A class of Petrov-Galerkin finite element methods for the numerical solution of the stationary convection-diffusion equation.

Perella, Andrew James

A Class of Petrov–Galerkin Finite Element Methods for the Numerical Solution of the Stationary Convection–Diffusion Equation.

Andrew James Perella

A thesis submitted in partial fulfilment
of the requirements for the degree of Doctor of Philosophy.

Department of Mathematical Sciences,
University of Durham,
Durham. DH1 3LE.
England.

September, 1996

To my wife Maria, my Mother and Father. Thanks for the support.

Abstract

A Class of Petrov–Galerkin Finite Element Methods
for the Numerical Solution of the
Stationary Convection–Diffusion Equation.

Andrew James Perella

A thesis submitted in partial fulfilment
of the requirements for the degree of Doctor of Philosophy.
September, 1996.

A class of Petrov–Galerkin finite element methods is proposed for the numerical solution of the n dimensional stationary convection–diffusion equation. After an initial review of the literature we describe this class of methods and present both asymptotic and nonasymptotic error analyses. Links are made with the classical Galerkin finite element method and the cell vertex finite volume method. We then present numerical results obtained for a selection of these methods applied to some standard test problems. We also describe extensions of these methods which enable us to solve accurately for derivative values of the solution.

Preface

The work in this thesis is based on research carried out, at Durham University, between October 1992 and October 1995. The material has not been submitted previously for any other degree in this or any other University.

Chapter One is an introduction to the stationary convection–diffusion equation. Chapter Two is a review of other work on the subject of this thesis and no claim of originality is made for these two chapters. Chapters Three to Five describe the class of numerical methods for the stationary convection–diffusion equation and contain theoretical and numerical analyses of them. Much of the work contained in these chapters was presented for the final of the 1995 Bill Morton Computational Fluid Dynamics prize at the 1995 International Computational Fluid Dynamics Conference held at Oxford University and is published in the proceedings of that conference[37]. Chapter Six extends the ideas of the previous chapters by creating difference schemes for boundary value problems which yield exact nodal solutions and also exact nth order derivative nodal values for n no greater than the order of the equation. In Chapter Seven we give conclusions and suggest ideas for future research.

I am extremely grateful to my supervisor, Dr. Alan Craig, for his help and his enthusiastic interest in the research. I am also grateful to Doctors John Coleman, James Blowey and Tony Ware, for words of encouragement

and support. I would like to thank the Engineering and Physical Sciences Research Council (previously the Science and Engineering Research Council) and British Gas plc for the sponsorship I have received.

Statement of Copyright

I declare that no part of this thesis has been submitted for any part of any degree at this or any other university.

Contents

List of Figures

List of Tables

Chapter 1

Stationary Convection–Diffusion Equation

1.1 Introduction

Many physical processes, in particular those arising from fluid flow problems, can be modelled as systems of partial differential equations. Consequently there is a great deal of interest in producing numerical methods which can approximate the solutions of these equations.

However it has been well-known for some time that if there is a dominant convective term in the partial differential equation then standard numerical methods often fail to work well. The rapidly changing solutions which are common features in problems of this type cause standard classical numerical schemes (such as centred finite difference methods and Galerkin finite element methods) to yield solutions which suffer from very large, unrealistic, oscillations. Thus, many numerical methods have been devised to tackle these problems. Of these, the streamline diffusion method [23] and the cell vertex finite volume method [27] have been particularly successful. Although all of the specially designed methods have their strengths none are ideally suited to a large range of problems; indeed they are often designed with view to solving only a small class of problems.

1.2 Convection–Diffusion Equation

The incompressible Navier-Stokes equations [38]:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla P - a\nabla^2 \mathbf{u} = f,$$

$$\nabla \cdot \mathbf{u} = 0,$$

where $\mathbf{u}$ is a velocity field, P is the pressure and $1/a$ is the Reynolds/Péclet number, in conjunction with suitable boundary conditions (such as prescribed boundary velocities or normal derivatives of boundary velocities) model many important fluid-like flow problems (such as transonic airflow around an aircraft). Unfortunately the numerical solution of these equations has proved extremely difficult especially in problems where the Péclet number is very large (typically this can be as large as 10^6 or greater in transonic airflow problems.).

While it is clearly desirable to be able to solve these equations we can have little hope of success over a whole range of problems unless we can adequately solve a simpler linear model. One such linear model is the convection–diffusion equation, an equation interesting in its own right. Like the Navier-Stokes equations the convection-diffusion equation is nonself-adjoint (there is a directionality introduced by the presence of an odd order derivative) and as a result it inherits many of the important properties and difficulties of the more general nonlinear problem.

In this thesis we shall consider new methods for the numerical solution of the stationary n-dimensional convection–diffusion equation. For the purpose of describing and analysing our methods we present the equation only with constant coefficients, although the methods are trivially applicable to the more general case.

The convection-diffusion equation is:

$$-\nabla.(a\nabla u) + \nabla.(\mathbf{b}u) = f \quad \text{in} \quad \Omega \subset R^n, \tag{1.1}$$

where $a(\cdot)$ is a constant $n \times n$ diffusion matrix and $\mathbf{b}(\cdot)$ is a constant $n \times 1$ convection vector, with, for ease, homogeneous Dirichlet boundary conditions.

To gain insight into the mechanisms involved in the convection–diffusion

equation we consider the following one dimensional example [44]:

$$-au''(x) + u'(x) = 1 \quad \text{for } 0 < x < 1,$$

with $u(0) = u(1) = 0$. In particular we are interested in the case when $0 < a << 1$ as it is in this case that standard numerical methods fail to model the solution accurately.

This equation yields the analytical solution

$$\begin{aligned} u(x) &= x + \frac{e^{-1/a} - e^{-(1-x)/a}}{1 - e^{-1/a}} \\ &= x - e^{-(1-x)/a} + O(e^{-1/a}). \end{aligned}$$

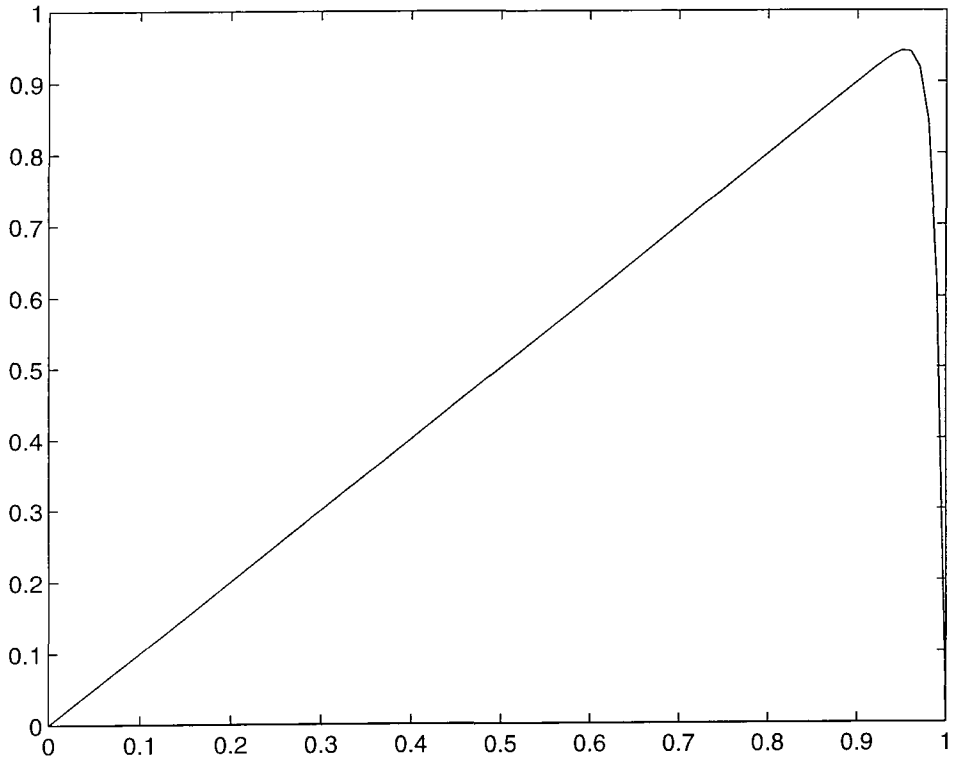


Figure 1.1: Solution plot for $u(x)$ when $a = 1/100$

The first term is the solution of the initial value problem

$$u'(x) = 1 \quad \text{on } (0,1) \text{ with } u(0) = 0.$$

The second term is negligible for small a unless x is near 1. This term enables the second boundary condition to be satisfied. The third term is small.

So we see that for small a the solution essentially satisfies a pure convection problem except in a region close to 1 and we say the solution exhibits an *exponential boundary layer* at $x = 1$.

In higher dimensions, more complicated situations can occur. In particular *internal shear layers* can be present, often due to discontinuous boundary data or to changes in sign of a [23]. These shear layers are characterized by a jump in the solution across the convective flow lines. *Parabolic layers* can also be present at boundaries where he flow is parallel to the boundary. These layers are called 'parabolic' due to the parabolic nature of the differential equation in such regions.

The following subsections explore the existence of solutions to the convection–diffusion equation and also their uniqueness.

1.2.1 Weak Form

We shall introduce the concept of *weak forms* in the following abstract setting as described in [41].

Given a open and bounded region Ω in R^n with polygonal boundary $\Gamma = \Gamma_1 \cup \Gamma_2$ such that Γ_1 and Γ_2 are nonoverlapping, then if L is a second order

is C^n continuous. The advantage of using *weak solutions* if they are unique is that it is easier to prove the existence of weak solutions than the existence of classical solutions. Moreover, since all strong solutions are also weak solutions, if we have a unique weak solution then there must be no more than one strong solution.

1.2.2 Existence and Uniqueness

Theorem 1 *The Generalised Lax-Milgram Theorem [29]*

Suppose that $B(\cdot,\cdot)$ is continuous, coercive, bilinear form on $H_1 \times H_2$, where H_1 and H_2 are real Hilbert spaces. That is there exists two positive constants C_1 and C_2 such that

$$|B(v,w)| \leq C_1||v||_{H_1}||w||_{H_2} \ \forall v \in H_1, \forall w \in H_2,$$

$$\inf_{v\in H_1} \sup_{w\in H_2} \frac{|B(v,w)|}{||v||_{H_1}||w||_{H_2}} \geq C_2$$

and

$$\sup_{v\in H_1} |B(v,w)| > 0 \ \forall w \neq 0.$$

Then $\forall f \in H_2'$ (the dual space of H_2, that is the space of continuous linear functionals on H_2), there exists a unique $u_0 \in H_1$ such that

$$B(u_0,w) = (f,w) \ \forall w \in H_2$$

and

$$||u_0||_{H_1} \leq \frac{1}{C_2}||f||_{H_2'}.$$

Note that

$$||l||_{H_2'} = \sup_{u \in H_2, u \neq 0} \frac{|(l,u)_{H_2}|}{||u||_{H_2}}.$$

Theorem 2 *Existence and Uniqueness of the Convection–Diffusion Equation [29]*

Suppose that for problem (1.2) with

$$Lu \equiv -\nabla.(a\nabla u - \mathbf{b}u)$$

with $a > 0, f \in H^{-1}, \mathbf{b} \in (H^1)^n, \mathbf{n}.\mathbf{b} \geq 0$ on Γ_2 and $\nabla \cdot \mathbf{b} = 0$ (divergence free flow), then a unique *weak* solution u exists and satisfies

$$B(u,v) = (f,v) \; \forall v \in H_0^1.$$

Proof The proof follows from application of the Lax-Milgram theorem above with $H_1 = H_2 = H_0^1$.

1.3 Motivation Behind Numerical Schemes

Many different schemes exist for the numerical solution of the convection–diffusion equation. These are described in the next chapter.

It is apparent that while many of these schemes are very similar, the motivation behind their construction can be very different. Many of these methods can be split into two groups; one in which methods designed to work well for pure diffusion problems are modified (e.g. Galerkin finite element methods / cell–centred finite volume methods / central finite differences) and the other

where methods designed for pure convection problems are modified (Cell–vertex finite volume methods/ upwind differences). A further motivation for most of these methods is the creation of a scheme which is nodally exact for a certain one-dimensional model problem (often where the source function f is zero.). In many cases however these methods are then used for problems bearing little or no resemblance to that original model problem rendering the motivation dubious or fallacious. Other methods have been designed with view to ensuring that they converge in a similar way for all problems ranging from pure diffusion to pure convection problems. These are known as globally uniform convergent schemes.

In this thesis the motivation for the methods described in chapter three is that we aim to produce highly accurate solutions on subregions of one dimension less than the original region. More explicitly, for one dimensional problems we aim to produce accurate solutions at a collection of nodes; in two dimensions we aim to produce accurate solutions on the element boundaries. This is a more realistic aim in n dimensions than producing nodally accurate solutions, as the concept of a nodal value has no meaning for functions in H^1 when $n > 1$. Boundary values, however, are well defined for all n for functions in H^1 via the *trace theorem* [5]. A byproduct of this paradigm is that the quality of the approximation is insensitive to the mesh; a highly sought after quality.

Chapter 2

A Review of Numerical Methods

2.1 Introduction

In this section we give a brief review of other work in this area. For clarity we have grouped methods into various classes. Firstly we describe finite difference and finite volume methods. Then we describe both conforming and nonconforming finite element methods. Many of these methods have been compared in [31]. An entertaining survey paper can be found in [26] which vigorously attacks 'upwind' methods (We shall classify as 'upwind', any method whose difference scheme weights nodes in the upwind direction more than the downwind direction where the direction of the 'wind' is defined by the convective term.). These attacks, although not without some justification, are proved unfounded by the many successes of modern methods (see for example [21]).

2.2 Finite Difference Methods

The method of finite differences aims to generate an approximation U to the solution of the strong form of the equation:

$$-\nabla.(a\nabla u) + \nabla.(\mathbf{b}u) = f \quad \text{in} \quad \Omega \subset R^n, \tag{2.1}$$

at a set of points $x_i \in \Omega$ called nodes. An equation for U_i (U_i is the approximation to $u(x_i)$) is found by replacing each of the derivatives of u by some approximation involving values of U_i. Usually the nodes are arranged in a regular grid pattern for both ease of use and higher accuracy. An equation like this is generated for every node and then the solution is found by solving these equations together simultaneously.

For the purposes of this subsection we consider the following constant co-

efficient one dimensional problem:

$$-au'' + bu' = f, \tag{2.2}$$

posed on $[0,1]$ and discretized by the nodes $x_i = ih$,
$i = 0, ..., n$ where $h = 1/n$ is the constant mesh spacing.

2.2.1 Classical Finite Difference Method

The classical central difference approach is to replace

$$\begin{aligned} u''(x) \quad &\text{by} \quad \frac{U_{i+1} - 2U_i + U_{i-1}}{h^2}, \text{and} \\ u'(x) \quad &\text{by} \quad \frac{U_{i+1} - U_{i-1}}{2h}. \end{aligned}$$

A standard Taylor expansion argument shows that both of these approximations are $O(h^2)$ accurate. From an M-matrix analysis [36], [44] we find that for stability of this scheme it is necessary to have $h < 2a/b$. When $h < 2a/b$ this scheme does indeed produce reasonable results, however when $h >> 2a/b$ as is almost always the case (for computational reasons), the scheme usually produces wildly inaccurate oscillations. Stability can be regained for the method by the use of specially defined meshes (see section 2.2.6.)

2.2.2 Upwind Methods

A common technique employed to overcome the stability problems inherent in classical techniques is that of numerical *upwinding*, so called due to the fact stability can be achieved by taking a one–sided approximation of the first

derivative in the upstream/upwind direction. That is we replace (for the one dimensional case with node x_i downstream from x_{i-1})

$$u''(x) \quad \text{by} \quad \frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} \text{ as before, and}$$
$$u'(x) \quad \text{by} \quad \frac{U_i - U_{i-1}}{h}.$$

Analysis shows that the resulting scheme is stable [44] independent of the mesh spacing. However, by introducing this stability we lose an order of accuracy.

Often a Hybrid Upwind Difference scheme is used where central differences are used for mesh Péclet numbers bh/a less than two, and upwind differencing otherwise.

2.2.3 Exponentially Fitted Schemes

One remedy for the lower accuracy of the upwind method is to still use upwind one-sided differences for the first order term, but to modify the diffusion term by an exponentially fitted parameter which is chosen so that the method yields exact nodal solutions for one dimensional problems without any source functions on the right hand side of the equation. This method suffers in higher dimensions due to increased diffusion perpendicular to the flow direction. This is known as the Allen and Southwell scheme [2]. We also mention here, fitted methods [22] on arbitrary meshes (in one dimension) where the coefficients of all the terms in the difference scheme are chosen to ensure uniform convergence in some appropriate norm with respect to the diffusion parameter.

2.2.4 Interpolation/Extrapolation Techniques

Many techniques exist, where the convective term is obtained by interpolation of a given function through a number of upwind and downwind nodes. The diffusion term is usually treated by standard central differences. Examples of such schemes are the QUICK scheme [42] and the LUE/LLUE schemes [42]. The quadratic upwind interpolations (QUICK) scheme employs a quadratic variation fitted to the nodal values of the two closest upwind nodes and the closest downwind node. In the linear upwind extrapolation (LUE) scheme, nodal solution values are found from the linear function fitted to the two upwind nodes. Note that the standard upwind scheme also falls into this category.

2.2.5 Artificial Diffusion

The artificial diffusion approach is to solve a modified equation by the classical finite difference approach. Here an extra diffusion term $(h/2)u''$ is added onto the convection-diffusion operator. For the one–dimensional problem, the resulting matrix equation is identical to that of the upwind scheme. Due to this order h perturbation this is limited only to first order accuracy. This approach can be extended more easily than the upwind method to higher dimensions where the extra diffusion can be added only in the direction of the streamlines. This can be derived as a finite element method within whose framework it is possible to prove greater than first order accuracy; See section 2.5.2.

2.2.6 Specially Designed Meshes

It has been shown (see for example the references in [6]) that success can be found in using standard finite difference techniques by applying them on specially designed meshes. Notable success has been found in using piecewise uniform meshes [12]. Also the methods of exponential fitting can be applied in conjunction with specially designed meshes. All these methods are designed with a view to producing uniform convergence in some norm (usually the discrete L^∞ norm, where $||f||_{L^\infty} = \sup_{a<x<b} |f(x)|$) with respect to the diffusion parameter.

2.3 Finite Volume Methods

Finite Volume methods [27] have been highly successful in the numerical solution of partial differential equations, particularly in the solution of conservation laws such as those that occur in fluid mechanics. These methods, broadly speaking, can be divided into two types: cell-centred and cell–vertex methods. These methods try to solve the equation in conservation form - that is they approximate the solution of

$$\int \nabla.(a\nabla u + \mathbf{b}u)\ d\Omega = \int f\ d\Omega.$$

To solve this, finite volume methods split the domain Ω into subregions Ω_i (intervals in one dimension , quadrilaterals in two dimensions) where $\bigcup \Omega_i = \Omega$. The integral equation is then posed on each of these domains (often called cells or finite volumes) individually and these equations are then solved by using Gauss' theorem to convert the integrals to surface integrals. It is in the approximation of u on these surfaces that differences appear between cell-

centred and cell-vertex methods.

2.3.1 Cell Centred Finite Volume Method

In the cell-centred finite volume method values for the approximate solution u_h are held at the centres of these cells. Values of u_h on the surfaces are approximated by interpolation between these cell-centred values. Although this method is highly suited to the solution of pure diffusion equations as there is no directionality expressed, it has also been used successfully for convection–diffusion problems.

2.3.2 Cell Vertex Finite Volume Method

Cell-vertex methods are particular useful for highly convective problems (such as the convection–diffusion equation with low diffusion) and so are more relevant to the content of this thesis. The cell-vertex finite volume method stores unknowns at the cell-vertices. The surface integrals are then calculated numerically by the trapezium rule for the approximation of u and by some suitable difference scheme for the approximation of ∇u. The cell vertex finite volume method for a pure convective problem (so no approximation for ∇u is necessary) can be viewed as a Petrov–Galerkin finite element method [34]. Petrov Galerkin methods are described later in this chapter.

Unlike cell-centred schemes the cell-vertex method suffers from 'counting' problems in that the number of unknowns will not, in general, match the number of cells (there will be one equation per cell). To overcome this problem either selected cells must be left out of the integration or some cells must be split. See [7] for more details.

2.4 Conforming Finite Element Methods

The standard weak form of the convection–diffusion equation is:

$$\text{find } u \in H_0^1(\Omega) \quad : \quad \mathbf{a}(u,v) + \mathbf{c}(u,v) = (f,v) \quad \forall\, v \in H_0^1(\Omega), \tag{2.3}$$

$$\begin{aligned} \text{where} \quad \mathbf{a}(u,v) &= \int_\Omega a \nabla u . \nabla v \; d\Omega, \\ \mathbf{c}(u,v) &= \int_\Omega \nabla . (\mathbf{b}u) v \; d\Omega, \\ (f,v) &= \int_\Omega f v \; d\Omega, \end{aligned}$$

and $H_0^1(\Omega)$ is the usual Sobolev space.

The Petrov-Galerkin method consists of choosing two finite dimensional spaces $\mathcal{V}, \mathcal{W} \subset H_0^1(\Omega)$ such that $\dim(\mathcal{V})$=$\dim(\mathcal{W})$, and solving

$$\text{find } U \in \mathcal{V} \text{ such that } \mathbf{a}(U,w) + \mathbf{c}(U,w) = (f,w) \; \forall\, w \in \mathcal{W}. \tag{2.4}$$

The space $\mathcal{V}$ is known as the *trial* space and $\mathcal{W}$ is known as the *test* space. The *finite element* Petrov-Galerkin method consists of producing $\mathcal{V}$ and $\mathcal{W}$ to contain piecewise functions (usually continuous piecewise polynomials) defined over a mesh.

These methods are referred to as 'conforming' finite elements as $\mathcal{V}, \mathcal{W} \subset H_0^1(\Omega)$.

For a good review and a unifying approach to finite element methods for convection-diffusion problems see [30].

2.4.1 Galerkin Finite Element Method

The Petrov–Galerkin finite element method reduces to the Galerkin finite element method when $\mathcal{V} = \mathcal{W}$. When a piecewise n-linear trial and test space is used (in n dimensions), this creates a difference scheme which is qualitatively similar to that of central differences (although more points in the stencil are used) and so leads to oscillatory solutions in cases of low diffusion.

The scheme reduces to finding $U \in \mathcal{V}$ such that

$$B(U, v) = (f, v) \; \forall v \in \mathcal{V},$$

where $B(v, w) = \mathbf{a}(v, w) + \mathbf{c}(v, w)$.

Defining the norm

$$||v||_{B_1} = (\mathbf{a}(v, v))^{\frac{1}{2}},$$

then for the case $\mathbf{b} = \mathbf{0}$, the approximate solution satisfies

$$||u - U||_{B_1} = \inf_{v \in \mathcal{V}} ||u - v||_{B_1}.$$

This optimal approximation property of the Galerkin method does not follow however in the case of non zero convection. For this case the error bound becomes

$$||u - U||_{B_1} = (1 + C_1/C_2) \inf_{v \in \mathcal{V}} ||u - v||_{B_1},$$

where $C_1 = a + ||\mathbf{b}||_{(L^\infty)^n}$ and $C_2 = aC(\Omega)$ (see [32]) are the constants in the Lax-Milgram theorem. By making standard assumptions about the approximation properties of the trial space, the constant of optimality $(1 + C_1/C_2)$ can be refined (see [32]) to

$$(1 + ||\mathbf{b}||_{(L^\infty)^n} h/a),$$

which depends on the size of the *mesh* Péclet number.

2.4.2 Liouville Transform

The one dimensional convection–diffusion equation can be symmetrized by the transform [29]

$$w(x) = \exp(-bx/(2a))u(x)$$

to remove the first order derivative term. The weak form can then be solved efficiently by the Galerkin finite element method. The problem with this seemingly powerful approach is that transforming back to the original variable is ill-conditioned. Any errors will be amplified by $e^{\frac{bx}{2a}}$. Another problem of this approach is that the method produces best approximations in a weighted H^1 norm where the weighting is such that the approximate solution is more accurate upstream (where the solution should be accurate anyway as it is here that Dirichlet data is prescribed.) and less accurate downstream.

2.4.3 Exponential Trial/Test Space Methods

The exponential trial space method [29] is based on choosing a trial space in a Petrov-Galerkin finite element method so that the solution is exact for a model problem with nonzero Dirichlet data and no forcing function. Either the same space as used for the trial space is used for the test space (in which case we have a Galerkin method) or a standard piecewise linear test space can be used [11], [14] and [28]. The exponential test space method [17] relies on using a test space which contains the global Green's functions associated with each node. In one dimension these global functions can be split up into a sum of local exponential basis functions. This one dimensional method yields the

exact solution at the nodes.

An extension of these methods are the LAM (localised adjoint method) and ELLAM (Eulerian-Lagrangian localised adjoint method) [39] where the test function associated with node i is chosen to satisfy the homogeneous adjoint equation $L^*u = 0$ locally over each element. Note that the Green's function associated with the point x_i satisfies this equation globally except at the point x_i. The motivation for this comes directly form the one dimensional use of Green's functions but here the adjoint equation is allowed to be violated at more points than just the nodes. More specifically the equation is violated on all the element boundaries.

The methods that we describe in this thesis fall into this category although the motivation is very different.

2.4.4 Artificial Diffusion

The method of artificial diffusion can be used in conjunction with the Galerkin finite element method, where the diffusion parameter a is replaced by h whenever $a < h$. This has the obvious consequence of introducing extra diffusion which 'smears out' any sharp fronts in the solution. This method is at best first order accurate due to this order h perturbation to the original problem.

2.4.5 Polynomial Upwinding

Many Petrov-Galerkin methods have been produced which use as a basis for the test space, functions which are polynomial perturbations of the standard linear trial basis functions on each element. The size of perturbation can

be chosen so that the resulting difference operator becomes the Allen and Southwell difference operator [2] in one dimension. These are discussed in [3], [4], [8], [15], [16] and [47]. It has recently, however, been shown that simply extending these schemes to higher dimensions by using tensor products of the one dimensional schemes does not necessarily work well, especially when the convective field is variable [32].

As a one dimensional example we consider the following constant coefficient one dimensional problem:

$$-au'' + bu' = f, \tag{2.5}$$

posed on $[-1, 1]$ and discretized by the nodes $x_i = ih - 1$,
$i = 0, ..., n$ where $h = 2/n$ is the constant mesh spacing.

In figures 2.1 and 2.2 we show the test basis function $\psi(x)$ as used in [15] generated by a quadratic perturbation $\alpha\sigma_i(x)$ of the linear hat function $\phi_i(x)$.

$$\phi_i(x) = \begin{cases} (x - x_{i-1})/h \ , -1 \leq x \leq 0. \\ (x_{i+1} - x)/h \ , 0 < x \leq 1. \end{cases}$$

$$\sigma_i(x) = \begin{cases} -3(x - x_{i-1})(x_i - x)/h^2 \ , -1 \leq x \leq 0. \\ -3(x_{i+1} - x)(x - x_i)/h^2 \ , 0 < x \leq 1. \end{cases}$$

$$\alpha = \coth(bh/2a) - (2a/bh).$$

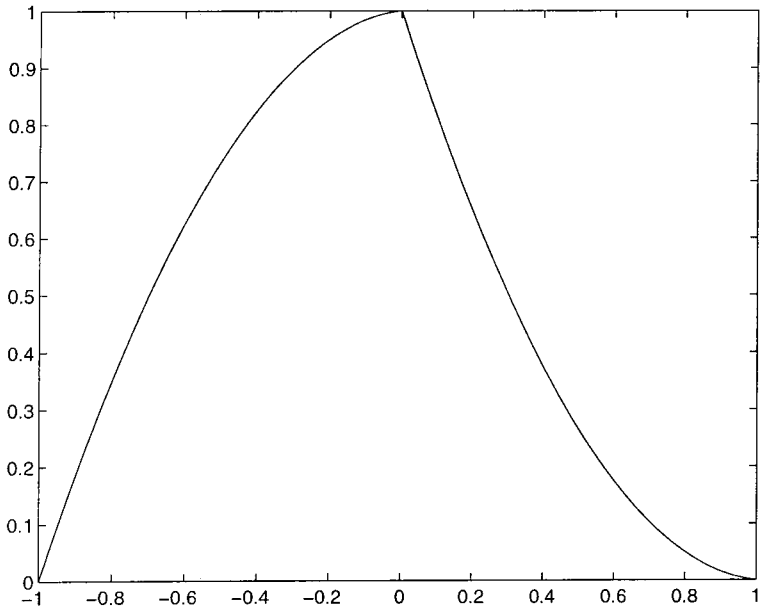


Figure 2.1: Quadratic test function on [-1,1] ($a = 1, b = 2, h = 1$)

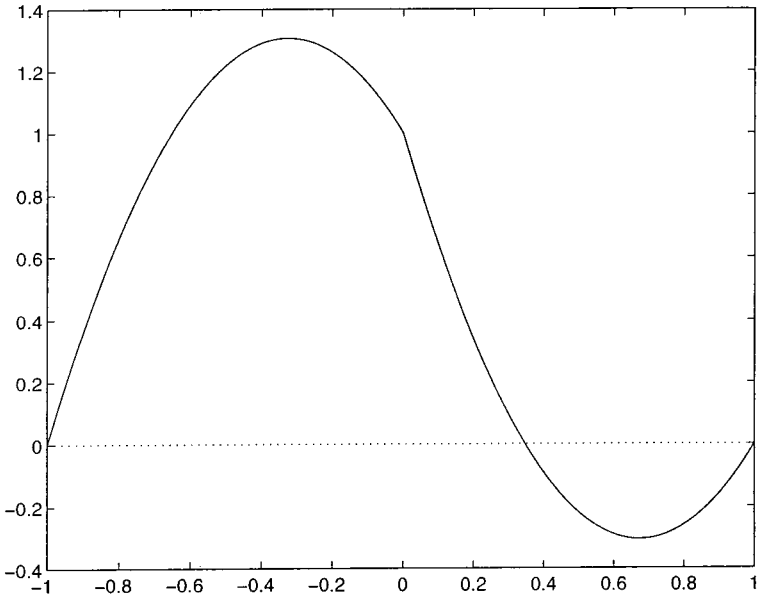


Figure 2.2: Quadratic test function on [-1,1] ($a = 1, b = 50, h = 1$)

2.5 Nonconforming Finite Element Methods

Nonconforming finite element methods are similar to the conforming methods described above, except the trial and test spaces are not restricted to being subspaces of $H_0^1(\Omega)$. Because of this, certain terms on the weak form do not mean anything in a strict mathematical sense. If we ignore such 'nonconforming' terms however the resulting numerical scheme can converge.

2.5.1 Exponential Fitting On Triangles

Recently a Galerkin finite element method has been developed [9] for the two dimensional convection–diffusion equation which uses a trial (and test) space based on a triangularisation of the domain. The trial functions are designed so that in each triangle they give the exact solution of the convection–diffusion equation with suitable boundary conditions. Due to the boundary conditions imposed, these functions are not continuous over element boundaries and so the method is nonconforming. The scheme has been modified in [40] to a nonconforming Petrov-Galerkin method by replacing the exponentially fitted test functions by the usual linear functions.

2.5.2 Streamline Diffusion Method

This is essentially an extension of the artificial diffusion idea. Here, extra diffusion is added only in the streamline direction, and so introduces much less crosswind diffusion. However this is still an order h perturbation of the original equation.

It is possible to generate this extra term without introducing an order h perturbation in the following way. We take piecewise n-linear trial space $\mathcal{V}$, but use a test space $\mathcal{W}$ consisting of test basis functions $v + \alpha \mathbf{b}.\nabla v$ corresponding to trial basis functions $v \in \mathcal{V}$. Then we have the following nonconforming Petrov Galerkin finite element method:

Find $U \in \mathcal{V}$ such that

$$-a\alpha(\nabla^2 U, \mathbf{b}.\nabla v) + a(\nabla U, \nabla v) + (\mathbf{b}.\nabla U, v + \alpha \mathbf{b}.\nabla v) = (f, v + \alpha \mathbf{b}.\nabla v) \; \forall v \in \mathcal{V}.$$

Often α is chosen so that $\alpha = 0, a \geq h$ and $\alpha = Ch, a < h$, where h is the mesh spacing and C is some sufficiently small constant to be chosen.

This is nonconforming as the test functions are not in H^1. This adds a term $-a\alpha(\nabla^2 U, \mathbf{b}.\nabla v)$ which has no meaning in a strict mathematical sense and is ignored. (The order of this term is much less than $O(h^2)$ [24]).

As a one dimensional example we consider the following constant coefficient one dimensional problem:

$$-au'' + bu' = f, \tag{2.6}$$

posed on $[-1, 1]$ and discretized by the nodes $x_i = ih - 1$,
$i = 0, ..., n$ where $h = 2/n$ is the constant mesh spacing.

In figures 2.3 and 2.4 we show the test basis functions for this problem.

