Loughborough University

A NOVEL BUBBLE FUNCTION SCHEME FOR THE FINITE ELEMENT SOLUTION OF ENGINEERING FLOW PROBLEMS

by

Alireza Yazdani

A doctoral thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy degree of Loughborough University

To Leila

Acknowledgement

I am indebted to my supervisor Professor Vahid Nassehi for his fruitful supervision of my PhD research. His help, support and invaluable advice are highly appreciated.

I am thankful to Professor Richard Wakeman for his constructive suggestions during this work.

Loughborough University is acknowledged for the financial support that made this work possible.

All my school and university teachers are remembered and acknowledged.

I wish to thank my family for their support and encouragement during the hard times.

And I would like to thank Leila for her constant help and encouragement that made possible, the start and the end of this work.

Abstract

This thesis is devoted to the study of some difficulties of practical implementation of finite element solution of differential equations within the context of multi-scale engineering flow problems. In particular, stabilized finite elements and issues associated with computer implementation of these schemes are discussed and a novel technique towards practical implementation of such schemes is presented. The idea behind this novel technique is to introduce elemental shape functions of the polynomial forms that acquire higher degrees and are optimized at the element level, using the least squares minimization of the residual. This technique provides a practical scheme that improves the accuracy of the finite element solution while using crude discretization. The method of residual free bubble functions is the point of our departure.

Residual free bubble functions yield accurate solutions for the problems of different scales of amplitude in the variations of the field unknown. These functions, however, are not readily derivable and due to their complex forms, they are not usually significant from a practical point of view. Computation of a residual free bubble function involves the solution of the local residual differential equations, which can be as difficult as the solution of the original problem. These will result in lack of flexibility or impracticality, especially in higher dimensions and non-symmetric problems.

We benefit from the advantages of polynomials that are continuous, differentiable and easily integrated and derive practical polynomial bubble functions that approximate the residual free bubble functions, using the method of least squares minimization. We employ our technique to solve several problems and show its practicality and superiority over the classical linear finite elements.

Table of Contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

Modelling the real world phenomena, in order to study and control them, requires mathematical formulation of their governing rules. This formulation is usually expressed in terms of Differential equations. Trying to solve these equations is trying to discover the hidden patterns ruling these phenomena, which results in control or in prediction as required. Patterns, however, are numerous and the solutions are rare. Solutions to the differential equations exist only for a limited number of equations and these solutions may not be readily applicable.

Methods of the treatment of differential equations usually fall in one of the following groups: analytical or exact solutions, approximate methods and numerical methods. Exact analytical methods solve a given differential equation strongly and present a closed form of the solution. This means that the solution satisfies the equation with no residuals within its definition domain and is expressed explicitly. Approximation methods are those methods that locally or globally approximate the solution. These methods provide a simple explicit form for the solution of the equation; however, a residual of the approximation is generated. Numerical methods are methods that obtain the global solution, based on joining the local point-wise solutions. A numerical solution of the differential equations is usually obtained from discretization of the problem domain and solving a set of local problems that will provide recursive formulas.

Several approaches there also exist that employ a combination of two or all of the above-mentioned methods to study certain problems.

The method of finite elements is a powerful numerical method to solve differential equations. This method is developed well, both in theory and techniques and is broadly employed to solve differential equations arising in the study of fluid dynamics, structure analysis, aerodynamics and so forth. This method employs a systematic procedure to solve the given problem and acquires considerable power and flexibility especially in coping with the complex geometry domains or problems with different degrees of desired precision over the entire domain or the so called multi-scale problems. The range of problems suitable for analysis by finite element method is clearly large.

A brief anatomy of the method is as follows: Consider a phenomenon expressed in terms of governing differential equation(s) over a prescribed domain. The first step in the finite elements solution of a differential equation is to perform a domain discretization. At this stage, the problem domain is discretized into a finite number of sub-domains (element domains). The second step is to assume an approximation of the solution that is written in a linear combination of a set of basis functions. These basis functions come from finite dimensional spaces associated to each element domain and have small supports. The next step is to insert this approximation into the weak form of the problem, the so-called weighted residual statement, and to force this residual to be zero in an average sense. Making the weighted residual zero, gives rise to a local system of linear equations. Performing the assembly process of the local equations along with the imposition of initial and boundary conditions and removal of the redundant equations, solves the global system of the equations that find the values of the unknown function at the selected element nodes. In simple words, the method

of finite elements is based on the reduction of an initial/boundary value differential equation to a matrix system of equations, which its solution provides the estimations to the solution at selected nodes.

1.1 Difficulties and common approaches with the exercise of finite element schemes

An important challenge in the exercise of a finite element procedure is to produce a stable numerical scheme, which prevents the errors in input data and intermediate approximations to accumulate and cause a meaningless solution. A discretization error, on the other hand, is likely to be generated due to the geometrical complexities.

In order to overcome these difficulties, several techniques are suggested and employed in the literature. All these techniques are based on the ground of one or both of the following: refining element domains (h-version finite element) or changing order of base functions (p-version finite element). In the h-version, the computational grid is refined at each mesh refinement level to improve the accuracy of approximation. In the p-version, however, the accuracy is improved using higher order elements. In addition, alterations to the h-version and p-version finite elements are used extensively, to approach several special cases.

Revisions to the p-version finite elements vary from use of higher order polynomials or sophisticated shape functions such as exponential or hyper-trigonometric functions, to modifying the weighted residual statements with upwinding [26].

One efficient method to achieve higher order shape functions is to introduce the hierarchical shape functions that contribute to the approximation by providing higher order refinements whereas the successive coefficients of the added terms are less important resulting in a larger tolerance of numerical inaccuracy [34].

Another selection is to assume a multi-component element consisting of standard approximation plus an additional part. The additional part is to be found exactly from the solution of the local residual equation generated by replacement of the linear part into the original equation. This approach is called enriched finite element method, which gives rise to the introduction of multi-scale functions and the residual free bubble functions. The method improves accuracy of the finite element solution, enriching the standard Galerkin approximation, and finds a stable and coarse-mesh accurate finite element discretization.

1.2 Notable examples of ongoing research: Multi-scale problems

Quantitative analysis of multi-scale problems has become an important issue in the engineering flow processes. Mathematical models of such problems are often expressed in terms of complex PDEs and their solution requires sophisticated numerical techniques. The basic concept of a multi-scale problem is explained below via comparisons between the free and porous flow regimes with different physical properties. A thorough analysis can be found in [22] and references therein.

Figure 1.1 shows a schematic diagram of a laminar plug flow where the domain is open to flow and its walls are not permeable). The flow is subject to perfect slip wall condition. In this case no stress is carried by the fluid. Such a free flow regime can be described mathematically by the use of Euler equations.

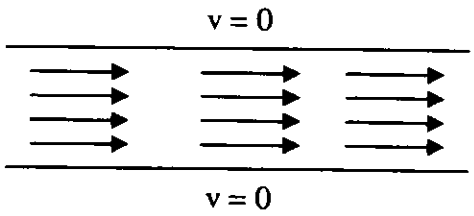


Figure 1.1

Plug flow regime (no slip wall)

In figure 1.2 the laminar free flow regime where the flow is subject to no slip wall conditions is shown. In this case the fluid carries all of the stress and becomes deformed. This flow regime can be described by Stokes or Navier-Stokes equation (depending on the Reynolds number).

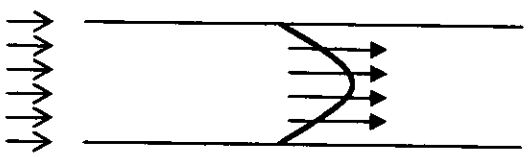


Figure 1.2

Velocity profile of Free flow regime (no permeability)

Figure 1.3 shows the physical features of a porous flow regime with high permeability (i.e. the domain consists of large pores). In this case the velocity at the walls is zero (i.e. no slip wall conditions). The fluid no longer carries all of the stress and some is borne by the porous medium. Such a porous flow phenomenon can be described mathematically by the Brinkman equation.

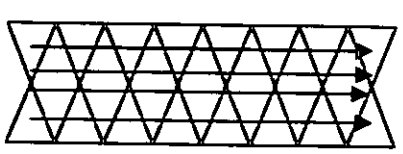


Figure 1.3

Porous flow regime (high permeability)

Figure 1.4 gives a representation of the physical aspects of a porous flow regime with very low permeability (i.e. the porous medium is dense and has very fine pores). In this case a slip wall condition is established and the velocity has a flat profile across the porous material. The stress is now carried completely by the solid matrix. Such a porous flow phenomenon can be described mathematically by the Darcy equation.

Note that although the velocity profile in this case will be similar to the one shown in figure 1.1 the mathematical representation of flow in the two cases will be very different. This is because the fluid viscosity plays no role in the free plug flow case and in contrast has a significant effect on the nature of a low permeability flow system.

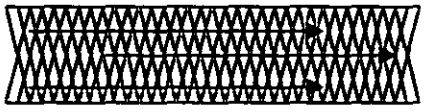


Figure 1.4
Porous flow regime (low permeability)

In figure 1.5 the typical velocity profile in a porous flow system where the permeability is high is shown. Amongst all of the regimes described here only the latter case can be regarded as a multi-scale flow problem. This is because the flow pattern at layers near the wall is very different in character to the established flow pattern within the domain. Inside the domain the profile will be plug flow but near the walls it will change very abruptly to a parabolic type.

Although Brinkman equation is able to characterize the flow in highly permeable porous medium with low Reynolds numbers the multi-scale nature of the flow makes

it necessary to use excessive domain discretization near the walls to obtain a good solution.

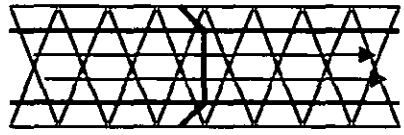


Figure 1.5

Velocity profile in a Porous flow regime with high permeability

The standard Galerkin finite element method is not a strong enough approach for transport models displaying multi-scale behaviour [10]. For these problems, a particular class of sub-grid scale models are proposed which are known as multi-scale methods [18].

This is mainly due to the fact that the representation of all physical scales needs a high level of discretization which is a common difficulty with these problems. If the discretization at a coarse level ignores the fine scale then the solution will be unstable and inaccurate. The influence of the fine scales must be incorporated into the model. If the flow occurs in highly permeable porous media the thickness of the boundary layer decreases by reducing the permeability. The discretization level must be less than the boundary layer thickness to achieve stable solution [27]. This problem can be satisfactorily resolved by the use of higher order approximation functions. Therefore if the problems related to 'numerical locking' can be resolved then methods based on such trial functions will be the appropriate technique for multi-scale flow problems.

Similar multi-scale behaviour can be seen in turbulent flows, large scale molecular dynamic simulations, weather forecasting, reaction and convection dominated transport problems.

1.3 Aim of this thesis; our approach

The main aim of this work is to contribute to the practical implementation of the p-version finite elements, especially within the framework of variational and multi-scale methods. The work is particularly concerned with the derivation of polynomial approximations of the residual free bubble functions. This is carried out in such a way that the accuracy of the solution with residual free bubble functions is not sacrificed by the selection of simple approximants. Indeed, several factors need to be taken into account, to this end.

The exact solutions to some differential equations (if available) are expressed in terms of sophisticated functions. This varies from the presence of special functions (e.g. Airy function) to the presence of oscillations in the solution of certain equations. Moreover, derivation of the residual free bubble functions involves the solution of local differential equations that can be as difficult as the original equation in some occasions. Therefore, adoption of a polynomial approximation seems to be a good choice to overcome these difficulties. This is because families of polynomials can, uniformly approximate functions of certain degree of smoothness. The polynomial bubble, however, has the property that it vanishes at the element boundary. Intuitively, this property confines the approximation error to the element level only. The next question to be answered is that what degree polynomial to use? It is clear that in order to capture the sharp drops or oscillations in the variations of the unknown in the problem domain, higher order polynomials are required. However, adoption of the high order polynomial might result in over-smoothness and over-convergence where a simple approximation is able to capture the variations. This gives rise to the analysis of the approximation error related to the size of the element. To make a balance, therefore, it is required to perform a moderate mesh refinement along with

the employment of the higher order approximation. This is to use both h-version and p-version finite elements at the same time. The above discussion implies that an optimal scheme is needed in order for an enriched finite element approximation to satisfy certain criteria of accuracy and practicality. The least squares polynomial approximation of the residual free bubble functions is a potential candidate, which meets the requirements of the criteria, highlighted above. In this thesis, we will introduce polynomials of the orders higher than the order of standard linear elements by them we enrich the standard finite elements. By minimizing the residual functional, generated from the replacement of these approximants into the original equation, we find the optimal polynomials that we will call them practical bubble functions. These practical bubble functions, along with a moderate refinement of computational mesh produce satisfactory results for the problems at hand.

1.4 Structure of the thesis

This thesis consists of five Chapters. In the introductory Chapter, we provide a brief background to this study and present the main motivations and issues associated with this work towards the justification of this research. In Chapter two, a comprehensive review of the literature is carried out. The main tools and ideas and novel techniques associated with the finite element solution of the engineering flow problems are presented and their advantages and disadvantages are counted. In Chapter three, a novel method for the derivation of bubble functions using the method of least squares is introduced, explained and tested for two-point boundary value problems and is compared to other existing methods. Chapter four of this thesis, studies the extension of the least squares bubble functions to one-dimensional time-dependent problems and derivation of such functions for the triangular and rectangular elements in multi-

dimensional analysis. Similar to Chapter three, Chapter four includes illustrative worked examples in order to demonstrate the efficiency of the suggested technique. Finally, in Chapter five conclusions of the present study and its possible extensions to obtain results that are more general and to solve more challenging and multi-dimensional problems are discussed. The thesis also includes list of references and appendices of the Maple works for the derivation of the bubble functions.

Chapter 2

Finite elements, Approximations and the Variational methods

This Chapter is devoted to the study of the approximation methods and, in particular, the development of novel finite element schemes for the solution of complex problems. The main aim of this chapter is to review different ideas and techniques employed in the variational formulation of finite elements. We start from introducing the most essential tools and ideas in the finite element method and carry out a literature survey on the problems that are practically important and consider the difficulties associated with them. We also present the ideas behind the variational methods, their benefits in coping with multi-scale problems and discuss the implementation and limitation of these methods.

2.1 A survey on principles of approximation and interpolation used in the finite element schemes

The main aim in approximation theory is to approximate functions from certain infinite-dimensional spaces (primarily $C[a,b]$) by means of simpler functions generally coming from a finite-dimensional space [24].

The approximation space should have certain properties:

It should be expandable to sufficiently large space to get a good approximation over there. On the other hand, it should be possible to get arbitrarily good approximations to a given function by making the dimension of approximating space sufficiently large, that is, the approximations should converge in some sense. Elements of the

approximation space should be simple so that they can be easily integrated and differentiated. There should be a well-developed theory to facilitate the analysis of the resulting computational procedures. Polynomials are an ideal choice on all three accounts. This fact is a result of a classical theorem of Weierstrass.

Weierstrass Approximation theorem: Let $f \in C[a,b]$. Then for any $\varepsilon>0$ there exists a polynomial p_n such that $\max_{a\le t\le b}|f(t)-p(t)| \le \varepsilon$, [24].

The best known method of approximation is polynomial interpolation, which consists of finding a polynomial $p_n(t)$ taking on pre-assigned values w_i at certain points t_i. This type of interpolation is called Lagrange interpolation. The Lagrange interpolation problem always has a unique solution with a simple representation

$$p_n(t)=\sum_{j=0}^{n} l_j(t)f(t_j) \tag{2.1}$$

where the l_j are the so-called fundamental polynomials

$$l_j(t)=\frac{(t-t_0)(t-t_1)...(t-t_{j-1})(t-t_{j+1})...(t-t_n)}{(t_j-t_0)(t_j-t_1)...(t_j-t_{j-1})(t_j-t_{j+1})...(t_j-t_n)}, \quad \text{j=0, 1, 2,..., n.} \tag{2.2}$$

for further discussion, see [24].

2.2 Numerical Integration

Numerical integration is the subject of approximating the value of an integral when the integrand is known either, as an expression, a table of values, or a computer subroutine and there is no straightforward method to calculate the exact value.

One of the key steps in the finite element schemes is where the derivation of element matrices for higher order elements is carried out. The complexity of the functions under the integral sign as well as the difficulties of evaluation of the derivatives in distorted integration domains makes it inevitable to approximate numerical

evaluations be used instead of the typical integrals [15]. In general, the strategy is to pass an approximant, say a polynomial, through the points defined by the function, and then integrate this polynomial approximation of the function. There exist several methods to do this such as midpoint method, trapezoidal method, Simpson method, and Newton-Cotes formulas in general. The theory of integration is a generalization of the theory of finite series into infinite series. The associated idea is as follows: for a given integral which is difficult to calculate, we restrict ourselves to the constructive part of the problem i.e. to the original finite summation, by which we can pass through to the integral formula via a formal procedure i.e. taking limit or supremum. This involves discretization of the integral domain, selection of a fast and suitable approximant and a method to reduce generated error. Formulation of this idea gives birth to the theory of quadratures.

A classical quadrature has the form $I(f) = \int_a^b f(x)dx = \sum_{i=1}^{n} w_i f(x_i) + E$, in which $w(x)$ is called a weight function, the set $\{x_i\}$ are the abscissa or nodes, the set $\{w_i\}$ are the weights, n is the point number and E the error term. Setting $Q(f) = \sum_{i=1}^{n} w_i f(x_i)$, Q is a functional which gives an approximate value for f, where the abscissa x_i and weights w_i are known constants depending on n, I, $w(x)$ and on the interval $[a, b]$, but independent of f (sometimes we write Q_n for an n-point formula). A formula is said to be of *m-th* degree or precision if it integrates exactly all polynomials of degree m or less but there exist some polynomials of degree $m+1$ (and higher) that the formula is not exact applied to them. We set $R(f) = I(f) - Q(f)$ and call it the truncation error. A formula then is called to be convergent if $R(f) \to 0$ as $n \to \infty$. An integrand is called to be rapidly oscillatory if it assumes numerous (say more than ten) local maxima and minima over a relatively small range of integration. Functions of

infinitely many local maxima and minima around a given point, frequently happens to appear in practice [15].

If a rule has abscissa $\{x_i\}$, none of which is equal to either of the end-points a or b, then the rule is called an open rule. Such open rule formulae can be used to evaluate integrals with integrands, which exhibit end-point singularities, as no function values are required at these points. If the end-points are included in the set of abscissa, then the rule is called a closed rule. Formulas in which the range of integration is partitioned into equal subintervals, and nodal values are predetermined, are called Newton-Cotes formulas [15]. The best-known examples of these integration formulas are: Trapezoidal Rule:

$$\int_a^b f(x)dx - \frac{b-a}{2}\{f(a)+f(b)\} = -\frac{(b-a)^3}{12}f^{(2)}(\xi), \quad a<\xi<b.$$

for $f \in C^2[a, b]$ and Simpson's Rule:

$$\int_a^b f(x)dx - \frac{b-a}{3}\{f(a)+4f(\frac{a+b}{2})+f(b)\} = -\frac{(b-a)^5}{90}f^{(4)}(\xi), \quad a<\xi<b.$$

for $f \in C^4[a, b]$.

For explicitly known integrands, we may use other methods to obtain a more accurate approximation. Such methods are called Gaussian integration methods. Remember the formula $\int_a^b f(x)dx = \sum_{i=1}^{n} w_i f(x_i) + E$. It is customary to shift the integration range to $[-1, 1]$ by setting $x = \frac{1}{2}(a(t-1)+b(t+1))$. Then we determine the n coefficients w_i and n nodes x_i so that the formula gives exact results for polynomials of degree k as high as possible. Since $n+n=2n$ is the number of coefficients of a polynomial of degree $2n-1$, it follows that $k \leq 2n-1$. Gauss has shown that exactness for polynomials of degree not exceeding $2n-1$ (instead of $n-1$ for predetermined nodes) can be attained,

and he has given the location of the x_i (the *i-th* zero of the Legendre polynomial P_n) and the coefficients w_i (which depend on n but not on f). This formula is called Gauss quadrature formula [15]. Gaussian quadratures are preferred to Newton-Cotes formulas for finite element applications because they have fewer function evaluations for a given order.

We will determine the parameters in the simple case of a two-term formula containing four unknown parameters: $\int_{-1}^{1} f(x) = af(x_1) + bf(x_2)$. Our formula is to be valid for any polynomial of degree 3. Hence, it will hold if $f(x) = x^3, f(x) = x^2, f(x) = x$, and $f(x) = 1$:

$$f(x) = x^3: \quad \int_{-1}^{1} x^3 dx = 0 = ax_1^3 + bx_2^3;$$

$$f(x) = x^2: \quad \int_{-1}^{1} x^2 dx = \frac{2}{3} = ax_1^2 + bx_2^2;$$

$$f(x) = x: \quad \int_{-1}^{1} x dx = 0 = ax_1 + bx_2;$$

$$f(x) = 1: \quad \int_{-1}^{1} dx = 2 = a + b.$$

We then find that

$$a = b = 1,$$

$$x_2 = -x_1 = \sqrt{\frac{1}{3}} = 0.5773,$$

$$\int_{-1}^{1} f(x) dx = f(-0.5773) + f(0.5733).$$

It is remarkable that adding these two values of the function gives the exact value for the integral of any cubic polynomial over the interval from *–1 to 1*.

Suppose our limit of integration are from a to b, and not –1 to 1 for which we derived this formula. To use the tabulated Gaussian quadrature parameters, we must change the interval of integration to *(-1,1)* by a change of variable. We replace the given

variable by another to which it is linearly related according to the following scheme:

If we let $t = \frac{(b-a)x+b+a}{2}$ so that $dt = (\frac{b-a}{2})dx$ then:

$$\int_a^b f(t)dt = \frac{b-a}{2}\int_{-1}^{1} f(\frac{(b-a)x+b+a}{2})dx.$$

As an example if $I = \int_0^{\frac{\pi}{2}} \sin x dx$ (the exact value of I is equal to 0.1), we change the variable of integration to make the limits of integration *(–1, 1)*.

Let $x = \frac{(\frac{\pi}{2})t + \frac{\pi}{2}}{2}$, so $dx = \frac{\pi}{4}dt$ and $I = \frac{\pi}{4}\int_{-1}^{1} \sin(\frac{\pi t + \pi}{4})dt$.

The Gaussian formula calculates the value of the new integral as a weighted sum of two values of the integrand, at *t=-0.5773* and *t=0.5773*. Therefore, $I = \frac{\pi}{4}[(1.0)(\sin(0.10566\pi) + (1.0)(\sin(0.39434\pi)] = 0.99847$ and the value of the error is 1.53×10^{-3} [15]. Gaussian quadrature can be extended beyond two terms. The formula is then given by $\int_{-1}^{1} f(x)dx = \sum_{i=1}^{n} w_i f(x_i)$ for n points. This formula is exact for polynomials of degree *2n-1* or less. Moreover, by extending the method we used previously for the 2-point formula, for each n we obtain a system of $2n$ equations:

$$w_1 x_1^k + \ldots + w_n x_n^k = \begin{cases} 0 & for\ k = 1,3,5,\ldots,2n-1 \\ \frac{2}{k+1} & for\ k = 0,2,4,\ldots 2n \end{cases} \qquad (2.4)$$

It turns out that the t_i for a given n are the roots of the nth-degree Legendre polynomial. The Legendre polynomials are defined by recursion: $(n+1)L_{n+1}(x) - (2n+1)xL_n(x) + nL_{n-1}(x) = 0$ with

$L_0(x) = 1, L_1(x) = x, L_2(x) = \frac{3}{2}x^2 - \frac{1}{2}$, and zeros at $\pm\sqrt{\frac{1}{3}} = \pm 0.5773$.

The following table [15] lists the zeros of Legendre polynomials up to degree 5, giving values that we need for Gaussian quadratures where the equivalent polynomial is up to degree 9. For example, $L_3(x)$ has zeros at $x=0, +0.77459667, -0.77459667$.

Number of terms	Values of t	Weighting factor	Valid to degree
2	-0.57735027 0.57735027	1.0 1.0	3
3	-0.77459667 0.0 0.77459667	0.55555555 0.88888889 0.55555555	5
4	-0.86113631 -0.33998104 0.33998104 0.86113631	0.34785485 0.65121451 0.65121451 0.34785485	7
5	-0.90617985 -0.53846931 0.0 0.53846931 0.90617985	0.23692689 0.47862867 0.56888889 0.47862867 0.23692689	9

Table 2.1
Gauss quadrature nodes and weights

The Legendre polynomials are orthogonal over the interval $[-1,1]$. That is,

$$\int_{-1}^{1} L_n(x)dx \begin{cases} = 0 \text{ if } n \neq m; \\ > 0 \text{ if } n = m. \end{cases}$$

Any polynomial of degree n can be written as a sum of the Legendre polynomials: $P_n(x) = \sum_{i=0}^{n} c_i L_i(x)$. The n roots of $L_n(x) = 0$ lie in the interval $[-1, 1]$.

It is a desired property of Gaussian quadratures that are well applicable to evaluate multiple integrals numerically, within the finite element packages. We consider first the case when the limits of integration are constant.

For a multiple integral over the domain $[-1,1]\times[-1,1]\times...\times[-1,1]$ we shall write:

$$I = \int_{-1}^{1}\int_{-1}^{1}...\int_{-1}^{1} f(x,y,...,z)dxdydz = \int_{-1}^{1}(\int_{-1}^{1}(...(\int_{-1}^{1} f(x,y,...,z)dx))dy...)dz\ .$$

Applying one dimensional Gaussian rule to each integral we find the approximate value $I \cong \sum_{i=1}^{n}\sum_{j=1}^{m}...\sum_{k=1}^{l} a_i a_j ...a_k f(x_i, y_j, ..., z_k)$, where n, m,..., l are the number of nodes used for approximation of each integral. If the limits of integration are not constant values (e.g. the triangular finite elements), the used procedure needs to be slightly modified.

Consider that any finite element domain can be approximately meshed into triangles or a number of rectangles (boxes, respectively) and a number of triangles (simplexes, respectively). If the region has curved boundaries then it cannot be completely filled up with boxes and simplexes.

Consider a two dimensional problem. The double integral of $f(x, y)$ over the domain R can be approximated by a sum $\sum_{i=1}^{N} w_i f(x_i, y_i)$ where the (x_i, y_i), $i = 1, \ldots, N$ are the centres of the rectangles which lie inside R and B_i is the area of the rectangle containing (x_i, y_i). One must decide what to do near the boundary. If the mesh is relatively small it would be reasonable to include in the sum those rectangles whose centres lie inside R and exclude those rectangles whose centres lie outside.

The alternative method is to write, if possible, a multiple integral as an iterated integral that is to write $\iint_R f(x,y)dxdy = \int_a^b [\int_{\Phi(x)}^{\Psi(x)} f(x,y)dy\,]dx$.

To approximate the iterated integral on the right side we proceed as follows. We select a one-variable formula for the variable x:

$$\int_a^b g(x)dx \cong \sum_{i=1}^{M_1} w_i^1 g(v_i) \qquad a \le v_1 < v_2 < ... < v_{M_1} \le b\ , \tag{2.5}$$

Select a second integration formula to approximate each of the following integrals:

$$\int_{\Phi(v_i)}^{\Psi(v_i)} f(v_i, y)dy \cong \sum_{j=1}^{M_2} w_{i,j}^2 f(v_i, \xi_{i,j}). \tag{2.6}$$

The above formula will be different for different i but usually one will pick one formula and transform it appropriately to each of the intervals $[\Phi(v_i), \Psi(v_i)]$. Finally we obtain:

$$\int_a^b \int_{\Phi(x)}^{\Psi(x)} f(x, y)dydx = \sum_{i=1}^{M_1} w_i^1 \sum_{j=1}^{M_2} w_{ij}^2 f(v_i, \xi_{ij}). \tag{2.7}$$

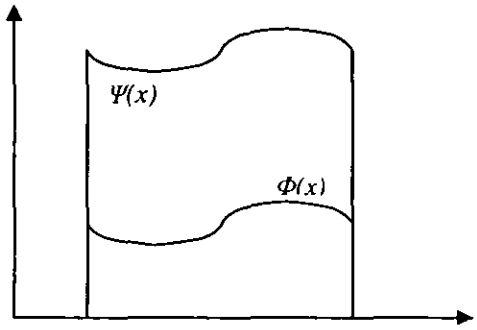


Figure 2.1
Variable integration limits

A similar method can be used in three dimensions [33].

Here, we construct product formulas for T_n the n-simplex with vertices

(0,0,0,...,0,0), (1,0,0,...,0,0), (0,1,0,...,0,0),..., (0,0,0,...,0.1).

T_2 is a triangle and T_3 is a tetrahedron. The integral of a monomial over T_n is

$$\int_0^1 \int_0^{1-x_1} \int_0^{1-x_1-x_2} \cdots \int_0^{1-x_1-\cdots-x_{n-1}} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} dx_n \cdots dx_1 \tag{2.8}$$

Let us transform the integral using the transformation

$x_1=y_1=y_1$

$x_2=y_2(1-y_1)=y_2(1-x_1)$

$x_3=y_3(1-y_2)(1-y_1)=y_3(1-x_1-x_2)$

$\vdots$

$x_n=y_n(1-y_{n-1})\ldots(1-y_1)=y_n(1-x_1-\ldots-x_{n-1})$

Since the limits of integration for the x_i are $0 \le x_i \le 1-x_1-\ldots-x_{i-1} \quad i=1,\ldots,n$, the limits for the y_i will be $0 \le y_i \le 1 \quad i=1,\ldots,n$.

Since the Jacobean of the transformation is $J=(1-y_1)^{n-1}(1-y_2)^{n-2}\ldots(1-y_{n-1})$. The monomial integral transforms into

$$\int_0^1\ldots\int_0^1(1-y_1)^{\beta_1}\ldots(1-y_{n-1})^{\beta_{n-1}}y_1^{\alpha_1}\ldots y_n^{\alpha_n}dy_1\ldots dy_n \tag{2.9}$$

$$\begin{aligned}\beta_1&=\alpha_2+\ldots+\alpha_n+n-1\\ \beta_2&=\alpha_3+\ldots+\alpha_n+n-2\\ &\vdots\\ \beta_{n-1}&=\alpha_n+1.\end{aligned}$$

This integral is a product of n single integrals, where the integral with respect to y_k has the form $\int_0^1(1-y_k)^{n-k}P_\alpha(y_k)dy_k \quad k=1,\ldots,n$ where $P_\alpha(y_k)=y_k^{\alpha_k}(1-y_k)^{\alpha_{k+1}+\ldots+\alpha_n}$. Therefore, if we have n one-variable formulas, each of degree d, of the form $\int_0^1(1-y_k)^{n-k}f(y_k)dy_k \equiv \sum_{i=1}^{M}w_{k,i}f(\mu_{k,i}) \quad k=1,\ldots,n$, these can be combined to give a formula of degree d for T_n [33]. For practical reasons, we evaluate the integrals of monomials over triangle T_2 and tetrahedron T_3. For a monomial over T_2 we find by direct integration $I=\int_0^1\int_0^{1-x}x^\alpha y^\beta dydx=\frac{\Gamma(\beta+2)\Gamma(\alpha+1)}{(\beta+1)\Gamma(\alpha+\beta+3)}$, where the Gamma function defined by $\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt=(x-1)\Gamma(x-1)$ is the generalized factorial function, that is, if x is an integer $n=1,2,3,\ldots$, then we have $\Gamma(n)=(n-1)\Gamma(n-1)=(n-1)(n-2)\Gamma(n-2)=\ldots=(n-1)(n-2)\ldots1=(n-1)!$.

Similarly, for a monomial over T_3 we have:

$$J=\int_0^1\int_0^{1-x}\int_0^{1-x-y}x^\alpha y^\beta z^\lambda dzdydx=\frac{\Gamma(\lambda+2)\Gamma(\beta+1)\Gamma(\alpha+1)}{(\lambda+1)\Gamma(\alpha+\beta+\lambda+4)}. \tag{2.10}$$

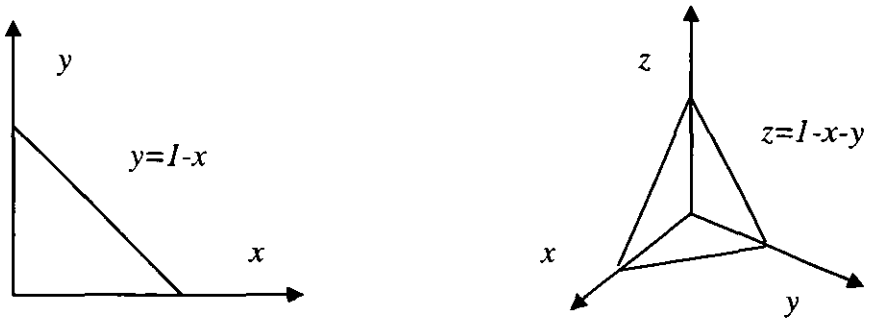


Fig 2.2
Standard triangular and tetrahedral regions

Using the Gaussian quadrature formula for numerical approximation of the above-mentioned integrals (over T_2 and T_3 respectively) we find

$$I = \int_0^1 \int_0^{1-x} x^\alpha y^\beta dydx = \int_0^1 \int_0^1 x^\alpha (1-x)^{\beta+1} t^\beta dtdx = (\int_0^1 x^\alpha (1-x)^{\beta+1} dx)(\int_0^1 t^\beta dt)$$

$$I = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} x_i^\alpha (1-x_i)^{\beta+1} t_j^\beta$$

and

$$J = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} x^\alpha y^\beta z^\lambda dzdxdy = \int_0^1 \int_0^{1-x} \int_0^1 x^\alpha y^\beta t^\lambda (1-x-y)^{\lambda+1} dtdydx$$

$$= \int_0^1 \int_0^1 \int_0^1 x^\alpha (1-x)^{\alpha+\beta+2} s^\beta (1-s)^{\lambda+1} t^\lambda dtdsdx = (\int_0^1 x^\alpha (1-x)^{\alpha+\beta+2} dx)(\int_0^1 s^\beta (1-s)^{\lambda+1} ds)(\int_0^1 t^\lambda dt)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{l} w_{ijk} x_i^\alpha (1-x_i)^{\alpha+\beta+2} (1-s_j)^{\lambda+1} s_j^\beta t_k^\lambda$$, and, m, n and l need not to be selected

equal. The nodal points and weights are given by tabulated values.

As an example the value of $I = \int_0^1 \int_0^{1-x} (xy + y^2) dydx$ is calculated as:

$$I = \int_0^1 \int_0^1 \{ xt(1-x) + t^2 (1-x)^2 \}(1-x) dtdx$$

$$I = \frac{1}{4} \int_{-1}^1 \int_{-1}^1 \{ (\frac{x+1}{2})(\frac{t+1}{2})(\frac{1-x}{2})^2 + (\frac{t+1}{2})^2 (\frac{1-x}{2})^3 \} dtdx$$

$$I = \frac{1}{4} \int_{-1}^1 \int_{-1}^1 (\frac{x+1}{2})(\frac{t+1}{2})(\frac{1-x}{2})^2 dtdx + \frac{1}{4} \int_{-1}^1 \int_{-1}^1 (\frac{t+1}{2})^2 (\frac{1-x}{2})^3 dtdx$$

$$I \cong \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}\left(\frac{x_j+1}{2}\right)\left(\frac{t_i+1}{2}\right)\left(\frac{1-x_j}{2}\right)^2 + \frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}\left(\frac{t_i+1}{2}\right)\left(\frac{1-x_j}{2}\right)^3$$

Taking $m=n=3$ and using table 2-1 we find $I \cong 0.121$ while $I = \frac{1}{8} = 0.125$ by direct integration.

2.3 Finite Elements, Variational Formulation

In this section, we briefly describe the construction of finite element method for boundary value problems and outline some of their properties.

The first step in the construction of a finite element method is to convert the problem into its weak formulation:

find $u \in V$ such that $a(u,v) = l(v) \qquad \forall v \in V$ (2.11)

where V is the solution space (e.g. $H_0^1(\Omega)$ for the homogeneous Dirichlet boundary value problem), $a(.,.)$ is a bilinear functional on $V \times V$ and $l(.)$ is a linear functional of V.

