Discontinuous Galerkin Methods on Polytopic Meshes

Thesis submitted for the degree of

Doctor of Philosophy

at the

University of Leicester

by

Zhaonan Dong

Department of Mathematics

University of Leicester

September 2016

"All exact science is dominated by the idea of approximation. When a man tells you that he knows the exact truth about anything, you are safe in infering that he is an inexact man."

Bertrand Arthur William Russell

Abstract

Discontinuous Galerkin Methods on Polytopic Meshes

Zhaonan Dong

This thesis is concerned with the analysis and implementation of the *hp*-version interior penalty discontinuous Galerkin finite element method (DGFEM) on computational meshes consisting of general polygonal/polyhedral (polytopic) elements. Two model problems are considered: general advection–diffusion–reaction boundary value problems and time dependent parabolic problems. New *hp*–version *a priori* error bounds are derived based on a specific choice of the interior penalty parameter which allows for edge/face–degeneration as well as an arbitrary number of faces and hanging nodes per element.

The proposed method employs elemental polynomial bases of total degree p ($\mathcal{P}_p$–bases) defined in the physical coordinate system, without requiring mapping from a given reference or canonical frame. A series of numerical experiments highlighting the performance of the proposed DGFEM are presented. In particular, we study the competitiveness of the p–version DGFEM employing a $\mathcal{P}_p$–basis on both polytopic and tensor–product elements with a (standard) DGFEM and FEM employing a (mapped) $\mathcal{Q}_p$–basis. Moreover, a careful theoretical analysis of optimal convergence rate in p for $\mathcal{P}_p$–basis is derived for several commonly used projectors, which leads to sharp bounds of exponential convergence with respect to degrees of freedom (dof) for the $\mathcal{P}_p$–basis.

Acknowledgements

I would like to acknowledge the help and support I have received from many friends during four years PhD life at University of Leicester.

Firstly, I would like to express my biggest gratitude to my first supervisor, Professor Emmanuil H. Georgoulis, for being a great tutor and advisor that everyone would hope for, and showing me how mathematicians should think. Without his constant kind support, ideas, advice and courage, I may possibly have lost all my enthusiasm and interest in Mathematics. Frankly speaking, if I did not meet him in my master study, I may not want to be a mathematician.

Secondly, I would also like to express my sincere gratitude to my second supervisor, Dr. Andrea Cangiani, for always being helpful, patient in talking about Mathematics, and his careful and rigorous attitude taught me that analysis can always be improved. Thanks for fighting together with me on the Christmas eve in 2014, when is the darkest time in my PhD.

I would like to express my gratitude to Prof. Paul Houston for giving me a lot of useful advices in implementation of DG and FEM method. Without those useful knowledge, I can not write a *hp*-version DG code for polytopic meshes smoothly. Meanwhile, I would like to thank Dr. Edward Hall for spending a lot of time on discussing about the anisotropic mesh refinement. Moveover, I would also like to thank for my master thesis supervisor Prof. Jeremy Levesley who helped me not only in Mathematics but also the reference letter supporting me to get the Graduate Teaching Assistant position.

Next, I would like to thank Oliver Sutton and Sam Cox not only for spending a lot of time on discussing the adaptivity of VEM and DG but also make me sounds more and more 'evil'. Also, Yangzhang Zhao and Qi Zhang deserves a special mention for being the best company, during the necessary coffee breaks, and for being there as good friends since since we were doing Masters. Speaking of good friends, special thanks to Younis Sabawi for his usual 'five minutes discuss', from which I learned a lot about the interface problem. I also thanks to Stephen Metcalfe and Mohammad Sabawi for discussing about the space-time adaptive DG method.

Special thanks to my parents for their love, faith and support, not only for my PhD study, but also for all of what they did for me in my life. Without their continued encouragement, I would not be brave enough to choose the life I like.

Last but by no means least, my deepest thanks go to my wife Shuheng Guo, for her everlasting love and support. Without her, my life will stop feeling happiness.

Contents

List of Figures

List of Tables

To Shuheng

Chapter 1

Introduction

1.1 Background

Mathematical modeling with ordinary differential equations (ODEs) and partial differential equations (PDEs) is widely used in diverse areas, from computational fluid dynamics, solid mechanics and optimal control, to finance, biology and geology. Many natural phenomena, e.g. diffusion, convection and reaction, can be accurately modeled by using PDEs. It is well known that there are only limited ways for finding closed form solutions of PDEs with appropriate boundary and initial conditions over particularly related solution domains. Therefore, the need to resort to numerical approximation to find the solutions of a large class of PDEs is apparent.

In the last six decades, finite element methods (FEMs) have been widely used and accepted by mathematicians and engineers as one of the most powerful tools for solving a wide range of PDEs problems. Historically, the first work in FEM was written by Richard Courant [82]. In the 1960s, finite element method began to be popular among the engineers due to its power in solving PDEs on complicated geometry with high-order approximation, as well as due to their solid mathematical foundations that has been developed for the analysis of their performance by mathematicians.

However, classical FEMs are known to lack sufficient stability properties for transport dominated PDE models. Various kind of stabilisation techniques have been designed for resolving this issue in the last 40 years, typically with the expense of the determination of a hard-to-evaluate user-defined parameter. On the other

hand, finite volume methods (FVMs) have been predominantly used for transport dominated problem in industrial software packages, especially, in computational fluid dynamics (CFD), due to their efficiency of implementation, particularly on parallel computer architectures and also their good stability for solving hyperbolic problem. However, the convergence of rate of FVMs is usually low, and their accuracy may deteriorate on irregular and/or highly stretched meshes.

Discontinuous Galerkin finite element methods (DGFEMs, for short) have enjoyed considerable success, especially during the last three decades, and are now considered as a standard variational framework for the numerical solution of many classes of problems involving partial differential equations. Loosely speaking, DGFEMs can be considered to be a hybrid between classical FEMs and FVMs. Indeed, just like in the FVM setting, information in DGFEMs is transmitted via the introduction of numerical fluxes. At the same time, DGFEMs are defined as Galerkin procedures just like FEMs, and they can easily employ approximation of arbitrary degree locally on each computational cell.

The origins of DGFEMs can be traced back to the early 1970s for the numerical solution of first–order hyperbolic problems by Reed & Hill [155]. This method is later analysed by Lesaint & Raviart [141] and by Johnson & Pitkäranta [132]; see, e.g., [79, 78, 76, 81, 97, 72, 41], and the volume [77]. In the context of elliptic PDEs, Nitsche's work on weak imposition of essential boundary conditions [149] for (classical) FEMs, allowed for the weak imposition of non-homeogeneous essential boundary conditions. This was subsequently studied by Baker [27] who proposed the first modern DGFEM for elliptic problems, later followed by Wheeler [187], Arnold [13], Baker et al. [28], and others. Also the related FEM with penalty of Babuška [26] is worth mentioning here.

In the late 1990's, a number of different DGFEMs have been developed by a number of researchers. These include the method of Bassi & Rebay [33, 34], the methods by Brezzi, Manzini, Marini, Pietra & Russo [48], and the generalisation of these ideas in the context of local discontinuous Galerkin methods (LDG) by Cockburn & Shu [80], and the so-called interior-penalty (IP) methods by Wheeler and co-workers [157, 156] and Houston, Schwab & Suli [175, 125]. Additionally, we also mention the DGFEM used by Baumann, Babuška & Oden [150, 19], which is a parameter free version of the IP method. The similarities between the above mentioned methods led Arnold, Brezzi, Cockburn & Marini to seek a unified framework for deriving and analysing DGFEMs [16, 71]. For reviews of some of

the main development before the year 2000, see monograph [77]. In recent years, DGFEMs have been applied to numerious boundary value and initial value problems, such as Stokes problems [161, 162], fourth order problems [176, 146, 107], Maxwell equation [122, 153, 121], Cahn-Hilliard equation [137, 100], Friedrichs' systems [128, 92, 93, 94] and more recently Hamilton-Jacobi-Bellman equations [168, 169], etc.

The interest in DGFEMs can be attributed to a number of factors: classical DGFEMs, such as interior penalty methods, have typically minimal communication, in the sense that only direct face-element neighbours are coupled through the exploitation of appropriate numerical fluxes; this has important advantages for imposing boundary conditions and also for parallel efficiency. Additionally, meshes containing hanging-nodes and elemental polynomial bases consisting of locally variable polynomial degrees are also admissible, owing to the lack of pointwise continuity requirements across the mesh-skeleton. This allows for the variation of the order of polynomials over the computational domain (p–refinement), which in combination with local mesh adaptation (h–refinement) leads to hp–version approximations. Furthermore, powerful solvers are now available for the resulting linear systems; indeed, both domain decomposition preconditioners, see, for example, [5, 6, 99, 140, 10, 9], and the references cited therein, as well as multigrid solvers, cf. [11, 12, 45, 44], have been developed.

On the other hand, many practitioners often object that DGFEMs are computationally expensive, as for a given mesh and polynomial order, DGFEMs lead to an increase in the number of degrees of freedom compared to classical FEM for comparable accuracy, typically when discretizing elliptic operators. This is a somehow simplistic argument, since it overlooks all the key aforementioned and other potential advantages of DGFEMs in terms of their applicability, versatility and mesh-flexibility. Indeed, as we shall see below, within the DGFEM framework, it is possible to employ the same underlying approximating space of piecewise polynomials, irrespective of the structure of the PDE of interest. Moreover, the flexibility offered by different choices of numerical fluxes allows for the design of DGFEMs with desirable conservation properties of important quantities (e.g., mass, momentum or energy conservation).

Additionally, DGFEM elemental bases can be constructed to contain fewer degrees of freedom than their (conforming) FEM counterparts for quadrilateral/hexahedral or, general, polytopic elements with more than d faces. The underlying idea

in this context is the use of physical frame (i.e., without resorting to the use of local element mappings) polynomial basis functions of *total* degree, say p, henceforth, denoted by $\mathcal{P}_p$, *independently* of the shape of the element; see, for example, in [31, 32, 30, 61]. This way, the order of convergence of the underlying method is independent of the element shape; cf., [14, 15] for a detailed discussion of this issue, when element mappings are employed. Indeed, as noted in our recent work [61], when the underlying mesh consists of tensor-product elements, e.g., quadrilaterals in 2D and hexahedra in 3D, the use of $\mathcal{P}_p$ polynomial spaces not only renders the underlying DGFEM more efficient than the standard DGFEM using tensor-product polynomials of degree p in each coordinate direction ($\mathcal{Q}_p$), but also more efficient than the standard FEM, as the polynomial degree p increases. Going one step further, the exploitation of DGFEMs using polynomial spaces defined in the physical frame, means that DGFEMs naturally allow for the use of computational meshes consisting of general polytopic elements.

This work is concerned with the theoretical analysis and practical performance of the *hp*-version interior penalty discontinuous Galerkin method (*hp*-IP DGFEM), for boundary value problems in non-negative characteristic form on general *polytopic* elements (polygonal/polyhedral elements in two/three space dimensions). Moreover, this work concerns with space-time *hp*-IP DGFEM for parabolic initial/boundary value problems over *prismatic* elements (polytopic spatial elements tensorised with a time interval).

Numerical methods on polytopic elements have gained substantial traction in recent years for a number of important reasons. A key underlying issue for all classes of FEMs/FVMs is the design of a suitable computational mesh upon which the underlying PDE problem will be discretized. The problem of good mesh design has to address two competing traits. On the one hand, the mesh should provide an accurate representation of the given computational geometry with sufficient resolution for accurate numerical approximations. On the other hand, there are cases where a 'coarse' mesh contains already too many degrees of freedom for computation, rendering the computations impractical, or even intractable. Such cases are often met in practice. Indeed, standard mesh generators typically generate grids consisting of triangular/ quadrilateral elements in 2D and tetrahedral/hexahedral/prismatic/pyramidal elements in 3D; these will be, henceforth, collectively termed a *standard element shapes*. In the presence of essentially lower-dimensional solution features (e.g., boundary layers), anisotropic meshing may be exploited. In areas of high curvature, however, the use of such highly-stretched elements may

lead to element self-intersection, unless the curvature of the geometry is carefully 'propagated' into the interior of the mesh through the use of (computationally expensive) isoparametric element mappings. These issues are particularly pertinent in the context of high–order methods, since in this setting, accuracy is often achieved by the use of coarse meshes combined with high order local basis; the flexibility in the shape of coarse meshes is, therefore, crucial in this context for the efficient approximation of localised geometrical features of the underlying solutions. Hence, it is obvious that, increasing dramatically the flexibility in the admissible element shapes in the mesh, can potentially deliver dramatic savings in computational cost.

An alternative approach is to exploit general meshes consisting of polytopic (i.e., polygonal in 2D and polyhedral in 3D) elements. In the context of discretizing PDEs in complicated geometries, Composite Finite Elements (CFEs) (both conforming and DGFEM versions), have been developed [116, 115, 8, 110], which exploit general meshes consisting of polytopic elements arising as agglomerates of standard shaped elements; cf., also the closely related (but more restrictive, in terms of the basis functions it employs, compared to [8, 110]), so-called, agglomerated DGFEM [30, 31, 32]; cf., also the unfitted discontinuous Galerkin Method [119, 35], which is one of the first works considering the computational issues related to the use of total degree basis over general shaped elements for DGFEMs to the best of author's knowledge. More recently, the Cut Finite Element Methods (both conforming and DGFEM versions), have been developed [56, 51, 55], which use a fixed background meshes to represent the geometry of the domain and build on a general finite element formulation for the approximation of PDEs, in the bulk and on surfaces, that can handle elements of complex shape and where boundary and interface conditions are built into the discrete formulation. In addition, the Hybrid High-Order methods have been first introduced [88] and developed [85, 86, 68, 73]. They support general polyhedral meshes and delivers an arbitrary-order accurate approximation by intermediating the cell-based discrete unknowns in addition to the face-based ones. The cell-based unknowns can be eliminated by static condensation which improves the efficiency.

In the context of the numerical simulation of evolution PDE problems, the resolution of time-dependent sharp solution features (layers, interfaces, shocks, etc.) remains a significant challenge in the quest of resolved computations, e.g., in

CFD. Mesh-geometry freedom, in conjunction with variable order local polynomial elemental degrees, is expected to achieve accurate approximation of lower–dimensional features, while simultaneously reducing significantly the sizes of the resulting linear systems required to be solved per time-step.

The use of polytopic meshes in the context of characteristic-based/Lagrange–Galerkin methods is also highly relevant. Such moving-meshing methodologies result to extremely general/irregular node configurations, which give rise to highly irregular element shapes. The practical relevance and potential impact of employing such general computational meshes is an extremely exciting topic which has witnessed a vast amount of research in recent years by a number of leading research groups. In the conforming setting, we mention the CFE method [116, 115], the Polygonal Finite Element Method [174], and the Extended Finite Element Method [101]. These latter two approaches achieve conformity by enriching/modifying the standard polynomial finite element spaces, in the spirit of the Generalized Finite Element framework of Babuška & Osborn in [23]. Typically, the handling of non-standard shape functions carries an increase in computational effort. The recently proposed Virtual Element Method [37, 40, 65, 177, 38], overcomes this difficulty, achieving the extension of conforming finite element methods to polytopic elements while maintaining the ease of implementation of these schemes; see also the closely related Mimetic Finite Difference method, e.g., [39, 47, 64].

We point out that in all of the above mentioned methods, the construction of the finite element space depends on the geometrical information of the underlying polytopic elements in various ways: the number of basis depends on the number of face of polytopic elements, or the stability of the method is lost when the measure of faces is degenerating.

In this work, we will introduce the mathematical construction and analysis of *hp*-version DGFEMs on meshes consisting of extremely general classes of polytopes. In particular, these meshes may contain d-dimensional polytopes with arbitrarily small $(d-k)$-dimensional faces, for $k = 1, \ldots, d-1$. Here, the construction of the proposed finite element space is *independent* of the number of faces per element. In the analysis presented below, stability and a priori error bounds will be deduced which are sharp with respect to face degeneration under a refined choice of the (user-defined) discontinuity-penalization parameter.

We briefly describe the mesh assumptions over the polytopic meshes which this work is based on. Due to the general geometry of the polytopic elements allowed,

we need to make assumptions on number of faces per elements and/or shape-regularity for polytopic elements in order to derive the stability and a-priori error bound depends explicitly on the geometrical information of the meshes.

The total number of $(d-1)$-dimensional faces of simplicial meshes and tensor product-type meshes are $(d+1)$ and 2^d, respectively. For $d = 2, 3$, the number of faces are uniformly bounded for the standard meshes.(The same condition holds for pyramidal and prismatic elements.) However, even for $d = 2$, polygons with arbitrary number of faces exist. So in order to extend the *hp*-IP DGFEM from standard meshes to polytopic meshes, it is very natural to start working on the polytopic meshes with bounded number of faces. We point out that polytopic elements satisfying above mesh assumption can still be shape regular.

On the other hand, there are some simple and "nice" shaped polytopic elements with unbounded number of faces which are excluded by the above mesh assumptions, 'nice' in the sense of satisfying shape regularity assumptions. In this work, a new shape regularity assumption which is stronger than the classical shape regularity assumptions for polytopic elements will be considered.

Finally, we will also present some new *hp*-approximation results for some commonly used projectors over standard tensor-product typed elements with $\mathcal{P}_p$ basis. By utilising the new results, we can prove for piecewise analytic problems in this work, DGFEMs with $\mathcal{P}_p$ basis has a steeper exponential convergence compared to DGFEMs with $\mathcal{Q}_p$ basis over tensor product elements, and the better convergence only depends on dimension. The sharpness of the approximation results is also verified by the numerical experiments.

1.2 Overview

In this thesis we present the a priori error analysis of *hp*-version DGFEMs on extremely general classes of meshes consisting of polytopic meshes, containing polytopes with arbitrary small $(d-k)$-dimensional faces, $k = 1, \ldots, d-1$. Additionally, series of numerical examples will be presented to conform the theoretical analysis. This work is structured as follows; the main results can also be found in [61, 59, 58, 7], as well as in the monograph [60] which is in preparation.

In Chapter 2, we start by introducing the functional space setting use to define the model problems and discontinuous Galerkin methods (Section 2.1). Next, the

discontinuous Galerkin method for first-order hyperbolic problems will be derived (Section 2.2). Then, we present the general discontinuous Galerkin approach to second order elliptic problems and derive the interior penalty DGFEMs from a particular choice of numerical fluxes (Section 2.3). Finally, we derive the IP DGFEM formulation for PDEs in non-negative characteristic form (Section 2.4).

In Chapter 3, we first fix a set of mesh assumptions allowing for very general polygonal meshes with a uniformly bounded number of faces per element (Section 3.1). Based on these assumptions, in Section 3.2 we derive inverse estimates over polytopic elements, making use of classical *hp*-version inverse estimates over standard simplical meshes cf. [167, 183, 185, 184]. The resulting inverse estimates are sharp with respect to the face degeneration by using ideas in [104]. In Section 3.3, we derive the *hp*-version polynomial approximation results over polytopic elements. The key technique is to use the classical Babuška-Suri operator [24] over a simplical spatial mesh covering.

In Chapter 4, we present the analysis for *hp*-version IP DGFEM for elliptic problems over polytopic elements. In Section 4.1, we define the elliptic problems satisfying the uniform ellipticity conditions. In Section 4.2, we prove coercivity and continuity of the IP DGFEM method assuming a uniformly bounded number of faces per polytopic element. Then we derive the *hp*-version a priori error bound by using the approximation results in Section 3.3. We emphasize that *the coercivity and continuity constants depend on the number of faces per element, but is independent of shape regularity of polytopic elements.* In Section 4.3, we present a new proof of coercivity and continuity conditions and we derive the *hp*-version a priori error bound with a different mesh assumption allowing for arbitrary number of faces per polytopic element. *In this case, the coercivity and continuity constants depend on the shape regularity of polytopic elements, but is independent of number of faces per element.* Although the a priori error bounds differs slightly under the two different mesh assumptions, both of the error bounds will be h optimal and p suboptimal by $1/2$ order, if we consider quasi-uniform meshes. In Section 4.4, several numerical examples are presented.

In Chapter 5, we present the analysis for *hp*-version DGFEM for partial differential equations with non-negative characteristic form over polytopic elements. For the sake of simplicity, we will use the bounded number of faces per element mesh assumption in this chapter. The reason for not using the new shape regularity

mesh assumption is due to lack of inverse estimates from H^1–seminorm to L^2–norm over the element satisfying the underlying assumption. In Section 5.1, we define the partial differential equations in non-negative characteristic form under the same setting of Houston, Schwab, Süli [125]. In Section 5.2, we derive a priori bounds for the *hp*–version IP DGFEM for this class of problems. Due to the lack of *hp*–approximation results for the local L^2–projection operator on polytopic elements, it is not possible to directly generalise the analysis from [125] to meshes consisting of such elements. To address this issue, we prove an inf-sup condition for the underlying DGFEM, with respect to a stronger streamline–diffusion type norm, for simple advection coefficients, thereby extending respective results from [133, 49, 18, 57] to the current setting. This naturally leads to a priori bounds for the *hp*-version DGFEM for this general class of linear PDE problems on very general polytopic element with possibly arbitrarily small/degenerate $(d-k)$–dimensional element facets, $k = 1, \ldots, d-1$. The resulting a priori bound will be h optimal and p suboptimal by $1/2$ order in pure hyperbolic cases and pure elliptic cases. In Section 5.3, a series of numerical examples is presented to test performance of the IP DGFEM. This analysis is also novel for classical simplicial or tensor-product type elements.

In Chapter 6, we present the analysis for *hp*-version space-time DGFEMs for parabolic problems over *prismatic space-time elements* under shape regularity mesh assumption for the spatial mesh. Moreover, we will define the new space-time finite element space in order to adapting to the space-time DGFEMs framework. In Section 6.1, we present the problem setting for parabolic PDEs. In Section 6.2, we will derive a priori bounds for the *hp*-version space-time DGFEMs for parabolic problems. Here, since total degree $\mathcal{P}_p$ basis is utilised over each prismatic space-time elements, there is no space-time tensor product structure in local basis. The classical stability proof [180] depends on utilising tensor product of spatial and temporal projectors, which is not possible under the current setting. We prove the unconditional stability of the new space-time DGFEMs, via the proof of an inf-sup condition for space-time elements with arbitrary aspect ratio between the time-step λ and the local spatial mesh-size h. The resulting inf-sup stability is independent of number of faces per element. Furthermore, under a *space-time shape-regularity* assumption, *hp*-a priori error bounds are proven in the $L^2(H^1)$– and $L^2(L^2)$–norms, combining the classical duality approach with careful use of approximation arguments to circumvent the fundamental impossibility of applying 'tensor-product' arguments (as is standard in this context [180]) in the present

setting. Instead, a new argument, based on judicious use of the space-time local degrees of freedom, eventually delivers the $L^2(H^1)$–norm and $L^2(L^2)$–norm error bound, with constants independent of number of faces per element. Here, the resulting a priori bound is h optimal and p suboptimal by half order in $L^2(H^1)$–norm, while the a priori bound is h suboptimal by half order and p suboptimal by $3/2$ order in $L^2(L^2)$–norm. In Section 6.3, extensive comparison among different combinations of spatial and temporal discretizations and the new approach are given through a series of numerical examples.

In Chapter 7, we present some *hp*-version polynomial approximation results for commonly used projectors onto the $\mathcal{P}_p$ basis and serendipity ($\mathcal{S}_p$) basis on tensor product elements. In Section 7.1, we will derive the sharp *hp*-bound for L^2-orthogonal projections and H^1-projections onto the $\mathcal{P}_p$ and $\mathcal{S}_p$, respectively, in several different norms. Classical *hp*-approximation theory depends on a tensor product $\mathcal{Q}_p$ basis, while the *hp*-approximation theory for $\mathcal{P}_p$ and $\mathcal{S}_p$ are usually constructed by using the fact that there always exist a $\mathcal{Q}_q$ basis, $q < p$, as a subspace of $\mathcal{P}_p$ or $\mathcal{S}_p$. We emphasize that *the resulting hp-bound for $\mathcal{P}_p$ and $\mathcal{S}_p$ are p-optimal when the underlying function has finite Sobolev regularity, and it is not p-optimal for piecewise analytic functions.* The new *hp*-approximation result for $\mathcal{P}_p$ and $\mathcal{S}_p$ bases are optimal in p, not only for functions with finite Sobolev regularity, but also for analytic functions. In fact, the analysis shows that *the extra basis functions in $\mathcal{Q}_p$ compared to $\mathcal{P}_p$ or $\mathcal{S}_p$ only reduce the constant in the error bound without improving the rate in p.* The main tools used in the proof are orthogonal polynomial expansions, together with judicious choice of the local basis. In Section 7.2, we will apply the new approximation results to prove the exponential convergence for DGFEMs with the $\mathcal{P}_p$ basis and FEMs with the $\mathcal{S}_p$ basis over standard tensor product elements for piecewise analytic problems. Here, the main proof is based on [125, 124, 167]. Moreover, we will prove that exponential convergence for DGFEMs with $\mathcal{P}_p$ basis is steeper than DGFEMs with $\mathcal{Q}_p$ basis in error against number of degrees of freedom under p-refinement, respectively, thereby highlighting that DGFEMs can be cheaper by standard FEM per dof in certain regimes. In Section 7.3, we present several examples to verify the sharpness of the theory.

In Chapter 8, we conclude this work and look at some possible future directions of further research.

Chapter 2

Discontinuous Galerkin Methods

In this chapter we will establish the general settings for this work is based on, and also we will introduce the discontinuous Galerkin finite element methods (DGFEMs).

2.1 Sobolev Spaces

Let Ω be a bounded open subset of $\mathbb{R}^d$, $d \geq 1$, with boundary $\partial\Omega$; moreover, we write $|\Omega|$ to denote the measure of the domain Ω. For $1 \leq p \leq \infty$, let $L^p(\Omega)$ denote the usual Lebesgue space of real–valued functions with norm $\|\cdot\|_{L^p(\Omega)}$, defined by

$$\|v\|_{L^p(\Omega)} := \left(\int_\Omega |v(\mathbf{x})|^p \, \mathrm{d}\mathbf{x}\right)^{1/p},$$

in the case $1 \leq p < \infty$, and in the case $p = \infty$

$$\|v\|_{L^\infty(\Omega)} := \operatorname{ess\,sup}_{\mathbf{x}\in\Omega} |v(\mathbf{x})|.$$

Given a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$, $\alpha_i \in \mathbb{N}_0$, $i = 1, \ldots, d$, of length $|\alpha| := \sum_{i=1}^d \alpha_i$, we let $D^\alpha := D_1^{\alpha_1} \ldots D_d^{\alpha_d}$ and $D_j = \partial/\partial x_j$ for $j = 1, \ldots, d$. For $m \in \mathbb{N}_0 \cup \{\infty\}$, we denote by $C^m(\Omega)$ the set of all continuous real–valued functions defined on Ω such that $D^\alpha v$ is continuous on Ω for all $|\alpha| \leq m$. In particular, when $m = 0$, we simply write $C(\Omega)$ instead of $C^0(\Omega)$. The subspace $C_0^m(\Omega)$ will denote the set of functions in $C^m(\Omega)$ which have compact support in Ω.

Next, we recall the definition of a Sobolev space (see, e.g., [4]); with a slight abuse of notation, we also write $D^\alpha v$ to denote the weak derivative of a sufficiently smooth function v.

