numerical method can be employed; see [29], [50], [48], [23], [27] and [88] for some examples and ideas.

From the result of Brezzi and Marini mentioned above, we deduce that the following general rules should dictate the choice of the approximation of the bubbles:

1. the bubbles need *not* be computed with high accuracy;

2. the approximate bubbles have to 'see' the fine scale that needs to be captured by the method.

Finally, since the bubble problems are completely independent of each other, it is obvious that

3. the bubbles can be computed in parallel.

A nice example highlighting the importance of this last item is given by the upscaling problems considered in the previous section. For such problems, if the subgrid scales are present globally, the (sequential) RFB method would not be faster to compute than the more accurate standard Galerkin method using the subgrids as patches of the global partition. Hence the RFB method should be interpreted as a way of parallelising the Galerkin method (see also [59], [60]).

To clarify the first two items in our list, let us consider again the instructive example of Section 2.4.2. We have seen that the algorithm only requires the computation of the implied SD–parameter τ_T. This, in turn, only depends on the elemental integral mean value of the bubble. Hence, a 'good' bubble is simply one with the correct average.

In this sense, if the Péclet number is large, a good approximation to b_T may be given by the solution $\widetilde{b}_T$ to the reduced (hyperbolic) problem

$$\begin{cases} \mathbf{a} \cdot \nabla \widetilde{b}_T = 1 \quad \text{in} T, \\ \widetilde{b}_T = 0 \quad \text{on} \, \partial T_-. \end{cases} \tag{2.60}$$

As in [29], in the case of $n = 2$, we define h_a to be the length of the longest segment parallel to $\boldsymbol{a}_{|_T}$ and contained in T. The solution $\widetilde{b}_T$ of the reduced problem (2.60) is simply the pyramid of base T and height $h_{\boldsymbol{a}}/|\boldsymbol{a}|$. Thus,

$$\tau_T = \frac{\int_T b_T \, \boldsymbol{x}}{|T|} \sim \frac{\int_T \widetilde{b}_T d\boldsymbol{x}}{|T|} = \frac{1}{|T|}\left(\frac{1}{3}|T|\frac{h_{\boldsymbol{a}}}{|\boldsymbol{a}|}\right) = \frac{h_{\boldsymbol{a}}}{3|\boldsymbol{a}|} = \widetilde{\tau}_T$$

The value $\widetilde{\tau}_T$ is straight–forward to compute and it gives a good approximation of τ_T in the convection–dominated regime; see the discussion on SD–parameters in the next chapter. The above idea is due to Brezzi and Russo [29] and is part of the adaptive algorithm considered in Chapter 5.

Chapter 3

The RFB method on anisotropic partitions

We present the error analysis of the RFB method applied to the solution of boundary value problems for a linear elliptic convection–diffusion equation on anisotropic meshes.

A motivation for this work is based on the fact that not much is known about the coupling of stabilisation methods and anisotropic mesh refinement for the solution of problems with thin layers. Relatively little is known, for example, about the problem of the optimal choice of the SD–parameter in the SUPG method on anisotropic meshes, although there have been interesting developments recently, see [73] and the references therein. Typically the approach is to derive the SD–parameter through *a priori* analysis. In most cases this leads to choosing the SD–parameter to be proportional to the height of the element with respect to its diameter, see, e.g., [5], [67] and [73]. Notably, the authors of [73] have been able to refine their choice by gaining inspiration from the SD–parameter implied by the RFB method. Indeed, in [72] the diameter of the element in the direction of convection is considered (see the end of last chapter). In Section 3.6 we better estimate the magnitude of the stabilisation parameter implied by the RFB method and propose, accordingly, a new choice of the SD–parameter.

The technique used here to obtain an *a priori* error bound for the RFB method is different from the one employed by Brezzi, Marini and Süli [28] to prove Theorem 2.3.1. Indeed, we will follow the approach taken by Sangalli [85] to subsequently re–derive the results presented in [28] without making use of the Besov space setting. The key idea of Sangalli's approach is to exploit the approximation properties of the augmented space V_{RFB}. To do so, Sangalli explicitly constructs a projector from H^1 onto the residual–free bubble space in a certain ε–weighted H^1 norm (ε being the diffusion coefficient). A similar approach is followed by Risch in [81].

3.1 Residual-Free Bubble discretisation on anisotropic triangulations

Let Ω be a bounded open axiparallel polygonal domain in $\mathbb{R}^2$. We consider the elliptic boundary–value problem

$$\begin{cases} -\varepsilon\Delta u + \boldsymbol{a}\cdot\nabla u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{3.1}$$

where ε is a positive parameter, $\boldsymbol{a}$ is a continuously differentiable field on $\bar{\Omega}$ and f belongs to $L^2(\Omega)$. Finally, we assume that

$$\operatorname{div}(\boldsymbol{a}) \le 0 \qquad \text{in } \Omega. \tag{3.2}$$

On $V = H_0^1(\Omega)$, we define the bilinear functional

$$\mathcal{L}(w,v) := \varepsilon \int_\Omega \nabla w\cdot\nabla v\, dx + \int_\Omega (\boldsymbol{a}\cdot\nabla w)\, v\, dx. \tag{3.3}$$

The variational formulation of (3.1) reads

$$\begin{cases} \text{find } u \in V \text{ such that} \\ \mathcal{L}(u,v) = (f,v) \qquad \forall v \in V. \end{cases} \tag{3.4}$$

The bilinear form $\mathcal{L}(\cdot,\cdot)$ is coercive in V since, for every $v \in V$, using the first Green formula and (3.2), we have

$$\begin{aligned} \mathcal{L}(v,v) &= \varepsilon \int_\Omega |\nabla v|^2 dx + \int_\Omega \boldsymbol{a}\cdot\nabla\left(\frac{v^2}{2}\right) dx \\ &= \varepsilon \int_\Omega |\nabla v|^2 dx - \int_\Omega \operatorname{div}(\boldsymbol{a})\,\frac{v^2}{2}\, dx \ge \varepsilon |v|_{1,\Omega}^2. \end{aligned} \tag{3.5}$$

Thus, since $|\cdot|_{1,\Omega}$ is a norm on V, by the Lax-Milgram lemma, (3.4) has a unique solution. (For a more general existence and uniqueness result, see [54], Theorem 8.6).

As in the previous chapter, for a fixed integer $k \ge 1$, we denote by $\mathcal{P}_k$ the space of algebraic polynomials in $\mathbb{R}^2$ of degree $\le k$ and by $\mathcal{Q}_k$ the space of algebraic polynomials in $\mathbb{R}^2$ of degree $\le k$ with respect to each variable; see the definitions (2.26) and (2.27), respectively.

We consider discretisations of (3.4) over *conforming* partitions $\mathcal{T}_h$ of Ω (i.e., any two elements either have a common edge or a common vertex, or they do not intersect at all) consisting of affine–equivalent quadrilateral or triangular elements. Given a reference element $\widehat{T}$, for any $T \in \mathcal{T}_h$ we shall denote by $F_T : \widehat{T} \to T$ the invertible affine transformation from $\widehat{T}$ onto T. (See Section 3.3 for a generalisation to isoparametric finite elements.)

The *residual–free bubble* space is defined as follows (cf. (2.14)):

$$V_{RFB} := \left\{ v \in V : v_{|_e} \in \mathcal{P}_k \text{ for each edge } e \text{ of } T \text{ and any element } T \in \mathcal{T}_h \right\}. \tag{3.6}$$

The RFB–approximation of (3.4) is, by definition, its Galerkin approximation in the space V_{RFB}, which reads

$$\begin{cases} \text{find } u_{RFB} \in V_{RFB} \text{ such that} \\ \mathcal{L}(u_{RFB}, v) = (f, v) \qquad \forall v \in V_{RFB}. \end{cases} \tag{3.7}$$

The space V_{RFB} is infinite–dimensional, admitting the representation

$$V_{RFB} = V_h + B_h, \tag{3.8}$$

where, for rectangular elements

$$V_h = \left\{ v_h \in H_0^1(\Omega) : w_{h|_T} \circ F_T \in \mathcal{Q}_k \quad \forall T \in \mathcal{T}_h \right\},$$

while if the partition consists of triangles, then

$$V_h = \left\{ v_h \in H_0^1(\Omega) : w_{h|_T} \in \mathcal{P}_k \quad \forall T \in \mathcal{T}_h \right\},$$

and

$$B_h = \bigoplus_{T \in \mathcal{T}_h} H_0^1(T).$$

Thus, V_{RFB} is the usual finite element space V_h augmented by the infinite–dimensional space of bubbles B_h.

We shall not assume that the partition $\mathcal{T}_h$ is shape regular, because we wish to allow anisotropic local refinements where special features of the exact solution, such as, boundary layers, are detected. Our only assumption will be that element edges are of at most the same order of magnitude as the diffusion coefficient; more precisely, we assume the existence of a positive constant $c \leq 1$ such that

$$\varepsilon \leq c h_\gamma, \tag{3.9}$$

for all edges γ; here h_γ represents the length of an edge. This is a reasonable assumption when dealing with convection–dominated problems. Moreover, it is clear that, if we can afford to solve the problem on finer grids, we do not need to use a stabilised method.

For any element $T \in \mathcal{T}_h$, let h_T be its diameter. We denote by λ_1 and λ_2 some characteristic dimensions of a generic element $T \in \mathcal{T}_h$ to be defined on a case–by–case basis, and use them to group the elements according to the following rule (which defines the sub–partitions $\mathcal{T}_1$ and $\mathcal{T}_2$):

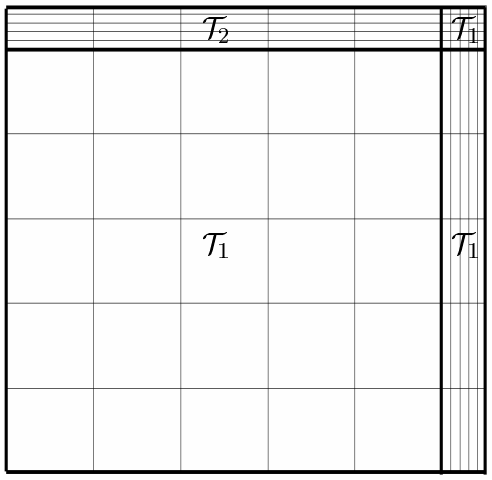


Figure 3.1: An anisotropic mesh designed to deal with two exponential boundary layers.

1. $T \in \mathcal{T}_1$ if $\lambda_1 \leq \lambda_2$

2. $T \in \mathcal{T}_2$ if $\lambda_2 < \lambda_1$,

An admissible structured mesh and its sub–partitions are shown in Figure 3.1.

In conducting the local *a priori* error analysis, we shall distinguish between the sub–partitions. In Chapter 2 we have seen that Brezzi *et al.* in [28] have proven that, if the mesh is shape regular and the solution $u \in H^{k+1}(\Omega)$, the energy norm of the error in the RFB solution satisfies

$$\varepsilon^{1/2}|u - u_{RFB}|_{1,\Omega} = O(h^{k+1/2}),$$

this result being a consequence of the usual scaling properties (here h represents the characteristic size of the partition). We would like to obtain the same order of convergence whenever the partition is shape regular. But, since shape regularity is not assumed, we make use of anisotropic approximation results to derive an *a priori* error bound in terms of appropriately weighted norms of directional derivatives of the exact solution u.

3.2 A–priori error analysis: structured quadrilateral meshes

We start with the case of rectangular elements, leaving the treatment of more general triangulations to subsequent sections.

In this case it is natural to define $\lambda_1 = h_1$ and $\lambda_2 = h_2$, where h_1 and h_2 denote the dimensions of the generic element $T \in \mathcal{T}$ in the x_1 and x_2 coordinate directions, respectively.

3.2.1 Preliminary results

Let $\widehat{T} = [-1,1]^2$ be the master element. Given a function $v \in H^1(T)$, we indicate by $\hat{v} \in H^1(\widehat{T})$ the function associated to v through the affine transformation F_T, hence

$$\hat{v} := v \circ F_T.$$

In this chapter we denote by $i^* = 3 - i$ the complementary index to i with respect to the set $\{1,2\}$.

It is easily seen that, for all $v \in H^1(T)$,

$$\|v\|^2_{0,T} = \frac{1}{4} h_1 h_2 \|\hat{v}\|^2_{0,\widehat{T}}, \tag{3.10}$$

$$\left\| \frac{\partial v}{\partial x_i} \right\|^2_{0,T} = \frac{h_{i^*}}{h_i} \left\| \frac{\partial \hat{v}}{\partial \widehat{x}_i} \right\|^2_{0,\widehat{T}}, \qquad i = 1,2. \tag{3.11}$$

We will also need some scaling properties for functions defined over edges of the elements $T \in \mathcal{T}_h$. The trace of functions belonging to the Hilbert space $H^k(T)$ and, more in general, the Sobolev space $W^{k,p}(T)$, is characterised as belonging to the so–called *fractional order* Sobolev space $W^{k-1/p,p}(\partial T)$. This can be defined using the real method of interpolation and, if $p = 2$, coincides with the Besov space (of the same order) mentioned in Section 2.3, see, e.g., Adams and Fournier [3].

The space $W^{s,p}(\partial T)$, $s > 0$, can also be characterised in terms of an intrinsically defined norm. For instance, for every $s \in (0,1)$, the norm and semi–norm of the Sobolev space of fractional order $H^s(\partial T)$ is defined as

$$\begin{aligned} \|v\|_{s,\partial T} &= \left\{ \|v\|^2_{0,\partial T} + \int_{\partial T} \int_{\partial T} \frac{|v(x) - v(y)|^2}{|x-y|^{1+2s}} d\sigma(x)\, d\sigma(y) \right\}^{1/2} \\ &= \left\{ \|v\|^2_{0,\partial T} + |v|^2_{s,\partial T} \right\}^{1/2}, \end{aligned} \tag{3.12}$$

where $d\sigma$ denotes the 1–dimensional curve measure of ∂T. This definition can be extended to portions of ∂T.

The Trace Theorem (again, see, e.g., [3]) ensures that the trace of a function $v \in H^k(T)$ belongs to $H^{k-1/2}(\partial T)$ and that there exists a constant C independent of v such that

$$\text{(Trace Theorem)} \qquad \|v\|_{k-1/2,\partial T} \le C \|v\|_{k,T}. \tag{3.13}$$

For a summary of the definitions and results of functional analysis used here, see [80].

Let γ be an edge of $T \in \mathcal{T}_h$. Scaling the Sobolev seminorm $|\cdot|_{s,\gamma}$ from $\widehat{T}$ to T we have

$$|v|^2_{s,\gamma} = \left(\frac{h_\gamma}{2} \right)^{1-2s} |\hat{v}|^2_{s,\widehat{\gamma}} \quad \forall v \in H^s(\gamma), \tag{3.14}$$

where, as before, $h_\gamma = |\gamma|$. The scaling property (3.14) is used to prove the following anisotropic trace inequalities which are refinements of the usual ones (see Section 4.3) valid for axiparallel domains.

Lemma 3.2.1 *Let $v \in H^1(T)$, where T is an an axiparallel rectangle in $\mathbb{R}^2$ and let γ_i be an edge of T parallel to the i-th axis. The following trace inequalities hold:*

$$\|v\|^2_{0,\gamma_i} \leq \frac{1}{h_{i^*}} \|v\|^2_{0,T} + 2 \|v\|_{0,T} \left\|v_{x_{i^*}}\right\|_{0,T}, \qquad i = 1, 2; \tag{3.15}$$

$$|v|^2_{1/2,\partial T} \leq C \left(\frac{1}{h_1 h_2} \|v\|^2_{0,T} + \frac{h_1}{h_2} \|v_{x_1}\|^2_{0,T} + \frac{h_2}{h_1} \|v_{x_2}\|^2_{0,T} \right), \tag{3.16}$$

where the constant C is independent of h_1 and h_2.

Proof. The proof of (3.15) can be found, for instance, in [53]. To prove (3.16), we apply (3.14) with $s = 1/2$ to scale from ∂T to $\partial\widehat{T}$, the Trace Theorem (3.13) to shift from $\partial\widehat{T}$ to $\widehat{T}$ and, finally, we use (3.10) and (3.11) to scale back to T:

$$\begin{aligned}
|v|^2_{1/2,\partial T} &= |\hat{v}|^2_{1/2,\partial\widehat{T}} \\
&\leq \|\hat{v}\|^2_{1/2,\partial\widehat{T}} \\
&\leq C \|\hat{v}\|^2_{1,\widehat{T}} \\
&= C \left(\|\hat{v}\|^2_{0,\widehat{T}} + \|\hat{v}_{x_1}\|^2_{0,\widehat{T}} + \|\hat{v}_{x_2}\|^2_{0,\widehat{T}} \right) \\
&= C \left(\frac{1}{h_1 h_2} \|v\|^2_{0,T} + \frac{h_1}{h_2} \|v_{x_1}\|^2_{0,T} + \frac{h_2}{h_1} \|v_{x_2}\|^2_{0,T} \right). \quad \square
\end{aligned}$$

The error analysis is based on the following lemma which was proved by Sangalli in [85].

Lemma 3.2.2 *Given a function $\hat{w}_0 \in H^{1/2}(\partial\widehat{T})$ and a real parameter t with $0 < t \leq 1$, there exists $\hat{w} \in H^1(\widehat{T})$ such that $\hat{w} = \hat{w}_0$ on $\partial\widehat{T}$ and*

$$t\,|\hat{w}|^2_{1,\widehat{T}} + t^{-1} \|\hat{w}\|^2_{0,\widehat{T}} \leq C \left(t\,|\hat{w}_0|^2_{1/2,\partial\widehat{T}} + \|\hat{w}_0\|^2_{0,\partial\widehat{T}} \right), \tag{3.17}$$

where the constant C does not depend on t and $\hat{w}_0$.

We recall that V_{RFB} is an infinite–dimensional space; indeed the elements of V_{RFB} are all the extensions to $H^1_0(\Omega)$ of a space of piecewise polynomials defined on the element boundaries. It is this property that motivates the use of Lemma 3.2.2.

The error analysis is based on the construction of a projection operator onto V_{RFB}.

Let us focus on a single element $T \in \mathcal{T}_h$ by considering a function $v \in H^1(T)$. To start with, since the elements of V_{RFB} are piecewise polynomials on the boundary of T, it makes sense to approximate the trace of v on ∂T by using the restriction onto ∂T of an interpolant of v.

Let $\Pi_k v$ be such an interpolant of v. We may complete the construction of an approximation $P_T v$ of v into T by carefully choosing an appropriate extension of $(\Pi_k v)_{|\partial T}$ into V_{RFB}. As we shall see, this is precisely where Lemma 3.2.2 comes into play.

3.2.2 H^1–projection error

Let

$$H^{1,1}(T) := \left\{ v \in L^2(T) : v_{x_1}, v_{x_2}, v_{x_1x_2} \in L^2(T) \right\}.$$

We begin by defining a suitable interpolant from $\mathcal{Q}_k$ of a generic function in $H^{1,1}(T)$. This will be the tensor–product H^1–projection operator Π_k as has been defined in [53] (see also [91]) by means of truncated Legendre expansions through the following definitions.

Definition 3.2.3 *Let L_n denote the Legendre polynomial of degree n defined on $I = (-1, 1)$. We define the* L_2–projection operator

$$\tilde{\pi}_k : L_2(I) \to \mathcal{P}_k(I)$$

by

$$\tilde{\pi}_k v(x) := \sum_{n=0}^{k} a_n L_n(x),$$

where

$$a_n = \frac{2n+1}{2} \int_I v(x) L_n(x) dx.$$

Further, we define the H^1–projection operator

$$\hat{\pi}_k : H^1(I) \to \mathcal{P}_k(I),$$

by setting, for any $v \in H^1(I)$,

$$\hat{\pi}_k v(x) := \int_{-1}^{x} \tilde{\pi}_{k-1}(v')(\eta) d\eta + v(-1), \qquad x \in (-1, 1).$$

The advantage of the above definition is that it can be easily extended to the multi–dimensional setting by means of a tensor–product construction; this is achieved at the cost of assuming some extra regularity.

Definition 3.2.4 *Let $\widehat{T} = (-1, 1)^2$. We define the* tensor-product projection operator

$$\widehat{\Pi}_k : H^{1,1}(\widehat{T}) \to \mathcal{Q}_k(\widehat{T})$$

by

$$\widehat{\Pi}_k := \hat{\pi}_k^{x_1} \circ \hat{\pi}_k^{x_2},$$

where $\hat{\pi}_k^{x_1}$, $\hat{\pi}_k^{x_2}$ denote the one-dimensional H^1-projection operators from Definition 3.2.3 and the superscripts x_i, $i = 1, 2$, indicate the directions in which the one-dimensional projections are applied.

Let F_T be the affine mapping that gives $T \in \mathcal{T}_h$ from $\widehat{T}$. The above definition is easily extended to a generic axiparallel rectangle T as follows.

Definition 3.2.5 *Let $T \in \mathcal{T}_h$. We define the* tensor product projection operator

$$\Pi_k : H^{1,1}(T) \to \mathcal{Q}_k(T)$$

by setting, for any $v \in H^{1,1}(T)$,

$$\Pi_k v := \widehat{\Pi}_k \hat{v} \circ F_T^{-1}.$$

Being of tensor–product type, the projection Π_k allows anisotropic error bounds. Furthermore, it is better–behaved than the L^2–projection operator when bounds on the derivatives of the interpolation error are needed. The relevant properties of Π_k as interpolation operator are summarised in the following lemma.

Lemma 3.2.6 *Let $v \in H^{1,1}(T)$ where T is an axiparallel rectangle. If $v \in H^{r+1}(T)$ with $1 \le r \le k$ then, for any s with $0 \le s \le r$, the following error estimate holds:*

$$\|v - \Pi_k v\|^2_{0,T} \le \Phi_2(k,s) \left(\left(\frac{h_1}{2}\right)^{2s+2} \|\partial_{x_1}^{s+1} v\|^2_{0,T} + \left(\frac{h_2}{2}\right)^{2s+2} \|\partial_{x_2}^{s+1} v\|^2_{0,T} \right)$$
$$+ \Phi_2(k,s-1) \min_{\substack{i,j=1,2 \\ i \ne j}} \left(\frac{h_i}{2}\right)^2 \left(\frac{h_j}{2}\right)^{2s} \|\partial_{x_j}^s \partial_{x_i} v\|^2_{0,T},$$

and, for any $i, j = 1, 2$, $i \ne j$,

$$\|\partial_{x_i}(v - \Pi_k v)\|^2_{0,T} \le \Phi_1(k,s) \left(\frac{h_i}{2}\right)^{2s} \|\partial_{x_i}^{s+1} v\|^2_{0,T} + \Phi_2(k,s-1) \left(\frac{h_j}{2}\right)^{2s} \|\partial_{x_j}^s \partial_{x_i} v\|^2_{0,T},$$

where

$$\Phi_1(k,s) = \left(\frac{\Gamma(k-s+1)}{\Gamma(k+s+1)}\right)^{1/2}, \qquad \Phi_2(k,s) = \frac{\Phi_1(k,s)}{\sqrt{k(k+1)}},$$

and Γ is the Gamma function.

The proof of the interpolation error bounds stated in the above lemma has been given by Georgoulis in [53] where such results are presented in a much more general setting.

3.2.3 Error bounds

We are now ready to complete the construction of the approximation operator in V_{RFB}. Given $v \in H^{1,1}(T)$, our aim is to define $P_T v \in V_{RFB}$ as a suitable extension of $(\Pi_k v)_{|\partial T}$. This is the crucial step in the error analysis, in which the main task is to bound the quantity

$$E_T^P(v) = \varepsilon |v - P_T v|^2_{1,T} + \varepsilon^{-1} \|v - P_T v\|^2_{0,T}. \tag{3.18}$$

Thus, to obtain the correct error bound we shall need a projector onto V_{RFB} with respect to the ε–weighted H^1 norm $\varepsilon^{1/2}|\cdot|_{1,T}+\varepsilon^{-1/2}\|\cdot\|_{0,T}$. This in turn motivates our choice for $P_T v$.

Assume that $T\in\mathcal{T}_i$, with $i\in\{1,2\}$. To bound (3.18) we proceed as follows. Let $P_{\widehat{T}}\hat{v}=P_T v\circ F_T$. Using (3.10) and (3.11) we get

$$\begin{aligned}E_T^P(v)&=\varepsilon\frac{h_i}{h_{i^*}}\|(\hat{v}-P_{\widehat{T}}\hat{v})_{\hat{x}_{i^*}}\|^2_{0,\widehat{I}}+\varepsilon\frac{h_{i^*}}{h_i}\|(\hat{v}-P_{\widehat{T}}\hat{v})_{\hat{x}_i}\|^2_{0,\widehat{I}}+\frac{\varepsilon^{-1}h_{i^*}h_i}{4}\|\hat{v}-P_{\widehat{T}}\hat{v}\|^2_{0,\widehat{I}}\\&\leq Ch_{i^*}\left(\frac{\varepsilon}{h_i}|\hat{v}-P_{\widehat{T}}\hat{v}|^2_{1,\widehat{I}}+\left(\frac{\varepsilon}{h_i}\right)^{-1}\|\hat{v}-P_{\widehat{T}}\hat{v}\|^2_{0,\widehat{I}}\right).\end{aligned}\tag{3.19}$$

As in [85], we choose $P_{\widehat{T}}\hat{v}$, and so $P_T v$, applying Lemma 3.2.2 to $(\hat{v}-\widehat{\Pi}_k\hat{v})_{|_{\partial\widehat{T}}}$ with

$$t=\frac{\varepsilon}{h_i}.$$

Notice that indeed we can make $t\leq 1$ due to assumption (3.9).

Let $\hat{w}_0=(\hat{v}-\widehat{\Pi}_k\hat{v})_{|_{\partial\widehat{T}}}$ and let $\hat{w}\in H^1(\widehat{T})$ be the extension to $\hat{w}_0$ given by Lemma 3.2.2, with respect to $t=\varepsilon/h_i$. We define $P_{\widehat{T}}\hat{v}$ and, consequently, $P_T v$, as follows:

$$\hat{v}-P_{\widehat{T}}\hat{v}=\hat{w},\tag{3.20}$$

$$P_T v=P_{\widehat{T}}\hat{v}\circ F_T^{-1}.\tag{3.21}$$

From (3.17) and (3.19) we have

$$E_T^P(v)\leq C\left(\varepsilon\frac{h_{i^*}}{h_i}|\hat{v}-\widehat{\Pi}_k\hat{v}|^2_{1/2,\partial\widehat{T}}+h_{i^*}\|\hat{v}-\widehat{\Pi}_k\hat{v}\|^2_{0,\partial\widehat{T}}\right).\tag{3.22}$$

We are now in a position to prove the following result.

Lemma 3.2.7 *Let $T\in\mathcal{T}$ and $v\in H^{r+1}(T)\cap H^{1,1}(T)$ with $0\leq r\leq k$. Given $P_T v$ defined as in (3.20) and (3.21), we consider the quantity $E_T^P(v)$ defined by (3.18). If $T\in\mathcal{T}_i$, $i\in\{1,2\}$, then*

$$\begin{aligned}E_T^P(v)\leq&\frac{C}{2^{2r+1}}\Bigg(\Phi_{12}(k,r)\left(h_i^{2r+1}\|\partial^{r+1}_{x_i}v\|^2_{0,T}+\frac{h_{i^*}^{2r+2}}{h_i}\|\partial^{r+1}_{x_{i^*}}v\|^2_{0,T}\right)\\&+\frac{5}{2}\Phi_2(k,r-1)\left(h_i^{2r-1}h_{i^*}^2\|\partial^r_{x_i}\partial_{x_{i^*}}v\|^2_{0,T}+h_ih_{i^*}^{2r}\|\partial_{x_i}\partial^r_{x_{i^*}}v\|^2_{0,T}\right)\Bigg),\end{aligned}\tag{3.23}$$

where $\Phi_{12}(k,r)=2\Phi_1(k,r)+\Phi_2(k,r)/2$.

Proof. Assume that $T\in\mathcal{T}_i$, $i\in\{1,2\}$, and let $\partial_{x_i}T$ and $\partial_{x_{i^*}}T$ be the collection of the edges of T parallel to the x_i and x_{i^*} coordinate directions, respectively. From (3.22), upon returning

to ∂T using (3.14) and applying the trace inequalities of Lemma 3.2.1, we have

$$\begin{aligned}
E_T^P(v) \leq & C\left(\varepsilon\frac{h_{i^*}}{h_i}|v-\Pi_k v|^2_{1/2,\partial T} + 4\|v-\Pi_k v\|^2_{0,\partial_{x_{i^*}}T} + 4\frac{h_{i^*}}{h_i}\|v-\Pi_k v\|^2_{0,\partial_{x_i}T}\right)\\
\leq & C\Bigg(\left(\frac{\varepsilon}{h_i^2}+\frac{1}{h_i}\right)\|v-\Pi_k v\|^2_{0,T} + \varepsilon\frac{h_{i^*}^2}{h_i^2}\|(v-\Pi_k v)_{x_{i^*}}\|^2_{0,T} + \varepsilon\|(v-\Pi_k v)_{x_i}\|^2_{0,T}\\
& + \frac{h_{i^*}}{h_i}\|v-\Pi_k v\|_{0,T}\|(v-\Pi_k v)_{x_{i^*}}\|_{0,T} + \|v-\Pi_k v\|_{0,T}\|(v-\Pi_k v)_{x_i}\|_{0,T}\Bigg)\\
\leq & C\Bigg(\left(\frac{\varepsilon}{h_i^2}+\frac{1}{h_i}\right)\|v-\Pi_k v\|^2_{0,T}\\
& + \left(\varepsilon\frac{h_{i^*}^2}{h_i^2}+\frac{h_{i^*}^2}{h_i}\right)\|(v-\Pi_k v)_{x_{i^*}}\|^2_{0,T} + (\varepsilon+h_i)\|(v-\Pi_k v)_{x_i}\|^2_{0,T}\Bigg).
\end{aligned}$$

With assumption (3.9) this bound may be written

$$E_T^P(v) \leq C\left(\frac{1}{h_i}\|v-\Pi_k v\|^2_{0,T} + \frac{h_{i^*}^2}{h_i}\|(v-\Pi_k v)_{x_{i^*}}\|^2_{0,T} + h_i\|(v-\Pi_k v)_{x_i}\|^2_{0,T}\right).$$

Thus we have bounded $E_T^P(v)$ in terms of the interpolation error. Finally, assuming that $v \in H^{r+1}(T)\cap H^{1,1}(T)$ with $0 \leq r \leq k$ and using the interpolation error bounds stated in Lemma 3.2.6, we get (3.23). □

We are now ready to prove the following *a priori* error bound in the energy norm $\varepsilon^{1/2}|\cdot|_{1,\Omega}$.

Theorem 3.2.8 *Let $u \in V$ be the solution of (3.4) and $u_{RFB} \in V_{RFB}$ the solution of the residual–free bubble problem (3.7). Assume that the partition $\mathcal{T}_h$ consists of axiparallel rectangles and that there exists a constant $c \in (0,1]$ such that, for any $T \in \mathcal{T}_h$, $\varepsilon \leq c\min\{h_1,h_2\}$. Finally, let $\mathcal{T}_1$ be the sub–partition given by all $T \in \mathcal{T}_h$ such that $h_1 \leq h_2$, and let $\mathcal{T}_2 = \mathcal{T}_h \setminus \mathcal{T}_1$.*

If $u \in H^{k+1}(\Omega)\cap H_0^1(\Omega)$, then there exists a positive constant C, independent of the mesh dimensions, of k and of ε, such that, for any $0 \leq r \leq k$,

$$\begin{aligned}
\varepsilon^{1/2}|u-u_{RFB}|_{1,\Omega} \leq C\frac{\bar{\Phi}(k,r)}{2^{r+1/2}}\sum_{i=1}^{2}\Bigg(\sum_{T\in\mathcal{T}_i}\Big(& h_i^{2r+1}\|\partial_{x_i}^{r+1}u\|^2_{0,T} + \frac{h_{i^*}^{2r+2}}{h_i}\|\partial_{x_{i^*}}^{r+1}u\|^2_{0,T}\\
& + h_i h_{i^*}^{2r}\|\partial_{x_i}\partial_{x_{i^*}}^{r}u\|^2_{0,T} + h_i^{2r-1}h_{i^*}^2\|\partial_{x_i}^{r}\partial_{x_{i^*}}u\|^2_{0,T}\Big)\Bigg)^{1/2}
\end{aligned} \tag{3.24}$$

where $\bar{\Phi}(r,k) = \max\{\Phi_{12}(k,r), \frac{5}{2}\Phi_2(k,r-1)\}$. The constant C only depends on the constant in the trace inequality (3.16) and on the constant in Lemma 3.2.2.

Proof. Let P be the V_{RFB} projector defined on the whole domain Ω using the local projector P_T defined above; that is, for $v \in H^2(\Omega)\cap H_0^1(\Omega)$ we define $Pv_{|T} = P_T v_{|T}$. We consider the decomposition $u - u_{RFB} = (u-Pu) + (Pu - u_{RFB})$.

Using the inequality (3.5) and Galerkin orthogonality, we have that

$$\begin{aligned}\varepsilon|u-u_{RFB}|^2_{1,\Omega} &\leq \mathcal{L}(u-u_{RFB},u-u_{RFB})\\ &= \mathcal{L}(u-u_{RFB},u-Pu).\end{aligned}$$

Furthermore, on applying the Cauchy–Schwarz inequality to the explicit definition of the bilinear form (3.3), we get

$$\begin{aligned}\varepsilon|u-u_{RFB}|^2_{1,\Omega} &\leq \sum_{T\in\mathcal{T}_h}\left(\varepsilon\int_T \nabla(u-u_{RFB})\cdot\nabla(u-P_Tu)\,dx\right.\\ &\qquad\left.+\int_T \boldsymbol{a}\cdot\nabla(u-u_{RFB})(u-P_Tu)\,dx\right)\\ &\leq C\sum_{T\in\mathcal{T}_h}\left(\varepsilon^{1/2}|u-u_{RFB}|_{1,T}\right)\left(\varepsilon^{1/2}|u-P_Tu|_{1,T}+\varepsilon^{-1/2}\|u-P_Tu\|_{0,T}\right)\\ &\leq C\varepsilon^{1/2}|u-u_{RFB}|_{1,\Omega}\left(\sum_{T\in\mathcal{T}_h}\left(\varepsilon^{1/2}|u-P_Tu|_{1,T}+\varepsilon^{-1/2}\|u-P_Tu\|_{0,T}\right)^2\right)^{1/2}.\end{aligned}$$

Next, we split the sum in the right–hand side among the sub–partitions $\mathcal{T}_1$ and $\mathcal{T}_2$ to obtain

$$\varepsilon^{1/2}|u-u_{RFB}|_{1,\Omega}\leq C\sum_{i=1,2}\Big(\sum_{T\in\mathcal{T}_i}E_T^P(u)\Big)^{1/2},$$

with $E_T^P(u)$ as in (3.18). The required bound now follows from the projection error bound (3.23). □

Remark. When the problem (3.1) is strongly convection–dominated, the solution is highly anisotropic. For this reason it is crucial that the error is bounded by appropriately weighted norms of directional derivatives of the solution, as it is for our error bound (3.24). Notice that, if the partition $\mathcal{T}_h$ is shape regular, since we are assuming that $\varepsilon\leq c\min\{h_1,h_2\}$, the *a priori* error bound (3.24) implies the error bound presented in Section 2.

3.3 Extension to isoparametric finite elements

We can extend the results discussed in the previous section to general quadrilateral partitions of Ω by considering isoparametric finite elements (see [33], Section 4.3).

Similarly to the case of affine–equivalent finite elements, a (Lagrange) *isoparametric finite element* $(\widetilde{T},\widetilde{P},\widetilde{N})$ is constructed by starting from a reference standard Lagrange finite element $(\widehat{T},\widehat{P},\widehat{N})$ (see, e.g., Ciarlet [33], pp. 80–81), through a mapping $F_{\widetilde{T}}:\widehat{T}\longrightarrow\widetilde{T}$ assuming that

all components of $F_{\widetilde{T}}$ belong to $\widehat{P}$. The finite element is then given by:

$$\begin{aligned}
&\widetilde{T} = F_{\widetilde{T}}(\widehat{T}),\\
&\widetilde{P} = \{p : \widetilde{T} \to \mathbb{R} : \quad p = \hat{p} \circ F_{\widetilde{T}}^{-1}, \quad \text{for some } \hat{p} \in \widehat{P}\},\\
&\widetilde{N} = \{n_{\hat{a}} : \ \begin{array}{l}\widetilde{P} \to \mathbb{R}\\ p \to p(F_{\widetilde{T}}(\hat{a}))\end{array}, \quad \text{for each node } \hat{a} \text{ of } \widehat{T}\}.
\end{aligned}$$

We remark that the mapping $F_{\widetilde{T}}$ is not necessarily invertible: this has to be assumed.

A family of finite elements $(\widetilde{T}, \widehat{P}, \widehat{N})$ is said to be an *isoparametric* family if all its elements are isoparametrically equivalent to a reference element $\widehat{T}$.

We will consider conforming isoparametric families of finite elements such that any $\widetilde{T} \in \widetilde{\mathcal{T}}_h$ is quadrilateral and $\widehat{P} = \mathcal{Q}_k|_{\widehat{T}}$. These are the natural generalisation to arbitrary quadrilateral meshes of the axiparallel meshes considered in the previous section.

Obviously, with our choice, each basis function $\hat{p}$ of the reference finite element $(\widetilde{T}, \widetilde{P}, \widetilde{N})$ vanishes along each face of any $\widehat{T}$ which does not contain the node associated with $\hat{p}$. This fact ensures the continuity of the finite elements along the edges, and hence we have a *conforming* finite element space

$$\widetilde{V}_h = \left\{ v \in \mathcal{C}^0(\Omega) : v|_T \in \widetilde{P} \quad \forall T \in \widetilde{\mathcal{T}}_h \right\},$$

where $\widetilde{\mathcal{T}}_h = \{\widetilde{T}\}$ is the partition defined by the finite element family $(\widehat{T}, \widehat{P}, \widehat{N})$.

Finally, the RFB space is defined, as before, as the sum

$$\widetilde{V}_{RFB} = \widetilde{V}_h + B_h.$$

To maintain the dimensional separation crucial to the anisotropic analysis, we shall follow the approach proposed in [53]. The isoparametric mapping $F_{\widetilde{T}}$ is decomposed into two mappings $A_T : \widehat{T} \longrightarrow T$ and $Q_T : T \longrightarrow \widetilde{T}$ such that:

1 T is a rectangle of dimensions $h_i > 0$, $i = 1, 2$;

2 A_T is an affine mapping,

3 The Jacobian J_{Q_T} of Q_T satisfies

$$\begin{aligned}
&C_1^{-1} \leq |\det J_{Q_T}| \leq C_1,\\
&\|(J_{Q_T})_{i,j}\|_{\infty,\widetilde{T}} \leq C_2, \qquad \forall i, j = 1, 2,
\end{aligned}$$

for some positive constants C_1, C_2 independent of the size of $\widetilde{T}$.

The last item ensures that the dimensions of $\widetilde{T}$ are similar to those of T, see Figure 3.2. In other words, the stretching from the reference element $\widehat{T}$ is related to the affine transformation A_T, while Q_T accounts for the shape of $\widetilde{T}$. Moreover, the 'intermediate' mesh

$$\mathcal{T}_h = \{T : T = Q_T^{-1}\widetilde{T}, \quad \text{for some } \widetilde{T} \in \widetilde{\mathcal{T}}_h\} \tag{3.25}$$

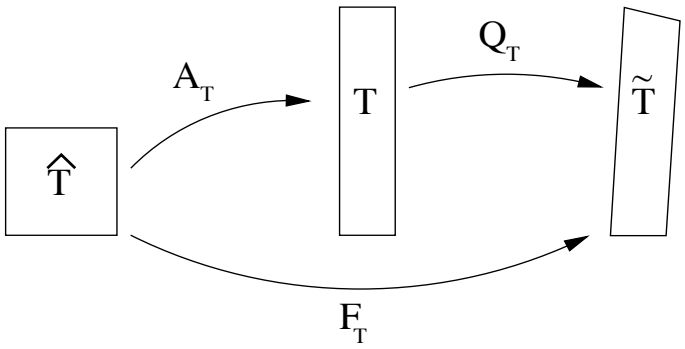


Figure 3.2: Decomposition of the isoparametric mapping $F_{\widetilde{T}} : \widehat{T} \to \widetilde{T}$ into the diagonal affine mapping A_T and the isoparametric dimension–preserving mapping Q_T.

is of the type discussed in the previous section.

