Locking-free discontinuous Galerkin methods for problems in elasticity, using linear and multilinear approximations

B. J. Grieshaber

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Mathematics and Applied Mathematics
UNIVERSITY OF CAPE TOWN

October 2013

Abstract

With interior penalty discontinuous Galerkin methods well established as locking-free for low-order triangular elements, and thus an effective alternative to the Standard Galerkin method for nearly incompressible materials, substantial numerical evidence in this work shows that this is not the case for quadrilateral elements. Direct comparisons to triangles illustrate the material dependence of three common interior penalty methods for bilinear quadrilaterals, with locking and other manifestations of poor approximations in the near-incompressible regime.

To understand this discrepancy with a view to providing a remedy for the problem, an existing convergence analysis for triangles is looked at for possible extension to the case of quadrilaterals. This highlights the need for a suitable interpolant for the error-splitting approach of the proof. To rectify the problem manifesting in the numerical results, a modification to the formulation or elements themselves is necessary, and a preliminary analysis with bilinear elements, assuming the existence of a suitable interpolant with some basic properties, indicates two modifications as potential remedies: edge-term under-integration, and the use of linear rather than multilinear elements.

A portion of the thesis is devoted to establishing what the required properties of an interpolant would be for proving uniform convergence in the case of rectangular elements (with the proposed modifications taken into consideration), and designing a new interpolant which fits these specifications.

Each proposed modification is then considered in turn, looking at its impact on every stage of the error analysis, specifically coercivity of the bilinear form and consistency of the formulation. In the case of the use of under-integration, a further small modification is incorporated as this is necessary to maintain stability. Both under-integration and the use of linear approximations are ultimately proven to lead to uniform convergence for rectangular elements.

The analysis, also extended to three-dimensional domains, is supported by numerical results on a range of benchmark problems, illustrating the effectiveness of the two remedies proposed.

Acknowledgements

I would like to express much appreciation to those who have contributed to my undertaking and completing this thesis.

Particular thanks to my supervisor, Prof. Daya Reddy, for his guidance throughout this research, and for sharing from his wealth of knowledge and expertise. He has also taught me much in his approach to problem-solving. Dr Andrew McBride, a significant teacher in my early stages in the field of finite elements, has given valued insight and encouragement during this project, and been a great help in practical areas, especially in the technical aspects of a number of software applications. Jean-Paul Pelteret has been a wonderful support, both as a fellow-student with expertise in mechanics and programming, and as a friend.

Thanks to Olivia Goodhind for all her help, especially with paperwork, making the administrative aspects of studying very smooth indeed.

I greatly appreciate the financial contribution to this project from the National Research Foundation.

Those who have been a support, encouragement and inspiration in a personal capacity: my parents, apart from having taught me valuable diligence and perseverance from a young age, have been an anchor for me while I have been working on this thesis; my uncle and aunt, Alan and Irene Cawdry, have been my family base in Cape Town; friends who have been closely alongside me from the start of this project, and whom I truly appreciate, are Karabo Rajuili, Benjamin Herr and Cheryn Baier. Thanks to Michele Berry for the discussions and perspective on managing important aspects of academic life.

Finally and ultimately, thanks to God, who gives me everything I need for life and productivity.

Contents

Chapter 1

Introduction

The finite element method is well established as a method for solving boundary value problems approximately. Numerical implementations are generally supported by rigorous analyses of the method, certainly for linear problems, and for an increasingly wide range of nonlinear problems.

In the context of fluid and solid mechanics, and in particular problems for elastic materials, the Standard Galerkin (SG) finite element method, while performing very well for compressible materials, may exhibit the phenomenon known as "locking" for materials that are nearly incompressible, if low-order (linear or bilinear, or trilinear in three dimensions) elements are used. This pathological behaviour results from the too-severe constraint placed on the solution by the incompressibility condition. The problem manifests itself particularly in the case of bending-dominated problems.

The problem may be circumvented by the use of high-order elements. Low-order elements remain an attractive option, though, and for this reason various alternatives to the SG method have been studied, and shown to be effective in remedying the locking problem when the lowest-order approximations for the displacement are used. One class is mixed methods, where more than one variable is directly solved for (see, for example, [11]). Some successful combinations, given the correct choice of element type in each variable, are pressure-displacement, stress-displacement (the Hellinger-Reissner formulation), and stress-strain-displacement (known commonly as the Hu-Washizu formulation). Related to the pressure-displacement mixed method is the method of selective reduced integra-

tion (SRI), also effective in producing locking-free results, and which can be shown to be equivalent to a mixed method. Finally, discontinuous Galerkin (DG) methods, specifically the range of interior penalty (IP) DG methods, have been used effectively with low-order elements, within a certain scope. It is the broadening of this scope that is the main concern of this thesis.

Within the general class of DG methods for linear elasticity are mixed and primal formulations. In the arena of mixed DG methods, Hansbo and Larson [24] present a pressure-displacement formulation and show that it produces optimal convergence rates, for both nearly and exactly incompressible materials. Another mixed DG method that is successful in circumventing locking is the Local Discontinuous Galerkin (LDG) method of Cockburn et al. [17], a strain-displacement-pressure formulation that can also accommodate both compressible and exactly incompressible elastic materials. The authors have shown optimal uniform convergence for simplices in three dimensions for both cases, and using a post-processing procedure obtain point-wise incompressible approximations for fully incompressible materials. For *hp*-adaptive mixed pressure-displacement DG methods, Wihler and Wirz [43] present an error analysis in three-dimensional domains, taking singularities into account. They use hexahedral elements of various polynomial orders, including trilinear approximations, and show the methods to be robust. Locking-free *a posteriori* estimates have been developed by Houston et al. [27], for both triangles and affine quadrilaterals, including for low-order polynomials.

Most DG mixed methods that have been proven to be robust for near-incompressible elasticity allow for meshes of quadrilateral or hexahedral elements, as well as simplicial. In contrast, DG primal formulations have been established as having optimal performance independent of material parameters for meshes of triangular elements, though not for meshes of quadrilateral elements. The distinction between one-field and mixed methods here is significant, primarily because even with continuous displacements there is a distinction in performance between one-field and mixed methods: that is, mixed methods may alleviate locking when it occurs in primal methods.

The study of one-field DG methods for linear elasticity problems can be divided into two categories: those investigations analysing the specific formulations broadly, without attention given to limiting values of material or other parameters; and those which do consider the effect of material properties, and are thereby able, for example, to predict

the presence or absence of the locking phenomenon for nearly incompressible materials.

Rivière and Wheeler [36] extended an IP method, NIPG (Nonsymmetric Interior Penalty Galerkin) method, which had recently been introduced for scalar elliptic problems by Rivière et al. [37], to the problem of elasticity. The authors have derived optimal estimates with respect to both mesh size and degree of approximation, in two and three dimensions. Lew et al. [28] used an extension of the formulation of Brezzi et al. [12], for the Poisson equation - in turn based on the method of Bassi and Rebay [6] for the compressible Navier-Stokes equations - and obtained optimal estimates with respect to mesh size. Both of these analyses consider the convergence of the error without specific reference to material parameters.

In an analysis taking material parameters into account, Wihler [41] proved the NIPG method locking-free for nearly incompressible materials when a mesh of simplicial elements is used, with low-order (linear) approximations, for polygonal domains in two dimensions. He showed that the method converges independently of the compressibility parameter, and that with the use of graded meshes is robust even on non-convex domains. Similarly, Hansbo and Larson [24] presented an extended and adapted SIPG (Symmetric Interior Penalty Galerkin) method - originally defined by Wheeler [40] for scalar elliptic equations - and showed that it is locking-free for nearly incompressible materials, on a two-dimensional domain using a mesh of simplices, with linear and higher-order approximations. They showed also that the adaptation they had made, viz. the addition of a second stabilization term, is a necessary requirement for stability of the method. Shortly thereafter, the same authors presented a similar method using a different stabilization, showing its relationship to a nonconforming method with a stabilized Crouzeix-Raviart element, and in particular proving that it is uniformly convergent in the incompressible limit for linear and higher-order simplices in two and three dimensions [25]. *A posteriori* analyses for DG methods on meshes of triangles have also been performed in [42] and [13], providing bounds that are robust with respect to the compressibility parameter.

Also considering the variation of material parameters in the context of low-order approximations was numerical work by Liu et al. [29], in three dimensions, showing the underperformance of the SG method and the superiority of the three IP methods (NIPG, SIPG and the Incomplete Interior Penalty Galerkin method, or IIPG, originally described by Dawson et al. [18] for flow problems), as well as of the OBB method, a penalty-free

version of NIPG, on a specific benchmark problem.

While all of the analyses for one-field DG show that their methods are locking-free, the analyses are restricted to the case of triangular elements, not merely in specified scope, but inherently in the chosen tools of proof: for example, based on the error-splitting technique, using an interpolant defined on a triangle.

In contrast, Liu et al. use hexahedral elements, though without an accompanying analysis. Moreover, while the benchmark problem does show an under-displacement due to the SG method, for near-incompressible materials, which the DG methods do not exhibit, the conditions are not those leading to the severest form of locking. With the focus of their investigation being on other aspects of the methods, such as variation in penalty parameters, convergence data for successive mesh refinements is also not given. Therefore, while the results are positive, questions are left open about the scope of the superior performance of the IP methods.

Since the use of quadrilateral elements is desirable both for the advantages that that type of mesh offers and for the increased variation allowed for by bilinear elements, an investigation of this option was deemed valuable, with a specific focus on the three IP methods.

Contrary to initial expectations, numerical experiments showed that bilinear quadrilateral elements, whilst performing much like triangular elements for highly compressible materials, perform quite differently from them for the nearly incompressible case. For all three IP methods, bilinear elements yield poor approximations for nearly incompressible materials, with recognisable locking-type behaviour, and they produce correlating poor convergence rates and errors, largely comparable to those of the SG method.

Using as a starting point the approach of Wihler [41] in his analysis for NIPG with triangular elements - although considerable deviation was required - an analysis for bilinear elements has been performed, incorporating all three of the IP methods. This analysis reveals where material properties contribute in a potentially problematic way to the error bound, showing that poor performance for nearly incompressible materials can, in fact, reasonably be expected when bilinear elements are used.

Two remedies are proposed and analysed for under-performing bilinear elements. The first is under-integration - on a minimal number of specifically selected terms. A care-

ful consideration of the analysis for bilinear elements indicates that the undesirable parameter-dependence enters at certain terms defined on element edges. These terms contain constant factors afforded by the linear approximations of triangular elements which play a role in avoiding this dependence, and using bilinear elements but with under-integration of those terms mimics this property of triangles. The complete analysis is presented, demonstrating independence of the potentially problematic parameter. Numerical results illustrate the effectiveness of this remedy.

An alternative remedy proposed here is to use linear approximations on the quadrilateral elements. This was done by Rivière and Wheeler [36] with the NIPG method, but not analysed for parameter-independence. The other IP methods have not been studied for this scenario. Immediately, the use of linear approximations on quadrilaterals might be expected to reproduce certain potentially beneficial properties of linear triangular elements. A complete analysis is again presented, with supporting numerical results, showing that the NIPG method with linear elements is locking-free, while the other two IP methods are locking-free if linear elements are used in conjunction with a minimal application of under-integration.

Finally, an essential component in the analyses, as is common in proving error estimates, is the choice of a suitable interpolation operator to facilitate the necessary mathematical manipulations. Often an interpolant from a nonconforming method is used (as in [41], for example) as it carries useful properties of the nonconforming elements. (This is also natural, since DG is related in quality to nonconforming methods.) In the current case, an interpolant with suitable properties was not found amongst nonconforming elements, and therefore a new interpolant has been constructed for the task. It is most closely related to the elements of Douglas et al. [22].

Chapter 2 contains details of the boundary value problem of linear elasticity in its continuous setting, a description of the DG framework, and the definition of the IP methods. In Chapter 3 the analysis is introduced, covering the fundamentals for the bilinear elements, and motivating the need for remedies to the methods. Chapter 4 deals with the new interpolant. In Chapter 5, the incorporation of under-integration and the use of linear elements are both analysed. An extension of the analyses to three-dimensional domains is provided in Chapter 6. Chapter 7 contains the numerical results, and conclusions are given in Chapter 8.

Chapter 2

Preliminaries

In this chapter, the boundary value problem of linear elasticity is defined and explained. The framework used for DG methods is then introduced, with various general and notational definitions, and a definition of the DG norm that will be used. Finally a statement of the IP formulation to be used in this work is given. Also included are historical notes on the DG method in general, and on the development of each of the IP methods specifically .

2.1 The boundary value problem of linear elasticity

2.1.1 Geometry and governing equations

Let a homogeneous, isotropic, linear elastic body occupy the bounded domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$, with the Lipschitz boundary $\partial\Omega$ consisting of a Dirichlet portion, Γ_D, of positive measure, and a Neumann portion, Γ_N, such that

$$\Gamma_D \cap \Gamma_N = \emptyset \text{ and } \Gamma_D \cup \Gamma_N = \partial\Omega,$$

and with outward unit normal $\mathbf{n}$.

A body force $\mathbf{f} \in \left[L^2(\Omega)\right]^d$ is applied, with prescribed displacement $\mathbf{g} \in \left[H^1(\Gamma_D)\right]^d$ on Γ_D and prescribed traction $\mathbf{h} \in \left[L^2(\Gamma_N)\right]^d$ on Γ_N. The resultant displacement is $\mathbf{u}$, and

the strain $\boldsymbol{\varepsilon}$ is expressed as a tensor defined in index notation as

$$\boldsymbol{\varepsilon}(\mathbf{u})_{ij} := \tfrac{1}{2}\left(u_{i,j} + u_{j,i}\right), \qquad 1 \leq i, j \leq d.$$

The stress $\boldsymbol{\sigma}(\mathbf{u})$ is related to the strain via the constitutive law

$$\boldsymbol{\sigma}(\mathbf{u}) := 2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda \mathrm{tr}\,\boldsymbol{\varepsilon}(\mathbf{u})\,\mathbf{1} = 2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\nabla\cdot\mathbf{u}\mathbf{1},$$

where λ and the shear modulus μ are known as the Lamé parameters, and $\mathbf{1}$ is the $d \times d$ identity tensor.

The governing equation of the system is

$$-\mathrm{div}\ \boldsymbol{\sigma}(\mathbf{u}) = \mathbf{f} \qquad \text{in } \Omega, \tag{2.1.1a}$$

which, with the boundary conditions

$$\mathbf{u} = \mathbf{g} \qquad \text{on } \Gamma_D, \tag{2.1.1b}$$

$$\boldsymbol{\sigma}(\mathbf{u})\,\mathbf{n} = \mathbf{h} \qquad \text{on } \Gamma_N, \tag{2.1.1c}$$

has a unique solution $\mathbf{u} \in \left[H^2(\Omega)\right]^d$.

2.1.2 Material parameters

The Lamé parameters λ and μ are assumed to be positive, and can be expressed in terms of the Young's modulus, E, and Poisson's ratio, ν, by

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)},$$
$$\mu = \frac{E}{2(1+\nu)}.$$

As $\nu \to \frac{1}{2}$, which corresponds to the incompressible limit, so $\lambda \to \infty$.

2.2 The DG framework

2.2.1 Background

The DG methods are a family of methods originating in 1971 in Nitsche's work [30], in which Dirichlet boundary conditions were weakly imposed through a penalty term. Weak

enforcement of some level of continuity between the discontinuous elements has been the defining characteristic of the methods from the start, whether through penalty terms such as in the early work by Douglas and Dupont [21] for elliptic and parabolic equations, or through other terms defined on inter-element edges, such as by Reed and Hill [34] for the neutron transport equation. There has been much development of DG methods since the 1970s, for elliptic, parabolic and hyperbolic equations, both linear and nonlinear (cf. [16]). The range of applications, using both primal and mixed formulations, includes fluid flow, elasticity, and plasticity, to name a few (see, for example, [6], [18], [39] and [20]). DG schemes for time-discretization have also been developed, such as in [1].

2.2.2 Discretization

The domain Ω is partitioned into a mesh of elements generically referred to as Ω_e, where $\mathcal{T}_h = \{\Omega_e\}$. It is assumed for the purpose of this work that the mesh is geometrically conforming (that is, it contains no hanging nodes), and that the elements are regular. Define $h_e := \text{diam}\,(\Omega_e)$, and $h := \max_{\Omega_e \in \mathcal{T}_h} (h_e)$. The outward unit normal of Ω_e is denoted by $\mathbf{n}_e$. Indices are changed if distinction between elements is necessary.

Here and in what follows, all of the definitions and notation are given for $d = 2$, but are equally applicable to $d = 3$ if "edge" is replaced with "face" in each instance.

Each element has a boundary $\partial\Omega_e$, consisting of edges E. Define $h_E := \text{diam}\,(E)$.

The union of all edges lying in the interior of the domain, rather than on the boundary, will be denoted by Γ_{int}. Define $\Gamma_{iD} := \Gamma_{int} \cup \Gamma_D$, and finally, $\Gamma := \Gamma_{int} \cup \Gamma_D \cup \Gamma_N$ is the union of all edges.

On an element boundary, define $\partial\Omega_e^{int} := \partial\Omega_e \cap \Gamma_{int}$, $\partial\Omega_e^{D} := \partial\Omega_e \cap \Gamma_D$, and $\partial\Omega_e^{iD} := \partial\Omega_e \cap \Gamma_{iD}$.

By abuse of notation, any symbol denoting a union of edges will also denote the corresponding set of edges.

2.2.3 Function spaces

The Sobolev space $H^m(\Omega)$, m a non-negative integer, has a norm defined by

$$\|v\|^2_{H^m(\Omega)} = \int_\Omega \sum_{|\alpha|\leq m} D^\alpha v(\boldsymbol{x})\, D^\alpha v(\boldsymbol{x})\, dx$$

and seminorm defined by

$$|v|^2_{H^m(\Omega)} = \int_\Omega \sum_{|\alpha|= m} D^\alpha v(\boldsymbol{x})\, D^\alpha v(\boldsymbol{x})\, dx,$$

with the conventional multi-index notation.

Use will also be made of the discrete Sobolev space

$$H^1(\mathcal{T}_h) := \left\{ v \in L^2(\Omega) : \quad v|_{\Omega_e} \in H^1(\Omega_e) \quad \forall\, \Omega_e \in \mathcal{T}_h \right\}.$$

2.2.4 Traces

The trace of a function $v \in H^1(\mathcal{T}_h)$ on any (boundary) edge $E \in \partial\Omega$ is single-valued, and will simply be denoted by v. On an (interior) edge $E \in \Gamma_{int}$, the trace of any $v \in H^1(\mathcal{T}_h)$ may be multivalued, as continuity between elements is not a constraint of that space. That is, evaluated on any edge E shared by elements Ω_e and Ω_f (denoted by Γ_{ef}), the trace of $v|_{\Omega_e}$ is v_e and of $v|_{\Omega_f}$ is v_f, where in general $v_e|_{\Gamma_{ef}} \neq v_f|_{\Gamma_{ef}}$.

2.2.5 Jumps and averages

Single-valued functions on Γ_{int} that connect neighbouring elements and are used in DG formulations to enforce weak continuity of some kind between the elements, are the jumps and averages of functions across edges. For a vector $\mathbf{v}$ and a tensor $\boldsymbol{\tau}$, on Γ_{ef}, the jumps are defined as

$$\begin{aligned} \llbracket \mathbf{v} \rrbracket &:= \mathbf{v}_e \otimes \mathbf{n}_e + \mathbf{v}_f \otimes \mathbf{n}_f && \text{(a tensor)}, \\ \llbracket \boldsymbol{\tau} \rrbracket &:= \boldsymbol{\tau}_e \mathbf{n}_e + \boldsymbol{\tau}_f \mathbf{n}_f && \text{(a vector)}, \\ \text{and} \qquad \llbracket \mathbf{v} \rrbracket &:= \mathbf{v}_e \cdot \mathbf{n}_e + \mathbf{v}_f \cdot \mathbf{n}_f && \text{(a scalar)}; \end{aligned}$$

and the averages are defined as

$$\{\!\{\mathbf{v}\}\!\} := \tfrac{1}{2}(\mathbf{v}_e + \mathbf{v}_f)$$
$$\text{and} \qquad \{\!\{\boldsymbol{\tau}\}\!\} := \tfrac{1}{2}(\boldsymbol{\tau}_e + \boldsymbol{\tau}_f).$$

The same notation used in reference to boundary edges is defined by

$$[\![\mathbf{v}]\!] = \mathbf{v} \otimes \mathbf{n}, \quad [\![\boldsymbol{\tau}]\!] = \boldsymbol{\tau}\mathbf{n},$$
$$[\![\mathbf{v}]\!] = \mathbf{v} \cdot \mathbf{n},$$
$$\{\!\{\mathbf{v}\}\!\} = \mathbf{v}, \quad \{\!\{\boldsymbol{\tau}\}\!\} = \boldsymbol{\tau}.$$

2.2.6 DG solution space

The DG solution space $V_h \subset H^1(\mathcal{T}_h)$ will be composed of functions that are element-wise polynomial, specifically linear or bilinear/trilinear.

With $\mathbb{P}_1(\Omega)$ the space of polynomials on Ω of maximum total degree one, and $\mathbb{Q}_1(\Omega)$ the space of polynomials on Ω with maximum degree one in each variable, let $\mathbb{V} = \mathbb{P}_1(\Omega_e)$ or $\mathbb{Q}_1(\Omega_e)$, to be specified in context, and define

$$V_h = \left[\mathbf{v} \in \left[L^2(\Omega)\right]^d : \ \mathbf{v}|_{\Omega_e} \in \mathbb{V}^d \ \ \forall\, \Omega_e \in \mathcal{T}_h\right]. \tag{2.2.2}$$

2.2.7 DG norm

The norm which will be used is that used by Wihler [41] (here with the assumption $\lambda > 0$):

$$\|\mathbf{u}\|_{\mathrm{DG}}^2 := \sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\mathbf{u})\|_{L^2(\Omega_e)}^2 + \tfrac{1}{2} \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \|[\![\mathbf{u}]\!]\|_{L^2(E)}^2 \tag{2.2.3}$$

This is equivalent to an element-wise H^1-norm, that is,

$$\left(\sum_{\Omega_e \in \mathcal{T}_h} \|\mathbf{u}\|_{H^1(\Omega_e)}^2 \right)^{1/2} \leq C \|\mathbf{u}\|_{\mathrm{DG}} \tag{2.2.4}$$

(cf. [41], [9]).

2.3 IP formulations

2.3.1 Development of the IP formulations

Douglas and Dupont [21] used interior penalties to enforce C^1 continuity weakly between conforming elements, for scalar elliptic and parabolic problems. This idea was used by Wheeler [40] in a collocation method, and later by Arnold [3] in a time-dependent formulation with discontinuous elements. These methods used bilinear forms that were symmetric in the part relating to the Laplace operator of the original equations, and used interior penalties, and so received the name "Symmetric Interior Penalty Galerkin", or SIPG. Hansbo and Larson [24], in studying SIPG for elasticity problems, showed that the method remained unstable even with the interior penalty term, and included a second stabilization term that penalised the jumps in the normal components only. This second penalisation was scaled by the Lamé parameter λ which is so significant in incompressible elasticity.

Oden et al. [31] removed the penalty and changed a sign in the bilinear form to make it non-symmetric, using it for the Poisson equation. This became known as the OBB formulation, and was later extended by two of the authors, Baumann and Oden, for the convection-diffusion equation [7]. Rivière et al. [37] combined the ideas of antisymmetry and interior penalty, to obtain the Nonsymmetric Interior Penalty Galerkin method (NIPG), essentially the OBB method with the reincorporation of a term penalising jumps. This was extended by two of the same authors in [36] for use in elasticity problems.

Dawson et al. [18] presented a third IP formulation, used in flow equations, omitting altogether the term that distinguished between the symmetric and nonsymmetric forms. This new method was called the Incomplete Interior Penalty Galerkin method (IIPG).

2.3.2 The formulation used in this work

All three IP methods will be studied in this work. Both stabilization terms will be included in the general formulation (either can be removed by setting the applicable

parameter to zero), and it will be established whether or not each is necessary in each of the methods.

With non-negative parameters k_μ and k_λ, and a switch θ to distinguish between methods, where $\theta = 1$ gives the NIPG method, $\theta = -1$ gives SIPG, and $\theta = 0$ gives IIPG, the complete general IP formulation is defined by the forms

$$\begin{aligned} a_h(\mathbf{u},\mathbf{v}) := & \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dx + \theta \sum_{E \in \Gamma_{iD}} \int_E [\![\mathbf{u}]\!] : \{\!\{\boldsymbol{\sigma}(\mathbf{v})\}\!\}\, ds \\ & - \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!]\, ds + k_\mu \mu \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \int_E [\![\mathbf{u}]\!] : [\![\mathbf{v}]\!]\, ds \\ & + k_\lambda \lambda \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \int_E [\![\mathbf{u}]\!]\, [\![\mathbf{v}]\!]\, ds \end{aligned} \tag{2.3.1}$$

and

$$\begin{aligned} l_h(\mathbf{v}) := & \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \mathbf{f} \cdot \mathbf{v}\, dx + \theta \sum_{E \in \Gamma_D} \int_E (\mathbf{g} \otimes \mathbf{n}) : \boldsymbol{\sigma}(\mathbf{v})\, ds + \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v}\, ds \\ & + k_\mu \mu \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E \mathbf{g} \cdot \mathbf{v}\, ds + k_\lambda \lambda \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E (\mathbf{g} \cdot \mathbf{n})(\mathbf{v} \cdot \mathbf{n})\, ds. \end{aligned} \tag{2.3.2}$$

A solution $\mathbf{u}_h \in V_h$ is required such that, for all $\mathbf{v} \in V_h$,

$$a_h(\mathbf{u}_h, \mathbf{v}) = l_h(\mathbf{v}). \tag{2.3.3}$$

Both Dirichlet and Neumann boundary conditions, (2.1.1b) and (2.1.1c), are imposed weakly through this formulation.

Chapter 3

Fundamentals of the analysis

This chapter explores and presents details of the distinction between using triangular elements with the IP methods, and using quadrilateral elements. Basic properties of the IP formulations, namely coercivity and consistency, are established first, and these are verified as being independent of the shape of the elements - aspects of these have been established previously by other authors, most significantly in the context of meshes of triangles. The initial stages of establishing an error bound for quadrilateral elements in particular are then dealt with, as previous work ([41], [24]) has been in the context of triangles, and the extension to meshes of quadrilaterals is not trivial. First, the approach to be used for bounding the error is described. A preliminary bound for bilinear elements is then reached, and the implications of this preliminary analysis are discussed with a view to establishing an improved bound through modifications.

3.1 Basic properties

3.1.1 Coercivity

The bilinear form associated with each IP method is considered separately, as the approach to establishing coercivity varies amongst the methods.

NIPG

Coercivity of the NIPG method with $k_\mu = 1, k_\lambda = 0$ has been shown by Wihler [41], and a specific coercivity constant obtained. This was done in the context of triangular elements but applies independently of element shape.

For the case of general stabilization parameters and also independent of the shape of the element, but without an explicit coercivity constant, $\forall\ \mathbf{v} \in \left[H^1\left(\mathcal{T}_h\right)\right]^2 \supset V_h$,

$$\begin{aligned} a_h\left(\mathbf{v},\mathbf{v}\right) &= 2\mu \sum_{\Omega_e\in\mathcal{T}_h} \left\|\boldsymbol{\varepsilon}\left(\mathbf{v}\right)\right\|^2_{L^2(\Omega_e)} + \lambda \sum_{\Omega_e\in\mathcal{T}_h} \left\|\nabla\cdot\mathbf{v}\right\|^2_{L^2(\Omega_e)} \\ &\qquad + k_\mu\mu \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|\left\llbracket\mathbf{v}\right\rrbracket\right\|^2_{L^2(E)} + k_\lambda\lambda \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|\left[\!\left[\mathbf{v}\right]\!\right]\right\|^2_{L^2(E)} \\ &\geq C\left(\sum_{\Omega_e\in\mathcal{T}_h} \left\|\boldsymbol{\varepsilon}\left(\mathbf{v}\right)\right\|^2_{L^2(\Omega_e)} + \tfrac{1}{2}\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|\left\llbracket\mathbf{v}\right\rrbracket\right\|^2_{L^2(E)}\right) \\ &= C\left\|\mathbf{v}\right\|^2_{\mathrm{DG}}. \end{aligned} \tag{3.1.1}$$

Here and elsewhere, C is a positive constant independent of h and λ.

The requirements on the stabilization parameters are $k_\mu > 0$, and $k_\lambda \geq 0$. That is, k_μ is essential for coercivity, and the second stabilization term is not required for the stability of NIPG.

SIPG

Hansbo and Larson [24] prove coercivity for the SIPG method for use on a mesh with triangular elements, showing that a λ-dependent stabilization term is necessary for stability of the method, which had in its original form ([40]) been used without this term. They show that, specifically, the parameters k_μ and k_λ must each be greater than a minimum value which can be calculated.

While their proof is also independent of the shape of the elements, the norm they use is different from the DG norm (2.2.3). The method of the proof detailed here with respect to the norm (2.2.3) is, nevertheless, similar to theirs.

The bilinear form is rewritten as a sum over all the elements in the partition, and interior edge terms are dealt with in this format by their contributions being split equally between the elements which share them.

That is,

$$\begin{aligned}
a_h(\mathbf{v},\mathbf{v}) &= 2\mu \sum_{\Omega_e\in\mathcal{T}_h} \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)} + \lambda \sum_{\Omega_e\in\mathcal{T}_h} \|\nabla\cdot\mathbf{v}\|^2_{L^2(\Omega_e)} - 2\sum_{E\in\Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{v})\}\!\} : \llbracket \mathbf{v} \rrbracket \, ds \\
&\quad + k_\mu \mu \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \|\llbracket \mathbf{v} \rrbracket\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|^2_{L^2(E)} \\
&= 2\mu \sum_{\Omega_e\in\mathcal{T}_h} \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)} + \lambda \sum_{\Omega_e\in\mathcal{T}_h} \|\nabla\cdot\mathbf{v}\|^2_{L^2(\Omega_e)} \\
&\quad - 2\left(\sum_{E\in\Gamma_{int}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{v})\}\!\} : \llbracket \mathbf{v} \rrbracket \, ds + \sum_{E\in\Gamma_D} \int_E \boldsymbol{\sigma}(\mathbf{v}) : (\mathbf{v}\otimes\mathbf{n}) \, ds \right) \\
&\quad + k_\mu \mu \left(\sum_{E\in\Gamma_{int}} \frac{1}{h_E} \|\llbracket \mathbf{v} \rrbracket\|^2_{L^2(E)} + \sum_{E\in\Gamma_D} \frac{1}{h_E} \|\mathbf{v}\otimes\mathbf{n}\|^2_{L^2(E)} \right) \\
&\quad + k_\lambda \lambda \left(\sum_{E\in\Gamma_{int}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|^2_{L^2(E)} + \sum_{E\in\Gamma_D} \frac{1}{h_E} \|\mathbf{v}\cdot\mathbf{n}\|^2_{L^2(E)} \right) \\
&= \sum_{\Omega_e\in\mathcal{T}_h} \Bigg(2\mu \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)} + \lambda \sum_{\Omega_e\in\mathcal{T}_h} \|\nabla\cdot\mathbf{v}\|^2_{L^2(\Omega_e)} \\
&\quad - \sum_{E\in\partial\Omega_e^{int}} \int_E \{\!\{2\mu\boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\nabla\cdot\mathbf{v}\mathbf{1}\}\!\} : \llbracket \mathbf{v} \rrbracket \, ds \\
&\quad - 2\sum_{E\in\partial\Omega_e^D} \int_E (2\mu\boldsymbol{\varepsilon}(\mathbf{v}) + \lambda\nabla\cdot\mathbf{v}\mathbf{1}) : (\mathbf{v}\otimes\mathbf{n}) \, ds \\
&\quad + \frac{k_\mu}{2}\mu \sum_{E\in\partial\Omega_e^{int}} \frac{1}{h_E} \|\llbracket \mathbf{v} \rrbracket\|^2_{L^2(E)} + k_\mu \mu \sum_{E\in\partial\Omega_e^D} \frac{1}{h_E} \|\mathbf{v}\otimes\mathbf{n}\|^2_{L^2(E)} \\
&\quad + \frac{k_\lambda}{2}\lambda \sum_{E\in\partial\Omega_e^{int}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E\in\partial\Omega_e^D} \frac{1}{h_E} \|\mathbf{v}\cdot\mathbf{n}\|^2_{L^2(E)} \Bigg).
\end{aligned} \tag{3.1.2}$$

Splitting this into terms involving μ and terms involving λ, write

$$a_h(\mathbf{v},\mathbf{v}) = T_\mu + T_\lambda, \tag{3.1.3}$$

where

$$T_\mu := 2\mu \sum_{\Omega_e \in \mathcal{T}_h} \left(\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)} - \sum_{E \in \partial\Omega_e^{int}} \int_E \{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\} : [\![\mathbf{v}]\!]\, ds \right.$$
$$- 2 \sum_{E \in \partial\Omega_e^D} \int_E \boldsymbol{\varepsilon}(\mathbf{v}) : (\mathbf{v} \otimes \mathbf{n})\, ds + \frac{k_\mu}{4} \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|^2_{L^2(E)}$$
$$\left. + \frac{k_\mu}{2} \sum_{E \in \partial\Omega_e^D} \frac{1}{h_E} \|(\mathbf{v} \otimes \mathbf{n})\|^2_{L^2(E)} \right),$$

$$T_\lambda := \lambda \sum_{\Omega_e \in \mathcal{T}_h} \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} - \sum_{E \in \partial\Omega_e^{int}} \int_E \{\!\{\nabla \cdot \mathbf{v}\}\!\} [\![\mathbf{v}]\!]\, ds \right.$$
$$- 2 \sum_{E \in \partial\Omega_e^D} \int_E (\nabla \cdot \mathbf{v})(\mathbf{v} \cdot \mathbf{n})\, ds + \frac{k_\lambda}{2} \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|^2_{L^2(E)}$$
$$\left. + k_\lambda \sum_{E \in \partial\Omega_e^D} \frac{1}{h_E} \|\mathbf{v} \cdot \mathbf{n}\|^2_{L^2(E)} \right).$$

From T_μ, on the interior edges,

$$\sum_{\Omega_e \in \mathcal{T}_h} \sum_{E \in \partial\Omega_e^{int}} \int_E \{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\} : [\![\mathbf{v}]\!]\, ds \leq \sum_{\Omega_e \in \mathcal{T}_h} \sum_{E \in \partial\Omega_e^{int}} \left| \int_E \{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\} : [\![\mathbf{v}]\!] \right| ds$$
$$\leq C \sum_{E \in \Gamma_{int}} \left| \int_E \{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\} : [\![\mathbf{v}]\!] \right| ds$$
$$\leq C \sum_{E \in \Gamma_{int}} \left\| h_E^{1/2} \{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\} \right\|_{L^2(E)} \left\| h_E^{-1/2} [\![\mathbf{v}]\!] \right\|_{L^2(E)}$$
$$\leq C \sum_{\Omega_e \in \mathcal{T}_h} \sum_{E \in \partial\Omega_e^{int}} \left\| h_E^{1/2} \boldsymbol{\varepsilon}(\mathbf{v}) \right\|_{L^2(E)} \left\| h_E^{-1/2} [\![\mathbf{v}]\!] \right\|_{L^2(E)}$$
$$\leq C \sum_{\Omega_e \in \mathcal{T}_h} \left\| h_E^{1/2} \boldsymbol{\varepsilon}(\mathbf{v}) \right\|_{L^2(\partial\Omega_e^{int})} \left\| h_E^{-1/2} [\![\mathbf{v}]\!] \right\|_{L^2(\partial\Omega_e^{int})},$$

and on the Dirichlet boundary,

$$\sum_{\Omega_e \in \mathcal{T}_h} \sum_{E \in \partial\Omega_e^D} \int_E \boldsymbol{\varepsilon}(\mathbf{v}) : (\mathbf{v} \otimes \mathbf{n})\, ds = \sum_{\Omega_e \in \mathcal{T}_h} \int_{\partial\Omega_e^D} \boldsymbol{\varepsilon}(\mathbf{v}) : (\mathbf{v} \otimes \mathbf{n})\, ds$$
$$\leq \sum_{\Omega_e \in \mathcal{T}_h} \left\| h_E^{1/2} \boldsymbol{\varepsilon}(\mathbf{v}) \right\|_{L^2(\partial\Omega_e^D)} \left\| h_E^{-1/2} \mathbf{v} \otimes \mathbf{n} \right\|_{L^2(\partial\Omega_e^D)}.$$

Together,

$$\begin{aligned}
&-\sum_{\Omega_e\in\mathcal{T}_h}\left(\sum_{E\in\partial\Omega_e^{int}}\int_E\{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\}:\,[\![\mathbf{v}]\!]\,ds+2\sum_{E\in\partial\Omega_e^{D}}\int_E\boldsymbol{\varepsilon}(\mathbf{v}):\,(\mathbf{v}\otimes\mathbf{n})\,ds\right)\\
&\geq -C\sum_{\Omega_e\in\mathcal{T}_h}\left\|h_E^{1/2}\boldsymbol{\varepsilon}(\mathbf{v})\right\|_{L^2(\partial\Omega_e^{iD})}\left\|h_E^{-1/2}[\![\mathbf{v}]\!]\right\|_{L^2(\partial\Omega_e^{iD})}\\
&\geq -C\sum_{\Omega_e\in\mathcal{T}_h}\|\boldsymbol{\varepsilon}(\mathbf{v})\|_{L^2(\Omega_e)}\left\|h_E^{-1/2}[\![\mathbf{v}]\!]\right\|_{L^2(\partial\Omega_e^{iD})}\\
&\geq -C\sum_{\Omega_e\in\mathcal{T}_h}\left(\epsilon_\mu\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)}+\frac{1}{\epsilon_\mu}\left\|h_E^{-1/2}[\![\mathbf{v}]\!]\right\|^2_{L^2(\partial\Omega_e^{iD})}\right) \qquad (3.1.4)
\end{aligned}$$

where $\epsilon_\mu>0$. Therefore

$$T_\mu\geq 2\mu\sum_{\Omega_e\in\mathcal{T}_h}\left[(1-C\epsilon_\mu)\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)}+\left(\frac{k_\mu}{4}-\frac{C}{\epsilon_\mu}\right)\left\|h_E^{-1/2}[\![\mathbf{v}]\!]\right\|^2_{L^2(\partial\Omega_e^{iD})}\right]. \qquad (3.1.5)$$

To set the coefficient of the first norm to a constant $m>0$, let $\epsilon_\mu=\frac{1}{C}(1-m)$, and restrict m to $0<m<1$ to ensure $\epsilon_\mu>0$.

Then

$$\begin{aligned}
&\frac{k_\mu}{4}-\frac{C}{\epsilon_\mu}\geq m\\
\iff\quad&\frac{k_\mu}{4}-\frac{C^2}{1-m}\geq m\\
\iff\quad&k_\mu\geq 4\left(m+\frac{C^2}{1-m}\right)
\end{aligned}$$

By this choice of k_μ,

$$\begin{aligned}
T_\mu&\geq 2\mu\, m\sum_{\Omega_e\in\mathcal{T}_h}\left(\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)}+\left\|h_E{}^{-1/2}[\![\mathbf{v}]\!]\right\|^2_{L^2(\partial\Omega_e^{iD})}\right)\\
&\geq C\left(\sum_{\Omega_e\in\mathcal{T}_h}\|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)}+1/2\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\|[\![\mathbf{v}]\!]\|^2_{L^2(E)}\right)\\
&=C\|\mathbf{v}\|^2_{\mathrm{DG}} \qquad (3.1.6)
\end{aligned}$$

All that is then required for coercivity is $T_\lambda\geq 0$. An identical procedure is followed to obtain

$$T_\lambda\geq\lambda\sum_{\Omega_e\in\mathcal{T}_h}\left[(1-\bar{C}\epsilon_\lambda)\|\nabla\cdot\mathbf{v}\|^2_{L^2(\Omega_e)}+\left(\frac{k_\lambda}{2}-\frac{\bar{C}}{\epsilon_\lambda}\right)\left\|h_E^{-1/2}[\![\![\mathbf{v}]\!]\!]\right\|^2_{L^2(\partial\Omega_e^{iD})}\right] \qquad (3.1.7)$$

for some $\epsilon_\lambda > 0$, where $\bar{C}$ is a positive constant independent of h and λ.

Set $\epsilon_\lambda = \frac{1}{\bar{C}}$, so that the first term vanishes. For a non-negative coefficient of the second norm in (3.1.7),

$$\begin{aligned} & \frac{k_\lambda}{2} - \frac{\bar{C}}{\epsilon_\lambda} = \frac{k_\lambda}{2} - \bar{C}^2 \geq 0 \\ \iff \quad & k_\lambda \geq 2\,\bar{C}^2, \end{aligned}$$

and consequently, with this choice of k_λ, $T_\lambda \geq 0$ and

$$a_h(\mathbf{v}, \mathbf{v}) \geq C \|\mathbf{v}\|_{\mathrm{DG}}^2 \tag{3.1.8}$$

for all $\mathbf{v} \in \left[H^1(\mathcal{T}_h)\right]^2 \supset V_h$.

From (3.1.5) and (3.1.7) it can be seen that in order to ensure a non-negative coefficient of the second norm in each case, as required to show coercivity of the form and thus stability of the method, both k_μ and k_λ need to have minimum positive values. Specifically, both stabilization terms are necessary to ensure the stability of the method.

IIPG

Coercivity with a λ-independent constant has not thus far been proven for the IIPG method, for any element shape, although a more general proof (with $k_\lambda = 0$) has been given in [35].

However, it can be shown exactly as for SIPG, once again independent of the shape of the elements.

The bilinear form is

$$\begin{aligned} a_h(\mathbf{v}, \mathbf{v}) = 2\mu \sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\mathbf{v})\|_{L^2(\Omega_e)}^2 + \lambda \sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \mathbf{v}\|_{L^2(\Omega_e)}^2 - \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{v})\}\!\} : [\![\mathbf{v}]\!]\, ds \\ + k_\mu \mu \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|_{L^2(E)}^2 + k_\lambda \lambda \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \|[\![\mathbf{v}]]\|_{L^2(E)}^2, \end{aligned} \tag{3.1.9}$$

where the coefficient in the third term is the only difference from the corresponding expression for SIPG, so that the proof is identical up to constants.

Again, in order to ensure a non-negative coefficient of the second norm in each case, as required to show coercivity of the form and thus stability of the method, both k_μ and k_λ need to have minimum positive values. Specifically, both stabilization terms are necessary to ensure the stability of the method.

3.1.2 Consistency

Demonstrating consistency of the IP methods is done for the three variants simultaneously, and is included here primarily for later reference. It is identical to the case for triangular elements, and has been shown by multiple other authors (see, for example, [29] and [35]).

Since the exact solution $\mathbf{u}$ lies in $\left[H^2(\Omega)\right]^2$, it is continuous over the domain Ω.