2.6 Transient Methods

Another, common approach to the solution of steady–state problems is to integrate a spatially discretised transient equation (with a u_t term) to steady

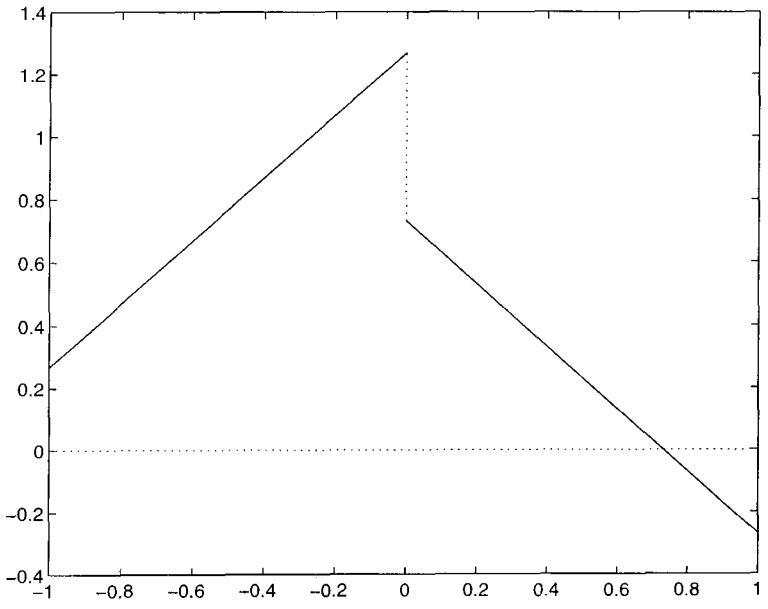


Figure 2.3: Streamline diffusion test function on [-1,1] ($a = 1, b = 4, h = 1$)

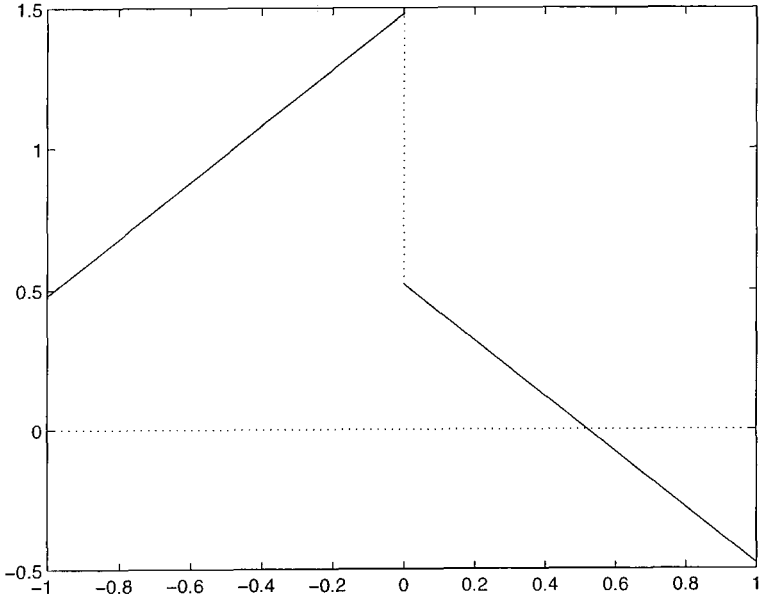


Figure 2.4: Streamline diffusion test function on [-1,1] ($a = 1, b = 50, h = 1$)

state. As we are only concerned with the steady–state solution, the accuracy of the time stepping is not too important as long as convergence to steady state is achieved. Often methods which do not perform well on steady state problems can still yield good results when treated in this manner.

2.7 Summary

Although finite difference methods are simple to implement, difficulties arise in treatment of the boundary conditions. These difficulties are not present in the finite element method where boundary conditions are treated in a natural manner. Although finite volume methods combine the advantages of both finite difference methods and finite element methods, the 'counting' problems can cause considerable difficulty. Clearly there is much greater cost in using a transient method for steady state problems. Many of these methods (exponential trial and test space methods, polynomial upwinding, streamline diffusion) are designed to reproduce the same difference operator for the one dimensional convection–diffusion operator as the Allen and Southwell scheme [2]. However, these methods do differ in how they treat the source function f. It is clear that these differences are very important and that it is perhaps naive to expect a method designed to perform well on a problem with $f = 0$ will work well on problems with nonzero f. It is extremely important that the source function f is sampled by the numerical scheme sufficiently well.

Chapter 3

A Class Of Petrov-Galerkin Finite Element Methods

3.1 Introduction

In this chapter we shall describe a class of Petrov–Galerkin Finite Element methods for the convection-diffusion equation, as described in section 2.4, posed on arbitrary polygonal domains. We shall note that this class contains the standard Galerkin Finite Element method in the case of pure diffusion, and the Cell Vertex Finite Volume method in the pure convective case so long as the flow is not directed along a mesh boundary. We shall give example of methods from this class and some classical theoretical results.

3.2 A Class Of Methods

3.2.1 Definition of the Test Space

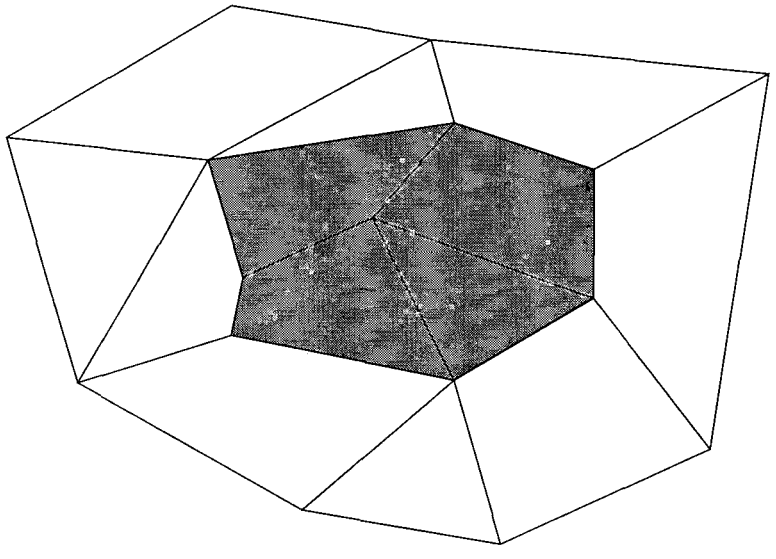


Figure 3.1: Section of mesh

Definition 3 With each node i we associate a function w_i that has the properties

i) w_i is continuous on Ω,

ii) $w_i = 1$ at node i,

iii) $w_i = 0$ on all elements not meeting at node i.

iv) within each element having node i as a vertex, we have

$$-\nabla(a\nabla w_i) - \mathbf{b}.\nabla w_i = 0,$$

(that is w_i is a solution of the homogeneous adjoint equations inside each element).

Remark 4 Functions w_i exist since the adjoint equation has solutions if (1.1) has a solution [10]. Note that the above definition does not uniquely define w_i, because we have not defined its values on all element boundaries.

Definition 5 Define Γ_i as the boundary of Ω_i, where Ω_i denotes element i. Define $\mathbf{n}_i$ as the outward unit normal vector on Γ_i.

Theorem 6 Given a finite element mesh define a continuous trial space $\mathcal{V}$ and a test space $\mathcal{W} = \text{span}\{w_i\}$ where the w_i belong to the class defined above, and with $\dim(\mathcal{V})=\dim(\mathcal{W})$. Then

$$\sum_{elements} \int_{\Gamma_i} (u - U)\nabla w.\mathbf{n}_i \, d\Gamma_i = 0 \;\forall\; w \in \mathcal{W}, \tag{3.1}$$

where the integration, for example, is taken in a counter-clockwise direction.

Remark 7 Theorem 6 says that the error on the boundaries projected (in the L^2 inner product) onto the jumps in the derivatives of the test functions is zero. Note, in particular that in one dimension the nodal solution is independent of the (continuous) trial space $\mathcal{V}$.

Proof of theorem 6 Setting $v = w$ in equation (2.3) and subtracting equation (2.4) gives

$$\mathbf{a}(u - U, w) + \mathbf{c}(u - U, w) = 0 \ \forall \ w \in \mathcal{W}.$$

We then break the integrals up into integrals over elements and write the bilinear forms explicitly:

$$\sum_i \int_{\Omega_i} a\nabla(u - U).\nabla w + \nabla.(\mathbf{b}(u - U))w \ d\Omega_i \ = \ 0 \ \forall \ w \in \mathcal{W},$$

and integrating by parts we obtain

$$\sum_i \int_{\Gamma_i} a(u - U)\nabla w.\mathbf{n}_i \ d\Gamma_i + \sum_i \int_{\Gamma_i} (u - U)w\mathbf{b}.\mathbf{n}_i \ d\Gamma_i - \sum_i \int_{\Omega_i} (u - U)(\nabla.(a\nabla w) + \mathbf{b}.\nabla w) \ d\Omega_i \ = \ 0 \ \forall \ w \in \mathcal{W}.$$

Use of property (iv) of definition 3 and noting that $(u - U)w$ is continuous with $w = 0$ on the boundary of Ω, completes the proof.

Assuming uniqueness of the approximate solution we can make the following statement:

Corollary 8 If the trial space contains a function v such that $u = v$ on the element boundaries, then we obtain the exact solution on the element boundaries. This result also follows from the non-asymptotic error estimate in chapter 4.

3.3 One Dimension

Theorem 6 contains the result that in one dimension we reproduce the exact solution at the nodes. The proof is in some sense less elegant than the standard Green's function approach in one dimension but is, we believe, the natural generalisation of that result. Using Green's functions in higher dimensions (to try to achieve nodal, rather than boundary accuracy) is not viable due to their singular nature. For the Green's function approach it is necessary to show that the local Green's functions are decompositions of the global Green's functions. In this method we need only to calculate the local adjoint solutions.

Remark 9 It is interesting to note that in one dimension the nodal solution values are independent of the trial space and in n dimensions the quality of the boundary solution depends more strongly on the test space than on the trial space. Note that we always assume that the trial space is continuous.

The following subsections describe methods of this type and also consider the effect of introducing a zeroth order term into the differential equation.

3.3.1 Linear Trial Space/Exponential Test Space

We need to solve the constant coefficient convection diffusion equation

$$-au'' + bu' = f \ in \ [0,1], \tag{3.2}$$

with some boundary conditions.

Theorem 6 indicates that we should use test functions w that satisfy the homogeneous adjoint equation,

$$-aw'' - bw' = 0, \tag{3.3}$$

on each element. So for a regular mesh, for example, in the elements $[(i-1)h, ih]$ and $[ih, (i+1)h]$ we impose

$$w((i-1)h) = w((i+1)h) = 0, w(ih) = 1$$

and solve in each element. This gives unique w for this one-dimensional case, a property that is not present in higher dimensions. Use of these functions as a basis for $\mathcal{W}$ gives the exact solution at the nodes.

3.3.2 Quadratic Trial Space/Exponential Test Space

Theorem 6 also suggests that, if for example we use a piecewise quadratic trial space, we might still use a piecewise linear test space (see fig. 3.2) for Poisson's equation. The standard quadratic Galerkin method in one dimension will give exact solutions to Poisson's equation on the element boundaries. This Petrov-Galerkin method will, additionally, give exact solutions at the internal

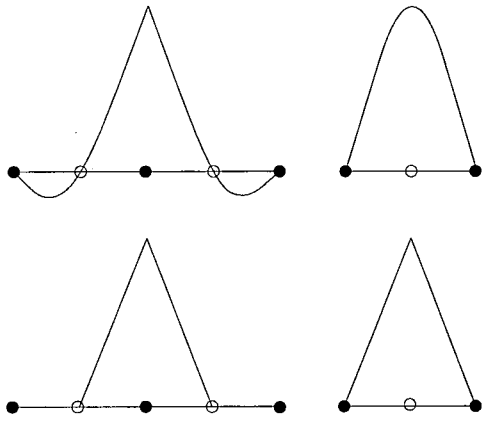


Figure 3.2: Quadratic trial space/linear test space

nodes. (Similarly for the convection-diffusion equation we can use a piecewise exponential test space with higher order trial spaces).

Remarkably the matrix resulting from discretisation in this way is still symmetric despite the trial and test spaces being different. This result can be generalised to the following theorem.

Theorem 10 If we discretise the equation

$$-au'' + cu = f$$

by a Petrov-Galerkin finite element method where the basis functions of both the test and the trial space are themselves symmetric then the matrix resulting from the discretisation is symmetric.

Proof The contribution from the term cu is clearly symmetric, so it suffices to consider the contribution from the au'' term.

Let $g(x)$ denote our test basis function centred at $x = h$ with compact support of width $2h$. Let $v(x)$ denote our trial basis function centred at $x = \lambda$ with compact support of width 2λ.

Let

$$g_1(x) = g(x), \quad g_2(x) = g(x+h) = g_1(h-x)$$

and

$$v_1(x) = v(x), \quad v_2(x) = v(x+\lambda) = v_1(\lambda - x).$$

For the matrix to be symmetric we need to show that $A = B$ and $C = D$ where

$$\begin{aligned}
A &= \int_0^h g_1'(x)v_1'(x)\,dx + \int_h^{2h} g_2'(x-h)v_1'(x)\,dx, \\
B &= \int_{\lambda-2h}^{\lambda-h} g_1'(x-(\lambda-2h))v_2'(x)\,dx + \int_{\lambda-h}^{\lambda} g_2'(x-(\lambda-h))v_2'(x)\,dx, \\
C &= \int_0^h g_2'(x)v_1'(x)\,dx, \\
D &= \int_{\lambda-h}^{\lambda} g_1'(x-(\lambda-h))v_2'(x)\,dx.
\end{aligned}$$

Proof that $A = B$ and $C = D$ follows immediately by the change of variable $y = \lambda - x$ and using the relationships $g_1(x) = g_2(h-x)$ and $v_1(x) = v_2(\lambda - x)$.

3.3.3 Effect of Introducing a Zero Order Term

It is sometimes necessary to solve a problem with a zero order term. If, for example, we need to solve the constant coefficient equation,

$$-au'' + bu' + cu = f \ in \ [0,1], \tag{3.4}$$

with some boundary conditions then we can proceed, as before, by using test functions w that satisfy the homogeneous adjoint equation,

$$-aw'' - bw' + cw = 0, \tag{3.5}$$

on each element. So for a regular mesh, for example, in the elements $[(i-1)h, ih]$ and $[ih, (i+1)h]$ we impose

$$w((i-1)h) = w((i+1)h) = 0, w(ih) = 1$$

and solve in each element.

Use of these functions as a basis for $\mathcal{W}$ will give the exact solution at the nodes.

In figures 3.3 to 3.14 we show the test functions with $a = 1, h = 1$ on $[-1, 1]$.

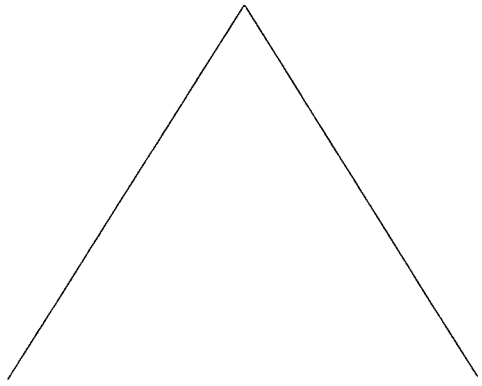


Figure 3.3: $b = 0, c = 0$

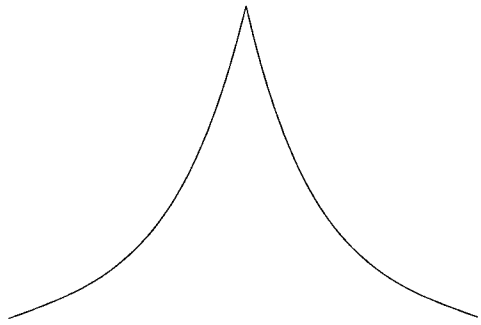


Figure 3.4: $b = 0, c = 10$

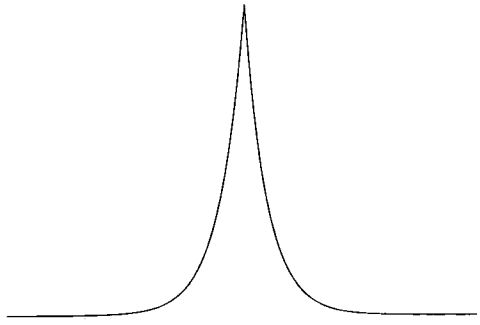


Figure 3.5: $b = 0, c = 100$

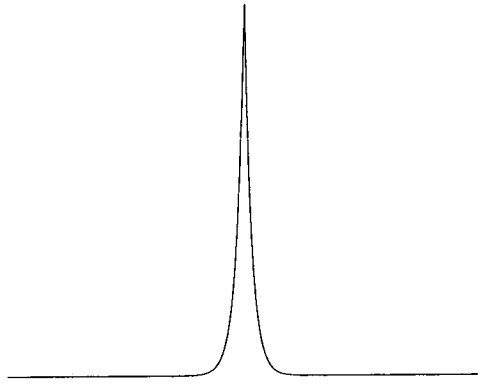


Figure 3.6: $b = 0, c = 1000$

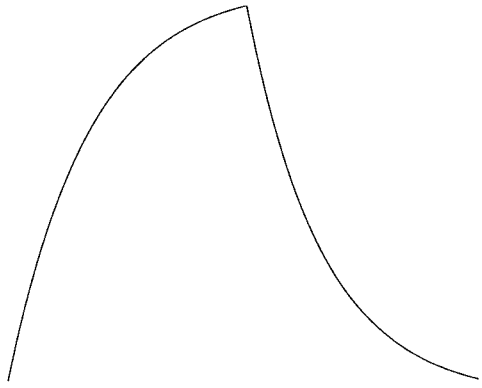


Figure 3.7: $b = 3, c = 0$

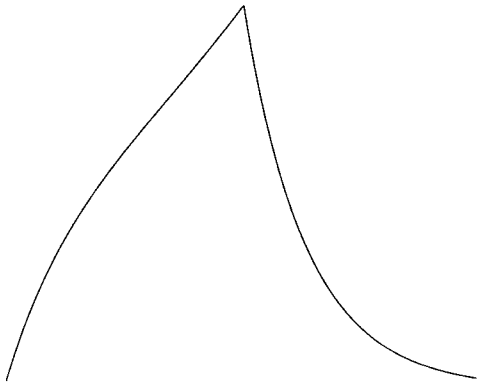


Figure 3.8: $b = 3, c = 3$

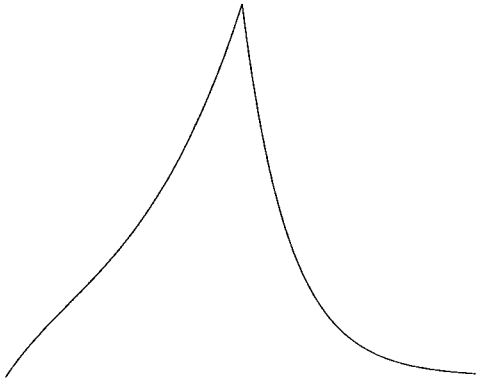


Figure 3.9: $b = 3, c = 10$

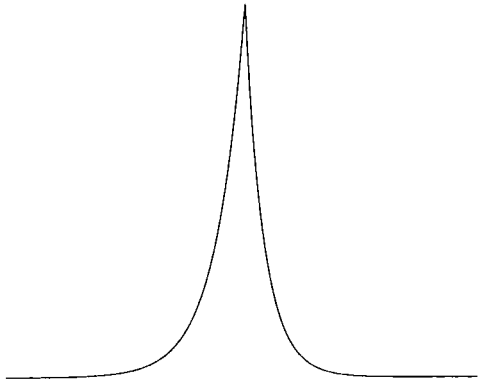


Figure 3.10: $b = 3, c = 100$

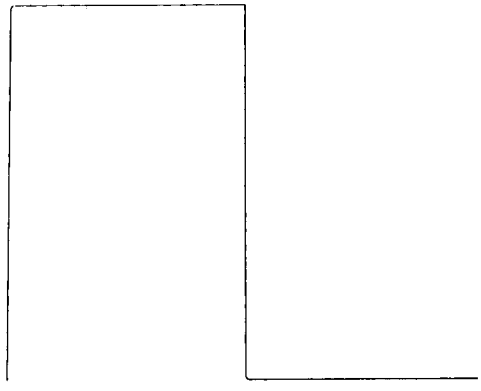


Figure 3.11: $b = 500, c = 0$

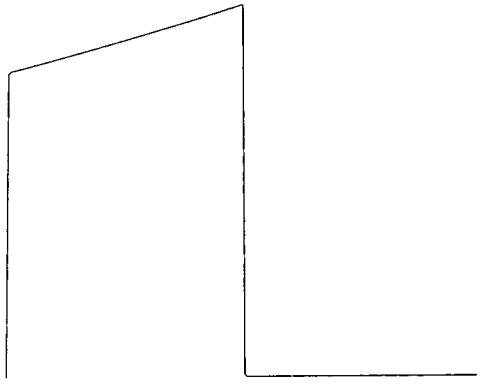


Figure 3.12: $b = 500, c = 100$

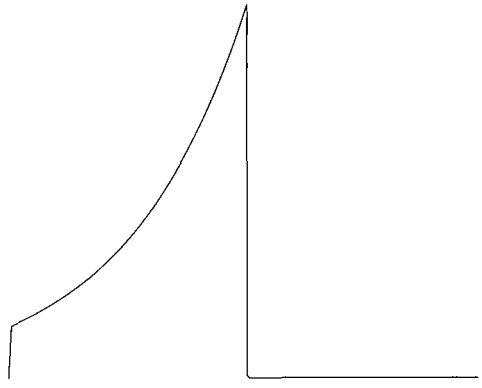


Figure 3.13: $b = 500, c = 1000$

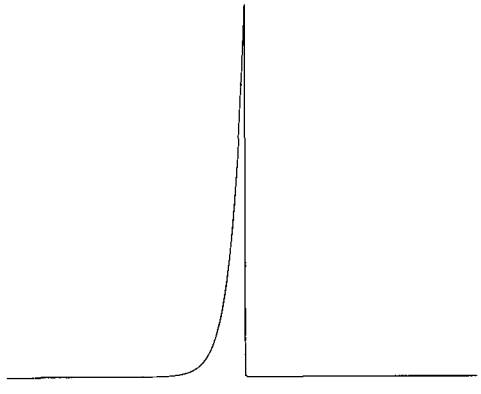


Figure 3.14: $b = 500, c = 10000$

3.4 Two Dimensions

Example 11

This example demonstrates that an exact solution can be obtained on the mesh lines with piecewise linear approximation if the solution is piecewise linear over the mesh lines. Take:

$$-\nabla^2 u + \nabla.(\mathbf{b}u) = f \text{ in } \Omega, \tag{3.6}$$

$$u = 0 \text{ on } \Gamma, \tag{3.7}$$

where $f(x, y; \mathbf{b})$ is chosen to give an exact solution of

$$u = \sin(\pi x)\sin(\pi y)$$

for arbitrary nonzero constant $\mathbf{b}$ and $\Omega = [-1, 1] \times [-1, 1]$. Note that u is zero on $x = 0$ and $y = 0$. We partition Ω into 4 elements as in figure 3.15.

We need to find our test functions as solutions of

$$-\nabla^2 w - \mathbf{b}.\nabla w = 0 \text{ on } \Omega_i, \tag{3.8}$$

with $w(0, 0) = 1$ and $w(x, y) = 0$ a.e. for (x, y) on Γ. There are infinitely many solutions, but we can, for example, choose one of the family of separable solutions as tensor products of the one-dimensional solutions. For example the test function has the form $w(x, y) = X(x)Y(y)$ over $[-1, 1] \times [-1, 1]$, where

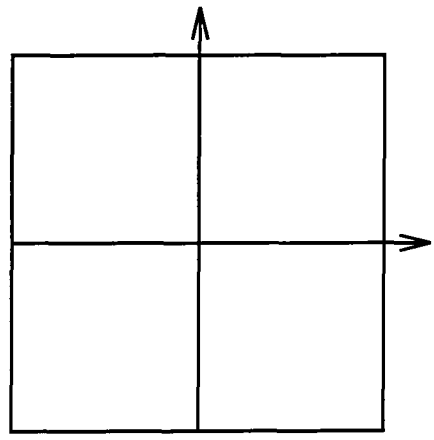


Figure 3.15: Mesh for example in two dimensions

$$X(x) = \begin{cases} \frac{\exp(-b_1 x) - \exp(b_1)}{1 - \exp(b_1)} & -1 \leq x \leq 0 \\ \frac{\exp(-b_1 x) - \exp(-b_1)}{1 - \exp(-b_1)} & 0 \leq x \leq 1. \end{cases}$$

$$Y(y) = \begin{cases} \frac{\exp(-b_2 y) - \exp(b_2)}{1 - \exp(b_2)}, & -1 \leq y \leq 0 \\ \frac{\exp(-b_1 y) - \exp(-b_2)}{1 - \exp(-b_2)} & 0 \leq y \leq 1. \end{cases}$$

Then we have,

$$\begin{aligned} (f, w) \;=\; & \int_{-1}^{1} \pi X(x) \sin(\pi x)\, dx \int_{-1}^{1} Y(y)(\pi \sin(\pi y) + b_2 \cos(\pi y))\, dy + \\ & \int_{-1}^{1} \pi Y(y) \sin(\pi y)\, dy \int_{-1}^{1} X(x)(\pi \sin(\pi x) + b_1 \cos(\pi x))\, dx. \end{aligned}$$

Let,

$$\begin{aligned} C_b^a &= \int_b^a (\pi \sin(\pi x) + b_1 \cos(\pi x))\, dx \\ &= [-\frac{1}{\pi}(\pi \cos(\pi x) - b_1 \sin(\pi x))]_b^a, \\ E_b^a &= \int_b^a (\pi \sin(\pi x) + b_1 \cos(\pi x)) e^{-b_1 x}\, dx \\ &= [-\cos(\pi x) e^{-b_1 x}]_b^a. \end{aligned}$$

Hence

$$\begin{aligned} \int_{-1}^{1} X(x)(\pi \sin(\pi x) + b_1 \cos(\pi x))\, dx &= E_{-1}^0 . \frac{1}{1-e^{b_1}} + E_0^1 . \frac{1}{1-e^{-b_1}} \\ &\quad + C_{-1}^0 . \frac{-e^{b_1}}{1-e^{b_1}} + C_0^1 . \frac{-e^{-b_1}}{1-e^{-b_1}} \\ &= -(1+e^{b_1}) . \frac{1}{1-e^{b_1}} \\ &\quad + (1+e^{-b_1}) . \frac{1}{1-e^{-b_1}} \\ &\quad - 2 . \frac{-e^{b_1}}{1-e^{b_1}} \\ &\quad + 2 . \frac{-e^{-b_1}}{1-e^{-b_1}} \\ &= 0. \end{aligned}$$

A similar result clearly holds for the term involving $Y(y)$. So we have that

$$(f, w) = 0$$

and hence our approximation $U = 0$. This is the exact solution on the mesh boundaries. A similar scheme (although considered as a difference scheme) has been considered by Hegarty, O'Riordan and Stynes [14] using a discrete H^1 norm. This method, motivated by the exponential nature of the solution, uses an exponential trial rather than test space. They then suggest, from experience

with one-dimensional problems, that one should use the scheme presented here. Our motivation for using this scheme is very different however.

3.4.1 Other Possible Schemes

Due to the infinity of possible test spaces (for $n > 1$), that fit into this framework, a choice has to be made as to which one to use. So far we have only described the scheme based on tensor products of the one dimensional test space. However another obvious choice is one of the other separable solutions of the adjoint equation.

So, for example in two dimensions, if our test functions are to satisfy

$$-a\nabla^2 w - \mathbf{b}.\nabla w = 0 \text{ on } \Omega_i, \tag{3.9}$$

then we can choose $w = X(x)Y(y)$ and obtain

$$-aX'' - b_1 X' = CX,$$

and

$$-aY'' - b_2 Y' = -CY.$$

We refer to the arbitrary constant C as the 'splitting' or 'separation' constant. Note that if we choose $C = 0$ we have our standard scheme based on tensor products of the one dimensional test functions. It is not obvious how to 'optimally' choose C but from experience from numerical experiments we have been able to develop automatic choices of C that seem to work well. By making a choice of $C = |b_2| - |b_1|$ we have found it is possible to greatly diminish or even remove any oscillations near parabolic boundary layers and

shear layers.

3.4.2 The Limits of Pure Convection and Pure Diffusion

This section describes the link between the methods described in this thesis and both the cell vertex finite volume method and the Galerkin finite element method. The link is made only for methods with zero splitting constant and for problems with no zeroth-order term, discretised on Cartesian product meshes. When links are made with the cell vertex finite volume method, we assume that the trial space in the Cell Vertex Finite Volume Method (CVFVM) is $(n-1)$-linear on the element boundaries (for an n dimensional problem). When links are made with the Galerkin finite element method we assume we are using an n-linear trial space.

For very high mesh Péclet numbers the test function resemble (see the figures in the next section) the test function used in the nonconforming Petrov-Galerkin formulation of the CVFVM (that is the characteristic function of the upwind cell), except for the cases $\mathbf{b} = \mathbf{0}$ and when the flow is along a mesh line. It is in these last two cases that the CVFVM behaves poorly (or is undefined) unless a different test function is used (usually by taking a weighted average of adjacent cells [7]). In fact in the limit of no diffusion with nonzero convection this method tends to the CVFVM:

Consider the difference stencil for node 5 on a 2×2 mesh on $[-1,1] \times [-1,1]$ with nodes numbered as shown below for both the CVFVM and the zero splitting constant method with bilinear trial space applied to a problem with no diffusion.

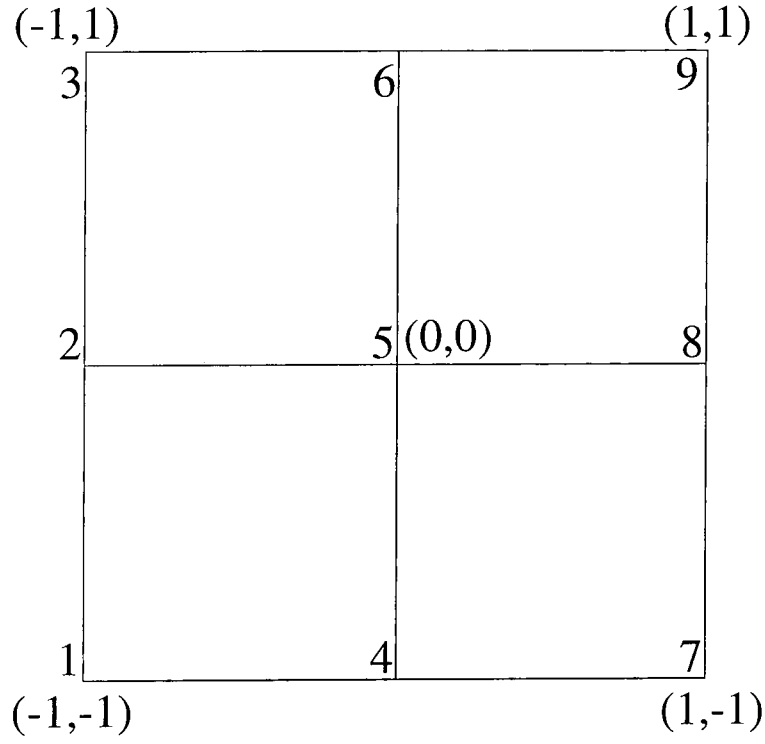


Figure 3.16: Section of mesh with numbered nodes

The difference scheme for the cell vertex finite volume method is

$$-\frac{(b_1+b_2)}{2}U_1 + \frac{(b_2-b_1)}{2}U_2 + 0.U_3 - \frac{(b_2-b_1)}{2}U_4 + \frac{(b_1+b_2)}{2}U_5 + 0.U_6 + 0.U_7 + 0.U_8 + 0.U_9 = \int_{[-1,0]\times[-1,0]} f \, d\Omega$$

This has been calculated from

$$\int_{[-1,0]\times[-1,0]} b_1 U_x + b_2 U_y \, d\Omega = \int_{[-1,0]\times[-1,0]} f \, d\Omega.$$

We are interested in the difference scheme for the zero splitting constant method with a bilinear trial space in the limit as the diffusion parameter a tends to 0. Firstly we note that terms of the form $\int a\nabla U\nabla w\ d\Omega$ can be ignored for we have with $w = X(x)Y(y)$ defined in a similar way to example 11 but with a general diffusion constant 'a'

$$\int a\nabla U\nabla w\ d\Omega = \int aU_xX(x)_xY(y) + aU_yY(y)_yX(x)\ d\Omega.$$

But,

$$\begin{aligned}\int_0^1 aX(x)_x\ dx &= -\frac{b_1}{(1-e^{-b_1/a})}\int_0^1 e^{-b_1x/a}\ dx \\ &= -\frac{b_1}{(1-e^{-b_1/a})}(\frac{-a}{b_1})(e^{-b_1/a}-1) \\ &= -a.\end{aligned}$$

Clearly as U_x is independent of x and U_y is independent of y (note that U is bilinear in x and y) and $X(x)$and $Y(y)$ are bounded above independent of a, we have that

$$|\int a\nabla U\nabla w\ d\Omega| \leq Ca$$

where C is a constant independent of a.