The second step in the construction is to replace V in (2.11) by a finite-dimensional subspace $v_h \in V$ which consists of continuous piecewise polynomial functions of a fixed degree associated with a subdivision of the computational domain;

Then consider the following approximation of (2.12):

find $u_h \in V_h$ such that $a(u_h, v_h) = l(v_h) \qquad \forall v_h \in V_h.$ (2.12)

Suppose, for example, that $\dim V_h = N(h)$ and $V_h = span\{\phi_1, ..., \phi_{N(h)}\}$, where the (linearly independent) base functions ϕ_i, $i = 1, ...N(h)$, have "small" support. Expressing the approximate solution u_h in terms of the basis functions, ϕ_i, we can write

$$u_h(x) = \sum_{i=1}^{N(h)} U_i \phi_i(x), \tag{2.13}$$

where U_i, $i = 1,...,N(h)$, are to be determined. Thus (2.12) can be written as follows:

find $U_1,...,U_{N(h)} \in \Re^{N(h)}$ such that

$$\sum_{i=1}^{N(h)} a(\phi_i, \phi_j) U_i = l(\phi_j), \qquad j = 1,...,N(h). \tag{2.14}$$

This is a system of a linear equations for $U = (U_1,...,U_{N(h)})^T$, with the matrix of the system $A = (a(\phi_j, \phi_i))$ of size $N(h) \times N(h)$. Because the ϕ_i's have small support, $a(\phi_j, \phi_i) = 0$ for most pairs of i and j, so the matrix A is sparse (in the sense that most of its entries are equal to 0); this property is crucial from the point of efficient solution- in particular, fast iterative methods are available for sparse linear systems. Once (2.14) has been solved for $U = (U_1,...,U_{N(h)})^T$, the expansion (2.13) provides the required approximation to u.

2.4 Finite element formulation of the fluid flow problems

The finite element method is widely used for the formulation and solution of fluid flow problems as well as solids structure problems. A major difference between the formulations for the analysis of fluid flows and of solids is the convective terms that give rise to the non-symmetry in the finite element coefficient matrix, and when the convection is dominant, the system of equations is strongly non-symmetric and then an additional numerical difficulty arises [3].

Before discussing this difficulty, we note that depending on the flow considered, as the convection dominance increases and when a certain range is reached, the flow

condition turns from laminar to turbulent. Under such condition, in order to model the details of turbulence, extremely fine discretization is required i.e. the resulting finite element systems become too large. For this reason, it is reasonable and more convenient to solve the governing equations by expressing the turbulence effects by means of turbulent viscosity and heat conductivity coefficients, and use wall functions (e.g. exponential fits) to describe the near-wall behaviours.

The modelling of turbulence is a very large and important field and the finite element procedures are in many regards directly applicable. In the following, we briefly address the difficulty of solving highly convection dominant flows. For this purpose, let us consider the simplest possible case that displays the difficulties that we encounter in general flow conditions. These difficulties arise from the magnitude of the convective terms when compared to the diffusive terms. We consider a model problem of one dimensional flow with prescribed velocity v. The temperature is prescribed at two points, which we label $x=0$ and $x=l$, and we want to calculate the temperature for $0<x<l$. The governing differential equation is:

$$\rho c_p \frac{du}{dx} v = k \frac{d^2 u}{dx^2} \tag{2.15}$$

With the boundary conditions

$$\begin{cases} u = u_l & at\ x = 0 \\ u = u_R & at\ x = l \end{cases} \tag{2.16}$$

and the left hand side in (2.15) represents the convective terms and the right hand side the diffusive terms.

The exact solution to the problem in (2.15) is given by:

$$\frac{u - u_l}{u_R - u_l} = \frac{\exp(\frac{P}{l} x) - 1}{\exp(P) - 1} \tag{2.17}$$

Where $P = vl/\alpha$ is called the Peclet number as $\alpha = k/\rho c_p$. It is well known that the numerical solution of (2.15) displays difficulties, as P increases, since the exact solution curve shows a strong boundary layer at $x = l$.

In order to demonstrate the difficulty in the finite element solution, let us use two node elements each of length h corresponding to a linearly varying temperature over each element. If we use the standard Galerkin method, for the finite element node i we get the governing equation

$$(-1-\frac{P^e}{2})u_{i-1} + 2u_i + (\frac{P^e}{2}-1)u_{i+1} = 0 \tag{2.18}$$

where $P^e = vh/\alpha$ is the element Peclet number. Therefore, we have:

$$u_i = \frac{1-P^e/2}{2}u_{i+1} + \frac{1+P^e/2}{2}u_{i-1} \tag{2.19}$$

This equation shows that for high values of P^e, unrealistic results are observed. For example, if $u_{i-1} = 0$ and $u_{i+1} = 100$, we have $u_i = 50(1-P^e/2)$, which gives a negative value if $P^e > 2$ [3].

The analytical solution of (2.15)-(2.16) shows that for a reasonably accurate solution P^e should be smaller than 2. This means that a very fine mesh is required when P^e is large. In practice, flows of very high Peclet numbers need to be solved. Therefore, the finite element discretization scheme must be amended to be applicable to such problems.

The shortcoming exposed above recognized and overcome by early researchers using finite difference method [3]. Considering (2.19), we realize that this equation is also obtained when central difference scheme is used to solve (2.15). Hence, the same solution inaccuracies are seen when the Euler's central difference method is used.

One remedy designed to overcome the above difficulties is to use upwind scheme. In the finite difference upwind scheme we use

$$\frac{du}{dx}\Big|_i = \frac{u_i - u_{i-1}}{h} \text{ if } v > 0 \text{ and } \frac{du}{dx}\Big|_i = \frac{u_{i+1} - u_i}{h} \text{ if } v < 0 \tag{2.20}$$

In the following discussion, we first assume $v > 0$ and then we generalize the results to consider any value of v.

If $v > 0$, the finite difference approximation of (2.15) is

$$(-1 - P^e)u_{i-1} + (2 + P^e)u_i - u_{i+1} = 0 \tag{2.21}$$

It can be seen that the results obtained with this upwinding is no longer oscillatory. This solution improvement is explained by the nature of the (exact) analytical solution: if the flow is in the positive x-direction, the values of u are influenced more by the upstream value u_l than by the downstream value u_R. Indeed, when P is large, the value of u is close to the upstream value u_l over much of the solution domain. The same observation holds when the flow is in the negative x-direction, but then u_R is of course the upstream value.

The intuitive implication of this observation is that in the finite difference discretization of (2.15), it should be appropriate to give more weight to the upstream value, and this is in essence accomplished in (2.21). Of course, it is desirable to further improve on the solution accuracy, and for the relatively simple (one dimensional) equation (2.15), such improvement is obtained using different approaches. We briefly present below some of the techniques that are actually closely related and result in good accuracy in one dimensional analysis cases. However, the generalization of these methods to obtain small solution errors using relatively coarse discretizations in general two and three dimensional flow conditions is difficult.

The first scheme that we consider is called exponential fitting. The basic idea of the exponential scheme is to match the numerical solution to the analytical (exact) solution, which is known in the case considered here.

To introduce the scheme, let us rewrite the (2.15) in the form

$$\frac{df}{dx} = 0 \tag{2.22}$$

where f is given by the convective minus diffusive parts,

$$f = vu - \alpha \frac{du}{dx} \tag{2.23}$$

The finite difference approximation of the relation in (2.22) for the *i-th* element gives

$$f\,|_{i+1/2} - f\,|_{i-1/2} = 0 \tag{2.24}$$

We now use the exact solution in (2.17) to express $f_{i+1/2}$ and $f_{i-1/2}$ in terms of the temperature values at the nodes $i-1, i, i+1$. Hence, using (2.17) for the interval i to $i+1$, we obtain

$$f_{i+1/2} = v[\,u_i + \frac{u_i - u_{i+1}}{exp(\,P^e\,) - 1}\,] \tag{2.25}$$

Similarly, we obtain an expression for $f_{i-1/2}$, and the relation (2.24) gives:

$$(-1-c)u_{i-1} + (2+c)u_i - u_{i+1} = 0 \tag{2.26}$$

where

$$c = exp(\,P^e - 1\,) \tag{2.27}$$

We notice that for $P^e = 0$ the relation (2.26) reduces to the use of the central difference method (and the Galerkin method) corresponding to the diffusive term only (because the convective term is zero) and that (2.26) has the form of (2.21) with c replacing P^e. This scheme, that is based on analytical solution of the problem, gives exact solution even when only very few elements are used in the discretization. The

scheme also yields very accurate solutions when the velocity v varies along the length of the domain and when source terms are included. A computational disadvantage is that the exponential functions need to be evaluated while in practice it is sufficiently accurate and more effective to use a polynomial approximation instead of the exact analytical solution.

The next scheme to be discussed is the method of Petrov-Galerkin, which is an extension of the classical Galerkin method. In the classical Galerkin method, the same trial functions are used to express the weighting and the solution. In principle, however, different functions may be employed and such an approach can provide increased solution accuracy, for certain types of problems.

In the Petrov-Galerkin method, weighting and trial base functions are selected to be different. Consider a finite element discretization of the problem domain and the weak formulation of problem within the $i-th$ element as:

$$\int_{-l}^{l} w_i v \frac{dh_j}{dx} u_j dx + \int_{-l}^{l} \frac{dw_i}{dx} \alpha \frac{dh_j}{dx} u_j dx = 0 \qquad j = i-1, i, i+1 \tag{2.28}$$

where w_i denotes the weighting function and the h_j are the usual functions of linear temperature distributions between nodes $i-1$, i and $i+1$.

The basic idea is now to choose w_i such as to obtain optimal accuracy. An efficient scheme is to use: $w_i = h_i + \gamma \frac{h}{2} \frac{dh_i}{dx}$ for $v > 0$ and $w_i = h_i - \gamma \frac{h}{2} \frac{dh_i}{dx}$ for $v < 0$.

Using this upwinding function in (2.28), we obtain for the case $v > 0$

$$[-1 - \frac{P^e}{2}(\gamma + 1)]u_{i-1} + (2 + \gamma P^e)u_i + [-\frac{P^e}{2}(\gamma - 1) - 1]u_{i+1} = 0 \tag{2.29}$$

We note that for $\gamma = 0$, the standard Galerkin finite element equation in (2.18) is recovered, and when $\gamma = 1$, the upwind finite difference scheme in (2.21) is obtained.

The variable γ can be evaluated such that nodal exact values are obtained for all values of P^e as $\gamma = coth(\frac{P^e}{2} - \frac{2}{P^e})$ [26]. The case $v < 0$ is solved similarly.

Another alternative is the Galerkin least squares method where the basic Galerkin equation is combined with a least squares expression to obtain good solution accuracy [3]. The least squares method applied to the equation (2.15) gives:

$$L[h_i] = v\frac{dh_i}{dx} - \alpha\frac{d^2h_i}{dx^2} \tag{2.30}$$

and

$$R_h = v\frac{du_h}{dx} - \alpha\frac{d^2u_h}{dx^2} \tag{2.31}$$

where the subscript h denotes the finite element solution corresponding to the mesh with elements of size h.

In the Galerkin least squares method the equation for the nodal variable u_i is generated by using the classical Galerkin expression and adding a factor τ times the least squares expression. The factor τ is optimized to obtain good solution accuracy.

Using for our problem the finite element discretization and evaluating the residual element by element (hence, the second derivative terms in (2.30) and (2.31) are zero), the i th equation is

$$\int_{-l}^{l} h_i v\frac{dh_j}{dx}u_j dx + \int_{-l}^{l}\frac{dh_i}{dx}\alpha\frac{dh_j}{dx}u_j dx + \int_{-l}^{l}(v\frac{dh_i}{dx})\tau(v\frac{dh_j}{dx}u_j)dx = 0,\ j = i-1, i, i+1 \tag{2.32}$$

where, in the last integral on the left hand side τ is the unknown parameter. To evaluate τ we can match the relation in (2.32) with the exact analytical solution (as we have done for the exponential scheme and the Petrov Galerkin method), and thus we obtain:

$$\tau = \frac{h}{2v} coth \frac{P^e}{2} - \frac{\alpha}{v^2} \tag{2.33}$$

It is interesting at this point to compare the exponential method, the Petrov Galerkin method and the Galerkin least squares procedures. Such a comparison shows that the equations (2.26), (2.29) and (2.32) with the optimal values of c, γ and τ, respectively, are identical within a factor (which has no effect because the right hand side of the equations is zero). For this reason, the solutions are identical. However, we should note that different solutions from the exponential scheme and the Petrov Galerkin method must in general be expected if a general source term is included in (2.15). For the linear approximation used here, the Galerkin least squares method gives the same solution as the Petrov Galerkin scheme [3].

An interesting observation and valuable interpolation is that all these methods are in essence equivalent to the Galerkin approximation with an additional diffusion term. If we write the Galerkin solution of (2.15) with an additional diffusion term $\alpha\beta$, we obtain:

$$\int_{-h}^{h} [\, h_i v \frac{dh_j}{dx} u_j + \frac{dh_i}{dx} (1+\beta) \alpha \frac{dh_j}{dx} u_j \,] dx = 0 \qquad j = i-1, i, i+1 \tag{2.34}$$

The solution of (2.34) is:

$$-(1+q)u_{i-1} + 2u_i - (1-q)u_{i+1} = 0 \tag{2.35}$$

where

$$q = \frac{P^e}{2} \frac{1}{1+\beta} \tag{2.36}$$

The value of β depends on the method which is used. For example, $\beta = 0$ yields the standard Galerkin technique and comparing (2.34) with (2.32), we find for the Galerkin least squares method $\beta = \frac{v\tau}{h} P^e$.

It is shown [6] that using the residual free bubble functions is, indeed, equivalent to introducing upwinding for convection-dominated problems. These functions couple only to the degree of freedom of the specific element considered. The principal idea is to compute the solution including the bubble functions and then, ignore the response in the bubbles. As an example, in one-dimensional analysis, we may use parabolic functions instead of linear elements as the parabolic variation, beyond the linear variation, corresponds to the bubble response [3]. The ideas of the residual free bubble functions and variational multi-scale methods are discussed in the following sections.

2.5 Residual free bubble functions

The development of variational multi-scale methods and the concept of residual free bubble functions have enabled researchers to cope with multi-scale problems beyond the power of classical finite elements. These techniques are particularly used to solve the finite element problems in which the chosen discretization level does not provide the stability properties. These methods are generally used in the transport problems with multi-scale behaviour in the form of interior and boundary layers such as turbulent flow, convection-diffusion equation and flow in porous media [27]. In this part, an intuitive description of these methods is presented and its superiorities over the classical finite elements are mentioned. The formulation of the scheme is presented in the following chapter.

To start, consider the approximate solution of the following boundary value problem:

$$\begin{cases} Lu = f, & in\ \Omega \\ B.C. & on\ \partial\Omega \end{cases} \tag{2.37}$$

as: $u = \bar{u} + u'$. The part $\bar{u}$ is a piecewise linear and u' is the analytical solution of the local residual differential equation:

$$\begin{cases} Lu' = -L\bar{u} + f & in\ \Omega_e \\ u' = 0 & on\ \partial\Omega_e \end{cases} \tag{2.38}$$

generated from insertion of $u - \bar{u}$ into the original equation (2.37), subject to homogeneous boundary conditions. The part $\bar{u}$ is called the residual free bubble function that strongly satisfies the equation (2.38). Now, consider a finite element discretization of the problem domain and repeat the above process within each element. We are interested to know the properties of the bubble functions and the processes involving their derivation. Bubbles are typically, higher order polynomials defined on the interior of finite elements that vanish on element boundaries [18]. The degrees-of-freedom associated with bubbles are eliminated by the well-known technique of static condensation. It is shown [18] that the element Green's function of the sub grid problem represents the ultimate residual free bubble. It follows that bubbles must somehow represent an approximation to the element Green's function. This idea goes back to 1980s where efforts took place to solve the Stokes problem. The concept of the bubble element was applied along with the Petrov-Galerkin method, which presented numerical schemes incorporating stability and accuracy of a higher degree than what was already in use. The variational multi-scale method was introduced in [18] and [19], through the procedures of modelling the multi-scale phenomena. It is motivated by the simple fact that straightforward application of Galerkin's method employing standard bases, such as Fourier series and finite elements, is not a robust approach in the presence of multi-scale phenomena. The variational multi-scale method seeks to rectify this situation. A simple description of the method is: some decomposition of the solution $u = \bar{u} + u'$ is sought, where we

think of solving for $\bar{u}$ numerically, but we attempt to determine u' analytically, eliminating it from the problem for $\bar{u}$. Indeed, $\bar{u}$ and u' may overlap or be disjoint, and u' may be globally or locally defined. The effect of u' on the problem for $\bar{u}$ will always be non-local. The part $\bar{u}$ represents coarse scales and the component u' fine scales [18]. Basic idea in the bubble function method is to decompose the solution of a given boundary value problem into the sum of a coarse scale solution and a fine scale one. The classical Galerkin finite element method is used for representation of the resolvable coarse scale part of the finite element mesh and bubbles are used for the fine scale part of the problem, which cannot be resolved by the crude finite element mesh. The idea of the sub-grid scale model is summarized as follows [19]:

1) $u = \bar{u} + u'$ (Overlapping sum decomposition).

2) u' is determined analytically on each element.

3) The effect of u' is non-local within each element.

4) The resulting problem for $\bar{u}$ can be solved numerically.

5) The multi-scale interpretation amounts to assuming that irresolvable fine scale behaviour exists within each element, but not on element at boundaries.

It was first indicated in [4], the importance of bubbles in finite element models in terms of enriching the finite element method. In classical linear finite element there are two interpolation functions associated to each element, while in the bubble enriched method, the approximation is made up of two interpolation functions plus an additional component vanishing at element boundaries. This additional component, in general, belongs to a functional space, which is orthogonal to the linear space. With the approximate form of the solution, the residual equation is solved to yield the bubble contribution.

2.6 Multi-scale problems- a general description

Multi-scale phenomena are those in which the field variables show different orders of magnitude in the scale of their variations. Fine scale variations usually demonstrate their affects during the abrupt changes in the behaviour of field unknowns. Examples of such behaviour are presentation of boundary layers in highly porous flows, and shocks that present sharp drops and non-smoothness in the variation of the field unknown within or around domain boundaries.

We are interested to study the convection-diffusion problem (2.15) more deeply. The problem is already presented in section (2.4) and common approaches for its solution are discussed. However, this problem is usually studied in the context of multi-scale phenomena and we are interested to consider the difficulties associated with its numerical solution using classical methods. To this end, we consider the following expression of the boundary value problem:

$$-\varepsilon\frac{d^2u}{dx^2}+\frac{du}{dx}=0 \tag{2.39}$$

$u(0)=1, u(1)=0$, where ε is a positive real number. It is convenient to assume that $\varepsilon \le 1$. The exact solution is:

$$u(x)=1-\frac{exp(\frac{x}{\varepsilon})-1}{exp(\frac{1}{\varepsilon})-1}. \tag{2.40}$$

It can be seen that when ε tends to zero, there is the onset of a boundary layer close to $x=1$. This is highlighted by the following fact: $\lim_{\varepsilon\to 0}\lim_{\substack{x\to 1\\ x<1}} u(x) \neq \lim_{\substack{x\to 1\\ x<1}}\lim_{\varepsilon\to 0} u(x)$.

Let us proceed with a straightforward discretization of (2.39) using finite element method. First rewrite (2.39) in a weak form, assuming that the exact solution $u\in H_0^1(0,1)=\{v\in H^1(0,1): v(0)=1, v(1)=0\}$ satisfies

$$a(u,v) := \varepsilon \int_0^1 \frac{du}{dx}\frac{dv}{dx}dx + \int_0^1 \frac{du}{dx}v dx = 0 \qquad \text{for all } v \in H_0^1(0,1) \tag{2.41}$$

For the domain *(0,1)*, the space $H^k(0,1) \subset L^2(0,1)$ is the set of functions with derivatives up to *k-th* order in $L^2(0,1)$ and $H_0^1(0,1)$ is the space of functions in $H^1(0,1)$ vanishing at the boundary $\{0,1\}$.

Now consider a discretization of the domain *(0, 1)* into finite elements by defining the nodal points $0 = x_0 < x_1 < \ldots < x_{N+1} = 1$, where $x_j = j/(N+1)$ and the mesh parameter $h = \frac{1}{N+1}$. The approximation space $V^h \subset V$ is defined the space of piecewise linear functions:

$V^h = \{ v^h \in V : v^h \textit{is linear in } (x_j, x_{j+1}) \textit{ for } j = 1,\ldots,N+1 \}$. The finite element approximation to u is $u^h \in V^h$ such that

$$a(u^h, v) = 0 \quad \forall v \in V_0^h \tag{2.42}$$

where $V_0^h = \{ v^h \in H_0^1(0,1) : v^h \textit{is piecewise linear} \}$. Note that u^h depends on ε, although this is not explicitly indicated in the notation. Numerical solution of the above problem, is well known to present unrealistic oscillations when ε becomes very small, unless excessive domain discretization is applied. Looking into the error analysis for this problem, gives the idea of where the difficulty arises. The constant that appears in our estimates is denoted by C and is independent of the parameters ε and h.

First we investigate the continuity of the bilinear form *a(.,.)*. In fact, it follows from its definition that

$$a(u,v) \le C \| u \|_{H^1(0,1)} \| v \|_{H^1(0,1)} \qquad \text{for all } u,v \in H_0^1(0,1) \tag{2.43}$$

The problem starts when we try to derive the coercivity estimate:

$$a(v,v)=\varepsilon\int_0^1(\frac{dv}{dx})^2dx+\int_0^1\frac{dv}{dx}vdx=\varepsilon\int_0^1(\frac{dv}{dx})^2dx\geq C\varepsilon\,\|v\|^2_{H^1(0,1)},\ \forall v\in H_0^1(0,1) \quad (2.44)$$

Integration by parts yields:

$$\int_0^1(\frac{dv}{dx})vdx=0 \text{ for } v\in H_0^1(0,1).$$

The Poincare's inequality is used at the last step [14].

Using (2.43), and then (2.44), we gather:

$$\begin{aligned}\|u-u^h\|^2_{H^1(0,1)}&\leq C\varepsilon^{-1}a(u-u^h,u-u^h)=C\varepsilon^{-1}a(u-u^hu-v^h)\\&\leq C\varepsilon^{-1}\|u-u^h\|_{H^1(0,1)}\|u-v^h\|_{H^1(0,1)} \text{ for all } v^h\in V^h\end{aligned} \quad (2.45)$$

Using standard interpolation estimates, we have that I^hu, the interpolator of u, satisfies: $\|u-u^h\|_{H^1(0,1)}\leq h\,|u|_{H^2(0,1)}$.

Making $v^h=I^hu$ in (2.45), we conclude that

$$\|u-u^h\|_{H^1(0,1)}\leq C\varepsilon^{-1}h\,|u|_{H^2(0,1)} \quad (2.46)$$

Let us interpret the error estimate just obtained. First, there is convergence in h. Indeed, for a fixed ε, the error goes to zero as the mesh size goes to zero.

The problem is that the convergence in h is not uniform in ε. Hence, for small ε, unless the mesh size is very small, the H^1 norm error estimate becomes large. The estimate is even worse since $|u|_{H^2(0,1)}=O(\varepsilon^{-3/2})$ which makes (2.46) and the traditional Galerkin method almost useless.

Another way to look at this problem is by first noticing that we would like to have $\lim_{\varepsilon\to0}u^h=\lim_{\varepsilon\to0}u=1$. After all, it would be desirable to have a method that converges (with ε) to the correct solution for a fixed mesh. This is not happening. Indeed, looking at the difference problem coming from (2.41), it can be seen that:

$$-\frac{\varepsilon}{h^2}(u_{j+1}-2u_j+u_{j-1})+\frac{1}{2h}(u_{j+1}-u_{j-1})=0 \qquad u_0=1,u_{N+1}=0 \quad (2.47)$$

where $u_j = u^h(x_j)$. Assume that N is an even number. As ε goes to zero, it follows that $u_{j+1} = u_{j-1}$. This and the boundary conditions originate the oscillatory behaviour of the approximate solution (figure 2.3).

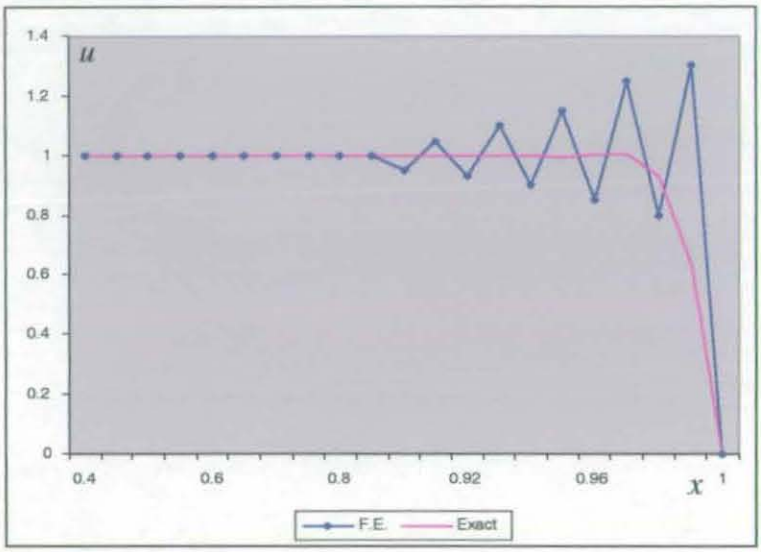


Figure 2.3
Unrealistic oscillation in finite element solution
of the multi-scale problems

Although we used a finite element scheme to derive (2.47), this scheme is also a finite difference scheme that uses a central difference approximation for the convective term $\frac{du}{dx}$. The more naive finite difference approximation:

$$-\frac{\varepsilon}{h^2}(u_{j+1} - 2u_j + u_{j-1}) + \frac{1}{h}(u_j - u_{j-1}) = 0, \quad u_0 = 1, \quad u_{N+1} = 0 \tag{2.48}$$

yields a better result. In fact for this scheme $u_j = u_{j-1}$, as ε goes to zero.

Since $u_0 = 1$, it holds that $u_j = 1$ in the $\varepsilon \to 0$ limit:

$$\lim_{\varepsilon \to 0} u^h(x_j) = \lim_{\varepsilon \to 0} u(x_j) = 1, \qquad \text{for } j = 1, \ldots, N.$$

The behaviour described above is typical in the PDEs where the onset of boundary layers is a common phenomenon. In some cases, as the small parameter goes to zero, and a careless method results in a wrong limit of computed solution.

Several numerical methods try to overcome these and other difficulties related to asymptotic limits. Looking at these difficulties (and their corresponding solutions) it becomes clear that it is important to have a full understanding of the solution's behaviour. This is useful not only to help designing new numerical methods, but also to analyse and estimate old ones.

2.7 Finite element approximations for reaction diffusion equation

In continuation of the above discussions, we consider a similar problem i.e. the so-called problem of reaction-diffusion and present an introductory discussion on how to use finite element techniques to approximate the solution of this problem. Consider the following boundary value problem:

$$\begin{cases} L^{\varepsilon}u := -\varepsilon^2 \Delta u + \sigma u = f \ in \ \Omega \\ u = 0 \quad on \quad \partial\Omega \end{cases} \tag{2.49}$$

where Ω is a two-dimensional bounded domain, ε is a positive constant and σ is a positive constant.

In what follows, we consider a partition of Ω into quadrilateral elements K. Finally, let $P^1(\Omega)$, be the space of continuous functions in Ω that are bilinear polynomials in each quadrilateral, and define $P_0^1(\Omega) = P^1(\Omega) \cap H_0^1(\Omega)$.

The failure of Classical Galerkin approximation is a well-known fact as $\varepsilon << 1$. In the Galerkin formulation, we seek $u^h \in P_0^1(\Omega)$, such that

$$a(u^h, v^h) = (f, v^h) \text{ for all } v^h \in P_0^1(\Omega). \tag{2.50}$$

We are interested in finding a finite element discretization for (2.49) that is stable and coarse mesh accurate for all ε. We use the approach of enriching the finite element space. The idea is to add special functions to the usual polynomial spaces to stabilize

and improve accuracy of the Galerkin method. This goes along with the philosophy of residual free bubble functions [4],[5],[6]. We use a Petrov-Galerkin formulation (i.e., the space of test functions differs from the trial space) and choose the space of test functions as polynomial plus bubbles, but with a different trial space. Consider:

$u^h = P_0^1(\Omega) \oplus E^*(\Omega)$, as the trial space, where $E^*(\Omega)$ is yet to be defined. As the test space, we set $P_0^1(\Omega) \oplus H_0^1(K)$, $K \in T$.

In this Petrov-Galerkin formulation: $u^h = u^1 + u^* \in U^h$, where $u^1 \in P_0^1(\Omega)$ and $u^* \in E^*(\Omega)$, and:

$$a(u^h, v^h) = (f, v^k) \text{ for all } v^h \in P_0^1(\Omega), \tag{2.51}$$

$$a(u^h, v) = (f, v) \text{ for all } v \in H_0^1(K) \text{ and all } K \in T \tag{2.52}$$

From (2.52), we conclude that, for every K,

$$Lu^* = f - Lu^1 \qquad \text{in } K. \tag{2.53}$$

The usual residual free bubble formulation subjects u^* to a homogeneous element boundary condition, i.e., $u^* = 0$ on ∂K, for all elements K. Herein, we replace this condition by a more sophisticated choice [14].

To determine u^* uniquely, we impose the boundary conditions:

$$u^* = 0 \text{ on } \partial K \text{ if } \partial K \in \partial\Omega, L_{\partial K} u^* = R(f - Lu^1) \text{ on } \partial K \text{ if } \partial K \notin \partial\Omega, \tag{2.54}$$

$$u^* = 0 \quad \text{on all vertices of } K \tag{2.55}$$

where R is the trace operator, and we choose

$$L_{\partial K} v = -\varepsilon^2 \partial_{ss} v + \partial v \tag{2.56}$$

where s denotes a variable that runs along ∂K. Note that the restriction of f to K must be regular enough so that its trace on ∂K makes sense. Henceforth, we assume that $f \in P^1(\Omega)$.

The choice of (2.56) is ad hoc, and by no means unique. However, it can be justified under the light of asymptotic analysis [14]. In some sense, the polynomial part of the approximation (u^1 in our case) captures the smooth behaviour of the exact solution. The local multi-scale behaviour is seen by the enrichment functions (u^* in our case), that adds its contribution to the final formulation, without making the method expensive. In other words, it is possible to describe the multi-scale characteristics of a solution for a singular perturbed PDE, without having to resolve with a refined mesh.

We can formally write the solution of (2.53)-(2.56) as

$$u^* = L_*^{-1}(f - L_T u^1) \in L^2(\Omega), \text{ where } \quad L_T = \sum_{K \in T} \chi_K L \tag{2.57}$$

and χ_K is the characteristic function of K. We finally set $E^*(\Omega) = L_*^{-1} P^1(\Omega)$.

Substituting (2.57) in (2.51), we gather that

$$a(I - L_*^{-1} L_T) u^1, v^h) = (f, v^h) - a(L_*^{-1} f, v^h) \quad \text{for all} \quad v^h \in P_0^1(\Omega) \tag{2.58}$$

Finally: $u^h = (I - L_*^{-1} L_T) u^1 + L_*^{-1} f$. Note nevertheless that, because of (2.55), $u^h = u^1$ at the nodal points, as in the usual polynomial Galerkin formulation.

Note that our particular choice of test space allowed the static condensation procedure, i.e., we were able to write u^* with respect to u^1 and f as in (2.57).

The matrix formulation can be obtained as follows. Under the assumption: $f \in P^1(\Omega)$, we write

$$f = \sum_{j \in J} f_j \psi_j \quad , \quad u^1 = \sum_{j \in J_0} u_j^1 \psi_j$$

where J and J_0 are the set of indexes of total and interior nodal points, $\{\psi_j\}_{j \in J}$, form a basis of $P^1(\Omega)$, and $\{\psi_j\}_{j \in J_0}$ form a basis of $P_0^1(\Omega)$. Substituting in (2.57), we have that

$$u^* = \sum_{j \in J} \frac{f_j}{\sigma} L_*^{-1} L_T \psi_j - \sum_{j \in J_0} u_j^1 L_*^{-1} L_T \psi_j \tag{2.59}$$

where we used that

$$L_T \psi_j = \sigma \psi_j \tag{2.60}$$

To write the variational formulation in an explicit form, it is convenient to define:

$$\lambda_j = (\sigma I - L_*^{-1} L_T)\psi_j .$$

Hence, (2.58) reads as

$$\sum_{j \in J_0} a(\lambda_j, \psi_j) u_j^1 = \sum_{j \in J} [(\psi_j, \psi_i) - a(L_*^{-1}\psi_j, \psi_i)] f_j \quad \text{for all } i \in J_0 \tag{2.61}$$

Using the definition of the bilinear form $a(.,.)$, and (2.60), yields

$$\sum_{j \in J_0} a(\lambda_j, \psi_i) u_j^1 = \sum_{j \in J} [(\lambda_j, \psi_i) - \frac{\varepsilon^2}{\sigma}(\nabla \psi_j, \nabla \psi_i) + \frac{\varepsilon^2}{\sigma}(\nabla \lambda_j, \nabla \psi_i)] f_j \forall i \in J_0. \tag{2.62}$$

Concrete computations of the matrix formulation follows. A core and troublesome issue in the present method is solving the local problems. From its definition, λ_j solves:

$$L\lambda_j = 0 \text{ in } K$$

$$L_{\partial K}\lambda_j = 0 \text{ on } \partial K,\ \lambda_j = \begin{cases} 1 \text{ on the jth vertice of T} \\ 0 \text{ on the other vertices of T} \end{cases} \tag{2.63}$$

In fact, we have $L_{\partial K}\lambda_j = L_{\partial K}\psi_j - L\psi_j$ on ∂K. Since we are assuming that ψ_j is bilinear over a rectangular mesh, we have that ψ_j is still linear over ∂K. Hence, $L_{\partial K}\psi_j = \partial \psi_j$. If we take a particular node $I \in J_0$, and look at all elements connected to this node, then the equation (2.53) can be used to illustrate the nodal shape functions λ_I.

Consider now a rectangular straight mesh. Our goal is to find λ_j. Without loss of generality, consider a rectangle K with vertices 1,...,4 at (0,0), $(h_x, 0)$, (h_x, h_y) and $(0, h_y)$ where h_x, h_y are the grid lengths in x and y directions. We have:

$$-\varepsilon^2\Delta\lambda_1+\sigma\lambda_1=0 \text{ in } K \tag{2.64}$$

On the side $y=0$, we have that

$$\begin{cases}-\varepsilon^2\partial_{xx}\lambda_1=0 & \text{for } x\in(0,1)\\ \lambda_1(0,0)=1, & \lambda_1(h_x,0)=0\end{cases}$$

Hence, $\lambda_1(x,0)=\mu_x(x,):=-\dfrac{\sinh(\varepsilon^{-1}\sqrt{\sigma}(x-h_x))}{\sinh(\varepsilon^{-1}\sqrt{\sigma}h_x)}$, similarly we find that:

$$\lambda_1(0,y)=\mu_y(y):=-\frac{\sinh(\varepsilon^{-1}\sqrt{\sigma}(y-h_y))}{\sinh(\varepsilon^{-1}\sqrt{\sigma}h_y)},\ \lambda_1(h_x,y)=\lambda_1(x,h_y)=0.$$

We propose two simple closed form for λ_1, none of which satisfy (2.54)-(2.56) exactly. If we set $\lambda_1(x,y)=\mu_x(x)\mu_y(y)$, then (2.55)-(2.56) holds, but $-\varepsilon^2\Delta\lambda_1+2\sigma\lambda_1=0$ *in* K thus, (2.54) is not satisfied. If we let

$$\lambda_1(x,y)=\frac{sinh(\varepsilon^{-1}\sqrt{\frac{\sigma}{2}}(x-h_x))sinh(\varepsilon^{-1}\sqrt{\frac{\sigma}{2}}(y-h_y))}{sinh(\varepsilon^{-1}\sqrt{\frac{\sigma}{2}}h_x)sinh(\varepsilon^{-1}\sqrt{\frac{\sigma}{2}}h_y)},$$

Then (2.54) holds, but the boundary conditions at $x=0$ and $y=0$ do not hold in this case [14].

2.8 Limitations of residual free bubble functions

In the final part of the previous section, the behaviour of the bubble functions which did not satisfy the original differential equation, is described. This reflects the difficulties associated with the derivation of bubble functions for multi-dimensional problems. In this section, we present an example to prove that such scheme may result in losing all advantages of using bubble enriched finite elements.

