

Geometrically unfitted finite element methods for the Helmholtz equation

Luke James Swift

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of the
University of London.

Department of Mathematics
University College London

January 30, 2018

Declaration

I, Luke James Swift confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Dedication

For Jasper James Swift.

Abstract

It is well known that the standard Galerkin finite element method experiences difficulties when applied to the Helmholtz equation in the medium to high wave number regime unless a condition of the form $hk^2 < C$ is satisfied, where h is the mesh parameter and k is the wave number. This condition becomes even more difficult to enforce when coupling multiple domains which may have different wave numbers. Numerous stabilizations have been proposed in order to make computations under the engineering rule of thumb $hk < C$ feasible. In this work I introduce a theoretical framework for analysing a class of stabilized finite element methods. I introduce two stabilized methods drawing inspiration from [34] and [22]. The methods both have an absolute stability property when considering Dirichlet or Neumann boundary conditions without condition on the mesh parameter or wave number and are shown to be stable for an appropriate choice of stabilization parameters when considering impedance boundary conditions. Numerically, I observe a reduction of the pollution error for given problems provided the stabilization parameters are chosen appropriately. These stabilizations are then extended to a fictitious domain setting using cut elements.

Using these methods provides a platform from which to analyse the effect of multi-domain coupling on the accuracy of solutions. The stabilizations proposed previously are used in the bulk of each domain and the coupling at the interface is handled using Nitsche's method [47]. The coupling parameters are chosen to be complex with positive or negative imaginary part depending on the sign in the Robin boundary condition. The new fitted and unfitted domain decomposition methods have similar properties to the original stabilized methods and both enter a similar theoretical framework. Interestingly the numerics seem to show that the introduction of the Nitsche coupling terms preserve the reduction of pollution effect noted earlier for the fitted domain decomposition methods for both unstructured and non-matching meshes.

Acknowledgements

The journey to completing a research degree is one that has many ups and downs and it is a journey that I could not have completed alone. Fortunately, I have had some great support along the way.

I would like to begin by thanking my supervisor Prof. E. Burman for his guidance throughout the project. His patience and mathematical expertise have been extremely important to my research and I cannot thank him enough for believing in me. The CutFEM library in FEniCs was an extremely important tool for the numerical studies carried out in this project and therefore I would like to extend my thanks to all that contributed to it, with special thanks going to Dr. S. Claus.

I am thankful to my parents, Denise and Carl who have always helped me as best they can and always encouraged me to follow my dreams. The values that they taught me have moulded me into the person I am today. I would also like to thank my late nan, Jill Tricker for her encouragement and ability to listen and offer advice when times seemed hardest. Thanks should also go to my grandad Jim Swift, who is always available on the end of the phone to make me smile.

Finally, I would like to thank my wife and biggest supporter Chloe who has experienced every up and down on this journey with me. The last four years have been tough but would have been practically impossible without Chloe's support and perpetual belief in me.

Contents

1	Introduction	14
2	Stabilized FEM for Wave equations	23
2.1	The Helmholtz equation	23
2.2	Stabilized finite element formulations	26
2.3	A priori error estimates and well-posedness of the discrete system	27
2.3.1	Galerkin Least Squares method with interior penalty on the normal gradient	33
2.3.2	Continuous Interior Penalty method with penalty on the jumps of first and second derivatives	44
2.4	Numerical examples	50
2.4.1	Bessel solution	52
2.4.2	Plane wave solution	56
3	Fictitious domain methods for Helmholtz equation using cut elements	72
3.1	Introduction	72
3.2	Generalized boundary conditions	74
3.2.1	Numerical Results	84
3.2.2	Plane wave solution	85
3.3	Fictitious domain method	88
3.3.1	Numerical Results	97
4	Multi-domain coupling for Helmholtz equations	101
4.1	Introduction	101
4.2	The coupled problem	102
4.3	The fitted domain decomposition method	105
4.3.1	Preliminaries	106
4.3.2	The formulation	107

4.3.3	Mathematical framework for the coupled problem	110
4.3.4	Mathematical tools	112
4.3.5	A priori error estimates	113
4.3.6	Discussion	123
4.3.7	Numerical Results	124
4.3.8	Conclusion	130
4.4	Unfitted domain decomposition	130
4.4.1	Preliminaries	131
4.4.2	The formulation	132
4.4.3	Mathematical tools	134
4.4.4	A priori error estimates	135
4.4.5	Numerical Results	143
5	Conclusion	146
	Appendices	147
A	Appendix	148
A	Useful Inequalities	148
	Bibliography	151

General Convention

Throughout this report, C will be used to denote a generic constant that is independent of wavenumber k and mesh size h , unless otherwise stated.

Nomenclature

$(u, v)_{0,\Omega}$	Complex inner product in $L^2(\Omega) : (\int_{\Omega} u\bar{v})^{1/2}$
\bar{u}	Complex conjugate of $u : \bar{u} = Re[u] - Im[u]i$
$[[\cdot]]$	Jump over element faces
$\langle u, v \rangle_{0,\partial\Omega}$	Complex inner product in $L^2(\partial\Omega) : (\int_{\partial\Omega} u\bar{v})^{1/2}$
$\ u\ _{0,\Omega}$	Norm in $L^2(\Omega) : (\int_{\Omega} u ^2)^{1/2}$
$\ u\ _{s,\Omega}$	Norm in $H^s(\Omega) : \left(\sum_{ \alpha \leq s} \ \partial^\alpha u\ _{0,\Omega}^2\right)^{1/2}$ where s is a non-negative integer
$ u _{s,\Omega}$	Semi norm in $H^s(\Omega) : u _{s,\Omega} = \sum_{ \alpha =s} \ \partial^\alpha u\ _{0,\Omega}$ where s is a non-negative integer
$\ u_h\ _{p,\mathcal{T}_h}$	Broken norm on the triangulation \mathcal{T}_h , given by $(\sum_{\tau\in\mathcal{T}_h} \ u_h\ _{p,\tau}^2)^{1/2}$
\mathbb{C}	Set of complex numbers
\mathbb{N}	Set of natural numbers
\mathbb{P}_p	Set of polynomial functions of order at most p , for $p \in \mathbb{N}$
\mathbb{R}	Set of real numbers
\mathcal{F}_{int}	Set of interior faces
\mathcal{T}_h	Triangulation of a domain
$meas_n(E)$	Lebesgue measure of $E \subset \mathbb{R}^n$
$\{\cdot\}$	Average over element faces
$H^s(\mathcal{T}_h)$	The space of functions whose element-wise derivatives up to order s are in $L^2(\tau)$ for all $\tau \in \mathcal{T}_h$ for $s \in \mathbb{N}$
$H^s(\Omega)$	Fractional Sobolev space, $W^{s,2}(\Omega)$ as defined in A.1, for $s > 0$
$H^s(\Omega)$	The space of functions whose derivatives up to order s are in $L^2(\Omega)$ for $s \in \mathbb{N}$

$H^{-s}(\Omega)$	Dual of $H_0^s(\Omega)$
i	Imaginary unit $i = \sqrt{-1}$
$Im[u]$	Imaginary part of u
$L^p(\Omega)$	The space of functions whose p^{th} power is Lebesgue integrable on Ω
$Re[u]$	Real part of u

List of Figures

2.1	Convergence plot for the Bessel Solution using piecewise linear elements on a structured mesh	53
2.2	Pollution study for Bessel Solution using piecewise linear elements on a structured mesh	53
2.3	Pollution study for Bessel Solution using piecewise quadratic elements on a structured mesh	54
2.4	Convergence plot for the Bessel Solution using piecewise linear elements on an unstructured mesh	54
2.5	Convergence plot for the Bessel Solution using piecewise quadratic elements on an unstructured mesh	55
2.6	Pollution study for Bessel Solution using piecewise linear elements on an unstructured mesh	55
2.7	Pollution study for Bessel Solution using piecewise quadratic elements on an unstructured mesh	56
2.8	Convergence plot for the plane wave solution using piecewise linear elements on a structured mesh with $k = 10$	57
2.9	Convergence plot for the plane wave solution using piecewise linear elements on a structured mesh with $k = 100$	58
2.10	Convergence plot for the plane wave solution using piecewise linear elements on an unstructured mesh with $k = 10$	59
2.11	Convergence plot for the plane wave solution using piecewise linear elements on an unstructured mesh with $k = 100$	59
2.12	Pollution error using piecewise linear elements on a structured mesh where $kh = 0.1$	60
2.13	Pollution error using piecewise linear elements on a structured mesh where $kh = 0.5$	61
2.14	Pollution error using piecewise linear elements on an unstructured mesh where $kh = 0.1$	62
2.15	Pollution error using piecewise linear elements on an unstructured mesh where $kh = 0.5$	62
2.16	Convergence plot for the plane wave solution using piecewise quadratic elements on a structured mesh with $k = 10$	63

2.17	Convergence plot for the plane wave solution using piecewise quadratic elements on a structured mesh with $k = 100$	63
2.18	Convergence plot for the plane wave solution using piecewise quadratic elements on an unstructured mesh with $k = 10$	64
2.19	Convergence plot for the plane wave solution using piecewise quadratic elements on an unstructured mesh with $k = 100$	64
2.20	Pollution error using piecewise quadratic elements on a structured mesh where $kh = 0.2$	65
2.21	Pollution error using piecewise quadratic elements on a structured mesh where $kh = 1.0$	65
2.22	Pollution error using piecewise quadratic elements on an unstructured mesh where $kh = 0.2$	66
2.23	Pollution error using piecewise quadratic elements on an unstructured mesh where $kh = 1.0$	66
2.24	Pollution error using piecewise quadratic elements on an unstructured mesh where $kh = 1.25$	67
2.25	Parameter sensitivity study for $Im[\delta_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$	67
2.26	Parameter sensitivity study for $Im[\gamma_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$	68
2.27	Parameter sensitivity study for $Re[\delta_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$	68
2.28	Parameter sensitivity study for $Re[\gamma_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$	69
2.29	Parameter sensitivity study for $Re[\beta_1]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$.	69
2.30	Parameter sensitivity study for $Im[\beta_1]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$.	70
2.31	Parameter sensitivity study for $Im[\gamma_{2,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$	70
2.32	Parameter sensitivity study for $Re[\gamma_{2,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$	70
2.33	Parameter sensitivity study for $Re[\beta_2]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$.	71
2.34	Parameter sensitivity study for $Im[\beta_2]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$.	71
3.1	Convergence plot for generalized boundary conditions using piecewise linear elements on a structured mesh	85
3.2	Convergence plot for generalized boundary conditions using piecewise linear elements on an unstructured mesh	86
3.3	Convergence plot for generalized boundary conditions using piecewise quadratic elements on a structured mesh	86
3.4	Convergence plot for generalized boundary conditions using piecewise quadratic elements on an unstructured mesh	87
3.5	Solution plots for $k \approx 70, N = 200, \theta = \frac{\pi}{5}$	88
3.6	Solution plots for $k \approx 173, N = 400, \theta = \frac{\pi}{5}$	88

3.7	Physical and computational domain	89
3.8	Convergence study for fictitious domain method using piecewise linear elements	99
3.9	Pollution study for fictitious domain method using piecewise linear elements	100
4.1	Solution plots for domain decomposition problem with matching meshes	125
4.2	Solution plots for domain decomposition problem with matching meshes	126
4.3	Convergence comparison for domain decomposition using piecewise linear elements with non-matching mesh	127
4.4	Example of a non-matching mesh	127
4.5	Convergence comparison for domain decomposition using piecewise linear elements with matching mesh	128
4.6	Example of a matching mesh	128
4.7	Convergence comparison for domain decomposition using piecewise quadratic elements with matching mesh	129
4.8	Comparison of domain decomposition methods with matching mesh for $kh = C$ which demonstrates the the stabilized methods' ability to reduce pollution	129
4.9	Comparison of domain decomposition methods using piecewise linear elements with a non-matching mesh for $kh = C$	130
4.10	Cut element example	132
4.11	Example of the domain of integration for cut elements, where $\tau_j \stackrel{\text{def}}{=} \tau \cap \Omega_j$	140
4.12	Convergence study for an unfitted domain decomposition method with no stabilization	144
4.13	Convergence study for an unfitted domain decomposition method with GLS/CIP stabi- lization	145
4.14	Convergence comparison between a stabilized and unstabilized unfitted domain decom- position method coupling domains with different wave numbers	145

Chapter 1

Introduction

Waves are naturally occurring phenomena that arise in numerous areas of Engineering and Physics. Acoustic waves are of particular interest in the areas of seismology and ocean acoustics. The speed of a wave is related to the density and compressibility of the medium through which it travels. Simply stated the more dense/ less compressible the medium the faster the wave. In many practical situations the wave being modelled will pass through numerous different materials which affect its speed and can give rise to coupled systems. It is therefore desirable to have a method for calculating the solutions of these problems reliably. In most practical applications it is possible to reduce the complexity of the wave equation by making certain assumptions about the solution. Under the assumption that waves are steady-state with constant frequency ω , the Helmholtz equation can be derived. Recall the wave equation

$$\frac{\partial^2}{\partial t^2} U(\underline{x}, t) = c^2 \Delta U(\underline{x}, t). \quad (1.1)$$

From our assumption solutions have the form $U(\underline{x}, t) = u(\underline{x})e^{i\omega t}$, which reduces the wave equation to

$$\begin{aligned} \frac{\partial^2}{\partial t^2} u(\underline{x})e^{i\omega t} &= c^2 \Delta u(\underline{x})e^{i\omega t} \\ u(\underline{x}) \frac{\partial^2}{\partial t^2} e^{i\omega t} &= c^2 e^{i\omega t} \Delta u(\underline{x}) \\ -\omega^2 u(\underline{x})e^{i\omega t} &= c^2 e^{i\omega t} \Delta u(\underline{x}) \\ -\Delta u(\underline{x}) - \frac{\omega^2}{c^2} u(\underline{x}) &= 0. \end{aligned}$$

Finally letting $k = \frac{\omega}{c}$ gives

$$-\Delta u(\underline{x}) - k^2 u(\underline{x}) = 0. \quad (1.2)$$

The resulting equation is referred to as the homogeneous Helmholtz equation. Equations of this form also occur in areas of physics involving Maxwell's equation and Schrödinger's equation.

A Partial Differential Equation (PDE) like Helmholtz equation is usually stated as a boundary value problem. That is, find a solution to the equation on a given domain that satisfies certain conditions on

the domain's boundary. For the resulting system to be considered well posed it is necessary to choose appropriate boundary conditions. When dealing with an interior problem, solutions of Helmholtz are said to be stationary wave solutions and can be considered with either Dirichlet or Neumann boundary conditions. For the interior problem well-posedness comes from the Fredholm alternative which says that

$$\left\{ \begin{array}{l} - \text{ either } k^2 \text{ is not an eigenvalue in the spectrum of } -\Delta \text{ and the Helmholtz equation admits a} \\ \text{ unique solution} \\ - \text{ or } k^2 \text{ is an eigenvalue in the spectrum of } -\Delta \text{ and the corresponding eigenfunctions are} \\ \text{ solutions of the homogeneous Helmholtz equation with zero right hand side hence the} \\ \text{ boundary value problem is not well posed} \end{array} \right.$$

Alternatively, one can also consider an exterior problem in this case solutions are referred to as progressive wave solutions. Here $-\Delta$ with Dirichlet or Neumann boundary conditions is not self-adjoint and does not have a compact inverse in $L^2(\Omega)$. For the exterior problem it is necessary to impose an extra condition to guarantee uniqueness of solutions. This condition is known as the Sommerfeld radiation condition and asks that

$$|\nabla u \cdot n + iku| \leq \frac{C}{|x|^{\frac{d+1}{2}}} \quad \text{at infinity,}$$

where n is taken to be the outward facing normal of a sphere centred at O of a radius R such that $|x| = R$ as R tends to ∞ and $d \in \{2, 3\}$ is taken to be the dimension of the problem. A proof that this leads to a well posed problem can be found in [46]. This condition essentially ensures that no energy can be reflected back into the system from infinity. I shall be using the first order approximation of the Sommerfeld radiation condition which is a Robin boundary condition.

When considering a PDE, like Helmholtz, on a complicated domain the Finite Element Method (FEM) is usually the engineering method of choice. The FEM is a method for approximating the solution of a PDE numerically and involves reformulating the PDE into a weak form. This manipulation involves multiplying by a test function and integrating over the domain of the problem. It is true that if the new equation is satisfied for all test functions in some test space then the element that satisfies this condition is the solution to the original PDE in some trial space. The FEM replaces the trial and test spaces with finite dimensional approximations and the solution is represented by the function in the finite dimensional trial space that satisfies the weak formulation for all test functions in the finite dimensional test space.

When dealing with elliptic, positive definite problems, well-posedness of the continuous and discrete problems are a direct result of the Lax-Milgram theorem. In this case the term positive definite is taken to mean that the sesquilinear form $A(\cdot, \cdot)$ is coercive or V -elliptic,

Definition 1.0.1 (Coercivity, V -ellipticity). Letting V be a Hilbert space, a sesquilinear form $A(\cdot, \cdot) :$

$V \times V \mapsto \mathbb{C}$ is said to be coercive if there exists a real constant $\alpha > 0$ such that

$$A(u, u) \geq \alpha \|u\|_V^2.$$

Lemma 1.1 (Lax-Milgram). *Let V be a Hilbert space. Then given a bounded, coercive sesquilinear form $A : V \times V \mapsto \mathbb{C}$ and a bounded linear functional f defined on V , there exists a unique element $u_0 \in V$ such that*

$$A(u_0, v) = (f, v) \quad \forall v \in V.$$

A proof of Lax-Milgram's lemma can be found for example in [56]. It can be shown that Helmholtz equation on a bounded domain in the low wave number regime satisfies the Lax-Milgram lemma when considered with Dirichlet boundary conditions. The coercivity assumption is broken as k increases therefore the user is forced to leave the Lax-Milgram framework. To illustrate the complications faced when attempting to prove coercivity let us consider the Helmholtz problem on a domain Ω with Dirichlet boundary conditions. The sesquilinear form $A(\cdot, \cdot)$ is then given by

$$A(u, v) = (\nabla u, \nabla v)_{0,\Omega} - k^2(u, v)_{0,\Omega} \quad \forall v \in V, \quad (1.3)$$

where we define our space V as

$$V \stackrel{\text{def}}{=} \{v \in H^1(\Omega) : v = 0 \text{ on } \partial\Omega\}. \quad (1.4)$$

It follows that

$$\begin{aligned} A(u, u) &= \|\nabla u\|_{0,\Omega}^2 - k^2 \|u\|_{0,\Omega}^2 \\ &= (1 - \alpha) \|\nabla u\|_{0,\Omega}^2 + \alpha \|\nabla u\|_{0,\Omega}^2 - k^2 \|u\|_{0,\Omega}^2, \quad \alpha \in [0, 1]. \end{aligned}$$

Since the space, V , is designed such that u vanishes on the boundary the following Poincaré inequality,

$$\|\nabla u\|_{0,\Omega}^2 \geq \frac{1}{C_p^2} \|u\|_{0,\Omega}^2,$$

holds for a positive constant C_p . Which gives

$$\begin{aligned} A(u, u) &\geq (1 - \alpha) \|\nabla u\|_{0,\Omega}^2 + \frac{\alpha}{C_p^2} \|u\|_{0,\Omega}^2 - k^2 \|u\|_{0,\Omega}^2 \\ &\geq \min\left\{(1 - \alpha), \frac{\alpha}{C_p^2} - k^2\right\} \|u\|_{1,\Omega}^2. \end{aligned}$$

So the form is coercive under the condition

$$k < C_p^{-1}.$$

In most practical applications this inequality does not hold and the problem is indefinite. This does not mean that the weak form of the Helmholtz problem is ill-posed but it does mean that the standard finite element analysis based on the Lax-Milgram lemma will not carry over from the continuous to the discrete case without some extra work. It can be shown that the Helmholtz equation on weak form is uniquely solvable using an argument found in [42]. The argument relies on the fact that the Helmholtz equation satisfies a Gårding inequality which implies existence if it can be shown that the solution of the adjoint problem is unique, see for example [53].

Consider the following Helmholtz equation on a Lipschitz bounded domain Ω

$$-\Delta u - k^2 u = f \quad \text{in } \Omega \quad (1.5)$$

$$\nabla u \cdot n + iku = g \quad \text{on } \partial\Omega, \quad (1.6)$$

for $f \in H^{-1}(\Omega)$ and $g \in H^{-\frac{1}{2}}(\partial\Omega)$, then letting $V \stackrel{\text{def}}{=} H^1(\Omega)$ the weak form is given by

$$A(u, v) = (\nabla u, \nabla v)_{0,\Omega} - k^2(u, v)_{0,\Omega} + ik \langle u, v \rangle_{0,\partial\Omega} = (f, v)_{0,\Omega} + \langle g, v \rangle_{0,\partial\Omega} \quad \forall v \in V, \quad (1.7)$$

and the adjoint problem is given by

$$\overline{A(v, u)} = (\nabla u, \nabla v)_{0,\Omega} - k^2(u, v)_{0,\Omega} - ik \langle u, v \rangle_{0,\partial\Omega} \quad \forall v \in V. \quad (1.8)$$

Then solutions are unique if

$$A(u, v) = 0 \forall v \in V \implies u = 0.$$

By letting $v = u$,

$$\|\nabla u\|_{0,\Omega}^2 - k^2 \|u\|_{0,\Omega}^2 + ik \|u\|_{0,\partial\Omega}^2 = 0. \quad (1.9)$$

Considering the imaginary part implies that $u = 0$ on $\partial\Omega$. This gives that $u \in H_0^1(\Omega)$ is the solution of the homogeneous problem

$$(\nabla u, \nabla v)_{0,\Omega} - k^2(u, v)_{0,\Omega} = 0 \quad \forall v \in V. \quad (1.10)$$

Notice that it is possible to extend the function u in the following manner

$$\hat{u}(x) = \begin{cases} u(x) & \text{for } x \in \Omega \\ 0 & \text{for } x \in \hat{\Omega} \setminus \Omega \end{cases}$$

The extended function now satisfies (1.10) on $\hat{\Omega}$. Therefore \hat{u} solves the homogeneous Helmholtz equation and vanishes on $\hat{\Omega} \setminus \Omega$. A continuation argument, see for example [40], then shows that $\hat{u} = 0$ on Ω also.

To prove existence and uniqueness of FE solutions there are two main approaches. The first is to show that the sesquilinear form obeys a discrete inf-sup condition. The inf-sup condition is a more general version of the coercivity condition in Lax-Milgram's theorem which gives information about the stability of a sesquilinear form. Firstly, let's introduce the Babuška theorem a proof of which can be found in [4] which acts as a generalization of Lax-Milgram.

Theorem 1 (Babuška). *Assume that sesquilinear form $A(\cdot, \cdot) : V_1 \times V_2 \mapsto \mathbb{C}$ on Hilbert spaces V_1, V_2 satisfies*

(B1) *Continuity*

$$\exists M > 0 : |A(u, v)| \leq M \|u\|_{V_1} \|v\|_{V_2}, \quad \forall u \in V_1, v \in V_2 \quad (1.11)$$

(B2) *inf-sup Condition:*

$$\exists \beta > 0 : \beta \leq \sup_{v \in V_2 \setminus \{0\}} \frac{|A(u, v)|}{\|u\|_{V_1} \|v\|_{V_2}}, \quad \forall u \in V_1 \setminus \{0\} \quad (1.12)$$

(B3) *“Transposed” inf-sup Condition:*

$$\sup_{u \in V_1 \setminus \{0\}} |A(u, v)| > 0, \quad \forall v \in V_2 \setminus \{0\} \quad (1.13)$$

and let $f : V_2 \mapsto \mathbb{C}$ be an antilinear bounded functional defined on V_2 . Then there exists a unique element $u_0 \in V_1$ such that

$$A(u_0, v) = f(v) \quad \forall v \in V_2.$$

The solution u_0 satisfies the bound

$$\|u_0\|_{V_1} \leq \frac{1}{\beta} \|f\|_{V_2'}. \quad (1.14)$$

From this theorem it is clear that properties (B2) and (B3) may not hold on a FE subspace V_h of the infinite dimensional space V as the supremum will generally decrease when taken over a subspace. This

leads to the discrete inf-sup condition. Essentially assume that conditions of Theorem 1 are satisfied for a sesquilinear form $A(\cdot, \cdot) : V_1 \times V_2 \mapsto \mathbb{C}$ on $V_1 \times V_2$ where V_1, V_2 are Hilbert spaces. Now let $V_1^h \subset V_1$, $V_2^h \subset V_2$ be subspaces, then if $A(\cdot, \cdot)$ also satisfies

(B2_h) the discrete inf-sup condition

$$\exists \beta_h > 0 : \beta_h \leq \sup_{v_h \in V_2^h \setminus \{0\}} \frac{|A(u_h, v_h)|}{\|u_h\|_{V_1} \|v_h\|_{V_2}}, \quad \forall u_h \in V_1^h \setminus \{0\}, \quad (1.15)$$

(B3_h) the transposed condition

$$\sup_{u_h \in V_1^h \setminus \{0\}} |A(u_h, v_h)| > 0 \quad \forall v_h \in V_2^h \setminus \{0\}, \quad (1.16)$$

then there exists a unique element $u_h^0 \in V_1^h$ such that

$$A(u_h^0, v_h) = f(v_h), \quad \forall v_h \in V_2^h.$$

It can be shown that the standard Galerkin FEM with piecewise linear elements applied to Helmholtz on weak form satisfies the discrete inf-sup condition on a bounded star-shaped domain with smooth boundary under the assumption $(1 + k^2)h < C$ see, for example, [42]. Under this assumption and replacing the standard H^1 norm with the k dependent norm,

$$\|v\|_{\mathcal{H}} = \|\nabla v\|_{0,\Omega} + |k| \|v\|_{0,\Omega}.$$

it is possible to obtain $\beta_h = O(k^{-1})$ and a quasi optimal error estimate

$$\|u - u_h\|_{\mathcal{H}} \leq C_3 \inf_{v_h \in V_h} \|u - v_h\|_{\mathcal{H}}.$$

This restriction is far too stringent for most practical applications but is necessary for stability. The problems arise in the high wave number regime where it is necessary to take a vast amount of mesh points to ensure that the discrete system is well posed.

Another technique for analysing the FEM for the solution of Helmholtz problems comes from an observation by A.Schatz [49] regarding indefinite bilinear forms. It is possible to use this method to deduce the standard FEM is stable and quasi-optimal, although this result only holds if the mesh size h is small enough. Schatz's argument relies on the associated adjoint problem being uniquely solvable and requires the following inequalities to hold

- Continuity

$$|A(u, v)| \leq M \|u\|_{1,\Omega} \|v\|_{1,\Omega} \quad \forall u, v \in H_0^1(\Omega).$$

- Gårding inequality

$$C_1 \|v\|_{1,\Omega}^2 - C_2 \|v\|_{0,\Omega}^2 \leq |A(v, v)| \quad \forall v \in H_0^1(\Omega).$$

Allowing $C_1 = 1$ and $C_2 = 1 + k^2$ it is obvious to see that Helmholtz equation satisfies the Gårding inequality. The argument of Schatz is as follows: Let $u \in H_0^1(\Omega)$ be the exact solution of the Helmholtz equation and $u_h \in V_h$ be the discrete solution given by the standard Galerkin method. From the Gårding inequality stated above it holds that

$$\|u - u_h\|_{1,\Omega}^2 - (1 + k^2) \|u - u_h\|_{0,\Omega}^2 \leq |A(u - u_h, u - u_h)|. \quad (1.17)$$

Using Galerkin Orthogonality and the continuity estimate given above, gives

$$|A(u - u_h, u - u_h)| = |A(u - u_h, u)| \quad (1.18)$$

$$\leq M \|u - u_h\|_{1,\Omega} \|u\|_{1,\Omega}. \quad (1.19)$$

Combining the previous two results and dividing through by $\|u - u_h\|_{1,\Omega}$ yields

$$\|u - u_h\|_{1,\Omega} - (1 + k^2) \frac{\|u - u_h\|_{0,\Omega}^2}{\|u - u_h\|_{1,\Omega}} \leq M \|u\|_{1,\Omega}. \quad (1.20)$$

Noticing that $\frac{\|u - u_h\|_{0,\Omega}}{\|u - u_h\|_{1,\Omega}} \leq 1$ gives the following estimate

$$\|u - u_h\|_{1,\Omega} - (1 + k^2) \|u - u_h\|_{0,\Omega} \leq M \|u\|_{1,\Omega}. \quad (1.21)$$

Now using Nitsche's trick it is possible to derive an estimate of the form

$$\|u - u_h\|_{0,\Omega} \leq Ch \|u - u_h\|_{1,\Omega}. \quad (1.22)$$

where h defines the mesh parameter. Using (1.21) in conjunction with (1.22) gives existence and uniqueness of discrete solutions again under the condition $h < \frac{C}{(1+k^2)}$. The argument uses the fact that under the condition $h < \frac{C}{(1+k^2)}$, $u = 0$ implies that $u_h = 0$. This asserts that the homogeneous equation has a unique solution. Then using the fact that V_h is finite dimensional implies that u_h exists and is unique for each $u \in H_0^1(\Omega)$. The argument, leads to the same condition as the inf-sup condition, ie $(1 + k^2)h < C$. As stated earlier this condition becomes too computationally restrictive for large k .

In practice one would like a method that works under the condition $kh < C$, meaning that regardless of the wavenumber a fixed number of points per wavelength is sufficient to achieve a certain accuracy. The problem with the standard FEM is that if only this condition is assumed to hold, stability is lost under increasing wavenumber. This loss of stability causes a couple of problems. The first is that existence and uniqueness of discrete solutions is no longer guaranteed and the second is that if a discrete solution does exist the method may not converge to that solution at an optimal rate. Numerous schemes have been devised to help alleviate these issues which have worked to great effect for the 1-D problem. In [34] Harari and Hughes introduced a Galerkin Least Squares (GLS) stabilization to improve the stability properties of the scheme which completely eliminates pollution in the 1-D case for an appropriate choice of penalty parameter. However, in 2-D this parameter is highly dependent on the angle of propagation of the wave which is not generally known a priori. This scheme was shown to be coercive for piecewise affine elements but the analysis does not carry to higher order elements. In [31] Feng and Wu applied a discontinuous Galerkin method with penalty on the jump of the solution and its derivatives over element faces to the problem. Since then Wu and Du have extended this idea in [28] to penalise the jumps of higher order terms for higher order methods using continuous finite element spaces. The CIP method was analysed in the 1-D case by Zhu, Burman and Wu in [22] and shown, for piecewise affine elements, to be absolutely stable. In this paper it is also shown, through a dispersion analysis, that the pollution effect can be eliminated for an appropriate choice of stabilization parameter. This was partly the motivation for Chapter 2, where I provide an abstract analysis for robust stabilized finite element methods which perform no worse than the standard FEM but are stable under the condition $kh < C$ for an appropriate choice of stabilization parameters. I present improved a priori error estimates of the same type as in [22] for piecewise linear elements for two stabilized methods, then explore the potential of these method to reduce the pollution effect numerically.

The focus of the first part of this project is to develop a stabilized FEM that is robust under the condition $kh < C$. This condition is the minimum requirement since waves are typically hard to discretize due to the large number of unknowns required to obtain an accurately resolved wavelength. The explanation comes from the Nyquist sampling theorem which when applied to Helmholtz equation states that in order to fully resolve each wave it is necessary to use a minimum of $10 \sim 12$ elements per wavelength. The FEM experiences two types of difficulties when $(1 + k^2)h$ is large, even if $kh < C$, firstly the linear system becomes ill-conditioned and could in principle have zero eigenvalues. Secondly, the finite element solution suffers from the so called “pollution effect”, i.e. the dispersion error is so large that the optimal approximation property of the FEM no longer holds in the pre asymptotic regime. These problems can be avoided by using higher order elements although this introduces its own issues. Increasing the polynomial order of the elements in your FE space has a negative impact on the conditioning of the system matrix and requires the use of an appropriate preconditioner. In practice these preconditioners

are usually low order methods which suffer from the problems discussed previously. Also for high order methods to be considered cost effective they must have superior convergence to low order methods, this requires additional regularity assumptions which are not always satisfied. The goal of this project is to ultimately develop a geometrically unfitted finite element method for the solution of Helmholtz equation which uses cut elements. These methods are known to only converge optimally for piecewise linear elements unless certain assumption are made about the interface. This motivates the need for a reliable low order method for solving the Helmholtz equation.

Once robust methods have been analysed the next step in the project is to provide a way of coupling two or more Helmholtz systems, using the technique of Nitsche introduced in [47] to weakly impose the coupling (or transmission) conditions across the interface. Chapter 3 introduces stabilized finite element methods for generalized boundary conditions using the technique of Nitsche to weakly enforce Dirichlet boundary conditions. The methods can be seen to enter the same theoretical framework introduced in Chapter 2. The final part of Chapter 3 introduces a new fictitious domain method using cut elements for the solution of Helmholtz equation. Classical fictitious domain methods are typically sub-optimal with poor accuracy and very few papers exist that consider the Helmholtz equation in this setting. The new method proposed is shown to be optimally convergent in the asymptotic regime and shown to be stable for an appropriate choice of stabilization parameters. The theoretical results are backed up numerically.

In the final chapter new stabilized fitted and unfitted domain decomposition methods are presented which can also be shown to enter a similar theoretical framework as proposed in Chapter 2. The multiple domain case highlights the need for robust methods as, in the case of coupled systems with different wave numbers, mesh restraints become harder to satisfy. The condition for the standard FEM becomes $\max\{k_i\}^2 h < C$ which may be too restrictive if each sub-domain has vastly different wave numbers. Problems of this type could typically arise in ocean acoustics where the water has a different density to the seabed. This causes waves to travel at different speeds through each medium and results in different wave numbers. It can be seen numerically that the newly proposed stabilized domain decomposition methods have advantages over the non-stabilized coupling techniques also proposed.

Chapter 2

Stabilized FEM for Wave equations

2.1 The Helmholtz equation

The numerical solution of the Helmholtz problem is challenging due to the indefinite character of the equation in weak form. In particular for high wave numbers the computational cost may be prohibitive due to the pollution effect. It is well known that the standard Galerkin discretization is not stable unless the condition k^2h small enough is satisfied [49, 38]. Indeed this condition has to be satisfied both to ensure that the system matrix is invertible and to ensure that optimal convergence is obtained. In the high wave number regime, optimal convergence is typically obtained only for k^2h small enough. This behaviour improves when high order approximation spaces are used and the dissipation relations have been carefully unearthed [1, 43]. The engineering rule of thumb on the other hand is that kh should be small enough, i.e. each wavelength should be resolved on a sufficiently large number of mesh cells, that ideally could be kept constant as the wave number increases. A remedy for the stability problems of the standard Galerkin method was suggested by Harari and Hughes in [34]. The idea was to stabilize the Helmholtz equation using the addition of a least squares perturbation similar to the SUPG method that had been so successful for the stabilization of convection-dominated flow problems. In [34] it was proven that for piecewise affine finite elements the Galerkin Least Squares (GLS) bilinear form, is coercive and hence the discrete problem is stable. This analysis however does not carry over to higher polynomial approximation. Using a multiscale argument Oberai and Pinsky [48] argued that a penalty on the normal gradient should be included in the formulation. Similar ideas were then used in combination with the discontinuous Galerkin (DG) method by Feng and Wu in order to stabilize the DG formulation of the Helmholtz equation [31]. Here a penalty was introduced on all the derivatives and it was shown that the discrete system is well posed without conditions on the mesh-parameter and the wave number. In recent work by Wu [54] the Continuous Interior Penalty (CIP) method, introduced by Douglas and Dupont [27] for advection dominated flows and later analysed by Burman and Hansbo [18], was used for the stabilization of the Helmholtz equation. The CIP-method has also been applied successfully to the

stabilization of different problems in elasticity [29, 19]. A very detailed analysis was carried out by Wu and co-workers in the papers [54, 57, 22], showing once again that the discrete system is well posed and that in one space dimension there is an interval in which the penalty parameter can be chosen so that the pollution effect is eliminated. In this chapter I propose two stabilized FEMs that have similar properties from the point of view of analysis. Drawing first on the ideas of [34, 48, 54] a GLS-type stabilization both on the element residual, and the jump of the normal gradient over element faces is proposed.

I then show that the least squares term in the interior of the elements may be replaced by a face oriented penalty term on the jump of the Laplacian, leading to a method similar to that proposed by Du and Wu [28] in the case of quadratic polynomials. I show that both the GLS method and the CIP method with penalty on jumps of first and second order derivatives are stable for all polynomial orders under the condition $kh < C$, where C is dependent on the choice of stabilization parameters, and that a priori error estimates of the same order as the standard Galerkin method under the condition $k^2h < C$ hold. In the case of piecewise linear elements I then prove an improved estimate for piecewise linear elements which shows that, although the estimates contain the standard pollution term, the error is upper bounded by data. As in [54] numerical evidence suggests that the method has some potential to reduce the pollution effect for a given problem if a good value of the stabilization parameter is known.

Consider the complex Helmholtz equation in $\Omega \subset \mathbb{R}^d$.

$$\left. \begin{aligned} \mathcal{L}(u) &= f & \text{in } \Omega \\ \mathcal{R}(u) &= g & \text{on } \partial\Omega \end{aligned} \right\} \quad (2.1)$$

where I have used $\mathcal{L}(\cdot) = -\Delta(\cdot) - k^2(\cdot)$ to denote the Helmholtz operator and $\mathcal{R}(\cdot) = \nabla(\cdot) \cdot n + ik(\cdot)$ to denote the Robin operator. The Robin operator, when stated in this way, can be thought of as a first order approximation of the Sommerfeld radiation condition which guarantees uniqueness of the Helmholtz equation. For the subsequent analysis it is useful to also consider the adjoint problem

$$\left. \begin{aligned} \mathcal{L}(z) &= \psi & \text{in } \Omega \\ \mathcal{R}^*(z) &= 0 & \text{on } \partial\Omega \end{aligned} \right\} \quad (2.2)$$

where $\mathcal{R}^*(\cdot) = \nabla(\cdot) \cdot n - ik(\cdot)$ denotes the Robin boundary operator for the adjoint problem. The adjoint problem comes from

$$(\mathcal{L}u, v)_{0,\Omega} = (u, \mathcal{L}^*v)_{0,\Omega}.$$

Letting \mathcal{L} denote the Helmholtz operator, choosing $g = 0$ and performing an integration by parts gives

$$(\mathcal{L}u, v)_{0,\Omega} = (-\Delta u - k^2 u, v)_{0,\Omega} \quad (2.3)$$

$$= (\nabla u, \nabla v)_{0,\Omega} - k^2 (u, v)_{0,\Omega} + ik \langle u, v \rangle_{0,\partial\Omega} \quad (2.4)$$

$$= (u, -\Delta v)_{0,\Omega} - k^2 (u, v)_{0,\Omega} + \underbrace{ik \langle u, v \rangle_{0,\partial\Omega} + \langle u, \nabla v \cdot n \rangle_{0,\partial\Omega}}_{(I)}. \quad (2.5)$$

Since $\langle u, v \rangle_{0,\partial\Omega} := \int_{\partial\Omega} u \bar{v}$ is a complex inner product the boundary terms (I) simplify to

$$(I) = \int_{\partial\Omega} u \nabla(\bar{v}) \cdot n + ik \int_{\partial\Omega} u \bar{v} \quad (2.6)$$

$$= \int_{\partial\Omega} u (\nabla(a - ib) \cdot n + ik(a - ib)) \quad (2.7)$$

$$= \int_{\partial\Omega} u \underbrace{(\nabla a \cdot n + kb - i\nabla b \cdot n + ika)}_{(w)}. \quad (2.8)$$

Taking the complex conjugate of w gives

$$w = \nabla a \cdot n + kb + i\nabla b \cdot n - ika \quad (2.9)$$

$$= \nabla(a + ib) \cdot n - ik(a + ib). \quad (2.10)$$

Therefore, the problem reduces to

$$(\mathcal{L}u, v)_{0,\Omega} = (u, \mathcal{L}^*v)_{0,\Omega} + \langle u, \nabla v \cdot n - ikv \rangle_{0,\partial\Omega} \quad (2.11)$$

This implies that the boundary condition for the adjoint problem should be

$$\mathcal{R}^*(v) = \nabla v \cdot n - ikv = 0$$

Assuming Ω is a bounded convex domain, it is known that the solution of (2.1) satisfies the following bounds, see [42]

$$k\|u\|_{0,\Omega} + \|u\|_{1,\Omega} \lesssim \mathcal{C}_{f,g}, \quad \|u\|_{2,\Omega} \lesssim k\mathcal{C}_{f,g}, \quad (2.12)$$

where $\mathcal{C}_{f,g} = \|f\|_{0,\Omega} + \|g_R\|_{0,\partial\Omega} + \frac{1}{k}\|g_R\|_{1/2,\partial\Omega}$. Under the same assumptions the adjoint problem satisfies similar bounds

$$k\|z\|_{0,\Omega} + \|z\|_{1,\Omega} \lesssim \|\psi\|_{0,\Omega}, \quad \|z\|_{2,\Omega} \lesssim k\|\psi\|_{0,\Omega}. \quad (2.13)$$

It is possible, if one assumes more regularity of f, g_R , to extend the estimates given above to higher order

Sobolev spaces. A proof of this can be found in [28], the idea uses the standard regularity estimate for Laplacian with Neumann boundary condition as well as an induction argument. The result generalizes to

$$\|u\|_{s,\Omega} \lesssim k^{s-1} C_{s-2,f,g}, \quad (2.14)$$

where $C_{s-2,f,g} \stackrel{\text{def}}{=} \|f\|_{0,\Omega} + \|g\|_{0,\Gamma} + \sum_{j=0}^{s-2} k^{-(j+1)} \left(\|f\|_{j,\Omega} + \|g\|_{j+\frac{1}{2},\Gamma} \right)$.

With the standard regularity results stated I would like to introduce an abstract formulation to denote Helmholtz equation given in weak form. Using this formulation the problem becomes: find $u \in H^1(\Omega)$ such that

$$A(u, v) = L(v), \quad \forall v \in H^1(\Omega), \quad (2.15)$$

where

$$A(u, v) := (\nabla u, \nabla v)_{0,\Omega} - k^2(u, v)_{0,\Omega} + i(ku, v)_{0,\partial\Omega}$$

and

$$L(v) := (f, v)_{0,\Omega} + \langle g, v \rangle_{0,\partial\Omega}.$$

2.2 Stabilized finite element formulations

The finite element method has been employed to solve a wide range of PDEs since its inception in the 1950's. The ability of the method to provide a good approximation to the underlying problem relies on two main properties. The first is that the discrete space V_h , in which solutions are sought, has reasonable approximation properties to the space V where solutions exist. The second property is that the problem can be approximated in a proper mathematical setting which guarantees consistency, stability and continuity. Often when issues arise in practice it can be attributed to one leaving this mathematical setting. The stabilized finite element methods that I will be introducing here are designed to be consistent with the underlying problem and also have additional continuity properties to make them stable for appropriate mesh parameters. Before I introduce these methods I introduce the finite element space. Let \mathcal{T}_h be a quasi-uniform triangulation of Ω where quasi-uniformity is as defined in (A.4) and let V_h be defined by

$$V_h \stackrel{\text{def}}{=} \{v \in H^1(\Omega) : v|_{\tau} \in \mathbb{P}_p(\tau) \forall \tau \in \mathcal{T}_h, p \in \mathbb{N}\},$$

where $\mathbb{P}_p(\tau)$ denotes the space of polynomials of order at most p when restricted to τ . Since the two stabilized methods I will be introducing each include at least one penalty term acting on the jump of some quantity over internal element faces it is also beneficial to define \mathcal{F}_{int} as the set of all interior faces,

$$\mathcal{F}_{int} \stackrel{\text{def}}{=} \{F : F = \tau_1 \cap \tau_2 \neq \emptyset \text{ where } \tau_1, \tau_2 \in \mathcal{T}_h\}.$$

Since I would like to analyse a class of stabilized FEMs which have similar properties, there is an advantage to introducing an abstract formulation. Using this abstract formulation the problem becomes:

Find $u_h \in V_h$ such that

$$A_h(u_h, v_h) = L_h(v_h) \quad \forall v_h \in V_h \quad (2.16)$$

where

$$A_h(u_h, v_h) = A(u_h, v_h) + s(u_h, v_h) \quad \text{and} \quad L_h(v_h) = L(v_h) + s(u, v_h).$$

Observe that since $s(u, v_h)$ appears in the right hand side of (2.16) this quantity must be known. This is typically the case if $s(\cdot, \cdot)$ is a residual based stabilization. $s(\cdot, \cdot)$ is assumed to be a sesquilinear form that may have a dependence on the mesh parameter h .

It is useful to note for the subsequent analysis that the abstract formulation stated in (2.16) obeys a Galerkin Orthogonality type relationship.

Lemma 2.1 (Galerkin Orthogonality). *Let u be the exact solution of (2.1) and u_h be the discrete solution of (2.16). Then*

$$A(u - u_h, v_h) = -s(u - u_h, v_h) \quad \forall v_h \in V_h. \quad (2.17)$$

Proof. The proof is standard

$$A(u, v_h) - A_h(u_h, v_h) = A(u, v_h) - A(u_h, v_h) - s(u_h, v_h) \quad (2.18)$$

$$= A(u - u_h, v_h) - s(u_h, v_h) \quad (2.19)$$

$$= L(v_h) - L(v_h) - s(u, v_h) \quad (2.20)$$

$$= -s(u, v_h) \quad (2.21)$$

which gives

$$A(u - u_h, v_h) - s(u_h, v_h) = -s(u, v_h). \quad (2.22)$$

The claim follows immediately. □

2.3 A priori error estimates and well-posedness of the discrete system

The main theorem of this section is a result about the well-posedness of the discrete system, the result relies upon a set of assumptions on the stabilization $s(\cdot, \cdot)$ and the discrete form $A_h(\cdot, \cdot)$ introduced previously. I show that under these assumptions a priori error estimates can be derived without making assumptions on k or h . This guarantees the discrete system is well-posed and also shows that under the condition $k^2 h < C$ the method converges optimally. To enter the framework of the theorem the

stabilization must be chosen appropriately. The first assumption on the stabilization is that there exists some type of energy semi-norm $|\cdot|_{\mathcal{J}}$ with the following properties $|\cdot|_{\mathcal{J}} : V_h \mapsto \mathbb{R}$ such that

$$C|u_h|_{\mathcal{J}}^2 \leq |A_h(u_h, u_h)|, \text{ for some } C > 0. \quad (2.23)$$

where u_h be the discrete solution of (2.16). It can be noted that if $|\cdot|_{\mathcal{J}}$ is taken to be a norm on V then this estimate is analogous to the coercivity assumption of Lax-Milgram. I also ask that the following Cauchy-Schwarz type inequality holds for some semi-norm $|\cdot|_s : V \mapsto \mathbb{R}$ which is associated to the stabilization,

$$|s(v, w)| \leq C|v|_{\mathcal{J}}|w|_s \leq |v|_{\mathcal{J}}|w|_{\mathcal{J}} \quad \forall v, w \in V. \quad (2.24)$$

It will become clear later that for the stabilizations I introduce these results hold by introducing stabilization parameters with non-zero complex components. The Cauchy-Schwarz type inequality ensures that the whole stabilization can be upper bounded by its Imaginary part.

The subsequent analysis also relies on the existence of an appropriate interpolation operator $\pi_h : V \mapsto V_h$. This operator should adhere to the standard interpolation estimates stated for piecewise linear elements

$$\|u - \pi_h u\|_{0,\Omega} + h\|\nabla(u - \pi_h u)\|_{0,\Omega} \leq Ch^2|u|_{2,\Omega}, \quad (2.25)$$

and

$$\|u - \pi_h u\|_{0,\Omega} + h\|\nabla(u - \pi_h u)\|_{0,\Omega} + \sum_{\tau \in \mathcal{T}_h} h_{\tau}^2 \|D^2(u - \pi_h u)\|_{0,\tau} \leq Ch^{p+1}|u|_{p+1,\Omega}, \quad (2.26)$$

for piecewise polynomial elements of order $p > 1$. The operator should also be chosen in conjunction with a norm $\|\cdot\|_*$ such that the following continuity estimate is satisfied

$$|A(u - \pi_h u, v_h)| + |s(u - \pi_h u, v_h)| + 2k\|u - \pi_h u\|_{0,\partial\Omega_R}\|v_h\|_{0,\partial\Omega_R} \leq M\|u - \pi_h u\|_*|v_h|_{\mathcal{J}}. \quad (2.27)$$

A second continuity assumption is posed purely on the sesquilinear form $A(\cdot, \cdot)$ and is designed to include the interpolation error of the adjoint problem. In this estimate z and ψ are taken to be the exact solution and source term of the adjoint problem as posed in (2.2), respectively.

$$|A(u - \pi_h u, z - \pi_h z)| \leq C(kh)h^p|u|_{p+1,\Omega}\|\psi\|_{0,\Omega}. \quad (2.28)$$

For the analysis to hold some control is needed over $\pi_h z$. This control is asserted through a stability estimate of the form

$$|\pi_h z|_s \leq Ckh\|\psi\|_{0,\Omega}. \quad (2.29)$$

which is required to show that the discrete solution converges to the exact solution in $L^2(\Omega)$. Finally the theorem requires interpolation estimates in $|\cdot|_{\mathcal{J}}$ and $\|\cdot\|_*$. Notice that these results highlight the similarity of the assumptions on the sesquilinear form.

$$\|u - \pi_h u\|_* \leq Ch^p |u|_{p+1, \Omega} \quad (2.30)$$

and

$$|u - \pi_h u|_{\mathcal{J}} \leq Ch^p |u|_{p+1, \Omega} \quad (2.31)$$

Now that the groundwork has been laid it is possible to construct our theorem

Theorem 2. *Let $u \in H^{p+1}(\Omega)$ be the unique solution of (2.15) and $u_h \in V_h$ be a solution of (2.16). If (2.16) satisfies properties (2.23)-(2.31) and $k > 1$ then the stabilized formulation admits a unique solution which satisfies the following a priori error estimates*

$$|u - u_h|_{\mathcal{J}} \leq Ch^p |u|_{p+1, \Omega} \quad (2.32)$$

$$\|u - u_h\|_{0, \Omega} \leq C(kh)h^p |u|_{p+1, \Omega} \quad (2.33)$$

$$\|\nabla(u - u_h)\|_{0, \Omega} \leq C(1 + (k^2 h)) h^p |u|_{p+1, \Omega}. \quad (2.34)$$

Considering the case $u \in H^2(\Omega)$ the following estimates hold

$$|u - u_h|_{\mathcal{J}} \leq Ckh\mathcal{C}_{f,g} \quad (2.35)$$

$$\|u - u_h\|_{0, \Omega} \leq C(kh)^2\mathcal{C}_{f,g} \quad (2.36)$$

$$\|\nabla(u - u_h)\|_{0, \Omega} \leq C(kh + k^3 h^2)\mathcal{C}_{f,g}. \quad (2.37)$$

To begin the proof of the theorem I first consider the convergence of $|u - u_h|_{\mathcal{J}}$.

Lemma 2.2 (Convergence of $|u - u_h|_{\mathcal{J}}$). *Under the assumptions of Theorem 2 the following convergence result holds*

$$|u - u_h|_{\mathcal{J}} \leq Ch^p |u|_{p+1, \Omega}. \quad (2.38)$$

Proof. Let $u - u_h = \underbrace{u - \pi_h u}_{\eta} + \underbrace{\pi_h u - u_h}_{\xi_h}$. From the triangle inequality and (2.31) it is clear that:

$$|u - u_h|_{\mathcal{J}} \leq C(|\eta|_{\mathcal{J}} + |\xi_h|_{\mathcal{J}}).$$

Turning our attention to $|\xi_h|_{\mathcal{J}}^2$ and using Galerkin Orthogonality followed by (2.27) reveals that

$$\begin{aligned}
C|\xi_h|_{\mathcal{J}}^2 &\leq |A_h(\xi_h, \xi_h)| \\
&= |A_h(\pi_h u, \xi_h) - (L(\xi_h) + s(u, \xi_h))| \\
&= |A_h(\eta, \xi_h)| \\
&\leq M\|\eta\|_* |\xi_h|_{\mathcal{J}},
\end{aligned}$$

dividing both sides by $|\xi_h|_{\mathcal{J}}$ and applying (2.30) yields the desired result,

$$|u - u_h|_{\mathcal{J}} \leq Ch^p |u|_{p+1, \Omega}.$$

□

With this result in hand it is now possible to estimate the rate of convergence in $L^2(\Omega)$. In order to obtain this result I have used a modified duality argument that uses the Galerkin orthogonality property stated in (2.17). The result also relies on the fact that

$$|\xi_h|_{\mathcal{J}} \leq Ch^p |u|_{p+1, \Omega}, \quad (2.39)$$

which comes directly from the proof of Lemma (2.2).

Lemma 2.3 (Convergence of $\|u - u_h\|_{0, \Omega}$). *Under the assumptions of Theorem 2 the following convergence result holds*

$$\|u - u_h\|_{0, \Omega} \leq C(kh)h^p |u|_{p+1, \Omega}. \quad (2.40)$$

Proof. The proof of convergence in the L^2 norm makes use of the adjoint problem given in (2.2). Let $z \in H^2(\Omega)$ and $\psi \in L^2(\Omega)$ be the exact solution and source term from (2.2), respectively. It follows using a modified duality argument and the Galerkin Orthogonality of the discrete sesquilinear form that

$$\begin{aligned}
(u - u_h, \psi)_{0, \Omega} &= A(u - u_h, z) \\
&= A(u - u_h, z - \pi_h z) - s(u - u_h, \pi_h z) \\
&\leq |A(u - u_h, z - \pi_h z) - s(u - u_h, \pi_h z)|.
\end{aligned}$$

Then using the same trick as before let $u - u_h = \eta + \xi_h$ the problem simplifies to

$$\begin{aligned}
(u - u_h, \psi)_{0, \Omega} &\leq |A(\eta, z - \pi_h z)| + |A(\xi_h, z - \pi_h z)| + |s(\eta, \pi_h z)| + |s(\xi_h, \pi_h z)| \\
&= I + II + III + IV.
\end{aligned}$$

It follows directly from (2.28) that

$$I \leq C(kh)h^p |u|_{p+1} \|\psi\|_{0,\Omega}.$$

II relies on the fact that

$$|A(\xi_h, z - \pi_h z)| = |\overline{A(\xi_h, z - \pi_h z)}|.$$

Expanding this gives

$$|\overline{A(\xi_h, z - \pi_h z)}| = |(\nabla(z - \pi_h z), \nabla \xi_h)_{0,\Omega} - k^2(z - \pi_h z, \xi_h)_{0,\Omega}| \quad (2.41)$$

$$- ik \langle z - \pi_h z, \xi_h \rangle_{0,\partial\Omega_R} | \quad (2.42)$$

$$= |A(z - \pi_h z, \xi_h) - 2ik \langle z - \pi_h z, \xi_h \rangle_{0,\partial\Omega_R}| \quad (2.43)$$

$$\leq |A(z - \pi_h z, \xi_h)| + |2k \langle z - \pi_h z, \xi_h \rangle_{0,\partial\Omega_R}|. \quad (2.44)$$

Which implies from (2.27), (2.30) and Lemma (2.2)

$$II \leq M |\xi_h|_{\mathcal{I}} \|z - \pi_h z\|_* \quad (2.45)$$

$$\leq Ckh (h^p |u|_{p+1,\Omega}) \|\psi\|_{0,\Omega}. \quad (2.46)$$

An application of the Cauchy Schwarz type inequality (2.24) to terms III and IV gives

$$III + IV \leq C|\eta|_{\mathcal{I}} |\pi_h z|_s + C|\xi_h|_{\mathcal{I}} |\pi_h z|_s \quad (2.47)$$

$$= C(|\eta|_{\mathcal{I}} + |\xi_h|_{\mathcal{I}}) |\pi_h z|_s. \quad (2.48)$$

Using the interpolation estimate (2.31) on the first term and the stability estimate (2.29) on the third term reduces to

$$III + IV \leq C(h^p |u|_{p+1,\Omega} + |\xi_h|_{\mathcal{I}}) kh \|\psi\|_{0,\Omega}. \quad (2.49)$$

Finally applying (2.39) to the second term gives the desired result

$$III + IV \leq C(h^p |u|_{p+1,\Omega} + h^p |u|_{p+1,\Omega}) kh \|\psi\|_{0,\Omega} \quad (2.50)$$

$$= Ckh (h^p |u|_{p+1,\Omega}) \|\psi\|_{0,\Omega}. \quad (2.51)$$

□

Lemma 2.4. *The stabilized form $A_h(\cdot, \cdot)$ defined in (2.16) satisfies a Gårding inequality of the form*

$$\|\nabla \xi_h\|_{0,\Omega}^2 \leq k^2 \|\xi_h\|_{0,\Omega}^2 + 2|A_h(\xi_h, \xi_h)|. \quad (2.52)$$

Proof. The weak formulation (2.15) shows that

$$\|\nabla \xi_h\|_{0,\Omega}^2 = k^2 \|\xi_h\|_{0,\Omega}^2 + \operatorname{Re} [A(\xi_h, \xi_h)].$$

Adding and subtracting $\operatorname{Re} [s(\xi_h, \xi_h)]$ gives

$$\|\nabla \xi_h\|_{0,\Omega}^2 = k^2 \|\xi_h\|_{0,\Omega}^2 + \operatorname{Re} [A_h(\xi_h, \xi_h)] - \operatorname{Re} [s(\xi_h, \xi_h)].$$

From (2.24)

$$\|\nabla \xi_h\|_{0,\Omega}^2 \leq k^2 \|\xi_h\|_{0,\Omega}^2 + |A_h(\xi_h, \xi_h)| + |s(\xi_h, \xi_h)| \quad (2.53)$$

$$\leq k^2 \|\xi_h\|_{0,\Omega}^2 + |A_h(\xi_h, \xi_h)| + |\xi_h|_3^2, \quad (2.54)$$

and since $|A_h(\xi_h, \xi_h)| \geq C|\xi_h|_3^2$ the Gårding inequality in the lemma holds. \square

Lemma 2.5 (Convergence of $\|\nabla(u - u_h)\|_{0,\Omega}$). *Let $u \in H^{p+1}(\Omega)$, for $p > 1/2$ be the exact solution of (2.1). It holds that the numerical solution $u_h \in V_h$, given by a stabilized method which conforms to the assumptions made in Theorem 2, satisfies the following a priori error estimate*

$$\|\nabla(u - u_h)\|_{0,\Omega} \leq C(1 + k^2 h) h^p |u|_{p+1,\Omega}. \quad (2.55)$$

Proof. From the triangle inequality

$$\|\nabla(u - u_h)\|_{0,\Omega} \leq \|\nabla \eta\|_{0,\Omega} + \|\nabla \xi_h\|_{0,\Omega}.$$

where η and ξ_h are as defined previously. From (2.25) or (2.26), $\|\nabla \eta\|_{0,\Omega} \leq Ch^p |u|_{p+1,\Omega}$ therefore it is enough to show that

$$\|\nabla \xi_h\|_{0,\Omega} \leq C(1 + k^2 h) h^p |u|_{p+1,\Omega}.$$

Using the Gårding inequality (2.52) it holds that

$$\|\nabla \xi_h\|_{0,\Omega}^2 \leq k^2 \|\xi_h\|_{0,\Omega}^2 + 2|A_h(\xi_h, \xi_h)|.$$

The first term is bounded using the triangle inequality followed by Lemma 2.3 and (2.25) or (2.26)

$$\|\xi_h\|_{0,\Omega} \leq \|u - u_h\|_{0,\Omega} + \|\eta\|_{0,\Omega} \quad (2.56)$$

$$\leq C((kh) + h)h^p|u|_{p+1,\Omega} \quad (2.57)$$

$$\leq C(kh)h^p|u|_{p+1,\Omega}. \quad (2.58)$$

The last line comes from the assumption that $k > 1$. The second term is handled in the same way as in the proof of (2.2).

$$|A_h(\xi_h, \xi_h)| \leq M\|\eta\|_*|\xi_h|_{\mathcal{D}} \quad (2.59)$$

$$\leq Ch^{2p}|u|_{p+1,\Omega}^2. \quad (2.60)$$

The result then follows

$$\|\nabla(u - u_h)\|_{0,\Omega} \leq C(1 + k^2h)h^p|u|_{p+1,\Omega}.$$

□

The well-posedness of the discrete system follows directly from this error estimate noticing that

$$\|u - u_h\|_{1,\Omega} \leq C(1 + kh + k^2h)\|u\|_{1,\Omega}. \quad (2.61)$$

Using the argument of Schatz [49], let u_h be the solution of (2.16) and $u = 0$ the solution of (2.1), then (2.61) implies that $u_h = 0$. This implies that the homogeneous equation has a unique solution and since V_h is finite dimensional u_h exists and is unique for all $u \in H^1(\Omega)$.