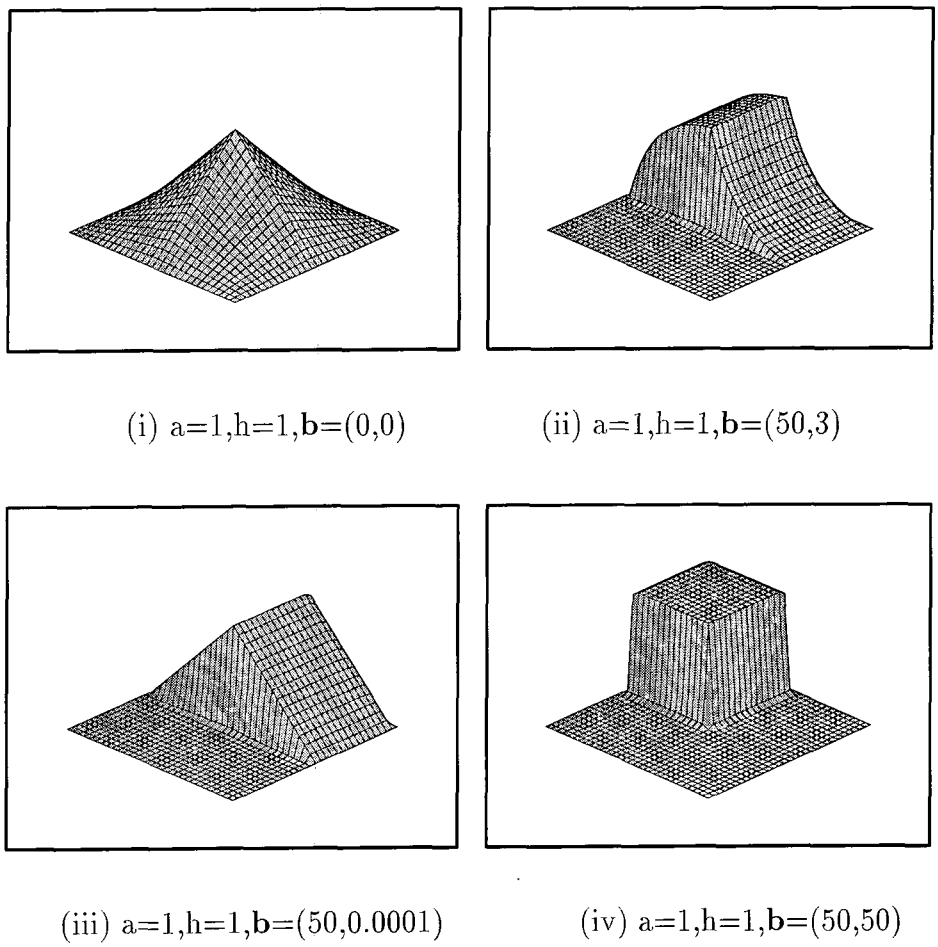
We now need to calculate $\lim_{a\to 0}\int \nabla U.\mathbf{b}w\ d\Omega$. This involves only integrals of linear functions times exponentials. Hence we can take the limit inside the integral. Note that for b_1 and b_2 both nonzero, $w = X(x)Y(y)$ in the limit as $a \to 0$ takes the value 1 on $[-1,0]\times[-1,0]$ and is zero elsewhere. Hence for non-mesh-aligned flow we have

$$\int \nabla U.\mathbf{b}(\lim_{a\to 0} w)\ d\Omega = \int_{[-1,0]\times[-1,0]} b_1U_x + b_2U_y\ d\Omega$$

Hence our scheme becomes identical to that of the CVFVM in these cases.

Note that in limit as $a \rightarrow 0$ with the flow along a mesh line (that is one of the components of $\mathbf{b}$ vanishes) we have a nonunique limit (depending on the ratio of diffusion to convection along the mesh as a varying flow moves towards alignment with the mesh). Figure 3.17 show various test functions on $[0,1]^2$ for this method. In particular we note how figures 3.17(ii) and 3.17(iii) differ greatly despite a very similar flow. In the limiting case we suggest the symmetric limit (see figure 3.17 (iii)). This interpretation of our method in the limit of no diffusion indicates that the CVFVM is 'overwinded' for problems involving diffusion. In fact the averaging that is sometimes performed (see [7]) could be viewed as 'downwinding' the 'overwinded' test functions. These observations also hold for the Cell Centred Finite Volume Method formulated as a Petrov-Galerkin method where the test space is based on a dual box mesh [45].

We note for clarity that for both nonzero splitting constant and for a nonzero zero-order term these test functions do not tend to the finite volume test functions in the limit of no diffusion. In the limiting case of pure diffusion, the method becomes the standard Galerkin finite element method.



(i) a=1,h=1,**b**=(0,0) (ii) a=1,h=1,**b**=(50,3)

(iii) a=1,h=1,**b**=(50,0.0001) (iv) a=1,h=1,**b**=(50,50)

Figure 3.17: Example tensor product test functions in two dimensions for various flows

3.4.3 Test Functions In Two Dimensions

Presented here in figures 3.18 to 3.35 are test functions with support $[0,1]^2$ for varying $\mathbf{b}$ and splitting constant C. In each case $a = 1$. When $b_1 = b_2$ we show the test functions for $C = 0, C = 10$ and $C = -10$. When $b_1 \neq b_2$ we show the test functions for $C = 0, C = |b_2| - |b_1|$ and $C = |b_1| - |b_2|$. Note, from our experience, we recommend using $C = |b_2| - |b_1|$.

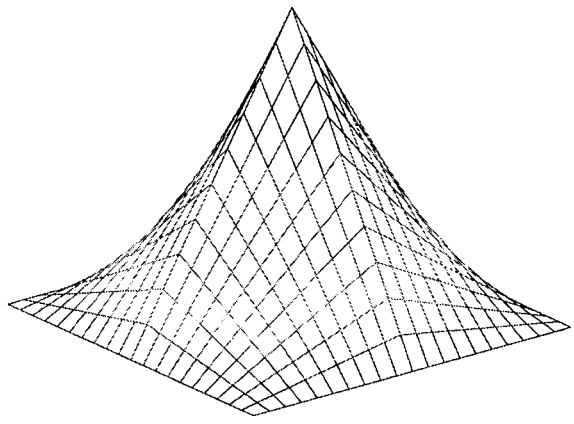


Figure 3.18: Test function : $\mathbf{b} = (0,0), C = 0$

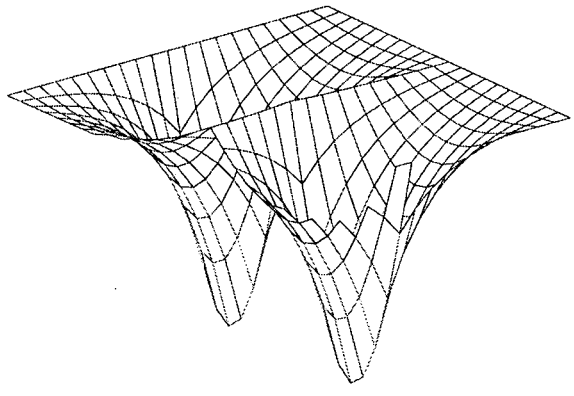


Figure 3.19: Test function : $\mathbf{b} = (0,0), C = 10$

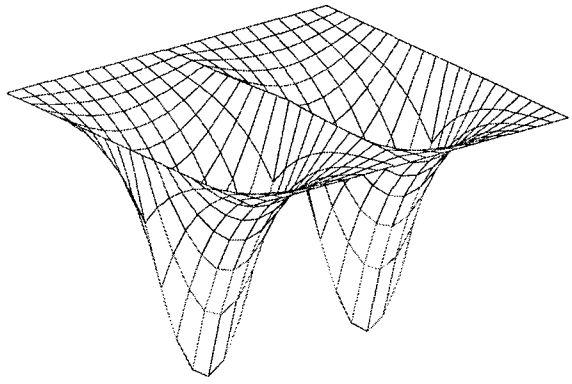


Figure 3.20: Test function : $\mathbf{b} = (0,0), C = -10$

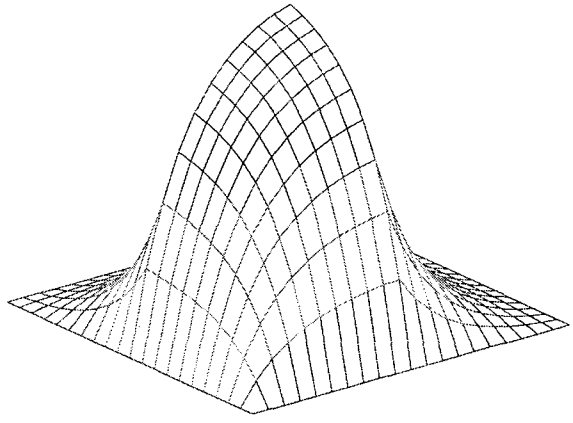


Figure 3.21: Test function : $\mathbf{b} = (3,3), C = 0$

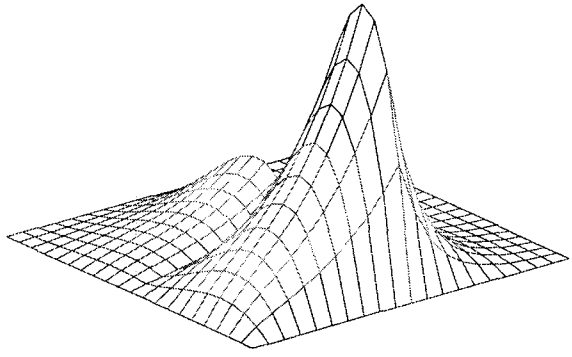


Figure 3.22: Test function : $\mathbf{b} = (3,3), C = 10$

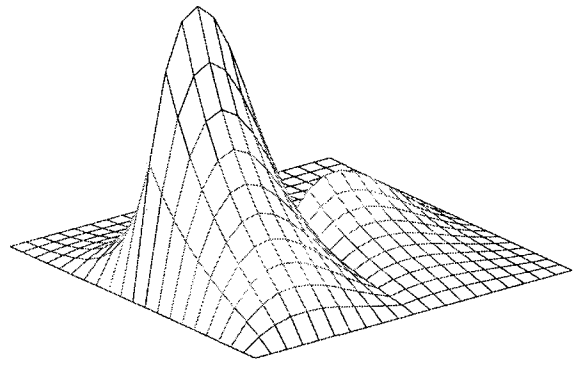


Figure 3.23: Test function : $\mathbf{b} = (3,3), C = -10$

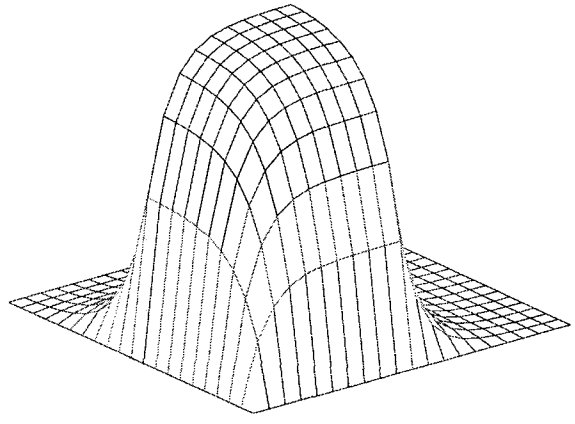


Figure 3.24: Test function : $\mathbf{b} = (7,5), C = 0$

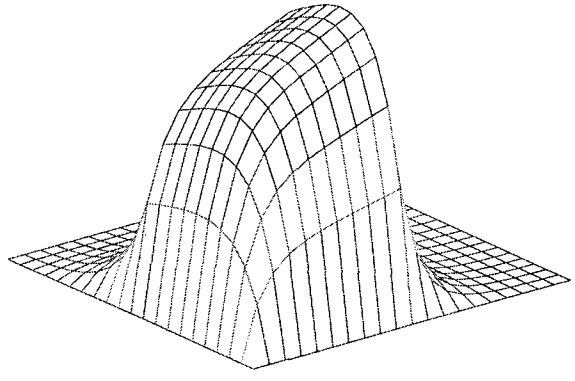


Figure 3.25: Test function : $\mathbf{b} = (7,5), C = 2$

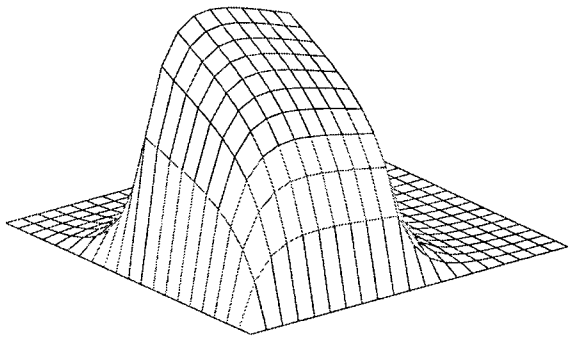


Figure 3.26: Test function : $\mathbf{b} = (7, 5), C = -2$

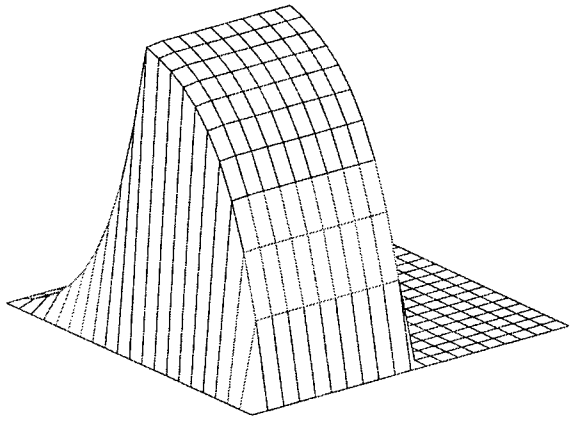


Figure 3.27: Test function : $\mathbf{b} = (50, 3), C = 0$

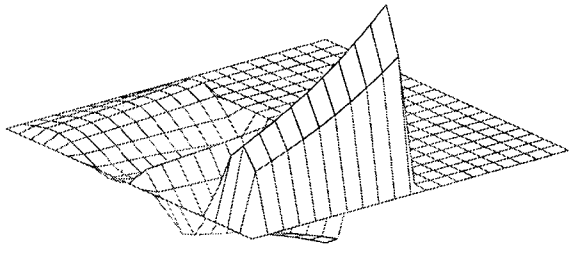


Figure 3.28: Test function : $\mathbf{b} = (50, 3), C = 47$

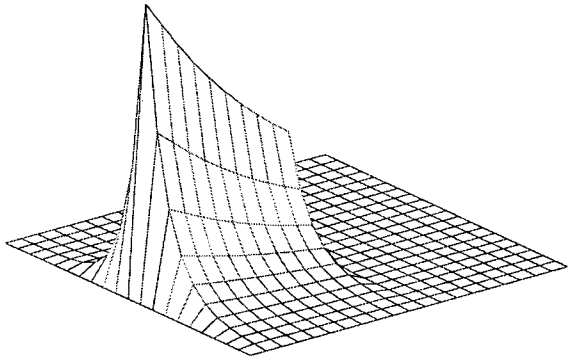


Figure 3.29: Test function : $\mathbf{b} = (50, 3), C = -47$

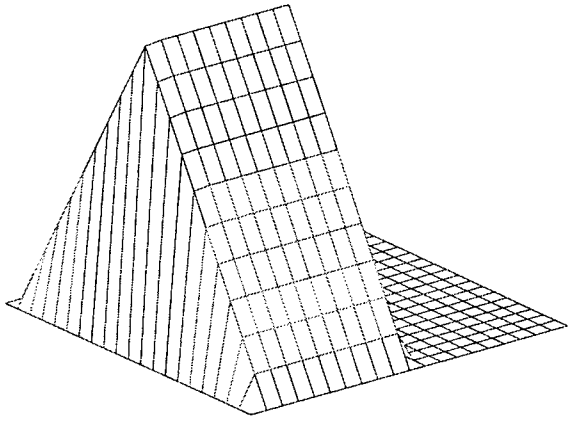


Figure 3.30: Test function : $\mathbf{b} = (50, 0), C = 0$

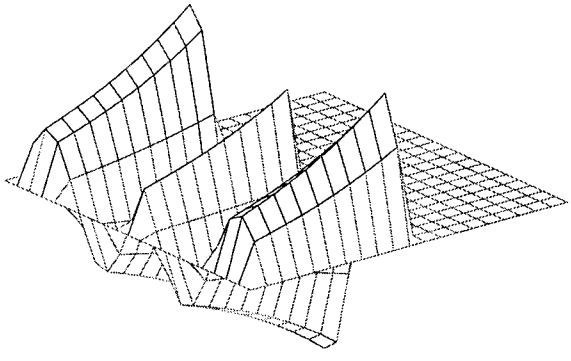


Figure 3.31: Test function : $\mathbf{b} = (50, 0), C = 50$

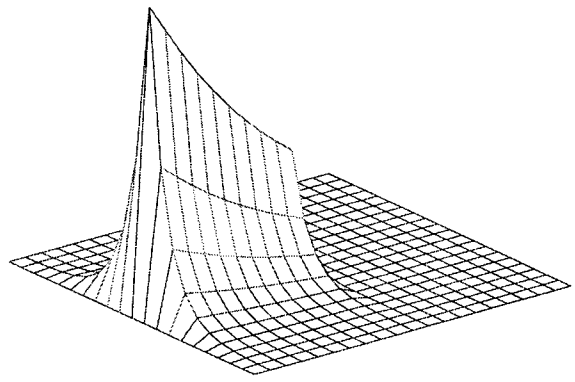


Figure 3.32: Test function : $\mathbf{b} = (50, 0), C = -50$

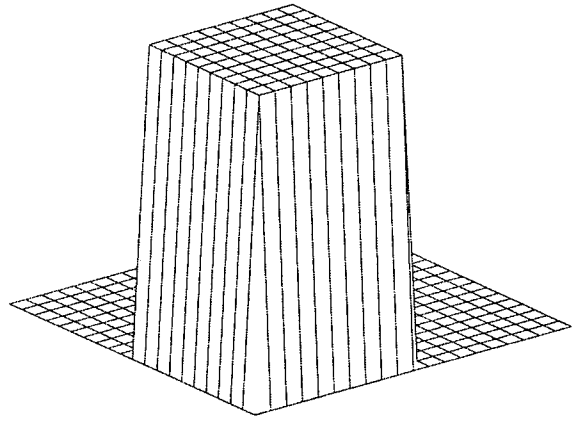


Figure 3.33: Test function : $\mathbf{b} = (500, 300), C = 0$

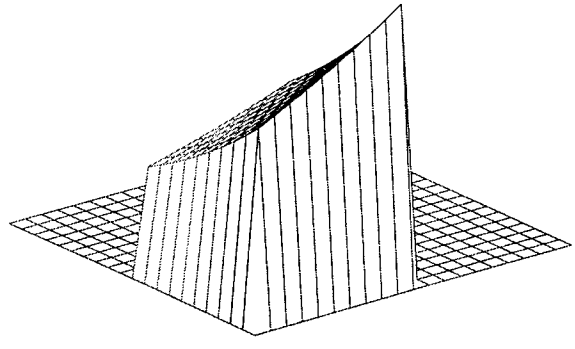


Figure 3.34: Test function : $\mathbf{b} = (500, 300), C = 200$

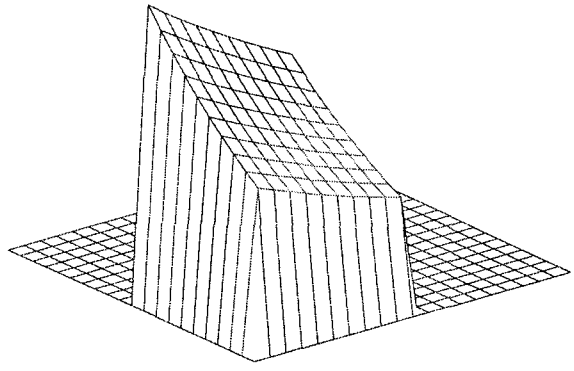


Figure 3.35: Test function : $\mathbf{b} = (500, 300), C = -200$

3.4.4 An Approximate Scheme For General Quadrilateral Meshes

Despite the existence of test spaces that fit this formulation, it is not possible to describe them in closed form for general quadrilaterals. One solution to this problem would be to solve the adjoint equation with some suitable boundary conditions on each element by some approximate method. This, however, is time consuming on serial computers. This method would be suited to a parallel computer implementation as each of these problems is entirely local. We have adopted a simpler solution for our numerical experiments. This consists of using test functions which do not satisfy the homogeneous adjoint equation exactly, but are 'close' enough for our purposes. For each quadrilateral containing a particular node as a vertex, we construct a rectangular region which is in some sense similar to that irregular quadrilateral. The test functions can then be defined on these rectangles in the usual way by using a local coordinate system defined by the sides of the rectangle. The functions are then mapped back to the quadrilateral in the usual way [5].

To generate the rectangle $(\bar{A}\bar{B}\bar{C}\bar{D})$ from the quadrilateral $(ABCD)$ we initially construct the diagonals of the quadrilateral (AC, BD) (see fig. 3.36). We then calculate

$$\lambda = \frac{(length\ of\ AC) + (length\ of\ BD)}{4}.$$

We then choose the points $\bar{A}\bar{B}\bar{C}$ and $\bar{D}$ at a distance λ from O along the line segments OA, OB, OC and OD, where O is the point of intersection of the diagonals. Note that if the original quadrilateral is also a rectangle then $\bar{A} = A, \bar{B} = B, \bar{C} = C$ and $\bar{D} = D$.

It is clear that the standard Galerkin finite element method based on ir-

regular quadrilateral grids is a method of this approximate type.

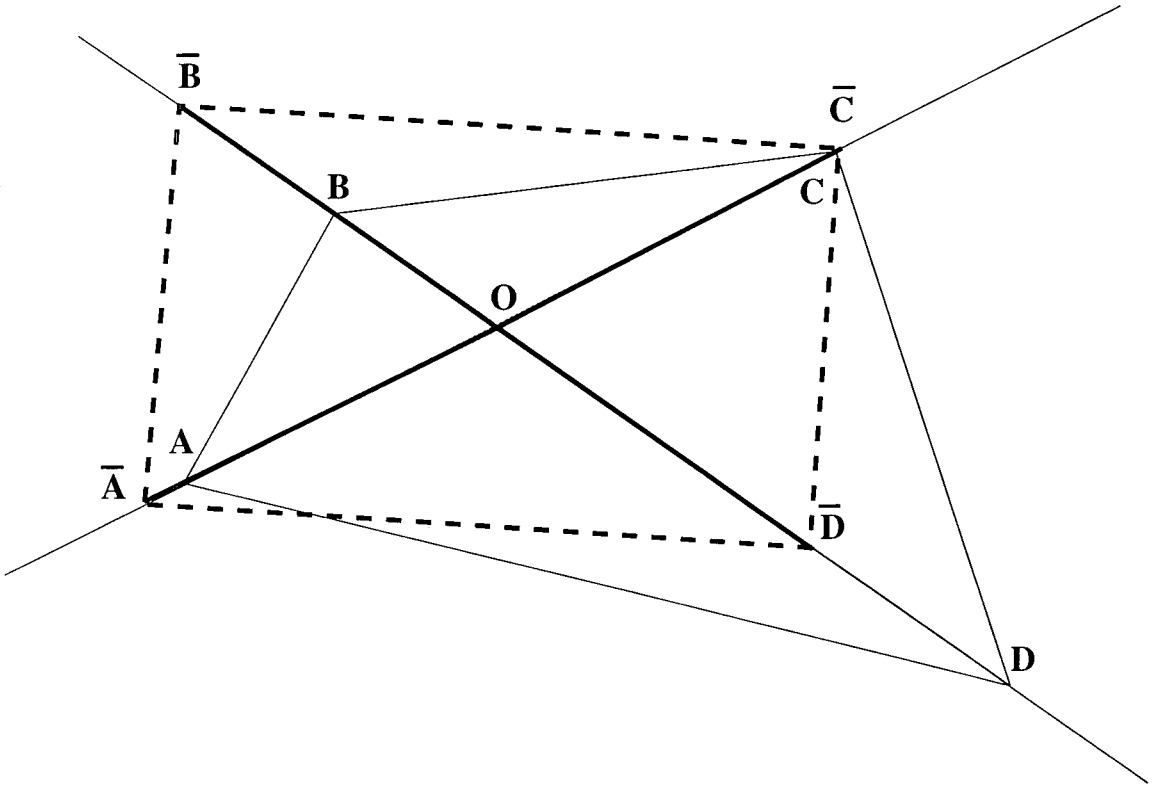


Figure 3.36: Rectangle construction from a general quadrilateral

Chapter 4

Error Analysis

4.1 Introduction

In this chapter we present asymptotic (mesh spacing $h \rightarrow 0$), nonasymptotic and truncation error analyses. Initially however we discuss, at a non technical level, various means at our disposal to produce error estimates. We present three methods, valid under certain assumptions.

In the following discussion the notation $||.||$ denotes some norm which may be different in separate occurrences. However when necessary it is assumed that two different $||.||$ are equivalent. We assume also that the space $\mathcal{V}$ consists of piecewise n-linear functions defined over the mesh.

We are trying to obtain an estimate for the difference between the exact solution u and the solution (assuming existence and uniqueness) to the approximate problem defined in section 2.4: find $U \in \mathcal{V}$ such that

$$B(U, w) = (f, w) \forall w \in \mathcal{W}.$$

We assume that $B(.,.)$ continuous, that is there is a positive constant β such that

$$|B(v, w)| \leq \beta ||v|| ||w||$$

Note that the error $u - U$ satisfies

$$B(u - U, w) = 0 \ \forall w \in \mathcal{W}.$$

4.1.1 Error bound assuming ellipticity of $B(.,.)$ and a bounded mapping from $\mathcal{V}$ to $\mathcal{W}$

The existence of a bounded mapping from the trial to the test space is very important in the error analysis of Petrov–Galerkin finite element methods. This is discussed in some depth in [31].

Assume there exists a constant $\alpha > 0$ and a mapping $g : v \rightarrow w, (v \in \mathcal{V}, w \in \mathcal{W})$ such that

$$|B(v, g(v))| \geq \alpha ||v||^2 \forall v \in \mathcal{V}$$

and

$$||g(v)|| \leq \gamma ||v||, v \in \mathcal{V}.$$

Then by defining u_2 as the best approximation in $\mathcal{V}$ of u in the $||.||$ norm, we have

$$\begin{aligned} ||u_2 - U||^2 &\leq (1/\alpha)|B(u_2 - U, g(u_2 - U))| \\ &\leq (1/\alpha)|B(u_2 - u, g(u_2 - U))| \\ &\leq (\gamma\beta/\alpha)||u_2 - u||||u_2 - U|| \end{aligned}$$

Hence,

$$||u - U|| \leq (1 + \gamma\beta/\alpha)||u_2 - u||.$$

4.1.2 Error bound assuming stability of the dual problem

We would like to be able to solve the problem: find $u \in \mathcal{W}$ such that

$$B(v,u) = (f,v)\ \forall v \in \mathcal{V}.$$

For this analysis we take the norm $||.||$ to be the $(L^2)^n$ norm. Assume we have stability of this dual problem. That is that there is a constant $\alpha > 0$ such that

$$||u|| \leq (1/\alpha)||f||. \tag{4.1}$$

Then (assuming existence and uniqueness) we choose $d \in \mathcal{W}$ such that

$$B(v,d) = (u_2 - U, v)\ \forall v \in \mathcal{V}$$

where u_2 is the best approximation to u from $\mathcal{V}$ in the $(L^2)^n$ norm.

Then by choosing $v = u_2 - U$ in the above equation we obtain

$$\begin{aligned} |u_2 - U||^2 &= B(u_2 - U, d) \\ &= B(u_2 - u, d) \\ &\leq (\beta/\alpha)||u_2 - u||||u_2 - U||, \end{aligned}$$

where we have used equation 4.1 to bound d above by $(1/\alpha)||u_2 - U||$. Hence,

$$||u - U|| \leq (1 + \beta/\alpha)||u_2 - u||.$$

4.1.3 Truncation Error estimate assuming stability

The standard truncation error analysis approach is based on the observation that given stability of the discrete problem, a small truncation error will yield a small global error. Under reasonable stability conditions (see [28] for a discussion of this) it can be shown that a local truncation error of order h^4 yields an error bound of order h^2. In section 4.5 we show that the local truncation error is of this order for certain methods.

4.2 Asymptotic Error Analysis ($h \rightarrow 0$)

Presented here is an asymptotic error analysis for a tensor product (with zero splitting constant) Petrov–Galerkin finite element method. It is presented in two dimensions on a unit square mesh $[0,1]^2$ although the technique is clearly applicable in higher dimensions. We write the convection vector $\mathbf{b} = (b_1, b_2)^T$ and for simplicity take $a = 1$.

Let $\Omega^h = \{(x_i, y_j) : i, j = 0, 1, ..., N\}$ be the set of mesh points, and let h be the mesh spacing which for simplicity we assume to be constant.

The functions $\phi_i(x)$ and $\psi_i(x)$ defined by the following expressions are the definitions of the equivalent trial and test space (respectively) basis functions.

Let,

$\phi_1(x)$ satisfy $\phi_{1_{xx}} = 0$ on $(0, h)$ with $\phi_1(0) = 0$ and $\phi_1(h) = 1$. That is $\phi_1(x) = x/h$ on [0, h].

$\phi_2(x)$ satisfy $\phi_{2_{xx}} = 0$ on $(0, h)$ with $\phi_2(0) = 1$ and $\phi_2(h) = 0$. That is $\phi_2(x) = (h - x)/h$ on [0, h].

$\phi^1(y)$ satisfy $\phi^1_{yy} = 0$ on $(0,h)$ with $\phi^1(0) = 0$ and $\phi^1(h) = 1$.

$\phi^2(y)$ satisfy $\phi^2_{yy} = 0$ on $(0,h)$ with $\phi^2(0) = 1$ and $\phi^2(h) = 0$.

$\psi_1(x,b_1) \in H^1_0(0,h)$ satisfy $\phi_1(0) = 0$ and $\phi_1(h) = 1$.

$\psi_2(x,b_1) \in H^1_0(0,h)$ satisfy $\phi_2(0) = 1$ and $\phi_2(h) = 0$.

$\psi^1(y,b_2) \in H^1_0(0,h)$ satisfy $\phi^1(0) = 0$ and $\phi^1(h) = 1$.

$\psi^2(y,b_2) \in H^1_0(0,h)$ satisfy $\phi^2(0) = 1$ and $\phi^2(h) = 0$.

Given $v \in \mathcal{V}$, let $\bar{v} \in \mathcal{W}$ be such that $v = \bar{v}$ at the nodes. Then, with obvious abuse of notation,

$$\begin{aligned} B(v,\bar{v}) &= \sum \bar{v}_{ij} B(v, \psi^i \psi_j) \\ &= \sum \bar{v}_{ij} \int [(v_x, \psi^i_x \psi_j + b_1 \psi^i \psi_j) + (v_y, \psi^i \psi_{j_y} + b_2 \psi^i \psi_j)] \, d\Omega \end{aligned}$$

Theorem 12 *If we have two constants $\gamma_1(h,\mathbf{b}) > 0$ and $\gamma_2(h,\mathbf{b}) > 0$ such that*

$$|B(v,\bar{v})| \geq \gamma_2 |v|_1^2 \ \forall v \in \mathcal{V}, \bar{v} \in \mathcal{W} : v = \bar{v} \text{ at the nodes},$$

and,

$$|B(v,\bar{w})| \leq \gamma_1 |v|_1 |w|_1 \ \forall v \in \mathcal{V}, \bar{w} \in \mathcal{W}$$

where $w \in \mathcal{V}$ is defined suchthat $w = \bar{w}$ at the nodes.

Then we have,

$$|u - U|_1 \leq (1 + \frac{\gamma_1}{\gamma_2}) \inf_{v \in \mathcal{V}} |u - v|_1$$

Proof *Define* u^* *such that* $|u-u^*|_1 \leq |u-v|_1 \ \forall v \in \mathcal{V}$. *We choose* $\bar{w} \in \mathcal{W}$ *such that* $\bar{w} = u^* - U$ *at the nodes. Then we have,*

$$\begin{aligned} \gamma_2 |u^* - U|_1^2 &\leq |B(u^* - U, \bar{w})| \\ &= |B(u^* - u, \bar{w})| \\ &= \gamma_1 |u^* - u|_1 |u^* - U|_1, \end{aligned}$$

The result follows from the triangle inequality.

To apply this theorem we need to show the existence of these two constants. The existence of a γ_1 is trivial, but the existence of γ_2 is more complicated.

To calculate γ_2 it is helpful to make the following definitions:

Definition 13 *Let* $w_1, w^1, w_2, w^2, z_1, z^1, z_2, z^2$ be defined from the following 8 relations.

$$a\psi_{1x}(x) + b_1\psi_1(x) = \phi_{1x}(x) w_1(b_1, x)$$

$$a\psi_{2x}(x) + b_1\psi_2(x) = \phi_{2x}(x) w_2(b_1, x)$$

$$a\psi^{1y}(y) + b_2\psi^1(y) = \phi^1_y(y) w^1(b_2, y)$$

$$a\psi^{2y}(y) + b_2\psi^2(y) = \phi^2_y(y) w^2(b_2, y)$$

$$\psi_1(x) = \phi_1(x) z_1(b_1, x)$$

$$\psi_2(x) = \phi_2(x) z_2(b_1, x)$$

$$\psi^1(y) = \phi^1(y) z^1(b_2, y)$$

$$\psi^2(y) = \phi^2(y) z^2(b_2, y)$$

The motivation for the above definitions is that we would like $B(v, \bar{v})$ to be similar to the square of the H^1 seminorm.

Notation We will use the following notation for the standard L^2 inner product : $(f,g) = \int_0^1 f(x)g(x)\,dx$.