Thus

$$\begin{aligned} &\llbracket \mathbf{u} \rrbracket = \mathbf{0}, \quad \lbrace\!\lbrace \mathbf{u} \rbrace\!\rbrace = 0 && \text{on } \Gamma_{int}, \\ &\llbracket \mathbf{u} \rrbracket = \mathbf{g} \otimes \mathbf{n}, \quad \lbrace\!\lbrace \mathbf{u} \rbrace\!\rbrace = \mathbf{g} \cdot \mathbf{n} && \text{on } \Gamma_D. \end{aligned} \tag{3.1.10}$$

It follows that

$$\begin{aligned} a_h(\mathbf{u},\mathbf{v}) - l_h(\mathbf{v}) = &\sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dx + \theta \sum_{E \in \Gamma_{iD}} \int_E \llbracket \mathbf{u} \rrbracket : \lbrace\!\lbrace \boldsymbol{\sigma}(\mathbf{v}) \rbrace\!\rbrace\, ds \\ &- \sum_{E \in \Gamma_{iD}} \int_E \lbrace\!\lbrace \boldsymbol{\sigma}(\mathbf{u}) \rbrace\!\rbrace : \llbracket \mathbf{v} \rrbracket\, ds + k_\mu \mu \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \int_E \llbracket \mathbf{u} \rrbracket : \llbracket \mathbf{v} \rrbracket\, ds \\ &+ k_\lambda \lambda \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \int_E \llbracket \mathbf{u} \rrbracket \llbracket \mathbf{v} \rrbracket\, ds - \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \mathbf{f} \cdot \mathbf{v}\, dx \\ &- \theta \sum_{E \in \Gamma_D} \int_E (\mathbf{g} \otimes \mathbf{n}) : \boldsymbol{\sigma}(\mathbf{v})\, ds - \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v}\, ds \\ &- k_\mu \mu \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E \mathbf{g} \cdot \mathbf{v}\, ds - k_\lambda \lambda \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E (\mathbf{g} \cdot \mathbf{n})(\mathbf{v} \cdot \mathbf{n})\, ds \end{aligned}$$

$$
\begin{aligned}
&= \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dx + \theta \sum_{E \in \Gamma_D} \int_E (\mathbf{g} \otimes \mathbf{n}) : \boldsymbol{\sigma}(\mathbf{v})\, ds \\
&\quad - \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!]\, ds + k_\mu \mu \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E (\mathbf{g} \otimes \mathbf{n}) : (\mathbf{v} \otimes \mathbf{n})\, ds \\
&\quad + k_\lambda \lambda \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E (\mathbf{g} \cdot \mathbf{n})(\mathbf{v} \cdot \mathbf{n})\, ds - \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \mathbf{f} \cdot \mathbf{v}\, dx \\
&\quad - \theta \sum_{E \in \Gamma_D} \int_E (\mathbf{g} \otimes \mathbf{n}) : \boldsymbol{\sigma}(\mathbf{v})\, ds - \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v}\, ds \\
&\quad - k_\mu \mu \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E \mathbf{g} \cdot \mathbf{v}\, ds - k_\lambda \lambda \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E (\mathbf{g} \cdot \mathbf{n})(\mathbf{v} \cdot \mathbf{n})\, ds \\
&= \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dx - \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!]\, ds \\
&\quad - \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \mathbf{f} \cdot \mathbf{v}\, dx - \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v}\, ds. \qquad (3.1.11)
\end{aligned}
$$

By integrating formally, and since the exact solution $\mathbf{u}$ satisfies the weak form of the governing equation (2.1.1a),

$$
\begin{aligned}
\sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dx &= - \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \operatorname{div} \boldsymbol{\sigma}(\mathbf{u}) \cdot \mathbf{v}\, dx + \sum_{\Omega_e \in \mathcal{T}_h} \int_{\partial \Omega_e} \boldsymbol{\sigma}(\mathbf{u})\, \mathbf{n} \cdot \mathbf{v}\, ds \\
&= \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \mathbf{f} \cdot \mathbf{v}\, dx + \sum_{\Omega_e \in \mathcal{T}_h} \int_{\partial \Omega_e} \boldsymbol{\sigma}(\mathbf{u})\ \mathbf{n} \cdot \mathbf{v}\, ds.
\end{aligned}
$$

Thus

$$
a_h(\mathbf{u}, \mathbf{v}) - l_h(\mathbf{v}) = \sum_{\Omega_e \in \mathcal{T}_h} \int_{\partial \Omega_e} \boldsymbol{\sigma}(\mathbf{u})\ \mathbf{n} \cdot \mathbf{v}\, ds - \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!]\, ds - \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v}\, ds.
$$

By the identity (A.1.1),

$$
\begin{aligned}
&\sum_{\Omega_e \in \mathcal{T}_h} \int_{\partial \Omega_e} \boldsymbol{\sigma}(\mathbf{u})\ \mathbf{n} \cdot \mathbf{v}\, ds \\
&= \sum_{E \in \Gamma} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!]\, ds + \sum_{E \in \Gamma_{int}} \int_E [\![\boldsymbol{\sigma}(\mathbf{u})]\!] \cdot \{\!\{\mathbf{v}\}\!\}\, ds \\
&= \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!]\, ds + \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v}\, ds + \sum_{E \in \Gamma_{int}} \int_E [\![\boldsymbol{\sigma}(\mathbf{u})]\!] \cdot \{\!\{\mathbf{v}\}\!\}\, ds,
\end{aligned}
$$

and since

$$[\![\boldsymbol{\sigma}(\mathbf{u})]\!]|_E = \mathbf{0} \qquad \forall\, E \in \Gamma_{int},$$

it follows that

$$a_h(\mathbf{u}, \mathbf{v}) - l_h(\mathbf{v}) = 0. \tag{3.1.12}$$

Thus the formulation is consistent for all three IP methods under consideration.

3.2 Error bound: approach and outline

Numerical results, which will be detailed in Chapter 7, indicate that the quality of approximations using bilinear elements is negatively λ-dependent.

This section gives a preliminary form of the error analysis for bilinear elements. It starts with an outline of the proof of the error bound obtained by Wihler in [41] for NIPG for triangular elements, and then highlights key components of the analysis that do not, as will be seen, automatically hold for bilinear elements. Some of these components are directly related to the fact that linear approximations are being used, but others revolve around the interpolant used for error-splitting. Next, an interpolant for the error-splitting in the proof for bilinear elements, with a minimum number of specified properties, is assumed to exist. The preliminary analysis is performed using this hypothetical interpolant, and a bound reflecting the expected λ-dependence is obtained.

Following this is a discussion of how modifications to the methods provide the potential for obtaining a λ-independent bound, on the condition that the interpolant possesses particular properties suggested by the preliminary analysis.

The details of the interpolant, and the details of the modifications and the corresponding analyses, are left for later chapters.

3.2.1 Introduction to analysis approach

The approach used by Wihler [41] in bounding the approximation error of NIPG for triangular meshes is used as a starting point. Wihler's analysis allows for singularities

in the exact solution, and graded meshes, both of which are outside the scope of this thesis; however the key steps in his analysis, and tools he uses especially in obtaining an optimal error bound, are relevant. The special case (and simple case) of an exact solution in $H^2(\Omega)$ and a uniform mesh is therefore considered in what follows.

3.2.2 Sketch of Wihler's error analysis

The error analysis makes use of the Crouzeix-Raviart interpolant.

On each element, for $\mathbf{u} \in \left[H^2(\Omega_e)\right]^2$, the interpolant $\pi_e \mathbf{u} \in \left[\mathbb{P}_1(\Omega_e)\right]^2$ is uniquely defined relating the midpoint $\mathbf{m}_E$ of edge E to the mean value of $\mathbf{u}$ on E:

$$\pi_e \mathbf{u}(\mathbf{m}_E) := \frac{1}{h_E} \int_E \mathbf{u}\, ds \qquad \forall\ E \subset \partial\Omega_e. \tag{3.2.1}$$

From this definition follow the properties

$$\text{(a)} \quad \int_E (\mathbf{u} - \pi_e \mathbf{u})\ ds = \mathbf{0}, \tag{3.2.2a}$$

$$\text{(b)} \quad \int_E (\mathbf{u} - \pi_e \mathbf{u}) \cdot \mathbf{n}_e\, ds = 0 \qquad \forall\ E \subset \partial\Omega_e, \tag{3.2.2b}$$

$$\text{(c)} \quad \int_{\Omega_e} \nabla \cdot (\mathbf{u} - \pi_e \mathbf{u})\ dx = 0 \tag{3.2.2c}$$

The following interpolation error estimates hold:

$$\|\mathbf{u} - \pi_e \mathbf{u}\|_{L^2(\Omega_e)} + h_e \left|\mathbf{u} - \pi_e \mathbf{u}\right|_{H^1(\Omega_e)} \leq C h_e^2 \left|\mathbf{u}\right|_{H^2(\Omega_e)} \tag{3.2.3a}$$

$$\left|\mathbf{u} - \pi_e \mathbf{u}\right|_{H^2(\Omega_e)} \leq \left|\mathbf{u}\right|_{H^2(\Omega_e)} \tag{3.2.3b}$$

$$\|\nabla \cdot (\mathbf{u} - \pi_e \mathbf{u})\|_{L^2(\Omega_e)} \leq C h_e \left|\nabla \cdot \mathbf{u}\right|_{H^1(\Omega_e)} \tag{3.2.3c}$$

$$\left|\nabla \cdot (\mathbf{u} - \pi_e \mathbf{u})\right|_{H^1(\Omega_e)} \leq \left|\nabla \cdot \mathbf{u}\right|_{H^1(\Omega_e)} \tag{3.2.3d}$$

where $C > 0$ is a constant independent of h_e and $\mathbf{u}$.

The global interpolant $\pi : \left[H^2(\Omega)\right]^2 \longrightarrow V_h$ is defined by

$$\pi \mathbf{u}|_{\Omega_e} = \pi_e \mathbf{u} \qquad \forall\ \Omega_e \in \mathcal{T}_h.$$

The approximation error for the DG method is written

$$\boldsymbol{e} = \boldsymbol{\eta} + \boldsymbol{\xi}, \tag{3.2.4}$$

where

$$\boldsymbol{\eta} := \mathbf{u} - \pi\mathbf{u} \tag{3.2.5a}$$

$$\boldsymbol{\xi} := \pi\mathbf{u} - \mathbf{u}_h. \tag{3.2.5b}$$

Thus the norm of the error is

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq \|\boldsymbol{\eta}\|_{\mathrm{DG}} + \|\boldsymbol{\xi}\|_{\mathrm{DG}}\,. \tag{3.2.6}$$

Firstly, the bound

$$\|\boldsymbol{\eta}\|_{\mathrm{DG}} \leq Ch\left(\|\mathbf{u}\|_{H^2(\Omega)} + \lambda\,\|\nabla\cdot\mathbf{u}\|_{H^1(\Omega)}\right) \tag{3.2.7}$$

is derived.

Next, obtaining the bound on $\|\boldsymbol{\xi}\|_{\mathrm{DG}}$ has a number of key aspects.

By the coercivity of the bilinear form,

$$C\,\|\boldsymbol{\xi}\|_{\mathrm{DG}}^2 \leq a_h\left(\boldsymbol{\xi},\boldsymbol{\xi}\right).$$

By the consistency of the formulation,

$$a_h\left(\boldsymbol{e},\boldsymbol{\xi}\right) = 0,$$

so that

$$a_h\left(\boldsymbol{\xi},\boldsymbol{\xi}\right) = a_h\left(\boldsymbol{e}-\boldsymbol{\eta},\boldsymbol{\xi}\right) = -a_h\left(\boldsymbol{\eta},\boldsymbol{\xi}\right).$$

Thus,

$$C\,\|\boldsymbol{\xi}\|_{\mathrm{DG}}^2 \leq \left|a_h\left(\boldsymbol{\eta},\boldsymbol{\xi}\right)\right|. \tag{3.2.8}$$

The right-hand side can be bounded by the absolute values of each of the integrals in the bilinear form, using the triangle inequality, and each term bounded individually.

The strategy is to extract a factor of $\|\boldsymbol{\xi}\|_{\mathrm{DG}}$ from each term (if the term does not vanish), to obtain an expression of the form $\|\boldsymbol{\xi}\|_{\mathrm{DG}}\,\phi\left(\boldsymbol{\eta}\right)$, for some function ϕ. Then

$$\|\boldsymbol{\xi}\|_{\mathrm{DG}}^2 \leq C\,\|\boldsymbol{\xi}\|_{\mathrm{DG}}\,\phi\left(\boldsymbol{\eta}\right) \tag{3.2.9a}$$

$$\implies \|\boldsymbol{\xi}\|_{\mathrm{DG}} \leq C\phi\left(\boldsymbol{\eta}\right) \tag{3.2.9b}$$

$$\implies \|\boldsymbol{\xi}\|_{\mathrm{DG}}^2 \leq C\left(\phi\left(\boldsymbol{\eta}\right)\right)^2. \tag{3.2.9c}$$

Following that, $\phi(\boldsymbol{\eta})$ is bounded by norms of the exact solution $\mathbf{u}$ and the overall error bound obtained:

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq Ch\left(\|\mathbf{u}\|_{H^2(\Omega)} + \lambda \|\nabla \cdot \mathbf{u}\|_{H^1(\Omega)}\right). \tag{3.2.10}$$

For domains satisfying the regularity condition (A.6.1), the right-hand side of this can be bounded by a finite constant, so that the optimal bound

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq Ch \tag{3.2.11}$$

holds.

3.2.3 Some key points in the bounding process

Several points that are very significant in the bounding process warrant highlighting.

1. $\pi_e \mathbf{u} \in [\mathbb{P}_1]^2 \implies \pi\mathbf{u} \in V_h$ means that $\boldsymbol{\xi} \in V_h$. Since the formulation is consistent, by Galerkin orthogonality $a_h(\boldsymbol{e}, \boldsymbol{\xi}) = 0$, which is crucial in obtaining the bound (3.2.8) for $\|\boldsymbol{\xi}\|_{\mathrm{DG}}$.
2. $\boldsymbol{\xi}|_{\Omega_e} \in [\mathbb{P}_1]^2$, implies that $\nabla \cdot \boldsymbol{\xi}$ and all the components of $\boldsymbol{\sigma}(\boldsymbol{\xi})$ are element-wise constant, a property that is useful in simplifying integrals.
3. The properties of the Crouzeix-Raviart interpolant include the vanishing of particular integrals involving $\boldsymbol{\eta}$, as seen in (3.2.2).

The second and third properties are responsible for the elimination of otherwise potentially problematic terms. For example, a term appearing in the bilinear form that vanishes due to the properties mentioned is

$$\begin{aligned}
&\sum_{E \in \Gamma_{iD}} \int_E [\![\boldsymbol{\eta}]\!] : \{\!\{\boldsymbol{\sigma}(\boldsymbol{\xi})\}\!\}\, ds \\
&= \sum_{E \in \Gamma_{iD}} \{\!\{\boldsymbol{\sigma}(\boldsymbol{\xi})\}\!\} : \int_E [\![\boldsymbol{\eta}]\!]\, ds \qquad (\boldsymbol{\sigma}(\boldsymbol{\xi}) \text{ constant}) \\
&= 0 \qquad \text{(by property (a) of the interpolant),}
\end{aligned} \tag{3.2.12}$$

since

$$\int_E \boldsymbol{\eta}\, ds = 0 \Longrightarrow \int_E [\![\boldsymbol{\eta}]\!]\, ds = 0. \tag{3.2.13}$$

The terms that remain either are not λ-dependent, or contain λ in correspondence with $\nabla \cdot \boldsymbol{\eta}$, which can be bounded by a norm of $\nabla \cdot \mathbf{u}$ using (3.2.3c) or (3.2.3d). Thus the regularity result can be used and λ-dependence does not enter the final error bound.

3.2.4 Approach to bounding the error for quadrilateral elements

To follow a similar approach in obtaining an error bound for quadrilateral elements, a suitable function for splitting the error is needed. This function would ideally lie in V_h, have the necessary associated error estimates, and possess any other properties (for example, orthogonality properties) necessary or helpful in eliminating potentially problematic terms.

Moreover, since bilinear elements are being considered, $\boldsymbol{\xi} \in [\mathbb{Q}_1]^2$ element-wise, and therefore does not have element-wise constant derivatives as in the case of the triangular elements. Alternative approaches to bounding the affected terms will therefore be necessary.

3.2.5 A preliminary bound for bilinear elements

In order to clarify more detail about the hypothetical "suitable" error-splitting function described necessarily vaguely in §3.2.4, as well as the type of error bound that may reasonably be expected, a preliminary bound is derived. This is done by assuming the existence of an error-splitting function that satisfies certain difference estimates, and using it to obtain an error bound. The procedure and result can then be investigated to obtain insight into various components of the analysis. This proof overlaps in portions with that of [41]; the common details will nevertheless be described in what follows.

Splitting the error

Suppose there existed a $\mathbf{u}_P \in V_h$ satisfying the estimates

$$\|\mathbf{u} - \mathbf{u}_P\|_{L^2(\Omega_e)} \leq C h_e^2 \, |\mathbf{u}|_{H^2(\Omega_e)} \,, \tag{3.2.14a}$$

$$|\mathbf{u} - \mathbf{u}_P|_{H^1(\Omega_e)} \leq C h_e \, |\mathbf{u}|_{H^2(\Omega_e)} \,, \tag{3.2.14b}$$

$$|\mathbf{u} - \mathbf{u}_P|_{H^2(\Omega_e)} \leq C \, |\mathbf{u}|_{H^2(\Omega_e)} \,, \tag{3.2.14c}$$

and

$$\|\nabla \cdot (\mathbf{u} - \mathbf{u}_P)\|_{L^2(\Omega_e)} \leq C h_e \, |\nabla \cdot \mathbf{u}|_{H^1(\Omega_e)} \,, \tag{3.2.15a}$$

$$|\nabla \cdot (\mathbf{u} - \mathbf{u}_P)|_{H^1(\Omega_e)} \leq C \, |\nabla \cdot \mathbf{u}|_{H^1(\Omega_e)} \,. \tag{3.2.15b}$$

To split the error, define

$$\boldsymbol{\gamma} := \mathbf{u} - \mathbf{u}_P \tag{3.2.16a}$$

$$\boldsymbol{w} := \mathbf{u}_P - \mathbf{u}_h, \tag{3.2.16b}$$

so that

$$\boldsymbol{e} = \boldsymbol{\gamma} + \boldsymbol{w}. \tag{3.2.17}$$

Then

$$C \, \|\boldsymbol{e}\|_{\mathrm{DG}}^2 \leq \|\boldsymbol{\gamma}\|_{\mathrm{DG}}^2 + \|\boldsymbol{w}\|_{\mathrm{DG}}^2 \,. \tag{3.2.18}$$

The first term

By definition of the DG norm,

$$\|\boldsymbol{\gamma}\|_{\mathrm{DG}}^2 = \sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\|_{L^2(\Omega_e)}^2 + \tfrac{1}{2} \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \, \|[\![\boldsymbol{\gamma}]\!]\|_{L^2(E)}^2 \,. \tag{3.2.19}$$

Firstly,

$$\begin{aligned} \sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\|_{L^2(\Omega_e)}^2 &\leq \sum_{\Omega_e \in \mathcal{T}_h} |\boldsymbol{\gamma}|_{H^1(\Omega_e)}^2 \\ &\leq C \sum_{\Omega_e \in \mathcal{T}_h} h_e^2 \, |\mathbf{u}|_{H^2(\Omega_e)}^2 \,, \end{aligned} \tag{3.2.20}$$

and secondly,

$$\begin{aligned}\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\left\|[\![\boldsymbol{\gamma}]\!]\right\|^2_{L^2(E)} &\leq C\sum_{\Omega_e\in\mathcal{T}_h}\sum_{E\in\partial\Omega_e^{iD}}\frac{1}{h_E}\left\|\boldsymbol{\gamma}\right\|^2_{L^2(E)}\\ &\leq C\sum_{\Omega_e\in\mathcal{T}_h}\left(h_e^{-2}\left\|\boldsymbol{\gamma}\right\|^2_{L^2(\Omega_e)}+\left|\boldsymbol{\gamma}\right|^2_{H^1(\Omega_e)}\right)\\ &\leq C\sum_{\Omega_e\in\mathcal{T}_h}h_e^2\left|\mathbf{u}\right|^2_{H^2(\Omega_e)}, \end{aligned}\tag{3.2.21}$$

so that

$$\left\|\boldsymbol{\gamma}\right\|^2_{\mathrm{DG}}\leq C\sum_{\Omega_e\in\mathcal{T}_h}h_e^2\left\|\mathbf{u}\right\|^2_{H^2(\Omega_e)}.\tag{3.2.22}$$

The second term: using the bilinear form

Since the bilinear form is coercive with respect to the DG norm,

$$\begin{aligned}\left\|\boldsymbol{w}\right\|^2_{\mathrm{DG}} &\leq a_h\left(\boldsymbol{w},\boldsymbol{w}\right)\\ &= \left|a_h\left(\boldsymbol{e}-\boldsymbol{\gamma},\boldsymbol{w}\right)\right|.\end{aligned}\tag{3.2.23}$$

By assumption, $\mathbf{u}_P\in V_h$, and consequently $\boldsymbol{w}=\mathbf{u}_P-\mathbf{u}_h\in V_h$.

By consistency of the formulation,

$$a_h\left(\mathbf{u},\boldsymbol{w}\right)=l_h\left(\boldsymbol{w}\right),\tag{3.2.24}$$

and since $\boldsymbol{w}\in V_h$,

$$a_h\left(\mathbf{u}_h,\boldsymbol{w}\right)=l_h\left(\boldsymbol{w}\right).\tag{3.2.25}$$

Thus

$$a_h(\boldsymbol{e},\boldsymbol{w})=0,\tag{3.2.26}$$

and

$$\begin{aligned}
\|\boldsymbol{w}\|_{\mathrm{DG}}^2 &\leq |-a_h(\boldsymbol{\gamma},\boldsymbol{w})| \\
&\leq 2\mu\left|\sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\boldsymbol{\varepsilon}(\boldsymbol{\gamma}):\boldsymbol{\varepsilon}(\boldsymbol{w})\,dx\right| + \lambda\left|\sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}(\nabla\cdot\boldsymbol{\gamma})(\nabla\cdot\boldsymbol{w})\,dx\right| \\
&\quad + 2\mu\left|\theta\sum_{E\in\Gamma_{iD}}\int_E[\![\boldsymbol{\gamma}]\!]:\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\,ds\right| + \lambda\left|\theta\sum_{E\in\Gamma_{iD}}\int_E[\![\boldsymbol{\gamma}]\!]:\{\!\{\nabla\cdot\boldsymbol{w}\mathbf{1}\}\!\}\,ds\right| \\
&\quad + 2\mu\left|\sum_{E\in\Gamma_{iD}}\int_E\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\}\!\}:[\![\boldsymbol{w}]\!]\,ds\right| + \lambda\left|\sum_{E\in\Gamma_{iD}}\int_E\{\!\{\nabla\cdot\boldsymbol{\gamma}\mathbf{1}\}\!\}:[\![\boldsymbol{w}]\!]\,ds\right| \\
&\quad + k_\mu\mu\left|\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\int_E[\![\boldsymbol{\gamma}]\!]:[\![\boldsymbol{w}]\!]\,ds\right| + k_\lambda\lambda\left|\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\int_E[\![\boldsymbol{\gamma}]\!][\![\boldsymbol{w}]\!]\,ds\right| \\
&= I+II+III+IV+V+VI+VII+VIII
\end{aligned} \tag{3.2.27}$$

by labelling the integrals $I-VIII$.

The strategy is again to extract a factor of $\|\boldsymbol{w}\|_{\mathrm{DG}}$ from each term (if the term does not vanish), to obtain an expression of the form $\|\boldsymbol{w}\|_{\mathrm{DG}}\,\phi(\boldsymbol{\gamma})$, for some function ϕ. Then

$$\|\boldsymbol{w}\|_{\mathrm{DG}}^2 \leq C\,\|\boldsymbol{w}\|_{\mathrm{DG}}\,\phi(\boldsymbol{\gamma}) \tag{3.2.28a}$$

$$\implies \|\boldsymbol{w}\|_{\mathrm{DG}} \leq C\phi(\boldsymbol{\gamma}) \tag{3.2.28b}$$

$$\implies \|\boldsymbol{w}\|_{\mathrm{DG}}^2 \leq C\left(\phi(\boldsymbol{\gamma})\right)^2. \tag{3.2.28c}$$

Following that, $\phi(\boldsymbol{\gamma})$ will be bounded by norms of the exact solution $\mathbf{u}$.

The second term: Extracting the factor $\|\boldsymbol{w}\|_{\mathrm{DG}}$

The terms of (3.2.27) are considered one by one.

I:

$$\begin{aligned} I &= 2\mu \left| \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\varepsilon}(\boldsymbol{\gamma}) : \boldsymbol{\varepsilon}(\boldsymbol{w}) \, dx \right| \\ &\leq C \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\boldsymbol{w})\|^2_{L^2(\Omega_e)} \right)^{\frac{1}{2}} \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\|^2_{L^2(\Omega_e)} \right)^{\frac{1}{2}} \\ &\leq C \|\boldsymbol{w}\|_{\mathrm{DG}} \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\|^2_{L^2(\Omega_e)} \right)^{\frac{1}{2}}. \end{aligned} \tag{3.2.29}$$

II:

$$\begin{aligned} II &= \lambda \left| \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} (\nabla \cdot \boldsymbol{\gamma})(\nabla \cdot \boldsymbol{w}) \, dx \right| \\ &\leq \lambda \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \boldsymbol{w}\|^2_{L^2(\Omega_e)} \right)^{\frac{1}{2}} \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \boldsymbol{\gamma}\|^2_{L^2(\Omega_e)} \right)^{\frac{1}{2}}. \end{aligned} \tag{3.2.30}$$

Since

$$\begin{aligned} \sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \boldsymbol{w}\|^2_{L^2(\Omega_e)} &\leq C \sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\boldsymbol{w})\|^2_{L^2(\Omega_e)} \\ &\leq C \|\boldsymbol{w}\|^2_{\mathrm{DG}}, \end{aligned} \tag{3.2.31}$$

it follows that

$$II \leq C\lambda \|\boldsymbol{w}\|_{\mathrm{DG}} \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \boldsymbol{\gamma}\|^2_{L^2(\Omega_e)} \right)^{\frac{1}{2}}. \tag{3.2.32}$$

III:

For $\theta = 0$, this term vanishes. For $\theta = \pm 1$,

$$
\begin{aligned}
III &= 2\mu \left| \sum_{E\in\Gamma_{iD}} \int_E [\![\boldsymbol{\gamma}]\!] : \{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\, ds \right| \\
&\leq C \sum_{E\in\Gamma_{iD}} \|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\|_{L^2(E)} \, \|[\![\boldsymbol{\gamma}]\!]\|_{L^2(E)} \\
&= C \sum_{E\in\Gamma_{iD}} h_E^{1/2} \|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\|_{L^2(E)} \, h_E^{-1/2} \|[\![\boldsymbol{\gamma}]\!]\|_{L^2(E)} \\
&\leq C \left(\sum_{E\in\Gamma_{iD}} h_E \|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\|^2_{L^2(E)} \right)^{\frac{1}{2}} \left(\sum_{E\in\Gamma_{iD}} h_E^{-1} \|[\![\boldsymbol{\gamma}]\!]\|^2_{L^2(E)} \right)^{\frac{1}{2}}. \qquad (3.2.33)
\end{aligned}
$$

Now,

$$
\begin{aligned}
\sum_{E\in\Gamma_{iD}} h_E \|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\|^2_{L^2(E)} &\leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e \|\boldsymbol{\varepsilon}(\boldsymbol{w})\|^2_{L^2(\partial\Omega_e)} \\
&\leq C \sum_{\Omega_e\in\mathcal{T}_h} \|\boldsymbol{\varepsilon}(\boldsymbol{w})\|^2_{L^2(\Omega_e)} \\
&\leq C \|\boldsymbol{w}\|^2_{\mathrm{DG}}, \qquad (3.2.34)
\end{aligned}
$$

giving

$$
III \leq C \|\boldsymbol{w}\|_{\mathrm{DG}} \left(\sum_{E\in\Gamma_{iD}} h_E^{-1} \|[\![\boldsymbol{\gamma}]\!]\|^2_{L^2(E)} \right)^{\frac{1}{2}}. \qquad (3.2.35)
$$

IV:

Again, for $\theta = 0$, this term vanishes. For $\theta = \pm 1$,

$$
\begin{aligned}
IV &= \lambda \left| \sum_{E\in\Gamma_{iD}} \int_E [\![\boldsymbol{\gamma}]\!] : \{\!\{\nabla\cdot\boldsymbol{w}\mathbf{1}\}\!\}\, ds \right| \\
&\leq C\lambda \left(\sum_{E\in\Gamma_{iD}} h_E \|\{\!\{\nabla\cdot\boldsymbol{w}\}\!\}\|^2_{L^2(E)} \right)^{\frac{1}{2}} \left(\sum_{E\in\Gamma_{iD}} h_E^{-1} \|[\![\boldsymbol{\gamma}]\!]\|^2_{L^2(E)} \right)^{\frac{1}{2}} \qquad (3.2.36)
\end{aligned}
$$

as in bounding III.

Since

$$
\|\{\!\{\nabla\cdot\boldsymbol{w}\}\!\}\|^2_{L^2(E)} \leq C \|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\|^2_{L^2(E)}, \qquad (3.2.37)
$$

the bound follows as for III:

$$\sum_{E\in\Gamma_{iD}} h_E \left\|\{\!\{\nabla\cdot\boldsymbol{w}\}\!\}\right\|^2_{L^2(E)} \leq C \sum_{E\in\Gamma_{iD}} h_E \left\|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{w})\}\!\}\right\|^2_{L^2(E)} \leq C\left\|\boldsymbol{w}\right\|^2_{\mathrm{DG}}, \tag{3.2.38}$$

therefore

$$IV \leq C\lambda \left\|\boldsymbol{w}\right\|_{\mathrm{DG}} \left(\sum_E h_E^{-1} \left\|[\![\boldsymbol{\gamma}]\!]\right\|^2_{L^2(E)}\right)^{\frac{1}{2}}. \tag{3.2.39}$$

V:

$$\begin{aligned} V &= 2\mu \left|\sum_{E\in\Gamma_{iD}} \int_E \{\!\{\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\}\!\} : [\![\boldsymbol{w}]\!]\, ds\right| \\ &\leq C \sum_{E\in\Gamma_{iD}} h_E^{1/2} \left\|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\}\!\}\right\|^2_{L^2(E)} h_E^{-1/2} \left\|[\![\boldsymbol{w}]\!]\right\|^2_{L^2(E)} \\ &\leq C \left(\sum_{E\in\Gamma_{iD}} h_E \left\|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\}\!\}\right\|^2_{L^2(E)}\right)^{\frac{1}{2}} \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|[\![\boldsymbol{w}]\!]\right\|^2_{L^2(E)}\right)^{\frac{1}{2}} \\ &\leq C \left\|\boldsymbol{w}\right\|_{\mathrm{DG}} \left(\sum_{E\in\Gamma_{iD}} h_E \left\|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\}\!\}\right\|^2_{L^2(E)}\right)^{\frac{1}{2}}. \end{aligned} \tag{3.2.40}$$

VI:

Similarly,

$$\begin{aligned} VI &= \lambda \left|\sum_{E\in\Gamma_{iD}} \int_E \{\!\{\nabla\cdot\boldsymbol{\gamma}\mathbf{1}\}\!\} : [\![\boldsymbol{w}]\!]\, ds\right| \\ &\leq C\lambda \left\|\boldsymbol{w}\right\|_{\mathrm{DG}} \left(\sum_{E\in\Gamma_{iD}} h_E \left\|\{\!\{\nabla\cdot\boldsymbol{\gamma}\mathbf{1}\}\!\}\right\|^2_{L^2(E)}\right)^{\frac{1}{2}}. \end{aligned} \tag{3.2.41}$$

VII:

$$\begin{aligned} VII &= k_\mu \mu \left| \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \int_E \llbracket \boldsymbol{\gamma} \rrbracket : \llbracket \boldsymbol{w} \rrbracket \, ds \right| \\ &\leq C \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| \llbracket \boldsymbol{\gamma} \rrbracket \right\|^2_{L^2(E)} \right)^{\frac{1}{2}} \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| \llbracket \boldsymbol{w} \rrbracket \right\|^2_{L^2(E)} \right)^{\frac{1}{2}} \\ &\leq C \left\| \boldsymbol{w} \right\|_{\mathrm{DG}} \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| \llbracket \boldsymbol{\gamma} \rrbracket \right\|^2_{L^2(E)} \right)^{\frac{1}{2}} . \end{aligned} \tag{3.2.42}$$

$VIII$:

$$\begin{aligned} VIII = \; & k_\lambda \lambda \left| \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \int_E [\![\boldsymbol{\gamma}]\!] [\![\boldsymbol{w}]\!] \, ds \right| \\ \leq \; & k_\lambda \, \lambda \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| [\![\boldsymbol{\gamma}]\!] \right\|^2_{L^2(E)} \right)^{\frac{1}{2}} \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| [\![\boldsymbol{w}]\!] \right\|^2_{L^2(E)} \right)^{\frac{1}{2}} . \end{aligned} \tag{3.2.43}$$

Then

$$\begin{aligned} \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| [\![\boldsymbol{w}]\!] \right\|^2_{L^2(E)} &= \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| \mathbf{1} : \llbracket \boldsymbol{w} \rrbracket \right\|^2_{L^2(E)} \\ &\leq C \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| \llbracket \boldsymbol{w} \rrbracket \right\|^2_{L^2(E)} \\ &\leq C \left\| \boldsymbol{w} \right\|^2_{\mathrm{DG}} , \end{aligned} \tag{3.2.44}$$

giving

$$VIII \leq \lambda C \left\| \boldsymbol{w} \right\|_{\mathrm{DG}} \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| [\![\boldsymbol{\gamma}]\!] \right\|^2_{L^2(E)} \right)^{\frac{1}{2}} . \tag{3.2.45}$$

Combining these results gives the form aimed for in (3.2.28).

The second term: Bounding the expressions containing $\boldsymbol{\gamma}$

The terms containing $\boldsymbol{\gamma}$ will again be dealt with one by one, each integral's contribution to the final result then being evident. Each term is manipulated to the form of an expression in $\boldsymbol{\gamma}$ that can be bounded using (3.2.14) and (3.2.15).

I:

$$\|\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\|_{L^2(\Omega_e)} \leq \|\nabla\boldsymbol{\gamma}\|_{L^2(\Omega_e)} = |\boldsymbol{\gamma}|_{H^1(\Omega_e)} \leq Ch_e\,|\mathbf{u}|_{H^2(\Omega_e)} \tag{3.2.46}$$

II:

$$\|\nabla\cdot\boldsymbol{\gamma}\|_{L^2(\Omega_e)} \leq Ch_e\,|\nabla\cdot\mathbf{u}|_{H^1(\Omega_e)} \tag{3.2.47}$$

with an associated coefficient of λ.

III:

$$\begin{aligned}\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\|[\![\boldsymbol{\gamma}]\!]\|^2_{L^2(E)} &\leq C\sum_{\Omega_e\in\mathcal{T}_h}\left(h_e^{-2}\|\boldsymbol{\gamma}\|^2_{L^2(\Omega_e)} + |\boldsymbol{\gamma}|^2_{H^1(\Omega_e)}\right)\\ &\leq C\sum_{\Omega_e\in\mathcal{T}_h} h_e^2\,|\mathbf{u}|^2_{H^2(\Omega_e)}\end{aligned} \tag{3.2.48}$$

IV:

Exactly as in bounding III,

$$\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\|[\![\boldsymbol{\gamma}]\!]\|^2_{L^2(E)} \leq C\sum_{\Omega_e\in\mathcal{T}_h} h_e^2\,|\mathbf{u}|^2_{H^2(\Omega_e)}\,. \tag{3.2.49}$$

Here, there is also an associated coefficient of λ.

V:

$$\begin{aligned}\sum_{E\in\Gamma_{iD}} h_E\|\{\!\{\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\}\!\}\|^2_{L^2(E)} &\leq \sum_{\Omega_e\in\mathcal{T}_h}\sum_{E\in\partial\Omega_e^{iD}} h_E\|\boldsymbol{\varepsilon}(\boldsymbol{\gamma})\|^2_{L^2(E)}\\ &\leq \sum_{\Omega_e\in\mathcal{T}_h}\sum_{E\in\partial\Omega_e^{iD}} h_E\|\nabla\boldsymbol{\gamma}\|^2_{L^2(E)}\\ &\leq C\sum_{\Omega_e\in\mathcal{T}_h}\left(\|\nabla\boldsymbol{\gamma}\|^2_{L^2(\Omega_e)} + h_e^2\,|\nabla\boldsymbol{\gamma}|^2_{H^1(\Omega_e)}\right)\\ &= C\sum_{\Omega_e\in\mathcal{T}_h}\left(|\boldsymbol{\gamma}|^2_{H^1(\Omega_e)} + h_e^2\,|\boldsymbol{\gamma}|^2_{H^2(\Omega_e)}\right)\\ &\leq C\sum_{\Omega_e\in\mathcal{T}_h} h_e^2\,|\mathbf{u}|^2_{H^2(\Omega_e)}\,.\end{aligned} \tag{3.2.50}$$

VI:

Similarly to V, but using a bound on the divergence,

$$\begin{aligned}\sum_{E\in\Gamma_{iD}} h_E \left\|\{\!\{\nabla\cdot\boldsymbol{\gamma}\mathbf{1}\}\!\}\right\|^2_{L^2(E)} &\leq C \sum_{\Omega_e\in\mathcal{T}_h}\sum_{E\in\partial\Omega_e^{iD}} h_E \left\|\nabla\cdot\boldsymbol{\gamma}\right\|^2_{L^2(E)} \\ &\leq C \sum_{\Omega_e\in\mathcal{T}_h} \left(\left\|\nabla\cdot\boldsymbol{\gamma}\right\|^2_{L^2(\Omega_e)} + h_e^2 \left|\nabla\cdot\boldsymbol{\gamma}\right|^2_{H^1(\Omega_e)}\right) \\ &\leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e^2 \left|\nabla\cdot\mathbf{u}\right|^2_{H^1(\Omega_e)} . \end{aligned} \tag{3.2.51}$$

with an associated coefficient of λ.

VII:

Exactly as in bounding III,

$$\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|[\![\boldsymbol{\gamma}]\!]\right\|^2_{L^2(E)} \leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e^2 \left|\mathbf{u}\right|^2_{H^2(\Omega_e)} . \tag{3.2.52}$$

$VIII$:

$$\begin{aligned}\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|[\![\gamma]\!]\right\|^2_{L^2(E)} &\leq C \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|[\![\boldsymbol{\gamma}]\!]\right\|^2_{L^2(E)} \\ &\leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e^2 \left\|\mathbf{u}\right\|^2_{H^2(\Omega_e)} , \end{aligned} \tag{3.2.53}$$

again with an associated coefficient of λ.

The overall bound for the second term

Combining the results of extracting the factor $\|\boldsymbol{w}\|_{\mathrm{DG}}$, using the manipulation of (3.2.28), and bounding the expressions involving $\boldsymbol{\gamma}$, one obtains

$$\left\|\boldsymbol{w}\right\|^2_{\mathrm{DG}} \leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e^2 \left(\left\|\mathbf{u}\right\|^2_{H^2(\Omega_e)} + \lambda^2 \left\|\mathbf{u}\right\|^2_{H^2(\Omega_e)} + \lambda^2 \left\|\nabla\cdot\mathbf{u}\right\|^2_{H^1(\Omega_e)}\right) . \tag{3.2.54}$$

The second term in the brackets has contributions from IV and $VIII$.

The final error bound

With $\|\boldsymbol{\gamma}\|^2_{\mathrm{DG}}$ and $\|\boldsymbol{w}\|^2_{\mathrm{DG}}$ bounded, it follows that

$$\begin{aligned}\|\boldsymbol{e}\|^2_{\mathrm{DG}} &\leq C \sum_{\Omega_e \in \mathcal{T}_h} h_e^2 \left(\|\mathbf{u}\|^2_{H^2(\Omega_e)} + \lambda^2 \|\mathbf{u}\|^2_{H^2(\Omega_e)} + \lambda^2 \|\nabla \cdot \mathbf{u}\|^2_{H^1(\Omega_e)} \right) \\ &\leq Ch^2 \left(\|\mathbf{u}\|^2_{H^2(\Omega)} + \lambda^2 \|\mathbf{u}\|^2_{H^2(\Omega)} + \lambda^2 \|\nabla \cdot \mathbf{u}\|^2_{H^1(\Omega)} \right). \end{aligned} \tag{3.2.55}$$

Finally, for domains satisfying the regularity result

$$\|\mathbf{u}\|_{H^2(\Omega)} + |\lambda| \, \|\nabla \cdot \mathbf{u}\|_{H^1(\Omega)} \leq CF\left(\mathbf{f}, \mathbf{g}_\Omega, \mathbf{h}_\Omega\right) \tag{3.2.56}$$

(cf. A.6), where $F\left(\mathbf{f}, \mathbf{g}_\Omega, \mathbf{h}_\Omega\right)$ is bounded,

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq C\lambda \, h. \tag{3.2.57}$$

Here, λ enters the estimate as a result of the second term in the brackets in (3.2.55).

3.2.6 Discussion of the preliminary analysis

This bound has made use of the fact that the hypothetical $\mathbf{u}_P$ lies in V_h, and that it satisfies certain difference estimates, but uses no other properties, as none have yet been specified.

The estimate that has been obtained is optimal with regard to mesh-size, but indicates that, unlike for linear triangular elements, the approximation error depends on λ, and therefore that the approximation is prone to being very poor for nearly incompressible materials, where λ is very large.

Specifically, the terms contributing to the poorness of the result are IV and $VIII$.

(Since for the coercivity of the bilinear form $k_\lambda > 0$ when $\theta = 0$ or -1, the situation of both k_λ and θ being zero - and thus both IV and $VIII$ vanishing - does not arise. Therefore the undesirable λ-dependence will always be present.)

The potentially problematic terms having been identified, investigating them suggests a way of circumventing the problem they cause.

In both terms, if the factor involving $\boldsymbol{w}$ in the integrand were edge-wise constant rather than linear, it could be factorised out of the integral, isolating $[\![\boldsymbol{\gamma}]\!]$ within the edge integral. If, then, a $\mathbf{u}_P$ could be found that had the property that the integral of that jump term vanished on edges (as in the case of the Crouzeix-Raviart interpolant, on triangles), the terms would vanish completely, thus eliminating the λ-dependence from the error bound.

3.2.7 Preliminary proposed solutions

With bilinear elements in the IP formulation as it stands, $\boldsymbol{w} \in V_h$ cannot in general lead to edge-wise constant values as would be convenient, and the λ-dependence cannot thus be eliminated.

However, there are two ways to obtain constant factors involving $\boldsymbol{w}$. The first is to modify the formulation by replacing each of those factors with its projection onto the space of constants. The second is to consider linear approximations rather than bilinear approximations, so that $\nabla \cdot \boldsymbol{w}$ would be element-wise constant - this approach would be helpful in term IV, although not in $VIII$.

In each case, a $\mathbf{u}_P$ would need to found, with adequate properties. A further necessary step would then be to ascertain whether the modification, whether of the formulation or of the approximation type, affected any other aspects of the analysis. Should difficulties arise, these would need to be addressed.

Each of the suggested modifications will be considered in detail. First, however, an appropriate $\mathbf{u}_P$ will be found. This will be the subject of Chapter 4.

Chapter 4

Constructing $\mathbf{u}_P$

The preliminary analysis for bilinear elements, detailed in the previous chapter, assumes the existence of a suitable error-splitting function, $\mathbf{u}_P$, assuming certain properties and indicating that others are desirable. This chapter is devoted to constructing a function with all of these properties.

It begins with an outline of the requirements on the function. The space in which it must lie is carefully considered, and following this is a motivation for using, as the splitting function, the projection of an interpolant onto the space of linear polynomials (element-wise), rather than the interpolant itself.

With this strategy in place, the required properties of the underlying interpolant are established, several potential interpolants are discussed and eliminated as options, and a new interpolant is finally constructed, designed specifically to fulfil the requirements.