The error analysis is carried out as before, mapping the projection error to the reference element. But, in order to obtain bounds which do not mix the dimensions h_1 and h_2, which are ultimately the dimensions of $\widetilde{T}$, when scaling back we stop at the 'intermediate' mesh $\mathcal{T}_h$.

Theorem 3.3.1 *Let $\tilde{u} \in V$ be the solution of (3.4) and $\tilde{u}_{RFB} \in \widetilde{V}_{RFB}$ the solution of the residual–free bubble problem (3.7). Assume that $(\widetilde{T}, \widetilde{P}, \widetilde{N})$ is a conforming family of isoparametric finite elements and let $(\widehat{T}, \widehat{P}, \widehat{N})$ be the reference finite element. Given the isoparametric mapping $F_{\widetilde{T}} : \widehat{T} \to \widetilde{T}$, let $F_{\widetilde{T}} = Q_T \circ A_T$ with Q_T and A_T satisfying conditions 1, 2 and 3 above. Moreover, let $\mathcal{T}_h$ be the partition (3.25) related to $\widetilde{\mathcal{T}}_h$ through Q_T and assume the existence of a constant $c \in (0,1]$ such that, for any $T \in \mathcal{T}_h$, $\varepsilon \leq c \min\{h_1, h_2\}$. Finally, let $\mathcal{T}_1$ be the sub–partition given by all $T \in \mathcal{T}_h$ such that $h_1 \leq h_2$, and let $\mathcal{T}_2 = \mathcal{T}_h \setminus \mathcal{T}_1$.*

If $\tilde{u} \in H^{k+1}(\Omega) \cap H_0^1(\Omega)$, then there exists a positive constant C, independent of the mesh dimensions, of k and of ε, such that, for any $0 \leq r \leq k$,

$$\varepsilon^{1/2} |\tilde{u} - \tilde{u}_{RFB}|_{1,\Omega} \leq C \frac{\bar{\Phi}(k,r)}{2^{r+1/2}} \sum_{i=1}^{2} \left(\sum_{T \in \mathcal{T}_i} \left(h_i^{2r+1} \|\partial_{x_i}^{r+1} u\|_{0,T}^2 + \frac{h_{i^*}^{2r+2}}{h_i} \|\partial_{x_{i^*}}^{r+1} u\|_{0,T}^2 \right.\right.$$
$$\left.\left. + h_i h_{i^*}^{2r} \|\partial_{x_i} \partial_{x_{i^*}}^{r} u\|_{0,T}^2 + h_i^{2r-1} h_{i^*}^2 \|\partial_{x_i}^{r} \partial_{x_{i^*}} u\|_{0,T}^2 \right) \right)^{1/2} \tag{3.26}$$

where for every $T \in \mathcal{T}_h$, $u = \tilde{u}|_{\widetilde{T}} \circ Q_T$ and $\bar{\Phi}(r,k) = \max\{\Phi_{12}(k,r), \frac{5}{2}\Phi_2(k,r-1)\}$. The constant C only depends on the constant in the trace inequality (3.16) and on the constant in Lemma 3.2.2.

Proof. We scale the projection error, say $E_{\widetilde{T}}^P(\tilde{u})$, from the element $\widetilde{T}$ to the intermediate element T and then simply refer to the proof in Section 3.2. In order to do so, we need to make sure that the space $\widetilde{V}_{RFB}|_{\widetilde{T}}$ is transformed into $V_{RFB}|_T = V_h|_T + H_0^1(T)$. Certainly

$V_h|_T = \mathcal{Q}_k|_T$ as before. As for the bubble component, this is preserved since the mapping Q_T is smooth with a smooth inverse (see Theorem 4.3.2 in [33]).

Let $\tilde{v} \in V$ and $P_{\widetilde{T}} = P_T \circ Q_T^{-1}$. We consider the quantity $E^P_{\widetilde{T}}(v)$ given by

$$\begin{aligned} E^P_{\widetilde{T}}(\tilde{v}) &= \varepsilon|\tilde{v} - P_{\widetilde{T}}\tilde{v}|^2_{1,\widetilde{T}} + \varepsilon^{-1}\|\tilde{v} - P_{\widetilde{T}}\tilde{v}\|^2_{0,\widetilde{T}} \\ &\leq C\left(\varepsilon|v - P_T v|^2_{1,T} + \varepsilon^{-1}\|v - P_T v\|^2_{0,T}\right) = C E^P_T(v), \end{aligned}$$

where the constant C only depends on the Jacobian of Q_T and its inverse (this is easy to check using the chain rule), and $v = \tilde{v} \circ Q_T$. We now conclude using (3.23) as before to bound $E^P_T(\tilde{u}_{RFB} \circ Q_T)$ and applying again the argument in the proof of Theorem 3.2.8 to the RFB solution $\tilde{u}_{RFB}$ belonging to the new space $\widetilde{V}_{RFB}$. $\square$.

3.4 A different approach: dealing with affine triangulations

We now discuss the case of triangulations $\mathcal{T}_h$ consisting of affine–equivalent (triangular or quadrilateral) elements. As before, our only assumption on the triangulation is conformity, i.e that any two elements in $\mathcal{T}_h$ either have a common edge or a common vertex, or they do not intersect at all.

The following *a priori* error analysis is based on Lemma 3.2.2 and on the technique introduced by Formaggia & Perotto [44] (see also the references therein and Micheletti, Perotto and Picasso [73]), to prove anisotropic error estimates for the interpolation error. More precisely, we will employ some *ad hoc* scaling properties derived in [44] in terms of some characteristic quantities of the affine transformation $F_T : \widehat{T} \longrightarrow T$ which we will define in a moment, where $\widehat{T}$ is the reference element (the unit simplex or the square $[-1,1]^2$) and T is a generic triangle in $\mathcal{T}_h$.

Notice that [44] and [73] deal with triangular elements, but the tools which we borrow from those papers are generally valid as long as the mapping F_T is affine. On the other hand, a limitation of the approach is that only an *a priori* error bound in terms of the H^2–seminorm can be obtained, so this analysis is relevant only in the case of $\mathcal{P}_1$ and $\mathcal{Q}_1$ finite elements.

Let $F_T(\hat{\boldsymbol{x}}) = M\hat{\boldsymbol{x}} + \boldsymbol{t}$ (we omit the dependence of M and t on T to simplify the notation). The matrix M is invertible, hence it admits a unique *polar decomposition* into $M = BZ$, where B is symmetric and positive definite and Z is orthonormal.

Further, B is factorized as $B = R^T \Lambda R$, where Λ is diagonal with positive decreasing entries (the eigenvalues of B) and R is orthonormal (with rows which are the eigenvectors of B). Hence,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \qquad R = \begin{bmatrix} \boldsymbol{r}_1^T \\ \boldsymbol{r}_2^T \end{bmatrix},$$

where $\lambda_1 \geq \lambda_2$ and $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ are the eigenvalues and eigenvectors of B, respectively.

The eigenvalues λ_1 and λ_2 of the scaling transformation replace h_1 and h_2 of the previous sections in describing the dimensions of the element T, with the advantage of being independent of the coordinate axes (we notice that, if the mesh is axiparallel, then the two definitions coincide up to a constant depending on the reference element).

With this new notation, we get the following scaling rules, which are the counterparts of (3.10) and (3.11):

$$\|v\|_{0,T}^2 = \lambda_1\lambda_2\|\hat{v}\|_{0,\widehat{T}}^2 \tag{3.27}$$

$$|v|_{1,T}^2 \leq \frac{\lambda_1}{\lambda_2}|\hat{v}|_{1,\widehat{T}}^2. \tag{3.28}$$

The equality (3.27) is elementary. As for (3.28), this is proved in [44] as Lemma 2.2.

To scale back from the reference element we shall use the following identity which is Lemma 2.2 in [73], see also the proof of Lemmas 2.1 and 2.1 in [44]:

$$|\hat{v}|_{2,\widehat{T}}^2 = \frac{\lambda_1^3}{\lambda_2}L_{11}v + \frac{\lambda_2^3}{\lambda_1}L_{22}v + 2\lambda_1\lambda_2 L_{12}v, \tag{3.29}$$

where

$$L_{ij}v = \int_T \left(\boldsymbol{r}_i^T H(v)\boldsymbol{r}_j\right)^2 d\boldsymbol{x}, \qquad \text{with } i,j = 1,2, \tag{3.30}$$

and $H(v)$ is the Hessian matrix associated with the function v, that is,

$$H(v) = \begin{bmatrix} \frac{\partial^2 v}{\partial x_1^2} & \frac{\partial^2 v}{\partial x_1 \partial x_2} \\ \frac{\partial^2 v}{\partial x_1 \partial x_2} & \frac{\partial^2 v}{\partial x_2^2} \end{bmatrix}.$$

Theorem 3.4.1 *Let $u \in V$ be the solution of (3.4) and $u_{RFB} \in V_{RFB}$ the solution of the residual–free bubble problem (3.7). Consider a conforming affine–equivalent triangulation $\mathcal{T}_h$ assuming that there exists a constant $c \in (0,1]$ such that, for every $T \in \mathcal{T}_h$, $\varepsilon \leq c\lambda_2$, where $\lambda_1 \geq \lambda_2$ are the characteristic dimensions of T defined above.*

If $u \in H^2(\Omega) \cap H_0^1(\Omega)$, then there exists a positive constant C independent of the mesh dimensions and of ε such that

$$\varepsilon^{1/2}|u - u_{RFB}|_{1,\Omega} \leq C\left(\sum_{T\in\mathcal{T}_h}\left(\frac{\lambda_1^4}{\lambda_2}L_{11}u + \lambda_2^3 L_{22}u + 2\lambda_1^2\lambda_2 L_{12}u\right)\right)^{1/2}, \tag{3.31}$$

where the terms L_{ij}, $i,j = 1,2$, are defined elementwise as in (3.30) in terms of the Hessian of u.

Proof. Let $T \in \mathcal{T}_h$. As in the previous sections, we need to bound the quantity given by equation (3.18), that is,

$$E_T^I(v) = \varepsilon|v - P_T v|_{1,T}^2 + \varepsilon^{-1}\|v - P_T v\|_{0,T}^2,$$

where $v \in V$ and $P_T v = Pv|_T$ is to be chosen under the constraint that $Pv \in V_{RFB}$. As before, we start by scaling $E_T^I(v)$ to the reference element $\widehat{T}$. Using (3.27) and (3.28) we get:

$$\begin{aligned} E_T^I(v) &\leq \varepsilon \frac{\lambda_1}{\lambda_2} |\hat{v} - P_{\widehat{T}}\hat{v}|^2_{1,\widehat{T}} + \varepsilon^{-1}\lambda_1\lambda_2 \|\hat{v} - P_{\widehat{T}}\hat{v}\|^2_{0,\widehat{T}} \\ &= \lambda_1 \left(\frac{\varepsilon}{\lambda_2} |\hat{v} - P_{\widehat{T}}\hat{v}|^2_{1,\widehat{T}} + \left(\frac{\varepsilon}{\lambda_2}\right)^{-1} \|\hat{v} - P_{\widehat{T}}\hat{v}\|^2_{0,\widehat{T}} \right). \end{aligned} \tag{3.32}$$

We then apply Lemma 3.2.2, this time to $(\hat{v} - \hat{\pi}_k\hat{v})_{|\partial\widehat{T}}$, where $\hat{\pi}_k$ is the standard Lagrange interpolant defined on the reference triangle $\widehat{T}$, and with $t = \varepsilon/\lambda_2$. In this way we get

$$E_T^I(v) \leq C \left(\varepsilon \frac{\lambda_1}{\lambda_2} |\hat{v} - \hat{\pi}_k\hat{v}|^2_{1/2,\partial\widehat{T}} + \lambda_1 \|\hat{v} - \hat{\pi}_k\hat{v}\|^2_{0,\partial\widehat{T}} \right).$$

Instead of scaling back to the boundary of the element T as was done previously, we now proceed by applying the Trace Theorem (3.13) and the standard Lagrange interpolation error bounds on $\widehat{T}$ (see Ciarlet [33] for the full description or Lemma 4.3.4 for the results used here). Since $\lambda_2 \leq \lambda_1$ and $\varepsilon \leq c\lambda_2$ with $c \in (0, 1]$ we get

$$\begin{aligned} E_T^I(v) &\leq C \left(\varepsilon \frac{\lambda_1}{\lambda_2} + \lambda_1 \right) \|\hat{v} - \hat{\pi}_k\hat{v}\|^2_{1,\widehat{T}} \\ &\leq C\lambda_1 |\hat{v}|^2_{2,\widehat{T}} \\ &\leq C \left(\frac{\lambda_1^4}{\lambda_2} L_{11}v + \lambda_2^3 L_{22}v + 2\lambda_1^2\lambda_2 L_{12}v \right), \end{aligned} \tag{3.33}$$

the last bound being due to (3.29).

The desired error bound now follows by repeating the steps in the proof of Theorem 3.2.8. □

As we mentioned earlier, if the triangulation $\mathcal{T}_h$ is axiparallel, then $\lambda_i = h_i/c_i$, with h_i and c_i, $i = 1, 2$, being the dimensions along the coordinate axes of T and $\widehat{T}$, respectively. In this case Theorem 3.4.1 collapses to the *a priori* error bound (3.24), with the only limitation that $r = 1$.

Finally, if the partition of Ω is affine quadrilateral, a special case of the isoparametric elements discussed in the previous section, then the bound (3.31) can be seen as a more explicit version of the bound obtained previously.

3.5 A numerical example

We consider the following simple model problem with mixed boundary conditions:

$$\begin{cases} -\varepsilon\Delta u + u_{x_2} = 0 & \text{in } \Omega = (0,1)^2, \\ u(x_1, 0) = 0; \quad u(x_1, 1) = 1, & x_1 \in [0,1] \\ u_{x_1} = 0 & \text{on } \Gamma_N = (\{0\} \times (0,1)) \cup (\{1\} \times (0,1)). \end{cases}$$

Due to the homogeneous Neumann condition on the two vertical edges of Ω, this is 'virtually' a 1–D problem, its solution being given by

$$u(\boldsymbol{x}) = \frac{\mathrm{e}^{x_2/\varepsilon} - 1}{\mathrm{e}^{1/\varepsilon} - 1}.$$

The RFB discretisation of boundary value problems with mixed boundary conditions is not dissimilar from that of problems with Dirichlet boundary condition already discussed in detail. To start with, we remove the non–homogeneous Dirichlet boundary condition acting on $x_2 = 1$ by subtracting a function $w \in H^1(\Omega)$ obeying the boundary conditions imposed in the definition of the problem. Having done that, the space V_{RFB} is still given by (3.6) only that, this time, the solution space is given by

$$V = \left\{ v \in H^1(\Omega) : v|_{\Gamma_D} = 0 \right\},$$

where $\Gamma_D = \partial\Omega \setminus \Gamma_N$.

We consider axiparallel uniform rectangular grids $\mathcal{T}_h$ of dimensions h_1 and h_2. Clearly, if $h_2 < h_1$, then we may set $\mathcal{T}_h = \mathcal{T}_2$, if $h_1 < h_2$, then $\mathcal{T}_h = \mathcal{T}_1$ and finally, if $h_1 = h_2$, then $\mathcal{T}_h = \overline{\mathcal{T}}$. Since all derivatives of the solution u in the direction of the x_1–axis are identically zero, the error bound (3.24) simply becomes:

$$\varepsilon^{1/2} |u - u_{RFB}|_{1,\Omega} \leq C \left(\sum_{T \in \overline{\mathcal{T}}} \max \left\{ h_2^3, \frac{h_2^4}{h_1} \right\} \|\partial_{x_2}^2 u\|_{0,T}^2 \right)^{1/2}.$$

The purpose of the experiment is to test the validity of the error bound, either by fixing h_1 while halving h_2 (*correct refinement*), or fixing h_2 while halving h_1 (*wrong refinement*).

Performing the correct refinement is, of course, not too different from solving the related sequence of 1–D problems. The error and error bound are shown in the loglog plot on the left in Figure 3.3, for $\varepsilon = 10^{-2}$. The asymptotic behaviour of the error shows first–order convergence, as predicted by the analysis of Brezzi *et al.* [28] presented in the previous Chapter.

The similarity of the numerical solution to our 2–D model problem with the 1-D numerical solution is lost when the wrong refinement is performed (notice that this does not happen when applying the standard affine Galerkin method). As predicted by the error bound, the accuracy of the solution actually deteriorates under the wrong refinement; see the loglog plot on the right in Figure 3.3. This is due to the peculiar definition of the RFB finite element space. Mesh refinement corresponds to a relative impoverishment of the bubble subspace, and an enrichment of the piecewise polynomial subspace. If the latter subspace, as is the case with our wrong refinement, is ineffective, then the overall approximation properties of V_{RFB} will be worse than on a coarser mesh.

In the limit as $h_1 \to 0$ the solution becomes constant along x_1, that is, it tends to the piecewise $\mathcal{Q}_1$ standard Galerkin solution, which is unaffected by the reduction of h_1, see Figure 3.4. Asymptotically in the case of the wrong refinement (with $h_1 \to 0$) the error is, consequently, $O(1)$ (cf. Figure 3.3 (right)).

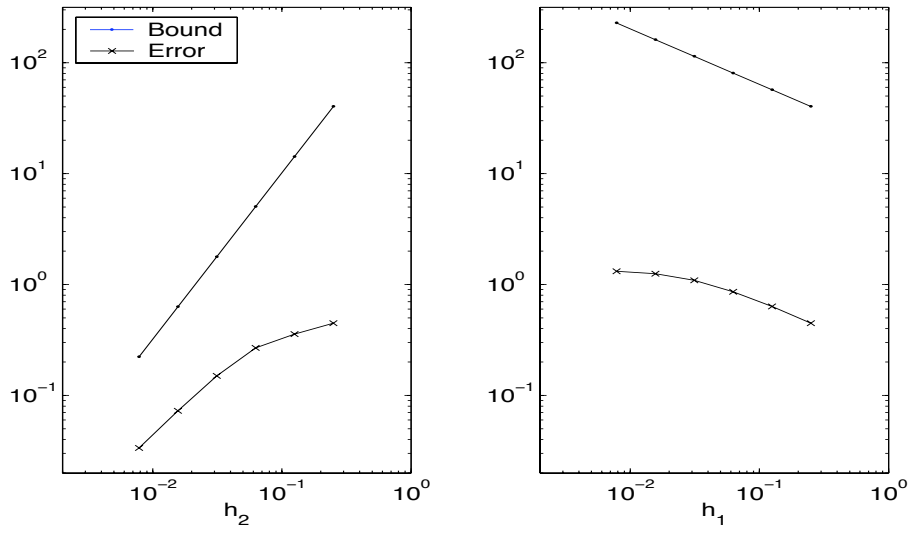


Figure 3.3: $\varepsilon^{1/2}$–weighted H^1–seminorm error and error bound under h_2 refinement (left) and h_1 refinement (right); $\varepsilon = 10^{-2}$

Another way of seeing this is that since the bubble part of the solution is forced to tend to zero as $h_1 \to 0$, its stabilising effect is diminished until, in the limit, it vanishes and the RFB method collapses to the standard Galerkin FEM.

The detailed error analysis of the RFB method on shape–regular triangulations in the next chapter is meant to clarify the approximation properties of the method in the pre–asymptotic regime when $\varepsilon \leq ch$. In particular, we shall relate the phenomenon just observed to the inadequacy of V_h in approximating, along the skeleton of the triangulation, the typical exponential behaviour of the solution in the boundary layer.

3.6 Remarks on the tuning of the SD–parameter

In the introduction to this chapter we pointed out that the analysis of stabilised finite element methods on highly anisotropic meshes is a relatively young field of research. In particular we refered to the critical issue of classical stabilised methods, namely the tuning of the stability parameters; to the best of our knowledge, the first study addressing this topic was the article, published in 1996, by Apel and Lube [5].

Ideally one performs adaptive mesh refinement by starting with a rather coarse mesh on which stabilisation is necessary, possibly turning off the stabilisation locally in the course of the refinement process. In many cases, and convection–dominated diffusion problems are certainly one such case, an effectively refined mesh consists of extremely stretched elements.

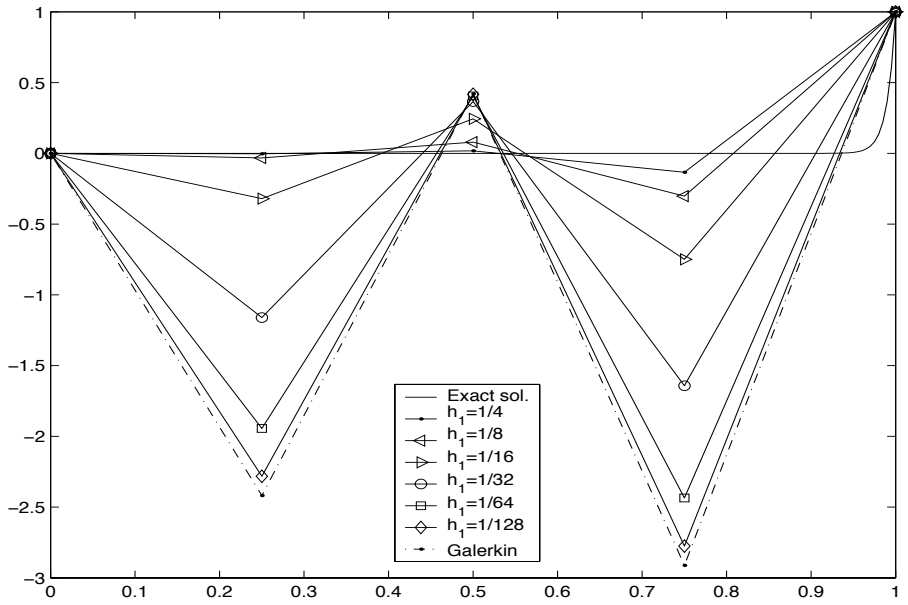


Figure 3.4: Profile of the solution along $x_1 = 0$ under h_1–refinement (as in the right plot in Figure 3.3). The lowest profile represents the piecewise $\mathcal{Q}_1$ standard Galerkin FEM solution. The exact solution is also plotted for comparison.

Let us consider, for example, the boundary–value problem

$$\begin{cases} -\varepsilon\Delta u + (2,1)^T \cdot \nabla u = 0 & \text{in } \Omega = (0,1)^2, \\ u(x_1,0) = u(1,x_2) = 0, & x_1, x_2 \in [0,1], \\ u(x_1,1) = u(0,x_2) = 1, & x_1, x_2 \in [0,1]. \end{cases} \tag{3.34}$$

The solution of (3.34) exhibits an internal layer emanating from the origin of the coordinate axis and a boundary layer situated along $x_1 = 1$. A typical mesh that is appropriately refined to capture the thin layers in the solution is the one depicted in Figure 3.5. The elements are stretched in the direction orthogonal and tangential to the convection field in the vicinity of the boundary and internal (characteristic) layer, respectively.

We have established through *a priori* analysis that the RFB method can be applied successfully on anisotropic triangulations. Since this method is parameter–free, it can be used to assess the choices of stabilisation parameters in classical stabilised methods proposed in the literature.

Let us focus our attention on the design of an optimal SD–parameter for the linear SDFEM on triangular meshes. The name SDFEM (*streamline–diffusion finite element method*) is used to collectively refer to the most studied classical stabilised methods, namely SUPG, GALS and DWG. Indeed, these methods, given by (2.33), (2.34) and (2.35) respectively, are equivalent in the case of linear elements.

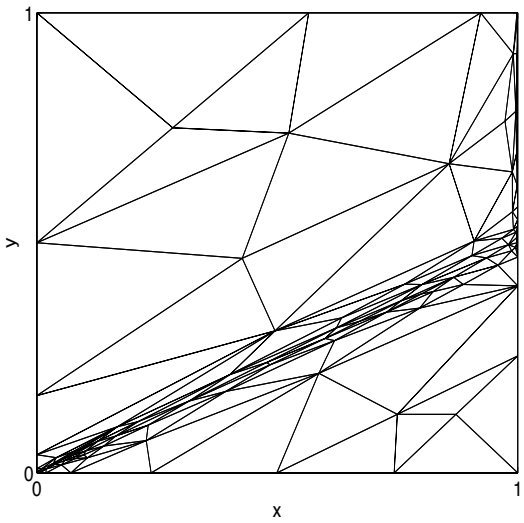


Figure 3.5: A mesh refined anisotropically to capture an internal and a boundary layer. Courtesy of M. Picasso [73].

We have seen in Section 2.4 that under the assumption that the coefficients in the differential equation are piecewise constant over Ω, the linear RFB method reproduces the stabilisation term of the linear SDFEM method, that is

$$\sum_{T\in\mathcal{T}_h} \tau_T \left(\boldsymbol{a}\cdot\nabla u_h, \boldsymbol{a}\cdot\nabla v_h\right)_T, \tag{3.35}$$

with the stabilisation parameter given through the process of static condensation by

$$\tau_T = \frac{1}{|T|}\int_T b_T d\boldsymbol{x}.$$

This choice of τ_T is predetermined by solving (2.39) for b_T in $H_0^1(T)$. In practice, b_T is obtained by solving the local b.v.p. for the bubble numerically, thus yielding approximations to the 'optimal' value τ_T. Here we will consider two instructive choices: a very accurate approximation of the optimal value, still referred to as τ_T, and the value obtained by solving the reduced (locally hyperbolic) problem (2.60), namely

$$\tilde{\tau}_T = \frac{1}{|T|}\int_T \tilde{b}_T\, d\boldsymbol{x} = \frac{h_{\boldsymbol{a}}}{3|\boldsymbol{a}|},$$

where $h_{\boldsymbol{a}}$ is the elemental diameter in the direction of $\boldsymbol{a}$.

On shape regular meshes it is well established (see, for example, Section 2.4 or [82]) that the optimal SD–parameter should be proportional to the characteristic dimension h_T of the element in the convection–dominated regime and proportional to h_T^2/ε in the diffusion dominated regime. Moreover, the correct switch between the two regimes is given by the local

Péclet number. For example in [72], following [46], it is defined as:

$$Pe_T := h_T \frac{\|\boldsymbol{a}\|_{\infty,T}}{6\varepsilon}; \qquad \tau_T^{sr} = h_T \frac{\min(Pe_T, 1)}{2\|\boldsymbol{a}\|_{\infty,T}}.$$

The RFB coefficient τ_T has precisely the same behaviour; see the example below.

Turning to anisotropic meshes, since for every element we have two characteristic dimensions λ_1 and λ_2, there is no obvious definition of the mesh Péclet number in the first place.

The *a priori* analysis performed in [5], [67] and [73] leads to the replacement of h_T with the shorter dimension of the element, namely λ_2. For example in [73] we find the following definition:

$$Pe_T := \lambda_2 \frac{\|\boldsymbol{a}\|_{\infty,T}}{6\varepsilon}; \qquad \tau_T^{(1)} = \lambda_2 \frac{\min(Pe_T, 1)}{2\|\boldsymbol{a}\|_{\infty,T}}.$$

Considering the definition of $\tilde{\tau}_T$, we see that the RFB method suggests, instead, to base the definition of the mesh Péclet number on $h_{\boldsymbol{a}}$. Inspired by this observation, the authors of [73] refined their analysis (see [72]) and defined the new characteristic length

$$\lambda_{\boldsymbol{a}} = \frac{1}{\|\Lambda^{-1} R \frac{\boldsymbol{a}}{|\boldsymbol{a}|}\|_{\infty,T}},$$

which is comparable with $h_{\boldsymbol{a}}$. The Péclet number is then chosen accordingly. As for the SD–parameter, this is still required to behave like λ_2^2/ε in the diffusion–dominated regime, while in the convection–dominated regime it is chosen to depend on $\lambda_{\boldsymbol{a}}$:

$$Pe_T := \lambda_{\boldsymbol{a}} \frac{\|\boldsymbol{a}\|_{\infty,T}}{6\varepsilon}; \qquad \tau_T^{(2)} = \lambda_a \frac{1}{2\|\boldsymbol{a}\|_{\infty,T}} \cdot \begin{cases} \frac{\lambda_2^2}{\lambda_{\boldsymbol{a}}^2} Pe_T & \text{if} \quad Pe_T < 1, \\ 1 & \text{if} \quad Pe_T \geq 1. \end{cases}$$

This definition is dependent on the direction of $\boldsymbol{a}$: $\tau_T^{(2)}$ is qualitatively different from (i.e. larger than) $\tau_T^{(1)}$ if the triangle T is aligned with the convection field. This case corresponds to the wrong refinement discussed in the previous section if T is in the boundary layer, while it corresponds to a correct placement of the element T if this is in the proximity of a characteristic layer.

Hence we expect the algorithm using $\tau_T^{(2)}$ to be more robust (stable) under the wrong refinement mentioned above, possibly over–stabilising in the presence of characteristic layers. Otherwise the two choices are qualitatively indistinguishable (see below).

Here we propose a further definition with the same asymptotic behaviour as $\tau_T^{(2)}$, but with a different turning point, namely:

$$Pe_T := \frac{\lambda_2^2}{\lambda_{\boldsymbol{a}}} \frac{\|\boldsymbol{a}\|_{\infty,T}}{6\varepsilon}; \qquad \tau_T^{(3)} = \lambda_a \frac{\min(Pe_T, 1)}{2\|\boldsymbol{a}\|_{\infty,T}}.$$

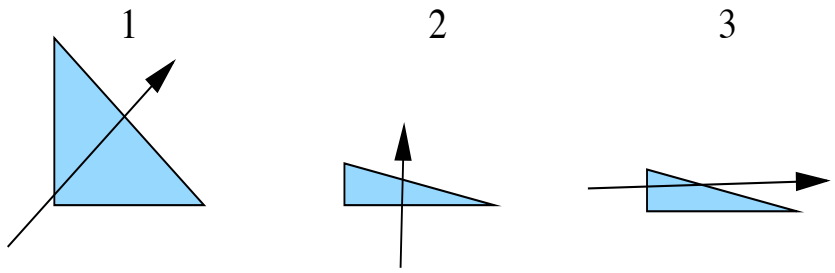


Figure 3.6: Sketch of the triangular shapes and velocity fields $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$ and their mutual orientations: 1. Isosceles triangle with $\alpha = \pi/4$, 2. Right–angle triangle with $\alpha = \pi/2 - 1°$, 3. Right–angle triangle with $\alpha = 1°$.

This choice is closer than $\tau_T^{(2)}$ to the 'optimal' RFB parameter τ_T. The reason for this is ultimately related to the fact that $\tilde{\tau}_T$ is a bad approximation to τ_T on triangles that are stretched in the direction of the convection field: it turns out that in this case $\tau_T^{(2)}$ is closer to $\tilde{\tau}_T$ while $\tau_T^{(3)}$ follows τ_T.

We can see this by calculating τ_T, $\tilde{\tau}_T$, $\tau_T^{(1)}$, $\tau_T^{(2)}$ and $\tau_T^{(3)}$ on a sample of triangular shapes and convective fields.

Let T be a right–angle triangle oriented along the coordinate axes, so that $\lambda_1 = h_1$ and $\lambda_2 = h_2$ are its characteristic dimensions. We calculate the SD–parameters defined above on a sequence of triangles T obtained by halving of h_2 starting from $h_2 = 1/2$. We do this three times in succession, while varying h_1 and the field $\boldsymbol{a} = (\cos\alpha, \sin\alpha)$ as follows (see Figure 3.6):

1. $h_1 = h_2$ and $\alpha = \pi/4$ (isosceles shape–regular triangle);
2. $h_1 = 1/2$ and $\alpha = \pi/2 - 1°$ (right–angle triangle stretched orthogonally to $\boldsymbol{a}$);
3. $h_1 = 1/2$ and $\alpha = 1°$ (right–angle triangle stretched in direction of $\boldsymbol{a}$).

This sample is quite comprehensive. The three experiments correspond to the three loglog plots in Figure 3.7. In all cases the value $\varepsilon = 10^{-4}$ is used for the diffusion coefficient.

In the first two experiments all values $\tau_T^{(i)}$, $i = 1, 2, 3$, behave as τ_T (we have not plotted $\tau_T^{(3)}$ in this cases because it coincides with τ_2). While, as it should be expected, $\tilde{\tau}_T$ is close to τ_T only in the convection–dominated regime.

As we suggested earlier, $\tilde{\tau}_T$ separates from τ_T much earlier when the element is stretched in the direction of $\boldsymbol{a}$ (see Figure 3.8 for a zoom of the bottom plot in Figure 3.7), hence showing that the concept of *regime* (and *local Péclet number*) should indeed be dependent on the relative orientation of the triangle and the convective field.

Consequently, only in this case, do $\tau_T^{(2)}$ and $\tau_T^{(3)}$ differ (as for $\tau_T^{(1)}$, this is definitely too small: after all, the fact that $\tau_T^{(1)}$ under–stabilizes was the reason for introducing the new parameters). As an aside, we note that $\tau_T^{(3)}$ follows more smoothly τ_T if in its definition $\lambda_{\boldsymbol{a}}$ is substituted with $h_{\boldsymbol{a}}$.

We have tested the two choices $\tau_T^{(2)}$ and $\tau_T^{(3)}$ by solving the boundary–value problem (3.34) using the SDFEM on the mesh in Figure 3.5. The elevation and contour plots of the solution are shown in Figure 3.9. Since the mesh is correctly refined, there is no obvious difference between the two numerical solutions; in fact the solution obtained by using $\tau_T^{(3)}$ is slightly less diffused and more oscillatory across the internal layer (cf. the contour plots in Figure 3.9).

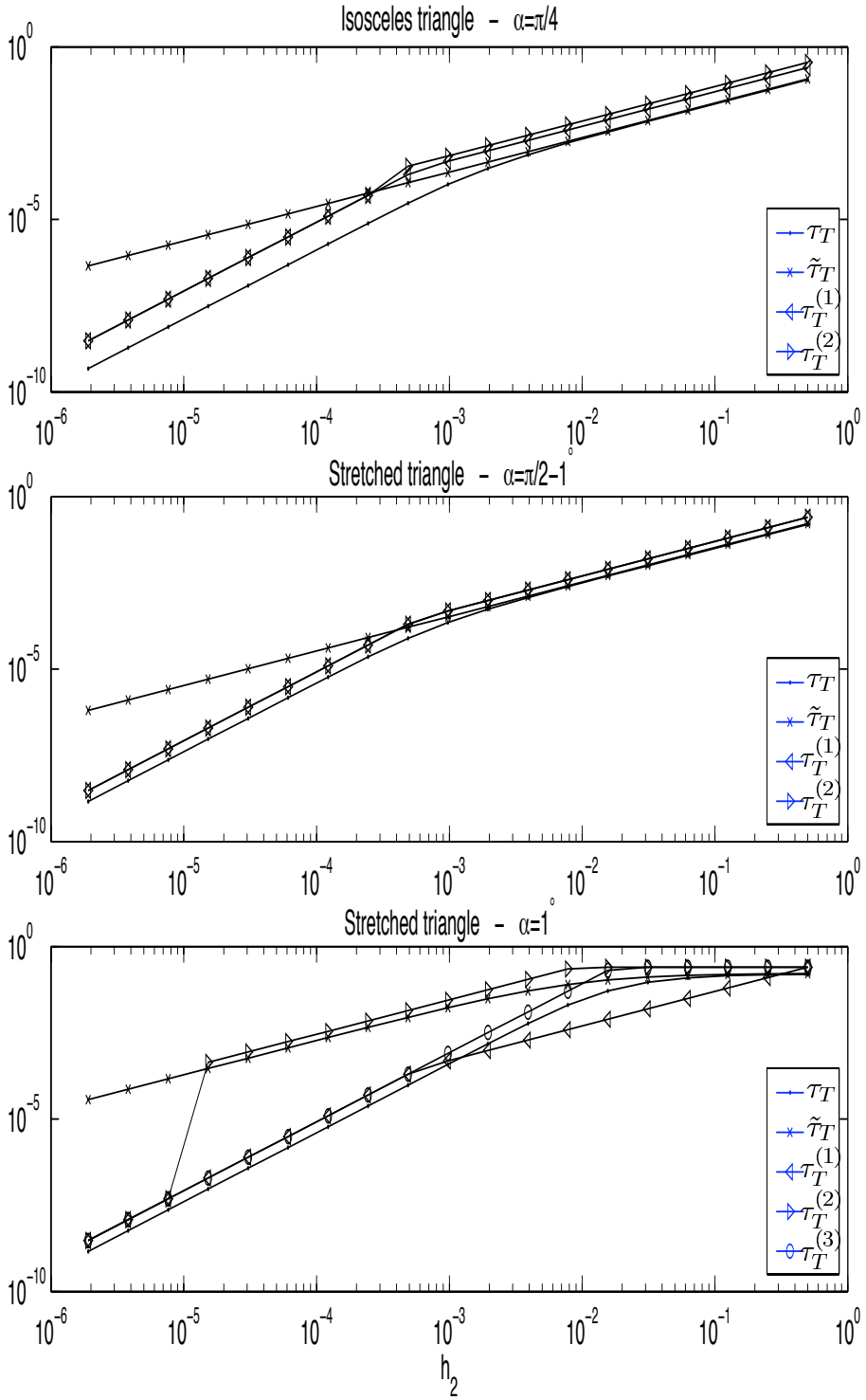


Figure 3.7: Different choices of the SD–parameter plotted against h_2 for the three configurations depicted in Figure 3.6. The parameter τ_3 coincides with τ_2 in the two upper–most plots.

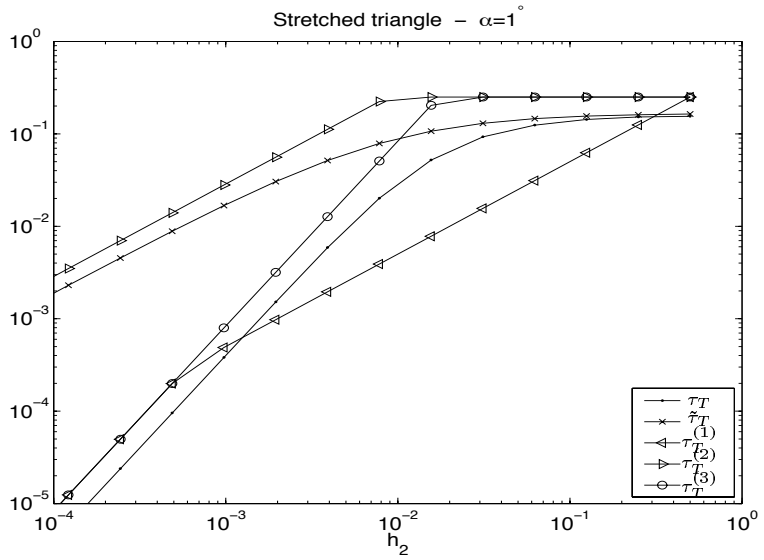


Figure 3.8: Different choices for the SD–parameter: zoom from the bottom plot in Figure 3.7

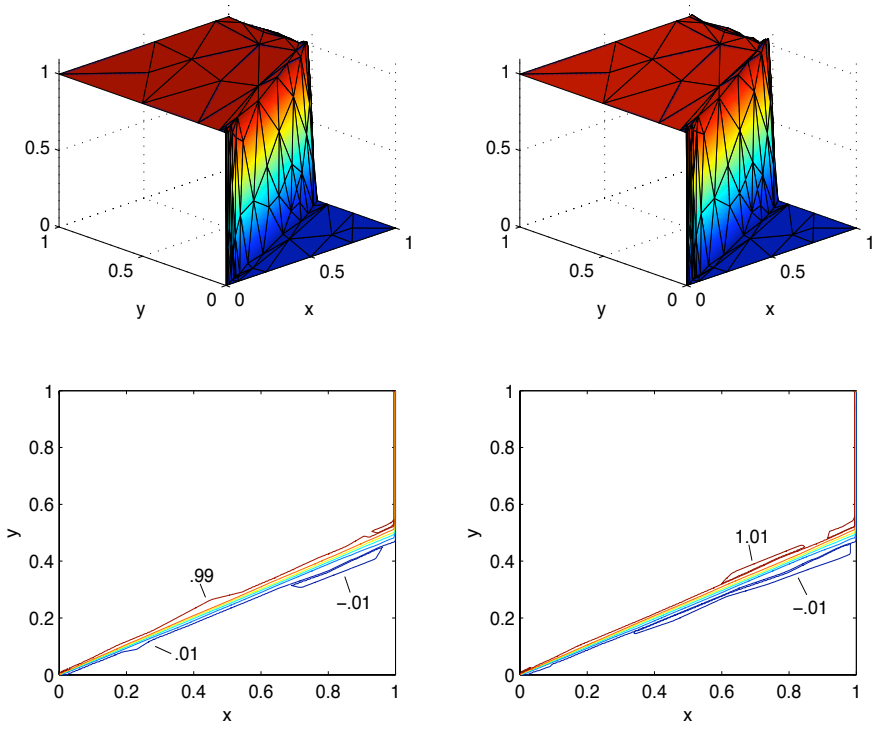


Figure 3.9: SDFEM solution of (3.34) using $\tau_T^{(2)}$ (left plots) and $\tau_T^{(3)}$ (right plots).

