Efficient Spectral Element Methods for Partial Differential Equations.

Ahmed Alshehri

Supervised by Prof. Tim N. Phillips

A thesis submitted for the degree of
Doctor of Philosophy

16th December 2019

School of Mathematics

Cardiff University

STATEMENTS AND DECLARATIONS TO BE SIGNED BY THE CANDIDATE AND INCLUDED IN THE THESIS

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD).

Signed . . . . . .

Date . . . . . .

STATEMENT 2

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed . . . . . .

Date . . . . . .

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed . . . . . .

Date . . . . . .

DECLARATION

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students Procedure.

Signed . . . . . .

Date . . . . . .

WORD COUNT . . . . . .

(Excluding summary, acknowledgments, declarations, contents pages, appendices , tables, diagrams and figures, references, bibliography, footnotes and endnotes)

Dedication

I would like to dedicate my thesis to my sweetheart my wife.

Abstract

In this thesis we applied a spectral element approximation to some elliptic partial differential equations. We demonstrated the difficulties related to the approximation of a discontinuous function in which the discontinuity is not fitted to the computational mesh. Such a situation gives rise to the Gibbs phenomenon. A $h-p$ spectral element equivalent of the eXtended Finite Element Method (XFEM), which we termed the eXtended Spectral Element Method (XSEM) was developed. This was applied to some model problems. XSEM removes some of the oscillations caused by Gibbs phenomenon. We then explained that when approximating a discontinuous function, XSEM is able to capture the discontinuity precisely. We derive spectral element error estimates. The convergence of the approximations is studied.
We have introduced several enrichment functions with the purpose of improving the approximation of discontinuous functions. In particular we have considered the two-dimensional Poisson equation. Unfortunately, this implementation of XSEM was not able to recover spectral convergence. An alternative idea in which the discontinuity is constrained within a spectral element produces accurate SEM approximation.

The Stokes problem was considered and solved using SEM coupled with an iterative PCG method. The zero volume condition on the pressure is satisfied identicaly using an alternative formulation of the continuity equation. Furthermore, we investigated the dependence of the accurency of the spectral element approximation on the weighting factor as well as the convergence properties of the preconditioner. An efficient and robust preconditioner is constructed for the Stokes problem. Exponential convergence was attained.

Acknowledgements

First of all I would like to thank my supervisor, Prof. Tim Phillips, for his excellent guidence, understanding, support and patience during the completion of this work. I would also like to thank him for the support and understanding he gave during the difficult period when arriving first time to Cardiff University. I genuinely cannot thank you enough for everything you have done over the last few years. It has been a pleasure to work with you and I hope to be able to do so again in the future.

I am very grateful to Prof. Tim Phillips who encouraged me to pursue this topic and spent extra time helping me to achieve a clearer structure. I acknowledge, with gratitude, members of the Applied Maths group at Cardiff for the friendly atmosphere and stimulating research environment. I would like acknowledge the financial support of the Saudi Ministry of Foreign Affairs.

Finally, I would like to thank my family for all the support you have given me over the last few years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Polynomial Interpolation

Differential equations are incorporated into a variety of different scientific disciplines, including biology, chemistry, physics, and economics. It is not possible to obtain closed form or analytical solutions to partial differential equations for the majority of problems. Therefore, scientific and engineering experts have designed numerical techniques, like the finite difference method (FDM) [79], the finite element method (FEM) [5, 109], meshless methods [40, 61], spectral methods [13], boundary element methods [88], discrete element methods [71], and Lattice Boltzmann methods [97] among others in order to determine approximations to the solution of differential equations.

The aforementioned numerical methods are employed to determine numerical approximations to the solutions of ordinary differential equations (ODEs) and partial differential equations (PDEs). The application of such methods is additionally called "numerical integration", but this term is often interpreted as the calculation of integrals.

It is possible to categorise such numerical methods into either local or global groups. Both finite difference and finite-element methods are founded on local approximations, while the spectral technique is characterised by its global nature. In practical terms, finite-element methods have increased suitability for complex geometric problems, while spectral methods can offer greater precisions, while sacrificing domain flexibility. It is emphasised that there are also numerical methods, like *hp* fi-

nite elements and spectral-elements, which amalgamate the benefits offered by local and global techniques.

The finite element method (FEM) has been employed for many years as a numerical instrument to solve a variety of problems in the field of engineering as it is capable of managing the complex challenges associated with geometric and material properties [31]. Computational mechanics based on FEM is an important element of numerous scientific and engineering fields. FEM does not operate on the strong form of the differential equations; rather, the continuous boundary and initial value problems are reformulated into similar variational forms. When applying the FEM, it is necessary for the domain to be separated into regions that do not overlap, known as elements.

Within FEM, a topological map, also known as a mesh, creates connections between each of the elements, while local polynomial representation is utilised for the fields inside every element. The resulting solution is dependent on the mesh quality, and the basic necessity is that the mesh conforms to the geometry. The primary benefit of the FEM is that it has the capability to easily manage complicated boundaries. It should be noted that certain solution approaches are targeted at dealing with multiple limitations of the FEM, such as meshfree methods [40, 61], and the recently developed Smoothed Finite Element Method (SFEM) [12, 62, 73]. It has been observed that FEM using piecewise polynomials is not efficient at addressing singularities or high gradients within the domain.

One approach involves the enrichment of the FEM approximation basis with more functions [95]. It is possible to combine certain developed methods with enrichment methods for resolving problems that involve singularities or high gradients. An example of a numerical method founded on the generalised finite element method (GFEM) and the partition of unity method (PUM) is the extended finite element method (XFEM). It expands on the traditional finite element method (FEM) methodology through the enrichment of the solution space to differential equations with prominent non-smooth properties in localised regions within the computational domain, such as

close to singularities or discontinuities.

When numerically solving partial differential equations, the spectral element method is formulated on the basis of the finite element method, which utilises high-degree piecewise polynomials as the basis functions, as suggested by Patera in 1983 [80].

The two methods involve the decomposition of the computational domain into sub-domains that have a suitably small size, followed by approximation of the solution. FEM employs low-order expansion and is capable of producing outcomes for domains characterised by their highly complex nature, although the accuracy is reduced. On the other hand, the advantage of SEM is enhanced accuracy, which is not easy to achieve when using low-order methods.

The application of domain decomposition methodology is similar to the finite element method. The whole computational domain is separated into distinct sub-domains (elements). Due to the fact that the integration process is performed on a standard (or reference) sub-domain (in increased dimensions $[-1,1]^d$, in which d denotes the physical dimension), the transition of every element to standard (or reference) domain is done through the coordinate transformation process. The Gauss-Lobatto-Legendre (GLL) integration will be applied. All computations will be implemented on an arbitrary (quadrilateral) domain, whereas the integration is performed on $[-1,1]^2$.

The spectral and finite element methods are compared. One of the similar features of these two techniques is their weighted residual foundation, which facilitates the combination of the methods. However, differences can be observed when the weighted residual framework is extended to the final series of discrete equations. Spectral methods utilise high order indefinitely differentiable basis functions, which are predominantly Chebyshev or Legendre polynomials. Conversely, finite element methods employ low order basis functions. Additionally, spectral methods have a global basis, which is defined across the entire given domain, while in the finite elements method (as with each of the methods involving domain decomposition), the basis functions are characterised as being local. The differentiating factor mentioned above has sig-

nificant repercussions; when using spectral methods, the approximate solution of the partial differential equation is converged via a higher spectral expansion order. Contrastingly, in the finite element methodology, this is accomplished via a greater amount of elements. An additional significant ramification is that while spectral expansions are characterised by their spectral or exponential precision, the limitation of finite element methods is that they only provide a maximum of algebraic convergence. However, the utilisation of global approximations with a higher order generates full system matrices in the resulting series of discrete equations. Finite element systems are sparse as a result of the local nature of the approximation. Element separation that is performed in the finite element method facilitates the resolution of highly complex geometric problems, as opposed to spectral techniques. Furthermore, the potential to refine the local mesh when performing finite element methods allows complicated physical phenomena to be addressed, including powerful solution discontinuities, which is a benefit in comparison to the global spectral methods common to the spectral element methods. In conclusion, spectral methods are highly compatible with problems where both the solution and data are characterised by their regularity and the complexity of the domain is low. Nevertheless, where the geometry involves either strong data discontinuities or complexities, it is evident that the implementation of finite element methods presents fewer difficulties.

1.2 Convergence of Approximations

With trial functions such as Chebyshev or Legendre polynomials and if the solution is m times differentiable $(u \in H^m(\Omega))$, it can be demonstrated that a constant C exists, where the following bound is satisfied by the approximation:

$$\|u - u_N\|_{L^2} \leq CN^{-m}\|u\|_{H^m}$$

where N is the order of polynomial. It should be noted that for functions that have infinite smoothness, this "truncation" flaw has the attribute of exponential convergence

when N is increased as it applies for each m. However, it should be emphasised that the estimation is asymptotic, meaning that it is applicable only for a sufficiently large N. The convergence theory for the Legendre series is practically the same as in Chebyshev. When comparing these two sets, weaker estimations are produced for the maximum pointwise error. For example, the estimation is poorer for the Legendre series in comparison to the Chebyshev series when it has an identical order, truncated subsequent to N terms by $O(N^{1/2})$.

It could be questioned why the Legendre series is being discussed when it has been determined that the performance of Chebyshev is better. This can be explained by the weight function, which produces all of these orthogonal polynomial sets. Although improved spectral convergence of the expansion coefficients as well as generation of the optimal polynomial approximation of continuous functions is achieved with Chebyshev polynomials, the Legendre set is selected as the attributes have strong similarities to the Chebyshev, while their generating weight function is $w_C(x) = 1$, rather than $w_C(x) = \dfrac{1}{\sqrt{1-x^2}}$ for Chebyshev. This property associated with Legendre polynomials allows the weak formulation of the differential equation to be simplified. Spectral element techniques also utilise the weak form, and this will be discussed in the following chapter; thus, only the Legendre polynomials will be used as the basis for the numerical schemes from this point.

The definition of a discontinuity is a swift alteration of a field quantity across a length, which is minimal in comparison to the observed domain dimensions. In reality, discontinuities are observed on a frequent basis. For examples, both stresses and strains in solids are discontinuous across material interfaces, while at cracks, displacements are continuous. Tangential displacements are discontinuous across shear bands. In regard to fluids, both pressure and velocity fields could incorporate discontinuities where two fluids are interfaced.

If $\Omega \in \mathbb{R}^n$ is considered to be the computational domain including two distinct immiscible incompressible phases and $\partial\Omega$ is the boundary of Ω. The sub-domains in

which the two phases are contained are defined as Ω_1 and Ω_2 with $\Omega = \Omega_1 \cup \Omega_2$ and $\Omega_1 \cap \Omega_2 = \emptyset$. The assumption is made that there is a connection between Ω_1 and Ω_2.

Within an element, $\Omega_e, e = 1,2$, the approximation for a function f is continuous. Nevertheless, in the event that the approximated function contains a discontinuity, it is possible to observe spurious oscillations in the approximation. This kind of phenomenon is widely recognised, and is defined as the Gibbs phenomenon; see for instance, [10, 11].

It is possible to implement a formal categorisation of the Gibbs phenomenon as its lack of ability to perform the approximation of a discontinuity by using continuous functions. The spectral element interpolation of a discontinuous function is addressed based on a grid where the points have uniform spacing so as to provide an explanation for this phenomenon. Certain complexities arise when using spectral methods for the approximation of a discontinuous function. For example, the spurious oscillations have the potential to contaminate the additional variables included in the computation, particularly if the discontinuity is allowed unrestricted movement within the computational domain. The next section will present a method that is capable of moderating these oscillations.

1.3 Weak and Strong Discontinuities

Research into strong and weak discontinuities performed by Cheng and Fries [20] was aimed at developing the sub-optimal order of convergence that was found in the application of a higher-order XFEM to curved discontinuities. The source of the sub-optimal order of convergence was the approximations in the quadrature, as demonstrated by Legay et al. [60]; hence, Cheng and Fries [20] supported a different quadrature scheme that involved the subdivision of an element containing the discontinuity into elements of smaller size where one of the sides of sub-element is curved. This infers that the curved side of the element has an increased amount of nodes in comparison to the non-curved sides. A different quadrature scheme is used for the purposes of this thesis.

Additionally, Cheng and Fries [20] recommended that a modified higher-order XFEM should be applied. In terms of the rates of convergence for curved weak discontinuities, the conventional higher-order XFEM did not produce optimal rates. However, Cheng and Fries [20] determined that it was possible to obtain maximum rates when equal-order basis functions were applied for both the conventional and extended aspects of the enriched approximation. Moreover, due to the fact that they adopted a modified abs-enrichment (where the enrichment function had absolute-value), they determined that a sub-optimal order of convergence emerged for curved discontinuities. Furthermore, the rates of convergence considered were all h-type, implying convergence in regard to the optimal polynomial degree and the width of the mesh, which was four.

The required steps included in the application of the XFEM are:

1. It is possible to explicitly represent the discontinuity or interface via line segments, or implicitly through the use of the level set technique (LSM) [76, 90].

2. Where local enrichment is involved, only a subgroup of the nodes nearby the region in question is enriched. Selection of the nodes that will be enriched is performed on the basis of an area criterion or based on the nodal values of the level set function.

3. Based on the physics of the issue, various enrichment functions can be employed.

4. An outcome of including customised enrichment in the FE approximation basis.

If we consider an n-dimensional domain $\Omega \in \mathbb{R}^n$ that is discretised by n^{el} elements, assigned numbers from 1 to n^{el}, I is defined as the set including each of the nodes within the domain, and I^* denotes the nodal subgroup of the enrichment ($I^* \subset I$). A conventional extended finite element approximation of function $u(x)$ can be formulated as

$$\begin{aligned} u^h(x) &= u^h_{FEM}(x) + u^h_{Enr}(x) \\ &= \sum_{i\in I} N_i(x)u_i + \sum_{j\in I^*} N^*_j(x)\rho(x)a_j. \end{aligned} \tag{1.1}$$

For the purposes of simplification, consideration is only shown for a single enrichment term. The approximation is comprised of a standard finite element (FE) in addition to the enrichment. The individual variables can be defined as follows:

- $u^h(x)$: approximated function,
- $N_i(x)$: Standard FE function of node i,
- u_i: unknown of the Standard FE part at node i,
- $N^*_j(x)$: conventional FE shape functions that are not automatically identical to the ones used in the standard section of the approximation ($N_i(x)$).
- $\rho(x)$: global enrichment function. The enrichment function $\rho(x)$ contains the essence of the solution or data regarding the fundamental physics associated with the problem; for instance, $\rho(x) = H$, is utilised in the capturing of robust discontinuities, while H defines the Heaviside function
- a_j: unknown of the enrichment at node j.

Equation (1.1) provides a general definition of the XFEM. To achieve a specific realisation of the XFEM, it is important to define the global enrichment function $\rho(x)$and the division of unity functions $N^*_i(x)$ along with the selection of the nodal subset $I^\star$.

With respect to the debate regarding strong discontinuities, Legay et al. [60] determined that additional consideration was not necessary in the blending elements. Nonetheless, in situations where the discontinuities are not strong, it was found by Legay et al. that terms of higher order that arise inside the blending elements should be removed. According to their observations, the elimination of the terms of higher order inside the blending elements was enough when using polynomials of degree $N-1$

in the enrichment process. Although Legay et al. [60] reviewed spectral basis functions, the researches addressed XSEM on the basis of a high-order FEM perspective. In terms of weak straight discontinuities, Legay et al. [60] found that this method generates an order of convergence that neared the optimum. Nonetheless, in relation to weak curved discontinuities, the order of convergence generated by this method was sub-optimal. The scheme employed by Legay et al. partitioned the element containing the discontinuity into smaller elements. If it is determined that the discontinuity is contained within one of these elements, then the smaller element is additionally partitioned into triangles thus generating a linear approximation of the discontinuity contained within the element.

As an illustration of this phenomenon, the spectral element interpolation of a discontinuous function on a grid of points with uniform spacing is considered. In order to simplify the process, the assumption is made that the domain $\Omega \subset \mathbb{R}$ and there is only *one* spectral element. The domain is defined as $\Omega = [-1,1]$ and we presume that the function we will interpolate is piecewise constant:

$$f(x) = \begin{cases} x+1 & \forall x \in [-1, 10^{-4}) \\ x & \forall x \in [10^{-4}, 1] \end{cases} \tag{1.2}$$

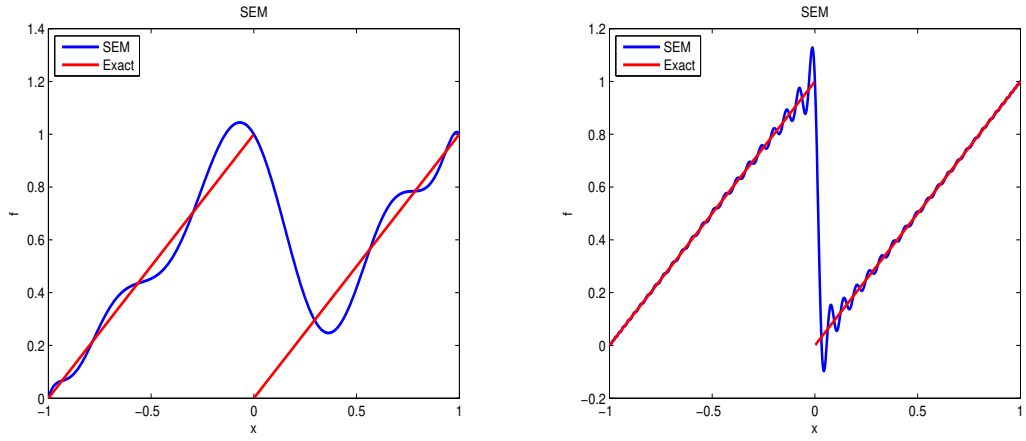
It should be noted that the point 10^{-4} is selected as the intersection point between the two sub-domains' boundaries as 0 belongs to the Gauss-Lobatto Legendre grid. The spectral element of this function is constructed on a grid with uniform spacing utilising only one element. The grid with uniform spacing is formulated as:

$$D_u = [x_0, x_1] \cup [x_1, x_2] \ldots \cup [x_{M-1}, x_M] = \bigcup_{k=1}^{M} [x_{k-1}, x_k]$$

where $x_0 = -1$ and $x_M = 1$ and M is the total number of points with uniform spacing. Hence, the interpolant is:

$$f_N(x) = \sum_{i=0}^{N} f_i h_i(x), \qquad \forall x \in D_u \tag{1.3}$$

where the polynomials $h_i, i = 0, ..., N$, denote the Lagrange interpolants. In Fig. 1.1, one can observe the spurious oscillations that exist surrounding the discontinuity for $M = 1000$ when $N = 10$ and $N = 100$. As N rises from 10 to 100, it can be observed that the oscillation frequency rises around the discontinuity. External to the discontinuity, it can be observed the oscillation amplitude declines in line with the increase in N. Hence, the Gibbs phenomenon becomes increasingly local as N becomes larger (see Fig. 1.1). This suggests that if we allow $N \to \infty$, this would ultimately lead to a convergence to the solution.



(a) $N = 10$ (b) $N = 100$

Figure 1.1: Spectral element interpolation of a discontinuous function on a grid of uniformly spaced points for $M = 1000$. The red line is the function f and the blue line is the interpolant.

While the above example is relatively simplistic, it is illustrative of problems that arise in relation to the use of spectral techniques when approximating a discontinuous function. The spurious oscillations have the potential to contaminate the additional variables within the computation. Such a phenomenon can additionally be observed in finite elements. Nevertheless, the severity is lower compared with spectral techniques as a result of the lower order polynomial interpolants. It could be argued that it is necessary to conform the discontinuity to the mesh. Nevertheless, this could make

the time required for computation significantly longer, specifically in situations where the discontinuity is permitted free movement inside the computational domain. In the following part, a technique that can mitigate these oscillations will be discussed.

1.4 eXtended Spectral Element Method

The approximation of a discontinuity combined with the computational mesh through the application of continuous polynomials could lead to spurious oscillations around the discontinuity. These could diffuse throughout the computational domain, thus contaminating the approximation distant from the discontinuity. In terms of the finite elements, it was proposed by Moes et al. [8, 68] that a method defined as the Extended Finite Element Method (XFEM) be adopted to resolve this issue. In situations where there is a strong discontinuity (as is the case in this thesis), the fundamental idea underlying XFEM, according to the specific formal definition, involves boosting the initial finite element space of admissible functions via a discontinuous event, this allows the discontinuity to be captured by the numerics, thus obtaining the optimum order of convergence for functions that are less regular. In the context of the present thesis, this approach is adapted for spectral elements, with the consequence that is renamed the eXtended Spectral Element Method (XSEM) and the objective is to optimise the convergence order in terms of the polynomial degree, based on which it will be possible to deduce spectral precision for functions containing discontinuities.

The rationale behind the below example is to numerically analyse the eXtended Spectral Element Method (XSEM). Firstly, the approximation of a discontinuous function is considered.

Let $\Omega = [-1,1]$ include two sub-domains $\Omega_1 = [-1, 10^{-4})$ and $\Omega_2 = [10^{-4}, 1]$ such that

$$\Gamma = \partial\Omega_1 \cap \partial\Omega_2 = \{10^{-4}\}$$

Represents the interface between the pair of regions. Γ stands for the discontinuity contained within the function. Reconsider the discontinuous function $f : \Omega \to \mathbb{R}$

defined by:

$$f(x) = \begin{cases} x+1 & \forall x \in \Omega_1 \\ x & \forall x \in \Omega_2 \end{cases} \tag{1.4}$$

For the purposes of simplification, only one element is considered. Hence, the spectral element and extended spectral element approximations, represented by f_N and f_N^Γ , respectively, on domain Ω are formulated by

$$\begin{aligned} f_N(x) &= \sum_{i=0}^{N} f_i h_i(x) \\ f_N^\Gamma(x) &= \sum_{i=0}^{N} f_i h_i(x) + \sum_{i=0}^{N} \alpha_i h_i(x) \phi_i(x) \end{aligned} \tag{1.5}$$

where $f_i = f(x_i), i = 0,...,N, h_i$ denote the Lagrange interpolants defined in (A.3), α_i represent the additional degrees of freedom resulting from the enrichment and ϕ_i is the enrichment function formulated as :

$$\phi_i(x) = H(x) - H(x_i) \tag{1.6}$$

where $H(x)$ denotes the Heaviside step-function formulated as :

$$H(x) = \begin{cases} 0 & \forall x \in \Omega_1 \\ 1 & \forall x \in \Omega_2 \end{cases} \tag{1.7}$$

The coefficients α_i are totally unknown. Consider a grid with uniform spacing formulated as:

$$D_u = [x_0, x_1] \cup [x_1, x_2] ... \cup [x_{M-1}, x_M] = \bigcup_{k=1}^{M} [x_{k-1}, x_k]$$

where $x_0 = -1$ and $x_M = 1$ and M denote the overall number of points with uniform spacing.

For the purpose of calculation, it is assumed that $f_N^\Gamma(x_k) \equiv f(x_k), \forall x_k \in D_u$. The α_i's can therefore be calculated based on residual of the standard SEM approximation:

$$\sum_{i=0}^{N} h_i(x_k) \phi_i(x_k) \alpha_i = \sum_{i=0}^{N} B_{ki} \alpha_i = F(x_k) = f(x_k) - \sum_{i=0}^{N} f_i h_i(x_k) \tag{1.8}$$

where the matrix entries are given by $B_{ki} = h_i(x_k)\phi_i(x_k)$.It should be noted that the number of points with uniform spacing will normally be greater than the polynomial degree ($M > N$), hence, the matrix B is not square, but its size is $(M+1) \times (N+1)$. Thus, it can be inverted through a process in which it is multiplied by its transpose B^T to generate a square matrix $B^T B$ with size $(N+1) \times (N+1)$ and can then be inverted :

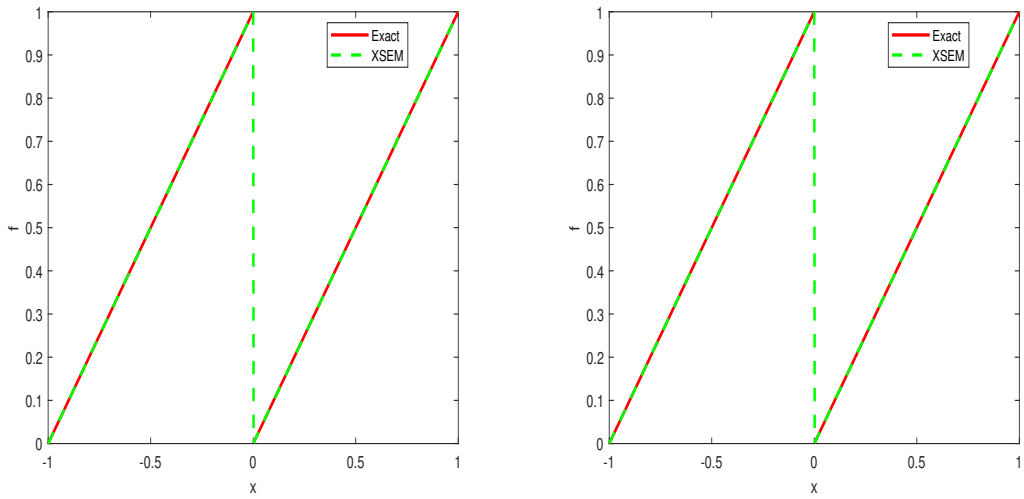
$$B^T B\alpha = B^T F \tag{1.9}$$

As N rises the conventional SEM approximation approaches the precise solution, meaning that the right-hand side of (1.8) will converge to zero. Consequently, the matrix B will become more singular in line with the increase in N. Deductively, this outcome can be interpreted as a lowering in the amount of enrichment necessary

Table 1.1 shows the convergence orders for both SEM and XSEM in terms of theL^2 norm. For the calculation of the L^2 norm, due to the fact that high order polynomials are being used, the Gauss-Lobatto-Legendre (GLL) quadrature is considered. Function $f$is interpolated utilising SEM and XSEM on a finer grid that has an order$M = 1000$. The resulting grid is subsequently employed in the quadrature for the L^2 norm. It is evident that SEM experiences difficulties with obtaining an analytical solution as a result of the existence of the discontinuity (see Fig 1.3). It can clearly be observed that there are oscillations local to the discontinuity (where $N = 10$). In Table 1.1 , the convergence orders for both the SEM and XSEM approximation of discontinuous $f$are shown. First, it is clear that the SEM approximation experiences problems and the error in fact grows when $N = 8, \ldots, 64$ (see second column of Table 1.1). It is probable that the oscillations observed in Fig 1.3 are caused by the Gibbs phenomenon. The convergence order for the XSEM approximation of a discontinuous function is identical to that obtained when it approximates a continuous function. This exemplifies the strength of an enriched method. For functions with reduced regularity, the desired high convergence order can be maintained.

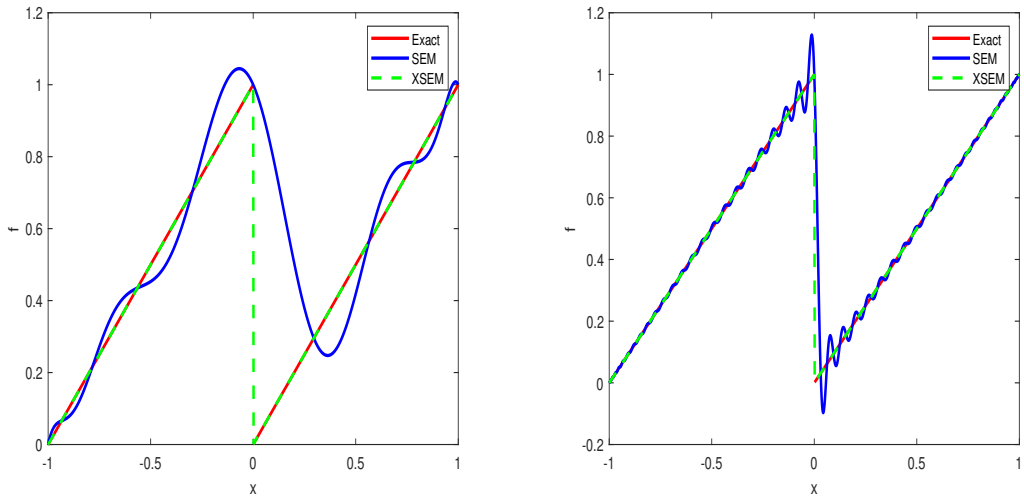
N	$\|f - f_N\|_{L^2(\Omega)}$	$\|f - f_N^\Gamma\|_{L^2(\Omega)}$
1	18.2574	5.740×10^{-15}
2	29.4052	3.884×10^{-14}
4	22.2915	2.958×10^{-13}
8	16.4059	2.613×10^{-10}
16	11.8316	3.767×10^{-7}
32	8.4116	1.786×10^{-7}
64	5.9035	1.290×10^{-7}

Table 1.1: Convergence of the error with respect to N for both SEM and XSEM.



(a) $N = 10$ (b) $N = 100$

Figure 1.2: XSEM interpolation of a discontinuous function on a grid of uniformly spaced points for $M = 1000$ and $N = 10$. The green line is the XSEM interpolant, the red line (underneath the green line) is the function f.



(a) $N = 10$ (b) $N = 100$

Figure 1.3: Comparison of the SEM and XSEM approximation (f_N and f_N^Γ) of discontinuous f against the analytical solution for varying values of N.

In the present thesis, a spectral element equivalent of the *hp* eXtended Finite Element Method (XFEM) is developed, which is defined as the eXtended Spectral Element Method (XSEM). A certain amount of the oscillations generated by the Gibbs phenomenon are eliminated via the XSEM. In combination with certain iterative methods (Preconditioned Conjugate Gradient), this approach was implemented in both Poisson and Stokes equations. Spectral element error estimates are derived and the convergence of the approximations is investigated.

Chapter 2

Spectral element method (SEM)

2.1 Introduction

Spectral element techniques, which were introduced by Patera (1984), are high order weighted residual methods that can be applied to solve partial differential equations. They amalgamate both the geometric flexible *h*-type (Axelsson and Barker(1984), Cuvelier et al.(1986), Girault and Raviart(1986)), and *p*-type, (Babuska and Dorr(1981), Babuska et al.(1981)) finite element techniques with spectral methods, and have the potential to yield high levels of precision (Canuto et al.(1988), Gottlieb and Orszag(1977), Gottlieb et al.(1984)). Subsequently, Maday and Patera (1989) developed this approach and presented a theoretical basis for the method.

In the spectral element discretisation, the domain is decomposed into a finite number of spectral elements. In each element, the dependent variables are approximated using high order polynomial expansions. With the aim of decreasing the need for explicitly imposing continuity at the interfaces of the elements, we write the partial differential equation in its equivalent variational form. Thus, this variational formulation is assumed to be the foundation of the discretisation procedure, in which an important role is played by high order Gauss-type numerical quadrature (Davis and Rabinowitz [**?**]). For *p*-covergence we expect the discrete approximation to converge as the order of polynomial approximations increases while keeping the number of elements constant.

The spectral element decomposition allows problems that incorporate geometries with high complexity to be approximated accurately. Hence, spectral element techniques can deal with complexities that are both geometric and physical in nature. Due to the fact that spectral element estimation can be regarded as a spectral collocation technique for fixed elements, the rates of convergence generated by spectral element techniques are comparable to those produced by spectral techniques and are dependent on the smoothness properties of the function that will be approximated. Therefore, in the case of analytic solutions, spectral element techniques achieve either spectral or exponential convergence, which means they are highly compatible for problems where high order regularity is not excluded, such as incompressible fluid mechanics (Korczak and Patera (1986), Ronquist and Patera (1988)). In the event that the regularity of the solution is diminished, spectral element techniques may become less desirable in comparison to h-type finite element techniques. A similar assertion can be made when the allowable level of error is comparably elevated.

2.2 Some examples of spectral methods

Spectral techniques are not only characterised on the basis of the method type (i.e., Galerkin, collocation or tau), but also the specific selection of the trial functions. The most widely utilised types of trial function are trigonometric, Chebyshev and Legendre. In the following sections, the fundamental tenets of the Legendre technique and the basic attributes of the polynomial set will be presented through an in-depth examination of a specific spectral technique based on its application to common kinds of equations.

2.2.1 The 1D spectral element mesh

Basis functions that are defined over the entire domain are impractical for complex geometries (such as the propagation of seismic waves inside a sedimentary basin) or models that do not have continuous physical attributes (such as the wave equation in layered media). The spectral element technique, which largely mirrors FEM, involves

the decomposition of the domain into separate elements, followed by the application of the spectral technique inside each of the elements. The continuity of the global approximation across inter-element borders can be accomplished in an efficient manner by imposing continuity on the interface points. The domain $[a,b]$ is divided into n_e elements $E_j = [x_{j-1}, x_j], j = 1, ..., n_e$, with $x_0 = a$ and $x_{n_e} = b$. Additionally, a mapping χ^j is determined which maps every element ($x \in E_j$) into the reference element ($\xi \in [-1,1]$):

$$x = \chi^j(\xi) = \frac{(1-\xi)}{2} x_{j-1} + \frac{(1+\xi)}{2} x_j, \quad \xi \in [-1,1]. \tag{2.1}$$

$$\xi = (\chi^j)^{-1}(x) = \frac{2x - (x_{j-1} + x_j)}{x_j - x_{j-1}}, \quad x \in E_j. \tag{2.2}$$

Each element is discretised using a Gauss-Lobatto Legendre (GLL) sub-grid. The i-th GLL node of the j-th element is located at

$$x_i^j = \chi^j(\xi_i). \tag{2.3}$$

The non-redundant list of these nodes (i.e. inter-element nodes counted only once) form a set of $n_e \times N + 1$ global nodes

$$x_I = x_i^j \quad \text{with} \quad I = \mathscr{I}(i,j) = (j-1)N + i, \quad \begin{array}{r} i = 1, \cdots, N, j = 1, \cdots, n_e - 1; \\ i = 1, \cdots, N+1, j = n_e. \end{array} \tag{2.4}$$

The table $\mathscr{I}(i,j)$ is the local-to-global index map table.

2.2.2 The assembly operator

Consider a set of local quantities defined on an element-by-element basis: a set of local vectors $\{\mathbf{a}^j\}_{j=1}^{n_e}$ each of size $N+1$, or a set of local matrices $\{A^j\}_{j=1}^{n_e}$ each of size $(N+1) \times (N+1)$. To assemble the global quantity we need to add the local contributions from each element to form the global array. The assembled vector $\boldsymbol{a}$ of size $Nn_e + 1$, by definition has the following components

$$a_I = \sum_{(i,j),\, I\equiv(i,j)} a_i^j \;=\; \begin{cases} a_i^j & \text{if } I \equiv (i,j) \text{ with } i \in [1, N-1] \\ & \text{(if } I \text{ is an interior node)} \\ a_N^{j-1} + a_0^j & \text{if } I \equiv (N, j-1) \equiv (0, j) \\ & \text{(if } I \text{ is a boundary node)} \end{cases} \tag{2.5}$$

The components of an assembled matrix A of size $(Nn_e+1)\times(Nn_e+1)$, are

$$A_{PQ} = \sum_{(p,q,j),\, P\equiv(p,j),\, Q\equiv(q,j)} A_{pq}^j \;=\; \begin{cases} A_{pq}^j & \text{if } P \equiv (p,j) \text{ with } p \in [1, N-1] \\ & \text{if } Q \equiv (q,j) \text{ with } q \in [1, N-1] \\ & \text{(if } I \text{ or } J \text{ are interior nodes)} \\ A_{N0}^{j-1} + A_{0N}^j & \text{if } P = Q \equiv (N, j-1) \equiv (0, j) \\ & \text{(if } P = Q \text{ and is a boundary node)} \end{cases} \tag{2.6}$$

For $n_e = 2$ and $N = 3$, the assembled matrix has the following form :

$$A = \begin{pmatrix} \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & 0 & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & \bullet & \bullet & \bullet & \bullet \end{pmatrix} \tag{2.7}$$

2.2.3 Basis functions

A set of global basis functions is defined by gluing together the spectral basis functions based on the GLL nodes of each element:

$$h_I(x) = \begin{cases} h_i^k(x) & \text{if } I \equiv (i,k) \text{ and } x \in E_k \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

where

$$h_i^k(x) = h_i[(\chi_k)^{-1}(x)], \quad \text{for } x \in E_k$$

These basis functions are continuous across inter-element boundaries. This efficiently enforces the continuity of the approximation.

2.2.4 One-dimensional Model Problem

Let $f \in L^2([a,b])$ and $\alpha, \beta, \gamma, a_1, a_2 \in \mathbb{R}$. We are seeking a solution u of the following second-order boundary value problem with Dirichlet boundary conditions:

$$\begin{cases} -\alpha u'' + \beta u' + \gamma u = f, \; x \in [a,b], \\ \\ u(a) = a_1, \quad u(b) = a_2. \end{cases} \tag{2.9}$$

2.2.5 Weak formulation

In order to outline the spectral element method, we first start with the variational formulation of the problem. The solution u is sought in the following admissible space:

$$\mathscr{U} = \left\{ u \in H^1([a,b]); \quad u(a) = a_1, \, u(b) = a_2 \right\}.$$

where H^1 is the classical Sobolev space that denotes the space of square-integrable functions with square-integrable generalized first derivatives. The test-functions are choosen in $\mathscr{V} = H_0^1([a,b])$. We define the L^2-scalar product

$$(u,v) = \int_a^b u(x)\, v(x)\, dx$$

Multiply the first equation of (2.9) by a test function $v \in \mathscr{V}$ integrate by parts, then one can easily obtain the weak form :

$$\begin{cases} \text{Find} \quad u \in \mathscr{U} \quad \text{such that} \\ \\ a(u,v) = (f,v), \qquad \forall v \in \mathscr{V} \end{cases} \tag{2.10}$$

where the bilinear form $a(.,.)$ is defined by

$$a(u,v) \;\; = \;\; \alpha(u',v') + \beta(u',v) + \gamma(u,v)$$

Discrete problem

Let $u_i \approx u(x_i)$. Then, in the case of a single element, the solution u is expanded in terms of the Lagrange interpolants based on the Gauss-Lobatto Legendre (GLL) points

i.e.

$$u_N(x) = \sum_{k=0}^{N} u_k h_k(x) \tag{2.11}$$

where

$$h_k(x) = \prod_{i=1, i\neq k}^{N+1} \frac{(x-x_i)}{(x_k-x_i)}, \quad k = 1,\ldots,N+1, \tag{2.12}$$

is defined on the set of interpolation nodes $\{x_k\}_{k=1}^{N+1}$. Let $\mathbb{P}_N([a,b])$ be the set of polynomials of degree N defined on $[a,b]$.

The Galerkin approximation is to solve the discrete weak problem:

Find $u_N \in \mathscr{U}^N = \{u_N \in \mathbb{P}_N([a,b]); \quad u_N(a) = a_1,\ u_N(b) = a_2\}$ such that

$$\alpha(u_N', v_N')_N + \beta(u_N', v_N)_N + \gamma(u_N, v_N)_N = (f, v_N)_N,\ \forall v_N \in \mathscr{U}^N \cap H_0^1([a,b]) \tag{2.13}$$

where the discrete inner product $(\cdot,\cdot)_N$ is defined by

$$(\phi, \psi)_N = \sum_{i=0}^{N} w_i \phi(x_i)\psi(x_i)$$

where the weights for Legendre-Gauss-Lobatto numerical integration are given by:

$$w_i = \frac{2}{N(N+1)} \frac{1}{L_N^2(x_i)}, i = 1,\ldots,N+1.$$

By injecting the approximation u_N given by (2.11) into the variational formulation (2.13) and by choosing the test functions as $v_N = h_j \ \ (1 \leq j \leq N-1)$, one obtain the following discrete system :

$$\alpha \sum_{k=0}^{N} u_k(h_k', h_j')_N + \beta \sum_{k=0}^{N} u_k(h_k', h_j)_N + \gamma \sum_{k=0}^{N} u_k(h_k, h_j)_N \quad = (f, h_j)_N, \quad j = 1,\ldots,N-1.$$

2.2.6 Legendre differentiation matrix

Define D to be the so-called Legendre differentiation matrix of dimension $(N+1) \times (N+1)$ given by $D_{ij} = h_j'(x_i)$, and with entries given explicitly by

$$\begin{cases} D_{00} = -\dfrac{N(N+1)}{4} \\ D_{NN} = \dfrac{N(N+1)}{4} \\ D_{ii} = 0,\ i = 1,\ldots,N \\ D_{ij} = \dfrac{L_N(x_i)}{L_N(x_j)(x_i - x_j)},\ i,j = 0,\ldots,N, i \neq j \end{cases}$$

2.2.7 Stiffness matrix

The stiffness matrix is full with entries given by

$$\begin{aligned} A_{ij} &= \alpha(h_i', h_j')_N + \beta(h_i', h_j)_N + \gamma(h_i, h_j)_N \\ &= \alpha \sum_{p=0}^{N} w_p h_i'(\xi_p) h_j'(\xi_p) + \beta \sum_{p=0}^{N} w_p h_i'(x_p) h_j(x_p) + \gamma \sum_{p=0}^{N} w_p h_i(x_p) h_j(x_p) \\ &= \alpha \sum_{p=0}^{N} w_p D_{pi} D_{pj} + \beta \sum_{p=0}^{N} w_p D_{pi} \delta_{pj} + \gamma \sum_{p=0}^{N} w_p \delta_{pi} \delta_{pj} \\ &= \alpha \sum_{p=0}^{N} w_p D_{pi} D_{pj} + \beta w_j D_{ji} + \gamma w_i \delta_{ij}. \end{aligned}$$

2.2.8 Elemental mass matrix

A major practical advantage is that the mass matrix is diagonal by construction

$$\begin{aligned} M_{ij} &= (h_i, h_j)_N \\ &= \sum_{p=0}^{N} w_p h_i(\xi_p) h_j(\xi_p) \\ &= \sum_{p=0}^{N} w_p \delta_{ip} \delta_{jp} \\ &= w_i \delta_{ij}. \end{aligned}$$

2.2.9 Right-hand side

The right-hand side is given by

$$\begin{aligned} F_j &= (f, h_j)_N \\ &= \sum_{p=0}^{N} w_p f(\xi_p) h_j(\xi_p) \\ &= \sum_{p=0}^{N} w_p f(\xi_p) \delta_{jp} \\ &= w_j f(\xi_j). \end{aligned}$$

Then the discrete problem is given by :

$$A\boldsymbol{U} = \boldsymbol{F} \tag{2.14}$$

where $\boldsymbol{U}$ is the unknown vector of components $(u_k)_{1\leq k\leq N-1}$ and $\boldsymbol{F}$ is the vector of components $F_j = (f,h_j)_N = \sum_{i=0}^{N} w_i f(x_i)h_j(x_i) = w_j f(x_j), 0 \leq j \leq N$. Let W be the diagonal matrix of element w_i written as

$$\begin{pmatrix} w_0 & 0 & . & 0 \\ 0 & w_1 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & w_N \end{pmatrix}$$

Then the matrix A is given by

$$A = \alpha D^T W D + \beta W D + \gamma W$$

where D is the Legendre differentiation matrix and D^T is the transpose of D.