Then,

$$\begin{aligned}
|B(v,\bar{v})| = \quad & \sum_{i=0}^{N}\sum_{j=0}^{N} v_{i,j}^2(W_1+W_2) \\
+ \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N} v_{i,j}v_{i+1,j}(A_1+A_2) \\
+ \quad & \sum_{i=0}^{N}\sum_{j=0}^{N-1} v_{i,j}v_{i,j+1}(B_1+B_2) \\
+ \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} v_{i,j}v_{i+1,j+1}(C_1+C_2) \\
+ \quad & \sum_{i=0}^{N-1}\sum_{j=1}^{N} v_{i,j}v_{i+1,j-1}(D_1+D_2)
\end{aligned}$$

where,

$$
W_1 = \begin{pmatrix} & (\phi_{1x}, \phi_{1x}w_1)(\phi^1, \phi^1 z^1) \\ + & (\phi_{1x}, \phi_{1x}w_1)(\phi^2, \phi^2 z^2) \\ + & (\phi_{2x}, \phi_{2x}w_2)(\phi^1, \phi^1 z^1) \\ + & (\phi_{2x}, \phi_{2x}w_2)(\phi^2, \phi^2 z^2) \end{pmatrix} \quad W_2 = \begin{pmatrix} & (\phi^{1y}, \phi^{1y}w^1)(\phi_1, \phi_1 z_1) \\ + & (\phi^{1y}, \phi^{1y}w^1)(\phi_2, \phi_2 z_2) \\ + & (\phi^{2y}, \phi^{2y}w^2)(\phi_1, \phi_1 z_1) \\ + & (\phi^{2y}, \phi^{2y}w^2)(\phi_2, \phi_2 z_2) \end{pmatrix}
$$

$$
A_1 = \begin{pmatrix} & (\phi_{2x}, \phi_{1x}w_1)(\phi^1, \phi^1 z^1) \\ + & (\phi_{2x}, \phi_{1x}w_1)(\phi^2, \phi^2 z^2) \\ + & (\phi_{1x}, \phi_{2x}w_2)(\phi^1, \phi^1 z^1) \\ + & (\phi_{1x}, \phi_{2x}w_2)(\phi^2, \phi^2 z^2) \end{pmatrix} \quad A_2 = \begin{pmatrix} & (\phi^{1y}, \phi^{1y}w^1)(\phi_2, \phi_1 z_1) \\ + & (\phi^{1y}, \phi^{1y}w^1)(\phi_1, \phi_2 z_2) \\ + & (\phi^{2y}, \phi^{2y}w^2)(\phi_2, \phi_1 z_1) \\ + & (\phi^{2y}, \phi^{2y}w^2)(\phi_1, \phi_2 z_2) \end{pmatrix}
$$

$$
B_1 = \begin{pmatrix} & (\phi_{1x}, \phi_{1x}w_1)(\phi^2, \phi^1 z^1) \\ + & (\phi_{1x}, \phi_{1x}w_1)(\phi^1, \phi^2 z^2) \\ + & (\phi_{2x}, \phi_{2x}w_2)(\phi^2, \phi^1 z^1) \\ + & (\phi_{2x}, \phi_{2x}w_2)(\phi^1, \phi^2 z^2) \end{pmatrix} \quad B_2 = \begin{pmatrix} & (\phi^{2y}, \phi^{1y}w^1)(\phi_1, \phi_1 z_1) \\ + & (\phi^{2y}, \phi^{1y}w^1)(\phi_2, \phi_2 z_2) \\ + & (\phi^{1y}, \phi^{2y}w^2)(\phi_1, \phi_1 z_1) \\ + & (\phi^{1y}, \phi^{2y}w^2)(\phi_2, \phi_2 z_2) \end{pmatrix}
$$

$$
C_1 = \begin{pmatrix} & (\phi_{2x}, \phi_{1x}w_1)(\phi^2, \phi^1 z^1) \\ + & (\phi_{1x}, \phi_{2x}w_2)(\phi^1, \phi^2 z^2) \end{pmatrix} \quad C_2 = \begin{pmatrix} & (\phi^{2y}, \phi^{1y}w^1)(\phi_2, \phi_1 z_1) \\ + & (\phi^{1y}, \phi^{2y}w^2)(\phi_1, \phi_2 z_2) \end{pmatrix}
$$

$$
D_1 = \begin{pmatrix} & (\phi_{2x}, \phi_{1x}w_1)(\phi^1, \phi^2 z^2) \\ + & (\phi_{1x}, \phi_{2x}w_2)(\phi^2, \phi^1 z^1) \end{pmatrix} \quad D_2 = \begin{pmatrix} & (\phi^{2y}, \phi^{1y}w^1)(\phi_1, \phi_2 z_2) \\ + & (\phi^{1y}, \phi^{2y}w^2)(\phi_2, \phi_1 z_1) \end{pmatrix}
$$

Theorem 14 *The following conditions are sufficient for the existence of a positive constant γ_2 satisfying the first condition of theorem 12.*

$$w_1, w_2, z_1, z_2, w^1, w^2, z^1, z^2 > 0,$$

$$A_1 + A_2 < 0$$

and

$$B_1 + B_2 < 0.$$

Proof of theorem 14.

We can rewrite $|B(v, \bar{v})|$ in a more useful manner:

$$
\begin{aligned}
|B(v,\bar{v})| = \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N}(v_{i,j}+v_{i+1,j})^2(-\frac{A_1+A_2}{2}) \\
+ \quad & \sum_{i=0}^{N}\sum_{j=0}^{N-1}(v_{i,j}-v_{i,j+1})^2(-\frac{B_1+B_2}{2}) \\
+ \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(v_{i,j}-v_{i+1,j+1})^2(-\frac{C_1+C_2}{2}) \\
+ \quad & \sum_{i=0}^{N-1}\sum_{j=1}^{N}(v_{i,j}-v_{i+1,j-1})^2(-\frac{D_1+D_2}{2}) \\
+ \quad & \sum_{i=1}^{N-1}\sum_{j=1}^{N-1}v_{i,j}^2(W_1+A_1+B_1+C_1+D_1+W_2+A_2+B_2+C_2+D_2) \\
+ \quad & \sum_{j=1}^{N-1}(v_{0,j}^2+v_{N,j}^2)(W_1+W_2+\frac{A_1+A_2}{2}+B_1+B_2+\frac{C_1+C_2}{2}+\frac{D_1+D_2}{2}) \\
+ \quad & \sum_{i=1}^{N-1}(v_{i,0}^2+v_{i,N}^2)(W_1+W_2+A_1+A_2+\frac{B_1+B_2}{2}+\frac{C_1+C_2}{2}+\frac{D_1+D_2}{2}) \\
+ \quad & (v_{0,N}^2+v_{N,0}^2)(W_1+W_2+\frac{A_1+A_2}{2}+\frac{B_1+B_2}{2}+\frac{D_1+D_2}{2}) \\
+ \quad & (v_{N,N}^2+v_{0,0}^2)(W_1+W_2+\frac{A_1+A_2}{2}+\frac{B_1+B_2}{2}+\frac{C_1+C_2}{2})
\end{aligned}
$$

Firstly we note that,

$$
\begin{aligned}
W_1+A_1+B_1+C_1+D_1 \quad = \quad & (\phi_{1x},\phi_{1x}w_1)((\phi^1,\phi^1z^1)+(\phi^2,\phi^2z^2)+(\phi^2,\phi^1z^1)+(\phi^1,\phi^2z^2)) \\
& +(\phi_{2x},\phi_{2x}w_2)((\phi^1,\phi^1z^1)+(\phi^2,\phi^2z^2)+(\phi^2,\phi^1z^1)+(\phi^1,\phi^2z^2)) \\
& +(\phi_{2x},\phi_{1x}w_1)((\phi^1,\phi^1z^1)+(\phi^2,\phi^2z^2)+(\phi^2,\phi^1z^1)+(\phi^1,\phi^2z^2)) \\
& +(\phi_{1x},\phi_{2x}w_2)((\phi^1,\phi^1z^1)+(\phi^2,\phi^2z^2)+(\phi^2,\phi^1z^1)+(\phi^1,\phi^2z^2)) \\
= \quad & (\phi_{1x},\phi_{1x}w_1)(\phi^1+\phi^2,\phi^1z^1+\phi^2z^2)
\end{aligned}
$$

$$
\begin{aligned}
&+(\phi_{2x}, \phi_{2x}w_2)(\phi^1 + \phi^2, \phi^1 z^1 + \phi^2 z^2) \\
&+(\phi_{2x}, \phi_{1x}w_1)(\phi^1 + \phi^2, \phi^1 z^1 + \phi^2 z^2) \\
&+(\phi_{1x}, \phi_{2x}w_2)(\phi^1 + \phi^2, \phi^1 z^1 + \phi^2 z^2) \\
= \;& (\phi_{1x} + \phi_{2x}, \phi_{1x}w_1 + \phi_{2x}w_2)(1, \phi^1 z^1 + \phi^2 z^2) \\
= \;& 0.
\end{aligned}
$$

(We have used, in the above expressions, the fact that $\phi^1 + \phi^2 = 1$ and also that $\phi_{1x} + \phi_{2x} = 0$.)

Similarly, $W_2 + A_2 + B_2 + C_2 + D_2 = 0$.

Also since $\phi_{1x} + \phi_{2x} = 0$, we have $W_1 = -A_1$ and $W_2 = -B_2$.

Combining these give $B_1 + C_1 + D_1 = A_2 + C_2 + D_2 = 0$.

Use of these expressions yields,

$$
\begin{aligned}
|B(v, \bar{v})| = \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N}(v_{i,j} + v_{i+1,j})^2(-\frac{A_1 + A_2}{2}) \\
+ \quad & \sum_{i=0}^{N}\sum_{j=0}^{N-1}(v_{i,j} - v_{i,j+1})^2(-\frac{B_1 + B_2}{2}) \\
+ \quad & \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(v_{i,j} - v_{i+1,j+1})^2(-\frac{C_1 + C_2}{2}) \\
+ \quad & \sum_{i=0}^{N-1}\sum_{j=1}^{N}(v_{i,j} - v_{i+1,j-1})^2(-\frac{D_1 + D_2}{2}) \\
+ \quad & \sum_{j=1}^{N-1}(v_{0,j}^2 + v_{N,j}^2)\frac{W_1 + B_1}{2} \\
+ \quad & \sum_{i=1}^{N-1}(v_{i,0}^2 + v_{i,N}^2)\frac{W_2 + A_2}{2} \\
+ \quad & (v_{0,N}^2 + v_{N,0}^2)\frac{W_1 - C_1}{2} + \frac{W_2 - C_2}{2}
\end{aligned}
$$

$$+ \quad (v_{N,N}^2 + v_{0,0}^2)\frac{W_1 - D_1}{2} + \frac{W_2 - D_2}{2}$$

As $w_1, w_2, z_1, z_2, w^1, w^2, z^1, z^2 > 0$ by assumption and by the definition of $\phi_i(x)$ the product $\phi_{1x}\phi_{2x} < 0$ we have that C_1, C_2, D_1 and D_2 are negative and W_1, W_2, B_1 and A_2 are positive.

Hence,

$$C_1 + C_2 < 0,$$

$$D_1 + D_2 < 0,$$

$$W_1 + B_1 > 0,$$

$$W_2 + A_2 > 0,$$

$$W_1 - C_1 > 0,$$

$$W_2 - C_2 > 0,$$

$$W_1 - D_1 > 0,$$

$$W_2 - D_2 > 0.$$

As all the individual terms in the equation for $|B(v, \bar{v})|$ are positive we can bound $|B(v, \bar{v})|$ below by

$$|B(v, \bar{v})| \geq C_2 |v|_1^2.$$

Example 15 *The tensor product methods (with splitting constant $C = 0$ and $a = 1$) described in the last chapter employ the following definitions:*

$\psi_1(x, b_1)$ satisfies
$-\psi_{1_{xx}} - b_1\psi_{1_x} = 0$ on $(0, h)$ with $\phi_1(0) = 0$ and $\phi_1(h) = 1$.

$\psi_2(x, b_1)$ *satisfies*
$-\psi_{2_{xx}} - b_1\psi_{2_x} = 0$ *on* $(0, h)$ *with* $\phi_2(0) = 1$ *and* $\phi_2(h) = 0$.

$\psi^1(y, b_2)$ *satisfies*
$-\psi^2_{yy} - b_2\psi^1_y = 0$ *on* $(0, h)$ *with* $\phi^1(0) = 0$ *and* $\phi^1(h) = 1$.

$\psi^2(y, b_2)$ *satisfies*
$-\psi^2_{yy} - b_2\psi^2_y = 0$ *on* $(0, h)$ *with* $\phi^2(0) = 1$ *and* $\phi^2(h) = 0$.

If we now choose $b_1 = b_2 = b$ *then we can obtain,*

$$\begin{aligned} w_1 &= \frac{bh}{1 - e^{-bh}} \\ w_2 &= \frac{bhe^{-bh}}{1 - e^{-bh}} \\ z_1 &= \frac{h}{x}\left(\frac{e^{-bx} - 1}{e^{-bh} - 1}\right) \\ z_2 &= \frac{h}{h - x}\left(\frac{e^{-bx} - e^{-bh}}{1 - e^{-bh}}\right) \end{aligned}$$

The above expressions are the products of positive factors and so satisfy the conditions of theorem 14

Using $\phi_{1x} = -\phi_{2x}$ *and similar expressions we have that:*

$$\begin{aligned} B_1 + B_2 &= A_1 + A_2. \\ A_1 + A_2 &= ((\phi_{1x}, \phi_{1x}w_1) + (\phi_{2x}, \phi_{2x}w_2))((\phi_2, \phi_1 z_1) + (\phi_1, \phi_2 z_2) - (\phi_1, \phi_1 z_1) - (\phi_2, \phi_2 z_2)) \\ &= ((\phi_{1x}, \phi_{1x}w_1) + (\phi_{2x}, \phi_{2x}w_2))(\phi_1 z_1 - \phi_2 z_2, \phi_2 - \phi_1). \end{aligned}$$

Writing $p = bh$, *we can show that*

$$(\phi_1 z_1 - \phi_2 z_2, \phi_2 - \phi_1) = \frac{-2}{p^2(1 - e^{-p})}(pe^{-p} + 2e^{-p} + p - 2)$$

Limit of p	$p = 0$	$p \to \infty$
$A_1 + A_2$	$\frac{-2}{3}$	-2
$B_1 + B_2$	$\frac{-2}{3}$	-2
$C_1 + C_2$	$\frac{-2}{3}$	$2 - p$
$D_1 + D_2$	$\frac{-2}{3}$	0
$W_1 + B_1$	2	p
$W_2 + B_2$	2	p
$W_1 - C_1$	$\frac{5}{3}$	p
$W_2 - C_2$	$\frac{5}{3}$	p
$W_1 - D_1$	$\frac{5}{3}$	$1 + \frac{p}{2}$
$W_2 - D_2$	$\frac{5}{3}$	$1 + \frac{p}{2}$

Table 4.1: Constant bounds above and below

$$\begin{aligned} &= \frac{-2}{p^2}\left(\frac{p(1+exp(-p)}{(1-exp(-p))} - 2\right) \\ &= \frac{-2}{p^2}\frac{2}{\tanh(p/2)}(p/2 - \tanh(p/2)) \\ &< 0. \end{aligned}$$

Therefore $A_1 + A_2 < 0$ *and similarly* $B_1 + B_2 < 0$.

The expressions in table 4.1 were calculated in the computer package MAPLE and their limits in the mesh Péclet number $p = bh$ *calculated. They are all monotonic in* p.

4.3 Nonasymptotic Error Analysis

Asymptotic results are of limited value as often a method will not demonstrate the asymptotic rate of convergence until the mesh spacing h becomes very small. In fact asymptotic bounds are available for the standard Galerkin method, but are of little use as the method works poorly for reasonable mesh sizes. Of more practical use is an error estimate which gives the error in any given problem as a multiple of the smallest possible error for that problem.

We present an error analysis for the two dimensional case, although the technique is trivially extendible to higher dimensions.

4.3.1 A Mesh Dependent Inner Product and Norm

Definition 16 Setting,

$$\Gamma \equiv \bigcup_i \Gamma_i,$$

we define the following mesh dependent inner product and its associated discrete L^2 norm,

$$(u, v)_h = h \int_\Gamma uv \; d\Gamma,$$

$$\|v\|^2 = h \int_\Gamma v^2 \; d\Gamma. \tag{4.2}$$

If $v \in H^1$ then $\|v\|$ is well defined which is not the case for the usual discrete L^2 norm.

4.3.2 Analysis

Definition 17 Let $\mathcal{D}$ be the space of functions defined on the mesh boundaries by taking the jump in the normal derivative across Γ of functions $w \in \mathcal{W}$. Note that $\mathcal{D}$ is not continuous at the nodal points.

For graphical examples of basis functions for various spaces $\mathcal{D}$ see section 4.4.

Definition 18 Let $\mathcal{C}$ be the space of continuous piecewise linear functions defined on Γ such that $c \in \mathcal{C}$ vanishes on the boundary of the domain Ω.

The projection given in theorem 6 now becomes,

$$\int_\Gamma (u - U)d \, d\Gamma = 0 \; \forall \; d \in \mathcal{D}.$$

Given $c \in \mathcal{C}$ we define $d^* \in \mathcal{D}$ to be its projection in $(\cdot, \cdot)_h$ into $\mathcal{D}$, that is

$$(d^*, d)_h = (c, d)_h \; \forall d \in \mathcal{D}.$$

Note that $\|d^*\| \leq \|c\|$. Also we can define [11]

$$k = \inf_{d \in \mathcal{D}} \sup_{c \in \mathcal{C}} \frac{\|c - d\|}{\|c\|}, c \neq 0$$

so that $\forall c \in \mathcal{C}$ we have

$$\|c - d^*\| \leq k\|c\|. \tag{4.3}$$

We have $0 \leq k \leq 1$. From now on we assume $k < 1$.

Given the restriction of $(u-U)$ to the element boundaries (which we will refer to, by abuse of notation, as $(u-U)$, we define c^* to be its projection in $(\cdot,\cdot)_h$ into $\mathcal{C}$, that is

$$(c^*, c)_h = (u - U, c)_h \ \forall c \in \mathcal{C}.$$

Note that

$$\|c^*\| \leq \|u - U\|. \tag{4.4}$$

Now,

$$\begin{aligned} \|u-U\|^2 &= h\int_\Gamma (u-U)(u-U-d)\ d\Gamma \ \forall d \in \mathcal{D} \\ &\leq \|u-U\|\|u-U-d\| \ \forall\ d \in \mathcal{D}. \end{aligned}$$

Hence,

$$\|u-U\| \leq \|u-U-d\| \ \forall\ d \in \mathcal{D}$$

so we have that

$$\|u-U\| \leq \|u-U-c^*\| + \|c^*-d^*\|.$$

Therefore using equations 4.3 and 4.4

$$\|u-U\| \leq \|u-U-c^*\| + k\|u-U\|,$$

which implies that,

$$\|u-U\| \leq \frac{1}{1-k}\|u-v\| \ \forall v \in \mathcal{C}.$$

Note that as $k < 1$ this implies that we get the exact solution on the mesh boundaries if it is possible to attain it. For any given problem, we are able to calculate this constant.

4.3.3 Evaluation of the Optimal Constants

From the definition of d^* we have (setting $d = d^*$),

$$\int_\Gamma c d^* \, d\Gamma = \int_\Gamma d^{*2} \, d\Gamma.$$

Hence we may write

$$\begin{aligned} \|c - d^*\|^2 &= \|c\|^2 + \|d^*\|^2 - 2\int_\Gamma c d^* \, d\Gamma \\ &= \|c\|^2 + \|d^*\|^2 - 2\int_\Gamma d^* d^* \, d\Gamma \\ &= \|c\|^2 - \|d^*\|^2. \end{aligned}$$

This allows us to calculate k^2 in the following way [11].

Let $\{\phi_i\}$ and $\{\psi_i\}$ be a basis for $\mathcal{C}$ and $\mathcal{D}$ respectively. We define three matrices

$$A_{ij} = h\int_\Gamma \psi_i \psi_j \, d\Gamma,$$

$$B_{ij} = h\int_\Gamma \psi_i \phi_j \, d\Gamma,$$

$$C_{ij} = h\int_\Gamma \phi_i \phi_j \, d\Gamma.$$

Then we obtain,

$$k^2 = \sup_{v \in \mathcal{C}} \left(1 - \frac{v^T (B^T A^{-1} B) v}{v^T C v}\right)$$

Since C is symmetric and positive definite we can compute the smallest

n	$b_1 = 0, b_2 = 0$	$b_1 = 10, b_2 = 10$	$b_1 = 1000, b_2 = 1$
3	1.0	3.0	3.5
4	1.7	7.5	5.0
5	1.9	12.5	8.5
6	2.1	18.2	12.9
7	2.2	24.1	17.7

Table 4.2: Values of $\frac{1}{1-k}$ for various problems

eigenvalue λ satisfying the generalised eigenvalue problem $B^T A^{-1} B v = \lambda C v$ (see for example [46]) and hence find $k^2 = 1 - \lambda$.

The constant $\frac{1}{1-k}$ measures how far the approximation error is from the best possible error in our mesh dependent L_2-like norm.

The following table shows the value of this constant for grids of n by n points on $[-1, 1] \times [-1, 1]$. In all cases $a = 1$, and the tensor product test space was used.

Remark 19 For Laplace's equation with $n = 3$ we see that we obtain the best possible L^2 solution on the mesh. This is due to the fact that $\mathcal{D} = \mathcal{C}$ in this case.

4.4 On The Nature of $\mathcal{D}$

Insight can be gained into the working of the these methods by examining the problem as an L^2 projection as we have done in the preceding analysis. It is interesting to examine more closely the form that the space $\mathcal{D}$ takes. Figures 4.1 to 4.13 depict single basis functions (corresponding to the usual bilinear 'hat' basis functions) for various spaces $\mathcal{D}$ with flow direction $\mathbf{b}$ and splitting

constant C. In all examples, the diffusion coefficient $a = 1$ and the mesh spacing parameter $h = 1$.

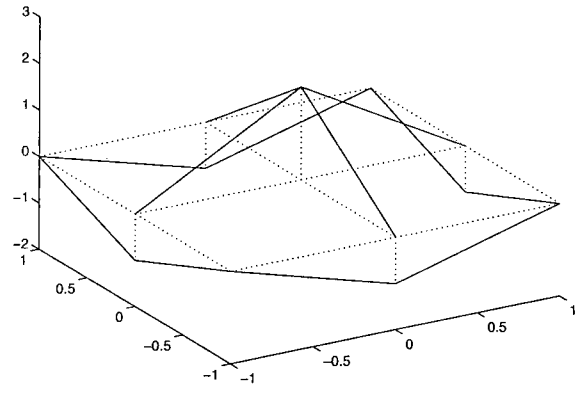


Figure 4.1: Test function boundary jumps : $\mathbf{b} = (0, 0), C = 0$

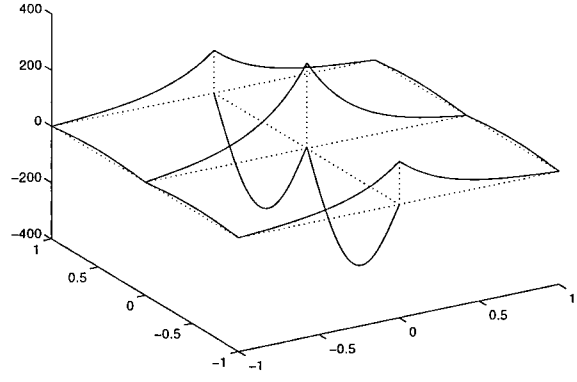


Figure 4.2: Test function boundary jumps : $\mathbf{b} = (0, 0), C = 10$

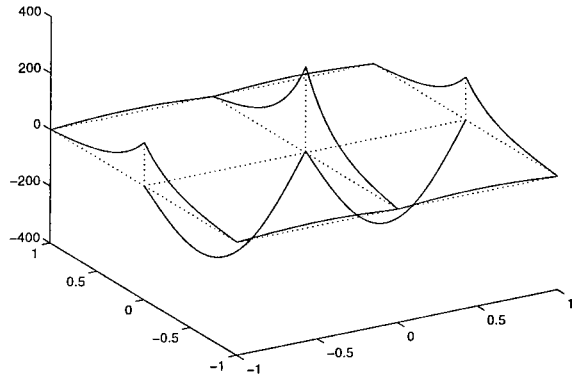


Figure 4.3: Test function boundary jumps : $\mathbf{b} = (0, 0), C = -10$

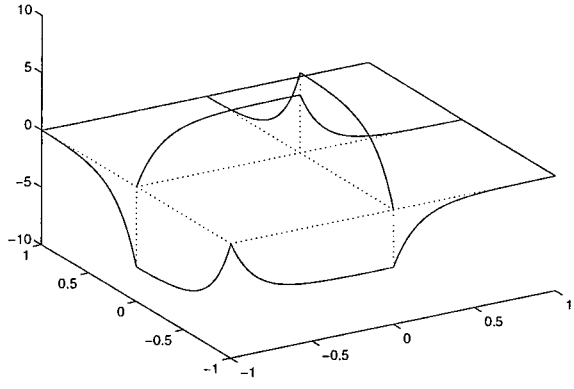


Figure 4.4: Test function boundary jumps : $\mathbf{b} = (7, 5), C = 0$

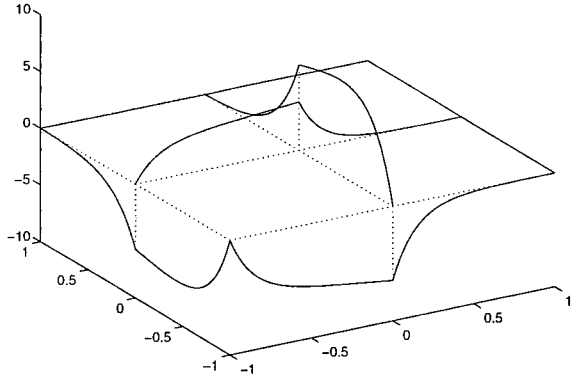


Figure 4.5: Test function boundary jumps : $\mathbf{b} = (7, 5), C = 2$

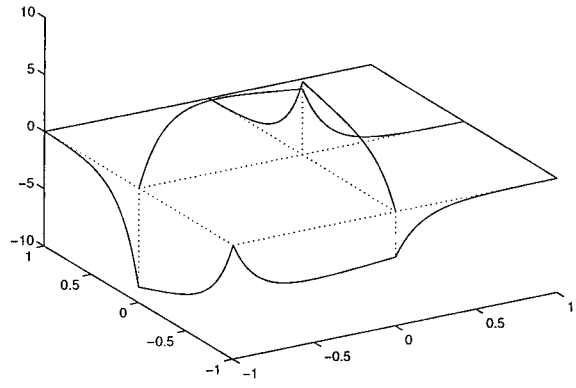


Figure 4.6: Test function boundary jumps : $\mathbf{b} = (7, 5), C = -2$

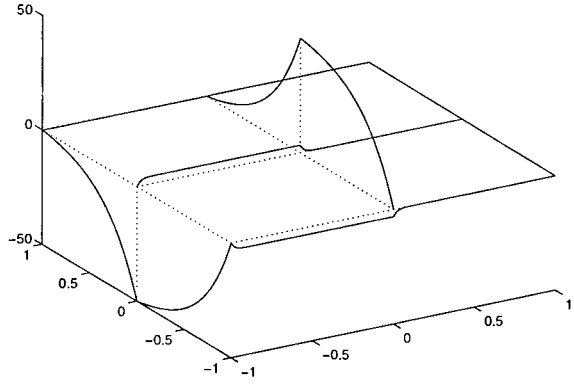


Figure 4.7: Test function boundary jumps : $\mathbf{b} = (50, 3), C = 0$

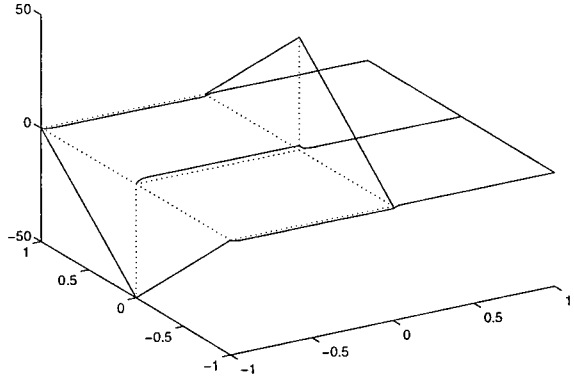


Figure 4.8: Test function boundary jumps : $\mathbf{b} = (50, 0), C = 0$

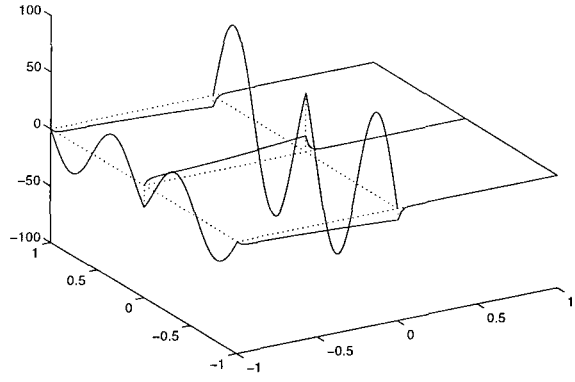


Figure 4.9: Test function boundary jumps : $\mathbf{b} = (50, 0), C = 50$

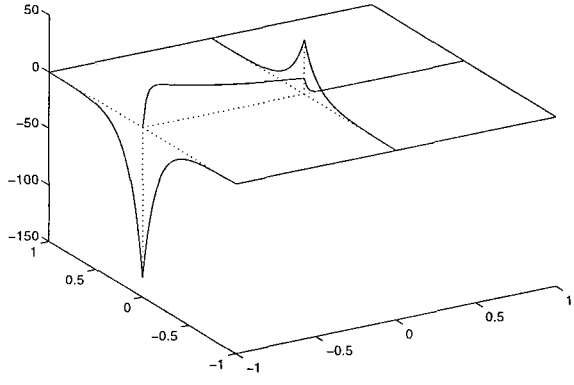


Figure 4.10: Test function boundary jumps : $\mathbf{b} = (50, 0), C = -50$

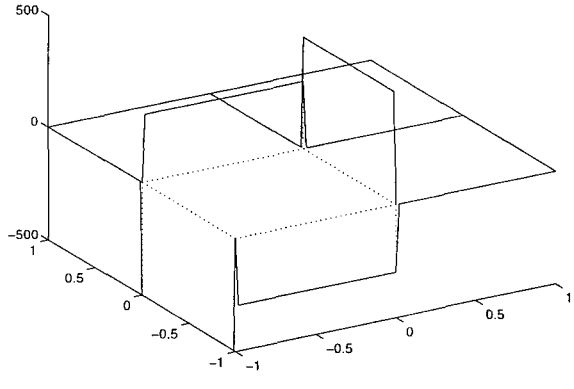


Figure 4.11: Test function boundary jumps : $\mathbf{b} = (500, 300), C = 0$

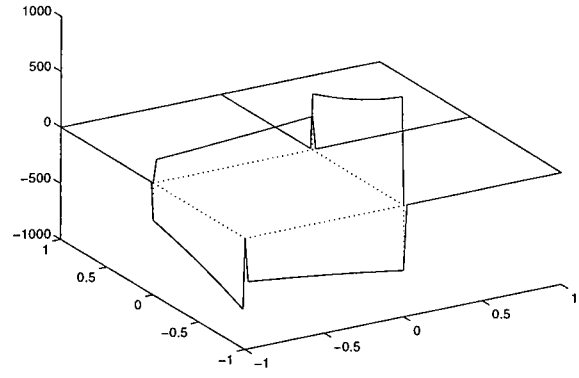


Figure 4.12: Test function boundary jumps : $\mathbf{b} = (500, 300), C = 200$

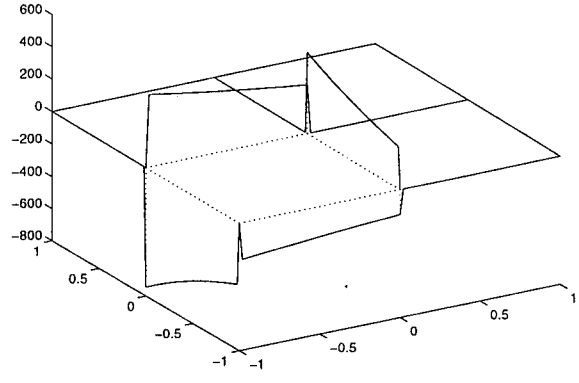


Figure 4.13: Test function boundary jumps : $\mathbf{b} = (500, 300), C = -200$

4.5 Truncation Error Analysis

In [28] a truncation error analysis is presented on a more general problem involving non constant coefficients. Presented here is a simpler estimate of the local truncation error for the problem in 2 dimensions.

This local truncation error

$$\tau_{I_h,k} = B((I_h - I)u, w_k)$$

where I_h is the bilinear interpolation operator and w_k is the local test basis function centred at node k.

In evaluating this, for simplicity of exposition we consider this integral over the 4 elements shown in figure 4.14 and calculate the bilinear form from the boundary value formulation. ie

$$B(u, w_k) = \int_\Gamma u d \, d\Gamma$$

where Γ is the mesh shown in figure 4.14 and d is the function obtained by taking the jumps in the normal derivative of w_k across Γ as defined in section 4.3.2.

For simplicity we now refer to the interpolation error $(I_h - I)u$ as e. We calculate $\tau_{I_h,k}$ as the sum of four components τ^1, τ^2, τ^3 and τ^4.

$$\tau^1 = \int_0^h e(-h,y)d(-h,y) + e(0,y)d(0,y) + e(h,y)d(h,y) \, dy$$

$$\tau^2 = \int_{-h}^0 e(-h,y)d(-h,y) + e(0,y)d(0,y) + e(h,y)d(h,y) \, dy$$

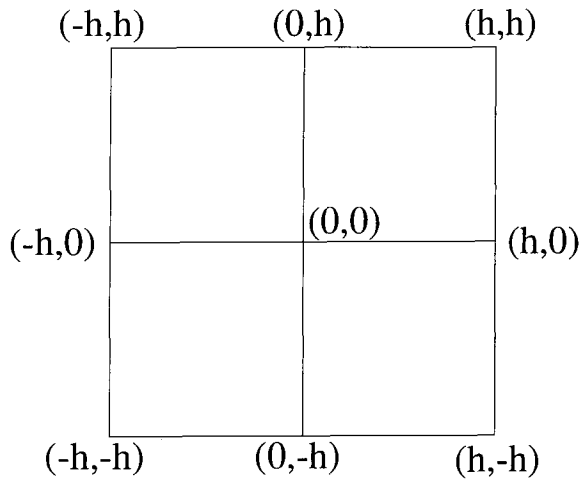


Figure 4.14: Integration region for truncation error analysis

$$\tau^3 = \int_0^h e(x,-h)d(x,-h) + e(x,0)d(x,0) + e(x,h)d(x,h)\ dx$$

$$\tau^4 = \int_{-h}^0 e(x,-h)d(x,-h) + e(x,0)d(x,0) + e(x,h)d(x,h)\ dx$$

As $\tau^i, i = 1,2,3,4$ are all of a similar form it suffices to calculate only τ^1 for general flow directions.