Chapter 4

The RFBe method

The underlying principle behind the use of *bubble methods* (see, e.g. [21]), is that of enriching the Galerkin finite element space with functions (bubbles) having compact support on every element of the given triangulation. The bubbles are successively eliminated through static condensation, leaving a generalised Galerkin scheme for the original finite element space which is expected to have improved approximation properties. In general, this procedure is effective when the numerical solution of the problem under consideration is sensitive to features present on scales that cannot be represented on the given mesh (*subgrid scales*).

Following [17], we note that a limitation of bubble methods in general, and of RFB in particular, is that only those subgrid features that do not cross the inter–element boundaries can be modelled. We propose to further enrich the RFB method with *edge bubbles*, i.e. bubbles with compact support on pairs of elements. The number of such edge bubbles should be small relative to the total number of degrees of freedom, the idea being that they should be introduced only on those edges where it is crucial that the subgrid scales are captured.

We show through *a priori* analysis, and illustrate by numerical experiments, that the resulting procedure, the *enhanced residual–free bubble method* (RFBe), can be successfully applied to improve the RFB resolution of boundary layers in convection–dominated diffusion problems. A crucial property of the method is that the locally introduced edge bubbles give global improvement of the solution, showing that they have a stabilising effect on the scheme. Moreover, the RFBe numerical solution is able to capture the correct behaviour inside boundary layers.

4.1 A framework for the enrichment of the RFB finite element space

Given a bounded polygonal domain Ω in $\mathbb{R}^2$, let $\mathcal{L}(\cdot,\cdot)$ be a bounded coercive bilinear functional on $V = H_0^1(\Omega)$ and let $f \in L^2(\Omega)$. We consider the elliptic boundary value problem in

variational form:

$$\begin{cases} \text{find } u \in V \text{ such that} \\ \mathcal{L}(u,v) = (f,v) \qquad \forall v \in V. \end{cases} \tag{4.1}$$

Assume that we are given a conforming and shape–regular partition $\mathcal{T}_h$ of Ω such that any $T \in \mathcal{T}_h$ is affine–equivalent to a reference element $\widehat{T}$ through an affine map F_T.

As usual, we define over $\mathcal{T}_h$ the standard finite element space V_h restricting ourselves to the cases of piecewise linear elements (if $\widehat{T}$ is a triangle) or piecewise bilinear elements (if $\widehat{T}$ is a rectangle). Moreover, let B_h be the space of residual–free bubbles according to definition (2.13), i.e. the set of all functions in V that have support on every element $T \in \mathcal{T}_h$. That is, we define

$$V_h = \left\{ v_h \in H_0^1(\Omega) : w_{h|_T} \in \mathcal{P}_1(T) \ (\text{or } w_{h|_T} \circ F_T \in \mathcal{Q}_1(\widehat{T})) \quad \forall T \in \mathcal{T}_h \right\}, \tag{4.2}$$

$$B_h = \bigoplus_{T \in \mathcal{T}_h} H_0^1(T). \tag{4.3}$$

The *residual–free bubble* finite element space is then defined as $V_{RFB} = V_h \oplus B_h$, the direct sum being ensured since V_h does not contain nontrivial bubbles, i.e. $V_h \cap B_h = \{0\}$.

Finally, as usual, the residual–free bubble method (RFB) is given by restricting (4.1) to the subspace V_{RFB}:

$$\begin{cases} \text{find } u_{RFB} \in V_{RFB} \text{ such that} \\ \mathcal{L}(u_{RFB},v) = (f,v) \qquad \forall v \in V_{RFB}. \end{cases} \tag{4.4}$$

As discussed previously, the bubble part of the solution can be eliminated, at least formally, from the RFB formulation through a static condensation procedure which leads to a generalised Galerkin formulation over the standard finite element space V_h. In other words, the RFB method can be formulated as a two–level algorithm which corresponds to a sort of divide–and–conquer principle: the space of bubbles B_h should take into account the fine scales of the problem while the solution on V_h gives an approximation of the global behaviour of u.

The draw–back of such a two–level procedure is that only the small scales that do not cross the boundary of any element of the partition will be taken into account.

Following an idea by Brezzi [17], we propose to further enrich the RFB space with the aim of improving the approximation properties on selected edges of the partition.

The framework is that of the *general augmented space* formulation proposed by Brezzi and Marini in [25] which we have introduced in Section 2.2.1. Given a finite dimensional space of traces Φ_h over the skeleton Σ of the partition, we define the *augmented space*

$$V_a := \left\{ v \in V \ : \ v_{|_\Sigma} \in \Phi_h \right\}. \tag{4.5}$$

The associated *general augmented space formulation* is given by the approximation of the variational problem (4.1) into V_a (c.f. (2.19)):

$$\begin{cases} \text{find } u_a \in V_a \text{ such that} \\ \mathcal{L}(u_a, v_a) = (f, v_a) \qquad \forall v_a \in V_a. \end{cases} \tag{4.6}$$

As explained in Section 2.2.1, V_a admits the representation

$$V_a = V_l \oplus B_h,$$

where

$$V_l := \{v_l \in V_a \,:\, \mathcal{L}(v_l, v_b) = 0 \quad \forall v_b \in B_h\}, \tag{4.7}$$

and B_h is the residual–free bubble space given by (4.3).

4.1.1 General definition of the RFBe method

The formulation (4.6) represents a framework for the modification of the RFB method. We mentioned in Section 2.2.1 that the RFB method is obtained from the augmented space formulation by choosing $\Phi_h = V_h|_\Sigma$: we now propose to modify the RFB method by augmenting $V_h|_\Sigma$ locally and considering the associated augmented space formulation.

This can be done by considering as space of traces Φ_h the traces of V_h plus a relatively low–dimensional space of traces Φ_e assuming that any element of Φ_e has support on a single edge. In this way we ensure that the space

$$\Phi_h := V_h|_\Sigma \oplus \Phi_e,$$

is defined through direct sum of its components. Indeed, on any given partition edge γ, the elements of Φ_e are either identically zero, or one–dimensional bubbles, while, by definition, $V_h|_\gamma$ does not contains bubbles. (For the definition of *bubbles*, see (2.1)). Hence $V_h|_\gamma \cap \Phi_e|_\gamma = \{0\}$, for any edge γ of the partition.

We then consider the augmented space V_a given by (4.5), and name the corresponding augmented space formulation (4.6) a RFBe (*enhanced residual–free bubble*) method.

Moreover, given the decomposition $V_a = V_l \oplus B_h$ discussed above, we define

$$\widetilde{V}_h := \{v \in V_l : v|_\Sigma \in V_h|_\Sigma\},$$
$$E_h := \{v \in V_l : v|_\Sigma \in \Sigma_e\},$$

and name *edge bubbles* the elements of E_h. We clearly have that

$$V_a = \widetilde{V}_h \oplus E_h \oplus B_h. \tag{4.8}$$

Further, we notice that, by construction, the space of edge bubbles E_h admits a basis whose elements have support on the couple of elements sharing a given edge.

The choice of the edge bubbles or, more precisely, of their traces in Σ, should be dependent on the problem under consideration; indeed the edge bubbles may be chosen by exploiting the differential equation or, alternatively, some information about the solution obtained through a previous computation.

In the next section we present an example of a successful application of the RFBe framework and show how an *a priori* error analysis can be used to determine a suitable choice of edge bubbles.

4.2 The RFBe method for the resolution of boundary layers

We exemplify edge bubbles designed to capture boundary layer behaviour and we prove, through *a priori* error analysis, that the particular choice of edge bubbles results in reduction of the discretisation error beyond that of the classical RFB method.

Consider the following boundary value problem for the convection–diffusion equation:

$$\begin{cases} Lu := -\varepsilon\Delta u + \boldsymbol{a}\cdot\nabla u = f & \text{in}\,\Omega = (0,1)^2, \\ u = 0 & \text{on}\,\partial\Omega, \end{cases} \tag{4.9}$$

where the convection field $\boldsymbol{a} \in [\mathcal{C}^1(\bar{\Omega})]^2$.

We define *outflow boundary* the subset of $\partial\Omega$ given by

$$\partial\Omega_+ = \{\boldsymbol{x} \in \partial\Omega : \boldsymbol{a}\cdot\boldsymbol{n} > 0\}.$$

In the convection–dominated regime, the solution exhibits a normal boundary layer in correspondence to the outflow boundary.

We aim to obtain increased resolution of the boundary layer behaviour over a standard Galerkin finite element method and over RFB by careful selection of edge bubbles.

4.2.1 Definition of the RFBe method

Let $\mathcal{T}_h$ be a shape regular axiparallel rectangular partition of $\bar{\Omega}$ (as always, h represents the maximum diameter over the elements of the partition). Thus, the mesh is the tensor–product of the subdivisions $0 = x_0 < x_1 < \cdots < x_m = 1$ and $0 = y_0 < y_1 < \cdots < y_n = 1$. Moreover, let Γ be the set of all edges in the partition.

We define as the *boundary layer region* a neighbourhood of the outflow boundary of width

$$\kappa = \varepsilon \ln(1/\varepsilon), \tag{4.10}$$

in the direction orthogonal to the boundary (see Figure 4.1). Accordingly, we distinguish among those elements that intersect the boundary layer region and those that do not by introducing the subpartitions:

$$\mathcal{T}_{bl} = \{T \in \mathcal{T}_h : \text{dist}(T, \partial\Omega_+) \le \kappa\} \tag{4.11}$$

$$\mathcal{T}_{outer} = \mathcal{T}_h \setminus \mathcal{T}_{bl}. \tag{4.12}$$

Similarly, we decompose the set of edges Γ onto

$$\Gamma_{bl} = \{\gamma \subset \Gamma : \mathrm{dist}(\gamma, \partial\Omega_+) \leq \kappa\}$$
$$\Gamma_{outer} = \Gamma \setminus \Gamma_{bl}.$$

Finally, given an edge γ we let $h_\gamma := |\gamma|$.

We make the assumption that $h \gtrsim \kappa$ an say that the boundary layer is not resolved by the given mesh.

We propose to introduce one edge bubble for any edge that cross the boundary layer. Under the above assumption, the number of such edges is $m + n - 2$ which in turn is bounded by Kh^{-1} where K is a constant depending only on the shape–regularity of the mesh. In other words,

$$\#(\Gamma_{bl}) = O(1/h); \qquad \#(\Gamma_{outer}) = O(1/h^2). \tag{4.13}$$

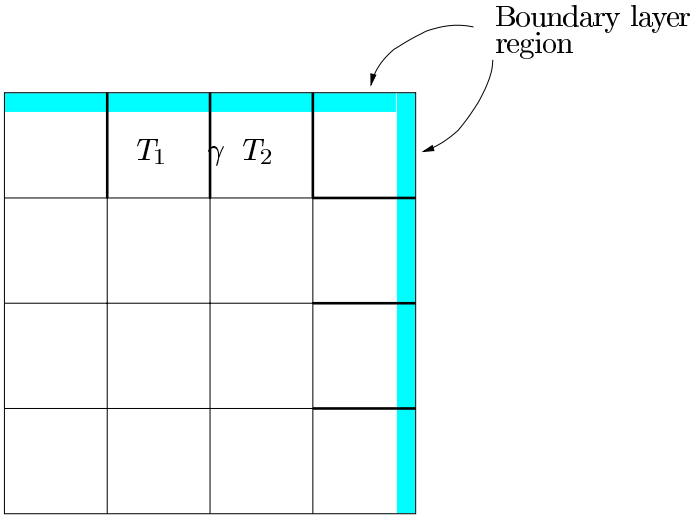


Figure 4.1: Axiparallel mesh. The edges where an edge–bubble is defined are marked.

Let T_i, $i = 1, 2$, be the two elements sharing a given edge $\gamma \in \Gamma_{bl}$. We define the edge bubble e_γ as follows:

- $e_\gamma \in H_0^1(\Omega)$ and $\mathrm{supp}(e_\gamma) \subset \left(\overline{T_1} \cup \overline{T_2}\right)^\circ$;
- the value of e_γ on γ is given by the solution of the one–dimensional boundary value problem

$$\begin{cases} L_\gamma w = 1 & \text{in } \gamma, \\ w_\gamma = 0 & \text{on } \partial\gamma, \end{cases} \tag{4.14}$$

where L_γ is a differential operator obtained from L by considering directional derivatives along γ. In our case, if for example γ is parallel to the y–axis as in Figure 4.1, then L_γ is obtained on the associated boundary layer edge by discarding all terms with derivatives in the x–direction.

- inside T_i, $i = 1, 2$, e_γ satisfies

$$\mathcal{L}_{T_i}(e_\gamma, v) = 0 \qquad \forall v \in H_0^1(T_i), \tag{4.15}$$

where $\mathcal{L}_{T_i}$ is the restriction of $\mathcal{L}$ to T_i.

Remark. More generally, when γ is not aligned with a coordinate direction, since the Laplacian is rotation–invariant, we would define $L_\gamma w = -\varepsilon w''(t) + a_\gamma w'(t)$, where a_γ is the projection of $\boldsymbol{a}$ along γ.

To handle the case of convection–diffusion equations with symmetric tensor diffusion coefficient, we would need to freeze the diffusion coefficient to an edge–wise constant tensor. Then a local rotation of the coordinate system and stretching along the rotated coordinate axis can be used to transform the diffusion term into the Laplacian.

We now define the edge bubble space

$$E_h := \operatorname{span}\{e_\gamma : \gamma \in \Gamma_{bl}\},$$

and the RFBe method follows as the Galerkin formulation on the associated augmented space (4.8).

Alternatively, in order to take advantage of the well–known approximation properties of the underlying finite element space V_h, we can consider the following equivalent definition.

As shown in the previous section, the space of edge bubbles E_h in direct sum with B_h and V_h. Thus, as already discussed in the case of the RFB method, we have the following alternative splitting of the augmented space:

$$V_a = V_h \oplus E_h \oplus B_h =: V_P \oplus B_h. \tag{4.16}$$

Since in (4.16) we have a direct sum, every $v_a \in V_a$ can be uniquely decomposed as a sum over the three subspaces. Hence, the RFBe formulation reads

$$\begin{cases} \text{find } u_a = u_h + u_e + u_b \in V_a = V_h \oplus E_h \oplus B_h \text{ such that} \\ \mathcal{L}(u_h + u_e + u_b, v_a) = (f, v_a) \qquad \forall v_a \in V_a. \end{cases} \tag{4.17}$$

As before, testing in (4.17) with $v_b \in B_h$, we obtain the bubble equation

$$\mathcal{L}(u_b, v_b) = (f, v_b) - \mathcal{L}(u_h, v_b) \qquad \forall v_b \in B_h, \tag{4.18}$$

where we have used the 'orthogonality' of E_h and B_h with respect to $\mathcal{L}$ expressed by (4.15). Starting from the standard RFB method formulation (4.4), we would obtain exactly the same

bubble equation. That is to say, the introduction of the edge bubbles leaves the static condensation procedure unchanged.

Formally, static condensation is carried out as follows. Letting φ_j, $j = 1, \ldots, N_T$, be the local basis functions for V_h on a generic element $T \in \mathcal{T}_h$ and considering the local decomposition $u_{h|_T} = \sum_{j=1}^{N_T} U_j \varphi_j$, we have from (4.18):

$$u_{b|_T} = \sum_{j=1}^{N_T} U_j b_j + b_f, \tag{4.19}$$

where

$$\begin{cases} b_j \in H_0^1(T) \text{ such that} \\ \mathcal{L}(b_j, v) = -\mathcal{L}(\varphi_j, v) \qquad \forall v \in H_0^1(T), \end{cases} \tag{4.20}$$

and

$$\begin{cases} b_f \in H_0^1(T) \text{ such that} \\ \mathcal{L}(b_f, v) = (f, v) \qquad \forall v \in H_0^1(T). \end{cases} \tag{4.21}$$

Assume that the variational problems (4.20) and (4.21) have been solved. Testing in (4.17) with $v \in V_P$ and substituting u_b using (4.19), we arrive at the following problem on the space V_P:

$$\begin{cases} \text{find } u_h + u_e \in V_P \text{ such that} \\ \mathcal{L}(u_h + u_e, v) + \displaystyle\sum_{T \in \mathcal{T}_h} \left(\sum_{j=1}^{N_T} U_j \mathcal{L}_T(b_j, v) \right) = (f, v) - \sum_{T \in \mathcal{T}_h} \mathcal{L}_T(b_f, v) \qquad \forall v \in V_P. \end{cases} \tag{4.22}$$

Thus, as for the standard RFB method (2.11), the RFBe method results in a generalised Galerkin formulation except that this generalised Galerkin formulation is now over the locally augmented finite element space $V_P = V_h \oplus E_h$.

We justify the choice of the edge bubbles through *a priori* analysis.

4.3 *A priori* error analysis

In the following pages, C represents a constant whose actual value may change value at different occurrences; C may be dependent on $\boldsymbol{a}$ but is always independent of ε and the mesh size h.

We emphasise that the following analysis is valid under the hypothesis that $h \gtrsim \kappa$, where κ is the width of the boundary layer defined by equation (4.10). Of course, as the mesh is refined, the relative improvement in accuracy due to the introduction of edge bubbles gradually diminishes (see the numerical examples below) and as h becomes smaller the standard RFB method will yield equally satisfactory results. Indeed, when $h \lesssim \kappa$, even a standard Galerkin

finite element method will provide good accuracy. However, our concern here is the case when $h \gtrsim \kappa$.

To proceed, we shall strengthen our hypotheses on the mesh by assuming that it is quasi–uniform, i.e. that there exists a constant $c_0 > 0$ such that

$$c_0 h \leq h_T \leq h = \max_{T \in \mathcal{T}_h} h_T, \tag{4.23}$$

for every $T \in \mathcal{T}_h$, where h_T is the diameter of the element T.

As regards the data of the problem under consideration (4.9), we make the further assumption that

$$\boldsymbol{a} = (a_1, a_2), \qquad \text{where } a_1, a_2 \geq c_a > 0, \tag{4.24}$$

so that the outflow boundary $\partial\Omega_+$ coincides with the union of the two sides $x = 1$ and $y = 1$. The analysis below is valid as long as internal layers are absent. For this purpose, we assume that f is a smooth, say, $f \in W^1_\infty(\Omega)$. Finally, we assume that $\operatorname{div} \boldsymbol{a} \leq 0$ in Ω to ensure coercivity of the bilinear functional associated with L. Under the above hypothesis, the boundary value problem (4.9) admits a unique weak solution $u \in H^1_0(\Omega)$.

4.3.1 Properties of an asymptotic approximation

In order to analyse the RFBe method, we need to use information about the behaviour of the solution based on performing an asymptotic analysis, in much the same way as in [60].

We note that the availability of an asymptotic expansion of the analytical solution u in terms of the small parameter ε is only required for analytical purposes and is not needed in the practical implementation of the method.

The reduced problem, corresponding to putting $\varepsilon = 0$in (4.9), is defined by

$$\begin{cases} \boldsymbol{a} \cdot \nabla u_0 = f & \text{in}\,\Omega, \\ u_0|_{x=0} = u_0|_{y=0} = 0, \end{cases} \tag{4.25}$$

with solution $u_0 \in W^1_\infty(\Omega)$; first–order partial derivatives of u_0 exhibit discontinuities across the characteristic passing through the inflow corner $(0,0)$ of Ω (see, e.g., [82], pp. 178).

Let Γ_{ns} be the set of those edges γ such that $\overset{\circ}{\gamma}$ is crossed by the characteristic through $(0,0)$, and let $\Gamma_{sm} = \Gamma_{outer} \setminus \Gamma_{ns}$. Under assumption (4.24), we clearly have

$$\#(\Gamma_{ns}) = O(1/h), \tag{4.26}$$

$$\#(\Gamma_{sm}) = O(1/h^2). \tag{4.27}$$

We observe that, for every edge $\gamma \in \Gamma_{ns}$, $u_0|_\gamma \in H^1(\gamma)$; moreover,

$$|u_0|^2_{1,\gamma} \leq C h_\gamma. \tag{4.28}$$

Indeed, by hypothesis, the characteristic through $(0,0)$ cuts γ into two disjoint open intervals γ_1 and γ_2. Since $u_{0|_{\gamma_i}} \in W^1_\infty(\gamma_i)$, $i=1,2$, we still have (assuming that γ is parallel with the x–axis):

$$\begin{aligned} |u_0|^2_{1,\gamma} &= |u_0|^2_{1,\gamma_1} + |u_0|^2_{1,\gamma_2} \\ &\leq \mathrm{meas}(\gamma_1)\|u_{0,x}\|^2_{\mathrm{L}^\infty(\gamma_1)} + \mathrm{meas}(\gamma_2)\|u_{0,x}\|^2_{\mathrm{L}^\infty(\gamma_2)} \\ &\leq \max\{\|u_{0,x}\|^2_{\mathrm{L}^\infty(\gamma_1)}, \|u_{0,x}\|^2_{\mathrm{L}^\infty(\gamma_2)}\}h_\gamma \\ &= Ch_\gamma. \end{aligned}$$

An identical bound holds when γ is parallel with the y–axis.

The same is true, of course, if $\gamma \in \Gamma_{sm}$. In this case, the eventual extra regularity of $u_0|_\gamma$ depends on the regularity of f. If, in particular, $f \in W^2_\infty(\Omega)$, then $u_0|_\gamma \in W^2_\infty(\gamma)$, for any $\gamma \in \Gamma_{sm}$. Hence, with a reasoning similar to he proof of (4.28), we get $u_0|_\gamma \in H^2(\gamma)$ and

$$|u_0|^2_{2,\gamma} \leq Ch_\gamma. \tag{4.29}$$

We consider the asymptotic approximation of u given by

$$\begin{aligned} u_{as}(x,y) =& u_0(x,y) \\ &- u_0(1,y)\mathrm{e}^{-a_1(1,y)\frac{1-x}{\varepsilon}} - u_0(x,1)\mathrm{e}^{-a_2(x,1)\frac{1-y}{\varepsilon}} + u_0(1,1)\mathrm{e}^{-a_1(1,1)\frac{1-x}{\varepsilon}}\mathrm{e}^{-a_2(1,1)\frac{1-y}{\varepsilon}} \\ \equiv& u_0(x,y) + u_c(x,y), \end{aligned} \tag{4.30}$$

where the last term in u_c is the *corner layer* correction, relevant in the vicinity of $(1,1)$ where the two boundary layers intersect (see, e.g. [38]).

The accuracy of the asymptotic approximation u_{as} depends on the smoothness of the reduced solution. In particular, the following result has been proven by Schieweck ([89], Lemma 4.4, pp. 33; see also [82], pp. 184).

Lemma 4.3.1 *Assume that $a \in [\mathcal{C}^1(\bar\Omega)]^2$, $f \in W^1_\infty(\Omega)$ and that (4.24) holds. Given $u \in H^1_0(\Omega)$, the solution of (4.9), and u_{as} as in (4.30), it holds*

$$\varepsilon|u-u_{as}|^2_{1,\Omega} + \|u-u_{as}\|^2_{0,\Omega} \leq C\varepsilon, \tag{4.31}$$

where C is a constant independent of ε.

Along the *a priori* error analysis, we will often use the following bounds proven in Appendix A.3.

Lemma 4.3.2 *Given an axiparallel rectangular mesh $\mathcal{T}_h$ satisfying (4.23), consider the subpartitions $\mathcal{T}_{bl}$ and $\mathcal{T}_{outer}$ given by (4.11) and (4.12), respectively. Let u_c be as in (4.30), under the hypothesis of Lemma 4.3.1. If $\varepsilon \leq ch$ for some constant $c \in \mathbb{R}^+$ and $\varepsilon \leq 1/\mathrm{e}$, then*

$$\sum_{\Gamma\in\Gamma_{outer}} |u_c|^2_{1,\Gamma} \leq C\frac{\mathrm{e}^{-2c_0c_ah/\varepsilon}}{\varepsilon^2}; \qquad \sum_{\Gamma\in\Gamma_{outer}} \|u_c\|^2_{0,\Gamma} \leq C\mathrm{e}^{-2c_0c_ah/\varepsilon}, \tag{4.32}$$

and

$$\sum_{\Gamma\in\Gamma_{bl}} |u_c|^2_{1,\Gamma} \leq C\frac{1}{\varepsilon h}; \qquad \sum_{\Gamma\in\Gamma_{bl}} ||u_c||^2_{0,\Gamma} \leq C\varepsilon h^{-1}. \tag{4.33}$$

4.3.2 *A priori* error bound

We recall some useful inequalities.

Lemma 4.3.3 *Let $T \in \mathcal{T}_h$. For any $v \in H^1(T)$, the trace inequalities*

$$|v|^2_{1/2,\partial T} \leq C\left(\frac{1}{h_T^2}||v||^2_{0,T} + |v|^2_{1,T}\right), \tag{4.34}$$

and

$$||v||^2_{0,\partial T} \leq C\left(h_T^{-1}||v||^2_{0,T} + \alpha^{-1}||v||^2_{0,T} + \alpha|v|^2_{1,T}\right), \tag{4.35}$$

hold, for any $\alpha > 0$, with C independent of v, h_T and α.

Furthermore, for any edge γ of T and for any $v \in H^1(\gamma)$, there exists a constant C independent of v and h_γ such that

$$|v|^2_{1/2,\gamma} \leq C\left(h_\gamma^{-1}||v||^2_{0,\gamma} + |v|_{1,\gamma}||v||_{0,\gamma}\right), \tag{4.36}$$

Proof. All the stated results are obtained by standard scaling arguments. In particular, (4.34) and (4.35) are proven scaling, respectively, the Trace Theorem (3.13) and the trace inequality

$$|\hat{v}|^2_{0,\partial\hat{T}} \leq C(||\hat{v}||^2_{0,\hat{T}} + ||\hat{v}||_{0,\hat{T}}|\hat{v}|_{1,\hat{T}}),$$

(this standard result can be found, for instance, in Brenner and Scott [15]). As for (4.36), this is obtained scaling the inequality

$$|\hat{v}|^2_{1/2,\hat{\gamma}} \leq C||\hat{v}||_{0,\hat{\gamma}}||\hat{v}||_{1,\hat{\gamma}}, \tag{4.37}$$

expressing the fact that the Hilbert space of fractional order $H^{1/2}(\gamma)$ is *exact of exponent* $1/2$ (see, e.g., Bergh and Löfström [14], pp. 27). □

The restriction of a function belonging to V_{RFB} onto an edge of the partition is a linear polynomial. This property will be used extensively in the following proof. In particular, we use the following well known properties of the linear Lagrange interpolant operator in one spatial dimension.

Lemma 4.3.4 *Let $\gamma \subset \mathbb{R}$ be an open interval of the real line and let $\pi_1 : \mathcal{C}^0(\bar{\gamma}) \to \mathcal{P}_1$ be the linear Lagrange interpolation operator. Then, for any $v \in H^m(\gamma)$, $m = 1, 2$, the following interpolation error bound holds:*

$$|u - \pi_1 u|_{i,\gamma} \leq Ch_\gamma^{m-i}|u|_{m,\gamma}, \tag{4.38}$$

for $i \leq m$. Moreover,

$$|u - \pi_1 u|_{1/2,\gamma} \leq C h_\gamma^{m-1/2} |u|_{m,\gamma}. \tag{4.39}$$

Proof. The result (4.38) is classical; see, e.g., Brenner and Scott [15]. As for (4.39), this is obtained by applying (4.36) to $v - \pi_1 v$ and using (4.38). □

Before developing the *a priori* analysis, we show that the global error of a locally residual–free finite element method is governed by the approximation properties of the augmented space V_a on the skeleton of the subdivision.

To switch to norms defined over the skeleton of the triangulation, we shall apply the following result from [85].

Lemma 4.3.5 *Assume that $\varepsilon < h \leq 1$, that (4.23) holds true and that the partition $\mathcal{T}_h$ is shape regular. Then, for each $T \in \mathcal{T}_h$, each $v_0 \in H^{1/2}(\partial T)$, and each $w \in H^1(T)$, there exists a function $v \in H^1(T)$ with $v = v_0$ on ∂T and satisfying*

$$\varepsilon \, |w - v|_{1,T}^2 + \varepsilon^{-1} \, \|w - v\|_{0,T}^2 \leq C \left(\varepsilon \, |w - v_0|_{1/2,\partial T}^2 + \|w - v_0\|_{0,\partial T}^2 \right), \tag{4.40}$$

where C depends only on $\mathcal{T}_h$.

Proof. This result follows easily by a scaling argument applying Lemma 3.2.2 to $(w - v_0) \circ F_T$ with $t = \varepsilon h^{-1}$, where F_T is the affine transformation from the reference element $\widehat{T}$ onto T. □

We are now ready to prove the following lemma.

Lemma 4.3.6 *Let $u \in V$ and $u_a \in V_a$ be, respectively, the exact solution of (4.1) and its numerical approximation, the solution of (4.17). Then the following bound holds:*

$$\varepsilon |u - u_a|_{1,\Omega}^2 \leq C \inf_{v_P \in V_P} \sum_{T \in \mathcal{T}_h} \left(\varepsilon \, |u - v_P|_{1/2,\partial T}^2 + \|u - v_P\|_{0,\partial T}^2 \right). \tag{4.41}$$

Proof. For any $v \in V_a$, using the appropriate coercivity inequality and Galerkin orthogonality, we have

$$\begin{aligned} \varepsilon |u - u_a|_{1,\Omega}^2 &\leq \mathcal{L}(u - u_a, u - u_a) = \mathcal{L}(u - u_a, u - v) \\ &= \varepsilon \, (\nabla(u - u_a), \nabla(u - v))_\Omega + (\boldsymbol{a} \cdot \nabla(u - u_a), u - v)_\Omega \\ &\leq C \varepsilon |u - u_a|_{1,\Omega} \left(|u - v|_{1,\Omega} + \varepsilon^{-1} \|u - v\|_{0,\Omega} \right). \end{aligned}$$

Thus, for any $v \in V_a$,

$$\varepsilon |u - u_a|_{1,\Omega}^2 \leq C \left(\varepsilon |u - v|_{1,\Omega}^2 + \varepsilon^{-1} \|u - v\|_{0,\Omega}^2 \right). \tag{4.42}$$

We fix $v_P \in V_P$ and consider its trace over the skeleton of the triangulation $\mathcal{T}_h$. Applying (4.40) with $w = u$ and $v_0 = v_P|_{\partial T}$ we have that, for some $v \in H_0^1(\Omega)$ which is equal to $v_P|_{\partial T}$ on the skeleton of $\mathcal{T}_h$,

$$\varepsilon \left|u - v\right|_{1,T}^2 + \varepsilon^{-1} \left\|u - v\right\|_{0,T}^2 \leq C \left(\varepsilon \left|u - v_P\right|_{1/2,\partial T}^2 + \left\|u - v_P\right\|_{0,\partial T}^2\right),$$

for every $T \in \mathcal{T}_h$. Hence, since $v \in V_a$ and recalling (4.42), from the arbitrarily of $v_P \in V_P$ we conclude that (4.41) holds true. □

The residual–free bubble finite element error bound (4.41) justifies the suggestion that the error can be reduced by enriching the trial space on the edges and represents the starting point for the *a priori* error analysis of RFBe.

Theorem 4.3.7 *Let $u \in H_0^1(\Omega)$ be the solution of the boundary–value problem (4.9) assuming that $\varepsilon \in \mathbb{R}^+$, $f \in W_\infty^2(\Omega)$ and $\boldsymbol{a} = (a_1, a_2) \in [\mathcal{C}^1(\bar{\Omega})]^2$, with $\operatorname{div} \boldsymbol{a} \leq 0$ and $a_1, a_2 \geq c_a > 0$. Moreover, let $\mathcal{T}_h$ be an axiparallel rectangular mesh satisfying (4.23) and $u_a \in V_a = V_h \oplus E_h \oplus B_h$ be the RFBe solution where the edge bubbles are defined according to (4.14) and (4.15). Then, as long as $h \geq (1/c_0 c_a)\,\kappa$ and $\varepsilon \leq 1/\mathrm{e}$, the following error bound holds:*

$$\varepsilon^{1/2} |u - u_a|_{1,\Omega} + h^{-1/2} \|\boldsymbol{a} \cdot \nabla(u - u_a)\|_{-1,\Omega} \leq C_1 \max\left(\varepsilon^{1/2} h^{-1/2}, \varepsilon^{1/4}\right) + C_2 h. \tag{4.43}$$

The constants C_1 and C_2 are independent of the mesh size h and of ε, but may depend on $\boldsymbol{a}$.

Proof. Let u_{as} be the asymptotic approximation of u given by (4.30). From (4.41) it follow that, for any $v_P \in V_P$,

$$\begin{aligned}
\varepsilon |u - u_a|_{1,\Omega}^2 &\leq C \sum_{T \in \mathcal{T}_h} \left(\varepsilon |u - v_P|_{1/2,\partial T}^2 + \|u - v_P\|_{0,\partial T}^2\right) \\
&\leq C \left(\sum_{T \in \mathcal{T}_h} \left(\varepsilon |u - u_{as}|_{1/2,\partial T}^2 + \|u - u_{as}\|_{0,\partial T}^2\right) \right. \\
&\qquad \left. + \sum_{T \in \mathcal{T}_h} \left(\varepsilon |u_{as} - v_P|_{1/2,\partial T}^2 + \|u_{as} - v_P\|_{0,\partial T}^2\right) \right) \\
&= C(I + II).
\end{aligned}$$

From our assumptions on ε and h it follows that $h \geq c\varepsilon$ with $c = c_0 c_a$. Thus, using the trace inequalities (4.34) and (4.35) with $\alpha = \min\left(\varepsilon^{1/2}, h\right)$, and the bound of the asymptotic

approximation (4.31), we get

$$\begin{aligned} I &\leq C \sum_{T\in\mathcal{T}_h} \left((h_T^{-1}+\alpha^{-1})||u-u_{as}||_{0,T}^2 + (\varepsilon+\alpha)|u-u_{as}|_{1,T}^2\right) \\ &\leq C\left((h^{-1}+\alpha^{-1})||u-u_{as}||_{0,\Omega}^2 + (\varepsilon+\alpha)|u-u_{as}|_{1,\Omega}^2\right) \\ &\leq C\left(\varepsilon h^{-1}+\varepsilon\alpha^{-1}+\alpha\right) \\ &\leq C\left(\varepsilon h^{-1}+\min\left(\varepsilon^{1/2},h\right)\right) \\ &\leq C\max\left(\varepsilon h^{-1},\varepsilon^{1/2}\right). \end{aligned}$$

Concerning II, we separate the contributions from those edges that cross the boundary layer (in which the edge bubbles are defined), and those that do not:

$$\begin{aligned} II &\leq 2\sum_{\gamma\in\Gamma_{outer}} \left(\varepsilon|u_{as}-v_P|_{1/2,\gamma}^2 + ||u_{as}-v_P||_{0,\gamma}^2\right) \\ &+ 2\sum_{\gamma\in\Gamma_{bl}} \left(\varepsilon|u_{as}-v_P|_{1/2,\gamma}^2 + ||u_{as}-v_P||_{0,\gamma}^2\right) \\ &= III + IV, \end{aligned}$$

the factor 2 being due to the fact that we are now summing over edges γ rather than through elements T as in the original definition of II.

To bound III, we consider the decomposition of the asymptotic approximation $u_{as} = u_0 + u_c$ introduced in (4.30). On any $\gamma \in \Gamma_{outer}$, since $V_P|_\gamma = V_h|_\gamma$ and

$$u_c \leq C\varepsilon^q \qquad \forall q \in \mathbb{N}, \tag{4.44}$$

it is, therefore, natural to choose, $v_P|_\gamma = \pi_1(u_0|_\gamma)$.

Using the triangle inequality, (4.39), (4.36), (4.28), (4.13) and (4.32) we get

$$\begin{aligned} \varepsilon\sum_{\gamma\in\Gamma_{outer}} |u_{as}-v_P|_{1/2,\gamma}^2 &\leq 2\varepsilon\sum_{\gamma\in\Gamma_{outer}} \left(|u_0-\pi_1 u_0|_{1/2,\gamma}^2 + |u_c|_{1/2,\gamma}^2\right) \\ &\leq C\varepsilon\left(\sum_{\gamma\in\Gamma_{outer}} h_\gamma |u_0|_{1,\gamma}^2 + \sum_{\gamma\in\Gamma_{outer}} \left(h_\gamma^{-1}||u_c||_{0,\gamma}^2 + |u_c|_{1,\gamma}||u_c||_{0,\gamma}\right)\right) \\ &\leq C\varepsilon\left(\sum_{\gamma\in\Gamma_{outer}} h_\gamma^2 + \varepsilon\sum_{\gamma\in\Gamma_{outer}} |u_c|_{1,\gamma}^2 + \varepsilon^{-1}\sum_{\gamma\in\Gamma_{outer}} ||u_c||_{0,\gamma}^2\right) \\ &\leq C\left(\varepsilon + \mathrm{e}^{-2c_0c_a h/\varepsilon}\right) \leq C\varepsilon, \end{aligned}$$

assuming, to obtain the last bound, that $h \geq (1/2c_0c_a)\,\varepsilon\ln(1/\varepsilon)$, which is a slightly less restrictive assumption then what is required by the statement of the theorem.

For the L_2–norm term in III, proceeding in the same fashion, we obtain

$$\begin{aligned}\sum_{\gamma\in\Gamma_{outer}} ||u_{as}-v_P||^2_{0,\gamma} &\leq \sum_{\gamma\in\Gamma_{outer}} \left(||u_0-\pi_1 u_0||^2_{0,\gamma}+||u_c||^2_{0,\gamma}\right)\\ &\leq C\left(\sum_{\gamma\in\Gamma_{outer}} h_\gamma^2|u_0|^2_{1,\gamma}+\mathrm{e}^{-2c_0c_ah/\varepsilon}\right)\\ &\leq C\left(\sum_{\gamma\in\Gamma_{outer}} h_\gamma^3+\mathrm{e}^{-2c_0c_ah/\varepsilon}\right)\leq C\left(h+\mathrm{e}^{-2c_0c_ah/\varepsilon}\right)\leq Ch.\end{aligned}$$

After comparing the two bounds we conclude that

$$III\leq Ch. \tag{4.45}$$

Notice that we can obtain sharper bounds by carefully distinguishing between edges where u_0 has different regularity properties (as we have already mentioned, u_0 may only fail to be a C^1 function across the characteristic curve that passes through the corner $(0,0)$ on the inflow boundary). Let Γ_{ns} and Γ_{sm} be as in Section 4.3.1. Then, using (4.26), (4.27), (4.28) and (4.29), we get

$$\begin{aligned}\varepsilon\sum_{\gamma\in\Gamma_{outer}} |u_0-v_P|^2_{1/2,\gamma} &= \varepsilon\left(\sum_{\gamma\in\Gamma_{sm}} |u_0-\pi_1u_0|^2_{1/2,\gamma}+\sum_{\gamma\in\Gamma_{ns}} |u_0-\pi_1u_0|^2_{1/2,\gamma}\right)\\ &\leq C\varepsilon\left(\sum_{\gamma\in\Gamma_{sm}} h_\gamma^3|u_0|^2_{2,\gamma}+\sum_{\gamma\in\Gamma_{ns}} h_\gamma|u_0|^2_{1,\gamma}\right)\\ &\leq C\varepsilon\left(\sum_{\gamma\in\Gamma_{sm}} h_\gamma^4+\sum_{\gamma\in\Gamma_{ns}} h_\gamma^2\right)\\ &\leq C\varepsilon(h^2+h)\leq C\varepsilon h,\end{aligned}$$

due also to the interpolation error bound (4.38) and (4.39). Similarly, we gain a factor of h in the bound of the L_2–norm, and we conclude that

$$III\leq Ch^2.$$

We now come to the analysis of IV, i.e. the term containing norms of $u_{as}-v_P$ over the edges γ belonging to Γ_{bl}: notice that on such edges $V_P=V_h\oplus E_h$.