2.3 Numerical tests

In this section, various examples of spectral element simulations will be given in order to numerically show the expected spectral precision. Firstly, an application of the one-dimensional boundary value problem (2.9).

Let $a = -1, b = 1, \alpha = \gamma = 1$ and $\beta = 0$. We take $f(x) = (1+\pi^2)\sin(\pi x)$ so that the exact solution that satisfies $u(\pm 1) = 0$ is $u(x) = \sin(\pi x)$.

A Matlab code was written to calculate the solution of the problem (2.13) by solving the discrete problem (2.14) satisfying the boundary conditions.
The L^2- norm of the error is tabulated as a function of N in Table 2.1.
We can see a rapid convergence with machine precision error is obtained for $N = 18$ in Fig. 2.1.

N	L^2 error
3	0.355
4	0.051
6	8.669×10^{-4}
8	1.242×10^{-5}
10	$1,446 \times 10^{-7}$
12	1.345×10^{-9}
14	$1,012 \times 10^{-11}$
16	6.267×10^{-14}
18	$1,085 \times 10^{-15}$

Table 2.1: L^2- norm of the error as a function of N.

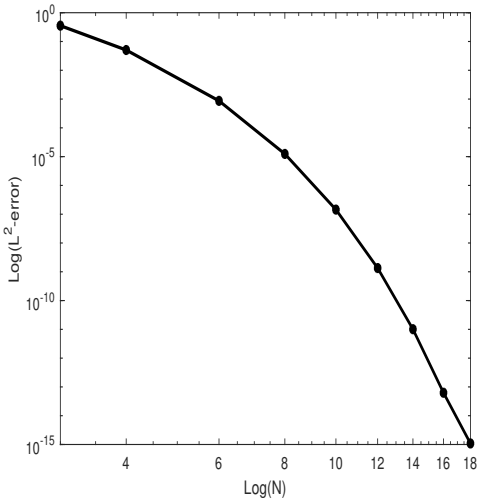


Figure 2.1: L^2 error convergence with respect to polynomial order

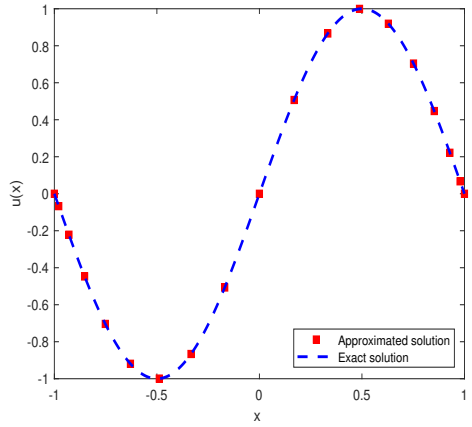


Figure 2.2: Exact and approximated solutions for $N = 18$

Let us now solve a second equation in one dimension, with mixed boundary conditions. We seek the solution of

$$\frac{d^2u(x)}{dx^2} = f(x),$$

in the region $-1 \leq x \leq 1$, with $f(x) = x^3$ with homogenous Dirichlet boundary conditions at $x_0 = -1$ and $x_{N+1} = 1$. Note that the exact solution can be calculated analytically and is given by $u(x) = \frac{x}{20} - \frac{x^5}{20}$. Again, a rapid convergence with machine precision error is obtained for $N = 8$. The exact and approximate solution are plotted for $N = 20$ in Fig. 2.3.

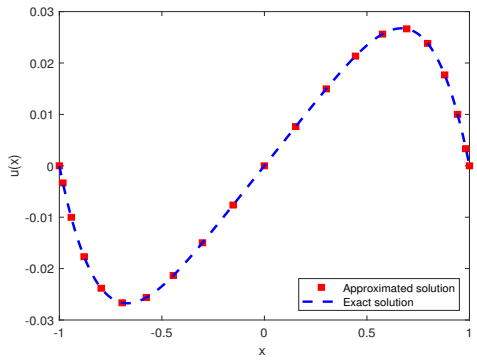


Figure 2.3: Exact and approximate solution for $N = 20$

Let us now reconsider the first example where we seek the solution of

$$-\frac{d^2u(x)}{dx^2} + u(x) = f(x),$$

in the region $-1 \leq x \leq 1$, with $f(x) = (1+\pi^2)\sin(\pi x)$ with homogenous Dirichlet boundary conditions. The domain is divided into more than one element. Again, a rapid convergence with machine precision error is obtained for $N = 8$ with 2 and 3 elements, see Table 2.2. The exact and approximate solution are plotted for $N = 16$ in Fig. 2.4 for 2 elements (left) and 3 elements (right).

N	L^2 error 1 element	L^2 error 2 elements	L^2 error 3 elements
3	0.355	8.25×10^{-3}	8.84×10^{-4}
4	0.051	1.07×10^{-3}	6.95×10^{-5}
6	8.67×10^{-4}	6.68×10^{-4}	1.92×10^{-7}
8	1.24×10^{-5}	3.20×10^{-8}	4.05×10^{-10}
10	1.45×10^{-7}	1.15×10^{-10}	6.44×10^{-13}
12	1.35×10^{-9}	3.16×10^{-13}	5.31×10^{-15}
14	$1{,}01 \times 10^{-11}$	3.14×10^{-15}	4.46×10^{-15}
16	6.27×10^{-14}	5.18×10^{-15}	2.16×10^{-14}
18	$1{,}09 \times 10^{-15}$	5.56×10^{-15}	1.30×10^{-14}

Table 2.2: L^2- norm of the error as a function of N.

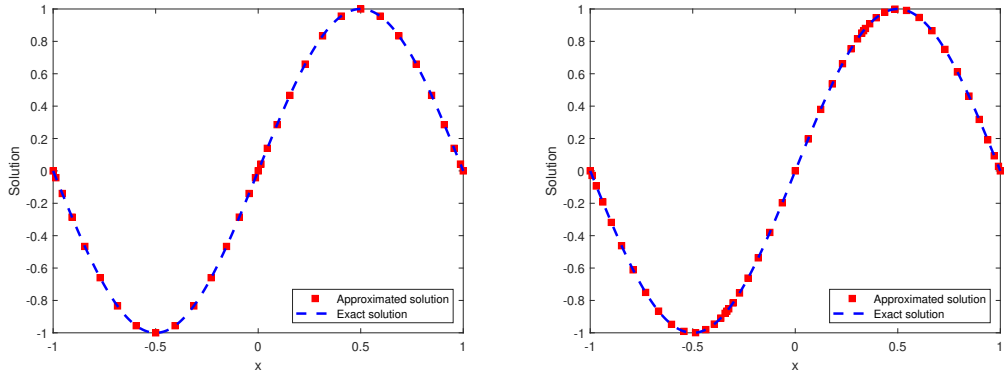


Figure 2.4: Exact and approximate solution for $N = 16$ with 2 elements(left) and 3 elements (right)

2.4 Conclusion

Based on the findings in this chapter, it can be concluded that SEM amalgamates the positive attributes of both the domain decomposition technique (finite element method) and spectral techniques. The spectral element technique utilises a variational formulation equivalent to the original partial differential equation. Legendre Gauss-type quadrature is used to approximate the corresponding variational problem. The specific formulation guarantees that for smooth problems, SE estimation accomplished exponential convergence to the exact solution of the given problem with comparatively few degrees of freedom. Numerical examples were considered to show the outcomes for linear elliptic equations. The L^2-error converges exponentially as can be seen in Fig. 2.1.

Chapter 3

Approximation of Poisson equation using SEM

The Poisson equation

$$-\nabla \cdot (\mu \nabla u) = f \quad \text{on} \quad \Omega \tag{3.1}$$

is the simplest and most famous elliptic partial differential equation. The source (or load) function f is given on some two- or three-dimensional domain denoted by $\Omega \in \mathbb{R}^2$ or $\mathbb{R}^3$. A solution u satisfying (3.1) must also satisfy boundary conditions on the boundary $\partial\Omega$ of Ω; for example the most general form of which is given by

$$\alpha u + \beta \frac{\partial u}{\partial n} = g \quad \text{on} \quad \partial\Omega \tag{3.2}$$

where $\dfrac{\partial u}{\partial n}$ stands for the directional derivative normal to the boundary $\partial\Omega$ (traditionally in an outwards direction) and both α and β are constants, while it is also possible to have coefficients that have variability. In practice, u could denote the temperature field in Ω that is subjected to the source of heat, f. Additional physical models that have importance are gravitation, electromagnetism, elasticity and inviscid fluid mechanics, see Ockendon et al. [74, chap. 5].

When (3.1) and (3.2) are combined, this is defined as a boundary value problem. Where the constant β in (3.2) equals zero, the type of boundary condition is a Dirichlet, while the boundary value problem is considered to be the Dirichlet problem for the Poisson equation. On the other hand, where the constant α equals zero, then it is

accordingly considered to be a Neumann boundary condition and therefore a Neumann problem. A third alternative is that the Dirichlet conditions are satisfied on a section of the boundary $\partial\Omega_d$, while the Neumann conditions (or in fact blended conditions in which both α and β are not zero) are satisfied for the rest $\partial\Omega_n = \partial\Omega \setminus \partial\Omega_d$.

3.1 Single-domain Discretizations

Let $\Omega = [x_a, x_b] \times [y_a, y_b], f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega)$. We seek a solution u of the following system :

$$\begin{cases} -\nabla \cdot (\mu \nabla u) = f \quad \text{on} \quad \Omega \\ \\ u = g \quad \text{on} \quad \partial\Omega \end{cases} \tag{3.3}$$

where $\mu(x,y)$ is a continuous function on Ω.

3.1.1 Weak formulation

In order to outline the spectral element method, we first start with the variational formulation of the problem (3.3) for the case when $\Omega = [-1,1] \times [-1,1]$ and $g = 0$. The solution u is sought in $H_0^1(\Omega)$ and then one can easily obtain the weak form :

$$\begin{cases} \text{Find} \quad u \in H_0^1(\Omega) \quad \text{such that} \\ \\ a(u,v) = (f,v), \qquad \forall v \in H_0^1(\Omega) \end{cases} \tag{3.4}$$

where the bilinear form $a(.,.)$ is defined by

$$a(u,v) \quad = \int_\Omega \mu \nabla u \cdot \nabla v$$

and the linear form is just the L^2 inner product

$$(f,v) \quad = \quad \int_\Omega fv$$

3.1.2 Discrete problem

Let N denote the degree of polynomial interpolation and let (x_i, y_j), $i, j = 1, \ldots, N+1$ denote the 2D GLL grid formed by the tensor product of the 1D GLL grids in the x and y directions.

The weights for Legendre-Gauss-Lobatto numerical integration are given by:

$$w_i = \frac{2}{N(N+1)} \frac{1}{L_N^2(x_i)}, \quad i = 1, \ldots, N+1,$$

so that

$$\int_{-1}^{1} f(x)dx \approx \sum_{i=1}^{N+1} w_i f(x_i)$$

with equality if f is a polynomial of degree $2N-1$ or less.

Denote by $u_{ij} = u_N(x_i, y_j), f_{ij} = f(x_i, y_j)$ and $\mu_{ij} = \mu(x_i, y_j), i, j = 1, \ldots, N+1$, where u_N is the approximation to u expanded in terms of the Lagrange interpolants based on the Gauss-Lobatto Legendre points, i.e.

$$u_N(x,y) = \sum_{i,j=1}^{N+1} u_{ij} h_i(x) h_j(y) \tag{3.5}$$

where h_i are the 1D Lagrange interpolants defined by

$$h_i(x) = \prod_{k=1, k\neq i}^{N+1} \frac{(x - x_k)}{(x_i - x_k)}, \quad i = 1, \ldots, N+1 \tag{3.6}$$

defined on the set of interpolation nodes $\{x_i\}_{i=1}^{N+1}$.

The Galerkin approximation is to solve the discrete weak problem:
Find $u_N \in \mathscr{V}^N$ such that

$$\Big(\mu \nabla u_N, \nabla v_N\Big)_N = \Big(f, v_N\Big)_N, \ \forall v_N \in \mathscr{V}^N$$

where the discrete inner product $\Big(.,.\Big)_N$ is defined by

$$\Big(\varphi, \psi\Big)_N = \sum_{m,n=1}^{N+1} w_m w_n \varphi(x_m, y_n) \psi(x_m, y_n)$$

Note that

$$\nabla u_N(x,y) = \begin{pmatrix} \sum_{i,j=1}^{N+1} u_{ij} h_i'(x) h_j(y) \\ \\ \sum_{i,j=1}^{N+1} u_{ij} h_i(x) h_j'(y) \end{pmatrix}$$

and by choosing the test functions as $v_N(x,y) = h_k(x)h_l(y)$ $(2 \le k,l \le N)$, one obtains

$$\nabla v_N(x,y) = \begin{pmatrix} h_k'(x) h_l(y) \\ \\ h_k(x) h_l'(y) \end{pmatrix}.$$

By substituting the approximation u_N given by (3.5) and the chosen test function into the variational formulation (3.7), one obtains the following discrete system of equations for the unknowns $u_{ij}, 2 \le i,j \le N$:

$$\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=1}^{N+1} \Big(h_i'(x_m) h_j(y_n) h_k'(x_m) h_l(y_n) + h_i(x_m) h_j'(y_n) h_k(x_m) h_l'(y_n) \Big) u_{ij} = \sum_{m,n=1}^{N+1} w_m w_n f_{mn} h_k(x_m) h_l(y_n), \quad 2 \le k,l \le N,$$

Since u_N satisfies homogenous Dirichlet boundary condition the expression (3.5) can be simplified to

$$u_N(x,y) = \sum_{i,j=2}^{N} u_{ij} h_i(x) h_j(y) \tag{3.7}$$

and the discrete system is equivalent to

$$\sum_{i,j=2}^{N} u_{ij} \sum_{m,n=1}^{N+1} \Big(h_i'(x_m) h_j(y_n) h_k'(x_m) h_l(y_n) + h_i(x_m) h_j'(y_n) h_k(x_m) h_l'(y_n) \Big) w_m w_n \mu_{mn} = \sum_{m,n=1}^{N+1} w_m w_n f_{mn} h_k(x_m) h_l(y_n)$$

Using the Kronecker delta property $h_i(x_m) = \delta_{mi}$ and defining $h_i'(x_m) = D_{mi}$ one deduces

$$\sum_{i,j=2}^{N} u_{ij} \sum_{m,n=1}^{N+1} \Big(D_{mi} \delta_{nj} D_{mk} \delta_{nl} + \delta_{mi} D_{nj} \delta_{mk} D_{nl} \Big) w_m w_n \mu_{mn} = \sum_{m,n=1}^{N+1} w_m w_n f_{mn} \delta_{mk} \delta_{nl}$$

Let $I = (l-2)(N-1)+(k-1)$ and $J = (j-2)(N-1)+(i-1)$. Then the system reduces to

$$\sum_{J=1}^{(N-1)^2} A_{IJ}u_J = F_I, \quad I = 1, \cdots, (N-1)^2. \tag{3.8}$$

where

$$A_{IJ} \quad = \delta_{lj}w_l \sum_{m=1}^{N+1} D_{mi}D_{mk}w_m\mu_{ml} + \delta_{ki}w_k \sum_{n=1}^{N+1} D_{nj}D_{nl}\mu_{kn}w_n$$

and

$$F_I = w_k w_l f_{kl}.$$

3.2 Multidomain Discretizations

Just over a decade after the ground-breaking studies of Orszag (1969) and Kreiss and Oliger (1972) on (global) spectral methods, the initial heuristic efforts were made in the attempts to combine spectral methods with domain decomposition techniques. Both Orszag (1980) and Morchoisne (1983) developed spectral domain decomposition approaches for solving basic elliptic problems on the basis of the stronger form (patching) of the equations. In both works, conventional spectral collation approaches were adopted inside the sub-domains as well as on the entire boundary of the domain. Nevertheless, Orszag (1980) utilised domains with no overlap and implemented conditions at every collocation node on the sub-domain interfaces to ensure that the solution and its standard derivative are continuous, while Morchoisne (1983) utilised domains with overlap (the Schwarz technique) and the solution continuity was only enforced at the internal boundaries of the sub-domain. Conversely, Patera (1984) made use of a weak (variational) calculation as the discretisation basis for his spectral element approach for decomposing the spectral domain. The primary incentives for developing spectral domain decomposition methods were to expand existing spectral methods to domains for which it was not possible to map the entire domain onto an individual reference domain in order to facilitate refinement of the local grid or potentially basic domains, to exploit

the distinct behaviours of the solution observed in various areas of the domain, and potentially to partially resolve some of the problems caused by the significant time-step constrains experienced with single-domain spectral methods utilising time discretization. An analogous viewpoint underpins the advancement of hp finite-element methods (see, for example Babuˇska and Suri (1994), Schwab (1998), Babuˇska and Strouboulis (2001), Melenk (2003)). When applied to spectral approximations, the domain decomposition technique enables the user to take advantage of local tensor-product bases.

In the following sections, we present a general overview of methods used for discretisation along with theoretical analysis of spectral methods used in complex geometrical problems. The scope of the thesis is predominantly restricted to model problems that are illustrative of the fundamentals and define the association between traditional spectral methods and their domain decomposition offshoots. Methods of domain decomposition that are founded on the weak formulation are more broadly used and their history has greater breadth and complexity. Resultantly, increased focus will be applied to these methods in comparison to those founded on the equation's strong form. Nevertheless, the similarities between the strong and weak approaches will also be identified. In-depth explanations of domain decomposition methods are presented in the works of Smith et al. (1996), Quarteroni and Valli (1999), Toselli and Widlund (2005), and Wohlmuth (2001). Researchers who have specifically focused on spectral element methods include Karniadakis and Sherwin (1999, 2005) and Deville, Fischer and Mund (2002). Readers of this thesis are recommended to depend on works like these to achieve a more comprehensive understanding of the various facets of this topic. In the following chapter, we will detail how it is possible to adapt high-order spectral methods for the approximation of differential problems that are located within a computational domain whose form is complex. More precisely, we will consider a domain $\Omega \subset \mathbb{R}^2$, which can be depicted as the union of sub-domains $\Omega_m, m = 1, \cdots, K$ (for a suitable integer $K \geq 2$). All of the sub-domains can be acquired via a process of mapping from a reference domain (alternatively known as a parent or master domain) $\hat{\Omega}$.

The division of Ω could be either geometrically conforming, in situations where either a vertex of whole face or edge is shared by the adjacent sub-domains, or geometrically non-conforming, in instances where there is not a complete match between interfaces. In the first case, a spectral element method (SEM) will be introduced. The solution continuity will be inferred by the selection of trial functions, while weak (integral formulation) will automatically account for the flux continuity, similar to finite-elements.

For the purposes of defining SEM on a multidimensional domain $\Omega \subset \mathbb{R}^2$, a division $T = \{\Omega_m\}$ of Ω is introduced. Every element Ω_m is acquired by transforming F_m from a reference (or parent) element $\hat{\Omega}$, which could take the form of either a reference 2D-cube (a square)

$$\hat{\Omega}_C = \{\hat{x} = (\hat{x}_1, \hat{x}_2) : -1 < \hat{x}_1, \hat{x}_2 < 1\} = (-1,1)^2,$$

or the reference 2D-simplex (a triangle)

$$\hat{\Omega}_S = \{\hat{x} = (\hat{x}_1, \hat{x}_2) : -1 < \hat{x}_1, \hat{x}_2, \hat{x}_1 + \hat{x}_2 < 0\}.$$

The transformation F_m is a bijection that is differentiable, where Ω_m denotes a quadrilateral element with straight edges, $F_m : \hat{\Omega}_C \mapsto \Omega_m$ represents a bilinear map.

For curvilinear elements, it is possible to construct the transformation F_m via the application of the Gordon–Hall map, in which the faces and edges are parameterised with polynomials that have identical degrees of freedom to those utilised in the construction of the SEM solution. If Ω_m denotes a simplex (a triangle or a tetrahedron) comprised of straight faces or edges, then $F_m : \hat{\Omega}_C^2 \mapsto \Omega_m$ defines the affine map

$$x = F_m(\hat{x}) = B_m \hat{x} + b_m,$$

where b_m is a vector with 2 components, whereas B_m is an invertible 2×2 matrix.

For every $\Omega_m \in T$, a basis $\{\hat{\phi}_i^{(m)}\}$ for $V_{N_m}(\Omega_m)$ is gained as the image of a appropriately selected boundary-adapted basis $\{\hat{\phi}_i\}$ of $\hat{P}_N$, that is,

$$\phi_i^{(m)} = \hat{\phi}_i \circ F_m^{-1}, \text{ or } \phi_i^{(m)}(x) = \hat{\phi}_i(\hat{x}) \text{ with } x = F_m(\hat{x})$$

A basis for the entire space X_δ can thus be acquired through the unification of the elemental basis functions on every element Ω_m to ensure that global continuity is achieved. Every basis function inside an element abruptly produces a global basis function by extending it by zero external to the element. Suitable matching of vertices and edges is performed to produce global basis functions.

The matching is relatively insignificant if, as is the case with the majority of SEM realisations, the degree of the polynomial on both contiguous domains is identical. To fix concepts, assume that there is a common edge shared by Ω_m and Ω_k , which can be denoted by $\Gamma_{km} = \partial\Omega_m \cap \partial\Omega_k$. If the nodal basis is utilised in every element, then the nodes on Γ_{km} are identical for each of the elements. Resultantly, the pair of basis functions (characteristic Lagrange polynomials) connected with the identical node on Γ_{km}, such as say $\psi_i^{(m)}$ existing in Ω_m and $\psi_j^{(k)}$ existing in Ω_k, intersect on Γ_{km}; therefore, they generate a continuous function throughout Γ_{km}. For $x \in \Gamma_{km}$, the function

$$\psi = \begin{cases} \psi_i^{(m)} & \text{in } \bar{\Omega}_m \\ \psi_j^{(k)} & \text{in } \bar{\Omega}_k \\ 0 & \text{elsewhere,} \end{cases}$$

represents the global basis function of the nodal type related to the node x. However, where x is a vertex, the global basis function can be acquired by unifying each of the local basis functions related to x and then expanding the resultant function by zero external to the area of elements that contain x. The process of constructing the global basis bears certain similarities when the type of the nodal bases is modal. The only nuance is that face or edge basis functions that possess identical wave numbers could still have conflicting signs as a result of the distinct local interface orientations; thus, it may be necessary to adjust the sign prior to unifying the local functions. The aforemen-

tioned matching process is also applicable if a distinct polynomial degree is allowed in each direction for every element, as long as there is agreement between the polynomial degrees across an interface among two elements. A basic example demonstrating this case is provided in Fig. 3.1

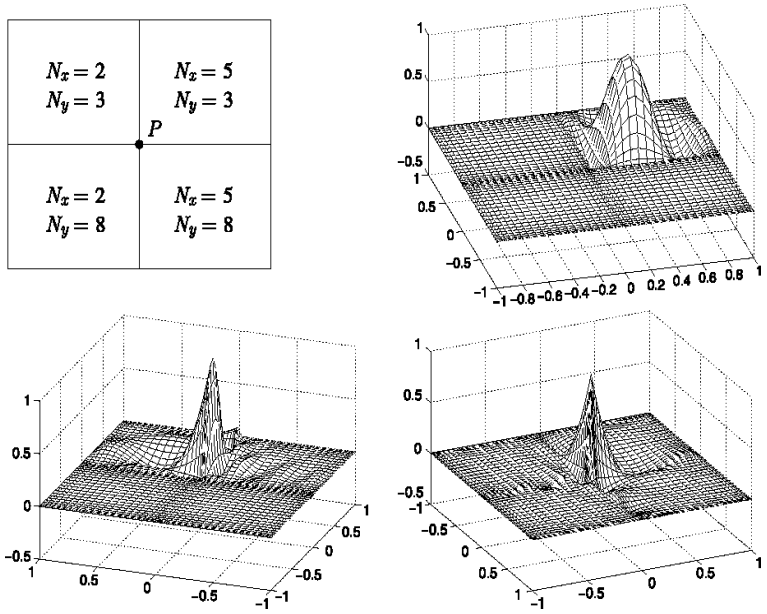


Figure 3.1: An example of spectral element discretization with different polynomial degrees in the different elements (top left) and three basis functions associated with an internal node (top right), an edge node (bottom left) and the cross-point P (bottom right)

A situation that has greater significance is one in which completely different polynomial degrees are permitted in neighbouring elements (unless we only add higher order polynomial bubbles to particular sub-domains). Continuity through an interface Γ_{km} among a pair of sub-domains Ω_k and Ω_m with, for example, $N_m < N_k$, necessitates that the limitation to Γ_{km} of all functions in $V_{N_k}(\Omega_k)$ must be (the image of) a polynomial with a degree of N_m. When the modal bases are utilised in every element, this necessity can easily be met: only the edge basis functions in Ω_m related to Γ_{km} make a contribution to the global basis functions; they are extended to the contiguous domain Ω_k as described above. In other words, one discards the contribution from the

edge basis functions in Ω_k associated with Γ_{km}, having wavenumber higher than N_m. Algebraically, the matching columns and rows of the local stiffness matrix of Ω_k are eliminated. Conversely, if one uses the nodal bases instead, the following process is applied: $\psi_i(m), i \in I$, is assumed to be the basis functions in Ω_m related to the nodes on Γ_{km}; $\pi_j(k), j \in J$ is defined in a similar manner for the additional domain. Due to the fact the constraint of $\psi_i(m)$ to Γ_{km} is (the image of) a polynomial with a degree $\leq N_m$, coefficients r_{ij} exist such that

$$\psi_i^{(m)} = \sum_{j \in J} r_{ij} \psi_j^{(k)} \text{ on } \Gamma_{km} \tag{3.9}$$

$\Big($ more precisely, $r_{ij} = \psi_j^{(m)}(x_j^{(k)})$, where $x_j^{(k)}$ denotes the node on $\Gamma_{km} \subset \Omega_k$ related to $\psi_j^{(k)}\Big)$. Hence, $\psi_j^{(m)}$is glued throughout Γ_{km} with the specific linear combination of edge basis functions in Ω_k by the right side of (3.9). Algebraically speaking, this amounts to the proper condensing of the stiffness matrix of Ω_k : the group of rows with indices$j \in J$ is substituted with their linear combinations with coefficients r_{ij} , for all $i \in I$; an analogous transformation process is utilized to the group of columns with indices $j \in J$. It can be observed that the matching process explained previously is merely a particular state of the mortar matching process. A necessary condition is that there is agreement between the pair of functions $v_m \in V_{N_k}(\Omega_m)$ and $v_k \in V_{N_k}(\Omega_k)$ such that

$$\int_{\Gamma_{km}} (v_m - v_k)\phi d\gamma = 0 \text{ for all } \phi \in Y_{km} \tag{3.10}$$

where Y_{km} represents an appropriate function space on Γ_{km}. When Y_{km} is the restriction space of $V_{N_k}(\Omega_k)$ on Γ_{km}, then there must be a coincidence between v_k and v_m on Γ_{km} (meaning that the degree of v_k is reduced to the same as v_m). In the event that Y_{km} coincides with the restriction space of $V_{N_k}(\Omega_m)$ on Γ_{km} (or potentially with a space that is even smaller), condition (3.10) would only infer that there is a continuity in the sense of least squares. In fact, the latter is the frequently used kind of matching in relation to the mortar method.

3.2.1 SEM Formulation

At this point, all the necessary components are present to establish the spectral element approximation related to boundary-value problems through the application of a multidomain discretisation. The weak formulation of (3.3) can be written in the following form:

$$\begin{cases} \text{Find} \quad u \in H_0^1(\Omega) \quad \text{such that} \\ \\ \sum_m a_{\Omega_m}(u,v) = \sum_m (f,v)_{\Omega_m}, \qquad \forall v \in H_0^1(\Omega) \end{cases} \tag{3.11}$$

where the bilinear form is

$$a_{\Omega_m}(u,v) \quad = \int_{\Omega_m} \mu \nabla u \cdot \nabla v d\Omega$$

and the linear form is

$$(f,v)_{\Omega_m} \quad = \quad \int_{\Omega_m} f v d\Omega$$

3.2.2 Algebraic Aspects of SEM

SEM produces an algebraic system in which the entries of the stiffness matrix A are:

$$A_{ij} = \sum_m a_{\Omega_m}(\phi_i^m, \phi_i^m)$$

It is possible to construct the matrix A by gathering the local stiffness matrices associated with each element Ω_m. If we consider the particular state where Ω is a square divided into K number of squares $\Omega_m, m = 1, \cdots, K$, of same size, as well as the same polynomial degree N is consistently utilised, this implies that $V_N(\Omega_m) = \mathbb{Q}_N$ is employed for each m.

Each of the elements contains boundary nodes that define their geometry and enable connectivity with adjacent elements. The process of gathering elements necessitates that the values of the main variables in nodes shared by neighbouring elements

are equal. The assemblage of specific sub-domains into the whole domain is a procedure called direct stiffness, which needs a universal or global system of nodes to be identified.

3.3 Numerical simulations

In the case of two-dimensional problems, the cost of using direct solvers on particularly fine meshes can be prohibitive. This is because of the costs associated with the assembly and factorisation of the global matrix. In such situations, it is necessary to use iterative methods. Such methods do not necessitate a large matrix to be stored or inverted. Rather, the fundamental computational procedure is founded on matrix-vector multiplications. There are numerous iterative methods that can be used to solve large systems of linear equations. Nevertheless, the critical factor is that a method must be found that is suitable for the given problem. An incorrect decision could cause slow convergence, or potential divergence.

There are numerous iterative methods that could be employed in the process of solving this system. The conjugate gradient (CG) method is the most traditional and celebrated member of the category of non-stationary iterative methods. In such methods, there is no element of choice in the determination of the iteration parameters. They are selected in a dynamic manner during every iteration in a way that minimises the error in a particular norm. The method was designed to solve symmetric, positive definite systems of linear equations. Convergence is accomplished in a rapid manner in situations where the eigenvalues of A are in a cluster or are located in specific clustered groups.

The rate of convergence for the CG method depends on the condition number of the coefficient matrix. If there is no clustering of the eigenvalues of A, then the CG method will produce slower convergence. This problem could be resolved by preconditioning the system using an appropriate non-singular matrix P. The main concept that under-

pins the preconditioning approach is that the initial system is transformed into a similar system with improved conditioning

$$P^{-1}A\boldsymbol{x} = P^{-1}\boldsymbol{b}, \tag{3.12}$$

where the preconditioner is defined as P. The preconditioner, P is selected to be an approximation to A, in a certain sense, which is less complex and more cost-effective to invert in comparison to A when the eigenvalues of $P^{-1}A$ are clustered close to unity. Preferably, the properties of the preconditioner should be analogous to those of the initial matrix and should additionally be sparse so that the efficiency of construction and storage is enhanced.

3.3.1 Preconditioned Conjugate Gradient method

In the field of mathematics, the conjugate gradient method represents an algorithm for numerically solving specific systems of linear equations, particular the ones that have a symmetric and positive-definite coefficient matrix, as can be observed here. The conjugate gradient method is frequently applied in the form of an iterative algorithm, which is suitable for sparse systems whose size restricts them from being addressed via a direct application or alternative direct techniques like the Cholesky decomposition. Large sparse systems frequently emerge in the numerical solutions of partial differential equations or problems involving optimisation.

In the majority of situations, preconditioning is required to ensure the conjugate gradient method rapidly converges. The Preconditioned Conjugate Gradient (PCG) consists of the steps shown below:

1. Choose an initial guess $\boldsymbol{x}_0$ and compute $\boldsymbol{r}_0 = \boldsymbol{b} - A\boldsymbol{x}_0$.

2. Solve $P\boldsymbol{z}_0 = \boldsymbol{r}_0$. Set $\boldsymbol{p}_0 = \boldsymbol{z}_0$.

3. For $n = 1, 2, ...$, compute

$$\begin{aligned}
\alpha_n &= \boldsymbol{r}_{n-1}^T \boldsymbol{z}_{n-1} / \boldsymbol{p}_{n-1}^T A \boldsymbol{p}_{n-1} \\
\boldsymbol{x}_n &= \boldsymbol{x}_{n-1} + \alpha_n \boldsymbol{p}_{n-1} \\
\boldsymbol{r}_n &= \boldsymbol{r}_{n-1} - \alpha_n A p_{n-1} \\
\boldsymbol{z}_n &= P^{-1} \boldsymbol{r}_n \\
\beta_n &= \boldsymbol{r}_n^T \boldsymbol{z}_n / \boldsymbol{r}_{n-1}^T \boldsymbol{z}_{n-1} \\
\boldsymbol{p}_n &= z_n + \beta_n \boldsymbol{p}_{n-1}
\end{aligned}$$

until convergence i.e. the stopping criterion is satisfied.

In our case, the preconditioner is chosen as follows: $P = A_{FEM}$ where A_{FEM} is the stiffness matrix calculated using the finite element method (FEM).

We choose forcing $f = 0$ and diffusivity such that

$$\mu(x,y) = \begin{cases} \mu_1, & y \leq \bar{y}, \\ \mu_2, & y > \bar{y}, \end{cases}$$

and $A_1, A_2, \eta_1, \eta_2, \mu_1$ and μ_2 are constants. The analytical solution is then

$$u(x,y) = \begin{cases} \sin(\pi x) A_1 \left(e^{\eta_1 y} - e^{-\eta_1 y} \right), & y \leq \bar{y}, \\ \sin(\pi x) \left[A_2 \left(e^{\eta_2 y} - e^{\eta_2 (2-y)} \right) + e^{\eta_2 (1-y)} \right], & y > \bar{y}. \end{cases}$$

Details of this model problem can be found in [58, 89]. We choose $\mu_1 = \mu_2 = 1$ and $\bar{y} = 0.5001$ so that the solution is continuous at $y = 0.5001$. The constants A_1 and A_2 are given by $A_1 = 0.043295$, $A_2 = -0.043295$.

3.3.2 Single-domain

Here we consider a single domain. The linear system was solved using the PCG-algorithm described in the previous section. The exact and approximated solutions are plotted in Figs. 3.2 - 3.3 for $N = 10$. The L^2-error converges exponentially as can be

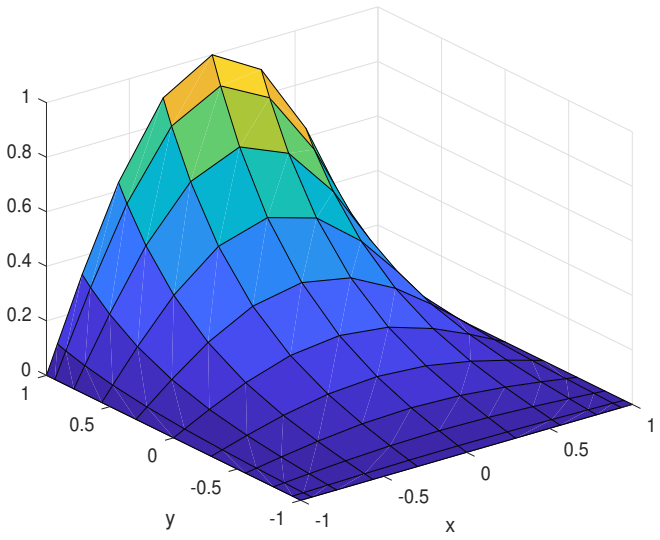


Figure 3.2: Analytical solution on mesh with $N = 10$ and a single element.

seen in Fig. 3.4. The ratio of the largest to smallest eigenvalue was plotted with respect to N in Fig. 3.5. A linear dependence can be seen. Similarly, the condition number of $P^{-1}A$ was plotted with respect to the mesh size N in Fig. 3.5 where a linear relationship can be seen. Finally, in Fig. 3.6, the number of iterations was plotted with respect to N. The number of iterations increases with the mesh size N.

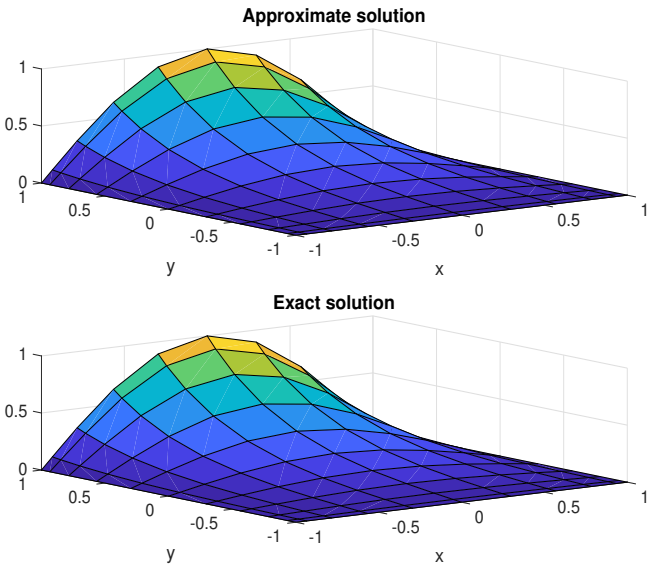


Figure 3.3: Approximate and exact solution on mesh with $N = 10$ and $K = 1$ for a stopping criteria $\varepsilon = 10^{-16}$.

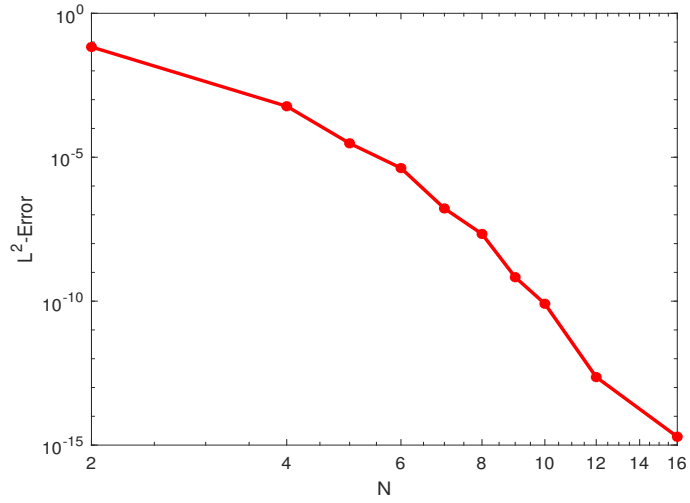


Figure 3.4: Convergence of the L^2-norm of the error with respect to N for $K = 1$ for a stopping criteria $\varepsilon = 10^{-16}$.

The L^2- norm of the error is tabulated as a function of N in Table 3.1. As it can be seen, a rapid convergence is obtained with machine precision error reached for $N = 12$.

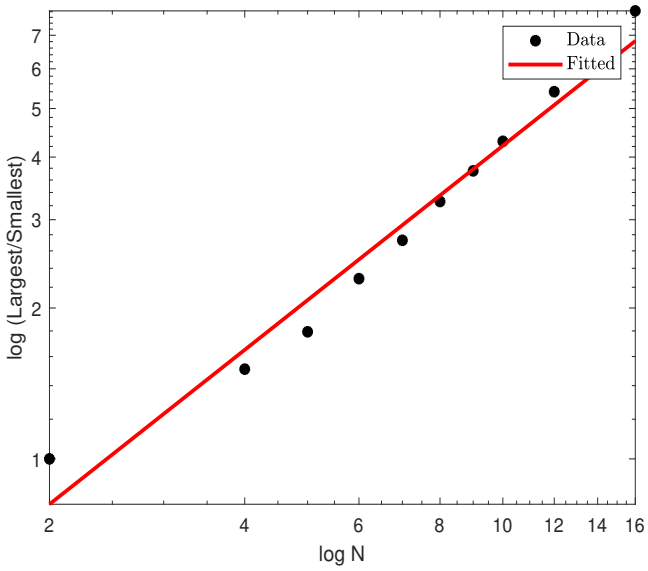


Figure 3.5: The ratio of the largest to smallest eigenvalue with respect to N for $K = 1$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.023.

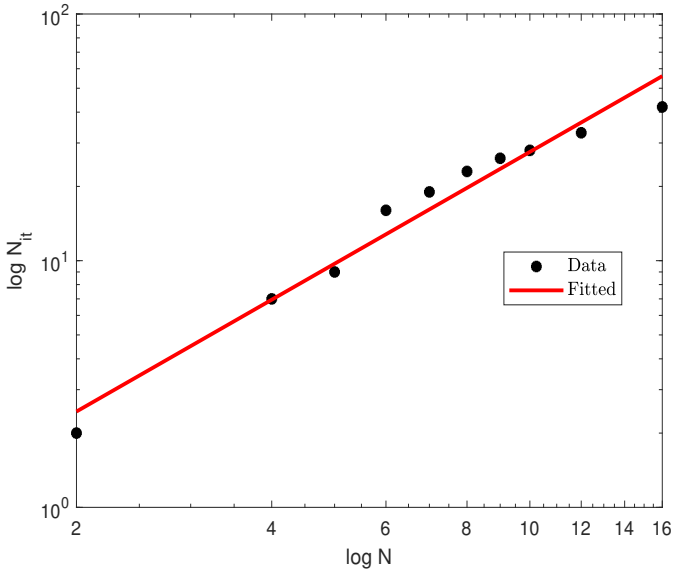


Figure 3.6: Number of iterations for convergence with respect to N for $K = 1$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.5.

N	Number of iterations	L^2 error
2	2	0.029
4	7	5.90×10^{-4}
5	9	3.05×10^{-5}
6	13	4.2×10^{-6}
7	14	1.66×10^{-7}
8	17	2.16×10^{-8}
9	26	6.8×10^{-10}
10	28	8.1×10^{-11}
12	33	2.30×10^{-13}
16	42	1.95×10^{-15}

Table 3.1: L^2- norm of the error as a function of N for $K = 1$ and a stopping criteria $\varepsilon = 10^{-16}$.