Consider the following boundary value problem:

$$\begin{cases} -u'' = f \quad on(0,1) \\ u(0) = u(1) = 0 \end{cases} \tag{2.65}$$

The weak formulation for the equation (2.65) seeks $u \in H_0^1(\Omega)$ such that

$$(u', v') = (f, v), \quad \forall v \in H_0^1(\Omega) \tag{2.66}$$

where $(.,.)$ denotes the integral on $(0,1)$, and $H_0^1(\Omega)$ is the space of functions satisfying (2.65) with square integrable value and derivative on the unit interval.

The standard Galerkin finite element method is obtained by computing with the weak formulation (2.66) on a subspace of $H_0^1(\Omega)$ consisting of continuous functions that are piecewise polynomials on a partition of the unit interval. We define elemental basis functions using fixed reference coordinate ξ on $(-1,1)$. For piecewise linear approximation, the two basis functions on each element are:

$$\psi_1(\varepsilon) = \frac{1}{2}(1-\xi), \psi_2(\xi) = \frac{1}{2}(1+\xi) \tag{2.67}$$

For a quadratic approximation, the basis functions are:

$$\varphi_1(\xi) = \frac{1}{2}\xi(\xi-1), \varphi_2(\xi) = 1-\xi^2, \varphi_3(\xi) = \frac{1}{2}\xi(\xi+1) \tag{2.68}$$

Note that φ_2 is zero on the element boundaries, and is referred to as a bubble function. We can write

$$\varphi_1(\xi) = \psi_1(\xi) - \frac{1}{2}\varphi_2(\xi), \quad \varphi_3(\xi) = \psi_2(\xi) - \frac{1}{2}\varphi_2(\xi) \tag{2.69}$$

The global stiffness matrix of the problem can be obtained from performing the assembly process. For the linear approximation, the elemental stiffness matrix is given by:

$$A_K^1 = \begin{bmatrix} (\psi_1, \psi_1)_K & (\psi_1, \psi_2)_K \\ (\psi_1, \psi_2)_K & (\psi_2, \psi_2)_K \end{bmatrix} \tag{2.70}$$

where the subscript K on the bilinear form $(.,.)$ denotes the range of the integral of $(.,.)$ is on the subinterval K instead of $(0,1)$. The element load vector is:

$$F_K^1 = \begin{bmatrix} (f,\psi_1)_K \\ (f,\psi_2)_K \end{bmatrix} \tag{2.71}$$

with the same convention on subscript K. For the quadratic approximation, we have:

$$A_K^q = \begin{bmatrix} (\varphi_1,\varphi_1)_K & (\varphi_1,\varphi_2)_K & (\varphi_1,\varphi_3)_K \\ (\varphi_1,\varphi_2)_K & (\varphi_2,\varphi_2)_K & (\varphi_2,\varphi_3)_K \\ (\varphi_1,\varphi_3)_K & (\varphi_2,\varphi_3)_K & (\varphi_3,\varphi_3)_K \end{bmatrix} \tag{2.72}$$

and

$$F_K^q = \begin{bmatrix} (f,\varphi_1)_K \\ (f,\varphi_2)_K \\ (f,\varphi_3)_K \end{bmatrix} \tag{2.73}$$

Using static condensation, the unknown value corresponding to the bubble function φ_2 can be eliminated at the element level. More precisely, we can take $v = \varphi_2$ on K and zero elsewhere to write:

$$(u_h,\varphi_2)_K = (f,\varphi_2)_K \tag{2.74}$$

for each element K, where u_h is the solution of the Galerkin method that can be written on an element K as

$$u_h|_K = \sum_{a=1}^{3} \varphi_a u_a \tag{2.75}$$

with u_a, $a = 1,2,3$ denoting the unknown values at the nodes corresponding to each shape function defined on element K [12].

Substituting (2.75) into (2.74) yields:

$$\sum_{a=1}^{3} (\varphi_a,\varphi_2)_K u_a = (f,\varphi_2)_K \tag{2.76}$$

or solving for u_2

$$u_2 = \frac{1}{(\varphi_2,\varphi_2)_K}((f,\varphi_2)_K - (\varphi_2,\varphi_1)_K u_1 - (\varphi_2,\varphi_3)_K u_3) \tag{2.77}$$

Equation (2.77) gives an explicit formula for computing the unknown value associated with the mid-side node, and it clearly holds for any element K in our partition [12]. Simplifying (2.77) using (2.69) gives:

$$(\varphi_2,\varphi_1)_K = (\varphi_2,\psi_1 - \frac{1}{2}\varphi_2)_K = -\frac{1}{2}(\varphi_2,\varphi_2)_K \tag{2.78}$$

and the integration by parts results in:

$$(\varphi_2,\psi_1)_K = (\varphi_2',\psi_1')_K = -(\varphi_2,\psi_1'')_K + (\varphi_2,\psi_1')_{\partial K} = 0 \tag{2.79}$$

Similarly:

$$(\varphi_2,\varphi_3)_K = (\varphi_2,\psi_2 - \frac{1}{2}\varphi_2)_K = -\frac{1}{2}(\varphi_2,\varphi_2)_K \tag{2.80}$$

Therefore:

$$u_2 = \frac{1}{2}(u_1 + u_3) + \frac{(f,\varphi_2)_K}{(\varphi_2,\varphi_2)_K} \tag{2.81}$$

Equation (2.81) shows that if $(f,\varphi_2)_K = 0$, then the bubble unknown coefficient u_2 is the average of the vertex nodes. This occurs when there are point loads applied to vertex nodes, and in this particular circumstances, there is no advantage in using quadratic approximation.

However, even when $(f,\varphi_2)_K \neq 0$, the vertex unknowns are the same as we had computed with linears. Let us show this by further examining the element stiffness matrix and the element load vector. Using (2.69), matrices (2.72)-(2.73) reduce to:

$$A_K^q = \begin{bmatrix} (\psi_1,\psi_1)_K + \frac{1}{4}(\varphi_2,\varphi_2)_K & -\frac{1}{2}(\varphi_2,\varphi_2)_K & (\psi_1,\psi_2)_K + \frac{1}{4}(\varphi_2,\varphi_2)_K \\ -\frac{1}{2}(\varphi_2,\varphi_2)_K & (\varphi_2,\varphi_2)_K & -\frac{1}{2}(\varphi_2,\varphi_2)_K \\ (\psi_1,\psi_2)_K + \frac{1}{4}(\varphi_2,\varphi_2)_K & -\frac{1}{2}(\varphi_2,\varphi_2)_K & (\psi_2,\psi_2)_K + \frac{1}{4}(\varphi_2,\varphi_2)_K \end{bmatrix} \tag{2.82}$$

and

$$F_K^q = \begin{bmatrix} (f,\psi_1)_K - \frac{1}{2}(f,\varphi_2)_K \\ (f,\varphi_2)_K \\ (f,\psi_2)_K - \frac{1}{2}(f,\varphi_2)_K \end{bmatrix} \tag{2.83}$$

If we now take $A_K^q u$, where $u = \{u_1\ u_2\ u_3\}^T$, then the second component is just the left-hand-side of (2.76), which we know by (2.76) equals the second component of F_K^q. The first component of $A_K^q u$ is given by:

$$\begin{aligned}(A_K^q u)_1 = [(\psi_1,\psi_1)_K + \frac{1}{4}(\varphi_2,\varphi_2)_K]u_1 - \frac{1}{2}(\varphi_2,\varphi_2)_K u_2 \\ + [(\psi_1,\psi_2)_K + \frac{1}{4}(\varphi_2,\varphi_2)_K]u_3\end{aligned} \tag{2.84}$$

which using (2.81) reduces to:

$$(A_K^q u)_1 = (\psi_1,\psi_1)_K u_1 + (\psi_1,\psi_2)_K u_3 - \frac{1}{2}(f,\varphi_2)_K \tag{2.85}$$

Similarly,

$$(A_K^q u)_3 = (\psi_2,\psi_1)_K u_1 + (\psi_2,\psi_2)_K u_3 - \frac{1}{2}(f,\varphi_2)_K \tag{2.86}$$

Therefore, neglecting the second row of A_K^q, which has been used to obtain (2.85), (2.86), we can write:

$$(A_K^q u) = \begin{bmatrix} (\psi_1,\psi_1)_K & (\psi_1,\psi_2)_K \\ (\psi_1,\psi_2)_K & (\psi_2,\psi_2)_K \end{bmatrix} \begin{bmatrix} u_1 \\ u_3 \end{bmatrix} - \frac{1}{2}(f,\varphi_2)_K \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{2.87}$$

which is equal to the linear element matrix, equation (2.70), times the vertex unknown minus the contribution to the right-hand-side that is also in (2.83).

Finally, assembling (2.87) and (2.83) to obtain the global equations, results in the same matrix and right-hand-side vector as if we had employed linear functions from

the beginning. Therefore, the vertex unknown on each element u_1 and u_3 will have the same value as if we had used linears, for any function $f(x)$.

An explanation of why there is no advantage in using higher order interpolation for this particular model equation is the well-known super-convergence result that the finite element approximation, interpolates the exact solution at the nodes. We observed that static condensation of the interior node of the quadratic polynomial yields the same matrix problem as the linear approximation. Therefore, the vertex node unknowns of the quadratic approximation give identical values as the node unknowns of the linear approximation, for any source function $f(x)$ [12].

2.9 Application of residual free bubble functions to solid deformation problems

So far, we have observed several applications of the residual free bubble functions in the analysis of the fluid flow problems and considered their advantages, limitations and extent of practicalities. However, it is proved that the method of residual free bubble functions improves the procedure and the degree of accuracy of the solution when applied to solid problems. Here, we give an example of such a problem that is solved using the residual free bubbles and present the method's capabilities in facilitating the solution procedure.

The Timoshenko model describes the deflection of a beam taking into account bending and shear deformations [12]. Standard Galerkin finite element method using equal-order piecewise linear approximations for the unknown dependent variables rotation θ and displacement w yields spurious oscillations.

However, the standard piecewise linears, enriched with residual-free bubbles show that the Galerkin method, without using the tricks of using full integration, produces

reduced integration with a coefficient for the shear. The load term gets some correction as well and the final formulation is nodally exact.

The governing equations of this model are:

$$\begin{cases} -\theta'' - \frac{1}{\varepsilon^2}(w' - \theta) = 0 \\ -\frac{1}{\varepsilon^2}(w'' - \theta') = f \end{cases} \quad in\ \Omega \tag{2.89}$$

where $\Omega = (0,1)$, θ and w are the rotation and displacement variables, f is the load and ε is a non-dimensional parameter proportional to the beam thickness.

To (2.89) we append the following boundary conditions, without loss of generality:

$$\begin{cases} w(0) = w(1) = 0 \\ \theta(0) = \theta(1) = 0 \end{cases} \tag{2.90}$$

The variational formulation, corresponding to the above equation and its boundary condition is to: Find $\{\theta, w\} \in H_0^1(\Omega)^2$ such that

$$(\theta', \psi') + \frac{1}{\varepsilon^2}(w' - \theta, v' - \psi) = (f, v), \quad \forall \{\psi, v\} \in H_0^1(\Omega)^2 \tag{2.91}$$

where $H_0^1(\Omega)$ is the space of functions with square-integrable value and derivative in Ω satisfying (2.90) and we use the notation $(f, g) = \int_\Omega fg d\Omega$.

Consider a partition of Ω into non-overlapping elements in the usual way. Also assume that the exact solution of our problem can be decomposed into

$$\begin{cases} \theta = \theta_l + \theta_b \\ w = w_l + w_b \end{cases} \tag{2.92}$$

where θ_l and w_l are spanned by the standard continuous piecewise linears of finite element method, and θ_b and w_b are assumed to satisfy the following differential equations in each element K:

$$\begin{cases} -\theta_b'' - \frac{1}{\varepsilon^2}(w_b' - \theta_b) = -(-\theta_1'' - \frac{1}{\varepsilon^2}(w_1' - \theta_1)) \\ -\frac{1}{\varepsilon^2}(w_b'' - \theta_b') = -(-\frac{1}{\varepsilon^2}(w_1'' - \theta_1') - f) \end{cases} \tag{2.93}$$

subject to the boundary conditions:

$$\theta_b = w_b = 0 \quad on\ \partial K. \tag{2.94}$$

Note that $\theta_1''' = w_1'' = 0 \quad in\ K$, therefore equations (2.94) can be written as

$$\begin{cases} -\varepsilon^2\theta_b'' + \theta_b - w_b' = w_1' - \theta_1 \\ \theta_b' - w_b'' = -\theta_1' + \varepsilon^2 f \end{cases}. \tag{2.95}$$

From (2.94) we get $\theta_b - w_b' = w_1' - \theta_1 + \varepsilon^2\theta_b''$ and combining with (2.95) yields:

$$\theta_b''' = f \quad in\ K \tag{2.96}$$

If we integrate three times with respect to the local variable in the element (i.e. $\xi \in [0, h_k], h_k = x_{i+1} - x_i, \xi = x - x_i$) and assume piecewise constant load f, and for notation's sake drop the subscripts of h and f (nowhere we need to assume that h_k is constant in what follows) we obtain

$$\theta_b(\xi) = \frac{\xi^3}{6} f + c_1 \frac{\xi^2}{2} + c_2\xi + c_3. \tag{2.97}$$

Applying the boundary conditions $\theta_b(0) = \theta_b(h) = 0$ above gives:

$$\theta_b(\xi) = \frac{\xi}{6} f(\xi^2 - h^2) + c_1 \frac{\xi}{2}(\xi - h) \tag{2.98}$$

Using this expression into the first equation of (2.95) after one integration, we get

$$\begin{aligned} w_b(\xi) &= \int_0^\xi \theta_1(t)dt - w_1(\xi) - \varepsilon^2[\frac{f}{6}(3\xi^2 - h^2) + \frac{c_1}{2}(2\xi - h)] \\ &+ \frac{f}{6}[\frac{\xi^4}{4} - \frac{\xi^2}{2}h^2] - \frac{c_1}{12}\xi^2(3h - 2\xi) + c_4 \end{aligned} \tag{2.99}$$

Applying the boundary conditions $w_b(0) = w_b(h) = 0$ in (2.99), we get expressions for the remaining constants c_1 and c_4. The expressions for the residual-free bubble functions are given by:

$$\theta_b(\xi) = f\{\frac{\xi}{6}(\xi^2 - h^2) + \frac{h\xi}{4}(h - \xi)\} + \frac{1}{\varepsilon^2 + (\frac{h^2}{12})}\frac{\xi(\xi - h)}{2}\{\theta_1(\frac{h}{2}) - \frac{w_1(h) - w_1(0)}{h}\} \tag{2.100}$$

$$w_b(\xi) = \xi(1 - \frac{\xi}{2h})\theta_1(0) + \frac{\xi^2}{2h}\theta_1(h) + \frac{\xi}{h}[w_1(0) - w_1(h)] - \xi[\varepsilon^2 - \frac{\xi^2}{6} + \frac{h\xi}{4}]\{\frac{1}{\varepsilon^2 + (\frac{h^2}{12})}[\theta_1(\frac{h}{2}) - \frac{w_1(h) - w_1(0)}{h}] - \frac{hf}{2} + \frac{f\xi^2}{2}[-\varepsilon^2 + \frac{\varepsilon^2}{12} - \frac{h^2}{6}] \tag{2.101}$$

If we take the test functions $\psi = \psi_1$ and $v = v_1$, where ψ_1 and v_1 are spanned by continuous piecewise linears, then using decomposition (2.92) the variational formulation (2.91) can be rewritten as

$$(\theta_1', \psi_1') + \frac{1}{\varepsilon^2}(w_1' - \theta_1, v_1' - \psi_1) - (f, v_1) + \frac{1}{\varepsilon^2}(w_b' - \theta_b, v_1' - \psi_1) = 0 \tag{2.102}$$

where, by integration by parts, we used that:

$$(\theta_b', \psi_1') = \sum_K (\theta_b', \psi_1')_K = \sum_K [(\theta_b, \psi_1')_{\partial K} - (\theta_b, \psi_1'')_K] = 0$$

Note that (2.102) consists of the Galerkin method for equal-order piecewise linear approximation for θ and w (without tricks, using full integration) plus a 'perturbation term' that we need to compute based on the bubble functions given by (2.100) and (2.101). First, by (2.100) and (2.101) we compute:

$$w_b' - \theta_b = \theta_1(0) + \frac{\xi}{h}[\theta_1(h) - \theta_1(0)] - \frac{w_1(h) - w_1(0)}{h}$$
$$- \frac{\varepsilon^2}{\varepsilon^2 + (\frac{h^2}{12})}[\theta_1(\frac{h}{2}) - \frac{w_1(h) - w_1(0)}{h}] + \varepsilon^2 f(\frac{h}{2} - \xi) \tag{2.103}$$

Note also that

$$w_1' - \theta_1 = \frac{w_1(h) - w_1(0)}{h} - (1 - \frac{\xi}{h})\theta_1(0) - \frac{\xi}{h}\theta_1(h) \tag{2.104}$$

Thus, assuming (2.103) to (2.104)

$$w_1' - \theta_1 + w_b' - \theta_b = \varepsilon^2 f(\frac{h}{2} - \xi) - \frac{\varepsilon^2}{\varepsilon^2 + (\frac{h^2}{12})}[\theta_1(\frac{h}{2}) - \frac{w_1(h) - w_1(0)}{h}] \tag{2.105}$$

Therefore, using (2.105), the variational formulation given by (2.102) reduces to

$$(\theta_1', \psi_1') + \sum_K \frac{1}{\varepsilon^2 + (\frac{h_K^2}{12})}(\frac{w_1(h_k) - w_1(0)}{h_k} - \theta_1(\frac{h_k}{2}), v_1' - \psi_1)_K$$
$$= (f, v_1) + \sum_K f_K(\xi - \frac{h_k}{2}, v_1' - \psi_1)_K \tag{2.106}$$

where we reintroduced the subscripts for h and the piecewise constant load f. This can also be rewritten as

$$(\theta_1', \psi_1') + \sum_K \frac{1}{\varepsilon^2 + (\frac{h_k^2}{12})}(w_1' - R\theta_1, v_1' - \psi_1)_k = (f, v_1) + \sum_K f_k(\xi - \frac{h_k}{2}, v_1' - \psi_1)_K \tag{2.107}$$

where R stands for a reduced integration operator. Formulation (2.107) was derived using full integration throughout and by construction, its solution is nodally exact. The final form is identical to application of the following ideas to the standard variational formulation: to use one-point reduced integration on the shear term, to replace its coefficient $\frac{1}{\varepsilon^2}$ by $\frac{1}{(\varepsilon^2 + \frac{h_k^2}{12})}$ in each element and to correct the right-hand side as in equation (2.107) for piecewise-constant loads.

To emerge with these collections of ideas requires ingenuity and for the first two ideas, some authors have given different arguments [12]. We wish to point out that the residual-free bubble provides us with a systematic approach to construct discretization procedures that may possibly improve the existing schemes.

From what presented in this chapter, we can see that the multi-scale variational approach presents a reliable substance for the study of multi-scale problems. As far as the manual solution of ordinary or partial differential equations is concerned, these techniques are stable and sufficiently accurate. However, the practical implementation of these methods is not a trivial matter. Derivation of the residual free bubble functions may result in the analytical solution of a differential equation, which becomes a cumbersome procedure in multi-dimensional problems. Solution of the analogous ordinary problem is not always extendable to higher dimensions. As a result, it remains an important issue to implement the idea of the bubble function within a practical scheme in order to automate the method and make it case-independent to some extent. We will pursue this ambition in the following Chapters.

Chapter 3

A novel method for the derivation of bubble functions for the finite element solution of two-point boundary value problems

In this chapter, first the classical Galerkin approximation and the residual free bubble methods are introduced within the context of variational formulation of differential equations. The general second-order boundary value problem with scalar coefficients is considered and its associated practical bubbles are derived using the method of least squares minimization. This method is regarded to be an alternative to the previously available techniques. The advantages of this technique over the existing procedures are discussed. Benchmark problems are solved using the quadratic polynomial bubble enriched finite elements and the results are tested against standard finite element and exact solutions.

3.1 Classical Galerkin approximation and residual free bubbles

Consider the following boundary value problem:

$$\begin{cases} Lu = f, & in\ \Omega \\ B.C. & on\ \partial\Omega \end{cases} \tag{3.1}$$

where Ω is a domain in R^n (initially assumed to be an interval in R) with a boundary of $\partial\Omega$, L is a linear differential operator, u is an unknown scalar or vector valued function and f is a given source function. We assume that L is such that the problem is well posed (i.e. it has a unique solution and that continuously depends on the initial values).

The weak formulation of system (3.1) requires the solution to satisfy the variational formulation of the above equation. This involves the multiplying of the residual by a set of test functions and integration of the resulting functions by parts (Green's integral theorem). To develop a classical Galerkin finite element scheme for equation (3.1) we assume a partition of Ω into elements Ω_e such that neither overlapping of elements nor gaps between them occurs. The approximation space V_h is chosen from a finite dimensional space related to the partition that satisfies $V_h \subset V$, where V_h is the space of functions in which we seek a solution to the continuous variational problem. We set h=max {diam (Ω_e)} as the partition diameter.

The Galerkin method states that: Find $u \in V_h$ such that

$$\langle Lu, v\rangle = \langle f, v\rangle \quad \forall v \in V_h \tag{3.2}$$

where $\langle .,. \rangle$ is a bilinear form of the variational problem (3.1). To enrich the standard Galerkin scheme by the use of bubble functions each $u \in V_h$ is taken as the sum of standard piecewise linear parts and bubble functions. To specify this we write $u = u_1 + u_b \ \forall u \in V_h$, where u_1 and u_b are the linear and the bubble components, respectively. We require the bubble function to vanish on the element boundaries, a well-established method towards the validity of the incorporation of a bubble function in finite element scheme. For the residual free bubble, we seek the bubble component to strongly satisfy the residual equation within each element, i.e.:

$$\begin{cases} Lu_b = -Lu_1 + f & in\ \Omega_e \\ u_b = 0 & on\ \partial\Omega_e \end{cases} \tag{3.3}$$

The vanishing of bubbles on each element boundary allows the use of static condensation [27] that makes possible the following selection of

$$v = \begin{cases} v_b \ on \ \Omega_e \\ 0 \ \ else \end{cases} \text{ in (3.2).}$$

Therefore, after solving equation (3.3) exactly or approximately we have:

$$\langle Lu_1 + Lu_b, v_1 \rangle + \langle Lu_1 + Lu_b, v_b \rangle = \langle f, v_1 \rangle + \langle f, v_b \rangle \quad v_1 + v_b \in V_h \ on \ \Omega_e \tag{3.4}$$

Therefore, it is sufficient to find $u = u_1 + u_b$ such that:

$$\langle Lu_1 + Lu_b, v_1 \rangle = \langle f, v_1 \rangle \quad \forall v_1 \in V_h \ on \ \Omega_e \tag{3.5}$$

The approximate solution of problem (3.1) can be written as:

$$u = u_1 + u_b = \sum_{i=1}^{n} u_i(\phi_i + \psi_i) + \phi_f \tag{3.6}$$

where ϕ and ψ are bubble and linear shape functions, respectively and n is the number of nodes per element. At this point, application of the finite element procedure to obtain the unknown nodal values u_i will result in the solution of the problem.

3.2 Polynomial bubble functions

The residual free bubble method, mentioned in previous section, offers a practical approach towards the solution of multi-scale problems in which both fine and coarse scale variations need to be taken into account. In order for the numerical methods to capture fine scale variation of a multi-scale problem, one requires to excessively refine the discretization of the problem domain. This generally requires the size of elements that are smaller than the thickness of boundary layers and results in a huge number of equations to be solved. The main idea behind the described residual free method is to solve the residual equation (3.3) subject to homogeneous boundary conditions. This, however, can be as difficult as solving the original differential equation especially in two and three-dimensional problems. Another difficulty is that

the typical complex form of residual free bubbles cannot be directly used in the evaluation of elemental integrals in finite element procedures. Therefore, in order to offset this loss of flexibility, it is desirable to convert them into simpler approximate forms to make it possible to use quadrature methods in a finite element program. In [27] and [28] the polynomial approximation of bubble functions is introduced and satisfactorily accurate and stable results are obtained for a multi-scale convection-diffusion problem. In this scheme Taylor series approximation of bubble functions within each element is carried out and the higher dimensional generalization is derived based on the analytical solution of analogous ODE, Taylor series approximation of the residual free bubble and tensor product of one-dimensional bubble functions. In what follows, alternative approaches to the residual free method are discussed and a novel least squares based method for the generation of bubbles is presented. It is worth however, before proceeding any further, to compare the least squares fit with series approximation of a given function. Application of the method of power series is justified, while the solution of differential equation is assumed sufficiently smooth over the element domain. This, however, cannot be known a priory, in many problems. It is preferred sometimes to work with a practical simple approximation of the solution rather than working with complicated solutions expressed in terms of special or sophisticated functions. Polynomials are among the best options to be used as bubble functions under such conditions. They can be easily integrated or differentiated and can be expressed as orthogonal families with respect to suitable norms and weighting functions. Every continuous function, over a closed interval can be expressed in terms of polynomials, which is a significant property in the development of approximate finite element schemes for differential equations.

Power series expansion of a function $y(x)=\sum_{n=0}^{\infty}\frac{y^n(0)x^n}{n!}$ when truncated to

$y_N(x)=\sum_{n=0}^{N}\frac{y^n(0)x^n}{n!}$ always gives a good approximation near $x= 0$, but as x increases (or decreases), the approximation tends to get worse. This is also the case in general point-wise approximations. The least squares fit, on the other hand, results in an almost uniform approximation for an interval, depending only on the weight function $w(x)$. Another fundamental issue is that power series fit at a point whereas the least squares method generates fits over an interval. Thus, the least square method provides the weighted average of the unknowns within the range of any scattered sample of data. In numerical computations, it is not generally possible to accurately estimate the derivatives at a point from the samples scattered in an interval. Therefore, exact matching interpolating polynomials or least squares polynomials are frequently used instead of truncated power series. However, one drawback of the least squares fitting is that the initial values of data points outside of a range could deviate the model from correct predictions. This problem should be relaxed by methods such that the experimental data used for generating the points are accurate. In practical modelling, those models that rely on lower number of parameters are more desired and the model calibrations are more straightforward. In such situations, the method of least squares has mainly the potential for dealing with either continuous variables or large samples of discrete variables.

3.3 Derivation of residual free bubble: A convection diffusion problem

In this section, a scalar convection diffusion ODE is studied and the corresponding residual free bubble is derived. The convection diffusion model in one dimension is formulated by the following differential equation:

$$\frac{d^2u}{dx^2} - k(x)\frac{du}{dx} = f(x) \quad in\ \Omega \tag{3.7}$$

subject to essential boundary conditions, where $x \in \Omega$ and Ω is the open interval *(0, 1)*. Assume k a constant, the corresponding equation is:

$$\frac{d^2u}{dx^2} - k\frac{du}{dx} = f(x) \quad in\ \Omega \tag{3.8}$$

Consider a discretization of Ω into finite elements and let the domain and boundary of element be Ω_e , $\partial\Omega_e$ respectively. Corresponding to the case $f = 0$, the suggested solution of the problem is written as:

$$u = u_1 + u_b \tag{3.9}$$

where u_1 is the standard continuous piecewise linear finite element approximation of u , and u_b is the bubble part, satisfying strongly the elemental residual differential equation (note that $\frac{d^2u_1}{dx^2} = 0\ in\ \Omega_e$):

$$\begin{cases} \dfrac{d^2u_b}{dx^2} - k\dfrac{du_b}{dx} = k\dfrac{du_1}{dx} & in\ \Omega_e \\ u_b = 0\ on\ \partial\Omega_e \end{cases} \tag{3.10}$$

Consider a formal power series representation of the trial solution of equation (3.10):

$$u_b(\eta) = \sum_{n=0}^{\infty} a_n \eta^n \tag{3.11}$$

with respect to the local element variable: $\eta \in [0, l_e], l_e = x_{i+1} - x_i, \eta = x - x_i$, (denoting $\frac{du_1}{d\eta} = \frac{u_1(l_e) - u_1(0)}{h_e} = \lambda_e\ in\ \Omega_e$). We find:

$$\sum_{n=2}^{\infty} n(n-1)a_n\eta^{n-2} - \sum_{n=1}^{\infty} kna_n\eta^{n-1} \equiv k\lambda_e \tag{3.12}$$

and:

$$\sum_{n=0}^{\infty}(n+2)(n+1)a_{n+2}\eta^n-\sum_{n=0}^{\infty}k(n+1)a_{n+1}\eta^n\equiv k\lambda_e \tag{3.13}$$

with the boundary condition, $u_b(0)=0$ we obtain the recursive formula:

$$a_0=0, a_2=\frac{k\lambda_e+ka_1}{2}, a_3=\frac{ka_2}{3}, \ldots, a_{n+1}=\frac{ka_n}{n+1}=\frac{k^n(a_1+\lambda_e)}{(n+1)!}. \tag{3.14}$$

using $\phi_b(l_e)=0$ we get

$$a_1=\lambda_e\left(\left(\frac{kl_e}{e^{kl_e}-1}\right)-1\right) \tag{3.15}$$

Therefore:

$$\begin{aligned} u_b(\eta) &= a_1\eta+\frac{1}{k}\sum_{n=2}^{\infty}\frac{k^n\eta^n}{n!}(a_1+\lambda_e) \\ &= \left(\frac{l_e}{e^{kl_e}-1}\right)\left(k\eta+\frac{\eta}{l_e}-\frac{\eta e^{kl_e}}{l_e}+\sum_{n=2}^{\infty}\frac{k^n\eta^n}{n!}\right)\left(\frac{u(l_e)-u(0)}{l_e}\right) \end{aligned} \tag{3.16}$$

The uniform convergence of the series is guaranteed everywhere in Ω_e thanks to the nature of the problem and the ratio test applied to (3.14). Truncation of the formal power series in (3.16) provides an approximate solution of (3.8):

$$\begin{aligned} \bar{u}(\eta) &= \frac{\eta-l_e}{l_e}u(0)+\frac{\eta}{l_e}u(l_e) \\ &+\left(\frac{l_e}{e^{kl_e}-1}\right)\left(k\eta+\frac{\eta}{l_e}-\frac{\eta e^{kl_e}}{l_e}+\sum_{n=2}^{N}\frac{k^n\eta^n}{n!}\right)\left(\frac{u(l_e)-u(0)}{l_e}\right) \end{aligned} \tag{3.17}$$

By direct solution of (3.10), the following closed form expression for the residual free bubble function is derived:

$$u_b(\eta)=\frac{\lambda_e l_e}{1-e^{kl_e}}(1-e^{k\eta})-\lambda_e\eta. \tag{3.17}$$

This, in fact, is equivalent to the power series (3.16).

3.4 Least squares approximation used to generate residual free bubble functions

In this section, a novel finite element method for the solution of previously described boundary value problems is presented by the use of polynomial approximation of bubble functions obtained from the method of least squares minimization. Here, the method is introduced and comprehensively described within the content of an example and the general cases are investigated in later sections of the thesis.

Consider the following two-point boundary value problem:

$$\begin{cases} u'' + u = 0, \quad on\ \Omega = [0,2] \\ u(0) = 0, u(2) = 1. \end{cases} \tag{3.18}$$

The exact solution of the above problem is $u(x) = \dfrac{sin(x)}{sin(2)}$. Considering a numerical solution of equation (3.18), we study a discretization of the interval [0, 2] into N sub-domains each of the length $l_j = x_{j+1} - x_j, where \quad 0 = x_0 < x_1 < ... < x_N = 1, j = 1,...,N$. Let us assume that the local approximation of u in the *j-th* subinterval $I_j = [x_j, x_{j+1}]$ is of the closed form:

$$\tilde{u}(x) = \frac{x - x_j}{l_j} u_{j+1} + \frac{x_{j+1} - x}{l_j} u_j + c_j x(l_j - x) \tag{3.19}$$

in which $B_{2,j}(x) = c_j x(l_j - x)$ is the quadratic polynomial bubble of the element I_j added to the classical Lagrange element. We agree to extend $B_{2,j}$ to zero outside the element I_j. The bubble coefficient c_j shall be determined in a way to reduce the approximation error without the need to incur mesh refinements. The following polynomial approximation can be readily applied to the polynomial bubbles of as higher order as desired, with the assumption that the bubble part vanishes at element boundaries.

To start, substitute the approximate form (3.19) into the equation (3.18) in the local interval I_j to get the residual R. For the sake of simplicity, adopt the standard interval $[0, l]$ and the associated shape functions. Hence (3.19) is written as:

$$\tilde{u}_j(x) = \frac{l-x}{l}u(0) + \frac{x}{l}u(l) + c_j x(l-x) \tag{3.20}$$

Setting $u_0 = u_j(0), u_l = u_j(l)$ and replacing (3.20) into equation (3.18) one has:

$$R_B = -2c_j + (\frac{l-x}{l})u_0 + (\frac{x}{l})u_l + c_j x(l-x) \tag{3.21}$$

In order to find the optimal value of c_j, we introduce the functional:

$$J_B = \int_0^l R_B{}^2 dx \tag{3.22}$$

Minimizing J_B with respect to c_j using the least square technique yields:

$$\begin{aligned} 0 = \frac{\partial J_B}{\partial c_j} &= \int_0^l 2R_B \frac{\partial R_B}{\partial c_j} dx \\ &= \int_0^l \{-2 + x(l-x)\}\{-2c_j + (\frac{l-x}{l})u_0 + (\frac{x}{l})u_l + c_j x(l-x)\}dx \end{aligned} \tag{3.23}$$

Having $c_j = -\frac{5}{2}(\frac{l^2-12}{20+(l^2-10)^2})(u_0+u_l)$, the approximate solution is:

$$\tilde{u}_j(x) = \frac{l-x}{l}u_0 + \frac{x}{l}u_l - \frac{5}{2}(\frac{l^2-12}{20+(l^2-10)^2})(u_0+u_l)(x(l-x)) \tag{3.24}$$

Setting $a = -\frac{5}{2}(\frac{l^2-12}{20+(l^2-10)^2})$ provides:

$$\tilde{u}_j(x) = \{\frac{l-x}{l} + ax(l-x)\}u_0 + \{\frac{x}{l} + ax(l-x)\}u_l \tag{3.25}$$

The above derivation is valid as it results in a unique minimum value due to the differentiability, non-negativity and the parabolic structure of the functional J. In general, where there are more than only one bubble coefficient to be determined,

equation (3.23) will result in a system of equations and solving the system will provide the bubble coefficients. This case will be discussed in following sections.

Approximation (3.25) is a quadratic approximation, whereas using the static condensation we only need to determine the nodal values of $u_0 = u(0), u_l = u(l)$ in the finite element scheme. Another significant advantage of the least squares derivation of approximate bubble function is that the analytical solution of residual equation (3.3) is avoided which is crucial in the automation of the approximation process by a computer code.

In order to determine the nodal values u_0, u_l the weighted residual procedure will be applied to approximation (3.25).

Rewriting (3.25), in terms of shape functions, within the space of finite element gives: $\tilde{u}_j(x) = N_0 u_0 + N_1 u_l$ where $N_0(x) = \frac{l-x}{l} + ax(l-x), N_1(x) = \frac{x}{l} + ax(l-x)$.

The weighted residual statement (weak formulation) with respect to weight function $w(x)$ is:

$$0 = \int_0^l w(x)(\tilde{u}_j''(x) + \tilde{u}_j(x))dx = \int_0^l (-w'(x)\tilde{u}_j'(x) + w(x)\tilde{u}_j(x))dx + \{w(x)\tilde{u}_j'(x)\}\Big|_0^l .$$

Similar to the standard Galerkin method we make the following selection of weight functions $w_0(x) = N_0(x) = \frac{l-x}{l} + ax(l-x), w_1(x) = N_1(x) = \frac{x}{l} + ax(l-x)$.