Let us concentrate on the boundary layer associated with the boundary $y=1$. For each edge $\gamma_i=x_i\times[1-h_\gamma,1]$, $i=1,\ldots,m-1$, let e_{γ_i} be the edge bubble associated to γ_i as it is defined in Section 4.2. Since we have already fixed $v_P(x_i,1-h_\gamma)=u_0(x_i,1-h_\gamma)$, we have

$$\begin{aligned}v_{P|_{\gamma_i}} &= \pi_i u_0|_{\gamma_i}+g_ie_{\gamma_i}|_{\gamma_i}\\ &= u_0(x_i,1-h_{\gamma_i})\frac{1-y}{h_\gamma}+g_ie_{\gamma_i}|_{\gamma_i},\end{aligned} \tag{4.46}$$

for some coefficient $g_i \in \mathbb{R}$ which we are free to choose.

We consider the following decomposition of the asymptotic approximation $u_{as}(x,y)$ of $u(x,y)$:

$$\begin{aligned} u_{as}(x_i,y) &= u_0(x_i,y) - u_0(x_i,1)\mathrm{e}^{-a_2(x_i,1)\frac{1-y}{\varepsilon}} \\ &+ \left(-u_0(1,y)\mathrm{e}^{-a_1(1,y)\frac{1-x_i}{\varepsilon}} + u_0(1,1)\mathrm{e}^{-a_1(1,1)\frac{1-x_i}{\varepsilon}}\mathrm{e}^{-a_2(1,1)\frac{1-y}{\varepsilon}}\right) \\ &= u_0^i(y) + u_{c_y}^i(y) + u_c^i(y), \end{aligned}$$

where we have named

$$u_0^i(y) = u_0(x_i,y), \qquad u_{c_y}^i(y) = u_0(x_i,1)\mathrm{e}^{-a_2(x_i,1)\frac{1-y}{\varepsilon}},$$

and the function u_c^i collects all the remaining correction terms, which are related to the boundary layer at $x=1$ and to the corner layer. Since u_c^i is exponentially small in the sense made precise by (4.44), the main task is to define $v_{P|_{\gamma_i}}$ as a good approximation of $u_0^i + u_{c_y}^i$.

Consider the decomposition

$$u_0^i + u_{c_y}^i = \left(u_0 + \pi_1(u_{c_y}^i)\right) + \left(u_{c_y}^i - \pi_1(u_{c_y}^i)\right).$$

We observe that $(u_{c_y}^i - \pi_1(u_{c_y}^i))|_{\gamma_i}$ is the solution of the boundary–value problem

$$\begin{cases} L_{\gamma_i} w = a_2(x_i,1)\frac{u_{c_y}^i(1-h_\gamma) - u_{c_y}^i(1)}{h_{\gamma_i}} & \text{on } \gamma_i\,, \\ w_{|\partial\gamma_i} = 0, \end{cases}$$

where

$$L_{\gamma_i} w := -\varepsilon w'' + a_2(x_i,1)w'. \tag{4.47}$$

In other words, it coincides with a multiple of the edge bubble e_{γ_i}, as defined in Section 4.2. Thus we obtain

$$g_i e_{\gamma_i}|_{\gamma_i} = u_{c_y}^i - \pi_1(u_{c_y}^i)$$

by setting

$$g_i := a_2(x_i,1)\frac{u_{c_y}^i(1-h_{\gamma_i}) - u_{c_y}^i(1)}{h_{\gamma_i}}.$$

In this way, recalling the definition of $v_P|_{\gamma_i}$, we get

$$\begin{aligned} u_0^i + u_{c_y}^i - v_{P|_{\gamma_i}} &= \left(u_0^i + \pi_1(u_{c_y}^i)\right) + \left(u_{c_y}^i - \pi_1(u_{c_y}^i)\right) - \left(u_0^i(1-h_{\gamma_i})\frac{1-y}{h_{\gamma_i}} + g_i e_{\gamma_i}|_{\gamma_i}\right) \\ &= u_0^i + \pi_1(u_{c_y}^i) - u_0^i(1-h_{\gamma_i})\frac{1-y}{h_{\gamma_i}} \\ &= u_0^i + u_{c_y}^i(1)\frac{y-1+h_\gamma}{h_{\gamma_i}} + u_{c_y}^i(1-h_{\gamma_i})\frac{1-y}{h_{\gamma_i}} - u_0^i(1-h_{\gamma_i})\frac{1-y}{h_{\gamma_i}} \\ &= u_0^i - \pi_1(u_0^i) + u_{c_y}^i(1-h_\gamma)\frac{1-y}{h_{\gamma_i}}, \end{aligned}$$

since $u^i_{c_y}(1) = -u^i_0(1)$.

Finally, we notice that by definition of the $H^{1/2}$–seminorm,

$$\begin{aligned} |u^i_{c_y}(1-h_{\gamma_i})\frac{1-y}{h_{\gamma_i}}|^2_{1/2,\gamma} &= \frac{|u^i_{c_y}(1-h_\gamma)|^2}{h^2_\gamma}\iint_\gamma \frac{|(1-\xi)-(1-\eta)|^2}{|\xi-\eta|^2}\,d\xi d\eta \\ &= |u^i_{c_y}(1-h_{\gamma_i})|^2 \\ &= |u^i_0(1)|^2 \mathrm{e}^{-2a_2(x_i,1)h_{\gamma_i}/\varepsilon}. \end{aligned} \tag{4.48}$$

We are now ready to bound IV. For any $q \in \mathbb{N}$, by the triangle inequality, (4.48), (4.39) and (4.28), we have

$$\begin{aligned} &\varepsilon \sum_{i=1}^{m-1} |u_{as}(x_i,\cdot) - v_P(x_i,\cdot)|^2_{1/2,\gamma_i} \\ &\quad \le 2\varepsilon \sum_{i=1}^{m-1} \left(|u^i_0 + u^i_{c_y} - v_P|^2_{1/2,\gamma_i} + |u^i_c|^2_{1/2,\gamma_i} \right) \\ &\quad = 2\varepsilon \sum_{i=1}^{m-1} \left(\left| u^i_0 - \pi_1(u^i_0) + u^i_{c_y}(1-h_{\gamma_i})\frac{1-y}{h_{\gamma_i}} \right|^2_{1/2,\gamma_i} + |u^i_c|^2_{1/2,\gamma_i} \right) \\ &\quad \le 4\varepsilon \sum_{i=1}^{m-1} \left(|u^i_0 - \pi_1(u^i_0)|^2_{1/2,\gamma_i} + |u^i_0(1)|^2 \mathrm{e}^{-2a_2(x_i,1)h_{\gamma_i}/\varepsilon} + |u^i_c|^2_{1/2,\gamma_i} \right) \\ &\quad \le C\varepsilon \sum_{i=1}^{m-1} \left(h_{\gamma_i} |u^i_0|^2_{1,\gamma_i} + \mathrm{e}^{-2c_0 c_a h/\varepsilon} + |u^i_c|^2_{1/2,\gamma_i} \right) \\ &\quad \le C(\varepsilon h + \varepsilon^3 h^{-1} + \varepsilon^q) \le C\varepsilon h, \end{aligned}$$

having chosen $q \ge 2$ and using $h \ge c\varepsilon$. Similarly,

$$\begin{aligned} \sum_{i=1}^{m-1} &\quad \|u_{as}(x_i,\cdot) - v_P(x_i,\cdot)\|^2_{0,\gamma_i} \\ &\le C \sum_{i=1}^{m-1} \left(h^2_{\gamma_i} |u^i_0|^2_{1,\gamma_i} + h_{\gamma_i} \mathrm{e}^{-2a_2(x_i,1)h_{\gamma_i}/\varepsilon} + \|u^i_c\|^2_{0,\gamma_i} \right) \\ &\le C(h^2 + \varepsilon^2) \le Ch^2. \end{aligned}$$

Thus we have shown that

$$IV \le Ch^2.$$

Finally, recalling from [87] that the streamline derivative of the error

$$\|\boldsymbol{a}\cdot\nabla(u-u_a)\|_{-1,\Omega} := \sup_{v\in H^1(\Omega)} \frac{\int_\Omega \boldsymbol{a}\cdot\nabla(u-u_a)v\,d\Omega}{|v|_{1,\Omega}}$$

is controlled by the energy norm error, in the sense that

$$h^{-1/2}||\boldsymbol{a}\cdot\nabla(u-u_a)||_{-1,\Omega}\leq C\varepsilon^{1/2}|u-u_a|_{1,\Omega},$$

we deduce that (4.43) holds. □

Remark. The error bounds stated in the theorem are still valid as long as the number of edges crossed by the characteristic curve through $(0,0)$ is of order $O(1/h)$, which is true under the hypothesis that $a_1, a_2 > 0$. In the absence of this assumption and requiring that $f \in W^1_\infty(\Omega)$ we obtain, by virtue of (4.45),

$$\varepsilon^{1/2}|u-u_a|_{1,\Omega}+h^{-1/2}||\boldsymbol{a}\cdot\nabla(u-u_a)||_{-1,\Omega}\leq C_1\left(\varepsilon^{1/2}h^{-1/2}+h^{1/2}\right)+C_2h.$$

Corollary 4.3.8 *Let $u \in H^1_0(\Omega)$ be the solution of the boundary–value problem (4.9) and let $u_{RFB} \in V_{RFB}$ be the numerical solution given by (4.4). Under the hypothesis of Theorem 4.3.7, we have*

$$\varepsilon^{1/2}|u-u_a|_{1,\Omega}+h^{-1/2}||\boldsymbol{a}\cdot\nabla(u-u_a)||_{-1,\Omega}\leq C_1\max\left(\varepsilon^{1/2}h^{-1/2},\varepsilon^{1/4}\right)+C_2. \tag{4.49}$$

Proof. The proof of Theorem 4.3.7 can be repeated for the RFB method, with the only difference that, since no edge bubbles are included in the RFB finite element space, when bounding IV, we have $V_P = V_h$ on every edge $\gamma \in \Gamma_{bl}$.

In this case, to bound IV we proceed as follows. Since the first derivative of u_{as} is very large within boundary layers, it is not advisable to make use of the standard bounds on the error in the Lagrange linear interpolation to estimate $(u_{as}-v_P)|_\gamma$ when $\gamma \in \Gamma_{bl}$. Instead, we shall simply apply the triangle inequality and bound each term individually.

As before, let us concentrate on the boundary layer associated with the boundary $y = 1$. For each edge $\gamma_i = x_i \times [1-h_{\gamma_i}, 1]$, $i = 1, \ldots, m-1$, we have

$$v_P|_{\gamma_i} = u_0(x_i, 1-h_{\gamma_i})\frac{1-y}{h_{\gamma_i}},$$

(cf. (4.46), which applies to the RFBe method). By recalling the definition of the seminorm $|\cdot|_{1/2,\gamma_i}$, we easily see that

$$|v_P|^2_{1/2,\gamma_i}\leq C,$$

(cf. the proof of (4.48)). Of course, an analogous result holds on the edges adjacent to the boundary $x = 1$. Thus, writing $u_{as}-v_P = u_0 - v_P + u_c$, applying (4.36), (4.13) and (4.33) we

obtain:

$$\begin{aligned}
\varepsilon \sum_{\gamma\in\Gamma_{bl}} |u_{as}-v_P|^2_{1/2,\gamma} &\leq 3\varepsilon \sum_{\gamma\in\Gamma_{bl}} \left(|u_0|^2_{1/2,\gamma}+|v_P|^2_{1/2,\gamma}+|u_c|^2_{1/2,\gamma}\right)\\
&\leq 3\varepsilon \sum_{\gamma\in\Gamma_{bl}} \left(Ch_{\gamma_i}+C+h^{-1}_{\gamma_i}||u_c||^2_{0,\gamma}+|u_c|_{1,\gamma}||u_c||_{0,\gamma}\right)\\
&\leq C\left(\varepsilon h^{-1}+\varepsilon^2\sum_{\gamma\in\Gamma_{bl}}|u_c|^2_{1,\gamma}+\sum_{\gamma\in\Gamma_{bl}}||u_c||^2_{0,\gamma}\right)\\
&\leq C\varepsilon h^{-1}.
\end{aligned}$$

Similarly, for the L_2–norm term we obtain

$$\begin{aligned}
\sum_{\gamma\in\Gamma_{bl}} ||u_{as}-v_P||^2_{0,\gamma} &\leq 3\sum_{\gamma\in\Gamma_{bl}} \left(||u_0||^2_{0,\gamma}+||v_P||^2_{0,\gamma}+||u_c||^2_{0,\gamma}\right)\\
&\leq C\sum_{\gamma\in\Gamma_{bl}} h_{\gamma_i}+\sum_{\gamma\in\Gamma_{bl}}||u_c||^2_{0,\gamma}\\
&\leq C\left(1+\varepsilon h^{-1}\right)\leq C.
\end{aligned}$$

Thus, for the RFB method,

$$IV \leq C, \tag{4.50}$$

in contrast to $IV \leq Ch$ for the RFBe method. □

We conclude this section with some comments. The correct choice of the value of edge bubbles on the edges that intersect the boundary layer is hinted by the error analysis. In particular, the value on any such edge is chosen to be equal to the boundary layer correction term in the asymptotic approximation of the solution while bounding the error $u_{as}-v_P$. This choice resulted in the $O(h)$ term in the error bound (4.43).

In the context of convection–diffusion problems, we say that a method is uniformly convergent of order α with respect to some norm $||\cdot||_\varepsilon$ (which may depend on ε), if an error bound of the form

$$||u-u_h||_\varepsilon \leq Ch^\alpha,$$

holds for some positive constant α that is independent of ε as $h\to 0$. In the pre–asymptotic regime, when $\varepsilon \lesssim h$, we observe for the RFBe method that $\sqrt{\varepsilon}|u-u_a|_{1,\Omega}\leq Ch$, with C independent of ε.

It is interesting to note that all methods studied in the literature that achieve uniform convergence on shape regular partitions are based on the use of ad–hoc exponential basis functions (see [82], pp. 273-8). In particular, we refer to the conforming methods in [78] and [90] and the non-conforming method in [2]; the methods involve exponentially fitted splines

in the definition of the trial and test spaces, constructed as tensor products of solutions of the restriction of the original equation on the edges of the triangulation. They all achieve a uniform rate of convergence with $\alpha = 1/2$ in the $\sqrt{\varepsilon}$–weighted energy norm.

The method by Schieweck [90], similarly to our method, considers basis functions constructed using exponentials only on those edges which cross the boundary layer. Elsewhere, it uses a standard bilinear approximation. Schieweck proves that, for his method,

$$\varepsilon^{1/2}|u - u_h|_{1,\Omega} \leq C\left((\varepsilon/h)^{1/2} + h^{1/2}\right).$$

Our method can be seen as a combination of the residual–free bubble method with exponential fitting on edges contained in the boundary layer.

We know from the local *a priori* error analysis of Sangalli [85] that, for $\varepsilon \lesssim h$, away from the boundary layer region, the RFB method is also $O(h)$ accurate in the energy norm. The error bound (4.50) suggests that an identical result does not hold in the energy norm on the whole of Ω, unless the approximation properties in the layers are improved. Moreover, we have identified the inferior approximation of the boundary layer behaviour along the edges of the triangulation as the main source of the inaccuracy of the RFB method. This observation will also be confirmed by our numerical experiments.

Figure 4.2 may help to clarify why refining the mesh need not improve the accuracy of the RFB method in the pre–asymptotic regime of $\varepsilon \lesssim h$. The plots show the RFB approximation to the solution of the boundary–value problem described in Example 2, page 85, on two subsequent uniform meshes. The lengths of the edges crossing the boundary layer are halved as we half the mesh size, but the number of such edges is doubled. Hence convergence is impeded until the mesh starts to resolve the boundary layer.

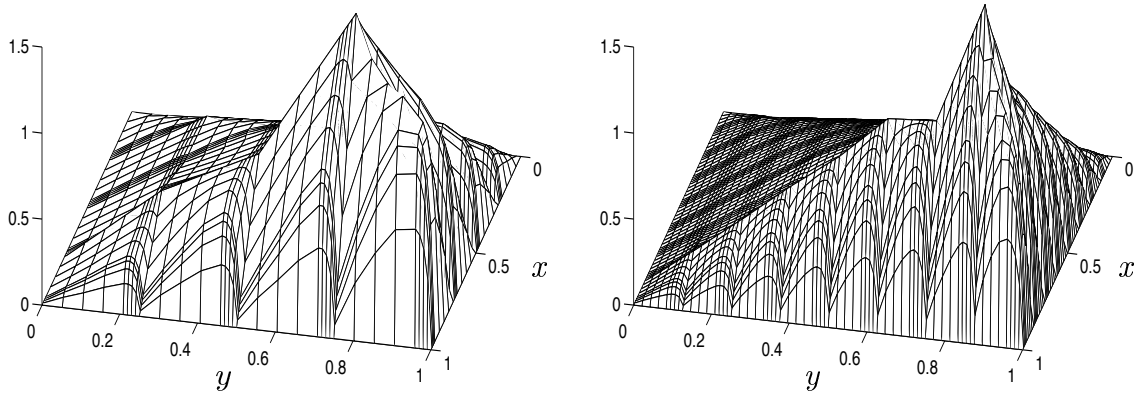
This argument, of course, can also be seen as further evidence of the advantages of anisotropic mesh refinement. Indeed, returning to the bound (4.50) on term IV which does not imply convergence of term IV to zero under mesh refinement, we see that this is due to the fact that

$$\sum_{\gamma \in \Gamma_{bl}} h_{\gamma_i} = h_{\gamma_i} \#(\Gamma_{bl}).$$

Hence, on a succession of anisotropically refined meshes, graded in the normal direction to the boundary layer, we achieve reduction of h_{γ_i} (whilst keeping $\#(\Gamma_{bl})$ fixed), and hence we should expect improvement in the accuracy of the solution even in the pre–asymptotic regime.

4.4 Full discretisation and numerical examples

As discussed in Section 2.2.1 and at the beginning of this chapter, one can consider different splittings of V_{RFB}, the only constraint being that each element of V_{RFB} must be linear on every edge of the partition.



(a) $h = 1/4$ (b) $h = 1/8$

Figure 4.2: The RFB solution of the problem of Example 2. The problem parameters are $\varepsilon = 10^{-2}$, $\boldsymbol{a} = (\cos(\pi/4), \sin(\pi/4))$.

For each $v_h \in V_h$ we define $\widetilde{v}_h \in V_{RFB}$ such that

$$\begin{cases} \widetilde{v}_h - v_h \in B_h & \text{and} \\ \mathcal{L}(\widetilde{v}_h, v_b) = 0 & \forall v_b \in B_h. \end{cases} \tag{4.51}$$

In this way we construct a new subspace $\widetilde{V}_h$ which coincides with the space V_l as defined in (2.20). Moreover, we still have

$$V_{RFB} = \widetilde{V}_h \oplus B_h,$$

which again leads to the augmented space formulation discussed in Section 2.2.1. Similarly, $V_a = \widetilde{V}_h \oplus E_h \oplus B_h$ where, this time, $\widetilde{V}_h \oplus E_h = V_l$, and the unique solution of (4.17) can be rewritten as $u_a = \widetilde{u}_h + u_e + u_b^f$ where, as we have seen in Section 2.2.1, u_b^f is the solution of

$$\mathcal{L}(u_b^f, v_b) = (f, v_b) \qquad \forall v_b \in B_h.$$

Equivalently, $u_b^f{}_{|T} = b_f$ for any $T \in \mathcal{T}_h$, where b_f solves (4.21). Further, testing with $w_h \in \widetilde{V}_h \oplus E_h$, we have that $\widetilde{u}_h + u_e$ satisfies

$$\mathcal{L}(\widetilde{u}_h + u_e, w_h) = (f, w_h) - \sum_{T \in \mathcal{T}_h} \mathcal{L}(b_f, w_h) \qquad \forall w_h \in \widetilde{V}_h \oplus E_h, \tag{4.52}$$

which is equivalent to (2.23).

The formulation (4.52) is similar to the *multiscale finite element method* (MFEM) defined in [59] for the solution of symmetric elliptic problems, enhanced by edge bubbles. The MFEM method (see Section 2.5) is a Galerkin method in which the local basis functions are defined by solving boundary value problems for the original equation. In our case, given an element $T \in \mathcal{T}_h$ and letting $\{\varphi_i\}_{i=1}^4$ be the standard bilinear basis functions related to T, the corresponding basis functions $\{\widetilde{\varphi}_i\}_{i=1}^4$ for the space $\widetilde{V}_h$ are obtained, according to (4.51), by solving, for $i = 1, 2, 3, 4$, the local boundary-value problems

$$\begin{cases} \widetilde{\varphi}_i \in H^1(T) \ \text{ such that} \\ \mathcal{L}(\widetilde{\varphi}_i, v_b) = 0 \qquad \forall v_b \in H_0^1(T) \\ \mathrm{Tr}(\widetilde{\varphi}_i) = \varphi_i \qquad \text{on } \partial T. \end{cases} \tag{4.53}$$

Since the basis functions have to be evaluated numerically, the algorithm is, clearly, of a two-level type.

We have chosen to use the formulation (4.52)-(4.53) as the starting point for the full discretisation of the method. Once more, we would like to stress the fact that the two formulations, (4.22) and (4.52), are equivalent (more on this in Section 2.5). Moreover the two approaches involve the same number of subgrid computations. Nevertheless, we have found that the formulation given by (4.52) is simpler to code and results in slightly faster computations.

In order to solve the boundary–value problems that define the basis functions of V_a, we need to introduce a subgrid. We do so by considering a sub-partition $\mathcal{T}_N$ of the original partition, where N is the discretization parameter of the new partition. The restriction of $\mathcal{T}_N$ to any element $T \in \mathcal{T}_h$ defines a partition of T that we use to solve the problems (4.53), (4.21) and (4.15) which define, respectively, the local basis for the space $\widetilde{V}_h$, the bubble related to the forcing term and, finally, the edge bubbles.

Accordingly, on selecting V^N as the bilinear finite element space over $\mathcal{T}_N$, we define:

1. the discrete counterpart of the bubble space (4.3) as
$$B^N = B_h \cap V^N;$$
2. the space of edge bubbles as the subspace of V^N given by
$$E^N = \mathrm{span}\,\{e_j^N,\, j = 1, \ldots, N_e\} \subset V^N,$$
where e_j^N is obtained by solving in V^N the boundary value problem (4.15). The relevant boundary conditions (i.e. the value of the edge bubble on the edges) are defined as the piecewise linear interpolant of the solution to (4.14);
3. the discrete counterpart $\widetilde{V}_h^N$ of $\widetilde{V}_h$ as
$$\mathrm{span}\,\{\widetilde{\varphi}_i^N,\, i = 1, \ldots, 4\} \qquad \forall T \in \mathcal{T}_h,$$
where $\widetilde{\varphi}_i^N = \widetilde{\varphi}_i^{N,T}$ is the solution in V^N of (4.53) on $T \in \mathcal{T}_h$.

The fully discrete RFBe space is then defined as

$$V_a^N = \widetilde{V}_h^N \oplus E^N \oplus B^N,$$

and the fully discrete RFBe formulation reads

$$\begin{cases} \text{find } u_a^N \in V_a^N \text{ such that} \\ \mathcal{L}(u_a^N, v_a^N) = (f, v_a^N) \qquad \forall v_a^N \in V_a^N, \end{cases} \tag{4.54}$$

and can be again rewritten in the form (4.52).

The analytical study of the size of the additional error due to the introduction of the subgrid (*subgrid discretisation error*) in the fully discrete formulation (4.54) does not appear to be straightforward; the main difficulty being that Lemma 4.3.5 is no longer applicable at the fully discrete level. Thus, we shall assess the impact of subgrid discretisation on the accuracy of the method through numerical experiments.

4.4.1 Numerical examples

We now present some numerical experiments, carried out in MATLAB, using the formulation (4.54) in order to validate the *a priori* error analysis and assess the size of the numerical error due to the discretisation at the subgrid level.

Example 1. We consider the boundary–value problem

$$\begin{cases} Lu := -\varepsilon \Delta u + \boldsymbol{a} \cdot \nabla u = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \end{cases}$$

where

$$\boldsymbol{a} = \frac{3 - xy}{2}(1,1); \qquad f = 2; \qquad g = \begin{cases} \frac{1}{2}(1 + \cos(5\pi x)) & \text{if } y = 0 \\ \frac{1}{2}(1 + \cos(3\pi y)) & \text{if } x = 0 \\ 0 & \text{otherwise;} \end{cases}$$

see Figure 4.3 for a sample solution corresponding to $\varepsilon = 1/50$.

We solve this model problem on a sequence of uniform meshes using both the RFB and the RFBe methods. For each computation, the subgrid mesh is axiparallel and of Shishkin type with turning point $\lambda = c_s \frac{\varepsilon}{c_a} \ln N$ and the same value of the Shishkin parameter $c_s = 1/4$; see Appendix A.2 for more details regarding Shishkin meshes.

We wish to confirm the *a priori* bounds on the energy–norm error both in terms of the mesh size h and the diffusion parameter ε. In order to do so, we need to ensure that the subgrid computations are accurate enough.

Figures 4.4(a) and 4.4(b) show the global error in the $\sqrt{\varepsilon}$–weighted energy norm for a fixed mesh size but different values of the subgrid mesh parameter N for, respectively, $\varepsilon = 1/50$ and $\varepsilon = 10^{-2}$.

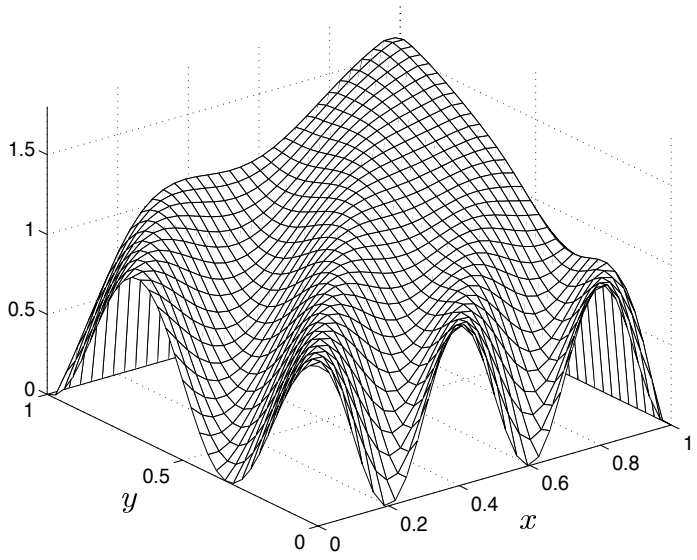


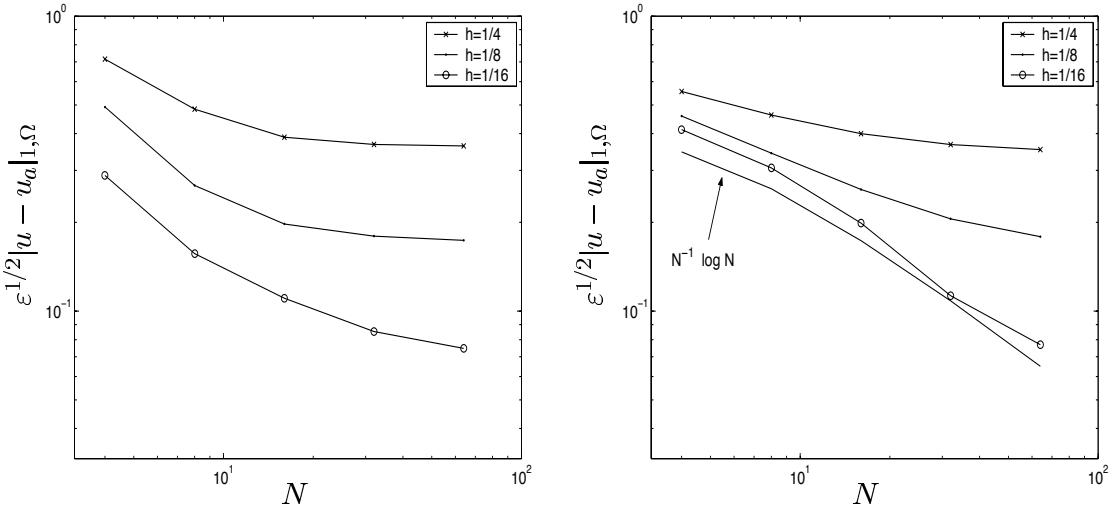
Figure 4.3: Example 1. Solution on a 32×32 uniform mesh, $\varepsilon = 1/50$.

We notice that a substantial reduction of the error is obtained just by refining the subgrid, until the macro discretisation error becomes dominant and refining the subgrid thereafter yields no improvement in the overall accuracy. This happens later for smaller values of ε and h. For instance, for $\varepsilon = 10^{-2}$ and $h = 1/16$ (see Figure 4.4(b)), we initially observe the characteristic $N^{-1} \log N$ convergence rate on Shishkin meshes (see Appendix A.2), showing that the subgrid discretisation error dominates the overall computational error in this case.

Based on such computational evidence, we may infer that the error bound for the fully discrete method should include a term proportional to the error in the numerical approximation of the elemental basis functions. This would be in accordance with the findings of Sangalli [87] on a different augmented RFB formulation applied to a symmetric problem. (For the problems considered in [87], Sangalli's argument is applicable to our method as well).

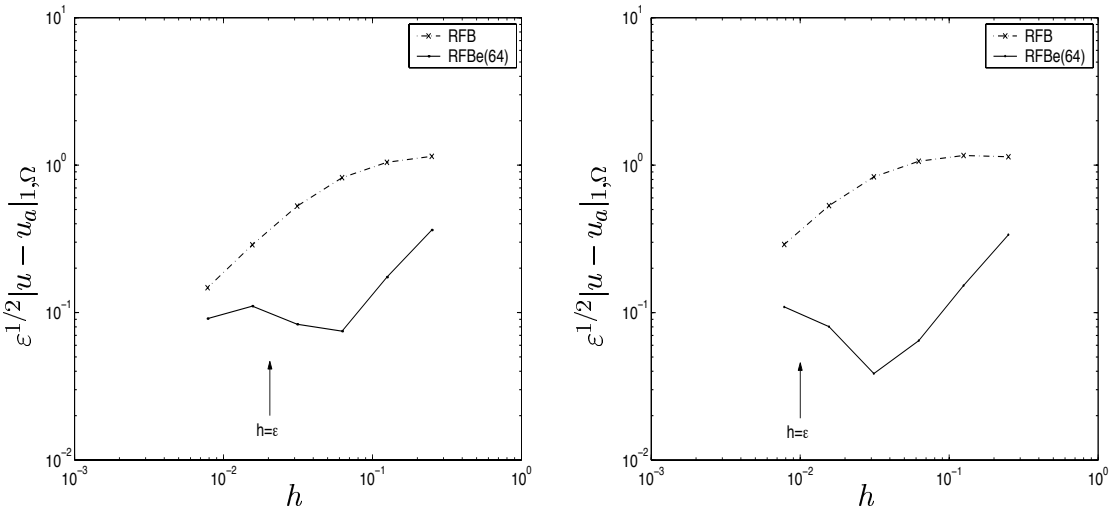
The convergence in terms of the mesh parameter h is shown in the loglog plots of Figure 4.5. Since we do not have at hand the exact solution, the error is evaluated using a reference solution given by Richardson extrapolation using two numerical solutions obtained on Shishkin–type meshes with, respectively, 256 and 512 nodes in each coordinate direction.

The RFBe solution initially converges to the reference solution with rate 1. As we keep refining, the slope of the error curve changes sign (the rate becomes approximately $-1/2$) until the error curve joins the corresponding error curve for the RFB method. See also the rates are listed in Table 4.1). These are defined by confronting two successive results as follows. Let e_{h_1} and e_{h_1} be the errors recorded by solving with mesh sizes $h_1 > h_2$. Then the rate is given by



(a) Error vs. N, $\varepsilon = 1/50$ (b) Error vs. N, $\varepsilon = 10^{-2}$

Figure 4.4: Example 1. $\sqrt{\varepsilon}$–weighted energy norm error as a function of N.



(a) Error vs. h, $\varepsilon = 1/50$ (b) Error vs. h, $\varepsilon = 10^{-2}$

Figure 4.5: Example 1. $\sqrt{\varepsilon}$–weighted energy norm error as a function of h: RFB method and RFBe method with $N = 64$ (RFBe(64)).

	$\varepsilon = 1/50$				$\varepsilon = 10^{-2}$			
h	Energy	rate	L_2	rate	Energy	rate	L_2	rate
1/4	0.363		0.141		0.337		0.17	
1/8	0.174	1.06	0.043	1.7	0.152	1.14	0.051	1.73
1/16	0.074	1.22	0.011	1.92	0.064	1.24	0.0138	1.89
1/32	0.083	−0.15	0.0042	1.43	0.038	0.74	0.0036	1.93
1/64	0.11	−0.41	0.0025	0.77	0.08	−1.05	0.0025	0.53
1/128	0.091	0.28	0.00103	1.28	0.109	−0.45	0.0017	0.55

Table 4.1: Example 1. Error and convergence rate for RFBe(64).

the exponent $\alpha \in \mathbb{R}$ satisfying the equality

$$\frac{e_{h_1}}{e_{h_2}} = \left(\frac{h_1}{h_2}\right)^{\alpha}.$$

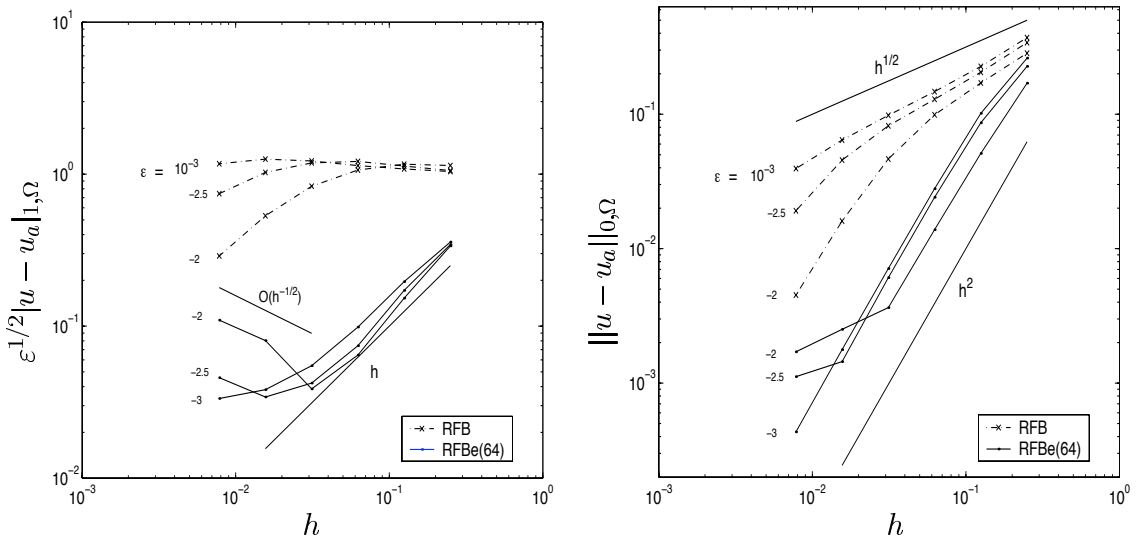
The relatively inaccurate results obtained when $h \sim \varepsilon$ are easy to explain. The region Ω_{bl}, i.e. the union of the elements enriched with the edge bubbles, is properly contained in the boundary layer region, hence part of the boundary layer behaviour cannot be accurately captured. Eventually, the mesh is fine enough to resolve the layer, and the asymptotic convergence rate of the method is then recovered.

As for the subgrid computations, we have used the value $N = 64$ (hence the notation RFBe(64)). As we see from Figure 4.4, this choice ensures that the subgrid discretisation error is of higher order. For smaller values of ε we can still identify the rate of convergence predicted by the *a priori* analysis; see Figure 4.6(a).

As regards the L_2–norm error, not covered by our theorem, the RFBe method seems to be second–order 'convergent' in the preasymptotic regime when $\varepsilon \ll h$. The local *a priori* analysis of Sangalli [85] predicts the same rate of convergence for RFB in the outside region.

So far we have not taken into account computational cost. The RFBe method is computationally more expensive then RFB on the same triangulation since it involves a larger number of degrees of freedom, and because the extra d.o.f., the edge bubbles, need to be computed.

Assume, for example, that the triangulation is uniform with $n \times n$ elements; the number of edge bubbles is then $2(n-1)$, while the total number of d.o.f. is of $O(n^2)$. Moreover, assume that the basis functions (4.53) are calculated using a 4×4 subgrid, as in all computations presented so far. As for the edge bubbles, we are free to consider finer subgrids. We then have a range of possible scenarios. If, for example, the edge bubbles are also computed on 4×4 subgrids, the computational times for the two methods (RFB and RFBe) are almost identical. On the other hand, if relatively fine subgrids are used for the edge bubbles (32×32 in our example below), the CPU time is dominated by the computation of the edge bubbles, hence RFBe becomes much more expensive.



(a) $\sqrt{\varepsilon}$–weighted energy norm error (b) L_2-norm error

Figure 4.6: Example 1. Error as a function of h for different values of ε: the RFB method and the RFBe method with subgrid discretisation parameter $N = 64$ (RFBe(64)).

Nevertheless, since the boundary layer is a major source of error, the RFBe method can be more effective even if the edge bubbles are calculated with high precision. This is particularly true if we require a certain accuracy in the energy norm of the solution. In the Table 4.2, we compare the computational time required by a sequential code implementing the RFBe(32), RFBe(4), RFB methods and a standard Galerkin finite element method to compute the solution to a fixed tolerance (TOL) in the $\sqrt{\varepsilon}$–weighted energy norm, for $\varepsilon = 10^{-2}$. (The procedure would be speeded up through parallelisation of the subgrid computations). The table reports the CPU time in seconds on a Pentium III 800 MHz processor; the corresponding number of elements on each coordinate direction is given by the numbers in square brackets. We have left the entries of the table blank if to achieve the required accuracy a *resolving* mesh was necessary, i.e. if $n > 100$.

We notice that the RFB and Galerkin methods are unable to provide reasonable accuracy in the energy norm on unresolving meshes. This situation would become even more evident if we were to consider problems with smaller values of ε. Having said this, the accuracy of the RFBe method on unresolving meshes is also limited albeit to a much lesser extent than the accuracy of RFB and standard Galerkin methods (cf. Figure 4.6).

Similarly, Table 4.3 reports the computational time to achieve a fixed accuracy in the L_2–norm, for $\varepsilon = 10^{-2}$, 10^{-3} and 10^{-4}.

Again, the RFBe method is the most effective in almost all the cases considered, in partic-

	$\varepsilon = 10^{-2}$			
TOL	RFBe(32)	RFBe(4)	RFB	Galerkin
1/2	6 [3]	2.2 [10]	77 [68]	14 [94]
1/5	16 [7]	59 [58]		
1/10	46 [18]			

Table 4.2: Example 1. Computational time (in seconds) to achieve, on a uniform mesh, a given accuracy (TOL) in the $\sqrt{\varepsilon}$–weighted H^1–norm. We indicate in square brackets the number of elements used in each coordinate direction on the global uniform mesh. RFBe(N) indicates that an $N \times N$ subgrid was used for computing the edge bubbles. A 4×4 subgrid is used to compute the RFB (internal) bubbles in each case.

ular for the smaller values of ε. More precisely, we notice the following:

1. For RFBe(32), the evaluation of the edge bubbles takes almost all the CPU time.
2. While this is the case, the growth of the computational time is roughly linear, while the method is converging quadratically (again, see Figure 4.6).
3. On the other hand, the computational cost of all the other methods considered grows quadratically, while their rate of convergence is less then quadratic.

Hence, as the tolerance becomes tighter, the RFBe(32) method requires a smaller amount of CPU time than the other methods.

Finally, we remark that the mesh used to evaluate the basis functions (the internal bubbles if we think in terms of RFB) could also be adjusted. The presence of the edge bubbles improves the stability of the RFB method (as we shall show in the next example), hence the same accuracy in the outside region can be achieved using poorer approximation to the internal bubbles. If we switch off the bubbles in the outside region, we obtain a method similar to the locally exponentially fitted finite element method in [90]; if no other layers are present (as in this example), this yields almost as accurate results as the full RFB method.