3.3.3 Multi-domain

In this sub-section we decompose the domain into more than one element (2, 3 and 4 elements) and initially we fix the order of polynomial approximation with $N = 10$. The solution of the system (3.8) was performed using the PCG-algorithm as in the previous section. The exact and approximated solutions are plotted in Fig. 3.7 for 2 elements, in Fig. 3.8 for 3 elements and in Fig. 3.9 for 4 elements. Again, the L^2-norm of the error converges exponentially as it can be seen in Fig. 3.10 for 2 elements, in Fig. 3.11 for 3 elements and in Fig. 3.12 for 4 elements. In Fig. 3.13 for 2 elements, in Fig. 3.14 for 3 elements and in Fig. 3.15 for 4 elements the number of iterations was plotted with respect to N. The number of iterations increases with the mesh size N. Similar, the condition number of $P^{-1}A$ was plotted with respect to the mesh size N in Fig. 3.16 for 2 elements, in Fig. 3.17 for 3 elements and in Fig. 3.18 for 4 elements where again a linear relationsheep can be seen. Finally, the ratio of the largest to smallest eigenvalue was plotted with respect to N in Fig. 3.19 for 2 elements, in Fig. 3.20 for 3 elements

and in Fig. 3.21 for 4 elements. A linear dependence can be seen in each case.

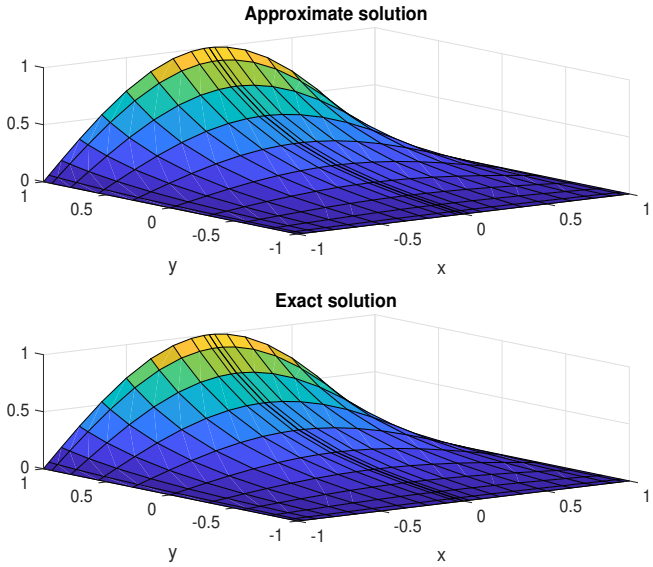


Figure 3.7: Approximate and exact solution for $N = 10$ with $K = 2$ and a stopping criteria $\varepsilon = 10^{-16}$.

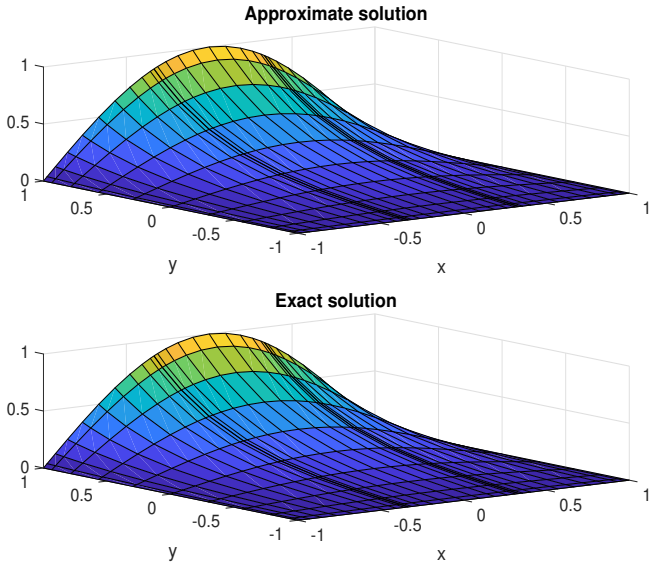


Figure 3.8: Approximate and exact solution for $N = 10$ with $K = 3$ and a stopping criteria $\varepsilon = 10^{-16}$.

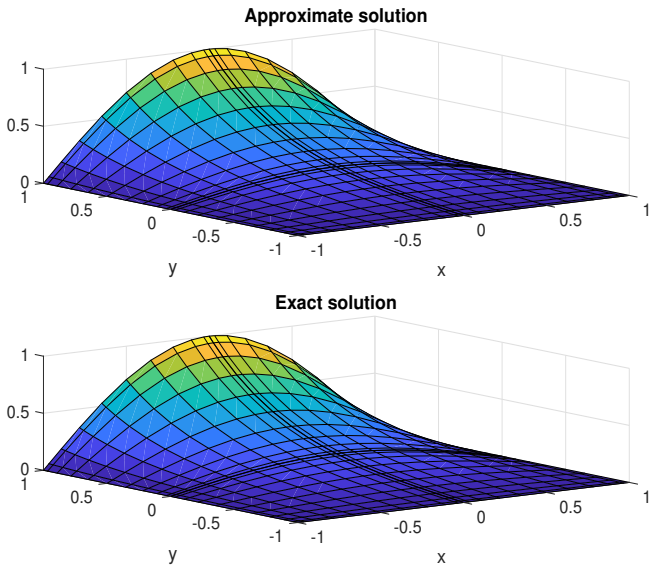


Figure 3.9: Approximate and exact solution for $N = 10$ with $K = 4$ and a stopping criteria $\varepsilon = 10^{-16}$.

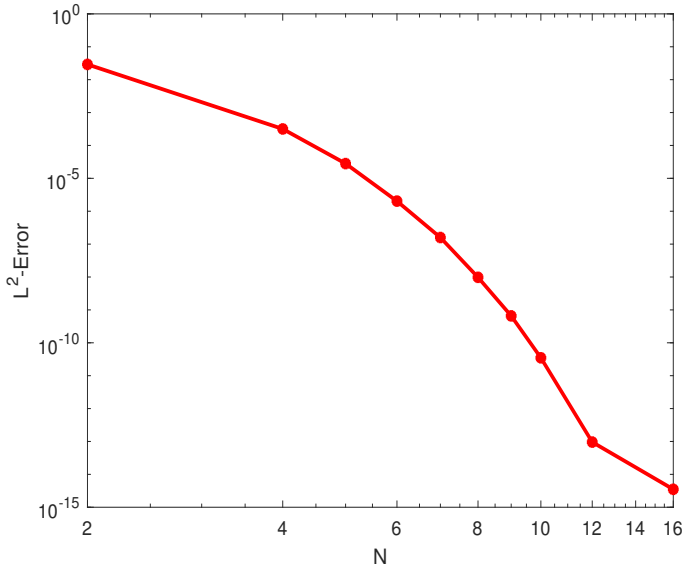


Figure 3.10: Convergence of the L^2-norm of the error with respect to N for $K = 2$ and a stopping criteria $\varepsilon = 10^{-16}$.

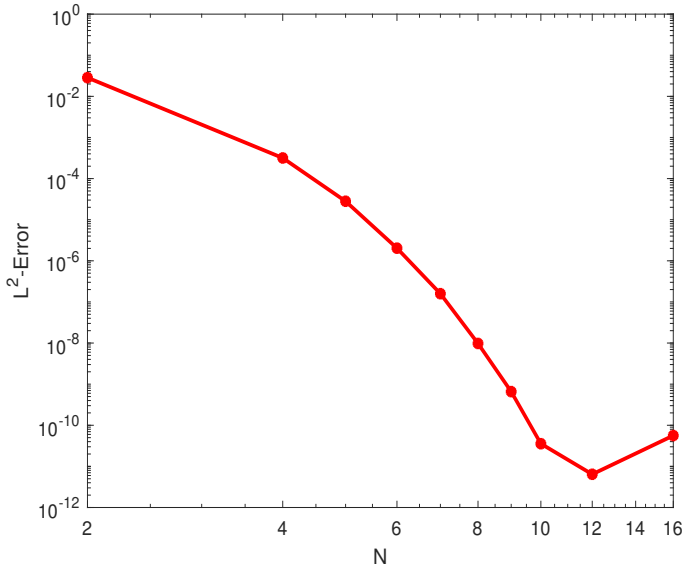


Figure 3.11: Convergence of the L^2-norm of the error with respect to N for $K = 3$ and a stopping criteria $\varepsilon = 10^{-16}$.

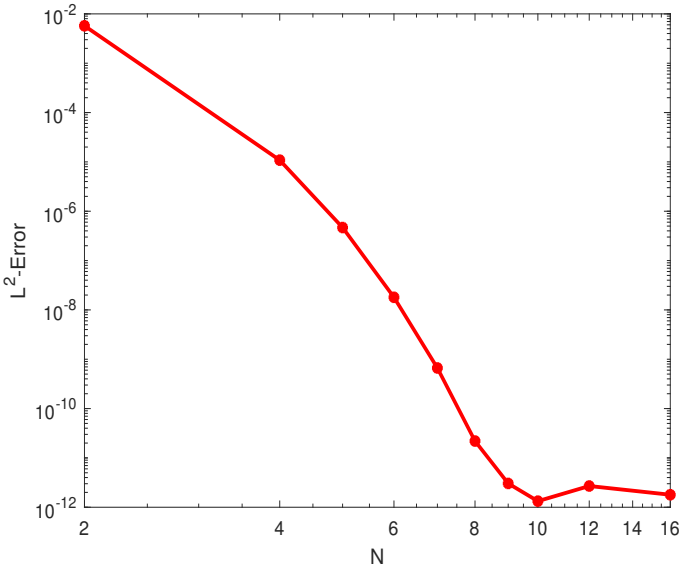


Figure 3.12: Convergence of the L^2-norm of the error with respect to N for $K = 4$ and a stopping criteria $\varepsilon = 10^{-16}$.

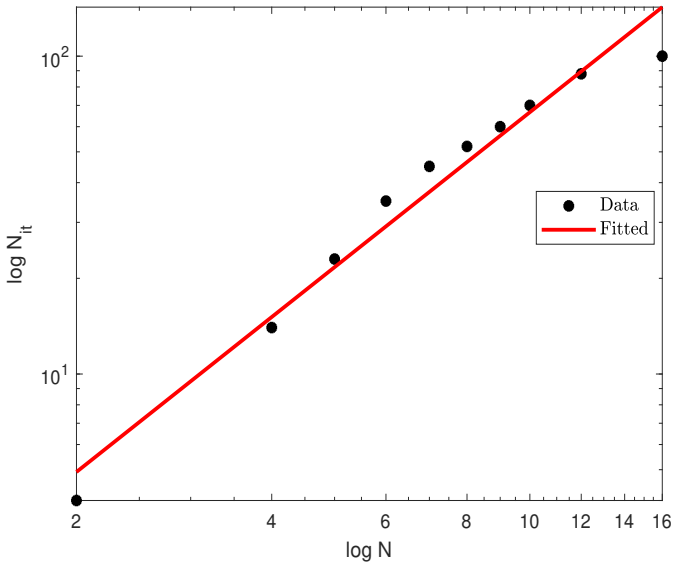


Figure 3.13: Dependence of the number of iterations for convergence with respect to N for $K = 2$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.6.

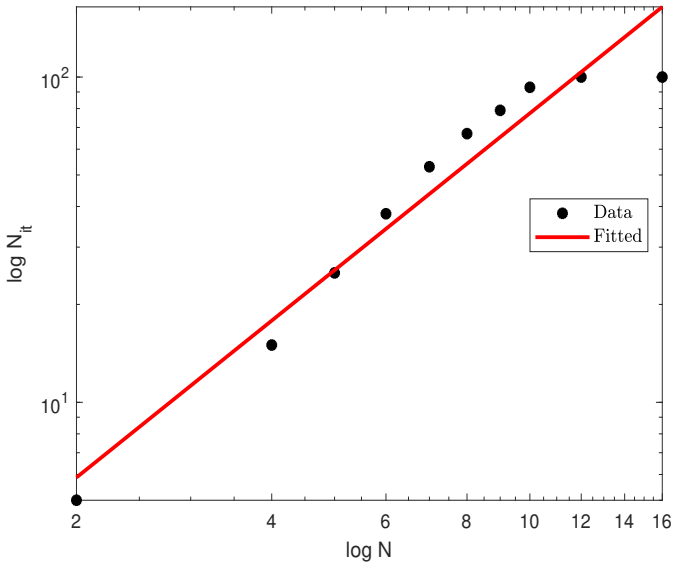


Figure 3.14: Dependence of the number of iterations for convergence with respect to N for $K = 3$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.6.

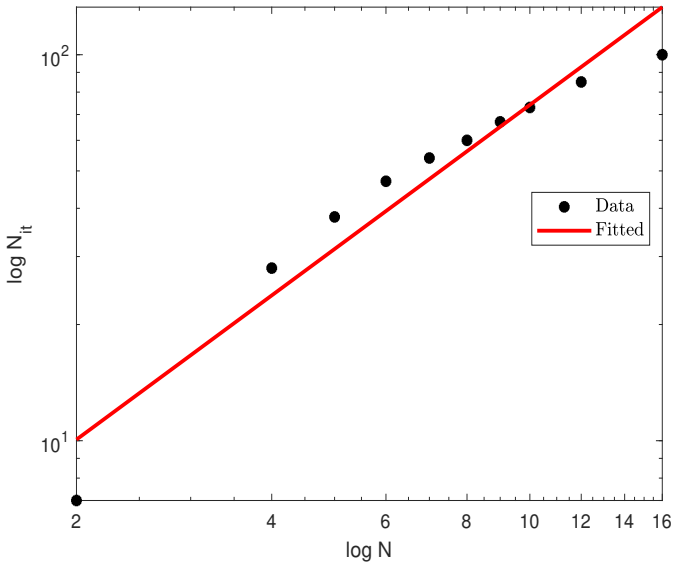


Figure 3.15: Dependence of the number of iterations for convergence with respect to N for $K = 4$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.2.

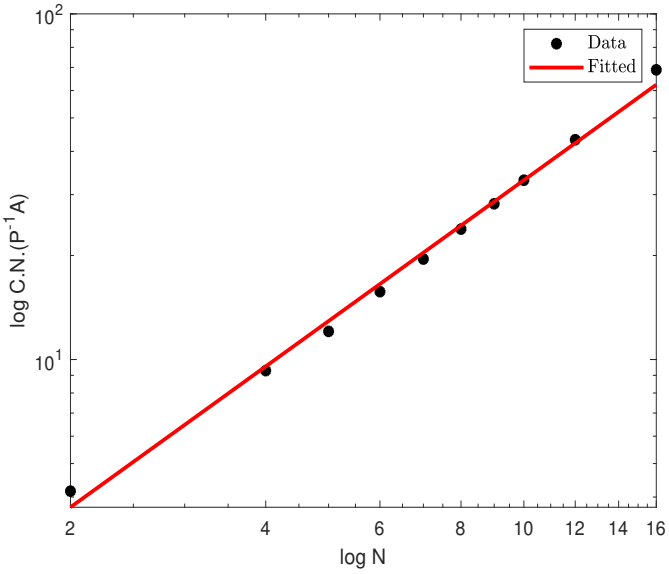


Figure 3.16: Dependence of the condition number of $P^{-1}A$ with respect to N for $K = 2$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.4.

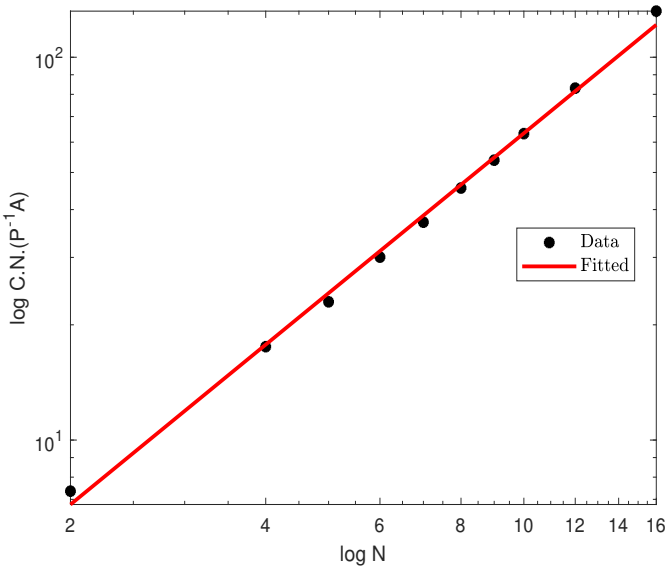


Figure 3.17: Dependence of the condition number of $P^{-1}A$ with respect to N for $K = 3$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.4.

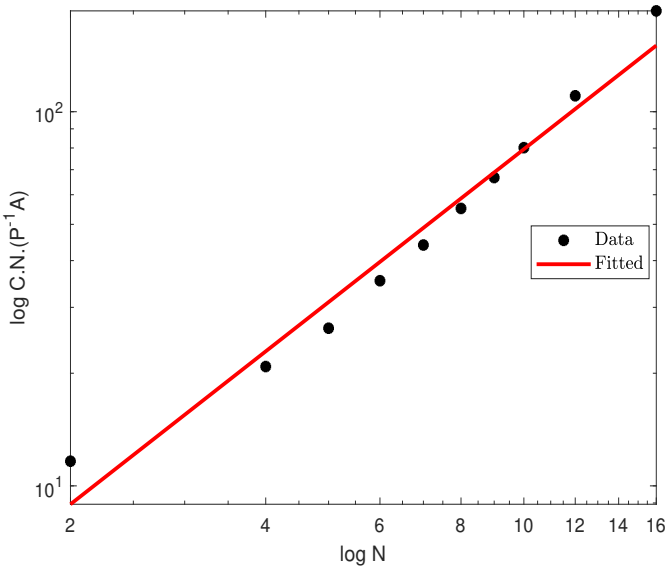


Figure 3.18: Dependence of the condition number of $P^{-1}A$ with respect to N for $K = 4$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.4.

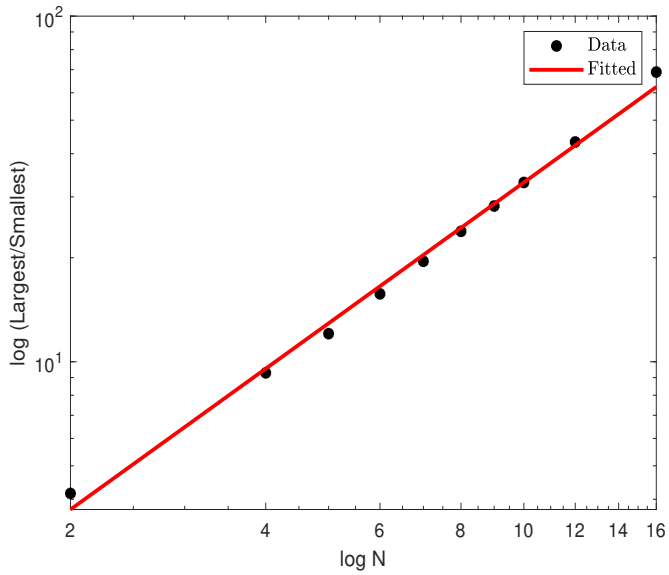


Figure 3.19: Dependence of the ratio of the largest to smallest eigenvalue with respect to N for $K = 2$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.4.

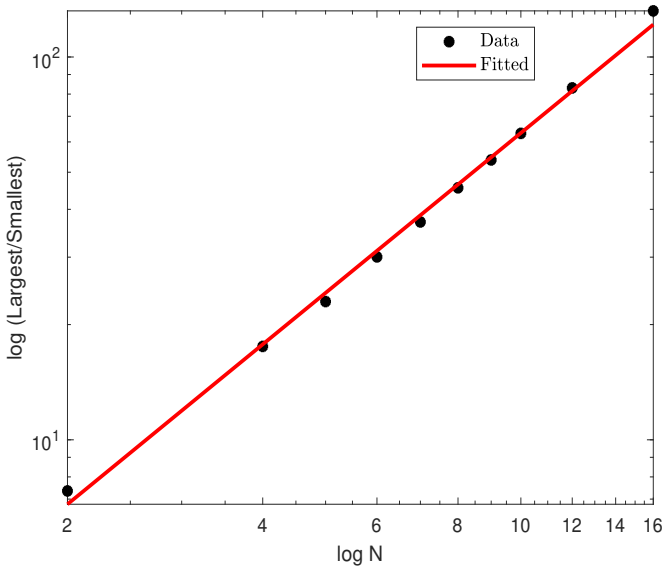


Figure 3.20: Dependence of the ratio of the largest to smallest eigenvalue with respect to N for $K = 3$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.4.

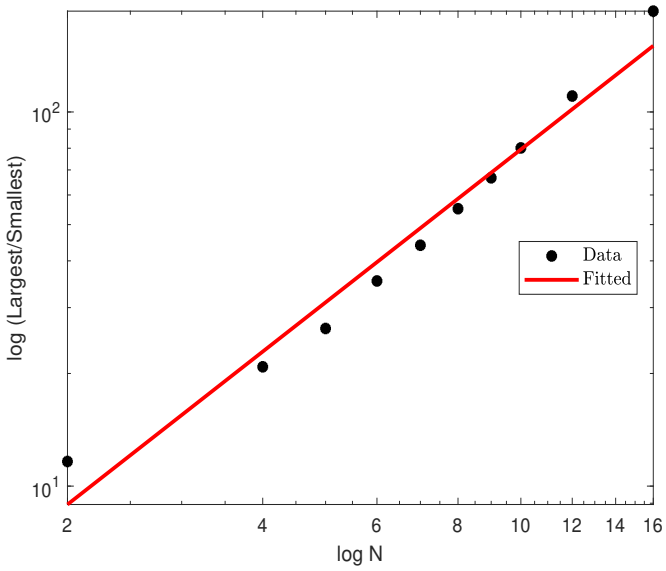


Figure 3.21: Dependence of the ratio of the largest to smallest eigenvalue with respect to N for $K = 4$ and a stopping criteria $\varepsilon = 10^{-16}$. The slope is about 1.4.

N	Number of iterations	L^2 error
2	3	0.029
4	13	3.17×10^{-4}
5	21	2.81×10^{-5}
6	28	2.03×10^{-6}
7	35	1.59×10^{-7}
8	42	9.77×10^{-9}
9	46	6.58×10^{-10}
10	52	3.47×10^{-11}
12	62	2.86×10^{-12}
16	81	2.46×10^{-12}

Table 3.2: L^2- norm of the error as a function of N for $K = 2$ and a stopping criteria $\varepsilon = 10^{-16}$.

N	Number of iterations	L^2 error
2	3	0.028
4	13	3.17×10^{-4}
5	22	2.81×10^{-5}
6	35	2.03×10^{-6}
7	44	1.5910^{-7}
8	54	9.77×10^{-9}
9	63	6.58×10^{-10}
10	73	3.56×10^{-11}
12	88	6.42×10^{-12}
16	100	5.62×10^{-11}

Table 3.3: L^2- norm of the error as a function of N for $K = 3$ and a stopping criteria $\varepsilon = 10^{-16}$.

N	Number of iterations	L^2 error
2	6	5.74×10^{-3}
4	24	1.08×10^{-5}
5	29	4.67×10^{-7}
6	34	1.8×10^{-8}
7	39	6.65×10^{-10}
8	43	2.2×10^{-11}
9	47	3.04×10^{-12}
10	53	1.33×10^{-12}
12	61	2.69×10^{-12}
16	79	1.79×10^{-12}

Table 3.4: L^2- norm of the error as a function of N for $K = 4$ and a stopping criteria $\varepsilon = 10^{-16}$.

3.4 Conclusions

In this chapter, we provide a general introduction to discretization methods and theoretical analysis of spectral methods. Our scope is mostly confined to model problems that illustrate the fundamentals and establish the relationship between classical spectral methods and their domain decomposition progeny. We discussed the discretization of Poisson's equation using spectral domain decomposition techniques. In particular, a detailed description of the spectral element discretization was provided. The obtained linear system was solved using the PCG-algorithm. As we can see from the previous figures and for single and multidomain discretization of differential equation the L^2– norm of the error of spectral element Method (SEM) depends on both number of elements (the domain was decomposed into one or more than one element (2, 3 and 4 elements)) and the degree of the approximation N (varying from 2 to 32). So, increasing one of these two parameters leads to convergence of SEM. The spectral element approximation achieved exponential convergence to the analytic solution of the problem with relatively few degrees of freedom. The L^2-norm of the error converges exponentially as can be seen in Fig. 3.4 for one element, in Fig. 3.10 for 2 elements, in Fig. 3.11 for 3 elements and in Fig. 3.12 for 4 elements. The ratio of the largest to smallest eigenvalue was plotted with respect to N in Figs. 3.5, 3.19, 3.20 and 3.21. A linear dependence can be seen. Similarly, the condition number of $P^{-1}A$ was plotted with respect to the mesh size N in Fig. 3.5 for one element, in Fig. 3.16 for 2 elements, in Fig. 3.17 for 3 elements and in Fig. 3.18 for 4 elements where a linear relationship can be seen. Finally, the number of iterations was plotted with respect to N. The number of iterations increases with the mesh size N. the number of iterations was plotted with respect to N in Fig. 3.6 for one element, Fig. 3.13 for 2 elements, in Fig. 3.14 for 3 elements and in Fig. 3.15 for 4 elements . The number of iterations increases linearly with the mesh size N. The L^2- norm of the error is tabulated as a function of N and number of iterations in Tables 3.1, 3.2, 3.3 and 3.4. As can be seen, a rapid convergence to machine precision error is obtained for $N = 10$.

Chapter 4

Approximation of Poisson equation using XSEM

Let $\Omega = [x_a, x_b] \times [y_a, y_b] = \Omega_1 \cup \Omega_2, f \in L^2(\Omega)$ and $u_d \in L^2(\partial\Omega)$ where $\Omega_1 \cap \Omega_2 = \Gamma$ is the interface between Ω_1 and Ω_2. We are seeking solution u of the following system :

$$\begin{cases} -\nabla \cdot (\mu \nabla u) = f \quad \text{on} \quad \Omega \\ \\ u = g \quad \text{on} \quad \partial\Omega \end{cases} \tag{4.1}$$

where

$$\mu(x,y) = \begin{cases} \mu_1 \quad \text{on } \Omega_1 \\ \\ \mu_2 \quad \text{on } \Omega_2 \end{cases}$$

μ_1 and μ_2 are two constants such that $\mu_1 \neq \mu_2$.

4.1 Weak formulation

In order to outline the extended spectral element method, we first start with the variational formulation of the problem (4.1) for the case when $\Omega = [x_a, x_b] \times [y_a, y_b]$ and $g = 0$. The solution u is sought in $H_0^1(\Omega)$ and then one can easily obtain the weak form

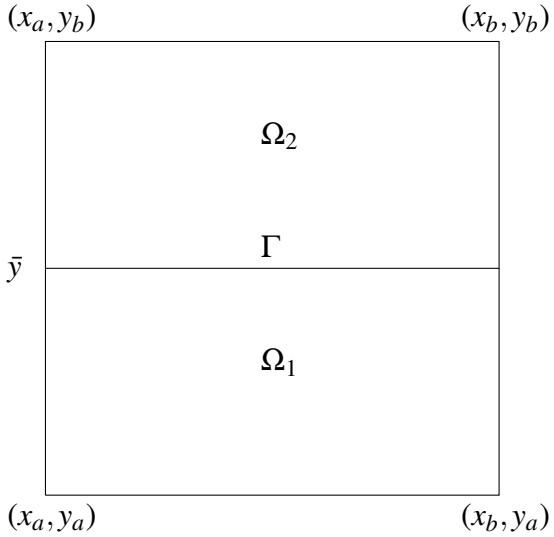


Figure 4.1: Two-phase domain.

:

$$\begin{cases} \text{Find} \quad u \in H_0^1(\Omega) \quad \text{such that} \\ \\ a(u,v) = (f,v), \qquad \forall v \in H_0^1(\Omega) \end{cases} \tag{4.2}$$

where the bilinear form is

$$a(u,v) \;\; = \int_\Omega \mu \nabla u \cdot \nabla v \, d\Omega$$

and the linear form is

$$(f,v) \;\; = \;\; \int_\Omega f v \, d\Omega$$

4.2 Discrete problem

If we denote by N the degree of interpolation and x_i and y_i, $i = 1,...,N+1$ are associated nodes, known as the Gauss-Lobatto Legendre points, which are the zeros of $(1-x^2)L'_N(x)$ and $(1-y^2)L'_N(y)$, respectively.

The weights for the Legendre-Gauss-Lobatto numerical integration rule are given by:

$$w_i = \frac{2}{N(N+1)} \frac{1}{L_N^2(x_i)}, \; i = 1,...,N+1.$$

Denote by $u_{ij} = u(x_i, y_j), f_{ij} = f(x_i, y_j)$ and $\mu_{ij} = \mu(x_i, y_j), i, j = 1, ..., N+1$.

The idea is based on decomposing the domain into three elements such that the middle element contains the discontinuity.

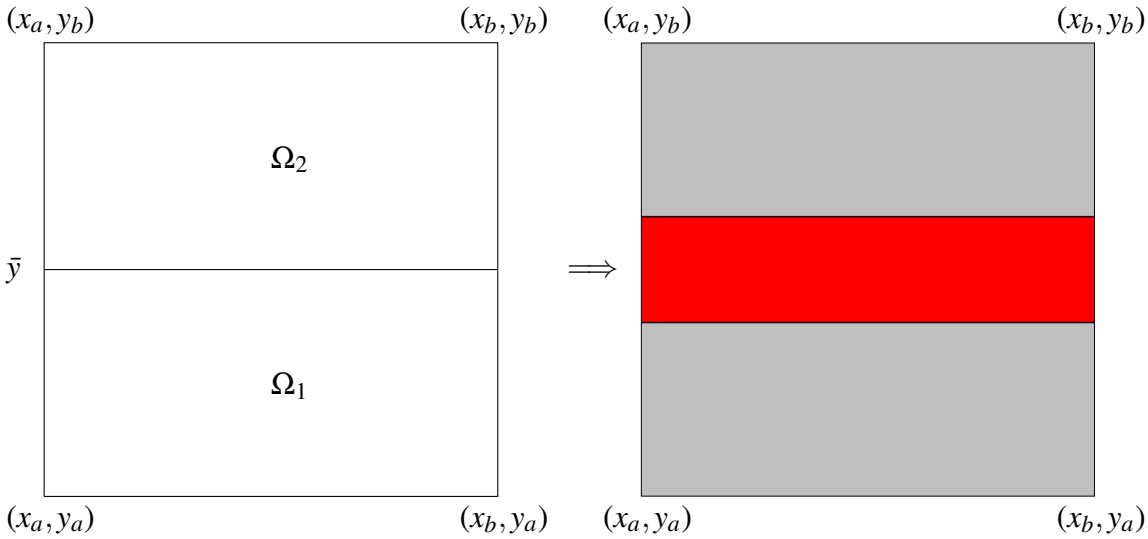


Figure 4.2: The idea is based on decomposing the two-phase domain into three elements such that the middle element contains the discontinuity.

On the first and third elements we used an SEM approach. On the middle element, u is expanded in terms of the Lagrange interpolants based on the Gauss-Lobatto Legendre points together with a supplementary enrichment term

$$u_N(x,y) = \underbrace{\sum_{i,j=1}^{N+1} u_{ij} h_i(x) h_j(y)}_{\text{strd. SE approx.}} + \underbrace{\sum_{i=2}^{N} \sum_{j=s}^{s+1} \alpha_{ij} \tilde{h}_i(x) \tilde{h}_j(y) \phi(x,y)}_{\text{enrichment}}$$

where $y_2 \leq y_s \leq \bar{y} \leq y_{s+1} \leq y_N$. Here the polynomials $h_i(x), i = 1, \ldots, N+1$ are the Lagrange interpolants and $\tilde{h}_i$ is a polynomial of degree $(N-2)$ that interpolates the interior $(N-1)$ GLL points, α_{ij} are the additional degrees of freedom and $\phi(x,y)$ is an enrichment function.

It can be a difficult task to analyse the enriched methods of the kind which we have discussed. As far as we are aware, it is only in Reusken [85] that we are able to find

any estimates of error for the extended finite element method (XFEM) . One reason for analysis of this method being difficult is the reliance on the estimate of the enrichment function ϕ . This function can be subject to considerable variation according to which enrichment type is needed. The enrichment function is totally different, even when the discontinuities at the interface are strong or weak. Consequently, it is no easy task to analyse this method in a unified way, and separate from the enrichment function which has been addressed. Reusken [85] inaugurated a structure which supplies an integrated treatment of the XFEM approximation for functions having strong discontinuities which was achieved by eliminating any reference to the enrichment function.

Let the space of functions V be a broken Sobolev space of order $m \geq 1$. When $m = 0$, we define:

$$H^0(\Omega_1 \cup \Omega_2) = L^2(\Omega_1 \cup \Omega_2) = \left\{ u \in L^2(\Omega) | u \in L^2(\Omega_i), i = 1,2 \right\} = L^2(\Omega)$$

In Reusken's structure [85], the construction of the approximation space shows a minor difference from that previously considered.

Therefore, let N_V denote the dimension of V_N and let $\Psi_i, i \in \mathscr{I}$, where $\mathscr{I} = 1, ..., N_V$, indicate the global basis functions which span V_N, where $V_N = V \cap [\phi;\ \phi_{|\Omega_e} \in P_N(\Omega_e)]^d$. Furthermore, let $X = x_k, k \in I$ be the set of all nodal points. We define the enriched approximation space as the restriction of the original approximation space to each side of the interface Γ. We define the restriction operator, $R_i : L^2(\Omega) \to L^2(\Omega), i = 1,2$, as:

$$R_i u = \begin{cases} u_{|\Omega_i} & \text{in } \Omega_i \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

Furthermore, it was demonstrated that the approximation may be written as:

$$V_N^\Gamma = R_1 V_N \oplus R_2 V_N.$$

Reusken [85] showed that

$$V^\Gamma := V \oplus V_1^\Gamma \oplus V_2^\Gamma.$$

Additionally, it was shown that one may write the approximation $u_N^\Gamma \in V_N^\Gamma$ in the form:

$$u_N^\Gamma = u_N + \sum_{k \in \mathscr{I}_1^\Gamma} \beta_k^{(1)} R_1 \phi_k + \sum_{k \in \mathscr{I}_2^\Gamma} \beta_k^{(2)} R_2 \phi_k \tag{4.4}$$

where $u_N \in V_N$ is the standard continuous approximation and $\beta_k^{(i)}, i = 1,2$, are additional degrees of freedom. We present the XSEM equivalent of the approximation results derived by Reusken [85] for the case of a single spectral element.

Let $\Omega = [-1,1]^d$ with $d = 1$ or $d = 2$. Define $\varepsilon_i^m : H^m(\Omega_i) \to H^m(\Omega)$ to be an extension operator such that : $(\varepsilon_i^m v)_{|\Omega_i} = v$ and $\|\varepsilon_i^m v\|_{H^m(\Omega)} \leq c\|v\|_{H^m(\Omega_i)}$, $\forall\, v \in H^m(\Omega_i)$. Let $\pi_N^m : H^m(\Omega) \to V_N := H^m(\Omega) \cap P_N(\Omega)$ be a projection operator satisfying

$$\begin{aligned} \|w - \pi_N^m w\|_{L^2(\Omega)} &\leq cN^{-m}\|w\|_{H^m(\Omega)}, \quad m \geq 0,\ \forall\, w \in H^m(\Omega) \\ \|w - \pi_N^m w\|_{H^1(\Omega)} &\leq cN^{1-m}\|w\|_{H^m(\Omega)}, \quad m \geq 2,\ \forall\, w \in H^m(\Omega) \end{aligned} \tag{4.5}$$

then one has the following corollary :

Corollary 1.

$$\begin{aligned} \inf_{u_N^\Gamma \in V_N^\Gamma} \|u - u_N^\Gamma\|_{L^2(\Omega_1 \cup \Omega_2)} &\leq cN^{-m}\|u\|_{H^m(\Omega_1 \cup \Omega_2)}, \quad m \geq 0,\ \forall\, w \in H^m(\Omega_1 \cup \Omega_2) \\ \inf_{u_N^\Gamma \in V_N^\Gamma} \|u - u_N^\Gamma\|_{H^1(\Omega)} &\leq cN^{1-m}\|w\|_{H^m(\Omega_1 \cup \Omega_2)}, \quad m \geq 2,\ \forall\, w \in H^m(\Omega_1 \cup \Omega_2) \end{aligned} \tag{4.6}$$

The steps needed to prove the result of the aforementioned approximation are congruous to the stages which Reusken utilised [85].

.

Define extension operators

$$\varepsilon_i^m : H^m(\Omega_i) \to H^m(\Omega), i = 1,2$$

with $(\varepsilon_i^m w)_{|\Omega_i} = w$ and

$$\|\varepsilon_i^m w\|_m \leq c\|w\|_{m,\Omega_i}.$$

Let $m \in 1,2$ and $u \in H^m(\Omega_1 \cup \Omega_2)$ be given. Define $v^* \in V^\Gamma$ by

$$v^* = R_1 \pi_N^m \varepsilon_1^m R_1 u + R_2 \pi_N^m \varepsilon_2^m R_2 u. \tag{4.7}$$

For this approximation we obtain

$$\begin{aligned}
\|u-v^*\|^2_{L^2(\Omega_1\cup\Omega_2)} &= \sum_{i=1}^{2} \|u-v^*\|^2_{L^2(\Omega_i)} \\
&= \sum_{i=1}^{2} \|u-\pi_N^m \varepsilon_i^m R_i u\|^2_{L^2(\Omega_i)} \\
&= \sum_{i=1}^{2} \|\varepsilon_i^m R_i u-\pi_N^m \varepsilon_i^m R_i u\|^2_{L^2(\Omega_i)} \\
&\leq \sum_{i=1}^{2} \|\varepsilon_i^m R_i u-\pi_N^m \varepsilon_i^m R_i u\|^2_{L^2(\Omega_1\cup\Omega_2)} \\
&\leq cN^{-m} \sum_{i=1}^{2} \|\varepsilon_i^m R_i u\|^2_m \\
&\leq cN^{-m} \sum_{i=1}^{2} \|R_i u\|^2_{m,\Omega_i} = cN^{-m}\|u\|^2_{m,\Omega_1\cup\Omega_2}
\end{aligned} \tag{4.8}$$

This verifies the first approximation result, the second being similarly verified by applying the H^1-semi norm. ■

To illustrate this phenomenon we consider the spectral element interpolation of a discontinuous function on a grid of uniformly spaced points. For simplicity, we will assume that our domain $\Omega = (-1,1)^2 \subset \mathbb{R}^2$ and that we only have a *single* spectral element.

Define the 1D interface

$$\Gamma = \{(x,y) \in \Omega; \quad y = 0.05\}, \quad \Omega_1 = \{(x,y) \in \Omega; \quad y < 0.05\}, \quad \Omega_2 = \Omega \setminus \Omega_1.$$

Note that the line $y = 0.05$ is chosen as the intersection between the boundaries of the two subdomains because 0 is a member of the Gauss-Lobatto Legendre grid.

Let u be given by

$$u(x,y) = \begin{cases} x^2+y^2 & \text{in} \quad \Omega_1, \\ \\ 3x^2+y^2+2 & \text{in} \quad \Omega_2. \end{cases} \tag{4.9}$$

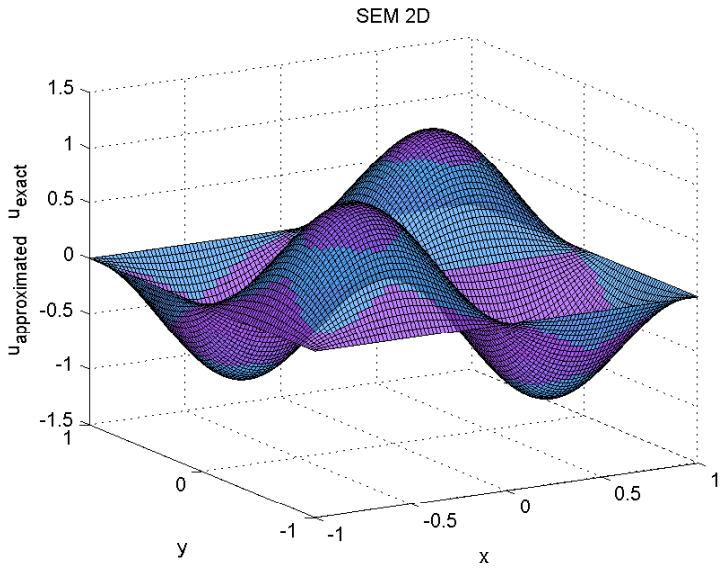


Figure 4.3: SEM for continuous function

We construct the spectral element interpolant of this function on a uniformly spaced grid using a single element. The uniformly spaced grid is denoted:

$$D_u = \bigcup_{k=1}^{M_x} [x_{k-1}, x_k] \times \bigcup_{m=1}^{M_y} [y_{m-1}, y_m]$$

where $x_0 = y_0 = -1$ and $x_{M_x} = y_{M_y} = 1$ and $(M_x+1)\times(M_y+1)$ is the total number of uniformly spaced points. Therefore the spectral element and extended spectral element approximations, denoted SEM and $XSEM$ respectively, on the domain Ω are given by:

$$\begin{aligned} SEM(x,y) &= \sum_{i,j=0}^{N} u_{ij} h_i(x) h_j(y) \\ XSEM(x,y) &= \sum_{i,j=0}^{N} u_{ij} h_i(x) h_j(y) + \sum_{i,j=1}^{N-1} \alpha_{ij} \tilde{h}_i(x) \tilde{h}_j(y) \phi_{ij}(x,y) \end{aligned} \tag{4.10}$$

where $u_{ij} = u(x_i, y_j), i, j = 0, ..., N, h_i$ are the Lagrange interpolants, α_{ij} are the additional degrees of freedom.