Firstly we note that $e(.,.)$ is zero at the mesh nodes. Hence we can define a function

$$g(x,y) = \frac{e(x,y)}{y(y-h)}.$$

Assume that e is such that $g_x(x,y)$ and $g_{xx}(x,y)$ for $y \in [0,h], x = -h, 0, h$ can be bounded above by a constant independent of h.

Then,

$$\tau^1 = \int_0^h y(y-h)[g(-h,y)d(-h,y) + g(0,y)d(0,y) + g(h,y)d(h,y)]\ dy.$$

But,

$$g(-h,y) = g(0,y) - hg_x(0,y) + \frac{h^2}{2}g_{xx}(\sigma_1,y),\ -h < \sigma_1 < 0,$$

and

$$g(h,y) = g(0,y) + hg_x(0,y) + \frac{h^2}{2}g_{xx}(\sigma_2,y),\ 0 < \sigma_2 < h.$$

So,

$$\begin{aligned}\tau^1 = \int_0^h y(y-h)[& (d(-h,y) + d(0,y) + d(h,y))g(0,y) \\ + & (d(h,y) - d(-h,y))hg_x(0,y) \\ + & (d(h,y) + d(h,y))\frac{h^2}{2}(g_{xx}(\sigma_1,y) + g_{xx}(\sigma_2,y))]\ dy.\end{aligned}$$

If we consider the standard tensor product method with zero splitting constant we can easily calculate that

$$d(-h,y) = \psi^2(y,b_2)\frac{b_1 e^{\frac{b_1 h}{a}}}{(-1+e^{\frac{b_1 h}{a}})a}$$

$$d(0,y) = -\psi^2(y,b_2)\frac{b_1\left(e^{-\frac{b_1 h}{a}} - e^{\frac{b_1 h}{a}}\right)}{\left(-1+e^{\frac{b_1 h}{a}}\right)a\left(-1+e^{-\frac{b_1 h}{a}}\right)}$$

$$d(h,y) = \psi^2(y,b_2)\frac{b_1 e^{-\frac{b_1 h}{a}}}{\left(1-e^{-\frac{b_1 h}{a}}\right)a}$$

where $\psi^2(y,b_2)$ is defined in section 4.2.

Consequently it is easy to show that,

$$(d(-h,y) + d(0,y) + d(h,y)) = 0,$$

$$|d(h,y) - d(-h,y)| = \psi^2(y,b_2)|b_1|/a$$

and

$$|d(h,y) + d(-h,y)| \leq \psi^2(y,b_2)(2/h + |b_1|/a)$$

.

Then it is immediately clear that

$$\tau^1 \leq Ch^4,$$

where $C(a,b)$ is a generic constant independent of h.

Chapter 5

Numerical Results

5.1 Introduction

We begin this chapter with a discussion of implementing the methods described so far on modern computer systems. Although so far we have considered only constant convection parameters, in practice the convective flow may vary. We carefully consider how to treat the convection term, especially in cases where the flow varies rapidly over the space of a few elements or even where the flow (the direction of the convective field) is discontinuous. We show how correct treatment of this term can give us a method of producing exact nodal solutions to the pure diffusion equation in one dimension even when the convection term is a piecewise constant function. We then present numerical results for some standard test problems. Results are given for a variety of both zero and non–zero splitting constant methods.

5.2 Computer Implementation

It has long been thought impractical to employ the exponential test functions in one dimension due to the high order quadrature formulae that have to be used to calculate the large gradients involved with high mesh Péclet numbers. We have tried three different methods of coping with this.

Firstly, integrals corresponding to terms in the bilinear form can be calculated explicitly and 'hard coded' into the program, thus leaving the relatively inexpensive and easier task of evaluating them. We have successfully employed this method for both one and two dimensional problems. Alternatively the terms in the bilinear form can be calculated as boundary integrals[18] by explicitly calculating the jumps of the test functions across element boundaries and then performing standard numerical integration techniques. A third

method which I have found to be the simplest (though still accurate) is to calculate the terms in the bilinear form via standard numerical integration using many integration points.

We note that due to the widespread availability of parallel computers, we are now prepared to spend a long time generating the finite element matrices accurately in order to obtain a good solution on a relatively small mesh. The matrix assembly process is entirely parallelisable and needs no communication between the different processors/computers.

Solution of the final matrix equation is found either directly by Gaussian elimination or iteratively by Gauss-Seidel [43]. We note here that in our experience the Gauss-Seidel method converges in surprisingly few iterations when applied to the matrix equations produced by these methods which suggests that the method is well conditioned. This is in stark contrast to the matrix systems produced by classical methods for the convection–diffusion equation which invariably are very ill–conditioned.

We remark here on another method we have successfully employed to reduce oscillations in the solution to some problems. Problems involving both very high convective parameters and discontinuous boundary conditions give rise to large oscillations in the finite volume solution (and hence in our zero splitting constant methods). Although, as discussed earlier, employing a test space with a nonzero splitting constant can vastly reduce these problems, it is also possible to smooth out the solution by using a test space with zero splitting constant but which satisfies the homogeneous adjoint equation but with a larger diffusion term. This is in some way similar (but perhaps more natural) to adding 'artificial' diffusion into the original equation. Results are not presented here for this method.

5.3 Numerical Treatment of the Convection Term

As a reminder, we state the problem we are solving,

$$-\nabla.(a\nabla u) + \nabla.(\mathbf{b}u) = f \quad \text{in} \quad \Omega \subset R^n. \tag{5.1}$$

Most finite element methods take a piecewise constant approximation to the function $\mathbf{b}$ (or some other simple approximation) when solving this equation. This is acceptable if $\mathbf{b}$ does not vary too rapidly. If however $\mathbf{b}$ is truly a piecewise constant function then great care must be taken to evaluate this term. In the weak form, this term becomes $(\nabla.(\mathbf{b}U), w)$.

This term should be integrated by parts to give $-(\mathbf{b}.\nabla w, U)$ so that approximation of $\mathbf{b}$ does not lose vital information about its gradient.

In fact if this integration by parts is not performed, and if we position element boundaries along the discontinuity, we will instead obtain the solution of the vastly different equation

$$-\nabla.(a\nabla u) + \mathbf{b}.\nabla u = f \quad \text{in} \quad \Omega \subset R^n. \tag{5.2}$$

To stress the important differences between these two forms we state and plot the exact solutions to a one dimensional test problem in both cases. The problem is posed on $[0,1]$ with boundary conditions $u(0) = 0$ and $u(1) = 1$. We chose a diffusion coefficient $a = 1$ and let the convection parameter take

the discontinuous form

$$b(x) = \begin{cases} 0 & for \quad 0 \le x < (1-d)/2, \\ \frac{1}{d} & for \quad (1-d)/2 \le x \le (1+d)/2, \\ 0 & for \quad (1+d)/2 < x \le 1. \end{cases}$$

The solution in both cases takes the form

$$u(x) = \begin{cases} \frac{2Ax}{1-d} & for \quad 0 \le x < (1-d)/2, \\ \frac{B\exp((1-d)/(2d)) - A\exp((d+1)/(2d)) + (A-B)\exp(x/d)}{\exp((1-d)/(2d)) - \exp((d+1)/(2d))} & for \quad (1-d)/2 \le x \le (1+d)/2, \\ \frac{2x-1-d-2Bx+2B}{1-d} & for \quad (1+d)/2 < x \le 1. \end{cases}$$

where A and B are constants which differ in the two forms.

Case 1

$$-u'' + (bu)' = 0.$$

$$A = \frac{(d-1)\exp((1-d)/(2d))}{3d\exp((1-d)/(2d)) - \exp((1-d)/(2d)) - d\exp((d+1)/(2d)) - \exp((d+1)/(2d))},$$

$$B = \frac{(-d\exp((d+1)/(2d)) + 2d\exp((1-d)/(2d)) - \exp((d+1)/(2d)))}{3d\exp((1-d)/(2d)) - \exp((1-d)/(2d)) - d\exp((d+1)/(2d)) - \exp((d+1)/(2d))},$$

$$= 1 - A.$$

Case 2

$$-u'' + b(u)' = 0.$$

$$A = \frac{-(d-1)\exp((1-d)/(2d))}{d\exp((d+1)/(2d)) + \exp((d+1)/(2d)) - 3d\exp((1-d)/(2d)) + \exp((1-d)/(2d))},$$

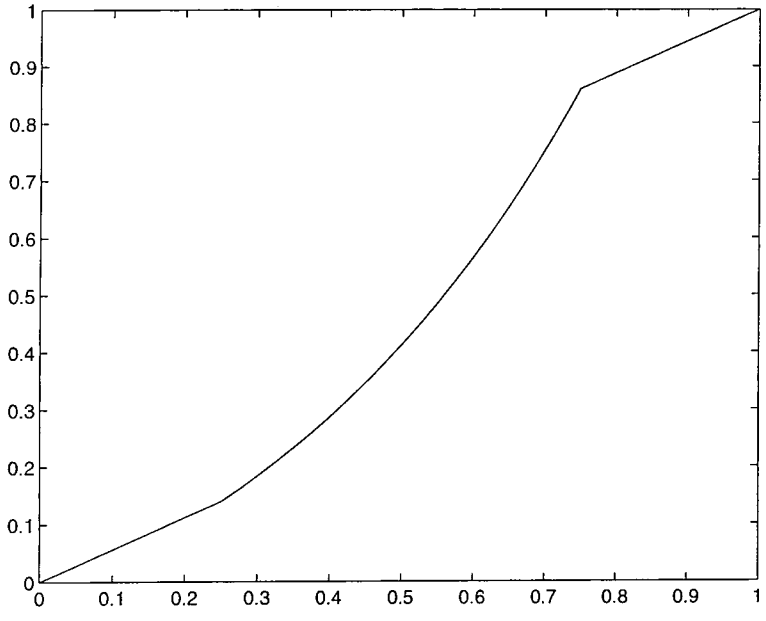


Figure 5.1: Case 1 $d = 0.5$

$$B = \frac{(-3d\exp((1-d)/(2d)) + 2d\exp((d+1)/(2d)) + \exp((1-d)/(2d)))}{d\exp((d+1)/(2d)) + \exp((d+1)/(2d)) - 3d\exp((1-d)/(2d)) + \exp((1-d)/(2d))}.$$

In figures 5.1 to 5.8 plot the solution for $d = 0.5, d = 0.2, d = 0.1, d = 0.01$ and for both forms of the equation.

It is enlightening to investigate exactly how a numerical scheme will differ when solving these two different forms of the equation. For simplicity we pose the problem on $[-1, 1]$ with $b = \beta_1$ when $x < 0$ and $b = \beta_2$ when $x > 0$.

When discretised by the standard Galerkin finite element method with two elements the term $(bu)'w$ becomes (writing the approximation to $u(x)$ at $x_i = ih - 1$ as U_i)

$$-\frac{\beta_1}{2}U_0 + \frac{(\beta_1 - \beta_2)}{2}U_1 + \frac{\beta_2}{2}U_2.$$

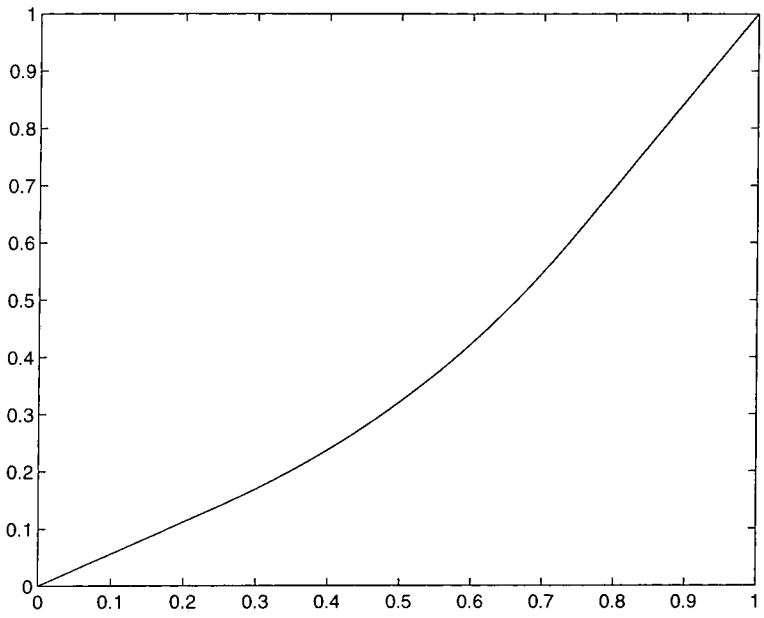


Figure 5.2: Case 2 $d = 0.5$

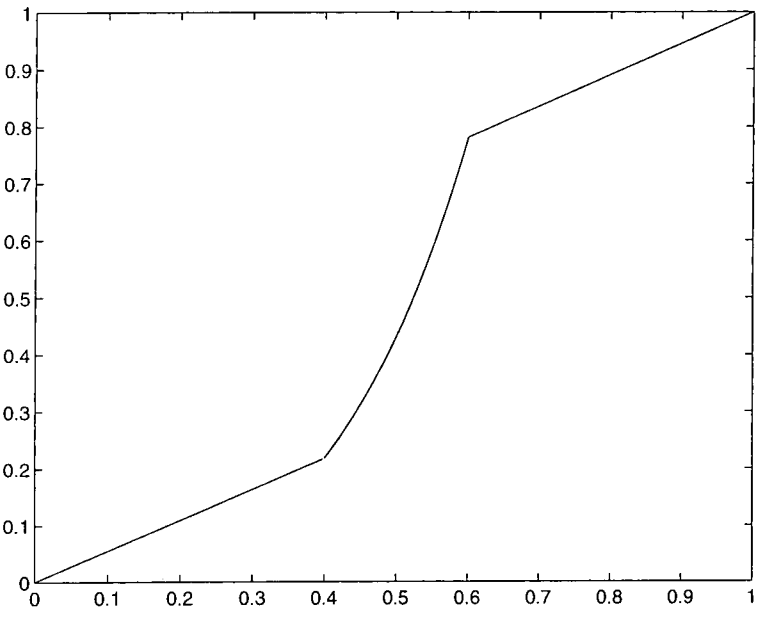


Figure 5.3: Case 1 $d = 0.2$

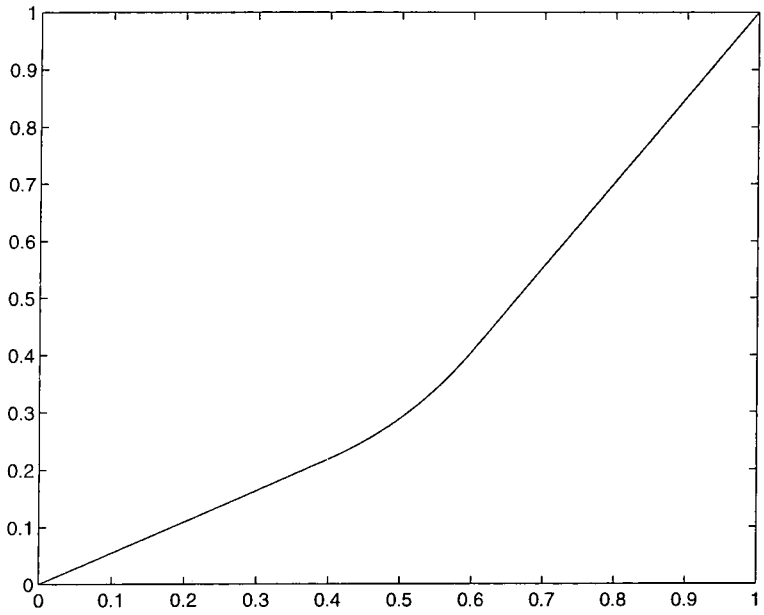


Figure 5.4: Case 2 $d = 0.2$

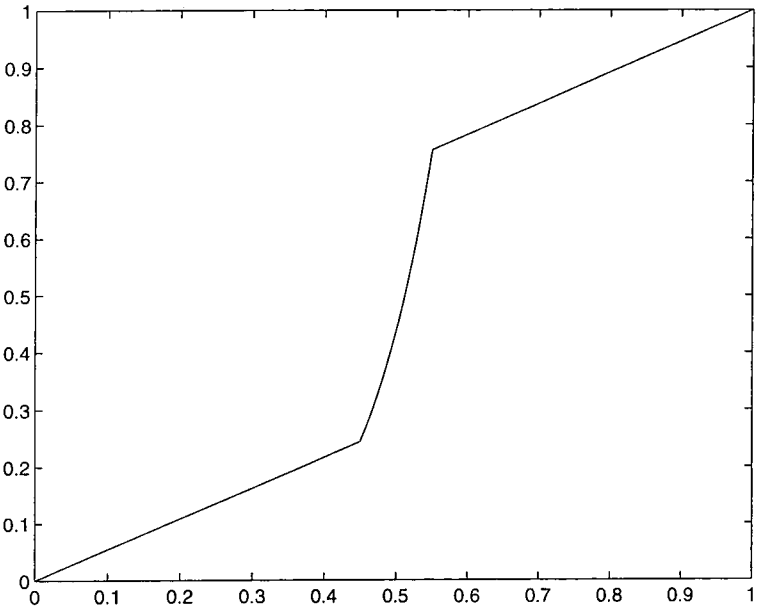


Figure 5.5: Case 1 $d = 0.1$

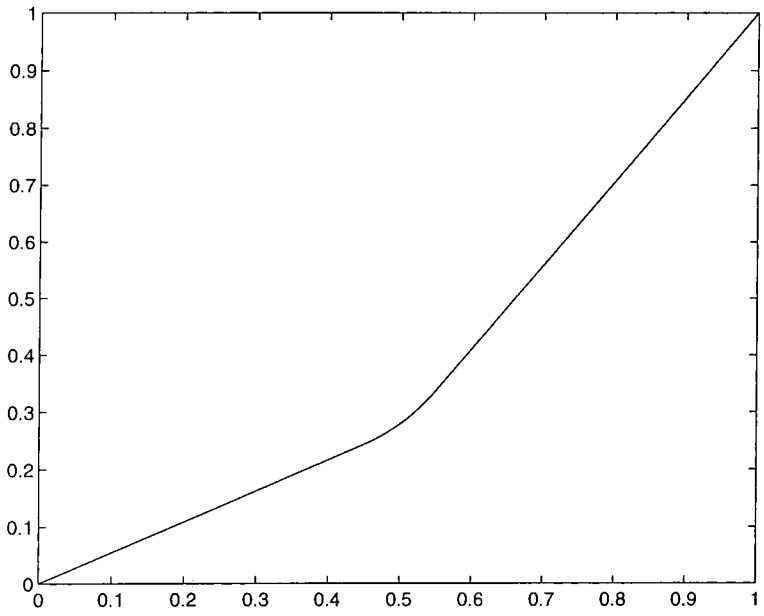


Figure 5.6: Case 2 $d = 0.1$

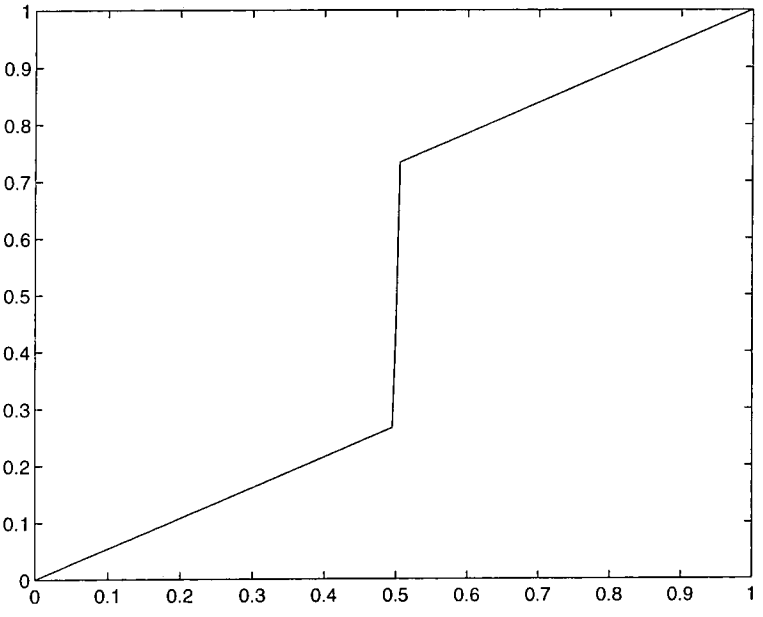


Figure 5.7: Case 1 $d = 0.01$

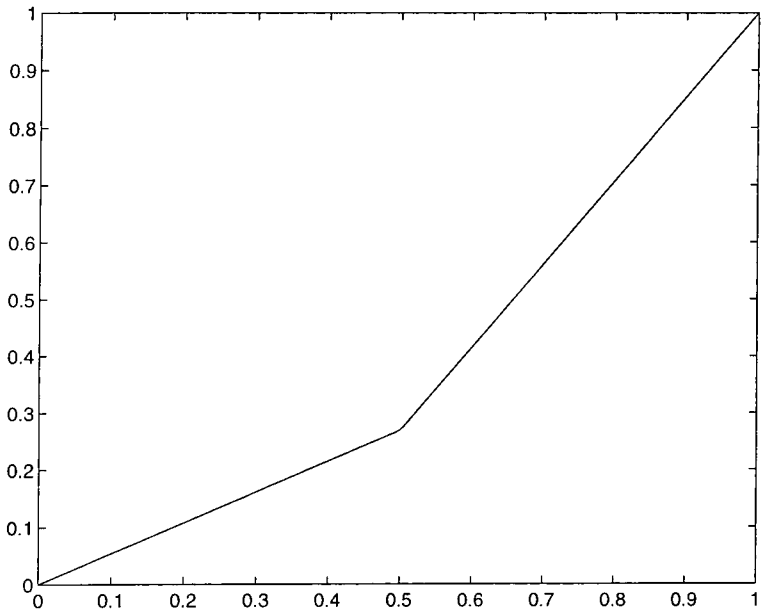


Figure 5.8: Case 2 $d = 0.01$

However $-buw'$ becomes

$$-\frac{\beta_1}{2}U_0 + \frac{(\beta_2 - \beta_1)}{2}U_1 + \frac{\beta_2}{2}U_2.$$

Of course if $\beta_1 = \beta_2$ these two forms are identical, but otherwise they differ greatly. Similar differences appear in other discretisations.

5.4 The limit of no diffusion with discontinuous convection parameters

If we calculate the difference equations resulting from discretising the convection diffusion equation in one dimension by the method described in this thesis, we can generate a scheme for the pure convection equation by letting

the diffusion parameter tend to zero.

As in the last section we shall consider a problem over two elements where the convection term b is piecewise constant over each element. Discretisation of $(bu)'w$ yields the standard cell vertex finite volume approximation

$$\beta_0 U_1 - \beta_0 U_0.$$

However, discretisation of $-buw'$ yields the subtly different

$$\beta_1 U_1 - \beta_0 U_0.$$

Let us consider the equation

$$(bu)' = 1, \quad u(0) = 0.$$

where $b = \beta_i$ when $x \in (i, i+1)$. We shall solve for U_i our approximation to $u(x_i)$ where $x_i = i$. The exact solution is discontinuous.

The finite volume method would produce

$$U_0 = 0, \quad U_1 = 1/\beta_0, \quad U_2 = 1/\beta_0 + 1/\beta_1, \quad U_3 = 1/\beta_0 + 1/\beta_1 + 1/\beta_2 \quad \ldots$$

However our method would produce

$$U_0 = 0, \quad U_1 = 1/\beta_1, \quad U_2 = 2/\beta_2, \quad U_3 = 3/\beta_3 \quad \ldots$$

This solution is nodally exact in that it produces the right–sided limit of the solution at x_i. The exact left–sided limit can be found from these values by

using the finite volume discretisation inside each of the unit intervals.

5.5 Semiconductor Test Problems

Presented here are two test problems similar to the one dimensional example described in the last section involving discontinuous convective fields. Both problems involve two regions of pure diffusion separated by a region of high convection with a little diffusion. They are:

5.5.1 Test Problem One

We have

$$\Omega = \{(x,y)|0 < x < 1, 0 < y < 1\} \tag{5.3}$$

with a convective field of

$$\mathbf{b}(x,y) = (c(x),0)^T, \tag{5.4}$$

where

$$c(x) = \begin{cases} 0 \quad for \quad 0 \leq x < \frac{1-d}{2}, \\ \frac{1}{d} \quad for \quad \frac{1-d}{2} \leq x \leq \frac{1+d}{2}, \\ 0 \quad for \quad \frac{1+d}{2} < x \leq 1. \end{cases}$$

The boundary conditions imposed are $u(0,y) = 0, u(1,y) = 1$ with homogenous Neumann conditions on the other two edges.

In figures 5.9 to 5.17 we present results for both forms of the equation (with the convection parameter inside and outside the derivative) when $a = 1$, $d = 0.2, d = 0.05, d = 0.01$ with a mesh spacing of $h_1 = 1/3, h_2 = 1/20$. In all cases we choose a splitting constant of zero. We shall refer to the two different forms of the problem as Case 1 and Case 2 as in the last section.

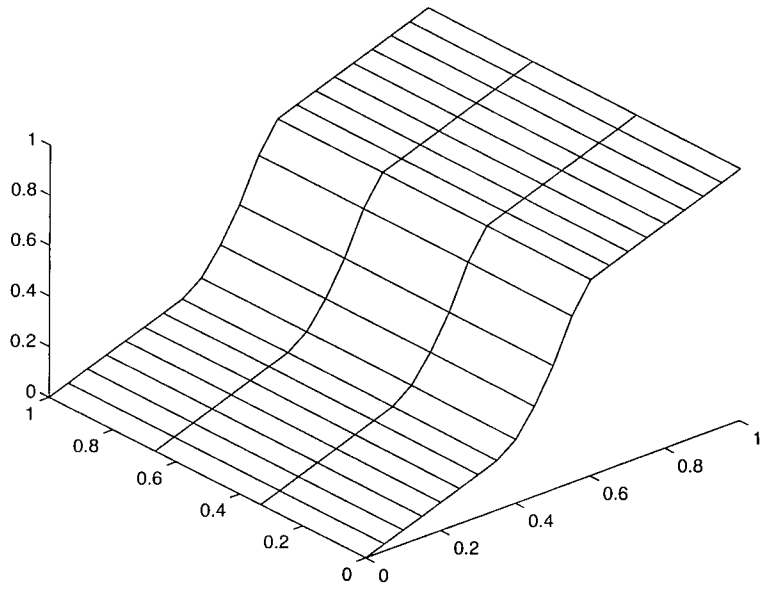


Figure 5.9: Solution to semiconductor test problem one (Case 1) with $d = 0.2, a = 1$

We also present the solution of Case 1 for these values of d, h_1 and h_2 but with $a = 0.001$.

We note upon the high accuracy of these solutions even though the layer of convection is not even resolved by the mesh in some cases.

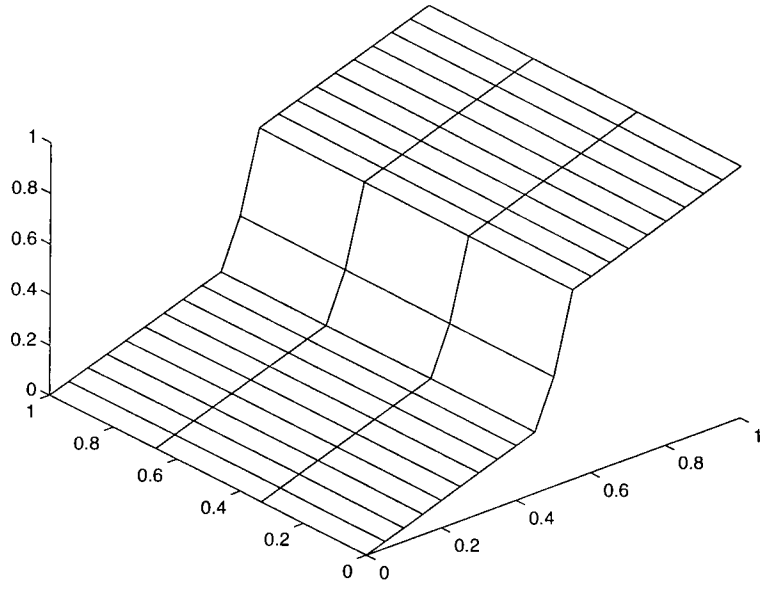


Figure 5.10: Solution to semiconductor test problem one (Case 1) with $d = 0.05, a = 1$

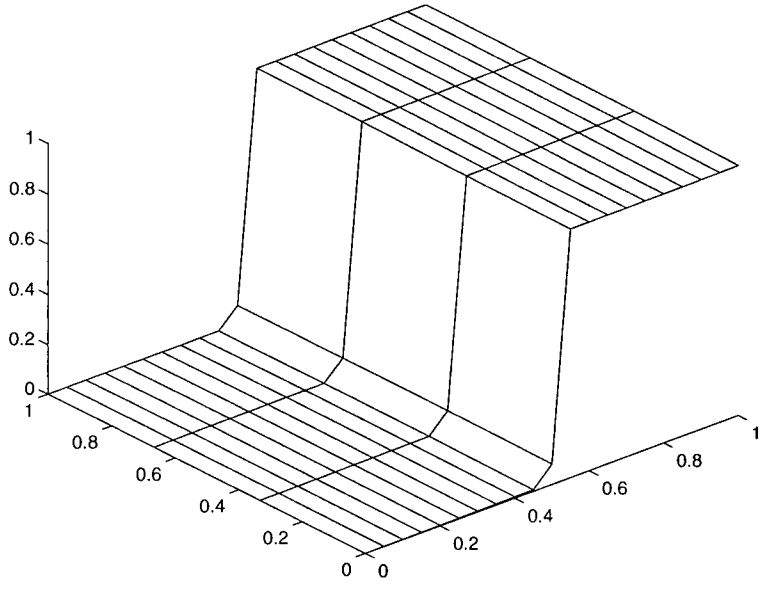


Figure 5.11: Solution to semiconductor test problem one (Case 1) with $d = 0.01, a = 1$

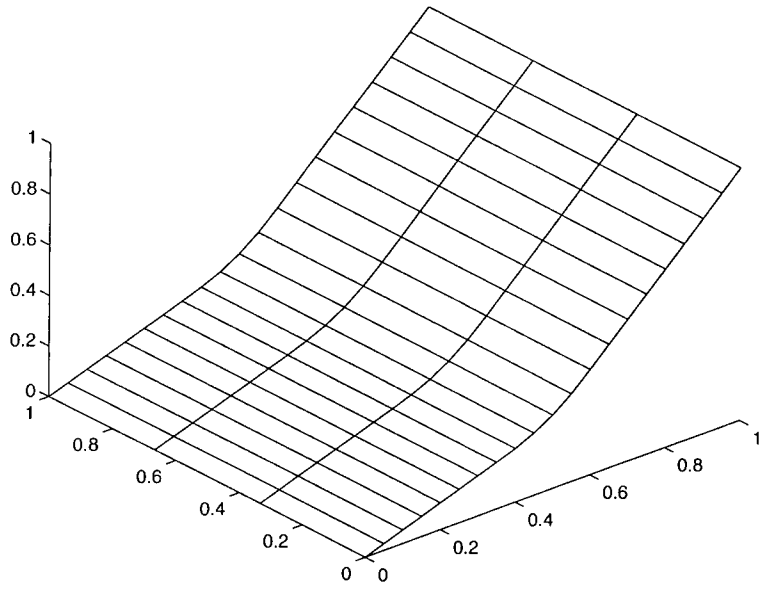


Figure 5.12: Solution to semiconductor test problem one (Case 2) with $d = 0.2, a = 1$

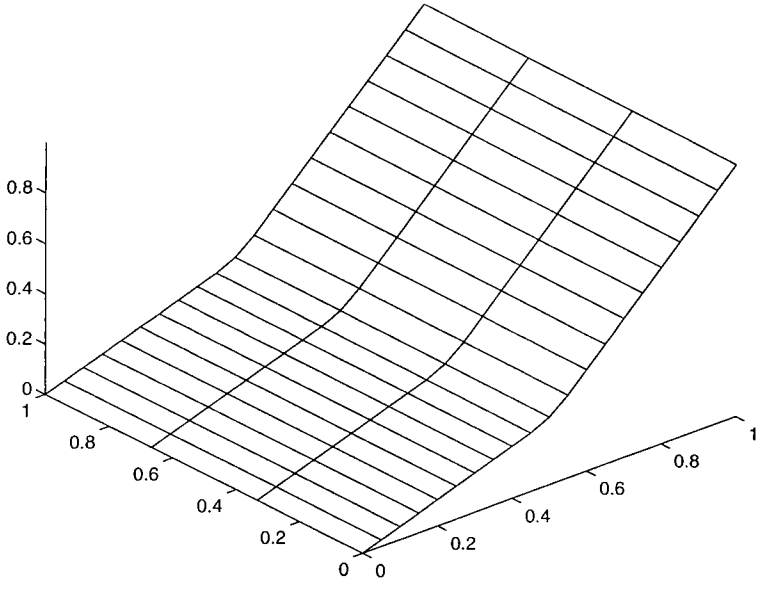


Figure 5.13: Solution to semiconductor test problem one (Case 2) with $d = 0.05, a = 1$