Definition 2.1 (Sobolev space)**.** For $m \in \mathbb{N}_0$, we define the *Sobolev space* $W^{m,p}(\Omega)$ over an open domain $\Omega \subset \mathbb{R}^d$, by

$$W^{m,p}(\Omega) := \{u \in L^p(\Omega) : D^\alpha u \in L^p(\Omega) \text{ for } |\alpha| \leq m\}, \tag{2.1}$$

with associated norm $\|\cdot\|_{W^{m,p}(\Omega)}$ and seminorm $|\cdot|_{W^{m,p}(\Omega)}$ given by:

$$\|u\|_{W^{m,p}(\Omega)} := \Big(\sum_{|\alpha|\leq m} \|D^\alpha u\|^p_{L^p(\Omega)}\Big)^{1/p}, \quad |u|_{W^{m,p}(\Omega)} := \Big(\sum_{|\alpha|=m} \|D^\alpha u\|^p_{L^p(\Omega)}\Big)^{1/p},$$

for $p \in [1, \infty)$, and

$$\|u\|_{W^{m,\infty}(\Omega)} := \max_{|\alpha|\leq m} \|D^\alpha u\|_{L^\infty(\Omega)}, \quad |u|_{W^{m,\infty}(\Omega)} := \max_{|\alpha|=m} \|D^\alpha u\|_{L^\infty(\Omega)},$$

for $p = \infty$, respectively.

For $p = 2$, we write $H^m(\Omega)$ to denote Hilbertian Sobolev spaces. Further, we define $H^m_0(\Omega)$ in the following way.

$$H^m_0(\Omega) := \{u : \|u\|_{H^m(\Omega)} < \infty, \text{ and } D^\alpha u|_{\partial\Omega} = 0 \text{ for } |\alpha| \leq m-1\}, \tag{2.2}$$

Next, we give the definition for dual norm of Sobolev spaces

$$\|u\|_{H^{-m}(\Omega)} := \sup_{v\in H^m_0(\Omega)} \frac{(u,v)_{L^2(\Omega)}}{\|v\|_{H^m(\Omega)}}, \tag{2.3}$$

where $(u,v)_{L^2(\Omega)} = \int_\Omega uv \,\mathrm{d}\mathbf{x}$ denotes the standard L^2 inner product. We compress the notation of the L^2 product and norm by $(\cdot,\cdot)_{L^2(\Omega)} = (\cdot,\cdot)$ and $\|\cdot\|_{L^2(\Omega)} = \|\cdot\|$ respectively, on Ω.

Definition 2.2. For $m \in \mathbb{N}_0$, we define the dual space of the Sobolev space H^m_0, by

$$H^{-m}(\Omega) := \{u : \|u\|_{H^{-m}(\Omega)} < \infty\}, \tag{2.4}$$

We point out that fractional order Sobolev spaces, i.e., where the *Sobolev index* $m \in \mathbb{R}$ are defined by (standard) function-space interpolation procedures; for more details concerning these techniques, we refer to [4], for example.

Finally, we introduce the Bochner spaces needed for time dependent problems. For $1 \leq p \leq \infty$, we define the spaces $L^p(0,T;X)$, with X being a real Banach space with norm $\|\cdot\|_X$, consisting of all measurable functions $v : [0,T] \to X$, for which

$$\|v\|_{L^p(0,T;X)} := \left(\int_0^T \|v(t)\|_X^p \, \mathrm{d}t \right)^{\frac{1}{p}} < \infty, \qquad 1 \leq p < \infty, \tag{2.5}$$

$$\|v\|_{L^\infty(0,T;X)} := \operatorname{ess} \sup_{t \in [0,T]} \|v(t)\|_X < \infty, \qquad p = \infty. \tag{2.6}$$

We denote by $C(0,T;X)$ the space of continuous function $v : [0,T] \to X$ with bounded norms

$$\|v\|_{C(0,T;X)} := \max_{t \in [0,T]} \|v(t)\|_X. \tag{2.7}$$

Throughout this work, we denote by $\mathcal{T}_h$ a subdivision of the domain Ω into disjoint open elements κ such that $\bar{\Omega} = \cup_{\kappa \in \mathcal{T}_h} \bar{\kappa}$. Moreover, for $\kappa \in \mathcal{T}_h$, we define $h_\kappa := \operatorname{diam}(\kappa)$ to be the diameter of the element κ. We stress that when $\kappa \in \mathcal{T}_h$ is polytopic, it is possible to be shape-regular in the sense of [70] *and* have faces with arbitrarily small diameters compared to h_κ. The detailed mesh assumptions will be presented at the beginning of following chapters.

On the basis of the subdivision $\mathcal{T}_h$ we define the broken Sobolev space $H^{\mathbf{l}}(\Omega, \mathcal{T}_h)$, up to composite order $\mathbf{s}$, by

$$H^{\mathbf{l}}(\Omega, \mathcal{T}_h) = \{ u \in L^2(\Omega) : u|_\kappa \in H^{l_\kappa}(\kappa) \quad \forall \kappa \in \mathcal{T}_h \}.$$

Moreover, for $v \in H^1(\Omega, \mathcal{T}_h)$, we define the broken gradient $\nabla_h v$ by $(\nabla_h v)|_\kappa = \nabla(v|_\kappa), \kappa \in \mathcal{T}_h$.

2.2 Discretization of first–order hyperbolic PDEs

To highlight the key aspects concerning the construction of DGFEMs, while keeping notation to a minimum, we first consider the discretization of a first–order linear Cauchy problem. To this end, let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, a bounded Lipschitz domain with boundary $\partial\Omega$, $c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$, and $\mathbf{b} := (b_1, b_2, \ldots, b_d)^\top \in [W^{1,\infty}(\Omega)]^d$. Furthermore, the inflow and outflow boundaries of the domain Ω are

denoted, respectively, by

$$\partial_-\Omega = \{\mathbf{x} \in \partial\Omega : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0\}, \qquad \partial_+\Omega = \{\mathbf{x} \in \partial\Omega : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) > 0\},$$

where $\mathbf{n}$ denotes the unit outward normal to $\partial\Omega$. Upon defining the graph space

$$G_b(\Omega) := \{v \in L^2(\Omega) : \mathbf{b} \cdot \nabla v \in L^2(\Omega)\},$$

we seek $u \in G_b(\Omega)$ such that

$$\mathbf{b} \cdot \nabla u + cu = f \quad \text{in } \Omega, \tag{2.8}$$

$$u = g \quad \text{on } \partial_-\Omega. \tag{2.9}$$

From the well-posedness of the above problem in graph space we know that the boundary $\partial_*\Omega = \{\mathbf{x} \in \partial\Omega : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = 0\}$ will play no role, see [84, Chapter 2] for details.

Before introducing the DGFEM approximation of (2.8), (2.9), we first consider a standard (conforming) FEM discretization based on employing *weakly imposed* boundary conditions, cf. [130], for example.

To approximate the solution of (2.8), (2.9) with a FEM, we first consider a shape–regular mesh $\mathcal{T}_h$ of the computational domain Ω, assuming, for simplicity, that $\mathcal{T}_h$ consists of d–dimensional simplicial elements $\kappa \in \mathcal{T}_h$. Letting $p \geq 1$ denote the finite element polynomial degree, we introduce the finite element space

$$V_C^p(\mathcal{T}_h) = \{u \in C(\Omega) : u|_\kappa \in \mathcal{P}_p(\kappa), \kappa \in \mathcal{T}_h\},$$

where $\mathcal{P}_p(\kappa)$ denotes the space of polynomials of total degree p on κ.

The FEM reads: find $u_h \in V_C^p(\mathcal{T}_h)$ such that

$$\int_\Omega (\mathbf{b} \cdot \nabla u_h + cu_h)\, v_h \,\mathrm{d}\mathbf{x} - \int_{\partial_-\Omega} \mathbf{b} \cdot \mathbf{n}\, u_h v_h \,\mathrm{d}s = \int_\Omega f v_h \,\mathrm{d}\mathbf{x} - \int_{\partial_-\Omega} \mathbf{b} \cdot \mathbf{n}\, g v_h \,\mathrm{d}s \tag{2.10}$$

for all $v_h \in V_C^p(\mathcal{T}_h)$. It is well-known that this method may exhibit numerical instabilities in the form of spurious oscillations [130]. Even in cases where meaningful solutions (free of spurious oscillations) are computed by (2.10), these typically converge at suboptimal rates when compared with the approximation power of $V_C^p(\mathcal{T}_h)$ [130].

To address these concerns, (2.10) should be supplemented by appropriate numerical stabilization in order to render the underlying scheme stable, e.g., by employing the so-called streamline-diffusion FEM, whereby the test functions arising in the volume integrals can be replaced by $v_h + \delta \mathbf{b} \cdot \nabla v_h$. In the h–version setting, i.e., when the polynomial degree p is kept fixed, the analysis undertaken in [131] indicates that $\delta = \mathcal{O}(h)$; the generalisation to the hp–setting outlined in [124] shows that $\delta = \mathcal{O}(h/p)$. Another choice is the so-called continuous interior penalty method [50, 53, 52, 54].

The essential idea behind the DGFEM discretization of (2.8)-(2.9) is to employ the scheme (2.10) elementwise, subject to a prescribed boundary condition on the inflow boundary of each element. This way we enhance the numerical stability of the approximation a the expense of introducing more degrees of freedom (in this d-simplicial mesh) as we will be seeking discontinuous approximations belonging to the DGFEM space

$$V^p(\mathcal{T}_h) = \{u \in L^2(\Omega) : u|_\kappa \in \mathcal{P}_p(\kappa), \kappa \in \mathcal{T}_h\},$$

defined for $p \geq 0$.

To make this precise, we first need to introduce some notation. For $p \geq 0$ we introduce the DGFEM space

$$V^p(\mathcal{T}_h) = \{u \in L^2(\Omega) : u|_\kappa \in \mathcal{P}_p(\kappa), \kappa \in \mathcal{T}_h\}.$$

(For simplicity of the exposition here, we only consider a uniform polynomial degree distribution over the mesh $\mathcal{T}_h$; the general hp–version case will be treated in the chapters below.) For any element $\kappa \in \mathcal{T}_h$, we denote by $\partial\kappa$ the union of $(d-1)$–dimensional open faces of κ. Then, the inflow and outflow parts of $\partial\kappa$ are defined as:

$$\partial_-\kappa = \{\mathbf{x} \in \partial\kappa, \quad \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}_\kappa(\mathbf{x}) < 0\}, \quad \partial_+\kappa = \{\mathbf{x} \in \partial\kappa, \quad \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}_\kappa(\mathbf{x}) > 0\},$$

respectively, where $\mathbf{n}_\kappa(\mathbf{x})$ denotes the unit outward normal vector to $\partial\kappa$ at $\mathbf{x} \in \partial\kappa$.

Given $\kappa \in \mathcal{T}_h$, we denote by v_κ^+, the trace of a function $v \in H^1(\Omega, \mathcal{T}_h)$ on $\partial\kappa$, relative to κ. Then for almost every $\mathbf{x} \in \partial\kappa \backslash \partial\Omega$, there exists a unique element $\kappa' \in \mathcal{T}_h$ such that $\mathbf{x} \in \partial\kappa'$; thereby, the outer or exterior trace v_κ^- of v on $\partial\kappa \backslash \partial\Omega$, relative to κ, is defined as the inner trace $v_{\kappa'}^+$ relative to the element(s) κ' such that the intersection of $\partial\kappa'$ with $\partial\kappa \backslash \partial\Omega$ has positive $(d-1)$–dimensional measure.

Then, the *upwind jump* of u across $\partial_-\kappa\backslash\partial\Omega$ is defined by

$$\lfloor v \rfloor_\kappa := v_\kappa^+ - v_\kappa^-. \tag{2.11}$$

We note that the sign of above *upwind jump* depends on the direction of the flow over each element $\kappa \in \mathcal{T}_h$. In the following, when it is clear from the context to which element κ in the subdivision $\mathcal{T}_h$ the quantities $v_\kappa^\pm$ correspond to, for the sake of notational simplicity we shall suppress the letter κ in the subscript and write, respectively, $v^\pm$ instead.

With this notation, motivated by (2.10), we may introduce the following local FEM formulation: for each $\kappa \in \mathcal{T}_h$, find $u_h \in V^p(\mathcal{T}_h)$, such that

$$\begin{aligned}\int_\kappa (\mathbf{b}\cdot\nabla u_h + c u_h)\, v_h \,\mathrm{d}\mathbf{x} - \int_{\partial_-\kappa} \mathbf{b}\cdot\mathbf{n}_\kappa\, u_h^+ v_h^+ \,\mathrm{d}s \\ = \int_\kappa f v_h \,\mathrm{d}\mathbf{x} - \int_{\partial_-\kappa} \mathbf{b}\cdot\mathbf{n}_\kappa\, \hat{g} v_h^+ \,\mathrm{d}s,\end{aligned} \tag{2.12}$$

for all $v_h \in V^p(\mathcal{T}_h)$, where

$$\hat{g}(\mathbf{x}) = \begin{cases} u_h^-(\mathbf{x}), & \mathbf{x} \in \partial_-\kappa\backslash\partial\Omega, \\ g(\mathbf{x}), & \mathbf{x} \in \partial_-\kappa \cap \partial\Omega. \end{cases}$$

Summing (2.12) over $\kappa \in \mathcal{T}_h$ and employing the definition of $\hat{g}$, the DGFEM approximation to (2.8), (2.9) is given by: find $u_h \in V^p(\mathcal{T}_h)$ such that

$$\begin{aligned}\sum_{\kappa\in\mathcal{T}_h} \Bigg\{ \int_\kappa (\mathbf{b}\cdot\nabla u_h + c u_h)\, v_h \,\mathrm{d}\mathbf{x} - \int_{\partial_-\kappa\backslash\partial\Omega} \mathbf{b}\cdot\mathbf{n}_\kappa\, \lfloor u_h \rfloor v_h^+ \,\mathrm{d}s \\ - \int_{\partial_-\kappa\cap\partial\Omega} \mathbf{b}\cdot\mathbf{n}_\kappa\, u_h^+ v_h^+ \,\mathrm{d}s \Bigg\} = \sum_{\kappa\in\mathcal{T}_h} \left\{ \int_\kappa f\, v_h \,\mathrm{d}\mathbf{x} - \int_{\partial_-\kappa\cap\partial\Omega} \mathbf{b}\cdot\mathbf{n}_\kappa\, g v_h^+ \,\mathrm{d}s \right\},\end{aligned} \tag{2.13}$$

for all $v_h \in V^p(\mathcal{T}_h)$. Integrating the first term in (2.13) by parts gives rise to the following equivalent formulation: find $u_h \in V^p(\mathcal{T}_h)$ such that

$$\begin{aligned}\sum_{\kappa\in\mathcal{T}_h} \Bigg\{ \int_\kappa ((c - \nabla\cdot\mathbf{b}) u_h v_h - u_h \mathbf{b}\cdot\nabla v_h) \,\mathrm{d}\mathbf{x} + \int_{\partial_-\kappa\backslash\partial\Omega} \mathbf{b}\cdot\mathbf{n}_\kappa\, u_h^- v_h^+ \,\mathrm{d}s \\ + \int_{\partial_+\kappa} \mathbf{b}\cdot\mathbf{n}_\kappa\, u_h^+ v_h^+ \,\mathrm{d}s \Bigg\} = \sum_{\kappa\in\mathcal{T}_h} \left\{ \int_\kappa f\, v_h \,\mathrm{d}\mathbf{x} - \int_{\partial_-\kappa\cap\partial\Omega} \mathbf{b}\cdot\mathbf{n}_\kappa\, g v_h^+ \,\mathrm{d}s \right\},\end{aligned} \tag{2.14}$$

for all $v_h \in V^p(\mathcal{T}_h)$.

To motivate why the above method has the potential of yielding significant improvement in the stability of the approximate solution it computes, let us consider a component-wise constant wind $\mathbf{b}$ across Ω. We observe that, then, $v_h+\delta\mathbf{b}\cdot\nabla v_h \in V^p(\mathcal{T}_h)$ for all $\delta > 0$ when $v_h \in V^p(\mathcal{T}_h)$. Therefore, the fact that such a function belongs to the element-wise discontinuous space $V^p(\mathcal{T}_h)$ allows for partial derivatives of the basis functions to be included in the finite element space, which gives the control of the derivative along the advective direction. This, in conjunction with the weak imposition of the elemental boundary conditions, has the effect of enhancing stability.

An alternative approach to derive the method (2.14), which is more generally applicable for the discretization of first–order nonlinear hyperbolic conservation laws, is to employ the concept of numerical fluxes, exploited widely within FVMs, see, e.g., [118]. In this approach, we begin again by the local weak formulation of (2.8), (2.9), and we integrate by parts the leading order term. (Notice that if $\mathbf{b}$ depends on the solution u also, the aforementioned integration by parts avoids the presence of, potentially cumbersome, derivatives of $\mathbf{b}$ in the numerical method.)

With this in mind, multiplying (2.8) by a smooth test function v and integrating over a single element $\kappa \in \mathcal{T}_h$ gives: find $u|_\kappa$ such that $u|_{\partial\_\Omega} = g$ and

$$\int_\kappa ((c - \nabla\cdot\mathbf{b})uv - u\mathbf{b}\cdot\nabla v)\,\mathrm{d}\mathbf{x} + \int_{\partial\kappa} \mathbf{b}\cdot\mathbf{n}_\kappa\, u^+ v^+ \,\mathrm{d}s = \int_\kappa fv\,\mathrm{d}\mathbf{x}. \tag{2.15}$$

The DGFEM discretization of (2.15) is then based on replacing the analytical solution u by the DGFEM approximation u_h and the test function v by v_h, where both u_h and v_h belong to $V^p(\mathcal{T}_h)$. Additionally, since u_h is discontinuous between neighbouring elements, we must replace the flux $\mathbf{b}\cdot\mathbf{n}_\kappa\, u^+$ by a numerical flux function $\mathcal{H}(u_h^+, u_h^-, \mathbf{n}_\kappa)$, which depends on both the inner– and outer–trace of u_h on $\partial\kappa$, $\kappa \in \mathcal{T}_h$, and on the unit outward normal vector $\mathbf{n}_\kappa$ to $\partial\kappa$. Summing over the elements κ in the mesh $\mathcal{T}_h$ yields the DGFEM: find $u_h \in V^p(\mathcal{T}_h)$ such that

$$\sum_{\kappa\in\mathcal{T}_h} \left\{ \int_\kappa ((c - \nabla\cdot\mathbf{b})u_h v_h - u_h\mathbf{b}\cdot\nabla v_h)\,\mathrm{d}\mathbf{x} + \int_{\partial\kappa} \mathcal{H}(u_h^+, u_h^-, \mathbf{n}_\kappa) v^+ \,\mathrm{d}s \right\} = \sum_{\kappa\in\mathcal{T}_h} \int_\kappa fv\,\mathrm{d}\mathbf{x}, \tag{2.16}$$

for all $v_h \in V^p(\mathcal{T}_h)$.

We emphasize that the choice of the numerical flux function is *independent* of the finite element space employed. Indeed, the two key properties that the numerical flux function $\mathcal{H}(\cdot,\cdot,\cdot)$ should satisfy are:

1. *Consistency:* for each $\kappa \in \mathcal{T}_h$ we require that $\mathcal{H}(v,v,\mathbf{n}_\kappa)|_{\partial\kappa} = (\mathbf{b}v)\cdot\mathbf{n}_\kappa$.

2. *Conservation:* given any two neighbouring elements κ and κ' from the finite element mesh $\mathcal{T}_h$, at each point $\mathbf{x} \in \partial\kappa \cap \partial\kappa' \neq \emptyset$, noting that $\mathbf{n}_{\kappa'} = -\mathbf{n}_\kappa$, we have that $\mathcal{H}(v,w,\mathbf{n}_\kappa) = -\mathcal{H}(w,v,-\mathbf{n}_\kappa)$.

A classical and very natural choice is the upwind numerical flux, given by

$$\mathcal{H}(u_h^+,u_h^-,\mathbf{n}_\kappa)|_{\partial\kappa} = \begin{cases} \mathbf{b}\cdot\mathbf{n}_\kappa \lim_{s\to 0^+} u_h(\mathbf{x}-s\mathbf{b}) & \mathbf{x}\in\partial\kappa\backslash\partial_-\Omega, \\ \mathbf{b}\cdot\mathbf{n}_\kappa\, g(\mathbf{x}) & \mathbf{x}\in\partial\kappa\cap\partial_-\Omega, \end{cases} \tag{2.17}$$

for $\kappa \in \mathcal{T}_h$; indeed, upon substituting (2.17) into (2.16), we immediately recover the DGFEM scheme given in (2.14) through an integration by parts. For further details, and indeed for the construction of appropriate numerical flux functions for nonlinear first–order hyperbolic conservation laws, we refer to, e.g., [138, 181].

2.3 Discretization of second–order elliptic PDEs

The DGFEM discretization of general second–order elliptic PDEs is based on the following key steps:

1. Rewrite the underlying PDE as a first–order system of equations and derive an elemental weak formulation.

2. Introduce appropriate numerical flux functions in a similar fashion to that undertaken in the previous section; this gives rise to the so-called *flux formulation*.

3. Finally, the auxiliary variables introduced in step 1. may be eliminated to yield the underlying *primal formulation*.

To demonstrate each of these steps in a clear fashion, here we consider the model elliptic problem of the Poisson equation with essential boundary conditions, given

by: given $\Omega \subset \mathbb{R}^d$, $d \geq 1$, and $f \in L^2(\Omega)$, find $u \in H^1(\Omega)$ such that

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, && (2.18)\\ u &= g \quad \text{on } \partial\Omega, && (2.19) \end{aligned}$$

in the weak sense, i.e., we seek solution $u \in H^1(\Omega)$ with $u|_{\partial\Omega} = g$ (in the sense of trace) of the above problem posed in the weak form:

$$\int_\Omega \nabla u \cdot \nabla v \,\mathrm{d}\mathbf{x} = \int_\Omega f v \,\mathrm{d}\mathbf{x} \quad \text{for all } v \in H_0^1(\Omega). \tag{2.20}$$

Step 1. We rewrite (2.18) as the first–order system:

$$\mathbf{s} - \nabla u = 0, \qquad -\nabla \cdot \mathbf{s} = f. \tag{2.21}$$

Upon multiplication by test functions τ and v, and intergration by parts, the element-wise formulation is given by: for each $\kappa \in \mathcal{T}_h$, find $u|_\kappa \in H^1(\kappa)$ and $\mathbf{s}|_\kappa \in [L^2(\kappa)]^d$, such that $u|_{\partial\Omega} = g$ and

$$\begin{aligned} \int_\kappa \mathbf{s} \cdot \tau \,\mathrm{d}\mathbf{x} + \int_\kappa u \nabla \cdot \tau \,\mathrm{d}\mathbf{x} - \int_{\partial\kappa} u\tau \cdot \mathbf{n}_\kappa \,\mathrm{d}s &= 0,\\ \int_\kappa \mathbf{s} \cdot \nabla v \,\mathrm{d}\mathbf{x} - \int_{\partial\kappa} \mathbf{s} \cdot \mathbf{n}_\kappa v \,\mathrm{d}s &= \int_\kappa f v \,\mathrm{d}\mathbf{x}. \end{aligned}$$

Step 2. To arrive to the *flux formulation*, we introduce the numerical flux functions $\hat{u} = \hat{u}(u_h)$ and $\hat{\mathbf{s}} = \hat{\mathbf{s}}(u_h, \nabla_h u_h)$ which represent approximations to u and $\mathbf{s}$, respectively, on the boundary of each element κ in the computational mesh $\mathcal{T}_h$. Thereby, replacing $(u, \mathbf{s})$ by $(u_h, \mathbf{s}_h) \in V^p(\mathcal{T}_h) \times \boldsymbol{\Sigma}^p(\mathcal{T}_h)$, $\boldsymbol{\Sigma}^p(\mathcal{T}_h) = [V^p(\mathcal{T}_h)]^d$, and (v, τ) by $(v_h, \tau_h) \in V^p(\mathcal{T}_h) \times \boldsymbol{\Sigma}^p(\mathcal{T}_h)$, and summing over $\kappa \in \mathcal{T}_h$ gives rise to the DGFEM: find $(u_h, \mathbf{s}_h) \in V^p(\mathcal{T}_h) \times \boldsymbol{\Sigma}^p(\mathcal{T}_h)$ such that

$$\sum_{\kappa\in\mathcal{T}_h} \int_\kappa \mathbf{s}_h \cdot \tau_h \,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h} \int_\kappa u_h \nabla \cdot \tau_h \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{u}\tau_h^+ \cdot \mathbf{n}_\kappa \,\mathrm{d}s = 0, \tag{2.22}$$

$$\sum_{\kappa\in\mathcal{T}_h} \int_\kappa \mathbf{s}_h \cdot \nabla v_h \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \hat{\mathbf{s}} \cdot \mathbf{n}_\kappa v_h^+ \,\mathrm{d}s = \sum_{\kappa\in\mathcal{T}_h} \int_\kappa f v_h \,\mathrm{d}\mathbf{x} \tag{2.23}$$

for all $(v_h, \tau_h) \in V^p(\mathcal{T}_h) \times \boldsymbol{\Sigma}^p(\mathcal{T}_h)$.

The flux formulation given in (2.22), (2.23) involves the additional (auxiliary) unknowns $\mathbf{s}_h$; these may be eliminated in the following manner. Setting $\tau_h|_\kappa =$

$\nabla(v_h|_\kappa)$, $\kappa \in \mathcal{T}_h$, in (2.22) and integrating by parts gives

$$\sum_{\kappa\in\mathcal{T}_h}\int_\kappa \mathbf{s}_h\cdot\nabla v_h\,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h}\int_\kappa \nabla u_h\cdot\nabla v_h\,\mathrm{d}\mathbf{x} + \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa}(u_h^+-\hat{u})\nabla v_h^+\cdot\mathbf{n}_\kappa\,\mathrm{d}s \;=\; 0. \tag{2.24}$$

Inserting (2.24) into (2.23) gives rise to the primal DGFEM formulation: find $u_h \in V^p(\mathcal{T}_h)$ such that

$$\sum_{\kappa\in\mathcal{T}_h}\int_\kappa \nabla u_h\cdot\nabla v_h\,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa}(u_h^+-\hat{u})\nabla v_h^+\cdot\mathbf{n}_\kappa\,\mathrm{d}s - \sum_{\kappa\in\mathcal{T}_h}\int_{\partial\kappa}\hat{\mathbf{s}}\cdot\mathbf{n}_\kappa v_h^+\,\mathrm{d}s = \sum_{\kappa\in\mathcal{T}_h}\int_\kappa f v_h\,\mathrm{d}\mathbf{x} \tag{2.25}$$

for all $v_h \in V^p(\mathcal{T}_h)$.

Before we consider the choice of the numerical flux functions $\hat{u}$ and $\hat{\mathbf{s}}$, we first rewrite (2.25) in terms of integrals arising on each face in the underlying mesh $\mathcal{T}_h$. To this end, we introduce the following notation. We denote by $\mathcal{F}_h$ the set of open $(d-1)$–dimensional element faces associated with $\mathcal{T}_h$. Further, we write $\mathcal{F}_h = \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{B}}$, where $\mathcal{F}_h^{\mathcal{I}}$ denotes the set of all open $(d-1)$–dimensional element faces $F \in \mathcal{F}_h$ that are contained in Ω, and $\mathcal{F}_h^{\mathcal{B}}$ is the set of element boundary faces, i.e., $F \subset \partial\Omega$ for $F \in \mathcal{F}_h^{\mathcal{B}}$. The boundary $\partial\kappa$ of an element κ and the sets $\partial\kappa \setminus \partial\Omega$, $\partial\kappa \cap \partial\Omega$ will be identified in a natural way with the corresponding subsets of $\mathcal{F}_h$.