2.3.1 Galerkin Least Squares method with interior penalty on the normal gradient

The first method to be analysed is an adaptation of the popular Galerkin Least Squares (GLS) method. The GLS method has been popularized by Hughes et al. and has been used to great effect in the approximation of advection-diffusion equations. In [34] Harari and Hughes introduced the GLS method for the numerical solution of Helmholtz equation. It was shown that the method is stable for piecewise linear elements however this analysis does not carry over to higher order elements. The method was introduced to improve the stability of the formulation and was shown to greatly reduce the pollution error in 1D. In fact it can be shown, using dispersion analysis, that the GLS method can eliminate the pollution term in the 1D problem for an appropriate choice of stabilization parameter. Unfortunately, this is not the case for problems in higher dimensions as the stabilization parameter is strongly dependent on the wave direction. In fact it can be seen numerically that if the underlying solution does not have a dominant

direction then the GLS has similar performance to the standard Galerkin method for the approximation of Helmholtz equation.

From this it is clear that the GLS method, on its own, is fairly limited in its ability to solve Helmholtz in \mathbb{R}^d for $d > 1$. In [48] Oberai and Pinsky added an additional stabilization term which took into account the residuals of inter-element boundaries in an attempt to overcome the problems experienced by the GLS method. The paper uses dispersion analysis to show that the method displays minimal phase error for bilinear quadrilateral finite elements, although does not look at the numerical stability of the method. Drawing inspiration from these papers I add a penalty term on the jump of the normal gradient over element faces and a least squares term giving additional control of the boundary condition to the standard GLS method proposed in [34]. This stabilization will be shown to satisfy the assumptions of Theorem 2. In addition to this the method can be shown to satisfy the following improved a priori error estimates.

$$\|k(u - u_h)\|_{0,\Omega} \leq C \min\{1, k^3 h^2\} \mathcal{C}_{f,g}, \quad (2.62)$$

$$\|\nabla(u - u_h)\|_{0,\Omega} \leq C(kh + \min\{1, k^3 h^2\}) \mathcal{C}_{f,g}, \quad (2.63)$$

for piecewise linear elements.

Introducing the method in the abstract formulation stated previously, the problem becomes: find $u_h \in V_h$ such that

$$A_h(u_h, v_h) = L_h(v_h) \quad \forall v_h \in V_h,$$

where

$$\begin{aligned} s(u_h, v_h) = & \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_{1,\tau} (\mathcal{L}(u_h), \mathcal{L}(v_h))_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} h_\tau \gamma_{1,\tau} \langle \llbracket \nabla u_h \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_{0,F} \\ & + \beta_1 h \langle \mathcal{R}(u_h), \mathcal{R}^*(v_h) \rangle_{0,\partial\Omega_R}, \end{aligned} \quad (2.64)$$

where $\delta_{1,\tau}, \gamma_{1,\tau}, \beta_1 \in \mathbb{C}$ and $\llbracket \cdot \rrbracket$ defines the jump over element faces such that for $\tau_1, \tau_2 \in \mathcal{T}_h$ with $\tau_1 \cap \tau_2 \neq \emptyset$

$$\llbracket \nabla v_h \rrbracket = \nabla v_h \cdot n|_{\tau_1} + \nabla v_h \cdot n|_{\tau_2}.$$

The first term given in the stabilization is referred to as the GLS term. Notice that I am allowed to take second order derivatives since u_h is C^∞ when considered element-wise. Summing over all elements then gives the GLS term as seen in the stabilization. The fact that the exact solution u must satisfy (2.1) on all subdomains of Ω is then used to make the term consistent.

$$\sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_{1,\tau} (\mathcal{L}(u), \mathcal{L}(v_h))_{0,\tau} = \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_{1,\tau} (f, \mathcal{L}(v_h))_{0,\tau}.$$

The second term in the stabilization is referred to as the CIP term. This term penalises the jump of the gradient over element faces. Since u is assumed to be smooth

$$\sum_{F \in \mathcal{F}_{int}} h_\tau \gamma_{1,\tau} \langle \llbracket \nabla u \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_{0,F} = 0.$$

The third and final term is a GLS term on the boundary and, in a similar manner to the first term, it holds that

$$\beta_1 h \langle \mathcal{R}(u), \mathcal{R}^*(v_h) \rangle_{\partial\Omega_R} = \beta_1 h \langle g_R, \mathcal{R}^*(v_h) \rangle_{0,\partial\Omega_R}.$$

Since the stabilization can be evaluated for the exact solution u it can be concluded that the method can be expressed using the abstract formulation (2.16).

To show that the formulation enters the mathematical framework of the theorem it is necessary to pick an appropriate operator $\pi_h : V \mapsto V_h$. In the following analysis I choose the standard Lagrange interpolant, a definition of which can be found in [30]

$$\pi_h \stackrel{\text{def}}{=} \mathcal{I}_h : C^0(\bar{\Omega}) \mapsto V_h.$$

It is shown in [30] that Assumptions (2.25) and (2.26) hold for this choice of π_h . The following lemma asserts the weak coercivity property and Cauchy-Schwarz type inequality hold for the stabilization for $|\cdot|_{\mathfrak{J}}^2 : V_h \mapsto \mathbb{R}$ defined as

$$\begin{aligned} |u_h|_{\mathfrak{J}}^2 &\stackrel{\text{def}}{=} k(1 - \text{Im}[\beta_1]hk) \|u_h\|_{0,\partial\Omega_R}^2 + \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \text{Im}[\delta_{1,\tau}] \|\mathcal{L}(u_h)\|_{0,\tau}^2 \\ &\quad + \sum_{F \in \mathcal{F}_{int}} h_\tau \text{Im}[\gamma_{1,\tau}] \|\llbracket \nabla u_h \rrbracket\|_{0,F}^2 + \text{Im}[\beta_1]h \|\nabla(u_h) \cdot n\|_{0,\partial\Omega_R}^2 \end{aligned} \quad (2.65)$$

Lemma 2.6. *Under the conditions $0 < \text{Im}[\beta_1]hk < 1$, $\text{Re}[\beta_1] = 0$ and given $\text{Im}[\delta_{1,\tau}]$, $\text{Im}[\gamma_{1,\tau}]$ and $\text{Im}[\beta_1] > 0, \forall \tau \in \mathcal{T}_h$ the form $|\cdot|_{\mathfrak{J}}$ given by (2.65) is a semi-norm on V_h , moreover the following identity holds*

$$|u_h|_{\mathfrak{J}}^2 = \text{Im}[A_h(u_h, u_h)]. \quad (2.66)$$

Proof. The initial claim follows by inspection. The identity follows by expanding $\text{Im}[A_h(u_h, u_h)]$

$$\begin{aligned} \text{Im}[A_h(u_h, u_h)] &= k(u_h, u_h)_{0,\partial\Omega_R} + \text{Im}[s(u_h, u_h)] \\ &= k \|u_h\|_{0,\partial\Omega_R}^2 + \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \text{Im}[\delta_{1,\tau}] \|\mathcal{L}(u_h)\|_{0,\tau}^2 + \sum_{F \in \mathcal{F}_{int}} h_\tau \text{Im}[\gamma_{1,\tau}] \|\llbracket \nabla u_h \rrbracket\|_{0,F}^2 \\ &\quad + \text{Im}\left[\beta_1 h \langle \mathcal{R}(u_h), \mathcal{R}^*(u_h) \rangle_{0,\partial\Omega_R}\right] \end{aligned}$$

Considering the final term gives

$$\begin{aligned} \operatorname{Im} \left[\beta_1 h \langle \mathcal{R}(u_h), \mathcal{R}^*(u_h) \rangle_{0, \partial\Omega_R} \right] &= \operatorname{Im} \left[\beta_1 h \left(\|\nabla(u_h) \cdot n\|_{0, \partial\Omega_R}^2 - k^2 \|u_h\|_{0, \partial\Omega_R}^2 \right. \right. \\ &\quad \left. \left. + \langle \nabla(u_h) \cdot n, ik u_h \rangle_{0, \partial\Omega_R} + \langle ik u_h, \nabla(u_h) \cdot n \rangle_{0, \partial\Omega_R} \right) \right] \end{aligned}$$

Since the last two terms are the complex conjugates of one another the expression simplifies to

$$\operatorname{Im} \left[\beta_1 h \langle \mathcal{R}(u_h), \mathcal{R}^*(u_h) \rangle_{0, \partial\Omega_R} \right] = \operatorname{Im}[\beta_1] h \left(\|\nabla(u_h) \cdot n\|_{0, \partial\Omega_R}^2 - k^2 \|u_h\|_{0, \partial\Omega_R}^2 \right)$$

With this result it is clear that under the conditions of the lemma, (2.23) holds. \square

The theorem also asks that the operator \mathcal{I}_h is stable with respect to $|\cdot|_s : V_h \mapsto \mathbb{R}$, which I define as

$$|\cdot|_s^2 \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} |\delta_{1, \tau}| h_\tau^2 \|\mathcal{L}(\cdot)\|_{0, \tau}^2 + \sum_{F \in \mathcal{F}_{int}} |\gamma_{1, \tau}| h_\tau \|\llbracket \nabla(\cdot) \cdot n \rrbracket\|_{0, F}^2 + |\beta_1| h \|\mathcal{R}^*(\cdot)\|_{0, \partial\Omega_R}^2, \quad (2.67)$$

and that the operator has the required interpolation properties in $|\cdot|_J$ and $\|\cdot\|_*$.

Lemma 2.7. *Under the conditions $hk < \operatorname{Im}[\beta_1^{-1}]$, $\operatorname{Re}[\beta_1] = 0$ and given $\operatorname{Im}[\delta_{1, \tau}]$, $\operatorname{Im}[\gamma_{1, \tau}]$ and $\operatorname{Im}[\beta_1] > 0$ where $\operatorname{Im}[\delta_{1, \tau}] > C \operatorname{Re}[\delta_{1, \tau}]$, $\operatorname{Im}[\gamma_{1, \tau}] > C \operatorname{Re}[\gamma_{1, \tau}] \forall \tau \in \mathcal{T}_h$ the GLS/CIP stabilization satisfies assumptions (2.23), (2.24), (2.29), (2.30) and (2.31) for*

$$\|\cdot\|_* \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} h_\tau^{-1} |\delta_{1, \tau}|^{-1/2} \|\cdot\|_{0, \tau} + \sum_{F \in \mathcal{F}_{int}} (|h_\tau \gamma_{1, \tau}|)^{-1/2} \|\cdot\|_{0, F} + (|h \beta_1|)^{-1/2} \|\cdot\|_{\partial\Omega_R} + |\cdot|_J,$$

and π_h denoted by \mathcal{I}_h as the standard Lagrange interpolant.

Proof. Let us begin our analysis by stating that (2.23) is clearly satisfied for an appropriate choice of stabilization parameters. Since

$$\begin{aligned} |u_h|_J^2 &= \operatorname{Im}[A_h(u_h, u_h)] \\ &\leq |A_h(u_h, u_h)|. \end{aligned}$$

Assumption (2.24) holds after a simple application of the Cauchy-Schwarz inequality followed by an

application of the triangle inequality.

$$\begin{aligned}
|s(v_h, w_h)| &= \left| \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_{1,\tau} (\mathcal{L}(v_h), \mathcal{L}(w_h))_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} h_\tau \gamma_{1,\tau} \langle \llbracket \nabla v_h \rrbracket, \llbracket \nabla w_h \rrbracket \rangle_{0,F} \right. \\
&\quad \left. + \beta_1 h \langle \mathcal{R}(v_h), \mathcal{R}^*(w_h) \rangle_{0,\partial\Omega_R} \right| \\
&\leq \sum_{\tau \in \mathcal{T}_h} h_\tau^2 |\delta_{1,\tau}| \|\mathcal{L}(v_h)\|_{0,\tau} \|\mathcal{L}(w_h)\|_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} h_\tau |\gamma_{1,\tau}| \|\llbracket \nabla v_h \rrbracket\|_{0,F} \|\llbracket \nabla w_h \rrbracket\|_{0,F} \\
&\quad + h |\beta_1| \|\mathcal{R}(v_h)\|_{0,\partial\Omega_R} \|\mathcal{R}^*(w_h)\|_{0,\partial\Omega_R} \\
&\leq \sum_{\tau \in \mathcal{T}_h} h_\tau^2 |\delta_{1,\tau}| \|\mathcal{L}(v_h)\|_{0,\tau} \|\mathcal{L}(w_h)\|_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} h_\tau |\gamma_{1,\tau}| \|\llbracket \nabla v_h \rrbracket\|_{0,F} \|\llbracket \nabla w_h \rrbracket\|_{0,F} \\
&\quad + h |\beta_1| (\|\nabla(v_h) \cdot n\|_{0,\partial\Omega_R} + k \|v_h\|_{0,\partial\Omega_R}) \|\mathcal{R}^*(w_h)\|_{0,\partial\Omega_R} \\
&\leq C |v_h|_{\mathfrak{J}} |w_h|_s.
\end{aligned}$$

To obtain the second inequality one simply applies the triangle inequality to the term $\|\mathcal{R}^*(w_h)\|_{0,\partial\Omega_R}$. Assumption (2.31) is the interpolation estimate in the \mathfrak{J} semi-norm. To show that this holds, let $\eta = u - \mathcal{I}_h u$ and $\beta = \text{Im}[\beta_1]$ then recall

$$\begin{aligned}
|\eta|_{\mathfrak{J}}^2 &= \underbrace{k(1 - \beta h k) \|\eta\|_{0,\partial\Omega_R}^2}_{(1a)} + \underbrace{\sum_{\tau \in \mathcal{T}_h} h_\tau^2 \text{Im}[\delta_{1,\tau}] \|\mathcal{L}(\eta)\|_{0,\tau}^2}_{(1b)} + \underbrace{\sum_{F \in \mathcal{F}_{int}} h_\tau \text{Im}[\gamma_{1,\tau}] \|\llbracket \nabla \eta \rrbracket\|_{0,F}^2}_{(1c)} \\
&\quad + \underbrace{\beta h \|\nabla \eta \cdot n\|_{0,\partial\Omega_R}^2}_{(1d)}.
\end{aligned}$$

Applying the Cauchy-Schwarz inequality and the trace inequality given in the Appendix (A:4)

$$(1a) \leq \sum_{\tau \in \mathcal{T}_h} C_T^2 k(1 - h k \beta) \left(h_\tau^{-1/2} \|\eta\|_{0,\tau} + h_\tau^{1/2} \|\nabla \eta\|_{0,\tau} \right)^2.$$

Letting $h = \max_{\tau \in \mathcal{T}_h} \{h_\tau\}$ and applying (2.25) or (2.26) implies

$$(1a) \leq C_T^2 (kh)(1 - kh\beta) (Ch^{2p} |u|_{p+1,\Omega}^2).$$

Since I have assumed $kh < C$ the required result follows. The proof of (1b), (1c) and (1d) follows in a similar manner. It is possible to use the linearity of the operator $\mathcal{L}(\cdot)$ followed by the triangle inequality to expand (1b) as such,

$$\begin{aligned}
(1b) &\leq \sum_{\tau \in \mathcal{T}_h} h_\tau^2 |\delta_{1,\tau}| \left[\|\Delta \eta\|_{0,\tau}^2 + 2 \|k^2 \eta\|_{0,\tau} \|\Delta \eta\|_{0,\tau} + \|k^2 \eta\|_{0,\tau}^2 \right] \\
&\leq C \max\{|\delta_{1,\tau}|\} (h^{2p} |u|_{p+1,\Omega}^2 + 2(kh)^2 h^{2p} |u|_{p+1,\Omega}^2 + (kh)^4 h^{2p} |u|_{p+1,\Omega}^2) \\
&= C \max\{|\delta_{1,\tau}|\} ((kh)^2 + 1)^2 h^{2p} |u|_{p+1,\Omega}^2.
\end{aligned}$$

Once expanded the result follows from applying the interpolation estimates given by (2.26) for elements of polynomial order greater than 1. The same result holds for piecewise linear polynomials after noting that the terms

$$\|\Delta(u - \mathcal{I}_h u)\|_{0,\tau} = \|\Delta u\|_{0,\tau} \leq C\|u\|_{2,\tau}.$$

(1c) is dealt with by a simple use of the trace inequality (A:4)

$$\begin{aligned} (1c) &\leq C \sum_{\tau \in \mathcal{T}_h} \sum_{F \in \partial\tau} h_\tau |\gamma_{1,\tau}| \|\nabla\eta \cdot n\|_{0,F}^2 \\ &\leq C \max\{|\gamma_{1,\tau}|\} h^{2p} |u|_{p+1,\Omega}^2. \end{aligned}$$

Then using the trace inequality and the interpolation estimates (2.25) and (2.26) gives the required result.

$$\begin{aligned} (1d) &\leq |\beta_1| h \sum_{F \subset \partial\Omega_R} [\|\nabla\eta \cdot n\|_{0,F}^2] \\ &\leq C |\beta_1| h^{2p} |u|_{p+1,\Omega}^2, \end{aligned}$$

which gives that

$$|\eta|_{\mathcal{J}} \leq C_{\mathcal{J},GLS} h^p |u|_{p+1,\Omega},$$

where

$$C_{\mathcal{J},GLS} \stackrel{\text{def}}{=} kh(1 - kh\beta_1) + C (\max\{|\delta_{1,\tau}|\}) ((kh)^2 + 1)^2 + \max\{|\gamma_{1,\tau}|\} + |\beta_1|)^{1/2}.$$

Assumption (2.29) asks that the operator \mathcal{I}_h is stable with respect to $|\cdot|_s$. It is possible to justify this claim after noting that for a sufficiently smooth z

$$|z - \pi_h z|_s \leq |z - \pi_h z|_{\mathcal{J}} \leq C_{\mathcal{J},GLS} h |z|_{2,\Omega}$$

which using the regularity estimate given in (2.13) implies that

$$|z - \pi_h z|_s^2 \leq C_{\mathcal{J},GLS}^2 h^2 |z|_{2,\Omega}^2 \tag{2.68}$$

$$|\pi_h z|_s^2 - |z|_s^2 \leq C_{\mathcal{J},GLS}^2 (hk)^2 \|\psi\|_{0,\Omega}^2 \tag{2.69}$$

$$|\pi_h z|_s^2 \leq C_{\mathcal{J},GLS}^2 (hk)^2 \|\psi\|_{0,\Omega}^2 + |z|_s^2. \tag{2.70}$$

Since z is the solution of the dual problem it is assumed to be smooth this infers the CIP terms in the stabilization vanish. Notice that the penalty on the Robin boundary condition is also zero from how the

problem is posed. Therefore the only term that remains is the GLS term. Which gives

$$|\pi_h z|_s^2 \leq C_{\mathfrak{J},GLS}^2 (hk)^2 \|\psi\|_{0,\Omega}^2 + \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_{1,\tau} \|\mathcal{L}(z)\|_{0,\tau}^2 \quad (2.71)$$

$$|\pi_h z|_s^2 \leq C_{\mathfrak{J},GLS}^2 (hk)^2 \|\psi\|_{0,\Omega}^2 + h^2 \max\{\delta_{1,\tau}\} \|\psi\|_{0,\Omega}^2 \quad (2.72)$$

$$|\pi_h z|_s \leq Chk \|\psi\|_{0,\Omega}. \quad (2.73)$$

It only remains to prove assumption (2.30), which is shown using the following argument.

$$\|\eta\|_*^2 = \underbrace{\sum_{\tau \in \mathcal{T}_h} h_\tau^{-2} |\delta_{1,\tau}|^{-1} \|\eta\|_{0,\tau}^2}_{(2a)} + \underbrace{\sum_{F \in \mathcal{F}_{int}} (h_\tau |\gamma_{1,\tau}|)^{-1} \|\eta\|_{0,F}^2}_{(2b)} + \underbrace{(\beta_1 h)^{-1} \|\eta\|_{0,\partial\Omega_R}^2}_{(2c)} + \underbrace{|\eta|_{\mathfrak{J}}^2}_{(2d)}.$$

(2a) is bounded using the interpolation estimates proposed in (2.25) or (2.26). Using the trace inequality (A:4) shows that (2b) satisfies the required bound.

$$(2b) \leq \sum_{\tau \in \mathcal{T}_h} (h_\tau |\gamma_{1,\tau}|)^{-1} \|\eta\|_{0,\partial\tau}^2 \quad (2.74)$$

$$\leq C \sum_{\tau \in \mathcal{T}_h} (h_\tau |\gamma_{1,\tau}|)^{-1} \left(h_\tau^{-1/2} \|\eta\|_{0,\tau} + h_\tau^{1/2} \|\nabla \eta\|_{0,\tau} \right)^2 \quad (2.75)$$

$$\leq C \sum_{\tau \in \mathcal{T}_h} |\gamma_{1,\tau}|^{-1} \left(h_\tau^{-1} \|\eta\|_{0,\tau} + \|\nabla \eta\|_{0,\tau} \right)^2 \quad (2.76)$$

$$\leq C \max\{|\gamma_{1,\tau}|\}^{-1} h^{2p} |u|_{p+1,\Omega}^2. \quad (2.77)$$

Using a similar argument (2c) can be shown to satisfy

$$(2c) \leq C |\beta_1|^{-1} h^{2p} |u|_{p+1,\Omega}^2. \quad (2.78)$$

Finally (2d) follows as shown previously. Keeping track of our constant gives

$$C_{*,GLS} = C \left(|\beta_1|^{-1/2} + |\delta_1|^{-1/2} + |\gamma_1|^{-1/2} + C_{\mathfrak{J},GLS} \right).$$

Since all of the assumptions in the lemma hold the proof is complete. \square

The other assumptions that are made are continuity results on the discrete form $A_h(\cdot, \cdot)$ and $A(\cdot, \cdot)$.

Proposition 1 (Continuity). *Let $u \in H^{p+1}(\Omega)$ and $\mathcal{I}_h : C^0(\bar{\Omega}) \mapsto V_h$ be the standard Lagrange interpolant then it follows that the GLS/CIP discrete sesquilinear form satisfies Assumptions (2.27) and*

(2.28). Let $\eta \stackrel{\text{def}}{=} u - \mathcal{I}_h u$ then,

$$\underbrace{|A(\eta, v_h)| + |s(\eta, v_h)|}_{(3)} + 2k\|\eta\|_{0,\partial\Omega}\|v_h\|_{0,\partial\Omega} \leq C\|\eta\|_*|v_h|_{\mathcal{J}} \quad \forall v_h \in V_h. \quad (2.79)$$

Proof. It is simple to show that the last term on the left hand side respects the continuity estimate therefore it is enough to analyse (3). Performing an element-by-element integration by parts on the first term in the formulation, gives

$$\begin{aligned} (3) &= \left| \sum_{\tau \in \mathcal{T}_h} (\eta, -\Delta v_h)_{0,\tau} - \sum_{\tau \in \mathcal{T}_h} k^2 (\eta, v_h)_{0,\tau} + \sum_{\tau \in \mathcal{T}_h} \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\tau} + ik \langle \eta, v_h \rangle_{0,\partial\Omega_R} \right| \\ &\quad + |s(\eta, v_h)| \\ &= \left| \sum_{\tau \in \mathcal{T}_h} (\eta, -\Delta v_h - k^2 v_h)_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} \langle \eta, \llbracket \nabla v_h \rrbracket \rangle_{0,F} + \langle \eta, \nabla v_h \cdot n + ik v_h \rangle_{0,\partial\Omega_R} \right| \\ &\quad + |s(\eta, v_h)|. \end{aligned}$$

Which after using the triangle and Cauchy-Schwarz inequalities implies that

$$(3) \leq \sum_{\tau \in \mathcal{T}_h} \|\eta\|_{0,\tau} \|\mathcal{L}(v_h)\|_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} \|\eta\|_{0,F} \|\llbracket \nabla v_h \rrbracket\|_{0,F} + \|\eta\|_{0,\partial\Omega_R} \|\mathcal{R}^*(v_h)\|_{0,\partial\Omega_R} + |\eta|_{\mathcal{J}} |v_h|_{\mathcal{J}}.$$

After noticing that the right hand norms in the first, second and fourth terms appear in $|v_h|_{\mathcal{J}}$ and the η terms appear in $\|\eta\|_*$, where $\|\cdot\|_*$ is defined as in (2.7) it is necessary to show the third term is controlled by $|\cdot|_{\mathcal{J}}$. This is shown by a simple application of the triangle inequality

$$\|\mathcal{R}^*(v_h)\|_{0,\partial\Omega_R} \leq \|\nabla v_h \cdot n\|_{0,\partial\Omega_R} + k\|v_h\|_{0,\partial\Omega_R} \quad (2.80)$$

$$\leq C|v_h|_{\mathcal{J}}. \quad (2.81)$$

with this result it follows that

$$|A_h(\eta, v_h)| \leq M\|\eta\|_*|v_h|_{\mathcal{J}}.$$

The final result for the discrete sesquilinear form (2.28) is to show an additional boundedness result.

$$\underbrace{|A(u - \mathcal{I}_h u, z - \mathcal{I}_h z)|}_{(*)} \leq Ch^p |u|_{p+1} \|\psi\|_{0,\Omega}.$$

Applying the triangle and Cauchy-Schwarz inequalities to the first term gives

$$\begin{aligned}
(*) &\leq \|\nabla(u - \mathcal{I}_h u)\|_{0,\Omega} \|\nabla(z - \mathcal{I}_h z)\|_{0,\Omega} + k^2 \|u - \mathcal{I}_h u\|_{0,\Omega} \|z - \mathcal{I}_h z\|_{0,\Omega} \\
&\quad + k \|u - \mathcal{I}_h u\|_{0,\partial\Omega} \|z - \mathcal{I}_h z\|_{0,\partial\Omega} \\
&\leq C(kh)h^p |u|_{p+1} \|\psi\|_{0,\Omega} + C(kh)^3 h^p |u|_{p+1} \|\psi\|_{0,\Omega} + C(kh)^2 h^p |u|_{p+1} \|\psi\|_{0,\Omega} \\
&\leq C(kh)h^p |u|_{p+1} \|\psi\|_{0,\Omega}.
\end{aligned}$$

I have used the trace inequality (A:4) to deal with the term on the boundary. The second line then follows by an application of our interpolation estimates (2.25) or (2.26) depending on the polynomial order of our elements. Note that both obtain the same result. \square

The preceding Lemma and Proposition show that the GLS/CIP method fits into the framework proposed by Theorem 2. It is possible to improve the a priori error estimates given by the Theorem when using piecewise linear elements.

Proposition 2. (*Improved Error Estimates*) Let $u \in H^{p+1}(\Omega)$, for $p > 1/2$, be the unique solution of (2.15) and $u_h \in V_h$, where V_h is the space of piecewise linear functions then for $kh \leq 1$ the GLS/CIP method satisfies the following improved a priori error estimates

$$\|k(u - u_h)\|_{0,\Omega} \leq C \min\{1, k^3 h^2\} \mathcal{C}_{f,g},$$

$$\|\nabla(u - u_h)\|_{0,\Omega} \leq C(kh + \min\{1, k^3 h^2\}) \mathcal{C}_{f,g},$$

Proof. It was shown earlier in (2.56) that the following discrete approximation result holds,

$$\|\xi_h\|_{0,\Omega} \leq C(kh)h^p |u|_{p+1,\Omega},$$

which assuming sufficient regularity gives

$$\|\xi_h\|_{0,\Omega} \leq (kh)^2 \mathcal{C}_{f,g},$$

multiplying both sides by k implies that

$$\|k\xi_h\|_{0,\Omega} \leq k^3 h^2 \mathcal{C}_{f,g}. \quad (2.82)$$

From the definition of the formulation

$$Re[A(\xi_h, \xi_h)] = \|\nabla\xi_h\|_{0,\Omega}^2 - \|k\xi_h\|_{0,\Omega}^2,$$

rearranging this gives

$$\|k\xi_h\|_{0,\Omega}^2 = \|\nabla\xi_h\|_{0,\Omega}^2 - Re[A(\xi_h, \xi_h)].$$

Using the fact that $Re[A(u_h, u_h)] \leq |A_h(u_h, u_h)|$

$$\|k\xi_h\|_{0,\Omega}^2 \leq \underbrace{\|\nabla\xi_h\|_{0,\Omega}^2}_{(4a)} + \underbrace{|A_h(\xi_h, \xi_h)|}_{(4b)}. \quad (2.83)$$

Considering (4a) element-wise and performing an integration by parts gives

$$(4a) = (\nabla\xi_h, \nabla\xi_h)_{0,\Omega} \quad (2.84)$$

$$= \sum_{\tau \in \mathcal{T}_h} \left(\langle \xi_h, \nabla\xi_h \cdot n \rangle_{0,\partial\tau} - (\xi_h, \Delta\xi_h)_{0,\tau} \right) \quad (2.85)$$

$$= Re \left[\sum_{\tau \in \mathcal{T}_h} \left(\langle \xi_h, \nabla\xi_h \cdot n \rangle_{0,\partial\tau} - (\xi_h, \Delta\xi_h)_{0,\tau} \right) \right]. \quad (2.86)$$

For piecewise linear elements the second order derivatives vanish so this simplifies to

$$(4a) = Re \left[\sum_{F \in \mathcal{F}_{int}} \langle \xi_h, \nabla\xi_h \cdot n \rangle_{0,F} + \langle \xi_h, \nabla\xi_h \cdot n \rangle_{0,\partial\Omega_R} \right]. \quad (2.87)$$

Since $Re[\langle \xi_h, ik\xi_h \rangle_{0,\partial\Omega_R}] = 0$ it is possible to add this term to deal with the Robin condition.

$$(4a) = Re \left[\sum_{F \in \mathcal{F}_{int}} \langle \xi_h, \nabla\xi_h \cdot n \rangle_{0,F} + \langle \xi_h, \nabla\xi_h \cdot n + ik\xi_h \rangle_{0,\partial\Omega_R} \right] \quad (2.88)$$

$$\leq \underbrace{\sum_{F \in \mathcal{F}_{int}} \|k\xi_h\|_{0,F} \|k^{-1}[\nabla\xi_h]\|_{0,F}}_{(4a')} + \underbrace{\|k\xi_h\|_{0,\partial\Omega} \|k^{-1}(\nabla\xi_h \cdot n + ik\xi_h)\|_{0,\partial\Omega_R}}_{(4a'')}. \quad (2.89)$$

Applying the trace inequality (A:4) and the local inverse inequality (A:5) reveals

$$\begin{aligned} (4a') &\leq C \sum_{\tau \in \mathcal{T}_h} \left(h_\tau^{-1/2} \|k\xi_h\|_{0,\tau} + h_\tau^{1/2} \|\nabla k\xi_h\|_{0,\tau} \right) \sum_{F \in \mathcal{F}_{int}} \|k^{-1}[\nabla\xi_h]\|_{0,F} \\ &\leq Ch^{-1/2} \|k\xi_h\|_{0,\Omega} \left(\sum_{F \in \mathcal{F}_{int}} \|k^{-1}[\nabla\xi_h]\|_{0,F}^2 \right)^{1/2}. \end{aligned}$$

Noticing that the jump terms appear in the semi-norm $|\cdot|_{\mathcal{J}}$ using a result from Lemma 2.2 yields

$$\begin{aligned} (4a') &\leq Ch^{-1/2} \|k\xi_h\|_{0,\Omega} \left((h \max\{|\gamma_{1,\tau}|\})^{-1} k^{-1} |\xi_h|_{\mathcal{J}}^2 \right)^{1/2} \\ &\leq C(hk)^{-1} \max\{|\gamma_{1,\tau}|\}^{-1/2} \|k\xi_h\|_{0,\Omega} |\xi_h|_{\mathcal{J}} \\ &\leq C \max\{|\gamma_{1,\tau}|\}^{-1/2} \|k\xi_h\|_{0,\Omega} \mathcal{C}_{f,g}. \end{aligned}$$

Applying the same ideas to (4a'') shows that

$$(4a'') \leq C|\beta_1|^{-1/2} \|k\xi_h\|_{0,\Omega} \mathcal{C}_{f,g}.$$

It follows from Theorem 2 that

$$(4b) \leq (kh)^2 \mathcal{C}_{f,g}.$$

Now note that from (4a) and (4b)

$$\|k\xi_h\|_{0,\Omega}^2 \leq C' \|k\xi_h\|_{0,\Omega} \mathcal{C}_{f,g} + (kh)^2 \mathcal{C}_{f,g}^2. \quad (2.90)$$

Applying the arithmetic-geometric inequality (A:1) to the first term on the right hand side gives

$$\|k\xi_h\|_{0,\Omega}^2 \leq \frac{1}{2} \|k\xi_h\|_{0,\Omega}^2 + \frac{1}{2} C'^2 \mathcal{C}_{f,g}^2 + (kh)^2 \mathcal{C}_{f,g}^2 \quad (2.91)$$

$$\|k\xi_h\|_{0,\Omega}^2 \leq C'^2 \mathcal{C}_{f,g}^2 + 2(kh)^2 \mathcal{C}_{f,g}^2 \quad (2.92)$$

$$\|k\xi_h\|_{0,\Omega}^2 \leq \left(C'^2 + 2(kh)^2 \right) \mathcal{C}_{f,g}^2. \quad (2.93)$$

Taking $kh \leq 1$ and combining the result with (2.82) gives the estimate

$$\|k\xi_h\|_{0,\Omega} \leq C \min\{1, k^3 h^2\} \mathcal{C}_{f,g}.$$

Which then implies using Gårding inequality that

$$\|\nabla \xi_h\|_{0,\Omega} \leq (C \min\{1, k^3 h^2\} + kh) \mathcal{C}_{f,g}. \quad (2.94)$$

□

It is worth noting that although the estimate still contains the standard pollution term the errors are both upper bounded by data. So to conclude the GLS/CIP method proposed here is stable if the stabilization parameters are chosen to have strictly positive imaginary components and the Robin stabilization parameter is chosen appropriately. Also the method satisfies improved a priori error estimates for piecewise linear elements. Numerical evidence seems to suggest that varying the real part of the stabilization parameters can reduce the pollution effect for certain problems. It would appear that the amplitude of the numerical solution is controlled by the imaginary part of the stabilization, whilst the real part of the stabilization effects the numerical dispersion. To really understand the capability of the method to reduce the pollution effect one would need to perform a dispersion analysis and choose stabilization parameters to minimize phase lag. Although this is interesting it is not the focus of this particular work and the

interested reader is advised to read [1] to obtain an understanding of this. One dimensional dispersion relations have been studied by Wu and co workers in [22] for the CIP method and in [28] the authors use the optimal parameters for the one dimensional problem and see a reduction in pollution in the two dimensional case. A two dimensional dispersion analysis is performed for the GLS method on regular and hexagonal meshes in [7].

2.3.2 Continuous Interior Penalty method with penalty on the jumps of first and second derivatives

The CIP method was first introduced by Douglas and Dupont for elliptic and parabolic equations in [27]. The method has since been applied with some success to a number of problems by Burman and co-workers, including Oseen's Equations [16] and the Convection-Diffusion equation [11], [18]. More recently the CIP method has been applied to Helmholtz equation by Wu and co-workers and was shown to eliminate the pollution effect in 1 dimension for an appropriate choice of stabilization parameters. The correct choice of parameter is given by a dispersion analysis performed in [22]. Unfortunately this analysis becomes trickier in higher dimensions and makes strong use of the structure of the underlying mesh. In the paper [28] by Du and Wu the authors suggest introducing a penalty on the jumps of higher order derivatives up to the polynomial order of the elements used. The parameters are chosen by making the 1 dimensional problem pollution free. The numerical results in this paper look promising and show that the method is capable of reducing the pollution effect present in the standard Galerkin method. However it appears that the effect of the higher order derivatives becomes less important as the polynomial order increases. This could be attributed to the fact that the standard Galerkin method performs well for higher order elements.

The second method that I have chosen to analyse is similar to the method introduced in the paper by Du and Wu. For piecewise linear and piecewise quadratic elements the methods are almost identical, the only difference being that I choose to add a GLS term to penalise the Robin-boundary condition so that the method fits into my abstract framework. The method that I present has the advantage of achieving the same theoretical convergence and stability as the method of Du and Wu without needing to introduce additional penalty terms on the jump over element faces of higher order derivatives when the polynomial order of elements is increased. The method differs from the GLS/CIP method that I introduced earlier in that the penalty terms only act on the skeleton of the mesh. Since the first order CIP terms and the jump of the Laplacian is assumed to be zero for the exact solution u there are less terms to be evaluated on the right hand side. An interesting observation regarding the CIP method that I will use is that it satisfies the same improved a priori error estimates for piecewise linear elements as was shown for the GLS/CIP method.

Introducing our CIP stabilization the abstract formulation becomes: find $u_h \in V_h$ such that

$$A_h(u_h, v_h) = L_h(v_h),$$

where

$$\begin{aligned} s(u_h, v_h) = & \sum_{F \in \mathcal{F}_{int}} h^3 \delta_{2,\tau} \langle \llbracket \Delta u_h \rrbracket, \llbracket \Delta v_h \rrbracket \rangle_{0,F} + \sum_{F \in \mathcal{F}_{int}} h \gamma_{2,\tau} \langle \llbracket \nabla u_h \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_{0,F} \\ & + \beta_2 h \langle \mathcal{R}(u_h), \mathcal{R}^*(v_h) \rangle_{0,\partial\Omega}, \end{aligned} \quad (2.95)$$

where $\delta_{2,\tau}, \gamma_{2,\tau}, \beta_2 \in \mathbb{C}$. The first term in $s(\cdot, \cdot)$ represents the jump of the Laplacian over element faces. Notice that again this is allowed since elements $u_h \in C^\infty(\tau)$ when restricted to the element τ , for piecewise linear elements this term is zero. As I mentioned earlier that assuming u smooth gives

$$\sum_{F \in \mathcal{F}_{int}} h^3 \delta_{2,\tau} \langle \llbracket \Delta u \rrbracket, \llbracket \Delta v_h \rrbracket \rangle_{0,F} = 0.$$

The last two terms in the stabilization are also present in the GLS/CIP method so I will not discuss them any further.

Now that I have introduced the stabilization I would like to show that it enters the mathematical framework of Theorem 2. It is necessary to choose an appropriate operator $\pi_h : V \mapsto V_h$ that satisfies the assumptions of the Theorem. For this purpose I choose $\pi_h : L^2(\Omega) \mapsto V_h$ to denote the L^2 orthogonal projection defined in the Appendix A.5. The reason for this choice will be made clear in the subsequent analysis. It is well known that the L^2 orthogonal projection satisfies the interpolation estimates of Assumptions (2.25) or (2.26). As for the case with the GLS/CIP stabilization the following lemma asserts that the weak coercivity property on $|\cdot|_s^2 : V \mapsto \mathbb{R}$ defined by

$$\begin{aligned} |u_h|_s^2 & \stackrel{\text{def}}{=} k(1 - \text{Im}[\beta_2]hk) \|u_h\|_{0,\partial\Omega_R}^2 + \sum_{F \in \mathcal{F}_{int}} h_\tau^3 \text{Im}[\delta_{2,F}] \|\llbracket \Delta(u_h) \rrbracket\|_{0,\tau}^2 \\ & + \sum_{F \in \mathcal{F}_{int}} h_\tau \text{Im}[\gamma_{2,\tau}] \|\llbracket \nabla u_h \rrbracket\|_{0,F}^2 + \text{Im}[\beta_2]h \|\nabla(u_h) \cdot n\|_{0,\partial\Omega_R}^2, \end{aligned} \quad (2.96)$$

and Cauchy-Schwarz type inequalities hold for the CIP stabilization. The Theorem also asks that the interpolation estimates (2.30) and (2.31) hold for the L^2 projection and that the L^2 projection adheres to the stability estimate (2.29) for

$$|\cdot|_s^2 \stackrel{\text{def}}{=} \sum_{F \in \mathcal{F}_{int}} |\delta_{2,\tau}| h_\tau^3 \|\llbracket \Delta(\cdot) \rrbracket\|_{0,F}^2 + \sum_{F \in \mathcal{F}_{int}} |\gamma_{2,\tau}| h_\tau \|\llbracket \nabla(\cdot) \cdot n \rrbracket\|_{0,F}^2 + |\beta_2| h \|\mathcal{R}^*(\cdot)\|_{0,\partial\Omega_R}^2, \quad (2.97)$$

Lemma 2.8. *Let $k > 1$, $kh < \text{Im}[\beta_2]^{-1}$, $\text{Re}[\beta_2] = 0$, $\text{Im}[\delta_2], \text{Im}[\gamma_2] > 0$, $\text{Im}[\delta_2] > C \text{Re}[\delta_2]$ and*

$Im[\gamma_2] > CRe[\gamma_2]$ then the CIP stabilization, taken on a quasi uniform mesh, satisfies Assumptions (2.23), (2.24), (2.29), (2.30) and (2.31) for

$$\|\eta\|_* \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} (h_\tau^2 |\delta_{2,\tau}|)^{-\frac{1}{2}} \|\eta\|_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} (h_\tau |\gamma_{2,\tau}|)^{-\frac{1}{2}} \|\eta\|_{0,F} + (h |\beta_2|)^{-\frac{1}{2}} \|\eta\|_{0,\partial\Omega_R} + |\eta|_{\mathcal{J}}^{\frac{1}{2}},$$

and π_h as the standard L^2 orthogonal projection.

Proof. To begin the proof of the lemma I start by showing the weak coercivity property as in Assumption (2.23). Once again the proof follows from noticing that under the conditions of the lemma $|\cdot|_{\mathcal{I}}$ is a seminorm on V_h and that

$$|u_h|_{\mathcal{I}}^2 = Im [A_h(u_h, u_h)]$$

Expanding the right hand side confirms the claim

$$\begin{aligned} Im [A_h(u_h, u_h)] &= k(u_h, u_h)_{0,\partial\Omega_R} + Im [s(u_h, u_h)] \\ &= k \|u_h\|_{0,\partial\Omega_R}^2 + \sum_{F \in \mathcal{F}_{int}} h_\tau^3 Im [\delta_{2,\tau}] \|[\Delta u_h]\|_{0,F}^2 + \sum_{F \in \mathcal{F}_{int}} h_\tau Im [\gamma_{2,\tau}] \|[\nabla u_h]\|_{0,F}^2 \\ &\quad + Im [\beta_2 h \langle \mathcal{R}(u_h), \mathcal{R}^*(u_h) \rangle_{0,\partial\Omega_R}]. \end{aligned}$$

Considering the final term in a similar way to that of the GLS method concludes the claim. Next I show that the stabilization satisfies a type of Cauchy-Schwarz inequality with respect to $|\cdot|_s$ ie Assumption (2.24) holds.

$$\begin{aligned} s(v_h, w_h) &= \sum_{F \in \mathcal{F}_{int}} h^3 \delta_{2,\tau} \langle [\Delta v_h], [\Delta w_h] \rangle_F + \sum_{F \in \mathcal{F}_{int}} h \gamma_{2,\tau} \langle [\nabla v_h], [\nabla w_h] \rangle_F \\ &\quad + \beta_2 h \langle \mathcal{R}(v_h), \mathcal{R}^*(w_h) \rangle_{\partial\Omega_R} \\ &\leq \sum_{F \in \mathcal{F}_{int}} h_\tau^2 |\delta_{2,\tau}| \|[\Delta v_h]\|_{0,F} \|[\Delta w_h]\|_{0,F} + \sum_{F \in \mathcal{F}_{int}} h_\tau |\gamma_{2,\tau}| \|[\nabla v_h]\|_{0,F} \|[\nabla w_h]\|_{0,F} \\ &\quad + |\beta_2| h \| \mathcal{R}(v_h) \|_{0,\partial\Omega_R} \| \mathcal{R}^*(w_h) \|_{0,\partial\Omega_R} \\ &\leq \sum_{F \in \mathcal{F}_{int}} h_\tau^2 |\delta_{2,\tau}| \|[\Delta v_h]\|_{0,F} \|[\Delta w_h]\|_{0,F} + \sum_{F \in \mathcal{F}_{int}} h_\tau |\gamma_{2,\tau}| \|[\nabla v_h]\|_{0,F} \|[\nabla w_h]\|_{0,F} \\ &\quad + |\beta_2| h (\| \nabla(v_h) \cdot n \|_{0,\partial\Omega_R} + k \|v_h\|_{0,\partial\Omega_R}) \| \mathcal{R}^*(w_h) \|_{0,\partial\Omega_R} \\ &\leq C |v_h|_{\mathcal{I}} |w_h|_s. \end{aligned}$$

The proof uses the Cauchy-Schwarz inequality in L^2 and asks that the real parts of the stabilization parameters can be bounded by their respective imaginary parts. To show the stability of the L^2 projection in $|\cdot|_s$, Assumption (2.29), it is beneficial (for the sake of notation) to treat individual terms separately,

consider

$$|\pi_h z|_s = \left(\underbrace{\sum_{F \in \mathcal{F}_{int}} h_\tau^3 \text{Im}[\delta_{2,\tau}] \|\llbracket \Delta \pi_h z \rrbracket\|_{0,F}^2}_{(5a)} + \underbrace{\sum_{F \in \mathcal{F}_{int}} h_\tau \text{Im}[\gamma_{2,\tau}] \|\llbracket \nabla \pi_h z \rrbracket\|_{0,F}^2}_{(5b)} + \underbrace{\text{Im}[\beta_2] h \|\mathcal{R}^*(\pi_h z)\|_{0,\partial\Omega_R}^2}_{(5c)} \right)^{1/2}.$$

To show (5a), let us once again make use of the trace and inverse inequalities given by (A:4) and (A:5) respectively and the stability of the L^2 projection. It should be noted that this term is zero for piecewise linear approximation spaces. Since the underlying mesh is assumed quasi-uniform the L^2 projection is known to be stable in the broken $H^2(\mathcal{T}_h)$ norm, a definition for this norm can be found in the appendix section.

$$(5a) \leq \sum_{\tau \in \mathcal{T}_h} \sum_{F \in \partial\tau} h_\tau^3 |\delta_{2,\tau}| \|\Delta \pi_h z\|_{0,F}^2 \quad (2.98)$$

$$\leq \sum_{\tau \in \mathcal{T}_h} C h_\tau^3 |\delta_{2,\tau}| h^{-1} \|\Delta \pi_h z\|_{0,\tau}^2 \quad (2.99)$$

$$\leq C h^2 |\delta_{2,\tau}| \|\Delta \pi_h z\|_{0,\mathcal{T}_h}^2 \quad (2.100)$$

$$\leq C |\delta_{2,\tau}| h^2 |z|_{2,\Omega}^2 \quad (2.101)$$

$$\leq C |\delta_{2,\tau}| (hk)^2 \|\psi\|_{0,\Omega}^2. \quad (2.102)$$

To bound (5b) in the required way it is useful to note that since the exact solution of the dual problem z is assumed to be smooth the CIP terms of first order are zero. Adding this term followed by the use of the trace inequality (A:4) and interpolation estimates (2.25) or (2.26) the desired result is obtained

$$(5b) = \sum_{F \in \mathcal{F}_{int}} h_\tau \text{Im}[\gamma_{2,\tau}] \|\llbracket \nabla(z - \pi_h z) \rrbracket\|_{0,F}^2 \quad (2.103)$$

$$\leq h \max\{|\gamma_{2,\tau}|\} \sum_{\tau \in \mathcal{T}_h} \sum_{F \in \partial\tau} \|\nabla(z - \pi_h z) \cdot n\|_{0,F}^2 \quad (2.104)$$

$$\leq C |\gamma_{2,\tau}| h \left(h^{-1/2} \|z - \pi_h z\|_{1,\mathcal{T}_h} + h^{1/2} \|z - \pi_h z\|_{2,\mathcal{T}_h} \right)^2 \quad (2.105)$$

$$\leq C |\gamma_{2,\tau}| h^2 |z|_{2,\Omega}^2 \quad (2.106)$$

$$\leq C |\gamma_{2,\tau}| (hk)^2 \|\psi\|_{0,\Omega}^2. \quad (2.107)$$

(5c) follows by noticing that $\mathcal{R}^*(z) = 0$ on the boundary. Subtracting this and then applying the trace

inequality (A:4) and interpolation estimates (2.25) or (2.26) gives

$$\begin{aligned}
(5c) &\leq |\beta_2| h \sum_{FC\partial\Omega_R} (\|\nabla(z - \pi_h z) \cdot n - ik(z - \pi_h z)\|_{0,F}^2) \\
&\leq |\beta_2| h \sum_{FC\partial\Omega_R} (\|\nabla(z - \pi_h z) \cdot n\|_{0,F}^2 + 2k\|z - \pi_h z\|_{0,F} \|\nabla(z - \pi_h z) \cdot n\|_{0,F} \\
&\quad + \|k(z - \pi_h z)\|_{0,F}^2) \\
&\leq C|\beta_2|(1 + kh)^2 h^2 |z|_{2,\Omega}^2 \\
&\leq C|\beta_2|(1 + kh)^2 (hk)^2 \|\psi\|_{0,\Omega}^2.
\end{aligned}$$

Combining these results proves Assumption (2.29). The proof of Assumption (2.31) follows in a similar manner, however since $u \in H^{p+1}(\Omega)$ we are not limited to $H^2(\Omega)$ regularity. The only additional terms that must be investigated are the boundary terms. Both terms conform after an application of the trace inequality and the interpolation results for the L^2 projection.

$$k\|\eta\|_{0,\partial\Omega_R}^2 = k \sum_{FC\partial\Omega_R} \|\eta\|_{0,F}^2 \quad (2.108)$$

$$\leq Ck \left(h^{-1/2} \|\eta\|_{0,\mathcal{T}_h} + h^{1/2} |\eta|_{1,\mathcal{T}_h} \right)^2 \quad (2.109)$$

$$\leq Ck \left(h^{p+1/2} |u|_{p+1,\Omega} \right)^2 \quad (2.110)$$

$$\leq C(kh)h^{2p} |u|_{p+1,\Omega}^2. \quad (2.111)$$

The second inequality follows similarly.

$$\beta_2 h \|\nabla\eta \cdot n\|_{0,\partial\Omega_R}^2 = \beta_2 h \sum_{FC\partial\Omega_R} \|\nabla\eta\|_{0,F}^2 \quad (2.112)$$

$$\leq C\beta_2 h \left(h^{-1/2} \|\nabla\eta\|_{0,\mathcal{T}_h} + h^{1/2} |\nabla\eta|_{1,\mathcal{T}_h} \right)^2 \quad (2.113)$$

$$\leq C\beta_2 (h^p |u|_{p+1,\Omega})^2 \quad (2.114)$$

$$\leq C\beta_2 h^{2p} |u|_{p+1,\Omega}^2. \quad (2.115)$$

The final assumption to verify is the interpolation estimate in the *-norm, (2.30). I shall omit the proof of this since it follows in the same vein as it did for GLS/CIP case. \square

The previous lemma has verified that the CIP stabilization satisfies the interpolation assumptions in Theorem 2 as well as assumptions about the stability of the L^2 projection in $|\cdot|_s$ and the weak coercivity property. The only assumptions left to verify are results regarding the continuity of the forms $A_h(\cdot, \cdot)$ and $A(\cdot, \cdot)$.

Proposition 3 (Continuity). *Let $u \in H^{p+1}(\Omega)$ for $p > 1/2$ and $\pi_h : L^2(\Omega) \mapsto V_h$ be the standard L^2 orthogonal projection then it follows that the CIP discrete sesquilinear form satisfies Assumptions (2.27) and (2.28)*

Proof. Let $\eta = u - \pi_h u$, then as before since functions in V_h are C^∞ on each element $\tau \in \mathcal{T}_h$ then it is possible to perform an integration by parts over each τ to obtain

$$\begin{aligned} |A_h(\eta, v_h)| &= \left| \sum_{\tau \in \mathcal{T}_h} (\eta, -\Delta v_h)_{0,\tau} - k^2(\eta, v_h)_{0,\Omega} + \sum_{\tau \in \mathcal{T}_h} \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\tau} + \langle ik\eta, v_h \rangle_{0,\partial\Omega_R} \right. \\ &\quad \left. + s(\eta, v_h) \right| \\ &= \left| \sum_{\tau \in \mathcal{T}_h} (\eta, -\Delta v_h)_\tau + \sum_{F \in \mathcal{F}_{int}} \langle \eta, \llbracket \nabla v_h \cdot n \rrbracket \rangle_F + \langle \eta, \nabla v_h \cdot n + ikv_h \rangle_{0,\partial\Omega_R} + s(\eta, v_h) \right|. \end{aligned}$$

The omission of the low order term $k^2(\eta, v_h)_{0,\tau}$ comes from the definition of the L^2 projection

$$(u - \pi_h u, v_h)_{0,\Omega} = 0 \quad \forall v_h \in V_h,$$

exploiting this property again allows us the freedom to add a term $(\eta, x_h)_{0,\tau}$, where $x_h \in V_h$

$$\begin{aligned} |A_h(\eta, v_h)| &= \left| \sum_{\tau \in \mathcal{T}_h} (\eta, -\Delta v_h + x_h)_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} \langle \eta, \llbracket \nabla v_h \cdot n \rrbracket \rangle_{0,F} \right. \\ &\quad \left. + \langle \eta, \nabla v_h \cdot n + ikv_h \rangle_{0,\partial\Omega_R} + s(\eta, v_h) \right| \\ &\leq \sum_{\tau \in \mathcal{T}_h} \|\eta\|_{0,\tau} \|\Delta v_h - x_h\|_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} \|\eta\|_{0,F} \|\llbracket \nabla v_h \cdot n \rrbracket\|_{0,F} \\ &\quad + \|\eta\|_{0,\partial\Omega_R} \|\nabla v_h \cdot n + ikv_h\|_{0,\partial\Omega_R} + |s(\eta, v_h)|. \end{aligned}$$

Since this statement holds for any $x_h \in V_h$ I can take the infimum over V_h of the first term

$$\begin{aligned} |A_h(\eta, v_h)| &\leq \inf_{x_h \in V_h} \sum_{\tau \in \mathcal{T}_h} \|\eta\|_{0,\tau} \|\Delta v_h - x_h\|_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} \|\eta\|_{0,F} \|\llbracket \nabla v_h \cdot n \rrbracket\|_{0,F} \\ &\quad + \|\eta\|_{0,\partial\Omega_R} \|\nabla v_h \cdot n - ikv_h\|_{0,\partial\Omega_R} + |s(\eta, v_h)| \\ &\leq \|\eta\|_* |v_h|_{\mathfrak{J}}. \end{aligned}$$

The last step comes from using the fact that

$$\inf_{x_h \in V_h} \sum_{\tau \in \mathcal{T}_h} \|\Delta v_h - x_h\|_{0,\tau}^2 \leq C \sum_{F \in \mathcal{F}_{int}} h_\tau \|\llbracket \Delta v_h \rrbracket\|_{0,F}^2,$$

a proof for which can be found in [15]. Finally, Assumption (2.28) follows in a similar manner to the

GLS/CIP case

$$\begin{aligned}
|A(\eta, z - \pi_h z)| &= |(\nabla \eta, \nabla(z - \pi_h z))_{0,\Omega} - k^2(\eta, z - \pi_h z)_{0,\Omega} + \langle ik\eta, z - \pi_h z \rangle_{\partial\Omega_R}| \\
&\leq \|\nabla \eta\|_{0,\Omega} \|\nabla(z - \pi_h z)\|_{0,\Omega} + \|k\eta\|_{0,\Omega} \|k(z - \pi_h z)\|_{0,\Omega} \\
&\quad + \|k^{\frac{1}{2}}\eta\|_{\partial\Omega_R} \|k^{\frac{1}{2}}(z - \pi_h z)\|_{\partial\Omega_R}.
\end{aligned}$$

The result now follows immediately after an application of the trace inequality (A:4) to the last term and then using the interpolation estimates for the L^2 projection given by (2.25) or (2.26). \square

An interesting observation is that for piecewise linear elements this method and the GLS/CIP method proposed earlier are almost identical. The only difference being that the low order term in the least squares stabilization is not present in this method. However from the point of view of analysis the improved a priori error estimates hold in both cases. The proof that the CIP method proposed in this section satisfies the bounds given by (2.62) and (2.63) is identical to the proof for the GLS/CIP method.