4.1 Requirements on $\mathbf{u}_P$

The properties of $\mathbf{u}_P$ assumed in the preliminary analysis of the previous chapter are

1. that it lies in V_h,
2. that it satisfies the basic interpolation error bounds (3.2.14), and

3. that it satisfies the bounds on the divergence of the interpolation error (3.2.15).

The analysis indicates that a further useful property, needed specifically for obtaining λ-independent convergence if the proposed modifications are effective, would be

4. that it satisfies

$$\int_E [\![\mathbf{u} - \mathbf{u}_P]\!] \, ds = 0 \tag{4.1.1}$$

for all $E \in \Gamma_{iD}$.

4.2 The interpolation space

4.2.1 Possible candidates for $\mathbf{u}_P$

In the search for an interpolant that could be used as the splitting function $\mathbf{u}_P$, a number have been considered, but none found that satisfies the requirements given in §4.1.

Beginning with the requirement that $\mathbf{u}_P$ lies in V_h, various interpolants that are locally either in $\mathbb{Q}_1$ or in a subspace of $\mathbb{Q}_1$ have been considered for use as $\mathbf{u}_P$. The possibility of using linear rather than bilinear approximations, as suggested in the preceding chapter, also motivates trying to identify a suitable interpolant that is element-wise specifically in $\mathbb{P}_1$. Such an interpolant would be appropriate for both linear and bilinear approximations. Examples are the linear interpolant of Girault, and the lowest-order Raviart-Thomas and BDM interpolants for quadrilateral elements.

Girault [23] presents a linear interpolant on quadrilaterals, the L^2-orthogonal projection onto $\mathbb{P}_1$ on each element. While this interpolant satisfies the desired bounds, sufficient element boundary information for evaluating the jumps in normal components on each edge is not available.

The lowest-order Raviart-Thomas and BDM interpolants for quadrilateral elements (see [11]), designed as non-conforming elements, are examples of interpolants that lie within even smaller subspaces of V_h. These, however, satisfy only the bounds (4.5.1a) and (4.5.1b). In general, any interpolant that lies in a space smaller than $\mathbb{P}_1$ element-wise

will not preserve linear polynomials, and therefore will not satisfy all the necessary error bounds (cf. A.5).

On the other hand, higher-order Raviart-Thomas and BDM interpolants (see [11]), while providing all the desired error bounds, are not fully contained within V_h as required. Other interpolants also with higher-order terms are those based on the rotated bilinear elements of Rannacher and Turek [32], and the elements of Douglas et al. [22], where nodal values are defined at edge midpoints. A bilinear element with nodal values at midpoints cannot be defined, as a singular system is obtained.

Thus an interpolant satisfying all the requirements on $\mathbf{u}_P$ is not readily available.

4.2.2 Evaluating the requirement $\mathbf{u}_P \in V_h$

While it was assumed in the preceding chapter that $\mathbf{u}_P$ would lie in V_h, this assumption was given without motivation. V_h is the relevant space in the analysis for triangular elements, and it appears to be a natural assumption to make. Nevertheless, the question arises as to whether or not this assumption is necessary.

Suppose that one has a function $\pi\mathbf{u} \in X$, where $V_h \subset X$, and write the approximation error $\boldsymbol{e}$ as

$$\boldsymbol{e} = \boldsymbol{\eta} + \boldsymbol{\xi} \quad \text{where} \quad \begin{array}{l} \boldsymbol{\eta} = \mathbf{u} - \pi\mathbf{u} \\ \boldsymbol{\xi} = \pi\mathbf{u} - \mathbf{u}_h \end{array}, \tag{4.2.1}$$

as was done in the context of triangular elements.

Since the bilinear form of the IP method is coercive for all functions in $H^1(\mathcal{T}_h)$,

$$\begin{aligned} C \left\| \boldsymbol{\xi} \right\|_{\mathrm{DG}}^2 &\leq a_h(\boldsymbol{\xi}, \boldsymbol{\xi}) \\ &= a_h(\boldsymbol{e}, \boldsymbol{\xi}) - a_h(\boldsymbol{\eta}, \boldsymbol{\xi}). \end{aligned} \tag{4.2.2}$$

However, Galerkin orthogonality holds only with respect to the space V_h, which means that in general $a_h(\boldsymbol{e}, \boldsymbol{\xi}) \neq 0$. Writing $\boldsymbol{\xi} = \boldsymbol{\xi}_V + \boldsymbol{\xi}_X$, where $\boldsymbol{\xi}_V$ is the part of $\boldsymbol{\xi}$ lying in V_h and $\boldsymbol{\xi}_X$ is the remainder, and using Galerkin orthogonality on the applicable portion,

one obtains

$$\begin{aligned} a_h(\boldsymbol{e}, \boldsymbol{\xi}) &= a_h(\boldsymbol{e}, \boldsymbol{\xi}_V) + a_h(\boldsymbol{e}, \boldsymbol{\xi}_X) \\ &= a_h(\boldsymbol{e}, \boldsymbol{\xi}_X), \end{aligned} \tag{4.2.3}$$

giving

$$C \|\boldsymbol{\xi}\|_{\mathrm{DG}}^2 \leq |a_h(\boldsymbol{e}, \boldsymbol{\xi}_X)| + |a_h(\boldsymbol{\eta}, \boldsymbol{\xi})|. \tag{4.2.4}$$

Following the usual bounding strategy (as contained in §3.2.5) for the second term on the right-hand side, one would obtain

$$C \|\boldsymbol{\xi}\|_{\mathrm{DG}}^2 \leq |a_h(\boldsymbol{e}, \boldsymbol{\xi}_X)| + \|\boldsymbol{\xi}\|_{\mathrm{DG}} \phi(\boldsymbol{\eta}). \tag{4.2.5}$$

However, the first term on the right-hand side becomes problematic, and hinders the bounding process. This arises directly as a result of splitting the error with a function that does not lie completely in the DG solution space V_h. That is, $\mathbf{u}_P = \pi\mathbf{u} \notin V_h$ is problematic, and therefore the assumption $\mathbf{u}_P \in V_h$ will be retained.

4.2.3 A projection onto $\mathbb{P}_1$

Returning to the coercivity of the bilinear form, consider instead a bound on the norm of $\boldsymbol{\xi}_V$:

$$\begin{aligned} C \|\boldsymbol{\xi}_V\|_{\mathrm{DG}}^2 &\leq a_h(\boldsymbol{\xi}_V, \boldsymbol{\xi}_V) \\ &= a_h(\boldsymbol{e} - (\boldsymbol{\eta} + \boldsymbol{\xi}_X), \boldsymbol{\xi}_V) \\ &\leq |a_h(\boldsymbol{e}, \boldsymbol{\xi}_V)| + |a_h(\boldsymbol{\eta} + \boldsymbol{\xi}_X, \boldsymbol{\xi}_V)| \\ &= |a_h(\boldsymbol{\eta} + \boldsymbol{\xi}_X, \boldsymbol{\xi}_V)|, \end{aligned} \tag{4.2.6}$$

where Galerkin orthogonality has been used.

The usual bounding strategy then leads to

$$C \|\boldsymbol{\xi}_V\|_{\mathrm{DG}}^2 \leq \|\boldsymbol{\xi}_V\|_{\mathrm{DG}} \phi(\boldsymbol{\eta} + \boldsymbol{\xi}_X). \tag{4.2.7}$$

If, then, appropriate bounds could be found on $\boldsymbol{\eta} + \boldsymbol{\xi}_X$, rather than on $\boldsymbol{\eta}$ on its own, bounding would proceed smoothly.

This process is equivalent to splitting the approximation error with a function in V_h, in this case the direct projection of the interpolant onto V_h, and avoids the problem arising in (4.2.5).

Therefore, needing a splitting-function $\mathbf{u}_P \in V_h$, and not having a suitable available interpolant, consider instead using a projection of an interpolant onto V_h or a subspace of V_h as $\mathbf{u}_P$.

Specifically, consider an interpolant $\pi\mathbf{u} \in X$, where X contains $\mathbb{P}_1$ locally, and with Π the direct projection onto the space of element-wise linear polynomials, define $\mathbf{u}_P$ by

$$\mathbf{u}_P := \Pi \circ \pi\mathbf{u}. \tag{4.2.8}$$

Then

$$\pi\mathbf{u} = \mathbf{u}_P + \mathbf{u}_X,$$

where, element-wise,

$$\mathbf{u}_P \in \mathbb{P}_1,$$
$$\mathbf{u}_X \in X \setminus \mathbb{P}_1.$$

The choice of $\mathbb{P}_1$ rather than $\mathbb{Q}_1$ as the local space for $\mathbf{u}_P$ is both for simplicity, as it excludes cross-terms of the components, and to allow for the use of a solution space of linear approximations as well as one of bilinear approximations.

Now, the first requirement, $\mathbf{u}_P \in V_h$, holds, and the rest need to be satisfied by an appropriate choice of the underlying interpolant $\pi\mathbf{u}$.

4.3 Desired properties of $\pi\mathbf{u}$

The first step in finding a suitable underlying interpolant $\pi\mathbf{u}$ is to identify sufficient conditions on $\pi\mathbf{u}$ for the requirements on $\mathbf{u}_P$ to be satisfied.

4.3.1 To obtain the basic error bounds

For $\mathbf{u}_P$ to satisfy the basic error bounds (3.2.14), that is,

$$|\mathbf{u} - \Pi \circ \pi \mathbf{u}|_{H^m(\Omega_e)} \leq C h_e^{2-m} |\mathbf{u}|_{H^2(\Omega_e)} \tag{4.3.1}$$

for $m = 0, 1, 2$, it is required by a theorem in Ciarlet [15] (see A.5 for details) that $\Pi \circ \pi$ preserve first-order polynomials locally. This holds for the composition if it holds for both Π and π. By definition, Π satisfies the requirement. For π to satisfy it, a necessary condition is that $\pi\mathbf{u}$ should lie in a space fully containing $\mathbb{P}_1$, element-wise.

4.3.2 To obtain the error bounds on the divergence

The bounds on the divergence, (3.2.15), can be expressed as

$$|\nabla \cdot \mathbf{u} - \nabla \cdot \mathbf{u}_P|_{H^m(\Omega_e)} \leq C h_e^{1-m} |\nabla \cdot \mathbf{u}|_{H^1(\Omega_e)} \tag{4.3.2}$$

for $m = 0, 1$.

Define Π_C to be the L^2-orthogonal projection operator onto constants on the element. If the identity

$$\nabla \cdot \mathbf{u}_P = \Pi_C \nabla \cdot \mathbf{u} \tag{4.3.3}$$

holds, then because Π_C is a constant-preserving projection operator, these bounds will hold ([15], see A.5).

Since $\nabla \cdot \mathbf{u}_P \in \mathbb{P}_0(\Omega_e)$, (4.3.3) is equivalent to

$$\Pi_C \nabla \cdot \mathbf{u}_P = \Pi_C \nabla \cdot \mathbf{u}, \tag{4.3.4}$$

which will hold if

$$\int_{\Omega_e} \nabla \cdot (\mathbf{u} - \mathbf{u}_P)\, dx = 0. \tag{4.3.5}$$

If $\mathbf{u}_P$ is continuous within the closure of each element Ω_e, then by the divergence theorem (4.3.5) is equivalent to

$$\int_{\partial\Omega_e} (\mathbf{u} - \mathbf{u}_P) \cdot \mathbf{n}\, ds = 0. \tag{4.3.6}$$

This would follow from an interpolant $\pi\mathbf{u}$ being defined such that both

$$\int_{\partial\Omega_e} (\mathbf{u} - \pi\mathbf{u}) \cdot \mathbf{n}\, ds = 0 \tag{4.3.7}$$

and

$$\int_{\partial\Omega_e} (\pi\mathbf{u} - \Pi \circ \pi\mathbf{u}) \cdot \mathbf{n}\, ds = 0 \tag{4.3.8}$$

are satisfied. This final equation states that the normal components of the higher order terms of the interpolant vanish when integrated around an element boundary.

In summary, if an interpolant $\pi\mathbf{u}$ is defined such that (4.3.7) and (4.3.8) hold and $\Pi \circ \pi\mathbf{u}$ is continuous within the closure of each element Ω_e, then the desired bounds (3.2.15) will be satisfied.

4.3.3 To satisfy equation (4.1.1)

The final requirement on $\mathbf{u}_P$ will be fulfilled if, on every edge E of every element,

$$\int_E (\mathbf{u} - \pi\mathbf{u}) \cdot \mathbf{n}\, ds = 0 \tag{4.3.9}$$

and

$$\int_E (\pi\mathbf{u} - \Pi \circ \pi\mathbf{u}) \cdot \mathbf{n}\, ds = 0. \tag{4.3.10}$$

Moreover, if these hold, then (4.3.7) and (4.3.8) of §4.3.2 follow automatically by summing over the edges of each element.

4.3.4 Summary of requirements on the interpolant

In summary, sufficient conditions on the interpolant are that:

1. π preserves $\mathbb{P}_1$ in each component;
2. $\Pi \circ \pi\mathbf{u}$ is continuous on the closure of each element;
3. $\pi\mathbf{u}$ satisfies

$$\int_E (\mathbf{u} - \pi\mathbf{u}) \cdot \mathbf{n} = 0 \tag{4.3.11}$$

and

$$\int_E (\pi\mathbf{u} - \Pi \circ \pi\mathbf{u}) \cdot \mathbf{n} = 0. \tag{4.3.12}$$

4.4 Finding a satisfactory interpolant

Having ascertained what properties are desirable in an interpolant, that is, sufficient conditions so that the requirements on the linear portion, $\mathbf{u}_P$, will be satisfied, an interpolant with these properties needs to be identified, either found or constructed.

4.4.1 Existing interpolants

One interpolant that has properties closely matching those described in §4.3 is the $BDM_{[1]}$ interpolant ([11]). The local basis is a direct sum of the space of linear polynomials in each component, and two curl functions (introducing coupling between the components). The linear projection of the interpolant satisfies the basic bounds (3.2.14). The equation (4.3.7) holds on each element boundary. By the property of the curl, the higher-order terms are locally divergence-free, so that (4.3.8) is satisfied by application of the divergence theorem. However, while these two integrals vanish over entire boundaries of elements, they do not vanish edge-wise as in (4.3.9) and (4.3.10). Therefore, while the divergence bounds (3.2.15) hold, the final property, which would hopefully allow for an a λ-independent convergence result after the proposed modifications have been made, is absent.

The elements defined by Douglas et al. [22] have the Crouzeix-Raviart-like property that the mean edge values are equal to the corresponding edge midpoint values, which leads to a useful orthogonality property. These elements are considered in detail in the upcoming section, and will be drawn on for constructing the interpolant that will ultimately be used in the convergence analysis.

4.4.2 The elements of Douglas, Santos, Sheen and Ye (DSSY)

The elements described by Douglas et al. in [22] (one of which is also used in [14]) assume no coupling between the components, each of which lies in

$$\hat{Q} = \text{span}\left(\{1, x, y, \Theta\left(x, y\right)\}\right) \tag{4.4.1}$$

on the reference element $[-1,1]^2$, and can be written in the form

$$\pi\mathbf{u} = \begin{pmatrix} a + bx + cy + d\,\Theta(x,y) \\ e + fx + gy + h\,\Theta(x,y) \end{pmatrix} \tag{4.4.2}$$

The elements are formed from two possibilities for $\Theta(x,y)$, which are motivated as variations of the rotated multilinear nonconforming element of Rannacher and Turek [32], for which $\Theta(x,y) := x^2 - y^2$.

Properties of the DSSY element

The element(s) by Rannacher and Turek can be defined either by nodal values being the mean values of the function along each of the edges, or by nodal values at midpoints of the edges. Either definition gives unisolvence, but the spaces they lead to are different. The element lacks certain orthogonality properties which Douglas et al. show are critical to show an extension of the optimal-order convergence result to general quadrilaterals. This led them to develop two alternative elements by redefining $\Theta(x,y) := \theta(x) - \theta(y)$ with

$$\theta_1(x) = x^2 - \frac{5}{3}x^4 \quad \text{and} \tag{4.4.3}$$

$$\theta_2(x) = x^2 - \frac{25}{6}x^4 + \frac{7}{2}x^6. \tag{4.4.4}$$

For either of these, nodal values can be defined at the edge midpoints $\mathbf{m}_i$ $(i = 1, ..., 4)$ with unisolvence, resulting in the basis functions $\phi_i(x,y)$. (All calculations are shown for the reference element, and edges are numbered anticlockwise, starting with the edge with midpoint $(0,-1)$.)

Calculations show that for either choice of θ,

$$\int_{-1}^{1} \theta(x)\,dx = 0 \tag{4.4.5}$$

(which does not hold for Rannacher and Turek's element, where $\theta(x) := x^2$), and

$$\int_{-1}^{1} x\,\theta(x)\,dx = 0; \tag{4.4.6}$$

that is, $\theta(x)$ is orthogonal to linear polynomials.

From (4.4.5),

$$\begin{aligned} &\int_{-1}^{1} (\theta(x) - \theta(y))\,|_{y=\pm 1} dx = -2\,\theta(\pm 1)\,, \\ \Longrightarrow \quad &\tfrac{1}{2}\int_{-1}^{1} \Theta(x, \pm 1)\; dx = -\theta(\pm 1)\,, \end{aligned} \tag{4.4.7}$$

the mean values of the edges with midpoints $(0, \pm 1)$.

Since $\theta(0) = 0$, the values at midpoints $(0, \pm 1)$ are

$$\Theta(0, \pm 1) = \theta(0) - \theta(\pm 1) = -\,\theta(\pm 1)\,. \tag{4.4.8}$$

Thus, the midpoint value of the higher-order function is equal to its mean value along each edge, a property not satisfied by the rotated elements of Rannacher and Turek. The same will hold for the edges with midpoints $(\pm 1, 0)$.

Since the remaining terms in the space are linear, the midpoint value of the interpolant will equal the mean value along each edge, so that for each basis function $\phi_i(x, y)$,

$$\phi_i(\mathbf{m}_j) = \frac{1}{h_{E^j}} \int_{E^j} \phi_i(x, y)\, ds = \delta_{ij}. \tag{4.4.9}$$

Let $Q(\Omega_e)$ be the mapping of the space $\hat{Q}$ onto element Ω_e in the real domain.

If an operator $\pi_e : H^1(\Omega_e) \to Q(\Omega_e)$ is defined by

$$\int_{E^j} \pi_e v\, ds = \int_{E^j} v\, ds \tag{4.4.10}$$

for $j = 1, ..., 4$, then, by the preservation of (4.4.9) under affine mapping,

$$\pi_e v(\mathbf{m}_j) = \frac{1}{h_{E^j}} \int_{E^j} \pi_e v(x, y)\, ds = \frac{1}{h_{E^j}} \int_{E^j} v(x, y)\, ds \tag{4.4.11}$$

for each edge $E^j,\ j = 1.., 4$.

Let NC_h be the nonconforming space resulting from the mapping of the given basis to each element in the domain, and requiring continuity at midpoints of interior edges and vanishing at midpoints of boundary edges.

Extending the operator to a global operator $\pi : H_0^1(\Omega) \to NC_h$ such that

$$\pi v|_{\Omega_e} = \pi_e v, \tag{4.4.12}$$

one has

$$\int_E \pi v \, ds = \int_E v \, ds \qquad \forall \; E \in \Gamma. \tag{4.4.13}$$

Additional properties

Referring to the requirements in §4.3: the interpolants described are from a space greater than $\mathbb{P}_1$, and give continuous functions πv on element closures. The next point to consider is what properties a projection of πv onto $\mathbb{P}_1$ has.

Define $\bar{\Pi}$ as the L_2-orthogonal projection onto the reference element, that is, for any linear polynomial p_1,

$$\int_{-1}^{1}\int_{-1}^{1} \bar{\Pi} w \; p_1 \; dx \, dy = \int_{-1}^{1}\int_{-1}^{1} w \; p_1 \; dx \, dy, \tag{4.4.14}$$

so that, specifically,

$$\int_{-1}^{1}\int_{-1}^{1} \bar{\Pi} \circ \pi v \; p_1 \; dx \, dy = \int_{-1}^{1}\int_{-1}^{1} \pi v \; p_1 \; dx \, dy. \tag{4.4.15}$$

Since $\theta(x)$ is orthogonal to linear polynomials,

$$\int_{-1}^{1}\int_{-1}^{1} (\theta(x) - \theta(y)) \; p_1 \; dx \, dy = 0, \tag{4.4.16}$$

and consequently $\bar{\Pi} \circ \pi v$ is simply the linear part of πv. This implies that $\bar{\Pi} = \Pi$, and $\pi v - \Pi \circ \pi v = k \, \Theta(x, y)$ for some constant k.

Evaluation of $\int_E (\pi v - \Pi \circ \pi v) \; ds = \int_E k \, \Theta(x, y) \; ds$ on each edge E of the reference element gives *nonzero* results on each edge, these results totalling zero around the boundary of the element.

That is,

$$\int_E (\pi v - \Pi \circ \pi v) \; ds \neq 0 \tag{4.4.17}$$

but

$$\int_{\partial\Omega_e} (\pi v - \Pi \circ \pi v) \; ds = 0, \tag{4.4.18}$$

so that (4.3.8) is satisfied, but (4.3.10) is not.

Therefore, once again, as in the case of the projection of the interpolant $BDM_{[1]}$, all the necessary bounds are valid, but the final property to be used in obtaining λ-independence is absent.

4.4.3 Designing a new basis

While the DSSY elements do not have satisfactory properties for the purpose of the convergence analysis, their construction provides useful insights. Specifically, it is useful to have an interpolant with mean edge values equal to edge midpoint values. On the other hand, what is necessary (and missing) is that the normal components of the higher-order terms have vanishing mean values on each edge.

Taking all of this into consideration, a new interpolant is designed specifically to satisfy the required properties. The reference element $[-1,1]^2$ is used for all calculations unless otherwise noted. The process of design begins with identifying the necessary properties of the basis:

1. Begin with the requirement that the interpolant lies locally in

$$\text{span}\left(\left\{\begin{array}{c} 1, x, y, \Theta(x,y) \\ 1, x, y, \Psi(x,y) \end{array}\right\}\right), \tag{4.4.19}$$

   with the distinction between components to allow for differing bases if necessary.

2. To ensure unisolvence for midpoint nodal values, $\Theta(\mathbf{m}_i) \neq 0$ for at least one of $i = 1, .., 4$, that is, it must not vanish at all midpoints; similarly Ψ must not vanish at all midpoints.

   This requires that $\Theta(x,y)$ and $\Psi(x,y)$ must each not be a product of x and y. The forms

$$\begin{aligned} \Theta(x,y) &= \theta(x) + \kappa(y), \\ \Psi(x,y) &= \rho(x) + \chi(y) \end{aligned} \tag{4.4.20}$$

   are therefore suitable.

3. For equivalence of the midpoint value and mean value on each edge (independently satisfied by the linear terms in each component):

$$\begin{aligned}\Theta(\mathbf{m}_i) &= \tfrac{1}{2}\int_{E_i}\Theta(x,y)\,ds\\ \Psi(\mathbf{m}_i) &= \tfrac{1}{2}\int_{E_i}\Psi(x,y)\,ds\end{aligned} \tag{4.4.21}$$

for $i = 1, .., 4$.

4. For the integral of the normal component of the nonlinear terms in the function to vanish on each edge:

$$\begin{aligned}&\int_{-1}^{1}\Psi(x,\pm 1)\,dx = 0\\ \text{and}\quad &\int_{-1}^{1}\Theta(\pm 1,y)\,dx = 0,\end{aligned} \tag{4.4.22}$$

which is equivalent to requiring

$$\begin{aligned}\Theta(\mathbf{m}_2) = 0 = \Theta(\mathbf{m}_4),\\ \Psi(\mathbf{m}_1) = 0 = \Psi(\mathbf{m}_3).\end{aligned} \tag{4.4.23}$$

After calculating the specifics, the requirements simplify to

$$\chi(1) = \chi(-1) = -\tfrac{1}{2}\int_{-1}^{1}\rho(x)\,dx = -\rho(0), \tag{4.4.24a}$$

$$\theta(1) = \theta(-1) = -\tfrac{1}{2}\int_{-1}^{1}\kappa(y)\,dy = -\kappa(0), \tag{4.4.24b}$$

midpoint-mean equality, and that Θ and Ψ do not vanish at all midpoints.

The similarities between (4.4.24a) and (4.4.24b) suggest choosing $\chi = \theta$ and $\rho = \kappa$, so that

$$\begin{aligned}\Theta &= \theta(x) + \kappa(y),\\ \Psi &= \kappa(x) + \theta(y).\end{aligned} \tag{4.4.25}$$

The requirements now become

$$\theta(1) = \theta(-1) = -\kappa(0) = -\tfrac{1}{2}\int_{-1}^{1} \kappa(x)\,dx \tag{4.4.26a}$$

$$\theta(0) = \tfrac{1}{2}\int_{-1}^{1} \theta(x)\,dx \tag{4.4.26b}$$

$$\kappa(1) \neq -\theta(0), \kappa(-1) \neq -\theta(0). \tag{4.4.26c}$$

These conditions are all satisfied by

$$\theta(x) := 3x^2 - 10x^4 + 7x^6 \tag{4.4.27a}$$

$$\kappa(x) := -3x^2 + 5x^4. \tag{4.4.27b}$$

(Without the third term in $\theta(x)$, the requirements cannot be satisfied. Certainly neither $\theta = \kappa$ nor $\theta = -\kappa$, as used in the DSSY elements, is an option. The distinction between Θ and Ψ is necessary, because if they were identical, requiring the specified edge integrals to vanish would be equivalent to requiring $\Theta = \Psi$ to vanish at all midpoints – because of the equivalence between the mean values and the midpoint values – and unisolvence would be lost.)

4.4.4 The new interpolant

The new interpolant is now constructed by the same procedure as in [22].

Review of basic properties on the reference element

Begin with a space on the reference element $[-1,1]^2$,

$$\hat{Q} = \operatorname{span}\left(\left\{\begin{array}{c} 1, x, y, \Theta(x,y) \\ 1, x, y, \Psi(x,y) \end{array}\right\}\right), \tag{4.4.28}$$

where

$$\Theta(x,y) := \theta(x) + \kappa(y), \tag{4.4.29}$$

$$\Psi(x,y) := \kappa(x) + \theta(y), \tag{4.4.30}$$

$$\theta(x) := 3x^2 - 10x^4 + 7x^6, \tag{4.4.31}$$

$$\kappa(x) := -3x^2 + 5x^4. \tag{4.4.32}$$

Calculation shows that

$$\begin{aligned} \theta(0) &= 0 = \kappa(0), \\ \theta(\pm 1) &= 0, \ \kappa(\pm 1) = 2, \end{aligned} \tag{4.4.33}$$

so that

$$\begin{aligned} \Theta(\mathbf{m}_1) &= \Theta(\mathbf{m}_3) = 2, \ \Theta(\mathbf{m}_2) = \Theta(\mathbf{m}_4) = 0, \\ \Psi(\mathbf{m}_1) &= \Psi(\mathbf{m}_3) = 0, \ \Psi(\mathbf{m}_2) = \Psi(\mathbf{m}_4) = 2. \end{aligned} \tag{4.4.34}$$

Define the nodal values at the edge midpoints $\mathbf{m}_i$ $(i = 1, ..., 4)$. There are unique sets of resulting basis functions $\phi_i^x(x, y)$ and $\phi_i^y(x, y)$ for the x and y components respectively (see B.1.1).

Further calculation gives

$$\begin{aligned} \int_{-1}^{1} \theta(x)\, dx &= 0, \\ \int_{-1}^{1} \kappa(x)\, dx &= 0, \end{aligned} \tag{4.4.35}$$

and

$$\begin{aligned} \tfrac{1}{2} \int_{E^i} \Theta(x, y)\, ds &= \Theta(\mathbf{m}_i), \\ \tfrac{1}{2} \int_{E^i} \Psi(x, y)\, ds &= \Psi(\mathbf{m}_i), \end{aligned} \tag{4.4.36}$$

for $i = 1, .., 4$.

Therefore, on each edge E_j,

$$\phi_i^x(\mathbf{m}_j) = \frac{1}{h_{E^j}} \int_{E^j} \phi_i^x(x, y)\, ds \tag{4.4.37}$$

for $i = 1, .., 4$, and similarly for all ϕ_i^y.

The interpolant at the global level

Let $Q(\Omega_e)$ be the mapped space $\hat{Q}$ onto Ω_e, and its components be denoted by $Q^x(\Omega_e)$ and $Q^y(\Omega_e)$.

Let NC_h be the nonconforming space resulting from the mapping of $\hat{Q}$ to each element in the domain, with the requirement of continuity at midpoints of interior edges and equality to the edge mean of the Dirichlet value $\mathbf{g}$ at the midpoints of Dirichlet boundary edges. Let NC_h^x and NC_h^y denote the spaces of the x and y components of NC_h respectively.

Define the operator $\pi_e^x : H^1(\Omega_e) \to Q^x(\Omega_e)$ by

$$\int_{E^j} \pi_e^x v \, ds = \int_{E^j} v \, ds \tag{4.4.38}$$

for $j = 1, ..., 4$.

Extending the operator π_e^x to a global operator $\pi^x : H^1_{\Gamma_D^x}(\Omega) \to NC_h^x$ such that

$$\pi^x v|_{\Omega_e} = \pi_e^x v, \tag{4.4.39}$$

one has

$$\int_E \pi^x v \, ds = \int_E v \, ds \qquad \forall \; E \in \Gamma. \tag{4.4.40}$$

Similarly, defining the analogous operators π_e^y and π^y, one has

$$\int_E \pi^y v \, ds = \int_E v \, ds \qquad \forall \; E \in \Gamma. \tag{4.4.41}$$

Since the equivalence between midpoint and mean values on edges is preserved under affine mapping,

$$\begin{aligned} \pi^x v(\mathbf{m}_i) &= \frac{1}{h_{E^i}} \int_{E^i} v \, ds, \\ \pi^y v(\mathbf{m}_i) &= \frac{1}{h_{E^i}} \int_{E^i} v \, ds \end{aligned} \tag{4.4.42}$$

hold for each edge E^i of Ω_e.

Define $\pi : H^1_{\Gamma_D^x}(\Omega) \times H^1_{\Gamma_D^y}(\Omega) \to NC_h$ as the concatenation of π^x and π^y, so that

$$\pi \mathbf{u} = \begin{bmatrix} \pi^x u_x \\ \pi^y u_y \end{bmatrix}. \tag{4.4.43}$$

4.5 The error-splitting function $\mathbf{u}_P$: the projection of the constructed interpolant

The interpolant in the preceding section (§4.4.4) was designed specifically to satisfy sufficient conditions (summarised in §4.3.4) for its projection to be used as an error-splitting function $\mathbf{u}_P$ that would have the properties described in §4.1. The outline below of the deduction of these properties, beginning with the interpolant and its projection, confirms that this end has been achieved.

The interpolant $\pi\mathbf{u}$ and its projection $\Pi \circ \pi\mathbf{u}$ onto the space of element-wise linear polynomials are defined as in the preceding sections.

1. By definition of Π, $\mathbf{u}_P = \Pi \circ \pi\mathbf{u}$ lies in V_h.

2. The operator π preserves Q on each element, and therefore, because $[\mathbb{P}_1]^2 \subset Q$, preserves linear polynomials on each element. By definition, Π preserves linear polynomials on each element, therefore the composition does so. The bounds

$$\|\mathbf{u} - \mathbf{u}_P\|_{L^2(\Omega_e)} \leq C h_e^2 \, |\mathbf{u}|_{H^2(\Omega_e)} \,, \tag{4.5.1a}$$

$$|\mathbf{u} - \mathbf{u}_P|_{H^1(\Omega_e)} \leq C h_e \, |\mathbf{u}|_{H^2(\Omega_e)} \,, \tag{4.5.1b}$$

$$|\mathbf{u} - \mathbf{u}_P|_{H^2(\Omega_e)} \leq C \, |\mathbf{u}|_{H^2(\Omega_e)} \,, \tag{4.5.1c}$$

   follow (as detailed in §4.3.1).

3. Writing, on the reference element,

$$\pi\mathbf{u} = \begin{pmatrix} a + bx + cy + d\,\Theta\,(x,y) \\ e + fx + gy + h\,\Psi\,(x,y) \end{pmatrix}, \tag{4.5.2}$$

   it follows that

$$\Pi \circ \pi\mathbf{u} = \begin{pmatrix} a + bx + cy \\ e + fx + gy \end{pmatrix}. \tag{4.5.3}$$

   Denote the difference between these two vectors by

$$\mathbf{q} = \pi\mathbf{u} - \Pi \circ \pi\mathbf{u} = \begin{pmatrix} d\,\Theta\,(x,y) \\ h\,\Psi\,(x,y) \end{pmatrix}. \tag{4.5.4}$$

Of specific interest are the integrals of $\mathbf{q}\cdot\mathbf{n}$ on individual edges of the element. These are

$$\begin{aligned}
\int_{E_1}\mathbf{q}\cdot\mathbf{n}\,ds &= -h\int_{-1}^{1}\Psi\left(x,-1\right)dx,\\
\int_{E_2}\mathbf{q}\cdot\mathbf{n}\,ds &= d\int_{-1}^{1}\Theta\left(1,y\right)dy,\\
\int_{E_3}\mathbf{q}\cdot\mathbf{n}\,ds &= h\int_{-1}^{1}\Psi\left(x,1\right)dx,\\
\int_{E_4}\mathbf{q}\cdot\mathbf{n}\,ds &= -d\int_{-1}^{1}\Theta\left(-1,y\right)dy,
\end{aligned}\tag{4.5.5}$$

which all vanish (see the calculations in §4.4.4). That is, on each E, (4.3.10) holds. By the definition of π,

$$\int_{E}\left(\mathbf{u}-\pi\mathbf{u}\right)\cdot\mathbf{n}\,ds = 0,\tag{4.5.6}$$

that is, (4.3.9), holds. Therefore

$$\int_{E}\left(\mathbf{u}-\mathbf{u}_P\right)\cdot\mathbf{n}\,ds = 0,\tag{4.5.7}$$

so that (4.1.1) is satisfied, which is desirable for λ-independent convergence.

4. Moreover, by summing the integrals over edges of the element,

$$\int_{\partial\Omega_e}\left(\mathbf{u}-\mathbf{u}_P\right)\cdot\mathbf{n}\,ds = 0.\tag{4.5.8}$$

Because $\mathbf{u}_P$ is continuous on the closure of the element, by the divergence theorem the bounds on $\nabla\cdot\left(\mathbf{u}-\mathbf{u}_P\right)$ are obtained (as detailed in §4.3.2), viz.

$$\left\|\nabla\cdot\left(\mathbf{u}-\mathbf{u}_P\right)\right\|_{L^2(\Omega_e)} \leq Ch_e\left|\nabla\cdot\mathbf{u}\right|_{H^1(\Omega_e)},\tag{4.5.9a}$$

$$\left|\nabla\cdot\left(\mathbf{u}-\mathbf{u}_P\right)\right|_{H^1(\Omega_e)} \leq C\left|\nabla\cdot\mathbf{u}\right|_{H^1(\Omega_e)}.\tag{4.5.9b}$$

Thus $\mathbf{u}_P$ satisfies all the properties assumed and suggested by the preliminary analysis in the preceding chapter.

Chapter 5

Remedies

This chapter is concerned with remedies for the problematic λ-dependent error bound obtained in Chapter 3 for the IP methods used with bilinear elements. The first remedy proposed and investigated is under-integration of specific edge terms in the formulation; its effects on coercivity and consistency are accounted for in order to obtain, finally, a proof of uniform convergence with respect to λ when the necessary modifications are imposed. The use of linear approximations, instead of bilinear, is then investigated as an alternative possible remedy, as is a combination of linear approximations and under-integration, and the cases leading to uniform convergence for each of the three IP methods are identified.

5.1 The error bound for bilinear elements

With the error-splitting function $\mathbf{u}_P$ having been defined and satisfying the assumptions of the preliminary analysis in Chapter 3, the λ-dependent error bound of that analysis is confirmed. That is, with bilinear elements, the approximation error of each of the three IP methods is bounded as

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq C\lambda\, h, \tag{5.1.1}$$

indicating that, in the incompressible limit, poor approximations and perhaps locking-type behaviour should be expected.

The terms IV and $VIII$ defined in (3.2.27) contribute specifically to the λ-dependence, as detailed in §3.2.5.

5.2 Under-integration (UI)

From the preliminary analysis of the IP formulation in the context of bilinear elements, it was seen that a projection of the factors of certain integrands onto the space of constants might lead to the desired, λ-independent error bound. Arnold [2] shows that this type of modification to a formulation is equivalent to the use of under-integration in the numerical implementation. Therefore, the first proposed remedy for the IP methods used with bilinear elements is under-integration of specific edge terms in the formulation, beginning with those identified as contributing to λ-dependence in the unmodified case.

While historically this technique has most commonly been used in conjunction with conforming methods, with selected integrals over the element domains being under-integrated, here it will be integrals on *edges only*, not on element domains, that are selected.

A similar application of under-integration was, effectively (although not under that description), employed by Hansbo and Larson [25], in an SIPG-like method for linear simplicial elements, where a λ-dependent stabilization term was used with projections onto constants. In their previous method, in [24], the λ-dependent stabilization term involved only the normal components of jumps. In [25], however, the authors use the L^2-orthogonal projections onto constants of the full jumps as part of their stabilization, effectively applying under-integration to a term that would otherwise be highly constraining for nearly incompressible materials.

5.2.1 Incorporating under-integration into the DG formulation

Under-integration (or reduced integration) refers to the technique of using, in the numerical evaluation of an integral, a lower order of quadrature than would be necessary for an exact result. Arnold [2] shows for one dimension that lowering the order of integrands using projection operators is equivalent to applying under-integration. The details of

the proof in [2] (cf. [33]) are described here for the case of integrating a product of linear polynomials using only a single quadrature point.

Given $f \in \mathbb{P}_1$, define $\tilde{f}$ as the constant interpolant based on the integration point x_{IP} of a one-point integration rule. Define $I_1\{\phi\}$ as numerical integration of ϕ over the appropriate domain using a one-point rule, and Π_0 as the L_2-orthogonal projection operator onto constants on the interval. Let $g \in \mathbb{P}_0$ be arbitrary.

$$\begin{aligned}
\int \tilde{f} g \, dx &= I_1\left\{\tilde{f} g\right\} && \text{since the integrand is constant,} \\
&= \tilde{f}(x_{\text{IP}})\, g(x_{\text{IP}})\, w && \text{where } w \text{ is the weight,} \\
&= f(x_{\text{IP}})\, g(x_{\text{IP}})\, w && \text{by definition of } \tilde{f}, \\
&= I_1\{fg\} \\
&= \int fg \, dx && \text{since this integrand is linear,} \\
&= \int (\Pi_0 f)\, g \, dx && \text{since } g \text{ is constant.}
\end{aligned} \tag{5.2.1}$$

Thus

$$\begin{aligned}
& \int \tilde{f} g \, dx = \int (\Pi_0 f)\, g \, dx \\
\implies & \int \left(\tilde{f} - \Pi_0 f\right) g \, dx = 0 \\
\implies & \Pi_0\left(\tilde{f} - \Pi_0 f\right) = 0 \\
\implies & \tilde{f} = \Pi_0 f.
\end{aligned} \tag{5.2.2}$$

Then, given $f, h \in \mathbb{P}_1$, with constant interpolants $\tilde{f}$ and $\tilde{h}$ respectively,

$$\begin{aligned}
\int (\Pi_0 f)(\Pi_0 h)\, dx &= \int \tilde{f}\tilde{h} \, dx \\
&= I_1\left\{\tilde{f}\tilde{h}\right\} \\
&= \tilde{f}(x_{\text{IP}})\, \tilde{h}(x_{\text{IP}})\, w \\
&= f(x_{\text{IP}})\, h(x_{\text{IP}})\, w \\
&= I_1\{fh\}
\end{aligned} \tag{5.2.3}$$

by definition of one-point integration.

Therefore, using L_2-projections onto the space of constants is equivalent to using one-point integration, for products of functions in $\mathbb{P}_1$.

5.2.2 Choice of terms for under-integration: a first guess

The terms to be considered for under-integration are firstly those that lead to undesirable λ-dependence in the error bound, that is IV and $VIII$ of (3.2.27).

Referring to the preliminary analysis, with $\mathbf{u}_P$ as defined in Chapter 4, the term IV is replaced with

$$\begin{aligned} IV^{\mathrm{UI}} &= \lambda \left| \theta \sum_{E\in\Gamma_{iD}} \int_E \Pi_0[\![\boldsymbol{\gamma}]\!]\ \Pi_0\{\!\{\nabla\cdot\boldsymbol{w}\}\!\}\, ds \right| \\ &= \lambda \left| \theta \sum_{E\in\Gamma_{iD}} \int_E [\![\boldsymbol{\gamma}]\!]\ \Pi_0\{\!\{\nabla\cdot\boldsymbol{w}\}\!\}\, ds \right| \\ &= \lambda \left| \theta \sum_{E\in\Gamma_{iD}} \Pi_0\{\!\{\nabla\cdot\boldsymbol{w}\}\!\}|_E \int_E [\![\boldsymbol{\gamma}]\!]\, ds \right|. \end{aligned} \tag{5.2.4}$$

By (4.1.1),

$$\int_E [\![\boldsymbol{\gamma}]\!]\, ds = 0 \tag{5.2.5}$$

whether E is an interior or a boundary edge, so that $IV^{\mathrm{UI}} = 0$, and can have no contribution to unwanted λ-dependence.

Similarly, the term $VIII$ is replaced with

$$\begin{aligned} VIII^{\mathrm{UI}} &= k_\lambda\,\lambda \left| \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \int_E \Pi_0\,[\![\boldsymbol{\gamma}]\!]\,\Pi_0\,[\![\boldsymbol{w}]\!]\, ds \right| \\ &= k_\lambda\,\lambda \left| \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \int_E [\![\boldsymbol{\gamma}]\!]\,\Pi_0\,[\![\boldsymbol{w}]\!]\, ds \right| \\ &= k_\lambda\,\lambda \left| \sum_{E\in\Gamma_{iD}} \frac{1}{h_E}\,\Pi_0\,[\![\boldsymbol{w}]\!]|_E \int_E [\![\boldsymbol{\gamma}]\!]\, ds \right|, \end{aligned} \tag{5.2.6}$$

which again vanishes. Thus $VIII^{\mathrm{UI}}$ also can contribute nothing to undesirable λ-dependence.

Thus under-integration could eliminate the contributions of the two problematic terms from the error bound. However, since this procedure involves modification of the IP formulation itself, the coercivity of the modified bilinear form and consistency of the modified formulation need to be considered as well.

5.2.3 Coercivity and a further modification

Coercivity of the modified bilinear form is dealt with for each IP method individually. As necessary, variations are considered for each method, to ascertain the effect of selecting different terms for under-integration.

NIPG

The original bilinear form used for the coercivity proof for NIPG is

$$a_h(\mathbf{v},\mathbf{v}) = 2\mu \sum_{\Omega_e\in\mathcal{T}_h} \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)} + \lambda \sum_{\Omega_e\in\mathcal{T}_h} \|\nabla\cdot\mathbf{v}\|^2_{L^2(\Omega_e)} + k_\mu\mu \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \|\{\!\{\mathbf{v}\}\!\}\|^2_{L^2(E)} + k_\lambda\lambda \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|^2_{L^2(E)}\,. \quad (5.2.7)$$

Under-integrating the two terms corresponding to IV and $VIII$ respectively gives the less simple

$$a_h^{\mathrm{UI}}(\mathbf{v},\mathbf{v}) = 2\mu \sum_{\Omega_e\in\mathcal{T}_h} \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)} + \lambda \sum_{\Omega_e\in\mathcal{T}_h} \|\nabla\cdot\mathbf{v}\|^2_{L^2(\Omega_e)} + \lambda \sum_{E\in\Gamma_{iD}} \int_E \left(\Pi_0[\![\mathbf{v}]\!]\ \Pi_0\{\!\{\nabla\cdot\mathbf{v}\}\!\} - [\![\mathbf{v}]\!]\{\!\{\nabla\cdot\mathbf{v}\}\!\}\right) ds + k_\mu\mu \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \|\{\!\{\mathbf{v}\}\!\}\|^2_{L^2(E)} + k_\lambda\lambda \sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \|\Pi_0[\![\mathbf{v}]\!]\|^2_{L^2(E)}\,. \quad (5.2.8)$$

Here and elsewhere, define T_μ as the sum of terms involving μ, and T_λ as the sum of terms involving λ.