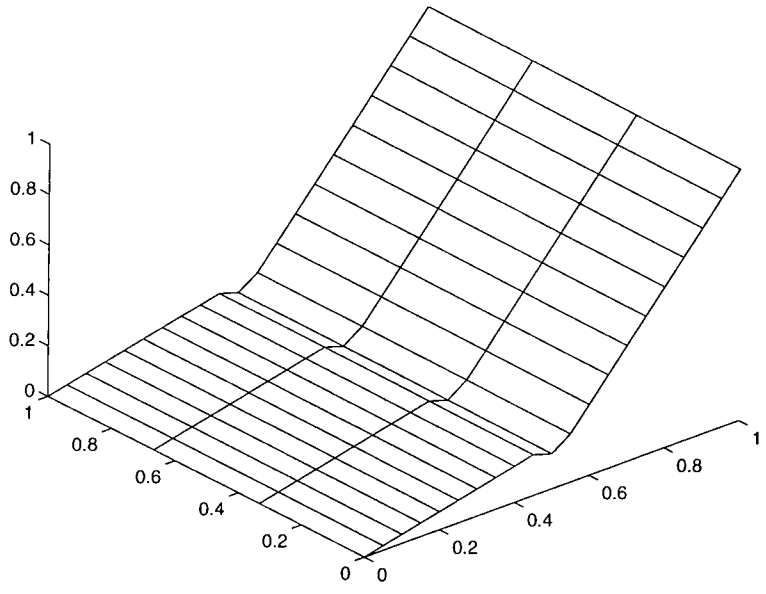


Figure 5.14: Solution to semiconductor test problem one (Case 2) with $d = 0.01, a = 1$

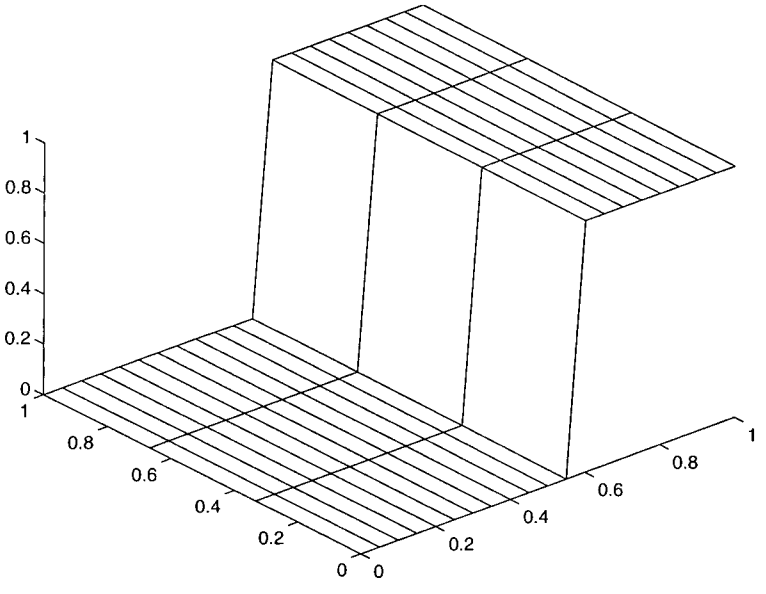


Figure 5.15: Solution to semiconductor test problem one (Case 1) with $d = 0.2, a = 0.001$

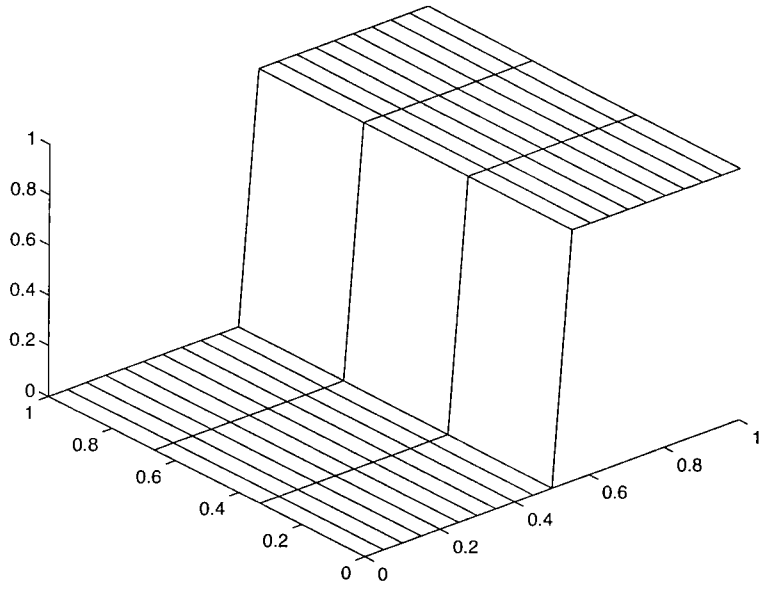


Figure 5.16: Solution to semiconductor test problem one (Case 1) with $d = 0.05, a = 0.001$

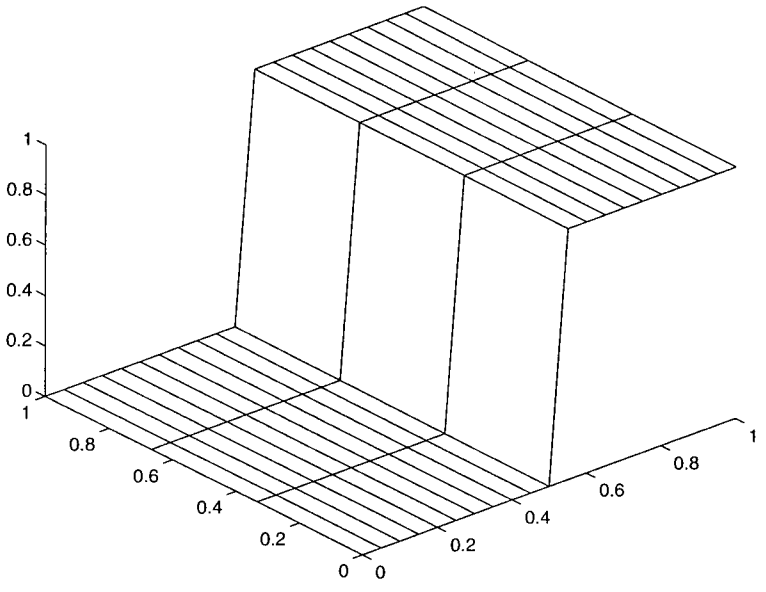


Figure 5.17: Solution to semiconductor test problem one (Case 1) with $d = 0.01, a = 0.001$

5.5.2 Test Problem Two

We have

$$\Omega = \{(x, y) \quad | \quad 0 < x < 1, 0 < y < 1\} \tag{5.5}$$

with a convective field of

$$\mathbf{b}(x, y) = (c(r)x/r, c(r)y/r)^T, \tag{5.6}$$

where $r^2 = x^2 + y^2$, and

$$c(r) = \begin{cases} 0 \quad for \quad 0 \leq r < \frac{1-2d}{4}, \\ \frac{1}{d} \quad for \quad \frac{1-2d}{4} \leq x \leq \frac{1+2d}{4}, \\ 0 \quad for \quad \frac{1+2d}{4} < x \leq 1. \end{cases}$$

The boundary conditions imposed are of homogenous Neumann type everywhere except for Dirichlet data of $u(x, y) = 1$ for $x < 0.25, y = 0$ and $x = 0, y < 0.25$.

In figures 5.18 to 5.23 we present results for $d = 0.2, d = 0.05, d = 0.01$ with a mesh spacing of $h_1 = 1/20, h_2 = 1/20$ and for both $a = 1$ and $a = 0.00001$.

We note that this problem is not radially symmetric. The Dirichlet conditions are imposed where the convection term is nonzero leading to a (small) parabolic layer. We also note that due to the Neumann condition we cannot hope to calculate the solution on the outflow boundary accurately unless we capture the region of nonzero convection accurately with the mesh. The results we show here do not attempt to resolve this region accurately but serve to give qualitative information about the form of the solution.

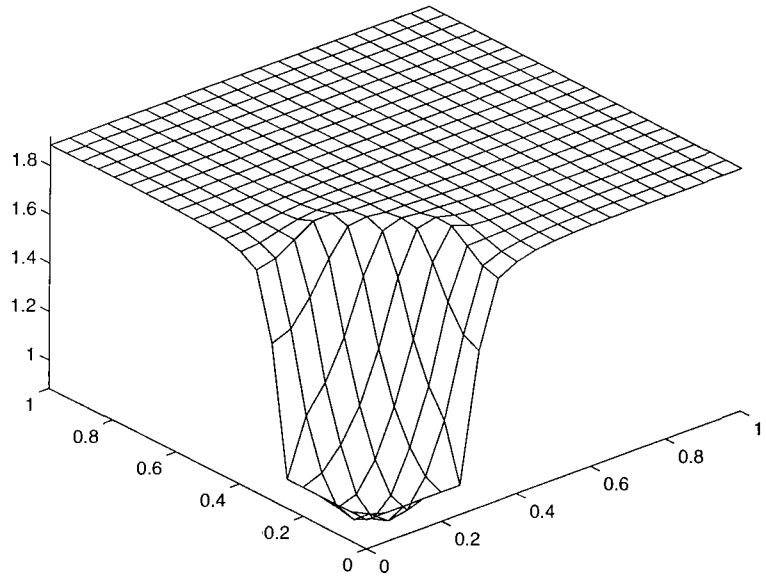


Figure 5.18: Solution to semiconductor test problem two with $d = 0.2, a = 1$

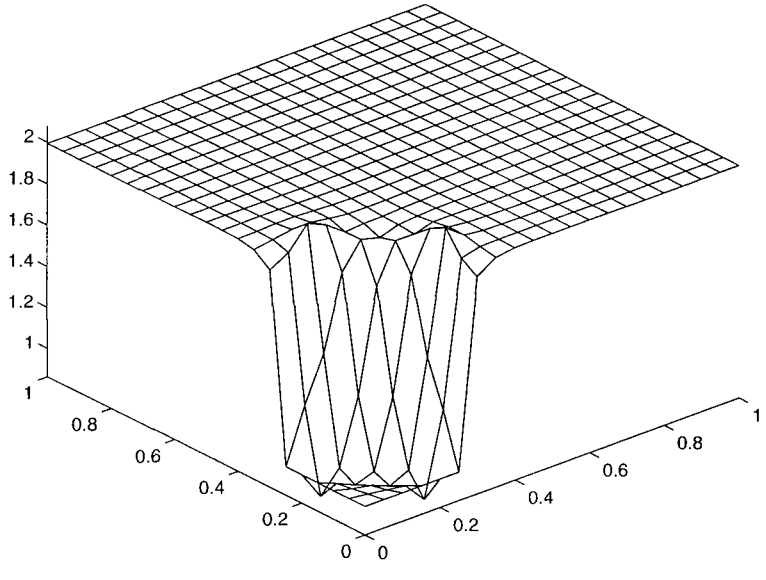


Figure 5.19: Solution to semiconductor test problem two with $d = 0.05, a = 1$

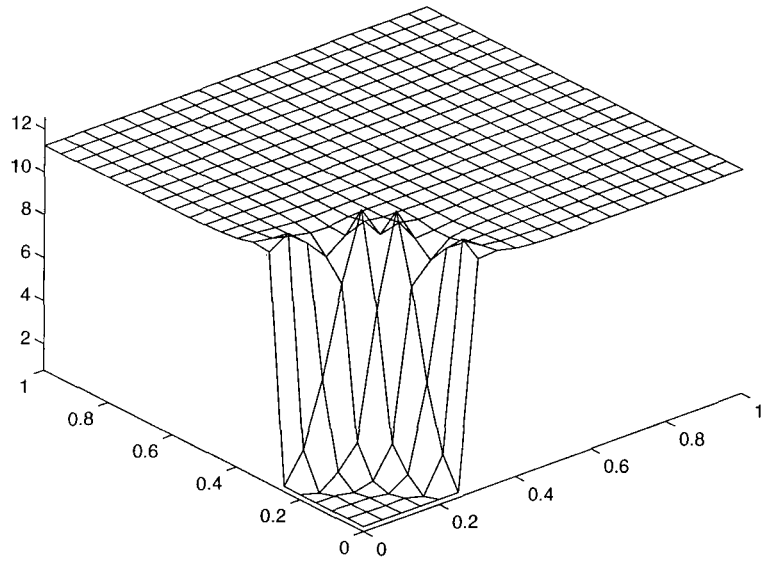


Figure 5.20: Solution to semiconductor test problem two with $d = 0.01, a = 1$

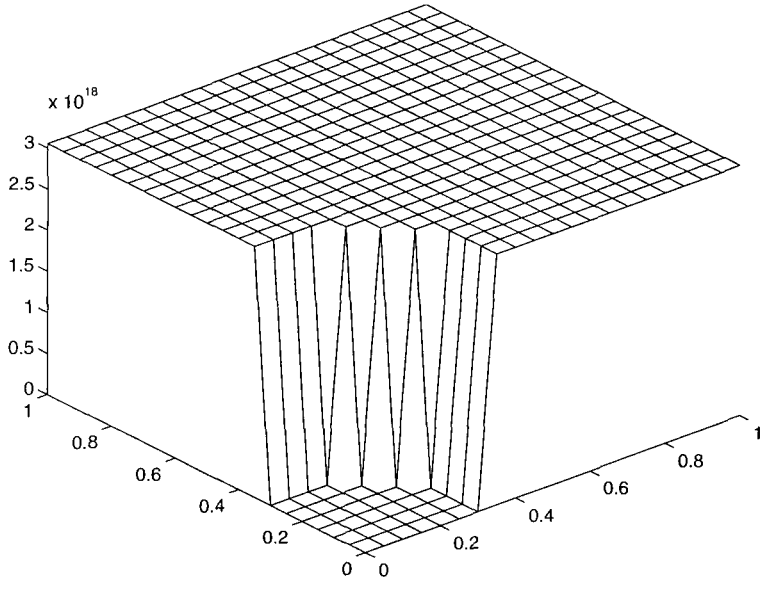


Figure 5.21: Solution to semiconductor test problem two with $d = 0.2, a = 0.00001$

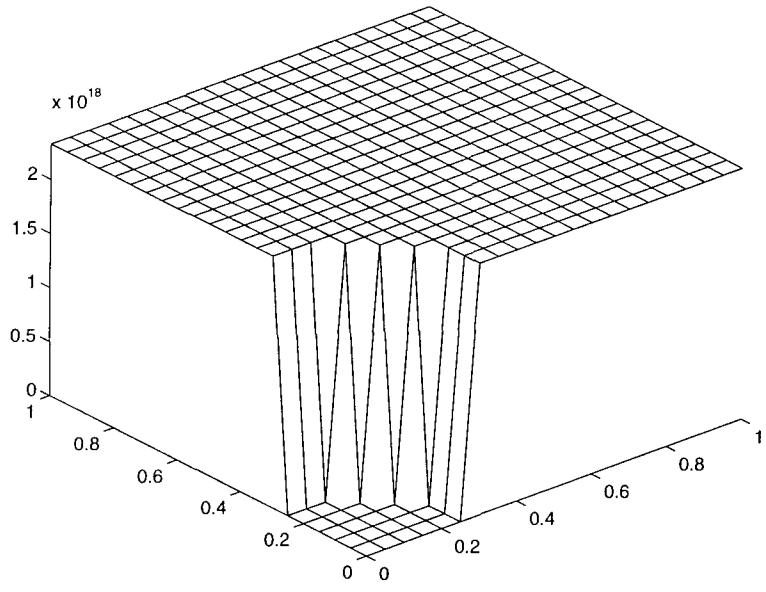


Figure 5.22: Solution to semiconductor test problem two with $d = 0.05, a = 0.00001$

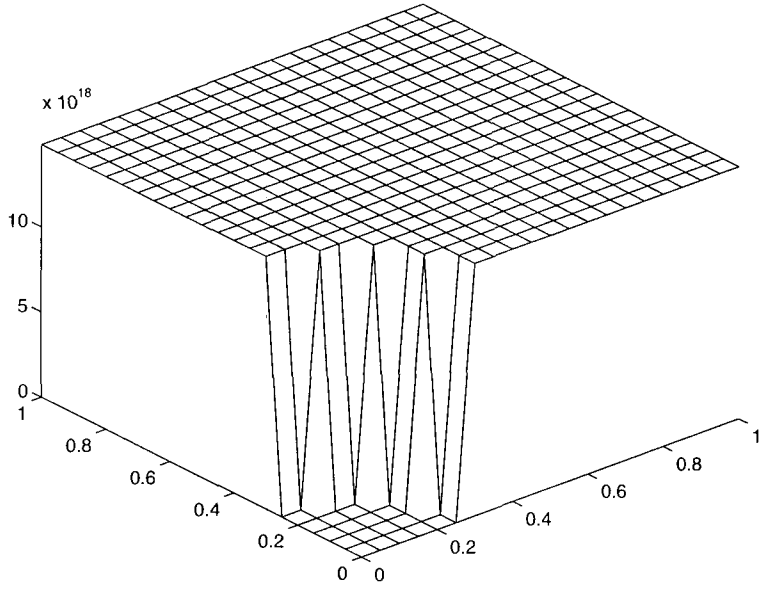


Figure 5.23: Solution to semiconductor test problem two with $d = 0.01, a = 0.00001$

5.6 IAHR/CEGB Test Problems

Presented here are the numerical results for the standard test problems devised by the Third Meeting of the International Association for Hydraulic Research Working Group on Refined Modelling of Flow[42].

For these test problems we have

$$\Omega = \{(x, y) | -1 < x < 1, 0 < y < 1\} \tag{5.7}$$

with convective field

$$\mathbf{b}(x, y) = (2y(1 - x^2), -2x(1 - y^2))^T, \tag{5.8}$$

and with $a = 1 \times 10^{-1}, 1 \times 10^{-2}, 2 \times 10^{-3}, 1 \times 10^{-6}$. All solutions are obtained with the zero splitting constant test space.

For the results obtained by other methods for these problems see [42] and [27]. There is little to say about our results presented in figures 5.25 to 5.30 other than to remark on their exceptional accuracy - even on a 10 by 5 mesh. We draw attention to the high quality of the solution even away from the outflow boundary. We draw attention to the accuracy of the method over the whole range of mesh Péclet numbers. We remark here that the oscillations that occur at the internal layer in the first test problem can be very much reduced by using a test space generated with a splitting constant.

5.6.1 Test Problem One

For this problem the inlet boundary condition along

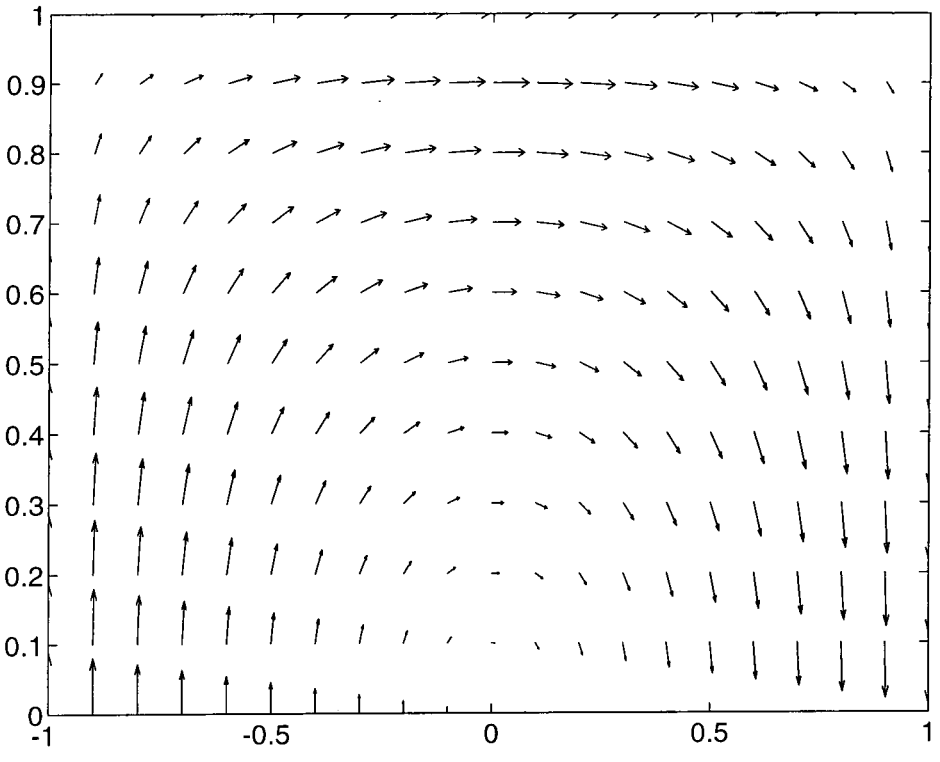


Figure 5.24: Streamlines for the IAHR/CEGB test problems

$-1 \leq x \leq 0,\ y = 0$ is given by

$$U(x,0) = 1 + \tanh[20x + 10]. \tag{5.9}$$

The boundary condition on the tangential boundaries,

$x = -1, y = 1$ and $x = 1$ is given by $U = 1 - \tanh 10 = 0(8d.p.)$, and a homogeneous Neumann boundary condition is placed on the outflow boundary.

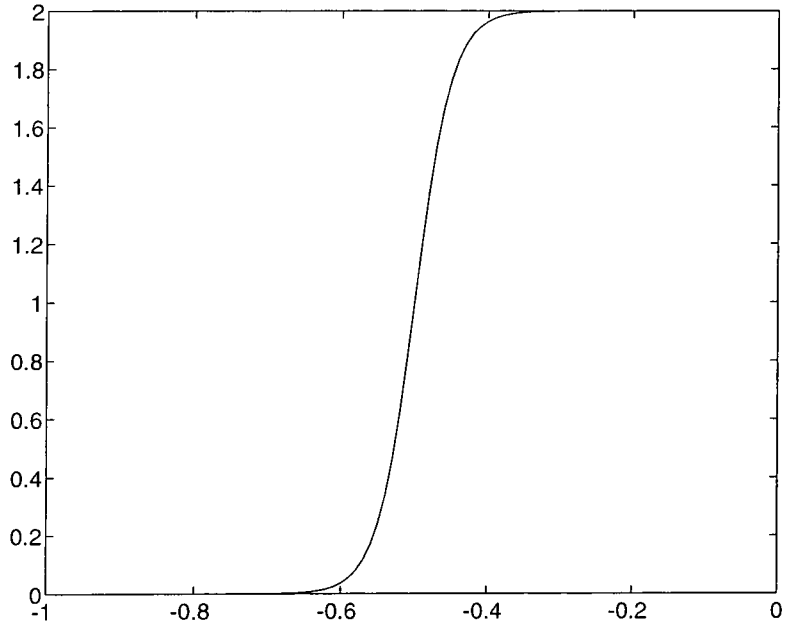
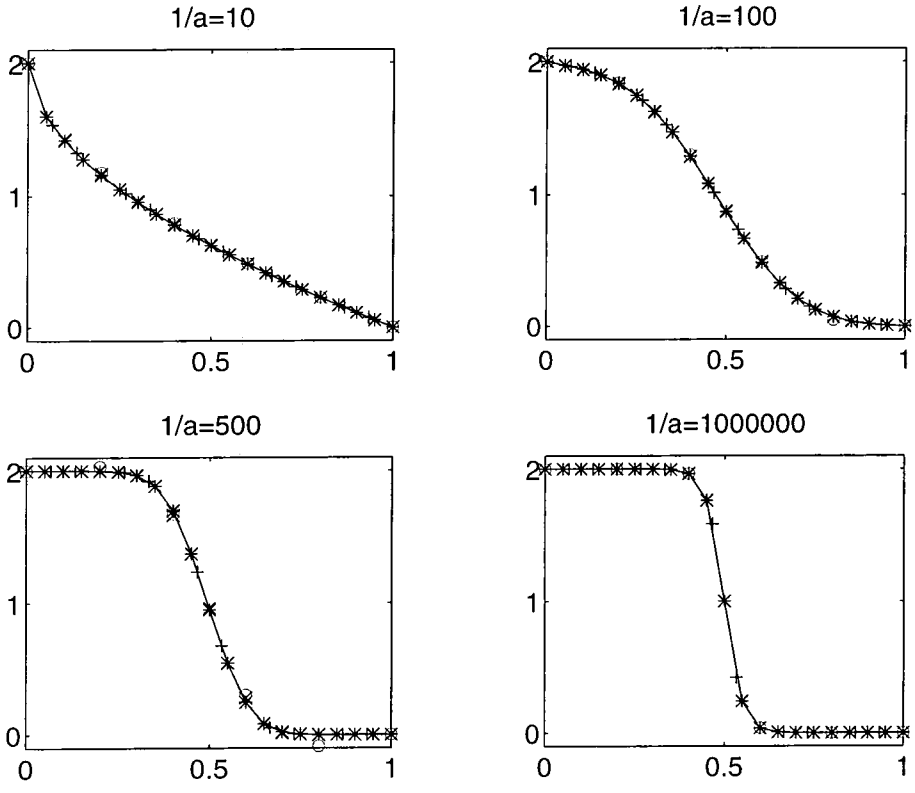


Figure 5.25: Inflow profile for IAHR/CEGB test problem 1

5.6.2 Test Problem Two

In this second test problem the boundary conditions are $U = 0$ on all boundaries except for the outflow which is left as before, and on $x = 1$, where $U = 100$.



∘ h=0.2, × h=0.1, + h=0.0667, * h=0.05

Figure 5.26: Outflow profiles for IAHR/CEGB test problem 1

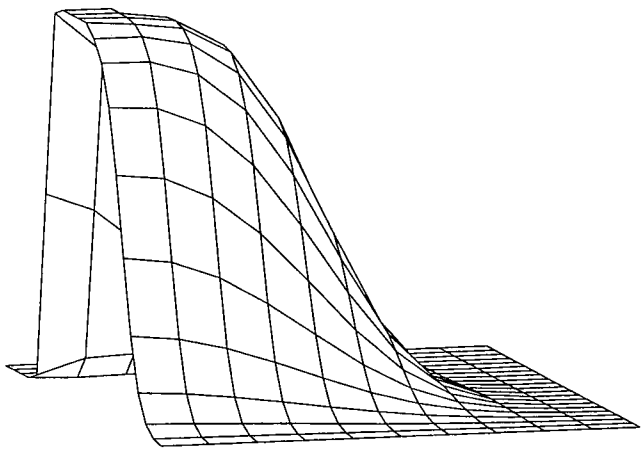


Figure 5.27: CEGB test problem 1 with $a = 0.01, h = 0.1$

5.7 Parabolic Layer Problems

We present here in example to show how the use of a 'splitting' constant can dramatically help in resolving parabolic boundary layers. Parabolic boundary layers can form when we have a flow which is parallel to a Dirichlet boundary. The test problem presented is taken from [13] where the solution is found with central differencing based on a specially grades mesh near the boundaries.

The test problem is posed on a domain $\Omega = [0,1] \times [0,1]$ with flow field

$$\mathbf{b} = (-1 - x^2 - y^2, 0)^T,$$

a diffusion parameter $a = 0.0001$ and with $f = 0$. The boundary conditions

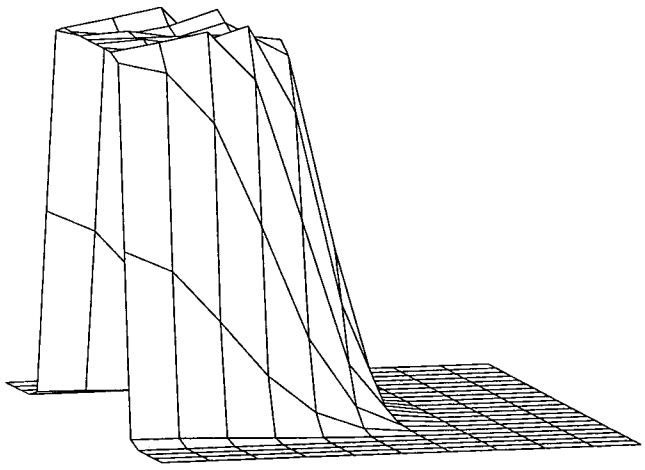


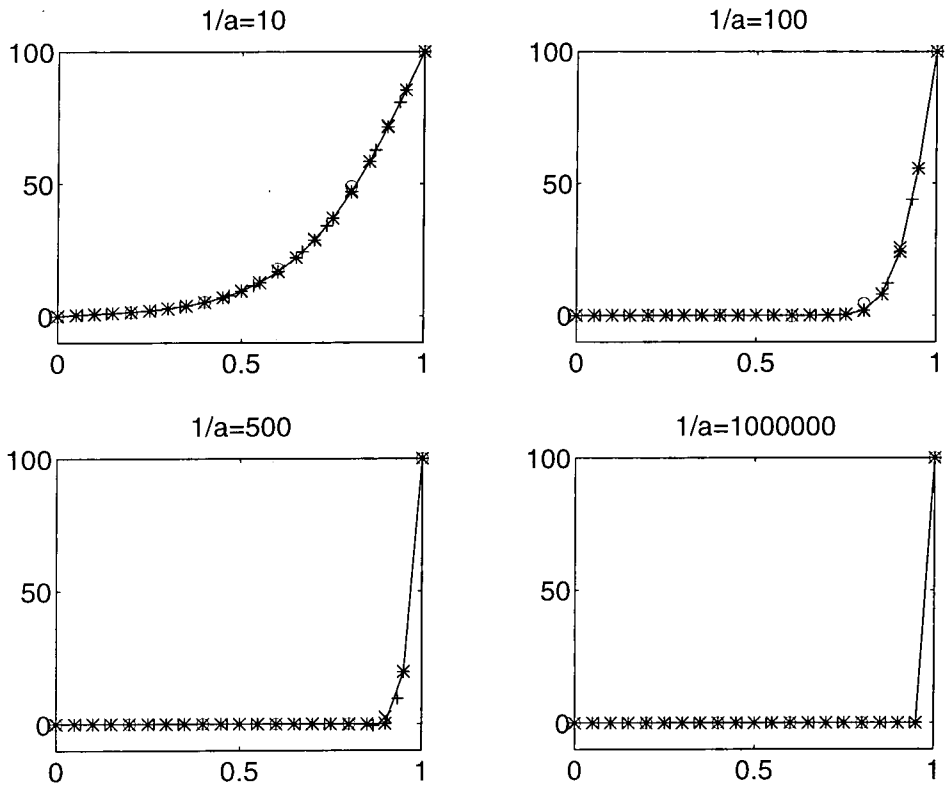
Figure 5.28: CEGB test problem 1 with $a = 0.000001, h = 0.1$

are

$$u(x,0) = x^3; u(x,1) = x^2; u(0,y) = 0; u(1,y) = 1.$$

The solution to this problem is approximately 1 everywhere except 'close' to the boundaries at $x = 0, y = 0$ and $y = 1$. At the latter two boundaries a parabolic layer forms.

We solve the problem on 6×6 and 10×10 regular meshes. The problem is solved twice on each mesh - with a zero and nonzero splitting constant and are presented in figures 5.31 to 5.34. The value of **b** used is calculated as an average of the four nodal values for each element. Note that in each figure, the labels on the 'x' and 'y' axes are values for $i + 1$ and $j + 1$ rather than x_i and y_j. We note that the nodal errors are effectively zero when we use an appropriate splitting constant but when we have a zero splitting constant, overshoots appear in the vicinity of the parabolic layers.