Therefore, corresponding weighted residual statements are:

$$\begin{cases} \int_0^l (N_I(N_I u_I + N_{II} u_{II}) - N_I'(N_I' u_I + N_{II}' u_{II}) = -\{N_I \tilde{u}_j'(x)\}\Big|_0^l \\ \int_0^l (N_{II}(N_I u_I + N_{II} u_{II}) - N_{II}'(N_I' u_I + N_{II}' u_{II}) = -\{N_{II} \tilde{u}_j'(x)\}\Big|_0^l . \end{cases} \quad (3.26)$$

The matrix representation of the above system of equation is:

$$\begin{bmatrix} -\int_0^l (N_I' N_I' - N_I N_I) & -\int_0^l (N_{II}' N_I' - N_{II} N_I) \\ -\int_0^l (N_{II}' N_I' - N_{II} N_I) & -\int_0^l (N_{II}' N_{II}' - N_{II} N_{II}) \end{bmatrix} \begin{bmatrix} u_I \\ \\ u_{II} \end{bmatrix} = \begin{bmatrix} -N_I \phi|_0^l \\ \\ -N_{II} \phi|_0^l \end{bmatrix}$$

(3.27)

where ϕ is the boundary line term. Substitution of shape functions into the above matrix system and evaluation of integrals and boundary line terms give:

$$\begin{bmatrix} (\frac{l}{3}-\frac{1}{l})+(\frac{a}{6}-\frac{a^2}{3})l^3+\frac{a^2}{30}l^5 & (\frac{l}{6}+\frac{1}{l})+(\frac{a}{6}-\frac{a^2}{3})l^3+\frac{a^2}{30}l^5 \\ (\frac{l}{6}+\frac{1}{l})+(\frac{a}{6}-\frac{a^2}{3})l^3+\frac{a^2}{30}l^5 & (\frac{l}{3}-\frac{1}{l})+(\frac{a}{6}-\frac{a^2}{3})l^3+\frac{a^2}{30}l^5 \end{bmatrix} \begin{bmatrix} u_I \\ \\ u_{II} \end{bmatrix} = \begin{bmatrix} q_I \\ \\ -q_{II} \end{bmatrix} \quad (3.28)$$

equivalent to the decomposed matrix form:

$$\begin{bmatrix} (\frac{l}{3}-\frac{1}{l}) & (\frac{l}{6}+\frac{1}{l}) \\ (\frac{l}{6}+\frac{1}{l}) & (\frac{l}{3}-\frac{1}{l}) \end{bmatrix} \begin{bmatrix} u_I \\ \\ u_{II} \end{bmatrix} + \{(\frac{a}{6}-\frac{a^2}{3})l^3+\frac{a^2}{30}l^5\} \begin{bmatrix} 1 & 1 \\ \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u_I \\ \\ u_{II} \end{bmatrix} = \begin{bmatrix} q_I \\ \\ -q_{II} \end{bmatrix} \quad (3.29)$$

where $a = -\frac{5}{2}(\frac{l^2-12}{20+(l^2-10)^2})$. This illustrates the fact that the bubble part belongs to the quadratic space, orthogonal to the space of standard linear elements.

Returning to equation (3.18) and setting the refinement of *[0,2]* into divisions with equal size of l =*0.2*, one gets $a = 0.25083$ and $(\frac{a}{6}-\frac{a^2}{3})l^3+\frac{a^2}{30}l^5 = 0.0001673$ and the general local stiffness matrix (3.28) is evaluated as:

$$\begin{bmatrix} -4.93316 & 5.03350 \\ 5.03350 & -4.93316 \end{bmatrix} \quad (3.29)$$

Therefore, the global stiffness matrix obtained from assembling local matrices is

$$\begin{bmatrix} -4.9331 & 5.0335 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5.0335 & -9.8663 & 5.0335 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5.0335 & -9.8663 & 5.0335 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5.0335 & -9.8663 & 5.0335 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.0335 & -9.8663 & 5.0335 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5.0335 & -9.8663 & 5.0335 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5.0335 & -9.8663 & 5.0335 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5.0335 & -9.8663 & 5.0335 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.0335 & -9.8663 & 5.0335 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.0335 & -9.8663 & 5.0335 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.0335 & -4.9331 \end{bmatrix}$$

At this point, imposition of boundary conditions i.e. $u(0)=0, u(2)=1$ and elimination of redundant equations corresponding to the boundary values will result into solution of system and the nodal values. Nodal values of the quadratic bubble enriched method, the analytical solution and the 10-point linear finite element results are listed in the following table for comparison:

x	10-Point FE	Bubble Enriched	Exact
0	0	0	0
0.2	0.2178	0.218491	0.218487
0.4	0.4269	0.428263	0.428263
0.6	0.6191	0.620948	0.620966
0.8	0.7867	0.788866	0.788912
1	0.9230	0.925326	0.925408
1.2	1.0227	1.0249	1.02501
1.4	1.0817	1.083629	1.083749
1.6	1.0977	1.099178	1.099281
1.8	1.0709	1.0709291	1.070989
2	1	1	1

Table 3.1
Comparison of standard, bubble enriched and exact solutions

The quadratic bubble enriched finite element method provides relatively better results as compared to the 10-point standard Lagrange elements. The approximation is more accurate not only within the elements, but also at nodal points. As soon as the nodal values are found, the approximation (3.25) can be used to determine the intermediate elemental values:

$$\begin{aligned}\bar{u}_j(x) &= \{1-5x+0.25083x(0.2-x)\}u_0 + \{5x+0.25083x(0.2-x)\}u_i \\ &= \{1-4.94983x-0.25083x^2\}u_0 + \{5.05016x-0.25083x^2\}u_i\end{aligned} \tag{3.30}$$

The following diagrams show the performance of the linear finite element and the quadratic practical bubble solution of the above problem against the exact solution.

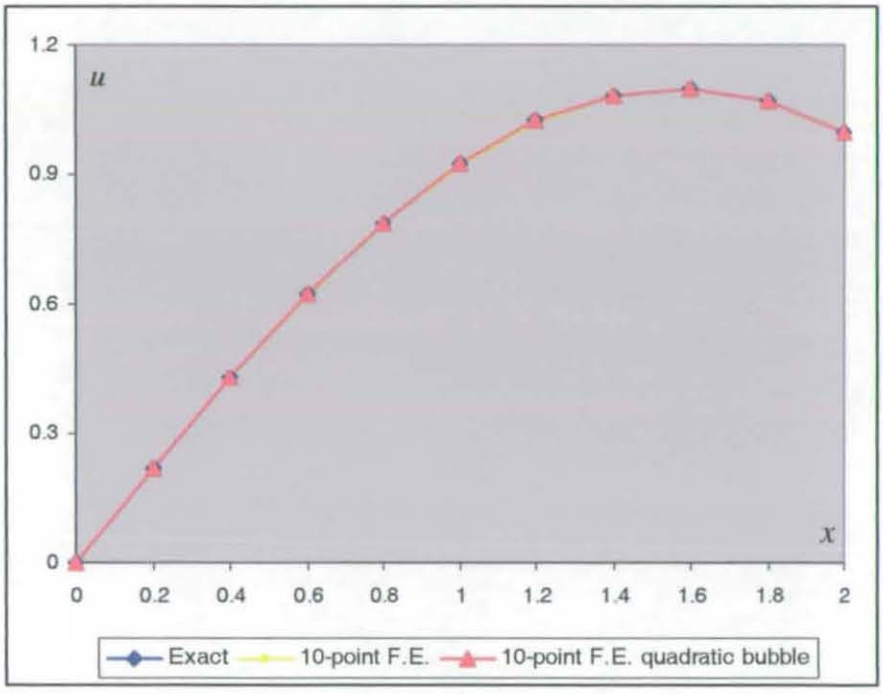


Figure 3.1
10-point FE, 10-point bubble enriched FE, Exact solution
Equi-distant nodes

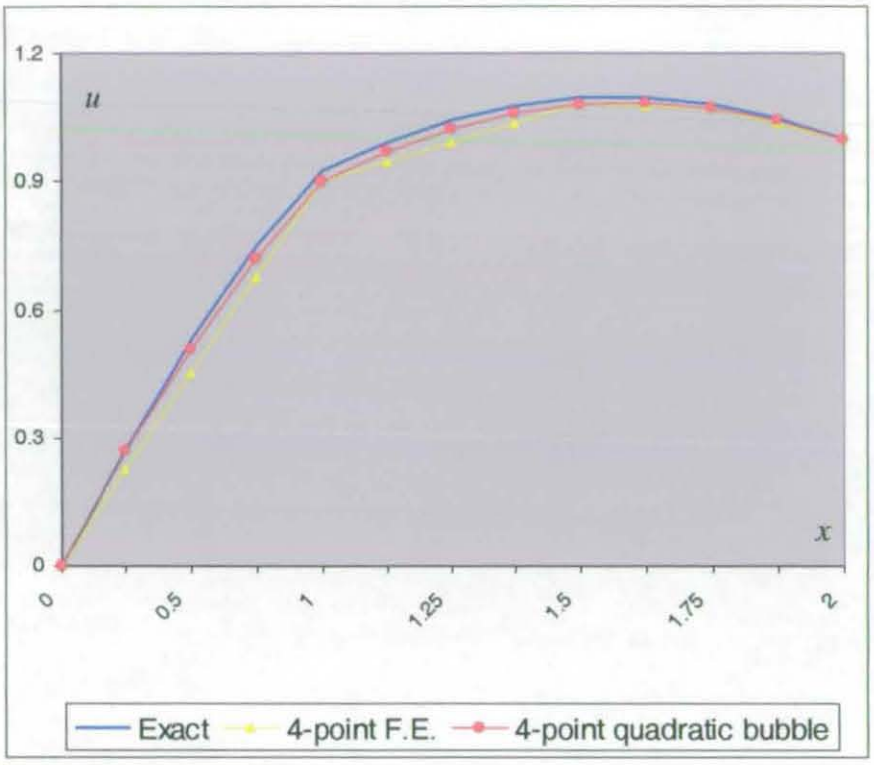


Figure 3.2
4-point FE, 4-point bubble enriched FE, Exact solution
Uneven mesh with node at x=0, 1, 1.25, 1.75, 2

3.5 The use of the least squares method to develop a practical scheme for bubble function generation: general case

We are now in a position to extend our method to more general case of two-point-scalar boundary value problems, represented by ODEs of second order. The following results are also applicable to the weakly nonlinear case and some types of nonlinear problems. However, due to the existence of nonlinearities, such cases require individual treatment of each problem, which is beyond the scope of the present thesis. Therefore, the following discussion only considers a scalar two-point boundary value problem given as:

$$\begin{cases} \varepsilon u'' + \kappa u' + \lambda u = f, \quad on\ \Omega = [a,b] \\ u(a) = \alpha, u(b) = \beta. \end{cases} \tag{3.31}$$

Assume, for the sake of simplicity, that $f=0$. Once again, the problem domain is discretized into N sub-intervals $a = x_0 < x_1 < ... < x_N = b$, $j = 1,...,N$ and the elemental quadratic approximation of the solution of (3.31) is taken to be $\tilde{u}(x) = \frac{x - x_j}{x_{j+1} - x_j} u_{j+1} + \frac{x_{j+1} - x}{x_{j+1} - x_j} u_j + c_j (x - x_j)(x_{j+1} - x)$ over the *j-th* sub-interval. Transforming the approximation to the standard interval [0 l] provides the more convenient form of: $\tilde{u}_j(x) = \frac{l - x}{l} u(0) + \frac{x}{l} u(l) + c_j x(l - x)$. The residual, generated from the insertion of this form into equation (3.31), is written as:

$$R_B = -2\varepsilon c_j + \kappa\left(\frac{u_1 - u_0}{l} + c_j(l - x) - c_j x\right) + \lambda\left(\left(\frac{l - x}{l}\right)u_0 + \left(\frac{x}{l}\right)u_l + c_j x(l - x)\right) \tag{3.32}$$

In a fashion quite similar to what discussed in previous section, the following functional is considered:

$$J_B = \int_0^l R_B{}^2 dx \tag{3.33}$$

minimization of the residual functional with respect to c_j using the least square technique yields:

$$c_j = \frac{5}{2} \frac{(-\lambda^2 l^3 + 12\varepsilon\lambda l)(u_l + u_0) + 24\varepsilon\kappa(u_l - u_0)}{\lambda^2 l^5 - 20\lambda\varepsilon l^3 + 10\kappa^2 l^3 + 120 l\varepsilon^2} \tag{3.34}$$

Note that the previous example was a special case of current discussion, corresponding to the selection of $\varepsilon = 1, \kappa = 0, \lambda = 1$.

The approximation to the solution of (3.31) is written as:

$$\begin{cases} \tilde{u}_j(x) = \{\dfrac{l-x}{l} + (a-b)x(l-x)\}u(0) + \{\dfrac{x}{l} + (a+b)x(l-x)\}u(l) \\ a = \dfrac{5}{2}\dfrac{(-\lambda^2 l^3 + 12\varepsilon\lambda l)}{\lambda^2 l^5 - 20\lambda\varepsilon l^3 + 10\kappa^2 l^3 + 120 l\varepsilon^2} \\ b = \dfrac{5}{2}\dfrac{24\varepsilon\kappa}{\lambda^2 l^5 - 20\lambda\varepsilon l^3 + 10\kappa^2 l^3 + 120 l\varepsilon^2} \end{cases} \tag{3.35}$$

The general weighted residual statement corresponding to the weight function $w(x)$ is written as:

$$\begin{aligned} &0 = \int_0^l w(x)(\varepsilon \tilde{u}_j''(x) + \kappa \tilde{u}_j'(x) + \lambda \tilde{u}_j(x))dx \Rightarrow \\ &\varepsilon\int_0^l -w'(x)\tilde{u}'_j(x)dx + \kappa\int_0^l w(x)\tilde{u}'_j(x)dx + \lambda\int_0^l w(x)\tilde{u}_j(x)dx = -\{\varepsilon w(x)\tilde{u}'_j(x)\}\Big|_0^l. \end{aligned} \tag{3.36}$$

Making selection of the Galerkin weight functions equal to the approximation shape functions:

$w_0(x) = N_I(x) = \dfrac{l-x}{l} + (a-b)x(l-x)$ and $w_1(x) = N_{II}(x) = \dfrac{x}{l} + (a+b)x(l-x)$, the weighted residual statements are written as:

$$\begin{cases} -\varepsilon\displaystyle\int_0^l N'_I(N'_I u_I + N'_{II}u_{II})dx + \kappa\int_0^l N_I(N'_I u_I + N'_{II}u_{II})\,dx \\ \qquad + \lambda\displaystyle\int_0^l N_I(N_I u_I + N_{II}u_{II})dx = -\{\varepsilon N_I\phi\}\Big|_0^l \\ -\varepsilon\displaystyle\int_0^l N'_{II}(N'_I u_I + N'_{II}u_{II})dx + \kappa\int_0^l N_{II}(N'_I u_I + N'_{II}u_{II})\,dx \\ \qquad + \lambda\displaystyle\int_0^l N_{II}(N_I u_I + N_{II}u_{II})dx = -\{\varepsilon N_{II}\phi\}\Big|_0^l \end{cases} \tag{3.37}$$

where, ϕ is the boundary line term. The matrix representation of the above system is:

$$\begin{bmatrix} -\varepsilon\int_0^l N'_I N'_I + \kappa\int_0^l N_I N'_I + \lambda\int_0^l N_I N_I & -\varepsilon\int_0^l N'_{II}N'_I + \kappa\int_0^l N_{II}N'_I + \lambda\int_0^l N_{II}N_I \\ -\varepsilon\int_0^l N'_{II}N'_I + \kappa\int_0^l N_{II}N'_I + \lambda\int_0^l N_{II}N_I & -\varepsilon\int_0^l N'_I N'_I + \kappa\int_0^l N_I N'_I + \lambda\int_0^l N_I N_I \end{bmatrix} \begin{bmatrix} u_I \\ u_{II} \end{bmatrix} = \begin{bmatrix} -\{\varepsilon N_I\phi\}\Big|_0^l \\ -\{\varepsilon N_{II}\phi\}\Big|_0^l \end{bmatrix} \tag{3.38}$$

Substitution of shape functions into the above matrix system and evaluation of integrals and boundary line terms give the following general elemental stiffness matrix:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} u_I \\ u_{II} \end{bmatrix} = \begin{bmatrix} q_I \\ -q_{II} \end{bmatrix}$$

$$\begin{aligned} A &= \frac{-30\varepsilon + 10\lambda l^2 - 15\kappa l + \lambda l^6 (a-b)^2 + 5\lambda l^4 (a-b) - 10\varepsilon l^4 (a-b)^2}{30l} \\ B &= \frac{60\varepsilon + 10\lambda l^2 + 30\kappa l + 2\lambda l^6 (a^2 - b^2) + 10\lambda l^4 a + 20\kappa l^3 a - 20\varepsilon l^4 (a^2 - b^2)}{60l} \\ C &= \frac{60\varepsilon + 10\lambda l^2 - 30\kappa l + 2\lambda l^6 (a^2 - b^2) + 10\lambda l^4 a - 20\kappa l^3 a - 20\varepsilon l^4 (a^2 - b^2)}{60l} \\ D &= \frac{-30\varepsilon + 10\lambda l^2 + 15\kappa l + \lambda l^6 (a+b)^2 + 5\lambda l^4 (a+b) - 10\varepsilon l^4 (a+b)^2}{30l} \end{aligned} \tag{3.39}$$

The selection of parameters $\varepsilon, \kappa, \lambda, l$ in model (3.31), provides the numerical values for the local stiffness matrices and the finite element method can be applied accordingly.

3.6 Higher order practical bubble functions and the approximation error

Questions that arise naturally in the context of polynomial approximation are on the possibility of employing higher order approximations and sufficiency of a selected approximation. Answering these questions enables us to become closer to the actual optimum of least computational cost while achieving higher accuracy. In this section, we address these issues briefly. First, we introduce higher order bubble functions derived using the method of least squares. We will then consider a general error analysis of the least squares approximation using a family of orthogonal functions over the element domain.

The following consideration, deals with the possibilities of increasing the order of polynomial bubbles. Remembering the equation (3.31), let:

$$\bar{u}_j(x)=\frac{l-x}{l}u(0)+\frac{x}{l}u(l)+c_j x(l-x)+f_j x^2(l-x) \tag{3.47}$$

be the cubic approximation of u within the *j-th* element. Note that the above formal representation of the approximate solution is valid if the unknown function u is assumed sufficiently smooth. We insert the approximation (3.47) into the equation (3.31) and introduce the residual functional in a quite similar fashion as what we did in (3.33). Minimization of the residual functional with respect to the unknown parameters c_j *and* f_j yields a linear system of two equations and two unknowns that its solution provides us with the values of c_j *and* f_j expressed in terms of $u(0)$ *and* $u(l)$:

$$c_j=\frac{1}{l}\{\frac{l^7\lambda^4(u(l)-6u(0))-40l^5\lambda^3\varepsilon(u(l)-13u(0))-70l^5\lambda^2\kappa^2(u(l)+2u(0))-60l^4\lambda^2\kappa\varepsilon(13u(l)+22u(0))}{(l^8\lambda^4+52l^6\lambda^2(\kappa^2-2\lambda\varepsilon)+l^4(4320\lambda^2\varepsilon^2-1680\lambda\kappa^2\varepsilon+420\kappa^4)+l^2\varepsilon^2(5040\kappa^2-60480\lambda\varepsilon)+302400\varepsilon^4)}$$
$$+\frac{-840l^3\lambda^2\varepsilon^2(5u(l)-16u(0))+840l^3\lambda\varepsilon\kappa^2(4u(0)-u(l))+5040l^2\varepsilon^2\kappa\lambda(6u(0)-u(l))+2520l^2\kappa^3\varepsilon(u(l)-u(0))}{(l^8\lambda^4+52l^6\lambda^2(\kappa^2-2\lambda\varepsilon)+l^4(4320\lambda^2\varepsilon^2-1680\lambda\kappa^2\varepsilon+420\kappa^4)+l^2\varepsilon^2(5040\kappa^2-60480\lambda\varepsilon)+302400\varepsilon^4)}$$
$$+\frac{50400l\lambda\varepsilon^3(u(l)+2u(0))+25200l\kappa^2\varepsilon^2(u(l)-u(0))+151200\kappa\varepsilon^3(u(l)-u(0))}{(l^8\lambda^4+52l^6\lambda^2(\kappa^2-2\lambda\varepsilon)+l^4(4320\lambda^2\varepsilon^2-1680\lambda\kappa^2\varepsilon+420\kappa^4)+l^2\varepsilon^2(5040\kappa^2-60480\lambda\varepsilon)+302400\varepsilon^4)}\}$$

and

$$f_j=\frac{7}{l}\{\frac{l^6\lambda^4(u(0)-u(l))-80l^4\lambda^3\varepsilon(u(0)-u(l))+10l^4\lambda^2\kappa^2(u(0)-u(l))+300l^3\lambda^2\kappa\varepsilon(u(l)+u(0))}{(l^8\lambda^4+52l^6\lambda^2(\kappa^2-2\lambda\varepsilon)+l^4(4320\lambda^2\varepsilon^2-1680\lambda\kappa^2\varepsilon+420\kappa^4)+l^2\varepsilon^2(5040\kappa^2-60480\lambda\varepsilon)+302400\varepsilon^4)}$$
$$+\frac{1320l^2\lambda^2\varepsilon^2(u(0)-u(l))-600l^2\lambda\varepsilon\kappa^2(u(0)-u(l))-3600l\varepsilon^2\kappa\lambda(u(0)+u(l))+2520l^2\kappa^3\varepsilon(u(l)-u(0))}{(l^8\lambda^4+52l^6\lambda^2(\kappa^2-2\lambda\varepsilon)+l^4(4320\lambda^2\varepsilon^2-1680\lambda\kappa^2\varepsilon+420\kappa^4)+l^2\varepsilon^2(5040\kappa^2-60480\lambda\varepsilon)+302400\varepsilon^4)}$$
$$+\frac{-7200\lambda\varepsilon^3(u(0)-u(l))+7200\kappa^2\varepsilon^2(u(0)-u(l))}{(l^8\lambda^4+52l^6\lambda^2(\kappa^2-2\lambda\varepsilon)+l^4(4320\lambda^2\varepsilon^2-1680\lambda\kappa^2\varepsilon+420\kappa^4)+l^2\varepsilon^2(5040\kappa^2-60480\lambda\varepsilon)+302400\varepsilon^4)}\}$$

corresponding to the cubic bubble function.

Numerical values of the bubble function coefficients can be determined given that the parameters $\varepsilon,\kappa,\lambda,l$ are known in the model equation. As it can be seen from the above discussion, there is no major difficulty associated with the derivation of higher order polynomial bubble functions using least squares fit. However, we consider that the general bubble coefficients are complicated and that the parameters $\varepsilon,\kappa,\lambda,l$ are model dependent. Therefore, we would rather prefer to derive the high order bubble

polynomials while solving actual equations in which the model parameters are given in advance.

We are now in a position to discuss the second issue associated with the least squares polynomial approximation of bubble functions that is what degree polynomial to use. To this end, first consider the notion of orthogonality of an independent set of functions $\{f_i\}_{i=1}^{N}$ with respect to the weight function $w(x)$ over the interval [a b]:

$$\int_a^b w(t) f_i(t) f_j(t) dt = \begin{cases} o & i \neq j \\ \lambda_i > 0 & i = j \end{cases} \tag{3.40}$$

If $\int_a^b w(t) f_i^2(t) dt = 1$ then the system is said to be orthonormal. A well-known example, is the orthogonal set of Legendre polynomials $\{P_i\}_{i=1}^{N}$ with respect to $w(x)=1$ over the standard interval $[-1\ +1]$ where:

$$\int_{-1}^{1} P_m(x) P_n(x) dx = \begin{cases} 0 & m \neq n \\ 2/(2n+1) & m = n \end{cases} \tag{3.41}$$

that is defined by $(n+1)P_{(n+1)}(x) - (2n+1)xP_n(x) + nP_{n-1}(x) = 0$.

Now, suppose that $\{f_i\}_{i=1}^{N}$ is a set of orthonormal polynomials over the interval $[0\ l]$, with respect to weight function $w(x)$. The least squares approximation of the function u on $[0\ l]$ in terms of the base functions is:

$$u(x) \cong \sum_{i=1}^{N} c_i f_i(x). \tag{3.42}$$

According to the Bessel's theorem [17], the following equality holds:

$$\int_0^l w(x) \varepsilon^2(x) dx = \int_0^l w(x) u^2(x) dx - \sum_{j=1}^{N} c_j^2 \tag{3.43}$$

where

$$\begin{cases} \varepsilon(x) = u(x) - \sum_{i=1}^{N} c_i f_i(x) \\ c_j = \int_0^l w(x) u(x) f_j(x) dx. \end{cases} \tag{3.44}$$

The coefficients c_j in (3.42) are called Fourier coefficients. Employing the notation of $\|u\|_{L^2}^2 = \int_0^l w(x) u^2(x) dx$ for the functional space norm of u, it can be seen that:

$$\|\varepsilon\|_{L^2}^2 = \|u\|_{L^2}^2 - \sum_{j=1}^{N} c_j^2 \tag{3.45}$$

In order for the square of the approximation error to be less than a prescribed value δ, it is sufficient that N is found so that:

$$\|u\|_{L^2}^2 \le \delta + \sum_{j=1}^{N} c_j^2 \tag{3.46}$$

On the other hand, if $\|u\|_\infty = sup\{|u(x)|; 0 \le x \le l\}$, a sufficient condition to satisfy (3.46) is to choose N large enough for:

$$\|u\|_\infty^2 \le \frac{1}{l}(\delta + \sum_{j=1}^{N} c_j^2). \tag{3.47}$$

Error bound (3.47) shows that smaller element size l, the lower degree polynomial is required. Although, the inequality (3.46) is sharper than (3.47), they are useful when one is provided with the upper and lower bounds of the unknown function.

Plotting this quantity (figure 3.3) as a function of the degree N of the polynomial being used gives a clue[17]:

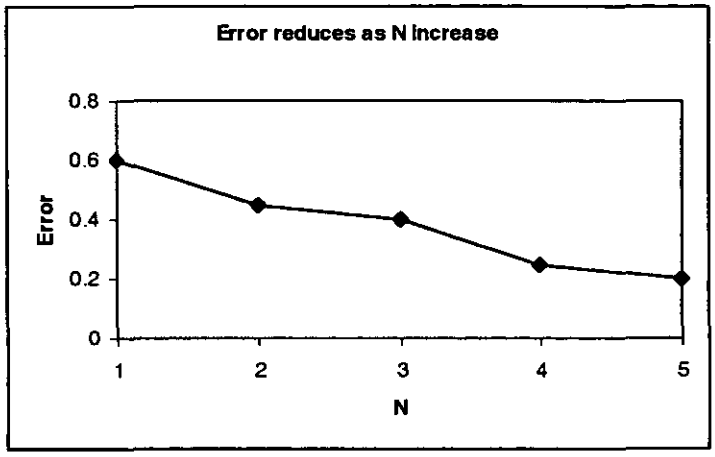


Figure 3.3
Representation of error as a function of degree of polynomial

A remarkable property of least squares fit with orthogonal functions is that each coefficient c_j is determined independently of all the others. Therefore, the use of higher order polynomials is facilitated, as no re-determination of the Fourier coefficients is required.

3.7 Numerical solution of a reaction-diffusion problem using least squares bubble functions: a worked example

From what we discussed in the previous sections, it can be concluded that the least squares bubble enrichment of the finite elements is a potential technique, in which we are able to use crude meshes while improving the element-level accuracy, without assuming any additional elemental midpoints. This technique, in a sense, is similar to the selection of hierarchical finite elements to contribute to the approximation. However, the difference is that, in our approach prior to the implementation of the finite element scheme to evaluate the nodal values, the enriched shape functions are worked out initially in an independent process and in such a way that the optimal bubble coefficients are derived.

In order to observe the effects of the bubble enrichment of the finite elements in improving the approximation with the use of a crude mesh, we consider the so-called diffusion-reaction problem that is used to model several phenomena such as the study of chemical processes or concentration of the pollutants. Consider the following scalar boundary value problem:

$$\begin{cases} \frac{-1}{100}u'' + u = 0, \quad on\ \Omega = [0,10] \\ u(0) = \frac{3}{2}, \frac{du}{dx}(x = 10) = 0 \end{cases} \tag{3.48}$$

having the exact solution:

$$u(x) = \frac{3}{2}\{\frac{\exp(100)\exp(-10x)}{\exp(-100)+\exp(100)} + \frac{\exp(-100)\exp(10x)}{\exp(-100)+\exp(100)}\} \tag{3.49}$$

As discussed earlier in section (2.7), due to the smallness of the diffusion coefficient 0.01, compared to the reaction coefficient 1) this problem presents a sharp drop (steep gradient in the solution) near the boundary wall (figure 3.4) that is called a boundary layer. This creates error in approximation by the use of linears only and hence the failure of the classical Galerkin finite element follows. The problem worsens when the ratio becomes even smaller.

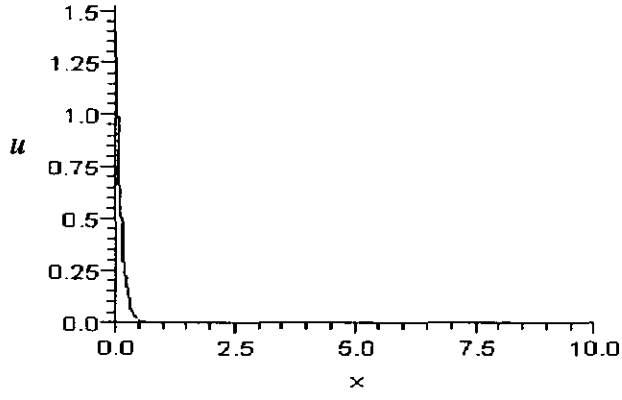


Figure 3.4
The exact solution of equation (3.48)

Now consider the finite element discretization of the problem domain and the enriched quadratic approximation: $\tilde{u}_j(x)=\frac{l-x}{l}u(0)+\frac{x}{l}u(l)+c_j x(l-x)$ within the $j-th$ element. From the discussion in section 3.5 and using the equation (3.34), we get the bubble coefficient:

$$c_j=\frac{-25(25l^2+3)}{250l^4+50l^2+3}(u(0)+u(l)) \tag{3.50}$$

and the elemental stiffness matrix becomes:

$$\begin{bmatrix} \frac{0.3+10l^2+l^6a^2+5l^4a+0.1l^4a^2}{30l} & \frac{-0.6+10l^2+2l^6a^2+10l^4a+0.2l^4a^2}{60l} \\ \frac{-0.6+10l^2+2l^6a^2+10l^4a+0.2l^4a^2}{60l} & \frac{0.3+10l^2+l^6a^2+5l^4a+0.1l^4a^2}{30l} \end{bmatrix} \begin{bmatrix} u_I \\ u_{II} \end{bmatrix} = \begin{bmatrix} q_I \\ -q_{II} \end{bmatrix} \tag{3.51}$$

The remaining step in the finite element solution of this equation is to perform the assembly process in order to obtain the global system of equations. Solution of this system yields the numerical values for the selected nodes. In our example, we have used different meshes of 30 and 50 equal size elements, respectively, and obtained the numerical solutions corresponding to the use of the linear finite elements and quadratic bubble-enriched finite elements. The comparison of the results from the standard linear and the bubble-enriched finite elements against one another and against the exact solution shows that the approximation difficulty is overcome by the use of quadratic bubble functions, whereas a coarser mesh is used (figure 3.5).

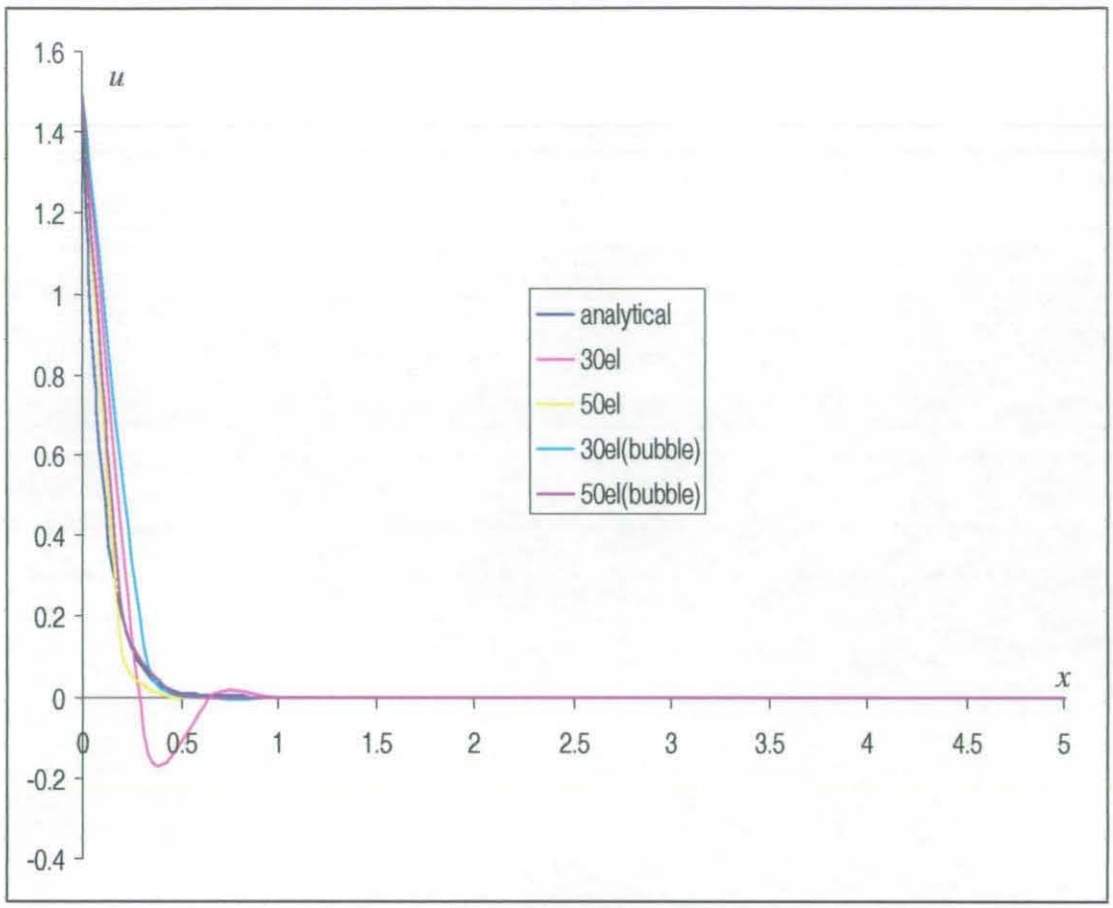


Figure 3.5
Linear, bubble enriched and exact analytical solutions

We have seen in this chapter how higher order element shape functions could be generated in a straightforward way. These higher order bubble functions are in particular important due to the reasons: they possess the merit of simplicity and computational ease due to having polynomial forms. With the least squares bubble functions, the finite element approximation is improved using a p-convergence rather than an h-convergence. We assumed unknown parameters in our definition of the bubble functions, however, the straightforward evaluation of these parameters allowed us to improve the approximation without introducing any additional nodes. Finally, compared to the method of residual free bubble functions, we avoided solving directly the residual differential equation, which increases the practicality of the presented technique. Since the derivation and implementation of the residual free

bubble functions is not a trivial matter in many problems, in practice it becomes inevitable to introduce small elements and use the quadratic or cubic least squares bubble functions to approximate the ideal residual free bubble functions.

Chapter 4

Derivation of Bubble functions for unsteady problems, and extension of the method to multi-dimensional case

In this Chapter, the extension of the main concepts of the least squares bubble functions to transient boundary value problems (i.e. initial value problems) and multi-dimensional cases are described. Here, we introduce a partial discretization method used for the solution of one-dimensional time-dependent problems. We present the quadratic least squares fit for the bubble-enriched finite element solution of a time-dependent initial-boundary problem of the diffusion-reaction type and compare the obtained results with the results generated using linear finite elements and also the exact analytical solutions. Later in this Chapter, higher order least squares bubble functions for unsteady one-dimensional problems are investigated and some basic methods for the derivation of the multi-dimensional bubble functions defined over rectangular and triangular elements are presented.

4.1 Least squares bubble functions for transient problems; partial discretization method

In the Previous Chapters, we studied the steady-state two-point boundary value problems and derived the elemental bubble functions for their solutions based on the method of least squares. However, in a great number of practical problems the conditions of the unknown functions change with time. Therefore, for such problems, dependency on the time has to be taken into account. In such problems, we are

provided with the state of the model at an initial time $t=0$ as well as certain conditions along the exterior boundaries, and in general, the determination of the state of the unknowns at subsequent times is required.