Example 2. We solve the boundary value problem with constant coefficients

$$\begin{cases} -\varepsilon\Delta u + \left(\cos(\frac{\pi}{4}), \sin(\frac{\pi}{4})\right) \cdot \nabla u = 1 & \text{in } \Omega = (0,1)^2, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{4.55}$$

In Figure 4.7 we see the numerical solution obtained by using an RFBe method on a uniform mesh with $h = 1/4$ and $\varepsilon = 10^{-2}$ as a sum of its components in the spaces B_h, E_h and V_h.

The convergence in terms of the mesh parameter h in the $\sqrt{\varepsilon}$–weighted energy norm for $\varepsilon = 10^{-2}$ is shown in Figure 4.9. As for the subgrid computations, we have used axiparallel Shishkin subgrids with $N = 20$ and $N = 4$ and the Shishkin parameter $c_s = 1/2$.

	$\varepsilon = 10^{-4}$			
TOL	RFBe(32)	RFBe(4)	RFB	Galerkin
1/5	14 [5]	1 [6]	1.1 [8]	272 [200]
1/10	26 [9]	2.2 [10]	14 [30]	
1/50	52 [20]	105 [78]		
1/100	92 [28]			

	$\varepsilon = 10^{-3}$			
TOL	RFBe(32)	RFBe(4)	RFB	Galerkin
1/5	14 [5]	1 [6]	1.8 [10]	5 [62]
1/10	22 [8]	2 [9]	17 [32]	32 [120]
1/50	49 [19]	39 [46]		
1/100	71 [26]	124 [84]		

	$\varepsilon = 10^{-2}$			
TOL	RFBe(32)	RFBe(4)	RFB	Gal
1/5	6 [3]	0.6 [4]	1 [6]	0.3 [17]
1/10	14 [6]	1 [6]	2 [16]	0.8 [28]
1/50	35 [14]	4.5 [15]	52 [56]	7 [74]
1/100	52 [20]	9 [22]	137 [90]	21 [106]

Table 4.3: Example 1. Computational time (in seconds) to achieve, on a uniform mesh, a given accuracy (TOL) in the L_2–norm. We indicate in square brackets, the number of elements used in each coordinate direction on the global uniform mesh.

We notice that, for this problem, the error is almost completely concentrated in the boundary layer, so mesh refinement, particularly in the outside region, is of secondary importance in the regime $h > \varepsilon$.

In contrast with this, as can be seen in Figure 4.9(b) which reports the error in the outside region $\Omega_{out} = (0, 1-h)^2$, the accuracy of the RFBe method away from the boundary layer is largely independent from that of the subgrid computations. As long as $\varepsilon < h$, the numerical solutions obtained by using the RFBe method are more accurate than those delivered by the RFB method: the edge bubbles have the effect of eliminating the over– and under–shooting near the boundary layer typical of RFB (and of most stabilised finite element methods). This can be seen by comparing Figures 4.2 and 4.7.

A similar result was obtained by Mizukami and Hughes [75] with their shock–capturing method which has the additional property of satisfying the discrete maximum principle. Example 4.1 in [75] ($f = 0$, $\boldsymbol{a} = (\cos(\pi/6), \sin(\pi/6))$, $\varepsilon = 10^{-7}$ and homogeneous Dirichlet

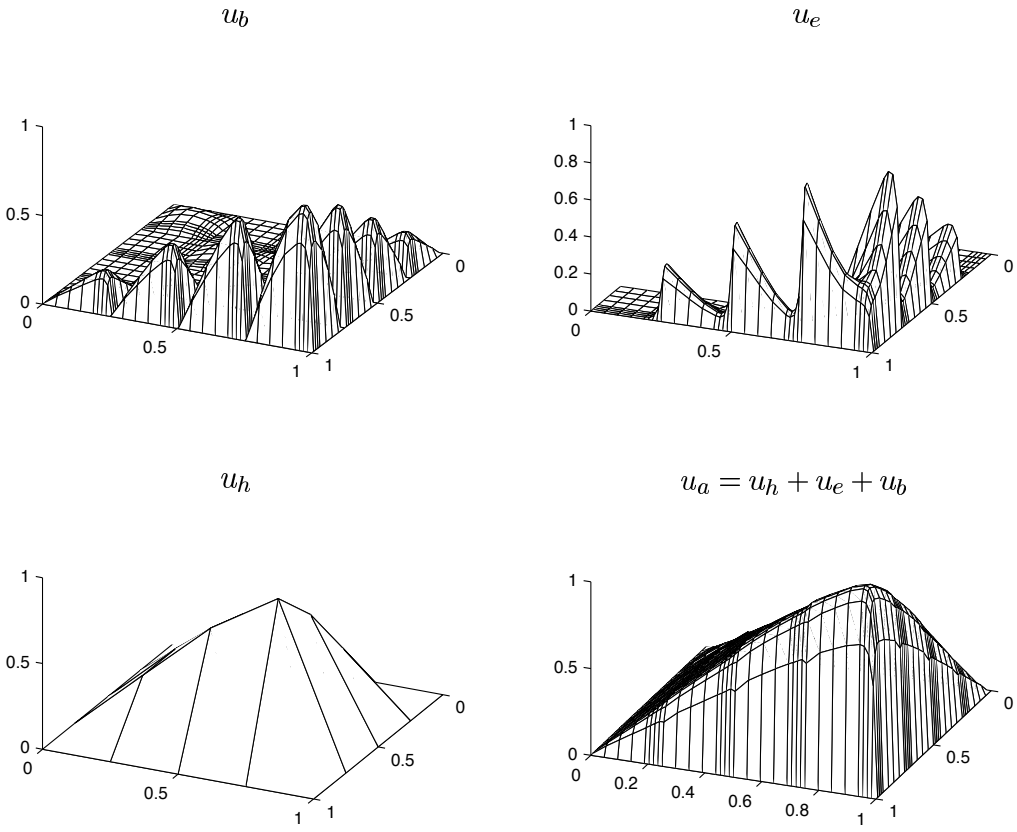
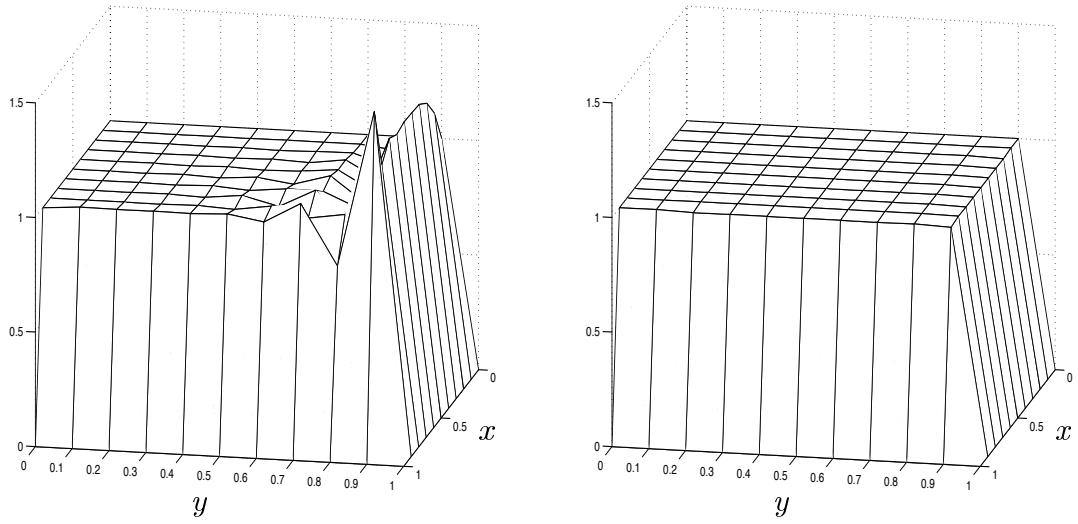


Figure 4.7: Example 2. The RFBe solution of the boundary value problem (4.55) on a 4×4 uniform mesh for $\varepsilon = 10^{-2}$. The bubbles are computed using a 8×8 Shishkin subgrid.

boundary conditions at $x = 1$ and $y = 1$; Dirichlet boundary condition $u = 1$ otherwise) highlights the problem of over– and under–shooting near the boundary layer. The RFBe solution to this problem is nodally very accurate; see Figure 4.8.

The energy norm error of the numerical solution of (4.55) for different values of ε on a fixed uniform mesh of size $h = 1/8$ is depicted in Figure 4.10(a). For small values of ε, since we cannot afford to use very fine subgrids, the error due to the approximate computation of the bubbles is dominant. On the other hand, the lower bound on the error given by the error in the outside region, which is largely independent of the subgrid size, permits us to confirm the *a priori* bounds (see Figure 4.10(b)).

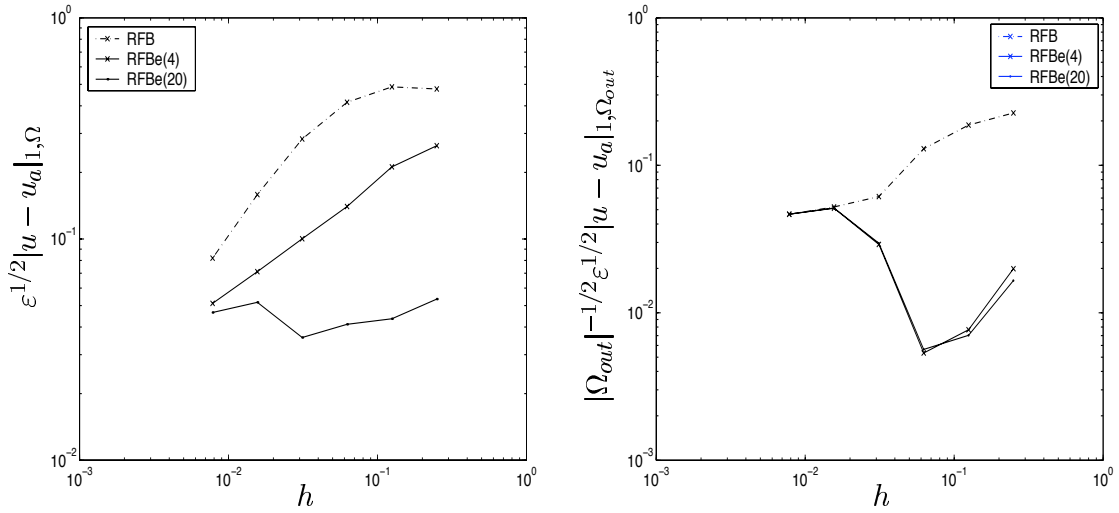
We have verified that the overall accuracy is improved by the introduction of the edge bubbles. The method achieves both an increased local resolution and a global improvement in accuracy in comparison with RFB (see Figure 4.9).



(a) RFB solution

(b) RFBe solution

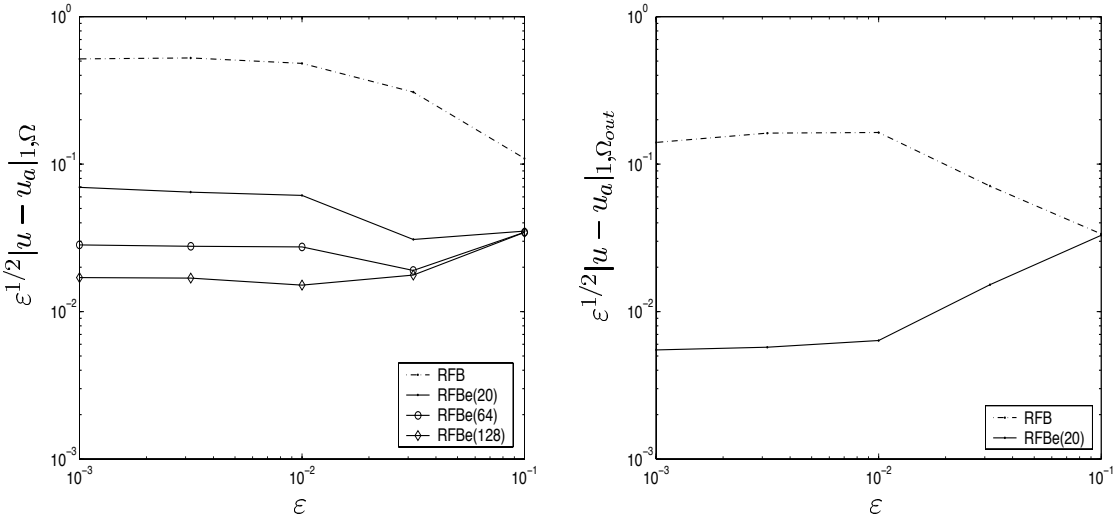
Figure 4.8: Solution of Example 4.1 in [75]: $f = 0$, $\boldsymbol{a} = (\cos(\pi/6), \sin(\pi/6))$, $\varepsilon = 10^{-7}$ and Dirichlet homogeneous boundary conditions at $x = 1$ and $y = 1$; Dirichlet boundary condition 1 otherwise. The edge bubbles were computed using an 8×8 subgrid.



(a) Energy error in Ω

(b) Error away from the layers

Figure 4.9: Example 2. $\sqrt{\varepsilon}$–weighted energy norm error as a function of h.



(a) Error vs ε in Ω (b) Error vs ε in Ω_{out}

Figure 4.10: RFB and RFBe $\sqrt{\varepsilon}$–weighted energy norm error on a uniform mesh of size $h = 1/8$ for different values of ε.

In practice, the bases for the bubble spaces must be evaluated numerically by introducing a subgrid. Since the edge bubbles are not eliminated via static condensation, one may be led to think that the method should be quite sensitive to the accuracy to which the numerical solution is to be computed. In fact, this seems to be true only inasmuch as accuracy within the layer is concerned, as we can see by comparing the plots of Figure 4.9(a) and (b).

Remark. In general, we may not know *a priori* where the inclusion of the edge bubbles is required. Thus, we believe that the RFBe method should be thought of as a corrector in an iteration whose predictor is the RFB method, followed by a loop of *a posteriori* error estimation aimed at locating elements where edge bubbles need to be inserted; work in this direction is in progress and will be discussed in Chapter 5.

We conclude with an example taken from [22] and with a remark concerning the application of the RFBe method to problems that exhibit internal layers.

Example 3. We solve the boundary value problem considered in [22] using the RFB

method:

$$\begin{cases} -\epsilon\Delta u + (\cos(\pi/3), \sin(\pi/3)) \cdot \nabla u = 1 & \text{in } \Omega = (0,1)^2 \\ u = 1 \quad \text{for} \begin{cases} x \leq 1/2, y = 0 \\ x = 0 \end{cases} \\ u = 0 \quad \text{otherwise.} \end{cases} \tag{4.56}$$

This problem is not covered by our analysis: because of the presence of an internal layer, the asymptotic approximation we used does not satisfy (4.31). Still, the RFBe results show a considerable improvement over the results obtained by the RFB method.

The solutions obtained by means of the RFB and RFBe methods on a uniform mesh with $h = 1/20$ for $\varepsilon = 10^{-6}$ are shown in Figure 4.11. We notice that the oscillations near the boundary layer and the spike in the corner layer present in the RFB solution are absent from the RFBe solution. This highlights the fact that such undesirable features of the RFB method are due to poor resolution of the boundary layer on the skeleton of the triangulation.

On the other hand, no visible improvement over the accuracy of RFB is obtained along the internal layer propagating from the point of discontinuity in the boundary condition. One would expect that carefully chosen edge bubbles defined near the layer itself will eliminate this numerical inaccuracy.

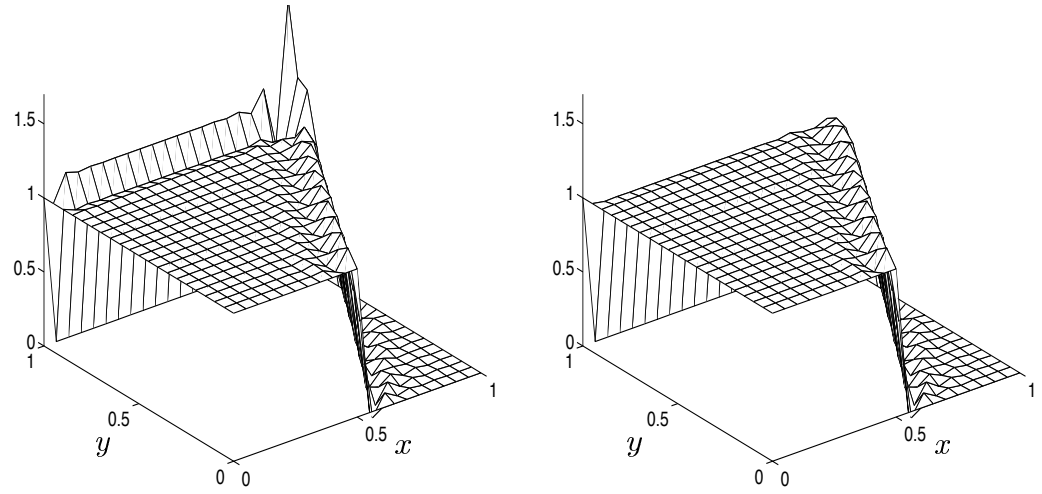
4.5 Closing remarks

We have shown how a small number of edge bubbles can be defined to improve the resolution of boundary layers of the RFB method in the context of convection–dominated convection–diffusion problems.

The resulting scheme has better accuracy properties then RFB in the regime $\varepsilon < h$. Indeed, both our *a priori* error analysis and our numerical experiments show that on coarse meshes the RFB method shows little or no improvement of accuracy in the energy norm, while RFBe exhibits the optimal asymptotic rate of convergence in the energy norm on Ω; in the pre–asymptotic regime $\varepsilon < h$ standard RFB is only capable of reproducing the same level of accuracy as RFBe only away from the layers.

Moreover, we have noticed that, although we are acting locally, the RFBe scheme exhibits increased accuracy globally, indicating that the introduction of the edge bubbles has a stabilising effect. More precisely, our method, similarly to the shock–capturing method presented in [75], shows no over– and under–shooting near the boundary layer. The observed improvement is obtained robustly with respect to the subgrid size and hence at (almost) no extra computational cost.

Another characteristic feature of RFBe is that accuracy inside the boundary layer region is sensitive to the accuracy to which the edge bubbles have been computed; thus, only if high precision in such region is required, should one consider performing expensive and accurate



(a) RFB (b) RFBe

Figure 4.11: Example 3. Solution of a problem with an internal layer on a uniform mesh of size $h = 1/20$. The problem parameters are $\varepsilon = 10^{-6}$, $\boldsymbol{a} = (\cos(\pi/3), \sin(\pi/3))$; as for the subgrid, a Shishkin mesh with $N = 4$ has been used.

calculations of the edge bubbles.

Following the ideas of Brezzi & Marini [25], we presented the RFBe method in a general form, suggesting that other multiscale problems for which the fine scale features are only locally present may benefit from the introduction of edge bubbles. The principle is that we are ready to afford the introduction of only a small number of new degrees of freedom; hence, for the procedure to work efficiently, we must assume that it is known *a priori* where the bubbles have to be added. The *a priori* information may be obtained from previous computations with, or without, the use of *aposteriori* error bounds (see the next chapter and in particular Section 5.4).

Chapter 5

A posteriori error estimators and RFB

This chapter is devoted to the *a posteriori* analysis of the RFB method in its generalised Galerkin (or stabilised) formulation (2.11). The efficiency of an adaptive mesh refinement algorithm based on the obtained *a posteriori* error bounds is also discussed.

Previous results have been obtained by Russo [84] and Sangalli [86] specifically on the subject of *a posteriori* evaluation of norm error.

Here we consider both linear functional error evaluation and norm error evaluation. In both cases, we find that the *a posteriori* error bound is composed of three terms: the two classical residual–based terms of the Galerkin formulation (internal residual and boundary jump of the gradient), as well as a third term due to static condensation of the bubbles.

In Section 5.4 we propose to use the relative magnitude of such terms to explore whether the bubble (stabilisation) can be phased off locally. A new *hb*–adaptive algorithm is studied: the idea is to turn off the bubble locally (*b*–derefinement) while refining the mesh (*h*–refinement). We investigate the robustness of the algorithm through numerical examples.

5.1 *A posteriori* dual–weighted error bounds

Our aim is to combine the use of the RFB method for the numerical solution of boundary value problems for convection–dominated convection–diffusion equations with mesh adaptation techniques.

Let us consider the convection–diffusion operator

$$Lu := -\varepsilon\Delta u + \boldsymbol{a}\cdot\nabla u,$$

where ε is a positive constant and the velocity field $\boldsymbol{a} \in L^\infty(\Omega)^2$ is divergence-free in an open polygonal domain $\Omega \subset \mathbb{R}^2$. Given a function $f \in L_2(\Omega)$, we consider the associated

homogeneous boundary-value problem in variational form

$$\begin{cases} \text{find } u \in V \text{ such that} \\ \mathcal{L}(u,v) = (f,v) \qquad \forall v \in V, \end{cases} \tag{5.1}$$

where $V = H_0^1(\Omega)$ and the bilinear form $\mathcal{L}(\cdot,\cdot)$ is defined on $V \times V$ as

$$\mathcal{L}(w,v) = \varepsilon \int_\Omega \nabla w \cdot \nabla v \, d\boldsymbol{x} + \int_\Omega (\boldsymbol{a} \cdot \nabla w)\, v \, d\boldsymbol{x}.$$

It is often the case that the quantity of interest is not the solution u itself but a linear functional $u \to J(u)$ of the solution (e.g. a point value, a flux, an average, etc.). *A posteriori* bounds on the error $J(u) - J(u_h)$, where u_h is the computed solution, can be obtained through duality arguments. Here we assume that

$$J(u) = (u,g), \qquad g \in L_2(\omega),$$

leaving the treatment of other linear functionals such as fluxes through a Neumann boundary to a later example.

We define the following dual problem:

$$\begin{cases} \text{find } z \in V \text{ such that} \\ \mathcal{L}(w,z) = J(w) \qquad \forall w \in V. \end{cases} \tag{5.2}$$

The (adjoint) differential operator involved in (5.2) can be recovered from (5.2) through integration by parts; for the partial differential operator L under consideration, it will be found to be defined by

$$L^* z := -\varepsilon \Delta z - \boldsymbol{a} \cdot \nabla z \qquad z \in V.$$

We shall perform such an *a posteriori* error analysis for the RFB method seen as a generalised Galerkin approximation for the polynomial part of the solution. That is, we assume that static condensation of the bubble part of the solution has been performed resulting in a stabilised finite element approximation on the given piecewise polynomial space V_h and triangulation $\mathcal{T}_h$; i.e., a formulation which we can write as

$$\begin{cases} \text{find } u_h \in V_h \text{ such that} \\ \mathcal{L}(u_h, v_h) + \sum\limits_{T \in \mathcal{T}_h} (u_b, L^* v_h)_T = (f, v_h) \qquad \forall v_h \in V_h. \end{cases} \tag{5.3}$$

Here, the family of triangulations $\mathcal{T}_h$, $h > 0$, is *admissible* (*conforming*) if any two triangles in $\mathcal{T}_h$ either have a common edge or common vertex, or they do not intersect at all.

In the sequel, we denote by $\boldsymbol{n}_T$ the unit outward normal vector to ∂T defined on the edges of any element T. Further, we denote by $\boldsymbol{n}_T \cdot [\nabla u_h]$ the jump of the normal derivative of u_h across the given edge.

We have

$$\begin{aligned} J(u) - J(u_h) &= J(u - u_h) \\ &= \mathcal{L}(u - u_h, z) \\ &= \mathcal{L}(u - u_h, z - z_h) + \sum_{T \in \mathcal{T}_h} (u_b, L^* z_h)_T \\ &= \sum_{T \in \mathcal{T}_h} ((f - Lu_h, z - z_h)_T - (\varepsilon \boldsymbol{n}_T \cdot \nabla u_h, z - z_h)_{\partial T \cap \Omega}) + \sum_{T \in \mathcal{T}} (u_b, L^* z_h)_T \\ &= \sum_{T \in \mathcal{T}_h} \left((R_T(u_h), z - z_h)_T - \frac{1}{2} (\varepsilon \boldsymbol{n}_T \cdot [\nabla u_h], z - z_h)_{\partial T \cap \Omega} + (u_b, L^* z_h)_T \right) \\ &= \sum_{T \in \mathcal{T}_h} \left(\left(\rho_T^{(1)}, \omega_T^{(1)} \right)_T + \left(\rho_T^{(2)}, \omega_T^{(2)} \right)_{\partial T \cap \Omega} + \left(\rho_T^{(3)}, \omega_T^{(3)} \right)_T \right), \end{aligned} \tag{5.4}$$

having denoted the elemental residual terms by

$$\rho_T^{(1)} = R_T(u_h), \qquad \rho_T^{(2)} = -\frac{1}{2} \varepsilon \boldsymbol{n}_T \cdot [\nabla u_h], \qquad \rho_T^{(3)} = u_b,$$

and defined the weights

$$\omega_T^{(1)} = \omega_T^{(2)} = z - z_h, \qquad \omega_T^{(3)} = L^* z_h.$$

Thus, the error representation formula (5.4) is a sum of three terms: the two classical residual–based terms and a third one due to the stabilisation term in (5.3). This identity is analogous to the error representation formulas for stabilised finite element approximations of first–order hyperbolic problems presented by Houston *et al.* in [61]. In particular, we will see later that, when the RFB method is equivalent to the stabilised finite element method considered by the authors of [61] in the context of first–order hyperbolic PDEs, then the identity (5.4) corresponds to their first error representation formula for functionals.

We now discuss a mesh adaptation algorithm based on the *a posteriori* error representation (5.4).

Given a positive tolerance `TOL`, the goal is the computation of $u_h \in V_h$ such that

$$|J(u) - J(u_h)| \leq \texttt{TOL}. \tag{5.5}$$

From (5.4) we have that

$$|J(u) - J(u_h)| = \left| \sum_{T \in \mathcal{T}_h} \left(\left(\rho_T^{(1)}, \omega_T^{(1)} \right)_T + \left(\rho_T^{(2)}, \omega_T^{(2)} \right)_{\partial T \cap \Omega} + \left(\rho_T^{(3)}, \omega_T^{(3)} \right)_T \right) \right| =: \mathcal{E}_1(u_h). \tag{5.6}$$

Thus, the constraint (5.5) is satisfied as soon as

$$\mathcal{E}_1(u_h) \leq \texttt{TOL}.$$

This will be our *stopping criterion*.

We now need to choose a *refinement criterion*, i.e. a marking strategy for the refinement of the elements in the mesh, and establish a way to actually compute the error representation formula (5.6).

The design of a refinement criterion is based on the localisation of $\mathcal{E}_1(u_h)$. Putting the absolute value sign under the summation sign, we get

$$|J(u) - J(u_h)| \leq \sum_{T\in\mathcal{T}_h} \left| \left(\rho_T^{(1)}, \omega_T^{(1)}\right)_T + \left(\rho_T^{(2)}, \omega_T^{(2)}\right)_{\partial T\cap\Omega} + \left(\rho_T^{(3)}, \omega_T^{(3)}\right)_T \right| =: \sum_{T\in\mathcal{T}_h} \eta_T =: \mathcal{E}_2(u_h),$$

and a decision as to which elements to refine can now be taken depending on the magnitude of the *local error indicator* η_T.

There are many possible refinement criterions; see [11] for a review. An optimal strategy, known as *error per cell strategy*, would be to equilibrate the local error indicator η_T by refining or coarsening according to the criterion

$$\eta_T \approx \frac{\texttt{TOL}}{N_{\text{el}}},$$

where N_{el} is the number of elements in the subdivision.

A criterion that may be more suitable if coarsening is not considered, is the *fixed fraction strategy* in which the elements are ordered according to the size of η_T and then some portion of those with largest η_T is refined. We have chosen to use the fixed fraction strategy proposed by Papastavrou and Verfürth in their article dedicated to the comparison of *a posteriori* error estimators for convection–diffusion problems [79]. The authors of [79] suggest to refine those elements for which

$$\eta_T \geq c_{\text{ref}}\,\bar{\eta},$$

for some user–selected threshold parameter $c_{\text{ref}} \in (0,1)$; the reference value $\bar{\eta}$ is taken to be the maximum of η_T after cutting the upper 10% or 5% of the values (in order to preclude runaway values). That is, a second parameter p_{ref}, usually fixed to 0.1 or 0.05, is defined and $\bar{\eta}$ is obtained as the maximum of η_T after discarding the $\lfloor p_{\text{ref}} N_{\text{el}} \rfloor$ elements with the largest η_T.

Regarding the computation of (5.6), the difficulty is in the evaluation of the dual solution. Here the dual solution z is computed using a new mesh $\mathcal{T}_H$, different from $\mathcal{T}_h$. Given the finite element space V_H corresponding to $\mathcal{T}_H$, and the RFB solution $z_H \in V_H$ $z_H \in V_H$ of the dual problem (5.2), the approximation z_h is taken as the projection or the interpolant of the computed z_H from the primal finite element space.

We reconsider the error representation formula (5.6). By decomposing

$$\omega_T^{(1)} = \omega_T^{(2)} = (z - z_h)_{|T} = (z_H - z_h)_{|T} + (z - z_H)_{|T} = \widetilde{\omega}_T + \bar{\omega}_T,$$

we get

$$\begin{aligned} |J(u)-J(u_h)| &\leq \left| \sum_{T\in\mathcal{T}_h} \left(\left(\rho_T^{(1)}, \widetilde{\omega}_T\right)_T + \left(\rho_T^{(2)}, \widetilde{\omega}_T\right)_{\partial T\cap\Omega} + \left(\rho_T^{(3)}, \omega_T^{(3)}\right)_T \right) \right| \\ &+ \left| \sum_{T\in\mathcal{T}_h} \left(\left(\rho_T^{(1)}, \bar{\omega}_T\right)_T + \left(\rho_T^{(2)}, \bar{\omega}_T\right)_{\partial T\cap\Omega} \right) \right| \\ &= \widetilde{\mathcal{E}}_1(u_h) + \bar{\mathcal{E}}_1(u_h). \end{aligned} \tag{5.7}$$

In this way we have isolated in $\bar{\mathcal{E}}_1(u_h)$ the uncomputable terms of the error bound.

5.1.1 Adaptive algorithm

Let us assume for a moment that $\bar{\mathcal{E}}_1(u_h)$, i.e. the term depending on the difference $z - z_H$, is such that $\bar{\mathcal{E}}_1(u_h) \ll \widetilde{\mathcal{E}}_1(u_h)$. The validity of this hypothesis will be discussed in Section 5.2 below. We then define the new local refinement indicator

$$\widetilde{\eta}_T = \left| (\rho_T^1, \widetilde{\omega}_T)_T + (\rho_T^2, \widetilde{\omega}_T)_{\partial T\cap\Omega} + (\rho_T^3, \omega_T^3)_T \right| \qquad \forall T \in \mathcal{T}_h, \tag{5.8}$$

and redefine $\bar{\eta}$ accordingly. Further, we define the new stopping criterion

$$\widetilde{\mathcal{E}}_1(u_h) \leq \frac{\texttt{TOL}}{C}, \tag{5.9}$$

for some constant $1 < C < 2$.

We may then consider the following adaptive algorithm:

1. Define an initial mesh;
2. Calculate u_h and the dual solutions z_H and z_h on the current meshes;
3. Check the stopping criterion: `IF` $\widetilde{\mathcal{E}}_1(u_h) \leq$ `TOL`$/C$ then `STOP`;
4. Apply the refinement criterion: refine those elements T whose local error indicator η_T exceeds $c_{\text{ref}}\bar{\eta}$ and `GOTO` 2.

5.1.2 Approximation of the dual solution

To ensure that the stopping criterion (5.9) is reliable, i.e. that the approximation error is below the given tolerance, we need to control the size of $\bar{\mathcal{E}}_1(u_h)$. To this end, we observe that the global–residual

$$R(u_h) : v \longrightarrow (f, v) - \mathcal{L}(u_h, v),$$

is a bounded linear functional in V. Hence, returning to (5.4), we see that we can write the term $\bar{\mathcal{E}}_1(u_h)$ as

$$\begin{aligned} \mathcal{L}(u-u_h, z-z_H) &= \mathcal{L}(u, z-z_H) - \mathcal{L}(u_h, z-z_H) \\ &= (f, z-z_H)_\Omega - \mathcal{L}(u_h, z-z_H) \\ &= \langle R(u_h), z-z_H \rangle, \end{aligned} \tag{5.10}$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between V and its dual space V'. We now notice that the right–hand side in (5.10) defines a new linear functional

$$N(v) = \langle R(u_h), v \rangle.$$

Thus, we can estimate the error terms in $\bar{\mathcal{E}}_1(u_h)$ by performing an *a posteriori* analysis of the error $N(z) - N(z_H)$; for this purpose we consider the dual of the dual problem

$$\begin{cases} \text{find } t \in V \text{ such that} \\ \mathcal{L}(t, v) = N(v) \qquad \forall v \in V. \end{cases} \tag{5.11}$$

Let $t_H \in V_H$ be some approximation of t. We have

$$\begin{aligned} N(z) - N(z_H) &= N(z - z_H) \\ &= \mathcal{L}(t, z - z_H) \\ &= \mathcal{L}(t - t_H, z - z_H) + \sum_{K \in \mathcal{T}_H} (Lt_H, z_b)_K \\ &= \sum_{K \in \mathcal{T}_H} \Big((t - t_H, R_K(z_H))_K - \frac{1}{2} \varepsilon \, (t - t_H, \boldsymbol{n}_K \cdot [\nabla z_H])_{\partial K \cap \Omega} + (Lt_H, z_b)_K \Big) \end{aligned}$$

where, as before, the elemental–residual is defined as $R_K(z_H) = (g - L^* z_H)_{|K}$. Finally, z_b represents the bubble part of the RFB solution to the dual problem.

The new error representation just obtained is in terms of $t - t_H$ which is just as uncomputable as $z - z_H$. To avoid a possibly infinite sequence of duality arguments, we bound $t - t_H$ in terms of a stability constant. The bound obtained in this way need not be sharper then the one we would obtain if instead we were to bound $z - z_H$ directly as was done by Eriksson *et al.* in [41] and [42]. From the practical point of view, though, the crudeness of the bound of $\bar{\mathcal{E}}_1(u_h)$ is not of particular concern since all we need $\bar{\mathcal{E}}_1(u_h)$ for is to generate an adequate sequence of finite element approximations z_H which we can use to compute $\widetilde{\mathcal{E}}_1(u_h)$.

From now on, we assume that linear finite elements are used, so

$$V_H = \left\{ \varphi \in C(\bar{\Omega}) : \varphi|_K \in \mathcal{P}_1 \quad \forall K \in \mathcal{T}_H \right\},$$

and assume that for any triangulation $\mathcal{T}_H$ and any element $T \in \mathcal{T}_H$ the number of neighbours of T is bounded.

We choose $t_H = P_H t$, where P_H is the modification of the quasi–interpolation operator of Clémant [35] analysed by Verfürth [95]. With this choice we can take advantage of explicit interpolation error bounds in terms of the H^1–seminorm.

The definition of the modified Clémant's quasi–interpolant requires the introduction of the following notational conventions. Let $\mathcal{E}_H$ and $\mathcal{N}_H$ be the sets of all the edges and all the vertices in $\mathcal{T}_H$, respectively. Further, let

$$\mathcal{E}_H = \mathcal{E}_{H,\Omega} \cup \mathcal{E}_{H,D}, \qquad \mathcal{N}_H = \mathcal{N}_{H,\Omega} \cup \mathcal{N}_{H,D},$$

be the decompositions of $\mathcal{E}_H$ and $\mathcal{N}_H$ into the subsets of internal and boundary edges and vertices, respectively. For any $S \in \mathcal{T}_H \cup \mathcal{E}_H$, let $\mathcal{N}(S)$ be the set of its vertices; and for any $K \in \mathcal{T}_H$, let $\mathcal{E}(K)$ be the set of all faces of K which are internal to Ω. Finally, for any vertex $x \in \mathcal{N}_H$, denote by ω_x the union of all triangles which have x as a vertex.

The following definition and bounds can be extended to $H^1_{\Gamma_D}(\Omega)$ for any $\Gamma_D \subset \partial\Omega$, essentially by interpreting the nodes sitting on the Neumann boundary $\partial\Omega \setminus \Gamma_D$ as internal.

Given a function $u \in L_2(\omega_x)$ we associate to any $x \in \mathcal{N}_H$ the value

$$\pi_x u = \frac{1}{|\omega_x|} \int_{\omega_x} u.$$

The quasi–interpolation operator $P_H : V \to V_H$ is defined as follows:

$$P_H u = \sum_{x \in \mathcal{N}_{H,\Omega}} (\pi_x u)\, \varphi_x,$$

where $\varphi_x \in V_H$ is the finite element basis function associated with x. Notice that P_H is *not* a projection operator.

Lemma 5.1.1 *For all $v \in V$, all $K \in \mathcal{T}_H$ and all $E \in \mathcal{E}_H$ we have*

$$\|v - P_H v\|_{0,K} \le \sum_{x \in \mathcal{N}(K)} C_{K;x} H_x |v|_{1,\omega_x}, \tag{5.12}$$

$$\|v - P_H v\|_{0,E} \le \sum_{x \in \mathcal{N}(E)} C_{E;x} H_x^{1/2} |v|_{1,\omega_x}, \tag{5.13}$$

where H_x is the maximum length of an edge having x as an end point. The values of the constants $C_{K;x}$ and $C_{E;x}$ are given explicitly by Verfürth [95] in terms of the following mesh–quality related quantities

$$\kappa_{3,x} := \max_{\substack{K_1, K_2 \in \mathcal{T}_H \\ x \in \mathcal{N}(K_1) \cap \mathcal{N}(K_2)}} \frac{|K_1|}{|K_2|}, \qquad \kappa_{4,x} := \max_{\substack{K \in \mathcal{T}_H, E \in \mathcal{E}_H \\ x \in \mathcal{N}(K) \cap \mathcal{N}(E)}} \frac{|E| H_K}{|K|}.$$

Moreover, the following bound holds on the interpolation error in the H^1–seminorm:

$$|v - P_H v|_{1,K} \le \sum_{x \in \mathcal{N}(K)} C'_{K;x} |v|_{1,\omega_x}, \tag{5.14}$$

with the constant $C'_{K;x}$ depending explicitly on $\kappa_{3,x}$ and on the shape–regularity of K, i.e. the ratio $\kappa_K = H_K/\rho_K$ between the diameter of K and the diameter of the largest circle inscribed into K.

Proof. The bounds (5.12) and (5.13) are due to Verfürth, see [95]. We prove (5.14) by adapting Verfürth's proof of (5.12). To start with, (5.14) is to be shown for an operator which does not need to satisfy the Dirichlet boundary conditions. That is, a new operator $\bar{P}_H$, a modification of P_H, is defined as

$$\bar{P}_H u = \sum_{x \in \mathcal{N}_H} (\pi_x u)\varphi_x.$$

Fix an arbitrary $K \in \mathcal{K}_H$. Since $\sum_{x \in \mathcal{N}(K)} \varphi_x = 1$ on K, we have

$$\begin{aligned} \|\nabla(u - \bar{P}_H u)\|_{0,K} &= \left\| \sum_{x \in \mathcal{N}(K)} ((u - \pi_x u)\nabla\varphi_x + (\nabla u)\varphi_x) \right\|_{0,K} \\ &\leq \sum_{x \in \mathcal{N}(K)} \|\nabla\varphi_x\|_{\infty,K} \|u - \pi_x u\|_{0,K} + \sum_{x \in \mathcal{N}(K)} \|\varphi_x\|_{\infty,K} \|\nabla u\|_{0,K} \\ &\leq \sum_{x \in \mathcal{N}(K)} \frac{\kappa_K}{H_K} \|u - \pi_x u\|_{0,K} + \sum_{x \in \mathcal{N}(K)} \|\nabla u\|_{0,K}. \end{aligned} \tag{5.15}$$

We have now reached the point at which this part of the proof follows that of (5.12) given in [95].

The following result is Lemma 4.3 in [95]: for any $x \in \mathcal{N}_H$ and $u \in H^1(\omega_x)$ we have

$$\|u - \pi_x u\|_{0,\omega_x} \leq C_x H_x \|\nabla u\|_{0,\omega_x}, \tag{5.16}$$

where C_x is an explicit constant depending on the ratio

$$\kappa_{1,x} = \frac{H_x}{\rho_x},$$

and ρ_x is the minimum length of an edge having x as an end point (in particular, if ω_x is convex, then $C_x = 2/\pi$).

Applying (5.16) in (5.15) we obtain the desired result for $\bar{P}_H$.