The μ-dependent terms are unaffected by the under-integration, and

$$\begin{aligned} T_\mu &\geq C \left(\sum_{\Omega_e \in \mathcal{T}_h} \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2_{L^2(\Omega_e)} + \tfrac{1}{2} \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \|[\![\mathbf{v}]\!]\|^2_{L^2(E)} \right) \\ &= C \|\mathbf{v}\|^2_{\mathrm{DG}} . \end{aligned} \tag{5.2.9}$$

Dealing with T_λ as in the coercivity proofs for the IIPG and SIPG methods, in §3.1.1, write it as the sum of element contributions:

$$\begin{aligned} T_\lambda = \sum_{\Omega_e \in \mathcal{T}_h} \Bigg[& \lambda \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} - \frac{\lambda}{2} \sum_{E \in \partial\Omega_e^{int}} \int_E \left([\![\mathbf{v}]\!] \{\!\{\nabla \cdot \mathbf{v}\}\!\} - \Pi_0 [\![\mathbf{v}]\!] \, \Pi_0 \{\!\{\nabla \cdot \mathbf{v}\}\!\} \right) ds \\ & - \lambda \sum_{E \in \partial\Omega_e^D} \int_E \left[(\mathbf{v} \cdot \mathbf{n}) (\nabla \cdot \mathbf{v}) - \Pi_0 (\mathbf{v} \cdot \mathbf{n}) \; \Pi_0 (\nabla \cdot \mathbf{v}) \right] ds \\ & + \frac{k_\lambda}{2} \lambda \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \|\Pi_0 [\![\mathbf{v}]\!]\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E \in \partial\Omega_e^D} \frac{1}{h_E} \|\Pi_0 \mathbf{v} \cdot \mathbf{n}\|^2_{L^2(E)} \Bigg] . \end{aligned} \tag{5.2.10}$$

By the definition of Π_0,

$$\begin{aligned} & \int_E \Pi_0 [\![\mathbf{v}]\!] \, \Pi_0 \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds = \int_E \Pi_0 [\![\mathbf{v}]\!] \, \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds \\ \implies & \int_E \left([\![\mathbf{v}]\!] \{\!\{\nabla \cdot \mathbf{v}\}\!\} - \Pi_0 [\![\mathbf{v}]\!] \, \Pi_0 \{\!\{\nabla \cdot \mathbf{v}\}\!\} \right) ds = \int_E \left([\![\mathbf{v}]\!] - \Pi_0 [\![\mathbf{v}]\!] \right) \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds. \end{aligned} \tag{5.2.11}$$

Following the same method as in §3.1.1 gives

$$\begin{aligned} & \sum_{\Omega_e \in \mathcal{T}_h} \sum_{E \in \partial\Omega_e^{int}} \int_E \left([\![\mathbf{v}]\!] - \Pi_0 [\![\mathbf{v}]\!] \right) \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds \\ & = \sum_{\Omega_e \in \mathcal{T}_h} \int_{\partial\Omega_e^{int}} \left([\![\mathbf{v}]\!] - \Pi_0 [\![\mathbf{v}]\!] \right) \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds \\ & \leq C \sum_{\Omega_e \in \mathcal{T}_h} \left\| h_E^{-\frac{1}{2}} \left([\![\mathbf{v}]\!] - \Pi_0 [\![\mathbf{v}]\!] \right) \right\|_{L^2(\partial\Omega_e^{int})} \|\nabla \cdot \mathbf{v}\|_{L^2(\Omega_e)} , \end{aligned}$$

and similarly for the Dirichlet boundary terms,

$$\begin{aligned} & \sum_{\Omega_e \in \mathcal{T}_h} \sum_{E \in \partial\Omega_e^D} \int_E \left[(\mathbf{v} \cdot \mathbf{n}) (\nabla \cdot \mathbf{v}) - \Pi_0 (\mathbf{v} \cdot \mathbf{n}) \, \Pi_0 (\nabla \cdot \mathbf{v}) \right] ds \\ & \leq C \sum_{\Omega_e \in \mathcal{T}_h} \left\| h_E^{-\frac{1}{2}} \left([\![\mathbf{v}]\!] - \Pi_0 [\![\mathbf{v}]\!] \right) \right\|_{L^2(\partial\Omega_e^D)} \|\nabla \cdot \mathbf{v}\|_{L^2(\Omega_e)} . \end{aligned}$$

Since

$$\left\| h_E^{-\frac{1}{2}} \left([\![\mathbf{v}]\!] - \Pi_0 [\![\mathbf{v}]\!] \right) \right\|_{L^2(\partial\Omega_e^{int})} \leq \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|_{L^2(\partial\Omega_e^{int})} + \left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|_{L^2(\partial\Omega_e^{int})}$$
$$\leq C \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|_{L^2(\partial\Omega_e^{int})}, \tag{5.2.12}$$

and similarly for $\partial\Omega_e^D$, it follows that

$$\begin{aligned} &- \sum_{\Omega_e \in \mathcal{T}_h} \left[\frac{1}{2} \sum_{E \in \partial\Omega_e^{int}} \int_E \left([\![\mathbf{v}]\!] \{\!\{ \nabla \cdot \mathbf{v} \}\!\} - \Pi_0 [\![\mathbf{v}]\!] \, \Pi_0 \{\!\{ \nabla \cdot \mathbf{v} \}\!\} \right) ds \right. \\ &\qquad \left. + \sum_{E \in \partial\Omega_e^D} \int_E \left((\mathbf{v} \cdot \mathbf{n}) (\nabla \cdot \mathbf{v}) - \Pi_0 (\mathbf{v} \cdot \mathbf{n}) \, \Pi_0 (\nabla \cdot \mathbf{v}) \right) ds \right] \\ \geq &- C \sum_{\Omega_e \in \mathcal{T}_h} \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|_{L^2(\partial\Omega_e^{iD})} \| \nabla \cdot \mathbf{v} \|_{L^2(\Omega_e)} \\ \geq &- C \sum_{\Omega_e \in \mathcal{T}_h} \left(\epsilon_\lambda \| \nabla \cdot \mathbf{v} \|^2_{L^2(\Omega_e)} + \frac{1}{\epsilon_\lambda} \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \right). \end{aligned} \tag{5.2.13}$$

Then

$$T_\lambda \geq \lambda \sum_{\Omega_e \in \mathcal{T}_h} \left[(1 - C\epsilon_\lambda) \| \nabla \cdot \mathbf{v} \|^2_{L^2(\Omega_e)} + \frac{k_\lambda}{2} \left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} - \frac{C}{\epsilon_\lambda} \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \right]. \tag{5.2.14}$$

The first term in brackets can shown to be positive by an appropriate choice of ϵ_λ.

However, since

$$\left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \leq \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})}, \tag{5.2.15}$$

the expression

$$\frac{k_\lambda}{2} \left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} - \frac{C}{\epsilon_\lambda} \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \tag{5.2.16}$$

can in general not be enforced as non-negative by a subsequent appropriate choice of k_λ. Consequently, T_λ cannot be enforced to be non-negative, and *coercivity of the bilinear form is therefore not established.*

It is worth noting that choosing $k_\lambda = 0$ (that is, not including the second stabilization term) or $k_\lambda > 0$ is irrelevant to this result.

On the other hand, if all three λ-dependent edge terms in (3.2.27) (that is, those corresponding to IV and VI, and $VIII$ if $k_\lambda > 0$) were under-integrated, then the first two would cancel each other out, so that

$$T_\lambda = \lambda \sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} + k_\lambda \lambda \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \|\Pi_0 [\![\mathbf{v}]\!]\|^2_{L^2(E)} \geq 0 \tag{5.2.17}$$

and the form would be coercive. This holds for $k_\lambda \geq 0$.

IIPG

As the μ-dependent terms in the formulation are unaffected by under-integration, it follows as in §3.1.1 that $T_\mu \geq C \|\mathbf{v}\|^2_{\mathrm{DG}}$, and what remains to establish coercivity is to show that $T_\lambda \geq 0$.

Since the term corresponding to IV does not appear in IIPG, under-integration of that term is not considered. Under-integration, then, of only the stabilization term gives

$$T_\lambda = \sum_{\Omega_e \in \mathcal{T}_h} \left[\lambda \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} - \frac{\lambda}{2} \sum_{E \in \partial\Omega_e^{int}} \int_E [\![\mathbf{v}]\!] \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds - \lambda \sum_{E \in \partial\Omega_e^D} \int_E (\mathbf{v} \cdot \mathbf{n}) (\nabla \cdot \mathbf{v}) \, ds \right.$$
$$\left. + \frac{k_\lambda}{2} \lambda \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \|\Pi_0 [\![\mathbf{v}]\!]\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E \in \partial\Omega_e^D} \frac{1}{h_E} \|\Pi_0 \mathbf{v} \cdot \mathbf{n}\|^2_{L^2(E)} \right]. \tag{5.2.18}$$

This again leads to

$$T_\lambda \geq \lambda \sum_{\Omega_e \in \mathcal{T}_h} \left[(1 - C\epsilon_\lambda) \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} + \frac{k_\lambda}{2} \left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} - \frac{C}{\epsilon_\lambda} \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \right] \tag{5.2.19}$$

(cf. (5.2.14)), which cannot be shown to be positive in general, and *coercivity is not established*. As in the case of NIPG, neither choosing k_λ to be positive nor choosing it to be zero will have any benefit for this result.

If, however, the term corresponding to VI is also under-integrated, then

$$\begin{aligned} T_\lambda = \sum_{\Omega_e \in \mathcal{T}_h} \Bigg[& \lambda \left\| \nabla \cdot \mathbf{v} \right\|^2_{L^2(\Omega_e)} - \frac{\lambda}{2} \sum_{E \in \partial\Omega_e^{int}} \int_E \Pi_0 [\![\mathbf{v}]\!] \Pi_0 \{\!\{ \nabla \cdot \mathbf{v} \}\!\} \, ds \\ & - \lambda \sum_{E \in \partial\Omega_e^D} \int_E \Pi_0 \left(\mathbf{v} \cdot \mathbf{n} \right) \Pi_0 \left(\nabla \cdot \mathbf{v} \right) \, ds \\ & + \frac{k_\lambda}{2} \lambda \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \left\| \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E \in \partial\Omega_e^D} \frac{1}{h_E} \left\| \Pi_0 \mathbf{v} \cdot \mathbf{n} \right\|^2_{L^2(E)} \Bigg] \\ \geq \sum_{\Omega_e \in \mathcal{T}_h} \lambda & \left[(1 - C\epsilon_\lambda) \left\| \nabla \cdot \mathbf{v} \right\|^2_{L^2(\Omega_e)} + \left(\frac{k_\lambda}{2} - \frac{C}{\epsilon_\lambda} \right) \left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \right]. \end{aligned} \tag{5.2.20}$$

The key difference between this case and the previous is that, here, every jump term is projected, so that there is one norm involving the jump terms, with one combined coefficient that can be enforced to be positive by a suitable choice of k_λ; while in the previous case, where only the jumps in the stabilization term are projected, there are two separate norms involving jump terms. In this case, then, as in §3.1.1, appropriate choices of ϵ_λ and k_λ will give $T_\lambda \geq 0$, and thus for this combination of under-integration, the bilinear form is coercive.

SIPG

Again, as the μ-dependent terms in the formulation are unaffected by under-integration, it follows as in §3.1.1 that $T_\mu \geq C \left\| \mathbf{v} \right\|^2_{\mathrm{DG}}$, and what remains to establish coercivity is to show that $T_\lambda \geq 0$.

Under-integration of the terms corresponding to IV and $VIII$ gives

$$\begin{aligned} T_\lambda = \sum_{\Omega_e \in \mathcal{T}_h} \Bigg[& \lambda \left\| \nabla \cdot \mathbf{v} \right\|^2_{L^2(\Omega_e)} - \frac{\lambda}{2} \sum_{E \in \partial\Omega_e^{int}} \int_E \left(\Pi_0 [\![\mathbf{v}]\!] \; \Pi_0 \{\!\{ \nabla \cdot \mathbf{v} \}\!\} + [\![\mathbf{v}]\!] \{\!\{ \nabla \cdot \mathbf{v} \}\!\} \right) ds \\ & - \lambda \sum_{E \in \partial\Omega_e^D} \int_E \left[\Pi_0 \left(\mathbf{v} \cdot \mathbf{n} \right) \; \Pi_0 \left(\nabla \cdot \mathbf{v} \right) + \left(\mathbf{v} \cdot \mathbf{n} \right) \left(\nabla \cdot \mathbf{v} \right) \right] ds \\ & + \frac{k_\lambda}{2} \lambda \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \left\| \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E \in \partial\Omega_e^D} \frac{1}{h_E} \left\| \Pi_0 \mathbf{v} \cdot \mathbf{n} \right\|^2_{L^2(E)} \Bigg]. \end{aligned} \tag{5.2.21}$$

The second term gives

$$\begin{aligned}
&\sum_{E\in\partial\Omega_e^{int}}\int_E \left(\Pi_0[\![\mathbf{v}]\!]+[\![\mathbf{v}]\!]\right)\{\!\{\nabla\cdot\mathbf{v}\}\!\}\,ds\\
&\leq C\left\|h_E^{-\frac{1}{2}}\left(\Pi_0[\![\mathbf{v}]\!]+[\![\mathbf{v}]\!]\right)\right\|_{L^2(\partial\Omega_e^{int})}\|\nabla\cdot\mathbf{v}\|_{L^2(\Omega_e)}\\
&\leq C\left(\left\|h_E^{-\frac{1}{2}}\Pi_0[\![\mathbf{v}]\!]\right\|_{L^2(\partial\Omega_e^{int})}+\left\|h_E^{-\frac{1}{2}}[\![\mathbf{v}]\!]\right\|_{L^2(\partial\Omega_e^{int})}\right)\|\nabla\cdot\mathbf{v}\|_{L^2(\Omega_e)}\\
&\leq C\left\|h_E^{-\frac{1}{2}}[\![\mathbf{v}]\!]\right\|_{L^2(\partial\Omega_e^{int})}\|\nabla\cdot\mathbf{v}\|_{L^2(\Omega_e)},
\end{aligned}\tag{5.2.22}$$

and the third term can be bounded similarly, so that again

$$T_\lambda\geq\sum_{\Omega_e\in\mathcal{T}_h}\lambda\left[(1-C\epsilon_\lambda)\|\nabla\cdot\mathbf{v}\|^2_{L^2(\Omega_e)}+\frac{k_\lambda}{2}\left\|h_E^{-\frac{1}{2}}\Pi_0[\![\mathbf{v}]\!]\right\|^2_{L^2(\partial\Omega_e^{iD})}-\frac{C}{\epsilon_\lambda}\left\|h_E^{-\frac{1}{2}}[\![\mathbf{v}]\!]\right\|^2_{L^2(\partial\Omega_e^{iD})}\right],\tag{5.2.23}$$

which cannot be shown to be positive in general.

The case of under-integrating all three λ-dependent terms is similar to the final case of IIPG: a lower bound is obtained in which every jump term is projected, so that the stabilization parameter can be used to give a positive coefficient, and coercivity is thus shown.

Summary of coercivity results

Lemma 1 *With the bilinear form defined by*

$$\begin{aligned}
a_h^{\mathrm{UI}}(\mathbf{u},\mathbf{v})=&\sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\boldsymbol{\sigma}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{v})\,dx\\
&+\theta\,2\mu\sum_{E\in\Gamma_{iD}}\int_E[\![\mathbf{u}]\!]:\{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\}\,ds+\theta\,\lambda\sum_{E\in\Gamma_{iD}}\int_E\Pi_0[\![\mathbf{u}]\!]:\Pi_0\{\!\{\nabla\cdot\mathbf{v}\mathbf{1}\}\!\}\,ds\\
&-2\mu\sum_{E\in\Gamma_{iD}}\int_E\{\!\{\boldsymbol{\varepsilon}(\mathbf{u})\}\!\}:[\![\mathbf{v}]\!]\,ds-\lambda\sum_{E\in\Gamma_{iD}}\int_E\Pi_0\{\!\{\nabla\cdot\mathbf{u}\mathbf{1}\}\!\}:\Pi_0[\![\mathbf{v}]\!]\,ds\\
&+k_\mu\,\mu\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\int_E[\![\mathbf{u}]\!]:[\![\mathbf{v}]\!]\,ds+k_\lambda\,\lambda\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\int_E\Pi_0[\![\mathbf{u}]\!]\Pi_0[\![\mathbf{v}]\!]\,ds,
\end{aligned}\tag{5.2.24}$$

for $\theta = \pm 1$ or 0 (that is, for the NIPG, SIPG and IIPG methods), and $\mathbb{V} = \mathbb{Q}_1(\Omega_e)$,

$$a_h^{\mathrm{UI}}(\mathbf{v},\mathbf{v}) \geq C \|\mathbf{v}\|_{\mathrm{DG}}^2 \qquad \forall\ \mathbf{v} \in V_h, \tag{5.2.25}$$

provided that, when $\theta = 1$, $k_\mu > 0$ and $k_\lambda \geq 0$, and when $\theta = 0$ or -1, $k_\mu \geq C_\mu$ and $k_\lambda \geq C_\lambda$, where C_μ and C_λ are positive constants to be calculated.

If all the conditions remain the same, but the bilinear form is instead defined by

$$\begin{aligned} a_h^{\mathrm{UI}}(\mathbf{u},\mathbf{v}) = & \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dx \\ & + \theta\ 2\mu \sum_{E \in \Gamma_{iD}} \int_E \llbracket \mathbf{u} \rrbracket : \{\!\{\boldsymbol{\varepsilon}(\mathbf{v})\}\!\}\, ds + \theta\ \lambda \sum_{E \in \Gamma_{iD}} \int_E \Pi_0 \llbracket \mathbf{u} \rrbracket : \Pi_0 \{\!\{\nabla \cdot \mathbf{v1}\}\!\}\, ds \\ & - 2\mu \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\varepsilon}(\mathbf{u})\}\!\} : \llbracket \mathbf{v} \rrbracket\, ds - \lambda \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\nabla \cdot \mathbf{u1}\}\!\} : \llbracket \mathbf{v} \rrbracket\, ds \\ & + k_\mu\, \mu \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \int_E \llbracket \mathbf{u} \rrbracket : \llbracket \mathbf{v} \rrbracket\, ds + k_\lambda\, \lambda \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \int_E \Pi_0 [\![\mathbf{u}]\!] \Pi_0 [\![\mathbf{v}]\!]\, ds, \end{aligned} \tag{5.2.26}$$

then coercivity is not guaranteed.

5.2.4 Consistency

The level of consistency of the modified formulation needs to be ascertained.

For the case in which all three terms suggested so far (in §5.2.2 and §5.2.3) are under-integrated, the bilinear form is defined by (5.2.24). After applying the continuity of the

exact solution $\mathbf{u}$ and the Dirichlet boundary conditions which it satisfies,

$$\begin{aligned}
a_h^{\mathrm{UI}}(\mathbf{u},\mathbf{v}) - l_h(\mathbf{v}) = & \sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\boldsymbol{\sigma}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{v})\,dx \\
& + \theta\ 2\mu\sum_{E\in\Gamma_D}\int_E(\mathbf{g}\otimes\mathbf{n}):\boldsymbol{\varepsilon}(\mathbf{v})\,ds + \theta\ \lambda\sum_{E\in\Gamma_D}\int_E\Pi_0(\mathbf{g}\otimes\mathbf{n}):\Pi_0(\nabla\cdot\mathbf{v}\mathbf{1})\,ds \\
& - 2\mu\sum_{E\in\Gamma_{iD}}\int_E\{\!\{\boldsymbol{\varepsilon}(\mathbf{u})\}\!\}:[\![\mathbf{v}]\!]\,ds - \lambda\sum_{E\in\Gamma_{iD}}\int_E\Pi_0\{\!\{\nabla\cdot\mathbf{u}\mathbf{1}\}\!\}:\Pi_0[\![\mathbf{v}]\!]\,ds \\
& + k_\mu\,\mu\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E(\mathbf{g}\otimes\mathbf{n}):(\mathbf{v}\otimes\mathbf{n})\,ds + k_\lambda\,\lambda\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E\Pi_0(\mathbf{g}\cdot\mathbf{n})\,\Pi_0(\mathbf{v}\cdot\mathbf{n})\,ds \\
& - \sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\mathbf{f}\cdot\mathbf{v}\,dx - \sum_{E\in\Gamma_N}\int_E\mathbf{h}\cdot\mathbf{v}\,ds \\
& - \theta\ 2\mu\sum_{E\in\Gamma_D}\int_E(\mathbf{g}\otimes\mathbf{n}):\boldsymbol{\varepsilon}(\mathbf{v})\,ds - \theta\ \lambda\sum_{E\in\Gamma_D}\int_E(\mathbf{g}\otimes\mathbf{n}):(\nabla\cdot\mathbf{v}\,\mathbf{1})\,ds \\
& - k_\mu\,\mu\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E(\mathbf{g}\otimes\mathbf{n}):(\mathbf{v}\otimes\mathbf{n})\,ds - k_\lambda\,\lambda\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E(\mathbf{g}\cdot\mathbf{n})(\mathbf{v}\cdot\mathbf{n})\,ds. \qquad (5.2.27)
\end{aligned}$$

For consistency, this expression must vanish. Instead of using $l_h(\mathbf{v})$, use a modified linear form $l_h^{\mathrm{UI}}(\mathbf{v})$ with under-integration corresponding to that of $a_h^{\mathrm{UI}}(\mathbf{u},\mathbf{v})$, that is,

$$\begin{aligned}
l_h^{\mathrm{UI}}(\mathbf{v}) = & \sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\mathbf{f}\cdot\mathbf{v}\,dx + \sum_{E\in\Gamma_N}\int_E\mathbf{h}\cdot\mathbf{v}\,ds + \theta\ 2\mu\sum_{E\in\Gamma_D}\int_E(\mathbf{g}\otimes\mathbf{n}):\boldsymbol{\varepsilon}(\mathbf{v})\,ds \\
& + \theta\ \lambda\sum_{E\in\Gamma_D}\int_E\Pi_0(\mathbf{g}\otimes\mathbf{n}):\Pi_0(\nabla\cdot\mathbf{v}\,\mathbf{1})\,ds + k_\mu\,\mu\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E(\mathbf{g}\otimes\mathbf{n}):(\mathbf{v}\otimes\mathbf{n})\,ds \\
& + k_\lambda\,\lambda\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E\Pi_0(\mathbf{g}\cdot\mathbf{n})\,\Pi_0(\mathbf{v}\cdot\mathbf{n})\,ds. \qquad (5.2.28)
\end{aligned}$$

This gives

$$\begin{aligned}
a_h^{\mathrm{UI}}(\mathbf{u},\mathbf{v}) - l_h^{\mathrm{UI}}(\mathbf{v}) = & \sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\boldsymbol{\sigma}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{v})\,dx \\
& - 2\mu\sum_{E\in\Gamma_{iD}}\int_E\{\!\{\boldsymbol{\varepsilon}(\mathbf{u})\}\!\}:[\![\mathbf{v}]\!] - \lambda\sum_{E\in\Gamma_{iD}}\int_E\Pi_0\{\!\{\nabla\cdot\mathbf{u}\mathbf{1}\}\!\}:\Pi_0[\![\mathbf{v}]\!]\,ds \\
& - \sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\mathbf{f}\cdot\mathbf{v}\,dx - \sum_{E\in\Gamma_N}\int_E\mathbf{h}\cdot\mathbf{v}\,ds. \qquad (5.2.29)
\end{aligned}$$

As in the original case, in §3.1.2,

$$\sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \mathbf{f} \cdot \mathbf{v} \, dx = \sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v}) \, dx - 2\mu \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\varepsilon}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!] \, ds$$
$$-\lambda \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\nabla \cdot \mathbf{u}\, \mathbf{1}\}\!\} : [\![\mathbf{v}]\!] \, ds - \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v} \, ds.$$

Substitution of this into (5.2.29) gives

$$a_h^{\mathrm{UI}}(\mathbf{u}, \mathbf{v}) - l_h^{\mathrm{UI}}(\mathbf{v}) = -\lambda \sum_{E \in \Gamma_{iD}} \int_E \left(\Pi_0 \{\!\{\nabla \cdot \mathbf{u}\, \mathbf{1}\}\!\} : \Pi_0 [\![\mathbf{v}]\!] - \{\!\{\nabla \cdot \mathbf{u}\, \mathbf{1}\}\!\} : [\![\mathbf{v}]\!]\right) ds$$
$$=: E_h^{\mathrm{UI}}(\mathbf{u}, \mathbf{v}), \tag{5.2.30}$$

a consistency error.

Throughout this simplification procedure, the effect of under-integration of any one term in the bilinear form is independent of the under-integration of any other term (provided that a corresponding term in the linear form, should it exist, is also under-integrated).

The consistency error that is produced by the modified formulation is specifically related to the under-integration of the term corresponding to VI; under-integrating only the terms corresponding to IV and $VIII$ leads to a consistent formulation. However, the under-integration of VI is necessary for coercivity, as was shown in §5.2.3 (Lemma 1). This applies to all three IP methods, insofar as the terms mentioned appear.

5.2.5 An error bound

The attempted direct elimination of unwanted λ-dependence, the clarification of conditions for coercivity, and the matter of a potentially inconsistent formulation are all included in obtaining an overall error bound for the modified formulation.

The procedure of bounding the error is similar to that of the preliminary analysis of Chapter 3.

Initial steps

With $\mathbf{u}_P$ as defined in Chapter 4, the approximation error is again written as $\boldsymbol{e} = \boldsymbol{\gamma} + \boldsymbol{w}$, so that

$$C \|\boldsymbol{e}\|_{\mathrm{DG}}^2 \leq \|\boldsymbol{\gamma}\|_{\mathrm{DG}}^2 + \|\boldsymbol{w}\|_{\mathrm{DG}}^2 .$$

The first term is bounded exactly as before, giving

$$\|\boldsymbol{\gamma}\|_{\mathrm{DG}}^2 \leq C \sum_{\Omega_e \in \mathcal{T}_h} h_e^2 \|\mathbf{u}\|_{H^2(\Omega_e)}^2 .$$

Requiring coercivity

For $\mathbb{Q}_1$ elements, for all three IP methods, while only two terms (if they appeared in the formulation) were initially identified as contributing to λ-dependence, coercivity of the modified formulation required that all three λ-dependent edge terms were under-integrated (Lemma 1). That case is therefore the one that is considered.

By coercivity,

$$\begin{aligned} \|\boldsymbol{w}\|_{\mathrm{DG}}^2 &\leq a_h^{\mathrm{UI}}(\boldsymbol{w}, \boldsymbol{w}) \\ &= \left| a_h^{\mathrm{UI}}(\boldsymbol{e} - \boldsymbol{\gamma}, \boldsymbol{w}) \right| . \end{aligned} \tag{5.2.31}$$

The effect of inconsistency

Because the term corresponding to VI is under-integrated, the formulation is inconsistent, giving

$$\begin{aligned} a_h^{\mathrm{UI}}(\mathbf{u}, \boldsymbol{w}) &= l_h^{\mathrm{UI}}(\boldsymbol{w}) + E_h^{\mathrm{UI}}(\mathbf{u}, \boldsymbol{w}) \\ &= a_h^{\mathrm{UI}}(\mathbf{u}_h, \boldsymbol{w}) + E_h^{\mathrm{UI}}(\mathbf{u}, \boldsymbol{w}) , \end{aligned} \tag{5.2.32}$$

so that

$$a_h^{\mathrm{UI}}(\boldsymbol{e}, \boldsymbol{w}) = E_h^{\mathrm{UI}}(\mathbf{u}, \boldsymbol{w}) \tag{5.2.33}$$

and

$$\|\boldsymbol{w}\|_{\mathrm{DG}}^2 \leq \left| E_h^{\mathrm{UI}}(\mathbf{u}, \boldsymbol{w}) \right| + \left| -a_h^{\mathrm{UI}}(\boldsymbol{\gamma}, \boldsymbol{w}) \right| . \tag{5.2.34}$$

For the second term on the right-hand side, the strategy, as before, is to extract a factor of $\|\boldsymbol{w}\|_{\mathrm{DG}}$ from each term in $a_h^{\mathrm{UI}}(\boldsymbol{\gamma}, \boldsymbol{w})$, and then to bound the remaining expressions, containing $\boldsymbol{\gamma}$, by norms of the exact solution.

The same procedure of factorising and bounding will be applied to the term contributed by the consistency error as well, in order to obtain a bound for the approximation error.

The terms of the bilinear form

The terms not affected by under-integration are bounded as for the unmodified formulation.

As shown in §5.2.2, under-integrating the two terms seen initially to be problematic eliminates their contribution to the error, and thereby specifically to λ-dependence.

Under-integrating the third edge term (to maintain coercivity) means that VI is replaced with

$$\begin{aligned} VI^{\mathrm{UI}} &= \lambda \left| \sum_{E \in \Gamma_{iD}} \int_E \Pi_0 \{\!\{ \nabla \cdot \boldsymbol{\gamma} \mathbf{1} \}\!\} : \Pi_0 [\![\boldsymbol{w}]\!] \, ds \right| \\ &\leq \lambda \sum_{E \in \Gamma_{iD}} \|\Pi_0 [\![\boldsymbol{w}]\!]\|_{L^2(E)} \, \|\Pi_0 \{\!\{ \nabla \cdot \boldsymbol{\gamma} \mathbf{1} \}\!\}\|_{L^2(E)} \\ &\leq \lambda \sum_{E \in \Gamma_{iD}} \|[\![\boldsymbol{w}]\!]\|_{L^2(E)} \, \|\{\!\{ \nabla \cdot \boldsymbol{\gamma} \mathbf{1} \}\!\}\|_{L^2(E)} \,, \end{aligned} \tag{5.2.35}$$

which is what was obtained in the analysis of the unmodified formulation (see the working leading up to (3.2.41)). This therefore leads to the same bounds as before, where the λ-dependence is not problematic.

Thus the bilinear form is bounded:

$$|a_h(\boldsymbol{\gamma}, \boldsymbol{w})| \leq C \|\boldsymbol{w}\|_{\mathrm{DG}} \left[\sum_{\Omega_e \in \mathcal{T}_h} h_e^2 \left(\|\mathbf{u}\|_{H^2(\Omega_e)}^2 + \lambda^2 \|\nabla \cdot \mathbf{u}\|_{H^1(\Omega_e)}^2 \right) \right]^{\frac{1}{2}} . \tag{5.2.36}$$

The consistency error

The consistency error should ultimately be bounded by an expression of the same form.

$$
\begin{aligned}
\left|E_h^{\mathrm{UI}}(\mathbf{u},\boldsymbol{w})\right| &= \lambda \left| \sum_{E\in\Gamma_{iD}} \int_E \left(\Pi_0 \{\!\{\nabla\cdot\mathbf{u1}\}\!\} : \Pi_0 \llbracket \boldsymbol{w} \rrbracket - \{\!\{\nabla\cdot\mathbf{u1}\}\!\} : \llbracket \boldsymbol{w} \rrbracket\right) ds \right| \\
&= \lambda \left| \sum_{E\in\Gamma_{iD}} \int_E \left(\Pi_0 \{\!\{\nabla\cdot\mathbf{u1}\}\!\} : \llbracket \boldsymbol{w} \rrbracket - \{\!\{\nabla\cdot\mathbf{u1}\}\!\} : \llbracket \boldsymbol{w} \rrbracket\right) ds \right| \\
&= \lambda \left| \sum_{E\in\Gamma_{iD}} \int_E \left(\Pi_0 \{\!\{\nabla\cdot\mathbf{u1}\}\!\} - \{\!\{\nabla\cdot\mathbf{u1}\}\!\}\right) : \llbracket \boldsymbol{w} \rrbracket \, ds \right| \\
&\leq \lambda \sum_{E\in\Gamma_{iD}} h_E^{1/2} \left\| \Pi_0 \{\!\{\nabla\cdot\mathbf{u1}\}\!\} - \{\!\{\nabla\cdot\mathbf{u1}\}\!\} \right\|_{L^2(E)} h_E^{-1/2} \left\| \llbracket \boldsymbol{w} \rrbracket \right\|_{L^2(E)} \\
&\leq C\,\lambda \left(\sum_{\Omega_e\in\mathcal{T}_h} \sum_{E\in\partial\Omega_e^{iD}} h_E \left\| \Pi_0\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u} \right\|^2_{L^2(E)} \right)^{1/2} \\
&\qquad\qquad \cdot \left(\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\| \llbracket \boldsymbol{w} \rrbracket \right\|^2_{L^2(E)} \right)^{1/2} \\
&\leq C\,\lambda \left\| \boldsymbol{w} \right\|_{\mathrm{DG}} \left(\sum_{\Omega_e\in\mathcal{T}_h} \sum_{E\in\partial\Omega_e^{iD}} h_E \left\| \Pi_0\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u} \right\|^2_{L^2(E)} \right)^{1/2} . \qquad (5.2.37)
\end{aligned}
$$

The splitting function $\mathbf{u}_P$, element-wise linear, has constant derivatives, and specifically divergence, on each element, so that on each edge

$$\Pi_0\nabla\cdot\mathbf{u}_P - \nabla\cdot\mathbf{u}_P = 0. \qquad (5.2.38)$$

Therefore

$$
\begin{aligned}
\left\| \Pi_0\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u} \right\|^2_{L^2(E)} &= \left\| \left(\Pi_0\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}\right) - \left(\Pi_0\nabla\cdot\mathbf{u}_P - \nabla\cdot\mathbf{u}_P\right) \right\|^2_{L^2(E)} \\
&= \left\| \Pi_0 \left(\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}_P\right) - \left(\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}_P\right) \right\|^2_{L^2(E)} \\
&\leq C \left\| \nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}_P \right\|^2_{L^2(E)} \\
&\leq C \left(h_e^{-1} \left\| \nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}_P \right\|^2_{L^2(\Omega_e)} + h_e \left| \nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}_P \right|^2_{H^1(\Omega_e)} \right) \\
&\leq C h_e \left| \nabla\cdot\mathbf{u} \right|^2_{H^1(\Omega_e)} , \qquad (5.2.39)
\end{aligned}
$$

using (4.5.9), so that

$$\sum_{\Omega_e\in\mathcal{T}_h} \sum_{E\in\partial\Omega_e^{iD}} h_E \left\| \Pi_0\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u} \right\|^2_{L^2(E)} \leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e^2 \left| \nabla\cdot\mathbf{u} \right|^2_{H^1(\Omega_e)} \qquad (5.2.40)$$

and the bound follows:

$$\left|E^{\mathrm{UI}}(\mathbf{u},\boldsymbol{w})\right| \leq C\,\lambda\,\|\boldsymbol{w}\|_{\mathrm{DG}} \left(\sum_{\Omega_e\in\mathcal{T}_h} h_e^2\,|\nabla\cdot\mathbf{u}|^2_{H^1(\Omega_e)}\right)^{\frac{1}{2}}. \tag{5.2.41}$$

The final error bound

Combining the bounds in the previous two sections gives

$$\begin{aligned} \|\boldsymbol{w}\|^2_{\mathrm{DG}} &\leq C\,\|\boldsymbol{w}\|_{\mathrm{DG}} \left[\sum_{\Omega_e\in\mathcal{T}_h} h_e^2 \left(\|\mathbf{u}\|^2_{H^2(\Omega_e)} + \lambda^2\,\|\nabla\cdot\mathbf{u}\|^2_{H^1(\Omega_e)}\right)\right]^{\frac{1}{2}} \\ \implies \|\boldsymbol{w}\|^2_{\mathrm{DG}} &\leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e^2 \left(\|\mathbf{u}\|^2_{H^2(\Omega_e)} + \lambda^2\,\|\nabla\cdot\mathbf{u}\|^2_{H^1(\Omega_e)}\right). \end{aligned} \tag{5.2.42}$$

Thus the approximation error is bounded

$$\begin{aligned} \|\boldsymbol{e}\|_{\mathrm{DG}} &\leq Ch\left(\|\mathbf{u}\|_{H^2(\Omega)} + \lambda\,\|\nabla\cdot\mathbf{u}\|_{H^1(\Omega)}\right) \\ &\leq ChF\left(\mathbf{f},\mathbf{g}_\Omega,\mathbf{h}_\Omega\right) \\ &\leq Ch, \end{aligned} \tag{5.2.43}$$

showing optimal convergence with respect to mesh size, uniformly with respect to λ.

5.2.6 Summary

In summary, under-integration of all of the λ-dependent edge integrals in the IP formulation leads to an optimal error bound, and consequently a locking-free method, when bilinear elements are used. However, application of under-integration on fewer integrals leads to the same poor performance as before, and potentially to instability.

The key result is restated in the following theorem:

Theorem 1 *Let the bilinear form be defined as in (5.2.24) and the linear form as in (5.2.28), for $\theta = \pm 1$ or $\theta = 0$, with k_μ and k_λ satisfying the requirements of Lemma 1, and let $\mathbb{V} = \mathbb{Q}_1(\Omega_e)$. Then the approximation error $\boldsymbol{e}$ of the DG method defined by*

$$a_h^{\mathrm{UI}}(\mathbf{u}_h,\mathbf{v}) = l_h^{\mathrm{UI}}(\mathbf{v}) \qquad \forall\ \mathbf{v}\in V_h \tag{5.2.44}$$

for solving (2.1.1) is bounded as

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq Ch, \tag{5.2.45}$$

where C is a constant independent of h and λ, provided that the domain is such that it satisfies the regularity condition (A.6.1), and that the mesh is geometrically conforming with regular rectangular elements.

Thus under-integration of these specific edge terms is a successful remedy for the poor approximations caused by the use of bilinear elements with the unmodified IP formulation.

5.3 Linear approximations

The second, alternative modification suggested by the preliminary analysis of Chapter 3 is to use linear ($\mathbb{P}_1$) approximations on quadrilateral elements.

5.3.1 Overall comparison to bilinear elements

Using linear approximations, much of the convergence analysis of Chapter 3 stays the same.

The coercivity of the form holds on the DG norm, as before, and the formulation is similarly consistent.

The error is split in exactly the same way, where once again $\boldsymbol{w} \in V_h$, this time with $\mathbb{V} = \mathbb{P}_1(\Omega_e)$.

Since $\mathbf{u}_P$ as defined in Chapter 4 is unaffected by the change in the approximation space, the bounds on $\boldsymbol{\gamma}$ hold as before; and the manipulation of terms performed in §3.2.5 can be performed in the same way. Thus the entire convergence analysis holds as before.

However, having a linear approximation means that, element-wise, $\boldsymbol{w} \in [\mathbb{P}_1]^2$, rather than $\boldsymbol{w} \in [\mathbb{Q}_1]^2$, and consequently $\nabla \cdot \boldsymbol{w} \in \mathbb{P}_0$ rather than $\nabla \cdot \boldsymbol{w} \in \mathbb{P}_1$. Following the reasoning of §3.2.7, this is expected potentially to allow for a better bound.

5.3.2 The effect on specific terms

The two terms that were shown to be problematic in the analysis for bilinear elements, that is, that contributed unwanted λ-dependence, were

$$IV = \lambda |\theta| \left| \sum_{E \in \Gamma_{iD}} \int_E [\![\boldsymbol{\gamma}]\!] \{\!\{ \nabla \cdot \boldsymbol{w} \}\!\} \, ds \right| \tag{5.3.1}$$

and

$$VIII = k_\lambda \lambda \left| \sum_{E \in \Gamma_{iD}} \int_E [\![\boldsymbol{\gamma}]\!] [\![\boldsymbol{w}]\!] \, ds \right|. \tag{5.3.2}$$

In IV, the use of linear approximations means that

$$\sum_{E \in \Gamma_{iD}} \int_E [\![\boldsymbol{\gamma}]\!] \{\!\{ \nabla \cdot \boldsymbol{w} \}\!\} \, ds = \sum_{E \in \Gamma_{iD}} \{\!\{ \nabla \cdot \boldsymbol{w} \}\!\}|_E \int_E [\![\boldsymbol{\gamma}]\!] \, ds. \tag{5.3.3}$$

As before, by (4.1.1), this vanishes, so that the contribution of this term to undesirable dependence on λ is removed.

However, considering $VIII$ in the same light shows that no improvement is made by using linear elements, as $[\![\boldsymbol{w}]\!] \in \mathbb{P}_1(E)$ on each edge whether bilinear or linear approximations are used, and the expression can thus not be manipulated in the same way.

5.3.3 Error bound depending on IP method

In the case of the SIPG and IIPG methods, where $k_\lambda > 0$ is necessary for the coercivity of the bilinear form, term $VIII$ is an unavoidable part of the formulation. Consequently, even if linear elements are used,

$$\begin{aligned} \|\boldsymbol{e}\|_{\mathrm{DG}} &\leq Ch \left(\|\mathbf{u}\|_{H^2(\Omega)} + \lambda \|\mathbf{u}\|_{H^2(\Omega)} + \lambda \|\nabla \cdot \mathbf{u}\|_{H^1(\Omega)} \right) \\ &\leq C\lambda h F(\mathbf{f}, \mathbf{g}_\Omega, \mathbf{h}_\Omega) \\ &\leq C\lambda h \end{aligned} \tag{5.3.4}$$

as before, and once again the approximation is prone to being very poor for nearly incompressible materials.

In fact, for the IIPG method, the use of linear elements has no effect on any part of the error bound, as $\theta = 0$ means that IV, the one term that is affected, does not appear in the formulation at all.

In the case of NIPG, if $k_\lambda \neq 0$ is chosen, a poor approximation will be expected as before, for the same reason.

However, since the second stabilization term is unnecessary in the NIPG method, k_λ can be chosen to be zero. In that case, using linear elements eliminates the only remaining problematic term, so that the error bound becomes

$$\begin{aligned} \|\boldsymbol{e}\|_{\mathrm{DG}} &\leq Ch \left(\|\mathbf{u}\|_{H^2(\Omega)} + \lambda \|\nabla \cdot \mathbf{u}\|_{H^1(\Omega)} \right) \\ &\leq ChF\left(\mathbf{f}, \mathbf{g}_\Omega, \mathbf{h}_\Omega\right) \\ &\leq Ch, \end{aligned} \tag{5.3.5}$$

which shows optimal convergence independent of λ, as desired.

In summary, using linear elements will alleviate the locking problem and other manifestations of poor approximations if the NIPG method is used with $k_\lambda = 0$ specifically, but has no remedial effect on the other methods.