The local enriched approximation is given by :

$$\begin{aligned} XSEM(x,y) &= SEM(x,y) + Enr(x,y) \\ &= \underbrace{\sum_{i,j=1}^{N+1} u_{ij} h_i(x) h_j(y)}_{\text{strd. SE approx.}} + \underbrace{\sum_{i,j=1}^{N-1} \alpha_{ij} \tilde{h}_i(x) \tilde{h}_j(y) \phi_{ij}(x,y)}_{\text{enrichment}} \end{aligned} \tag{4.11}$$

It is clear that this function is dependent on the kind of discontinuity or singularity which is being enriched. It is possible to utilise the same global enrichment function as that which is defined in [41], i.e. $\phi_{ij}(x,y) = H(x,y) - H(x_i,y_j)$, where $H(x,y)$ is the Heaviside function defined by:

$$H(x,y) = \begin{cases} 0, & (x,y) \in \Omega_1 \\ 1, & (x,y) \in \Omega_2 \end{cases} \tag{4.12}$$

The coefficients $\alpha_{ij} = \alpha(i,j)$ are totally unknown. Therefore, we need to presuppose that $XSEM(x_k,y_m) \equiv u(x_k,y_m), \forall (x_k,y_m) \in D_u$ for the purpose of calculating them. We subsequently found the coefficients $\alpha(i,j)$ from the residual of the standard SEM approximation:

$$\begin{aligned} Enr(x_k,y_m) &= \sum_{i,j=0}^{N} \alpha_{ij} h_i(x_k) h_j(y_m) \phi_{ij}(x_k,y_m) \\ &= u(x_k,y_m) - SEM(x_k,y_m) \\ &= u(x_k,y_m) - \sum_{i,j=0}^{N} u_{ij} h_i(x_k) h_j(y_m) \end{aligned} \tag{4.13}$$

In order to calculate them, one can write the matrix $(\alpha(i,j))$ in a vector form as follows

$$(V_\alpha)_I = \alpha(i,j) \text{ for } I = (i-1)(N+1) + j, i, j = 0, \ldots, N. \tag{4.14}$$

Then one obtains a linear system of the form

$$AV_\alpha = F$$

where F is the right-hand side. Note that the number of uniformly spaced points will, in general, be larger than the degree of the polynomial ($M > N$), thus the matrix A

of size $(M+1)^2 \times (N+1)^2$ is not square. Therefore it is inverted by multiplying by its transpose A^T of size $(N+1)^2 \times (M+1)^2$ to produce a square matrix $A^T A$ of size $(N+1)^2 \times (N+1)^2$ which is then inverted.

$$A^T A V_\alpha = A^T F \tag{4.15}$$

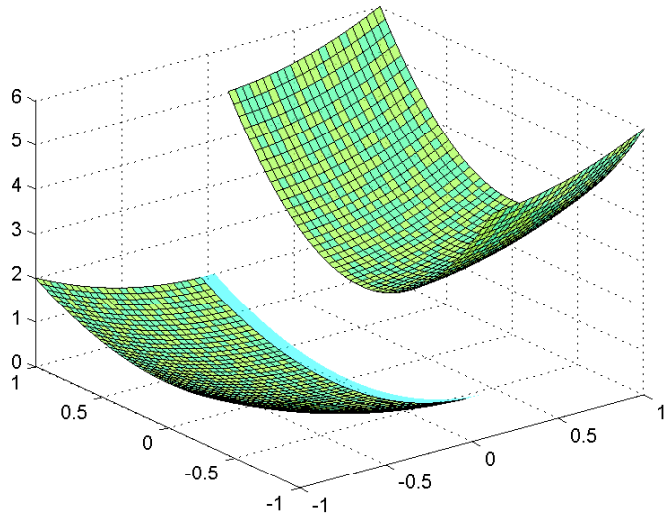


Figure 4.4: XSEM for discontinuous functions

The Galerkin approximation is to solve the discrete weak problem (4.2):
Find $u_N \in \mathscr{V}^N$ such that

$$\Big(\mu \nabla u_N, \nabla v_N\Big)_N = \Big(f, v_N\Big)_N, \ \forall v_N \in \mathscr{V}^N \tag{4.16}$$

where the discrete inner product $\Big(.,.\Big)_N$ is defined by

$$\Big(\varphi, \psi\Big)_N = \sum_{m,n=1}^{N+1} w_m w_n \varphi(x_m, y_n) \psi(x_m, y_n)$$

Note that

$$\nabla u_N(x,y) = \begin{pmatrix} \displaystyle\sum_{i,j=1}^{N+1} u_{ij} h_i'(x) h_j(y) + \sum_{i=2}^{N} \sum_{j=s}^{s+1} \alpha_{ij} \Big(\tilde{h}_i'(x) \phi(x,y) + \tilde{h}_i(x) \frac{\partial \phi}{\partial x}(x,y) \Big) \tilde{h}_j(y) \\ \\ \displaystyle\sum_{i,j=1}^{N+1} u_{ij} h_i(x) h_j'(y) + \sum_{i=2}^{N} \sum_{j=s}^{s+1} \alpha_{ij} \Big(\tilde{h}_j'(y) \phi(x,y) + \tilde{h}_j(y) \frac{\partial \phi}{\partial y}(x,y) \Big) \tilde{h}_i(x) \end{pmatrix}$$

and by choosing the test functions as $v_N(x,y) = h_k(x)h_l(y)$ $(2 \leq k,l \leq N)$, one obtains

$$\nabla v_N(x,y) = \begin{pmatrix} h_k'(x)h_l(y) \\ \\ h_k(x)h_l'(y) \end{pmatrix}.$$

By injecting the solution's decomposition u_N and the chosen test function into the variational formulation (4.16), one obtains the following discrete system :

$$\begin{aligned}
&\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=1}^{N+1} u_{ij} \Big(h_i'(x_m)h_j(y_n)h_k'(x_m)h_l(y_n) + h_i(x_m)h_j'(y_n)h_k(x_m)h_l'(y_n) \Big) + \\
&\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i=2}^{N} \sum_{j=s}^{s+1} \alpha_{ij} \Big(\tilde{h}_i'(x_m)\phi(x_m,y_n) + \tilde{h}_i(x_m)\frac{\partial \phi}{\partial x}(x_m,y_n) \Big) \tilde{h}_j(y_n)h_k'(x_m)h_l(y_n) + \\
&\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i=2}^{N} \sum_{j=s}^{s+1} \alpha_{ij} \Big(\tilde{h}_j'(y_n)\phi(x_m,y_n) + \tilde{h}_j(y_n)\frac{\partial \phi}{\partial y}(x_m,y_n) \Big) \tilde{h}_i(x_m)h_k(x_m)h_l'(y_n) \\
&\qquad = \sum_{m,n=1}^{N+1} w_m w_n f_{mn} h_k(x_m)h_l(y_n)
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\sum_{i,j=1}^{N+1} u_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \Big(h_i'(x_m)h_j(y_n)h_k'(x_m)h_l(y_n) + h_i(x_m)h_j'(y_n)h_k(x_m)h_l'(y_n) \Big) + \\
&\sum_{i=2}^{N} \sum_{j=s}^{s+1} \alpha_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \Big(\tilde{h}_i'(x_m)\phi(x_m,y_n) + \tilde{h}_i(x_m)\frac{\partial \phi}{\partial x}(x_m,y_n) \Big) \tilde{h}_j(y_n)h_k'(x_m)h_l(y_n) + \\
&\sum_{i=2}^{N} \sum_{j=s}^{s+1} \alpha_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \Big(\tilde{h}_j'(y_n)\phi(x_m,y_n) + \tilde{h}_j(y_n)\frac{\partial \phi}{\partial y}(x_m,y_n) \Big) \tilde{h}_i(x_m)h_k(x_m)h_l'(y_n) \\
&\qquad = \sum_{m,n=1}^{N+1} w_m w_n f_{mn} h_k(x_m)h_l(y_n)
\end{aligned}$$

By choosing

$$\phi(x,y) = \sum_{p,q=1}^{N+1} \Big|y_q - \bar{y}\Big| h_p(x)h_q(y) - \Big|\sum_{p,q=1}^{N+1} (y_q - \bar{y})\, h_p(x)h_q(y)\Big|$$

where $-1 \leq \bar{y} \leq 1$, arbitrarily chosen. The values of ϕ on the nodes are given by

$$\begin{aligned}\phi(x_i,y_j) &= \sum_{p,q=1}^{N+1} \Big|y_q - \bar{y}\Big| h_p(x_i)h_q(y_j) - \Big|\sum_{p,q=1}^{N+1} (y_q - \bar{y})\, h_p(x_i)h_q(y_j)\Big| \\ &= |y_j - \bar{y}| - |y_j - \bar{y}|\end{aligned}$$

so that $\phi(x_i,y_j) = 0, 1 \leq i,j \leq N+1$

Remark 1. *We considered different enrichment functions given hereafter*

- $\phi_1(x,y) = ||(x,y) - (x,\bar{y})|| = |y - \bar{y}|$ *where* $\bar{y}$*, arbitrarily chosen.*
- $\phi_2(x,y) = \sum_{p,q=1}^{N+1} \Big|(x_p,y_q) - (x_p,\bar{y})\Big| h_q(y)$ *where* $\bar{y}$*, arbitrarily chosen.*
- $\phi_3(x,y) = ||(x,y) - (\bar{x},\bar{y})|| = \sum_{p,q=1}^{N+1} ||(x_p,y_q) - (\bar{x},\bar{y})||\, h_p(x)h_q(y)$

  where $(\bar{x},\bar{y})$ *is a point on the interface, arbitrarily chosen. It is taken to be somewhere about the midpoint along the interface which is easy to calculate in the case of a straight interface.*
- $\phi_4(x,y) = |y - \bar{y}| - \Big|\sum_{p,q=1}^{N+1} (y_q - \bar{y})\, h_p(x)h_q(y)\Big|$ *where* $-1 \leq \bar{y} \leq 1$*, arbitrarily chosen.*
- $\phi_5(x,y) = \sum_{p,q=1}^{N+1} \Big|y_q - \bar{y}\Big| h_p(x)h_q(y) - \Big|\sum_{p,q=1}^{N+1} (y_q - \bar{y})\, h_p(x)h_q(y)\Big|$ *where* $-1 \leq \bar{y} \leq 1$*, arbitrarily chosen.*

By applying these functions in the local enriched approximation (4.11), the results using each of the above enrichment functions do not improve the accurcy of the enriched XSEM approximation.

One can easily calculate

$$\frac{\partial \phi}{\partial x}(x,y) = \sum_{p,q=1}^{N+1} \Big|y_q-\bar{y}\Big| h_p'(x)h_q(y) - \sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p'(x)h_q(y)\ \text{sign}\Big(\sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p(x)h_q(y)\Big)$$

and their values on the nodes are given by

$$\begin{aligned}\frac{\partial \phi}{\partial x}(x_i,y_j) &= \sum_{p,q=1}^{N+1} \Big|y_q-\bar{y}\Big| h_p'(x_i)h_q(y_j) - \sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p'(x_i)h_q(y_j)\ \text{sign}\Big(\sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p(x_i)h_q(y_j)\Big)\\ &= |y_j-\bar{y}|\sum_{p=1}^{N+1} h_p'(x_i) - \ \text{sign}\,(y_j-\bar{y})(y_j-\bar{y})\sum_{p=1}^{N+1} h_p'(x_i)\\ &= 0\end{aligned}$$

and also the derivative with respect to y is given by

$$\frac{\partial \phi}{\partial y}(x,y) = \sum_{p,q=1}^{N+1} \Big|y_q-\bar{y}\Big| h_p(x)h_q'(y) - \sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p(x)h_q'(y)\ \text{sign}\Big(\sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p(x)h_q(y)\Big)$$

and then their values on the nodes are given by

$$\begin{aligned}\frac{\partial \phi}{\partial y}(x_i,y_j) &= \sum_{p,q=1}^{N+1} \Big|y_q-\bar{y}\Big| h_p(x_i)h_q'(y_j) - \sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p(x_i)h_q'(y_j)\text{sign}\Big(\sum_{p,q=1}^{N+1}(y_q-\bar{y})h_p(x_i)h_q(y_j)\Big)\\ &= \sum_{q=1}^{N+1} \Big|y_q-\bar{y}\Big| D_{jq} - \text{sign}(y_j-\bar{y})\sum_{q=1}^{N+1}(y_q-\bar{y})D_{jq}\end{aligned}$$

then the discrete system becomes

$$\begin{aligned}&\sum_{i,j=1}^{N+1} u_{ij}\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}\Big(h_i'(x_m)h_j(y_n)h_k'(x_m)h_l(y_n) + h_i(x_m)h_j'(y_n)h_k(x_m)h_l'(y_n)\Big) +\\ &\sum_{i,j=2}^{N} \alpha_{ij}\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}\tilde{h}_i(x_m)\tilde{h}_j(y_n)h_k(x_m)h_l'(y_n)\frac{\partial \phi}{\partial y}(x_m,y_n) = \sum_{m,n=1}^{N+1} w_m w_n f_{mn}h_k(x_m)h_l(y_n)\end{aligned}$$

Using the notation $h_i(x_m)=\delta_{mi}$, $\tilde{h}_i(x_m)=G_{mi}$ and $h_i'(x_m)=D_{mi}$, one deduces that

$$\begin{aligned}&\sum_{i,j=1}^{N+1} u_{ij}\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}\Big(D_{mi}\delta_{nj}D_{mk}\delta_{nl} + \delta_{mi}D_{nj}\delta_{mk}D_{nl}\Big) +\\ &\sum_{i=2}^{N}\sum_{j=2}^{N} \alpha_{ij}\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}G_{mi}G_{nj}\delta_{mk}D_{nl}\frac{\partial \phi}{\partial y}(x_m,y_n) = \sum_{m,n=2}^{N} w_m w_n f_{mn}\delta_{mk}\delta_{nl}\end{aligned}$$

Let $I=(l-1)(N+1)+k$ and $J=(j-1)(N+1)+i$. Then the system reduces to

$$\sum_{J=1}^{(N+1)^2} A_{IJ}u_J + \sum_{J=2}^{N^2} B_{IJ}\alpha_J = F_I \tag{4.17}$$

where

$$A_{IJ} = \delta_{lj} \sum_{m=1}^{N+1} D_{mi} D_{mk} w_m \mu_{ml} w_l + \delta_{ki} \sum_{n=1}^{N+1} D_{nj} D_{nl} w_k \mu_{kn} w_n,$$

$$B_{IJ} = \sum_{n=1}^{N+1} w_k w_n \mu_{kn} G_{ki} G_{nj} D_{nl} \frac{\partial \phi}{\partial y}(x_k, y_n)$$

and

$$F_I = w_k w_l f_{kl}.$$

Now by choosing another test function $v_N = \tilde{h}_k(x)\tilde{h}_l(y)\phi(x,y)$ $(2 \leq k,l \leq N)$, one obtains

$$\nabla v_N(x,y) = \begin{pmatrix} \left(\tilde{h}'_k(x)\phi(x,y) + \tilde{h}_k(x)\frac{\partial \phi}{\partial x}(x,y)\right)\tilde{h}_l(y) \\ \tilde{h}_k(x)\left(\tilde{h}'_l(y)\phi(x,y) + \tilde{h}_l(y)\frac{\partial \phi}{\partial y}(x,y)\right) \end{pmatrix}.$$

Then the discrete system becomes:

$$\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=1}^{N+1} u_{ij}\, h_i(x_m) h'_j(y_n) \tilde{h}_k(x_m) \tilde{h}_l(y_n) \frac{\partial \phi}{\partial y}(x_m, y_n) +$$
$$\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=2}^{N} \alpha_{ij}\, \tilde{h}_i(x_m) \tilde{h}_k(x_m) \tilde{h}_j(y_n) \frac{\partial \phi}{\partial y}(x_m, y_n) \tilde{h}_l(y_n) \frac{\partial \phi}{\partial y}(x_m, y_n) = 0$$

which is equivalent to

$$\sum_{i,j=1}^{N+1} u_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}\, h_i(x_m) h'_j(y_n) \tilde{h}_k(x_m) \tilde{h}_l(y_n) \frac{\partial \phi}{\partial y}(x_m, y_n) +$$
$$\sum_{i,j=2}^{N} \alpha_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \tilde{h}_i(x_m) \tilde{h}_k(x_m) \tilde{h}_j(y_n) \frac{\partial \phi}{\partial y}(x_m, y_n) \tilde{h}_l(y_n) \frac{\partial \phi}{\partial y}(x_m, y_n) = 0$$

$$\sum_{i,j=1}^{N+1} u_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}\, h_i(x_m) h'_j(y_n) \tilde{h}_k(x_m) \tilde{h}_l(y_n) \frac{\partial \phi}{\partial y}(x_m, y_n) +$$
$$\sum_{i,j=2}^{N} \alpha_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \tilde{h}_i(x_m) \tilde{h}_j(y_n) \tilde{h}_k(x_m) \tilde{h}_l(y_n) (\frac{\partial \phi}{\partial y})^2(x_m, y_n) = 0$$

using notations as $h_i(x_m) = \delta_{mi}$ and $\tilde{h}_i(x_m) = G_{mi}$, one deduces

$$\sum_{i,j=1}^{N+1} u_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}\, \delta_{mi} D_{nj} G_{mk} G_{nl} \frac{\partial \phi}{\partial y}(x_m, y_n) +$$
$$\sum_{i,j=2}^{N} \alpha_{ij} \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} G_{mi} G_{nj} G_{mk} G_{nl} (\frac{\partial \phi}{\partial y})^2(x_m, y_n) = 0$$

Let $I=(l-2)(N-1)+k-1$ and $J=(j-1)(N-1)+i$. Then the system reduces to

$$\sum_{J=1}^{(N+1)^2} C_{IJ}u_J + \sum_{J=1}^{(N-1)^2} E_{IJ}\alpha_J = 0 \tag{4.18}$$

where

$$C_{IJ} = \sum_{n=1}^{N+1} w_i w_n \mu_{in}\, D_{nj} G_{ik} G_{nl} \frac{\partial \phi}{\partial y}(x_i, y_n)$$

$$E_{IJ} = \sum_{m,n=1}^{N+1} w_m w_n \mu_{mn}\, G_{mi} G_{nj} G_{mk} G_{nl} \left(\frac{\partial \phi}{\partial y}(x_m, y_n)\right)^2$$

Now combining the two systems (4.17) and (4.18), one obtains the system of equations

$$\begin{cases} Au + B\alpha = F \\ \\ Cu + E\alpha = 0 \end{cases} \tag{4.19}$$

Let $G = \begin{pmatrix} A & B \\ C & E \end{pmatrix}$ be a square matrix, $U = \begin{pmatrix} u \\ \alpha \end{pmatrix}$ and $FF = \begin{pmatrix} F \\ \mathbf{0} \end{pmatrix}$ then the system (4.19) in matrix-vector form is

$$GU = FF \tag{4.20}$$

Direct solvers for two-dimensional problems are excessively costly on the exceptionally fine meshes which are occasionally needed in order to solve the singularities as well as the steep stress boundary layers which are shown by several constitutive equations as applied in numerical simulations. This is caused by the expense involved in factorising and building the global matrix. A fine computational mesh with a corresponding increase in the magnitude of the algebraic system is needed to resolve the boundary layers. Iterative techniques are necessary in this situation. Such techniques need no storage or inversion of a large matrix, but rather, the fundamental computational procedure is on the basis of matrix-vector multiplications. There are several

iterative techniques to large systems of linear equations. Nevertheless, it is a crucial matter to discover the most effective technique for the current problem, and an inappropriate selection may result in slow convergence or even divergence.

In our case, we used the following iterative algorithm. For given initial values $u^{(0)}$ and $\alpha^{(0)}$, calculate for $k \geq 1$

$$\begin{cases} Au^{(k)} + B\alpha^{(k-1)} = F \\ \\ B^T u^{(k)} + E\alpha^{(k)} = 0 \end{cases} \tag{4.21}$$

In many applications the error cannot be evaluated since the analytical solution is not known and then a stopping tolerance ε is imposed such that

$$\|u^{(k)} - u^{(k-1)}\| < \varepsilon$$

where ε is a sufficiently small constant.

4.3 Numerical simulations

Consider the system (4.1) on a domain $\Omega = [-1,1] \times [-1,1]$ with Dirichlet boundary condition. We choose a forcing $f = 0$ and diffusivity such that

$$\mu(x,y) = \begin{cases} \mu_1, & y \leq \bar{y}, \\ \mu_2, & y > \bar{y}. \end{cases}$$

$A_1, A_2, \eta_1, \eta_2, \mu_1$ and μ_2 are constants. The analytical solution is then

$$u(x,y) = \begin{cases} \sin(\pi x) A_1 \left(e^{\eta_1 y} - e^{-\eta_1 y}\right), & y \leq \bar{y}, \\ \sin(\pi x) \left[A_2 \left(e^{\eta_2 y} - e^{\eta_2 (2-y)}\right) + e^{\eta_2 (1-y)}\right], & y > \bar{y}. \end{cases}$$

Details can be found in [58, 89]. Let us choose $\mu_1 = 1, \mu_2 = 2, \eta_1 = 3.1416$, $\eta_2 = -3.1416$ and $\bar{y} = 0.5001$ so that the solution is discontinuous at $y = 0.5001$. The constant A_1 and A_2 are defined by $A_1 = 0.043295$, $A_2 = -0.043295$. The discrete system was solved using either a direct method or the PCG-algorithm. The exact and approximated solutions are plotted in Fig. 4.5. The L^2-error with respect to the polynomial degree N converges linearly as it can be seen in Fig. 4.6.

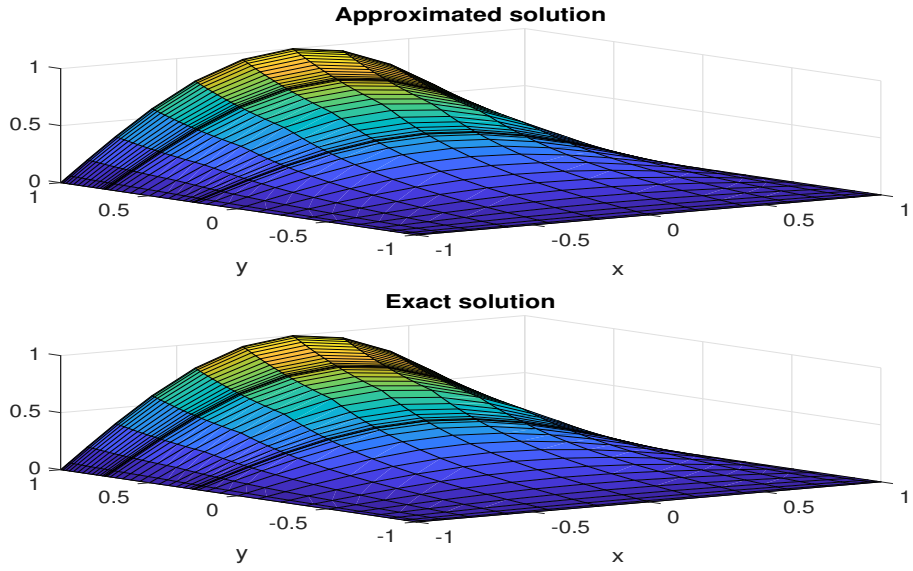


Figure 4.5: Exact and approximated solution for $N = 18$.

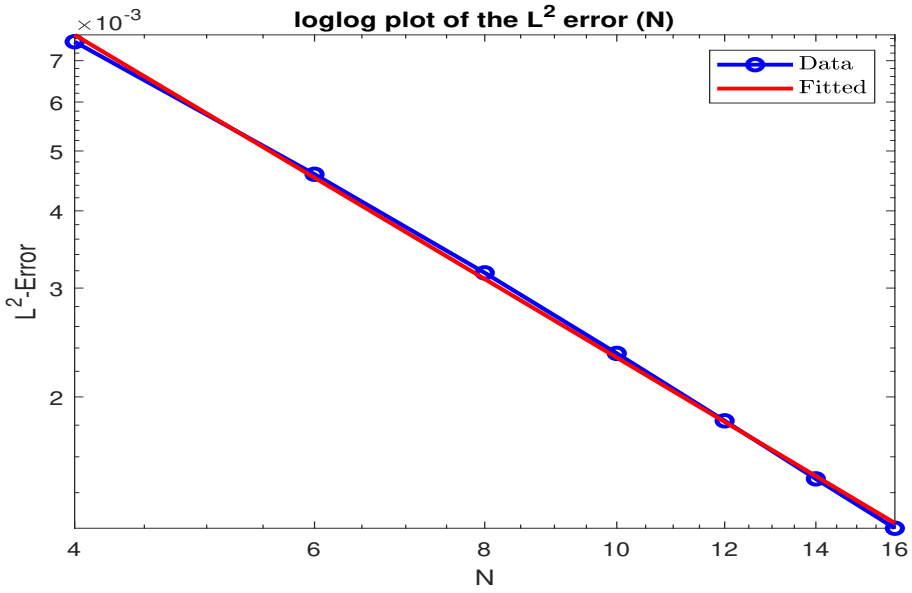


Figure 4.6: L^2-error with respect to N.

4.4 Discontinuity and domain decomposition

In this section we propose an alternative way to resolve numerically the system (4.1) using the spectral element method coupled with a particular domain decomposition. Let $\Omega = [x_a, x_b] \times [y_a, y_b] = \Omega_1 \cup \Omega_2$, $f \in L^2(\Omega)$ and $u_d \in L^2(\partial\Omega)$ where $\Omega_1 \cap \Omega_2 = \Gamma$ is the interface between Ω_1 and Ω_2. We are looking for finding solution u of the following system :

$$\begin{cases} -\nabla(\mu(x,y)\nabla u) = f & \text{on} \quad \Omega \\ u = u_d & \text{on} \quad \partial\Omega \end{cases} \tag{4.22}$$

where

$$\mu(x,y) = \begin{cases} \mu_1 & \text{on } \Omega_1 \\ \mu_2 & \text{on } \Omega_2 \end{cases}$$

μ_1 and μ_2 are two constants such that $\mu_1 \neq \mu_2$.

The idea is based on decomposing the domain into three elements such that the middle element contains the discontinuity and is of size 2ε. This means that the domain

is decomposed as follows

$$\Omega = [x_a, x_b] \times [y_a, y_b] = [x_a, x_b] \times [y_a, \bar{y} - \varepsilon] \cup [x_a, x_b] \times [\bar{y} - \varepsilon, \bar{y} + \varepsilon] \cup [x_a, x_b] \times [\bar{y} + \varepsilon, y_b]$$

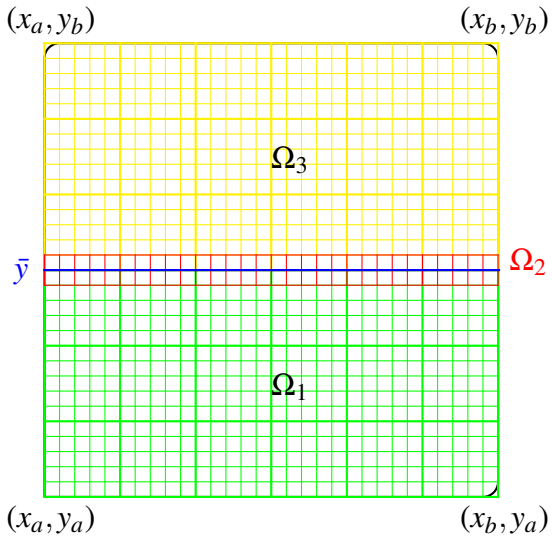


Figure 4.7: Domain two-phase

The spectral element approximation in each of the three elements is:

$$u_N(x,y) = \begin{cases} \sum_{i,j=1}^{N+1} u_{ij}^{(1)} h_i\Big(\dfrac{2x-(x_a+x_b)}{x_b-x_a}\Big) h_j\Big(\dfrac{2y-(y_a+\bar{y}-\varepsilon)}{(\bar{y}-y_a)}\Big) & \text{in} \quad \Omega_1 \\ \sum_{i,j=1}^{N+1} u_{ij}^{(2)} h_i\Big(\dfrac{2x-(x_a+x_b)}{x_b-x_a}\Big) h_j\Big(\dfrac{y-\bar{y}}{\varepsilon}\Big) & \text{in} \quad \Omega_2 \\ \sum_{i,j=1}^{N+1} u_{ij}^{(3)} h_i\Big(\dfrac{2x-(x_a+x_b)}{x_b-x_a}\Big) h_j\Big(\dfrac{2y-(\bar{y}+\varepsilon+y_b)}{y_b-(\bar{y}+\varepsilon)}\Big) & \text{in} \quad \Omega_3 \end{cases}$$

where the solution u_N is expanded in terms of Lagrange interpolants based on the Gauss-Lobatto Legendre points.

The discrete system was solved using either a direct method or the PCG-algorithm. We used the same exact solution as in the previous section. The L^2-error converges exponentially with respect to the parameter ε as it can be seen in Fig. 4.8 for $N = 12$ and in Fig. 4.9 for $N = 14$. Now by fixing $\varepsilon = 10^{-6}$, in Fig. 4.10 the L^2-error converges exponentially with respect to N.

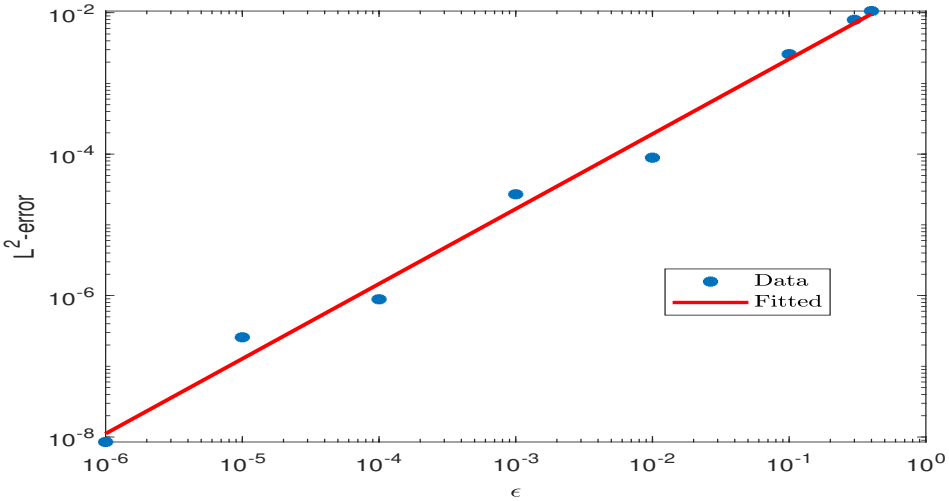


Figure 4.8: log-log plot of the L^2-norm of the error with respect to the value of ε for $N = 12$. The slope is about 1.06.

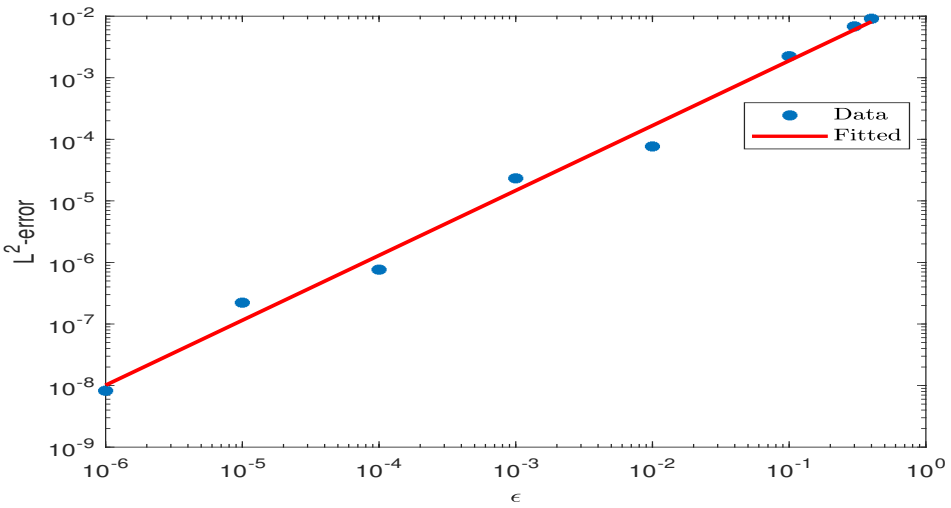


Figure 4.9: log-log plot of the L^2-norm of the error with respect to the value of ε for $N = 14$. The slope is about 1.05.

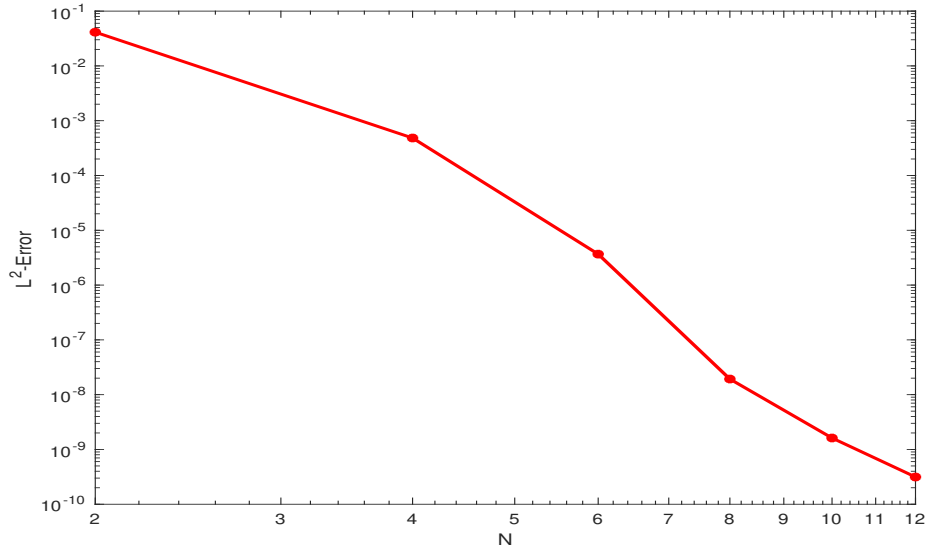


Figure 4.10: L^2-error with respect to N for $\varepsilon = 10^{-6}$.

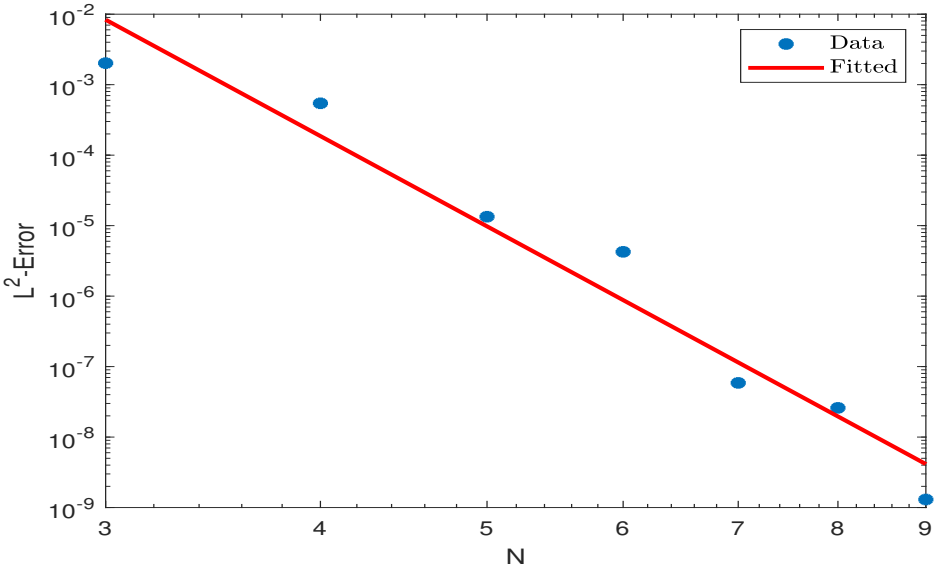


Figure 4.11: log-log plot of the L^2-norm of the error with respect to N for $\varepsilon = 10^{-6}$.

4.5 Conclusion

In this chapter we introduced the spectral element method and illustrated the problems associated with approximating a function which is discontinuous, where the discontinuity is unfitted to the computational mesh. We consider how the presence of a discontinuity influences the convergence and accuracy of the spectral element method (SEM). So we consider a rectangular domain subdivided into two sub-domains (two rectangular) with different characteristics. By applying the spectral element method, a Gibbs phenomenon can be observed. To overcome this we propose an extended spectral element method (XSEM) by adding an enrichment to the classical spectral element method which introduces two additional sparse matrices to the linear system which increase the condition number quite significantly. We showed that when approximating a discontinuous function, XSEM can capture the discontinuity exactly. Using the framework of Reusken [85] we were able to obtain the spectral equivalents of his error estimates. Therefore we applied XSEM to Poisson's equation. We found that the majority of Gibbs phenomena was removed. However, the results were not satisfactory. It is clear from Fig. 4.6 that the XSEM approximation does not improve the solution. This is quite disappointing, however - to a certain degree - not too surprising. More research is required to determine the reason why XSEM does not improve the solution in this case.

Then we propose an alternative way by applying spectral element method on special domain decomposition. The idea is based on decomposing the domain into three elements such that the middle element contains the discontinuity and is of size 2ε. The L^2-error converges exponentially with respect to the parameter ε (size of the domain decomposition) and by fixing ε small enough, the L^2-error converges exponentially with respect to mesh size.

Chapter 5

Approximation of the Stokes Problem using Spectral Element Method

5.1 Introduction

This chapter investigates the approximation of the Stokes problem as well as the consequent conditioning of the discrete problem . The well-posedness of the variational formulation of the Stokes problem and, in particular, the uniqueness of the pressure has been demonstrated when the subspace of square-integrable functions having vanishing mean is chosen for the pressure space. A conforming subspace is selected in the discrete setting for the analysis of the discrete problem as well as the discrete pressure space according to this selection. Nevertheless, there is no utilisation of this space in the practical applications of spectral or finite element methods in the case of the Stokes problem. In this chapter, we demonstrate the means of accommodating the zero volume condition on the pressure within the trial space by modifying the continuity equation in a consistent manner. Furthermore, we investigate the precision of the spectral element approximation on the weighting factor as well as the dependence of the condition number of the preconditioned linear system on N.

5.2 Stokes Problem

Let $\Omega = [x_a, x_b] \times [y_a, y_b]$ be a bounded, connected domain in $\mathbb{R}^2$ with presupposed boundary Γ and $\boldsymbol{f} = (f_1, f_2)^T \in \left(L^2(\Omega)\right)^2$. We search for a solution $(\boldsymbol{u}, p)$ to the fol-

lowing Stokes system :

$$\begin{cases} -\nabla\cdot(\mu\nabla\boldsymbol{u})+\nabla p &= \boldsymbol{f} \quad \text{on} \quad \Omega \\ -\operatorname{div}\boldsymbol{u} &= \boldsymbol{0} \quad \text{on} \quad \Omega \\ \boldsymbol{u} &= \boldsymbol{g} \quad \text{on} \quad \Gamma \end{cases} \tag{5.1}$$

where $\boldsymbol{u}=(u_1,u_2)^T$ indicates the velocity field of an incompressible fluid motion, and p indicates the associated pressure , the function μ is the viscosity of the fluid. In simple terms, we treat the first stage as being homogeneous Dirichlet boundary conditions for the velocity, which is $\boldsymbol{u}|_\Gamma=\boldsymbol{0}$.

Green's theorem tells us that the inhomogeneous Dirichlet data $\boldsymbol{g}$ is required to satisfy the condition

$$\int_\Gamma \boldsymbol{g}\cdot\boldsymbol{n}\,ds=0 \tag{5.2}$$

where $\boldsymbol{n}$ being the outward unit normal vector to Γ. The Stokes problem is of fundamental importance in the study of incompressible fluid flow. Not only does it appear in a primary role as a model for the slow flow of fluids with very high viscosity but it also appears in a supporting capacity as part of the solution process for the Navier–Stokes equations when a time-splitting or semi-implicit approach is adopted.
We may regard the pressure to be a restriction which guarantees the velocity field to be divergence free. The only occurrence of pressure in the Stokes problem is as a gradient in the momentum equation (5.1). Therefore, the pressure can only be determined up to a constant. Two techniques may determine a unique pressure solution. The first of these is to establish the pressure at a specific point in Ω, normally a boundary point or a node, . thereby eliminating the null space. If this method is used, the matching discrete technique for the pressure is non-singular, which we may solve by applying standard direct methods. We should be aware that if the problem's boundary conditions differ, it may not be necessary to impose a zero mean on the pressure. For instance, in

the case of a zero mean condition being enforced on the pressure regarding a problem having an outflow boundary on which a natural, zero normal stress boundary condition is imposed, this will lead to an over-determined problem. When the second technique is enforced, we calculate the pressure from a consistent singular system by utilising a minimisation involving an iterative solver; for instance, the conjugate gradient method. There are many reasons why this method is unpopular in the finite element community. However, the spectral clement community approves of it, where there appears to be less of a reluctance to solve singular systems. Nevertheless, the expansion of round-off may result in the singular method becoming inconsistent, thereby having severe repercussions for the conjugate gradient method convergence.

The analysis of the conventional velocity–pressure formulation of the Stokes problem is based on its reformulation as a saddle point problem. The uniqueness and existence theory for this type of problem was developed by Brezzi [14]. The basis of this analysis is the satisfaction of a compatibility condition between the pressure and velocity function spaces. We sometimes call this the inf-sup or Ladyzhenskaya–Babuška–Brezzi (LBB) [3, 14, 57] condition which is required in order to guarantee a unique pressure solution $L_0^2(\Omega)$, being the space of the square-integrable functions according to the Ω domain with a vanishing mean, the velocity space being $(H^1(\Omega))^2$. Furthermore, the corresponding discrete problem analysis needs the compatibility condition between the pressure spaces and the discrete velocity to be satisfied. False pressure modes can be caused by the violation of this condition. Modes of this kind can affect the pressure approximation precision. Modes of this kind can affect the pressure approximation precision. We are aware that it is insufficient for the velocity and pressure approximation spaces to conform to the subspaces of $(H^1(\Omega))^2$ and $L_0^2(\Omega)$respectively which attains a well-posed discrete problem. However, most theoretical finite element or spectral element monographs do not describe the practical details of implementing conforming finite element or spectral element approximations in this framework. For example, the nodal basis functions that are generally used in the spectral element approximation of pressure do not lie in $L_0^2(\Omega)$, i.e. the nodal basis functions do not have vanishing mean.

The spectral element literature does not document how the problem of implementing conforming approximations is resolved in practice, although a common practice seems to be to modify the pressure approximation a posteriori by adding a constant chosen so that it has zero mean. Note, however, that modal hierarchical basis functions constructed using the Legendre polynomials do belong to $L_0^2(\Omega)$ apart from the constant mode. For finite dimensional approximations that are local in their influence, such as those based on finite elements or certain spectral elements the global conditions on the average of pressure may be inconvenient to use. In this chapter we introduce and analyse a regularized saddle point problem that has a unique pressure solution. Although this formulation of the Stokes problem does not explicitly require the pressure to have vanishing mean, the solution to the regularized problem does possess this property. An additional benefit of this approach is that it results in a positive definite system for the pressure whereas the original formulation produced a positive semi-definite system. A review of the classical statement of the Stokes problem is provided. An alternative statement of the Stokes problem is given and analyzed. The spectral element discretization of the Stokes problem is developed. Numerical results demonstrating the performance of the PCG method and the efficiency of the preconditioners are presented.