Next, we introduce some trace operators. Let κ_i and κ_j be two adjacent elements of $\mathcal{T}_h$ and let $\mathbf{x}$ be an arbitrary point on the interior face $F \in \mathcal{F}_h^{\mathcal{I}}$ given by $F = \partial\kappa_i \cap \partial\kappa_j$. We write $\mathbf{n}_{\kappa_i}$ and $\mathbf{n}_{\kappa_j}$ to denote the outward unit normal vectors on F, relative to $\partial\kappa_i$ and $\partial\kappa_j$, respectively. Furthermore, let v and $\mathbf{q}$ be scalar- and vector-valued functions, which are smooth inside each element κ_i and κ_j. Using the above notation, we write $(v_{\kappa_i}^+, \mathbf{q}_{\kappa_i}^+)$ and $(v_{\kappa_j}^+, \mathbf{q}_{\kappa_j}^+)$, we denote the traces of $(v, \mathbf{q})$ on F taken from within the interior of κ_i and κ_j, respectively. The averages of v and $\mathbf{q}$ at $\mathbf{x} \in F \in \mathcal{F}_h^{\mathcal{I}}$ are given by

$$\{\!\{v\}\!\} = \frac{1}{2}(v_{\kappa_i}^+ + v_{\kappa_j}^+), \quad \{\!\{\mathbf{q}\}\!\} = \frac{1}{2}(\mathbf{q}_{\kappa_i}^+ + \mathbf{q}_{\kappa_j}^+),$$

respectively. Similarly, the jumps of v and $\mathbf{q}$ at $\mathbf{x} \in F \in \mathcal{F}_h^{\mathcal{I}}$ are given by

$$[\![v]\!] = v_{\kappa_i}^+\,\mathbf{n}_{\kappa_i} + v_{\kappa_j}^+\,\mathbf{n}_{\kappa_j}, \quad [\![\mathbf{q}]\!] = \mathbf{q}_{\kappa_i}^+\cdot\mathbf{n}_{\kappa_i} + \mathbf{q}_{\kappa_j}^+\cdot\mathbf{n}_{\kappa_j},$$

respectively. On a boundary face $F \in \mathcal{F}_h^{\mathcal{B}}$, such that $F \subset \partial\kappa_i$, $\kappa_i \in \mathcal{T}_h$, we set

$$\{\!\{v\}\!\} = v_{\kappa_i}^+, \quad \{\!\{\mathbf{q}\}\!\} = \mathbf{q}_{\kappa_i}^+, \quad [\![v]\!] = v_{\kappa_i}^+ \mathbf{n}_{\kappa_i} \quad [\![\mathbf{q}]\!] = \mathbf{q}_{\kappa_i}^+ \cdot \mathbf{n}_{\kappa_i},$$

with $\mathbf{n}_{\kappa_i}$ denoting the unit outward normal vector on the boundary $\partial\Omega$. Here, we point out that the jump operator here is different compared to the upwind jump operator $\lfloor\cdot\rfloor$ defined in the previous section. Here the sign of the upwind jump $\lfloor\cdot\rfloor$ depends on the direction of the flow, whereas in the $[\![\cdot]\!]$ case it only depends on the element-numbering.

With this notation, we note that the following elementary identity holds:

$$\sum_{\kappa\in\mathcal{T}_h} \int_{\partial\kappa} \mathbf{q}^+ \cdot \mathbf{n}^+ \, v^+ \, \mathrm{d}s = \sum_{F\in\mathcal{F}_h} \int_F \{\!\{\mathbf{q}\}\!\} \cdot [\![v]\!] \, \mathrm{d}s + \sum_{F\in\mathcal{F}_h^{\mathcal{I}}} \int_F [\![\mathbf{q}]\!] \{\!\{v\}\!\} \, \mathrm{d}s, \tag{2.26}$$

cf. [16]. Exploiting (2.26), the primal formulation of the DGFEM (2.25) may be rewritten in the following equivalent manner: find $u_h \in V^p(\mathcal{T}_h)$ such that

$$\begin{aligned} \sum_{\kappa\in\mathcal{T}_h} \int_\kappa \nabla u_h \cdot \nabla v_h \, \mathrm{d}\mathbf{x} + \sum_{F\in\mathcal{F}_h} \int_F ([\![\hat{u} - u_h]\!] \cdot \{\!\{\nabla v_h\}\!\} - \{\!\{\hat{\mathbf{s}}\}\!\} \cdot [\![v_h]\!]) \, \mathrm{d}s \\ + \sum_{F\in\mathcal{F}_h^{\mathcal{I}}} \int_F (\{\!\{\hat{u} - u_h\}\!\} [\![\nabla v_h]\!] - [\![\hat{\mathbf{s}}]\!] \{\!\{v_h\}\!\}) \, \mathrm{d}s = \sum_{\kappa\in\mathcal{T}_h} \int_\kappa f v_h \, \mathrm{d}\mathbf{x} \end{aligned} \tag{2.27}$$

for all $v_h \in V^p(\mathcal{T}_h)$.

The choice of the numerical flux functions $\hat{u}$ and $\hat{\mathbf{s}}$ arising in the DGFEM (2.27) has been studied extensively: different choices of numerical flux functions lead to discontinuous Galerkin schemes with quite different consistency, stability, and convergence properties; for a review, we refer to [16]. In the interest of simplicity of the presentation, in this work, we consider one popular family of schemes, referred to as *interior penalty (IP) methods*. We stress, however, that the theoretical developments presented below are applicable to many other discontinuous Galerkin schemes. For IP methods, we select

$$\hat{u} = \hat{u}(u_h) = \begin{cases} \{\!\{u_h\}\!\} + \frac{1+\theta}{2} \mathbf{n}_F \cdot [\![u_h]\!] & \text{on } \ F \in \mathcal{F}_h^{\mathcal{I}}, \\ (1+\theta) u_h - \theta g & \text{on } \ F \in \mathcal{F}_h^{\mathcal{B}}, \end{cases}$$

$$\hat{\mathbf{s}} = \hat{\mathbf{s}}(u_h, \nabla_h u_h) = \begin{cases} \{\!\{\nabla_h u_h\}\!\} - \sigma [\![u_h]\!] & \text{on } \ F \in \mathcal{F}_h^{\mathcal{I}}, \\ \nabla u_h - \sigma (u_h - g) \mathbf{n} & \text{on } \ F \in \mathcal{F}_h^{\mathcal{B}}, \end{cases}$$

where $\theta \in [-1,1]$ and for $F \in \mathcal{F}_h^{\mathcal{I}}$, $F \subset \partial\kappa_i \cap \partial\kappa_j$, $\mathbf{n}_F = \mathbf{n}_{\kappa_i}$. Moreover, $\sigma : \mathcal{F}_h \mapsto \mathbb{R}$ is referred to as the *discontinuity penalization* function; the precise definition of σ depends on the local mesh size and local polynomial degree. In the current setting, i.e., assuming that the underlying simplicial mesh $\mathcal{T}_h$ is shape–regular and that the polynomial degree p is constant over $\mathcal{T}_h$, then the analysis undertaken in [125], for example, indicates that $\sigma = \mathcal{O}(p^2/h)$. The precise definition for general polytopic meshes and variable elemental polynomial degrees is a key question in this work and will be discussed in detail in Chapter 4.

Given the above definition of $\hat{u}$ and $\hat{\mathbf{s}}$, we deduce the following family of IP-DGFEMs: find $u_h \in V^p(\mathcal{T}_h)$ such that

$$\sum_{\kappa\in\mathcal{T}_h}\int_\kappa \nabla u_h\cdot\nabla v_h\,\mathrm{d}\mathbf{x} + \sum_{F\in\mathcal{F}_h}\int_F \left(-\{\!\{\nabla u_h\}\!\}\cdot[\![v_h]\!] + \theta\{\!\{\nabla v_h\}\!\}\cdot[\![u_h]\!]\right)\mathrm{d}s$$
$$+\sum_{F\in\mathcal{F}_h}\int_F \sigma[\![u_h]\!]\cdot[\![v_h]\!]\,\mathrm{d}s = \sum_{\kappa\in\mathcal{T}_h}\int_\kappa f v_h\,\mathrm{d}\mathbf{x} + \sum_{F\in\mathcal{F}_h^{\mathcal{B}}}\int_F g(\theta\nabla v_h\cdot\mathbf{n}+\sigma v_h)\,\mathrm{d}s \qquad (2.28)$$

for all $v_h \in V^p(\mathcal{T}_h)$. Selecting the parameter $\theta = 1$ gives rise to the so–called Nonsymmetric Interior Penalty (NIP) DGFEM, $\theta = 0$ is the Incomplete Interior Penalty (IIP) DGFEM, while setting $\theta = -1$ yields the Symmetric Interior Penalty (SIP) scheme.

On the basis of the schemes (2.16) and (2.28) the DGFEM discretization of general classes of second–order PDEs with so–called non-negative characteristic form may be defined; see Chapter 5 for details, cf., also, [125]. Before we embark on this topic, in the next chapter we first introduce the key technical results required to study the stability and convergence properties of DGFEMs defined over general mesh partitions.

2.4 PDEs with non-negative characteristic form

To highlight the versatility of the DGFEMs described above, we also consider the general class of *linear second order equations with non-negative characteristic form* in the form of respective initial/boundary value problems.

Given Ω is a bounded Lipschitz domain in $\mathbb{R}^d$, $d \geq 1$, we consider the following PDE: find u such that

$$-\nabla \cdot (a\nabla u) + \mathbf{b} \cdot \nabla u + cu \;=\; f \quad \text{in } \Omega. \tag{2.29}$$

Here, $a = \{a_{ij}\}_{i,j=1}^{d}$ with $a_{ij} \in L^\infty(\Omega)$ and $a_{ij} = a_{ji}$, for $i, j = 1, \ldots, d$, $\mathbf{b} = (b_1, \ldots, b_d) \in [W^{1,\infty}(\Omega)]^d$, $c \in L^\infty(\Omega)$ and $f \in L^2(\Omega)$. The PDE (2.29) is referred to as an equation with nonnegative characteristic form on the set $\Omega \subset \mathbb{R}^d$ if, at each $\mathbf{x}$ in $\bar{\Omega}$,

$$\sum_{i,j=1}^{d} a_{ij}(\mathbf{x})\xi_i\xi_j \geq 0 \tag{2.30}$$

for any vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ in $\mathbb{R}^d$.

This class of equations includes second–order elliptic and parabolic PDEs, ultra-parabolic equations, first–order hyperbolic problems, the Kolmogorov–Fokker–Planck equations of Brownian motion (cf. [29], for example), the equations of boundary layer theory in hydrodynamics, and various other degenerate equations. More generally, according to a well-known result of Hörmander [151], second–order hypoelliptic operators have nonnegative characteristic form at each point of the domain Ω, after possible multiplication by -1, so they all into this category.

To supplement (2.29) with suitable boundary conditions, following [151, 126], we first subdivide the boundary $\partial\Omega$ of the computational domain Ω into appropriate subsets. To this end, we let

$$\partial_0\Omega = \Big\{\mathbf{x} \in \partial\Omega : \sum_{i,j=1}^{d} a_{ij}(\mathbf{x})n_i n_j > 0\Big\},$$

where $\mathbf{n} = (n_1, \ldots, n_d)$ denotes the unit outward normal vector to $\partial\Omega$. Loosely speaking, we may think of $\partial_0\Omega$ as being the 'elliptic' portion of the boundary $\partial\Omega$. On the 'hyperbolic' portion of the boundary $\partial\Omega\backslash\partial_0\Omega$, we define the inflow and outflow boundaries $\partial_-\Omega$ and $\partial_+\Omega$, respectively, in the standard manner:

$$\begin{aligned} \partial_-\Omega &= \{\mathbf{x} \in \partial\Omega\backslash\partial_0\Omega : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0\}, \\ \partial_+\Omega &= \{\mathbf{x} \in \partial\Omega\backslash\partial_0\Omega : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) > 0\}. \end{aligned}$$

If $\partial_0\Omega$ is nonempty, we shall further divide it into disjoint subsets $\partial\Omega_{\mathrm{D}}$ and $\partial\Omega_{\mathrm{N}}$ whose union is $\partial_0\Omega$, with $\partial\Omega_{\mathrm{D}}$ nonempty and relatively open in $\partial\Omega$. It is evident

from these definitions that $\partial\Omega = \partial\Omega_{\rm D} \cup \partial\Omega_{\rm N} \cup \partial_-\Omega \cup \partial_+\Omega$. Assuming the (physically reasonable) hypothesis that $\mathbf{b}\cdot\mathbf{n} \geq 0$ on $\partial\Omega_{\rm N}$ whenever $\partial\Omega_{\rm N}$ is nonempty, we supplement (2.29) with the following, respectively, Dirichlet and Neumann boundary conditions:

$$u = g_{\rm D} \quad \text{on } \partial\Omega_{\rm D} \cup \partial_-\Omega, \qquad \mathbf{n}\cdot(a\nabla u) = g_{\rm N} \quad \text{on } \partial\Omega_{\rm N}. \tag{2.31}$$

Additionally, we assume that the following positivity hypothesis holds: there exists a constant vector $\boldsymbol{\xi} \in \mathbb{R}^d$ such that

$$c(\mathbf{x}) - \frac{1}{2}\,\nabla\cdot\mathbf{b}(\mathbf{x}) + \mathbf{b}(\mathbf{x})\cdot\boldsymbol{\xi} \geq \gamma_0 \ \text{ a.e. } \mathbf{x} \in \Omega, \tag{2.32}$$

where $\gamma_0 > 0$ is a constant. The well–posedness of the boundary value problem (2.29), (2.31), in the case of homogeneous boundary conditions, has been studied in [126], cf. also [151].

It is possible to introduce an IP-DGFEM discretization for the general (2.29), (2.31), thereby, treating numerically this very general class of equations in a stable fashion.

We consider a mesh $\mathcal{T}_h$, which is subordinate to the mixed boundary conditions (2.31), in the sense that the set of boundary faces $\mathcal{F}_h^{\mathcal{B}}$ of $\mathcal{T}_h$ can be subdivided as $\mathcal{F}_h^{\mathcal{B}} = \mathcal{F}_h^{\mathcal{D}} \cup \mathcal{F}_h^{\mathcal{N}} \cup \mathcal{F}_h^- \cup \mathcal{F}_h^+$, covering (almost everywhere) the Dirichlet, Neumann, inflow and outflow parts of boundary, respectively. We define the IP-DGFEMs: find $u_h \in V^p(\mathcal{T}_h)$ such that

$$\begin{aligned}
&\int_\Omega \big(a\nabla_h u_h\cdot\nabla_h v_h + (\mathbf{b}\cdot\nabla_h u_h)v_h + c u_h\, v_h\big)\,\mathrm{d}\mathbf{x} \\
&+ \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \big(-\{\!\{a\nabla u_h\}\!\}\cdot[\![v_h]\!] + \theta\{\!\{a\nabla v_h\}\!\}\cdot[\![u_h]\!] + \sigma[\![u_h]\!]\cdot[\![v_h]\!]\big)\,\mathrm{d}s \\
&- \sum_{F\in\mathcal{F}_h^-\setminus\mathcal{F}_h^{\mathcal{B}}} \int_F \mathbf{b}\cdot\mathbf{n}_\kappa\, \lfloor u_h\rfloor v_h^+\,\mathrm{d}s - \sum_{F\in\mathcal{F}_h^-\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \mathbf{b}\cdot\mathbf{n}_\kappa\, u_h^+ v_h^+\,\mathrm{d}s \\
=\ &\int_\Omega f\, v_h\,\mathrm{d}\mathbf{x} + \sum_{F\in\mathcal{F}_h^{\mathcal{D}}} \int_F g_{\rm D}(\theta\nabla v_h\cdot\mathbf{n} + \sigma v_h)\,\mathrm{d}s \\
&- \sum_{F\in\mathcal{F}_h^-\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \mathbf{b}\cdot\mathbf{n}_\kappa\, g_{\rm D} v_h^+\,\mathrm{d}s + \sum_{F\in\mathcal{F}_h^{\mathcal{N}}} \int_F g_{\rm N} v_h\,\mathrm{d}s
\end{aligned}$$

for all $v_h \in V^p(\mathcal{T}_h)$.

Chapter 3

Polynomial Approximation and Inverse Estimates

In this chapter we develop the key mathematical tools needed to study the stability and convergence properties of *hp*–version DGFEMs; these estimates will be exploited in the following chapters for IP-DGFEM discretizations, with the ultimate goal of tackling general second–order PDEs with non-negative characteristic form. While results of this type are readily available within the literature for standard element types, e.g., simplices and tensor product elements, cf., for example, [24, 25, 69, 147, 167], in this chapter we concentrate on the extension of these bounds to general meshes consisting of polytopic elements. A key issue in this setting is that general shape–regular polytopic meshes may, under refinement, possess elements with $(d-k)$–dimensional facets, $k = 1, 2, \ldots, d-1$, which degenerate as the mesh size tends to zero. Therefore, care must be taken to ensure that the resulting inverse estimates and polynomial approximation results are sensitive to this type of degeneracy. The key approach adopted here is to exploit known results for standard elements, both within an L^2– and L^∞–setting, and to take the minimum of the resulting bounds, cf. [61, 59, 58, 7]. In this way, bounds which are optimal in both the *h*–version and *p*–version setting may be deduced, which directly account for $(d-k)$–dimensional facet degeneration, $k = 1, 2, \ldots, d-1$.

Firstly, we begin by introducing the classes of meshes which may be admitted in the analysis presented below, cf. Section 3.1. Under these assumptions in Sections 3.2 and 3.3 we derive *hp*–version inverse and approximation results, respectively.

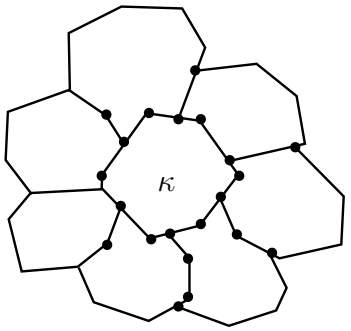


FIGURE 3.1: Polygonal element κ, $\kappa \in \mathcal{T}_h$, and its face–wise neighbours; hanging nodes are highlighted with $\bullet$.

3.1 Mesh assumptions

We introduce a very general class of computational meshes consisting of polytopic elements, satisfying some technical assumptions. The notation introduced here will be employed throughout the rest of this work.

To this end, we let $\mathcal{T}_h$ be a subdivision of the computational domain $\Omega \subset \mathbb{R}^d$, $d > 1$, into disjoint open *polygonal/polyhedral* (polytopic) elements κ constructed in such a manner that the union of the closures of the elements $\kappa \in \mathcal{T}_h$ forms a covering of the closure of Ω, i.e., $\bar{\Omega} = \cup_{\kappa \in \mathcal{T}_h} \bar{\kappa}$. Furthermore, we denote by h_κ the diameter of $\kappa \in \mathcal{T}_h$, i.e., $h_\kappa := \mathrm{diam}(\kappa)$.

From a mesh adaptation point of view DGFEMs are advantageous in the sense that they can naturally handle meshes which contain irregular/hanging nodes. With this in mind, we allow $\mathcal{T}_h$ to consist of general elements which may possess several hanging nodes on their $(d-k)$–dimensional facets, $k = 1, 2, \ldots, d-1$, cf. Figure 3.1. As noted above, the stability and approximation results developed in this chapter rely on employing results for standard element shapes; in fact, here we shall rely on *hp*–version bounds for simplices. For this reason, we introduce the notion of both element interfaces and element faces; the latter being assumed to be simplices in $\mathbb{R}^{d-1}$.

To this end, and to facilitate the presence of hanging nodes, we define the *interfaces* of the computational mesh $\mathcal{T}_h$ to be the intersection of the $(d-1)$–dimensional facets of neighbouring elements. In the two–dimensional setting, i.e., when $d = 2$, the interfaces of $\mathcal{T}_h$ are simply piecewise linear line segments, i.e., they consist of a set of $(d-1)$–dimensional simplices. However, in general for $d = 3$, (or indeed $d \geq 3$), the interfaces of $\mathcal{T}_h$ will consist of general polygonal surfaces in $\mathbb{R}^3$ (or

polyhedral surfaces in $\mathbb{R}^d$, respectively). Thereby, we assume that each planar section of each interface of an element $\kappa \in \mathcal{T}_h$ may be subdivided into a set of co-planar triangles ($(d-1)$–dimensional simplices). We refer to these $(d-1)$–dimensional simplices, whose union form the interfaces of $\mathcal{F}_h$, as *faces*.

In the following sections we outline the key assumptions required to be satisfied by the computational mesh $\mathcal{T}_h$ in order to derive suitable inverse inequalities and approximation results for general polytopic elements. Firstly, however, we introduce the following assumption, which guarantees that the number of faces each element possesses remains bounded under mesh refinement; we shall return to this issue in Chapter 4 when we consider the coercivity of the IP-DGFEM.

Assumption 3.1.1. *For each element $\kappa \in \mathcal{T}_h$, we define*

$$C_\kappa = card\Big\{F \in \mathcal{F}_h : F \subset \partial\kappa\Big\}.$$

In the following we assume there exists a positive constant C_F, independent of the mesh parameters, such that

$$\max_{\kappa\in\mathcal{T}_h} C_\kappa \leq C_F.$$

3.2 Inverse estimates

One of the key mathematical tools required for the analysis of DGFEMs are inverse inequalities; results of this type for standard element shapes are well–known the literature, cf., for example, [167, 183, 185].

Lemma 3.1. *Given T is a simplex in $\mathbb{R}^d$, $d = 2, 3$, we write $F \subset \partial T$ to denote one of its faces. Then, for $v \in \mathcal{P}_p(T)$, the following inverse inequalities hold*

$$\|v\|^2_{L^2(F)} \leq C_{\mathrm{inv},1} p^2 \frac{|F|}{|T|} \|v\|^2_{L^2(T)}, \tag{3.1}$$

$$\|v\|^2_{L^\infty(T)} \leq C_{\mathrm{inv},2} \frac{p^{2d}}{|T|} \|v\|^2_{L^2(T)}, \tag{3.2}$$

$$\|\nabla v\|^2_{L^2(T)} \leq C_{\mathrm{inv},3} \frac{p^4}{h_T^2} \|v\|^2_{L^2(T)}, \tag{3.3}$$

where $C_{\mathrm{inv},i}$, $i = 1, 2, 3$, are positive constants which are independent of v, p, and h_T. $C_{\mathrm{inv},3}$ depends on the shape regularity of T.

Proof. The detailed proof of (3.1) can be found in [185], and is based on solving the eigenvalue problem for polynomial functions. We point out that the exact bound is:

$$\|v\|^2_{L^2(F)} \leq \frac{(p+1)(p+d)}{d}\frac{|F|}{|T|}\|v\|^2_{L^2(T)}. \tag{3.4}$$

The proof of (3.2) can be found in [167]. For (3.3), the proof can be found in [167]. Here, we emphasize that constant $C_{\text{inv},3}$ depends on the shape regularity of simplex T. □

We shall consider the generalization of (3.1) and (3.3) to general meshes consisting of polytopic elements. We remark that (3.1) is required to establish stability of IP-DGFEM approximations of second–order elliptic PDEs, cf. Lemma 4.2, while (3.3) will be utilized to determine an inf-sup condition in the presence of first–order transport terms, cf. Theorem 5.5. In order to generalize (3.1) to general polytopic elements κ, $\kappa \in \mathcal{T}_h$, we first introduce the following family of (overlapping) simplices associated with each face $F \subset \partial\kappa$. Note that this is precisely the reason why we require that each face F is a $d-1$-dimensional simplex.

Definition 3.2. For each element κ in the computational mesh $\mathcal{T}_h$, we define the family $\mathcal{F}_\flat^\kappa$ of all possible d–dimensional simplices contained in κ and having at least one face in common with κ. Moreover, we write $\kappa_\flat^F$ to denote a simplex belonging to $\mathcal{F}_\flat^\kappa$ which shares with $\kappa \in \mathcal{T}_h$ the specific face $F \subset \partial\kappa$.

With the above definition, we may now employ (3.1) directly to deduce the corresponding inverse estimate on a general polytopic element. To this end, given $\kappa \in \mathcal{T}_h$ and the face $F \in \mathcal{F}_h$ such that $F \subset \partial\kappa$, consider $\kappa_\flat^F \in \mathcal{F}_\flat^\kappa$ given in Definition 3.2. Then, for $v \in \mathcal{P}_p(\kappa)$, applying (3.1) on $\kappa_\flat^F$, we immediately deduce that

$$\|v\|^2_{L^2(F)} \leq C_{\text{inv},1}p^2\frac{|F|}{|\kappa_\flat^F|}\|v\|^2_{L^2(\kappa_\flat^F)} \leq C_{\text{inv},1}p^2\frac{|F|}{|\kappa_\flat^F|}\|v\|^2_{L^2(\kappa)}, \tag{3.5}$$

where $C_{\text{inv},1}$ is a positive contant, independent of v, $|F|$, $|\kappa_\flat^F|$, and p.

Clearly, the choice of $\kappa_\flat^F$ is not unique; thereby, we may select $\kappa_\flat^F$ to have the largest possible measure $|\kappa_\flat^F|$. Hence, on the basis of (3.5), the following inverse inequality holds:

$$\|v\|^2_{L^2(F)} \leq C_{\text{inv},1}p^2\frac{|F|}{\sup_{\kappa_\flat^F \subset \kappa}|\kappa_\flat^F|}\|v\|^2_{L^2(\kappa)}. \tag{3.6}$$

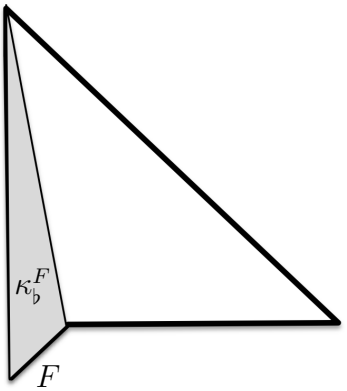


FIGURE 3.2: Illustration of the quadrilateral in Example 3.1

We point out that for a fixed element size h_κ, the inverse inequality (3.6) is sharp with respect to the polynomial degree p, cf. [167]. However, for fixed polynomial order p, (3.6) lacks sharpness with respect to $(d-k)$–dimensional facet degeneration, $k = 1, \ldots, d-1$; or more precisely, it is not sensitive to the magnitude of the face measure relative to the measure of the polytopic element κ. To illustrate this in a clear manner, we consider the two–dimensional example presented in [61].