The proof is completed by bounding the norm of the difference between $\nabla P_H u$ and $\nabla \bar{P}_H u$:

$$\begin{aligned} \|\nabla(P_H u - \bar{P}_H u)\|_{0,K} &= \left\| \sum_{x \in \mathcal{N}(K) \cap \mathcal{N}_{H,D}} \pi_x u \, \nabla\varphi_x \right\|_{0,K} \\ &\leq \sum_{x \in \mathcal{N}(K) \cap \mathcal{N}_{H,D}} |\pi_x u| \, \|\nabla\varphi_x\|_{0,K} \\ &\leq \sum_{x \in \mathcal{N}(K) \cap \mathcal{N}_{H,D}} \frac{\sqrt{12}\kappa_K}{H_K} |\pi_x u| \, \|\varphi_x\|_{0,K}. \end{aligned}$$

We are again within the framework of the proof of (5.12) given in [95]. This time, the proof is completed by using the following result (equation (5.6) in [95]):

$$|\pi_x u|\,\|\varphi_x\|_{0,K} \leq \frac{1}{2}\left(\frac{|K|}{|K_x|}\right)^{1/2}\left(\|u-\pi_x u\|_{0,K_x}+H_x\|\nabla u\|_{0,K_x}\right),$$

(here K_x represents any triangle K with the following properties: x is a vertex of K and K shares an entire edge with $\partial\Omega$), and applying again (5.16). □

We are now ready to obtain a computable bound on $|N(z)-N(z_H)|$. By applying the Cauchy–Schwarz inequality and the interpolation error bounds (5.12), (5.13) and (5.14) after noting that $LP_Ht = \boldsymbol{a}\cdot\nabla P_Ht$, we have

$$\begin{aligned}
|N(z)-N(z_H)| \leq \sum_{K\in\mathcal{T}_H}\Bigg(&\|R_K(z_H)\|_{0,K}\|t-P_Ht\|_{0,K}\\
&+\sum_{E\in\mathcal{E}(K)}\|\frac{1}{2}\varepsilon\,\boldsymbol{n}_K\cdot[\nabla z_H]\|_{0,E}\|t-P_Ht\|_{0,E}+\|z_b\|_{0,K}\|\boldsymbol{a}\cdot\nabla P_Ht\|_{0,K}\Bigg)\\
\leq \sum_{K\in\mathcal{T}_H}\Bigg(&\|R_K(z_H)\|_{0,K}\sum_{x\in\mathcal{N}(K)}C_{K;x}H_x|t|_{1,\omega_x}\\
&+\sum_{E\in\mathcal{E}(K)}\Big(\|\frac{1}{2}\varepsilon\boldsymbol{n}_K\cdot[\nabla z_H]\|_{0,E}\sum_{x\in\mathcal{N}(E)}C_{E;x}H_x^{1/2}|t|_{1,\omega_x}\Big)\\
&+\|\boldsymbol{a}\|_{\infty,K}\|z_b\|_{0,K}\Big(|t|_{1,K}+\sum_{x\in\mathcal{N}(K)}C'_{K;x}|t|_{1,\omega_x}\Big)\Bigg).
\end{aligned}$$

The solution t_H has now been removed from the bound, but t is still present. We can eliminate t as follows, at the expense of breaking up the sum over the elements of the triangulation. Let

$$\begin{aligned}
C_K &= \max_{x\in\mathcal{N}(K)}(C_{K;x}\kappa_{1,x}),\\
C_{\partial K} &= \max_{E\in\mathcal{E}(K)}\max_{x\in\mathcal{N}(E)}(C_{E;x}\kappa_{1,x}^{1/2}),\\
C'_K &= 1+\max_{x\in\mathcal{N}(K)}C'_{K;x}.
\end{aligned}$$

Then, since the number of neighbours of any $T \in \mathcal{T}_H$ is bounded, we get

$$\begin{aligned}
&|N(z)-N(z_H)| \\
&\quad \le \sum_{K\in\mathcal{T}_H}\Bigg(\Big(C_K H_K\|R_K(z_H)\|_{0,K} + 2C_{\partial K}H_K^{1/2}\|\frac{1}{2}\varepsilon\boldsymbol{n}_K\cdot[\nabla z_H]\|_{0,\partial K\cap\Omega}\Big)\sum_{x\in\mathcal{N}(K)}|t|_{1,\omega_x}\Bigg) \\
&\qquad + C_K'\|\boldsymbol{a}\|_{\infty,K}\|z_b\|_{0,K}\sum_{x\in\mathcal{N}(K)}|t|_{1,\omega_x}\Bigg) \\
&\quad \le 3\Bigg(\sum_{K\in\mathcal{T}_H}\Big(C_K H_K\|R_K(z_H)\|_{0,K} + C_{\partial K}H_K^{1/2}\|\varepsilon\boldsymbol{n}_K\cdot[\nabla z_H]\|_{0,\partial K\cap\Omega} \\
&\qquad + C_K'\|\boldsymbol{a}\|_{\infty,K}^2\|z_b\|_{0,K}^2\Big)^2\Bigg)^{1/2}|t|_{1,\Omega}.
\end{aligned}$$

To quantify $|t|_{1,\Omega}$, we reconsider the problem (5.11) in strong form with forcing term $R(u_h) \in H^{-1}(\Omega)$:

$$\begin{cases} Lt = R(u_h) & \text{in } \Omega, \\ t = 0 & \text{on } \partial\Omega. \end{cases} \tag{5.17}$$

Multiplying the first equation in (5.17) by t and integrating by parts over Ω, we get

$$\varepsilon|t|_{1,\Omega}^2 = \langle R(t_h), t\rangle \le |t|_{1,\Omega}\sup_{v\in V\setminus\{0\}}\frac{\langle R(u_h), v\rangle}{|v|_{1,\Omega}}.$$

Hence, we have the stability estimate

$$|t|_{1,\Omega} \le \varepsilon^{-1}\|R(u_h)\|_{-1,\Omega},$$

and we conclude that

$$\begin{aligned}
&|N(z) - N(z_H)| \\
&\quad \le 3\Bigg(\sum_{K\in\mathcal{T}_H}\Big(C_K H_K\|\varepsilon^{-1}R_K(z_H)\|_{0,K} + C_{\partial K}H_K^{1/2}\|\boldsymbol{n}_K\cdot[\nabla z_H]\|_{0,\partial K\cap\Omega} \\
&\qquad + \varepsilon^{-1}C_K'\|\boldsymbol{a}\|_{\infty,K}^2\|z_b\|_{0,K}^2\Big)^2\Bigg)^{1/2}\|R(u_h)\|_{-1,\Omega} \\
&\quad := \bar{\mathcal{E}}_2(u_h).
\end{aligned} \tag{5.18}$$

In this way, we have obtained the following computable error bound:

$$|J(u) - J(u_h)| \le \widetilde{\mathcal{E}}_1(u_h) + \bar{\mathcal{E}}_2(u_h) \tag{5.19}$$

Two remarks are in order. First, the number $\|R(u_h)\|_{-1,\Omega}$ is not directly computable from u_h. We can still compute it, but at the expense of the solution of an auxiliary problem; see the next section for details. Finally, the derivation of sharp and computable *a posteriori* error bounds is a nontrivial task. Our bound is no exception in this respect: by bounding $z - z_H$ we achieved computability and reliability of the bound, but we anticipate a loss in terms of sharpness.

In the next section, we investigate the sharpness of the error bound (5.19), and study the reliability and effectivity of the alternative error estimator $\bar{\mathcal{E}}_1(u_h)$.

5.2 Implementation

We experiment with the *a posteriori* error bound described in the previous section by considering linear finite elements on triangles for both the primal and dual computations. We have chosen to define $\mathcal{T}_H$ as the triangulation obtained by subdividing every triangle in $\mathcal{T}_h$ into the four triangles obtained by bisection of its edges. A survey of other possible choices can be found, for example, in [11]; see also the comments in the review article by Giles and Süli [55].

Let $V_h \subset V = H_0^1(\Omega)$ denote the usual space of linear finite elements

$$V_h = \left\{\varphi \in C(\overline{\Omega}) \,:\, \varphi_{|T} \in \mathcal{P}_1 \quad \forall T \in \mathcal{T}_h\right\},$$

and similarly for V_H.

To implement the RFB formulation (5.3) over V_h we need to calculate the bubble part of the solution u_b. As we have explained in Section 2.1, this is given on every element $T \in \mathcal{T}_h$ as the solution in $H_0^1(T)$ of the *bubble equation*

$$L_T u_b|_T = (f - Lu_h)|_T,$$

where $L_T : H_0^1(T) \to H^{-1}(T)$ denotes the restriction of the operator L to T.

In what follows we assume that, in the evaluation of the bubble term of the RFB formulation (5.3), the velocity field $\boldsymbol{a}$ and the forcing term f can be treated as piecewise constant functions on $\mathcal{T}_h$: these assumptions ensure that the RFB method is equivalent to SDFEM; see Section 2.4.2. In this case, the bubble part of the solution of the RFB formulation is given by

$$u_{b|T} = (f - \boldsymbol{a}\cdot\nabla u_h)_{|T} b_T,$$

where b_T is the solution of the local problem

$$\begin{cases} -\varepsilon\Delta b_T + \boldsymbol{a}\cdot\nabla b_T = 1 & \text{in } T, \\ b_T = 0 & \text{on } \partial T. \end{cases}$$

Thus, for any $T \in \mathcal{T}_h$, as we proved in Section 2.4.2, the static condensation of the bubble results in the classical streamline–diffusion stabilisation term, since

$$(u_b, L^* v_h)_T = \tau_T(\boldsymbol{a}\cdot\nabla u_h - f, \boldsymbol{a}\cdot\nabla v_h)_T,$$

where the SD–parameter τ_T is defined via the bubble b_T as

$$\tau_T = \frac{1}{|T|}\int_T b_T \, d\boldsymbol{x}.$$

Moreover, as in Section 2.6, assuming that $\varepsilon \ll |\boldsymbol{a}|$, i.e. that the equation is convection–dominated, we can approximate the integral average of the bubble b_T by integrating, instead, the solution $\widetilde{b}_T$ of the reduced problem:

$$\begin{cases} \boldsymbol{a}\cdot\nabla\widetilde{b}_T = 1 & \text{in } T, \\ \widetilde{b}_T = 0 & \text{on } \partial T_-, \end{cases}$$

where ∂T_- denotes the inflow boundary of T. Further, let us consider the approximation

$$\tau_T \approx \widetilde{\tau}_T := \frac{1}{|T|}\int_T \widetilde{b}_T \, d\boldsymbol{x} = \frac{h_{\boldsymbol{a}}}{3|\boldsymbol{a}|}, \tag{5.20}$$

where $h_{\boldsymbol{a}}$ is the length of the longest segment contained in T in the direction of $\boldsymbol{a}$.

The third term in the *a posteriori* error estimator $\widetilde{\mathcal{E}}_1(u_h)$ defined in (5.7) can be approximated in exactly the same way, since

$$\begin{aligned} (\rho_T^3, \omega_T^3)_T &= (u_b, L^* z_h)_T = \tau_T (f - \boldsymbol{a}\cdot\nabla u_h, -\boldsymbol{a}\cdot\nabla z_h)_T \\ &\approx \widetilde{\tau}_T |T| (\boldsymbol{a}\cdot\nabla u_h - f)_{|T} (\boldsymbol{a}\cdot\nabla z_h)_{|T}. \end{aligned}$$

We notice that the first line above coincides with the third term in the first error estimator for functionals defined in [61]. That is, in the special case in which the RFB method coincides with the streamline–diffusion method, our estimator $\widetilde{\mathcal{E}}_1(u_h)$ is identical to that for the streamline–diffusion method.

Regarding the computation of $\|R(u_h)\|_{-1,\Omega}$ in $\bar{\mathcal{E}}_2(u_h)$ (see (5.18)), we may proceed as follows. We define the auxiliary problem

$$\begin{cases} -\Delta\phi = R(u_h) & \text{in } \Omega, \\ \phi = 0 & \text{on } \partial\Omega. \end{cases} \tag{5.21}$$

By definition of the norm of the dual space V', we have

$$\begin{aligned} \|R(u_h)\|_{-1,\Omega} &= \sup_{\psi\in V} \frac{(R(u_h),\psi)}{\|\nabla\psi\|_{0,\Omega}} \\ &= \sup_{\psi\in V} \frac{(-\Delta\phi,\psi)}{\|\nabla\psi\|_{0,\Omega}} \\ &= \sup_{\psi\in V} \frac{(\nabla\phi,\nabla\psi)}{\|\nabla\psi\|_{0,\Omega}} \\ &= \|\nabla\phi\|_{0,\Omega}. \end{aligned}$$

Thus, we have reduced the problem of the compuation of $\|R(u_h)\|_{-1,\Omega}$ to that of the computation of $\|\nabla\phi\|_{0,\Omega}$, where ϕ is the solution of (5.21). This will be done approximately, by

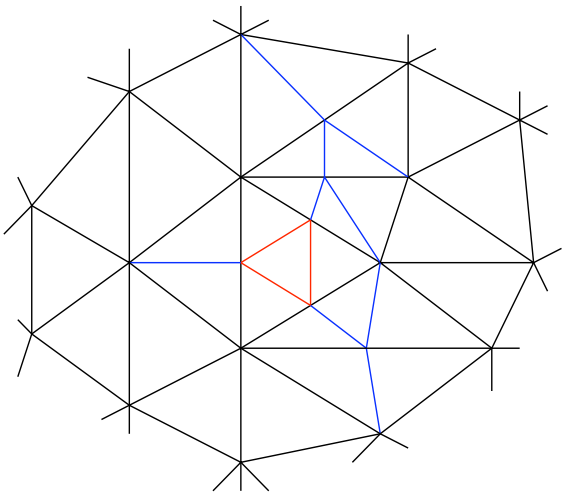


Figure 5.1: Example of red refinement (in red) and subsequent refinement of the neighbours (in blue). The rule for blue refinement is that the longest edge is always bisected.

considering (5.21) in weak form and solving it by a finite element method on the triangulation $\mathcal{T}_h$. That is, we will solve the problem

$$\begin{cases} \text{find } \phi_h \in V_h \text{ such that} \\ (\nabla\phi_h, \nabla v_h) = \langle R(u_h), v_h\rangle = (f, v_h) - \mathcal{L}(u_h, v_h) \qquad \forall v_h \in V_h, \end{cases} \tag{5.22}$$

and compute $\|\nabla\phi_h\|_{0,\Omega}$.

At every iteration of the adaptive algorithm we need to refine the triangulation, basing the refinement on the marking strategy described in the previous section. In our numerical computations we utilise the MATLAB pde–toolbox refinement routine `refinemesh`, which performs *red refinement* on the marked elements. That is, every element marked for refinement is subdivided into four triangles (sons) by connecting the mid–points of the edges. To avoid the creation of hanging–nodes, and hence to comply with the constraint of admissibility of the triangulations given before, the neighbours of the marked elements are subjected to blue refinement following the longest edge bisection scheme [83], as shown in Figure 5.1. This technique is not ideal, since it involves refinement of distant neighbours, but it has the advantage of ensuring the shape–regularity of the triangulation.

Example 1. We consider the boundary value problem

$$\begin{cases} -\varepsilon\Delta u + u_x + u_y = f & \text{in } \Omega = (0,1)^2, \\ u = 0 & \text{in } \partial\Omega, \end{cases} \tag{5.23}$$

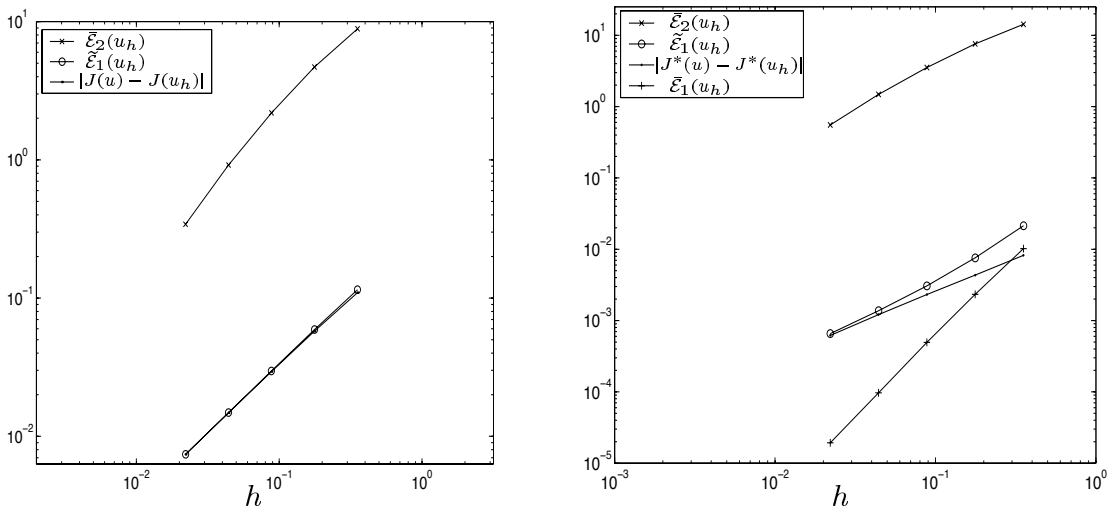


Figure 5.2: Example 1. The error and the *a posteriori* error bounds under successive uniform refinements with respect to the functional $J(\cdot)$ (a) and $J^*(\cdot)$ (b), with $\varepsilon = 10^{-2}$.

with f defined in such a way that the exact solution is given by

$$u(x,y) = 2\sin(x)\, y^2(1 - \mathrm{e}^{-(1-x)/\varepsilon})(1 - \mathrm{e}^{-(1-y)/\varepsilon}).$$

Our aim is the computation of the mean–flow over Ω, i.e.

$$J(u) = \int_\Omega u \, d\boldsymbol{x}, \tag{5.24}$$

with the aim to ensure that the error does not exceed a given tolerance `TOL`.

To start with, we compare the true error $|J(u) - J(u_h)|$ and the *a posteriori* error bound (5.19) on successively refined uniform meshes. We recall that the r.h.s. of (5.19) consists of two terms: $\widetilde{\mathcal{E}}_1(u_h)$ which is related to the difference $z_H - z_h$, and $\bar{\mathcal{E}}_2(u_h)$ which is an upper bound for $\bar{\mathcal{E}}_1(u_h)$ which refers to $z - z_H$.

The results for $\varepsilon = 10^{-2}$ are shown in Figure 5.2(a): the log–log plot makes it apparent that $\bar{\mathcal{E}}_2(u_h)$ is over–estimating the true error. On the other hand, the term $\widetilde{\mathcal{E}}_1(u_h)$ alone agrees remarkably well with the error, see also the effectivity indices reported in Table 5.1.

In order to conclude that $\widetilde{\mathcal{E}}_1(u_h)$ can be used in an adaptive algorithm as an *a posteriori* error bound, we still need to ascertain its reliability, which depends on whether or not $\bar{\mathcal{E}}_1(u_h)$, i.e. the omitted term in (5.7), is of higher order.

We can compare $\widetilde{\mathcal{E}}_1(u_h)$ and $\bar{\mathcal{E}}_1(u_h)$ for a slightly different problem, in which the target linear functional is chosen so that the dual solution z is known. This is achieved by performing

the change of variables $(x, y) \to (1-x, 1-y)$ in the primal problem. Thus we define

$$z(x,y) = 2\sin(1-x)(1-y)^2(1-\mathrm{e}^{-x/\varepsilon})(1-\mathrm{e}^{-y/\varepsilon}),$$

and fix the new target functional J^* consequently. The results obtained for this new problem are shown in Figure 5.2(b). The bound $\bar{\mathcal{E}}_2(u_h)$ is still over–estimating the error in approximately the same way as before, while we observe that (except on the coarsest grid) the sharper bound $\bar{\mathcal{E}}_1(u_h)$ is indeed of higher order than the true error.

We conclude that the adaptive algorithm described in Section 5.1.1, which uses $\widetilde{\mathcal{E}}_1(u_h)$ as error estimator, is, for this problem at least, reliable and efficient.

We can further investigate the relative magnitudes of the terms in the *a posteriori* error bound by considering separately the three terms comprising $\widetilde{\mathcal{E}}_1(u_h)$. Define

$$\begin{aligned} D_{\text{res}} &= \Big| \sum_{T\in\mathcal{T}_h} (R_T(u_h), z_H - z_h)_T \Big|, \\ D_{\text{jump}} &= \Big| \frac{1}{2} \sum_{T\in\mathcal{T}_h} (\varepsilon\, \boldsymbol{n}\cdot[\nabla u_h], z_H - z_h)_T \Big|, \\ D_{\text{bubl}} &= \Big| \sum_{T\in\mathcal{T}_h} (u_b, L^* z_h)_T \Big|. \end{aligned}$$

The behaviour of these three terms can be appreciated from the numbers in Table 5.1 and the

h	$\lvert J(u)-J(u_h)\rvert$	$\widetilde{\mathcal{E}}_1(u_h)$	*eff*	D_{bubl}	D_{res}	D_{jump}	$\bar{\mathcal{E}}_2(u_h)$
1/4	1.09×10^{-1}	1.15×10^{-1}	1.05	1.08×10^{-1}	7.5×10^{-3}	5.8×10^{-4}	8.88
1/8	5.76×10^{-2}	5.91×10^{-2}	1.025	5.8×10^{-2}	1.4×10^{-3}	2.5×10^{-4}	4.69
1/16	2.93×10^{-2}	2.97×10^{-2}	1.01	2.95×10^{-2}	3.1×10^{-4}	9.0×10^{-5}	2.187
1/32	1.47×10^{-2}	1.48×10^{-2}	1.003	1.48×10^{-2}	7.7×10^{-5}	2.6×10^{-5}	0.917
1/64	7.39×10^{-3}	7.4×10^{-3}	1.001	7.39×10^{-3}	1.9×10^{-5}	6.3×10^{-6}	0.342

Table 5.1: Example 1. Convergence of $|J(u)-J(u_h)|$ and $\widetilde{\mathcal{E}}_1(u_h)$ and its components, with $\varepsilon = 10^{-2}$.

graphs in Figure 5.3. We observe that D_{res} and D_{jump} are of higher order and the true error is well approximated by the term D_{bubl} alone. Given that D_{bubl} is computable (from u_b and z_h), this fact suggests to use D_{bubl} as a correction term by moving it across to the left–hand side of the error representation formula (this viewpoint is discussed in Giles & Süli [55]). In other words, the quantity

$$J_{\text{cor}}(u_h) = J(u_h) - \sum_{T\in\mathcal{T}_h} (u_b, L^* z_h)_T\,, \tag{5.25}$$

should give a better approximation to $J(u)$ than $J(u_h)$. The error representation formula now becomes:

$$J(u) - J_{\text{cor}} = \mathcal{L}(u-u_h, z-z_h).$$

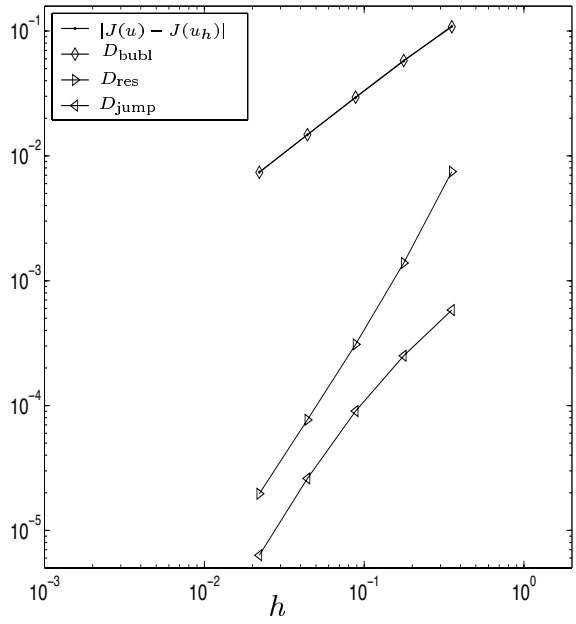


Figure 5.3: Example 1. The error and the three terms of the *a posteriori* error estimator $\widetilde{\mathcal{E}}_1$ under successive uniform refinements with respect to the functional (5.24); $\varepsilon = 10^{-2}$.

For this we can rewrite the *a posteriori* error bound in the following form:

$$|J(u) - J_{\text{cor}}| \leq D_{\text{res}} + D_{\text{jump}} + \bar{\mathcal{E}}_2(u_h).$$

We know already that this error bound is not sharp, due to the third term on the right–hand side. But, this time the term $\bar{\mathcal{E}}_1(u_h)$ is no longer negligible in comparison with $D_{\text{res}} + D_{\text{jump}}$ alone. Indeed the quantity $D_{\text{res}} + D_{\text{jump}}$ under–estimates the true error, and so it cannot be used as an *a posteriori* error bound in an adaptive algorithm.

To show this, we have run the adaptive algorithm described in the previous section, using the following values of the parameters in the adaptive algorithm:

$$c_{\text{ref}} = .5, \qquad p_{\text{ref}} = .1, \qquad \texttt{TOL} = 10^{-3}. \tag{5.26}$$

The results are shown in Table 5.2, for $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-6}$. We observe that:

1. The error estimator $\widetilde{\mathcal{E}}_1(u_h)$ is very effective in predicting the error $J(u) - J(u_h)$, robustly with respect to ε;

2. The corrected quantity (5.25) gives, as expected, a considerably better approximation of the target quantity $J(u)$;

3. $D_{\text{res}} + D_{\text{jump}}$ under–estimates the true error, as indicated by the related effectivity index

$$\mathit{eff}_{\text{cor}} = \frac{D_{\text{res}} + D_{\text{jump}}}{|J(u) - J_{\text{cor}}(u_h)|}.$$

We conclude that the quantity (5.25) is best used as a more accurate approximation of $J(u)$ after the last step in an adaptive algorithm based on the error estimator $\widetilde{\mathcal{E}}_1(u_h)$.

$N_{\rm el}$	$\lvert J(u)-J(u_h)\rvert$	$\widetilde{\mathcal{E}}_1(u_h)$	*eff*	$\lvert J(u)-J_{\rm cor}(u_h)\rvert$	$D_{\rm res}+D_{\rm jump}$	$eff_{\rm cor}$
32	1.095×10^{-1}	1.15×10^{-1}	1.05	1.1388×10^{-3}	8.05×10^{-3}	0.43
69	5.667×10^{-2}	5.93×10^{-2}	1.046	2.192×10^{-3}	1.62×10^{-3}	0.27
155	2.736×10^{-2}	2.935×10^{-2}	1.073	3.038×10^{-3}	1.05×10^{-3}	0.19
337	1.332×10^{-2}	1.499×10^{-2}	1.125	3.07×10^{-3}	1.39×10^{-3}	0.18
779	6.8×10^{-3}	7.45×10^{-3}	1.097	1.071×10^{-3}	4.09×10^{-4}	0.15
1680	3.543×10^{-3}	3.87×10^{-3}	1.092	6.017×10^{-4}	2.74×10^{-4}	0.14
4644	1.995×10^{-3}	2.144×10^{-3}	1.075	2.246×10^{-4}	7.55×10^{-5}	0.06

$N_{\rm el}$	$\lvert J(u)-J(u_h)\rvert$	$\widetilde{\mathcal{E}}_1(u_h)$	*eff*	$\lvert J(u)-J_{\rm cor}(u_h)\rvert$	$D_{\rm res}+D_{\rm jump}$	$eff_{\rm cor}$
32	1.119×10^{-1}	1.194×10^{-1}	1.067	2.44×10^{-3}	5.12×10^{-3}	2.1
69	5.825×10^{-2}	6.2×10^{-2}	1.064	2.84×10^{-3}	9.07×10^{-4}	0.32
166	2.835×10^{-2}	3.11×10^{-2}	1.099	2.94×10^{-3}	1.32×10^{-4}	0.045
345	1.376×10^{-2}	1.64×10^{-2}	1.192	2.64×10^{-3}	8.1×10^{-6}	0.003
763	6.99×10^{-3}	8.36×10^{-3}	1.193	1.3×10^{-3}	4.69×10^{-5}	0.036
1595	3.28×10^{-3}	4.1×10^{-3}	1.25	8.41×10^{-4}	1.93×10^{-5}	0.023
3296	1.648×10^{-3}	2.03×10^{-3}	1.234	3.98×10^{-4}	1.28×10^{-5}	0.032
6552	7.514×10^{-4}	1.02×10^{-3}	1.357	2.92×10^{-4}	2.35×10^{-5}	0.08

Table 5.2: Example 1. The error under successive refinements with respect to the functional $J(u)=\int_\Omega u\,d\boldsymbol{x}$; $\varepsilon=10^{-2}$ (above) and $\varepsilon=10^{-6}$ (below).

Repeating the same experiment for the modified target functional J^* we have observed that the terms in $\bar{\mathcal{E}}_1(u_h)$ are quantitatively comparable to $D_{\rm res}$ and $D_{\rm jump}$, and hence they cannot be neglected.

Example 2. We consider the boundary value problem

$$\begin{cases} -\varepsilon\Delta u + u_x + u_y = f & \text{in } \Omega=(0,1)^2,\\ u=0 & \text{on } \partial\Omega, \end{cases} \tag{5.27}$$

with f defined in such a way that the exact solution is given by

$$u(x,y)=2xy(1-\mathrm{e}^{-(1-x)/\varepsilon})(1-\mathrm{e}^{-(1-y)/\varepsilon}).$$

Notice that the function u is symmetric with respect to the line $x=y$.

The objective is the computation of the solution at a given point $P=P(x_0,y_0)$ with the aim to ensure that the error between $u(x_0,y_0)$ and $u_h(x_0,y_0)$ does not exceed a given tolerance

$N_{\rm el}$	$\lvert J(u)-J(u_h)\rvert$	$\widetilde{\mathcal{E}}_1(u_h)$	*eff*
32	2.12×10^{-2}	9.17×10^{-2}	4.3
64	2.5×10^{-3}	8.7×10^{-3}	3.5
146	1.7×10^{-3}	2.61×10^{-3}	1.5
290	8.78×10^{-4}	1.24×10^{-3}	1.4
616	2.1×10^{-4}	4.06×10^{-4}	1.9
1396	7.43×10^{-5}	1.24×10^{-4}	1.6

$N_{\rm el}$	$\lvert J(u)-J(u_h)\rvert$	$\widetilde{\mathcal{E}}_1(u_h)$	*eff*
32	2.58×10^{-2}	1.28×10^{-1}	4.98
82	1.02×10^{-3}	1.13×10^{-3}	1.1
188	3.77×10^{-3}	4.61×10^{-3}	1.2
378	2.14×10^{-4}	6.52×10^{-3}	3.04
708	5.08×10^{-4}	2.09×10^{-3}	4.1
1262	2.0×10^{-4}	7.04×10^{-4}	3.5
2176	1.19×10^{-4}	3.2×10^{-4}	2.7
3858	9.39×10^{-5}	2.82×10^{-4}	3

Table 5.3: Example 2. The error under successive refinements with respect to the functional $J(u)=u((.49,.49))$; $\varepsilon=10^{-2}$ (left) and $\varepsilon=10^{-6}$ (right).

TOL. We apply the algorithm described in Section 5.1.1 using the same parameter values as in (5.26).

The successive meshes produced by the algorithm to calculate the solution at $P=(.49,.49)$ with $\varepsilon=10^{-2}$ and $\varepsilon=10^{-6}$ are depicted in Figure 5.6 and Figure 5.7, respectively.

The meshes respect the symmetry of the problem. Moreover, we notice that initially the mesh is refined down–wind of P: the algorithm recognises that some resolution of the boundary–layers is necessary in order to ensure any accuracy at the point of interest. In subsequent refinement the boundary layer zone is left unchanged, the refinements being concentrated upwind of the point, along the subcharacteristic curve passing through P.

The effectivity of the *a posteriori* error estimator $\widetilde{\mathcal{E}}_1(u_h)$ is reported in Table 5.3: the estimator is robust with respect to the diffusion parameter.

Example 3. We solve the boundary value problem with discontinuous boundary conditions

$$\begin{cases} -\varepsilon\Delta u+(\cos(\pi/3),\sin(\pi/3))^T\cdot\nabla u=1 & \text{in }\Omega=(0,1)^2,\\ u=1 & \text{for } \begin{cases} x\le 1/2, \quad y=0,\\ x=0,\end{cases}\\ u=0 & \text{otherwise.}\end{cases} \tag{5.28}$$

The solution of this problem has an internal layer propagating across Ω from the discontinuity in the boundary condition at $(0.5,0)\in\partial\Omega$. As for Example 1, we defined the target functional to be the mean–flow in the entire domain Ω; that is $J(u)=\int_\Omega u\,d\boldsymbol{x}$ and fixed $\varepsilon=10^{-6}$. The algorithm is again the one described in Section 5.1.1, which employs the computable error estimator $\widetilde{\mathcal{E}}_1(u_h)$. The values for the parameters in the adaptive algorithm are:

$$c_{\rm ref}=.5,\qquad p_{\rm ref}=.05,\qquad \texttt{TOL}=10^{-3}.$$

As we can see in Figure 5.8, first the mesh gets refined in the boundary layer, which is the major source of error. Only when the boundary layer has been partially resolved, is the mesh

refined along the internal layer, showing that the indicator correctly identifies locations in the mesh that most affect the accuracy of the approximation of the functional. The final mesh consists of 7405 triangles, approximately 5 out of 7 of which are located in the proximity of the boundary layer given by $y > 7/8$.

Example 4. We consider the mixed boundary–value problem for the convection–diffusion equation specified in Figure 5.4 (top left). Homogeneous Neumann boundary condition is imposed on $\Gamma_N = \{(x, y) \in \Gamma : y = 0 \text{ or } x = 1\}$, while on $\Gamma_D = \partial\Omega \backslash \Gamma_N$ a Dirichlet boundary–condition is given.

The objective is to evaluate the mean–flow over the Neumann boundary. That is,

$$J(u) = \int_{\Gamma_N} u \, dx.$$

The corresponding dual–problem is given by

$$\begin{cases} -\varepsilon \Delta z - \boldsymbol{a} \cdot \nabla z = 0 & \text{in } \Omega = (0,1)^2 \\ z = 0 & \text{on } \Gamma_D \\ \varepsilon \boldsymbol{n} \cdot \nabla z + \boldsymbol{n} \cdot \boldsymbol{a} z = 1 & \text{on } \Gamma_N. \end{cases} \tag{5.29}$$

This time, the *a posteriori* analysis proceeds as follows:

$$\begin{aligned} J(u) - J(u_h) &= J(u - u_h) \\ &= \mathcal{L}(u - u_h, z) \\ &= \mathcal{L}(u - u_h, z - z_h) + \sum_{T \in \mathcal{T}_h} (u_b, L^* z_h)_T \\ &= \sum_{T \in \mathcal{T}_h} \left((f - L u_h, z - z_h)_T - (\varepsilon \boldsymbol{n}_T \cdot \nabla u_h, z - z_h)_{\partial T \cap (\Omega \cup \Gamma_N)} \right) \\ &\quad + \sum_{T \in \mathcal{T}} (u_b, L^* z_h)_T \\ &= \sum_{T \in \mathcal{T}_h} \Big((R_T(u_h), z - z_h)_T - \frac{1}{2} (\varepsilon \, \boldsymbol{n}_T \cdot [\nabla u_h], z - z_h)_{\partial T \cap \Omega} \\ &\quad - (\varepsilon \, \boldsymbol{n}_T \cdot \nabla u_h, z - z_h)_{\partial T \cap \Gamma_N} + (u_b, L^* z_h)_T \Big), \end{aligned}$$

with the new term related to the presence of the Neumann boundary.

The results obtained using the error estimator $\bar{\mathcal{E}}_1$, modified to include the Neumann term, are shown in Figure 5.4. We notice that the refinement is mainly concentrated along the internal layer in the primal solution, but some refinement is performed where features of the dual solution are present. This should be compared with the output of mesh adaptation driven by norm error examined in the next section.

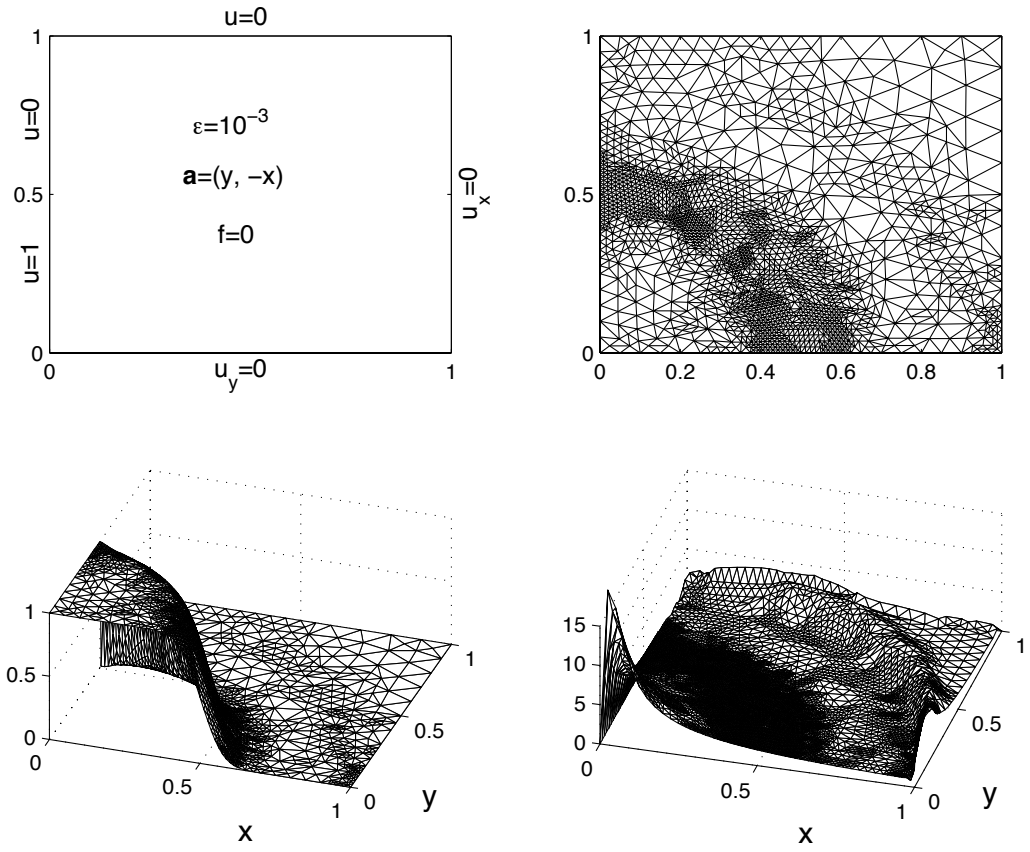


Figure 5.4: Example 4. Above: problem specifications (left) and mesh after five refinements (right) for $J(u) = \int_{\Gamma_N} u dx$ and $\varepsilon = 10^{-3}$. Below: The corresponding solution (left) and dual solution (right).

5.3 The bubble as error estimator

In this section we discuss *a posteriori* error estimation with respect to the energy norm on shape regular triangulations.

We consider the following model problem with mixed Dirichlet and Neumann boundary conditions:

$$\begin{cases} -\varepsilon\Delta u + \boldsymbol{a}\cdot\nabla u = f & \text{in } \Omega, \\ u = 0 & \text{on } \Gamma_D, \\ \varepsilon \boldsymbol{n}\cdot\nabla u = g & \text{on } \Gamma_N, \end{cases} \tag{5.30}$$

with $\partial\Omega = \Gamma_D \cap \Gamma_N$ and $\Gamma_D \cup \Gamma_N = \emptyset$, assuming that Γ_D is closed and has nonzero measure.

Let $\mathcal{T}_h$, $h > 0$ be a family of triangulations of Ω such that the following conditions hold.

1. *Conformity*: Any two triangles in $\mathcal{T}_h$ either have a common edge or common vertex, or they do not intersect at all;

2. *Consistency*: Any triangle–edge lying on the boundary is contained either in Γ_D or in Γ_N;

3. *Shape regularity*: For any triangle, the ratio of the largeest circumscribed circle to that of the smallest inscribed circle is bounded by a constant which does not depend on the triangle and on h.

As we have mentioned in Section 5.1, the bubble part of the RFB finite element solution is given element–wise as the function $u_b|_T \in H_0^1(T)$ satisfying

$$-\varepsilon \Delta u_b|_T + \boldsymbol{a} \cdot \nabla u_b|_T = (f - Lu_h)|_T \qquad \forall T \in \mathcal{T}_h. \tag{5.31}$$

Thus, we can bound the L_2–norm of the bubble u_b using the following stability result for the convection–diffusion equation.