5.3 Classical statement of the Stokes problem

The Stokes problem comprises the conservation laws of momentum and mass (5.1) in which $\boldsymbol{u}$ is the velocity, p is the pressure and $\boldsymbol{f}$ is the body force. Consider the solution of the problem (5.1) in some bounded, connected domain Ω in $\mathbb{R}^2$, with a Lipschitz continuous boundary Γ. In this section we provide a review of the classical statement of the Stokes problem, including existence and uniqueness results. The following existence and uniqueness result holds:

Theorem 1. *Let Ω be a bounded and connected subset of $\mathbb{R}^2$ with a Lipschitz continuous boundary Γ. Let $\boldsymbol{f}$ and $\boldsymbol{g}$ be two given functions in $(H^{-1}(\Omega))^2$ and $(H^{1/2}(\Gamma))^2$,*

respectively, such that $\int_\Gamma \boldsymbol{g}\cdot\boldsymbol{n}\,ds = 0$ *where* $\boldsymbol{n}$ *is the unit outward normal to* Γ. *Then, there exists a unique solution* $(\boldsymbol{u},p)\in(H^1(\Omega))^2\times L^2_0(\Omega)$ *satisfying (5.1).*

The proof of this theorem is based on the equivalence of the solution to this problem with the solution to a corresponding variational problem, and involves an application of the theory of saddle point problems (see Chorin [22]). If the boundary data $\boldsymbol{g}$ is a function in $(H^{1/2}(\Gamma))^2$ satisfying (5.1) then Temam [100] has shown that there exists a function $\boldsymbol{u_0}\in(H^1(\Omega))^2$ such that

$$\operatorname{div}\boldsymbol{u_0}=\boldsymbol{0} \quad \text{on} \quad \Omega \quad \text{and} \quad \boldsymbol{u_0}=\boldsymbol{g} \quad \text{on} \quad \Gamma \tag{5.3}$$

This outcome allows one to articulate the inhomogeneous Stokes problem (5.1) with regard to an equivalent one having homogeneous Dirichlet boundary conditions:

$$\begin{cases} -\nabla\cdot(\mu\nabla\boldsymbol{w})+\nabla p &= \boldsymbol{f}+\nabla\cdot(\mu\nabla\boldsymbol{u_0}) \quad \text{on} \quad \Omega \\ -\operatorname{div}\boldsymbol{w} &= \boldsymbol{0} \quad \text{on} \quad \Omega \\ \boldsymbol{w} &= \boldsymbol{0} \quad \text{on} \quad \Gamma \end{cases} \tag{5.4}$$

Therefore, the solution to (5.1) is $\boldsymbol{u}=\boldsymbol{w}+\boldsymbol{u_0}$. In this chapter, in the discussion of the theoretical treatment of the Stokes problem, we presuppose that, without any loss of generality, the velocity satisfies homogeneous Dirichlet boundary conditions, that is $\boldsymbol{u}=\boldsymbol{0}$ on Γ.

In order to implement the variational formulation of this problem, we apply:

$$\begin{cases} -\nabla\cdot(\mu\nabla\boldsymbol{u})+\nabla p &= \boldsymbol{f} \quad \text{on} \quad \Omega \\ -\operatorname{div}\boldsymbol{u} &= \boldsymbol{0} \quad \text{on} \quad \Omega \\ \boldsymbol{u} &= \boldsymbol{0} \quad \text{on} \quad \Gamma \end{cases} \tag{5.5}$$

It is necessary to implement function spaces for both velocity and pressure as indicated by V and Q, respectively. With regard to the conventional statement of the Stokes problem, these are provided as follows:

$$V = \left(H_0^1(\Omega)\right)^2 \quad \text{and} \quad Q = L_0^2(\Omega) = \left\{q \in L^2(\Omega): \quad \int_\Omega q\, d\Omega = 0\right\}$$

We need to be aware that the selection of pressure space as aforementioned eliminates any indeterminacy in the pressure. Furthermore, this may be attained through several other means, which include the pressure specification at one point within the domain. The particular choice of Q, as mentioned previously is appropriate when the variational formulation is discretised by applying spectral methods. In order to obtain a discrete pressure solution with zero average, it is basically necessary to ignore its lowest Fourier coefficient which is the one that is connected to the constant polynomial [42].

5.3.1 Weak fomulation

We multiply the momentum equation by the test function $\boldsymbol{v} \in \left(H_0^1(\Omega)\right)^2$ and the mass (continuity) equation by the test function $q \in L_0^2(\Omega)$. For the momentum equation, we use integration by parts in order to get the weak formulation of the Stokes problem: Find $\boldsymbol{u} \in \left(H_0^1(\Omega)\right)^2$ and $p \in L_0^2(\Omega)$ such that

$$\begin{cases} (\mu\,\nabla\boldsymbol{u},\nabla\boldsymbol{v}) - (p, \operatorname{div}\boldsymbol{v}) = (\boldsymbol{f},\boldsymbol{v}), & \forall \boldsymbol{v} \in \left(H_0^1(\Omega)\right)^2 \\ -(\operatorname{div}\boldsymbol{u}, q) = 0, & \forall q \in L^2(\Omega) \end{cases} \tag{5.6}$$

The required conditions for the well-posedness of a saddle point system are called inf-sup conditions. The setting for the Stokes equations is as follows:

- Define the spaces $\mathbf{V} = \left(H_0^1(\Omega)\right)^2$ with semi-norm $|\boldsymbol{v}|_1 = \|\nabla \boldsymbol{v}\|$, and $\mathbf{Q} = L_0^2(\Omega) = \left\{q \in L^2(\Omega), \int_\Omega q = 0\right\}$, with norm $\|p\|$.
- Define the bilinear forms $a(\boldsymbol{u},\boldsymbol{v}) = \int_\Omega \mu\, \nabla \boldsymbol{u} : \nabla \boldsymbol{v}$ and $b(\boldsymbol{v},q) = -\int_\Omega (\operatorname{div} \boldsymbol{v})\, q$.

For the purpose of summarising the spectral element method, we begin with the problem's variational formulation (5.5). We seek the solution $(\boldsymbol{u},p)$ in $\mathbf{V}\times\mathbf{Q}$, being the function spaces for the velocity and pressure respectively. Subsequently, we can attain the weak formulation:

$$\begin{cases} \text{Find} \quad (\boldsymbol{u},p) \in \mathbf{V}\times\mathbf{Q} \quad \text{such that} \\ \\ a(\boldsymbol{u},\boldsymbol{v}) + b(\boldsymbol{v},p) = (\boldsymbol{f},\boldsymbol{v}), \qquad \forall \boldsymbol{v} \in \mathbf{V} \\ \\ b(\boldsymbol{u},q) = 0, \qquad \forall q \in \mathbf{Q} \end{cases} \tag{5.7}$$

If this problem is to be solved, it is necessary to verify the following conditions to have a unique solution for the problem (5.7) :

1. The coercivity condition of the bilinear form a is given by
$$\inf_{\boldsymbol{u}\in\mathbf{V}} \sup_{\boldsymbol{v}\in\mathbf{V}} \frac{a(\boldsymbol{u},\boldsymbol{v})}{|\boldsymbol{u}|_1|\boldsymbol{v}|_1} = \inf_{\boldsymbol{v}\in\mathbf{V}} \sup_{\boldsymbol{u}\in\mathbf{V}} \frac{a(\boldsymbol{u},\boldsymbol{v})}{|\boldsymbol{u}|_1|\boldsymbol{v}|_1} = \alpha > 0.$$
2. The inf-sup condition is given by
$$\inf_{q\in\mathbf{Q}} \sup_{\boldsymbol{v}\in\mathbf{V}} \frac{b(\boldsymbol{v},q)}{|\boldsymbol{v}|_1\|q\|} = \beta > 0.$$
3. The bilinear forms continuity is given by
$$a(\boldsymbol{u},\boldsymbol{v}) \le k_1|\boldsymbol{u}|_1|\boldsymbol{v}|_1, \qquad \forall \boldsymbol{u},\boldsymbol{v} \in \mathbf{V}$$
$$b(\boldsymbol{v},q) \le k_2|\boldsymbol{v}|_1\|q\|, \quad \forall \boldsymbol{v} \in \mathbf{V}, \quad \forall q \in \mathbf{Q}$$
where $k_1, k_2 > 0$ and $|\cdot|_1 = \|\nabla \cdot\|$ is the H^1 semi-norm.

5.3.2 Alternative continuity equation

We have addressed, in the previous sections, the normal continuity equation for an incompressible fluid, $\operatorname{div} \boldsymbol{u} = 0$. Gwynllyw and Phillips [42], in 2006, investigated an alternative formulation of the Stokes problem by replacing the continuity equation by:

$$-\operatorname{div} \boldsymbol{u} = \lambda \int_{\Omega} p \, d\Omega \quad \text{in} \quad \Omega \tag{5.8}$$

where λ is a positive constant.

The statement continues to be equivalent to the original equation since the pressure has a vanishing mean over the domain Ω. We may see this by integrating (5.8) over Ω and using Green's Theorem. It is advantageous that the vanishing mean property is not certain to be satisfied in one element of the domain. However, this alternative formulation ensures that the pressure possesses the vanishing mean across the domain; therefore, it is certain that $p \in L_0^2(\Omega)$. This means that a unique solution for the pressure may be determined instead of a solution up to a constant for the conventional formulation. Although this formulation was initially applied to the Stokes problem, it is clear that the alternative formulation is also valid for the Navier-Stokes equations. Furthermore, another advantage for the spectral element method is the enhancement in the conditioning of the linear system resulting from the spatial discretisation. We shall address this in a later part of the chapter. We shall now use the alternative form of the continuity equation,(5.8) since we are aware that we may recover the original formulation by setting $\lambda = 0$.

5.4 Approximation using Spectral Element Method

The Spectral Element Method (SEM) is used to discretize the Stokes problem's weak formulation of the Stokes problem.

We define the nodes $x_i = y_i$ and weights w_i, $i = 1, \ldots, N+1$ as in Chapter 1.

We need to choose conforming discrete subspaces for pressure and velocity when applying the spectral element method. Let $\mathbf{V}_N \subset \mathbf{V}$ and $\mathbf{Q}_N \subset \mathbf{Q}$ indicate these subspaces, respectively. Let $P_N(\Omega)$ indicate the space of all polynomials on Ω of degree less than or equal to N and define:

$$P_N(\Omega) := \{\varphi : \ \varphi|_\Omega \in P_N(\Omega)\}$$

We may define the velocity and pressure approximation spaces as:

$$\mathbf{V}_N := \mathbf{V} \cap \Big[P_N(\Omega)\Big]^2, \qquad \mathbf{Q}_N := \mathbf{Q} \cap P_{N-2}(\Omega)$$

Subsequently, the velocity $\boldsymbol{u} = (u_1, u_2)^T$ is expanded in terms of Lagrange interpolants based on the Gauss-Lobatto Legendre points

$$u_{1N}(x,y) = \sum_{i,j=1}^{N+1} u_{1ij} h_i(x) h_j(y)$$

$$u_{2N}(x,y) = \sum_{i,j=1}^{N+1} u_{2ij} h_i(x) h_j(y)$$

where $u_{1ij} = u_1(x_i, y_j), u_{2ij} = u_2(x_i, y_j), i, j = 1, \ldots, N+1$, and h_i are the Lagrange interpolants given by

$$h_i(x) = \prod_{\substack{k=1,\\ k\neq i}}^{N+1} \frac{(x-x_k)}{(x_i-x_k)} = \frac{-(1-x^2)L'_N(x)}{N(N+1)L_N(x_i)(x-x_i)} = \frac{L_{N+1}(x) - L_{N-1}(x)}{(2N+1)L_N(x_i)(x-x_i)}, i = 1, \ldots, N+1$$

defined on the set of GLL interpolation nodes $\{x_i\}_{i=1}^{N+1}$.
The pressure p is expanded in terms of the Lagrange interpolants based on the interior Gauss-Lobatto Legendre points

$$p_N(x,y) = \sum_{i,j=2}^{N} p_{ij} \tilde{h}_i(x) \tilde{h}_j(y)$$

where $p_{ij} = p(x_i, y_j), i, j = 2, \ldots, N$ and $\tilde{h}_i$ are the Lagrange interpolants given by

$$\tilde{h}_i(x) = \frac{(1-x_i^2)}{(1-x^2)} h_i(x) = \prod_{\substack{k=2\\ k\neq i}}^{N} \frac{(x-x_k)}{(x_i-\tilde{x}_k)} = \frac{-(1-x_i^2)L'_N(x)}{N(N+1)L_N(x_i)(x-x_i)}, i = 2, \ldots, N$$

defined on the set of the interior interpolation nodes $\{x_i\}_{i=2}^{N}$.

The only difference between these formulae is the selection of the interpolating nodes. The interpolation points are the nodes within the corresponding GLL quadrature formula, and the Stokes equation spectral element approximation produces the following semi-discrete problem: Find $\boldsymbol{u}_N = (u_{1N}, u_{2N}) \in \mathbf{V}_N$ and $p_N \in \mathbf{Q}_N$ such that

$$\begin{cases} (\mu \nabla \boldsymbol{u}_N, \nabla \boldsymbol{v}_N)_N - (\nabla \cdot \boldsymbol{v}_N, p_N)_N &= (\boldsymbol{f}_N, \boldsymbol{v}_N)_N, \ \forall \boldsymbol{v}_N \in \mathbf{V}_N \\ -(\nabla \cdot \boldsymbol{u}_N, q_N)_N &= \lambda (p_N, 1)_N (q_N, 1)_N, \ \forall q_N \in \mathbf{Q}_N \end{cases} \tag{5.9}$$

where $\boldsymbol{u}_N = (u_{1N}, u_{2N}), p_N$ and $\boldsymbol{f}_N$ indicate the approximations of the velocity, pressure and source term, respectively, μ denotes the dimensionless viscosity and $(\cdot,\cdot)$ the standard $L^2(\Omega)$ inner product. We define the discrete inner product by: $(\cdot,\cdot)_N$

$$(\phi, \psi)_N = \sum_{m,n=1}^{N+1} w_m w_n \phi(x_m, y_n) \psi(x_m, y_n)$$

Subsequently, we define the discrete L^2-norm by

$$\|\phi\|_{L^2}^2 = \sum_{m,n=1}^{N+1} w_m w_n \big[\phi(x_m, y_n)\big]^2$$

The discrete weak formulation then becomes: find $\boldsymbol{u}_N = (u_{1N}, u_{2N}) \in \mathbf{V}_N$ and $p_N \in \mathbf{Q}_N$ such that

$$\begin{cases} (\mu \nabla u_{1N}, \nabla v_{1N})_N - \left(\dfrac{\partial v_{1N}}{\partial x}, p_N\right)_N &= (f_{1N}, v_{1N})_N, \\ (\mu \nabla u_{2N}, \nabla v_{2N})_N - \left(\dfrac{\partial v_{2N}}{\partial y}, p_N\right)_N &= (f_{2N}, v_{2N})_N, \ \forall (v_{1N}, v_{2N}) \in \mathbf{V}_N \\ -\left(\dfrac{\partial u_{1N}}{\partial x} + \dfrac{\partial u_{2N}}{\partial y}, q_N\right)_N &= \lambda (p_N, 1)_N (q_N, 1)_N, \ \forall q_N \in \mathbf{Q}_N \end{cases} \tag{5.10}$$

The optimal a priori error estimate [42] is given by

$$\|\boldsymbol{u} - \boldsymbol{u}_N\|_{\mathbf{V}} + \|p - p_N\|_{\mathbf{Q}} \leq L\Big(\inf_{\boldsymbol{w}_N \in \mathbf{V}} \|\boldsymbol{u} - \boldsymbol{w}_N\|_{\mathbf{V}} + \inf_{r_N \in \mathbf{Q}} \|p - r_N\|_{\mathbf{Q}}\Big) \tag{5.11}$$

By injecting the decomposition of $\boldsymbol{u}_N = (u_{1N}, u_{2N})$ and p_N into the variational formulation (5.9) and choose the test functions as $v_{1N} = v_{2N} = h_k(x)h_l(y) \;\; (2 \leq k,l \leq N)$ and $q_N = \tilde{h}_k(x)\tilde{h}_l(y) \;\; (2 \leq k,l \leq N)$, one attains the following discrete system:

$$\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=1}^{N+1} \Big(h_i'(x_m)h_j(y_n)h_k'(x_m)h_l(y_n) + h_i(x_m)h_j'(y_n)h_k(x_m)h_l'(y_n) \Big) u_{1ij}$$

$$- \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=2}^{N} \tilde{h}_i(x_m)\tilde{h}_j(y_n)h_k'(x_m)h_l(y_n)p_{ij} = \sum_{m,n=1}^{N+1} w_m w_n f_{1mn} h_k(x_m)h_l(y_n),$$

$$\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=1}^{N+1} \Big(h_i'(x_m)h_j(y_n)h_k'(x_m)h_l(y_n) + h_i(x_m)h_j'(y_n)h_k(x_m)h_l'(y_n) \Big) u_{2ij}$$

$$- \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=2}^{N} \tilde{h}_i(x_m)\tilde{h}_j(y_n)h_k(x_m)h_l'(y_n)p_{ij} = \sum_{m,n=1}^{N+1} w_m w_n f_{2mn} h_k(x_m)h_l(y_n),$$

$$- \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=1}^{N+1} h_i'(x_m)h_j(y_n)\tilde{h}_k(x_m)\tilde{h}_l(y_n)u_{1ij}$$

$$- \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=1}^{N+1} h_i(x_m)h_j'(y_n)\tilde{h}_k(x_m)\tilde{h}_l(y_n)u_{2ij}$$

$$= \lambda \left(\sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=2}^{N} \tilde{h}_i(x_m)\tilde{h}_j(y_n)p_{ij} \right) \left(\sum_{m,n=1}^{N+1} w_m w_n \tilde{h}_k(x_m)\tilde{h}_l(y_n) \right),$$

which is equivalent to

$$\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=1}^{N+1} \Big(D_{mi} D_{mk} \delta_{nlj} + D_{nj} D_{nl} \delta_{mik} \Big) u_{1ij}$$

$$- \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=2}^{N} D_{mk} \delta_{mi} \delta_{njl} p_{ij} = \sum_{m,n=1}^{N+1} w_m w_n f_{1mn} \delta_{mk} \delta_{nl},$$

$$\sum_{m,n=1}^{N+1} w_m w_n \mu_{mn} \sum_{i,j=1}^{N+1} \Big(D_{mi} D_{mk} \delta_{nlj} + D_{nj} D_{nl} \delta_{mik} \Big) u_{2ij}$$

$$- \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=2}^{N} D_{nl} \delta_{mik} \delta_{nj} p_{ij} = \sum_{m,n=1}^{N+1} w_m w_n f_{2mn} \delta_{mk} \delta_{nl},$$

$$- \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=1}^{N+1} D_{mi} \delta_{mk} \delta_{njl} u_{1ij} - \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=1}^{N+1} D_{nj} \delta_{mik} \delta_{nl} u_{2ij}$$

$$= \lambda \sum_{m,n=1}^{N+1} w_m w_n \sum_{i,j=2}^{N} \delta_{mi} \delta_{nj} p_{ij} \sum_{m,n=1}^{N+1} w_m w_n \delta_{mk} \delta_{nl}.$$

Making use of the Kronecker delta property, the equations reduce to the following

$$\sum_{m=1}^{N+1} w_m w_l \mu_{ml} \sum_{i=1}^{N+1} D_{mi} D_{mk} u_{1il} + \sum_{n=1}^{N+1} w_k w_n \mu_{kn} \sum_{j=1}^{N+1} D_{nj} D_{nl} u_{1kj} - \sum_{i=2}^{N} w_i w_l D_{ik} p_{il} = w_k f_{1kl} w_l,$$

$$\sum_{m=1}^{N+1} w_m w_l \mu_{ml} \sum_{i=1}^{N+1} D_{mi} D_{mk} u_{2il} + \sum_{n=1}^{N+1} w_k w_n \mu_{kn} \sum_{j=1}^{N+1} D_{nj} D_{nl} u_{2kj} - \sum_{j=2}^{N} w_k w_j D_{jl} p_{kj} = w_k f_{2kl} w_l,$$

$$- \sum_{i=1}^{N+1} w_k w_l D_{ki} u_{1il} - \sum_{j=1}^{N+1} w_k w_l D_{lj} u_{2kj} - \lambda \; w_k w_l \sum_{i,j=2}^{N} w_i w_j p_{ij} = 0.$$

Since u_{1N} and u_{2N} satisfy homogenous Dirichlet boudary conditions, their expression can be simplified to

$$u_{1N}(x,y) = \sum_{i,j=2}^{N} u_{1ij} h_i(x) h_j(y) \text{ and } u_{2N}(x,y) = \sum_{i,j=2}^{N} u_{2ij} h_i(x) h_j(y) \tag{5.12}$$

Let $I = (l-2)(N-1)+(k-1)$ and $J = (j-2)(N-1)+(i-1)$. Therefore, the system can be written in the more compact form

$$\begin{aligned}
&\sum_{J=1}^{(N-1)^2} A_{IJ}u_{1J} - \sum_{J=1}^{(N-1)^2} C_{3IJ}p_J = F_{1I}, \quad I = 1,\cdots,(N-1)^2,\\
&\sum_{J=1}^{(N-1)^2} A_{IJ}u_{2J} - \sum_{J=1}^{(N-1)^2} C_{4IJ}p_J = F_{2I}, \quad I = 1,\cdots,(N-1)^2,\\
&-\sum_{J=1}^{(N-1)^2} C_{1IJ}u_{1J} - \sum_{J=1}^{(N-1)^2} C_{2IJ}u_{2J} - \lambda \sum_{J=1}^{(N-1)^2} B_{IJ}p_J = \mathbf{0}, \quad I = 1,\cdots,(N-1)^2.
\end{aligned}$$

where

$$\begin{aligned}
A_{IJ} &= \delta_{lj}\sum_{m=1}^{N+1} D_{mi}D_{mk}w_m\mu_{ml}w_l + \delta_{ki}\sum_{n=1}^{N+1} D_{nj}D_{nl}w_k\mu_{kn}w_n\\
C_{1IJ} &= \delta_{lj}D_{ki}w_kw_l\\
C_{2IJ} &= \delta_{ki}D_{lj}w_kw_l\\
C_{3IJ} &= \delta_{lj}D_{ik}w_iw_l\\
C_{4IJ} &= \delta_{ki}D_{jl}w_kw_j\\
B_{IJ} &= w_kw_l\ w_iw_j
\end{aligned}$$

and

$$F_{iI} = w_kw_lf_{ikl}, \quad i = 1,2.$$

Note that $C_3 = C_1^T$ and $C_4 = C_2^T$, then the system is equivalent to

$$\begin{cases}
Au_1 & & -C_1^Tp &=& F_1\\
& Au_2 & -C_2^Tp &=& F_2\\
-C_1u_1 & -C_2u_2 & -\lambda Bp &=& 0
\end{cases} \tag{5.13}$$

or in matrix-vector form

$$\begin{pmatrix} A & 0 & -C_1^T\\ 0 & A & -C_2^T\\ -C_1 & -C_2 & -\lambda B \end{pmatrix}\begin{pmatrix} u_1\\ u_2\\ p \end{pmatrix} = \begin{pmatrix} F_1\\ F_2\\ 0 \end{pmatrix} \tag{5.14}$$

One method of decoupling the set of equations as suggested by Maday and Patera (1989) is by writing

$$\begin{cases} u_1 = A^{-1}(C_1^T p + F_1) \\ \\ u_2 = A^{-1}(C_2^T p + F_2) \\ \\ C_1 u_1 + C_2 u_2 + \lambda B p = 0 \end{cases}$$

this we can do since A is symmetric and positive definite and subsequently A^{-1} exists. The multiplication of the first two equations by C_1 and C_2, respectively and substitution into the third equation produces:

$$C_1 u_1 + C_2 u_2 + \lambda B p = C_1 A^{-1}(C_1^T p + F_1) + C_2 A^{-1}(C_2^T p + F_2) + \lambda B p = 0$$

therefore, we may obtain an equation for the pressure:

$$\left(C_1 A^{-1} C_1^T + C_2 A^{-1} C_2^T + \lambda B\right) p = -\left(C_1 A^{-1} F_1 + C_2 A^{-1} F_2\right). \tag{5.15}$$

As soon as the pressure has been determined; for example, by utilising an iterative solver for (5.15), we can calculate the velocity by solving the decoupled systems:

$$\begin{cases} A u_1 = F_1 + C_1^T p \\ \\ A u_2 = F_2 + C_2^T p \end{cases} \tag{5.16}$$

We remove any columns or rows corresponding to a Dirichlet node in the global matrix for the purpose of applying the boundary conditions. Subsequently, we substitute the known values into the solution at these nodes and undertake the matrix-vector calculations with the corresponding rows with the global matrix. This contributes to subtracting from the right hand side of the linear system.

5.5 Preconditioning

Direct solvers' cost becomes prohibitive, even for two-dimensional problems, on the extremely fine meshes which are sometimes required to resolve the steep stress boundary layers in addition to the singularities shown by several constitutive equations which are used in numerical simulations. The reason for this is the cost of the assembly and factorisation of the global matrix. A fine computational mesh with a corresponding increase in the size of the algebraic system is needed for resolving the boundary layers, and in this situation, iterative methods are necessary. In the case of iterative methods, the inversion or storage of a large matrix is not necessary, but rather, the fundamental computational procedure matrix-vector multiplications. There are numerous iterative methods for solving large systems of linear equations. Nevertheless, the critical point is to discover the most efficient method for the problem which is being considered. An inappropriate choice could result in slow convergence or even in divergence. The inf-sup stability condition, which is linked to the mixed approximation of the Stokes Problem, is of major significance regarding the discovery of fast and dependable iterative solution methods. Although it is certainly unnecessary to apply stable approximations to calculate a precise velocity field at all times, it is essential to do so for the construction of fast convergent iterative methods. This topic is addressed in this section.
We desire to solve the following system:

$$K\boldsymbol{x} = \boldsymbol{b}, \tag{5.17}$$

where K is an $n \times n$ matrix and $\boldsymbol{x}$ and $\boldsymbol{b}$ are n-dimensional column vectors have a certain structure and given by:

$$K = \begin{pmatrix} A & 0 & -C_1^T \\ 0 & A & -C_2^T \\ -C_1 & -C_2 & -\lambda B \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} u_1 \\ u_2 \\ p \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} F_1 \\ F_2 \\ 0 \end{pmatrix}. \tag{5.18}$$

where K is symmetric, and reflects the self-adjointness of the continuous Stokes operator, but is always indefinite, having both positive and negative eigenvalues. We note that the Laplacian matrix A is always positive, definite and has the dimension $(N-1)^2 \times (N-1)^2$.

Several iterative methods may be applied in order to solve this system. The oldest and best-known member of the non-stationary iterative methods is the conjugate gradient (CG) method. In a nonstationary iterative method there is no element of choice in the determination of the iteration parameters. They are dynamically selected at each iteration in order to minimise the error in a particular norm (L^2-norm, for example). This method was developed in order to resolve symmetric, positive definite systems of linear equations. It converges extremely rapidly when the eigenvalues of K are clustered or lie within distinct clustered groups. The convergence rate for the CG method is dependent on the condition number of the coefficient matrix. In a situation where the eigenvalues of K are not clustered, the convergence of the CG method is slow. However, we may enhance this by preconditioning the system with an appropriate non-singular matrix P. The concept which underlies the preconditioning philosophy is to change the initial system into an equivalent and improved conditioned system

$$P^{-1}K\boldsymbol{x} = P^{-1}\boldsymbol{b}, \tag{5.19}$$

where P is the preconditioner, which is chosen as an approximation of K with the eigenvalues of $P^{-1}K$ clustered close to unity. In an ideal situation, the properties of the preconditioner ought to be similar to the original matrix as well as being sparse in order to be efficient to construct and to store. It is essential to consider the trade-off between application and construction cost of the preconditioner as well as the anticipated increase in the convergence speed of the iterative method. The inverse P^{-1} is not constructed explicitly, but rather, the linear system of the form

$$P\boldsymbol{y} = \boldsymbol{c}, \tag{5.20}$$

are solved. However, we are indeed aware of the existence of two extreme cases. lf $P = K$ then (5.20) is the same except for there being a possible change of the right side

vector. Therefore, (5.17) the application of the preconditioner is equally as difficult as finding a solution to the initial problem. If $P = I$, then (5.19) is equal to (5.17) and it is trivial to apply the preconditioning because it is still necessary to invert K.

However, the linear system (5.19) transformation (5.17) is not used in practice in the calculations because $P^{-1}K$ may not necessarily be symmetric and positive definite, but rather the preconditioner is decomposed in the form $P = HH^T$ and the transformed system is written as

$$H^{-1}KH^{-T}(H^T\boldsymbol{x}) = H^{-1}\boldsymbol{b}. \tag{5.21}$$

The convergence behaviour of the preconditioned method is dependent on the spectrum of $R = H^{-1}KH^{-T}$. A given iterative method is preconditioned as follows:

1. Transform the right-hand side vector according to

$$\tilde{\boldsymbol{b}} = H^{-1}\boldsymbol{b}$$

2. Apply the iterative method to the system

$$R\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}},$$

where $\tilde{\boldsymbol{x}} = H^T\boldsymbol{x}$.

3. Compute

$$\boldsymbol{x} = H^{-T}\tilde{\boldsymbol{x}}$$

In practice, the decomposition of P, as given in (5.21) is is not required. The stages of the conjugate gradient method may be rewritten to enable the preconditioner to be applied in its entirety:

The Preconditioned Conjugate Gradient (PCG) algorithm

1. Select an initial guess $\boldsymbol{x}_0$ and compute $\boldsymbol{r}_0 = \boldsymbol{b} - K\boldsymbol{x}_0$.

2. Solve $P\boldsymbol{z}_0 = \boldsymbol{r}_0$. Set $\boldsymbol{p}_0 = \boldsymbol{z}_0$.

3. For $n = 1, 2, \ldots$, compute

$$\begin{aligned}
\alpha_n &= \boldsymbol{r}_{n-1}^T \boldsymbol{z}_{n-1} / \boldsymbol{p}_{n-1}^T K \boldsymbol{p}_{n-1} \\
\boldsymbol{x}_n &= \boldsymbol{x}_{n-1} + \alpha_n \boldsymbol{p}_{n-1} \\
\boldsymbol{r}_n &= \boldsymbol{r}_{n-1} - \alpha_n K p_{n-1} \\
\boldsymbol{z}_n &= P^{-1} \boldsymbol{r}_n \\
\beta_n &= \boldsymbol{r}_n^T \boldsymbol{z}_n / \boldsymbol{r}_{n-1}^T \boldsymbol{z}_{n-1} \\
\boldsymbol{p}_n &= \boldsymbol{z}_n + \beta_n \boldsymbol{p}_{n-1}
\end{aligned}$$

until convergence stopping criterion is satisfied.

The convergence behaviour of the preconditioned method depends on the spectrum of the matrix $H^{-1}KH^{-T}$. The theory in [33] shows that convergence of the preconditioned CG iteration depends on the eigenvalues of $H^{-1}KH^{-T}$, which are identical to the eigenvalues of $P^{-1}K$ because of the similarity transformation $H^{-T}H^{-1}KH^{-T}H^T = P^{-1}K$. Thus, introducing the loose notation

$$\kappa(P^{-1}K) = \frac{\lambda_{\max}(H^{-1}KH^{-T})}{\lambda_{\min}(H^{-1}KH^{-T})}$$

we are able to demonstrate that $\frac{1}{2}\sqrt{\kappa(P^{-1}K)}|\log(\varepsilon)/2|$ preconditioned CG iterations will be needed if we are to reduce the K-norm of the error by ε (pre-selected tolerance). In particular, if a preconditioner P can be found such that $\kappa(P^{-1}K)$ is bounded independently of N then, for a fixed convergence tolerance ε, the number of required iterations will not increase when we search for more accurate solutions by applying more refined grids.

The bound based on the discrete eigenvalues λ_j of the operator $P^{-1}K$ is given by

[33]

$$\frac{\|\boldsymbol{r}_k\|_{P^{-1}}}{\|\boldsymbol{r}_0\|_{P^{-1}}} \leq \min_{p_k \in \Pi_k, p_k(0)=1} \max_j |p_k(\lambda_j)| \tag{5.22}$$

where Π_k is the set of real polynomials of degree less than or equal to k and p_k is a real polynomial of degree less than or equal to k. This bound shows that the rate of convergence is dependent on the eigenvalues of the generalised eigenvalue problem

$$K\boldsymbol{x} = \lambda P\boldsymbol{x} \tag{5.23}$$

5.6 General strategies for preconditioning

For the Stokes problem, we assess the discretization error in the energy norm for velocities and in the L_2 norm for pressure. Therefore, the natural matrix norm is $\|\boldsymbol{e}^{(k)}\|_E$ where

$$E = \begin{pmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & Q \end{pmatrix} \tag{5.24}$$

where Q is the pressure mass matrix with entries given by:

$$Q_{IJ} = \int\int \tilde{h}_i(x)\tilde{h}_j(y)\tilde{h}_k(x)\tilde{h}_l(y)dxdy.$$

where $I = (l-2)(N-1)+(k-1)$ and $J = (j-2)(N-1)+(i-1)$. In the case of SEM, its approximation obtained by means of a GLL quadrature rule given by

$$Q_{IJ} = \sum_{m,n=1}^{N+1} w_m w_n \tilde{h}_i(x_m)\tilde{h}_j(y_n)\tilde{h}_k(x_m)\tilde{h}_l(y_n).$$

Using the fact that

$$\sum_{m,n=1}^{N+1} a_{m,n} = \sum_{m,n=2}^{N} a_{m,n} + \sum_{m=2}^{N} a_{m,1} + \sum_{m=2}^{N} a_{m,N+1} + \sum_{n=2}^{N} a_{1,n} + \sum_{n=2}^{N} a_{N+1,n}$$

$$+a_{1,1}+a_{1,N+1}+a_{N+1,1}+a_{N+1,N+1}$$

we can express the entries of Q in the form

$$\begin{aligned}
Q_{IJ} = & \sum_{m,n=2}^{N} w_m w_n \tilde{h}_i(x_m)\tilde{h}_j(y_n)\tilde{h}_k(x_m)\tilde{h}_l(y_n) \\
& + \sum_{m=2}^{N} w_m w_1 \tilde{h}_i(x_m)\tilde{h}_j(y_1)\tilde{h}_k(x_m)\tilde{h}_l(y_1) \\
& + \sum_{m=2}^{N} w_m w_{N+1} \tilde{h}_i(x_m)\tilde{h}_j(y_{N+1})\tilde{h}_k(x_m)\tilde{h}_l(y_{N+1}) \\
& + \sum_{n=2}^{N} w_1 w_n \tilde{h}_i(x_1)\tilde{h}_j(y_n)\tilde{h}_k(x_1)\tilde{h}_l(y_n) \\
& + \sum_{n=2}^{N} w_{N+1} w_n \tilde{h}_i(x_{N+1})\tilde{h}_j(y_n)\tilde{h}_k(x_{N+1})\tilde{h}_l(y_n) \\
& + w_1 w_1 \tilde{h}_i(x_1)\tilde{h}_j(y_1)\tilde{h}_k(x_1)\tilde{h}_l(y_1) \\
& + w_1 w_{N+1} \tilde{h}_i(x_1)\tilde{h}_j(y_{N+1})\tilde{h}_k(x_1)\tilde{h}_l(y_{N+1}) \\
& + w_{N+1} w_1 \tilde{h}_i(x_{N+1})\tilde{h}_j(y_1)\tilde{h}_k(x_{N+1})\tilde{h}_l(y_1) \\
& + w_{N+1} w_{N+1} \tilde{h}_i(x_{N+1})\tilde{h}_j(y_)N+1\tilde{h}_k(x_{N+1})\tilde{h}_l(y_{N+1})
\end{aligned}$$

which reduces to

$$\begin{aligned}
Q_{IJ} = & \; w_k w_l \delta_{ik}\delta_{jl} \\
& + w_k w_1 \delta_{ik}\tilde{h}_j(y_1)\tilde{h}_l(y_1) \\
& + w_k w_{N+1} \delta_{ik}\tilde{h}_j(y_{N+1})\tilde{h}_l(y_{N+1}) \\
& + w_1 w_l \delta_{jl}\tilde{h}_i(x_1)\tilde{h}_k(x_1) \\
& + w_{N+1} w_l \delta_{jl}\tilde{h}_i(x_{N+1})\tilde{h}_k(x_{N+1}) \\
& + w_1 w_1 \tilde{h}_i(x_1)\tilde{h}_j(y_1)\tilde{h}_k(x_1)\tilde{h}_l(y_1) \\
& + w_1 w_{N+1} \tilde{h}_i(x_1)\tilde{h}_j(y_{N+1})\tilde{h}_k(x_1)\tilde{h}_l(y_{N+1}) \\
& + w_{N+1} w_1 \tilde{h}_i(x_{N+1})\tilde{h}_j(y_1)\tilde{h}_k(x_{N+1})\tilde{h}_l(y_1) \\
& + w_{N+1} w_{N+1} \tilde{h}_i(x_{N+1})\tilde{h}_j(y_{N+1})\tilde{h}_k(x_{N+1})\tilde{h}_l(y_{N+1})
\end{aligned}$$

Since $K\boldsymbol{e}^{(k)} = \boldsymbol{r}^{(k)}$, in terms of the residual this is

$$\|\boldsymbol{e}^{(k)}\|_P^2 = < EK^{-1}\boldsymbol{r}^{(k)}, K^{-1}\boldsymbol{r}^{(k)} > = \|\boldsymbol{r}^{(k)}\|^2_{K^{-1}EK^{-1}}$$

For the Stokes problem, the relevant matrix is

$$
\begin{aligned}
K^{-1}EK^{-1} &= (KE^{-1}K)^{-1} \\
&= \left[\begin{pmatrix} A & 0 & -C_1^T \\ 0 & A & -C_2^T \\ -C_1 & -C_2 & -\lambda B \end{pmatrix} \begin{pmatrix} A^{-1} & 0 & 0 \\ 0 & A^{-1} & 0 \\ 0 & 0 & Q^{-1} \end{pmatrix} \begin{pmatrix} A & 0 & -C_1^T \\ 0 & A & -C_2^T \\ -C_1 & -C_2 & -\lambda B \end{pmatrix} \right]^{-1} \\
&= \left[\begin{pmatrix} I & 0 & -C_1^T Q^{-1} \\ 0 & I & -C_2^T Q^{-1} \\ -C_1 A^{-1} & -C_2 A^{-1} & -\lambda B Q^{-1} \end{pmatrix} \begin{pmatrix} A & 0 & -C_1^T \\ 0 & A & -C_2^T \\ -C_1 & -C_2 & -\lambda B \end{pmatrix} \right]^{-1} \\
&= \begin{pmatrix} A + C_1^T Q^{-1} C_1 & C_1^T Q^{-1} C_2 & -C_1^T + \lambda C_1^T Q^{-1} B \\ C_2^T Q^{-1} C_1 & A + C_2^T Q^{-1} C_2 & -C_2^T + \lambda C_2^T Q^{-1} B \\ -C_1 + \lambda B Q^{-1} C_1 & -C_2 + \lambda B Q^{-1} C_2 & C_1 A^{-1} C_1^T + C_2 A^{-1} C_2^T + \lambda^2 B Q^{-1} B \end{pmatrix}^{-1}
\end{aligned}
$$

Since the PCG method reduces $\|\boldsymbol{r}^{(k)}\|$, it is apparent that a wise choice of preconditioner would be the positive-definite matrix

$$
P = \begin{pmatrix} A + C_1^T Q^{-1} C_1 & C_1^T Q^{-1} C_2 & -C_1^T + \lambda C_1^T Q^{-1} B \\ C_2^T Q^{-1} C_1 & A + C_2^T Q^{-1} C_2 & -C_2^T + \lambda C_2^T Q^{-1} B \\ -C_1 + \lambda B Q^{-1} C_1 & -C_2 + \lambda B Q^{-1} C_2 & C_1 A^{-1} C_1^T + C_2 A^{-1} C_2^T + \lambda^2 B Q^{-1} B \end{pmatrix}
$$

For uniformly stabilized approximation ($\lambda = 0$), this has the form

$$P = \begin{pmatrix} A + C_1^T Q^{-1} C_1 & C_1^T Q^{-1} C_2 & -C_1^T \\ C_2^T Q^{-1} C_1 & A + C_2^T Q^{-1} C_2 & -C_2^T \\ -C_1 & -C_2 & C_1 A^{-1} C_1^T + C_2 A^{-1} C_2^T \end{pmatrix}$$

The form of the Galerkin matrix K and the desired norm on the basis of matrix E implies that it is essential that the block structure should be considered during preconditioning. Therefore, we consider the block diagonal preconditioning matrices of the form

$$P = \begin{pmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & T \end{pmatrix} \tag{5.25}$$

where $P_1, P_2, T \in \mathbb{R}^{(N-1)^2 \times (N-1)^2}$ are positive-definite and symmetric.

The PCG convergence speed is dependent on the eigenvalues μ of the generalized eigenvalue problem

$$\begin{pmatrix} A & 0 & -C_1^T \\ 0 & A & -C_2^T \\ -C_1 & -C_2 & -\lambda B \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ p \end{pmatrix} = \mu \begin{pmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & T \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ p \end{pmatrix} \tag{5.26}$$

If $P_1 = P_2 = A$, then the generalized eigenvalue problem becomes

$$\begin{pmatrix} (1-\mu)A & 0 & -C_1^T \\ 0 & (1-\mu)A & -C_2^T \\ -C_1 & -C_2 & -\lambda B-\mu T \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ p \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{5.27}$$

or, in component form,

$$\begin{cases} (1-\mu)Au_1 = C_1^T p \\ (1-\mu)Au_2 = C_2^T p \\ C_1u_1 + C_2u_2 + (\lambda B+\mu T)p = 0 \end{cases} \tag{5.28}$$

It is easily observed that if $P_1 = P_2 = A$, then 1 is an eigenvalue of multiplicity being at least $2(N-1)^2 - (N-1)^2$ corresponding to any eigenvector $[u_1^T, u_2^T, 0^T]^T$ with $C_1u_1 + C_2u_2 = 0$. The multiplicity is derived from the size of the right null space of the rectangular matrix $[C_1, C_2]$; therefore, if $[C_1, C_2]$ is of full rank, $(N-1)^2$, then the multiplicity of 1 is exactly $2(N-1)^2 - (N-1)^2 = (N-1)^2$.