Although a combined discretization of both temporal and spatial derivatives is, in general, possible, however, this approach is complex and shall not be used. We will use the Kantorovich method for partial discretization. Therefore, in the development of the technique for the solution of the initial boundary value problems, the initial and boundary conditions have no effects towards derivation of the bubble functions, as for the case of one-dimensional problems. A justification of this approach can be given as follows: in the partial discretization method, only spatial discretization is carried out which makes the problem, equivalent to a steady state case. This is fulfilled by the procedure that temporal variations in multi-scale problems are captured by the solution of a system of ordinary differential equations with given initial conditions. In these problems, the spatial variables, however, undergo abrupt changes which makes it inevitable for higher resolution to be taken into account in space directions. In other words, in multi-dimensional multi-scale problems higher resolution is required at any directions in space, while in one-dimensional transient cases, increasing the degree of resolution is important in spatial discretization only.

The main interest is to investigate the effects of the bubble enrichment of the finite elements on both local and global stiffness matrices. Therefore, it will be more illustrative to first describe the application of the partial discretization method to transient boundary value problems and then use the least squares bubble functions to enrich the finite element scheme.

Let us, consider the following time-dependent initial-boundary value problem:

$$\begin{cases} \beta\dfrac{\partial u}{\partial t}+\varepsilon\dfrac{\partial^2 u}{\partial x^2}+\lambda u=g(x) & \\ u(x,t)=f(x) & t=0 \\ u(x,t)=0 & x=a, t\geq 0 \\ u(x,t)=0 & x=b, t\geq 0 \end{cases} \tag{4.1}$$

where $\beta,\varepsilon,\lambda$ are constants and we take $g=0$ for the sake of simplicity. We proceed with the finite element discretization of the domain in x-direction and set: $a=x_0<x_1<\ldots<x_n=b$.

Consider the trial function approximation:

$$\hat{u}_j(x,t)=\alpha_{1_j}(t)\frac{x}{l_j}+\alpha_{2_j}(t)\frac{l_j-x}{l_j} \tag{4.2}$$

defined within the *j-th* element of domain discretization and zero elsewhere (we shall drop the subscript *j*).

The weighted residual statement requires the residual generated from replacing identity (4.2) into equation (4.1) to be zero on average i.e.:

$$0=\int_0^l w(\beta\frac{\partial\hat{u}}{\partial t}+\varepsilon\frac{\partial^2 u}{\partial x}+\lambda u)dx \tag{4.3}$$

Use of the Green's integral theorem (integration by part) results in the weak form:

$$0=\int_0^l \beta w(x)\frac{\partial\hat{u}}{\partial t}dx-\int_0^l \varepsilon\frac{\partial w}{\partial x}\frac{\partial u}{\partial x}dx+\int_0^l \lambda w(x).udx=-\varepsilon w(x)\frac{\partial u}{\partial x}|_0^l \tag{4.4}$$

We make a selection of the Galerkin's method and set the weight functions:

$w_I(x)=\dfrac{x}{l}$ *and* $w_{II}(x)=\dfrac{l-x}{l}$, respectively.

We drop the index *j* from the element length *l*, making the assumption of elements with equal size. Inserting the approximation (4.2) into formulation (4.4) gives the following elemental system of equations:

$$\beta\begin{bmatrix} \frac{l}{3} & \frac{l}{6} \\ \frac{l}{6} & \frac{l}{3} \end{bmatrix}\begin{bmatrix} \alpha'_{1j}(t) \\ \alpha'_{2j}(t) \end{bmatrix}+\begin{bmatrix} \frac{\lambda l}{3}-\frac{\varepsilon}{l} & \frac{\lambda l}{6}+\frac{\varepsilon}{l} \\ \frac{\lambda l}{6}+\frac{\varepsilon}{l} & \frac{\lambda l}{3}-\frac{\varepsilon}{l} \end{bmatrix}\begin{bmatrix} \alpha_{1j}(t) \\ \alpha_{2j}(t) \end{bmatrix}=\begin{bmatrix} \phi_I \\ -\phi_{II} \end{bmatrix} \tag{4.5}$$

where ϕ_I, ϕ_{II} are the boundary integral terms.

Equation set (4.5) is written in the form:

$$\beta\vec{C}\frac{d\alpha}{dt}+\vec{K}\alpha=\vec{f} \tag{4.6}$$

with the elemental matrices:

$$\alpha=\begin{bmatrix} \alpha_{1j} \\ \alpha_{2j} \end{bmatrix}, C=\begin{bmatrix} \frac{l}{3} & \frac{l}{6} \\ \frac{l}{6} & \frac{l}{3} \end{bmatrix}, K=\begin{bmatrix} \frac{\lambda l}{3}-\frac{\varepsilon}{l} & \frac{\lambda l}{6}+\frac{\varepsilon}{l} \\ \frac{\lambda l}{6}+\frac{\varepsilon}{l} & \frac{\lambda l}{3}-\frac{\varepsilon}{l} \end{bmatrix}, f=\begin{bmatrix} -\varepsilon.\frac{\partial\alpha_1}{\partial x}.N_I \\ \varepsilon\frac{\partial\alpha_2}{\partial x}N_{II} \end{bmatrix} \tag{4.7}$$

Performing the assembly process and applying essential boundary conditions to the obtained global system of equations, along with removal of the equations corresponding to the first and the last time-dependent functions, results in a set of first order ordinary differential equations analogous to (4.6) for each α_j i.e. we get:

$$\alpha_1=\alpha_n=0, \quad c_j\frac{d\alpha_j}{dt}+k_j\alpha_j=0 \tag{4.8}$$

With the solution $\alpha_j(t)=A_j e^{-\frac{k_j}{c_j}t}$ and the values A_j are found according to the local boundary conditions. Finally, the approximation to the solution of (4.1) is written:

$$\hat{u}_j(x,t)=\sum_{j=1}^{N}\alpha_{1_j}(t)\frac{x}{l}+\alpha_{2_j}(t)\frac{l-x}{l}. \tag{4.9}$$

With the short description given above, we are ready to employ our technique of the least squares bubble functions to enrich the finite elements for the transient problems.

Now consider the following equation:

$$\begin{cases} \dfrac{\partial u}{\partial t}+\varepsilon\dfrac{\partial^2 u}{\partial x^2}+u=0 & \\ u(x,t)=f(x) & t=0 \\ u(x,t)=0 & x=a, t\geq 0 \\ u(x,t)=0 & x=b, t\geq 0 \end{cases} \tag{4.10}$$

and for the finite element discretization of the domain in x-direction set: $a=x_0<x_1<\ldots<x_n=b$. Consider the trial function approximation:

$$\hat{u}_j(x,t)=\alpha_{1_j}(t)\{\frac{x}{l}+cx(l-x)\}+\alpha_{2_j}(t)\{\frac{l-x}{l}+cx(l-x)\} \tag{4.11}$$

defined within the *j-th* element of domain discretization and zero elsewhere. Note that we have made the selection of quadratic least squares bubble function as developed in (3.31)-(3.35) of previous Chapter to deal with the ordinary differential equation $\varepsilon\dfrac{d^2u}{dx^2}+u=0$. In this case, the unknown bubble coefficient is found:

$$c=\frac{-5}{2}\frac{(l^2-12\varepsilon)}{l^4-20\varepsilon l^2+120\varepsilon^2} \tag{4.12}$$

The Galerkin's weighted residual statements corresponding to the weight functions $w_0(x)=\dfrac{l-x}{l}+cx(l-x)$ *and* $w_1(x)=\dfrac{x}{l}+cx(l-x)$ give rise to the following system of equations:

$$\begin{bmatrix} \dfrac{c^2l^6+5cl^4+10l^2}{30l} & \dfrac{c^2l^6+5cl^4+5l^2}{30l} \\ \dfrac{c^2l^6+5cl^4+5l^2}{30l} & \dfrac{c^2l^6+5cl^4+10l^2}{30l} \end{bmatrix} \begin{bmatrix} \alpha'_{1j}(t)+\alpha_{1j}(t) \\ \alpha'_{2j}(t)+\alpha_{2j}(t) \end{bmatrix}$$
$$+\begin{bmatrix} \dfrac{-\varepsilon(10c^2l^4+30)}{30l} & \dfrac{-\varepsilon(10c^2l^4-30)}{30l} \\ \dfrac{-\varepsilon(10c^2l^4-30)}{30l} & \dfrac{-\varepsilon(10c^2l^4+30)}{30l} \end{bmatrix} \begin{bmatrix} \alpha_{1j}(t) \\ \alpha_{2j}(t) \end{bmatrix} = \begin{bmatrix} -\varepsilon\dfrac{\partial\hat{u}}{\partial x}\Big|_0^l \\ \varepsilon\dfrac{\partial\hat{u}}{\partial x}\Big|_0^l \end{bmatrix}. \tag{4.13}$$

For the sake of notation simplicity, we set:

$$E=\frac{c^2l^6+5cl^4+10l^2}{30l},F=\frac{c^2l^6+5cl^4+5l^2}{30l}$$
$$G=\frac{-\varepsilon(10c^2l^4+30)}{30l},H=\frac{-\varepsilon(10c^2l^4-30)}{30l} \tag{4.14}$$
$$\begin{bmatrix}E & F\\ F & E\end{bmatrix}\begin{bmatrix}\alpha'_{1j}(t)+\alpha_{1j}(t)\\ \alpha'_{2j}(t)+\alpha_{2j}(t)\end{bmatrix}+\begin{bmatrix}G & H\\ H & G\end{bmatrix}\begin{bmatrix}\alpha_{1j}(t)\\ \alpha_{2j}(t)\end{bmatrix}=\begin{bmatrix}\phi_I\\ -\phi_{II}\end{bmatrix}.$$

Performing the assembly process results in the global matrix system:

$$\begin{bmatrix}E & F & 0 & 0 & . & 0\\ F & 2E & F & 0 & . & 0\\ 0 & . & . & . & . & 0\\ 0 & . & . & . & . & .\\ . & . & . & . & 2E & F\\ 0 & 0 & 0 & . & F & E\end{bmatrix}\begin{bmatrix}\alpha'_{11}(t)+\alpha_{11}(t)\\ \alpha'_{21}(t)+\alpha_{21}(t)\\ .\\ .\\ \alpha'_{1n}(t)+\alpha_{1n}(t)\\ \alpha'_{2n}(t)+\alpha_{2n}(t)\end{bmatrix}$$
$$+\begin{bmatrix}G & H & 0 & 0 & . & 0\\ H & 2G & H & . & . & 0\\ 0 & . & . & . & . & 0\\ 0 & . & . & . & . & .\\ . & . & . & . & 2G & H\\ 0 & 0 & 0 & . & H & G\end{bmatrix}\begin{bmatrix}\alpha_{11}(t)\\ \alpha_{21}(t)\\ .\\ .\\ \alpha_{1n}(t)\\ \alpha_{2n}(t)\end{bmatrix}=\begin{bmatrix}\phi_1\\ 0\\ 0\\ .\\ .\\ \phi_n\end{bmatrix} \tag{4.15}$$

Imposition of the essential boundary conditions $u(a,t)=0$ *and* $u(b,t)=0$ and elimination of redundant equations corresponding to the boundary values will result into solution of system and evaluation of the nodal values.

4.2 Application of the method to a transient problem

In order to illustrate what described in section (4.1), we shall consider the following initial boundary value problem:

$$\begin{cases}\dfrac{\partial u}{\partial t}-\dfrac{\partial^2 u}{\partial x^2}=0\\ u(x,t)=x(1-x) & t=0\\ u(x,t)=0 & x=0,t\geq 0\\ u(x,t)=0 & x=1,t\geq 0\end{cases} \tag{4.16}$$

To start, consider first the linear approximation $\hat{u}_j(x,t) = \alpha_{1_j}(t)\frac{x}{l} + \alpha_{2_j}(t)\frac{l-x}{l}$, within the *j-th* element. The weighted residual statement is:

$$\int_0^l w_j\left(\frac{\partial^2 \hat{u}}{\partial x^2} - \frac{\partial \hat{u}}{\partial t}\right)dx = 0,\ j = 1,2,...,M+1 \tag{4.17}$$

and using the integration by parts gives the weak form:

$$\int_0^l \left(\frac{\partial \hat{u}}{\partial x}\frac{\partial w_j}{\partial x} + \frac{\partial \hat{u}}{\partial t}w_j\right)dx - w_j\frac{\partial \hat{u}}{\partial x}\Big|_0^l = 0,\ j = 1,2,...,M+1. \tag{4.18}$$

The Galerkin weighted residual statement corresponding to the selection of weight functions equal to the shape functions $w_I(x) = \frac{x}{l}$ *and* $w_{II}(x) = \frac{l-x}{l}$ and setting *M=2*, *l=0.5* gives rise to the following assembled system of global equations:

$$\frac{1}{2}\begin{bmatrix} \frac{1}{3} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} \alpha_1'(t) \\ \alpha_2'(t) \\ \alpha_3'(t) \end{bmatrix} + 2\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} \alpha_1(t) \\ \alpha_2(t) \\ \alpha_3(t) \end{bmatrix} = \begin{bmatrix} \phi_I \\ 0 \\ -\phi_{III} \end{bmatrix} \tag{4.19}$$

The imposition of essential boundary conditions at $x = 0, x = 1$ yields:

$$\frac{1}{2}\begin{bmatrix} \frac{1}{3} & \frac{1}{6} & 0 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 0 \\ \alpha_2'(t) \\ 0 \end{bmatrix} + 2\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ \alpha_2(t) \\ 0 \end{bmatrix} = \begin{bmatrix} \phi_I \\ 0 \\ -\phi_{III} \end{bmatrix} \tag{4.20}$$

and the only equation to be solved is the ordinary differential equation:

$$\frac{1}{3}\alpha_2'(t) + 4\alpha_2(t) = 0 \tag{4.21}$$

which its solution is $\alpha_2(t) = \frac{1}{4}e^{-12t}$, considering that $\hat{u}(\frac{1}{2},0) = \frac{1}{4}$.

Now, let $\hat{u}_j(x,t) = \alpha_{1_j}(t)\{\frac{x}{l} + cx(l-x)\} + \alpha_{2_j}(t)\{\frac{l-x}{l} + cx(l-x)\}$ be the trial approximation of the solution of (4.16) where $c = \frac{-5}{2}\frac{(l^2+12)}{l^4+20l^2+120}$. In a quite similar fashion to what described in (4.17)-(4.19) we obtain:

$$\begin{bmatrix} 0.1616 & 0.0782 \\ 0.0782 & 0.1616 \end{bmatrix}\begin{bmatrix} \alpha_1'(t) \\ \alpha_2'(t) \end{bmatrix} + \begin{bmatrix} 2.002 & -1.9975 \\ -1.9975 & 2.002 \end{bmatrix}\begin{bmatrix} \alpha_{1j}(t) \\ \alpha_{2j}(t) \end{bmatrix} = \begin{bmatrix} \phi_I \\ -\phi_{II} \end{bmatrix} \tag{4.22}$$

and the global stiffness matrix is:

$$\begin{bmatrix} 0.1616 & 0.0782 & 0 \\ 0.0782 & 0.3232 & 0.0782 \\ 0 & 0.0782 & 0.1616 \end{bmatrix}\begin{bmatrix} \alpha_1'(t) \\ \alpha_2'(t) \\ \alpha_3'(t) \end{bmatrix} + \begin{bmatrix} 2.002 & -1.9975 & 0 \\ -1.9975 & 4.004 & -1.9975 \\ 0 & -1.9975 & 2.002 \end{bmatrix}\begin{bmatrix} \alpha_1(t) \\ \alpha_2(t) \\ \alpha_3(t) \end{bmatrix} = \begin{bmatrix} \phi_I \\ 0 \\ -\phi_{III} \end{bmatrix} \tag{4.23}$$

and the imposition of the boundary conditions results:

$$\begin{bmatrix} 0.1616 & 0.0782 & 0 \\ 0.0782 & 0.3232 & 0.0782 \\ 0 & 0.0782 & 0.1616 \end{bmatrix}\begin{bmatrix} 0 \\ \alpha_2'(t) \\ 0 \end{bmatrix} + \begin{bmatrix} 2.002 & -1.9975 & 0 \\ -1.9975 & 4.004 & -1.9975 \\ 0 & -1.9975 & 2.002 \end{bmatrix}\begin{bmatrix} 0 \\ \alpha_2(t) \\ 0 \end{bmatrix} = \begin{bmatrix} \phi_I \\ 0 \\ -\phi_{III} \end{bmatrix} \tag{4.24}$$

which provides the equation:

$$0.3232\alpha_2'(t) + 4.004\alpha_2(t) = 0 \tag{4.25}$$

and we obtain $\alpha_2(t) = \frac{1}{4}e^{(-12.38)t}$.

4.3 The least squares bubble function: A worked example

Consider the problem of one-dimensional unsteady heat conduction with the reaction term in the region $0 \le x \le \pi$ is governed by the following initial-boundary equation:

$$\begin{cases} \dfrac{\partial u}{\partial t} - \dfrac{\partial^2 u}{\partial x^2} + u = 0 & \\ u(x,t) = sin(x) & t = 0 \\ u(x,t) = 0 & x = 0, t \geq 0 \\ u(x,t) = 0 & x = \pi, t \geq 0 \end{cases} \tag{4.26}$$

with the exact solution given as: $u(x,t) = sin(x)e^{-2t}, x \in [0,\pi]$ *and* $t \geq 0$.

We solve this equation, using the partial discretization method and with the use of standard linear and bubble enriched finite elements, respectively, and compare our results against the exact values.

Consider the finite element discretization of the equation domain in the x direction as $0 = x_0 < x_1 < ... < x_n = \pi$ and let the trial approximation within the *j-th* element be written as: $\hat{u}_j(x,t) = \alpha_{1_j}(t)\dfrac{x}{l} + \alpha_{2_j}(t)\dfrac{l-x}{l}$.

The weighted residual statement corresponding to the weight function w_j is: $\int_0^l w_j(\hat{u} - \dfrac{\partial^2 \hat{u}}{\partial x^2} + \dfrac{\partial \hat{u}}{\partial t})dx = 0,\ j = 1,2,...,M+1$ and using integration by parts requires the weighted average to be zero i.e.:

$$\int_0^l (\hat{u}w_j + \frac{\partial \hat{u}}{\partial x}\frac{\partial w_j}{\partial x} + \frac{\partial \hat{u}}{\partial t}w_j)dx - w_j \frac{\partial \hat{u}}{\partial x}\Big|_0^l = 0,\ j = 1,2,...,M+1. \tag{4.27}$$

With the selection of the weight functions equal to the linear shape functions as in previous examples i.e. $w_I(x) = \dfrac{x}{l}$ *and* $w_{II}(x) = \dfrac{l-x}{l}$, we obtain:

$$\begin{cases} \int_0^l (\dfrac{x}{l}\hat{u} + \dfrac{1}{l}\dfrac{\partial \hat{u}}{\partial x} + \dfrac{x}{l}\dfrac{\partial \hat{u}}{\partial t})dx = \dfrac{x}{l}\dfrac{\partial \hat{u}}{\partial x}\Big|_0^l \\ \int_0^l (\dfrac{l-x}{l}\hat{u} - \dfrac{1}{l}\dfrac{\partial \hat{u}}{\partial x} + \dfrac{l-x}{l}\dfrac{\partial \hat{u}}{\partial t})dx = \dfrac{l-x}{l}\dfrac{\partial \hat{u}}{\partial x}\Big|_0^l \end{cases} \tag{4.28}$$

which results in the following local system of equations:

$$\begin{bmatrix} \frac{l}{3} & \frac{l}{6} \\ \frac{l}{6} & \frac{l}{3} \end{bmatrix} \begin{bmatrix} \alpha'_{1j}(t) \\ \alpha'_{2j}(t) \end{bmatrix} + \begin{bmatrix} \frac{l}{3}+\frac{1}{l} & \frac{l}{6}-\frac{1}{l} \\ \frac{l}{6}-\frac{1}{l} & \frac{l}{3}+\frac{1}{l} \end{bmatrix} \begin{bmatrix} \alpha_{1j}(t) \\ \alpha_{2j}(t) \end{bmatrix} = \begin{bmatrix} \left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \\ -\left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \end{bmatrix} \tag{4.29}$$

If we restrict the consideration to two linear elements of equal lengths, that is, $M = 2, l = \frac{\pi}{2}$, then (4.29) can be re-written as:

$$\begin{bmatrix} \frac{\pi}{6} & \frac{\pi}{12} \\ \frac{\pi}{12} & \frac{\pi}{6} \end{bmatrix} \begin{bmatrix} \alpha'_{1j}(t) \\ \alpha'_{2j}(t) \end{bmatrix} + \begin{bmatrix} \frac{\pi}{6}+\frac{2}{\pi} & \frac{\pi}{12}-\frac{2}{\pi} \\ \frac{\pi}{12}-\frac{2}{\pi} & \frac{\pi}{6}+\frac{2}{\pi} \end{bmatrix} \begin{bmatrix} \alpha_{1j}(t) \\ \alpha_{2j}(t) \end{bmatrix} = \begin{bmatrix} \left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \\ -\left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \end{bmatrix} \tag{4.30}$$

and the assembled system of equations is:

$$\begin{bmatrix} \frac{\pi}{6} & \frac{\pi}{12} & 0 \\ \frac{\pi}{12} & \frac{\pi}{3} & \frac{\pi}{12} \\ 0 & \frac{\pi}{12} & \frac{\pi}{6} \end{bmatrix} \begin{bmatrix} \alpha'_1(t) \\ \alpha'_2(t) \\ \alpha'_3(t) \end{bmatrix} + \begin{bmatrix} \frac{\pi}{6}+\frac{2}{\pi} & \frac{\pi}{12}-\frac{2}{\pi} & 0 \\ \frac{\pi}{12}-\frac{2}{\pi} & \frac{\pi}{3}+\frac{4}{\pi} & \frac{\pi}{12}-\frac{2}{\pi} \\ 0 & \frac{\pi}{12}-\frac{2}{\pi} & \frac{\pi}{6}+\frac{2}{\pi} \end{bmatrix} \begin{bmatrix} \alpha_1(t) \\ \alpha_2(t) \\ \alpha_3(t) \end{bmatrix} = \begin{bmatrix} \left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \\ 0 \\ -\left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \end{bmatrix} \tag{4.31}$$

However, the imposition of the boundary conditions from (4.26) requires the values of α_1 *and* α_3 to be zero that gives rise to the following ordinary differential equation:

$$\frac{\pi}{3}\alpha'_2(t) + (\frac{\pi}{3}+\frac{4}{\pi})\alpha_2(t) = 0 \tag{4.32}$$

Since $u(\frac{\pi}{2},0) = 1 = sin(\frac{\pi}{2})$ we find $\alpha_2(t) = \exp(-t - \frac{12t}{\pi^2}) = \exp(-2.216t)$.

Moreover, the piecewise global approximation is found to be:

$$u(x,t) = \begin{cases} \frac{2x}{\pi}\exp(-2.216t) & if\ x \in [0, \frac{\pi}{2}] \\ \frac{2(\pi - x)}{\pi}\exp(-2.216t) & if\ x \in [\frac{\pi}{2}, \pi] \end{cases} \tag{4.33}$$

Now, in order to examine the performance of our second approach (i.e. the least squares bubble function enrichment of finite elements) we remember the discussion

leading to the selection of the quadratic shape functions (4.11) and the bubble coefficient (4.12). For this purpose, let us assume the approximate solution of equation (4.26) be:

$$\hat{u}_j(x,t) = \alpha_{1_j}(t)\{\frac{x}{l} + cx(l-x)\} + \alpha_{2_j}(t)\{\frac{l-x}{l} + cx(l-x)\} \tag{4.34}$$

within the *j-th* element and zero elsewhere. With the quadratic polynomial bubble coefficient introduced in (4.12) and corresponding to $\varepsilon = -1$ we find that:

$$c = \frac{-5}{2}\frac{(l^2+12)}{l^4+20l^2+120} \tag{4.35}$$

The Galerkin's weighted residual statements corresponding to the weight functions $w_0(x) = \frac{l-x}{l} + cx(l-x)$ *and* $w_1(x) = \frac{x}{l} + cx(l-x)$ result:

$$\begin{bmatrix} \frac{c^2l^6+5cl^4+10l^2}{30l} & \frac{c^2l^6+5cl^4+5l^2}{30l} \\ \frac{c^2l^6+5cl^4+5l^2}{30l} & \frac{c^2l^6+5cl^4+10l^2}{30l} \end{bmatrix} \begin{bmatrix} \alpha'_{1j}(t)+\alpha_{1j}(t) \\ \alpha'_{2j}(t)+\alpha_{2j}(t) \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{-\varepsilon(10c^2l^4+30)}{30l} & \frac{-\varepsilon(10c^2l^4-30)}{30l} \\ \frac{-\varepsilon(10c^2l^4-30)}{30l} & \frac{-\varepsilon(10c^2l^4+30)}{30l} \end{bmatrix} \begin{bmatrix} \alpha_{1j}(t) \\ \alpha_{2j}(t) \end{bmatrix} = \begin{bmatrix} -\varepsilon\frac{\partial\hat{u}}{\partial x}\Big|_0^l \\ \varepsilon\frac{\partial\hat{u}}{\partial x}\Big|_0^l \end{bmatrix}. \tag{4.36}$$

Similarly, if we select $M = 2, l = \frac{\pi}{2}$ then $c = 0.2061$ and the system (4.36) becomes:

$$\begin{bmatrix} 0.67027 & 0.40848 \\ 0.40848 & 0.67027 \end{bmatrix} \begin{bmatrix} \alpha'_{1j}(t) \\ \alpha'_{2j}(t) \end{bmatrix} + \begin{bmatrix} 1.36174 & -0.17432 \\ -0.17432 & 1.36174 \end{bmatrix} \begin{bmatrix} \alpha_{1j}(t) \\ \alpha_{2j}(t) \end{bmatrix} = \begin{bmatrix} -\varepsilon\frac{\partial\hat{u}}{\partial x}\Big|_0^l \\ \varepsilon\frac{\partial\hat{u}}{\partial x}\Big|_0^l \end{bmatrix}. \tag{4.37}$$

The assembled system of equations is:

$$\begin{bmatrix} 0.67027 & 0.40848 & 0 \\ 0.40848 & 1.34054 & 0.40848 \\ 0 & 0.40848 & 0.67027 \end{bmatrix} \begin{bmatrix} \alpha_1'(t) \\ \alpha_2'(t) \\ \alpha_3'(t) \end{bmatrix}$$

$$+ \begin{bmatrix} 1.36174 & -0.17432 & 0 \\ -0.17432 & 2.72348 & -0.17432 \\ 0 & -0.17432 & 1.36174 \end{bmatrix} \begin{bmatrix} \alpha_1(t) \\ \alpha_2(t) \\ \alpha_3(t) \end{bmatrix} = \begin{bmatrix} \left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \\ 0 \\ -\left.\frac{\partial \hat{u}}{\partial x}\right|^{x=l} \end{bmatrix} \tag{4.38}$$

The imposition of boundary conditions implies that $\alpha_1 = \alpha_3 = 0$. After the elimination of the redundant equations from system (4.38), we find α_2 satisfying the ordinary differential equation:

$$1.34054\,\alpha_2'(t) + 2.72348\,\alpha_2(t) = 0 \tag{4.39}$$

which its solution is $\alpha_2(t) = exp(-2.0316t)$.

The piecewise global approximation is found as:

$$u(x,t) = \begin{cases} \{\frac{2x}{\pi} + 0.2061x(\frac{\pi}{2} - x)\}\exp(-2.0316t) & if\ x \in [0, \frac{\pi}{2}] \\ \{\frac{2(\pi - x)}{\pi} + 0.2061(\pi - x)(x - \frac{\pi}{2})\}\exp(-2.0316t) & if\ x \in [\frac{\pi}{2}, \pi] \end{cases} \tag{4.40}$$

corresponding to the selection of two elements of equal length given above.

Feasibility of the generalization of the least squares bubble functions to the solution of time dependent equations, and its similarity to the one dimensional case, is illustrated in the above development. Furthermore, a comparison between the exact solution of the initial-boundary equation (4.26) and the approximations that we have just constructed, demonstrates the efficiency of the approximation by least squares bubble functions and its superiority over the linear finite elements in the solution of the unsteady partial differential equation. It can be observed that the quadratic bubble-enriched finite elements, provide closer estimation of the behaviour of the exact analytical solution, even though a crude mesh is adopted to test the technique. The

following tables and figures presented at different cross sections of the solution, provide an intuitive proof of this statement. These results are selected from the solution profiles at different time steps and space level for $t = 0, t = 0.9$ *and* $x = \frac{7\pi}{8}$, respectively.

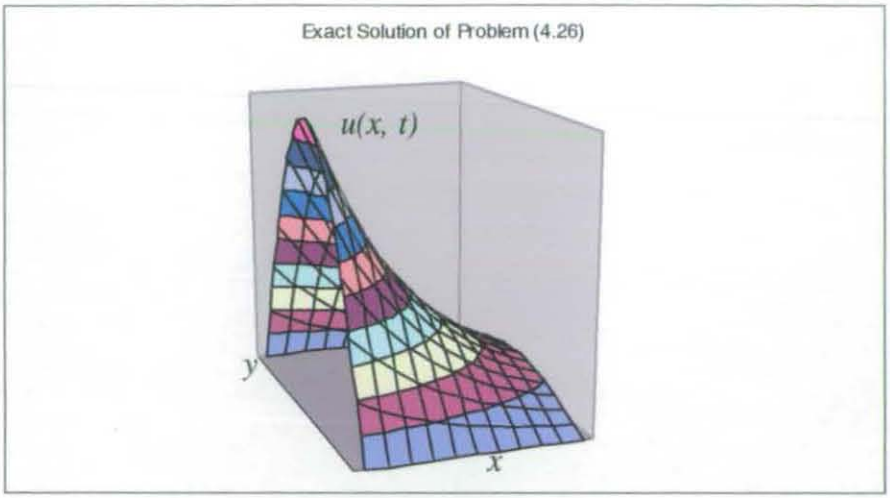


Figure 4.1(a)
Exact solution of problem (4.26)

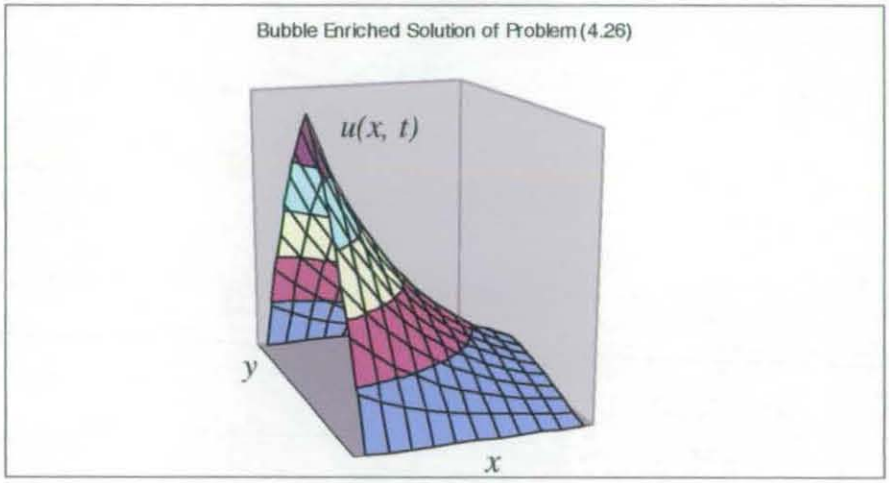


Figure 4.1(b)
Bubble enriched solution of problem (4.26)

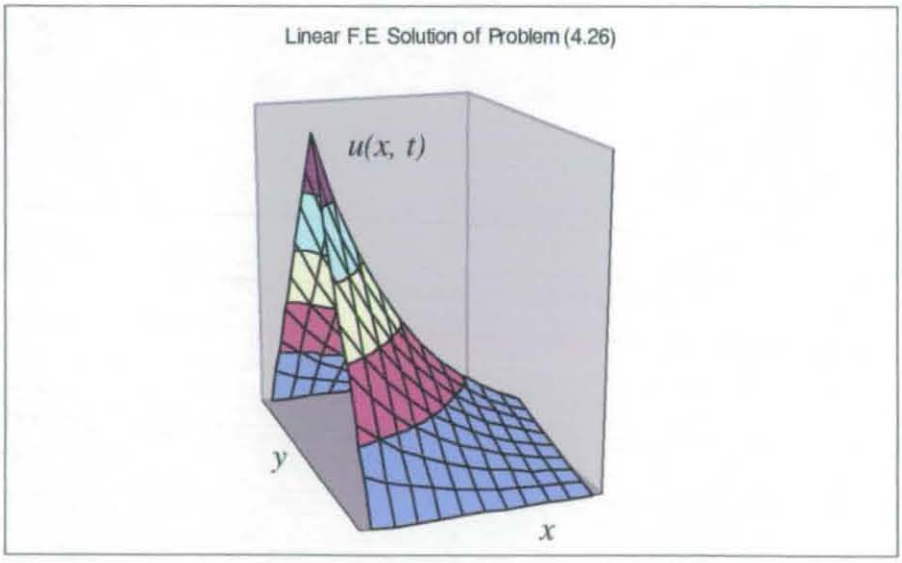


Figure 4.1(c)
Linear F.E. solution of problem (4.26)

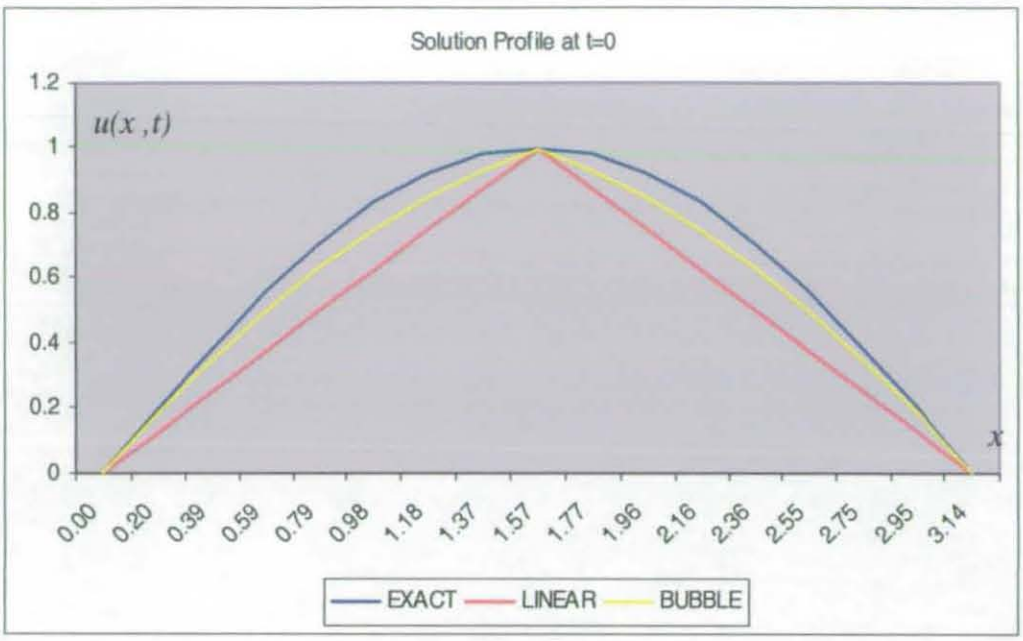


Figure 4.2
Solution profile at $t = 0, 0 \leq x \leq \pi$

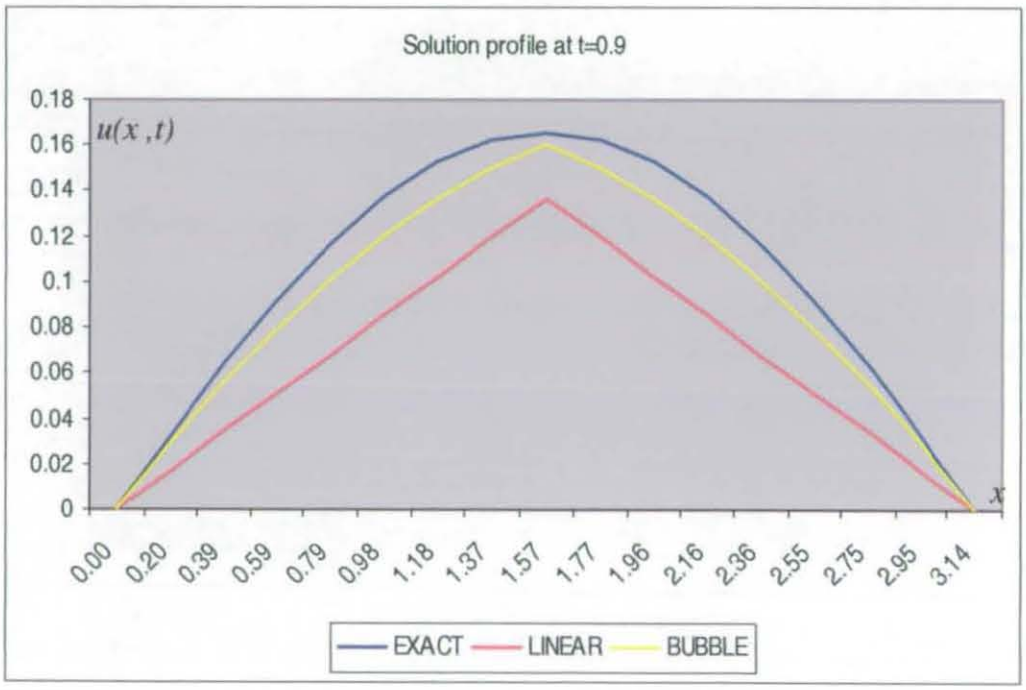


Figure 4.3
Solution profile at $t = 0.9, 0 \leq x \leq \pi$

t=0	u(x, t)	u(x, t)	u(x, t)
x	Exact	Bubble	Linear
0	0	0	0
$\frac{\pi}{16}$	0.195090322	0.180620618	0.125
$\frac{\pi}{8}$	0.382683432	0.345349631	0.25
$\frac{3\pi}{16}$	0.555570233	0.494187039	0.375
$\frac{\pi}{4}$	0.707106781	0.627132842	0.5
$\frac{5\pi}{16}$	0.831469612	0.744187039	0.625
$\frac{3\pi}{8}$	0.923879533	0.845349631	0.75
$\frac{7\pi}{16}$	0.98078528	0.930620618	0.875
$\frac{\pi}{2}$	1	1	1
$\frac{9\pi}{16}$	0.98078528	0.930620618	0.875
$\frac{\pi}{8}$	0.923879533	0.845349631	0.75
$\frac{11\pi}{16}$	0.831469612	0.744187039	0.625
$\frac{3\pi}{4}$	0.707106781	0.627132842	0.5
$\frac{13\pi}{16}$	0.555570233	0.494187039	0.375
$\frac{7\pi}{8}$	0.382683432	0.345349631	0.25
$\frac{15\pi}{16}$	0.195090322	0.180620618	0.125
π	0	0	0