Theorem 2 *Let the bilinear form be defined by*

$$\begin{aligned} a_h(\mathbf{u}, \mathbf{v}) = &\sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v})\, dx + \sum_{E \in \Gamma_{iD}} \int_E [\![\mathbf{u}]\!] : \{\!\{\boldsymbol{\sigma}(\mathbf{v})\}\!\}\, ds \\ &- \sum_{E \in \Gamma_{iD}} \int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\} : [\![\mathbf{v}]\!]\, ds + k_\mu \mu \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \int_E [\![\mathbf{u}]\!] : [\![\mathbf{v}]\!]\, ds \end{aligned} \tag{5.3.6}$$

and the linear form by

$$\begin{aligned} l_h(\mathbf{v}) = &\sum_{\Omega_e \in \mathcal{T}_h} \int_{\Omega_e} \mathbf{f} \cdot \mathbf{v}\, dx + \sum_{E \in \Gamma_D} \int_E (\mathbf{g} \otimes \mathbf{n}) : \boldsymbol{\sigma}(\mathbf{v})\, ds + \sum_{E \in \Gamma_N} \int_E \mathbf{h} \cdot \mathbf{v}\, ds \\ &+ k_\mu \mu \sum_{E \in \Gamma_D} \frac{1}{h_E} \int_E \mathbf{g} \cdot \mathbf{v}\, ds \end{aligned} \tag{5.3.7}$$

(that is, (2.3.1) and (2.3.2) with $\theta = 1$ and $k_\lambda = 0$), with $k_\mu > 0$, and let $\mathbb{V} = \mathbb{P}_1(\Omega_e)$. Then the approximation error $\boldsymbol{e}$ of the DG method defined by

$$a_h(\mathbf{u}_h, \mathbf{v}) = l_h(\mathbf{v}) \qquad \forall\ \mathbf{v} \in V_h \tag{5.3.8}$$

for solving (2.1.1) is bounded as

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq Ch, \tag{5.3.9}$$

where C is a constant independent of h and λ, provided that the domain is such that it satisfies the regularity condition (A.6.1), and that the mesh is geometrically conforming with regular rectangular elements.

5.4 Linear approximations with under-integration

As the use of linear approximations produces uniform convergence for the NIPG method but not for the other two IP methods, under-integration, particularly of the second stabilization term, is introduced. As in the case of bilinear elements, under-integration of this stabilization term eliminates the contribution of this term to the λ-dependence of the error bound. However, as before, this involves a modification of the formulation itself, and therefore the error analysis as a whole needs to be examined for possible effects. Specifically, coercivity and consistency will be considered, as was done in the case of the bilinear elements.

5.4.1 Choice of terms for under-integration: a first guess

Only one term contributes to the λ-dependence of the error bound, with linear approximations, so this term, $VIII$, is given primary consideration.

5.4.2 Coercivity

Coercivity of the modified bilinear form is again dealt with for each IP method individually.

As before, the μ-dependent terms are unaffected, and what is required for coercivity in each case is $T_\lambda \geq 0$.

NIPG

While the term corresponding to $VIII$ is in general superfluous to the NIPG formulation, under-integration of this term is considered for the sake of gaining a complete picture.

Under-integration gives

$$T_\lambda = \lambda \sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} + k_\lambda \lambda \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \|\Pi_0 [\![\mathbf{v}]\!]\|^2_{L^2(E)} \geq 0. \tag{5.4.1}$$

Coercivity holds, then, for $k_\lambda \geq 0$.

IIPG

For the IIPG formulation, under-integration of only the stabilization term gives

$$T_\lambda = \sum_{\Omega_e \in \mathcal{T}_h} \left[\lambda \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} - \frac{\lambda}{2} \sum_{E \in \partial\Omega_e^{int}} \int_E [\![\mathbf{v}]\!] \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds - \lambda \sum_{E \in \partial\Omega_e^D} \int_E (\mathbf{v} \cdot \mathbf{n}) (\nabla \cdot \mathbf{v}) \, ds \right.$$
$$\left. + \frac{k_\lambda}{2} \lambda \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \|\Pi_0 [\![\mathbf{v}]\!]\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E \in \partial\Omega_e^D} \frac{1}{h_E} \|\Pi_0 \mathbf{v} \cdot \mathbf{n}\|^2_{L^2(E)} \right]. \tag{5.4.2}$$

In the case of bilinear elements, this would lead to

$$T_\lambda \geq \lambda \sum_{\Omega_e \in \mathcal{T}_h} \left[(1 - C\epsilon_\lambda) \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} + \frac{k_\lambda}{2} \left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} - \frac{C}{\epsilon_\lambda} \left\| h_E^{-\frac{1}{2}} [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \right], \tag{5.4.3}$$

and coercivity would not be shown.

In contrast, however, if linear elements are used, the identity

$$\int_E [\![\mathbf{v}]\!] \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds = \int_E \Pi_0 [\![\mathbf{v}]\!] \{\!\{\nabla \cdot \mathbf{v}\}\!\} \, ds \tag{5.4.4}$$

gives

$$T_\lambda = \sum_{\Omega_e \in \mathcal{T}_h} \Bigg[\lambda \left\| \nabla \cdot \mathbf{v} \right\|^2_{L^2(\Omega_e)} - \frac{\lambda}{2} \sum_{E \in \partial\Omega_e^{int}} \int_E \Pi_0 [\![\mathbf{v}]\!] \{\!\{ \nabla \cdot \mathbf{v} \}\!\} \, ds - \lambda \sum_{E \in \partial\Omega_e^{D}} \int_E \Pi_0 \left(\mathbf{v} \cdot \mathbf{n} \right) \left(\nabla \cdot \mathbf{v} \right) \, ds$$
$$+ \frac{k_\lambda}{2} \lambda \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \left\| \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E \in \partial\Omega_e^{D}} \frac{1}{h_E} \left\| \Pi_0 \mathbf{v} \cdot \mathbf{n} \right\|^2_{L^2(E)} \Bigg]$$
$$\geq \sum_{\Omega_e \in \mathcal{T}_h} \lambda \left[(1 - C\epsilon_\lambda) \left\| \nabla \cdot \mathbf{v} \right\|^2_{L^2(\Omega_e)} + \left(\frac{k_\lambda}{2} - \frac{C}{\epsilon_\lambda} \right) \left\| h_E^{-\frac{1}{2}} \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(\partial\Omega_e^{iD})} \right]. \tag{5.4.5}$$

Here, appropriate choices of ϵ_λ and k_λ will give $T_\lambda \geq 0$, and thus the modified form is coercive if the approximations are linear.

SIPG

Under-integration of only the stabilization term is similar to the corresponding case for IIPG, with the difference being nothing more than the value of the constants in particular coefficients:

$$T_\lambda = \sum_{\Omega_e \in \mathcal{T}_h} \Bigg[\lambda \left\| \nabla \cdot \mathbf{v} \right\|^2_{L^2(\Omega_e)} - \lambda \sum_{E \in \partial\Omega_e^{int}} \int_E [\![\mathbf{v}]\!] \{\!\{ \nabla \cdot \mathbf{v} \}\!\} \, ds - 2\lambda \sum_{E \in \partial\Omega_e^{D}} \int_E \left(\mathbf{v} \cdot \mathbf{n} \right) \left(\nabla \cdot \mathbf{v} \right) \, ds$$
$$+ \frac{k_\lambda}{2} \lambda \sum_{E \in \partial\Omega_e^{int}} \frac{1}{h_E} \left\| \Pi_0 [\![\mathbf{v}]\!] \right\|^2_{L^2(E)} + k_\lambda \lambda \sum_{E \in \partial\Omega_e^{D}} \frac{1}{h_E} \left\| \Pi_0 \mathbf{v} \cdot \mathbf{n} \right\|^2_{L^2(E)} \Bigg]. \tag{5.4.6}$$

Using (5.4.4) leads to being able to obtain $T_\lambda \geq 0$ in exactly the same way as for IIPG, with appropriate choices of the parameters, so that for linear approximations the form is coercive.

Summary of coercivity results

Lemma 2 *With the bilinear form defined by*

$$a_h^{\mathrm{UI}}(\mathbf{u},\mathbf{v}) = \sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\boldsymbol{\sigma}(\mathbf{u}):\boldsymbol{\varepsilon}(\mathbf{v})\,dx + \theta\sum_{E\in\Gamma_{iD}}\int_E \llbracket\mathbf{u}\rrbracket:\{\!\{\boldsymbol{\sigma}(\mathbf{v})\}\!\}\,ds - \sum_{E\in\Gamma_{iD}}\int_E \{\!\{\boldsymbol{\sigma}(\mathbf{u})\}\!\}:\llbracket\mathbf{v}\rrbracket\,ds + k_\mu\,\mu\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\int_E \llbracket\mathbf{u}\rrbracket:\llbracket\mathbf{v}\rrbracket\,ds + k_\lambda\,\lambda\sum_{E\in\Gamma_{iD}}\frac{1}{h_E}\int_E \Pi_0\llbracket\mathbf{u}\rrbracket\Pi_0\llbracket\mathbf{v}\rrbracket\,ds, \tag{5.4.7}$$

for $\theta=\pm1$ or 0 (that is, for the NIPG, SIPG and IIPG methods), and $\mathbb{V}=\mathbb{P}_1(\Omega_e)$,

$$a_h^{\mathrm{UI}}(\mathbf{v},\mathbf{v}) \geq C\,\|\mathbf{v}\|_{\mathrm{DG}}^2 \qquad \forall\ \mathbf{v}\in V_h, \tag{5.4.8}$$

provided that, when $\theta=1$, $k_\mu>0$ and $k_\lambda\geq 0$, and when $\theta=0$ or -1, $k_\mu\geq C_\mu$ and $k_\lambda\geq C_\lambda$, where C_μ and C_λ are positive constants to be calculated.

5.4.3 Consistency

Evaluation of the consistency of the modified formulation follows exactly as in §5.2.4: under-integration of the term corresponding to $VIII$ does not produce a consistency error with the linear form under-integrated in the term associated with λ-stabilization, that is,

$$l_h^{\mathrm{UI}}(\mathbf{v}) = \sum_{\Omega_e\in\mathcal{T}_h}\int_{\Omega_e}\mathbf{f}\cdot\mathbf{v}\,dx + \theta\sum_{E\in\Gamma_D}\int_E(\mathbf{g}\otimes\mathbf{n}):\boldsymbol{\sigma}(\mathbf{v})\,ds + \sum_{E\in\Gamma_N}\int_E\mathbf{h}\cdot\mathbf{v}\,ds + k_\mu\mu\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E\mathbf{g}\cdot\mathbf{v}\,ds + k_\lambda\,\lambda\sum_{E\in\Gamma_D}\frac{1}{h_E}\int_E\Pi_0(\mathbf{g}\cdot\mathbf{n})\,\Pi_0(\mathbf{v}\cdot\mathbf{n})\,ds. \tag{5.4.9}$$

5.4.4 Bounding the error

Since, with linear elements, neither coercivity nor consistency is compromised by under-integrating the second stabilization term, the error is bounded exactly as in §5.3 until the contribution of $VIII$ is considered. With under-integration, the term $VIII$ is replaced

with $VIII^{\mathrm{UI}}$, which vanishes, and therefore the final error bound is

$$\begin{aligned}\|\boldsymbol{e}\|_{\mathrm{DG}} &\leq Ch\left(\|\mathbf{u}\|_{H^2(\Omega)} + \lambda\,\|\nabla\cdot\mathbf{u}\|_{H^1(\Omega)}\right)\\ &\leq Ch.\end{aligned} \tag{5.4.10}$$

This shows that this application of under-integration gives optimal convergence with respect to mesh size, uniformly with respect to λ, when linear approximations are used:

Theorem 3 *Let the bilinear form be defined as in (5.4.7) and the linear form as in (5.4.9), for $\theta = \pm 1$ or $\theta = 0$, with k_μ and k_λ satisfying the requirements of Lemma 2, and let $\mathbb{V} = \mathbb{P}_1(\Omega_e)$. Then the approximation error $\boldsymbol{e}$ of the DG method defined by*

$$a_h^{\mathrm{UI}}(\mathbf{u}_h, \mathbf{v}) = l_h^{\mathrm{UI}}(\mathbf{v}) \qquad \forall\ \mathbf{v} \in V_h \tag{5.4.11}$$

for solving (2.1.1) is bounded as

$$\|\boldsymbol{e}\|_{\mathrm{DG}} \leq Ch, \tag{5.4.12}$$

where C is a constant independent of h and λ, provided that the domain is such that it satisfies the regularity condition (A.6.1), and that the mesh is geometrically conforming with regular rectangular elements.

5.5 Other cases of under-integration

In the earlier sections, §5.2 and §5.4, the terms ultimately chosen for under-integration were those that were necessary and sufficient to guarantee uniformly convergent formulations. In the process of ascertaining which those were, other combinations of terms were considered in §5.2.

This section outlines the results of under-integrating various combinations of terms, including several not covered in the preceding sections. A range of cases are described briefly, highlighting what is necessary for coercivity of a form, or what undermines coercivity. The results pertaining to which terms need to be under-integrated because of their direct contribution to λ-dependence in the error estimate, and the result that under-integration of term VI is the cause of lack of consistency in the modification of

the formulation, will not be referred to in the descriptions, as they have been covered sufficiently in preceding sections.

The accompanying tables summarise the results, displaying the various factors that contribute to the final error bounds.

5.5.1 Under-integrating the stabilization term

The NIPG formulation stands alone in that for bilinear elements, under-integration of the second stabilization term does not compromise coercivity - the proof is the same as for linear elements.

For IIPG and SIPG, for bilinear elements, under-integration of only that term produces the situation shown in (5.4.2) and (5.4.3): the stabilization term involves a projection of the jump, while the other integral involving a jump does not include a projection. Therefore the two norms remain separate, and the constants cannot in general be chosen to ensure a positive expression overall. For linear elements, as was demonstrated, the use of (5.4.4) eliminates this discrepancy, which is why the two norms can be written as one term and a positive coefficient can be found. This is, in most cases, the situation distinguishing the forms that can be proven coercive from those that cannot.

5.5.2 Under-integrating terms IV and $VIII$

Under-integrating a second term in the case of NIPG for bilinear elements was dealt with in §5.2.3: when $VIII$ was present, the situation described above occurred; when it was not present, there was no term available to compensate for the negative coefficient of the norm of the jump term.

In contrast, for linear elements,

$$\int_E [\![\mathbf{v}]\!] \{\!\{\nabla \cdot \mathbf{v}\}\!\}\, ds = \int_E \Pi_0 [\![\mathbf{v}]\!]\, \Pi_0 \{\!\{\nabla \cdot \mathbf{v}\}\!\}\, ds, \tag{5.5.1}$$

so that $IV^{\mathrm{UI}} = -VI$, and the edge terms cancel each other out, leaving

$$\begin{aligned} T_\lambda &= \lambda \sum_{\Omega_e \in \mathcal{T}_h} \|\nabla \cdot \mathbf{v}\|^2_{L^2(\Omega_e)} + k_\lambda \lambda \sum_{E \in \Gamma_{iD}} \frac{1}{h_E} \|\Pi_0 [\![\mathbf{v}]\!]\|^2_{L^2(E)} \\ &\geq 0. \end{aligned} \tag{5.5.2}$$

That is, under-integration that undermines coercivity for bilinear elements does not, in this case, do so for linear elements. Therefore, while this extra under-integration is unnecessary for eliminating locking, it is not detrimental to the performance of the method.

For the IIPG method, term IV does not appear.

For the SIPG method, under-integration of the two terms leads to the situation described above, where for bilinear elements, the two jump-related norms are necessarily separate, and an overall positive bound cannot be established, while for linear elements, the norms are combined and a positive coefficient can be found by a suitable choice of the parameters, particularly k_λ. Once again, it is unnecessary to under-integrate both terms for linear elements, but doing so does not compromise coercivity, and therefore does not have a negative effect on the performance of the method.

5.5.3 Under-integrating terms IV, VI and $VIII$

Under-integration of all three terms was dealt with in detail for bilinear elements in §5.2.3, with coercivity being obtained in every case.

For linear elements, the proofs are exactly the same, demonstrating again in this final case that excess under-integration, while restricted to the edge integrals containing λ, is not detrimental to the performance of the methods.

5.6 Summary

The tables summarise the results of each of the IP methods for increments of under-integration: none at all ("full" integration), or under-integration of one, two or three

terms insofar as they appear in the formulation in question (later referred to as UI1, UI2 or UI3), for bilinear and linear approximations. Results for the NIPG method are presented for the case $k_\lambda = 0$, omitting the superfluous second stabilization term, while those for the IIPG and SIPG methods assume $k_\lambda > 0$. "NIPG+" refers to the NIPG method with the second stabilization term included, that is, with $k_\lambda > 0$.

A property that is absent is marked "x". In the case of consistency, the symbol "$\otimes$" denotes that the formulation is not consistent, but that this is not problematic in obtaining the desired error bound, as the consistency error can itself be bounded, as was shown in §5.2.5. λ-independence refers to whether the individual terms of a formulation can be bounded in a way that would allow for a λ-independent result, not to the ultimate result itself (as the final result is dependent on coercivity, consistency - or the bounding of the consistency error - and boundability of each term). The final column gives the overall results, indicating whether or not uniform convergence is proven.

$\mathbb{Q}_1$		Coercivity	Consistency	λ-indep.	Optimal conv.
NIPG	Full			x	x
	UI of IV	x			x
	UI of IV,VI		$\otimes$		$\surd$
NIPG+	Full			x	x
	UI of $VIII$	x		x	x
	UI of $VIII$, IV	x			x
	UI of $VIII$, IV, VI		$\otimes$		$\surd$
IIPG	Full			x	x
	UI of $VIII$	x			x
	UI of $VIII$, VI		$\otimes$		$\surd$
SIPG	Full			x	x
	UI of $VIII$	x		x	x
	UI of $VIII$, IV	x			x
	UI of $VIII$, IV, VI		$\otimes$		$\surd$

Table 5.1: Properties of the IP methods leading to uniform convergence with quadrilateral elements and bilinear approximations

$\mathbb{P}_1$		Coercivity	Consistency	λ-indep.	Optimal conv.
NIPG	Full				$\surd$
	UI of IV				$\surd$
	UI of IV, VI		$\otimes$		$\surd$
NIPG+	Full			x	x
	UI of $VIII$				$\surd$
	UI of $VIII$, IV				$\surd$
	UI of $VIII$, IV, VI		$\otimes$		$\surd$
IIPG	Full			x	x
	UI of $VIII$				$\surd$
	UI of $VIII$, VI		$\otimes$		$\surd$
SIPG	Full			x	x
	UI of $VIII$				$\surd$
	UI of $VIII$, IV				$\surd$
	UI of $VIII$, IV, VI		$\otimes$		$\surd$

Table 5.2: Properties of the IP methods leading to uniform convergence with quadrilateral elements and linear approximations

Chapter 6

Extension to three-dimensional domains

In this chapter, the theory presented in Chapters 3, 4 and 5 for a two-dimensional domain is extended to three-dimensional domains, many of the details remaining the same. Particularly, a three-dimensional interpolant is described, along with the properties of its projection onto linear polynomials, to provide a error-splitting function $\mathbf{u}_P$ with analogous properties to the two-dimensional function $\mathbf{u}_P$ of Chapter 4; and the equivalence between under-integration and projections onto constants is extended from the context of one-dimensional edges to that of two-dimensional faces.

In a three-dimensional domain Ω, let E represent a face of the 3-rectangle Ω_e. Let h_e be the element measure and h_E be the measure of face E.

6.1 An error-splitting function in three dimensions

To extend the analysis presented for the two-dimensional setting to provide for three-dimensional domains, a function $\mathbf{u}_P$ satisfying the three-dimensional analogue of the requirements of §4.1 is desired. As before, a new interpolant $\pi\mathbf{u}$ in a local space greater than $[\mathbb{P}_1]^3$ is constructed, and its direct projection onto the space of element-wise linear polynomials is used as $\mathbf{u}_P$, that is, $\mathbf{u}_P = \Pi \circ \pi\mathbf{u}$.

The requirements on the interpolant $\pi\mathbf{u}$ are derived from those on $\mathbf{u}_P$ in exactly the same way as in Chapter 4, so that what is required is an interpolant that has the properties described in §4.3.4.

By construction, then, $\mathbf{u}_P$ will satisfy the requirements given in §4.1.

The new interpolant, designed to fulfil the requirements of §4.3.4, is defined below, but the process of design and construction, similar to that of the two-dimensional case, is omitted. Some additional details are included in Appendix B.

6.1.1 The interpolant

Define the space $\hat{Q}$ on the reference element $[-1,1]^3$ by

$$\hat{Q} = \text{span}\left(\left\{\begin{array}{l} 1, x, y, z, \theta(x) + \kappa(y), \theta(x) + \kappa(z) \\ 1, x, y, z, \theta(y) + \kappa(z), \kappa(x) + \theta(y) \\ 1, x, y, z, \kappa(x) + \theta(z), \kappa(y) + \theta(z) \end{array}\right\}\right) \tag{6.1.1}$$

or

$$\hat{Q} = \text{span}\left(\left\{\begin{array}{l} 1, x, y, z, L_1(x,y), L_2(x,z) \\ 1, x, y, z, M_1(y,z), M_2(x,y) \\ 1, x, y, z, N_1(x,z), N_2(y,z) \end{array}\right\}\right), \tag{6.1.2}$$

where L_1, L_2, M_1, M_2, N_1 and N_2 are defined as the correspondence indicates, and θ and κ are defined as before.

Number the six faces beginning with the face with normal $(1,0,0)$, then its opposite, with normal $(-1,0,0)$, then with the normals in the y-direction and finally in the z, positive before negative in each case. (These are listed by number in B.2.1.)

Evaluating each of the higher-order functions at the midpoint of each face of $[-1,1]^3$, and calculating the mean value of each on each face give corresponding results: that is, the mean values are equal to the midpoint values in each case. (Details are provided in B.2.2 and B.2.3.)

Since the same holds for the linear and constant basis functions, any function in $\hat{Q}$ will have its mean value on a face equal to its value at the midpoint of that face.

Using the midpoints as nodal values gives a unique set of basis functions for each component: $\phi_i^x(x,y,z)$, $\phi_i^y(x,y,z)$ and $\phi_i^z(x,y,z)$, $i=1,...,6$. (These are detailed in B.2.4.)

By definition,

$$\phi_i^x(\mathbf{m}_j) = \delta_{ij} \tag{6.1.3}$$

for $i,j=1,...,6$, and by the relationship between mean and midpoint values,

$$\tfrac{1}{2}\int_{E^j} \phi_i^x(x,y,z)\, ds = \delta_{ij} \tag{6.1.4}$$

for each face E^j of the reference element. Similar results hold for all ϕ_i^y and ϕ_i^z.

As before, an affine map to the real domain gives the space $Q(\Omega_e)$ on each element, which leads to a nonconforming space NC_h, continuous at the midpoints of the element faces, and on each Dirichlet boundary face having a midpoint value equal to the mean of the Dirichlet function on that face. For the x-component, the operator $\pi_e^x : H^1(\Omega_e) \to Q^x(\Omega_e)$ is defined by

$$\int_{E^j} \pi_e^x v\, ds = \int_{E^j} v\, ds \tag{6.1.5}$$

for each face E^j of Ω_e, and $\pi^x : H^1_{\Gamma_D^x}(\Omega) \to NC_h^x$ is defined such that

$$\pi^x v|_{\Omega_e} = \pi_e^x v, \tag{6.1.6}$$

and similarly for the other components. Again, by concatenation one obtains $\pi_e : \left[H^1(\Omega_e)\right]^3 \to Q(\Omega_e)$ and $\pi : H^1_{\Gamma_D^x}(\Omega) \times H^1_{\Gamma_D^y}(\Omega) \times H^1_{\Gamma_D^z}(\Omega) \to NC_h$.

Because of the affine map, the identity

$$\pi_e \mathbf{v}(\mathbf{m}_j) = \frac{1}{h_{E^j}} \int_{E^j} \pi_e \mathbf{v}(x,y,z)\, ds = \frac{1}{h_{E^j}} \int_{E^j} \mathbf{v}(x,y,z)\, ds \tag{6.1.7}$$

again holds for every face E^j of Ω_e, and with midpoint continuity,

$$\pi \mathbf{v}(\mathbf{m}_E) = \frac{1}{h_E} \int_E \pi \mathbf{v}(x,y,z)\, ds = \frac{1}{h_E} \int_E \mathbf{v}(x,y,z)\, ds \tag{6.1.8}$$

for every $E \in \Gamma$.

6.1.2 The linear projection of the interpolant: $\mathbf{u}_P$

As in the case of the two-dimensional domain, the function $\mathbf{u}_P$ for error-splitting is the projection of the interpolant onto the space of linear polynomials on the element. The function $\mathbf{q}$ is defined as before to be the higher-order portion of $\pi\mathbf{u}$.

Evaluation of the integrals of the higher order basis functions on each face shows that on the faces with normals in the x-direction, the functions in the x-component (L_1 and L_2) integrate to zero; similarly, on those with the normals in the y- and z-directions respectively, the y- and z-component functions respectively integrate to zero. Thus

$$\int_E \mathbf{q} \cdot \mathbf{n}\, ds = 0 \tag{6.1.9}$$

on each face of $[-1,1]^3$, showing that the desired property (4.3.10) holds. This, together with (6.1.8), gives

$$\int_E (\mathbf{u} - \mathbf{u}_P) \cdot \mathbf{n}\, ds = 0 \tag{6.1.10}$$

for every $E \in \Gamma$ so that (4.1.1) is satisfied and

$$\int_{\partial\Omega_e} (\mathbf{u} - \mathbf{u}_P) \cdot \mathbf{n}\, ds = 0 \tag{6.1.11}$$

for every element Ω_e.

Smoothness of the projection leads to

$$\int_{\Omega_e} \nabla \cdot (\mathbf{u} - \mathbf{u}_P)\, dx = 0, \tag{6.1.12}$$

from which follow the bounds on $\nabla \cdot (\mathbf{u} - \mathbf{u}_P)$ as in §4.3.2; the bounds (4.5.1) on $\mathbf{u}_P$ are also obtained as before.

Thus $\mathbf{u}_P$ satisfies the requirements listed in §4.1.

6.2 Equivalence of under-integration to projections

For three-dimensional elements, under-integration will take place on specified face terms in the IP formulation, so an equivalence result in two dimensions similar to that of §5.2.1 (which is in one dimension) is required.

In two dimensions, define I_1^x and I_1^y as numerical integration operators in the x and y directions respectively, and w^x and w^y as the corresponding weights. Here, $f \in \mathbb{Q}_1$, which means it is linear in each component; $\tilde{f}$ is its constant interpolant and $g \in \mathbb{P}_0$. Π_0 is the projection onto constants on the two-dimensional surface.

Using similar arguments to those used in the one-dimensional case in §5.2.1,

$$\begin{aligned}
\int\int \tilde{f}g \; dx \; dy &= \int \left(\int \tilde{f}g \; dx \right) \; dy \\
&= \int \left(I_1^x \left\{ \tilde{f}g \right\} \right) \; dy \\
&= I_1^y \left\{ I_1^x \left\{ \tilde{f}g \right\} \right\} \\
&= I_1^y \left\{ \tilde{f}\left(x_{\mathrm{IP}}, y\right) g\left(x_{\mathrm{IP}}, y\right) w^x \right\} \\
&= \tilde{f}\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) g\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) w^x w^y \\
&= f\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) g\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) w^x w^y \\
&= I_1^y \left\{ f\left(x_{\mathrm{IP}}, y\right) g\left(x_{\mathrm{IP}}, y\right) w^x \right\} \\
&= I_1^y \left\{ I_1^x \left\{ f\left(x, y\right) g\left(x, y\right) \right\} \right\} \\
&= I_1^y \left\{ \int fg \; dx \right\} \\
&= \int\int fg \; dx \; dy \\
&= \int\int \left(\Pi_0 f\right) g \; dx \; dy. \qquad (6.2.1)
\end{aligned}$$

Therefore, as before,

$$\tilde{f} = \Pi_0 f. \qquad (6.2.2)$$

Then, for $f, h \in \mathbb{Q}_1$, and $\tilde{f}$ and $\tilde{h}$ their constant interpolants,

$$\begin{aligned}\int\int \Pi_0 f \; \Pi_0 h \; dx \; dy &= \int\int \tilde{f}\tilde{h} \; dx \; dy \\ &= I_1^y \left\{ I_1^x \left\{ \tilde{f}\tilde{h} \right\} \right\} \\ &= I_1^y \left\{ \tilde{f}\left(x_{\mathrm{IP}}, y\right) \tilde{h}\left(x_{\mathrm{IP}}, y\right) w^x \right\} \\ &= \tilde{f}\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) \tilde{h}\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) w^x w^y \\ &= f\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) h\left(x_{\mathrm{IP}}, y_{\mathrm{IP}}\right) w^x w^y \\ &= I_1^y \left\{ f\left(x_{\mathrm{IP}}, y\right) h\left(x_{\mathrm{IP}}, y\right) w^x \right\} \\ &= I_1^y \left\{ I_1^x \left\{ f\left(x, y\right) h\left(x, y\right) \right\} \right\}. \end{aligned} \tag{6.2.3}$$

Therefore, as in the one-dimensional case, using L_2-projections onto the space of constants is equivalent to using one-point integration, for products of functions in $\mathbb{Q}_1$.

6.3 The convergence analysis

Considering the convergence analysis itself, much of the detail remains unchanged between two and three dimensions.

Firstly, coercivity and consistency of the methods are independent of the dimension of the problem.

With the interpolant and its projection defined, and their properties established, the approximation error is split as before. All the details of the manipulations used in obtaining the error bounds hold for the three-dimensional domain as for the two-dimensional domain, with the substitution of faces for edges (with occasional variation in the values of the constants only). Again, the domain is required to satisfy the regularity result (A.6.1), although no explicit specification of conditions under which this is achieved is available (see A.6).

Thus the approximation errors of the IP methods are bounded exactly as before, for trilinear approximations, for linear approximations, and for either of these used in conjunction with under-integration. That is, the results of Chapter 5, and specifically of Theorems 1, 2 and 3, hold exactly for the analagous cases in three-dimensional domains.

Chapter 7

Numerical results

This chapter presents the results of a range of numerical examples based on the IP methods described in previous chapters. After the benchmark problems are defined, and important aspects of the implementation are described, a spectrum of results are presented and discussed. The first set contains comparisons between the use of triangular and bilinear quadrilateral elements with the original, unmodified methods. The next set relates to the core of the analyses performed: the results of applying the remedies that have been proposed and analysed.

Results of under-integrating other non-central combinations of terms are then presented to illustrate the analytical conclusions regarding these cases.

The final two sections deal with the preliminary investigation of problems that could potentially form the basis of future, related work. The first is the case of meshes of non-rectangular elements, which is not covered in the theory developed here: a set of numerical results is used to begin to evaluate the potential effectiveness of the proposed remedies for this extension. The second regards the use of higher-order quadrilateral elements ($\mathbb{Q}_k$, with $k \geq 2$) with the IP methods as a proposed solution to the locking problem.

7.1 Model problems

Five model problems are used throughout this chapter, providing variation in boundary conditions, loading types, and number of dimensions. The problems in two-dimensional domains (the first four) use plane strain conditions, and standard units are used.

7.1.1 Cantilever beam (CB)

A square beam in two dimensions, with $E = 15000$, is subjected to a linearly varying force on the free end, with the maximum value $f = 3000$, as illustrated in Figure 7.1. The height of the beam is denoted by h.

The analytical solution (see [19]) is

$$u_1 = 2f\frac{1-\nu^2}{Eh}x\left(\frac{h}{2}-y\right)$$
$$u_2 = f\frac{1-\nu^2}{Eh}\left[x^2+\frac{\nu}{1-\nu}y\left(y-h\right)\right].$$

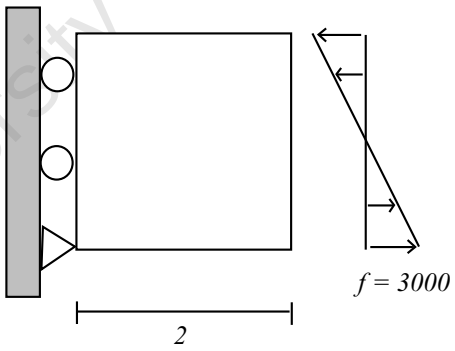


Figure 7.1: Cantilever beam with boundary conditions

While the Dirichlet boundary condition of constrained displacement in the x-direction is imposed weakly, according to the IP method as usual, the constraint of the pinned corner is included by setting the displacement degrees of freedom associated with that vertex to zero.

7.1.2 Square plate (SP)

The linear elastic unit square plate $[0,1]^2$ described in [8], with $\mu = 1$, is fixed on all its edges and subjected to an internal body force $\mathbf{f}$:

$$f_1 = 0.04\,\pi^2 \left[4 \sin 2\pi y \,(-1 + 2\cos 2\pi x) - \cos \pi\,(x+y) + \frac{2}{1+\lambda} \sin \pi x \sin \pi y\right],$$
$$f_2 = 0.04\,\pi^2 \left[4 \sin 2\pi x \,(1 - 2\cos 2\pi y) - \cos \pi\,(x+y) + \frac{2}{1+\lambda} \sin \pi x \sin \pi y\right].$$

The exact solution is

$$u_1 = 0.04 \left[\sin 2\pi y \,(-1 + \cos 2\pi x) + \frac{1}{1+\lambda} \sin \pi x \sin \pi y\right],$$
$$u_2 = 0.04 \left[\sin 2\pi x \,(1 - \cos 2\pi y) + \frac{1}{1+\lambda} \sin \pi x \sin \pi y\right].$$

7.1.3 Cook's membrane (CM)

A tapered panel, fixed on one edge, is subjected to a uniformly distributed shear force of 100 on the opposite edge (see Figure 7.2). Here $E = 250$. The mesh is non-affine.

7.1.4 T-shaped bracket (TB)

The bracket shown in Figure 7.3, constrained in one direction only on each relevant edge, is subjected to a uniform pressure of $p = 6000$ on the top edge. Here $E = 55 \times 10^6$. The mesh, as used in the three-dimensional version of the problem in [29], is non-uniform, and the elements rectangular in all but the curved regions.

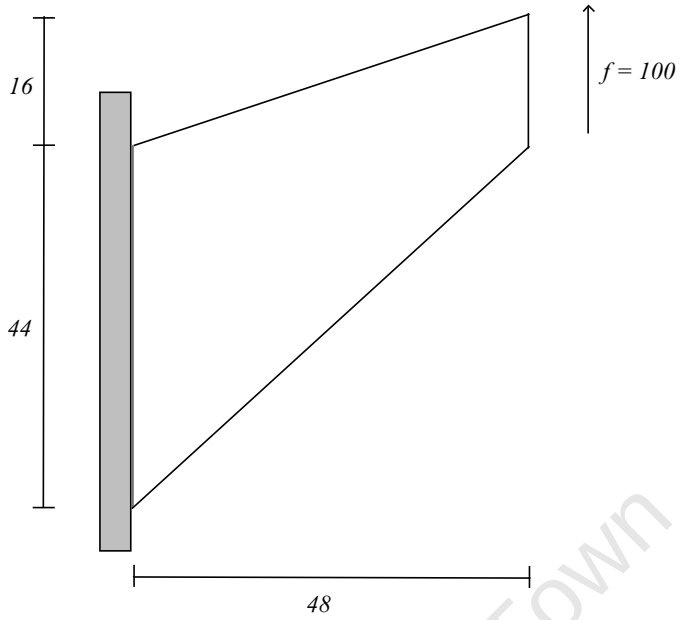


Figure 7.2: Cook's membrane with boundary conditions

7.1.5 Cube with trigonometric body force

The linear elastic unit cube $[0,1]^3$, with $E = 15 \text{ x } 10^6$, is fixed on all its faces and subjected to an internal body force $\mathbf{f}$:

$$
\begin{aligned}
f_1 &= \mu\pi^2 \left[(9\cos 2\pi x - 5)(\sin 2\pi y \sin \pi z - \sin \pi y \sin 2\pi z) + \frac{3}{1+\lambda} \sin \pi x \sin \pi y \sin \pi z \right] \\
&\quad + (\mu + \lambda)\pi^2 (\sin \pi x \sin \pi y \sin \pi z - \cos \pi x \cos \pi y \sin \pi z - \cos \pi x \sin \pi y \cos \pi z), \\
f_2 &= \mu\pi^2 \left[(9\cos 2\pi y - 5)(\sin 2\pi z \sin \pi x - \sin \pi z \sin 2\pi x) + \frac{3}{1+\lambda} \sin \pi x \sin \pi y \sin \pi z \right] \\
&\quad + (\mu + \lambda)\pi^2 (\sin \pi x \sin \pi y \sin \pi z - \cos \pi x \cos \pi y \sin \pi z - \sin \pi x \cos \pi y \cos \pi z), \\
f_3 &= \mu\pi^2 \left[(9\cos 2\pi z - 5)(\sin 2\pi x \sin \pi y - \sin \pi x \sin 2\pi y) + \frac{3}{1+\lambda} \sin \pi x \sin \pi y \sin \pi z \right] \\
&\quad + (\mu + \lambda)\pi^2 (\sin \pi x \sin \pi y \sin \pi z - \cos \pi x \sin \pi y \cos \pi z - \sin \pi x \cos \pi y \cos \pi z).
\end{aligned}
$$

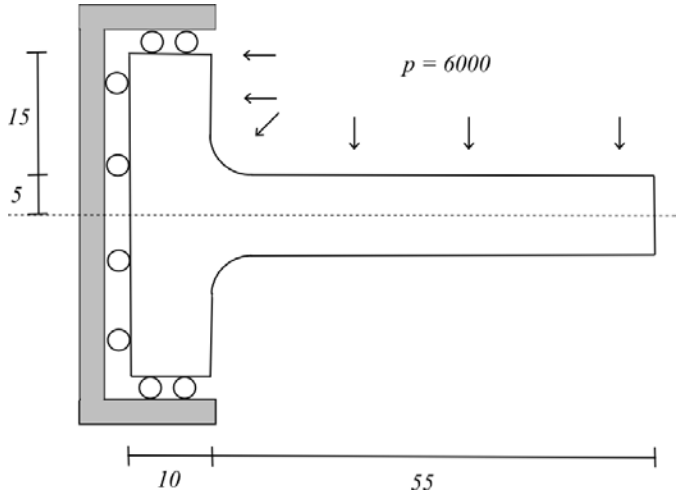


Figure 7.3: T-shaped bracket with boundary conditions

The exact solution is

$$u_1 = (\cos 2\pi x - 1)(\sin 2\pi y \sin \pi z - \sin \pi y \sin 2\pi z) + \frac{1}{1+\lambda} \sin \pi x \sin \pi y \sin \pi z,$$
$$u_2 = (\cos 2\pi y - 1)(\sin 2\pi z \sin \pi x - \sin \pi z \sin 2\pi x) + \frac{1}{1+\lambda} \sin \pi x \sin \pi y \sin \pi z,$$
$$u_3 = (\cos 2\pi z - 1)(\sin 2\pi x \sin \pi y - \sin \pi x \sin 2\pi y) + \frac{1}{1+\lambda} \sin \pi x \sin \pi y \sin \pi z.$$

7.2 Technical aspects

7.2.1 Parameters

Poisson's ratio is set to $\nu = 0.49995$ where near-incompressibility is investigated. For compressible materials, $\nu = 0.3$ is used.

The choice of values for the stabilization parameters k_μ and k_λ is based firstly on the theoretical considerations of stability discussed in earlier chapters. The NIPG method, stable with $k_\lambda \geq 0$, is by default used without the second stabilization term present, that is, with $k_\lambda = 0$. The IIPG and SIPG methods are by default used with the second stabilization term included. While all the methods require $k_\mu > 0$, there is no minimum positive value required for NIPG to be stable, while IIPG and SIPG require

both parameters to be greater than a positive constant which can be calculated (see, for example, [24]).

Within the constraints of these coercivity requirements, a single value for the parameters for all the methods was desirable for consistency of results, for the sake of useful comparison. A series of preliminary tests on two benchmark problems indicated that choosing $k_\mu = k_\lambda = 10$ gives a consistent reflection of the behaviour of the various methods, and therefore these values have been adopted.

7.2.2 Meshes

Each of the model problems above except that of the bracket (TB) has been solved on a sequence of meshes at increased refinement levels.

The bracket mesh is non-uniform, but consists almost entirely of rectangular elements.

The other four domains have each been refined isotropically from a single quadrilateral or cube element. Where triangular elements have been used, these have been obtained by cutting each quadrilateral element diagonally, at each refinement level.

7.2.3 Presentation of results

Deformed shapes are shown for only one refinement level in each case (32×32 elements for the two-dimensional problems, and $16\times16\times16$ for the three-dimensional problem, except where otherwise noted). For visualisation of the cube, the elements beyond a threshold, slanted plane have been extracted so that the deformation inside the cube can be viewed. (By this procedure, an element is removed in its entirety based on the relationship of its central point to the threshold plane, resulting sometimes in slight differences in deformation appearing greater than they are. This should be taken into account when the results are compared.) Plots of the H^1 error (that is, the approximation error in the broken H^1-norm) display the rates of convergence with refinement. This is the appropriate norm with which to measure the error, due to its equivalence to the DG norm that has been used in the analyses (see (2.2.4)).

The first (extreme right-hand) data point on the error plot for a two-dimensional problem represents the error of a 2×2-element mesh, and thus the deformed shapes shown correspond to the fifth point from the right, except where otherwise noted. The first data point for the cube represents a single element, so that the deformed shapes shown again correspond to the fifth point.

Where an analytical solution is available, it is displayed for comparison. Otherwise, a solution obtained using biquadratic elements with the Standard Galerkin (SG) method is shown.

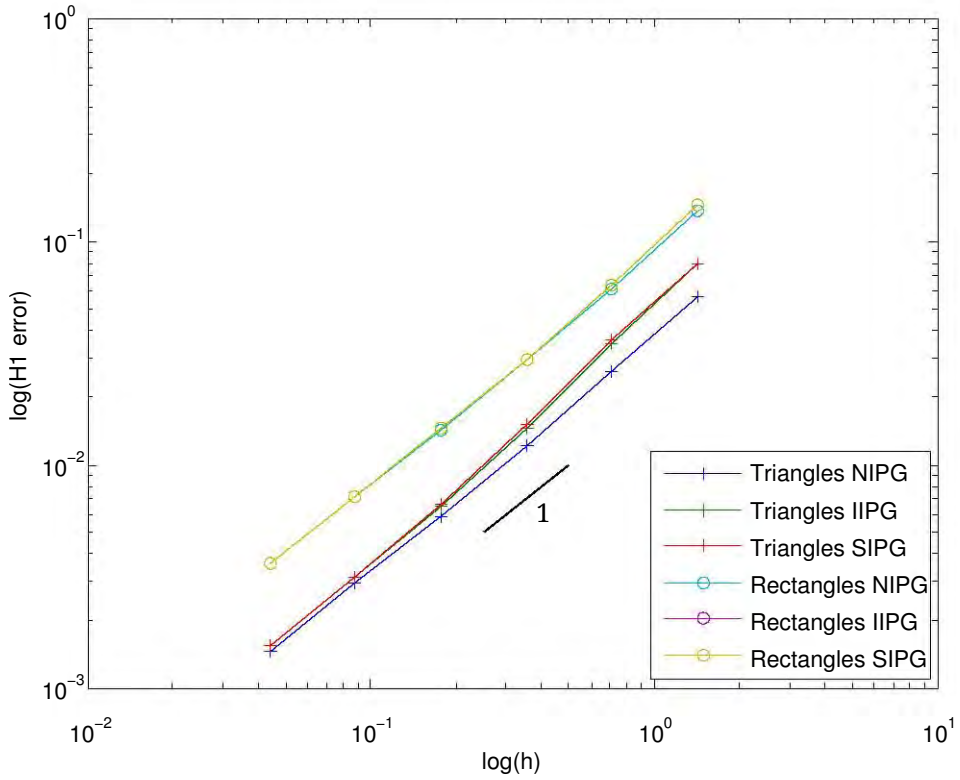
Scaling of the deformation has been applied in the case of the bracket for clearer visualisation, with the vertical displacement being magnified five-fold. The deformation of the cube has been scaled down to 10%.