Example 3.1. *In order to demonstrate the lack of sharpness of the inverse inequality* (3.6) *with respect to one of its lower–dimensional facets degenerating, we consider the quadrilateral domain* κ *given by*

$$\kappa := \{(x,y) \in \mathbb{R}^2 : x > 0, y > 0, x+y < 1\} \cup \{(x,y) \in \mathbb{R}^2 : x > 0, y \leq 0, x-y < \epsilon\},$$

for some $\epsilon > 0$*, cf. Figure 3.2. Given* $v \in \mathcal{P}_p(\kappa)$*, let* $F := \{(x,y) \in \mathbb{R}^2 : x-y = \epsilon\}$*, then exploiting (3.6) gives*

$$\|v\|^2_{L^2(F)} \leq C_{\mathrm{inv},1} \frac{\sqrt{2}p^2\epsilon}{|\kappa_\flat^F|} \|v\|^2_{L^2(\kappa)}, \tag{3.7}$$

where

$$\kappa_\flat^\kappa := \{(x,y) \in \mathbb{R}^2 : x > 0,\ x + \epsilon y < \epsilon,\ x - y < \epsilon\}.$$

Noting that $|\kappa_\flat^F| = \epsilon(1+\epsilon)/2$*, inequality (3.7) becomes*

$$\|v\|^2_{L^2(F)} \leq C_{\mathrm{inv},1} \frac{2\sqrt{2}p^2}{1+\epsilon} \|v\|^2_{L^2(\kappa)}.$$

Hence, if we let $\epsilon \to 0$*, the left-hand side* $\|v\|^2_{L^2(F)} \to 0$*, whereas the right-hand side* $\frac{2\sqrt{2}p^2}{1+\epsilon}\|v\|^2_{L^2(\kappa)} \to 2\sqrt{2}p^2\|v\|^2_{L^2(\kappa)} \neq 0$ *in general.*

The above example clearly indicates that the inverse inequality (3.6) may not be sharp, with respect to element facets of degenerating measure. In the context of employing such a bound to deduce the stability of a given DGFEM approximation of a given second–order elliptic PDE, cf. Section 4.2, will typically lead to an excessively large penalization term within the underlying scheme; this in turn may result in ill conditioning of the resulting system of equations.

To rectify this issue, we proceed by deriving an alternative inverse inequality, under suitable mesh assumptions, based on first noting that since $F \subset \partial\kappa_\flat^F$, by definition, we have that

$$\|v\|^2_{L^2(F)} \leq |F|\|v\|^2_{L^\infty(\kappa_\flat^F)}. \tag{3.8}$$

In order to bound the right-hand side of (3.8), we need to introduce some additional requirements on the elements $\kappa \in \mathcal{T}_h$. These are based on the following result which represents the generalization of Lemma 3.7 in [104].

Lemma 3.3. *Let K be a shape-regular simplex. Then, for each $v \in \mathcal{P}_p(K)$, there exists a simplex $\hat{\kappa} \subset K$, having the same shape as K and faces parallel to the faces of K, with* $\mathrm{dist}(\partial\hat{\kappa}, \partial K) > C_{as}\,\mathrm{diam}(K)/p^2$*, where C_{as} is a positive constant, independent of v, K and p, such that*

$$\|v\|_{L^2(\hat{\kappa})} \geq \frac{1}{2}\|v\|_{L^2(K)}.$$

Proof. For simplicity, we present here the proof for triangles, as the general case follows analogously, see the proof of Lemma 3.7 in [104] for more details.

We first consider the case of the reference triangle K of vertices $(0,0), (1,0)$, and $(0,1)$. We consider a splitting of K into 4 disjoint parts as follows, cf. Figure 3.3. We let $\hat{\kappa}$ be the triangle having same shape as K, faces parallel to K, and $\mathrm{dist}(\partial\hat{\kappa}, \partial K) = \delta$. Then, we also split $K\backslash\hat{\kappa}$ into 3 disjoint parts $\{\hat{\kappa}_i\}_{i=1}^3$. For $\hat{\kappa}_1$, we have

$$\begin{aligned}
\|v\|^2_{L^2(\hat{\kappa}_1)} &= \int_0^\delta \int_0^{\frac{1}{2}} v^2(x,y)\,\mathrm{d}x\,\mathrm{d}y + \int_\delta^{\frac{1}{2}} \int_0^\delta v^2(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&\leq \int_0^{\frac{1}{2}} \delta\|v(x,\cdot)\|^2_{L^\infty(0,\delta)}\,\mathrm{d}x + \int_\delta^{\frac{1}{2}} \delta\|v(\cdot,y)\|^2_{L^\infty(0,\delta)}\,\mathrm{d}y \\
&\leq \int_0^{\frac{1}{2}} \delta\|v(x,\cdot)\|^2_{L^\infty(0,\frac{1}{2})}\,\mathrm{d}x + \int_0^{\frac{1}{2}} \delta\|v(\cdot,y)\|^2_{L^\infty(0,\frac{1}{2})}\,\mathrm{d}y \\
&\leq \delta C_{\mathrm{inv},2} p^2 \|v\|^2_{L^2(A_1)},
\end{aligned} \tag{3.9}$$

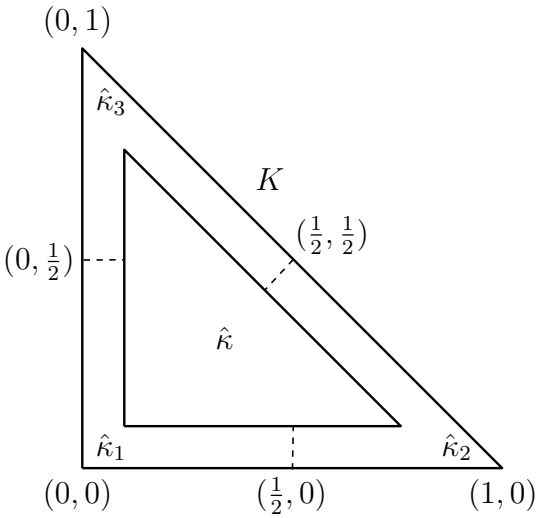


FIGURE 3.3: Splitting triangle K into $\hat{\kappa}$ and $\{\hat{\kappa}_i\}_{i=1}^3$.

where $A_1 = (0, \frac{1}{2})^2$, and in the last inequality we have used the one-dimensional analogue of the inverse inequality (3.2).

For $\hat{\kappa}_2$, we make the (linear) change of variables $(x, y) \to (\tilde{x}, \tilde{y})$, where $\tilde{x} = x + y$ and $\tilde{y} = y$. Then, we have

$$\begin{aligned}
\|v\|^2_{L^2(\hat{\kappa}_2)} &= \int_0^{\delta} \int_{\frac{1}{2}}^1 v^2(\tilde{x} - \tilde{y}, \tilde{y}) \, \mathrm{d}\tilde{x} \, \mathrm{d}\tilde{y} + \int_{\delta}^{\frac{1}{2}} \int_{1-\delta}^1 v^2(\tilde{x} - \tilde{y}, \tilde{y}) \, \mathrm{d}\tilde{x} \, \mathrm{d}\tilde{y} \\
&\leq \int_{\frac{1}{2}}^1 \delta \|v(\tilde{x} - \cdot, \cdot)\|^2_{L^\infty(0,\delta)} \, \mathrm{d}\tilde{x} + \int_{\delta}^{\frac{1}{2}} \delta \|v(\cdot - \tilde{y}, \tilde{y})\|^2_{L^\infty(1-\delta,1)} \, \mathrm{d}\tilde{y} \\
&\leq \int_{\frac{1}{2}}^1 \delta \|v(\tilde{x} - \cdot, \cdot)\|^2_{L^\infty(0,\frac{1}{2})} \, \mathrm{d}\tilde{x} + \int_0^{\frac{1}{2}} \delta \|v(\cdot - \tilde{y}, \tilde{y})\|^2_{L^\infty(\frac{1}{2},1)} \, \mathrm{d}\tilde{y} \\
&\leq \delta C_{\mathrm{inv},2} p^2 \|v\|^2_{L^2(A_2)}, \qquad (3.10)
\end{aligned}$$

where A_2 denotes the parallelogram with vertices $(\frac{1}{2}, 0), (1, 0), (\frac{1}{2}, \frac{1}{2}), (0, \frac{1}{2})$.

For $\hat{\kappa}_3$, we make the (linear) change of variables $(x, y) \to (\tilde{x}, \tilde{y})$, where $\tilde{x} = x$ and $\tilde{y} = x + y$. Then, completely analogously to the case of $\hat{\kappa}_2$, we obtain

$$\|v\|^2_{L^2(\hat{\kappa}_3)} \leq \delta C_{\mathrm{inv},2} p^2 \|v\|^2_{L^2(A_3)}, \qquad (3.11)$$

where A_3 denotes the parallelogram with vertices $(\frac{1}{2}, 0), (\frac{1}{2}, \frac{1}{2}), (0, 1), (0, \frac{1}{2})$. Combining (3.9), (3.10), and (3.11), we deduce

$$\|v\|^2_{L^2(K \setminus \hat{\kappa})} \leq 3\delta C_{\mathrm{inv},2} p^2 \|v\|^2_{L^2(K)}.$$

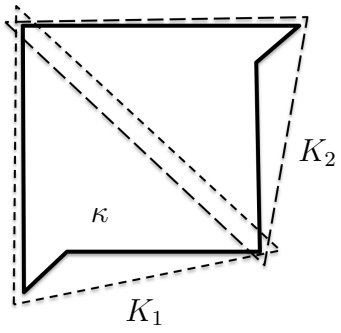


FIGURE 3.4: Illustration of quadrilateral in Definition 3.4

Selecting $\delta = (4C_{\text{inv},2}p^2)^{-1}$, we have $\|v\|^2_{L^2(K\setminus\hat{\kappa})} \leq \frac{3}{4}\|v\|^2_{L^2(K)}$. Using this, we have, respectively,

$$\|v\|^2_{L^2(\hat{\kappa})} = \|v\|^2_{L^2(K)} - \|v\|^2_{L^2(K\setminus\hat{\kappa})} \geq \|v\|^2_{L^2(K)} - \frac{3}{4}\|v\|^2_{L^2(K)} = \frac{1}{4}\|v\|^2_{L^2(K)}.$$

For general physical triangles, by using scaling arguments, it is easy to see that there must *exist* a triangle $\hat{\kappa}$ having same shape as K, faces parallel to K, with $\text{dist}(\partial\hat{\kappa}, \partial K) \geq C_{as}\,\text{diam}(K)/p^2$, satisfying the required statement. □

Motivated by the result of Lemma 3.3 we introduce the following definition.

Definition 3.4. An element $\kappa \in \mathcal{T}_h$ is said *p-coverable* with respect to $p \in \mathbb{N}$ if there exists a set of m_κ shape-regular simplices K_i, $i = 1, \ldots, m_\kappa$, $m_\kappa \in \mathbb{N}$, such that

$$\text{dist}(\kappa, \partial K_i) < C_{as}\frac{\text{diam}(K_i)}{p^2}, \qquad \text{and} \qquad |K_i| \geq c_{as}|\kappa| \tag{3.12}$$

for all $i = 1, \ldots, m_\kappa$, where C_{as} and c_{as} are positive constants, independent of κ and $\mathcal{T}_h$.

Following [61], in Figure 3.4 we present a polygonal element κ in $\mathbb{R}^2$ which may be covered by two triangles K_1 and K_2, i.e., $m_\kappa = 2$. We point out that Definition 3.4 admits very general polytopic elements $\kappa \in \mathcal{T}_h$ which may contain $(d-k)$–dimensional facets, $k = 1, \ldots, d-1$, whose measure is arbitrarily small, relative to the measure of κ itself. We point out that (3.12) can be considered as a restriction on the polynomial degree p for the proposed DGFEM over polytopic elements. Returning to Example 3.1, we note that the quadrilateral element κ depicted in Figure 3.2 is p-coverable when $\epsilon < C_{as}/p^2$ for some constant $C_{as} > 0$.

Equipped with (3.6), (3.8), Lemma 3.3, and Definition 3.4, we are now in a position to present *hp*–version inverse inequality for a general polytopic elements which directly accounts for elemental interface degeneration.

Lemma 3.5. *Let $\kappa \in \mathcal{T}_h$, $F \subset \partial\kappa$ denote one of its faces. Then, for each $v \in \mathcal{P}_p(\kappa)$, the following inverse inequality holds*

$$\|v\|^2_{L^2(F)} \leq C_{\mathrm{INV}}(p,\kappa,F)p^2\frac{|F|}{|\kappa|}\|v\|^2_{L^2(\kappa)}, \tag{3.13}$$

where

$$C_{\mathrm{INV}}(p,\kappa,F) := \begin{cases} C_{\mathrm{inv},4}\min\left\{\dfrac{|\kappa|}{\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|}, p^{2(d-1)}\right\}, & \text{if } \kappa \text{ is } p\text{-coverable} \\ C_{\mathrm{inv},1}\dfrac{|\kappa|}{\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|}, & \text{otherwise,} \end{cases} \tag{3.14}$$

and $\kappa_\flat^F \in \mathcal{F}_\flat^\kappa$ as in Definition 3.2. Furthermore, $C_{\mathrm{inv},1}$ and $C_{\mathrm{inv},4}$ are positive constants which are independent of $|\kappa|/\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|$, $|F|$, p, and v.

Proof. If κ is not p-coverable, then the above inverse inequality follows immediately from the bound (3.6). Thereby, we now consider the case when κ is p-coverable; recalling Definition 3.4, the element κ may be covered by shape-regular simplices K_i, $i = 1,\ldots,m_\kappa$. Hence, given $\kappa_\flat^F \in \mathcal{F}_\flat^\kappa$, $F \subset \partial\kappa$, cf. Definition 3.2, we note that

$$\kappa_\flat^F \subset \kappa \subset \cup_{i=1}^{m_\kappa} K_i,$$

with $|K_i| \geq c_{as}|\kappa|$, $i = 1,\ldots,m_\kappa$.

Employing the inverse estimate (3.2) on each K_i, $i = 1,\ldots,m_\kappa$, together with Definition 3.4, we deduce that

$$\begin{aligned} \|v\|^2_{L^\infty(\kappa_\flat^F)} &\leq \max_{i=1,\ldots,m_\kappa}\|v\|^2_{L^\infty(K_i)} \\ &\leq C_{\mathrm{inv},2}p^{2d}\max_{i=1,\ldots,m_\kappa}\frac{\|v\|^2_{L^2(K_i)}}{|K_i|} \\ &\leq \frac{C_{\mathrm{inv},2}}{c_{as}}\frac{p^{2d}}{|\kappa|}\max_{i=1,\ldots,m_\kappa}\|v\|^2_{L^2(K_i)}. \end{aligned} \tag{3.15}$$

We now define $\hat{\kappa}_i \subset K_i$ to denote the simplex relative to K_i defined in Lemma 3.3; hence, exploiting Lemma 3.3 and Definition 3.4, and noting by construction that

$\hat{\kappa}_i \subset \kappa \cap K_i \subset K_i$ and $K_i \cap \kappa \subset \kappa$, for each $i = 1, \ldots, m_\kappa$, we deduce that

$$\frac{1}{4}\|v\|^2_{L^2(K_i)} \leq \|v\|^2_{L^2(\hat{\kappa}_i)} \leq \|v\|^2_{L^2(K_i \cap \kappa)} \leq \|v\|^2_{L^2(\kappa)}. \tag{3.16}$$

Combining (3.15) and (3.16), we arrive at the inequality

$$\|v\|^2_{L^\infty(\kappa_\flat^F)} \leq \frac{4\, C_{\mathrm{inv},2}}{c_{as}} \frac{p^{2d}}{|\kappa|} \|v\|^2_{L^2(\kappa)}. \tag{3.17}$$

Inserting (3.22) into (3.8) gives

$$\|v\|^2_{L^2(F)} \leq \frac{4\, C_{\mathrm{inv},2}}{c_{as}} \frac{p^{2d}|F|}{|\kappa|} \|v\|^2_{L^2(\kappa)}. \tag{3.18}$$

Taking the minimum between (3.6) and (3.18), we deduce the desired result, with a positive constant $C_{\mathrm{inv},4} = \max\{C_{\mathrm{inv},1}, 4\, C_{\mathrm{inv},2}/c_{as}\}$. □

Remark 3.6. We point that for a fixed mesh size the inverse inequality stated in (3.13) is sharp with respect to the polynomial degree p; indeed, as $p \to \infty$ the minimum in (3.14) will be equal to $|\kappa|/\sup_{\kappa_\flat^F \subset \kappa} |\kappa_\flat^F|$. Moreover, (3.13) is sensitive with respect to the $(d-k)$–dimensional facet degeneration, $k = 1, \ldots, d-1$. Indeed, recalling Example 3.1, we observe that the left– and right–hand sides of (3.13) degenerate at the same rate as $\epsilon \to 0$.

We end this section by presenting a further inverse inequality which provides a bound on the $H^1(\kappa)$–norm of a polynomial function v, $\kappa \in \mathcal{T}_h$, with respect to the $L^2(\kappa)$–norm of v, cf. (3.3) for the case of simplices; this result will be required to deduce the inf-sup estimate derived in Theorem 5.5. *In this setting, it is necessary to assume shape–regularity of the polytopic mesh $\mathcal{T}_h$.*

Assumption 3.2.1. *We assume that the subdivision $\mathcal{T}_h$ is shape–regular in the sense of [70], i.e., there exists a positive constant C_{r}, independent of the mesh parameters, such that*

$$\forall \kappa \in \mathcal{T}_h, \quad \frac{h_\kappa}{\rho_\kappa} \leq C_{\mathrm{r}}.$$

with ρ_κ denoting the diameter of the largest ball contained in κ.

In addition to Assumption 3.2.1, we also require that each non p-coverable element also admits a shape–regular simplicial sub-partition; more precisely, we require that the following assumption holds.

Assumption 3.2.2. *We assume that each polytopic element which is not p-coverable admits a sub-partition into at most n_κ, $n_\kappa \in \mathbb{N}$, shape-regular simplices $\mathfrak{s}_i$, $i = 1, 2, \ldots, n_\kappa$, such that*

$$|\mathfrak{s}_i| \geq c_{\mathfrak{s}}|\kappa|, \qquad i = 1, \ldots, n_\kappa,$$

where $c_{\mathfrak{s}}$ is a positive constant, independent of κ and $\mathcal{T}_h$.

We note that the above assumptions have been commonly used in other polygonal discretization methods, see [85, 86, 63].

Lemma 3.7. *Given Assumptions 3.2.1 and 3.2.2, for $v \in \mathcal{P}_p(\kappa)$, the following inverse inequality holds*

$$\|\nabla v\|^2_{L^2(\kappa)} \leq C_{\mathrm{inv},5}\frac{p^4}{h_\kappa^2}\|v\|^2_{L^2(\kappa)}, \tag{3.19}$$

where $C_{\mathrm{inv},5}$ is a positive constant, which is independent of h_κ and p, but depends on the shape regularity of the covering of κ if κ is p-coverable, or the sub-partition of κ if κ is not p-coverable.

Proof. Let us first consider the case when κ is not p-coverable; then recalling the sub-triangulation introduced in Assumption 3.2.2, together with the inverse inequality (3.3), we note that

$$\|\nabla v\|^2_{L^2(\kappa)} = \sum_{i=1}^{n_\kappa}\|\nabla v\|^2_{L^2(\mathfrak{s}_i)} \leq C_{\mathrm{inv},3}p^4\sum_{i=1}^{n_\kappa}h_{\mathfrak{s}_i}^{-2}\|v\|^2_{L^2(\mathfrak{s}_i)}, \tag{3.20}$$

where $h_{\mathfrak{s}_i} = \mathrm{diam}(\mathfrak{s}_i)$, $i = 1, \ldots, n_\kappa$. Recalling the shape-regularity of the mesh $\mathcal{T}_h$, cf. Assumption 3.2.1, together with Assumption 3.2.2 and the trivial relation $h_\kappa^d \geq |\kappa| \geq \rho_\kappa^d$, we note that the following inequalities hold for $i = 1, \ldots, n_\kappa$:

$$h_{\mathfrak{s}_i}^d \geq |\mathfrak{s}_i| \geq c_{\mathfrak{s}}|\kappa| \geq c_{\mathfrak{s}}\rho_\kappa^d \geq \frac{c_{\mathfrak{s}}}{C_{\mathrm{r}}^d}h_\kappa^d.$$

Thereby, we deduce that

$$h_{\mathfrak{s}_i} \geq \frac{c_{\mathfrak{s}}^{1/d}}{C_{\mathrm{r}}}h_\kappa, \tag{3.21}$$

for $i = 1, \ldots, n_\kappa$. Inserting (3.21) into (3.20) we immediately deduce (3.19) with constant equal to $C_{\mathrm{inv},3}C_{\mathrm{r}}^2/c_{\mathfrak{s}}^{2/d}$.

Let us now consider the case when κ is p-coverable. From Definition 3.4 there exits a covering of κ by shape–regular simplices K_i, $i = 1, \ldots, m_\kappa$, such that $|K_i| \geq c_{as}|\kappa|$, $i = 1, \ldots, m_\kappa$. By proceeding in an analogous manner to the previous case, we note that $h_{K_i} \geq c_{as}^{1/d} h_\kappa / C_{\mathrm{r}}$, for $i = 1, \ldots, m_\kappa$, cf. (3.21) above.

Employing (3.3) and Definition 3.4, we deduce that

$$\begin{aligned} \|\nabla v\|^2_{L^2(\kappa)} \leq \sum_{i=1}^{m_\kappa} \|\nabla v\|^2_{L^2(K_i)} &\leq C_{\mathrm{inv},3} \sum_{i=1}^{m_\kappa} \frac{p^4}{h^2_{K_i}} \|v\|^2_{L^2(K_i)} \\ &\leq \frac{C_{\mathrm{inv},3} C_{\mathrm{r}}^2}{c_{as}^{2/d}} \frac{p^4}{h^2_\kappa} \sum_{i=1}^{m_\kappa} \|v\|^2_{L^2(K_i)}. \end{aligned} \tag{3.22}$$

Recalling (3.16) in the proof of Lemma 3.5, the inequality given in (3.22) may be bounded as follows:

$$\|\nabla v\|^2_{L^2(\kappa)} \leq \frac{4\, C_{\mathrm{inv},3} C_{\mathrm{r}}^2 m_\kappa}{c_{as}^{2/d}} \frac{p^4}{h^2_\kappa} \|v\|^2_{L^2(\kappa)},$$

as required. Thereby, the statement of the lemma holds with

$$C_{\mathrm{inv},5} = \max(C_{\mathrm{inv},3} C_{\mathrm{r}}^2 / c_{\mathfrak{s}}^{2/d}, 4\, C_{\mathrm{inv},3} C_{\mathrm{r}}^2 m_\kappa / c_{as}^{2/d}).$$

□

Remark 3.8. We point out that Assumption 3.2.1, which imposes the shape regularity of the mesh $\mathcal{T}_h$ is *only* needed for the proof of Lemma 3.7; this result extends the classical inverse estimate, bounding the H^1–seminorm of a polynomial function with its L^2–norm, to polytopic elements. We note, however, that such inverse estimates depend on the shape regularity of the elements, even in the case of simplicial elements, cf. [183]. Moreover, the Assumption 3.2.1 and Lemma 3.7 are only used for proving the inf-sup stability in Chapter 5.

3.3 *hp*–Approximation bounds

For the approximation theory undertaken in this section, we require the existence of a suitable covering of the mesh by an overlapping set of simplices in $\mathbb{R}^d$. More precisely, we introduce the following definition.

Definition 3.9. We define the *covering* $\mathcal{T}_h^\sharp = \{\mathcal{K}\}$ related to the computational mesh $\mathcal{T}_h$ as a set of open shape-regular d–simplices $\mathcal{K}$, such that for each $\kappa \in \mathcal{T}_h$,

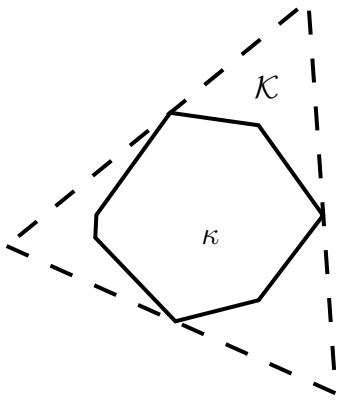


FIGURE 3.5: Polygonal element κ, $\kappa \in \mathcal{T}_h$, in $\mathbb{R}^2$ and the corresponding simplex $\mathcal{K} \in \mathcal{T}_h^\sharp$, $\kappa \subset \mathcal{K}$.

there exists a $\mathcal{K} \in \mathcal{T}_h^\sharp$, such that $\kappa \subset \mathcal{K}$. Given $\mathcal{T}_h^\sharp$, we denote by $\Omega_\sharp$ the *covering domain* given by $\bar{\Omega}_\sharp := \cup_{\mathcal{K} \in \mathcal{T}_h^\sharp} \bar{\mathcal{K}}$.

For clarity, in Figure 3.5 we show a single polygonal element κ, $\kappa \in \mathcal{T}_h$, in $\mathbb{R}^2$ and the corresponding simplex $\mathcal{K} \in \mathcal{T}_h^\sharp$ such that $\kappa \subset \mathcal{K}$. With the definition of the simplicial covering $\mathcal{T}_h^\sharp$ associated with the computational mesh $\mathcal{T}_h$ given in Definition 3.9, we make the following key assumption regarding the amount of allowable overlap between elements in $\mathcal{T}_h$ and the simplices present in $\mathcal{T}_h^\sharp$.

Assumption 3.3.1. *We assume that there exists a covering $\mathcal{T}_h^\sharp$ of $\mathcal{T}_h$ and a positive constant $\mathcal{O}_\Omega$, independent of the mesh parameters, such that*

$$\max_{\kappa \in \mathcal{T}_h} card\Big\{\kappa' \in \mathcal{T}_h : \kappa' \cap \mathcal{K} \neq \emptyset,\ \mathcal{K} \in \mathcal{T}_h^\sharp \ \ such\ that\ \ \kappa \subset \mathcal{K}\Big\} \leq \mathcal{O}_\Omega.$$

As a consequence, we deduce that

$$h_{\mathcal{K}} := \mathrm{diam}(\mathcal{K}) \leq C_{\mathrm{diam}} h_\kappa,$$

for each pair $\kappa \in \mathcal{T}_h$, $\mathcal{K} \in \mathcal{T}_h^\sharp$, with $\kappa \subset \mathcal{K}$, for a constant $C_{\mathrm{diam}} > 0$, uniformly with respect to the mesh size.

Remark 3.10. We point out that Assumption 3.3.1 requires shape–regularity of the mesh covering $\mathcal{T}_h^\sharp$, rather than the corresponding condition being assumed directly for the computational mesh $\mathcal{T}_h$.

In order to derive appropriate *hp*–version approximation estimates on general polytopic elements κ, $\kappa \in \mathcal{T}_h$, we note that standard results, cf. [167], for example, are

applicable by noting that each κ is a subset of a d–simplex belonging to the covering $\mathcal{T}_h^\sharp$ and that the local finite element spaces consist of polynomials *without* the use of element mappings to a reference/canonical element. With this in mind, we recall the following standard *hp*–approximation results (Babuška-Suri operator) on d–simpleces; see, for example, [24] for the case when $d = 2$ and [147] when $d = 3$. We also refer to [25] for similar results.

Lemma 3.11. *Let $\mathcal{K} \in \mathcal{T}_h^\sharp$ be a d–simplex, $d = 2, 3$, of diameter $h_\mathcal{K}$. Suppose further that $v|_\mathcal{K} \in H^l(\mathcal{K})$, for some $l \geq 0$. Then, for $p \in \mathbb{N}$, there exists $\Pi_p v \in \mathcal{P}_p(\mathcal{K})$, such that*

$$\|v - \Pi_p v\|_{H^q(\mathcal{K})} \leq C_{\mathrm{I},1} \frac{h_\mathcal{K}^{s-q}}{p^{l-q}} \|v\|_{H^l(\mathcal{K})}, \quad l \geq 0, \tag{3.23}$$

for $0 \leq q \leq l$, and

$$\|v - \Pi_p v\|_{L^\infty(\mathcal{K})} \leq C_{\mathrm{I},2} \frac{h_\mathcal{K}^{s-d/2}}{p^{l-d/2}} \|v\|_{H^l(\mathcal{K})}, \quad l > d/2. \tag{3.24}$$

Here, $s = \min\{p+1, l\}$ and $C_{\mathrm{I},1}$ and $C_{\mathrm{I},2}$ are positive constants which depend on the shape-regularity of $\mathcal{K}$, but are independent of v, $h_\mathcal{K}$, and p.