2.4 Numerical examples

This section aims to verify the theoretical results given previously both of the stabilized method's proposed are tested against the standard Galerkin method. The numerical results appear to verify the claims of the theorem and in some cases work better than anticipated. Since the parameter in the standard GLS technique proposed in [34] is known to rely heavily on the wave direction of the underlying solution it would be interesting to see if this is the case with the GLS/CIP and CIP stabilizations.

The stabilization parameters used in the following section have been chosen with the aim of reducing the pollution effect. A range of parameters have been considered and later I will discuss their sensitivity. The parameters have been chosen by testing an array of different selections using a series of “for loops” and using a search command to find the choice of parameters with the lowest L^2 error for a range of different k and h . The L^2 errors were stored in a large array and for this reason have not been included in the proceeding section. All computations in this section were performed using the UMFPACK solver in the FreeFEM++ package which is available from <http://www.freefem.org/>.

```

for Numerous  $k$  and  $h$  do
  initialization;
  for Imaginary part of parameters in a certain bound do
    for Real part of parameters in a certain bound do
      Solve ;
      print relative  $L^2$  error ;
    end
  end
  search Parameters with minimum relative error ;
  if Parameters for  $k_{n-1}, h_{n-1}$  match parameters for  $k_n, h_n$  then
    accept parameters;
  else
    try an average of the parameters;
  end
  n++;
end

```

Algorithm 1: Parameter Search

The parameters that have been chosen as optimal are:

The parameters for GLS/CIP stabilization using piecewise affine elements are

$$\gamma_{1,\tau} = -0.0287 + 0.00216i, \delta_{1,\tau} = -0.003 + 0.00025i, \beta_1 = 0 + 0.00025i.$$

and for piecewise quadratic elements

$$\gamma_{1,\tau} = -0.01435 + 1.08 \times 10^{-3}i, \delta_{1,\tau} = -7.5 \times 10^{-4} + 6.25 \times 10^{-5}i, \beta_1 = 0 + 6.25 \times 10^{-5}i.$$

The parameters for CIP stabilization using piecewise affine elements are

$$\gamma_{2,\tau} = -0.0287 + 0.00216i, \beta_2 = 0 + 0.00025i.$$

and for piecewise quadratic elements

$$\gamma_{2,\tau} = -0.01435 + 1.08 \times 10^{-3}i, \delta_{2,\tau} = -3.75 \times 10^{-4} + 3.125 \times 10^{-5}i, \beta_2 = 0 + 1.5625 \times 10^{-5}i.$$

The parameters have been optimized using the Bessel solution that I present below. The robustness of these parameters is then tested using a plane wave solution. Note that for our analysis to hold the sign of

the imaginary parts of the parameters must change when being applied to the plane wave solution since the sign in the Robin boundary condition is different.

2.4.1 Bessel solution

The Bessel function is an interesting case since the solution has no dominant wave direction. This example was studied in [28].

The problem is designed such that

$$u = \frac{\cos\left(k\sqrt{x^2 + y^2}\right)}{k} - CJ_0\left(k\sqrt{x^2 + y^2}\right). \quad (2.116)$$

where

$$C = \frac{\cos(k) + i \sin(k)}{k(J_0(k) + iJ_1(k))}.$$

and $J_0(\cdot)$ and $J_1(\cdot)$ are Bessel functions of the First Kind. ie:

$$J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+\alpha)!} \left(\frac{x}{2}\right)^{2m+\alpha} \quad \forall \alpha \in \mathbb{N}_0.$$

and

$$J_{-\alpha}(x) = (-1)^\alpha J_\alpha(x).$$

We pose the following problem on the unit square $\Omega = [0, 1] \times [0, 1]$

$$\begin{cases} -\Delta u - k^2 u = f & \text{in } \Omega \\ \nabla u \cdot n + iku = g_R & \text{on } \partial\Omega \end{cases}$$

For our ansatz to be the solution let

$$f = \frac{\sin(k\sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}.$$

Naturally, it is possible to obtain the Dirichlet boundary conditions by using the exact solution at the boundary. However, I must now find g_R .

$$\begin{aligned} g_R &= \frac{(xn_x + yn_y)}{\sqrt{x^2 + y^2}} \left[CkJ_1(k\sqrt{x^2 + y^2}) - \sin(k\sqrt{x^2 + y^2}) \right] \\ &\quad - ik \left[\frac{\cos(k\sqrt{x^2 + y^2})}{k} - CJ_0(k\sqrt{x^2 + y^2}) \right]. \end{aligned}$$

where n_x and n_y are the x and y components of the normal vector.

The following convergence plots were created using the optimal parameters stated above, on a structured mesh see for example [32].

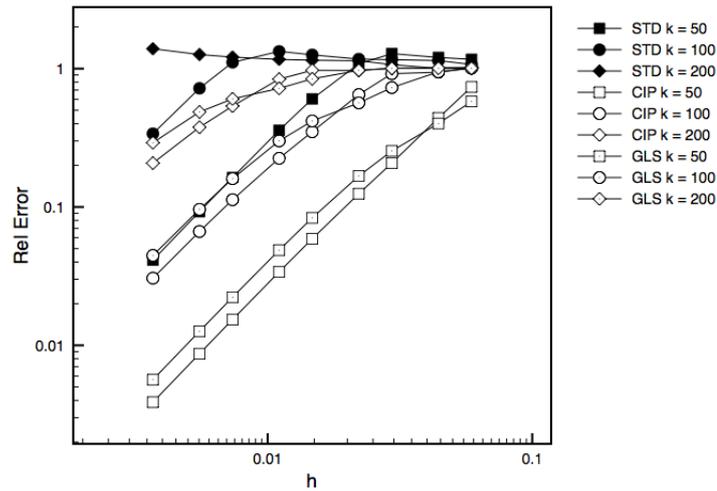


Figure 2.1: Convergence plot for the Bessel Solution using piecewise linear elements on a structured mesh

Figure 2.1 shows that the numerical methods all reach optimal convergence. The stabilized methods can be seen to begin converging for much coarser meshes than the standard Galerkin method. It is also interesting to note that the stabilized methods never obtain a relative error higher than 1 which appears to agree with the improved error estimates proven in the previous section. The standard Galerkin method does not respect this estimate which essentially means that there is no limit to how bad the solution can be.

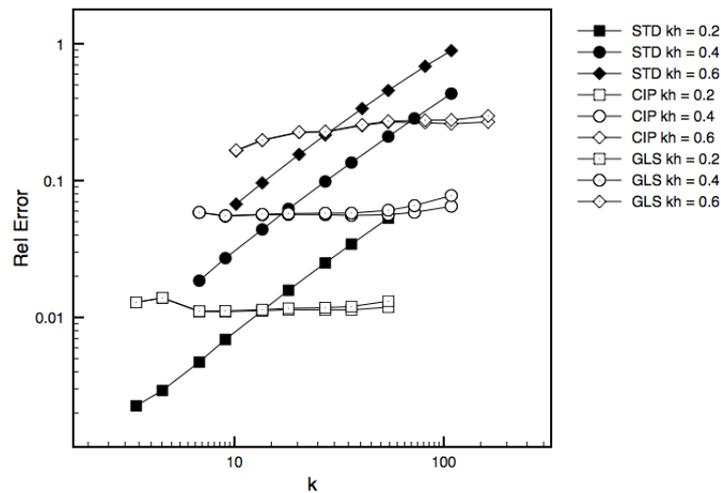


Figure 2.2: Pollution study for Bessel Solution using piecewise linear elements on a structured mesh

Figures 2.2 and 2.3 are designed to test the accuracy of the numerical methods under the constraint $kh = C$. The mesh is refined globally the idea being that as the mesh is refined the value of k increases. As expected, the relative error for the standard Galerkin method increases with k which is a direct consequence of numerical pollution. What is surprising with the plot showing the results using piecewise

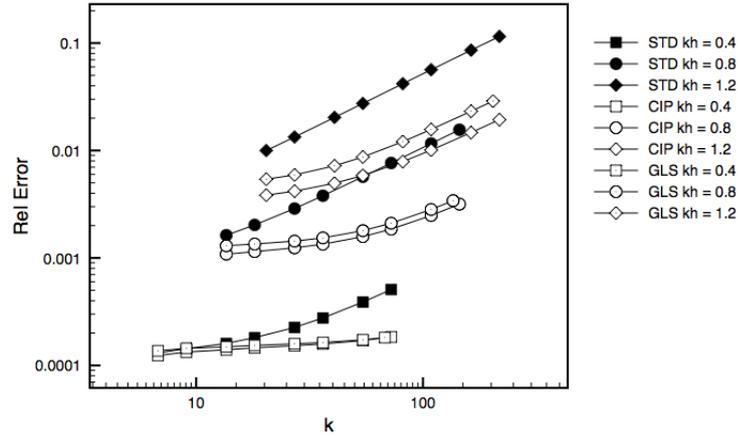


Figure 2.3: Pollution study for Bessel Solution using piecewise quadratic elements on a structured mesh

linear approximation is the fact that the stabilized methods appear to maintain a constant error in the $kh = C$ regime. This would appear to suggest that the methods have some potential to substantially reduce the pollution error for a given problem on a structured mesh. The piecewise quadratic plot is not as convincing but the methods do still seem to reduce the pollution effect as the gradients of the error curves for the stabilized methods appear less steep than for the standard Galerkin method.

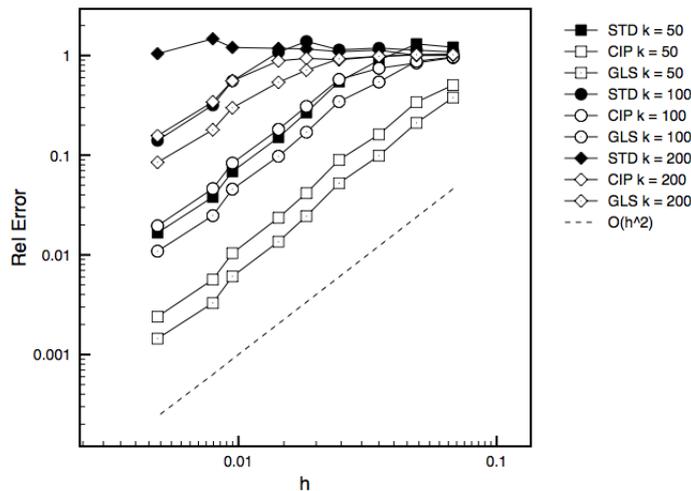


Figure 2.4: Convergence plot for the Bessel Solution using piecewise linear elements on an unstructured mesh

Figures 2.4 and 2.5 show that all methods converge at the rate that the theory predicts for unstructured meshes, see for example [32]. Again the stabilized methods appear to begin converging optimally for a coarser mesh when considering piecewise linear elements and seem to respect the improved error estimates in the same way as they did for the structured mesh example. As was the case previously, the advantages of the stabilizations are not as apparent in the case of piecewise quadratic elements.

Finally, Figures 2.6 and 2.7 show pollution studies for unstructured meshes using piecewise linear and

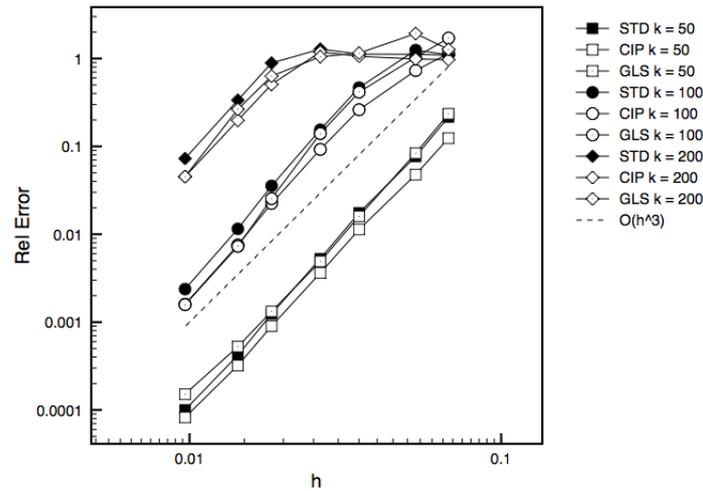


Figure 2.5: Convergence plot for the Bessel Solution using piecewise quadratic elements on an unstructured mesh

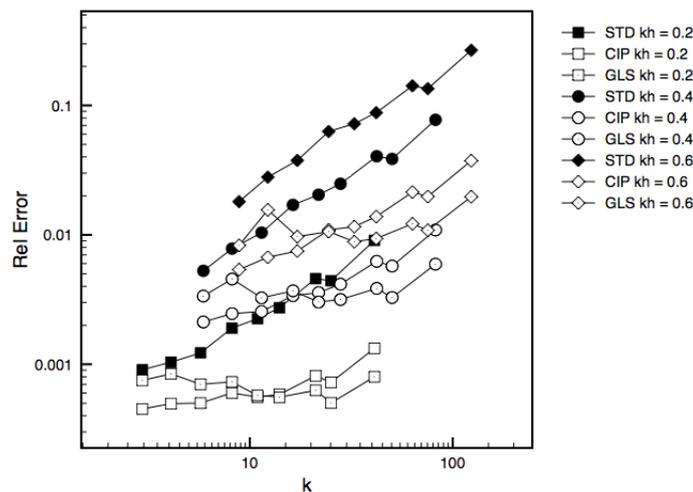


Figure 2.6: Pollution study for Bessel Solution using piecewise linear elements on an unstructured mesh

quadratic elements respectively. Figure 2.6 shows a reduction in the pollution effect for piecewise linear elements. The reduction is not as prominent as for the structured mesh since the stabilization parameters appear to be dependent on how the solution is oriented with respect to the mesh. Even though the image is not as clear as for the structured mesh there is still an argument for the stabilized methods ability to reduce the effect of numerical pollution. Interestingly enough, Figure 2.7 seems to show the stabilizations working well to reduce the pollution error. It would have been interesting to extend this to higher wavenumber k although the system becomes ill conditioned in this regime and appropriate preconditioners must be employed to ensure that iterative solvers converge to a reliable solution in an appropriate time frame.

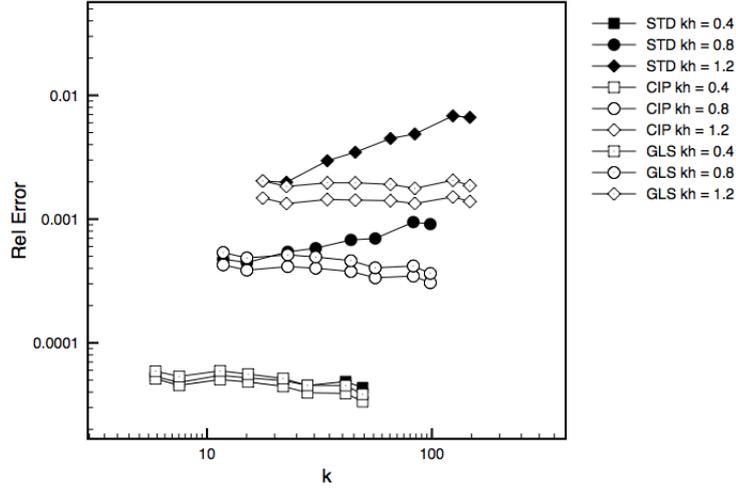


Figure 2.7: Pollution study for Bessel Solution using piecewise quadratic elements on an unstructured mesh

2.4.2 Plane wave solution

The plane wave solution has been constructed to test the robustness of the methods. As mentioned previously a clear limitation to the standard GLS method is the requirement to know a priori a wave direction to choose the correct stabilization parameter. This knowledge is not usually available in practice.

Let's begin by showing that $u = e^{i(k_1x+k_2y)}$ is a solution of the problem

$$\left\{ \begin{array}{ll} -\Delta u(x, y) - k^2 u(x, y) = 0 & (x, y) \in [0, 1] \times [0, 1] \\ \frac{\partial u(x, y)}{\partial y} = -ik_2 u(x, y) & \text{for } y = 0 \\ \frac{\partial u(x, y)}{\partial x} = -ik_1 u(x, y) & \text{for } x = 0 \\ \frac{\partial u(x, y)}{\partial y} = ik_2 u(x, y) & \text{for } y = 1 \\ \frac{\partial u(x, y)}{\partial x} = ik_1 u(x, y) & \text{for } x = 1 \end{array} \right.$$

It can be seen immediately that

$$\begin{aligned} \frac{\partial^2 u(x, y)}{\partial x^2} &= -k_1^2 e^{i(k_1x+k_2y)} \\ \frac{\partial^2 u(x, y)}{\partial y^2} &= -k_2^2 e^{i(k_1x+k_2y)}. \end{aligned}$$

Which when substituted into our problem gives

$$k_1^2 e^{i(k_1x+k_2y)} + k_2^2 e^{i(k_1x+k_2y)} - k^2 e^{i(k_1x+k_2y)} = 0.$$

So taking $k_1 = k \cos(\theta)$ and $k_2 = k \sin(\theta)$, where θ can be thought of as the angle of propagation (wave

direction), gives

$$\begin{aligned} & (k^2 \cos^2(\theta) + k^2 \sin^2(\theta) - k^2) e^{i(k_1 x + k_2 y)} = 0 \\ \implies & (k^2 (\cos^2(\theta) + \sin^2(\theta)) - k^2) e^{i(k_1 x + k_2 y)} = 0 \\ \implies & (k^2 - k^2) e^{i(k_1 x + k_2 y)} = 0. \end{aligned}$$

u clearly satisfies the boundary conditions and is therefore a solution to the Helmholtz equation.

Figure 2.8 shows the convergence of the methods for a moderate wave number $k = 10$ on a structured mesh. The graph verifies the theory as all methods achieve second order convergence for $\theta = \pi/n$ where $n \in \{2, 3, 4, 5\}$. The stabilized methods are seen to be more accurate than the standard Galerkin method with the GLS/CIP method slightly more accurate than the CIP method for a structured mesh. Natural logs have been used in all of the following log – log graphs.

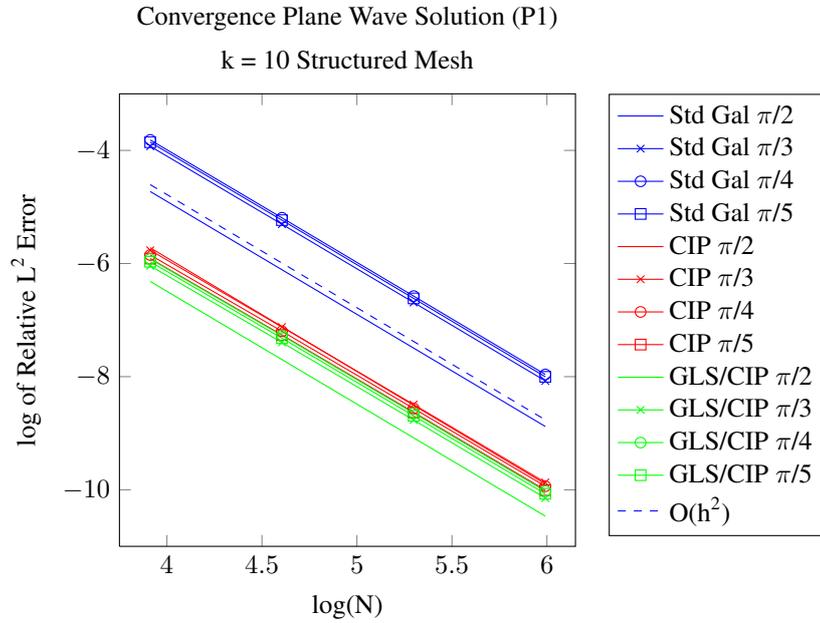


Figure 2.8: Convergence plot for the plane wave solution using piecewise linear elements on a structured mesh with $k = 10$

Figure 2.9 shows the convergence of the methods for $k = 100$ on a structured mesh. Initially none of the methods perform very well. This is because the mesh is not refined enough to fully resolve each wave length. I mentioned earlier the engineering rule of thumb is to have $10 \sim 12$ elements per wavelength. However

$$\lambda = \frac{2\pi}{k}, h = \frac{\sqrt{2}}{N} \quad (2.117)$$

$$\frac{\lambda}{h} = \sqrt{2}\pi \frac{N}{k}. \quad (2.118)$$

For Figure 2.9 $N \in \{50, 100, 200, 400\}$ and $k = 100$, this gives that $\frac{\lambda}{h} \in \{\frac{\pi}{\sqrt{2}}, \sqrt{2}\pi, 2\sqrt{2}\pi, 4\sqrt{2}\pi\}$ the rule of thumb is only met in the last case. The standard Galerkin method performs poorly and for $\theta \neq \frac{\pi}{2}$ does not begin to converge until each wavelength is fully resolved. The stabilized methods reach optimal convergence quicker than the standard Galerkin method and remain more accurate. An interesting observation is that all of the methods seem better equipped to handle the case $\theta = \frac{\pi}{2}$. This is not the case for the unstructured mesh and can be explained by the nature of the solution. Since the solution for $\theta = \frac{\pi}{2}$ travels along the boundary, the number of elements per wavelength is better represented by $\frac{\lambda}{h} \in \{\pi, 2\pi, 4\pi, 8\pi\}$ since $h \in \{\frac{1}{50}, \frac{1}{100}, \frac{1}{200}, \frac{1}{400}\}$ along the boundary. This means that the wave is better resolved in this case.

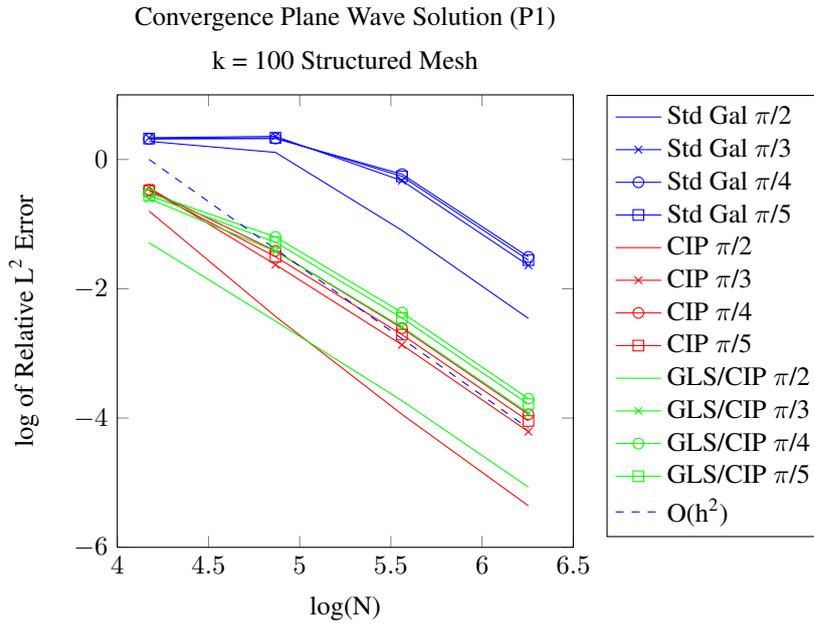


Figure 2.9: Convergence plot for the plane wave solution using piecewise linear elements on a structured mesh with $k = 100$

Figures 2.10 and 2.11 show the convergence of the methods for different values of k . Since the mesh is unstructured the direction of the wave has less influence on the accuracy of the solutions. The structure the mesh appears to have little effect on the accuracy of the CIP method but the GLS/CIP stabilization seems to perform better on the unstructured mesh. All methods can be seen to converge optimally.

The following figures aim to determine the effectiveness of the stabilizations in eliminating the pollution error from numerical solutions. Here the log of the relative error is plotted against $\log(N)$ with $kh = C$. The analysis says that the relative error should grow with k since $\|u - u_h\|_{0,\Omega} = O((kh)^2)$ and $\|u\|_{0,\Omega} = O(k^{-1})$ therefore the relative error is

$$\frac{\|u - u_h\|_{0,\Omega}}{\|u\|_{0,\Omega}} = O(k^3 h^2) \quad (2.119)$$

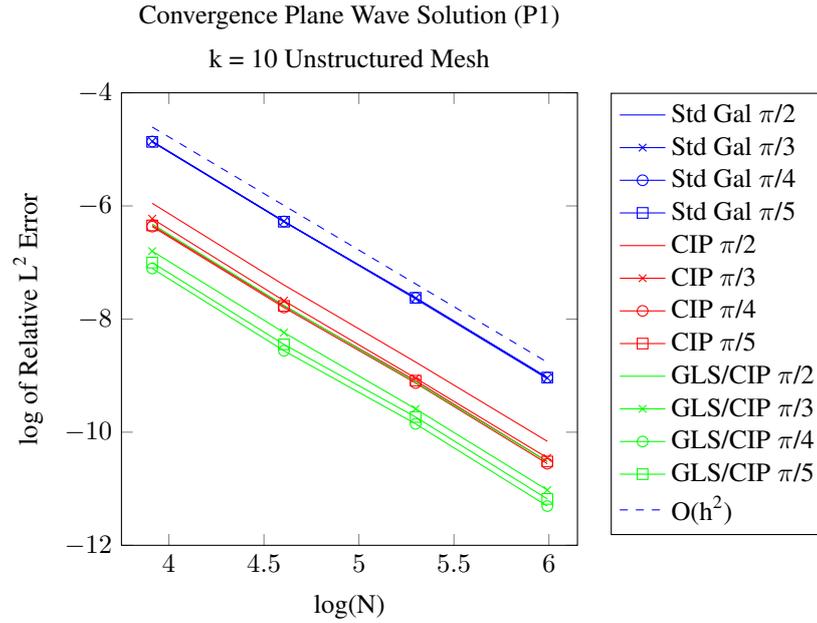


Figure 2.10: Convergence plot for the plane wave solution using piecewise linear elements on an unstructured mesh with $k = 10$

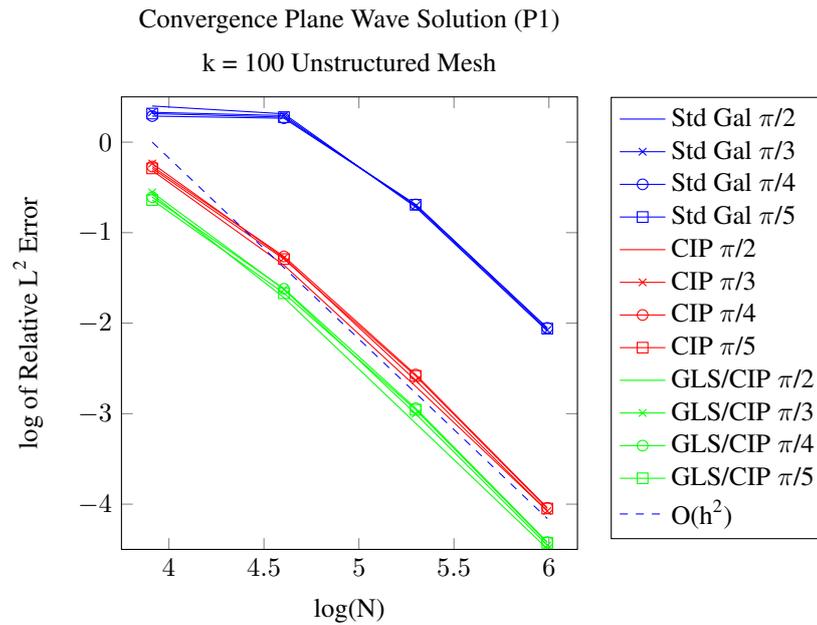


Figure 2.11: Convergence plot for the plane wave solution using piecewise linear elements on an unstructured mesh with $k = 100$

Since kh is taken to be constant the error is expected to grow with k . Figure 2.12 shows that the standard Galerkin method performs as would be expected, with the relative error growing with k . The two stabilized methods, on the other hand, remain constant which indicates the methods are pollution free in this case. What is even more interesting is the fact that this works for a range of θ .

Figure 2.13 shows a clear reduction in the pollution effect for $kh = 0.5$. The choice of $kh = 0.5$ is equivalent to approximately 12 elements per wavelength. In the figure $k \in \{12.5\sqrt{2}, 25\sqrt{2}, 50\sqrt{2}, 100\sqrt{2}\}$.

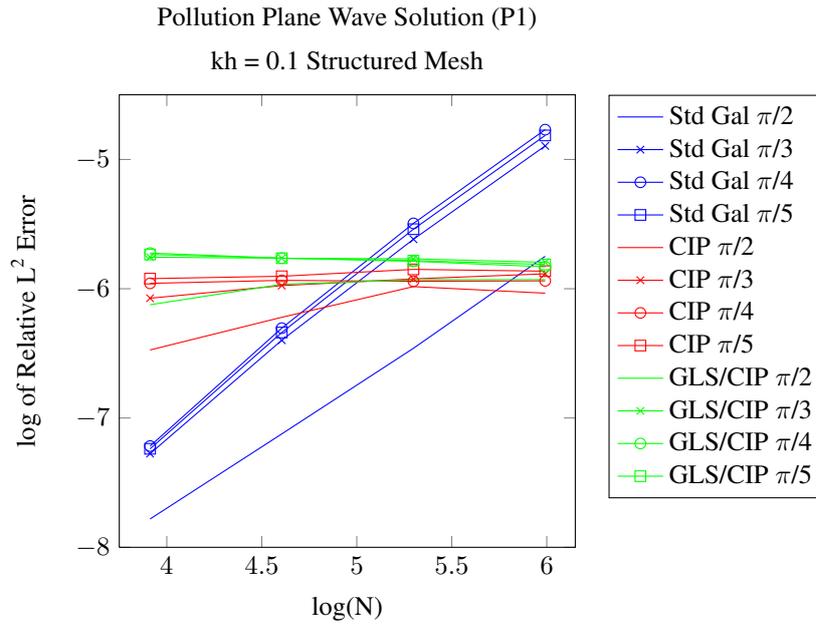


Figure 2.12: Pollution error using piecewise linear elements on a structured mesh where $kh = 0.1$

The figures would suggest that an appropriate choice of parameters can help reduce the pollution error for specific problems on a structured mesh. However, it is unsurprising that the same reduction is not present in the case of the unstructured mesh. Figures 2.14 and 2.15 show that the stabilized methods do not decrease the pollution error on the unstructured mesh. This seems logical since the pollution error in this case is caused by the wave number of the exact solution k not matching the discrete wave number k_h . This causes the numerical solution to become out of phase with the exact solution and leads to an error. For a more detailed discussion on this see [1]. On a structured mesh the stabilization parameters all act in the same way on the numerical solution and can adjust the discrete wave number to match that of the exact solution. In the case of the unstructured mesh the wave cuts the mesh at a variety of different angles and therefore it would seem less likely that the stabilizations would be able to alleviate the pollution effect for constant stabilization parameters.

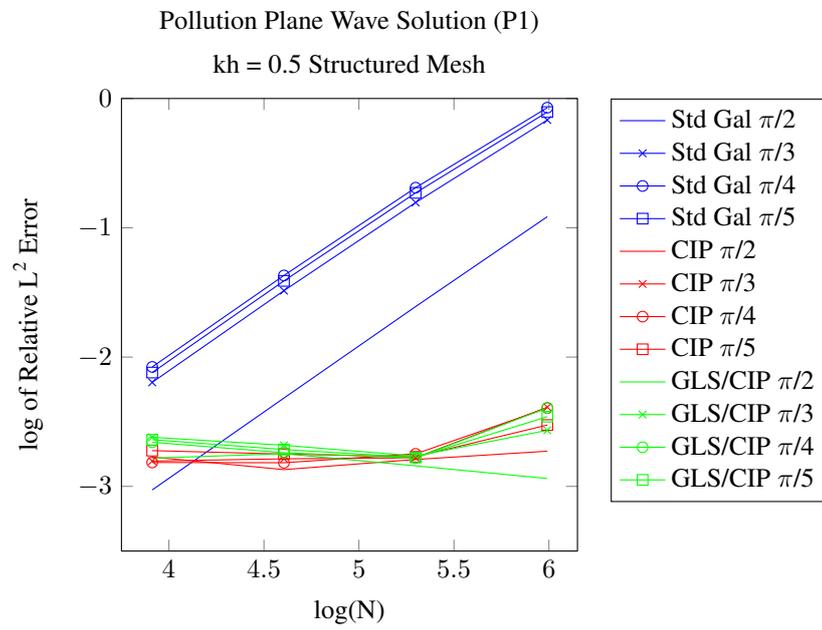


Figure 2.13: Pollution error using piecewise linear elements on a structured mesh where $kh = 0.5$

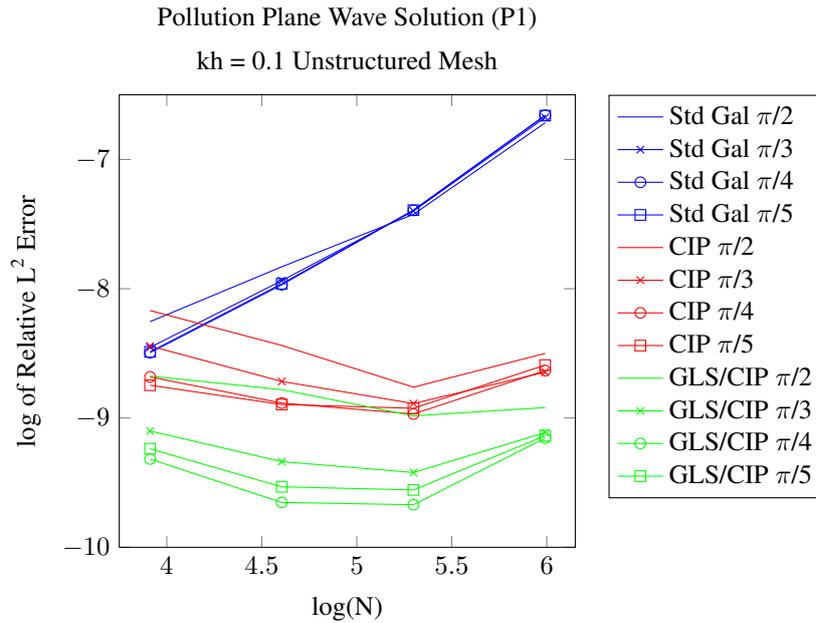


Figure 2.14: Pollution error using piecewise linear elements on an unstructured mesh where $kh = 0.1$

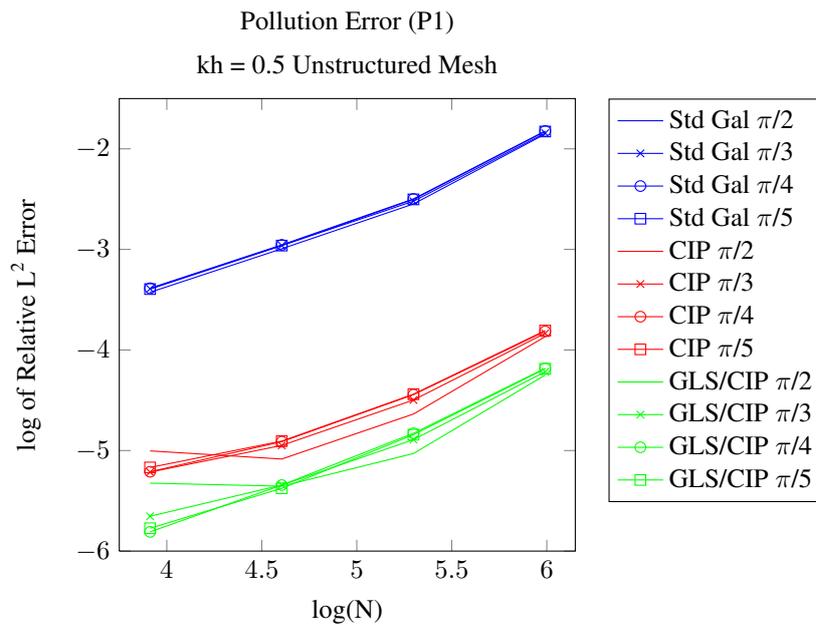


Figure 2.15: Pollution error using piecewise linear elements on an unstructured mesh where $kh = 0.5$

Figure 2.16 shows the convergence for all of the methods for piecewise quadratic elements at low wave number $k = 10$. The methods all appear to converge optimally in accordance with the theory and are almost indistinguishable from each other.

Figure 2.17 shows the convergence of the methods for $k = 100$ here the stabilized methods provide a better approximation than the standard Galerkin method but not by the same margin as for the piecewise linear elements.

Figure 2.18 shows that all the methods obtain the expected convergence for piecewise quadratic elements

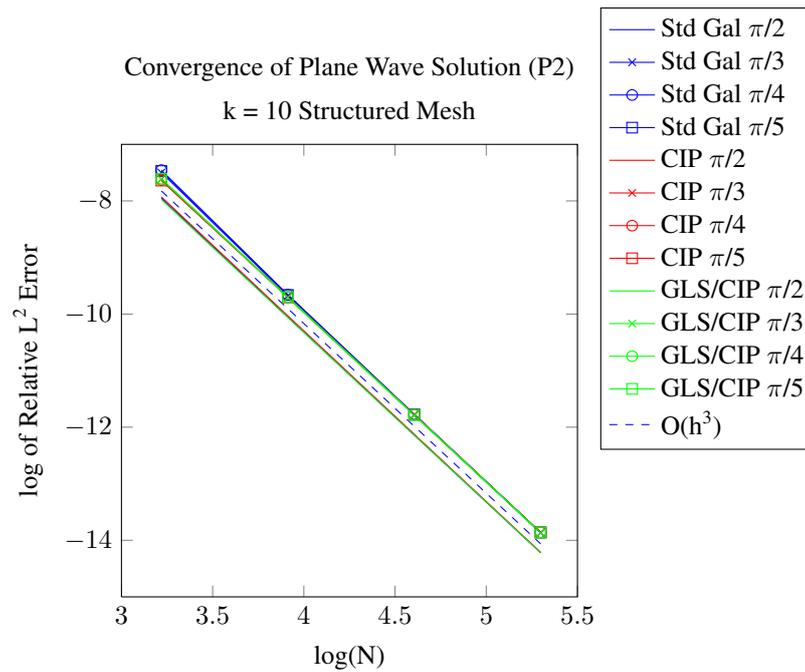


Figure 2.16: Convergence plot for the plane wave solution using piecewise quadratic elements on a structured mesh with $k = 10$

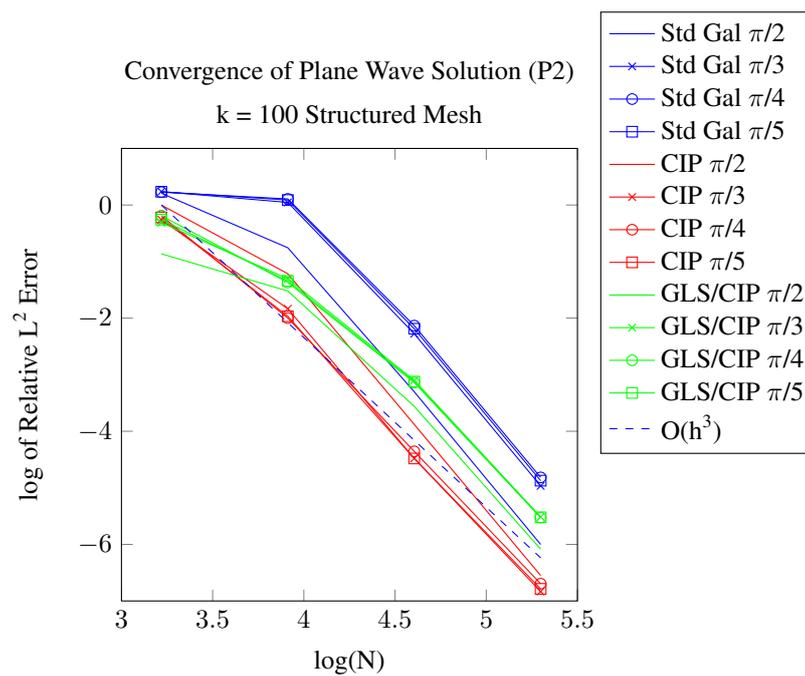


Figure 2.17: Convergence plot for the plane wave solution using piecewise quadratic elements on a structured mesh with $k = 100$

on an unstructured mesh for low wavenumber $k = 10$ and Figure 2.19 shows the pollution effect in the pre-asymptotic regime.

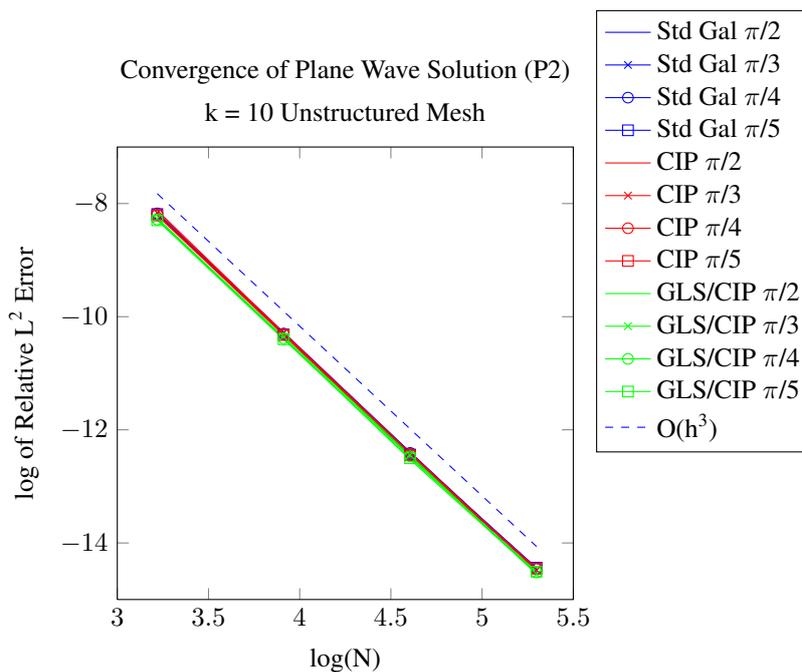


Figure 2.18: Convergence plot for the plane wave solution using piecewise quadratic elements on an unstructured mesh with $k = 10$

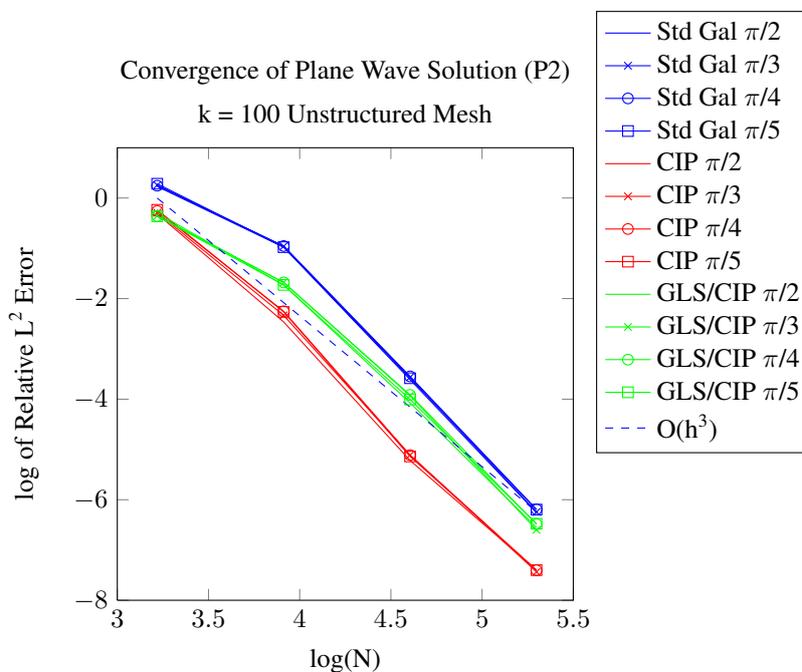


Figure 2.19: Convergence plot for the plane wave solution using piecewise quadratic elements on an unstructured mesh with $k = 100$

Figure 2.20 shows how the methods handle the pollution error for $kh = 0.2$ on structured mesh. The GLS/CIP method seems less affected by the pollution error than the standard Galerkin method and the CIP method appears to be pollution free. Figure 2.21 shows how the methods are able to control pollution for a more practical choice of $kh = 1.0$. Here the pollution error is clearly present for the standard Galerkin and the GLS/CIP methods. On the other hand, the CIP stabilization seems to damp the effects of the numerical pollution for the given choice of parameters.

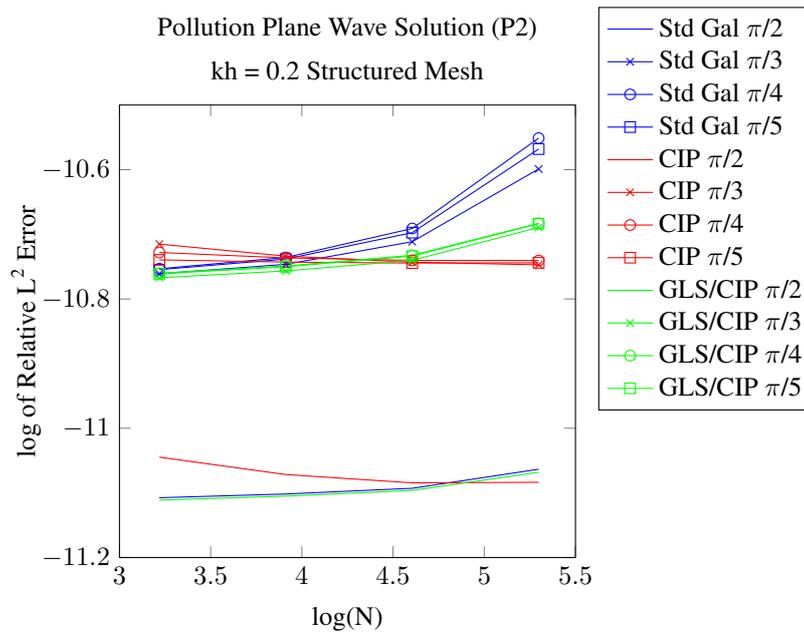


Figure 2.20: Pollution error using piecewise quadratic elements on a structured mesh where $kh = 0.2$

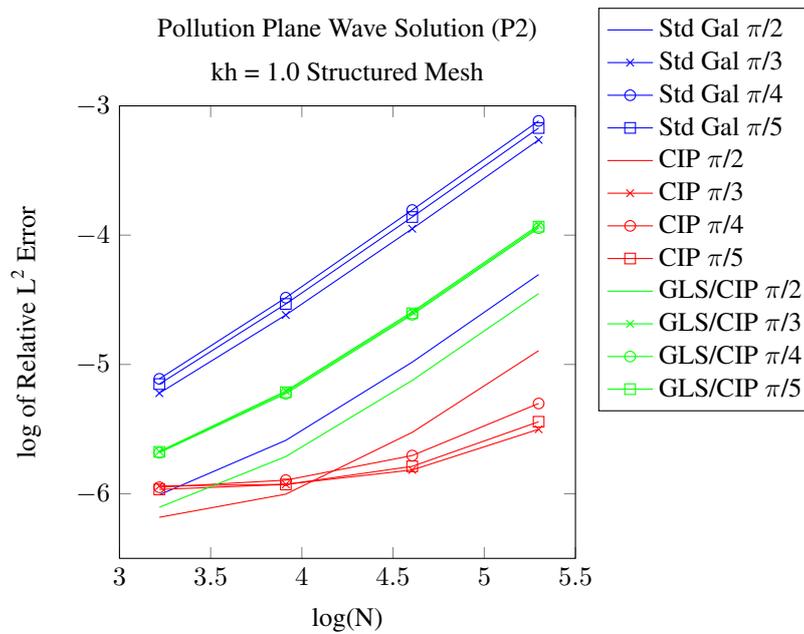


Figure 2.21: Pollution error using piecewise quadratic elements on a structured mesh where $kh = 1.0$

An interesting observation can be seen when applying our piecewise quadratic numerical methods to

unstructured meshes. Figures 2.22, 2.23 and 2.24 all exhibit strange behaviours. Figure 2.22 shows the methods converging as the mesh is refined holding kh constant. This is better than can be expected and it is unclear why this is the case. Taking kh closer to the engineering rule of thumb shows the standard Galerkin method acting as would be expected and the GLS/CIP method providing a more accurate but still polluted approximation of the exact solution. Interestingly the CIP method still converges in this case and the same can be said about Figure 2.24.

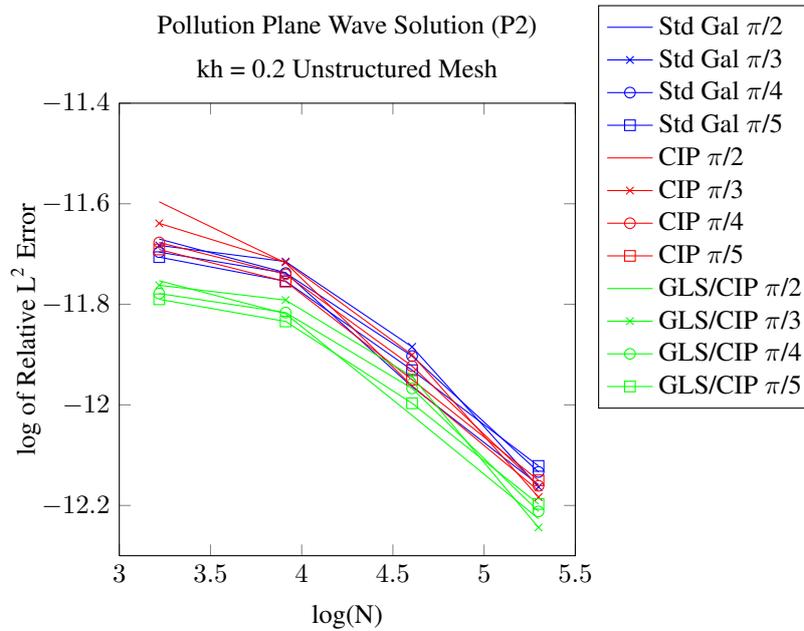


Figure 2.22: Pollution error using piecewise quadratic elements on an unstructured mesh where $kh = 0.2$

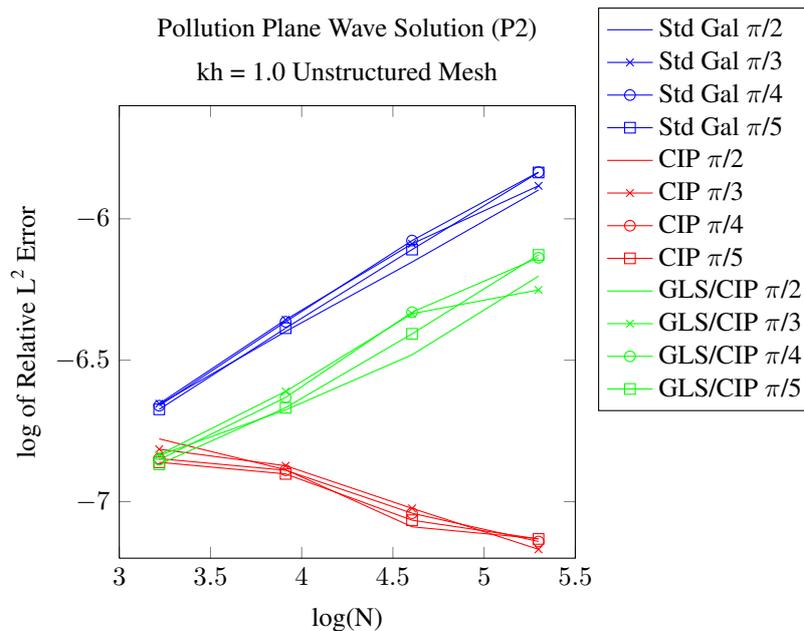


Figure 2.23: Pollution error using piecewise quadratic elements on an unstructured mesh where $kh = 1.0$

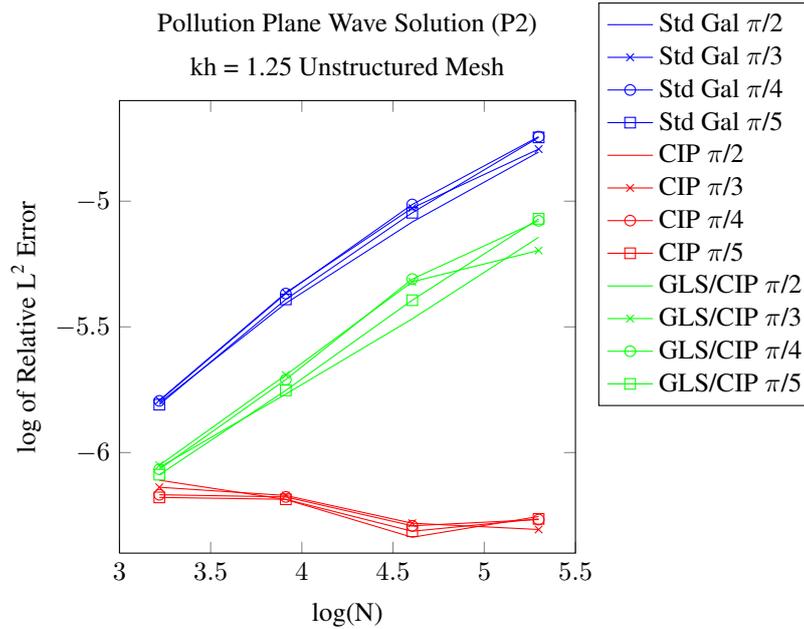


Figure 2.24: Pollution error using piecewise quadratic elements on an unstructured mesh where $kh = 1.25$

2.4.2.1 Sensitivity study of parameters

To study the sensitivity of the parameters I vary them individually by minimizing the error in the L^2 norm between the FE solution u_h and the exact solution u . I vary the parameters one at a time as this helps to give an indication of how sensitive the parameters are to the wave direction.

I begin by considering the GLS/CIP stabilization. Note that in the case studied in this section the theory states that the imaginary parts of the parameters should be less than zero since they should be chosen to match the sign in the Robin boundary condition. The first parameter I consider is the imaginary part of the GLS parameter $\delta_{1,\tau}$.

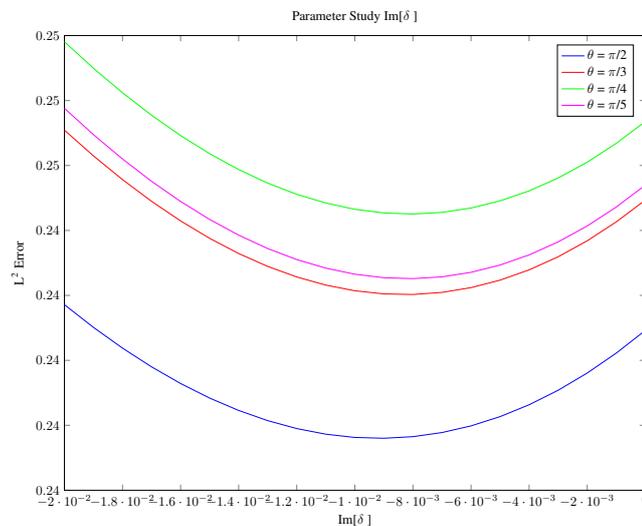


Figure 2.25: Parameter sensitivity study for $Im[\delta_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

Figure 2.25 shows distinct minima for each angle θ . The steepness of the curves gives an indication as to the sensitivity of the parameters. The choice of parameter that I have chosen is close enough to the minimum of the curves to still be accurate but is has been picked to reduce numerical pollution. In Figure 2.26 I consider the imaginary part of the CIP parameter $\gamma_{1,\tau}$. Here the parameter is much less sensitive and allows the user more freedom to vary the parameter and retain accuracy. In fact after considering Figures 2.27 and 2.28 it would seem that the choice of CIP parameter has much less impact on the error of the solution than its GLS counterpart. In practice this is much more desirable since it is very rare that you know the optimal choice of parameter for a particular problem a priori. Figures 2.29 and 2.30 seem to show that the method is rather sensitive to the choice of $Im[\beta_1]$ for the case $\theta = \pi/2$. The analysis suggests that this term is crucial to ensure the stability of the method.