7.2.4 Software for implementation

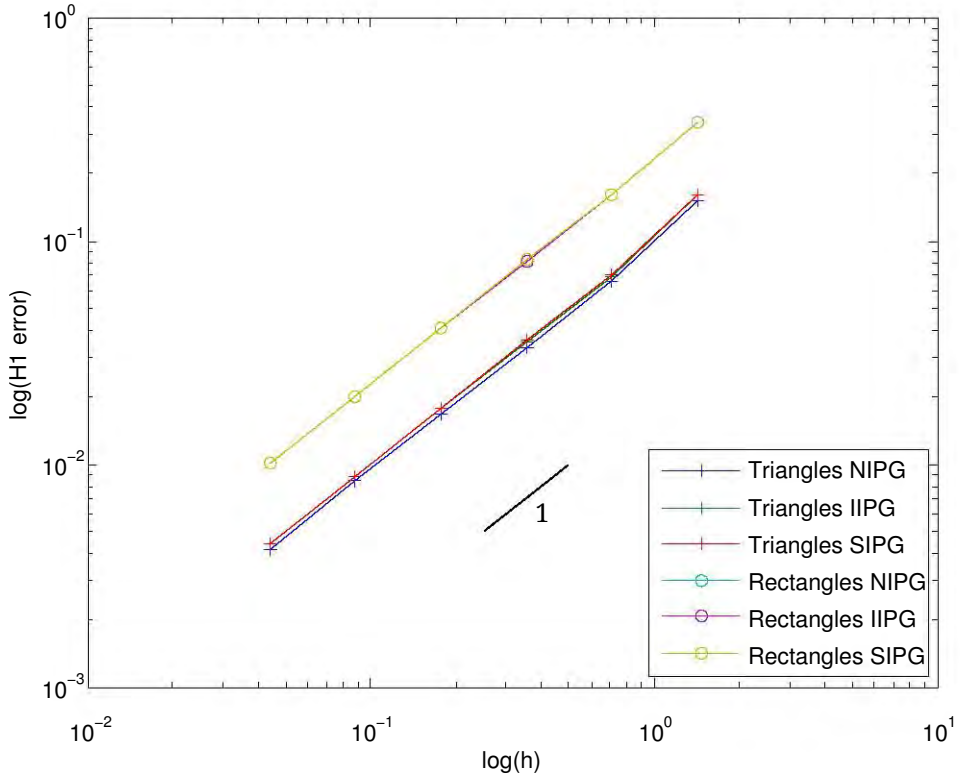
The DG methods were implemented primarily in C++, using the deal.II finite element analysis library ([4], [5]); Matlab was used for implementing the IP methods on meshes of triangular elements. ParaView and Matlab were used for post-processing and visualisation.

7.3 Comparing triangular to quadrilateral elements

This section contains the results that were the motivation for the investigation discussed in this thesis: the disparity between the use of triangular elements and the use of quadrilateral elements in solving near-incompressible elasticity problems with the IP methods. Three model problems (CB, SP, and CM) are solved with the NIPG, IIPG and SIPG methods in turn, and the results from using triangular linear elements are contrasted to those from using quadrilateral bilinear elements. In each case, deformed shapes are shown, with both an accurate solution and the result obtained using the Standard Galerkin method with $\mathbb{Q}_1$ elements included for comparison. Error convergence plots are shown for the CB and SP problems for both compressible and nearly incompressible materials.

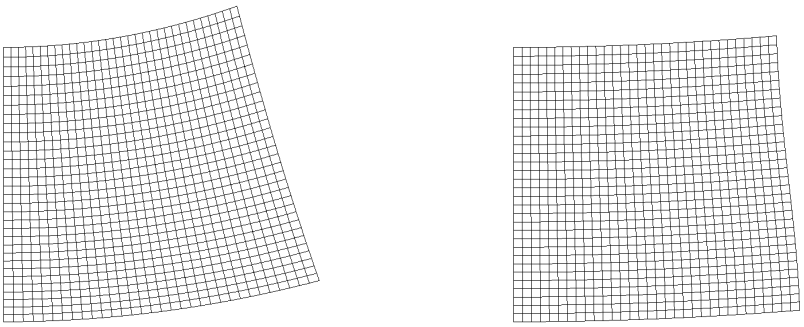


(a) Cantilever beam



(b) Square plate

Figure 7.4: Convergence of H^1 error using triangular or rectangular elements, $\nu = 0.3$



(a) Exact solution (b) SG with $\mathbb{Q}_1$ elements

Figure 7.5: Cantilever beam, $\nu = 0.49995$

For compressible materials, with $\nu = 0.3$, triangular and quadrilateral elements produce optimal convergence rates of the H^1 error (Figure 7.4), with all three IP methods. Thus, at a low value of λ, there is no significant disparity between the results.

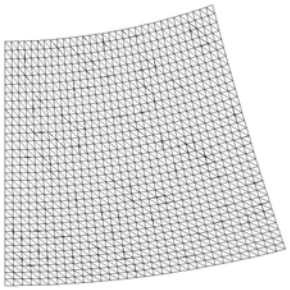
For nearly incompressible materials, in all three boundary value problems shown here, the deformed shapes resulting from using the IP methods with triangles (Figures 7.6, 7.9 and 7.12) are, to the eye, identical to those of the exact or higher-order solutions (Figures 7.5a, 7.8a and 7.11a), while the quadrilateral elements produce quite different results. The convergence plots (Figures 7.7 and 7.10) give corroborating information, showing optimal convergence rates when triangles are used, and poor convergence on rectangles.

For the cantilever beam (Figure 7.6), NIPG with rectangles produces a poor approximation for the near-incompressible case, with very little displacement in the vertical direction and deformation in the horizontal, qualitatively similar to the results of the IIPG and SIPG methods, but more exaggerated in the horizontal displacement. The IIPG and SIPG results are very much like those of the SG method with $\mathbb{Q}_1$ elements (Figure 7.5b). The error plots (Figure 7.7) show that, with rectangular elements, the IP methods produce results comparable to the SG method, that is, with very poor convergence rates.

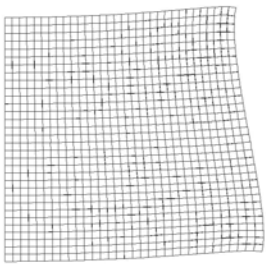
The second example, the square plate (Figure 7.9), again shows strong similarity between the IIPG and SIPG methods with rectangular elements with $\nu = 0.49995$, manifesting typical locking behaviour much like that produced by the SG method (Figure 7.8b), while the corresponding results with triangles are locking-free. The NIPG method produces different behaviour, exhibiting not locking but error in the jumps between elements, which can be seen from the lack of clarity at element edges in portions of the mesh. The errors of the IIPG and SIPG methods shown in Figure 7.10 are again comparable to that of the SG method, with only fractional decrease in error resulting from mesh refinement. The NIPG method with rectangles has a nearly optimal rate of convergence for the coarser meshes, but with further refinement the convergence rate decreases.

The final test problem shown here, Cook's membrane (Figure 7.12), shows typical locking behaviour for $\nu = 0.49995$ for all three IP methods with quadrilaterals, similar to the result using the SG method (Figure 7.11b), in contrast to the locking-free behaviour of the methods with triangular elements.

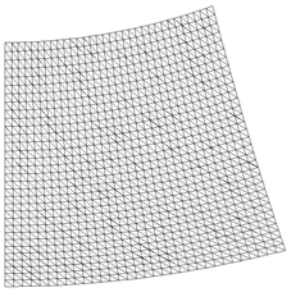
While the IP methods perform optimally for $\nu = 0.3$ with both triangular and quadrilateral elements, this quality of approximation is lost for quadrilateral elements in the near-incompressible regime, with, specifically, the well-known problem of locking manifesting in most of these poor approximations. This indicates a problematic dependence on the material parameters for quadrilaterals that is absent for triangles.



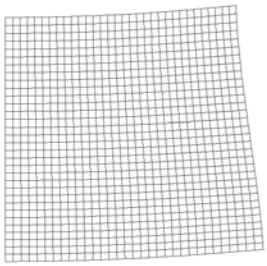
(a) NIPG with triangles



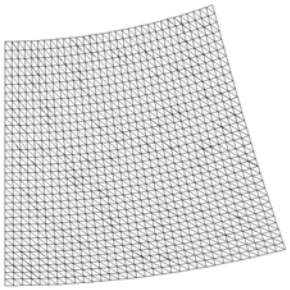
(b) NIPG with rectangles



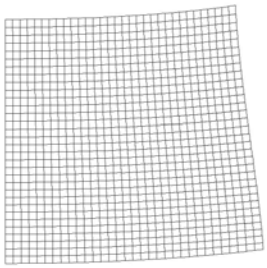
(c) IIPG with triangles



(d) IIPG with rectangles



(e) SIPG with triangles



(f) SIPG with rectangles

Figure 7.6: Cantilever beam: comparison of triangles and rectangles for the IP methods, $\nu = 0.49995$

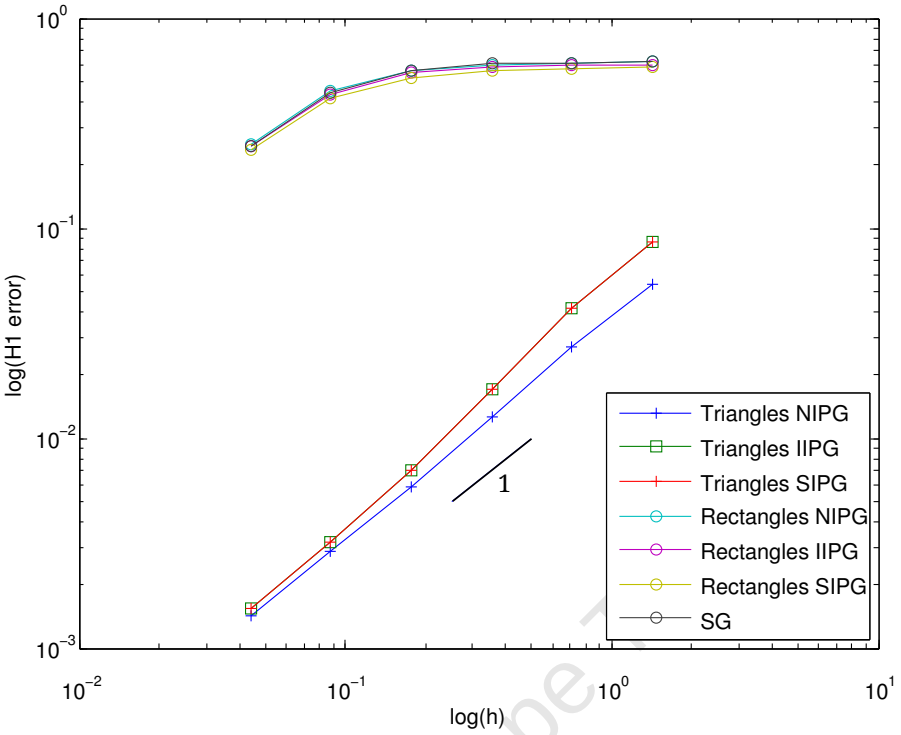
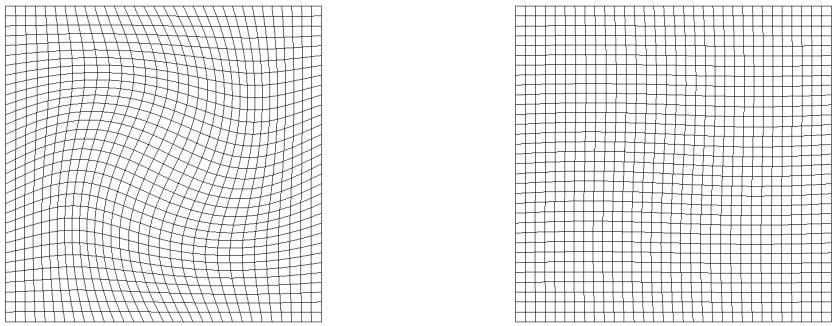


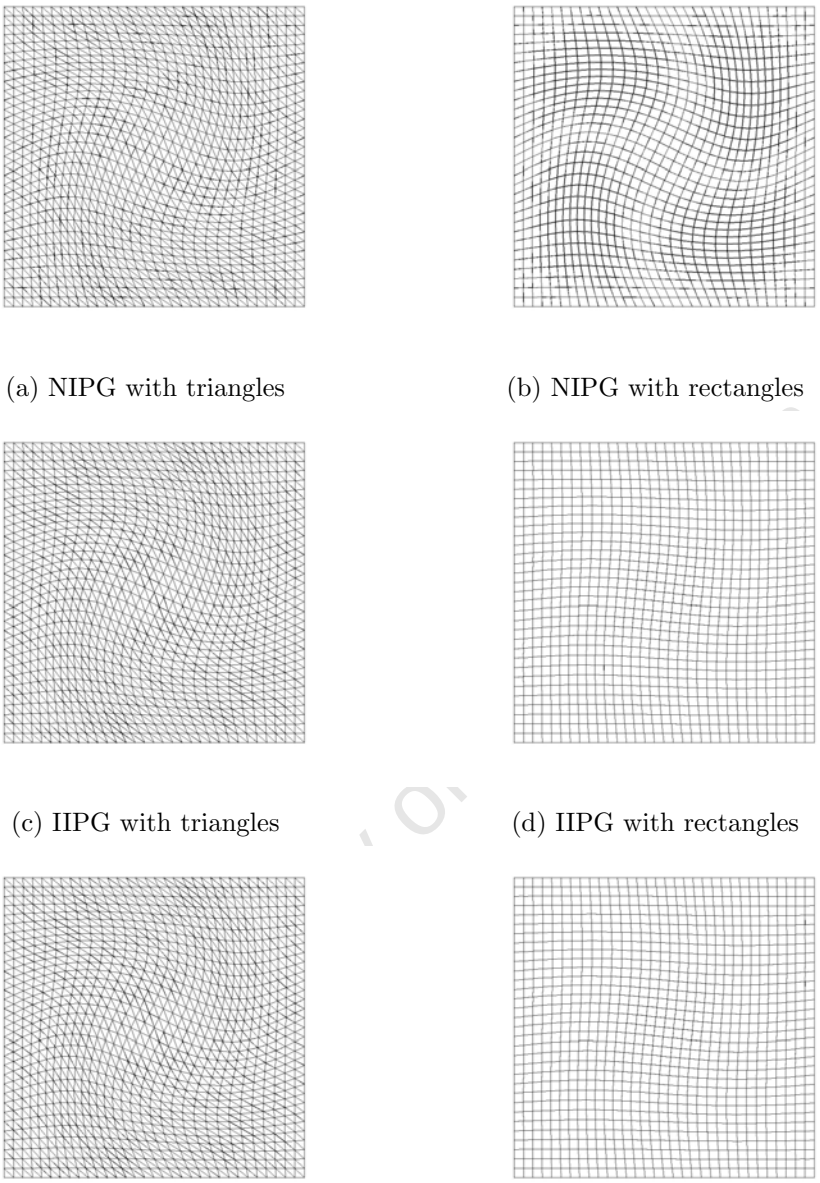
Figure 7.7: Convergence of H^1 error for the cantilever beam using triangular or rectangular elements, $\nu = 0.49995$



(a) Exact solution

(b) SG with $\mathbb{Q}_1$ elements

Figure 7.8: Square plate, $\nu = 0.49995$



(a) NIPG with triangles

(b) NIPG with rectangles

(c) IIPG with triangles

(d) IIPG with rectangles

(e) SIPG with triangles

(f) SIPG with rectangles

Figure 7.9: Square plate: comparison of triangles and rectangles for the IP methods, $\nu = 0.49995$

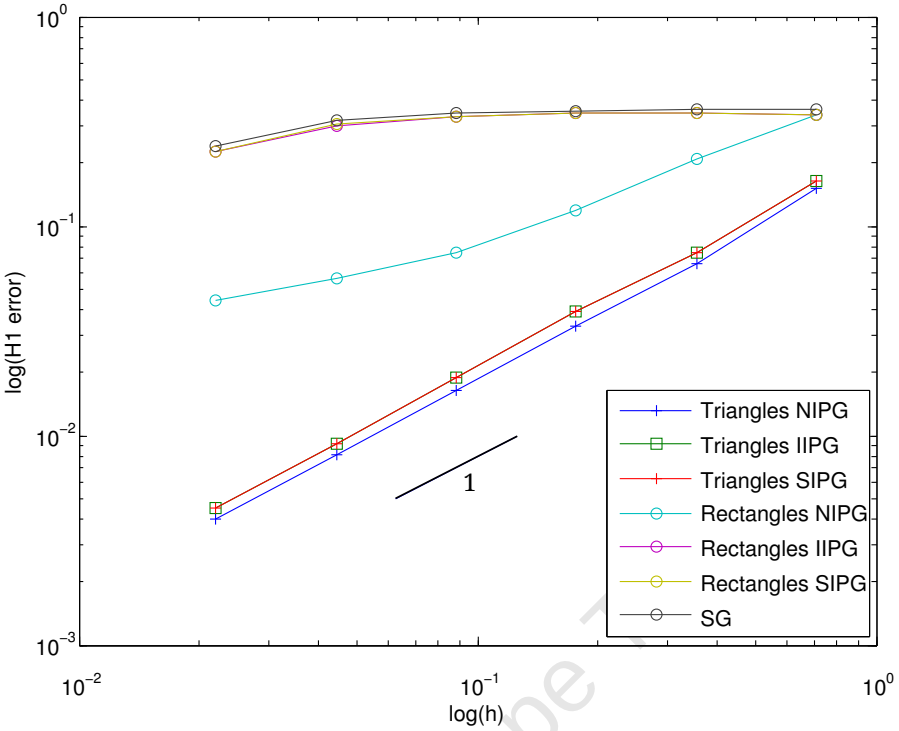
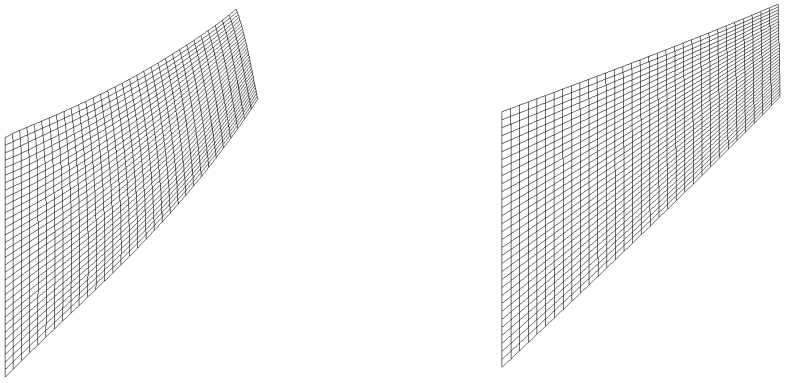


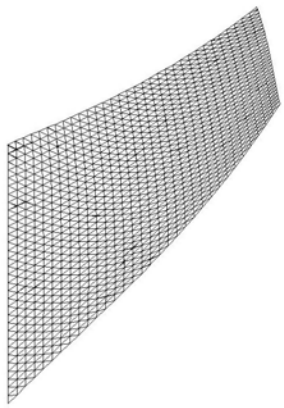
Figure 7.10: Convergence of H^1 error for the square plate using triangular or rectangular elements, $\nu = 0.49995$



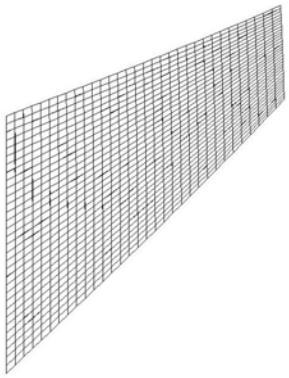
(a) SG with $\mathbb{Q}_2$ elements

(b) SG with $\mathbb{Q}_1$ elements

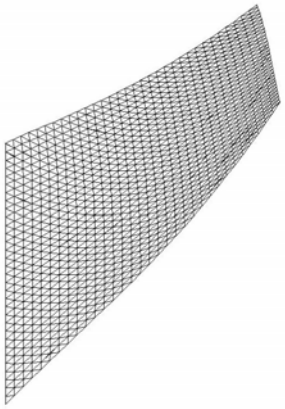
Figure 7.11: Cook's membrane, $\nu = 0.49995$



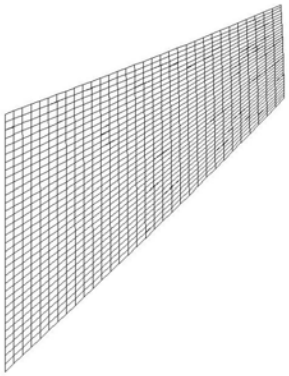
(a) NIPG with triangles



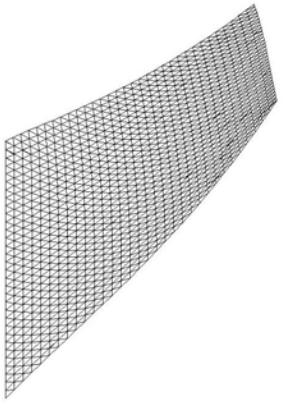
(b) NIPG with quadrilaterals



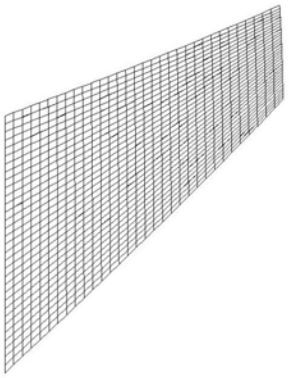
(c) IIPG with triangles



(d) IIPG with quadrilaterals



(e) SIPG with triangles



(f) SIPG with quadrilaterals

Figure 7.12: Cook's membrane: comparison of triangles and rectangles for the IP methods, $\nu = 0.49995$

7.4 Remedies for poor approximations

From the analyses of Chapter 5, the first proposed solution for the pathological behaviour observed is under-integration applied on the λ-dependent edge terms of the formulation, as prescribed by Theorem 1 in §5.2. The alternative proposal is the use of linear elements. As described in Theorem 2 in §5.3 and Theorem 3 in §5.4, linear elements with the NIPG method are expected to give optimal convergence and locking-free results, while with the IIPG and SIPG methods, the stabilization term requires under-integration when linear elements are used, for the same quality of approximations.

The results of under-integration on bilinear elements are shown for the cantilever beam problem, and the results of both of remedies are shown for three further boundary-value problems, SP, TB and the cube, all with $\nu = 0.49995$. (For the deformed shapes of CB and SP, the wireframe mesh depicts the IP approximation, while the solid colour represent the accurate – exact or higher-order – solution.) Error plots for each IP method compare the two remedies with the SG method ($\mathbb{Q}_1$ elements), with the IP method with bilinear elements with full-order integration, and with the IP method in conjunction with triangular elements.

(The reason for the omission of results for the beam with linear elements is related to the application of boundary conditions for this problem: the accurate conditions involve "pinning" a vertex, but using general linear elements on quadrilaterals, there are not naturally any degrees of freedom associated with the vertex where the zero-displacement constraint applies.)

In the first example, the beam, edge-term under-integration with $\mathbb{Q}_1$ elements gives very accurate solutions with all three IP methods (Figures 7.13, 7.15 and 7.17). The error plots indicate optimal convergence rates (Figures 7.14, 7.16 and 7.18).

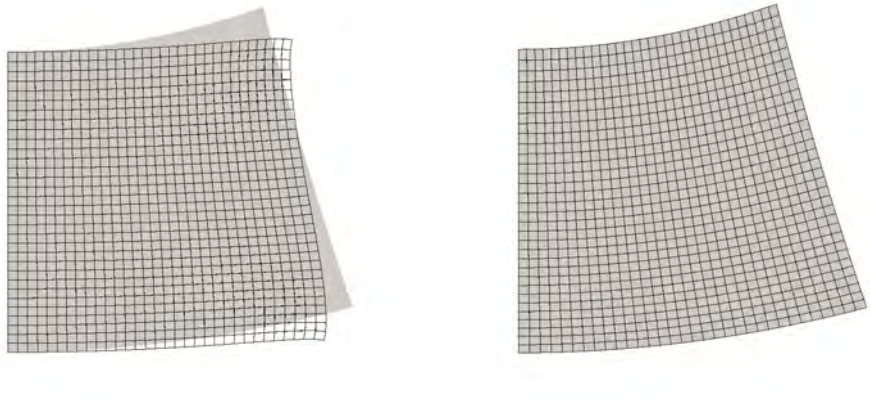
The square plate problems of locking with the IIPG and SIPG methods, and poor approximation with the NIPG method (Figures 7.19, 7.21 and 7.23) are seen to resolve with the application of under-integration to bilinear elements, and equally with the use of linear elements (with under-integration as necessary). The convergence plots (Figures 7.20, 7.22 and 7.24) demonstrate the effectiveness of both remedies, showing optimal rates of convergence for all three methods. The plots also show the poor convergence for the

SIPG and IIPG methods where linear elements are used with full-order integration in all terms, illustrating the need for under-integration of the stabilization term.

The bracket results are similar across the three IP methods, both remedies effective in relieving locking in all cases (Figures 7.25, 7.26 and 7.27).

The portions of the deformed cube (Figures 7.29, 7.31 and 7.33) show that for all three methods, under-integration and linear elements both produce results much like the exact solution (Figure 7.28), especially when compared to the locking behaviour with trilinear elements for the SIPG and IIPG methods. The deformed shape for the NIPG method does not show clearly that the approximation is poor, and indeed the error plot (Figure 7.30) indicates that it is relatively accurate for coarse meshes, but that the method is not robust, as the convergence rate is not consistently optimal. The rates for the under-integrated or linear elements are, in contrast, optimal. In Figures 7.32 and 7.34, the error plots for the IIPG and SIPG methods for full-order integration with either linear or trilinear elements show the negligible reduction in error with refinement that is typical of locking-type behaviour, while the convergence rates are optimal when the appropriate under-integration is applied in each case.

Across the scope of the boundary-value problems, it can be seen that the application of under-integration produces highly accurate results, as does the use of linear elements, with under-integration as appropriate. The error convergence plots display optimal rates of convergence when either remedy is used with any of the IP methods, unlike when bilinear elements are used with full-order integration, or linear elements with full-order integration for the IIPG or SIPG methods. This indicates that edge-term under-integration as prescribed in Theorem 1 is an effective remedy for the poor approximations produced by bilinear elements, as is the use of linear elements with under-integration as prescribed in Theorem 3, and illustrates the optimal, λ-independent convergence predicted in the three theorems.



(a) $\mathbb{Q}_1$ elements (b) With under-integration

Figure 7.13: Cantilever beam with NIPG, $\nu = 0.49995$

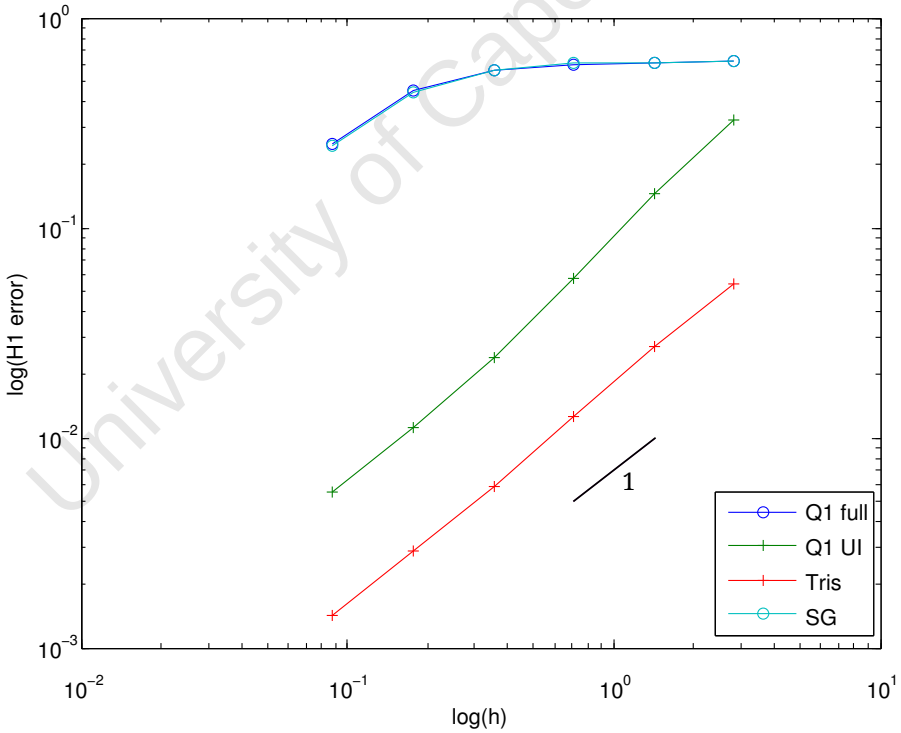
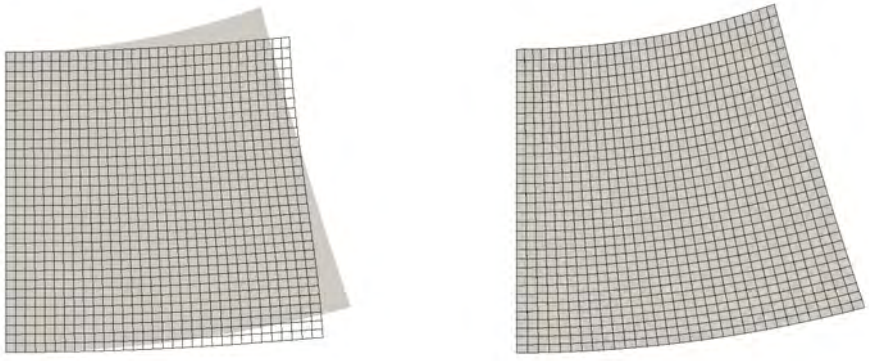


Figure 7.14: Comparison of H^1 errors for NIPG for the cantilever beam, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration

Figure 7.15: Cantilever beam with IIPG, $\nu = 0.49995$

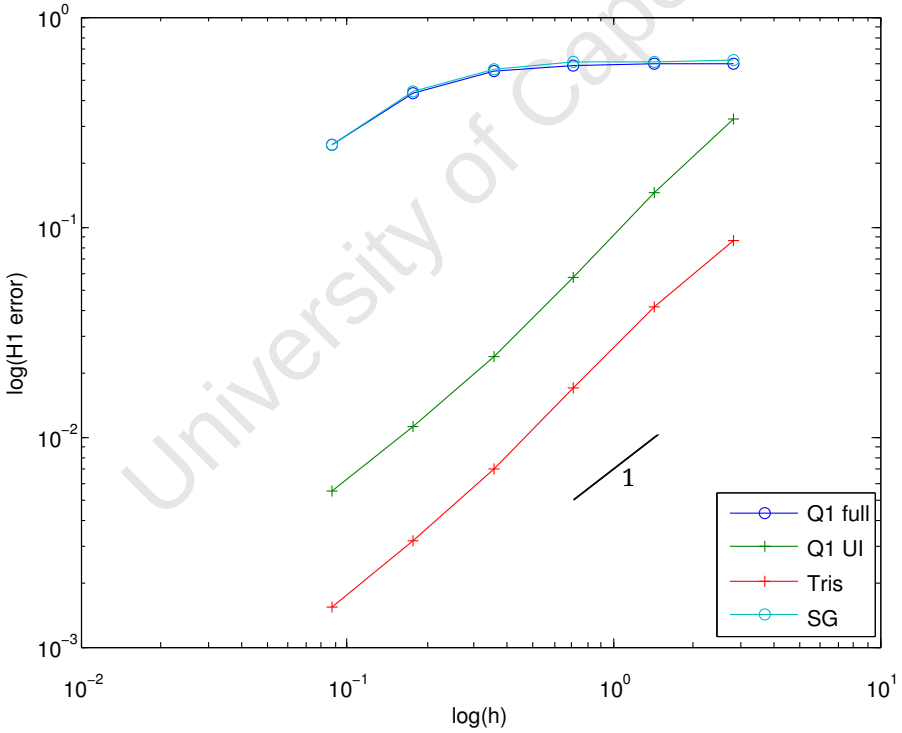
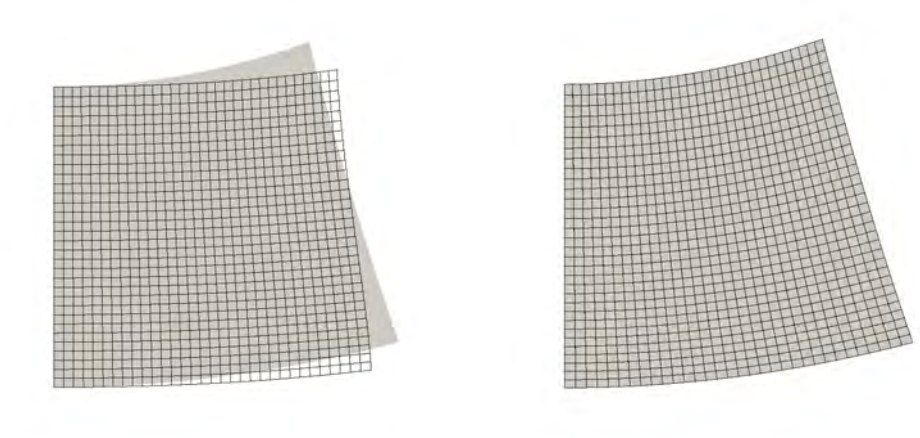


Figure 7.16: Comparison of H^1 errors for IIPG for the cantilever beam, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration

Figure 7.17: Cantilever beam with SIPG, $\nu = 0.49995$

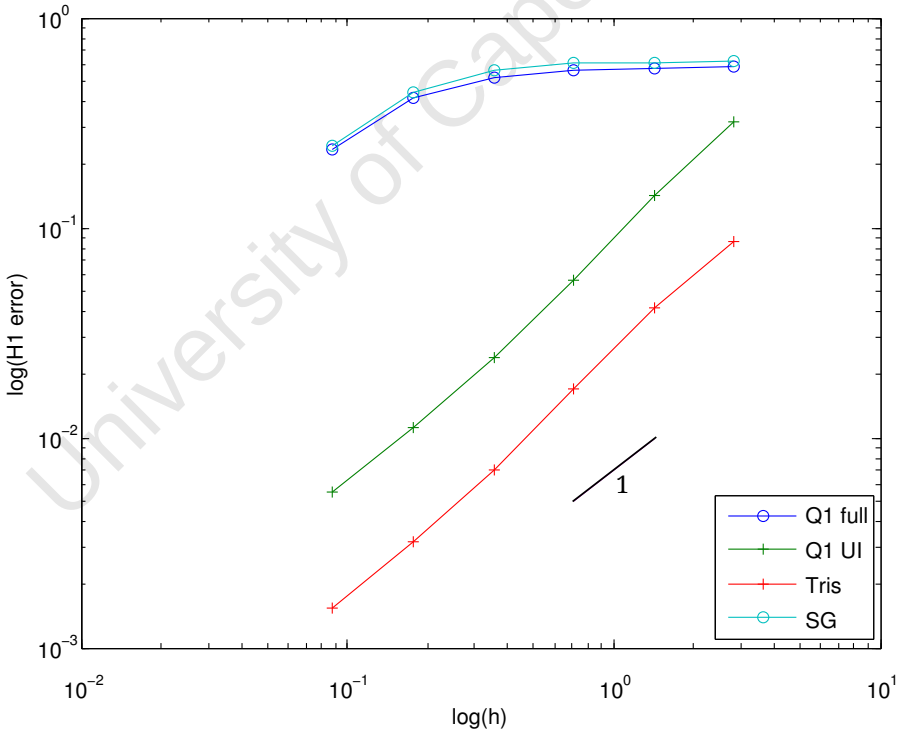
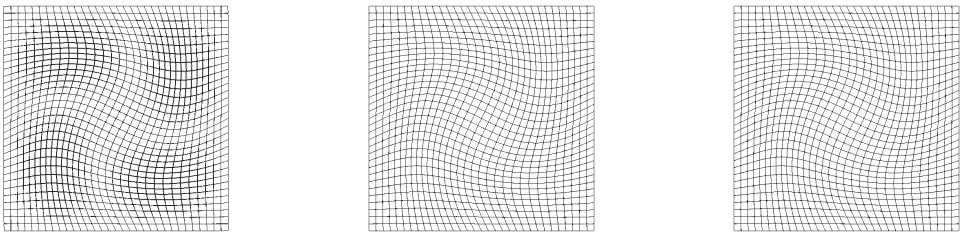


Figure 7.18: Comparison of H^1 errors for SIPG for the cantilever beam, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration (c) Linear elements

Figure 7.19: Square plate with NIPG, $\nu = 0.49995$

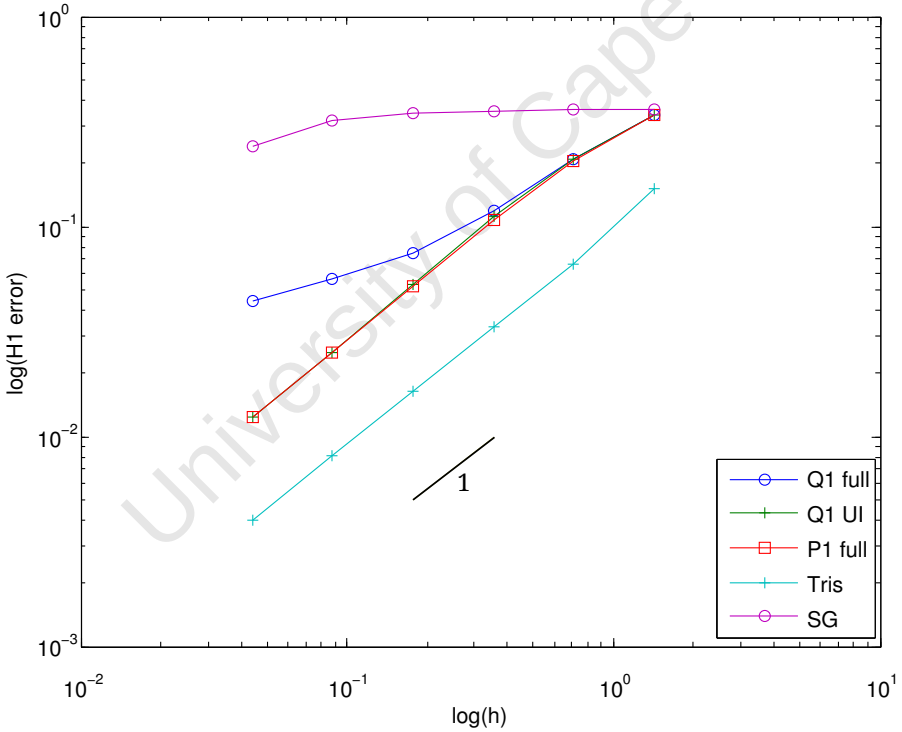
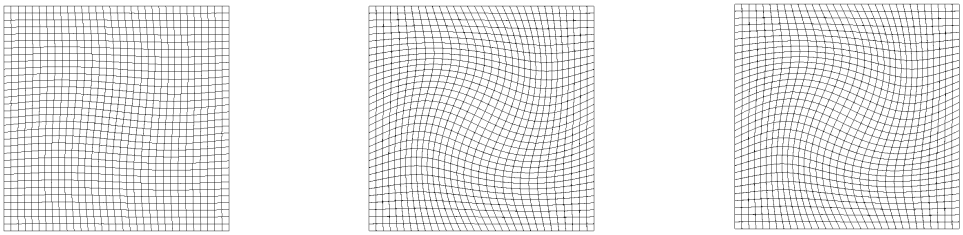


Figure 7.20: Comparison of H^1 errors for NIPG for the square plate, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration (c) Linear elements with under-integration

Figure 7.21: Square plate with IIPG, $\nu = 0.49995$

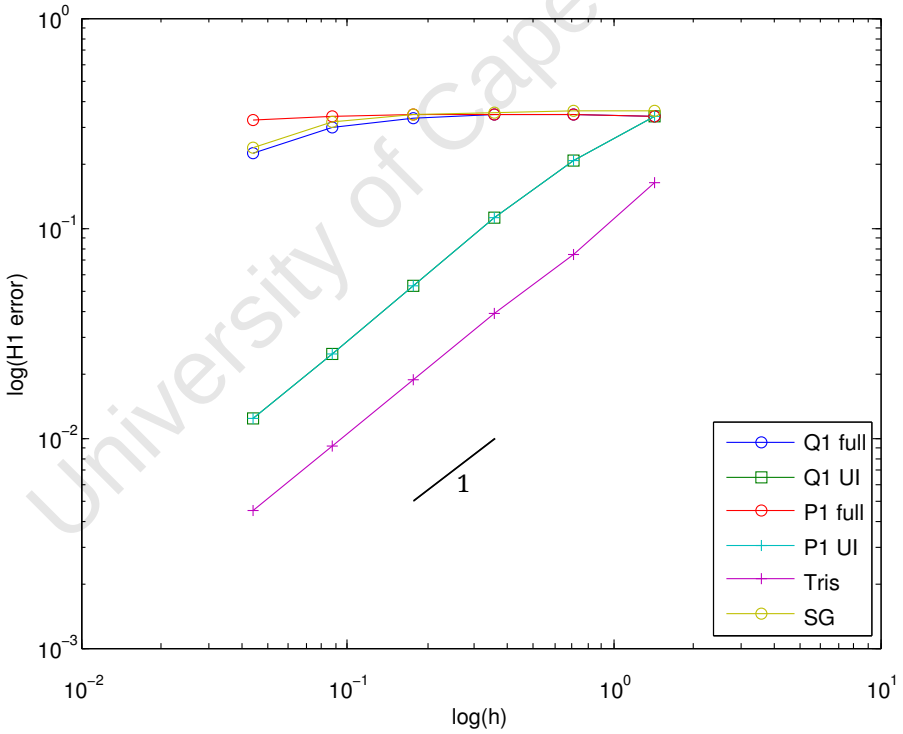
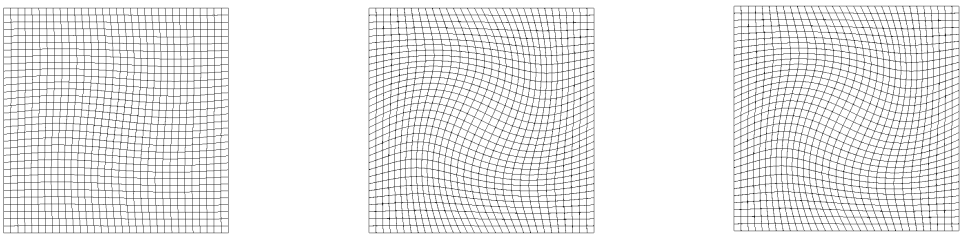


Figure 7.22: Comparison of H^1 errors for IIPG for the square plate, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration (c) Linear elements with under-integration