In the uniformly stable case ($B = 0$), if also $T = C_1A^{-1}C_1^T + C_2A^{-1}C_2^T$, then the remaining eigenvalues satisfy,

$$\begin{cases} (1-\mu)Au_1 = C_1^T p \\ (1-\mu)Au_2 = C_2^T p \\ C_1u_1 + C_2u_2 = -\mu T p = -\mu(C_1A^{-1}C_1^T + C_2A^{-1}C_2^T)p \end{cases} \tag{5.29}$$

or

$$\begin{cases} (1-\mu)C_1A^{-1}Au_1 = C_1A^{-1}C_1^T p \\ (1-\mu)C_2A^{-1}Au_2 = C_2A^{-1}C_2^T p \\ C_1u_1 + C_2u_2 = -\mu(C_1A^{-1}C_1^T + C_2A^{-1}C_2^T)p \end{cases} \tag{5.30}$$

or

$$\begin{cases} (1-\mu)C_1u_1 = C_1A^{-1}C_1^T p \\ (1-\mu)C_2u_2 = C_2A^{-1}C_2^T p \\ C_1u_1 + C_2u_2 = -\mu(C_1A^{-1}C_1^T + C_2A^{-1}C_2^T)p \end{cases} \tag{5.31}$$

Finally, multiplying the last equation in (5.31) by $(1-\mu)$ and using the other two equation to eliminate u_1 and u_2.

$$(1-\mu)(C_1u_1+C_2u_2) = (C_1A^{-1}C_1^T+C_2A^{-1}C_2^T)p = -\mu(1-\mu)(C_1A^{-1}C_1^T+C_2A^{-1}C_2^T)p$$

This reduces to

$$(\mu^2-\mu-1)(C_1A^{-1}C_1^T+C_2A^{-1}C_2^T)p = 0$$

Therefore, since in this case, the assumed inf-sup stability guarantees that $(C_1A^{-1}C_1^T + C_2A^{-1}C_2^T)$ is positive-definite, we deduce that $\mu = \frac{1}{2} \pm \frac{\sqrt{5}}{2}$ are the remaining eigenvalues, each with multiplicity $(N-1)^2$. From the convergence perspective of PCG, this situation is ideal because the preconditioned matrix has only three distinct eigenvalues. A cubic polynomial having these three roots means that PCG will terminate with the precise solution subsequent to three iterations, irrespective of the size of the discrete problem where the arithmetic is exact, meaning that there is no truncation error. This situation is ideal because the preconditioning operation with (5.25) needs the action of the inverses of A and also of the Schur complement $(C_1A^{-1}C_1^T + C_2A^{-1}C_2^T)$. These three iterations need three such computations. The operation involving the Schur complement is entirely impractical because this is a full matrix. Nevertheless, this special

selection of P implies what is really required, namely, an appropriately selected P_1 and P_2 which approximate A and a suitable P that approximates of the Schur complement $(C_1A^{-1}C_1^T + C_2A^{-1}C_2^T)$.

In order to verify that $\int pd\Omega \approx 0$, we approximated $\int pd\Omega$ using the L^2-norm of the vector Tp in (5.26) for different values of N for a range of values of λ.

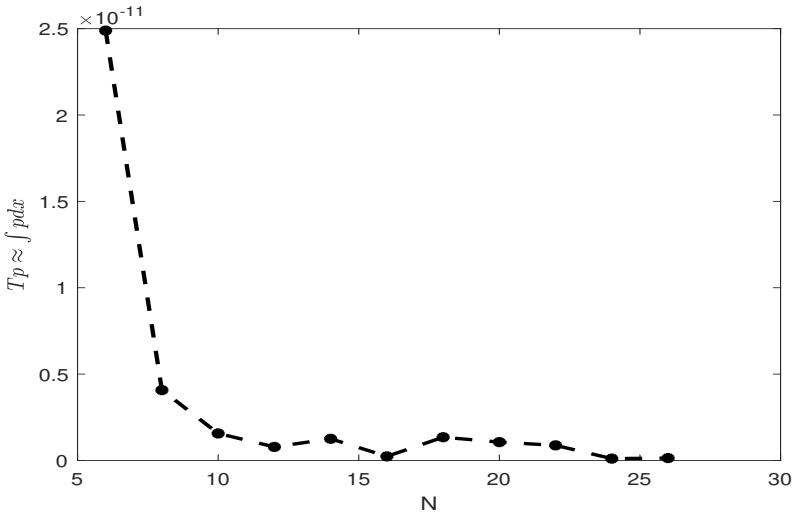


Figure 5.1: Mesh convergence of the approximation of $\int pd\Omega$ for $\lambda = 0.01$.

We can see that $\int p\, d\Omega = O(10^{-12})$ with respect to N for different values of λ. Hence, we can effectively neglect the term Tp in (5.29). Then for $\mu \neq 1$, we have

$$\begin{cases} (1-\mu)Au_1 = C_1^T p \\ (1-\mu)Au_2 = C_2^T p \\ C_1u_1 + C_2u_2 + \lambda Bp = 0 \end{cases} \tag{5.32}$$

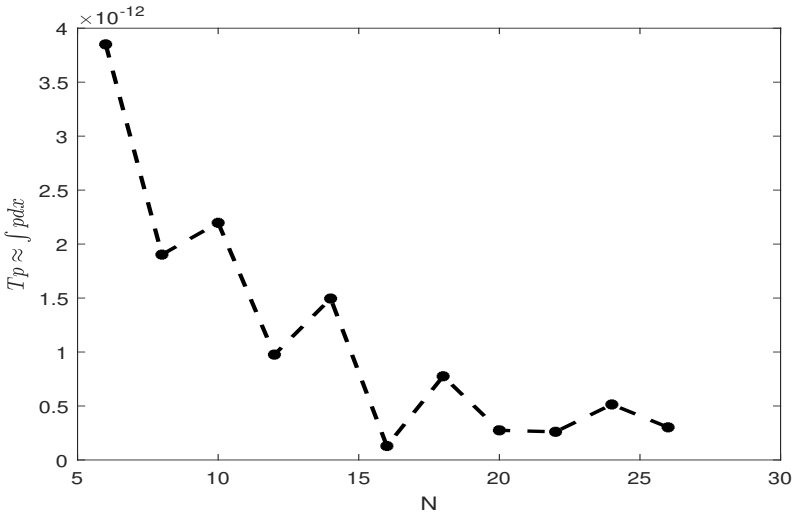


Figure 5.2: Mesh convergence of the approximation of $\int pd\Omega$ for $\lambda = 0.017291812$.

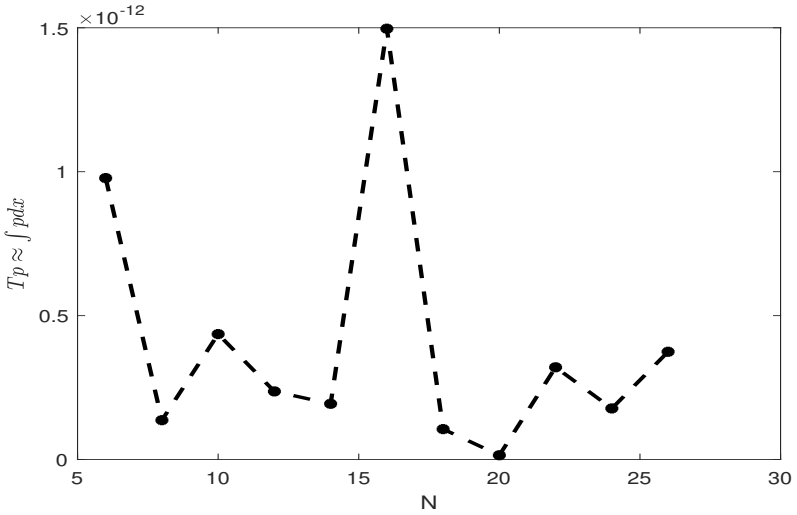


Figure 5.3: Mesh convergence of the approximation of $\int pd\Omega$ for $\lambda = 0.1$.

which, on eliminating u_1 and u_2 from the first two equations becomes

$$\begin{cases} u_1 = \dfrac{1}{(1-\mu)} A^{-1} C_1^T p \\ \\ u_2 = \dfrac{1}{(1-\mu)} A^{-1} C_2^T p \\ \\ \Big(C_1 A^{-1} C_1^T + C_2 A^{-1} C_2^T + \lambda (1-\mu) B \Big) p = 0 \end{cases} \tag{5.33}$$

In our case the Stokes system is

$$\begin{pmatrix} A & 0 & -C_1^T \\ 0 & A & -C_2^T \\ -C_1 & -C_2 & -\lambda B \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ p \end{pmatrix} = \begin{pmatrix} F_1 \\ F_2 \\ 0 \end{pmatrix} \tag{5.34}$$

and the preconditionner is chosen to be

$$P = \begin{pmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & T \end{pmatrix} \tag{5.35}$$

where A is the stiffness matrix (calculated either by SEM or FEM) and $T = diag(Q)$ where Q is the SEM pressure mass matrix.

5.7 Numerical simulations

For $\mu(x,y) = 1$, we examine the solution to a Stokes problem for which the exact solution is given by

$$u(x,y) = \begin{pmatrix} \sin(\pi x)\cos(\pi y) \\ \\ -\cos(\pi x)\sin(\pi y) \end{pmatrix}, \text{ and } p(x,y) = \sin(\pi x)\sin(\pi y).$$

The exact solution automatically satisfies the Stokes equations and

$$\int_{-1}^{1}\int_{-1}^{1} p(x,y)dxdy = 0.$$

The source term in the momentum equation is given by

$$f(x,y) = \begin{pmatrix} 2\pi^2 \sin(\pi x)\cos(\pi y) + \pi\cos(\pi x)\sin(\pi y) \\ \\ -2\pi^2 \cos(\pi x)\sin(\pi y) + \pi\sin(\pi x)\cos(\pi y) \end{pmatrix}$$

In this section, we presuppose every spectral element to be rectangular. The number of spectral elements in each of the Cartesian coordinate directions is defined by the ordered pair (E_x, E_y). This results in the number of spectral elements, N_e equals $E_x \times E_y$.

We calculated the L^2-norm of both pressure and velocity with regard to the spectral discretisation parameters N, which are plotted here on a log-log scale. We observed an exponential convergence of the L^2-norm to zero for both variables (velocity and pressure) (see Figures 5.4 and 5.5 for $(E_x, E_y) = (2,3)$ and $(E_x, E_y) = (1,3)$, respectively).

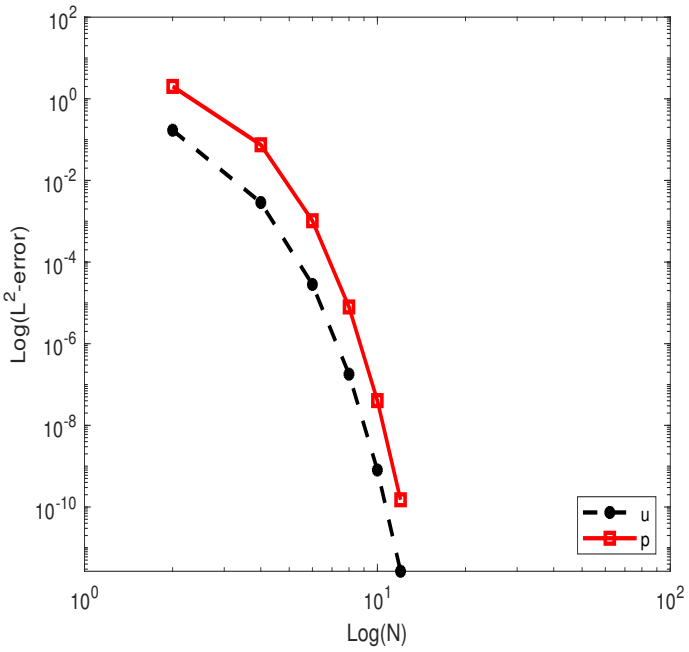


Figure 5.4: L^2-norm of the error with respect to N for $(E_x, E_y) = (2,3)$.

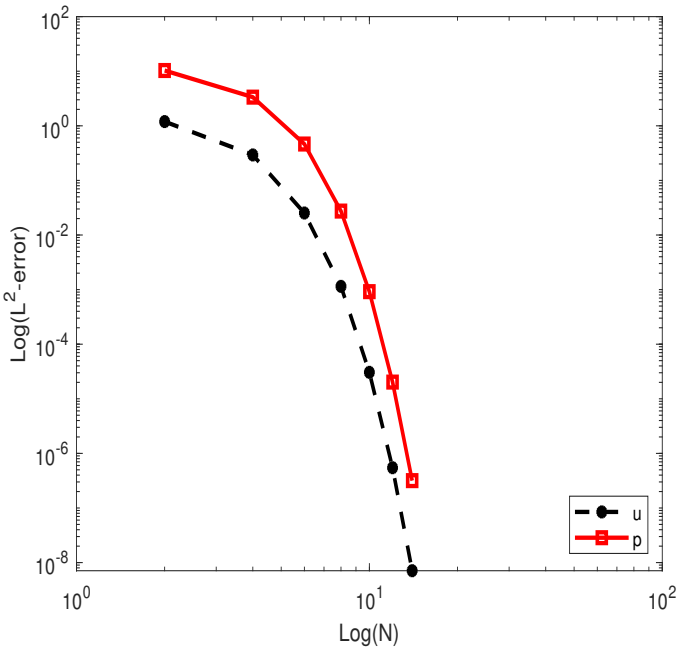


Figure 5.5: L^2-norm of the error with respect to N for $(E_x, E_y) = (1,3)$.

We examine the dependence of the condition number of the operator $P^{-1}K$ over a range of numerical and physical parameters. The condition numbers of $P^{-1}K$ is indicated by κ. Furthermore, we consider the impact of the spectral discretisation parameters N and the number of elements $N_e = E_x \times E_y$, the aspect ratio of the geometry α, and also the stabilisation parameter λ on this condition number.

5.7.1 Effect of the integral weighting factor, λ

The parameter λ, which multiplies the domain pressure integral is an arbitrary weighting factor in the pressure matrix. When $\lambda = 0$, the pressure level may be selected arbitrarily. However, when λ is large then the matrix λB will be dominant, which is also an undesirable situation. Consequently, it is interesting to study the impact of λ on the condition number κ.

The dependence of κ on λ is given for a single spectral element for the domain $[-1,1] \times [-1,1]$. The results for the three different values of N are given; namely,

$N = 8, 10, 12$. Figs. 5.6, 5.7 and 5.8 depict a plateau at the minimum value of κ. It is only values which are outside of this interval that affect the condition number κ.
The width and position of the plateau denote the interval where λ has no impact on κ.

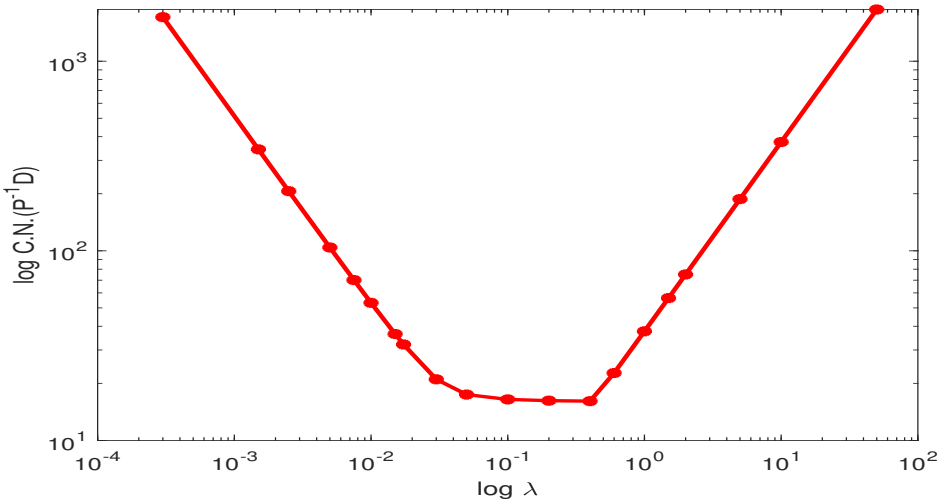


Figure 5.6: Condition number of $P^{-1}K$ when using SEM preconditioner with respect to λ for $N = 8$.

As can be observed in Fig. 5.6, the condition number κ, for $N = 8$, admits its minimum value (≈ 1.08 in log scale) on an interval for which $\lambda < 1$ and not close to zero.

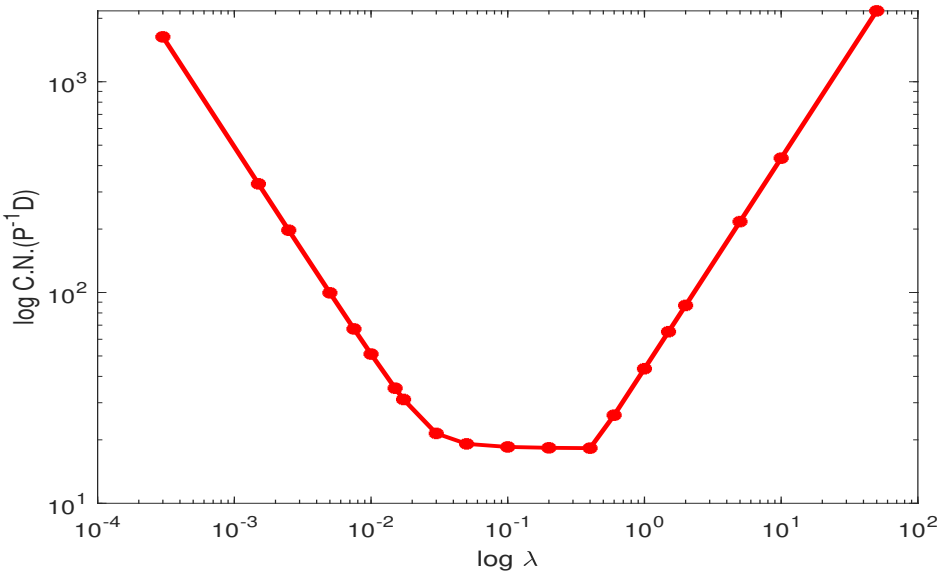


Figure 5.7: Condition number of $P^{-1}K$ when using SEM precondioner with respect to λ for $N = 10$.

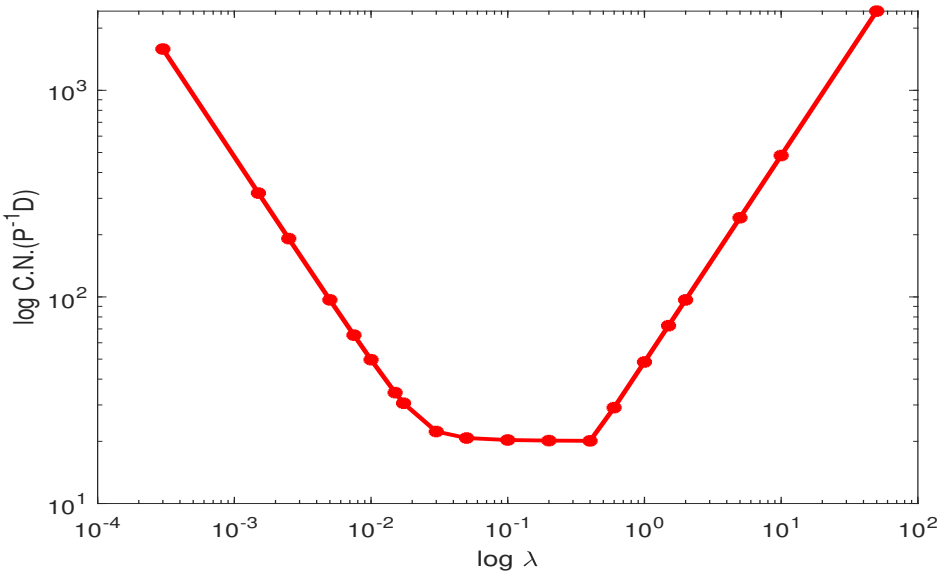


Figure 5.8: Condition number of $P^{-1}K$ when using SEM precondioner with respect to λ for $N = 12$.

Similarly to Figs. 5.7 and 5.8, the condition number κ, for $N = 10, 12$, admits almost the same minimum value for almost the same range of value of λ.

The problem is then well-conditioned if we choose our parameter λ in this interval ($\approx [0.05, 0.4]$). Values of λ in this interval provide the smallest condition number for the preconditioned system.

5.7.2 Effect of the spectral discretization parameter N

Condition Number

Furthermore, the discretisation parameter N has an important impact on the condition number of the operator $P^{-1}K$. In this part, only a single spectral element $(E_x, E_y) = (1, 1)$ is considered. The log-log plot of the condition number against N effectively produces straight lines. The dependence of κ upon N varies according to different values λ because the slope of the straight lines depends on λ.

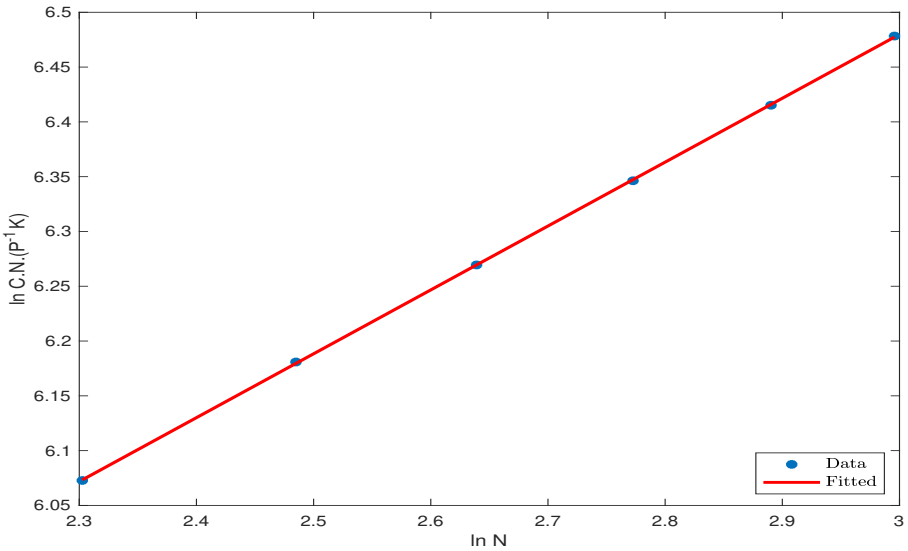


Figure 5.9: Condition number of $P^{-1}K$ when using the SEM preconditioner with respect to N on a log-log scale for $\lambda = 10$. Linear dependence of the condition number of $P^{-1}K$ with respect to N with a slope equal to 0.5939

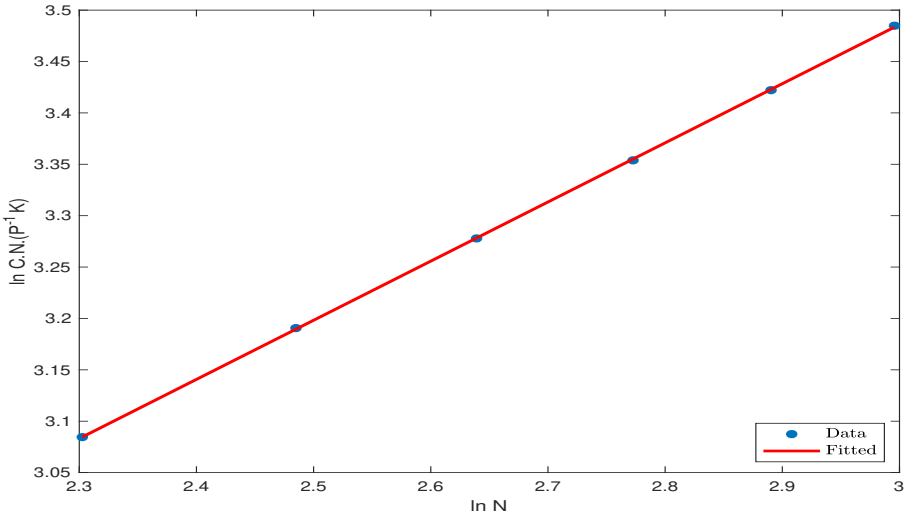


Figure 5.10: Condition number of $P^{-1}K$ when using the SEM preconditioner with respect to N on a log-log scale for $\lambda = 0.5$. Linear dependence of the condition number of $P^{-1}K$ with respect to N with a slope equal to 0.5878

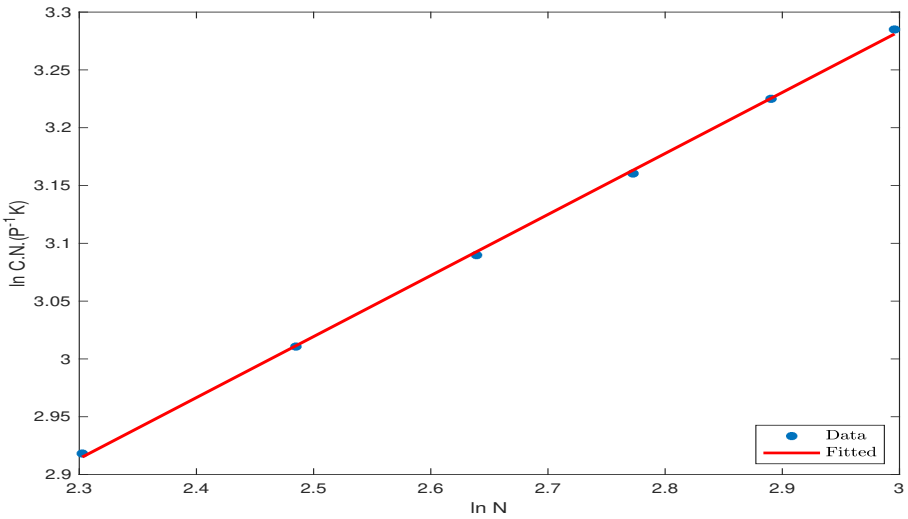


Figure 5.11: Condition number of $P^{-1}K$ when using the SEM preconditioner with respect to N on a log-log scale for $\lambda = 0.1$. Linear dependence of the condition number of $P^{-1}K$ with respect to N with a slope equal to 0.5499

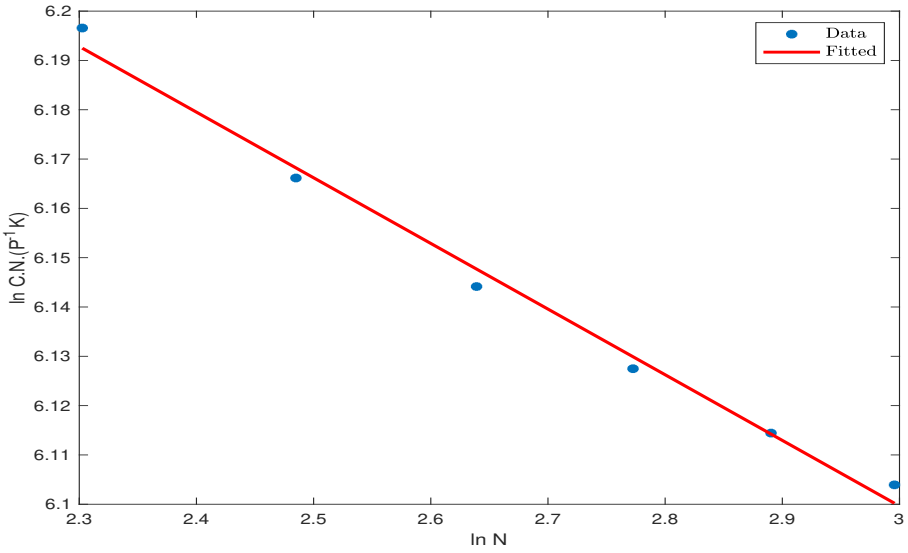


Figure 5.12: Condition number of $P^{-1}K$ when using the SEM preconditioner with respect to N on a log-log scale for $\lambda = 0.001$. Linear dependence of the condition number of $P^{-1}K$ with respect to N with a slope equal to -0.1178

We are aware that, for large values of λ, the condition number κ of $P^{-1}K$ increases with N. (see Figs. 5.9, 5.10, 5.11) . Nevertheless, for small values of λ, it decreases with respect to N (see Fig. 5.12). One possible relationship can take the form $\kappa = CN^{\alpha}$ where C and α are constants and subsequently, in a log-log plot, the relationship becomes $\ln\kappa = \alpha \ln N + \ln C$ where α is the slope of the linear dependence of $\ln\kappa$ with respect to $\ln N$.

Number of Outer PCG Iterations, N_{it}

In this section, results will be provided on the number of outer PCG iterations, N_{it}, which are needed to solve Eq. (5.26) by a termination criterion of a maximum relative difference of 10^{-12} between successive conjugate gradient iterates. This tolerance is sufficiently small to guarantee that it does not have any general impact on the precision of the spectral approximation.

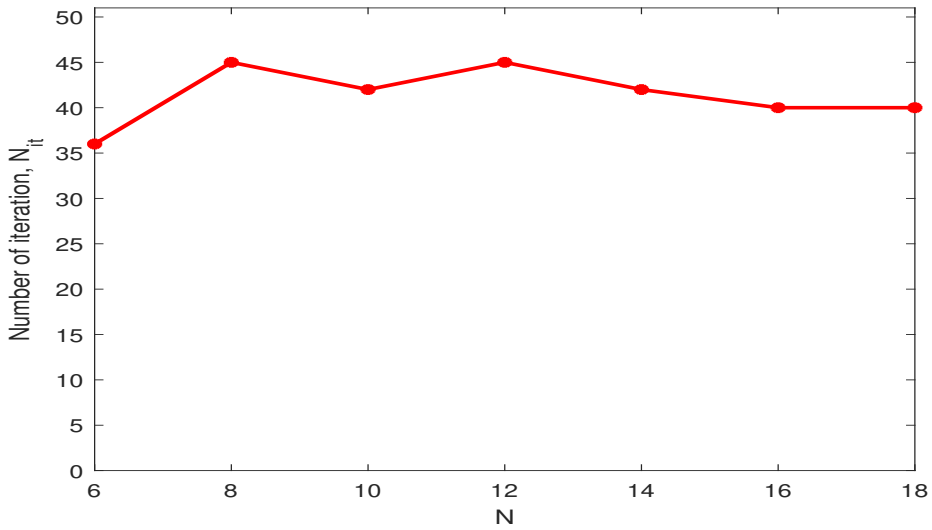


Figure 5.13: Number of iterations for convergence of the PCG method with respect to N for $\lambda = 10$.

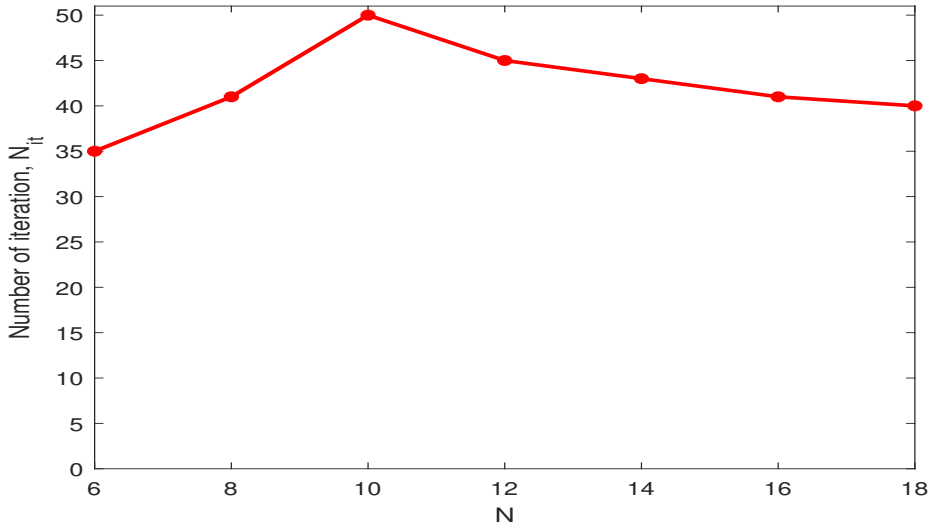


Figure 5.14: Number of iterations for convergence of the PCG method with respect to N for $\lambda = 0.5$.

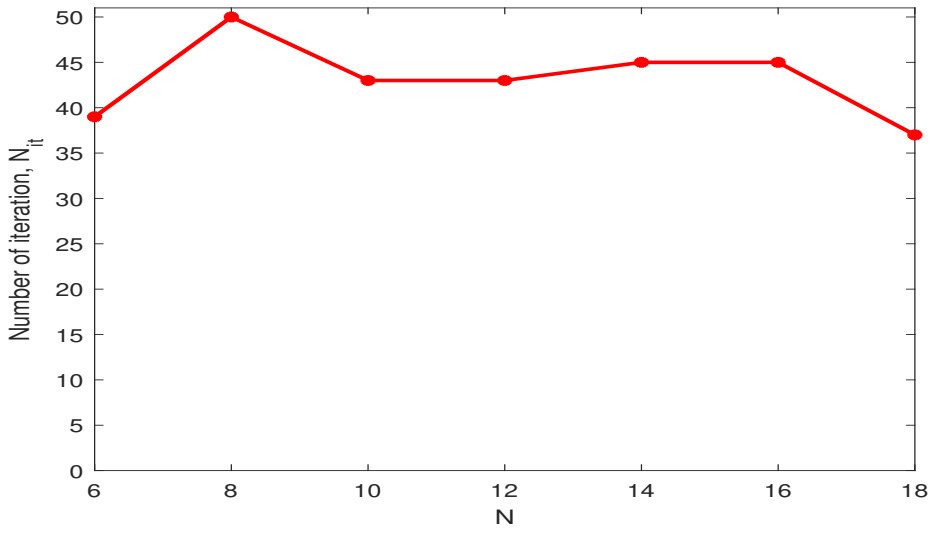


Figure 5.15: Number of iterations for convergence of the PCG method with respect to N for $\lambda = 0.1$.

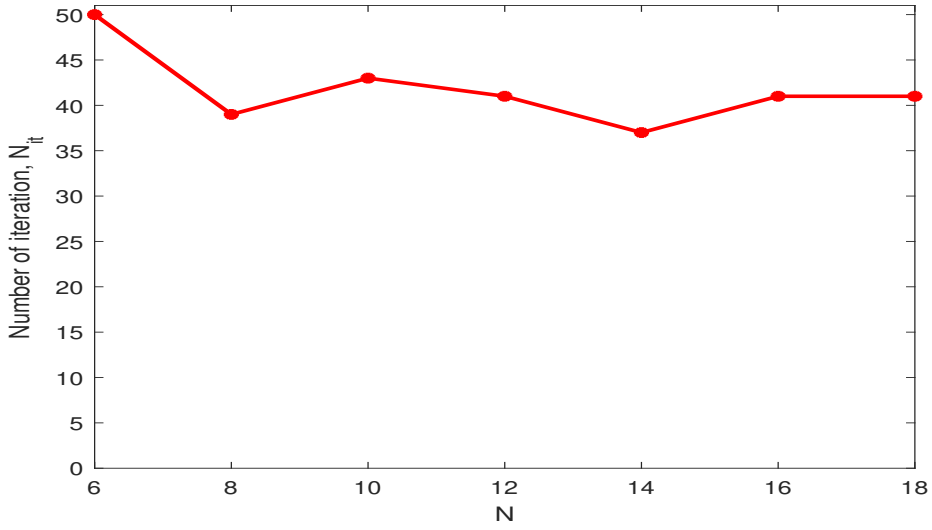


Figure 5.16: Number of iterations for convergence of the PCG method with respect to N for $\lambda = 0.001$.

The number of outer PCG iterations, N_{it}, is almost constant and is independent of the spectral discretization parameter, N, as can be seen in Figs. 5.13, 5.14, 5.15 and 5.16.

We solve the system (5.26) using the SEM with one element, where the matrix A_p used in the preconditioner P is calculated using either FEM or SEM preconditioner. The L^2-norm is used for both variables (velocity and pressure). As can be seen in Tables 5.2 and 5.4, the number of iterations grows slowly when the matrix A_p is calculated using SEM compared to the case when the matrix A_p is calculated using FEM.

	A_p calculated using SEM		A_p calculated using FEM	
N	u-error	p-error	u-error	p-error
4	0.255	1.832	0.255	1.832
6	3.32×10^{-2}	0.396	3.321×10^{-2}	0.396
8	1.4×10^{-3}	0.025	1.403×10^{-3}	0.025
10	3.729×10^{-5}	9.239×10^{-4}	3.729×10^{-5}	9.239×10^{-4}
12	6.671×10^{-7}	2.177×10^{-5}	6.671×10^{-7}	2.177×10^{-5}
14	8.653×10^{-9}	3.582×10^{-7}	8.654×10^{-9}	3.582×10^{-7}
16	8.527×10^{-11}	4.344×10^{-9}	8.536×10^{-11}	4.342×10^{-9}
18	6.603×10^{-13}	4.04×10^{-11}	1.431×10^{-12}	4.919×10^{-11}
20	4.492×10^{-15}	2.952×10^{-13}	2.657×10^{-12}	4.091×10^{-11}

Table 5.1: Resolution of the system (5.26) using the SEM with one element, where the matrix A_p used in the preconditioner P is calculated using SEM (left) or FEM (right). The L^2-norm is used for both variables (velocity and pressure).

	A_p calculated using SEM		A_p calculated using FEM	
N	N_{it}	CPU	N_{it}	CPU
4	21	0.047	24	0.141
6	45	0.125	53	0.031
8	56	0.078	72	0.063
10	54	0.313	88	0.063
12	55	0.500	102	0.234
14	55	1.484	124	1.063
16	57	2.984	147	1.359
18	57	5.500	169	2.828
20	61	9.969	194	5.734

Table 5.2: Resolution of the system (5.26) using the SEM with one element, where the matrix A_p used in the preconditioner P is calculated using both SEM (left) and FEM (right). Number of PCG iterations and CPU time.

Once again, the number of iterations is almost constant for $N \geq 8$ when the matrix A_p is calculated using SEM compared to the case when the matrix A_p is calculated using FEM. If we compare these results to the ones in the case with one element, we see that by increasing the number of elements, the convergence is more rapid (Table 5.3), and the number of iterations increases slowly in the case of SEM (see Table 5.4).

	A_p calculated using SEM		A_p calculated using FEM	
N	u-error	p-error	u-error	p-error
4	3.285×10^{-3}	7.154×10^{-2}	3.285×10^{-3}	7.154×10^{-2}
6	3.75×10^{-5}	1.01×10^{-3}	3.75×10^{-5}	1.01×10^{-3}
8	2.461×10^{-7}	8.706×10^{-6}	2.461×10^{-7}	8.706×10^{-6}
10	1.121×10^{-9}	4.893×10^{-8}	1.121×10^{-9}	4.893×10^{-8}
12	3.732×10^{-12}	1.936×10^{-10}	8.642×10^{-12}	2.371×10^{-10}
14	9.639×10^{-15}	5.723×10^{-13}	1.732×10^{-12}	5.832×10^{-11}
16	6.34×10^{-15}	7.392×10^{-14}	2.936×10^{-12}	1.988×10^{-10}
18	1.597×10^{-14}	6.275×10^{-13}	4.027×10^{-12}	1.118×10^{-10}
20	6.477×10^{-15}	7.197×10^{-14}	6.98×10^{-12}	2.196×10^{-10}

Table 5.3: Resolution of the system (5.26) using the SEM with four elements ($K = 4$), where the matrix A_p used in the preconditioner P is calculated using both SEM (left) and FEM (right).

	A_p calculated using SEM		A_p calculated using FEM	
N	N_{it}	CPU	N_{it}	CPU
4	75	0.063	126	0.031
6	90	0.125	219	0.172
8	85	0.719	271	0.550
10	85	2.125	343	1.828
12	93	3.813	403	6.531
14	89	6.406	480	9.500
16	97	10.219	561	21.375
18	103	16.297	653	46.063
20	113	27.188	765	63.469

Table 5.4: Resolution of the system (5.26) using the SEM with four elements ($K = 4$), where the matrix A_p used in the preconditioner P is calculated using both SEM (left) and FEM (right). Dependence of the number of PCG iterations and CPU time on N.

As can be seen in Fig. 5.17, there exists a critical value of the parameter $\lambda \approx 0.0173$, for which the condition number of $P^{-1}K$ is almost constant with respect to N.

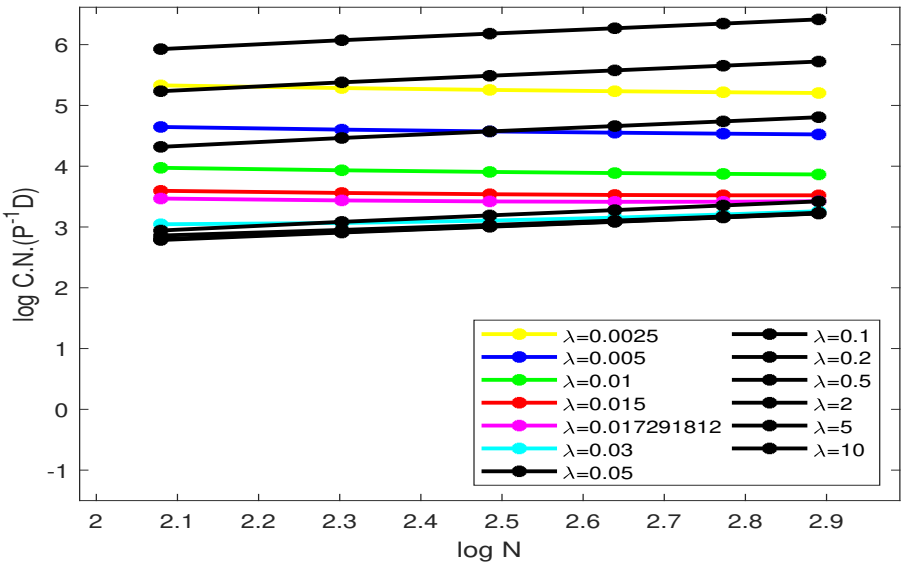


Figure 5.17: log-log plot of the condition number of $P^{-1}K$ when using SEM preconditioner with respect to N for different values of λ.

Next, we calculated the slope of the curve describing the relation between the condition number of $P^{-1}K$ and the discretization parameter, N for different values of λ. It can be seen that for large values of λ, the slope is almost constant which means that the condition number of $P^{-1}K$ is not influenced by the discretization parameter, N.

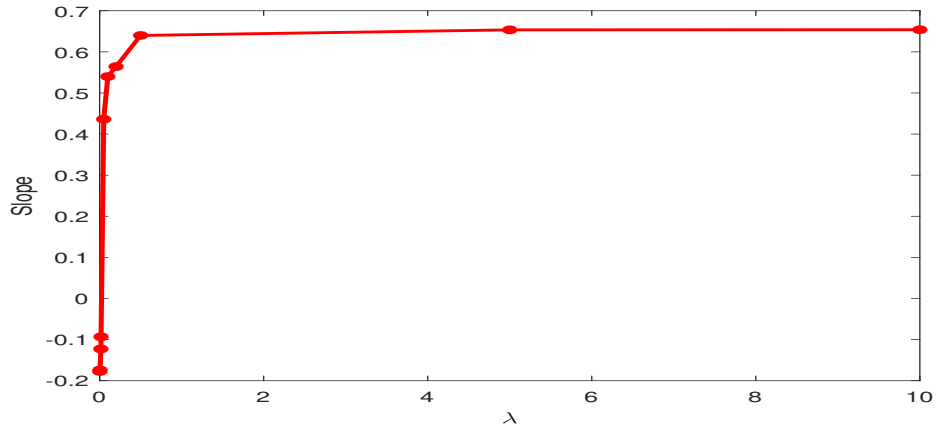


Figure 5.18: Dependence of the slope of the log-log relation between the condition number of $P^{-1}K$ and N on λ.