In order to generalize Lemma 3.11 to general polytopic elements, we first note that functions defined on Ω can be extended to the covering domain $\Omega_\sharp$ based on the employing the following standard extension operator.

Theorem 3.12. *Let Ω be a domain with a Lipschitz boundary. Then there exists a linear extension operator $\mathfrak{E} : H^s(\Omega) \mapsto H^s(\mathbb{R}^d)$, $s \in \mathbb{N}_0$, such that $\mathfrak{E}v|_\Omega = v$ and*

$$\|\mathfrak{E}v\|_{H^s(\mathbb{R}^d)} \leq C_\mathfrak{E} \|v\|_{H^s(\Omega)},$$

where $C_\mathfrak{E}$ is a positive constant depending only on s and Ω.

Proof. See [171]. □

We note that the assumptions stated in Theorem 3.12 on the domain Ω may be weakened. Indeed, [171] only requires that Ω is a domain with a minimally smooth boundary; the extension to domains which are simply connected, but may contain microscales, is treated in [158].

Secondly, we also recall the following multiplicative trace inequality for d–simplex.

Lemma 3.13. *Let T is a d-dimensional simplex and $F \subset \partial T$ denote one of its faces. Then, given $v \in H^1(T)$, the following inequality holds:*

$$\|v\|^2_{L^2(F)} \leq C_t|F| \left(\frac{1}{|T|}\|v\|^2_{L^2(T)} + \frac{h_T}{|T|}\|v\|_{L^2(T)}\|\nabla v\|_{L^2(T)} \right), \tag{3.25}$$

where C_t is a positive constant depends on d but independent of the mesh size h_T, $|T|$, $|F|$ and shape regularity .

Proof. The proof of (3.25) follows from Lemma 1.49 in [84], also see [145, 67], where the relation (3.25) is written to be independent of unknown constants in the following way.

$$\|v\|^2_{L^2(F)} \leq \left(\frac{|F|}{|T|}\|v\|^2_{L^2(T)} + \frac{2|F|h_T}{d|T|}\|v\|_{L^2(T)}\|\nabla v\|_{L^2(T)} \right), \tag{3.26}$$

Here, it is easy see positive constant C_t depends only on d. □

Given the projection operator Π_p defined in Lemma 3.11 and the extension operator $\mathfrak{E}$ given in Theorem 3.12, we now proceed to define a suitable projection operator on a general polytopic element κ, $\kappa \in \mathcal{T}_h$. To this end, for $v \in L^2(\Omega)$, we define $\tilde{\Pi}_p v \in \mathcal{P}_p(\kappa)$ as follows: for each $\kappa \in \mathcal{T}_h$ and given the associated element $\mathcal{K} \in \mathcal{T}_h^\sharp$, such that $\kappa \subset \mathcal{K}$, cf. Definition 3.9, we write

$$\tilde{\Pi}_p v := \Pi_p(\mathfrak{E}v|_{\mathcal{K}})|_\kappa, \tag{3.27}$$

where $\Pi_p : L^2(\mathcal{K}) \to \mathcal{P}_p(\mathcal{K})$ as defined in Lemma 3.11. With the definition of $\tilde{\Pi}_p$ given in (3.27) we now give the following hp–version approximation bounds.

Lemma 3.14. *Let $\kappa \in \mathcal{T}_h$, $F \subset \partial\kappa$ denote one of its faces, and $\mathcal{K} \in \mathcal{T}_h^\sharp$ the corresponding simplex such that $\kappa \subset \mathcal{K}$, cf. Definition 3.9. Suppose that $v \in L^2(\Omega)$ is such that $\mathfrak{E}v|_{\mathcal{K}} \in H^{l_\kappa}(\mathcal{K})$, for some $l_\kappa \geq 0$. Then, given Assumption 3.3.1 is satisfied, the following bounds hold*

$$\|v - \tilde{\Pi}_p v\|_{H^q(\kappa)} \leq C_{\mathrm{I},3}\frac{h_\kappa^{s_\kappa - q}}{p^{l_\kappa - q}}\|\mathfrak{E}v\|_{H^{l_\kappa}(\mathcal{K})}, \quad l_\kappa \geq 0, \tag{3.28}$$

for $0 \leq q \leq l_\kappa$, and

$$\|v - \tilde{\Pi}_p v\|_{L^2(F)} \leq C_{\mathrm{I},4}|F|^{1/2}\frac{h_\kappa^{s_\kappa - d/2}}{p^{l_\kappa - 1/2}}C_m(p,\kappa,F)^{1/2}\|\mathfrak{E}v\|_{H^{l_\kappa}(\mathcal{K})}, \; l_\kappa > d/2, \tag{3.29}$$

where

$$C_m(p,\kappa,F)=\min\Big\{\frac{h_\kappa^d}{\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|},\frac{1}{p^{1-d}}\Big\},$$

$s_\kappa=\min\{p+1,l_\kappa\}$ *and* $C_{\mathrm{I},3}$ *and* $C_{\mathrm{I},4}$ *are positive constants, which depend on the shape-regularity of* $\mathcal{K}$*, but are independent of* v*,* h_κ*, and* p*.*

Proof. To prove (3.28), we note that

$$\|v-\tilde\Pi_p v\|_{H^q(\kappa)}=\|\mathfrak{E}v-\Pi_p(\mathfrak{E}v)\|_{H^q(\kappa)}\le\|\mathfrak{E}v-\Pi_p(\mathfrak{E}v)\|_{H^q(\mathcal{K})}.$$

Thereby, upon application of (3.23) and noting that Assumption 3.3.1 holds, the desired bound follows immediately with $C_{\mathrm{I},3}=C_{\mathrm{I},1}C_{\mathrm{diam}}^{s_\kappa-q}$,

To prove (3.29), we let $\kappa_\flat^F\in\mathcal{F}_\flat^\kappa$, cf. Definition 3.2; then applying a standard scaling argument with respect to $\kappa_\flat^F$, the multiplicative trace inequality (3.25), and (3.28), we obtain

$$\begin{aligned}\|v-\tilde\Pi_p v\|_{L^2(F)}^2 &\le C_{\mathrm{t}}|F|\Big(\frac{1}{|\kappa_\flat^F|}\|v-\tilde\Pi_p v\|_{L^2(\kappa_\flat^F)}^2\\ &\quad+h_{\kappa_\flat^F}|\kappa_\flat^F|^{-1}\|v-\tilde\Pi_p v\|_{L^2(\kappa_\flat^F)}\|\nabla(v-\tilde\Pi_p v)\|_{L^2(\kappa_\flat^F)}\Big)\\ &\le C_{\mathrm{t}}\,C_{\mathrm{I},1}^2\,C_{\mathrm{diam}}^{2s_\kappa-1}\,\frac{|F|}{|\kappa_\flat^F|}\left(C_{\mathrm{diam}}\frac{h_\kappa}{p}+h_{\kappa_\flat^F}\right)\frac{h_\kappa^{2s_\kappa-1}}{p^{2l_\kappa-1}}\|\mathcal{E}v\|_{H^{l_\kappa}(\mathcal{K})}^2.\end{aligned}\tag{3.30}$$

Given that $h_{\kappa_\flat^F}\le h_\kappa$ and $\kappa_\flat^F$ is arbitrary, from (3.30) we conclude that

$$\|v-\tilde\Pi v\|_{L^2(F)}^2\le C_{\mathrm{t}}\,C_{\mathrm{I},1}^2\,C_{\mathrm{diam}}^{2s_\kappa-1}(1+C_{\mathrm{diam}})\frac{|F|}{\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|}\frac{h_\kappa^{2s_\kappa}}{p^{2l_\kappa-1}}\|\mathcal{E}v\|_{H^{l_\kappa}(\mathcal{K})}^2.\tag{3.31}$$

On the other hand, proceeding as in the proof of Lemma 3.5, we observe that

$$\|v-\tilde\Pi_p v\|_{L^2(F)}^2\le|F|\|v-\tilde\Pi_p v\|_{L^\infty(F)}^2.$$

Hence, employing the definition of the projection operator $\tilde\Pi_p$, together with (3.24) and Assumption 3.3.1, we deduce that

$$\|v-\tilde\Pi_p v\|_{L^2(F)}^2\le C_{\mathrm{I},2}^2\,C_{\mathrm{diam}}^{2s_\kappa-d}\,|F|\frac{h_\kappa^{2s_\kappa-d}}{p^{2l_\kappa-d}}\|\mathfrak{E}v\|_{H^{l_\kappa}(\mathcal{K})}^2.\tag{3.32}$$

Thereby, taking the minimum of the two bounds (3.31) and (3.32), the approximation result stated in (3.29) holds with

$$C_{\mathrm{I},4}=\max(C_{\mathrm{I},1}\,C_{\mathrm{diam}}^{s_\kappa-1/2}\,\sqrt{C_{\mathrm{t}}(1+C_{\mathrm{diam}})},C_{\mathrm{I},2}C_{\mathrm{diam}}^{s_\kappa-d/2}).$$

□

Remark 3.15. We note that (3.31) is also valid for the case when $l_\kappa \geq 1$; for simplicity of presentation, we have omitted this level of generality in the statement of Lemma 3.14.

Chapter 4

DGFEMs for Pure Diffusion PDEs

On the basis of the *hp*–version inverse and approximation bounds developed in the previous chapter, here we study the IP-DGFEM discretization of pure diffusion problems based on two different type mesh assumptions over polytopic meshes.

4.1 Model problem

Let Ω be a bounded Lipschitz domain in $\mathbb{R}^d$, $d \geq 1$, and let $\partial\Omega$ signify the union of its $(d-1)$-dimensional open faces. We consider the following PDE: find u as the solution of

$$
\begin{aligned}
-\nabla \cdot (a\nabla u) &= f \quad \text{in } \Omega, && (4.1)\\
u &= g_{\mathrm{D}} \quad \text{on } \partial\Omega_{\mathrm{D}}, && (4.2)\\
\mathbf{n} \cdot (a\nabla u) &= g_{\mathrm{N}} \quad \text{on } \partial\Omega_{\mathrm{N}}. && (4.3)
\end{aligned}
$$

Here, $f \in L^2(\Omega)$, $a = \{a_{ij}\}_{i,j=1}^d$ with $a_{ij} \in L^\infty(\Omega)$ and $a_{ij} = a_{ji}$, for $i, j = 1, \ldots, d$, at each $\mathbf{x}$ in $\bar{\Omega}$,

$$
\sum_{i,j=1}^{d} a_{ij}(\mathbf{x})\xi_i\xi_j \geq \theta|\boldsymbol{\xi}|^2 > 0, \tag{4.4}
$$

with θ a positive constant, for any vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ in $\mathbb{R}^d$. For the sake of simplicity, we divide $\partial\Omega$ into two disjoint subsets $\partial\Omega_{\mathrm{D}}$ and $\partial\Omega_{\mathrm{N}}$ whose union is

$\partial\Omega$, with $\partial\Omega_{\rm D}$ is non empty and relatively open in $\partial\Omega$. The well–posedness of the boundary value problem (4.1), (4.2), (4.3) under the *uniform ellipticity condition* (4.4), can be proved by using the Lax-Milgram theorem; see [70, 46].

Before we present the IP DGFEMs for elliptic problems, we will talk about two mesh assumptions used in this chapter. The first mesh assumption is given in Assumption 3.1.1, which can be interpreted as *each polytopic mesh has uniformly bounded number of faces without any shape regularity restrictions.* This mesh assumption first appeared in [61] and is already utilised in Chapter 3 for deriving inverse estimation and polynomial approximation results. We will keep on using this mesh assumption for DGFEMs to solve elliptic PDEs in Section 4.2 and also to solve PDEs in non-negative characteristic form in Chapter 5.

On the other hand, the second mesh assumption can be interpreted as *each polytopic mesh is allowed to have arbitrary number of faces if it satisfies a general shape regularity assumption.* This mesh assumption first appeared in [58]; its precise definition is given in Section 4.3. In this setting, we can simplify some inverse estimate and polynomial approximation results from Chapter 3. Then, these new results will be utilised for DGFEMs to solve elliptic PDEs in Section 4.3 and also to solve time dependent PDEs in Chapter 6.

4.2 DGFEMs for elliptic PDEs on polytopic meshes with bounded number of faces

Following Chapter 3, we write $\mathcal{T}_h$ to denote a subdivision of the computational domain $\Omega \subset \mathbb{R}^d$, $d > 1$, into disjoint open polytopic elements κ constructed such that $\bar{\Omega} = \cup_{\kappa\in\mathcal{T}_h}\bar{\kappa}$. Recalling that $\mathcal{F}_h$ denotes the set of open $(d-1)$–dimensional element faces associated with the computational mesh $\mathcal{T}_h$, employing the notation introduced in Chapter 2, we write $\mathcal{F}_h = \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{B}}$, where $\mathcal{F}_h^{\mathcal{I}}$ denotes the set of all open $(d-1)$–dimensional element faces $F \in \mathcal{F}_h$ that are contained in Ω, and $\mathcal{F}_h^{\mathcal{B}}$ is the set of element boundary faces, i.e., $F \subset \partial\Omega$ for $F \in \mathcal{F}_h^{\mathcal{B}}$. For simplicity, we assume that $\mathcal{T}_h$ respects the decomposition of $\partial\Omega$ in the sense that each $F \in \mathcal{F}_h^{\mathcal{B}}$ belongs to the interior of exactly one of $\partial\Omega_{\rm D}$ or $\partial\Omega_{\rm N}$. Hence we further denote by $\mathcal{F}_h^{\mathcal{D}}, \mathcal{F}_h^{\mathcal{N}} \subset \mathcal{F}_h^{\mathcal{B}}$ as the subsets of boundary faces belonging to $\partial\Omega_{\rm D}, \partial\Omega_{\rm N}$, respectively.

To facilitate *hp*–adaptivity, to each element $\kappa \in \mathcal{T}_h$, we write $p_\kappa \geq 1$ to denote the elementwise polynomial degree, and collect the p_κ in the vector $\mathbf{p} := (p_\kappa : \kappa \in \mathcal{T}_h)$. With this notation, we define the finite element space $S^{\mathbf{p}}_{\mathcal{T}_h}$ with respect to $\mathcal{T}_h$ and $\mathbf{p}$ by

$$S^{\mathbf{p}}_{\mathcal{T}_h} := \{u \in L^2(\Omega) : u|_\kappa \in \mathcal{P}_{p_\kappa}(\kappa), \kappa \in \mathcal{T}_h\},$$

where, we recall that $\mathcal{P}_p(\kappa)$ denotes the space of polynomials of total degree p on κ. We stress that, by construction, the local elemental polynomial spaces employed within the definition of $S^{\mathbf{p}}_{\mathcal{T}_h}$ are defined in the physical space, without the need to map from a given reference or canonical frame, as is typically necessary for classical finite element methods.

Following the derivation presented in Section 2.3, we introduce the following (symmetric) IP-DGFEM bilinear form

$$\begin{aligned} B_{\mathrm{d}}(u_h, v_h) \quad &= \sum_{\kappa \in \mathcal{T}_h} \int_\kappa a \nabla u_h \cdot \nabla v_h \, \mathrm{d}\mathbf{x} \\ &\quad - \sum_{F \in \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}} \int_F (\{\!\{a \nabla u_h\}\!\} \cdot [\![v_h]\!] + \{\!\{a \nabla v_h\}\!\} \cdot [\![u_h]\!] - \sigma [\![u_h]\!] \cdot [\![v_h]\!]) \, \mathrm{d}s, \end{aligned}$$

and linear functional

$$\ell(v_h) = \sum_{\kappa \in \mathcal{T}_h} \int_\kappa f v_h \, \mathrm{d}\mathbf{x} - \sum_{F \in \mathcal{F}_h^{\mathcal{D}}} \int_F g_{\mathrm{D}} (a \nabla v_h \cdot \mathbf{n} - \sigma v_h) \, \mathrm{d}s + \sum_{F \in \mathcal{F}_h^{\mathcal{N}}} \int_F g_{\mathrm{N}} v_h \, \mathrm{d}s,$$

for $u_h, v_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$. The corresponding DGFEM reads: find $u_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$ such that

$$B_{\mathrm{d}}(u_h, v_h) = \ell(v_h), \tag{4.5}$$

for all $v_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$.

The well-posedness and stability properties of the above formulation depend on the choice of the discontinuity penalization function σ. These are analysed in the next section based on the new *hp*–version inverse estimates presented in the previous chapter, whereby we anticipate that the choice of σ must be sensitive to the size of each face relative to that of the neighbouring elements.

4.2.1 The well-posedness of the IP-DGFEMs

To focus on the treatment of general polytopic subdivisions, we consider here the special case of piecewise constant diffusion tensors, i.e.,

$$a \in [V^0(\mathcal{T}_h)]^{d\times d}_{\mathrm{sym}}. \tag{4.6}$$

For the case of general positive semidefinite diffusion tensors, see [105] and [60]. We can consider $\sqrt{a}$ as the unique (positive definite) square-root of the symmetric matrix a and define $\bar{a}_\kappa := |\sqrt{a}|_2^2|_\kappa$, with $|\cdot|_2$ denoting the matrix l_2–norm.

Definition 4.1. Assume that (4.6) holds. The discontinuity penalization function $\sigma : \mathcal{F}_h \to \mathbb{R}$ is given by

$$\sigma(x) = \begin{cases} C_\sigma \max\limits_{\kappa\in\{\kappa_i,\kappa_j\}} \left\{ C_{\mathrm{INV}}(p_\kappa,\kappa,F)\ \dfrac{\bar{a}_\kappa p_\kappa^2 |F|}{|\kappa|} \right\}, & F \in \mathcal{F}_h^{\mathcal{I}},\ F = \partial\kappa_i \cap \partial\kappa_j, \\ C_\sigma C_{\mathrm{INV}}(p_\kappa,\kappa,F)\ \dfrac{\bar{a}_\kappa p_\kappa^2 |F|}{|\kappa|}, & F \in \mathcal{F}_h^{\mathcal{D}},\ F \subset \partial\kappa. \end{cases} \tag{4.7}$$

Here, C_{INV} is the constant of the inverse inequality of Lemma 3.5 and $C_\sigma > 0$ is a constant independent of p_κ, $|F|$, and of $|\kappa|$.

In accordance with the mesh settings laid out in Section 3.1, the value of the discontinuity penalization function σ on a given elemental interface is independently determined on each constituent $(d-1)$–dimensional simplicial mesh face. This way, the penalization function is independent of any local h or p quasi-uniformity or hanging nodes regularity assumption. In particular, for standard simplicial and tensor product meshes with hanging nodes, the independent piecewise constant definition of the penalization function allows for *irregular* hanging nodes with arbitrary positioning within the parent interface. This is in contrast with standard IP-DGFEM settings, whereby irregular hanging nodes are not permitted as the penalization function definition relies on the face and parent interface to be of size comparable to that of the element [125].

A first issue encountered when analysing (4.5) is that this formulation cannot be extended to functions in $H^1(\Omega)$. Indeed functions in $L^2(\Omega)$ do not have well-defined traces on $\mathcal{F}_h$ and hence the terms $\{\!\{\nabla v\}\!\}$ are not well defined for $v \in H^1(\Omega)$. Hence we are not allowed to test in (4.5) with the analytical solution of (4.1) unless we assume that this is regular enough. This issue can be overcome by introducing suitable extensions of the bilinear form $B_{\mathrm{d}}(\cdot,\cdot)$ and linear functional $\ell(\cdot)$.

Let $\mathbf{\Pi}_2 : [L^2(\Omega)]^d \to [S^{\mathbf{p}}_{\mathcal{T}_h}]^d$ denote the orthogonal L^2-projection onto the finite element space $[S^{\mathbf{p}}_{\mathcal{T}_h}]^d$. Following [152, 105], we define the bilinear form

$$\begin{aligned}\tilde{B}_{\mathrm{d}}(w,v) \;&:=\; \sum_{\kappa\in\mathcal{T}_h}\int_\kappa a\nabla u\cdot\nabla v\,\mathrm{d}\mathbf{x}\\ &\quad-\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F\left(\{\!\{a\mathbf{\Pi}_2(\nabla u)\}\!\}\cdot[\![v]\!]+\{\!\{a\mathbf{\Pi}_2(\nabla v)\}\!\}\cdot[\![u_h]\!]-\sigma[\![u]\!]\cdot[\![v]\!]\,\mathrm{d}s\right),\end{aligned} \tag{4.8}$$

and linear functional

$$\tilde{\ell}(v)=\sum_{\kappa\in\mathcal{T}_h}\int_\kappa fv\,\mathrm{d}\mathbf{x}-\sum_{F\in\mathcal{F}_h^{\mathcal{D}}}\int_F g_{\mathrm{D}}(a\mathbf{\Pi}_2(\nabla v)\cdot\mathbf{n}-\sigma v)\,\mathrm{d}s+\sum_{F\in\mathcal{F}_h^{\mathcal{N}}}\int_F g_{\mathrm{N}}v_h\,\mathrm{d}s,$$

for all $v, w \in \mathcal{S} := H^1(\Omega) + S^{\mathbf{p}}_{\mathcal{T}_h}$. Then the DGFEM formulation (4.5) is equivalent to: find $u_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$ such that

$$\tilde{B}_{\mathrm{d}}(u_h, v_h) = \tilde{\ell}(v_h), \tag{4.9}$$

for all $v_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$. This discrete problem is inconsistent with (4.1), hence Galerkin orthogonality does not hold. On the other hand, weaker regularity assumptions on the analytical solution are required allowing us to prove continuity and coercivity of the bilinear form $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$ on the whole of $\mathcal{S}$.

We analyse the DGFEM method in the associated energy norm

$$|||v|||^2_{\mathrm{DG}} := \sum_{\kappa\in\mathcal{T}_h}\|\sqrt{a}\nabla v\|^2_{L^2(\kappa)}+\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F\sigma|[\![v]\!]|^2\,\mathrm{d}s. \tag{4.10}$$

Here and in following chapters we shall often make use of the arithmetic–geometric mean inequality $ab \le a^2\epsilon + \frac{b^2}{4\epsilon}$, holding for any $a, b \in \mathbb{R}$ and $\epsilon > 0$.

Lemma 4.2. *Under Assumption 3.1.1 and with σ as in Definition 4.1 with C_σ large enough, the bilinear form given by* (4.8) *is coercive and continuous, that is*

$$\tilde{B}_{\mathrm{d}}(v,v) \ge C_{\mathrm{coer}}|||v|||^2_{\mathrm{DG}} \quad \textit{for all} \quad v\in\mathcal{S}, \tag{4.11}$$

and

$$\tilde{B}_{\mathrm{d}}(w,v) \le C_{\mathrm{cont}}|||w|||_{\mathrm{DG}}\,|||v|||_{\mathrm{DG}} \quad \textit{for all} \quad w,v\in\mathcal{S}, \tag{4.12}$$

where $C_{\rm coer}$ and $C_{\rm cont}$ are positive constants, independent of the mesh size h_κ and polynomial order p_κ.

Proof. The proof follows standard arguments [84], by employing the inverse inequality Lemma 3.5 in place of the standard inverse inequality for simplexes. From the definition of $\tilde{B}_{\rm d}$, for any $v \in \mathcal{S}$, we have

$$\tilde{B}_{\rm d}(v,v) \;=\; |||v|||^2_{\rm DG} - 2\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F \{\!\{a\boldsymbol{\Pi}_2(\nabla v)\}\!\}\cdot[\![v]\!]\,{\rm d}s, \tag{4.13}$$

and it remains to bound the second term on the right-hand side. For $F \in \mathcal{F}_h^{\mathcal{I}}$, the Cauchy–Schwarz inequality and arithmetic–geometric mean inequality imply

$$\begin{aligned}\int_F \{\!\{a\boldsymbol{\Pi}_2(\nabla v)\}\!\}\cdot[\![v]\!]\,{\rm d}s \;&\le\; \frac{1}{2}\left(\|\frac{1}{\sqrt{\sigma}}a\boldsymbol{\Pi}_2(\nabla v^+)\|_{L^2(F)} + \|\frac{1}{\sqrt{\sigma}}a\boldsymbol{\Pi}_2(\nabla v^-)\|_{L^2(F)}\right)\\ &\quad\times\|\sqrt{\sigma}[\![v]\!]\|_{L^2(F)}\\ &\le\; \epsilon\left(\|\frac{1}{\sqrt{\sigma}}a\boldsymbol{\Pi}_2(\nabla v^+)\|^2_{L^2(F)} + \|\frac{1}{\sqrt{\sigma}}a\boldsymbol{\Pi}_2(\nabla v^-)\|^2_{L^2(F)}\right)\\ &\quad+\frac{1}{8\epsilon}\|\sqrt{\sigma}[\![v]\!]\|^2_{L^2(F)}.\end{aligned}$$

Using the inverse inequality Lemma 3.5, the definition of the interior penalty parameter σ, the assumption of diffusion tensor in (4.6), and the L^2-stability of the projector $\boldsymbol{\Pi}_2$, we conclude that

$$\begin{aligned}\int_F \{\!\{a\boldsymbol{\Pi}_2(\nabla v)\}\!\}\cdot[\![v]\!]\,{\rm d}s \;&\le\; \epsilon\left(C_{\rm INV}(p_{\kappa^+},\kappa^+,F)\frac{\bar{a}_{\kappa^+}p^2_{\kappa^+}|F|}{|\kappa^+|}\|\frac{1}{\sqrt{\sigma}}\sqrt{a}\boldsymbol{\Pi}_2(\nabla v)\|^2_{L^2(\kappa^+)}\right.\\ &\quad\left.+C_{\rm INV}(p_{\kappa^-},\kappa^-,F)\frac{\bar{a}_{\kappa^-}p^2_{\kappa^-}|F|}{|\kappa^-|}\|\frac{1}{\sqrt{\sigma}}\sqrt{a}\boldsymbol{\Pi}_2(\nabla v)\|^2_{L^2(\kappa^-)}\right)\\ &\quad+\frac{1}{8\epsilon}\|\sqrt{\sigma}[\![v]\!]\|^2_{L^2(F)}\\ &\le\; \frac{\epsilon}{C_\sigma}\left(\|\sqrt{a}\nabla v\|^2_{L^2(\kappa^+)} + \|\sqrt{a}\nabla v\|^2_{L^2(\kappa^-)}\right)\\ &\quad+\frac{1}{8\epsilon}\|\sqrt{\sigma}[\![v]\!]\|^2_{L^2(F)}.\end{aligned} \tag{4.14}$$

Similarly, for $F \in \mathcal{F}_h^{\mathcal{D}}$, we have that

$$\int_F \{\!\{a\boldsymbol{\Pi}_2(\nabla v)\}\!\}\cdot[\![v]\!]\,{\rm d}s \le \frac{\epsilon}{C_\sigma}\|\sqrt{a}\nabla v\|^2_{L^2(\kappa^+)} + \frac{1}{4\epsilon}\|\sqrt{\sigma}[\![v]\!]\|^2_{L^2(F)}. \tag{4.15}$$

Inserting (4.27) and (4.15) into (4.13) immediately gives

$$\tilde{B}_{\mathrm{d}}(v,v) \;\geq\; \left(1-\frac{2C_F}{C_\sigma}\epsilon\right)\sum_{\kappa\in\mathcal{T}_h}\|\sqrt{a}\nabla v\|^2_{L^2(\kappa)}+\left(1-\frac{1}{2\epsilon}\right)\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\|\sqrt{\sigma}[\![v]\!]\|^2_{L^2(F)},$$

as the number of elemental faces is uniformly bounded by Assumption 3.1.1. Hence the bilinear form $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$ is coercive over $\mathcal{S}\times\mathcal{S}$ if $C_\sigma > 2C_F\epsilon$ for some $\epsilon > 1/2$.