Figure 7.23: Square plate with SIPG, $\nu = 0.49995$

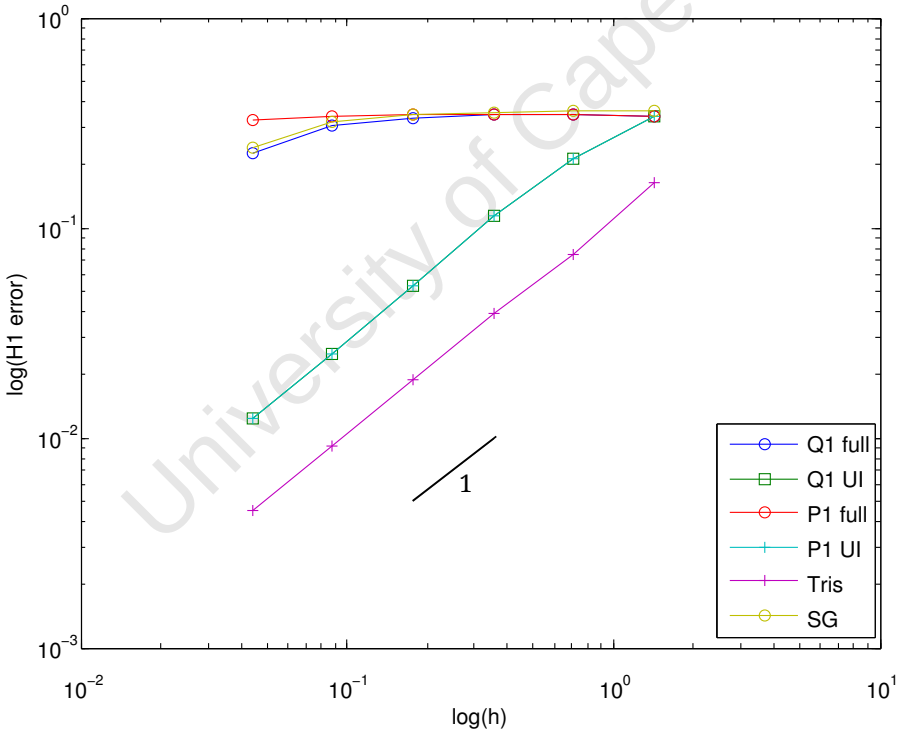
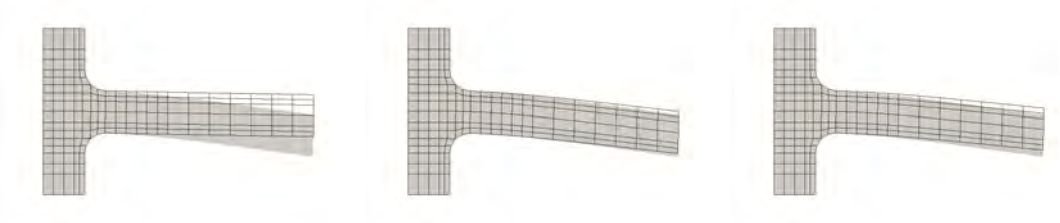
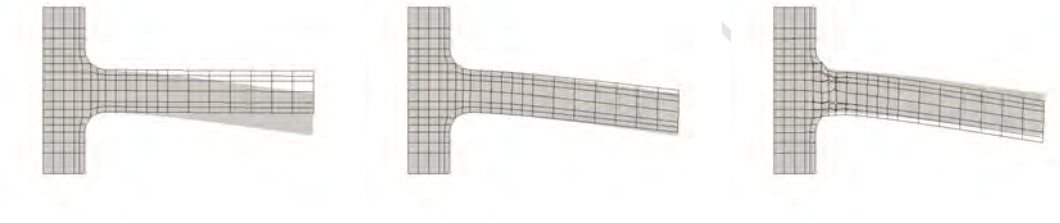


Figure 7.24: Comparison of H^1 errors for SIPG for the square plate, $\nu = 0.49995$



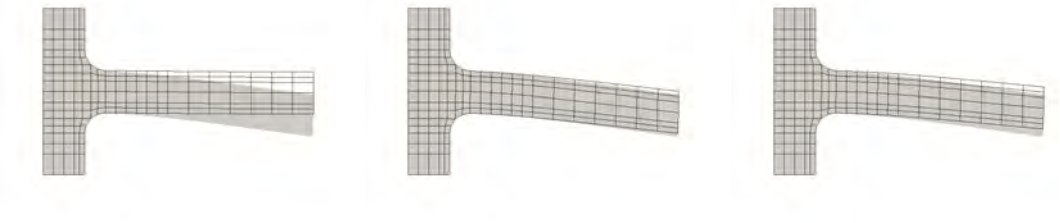
(a) $\mathbb{Q}_1$ elements (b) With under-integration (c) Linear elements

Figure 7.25: T-shaped bracket with NIPG, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration (c) Linear elements with UI

Figure 7.26: T-shaped bracket with IIPG, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration (c) Linear elements with UI

Figure 7.27: T-shaped bracket with SIPG, $\nu = 0.49995$

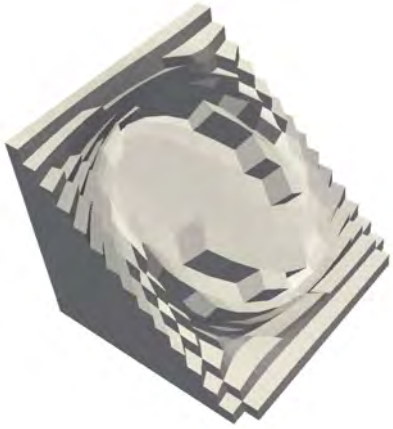
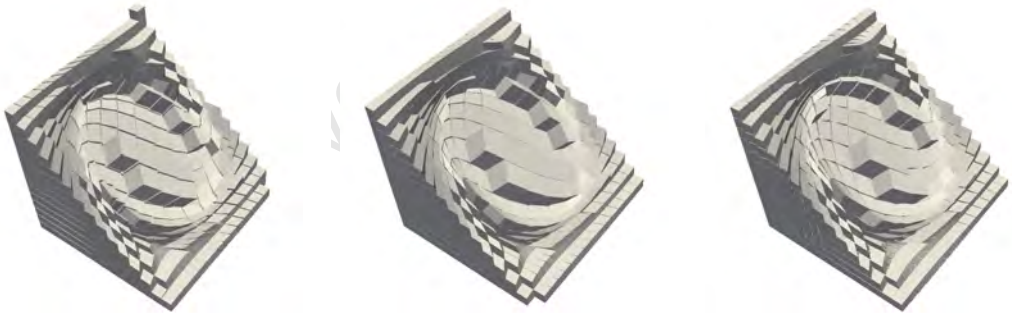


Figure 7.28: Cube: exact solution with $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements (b) With under-integration (c) Linear elements

Figure 7.29: Cube with NIPG, $\nu = 0.49995$

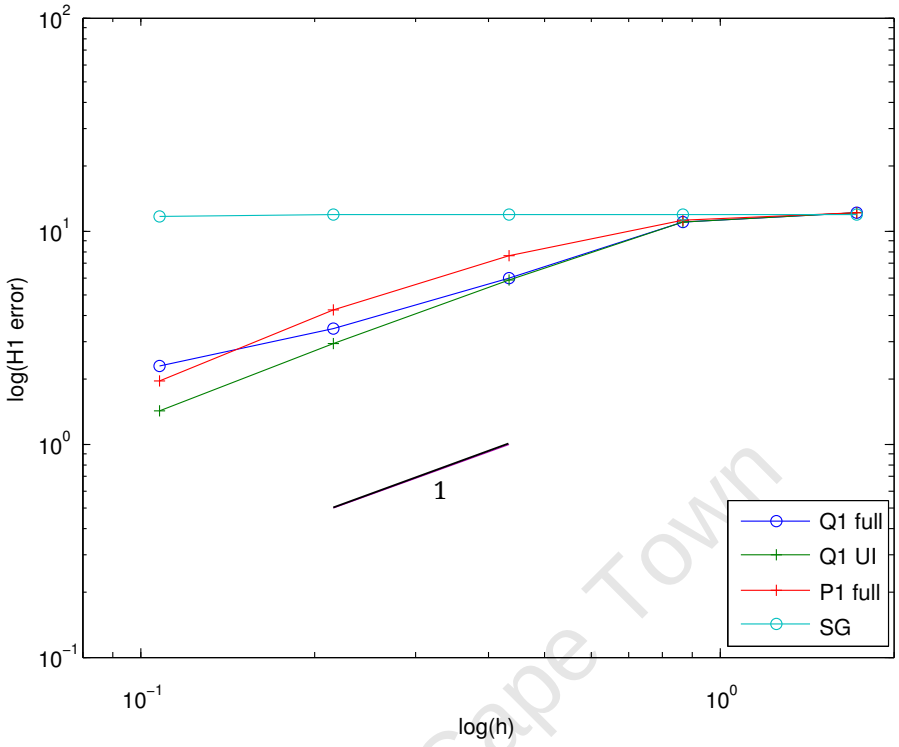
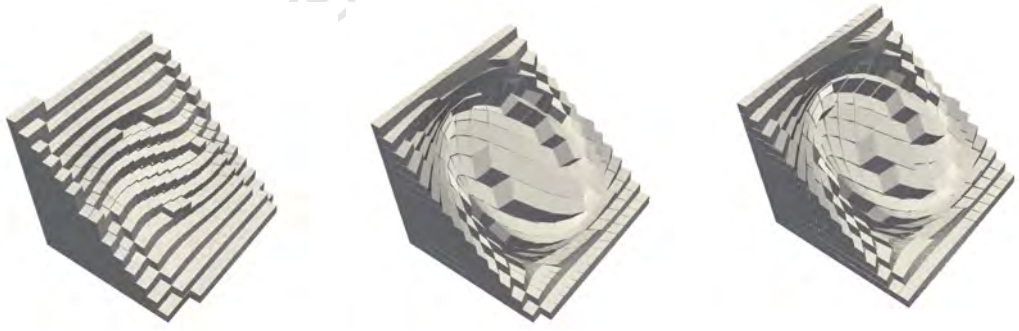


Figure 7.30: Comparison of H^1 errors for NIPG for the cube, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements

(b) With under-integration

(c) Linear elements with under-integration

Figure 7.31: Cube with IIPG, $\nu = 0.49995$

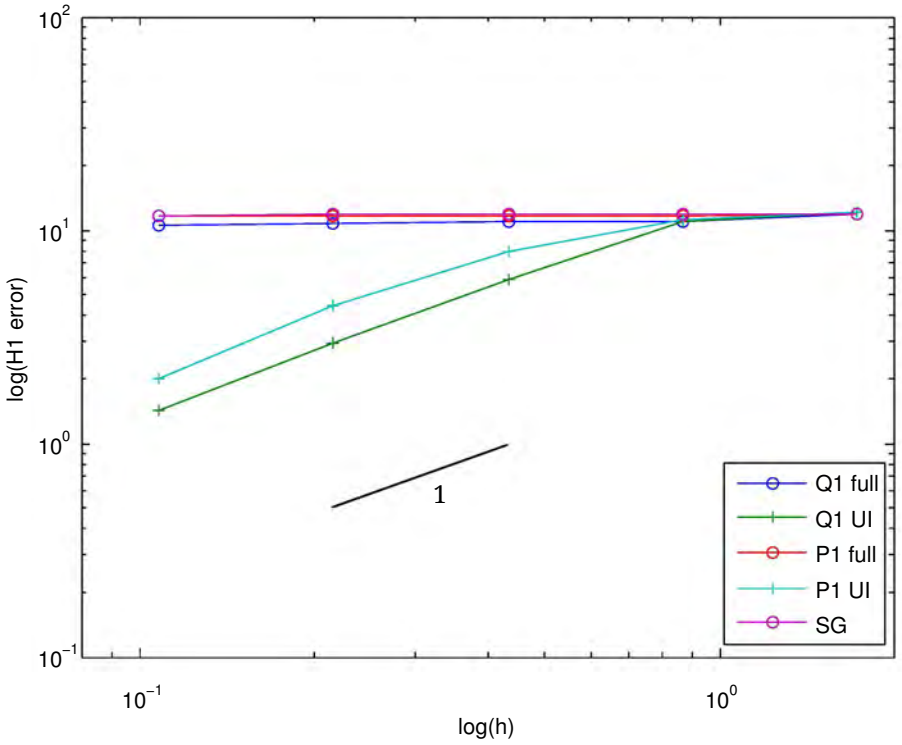
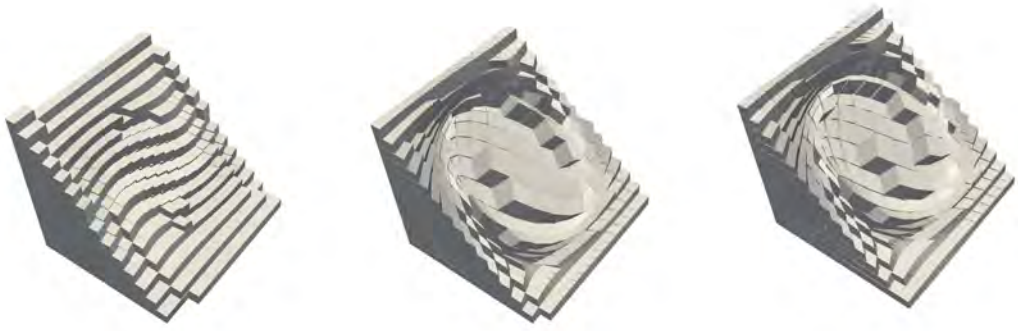


Figure 7.32: Comparison of H^1 errors for IIPG for the cube, $\nu = 0.49995$



(a) $\mathbb{Q}_1$ elements

(b) With under-integration

(c) Linear elements with under-integration

Figure 7.33: Cube with SIPG, $\nu = 0.49995$

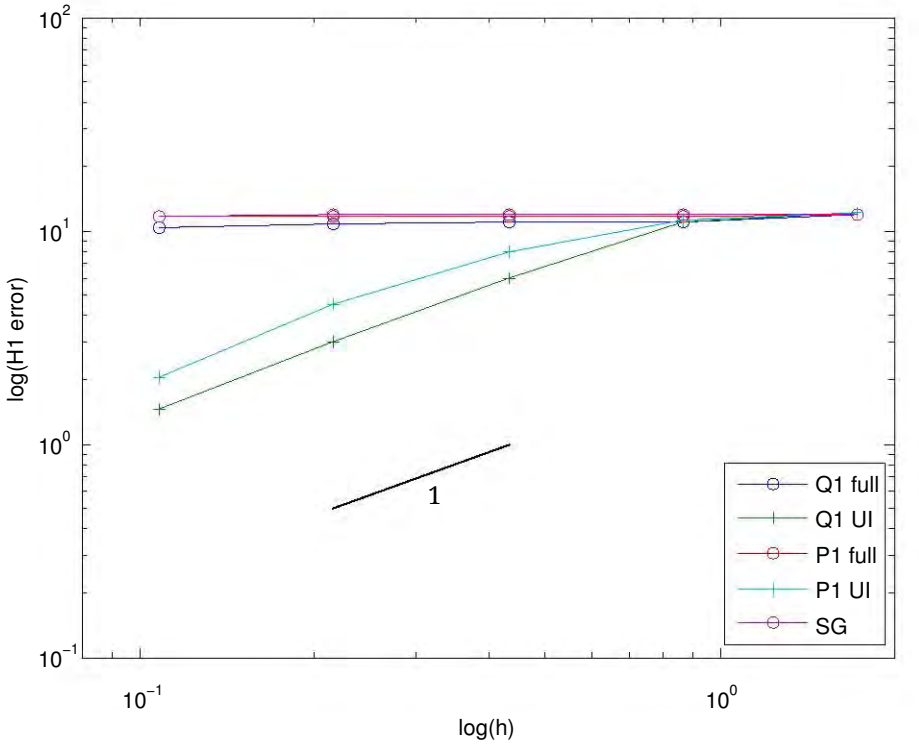


Figure 7.34: Comparison of H^1 errors for SIPG for the cube, $\nu = 0.49995$

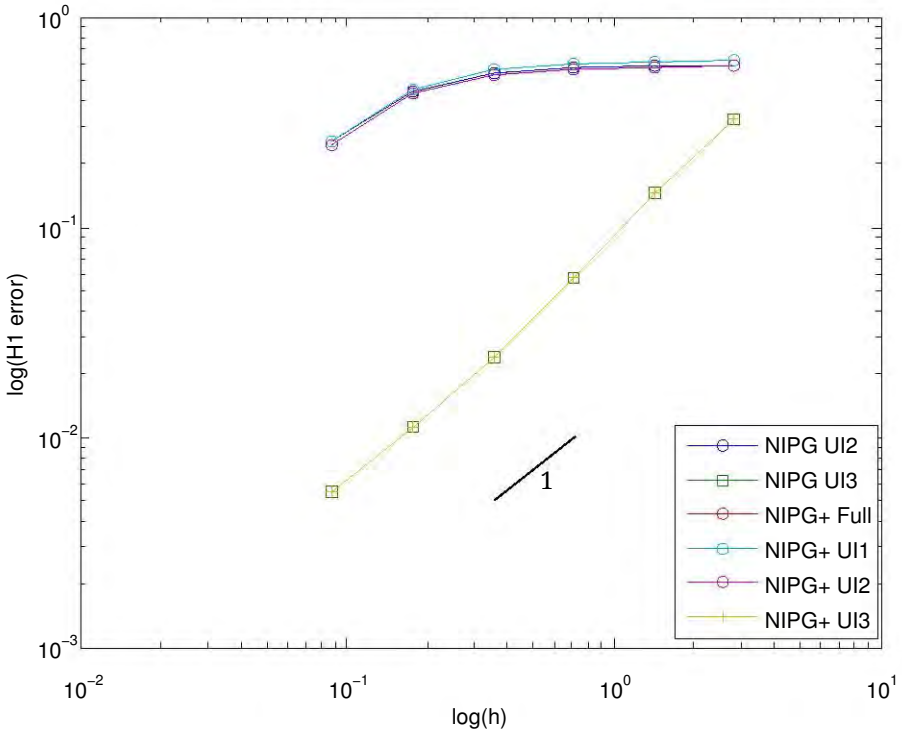
7.5 Other cases of under-integration

To illustrate the conclusions of §5.5, in which various combinations of under-integrated terms are considered with a focus on the effects on stability, results for four boundary value problems (CB, SP, TB and the cube) with nearly incompressible materials are presented here. These include results for the NIPG method with $k_\lambda > 0$, that is, with superfluous additional stabilization, referred to as before as NIPG+. (Results with sufficient under-integration are included in the error plots for benchmark comparison.)

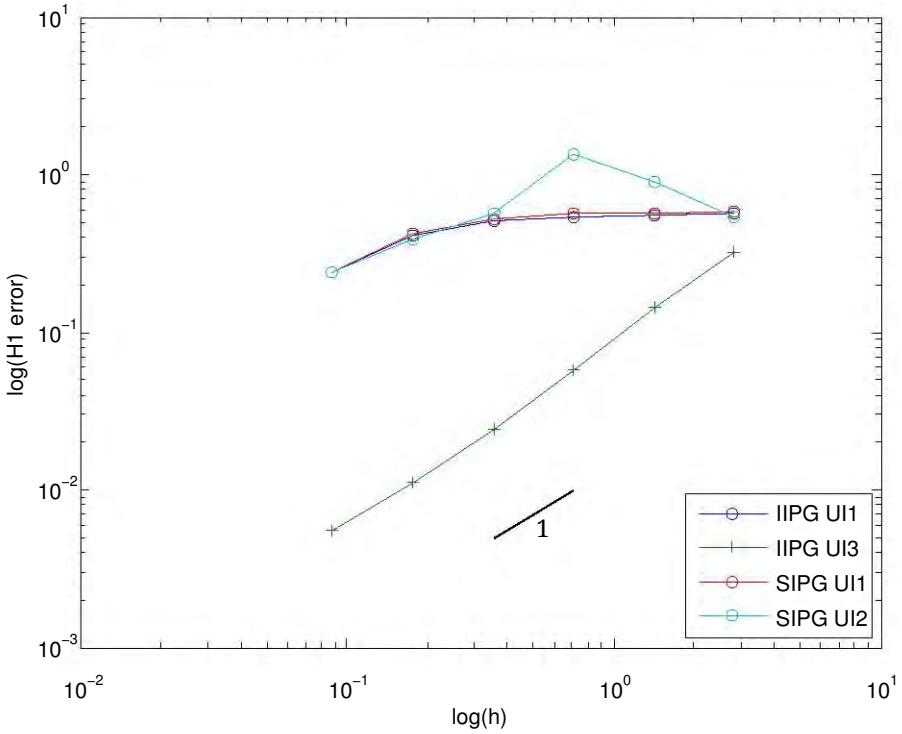
The summary in Tables 5.1 and 5.2 is useful for comparing the analysis aspects to the phenomena arising in these results.

The error plots in Figure 7.35 show the behaviour produced by the three IP methods with bilinear elements for the beam problem. Both the NIPG and the NIPG+ methods show poor convergence unless all λ-dependent terms are under-integrated (UI3), as expected from the analysis, although no obvious instability appears for the cases in which coercivity of the bilinear form is not proven (UI1, UI2). The IIPG and SIPG methods with under-integration of the stabilization term only (UI1) show large error and poor convergence while SIPG shows specific instability for the coarse meshes when two terms are under-integrated (UI2).

The example of the square plate shows similar convergence behaviour to that of the beam when bilinear elements are used, for most of the methods (Figure 7.36). Again, in the NIPG+ method, there is poor convergence for all combinations except where all three λ-dependent terms are under-integrated (UI3), where there are optimal results. However, the NIPG method without theoretically sufficient under-integration (UI2) performs as well as that with sufficient under-integration (UI3). The theory indicates possible lack of coercivity for UI2, although does not prove that it is always absent, which could explain the unexpectedly good performance of this case. The results for IIPG and SIPG are poor where the theory indicates likely insufficient under-integration, with overt divergence for SIPG UI2. Figure 7.39b shows a corresponding deformed shape for SIPG UI2, with an instability different from the usual locking that very low convergence rates commonly represent. An example of this locking is produced by SIPG UI1, and is shown in Figure 7.39a.

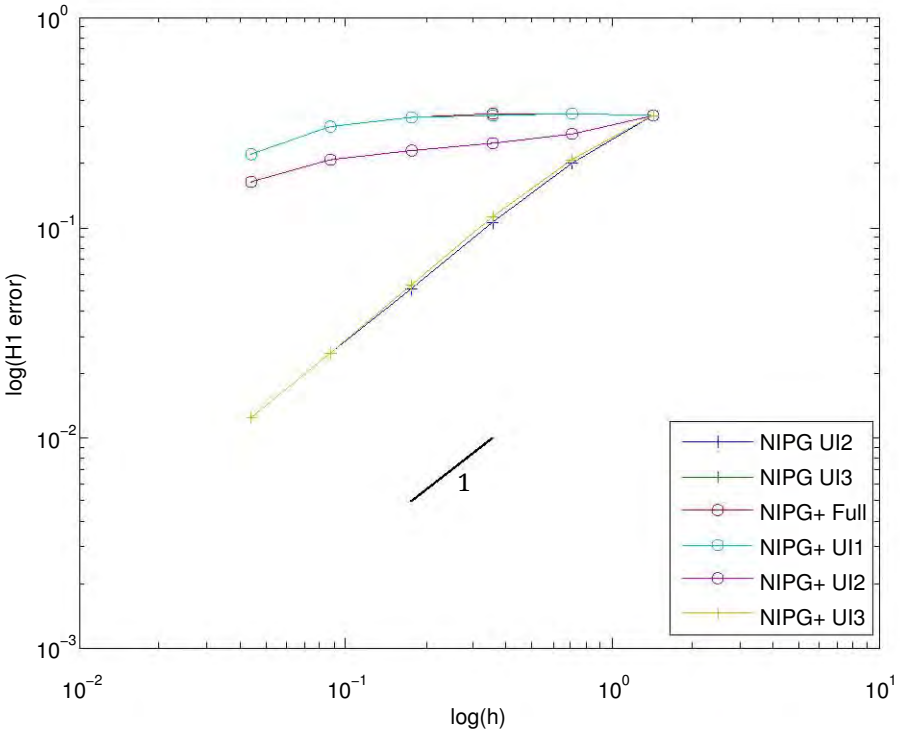


(a) NIPG and NIPG+

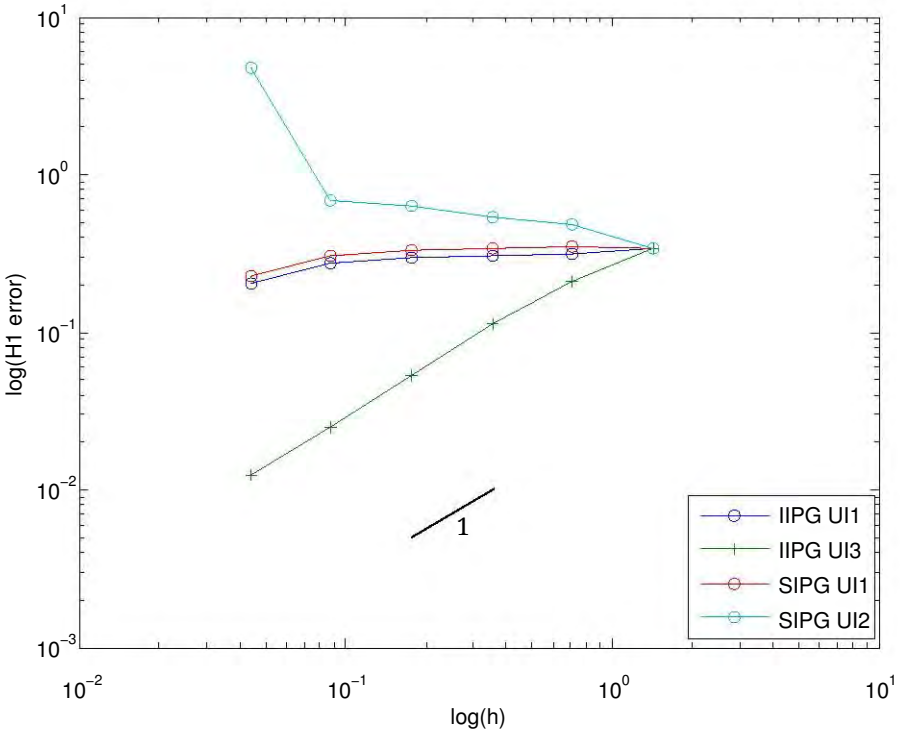


(b) IIPG and SIPG

Figure 7.35: Comparison of H^1 errors for various under-integration combinations with bilinear elements for the cantilever beam, $\nu = 0.49995$

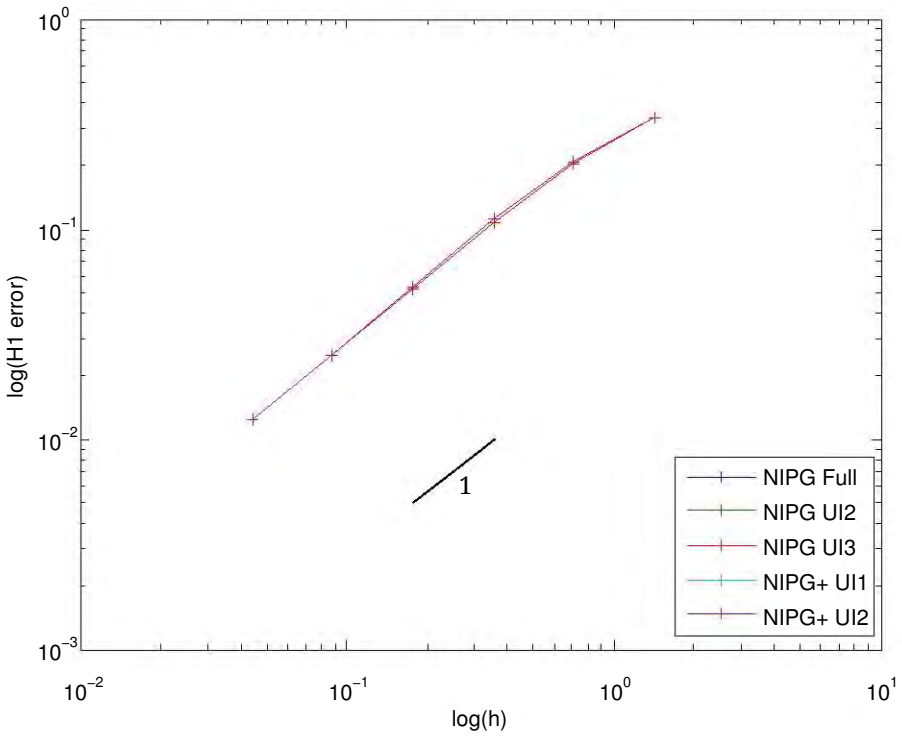


(a) NIPG and NIPG+

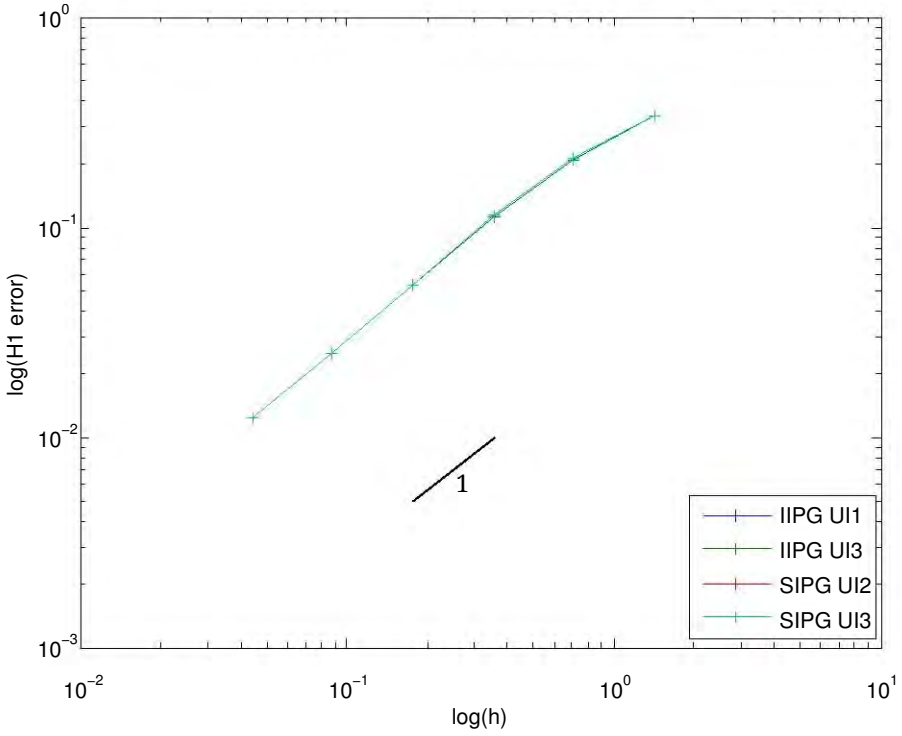


(b) IIPG and SIPG

Figure 7.36: Comparison of H^1 errors for various under-integration combinations with bilinear elements for the square plate, $\nu = 0.49995$

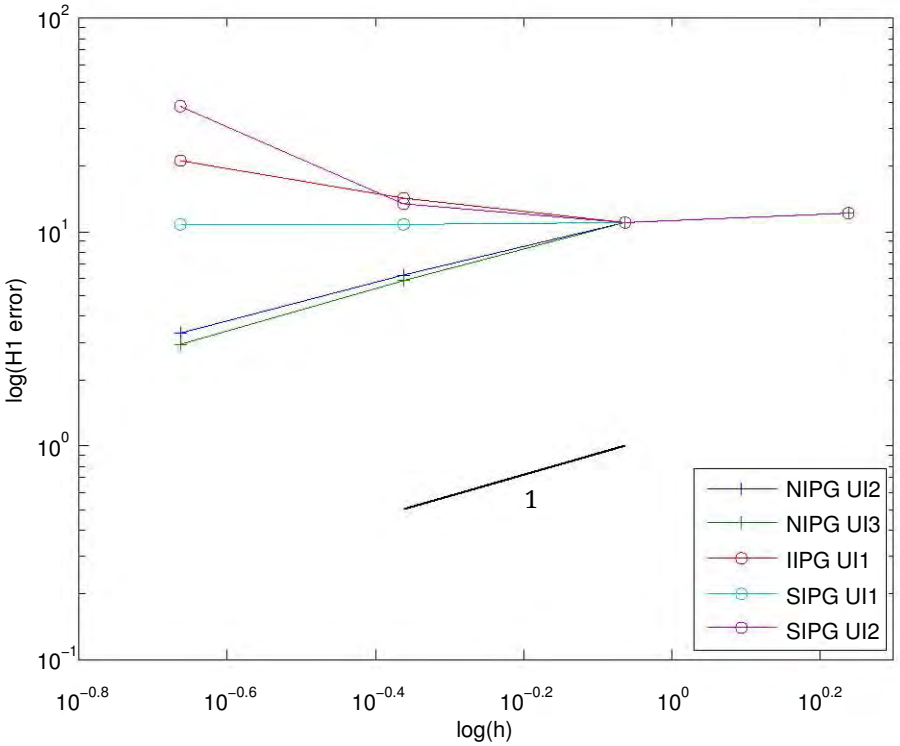


(a) NIPG and NIPG+

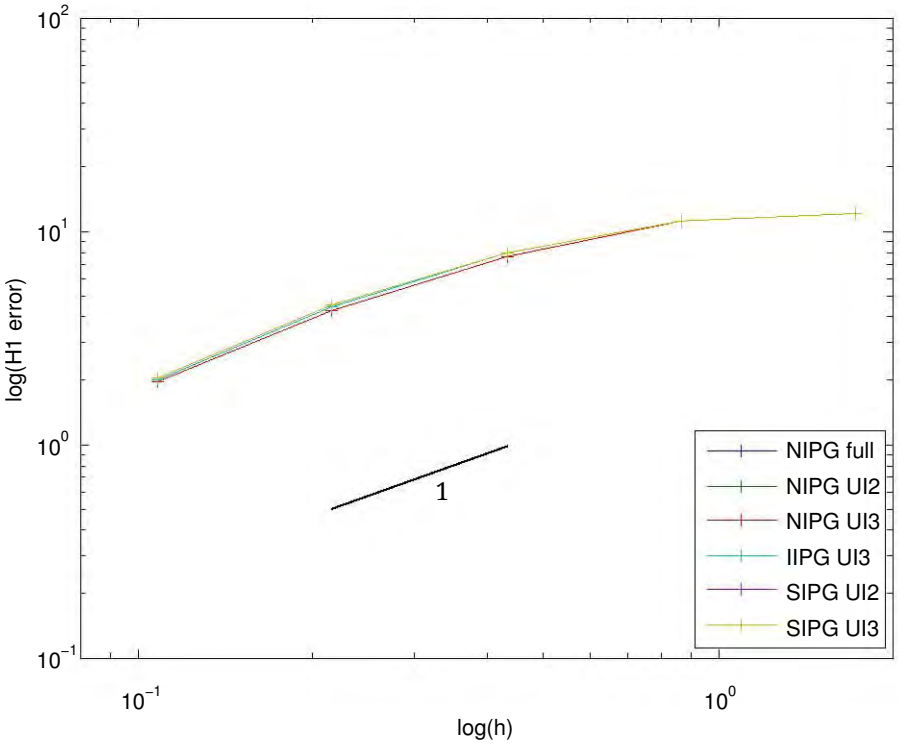


(b) IIPG and SIPG

Figure 7.37: Comparison of H^1 errors for various under-integration combinations with linear elements for the square plate, $\nu = 0.49995$

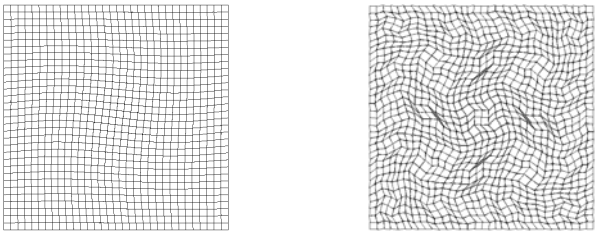


(a) Trilinear elements



(b) Linear elements

Figure 7.38: Comparison of H^1 errors for various under-integration combinations for the cube, $\nu = 0.49995$



(a) UI1 (b) UI2

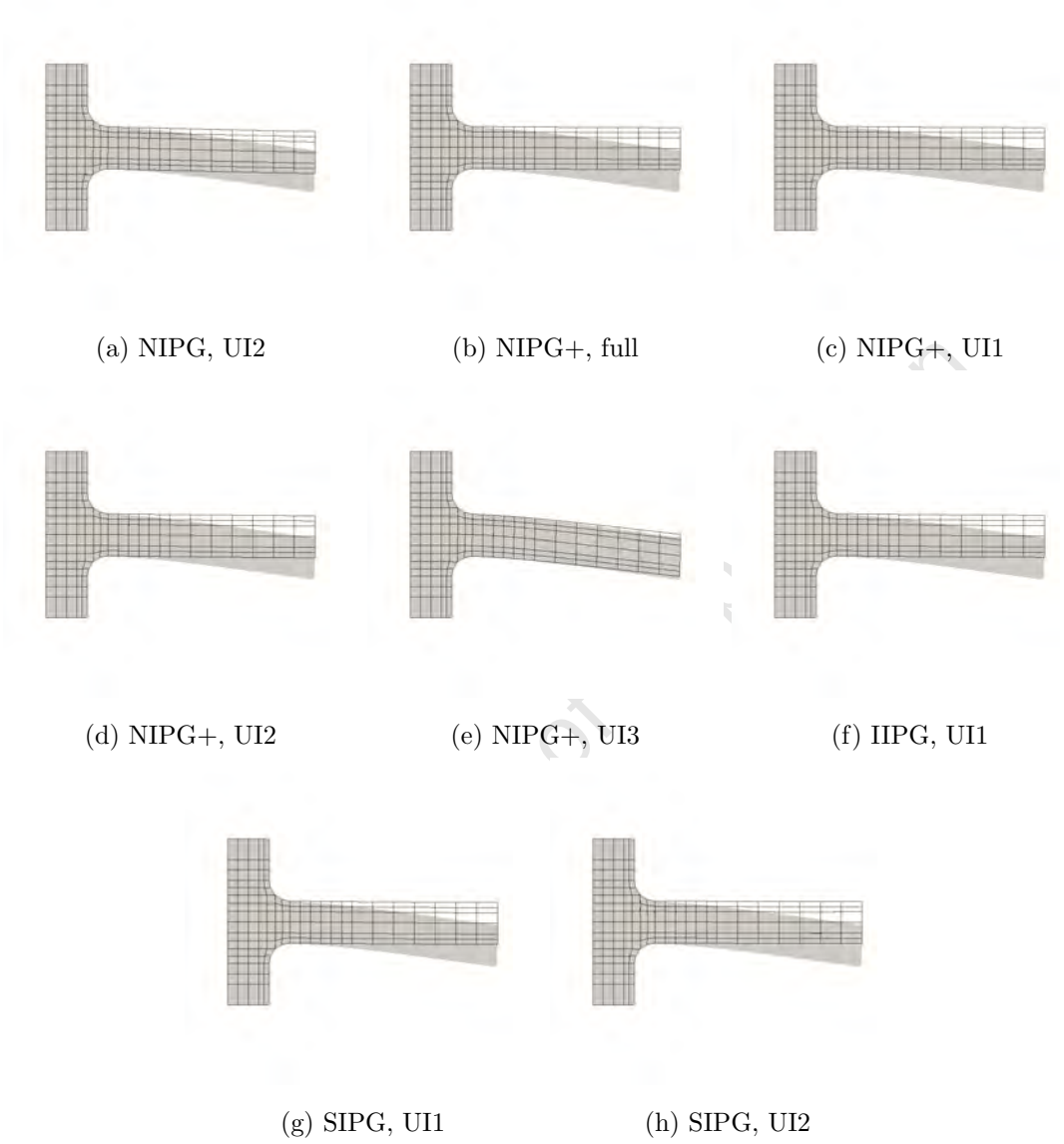
Figure 7.39: Examples of poor approximations: SIPG with bilinear elements without sufficient under-integration, for the square plate, $\nu = 0.49995$

When linear elements are used (Figure 7.37), however, the convergence rates are optimal for all of the methods that have at least the stabilization term under-integrated, as predicted by the theory.

For the cube subjected to a trigonometric body force, the results for trilinear elements with under-integration (Figure 7.38a) again show an approximation better than expected for the NIPG method where the theory indicates that there may be a lack of coercivity (UI2). The IIPG and SIPG methods produce poor convergence or overt divergence where coercivity is not proven. With linear elements (Figure 7.38b), all the convergence rates considered are optimal, as predicted.

Deformed shapes for the T-shaped bracket (Figures 7.40 and 7.41) show poor results, specifically locking, where the theory does not predict uniform convergence, generally for bilinear elements, and accurate deformations where they are expected according to the analysis, generally for linear elements. The exceptions are, as expected, NIPG+ with UI3 for bilinear elements (Figure 7.40e), and NIPG+ with full-order integration for linear elements (Figure 7.41c).

In general, the numerical results for these variations of under-integration follow the predictions of the analysis. Where there is apparent discrepancy, it can be explained by the fact that the theory provides sufficient conditions for stability rather than absolutely specifying necessary ones. In most cases, the poor approximations predicted manifest

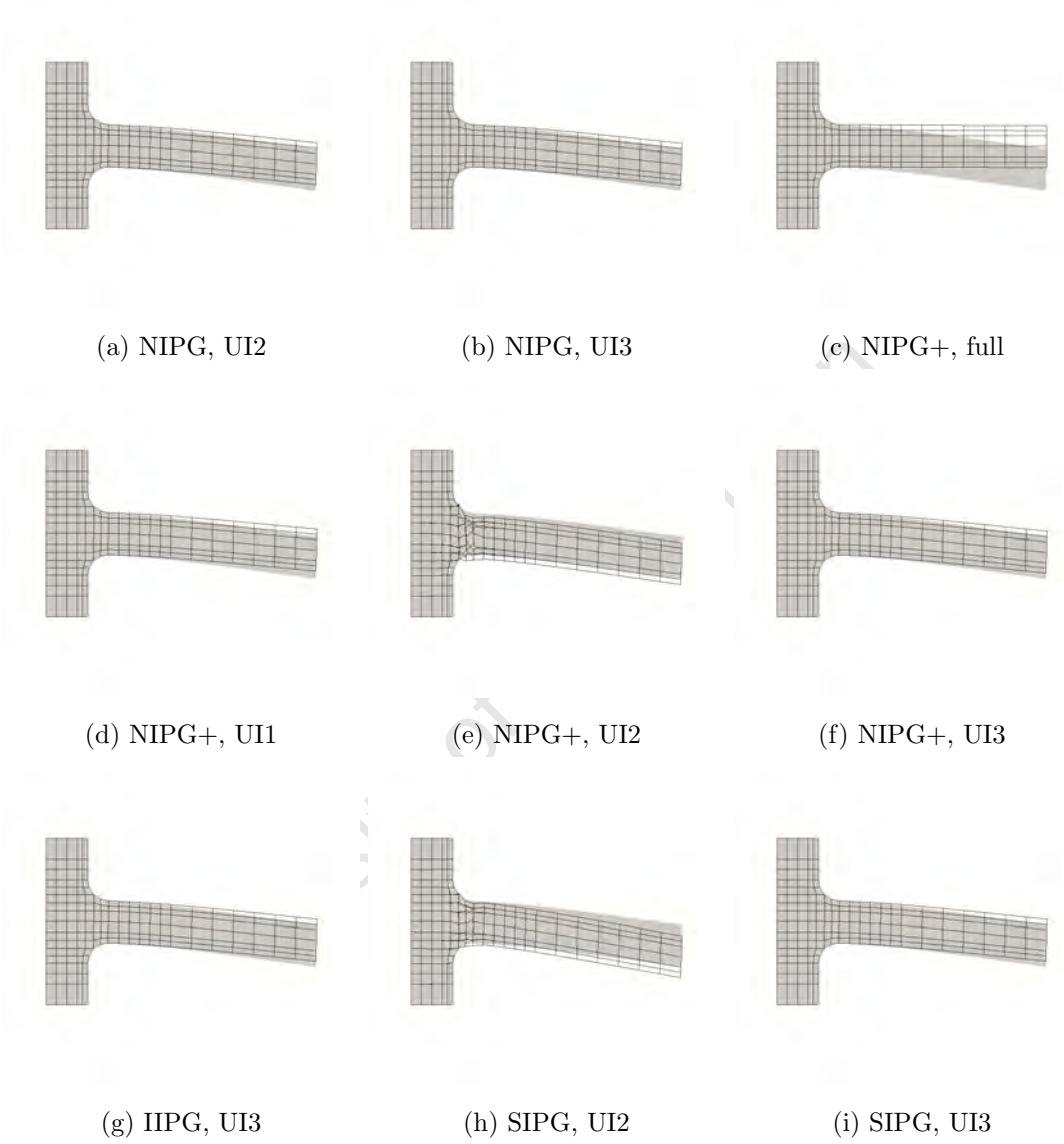


(a) NIPG, UI2 (b) NIPG+, full (c) NIPG+, UI1

(d) NIPG+, UI2 (e) NIPG+, UI3 (f) IIPG, UI1

(g) SIPG, UI1 (h) SIPG, UI2

Figure 7.40: T-shaped bracket, bilinear elements with various under-integration combinations, $\nu = 0.49995$



(a) NIPG, UI2 (b) NIPG, UI3 (c) NIPG+, full

(d) NIPG+, UI1 (e) NIPG+, UI2 (f) NIPG+, UI3

(g) IIPG, UI3 (h) SIPG, UI2 (i) SIPG, UI3

Figure 7.41: T-shaped bracket, linear elements with various under-integration combinations, $\nu = 0.49995$

in the form of locking, with extremely low rates of convergence. However, instability is sometimes visible in the form of divergent error plots and unusual deformation. In every case in which the theory predicts optimal convergence, this is supported by the results of the numerical experiments.