Let ω be a bounded Lipschitz domain in $\mathbb{R}^2$ and let $w \in H_0^1(\omega)$ be such that

$$\begin{cases} -\varepsilon \Delta w + \boldsymbol{a} \cdot \nabla w = f & \text{in } \omega, \\ w = 0 & \text{on } \partial\omega, \end{cases}$$

where $\varepsilon > 0$, $f \in L_2(\omega)$ and $\boldsymbol{a} \in \left(C^1(\omega)\right)^2$ satisfies

$$\begin{cases} \nabla \cdot \boldsymbol{a} \leq 0 & \text{in } \omega, \\ \boldsymbol{a} \text{ has no closed integral curves in } \bar{\omega}. \end{cases}$$

Then, according to [77], there exist a constant C dependent only on ω and on $\boldsymbol{a}$ such that

$$\varepsilon^{1/2} |w|_{1,\omega} + \|w\|_{0,\omega} \leq C \|f\|_{0,\omega}. \tag{5.32}$$

The following result is mentioned by Russo in [84] where the case of constant coefficients is treated.

Proposition 5.3.1 *Let $u_b|_T \in H_0^1(T)$ be the solution of (5.31). If $\nabla \cdot \boldsymbol{a} \leq 0$ and $\boldsymbol{a}$ has no closed integral curves in T, then*

$$\|u_b\|_{0,T} \leq C h_T \|f - \boldsymbol{a} \cdot \nabla u_h\|_{0,T}, \tag{5.33}$$

where C is independent of ε and h_T.

Proof. Upon scaling problem (5.31) by using the transformation $A : (x, y) \to h_T(\xi, \eta)$ we deduce that the image of $u_b|_T$, say $\hat{u}_b \in H_0^1(\widehat{T})$, where $\widehat{T} = A(T)$, satisfies

$$-\frac{\varepsilon}{h_T}\Delta\hat{u}_b(\xi,\eta) + \boldsymbol{a}\cdot\nabla\hat{u}_b(\xi,\eta) = h_T(f - \boldsymbol{a}\cdot\nabla u_h)(x(\xi,\eta), y(\xi,\eta)).$$

To obtain (5.33), we scale the stability estimate (5.32) applied to $\hat{u}_b$:

$$\|u_b\|_{0,T} \le Ch_T\|\hat{u}_b\|_{0,\widehat{T}} \le Ch_T^2\|(f - \boldsymbol{a}\cdot\nabla u_h)\circ A^{-1}\|_{0,\widehat{T}} \le Ch_T\|f - \boldsymbol{a}\cdot\nabla u_h\|_{0,T},$$

where C depends on $\boldsymbol{a}$ and the shape regularity of T. □

This result permits us to repeat, for the RFB formulation with linear finite elements, the *a posteriori* analysis carried out by Verfürth [94] for the SUPG method.

From the coercivity of $\mathcal{L}$ we have

$$\varepsilon^{1/2}|u - u_h|_{1,\Omega} \le \sup_{v\in V\setminus\{0\}} \frac{\mathcal{L}(u - u_h, v)}{\varepsilon^{1/2}|v|_{1,\Omega}}.$$

Now let $v \in V$ be such that $\varepsilon^{1/2}|v|_{1,\Omega} = 1$. For any such v we have

$$\varepsilon^{1/2}|u - u_h|_{1,\Omega} \le \mathcal{L}(u - u_h, v - P_hv) + \mathcal{L}(u - u_h, P_hv) = I + II.$$

Let $\mathcal{E}_{h,N}$ be the set of all edges contained in the Neumann boundary Γ_N. To bound the first term on the right–hand side we employ Green's formula element–wise:

$$\begin{aligned}
I &= \sum_{T\in\mathcal{T}_h}\left((f - Lu_h, v - P_hv)_T - \frac{1}{2}(\varepsilon\,\boldsymbol{n}_T\cdot[\nabla u_h], v - P_hv)_{\partial T\cap\Omega}\right) \\
&\quad + \sum_{E\in\mathcal{E}_{h,N}}(g - \varepsilon\boldsymbol{n}\cdot\nabla u_h, v - P_hv)_E \\
&\le \sum_{T\in\mathcal{T}_h}\left(\|f - Lu_h\|_{0,T}\|v - P_hv\|_{0,T} + \frac{1}{2}\|\varepsilon\,\boldsymbol{n}_T\cdot[\nabla u_h]\|_{0,\partial T\cap\Omega}\|v - P_hv\|_{0,\partial T\cap\Omega}\right) \\
&\quad + \sum_{E\in\mathcal{E}_{h,N}}\|g - \varepsilon\boldsymbol{n}\cdot\nabla u_h\|_{0,E}\|v - P_hv\|_{0,E} \\
&\le C\Bigg(\sum_{T\in\mathcal{T}_h}\left(h_T^2\varepsilon^{-1}\|f - Lu_h\|_{0,T}^2 + \frac{1}{2}h_T\varepsilon\|\boldsymbol{n}_T\cdot[\nabla u_h]\|_{0,\partial T\cap\Omega}^2\right) \\
&\quad + \sum_{E\in\mathcal{E}_{h,N}} h_T\varepsilon^{-1}\|g - \varepsilon\boldsymbol{n}\cdot\nabla u_h\|_{0,E}^2\Bigg)^{1/2},
\end{aligned} \tag{5.34}$$

having made use of the Cauchy–Schwarz inequality, the interpolant approximation properties (5.12) and (5.13) applied on T and the shape regularity of $\mathcal{T}_h$.

To bound II, we notice that since u is the solution of (5.1), u_h solves (5.3) and $\Delta P_h v = 0$, we have

$$\begin{aligned} II &= \sum_{T\in\mathcal{T}_h} (u_b, L^* P_h v)_T \\ &= -\sum_{T\in\mathcal{T}_h} (u_b, \boldsymbol{a}\cdot\nabla P_h v)_T \\ &\leq \sum_{T\in\mathcal{T}_h} \|u_b\|_{0,T} \|\boldsymbol{a}\cdot\nabla P_h v\|_{0,T} \\ &\leq C \sum_{T\in\mathcal{T}_h} h_T \varepsilon^{-1/2} \|f - \boldsymbol{a}\cdot\nabla u_h\|_{0,T} \|\boldsymbol{a}\|_{\infty,T} \varepsilon^{1/2} |P_h v|_{1,T} \\ &\leq C \left\{ \sum_{T\in\mathcal{T}_h} \|\boldsymbol{a}\|^2_{\infty,T} h_T^2 \varepsilon^{-1} \|f - \boldsymbol{a}\cdot\nabla u_h\|^2_{0,T} \right\}^{1/2} \end{aligned} \tag{5.35}$$

thanks to (5.33), (5.14) and the shape regularity of $\mathcal{T}_h$.

Thus, from (5.34) and (5.35) we conclude that the following error bound holds:

$$\varepsilon^{1/2} |u - u_h|_{1,\Omega} \leq C\, \mathcal{E}_S, \tag{5.36}$$

with the error estimator $\mathcal{E}_S$ given by

$$\begin{aligned} \mathcal{E}_S^2 &:= \sum_{T\in\mathcal{T}_h} \left(h_T^2 \varepsilon^{-1} \|f - Lu_h)\|^2_{0,T} + \frac{1}{2} h_T \varepsilon \|\boldsymbol{n}\cdot[\nabla u_h]\|^2_{0,\partial T\cap\Omega} \right) \\ &+ \sum_{E\in\mathcal{E}_{h,N}} \|\boldsymbol{a}\|^2_{\infty,T} h_T \varepsilon^{-1} \|g - \varepsilon \boldsymbol{n}\cdot\nabla u_h\|^2_{0,E}. \end{aligned}$$

Remark. The *a posteriori* error analysis in both the $\varepsilon^{1/2}$–weighted H^1–seminorm and the L_2–norm error of the RFB method based on general finite elements (i.e. not necessarily linear) was carried out by Sangalli [86] under the hypothesis that the vector field $\boldsymbol{a}$ has no closed integral curves on the whole domain $\overline{\Omega}$. Sangalli's result, though, applies to the component of the RFB solution in the space

$$\widetilde{W}_h = \{ v \in V_{\mathrm{RFB}} : L^* v = 0 \quad \text{in each element } T \in \mathcal{T}_h \},$$

instead of the piecewise polynomial component as is the case of our bound.

Under such a hypothesis on $\boldsymbol{a}$ and again restricting ourselves to the case of linear finite elements, we can repeat Sangalli's argument, this time applied to the polynomial component of the solution by employing (5.33). In this way the error bound (5.36) follows, with the extra control on the L_2–norm error.

Example 5. We consider again the boundary–value problem of Example 1 of the previous section. The outcome of the implementation of the error estimator (5.36) in an adaptive

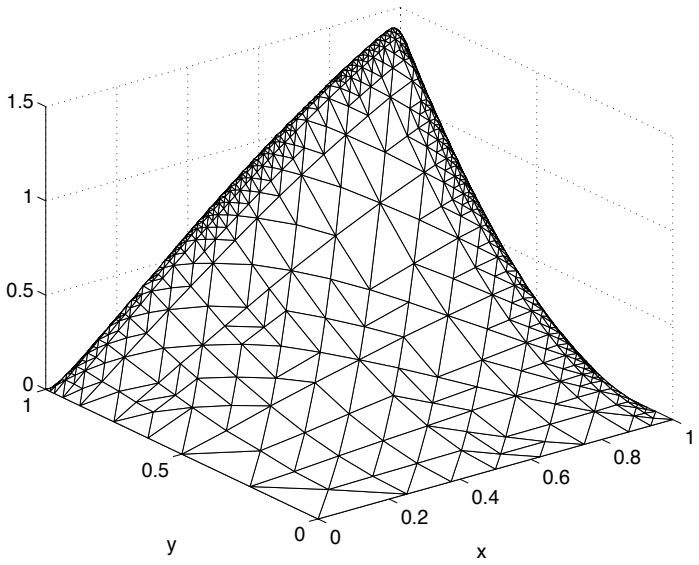


Figure 5.5: Example 5. The final solution produced by the algorithm based on the $\mathcal{E}_S$ error indicator.

algorithm is displayed in Table 5.4 (top). The quantities S_i and S_b reported in the table are the two terms in the error estimator which are relevant to this problem, i.e.

$$\begin{aligned} S_i^2 &= \sum_{T\in\mathcal{T}_h} h_T^2\varepsilon^{-1}\|f - Lu_h\|_{0,T}^2, \\ S_b^2 &= \frac{1}{2}\sum_{T\in\mathcal{T}_h} h_T\varepsilon\|\boldsymbol{n}\cdot[\nabla u_h]\|_{0,\partial T\cap\Omega}^2. \end{aligned}$$

In the computation the marking parameters and the tolerance were set to

$$c_{\text{ref}} = .3, \qquad p_{\text{ref}} = .1, \qquad \texttt{TOL} = 0.1.$$

The final solution and mesh can be seen in Figure 5.5. The intermediate meshes are similar to those shown in Figure 5.10 below.

We have repeated the same experiment but using as error indicator the L_2–norm of the bubble part of the RFB solution, which we know to be a representation of the internal residual. Thus, we define the new indicator

$$\mathcal{E}_B^2 = \sum_{T\in\mathcal{T}_h} \|\boldsymbol{a}\|_{\infty,T}^2\varepsilon^{-1}\|u_b\|_{0,T}^2.$$

The results obtained using this error indicator are shown in Table 5.4 (bottom). Of course, $\mathcal{E}_B$ is reliable only as long as the internal residual term S_i dominates the edge residual term S_b.

From the table we observe that this may be true initially, but not asymptotically. Indeed, it has been proved by Kunert and Verfürth [68] and Carstensen and Verfürth [32] that edge residuals *dominate* the error of the finite element method in the case of low order finite elements, by showing that the edge residuals yield global upper and local lower bounds on the error, both in the H^1– and the L_2–norm.

So $\mathcal{E}_B$ cannot alone be considered as a reliable *a posteriori* error bound. However, it gives a more effective estimate of the internal residual than S_i; see the effectivities reported in Table 5.4.

Alternatively, $\mathcal{E}_B$ may be used to decide wherever to locally turn off the stabilisation term. This idea is developed in the next section.

Example 6. Our final example of this section regards the mixed boundary value problem Example 4 in the previous section. The problem data are specified in Figure 5.4 (above–left): a discontinuity in the Dirichlet boundary condition at $(x, y) = (0, 1/2)$ is propagated (and smoothed) inside the domain and exits through the Neumann boundary.

Figure 5.9 shows the successive refinements obtained using the bubble error indicator $\mathcal{E}_B$, with the tolerance set to $\texttt{TOL} = 0.05$. The solution on the final mesh is also shown.

Refinement is limited to the areas with relatively strong variations of ∇u. Figure 5.9 should be compared with Figure 5.4 which reports the triangulation obtained by solving the same boundary–value problem but with the refinement driven by a linear functional of the solution as target.

5.4 *hb*–adaptivity

In this chapter we have examined adaptive mesh refinement algorithms for the stabilised Galerkin finite element formulation (5.3) derived from the RFB method. The stabilisation term in (5.3) depends on the bubble part of the solution u_b. Using the assumption of local constant coefficients, we have been able to reduce the complex task of the evaluation of such term to that of the evaluation of an average of the bubble. Further, the approximation (5.20) led to the even simpler task of the evaluation of the elemental diameter in the direction of the convective field. As a consequence, in terms of computational complexity, the resulting method in the case of piecewise linear elements is equivalent to the SDFEM, with the advantage that the stabilisation parameter is given by the method.

This way of proceeding, of course, has some limitations. The approximation (5.20) may not be sufficiently accurate if the coefficients in the p.d.e. cannot be treated as piecewise constant functions or if the mesh is sufficiently refined, as we have seen in Chapter 3. One may also wish to compute the bubbles more accurately, for example to include an edge–stabilisation of the sort discussed in the previous chapter. Finally, it is clear that in parts of the computational domain where the mesh has been sufficiently refined or where the solution is relatively flat, there is no reason for stabilising in the first place.

$N_{\rm el}$	$\|u-u_h\|_{0,\Omega}$	$\varepsilon^{1/2}\|u-u_h\|_{1,\Omega}$	$\mathcal{E}_S(u_h)$	S_i	S_b	*eff*$_{l2}$	*eff*$_{en}$
32	0.29	0.9	10.4	10.4	0.44	35.9	11.5
69	0.19	0.83	7.13	7.1	0.58	36.3	8.6
155	0.12	0.74	4.71	4.6	0.66	39.1	6.3
329	0.06	0.59	2.99	2.9	0.69	45	5
711	0.034	0.4	1.83	1.7	0.62	52.9	4.6
1501	0.018	0.24	1.1	0.1	0.5	58.8	4.6
3706	0.01	0.13	0.66	0.5	0.35	66.5	4.8
11272	0.005	0.07	0.38	0.3	0.22	72.7	5.06

$N_{\rm el}$	$\|u-u_h\|_{0,\Omega}$	$\varepsilon^{1/2}\|u-u_h\|_{1,\Omega}$	$\mathcal{E}_B(u_h)$	*eff*$_{l2}$	*eff*$_{en}$	S_i	S_b
32	0.29	0.9	3.01	10.3	3.3	10.4	0.44
69	0.19	0.83	2.03	10.3	2.4	7.1	0.58
143	0.12	0.7	1.33	10.9	1.79	4.7	0.67
310	0.068	0.59	0.83	12.2	1.4	3.03	0.69
650	0.036	0.4	0.49	13.6	1.21	1.8	0.62
1358	0.019	0.24	0.27	14.1	1.12	1.04	0.51
3137	0.011	0.14	0.15	14.2	1.07	0.6	0.37
9367	0.0058	0.08	0.08	14.4	1.02	0.3	0.24
28705	0.0032	0.04	0.04	14.6	0.98	0.19	0.15

Table 5.4: Example 5. Error over successive refinements in the L_2 and $\varepsilon^{1/2}$–weighted H^1–seminorm (energy norm); $\varepsilon = 10^{-2}$. $\mathcal{E}_S$ indicator (above) and bubble indicator $\mathcal{E}_B$ (below).

This justifies the idea of including in the mesh adaptation algorithms discussed so far (*h–refinement*) an automatic way of 'turning off' the stabilising term wherever this is no more required (*b–derefinement*).

To address the crucial issue of when and where to phase out the bubble stabilisation alongside the *h*–refinement process, we may use the particular residual term in the *a posteriori* error bound which is related to the stabilising term itself.

Let us consider, in particular, the $\varepsilon^{1/2}$–weighted H^1–seminorm error bound $\mathcal{E}_S$ in (5.36), discarding, for simplicity, any term related to Neumann boundary conditions. Looking back to (5.35) we see that $\mathcal{E}_S$ in fact consists of three terms, S_i, S_b and $\mathcal{E}_b$, related to the internal residual, the boundary jump and the stabilising term, respectively. Let S_i^T, S_b^T and $\mathcal{E}_b^T$ be the

local components of these terms, that is,

$$\begin{aligned} S_i^T &= h_T \varepsilon^{-1/2} \| f - (-\varepsilon \Delta u_h + \boldsymbol{a} \cdot \nabla u_h) \|_{0,T} \\ S_b^T &= \left(\frac{h_T \varepsilon}{2} \right)^{1/2} \| \boldsymbol{n} \cdot [\nabla u_h] \|_{0, \partial T \cap \Omega} \\ \mathcal{E}_b^T &= \| \boldsymbol{a} \|_{\infty,T} \varepsilon^{-1/2} \| u_b \|_{0,T}. \end{aligned}$$

We have seen that the term $\mathcal{E}_b^T$ is bounded by S_i^T. Moreover, we expect the term $\mathcal{E}_b^T$ to be relatively small wherever stabilisation is not crucial. Hence we use the relative magnitude of $\mathcal{E}_b^T$ with respect to S_i^T and S_b^T as an indicator for b–derefinement.

Let

$$\eta_T = S_i^T + S_b^T,$$

be the local error indicator and let $\bar{\eta}$ be the maximum of η_T after cutting the upper 10% or 5% of the values. Finally, let $c_{\text{ref}} \in (0,1)$ and $c_b > 0$ be some user–selected threshold parameters and `TOL` a given tolerance.

We propose the following algorithm, which is a modification of the one defined in Section 5.1.1:

1. Define an initial mesh;
2. Calculate u_h on the current mesh;
3. Check the stopping criterion: `IF` $\mathcal{E}_S \leq$ `TOL`$/C$ then `STOP`;
4. Apply the h–refinement criterion: refine those elements whose error estimator $\mathcal{E}_S^T = S_i^T + S_b^T$ exceeds $c_{\text{ref}} \bar{\eta}$;
5. Apply the b–derefinement criterion: `IF` $\mathcal{E}_b^T < c_b \left(S_i^T + S_b^T \right)$, turn off the stabilisation term on T or its newly defined sons and `GOTO` 2.

Since we are not assuming any *a priori* knowledge about the behaviour of the solution, the algorithm is started up with the stabilising term turned on everywhere. Later on, the fraction of elements selected for b–derefinement depends on the threshold parameter c_b. Its 'correct' value should be expected to be subject to the relative sharpness of the different components of the error bound. In both examples below the value $c_b = 0.1$ produced satisfactory results.

Example 7. It is instructive to apply the *hb*–adaptive algorithm to Example 5 above. We recall that we are solving:

$$\begin{cases} -\varepsilon \Delta u + u_x + u_y = f & \text{in } \Omega = (0,1)^2, \\ u = 0 & \text{in } \partial\Omega. \end{cases}$$

N_{el}	$\varepsilon^{1/2}\lvert u-u_h\rvert_{1,\Omega}$	$\mathcal{E}_S(u_h)$	eff_{en}	N_{el}	$\varepsilon^{1/2}\lvert u-u_h\rvert_{1,\Omega}$	$\mathcal{E}_S(u_h)$	eff_{en}
32	0.9	10.4	11.5	32	0.9	10.4	11.5
69	0.83	7.13	8.6	69	0.83	7.13	8.6
155	0.74	4.71	6.3	155	0.74	4.86	6.5
329	0.59	2.99	5	332	0.59	3.15	5.3
711	0.4	1.83	4.6	727	0.39	1.92	4.8
1501	0.24	1.1	4.6	1523	0.22	1.16	5.2
3706	0.13	0.66	4.8	3808	0.123	0.67	5.4
11272	0.07	0.38	5.06	11240	0.067	0.38	5.7

Table 5.5: Example 5. Error and $\mathcal{E}_S$ indicator over successive refinements in the $\varepsilon^{1/2}$–weighted H^1–seminorm (energy norm); $\varepsilon = 10^{-2}$. We show the output of the h–refinement algorithm (left) and hb–algorithm (right).

The exact solution exhibits a boundary layer at the outflow boundary $x = 1$ and $y = 1$ (see Figure 5.5). Ideally, stabilisation should be employed only in the layer. The sequence of meshes produced is shown in Figure 5.10: the yellow elements are those where stabilisation is turned on. We notice that stabilisation is soon removed away from the layer and, when this is resolved, also inside it.

The effectiveness of the new algorithm can be appreciated from the numbers displayed in Table 5.5: the outputs of the old algorithm are reproduced on the left (cf. Table 5.4), while, on the right, we present those of the hb–refinement algorithm. Similar accuracy is achieved with comparable meshes. The new algorithm uses more elements while removing the stabilisation from the layer. On the other hand, the final mesh, which is layer–resolving, is slightly more effective since the same accuracy is obtained with 32 elements less. Over all, if the parameter c_b is appropriately tuned, the numerical solution is not corrupted and the algorithm is robust.

Example 8. To conclude, we test the hb–refinement algorithm on a problem whose solution exhibits an internal layer, namely Example 6 above. The sequence of mesh refinements is depicted in Figure 5.11. Again we have marked in yellow the elements where stabilisation is present. To highlight more clearly such elements, the final mesh is plotted a second time at the bottom–right of the figure with the stabilised elements lifted out of the xy–plane. The meshes are very similar to those obtained without b–derefinement, cf Figure 5.9: this may have to do with the fact that the layer of the solution of this problem is less severe then the one of Example 7 above.

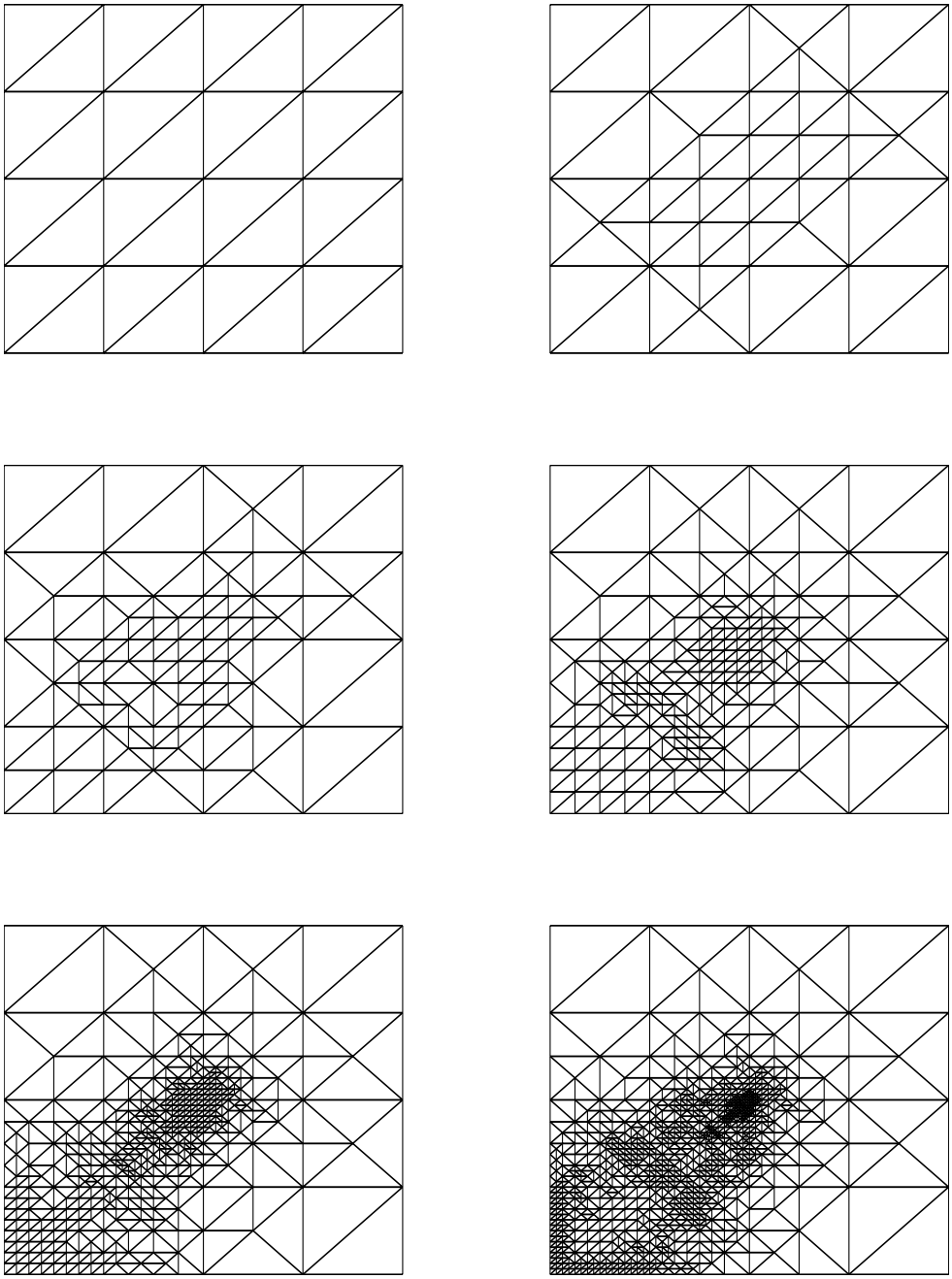


Figure 5.6: Example 2. Successive mesh refinements, $\varepsilon = 10^{-2}$, $P = (.49, .49)$

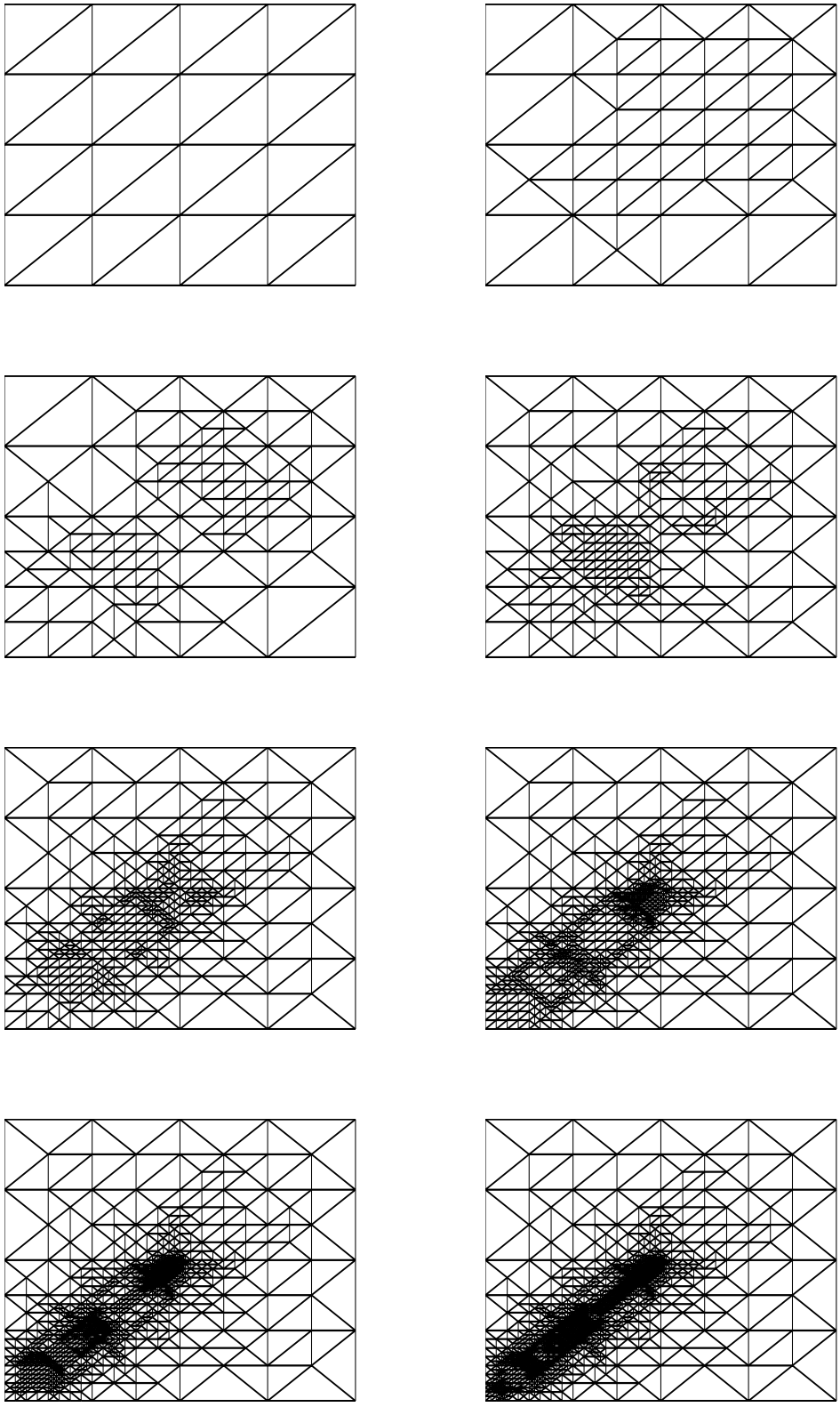


Figure 5.7: Example 2. Successive mesh refinements, $\varepsilon = 10^{-6}$, $P = (.49, .49)$

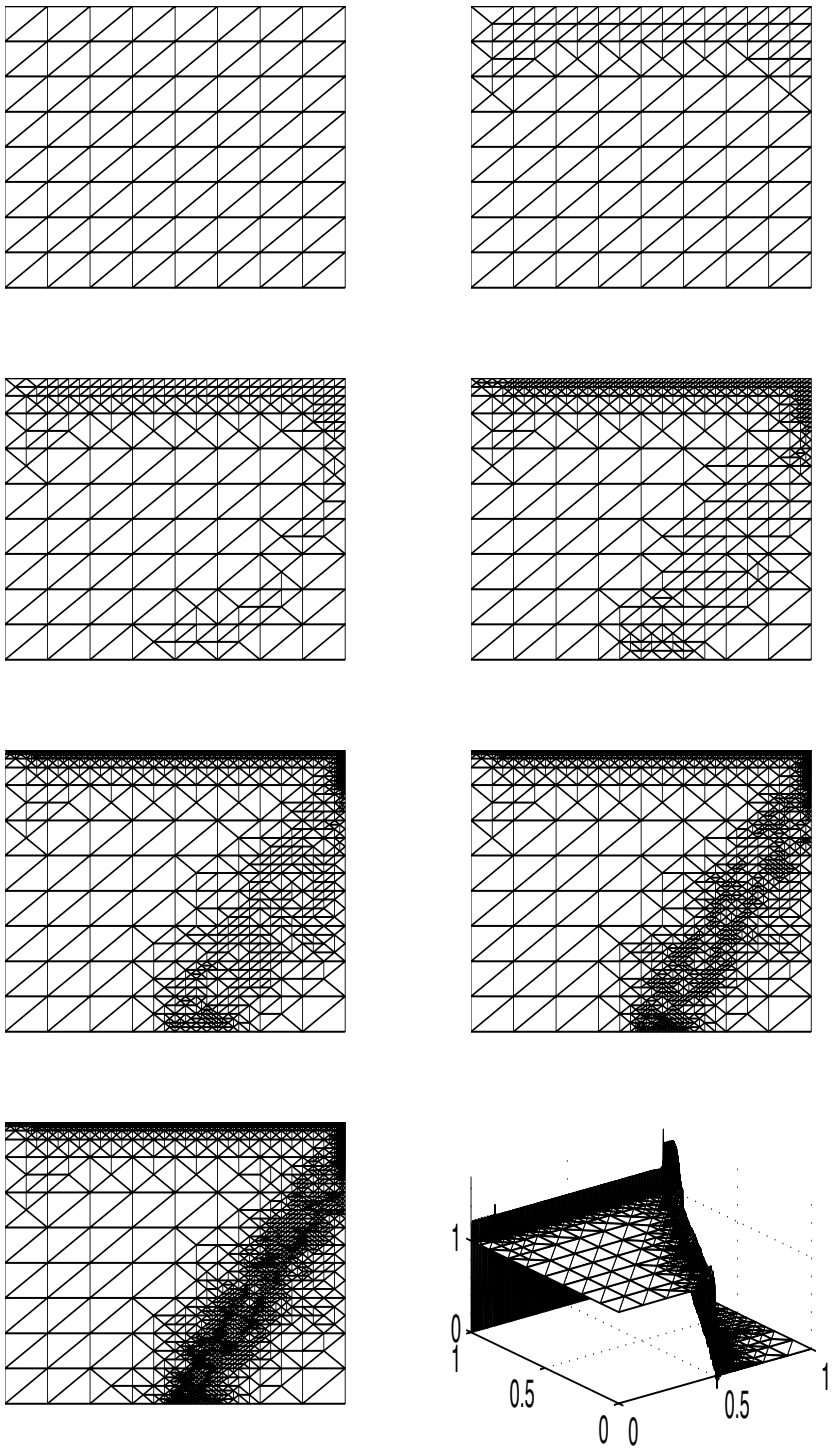


Figure 5.8: Example 3. Successive mesh refinement and final solution for $J(u) = \int_\Omega u$ and $\varepsilon = 10^{-6}$.

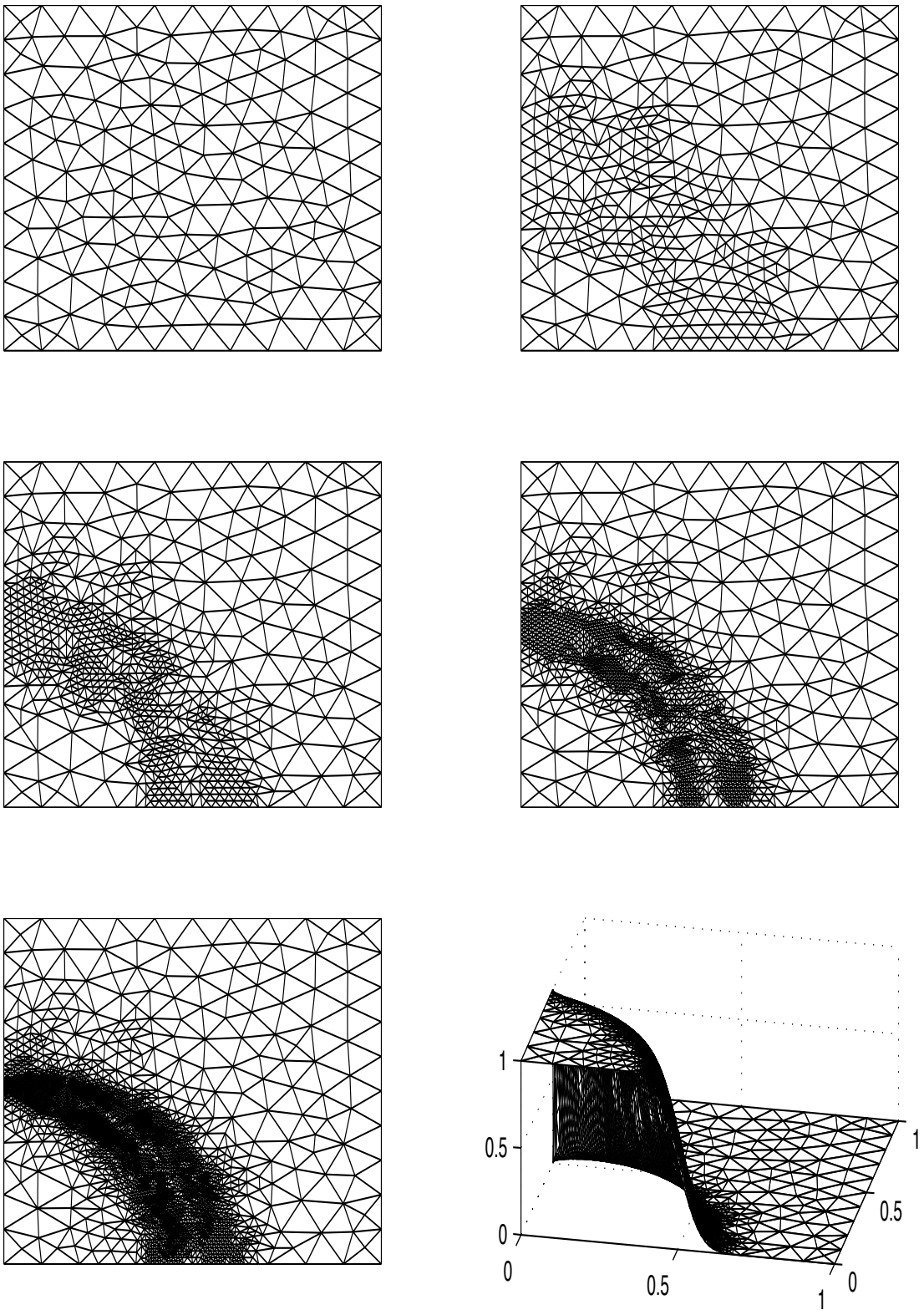


Figure 5.9: Example 6. Successive mesh refinement using the error estimator $\mathcal{E}_B$, $\varepsilon = 10^{-3}$.

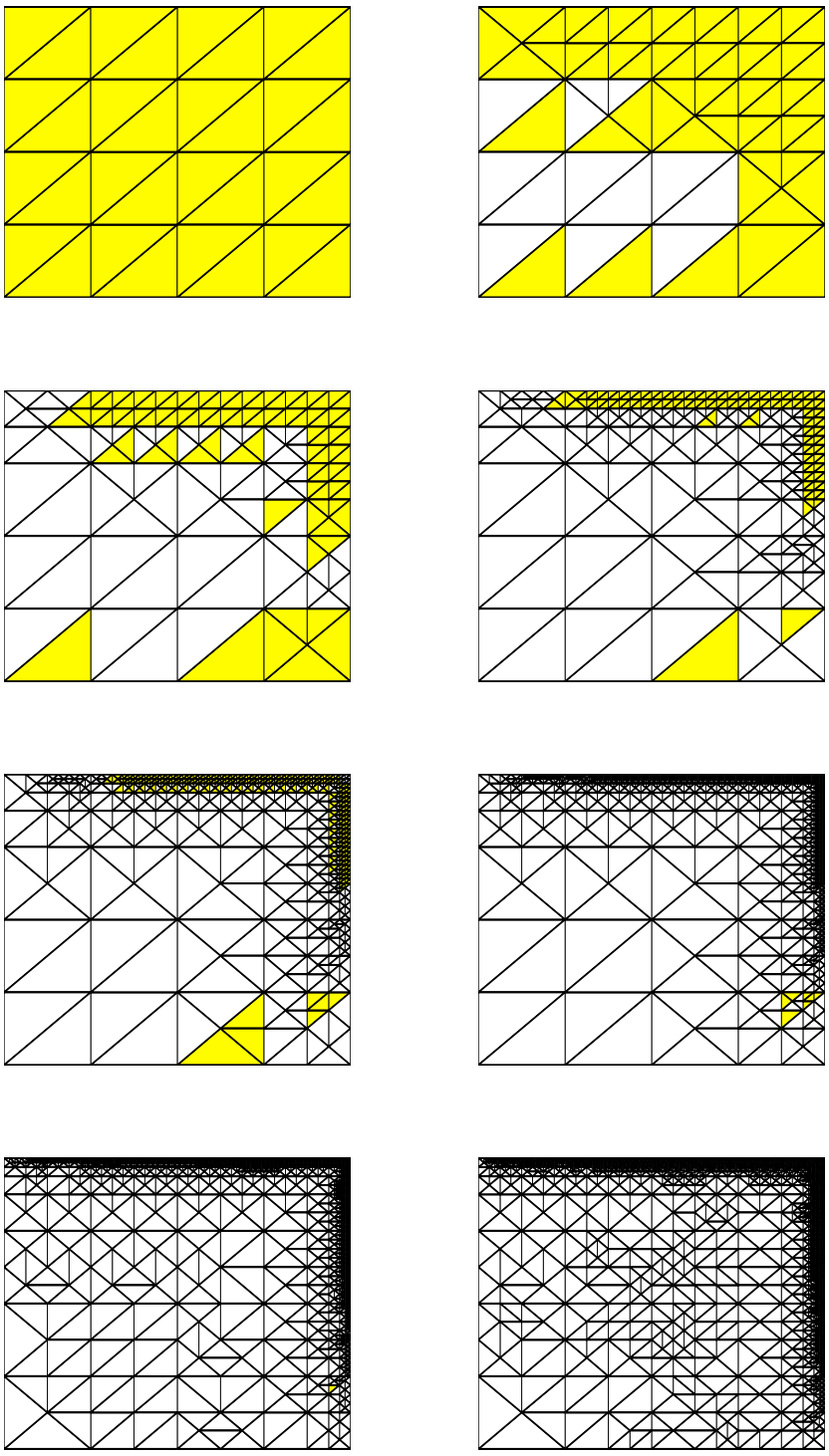


Figure 5.10: Example 7. Successive mesh refinement and bubble derefinement, $\varepsilon = 10^{-2}$.