Table 4.1
Comparison of the results at $t=0$

t=0.9	u(x, t)	u(x, t)	u(x, t)
x	Exact	Linear	Bubble
0	0	0	0
$\frac{\pi}{16}$	0.032248213	0.017027228	0.029019232
$\frac{\pi}{8}$	0.063257146	0.034054457	0.055485256
$\frac{3\pi}{16}$	0.091835142	0.051081685	0.079398071
$\frac{\pi}{4}$	0.116883965	0.068108914	0.100757676
$\frac{5\pi}{16}$	0.137441003	0.085136142	0.119564073
$\frac{3\pi}{8}$	0.15271626	0.102163371	0.13581726
$\frac{7\pi}{16}$	0.162122716	0.119190599	0.149517239
$\frac{\pi}{2}$	0.165298888	0.136217828	0.160664008
$\frac{9\pi}{16}$	0.162122716	0.119190599	0.149517239
$\frac{\pi}{8}$	0.15271626	0.102163371	0.13581726
$\frac{11\pi}{16}$	0.137441003	0.085136142	0.119564073
$\frac{3\pi}{4}$	0.116883965	0.068108914	0.100757676
$\frac{13\pi}{16}$	0.091835142	0.051081685	0.079398071
$\frac{7\pi}{8}$	0.063257146	0.034054457	0.055485256
$\frac{15\pi}{16}$	0.032248213	0.017027228	0.029019232
π	0	0	0

Table 4.2
Comparison of the results at $t=0.9$

$x = \frac{7\pi}{8}$	u(x, t)	u(x, t)	u(x, t)
t	Exact	Linear	Bubble
0	0.382683432	0.25	0.345349631
0.1	0.313314695	0.200328981	0.281856289
0.2	0.256520376	0.160526802	0.230036347
0.3	0.210021121	0.128632682	0.187743624
0.4	0.17195075	0.103075416	0.153226517
0.5	0.140781367	0.082595972	0.125055461
0.6	0.115262035	0.066185468	0.10206372
0.7	0.094368573	0.053035469	0.083299065
0.8	0.077262452	0.042498166	0.067984336
0.9	0.063257146	0.034054457	0.055485256
1	0.051790571	0.027288379	0.045284161

Table 4.3

Comparison of the results at $x = \frac{7\pi}{8}$

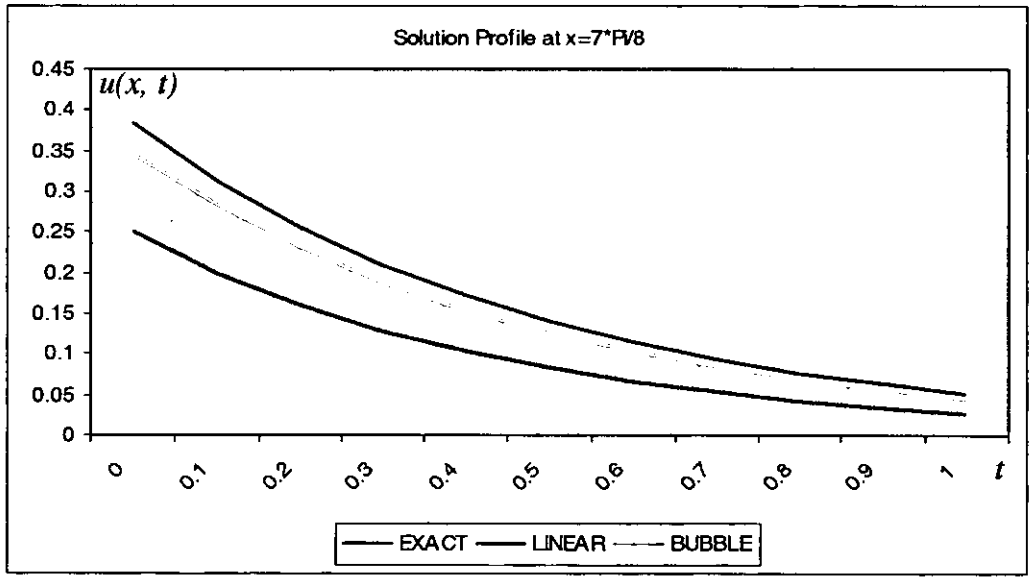


Figure 4.4

Solution profile at $x = \frac{7\pi}{8}, 0 \le t$

4.4 Multi-dimensional problems and bubble functions: general ideas of the extension

One important advantage of the method of finite elements is the ability to solve partial differential equations defined on geometrically complex domains in two or three dimensions. In two-dimensional analysis, triangles are particularly useful since every complex domain can be accurately triangulated using sufficient number of triangles. It is also customary to use combined meshes made up of rectangular elements for the interior points and triangular elements for the boundaries, wherever the degree of accuracy permits such a selection. The above discussion is well applicable to higher dimensional problems. The tensor product of one-dimensional linear shape functions usually generates multi-dimensional shape functions, such as bilinear shape functions. Higher order polynomial shape functions are constructed in a similar fashion.

The exponential fitting method involves the inclusion of the sophisticated functions, such as exponential functions, in the definition of the shape functions, so that the shape functions are able to capture the variations of the problem. However, as we have seen earlier, a systematic derivation of these functions at the element level requires the analytical solution of a local differential equation subject to homogeneous boundary conditions (residual free bubble functions). Solving such local problems analytically becomes tedious and far from trivial in multi-dimensional spaces. An alternative approach to get around this difficulty is to solve an analogous ordinary differential equation and then use the tensor product to generate the multi-dimensional exponential fitting. However, as we have already seen in section (2.7), the multiplicative model may or may not satisfy the original PDE and its boundary conditions. Solution of practical multi-scale problems addressed in [28] resort to the Taylor series approximation of the local residual free bubbles. The presented method

therein, is not residual free anymore; however, it owns the merit of practicality towards the solution of ill-natured problems. This approach shows that in order to obtain practical schemes with satisfactory solutions, it is possible to adopt a polynomial approximation of the residual free bubble functions combined with reasonably refined meshes. This, in turn, gives rise to the idea of approximating the residual free bubble functions with an optimal polynomial of a prescribed degree.

Using the method of least squares minimization, is therefore a natural selection when it comes to an optimal polynomial approximation. As we have seen in previous Chapters, the least squares polynomial approximation of the residual free bubble functions not only offers a practical solution to the problems under study, but also it yields the optimal polynomial approximation from all square integrable functions in an interval. The selection of such polynomial bubble functions in one dimension required the polynomial to vanish at the element boundary. The computations, cost less if the trial polynomials are selected from an orthogonal set of functions. The selection of multi-dimensional orthogonal families of polynomials is not as straightforward as in the one-dimensional case. However, the tensor product of the one-dimensional functions readily offers the multi-dimensional trial polynomials. For example, consider a typical two-dimensional element $\Omega_e = [0,l]\times[0,h]$ with the nodes situated on $(0,0),(l,0),(0,h),(l,h)$. The one-dimensional polynomial bubble functions in each direction are:

$$P_e(x)=\sum_{n=1}^{N} a_n x^n (l-x) \text{ and } Q_e(y)=\sum_{k=1}^{M} b_k y^k (h-y)$$

where $\{a_n\},\{b_k\}$ are found in the exercise of the least squares minimization. The two-dimensional trial polynomial in form of a complete polynomial is defined as:

$B_e(x,y)=\sum_{n=1}^{N}\sum_{k=1}^{M}a_n b_k x^n(l-x)y^k(h-y)$. Incomplete polynomials in terms of two variables x and y can be used providing that they vanish at element boundary. The numbers M and N are selected in such a way that the required resolution in x and y directions is maintained.

4.5 Derivation of least squares bubble functions for multi-dimensional problems: rectangular and triangular elements

From the discussion presented in section (4.4) we are able to draw conclusions towards the derivation of polynomial bubble functions for multi-dimensional problems. Two approaches seem possible: direct derivation of the bubble functions by the method of least squares and using the tensor product of the one dimensional least squares bubble functions. Here, we shall briefly discuss two methods and derive the bubble functions accordingly. Let us consider first, the idea of direct derivation of the bubble functions by the method of least squares.

We present the technique in the context of an example, though, this technique can be generalized to some extent to other cases. Consider the following two-dimensional boundary value problem:

$$\begin{cases} \dfrac{\partial^2 u}{\partial x^2}=\dfrac{\partial u}{\partial y} & \\ u(x,t)=f(x) & 0\le x\le l, t=0 \\ u(x,t)=0 & x=0, t\ge 0 \\ u(x,t)=0 & x=l, t\ge 0 \end{cases} \tag{4.56}$$

and let the finite element discretization of the equation domain in the x and y directions be $0=x_0<x_1<\ldots<x_n=a_x$ and $0=y_0<y_1<\ldots<y_m=b_y$, respectively. Also, consider a typical standard rectangular element of the mesh $(x,y)\in[0,l]\times[0,h]$

and let the standard two-dimensional approximation be defined in the usual way:

$\hat{u}(x,y) = \frac{l-x}{l}\frac{h-y}{h}u(0,0) + \frac{l-x}{l}\frac{y}{h}u(0,h) + \frac{x}{l}\frac{h-y}{h}u(l,0) + \frac{x}{l}\frac{y}{h}u(l,h)$ Similar to the one-dimensional case we add the bubble function contribution in such a way that the additional term vanishes at the element boundaries and its effects are optimal in the sense of generating least approximation error:

$$\begin{aligned}\hat{u}(x,y) = &\frac{l-x}{l}\frac{h-y}{h}u(0,0) + \frac{l-x}{l}\frac{y}{h}u(0,h) \\ &+ \frac{x}{l}\frac{h-y}{h}u(l,0) + \frac{x}{l}\frac{y}{h}u(l,h) + c\,x\,y(l-x)(h-y)\end{aligned} \tag{4.57}$$

It should be noted that the selection of the additional term is not unique in this case and any suitable polynomial function that vanishes at the element boundaries can be chosen, depending on the required resolution in each space directions.

Replacement of the approximation (4.57) into equation (4.56) generates the approximation error: $R = \frac{\partial^2 \hat{u}}{\partial x^2} - \frac{\partial \hat{u}}{\partial y}$, for which we introduce the two-dimensional analogous of the residual functional defined by equation (3.22) in Chapter 3 for one-dimensional cases:

$$J = \int_0^l \int_0^h R^2 dy\,dx. \tag{4.58}$$

Minimization of this residual functional within the limits of selected element and with respect to the unknown parameter c (i.e. $\frac{\partial J}{\partial c} = 0$), yields the bilinear bubble coefficient:

$$c = \frac{15\{u(0,0) - u(0,h) + u(l,0) - u(l,h)\}}{h\{l^4 + 12h^2\}}. \tag{4.59}$$

Higher order polynomial bubble functions can also be computed in a straightforward fashion, similarly. Equation (4.59) shows that the bubble function coefficient (and

hence the bubble function itself) can be written in a combination of four other elemental nodes that makes possible the process of static condensation.

In the example just given, we demonstrated the straightforward procedure of derivation of least squares quadratic bubble functions for standard rectangular elements. However, it is worthwhile to consider the eventualities in which the selected element is a triangle in the two-dimensional plane. As it will be shown, this process only slightly differs from the rectangular case. Employing the notations for our purpose, we recall that the standard triangular element is the area of all pairs $(x, y) \in [0,l]\times[0,h]$ subject to the constraint of $\frac{x}{l}+\frac{y}{h} \le 1$. In such a case, the standard approximation is given by $\hat{u}(x,y) = \frac{x}{l}u(l,0)+\frac{y}{h}u(0,h)+(1-\frac{x}{l}-\frac{y}{h})u(0,0)$. The simplest quadratic bubble function can therefore be augmented in the following way:

$$\hat{u}_T(x,y) = \frac{x}{l}.u(l,0)+\frac{y}{h}.u(0,h)+(1-\frac{x}{l}-\frac{y}{h}).u(0,0)+cxy(1-\frac{x}{l}-\frac{y}{h}) \qquad (4.60)$$

with the intention to find the optimal value of c, the bubble coefficient, by which we seek to reduce the approximation error within the element level. Substitution of the approximation (4.60) in the original differential equation (4.56) $R_T = \frac{\partial^2 \hat{u}_T}{\partial x^2} - \frac{\partial \hat{u}_T}{\partial y}$ results in the following residual functional:

$$J_T = \int_0^l \int_0^{h(1-\frac{x}{l})} R_T^2 \, dx\, dy \qquad (4.61)$$

to be minimized. Differentiating from (4.61) with respect to unknown parameter c and making the result identical to zero, we find the solution:

$$c = \frac{-60l\{u(0,h)-u(0,0)\}}{l^4-6hl^2+60h^2} \qquad (4.62)$$

which represents the bubble function coefficient corresponding to the particular selection of approximation (4.60).

The second approach that we will shortly describe, provides an alternative to the derivation of the polynomial bubble functions based on the technique proposed in [28]. We will briefly outline their approach towards the derivation of practical bubble functions and will introduce our technique afterwards.

To this end, consider the steady state diffusion-reaction equation in $\Omega \subset R^2$ written in dimensionless form as: $\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial x^2}-D_a u = f$, subject to the prescribed boundary conditions. The approach adopted for the derivation of practical bubble functions requires the analogous one-dimensional equations:

$$\begin{cases} \dfrac{d^2 N_1}{dx^2} - D_a N_1 = 0 \quad for\, x \in [0,l] \\ N_1(0) = 1,\ N_1(l) = 0 \end{cases}$$

and (4.63)

$$\begin{cases} \dfrac{d^2 N_2}{dx^2} - D_a N_2 = 0 \quad for\, x \in [0,l] \\ N_2(0) = 0,\ N_2(l) = 1 \end{cases}$$

to be solved at the element level, in order to extract the bubble shape functions. With the characteristic element length l, the solution of the above equation is:

$$\begin{cases} N_1(x) = \dfrac{\sinh(\sqrt{D_a}(l-x))}{\sinh(\sqrt{D_a}l)} \\ N_2(x) = \dfrac{\sinh(\sqrt{D_a}x)}{\sinh(\sqrt{D_a}l)} \end{cases} \tag{4.64}$$

In order to turn the elemental bubble functions into a practical format, the Taylor series expansion of the bubble functions are used and truncated after a selected number of terms. Hence, bubble functions are converted into polynomials, suitable for use in quadrature methods in finite element programs. Finally after the truncation,

bubble functions are derived expressed as polynomials, for example the third order polynomial bubble functions are:

$$\begin{cases} N_1(x) = \dfrac{l-x}{l} - \dfrac{x(l-x)(2l-x)}{l(\dfrac{6}{D_a}+l^2)} \\ N_2(x) = \dfrac{x}{l} - \dfrac{x(l-x)(l+x)}{l(\dfrac{6}{D_a}+l^2)} \end{cases} \tag{4.65}$$

Recalling the set of equations (4.63), our aim is to approximate the solution of this system with optimal least squares fitted polynomials (i.e. polynomials satisfying the best approximation property). Carrying out such an approximation, enables us to derive practical bubble functions expressed as polynomials and yet avoid from solving directly the set of equations (4.63). In order for the future comparisons, we will derive the third order least squares approximations to the solution of (4.63). Consider the following third order approximations:

$$\begin{cases} N_1(x) = \dfrac{l-x}{l} + \alpha x(l-x) + \beta x^2(l-x) \\ N_2(x) = \dfrac{x}{l} + \lambda x(l-x) + \mu x^2(l-x) \end{cases} \tag{4.66}$$

By replacing these approximations into the system (4.63) and minimizing the residual functionals by the method of least squares the following approximations are obtained:

$$\begin{cases} N_1(x) = \dfrac{l-x}{l} + \alpha x(l-x) + \beta x^2(l-x) \\ \alpha = -\dfrac{2D_a(3D_a^3l^6 + 260D_a^2l^4 + 6720D_al^2 + 50400)}{(D_a^2l^4 + 120 + 20D_al^2)(D_a^2l^4 + 2520 + 84D_al^2)} \\ \beta = \dfrac{7D_a(60 + D_al^2)}{l(D_a^2l^4 + 2520 + 84D_al^2)} \end{cases} \tag{4.67}$$

and

$$\begin{cases} N_2(x) = \dfrac{x}{l} + \lambda x(l-x) + \mu x^2(l-x) \\ \lambda = -\dfrac{D_a(D_a^3 l^6 + 40D_a^2 l^4 - 4200D_a l^2 - 50400)}{(D_a^2 l^4 + 120 + 20D_a l^2)(D_a^2 l^4 + 2520 + 84D_a l^2)} \\ \mu = -\dfrac{7D_a(60 + D_a l^2)}{l(D_a^2 l^4 + 2520 + 84D_a l^2)} \end{cases} \tag{4.68}$$

Higher order polynomial bubble functions are derived similarly. Multi-dimensional bubble functions, can be derived using the tensor product of one-dimensional functions, if the second approach is taken for derivation of the bubble functions.

4.6 A benchmark problem

We start with the analytical solution of a two dimensional convection-diffusion problem given in terms of the following PDE [22]:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} - k_1 \frac{\partial \phi}{\partial x} - k_2 \frac{\partial \phi}{\partial y} = 0 \tag{4.69}$$

where $\phi(x , y) \in \Re = [0,1] \times [0,1]$

and $k_1, k_2 >> 0$

Equation (4.69) should be solved subject to the boundary conditions of :

$\phi(0, y) = \phi(x, 0) = 1 \ (x , y \neq 1)$

$\phi(1, y) = \phi(x, 1) = 0 \ (x , y \neq 0)$

$\phi(0, 1) = \phi(1, 0) = ½$

Suppose that:

$$\phi(x, y) = X(x)Y(y) \tag{4.70}$$

Substituting equation (4.69) into equation (4.70) gives

$$\frac{X'' - k_1 X'}{X} = -\frac{Y'' + k_2 Y'}{Y} \tag{4.71}$$

which gives:

$$X''-k_1X'-MX=0 \tag{4.72}$$

$$Y''-k_2Y'-MY=0 \tag{4.73}$$

where M is a constant value.

These ordinary differential equations give the following

$$X(x)=e^{\frac{k_1x}{2}}\left[Ae^{\sqrt{k_1^2+4M}\,x/2}+Be^{-\sqrt{k_1^2+4M}\,x/2}\right] \tag{4.74}$$

$$Y(y)=e^{\frac{k_2y}{2}}\left[Ce^{\sqrt{k_2^2+4M}\,y/2}+De^{-\sqrt{k_2^2+4M}\,y/2}\right] \tag{4.75}$$

(i) $\phi = 0$

$\phi = 0$ | $\phi = 0$

$\phi = 1$

$\phi(x, 0) = 1 \quad 0 < x < 1$

$\phi(x, 1) = \phi(0, y) = \phi(1, y) = 0$

(ii) $\phi = 0$

$\phi = 1$ | $\phi = 0$

$\phi = 0$

$\phi(0, y) = 1 \quad 0 < x < 1$

$\phi(x, 0) = \phi(x, 1) = \phi(1, y) = 0$

The solution of (4.69) excluding the points (0,0), (0,1) (1,0) which are dealt with separately is given as the sum of the solutions of (i) and (ii).

Consider case (i) first – we have

$$X(0) = X(1) = 0 \tag{4.76}$$

The form of the solution in equation (4.74) is dictated by the sign of k_1^2+4M

If k_1^2+4M is positive we have

$$X(x)=e^{\frac{k_1x}{2}}\left[(A+B)\cosh\frac{\sqrt{k_1^2+4M}}{2}x+(A-B)\sinh\frac{\sqrt{k_1^2+4M}}{2}x\right] \tag{4.77}$$

from equation (4.76) X(0) = 0 gives A+B=0 and X(1) = 0 gives $\sinh\frac{\sqrt{k_1^2+4M}}{2}x=0$

in this case A–B=0 for non-trivial solution therefore in this case it is impossible to give any real answers.

If k_1^2+4M is negative we have from equation (4.74)

$$X(x)=e^{\frac{k_1x}{2}}\left[Ae^{i\sqrt{-k_1^2-4M}\,x/2}+Be^{-i\sqrt{-k_1^2-4M}\,x/2}\right] \tag{4.78}$$

or

$$X(x)=e^{\frac{k_1x}{2}}\left[(A+B)\cos\left(\sqrt{-k_1^2-4M}\,x/2\right)+(A-B)i\sin\left(\sqrt{-k_1^2-4M}\,x/2\right)\right] \tag{4.79}$$

From X(0) = 0, A+B=0 and X(1)=0 gives

$$Sin\left(\sqrt{-k_1^2-4M}\,x/2\right)=0$$

or $\frac{1}{2}\sqrt{-k_1^2-4M}=n\pi$ n = 1, 2, ... (4.80)

i.e. $M=-\frac{1}{4}(k_1^2-4n^2\pi^2)$

and we have

$$X(x)=(A-B)ie^{\frac{k_1x}{2}}\sin(n\pi x) \tag{4.81}$$

and $k_1^2+4M<0$ is the only possible case. We define

$$An=k_2^2-4M=k_1^2-k_2^2+4n^2\pi^2>0 \tag{4.82}$$

Similarly for constants C and D equation (4.75) gives

$$Y(y)=e^{\frac{k_2y}{2}}\left[C'\cosh\left(\sqrt{An}\,(1-y)/2\right)+D'\sinh\left(\sqrt{An}\,(1-y)/2\right)\right] \tag{4.83}$$

Using Y(1) = 0 gives C' = 0 thus

$$Y(y)=D'e^{\frac{k_2y}{2}}\sinh\left(\sqrt{An}\,(1-y)/2\right) \tag{4.84}$$

We do not require any condition on Y(0) since Y(0) = 0 ensures $\phi(0,0) = 0$. Thus for constant α

$$\phi_1(x,y) = \alpha e^{\frac{(k_1x+k_2y)}{2}} \sin(\pi x)\sinh\left(\sqrt{An}\,(1-y)/2\right) \tag{4.85}$$

The most general solution for boundary conditions given by (4.70) is a linear combination of the particular solution i.e.

$$\phi_1(x,y) = \sum_{n=1}^{\infty} C_n e^{(k_1x+k_2y)/2} \sin(\pi x)\sinh\left(\sqrt{An}\,(1-y)/2\right) \tag{4.86}$$

where the coefficients C_n (n=1,2,...) are determined from $\phi(x,0) = 1$ using orthogonal functions and Fourier Series techniques. Putting y=0 and $b_n = C_n \sinh\left(\sqrt{An}/2\right)$ we obtain:

$$\phi_1(x,0) = \sum_{n=1}^{\infty} b_n e^{\frac{k_1x}{2}} \sin(\pi x) = 1 \tag{4.87}$$

functions $e^{\frac{k_1x}{2}} \sin(n\pi x)$ and $e^{\frac{k_1x}{2}} \sin(m\pi x)$ are orthogonal with respect to e^{-k_1x} (positive weighting). Thus simplifying equation (4.87) by $e^{\frac{k_1x}{2}} \sin(m\pi x)$, weighting by e^{-k_1x} and integrating over intervals $0 \rightarrow 1$ gives

$$\sum b_n \int_0^1 e^{-k_1x} e^{\frac{k_1x}{2}} \sin(\pi x) e^{\frac{k_1x}{2}} \sin(m\pi x)dx = \int_0^1 1.e^{-k_1x} e^{\frac{k_1x}{2}} \sin(m\pi x)dx \tag{4.88}$$

this term on the left hand side of equation (4.88) except for m=n.

And

$$b_n \int_0^1 \sin^2(n\pi x)dx = \int_0^1 e^{\frac{k_1x}{2}} \sin(n\pi x)dx \tag{4.89}$$

$$b_n = \frac{\int_0^1 e^{\frac{k_1x}{2}} \sin(n\pi x)dx}{\int_0^1 \sin^2(n\pi x)dx} = \frac{8n\pi\left[1-(-1)^n e^{-k_1/2}\right]}{k_1^2 + 4n^2\pi^2} \tag{4.90}$$

and equation (4.86) becomes

$$\phi_1(x,y)=\sum_{n=1}^{\infty}\frac{8n\pi\left[1-(-1)^n e^{-k_1/2}\right]e^{(k_1x+k_2y)/2}\sin(\pi x)\sinh\left(\sqrt{An}\,(1-y)/2\right)}{k_1^2+4n^2\pi^2\sinh\left(\sqrt{An}/2\right)} \tag{4.91}$$

in an identical manner we obtain $\phi_2(x,y)$ the solution for case (ii). The general solution of equation (4.69) is then the sum of ϕ_1 and ϕ_2 plus the values of ϕ at the points (0,0), (0,1) and (1,0) and is given by

$\phi(0,0) = 1$, $\phi(0,1) = \phi(1,0) = ½$

$$\phi(x,y)=\sum_{n=1}^{\infty}\frac{8n\pi e^{(k_1x+k_2y)/2}}{\sinh\left(\sqrt{An}/2\right)}\left[\begin{array}{l}\left(\frac{\left[1-(-1)^n e^{-k_1/2}\right]}{k_1^2+4n^2\pi^2}\right)\sin(\pi x)\sinh\left(\sqrt{An}\,(1-y)/2\right)+\\ \left(\frac{\left[1-(-1)^n e^{-k_2/2}\right]}{k_1^2+4n^2\pi^2}\right)\sin(\pi y)\sinh\left(\sqrt{An}\,(1-x)/2\right)\end{array}\right] \tag{4.92}$$

Boundary conditions at points $\phi(0,0) = 1$, $\phi(0,1) = \phi(1,0) = ½$ must be fixed and k_1 is small

$$\sum_{1}^{\infty}\frac{f_n\frac{n\pi}{2}}{\frac{n\pi}{2}}e^{\frac{k_1}{4}}\left(1+e^{-\frac{k_1}{2}}\right)$$

$$\phi(x,0)=\sum_{n=1}^{\infty}\frac{8n\pi e^{\frac{k_1x}{2}}\left[1-(-1)^n e^{-\frac{k_1}{2}}\right]\sin(\pi x)}{k_1^2+4n^2\pi^2},\phi(0,y)=\sum_{n=1}^{\infty}\frac{8n\pi e^{\frac{k_2y}{2}}\left[1-(-1)^n e^{-\frac{k_2}{2}}\right]\sin(\pi y)}{k_2^2+4n^2\pi^2}$$

$$\sum_{n=1}^{\infty}\frac{n\sin(\pi x)}{k_2^2+4n^2\pi^2} \tag{4.93}$$

If the PDE has a term such as Q included in the RHS of equation (4.69) a complete solution will require the addition of the below source term to equation (4.93).

$$-\frac{1}{2}\left(\frac{x}{k_1}+\frac{y}{k_2}\right)Q.$$

4.7 Least squares bubble functions for convection-diffusion problem

A different solution for the problem represented by equation (4.69) can be obtained using a domain discretization. This has the advantage that it can be used to obtain solutions under more general boundary conditions than the specific conditions that were used to generate the previous results. The solution can also be extended to irregular domains. However we will use a square domain and boundary conditions similar to those given previously to be able to directly compare these solutions with the analytical results.

We shall derive the least squares bubble functions for the use in finite elements approximation by setting $k_1 = k_2$. Therefore, equation (4.69) becomes:

$$\begin{cases} \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} - k\left(\dfrac{\partial u}{\partial x} + \dfrac{\partial u}{\partial y}\right) = 0 \\ u(0, y) = u(x,0) = 1 \quad x \neq 1, y \neq 1 \\ u(1, y) = u(x,1) = 0 \quad x \neq 0, y \neq 0 \\ u(0,1) = u(1,0) = \dfrac{1}{2} \end{cases} \tag{4.94}$$

Consider the finite element discretization of the problem domain in the x and y directions with rectangular elements:

$0 = x_0 < x_1 < \ldots < x_n = 1, 0 = y_0 < y_1 < \ldots < y_m = 1$ where within a typical element we have $(x, y) \in [0,l] \times [0,h]$. Let us consider the standard linear approximation shape functions enriched with the bubble function given by:

$$\begin{aligned} \hat{u}(x, y) = & \frac{x}{l}\frac{y}{h}u(l,h) + \frac{l-x}{l}\frac{y}{h}u(0,h) \\ & + \frac{x}{l}\frac{h-y}{h}u(l,0) + \frac{l-x}{l}\frac{h-y}{h}u(0,0) + cxy(l-x)(h-y) \end{aligned} \tag{4.95}$$

The residual generated from replacement of approximation (4.95) into equation (4.94) and the functional to be minimized with respect to unknown bubble function coefficient are:

$$R = \frac{\partial^2 \hat{u}}{\partial x^2} + \frac{\partial^2 \hat{u}}{\partial y^2} - k\left(\frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{u}}{\partial y}\right)$$

and

$$J = \int_0^l \int_0^h R^2 dy dx$$

respectively. Minimizing J to find the bubble coefficient results in:

$$\begin{aligned} c = & \frac{5k}{l\,h(20h^2l^2 + 12(l^4 + h^4) + h^4l^2k^2 + h^2l^4k^2)} \cdot \\ & \{(3l^3 + 3h^3 + 3l^2 + kl^2h^2 + 3lh^2)u(0,0) \\ & + (-3l^3 + 3h^3 + 3l^2 - kl^2h^2 - 3lh^2)u(0,h) + (3l^3 - 3h^3 - 3l^2 - kl^2h^2 + 3lh^2)u(l,0) \\ & + (-3l^3 - 3h^3 - 3l^2 + kl^2h^2 - 3lh^2)u(l,h)\ \} \end{aligned} \tag{4.96}$$

We are interested to solve the equation (4.94) by setting $k=2.5$ and $h=l=0.2$ for a 5x5 finite element mesh made of equal sized rectangles. We will find the nodal values by use of linear and bubble enriched elements and compare the results against the analytical solution. The bubble coefficient in this case becomes:

$$c = \frac{78125}{178}u(0,0) - \frac{71875}{178}u(l,h) - \frac{3125}{178}u(0,h) - \frac{3125}{178}u(l,0) \tag{4.97}$$

and the elemental bubble enriched approximation becomes:

$$\begin{aligned} \hat{u}(x,t) = & \{25 - \frac{78125}{178}(0.2-x)(0.2-y)\}.x.y.u(l,h) \\ & + \{25 - \frac{3125}{178}x(0.2-y)\}.y.(0.2-x)u(0,h) \\ & + \{25 - \frac{78125}{178}y(0.2-x)\}.x.(0.2-y)u(l,0) \\ & + \{25 + \frac{78125}{178}x\,y\}(0.2-x)(0.2-y)\}u(0,0). \end{aligned} \tag{4.98}$$

Note that for the sake of simplicity, we have kept the notations indicating operation of function u on the nodal points (i.e. *u(l,h), u(l,0)*, etc).