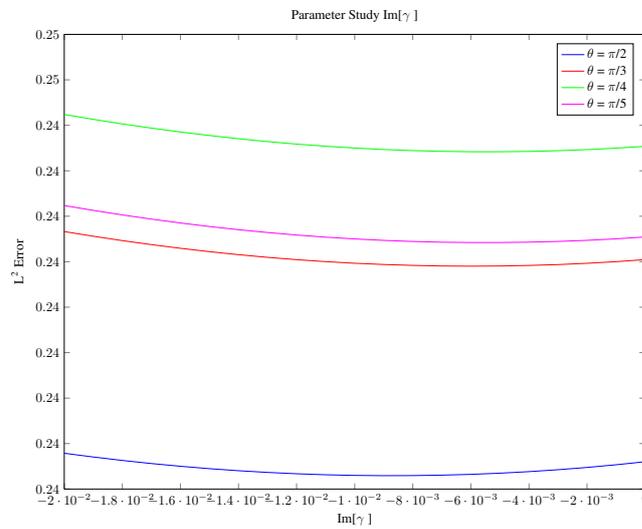


Figure 2.26: Parameter sensitivity study for $Im[\gamma_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

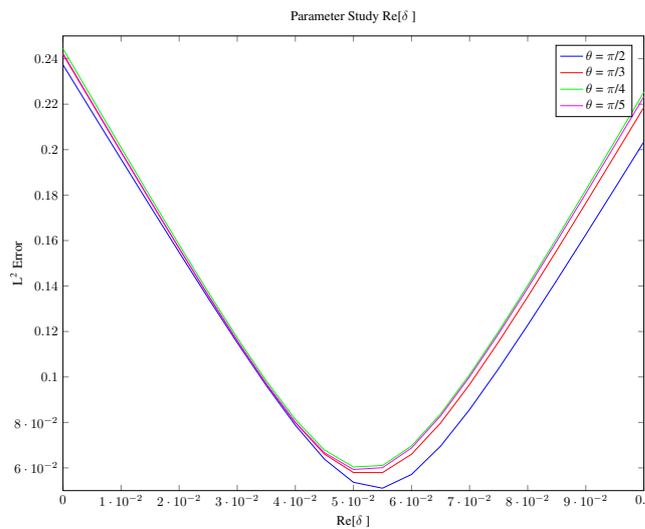


Figure 2.27: Parameter sensitivity study for $Re[\delta_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

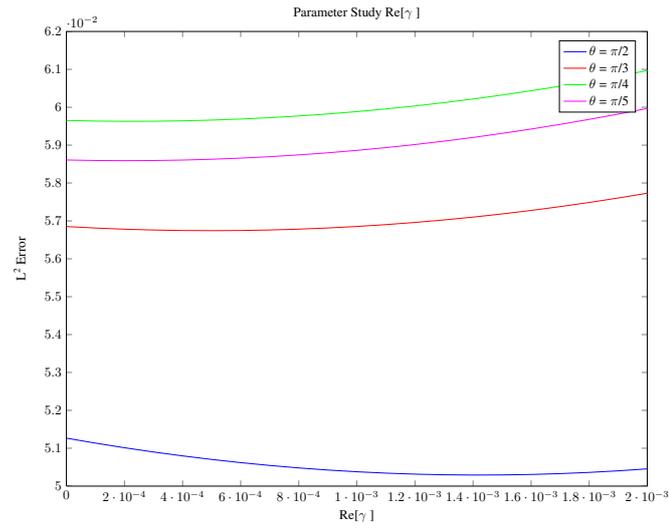


Figure 2.28: Parameter sensitivity study for $Re[\gamma_{1,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

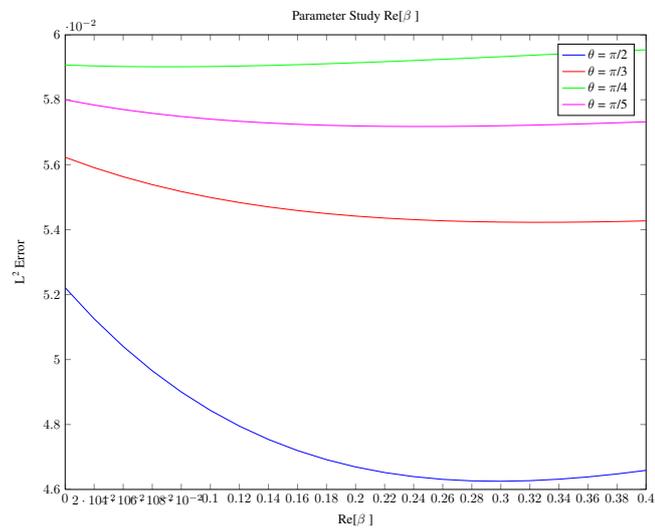


Figure 2.29: Parameter sensitivity study for $Re[\beta_1]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

Figures 2.31-2.34 show that without the GLS term the choice of CIP parameter becomes more sensitive. This can be attributed to the fact that for piecewise linear elements the CIP parameter is the only stabilization term present in the bulk.

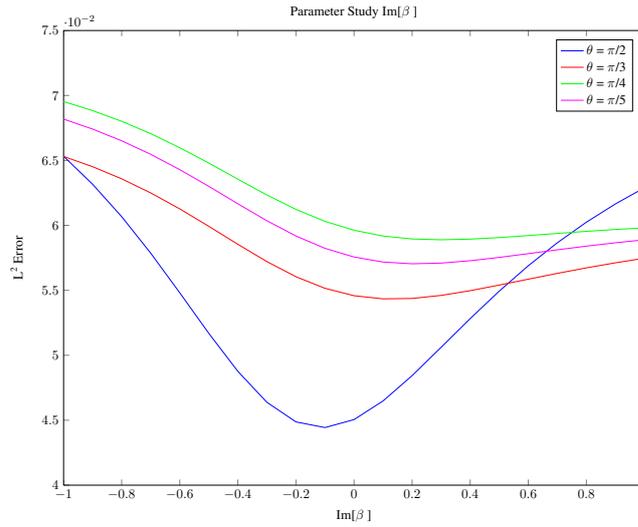


Figure 2.30: Parameter sensitivity study for $Im[\beta_1]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

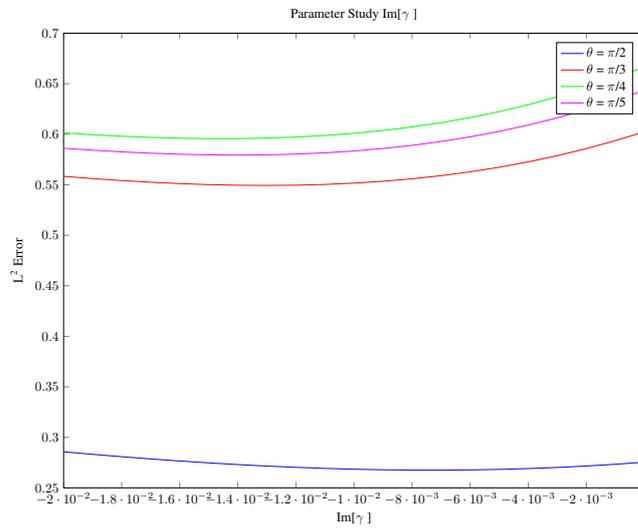


Figure 2.31: Parameter sensitivity study for $Im[\gamma_{2,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

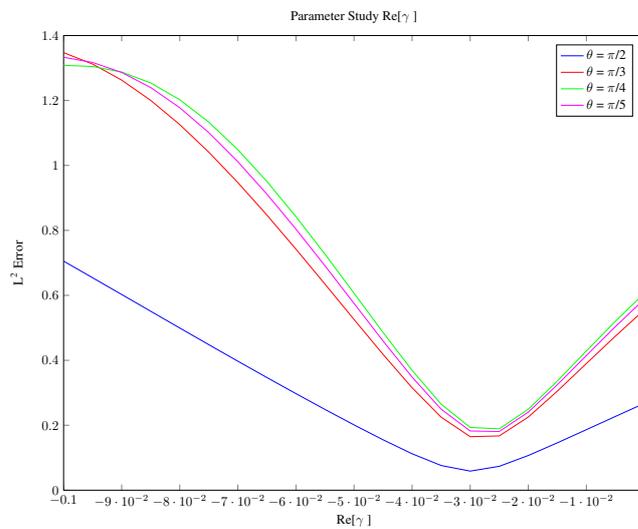


Figure 2.32: Parameter sensitivity study for $Re[\gamma_{2,\tau}]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

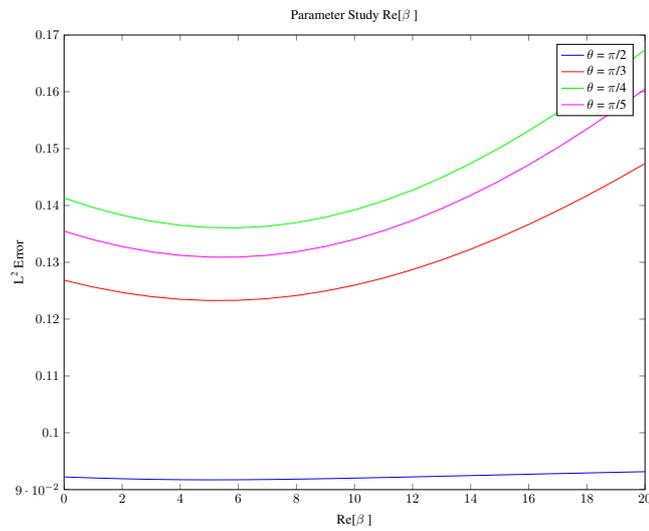


Figure 2.33: Parameter sensitivity study for $Re[\beta_2]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

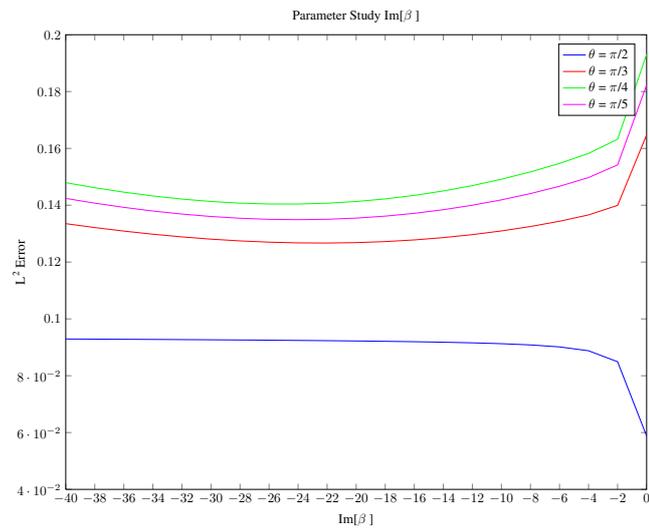


Figure 2.34: Parameter sensitivity study for $Im[\beta_2]$ using piecewise linear elements where $kh = \frac{1}{\sqrt{2}}$

Chapter 3

Fictitious domain methods for Helmholtz equation using cut elements

3.1 Introduction

In problems of practical importance the boundary for a given domain is often determined by data or some type of Computer Aided Design (CAD) software. This can lead to domains having complex geometries that can be costly to mesh. In order to avoid such costs a variety of methods have been introduced which aim to decouple the geometry of the physical domain from that of the background mesh. The Fictitious Domain (FiDo) method was first introduced in a FE framework by Glowinski in [26]. Previously, FiDo methods had been used numerically to extend the capabilities of the Finite Difference method by enabling applications to complex geometries. The method introduced by Glowinski used Lagrange multipliers to enforce the Dirichlet boundary conditions for the physical problem. The resulting saddle point problem was then solved iteratively using a conjugate gradient algorithm. The main motivations behind introducing fictitious domain methods in a FE setting are:

- The mesh does not have to fit the boundary of the physical domain which reduces the cost of calculations.
- The extended domain can be chosen such that the resultant mesh is ‘simple’ which can benefit the user when solving the linear system.
- The extended domain can be chosen to be fixed in time dependent problems even when the physical domain is allowed to evolve with time.

For the reasons stated above, fictitious domain methods have been of great interest when solving problems involving fluid structure interaction. Classical FiDo methods define a continuous extension of the underlying problem from the physical domain onto the computational domain. These classical techniques typically have poor accuracy due to the lack of approximation properties over the physical do-

main. In more recent work by Burman and Hansbo [20],[21] techniques were introduced that allow the solution to become discontinuous inside each element that is intersected by the boundary. In these papers the authors allow the boundary of the physical domain to arbitrarily cut the background mesh using Lagrange multipliers or Nitsche's method to weakly enforce Dirichlet boundary conditions. The problem is only evaluated on the physical domain which introduces cut elements leading to the name CutFEM methods. A key result in these papers relies on an important observation by Stein in [51] which I will discuss later in the chapter. The main fictitious domain techniques that have been proposed with cut elements are:

- Lagrange Multiplier methods ([26], [20])
- XFEM ([35], [24])
- Nitsche's method ([21],[14])

One thing that all of these methods have in common is the fact that they all impose boundary conditions using the weak formulation rather than building them into the solution space. One inherent advantage to this is that, since the solution space in this framework is much larger, the user is less likely to run into computational problems such as numerical locking when performing simulations. In this chapter I will be introducing a Nitsche-based fictitious domain method. The main advantages that Nitsche's method has over XFEM and Lagrange multipliers is that it is easy to implement and analyse. An advantage that this method has over existing fictitious domain methods designed for Helmholtz equation is that it does not require the introduction of Lagrange multipliers which add to the complexity of the method. Lagrange multiplier methods rely on introducing additional unknowns, increasing the complexity of the linear system and leading to saddle point problems requiring a careful choice of discrete spaces to ensure stability. Nitsche's method does not introduce any additional unknowns and can be solved directly. XFEM relies on introducing additional enrichments to the FE space, which can be difficult to construct if the spaces are to remain $H^1(\Omega)$ -conforming and have optimal approximation properties.

In this chapter I will first extend the stabilized methods introduced previously using Nitsche's method to consider general boundary conditions. The use of Nitsche's method is a popular alternative to classical Lagrange multiplier methods as it is a consistent, optimally convergent method that does not introduce extra variables to be solved for. I will show that the new methods are optimally convergent using a similar analysis to what was presented in the previous chapter. Then, numerical studies are presented to demonstrate the potential of these methods for reducing the pollution effect. Following the discussion of generalized boundary conditions, I introduce a new fictitious domain method for the solution of Helmholtz equation using the CutFEM technique introduced by Burman and Hansbo in [20] and [21]. After introducing some important notation and concepts, I analyse the new method and show that it is

optimally convergent. Finally, I present a numerical study to compare the stabilized fictitious domain methods introduced with a fictitious domain method recently proposed in [58].

3.2 Generalized boundary conditions

The boundary conditions for any PDE contain important information regarding the physical problem being solved. It is therefore an added benefit to have reliable numerical methods that are capable of handling generalized boundary conditions. When considering the problem of acoustic scattering the boundary conditions describe the physical properties of the scatterer. Dirichlet boundary conditions are used to model a surface that has very low acoustic impedance compared to the acoustic impedance of the carrier medium. This type of surface is called sound soft. Conversely, Neumann boundary conditions are used to model a surface that has a very high acoustic impedance compared to the acoustic impedance of the carrier medium. This type of surface is referred to as being sound hard. The Robin condition can be thought of as a blend of the pair where the acoustic impedance of the surface is of similar magnitude to that of the carrier medium. To ensure that the problem is well posed, it is necessary to also include the Sommerfeld Radiation condition which ensures that no energy is allowed to enter the system from infinity. This will once again be modelled using the first order approximation given as a Robin boundary condition.

Our problem is then the following

$$\left\{ \begin{array}{lll} \mathcal{L}(u) = f & \text{in } \Omega \\ u = g_D & \text{on } \partial\Omega_D \\ \nabla u \cdot n = g_N & \text{on } \partial\Omega_N \\ \mathcal{R}(u) = g_R & \text{on } \partial\Omega_R. \end{array} \right. \quad (3.1)$$

For the subsequent analysis it is useful to also consider the adjoint problem

$$\left\{ \begin{array}{lll} \mathcal{L}(z) = \psi & \text{in } \Omega \\ z = 0 & \text{on } \partial\Omega_D \\ \nabla z \cdot n = 0 & \text{on } \partial\Omega_N \\ \mathcal{R}^*(z) = 0 & \text{on } \partial\Omega_R. \end{array} \right. \quad (3.2)$$

To ensure well posedness of the problem I assume that $\text{meas}_{d-1}(\partial\Omega_R) > 0$.

Let Ω be a bounded domain with smooth boundary or a bounded convex domain such that there exists an $x_0 \in \Omega$ such that

$$(x - x_0) \cdot n \leq 0 \quad \forall x \in \partial\Omega_D \quad (3.3)$$

$$(x - x_0) \cdot n = 0 \quad \forall x \in \partial\Omega_N \quad (3.4)$$

$$(x - x_0) \cdot n \geq C \quad \forall x \in \partial\Omega_R \quad (3.5)$$

the solution of Helmholtz equation is known to satisfy the following bounds for any $f \in L^2(\Omega)$, $g_D \in H^{1/2}(\partial\Omega_D)$, $g_N \in H^{3/2}(\partial\Omega_N)$ and $g_R \in H^{3/2}(\partial\Omega_R)$ a proof can be found in [36].

$$k\|u\|_{0,\Omega} + \|u\|_{1,\Omega} \lesssim \mathcal{C}_{f,g}, \quad \|u\|_{2,\Omega} \lesssim (k+1)\mathcal{C}'_{f,g}, \quad (3.6)$$

where $\mathcal{C}_{f,g} = \|f\|_{0,\Omega} + \|g_R\|_{0,\partial\Omega_R} + \|g_N\|_{0,\partial\Omega_N} + k^{3/2}\|g_D\|_{0,\partial\Omega_D}$ and $\mathcal{C}'_{f,g} = \|f\|_{0,\Omega} + \|g_R\|_{1/2,\partial\Omega_R} + \|g_N\|_{1/2,\partial\Omega_N} + k\|g_D\|_{1/2,\partial\Omega_D}$.

Under the same assumptions the adjoint problem satisfies similar bounds

$$k\|z\|_{0,\Omega} + \|z\|_{1,\Omega} \lesssim \|\psi\|_{0,\Omega}, \quad \|z\|_{2,\Omega} \lesssim k\|\psi\|_{0,\Omega}. \quad (3.7)$$

Using standard FE techniques a problem posed with these boundary conditions would usually be solved by imposing the Dirichlet boundary conditions as essential boundary conditions. This means that the space V_0 , in which solutions are sought, is restricted to functions that satisfy the Dirichlet boundary conditions as given in (3.1) on $\partial\Omega_D$. The natural choice of V_0 would be:

$$V_0 \stackrel{\text{def}}{=} \{v \in H^1(\Omega) : v = 0 \text{ on } \partial\Omega_D\}.$$

The Neumann and Robin conditions are said to be imposed as natural boundary conditions and are implemented in the weak formulation. The idea of Nitsche's method is to impose the Dirichlet boundary conditions using the weak formulation and hence remove the restriction on the solution space V i.e.

$$V \stackrel{\text{def}}{=} \{v \in H^1(\Omega)\}.$$

Nitsche proposed enforcing the Dirichlet boundary conditions using a penalty method. To show this I will first derive the weak formulation. The problem becomes: Find $u \in V$ such that

$$a(u, v) = l(v) \forall v \in V.$$

where

$$a(u, v) = (\nabla u, \nabla v)_{0,\Omega} - k^2(u, v)_{0,\Omega} - \langle \nabla u \cdot n, v \rangle_{0,\partial\Omega}.$$

and

$$l(v) = (f, v)_{0,\Omega}.$$

Let \mathcal{T}_h be a quasi-uniform triangulation of the domain Ω where quasi-uniformity is as defined in (A.4) and

$$V_h \stackrel{\text{def}}{=} \{v_h \in V : v_h|_{\tau} \in \mathbb{P}^p(\tau) \forall \tau \in \mathcal{T}_h, p \in \mathbb{N}\}.$$

The finite element method writes: Find $u_h \in V_h$ such that

$$a(u_h, v_h) = l(v_h) \forall v_h \in V_h.$$

After deriving the weak formulation in the usual way and moving to a discrete setting, the symmetric method of Nitsche then proposes the addition of two extra terms in order to enforce the Dirichlet boundary condition. In the classical formulation proposed by Nitsche a symmetry term is introduced to preserve the symmetry of the method whilst also acting as a residual term to impose the Dirichlet boundary condition. Although this term preserves the symmetry of the method, the analysis shows that to ensure stability a penalty term, $\alpha \langle u_h, v_h \rangle_{0, \partial\Omega_D}$, must also be added. In recent work it has been shown that this term can be omitted by changing the sign of the symmetry term although I will not consider that case in this work.

For the problem to scale correctly and for the method to enter the analysis of Theorem 2 the penalty parameter should be taken such that $\alpha = i\beta_D h^{-1}$, where $\beta_D \in \mathbb{R}^+$. Notice that the method is consistent since $u = g_D$ on $\partial\Omega_D$.

$$\begin{aligned} & (\nabla u_h, \nabla v_h)_{0, \Omega} - k^2 (u_h, v_h)_{0, \Omega} - \underbrace{\langle \nabla u_h \cdot n, v_h \rangle_{0, \partial\Omega_D}}_{\text{consistency}} - \underbrace{\langle u_h, \nabla v_h \cdot n \rangle_{0, \partial\Omega_D}}_{\text{symmetry}} \quad (3.8) \\ & + \underbrace{\alpha \langle u_h, v_h \rangle_{0, \partial\Omega_D}}_{\text{penalty}} + \langle iku_h, v_h \rangle_{0, \partial\Omega_R} \\ = & l(v_h) + \alpha \langle g_D, v_h \rangle_{0, \partial\Omega_D} + \langle g_D, \nabla v_h \cdot n \rangle_{0, \partial\Omega_D} + \langle g_N, v_h \rangle_{0, \partial\Omega_N} + \langle g_R, v_h \rangle_{0, \partial\Omega_R}. \quad (3.9) \end{aligned}$$

It can be seen that the problem still conforms to the same abstract framework as before. Except in this case I would like to pose the stabilization form in two parts; one that describes the stabilization on the interior of the domain and the other that describes the stabilization on the boundary. It still follows that

$$A_h(u_h, v_h) = A(u_h, v_h) + s(u_h, v_h) = L(v_h) + s(u, v_h) = L_h(v_h). \quad (3.10)$$

However in this case

$$A_h(u_h, v_h) = a_h(u_h, v_h) - k^2 (u_h, v_h)_{0, \Omega} + \langle iku_h, v_h \rangle_{0, \partial\Omega_R},$$

and

$$s(u_h, v_h) = s_\Omega(u_h, v_h) + s_{\partial\Omega}(u_h, v_h), \quad (3.11)$$

where

$$a_h(u_h, v_h) = (\nabla u_h, \nabla v_h)_{0, \Omega} - \langle \nabla u_h \cdot n, v_h \rangle_{0, \partial\Omega_D} - \langle u_h, \nabla v_h \cdot n \rangle_{0, \partial\Omega_D},$$

and $s_\Omega(u_h, v_h)$ represents a stabilization over the interior of the domain, and $s_{\partial\Omega}(u_h, v_h)$ is the stabilization acting on the boundary. Again I consider two cases for the stabilization over the domain, namely the GLS/CIP and CIP methods proposed previously. The GLS/CIP stabilization is given by

$$s_\Omega^G(u_h, v_h) \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_{1,\tau} (\mathcal{L}(u_h), \mathcal{L}(v_h))_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} h_\tau \gamma_{1,\tau} \langle \llbracket \nabla u_h \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_{0,F}, \quad (3.12)$$

and the CIP stabilization is given by

$$s_\Omega^C(u_h, v_h) \stackrel{\text{def}}{=} \sum_{F \in \mathcal{F}_{int}} h_\tau^3 \delta_{2,\tau} \langle \llbracket \Delta u_h \rrbracket, \llbracket \Delta v_h \rrbracket \rangle_{0,F} + \sum_{F \in \mathcal{F}_{int}} h_\tau \gamma_{1,\tau} \langle \llbracket \nabla u_h \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_{0,F}. \quad (3.13)$$

The stabilization over the boundary shall be defined as

$$\begin{aligned} s_{\partial\Omega}(u_h, v_h) &= i\beta_N h \langle \nabla u_h \cdot n, \nabla v_h \cdot n \rangle_{\partial\Omega_N} + i\beta_R h \langle \mathcal{R}(u_h), \mathcal{R}^*(v_h) \rangle_{\partial\Omega_R} \\ &\quad + i\beta_D h^{-1} \langle u_h, v_h \rangle_{\partial\Omega_D}. \end{aligned}$$

I have also introduced a least squares penalty term for the Neumann boundary condition. This addition means that the proposed method will once again fit into the framework of Theorem 2 for $\beta_{j,D}, \beta_{j,N}, \beta_{j,R} \in \mathbb{R}^+$ where $j \in \{1, 2\}$. For the remainder of this section the subscript j is used to distinguish between the different coefficients used in the different stabilized methods, $j = 1$ is used for GLS/CIP stabilization and $j = 2$ for CIP method. As before I will show that the stabilized method satisfies all of the assumptions of the theorem by showing that the interpolation estimates hold and by proving the continuity of the form. Some of the results carry over from the analysis of the stabilizations with just Robin boundary conditions and to avoid repetition I shall just state where I have used a previous result.

For the following analysis it is useful to introduce the following definitions. Let $u_h \in V_h$

$$\begin{aligned} |u_h|_{3,G}^2 &\stackrel{\text{def}}{=} k(1 - \beta_{1,R}kh) \|u_h\|_{0,\partial\Omega_R}^2 + \max_{\tau \in \mathcal{T}_h} \{Im[\delta_{1,\tau}]\} h^2 \sum_{\tau \in \mathcal{T}_h} \|\mathcal{L}(u_h)\|_{0,\tau}^2 \\ &\quad + \max_{\tau \in \mathcal{T}_h} \{Im[\gamma_{1,\tau}]\} h \sum_{F \in \mathcal{F}_{int}} \|\llbracket \nabla u_h \rrbracket\|_{0,F}^2 + \beta_{1,D} h^{-1} \|u_h\|_{0,\partial\Omega_D}^2 \\ &\quad + \beta_{1,N} h \|\nabla u_h \cdot \mathbf{n}\|_{0,\partial\Omega_N}^2 + \beta_{1,R} h \|\nabla u_h \cdot \mathbf{n}\|_{0,\partial\Omega_R}^2, \end{aligned} \quad (3.14)$$

and

$$\begin{aligned}
|u_h|_{3,C}^2 &\stackrel{\text{def}}{=} k(1 - \beta_{2,R}kh)\|u_h\|_{0,\partial\Omega_R}^2 + \max_{\tau \in \mathcal{T}_h} \{Im[\delta_{2,\tau}]\} h^3 \sum_{\tau \in \mathcal{T}_h} \|[\Delta u_h]\|_{0,\tau}^2 \\
&+ \max_{\tau \in \mathcal{T}_h} \{Im[\gamma_{2,\tau}]\} h \sum_{F \in \mathcal{F}_{int}} \|[\nabla u_h]\|_{0,F}^2 + \beta_{2,D} h^{-1} \|u_h\|_{0,\partial\Omega_D}^2 \\
&+ \beta_{2,N} h \|\nabla u_h \cdot \mathbf{n}\|_{0,\partial\Omega_N}^2 + \beta_{2,R} h \|\nabla u_h \cdot \mathbf{n}\|_{0,\partial\Omega_R}^2. \tag{3.15}
\end{aligned}$$

Lemma 3.1. For $0 < \beta_{1,R}kh < 1$ and $\max_{\tau \in \mathcal{T}_h} \{Im[\delta_{1,\tau}]\}, \max_{\tau \in \mathcal{T}_h} \{Im[\gamma_{1,\tau}]\}, \beta_{1,D}, \beta_{1,N}, \beta_{1,R} > 0$ it follows that $|u_h|_{3,G}^2$, as defined in (3.14), is a norm on V_h .

Proof. The proof follows directly from the definition. \square

Lemma 3.2. For $0 < \beta_{2,R}kh < 1$ and $\max_{\tau \in \mathcal{T}_h} \{Im[\delta_{2,\tau}]\}, \max_{\tau \in \mathcal{T}_h} \{Im[\gamma_{2,\tau}]\}, \beta_{2,D}, \beta_{2,N}, \beta_{2,R} > 0$ it follows that $|u_h|_{3,C}^2$, as defined in (3.15) is a semi-norm on V_h .

Proof. The proof follows directly from the definition. \square

It is also useful to introduce the two additional semi-norms which are associated to the stabilizations

$|\cdot|_{s,G} : V_h \mapsto \mathbb{R}$ for GLS/CIP and $|\cdot|_{s,C} : V_h \mapsto \mathbb{R}$ for CIP which are defined as

$$\begin{aligned}
|\cdot|_{s,G}^2 &\stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} |\delta_{1,\tau}| h_\tau^2 \|\mathcal{L}(\cdot)\|_{0,\tau}^2 + \sum_{F \in \mathcal{F}_{int}} |\gamma_{1,\tau}| h_\tau \|[\nabla(\cdot) \cdot \mathbf{n}]\|_{0,F}^2 + |\beta_{1,D}| h^{-1} \|(\cdot)\|_{0,\partial\Omega_D}^2 \\
&+ |\beta_{1,N}| h \|\nabla(\cdot) \cdot \mathbf{n}\|_{0,\partial\Omega_N}^2 + |\beta_{1,R}| h \|\mathcal{R}^*(\cdot)\|_{0,\partial\Omega_R}^2. \tag{3.16}
\end{aligned}$$

and

$$\begin{aligned}
|\cdot|_{s,C}^2 &\stackrel{\text{def}}{=} \sum_{F \in \mathcal{F}_{int}} |\delta_{2,\tau}| h_\tau^3 \|[\Delta(\cdot)]\|_{0,F}^2 + \sum_{F \in \mathcal{F}_{int}} |\gamma_{2,\tau}| h_\tau \|[\nabla(\cdot) \cdot \mathbf{n}]\|_{0,F}^2 + |\beta_{2,D}| h^{-1} \|(\cdot)\|_{0,\partial\Omega_D}^2 \\
&+ |\beta_{2,N}| h \|\nabla(\cdot) \cdot \mathbf{n}\|_{0,\partial\Omega_N}^2 + |\beta_{2,R}| h \|\mathcal{R}^*(\cdot)\|_{0,\partial\Omega_R}^2. \tag{3.17}
\end{aligned}$$

Lemma 3.3. Under the assumption $hk < \beta_R^{-1}$ the Nitsche formulation given in (3.10) satisfies assumptions (2.23), (2.24), (2.29), (2.30) and (2.31) for

$$\begin{aligned}
\|(\cdot)\|_{*,G}^2 &\stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} |\delta_{1,\tau}|^{-1} h_\tau^{-2} \|(\cdot)\|_{0,\tau}^2 + \sum_{F \in \mathcal{F}_{int}} (|\gamma_{1,\tau}| h_\tau)^{-1} \|(\cdot)\|_{0,F}^2 + (|\beta_{1,N}| h)^{-1} \|(\cdot)\|_{0,\partial\Omega_N}^2 \\
&+ |\beta_{1,D}|^{-1} h \|\nabla(\cdot) \cdot \mathbf{n}\|_{0,\partial\Omega_D}^2 + (|\beta_{1,R}| h)^{-1} \|(\cdot)\|_{0,\partial\Omega_R}^2 + s_\Omega^G((\cdot), (\cdot)), \tag{3.18}
\end{aligned}$$

and π_h is the standard Lagrange interpolant when the GLS/CIP stabilization, proposed in (2.64), is used

over the interior of the domain Ω or

$$\begin{aligned} \|(\cdot)\|_{*,C}^2 &\stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} |\delta_{2,\tau}|^{-1} h_\tau^{-1} \|(\cdot)\|_{0,\tau}^2 + \sum_{F \in \mathcal{F}_{int}} (|\gamma_{2,\tau} h_\tau|)^{-1} \|(\cdot)\|_{0,F}^2 + (|\beta_{2,N} h|)^{-1} \|(\cdot)\|_{0,\partial\Omega_N}^2 \\ &\quad + |\beta_{2,D}|^{-1} h \|\nabla(\cdot) \cdot n\|_{0,\partial\Omega_D}^2 + (|\beta_{2,R} h|)^{-1} \|(\cdot)\|_{0,\partial\Omega_R}^2 + s_\Omega^C((\cdot), (\cdot)), \end{aligned} \quad (3.19)$$

and π_h is the standard L^2 projection when the CIP stabilization, proposed in (2.95), is used over the interior of the domain Ω .

Proof. The proof is a direct result of the fact that

$$|u_h|_J^2 = \text{Im}[A_h(u_h, u_h)]. \quad (3.20)$$

Recall that

$$\text{Im}[A_h(u_h, u_h)] = \text{Im}[A(u_h, u_h)] + \text{Im}[s_\Omega(u_h, u_h)] + \text{Im}[s_{\partial\Omega}(u_h, u_h)]. \quad (3.21)$$

Analysing term by term shows that $\text{Im}[A(u_h, u_h)] = k \|u_h\|_{0,\partial\Omega_R}^2$

$$\begin{aligned} \text{Im}[A(u_h, u_h)] &= \text{Im} \left[(\nabla u_h, \nabla u_h)_{0,\Omega} - k^2 (u_h, u_h)_{0,\Omega} - \langle \nabla u_h \cdot n, u_h \rangle_{0,\partial\Omega_D} \right. \\ &\quad \left. - \langle u_h, \nabla u_h \cdot n \rangle_{0,\partial\Omega_D} + ik \langle u_h, u_h \rangle_{0,\partial\Omega_R} \right] \\ &= \text{Im} \left[\|\nabla u_h\|_{0,\Omega}^2 - k^2 \|u_h\|_{0,\Omega}^2 - \langle \nabla u_h \cdot n, u_h \rangle_{0,\partial\Omega_D} - \langle u_h, \nabla u_h \cdot n \rangle_{0,\partial\Omega_D} \right. \\ &\quad \left. + ik \langle u_h, u_h \rangle_{0,\partial\Omega_R} \right] \\ &= \text{Im} \left[\|\nabla u_h\|_{0,\Omega}^2 - k^2 \|u_h\|_{0,\Omega}^2 - 2\text{Re}[\langle \nabla u_h \cdot n, u_h \rangle_{0,\partial\Omega_D}] + ik \|u_h\|_{0,\partial\Omega_R}^2 \right]. \end{aligned}$$

The last line comes from equation (A:11) in the Appendix. The term $\text{Im}[s_\Omega(u_h, u_h)]$ is straightforward to deduce for both of the interior stabilizations introduced earlier therefore I shall omit this step.

Analysing the final term

$$\begin{aligned} \text{Im}[s_{\partial\Omega}(u_h, u_h)] &= \text{Im} \left[i\beta_N h \langle \nabla u_h \cdot n, \nabla u_h \cdot n \rangle_{0,\partial\Omega_N} + i\beta_R h \langle \mathcal{R}(u_h), \mathcal{R}^*(u_h) \rangle_{0,\partial\Omega_R} \right. \\ &\quad \left. + i\beta_D h^{-1} \langle u_h, u_h \rangle_{0,\partial\Omega_D} \right] \end{aligned} \quad (3.22)$$

$$\begin{aligned} &= \text{Im} \left[i\beta_N h \|\nabla u_h \cdot n\|_{0,\partial\Omega_N}^2 + i\beta_R h \langle \mathcal{R}(u_h), \mathcal{R}^*(u_h) \rangle_{0,\partial\Omega_R} \right. \\ &\quad \left. + i\beta_D h^{-1} \|u_h\|_{0,\partial\Omega_D}^2 \right] \end{aligned} \quad (3.23)$$

$$\begin{aligned} &= \beta_N h \|\nabla u_h \cdot n\|_{0,\partial\Omega_N}^2 + \beta_R h (\|\nabla u_h \cdot n\|_{0,\partial\Omega_R}^2 - k^2 \|u_h\|_{0,\partial\Omega_R}^2) \\ &\quad + \beta_D h^{-1} \|u_h\|_{0,\partial\Omega_D}^2. \end{aligned} \quad (3.24)$$

Adding together all of the terms proves the claim (3.20). Assumption (2.23) follows after noticing

$$\operatorname{Im} [A(u_h, u_h)] < \|A(u_h, u_h)\|. \quad (3.25)$$

The proof of assumption (2.24) follows from noticing that

$$|s(v_h, w_h)| \leq |s_\Omega(v_h, w_h)| + |s_{\partial\Omega}(v_h, w_h)| \quad (3.26)$$

$$\leq |v_h|_{\mathcal{J}} |w_h|_s + |s_{\partial\Omega}(v_h, w_h)|, \quad (3.27)$$

so it is enough to show that $|s_{\partial\Omega}(v_h, w_h)| \leq |v_h|_{\mathcal{J}} |w_h|_s$. Once again this can be shown through an application of the Cauchy-Schwarz inequality followed by an application of the triangle inequality.

$$\begin{aligned} |s_{\partial\Omega}(v_h, w_h)| &\leq |\beta_N| h \|\nabla v_h \cdot n\|_{0, \partial\Omega_N} \|\nabla w_h \cdot n\|_{0, \partial\Omega_N} + \beta_D h^{-1} \|v_h\|_{0, \partial\Omega_D} \|w_h\|_{0, \partial\Omega_D} \\ &\quad + \beta_R h \|\nabla v_h \cdot n\|_{0, \partial\Omega} + k \|v_h\|_{0, \partial\Omega_R} \|\mathcal{R}^*(w_h)\|_{0, \partial\Omega_R} \\ &\leq |v_h|_{\mathcal{J}} |w_h|_s. \end{aligned}$$

Since assumption (2.29) has already been shown for the stabilization terms acting on the interior of the boundary it is enough to show that the additional boundary terms satisfy the assumption. Using the same trick as when dealing with the Robin boundary stabilization previously, an application of the trace inequality and assumptions (2.25) or (2.26) yields the desired result.

It has already been shown that Assumption (2.30) is satisfied by the stabilizations in the previous section. Therefore the result will hold if the additional terms in the norm $\|\cdot\|_*$ satisfy the same bound. This is indeed the case and the proof comes from an application of the trace inequality (A:4) and the interpolation estimates (2.25) or (2.26). The first additional term is the normal derivative on the Dirichlet boundary and is handled in the following way

$$\begin{aligned} |\beta_D|^{-1} h \|\nabla \eta \cdot n\|_{0, \partial\Omega_D}^2 &\leq |\beta_D|^{-1} h \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_D} C_T^2 \left(h_\tau^{-1/2} \|\nabla \eta\|_{0, \tau} + h_\tau^{1/2} |\nabla \eta|_{1, \tau} \right)^2 \\ &\leq |\beta_D|^{-1} h C h^{2p-1} |u|_{p+1, \Omega}^2 \\ &= |\beta_D|^{-1} C h^{2p} |u|_{p+1, \Omega}^2. \end{aligned}$$

The same steps are used in handling the term on the Neumann boundary

$$\begin{aligned} (|\beta_N| h)^{-1} \|\eta\|_{0, \partial\Omega_N}^2 &\leq (|\beta_N| h)^{-1} \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_N} C_T^2 \left(h_\tau^{-1/2} \|\eta\|_{0, \tau} + h_\tau^{1/2} |\eta|_{1, \tau} \right)^2 \\ &\leq (|\beta_N| h)^{-1} C h^{2p+1} |u|_{p+1, \Omega}^2 \\ &= |\beta_N|^{-1} C h^{2p} |u|_{p+1, \Omega}^2. \end{aligned}$$

Finally (2.31) follows from a similar argument. \square

So it has been shown that the interpolation results hold for the Nitsche formulation. It is now sufficient to show that the forms $A(\cdot, \cdot)$ and $A_h(\cdot, \cdot)$ have the continuity properties required by the theorem.

Proposition 4 (Continuity). *Let $u \in H^{p+1}(\Omega)$ for $p > 1/2$ and π_h be either the standard Lagrange interpolant or the standard L^2 orthogonal projection for the GLS/CIP or CIP interior stabilizations respectively. It follows that the discrete sesquilinear form given by (3.10) satisfies assumptions (2.27) and (2.28) for the respective star norms given by (3.18), (3.19).*

Proof. Assumption (2.27) can be satisfied for the GLS/CIP stabilization by letting $\eta = u - \mathcal{I}_h u$ where $\mathcal{I}_h : C^0(\bar{\Omega}) \mapsto V_h$ is the standard Lagrange interpolant. It follows that

$$|A(\eta, v_h)| + |s(\eta, v_h)| \leq |A(\eta, v_h)| + |s_{GLS/CIP}(\eta, v_h)| + |s_{\partial\Omega}(\eta, v_h)|. \quad (3.28)$$

It is immediate from the definitions in (3.14), (3.18) and (3.15), (3.19) that the last two terms can be decomposed in a way that satisfies the continuity assumption for either stabilization. The only question is whether the same can be said about the form $A(\cdot, \cdot)$

$$\begin{aligned} |A(\eta, v_h)| &= |(\nabla\eta, \nabla v_h)_{0,\Omega} - k^2(\eta, v_h)_{0,\Omega} - \langle \nabla\eta \cdot n, v_h \rangle_{0,\partial\Omega_D} - \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\Omega_D} \\ &\quad + ik \langle \eta, v_h \rangle_{0,\partial\Omega_R} | \\ &= | \sum_{\tau \in \mathcal{T}_h} \left((\eta, -\Delta v_h)_{0,\tau} - k^2(\eta, v_h)_{0,\tau} + \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\tau} \right) - \langle \nabla\eta \cdot n, v_h \rangle_{0,\partial\Omega_D} \\ &\quad - \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\Omega_D} + ik \langle \eta, v_h \rangle_{0,\partial\Omega_R} | \\ &= | \sum_{\tau \in \mathcal{T}_h} (\eta, -\Delta v_h - k^2 v_h)_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} \langle \eta, \llbracket \nabla v_h \rrbracket \rangle_{0,F} + \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\Omega_N} \\ &\quad - \langle \nabla\eta \cdot n, v_h \rangle_{0,\partial\Omega_D} + \langle \eta, \mathcal{R}^*(v_h) \rangle_{0,\partial\Omega_R} |. \end{aligned}$$

The last line uses the equality

$$\sum_{\tau \in \mathcal{T}_h} \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\tau} = \sum_{F \in \mathcal{F}_{int}} \langle \eta, \llbracket \nabla v_h \rrbracket \rangle_{0,F} + \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\Omega_D} + \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\Omega_N} + \langle \eta, \nabla v_h \cdot n \rangle_{0,\partial\Omega_R}.$$

The Dirichlet boundary term cancels with the boundary term from the integration by parts on $\partial\Omega_D$. Now recall the stabilizations

$$\begin{aligned} |s_{\partial\Omega}(\eta, v_h)| &= |i\beta_N h \langle \nabla\eta \cdot n, \nabla v_h \cdot n \rangle_{0,\partial\Omega_N} + i\beta_R h \langle \mathcal{R}(\eta), \mathcal{R}^*(v_h) \rangle_{0,\partial\Omega_R} \\ &\quad + i\beta_D h^{-1} \langle \eta, v_h \rangle_{\partial\Omega_D} | \\ &\leq |i\beta_N h \langle \nabla\eta \cdot n, \nabla v_h \cdot n \rangle_{0,\partial\Omega_N} + i\beta_R h \|\mathcal{R}(\eta)\|_{0,\partial\Omega_R} (\|\nabla v_h \cdot n\|_{0,\partial\Omega_R} + k\|v_h\|_{0,\partial\Omega_R}) \\ &\quad + i\beta_D h^{-1} \langle \eta, v_h \rangle_{\partial\Omega_D} |. \end{aligned}$$

In the case of the GLS/CIP method

$$|s_{\Omega}^G(\eta, v_h)| = \left| \sum_{\tau \in \mathcal{T}_h} h_{\tau}^2 \delta_{1,\tau} (\mathcal{L}(\eta), \mathcal{L}(v_h))_{0,\tau} + \sum_{F \in \mathcal{F}_{int}} h_{\tau} \gamma_{1,\tau} \langle \llbracket \nabla \eta \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_F \right|.$$

Then the result follows from noticing that the right hand term in each inner product appears in $|\cdot|_{\mathfrak{J},G}$ to give the last line. So it can be concluded, after an application of the Cauchy-Schwarz inequality in all terms, that

$$A_h(\eta, v_h) \leq \|\eta\|_{*,G} |v_h|_{\mathfrak{J},G}.$$

where

$$\begin{aligned} \|\eta\|_{*,G}^2 &= \sum_{\tau \in \mathcal{T}_h} |\delta_{1,\tau}|^{-1} h_{\tau}^{-2} \|\eta\|_{0,\tau}^2 + \sum_{F \in \mathcal{F}_{int}} (|\gamma_{1,\tau} h_{\tau}|)^{-1} \|\eta\|_{0,F}^2 + (|\beta_{1,N}|h)^{-1} \|\eta\|_{0,\partial\Omega_N}^2 \\ &\quad + |\beta_{1,D}|^{-1} h \|\nabla \eta \cdot n\|_{0,\partial\Omega_D}^2 + (|\beta_{1,R}|h)^{-1} \|\eta\|_{0,\partial\Omega_R}^2 + s_{\Omega}^G(\eta, \eta). \end{aligned}$$

The argument for the CIP case is similar, except now I use the L^2 -projection instead of the Lagrange interpolant which eliminates the k^2 term and is used to handle the Laplacian after an integration by parts as shown previously.

Assumption (2.28) requires that the discrete sesquilinear form satisfies the following inequality,

$$A(u - \pi_h u, z - \pi_h z) \leq C(kh)h^p |u|_{p+1,\Omega} \|\psi\|_{0,\Omega}.$$

Since it has been shown in the previous section that this is indeed the case for the majority of the terms in $A_h(\cdot, \cdot)$, it suffices to show that the extra boundary terms satisfy the condition.

Let $\eta = u - \pi_h u$ and $\eta' = z - \pi_h z$ where π_h is either chosen to be the Lagrange interpolant (GLS/CIP case) or the L^2 projection (CIP case). In fact π_h can be chosen as any operator that satisfies Assumption (2.25) for linear elements or Assumption (2.26) for higher order elements. The extra terms from the Nitsche form are as follows

$$\underbrace{\langle \nabla \eta \cdot n, \eta' \rangle_{0,\partial\Omega_D}}_{(A)} + \underbrace{\langle \eta, \nabla \eta' \cdot n \rangle_{0,\partial\Omega_D}}_{(B)} + \underbrace{|\beta_D| h^{-1} \langle \eta, \eta' \rangle_{0,\partial\Omega_D}}_{(C)} + \underbrace{|\beta_N| h \langle \nabla \eta \cdot n, \nabla \eta' \cdot n \rangle_{0,\partial\Omega_N}}_{(D)}.$$

The first term (A) can be shown to satisfy (2.28) after applying the Cauchy-Schwarz inequality followed by the trace inequality (A:4).

$$\begin{aligned} (A) &\leq \|\nabla \eta \cdot n\|_{0,\partial\Omega_D} \|\eta'\|_{0,\partial\Omega_D} \\ &\leq \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_D} C_T (h_{\tau}^{-1} \|\nabla \eta\|_{0,\tau}^2 + h_{\tau} \|\nabla \eta\|_{1,\tau}^2)^{1/2} \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_D} C_T (h_{\tau}^{-1} \|\eta'\|_{0,\tau}^2 + h_{\tau} \|\eta'\|_{1,\tau}^2)^{1/2} \\ &\leq C h^{p-1/2} |u|_{p+1,\Omega} h^{3/2} |z|_{2,\Omega} \\ &\leq C(hk)h^p |u|_{p+1,\Omega} \|\psi\|_{0,\Omega}. \end{aligned}$$

The last two lines come from the interpolation result (2.25) for linear elements or Assumption (2.26) for higher order elements. In the case of piecewise linear elements it holds that

$$|\nabla\eta|_{1,\tau} \leq |u|_{2,\tau},$$

since $D^2\pi_h u|_\tau = 0$.

Similarly, the second term (B) can be shown to satisfy (2.28) after applying the Cauchy-Schwarz inequality followed by the trace inequality (A:4).

$$\begin{aligned} (B) &\leq \|\eta\|_{0,\partial\Omega_D} \|\nabla\eta' \cdot n\|_{0,\partial\Omega_D} \\ &\leq \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_D} C_T (h_\tau^{-1} \|\eta\|_{0,\tau}^2 + h_\tau |\eta|_{1,\tau}^2)^{1/2} \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_D} C_T (h_\tau^{-1} \|\nabla\eta'\|_{0,\tau}^2 + h_\tau |\nabla\eta'|_{1,\tau}^2)^{1/2} \\ &\leq Ch^{1/2} |z|_{2,\Omega} h^{p+1/2} |u|_{p+1,\Omega} \\ &\leq C(hk)h^p |u|_{p+1,\Omega} \|\psi\|_{0,\Omega}. \end{aligned}$$

Once again the last two lines come from the interpolation results (2.25) for linear elements or Assumption (2.26) for higher order elements. In this case for piecewise linear elements use is made of the fact that $D^2\pi_h z|_\tau = 0$.

The third term (C) is upper bounded using the Cauchy-Schwarz and trace inequalities followed by the interpolation results given in (2.25) or (2.26). Since there are no second order derivatives required in this estimate, the trick used in the previous two cases is not necessary.

$$\begin{aligned} (C) &\leq |\beta_D| h^{-1} \|\eta\|_{0,\partial\Omega_D} \|\eta'\|_{\partial\Omega_D} \\ &\leq |\beta_D| h^{-1} \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_D} C_T (h_\tau^{-1} \|\eta\|_{0,\tau}^2 + h_\tau |\eta|_{1,\tau}^2)^{1/2} \sum_{\tau \in \mathcal{T}_h \cap \partial\Omega_D} C_T (h_\tau^{-1} \|\eta'\|_{0,\tau}^2 + h_\tau |\eta'|_{1,\tau}^2)^{1/2} \\ &\leq |\beta_D| h^{-1} Ch^{p+1/2} |u|_{p+1,\Omega} h^{3/2} |z|_{2,\Omega} \\ &\leq C(hk)h^p |u|_{p+1,\Omega} \|\psi\|_{0,\Omega}. \end{aligned}$$

Finally, the last term can be bounded in the same way. Making use of the fact that $D^2\pi_h u|_\tau = 0$ and $D^2\pi_h z|_\tau = 0$ for piecewise linear elements

$$\begin{aligned}
(D) &\leq |\beta_N| h \|\nabla \eta \cdot n\|_{0, \partial \Omega_N} \|\nabla \eta' \cdot n\|_{\partial \Omega_N} \\
&\leq |\beta_N| h \sum_{\tau \in \mathcal{T}_h \cap \partial \Omega_N} C_T (h_\tau^{-1} \|\nabla \eta\|_{0, \tau}^2 + h_\tau \|\nabla \eta\|_{1, \tau}^2)^{1/2} \sum_{\tau \in \mathcal{T}_h \cap \partial \Omega_N} C_T (h_\tau^{-1} \|\nabla \eta'\|_{0, \tau}^2 + h_\tau \|\nabla \eta'\|_{1, \tau}^2)^{1/2} \\
&\leq |\beta_N| h C h^{p-1/2} |u|_{p+1, \Omega} h^{1/2} |z|_{2, \Omega} \\
&\leq C(hk) h^p |u|_{p+1, \Omega} \|\psi\|_{0, \Omega}.
\end{aligned}$$

Rounding up all of the results concludes the proof. \square

It has been shown that the stabilized methods can now handle general boundary conditions without leaving the framework of Theorem 2. This flexibility is important when considering problems of practical importance as the boundary conditions contain information about the physical properties of the scatterer which I mentioned earlier. In conclusion, the stabilized methods I have introduced have been shown to be stable for general boundary conditions provided the stabilization parameters are chosen appropriately. The analysis also shows the method to have quasi-optimal a priori error estimates in the asymptotic regime ie for $k^2 h < C$.

3.2.1 Numerical Results

The numerical results presented in this section have been performed using the Bessel solution introduced in (2.116). All computations in this section were performed using the UMFPACK solver in the FreeFEM++ package which is available from <http://www.freefem.org/>.

I pose the following problem on the unit square $\Omega = [0, 1] \times [0, 1]$

$$\left\{ \begin{array}{ll} -\Delta u - k^2 u = f & \text{in } \Omega \\ u = g_D & \text{for } x = 0, y \in [0, 1] \\ u = g_D & \text{for } x \in [0, 1], y = 0 \\ \nabla u \cdot n = g_N & \text{for } x = 1, y \in [0, 1] \\ \nabla u \cdot n + iku = g_R & \text{for } x \in [0, 1], y = 1 \end{array} \right.$$

For our ansatz to be the solution let

$$f = \frac{\sin(k\sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}.$$

The plots are designed to compare the convergence rates for three numerical methods. The first method is the standard Galerkin method. This method can be obtained from either of the stabilized methods presented earlier by setting the stabilization parameters to 0. The other two methods analysed here are the CIP and GLS/CIP methods, using the stabilization parameters obtained in chapter 2, as presented with general boundary conditions. The additional Nitsche penalty parameters are taken to be 10.

Figure 3.1 compares the convergence rates of the different methods for piecewise linear elements on a structured mesh. The plot shows that the stabilized methods are more accurate than the standard Galerkin method for this particular problem and obtain optimal convergence for a coarser mesh.

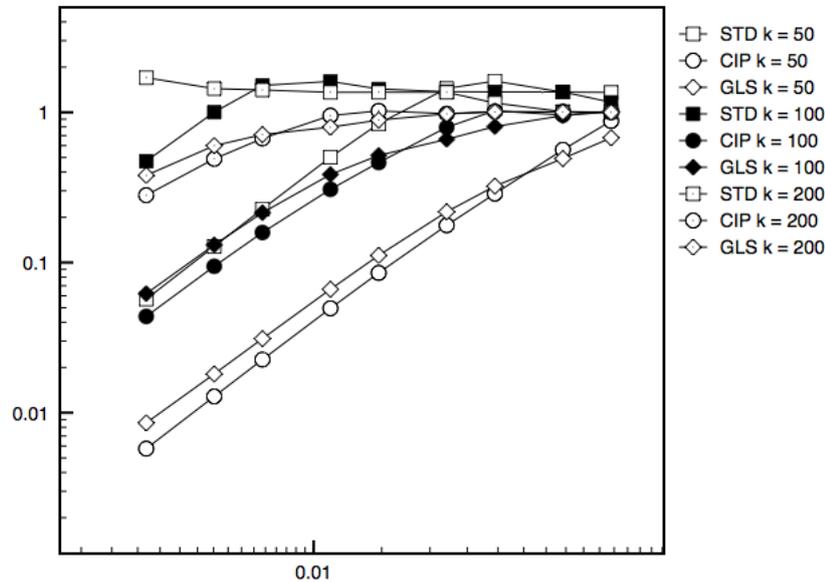


Figure 3.1: Convergence plot for generalized boundary conditions using piecewise linear elements on a structured mesh

Figure 3.2 compares the convergence of the three numerical methods using piecewise linear elements on an unstructured mesh. The plot, as expected, is fairly similar to the structured mesh case. The stabilized methods reach optimal convergence for a coarser mesh size than the standard Galerkin method and are also more accurate. Figures 3.1 and 3.2 seem to demonstrate the potential of the proposed stabilized method to reduce the pollution effect, even with weakly imposed boundary conditions, if the stabilization parameters are chosen optimally.

Figures 3.3 and 3.4 compare the convergence of the three numerical methods for piecewise quadratic elements on a structured and unstructured mesh respectively. The standard Galerkin method performs better (as expected) in these plots. The GLS method is the most accurate in the under resolved regime for $k = 100, 200$ but does not reach optimal convergence as quickly as the CIP method. For $k = 400$ the CIP method begins converging first and provides a more reliable solution than both other methods. It should be noted that although the stabilized methods appear to yield better results than the standard Galerkin method they are also more expensive to compute making the comparison difficult to quantify.

3.2.2 Plane wave solution

The following problem has been designed to demonstrate qualitatively the ability of the Nitsche's method to reduce pollution when solving Helmholtz equation.