The continuity of $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$ easily follows by applying the Cauchy-Schwarz inequality and then bounding the face terms by repeating the arguments leading to (4.27). □

Remark 4.3. The above analysis extends well known results for standard meshes to meshes made of general polytopes. It is based on exploiting the new inverse estimate of Lemma 3.5 to control the face terms and on requiring that the number of elemental faces is uniformly bounded, cf. Assumption 3.1.1, when summing up the contributions of all faces. This approach has the crucial advantage of permitting very general polytopic meshes as no mesh elements shape regularity is directly assumed.

Remark 4.4. It is possible to avoid the composition of a bound on the number of faces by requiring, instead, that the mesh satisfies a certain shape regularity assumption, as we shall show in Section 4.3. We note that the present approach and the one described below may be easily combined to produce stable DGFEM discretisations on very general mesh settings.

4.2.2 A priori error analysis

The following abstract error bound is an instance of Strang's second lemma [172, 70], whereby the error is controlled by the sum of a quasi-optimal approximation term and a residual term.

Lemma 4.5. *Let $u\in H^1(\Omega)$ be the weak solution of* (4.3) *and $u_h\in S^{\mathbf{p}}_{\mathcal{T}_h}$ be the DGFEM solution given by* (4.5)*. Under the hypotheses of Lemma 4.2, it holds*

$$|||u-u_h|||_{\mathrm{DG}} \leq \left(1+\frac{C_{\mathrm{cont}}}{C_{\mathrm{coer}}}\right)\inf_{v_h\in S^{\mathbf{p}}_{\mathcal{T}_h}}|||u-v_h|||_{\mathrm{DG}}+\frac{1}{C_{\mathrm{coer}}}\sup_{w_h\in S^{\mathbf{p}}_{\mathcal{T}_h}}\frac{|\tilde{B}_{\mathrm{d}}(u,w_h)-\tilde{\ell}(u,w_h)|}{|||w_h|||_{\mathrm{DG}}}.$$

Proof. By the triangle inequality,

$$|||u-u_h|||_{\mathrm{DG}} \leq |||u-v_h|||_{\mathrm{DG}}+|||v_h-u_h|||_{\mathrm{DG}},$$

for all $v_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$, and it remains to bound $|||v_h - u_h|||_{\mathrm{DG}}$. To this end, we use the coercivity on $S^{\mathbf{p}}_{\mathcal{T}_h}$ and continuity on $\mathcal{S}$ of $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$, to obtain

$$\begin{aligned}
|||u_h - v_h|||^2_{\mathrm{DG}} &\leq \frac{1}{C_{\mathrm{coer}}}\tilde{B}_{\mathrm{d}}(v_h - u_h, v_h - u_h) \\
&= \frac{1}{C_{\mathrm{coer}}}(\tilde{B}_{\mathrm{d}}(v_h - u, u_h - v_h) + \tilde{B}_{\mathrm{d}}(u - u_h, u_h - v_h)) \\
&\leq \frac{C_{\mathrm{cont}}}{C_{\mathrm{coer}}}|||v_h - u|||_{\mathrm{DG}}|||u_h - v_h|||_{\mathrm{DG}} \\
&\quad + \frac{1}{C_{\mathrm{coer}}}(\tilde{B}_{\mathrm{d}}(u, u_h - v_h) - \tilde{\ell}(u_h - v_h)),
\end{aligned}$$

and the required bound easily follows. □

The abstract error bound of Lemma 4.5 is used to derive convergence results for the method at hand. These depends on the availability of the *hp*–version approximation estimates of Lemma 3.11. Assume that the mesh $\mathcal{T}_h$ admits a shape regular covering $\mathcal{T}_h^\sharp = \{\mathcal{K}\}$, cf. Definition 3.9, satisfying Assumption 3.3.1, and further assume that $u|_\kappa \in H^{l_\kappa}(\kappa)$, for some $l_\kappa > 1 + d/2$, for each $\kappa \in \mathcal{T}_h$, so that, by Theorem 3.12, $\mathfrak{E}u|_{\mathcal{K}} \in H^{l_\kappa}(\mathcal{K})$, where $\mathcal{K} \in \mathcal{T}_h^\sharp$ with $\kappa \subset \mathcal{K}$. Then, the approximation estimates of Lemma 3.14 together with Assumption 3.1.1 give

$$\begin{aligned}
&\inf_{v \in S^{\mathbf{p}}_{\mathcal{T}_h}} |||u - v|||^2_{\mathrm{DG}} \leq |||u - \tilde{\Pi}_{p_\kappa} u|||^2_{\mathrm{DG}} \\
&\leq \sum_{\kappa \in \mathcal{T}_h} \left(\|\sqrt{a}\nabla(u - \tilde{\Pi}_{p_\kappa} u)\|^2_{L^2(\kappa)} + 2 \sum_{F \subset \partial\kappa \setminus \mathcal{F}_h^{\mathcal{N}}} \sigma \|(u - \tilde{\Pi}_{p_\kappa} u)|_\kappa\|^2_F \right) \\
&\leq C \sum_{\kappa \in \mathcal{T}_h} \frac{h_\kappa^{2(s_\kappa - 1)}}{p_\kappa^{2(l_\kappa - 1)}} \left(\bar{a}_\kappa + \frac{h_\kappa^{-d+2}}{p_\kappa} \sum_{F \subset \partial\kappa \setminus \mathcal{F}_h^{\mathcal{N}}} C_m(p_\kappa, \kappa, F)\sigma|F| \right) \|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}, \quad (4.16)
\end{aligned}$$

with $s_\kappa = \min\{p_\kappa + 1, l_\kappa\}$.

Similarly, we can bound the residual term as follows. First note that integration by parts elementwise together with the identity (2.26) and the fact that u is the solution of (4.1), gives

$$\begin{aligned}
\left|\tilde{B}_{\mathrm{d}}(u, w_h) - \tilde{\ell}(u, w_h)\right| &= \Big| \sum_{F \in \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}} \int_F \{\!\{a(\nabla u - \mathbf{\Pi}_2(\nabla u))\}\!\} \cdot [\![w_h]\!] \, \mathrm{d}s \Big| \\
&\leq \Big(\sum_{F \in \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}} \int_F \sigma^{-1} |\{\!\{a(\nabla u - \mathbf{\Pi}_2(\nabla u))\}\!\}|^2 \, \mathrm{d}s \Big)^{1/2} |||w_h|||_{\mathrm{DG}}.
\end{aligned}$$

Let $\tilde{\mathbf{\Pi}}$ denote the vector-valued *hp*–projection operator obtained by applying componentwise the operator $\tilde{\Pi}_{p_\kappa}$ given in (3.27). Adding and subtracting $\tilde{\mathbf{\Pi}}(\nabla u)$, we obtain

$$\begin{aligned}
&\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F \sigma^{-1}|\{\!\{a(\nabla u-\mathbf{\Pi}_2(\nabla u))\}\!\}|^2\,\mathrm{d}s\\
&\quad\le \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F 2\sigma^{-1}(|\{\!\{a(\nabla u-\tilde{\mathbf{\Pi}}(\nabla u))\}\!\}|^2+|\{\!\{a(\mathbf{\Pi}_2(\tilde{\mathbf{\Pi}}(\nabla u)-\nabla u))\}\!\}|^2)\,\mathrm{d}s.\\
&\quad\equiv \mathrm{I}+\mathrm{II}.
\end{aligned}$$

Using, as above, the approximation estimate (3.29) yields:

$$\mathrm{I}\le C\sum_{\kappa\in\mathcal{T}_h}\bar{a}_\kappa^2\frac{h_\kappa^{2(s_\kappa-1)}}{p_\kappa^{2(l_\kappa-1)}}\frac{h_\kappa^{-d}}{p_\kappa^{-1}}\sum_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}C_m(p_\kappa,\kappa,F)\sigma^{-1}|F|\|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}.$$

Similarly, the inverse inequality (3.13), the L^2-stability of the projector $\mathbf{\Pi}_2$, and the approximation estimate (3.28), yield:

$$\mathrm{II}\le C\sum_{\kappa\in\mathcal{T}_h}\bar{a}_\kappa^2\frac{h_\kappa^{2(s_\kappa-1)}}{p_\kappa^{2(l_\kappa-1)}}\frac{|\kappa|^{-1}}{p_\kappa^{-2}}\left(\sum_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}C_{\mathrm{INV}}(p_\kappa,\kappa,F)\sigma^{-1}|F|\right)\|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}.$$

Combining the above developments we arrive to the following bound of the residual term:

$$\begin{aligned}
&\sup_{w_h\in S^{\mathbf{p}}_{\mathcal{T}_h}}\frac{|\tilde{B}_{\mathrm{d}}(u,w_h)-\tilde{\ell}(u,w_h)|}{|||w_h|||_{\mathrm{DG}}}\le\Big(\mathrm{I}+\mathrm{II}\Big)^{1/2}\\
&\le C\Bigg(\sum_{\kappa\in\mathcal{T}_h}\bar{a}_\kappa^2\frac{h_\kappa^{2(s_\kappa-1)}}{p_\kappa^{2(l_\kappa-1)}}\\
&\qquad\times\left(\sum_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}\left(C_m(p_\kappa,\kappa,F)\frac{h_\kappa^{-d}}{p_\kappa^{-1}}+C_{\mathrm{INV}}(p_\kappa,\kappa,F)\frac{|\kappa|^{-1}}{p_\kappa^{-2}}\right)\sigma^{-1}|F|\right)\\
&\qquad\times\|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}\Bigg)^{1/2}.
\end{aligned}\tag{4.17}$$

Now the approximation bound (4.16) and residual bound (4.17) yield the following DGFEM convergence result.

Theorem 4.6. *Let $\mathcal{T}_h=\{\kappa\}$ be a subdivision of $\Omega\subset\mathbb{R}^d$, $d=2,3$, consisting of general polytopic elements satisfying Assumption 3.1.1 and Assumption 3.3.1 with $\mathcal{T}_h^\sharp=\{\mathcal{K}\}$ an associated covering of $\mathcal{T}_h$ consisting of shape-regular d–simplices, cf.*

Definition 3.9. Let $u_h \in S^{\mathfrak{p}}_{\mathcal{T}_h}$, *with* $p_\kappa \geq 1$ *for all* $\kappa \in \mathcal{T}_h$, *be the corresponding DGFEM solution defined by* (4.5) *with the discontinuity-penalization functions given by* (4.7). *If the exact solution* $u \in H^1(\Omega)$ *to* (4.1)–(4.3) *satisfies* $u|_\kappa \in H^{l_\kappa}(\kappa)$, $l_\kappa > 1 + d/2$, *for each* $\kappa \in \mathcal{T}_h$, *such that* $\mathfrak{E}u|_{\mathcal{K}} \in H^{l_\kappa}(\mathcal{K})$, *where* $\mathcal{K} \in \mathcal{T}_h^\sharp$ *with* $\kappa \subset \mathcal{K}$, *then*

$$|||u - u_h|||^2_{\mathrm{DG}} \leq C \sum_{\kappa \in \mathcal{T}_h} \frac{h_\kappa^{2(s_\kappa - 1)}}{p_\kappa^{2(l_\kappa - 1)}} \left(\bar{a}_\kappa + \mathcal{G}_\kappa(F, C_{\mathrm{INV}}, C_m, p_\kappa)\right) \|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}.$$

Here, $s_\kappa = \min\{p_\kappa + 1, l_\kappa\}$,

$$\mathcal{G}_\kappa(F, C_{\mathrm{INV}}, C_m, p_\kappa) = \bar{a}_\kappa^2 p_\kappa h_\kappa^{-d} \sum_{F \subset \partial\kappa \setminus \mathcal{F}_h^{\mathcal{N}}} C_m(p_\kappa, \kappa, F)\sigma^{-1}|F|$$
$$+\bar{a}_\kappa^2 p_\kappa^2 |\kappa|^{-1} \sum_{F \subset \partial\kappa \setminus \mathcal{F}_h^{\mathcal{N}}} C_{\mathrm{INV}}(p_\kappa, \kappa, F)\sigma^{-1}|F| + h_\kappa^{-d+2} p_\kappa^{-1} \sum_{F \subset \partial\kappa \setminus \mathcal{F}_h^{\mathcal{N}}} C_m(p_\kappa, \kappa, F)\sigma|F|,$$

and C *is a positive constant independent of the discretization parameters.*

Remark 4.7. The above result generalises well-known *a priori* bounds for IP-DGFEMs defined on standard element domains [125, 156] in two ways. Firstly, meshes comprising polytopic elements are allowed. Secondly, elemental faces are allowed to degenerate. For $d = 3$, this also implies that positive measure interfaces may have degenerating $(d-2)$–dimensional edges. In turns, this freedom is relevant to standard (simplicial/hexahedral) meshes with hanging nodes in that no regularity assumptions of the hanging nodes is required. If, on the other hand, the diameter of the faces of each element $\kappa \in \mathcal{T}_h$ is of comparable size to the diameter of the corresponding element, then the *a priori* error bound of Theorem 4.6 reduces to

$$|||u - u_h|||_{\mathrm{DG}} \leq C \frac{h^{s-1}}{p^{l-\frac{3}{2}}} \|u\|_{H^l(\Omega)}.$$

This coincides with the analogous result derived in [125] for standard meshes consisting of simplices or tensor-product elements. It is easy to check that the above a priori error bound is optimal in h and suboptimal in p by half an order.

Finally, we point out that *a priori* bounds depend on the mesh assumption 3.1.1, which allows shape irregular polytopic meshes but meshes should have uniformly bounded number of faces.

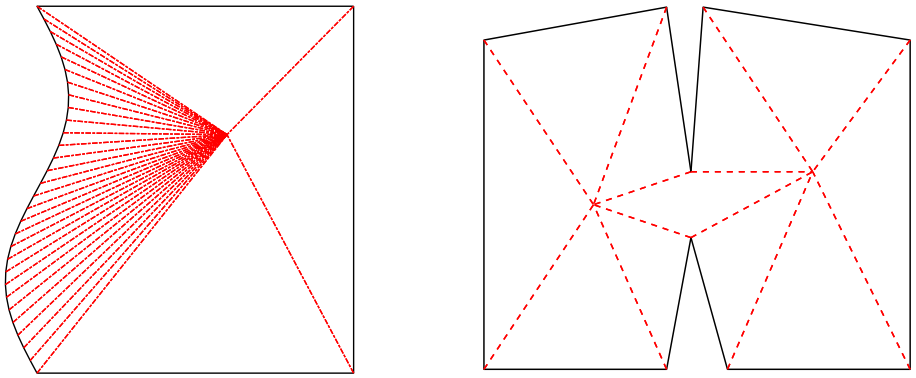


FIGURE 4.1: Polygons with a lot of tiny faces (left); star shaped polygon (right).

4.3 DGFEMs for elliptic PDEs on polytopic meshes with arbitrary number of faces

We recall the second mesh assumption.

Assumption 4.3.1 (Unbounded number of faces)**.** *For any $\kappa \in \mathcal{T}_h$, there exists a set of non-overlapping d-dimensional simplices $\{\kappa_\flat^F\}_{F\subset\partial\kappa} \subset \mathcal{F}_\flat^\kappa$ contained in κ, such that for all $F \subset \partial\kappa$, and*

$$h_\kappa \leq C_s \frac{d|\kappa_\flat^F|}{|F|}, \tag{4.18}$$

with $C_s > 0$ constant independent of the discretization parameters, the number of faces per element, and the face measures.

In Figure 4.1, we exemplify two different polygons satisfying the above mesh regularity assumption. We note that the assumption does not give any restrictions on neither the number nor the measure of the elemental faces. Indeed, shape irregular simplices $\kappa_\flat^F$, with base $|F|$ of small size compared to the corresponding height $d|\kappa_\flat^F|/|F|$, are allowed: the height, however, has to be comparable to h_κ; cf., the left polygon on Figure 4.1. Further, we note that the union of the simplices $\kappa_\flat^F$ does not need to cover the whole element κ, as in general it is sufficient to assume that

$$\cup_{F\subset\partial\kappa}\, \bar{\kappa}_\flat^F \subseteq \bar{\kappa}; \tag{4.19}$$

cf., the right polygon on Figure 4.1.

Remark 4.8. Meshes made of polytopes which are finite union of polytopes with the latter being uniformly star-shaped with respect to the largest inscribed circle will satisfy Assumption 4.3.1.

We point out that above mesh assumptions are can be viewed as a generalization of shape regularity assumption over polytopic meshes. It is easy to see that mesh assumption 4.3.1 is equivalent to the classical shape regularity assumptions in the sense of [70] for simplical meshes or tensor product meshes, if we take $\rho_\kappa = \min_{F\subset\partial\kappa} d|\kappa_\flat^F|/|F|$, where ρ_κ denotes radius of largest inscribed ball.

We will simplify the inverse estimate and polynomial approximation results for trace term based on Assumption 4.3.1.

Lemma 4.9. *Let $\kappa \in \mathcal{T}_h$, and Assumption 4.3.1 holds. Then, for each $v \in \mathcal{P}_p(\kappa)$, the following inverse inequality holds*

$$\|v\|^2_{L^2(\partial\kappa)} \leq C_s \frac{(p+1)(p+d)}{h_\kappa} \|v\|^2_{L^2(\kappa)}. \tag{4.20}$$

Constant C_s is defined in 4.18, independent of $|\kappa|/\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|$, $|F|$, p, and v.

Proof. The proof is straightforward applying inverse estimate (3.4) over each simplex $\kappa_\flat^F$ inside the κ, combined with relation (4.19).

$$\begin{aligned}\|v\|^2_{L^2(\partial\kappa)} &\leq \sum_{F\subset\partial\kappa} \frac{(p+1)(p+d)}{d}\frac{|F|}{|\kappa_\flat^F|}\|v\|^2_{L^2(\kappa_\flat^F)} \\ &= \sum_{F\subset\partial\kappa} C_s\frac{(p+1)(p+d)}{h_\kappa}\|v\|^2_{L^2(\kappa_\flat^F)} \leq C_s\frac{(p+1)(p+d)}{h_\kappa}\|v\|^2_{L^2(\kappa)}.\end{aligned}$$

Here, $\kappa_\flat^F \in \mathcal{F}_\flat^\kappa$ is defined in Definition 3.2, the proof is complete. □

Lemma 4.10. *Let $\kappa \in \mathcal{T}_h$, $F \subset \partial\kappa$ denote one of its faces, and $\mathcal{K} \in \mathcal{T}_h^\sharp$ the corresponding simplex such that $\kappa \subset \mathcal{K}$, cf. Definition 3.9. Suppose that $v \in L^2(\Omega)$ is such that $\mathfrak{E}v|_\mathcal{K} \in H^{l_\kappa}(\mathcal{K})$, for some $l_\kappa \geq 0$. Then, given Assumption 4.3.1 is satisfied, the following bound holds*

$$\|v - \tilde{\Pi}_p v\|_{L^2(\partial k)} \leq C_{\mathrm{I},5}\frac{h_\kappa^{s_\kappa-1/2}}{p^{l_\kappa-1/2}}\|\mathcal{E}v\|_{H^{l_\kappa}(\mathcal{K})},\ l_\kappa > 1/2, \tag{4.21}$$

where $s_\kappa = \min\{p+1, l_\kappa\}$ and $C_{\mathrm{I},5}$ are positive constants, which depend on constant C_s defined in 4.18 and shape-regularity of $\mathcal{K}$, but are independent of v, h_κ, p, and number of faces per element.

Proof. By employing Assumption 4.3.1, relation (4.18), (4.19), the multiplicative trace inequality 3.13 over simplices, arithmetic mean inequality, and bounds (3.28), we have

$$\begin{aligned}
\|v-\tilde{\Pi}_p v\|^2_{L^2(\partial\kappa)} &\leq \sum_{F\subset\partial\kappa} C_{\mathrm{t}}|F|\Big(\frac{1}{|\kappa_\flat^F|}\|v-\tilde{\Pi}_p v\|^2_{L^2(\kappa_\flat^F)} \\
&\quad +h_{\kappa_\flat^F}|\kappa_\flat^F|^{-1}\|v-\tilde{\Pi}_p v\|_{L^2(\kappa_\flat^F)}\|\nabla(v-\tilde{\Pi}_p v)\|_{L^2(\kappa_\flat^F)}\Big) \\
&\leq C_{\mathrm{t}}C_s d\sum_{F\subset\partial\kappa}\Big(\frac{1}{h_\kappa}\|v-\tilde{\Pi}_p v\|^2_{L^2(\kappa_\flat^F)} \\
&\quad +\|v-\tilde{\Pi}_p v\|_{L^2(\kappa_\flat^F)}\|\nabla(v-\tilde{\Pi}_p v)\|_{L^2(\kappa_\flat^F)}\Big) \\
&\leq C_{\mathrm{t}}C_s d\sum_{F\subset\partial\kappa}\Big(\frac{(p+1)}{h_\kappa}\|v-\tilde{\Pi}_p v\|^2_{L^2(\kappa_\flat^F)} \\
&\quad +\frac{h_k}{4p}\|\nabla(v-\tilde{\Pi}_p v)\|^2_{L^2(\kappa_\flat^F)}\Big) \\
&\leq C_{\mathrm{t}}C_s d\Big(\frac{(p+1)}{h_\kappa}\|v-\tilde{\Pi}_p v\|^2_{L^2(\kappa)}+\frac{h_k}{4p}\|\nabla(v-\tilde{\Pi}_p v)\|^2_{L^2(\kappa)}\Big) \\
&\leq C_{\mathrm{t}}C_s d C^2_{\mathrm{I},1}\,C^{2s_\kappa-1}_{\mathrm{diam}}\frac{h_\kappa^{2s_\kappa-1}}{p^{2l_\kappa-1}}\|\mathcal{E}v\|^2_{H^{l_\kappa}(\mathcal{K})}. \qquad (4.22)
\end{aligned}$$

Here, we have

$$C_{\mathrm{I},5}=C_{\mathrm{I},1}\,C^{s_\kappa-1/2}_{\mathrm{diam}}\sqrt{C_{\mathrm{t}}C_s d}.$$

□

Remark 4.11. We point out that the above two bounds are both independent of number of faces per element and measure of faces. The idea behind the above two bounds are simple. We are not applying the inverse estimate and approximation results from each individual faces F to the whole element κ, but applying those results from element boundary $\partial\kappa$ to the element κ. With this approach, we do not need to consider the L^∞–norm bounds for inverse estimate and approximation result. So if the mesh κ satisfies Assumption 4.3.1, each individual faces F is allowed to have arbitrarily small measure and each element κ is allowed to have arbitrary number of faces.

4.3.1 The stability and a priori error bound of IP DGFEM

Based on the above new mesh Assumption 4.3.1, we will derive the coercivity and continuity of proposed IP DGFEM. In this section, we will assume the diffusion tensor a is a general function which satisfies the *uniform ellipticity condition* (4.4).

Next, we shall often make use of the arithmetic–geometric mean inequality, together with inverse estimate (4.20) and relation (4.4)

Lemma 4.12. *Under Assumption 4.3.1 and with the discontinuity penalization function $\sigma : \mathcal{F}_h \to \mathbb{R}$ is given by*

$$\sigma(x) = \begin{cases} C_\sigma \max\limits_{\kappa\in\{\kappa_i,\kappa_j\}} \left\{ \dfrac{\bar{a}_\kappa^2(p_\kappa+1)(p_\kappa+d)}{h_\kappa} \right\}, & F \in \mathcal{F}_h^{\mathcal{I}},\ F = \partial\kappa_i \cap \partial\kappa_j, \\ \\ C_\sigma \dfrac{\bar{a}_\kappa^2(p_\kappa+1)(p_\kappa+d)}{h_\kappa}, & F \in \mathcal{F}_h^{\mathcal{D}},\ F \subset \partial\kappa, \end{cases} \tag{4.23}$$

with $C_\sigma > 0$ sufficiently large, independent of discretization parameters and the number of faces per element. The bilinear form given by (4.8) *is coercive and continuous, that is*

$$\tilde{B}_{\mathrm{d}}(v,v) \geq C_{\mathrm{coer}} |||v|||_{\mathrm{DG}}^2 \quad \textit{for all} \quad v \in \mathcal{S}, \tag{4.24}$$

and

$$\tilde{B}_{\mathrm{d}}(w,v) \leq C_{\mathrm{cont}} |||w|||_{\mathrm{DG}}\, |||v|||_{\mathrm{DG}} \quad \textit{for all} \quad w, v \in \mathcal{S}, \tag{4.25}$$

where C_{coer} and C_{cont} are positive constants, mesh size h_κ, polynomial order p_κ, the number of faces per element.

Proof. By employing the inverse inequality Lemma 4.9 and recalling the definition of $\tilde{B}_{\mathrm{d}}$, for any $v \in \mathcal{S}$, we have

$$\tilde{B}_{\mathrm{d}}(v,v) \;=\; |||v|||_{\mathrm{DG}}^2 - 2 \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \{\!\{ a\boldsymbol{\Pi}_2(\nabla v) \}\!\} \cdot [\![v]\!]\, \mathrm{d}s, \tag{4.26}$$

and it remains to bound the second term on the right-hand side. For $F \in \mathcal{F}_h^{\mathcal{I}}$, the Cauchy–Schwarz inequality and arithmetic–geometric mean inequality imply

$$\begin{aligned} \int_F \{\!\{ a\boldsymbol{\Pi}_2(\nabla v) \}\!\} \cdot [\![v]\!]\, \mathrm{d}s \;&\leq\; \frac{1}{2} \left(\| \frac{1}{\sqrt{\sigma}} a\boldsymbol{\Pi}_2(\nabla v^+) \|_{L^2(F)} + \| \frac{1}{\sqrt{\sigma}} a\boldsymbol{\Pi}_2(\nabla v^-) \|_{L^2(F)} \right) \\ &\quad \times \| \sqrt{\sigma} [\![v]\!] \|_{L^2(F)} \\ &\leq\; \epsilon \left(\| \frac{1}{\sqrt{\sigma}} a\boldsymbol{\Pi}_2(\nabla v^+) \|_{L^2(F)}^2 + \| \frac{1}{\sqrt{\sigma}} a\boldsymbol{\Pi}_2(\nabla v^-) \|_{L^2(F)}^2 \right) \\ &\quad + \frac{1}{8\epsilon} \| \sqrt{\sigma} [\![v]\!] \|_{L^2(F)}^2 . \end{aligned}$$

Similarly, for $F \in \mathcal{F}_h^{\mathcal{D}}$, we have that

$$\int_F \{\!\{a\mathbf{\Pi}_2(\nabla v)\}\!\} \cdot [\![v]\!] \, \mathrm{d}s \leq \epsilon \| \frac{1}{\sqrt{\sigma}} a\mathbf{\Pi}_2(\nabla v^+) \|^2_{L^2(F)} + \frac{1}{4\epsilon} \| \sqrt{\sigma} [\![v]\!] \|^2_{L^2(F)}.$$

Using the inverse inequality Lemma 4.9, the definition of the interior penalty parameter σ, the uniform ellipticity condition (4.4) of diffusion tensor, and the L^2-stability of the projector $\mathbf{\Pi}_2$, we conclude that

$$\begin{aligned}
&\sum_{F \in \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}} \int_F \{\!\{a\mathbf{\Pi}_2(\nabla v)\}\!\} \cdot [\![v]\!] \, \mathrm{d}s \\
&\leq \epsilon \sum_{\kappa \in \mathcal{T}_h} \sum_{F \subset \partial\kappa} \| \frac{1}{\sqrt{\sigma}} a\mathbf{\Pi}_2(\nabla v) \|^2_{L^2(F)} + \frac{1}{4\epsilon} \sum_{F \in \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}} \| \sqrt{\sigma} [\![v]\!] \|^2_{L^2(F)} \\
&\leq \epsilon \sum_{\kappa \in \mathcal{T}_h} \sum_{F \subset \partial\kappa} \sigma^{-1} \bar{a}_\kappa^2 \frac{(p_\kappa + 1)(p_\kappa + d)}{d} \frac{|F|}{|\kappa_\flat^F|} \| \mathbf{\Pi}_2(\nabla v) \|^2_{L^2(\kappa_\flat^F)} \\
&+ \frac{1}{4\epsilon} \sum_{F \in \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}} \| \sqrt{\sigma} [\![v]\!] \|^2_{L^2(F)} \\
&\leq \frac{\epsilon C_s}{\theta C_\sigma} \sum_{\kappa \in \mathcal{T}_h} \| \sqrt{a} \nabla v \|^2_{L^2(\kappa)} + \frac{1}{4\epsilon} \| \sqrt{\sigma} [\![v]\!] \|^2_{L^2(F)}. \qquad (4.27)
\end{aligned}$$

Inserting (4.27) into (4.26) immediately gives

$$\tilde{B}_{\mathrm{d}}(v, v) \;\geq\; \left(1 - \frac{2\epsilon C_s}{\theta C_\sigma}\right) \sum_{\kappa \in \mathcal{T}_h} \| \sqrt{a} \nabla v \|^2_{L^2(\kappa)} + \left(1 - \frac{1}{2\epsilon}\right) \sum_{F \in \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}} \| \sqrt{\sigma} [\![v]\!] \|^2_{L^2(F)},$$

as generalized shape regularity constant C_s is uniformly bounded by Assumption 4.3.1. Hence the bilinear form $\tilde{B}_{\mathrm{d}}(\cdot, \cdot)$ is coercive over $\mathcal{S} \times \mathcal{S}$ if $C_\sigma > 2C_s\epsilon/\theta$ for some $\epsilon > 1/2$. C_σ depends on the constant C_s, but is independent of number of faces per element.