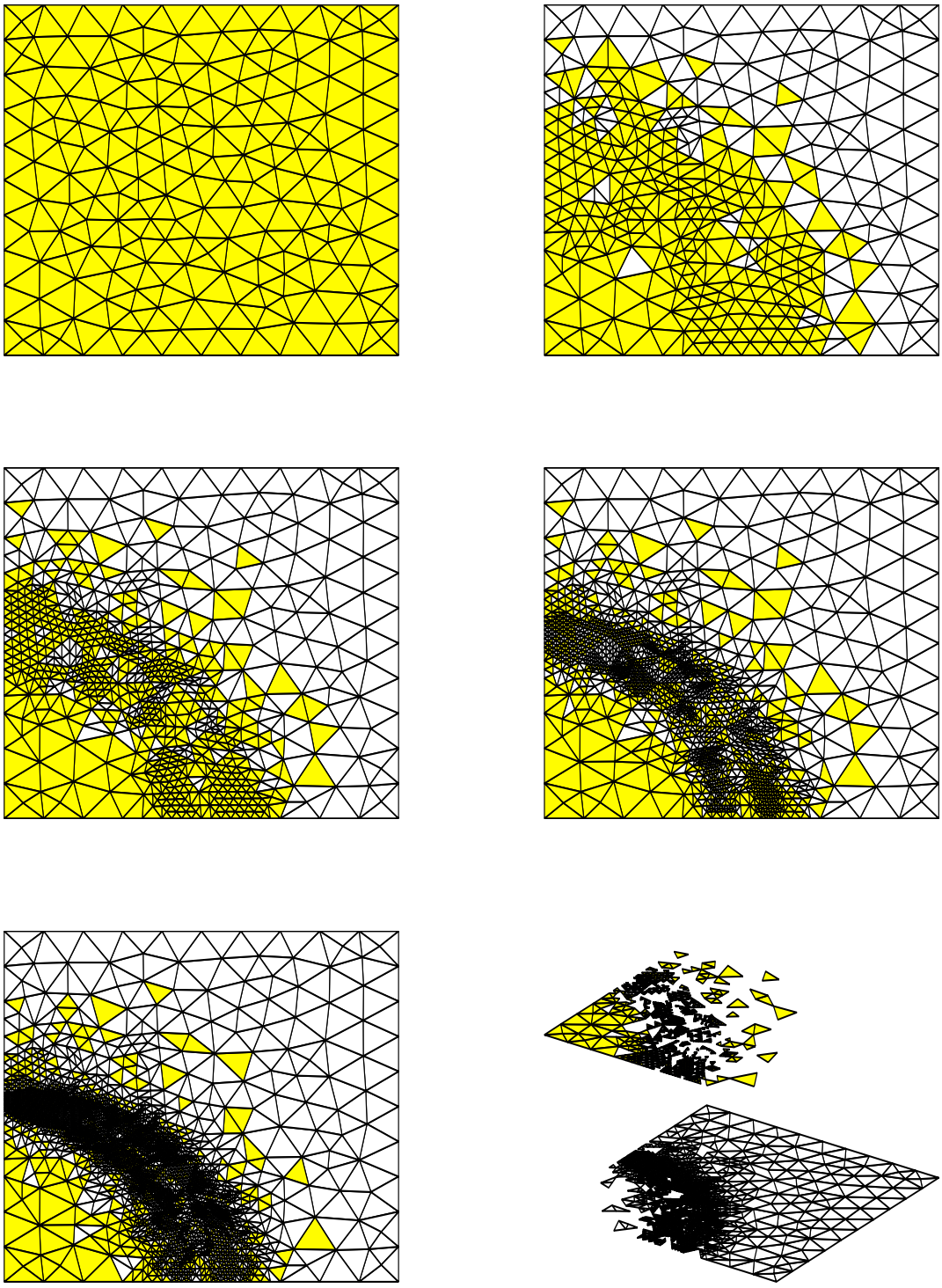


Figure 5.11: Example 8. Successive mesh refinement and bubble derefinement, $\varepsilon = 10^{-3}$.

Chapter 6

Conclusions and future work

In this chapter we summarise the work presented in this thesis and indicate directions for further research.

6.1 Conclusions

In this thesis we analysed the mathematical properties and discussed the implementation of the residual–free bubble (RFB) finite element method as a two–level procedure.

In particular, we studied the RFB method for the accurate and stable solution of steady–state convection–diffusion problems focusing on the convection–dominated regime.

Our analysis guarantees the successful application of the method to a range of multiscale problems. Indeed, the ideas developed are based on the interpretation of convection–dominated diffusion problems as multiscale problems in which the fine scales are highly localised while their effect on the accuracy of the numerical solution is global.

After an introductory survey of known properties in Chapter 2, we embarked in Chapter 3 on the *a priori* analysis of the energy–norm error of the RFB method on anisotropic meshes. The methodology is closely related to that introduced by Sangalli [85], defining an interpolation operator that fully exploits the richness of the RFB finite element V_{RFB}. In this way sharp error bounds are derived. When the problem is strongly convection–dominated, the solution is highly anisotropic. For this reason it is crucial that the error is bounded by appropriately weighted norms of directional derivatives of the solution. To this end, we used ad hoc anisotropic scaling results and interpolation error bounds.

Anisotropy also has to be taken into account in the tuning of the parameters appearing in *streamline–diffusion* type methods. We have used the stabilising term derived from the RFB method to re–define the mesh Péclet number and propose a new choice of the SD–parameter suited to the anisotropic nature of the partition.

In Chapter 4, we defined, analysed and presented an implementation of the RFBe (*enhanced residual–free bubble*) method. This method is designed to combine the stabilising effect of the

residual–free bubbles with an improved resolution of the boundary layer behaviour. The latter is obtained by further enriching the RFB method with *edge bubbles* inside the layer. The choice of the edge bubbles is justified by *a priori* analysis of a particular problem for which the asymptotic approximation of the solution is known.

A key point in the analysis is that the global error of a locally residual–free finite element method is governed by the approximation properties of the finite element space on the skeleton of the partition. For this reason it has been possible to express the interpolation error as a sum of the errors committed over the edges of the triangulation. Subsequently, the value of the edge bubble is fixed by matching with a first order asymptotic approximation. We obtained an *a priori* error bound explicit in the diffusion parameter ε. This shows that the RFBe scheme has better accuracy properties than RFB in the preasymptotic regime $\varepsilon \lesssim h$.

Moreover, numerical examples show that, although we are acting locally by introducing a few edge–based bubbles, the RFBe method is globally more accurate than the classical RFB. This fact indicates that the introduction of the edge bubbles has a stabilising effect on the method. In particular, we observed results as good as those obtained with the shock–capturing method [75]. The new method is able to achieve the same degree of accuracy as the RFB method in a considerably smaller amount of CPU time, even on a sequential computer.

Finally, in Chapter 5, we developed an h–adaptive algorithm based on our residual–based *a posteriori* error bounds. The error is represented in terms of the residual of the finite element approximation weighted by the solution of the dual problem. The error estimate depends on the choice of the dual problem. The numerical approximation of linear functionals of the solution as well as energy norm error estimation have been considered.

We have shown that the elimination of the dual solution from the *a posteriori* bound, obtained via strong stability estimates, leads to suboptimal bounds. For this reason, the algorithm proposed is based on, so called, Type I error bounds, i.e. *a posteriori* error bounds depending explicitly on the dual solution. The downside of this approach is that, in order to obtain a computable error bound, the dual solution need to be computed numerically. Moreover, for the error bound to be reliable, it is necessary that the dual problem is solved accurately. The recipe followed has been to use a dual mesh with mesh size half of that of the primal mesh. In the examples considered the algorithm was seen to be reliable and effective.

The *a posteriori* analysis of the energy norm error of the RFB method, also presented in Chapter 5, is similar to that of the SUPG method due to Verfürth [94]. The error bound obtained consists of three terms: the two classical residual–based terms of the Galerkin formulation, i.e. internal residual and boundary jump of the gradient, as well as a third term which stems from static condensation of the bubbles. We showed how the third term can be bounded by means of the other two terms. To this end, we used an appropriate stability result applied to the bubble part of the solution.

We also introduced a new *hb–adaptive* algorithm in which the bubble stabilisation is phased off locally according to the relative magnitude of the three terms mentioned above in the *a posteriori* error bound. In this way, while the mesh is refined, we avoid the evaluation of the

bubble part of the solution where this is not crucial.

6.2 Future work

We developed the RFBe method for the resolution of boundary layers in convection–dominated diffusion problems, assuming that the position of the layers is known *a priori*. The design of equally effective edge bubbles for the resolution of internal layers, not discussed in this thesis, should be considered. Moreover, the method should be combined with *a posteriori* error estimation aimed at locating elements where edge bubbles need to be inserted. To this end, the *hb*–adaptive algorithm proposed in Section 5.4 may be employed. Indeed, the local magnitude of the bubble part of the solution is a clear indicator of the presence of layers and hence of the regions where edge bubbles should be introduced. We are also interested in comparing an adaptive RFBe algorithm with other techniques for the solution of problems of practical relevance in which it is required that the layers are computed accurately.

Finally, the study of the RFBe framework applied to multiscale problems, different from convection–diffusion problems, remains open and is worthy of investigation.

The same applies to the other results obtained here for the classical RFB method, such as the anisotropic *a priori* error analysis and *a posteriori* error analysis.

As we mentioned in the Introduction, some work has been done on RFB applied to the Stokes problem, the incompressible Navier–Stokes equations, the Helmholtz equation and homogenization problems which seem worth pursuing further, although, for the latter, the use of nonconforming techniques like *oversampling* seems necessary, see Hou *et al.* [39].

The *h*–adaptive algorithm considered in Chapter 5 could be made more effective by allowing anisotropic mesh refinement. To this end, a more complex refinement strategy which allows for mesh derefinement is needed.

Appendix A

Relevant topics

A.1 Boundary and internal layers

What is the thickness of a boundary or internal layer? Acheson ([1], pp. 267-268), writing about fluid motion near a no-slip boundary, states that "the rapid variation of u (near the boundary) should be just sufficient to prevent the viscous term from being negligible, notwithstanding the small coefficient of viscosity ν. We may at once use this consideration to obtain an order of magnitude estimate of the boundary layer thickness". Indeed, the thickness of a boundary layer is defined through the mathematical notion of distinguished limit in the asymptotic expansion of the equation with respect to the perturbation parameter. The idea is to define the thickness of a boundary layer as the width of the region in which the magnitude of the higher order term matches that of other terms in the equations. For this purpose, the solution is represented by an asymptotic expansion in terms of ε.

We present such a method through a simple model problem. Consider the boundary value problem

$$\begin{cases} -\varepsilon\Delta u + u_y = f & \text{in } \Omega = (0,1)^2, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{A.1}$$

where ε is a positive constant. We begin by analysing the outflow boundary, given by $y = 1$, where we expect a boundary layer.

Let δ be the width of such layer, which in the context of matched asymptotic expansions is called *inner region*, whereas the remainder of the domain takes the name of *outer region*. We consider a perturbation expansion of the outer solution in the form

$$u_o(x, y, \varepsilon) = \sum_{j=0}^{m} \varepsilon^j u_j(x, y) + o(\varepsilon^m).$$

This has to be matched with an inner solution which is defined in the inner region in terms of

a local stretched coordinate:

$$u_i(x,y,\varepsilon) \equiv w(x,\eta,\varepsilon) = \sum_{j=0}^{m} \varepsilon^j w_j(x,\eta) + o(\varepsilon^m). \tag{A.2}$$

The choice of the stretched variable η gives the range of validity of the inner approximation and so the thickness of the boundary layer. Let $\eta = (1-y)/\varepsilon^\alpha$. Substituting (A.2) into (A.1) we have that w_0 must satisfy

$$-\varepsilon \frac{\partial^2 w_0}{\partial x^2} - \varepsilon^{1-2\alpha} \frac{\partial^2 w_0}{\partial \eta^2} - \varepsilon^{-\alpha} \frac{\partial w_0}{\partial \eta} = f. \tag{A.3}$$

We determine α through the notion of distinguished limit; that is, the correct α is the one that gives a dominant balance between two or more terms of the equation (A.3). The reason for this is that, as explained in [12], p. 437, "in general, only the distinguished limit gives a nontrivial boundary-layer structure which is asymptotically matchable to the outer solution". In this case, the correct choice is $\alpha = 1$ which gives, to leading order,

$$\begin{cases} \frac{\partial^2 w_0}{\partial \eta^2}(x,\eta) + \frac{\partial w_0}{\partial \eta}(x,\eta) = 0, \\ w_0(x,0) = 0. \end{cases}$$

This equation represents an exponential correction. We say that near Γ_+ there is an *exponential boundary layer* of width $\delta \sim \varepsilon$ as $\varepsilon \to 0$.

A similar analysis near the characteristic boundary $x = 0$ and $x = 1$ gives, in terms of the stretched variable $\xi = x/\varepsilon^{1/2}$, the leading term equation

$$-\frac{\partial^2 w_0}{\partial \xi^2} + \frac{\partial w_0}{\partial y} = 0,$$

which is of parabolic type. Near the characteristic boundary we have a *parabolic boundary layer* of thickness $\mathcal{O}(\varepsilon^{1/2})$ as $\varepsilon \to 0$.

The internal layers of the convection-diffusion equation are also parabolic, again of thickness $\mathcal{O}(\varepsilon^{1/2})$.

Since numerical methods are often formulated as *energy* minimisation techniques, it can be convenient to define the inner region as the region in which the gradient of u is bounded uniformly w.r.t. ε. For this reason, in the numerical analysis literature, we also find the layer thickness defined to be of order $\mathcal{O}(\varepsilon \log(1/\varepsilon))$ for an exponential layer and of order $\mathcal{O}(\varepsilon^{1/2} \log(1/\varepsilon))$ for a parabolic layer.

A.2 Shishkin interpolation

Given the one–dimensional boundary value problem

$$\begin{cases} -\varepsilon v_h'' + a v_h' = f & \text{in } I_h = (0,h), \\ v_h(0) = 0, & v_h(h) = 0, \end{cases}$$

we scale it back to the unit interval:

$$\begin{cases} -\varepsilon^* v'' + av' = hf \quad \text{in } I = (0,1), \\ v(0) = 0, \quad v(h) = 0, \end{cases} \tag{A.4}$$

where $\varepsilon^* = \varepsilon/h$, and consider Shishkin interpolation of v.

A Shishkin piecewise equidistant mesh consisting of N subdivisions (with N even) is defined as follows. Given the turning point

$$\lambda^* = c_s(\varepsilon^*/c_a)\ln N,$$

where c_s is a constant independent of ε^* and N, the mesh is taken to be uniform with $N/2$ subdivisions on the two subintervals $(0, 1-\lambda^*)$ and $(1-\lambda^*, 1)$. Thus, the mesh on $[0,1]$ is piecewise uniform. For the continuous piecewise linear interpolant v^I of the solution v of (A.4), the interpolation error over such mesh satisfies (see [82])

$$\varepsilon^* |v - v^I|^2_{1,I} + \|v - v^I\|^2_{0,I} \leq CN^{-2}\ln^2 N,$$

with the constant C independent of ε and N.

Scaling back to the interval I_h we obtain a Shishkin mesh with turning point $\lambda = c(\varepsilon/c_a)\ln N$ and the scaled interpolation error bound

$$\varepsilon |v - v^I|^2_{1,I_h} + h^{-1}\|v - v^I\|^2_{0,I_h} \leq CN^{-2}\ln^2 N.$$

Shishkin meshes on rectangles are constructed by taking a tensor product of 1-D meshes, and then similar approximation results apply.

A.3 Estimates for the asymptotic approximation

We prove the bounds stated in Lemma 4.3.2. Let $u_c = u_{\text{as}} - u_0$ be the collection of all correction terms in the asymptotic approximation u_{as} defined by (4.30). We prove that there exists a constant C independent of ε and h such that

$$\sum_{\Gamma \in \Gamma_{out}} |u_c|^2_{1,\Gamma} \leq C \frac{\mathrm{e}^{-2c_0 c_a h/\varepsilon}}{\varepsilon^2}$$

Let us concentrate on the L_2–norm of the derivative with respect to x (proceeding in a similar way, one can prove a similar bound on the y–derivative). It is sufficient to consider the first term of u_c, i.e.,

$$u_0(1,y)\,\mathrm{e}^{-a_1(1,y)\frac{1-x}{\varepsilon}}.$$

We start by fixing an edge $\Gamma_{ij} = [x_{i-1}, x_i] \times y_j$. We have:

$$\int_{x_{i-1}}^{x_i} \left| u_0(1,y_j)\frac{d}{dx}\mathrm{e}^{-a_1(1,y_j)\frac{1-x}{\varepsilon}} \right|^2 dx \leq C\frac{a_1(1,y_j)}{2\varepsilon}\left(\mathrm{e}^{-2a_1(1,y_j)\frac{1-x_i}{\varepsilon}} - \mathrm{e}^{-2a_1(1,y_j)\frac{1-x_{i-1}}{\varepsilon}}\right).$$

Thus, summing over all edges oriented along the x–axis, we get

$$\begin{aligned}\sum_{j=1}^{n-1}\sum_{i=1}^{m-1}\int_{x_{i-1}}^{x_i}\left|u_0(1,y_j)\frac{d}{dx}\mathrm{e}^{-a_1(1,y_j)\frac{1-x}{\varepsilon}}\right|^2 dx &\leq C\sum_{j=1}^{n-1}\frac{a_1(1,y_j)}{2\varepsilon}\left(\mathrm{e}^{-2a_1(1,y_j)h_{bl}/\varepsilon}-\mathrm{e}^{-2a_1(1,y_j)/\varepsilon}\right)\\ &\leq C\frac{\mathrm{e}^{-2c_0c_ah/\varepsilon}}{\varepsilon h},\end{aligned}$$

where $h_{bl}=1-x_{M-1}$. Let us now consider the edge $\Gamma_{ji}=x_j\times[y_{i-1},y_i]$. This time we have

$$\int_{y_{i-1}}^{y_i}\left|u_0(1,y)\left(\frac{d}{dx}\mathrm{e}^{-a_1(1,y)\frac{1-x}{\varepsilon}}\right)_{x=x_j}\right|^2 dy\leq C\frac{1}{\varepsilon^2}\int_{y_{i-1}}^{y_i}\mathrm{e}^{-2a_1(1,y)\frac{1-x_j}{\varepsilon}}dy\leq C\frac{h}{\varepsilon^2}\mathrm{e}^{-2c_a\frac{1-x_j}{\varepsilon}}.$$

Thus, summing over all edges oriented along the y–axis, we get

$$\begin{aligned}\sum_{j=1}^{m-1}\sum_{i=1}^{n-1}\int_{y_{i-1}}^{y_i}\left|u_0(1,y)\left(\frac{d}{dx}\mathrm{e}^{-a_1(1,y)\frac{1-x}{\varepsilon}}\right)_{x=x_j}\right|^2 dy &\leq C\frac{1}{\varepsilon^2}\sum_{j=1}^{m-1}\mathrm{e}^{-2c_a\frac{1-x_j}{\varepsilon}}\\ &\leq C\frac{\mathrm{e}^{-2c_0c_ah/\varepsilon}}{\varepsilon^2}\sum_{j=0}^{m-2}\mathrm{e}^{-2c_0c_a\frac{jh}{\varepsilon}}\\ &\leq C\frac{\mathrm{e}^{-2c_0c_ah/\varepsilon}}{\varepsilon^2}.\end{aligned}$$

A similar argument yields

$$\sum_{\Gamma\in\Gamma_{out}}||u_c||^2_{0,\Gamma}\leq C\mathrm{e}^{-2c_0c_ah/\varepsilon}.$$

As regards the H^1–seminorm of u_c over Γ_{bl}, we show that

$$\sum_{\Gamma\in\Gamma_{bl}}|u_c|^2_{1,\Gamma}\leq C\frac{1}{\varepsilon h}. \tag{A.5}$$

Let us fix the edge $\Gamma_j=[1-h,1]\times y_j$. The leading term is clearly given by

$$\int_{1-h}^{1}\left|\frac{d}{dx}\mathrm{e}^{-a_1(1,y_j)\frac{1-x}{\varepsilon}}\right|^2 dx=\frac{a_1(1,y_j)}{2\varepsilon}\left(1-\mathrm{e}^{-2a_1(1,y_j)h/\varepsilon}\right)\leq C\frac{1}{\varepsilon},$$

and summing over j, we get (A.5). Again we can proceed in a similar fashion to prove that

$$||u_c||^2_{0,\Gamma}\leq C\varepsilon,$$

and hence

$$\sum_{\Gamma\in\Gamma_{bl}}||u_c||^2_{0,\Gamma}\leq C\varepsilon h^{-1}.$$

Bibliography

[1] Acheson, D. J. *Elementary fluid dynamics.* The Clarendon Press, Oxford University Press, New York, 1990.

[2] Adam, D., Felgenhauer, A., Roos, H.-G., and Stynes, M. A nonconforming finite element method for a singularly perturbed boundary value problem. *Computing 54*, 1 (1995), 1–25.

[3] Adams, R. A., and Fournier, J. J. F. *Sobolev spaces* (Second Edition). Academic Press, New York-London, 2003. Pure and Applied Mathematics, Vol. 140.

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

[5] Apel, T., and Lube, G. Anisotropic mesh refinement in stabilized Galerkin methods. *Numer. Math. 74*, 3 (1996), 261–282.

[6] Arbogast, T. Numerical subgrid upscaling of two-phase flow in porous media. In *Numerical treatment of multiphase flows in porous media (Beijing, 1999)*, vol. 552 of *Lecture Notes in Phys.* Springer, Berlin, 2000, pp. 35–49.

[7] Babuška, I., Andersson, B., Smith, P. J., and Levin, K. Damage analysis of fiber composites. I. Statistical analysis on fiber scale. *Comput. Methods Appl. Mech. Engrg. 172*, 1-4 (1999), 27–77.

[8] Babuška, I., Caloz, G., and Osborn, J. E. Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM J. Numer. Anal. 31*, 4 (1994), 945–981.

[9] Babuška, I., and Osborn, J. Analysis of finite element methods for second order boundary value problems using mesh dependent norms. *Numer. Math. 34*, 1 (1980), 41–62.

[10] Baiocchi, C., Brezzi, F., and Franca, L. P. Virtual bubbles and Galerkin-least-squares type methods (Ga.L.S.). *Comput. Methods Appl. Mech. Engrg. 105*, 1 (1993), 125–141.

[11] BECKER, R., AND RANNACHER, R. A feed-back approach to error control in finite element methods: basic analysis and examples. *East-West J. Numer. Math. 4*, 4 (1996), 237–264.

[12] BENDER, C. M., AND ORSZAG, S. A. *Advanced mathematical methods for scientists and engineers*. McGraw-Hill Book Co., New York, 1978. International Series in Pure and Applied Mathematics.

[13] BENSOUSSAN, A., LION, J.-L., AND PAPANICOLAU, G. *Asymptotic Analysis for Periodic Structure*. Studies in Mathematics and Its Applications, Vol. 5, North-Holland, Amsterdam, 1978.

[14] BERGH, J., AND LÖFSTRÖM, J. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

[15] BRENNER, S. C., AND SCOTT, L. R. *The mathematical theory of finite element methods*. Springer-Verlag, New York, 1994.

[16] BREZZI, F. Interacting with the subgrid world. In *Numerical analysis 1999 (Dundee)*. Chapman & Hall/CRC, Boca Raton, FL, 2000, pp. 69–82.

[17] BREZZI, F. Recent results in the treatment of subgrid scales. In *Canum 2000, Actes du 32ème Congrès d'Analyse Numérique*, vol. 11. A. Blouza, I. Danaila, P. Joly, S.M. Kaber, B. Lucquin, F. Murat & M. Postel, Éditeurs ESAIM: Proceedings, 2002, pp. 61–84.

[18] BREZZI, F., BRISTEAU, M. O., FRANCA, L. P., MALLET, M., AND ROGÉ, G. A relationship between stabilized finite element methods and the Galerkin method with bubble functions. *Comput. Methods Appl. Mech. Engrg. 96*, 1 (1992), 117–129.

[19] BREZZI, F., AND FORTIN, M. *Mixed and hybrid finite element methods*, vol. 15 of *Springer Series in Computational Mathematics*. Springer-Verlag, New York, 1991.

[20] BREZZI, F., FRANCA, L. P., HUGHES, T. J. R., AND RUSSO, A. $b = \int g$. *Comput. Methods Appl. Mech. Engrg. 145*, 3-4 (1997), 329–339.

[21] BREZZI, F., FRANCA, L. P., HUGHES, T. J. R., AND RUSSO, A. Stabilization techniques and subgrid scales capturing. In *The State of the Art in Numerical Analysis (York, 1996)*. Oxford Univ. Press, New York, 1997, pp. 391–406.

[22] BREZZI, F., FRANCA, L. P., AND RUSSO, A. Further considerations on residual-free bubbles for advective-diffusive equations. *Comput. Methods Appl. Mech. Engrg. 166*, 1-2 (1998), 25–33.

[23] BREZZI, F., HAUKE, G., MARINI, L. D., AND SANGALLI, G. Link-cutting bubbles for the stabilization of convection-diffusion-reaction problems. *Math. Models Methods Appl. Sci. 13*, 3 (2003), 445–461. Dedicated to Jim Douglas, Jr. on the occasion of his 75th birthday.

[24] Brezzi, F., Hughes, T. J. R., Marini, L. D., Russo, A., and Süli, E. *A priori* error analysis of residual-free bubbles for advection-diffusion problems. *SIAM J. Numer. Anal. 36*, 6 (1999), 1933–1948 (electronic).

[25] Brezzi, F., and Marini, D. Augmented spaces, two-level methods and stabilising subgrids. In *Numerical methods for fluid dynamics. VII. Proceedings of the Seventh Conference on Computational Fluid Dynamics held at the University of Oxford, Oxford, March 2001*. Ed. Baines, M. J., The Clarendon Press Oxford University Press, New York, 2001.

[26] Brezzi, F., and Marini, D. Subgrid phenomena and numerical schemes. In *Mathematical modeling and numerical simulation in continuum mechanics (Yamaguchi, 2000)*, vol. 19 of *Lect. Notes Comput. Sci. Eng.* Springer, Berlin, 2002, pp. 73–89.

[27] Brezzi, F., Marini, D., and Russo, A. Applications of the pseudo residual-free bubbles to the stabilization of convection-diffusion problems. *Comput. Methods Appl. Mech. Engrg. 166*, 1-2 (1998), 51–63.

[28] Brezzi, F., Marini, D., and Süli, E. Residual-free bubbles for advection-diffusion problems: the general error analysis. *Numer. Math. 85*, 1 (2000), 31–47.

[29] Brezzi, F., and Russo, A. Choosing bubbles for advection-diffusion problems. *Math. Models Methods Appl. Sci. 4*, 4 (1994), 571–587.

[30] Brooks, A. N., and Hughes, T. J. R. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations. *Comput. Methods Appl. Mech. Engrg. 32*, 1-3 (1982), 199–259. FENOMECH '81, Part I (Stuttgart, 1981).

[31] Canuto, C., Russo, A., and van Kemenade, V. Stabilized spectral methods for the Navier-Stokes equations: residual-free bubbles and preconditioning. *Comput. Methods Appl. Mech. Engrg. 166*, 1-2 (1998), 65–83.

[32] Carstensen, C., and Verfürth, R. Edge residuals dominate *a posteriori* error estimates for low order finite element methods. *SIAM J. Numer. Anal. 36*, 5 (1999), 1571–1587 (electronic).

[33] Ciarlet, P. G. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam, 1978. Studies in Mathematics and its Applications, Vol. 4.

[34] Cioranescu, D., and Donato, P. *An introduction to homogenization*, vol. 17 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press Oxford University Press, New York, 1999.

[35] Clément, P. Approximation by finite element functions using local regularization. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér. 9*, R-2 (1975), 77–84.

[36] DURLOFSKY, L. J. Coarse scale models of two phase flow in heterogeneous reservoirs: volume averaged equations and their relationship to existing upscaling techniques. *Comput. Geosci. 2*, 2 (1998), 73–92.

[37] E, W., AND ENGQUIST, B. The heterogeneous multiscale methods. *Commun. Math. Sci. 1*, 1 (2003), 87–132.

[38] ECKHAUS, W. *Asymptotic analysis of singular perturbations*, vol. 9 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1979.

[39] EFENDIEV, Y. R., HOU, T. Y., AND WU, X.-H. Convergence of a nonconforming multiscale finite element method. *SIAM J. Numer. Anal. 37*, 3 (2000), 888–910 (electronic).

[40] ENGQUIST, B., AND RUNBORG, O. Wavelet-based numerical homogenization with applications. In *Multiscale and multiresolution methods*, vol. 20 of *Lect. Notes Comput. Sci. Eng.* Springer, Berlin, 2002, pp. 97–148.

[41] ERIKSSON, K., ESTEP, D., HANSBO, P., AND JOHNSON, C. Introduction to adaptive methods for differential equations. In *Acta numerica, 1995*, Acta Numer. Cambridge Univ. Press, Cambridge, 1995, pp. 105–158.

[42] ERIKSSON, K., ESTEP, D., HANSBO, P., AND JOHNSON, C. *Computational differential equations*. Cambridge University Press, Cambridge, 1996.

[43] FARMER, C. L. Upscaling: a review. *Internat. J. Numer. Methods Fluids 40*, 1-2 (2002), 63–78.

[44] FORMAGGIA, L., AND PEROTTO, S. New anisotropic *a priori* error estimates. *Numer. Math. 89*, 4 (2001), 641–667.

[45] FRANCA, L. P., FARHAT, C., MACEDO, A. P., AND LESOINNE, M. Residual-free bubbles for the Helmholtz equation. *Internat. J. Numer. Methods Engrg. 40*, 21 (1997), 4003–4009.

[46] FRANCA, L. P., FREY, S. L., AND HUGHES, T. J. R. Stabilized finite element methods. I. Application to the advective-diffusive model. *Comput. Methods Appl. Mech. Engrg. 95*, 2 (1992), 253–276.

[47] FRANCA, L. P., AND MACEDO, A. P. A two-level finite element method and its application to the Helmholtz equation. *Internat. J. Numer. Methods Engrg. 43*, 1 (1998), 23–32.

[48] FRANCA, L. P., NESLITURK, A., AND STYNES, M. On the stability of residual-free bubbles for convection-diffusion problems and their approximation by a two-level finite element method. *Comput. Methods Appl. Mech. Engrg. 166*, 1-2 (1998), 35–49.

[49] FRANCA, L. P., AND RUSSO, A. Approximation of the Stokes problem by residual-free macro bubbles. *East-West J. Numer. Math. 4*, 4 (1996), 265–278.

[50] FRANCA, L. P., AND RUSSO, A. Deriving upwinding, mass lumping and selective reduced integration by residual-free bubbles. *Appl. Math. Lett. 9*, 5 (1996), 83–88.

[51] FRANCA, L. P., AND TOBISKA, L. Stability of the residual free bubble method for bilinear finite elements on rectangular grids. *IMA J. Numer. Anal. 22*, 1 (2002), 73–87.

[52] GARIKIPATI, K., AND HUGHES, T. J. R. The variational multiple scale approach for strain localization problems. In *Computational mechanics (Buenos Aires, 1998)*. Centro Internac. Métodos Numér. Ing., Barcelona, 1998, pp. CD–ROM file.

[53] GEORGOULIS, E. H. *Discontinuous Galerkin methods on Shape–Regular and Anisotropic Meshes*. Oxford University Computing Laboratory D.Phil Thesis, 2003.

[54] GILBARG, D., AND TRUDINGER, N. S. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[55] GILES, M. B., AND SÜLI, E. Adjoint methods for PDEs: *a posteriori* error analysis and postprocessing by duality. *Acta Numer. 11* (2002), 145–236.

[56] GRISVARD, P. *Elliptic problems in nonsmooth domains*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[57] HACKBUSCH, W. *Multigrid methods and applications*, vol. 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1985.

[58] HOLMES, M. H. *Introduction to perturbation methods*, vol. 20 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1995.

[59] HOU, T. Y., AND WU, X.-H. A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys. 134*, 1 (1997), 169–189.

[60] HOU, T. Y., WU, X.-H., AND CAI, Z. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comp. 68*, 227 (1999), 913–943.

[61] HOUSTON, P., RANNACHER, R., AND SÜLI, E. em A posteriori error analysis for stabilised finite element approximations of transport problems. *Comput. Methods Appl. Mech. Engrg. 190*, 11-12 (2000), 1483–1508.

[62] HUGHES, T. J. R. Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Comput. Methods Appl. Mech. Engrg. 127*, 1-4 (1995), 387–401.

[63] HUGHES, T. J. R., AND BROOKS, A. A multidimensional upwind scheme with no crosswind diffusion. In *Finite element methods for convection dominated flows (Papers, Winter Ann. Meeting Amer. Soc. Mech. Engrs., New York, 1979)*, vol. 34 of *AMD*. Amer. Soc. Mech. Engrs. (ASME), New York, 1979, pp. 19–35.

[64] HUGHES, T. J. R., FEIJÓO, G. R., MAZZEI, L., AND QUINCY, J.-B. The variational multiscale method—a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Engrg. 166*, 1-2 (1998), 3–24.

[65] HUGHES, T. J. R., AND STEWART, J. R. A space-time formulation for multiscale phenomena. *J. Comput. Appl. Math. 74*, 1-2 (1996), 217–229.

[66] JOHNSON, C., AND NÄVERT, U. An analysis of some finite element methods for advection-diffusion problems. In *Analytical and numerical approaches to asymptotic problems in analysis (Proc. Conf., Univ. Nijmegen, Nijmegen, 1980)*, vol. 47 of *North-Holland Math. Stud.* North-Holland, Amsterdam, 1981, pp. 99–116.

[67] KUNERT, G. Robust *a posteriori* error estimation for a singularly perturbed reaction-diffusion equation on anisotropic tetrahedral meshes. *Adv. Comput. Math. 15*, 1-4 (2001), 237–259 (2002). *A posteriori* error estimation and adaptive computational methods.

[68] KUNERT, G., AND VERFÜRTH, R. Edge residuals dominate *a posteriori* error estimates for linear finite element methods on anisotropic triangular and tetrahedral meshes. *Numer. Math. 86*, 2 (2000), 283–303.

[69] MADDEN, N., AND STYNES, M. Efficient generation of Shishkin meshes in solving convection-diffusion problems. *Preprint of the Department of Mathematics, University College, Cork, Ireland no. 1995-2* (1995).

[70] MATACHE, A. M., BABUŠKA, I., AND SCHWAB, C. Generalized p-FEM in homogenization. *Numer. Math. 86*, 2 (2000), 319–375.

[71] MATACHE, A.-M., AND SCHWAB, C. Two-scale FEM for homogenization problems. *M2AN Math. Model. Numer. Anal. 36*, 4 (2002), 537–572.

[72] MICHELETTI, S., PEROTTO, S., AND PICASSO, M. Some remarks on the stability coefficients and bubble stabilization of fem on anisotropic meshes. *Technical Report, MOX, Dipartimento di Matematica, Politecnico di Milano, Italy, 2002; submitted to Comput. Visualization Sci.* (2002).

[73] MICHELETTI, S., PEROTTO, S., AND PICASSO, M. Stabilized finite elements on anisotropic meshes: *a priori* error estimates for the advection-diffusion and the Stokes problems. *SIAM J. Numer. Anal. 41*, 3 (2003), 1131–1162 (electronic).

[74] MITCHELL, A. R., AND GRIFFITHS, D. F. Generalised Galerkin methods for second order equations with significant first derivative terms. In *Numerical analysis (Proc. 7th Biennial Conf., Univ. Dundee, Dundee, 1977)*. Springer, Berlin, 1978, pp. 90–104. Lecture Notes in Math., Vol. 630.

[75] MIZUKAMI, A., AND HUGHES, T. J. R. A Petrov-Galerkin finite element method for convection-dominated flows: an accurate upwinding technique for satisfying the maximum principle. *Comput. Methods Appl. Mech. Engrg. 50*, 2 (1985), 181–193.

[76] MORTON, K. W. *Numerical solution of convection-diffusion problems*, vol. 12 of *Applied Mathematics and Mathematical Computation*. Chapman & Hall, London, 1996.

[77] NÄVERT, U. *A Finite Element Method for Convection–Diffusion Problems*. Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, 1982.

[78] O'RIORDAN, E., AND STYNES, M. A globally uniformly convergent finite element method for a singularly perturbed elliptic problem in two dimensions. *Math. Comp. 57*, 195 (1991), 47–62.

[79] PAPASTAVROU, A., AND VERFÜRTH, R. em A posteriori error estimators for stationary convection-diffusion problems: a computational comparison. *Comput. Methods Appl. Mech. Engrg. 189*, 2 (2000), 449–462.

[80] QUARTERONI, A., AND VALLI, A. *Numerical approximation of partial differential equations*. Springer-Verlag, Berlin, 1994.

[81] RISCH, U. Convergence analysis of the residual free bubble method for bilinear elements. *SIAM J. Numer. Anal. 39*, 4 (2001), 1366–1379 (electronic).

[82] ROOS, H.-G., STYNES, M., AND TOBISKA, L. *Numerical Methods for Singularly Perturbed Differential Equations*. Springer-Verlag, Berlin, 1996.

[83] ROSENBERG, I. G., AND STENGER, F. A lower bound on the angles of triangles constructed by bisecting the longest side. *Math. Comp. 29* (1975), 390–395.

[84] RUSSO, A. em A posteriori error estimators via bubble functions. *Math. Models Methods Appl. Sci. 6*, 1 (1996), 33–41.

[85] SANGALLI, G. Global and local error analysis for the residual–free bubbles method applied to advection–dominated problems. *SIAM J. Numer. Anal. 38*, 5 (2000), 1496–1522 (electronic).

[86] SANGALLI, G. A robust *a posteriori* estimator for the residual-free bubbles method applied to advection-diffusion problems. *Numer. Math. 89*, 2 (2001), 379–399.

[87] SANGALLI, G. Capturing small scales in elliptic problems using a residual–free bubbles finite element method. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal 1*, 3 (2003), 485–503.

[88] SANGALLI, G. A discontinuous residual-free bubble method for advection-diffusion problems. *J. Engrg. Math. 49*, 2 (2004), 149–162.

[89] SCHIEWECK, F. *Eine asymptotisch angepaßte Finite–Element–Methode für singulär gestörte elliptische Randwertaufgaben*. Dissertation, TH Magdeburg, 1986.

[90] SCHIEWECK, F. Numerische integration bei der finite-element-diskretisierung singulär gestörten elliptischer randwertaufgaben. *Wiss. Z. Techn. Universität Magdeburg* (1987).

[91] SCHWAB, C. *p- and hp-finite element methods.* The Clarendon Press, Oxford University Press, New York, 1998.

[92] SCHWAB, C., AND MATACHE, A.-M. Generalized FEM for homogenization problems. In *Multiscale and multiresolution methods*, vol. 20 of *Lect. Notes Comput. Sci. Eng.* Springer, Berlin, 2002, pp. 197–237.

[93] SCHWAB, C., AND SURI, M. The p and hp versions of the finite element method for problems with boundary layers. *Math. Comp. 65*, 216 (1996), 1403–1429.

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

[95] VERFÜRTH, R. Error estimates for some quasi-interpolation operators. *M2AN Math. Model. Numer. Anal. 33*, 4 (1999), 695–713.

[96] ZHOU, G. How accurate is the streamline diffusion finite element method? *Math. Comp. 66*, 217 (1997), 31–44.