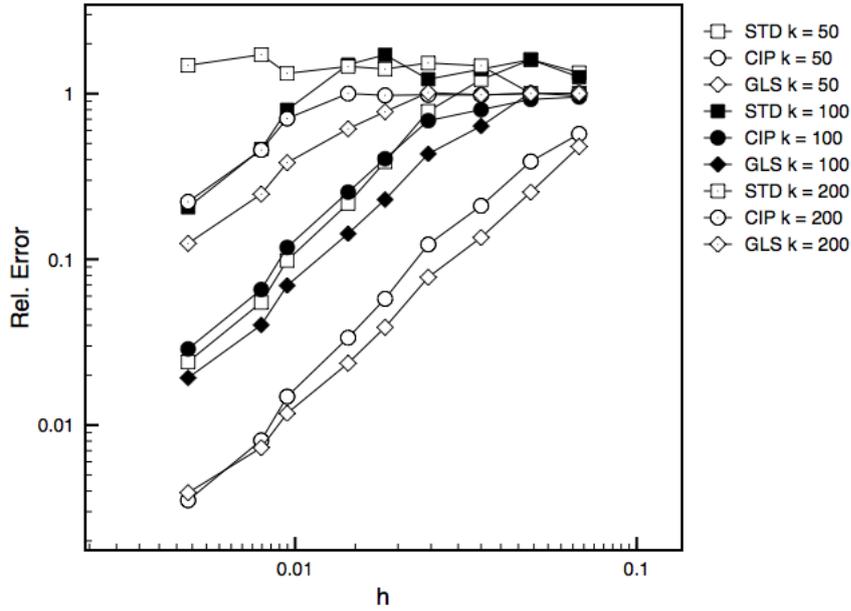


Figure 3.2: Convergence plot for generalized boundary conditions using piecewise linear elements on an unstructured mesh

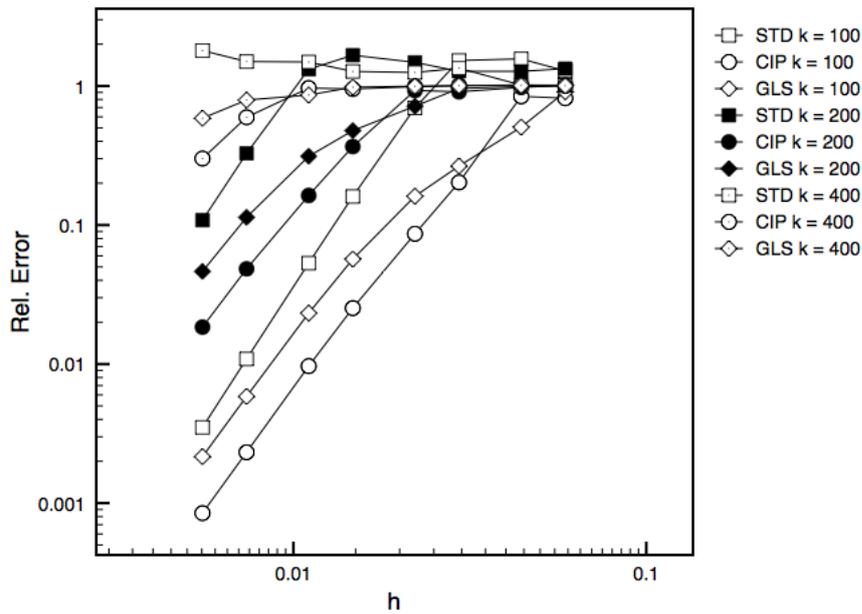


Figure 3.3: Convergence plot for generalized boundary conditions using piecewise quadratic elements on a structured mesh

$$\left\{ \begin{array}{l} -\Delta u(x, y) - k^2 u(x, y) = 0 \quad (x, y) \in [0, 1] \times [0, 1] \\ u(x, 0) = e^{ik_1 x} \\ u(0, y) = e^{ik_2 y} \\ \frac{\partial u(x, y)}{\partial y} = ik_2 u(x, y) \quad \text{for } y = 1 \\ \frac{\partial u(x, y)}{\partial x} = ik_1 u(x, y) \quad \text{for } x = 1 \end{array} \right.$$

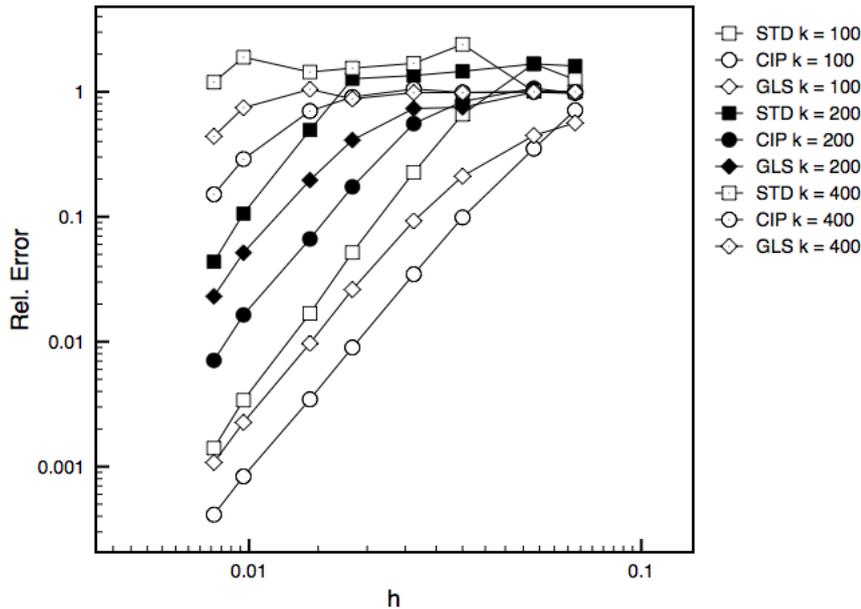
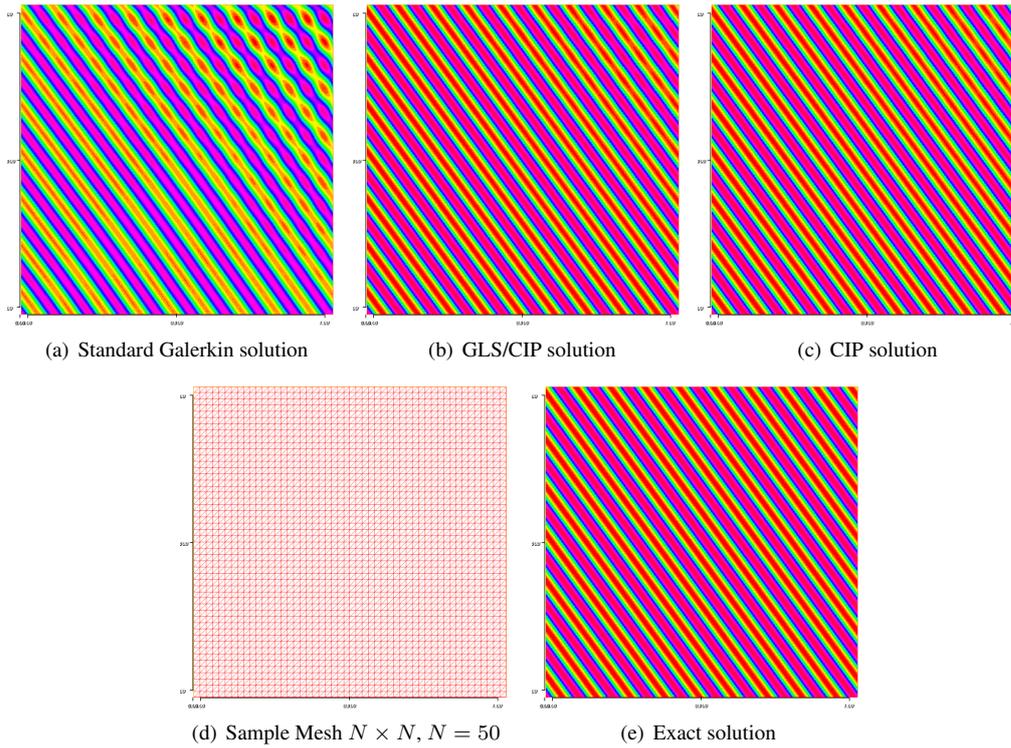
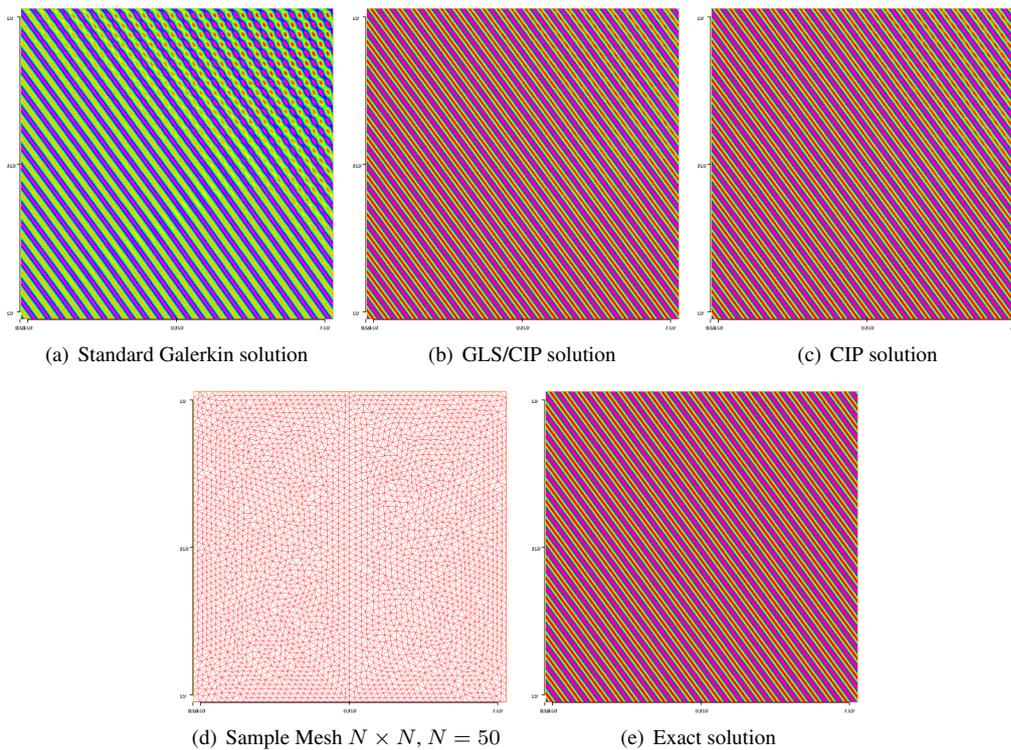


Figure 3.4: Convergence plot for generalized boundary conditions using piecewise quadratic elements on an unstructured mesh

Figure 3.5 shows the real part of the numerical solution of the plane wave problem given by the standard Galerkin method and the two stabilized methods introduced previously. The problem is solved using a structured mesh of the same form as shown, where N characterises the number of elements used to discretize the boundary, for example a 50×50 mesh describes a mesh on a rectangular domain having exactly 50 elements that share an edge/face with each side of the domain. The standard Galerkin method introduces spurious oscillations generated by a loss of stability which are not present in the exact solution. These oscillations do not appear in the stabilized methods presented.

Figure 3.6 aims to replicate similar results using an unstructured mesh. The structure of the mesh plays an important role in the calculation of numerical solutions to the plane wave problem. For this particular example the numerical solution is less accurate on the structured mesh which is why I have chosen to present a solution with larger wave number in the unstructured case. The standard Galerkin method can again be seen to lose stability causing it to differ from the exact solution. The stabilized methods seem to limit the introduction of unwanted oscillations and appear to provide a solution similar to the exact solution.

Figure 3.5: Solution plots for $k \approx 70$, $N = 200$, $\theta = \frac{\pi}{5}$ Figure 3.6: Solution plots for $k \approx 173$, $N = 400$, $\theta = \frac{\pi}{5}$

3.3 Fictitious domain method

In this section I introduce a fictitious domain method for the solution of Helmholtz equation using cut elements. In practice when solving PDEs, such as Helmholtz equation, engineers must often consider

problems on complex domains. The meshing of these complicated geometries can become computationally expensive especially when considering evolving domains in which the mesh must be updated at every time step. The fictitious domain method is used to embed the complex domain into a simpler one and, by using a fixed background mesh, can avoid these issues. These advantages are beyond the scope of this work but are used as additional motivations to the use of fictitious domain methods for the solution of Helmholtz equation. The main motivation in this instance is the application to multi-domain coupling and extension to include an unfitted interface.

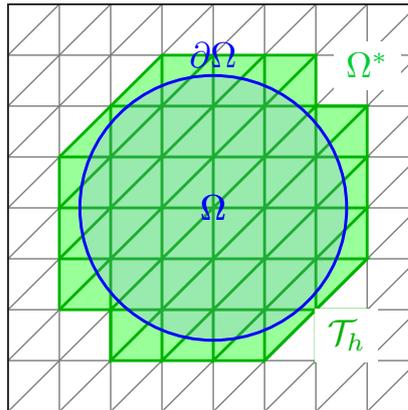


Figure 3.7: Physical and computational domain

Figure 3.5 shows an example of a cut mesh that I will be considering in this section. The physical domain Ω is embedded in the background mesh \mathcal{T}_h . Unlike conventional methods, the fictitious domain method that I consider does not fully discretize the domain Ω . The boundary $\partial\Omega$ of the physical domain is not discretized since the CutFEM method assumes that the cut elements can be integrated over exactly, this can be thought of as a semi-discretization. Ω^* is referred to as the extended domain. Its purpose will become evident during the analysis.

In [21] a stabilized Nitsche method was proposed by Burman and Hansbo to implement a fictitious domain method using cut elements for the solution of the Poisson equation. It was noted by the authors that in order to ensure that the method was robust a stabilization term known as the ghost penalty, introduced by Burman in [12], must be added to ensure that the method is coercive on the entire computational domain. The method introduced in this section will use the ideas from [21] to develop a fictitious domain method for the solution of Helmholtz equation. It is interesting to note that the new method does not require the addition of the ghost penalty for accuracy since I choose to take the stability parameters defined in Nitsche's method to be purely imaginary. If I had chosen to take these parameters to be real I would need to introduce the ghost penalty term. Another motivation for introducing the ghost penalty to this method would be that it can be used to improve the stability of iterative solvers. The new method will be shown to fit into the framework of Theorem 2 and numerical evidence will be presented to back up the claims of optimal convergence.

Considering the complex Helmholtz equation in $\Omega \subset \mathbb{R}^d$, where Ω is a smooth convex domain such that integrals over cut elements can be evaluated exactly, the Helmholtz equation is given as

$$\left. \begin{aligned} -\Delta u - k^2 u &= f && \text{in } \Omega \\ \nabla u \cdot n + iku &= g_R && \text{on } \partial\Omega \end{aligned} \right\}. \quad (3.29)$$

with the associated adjoint problem

$$\left. \begin{aligned} -\Delta z - k^2 z &= f && \text{in } \Omega \\ \nabla z \cdot n - ikz &= 0 && \text{on } \partial\Omega \end{aligned} \right\}. \quad (3.30)$$

The stabilized finite element method in this case can be written:

$$A_h(u_h, v_h) = A(u_h, v_h) + s(u_h, v_h), \quad (3.31)$$

where

$$A(u_h, v_h) = (\nabla u_h, \nabla v_h)_{0,\Omega} - k^2 (u_h, v_h)_{0,\Omega} + ik \langle u_h, v_h \rangle_{0,\Omega} \quad \forall v_h \in V_h, \quad (3.32)$$

and

$$\begin{aligned} s(u_h, v_h) &= \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau (\mathcal{L}(u_h), \mathcal{L}(v_h))_{0,\tau \cap \Omega} + \sum_{F \in \mathcal{F}_i} h_F \gamma_F \langle \llbracket \nabla u_h \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_{0,F \cap \Omega} \\ &\quad + h\beta \langle \mathcal{R}(u_h), \mathcal{R}^*(v_h) \rangle_{0,\partial\Omega}, \end{aligned} \quad (3.33)$$

where $\mathcal{L}(\cdot) \stackrel{\text{def}}{=} -\Delta(\cdot) - k^2(\cdot)$, $\mathcal{R}(\cdot) \stackrel{\text{def}}{=} \nabla(\cdot) \cdot n + ik(\cdot)$ and $\mathcal{R}^*(\cdot) \stackrel{\text{def}}{=} \nabla(\cdot) \cdot n - ik(\cdot)$. Let $\mathcal{T}_h \subset \mathbb{R}^2$ denote a computational mesh consisting of quasi-uniform, shape regular triangles τ , as defined in (A.4) and (A.3), such that $\Omega \subset \mathcal{T}_h$. The finite element space in this case is defined as

$$V_h \stackrel{\text{def}}{=} \{v_h \in C^0(\Omega) : v_h|_\tau \in \mathbb{P}^p(\tau) \forall \tau \in \mathcal{T}_h\}. \quad (3.34)$$

It is also useful to define the set of all interior faces as

$$\mathcal{F}_i \stackrel{\text{def}}{=} \{F \in \mathcal{T}_h : F \cap \Omega \neq \emptyset\}, \quad (3.35)$$

where F denotes an element face. Finally, I introduce the set of elements cut by the interface as

$$\mathcal{G}_h \stackrel{\text{def}}{=} \{\tau \in \mathcal{T}_h : \bar{\tau} \cap \Gamma \neq \emptyset\}, \quad (3.36)$$

For the following analysis to hold it is important that the boundary of the domain Ω is resolved to

a reasonable degree of accuracy. To ensure this is the case it is important to specify the necessary assumptions as presented in [14].

FD1: The intersection between $\partial\Omega$ and a face $F \in \mathcal{F}_i$ is connected. That is, the boundary does not intersect an interior face more than once.

FD2: For each element τ intersected by $\partial\Omega$ there exists a line (plane for $d = 3$) S_τ and a piecewise smooth parametrization $\Phi : S_\tau \cap \tau \mapsto \partial\Omega \cap \tau$.

FD3: Assume that there is an integer $N > 0$ such that element $\tau \in \mathcal{G}_h$ there exists an element $\tau' \in \mathcal{T}_h \setminus \mathcal{G}_h$ and at most N elements $\{\tau\}_{i=1}^N$ such that $\tau_1 = \tau, \tau_N = \tau'$ and $\tau_i \cap \tau_{i+1} \in \mathcal{F}_i, i = 1, \dots, N-1$ i.e. the amount of faces crossed in order to go from a cut element to an element completely inside the domain is uniformly bounded.

The following analysis relies on an important observation by Stein [51] which asserts the existence of a so called extension operator:

Definition 3.3.1 (Extension Operator). Let $u \in H^p(\Omega)$, where $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary. Then for $\Omega \subset \Omega^* \subset \mathbb{R}^d$ there exists an operator $\mathbb{E}_p : H^p(\Omega) \mapsto H^p(\Omega^*)$ such that

$$\mathbb{E}u|_{\Omega} = u,$$

and

$$\|\mathbb{E}_p u\|_{p, \Omega^*} \leq C \|u\|_{p, \Omega}.$$

Definition 3.3.1 allows me to analyse the CutFEM fictitious domain method on an extended domain. It is therefore beneficial to introduce Ω^* as the extended domain such that:

$$\Omega^* \stackrel{\text{def}}{=} \{\tau \in \mathcal{T}_h : \tau \cap \Omega \neq \emptyset\}.$$

The extension property also allows me to introduce notation to denote an extended interpolation operator. Let $\mathcal{I}_h : H^p(\Omega^*) \mapsto V_h$ be any interpolation operator. Then I define its associated extended operator $\mathcal{I}_h^* : H^p(\Omega) \mapsto V_h$ where

$$\mathcal{I}_h^* v \stackrel{\text{def}}{=} \mathcal{I}_h \mathbb{E}_p v,$$

for any $v \in H^p(\Omega)$.

It is useful to note that a variant of the standard interpolation estimates hold for this operator.

Lemma 3.4 (Interpolation Estimates). Let $u \in H^{p+1}(\Omega)$ and $\mathcal{I}_h : C^0(\bar{\tau}) \mapsto V(\tau)$ be the standard

Lagrange interpolant. It holds that for $0 \leq s \leq p$

$$|u - \mathcal{I}_h u|_{s,\tau} \leq Ch^{p-s+1} |u|_{p+1,\tau}. \quad (3.37)$$

Given that \mathcal{T}_h is shape regular it also holds that for $H^{p+1}(\Omega^*) \ni u^* \stackrel{\text{def}}{=} \mathbb{E}_{p+1} u$

$$\|u^* - \mathcal{I}_h^* u\|_{s,\Omega^*} \leq Ch^{p-s+1} \|u\|_{p+1,\Omega}. \quad (3.38)$$

Also under the same assumptions

$$\|u - \mathcal{I}_h^* u\|_{s,\Omega} \leq h^{p-s+1} \|u\|_{p+1,\Omega}. \quad (3.39)$$

Proof. A proof for (3.37) can be found in numerous Finite Element textbooks for example [30]. The proof of (3.38) follows from (3.37) since $u^* \in H^{p+1}(\Omega^*)$ implies

$$\|u^* - \mathcal{I}_h^* u\|_{s,\Omega^*} \leq Ch^{p-s+1} |u^*|_{p+1,\Omega^*}. \quad (3.40)$$

(3.38) then follows directly from the definition of the extension operator, see Definition 3.3.1. The final result is obtained from using the fact that

$$\mathbb{E}u|_{\Omega} = u,$$

which implies that

$$\|u - \mathcal{I}_h^* u\|_{s,\Omega} \leq \|u^* - \mathcal{I}_h^* u\|_{s,\Omega^*} \quad (3.41)$$

$$\leq Ch^{p-s+1} \|u\|_{p+1,\Omega}. \quad (3.42)$$

The last step is a result of the stability of the extension. \square

Since the fictitious domain method does not require the mesh to fit the boundary, it is necessary to introduce an additional trace inequality that handles the case where the intersection $\partial\Omega \cap \tau \neq \partial\tau$.

Lemma 3.5 (Trace Inequality). *Under the conditions given in FD1–FD3 the following trace inequality*

$$\|v\|_{0,\tau \cap \partial\Omega} \leq C_T \left(h_\tau^{-1/2} \|v\|_{0,\tau} + h_\tau^{1/2} \|\nabla v\|_{0,\tau} \right), \quad (3.43)$$

holds for all $v \in H^1(\Omega^*)$.

Proof. A proof can be found in [33] \square

With the preceding lemmas in place it is now possible to show that the new formulation fits into the

framework of Theorem 2. The following definitions are introduced to help simplify the subsequent analysis. Let $|u_h|_{\mathcal{J}}$ be defined as

$$\begin{aligned} |u_h|_{\mathcal{J}}^2 &\stackrel{\text{def}}{=} k(1 - \text{Im}[\beta]kh) \|u_h\|_{0,\partial\Omega_R}^2 + \max_{\tau \in \mathcal{T}_h} \{\text{Im}[\delta_\tau]\} h^2 \sum_{\tau \in \mathcal{T}_h} \|\mathcal{L}(u_h)\|_{0,\tau \cap \Omega}^2 \\ &\quad + \max_{\tau \in \mathcal{T}_h} \{\text{Im}[\gamma_\tau]\} h \sum_{F \in \mathcal{F}_i} \|\llbracket \nabla u_h \rrbracket\|_{0,F \cap \Omega}^2 + \text{Im}[\beta] h \|\nabla u_h \cdot \mathbf{n}\|_{0,\partial\Omega_R}^2. \end{aligned} \quad (3.44)$$

Lemma 3.6. *For $0 < \text{Im}[\beta]kh < 1$ and $\max_{\tau \in \mathcal{T}_h} \{\text{Im}[\delta_\tau]\}, \max_{\tau \in \mathcal{T}_h} \{\text{Im}[\gamma_\tau]\}, \text{Im}[\beta] > 0$ it follows that $|u_h|_{\mathcal{J},G}^2$, as defined in (3.44), is a norm on V_h .*

Proof. The proof follows directly from the definition. \square

The following norm is also useful in the subsequent analysis $|\cdot|_s : V_h \mapsto \mathbb{R}$.

$$|\cdot|_s^2 \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} |\delta_\tau| h_\tau^2 \|\mathcal{L}(\cdot)\|_{0,\tau \cap \Omega}^2 + \sum_{F \in \mathcal{F}_i} |\gamma_\tau| h_\tau \|\llbracket \nabla(\cdot) \cdot \mathbf{n} \rrbracket\|_{0,F \cap \Omega}^2 + h|\beta| \|\mathcal{R}^*(\cdot)\|_{0,\partial\Omega}^2, \quad (3.45)$$

and finally $\|\cdot\|_* : V \mapsto \mathbb{R}$

$$\|(\cdot)\|_*^2 \stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} \|h_\tau^{-1} \delta_\tau^{-1/2}(\cdot)\|_{0,\tau \cap \Omega}^2 + \sum_{F \in \mathcal{F}_i} \|(h_\tau \gamma_\tau)^{-1/2}(\cdot)\|_{0,F \cap \Omega}^2 + \|(h\beta)^{-1/2}(\cdot)\|_{0,\partial\Omega}^2. \quad (3.46)$$

Lemma 3.7. *Let $u \in H^{p+1}(\Omega)$ and $z \in H^2(\Omega)$ be the solution of the primal and adjoint Helmholtz problems given in (3.29) and (3.30) respectively. Then it follows that for $\text{Im}[\beta]^{-1} \geq kh$ the fictitious domain formulation given in (3.31) with stabilization given by (3.33) satisfies assumptions (2.23), (2.24), (2.29), (2.30) and (2.31) for the associated norms given by (3.44) - (3.46) and interpolant*

$$\pi_h \stackrel{\text{def}}{=} \mathcal{I}_h^* : C^0(\Omega) \mapsto V_h. \quad (3.47)$$

Proof. The proof of the first two assumptions follows in a similar manner as shown previously so to avoid repetition I will exclude the justification here. The proof of (2.30) follows using a similar methodology to that presented previously

$$\begin{aligned} |\mathcal{I}_h^* z|_s &= \left(\sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau \|\mathcal{L}(\mathcal{I}_h^* z)\|_{0,\tau \cap \Omega}^2 + \sum_{F \in \mathcal{F}_i} h_\tau \gamma_\tau \|\llbracket \nabla(\mathcal{I}_h^* z) \cdot \mathbf{n} \rrbracket\|_{0,F \cap \Omega}^2 \right. \\ &\quad \left. + h\beta \|\mathcal{R}^*(\mathcal{I}_h^* z)\|_{0,\partial\Omega}^2 \right)^{1/2}. \end{aligned} \quad (3.48)$$

Using the fact that z is the exact solution of the dual problem gives

$$|\mathcal{I}_h^* z|_s = \left(\sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau \|\mathcal{L}(z - \mathcal{I}_h^* z) - \psi\|_{0,\tau \cap \Omega}^2 + \sum_{F \in \mathcal{F}_i} h_\tau \gamma_\tau \|\llbracket \nabla(z - \mathcal{I}_h^* z) \rrbracket\|_{0,F \cap \Omega}^2 + h\beta \|\mathcal{R}^*(z - \mathcal{I}_h^* z)\|_{0,\partial\Omega}^2 \right)^{1/2}. \quad (3.49)$$

Applying the triangle inequality to the first term gives

$$\sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau \|\mathcal{L}(z - \mathcal{I}_h^* z) - \psi\|_{0,\tau \cap \Omega}^2 \leq \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau \|\mathcal{L}(z - \mathcal{I}_h^* z)\|_{0,\tau \cap \Omega}^2 + h_\tau^2 \delta_\tau \|\psi\|_{0,\tau \cap \Omega}^2.$$

All that is then left to show for the first term is

$$\sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau \|\mathcal{L}(z - \mathcal{I}_h^* z)\|_{0,\tau \cap \Omega}^2 \leq \sum_{\tau \in \mathcal{T}_h} (h_\tau^2 \delta_\tau \|\Delta(z - \mathcal{I}_h^* z)\|_{0,\tau \cap \Omega}^2 + (kh_\tau)^2 \delta_\tau \|z - \mathcal{I}_h^* z\|_{0,\tau \cap \Omega}^2).$$

The result follows from using the interpolation estimates given by (3.39) for elements of order $p > 1$ and for the piecewise linear case the result follows from noticing

$$\Delta \mathcal{I}_h^* z = 0,$$

on each element, reducing the first term to

$$\sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau \|\Delta z\|_{0,\tau \cap \Omega}^2 \leq Ch^2 \|z\|_{2,\Omega}^2.$$

The second term, after an application of the trace inequality given by (A:4), simplifies to

$$\begin{aligned} \sum_{F \in \mathcal{F}_i} h_\tau \gamma_\tau \|\llbracket \nabla(z - \mathcal{I}_h^* z) \rrbracket\|_{0,F \cap \Omega}^2 &\leq C_T^2 \sum_{\tau \in \mathcal{T}_h} h_\tau \gamma_\tau (h_\tau^{-1} \|z - \mathcal{I}_h^* z\|_{0,\tau}^2 + h_\tau \|\nabla(z - \mathcal{I}_h^* z)\|_{0,\tau}^2)^{1/2} \\ &\leq C(hk)^2 \|\psi\|_{0,\Omega}^2. \end{aligned}$$

Finally, the last term is bounded by using the triangle inequality

$$h\beta \|\mathcal{R}^*(z - \mathcal{I}_h^* z)\|_{0,\partial\Omega}^2 = h\beta \sum_{\tau \in \mathcal{T}_h} \|\mathcal{R}^*(z - \mathcal{I}_h^* z)\|_{0,\partial\Omega \cap \tau}^2 \quad (3.50)$$

$$\leq h\beta \sum_{\tau \in \mathcal{T}_h} \|\nabla(z - \mathcal{I}_h^* z) \cdot n\|_{0,\partial\Omega \cap \tau}^2 + \|ik(z - \mathcal{I}_h^* z)\|_{0,\partial\Omega \cap \tau}^2, \quad (3.51)$$

followed by an application of the trace inequality, given in (3.43), on the resulting terms giving

$$h\beta \sum_{\tau \in \mathcal{T}_h} (\|\nabla(z - \mathcal{I}_h^* z) \cdot n\|_{0,\partial\Omega \cap \tau}^2) \leq h\beta \sum_{\tau \in \mathcal{T}_h} C_T^2 (h_\tau^{-1} \|\nabla(z^* - \mathcal{I}_h^* z)\|_{0,\tau}^2 + \quad (3.52)$$

$$h_\tau |\nabla(z^* - \mathcal{I}_h^* z)|_{1,\tau}^2)^{1/2} \\ \leq Ch^2 \|z\|_{2,\Omega}^2, \quad (3.53)$$

and

$$h\beta \sum_{\tau \in \mathcal{T}_h} \|ik(z - \mathcal{I}_h^* z)\|_{0,\partial\Omega \cap \tau}^2 \leq h\beta \sum_{\tau \in \mathcal{T}_h} C_T^2 k^2 \left(h_\tau^{-1} \|z^* - \mathcal{I}_h^* z\|_{0,\tau}^2 + h_\tau^{1/2} \|\nabla(z^* - \mathcal{I}_h^* z)\|_{0,\tau}^2 \right)^{1/2} \\ \leq C(kh)^2 h^2 \|z\|_{2,\Omega}^2.$$

The proof is then concluded by using the regularity estimates for the dual problem. Next, the proof of Assumption (2.30). Let $\eta = u - \mathcal{I}_h^* u$

$$|\eta|_3^2 = k(1 - hk\beta) \|\eta\|_{0,\partial\Omega}^2 + \sum_{\tau \in \mathcal{T}_h} h_\tau^2 \delta_\tau \|\mathcal{L}(\eta)\|_{0,\tau \cap \Omega}^2 + \sum_{F \in \mathcal{F}_i} h_\tau \gamma_\tau \|\llbracket \nabla(\eta) \rrbracket\|_{0,F \cap \Omega}^2 + h\beta \|\nabla \eta \cdot n\|_{0,\partial\Omega}^2.$$

It is clear that each term can be shown to be bounded, as required, using similar techniques to the proof of (2.29). The final assumption is to show (2.31)

$$\|\eta\|_*^2 = \sum_{\tau \in \mathcal{T}_h} \|h_\tau^{-1} \delta_\tau^{-1/2} \eta\|_{0,\tau \cap \Omega}^2 + \sum_{F \in \mathcal{F}_i} \|(h_\tau \gamma_\tau)^{-1/2} \eta\|_{0,F \cap \Omega}^2 \\ + \|(h\beta)^{-1/2} \eta\|_{0,\partial\Omega}^2.$$

The first term can be easily bounded using the interpolation estimate given in (3.39) whilst the second and third terms follow from an application of the trace inequalities given in (A:4) and (3.43) respectively. Then an application of the interpolation estimate given in (3.39) can be applied to give the required result. \square

Proposition 5 (Continuity). *Let $u \in H^{p+1}(\Omega)$, for $p > 1$ and $\mathcal{I}_h^* : C^0(\Omega) \mapsto V_h$ be the extended Lagrange Interpolant. It follows that the discrete sesquilinear form $A_h(\cdot, \cdot)$ given by (3.31) with GLS/CIP stabilization satisfies assumptions (2.27) and (2.28) for the *-norm given by (3.46).*

Proof. The proof of assumption (2.27) follows by an integration by parts over the physical domain Ω .

Once again this operation must be performed element by element.

$$\begin{aligned}
|A(u - \pi_h^* u, v_h)| &= |(\nabla(u - \pi_h^* u), \nabla v_h)_{0,\Omega} - k^2(u - \pi_h^* u, v_h)_{0,\Omega} + ik \langle u - \pi_h^* u, v_h \rangle_{0,\partial\Omega}| \\
&= \left| \sum_{\tau \in \mathcal{T}_h} \left((u - \pi_h^* u, -\Delta v_h)_{0,\tau \cap \Omega} + \langle u - \pi_h^* u, \nabla v_h \cdot n \rangle_{0,\partial\Omega \cap \tau} \right) \right. \\
&\quad \left. - k^2(u - \pi_h^* u, v_h)_{0,\Omega} - \langle u - \pi_h^* u, ikv_h \rangle_{0,\partial\Omega} \right|.
\end{aligned}$$

This simplifies to

$$\begin{aligned}
|A(u - \pi_h^* u, v_h)| &= \left| \sum_{\tau \in \mathcal{T}_h} (u - \pi_h^* u, \mathcal{L}(v_h))_{0,\tau \cap \Omega} + \langle u - \pi_h^* u, \mathcal{R}^* v_h \rangle_{0,\partial\Omega \cap \tau} \right. \\
&\quad \left. + \sum_{F \in \mathcal{F}_i} \langle u - \pi_h^* u, \llbracket \nabla v_h \rrbracket \rangle_{0,F \cap \Omega} \right|.
\end{aligned}$$

An application of the Cauchy-Schwarz inequality reveals

$$\begin{aligned}
|A(u - \pi_h^* u, v_h)| &\leq \left| \sum_{\tau \in \mathcal{T}_h} \|h_\tau^{-1} \delta_\tau^{-1/2} (u - \pi_h^* u)\|_{0,\tau \cap \Omega} \|h_\tau \delta_\tau^{1/2} \mathcal{L}(v_h)\|_{0,\tau \cap \Omega} \right. \\
&\quad \left. + \|(h\beta)^{-1/2} (u - \pi_h^* u)\|_{0,\partial\Omega \cap \tau} \|(h\beta)^{1/2} \mathcal{R}^*(v_h)\|_{0,\partial\Omega \cap \tau} \right. \\
&\quad \left. + \sum_{F \in \mathcal{F}_i} \|(h_\tau \gamma_\tau)^{-1/2} (u - \pi_h^* u)\|_{0,F \cap \Omega} \|(h_\tau \gamma_\tau)^{1/2} \llbracket \nabla v_h \rrbracket\|_{0,F \cap \Omega} \right|.
\end{aligned}$$

Finally, the proof is complete by applying the triangle inequality to the adjoint Robin condition

$$\begin{aligned}
|A(u - \pi_h^* u, v_h)| &\leq \left| \sum_{\tau \in \mathcal{T}_h} \|h_\tau^{-1} \delta_\tau^{-1/2} (u - \pi_h^* u)\|_{0,\tau} \|h_\tau \delta_\tau^{1/2} \mathcal{L}(v_h)\|_{0,\tau \cap \Omega} \right. \\
&\quad \left. + \|(h\beta)^{-1/2} (u - \pi_h^* u)\|_{0,\partial\Omega \cap \tau} \|(h\beta)^{1/2} \nabla(v_h) \cdot n\|_{0,\partial\Omega \cap \tau} \right. \\
&\quad \left. + \|(h\beta)^{-1/2} (u - \pi_h^* u)\|_{0,\partial\Omega \cap \tau} \|(h\beta)^{1/2} kv_h\|_{0,\partial\Omega \cap \tau} \right. \\
&\quad \left. + \sum_{F \in \mathcal{F}_i} \|(h_\tau \gamma_\tau)^{-1/2} (u - \pi_h^* u)\|_{0,F \cap \Omega} \|(h_\tau \gamma_\tau)^{1/2} \llbracket \nabla v_h \rrbracket\|_{0,F \cap \Omega} \right. \\
&\leq \|u - \pi_h^* u\|_* |v_h|_{\mathfrak{J}}.
\end{aligned}$$

The proof of (2.28) follows from an application of the triangle and Cauchy-Schwarz inequalities such that

$$\begin{aligned}
|A(u - \mathcal{I}_h^* u, z - \mathcal{I}_h^* z)| &\leq \|\nabla(u - \mathcal{I}_h^* u)\|_{0,\Omega} \|\nabla(z - \mathcal{I}_h^* z)\|_{0,\Omega} + k^2 \|u - \mathcal{I}_h^* u\|_{0,\Omega} \|z - \mathcal{I}_h^* z\|_{0,\Omega} \\
&\quad + k \|u - \mathcal{I}_h^* u\|_{0,\partial\Omega} \|z - \mathcal{I}_h^* z\|_{0,\partial\Omega}.
\end{aligned}$$

The result then follows directly from our interpolation estimates. \square

The lemma and proposition assert that the method enters the framework of Theorem 2 without the need

of additional stabilization. The method is therefore stable for appropriate choices of stabilization parameters and will observe quasi-optimal convergence rates. It is worth noting that this theory can be extended to include the generalized boundary conditions discussed in the previous section.

3.3.1 Numerical Results

In the following section I introduce 3 fictitious domain methods for comparison. The analysis of the fictitious domain method presented in the previous section can be extended to consider the Dirichlet and Neumann using the techniques introduced in the first half of the chapter. The first method introduced below, (3.56), was analysed by Zou, Aquino and Harari in [58] and was shown to respect the same convergence estimates as the stabilized methods presented in this chapter. Unlike the stabilized methods that I have presented the results only appear to hold under the condition $k^2h < C$. The second method introduced below, (3.57), has not been analysed but can be thought of as a blend between the real Nitsche's method analysed by Zou et al and the Nitsche's method analysed earlier in this chapter. Finally, the last method (3.58) is the stabilized method that I introduced earlier.

$$A(u_h, v_h) \stackrel{\text{def}}{=} (\nabla u_h, \nabla v_h)_{0,\Omega} - k^2(u_h, v_h)_{0,\Omega} - \langle \nabla u_h \cdot \mathbf{n}, v_h \rangle_{0,\partial\Omega_D} + ik \langle u_h, v_h \rangle_{0,\partial\Omega_R}. \quad (3.54)$$

Definition 3.3.2 (Real Nitsche Method). Let u be the exact solution of the Helmholtz problem defined in $\beta_R, \hat{\beta}_R \in \mathbb{R}^+$ and $A(u_h, v_h)$ be as defined in (3.54). Then the Real Nitsche method is defined as

$$A_h(u_h, v_h) = L(u_h) + S_R(u, v_h). \quad (3.55)$$

$$\begin{aligned} S_R(u_h, v_h) &\stackrel{\text{def}}{=} - \langle u_h, \nabla v_h \cdot \mathbf{n} \rangle_{0,\partial\Omega_D} + \beta_R h^{-1} \langle u_h, v_h \rangle_{0,\partial\Omega_D} \\ &\quad + \hat{\beta}_R h \langle \nabla u_h \cdot \mathbf{n}, \nabla v_h \cdot \mathbf{n} \rangle_{0,\partial\Omega_D}. \end{aligned} \quad (3.56)$$

Definition 3.3.3 (Imaginary Nitsche Method).

$$\begin{aligned} S_I(u_h, v_h) &\stackrel{\text{def}}{=} - \langle u_h, \nabla v_h \cdot \mathbf{n} \rangle_{0,\partial\Omega_D} + h^{-1} i \beta_I \langle u_h, v_h \rangle_{0,\partial\Omega_D} \\ &\quad + i \hat{\beta}_I h \langle \nabla u_h \cdot \mathbf{n}, \nabla v_h \cdot \mathbf{n} \rangle_{0,\partial\Omega_D}. \end{aligned} \quad (3.57)$$

Definition 3.3.4 (Fully Stabilized Method). Let $\beta_I, \hat{\beta}_I \in \mathbb{R}^+$

$$\begin{aligned}
S_{\mathcal{S}}(u_h, v_h) &\stackrel{\text{def}}{=} \sum_{\tau \in \mathcal{T}_h} \delta_{\mathcal{S}, \tau} h_{\tau}^2 (\mathcal{L}(u_h), \mathcal{L}(v_h))_{0, \tau \cap \Omega} + \sum_{F \in \mathcal{F}_i} \gamma_{\mathcal{S}, \tau} h_{\tau} \langle \llbracket \nabla u_h \rrbracket, \llbracket \nabla v_h \rrbracket \rangle_{0, F \cap \Omega} \\
&\quad - \langle u_h, \nabla v_h \cdot \mathbf{n} \rangle_{0, \partial\Omega_D} + i\beta_{\mathcal{S}} h^{-1} \langle u_h, v_h \rangle_{0, \partial\Omega_D} + i\hat{\beta}_{\mathcal{S}} h \langle \nabla u_h \cdot \mathbf{n}, \nabla v_h \cdot \mathbf{n} \rangle_{0, \partial\Omega_D} \\
&\quad + \alpha_{\mathcal{S}} h \langle \mathcal{R}(u_h), \mathcal{R}^*(v_h) \rangle_{0, \partial\Omega_R}.
\end{aligned} \tag{3.58}$$

The numerical results presented in this section have been performed using the Bessel solution introduced in (2.116). All computations in this section were performed using the CutFEM library in the FEniCs software package which is available from <https://fenicsproject.org/>.

Let $\Omega \subset \mathbb{R}^2$ be a ball with radius 0.5 centred at $(0, 0)$

$$\begin{cases} -\Delta u - k^2 u = f & \text{in } \Omega \\ u = g_D & \text{on } \partial\Omega_D \\ \nabla u \cdot \mathbf{n} + iku = g_R & \text{on } \partial\Omega_R \end{cases}$$

where $\partial\Omega_D \stackrel{\text{def}}{=} \partial\Omega \cap [0, -0.5] \times [-0.5, 0.5]$ and $\partial\Omega_R \stackrel{\text{def}}{=} \partial\Omega \setminus \partial\Omega_D$. For our ansatz to be the solution let

$$f = \frac{\sin(k\sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}.$$

The domain Ω and background mesh are as presented in Figure 3.5. The stabilization parameters are chosen as the optimal parameters studied in the previous chapter. The newly introduced Nitsche parameters are both taken to be 1.0

Figure 3.8 shows a convergence study of three methods. The first method denoted ‘‘ImagNitsche’’ enforces Dirichlet boundary conditions on part of the boundary using Nitsche’s method. The penalty parameter introduced by Nitsche’s method is purely imaginary in this case. No additional stabilization is included in this method. The second method denoted ‘‘RealNitsche’’ also enforces Dirichlet boundary conditions on part of the boundary using Nitsche’s method although in this case the penalty parameter is taken as real. The final method presented, denoted ‘‘FullStab’’, is the fully stabilized fictitious domain method given by (3.31).

The plot shows that the stabilized method behaves as the theory would suggest, reaching optimal convergence when the mesh becomes sufficiently fine. ‘‘ImagNitsche’’ behaves well and converges optimally for low wave number but does not converge optimally for the case $k = 50$. Finally, the ‘‘RealNitsche’’ method behaves strangely even at low wavenumber which could perhaps be expected considering this case is the furthest from the framework of Theorem 2.

The second figure in this section is a pollution study that fixes $kh = C$. The mesh is then refined

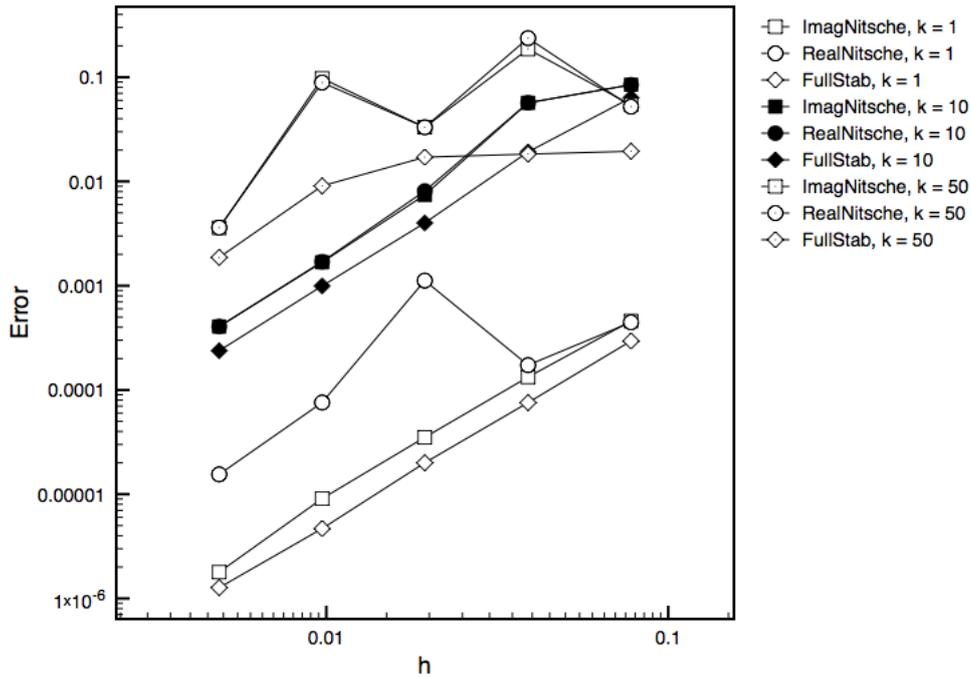


Figure 3.8: Convergence study for fictitious domain method using piecewise linear elements

(subsequently increasing k) and the relative L^2 error is plotted against the wavenumber. In the previous cases it was possible to tune the stabilization parameters to reduce the effect of numerical pollution for a given test case. Unfortunately, after trying a number of different combinations of stabilization parameters I was unable to replicate this result for the fictitious domain case. Figure 3.9 shows that the fully stabilized fictitious domain method is slightly more accurate but does little to reduce the pollution effect. All methods can be seen to be affected by numerical pollution. An interesting observation is that the “RealNitsche” method has an unexplained peak for the same mesh refinement as it did in the previous plot. This is most likely the result of an unfavourable cut.

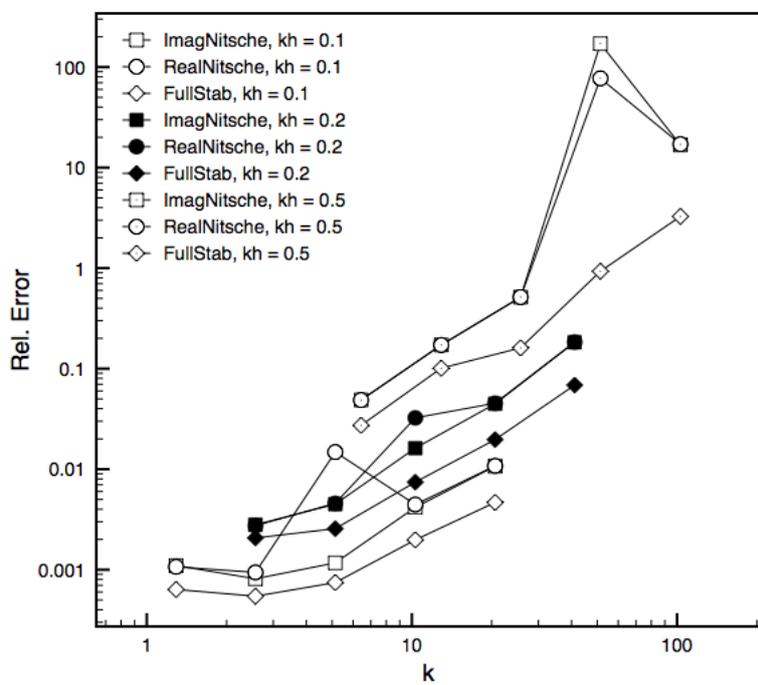


Figure 3.9: Pollution study for fictitious domain method using piecewise linear elements

Chapter 4

Multi-domain coupling for Helmholtz equations

4.1 Introduction

Multi-domain coupling is an area of practical importance in numerous fields of physics and engineering. Partial Differential Equations used to model physical phenomena often rely on various parameters which depend on the material properties of the underlying domain. For instance, the wave number in the Helmholtz equation is inversely proportional to the square root of the density of the medium in which it is being modelled. In areas such as Seismology it is necessary to model huge domains with varying underlying properties. This type of problem leads to a coupling of multiple Helmholtz type equations with different wave numbers. A natural way to model this type of problem mathematically is to use some type of domain decomposition method. Domain decomposition methods have numerous uses in the numerical solution of PDEs since they are highly adaptable and perfect candidates to extort the power of parallel computing. The goal of this work is to develop a robust domain decomposition method to study the coupling of multiple Helmholtz systems.

One of the most important factors to consider when designing a domain decomposition method is how to enforce continuity across the interfaces between sub-domains. The options fall into two main categories

- Iterative procedures
- Direct procedures

Iterative procedures, like the alternating Schwarz method presented by Lions in [41], enforce continuity across the interface by passing information from one sub-domain to the next iteratively. This process can be computationally expensive and when considering a problem, like Helmholtz equation with highly contrasting wave numbers, these computational costs can be either computationally prohibitive or lead to slow methods that take a long time to converge. On the other hand, direct methods using Lagrange

multiplier techniques lead to a solvable global system directly but also introduce additional unknowns that must be solved for. An alternative approach, which has become increasingly popular, is to enforce continuity across the interface using Nitsche's method. Nitsche's method has the advantage of being a direct method that does not introduce any additional unknowns and is consistent with the underlying PDE.

Nitsche's method was introduced in the context of multi-domain coupling in a paper by Hansbo and Hansbo [33] and has also been shown have a host of different applications including:

- Enforcing inter-element continuity in various DG methods [2].
- Fitted domain decomposition methods with matching and non-matching meshes [8],[52].
- Fictitious domain methods [20].
- Unfitted domain decomposition methods using cut elements [33],[58], [55],[13].

In this section I will introduce and analyse new domain decomposition methods capable of coupling multiple Helmholtz type equations with varying wave numbers. A variant of Nitsche's method is used to enforce the transmission conditions across the boundary. The method can be seen to fit into a similar mathematical framework as I proposed in Theorem 3 which I will present in this chapter. I will demonstrate the robustness of the method by applying it to both matching and non-matching grids in the fitted case and later extend the method to consider the case of an embedded interface that does not fit the computational mesh. Optimal convergence follows from the satisfaction of the assumption of Theorem 3 that I will verify below. The theoretical results are then verified numerically.

4.2 The coupled problem

Let Ω be a star-shaped bounded domain in \mathbb{R}^d with boundary $\partial\Omega$. Now let Ω be divided into two non-overlapping, polygonal sub-domains Ω_1 and Ω_2 with interface Γ such that

$$\Gamma = \bar{\Omega}_1 \cap \bar{\Omega}_2.$$

$$\text{Int} \left(\bigcup_j \Omega_j \cup \Gamma \right) = \Omega \text{ where } j \in \{1, 2\}. \quad (4.1)$$

It is also useful to introduce the following notation

$$\partial\Omega_j = \partial\Omega \cap \bar{\Omega}_j.$$

to define the intersection between the boundary of our domain and the boundary of our sub-domains.

Denote the space of square integrable functions on $\Omega_1 \cup \Omega_2$ by $L^2(\Omega_1 \cup \Omega_2)$. Let $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ be a multi-index with $|\alpha| = \sum_{i=1}^n \alpha_i$. For $m \in \mathbb{N}_0$, $H^m(\Omega_1 \cup \Omega_2)$ denotes the set of functions $u \in L^2(\Omega_1 \cup \Omega_2)$ such that all weak partial derivatives $\partial^\alpha u$ with $|\alpha| \leq m$ are also in $L^2(\Omega_1 \cup \Omega_2)$. Spaces $L^2(\Omega_1 \cup \Omega_2)$ and $H^m(\Omega_1 \cup \Omega_2)$ are equipped with norms:

$$\|u\|_{0, \Omega_1 \cup \Omega_2}^2 \stackrel{\text{def}}{=} \int_{\Omega_1 \cup \Omega_2} |u|^2. \quad (4.2)$$

$$\|u\|_{m, \Omega_1 \cup \Omega_2}^2 \stackrel{\text{def}}{=} \sum_{|\alpha| \leq m} \left(\int_{\Omega_1} |\partial^\alpha u|^2 + \int_{\Omega_2} |\partial^\alpha u|^2 \right). \quad (4.3)$$

In this setting it is now possible to define the coupled problem as defined in [9]. Consider two complex Helmholtz equations coupled in Ω . The problem then becomes: find $u = (u_1, u_2) \in H^p(\Omega_1 \cup \Omega_2)$, such that

$$\left. \begin{aligned} -\nabla \cdot (\mu_j \nabla u) - \omega^2 \rho_j u &= f && \text{in } \Omega_j \\ \nabla u \cdot n + i\omega \sqrt{\frac{\rho_j}{\mu_j}} u &= g && \text{on } \partial\Omega_j \\ u|_1 &= u|_2 && \text{on } \Gamma \\ \mu_1 \nabla u|_1 \cdot n_\Gamma &= \mu_2 \nabla u|_2 \cdot n_\Gamma && \text{on } \Gamma, \end{aligned} \right\} \quad (4.4)$$

where $j \in \{1, 2\}$, n_Γ is the outward facing normal with respect to Ω_1 , ω is the frequency of the harmonic oscillations and the coefficients μ_j and ρ_j are strictly positive and bounded functions which represent characteristics of the non-dispersive medium. I shall assume that

$$\frac{\max_j \{\mu_j\}}{\min_j \{\mu_j\}} < C \text{ and } \frac{\max_j \{\rho_j\}}{\min_j \{\rho_j\}} < C.$$

The values of these coefficients depend on the physics of the problem being solved. The additional boundary conditions imposed on the interfaces between sub-domains are introduced to enforce continuity over the global domain and are known as the transmission conditions. Transmission conditions of this type are popular in modern Discontinuous Galerkin methods where each element can be considered an individual sub-domain. It is useful in my analysis to note that the primal problem stated in (4.4) has a uniquely defined associated adjoint problem which is defined as

$$\left. \begin{aligned} -\nabla \cdot (\mu_j \nabla z) - \omega^2 \rho_j z &= \psi && \text{in } \Omega_j \\ \nabla z \cdot n - i\omega \sqrt{\frac{\rho_j}{\mu_j}} z &= 0 && \text{on } \partial\Omega_{j,R} \\ z|_1 &= z|_2 && \text{on } \Gamma \\ \mu_1 \nabla z|_1 \cdot n_\Gamma &= \mu_2 \nabla z|_2 \cdot n_\Gamma && \text{on } \Gamma, \end{aligned} \right\} \quad (4.5)$$

where $j \in \{1, 2\}$ and n_Γ is the outward facing normal with respect to Ω_1 . Notice that since μ_j is allowed to vary between sub-domains the Neumann transmission condition must be adjusted to include this variation. The adjoint Problem comes from

$$\sum_{j=1}^2 (\mathcal{L}^j(u), v)_{0, \Omega_j} = \sum_{j=1}^2 (u, \mathcal{L}^{j*}(v))_{0, \Omega_j}.$$

This give that

$$\begin{aligned} \sum_{j=1}^2 (\mathcal{L}^j(u), v)_{0, \Omega_j} &= \sum_{j=1}^2 (-\nabla \cdot (\mu_j \nabla u) - \omega^2 \rho_j u, v)_{0, \Omega_j} \\ &= \sum_{j=1}^2 (\mu_j \nabla u, \nabla v)_{0, \Omega_j} - \omega^2 \rho_j (u, v)_{0, \Omega_j} + i\omega \sqrt{\rho_j \mu_j} \langle u, v \rangle_{0, \partial \Omega_j} \\ &\quad - \langle \{\mu_j \nabla u \cdot n_\Gamma\}, \llbracket v \rrbracket \rangle_{0, \Gamma} \\ &= \sum_{j=1}^2 (u, -\nabla \cdot (\mu_j \nabla v))_{0, \Omega_j} - \omega^2 \rho_j (u, v)_{0, \Omega_j} \\ &\quad + \underbrace{i\omega \sqrt{\rho_j \mu_j} \langle u, v \rangle_{0, \partial \Omega_j} + \langle u, \mu_j \nabla v \cdot n \rangle_{0, \partial \Omega_j}}_{(I)} \\ &\quad - \langle \{\mu_j \nabla u \cdot n_\Gamma\}, \llbracket v \rrbracket \rangle_{0, \Gamma} + \langle \llbracket u \rrbracket, \{\mu_j \nabla v \cdot n_\Gamma\} \rangle_{0, \Gamma}. \end{aligned} \quad (4.6)$$

Since $\llbracket u \rrbracket = 0$ on Γ this simplifies to

$$\begin{aligned} \sum_{j=1}^2 (\mathcal{L}^j(u), v)_{0, \Omega} &= \sum_{j=1}^2 (u, \mathcal{L}^{j*}(v))_{0, \Omega_j} + \left\langle \mu_j u, \nabla v \cdot n - i\omega \sqrt{\frac{\rho_j}{\mu_j}} v \right\rangle_{0, \partial \Omega_j} \\ &\quad - \langle \{\mu_j \nabla u \cdot n_\Gamma\}, \llbracket v \rrbracket \rangle_{0, \Gamma}. \end{aligned}$$

This implies that the boundary condition for the adjoint problem should be

$$\mathcal{R}^{*,j}(v) = \nabla v \cdot n - ikv = 0,$$

and the jump over the interface should be

$$\llbracket v \rrbracket = 0.$$

Therefore the adjoint problem is

$$L^*(z) = -\nabla \cdot (\mu_j \nabla z) - \omega^2 \rho_j z = \psi \text{ in } \Omega_j \quad (4.7)$$

$$R^*(z) = \nabla z \cdot n - i\sqrt{\frac{\rho_j}{\mu_j}} z = 0 \text{ on } \partial \Omega_j \quad (4.8)$$

$$\llbracket z \rrbracket = 0 \text{ on } \Gamma \quad (4.9)$$

In the case that $\rho_j = \mu_j = 1$ the standard domain dependent regularity results hold via the methods found in either [42] or [25]. For Ω a bounded convex domain it holds that

$$\omega \|u\|_{0, \Omega} + \|u\|_{1, \Omega} \lesssim \mathcal{C}_{f,g}, \quad \|u\|_{2, \Omega_1 \cup \Omega_2} \lesssim \omega \mathcal{C}_{f,g}, \quad (4.10)$$

where $\mathcal{C}_{f,g} = \|f\|_{0,\Omega} + \|g_{R,j}\|_{0,\partial\Omega_j} + \frac{1}{\omega} \|g_{R,j}\|_{1/2,\partial\Omega_j}$. Under the same assumptions the adjoint problem satisfies similar bounds

$$\omega \|z\|_{0,\Omega} + \|z\|_{1,\Omega} \lesssim \|\psi\|_{0,\Omega}, \quad \|z\|_{2,\Omega} \lesssim \omega \|\psi\|_{0,\Omega}. \quad (4.11)$$

A regularity estimate for the case ρ_j discontinuous can be found in [37] where the authors use the ideas in [25] to derive a regularity result, which is of the same order as the classical results providing ρ_j has certain nice properties. The regularity for the high-contrast Laplacian was studied in [23]. More recently the regularity of the acoustic transmission problem has been studied by Moiola and Spence in [44]. The authors derive estimates for the case where Ω_1 is star shaped and $\Omega_2 = \mathbb{R}^n \setminus \bar{\Omega}_1$. The result relies on the source term having compact support and sufficient regularity. The authors remark that it is possible to extend the results of the paper to the truncated case where Ω_2 is a star shaped domain containing Ω_1 and the radiation condition is replaced by the impedance boundary condition.