We calculated the ratio of the largest to the smallest eigenvalue of $P^{-1}K$ with respect to N for different values of λ (see Figs. 5.19, 5.20, 5.21, 5.22).

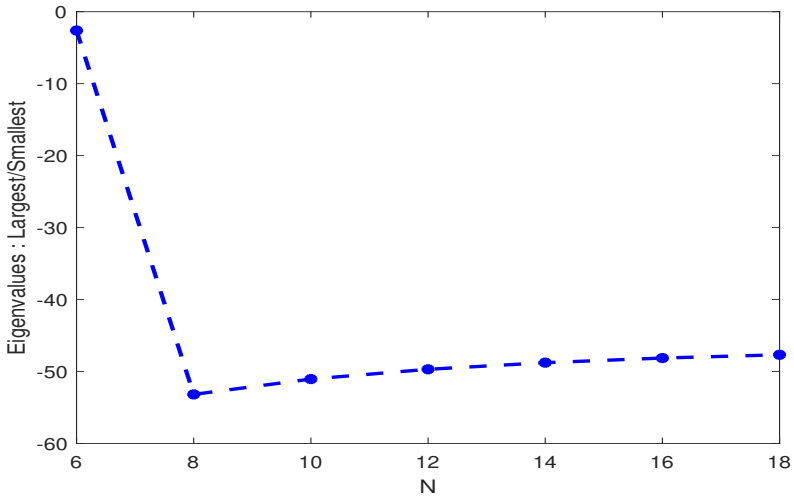


Figure 5.19: The ratio of the largest to the smallest eigenvalue of the preconditioned system (5.26) with respect to N for $\lambda = 0.01$.

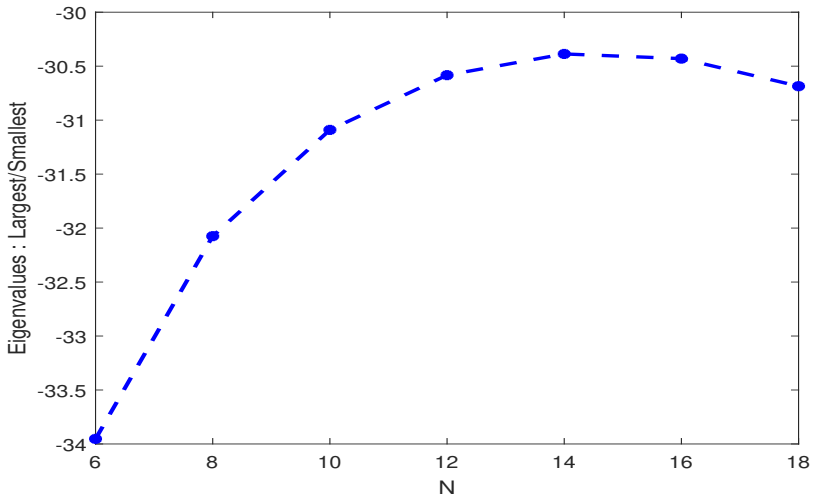


Figure 5.20: The ratio of the largest to the smallest eigenvalue of $P^{-1}K$ with respect to N for the critical value $\lambda = 0.017291812$.

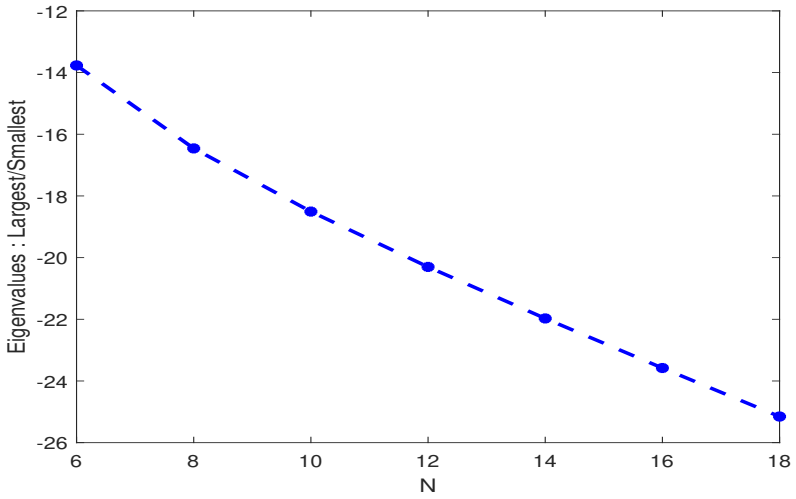


Figure 5.21: The ratio of the largest to the smallest eigenvalue of $P^{-1}K$ with respect to N for $\lambda = 0.1$.

It can be seen that the ratio of the largest to smallest eigenvalue with respect to N is increasing for $\lambda = 0.01$ for $N \geq 8$, almost constant for the critical value $\lambda = 0.017291812$ (note the scale on the vertical axis) and then decreasing for $\lambda = 0.1$.

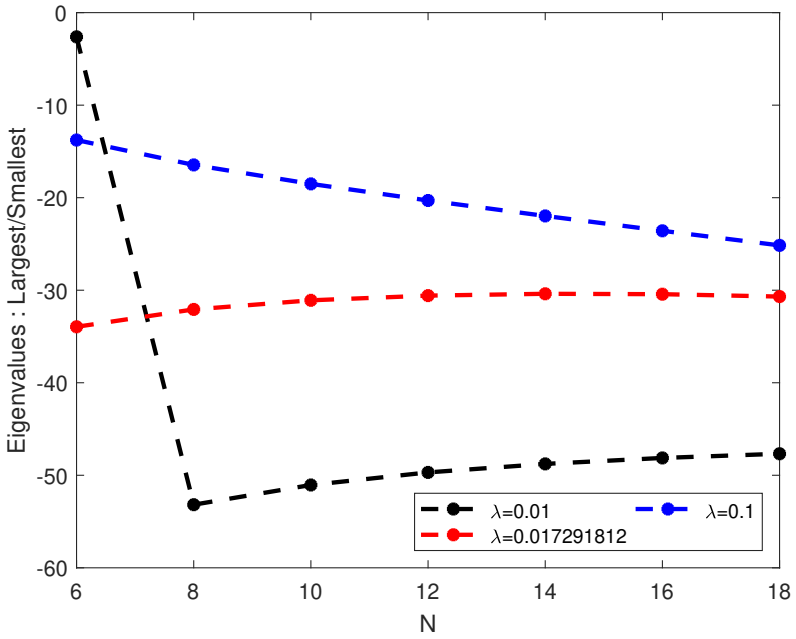


Figure 5.22: The ratio of the largest to the smallest eigenvalue of $P^{-1}K$ with respect to N for different values of λ.

The number of iterations was calculated with respect to discretization parameter, N, for different values of λ. It can be seen that the number of iterations is almost constant when varying the discretization parameter, N (see Tables 5.5-5.6). It is approximately 40 iterations for most cases, as can be seen also in Figures 5.13, 5.14, 5.15 and 5.16.

$N \setminus \lambda$	0.0025	0.005	0.01	0.015	0.0173	0.03	0.05	0.1	0.2	0.5
10	43	50	50	43	43	50	39	43	44	50
12	41	39	41	43	39	41	43	43	50	45
14	39	41	41	43	45	41	39	45	42	43
16	43	41	41	41	41	45	41	45	42	41
18	37	39	41	39	41	39	37	37	39	40
20	45	37	43	37	37	37	37	35	39	40

Table 5.5: Number of iterations for convergence of the PCG algorithm with respect to N for different values of λ.

$N \setminus \lambda$	2	5	10
10	45	44	42
12	44	41	45
14	45	45	42
16	47	43	40
18	44	38	40
20	40	38	37

Table 5.6: Number of iterations for convergence of the PCG algorithm with respect to N for different values of λ.

5.8 Stokes flow in contraction geometries and unbounded domains

In this part, we resolve Stokes flow numerically in a contraction which means an infinite channel with an abruptly changing width. Two approximations are involved in the numerical solution of these problems. First, the solution must be approximated by some representation. The unknowns in such a representation are determined by satisfying the differential equation and boundary conditions, in some sense. Secondly, it is possible to approximate the unbounded domain by a finite domain. We have a special interest in the impact on the solution of the truncation of the unbounded domain. It is well known that from the study of non-Newtonian fluids by means of tortuous geometries, the flow behaviour depends on entry and exit lengths [27]. It is necessary to increase the entry length as Weissenberg number or the elasticity parameter also increases in order to impose a fully-developed velocity profile on the entry and exit sections. The resolution of these problems numerically utilises domain truncation and also enforces an artificial boundary condition at a considerable but finite distance. Subsequently, the problem is solved in the finite domain.

The Stokes problem is

$$\begin{cases} -\nabla\cdot(\mu\nabla\boldsymbol{u})+\nabla p &= \boldsymbol{f} & \text{on} \quad \Omega \\ -\operatorname{div}\boldsymbol{u} &= \lambda\displaystyle\int_\Omega p\,d\Omega & \text{on} \quad \Omega \end{cases} \tag{5.36}$$

where $\boldsymbol{u}=(u_1,u_2)^T$ can be interpreted as the velocity field of an incompressible fluid motion, p is the associated pressure and the function μ is the viscosity of the fluid. We enforce some artificial Dirichlet boundary condition at the input (inflow, u_{in}) and the output (outflow, u_{out}) of the truncated domain. The flow rate is constant; therefore we require that $\int_{\Gamma_{in}} u_{in} = \int_{\Gamma_{out}} u_{out}$.

Again we assume that all spectral elements are rectangular. The domain was subdivided into six sub-domains (six elements). We calculated the L^2-norm of the error

of both velocity and pressure with respect to the spectral discretization parameter, N, plotted here on a log-log scale. An exponential convergence of the L^2-norm of the error for both variables (velocity and pressure) is again observed.

5.8.1 Mixed boundary condition

It is evident that the x-axis represents an axis of symmetry for the problem. We solve Eq. (5.36) subject to the following boundary conditions on $\boldsymbol{u}(x,y) = \big(u_1(x,y), u_2(x,y)\big)$:

$$\begin{cases} u_1(-a,y) = u_{in}(y), u_2(-a,y) = 0, & \forall -1 \leq y \leq 1 \\ u_1(b,y) = u_{out}(y), u_2(b,y) = 0, & \forall\, 0 \leq y \leq 1 \\ u_1(0,y) = u_2(0,y) = 0, & \forall -1 \leq y \leq 0 \\ u_1(x,1) = \dfrac{\partial u_2}{\partial n}(x,1) = 0, & \forall -a \leq x \leq b \\ u_1(x,-1) = u_2(x,-1) = 0, & \forall -a \leq x \leq 0 \\ u_1(x,0) = u_2(x,0) = 0, & \forall\, 0 \leq x \leq b \end{cases} \tag{5.37}$$

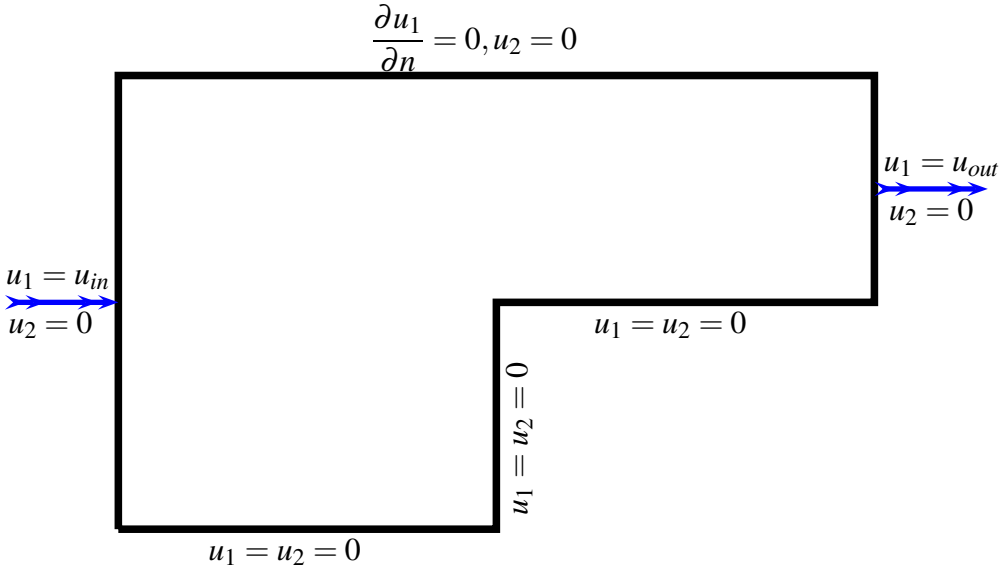


Figure 5.23: Stokes flow in contraction domain.

To ensure conservation of mass we must have

$$\int_{-1}^{1} u_{in}(-a,b)dy = \int_{0}^{1} u_{out}(b,y)dy$$

We separate the area of interest into two semi-infinite rectangular sub-regions, in each of which we represent the solution to Eq. (5.36) in a truncated series format which involves orthogonal polynomials. In these expansions, the unknown coefficients are established by satisfying the boundary conditions (5.37) and also those of the differential equation (5.36) at various selected points in the area (collocation points) and also the solutions in both sub-regions which are matched by enforcing continuity conditions over the interface.

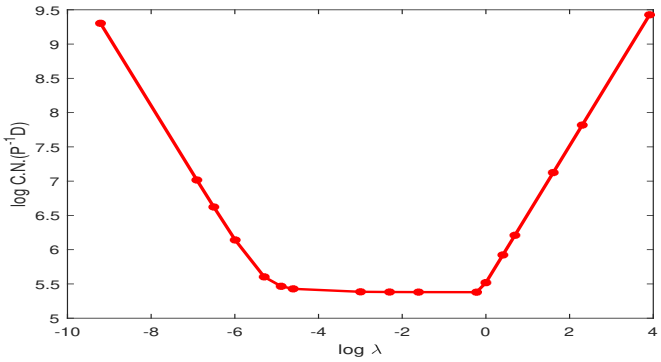


Figure 5.24: Condition number of $P^{-1}K$ when using SEM precondioner with respect to λ for $N = 6$.

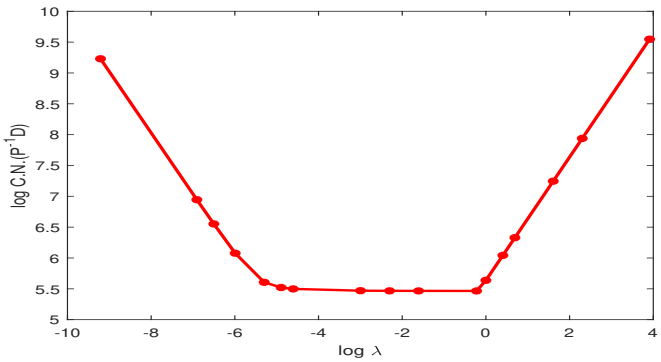


Figure 5.25: Condition number of $P^{-1}K$ when using SEM precondioner with respect to λ for $N = 8$.

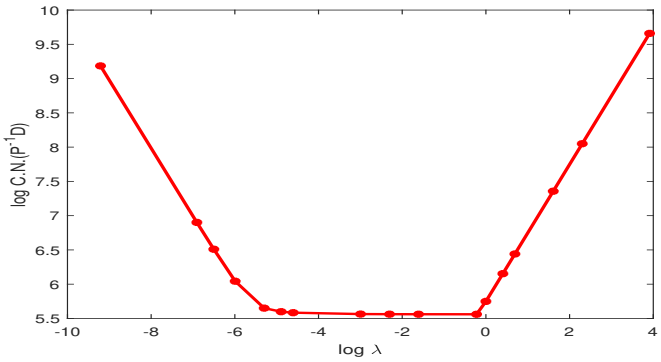


Figure 5.26: Condition number of $P^{-1}K$ when using SEM precondioner with respect to λ for $N = 10$.

As we can see from Figs. 5.24-5.26, the condition number κ admits its minimum value (≈ 5.5 on log scale) on an interval for which $\lambda < 1$ and not close to zero. At the minimum value of κ, a plateau is shown by Figs. 5.24, 5.25 and 5.26 in which the only values of λ that have an impact on the condition number κ are outside this interval. The plateau's position and the width show the interval within which λ has no discernible impact on κ. The log-log plot of the condition number against N produces lines which are almost straight in which their slopes are dependent on λ . It should be noted that for large values of λ, κ increases with N (see Fig. 5.27); nevertheless, for small values of λ, it decreases with respect to N(see Fig. 5.29).

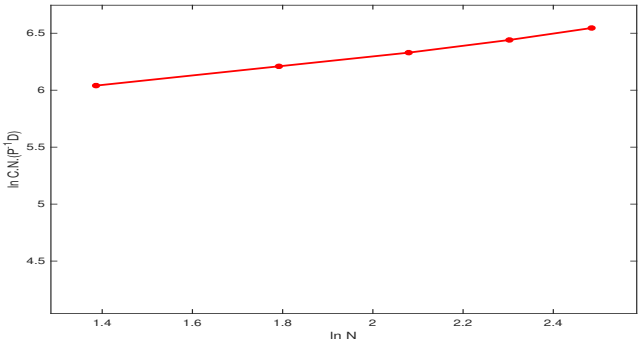


Figure 5.27: Condition number of $P^{-1}K$ when using SEM precondioner with respect to N on a log-log scale for $\lambda = 2$.

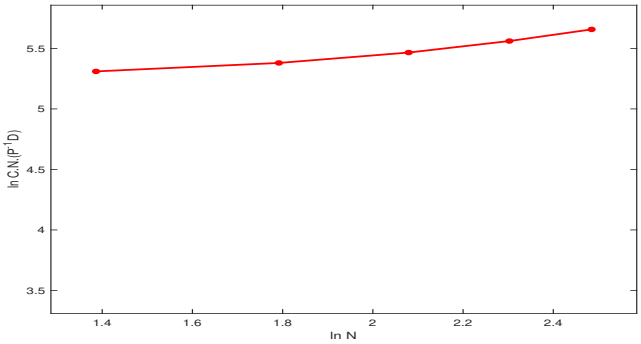


Figure 5.28: Condition number of $P^{-1}K$ when using SEM precondioner with respect to N on a log-log scale for $\lambda = 0.5$.

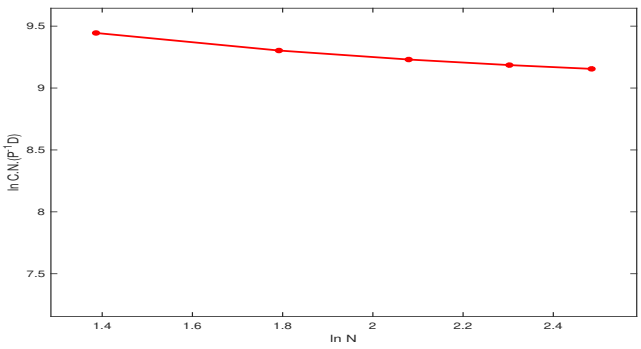


Figure 5.29: Condition number of $P^{-1}K$ when using SEM precondioner with respect to N on a log-log scale for $\lambda = 0.0001$.

We calculated the ratio of the largest to the smallest eigenvalue with respect to N for different values of λ (see Figs. 5.30, 5.32, 5.31). It can be seen that the ratio of the largest to smallest eigenvalue with respect to N is increasing for $\lambda = 0.00001$, almost constant for the critical value $\lambda = 0.5$ and then increasing for $\lambda = 2$.

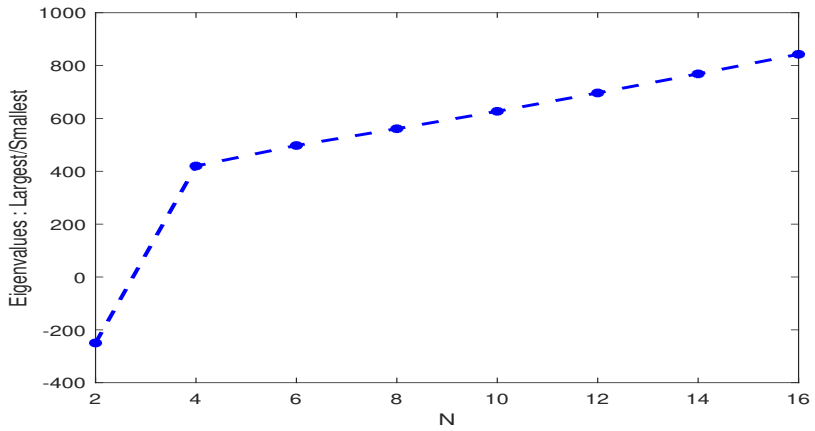


Figure 5.30: The ratio of the largest to smallest eigenvalue of $P^{-1}K$ with respect to N for $\lambda = 2$.

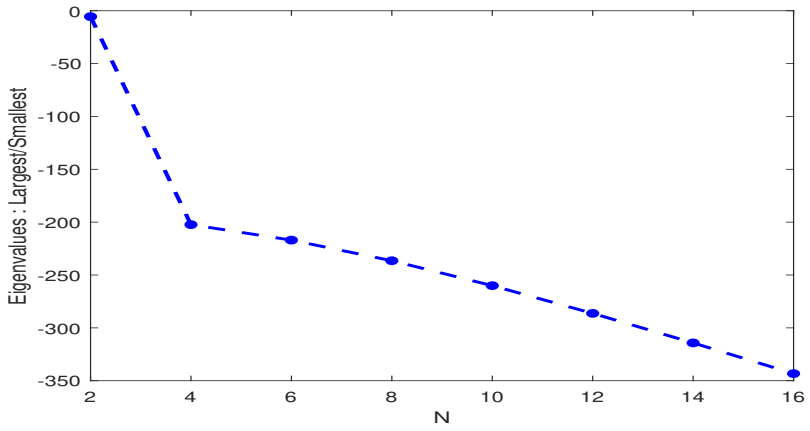


Figure 5.31: The ratio of the largest to smallest eigenvalue of $P^{-1}K$ with respect to N for the critical value $\lambda = 0.5$.

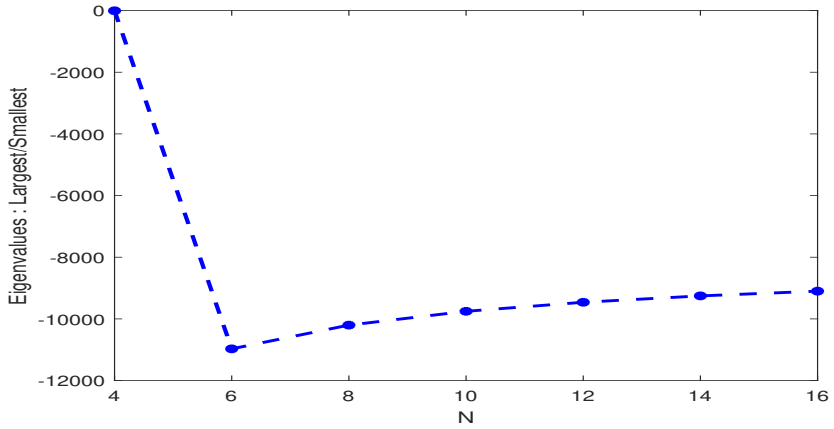


Figure 5.32: The ratio of the largest to smallest eigenvalue of $P^{-1}K$ with respect to N for $\lambda = 0.0001$.

The number of outer PCG iterations, N_{it}, is almost constant and increases slowly with respect to the spectral discretization parameter, N, as can be seen in Figs. 5.33 , 5.34 and 5.35.

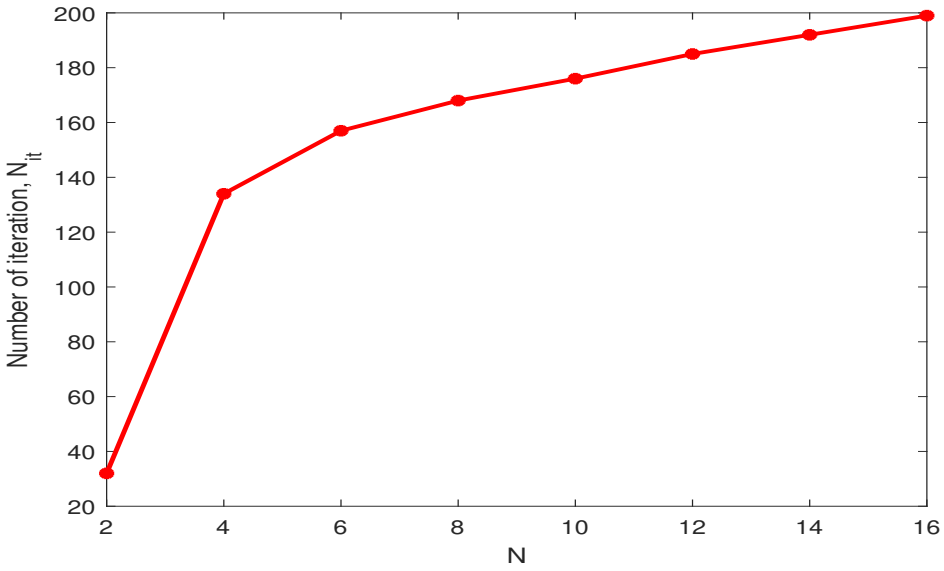


Figure 5.33: Number of iterations for the PCG method with respect to N for $\lambda = 2$.

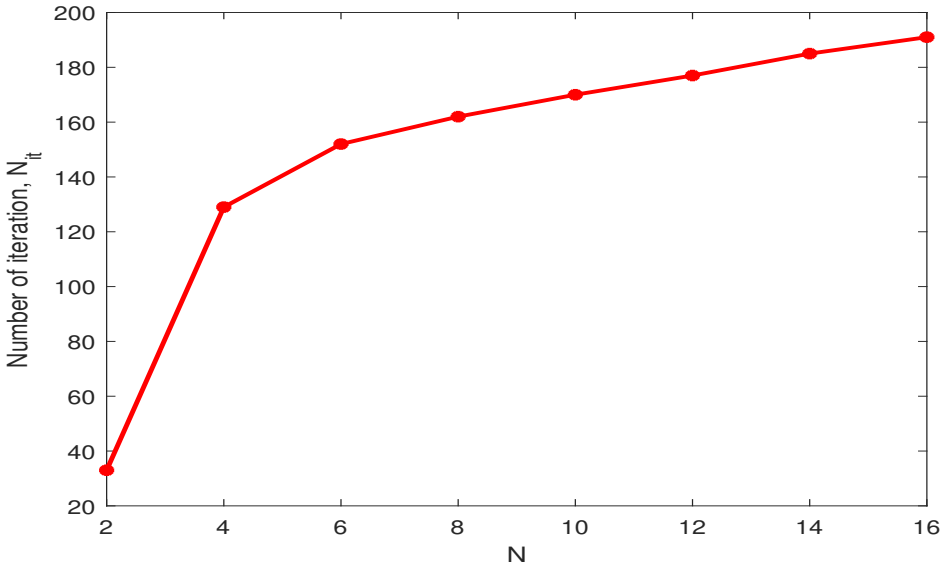


Figure 5.34: Number of iterations for the PCG method with respect to N for $\lambda = 0.5$.

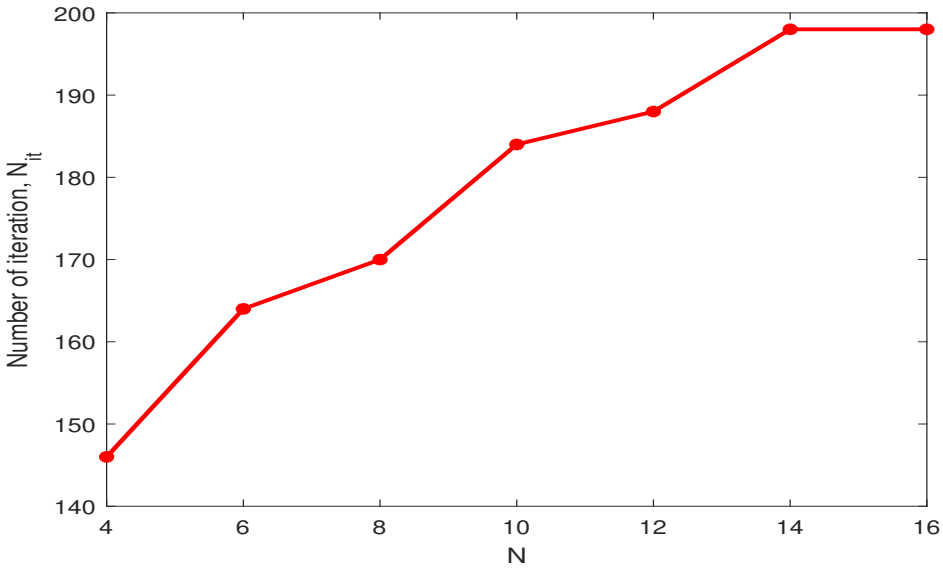


Figure 5.35: Number of iterations for the PCG method with respect to N for $\lambda = 0.0001$.

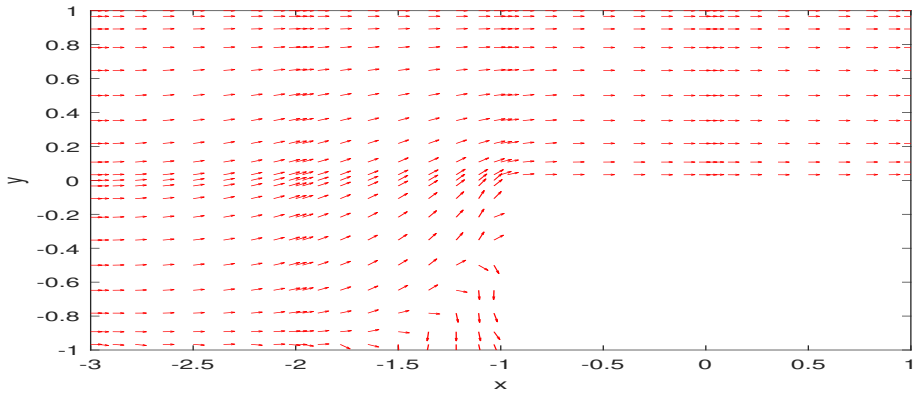
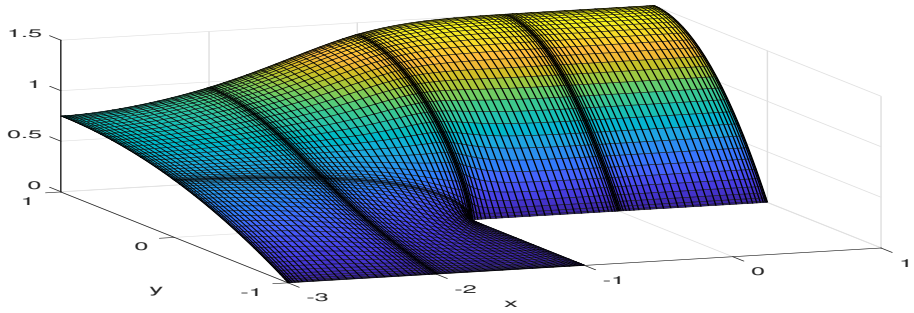
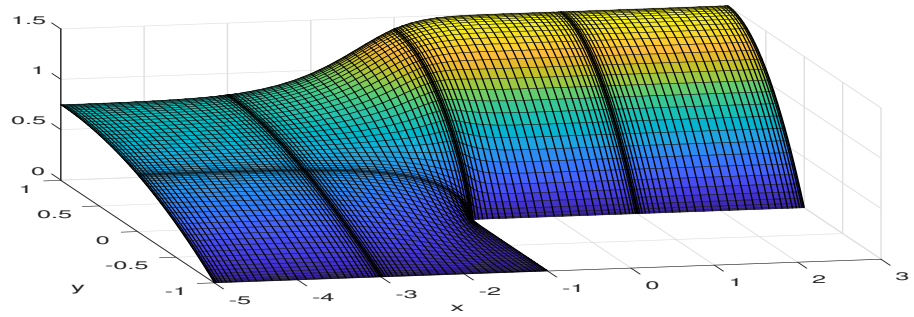


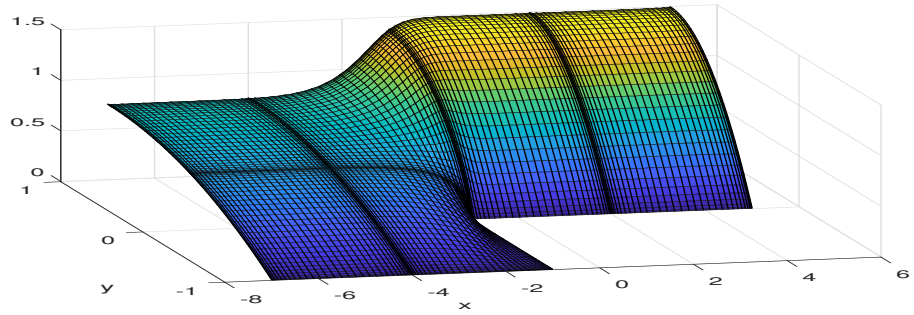
Figure 5.36: Velocity vector for $N = 10$ on a truncated domain.



(a) $\Omega = [-3,1] \times [-1,1]$

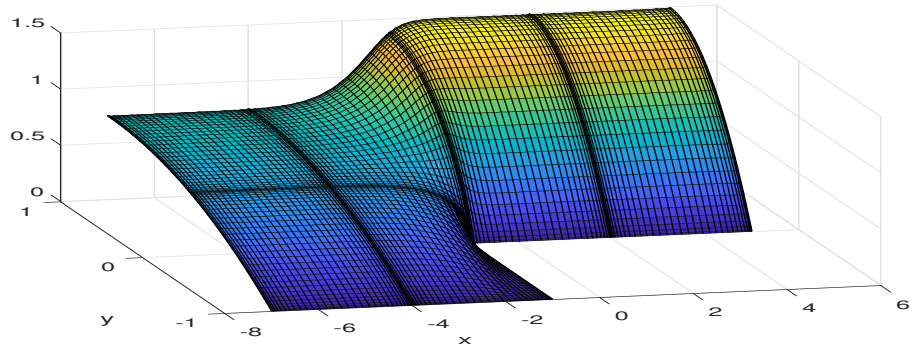


(b) $\Omega = [-5,3] \times [-1,1]$

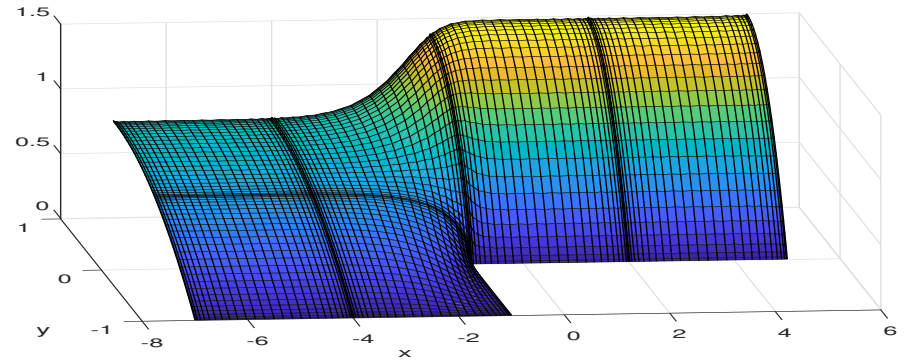


(c) $\Omega = [-7,5] \times [-1,1]$

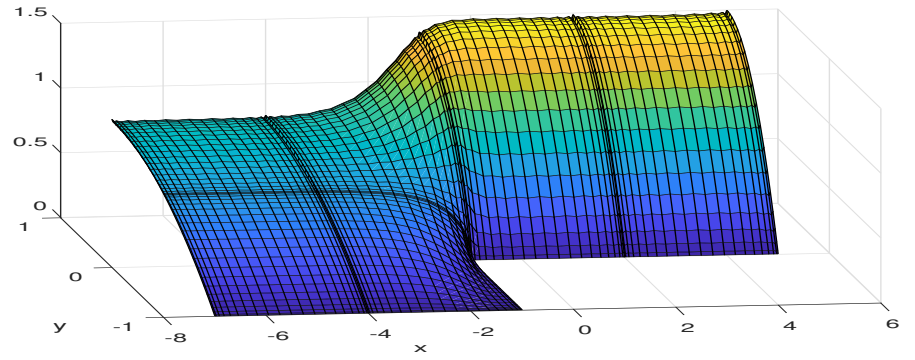
Figure 5.37: Approximated horizontal velocity component for $N = 30$ on different truncated domains.



(a) $N = 30$

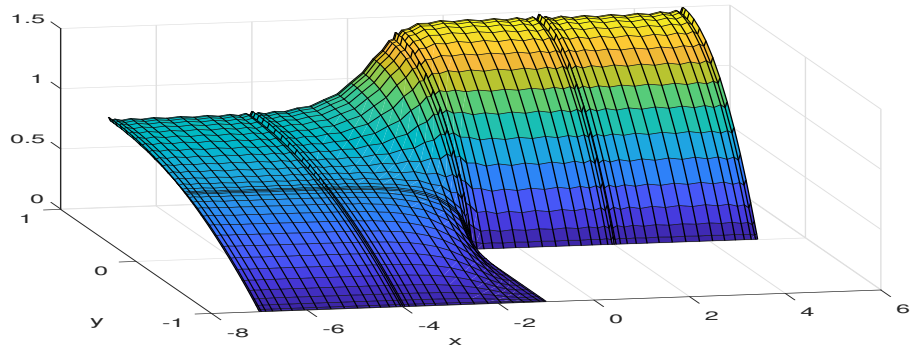


(b) $N = 26$

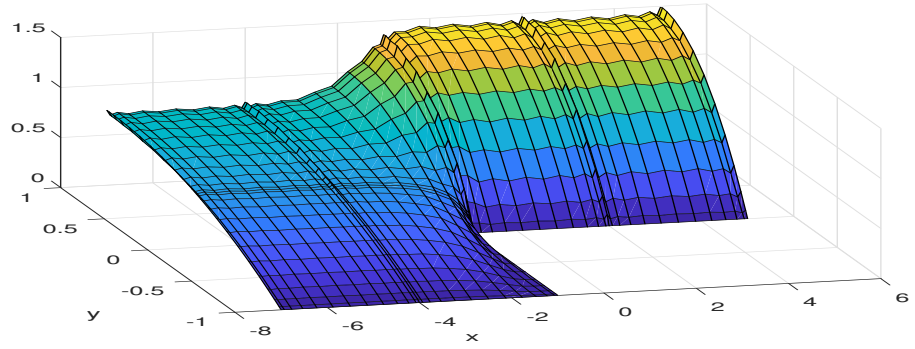


(c) $N = 20$

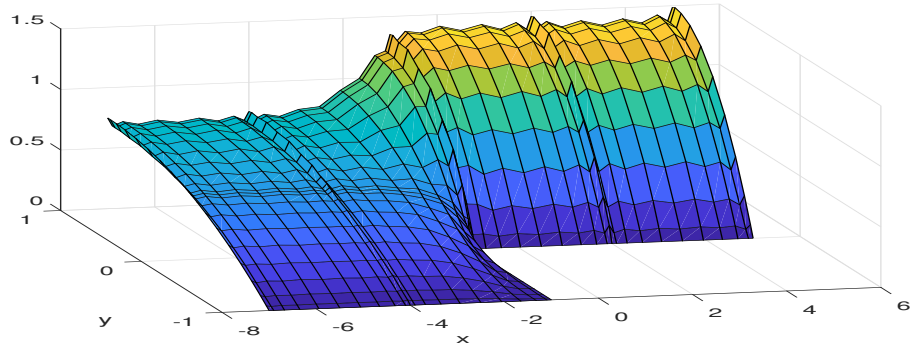
Figure 5.38: Approximated horizontal velocity component for $N = 30, 26$ and 20 on the truncated domain $\Omega = [-7,5] \times [-1,1] \setminus [-1,5] \times [-1,0]$.



(a) $N = 16$



(b) $N = 12$



(c) $N = 10$

Figure 5.39: Approximated horizontal velocity component for $N = 16, 12$ and 10 on the truncated domain $\Omega = [-7,5] \times [-1,1] \setminus [-1,5] \times [-1,0]$.

In these simulations, the exact solution is unknown and so we cannot calculate the error with respect to an exact solution but the iterative method was applied by imposing a criteria on the residual (relative error between the approximate solution obtained at iteration k and the approximate solution obtained at iteration $(k-1)$. As can be seen in Figs. 5.37-5.38-5.39 the smoothness of the approximate solution depends on the discretization parameter, N. The larger the value of N, the smoother the approximation becomes.

5.8.2 Dirichlet conditions on all boundaries

Consider the Stokes problem

$$\begin{cases} -\nabla\cdot(\mu\nabla\boldsymbol{u})+\nabla p &= \boldsymbol{f} & \text{on} \quad \Omega \\ -\operatorname{div}\boldsymbol{u} &= \lambda\displaystyle\int_\Omega p\,d\Omega & \text{on} \quad \Omega \\ \boldsymbol{u} &= \boldsymbol{g} & \text{on} \quad \Gamma \end{cases} \tag{5.38}$$

where $\boldsymbol{u}=(u_1,u_2)^T$ can be interpreted as the velocity field of an incompressible fluid, p is the associated pressure and the function μ is the viscosity of the fluid. We note that the x-axis represents an axis of symmetry for the problem. The problem geometry is shown in Fig. 5.40. We solve Eq. (5.38) subject to the following boundary conditions on $\boldsymbol{u}(x,y)=\big(u_1(x,y),u_2(x,y)\big)$:

$$\begin{cases} u_1(-a,y)=u_{in}(y), u_2(-a,y)=0, & \forall -1\le y\le 1 \\ u_1(b,y)=u_{out}(y), u_2(b,y)=0, & \forall -1\le y\le 1 \\ u_1(0,y)=u_2(0,y)=0, & \forall -1\le y\le 0 \\ u_1(x,1)=u_2(x,1)=0, & \forall -a\le x\le b \\ u_1(x,-1)=u_2(x,-1)=0, & \forall -a\le x\le 0 \\ u_1(x,0)=u_2(x,0)=0, & \forall\, 0\le x\le b \end{cases} \tag{5.39}$$

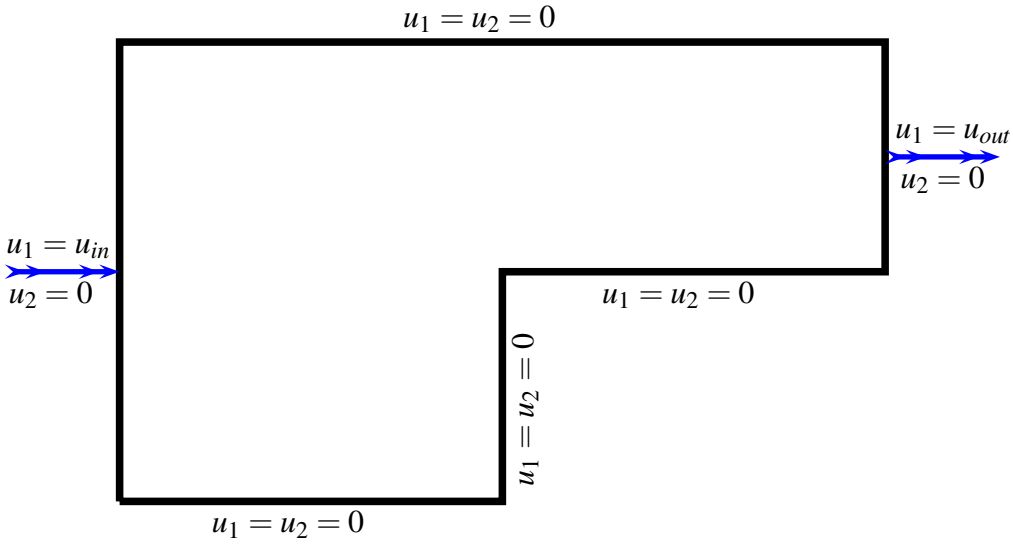


Figure 5.40: Stokes flow in contraction geometries and unbounded domain.