The nodal values $u(l,h), u(0,h), u(l,0), u(0,0)$ for each element should be calculated via the application of weighted residual technique and from given boundary conditions. The weak formulation of equation (4.94) is:

$$0 = \int_0^l \int_0^h w_j \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - k\left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\right)\right) dy dx \tag{4.99}$$

Application of Green's integral theorem in this case, converts the weak formulation into the following reduced form:

$$\int_0^l \int_0^h \left(\frac{\partial w_j}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial w_j}{\partial y}\frac{\partial u}{\partial y}\right) dy dx + k\int_0^l \int_0^h w_j \left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}\right)\right) dy dx = \int_\Gamma w_j \left(n_x \frac{\partial u}{\partial x} + n_y \frac{\partial u}{\partial y}\right) ds$$

where n_x, n_y, ds are the components of the unit normal vector $n = n_x \vec{i} + n_y \vec{j}$ on the boundary Γ^e, and ds is the arc-length of an infinitesimal line element along the boundary. Let us select the weight functions:

$$w_1 = \frac{x}{l}\frac{y}{h}, \quad w_2 = \frac{l-x}{l}\frac{y}{h}, \quad w_3 = \frac{x}{l}\frac{h-y}{h}, \quad w_4 = \frac{l-x}{l}\frac{h-y}{h}$$

identical to the linear (two-dimensional)shape functions. The weak formulation (4.99) with the exercise of Green's integral theorem result in the local element matrices as:

$$\begin{bmatrix} \frac{19349}{32040} & \frac{-4811}{32040} & \frac{-4811}{32040} & \frac{-9727}{32040} \\ \frac{-13}{60} & \frac{2}{3} & \frac{-1}{3} & \frac{-7}{60} \\ \frac{-13}{60} & \frac{-1}{3} & \frac{2}{3} & \frac{-7}{60} \\ \frac{-11873}{32040} & \frac{-5869}{32040} & \frac{-5869}{32040} & \frac{23611}{32040} \end{bmatrix} \begin{bmatrix} u(0,0) \\ u(0,h) \\ u(l,0) \\ u(l,h) \end{bmatrix} = \begin{bmatrix} \phi_I \\ \phi_{II} \\ \phi_{III} \\ \phi_{IV} \end{bmatrix} \tag{4.100}$$

It can be observed that we have obtained a *4×4* matrix corresponding to every node of the each rectangular element. Performing the assembly of local matrices is a tedious task if carried out by hand specially for large number of elements. However their evaluation can be easily accomplished using a computer program. With k= 2.5 the convection terms are comparatively small and the linear interpolation and the bubble

enriched elements obtain reasonable solutions as compared to the exact analytical results. It is a well known fact that, however, for higher values of k the linear and classical Galerkin treatments fail. The least squares bubble functions have the property of minimizing the approximation error in a uniform sense. Therefore, with the use of a relatively crude mesh in multi-dimensional problems low order least squares bubble functions (e.g. quadratic) turn to an awkward approximation. However, the least squares bubble functions prove to be very effective tools and superior to the standard approximations when employed in higher degrees (e.g. cubic, quintic, etc) and the computational mesh is not coarse. In such an environment, the least squares bubble functions act in a hierarchical way meaning that the contribution of very high order polynomials is of diminishing importance. The following tables and charts compare the performance of the linear standard and least squares bubble function approximations against the exact solution, for the above two-dimensional example.

x=0.05	*u(x,y)*	*u(x,y)*	*u(x,y)*
y	***Linear***	***Bubble***	***Exact***
0	1	1	1
0.05	0.8961721371	0.8865448553	0.886754727
0.15	0.8796370948	0.8742811656	0.87419473
0.25	0.8631020526	0.8595461702	0.859597615
0.35	0.8465670104	0.8423398692	0.842622801
0.45	0.8300319682	0.8226622625	0.822818793
0.55	0.8008270582	0.7995354646	0.799592276
0.65	0.7630845854	0.7734094675	0.772156048
0.75	0.7253421125	0.7398308440	0.73943447
0.85	0.6875996397	0.6987995942	0.699870348
0.95	0.6498571669	0.6503157180	0.650959086
1	0	0	0

Table 4.4
Solution profile at $x = 0.05$

x=0.95	*u(x,y)*	*u(x,y)*	*u(x,y)*
y	*Linear*	*Bubble*	*Exact*
0	1	1	1
0.05	0.6803819251	0.6802809517	0.680628372
0.15	0.6543645168	0.6545155494	0.654958596
0.25	0.6283471085	0.6295393891	0.630975318
0.35	0.6023297002	0.6053524710	0.60688213
0.45	0.5763122919	0.5819547950	0.581648018
0.55	0.5456988321	0.5544406338	0.554596691
0.65	0.5144204044	0.5247536262	0.525262075
0.75	0.4831419768	0.4926842798	0.493335749
0.85	0.4518635491	0.4582325946	0.458651484
0.95	0.4205851215	0.4213985707	0.42118422
1	0	0	0

Table 4.5

Solution profile at $x = 0.95$

y=0.95	*u(x,y)*	*u(x,y)*	*u(x,y)*
x	*Linear*	*Bubble*	*Exact*
0	1	1	1
0.05	0.6498571669	0.6503157180	0.650959086
0.15	0.626087601	0.6232363536	0.623430518
0.25	0.602318034	0.5983092201	0.599241791
0.35	0.578548468	0.5755343174	0.576774527
0.45	0.554778902	0.5549116455	0.554936298
0.55	0.529141098	0.5326911864	0.532802638
0.65	0.502002104	0.5099444765	0.509456763
0.75	0.47486311	0.4838134705	0.483892762
0.85	0.4477241155	0.4542981686	0.454938976
0.95	0.4205851215	0.4213985707	0.42118422
1	0	0	0

Table 4.6

Solution profile at $y = 0.95$

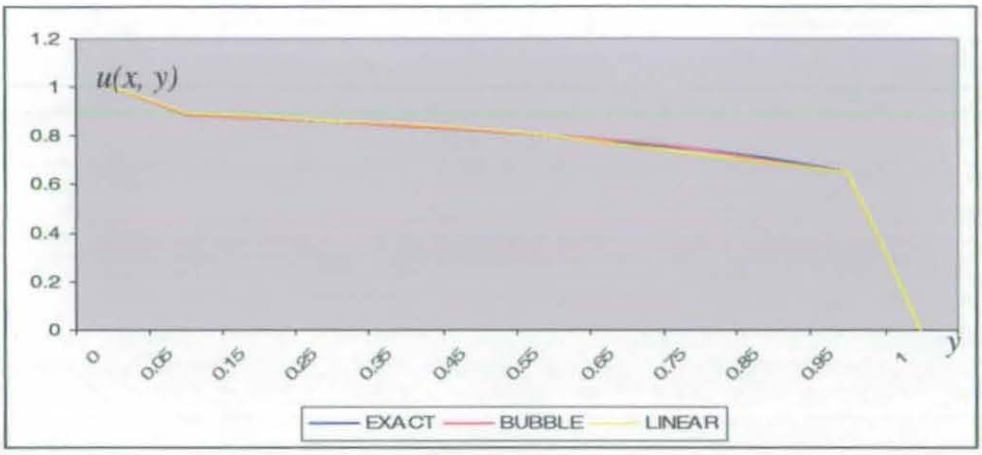


Figure 4.5
Solution profile at $x = 0.05$

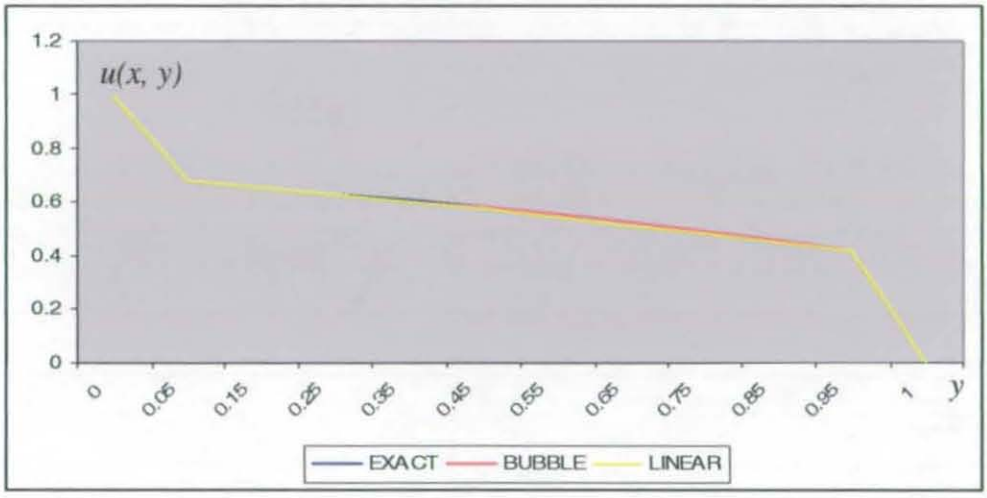


Figure 4.6
Solution profile at $x = 0.95$

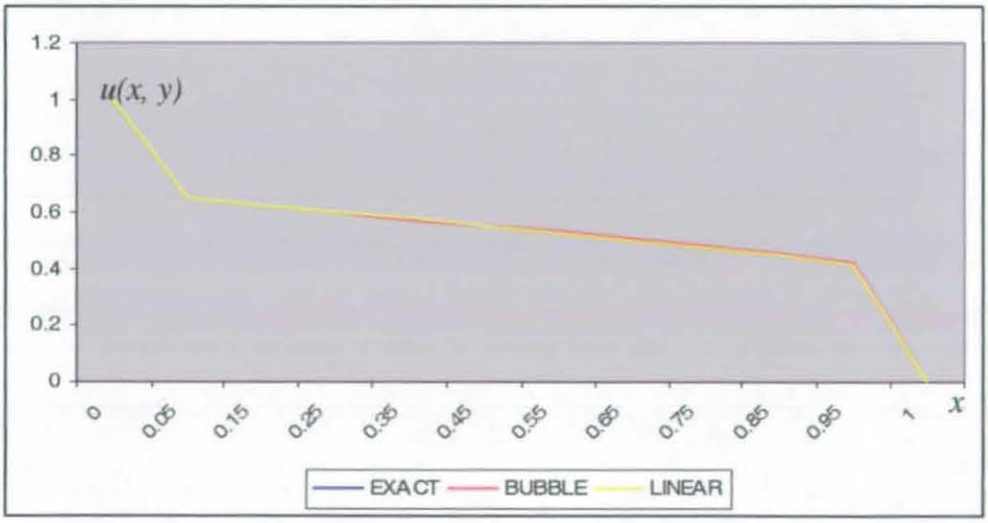


Figure 4.7
Solution profile at $y = 0.95$

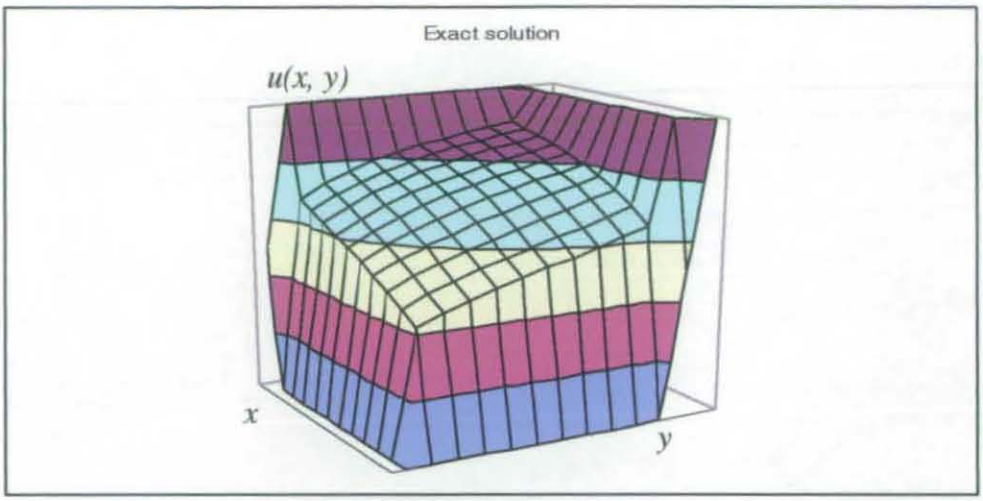


Figure 4.8
Exact solution of problem (4.94)

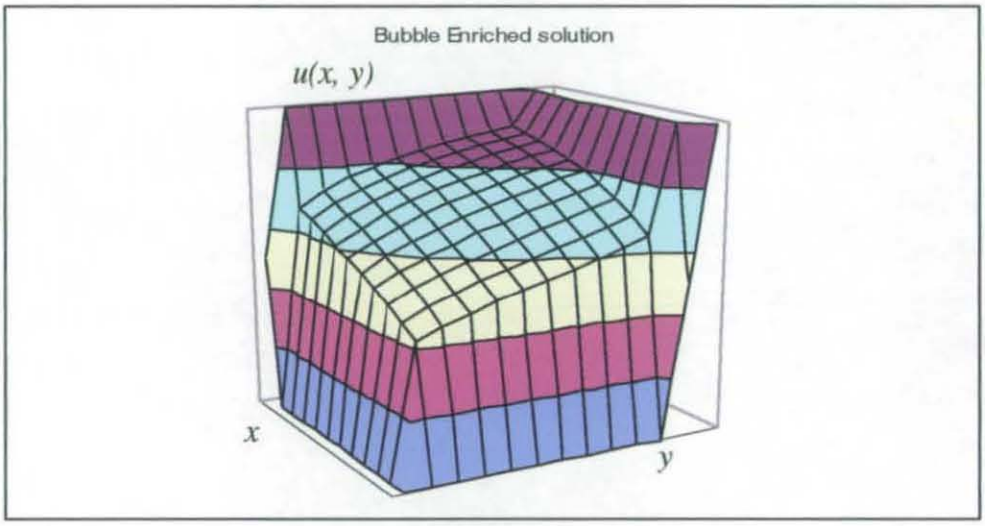


Figure 4.9
Bubble-enriched solution of problem (4.94)

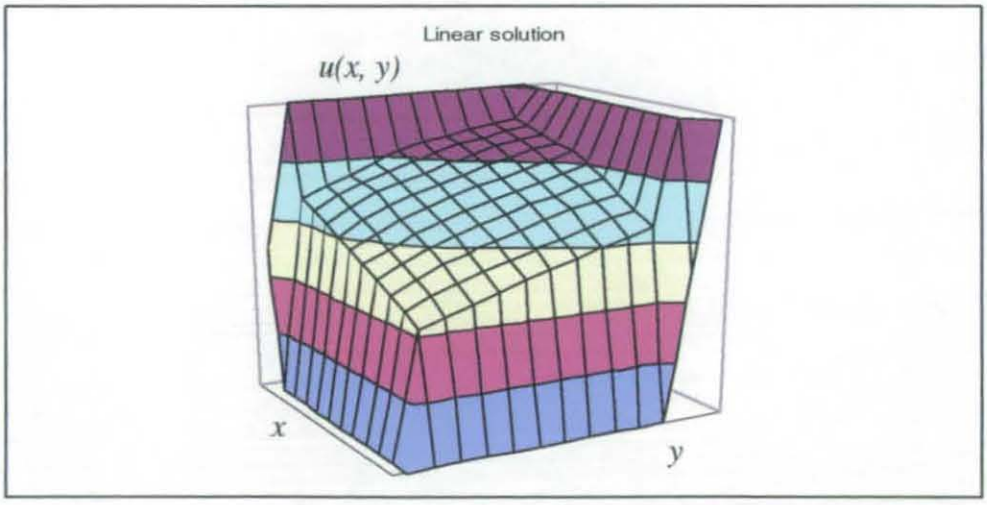


Figure 4.10
Linear finite element solution of problem (4.94)

Chapter 5

Conclusions and suggestions for future research

In this Chapter the main objectives, motivations and conclusions drawn from the present research are explained and discussed. Furthermore, potentials and possibilities for the continuation and extension of this work are highlighted.

5.1 Conclusions

In this thesis, a novel bubble function scheme for the finite element solution of engineering flow problems is developed and successfully applied to a range of problems encountered in the modelling of engineering processes. This scheme is based on the idea of the residual free bubble functions that is widely used towards the study and modelling of multi-scale problems. The scheme results in a feasible alternative for existing sophisticated techniques and acquires good degree of accuracy as well as computational time effectiveness. Theoretical multi-scale methods are not broadly applied to practical problems and implementation of these methods are hardly achieved. The main burden on these methods is their excessive refinement of computational mesh which turns to impracticality due to the required amount of time. The main aim in this research is to find (or at least get closer to) a point of equilibrium of the approaches using h-version finite elements and p-version finite elements. The present technique, incorporates the ideas of polynomial approximation with the method of least squares, the Galerkin finite element technique and the bubble enriched finite elements to develop an optimal practical enriched scheme. Finding a polynomial

fit for the elemental bubble function with the method of least squares error minimization, enables us to increase the accuracy of approximation without incurring excessive mesh refinement. On the other hand, the error analysis of the least squares polynomial fits, provides us with an estimation on the highness of the degree of approximant polynomial. This selection informs us of the potentially over-estimations such as over-smooth solutions or super convergent results.

In this work a novel technique for derivation of practical bubble functions, expressed as polynomials, is developed and applied to a range of flow problems including the problems belonging to the family of diffusion-reaction. Along with the progress of this work, the following conclusions are drawn:

- As far as the issues of practicality and applicability in finite element packages are concerned, polynomials are natural selections to approximate the residual free bubble functions.
- Since all the hyperbolic, exponential and other smooth transcendental functions can be approximated by the families of polynomials, construction of such families with higher convergence rate will produce sufficiently satisfactory approximation results.
- The generated error of replacing the approximant into the model equation, can be minimized in a uniform sense by the use of least squares minimization.
- Higher order least squares polynomial bubble functions can be derived with no major theoretical or practical difficulty.
- The least squares bubble functions, can be equally used for approximation of the solution of equations that are usually expressed as special functions.

- With the use of least squares bubble functions, the direct solution of local residual differential equations are avoided, which is a major benefit in multi-dimensional equations.

- Extension of the idea to multi dimensional problems is possible as direct least squares derivation of bubble functions or tensor product of the analogous one-dimensional equations can be used to construct two or three dimensional elements.

- The idea is also applied to the transient problems and is found superior to the classical linear finite elements.

5.2 Future work

The idea of minimization of the approximation error based on the method of least squares is supported by the substantial theory of best approximations in function spaces. Construction of least squares bubble functions, on the other hand, is carried out in a straightforward fashion. There are numerous model equations in engineering problems that are categorized as multi-scale problems. To achieve a satisfactory practical numerical solution for these problems it is required that the technique to be tested for different categories of models including different types of ODE and PDEs (hyperbolic, elliptic, etc) and linear or nonlinear problems. This assessment will shed light on the abilities and limitations of the method of practical bubble functions. Performing a comprehensive error analysis in order to work out the span and rate of convergence of the schemes that are produced by least squares minimization, will be illustrative when compared to other numerical schemes. This can lead us to the clever selection of polynomial families weighted in an appropriate way to be employed as

the base bubble functions. Extension of the least squares bubble functions to three-dimensional problems and tetrahedral elements is to be investigated in possible future research projects.

REFERENCES

REFERENCES

[1] T. M. Apostol; Mathematical Analysis, 1965, Addison-Wesley.

[2] D. N. Arnold, F. Brezzi, M. Fortin; A stable finite element for Stokes equation, Calcolo, 23 (1984) 337-344.

[3] K. J. Bathe; Finite Element Procedures, 1996, Prentice-Hall, Englewood Cliffs, NJ.

[4] C. Baiocchi, F. Brezzi, L. Franca; Virtual bubbles and the Galerkin-least-square method, Comput. Methods Appl. Mech. Engrg., 105 (1993) 125-141.

[5] J. F. Botha, G. F. Pinder; Fundamental Concepts in the Numerical Solution of Differential Equations.

[6] F. Brezzi, A. Russo; Choosing Bubbles for Advection-Diffusion Problems, Mathematical Models and Methods in Applied Sciences, Vol. 4, (1994) 571-587.

[7] A. J. Chorin, J. E. Marsden; A Mathematical Introduction to Fluid Mechanics, 1990, Springer –Verlag.

[8] L. Collatz; The Numerical Treatment of differential Equations, 1960, Springer Verlag.

[9] P. J. Davis, P.Rabinowitz; Methods of Numerical Integration, 1984, Academic press.

[10] J. Donea, A. Huerta; Finite Element Method for Flow Problems, 2003, John Wiley & Sons Ltd.

[11] M. J. Fagan; Finite Element Analysis, 1992, Longman Scientific & Technical.

[12] L. P. Franca; On the limitations of bubble functions, Comput. Methods Appl. Mech. Engrg. 117 (1994) 225-230.

[13] L. P. Franca, A. Russo; Unlocking with residual-free bubbles, Comput. Methods Appl. Mech. Engrg. 142 (1997) 361-364.

[14] L. P. Franca, A. L. Madureira, L. Tobiska, and F. Valentin; Convergence Analysis of a Multiscale Finite Element Method for singularly Perturbed Problems, SIAM Journal of Multiscale Modelling and Simulation 4, No. 3, (2005) 839-866.

[15] C. F. Gerald, P. O. Wheatley; Applied Numerical Analysis, 1999, Addison – Wesley.

[16] D. Greenspan, V. Cassulli; Numerical Analysis for Applied Mathematics, Science, and Engineering; 1988, Addison-Wesley.

[17] R. W. Hamming; Numerical Methods for Science and Engineering, 1986, Dover Publications.

[18] T. J. R. Hughes; Multiscale phenomena; Green's function, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods, Comput. Methods Appl. Mech. Engrg., 127 (1995) 387-401.

[19] T. J. R. Hughes, G. R. Feijo'o, L. Mazzei, J. B. Quincy; The variational Multiscale method- a paradigm for computational mechanics; Comput. Methods Appl. Mech. Engrg., 166 (1998) 3-24.

[20] D. Henwood, J. Bonet; Finite Element, A Gentle Introduction; 1996, Macmillan.

[21] Wilfred Kaplan; Advanced Mathematics for Engineers, 1981, Addison-wesley.

[22] A. Khan; PhD Thesis, 2007, Loughborough University.

[23] E. Koffman, F. L. Friedman; Fortran with Engineering Applications; 1993, Addison-Wesley.

[24] P. Linz; Theoretical Numerical Analysis, 1979, Dover Publications.

[25] A. Masud, R. A. Khurram; A Multiscale/stabilized finite element method for the advection-diffusion equation, Comput. Methods Appl. Mech. Engrg., 193 (2004) 1997-2018.

[26] V. Nassehi; Practical Aspects of Finite Element Modelling of Polymer Processing, 2002, John Wiley & Sons.

[27] M. Parvazinia, V. Nassehi, R.J. Wakeman; Multiscale finite element modelling using bubble function method for a convection-diffusion problem, Chemical Engineering Science, 61 (2006) 2742-2751.

[28] M. Parvazinia, V. Nassehi; Multiscale finite element modelling of diffusion-reaction equation using bubble functions with bilinear and triangular elements, Comput. Methods Appl. Mech. Engrg., 196 (2007) 1095-1107.

[29] J. Petera, V. Nassehi, J. F. Pittman; Petrov Galerkin Methods on Isoparametric Bilinear and Biquadratic Elements Tested for A Scalar Convection- Diffusion Problem; Int. J. Meth. Heat Fluid Flow, Vol.3 (1993) 205-221.

[30] J. N. Reddy; An Introduction to the Finite Element Method, 1984, McGraw-Hill.

[31] W. Rudin; Principles of Mathematical Analysis,1976, McGraw-Hill.

[32] M. R. Spiegel; Theory and Problems of Vector Analysis, 1959, Schaum Publishing Company.

[33] H. Stroud; Approximate Calculation of Multiple Integrals; 1971, Prentice-Hall.

APPENDIX A

List of presentations and publications

Paper One

A novel method for the derivation of bubble functions using the Method of Least Squares

International Congress of Mathematicians, ICM 2006 Madrid, Spain

A.Yazdani*, V. Nassehi, R. J. Wakeman

Department of Chemical Engineering, Loughborough University,

Loughborough, Leicestershire, LE11 3TU, UK

A.Yazdani@lboro.ac.uk*

The multi-scale nature of certain differential equations and the absence of analytical solutions to majority of these equations have given impetus to the advent of the bubble enriched finite element techniques. These elements are mainly constructed by the residual free bubbles (RFB) originally proposed by Hughes et. al. [1] and recently extended to practical problems by Parvazinia et. al [2].

The RFB method is based on decomposing the solution of a model differential equation into two parts within each element. These parts represent coarse and fine phenomena, respectively. The standard piecewise linear component represents the coarse scale, while the analytical solution of the equation using homogeneous boundary conditions (i.e. residual-free bubble) captures the fine scale.

This method yields stabilising parameters as well as a high degree of accuracy in the solution of ill-conditioned differential equations. However, the required work to derive the bubble which involves analytical solution of the fine scale becomes cumbersome to say the least.

In the present work a novel method for the derivation of the bubble function is suggested which employs the least squares minimisation of the error generated by

substitution of a trial solution into the primary boundary value problem. This method avoids the use of an analytical solution and therefore it provides a technique for the extension of the bubble function method to more realistic problems.

The method is used to solve a number of boundary value problems and is shown to be accurate and reliable.

References

[1] T. J. R. Hughes, G. R. Feijo`o, L. Mazzei, J-B. Quincy, The variational multiscale method-a paradigm for computational mechanics, Comput. Methods Appl. Mech. Engrg. 166 (1998) 3-24.

[2] M. Parvazinia, V. Nassehi, R. J. Wakeman, Multi-scale finite element modelling using bubble function method for a convection-diffusion problem, Chemical Engineering Science, 61 (2006) 2742-2751.

Paper Two

Derivation of the residual free bubbles using the Method of Least Squares

57th Midwest PDE Seminar, April 2006, University of Chicago Illinois, US

Alireza Yazdani

Department of Chemical Engineering, Loughborough University, Loughborough,

LE11 3TU, UK

A.Yazdani@lboro.ac.uk

Abstract. The Galerkin's method of weighted residuals is based on local approximation of the solution of a given differential equation, where this approximation is obtained from substitution of a piecewise linear interpolation into equation. Nodal values of the piecewise solution are obtained throughout a reduction of the problem to a linear system of equations. The locality is then resolved via an assembly process, which, along with other steps, form up the very concept of Finite Element Method.

The so called "bubble function" method is developed from the idea of enrichment of the interpolation base functions by addition a bubble function, normally coming from an infinite dimensional augmented linear space, so that this bubble function takes zero at element boundaries which in turn, confines the generated approximation error inside the element as well as relaxing the non-homogeneity of the original boundary conditions [1- 2].

Ideally, the bubble function is the analytical solution of the residual differential equation, subject to homogeneous boundary conditions. However, the analytical

solution is hardly obtainable in general, and as is the case in many practical situations, a simple polynomial approximate form is needed for computational purposes. Many people adopt the aforementioned approach to treat differential equations numerically within the context of finite element modelling of physical phenomena.

In this work we assume a polynomial form of the bubble function and derive the approximate polynomial form of the practical bubble, using the method of least squares. This method turns out to yield high degree of accuracy, capability of generalization to higher dimensions and non-linear differential equations as well as being computerizable, where suggest a benchmark for dealing with different classes of differential equations.

References

[1] A. Cangiani, E. Suli, Enhanced RFB method, Numerische Mathematik, (2005) 101, 273-308.

[2] T. J. R. Hughes, G. R. Feijo`o, L. Mazzei, J-B. Quincy, The variational multiscale method-a paradigm for computational mechanics, Comput. Methods Appl. Mech. Engrg. 166 (1998) 3-24.

Paper Three

Solution of a Scalar Convection-Diffusion Equation Using FEMLAB

FEMLAB users Conference, October 2005, Hyatt Regency, Boston, US

A. Yazdani* L. Shojai

Department of Chemical Engineering,

Loughborough University,

Loughborough, LE11 3TU, UK

A.Yazdani@lboro.ac.uk

ABSTRACT

A steady scalar convection-diffusion problem has been studied for one and two dimensional cases. The major problem of unrealistic oscillations of the convection dominated problems is relaxed thanks to the wide range of the elements FEMLAB 3.1 benefits. The FEMLAB 3.1 solution has been presented for the problems, unique features and illustrations of the software have been used and results have been tested against analytic solution.

Keywords Convection-Diffusion, Convection -dominated, FEMLAB 3.1

1. Introduction

Processes involving a combination of convection and diffusion are ubiquitously found in physical and engineering problems. These problems occur in many applications such as in the transport of air and ground water pollutants, oil reservoir flow, in the modelling of semiconductors, and so forth. Convection is a physical process by which some property is transported by the ordered motion of the flow, while diffusion

is the physical process by which the property is transported by the random motion of the molecules of the fluid. The behaviour of fluid undergoing mass, vorticity, or forced heat transfer is described by a set of partial differential equations which are mathematical formulations of one or more of the conservation laws of physics. These laws include those of conservation of mass, momentum, and energy. The numerical solution of convection diffusion equations whose first derivative have large coefficients (convection dominated) presents difficulties such as parasitic oscillation and instability. Several finite element treatments of the problem have ever been tried and developed, including upwinding techniques, Petrov-Galerkin approach and artificial diffusivity method, and the more recent stabilized methods. FEMLAB 3.1, takes the advantage of employing these methods in a very straightforward and user-friendly way. The obtained results are therefore, as reliable and accurate as the results based on the most recent techniques, bearing in mind that all code developments and programming works are already done.

2. Statement of the Problem

The general steady linear problem on a bounded domain is of the form

$-\varepsilon\nabla.(a\nabla u)+\nabla.(\vec{b}u)+cu=S \text{ in } \Omega$ (2-1) with boundary conditions

$u=u_B \text{ on } \partial\Omega_D,\ \frac{\partial u}{\partial n}=0 \text{ on } \partial\Omega_N$ (2-2)

where $\partial\Omega_D, \partial\Omega_N$ form a partition of the boundary of Ω in which $\partial\Omega_D$ is non-empty. We usually assume that the advective velocity field $\vec{b}$ is incompressible, $\nabla\vec{b}=0$, so that the convective term can also be written $\vec{b}.\nabla u$, and also $a(x)\geq 1, c(x)\geq 0$ while ε is a small positive constant [2].

First we consider the one dimensional differential equation:

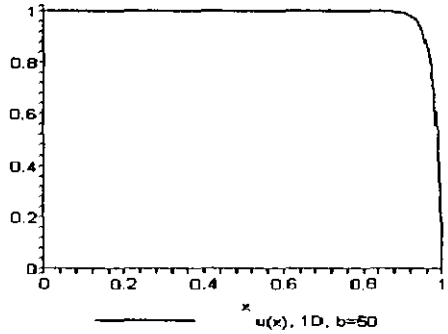
$$\frac{d^2u}{dx^2} - \bar{b}(x)\frac{du}{dx} = S(x) \ in \ \Omega = [0,1] \qquad (2\text{-}3)$$

subject to the given boundary conditions $u(0) = 1$, $u(1) = 0$. We assume that the convective term $\bar{b}(x)$ is a constant and there is no source term, i.e. $S(x) = 0$. In this case the analytic solution of equation (3-2) is given as:

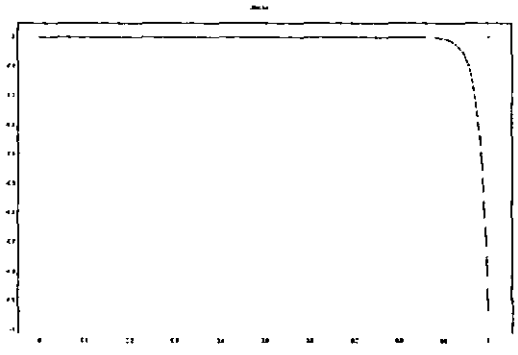
$$u(x) = \frac{e^{bx} - e^{b}}{1 - e^{b}}. \qquad (2\text{-}4)$$

Numerical schemes successfully cope with such a simplified linearized equation.

Setting b=50, and using predefined cubic Lagrange element, FEMLAB 3.1 returns: number of elements: 30, number of degrees of freedom: 91, solution time: 0.032 Seconds, for which the stationary analysis, Coefficient forms, PDE module, has been used. Figure 1. depicts similarity of analytic solution and FEMLAB result.



1.a. Analytic result b=50



1.b. FEMLAB result b=50

Figure 1. Comparison of analytic and FEMLAB results

The solution (using cubic Lagrange elements) is fast, accurate and highly reliable compared to the analytic solution.

3. Two dimensional test problem

Two dimensional convection-diffusion problem is represented by:

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}-b_1\frac{\partial u}{\partial x}-b_2\frac{\partial u}{\partial y}=S\,. \qquad (3\text{-}1)$$

Unlike the one dimensional case, it is not very easy to invent a range of two dimensional problems with ready analytical solutions. As a test problem we study the convection-diffusion model of (3-1) with constant b and no source term, over the unit square. For this purpose we set $b_1=b_2, S=0, \Omega=[0,1]\times[0,1]$ subject to the following Dirichlet boundary conditions:

$$\begin{cases} u(x,1)=0=u(1,y) \\ u(x,0)=1=u(0,y)\ . \\ u(0,1)=0.5=u(1,0) \end{cases} \qquad (3\text{-}2)$$

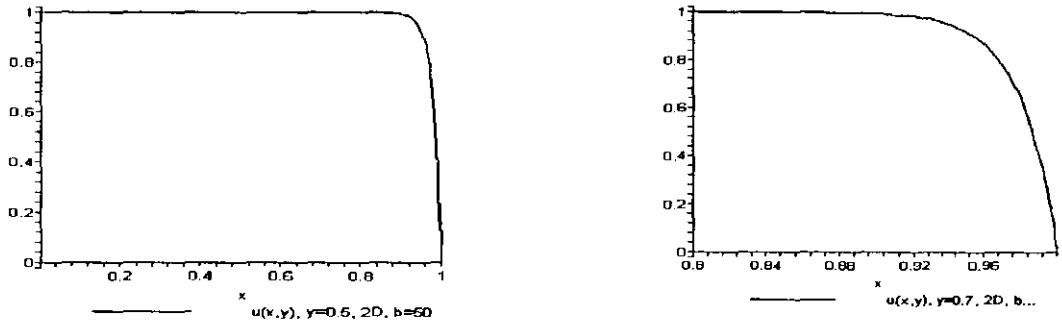
We shall simply write:

$$\begin{cases} \nabla^2 u-b\nabla u=0 \\ B.C. \end{cases} \qquad (3\text{-}3)$$

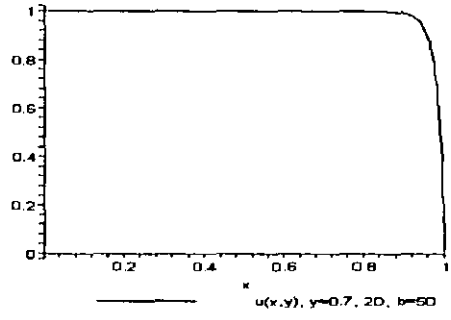
(which can be interpreted as the energy conservation equation with no heat source term). By the method of separation of variables the analytical solution of (3-2) is given by:

$$u(x,y)=\sum_{n=1}^{\infty}\left|\frac{(1-(-1)^n e^{\frac{-b}{2}})8n\pi e^{\frac{b(x+y)}{2}}}{\sinh(\frac{\sqrt{A_n}}{2})(b^2+4n^2\pi^2)}\right| \left|\sin(n\pi x)\sinh(\frac{\sqrt{A_n}(1-y)}{2})+\sin(n\pi y)\sinh(\frac{\sqrt{A_n}(1-x)}{2})\right| \qquad (3\text{-}4)$$

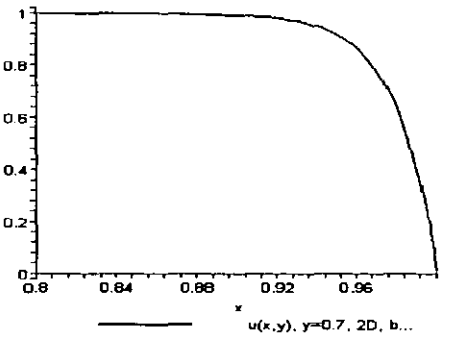
where $A_n = 2b^2 + 4n^2\pi^2 > 0$. The common higher degree interpolation functions give acceptable solution for value of $-40 \le b \le 40$, but show oscillatory behaviour for larger b, i.e. the convection dominated cases [3 & 5]. One of the treatments to the mentioned problem is to employ the so-called bubble function method, under the more general title of "stabilization techniques", in which the interpolation element includes some special element types in addition to the standard Lagrange elements. These elements are potentially useful for applications in fluid dynamics [1]. FEMLAB 3.1, employs these elements for application modes such as incompressible Navier-Stokes, Brinkman equations, non isothermal flow and many more, which is a novelty that could be extended to other eventualities. Setting $b = 50$ and y=0.5, 0.7 the analytic solution looks like what follows:



2.a Analytic result, layer y=0.5



2.b Analytic result, layer y=0.7, x=0..1



2.c Analytic result, layer y=0.7, x=0.8..1

Figure 2. Two dimensional test problem plotted at layers y=0.5 (2.a), y=0.7 (2.b& 2.c)

Solving the equation (3-3) using predefined quintic Lagrange element, FEMLAB returns within 4.438 seconds using a triangular mesh of 3976 elements with 50001 degrees of freedom. Figure 3 shows the plotted solution at different boundary layers:

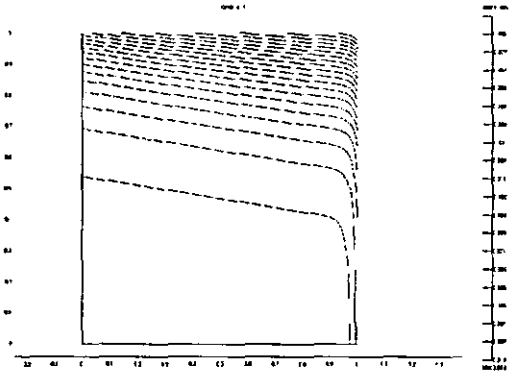


Figure 3. FEMLAB 3.1 result, b=50, Various layers

Equation (3-3) with large convection coefficients is known [5] to yield oscillatory results if treated by classical Galerkin finite element method. Different values of y should be interpreted on their own. FEMLAB results at this intermediate value of b is, however, quite satisfactory and accurate.

4. Conclusions

A steady, scalar convection-diffusion model has been studied, in which the convection coefficient was dominant, although not very large compared to unity. Simulated

results were highly reliable compared to the analytic solutions, processing and solution time was quite short and the software showed to be very easy to work with. The test problem was idealized in order to obtain analytic solutions, though, generalization to more complex geometries and problems is easily attainable. It was found that FEMLAB is a very successful modelling tool in terms of graphical features, coping with complex geometries, diversity of modules and models and specially being equipped with the stabilization techniques in many application modes.

Acknowledgment

The authors wish to thank Prof. V. Nassehi and Prof. R. J. Wakeman of Loughborough University, for their helps and supports.

References

[1] FEMLAB User's Guide, Version 3.1, pp. 474-490,COMSOL (2004).

[2] K. W. Morton, Numerical Solution of Convection-Diffusion problems, pp. 9-13, CHAPMAN & HALL, London (1996).