Therefore, taking Ω_1, Ω_2 star shaped with $\Omega_1 \subset \Omega_2$ and

$$\frac{\rho_1}{\rho_2} \leq 1 \leq \frac{\mu_1}{\mu_2}$$

I assume that the transmission problem as posed in (4.4), for $\omega \gg 0$, satisfies the following regularity estimate

$$\sum_j \mu_j \|\nabla u\|_{0,\Omega_j}^2 + \omega^2 \rho_j \|u\|_{0,\Omega_j}^2 \leq \sum_j \left(\frac{C(\Omega_j)}{\mu_j} + \frac{C'(\Omega_j)^2}{\rho_j} \right) \mathcal{C}'_{f,g}, \quad (4.12)$$

where $\mathcal{C}'_{f,g} = \|f_j\|_{0,\Omega_j}^2 + \|g_{R,2}\|_{0,\partial\Omega_2}^2 + \frac{1}{\omega^2} \|g_{R,2}\|_{1/2,\partial\Omega_2}^2$

Assuming sufficient regularity on u it is possible to derive the following estimates

$$\sum_j \|u\|_{2,\Omega_j}^2 \leq C \max_j \left\{ \frac{\omega^2 \rho_j}{\mu_j} \right\} \mathcal{C}_{f,g}, \quad (4.13)$$

and

$$\sum_j \|z\|_{2,\Omega_j}^2 \leq C \max_j \left\{ \frac{\omega^2 \rho_j}{\mu_j} \right\} \|\psi\|_{0,\Omega}^2. \quad (4.14)$$

4.3 The fitted domain decomposition method

As I have already mentioned domain decomposition methods exist in many forms. In this section I shall pose a new fitted domain decomposition method for the coupling of multiple Helmholtz systems. A fitted domain decomposition method refers to a method where the mesh is designed such that it fits the interface exactly. To experience the full benefits of this method the interface should be simple enough to be modelled accurately by the mesh and remain fixed to avoid re-meshing. One drawback of the method is that it requires the user to know the exact position of the interface in order to construct an appropriate

mesh.

When considering fitted domain decomposition methods it is possible to imagine two cases. The first is where the meshes on either side of the interface share the same common nodes. I will refer to this as a fitted domain decomposition method with matching mesh. The second case is where the meshes on either side of the interface do not share all of the same common nodes. I will refer to this as a fitted domain decomposition method with non-matching mesh. In this section I shall consider both matching and non-matching meshes and present numerical results for both cases. In [8] the authors showed the versatility of Nitsche's method for handling both matching and non-matching mesh cases for Poisson's equation. In this section I explore the capabilities of a similar Nitsche style coupling for applications involving Helmholtz equation.

4.3.1 Preliminaries

Before introducing the formulation it is useful to introduce the following notation. Let $\{\mathcal{T}_h^j\}_h$ denote a family of quasi-uniform and shape regular triangulations fitted to sub-domain Ω_j where shape regularity and quasi uniformity are defined as in (A.3) and (A.4) respectively.

These triangulations introduce a 'trace' mesh on the interface which is defined as

$$\mathcal{G}_h^j \stackrel{\text{def}}{=} \{E : E = \tau \cap \Gamma, \tau \in \mathcal{T}_h^j\}. \quad (4.15)$$

It is also useful to define the set of elements that have an edge in the trace mesh

$$\mathcal{T}_\Gamma = \{\tau \in \mathcal{T}_{1,h} \cup \mathcal{T}_{2,h} : \partial\tau_\Gamma = \bar{\tau} \cap \Gamma \neq \emptyset\}. \quad (4.16)$$

In order to derive optimal a priori error estimates it is necessary to define

$$h \stackrel{\text{def}}{=} \max\{h_\tau, h_E : \tau \in \mathcal{T}_h^j, E \in \mathcal{G}_h^j, j \in \{1, 2\}\}, \quad (4.17)$$

where h_τ and h_E denote the diameter of an element $\tau \in \mathcal{T}_h^j$ and $E \in \mathcal{G}_h^j$. The analysis requires that the meshes on both domains are of comparable size.

Assumption 1. *There exist positive constants C_1, C_2, C'_1 and C'_2 such that*

$$C_1 h_{E_1} \leq h_{E_2} \leq C_2 h_{E_1} \text{ and } C'_1 h_{\tau_1} \leq h_{\tau_2} \leq C'_2 h_{\tau_1}, \quad (4.18)$$

holds for all pairs $(E_1, E_2) \in \mathcal{T}_h^1 \times \mathcal{T}_h^2$, with $E_1 \cap E_2 \neq \emptyset$ and $(\tau_1, \tau_2) \in \mathcal{T}_h^1 \times \mathcal{T}_h^2$.

Finally, I introduce the space

$$\mathcal{F}_{int}^j \stackrel{\text{def}}{=} \{F : F = \tau_1 \cap \tau_2 \forall \tau_1, \tau_2 \in \mathcal{T}_h^j\}, \quad (4.19)$$

to denote the space of all internal faces in a triangulation \mathcal{T}_h^j .

On the interface I shall use the notation:

$$[[v]] \stackrel{\text{def}}{=} v_1 - v_2, \quad (4.20)$$

to define the jump of v over Γ .

$$\{v\} \stackrel{\text{def}}{=} \frac{v_1 + v_2}{2}, \quad (4.21)$$

to define the average of v over Γ .

Remark 1. *It is useful to note that if I choose to fix $n_\Gamma \stackrel{\text{def}}{=} n_1$ on the interface it is also true that $n_\Gamma \stackrel{\text{def}}{=} -n_2$ which implies that on the interface the jump average becomes*

$$\{\nabla v \cdot n_\Gamma\} \stackrel{\text{def}}{=} \frac{\nabla v_1 \cdot n_1 - \nabla v_2 \cdot n_2}{2}. \quad (4.22)$$

4.3.2 The formulation

The following analysis is restricted to the case of 2 sub-domains. The extension to a larger number of sub-domains is a fairly straightforward exercise. The following analysis only considers matching meshes but it is not too difficult to extend this to the non-matching case. Introducing the continuous space,

$$V^j \stackrel{\text{def}}{=} \{v_j \in H^1(\Omega_j) : \nabla v_j \cdot n \in L^2(\Gamma), v_j|_{\partial\Omega \cap \partial\Omega_j} = 0\}, \text{ for } j \in \{1, 2\}, \quad (4.23)$$

I define my finite element space as $V_h \stackrel{\text{def}}{=} V_h^1 \times V_h^2$

$$V_h^j \stackrel{\text{def}}{=} \{v_j \in V^j : v_j|_\tau \in \mathbb{P}^p(\tau) \forall \tau \in \mathcal{T}_h^j : p \in \mathbb{N}\}, \quad (4.24)$$

where \mathbb{P}^p denotes a polynomial of degree p . It is useful in the subsequent analysis to pose the numerical method using the following abstract formulation. The problem becomes find $u_h \in V_h$: such that

$$A_h(u_h, v_h) = A(u_h, v_h) + S(u_h, v_h) = L(v_h) + S(u, v_h) = L_h(v_h), \quad (4.25)$$

where the sesquilinear form $A(., .)$ represents the Nitsche formulation applied to the physical problem and is given by

$$A(u_h, v_h) = \sum_{j=1}^2 (\mu_j \nabla u_j, \nabla v_j)_{0, \Omega_j} - (\omega^2 \rho_j u_j, v_j)_{0, \Omega_j} + \langle i\omega \sigma_j u_j, v_j \rangle_{0, \partial \Omega_j} - \langle \{\mu \nabla u_h \cdot n_\Gamma\}, \llbracket v_h \rrbracket \rangle_{0, \Gamma} \\ - \langle \llbracket u_h \rrbracket, \{\mu \nabla v_h \cdot n_\Gamma\} \rangle_{0, \Gamma},$$

where $\sigma_j = \sqrt{\rho_j \mu_j}$ and the global stabilization term $S(., .)$ representing the additional non-physical terms introduced to improve stability is expressed as

$$S(u_h, v_h) = \langle i\lambda_1 h^{-1} \llbracket u_h \rrbracket, \llbracket v_h \rrbracket \rangle_{0, \Gamma} + \left\langle i\hat{\lambda}_1 h \llbracket \mu \nabla u_h \cdot n_\Gamma \rrbracket, \llbracket \mu \nabla v_h \cdot n_\Gamma \rrbracket \right\rangle_{0, \Gamma} + \sum_{j=1}^2 s_j(u_h, v_h).$$

The local stabilization $s_j(., .)$ is taken to be the GLS/CIP or CIP methods which act on the interior of the subdomain Ω_j as well as the boundary least squares term strengthening the control of the impedance boundary condition on $\partial \Omega_j$. The GLS/CIP stabilization form is defined as

$$s_j(u_h, v_h) \stackrel{\text{def}}{=} \sum_{\tau_j \in \mathcal{T}_h^j} \delta_{1, \tau}^j h_\tau^2 (\mathcal{L}^j(u_h), \mathcal{L}^j(v_h))_{0, \tau_j} + \sum_{F_j \in \mathcal{F}_{int}^j} \gamma_{1, \tau}^j h_\tau \langle \llbracket \mu_j \nabla u_h \cdot \mathbf{n} \rrbracket, \llbracket \mu_j \nabla v_h \cdot \mathbf{n} \rrbracket \rangle_{0, F_j} \\ + \beta_1 h \langle \mathcal{R}^j(u_h), \mathcal{R}^{*,j}(v_h) \rangle_{0, \partial \Omega_j}, \quad (4.26)$$

where $\mathcal{L}^j(\cdot) \stackrel{\text{def}}{=} -\nabla \cdot (\mu_j \nabla (\cdot)) - \omega^2 \rho_j (\cdot)$, $\mathcal{R}^j(\cdot) \stackrel{\text{def}}{=} \nabla(\cdot) \cdot \mathbf{n} + i\omega \sqrt{\frac{\rho_j}{\mu_j}}(\cdot)$ and $\mathcal{R}^{*,j}(\cdot) \stackrel{\text{def}}{=} \nabla(\cdot) \cdot \mathbf{n} - i\omega \sqrt{\frac{\rho_j}{\mu_j}}(\cdot)$. The CIP stabilization form is defined as

$$s_j(u_h, v_h) \stackrel{\text{def}}{=} \sum_{F_j \in \mathcal{F}_{int}^j} \delta_{0, \tau}^j h_\tau^3 \langle \llbracket \nabla \cdot (\mu_j \nabla u_h) \rrbracket, \llbracket \nabla \cdot (\mu_j \nabla v_h) \rrbracket \rangle_{0, F_j} \quad (4.27)$$

$$+ \sum_{F_j \in \mathcal{F}_{int}^j} \gamma_{0, \tau}^j h_\tau \langle \llbracket \mu_j \nabla u_h \cdot \mathbf{n} \rrbracket, \llbracket \mu_j \nabla v_h \cdot \mathbf{n} \rrbracket \rangle_{0, F_j} \quad (4.28)$$

$$+ \beta_0 h \langle \mathcal{R}^j(u_h), \mathcal{R}^{*,j}(v_h) \rangle_{0, \partial \Omega_j}, \quad (4.29)$$

Finally, the linear form $L(\cdot)$ on the RHS is defined as

$$L(v_h) \stackrel{\text{def}}{=} \sum_{j=1}^2 (f, v_h)_{0, \Omega_j} + \langle g, v_h \rangle_{0, \partial \Omega_j}. \quad (4.30)$$

Remark 2. The Nitsche term added to the formulation $A(., .)$ is equal to zero for u as the exact solution of (4.4).

$$\langle \llbracket u \rrbracket, \{\mu \nabla v \cdot n_\Gamma\} \rangle_{0, \Gamma} = 0 \quad \forall v \in V^1 \times V^2$$

this term is equivalent to the symmetric term added in the Nitsche formulation to weakly impose Dirichlet

boundary conditions. The consistency term in the Nitsche formulation is equivalent to the term

$$\langle \{\mu \nabla u_h \cdot n_\Gamma\}, \llbracket v_h \rrbracket \rangle_{0,\Gamma}$$

Observe that for u , as the exact solution of the coupled problem given in (4.4), the stabilization term $S(\cdot, \cdot)$ satisfies the following equality

$$S(u, v_h) = \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} \delta_{1,\tau}^j h_\tau^2 (f, \mathcal{L}^j(v_h))_{0,\tau} + \beta_1 h \langle g, \mathcal{R}^*(v_h) \rangle_{0,\partial\Omega_j} \right). \quad (4.31)$$

Lemma 4.1. *The solution $u = (u_1, u_2)$ to (4.4) satisfies*

$$A_h(u, v) = L_h(v) \quad \forall v \in V.$$

Proof. It is immediate from (4.4) that

$$A(u, v) + S(u, v) = L(v) + S(u, v) \quad \forall v \in V. \quad (4.32)$$

$$\implies A(u, v) = L(v) \quad \forall v \in V. \quad (4.33)$$

It can then be shown that

$$\begin{aligned} L(v) &= \sum_{j=1}^2 (f, v_j)_{0,\Omega_j} + \langle \mu_j g, v_j \rangle_{0,\partial\Omega_j} \\ &= \sum_{j=1}^2 \left[(\mu_j \nabla u_j, \nabla v_j)_{0,\Omega_j} - \omega^2 \rho_j (u_j, v_j)_{0,\Omega} - i\omega \sigma_j \langle u_j, v_j \rangle_{0,\partial\Omega_j} \right. \\ &\quad \left. - \langle \mu_j \nabla u_j \cdot n_j, v_j \rangle_{0,\Gamma} \right] \\ &= \sum_{j=1}^2 \left[(\mu_j \nabla u_j, \nabla v_j)_{0,\Omega_j} - \omega^2 \rho_j (u_j, v_j)_{0,\Omega} - i\omega \sigma_j \langle u_j, v_j \rangle_{0,\partial\Omega_j} \right] \\ &\quad - \langle \{\mu \nabla u \cdot n_\Gamma\}, \llbracket v \rrbracket \rangle_{0,\Gamma}. \end{aligned} \quad (4.34)$$

Since $\llbracket u \rrbracket = 0$ on Γ it holds that

$$0 = - \langle \llbracket u \rrbracket, \{\mu \nabla v \cdot n_\Gamma\} \rangle_{0,\Gamma}. \quad (4.35)$$

Using (4.34) and (4.35) concludes the proof. \square

4.3.3 Mathematical framework for the coupled problem

The first assumption on the stabilization is that there exists some type of energy semi-norm $|\cdot|_{\mathcal{J}}$ with the following properties $|\cdot|_{\mathcal{J}} : V_h \mapsto \mathbb{R}$ such that

$$|u_h|_{\mathcal{J}}^2 \leq |A_h(u_h, u_h)|. \quad (4.36)$$

It can be noted that if $|\cdot|_{\mathcal{J}}$ is taken to be a norm on V then this estimate is analogous to the coercivity assumption of Lax-Milgram. I also ask that the following Cauchy-Schwarz type inequality holds for some semi-norm $|\cdot|_s : V \mapsto \mathbb{R}$ which is associated to the stabilization,

$$|s(v, w)| \leq |v|_{\mathcal{J}} |w|_s \leq |v|_{\mathcal{J}} |w|_{\mathcal{J}} \forall v, w \in V. \quad (4.37)$$

The Cauchy-Schwarz type inequality ensures that the whole stabilization can be upper bounded by the \mathcal{J} semi norm.

Recall that the analysis also relies on the existence of an appropriate interpolation operator $\pi_h : V \mapsto V_h$ which should respect the following estimate. For piecewise linear elements

$$\|u - \pi_h u\|_{0,\Omega} + h \|\nabla(u - \pi_h u)\|_{0,\Omega} \leq Ch^2 |u|_{2,\Omega_1 \cup \Omega_2}, \quad (4.38)$$

and for piecewise polynomial elements of order $p > 1$

$$\|u - \pi_h u\|_{0,\Omega} + h \|\nabla(u - \pi_h u)\|_{0,\Omega} + \sum_{\tau \in \mathcal{T}_h} h_{\tau}^2 \|D^2(u - \pi_h u)\|_{0,\tau} \leq Ch^{p+1} |u|_{p+1,\Omega_1 \cup \Omega_2}. \quad (4.39)$$

I also assume the existence of a norm, $\|\cdot\|_*$, such that the following continuity estimate is satisfied

$$|A(u - \pi_h u, v_h)| + |s(u - \pi_h u, v_h)| + 2C(\rho_j, \mu_j) \omega \|u - \pi_h u\|_{0,\partial\Omega_R} \|v_h\|_{0,\partial\Omega_R} \leq M \|u - \pi_h u\|_* |v_h|_{\mathcal{J}}. \quad (4.40)$$

A second continuity assumption is posed purely on the sesquilinear form $A(\cdot, \cdot)$ and is designed to include the interpolation error of the adjoint problem. In this estimate z and ψ are taken to be the exact solution and source term of the adjoint problem, as posed in (4.5), respectively.

$$|A(u - \pi_h u, z - \pi_h z)| \leq C(\rho_j, \mu_j) (\omega h) h^p |u|_{p+1,\Omega_1 \cup \Omega_2} \|\psi\|_{0,\Omega}. \quad (4.41)$$

However, for the analysis to hold some control is needed over $\pi_h z$. This control is asserted through a stability estimate of the form

$$|\pi_h z|_s \leq C(\rho_j, \mu_j) \omega h \|\psi\|_{0,\Omega}, \quad (4.42)$$

which is required ensure sharpness of our $L^2(\Omega)$ convergence estimate. Finally, the theorem requires interpolation estimates in $|\cdot|_{\mathcal{J}}$ and $\|\cdot\|_*$

$$\|u - \pi_h u\|_* \leq Ch^p |u|_{p+1, \Omega_1 \cup \Omega_2}, \quad (4.43)$$

and

$$|u - \pi_h u|_{\mathcal{J}} \leq Ch^p |u|_{p+1, \Omega_1 \cup \Omega_2}. \quad (4.44)$$

Now that the groundwork has been laid it is possible to construct our theorem. The main differences between Theorem 3 and Theorem 2 are the additional coefficients which must be handled appropriately and the use of broken norms.

Theorem 3. *Let $u \in H^{p+1}(\Omega_1 \cup \Omega_2)$, for $p > 1/2$ be the unique solution of (2.15) that satisfies the assumed regularity given in (4.13) and $u_h \in V_h$ be the solution of (2.16). If (2.16) satisfies properties (4.36)-(4.44) then the formulation admits a unique solution which satisfies the following a priori error estimates*

$$|u - u_h|_{\mathcal{J}} \leq Ch^p |u|_{p+1, \Omega_1 \cup \Omega_2} \quad (4.45)$$

$$\|u - u_h\|_{0, \Omega} \leq C(\rho_j, \mu_j) (\omega h) h^p |u|_{p+1, \Omega_1 \cup \Omega_2} \quad (4.46)$$

$$\|\nabla(u - u_h)\|_{0, \Omega_1 \cup \Omega_2} \leq C(\rho_j, \mu_j) (1 + (\omega^2 h)) h^p |u|_{p+1, \Omega_1 \cup \Omega_2}. \quad (4.47)$$

Proof. The proof follows the same outline as Theorem 2. □

In the formulation the coupling terms on the interface are enforced using Nitsche's method which was introduced earlier and has become a popular way of coupling two or more domains. The jump and average terms come from applying Nitsche boundary conditions to the interface between Ω_1 and Ω_2 then fixing a normal direction n_Γ with respect to the interface. This technique has been applied to a number of different problems with some success. The novel aspect of the method introduced in (4.25) is the introduction of an additional penalty term which penalises the jump of the flux across the interface. This additional penalty parameter gives absolute stability when considering Dirichlet or Neumann boundary conditions and also ensures stability in the regime $\max_j \left\{ \frac{\omega \sigma_j \mu_j^2}{h} \right\} < \beta_0^{-1}$ for the CIP stabilization or $\max_j \left\{ \frac{\omega \sigma_j \mu_j^2}{h} \right\} < \beta_1^{-1}$ for the GLS/CIP formulation. It will become clear in the following analysis why this term has been introduced. In order for the Nitsche coupling terms to enter the analytic framework I have fixed the penalty parameters to be purely imaginary which is also a unique property of this method.

4.3.4 Mathematical tools

For the following analysis it is beneficial to introduce the following semi-norms which act on the stabilization. Let $|\cdot|_{\mathcal{J}} : V_h \mapsto \mathbb{R}$ then for $u_h \in V_h$

$$\begin{aligned}
|u_h|_{\mathcal{J}}^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\omega \sigma_j \left(1 - \frac{\omega \sigma_j}{\mu_j^2} \beta_\epsilon h \right) \|u_h\|_{0,\partial\Omega_j}^2 + \sum_{F \in \mathcal{F}_{int}^j} (1 - \epsilon) \delta_{\epsilon,\tau} h_\tau^3 \|\llbracket \nabla \cdot (\mu \nabla u_h) \rrbracket\|_{0,F}^2 \right. \\
&\quad \left. + \sum_{\tau \in \mathcal{T}_h^j} \epsilon h_\tau^2 \delta_{\epsilon,\tau} \|\mathcal{L}^j(u_h)\|_{0,\tau}^2 + \sum_{F \in \mathcal{F}_{int}^j} \gamma_{\epsilon,\tau} h_\tau \|\llbracket \nabla u_h \cdot n \rrbracket\|_{0,F}^2 + h \beta_\epsilon \|\nabla u_h \cdot n\|_{0,\partial\Omega_j}^2 \right) \\
&\quad + \lambda_\epsilon h^{-1} \|\llbracket u_h \rrbracket\|_{0,\Gamma}^2 + \hat{\lambda}_\epsilon h \|\llbracket \mu \nabla u_h \cdot n_\Gamma \rrbracket\|_{0,\Gamma}^2. \tag{4.48}
\end{aligned}$$

Another useful semi-norm is closely related $|\cdot|_s : V_h \mapsto \mathbb{R}$

$$\begin{aligned}
|u_h|_s^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\beta_{\epsilon,j} h \|\nabla u_h \cdot n - i \frac{\omega \sigma_j}{\mu_j} u\|_{0,\partial\Omega_j}^2 + \sum_{F \in \mathcal{F}_{int}^j} (1 - \epsilon) \delta_{\epsilon,\tau} h_\tau^3 \|\llbracket \nabla \cdot (\mu \nabla u_h) \rrbracket\|_{0,F}^2 \right. \\
&\quad \left. + \sum_{\tau \in \mathcal{T}_h^j} \epsilon \delta_{\epsilon,\tau} h_\tau^2 \|\mathcal{L}^j(u_h)\|_{0,\tau}^2 + \sum_{F \in \mathcal{F}_{int}^j} \gamma_{\epsilon,\tau} h_\tau \|\llbracket \nabla u_h \cdot n \rrbracket\|_{0,F}^2 \right) \\
&\quad + \lambda_\epsilon h^{-1} \|\llbracket u_h \rrbracket\|_{0,\Gamma}^2 + \hat{\lambda}_\epsilon h \|\llbracket \mu \nabla u_h \cdot n_\Gamma \rrbracket\|_{0,\Gamma}^2, \tag{4.49}
\end{aligned}$$

where $\epsilon \in \{0, 1\}$, this provides a switch between the CIP and GLS/CIP methods. When $\epsilon = 0$ the stabilization reduces to the CIP method and when $\epsilon = 1$ the stabilization is the GLS/CIP method.

Lemma 4.2. *Given $\epsilon \in \{0, 1\}$, $Re[\lambda_\epsilon] = Re[\hat{\lambda}_\epsilon] = Re[\beta_{\epsilon,j}] = 0$, $\min_j \{Im[\beta_{\epsilon,j}^{-1}]\} > h \frac{\omega \sigma_j}{\mu_j^2}$ and the imaginary parts of all penalty parameters are strictly positive, the sesquilinear form $A_h(u_h, u_h)$, defined in (4.25), satisfies the equality*

$$|\cdot|_{\mathcal{J}}^2 = Im[A_h(u_h, u_h)],$$

for $|\cdot|_{\mathcal{J}}$, as defined in (4.48), which under the conditions of the lemma defines a semi-norm on V_h .

Proof. The proof follows by noticing that

$$Im[A_h(u_h, u_h)] = Im[A(u_h, u_h)] + Im[S(u_h, u_h)]. \tag{4.50}$$

Inspecting these forms individually shows that

$$\begin{aligned}
A(u_h, u_h) &= \sum_{j=1}^2 \mu_j \|\nabla u_h^j\|_{0,\Omega_j}^2 - \omega^2 \rho_j \|u_h^j\|_{0,\Omega_j}^2 + i \omega \sigma_j \|u_h^j\|_{0,\partial\Omega_j}^2 \\
&\quad - 2Re \left[\langle \{\mu \nabla u_h \cdot n_\Gamma\}, \llbracket u_h \rrbracket \rangle_{0,\Gamma} \right],
\end{aligned}$$

where the last line follows from an application of (A:11). Taking the imaginary part of this gives

$$\operatorname{Im} [A(u_h, u_h)] = \omega \sigma_j \|u_h^j\|_{0, \partial \Omega_j}^2, \quad (4.51)$$

Investigating the stabilization term

$$S(u_h, u_h) = i \lambda_\epsilon h^{-1} \| [u_h] \|_{0, \Gamma}^2 + i \hat{\lambda}_\epsilon h \| [\mu \nabla u_h \cdot n_\Gamma] \|_{0, \Gamma}^2 + \sum_{j=1}^2 s_j (u_h^j, u_h^j), \quad (4.52)$$

and taking the imaginary parts gives

$$\operatorname{Im} [S(u_h, u_h)] = \lambda_\epsilon h^{-1} \| [u_h] \|_{0, \Gamma}^2 + \hat{\lambda}_\epsilon h \| [\mu \nabla u_h \cdot n_\Gamma] \|_{0, \Gamma}^2 + \operatorname{Im} \left[\sum_{j=1}^2 s_j (u_h^j, u_h^j) \right]. \quad (4.53)$$

The claim follows from an application of the Cauchy-Schwarz inequality to the term $\operatorname{Im} \left[\sum_{j=1}^2 s_j (u_h^j, u_h^j) \right]$ as shown previously. The claim that $|\cdot|_j$ defines a semi-norm on V_h is immediate, under the conditions of the lemma, from its definition. \square

It is also useful to introduce the following norms on V . For $\epsilon = 1$ let $\|(\cdot)\| : V \mapsto \mathbb{R}$ where

$$\begin{aligned} \|(\cdot)\|^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\left[\sum_{\tau_j \in \mathcal{T}_{j,h}} (h_{\tau_j}^2 |\delta_{1,\tau_j}|)^{-1} \|(\cdot)\|_{0,\tau_j}^2 + \sum_{F_j \in \mathcal{F}_{int}^j} (h_{\tau_j} |\gamma_{1,\tau_j}|)^{-1} \|(\cdot)\|_{0,F_j}^2 \right. \right. \\ &\quad \left. \left. + \mu_j (h |\beta_{1,j}|)^{-1} \|(\cdot)\|_{0,\partial \Omega_j}^2 \right] + h |\lambda_1|^{-1} \| \{ \mu \nabla(\cdot) \cdot n_\Gamma \} \|_{0,\Gamma}^2 + (h |\hat{\lambda}_1|)^{-1} \| \{ (\cdot) \} \|_{0,\Gamma}^2 \right. \\ &\quad \left. + |(\cdot)|_s^2 \right), \end{aligned} \quad (4.54)$$

and for $\epsilon = 0$ let $\|(\cdot)\|_* : V \mapsto \mathbb{R}$ where

$$\begin{aligned} \|(\cdot)\|_*^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\left[\sum_{\tau_j \in \mathcal{T}_{j,h}} (h_{\tau_j}^2 |\delta_{0,\tau_j}|)^{-1} \|(\cdot)\|_{0,\tau_j}^2 + \sum_{F_j \in \mathcal{F}_{int}^j} (h_{\tau_j} |\gamma_{0,\tau_j}|)^{-1} \|(\cdot)\|_{0,F_j}^2 \right. \right. \\ &\quad \left. \left. + \mu_j (h |\beta_{0,j}|)^{-1} \|(\cdot)\|_{0,\partial \Omega_j}^2 \right] + h |\lambda_0|^{-1} \| \{ \mu \nabla(\cdot) \cdot n_\Gamma \} \|_{0,\Gamma}^2 + (h |\hat{\lambda}_0|)^{-1} \| \{ (\cdot) \} \|_{0,\Gamma}^2 \right. \\ &\quad \left. + |(\cdot)|_s^2 \right). \end{aligned} \quad (4.55)$$

It is clear to see that these are both norms on V so I shall omit the proof.

4.3.5 A priori error estimates

In the following section the finite element scheme defined by (4.25) will be analysed using the theoretical framework introduced in Theorem 3. Well posedness of the discrete system is a direct result of the a priori error estimates derived in the analysis. In order to show that the coupled formulation fits into the framework of Theorem 3 I first show that the stabilization has the weak coercivity property and respects the interpolation estimates given by assumptions (4.36), (4.37), (4.42), (4.43) (4.44).

Lemma 4.3. *Under the assumptions given in Lemma 4.2 the domain decomposition method introduced in (4.25) satisfies the weak coercivity and Cauchy-Schwarz properties given by, (4.36) and (4.37) respectively, for $|\cdot|_{\mathcal{J}}$ defined in (4.48) and $|\cdot|_s$ defined in (4.49).*

Proof. The weak coercivity property follows directly from Lemma 4.2 since

$$|u|_{\mathcal{J}}^2 = \text{Im}[A_h(u_h, u_h)] \quad (4.56)$$

$$\leq |A_h(u_h, u_h)|. \quad (4.57)$$

The Cauchy-Schwarz style inequalities follow from

$$\begin{aligned} |S(v_h, w_h)| &= \gamma_0 h^{-1} \langle \llbracket v_h \rrbracket, \llbracket w_h \rrbracket \rangle_{0,\Gamma} + \gamma_1 h \langle \llbracket \nabla v_h \cdot n \rrbracket, \llbracket \nabla w_h \cdot n \rrbracket \rangle_{0,\Gamma} + s(v_h, w_h) \\ &\leq \gamma_0 h^{-1} \|\llbracket v_h \rrbracket\|_{0,\Gamma} \|\llbracket w_h \rrbracket\|_{0,\Gamma} + \gamma_1 h \|\llbracket \nabla v_h \cdot n \rrbracket\|_{0,\Gamma} \|\llbracket \nabla w_h \cdot n \rrbracket\|_{0,\Gamma} + \sum_{j=1}^2 |s(v_h^j, w_h^j)|. \end{aligned}$$

An application of the Cauchy-Schwarz inequality gives

$$|s(v_h, w_h)| \leq |s(v_h, v_h)|^{\frac{1}{2}} |s(w_h, w_h)|^{\frac{1}{2}}.$$

Which proves the result. \square

Lemma 4.4. *Under the same assumptions as Lemma 4.2 the semi-norm $|\cdot|_{\mathcal{J}}$, defined in (4.48), satisfies (4.42) and (4.44) for both GLS/CIP and CIP stabilizations introduced in Chapter 2 where π_h^0 is chosen to be the Lagrange interpolant proposed in (2.64) and π_h^1 the L^2 projection on each sub-domain proposed in (A.5) respectively. The norms (4.54) and (4.55) both satisfy (4.43) for the same respective choice of π_h .*

Proof. For the purpose of this proof it is beneficial to introduce some additional notation to handle the two different cases. Let

$$\delta_\tau^j = \epsilon \delta_{1,\tau}^j \text{ and } \hat{\delta}_\tau^j = (1 - \epsilon) \delta_{0,\tau}^j,$$

where $\epsilon \in \{0, 1\}$. I also introduce the interpolation operator π_h^ϵ where π_h^0 is the L^2 projection defined in (A.5) and π_h^1 is the standard Lagrange Interpolant. Using this notation allows me to study both the GLS/CIP and CIP methods introduced in Chapter 2 simultaneously. As stated earlier in the Chapter I will use ϵ as a switch to change between stabilizations, noting that when $\epsilon = 0$ the stabilization in the interior of each sub-domain is the CIP method and when $\epsilon = 1$ the stabilization in the interior of each

sub-domain is the GLS/CIP method. Therefore using this notation it holds that

$$|\pi_h^\epsilon z|_s^2 = \sum_{j=1}^2 \left[\sum_{F_j \in \mathcal{F}_{int}^j} \delta_\tau^j h_\tau^3 \|\llbracket \nabla \cdot (\mu \nabla \pi_h^\epsilon z) \rrbracket\|_{0,F_j}^2 + \sum_{\tau_j \in \mathcal{T}_h^j} \delta_\tau^j h_\tau^2 \|\mathcal{L}(\pi_h^\epsilon z)\|_{0,\tau_j}^2 \right] \quad (4.58)$$

$$+ \sum_{F_j \in \mathcal{F}_{int}^j} \gamma_{\epsilon,\tau}^j h_\tau \|\llbracket \mu \nabla \pi_h^\epsilon z \cdot \mathbf{n} \rrbracket\|_{0,F_j}^2 + \lambda_\epsilon h^{-1} \|\llbracket \pi_h^\epsilon z \rrbracket\|_{0,\Gamma}^2 \quad (4.59)$$

$$+ \left[\hat{\lambda}_\epsilon h \|\llbracket \mu \nabla \pi_h^\epsilon z \cdot \mathbf{n} \rrbracket\|_{0,\Gamma}^2 + \beta_{\epsilon,j} h \|\mathcal{R}^{*,j}(\pi_h^\epsilon z)\|_{0,\partial\Omega_j}^2 \right]. \quad (4.60)$$

The first, second, third and sixth terms all follow trivially by arguing in a similar way to Lemmas 2.7 and 2.8. In order to not repeat myself I will ignore these terms in the proof. Considering the fourth term and using the fact that the jump of z is zero over the interface Γ it follows that,

$$(IV) = \lambda_\epsilon h^{-1} \|\llbracket z - \pi_h^\epsilon z \rrbracket\|_{0,\Gamma}^2 \quad (4.61)$$

$$\leq \lambda_\epsilon h^{-1} \|z_1 - \pi_h^\epsilon z_1\|_{0,\Gamma}^2 + \gamma_{0,\Gamma} h^{-1} \|z_2 - \pi_h^\epsilon z_2\|_{0,\Gamma}^2 \quad (4.62)$$

$$\leq \lambda_\epsilon h^{-1} \sum_{j=1}^2 \sum_{\tau_j \in \mathcal{T}_{j,h}} \|z_j - \pi_h^\epsilon z_j\|_{0,\partial\tau_j}^2. \quad (4.63)$$

Applying the trace inequality (A:4) and using the interpolation estimates that the operator π_h is assumed to respect completes the proof for the fourth term

$$(IV) \leq C \lambda_\epsilon h^{-1} \sum_{j=1}^2 \sum_{\tau_j \in \mathcal{T}_{j,h}} \left(h_{\tau_j}^{-1/2} \|z_j - \pi_h^\epsilon z_j\|_{0,\tau_j} + h_{\tau_j}^{1/2} \|\nabla(z_j - \pi_h^\epsilon z_j)\|_{0,\tau_j} \right)^2 \quad (4.64)$$

$$\leq C \lambda_\epsilon h^{-1} \left(h^{3/2} |z|_{2,\Omega} \right)^2 \quad (4.65)$$

$$\leq C \lambda_\epsilon \max_j \left\{ \frac{\rho_j}{\mu_j} \right\} (\omega h)^2 \|\psi\|_{0,\Omega}^2. \quad (4.66)$$

The fifth term is bounded in a similar vein. Combining these results completes the claim. To prove the interpolation estimate for π_h^ϵ in the \mathfrak{J} semi-norm it is again easier to consider each term separately. Let $\eta^\epsilon = u - \pi_h^\epsilon u$

$$|\eta^\epsilon|_{\mathfrak{J}}^2 = \underbrace{(\lambda_\epsilon h^{-1}) \|\llbracket \eta^\epsilon \rrbracket\|_{0,\Gamma}^2}_{(A_{\mathfrak{J}})} + \underbrace{(\hat{\lambda}_\epsilon h) \|\llbracket \mu \nabla \eta^\epsilon \cdot \mathbf{n}_\Gamma \rrbracket\|_{\Gamma}}_{(B_{\mathfrak{J}})} + \underbrace{Im[s(\eta^\epsilon, \eta^\epsilon)]}_{(C_{\mathfrak{J}})} + \underbrace{\sum_{j=1}^2 \omega \sigma_j \|u^\epsilon\|_{0,\partial\Omega_j}^2}_{(D_{\mathfrak{J}})}.$$

The proof of the terms $(C_{\mathfrak{J}})$ and $(D_{\mathfrak{J}})$ are similar to the results seen previously so I shall not include them in the proof. The new terms are $(A_{\mathfrak{J}})$ and $(B_{\mathfrak{J}})$. The first term $(A_{\mathfrak{J}})$ can be shown to satisfy the estimate given in (4.44) by first applying the triangle inequality to expand the jump condition into two

separate terms on the trace of the elements intersected on the boundary

$$\begin{aligned}
(A_{\mathcal{T}}) &\leq (|\lambda_\epsilon| h^{-1}) \sum_{\tau \in \mathcal{T}_\Gamma} (\|\eta_1^\epsilon\|_{0,\partial\tau_\Gamma} + \|\eta_2^\epsilon\|_{0,\partial\tau_\Gamma})^2 \\
&\leq |\lambda_\epsilon| C_T^2 \sum_{\tau \in \mathcal{T}_\Gamma} (h_\tau^{-1} \|\eta^\epsilon\|_{0,\tau} + \|\nabla \eta^\epsilon\|_{0,\tau})^2 \\
&\leq \sum_{j=1}^2 |\lambda_\epsilon| C C_T^2 h^{2p} |u|_{p+1,\Omega_j}^2.
\end{aligned}$$

The second line comes from taking the trace inequality over all elements intersecting Γ . Finally, the estimate follows using the interpolation estimates that the operator π_h^ϵ is assumed to satisfy. The second term, $(B_{\mathcal{T}})$ is bounded in a similar way

$$\begin{aligned}
(B_{\mathcal{T}}) &\leq (|\hat{\lambda}_\epsilon| h) \sum_{\tau \in \mathcal{T}_\Gamma} (\|\mu_1 \nabla \eta_1^\epsilon \cdot \mathbf{n}_\Gamma\|_{0,\partial\tau_\Gamma} + \|\mu_2 \nabla \eta_2^\epsilon \cdot \mathbf{n}_\Gamma\|_{0,\partial\tau_\Gamma})^2 \\
&\leq |\hat{\lambda}_\epsilon| \max\{\mu_1, \mu_2\} C_T^2 \sum_{\tau \in \mathcal{T}_\Gamma} (\|\nabla \eta^\epsilon\|_{0,\tau} + h_\tau |\nabla \eta^\epsilon|_{1,\tau})^2 \\
&\leq \sum_{j=1}^2 |\hat{\lambda}_\epsilon| \max\{\mu_1, \mu_2\} C C_T h^{2p} |u|_{p+1,\Omega_j}^2.
\end{aligned}$$

Summing up the four terms and square rooting gives the required estimate. The final assumption (4.43) is a result about the interpolation properties of the operator π_h with respect to an appropriate norm. The proofs are similar for both the GLS/CIP case and the CIP case. With this in mind I shall prove the GLS/CIP case and just state the result for the CIP stabilization. Again for ease of notation I shall break the *-norm into components to be considered separately. Let $\eta \stackrel{\text{def}}{=} u - \pi_h^1 u$

$$\begin{aligned}
\|\eta^\epsilon\|^2 &= \sum_{j=1}^2 \left(\underbrace{\left[\sum_{\tau_j \in \mathcal{T}_{j,h}} (h_{\tau_j}^2 |\delta_{1,\tau_j}|)^{-1} \|\eta\|_{0,\tau_j}^2 + \sum_{F_j \in \mathcal{F}_{int}^j} (h_{\tau_j} |\gamma_{1,\tau_j}|)^{-1} \|\eta\|_{0,F_j}^2 + (\mu_j h |\beta_{1,j}|)^{-1} \|\eta\|_{0,\partial\Omega_j}^2 \right]}_{A_*} \right. \\
&\quad \left. + \underbrace{h |\lambda_1|^{-1} \|\{\mu \nabla \eta \cdot \mathbf{n}_\Gamma\}\|_{0,\Gamma}^2}_{B_*} + \underbrace{(h |\hat{\lambda}_1|)^{-1} \|\{\eta\}\|_{0,\Gamma}^2}_{C_*} + \underbrace{|\eta|_s^2}_{D_*} \right).
\end{aligned}$$

First I shall consider A_* . It is clear that the first term is bounded immediately by the interpolation results that π_h^ϵ is assumed to satisfy. The other two terms require an application of the trace inequality (A:4)

and then can be bounded in the same way as the first term.

$$\begin{aligned}
A_* &= \sum_{j=1}^2 \left[\sum_{\tau_j \in \mathcal{T}_{j,h}} (h_{\tau_j}^2 |\delta_{1,\tau_j}|)^{-1} \|\eta\|_{0,\tau_j}^2 + \sum_{F_j \in \mathcal{F}_{int}^j} (h_{\tau_j} |\gamma_{1,\tau_j}|)^{-1} \|\eta\|_{0,F_j}^2 + \mu_j (h |\beta_{1,j}|)^{-1} \|\eta\|_{0,\partial\Omega_j}^2 \right] \\
&\leq C \sum_{j=1}^2 \max_{\tau \in \mathcal{T}_h^j} \{|\delta_{1,\tau_j}|^{-1}\} h^{2p} |u|_{p+1,\Omega_j}^2 + \sum_{\tau_j \in \mathcal{T}_{j,h}} C_T^2 (|\gamma_{1,\tau_j}^{-1}| + |\mu_j \beta_{1,j}^{-1}|) \left(h_{\tau_j}^{-1} \|\eta\|_{0,\tau} + \|\nabla \eta\|_{0,\tau} \right)^2 \\
&\leq C \sum_{j=1}^2 h^{2p} |u|_{p+1,\Omega_j}^2.
\end{aligned}$$

The second component B_* involves an average which is handled in a similar way to the jump terms seen earlier. The average is split into two terms with an application of the triangle inequality. Then the trace inequality is used to enable the use of the interpolation estimates given in (4.38) and (4.39).

$$\begin{aligned}
B_* &\leq \frac{1}{4} h |\lambda_1|^{-1} \sum_{E \in \mathcal{G}_h^j} (\|\mu_1 \nabla \eta_1 \cdot n_\Gamma\|_{0,E} + \|\mu_2 \nabla \eta_2 \cdot n_\Gamma\|_{0,E})^2 \\
&\leq \frac{1}{4} |\lambda_1|^{-1} \max_j \mu_j^2 C_T^2 \sum_{\tau \in \mathcal{T}_\Gamma} (\|\nabla \eta\|_{0,\tau} + h_\tau |\nabla \eta|_{1,\tau})^2 \\
&\leq \sum_{j=1}^2 |\lambda_1|^{-1} C \max_j \mu_j^2 h^{2p} |u|_{p+1,\Omega_j}^2.
\end{aligned}$$

Once again care should be taken when considering the case of piecewise linear elements since $|\nabla \pi_h u|_{1,\tau} = 0$. This means that the final term must be bounded by noticing

$$|\nabla \eta|_{1,\tau} = |\nabla u|_{1,\tau} \leq C |u|_{2,\tau}.$$

The third component C_* is bounded in a similar way to B_* .

$$\begin{aligned}
C_* &\leq \frac{1}{4} h^{-1} |\hat{\lambda}_1|^{-1} \sum_{E \in \mathcal{G}_h^j} (\|\eta_1\|_{0,E} + \|\eta_2\|_{0,E})^2 \\
&\leq |\hat{\lambda}_1|^{-1} C_T^2 \sum_{\tau \in \mathcal{T}_\Gamma} (h_\tau^{-1} \|\eta\|_{0,\tau} + \|\nabla \eta\|_{0,\tau})^2 \\
&\leq \sum_{j=1}^2 |\hat{\lambda}_1|^{-1} C C_T^2 h^{2p} |u|_{p+1,\Omega_j}^2.
\end{aligned}$$

The bound on D_* follows as shown previously. It should be noted that the star norms are very similar for both cases; CIP and GLS/CIP and the proof for the CIP case is almost identical to that of the GLS/CIP case. So I just state the result

$$\|\eta\|_* \leq \sum_{j=1}^2 C C_T (|\gamma_2^{-1}| + |\beta_2^{-1}| + |\delta_2|^{-1} + |\gamma_0|^{-1} + |\gamma_1|^{-1} + C_s)^{\frac{1}{2}} h^p |u|_{p+1,\Omega_j},$$

where $\delta_2^{-1} = \max_{\tau_j \in \mathcal{T}_h^j} \{|\delta_{2,\tau_j}^{-1}|\}$, $\gamma_2^{-1} = \max_{\tau_j \in \mathcal{T}_h^j} \{|\gamma_{2,\tau_j}^{-1}|\}$, $\beta_2^{-1} = \max\{|\beta_{21}^{-1}|, |\beta_{22}^{-1}|\}$ and C_s is the constant given from D_* \square

Now that I have shown the stabilization fits into the theoretical framework of the theorem and also that the operator π_h has the correct interpolation properties all I must now show is that the forms $A_h(\cdot, \cdot)$ and $A(\cdot, \cdot)$ have the required continuity properties.

Proposition 6 (Continuity). *Let $u \in H^{p+1}(\Omega_1 \cup \Omega_2)$, for $p > 1/2$, and $z \in H^2(\Omega_1 \cup \Omega_2)$, be the exact solutions to the primal and dual problems defined by (4.4) and (4.5) respectively. It holds that the domain decomposition method given by (4.4) satisfies assumptions (4.40) and (4.41) of Theorem 3 for π_h^ϵ as defined in Lemma 4.4 and $\|\cdot\|_* \stackrel{\text{def}}{=} \epsilon\|\cdot\| + (1-\epsilon)\|\cdot\|_*$ as given in (4.55) and (4.54) for $\epsilon \in \{0, 1\}$*

Proof. The proof of Assumption (4.40) is possibly the most technical but follows the same ideas as shown previously. A couple of manipulations are required to deal with the terms acting on the interface. I begin once again by considering the first term of $A_h(u - \pi_h u, v_h)$ element wise and performing an integration by parts.

Let $\eta^\epsilon = u - \pi_h^\epsilon u$,

$$\begin{aligned} \sum_{j=1}^2 (\mu_j \nabla \eta_j^\epsilon, \nabla v_j)_{\Omega_j} &= \sum_{j=1}^2 \sum_{\tau_j \in \mathcal{T}_j} (\nabla^\epsilon \eta_j, \mu_j \nabla v_j)_{0,\tau_j} \\ &= \sum_{j=1}^2 \sum_{\tau_j \in \mathcal{T}_{j,h}} (\eta_j^\epsilon, -\nabla \cdot (\mu_j \nabla v_j))_{0,\tau_j} + \langle \eta_j^\epsilon, \mu_j \nabla v_j \cdot n \rangle_{0,\partial\tau_j} \\ &= \sum_{j=1}^2 \left(\sum_{\tau_j \in \mathcal{T}_j} (\eta_j^\epsilon, -\nabla \cdot (\mu_j \nabla v_j))_{0,\tau_j} + \sum_{F_j \in \mathcal{F}_{int}^j} \langle \eta_j^\epsilon, \llbracket \mu_j \nabla v_j \cdot n \rrbracket \rangle_{0,F_j} \right. \\ &\quad \left. + \langle \mu_j \eta_j^\epsilon, \nabla v_j \cdot n \rangle_{0,\partial\Omega_j} \right) + \langle \eta_1^\epsilon, \mu_1 \nabla v_1 \cdot n \rangle_{0,\Gamma} + \langle \eta_2^\epsilon, \mu_2 \nabla v_2 \cdot n \rangle_{0,\Gamma}. \end{aligned}$$

This gives rise to the usual terms on each subdomain with a couple of additional terms acting on the interface. These last two terms allow us to deal with the fifth term in $A_h(\eta^\epsilon, v_h)$. This statement becomes more obvious after a closer inspection of the fifth term in $A_h(\eta^\epsilon, v_h)$

$$\begin{aligned} -\langle \llbracket \eta^\epsilon \rrbracket, \{\mu \nabla v_h \cdot n_\Gamma\} \rangle_{0,\Gamma} &= -\frac{1}{2} \langle \eta_1^\epsilon, \mu_1 \nabla v_1 \cdot n_\Gamma \rangle_{0,\Gamma} - \frac{1}{2} \langle \eta_1^\epsilon, \mu_2 \nabla v_2 \cdot n_\Gamma \rangle_{0,\Gamma} \\ &\quad + \frac{1}{2} \langle \eta_2^\epsilon, \mu_1 \nabla v_1 \cdot n_\Gamma \rangle_{0,\Gamma} + \frac{1}{2} \langle \eta_2^\epsilon, \mu_2 \nabla v_2 \cdot n_\Gamma \rangle_{0,\Gamma}. \end{aligned}$$

Adding the two terms from our integration by parts and simplifying gives

$$\begin{aligned}
&\implies -\frac{1}{2} \langle \eta_1^\epsilon, \mu_1 \nabla v_1 \cdot n_\Gamma \rangle_{0,\Gamma} - \frac{1}{2} \langle \eta_1^\epsilon, \mu_2 \nabla v_2 \cdot n_\Gamma \rangle_{0,\Gamma} + \frac{1}{2} \langle \eta_2^\epsilon, \mu_1 \nabla v_1 \cdot n_\Gamma \rangle_{0,\Gamma} \\
&\quad + \frac{1}{2} \langle \eta_2^\epsilon, \mu_2 \nabla v_2 \cdot n_\Gamma \rangle_{0,\Gamma} + \langle \eta_1^\epsilon, \mu_1 \nabla v_1 \cdot n_\Gamma \rangle_{0,\Gamma} - \langle \eta_2^\epsilon, \mu_2 \nabla v_2 \cdot n_\Gamma \rangle_{0,\Gamma} \\
&= -\frac{1}{2} \langle \eta_1^\epsilon, \mu_2 \nabla v_2 \cdot n_\Gamma \rangle_{0,\Gamma} + \frac{1}{2} \langle \eta_2^\epsilon, \mu_1 \nabla v_1 \cdot n_\Gamma \rangle_{0,\Gamma} + \frac{1}{2} \langle \eta_1^\epsilon, \mu_1 \nabla v_1 \cdot n_\Gamma \rangle_{0,\Gamma} \\
&\quad - \frac{1}{2} \langle \eta_2^\epsilon, \mu_2 \nabla v_2 \cdot n_\Gamma \rangle_{0,\Gamma} \\
&= \frac{1}{2} \langle \eta_1^\epsilon, \llbracket \mu \nabla v \cdot n_\Gamma \rrbracket \rangle_{0,\Gamma} + \frac{1}{2} \langle \eta_2^\epsilon, \llbracket \mu \nabla v \cdot n_\Gamma \rrbracket \rangle_{0,\Gamma} \\
&= \langle \{\eta^\epsilon\}, \llbracket \mu \nabla v \cdot n_\Gamma \rrbracket \rangle_{0,\Gamma}.
\end{aligned}$$

It is this that motivates the addition of the non-standard penalty term in our stabilization. From this the sesquilinear form reduces to

$$\begin{aligned}
|A_h(\eta^\epsilon, v_h)| &= \left| \sum_{j=1}^2 \left[\sum_{\tau_j \in \mathcal{T}_j} [(\eta_j^\epsilon, -\nabla \cdot (\mu_j \nabla v_j))_{0,\tau_j} - (\eta_j^\epsilon, \omega^2 \rho_j v_j)_{0,\tau_j}] \right. \right. \\
&\quad \left. \left. + \sum_{F_j \in \mathcal{F}_{int}^j} \langle \eta_j^\epsilon, \llbracket \mu \nabla v_j \cdot n \rrbracket \rangle_{0,F_j} + \langle \mu_j \eta_j^\epsilon, \nabla v_j \cdot n + i\omega \sigma_j v_j \rangle_{0,\partial\Omega_j} \right] \right. \\
&\quad \left. - \langle \{\mu \nabla \eta^\epsilon \cdot n_\Gamma\}, \llbracket v_h \rrbracket \rangle_{0,\Gamma} + \langle \{\eta^\epsilon\}, \llbracket \mu \nabla v_h \cdot n_\Gamma \rrbracket \rangle_{0,\Gamma} + S(\eta^\epsilon, v_h) \right|.
\end{aligned}$$

The proof now follows in a similar vein as in the previous chapter and depends on which stabilization we choose. If $\epsilon = 1$ the stabilization on the interior of each sub-domain is equivalent to the GLS/CIP stabilization introduced earlier. Recalling that $\pi_h^1 : C^0(\bar{\Omega}) \mapsto V_h$ is the Lagrange interpolant and applying the Cauchy-Schwarz inequality it follows that

$$|A_h(\eta^1, v_h)| \leq M \|\eta^1\| \|v_h\|_{\mathcal{J}} \quad (4.67)$$

$$= M \|\eta^1\|_* \|v_h\|_{\mathcal{J}}. \quad (4.68)$$

Similarly recall that for $\epsilon = 0$ the stabilization on the interior of each sub-domain reduces to the CIP stabilization and $\pi_h^0 : L^2(\Omega) \mapsto V_h$ is the L^2 -projection. This eliminates the low-order terms and allows us to further exploit the orthogonality property to bound the Laplacian by the second order jumps in our stabilization. The only difference from before is that this argument must be used on both domains. Using the aforementioned technique and an application of the Cauchy-Schwarz inequality gives

$$|A_h(\eta^0, v_h)| \leq M \|\eta^0\|_* \|v_h\|_{\mathcal{J}} \quad (4.69)$$

$$= M \|\eta^0\|_* \|v_h\|_{\mathcal{J}}, \quad (4.70)$$

which completes the claim.

Assumption (4.41) follows using the same arguments as before on each subdomain and using

$$\|v\|_{0,\Omega_1 \cup \Omega_2}^2 = \|v\|_{0,\Omega_1}^2 + \|v\|_{0,\Omega_2}^2.$$

The interface terms are the only terms that have not been shown to satisfy the assumption, namely

$$\begin{aligned} & \underbrace{\left| \langle \{\mu \nabla \eta^\epsilon \cdot n_\Gamma\}, \llbracket \zeta^\epsilon \rrbracket \rangle_{0,\Gamma} \right|}_{(I)} + \underbrace{\left| \langle \llbracket \eta^\epsilon \rrbracket, \{\mu \nabla \zeta^\epsilon \cdot n_\Gamma\} \rangle_{0,\Gamma} \right|}_{(II)} + \underbrace{\left| \langle \lambda_\epsilon h^{-1} \llbracket \eta^\epsilon \rrbracket, \llbracket \zeta^\epsilon \rrbracket \rangle_{0,\Gamma} \right|}_{(III)} \\ & + \underbrace{\left| \langle \hat{\lambda}_\epsilon h \llbracket \mu \nabla \eta^\epsilon \cdot n_\Gamma \rrbracket, \llbracket \mu \nabla \zeta^\epsilon \cdot n_\Gamma \rrbracket \rangle_{0,\Gamma} \right|}_{(IV)}, \end{aligned} \quad (4.71)$$

where, as previously, $\eta^\epsilon = u - \pi_h^\epsilon u$ and $\zeta^\epsilon = z - \pi_h^\epsilon z$.