∘ h=0.2, × h=0.1, + h=0.0667, * h=0.05

Figure 5.29: Outflow profiles for IAHR/CEGB test problem 2

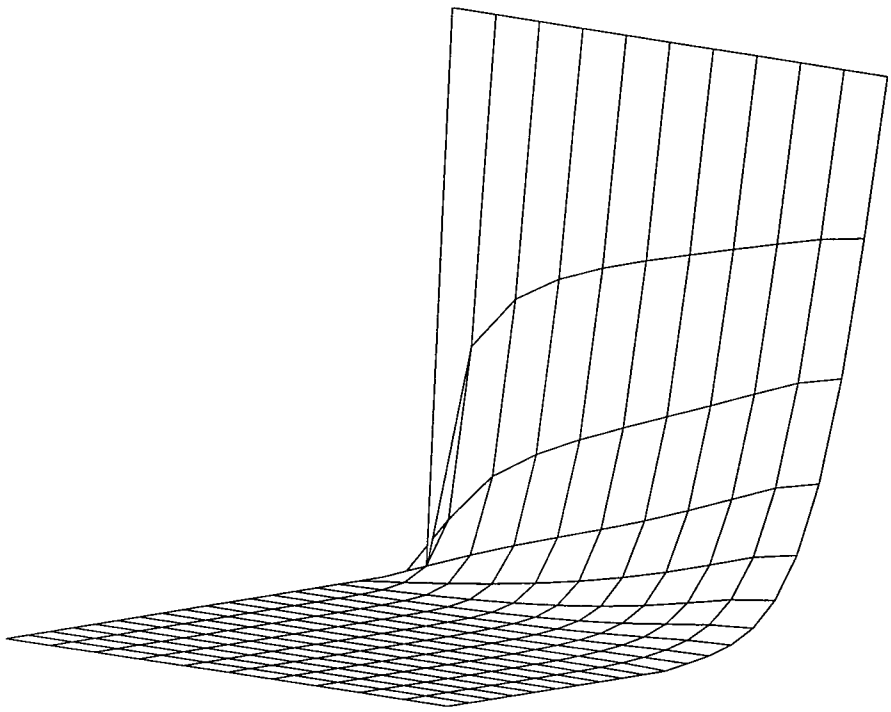


Figure 5.30: CEGB test problem 2 with $a = 0.1, h = 0.1$

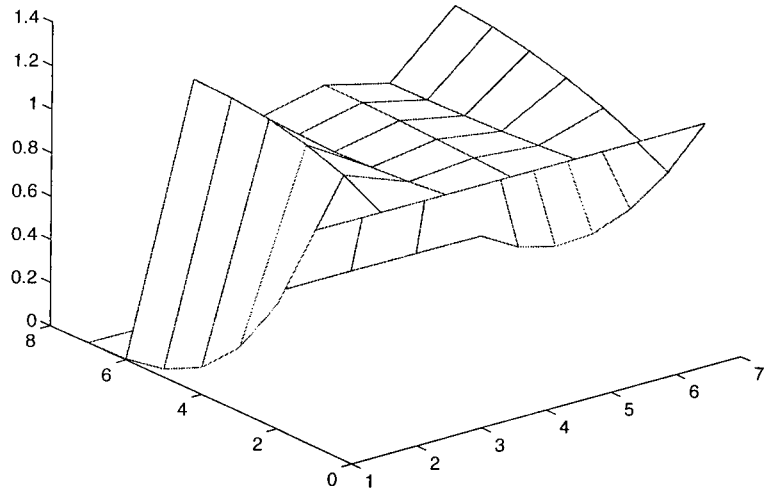


Figure 5.31: Parabolic layer test problem with splitting constant $C = 0$ and $h = 1/6$

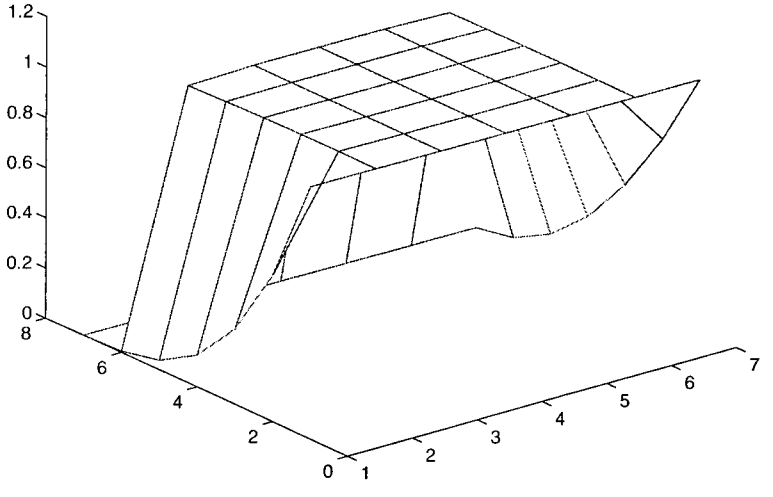


Figure 5.32: Parabolic layer test problem with splitting constant $C = |b_2| - |b_1|$ and $h = 1/6$

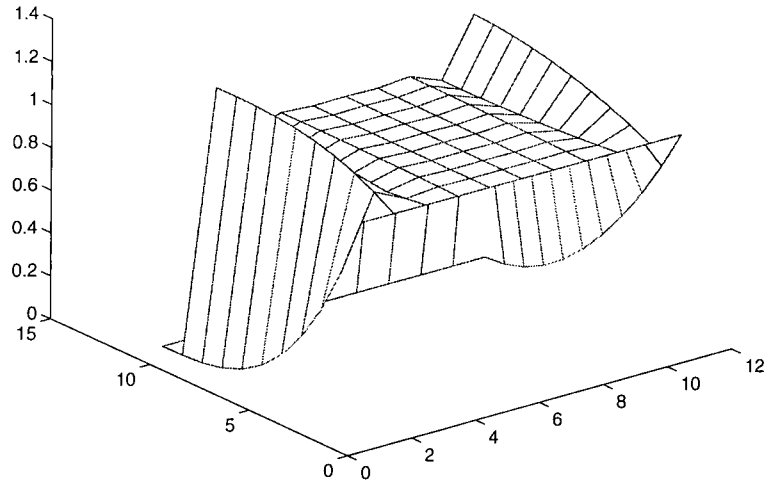


Figure 5.33: Parabolic layer test problem with splitting constant $C = 0$ and $h = 1/10$

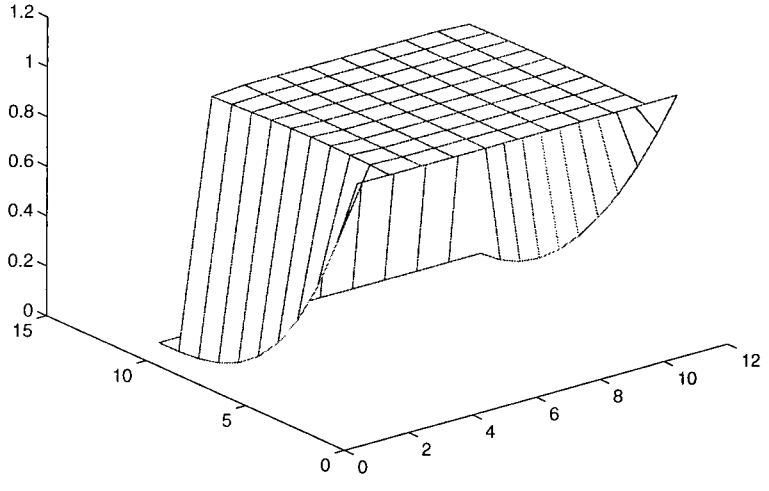


Figure 5.34: Parabolic layer test problem with splitting constant $C = |b_2| - |b_1|$ and $h = 1/10$

5.8 Three Dimensional Problems

In this section we present three dimensional results for a simple extension of the first CEGB test problem. Due to the difficulty in representing three dimensional solutions on paper we take two dimensional slices through the problem regions and display contour plots of the solution on those slices. We also give a contour plot of inflow data and outflow solution.

5.8.1 Extension of the CEGB Test Problem One

We present the following example to show how well the method copes with face–aligned flow. (The method copes equally well with non-face-aligned flow.)

For this test problem we have

$$\Omega = \{(x, y, z) | -1 < x < 1, 0 < y < 1, -1 < z < 1\} \tag{5.10}$$

with convective field (see figure 5.35)

$$\mathbf{b}(x, y, z) = (2y(1 - x^2), -2x(1 - y^2), 0)^T, \tag{5.11}$$

and with $a = 1 \times 10^{-1}, 1 \times 10^{-2}, 2 \times 10^{-3}, 1 \times 10^{-6}$.

The inflow profile ($x < 0$) is given by

$$u(x) = (1 + \tanh(20x + 10))(1 + \tanh(-20|z| + 10)).$$

Due to the high quality of the solution at very modest mesh spacing we

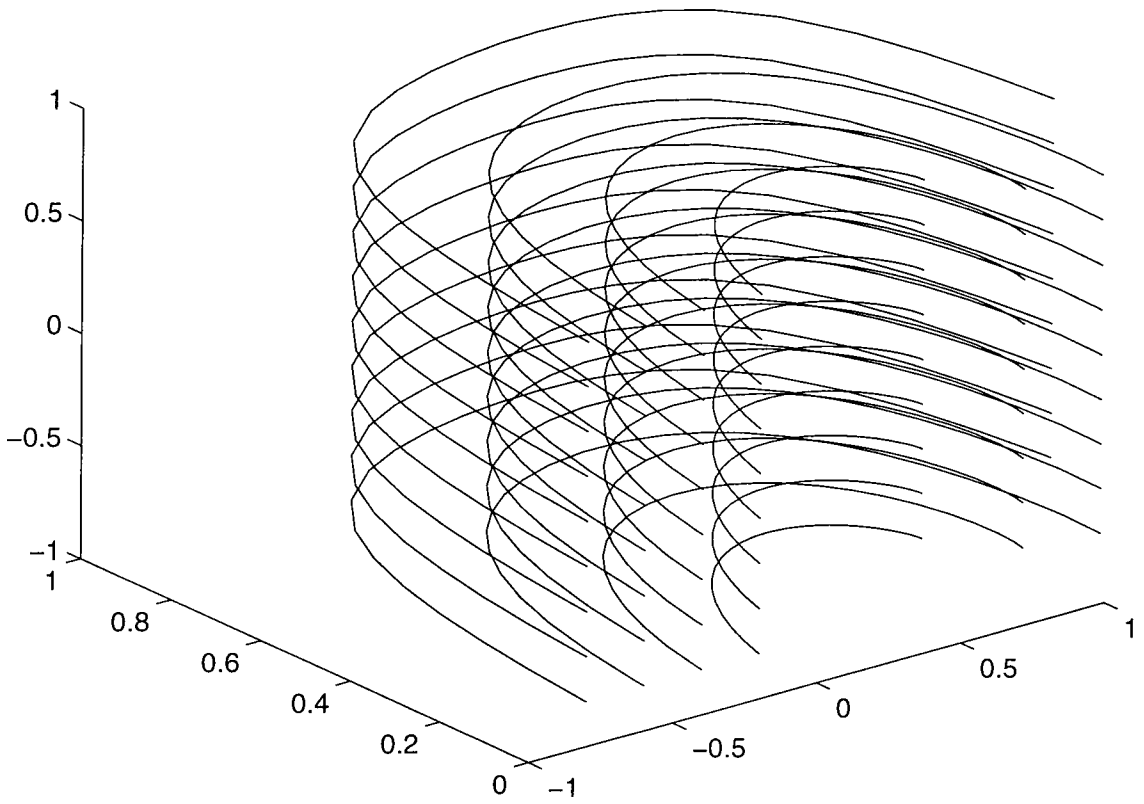


Figure 5.35: Convective field for three dimensional CEGB1 problem

present results (see figures 5.36 to 5.43) only for cubic elements of width 0.25 giving a mesh of $8 \times 4 \times 8$ elements. Contours are plotted at intervals of 0.25. Due to the symmetry about $z = 0$ (which is preserved by the numerical method) we show only slices through $z = -1, 0.75, -0.5, -0.25, 0$. We also present inflow/outflow contour plots. In each figure the labels on the 'x','y' and 'z' axes are for values of i,j and k rather than for x_i,y_j and z_k.