The continuity of $\tilde{B}_{\mathrm{d}}(\cdot, \cdot)$ easily follows by applying the Cauchy-Schwarz inequality and then bounding the face terms by repeating the arguments leading to (4.27). □

Remark 4.13. The coercivity constant may depend on the shape regularity constant C_s and on the uniform ellipticity constant θ. To avoid the dependence on the latter, it is possible to combine the present developments with the DGFEMs proposed in [105]; we refrain from doing so here, in the interest of simplicity of the presentation.

Remark 4.14. We point out that the stability of SIP-DGFEM may be lost when the diffusion tensor a has a high contrast. In that case, we should modify our method by using diffusivity-dependent weighted averages, see [95, 87] for details.

Finally, we will derive the *a priori* error bound under the Assumption 4.3.1 for the proposed IP DGFEM in $|||\cdot|||_{\mathrm{DG}}$. Here, we emphasize that only the error analysis related to trace term will be different in this section compared to that in Section 4.2. We detail here a different treatment of the trace terms to take advantages of the different mesh assumption used here. By employing relation (4.21) in approximation Lemma 4.10, we have

$$\begin{aligned}
&\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F \sigma[\![v-\tilde{\Pi}_p v]\!]^2\,\mathrm{d}s = \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\sigma\|[\![v-\tilde{\Pi}_p v]\!]\|^2_{L^2(F)}\\
&\le 2\sum_{\kappa\in\mathcal{T}_h}\sum_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}\sigma\|v-\tilde{\Pi}_p v\|^2_{L^2(F)} \le 2\sum_{\kappa\in\mathcal{T}_h}(\max_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}\sigma)\|v-\tilde{\Pi}_p v\|^2_{L^2(\partial\kappa)}\\
&\le C\sum_{\kappa\in\mathcal{T}_h}(\max_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}\sigma)\frac{h_\kappa^{2s_\kappa-1}}{p^{2l_\kappa-1}}\|\mathcal{E}v\|^2_{H^{l_\kappa}(\mathcal{K})},
\end{aligned}\tag{4.28}$$

where the constant $C>0$ is independent of number of faces per elements. Bounds for remaining trace and inconsistency terms can be derived in a completely analogous fashion. Then, we have the following DGFEM convergence result.

Theorem 4.15. *Let $\mathcal{T}_h=\{\kappa\}$ be a subdivision of $\Omega\subset\mathbb{R}^d$, $d=2,3$, consisting of general polytopic elements satisfying Assumption 4.3.1 and Assumption 3.3.1 with $\mathcal{T}_h^\sharp=\{\mathcal{K}\}$ an associated covering of $\mathcal{T}_h$ consisting of shape-regular d–simplexes, cf. Definition 3.9. Let $u_h\in S^{\mathbf{p}}_{\mathcal{T}_h}$, with $p_\kappa\ge 1$ for all $\kappa\in\mathcal{T}_h$, be the corresponding DGFEM solution defined by (4.5) with the discontinuity-penalization functions given by (4.23). If the exact solution $u\in H^1(\Omega)$ to (4.1)–(4.3) satisfies $u|_\kappa\in H^{l_\kappa}(\kappa)$, $l_\kappa>3/2$, for each $\kappa\in\mathcal{T}_h$, such that $\mathfrak{E}u|_{\mathcal{K}}\in H^{l_\kappa}(\mathcal{K})$, where $\mathcal{K}\in\mathcal{T}_h^\sharp$ with $\kappa\subset\mathcal{K}$, then*

$$|||u-u_h|||^2_{\mathrm{DG}}\le C\sum_{\kappa\in\mathcal{T}_h}\frac{h_\kappa^{2(s_\kappa-1)}}{p_\kappa^{2(l_\kappa-1)}}\left(\bar{a}_\kappa+\mathcal{G}_\kappa(h_\kappa,p_\kappa)\right)\|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})},$$

where, $s_\kappa=\min\{p_\kappa+1,l_\kappa\}$,

$$\begin{aligned}
\mathcal{G}_\kappa(h_\kappa,p_\kappa) &= \bar{a}_\kappa^2 p_\kappa h_\kappa^{-1}\max_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}\sigma^{-1}+\bar{a}_\kappa^2 p_\kappa^2 h^{-1}\max_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}\sigma^{-1}\\
&+ p_\kappa^{-1}h_\kappa\max_{F\subset\partial\kappa\setminus\mathcal{F}_h^{\mathcal{N}}}\sigma,
\end{aligned}$$

and the positive constant C is independent of the discretization parameters, number of faces per element and u.

Remark 4.16. The above *a priori* error bound holds *without* any assumptions on the relative size of the spatial faces F, $F \subset \partial\kappa$, and number of faces of a given polytopic element $\kappa \in \mathcal{T}_h$, i.e., elements with arbitrarily small faces and/or arbitrary number of faces are permitted, as long as they satisfy Assumption 4.3.1. We will extend the above results in Chapter 6 for analysing the IP DGFEM for parabolic time dependent problems.

4.4 Numerical examples

We present a series of computational examples to numerically investigate the asymptotic convergence behaviour of the proposed IP DGFEM on general meshes consisting of polygonal elements. Throughout this section the IP DGFEM solution u_h is computed with the constant $C_\sigma = 10$ appearing in the interior penalty parameter.

4.4.1 Example 1

In this first example, we investigate the computational efficiency of employing the IP DGFEM on standard tensor-product elements (quadrilaterals in 2D and hexahedra in 3D) with local polynomial bases consisting of either $\mathcal{P}_p$ or $\mathcal{Q}_p$ polynomials; in the following figures, these schemes will be denoted by DGFEM(P) and DGFEM(Q), respectively. Moreover, we shall compare both IP DGFEM approaches with the standard continuous Galerkin finite element method with $\mathcal{Q}_p$ basis, denoted by FEM(Q).

Firstly, we consider the following two–dimensional Poisson problem: let $\Omega = (0,1)^2$ and select $f = 2\pi^2 \sin(\pi x)\sin(\pi y)$, so that the analytical solution to (4.1) is given by $u = \sin(\pi x)\sin(\pi y)$. In Figure 4.2 we investigate the convergence behaviour of the three schemes, namely DGFEM(P), DGFEM(Q), and FEM(Q), under p–refinement for fixed h. Here, uniform square meshes consisting of 16, 64, and 256 elements are employed; for each mesh, we plot both the $L^2(\Omega)$–norm and $H^1(\Omega)$–seminorm error against the square root of the number of degrees of freedom in the underlying finite element space, as the polynomial degree p is uniformly increased. Here, we clearly observe exponential convergence of all three methods, in

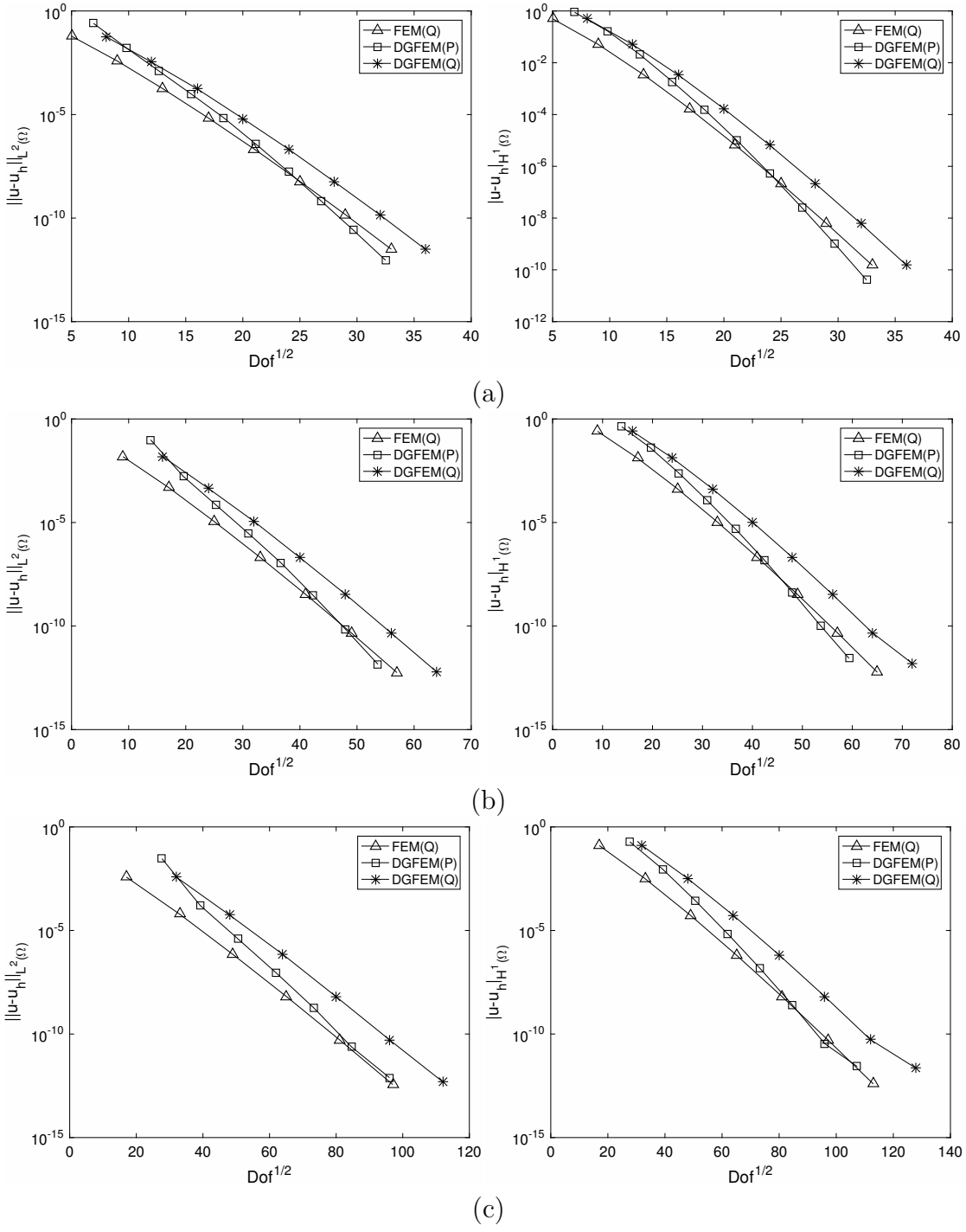


FIGURE 4.2: Example 1. Comparison between IP DGFEM exploiting local $\mathcal{Q}_p$ and $\mathcal{P}_p$ polynomial spaces with FEM under p–refinement on uniform meshes consisting of square elements on $(0,1)^2$ (2D). Left: $\|u-u_h\|_{L^2(\Omega)}$; Right: $|u-u_h|_{H^1(\Omega)}$; (a) 4×4 mesh; (b) 8×8 mesh; (c) 16×16 mesh.

the sense that, on the linear-log scale, the convergence plots become straight lines as p is increased. Moreover, we observe that the convergence lines for FEM(Q) and DGFEM(Q) are roughly parallel, with the former method being more efficient, in

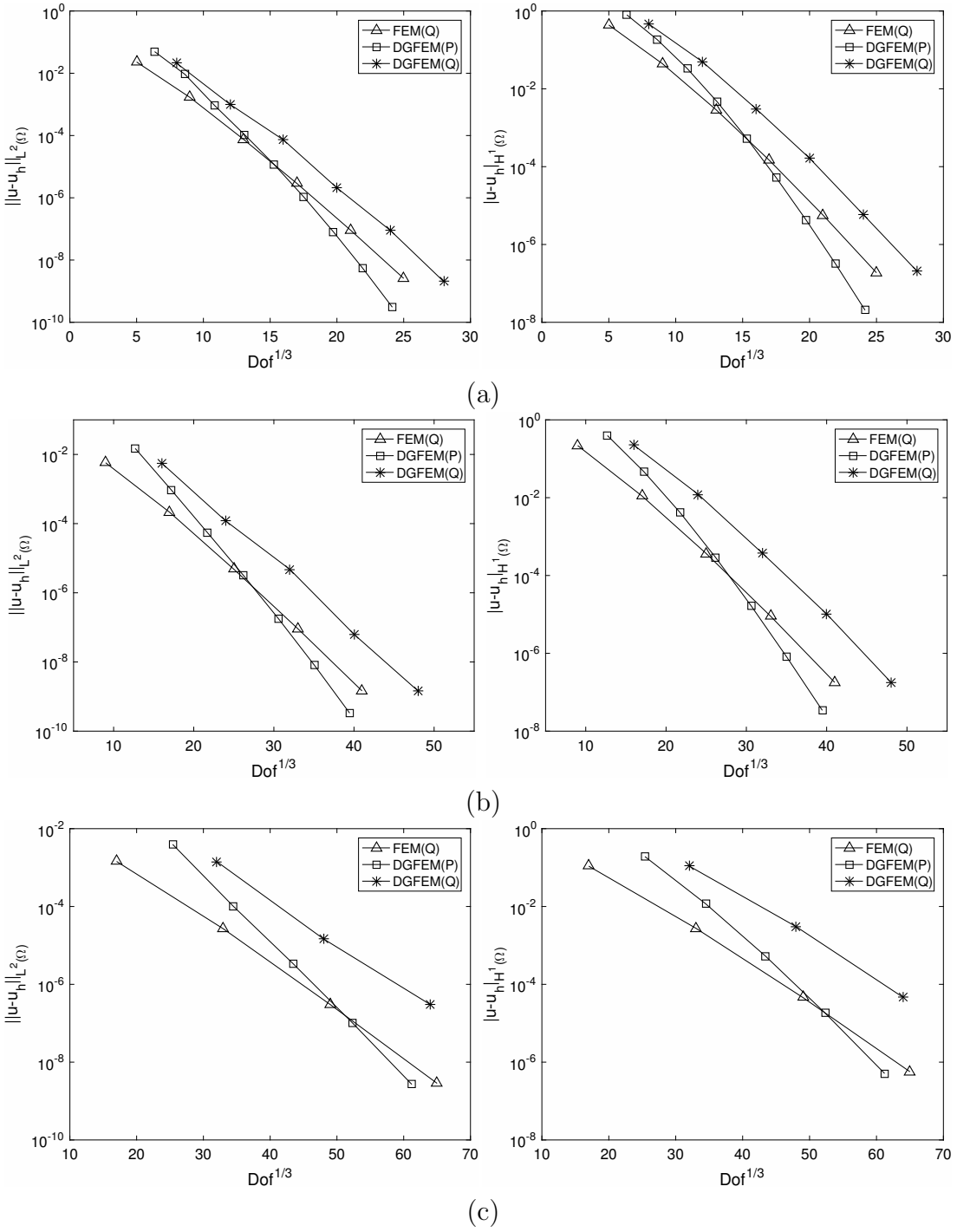


FIGURE 4.3: Example 1. Comparison between IP DGFEM exploiting local $\mathcal{Q}_p$ and $\mathcal{P}_p$ polynomial spaces with FEM under p–refinement on uniform meshes consisting of hexahedral elements on $(0,1)^3$ (3D). Left: $\|u-u_h\|_{L^2(\Omega)}$; Right: $|u-u_h|_{H^1(\Omega)}$; (a) $4\times4\times4$ mesh; (b) $8\times8\times8$ mesh; (c) $16\times16\times16$ mesh.

the sense that, for a given number of degrees of freedom (dof), the error measured with respect to both the $L^2(\Omega)$–norm and $H^1(\Omega)$–seminorm is less than the corresponding quantity computed for DGFEM(Q). However, one important observation

is that, for each mesh, the slope of the convergence line for DGFEM(P), i.e., the IP DGFEM employing local $\mathcal{P}_p$ polynomial bases, is actually steeper than the corresponding convergence line when local polynomial bases consisting of tensor-product $\mathcal{Q}_p$ polynomials are employed. Indeed, while for moderate p, we observe that the FEM(Q) method is more efficient than DGFEM(P), as the polynomial degree is increased, the convergence line for DGFEM(P) crosses the corresponding line for FEM(Q), at least on the coarser meshes.

To investigate this behaviour further, we now consider the three–dimensional variant of the above problem. To this end, we let $\Omega = (0,1)^3$ and select $f = 3\pi^2 \sin(\pi x)\sin(\pi y)\sin(\pi z)$, so that the analytical solution to (4.1) is given by $u = \sin(\pi x)\sin(\pi y)\sin(\pi z)$. In Figure 4.3 we consider the convergence of the DGFEM(P), DGFEM(Q), and FEM(Q) schemes under p–refinement on uniform hexahedral meshes consisting of 64, 512, and 4096 elements. As in the two–dimensional setting, we again observe that the convergence lines for both FEM(Q) and DGFEM(Q) are roughly parallel, with, again, the former method being more efficient in terms of leading to a smaller error for a given number of degrees of freedom. Moreover, the slope of convergence line for the DGFEM(P) scheme is not only steeper than the corresponding line for DGFEM(Q), but also that the cross–over point between DGFEM(P) becoming more efficient than FEM(Q) occurs much sooner.

We now turn our attention to investigate the asymptotic behaviour of the proposed IP DGFEM (DGFEM(P) using the introduced early notation) on a sequence of successively finer polygonal and square meshes for different values of the polynomial degree p; we point out that in both cases we employ local spaces consisting of polynomials of degree at most p on each element $\kappa \in \mathcal{T}_h$. The polygonal meshes are generated using the general-purpose mesh generator `PolyMesher`, cf. [179]. Typical meshes generated by `PolyMesher` are shown in Figure 4.4.

Here, we again consider the 2D Poisson example, we let $\Omega = (0,1)^2$ and select $f = 2\pi^2 \sin(\pi x)\sin(\pi y)$, so that $u = \sin(\pi x)\sin(\pi y)$. In Figure 4.5 we plot the error, measured in terms of both the $L^2(\Omega)$–norm and the DG–norm $|||\cdot|||_{\mathrm{DG}}$, against the square root of the number of degrees of freedom in the underlying finite element space $S^{\mathbf{p}}_{\mathcal{T}_h}$ for (uniform) p between 1 and 5. We clearly observe that the error $\|u - u_h\|_{L^2(\Omega)}$ and $|||u - u_h|||_{\mathrm{DG}}$ converge to zero at the optimal rates $\mathcal{O}(h^{p+1})$ and $\mathcal{O}(h^p)$, respectively, as the mesh size h tends to zero for each (fixed) p; these latter results clearly confirm the optimality of Theorem 4.6. In particular, we observe

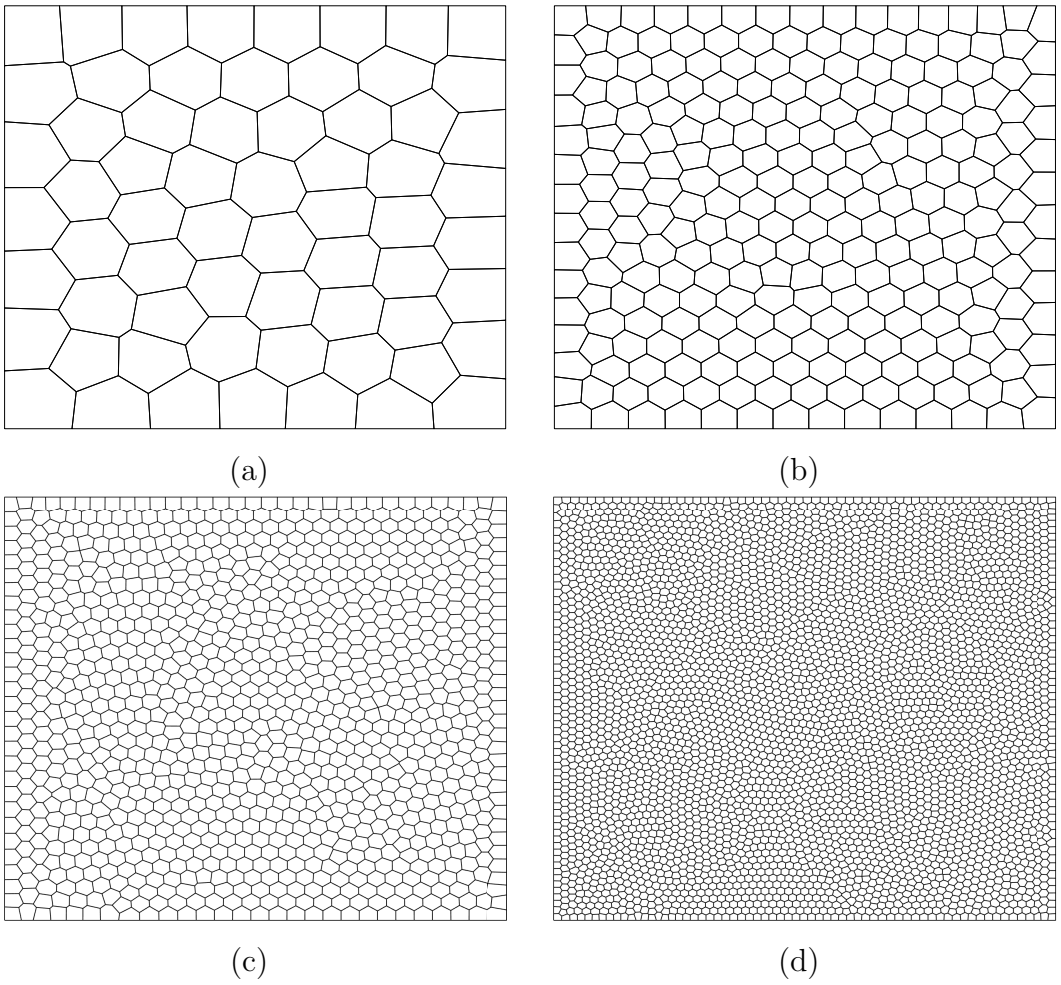
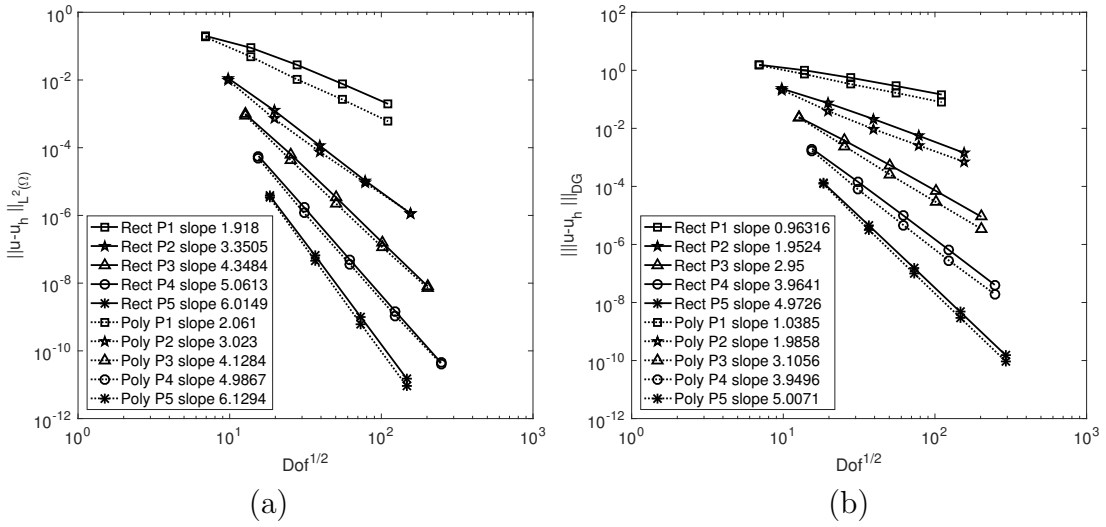


FIGURE 4.4: Example 1. Polygonal element meshes generated using `PolyMesher`. (a) Mesh with 64 elements; (b) Mesh with 256 elements; (c) Mesh with 1024 elements; (d) Mesh with 4096 elements.

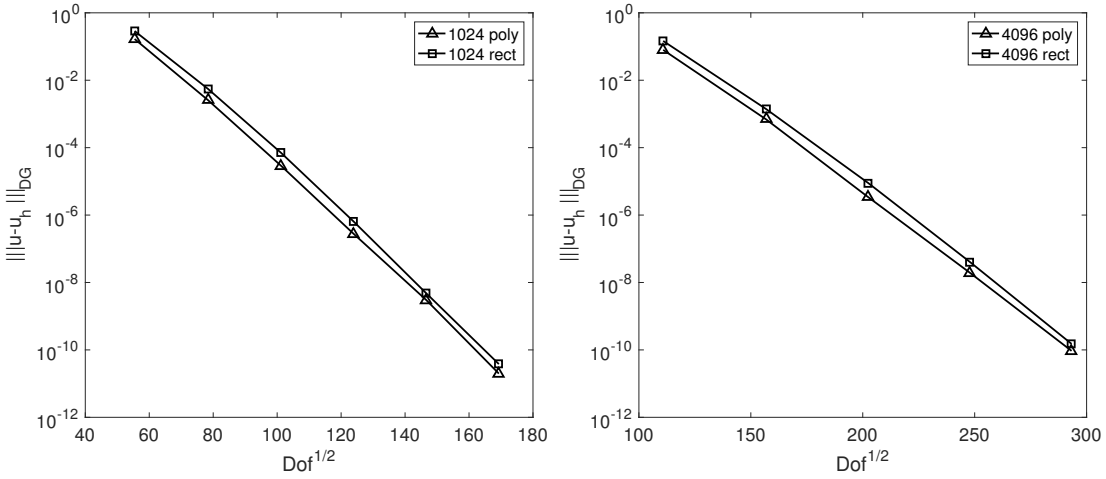
that the error in the underlying IP DGFEM is smaller when polygonal elements are employed, when compared to the corresponding quantity computed based on exploiting either uniform square elements; this behaviour is more pronounced when the error is computed with respect to the DG–norm.