[3] V. Nassehi & S. A. King, Finite Element Methods for the Convection Diffusion Equation, IRI Journal of Engineering, Vol. 4, pp. 93-99 (1991).

[4] V. Nassehi, Practical Aspects of Finite Element Modelling of Polymer Processing, John Wiley and Sons, Chichester (2002).

[5] J. Petera, V. Nassehi, J. F. Pittman, Petrov-Galerkin Methods On Isoparametric Bilinear and Biquadratic Element Tested for a Scalar Convection-Diffusion Problem, Int. J. Meth. Heat Fluid Flow, Vol. 3, pp. 205-221 (1993).

List of grants awarded to this work:

- IMA Small Grant Scheme: The Institute of Mathematics and its Applications, UK, August 2006.

- Grant from 57th Midwest PDE Seminar, University of Illinois at Chicago and NSF, USA, April 2006.

APPENDIX B

Glossary of Mathematical Terms and Formulas

Supremum and Infimum:

Let S be a subset of $\Re$, the set of real numbers. We say that S has a lower bound β , if for any number $\alpha \in S$ we have $\alpha \geq \beta$. A lower bound β is called the greatest lower bound or supremum for the set S if for any other lower bound γ we have $\beta \geq \gamma$. The concept of infimum or the least upper bound is defined similarly.

Open and closed sets:

We call $S \subseteq \Re$ an open subset of $\Re$ if for every $x \in S$ there is some $\varepsilon \geq 0$ such that the open interval $(x - \varepsilon , x + \varepsilon)$ is still a subset of S. We call

$F \subseteq \Re$ a closed set if and only if F^c (the complement of F) is an open set.

Compact sets in $\Re$:

$F \subseteq \Re$ is called compact if it is closed and bounded (i.e. F^c is open and F is included in a sufficiently large interval).

Generalization to $\Re^d$:

By $\Re^d$ we mean the Cartesian product of $\Re$, d times. Every closed(resp. open) subset of $\Re^d$ is of the form of a Cartesian product of d closed (open) subsets of $\Re$. $F \subseteq \Re^d$ is compact if and only if F is bounded and closed. A domain in $\Re^d$ is a bounded open set.

Example: The set (a, b)= $\{x \in \Re: a<x<b\}$ is a bounded open interval(and so a domain) in $\Re$. The set [a, b]= $\{x \in \Re: a \leq x \leq b\}$ is a bounded closed (and so compact) interval in $\Re$. Evidently the sets (a, b)×(c, d) and [a,b]×[c,d] are open and closed

subsets of $\Re^2$, respectively. The set $(a,b]=\{x\in\Re: a<x\leq b\}$ is neither a closed nor an open subset of $\Re$.

Vector Spaces:

Let V be a set with two operations addition(+) and scalar multiplication(.). We call V a vector space if for any x, y, z in V and for every $\lambda, \mu \in \Re$ we have:

1) $x + y \in V,\ \lambda \in V$
2) $x + (y + z) = x + y + z = (x + y) + z$
3) $\lambda.(x + y) = \lambda. x + \lambda. y$
4) $(\lambda + \mu).(x) = \lambda. x + \mu. x$
5) $(\lambda\ \mu).(x) = \lambda.(\mu .x)$
6) There is a unique vector 0 in V, with $x + 0 = x$.
7) For each $x \in V$ there is a unique vector $-x$, with $(-x) + x = 0$
8) $(1.x) = x$.

Normed Vector Spaces:

Let V be a vector space. A function N (or $\|.\|$) is called a norm on V (and V is then called a normed vector space) if:

1) $N: V \rightarrow \Re^+$
2) $N(x + y) \leq N(x) + N(y)$ for every x, y in V
3) $N(x) = 0$ if and only if $x = 0$
4) $N(\lambda. x) = |\lambda|\, N(x)$.

Example: The set of real numbers with its ordinary addition and multiplication operations (take N = |. |, the absolute value function, resp.) is a vector space (normed vector space, resp.).

Linear independence and Bases:

Let V be a vector space. The set $B \subseteq V$ is called a basis for V provided

1) For every $x \in V$ there are $\lambda_1, \lambda_2, \ldots, \lambda_n \in \Re$ and $x_1, x_2, \ldots, x_n \in B$ such that $x = \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n$ for some $n \in N$.

2) For any number m, any set $\{ x_1, x_2, \ldots, x_m \} \subseteq B$ of vectors and any set $\{ \lambda_1, \lambda_2, \ldots, \lambda_m \} \subseteq \Re$ of scalars, the equality $\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_m = 0$ implies that $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$.

The first property states that every member of V is a linear combination of members of B. The second property states that B is a linearly independent subset of vector space V.

Every subset of V with these two properties is called a basis for V.

Inner product and Orthogonality:

An inner product $< . , . >$ on a vector space V is a function that :

1) $< x , y + z > = < x , y > + < x , z >$

2) $< \lambda x , y > = \lambda < x , y >$

3) $< x , x > > 0$ if $x \neq 0$.

If $< x , y > = 0$ then we say that x and y are orthogonal. $B \subseteq V$ is called orthogonal if every distinct pair in B are orthogonal.

Linear operator:

Suppose that V and W are two vector spaces. A function $T : V \rightarrow W$ is a linear operator if

1) $T(x + y) = T(x) + T(y)$ for every $x, y \in V$,

2) $T(\lambda x) = \lambda T(x)$ for each $x \in V$ and $\lambda \in \Re$.

Projection:

Suppose that $T : V \rightarrow W$ is a linear operator. We define Kernel of T to be $Ker(T) = \{ x \in V: T(x) = 0 \}$. Ker(T) is then a vector space itself, by addition and scalar multiplication induced from V and we say that Ker(T) is a vector subspace of V. Image of the operator T may be defined by $Im(T) = \{ T(x): x \in V \}$ and is a vector subspace of W.

If $T:V \rightarrow V$ and $T(T(x)) = T(x)$ for every $x \in V$ then we say T to be a projection on V. Remind that $T(0) = 0$.

Example : $\Re^3$ is a vector space.

Define :

1) $(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$

2) $\lambda.(x, y, z) = (\lambda.x, \lambda.y, \lambda.z)$.

The set $\{ i, j, k \}$ is a basis for $\Re^3$. In fact $i = (1, 0, 0)$, $j = (0, 1, 0)$ and $k = (0, 0, 1)$ and we have $(x, y, z) = xi + yj + zk$.

The set $\Re^3$ is an inner product space if we define :

$\langle (x_1, y_1, z_1), (x_2, y_2, z_2) \rangle = x_1 x_2 + y_1 y_2 + z_1 z_2$.

Example :

The function T : $\Re^3 \rightarrow \Re^3$ defined by T(x , y , z) = (0 , y , z) is a linear operator on $\Re^3$. The set Ker(T) consists of the x- axis of $\Re^3$. The set Im(T) is a vector subspace of z-y plane of $\Re^3$.

Eigen-values and eigenvectors :

Eigen-values are a special set of scalars associated with a linear system of equations that are sometimes known as characteristic roots, proper values or latent roots. Each eigen-value is paired with a corresponding so-called eigenvector.

Let A be a linear transformation represented by a matrix A. If there is a vector $0 \neq X \in \Re^n$ Such that

$$AX = \lambda X \qquad (1)$$

for some scalar λ , then λ is called the eigen-value of A with corresponding the eigenvector X.

Letting A be a k x k square matrix:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \qquad (2)$$

with eigen-value λ, then the corresponding eigenvectors satisfy

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \qquad (3)$$

which is equivalent to the homogeneous system

$$\begin{bmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk}-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (4)$$

Equation (4) can be written compactly as

$$(A - \lambda I)X = 0 \qquad (5)$$

Where I is the identity matrix. As shown in Cramer's rule , a linear system of equations has nontrivial solutions iff the determinant vanishes, so the solutions of equation (5) are given by

$$\text{Det}(A - \lambda I) = 0 \qquad (6)$$

This equation is known as the characteristic equation of A, and the left-hand side is known as characteristic polynomial.

For example, for a 2 x 2 matrix, the eigen-values are

$$\lambda_{\pm} = 1/2\,(a_{11} + a_{22}) \pm \sqrt{4a_{12}a_{21} + (a_{11} - a_{22})^2}$$

which arises as the solution of the characteristic equation

$$x^2 - x(a_{11} + a_{22}) + (a_{11}a_{22} - a_{12}a_{21}) = 0 \qquad (8)$$

If all *k* eigen-values are different, then plugging these back in gives k-1 independent equations for the *k* components of each corresponding eigenvector and the system is said to be non-degenerate. If the eigen-values are n-fold degenerate, then the system is said to be degenerate and the eigenvectors are not linearly independent. In such cases, the additional constraint that the eigenvectors be orthogonal,

$$X_i X_j = |X_i|\,|X_j|\,\delta_{ij} \qquad (9)$$

Where δ_{ij} is the Kronecker delta, can be applied to yield n additional constraints, thus allowing solution for the eigenvectors.

Assume we know the eigen-value for

$$A X = \lambda X \qquad (10)$$

Adding a constant times the identity matrix to A,

$$(A + c I) = (\lambda + c) X = \lambda' X \qquad (11)$$

so the eigen-values equal the old plus c. multiplying A by a constant c

$$(cA) X = c (\lambda X) \equiv \lambda' X \qquad (12)$$

so the new eigen-values are the old multiplied by c.

Now consider a similarity transformation of A. Let $|A|$ be the determinant of A, then

$$|Z^{-1} AZ - \lambda I| = |Z^{-1} (I - \lambda I) Z|$$

$$= |Z| |A - \lambda I| |Z^{-1}| = |A - \lambda I| \qquad (13)$$

so the eigen-values are the same as for A.

Matrix norms:

A norm of a matrix A is of order n x n, written as $\|A\|$, is a single number. The norm is a function of the element of A, and the following relations hold:

1. $\|A\| \geq 0$ and $\|A\| = 0$ if and only if $A = 0$. (a)
2. $\|cA\| = |c| \|A\|$ for any scalar c. (b)
3. $\|A+B\| \leq \|A\| + \|B\|$ for matrices A and B. (c)
4. $\|AB\| \leq \|A\| \|B\|$ for matrices A and B. (d)

The relation in (c) is the triangle inequality. The additional condition in (d), which was not postulated in the definition of a vector norm, must be satisfied in order to be able to use matrix norms when matrix products occur.

The following are frequently used matrix norms:

$$\|A\|_\infty = \max \sum_{j=1}^{n} |a_{ij}| \qquad (e)$$

$$\|A\|_1 = \max \sum_{i=1}^{n} a_{ij} \qquad \text{(f)}$$

$$\|A\|_2 = \sqrt{\tilde{\lambda}_n} \;;\; \tilde{\lambda}_n = \text{maximum eigen-value of } A^T A \qquad \text{(g)}$$

where for a symmetric matrix A we have $\|A\|_\infty = \|A\|_1$ and $\|A\|_2 = \max |\lambda_i|$

The norm $\|A\|_2$ is called the spectral norm of A. Each of these norms satisfies the relations in (a) to (d). The proof that the relation in (d) is satisfied for the infinity norm is given in example 2.

Example 1. Calculate the ∞-, 1-, 2-norms of the matrix A, where A is :

$$A = \begin{bmatrix} 5 & -4 & -7 \\ -4 & 2 & -4 \\ -7 & -4 & 5 \end{bmatrix}$$

Using the definition given in (e) to (g) we have

$\|A\|_\infty = 5 + 4 + 7 = 16$

$\|A\|_1 = 5 + 4 + 7 = 16$

The 2-norm is equal to $|\lambda_3|$, and hence $\|A\|_2 = 12$.

Example 2. Show that for two matrices A and B, we have

$$\|AB\|_\infty \leq \|A\|_\infty \, \|B\|_\infty$$

Using the definition of the infinity matrix norm in (e), we have

$$\|AB\|_\infty = \max \sum_{j=1}^{n} \left| \sum_{k=1}^{n} a_{ik} b_{kj} \right|$$

but then

$$\|AB\|_\infty \le \max \sum_{j=1}^{n} \sum_{k=1}^{n} |a_{ik}| \, |b_{kj}|$$

$$= \max \sum_{k=1}^{n} |a_{ik}| \sum_{j=1}^{n} |b_{kj}|$$

$$\le \{\max \sum_{k=1}^{n} |a_{ik}|\} \{\max \sum_{j=1}^{n} |b_{kj}|\}$$

This proves the desired result.

Condition number:

For the invertible matrix A, the value $K(A) = \|A\| \|A^{-1}\|$ is called the condition number with respect to the norm $\|.\|$. A matrix A is called to be well-conditioned if the value K(A) is close to 1, and is called ill-conditioned if K(A) is considerably bigger than 1. Note that for any invertible matrix A we have :

$1 = \| I \| = \| A \cdot A^{-1} \| \le \| A \| \| A^{-1} \| = K(A)$.

Notations:

If $A \subset X$, then $f[A] = \{f(x) | x \in A\}$ is a subset of Y and $f^{-1}[B] = \{x | f(x) \in B\}$ is a subset of X if $B \subset Y$. f[X] will usually be called the range of f and will be denoted by Ran(f). X is called the domain of f.

If $A \subset X$ we define the characteristic function $\chi_A(x)$ as:

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.$$

Metric spaces:

A metric space is a set M and a real-valued function d(., .) on M × M which satisfies :

1) $d(x, y) \geq 0$

2) $d(x, y) = 0$ if and only if x=y

3) $d(x, y) = d(y, x)$

4) $d(x, z) \leq d(x, y) + d(y, z)$ [triangle inequality]

The function d is called a **metric(distance)** on M.

Sequences in metric spaces:

A sequence of elements $\{x_n\}^{\infty}_{n=1}$ of a metric space <M, d> is said to **converge** to an element $x \in M$, if $d(x, x_n) \to 0$ as $n \to \infty$. We will often denote this by $x_n \xrightarrow{d} x$ or $\lim_{n\to\infty} x_n = x$ and say that x is the limit of sequence $\{x_n\}^{\infty}_{n=1}$. The limit of a sequence in a metric space is unique.

A sequence of elements $\{x_n\}$ of a metric space <M, d> is called a **Cauchy sequence** if $(\forall \varepsilon > 0)(\exists N) n, m \geq N$ implies $d(x_n, x_m) < \varepsilon$.

Proposition : Any convergent sequence is Cauchy.

A metric space in which all Cauchy sequences converge is called **complete**. A set B in a metric space M is called **dense** if every $m \in M$ is a limit of a sequence of elements in B. A function f from a metric space <X, d> to a metric space <Y,ρ> is called continuous at x if $f(x_n) \xrightarrow{<y,\rho>} f(x)$ whenever $x_n \xrightarrow{<x,d>} x$.

Measure theory:

Let $\chi_i(x)$ be the characteristic function of $[x_{i-1},x_i)$ except for $\chi_n(x)$ which is the characteristic function of $[x_{n-1},x_n]$. A function on [a,b] of the form $\sum_{i=1}^{n} s_i\chi_i(x)$ with s_i real is called a **step function** on $[a,b]=\bigcup_{i=1}^{n}[x_{i-1},x_i]$ where $x_0=a$ and $x_1=b$.

The family of **Borel sets (Measurable sets)** of $\Re$ is the smallest family of subsets of $\Re$ with the following properties:

1) The family is closed under complements.
2) The family is closed under countable unions.
3) The family contains each open interval.

Let $\Im$ be the family of all countable unions of disjoint open intervals (which is just the family of open sets) and let

$$\mu\left(\bigcup_{i=1}^{\infty}(a_i,b_i)\right)\equiv\sum_{i=1}^{\infty}\mu(b_i-a_i)$$

(which may be infinite). For any Borel set $B\in\Im$, define

$\mu(B)=\text{Inf}\{\mu(I)|\ B\subseteq I\ ,\ I \text{ is an Interval}\}$.

The notion of size has four crucial properties:

1) $\mu(\phi)=0$

2) If $\{A_n\}_{n=1}^{\infty}$ is a family of Borel sets and the A_n are mutually disjoint

($A_n\cap A_m=\phi$, all $m\neq n$), then $\mu\left(\bigcup_{i=1}^{\infty}A_n\right)\equiv\sum_{i=1}^{\infty}\mu(A_n)$.

3) $\mu(B)=\text{Inf}\{\mu(I)|\ B\subseteq I\ ,\ I \text{ is open}\}$.

4) $\mu(B) = \text{Sup}\{\mu(C) | C \subseteq B, C \text{ is compact}\}$.

A function f is called a Borel function(measurable function) if and only if $f^{-1}[(a, b)]$ is a Borel set(measurable set) for all a, b.

Proposition:

(a) If f, g are Borel, then so are f+ g ,fg ,max{f, g} and min{f, g}. If f is Borel and $\lambda \in \Re$, λf is Borel.

(b) If each f_n is Borel, n=1,2,3,..., and $f_n(p) \to f(p)$ for all p, then f is Borel.

Since $|f| = \max\{f,-f\}$, $|f|$ is measurable if f is.

As we sketched above, given $f \geq 0$, one can define $\int f dx$ (which may be ∞).

If $\int |f| dx < \infty$,we write $f \in L^1$ and define $\int f dx = \int f_+ dx - \int f_- dx$ where $f_+ = \max\{f,0\}$; $f_- = \max\{-f,0\}$. $L^1(a, b)$ is the set of functions on (a, b) which are in L^1 if we extend them to the whole real line by defining them to be zero outside of (a, b). If $f \in L^1(a, b)$, we write $\int f dx = \int_a^b f dx$.

Theorem: Let f and g be measurable functions. Then:

(a) If $f, g \in L^1(a, b)$, so are f+ g and λf, for all $\lambda \in \Re$.

(b) If $|g| \leq f$ and $f \in L^1$, $g \in L^1$.

(c) $\int (f + g) dx = \int f dx + \int g dx$ if $f, g \in L^1$.

(d) $\left| \int f dx \right| \leq \int |f| dx$ if f is in L^1.

(e) If $f \leq g$ then $\int f dx \leq \int g dx$ if f and g are in L^1.

(f) If f is bounded and measurable on $-\infty<a<b<\infty$, then $f\in L^1$ and

$$\left|\int_a^b f dx\right| \le |b-a| (\sup |f(x)|) \quad (a<x<b).$$

Theorem(monotone convergence theorem): Let $f_n \ge 0$ be measurable. Suppose $f_n(p) \to f(p)$ for each p and that $f_{n+1}(p) \ge f_n(p)$ all p and n (in which case we write $f_n \to f$). If $\int f_n(p)dp < C$ for all n, then $f\in L^1$ and

$\int |f(p) - f_n(p)|dp \to 0$ as $n\to\infty$.

Theorem(dominated convergence theorem): Let $f_n(p)\to f(p)$ for each p and suppose that $|f_n(p)| \le G(P)$ for all n and some G in L^1. Then $f\in L^1$ and

$$\left|\int f(p) - f_n(p)\right| dp \to 0 \text{ as } n \to \infty.$$

Definition:

We say a condition C(x) holds almost everywhere (a. e.) if the set {x | C(x) is false} is a subset of a set of measure zero. We say two functions $f, g \in L^1$ are equivalent if f(x)=g(x) a. e. (this is the same as saying $\int |f-g| dx = 0$). The set of equivalence classes in L^1 is denoted by L^1. L^1 with the norm $\|f\|_1 = \int |f| dx$ is a normed linear space.

Theorem: L^1 is complete.

Generalization:

The spaces $L^2, \ldots, L^p$ ($L^1, \ldots, L^p$) are defined in a similar way i.e. If $\int |f|^p dx < \infty$,we write $f\in L^p$. L^p with the norm $\|f\|_p = (\int |f|^p)^{1/p} dx$ is a normed linear space.

Hilbert spaces:

Denote that every vector space V with an inner product is a normed one, if we define $\|x\| = \sqrt{< x, x >}$ for every x∈V. Defining d(x, y)= $\|x - y\|$, every normed space is a metric space in turn. A complete inner product space is called a Hilbert space.

Example:

The space L^2 plays an important role in many applications. Consider for example the space L^2[a, b] the set of complex-valued measurable functions on [a, b], and define <f,g>= $\int_a^b g(x)\,\overline{f(x)}\,dx$ where $\overline{g(x)}$ represents the complex conjugate of g(x). It can be shown that L^2(a, b) is a Hilbert space with norm and metric induced from the inner product. On the other hand if μ is a Borel measure on $\Re^n$ and $L^2(\Re^n, d\mu)$ is the set of complex-valued measurable functions on $\Re^n$ which satisfy $\int_{\Re^n} |f(x)|^2 d\mu < \infty$ then $L^2(\Re^n, d\mu)$ is a Hilbert space under the inner product $< f, g >= \int_{\Re^n} \overline{f(x)} g(x) d\mu$.

Miscellaneous

Grad≡ $\nabla \equiv i\ \partial/\partial x + j\ \partial/\partial y + k\ \partial/\partial z$ is a vector.

Suppose that $F= (F_1, F_2, F_3)= i\ F_1+ j\ F_2+ k\ F_3$ is a vector field, then

$Div(F)= \nabla.F= \partial F_1/\partial x + \partial F_2/\partial y + \partial F_3/\partial z$ is a scalar.

$Laplacian(F)= \nabla^2(F)= Div(Grad(F))$ is a scalar.

$Curl(F)= \nabla \times F= \begin{vmatrix} i & j & k \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ F_1 & F_2 & F_3 \end{vmatrix}$ is a vector.

A function is called sufficiently smooth if it has continuous derivatives of sufficient orders. Such a function is shown by C^n if its n-th derivative exists and is continuous.

Green's Formula: If ∂D is the counter-clockwise oriented boundary of a domain D and f and g are C^1 functions on D, then

$$\oint_{\partial D} f dx + g dy = \iint_D (\partial g/\partial x - \partial f/\partial y) dx dy.$$

Integration by parts:

$$\int_a^b f dg = fg\Big|_a^b - \int_a^b g df \quad \text{where } df = f'(x)dx.$$

APPENDIX C

Maple Calculations

1. Derivation of a cubic bubble function for 1-D nonlinear diffusion-reaction

Problem: least squares approximation of a special function

```
> ode := K·(d/dx)(d/dx)u(x) + x·u(x) = 0; BC := u(0) = 0, u(2) = 1;
```

$$ode := K\left(\frac{d^2}{dx^2}u(x)\right) + x\,u(x) = 0$$

$$BC := u(0) = 0, u(2) = 1$$

```
> dsolve({ode, BC}, u(x));
```

$$u(x) = \frac{3\,\mathrm{AiryAi}\left(-\left(\frac{1}{K}\right)^{(1/3)}x\right)}{3\,\mathrm{AiryAi}\left(-2\left(\frac{1}{K}\right)^{(1/3)}\right) - \sqrt{3}\,\mathrm{AiryBi}\left(-2\left(\frac{1}{K}\right)^{(1/3)}\right)} - \frac{\sqrt{3}\,\mathrm{AiryBi}\left(-\left(\frac{1}{K}\right)^{(1/3)}x\right)}{3\,\mathrm{AiryAi}\left(-2\left(\frac{1}{K}\right)^{(1/3)}\right) - \sqrt{3}\,\mathrm{AiryBi}\left(-2\left(\frac{1}{K}\right)^{(1/3)}\right)}$$

```
> N := (x/l)·u(b) + ((l − x)/l)·u(0) + c·x·(l − x) + f·x^2·(l − x);
```

$$N := \frac{x\,u(b)}{l} + \frac{(l-x)\,u(0)}{l} + c\,x\,(l-x) + f\,x^2\,(l-x)$$

```
> R := K·(d/dx)(d/dx)N + x·N;
```

$$R := K\,(-2\,c + 2\,f\,(l-x) - 4\,f\,x) + x\left(\frac{x\,u(b)}{l} + \frac{(l-x)\,u(0)}{l} + c\,x\,(l-x) + f\,x^2\,(l-x)\right)$$

```
> J := int(R^2, x = 0 .. l):
```

$$solve\left(\left\{\frac{\partial}{\partial c}\,J = 0,\ \frac{\partial}{\partial f}\,J = 0\right\},\ [c, f]\right);$$

$$\left[\left[c = -\frac{1}{5\,l^{12} - 952\,l^{9}\,K + 89712\,l^{6}\,K^2 - 2540160\,l^{3}\,K^3 + 25401600\,K^4}\,(4\,(\right.\right.$$

$$5\,l^{9}\,u(0) - 5\,l^{9}\,u(b) - 763\,l^{6}\,u(0)\,K + 637\,K\,u(b)\,l^{6} + 13020\,K^2\,u(b)\,l^{3}$$

$$+ 39900\,l^{3}\,u(0)\,K^2 - 529200\,K^3\,u(b) - 529200\,K^3\,u(0))\,l),\ f = (6\,(3\,l^{9}\,u(0)$$

$$- 392\,l^{6}\,u(0)\,K + 6160\,l^{3}\,u(0)\,K^2 + 1456\,K\,u(b)\,l^{6} - 53200\,K^2\,u(b)\,l^{3}$$

$$+ 705600\,K^3\,u(b) - 10\,l^{9}\,u(b)))/(5\,l^{12} - 952\,l^{9}\,K$$

$$\left.\left. + 89712\,l^{6}\,K^2 - 2540160\,l^{3}\,K^3 + 25401600\,K^4)\right]\right]$$

```
>
```

2. Derivation of a cubic least squares bubble function for a 1-D diffusion-reaction Problem

> $ode := \lambda \cdot \frac{d}{dx}\,\frac{d}{dx}\,u(x) + u(x) = 0;\ BC := u(0) = 1,\ u(1) = 0;$

$$ode := \lambda \left(\frac{d^2}{dx^2} u(x) \right) + u(x) = 0$$

$$BC := u(0) = 1,\ u(1) = 0$$

> $dsolve(\{ode, BC\}, u(x));$

$$u(x) = -\frac{\cos\left(\frac{1}{\sqrt{\lambda}}\right) \sin\left(\frac{x}{\sqrt{\lambda}}\right)}{\sin\left(\frac{1}{\sqrt{\lambda}}\right)} + \cos\left(\frac{x}{\sqrt{\lambda}}\right)$$

>

$$N := \left(\frac{l - x}{l}\right) \cdot u(0) + \left(\frac{x}{l}\right) \cdot u(l) + c \cdot x \cdot (l - x) + f \cdot x^2 \cdot (l - x)$$

$$N := \frac{(l - x)\,u(0)}{l} + \frac{x\,u(l)}{l} + c\,x\,(l - x) + f\,x^2\,(l - x)$$

> $R := simplify\left(-\frac{1}{10000} \cdot \frac{d}{dx}\,\frac{d}{dx}\,N + N \right)$

$$R := \frac{1}{5000}\,\frac{1}{l}\,(c\,l - f\,l^2 + 3\,f\,x\,l + 5000\,u(0)\,l - 5000\,u(0)\,x + 5000\,x\,u(l)$$

$$+ 5000\,c\,x\,l^2 - 5000\,c\,x^2\,l + 5000\,f\,x^2\,l^2 - 5000\,f\,x^3\,l)$$

```
> simplify(solve({diff(J,c)=0, diff(J,f)=0},[c,f]));
```

$$\left[\left[c = -\frac{1}{(3 + 5000\,l^2 + 2500000\,l^4)\,(2500000\,l^4 + 21000\,l^2 + 63)}\,(5000\,(\right.\right.$$

$$65000000\,l^4 u(0) + 7500000000\,l^6 u(0) - 1250000000\,l^6 u(l) - 5000000\,l^4 u(l)$$

$$+ 168000\,u(0)\,l^2 + 52500\,u(l)\,l^2 + 126\,u(0) + 63\,u(l))),\ f =$$

$$\left.\left.\frac{35000\,(3\,u(0) + 500\,u(0)\,l^2 - 500\,u(l)\,l^2 - 3\,u(l))}{(2500000\,l^4 + 21000\,l^2 + 63)\,l}\right]\right]$$

3. Derivation of a quadratic least squares bubble function for a 2-D convection-diffusion Problem

> *restart* :

> $pde := k \cdot \left(\frac{\partial}{\partial x} u(x,y) + \frac{\partial}{\partial y} u(x,y) \right) + \left(\frac{\partial}{\partial x} \frac{\partial}{\partial x} u(x,y) + \frac{\partial}{\partial y} \frac{\partial}{\partial y} u(x,y) \right) = 0;$

$$pde := k \left(\frac{\partial}{\partial x} u(x,y) + \frac{\partial}{\partial y} u(x,y) \right) + \frac{\partial^2}{\partial x^2} u(x,y) + \frac{\partial^2}{\partial y^2} u(x,y) = 0$$

> $N := \frac{l - x}{l} \cdot \frac{h - y}{h} \cdot u(0,0) + \frac{l - x}{l} \cdot \frac{y}{h} \cdot u(0,h) + \frac{x}{l} \cdot \frac{h - y}{h} \cdot u(l,0) + \frac{x}{l} \cdot \frac{y}{h} \cdot u(l,h) + c \cdot x \cdot y \cdot (l - x) \cdot (h - y);$

$$N := \frac{(l-x)\,(h-y)\,u(0,0)}{l\,h} + \frac{(l-x)\,y\,u(0,h)}{l\,h} + \frac{x\,(h-y)\,u(l,0)}{l\,h} + \frac{x\,y\,u(l,h)}{l\,h} + c\,x\,y\,(l-x)\,(h-y)$$

> $R := k \cdot \left(\frac{\partial}{\partial x} N + \frac{\partial}{\partial y} N \right) + \left(\frac{\partial}{\partial x} \frac{\partial}{\partial x} N + \frac{\partial}{\partial y} \frac{\partial}{\partial y} N \right)$:

> $J := \int_0^l \int_0^h R^2 \, dy \, dx$:

$$> \; c := solve\left(\frac{\partial}{\partial c} J, c\right)$$

$$c$$

$$= -\frac{1}{l\,h\,(12\,l^4 + h^2\,l^4\,k^2 + h^4\,l^2\,k^2 + 12\,h^4 + 20\,h^2\,l^2)}\,(5\,k\,(-l^2\,k\,h^2\,u(0,0)$$

$$+ l^2\,k\,h^2\,u(l,0) + l^2\,k\,h^2\,u(0,h) - l^2\,k\,h^2\,u(l,h)$$

$$+ 3\,l^2\,u(0,h)\,h - 3\,l^2\,u(l,h)\,h + 3\,l^2\,u(0,0)\,h - 3\,l^2\,u(l,0)\,h + 3\,l\,u(0,0)\,h^2$$

$$+ 3\,l\,u(l,0)\,h^2 - 3\,l\,u(0,h)\,h^2 - 3\,l\,u(l,h)\,h^2 + 3\,l^3\,u(l,0) + 3\,l^3\,u(0,0)$$

$$+ 3\,h^3\,u(0,h) - 3\,h^3\,u(l,0) - 3\,h^3\,u(l,h)$$

$$+ 3\,h^3\,u(0,0) - 3\,l^3\,u(l,h) - 3\,l^3\,u(0,h)))$$

4. Derivation of a cubic least squares bubble function for a 1-D general convection-diffusion-reaction Problem

> *restart;*

> $ode := \lambda \cdot \frac{d}{dx}\frac{d}{dx}u(x) + \beta \cdot \frac{d}{dx}u(x) + \varepsilon \cdot u(x) = 0;$

$$ode := \lambda\left(\frac{d^2}{dx^2}u(x)\right) + \beta\left(\frac{d}{dx}u(x)\right) + \varepsilon\, u(x) = 0$$

> $N := \frac{l-x}{l} \cdot u(0) + \frac{x}{l} \cdot u(l) + c \cdot x \cdot (l-x) + f \cdot x^2 \cdot (l-x);$

$$N := \frac{(l-x)\,u(0)}{l} + \frac{x\,u(l)}{l} + c\,x\,(l-x) + f x^2\,(l-x)$$

> $R := \lambda \cdot \frac{d}{dx}\frac{d}{dx}N + \beta \cdot \frac{d}{dx}N + \varepsilon \cdot N;$

$$R := \lambda\,(-2c + 2f(l-x) - 4fx) + \beta\left(-\frac{u(0)}{l} + \frac{u(l)}{l} + c\,(l-x) - c\,x + 2fx\,(l-x) - fx^2\right) + \varepsilon\left(\frac{(l-x)\,u(0)}{l} + \frac{x\,u(l)}{l} + c\,x\,(l-x) + fx^2\,(l-x)\right)$$

> $J := simplify\left(\int_0^l R^2\, dx\right):$

```
> solve( { diff(J,c)=0, diff(J,f)=0 }, [c,f] );
```

$$\left[\left[c = (-6\,l^7\varepsilon^4 u(0) + l^7\varepsilon^4 u(l) - 40\,l^5\varepsilon^3 u(l)\,\lambda\right.\right.$$

$$+ 520\,\lambda\,l^5\varepsilon^3 u(0) - 70\,l^5\varepsilon^2 u(l)\,\beta^2 - 140\,l^5\varepsilon^2 u(0)\,\beta^2 - 1320\,\lambda\,\beta\,u(0)\,l^4\varepsilon^2$$

$$- 780\,\lambda\,\beta\,u(l)\,l^4\varepsilon^2 - 13440\,\lambda^2 l^3\varepsilon^2 u(0) - 4200\,l^3\varepsilon^2 u(l)\,\lambda^2$$

$$- 840\,\lambda\,l^3\varepsilon\,u(l)\,\beta^2 + 3360\,\lambda\,l^3\varepsilon\,u(0)\,\beta^2 - 5040\,\lambda^2\beta\,u(l)\,l^2\varepsilon$$

$$+ 30240\,\lambda^2\beta\,u(0)\,l^2\varepsilon + 2520\,\lambda\,\beta^3 u(l)\,l^2 - 2520\,\lambda\,\beta^3 u(0)\,l^2$$

$$+ 100800\,\lambda^3 l\varepsilon\,u(0) + 50400\,\lambda^3 l\varepsilon\,u(l) - 25200\,\lambda^2\beta^2 u(0)\,l$$

$$+ 25200\,\lambda^2\beta^2 u(l)\,l - 151200\,\lambda^3\beta\,u(0) + 151200\,\lambda^3\beta\,u(l))\Big/(l\,(302400\,\lambda^4$$

$$+ l^8\varepsilon^4 + 420\,l^4\beta^4 - 1680\,l^4\lambda\varepsilon\beta^2 + 4320\,l^4\lambda^2\varepsilon^2 - 60480\,l^2\lambda^3\varepsilon - 104\,l^6\lambda\varepsilon^3$$

$$+ 5040\,l^2\lambda^2\beta^2 + 52\,l^6\varepsilon^2\beta^2)),\ f = -(7\,(-l^6\varepsilon^4 u(0)$$

$$+ 3600\,\lambda^2\beta\,u(0)\,l\varepsilon - 300\,\lambda\,\beta\,u(0)\,l^3\varepsilon^2 + 7200\,\lambda^2\beta^2 u(l)$$

$$+ 10\,l^4\varepsilon^2 u(l)\,\beta^2 - 80\,l^4\varepsilon^3 u(l)\,\lambda + 1320\,l^2\varepsilon^2 u(l)\,\lambda^2 - 1320\,\lambda^2 l^2\varepsilon^2 u(0)$$

$$+ 7200\,\lambda^3\varepsilon\,u(0) + 80\,\lambda\,l^4\varepsilon^3 u(0) - 7200\,\lambda^3\varepsilon\,u(l)$$

$$+ 3600\,\lambda^2\beta\,u(l)\,l\varepsilon - 300\,\lambda\,\beta\,u(l)\,l^3\varepsilon^2$$

$$+ 600\,\lambda\,l^2\varepsilon\,u(0)\,\beta^2 - 600\,\lambda\,l^2\varepsilon\,u(l)\,\beta^2 - 10\,l^4\varepsilon^2 u(0)\,\beta^2$$

$$\left.\left. + l^6\varepsilon^4 u(l) - 7200\,\lambda^2\beta^2 u(0))\right)\Big/(l\,(302400\,\lambda^4 + l^8\varepsilon^4\right.$$

$$+ 420\, f^4 \beta^4 - 1680\, f^4 \lambda\, \epsilon\, \beta^2 + 4320\, f^4 \lambda^2 \epsilon^2 - 60480\, f^2 \lambda^3 \epsilon - 104\, f^6 \lambda\, \epsilon^3$$

$$+ 5040\, f^2 \lambda^2 \beta^2 + 52\, f^6 \epsilon^2 \beta^2)) \Big] \Big]$$