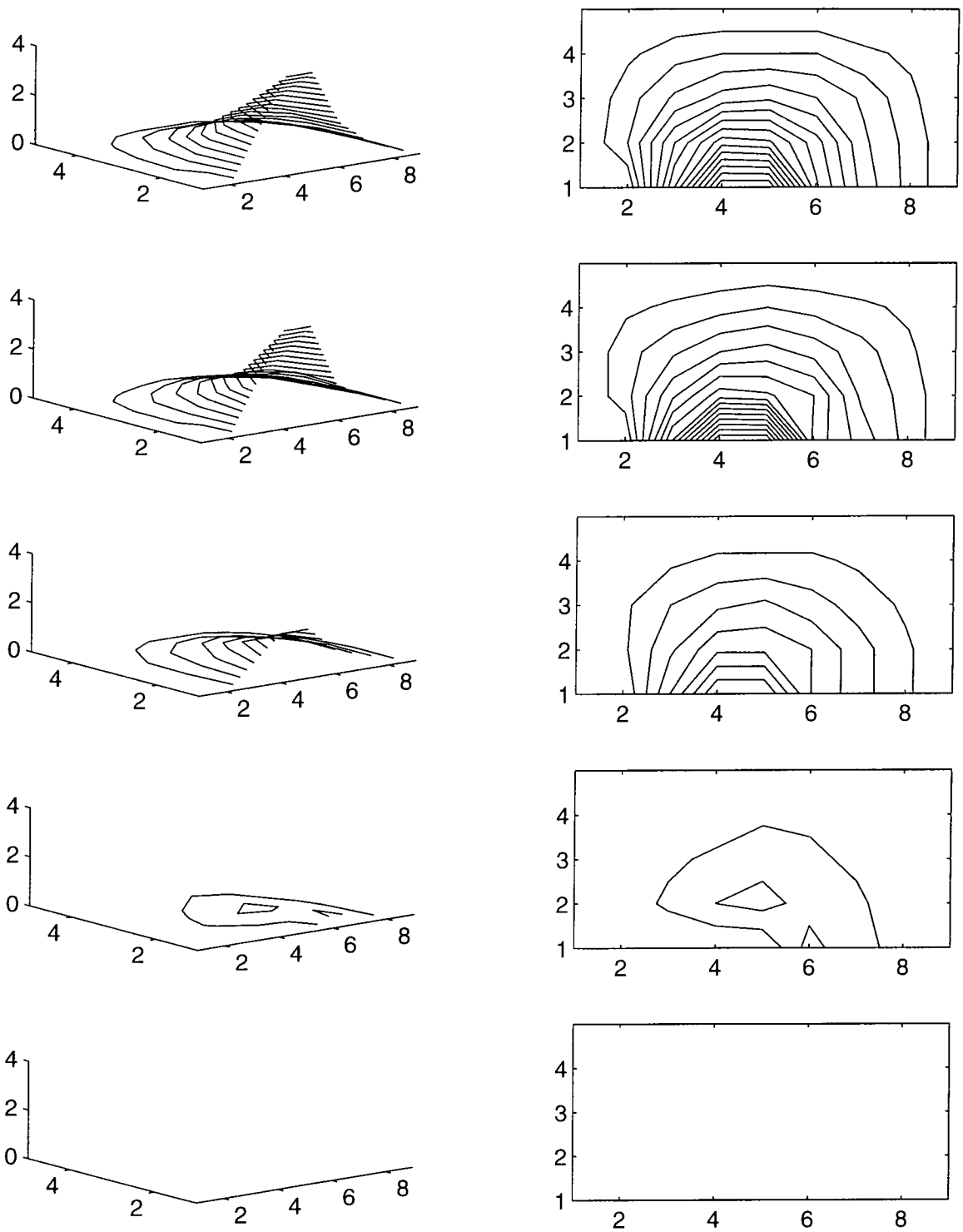


Figure 5.36: Three dimensional CEGB1 with $a = 0.1, h_1 = h_2 = h_3 = 0.25$

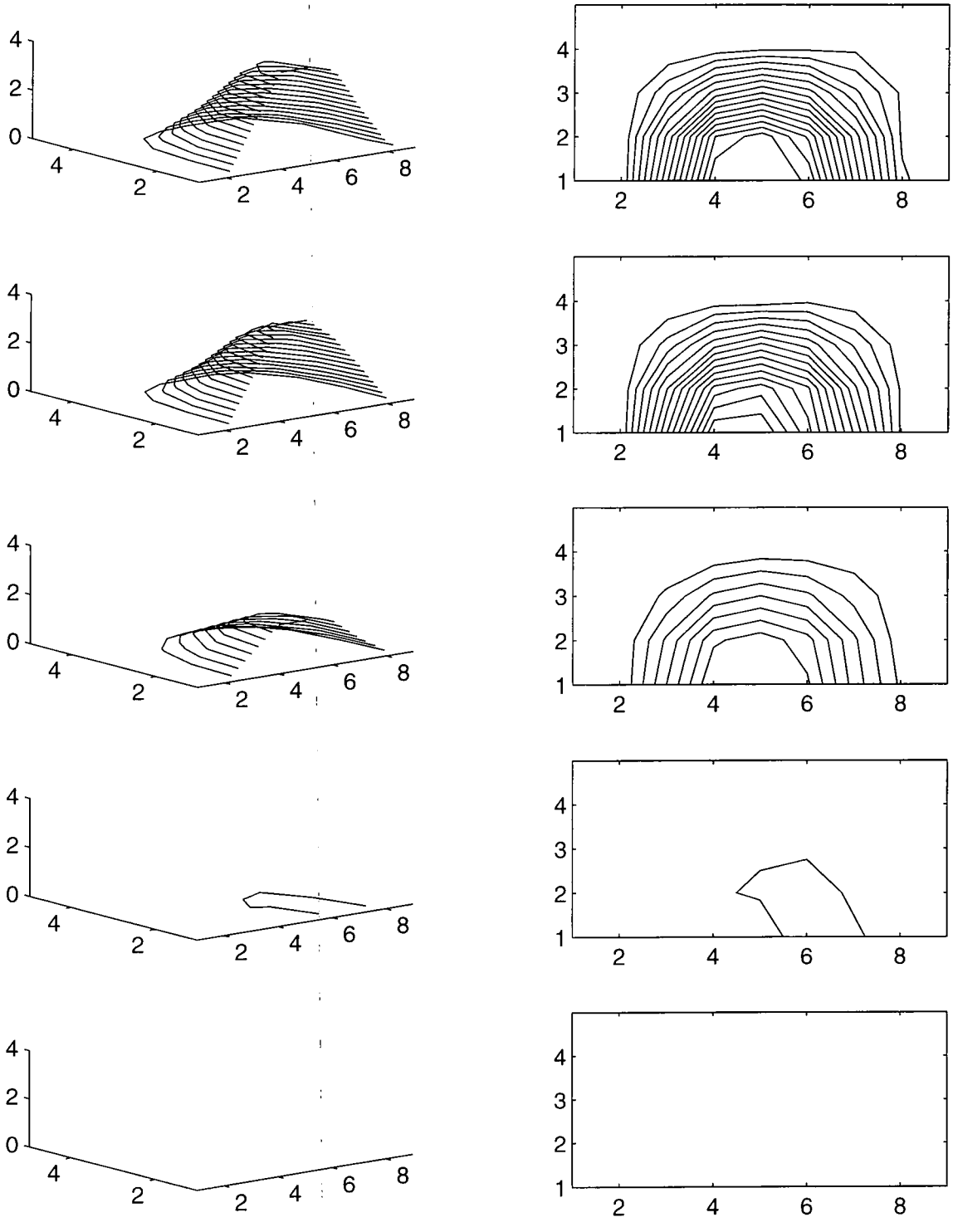


Figure 5.37: Three dimensional CEGB1 with $a = 0.01, h_1 = h_2 = h_3 = 0.25$

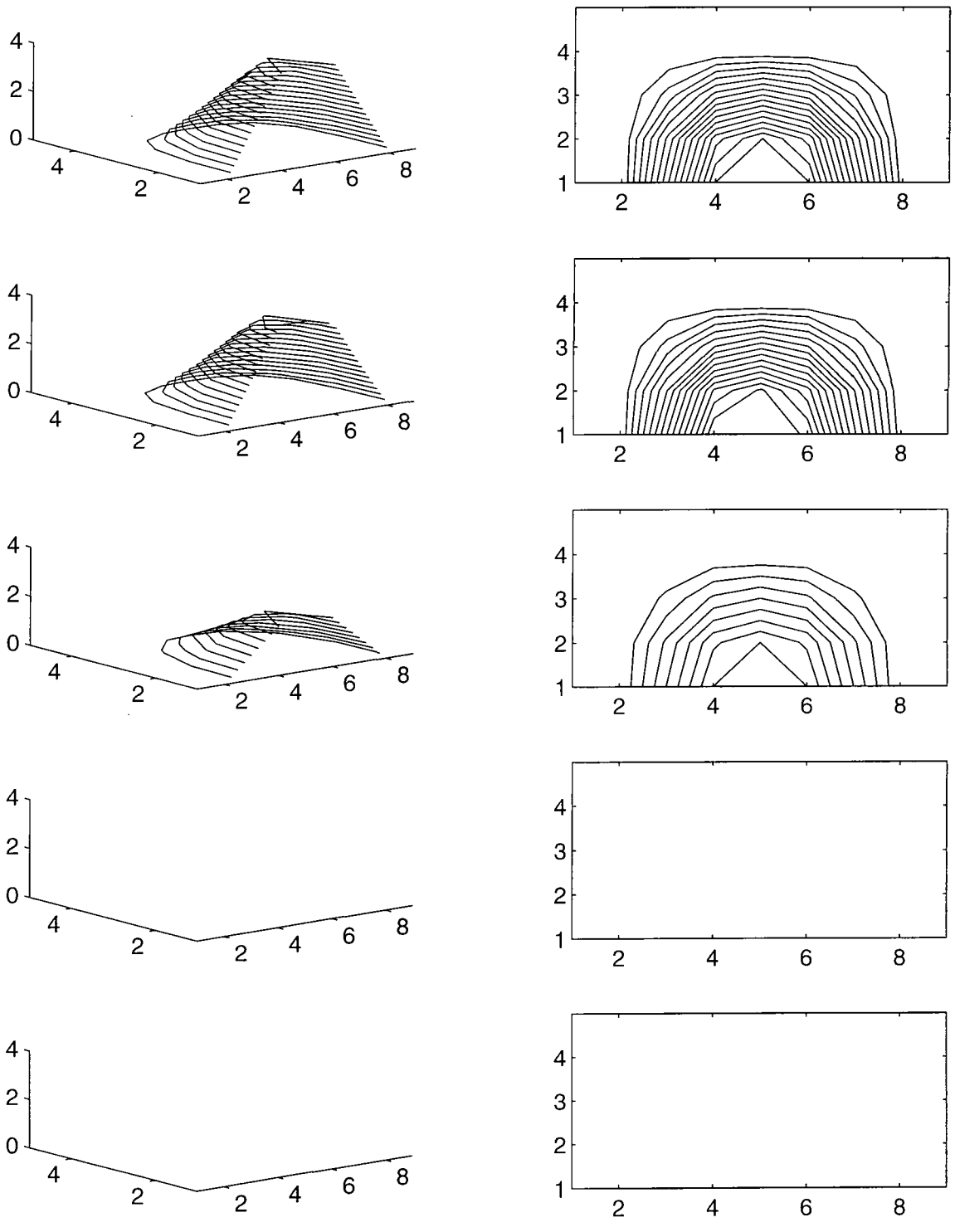


Figure 5.38: Three dimensional CEGB1 with $a = 0.002, h_1 = h_2 = h_3 = 0.25$

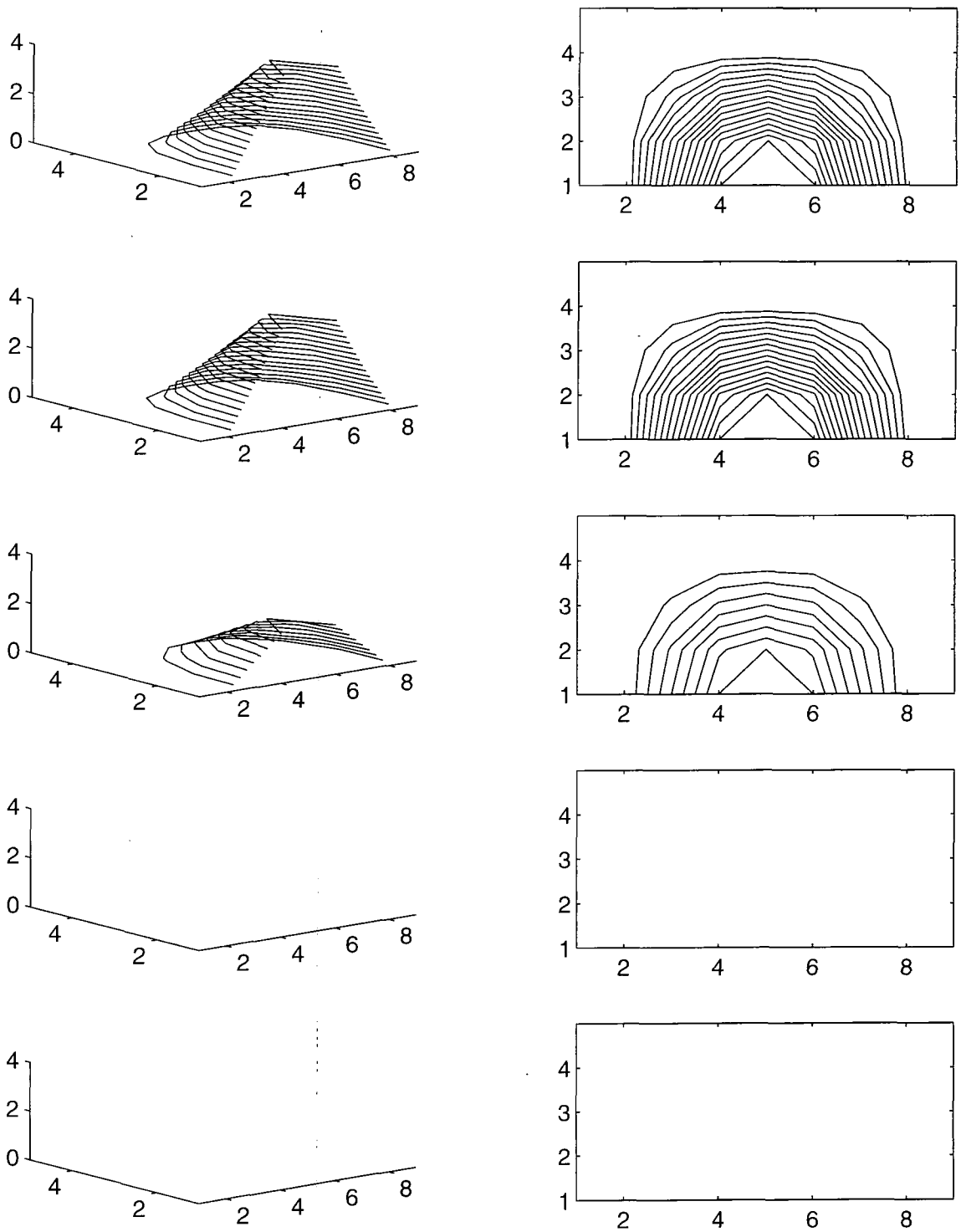


Figure 5.39: Three dimensional CEGB1 with $a = 0.000001, h_1 = h_2 = h_3 = 0.25$

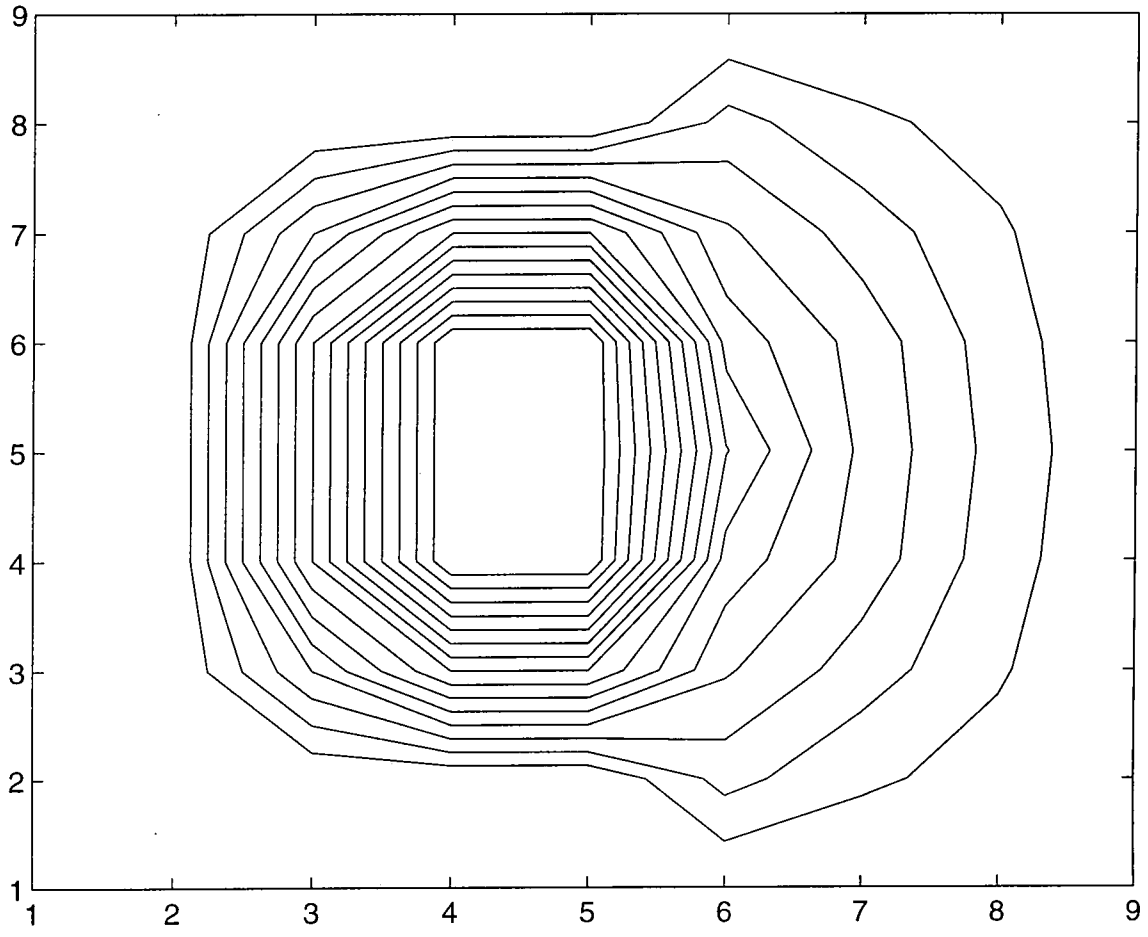


Figure 5.40: Inflow/Outflow contour with $a = 0.1, h_1 = h_2 = h_3 = 0.25$

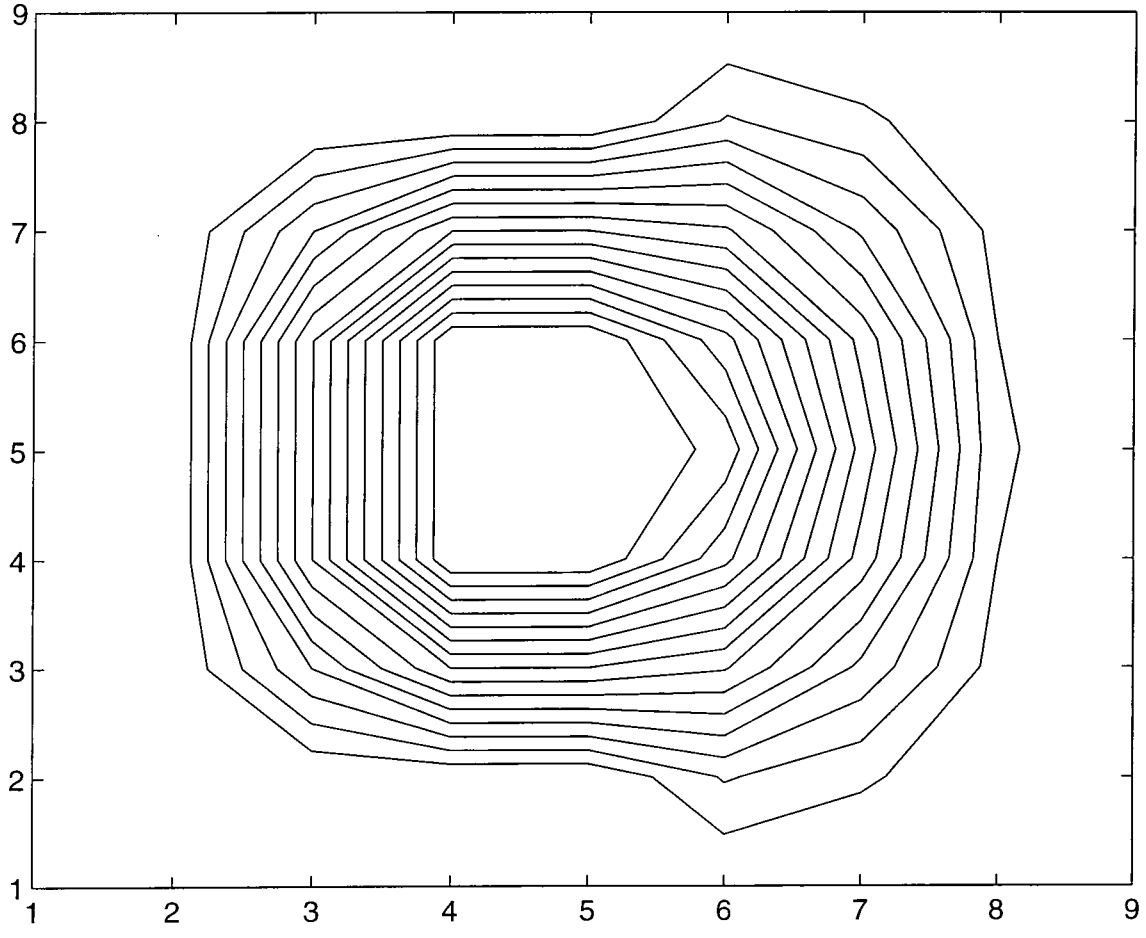


Figure 5.41: Inflow/Outflow with $a = 0.01, h_1 = h_2 = h_3 = 0.25$

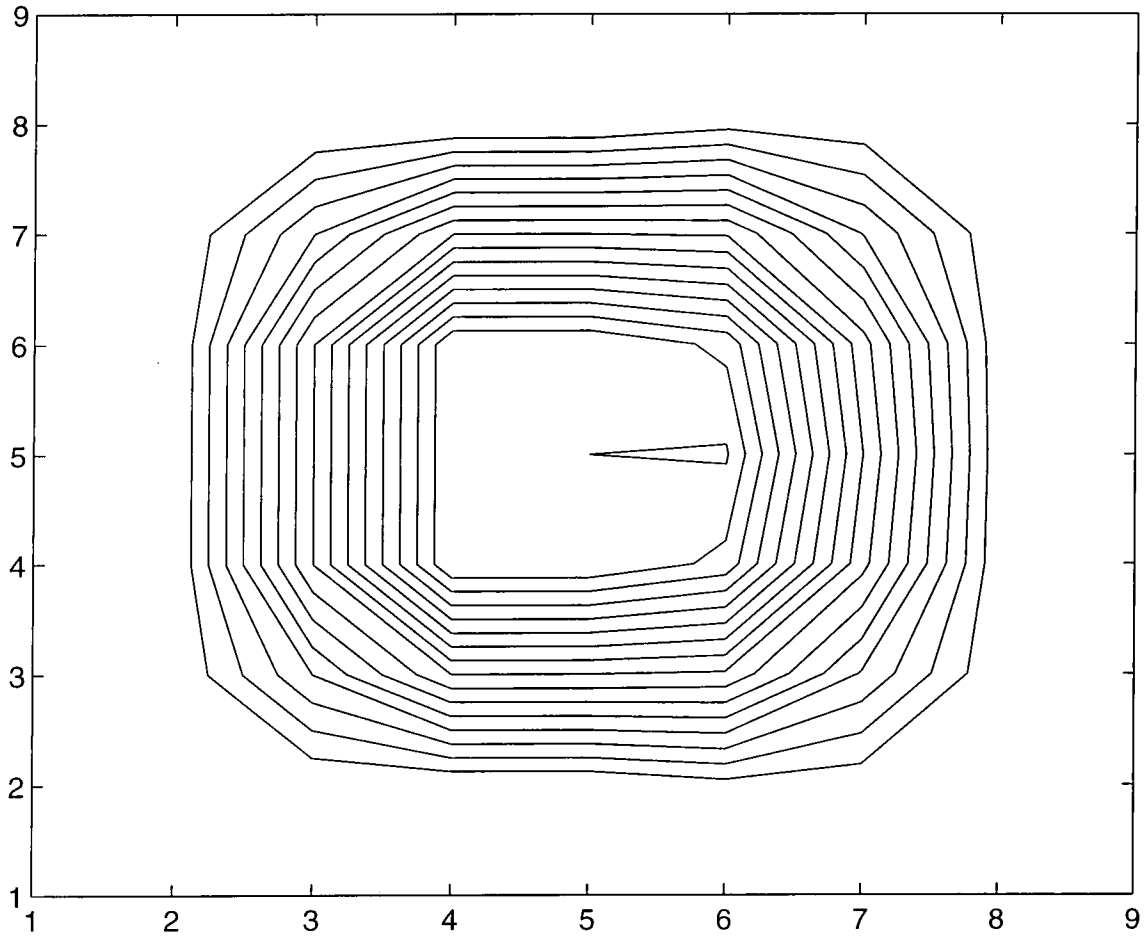


Figure 5.42: Inflow/Outflow CEGB1 with $a = 0.002, h_1 = h_2 = h_3 = 0.25$

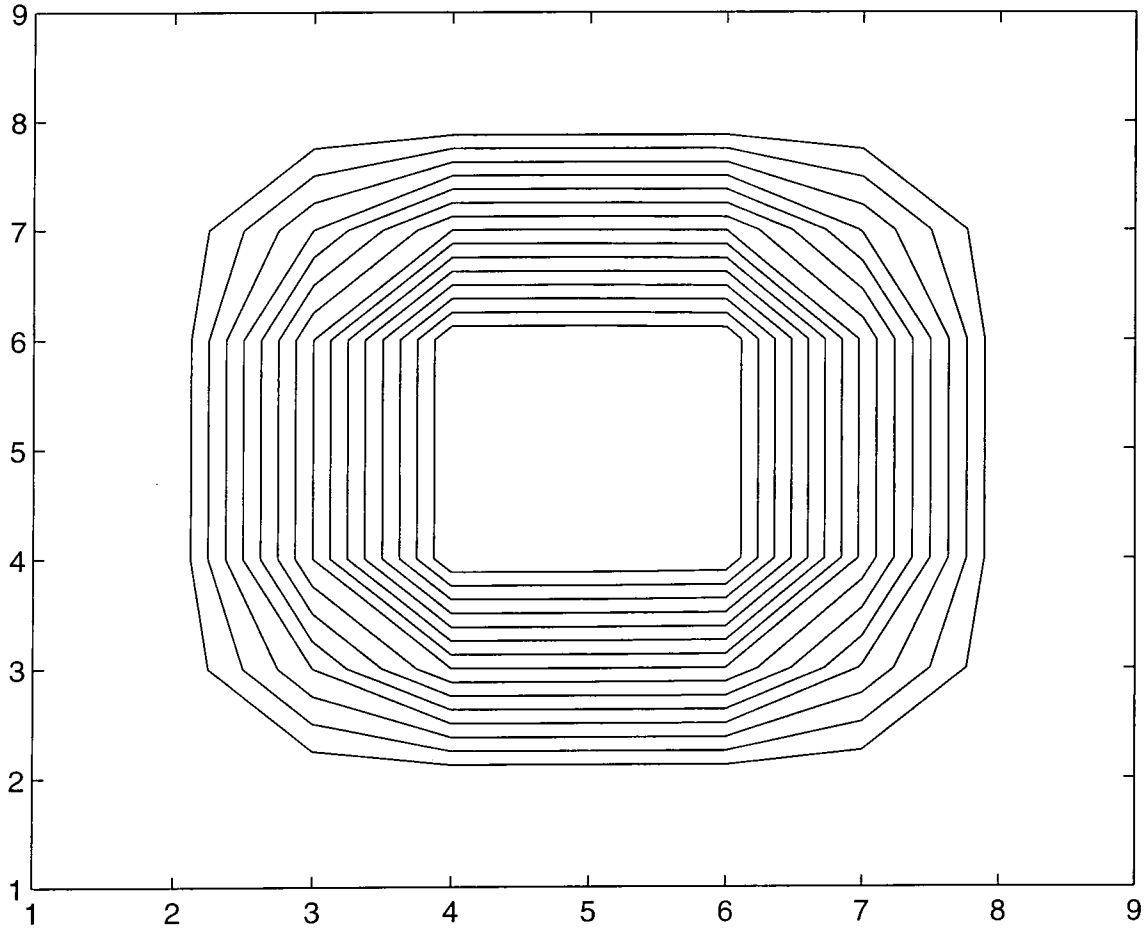


Figure 5.43: Inflow/Outflow with $a = 0.000001, h_1 = h_2 = h_3 = 0.25$

Chapter 6

Generating Exact Difference Schemes for Boundary Value Problems

6.1 Introduction

In this chapter we describe a new method for producing difference schemes for boundary value problems that yield exact values for the solution and all of its derivatives (up to the order of the equation) at a set of nodes. We first define the method and then prove the exact accuracy by the principle of mathematical induction.

We then describe an alternative derivation of this method and explore it in the context of the Poisson equation.

We then briefly discuss extensions to higher dimensional problems.

6.2 An exact difference scheme

Given an nth order linear differential operator $Lu = \sum_{i=0}^{n} \alpha_i u^{(i)}$ and the boundary value problem

$$Lu(x) = f(x) \text{ in } \Omega = [0, 1],$$

with appropriate boundary conditions we have the weak form by multiplying both sides of the equation by a test function w from some test space $\mathcal{W}$.

We can now write

$$\sum_{i=0}^{n} \alpha_i (u^{(i)}, w) = (f, w).$$

We partition Ω in to m elements by the nodes x_j, $j = 0, ...m$. If we restrict $\mathcal{W}$ by making all $w \in \mathcal{W}$ satisfy the homogeneous adjoint equation on each

element. That is,

$$\sum_{i=0}^{n}(-1)^i\alpha_i w^{(i)} = 0 \ for \ x \in (x_j, x_{j+1}) \ , j = 0, ..., m-1 \tag{6.1}$$

then we can state the following theorem.

Theorem 20 *Given a w as defined above, the problem $Lu(x) = f(x)$ in $\Omega = [0,1]$ can be written as*

$$\sum_{j=1}^{m-1} J_j \left(\sum_{i=1}^{n} \alpha_i \sum_{k=0}^{i-1} u^{(i-1-k)} w^{(k)} \right) = (f, w),$$

where

$$J_j(g) = g(x_{j+}) - g(x_{j-})$$

The proof of the above theorem is a trivial use of integration by parts and the application of equation 6.1.

If we chose a finite dimensional trial space $\mathcal{V}$ of the same dimension as $\mathcal{W}$ in such a way that all the integrals make sense (i.e. $U \in \mathcal{V}$ is smooth enough at the element boundaries) we have the following Petrov-Galerkin finite element method. Find $U \in \mathcal{V}$ such that

$$\sum_{j=1}^{m-1} J_j \left(\sum_{i=1}^{n} \alpha_i \sum_{k=0}^{i-1} U^{(i-1-k)} w^{(k)} \right) = (f, w), \ \forall \ w \in \mathcal{W}.$$

We split the test space $\mathcal{W}$ into n subspaces $\mathcal{W}_i$ $i = (0, ..., n-1)$ and chose $\mathcal{W}$ by choosing each $\mathcal{W}_i$ in the manner described in the following theorem.

Theorem 21 *Choose $\mathcal{W}_i$ so that for all $w_i \in \mathcal{W}_i$, $w_i^{(n-1-i)}$ is discontinuous*

at element boundaries and lower derivatives are continuous at element boundaries.

If we assume that the Petrov-Galerkin problem defined above has a unique solution then the approximate solution U and all of its derivatives up to $n-1$ will be exact at the element boundaries.

Proof

Let $e = U - u$ be the difference between the exact and approximate solutions.

Assume that $e^{(i)} = 0$ at the element boundaries for $i = 0, ..., k-1$. Then applying the conditions of $\mathcal{W}_k$ immediately yields

$$\sum_{i=1}^{m-1} J_i(e^{(k)} w_k^{(n-1-k)}) = 0.$$

Hence by the assumed uniqueness of the linear system obtained by varying w_k we have that $e^{(k)} = 0$ at the element boundaries.

However if we apply the conditions of $\mathcal{W}_0$ yields,

$$\sum_{i=1}^{m-1} J_i(e w_k^{(n-1)}) = 0.$$

Hence $e = 0$ at the element boundaries.

Proof follows immediately by the principle of mathematical induction.

We remark here how it is possible to generate these exact difference schemes without explicitly generating the test space [25]. This can be performed even if the equation is nonlinear.

6.3 An alternative derivation

Presented here is an alternative derivation of these methods designed to give exact derivative information. The motivation for this work comes from the observation that the exact solution $u(x)$ can be written in terms of the Greens function associated with the operator and the region of integration.

$$u(x) = \int_{\Omega} G(x,y) f(y)\ dy.$$

where $G(x,y)$ is the Greens function associated with the operator L and the region Ω.

If we now differentiate both sides of the equation with respect to x $n-1$ times we obtain,

$$\frac{d^i u(x)}{dx^i} = \int_{\Omega} \frac{d^i G(x,y)}{dx^i} f(y)\ dy, \quad i = 0, ..., n-1.$$

Hence if we construct a Petrov-Galerkin method based on a set of nodes x_i $(i = 0, ..., m)$ using a trial space $\mathcal{V}$ which is C^j $(0 \leq j < n)$ continuous at the nodes x_i and a test space $\mathcal{W} = \text{span}\{\frac{d^j G(x_i,y)}{dx^j},\ i = 1, ..., m-1\}$ then the method will yield a solution which has exact jth derivatives at the nodes x_i. This will be a Petrov-Galerkin finite element method if both the trial and test spaces can be constructed from a set of local basis functions defined over r elements. At first glance it is not clear that the test space can be written as a sum of local basis functions and while this is not necessary it is highly desirable for computational reasons. It is clear now that we can obtain $\frac{d^j G(x_i,y)}{dx^j}$, $j = 0, ..., n-1$ by using a test space $\mathcal{W} = \text{span}\{\frac{d^j G(x_i,y)}{dx^j},\ i = 1, ..., m-1, j = 0, ..., n-1\}$ and can recover values for $\frac{d^n G(x_i,y)}{dx^n}$ directly from the equation itself using the values for $\frac{d^j G(x_i,y)}{dx^j}$, $j = 0, ..., n-1$.

6.4 Exact derivative solution for the Poisson equation

For simplicity we first present the method in the context of the Poisson equation:.

$$-u''(x) = f(x),$$

on $[0, 1]$ with zero Dirichlet data.

We have,

$$G(y, x) = y(1 - y)\phi(y),$$

where $\phi(y)$ is the standard hat function which takes the value 0 at $x = 0, x = 1$ and the value 1 at $x = y$ (see figure 6.1).

In order to generate exact derivative values at $x = y$ we need to use the test function $\frac{dG(y,x)}{dy} = \psi(x)$ where $\psi(x) = -x, 0 \leq x < y$ and $\psi(x) = 1 - x, y < x \leq 1$ (see figure 6.2).

Choosing a set of internal nodes x_i $(i = 1, 2, ..., m-1)$ we define $\psi_i(x) = -x$ for $0 \leq x < x_i$ and $\psi_i(x) = 1 - x$ for $x_i < x \leq 1$. We shall refer to the standard hat functions centred at node x_i as $\phi_i(x)$.

Unfortunately it is not possible to find a local basis for the space $\mathcal{H} = \text{span}\{\psi_i(x)\}$ (although it is possible to arrange for all but one of the test functions to be local) so the problem of finding the derivative solution at the nodes is not computationally very efficient.

However if we solve for exact function values in addition to exact derivative values then we can produce a local basis. This is because we can use linear

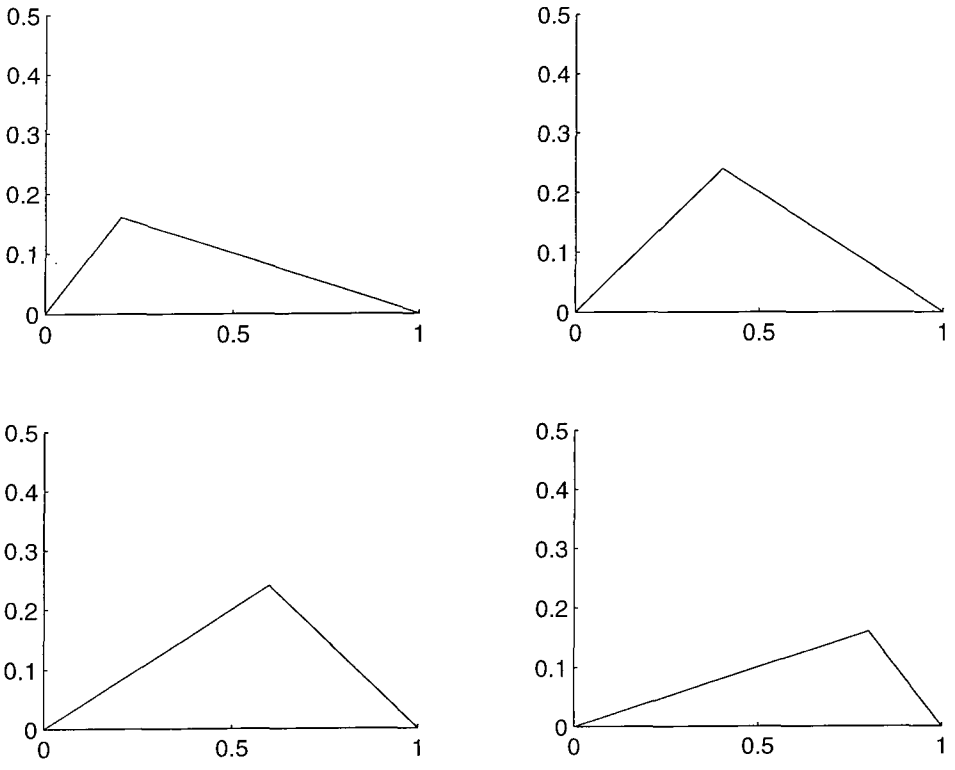


Figure 6.1: Plot of $G(y,x)$ for $y = 0.2, 0.4, 0.6$ and 0.8

combinations of both ϕ_i and ψ_i to produce a new basis. We generate a new local set of basis functions $\lambda_i(x) = \psi_i(x) + \sum_{j=1}^{m-1} a_j \phi_j(x)$. The coefficients a_j are chosen to make $\lambda_i(x)$ zero everywhere apart from $[x_{i-1}, x_{i+1}]$ (see figure 6.3).

We can make a further simplification by using the basis functions $\gamma_i(x) = b_i \lambda_i(x) + c_i \phi_i(x)$ where b_i and c_i are chosen so that the left and right limits of $\gamma_i(x)$ at $x = x_i$ are 1 and -1 (see figure 6.4).

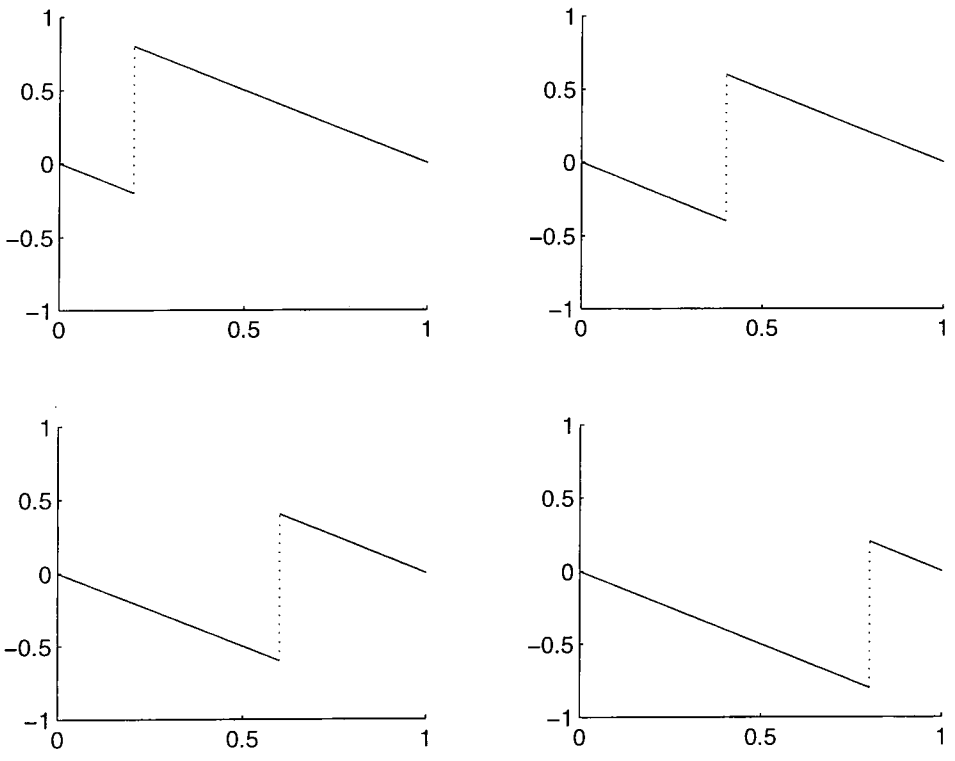


Figure 6.2: Plot of $\frac{dG(y,x)}{dy}$ for $y = 0.2, 0.4, 0.6$ and 0.8

6.5 Extensions to higher dimensional problems

We can extend this method into higher dimensions in exactly the same way that the methods in the previous chapters are extensions of the one dimensional nodally exact methods. In a similar way, we obtain an error equation posed on the element boundaries. In particular we note that if the trial space is capable of reproducing the exact solution and its derivatives on the element boundaries, then there will be zero error on the element boundaries.

What is more exciting is that as long as the numerical problem is reasonably stable and we can reproduce the solution values on the element boundaries

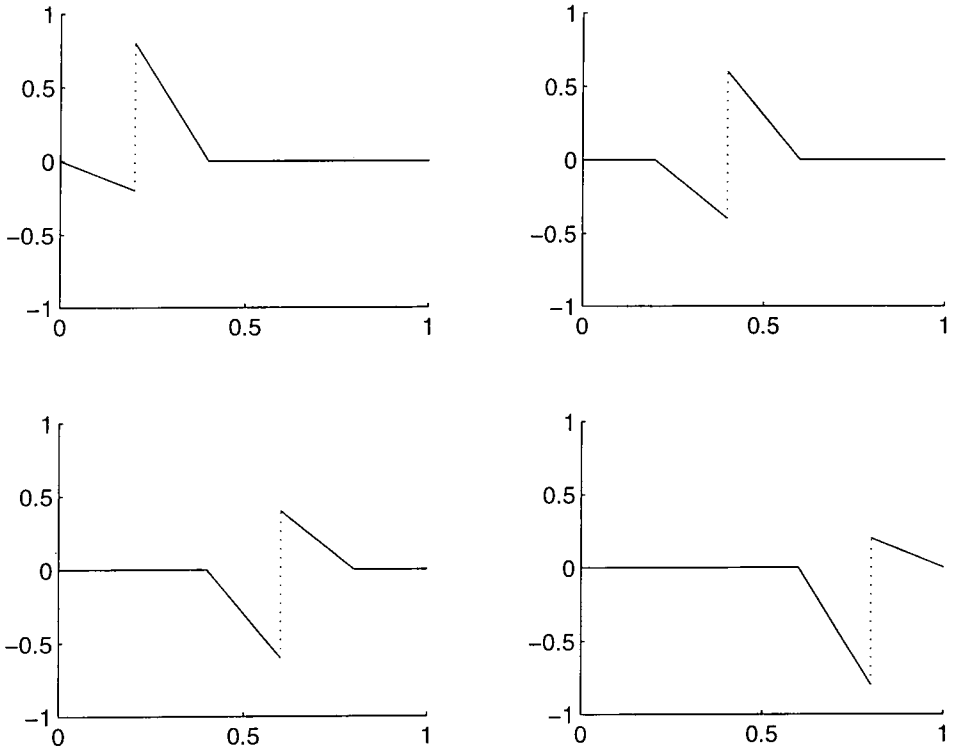


Figure 6.3: Plot of $\lambda_i(x)$ for $x_i = 0.2, 0.4, 0.6$ and 0.8

to a reasonable accuracy then we can also obtain good derivative values of the solution just in small critical areas by adding in just the 'derivative' test functions at those areas.

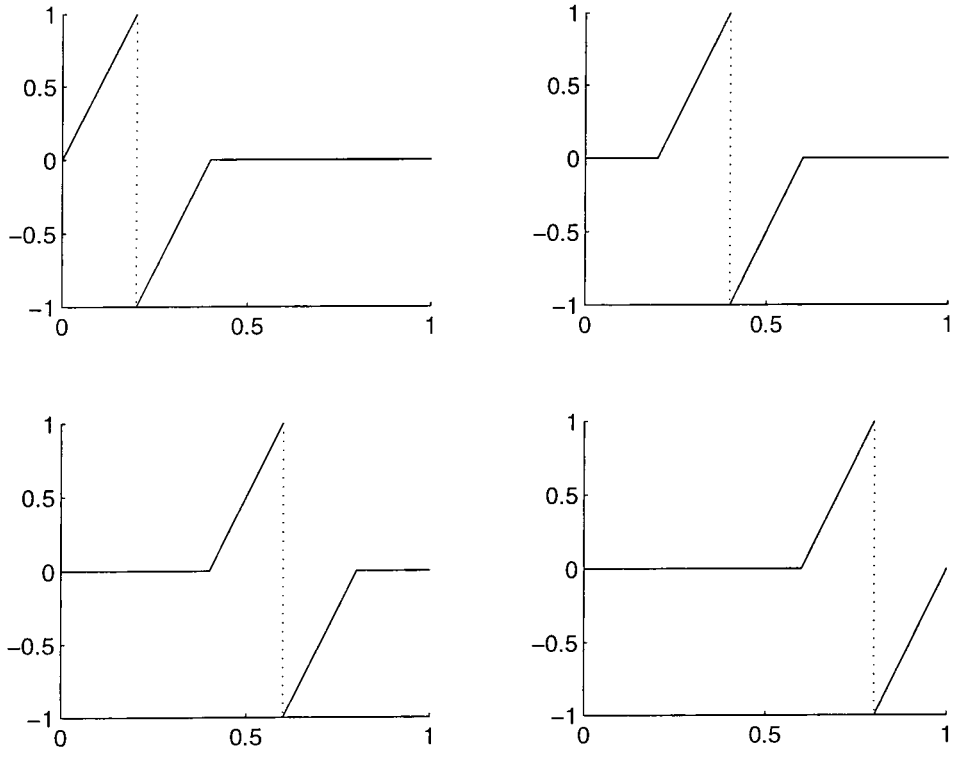


Figure 6.4: Plot of $\gamma_i(x)$ for $x_i = 0.2, 0.4, 0.6$ and 0.8

Chapter 7

Conclusions

7.1 Summary

In this thesis we have described the problems associated with the numerical solution of the convection-diffusion equation and described the various techniques used to overcome these problems.

It became clear that none of these methods was ideally suited for a wide range of problems. The need for a more generally applicable method prompted the creation of a class of Petrov-Galerkin finite element schemes designed to control the errors in the numerical solution on the element boundaries of the finite element mesh. We saw how a particular case of this class, the tensor product case, produced highly accurate solutions to various standard test problems. We also saw that two limiting cases of the method produced the standard Galerkin finite element method and also the cell vertex finite volume method. Asymptotic, nonasymptotic and truncation error analyses were performed on the method.

We then presented a class of schemes designed to produce exact n-th derivative values for the solution of the one dimensional boundary value problems.

7.2 Suggestions for Further Work

So far only a basic preliminary error analysis of the class of methods we have developed has been performed. It will be useful for the practical application of these methods to have a tight a posteriori estimate of the error. There is also much scope for extensions of the exact derivative methods to be extended to higher dimensional problems. This would be extremely useful for problems such as accurately calculating the stresses in structures. An exciting prop-

erty of the one dimensional method is that extra local basis functions can be placed where accurate derivative type information is required without globally adding basis functions where they are not needed. Extensions of this to higher dimensions with the aid of an a posteriori estimate of the error, could lead to an adaptive mesh refinement strategy resulting in a highly efficient widely applicable numerical solver.

Bibliography

[1] Aziz, A.K. and Babuska, I. (1972). Survey lectures on the mathematical foundation of the finite element method. *The mathematical foundations of the finite element method with applications to partial differential equations.* (Aziz, A.K. , Ed.), Academic press, New York, 3-363.

[2] Allen, D. and Southwell, R. (1955). Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder. *Quart. J. Mech. and Appl. Math., VIII*, 129-145.

[3] Christie, I., Griffiths, D.F. , Mitchell, A.R. and Zienkiewicz, O.C. (1976). Finite element methods for second order differential equations with significant first derivatives. *Int. J. Num. Meth. Engng., 10*, 1389-1396.

[4] Christie, I. and Mitchell A. R. (1978). Upwinding of high order Galerkin methods in conduction-convection problems. *Int. J. Num. Meth. Engng., 12*, 1764-1771.

[5] Ciarlet, P.G. (1978). *The finite element method for elliptic problems.* North-Holland, Amsterdam, New York, Oxford.

[6] Farrell, P.A. , Hemker, P.W. and Shishkin, G.I. (1995). Discrete Approximations for Singularly Perturbed Boundary Value Problems with Parabolic Layers. *Centrum voor Wiskunde en Informatica Report NM-R9502.*

[7] Field, M. R. (1992). The setting up and solution of the cell vertex equations. *Oxford University Computing Laboratory Numerical Analysis Group Report 92/7.*

[8] Gartling, D.K. (1978). Some comments on the paper by Heinrich, Huyakorn, Zienkiewicz and Mitchell. *Int. J. Num. Meth. Engng., 12*, 187-190.

[9] Gatti, E., Gotusso, L. and Sacco R. (1995). *Polytechnic of Milan, technical report.*

[10] Gilbarg, D. and Trudinger, N.S. (1977). *Elliptic partial differential equations of second order.*

[11] Griffiths, D. and Lorenz, J. (1978). An analysis of the Petrov-Galerkin finite element method. *Comp. Meth. Appl. Mech. Engng. 14*, 39-64.

[12] Hegarty, A.F., Miller, J.J.H., O'Riordan, E. and Shishkin, G.I. (1994). Special numerical methods for convection-dominated laminar flows at arbitrary Reynolds number. *East-West J. Numer. Math. 2,1*, 65-74.

[13] Hegarty, A.F., Miller, J.J.H., O'Riordan E. and Shishkin, G.I. (1994). Use of central-difference operators for solution of singularly perturbed problems. *Comm. Numer. Meth. Eng. 10*, 297-302

[14] Hegarty, A.F., O'Riordan, E., Stynes, M. (1993). A comparison of uniformly convergent difference schemes for two-dimensional convection-diffusion problems. *J. Comp. Phys., 105*, 24-42.

[15] Heinrich, J.C., Huyakorn, P.S., Zienkiewicz, O.C.and Mitchell, A.R. (1977). An Upwind finite element scheme for two-dimensional convective transport equation. *Int. J. Num. Meth. Engng., 11*, 131-143.

[16] Heinrich, J.C. and Zienkiewicz, O.C. (1977). Quadratic finite element schemes for two-dimensional convective-transport problems. *Int. J. Num. Meth. Engng.*, *11*, 1831-1844.

[17] Hemker, P.W. (1977). A numerical study of stiff two-point boundary problems. *Thesis, Mathematisch Centrum, Amsterdam.*

[18] Herrera I. (1985). Unified formulation of numerical methods. 1. Green's formulas for operators in discontinuous fields. *Numerical Methods for Partial Differential Equations.*, *1*, 25-44.

[19] Hubbard, M.E. (1993). A survey of genuinely multidimensional upwinding techniques. *University of Reading, technical report 7/93.*

[20] Hughes, T.J.R. (1978). A simple scheme for developing 'upwind' finite elements. *Int. J. Num. Meth. Engng.*, *12*, 1359-1365.

[21] Hughes, T.J.R. and Brooks, A. (1982). A theoretical framework for Petrov-Galerkin methods with discontinuous weighting functions: Application to the streamline-upwind procedure. *Finite elements in fluids, Vol. 4.* (Gallagher, R.H. , Norrie, D.H. , Oden, J.T. and Zienkiewicz, O.C. Eds.) John Wiley & Sons Ltd., 47-65.

[22] Il'in, A.M. (1969). Differencing scheme for a differential equation with a small parameter affecting the highest derivative. *Math. Notes Acad. Sc. USSR.*, *6*, 596-602.

[23] Johnson, C. (1987). *Numerical solution of partial differential equations by the finite element method.* Cambridge University Press, Cambridge.

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

[25] Laurie, D.P. and Craig, A. (1996). Exact difference formulas for linear differential operators. *Numerical Analysis, A.R. Mitchell* 75^{th} *Birthday Volume.* (Griffiths, D.F. and Watson, G.A. Ed.) World Scientific. 155-162.

[26] Leonard, B.P. (1979). A survey of finite differences of opinion on numerical muddling of the incomprehensible defective confusion equation. *Finite element methods for convection dominated flow.* (Hughes, T.J.H. ed.) AMD-Vol. 34, 1-18.

[27] Mackenzie, J.A. and Morton, K.W. (1992). Finite Volume Solutions of Convection-Diffusion Test Problems. *Math. Comp., 60,* 189-220.

[28] Manteuffel, T.A. and Otto, J.S. (1995). On the uniform consistency of an exponentially upwinded discretization for singular perturbation problems. *Research in progress.*

[29] Morton, K.W. (1981). Finite element methods for non-self-adjoint problems. *University of Reading, technical report 3/81.*

[30] Morton, K.W. (1982). Generalised Galerkin methods for steady and unsteady problems. *Numerical methods for fluid dynamics.* (Morton, K.W. and Baines, M.J. ed.) Academic Press, London and New York. 1-32.

[31] Morton, K.W. (1996). *Numerical Solution of Convection–Diffusion Problems.* Chapman and Hall.

[32] Morton, K.W., Murdoch, T. and Süli, E. (1992). Optimal error estimation for Petrov-Galerkin methods in two dimensions. *Numer. Math. 61,* 359-372.

[33] Morton, K.W. and Scotney, B.W. (1985). Petrov-Galerkin methods and diffusion-convection problems in 2D. *The Mathematics of Finite Elements and Applications, V MAFELAP 1984.* (J.R.Whiteman, ed.), Academic Press, London and New York. 343-366.

[34] Morton, K. W. and Süli, E. (1991). Finite Volume Methods and their Analysis. *IMA J. Numer. Anal. 11*, 241-260.

[35] Oden, J. T. and Reddy, J. N. (1976). *An Introduction To The Mathematical Theory Of Finite Elements.* John Wiley and Sons.

[36] Ortega, J. M. and Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables.* Academic Press, New York.

[37] Perella, A. J. (1996). Highly Accurate Solution of the Stationary Convection–Diffusion Equation. *Numerical Methods for Fluid Dynamics V.* (Morton, K.W. and Baines, M.J. ed.) Oxford Science Publications.

[38] Rae, J. (1982) Finite element solutions for Navier–Stokes equations. *Numerical methods for fluid dynamics.* (Morton, K.W. and Baines, M.J. ed.) Academic Press, London and New York. 81-96

[39] Russel, T.F. (1989). Eulerian-Lagrangian Localised Adjoint Methods for Advection-Dominated Problems. *Numerical Analysis 1989*

[40] Sacco, R. (1995). A nonconforming Petrov-Galerkin finite element method for stationary convection-diffusion equations. *Polytechnic of Milan, technical report.*

[41] Scotney, B.W. (1982). Error analysis and numerical experiments for Petrov-Galerkin methods. *University of Reading, technical report 11/82.*

[42] Smith, R.M. and Hutton, A.G. (1982). The numerical treatment of convection - a performance/comparison of current methods. *Num. Heat Trans. 5*, 439-461.

[43] Strang, G. (1976). *Linear algebra and its applications.* Academic Press, New York.

[44] Stynes, M. (1993). Numerical solution of convection-diffusion problems. *Department of Mathematics, University College Cork, Ireland Technical report.*

[45] Süli, E. (1991). Convergence of finite volume schemes for Poisson's equation on nonuniform meshes. *SIAM J. Numer. Anal. 28*, 5, 1419-1430.

[46] Wilkinson, J.H. *The algebraic eigenvalue problem.* Oxford University Press.

[47] Zienkiewicz, O.C. (1978). Reply by O. C. Zienkiewicz to comments by Gartling on a paper by Heinrch, Huyakorn, Zienkiewicz and Mitchell. *Int. J. Num. Meth. Engng., 12*, 191.