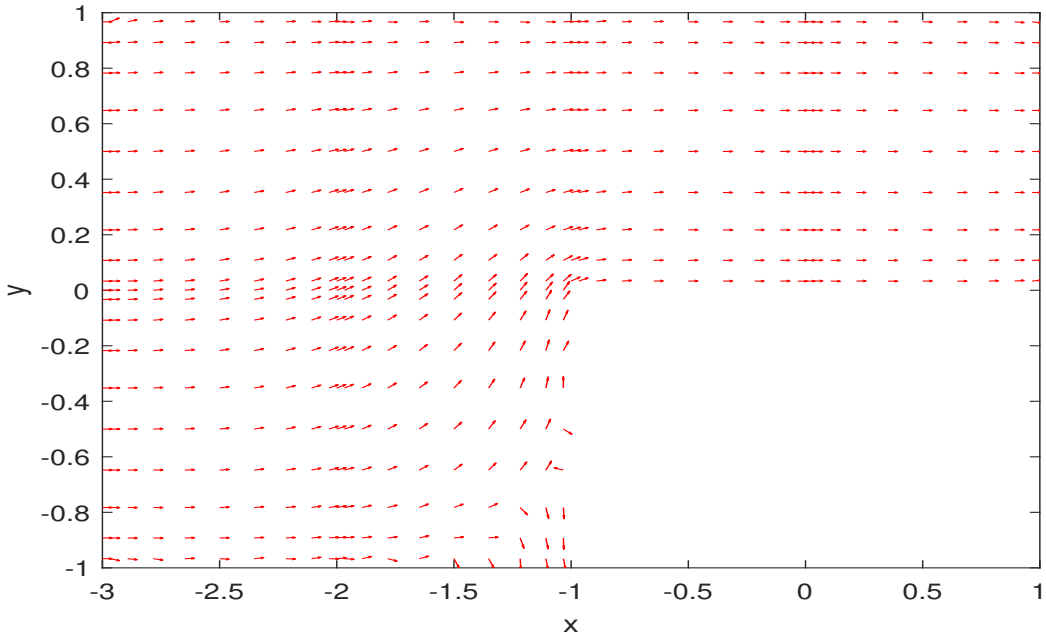
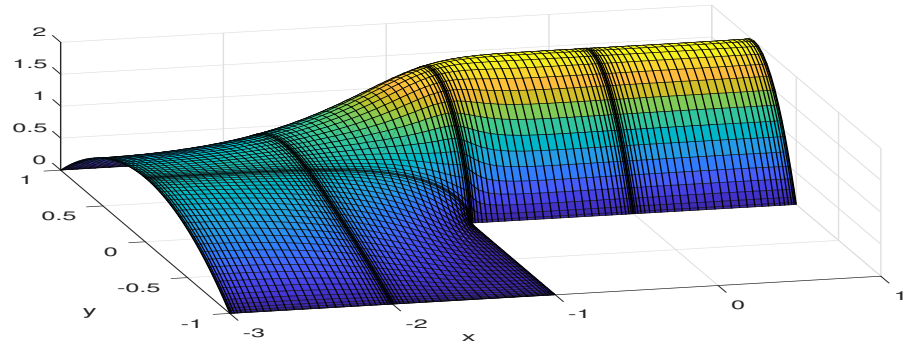
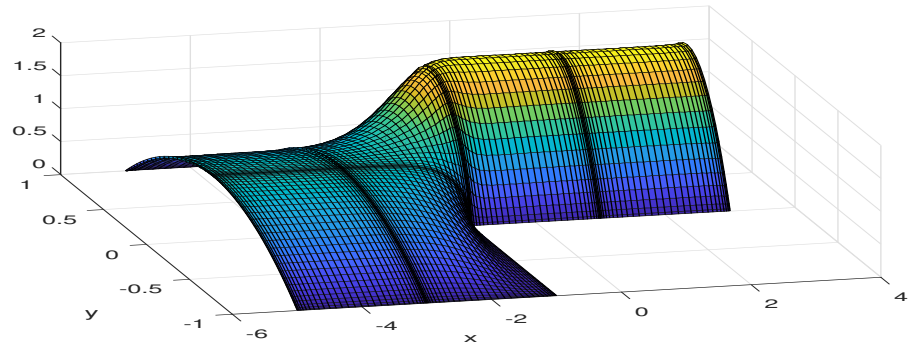


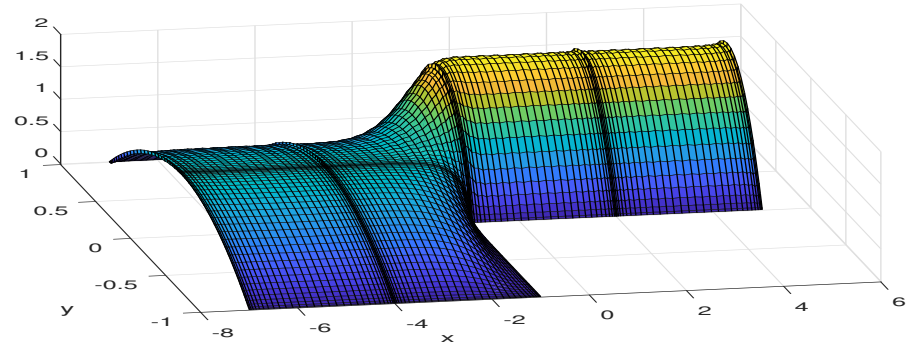
Figure 5.41: Field vector for $N = 10$ on a truncated domain $\Omega = [-3,1] \times [-1,1] \setminus [-1,1] \times [-1,0]$.



(a) $\Omega = [-3,1] \times [-1,1]$

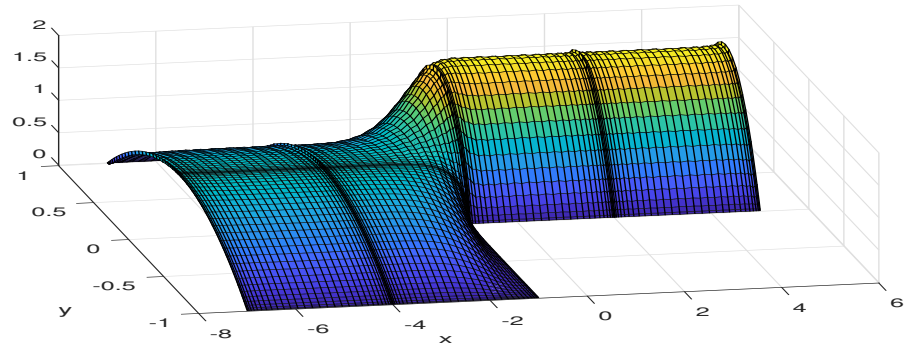


(b) $\Omega = [-5,3] \times [-1,1]$

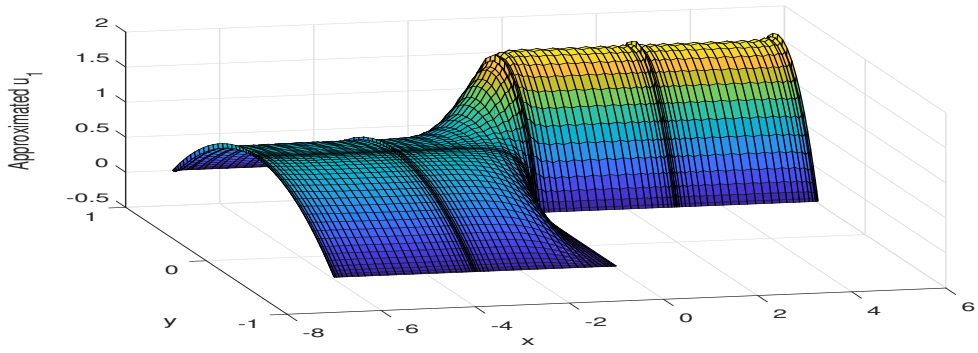


(c) $\Omega = [-7,5] \times [-1,1]$

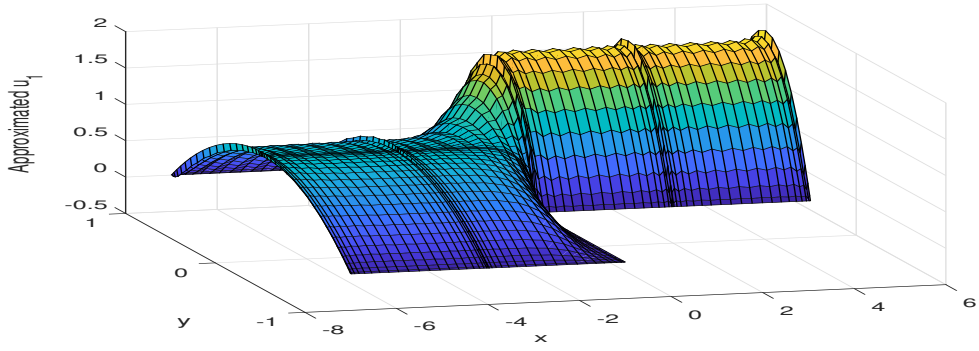
Figure 5.42: Approximated solution for u_1 for $N = 30$ on different truncated domains.



(a) $N = 30$

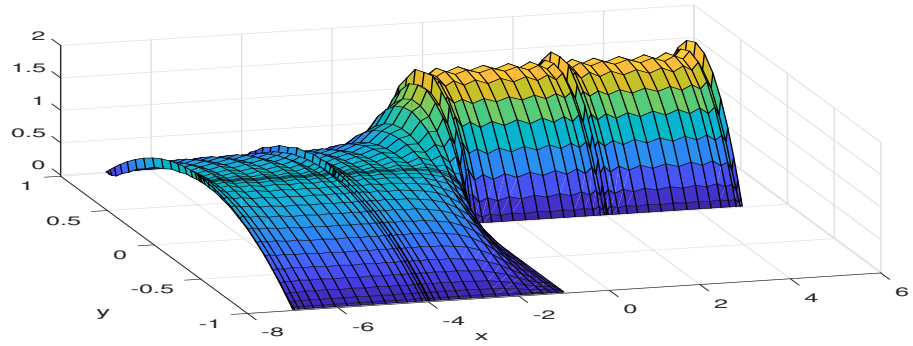


(b) $N = 26$

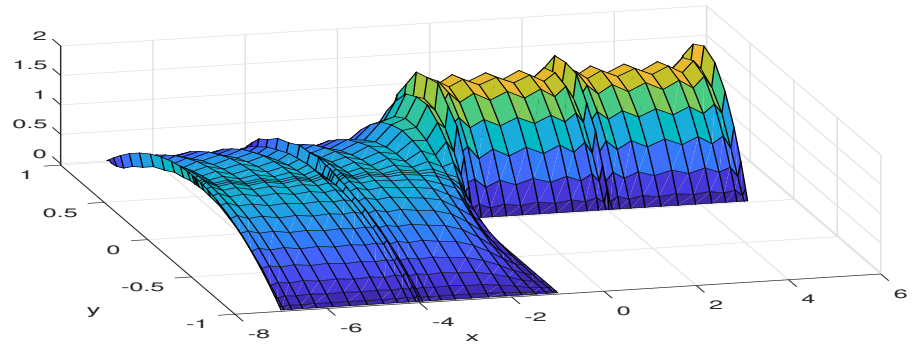


(c) $N = 20$

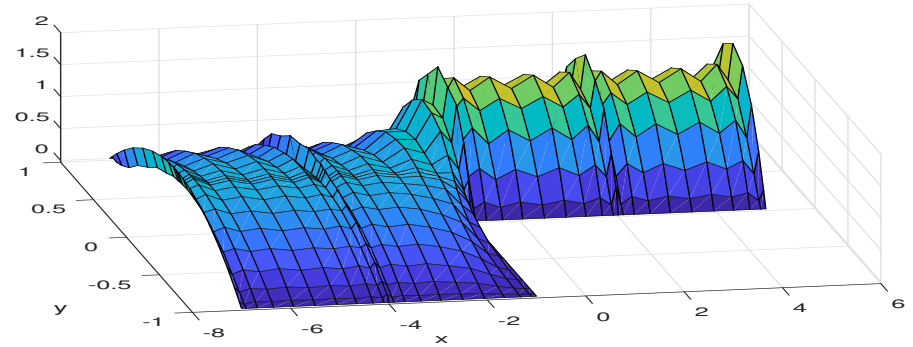
Figure 5.43: Approximated solution for u_1 for $N = 30, 26$ and 20 on the truncated domain $\Omega = [-7,5] \times [-1,1] \setminus [-1,5] \times [-1,0]$.



(a) $N = 16$



(b) $N = 12$



(c) $N = 10$

Figure 5.44: Approximated solution for u_1 for $N = 16, 12$ and 10 on the truncated domain $\Omega = [-7,5] \times [-1,1] \setminus [-1,5] \times [-1,0]$.

Again, the exact solution is unknown and then the iterative method was applied by imposing a criteria on the residual. As it can be noted in Figures 5.42-5.43-5.44 the smoothness of the approximate solution depends on the discretization parameter, N. The smoothness of the solution improves as N increases. By comparing the Dirichlet boudary condition to the mixed one, one can see that for the same discretization parameter, the mixed boundary condition produces a smoother approximate solution than the case of a Dirichlet boundary condition.

5.9 Conclusions

A new formulation of the Stokes problem has been presented and analysed in this chapter. Such a formulation guarantees a unique pressure solution to this problem as it possesses a vanishing mean. The traditional primitive variable formulation of this problem is adjusted by the addition of a scalar multiple of the domain integral of the pressure to the right side of the continuity equation. The modified Stokes problem has a weak formulation and is subject to unique pressure in addition to a unique velocity solution. A successful application has been made of a spectral element approximation (SEM) to the Stokes problem, and the PCG-algorithm was applied to resolve the obtained linear system. Moreover, the L^2-norm of the error converges exponentially. The effect of the scalar multiplier, λ, on the conditioning of the discrete problem and the precision of the spectral element approximations is explored. We discovered a range of values of λ for which the precision and effectiveness of the preconditioned conjugate gradient scheme is optimal. The effectiveness of the scheme can deteriorate significantly for values of λ outside this range, particularly if the operator is conditioned poorly. Furthermore, the precision of the pressure field can be powerfully influenced by the choice of λ, despite the fact that the velocity approximation error is less sensitive to the value of this parameter. Moreover, this chapter also investigates the Stokes flow in contraction geometries and unbounded domains. A primitive variable formulation is applied to present a spectral element technique over elements or semi-infinite rectan-

gular sub-regions. The area of interest is separated into two semi-infinite rectangular sub-regions, and within each of these the solution in a truncated series format involving orthogonal polynomials is presented. In these expansions, the unknown coefficients are established by satisfying the boundary conditions and differential equations at selected points in the area; moreover, the solutions in the two sub-regions are matched by enforcing continuity conditions over the interface. Furthermore, spectral element approximation (SEM) was used to resolve the Stokes problem, in combination with the PCG-algorithm. The L^2– norm of the spectral element Method (SEM) error is dependent on the level of the approximation N (between 2 to 32). Exponential convergence of the approximation to the solution of the Stokes problem was attained with relatively few degrees of freedom.

Chapter 6

Conclusions and Future Work

In this thesis, we have successfully applied a spectral element approximation to some partial differential equations. We introduced the extended spectral element method (XSEM) that we applied to some example problems that possessed a weak discontinuity.

In Chapter 2, we introduced the spectral element method in one dimension and illustrated some numerical examples for linear elliptic equations which demonstrated that the L^2-error converges exponentially with respect to the order of polynomial.

Chapter 3 contains a basic introduction to the spectral element discretisation in 2 D as well as a theoretical analysis of spectral methods. A discussion of the development of spectral domain decomposition methods is given and the subsequent development of spectral element methods showing the transition from the treatment of the strong formulation to that of the weak formulation. The extent of our study is mainly restricted to difficulties with paradigms which depict the basics and verify the connection between classical spectral methods and their domain decomposition progeny. Furthermore, the chapter considered the discretisation of Poisson's equation by applying spectral domain decomposition techniques. The chapter provides a comprehensive explanation of the spectral element discretisation. We applied the PCG-algorithm in order to resolve the derived linear system. The L^2-norm of the spectral element method (SEM) error is dependent on the degree of the approximation, N, as well as the number of elements, and it is shown that the L^2-norm of the error converges exponentially.

Chapter 4 began by introducing the spectral element method and subsequently demonstrated the difficulties related to determining an approximation to the solution of a problem that is discontinuous in the case when discontinuity is not fitted to the computational mesh. In a such situation, oscillations are observed local to the discontinuity that gives rise to the Gibbs phenomenon. Furthermore, the spectral element version of the extended finite element method was introduced, which we refer to as the extended spectral element method (XSEM). We then demonstrated that when approximating a discontinuous function, XSEM is able to approximate the discontinuity precisely. We modified the analysis of Reusken [85] to derive spectral equivalents of his error estimates, and following this, we discussed a possible inf-sup condition. We implemented several enrichment functions with the purpose of improving the approximation of the discontinuous functions, particularly for the two-dimensional Poisson problem. Unfortunately, we have not achieved the desired goal of overcoming the Gibbs phenomenon. However, an alternative approach in which the system (4.1) is solved numerically, by utilizing SEM accompanied with a particular domain decomposition is adopted. We observed that good results are obtained in this case, where the L^2-error converges exponentially for the parameter ε, which is a measure of the width of an element, that contains the discontinuity, perpendicular to the discontinuity.

In Chapter 5, a general study of Stokes problem was undertaken. Its importance in several scientific fields was discussed. The mathematical formulation of the Stokes problem was discussed and numerical methods based on the SEM were introduced. An alternative continuity equation was introduced and shown to be equivalent to the traditional formulation. Its effectiveness for finding a unique solution for pressure is highlighted. The discretization of the Stokes problem is performed using SEM. Iterative methods and their importance in solving algebraic equation systems in general, were presented. The PCG method is used to solve the linear systems and efficient preconditioners are constructed for the Stokes problem. The dependence of the number of iterations for convergence on the discretization parameters is studied. The aim

here was to obtain the optimal preconditioning strategy for the Stokes' system. The condition number of the coefficient matrix in the Stokes system is studied and the influence of parameters on the condition number is investigated. Finally, some problems in complex geometric are considered. The Stokes problem has been solved numerically in an unbounded contraction geometry. The boundary conditions are applied and the effect on both pressure and velocity variables is considered. Using what was previously presented and examined in this chapter we find that approximate solutions to the Stokes problem , which depends on smooth basis functions in the computational domain, utilizing SEM indicated exponential convergence. This is in comparison to the FEM which yields only algabraic convergence.

Future goals which we wish to consider in the future are :

- Application of the XSEM to more realistic example problems. To the best of our knowledge, the literature has not addressed the issue of errors in the spectral estimation of a function from an interpolation space or a broken Sobolev space. Consequently, it is our intention to consider this, and also XSEM, in more detail in the future.

- Treatment of higher-order (curved) interfaces and to compare use of the over-integration instead of the standard quadrature scheme used in the XFEM literature, which involves subdividing the element containing the discontinuity.

- Examination of the inf-sup condition for XSEM, the velocity-pressure and velocity-pressure-stress formulations, and also removing areas of small support as advocated by Grob and Reusken [85].

In our opinion, XSEM has great potential and it is important to continue to resolve some of the outstanding difficulties. We also have a desire to apply XSEM to more physically realistic problems .

Appendices

Appendix A

Legendre polynomials

For our numerical integration procedure we adopt Gauss-Lobatto Legendre quadrature. The fundamental concept of any numerical integration procedure is the approximation of the integral by a quadrature rule.

A.1 Gauss-Lobatto Legendre quadrature

The Legendre polynomials are polynomials that are orthogonal over the interval $[-1,1]$ with respect to weight function $w(x) = 1$. Legendre Polynomials are generated using the following recursive formula:

$$\begin{cases} L_0(x) = 1,\ L_1(x) = x \\ \\ (n+1)L_{n+1}(x) = (2n+1)xL_n(x) - nL_{n-1}(x), n = 1, \cdots. \end{cases}$$

The last relation, along with knowledge of the first two polynomials L_0 and L_1, allows the Legendre Polynomials to be generated recursively. Important properties are given by:

$$(2n+1)L_n(x) = L'_{n+1}(x) - L'_{n-1}(x) \tag{A.1}$$

and:

$$(1-x^2)L'_n(x) = nL_{n-1}(x) - nxL_n(x). \tag{A.2}$$

A.1.1 Legendre-Gauss-Lobatto- Lagrange interpolants

Denote by N the degree of polynomial interpolation and let x_i, $i = 0, \cdots, N$, denote the associated nodes, known as the Gauss-Lobatto Legendre points, which are the zeros of $(1-x^2)L'_N(x)$.

These roots can be found using the Newton-Raphson method. Given a function $f(x) = (1-x^2)L'_N(x)$ defined over the reals line, and its derivative $f'(x)$, we begin by determining a first guess, $x_i^{(0)}$ for a root of the function f. Provided the function satisfies all the assumptions made in the derivation of the formula, an improved approximation $x_i^{(1)}$ is

$$x_i^{(1)} = x_i^{(0)} - \frac{f(x_i^{(0)})}{f'(x_i^{(0)})}.$$

The process is repeated as

$$x_i^{(n+1)} = x_i^{(n)} - \frac{f(x_i^{(n)})}{f'(x_i^{(n)})}$$

until a sufficiently accurate approximation to x_i is reached.

It can be shown that $f(x_i) = L_{N-1}(x_i) - L_{N+1}(x_i) = 0$ which follows from the recurrence relation

$$f(x) = (1-x^2)L'_N(x) = \frac{N(N+1)}{2N+1}\Big(L_{N-1}(x) - L_{N+1}(x)\Big) = N\Big(L_{N-1}(x) - xL_N(x)\Big)$$

and by (A.1), one obtains

$$f'(x) = -N(N+1)L_N(x)$$

Then

$$\frac{f(x)}{f'(x)} = \frac{xL_N(x) - L_{N-1}(x)}{(N+1)L_N(x)}$$

An approximation of the zero of $L_N(x)$ is given by [92]:

$$\sigma_i = \Big[1 - \frac{N-1}{8N^3} - \frac{1}{384N^4}\left(39 - \frac{28}{\sin^2(\theta_i)}\right)\Big]\cos(\theta_i) + O(N^{-5})$$

where θ_i is given by:

$$\theta_i = \frac{4i-1}{4N+2}\pi, \quad 1 \le i \le N.$$

Notice that there exists exactly one zero of $L'_N(x)$ between two consecutive zeros of $L_N(x)$. As for an iterative method, it is essential to start with a good initial approximation, one can take the initial guess as

$$x_i^{(0)} = \frac{\sigma_i + \sigma_{i+1}}{2}, \quad 1 \leq i \leq N-1.$$

Notice that the roots of the Chebyshev polynomials can be used as initial guess $x_i^{(0)} = \cos\left(\frac{i\pi}{N}\right), \quad 1 \leq i \leq N-1.$

The Lagrange interpolant through Legendre-Gauss-Lobatto nodes is described by:

$$h_j(x) = \prod_{\substack{i=1 \\ i \neq j}}^{N} \frac{x - x_i}{x_j - x_i} = -\frac{(1-x^2)L'_N(x)}{N(N+1)L_N(x_j)(x-x_j)} = -\frac{L_{N-1}(x) - L_{N+1}(x)}{(2N+1)L_N(x_j)(x-x_j)}. \quad \text{(A.3)}$$

An important property is given by

$$h_j(x_i) = \delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{otherwise} \end{cases}$$

A.1.2 Weights for Legendre-Gauss-Lobatto numerical integration

Weights for Legendre-Gauss-Lobatto numerical integration are given by:

$$w_i = \frac{2}{N(N+1)} \frac{1}{L_N^2(x_i)}, i = 0, \cdots, N$$

A.1.3 Differentiation matrix

Denote by D the so-called differentiation matrix with dimension $(N+1) \times (N+1)$ with entries defined by $D_{ij} = h'_j(x_i)$, and given explicitly by

$$D_{ij} = \begin{cases} -\dfrac{N(N+1)}{4} & \text{if } i = j = 0 \\ \dfrac{N(N+1)}{4} & \text{if } i = j = N \\ 0 & \text{if } 1 < i = j < N \\ \dfrac{L_N(x_i)}{L_N(x_j)(x_i - x_j)} & \text{if } i \neq j \end{cases}$$

Bibliography

[1] W.F. Ames. *Numerical Methods for Partial Differential Equations*. Academic Press, 2014.

[2] J.P. Ampuero. *The spectral element method*. Lecture notes: GE 263 - Computational Geophysics, 2009.

[3] I. Babuška. The finite element method with Lagrangian multipliers. *Numer. Math.*, 20:179–192, 1973.

[4] I. Babuška, G. Caloz, and J. E. Osborn. Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM J. Numer. Anal.*, 31:945–981, 1994.

[5] K. J. Bathe. *Finite Element Procedures*. Cambridge, MA. Englewood Cliffs, NJ, 1996.

[6] T. Belytschko and T. Black. *Elastic crack growth in finite elements with minimal remeshing*, volume 45. Int. J. Numer. Meth. Engng., 1999.

[7] T. Belytschko, J. Fish, and B.E. Englemann. A finite element with embedded localization zones. *Comput. Meth. Appl. Mech. Engng.*, 70:59–89, 1988.

[8] T. Belytschko, N. Moes, S. Usui, and C. Parimi. Arbitrary discontinuities in finite elements. *Int. J. Numer. Meth. Engng.*, 50:993–1013, 2001.

[9] S. E. Benzley. Representation of singularities with isoparametric finite elements. *Int. J. Numer. Meth. Engng.*, 8:537–545, 1974.

[10] D. Boffi, N. Cavallini, F. Gardini, and L. Gastaldi. *Immersed Boundary Method: Performance Analysis of Popular Finite Element Spaces*. IV International Conference on Computational Methods for Coupled Problems in Science and Engineering, 2011.

[11] D. Boffi, N. Cavallini, F. Gardini, and L. Gastaldi. Local Mass Conservation of Stokes Finite Elements. *J. Sci. Comput.*, 52:383–400, 2012.

[12] S.P.A. Bordas and S.Natarajan. On the approximation in the smoothed finite element method (SFEM). *Int. J. Num. Meth. Engng.*, 81:660–670, 2010.

[13] J. P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications Inc, 2000.

[14] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers. *R.A.I.R.0., Anal. Numer.*, 2:129–151, 1974.

[15] J. Burgerscentrum. Iterative solution methods. *Applied Numerical Mathematics*, 51(4):437–450, 2011.

[16] C. Canuto, A.Quarteroni, M.Y.Hussaini, and T.A. Zang. *Spectral Methods : Evolution to Complex Geometries and Applications to Fluid Dynamics*. Springer: Scient. Comput., 2007.

[17] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral Methods: Fundamentals in Single Domains*. Scientific Computation, Springer-Verlag, 2006.

[18] C. Canuto, A. Quarteroni, M. Y. Hussaini, and T. A. Zang. *Spectral Methods in Fluid Dynamics*. Springer Ser. Comput. Phys.s, 1988.

[19] C. Carstensen. Clément Interpolation and Its Role in Adaptive Finite Element Error Control. *Operator Theory: Adv. Appl.*, 168:27–43, 2006.

[20] K. W. Cheng and T.-P. Fries. Higher-Order XFEM for Curved Strong and Weak Discontinuities. *Int. J. Numer. Meth. Engng.*, 82:564–590, 2010.

[21] J. Chessa, H.Wang, and T. Belytschko. On the Construction of Blending Elements for Local Partition of Unity Enriched Finite Elements. *Int. J. Numer. Meth. Engng.*, 57:1015–1038, 2003.

[22] A.J. Chorin. Numerical Solution of the Navier-Stokes Equations. *Math. Comp.*, 22:745–762, 1968.

[23] Ph.G. Ciarlet. The fimte element method for elhptic problems, 1978.

[24] Ph.G. Ciarlet. *Basic error estimates for elliptic problems*. Elsevier, 1991.

[25] Ph.G. Ciarlet. *The finite element method for elliptic problems*, volume 40. Siam, 2002.

[26] J. Claes. *Numerical solution of partial differential equations by the finite element method*. Courier Corporation, 2012.

[27] M. J. Crochet, A. R. Daves, and K. Walters. *Numerical Simulation of Non-Newtonian Flow*. Elsevier, Amsterdam, 1984.

[28] G. Dahlquist and A. Bjorck. *Numerical Methods*. Englewood Cliffs, N.J., Prentice-Hall series in automatic computation., 1974.

[29] G. Dahlquist and A. Bjorck. *Numerical Methods in Scientific Computing: Volume 1*, volume 103. Siam, 2008.

[30] P.J. Davis and P. Rabinowitz. Some geometrical theorems for abscissas and weights of Gauss type. *J.Math. Anal. Appl.*, 2(3):428–437, 1961.

[31] A. Deraemaeker, I. Babuška, and P. Bouillard. *Dispersion and Pollution of the FEM solution for the Helmholtz equation in one, two and three dimensions*, volume 46(4). 1999.

[32] I.S. Duff. Combining direct and iterative methods for the solution of large systems in different application areas. *Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, Toulouse, France, Tech. Rep. TR/-PA/04/128*, 2004.

[33] Howard C. Elman, David J. Silvester, and Andrew J. Wathen. *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*. Oxford University Press, Oxford, 2006.

[34] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. *Handbook of numerical analysis*, 7:713–1018, 2000.

[35] J. Fish. The s-version of the finite element method. *Computers and Structures*, 43:539–547, 1992.

[36] G.J. Fix. Higher-order Rayleigh-Ritz approximations. *J. Math. Mech.*, 18(7):645–657, 1969.

[37] G.J. Fix, S. Gulati, and G. I. Wakoff. On the use of singular functions with finite element approximations. *J. Comput. Phys.*, 13:209–228, 1973.

[38] O. Garcia, E. Fancello, C. Barcellos, and C. Duarte. Hp clouds in mindlin's thick plate model. *Int. J. Numer. Meth. Engng.*, 47:1381–1400, 2000.

[39] D. Gottlieb and S. A. Orszag. *Numerical analysis of spectral methods: theory and applications*, volume 26. SIAM-CMBS, 1977.

[40] M. Griebel and M. A. Schweitzer. *Meshfree Methods for Partial Differential Equations*. Springer, 2003.

[41] S. GroB and A. Reusken. An extended pressure finite element space for two-phase incompressible flows with surface tension. *J. Comput. Phys.*, 224:40–58, 2007.

[42] D.Rh. Gwynllyw and T.N. Phillips. On the enforcement of the zero mean pressure condition in the spectral element approximation of the Stokes problem. *Comput. Meth. Appl. Mech. Engng.*, 195 (9-12):1027–1049, 2006.

[43] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(1):409–436, 1952.

[44] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb. *Spectral Methods for Time-Dependent Problems*. Cambridge Monogr. Appl. Comput. Math., 2007.

[45] T. Hettich, A. Hung, and E. Ramm. Modeling of failure in composites by X-FEM and level-sets within a multiscale framework. *Comput. Meth. Appl. Mech. Engng.*, 197:414–424, 2008.

[46] M. H. Holmes. Introduction to numerical methods differential equations. 2007.

[47] 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. Meth. Appl. Mech. Engng.*, 127:387–401, 1995.

[48] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, and J.B. Quincy. The variational multiscale method - a paradigm for computational mechanics. *Comput. Meth. Appl. Mech. Engng.*, 166:3–24, 1995.

[49] M.Y. Hussaini, C.L. Streett, and T.A. Zang. *Spectral methods for partial differential equations*. NASA Langley Research Center; Hampton, VA, United States, 1983.

[50] M.Y. Hussaini and T.A. Zang. Spectral methods in fluid dynamics. *Annual Review of Fluid Mechanics*, 19(1):339–367, 1987.

[51] R. Johan and M. Bastiaans. *Transitional free convection flows induced by thermal line sources*. Eindhoven University of Technology, Faculty of Mechanical Engineering, 1993.

[52] A. Karageorghis and T. N. Phillips. Spectral Collocation Methods for Stokes Flow in Contraction Geometries and Unbounded Domains. *Numer. Math.*, 80:314–330, 1988.

[53] G. E. Karniadakis and S. J. Sherwin. *Spectral/hp element methods for computational fluid dynamics*. Oxford Universtity Press, Oxford, 2nd ed. edition, 2004.

[54] Amir R Khoei. *Extended Finite Element Method: Theory and Applications*. John Wiley & Sons, 2014.

[55] D.I Komatitsh, J.P. Violette, R. Vai, J.M. Castillo-Covarrubias, and F.J. Sanchez-Sesma. The spectral element method for elastic wave equations: Application to 2D and 3D seismic problems. *Int. J. Numer. Meth. Engng.*, 45:1139–1164, 1999.

[56] D. A. Kopriva. *Implementing Spectral Methods for Partial Differential Equations: Algorithms for Scientists and Engineers*. Springer, 2008.

[57] O.A. Ladyshenskaya. *The Mathematical Theory of Viscous Incompressible Flow*. 2nd edn, Gordon and Breach, New York, 1969.

[58] L.Ducker. *Level Set and Finite Element Method*. Master's Dissertation.School of Mathematics, The University of Manchester, 2006.

[59] U. Lee. *Spectral element method in structural dynamics*. John Wiley & Sons (Asia) Pte Ltd, 2009.

[60] A. Legay, H. W. Wang, and T. Belytschko. Strong and Weak Arbitrary Discontinuities in Spectral Finite Elements. *Int. J. Numer. Meth. Engng.*, 64:991–1008, 2005.

[61] S. Li and W. K. Liu. *Meshfree Particle Methods*. Springer, 2004.

[62] G.R. Liu, T.T. Nguyen, K.Y. Dai, and K.Y. Lam. Theoretical aspects of the smoothed finite element method (SFEM). *Int. J. Num. Meth. Engng.*, 71:902–930, 2007.

[63] T. Maral. *Spectral (h-p) element methods approach to the solution of Poisson and Helmotz equations using Matlab*. Ph.D. thesis. Middle East Technical University, 2006.

[64] J.M. Melenk. *On generalized finite element methods*. PhD thesis, University of Maryland, College Park, MD, 1995.

[65] J.M. Melenk and I. Babuška. The partition of unity finite element method: Basic theory and applications. *Comput. Meth. Appl. Mech. Engng.*, 139:289–314, 1996.

[66] J. Mergheim. A variational multiscale method to model crack propagation at finite strains. *Int. J. Numer. Meth. Engng.*, 80:269–289, 2009.

[67] N. Moes, M. Cloirec, P. Cartraud, and J.F. Remacle. A computational approach to handle complex microstructure geometries. *Comput. Meth. Appl. Mech. Engng.*, 192:3163–3177, 2003.

[68] N. Moes, J. Dolbow, and T. Belytschko. A finite element method for crack growth without remeshing. *Int. J. Numer. Meth. Engng.*, 46:131–150, 1999.

[69] S. Mohammadi. *Extended finite element method for fracture analysis of structures*. Blackwell Publishing Ltd, 2008.

[70] C.D. Mote. Glocal-local finite element. *Int. J. Numer. Meth. Engng.*, 3:565–574, 1971.

[71] A. A. Munjiza. *The combined finite-discrete element method*. Wiley Publishers, 2004.

[72] S. Natarajan. *Enriched Finite Element Methods: Advances & Applications*. Thesis. Institute of Mechanics and Advanced Materials Theoretical and Computational Mechanics, Cardiff University, 2011.

[73] H. Nguyen-Xuan, S.P.A. Bordas, and H. Nguyen-Dang. Smooth finite element methods: convergence, accuracy and properties. *Int. J. Num. Meth. Engng.*, 74:175–208, 2008.

[74] J. Ockendon, S. Howison, A. Lacey, and A. Movchan. *Applied Partial Differential Equations*. Oxford University Press, Oxford, 1999.

[75] J. Tinsley Oden. Finite elements: Introduction. *Handbook of Numerical Analysis Volime II: Finite Element Methods (Part1)*, 1991.

[76] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.*, 79:12–49, 1988.

[77] R.G. Owens and T.N. Phillips. *Computational Rheology*. Imperial College Press, 2002.

[78] M. Pais, N.H. Kim, and J. Peters. Discussions on modeling weak discontinuities independent of the finite element mesh. *10th US National Congress on Computational Mechanics, Columbus, Ohio*, 2009.

[79] S. V. Patankar. *Numerical Heat Transfer and Fluid Flow*. Taylor & Francis, 1980.

[80] A. T. Patera. A spectral element methods for fluid dynamics: laminar flow in a channel expansion. *J. Comput. Phys.*, 54:468–488, 1984.

[81] J. Pech. *Application of spectral element method in fluid dynamics*. Master's Thesis. Institute of Theoretical Physics, 2006.

[82] S. Pommier, A. Gravouil, A. Combescure, and N. Moes. *Extended Finite Element Method for Crack Propagation*. John Wiley & Sons ISTE Ltd, 2011.

[83] C. Pozrikidis. *Introduction to Finite and Spectral Element Methods Using MATLAB*. CRC Press, Taylor & Francis Group, 2014.

[84] T. Rabczuk, G. Zi, A. Gerstenberger, and W. A.Wall. A new crack tip element for the phantom node method with arbitrary cohesive cracks. *Int. J. Numer. Meth. Engng.*, 75:577–599, 2008.

[85] A. Reusken. Analysis of an eXtended Pressure Finite Element Space for Two-Phase Incompressible Flows. *Comput. Visual Sci.*, 11:293–305, 2008.

[86] J.E. Roberts and J.M. Thomas. *Mixed and Hybrid Methods in Handbook of Numerical Analysis II: Finite Element Methods (Part 1), PG Ciarlet and JL Lions Eds*. Amsterdam, North-Holland, 1991.

[87] C.F. Rowlatt. *Modelling Flows of Complex Fluids using the Immersed Boundary Method*. Ph.D thesis. School of Mathematics, Cardiff University, 2014.

[88] S. A. Sauter and C. Schwab. *Boundary element methods*. Springer-Verlag Berlin Heidelberg, 2011.

[89] B. Saxby. *High-order XFEM with applications to two-phase flows*. PhD, School of Mathematics, The University of Manchester, 2014.

[90] J. A. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fuid mechanics, Computer vision and Material science*. Cambridge University Press, 1999.

[91] A. Shalabney and I. Abdulhalim. Sensitivity-enhancement methods for surface plasmon sensors. *Laser & Photonics Reviews*, 5(4):571–606, 2011.

[92] J. Shen, T. Tang, and L.L. Wang. *Spectral element method : Algorithms, Analysis and Applications*. Springer Ser. Comput. Math., 2011.

[93] E. Siahlooei and S.A.S. Abolfazl. Two iterative methods for solving linear interval systems. *Applied Computational Intelligence and Soft Computing*, 2018, 2018.

[94] J.-H. Song, P. M. Areias, and T. Belytschko. A method for dynamic crack and shear band propagation with phantom nodes. *Int. J. Numer. Meth. Engng.*, 67:868–893, 2006.

[95] G. Strang and G. J. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, 1973.

[96] T. Strouboulis, I. Babuška, and K. Copps. The design and analysis of the generalized finite element method. *Comput. Meth. Appl. Mech. Engng.*, 181:43–69, 2000.

[97] M. C. Sukop. *Lattice Boltzmann Modeling: An Introduction for Geoscientists and Engineers*. Springer,, 2006.

[98] N. Sukumar, D.L. Chopp, N. Moes, and T. Belytschko. Modeling holes and inclusions by level sets in the extended finite-element method. *Comput. Meth. Appl. Mech. Engng.*, 190:6183–6200, 2001.

[99] M. Sussman, P. Smereka, and S. Osher. A level set approach for computing solutions to incompressible two-phase flow. *J. Comput. Phys.*, 114:146–159, 1994.

[100] R. Temam. Une méthode d'approximation de la solution des équations de Navier-Stokes. *Bull. Soc. Math. France*, 98:115–152, 1968.

[101] P. Tong, T.H.H. Pian, and S.J. Lasry. A hybrid-element approach to crack problems in plane elasticity. *Int. J. Numer. Meth. Engng.*, 7:297–308, 1973.

[102] T. Vidar. Finite difference methods for linear parabolic equations. *Handbook of numerical analysis*, 1:5–196, 1990.

[103] T. Vidar. From finite differences to finite elements a short history of numerical analysis of partial differential equations. In *Numerical Analysis: Historical Developments in the 20th Century*, pages 361–414. Elsevier, 2001.

[104] R. Wait and A.R. Mitchell. Corner singularities in elliptic problems by finite element methods. *J. Comput. Phys.*, 8:45–52, 1971.

[105] J.R. Whiteman and J. E. Akin. Finite elements, singularities and fracture : The Mathematics of Finite Elements and Applications III. *London ; New York : Academic Press*, pages 35–54, 1978.

[106] J. Wloka. *Partial Differential Equations*. Cambridge University Press, Cambridge, 1992.

[107] F.D. Xie and X.S. Gao. Exact traveling wave solutions for a class of nonlinear partial differential equations. *Chaos Solitons Fractals*, 19:1113–1117, 2004.

[108] Q.Z .Zhu, S.T .Gu, Julien Yvonnet, J.F Shao, and Q.C .He. Three-dimensional numerical modelling by XFEM of spring-layer imperfect curved interfaces with applications to linearly elastic composite materials. *Int. J. Numer. Meth. Engng.*, 88(4):307–328, 2011.

[109] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu. *The finite element method: its basics and fundamentals*. Elsevier Butterworth Heinemann, 6th edition, 2000.