I begin by expanding the jump and average terms of (I). Using the fact that the mesh fits the intersection Γ exactly, I can consider each term element wise. In the last line I use the triangle inequality to get an upper bound on (I)

$$\begin{aligned} (I) &= \left| \frac{1}{2} \langle \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \zeta_1^\epsilon \rangle_{0,\Gamma} - \frac{1}{2} \langle \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \zeta_2^\epsilon \rangle_{0,\Gamma} + \frac{1}{2} \langle \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \zeta_1^\epsilon \rangle_{0,\Gamma} \right. \\ &\quad \left. - \frac{1}{2} \langle \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \zeta_2^\epsilon \rangle_{0,\Gamma} \right| \\ &= \left| \frac{1}{2} \sum_{E \in \mathcal{G}_h^j} \langle \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \zeta_1^\epsilon \rangle_{0,E} - \langle \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \zeta_2^\epsilon \rangle_{0,E} + \langle \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \zeta_1^\epsilon \rangle_{0,E} \right. \\ &\quad \left. - \langle \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \zeta_2^\epsilon \rangle_{0,E} \right| \\ &\leq \frac{1}{2} \sum_{E \in \mathcal{G}_h} \underbrace{\left| \langle \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \zeta_1^\epsilon \rangle_{0,E} \right| + \left| \langle \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \zeta_2^\epsilon \rangle_{0,E} \right|}_{(a)} \\ &\quad + \underbrace{\left| \langle \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \zeta_1^\epsilon \rangle_{0,E} \right| + \left| \langle \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \zeta_2^\epsilon \rangle_{0,E} \right|}_{(b)}. \end{aligned}$$

It is then possible to use the Cauchy-Schwarz inequality and Trace inequality to bound each term and get

$$\begin{aligned} (a) &\leq \frac{1}{2} \sum_{\tau_j \in \mathcal{T}_\Gamma^j} C_T^2 \left(h_{\tau_1}^{-\frac{1}{2}} \|\mu_1 \nabla \eta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\mu_1 \nabla \eta_1^\epsilon|_{1,\tau_1} \right) \left(h_{\tau_1}^{-\frac{1}{2}} \|\zeta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\zeta_1^\epsilon|_{1,\tau_1} \right) \\ &\quad + C_T^2 \left(h_{\tau_2}^{-\frac{1}{2}} \|\mu_2 \nabla \eta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\mu_2 \nabla \eta_2^\epsilon|_{1,\tau_2} \right) \left(h_{\tau_2}^{-\frac{1}{2}} \|\zeta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\zeta_2^\epsilon|_{1,\tau_2} \right) \\ &\leq C \left(\max_j \{\mu_j\} \right) \max_j \left\{ \sqrt{\frac{\omega^2 \rho_j}{\mu_j}} \right\} h^{p+1} \sum_{j=1}^2 |u|_{p+1,\Omega_j} \|\psi\|_{0,\Omega}, \end{aligned}$$

and

$$\begin{aligned}
(b) &\leq \frac{1}{2} \sum_{\tau_j \in \mathcal{T}_\Gamma^j} C_T^2 \left(h_{\tau_2}^{-\frac{1}{2}} \|\mu_2 \nabla \eta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\mu_2 \nabla \eta_2^\epsilon|_{1,\tau_2} \right) \left(h_{\tau_1}^{-\frac{1}{2}} \|\zeta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\zeta_1^\epsilon|_{1,\tau_1} \right) \\
&\quad + C_T^2 \left(h_{\tau_1}^{-\frac{1}{2}} \|\mu_1 \nabla \eta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\mu_1 \nabla \eta_1^\epsilon|_{1,\tau_1} \right) \left(h_{\tau_2}^{-\frac{1}{2}} \|\zeta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\zeta_2^\epsilon|_{1,\tau_2} \right) \\
&\leq C \left(\max_j \{\mu_j\} \right) \max_j \left\{ \sqrt{\frac{\omega^2 \rho_j}{\mu_j}} \right\} h^{p+1} \sum_{j=1}^2 |u|_{p+1,\Omega_j} \|\psi\|_{0,\Omega},
\end{aligned}$$

where the final inequalities come from the interpolation estimates for both the Lagrange interpolant and L^2 -projections given by (4.38) or (4.39). It should be noted that once again in the piecewise linear case second order derivatives of elements of V_h are zero. However, it is possible to obtain the bound using the fact that u and z are in $H^2(\Omega_1 \cup \Omega_2)$. A similar argument holds for the (II),

$$\begin{aligned}
(II) &= \left| \frac{1}{2} \langle \eta_1^\epsilon, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \rangle_\Gamma + \frac{1}{2} \langle \eta_1^\epsilon, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \rangle_\Gamma - \frac{1}{2} \langle \eta_2^\epsilon, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \rangle_\Gamma \right. \\
&\quad \left. - \frac{1}{2} \langle \eta_2^\epsilon, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \rangle_\Gamma \right| \\
&= \left| \frac{1}{2} \sum_{E \in \mathcal{G}_h^j} \langle \eta_1^\epsilon, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \rangle_{0,E} + \langle \eta_1^\epsilon, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \rangle_{0,E} - \langle \eta_2^\epsilon, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \rangle_{0,E} \right. \\
&\quad \left. - \langle \eta_2^\epsilon, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \rangle_{0,E} \right| \\
&\leq \frac{1}{2} \sum_{E \in \mathcal{G}_h^j} \underbrace{|\langle \eta_1, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \rangle_{0,E}| + |\langle \eta_1, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \rangle_{0,E}|}_{(c)} \\
&\quad + \underbrace{|\langle \eta_2, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \rangle_{0,E}| + |\langle \eta_2, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \rangle_{0,E}|}_{(d)}.
\end{aligned}$$

Again applying the Cauchy-Schwarz and Trace Inequalities as well as our interpolation estimates simplifies to

$$\begin{aligned}
(c) &\leq \frac{1}{2} \sum_{\tau_j \in \mathcal{T}_\Gamma^j} C_T^2 \left(h_{\tau_1}^{-\frac{1}{2}} \|\eta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\eta_1^\epsilon|_{1,\tau_1} \right) \left(h_{\tau_1}^{-\frac{1}{2}} \|\mu_1 \nabla \zeta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\mu_1 \nabla \zeta_1^\epsilon|_{1,\tau_1} \right) \\
&\quad + C_T^2 \left(h_{\tau_1}^{-\frac{1}{2}} \|\eta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\eta_1^\epsilon|_{1,\tau_1} \right) \left(h_{\tau_2}^{-\frac{1}{2}} \|\mu_2 \nabla \zeta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\mu_2 \nabla \zeta_2^\epsilon|_{1,\tau_2} \right) \\
&\leq C \left(\max_j \{\mu_j\} \right) \max_j \left\{ \sqrt{\frac{\omega^2 \rho_j}{\mu_j}} \right\} h^{p+1} \sum_{j=1}^2 |u|_{p+1,\Omega_j} \|\psi\|_{0,\Omega},
\end{aligned}$$

and

$$\begin{aligned}
(d) &\leq \frac{1}{2} \sum_{\tau_j \in \mathcal{T}_\Gamma^j} C_T^2 \left(h_{\tau_2}^{-\frac{1}{2}} \|\eta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\eta_2^\epsilon|_{1,\tau_2} \right) \left(h_{\tau_1}^{-\frac{1}{2}} \|\mu_1 \nabla \zeta_1^\epsilon\|_{0,\tau_1} + h_{\tau_1}^{\frac{1}{2}} |\mu_1 \nabla \zeta_1^\epsilon|_{1,\tau_1} \right) \\
&\quad + C_T^2 \left(h_{\tau_2}^{-\frac{1}{2}} \|\eta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\eta_2^\epsilon|_{1,\tau_2} \right) \left(h_{\tau_2}^{-\frac{1}{2}} \|\mu_2 \nabla \zeta_2^\epsilon\|_{0,\tau_2} + h_{\tau_2}^{\frac{1}{2}} |\mu_2 \nabla \zeta_2^\epsilon|_{1,\tau_2} \right) \\
&\leq C \left(\max_j \{\mu_j\} \right) \max_j \left\{ \sqrt{\frac{\omega^2 \rho_j}{\mu_j}} \right\} h^{p+1} \sum_{j=1}^2 |u|_{p+1,\Omega_j} \|\psi\|_{0,\Omega}.
\end{aligned}$$

All that needs to be shown now is that the additional terms in the stabilization, the interface terms, are bounded in a similar way. Once again I begin by splitting the jump terms and considering them element-wise. The final line comes from an application of the triangle inequality followed by Cauchy-Schwarz.

$$\begin{aligned}
(III) &= | \langle \lambda_\epsilon h^{-1} \eta_1^\epsilon, \zeta_1^\epsilon \rangle_{0,\Gamma} - \langle \lambda_\epsilon h^{-1} \eta_1^\epsilon, \zeta_2^\epsilon \rangle_{0,\Gamma} - \langle \lambda_\epsilon h^{-1} \eta_2^\epsilon, \zeta_1^\epsilon \rangle_{0,\Gamma} + \langle \lambda_\epsilon h^{-1} \eta_2^\epsilon, \zeta_2^\epsilon \rangle_{0,\Gamma} | \\
&= \left| \sum_{E \in \mathcal{G}_h^j} \langle \lambda_\epsilon h^{-1} \eta_1^\epsilon, \zeta_1^\epsilon \rangle_{0,E} - \langle \lambda_\epsilon h^{-1} \eta_1^\epsilon, \zeta_2^\epsilon \rangle_{0,E} - \langle \lambda_\epsilon h^{-1} \eta_2^\epsilon, \zeta_1^\epsilon \rangle_{0,E} + \langle \lambda_\epsilon h^{-1} \eta_2^\epsilon, \zeta_2^\epsilon \rangle_{0,E} \right| \\
&\leq \sum_{E \in \mathcal{G}_h^j} |\lambda_\epsilon h_\tau^{-1}| \left(\underbrace{\|\eta_1^\epsilon\|_{0,E} \|\zeta_1^\epsilon\|_{0,E} + \|\eta_1^\epsilon\|_{0,E} \|\zeta_2^\epsilon\|_{0,E}}_{(A_s)} \right. \\
&\quad \left. + \underbrace{\|\eta_2^\epsilon\|_{0,E} \|\zeta_1^\epsilon\|_{0,E} + \|\eta_2^\epsilon\|_{0,E} \|\zeta_2^\epsilon\|_{0,E}}_{(B_s)} \right).
\end{aligned}$$

The terms A_s and B_s are bounded using an application of the trace inequality followed by the interpolation estimates for the operator π_h .

$$\begin{aligned}
(A_s) &\leq \sum_{\tau \in \mathcal{T}_\Gamma} \left(C_T^2 \left(h_\tau^{-1/2} \|\eta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \eta_1^\epsilon\|_{0,\tau} \right) \left(h_\tau^{-1/2} \|\zeta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \zeta_1^\epsilon\|_{0,\tau} \right) \right. \\
&\quad \left. + C_T^2 \left(h_\tau^{-1/2} \|\eta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \eta_1^\epsilon\|_{0,\tau} \right) \left(h_\tau^{-1/2} \|\zeta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \zeta_2^\epsilon\|_{0,\tau} \right) \right) \\
&\leq C \max_j \left\{ \sqrt{\frac{\omega^2 \rho_j}{\mu_j}} \right\} h^{p+2} \sum_{j=1}^2 |u|_{p+1,\Omega_j} \|\psi\|_{0,\Omega}.
\end{aligned}$$

$$\begin{aligned}
(B_s) &\leq \sum_{\tau \in \mathcal{T}_\Gamma} \left(C_T^2 \left(h_\tau^{-1/2} \|\eta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \eta_2^\epsilon\|_{0,\tau} \right) \left(h_\tau^{-1/2} \|\zeta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \zeta_1^\epsilon\|_{0,\tau} \right) \right. \\
&\quad \left. + C_T^2 \left(h_\tau^{-1/2} \|\eta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \eta_2^\epsilon\|_{0,\tau} \right) \left(h_\tau^{-1/2} \|\zeta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} \|\nabla \zeta_2^\epsilon\|_{0,\tau} \right) \right) \\
&\leq C \max_j \left\{ \sqrt{\frac{\omega^2 \rho_j}{\mu_j}} \right\} h^{p+2} \sum_{j=1}^2 |u|_{p+1,\Omega_j} \|\psi\|_{0,\Omega}.
\end{aligned}$$

The fourth and final term on the interface is the Neumann penalty. Once again I decompose the jump

terms and apply the triangle and Cauchy-Schwarz inequalities.

$$\begin{aligned}
(IV) &= \left| \left\langle \hat{\lambda}_\epsilon h \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \right\rangle_{0,\Gamma} - \left\langle \hat{\lambda}_\epsilon h \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \right\rangle_{0,\Gamma} \right. \\
&\quad \left. - \left\langle \hat{\lambda}_\epsilon h \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \right\rangle_{0,\Gamma} + \left\langle \hat{\lambda}_\epsilon h \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \right\rangle_{0,\Gamma} \right| \\
&= \left| \sum_{E \in \mathcal{G}_h^j} \left\langle \hat{\lambda}_\epsilon h \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \right\rangle_{0,E} - \left\langle \hat{\lambda}_\epsilon h \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \right\rangle_{0,E} \right. \\
&\quad \left. - \left\langle \hat{\lambda}_\epsilon h \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \right\rangle_{0,E} + \left\langle \hat{\lambda}_\epsilon h \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma, \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \right\rangle_{0,E} \right| \\
&\leq \sum_{E \in \mathcal{G}_h^j} |\hat{\lambda}_\epsilon h_\tau| \left(\underbrace{\left\| \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma \right\|_{0,E} \left\| \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \right\|_{0,E} + \left\| \mu_1 \nabla \eta_1^\epsilon \cdot n_\Gamma \right\|_{0,E} \left\| \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \right\|_{0,E}}_{(C_s)} \right. \\
&\quad \left. + \underbrace{\left\| \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma \right\|_{0,E} \left\| \mu_1 \nabla \zeta_1^\epsilon \cdot n_\Gamma \right\|_{0,E} + \left\| \mu_2 \nabla \eta_2^\epsilon \cdot n_\Gamma \right\|_{0,E} \left\| \mu_2 \nabla \zeta_2^\epsilon \cdot n_\Gamma \right\|_{0,E}}_{(D_s)} \right).
\end{aligned}$$

The resulting term can then be bounded using the trace inequality followed by an application of the interpolation estimate (4.39) for elements of polynomial order $p > 1$ or using the regularity of the solution u for $p = 1$.

$$\begin{aligned}
(C_s) &\leq C \max\{\mu_1, \mu_2\} \mu_1 \sum_{\tau \in \mathcal{T}_\Gamma} C_T^2 \left(h_\tau^{-1/2} \|\nabla \eta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \eta_1^\epsilon|_{1,\tau} \right) \left(h_\tau^{-1/2} \|\nabla \zeta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \zeta_1^\epsilon|_{1,\tau} \right) \\
&\quad + C_T^2 \left(h_\tau^{-1/2} \|\nabla \eta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \eta_1^\epsilon|_{1,\tau} \right) \left(h_\tau^{-1/2} \|\nabla \zeta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \zeta_2^\epsilon|_{1,\tau} \right) \\
&\leq CC_T^2 (\max\{\mu_1, \mu_2\} \mu_1) (\max_j \sqrt{\frac{\omega^2 \rho_j}{\mu_j}}) h^p \sum_{j=1}^2 |u|_{p+1, \Omega_j} \|\psi\|_{0,\Omega}.
\end{aligned}$$

$$\begin{aligned}
(D_s) &\leq C \max\{\mu_1, \mu_2\} \mu_2 \sum_{\tau \in \mathcal{T}_\Gamma} C_T^2 \left(h_\tau^{-1/2} \|\nabla \eta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \eta_2^\epsilon|_{1,\tau} \right) \left(h_\tau^{-1/2} \|\nabla \zeta_1^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \zeta_1^\epsilon|_{1,\tau} \right) \\
&\quad + C_T^2 \left(h_\tau^{-1/2} \|\nabla \eta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \eta_2^\epsilon|_{1,\tau} \right) \left(h_\tau^{-1/2} \|\nabla \zeta_2^\epsilon\|_{0,\tau} + h_\tau^{1/2} |\nabla \zeta_2^\epsilon|_{1,\tau} \right) \\
&\leq CC_T^2 (\max\{\mu_1, \mu_2\} \mu_2) (\max_j \sqrt{\frac{\omega^2 \rho_j}{\mu_j}}) h^p \sum_{j=1}^2 |u|_{p+1, \Omega_j} \|\psi\|_{0,\Omega}.
\end{aligned}$$

Summing all of these results concludes the proof of the proposition. \square

4.3.6 Discussion

Since all of the assumptions of Theorem 3 hold the domain decomposition method introduced is stable under the condition $\min_j \{\beta_{\epsilon,j}^{-1}\} > h \frac{\omega \sigma_j}{\mu_j^2}$ and respects the a priori error estimates given. This formulation enables the user to couple multiple Helmholtz systems under the condition $\max_j \left\{ \sqrt{\frac{\omega^2 \rho_j}{\mu_j}} \right\} h \leq C$ which has benefits in several areas of Physics and Engineering. One limitation of the method is the fact that the

user must know a priori where the interface is and design a mesh to track it. This provided the motivation to design a fictitious domain method that does not require the mesh to fit the interface. The purpose of the domain decomposition method presented here is to provide a platform to study the efficiency of the Nitsche coupling technique when solving Helmholtz. Since the stabilized methods presented have been able to reduce the pollution effect for certain problems it is interesting to study whether the introduction of the coupling terms introduces numerical pollution. The numerical evidence would suggest that the coupling terms do not introduce a substantial amount of pollution and seem to highlight the need for a method that reduces pollution in the bulk. All computations in this section were performed using the UMFPAK solver in the FreeFEM++ package which is available from <http://www.freefem.org/>.

4.3.7 Numerical Results

Let Ω_1 be $[0, 1] \times [0, 0.5]$ and Ω_2 be $[0, 1] \times [0.5, 1]$, the model problem is defined as

$$\left. \begin{aligned} -\nabla \cdot (\nabla u) - \omega^2 \rho_1 u &= f_1 && \text{in } \Omega_1 \\ -\nabla \cdot (\nabla u) - \omega^2 \rho_2 u &= f_1 + (\rho_1 \omega^2 - \rho_2 \omega^2) u && \text{in } \Omega_2 \\ u_1 &= g_{1,D} && \text{for } x \in [0, 1], y = 0 \\ u_1 &= g_{1,D} && \text{for } x = 0, y \in [0, 0.5] \\ u_2 &= g_{2,D} && \text{for } x = 0, y \in [0.5, 1] \\ \nabla u \cdot n + i\omega \sqrt{\rho_1} u &= g_{1,R} && \text{for } x = 1, y \in [0, 0.5] \\ \nabla u \cdot n + i\omega \sqrt{\rho_2} u &= g_{2,R} && \text{for } x = 1, y \in [0.5, 1] \\ \nabla u \cdot n + i\omega \sqrt{\rho_2} u &= g_{2,R} && \text{for } x \in [0, 1], y = 1 \\ u|_1 &= u|_2 && \text{on } \Gamma \\ \nabla u|_1 \cdot n_\Gamma &= \nabla u|_2 \cdot n_\Gamma && \text{on } \Gamma, \end{aligned} \right\} \quad (4.72)$$

Taking $\rho_j \omega^2 = k_j^2$ the problem is designed such that

$$u = \frac{\cos(k_1 \sqrt{x^2 + y^2})}{k_1} - C J_0(k_1 \sqrt{x^2 + y^2})$$

where

$$C = \frac{\cos(k_1) + i \sin(k_1)}{k_1 (J_0(k_1) + i J_1(k_1))}$$

and $J_0(\cdot)$ and $J_1(\cdot)$ are Bessel functions of the First Kind. For the stabilized methods proposed in (4.25) the calculations are performed using the same choice of parameters as taken in Chapter 2. The Nitsche coupling parameters are taken to be $\alpha_0 = \alpha_1 = 1$ which numerical evidence seems to suggest is a good choice. In order to present some comparison I also consider the case where all of the stabilization parameters acting on the bulk of each domain are taken to be 0 with $\alpha_0 = \alpha_1 = 1$, I refer to this case as the ‘no stabilization case’ or just ‘Nitsche’s’ case.

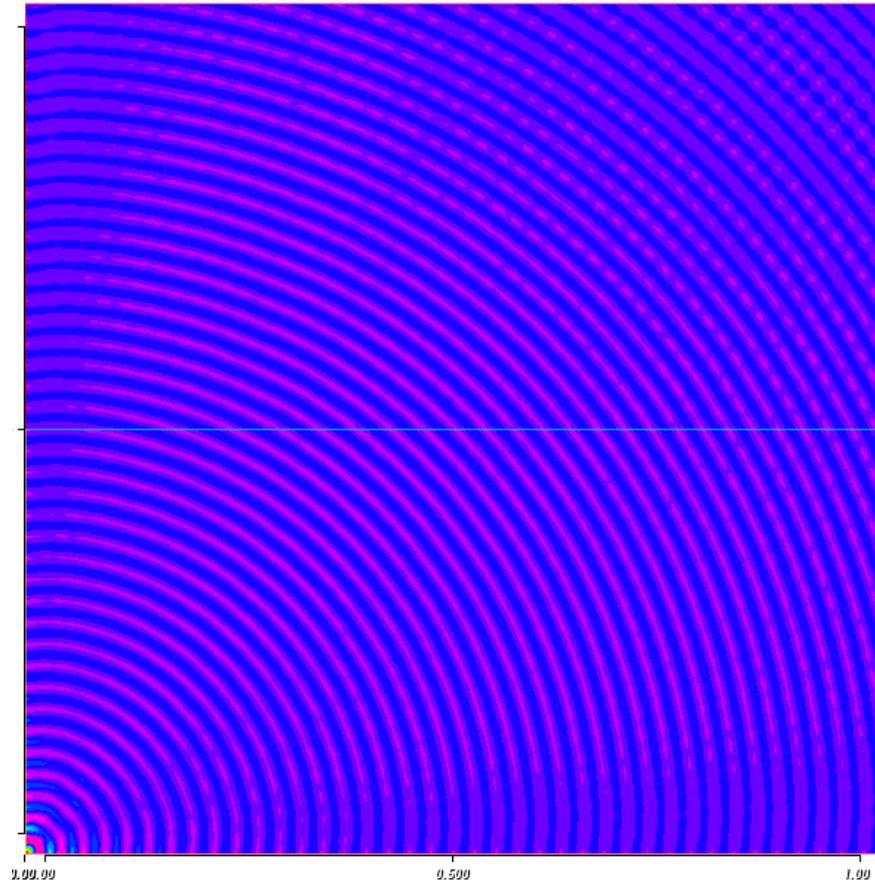
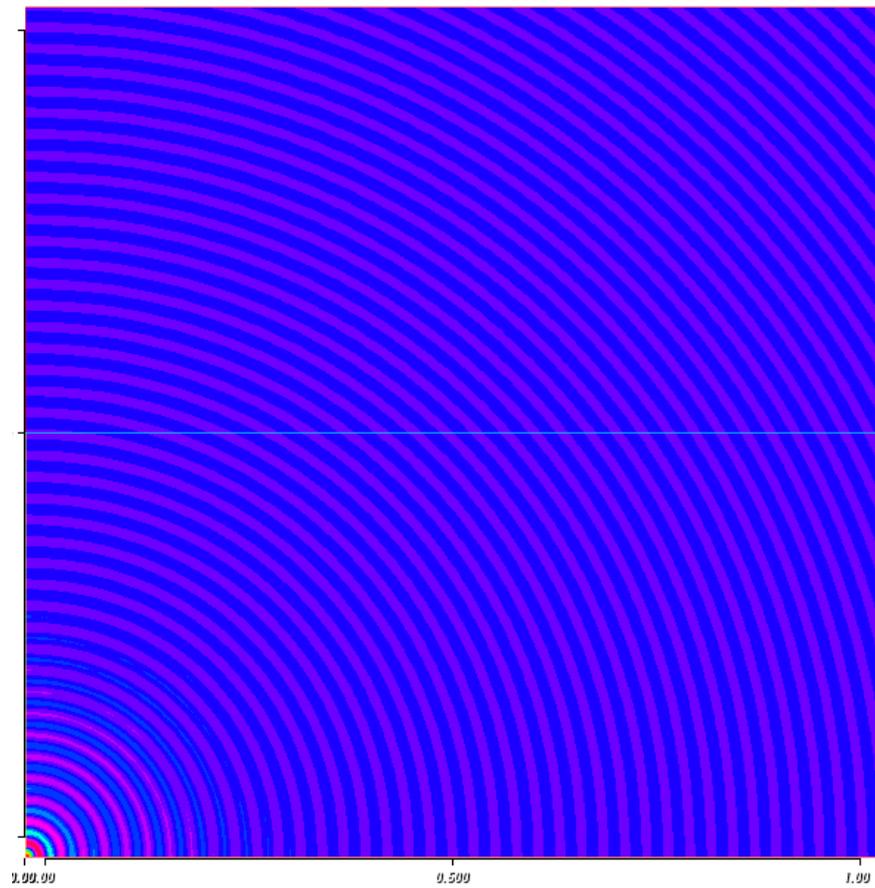
(a) Standard Galerkin solution $k_1 = k_2 = 250$ (b) GLS/CIP solution $k_1 = k_2 = 250$

Figure 4.1: Solution plots for domain decomposition problem with matching meshes

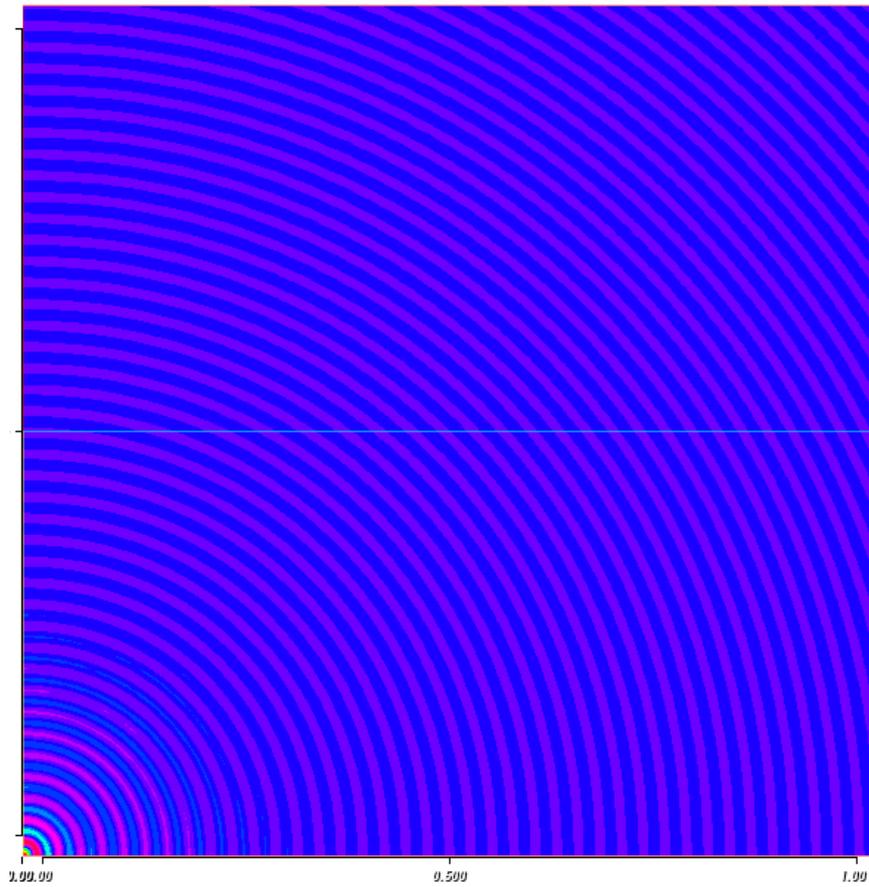
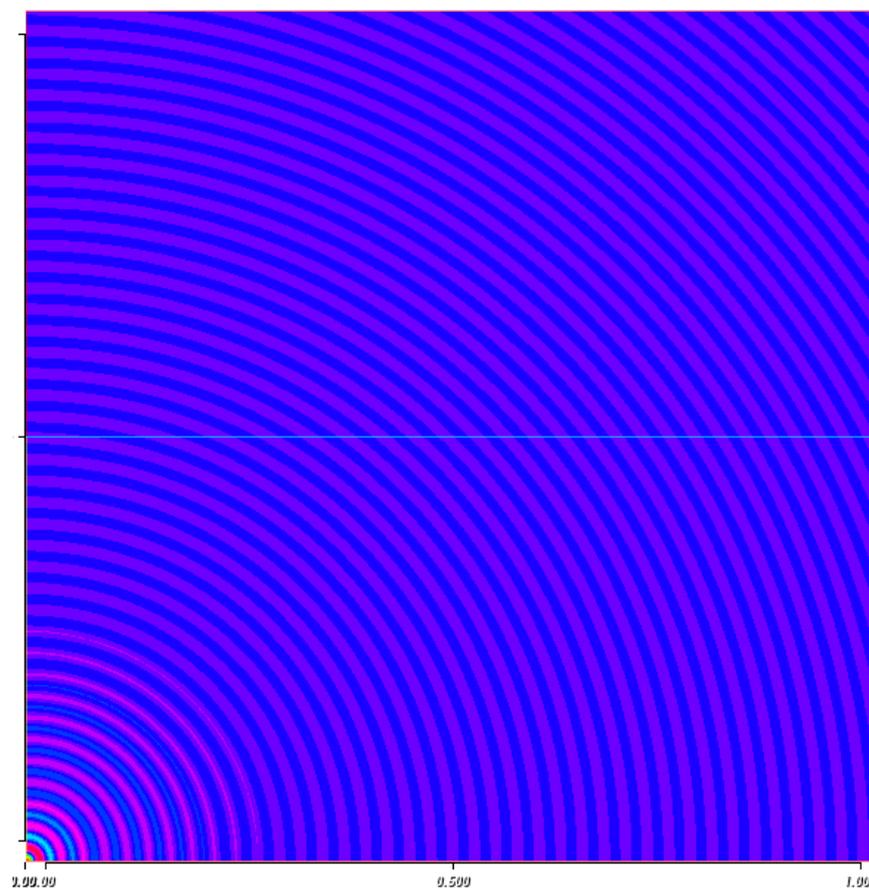
(a) CIP solution $k_1 = k_2 = 250$ (b) Exact solution $k_1 = k_2 = 250$

Figure 4.2: Solution plots for domain decomposition problem with matching meshes

Figures 4.1 and 4.2 show the numerical solution for the domain decomposition problem with matching meshes. The stabilized methods appear to conserve the correct wave number whilst the standard Galerkin solution becomes out of phase with the exact solution compare the phase of the solutions in the upper right corner.

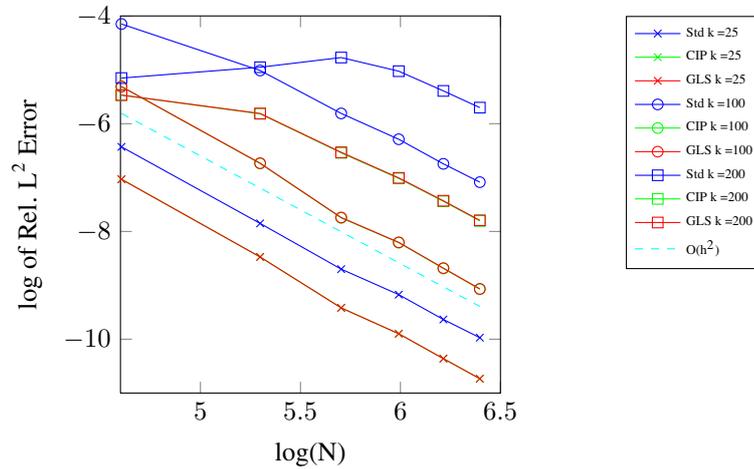


Figure 4.3: Convergence comparison for domain decomposition using piecewise linear elements with non-matching mesh

Figure 4.3 shows that the domain decomposition methods on non matching meshes achieve $O(h^2)$ convergence for piecewise linear elements which supports the theory in the previous section. The stabilized methods are almost indistinguishable in the figure which would appear to suggest that the GLS term is perhaps not that important in the domain decomposition case. Figure 4.4 gives an example of a non-matching mesh.

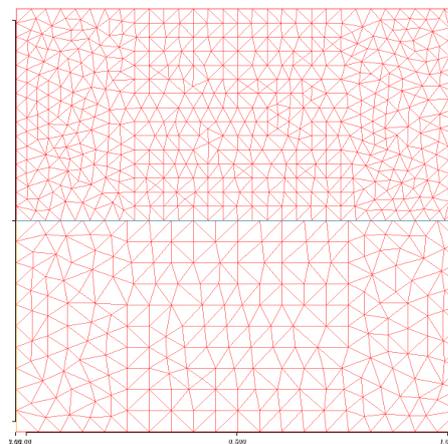


Figure 4.4: Example of a non-matching mesh

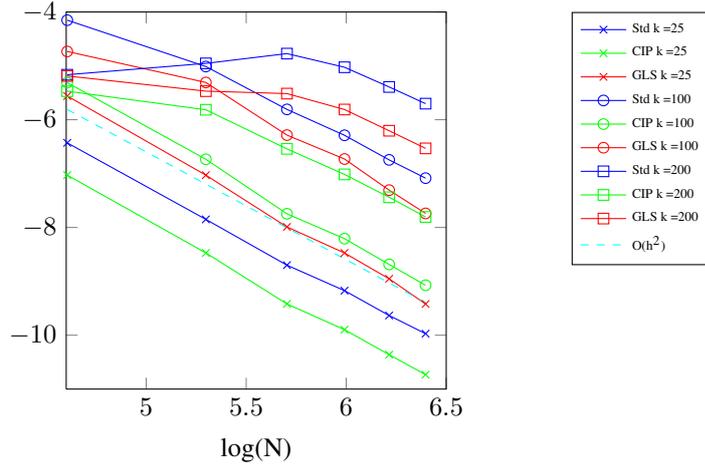


Figure 4.5: Convergence comparison for domain decomposition using piecewise linear elements with matching mesh

Figure 4.5 shows that the domain decomposition methods on matching meshes achieve $O(h^2)$ convergence for piecewise linear elements which supports the theory in the previous section. In this case it seems that the GLS term actually has a negative impact on the methods. The CIP method performs well and is the least affected by numerical pollution. The GLS/CIP method is less affected by numerical pollution than the standard Galerkin method. However, it does not perform particularly well in the case of low wave number k . Figure 4.6 gives an example of a matching mesh.

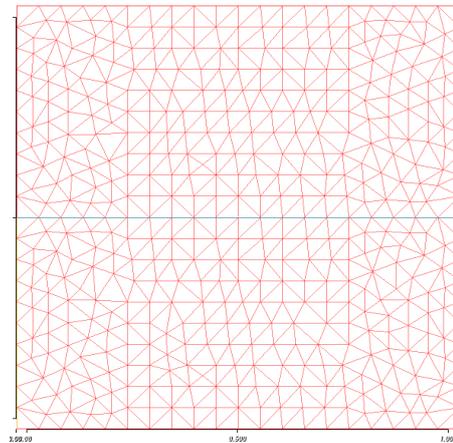


Figure 4.6: Example of a matching mesh

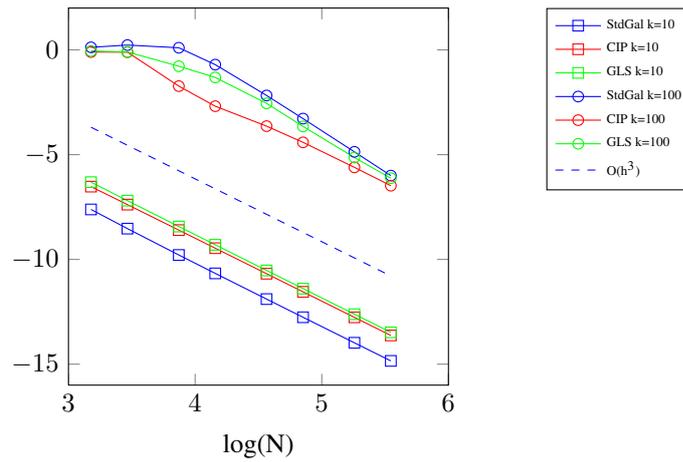


Figure 4.7: Convergence comparison for domain decomposition using piecewise quadratic elements with matching mesh

Figure 4.7 shows that the methods achieve $O(h^3)$ convergence on matching meshes for piecewise quadratic elements which also supports the theory in the previous section. The CIP method is the least affected by numerical pollution however all methods seem to perform well in the piecewise quadratic case.

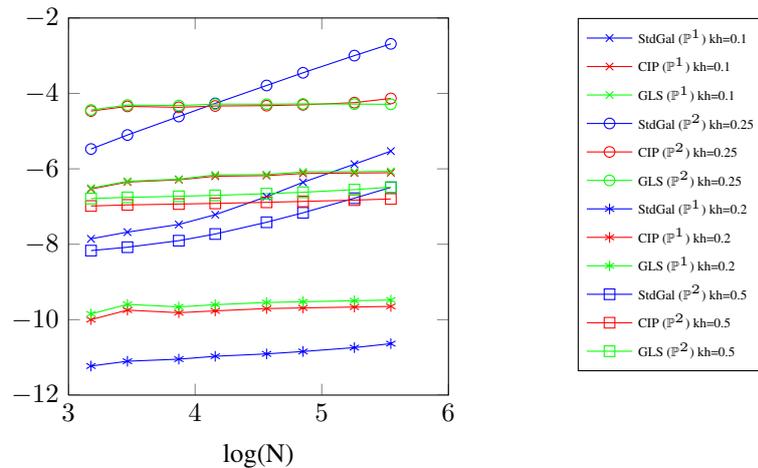


Figure 4.8: Comparison of domain decomposition methods with matching mesh for $kh = C$ which demonstrates the the stabilized methods' ability to reduce pollution

Figure 4.8 demonstrates the stabilized methods ability to reduce the pollution effect. It is worth noting that these results are similar to those presented without multi-domain coupling. This would seem to imply that the Nitsche coupling terms do not introduce any noticeable pollution.

In conclusion, the domain decomposition method does not appear to introduce any more numerical pollution than what is already present in the numerical methods, this is further verified when considering Figure 4.9 which seems to demonstrate a reduction in pollution using a non-matching mesh. This is an encouraging result since it would appear that if you have a suitable numerical method for the bulk you

can couple those methods effectively using the techniques presented here. The numerics would suggest that the CIP stabilization has a slight edge over the GLS/CIP method in the situations studied.

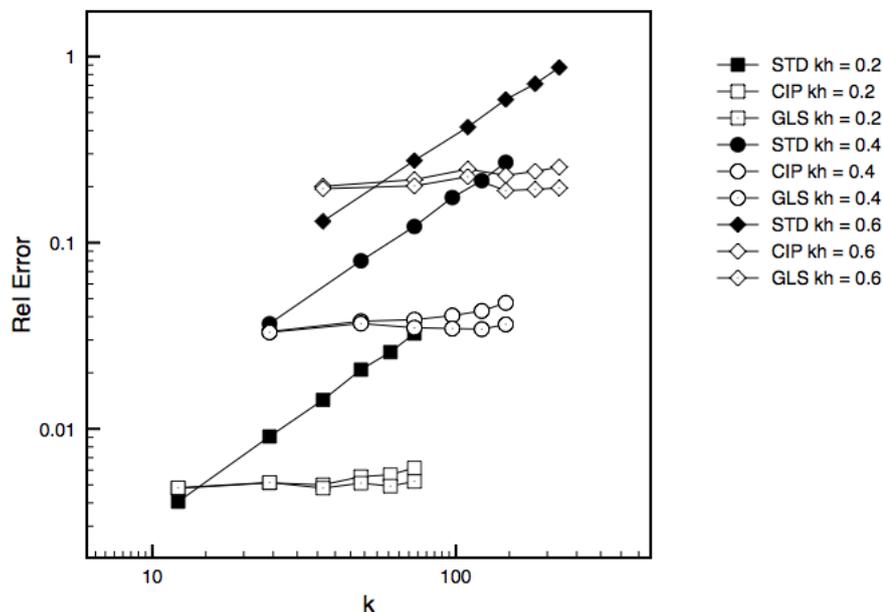


Figure 4.9: Comparison of domain decomposition methods using piecewise linear elements with a non-matching mesh for $kh = C$

4.3.8 Conclusion

The numerical results presented in this section use the same stabilization parameters as were presented for the Bessel solution previously. In addition these methods include two coupling parameters which are both taken to be $1.0i$. The numerics appear to show the domain decomposition method is more resilient to the pollution effect than the standard Galerkin method. Both of the stabilized methods converge optimally and perform well under the constraint $kh < C$. What is perhaps more surprising is how well the stabilized methods perform for non-matching meshes.

4.4 Unfitted domain decomposition

In the following section I propose an unfitted domain decomposition method using cut elements to solve Helmholtz equation on a domain with an acoustically permeable embedded interface. As mentioned previously, the wavenumber present in the Helmholtz equation is dependent upon the material properties of the medium through which it is being scattered. In a practical setting the location of the interface is usually determined by data and can be difficult to mesh exactly. It is also feasible that the interface could evolve with time which would require remeshing at every time step. Using the ideas of the fictitious domain method presented in Chapter 3, it is possible to design an unfitted domain decomposition method capable of modelling the coupling of multiple domains. In this section I analyse the potential

of using cut elements to solve a coupled Helmholtz system where the interface does not fit the mesh. The cut finite element method was introduced by Hansbo and Hansbo in [33] and has since been applied to Stoke's and the Convection-Diffusion equations with relative success. The Helmholtz problem has recently been addressed in this setting in [55] where the authors consider a Nitsche style approach to weakly enforce transmission conditions across the interface. The method that I propose differs from the aforementioned work in that it introduces a stabilization to obtain a method that has an absolute stability property for Dirichlet boundary conditions or Neumann boundary conditions and can be shown to be stable for an appropriate choice of stabilization parameters when considering impedance boundary conditions. The method introduced considers a semi-discretization of the domain in the sense that I assume that integration on the interface and over cut elements can be performed exactly.

I begin the section by introducing some useful notation to simplify the subsequent analysis. With the necessary definitions in place, I introduce the formulation for an unfitted domain decomposition method which I will show fits into a similar mathematical framework as proposed in Theorem 3. The method is shown to be stable under certain conditions on the penalty parameters introduced and is designed to reduce the dependence between the discretization and the geometric description of the problem whilst remaining optimally convergent and consistent. Optimal error estimates are a result of Theorem 3 and numerical evidence is then presented to back up the analysis.

4.4.1 Preliminaries

Considering the coupled problem, as defined in section 4.2, it is useful to introduce the following notation. Let $\{\mathcal{T}_h\}_h$ be a family of quasi-uniform, shape-regular triangulations fitted to Ω where shape regularity and quasi uniformity are defined as in (A.3) and (A.4) respectively. For each triangular element $\tau \in \mathcal{T}_h$ define $h_\tau \stackrel{\text{def}}{=} \text{diam}(\tau)$ as the diameter of τ . Let $h \stackrel{\text{def}}{=} \max_{\{\tau \in \mathcal{T}_h\}} h_\tau$ define the mesh parameter for a given triangulation \mathcal{T}_h . Let,

$$\mathcal{T}_h^j \stackrel{\text{def}}{=} \{\tau \in \mathcal{T}_h : \tau \cap \Omega_j \neq \emptyset\}. \quad (4.73)$$

It is useful to define the set of elements cut by the interface as

$$\mathcal{G}_h^j \stackrel{\text{def}}{=} \{\tau_j \in \mathcal{T}_h^j : \bar{\tau} \cap \Gamma \neq \emptyset\}, \quad (4.74)$$

and since the interface is no longer fitted to the mesh it is useful to introduce the following notation

$$\Gamma_h := \{\Sigma_h : \Sigma_h = \tau \cap \Gamma \forall \tau \in \mathcal{T}_h\}, \quad (4.75)$$

to denote the interface element-wise, Σ_h is depicted in 4.10. Using this notation it is possible to define the open discrete domains

$$\Omega_h^j \stackrel{\text{def}}{=} \text{Int} \left(\bigcup_{\tau_j \in \mathcal{T}_h^j} \bar{\tau}_j \right), \quad (4.76)$$

as well as the sets of corresponding interior faces restricted to

$$\mathcal{F}_h^j \stackrel{\text{def}}{=} \{F_j = \text{Int}(\partial\tau_1 \cap \partial\tau_2) : \tau_1, \tau_2 \in \mathcal{T}_h^j\}. \quad (4.77)$$

On the interface I shall use the notation:

$$[[v]] \stackrel{\text{def}}{=} v|_{\Omega_1} - v|_{\Omega_2}, \quad (4.78)$$

to define the jump of v over Γ .

$$\{v\} \stackrel{\text{def}}{=} \frac{v|_{\Omega_1} + v|_{\Omega_2}}{2}, \quad (4.79)$$

to define the average of v over Γ .

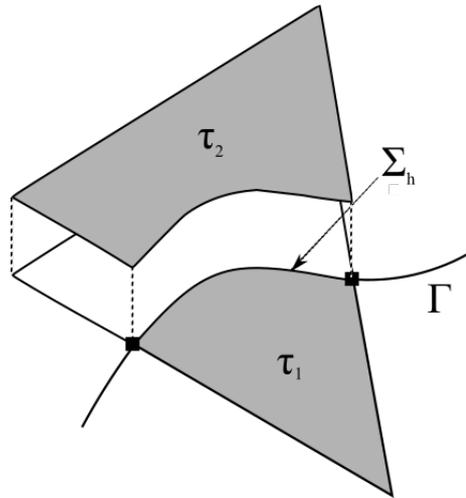


Figure 4.10: Cut element example

Remark 3. It is useful to note that if I choose to fix $n_\Gamma \stackrel{\text{def}}{=} n_1$ on the interface it is also true that $n_\Gamma \stackrel{\text{def}}{=} -n_2$ which implies that

$$\{\nabla v \cdot n_\Gamma\} \stackrel{\text{def}}{=} \frac{\nabla v|_{\Omega_1} \cdot n_1 - \nabla v|_{\Omega_2} \cdot n_2}{2}. \quad (4.80)$$

4.4.2 The formulation

Using the finite element space:

$$V_h^j \stackrel{\text{def}}{=} \{v \in H^1(\Omega_h^j) : v|_\tau \in \mathbb{P}^p(\tau), \text{ and } v|_{\partial\Omega_j \setminus \Gamma} \equiv 0\},$$

for $j \in \{1, 2\}$ it is possible to introduce an unfitted domain decomposition method using the following abstract formulation. The coupled problem posed in (4.4) becomes find $u_h \in V_h^1 \times V_h^2$ such that

$$A_h(u_h, v_h) = A(u_h, v_h) + S(u_h, v_h) = L(v_h) + S(u, v_h) = L_h(v_h) \quad \forall v_h \in V_h, \quad (4.81)$$

where $u \in H^{p+1}(\Omega_1 \cup \Omega_2)$ is the exact solution of the coupled problem defined in (4.4). Using the notation $v^j \stackrel{\text{def}}{=} v|_{\Omega_j}$ the discrete sesquilinear forms are given by

$$\begin{aligned} A(u_h, v_h) &= \sum_{j=1}^2 \left((\mu_j \nabla u_h^j, \nabla v_h^j)_{0, \Omega_j} - \omega^2 \rho_j (u_h^j, v_h^j)_{0, \Omega_j} + i\omega \sigma_j \langle u_h^j, v_h^j \rangle_{0, \partial\Omega_j} \right) \\ &\quad - \langle \{\mu \nabla u_h \cdot \mathbf{n}_\Gamma\}, \llbracket v_h \rrbracket \rangle_{0, \Gamma} - \langle \llbracket u_h \rrbracket, \{\mu \nabla v_h \cdot \mathbf{n}_\Gamma\} \rangle_{0, \Gamma}, \end{aligned}$$

and

$$S(u_h, v_h) = \sum_{j=1}^2 s_j(u_h^j, v_h^j) + \frac{i\alpha_0}{h} \langle \llbracket u_h \rrbracket, \llbracket v_h \rrbracket \rangle_{0, \Gamma} + i\alpha_1 h \langle \llbracket \mu \nabla u_h \cdot \mathbf{n}_\Gamma \rrbracket, \llbracket \mu \nabla v_h \cdot \mathbf{n}_\Gamma \rrbracket \rangle_{0, \Gamma},$$

where \mathbf{n}_Γ is taken to be the outward pointing normal from Ω_1 to Ω_2 and α_0, α_1 are real constants with the same sign as the imaginary parts of the stabilization parameters in $s_j(\cdot, \cdot)$. The local stabilization $s_j(\cdot, \cdot)$ is taken to be the GLS/CIP method which acts on the interior of the subdomain Ω_j as well as on the boundary given by $\partial\Omega_j$. The GLS/CIP stabilization form is defined as

$$\begin{aligned} s_j(u_h, v_h) &\stackrel{\text{def}}{=} \sum_{\tau_j \in \mathcal{T}_h^j} \delta_{1, \tau}^j h_\tau^2 (\mathcal{L}^j(u_h), \mathcal{L}^j(v_h))_{0, \tau_j \cap \Omega_j} + \beta_1 h \langle \mathcal{R}^j(u_h), \mathcal{R}^{*,j}(v_h) \rangle_{0, \partial\Omega_j} \\ &\quad + \sum_{F_j \in \mathcal{F}_h^j} \gamma_{1, \tau}^j \mu_j h_\tau \langle \llbracket \nabla u_h \cdot \mathbf{n} \rrbracket, \llbracket \mu \nabla v_h \cdot \mathbf{n} \rrbracket \rangle_{0, F_j}, \end{aligned} \quad (4.82)$$

where $\mathcal{L}^j(\cdot) \stackrel{\text{def}}{=} -\nabla(\mu_j \nabla(\cdot)) - \omega^2 \rho_j(\cdot)$, $\mathcal{R}^j(\cdot) \stackrel{\text{def}}{=} \nabla(\cdot) \cdot \mathbf{n} + i\omega \sqrt{\frac{\rho_j}{\mu_j}}(\cdot)$ and $\mathcal{R}^{*,j}(\cdot) \stackrel{\text{def}}{=} \nabla(\cdot) \cdot \mathbf{n} - i\omega \sqrt{\frac{\rho_j}{\mu_j}}(\cdot)$. Finally the linear form $L(\cdot)$ is given by

$$L(v_h) = (f, v_h)_{0, \Omega} + \sum_{j=1}^2 \langle g_j, v_h \rangle_{0, \partial\Omega_j}. \quad (4.83)$$

It was shown in (4.34) that the fitted domain decomposition method proposed earlier is consistent by letting $u_h = u$. Following a similar argument shows that the unfitted domain decomposition method is also consistent. The unfitted domain decomposition method is also adjoint consistent.

4.4.3 Mathematical tools

For the following analysis it is beneficial to introduce the following definitions which act on the stabilization introduced in (4.82). Let $|\cdot|_{\mathcal{J}} : V_h^1 \times V_h^2 \mapsto \mathbb{R}$ then for $u_h \in V_h^1 \times V_h^2$

$$\begin{aligned} |u_h|_{\mathcal{J}}^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\omega \sigma_j \left(1 - \frac{\omega \sigma_j}{\mu_j^2} \beta_j h \right) \|u_h\|_{0, \partial \Omega_j}^2 + \sum_{\tau \in \mathcal{T}_h^j} \delta_\tau \|\mathcal{L}^j(u_h)\|_{0, \tau \cap \Omega_j}^2 \right. \\ &\quad \left. + \sum_{F \in \mathcal{F}_h^j} \gamma_\tau h_\tau \|\llbracket \nabla u_h \cdot n \rrbracket\|_{0, F}^2 + h \beta_j \|\nabla u_h \cdot n\|_{0, \partial \Omega_j}^2 \right) \\ &\quad + \alpha_0 h^{-1} \|\llbracket u_h \rrbracket\|_{0, \Gamma}^2 + \alpha_1 h \|\llbracket \mu \nabla u_h \cdot n_\Gamma \rrbracket\|_{0, \Gamma}^2. \end{aligned} \quad (4.84)$$

It is also useful to define $|\cdot|_s : V_h^1 \times V_h^2 \mapsto \mathbb{R}$

$$\begin{aligned} |u_h|_s^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\beta_j h \|\nabla u_h \cdot n - i \frac{\omega \sigma_j}{\mu_j} u\|_{0, \partial \Omega_j}^2 + \sum_{\tau \in \mathcal{T}_h^j} h_\tau^2 \delta_\tau \|\mathcal{L}^j(u_h)\|_{0, \tau \cap \Omega_j}^2 \right. \\ &\quad \left. + \sum_{F \in \mathcal{F}_h^j} \gamma_\tau h_\tau \|\llbracket \nabla u_h \cdot n \rrbracket\|_{0, F}^2 \right) \\ &\quad + \alpha_0 h^{-1} \|\llbracket u_h \rrbracket\|_{0, \Gamma}^2 + \alpha_1 h \|\llbracket \mu \nabla u_h \cdot n_\Gamma \rrbracket\|_{0, \Gamma}^2. \end{aligned} \quad (4.85)$$

For an appropriate choice of stabilization parameters both $|\cdot|_{\mathcal{J}}$ and $|\cdot|_s$ define semi-norms on $V_h^1 \times V_h^2$.

Lemma 4.5. *Given $\text{Re}[\alpha_0] = \text{Re}[\alpha_1] = \text{Re}[\beta_j] = 0$, $\min_j \{\text{Im}[\beta_j^{-1}]\} > h \frac{\omega \sigma_j}{\mu_j^2}$ and the imaginary parts of all penalty parameters strictly positive $|\cdot|_{\mathcal{J}}$ is a semi-norm on $V_h^1 \times V_h^2$.*

Proof. The proof is immediate from the definition. \square

Lemma 4.6. *Under the same conditions as in Lemma 4.5 $|\cdot|_s$ is also a semi-norm on $V_h^1 \times V_h^2$.*

Proof. The proof is immediate from the definition. \square

In order to perform the following analysis it is necessary to choose extension operators as given in [51]

$\mathbb{E}_j^p : H^p(\Omega_j) \mapsto H^p(\Omega)$ such that $(\mathbb{E}_j^p v)|_{\Omega_j} = v$ and

$$\|\mathbb{E}_j^p v\|_{p, \Omega} \leq C \|v\|_{p, \Omega_j} \quad \forall v \in H^p(\Omega_j), \quad p \in \mathbb{N}_0. \quad (4.86)$$

Let \mathcal{I}_h be the standard Lagrange interpolant and define

$$\mathcal{I}_h^* v \stackrel{\text{def}}{=} (\mathcal{I}_{h,1}^* v_1, \mathcal{I}_{h,2}^* v_2), \quad \text{where } \mathcal{I}_{h,j}^* v_j \stackrel{\text{def}}{=} (\mathcal{I}_h \mathbb{E}_j^p v_j)|_{\Omega_j}. \quad (4.87)$$

I demonstrated earlier in (3.39) that the extended Lagrange interpolant satisfies the following estimate

$$\|u - \mathcal{I}_h^* u\|_{s,\Omega} \leq h^{p-s+1} \|u\|_{p+1,\Omega}. \quad (4.88)$$

It was shown in [33] that if the interface satisfies assumptions [FD1-FD3] given in Chapter 3 then the following lemma holds.

Lemma 4.7. *Under mesh assumptions [FD1-FD3] there exists a constant C_T , dependent on Γ , such that*

$$\|v\|_{0,\Sigma_h}^2 \leq C_T (h_\tau^{-1} \|v\|_{0,\tau}^2 + h_\tau \|\nabla v\|_{0,\tau}^2) \quad \forall v \in H^1(\tau), \quad (4.89)$$

where C_T is independent of the mesh and Σ_h is as defined in (4.76).

Finally, I introduce the norm $\|\cdot\|_* : V_h^1 \times V_h^2 \mapsto \mathbb{R}$

$$\begin{aligned} \|(\cdot)\|_*^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\left[\sum_{\tau \in \mathcal{T}_{j,h}} (h_\tau^2 |\delta_\tau|)^{-1} \|(\cdot)\|_{0,\tau \cap \Omega_j}^2 + \sum_{F \in \mathcal{F}_h^j} (h_{\tau_j} |\gamma_{\tau_j}|)^{-1} \|(\cdot)\|_{0,F}^2 \right. \right. \\ &\quad \left. \left. + \mu_j (h |\beta_j|)^{-1} \|(\cdot)\|_{0,\partial\Omega_j}^2 \right] + h |\alpha_0|^{-1} \|\{\mu \nabla(\cdot) \cdot n_\Gamma\}\|_{0,\Gamma}^2 + h^{-1} |\alpha_1|^{-1} \|\{(\cdot)\}\|_{0,\Gamma}^2 \right. \\ &\quad \left. + |(\cdot)|_{S^*}^2 \right), \end{aligned} \quad (4.90)$$

where

$$\begin{aligned} |u|_{S^*}^2 &\stackrel{\text{def}}{=} \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(\cdot)\|_{0,\tau \cap \Omega_j}^2 + \sum_{F \in \mathcal{F}_h^j} \gamma_\tau h_\tau \|\llbracket \nabla(\cdot) \cdot n \rrbracket\|_{0,F}^2 \right. \\ &\quad \left. + \beta_j h \|\mathcal{R}^j(\cdot)\|_{0,\partial\Omega_j}^2 \right) + \frac{\alpha_0}{h} \|\llbracket (\cdot) \rrbracket\|_{0,\Gamma}^2 + (\alpha_1 h) \|\llbracket \mu \nabla(\cdot) \cdot n \rrbracket\|_{0,\Gamma}^2. \end{aligned}$$

4.4.4 A priori error estimates

In the following section I will prove that the unfitted domain decomposition method proposed in (4.81) fits into the mathematical framework proposed earlier. An important observation is that since the framework does not rely on a coercivity result it is possible to omit the introduction of any additional stabilization terms such as the Ghost penalty which was introduced by Burman in [12] and has been used to great effect for fictitious domain methods involving cut elements, see Figure 4.10 for an example of a cut element. It is interesting to note that the Ghost penalty is actually contained in the CIP part of the stabilization over element faces. No additional stabilization is needed to ensure the integrity of data where the intersection of the cut element and the physical domain become small.