7.6 Meshes of non-rectangular elements

Central to the analyses of Chapter 5 is the interpolant defined in Chapter 4, which has properties dependent on the rectangular nature of the elements in the mesh. The implication of this is that the analyses hold for rectangular elements, but not necessarily for more general quadrilaterals (and similarly for the analogous situation in three dimensions, as described in Chapter 6). The results shown in the previous sections (§7.4 and §7.5) are for meshes consistently almost entirely of rectangular elements (or 3-rectangles, in three dimensions), and accurately illustrate the conclusions of the analyses.

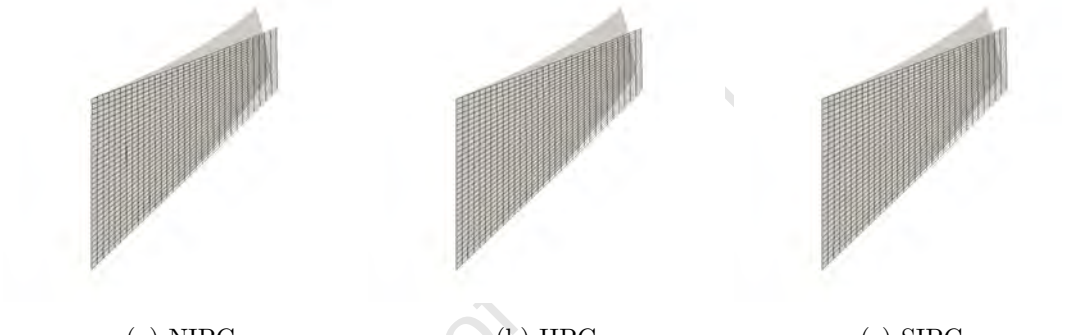
In order to ascertain the potential effectiveness of the remedies over a broader range of meshes, including those not within the scope of the analyses performed, numerical experiments have been performed using Cook's membrane as a test case, with $\nu = 0.49995$. Results from all three IP methods using the two proposed remedies (as described in Theorems 1, 2 and 3) are presented, with the wire mesh depicting the approximation produced by the IP method, and the solid colour depicting the accurate solution, in each case. Results for various other combinations of under-integration with linear elements are also presented.

Figure 7.42, depicting the results of edge-term under-integration with bilinear elements, indicates by the accuracy of the deformation approximations that this remedy is effective in overcoming the locking problem. While one test case is not sufficient for drawing bold conclusions in the absence of accompanying analysis, these results suggest that edge-term under-integration will be a broadly useful remedy applying to general quadrilaterals.

However, the results for linear elements do not show similar success in overcoming the locking problem, in this general case. While for rectangular elements IIPG and SIPG are still expected to lock when no under-integration is applied, NIPG has been seen to produce optimal results. In contrast, all three IP methods manifest locking with linear

(a) NIPG (b) IIPG (c) SIPG

Figure 7.42: Cook's membrane, $\mathbb{Q}_1$ elements with under-integration, $\nu = 0.49995$



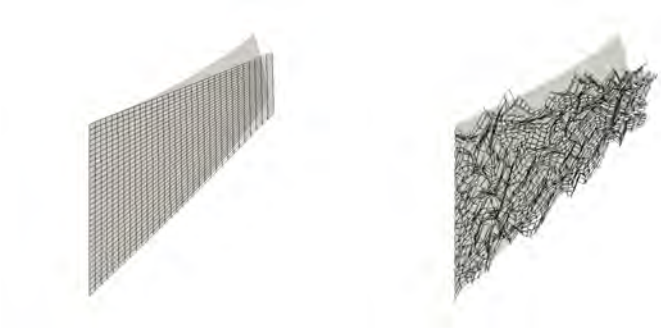
(a) NIPG (b) IIPG (c) SIPG

Figure 7.43: Cook's membrane, $\mathbb{P}_1$ elements, $\nu = 0.49995$

elements (Figure 7.43), and under-integration of the stabilization term, a remedy for IIPG and SIPG with rectangular elements, has no apparent effect on the IIPG method, while it produces instability for SIPG (Figure 7.44).

Simple preliminary experiments with other combinations of edge-term under-integration have not yielded results suggesting a potential direction for modification. For example, under-integration of all the λ-dependent edge terms (Figure 7.45) produces standard locking behaviour again.

These numerical investigations indicate that extension of the use of linear elements to general quadrilaterals is not straight-forward. An accompanying analysis would be necessary to isolate the cause of the λ-dependence and suggest a modification to remove



(a) IIPG (b) SIPG

Figure 7.44: Cook's membrane, $\mathbb{P}_1$ elements with under-integration (UI1), $\nu = 0.49995$

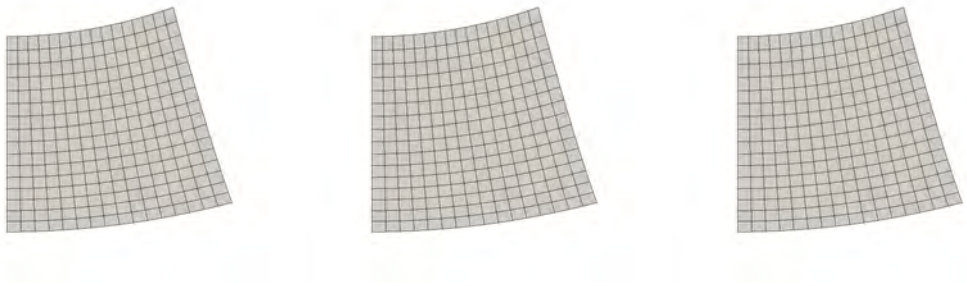
(a) NIPG (b) IIPG (c) SIPG

Figure 7.45: Cook's membrane, $\mathbb{P}_1$ elements with under-integration (UI3), $\nu = 0.49995$

it, if possible. Specifically, the identification or construction of an interpolant that has properties applicable to the general case would be necessary. Similarly, while the test case shown suggests that under-integration with bilinear elements will be broadly effective, a general analysis (and specifically a generally-applicable interpolant) would be necessary to establish the robustness of this remedy in the extended case.

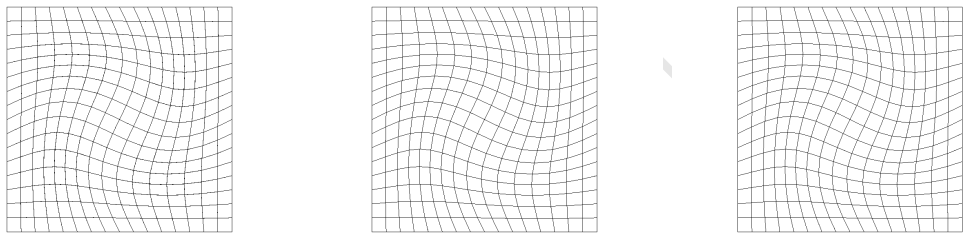
7.7 Higher-order elements

Higher-order elements are known to circumvent locking when used with the Standard Galerkin method, both on triangular and quadrilateral elements ($\mathbb{P}_k$ and $\mathbb{Q}_k$ respectively,



(a) NIPG (b) IIPG (c) SIPG

Figure 7.46: Cantilever beam with biquadratic elements, $\nu = 0.49995$



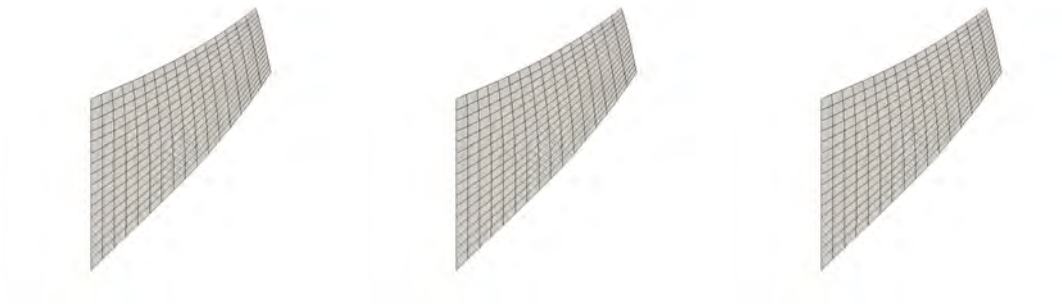
(a) NIPG (b) IIPG (c) SIPG

Figure 7.47: Square plate with biquadratic elements, $\nu = 0.49995$

with $k \geq 2$). In the realm of DG methods, particularly the IP methods, Hansbo and Larson [24] have shown that the analysis establishing the SIPG method as locking-free for low-order ($\mathbb{P}_1$) elements on triangles holds for higher-order elements: that is, $\mathbb{P}_k$ elements produce optimal convergence on triangles.

Here, results are presented from numerical experiments which are a preliminary investigation into the effectiveness of higher-order elements with the IP methods on meshes of quadrilaterals, with no modification to the methods.

Figures 7.46, 7.47 and 7.48 show that for all three IP methods, the use of $\mathbb{Q}_2$ elements produces accurate approximations with no signs of locking.



(a) NIPG (b) IIPG (c) SIPG

Figure 7.48: Cook's membrane with biquadratic elements, $\nu = 0.49995$

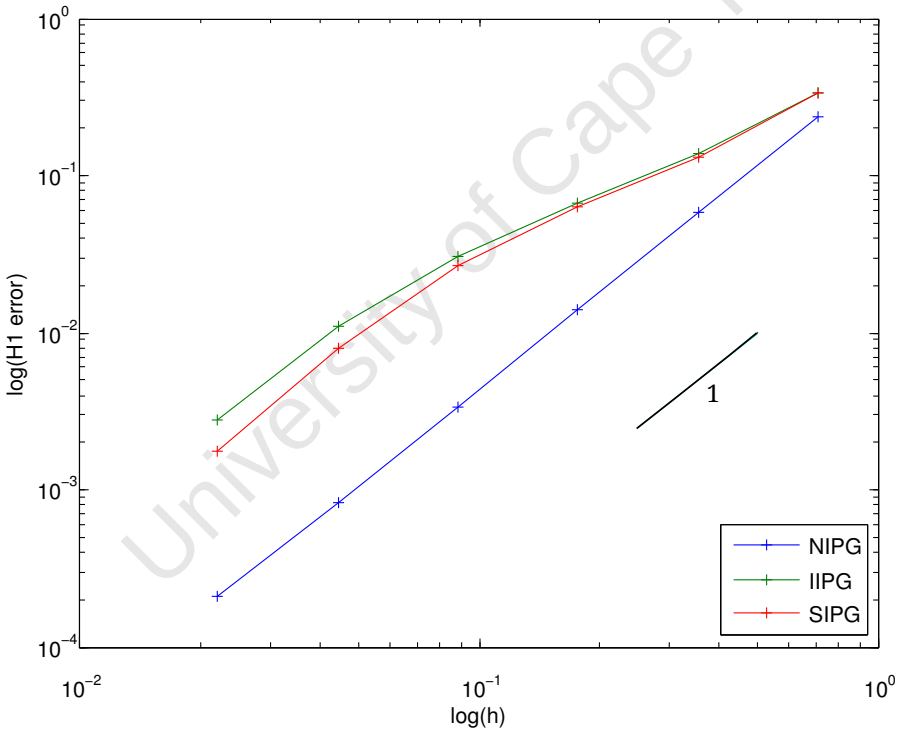


Figure 7.49: Convergence of H^1 error for the square plate, using $\mathbb{Q}_2$ elements, $\nu = 0.49995$

The meshes of the results displayed are each of 16 x 16 elements, one refinement level below that of the results displayed in previous sections for the low-order elements. For comparison, for every eighteen degrees of freedom at any particular refinement level for $\mathbb{Q}_2$ elements, there would be thirty-two degrees of freedom if $\mathbb{Q}_1$ elements were to be used at one refinement level more, or twenty-four if $\mathbb{P}_1$ elements were to be used at one refinement level more. The results are therefore comparable, and the accuracy of the results can be ascribed to the element type rather than the refinement level.

Convergence rates for the square plate problem are shown in Figure 7.49, and are seen to be optimal for all three IP methods. (The deformed shapes shown correspond to the fourth data point.)

These results indicate, in the absence of a full convergence analysis, that it is probable that higher-order quadrilateral elements with the IP methods can be used to circumvent locking, and that no modification to the methods, as in the case of lower-order elements, is necessary.

Chapter 8

Conclusions

Following the observation that the IP discontinuous Galerkin methods, widely acknowledged in the literature as locking-free, have until now been analysed only for meshes of triangles, substantial numerical evidence has been provided in this thesis to show that these results do not carry over to the use of bilinear quadrilateral elements. Locking and other forms of poor approximations have been shown to manifest when multilinear elements are used for nearly incompressible materials. This evidence is accompanied by an analytical investigation that considers an extension of a convergence analysis used for triangles to one for quadrilaterals, and shows that the discrepancy between triangle and quadrilateral results is reasonable.

A formulation incorporating specific edge-term under-integration, as outlined in Theorem 1, has been developed in this work as an effective remedy for the poor approximations produced with multilinear elements, for all three IP methods, in two and three dimensions. A convergence analysis has been performed, proving that the resulting methods are locking-free on rectangular multilinear elements, and numerical results have been presented, illustrating the outcomes of this analysis and the effectiveness of the remedy. Furthermore, the results of preliminary numerical tests suggest that the locking-free nature of the modified formulation extends to the non-affine case.

A second remedy, the use of linear ($\mathbb{P}_1$) approximations on quadrilateral elements, has been proposed for the NIPG method; and a similar solution for IIPG and SIPG, viz. linear elements in conjunction with minimal under-integration, has been developed. The

analysis presented, accompanied by supporting numerical results, proves that these give locking-free methods on meshes of rectangular elements, in two and three dimensions (Theorems 2 and 3).

The development of both remedies has involved investigating how to deal with terms contributing to undesirable λ-dependence in the error bound, establishing what is needed for the coercivity of the bilinear form, for each IP method, and considering the consistency of the modified formulations. Analytical observations and supporting numerical results are given for a range of cases of under-integration, providing insight into the aspects of the theory that are key in obtaining methods with uniform convergence.

For the more abstract mathematics of the convergence analysis, an alternative to the usual method for error-splitting has been used. Here, the approximation error is split with the linear portion of an interpolant, that is, a projection of an interpolant onto the space of linear polynomials, rather than with the interpolant itself. The advantage of this approach is that the resulting error-splitting function inherits critical properties from the underlying interpolant, but itself lies in a function space small enough to be useful in the convergence analysis.

A new interpolant with specialised orthogonality properties on rectangular elements has been constructed for use in this proof, with the specifically useful property that the mean edge value of the normal component of the nonlinear terms vanishes. This is related to the more general property, also possessed by the nonconforming DSSY elements ([22]), that the basis functions have mean values equal to the midpoint values on each edge of an element. A new interpolant in three dimensions has been designed with analogous properties.

Future work should include, firstly, an extension of the analyses of Chapters 5 and 6 to general quadrilateral elements. While the first remedy, under-integration on multilinear elements, is expected to be effective in the general case, only a convergence analysis will guarantee its robustness. Numerical results indicate that the second remedy does not apply to problems with non-affine meshes, positioning it as a remedy with limited scope. An extension of the analysis could provide insight into a useful modification that might allow this method to be used for general problems.

Secondly, the results of the use of $\mathbb{Q}_2$ elements suggest that higher-order quadrilateral

elements circumvent locking, and an analysis would be useful in establishing this.

Finally, there is scope for the extension of the formulations designed and described here into the arenas of finite-strain elasticity, nonhomogenous media, and other applications that have as a special case the fundamental linear elasticity which is the context of this work.

Appendix A

Useful identities, bounds and other theorems

A.1 The "magic formula"

For a vector $\boldsymbol{\varphi}$ and a second-order tensor $\boldsymbol{\tau}$,

$$\sum_{\Omega_e \in \mathcal{T}_h} \int_{\partial\Omega_e} (\boldsymbol{\tau}\mathbf{n}) \cdot \boldsymbol{\varphi}\, ds \;= \sum_{E \in \Gamma} \int_E \{\!\{\boldsymbol{\tau}\}\!\} : [\![\boldsymbol{\varphi}]\!]\, ds \;+ \sum_{E \in \Gamma_{int}} \int_E [\![\boldsymbol{\tau}]\!] \cdot \{\!\{\boldsymbol{\varphi}\}\!\}\, ds. \tag{A.1.1}$$

A.2 Young's inequality

$$ab \leq \tfrac{1}{2}\left(\epsilon a^2 + \frac{1}{\epsilon} b^2\right), \tag{A.2.1}$$

where $\epsilon > 0$.

A.3 Hölder's inequality

For scalar-valued functions f and g, with constants p and q such that $0 \leq p, q, \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$,

$$\int |fg| \leq \|f\|_{L^p} \|g\|_{L^q} \tag{A.3.1}$$

and

$$\left| \sum_{k=1}^{n} \int_{\Omega_k} fg \right| \leq \left(\sum_{k=1}^{n} \|f\|^2_{L^p(\Omega_k)} \right)^{\frac{1}{2}} \left(\sum_{k=1}^{n} \|g\|^2_{L^q(\Omega_k)} \right)^{\frac{1}{2}}, \tag{A.3.2}$$

of particular use when $p = q = 2$.

Similar results hold for the scalar products of vectors or second-order tensors $\mathbf{f}$, $\mathbf{g}$.

A.4 Edge term manipulation

A.4.1 Relating jumps of functions

$$\mathbf{1} : \llbracket \mathbf{v} \rrbracket = \llbracket \mathbf{v} \rrbracket. \tag{A.4.1}$$

A.4.2 Trace inequalities

From the inverse inequalities given in [38]: for an edge $\hat{E}$ of a reference element $\hat{\Omega}$, for a polynomial v,

$$\|v\|_{L^2(\hat{E})} \leq C \|v\|_{L^2(\hat{\Omega})}. \tag{A.4.2}$$

By a scaling argument,

$$\|v\|_{L^2(E)} \leq C h_e^{-1/2} \|v\|_{L^2(\Omega_e)}. \tag{A.4.3}$$

For $v \in H^1(\Omega_e)$,

$$\|v\|_{L^2(\partial\Omega_e)} \leq C \left(h_e^{-1/2} \|v\|_{L^2(\Omega_e)} + h_e^{1/2} |v|_{H^1(\Omega_e)} \right) \tag{A.4.4}$$

(cf. [26]).

A.4.3 Bounds involving jumps and averages

$$\sum_{E\in\Gamma_{iD}} h_E \left\|\{\!\{\phi\}\!\}\right\|^2_{L^2(E)} \leq C \sum_{\Omega_e\in\mathcal{T}_h} h_e \left\|\phi\right\|^2_{L^2(\partial\Omega_e^{iD})} \tag{A.4.5}$$

$$\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|\{\!\{\phi\}\!\}\right\|^2_{L^2(E)} \leq \sum_{\Omega_e\in\mathcal{T}_h} \sum_{E\in\partial\Omega_e^{iD}} \frac{1}{h_E} \left\|\phi\right\|^2_{L^2(E)}, \tag{A.4.6}$$

$$\sum_{E\in\Gamma_{iD}} \frac{1}{h_E} \left\|[\![\mathbf{v}]\!]\right\|^2_{L^2(E)} \leq \sum_{\Omega_e\in\mathcal{T}_h} \sum_{E\in\partial\Omega_e^{iD}} \frac{2}{h_E} \left\|\mathbf{v}\right\|^2_{L^2(E)}. \tag{A.4.7}$$

A.5 Polynomial-preserving projections

By Theorem 3.1.4 of Ciarlet [15], for a bounded linear operator $\mathbf{P} : H^{k+1}(\Omega_e) \to H^m(\Omega_e)$ that preserves polynomials of order k, for integers $0 \leq m \leq k+1$,

$$|v - \mathbf{P}v|_{H^m(\Omega_e)} \leq C h_e^{k+1-m} |v|_{H^{k+1}(\Omega_e)}, \tag{A.5.1}$$

where shape-regularity is assumed and the Ω_e are affine-equivalent domains.

(The original result is more general but nevertheless restricted to affine-equivalent domains.)

A.6 Regularity

1. The regularity estimate for planar elasticity is summarised in Brenner and Scott [10] as follows:

   Assume $\mathbf{u} \in \left[H^2(\Omega)\right]^2$. When

   - $\partial\Omega$ is smooth and $\bar{\Gamma}_D \cap \bar{\Gamma}_N = \emptyset$,

   or when

- Ω is a convex polygon and either Γ_D or Γ_N is empty,

then there exists a $C > 0$ independent of λ such that

$$\|\mathbf{u}\|_{H^2(\Omega)} + \lambda \|\nabla \cdot \mathbf{u}\|_{H^1(\Omega)} \leq C \left(\|\mathbf{f}\|_{L^2(\Omega)} + \|\mathbf{g}_\Omega\|_{H^2(\Omega)} + \|\mathbf{h}_\Omega\|_{H^1(\Omega)} \right), \quad \text{(A.6.1)}$$

where $\mathbf{g} = \mathbf{g}_\Omega|_{\Gamma_D}$ and $\mathbf{h} = \mathbf{h}_\Omega|_{\Gamma_N}$, provided the data have sufficient regularity.

2. For polygonal domains in general (including non-convex domains), an analogous estimate is given in terms of weighted Sobolev spaces in [41].

3. No similar result (showing explicitly the role of λ in the estimate) is known to be available for three-dimensional domains.

Appendix B

Details relating to the new interpolants

B.1 The new interpolant in two dimensions (Chapter 4)

B.1.1 Basis functions

For the x-component of the interpolant:

$$\begin{aligned}
\phi_1^x(x,y) &= -\tfrac{1}{2}y + \frac{1}{4}\Theta \\
\phi_2^x(x,y) &= \tfrac{1}{2} + \tfrac{1}{2}x - \frac{1}{4}\Theta \\
\phi_3^x(x,y) &= \tfrac{1}{2}y + \frac{1}{4}\Theta \\
\phi_4^x(x,y) &= \tfrac{1}{2} - \tfrac{1}{2}x - \frac{1}{4}\Theta
\end{aligned}$$

For the y-component of the interpolant:

$$\begin{aligned}
\phi_1^y(x,y) &= \tfrac{1}{2} - \tfrac{1}{2}y - \frac{1}{4}\Psi \\
\phi_2^y(x,y) &= \tfrac{1}{2}x + \frac{1}{4}\Psi \\
\phi_3^y(x,y) &= \tfrac{1}{2} + \tfrac{1}{2}y - \frac{1}{4}\Psi \\
\phi_4^y(x,y) &= -\tfrac{1}{2}x + \frac{1}{4}\Psi
\end{aligned}$$

B.2 The new interpolant in three dimensions (Chapter 6)

B.2.1 Numbering of faces of reference element

The midpoint of each face is equal to the outward unit normal on the face, that is,

Face	Midpoint or normal
1	$(1,0,0)$
2	$(-1,0,0)$
3	$(0,1,0)$
4	$(0,-1,0)$
5	$(0,0,1)$
6	$(0,0,-1)$

B.2.2 Values at midpoints of element faces

In evaluating the functions L_1, L_2, M_1, M_2, N_1 and N_2 at the midpoints of each of the faces of the reference element, since the midpoints are combinations of 0, 1 and -1, and $\theta(0) = \kappa(0) = \theta(\pm 1) = 0$, the only nonzero results will be from $\kappa(\pm 1) = 2$. Thus at the midpoints of faces 1 and 2, M_2 and N_1 will equal 2; for faces 3 and 4, L_1 and N_2; and for faces 5 and 6, L_2 and M_1. The rest vanish.

B.2.3 Mean values on element faces

In integrating the higher-order basis functions of $\hat{Q}$ over the faces of the reference element, as $\theta(0) = \kappa(0) = \theta(\pm 1) = 0$, and $\int_{-1}^{1} \theta(x)\,dx = \int_{-1}^{1} \kappa(x)\,dx = 0$, nonzero values will arise only when $\kappa(\pm 1)$ appears explicitly. Thus the same functions will be nonzero when integrated on a given face as evaluate as nonzero at the midpoint of that face. In each case, the value of the integrated function will be

$$\kappa(\pm 1) \int_{-1}^{1} \int_{-1}^{1} dE = 4.\kappa(\pm 1).$$

Thus the mean value on the face is $\kappa(\pm 1) = 2$.

B.2.4 Basis functions

For the x-component of the interpolant:

$$\phi_1^x(x,y,z) = \frac{1}{4}\left(2 + 2x - L_1(x,y,z) - L_2(x,y,z)\right)$$
$$\phi_2^x(x,y,z) = \frac{1}{4}\left(2 - 2x - L_1(x,y,z) - L_2(x,y,z)\right)$$
$$\phi_3^x(x,y,z) = \frac{1}{4}\left(2y + L_1(x,y,z)\right)$$
$$\phi_4^x(x,y,z) = \frac{1}{4}\left(-2y + L_1(x,y,z)\right)$$
$$\phi_5^x(x,y,z) = \frac{1}{4}\left(2z + L_2(x,y,z)\right)$$
$$\phi_6^x(x,y,z) = \frac{1}{4}\left(-2z + L_2(x,y,z)\right)$$

For the y-component of the interpolant:

$$\phi_1^y(x,y,z) = \frac{1}{4}\left(2x + M_2(x,y,z)\right)$$
$$\phi_2^y(x,y,z) = \frac{1}{4}\left(-2x + M_2(x,y,z)\right)$$
$$\phi_3^y(x,y,z) = \frac{1}{4}\left(2 + 2y - M_1(x,y,z) - M_2(x,y,z)\right)$$
$$\phi_4^y(x,y,z) = \frac{1}{4}\left(2 - 2y - M_1(x,y,z) - M_2(x,y,z)\right)$$
$$\phi_5^y(x,y,z) = \frac{1}{4}\left(2z + M_1(x,y,z)\right)$$
$$\phi_6^y(x,y,z) = \frac{1}{4}\left(-2z + M_1(x,y,z)\right)$$

For the z-component of the interpolant:

$$\phi_1^z(x,y,z) = \frac{1}{4}\left(2x + N_1(x,y,z)\right)$$
$$\phi_2^z(x,y,z) = \frac{1}{4}\left(-2x + N_1(x,y,z)\right)$$
$$\phi_3^z(x,y,z) = \frac{1}{4}\left(2y + N_2(x,y,z)\right)$$
$$\phi_4^z(x,y,z) = \frac{1}{4}\left(-2y + N_2(x,y,z)\right)$$
$$\phi_5^z(x,y,z) = \frac{1}{4}\left(2 + 2z - N_1(x,y,z) - N_2(x,y,z)\right)$$
$$\phi_6^z(x,y,z) = \frac{1}{4}\left(2 - 2z - N_1(x,y,z) - N_2(x,y,z)\right)$$

Notation

$\mathbf{1}$	identity tensor
$\otimes$	tensor product
$:$	scalar product of second-order tensors
∇	gradient operator
$\nabla\cdot$	divergence operator
E	Young's modulus
ν	Poisson's ratio
μ, λ	Lamé parameters
$\mathbf{u}$	displacement
$\boldsymbol{\varepsilon}(\mathbf{u})$	symmetric gradient of $\mathbf{u}$, also strain
$\boldsymbol{\sigma}(\mathbf{u})$	$2\mu\boldsymbol{\varepsilon}(\mathbf{u}) + \lambda\nabla\cdot\mathbf{u}\mathbf{1}$, also stress
$\mathbf{f}$	body force
$\mathbf{g}$	prescribed displacement (Dirichlet boundary function)
$\mathbf{h}$	prescribed traction (Neumann boundary function)
$\mathbf{g}_\Omega$	function on Ω such that $\mathbf{g}_\Omega\|_{\Gamma_D} = \mathbf{g}$
$\mathbf{h}_\Omega$	function on Ω such that $\mathbf{h}_\Omega\|_{\Gamma_N} = \mathbf{h}$
$\mathbf{u}_h$	DG approximation of $\mathbf{u}$
$\mathbf{u}_P$	error-splitting function
d	dimension
Ω	domain
$\partial\Omega$	domain boundary, or set of edges/faces on domain boundary
Γ_D	Dirichlet boundary, or set of edges/faces on Dirichlet boundary
Γ_N	Neumann boundary, or set of edges/faces on Neumann boundary
Γ_{int}	interior boundary, or set of edges/faces on interior boundary

Γ_{iD}	$\Gamma_{int} \cup \Gamma_D$
Γ	$\Gamma_{int} \cup \partial\Omega$
$\mathcal{T}_h$	partition
Ω_e	element e domain
$\hat{\Omega}$	reference element
E	edge in 2D, face in 3D
$\partial\Omega_e$	boundary of Ω_e or set of edges/faces on boundary of Ω_e
$\partial\Omega_e^{int}$	$\partial\Omega_e \cap \Gamma_{int}$
$\partial\Omega_e^{D}$	$\partial\Omega_e \cap \Gamma_D$
$\partial\Omega_e^{iD}$	$\partial\Omega_e \cap \Gamma_{iD}$
Γ_{ef}	$\partial\Omega_e \cap \partial\Omega_f$, that is, edge/face shared by elements Ω_e and Ω_f
$\mathbf{n}$	outward unit normal
$\mathbf{n}_e$	outward unit normal to element Ω_e
$\mathbf{m}_E$	midpoint of edge/face E
$\mathbf{m}_i$	midpoint of specific edge/face E^i
h_e	$\text{diam}(\Omega_e)$
h_E	$\text{diam}(E)$
h	maximum h_e in partition
$\{\!\{\mathbf{v}\}\!\}$	$\frac{1}{2}(\mathbf{v}_e + \mathbf{v}_f)$ on interior edges/faces $E = \Gamma_{ef}$, or $\mathbf{v}$ on edges/faces $E \in \partial\Omega$, for a vector $\mathbf{v}$
$\{\!\{\boldsymbol{\tau}\}\!\}$	$\frac{1}{2}(\boldsymbol{\tau}_e + \boldsymbol{\tau}_f)$ on interior edges/faces $E = \Gamma_{ef}$, or $\boldsymbol{\tau}$ on edges/faces $E \in \partial\Omega$, for a second-order tensor $\boldsymbol{\tau}$
$\llbracket\mathbf{v}\rrbracket$	$\mathbf{v}_e \otimes \mathbf{n}_e + \mathbf{v}_f \otimes \mathbf{n}_f$ on interior edges/faces $E = \Gamma_{ef}$, or $\mathbf{v} \otimes \mathbf{n}$ on edges/faces $E \in \partial\Omega$, for a vector $\mathbf{v}$
$[\![\mathbf{v}]\!]$	$\mathbf{v}_e \cdot \mathbf{n}_e + \mathbf{v}_f \cdot \mathbf{n}_f$ on interior edges/faces $E = \Gamma_{ef}$, or $\mathbf{v} \cdot \mathbf{n}$ on edges/faces $E \in \partial\Omega$, for a vector $\mathbf{v}$
$\llbracket\boldsymbol{\tau}\rrbracket$	$\boldsymbol{\tau}_e\mathbf{n}_e + \boldsymbol{\tau}_f\mathbf{n}_f$ on interior edges/faces $E = \Gamma_{ef}$, or $\boldsymbol{\tau}\mathbf{n}$ on edges/faces $E \in \partial\Omega$, for a second-order tensor $\boldsymbol{\tau}$
$\mathbb{P}_k(\Omega)$	space of polynomials on Ω with maximum total degree k
$\mathbb{Q}_k(\Omega)$	space of polynomials on Ω with maximum degree k in each variable
$\mathbb{V}$	$\mathbb{P}_1(\Omega_e)$ or $\mathbb{Q}_1(\Omega_e)$
V_h	DG approximation space
$H^m(\Omega)$	standard Sobolev space on Ω
$H^1(\mathcal{T}_h)$	broken H^1 space on the partition
$H_0^1(\Omega)$	space of functions in $H^1(\Omega)$ vanishing on $\partial\Omega$
$H^1_{\Gamma_D^x}(\Omega)$	space of functions in $H^1(\Omega)$ equal to g_x on Γ_D^x (also $H^1_{\Gamma_D^y}(\Omega)$, also $H^1_{\Gamma_D^z}(\Omega)$)
$\hat{Q}$	interpolation space on the reference element
Q	interpolation space on the real element (also Q^x, Q^y, Q^z)
NC_h	global interpolation space (also NC_h^x, NC_h^y, NC_h^z)

$\|\cdot\|_{H^m(\Omega)}$	Sobolev norm
$\|\cdot\|_{H^m(\Omega)}$	Sobolev seminorm
$\|\cdot\|_{\mathrm{DG}}$	DG norm
ϕ_i	basis function (also ϕ_i^x, ϕ_i^y, ϕ_i^z)
Θ, Ψ	higher-order functions in the 2D interpolation space
L_i, M_i, N_i	higher-order functions in the 3D interpolation space
$\theta(\cdot)$, $\kappa(\cdot)$	functions contributing towards interpolant basis functions
π	global projection (also π^x, π^y)
π_e	projection on Ω_e (also π_e^x, π_e^y)
Π	direct projection onto $\mathbb{P}_1(\Omega_e)$
Π_0	L^2-orthogonal projection onto $\mathbb{P}_0(E)$
Π_C	L^2-orthogonal projection onto $\mathbb{P}_0(\Omega_e)$
$\bar{\Pi}$	L^2-orthogonal projection onto $\mathbb{P}_1\left(\hat{\Omega}\right)$
$a(\cdot,\ \cdot)$	bilinear form
$a_h(\cdot,\ \cdot)$	bilinear form for the discrete problem
$a_h^{\mathrm{UI}}(\cdot,\ \cdot)$	bilinear form for the discrete problem with under-integration
$l(\cdot)$	linear form
$l_h(\cdot)$	linear form for the discrete problem
$l_h^{\mathrm{UI}}(\cdot)$	linear form for the discrete problem with under-integration
θ	switch parameter for distinguishing between IP methods
k_μ, k_λ	stabilization parameters
T_μ, T_λ	portions of the bilinear form dependent on μ and λ respectively
$\boldsymbol{e}$	approximation error
$\boldsymbol{\eta}$	interpolation error $\mathbf{u} - \pi\mathbf{u}$
$\boldsymbol{\xi}$	$\pi\mathbf{u} - \mathbf{u}_h$
$\boldsymbol{\gamma}$	$\mathbf{u} - \mathbf{u}_P$
$\boldsymbol{w}$	$\mathbf{u}_P - \mathbf{u}_h$
$\mathbf{q}$	$\pi\mathbf{u} - \Pi \circ \pi\mathbf{u}$

Common abbreviations

SG	Standard Galerkin (Method)
DG	Discontinuous Galerkin (Method)
IP	interior penalty
NIPG	Nonsymmetric Interior Penalty Galerkin (Method)
IIPG	Incomplete Interior Penalty Galerkin (Method)
SIPG	Symmetric Interior Penalty Galerkin (Method)
UI	under-integration
UI1	under-integration of term $VIII$ of (3.2.27), if it appears
UI2	under-integration of terms $VIII$ and IV of (3.2.27), if they appear
UI3	under-integration of terms $VIII$, IV and VI of (3.2.27), if they appear
CB	Cantilever beam (model problem)
SP	Square plate (model problem)
CM	Cook's membrane (model problem)
TB	T-shaped bracket (model problem)

Bibliography

[1] J. Alberty and C. Carstensen. Discontinuous Galerkin time discretization in elastoplasticity: motivation, numerical algorithms, and applications. *Computer Methods in Applied Mechanics and Engineering*, 191:4949–4968, 2002.

[2] D. N. Arnold. Discretization by finite elements of a model parameter dependent problem. *Numerische Mathematik*, 37:405–421, 1981.

[3] D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM Journal on Numerical Analysis*, 19:742–760, 1982.

[4] W. Bangerth, T. Heister, G. Kanschat, et al. `deal.II` *Differential Equations Analysis Library, Technical Reference*. `http://www.dealii.org`.

[5] W. Bangerth, R. Hartmann, and G. Kanschat. deal.II – a general purpose object oriented finite element library. *ACM Transactions on Mathematical Software*, 33: 24/1–24/27, 2007.

[6] F. Bassi and S. Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations. *Journal of Computational Physics*, 131:267–279, 1997.

[7] C. E. Baumann and J. T. Oden. A discontinuous *hp* finite element method for convection-diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 175:311–341, 1999.

[8] S. C. Brenner. A nonconforming mixed multigrid method for the pure displacement problem in planar linear elasticity. *SIAM Journal on Numerical Analysis*, 30:116–135, 1993.

[9] S. C. Brenner. Korn's inequalities for piecewise H^1 vector fields. *Mathematics of Computation*, 73:1067–1087, 2003.

[10] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, Second edition, 2002.

[11] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, 1991.

[12] F. Brezzi, G. Manzini, D. Marini, P. Pietra, and A. Russo. Discontinuous Galerkin approximations for elliptic problems. *Numer. Methods Partial Differential Equations*, 16:365–378, 2000.

[13] L. J. Bridgeman and T. P. Wihler. Stability and a posteriori error analysis of discontinuous Galerkin methods for linearized elasticity. *Computer Methods in Applied Mechanics and Engineering*, 200:1543–1557, 2011.

[14] Z. Cai, J. Douglas, and X. Ye. A stable nonconforming quadrilateral finite element method for the stationary Stokes and Navier–Stokes equations. *CALCOLO*, 36: 215–232, 1999.

[15] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland Publishing Company, 1980.

[16] B. Cockburn, G. Karniadakis, and C.-W. Shu, editors. *Discontinuous Galerkin Methods: Theory, Computation and Applications*. Lecture Notes in Computational Science and Engineering. Springer-Verlag, 2000.

[17] B. Cockburn, D. Schötzau, and J. Wang. Discontinuous Galerkin methods for incompressible elastic materials. *Computer Methods in Applied Mechanics and Engineering*, 195:3184–3204, 2006.

[18] C. Dawson, S. Sun, and M. F. Wheeler. Compatible algorithms for coupled flow and transport. *Computer Methods in Applied Mechanics and Engineering*, 193: 2565–2580, 2004.

[19] J. K. Djoko, B. P. Lamichhane, B. D. Reddy, and B. I. Wohlmuth. Conditions for equivalence between the Hu-Washizu and related formulations, and computational

behavior in the incompressible limit. *Computer Methods in Applied Mechanics and Engineering*, 195:4161–4178, 2006.

[20] J. K. Djoko, F. Ebobisse, A. T. McBride, and B. D. Reddy. A discontinuous Galerkin formulation for classical and gradient plasticity – Part 1: Formulation and analysis. *Computer Methods in Applied Mechanics and Engineering*, 196:3881–3897, 2007.

[21] J. Douglas and T. Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. In *Lecture Notes in Physics 58*. Springer-Verlag, 1976.

[22] J. Douglas, J. E. Santos, D. Sheen, and X. Ye. Nonconforming Galerkin methods based on quadrilateral elements for second order elliptic problems. *Mathematical Modelling and Numerical Analysis*, 33:747–770, 1999.

[23] V. Girault. A local projection operator for quadrilateral finite elements. *Mathematics of Computation*, 64:1421–1431, 1995.

[24] P. Hansbo and M. G. Larson. Discontinuous Galerkin methods for incompressible and nearly incompressible elasticity by Nitsche's method. *Computer Methods in Applied Mechanics and Engineering*, 191:1895–1908, 2002.

[25] P. Hansbo and M. G. Larson. Discontinuous Galerkin and the Crouzeix-Raviart element: application to elasticity. *Mathematical Modelling and Numerical Analysis*, 37:63–72, 2003.

[26] C. O. Horgan. Eigenvalue estimates and the trace theorem. *Journal of Mathematical Analysis and Applications*, 69:231–242, 1979.

[27] P. Houston, D. Schötzau, and T. P. Wihler. An *hp*-adaptive mixed discontinuous Galerkin FEM for nearly incompressible linear elasticity. *Computer Methods in Applied Mechanics and Engineering*, 195:3224–3246, 2006.

[28] A. Lew, P. Neff, D. Sulsky, and M. Ortiz. Optimal BV estimates for a discontinuous Galerkin method for linear elasticity. *Applied Mathematics Research Express*, 3:73–106, 2004.

[29] R. Liu, M. F. Wheeler, and C. N. Dawson. A three-dimensional nodal-based implementation of a family of discontinuous Galerkin methods for elasticity problems. *Computer and Structures*, 87:141–150, 2009.

[30] J. A. Nitsche. Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind. *Abh. Math. Univ. Hamburg*, 36:9–15, 1971.

[31] J. T. Oden, I. Babuška, and C. E. Baumann. A discontinuous *hp* finite element method for diffusion problems. *Journal of Computational Physics*, 146:491–519, 1998.

[32] R. Rannacher and S. Turek. Simple nonconforming quadrilateral Stokes element. *Numerical Methods for Partial Differential Equations*, 8:97–111, 1992.

[33] B. D. Reddy. Mixed finite element methods for one-dimensional problems: a survey. *Quaestiones Mathematicae*, 15:233–259, 1992.

[34] W. H. Reed and T. R. Hill. Triangular mesh methods for the neutron transport equation. Technical report, Los Alamos Scientific Laboratory, 1973.

[35] B. Rivière. *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*. SIAM, 2008.

[36] B. Rivière and M. F. Wheeler. Optimal error estimates for discontinuous Galerkin methods applied to linear elasticity problems. Technical report, Texas Institute for Computational and Applied Mathematics, 2000.

[37] B. Rivière, M. F. Wheeler, and V. Girault. Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. Part I. *Computational Geosciences*, 3:337–360, 1999.

[38] C. Schwab. *p- and hp- Finite Element Methods: Theory and Applications in Solid and Fluid Mechanics*. Clarendon Press, 1998.

[39] A. Ten Eyck and A. Lew. Discontinuous Galerkin methods for non-linear elasticity. *International Journal for Numerical Methods in Engineering*, 67:1204–1243, 2006.

[40] M. F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM Journal on Numerical Analysis*, 15:152–161, 1978.

[41] T. P. Wihler. Locking-free DGFEM for elasticity problems in polygons. *IMA Journal of Numerical Analysis*, 24:45–75, 2004.

[42] T. P. Wihler. Locking-free adaptive discontinuous Galerkin FEM for linear elasticity problems. *Mathematics of Computation*, 75:1087–1102, 2006.

[43] T. P. Wihler and M. Wirz. Mixed *hp*-discontinuous Galerkin FEM for linear elasticity and Stokes flow in three dimensions. *Mathematical Models and Methods in Applied Sciences*, 22(8), 2012.