Lemma 4.8. *Under the assumptions on the stabilization parameters given in (4.5), the unfitted domain decomposition method introduced in (4.81) satisfies the weak coercivity and Cauchy-Schwarz estimates given in (4.36) and (4.37) respectively for $|\cdot|_\mathcal{I}$ defined in (4.84) and $|\cdot|_s$ defined in (4.85).*

Proof. Under the assumptions of the Lemma $|\cdot|_{\mathcal{I}}$ it is simple to show that,

$$|u_h|_{\mathcal{I}}^2 = \text{Im} [A_h(u_h, u_h)] \quad (4.91)$$

$$\leq |A_h(u_h, u_h)|, \quad (4.92)$$

(4.36) follows trivially. The Cauchy-Schwarz type inequality in (4.37) asks that the stabilization respects an estimate of the form,

$$(a) \stackrel{\text{def}}{=} |S(v_h, w_h)| \leq |v_h|_{\mathcal{I}} |w_h|_S,$$

the result follows after an application of the triangle and Cauchy-Schwarz inequalities

$$\begin{aligned} (a) &\leq \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} \delta_{\tau} h_{\tau}^2 \|\mathcal{L}^j(v_h^j)\|_{0,\tau \cap \Omega_j} \|\mathcal{L}^j(w_h^j)\|_{0,\tau \cap \Omega_j} + \beta_j h \|\mathcal{R}(v_h^j)\|_{0,\partial \Omega_j} \|\mathcal{R}^*(w_h^j)\|_{0,\partial \Omega_j} \right. \\ &\quad \left. + \sum_{F \in \mathcal{F}_h^j} \gamma_{\tau} h_{\tau} \|\llbracket \nabla v_h^j \cdot n \rrbracket\|_{0,F_j} \|\llbracket \nabla w_h^j \cdot n \rrbracket\|_{0,F_j} \right) + \frac{\alpha_0}{h} \|\llbracket v_h \rrbracket\|_{0,\Gamma} \|\llbracket w_h \rrbracket\|_{0,\Gamma} \\ &\quad + \alpha_1 h \|\llbracket \mu \nabla v_h \cdot n_{\Gamma} \rrbracket\|_{0,\Gamma} \|\llbracket \mu \nabla w_h \cdot n_{\Gamma} \rrbracket\|_{0,\Gamma} \\ &\leq |v_h|_{\mathcal{I}} |w_h|_S. \end{aligned}$$

Using a similar proof it is possible to show

$$|S(v_h, w_h)| \leq |v_h|_{S^*} |w_h|_{\mathcal{I}}.$$

□

Lemma 4.9. *Let $u \in H^{p+1}(\Omega_1 \cup \Omega_2)$, for $p > 1/2$, and $z \in H^2(\Omega_1 \cup \Omega_2)$, be the exact solutions to the primal and dual problems defined by (4.4) and (4.5) respectively. Under the assumptions on the stabilization parameters given in (4.5), the semi norm $|\cdot|_{\mathcal{I}}$ defined in (4.84) satisfies (4.42) and (4.44) for the unfitted domain decomposition method introduced in (4.81) with interpolation operator \mathcal{I}_h^* taken to be the extended Lagrange interpolant proposed in (4.87). For the same choice of \mathcal{I}_h^* the norm introduced in (4.54) satisfies (4.43).*

Proof. I begin the proof by first investigating the stability of the extended Lagrange interpolant in $|\cdot|_S$. Recall that since the exact solution, z , of the adjoint problem (4.5) is assumed to be smooth in each sub-domain and the adjoint Robin boundary condition applied to z vanishes on $\partial \Omega_j$, I can add it into the CIP and Robin boundary penalty terms. This is also true for the penalty term on the interface since the exact solution, z , is assumed to satisfy the interface conditions for the adjoint problem. Using this trick

of adding zero the following holds

$$\begin{aligned} |\mathcal{I}_h^* z|_S^2 &= \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(\mathcal{I}_h^* z)\|_{0,\tau \cap \Omega_j}^2 + \sum_{F \in \mathcal{F}_{int}^j} \gamma_\tau h_\tau \mu_j \|\llbracket \nabla(z - \mathcal{I}_h^* z) \cdot n \rrbracket\|_{0,F}^2 \right. \\ &\quad \left. + \beta h \|\mathcal{R}^*(z - \mathcal{I}_h^* z)\|_{0,\partial \Omega_j}^2 \right) + \frac{\alpha_0}{h} \|\llbracket z - \mathcal{I}_h^* z \rrbracket\|_{0,\Gamma}^2 + \alpha_1 h \|\llbracket \mu \nabla(z - \mathcal{I}_h^* z) \cdot n_\Gamma \rrbracket\|_{0,\Gamma}^2. \end{aligned}$$

With the addition of the exact solution z , the estimate follows using the standard interpolation results which can be shown to hold for the extended Lagrange interpolant for each term bar the GLS term. To conclude the proof note that

$$\begin{aligned} \sum_{j=1}^2 \sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(\mathcal{I}_h^* z - z)\|_{0,\tau \cap \Omega_j}^2 &\geq \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(\mathcal{I}_h^* z)\|_{0,\tau \cap \Omega_j}^2 - \sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(z)\|_{0,\tau \cap \Omega_j}^2 \right) \\ &= \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(\mathcal{I}_h^* z)\|_{0,\tau \cap \Omega_j}^2 - \sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\psi\|_{0,\tau \cap \Omega_j}^2 \right). \end{aligned}$$

From this it follows that

$$\underbrace{\sum_{j=1}^2 \sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(\mathcal{I}_h^* z)\|_{0,\tau \cap \Omega_j}^2}_{(b)} \leq \left(\sum_{j=1}^2 \sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\mathcal{L}^j(\mathcal{I}_h^* z - z)\|_{0,\tau \cap \Omega_j}^2 + \sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \|\psi\|_{0,\tau \cap \Omega_j}^2 \right).$$

The proof is completed using the triangle inequality and interpolation estimates as well as H^2 stability of the interpolant.

$$\begin{aligned} (b) &\leq \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} \delta_\tau h_\tau^2 \left(\|\Delta(\mathcal{I}_h^* z)\|_{0,\tau \cap \Omega_j}^2 + \|\Delta z\|_{0,\tau \cap \Omega_j}^2 \right. \right. \\ &\quad \left. \left. + \|\omega^2 \rho_j(z - \mathcal{I}_h^* z)\|_{0,\tau \cap \Omega_j}^2 + \|\psi\|_{0,\tau \cap \Omega_j}^2 \right) \right) \\ &\leq \sum_{j=1}^2 \left(\delta h^2 \left(|z|_{2,\Omega_j} + \max_j \{\rho_j^2\} (\omega h)^4 |z|_{2,\Omega_j} + \|\psi\|_{0,\Omega_j}^2 \right) \right) \\ &\leq \sum_{j=1}^2 \left(\delta \left(\max_j \{\rho_j\} (\omega h)^2 + (\max_j \{\rho_j\})^3 (\omega h)^6 + h^2 \right) \|\psi\|_{0,\Omega}^2 \right). \end{aligned}$$

For higher order elements the result follows as a direct consequence of the interpolation estimates for the extended Lagrange interpolant.

To complete the proof of the proposition I must show the additional interpolation estimates given by (4.43) and (4.44).

$$|u - \mathcal{I}_h^* u|_X \leq C \sum_{j=1}^2 h^p \|u\|_{p+1,\Omega_j}.$$

Let $\eta_j^* := u^j - \mathcal{I}_h^* u^j$

$$\begin{aligned} |\eta_j^*|_{\mathcal{I}}^2 &= \sum_{j=1}^2 \left(\omega \sqrt{\rho_j} (1 - h\omega \sqrt{\rho_j} \beta_j) \|\eta_j^*\|_{0,\partial\Omega_j}^2 + \sum_{\tau \in \mathcal{T}_h^j} \text{Im}[\delta_\tau] h_\tau^2 \|\mathcal{L}^j(\eta_j^*)\|_{0,\tau \cap \Omega_j}^2 \right. \\ &\quad \left. + \sum_{F \in \mathcal{F}_h^j} \text{Im}[\gamma_\tau] h_\tau \|\llbracket \nabla \eta_j^* \cdot n \rrbracket\|_{0,F}^2 + \beta_j h \|\nabla \eta_j^* \cdot n\|_{0,\partial\Omega_j}^2 \right) \\ &\quad + \frac{\alpha_0}{h} \|\llbracket \eta_j^* \rrbracket\|_{0,\Gamma}^2 + \alpha_1 h \|\llbracket \mu \nabla(\eta_j^*) \cdot n_\Gamma \rrbracket\|_{0,\Gamma}^2. \end{aligned}$$

It is possible to bound the terms in the bracket above by applying a combination of the standard trace inequality (A:4), triangle inequality and extended interpolation results (4.88). However, the terms on the interface must be treated slightly differently. Since the interface is allowed to arbitrarily cut the background mesh the user is left with cut elements where only a section of the element belongs to a domain Ω_j . Cut elements are defined as elements that are intersected by the interface Γ . In this scenario the interface is no longer described on the boundary of elements and the standard trace inequality is no longer valid. Therefore, in order to prove the result, I must use the cut element trace inequality (4.89) stated earlier.

$$\begin{aligned} \frac{\alpha_0}{h} \|\llbracket u - \mathcal{I}_h^* u \rrbracket\|_{0,\Gamma}^2 &\leq C_T^2 \frac{\alpha_0}{h} \sum_{j=1}^2 \sum_{\tau \in \mathcal{G}_h^j} (h_\tau^{-1} \|u_j - \mathcal{I}_h^* u_j\|_{0,\tau}^2 \\ &\quad + h_\tau \|\nabla(u_j - \mathcal{I}_h^* u_j)\|_{0,\tau}^2) \\ &\leq C \sum_{j=1}^2 \alpha_0 h^{2p} \|u\|_{p+1,\Omega_j}^2. \end{aligned}$$

After an application of the cut element trace inequality, the results follow from the extended Lagrange interpolation estimates given in (4.88).

$$\begin{aligned} \alpha_1 h \|\llbracket \mu \nabla(u - \mathcal{I}_h^* u) \cdot n_\Gamma \rrbracket\|_{0,\Gamma} &\leq C_T^2 \alpha_1 h \sum_{j=1}^2 \sum_{\tau \in \mathcal{G}_h^j} (h_\tau^{-1} \|\mu_j \nabla(u_j - \mathcal{I}_h^* u_j)\|_{0,\tau}^2 \\ &\quad + h_\tau \|\{\mu_j \nabla(u_j - \mathcal{I}_h^* u_j)\}_{1,\tau}\|_{0,\tau}^2) \\ &\leq C \sum_{j=1}^2 \alpha_1 \max \mu_j^2 h^{2p} \|u\|_{p+1,\Omega_j}^2. \end{aligned}$$

The final interpolation result to show is (4.43), which demonstrates the approximability of u in the star norm. That is

$$\|\eta^*\|_* \leq Ch^p \|u\|_{0,p+1}.$$

Recall the definition of the star norm

$$\begin{aligned} \|\eta^*\|_* &= \sum_{j=1}^2 \left(\left[\sum_{\tau \in \mathcal{T}_{j,h}} (h_\tau^2 |\delta_\tau|)^{-1} \|\eta_j^*\|_{0,\tau \cap \Omega_j}^2 + \sum_{F \in \mathcal{F}_h^j} (h_\tau |\gamma_\tau|)^{-1} \|\eta_j^*\|_{0,F}^2 \right. \right. \\ &\quad \left. \left. + \mu_j (h |\beta_j|)^{-1} \|\eta_j^*\|_{0,\partial\Omega_j}^2 \right] + h |\alpha_0|^{-1} \|\{\mu \nabla \eta^* \cdot n_\Gamma\}\|_{0,\Gamma}^2 + h^{-1} |\alpha_1|^{-1} \|\{\eta^*\}\|_{0,\Gamma}^2 \right. \\ &\quad \left. + |\eta^*|_{S^*}^2 \right)^{\frac{1}{2}}. \end{aligned} \quad (4.93)$$

Similarly to the previous estimate the majority of terms can be shown to satisfy the result using the standard trace inequality (A:4), triangle inequality and extended interpolation results (4.88). Once again it is only the cut elements on the interface that require special attention and the estimate follows trivially after an application of the cut element trace inequality (4.89). I state the results below for completeness.

$$\begin{aligned} \left(\frac{h}{\alpha_0}\right) \|\{\nabla(u - \mathcal{I}_h^* u) \cdot n\}\|_{0,\Gamma}^2 &\leq C_T^2 \frac{1}{4} \left(\frac{h}{\alpha_0}\right) \sum_{j=1}^2 \sum_{\tau \in \mathcal{G}_h^j} (h_\tau^{-1} \|\nabla(u_j - \mathcal{I}_h^* u_j)\|_{0,\tau}^2 \\ &\quad + h_\tau \|\nabla(u_j - \mathcal{I}_h^* u_j)\|_{1,\tau}^2) \\ &\leq C \sum_{j=1}^2 \alpha_0^{-1} h^{2p} \|u\|_{p+1,\Omega_j}^2. \end{aligned}$$

$$\begin{aligned} (\alpha_1 h)^{-1} \|\{(u - \mathcal{I}_h^* u)\}\|_{0,\Gamma}^2 &\leq C_T^2 \frac{1}{4} (\alpha_1 h)^{-1} \sum_{j=1}^2 \sum_{\tau \in \mathcal{G}_h^j} (h_\tau^{-1} \|(u_j - \mathcal{I}_h^* u_j)\|_{0,\tau}^2 \\ &\quad + h_\tau \|\nabla(u_j - \mathcal{I}_h^* u_j)\|_{0,\tau}^2) \\ &\leq C \sum_{j=1}^2 \alpha_1^{-1} h^{2p} \|u\|_{p+1,\Omega_j}^2. \end{aligned}$$

Collecting the results concludes the proof. □

Now that the interpolation and stability estimates have been proven the only thing left to show are the continuity results of Theorem 3.

Proposition 7 (Continuity). *Let $u \in H^{p+1}(\Omega_1 \cup \Omega_2)$, for $p > 1/2$, and $z \in H^2(\Omega_1 \cup \Omega_2)$, be the exact solution to the primal and dual problems posed in (4.4) and (4.5), respectively. Then it holds that the unfitted domain decomposition method given by (4.81) satisfies assumptions (4.40) and (4.41) of Theorem 3 for π_h defined as the extended Lagrange interpolant, \mathcal{I}_h^* given in (4.87), and $\|\cdot\|_*$ as given in (4.90).*

Proof. To prove (4.40) it is enough to show that

$$\underbrace{|A(u - \mathcal{I}_h^* u, v_h)|}_{(A)} + \underbrace{|s(u - \mathcal{I}_h^* u, v_h)|}_{(B)} \leq C \|u - \mathcal{I}_h^* u\|_* |v_h|_{\mathcal{I}}.$$

Since,

$$\sum_{j=1}^2 2\sqrt{\rho_j \omega} \|u - \mathcal{I}_h^* u\|_{0, \partial\Omega_R} \|v_h\|_{0, \partial\Omega_R} \leq C \|u - \mathcal{I}_h^* u\|_* |v_h|_{\mathcal{I}},$$

follows trivially.

To show that (A) satisfies the estimate I begin by performing an element-wise application of integration by parts to the first term in a similar way to how I treated the domain decomposition method. However, in the case of the unfitted domain decomposition method the interface introduces cut elements. The integration by parts is performed element-wise up to the interface as shown in Figure 4.11. Note that this is perfectly acceptable as the element is allowed to be discontinuous across the interface.

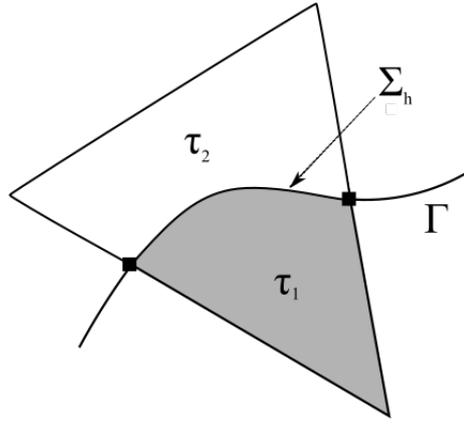


Figure 4.11: Example of the domain of integration for cut elements, where $\tau_j \stackrel{\text{def}}{=} \tau \cap \Omega_j$

This integration introduces additional terms at the interface in a similar way to the domain decomposition method described earlier. The important distinction being that the interface is no longer fitted to the mesh.

Performing the integration shows

$$\begin{aligned} \sum_{j=1}^2 \sum_{\tau \in \mathcal{T}_h^j} (\mu_j \nabla \eta_j^*, \nabla v_h)_{0, \tau \cap \Omega_j} &= \sum_j \sum_{\tau \in \mathcal{T}_h^j} \left((\eta_j^*, -\nabla \cdot (\mu_j \nabla v_h))_{0, \tau \cap \Omega_j} + \langle \eta_j^*, \mu_j \nabla v_h \cdot \mathbf{n} \rangle_{0, \partial\tau \cap \Omega_j} \right) \\ &= \sum_{j=1}^2 \left[\sum_{\tau \in \mathcal{T}_h^j} (\eta_j^*, -\nabla \cdot (\mu_j \nabla v_h))_{0, \tau \cap \Omega_j} + \sum_{F \in \mathcal{F}_h^j} \langle \eta_j^*, \mu_j \llbracket \nabla v_h \cdot \mathbf{n} \rrbracket \rangle_{0, F} \right. \\ &\quad \left. + \langle \mu_j \eta_j^*, \mathcal{R}_j^*(v_h) \rangle_{0, \partial\Omega_j} + \sum_{\Sigma_h \in \Gamma_h} \langle \eta_j^*, \mu_j \nabla v_h^j \cdot \mathbf{n} \rangle_{0, \Sigma_h} \right]. \quad (4.94) \end{aligned}$$

Comparing the new interface terms present in the integration by parts with the existing interface terms

in the formulation gives

$$(I) = \sum_{\Sigma_h \in \Gamma_h} \langle \eta_1^*, \mu_1 \nabla v_1 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} - \langle \eta_2^*, \mu_2 \nabla v_2 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} - \langle \llbracket \eta^* \rrbracket, \{\mu \nabla v_h \cdot \mathbf{n}_\Gamma\} \rangle_{0, \Sigma_h} \quad (4.95)$$

$$\begin{aligned} &= \sum_{\Sigma_h \in \Gamma_h} \langle \eta_1^*, \mu_1 \nabla v_1 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} - \langle \eta_2^*, \mu_2 \nabla v_2 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} - \frac{1}{2} \langle \eta_1^*, \mu_1 \nabla v_1 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} \\ &\quad + \frac{1}{2} \langle \eta_2^*, \mu_1 \nabla v_1 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} - \frac{1}{2} \langle \eta_1^*, \mu_2 \nabla v_2 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} \\ &\quad + \frac{1}{2} \langle \eta_2^*, \mu_2 \nabla v_2 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} \end{aligned} \quad (4.96)$$

$$\begin{aligned} &= \sum_{\Sigma_h \in \Gamma_h} \frac{1}{2} \langle \eta_1^*, \mu_1 \nabla v_1 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} - \frac{1}{2} \langle \eta_2^*, \mu_2 \nabla v_2 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} + \frac{1}{2} \langle \eta_2^*, \mu_1 \nabla v_1 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h} \\ &\quad - \frac{1}{2} \langle \eta_1^*, \mu_2 \nabla v_2 \cdot \mathbf{n}_\Gamma \rangle_{0, \Sigma_h}. \end{aligned} \quad (4.97)$$

After collecting like terms the interface terms reduce to

$$(I) = \sum_{\Sigma_h \in \Gamma_h} \frac{1}{2} \langle \eta_1^*, \llbracket \mu \nabla v_h \cdot \mathbf{n}_\Gamma \rrbracket \rangle_{0, \Sigma_h} + \frac{1}{2} \langle \eta_2^*, \llbracket \mu \nabla v_h \cdot \mathbf{n}_\Gamma \rrbracket \rangle_{0, \Sigma_h} \quad (4.98)$$

$$= \sum_{\Sigma_h \in \Gamma_h} \langle \{\eta^*\}, \llbracket \mu \nabla v_h \cdot \mathbf{n}_\Gamma \rrbracket \rangle_{0, \Sigma_h}. \quad (4.99)$$

The jump term in the right hand argument is present in the stabilization. As was the case for the domain decomposition method presented earlier, this term emphasises the need to stabilize continuity of the fluxes over the interface. Collecting results shows that the physical form $A(., .)$ respects the continuity estimate given by (4.40).

$$\begin{aligned} |A(\eta^*, v_h)| &= \left| \sum_{j=1}^2 \left(\sum_{\tau \in \mathcal{T}_h^j} (\eta_j^*, \mathcal{L}^j(v_h))_{0, \tau \cap \Omega_j} + \sum_{F \in \mathcal{F}_h^j} \langle \eta_j^*, \llbracket \mu \nabla v_h \cdot \mathbf{n} \rrbracket \rangle_{0, F} + \langle \mu_j \eta_j^*, \mathcal{R}_j^*(v_h) \rangle_{0, \partial \Omega_j} \right) \right. \\ &\quad \left. - \langle \{\mu \nabla \eta^* \cdot \mathbf{n}\}, \llbracket v_h \rrbracket \rangle_{0, \Gamma} - \langle \{\eta^*\}, \llbracket \mu \nabla v_h \cdot \mathbf{n} \rrbracket \rangle_{0, \Gamma} \right| \end{aligned} \quad (4.100)$$

$$\leq \| \eta^* \|_* |v_h|_{\mathcal{I}}. \quad (4.101)$$

Finally, the stabilization term $s(., .)$ can be shown to satisfy (4.40) after an application of Cauchy-Schwarz. Recall the continuity estimate (4.40)

$$|A(u - \pi_h^* u, z - \pi_h^* z)| \leq C(\mu_j, \rho_j) h \omega(h^p) \|u\|_{p+1, \Omega_1 \cup \Omega_2} \|\psi\|_{0, \Omega},$$

confirmation of this claim follows from the interpolation results that hold for the extended Lagrange interpolant as well as the assumed regularity estimate associated with the adjoint problem. Continuity of

the first term in $A(\cdot, \cdot)$ follows directly in this manner.

$$\begin{aligned} \sum_{j=1}^2 (\mu_j \nabla(u - \pi_h^* u), \nabla(z - \pi_h^* z))_{0, \Omega_j} &\leq \sum_{j=1}^2 \|\mu_j \nabla(u - \pi_h^* u)\|_{0, \Omega_j} \|\nabla(z - \pi_h^* z)\|_{0, \Omega_j} \\ &\leq \sum_{j=1}^2 C'(h\omega) h^p \|u\|_{p+1, \Omega_j} \|\psi\|_{0, \Omega}, \end{aligned}$$

where $C' = C \max \left\{ \sqrt{\frac{\rho_j}{\mu_j}} \right\} \max \{\mu_j\}$. Notice that, like in the fitted domain decomposition method, the constant in the continuity estimate is related to the contrast in constants μ_j associated with the Laplace operators in the coupled problem. The credibility of the result relies on the assumption that this constant is bounded. The low order term follows in an equally straightforward manner,

$$\begin{aligned} \sum_{j=1}^2 \omega^2 \rho_j (u - \pi_h^* u, z - \pi_h^* z)_{0, \Omega_j} &\leq \sum_{j=1}^2 \omega^2 \rho_j \|u - \pi_h^* u\|_{0, \Omega_j} \|z - \pi_h^* z\|_{0, \Omega_j} \\ &\leq \sum_{j=1}^2 C \max_j \left\{ \sqrt{\frac{\rho_j^3}{\mu_j}} \right\} (h\omega)^3 h^p \|u\|_{p+1, \Omega_j} \|\psi\|_{0, \Omega}. \end{aligned}$$

The terms on the fitted boundary are treated in the same way as for the fitted domain decomposition method presented earlier. The result is presented here for completeness and follows after an application of the trace inequality.

$$\begin{aligned} \sum_{j=1}^2 \omega \sigma_j (u - \pi_h^* u, z - \pi_h^* z)_{0, \partial \Omega_j} &\leq \sum_{j=1}^2 \omega \sigma_j \|u - \pi_h^* u\|_{0, \partial \Omega_j} \|z - \pi_h^* z\|_{0, \partial \Omega_j} \\ &\leq \sum_{j=1}^2 \sum_{\tau \in \mathcal{T}_h^j} C_T^2 \omega \sigma_j (h_\tau^{-1} \|u - \pi_h^* u\|_{0, \tau}^2 + h_\tau \|u - \pi_h^* u\|_{1, \tau}^2)^{1/2} \\ &\quad \left(h_\tau^{-1} \|z - \pi_h^* z\|_{0, \tau}^2 + h_\tau \|z - \pi_h^* z\|_{1, \tau}^2 \right)^{1/2} \\ &\leq C \max \left\{ \sigma_j \sqrt{\frac{\rho_j}{\mu_j}} \right\} (h\omega)^2 h^p \|u\|_{p+1, \Omega} \|\psi\|_{0, \Omega}. \end{aligned}$$

The interface terms follow in a similar way to the fitted domain decomposition case. An application of the Cauchy-Schwarz inequality gives

$$\langle \{\nabla(u - \pi_h^* u) \cdot \mathbf{n}_\Gamma\}, \llbracket z - \pi_h^* z \rrbracket \rangle_{0, \Gamma} \leq \| \{\nabla(u - \pi_h^* u) \cdot \mathbf{n}_\Gamma\} \|_{0, \Gamma} \| \llbracket z - \pi_h^* z \rrbracket \|_{0, \Gamma}.$$

The result follows after an element-wise application of the cut element trace inequality followed by the

use of the extended interpolation estimates derived earlier.

$$\begin{aligned} \|\{\nabla(u - \pi_h^* u) \cdot \mathbf{n}_\Gamma\}\|_{0,\Gamma} &\leq \frac{1}{2} C_T \sum_{j=1}^2 \sum_{\tau \in \mathcal{G}_h^j} (h_\tau^{-1} \|\nabla(u_j - \pi_h^* u_j)\|_{0,\tau}^2 \\ &\quad + h_\tau \|\nabla(u_j - \pi_h^* u_j)\|_{1,\tau}^2)^{\frac{1}{2}} \\ &\leq \sum_{j=1}^2 C_T h^{p-\frac{1}{2}} \|u\|_{p+1,\Omega_j}. \end{aligned}$$

For completeness I state the results.

$$\begin{aligned} \| [z - \pi_h^* z] \|_{0,\Gamma} &\leq C_T \sum_{j=1}^2 \sum_{\tau \in \mathcal{G}_h^j} \left(h_{\tau_j}^{-1} \|z_j - \pi_h^* z_j\|_{0,\tau}^2 \right. \\ &\quad \left. + h_\tau \|z_j - \pi_h^* z_j\|_{1,\tau}^2 \right)^{1/2} \\ &\leq C_T \max_j \left\{ \sqrt{\frac{\rho_j}{\mu_j}} \right\} \omega h^{\frac{3}{2}} \|\psi\|_{0,\Omega}. \end{aligned}$$

Repeating a similar process for the second interface term

$$\langle [u - \pi_h^* u], \{\nabla(z - \pi_h^* z) \cdot \mathbf{n}_\Gamma\} \rangle_{0,\Gamma} \leq \| [u - \pi_h^* u] \|_{0,\Gamma} \| \{\nabla(z - \pi_h^* z) \cdot \mathbf{n}_\Gamma\} \|_{0,\Gamma},$$

completes the proof. \square

The unfitted domain decomposition method proposed in (4.82) has been shown to enter the same mathematical framework as the fitted domain decomposition method proposed in the previous section. It can be seen that for sufficiently regular u results are optimal.

4.4.5 Numerical Results

Let $\Omega \subset \mathbb{R}^2$ be $[-0.6, 0.6] \times [-0.6, 0.6]$. Define Ω_1 as a ball of radius 0.5 centred at $(0, 0)$ and $\Omega_2 \stackrel{\text{def}}{=} \Omega \setminus \bar{\Omega}_1$. The problem is defined as

$$\left. \begin{aligned} -\nabla \cdot (\nabla u) - \omega^2 \rho_1 u &= f_1 && \text{in } \Omega_1 \\ -\nabla \cdot (\nabla u) - \omega^2 \rho_2 u &= f_1 + (\rho_1 \omega^2 - \rho_2 \omega^2) u && \text{in } \Omega_2 \\ u_2 &= g_{2,D} && \text{for } x \in [-0.6, 0.6], y = -0.6 \\ u_2 &= g_{2,D} && \text{for } x = -0.6, y \in [-0.6, 0.6] \\ \nabla u \cdot \mathbf{n} + i\omega \sqrt{\rho_2} u &= g_{2,R} && \text{for } x = 0.6, y \in [-0.6, 0.6] \\ \nabla u \cdot \mathbf{n} + i\omega \sqrt{\rho_2} u &= g_{2,R} && \text{for } x \in [-0.6, 0.6], y = 0.6 \\ u|_1 &= u|_2 && \text{on } \Gamma \\ \nabla u|_1 \cdot \mathbf{n}_\Gamma &= \nabla u|_2 \cdot \mathbf{n}_\Gamma && \text{on } \Gamma, \end{aligned} \right\} \quad (4.102)$$

Taking $\rho_j \omega^2 = k_j^2$ the problem is designed such that

$$u = \frac{\cos(k_1 \sqrt{x^2 + y^2})}{k_1} - C J_0(k_1 \sqrt{x^2 + y^2})$$

where

$$C = \frac{\cos(k_1) + i \sin(k_1)}{k_1 (J_0(k_1) + i J_1(k_1))}$$

and $J_0(\cdot)$ and $J_1(\cdot)$ are Bessel functions of the First Kind. For the stabilized method proposed in (4.81) the calculations are performed using the same choice of parameters as taken in Chapter 2. The Nitsche coupling parameters are taken to be $\alpha_0 = \alpha_1 = 1$ which numerical evidence seems to suggest is a good choice. In order to present some comparison I also consider the case where all of the stabilization parameters acting on the bulk of each domain are taken to be 0 with $\alpha_0 = \alpha_1 = 1$, I refer to this case as the ‘no stabilization case’ or just ‘Nitsche’s’ case. All computations in this section were performed using the CutFEM library in the FEniCs software package which is available from <https://fenicsproject.org/>.

The first case to consider is the case where $k_1 = k_2 = k$. Figures 4.12 and 4.13 show that both methods reach optimal convergence in the low wave number regime. The stabilized method seems to reach the asymptotic regime for $k = 50$ faster than its non stabilized counterpart. Interestingly, the non stabilized case performs rather well although I do not have an analysis to back this up. In this regime the unfitted case achieves optimal convergence for piecewise linear elements.

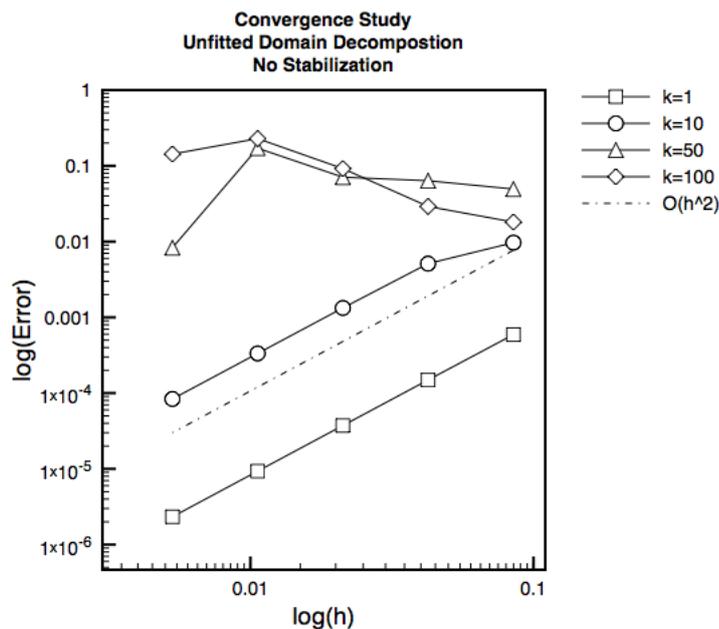


Figure 4.12: Convergence study for an unfitted domain decomposition method with no stabilization

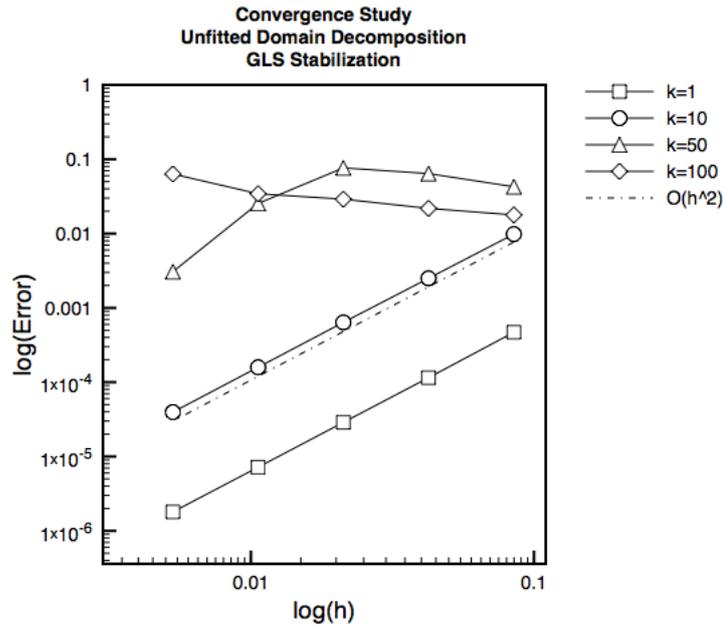


Figure 4.13: Convergence study for an unfitted domain decomposition method with GLS/CIP stabilization

In the case when $k_1 \neq k_2$ Figure 4.14 shows that the unfitted domain decomposition method with GLS stabilization still reaches optimal convergence. An unexplained observation is that the Nitsche’s method seems to achieve super convergence in this regime. It has been observed that the penalty-free Nitsche’s method obtains super convergence of $O(h^{1/2})$ in [10]. However, it is still an unexplained phenomenon.

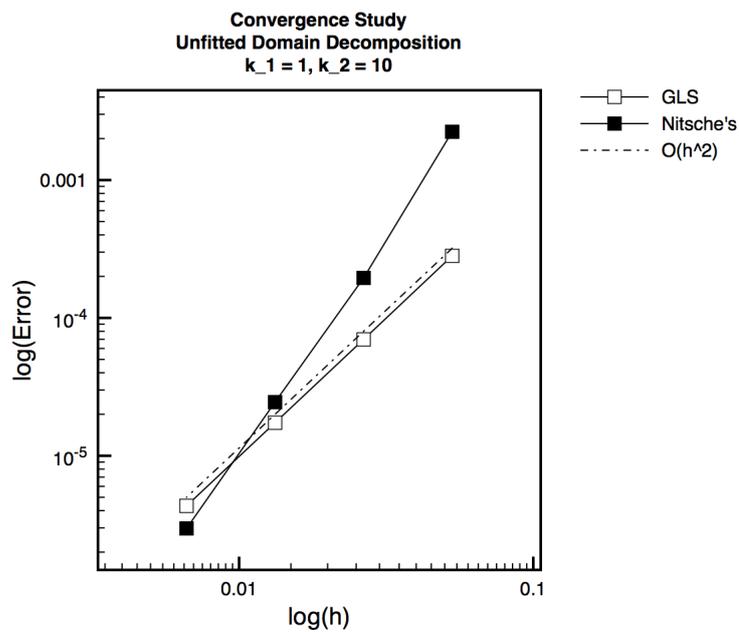


Figure 4.14: Convergence comparison between a stabilized and unstabilized unfitted domain decomposition method coupling domains with different wave numbers

Chapter 5

Conclusion

In conclusion, I have introduced an abstract analysis with which to analyse a class of residual-based stabilized FEMs for the solution of Helmholtz problems. This analysis was then extended to consider the coupling of multiple Helmholtz equations which resulted in a new geometrically unfitted domain decomposition method capable of solving Helmholtz equation with an embedded interface.

In Chapter 2 new stabilized methods were introduced and shown to fit into the abstract framework for an appropriate choice of stabilization parameters. The methods were shown to be optimally convergent in the asymptotic regime and improved error estimates were also presented when considering piecewise linear elements. Numerical evidence suggested that the stabilized methods showed some promise for reducing the pollution effect for a couple of specific problems for an appropriate choice of stabilization parameter. The reason for this improvement is an open question and in order to investigate this, at least in the case of structured meshes, one would need to perform a detailed 2 dimensional dispersion analysis. The reduction in pollution is problem specific and evidence does not suggest that these methods can be considered pollution free.

In Chapter 3 the stabilized methods are extended to handle generalized boundary conditions using a modified version of the technique of Nitsche [47] to weakly impose essential boundary conditions. The stabilized methods are seen to be absolutely stable when considering sound hard and sound soft interior problems and respect the same a priori error estimates presented in Chapter 2. Numerical evidence was presented to back up the theoretical results and the same reduction of pollution was observed. In the second part of Chapter 3 a new fictitious domain method using cut elements was proposed for the solution of Helmholtz equation. This method was shown to be optimally convergent but did not show the same potential for reducing the pollution effect as the methods proposed previously. To the author's knowledge this is the first fictitious domain method proposed for Helmholtz, in a finite element setting, that has been shown to be stable under the condition $kh < C$.

Finally, in Chapter 4, I expand the abstract framework introduced in Chapter 2 to include the case of multi-domain coupling. In the first part of the chapter the modified Nitsche's method introduced in

Chapter 3 was used to weakly impose continuity across an interface between multiple domains where the mesh was assumed to fit the interface exactly. The method was shown to enter the mathematical framework and optimal a priori error estimates were obtained. Surprisingly, numerical evidence was presented which suggested that the new method retained the ability of the stabilized methods to reduce numerical pollution for certain problems. In the second part of Chapter 4 a new geometrically unfitted domain decomposition method using cut elements was presented which was shown to be optimally convergent and stable under the condition $\max_j \sqrt{\omega^2 \rho_j} h < C$. Unsurprisingly, the reduction in numerical pollution was not present when considering cut elements.

The methods presented in this report have shown potential for solving the Helmholtz equation and have culminated in the introduction of two methods that are designed to reduce the dependence between the geometric description of the PDE and its discretization. Possible future work would be to test the potential of these methods in a practical setting where the interface is evolving. Another interesting topic would be to explore the potential of extending these methods to the elastic Helmholtz equation.

Appendix A

Appendix

A Useful Inequalities

I also introduce the following inequalities, proofs of which can be found in [30] which I will be using repeatedly in this paper.

Lemma A.1 (Arithmetic-Geometric Inequality). *Let $x, y \in \mathbb{R}$ and $\epsilon > 0$. Then*

$$\frac{x^2}{\epsilon} + \epsilon y^2 \geq 2xy. \quad (\text{A:1})$$

The proof for this is relatively simple

$$\left(\frac{x}{\sqrt{\epsilon}} - y\sqrt{\epsilon}\right)^2 \geq 0 \quad (\text{A:2})$$

$$\frac{x^2}{\epsilon} + \epsilon y^2 - 2xy \geq 0. \quad (\text{A:3})$$

Lemma A.2 (Trace Inequality). *Let τ be a finite element in \mathbb{R}^d with diameter h_τ . Then there exists a constant C_T such that*

$$\|v\|_{0,\partial\tau} \leq C_T(h_\tau^{-1/2}\|v\|_{0,\tau} + h_\tau^{1/2}\|\nabla v\|_{0,\tau}), \quad (\text{A:4})$$

for all $v \in H^1(\tau)$.

Proof. A proof can be found in [30] □

Lemma A.3 (Inverse Inequality). *Let $v \in V_h$. Then*

$$\|\nabla v\|_{0,\tau} \leq C_I h_\tau^{-1} \|v\|_{0,\tau}. \quad (\text{A:5})$$

Proof. A proof can be found in [30] □

Definition A.1 (Fractional Sobolev spaces). For $0 < s < 1$ and $1 \leq p < +\infty$, the so-called Sobolev

space with fractional exponent is defined as

$$W^{s,p}(\Omega) = \left\{ u \in L^p(\Omega) : \frac{|u(x) - u(y)|}{|x - y|^{s + \frac{d}{p}}} \in L^p(\Omega \times \Omega) \right\}. \quad (\text{A:6})$$

Furthermore, when $s > 1$ is not an integer, letting $\sigma = s - [s]$ where $[s]$ is the integer part of s , $W^{s,p}(\Omega)$ is defined as

$$W^{s,p}(\Omega) = \{u \in W^{[s],p}(\Omega) : \partial^\alpha u \in W^{\sigma,p}(\Omega), \forall \alpha, |\alpha| = [s]\}. \quad (\text{A:7})$$

Definition A.2 (Star-shaped domain). Let $\Omega \subset \mathbb{R}^N$ be a bounded Lipschitz domain. Ω is said to be star-shaped if there exists a real constant $C > 0$ and a point $x_0 \in \Omega$ such that

$$(x - x_0) \cdot n > C \forall x \in \partial\Omega$$

where n denotes the outward facing normal with respect to Ω . Ω is said to be star-shaped with respect to the set B if, for all $x \in \Omega$, the closed convex hull of $\{x\} \cup B$ is a subset of Ω

Definition A.3 (Shape regularity). A family of meshes $\{\mathcal{T}_h\}_{h>0}$ is said to be shape-regular if there exists σ_0 such that

$$\forall h, \forall \tau \in \mathcal{T}_h, \quad \sigma_\tau = \frac{h_\tau}{\rho_\tau} \leq \sigma_0$$

where ρ_τ is the diameter of the largest ball B_τ contained in τ such that τ is star-shaped with respect to B_τ .

Definition A.4 (Quasi-uniformity). A family of meshes $\{\mathcal{T}_h\}_{h>0}$ is said to be quasi-uniform if and only if it is shape-regular and there is C such that

$$\forall h, \forall \tau \in \mathcal{T}_h, \quad h_\tau \geq Ch$$

Definition A.5 (L^2 Orthogonal Projection). Let $(\cdot, \cdot)_{0,\Omega}$ denote the $L^2(\Omega)$ scalar product with associated norm $\|\cdot\|_{0,\Omega}$. Let V be a closed subspace of $L^2(\Omega)$. The orthogonal projection from $L^2(\Omega)$ to V is defined to be the operator $\pi_h : L^2(\Omega) \mapsto V$ such that

$$(\pi_h u, v) = (u, v) \forall v \in V \quad (\text{A:8})$$

Lemma A.4 (Interpolation Estimate for the standard Lagrange Interpolant). Let $\mathcal{I}_h : C^0(\bar{\tau}) \mapsto V(\tau)$ be the standard Lagrange interpolant and $u \in H^{s+1}(\tau)$. Then

$$\|u - \mathcal{I}_h u\|_{0,\tau} + \sum_{j=1}^s h^j |u - \mathcal{I}_h u|_{j,\tau} \leq h^{s+1} |u|_{s+1,\tau} \quad (\text{A:9})$$

Proof. A proof can be found in [30] □

Lemma A.5 (Interpolation Estimate for the standard L^2 Orthogonal Projection). *Let $\pi_h : L^2(\Omega) \mapsto V(\Omega)$ be the standard L^2 orthogonal projection and $u \in H^{s+1}(\Omega)$. Then*

$$\|u - \pi_h u\|_{0,\Omega} + \sum_{j=1}^s \left(\sum_{\tau \in \mathcal{T}_h} h^j |u - \pi_h u|_{j,\tau}^2 \right)^{1/2} \leq h^{s+1} |u|_{s+1,\tau} \quad (\text{A:10})$$

Proof. A proof can be found in [30] □

Lemma A.6 (Addition of complex inner products). *Let $A, B \in \mathbb{C}$ and $\langle \cdot, \cdot \rangle$ denote the complex L^2 inner product. It holds that*

$$\langle A, B \rangle + \langle B, A \rangle = 2\text{Re}[\langle A, B \rangle] \quad (\text{A:11})$$

Similarly

$$\langle A, B \rangle - \langle B, A \rangle = 2i\text{Im}[\langle A, B \rangle] \quad (\text{A:12})$$

Proof.

$$\begin{aligned} & \langle A, B \rangle + \langle B, A \rangle \\ &= \int (A_R + A_I i)(B_R - B_I i) + \int (B_R + B_I i)(A_R - A_I i) \\ &= \int (A_R B_R + A_I B_R i - A_R B_I i + A_I B_I + A_R B_R - A_I B_R i + A_R B_I i + A_I B_I) \\ &= \int 2(A_R B_R + A_I B_I) \end{aligned}$$

and

$$\begin{aligned} & \langle A, B \rangle - \langle B, A \rangle \\ &= \int (A_R + A_I i)(B_R - B_I i) - \int (B_R + B_I i)(A_R - A_I i) \\ &= \int (A_R B_R + A_I B_R i - A_R B_I i + A_I B_I - A_R B_R + A_I B_R i - A_R B_I i - A_I B_I) \\ &= \int 2i(A_I B_R - A_R B_I) \end{aligned}$$

□

Lemma A.7 (Parallelogram Law). *Let V be a complex inner product space, then it holds that $\forall x, y \in V$*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2). \quad (\text{A:13})$$

Proof.

$$\begin{aligned}
\|x + y\|^2 + \|x - y\|^2 &= \langle x + y, x + y \rangle + \langle x - y, x - y \rangle \\
&= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle + \langle x, x \rangle - \langle x, y \rangle - \langle y, x \rangle + \langle y, y \rangle \\
&= 2(\langle x, x \rangle + \langle y, y \rangle) \\
&= 2(\|x\|^2 + \|y\|^2).
\end{aligned}$$

□

Lemma A.8 (Polarisation Identity). *Let V be a complex inner product space, then it holds that $\forall x, y \in V$*

$$\langle x, y \rangle = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2). \quad (\text{A:14})$$

Proof. Expanding each term on the RHS shows that

$$\begin{aligned}
\|x + y\|^2 &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle, \\
-\|x - y\|^2 &= -\langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle - \langle y, y \rangle, \\
i\|x + iy\|^2 &= i\langle x, x \rangle + \langle x, y \rangle - \langle y, x \rangle + i\langle y, y \rangle, \\
-i\|x - iy\|^2 &= -i\langle x, x \rangle + \langle x, y \rangle - \langle y, x \rangle - i\langle y, y \rangle.
\end{aligned}$$

Summing each of these components gives

$$\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2 = 4\langle x, y \rangle.$$

The proof is concluded by dividing both sides by 4.

□

Acknowledgment The author received funding from the UK research council grant EPSRC EP/J002313/1.

Bibliography

- [1] M. AINSWORTH, *Discrete dispersion relation for hp-version finite element approximation at high wave number*, SIAM J. Numer. Anal., 42 (2004), pp. 553–575.
- [2] D.N. ARNOLD, F. BREZZI, B. COCKBURN, L.D. MARINI, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal. 39 (2002) 1749-1779
- [3] C. ATHANASIADIS AND I.G. STRATIS, *On some elliptic transmission problems*, Ann. Pol. Math. 63 (1996), 137-154
- [4] I. BABUŠKA AND A.K. AZIZ, *The mathematical foundations of the finite element method*, In: A.K. Aziz (ed.), *The mathematical foundations of the finite element method with applications to partial differential equations*, Academic Press, New York, 1972, 5-359.
- [5] I.M. BABUŠKA AND S.A. SAUTER, *Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?*, SIAM Rev., 42 (2000), pp. 451–484.
- [6] I. BABUŠKA AND M. ZLÁMAL, *Nonconforming elements in the finite element method with penalty*, SIAM J. Numer. Anal., 10 (1973), pp. 863–875.
- [7] J. BAIGES AND R. CODINA, *A variational multiscale method with subscales on the element boundaries for the Helmholtz equation*, Internat. J. Numer. Methods Engrg., 93 (2013), pp. 664–684.
- [8] R. BECKER, P. HANSBO AND R. STENBERG, *A Finite Element Method for Domain Decomposition with non-matching grids*, ESAIM: Math. Model. Num., 37 (2003) 209-225.
- [9] J.D. BENAMOU AND B. DESPRÈS, *A Domain Decomposition Method for the Helmholtz equation and related Optimal Control Problems*, J. Comput. Phys. 136 (1997), no. 1, 6882.
- [10] T. BOIVEAU, *Fitted and unfitted domain decomposition using penalty free Nitsche method for the Poisson problem with discontinuous material parameters*, arXiv preprint: <https://arxiv.org/abs/1509.00542v1>
- [11] E. BURMAN, *A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty*, SIAM J. Numer. Anal., 43 (2005), pp. 2012–2033.

- [12] E. BURMAN, *Ghost Penalty*, C. R. Math. Acad. Sci. Paris, 348 (2010), pp. 1217–1220.
- [13] E. BURMAN, S. CLAUS, P. HANSBO, M.G. LARSON, A. MASSING, *CutFEM: Discretizing geometry and partial differential equations*, Internat. J. Numer. Methods Engrg. 104 (2015) 472-501.
- [14] E. BURMAN, S. CLAUS AND A. MASSING, *A stabilized cut finite element method for the three field Stokes problem*, SIAM J. Sci. Comput. ,37 (2015), No. 4, pp. A1705–A1726.
- [15] E. BURMAN AND A. ERN, *Continuous Interior Penalty hp-Finite Element Methods For Transport Operators*, Math. Comp., 76 (2007), pp. 1119–1140.
- [16] E. BURMAN, M.A. FERNÁNDEZ AND P. HANSBO, *Continuous Interior Penalty Finite Element Method for Oseen's Equations*, SIAM journal on numerical analysis, 44 (2006), pp.1248-1274.
- [17] E. BURMAN, J. GUZMÁN, M.A. SÁNCHEZ AND M. SARKIS, *Robust flux error estimation of Nitsche's method for high contrast interface problems*, IMA J. Numer. Anal., (2017) doi: <https://doi.org/10.1093/imanum/drx017>.
- [18] E. BURMAN AND P. HANSBO, *Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 1437–1453.
- [19] E. BURMAN AND P. HANSBO, *Edge stabilization for the generalized Stokes problem: a continuous interior penalty method.*, Comput. Methods Appl. Mech. Engrg., 195 (2006), no. 19-22, 23932410.
- [20] E. BURMAN AND P. HANSBO, *Fictitious domain finite element methods using cut elements: I. A stabilized Lagrange multiplier method*, Comput. Methods Appl. Mech. Engrg., 199(2010):41-44, pp. 2680-2686
- [21] E. BURMAN AND P. HANSBO, *Fictitious domain finite element methods using cut elements: II. A stabilized Nitsche Method*, Applied Numerical Mathematics., 62 (2012), pp. 328–341.
- [22] E. BURMAN, H. WU AND L. ZHU, *Linear continuous interior penalty finite element method for Helmholtz equation with high wave number: one-dimensional analysis*, Numer. Methods Partial Differential Equations, 32 (2016), pp. 1378–1410.
- [23] C.-C. CHU, I.G. GRAHAM AND T.-Y. HOU, *A new multiscale finite element method for high-contrast Elliptic interface problems*, Math. Comp., 79(272):1915-1955, (2010).
- [24] S. COURT, M. FOURNIÉ AND A. LOZINSKI, *A fictitious domain approach for the Stokes problem based on the extended finite element method.*, Internat. J. Numer. Methods Fluids 74 (2014), no. 2, 7399.

- [25] P. CUMMINGS AND X. FENG, *Sharp regularity coefficient estimates for complex-valued acoustic and elastic Helmholtz equations*, *Mathematical Models and Methods in Applied Sciences*, 16 (2006), pp. 139–160
- [26] Q.V. DINH, R. GLOWINSKI, J. HE, V. KWOCK, T.W. PAN AND J. PERIAUX, *Lagrange multiplier approach to fictitious domain methods: Application to fluid dynamics and electromagnetics*, *C. R. Math. Acad. Sci. Paris*, 348 (2010), pp. 1217–1220.
- [27] J. DOUGLAS JR AND T. DUPONT, *Interior Penalty Procedures for Elliptic and Parabolic Galerkin methods*, *Lecture Notes in Phys.* 58, Springer-Verlag, Berlin, 1976.
- [28] Y. DU, H. WU, *Preasymptotic error analysis of higher order FEM and CIP-FEM for Helmholtz equation with high wave number*, *SIAM J. Numer. Anal.*, 53 (2015), pp. 782–804.
- [29] G. ENGEL, K. GARIKIPATI, T. J. R. HUGHES, M. G. LARSON, L. MAZZEI, AND R. L. TAYLOR, *Continuous/discontinuous finite element approximations of fourth-order elliptic problems in structural and continuum mechanics with applications to thin beams and plates, and strain gradient elasticity*, *Comput. Methods Appl. Mech. Engrg.*, 191 (2002), pp. 3669–3750.
- [30] A. ERN AND J.-L. GUERMOND, *Theory and Practice of Finite Elements*, volume 159 of, *Applied Mathematical Sciences*. Springer-Verlag, New York, NY, (2004).
- [31] F. FENG AND H. WU, *hp-discontinuous Galerkin methods for the Helmholtz equation with large wave number*, *Math. Comp.*, 80 (2011), pp. 1997–2024.
- [32] P.J. FREY AND P.L. GEORGE, *Mesh Generation: Application to Finite Elements*, ISTE, London, (2008)
- [33] A. HANSBO AND P. HANSBO, *An unfitted finite element method, based on Nitsche’s method, for elliptic interface problems*, *Comput. Methods Appl. Mech. Engrg.*, 191 (2002), pp. 5537–5552.
- [34] I. HARARI AND T.J.R. HUGHES, *Finite element method for the Helmholtz equation in an exterior domain: Model problems*, *Comput. Methods Appl. Mech. Engrg.*, 87 (1991), pp. 59–96.
- [35] J. HASLINGER AND Y. RENARD, *A new fictitious domain approach inspired by the Extended Finite Element Method*, *SIAM Journal on Numerical Analysis*. 47 (2009), pp. 1474–1499
- [36] U. HETMANIUK, *Stability Estimates for a Class of Helmholtz Problems*, *Commun. Math. Sci.*, Vol. 5, No. 3, (2007), pp. 665–678.
- [37] Z. HUANG AND X. YANG, *Tailored finite cell method for solving Helmholtz equation in Layered heterogeneous medium*, *Journal of Computational Mathematics*. Vol.30, No.4, (2012) pp. 381–391.

- [38] F. IHLENBURG AND I. BABUŠKA, *Finite element solution of the Helmholtz equation with high wave number. I. The h-version of the FEM*, *Comput. Math. Appl.*, 30 (1995), pp. 9–37.
- [39] ———, *Finite element solution of the Helmholtz equation with high wave number. II. The h-p version of the FEM*, *SIAM J. Numer. Anal.*, 34 (1997), pp. 315–358.
- [40] R. LEIS, *Initial Boundary Value Problems in Mathematical Physics*, Dover Publications, 2013
- [41] P.L. LIONS, *On the Schwarz alternating method III: a variant for non-overlapping subdomains*, Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, T.F. Chan, R. Glowinski, J. Periaux and O.B. Widlund Eds., SIAM (1989)
- [42] J.M. MELENK, *On Generalized Finite Element Methods*, Ph.D. Thesis, The University of Maryland, (1995).
- [43] J.M. MELENK AND S. SAUTER, *Wavenumber explicit convergence analysis for Galerkin discretizations of the Helmholtz equation*, *SIAM J. Numer. Anal.*, 49 (2011), pp. 1210–1243.
- [44] A. MOIOLA AND E.A. SPENCE, *Acoustic transmission problems: wavenumber-explicit bounds and resonance-free regions*, arXiv preprint: <http://arxiv.org/pdf/1702.00745v3>
- [45] L. MU, J. WANG, G. WEI, X. YE AND S. ZHAO, *Weak Galerkin methods for second order elliptic interface problems*, *J. Comput. Phys.*, 250 (2013), pp. 106–125.
- [46] J.C. NEDELEC, *Acoustic and electromagnetic equations: integral representations for harmonic problems, volume 144 of, Applied Mathematical Sciences*. Springer-Verlag, New York, NY, (2001).
- [47] J. NITSCHKE, *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind*, *Abh. Math. Sem. Univ. Hamburg*, 36 (1971), pp. 9–15.
- [48] A.A. OBERAI AND P.M. PINSKY, *A residual-based finite element method for the Helmholtz equation*, *Internat. J. Numer. Methods Engrg.*, 49 (2000), pp. 399–419.
- [49] A.H. SCHATZ, *An observation concerning Ritz–Galerkin methods with indefinite bilinear forms*, *Math. Comp.*, 28 (1974), pp. 959–962.
- [50] B.G. SMITH, B.L. VAUGHAN JR. AND D.L. CHOPP, *The Extended Finite Element Method for boundary layer problems in Biofilm growth*, *Comm. App. Math. and Comp. Sci.*, Vol. 2, No. 1 (2007)
- [51] E. STEIN, *Singular Integrals and Differentiability Properties of Functions.*, Princeton University Press (1970)

- [52] R. STENBERG, *Mortaring by a method of J. A. Nitsche*, in: S. Idelsohn, E. Oñate, E. Dvorkin (Eds.), *Computational Mechanics: New Trends and Applications*, CIMNE, Barcelona, Spain, 1998.
- [53] J. WLOKA, *Partial Differential Equations*, Cambridge University Press, 1987.
- [54] H. WU, *Pre-asymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. Part I: Linear version*, IMA J. Numer. Anal. 34 (2013), pp. 1266–1288
- [55] E.L. YEDEG, E. WADBRO, P. HANSBO, M. LARSON AND M. BERGGREN, *A Nitsche-type Method for Helmholtz Equation with an Embedded Acoustically Permeable Interface*, Comp. Meth. Appl. Mech. and Engrg., 304 (2016), pp. 479–500
- [56] K. YOSIDA, *Functional Analysis*, Springer-Verlag, Berlin, (1980).
- [57] L. ZHU AND H. WU, *Pre-asymptotic error analysis of CIP-FEM and FEM for Helmholtz equation with high wave number. Part II: hp version*, SIAM J. Numer. Anal., 51 (2013), pp. 1828–1852
- [58] Z. ZOU, W. AQUINO AND I. HARARI, *Nitsche's method for Helmholtz problems with embedded interfaces*, Internat. J. Numer. Methods Engrg., 110 no. 7 (2017), pp. 618–636