We remark that similar behaviour was observed in [124] when the DG–norm of the error was computed on irregular quadrilateral meshes constructed by randomly splitting each of the interior nodes by a displacement of up to 10% of the local mesh size. As in [124], we attribute the improvement in the computed error, when polygonal elements are employed, to the increase in interelement communication. Indeed, uniform square elements may only communicate with their four immediate neighbours, while polygonal elements possess a much greater stencil due to the increase in the number of local element faces.



(a) (b)

FIGURE 4.5: Example 1. Convergence of the IP DGFEM with $\mathcal{P}_p$ basis under h–refinement: (a) $\|u - u_h\|_{L^2(\Omega)}$; (b) $|||u - u_h|||_{\mathrm{DG}}$.



(a) (b)

FIGURE 4.6: Example 1. Convergence of the IP DGFEM with $\mathcal{P}_p$ basis under p–refinement in DG–norm: (a) 1024 elements; (b) 4096 elements.

Finally, we investigate the convergence of the IP DGFEM under p–refinement for fixed h. To this end, in Figure 4.6 we plot the DG–norm of the error against number of degrees of freedom on rectangle and polygonal meshes. In each case, we observe that on the linear-log scale, the convergence plots become straight lines as the degree of the approximating polynomial is increased, thereby indicating exponential convergence in p.

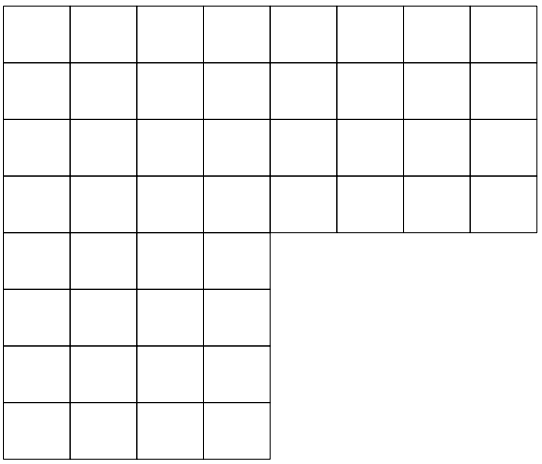


FIGURE 4.7: Example 2: Uniform square mesh, consisting of 48 elements.

4.4.2 Example 2

Following on from the previous numerical example, here we investigate the convergence behaviour of the DGFEM(P) and DGFEM(Q) approaches for a non-smooth problem on fixed computational meshes under p-refinement. To this end, we let Ω be the L-shaped domain $(-1,1)^2 \setminus [0,1) \times (-1,0]$. Uniform square meshes consisting of 48 elements are used, see Figure 4.7. Then, writing (r,φ) to denote the system of polar coordinates, we impose an appropriate inhomogeneous boundary condition for u so that

$$u = r^{2/3}\sin(2\varphi/3);$$

cf. [189]. We note that u is analytic in $\overline{\Omega}\setminus\{\mathbf{0}\}$, but ∇u is singular at the origin; indeed, here $u \notin H^2(\Omega)$. This example reflects the typical (singular) behaviour that solutions of elliptic boundary value problems exhibit in the vicinity of reentrant corners in the computational domain.

In fact, $u \in H^{\frac{5}{3}-\epsilon}(\Omega)$, $\epsilon > 0$ an arbitrary small real number. We investigate the convergence rate of the DGFEM(P) and DGFEM(Q) under p-refinement for this problem. In Table 4.1, we list the DG–norm error and also the convergence rate of DGFEM(P) and DGFEM(Q) with polynomial order $p = 1,\ldots,40$. We point out that due to the singularity at the origin, geometrically graded quadrature points towards the origin are used in order to get the desired accuracy. As we can see, the convergence rate in p for both DGFEM(P) and DGFEM(Q) is approximately:

$$|||u-u_h|||_{\mathrm{DG}} \leq Cp^{-\frac{4}{3}},$$

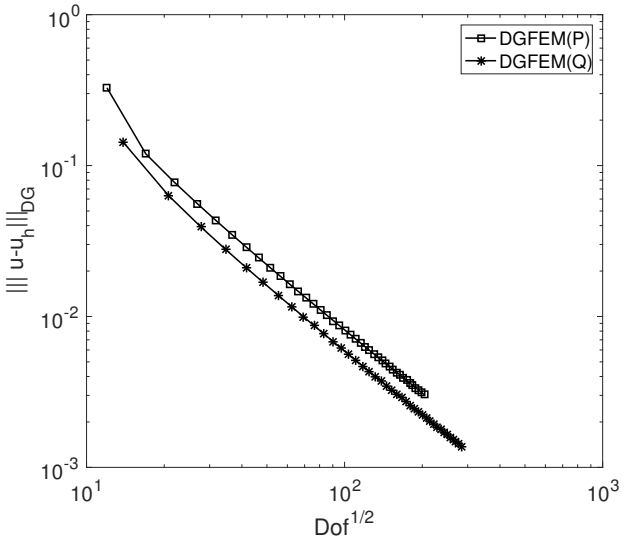


FIGURE 4.8: Example 2: Convergence of the IP DGFEM with $\mathcal{P}_p$ and $\mathcal{Q}_p$ basis under p–refinement in DG norm.

where the constant C is independent of p. The convergence rate in p is double the theoretical rate in Theorem 4.15. This is the *doubling order* convergence in the p-version finite element, see [25] for details. The reason of this doubling order convergence in p is related to the fact that Sobolev space can not optimally characterize the singularity of $r^\gamma \log^\nu r$ type, $\gamma \in \mathbb{R}^+$, $\nu \in \mathbb{N}$; indeed from [21, 22], we know that the modified Jacobi-weighted Besov spaces provide a sharper function space setting to characterize such singular functions.

Finally, we present comparisons for error against Dofs between DGFEM(P) and DGFEM(Q) under under p–refinement for fixed h. In Figure 4.8, observe linear convergence on the log-log scale between DG–norm error and Dofs, which shows that the convergence rate is only algebraic. Interestingly, the convergence of DGFEM(P) is as steep as the convergence of DGFEM(Q), and DGFEM(P) is always larger by a fixed constant. This situation is quite different from the smooth example.

p	DGFEM(P)		DGFEM(Q)		Ratio of Error
	$\|\|\|u-u_h\|\|\|_{\mathrm{DG}}$	p-rate	$\|\|\|u-u_h\|\|\|_{\mathrm{DG}}$	p-rate	DG(P)/DG(Q) of error
1	3.28E-01		1.43E-01		2.2865
2	1.20E-01	1.4538	6.33E-02	1.1791	1.8899
3	7.74E-02	1.0726	3.93E-02	1.1764	1.9712
4	5.61E-02	1.1209	2.77E-02	1.2081	2.0213
5	4.32E-02	1.1656	2.11E-02	1.2351	2.0529
6	3.48E-02	1.1927	1.67E-02	1.2557	2.0766
7	2.88E-02	1.213	1.38E-02	1.2714	2.0954
8	2.45E-02	1.2312	1.16E-02	1.2837	2.1101
9	2.11E-02	1.2462	9.96E-03	1.2933	2.1218
10	1.85E-02	1.2576	8.68E-03	1.301	2.1316
11	1.64E-02	1.2674	7.67E-03	1.3071	2.1397
12	1.47E-02	1.2749	6.84E-03	1.3121	2.1466
13	1.33E-02	1.2816	6.16E-03	1.3162	2.1525
14	1.20E-02	1.2869	5.58E-03	1.3196	2.1578
15	1.10E-02	1.2916	5.10E-03	1.3223	2.1623
16	1.01E-02	1.2954	4.68E-03	1.3247	2.1664
17	9.37E-03	1.2989	4.32E-03	1.3266	2.1701
18	8.70E-03	1.3017	4.00E-03	1.3282	2.1733
19	8.10E-03	1.3044	3.72E-03	1.3296	2.1763
20	7.58E-03	1.3065	3.48E-03	1.3308	2.179
21	7.11E-03	1.3086	3.26E-03	1.3318	2.1815
22	6.69E-03	1.3103	3.06E-03	1.3327	2.1838
23	6.31E-03	1.3119	2.89E-03	1.3334	2.1878
24	5.97E-03	1.3133	2.73E-03	1.334	2.1878
25	5.66E-03	1.3146	2.58E-03	1.3346	2.1896
26	5.37E-03	1.3157	2.45E-03	1.335	2.1912
27	5.11E-03	1.3168	2.33E-03	1.3354	2.1928
28	4.87E-03	1.3177	2.22E-03	1.3358	2.1942
29	4.65E-03	1.3186	2.12E-03	1.3361	2.1956
30	4.45E-03	1.3193	2.02E-03	1.3363	2.1968
31	4.26E-03	1.3201	1.94E-03	1.3366	2.198
32	4.08E-03	1.3207	1.86E-03	1.3368	2.1991
33	3.92E-03	1.3214	1.78E-03	1.3369	2.2002
34	3.77E-03	1.3219	1.71E-03	1.3371	2.2012
35	3.63E-03	1.3225	1.65E-03	1.3372	2.2021
36	3.50E-03	1.3229	1.59E-03	1.3373	2.203
37	3.37E-03	1.3234	1.53E-03	1.3374	2.2039
38	3.25E-03	1.3238	1.48E-03	1.3375	2.2047
39	3.14E-03	1.3242	1.43E-03	1.3375	2.2054
40	3.04E-03	1.3245	1.38E-03	1.3376	2.2062

TABLE 4.1: Example 2: Convergence rate in p of the IP DGFEM with $\mathcal{P}_p$ and $\mathcal{Q}_p$ basis in DG–norm, and the ratio of error.

Chapter 5

DGFEMs for PDEs with Nonnegative Characteristic Form

On the basis of the *hp*–version inverse and approximation bounds developed in Chapter 3, together with the IP-DGFEM scheme for pure diffusion problems in Section 4.2 of Chapter 4, here we study the IP-DGFEM discretization of a general class of second–order PDEs with non-negative characteristic form, following the bounded number of faces per element mesh Assumption 3.1.1. The work contained in this chapter is drawn from [59].

5.1 Model problem

Given Ω a bounded Lipschitz domain in $\mathbb{R}^d$, $d \geq 1$, we consider the PDE: find u such that

$$-\nabla \cdot (a\nabla u) + \mathbf{b} \cdot \nabla u + cu \;=\; f \quad \text{in } \Omega, \tag{5.1}$$

where, $a = \{a_{ij}\}_{i,j=1}^{d}$ with $a_{ij} \in L^\infty(\Omega)$ and $a_{ij} = a_{ji}$, for $i, j = 1, \ldots, d$, $\mathbf{b} = (b_1, \ldots, b_d) \in [W^{1,\infty}(\Omega)]^d$, $c \in L^\infty(\Omega)$ and $f \in L^2(\Omega)$. The PDE (5.1) is referred to as an equation with nonnegative characteristic form on the set $\Omega \subset \mathbb{R}^d$ if, at each $\mathbf{x}$ in $\bar{\Omega}$, we have

$$\sum_{i,j=1}^{d} a_{ij}(\mathbf{x})\xi_i\xi_j \geq 0, \tag{5.2}$$

for any vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_d)$ in $\mathbb{R}^d$.

In order to supplement (5.1) with suitable boundary conditions, following [151, 126], we first subdivide the boundary $\partial\Omega$ of the computational domain Ω into appropriate subsets: we let

$$\partial_0\Omega = \left\{ \mathbf{x} \in \partial\Omega : \sum_{i,j=1}^{d} a_{ij}(\mathbf{x}) n_i n_j > 0 \right\}, \tag{5.3}$$

where $\mathbf{n} = (n_1, \ldots, n_d)$ denotes the unit outward normal vector to $\partial\Omega$. Loosely speaking, we may think of $\partial_0\Omega$ as being the 'elliptic' portion of the boundary $\partial\Omega$. On the 'hyperbolic' portion of the boundary $\partial\Omega\backslash\partial_0\Omega$, we define the inflow and outflow boundaries $\partial_-\Omega$ and $\partial_+\Omega$, respectively, in the standard manner:

$$\begin{aligned} \partial_-\Omega &= \{\mathbf{x} \in \partial\Omega\backslash\partial_0\Omega : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0\}, \\ \partial_+\Omega &= \{\mathbf{x} \in \partial\Omega\backslash\partial_0\Omega : \mathbf{b}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) > 0\}. \end{aligned} \tag{5.4}$$

If $\partial_0\Omega$ is nonempty, we shall further divide it into disjoint subsets $\partial\Omega_{\mathrm{D}}$ and $\partial\Omega_{\mathrm{N}}$ whose union is $\partial_0\Omega$, with $\partial\Omega_{\mathrm{D}}$ nonempty and relatively open in $\partial\Omega$; cf. Figure 5.1 It is evident from these definitions that $\partial\Omega = \partial\Omega_{\mathrm{D}} \cup \partial\Omega_{\mathrm{N}} \cup \partial_-\Omega \cup \partial_+\Omega$. Assuming the (physically reasonable) hypothesis that $\mathbf{b} \cdot \mathbf{n} \geq 0$ on $\partial\Omega_{\mathrm{N}}$ whenever $\partial\Omega_{\mathrm{N}}$ is nonempty, we supplement (5.1) with the following boundary conditions:

$$u = g_{\mathrm{D}} \quad \text{on } \partial\Omega_{\mathrm{D}} \cup \partial_-\Omega, \qquad \mathbf{n} \cdot (a\nabla u) = g_{\mathrm{N}} \quad \text{on } \partial\Omega_{\mathrm{N}}. \tag{5.5}$$

The extension of this setting can be found in [62].

Additionally, we assume that the following positivity hypothesis holds: there exists a constant vector $\boldsymbol{\xi} \in \mathbb{R}^d$ such that

$$c(\mathbf{x}) - \frac{1}{2} \nabla \cdot \mathbf{b}(\mathbf{x}) + \mathbf{b}(\mathbf{x}) \cdot \boldsymbol{\xi} \geq \gamma_0 \ \text{ a.e. } \ \mathbf{x} \in \Omega, \tag{5.6}$$

where $\gamma_0 > 0$ is a constant. For simplicity of presentation, following [125] we shall assume throughout that (5.6) may be satisfied with $\boldsymbol{\xi} \equiv \mathbf{0}$; we then define the positive function c_0 by

$$(c_0(\mathbf{x}))^2 = c(\mathbf{x}) - \frac{1}{2} \nabla \cdot \mathbf{b}(\mathbf{x}) \ \text{ a.e. } \ \mathbf{x} \in \Omega. \tag{5.7}$$

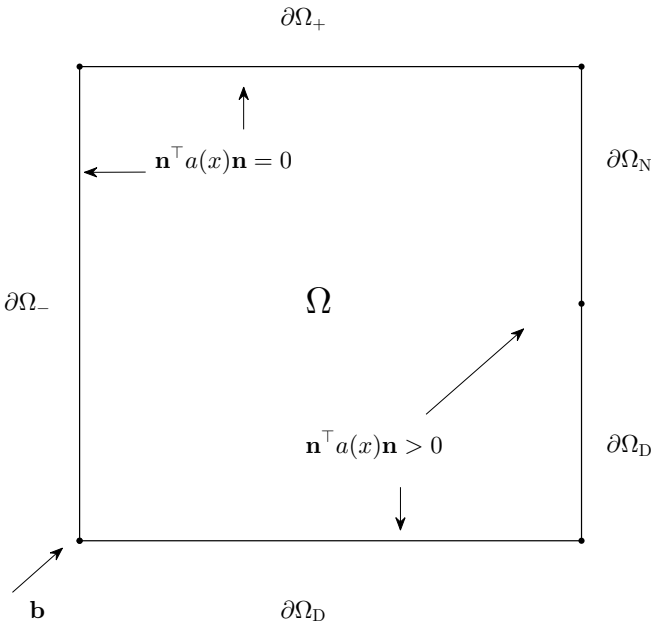


FIGURE 5.1: Boundary Conditions

The well–posedness of the boundary value problem (5.1), (5.5), in the case of homogeneous boundary conditions, has been studied in [126], cf. also [151].

In next section, we will introduce the IP-DGFEM discretization of (5.1), (5.5). We will prove an inf-sup stability condition based on the analysis for pure diffusion problem undertaken in Sections 4.2, and we present *hp*–version *a priori* bounds for the IP-DGFEM discretization of (5.1), (5.5) in Section 5.2.

5.2 DGFEMs

In this section, we will consider the IP-DGFEM discretization of the PDE with nonnegative characteristic form introduced above. Due to the general boundary conditions (5.5), we need to overload boundary faces notation $\mathcal{F}_h^{\mathcal{B}}$ in this section.

Recalling (5.3) and (5.4), we have $\partial\Omega = \partial\Omega_{\mathrm{D}} \cup \partial\Omega_{\mathrm{N}} \cup \partial_-\Omega \cup \partial_+\Omega$. Similarly, we also define $\mathcal{F}_h^{\mathcal{B}} = \mathcal{F}_h^- \cup \mathcal{F}_h^+ \cup \mathcal{F}_h^{\mathcal{D}} \cup \mathcal{F}_h^{\mathcal{N}}$, where $\mathcal{F}_h^{\mathcal{B}}$ denotes the set of all open $(d-1)$–dimensional element faces $F \in \mathcal{F}_h$ that are contained in $\partial\Omega$. For simplicity, we assume that $\mathcal{T}_h$ respects the decomposition of $\partial\Omega$ in the sense that each $F \in \mathcal{F}_h^{\mathcal{B}}$ belongs to the interior of exactly one of $\partial_-\Omega$, $\partial_+\Omega$, $\partial\Omega_{\mathrm{D}}$ and $\partial\Omega_{\mathrm{N}}$. Hence we further

denote by $\mathcal{F}_h^-, \mathcal{F}_h^+, \mathcal{F}_h^{\mathcal{D}}, \mathcal{F}_h^{\mathcal{N}} \subset \mathcal{F}_h^{\mathcal{B}}$ as the subsets of boundary faces belonging to $\partial_-\Omega$, $\partial_+\Omega$, $\partial\Omega_{\mathrm{D}}, \partial\Omega_{\mathrm{N}}$, respectively.

Next, we define the finite element space $S^{\mathbf{p}}_{\mathcal{T}_h}$ with respect to $\mathcal{T}_h$ and $\mathbf{p}$ by

$$S^{\mathbf{p}}_{\mathcal{T}_h} := \{u \in L^2(\Omega) : u|_\kappa \in \mathcal{P}_{p_\kappa}(\kappa), \kappa \in \mathcal{T}_h\},$$

where we recall that $\mathcal{P}_p(\kappa)$ denotes the space of polynomials of total degree p on κ. We stress that, by construction, the local elemental polynomial spaces employed within the definition of $S^{\mathbf{p}}_{\mathcal{T}_h}$ are defined in the physical space, without the need to map from a given reference or canonical frame, as is typically necessary for classical finite element methods.

We introduce the following (symmetric) IP-DGFEM bilinear form

$$B(u_h, v_h) = \ell(v_h) \tag{5.8}$$

for all $v_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$. Here, the bilinear form $B(\cdot,\cdot) : S^{\mathbf{p}}_{\mathcal{T}_h} \times S^{\mathbf{p}}_{\mathcal{T}_h} \to \mathbb{R}$ is defined as the sum of two parts:

$$B(u, v) := B_{\mathrm{ar}}(u, v) + B_{\mathrm{d}}(u, v),$$

where the bilinear form $B_{\mathrm{ar}}(\cdot,\cdot)$ accounts for the advection and reaction terms:

$$\begin{aligned} B_{\mathrm{ar}}(u, v) := &\sum_{\kappa\in\mathcal{T}_h} \int_\kappa \Big(\mathbf{b}\cdot\nabla u + cu\Big) v \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}} (\mathbf{b}\cdot\mathbf{n}) \lfloor u \rfloor v^+ \,\mathrm{d}s \\ &- \sum_{\kappa\in\mathcal{T}_h} \int_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)} (\mathbf{b}\cdot\mathbf{n}) u^+ v^+ \,\mathrm{d}s. \end{aligned} \tag{5.9}$$

The bilinear form $B_{\mathrm{d}}(\cdot,\cdot)$ takes care of the diffusion term:

$$\begin{aligned} B_{\mathrm{d}}(u, v) \;:=\; &\sum_{\kappa\in\mathcal{T}_h} \int_\kappa a\nabla u\cdot\nabla v \,\mathrm{d}\mathbf{x} + \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \sigma [\![u]\!]\cdot[\![v]\!] \,\mathrm{d}s \\ &- \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \Big(\{\!\{a\nabla u\}\!\}\cdot[\![v]\!] + \{\!\{a\nabla v\}\!\}\cdot[\![u]\!] \Big) \,\mathrm{d}s. \end{aligned} \tag{5.10}$$

Furthermore, the linear functional $\ell : S^{\mathbf{p}}_{\mathcal{T}_h} \to \mathbb{R}$ is defined by

$$\begin{aligned} \ell(v) \;:=\; &\sum_{\kappa\in\mathcal{T}_h} \int_\kappa f v \,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h} \int_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)} (\mathbf{b}\cdot\mathbf{n}) g_{\mathrm{D}} v^+ \,\mathrm{d}s \\ &- \sum_{F\in\mathcal{F}_h^{\mathcal{D}}} \int_F g_{\mathrm{D}} \Big((a\nabla v)\cdot\mathbf{n} - \sigma v \Big) \,\mathrm{d}s + \sum_{F\in\mathcal{F}_h^{\mathcal{N}}} \int_F g_{\mathrm{N}} v \,\mathrm{d}s. \end{aligned} \tag{5.11}$$

The nonnegative function $\sigma \in L_\infty(\mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}})$ appearing in (5.10) and (5.11) is the same discontinuity-penalization functions defined in Definition 4.7, which plays an important role for proving the inf-sup stability of the proposed DGFEM in next section.

5.2.1 Inf-Sup Stability of IP-DGFEMs

In this section, the diffusion tensor a is assumed to satisfy (4.6), i.e.,

$$a \in [V^0(\mathcal{T}_h)]^{d\times d}_{\mathrm{sym}}. \tag{5.12}$$

Moreover, we are going to use the discontinuity penalisation function σ as in Definition 4.1. The proof of inf-sup stability will employ an inconsistent formulation of the diffusion part of the bilinear form as in the previous section. We define, for all $u, v \in \mathcal{S} := H^1(\Omega) + S^{\mathbf{p}}_{\mathcal{T}_h}$, the bilinear form

$$\tilde{B}(u,v) := B_{\mathrm{ar}}(u,v) + \tilde{B}_{\mathrm{d}}(u,v), \tag{5.13}$$

where

$$\begin{aligned}\tilde{B}_{\mathrm{d}}(u,v) \;:=\; & \sum_{\kappa\in\mathcal{T}_h}\int_\kappa a\nabla u\cdot\nabla v\,\mathrm{d}\mathbf{x} + \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F \sigma[\![u]\!]\cdot[\![v]\!]\,\mathrm{d}s \\ & - \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F \Big(\{\!\{a\mathbf{\Pi}_2(\nabla u)\}\!\}\cdot[\![v]\!] + \{\!\{a\mathbf{\Pi}_2(\nabla v)\}\!\}\cdot[\![u]\!]\Big)\,\mathrm{d}s,\end{aligned}$$

and the linear functional $\tilde{\ell} : S^{\mathbf{p}}_{\mathcal{T}_h} \to \mathbb{R}$ by

$$\begin{aligned}\tilde{\ell}(v) \;:=\; & \sum_{\kappa\in\mathcal{T}_h}\int_\kappa fv\,\mathrm{d}\mathbf{x} - \sum_{\kappa\in\mathcal{T}_h}\int_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}(\mathbf{b}\cdot\mathbf{n})g_{\mathrm{D}}v^+\,\mathrm{d}s \\ & - \sum_{F\in\mathcal{F}_h^{\mathcal{D}}}\int_F g_{\mathrm{D}}\Big(a\mathbf{\Pi}_2(\nabla v)\cdot\mathbf{n} - \sigma v\Big)\,\mathrm{d}s + \sum_{F\in\mathcal{F}_h^{\mathcal{N}}}\int_F g_{\mathrm{N}}v\,\mathrm{d}s.\end{aligned}$$

Here, $\mathbf{\Pi}_2 : [L^2(\Omega)]^d \to [S^{\mathbf{p}}_{\mathcal{T}_h}]^d$ denotes the orthogonal L^2–projection onto $[S^{\mathbf{p}}_{\mathcal{T}_h}]^d$.

We then rewrite the discrete problem with inconsistent formulation in the equivalent form: find $u_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$ such that

$$\tilde{B}(u_h, v_h) = \tilde{l}(v_h) \quad \forall v_h \in S^{\mathbf{p}}_{\mathcal{T}_h}. \tag{5.14}$$

Note that the above IP DGFEM formulation is generally not consistent due to the discrete nature of L^2-orthogonal projector, but is consistent for $u_h, v_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$ when the diffusion tensor a is element-wise constant.

In view of the error analysis, we introduce the DGFEM–norm $|||\cdot|||_{\mathrm{DG}}$ as the sum of two parts as follows:

$$|||v|||^2_{\mathrm{DG}} := |||v|||^2_{\mathrm{ar}} + |||v|||^2_{\mathrm{d}},$$

where

$$|||v|||^2_{\mathrm{ar}} := \sum_{\kappa\in\mathcal{T}_h} \Big(\|c_0 v\|^2_{L^2(\kappa)} + \frac{1}{2}\|v^+\|^2_{\partial_-\kappa\cap(\mathcal{F}^{\mathcal{D}}_h\cup\mathcal{F}^-_h)} + \frac{1}{2}\|v^+ - v^-\|^2_{\partial_-\kappa\setminus\mathcal{F}^{\mathcal{B}}_h} + \frac{1}{2}\|v^+\|^2_{\partial_+\kappa\cap\mathcal{F}^{\mathcal{B}}_h}\Big), \tag{5.15}$$

with c_0 as in (5.7), and

$$|||v|||^2_{\mathrm{d}} := \sum_{\kappa\in\mathcal{T}_h} \|\sqrt{a}\nabla v\|^2_{L^2(\kappa)} + \sum_{F\in\mathcal{F}^{\mathcal{I}}_h\cup\mathcal{F}^{\mathcal{D}}_h} \int_F \sigma|[\![v]\!]|^2\,\mathrm{d}s. \tag{5.16}$$

Here, $\|\cdot\|_\tau$, $\tau\subset\partial\kappa$, denotes the (semi)norm associated with the (semi)inner product $(v,w)_\tau = \int_\tau |\mathbf{b}\cdot\mathbf{n}|vw\,\mathrm{d}s$.

The following relation holds

$$B_{\mathrm{ar}}(v,v) = |||v|||^2_{\mathrm{ar}}, \tag{5.17}$$

for all $v\in\mathcal{S}$, cf. [125]. The continuity and coercivity of the inconsistent diffusion bilinear form $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$, with respect to the DGFEM–norm $|||\cdot|||_{\mathrm{d}}$, is established in Lemma 4.2.

Before we prove the inf-sup condition, we briefly discuss the reasons why the inf-sup condition is essential. The *hp*–version *a priori* error analysis presented in [125] relies on the derivation of optimal *hp*–approximation results for the trace of the local L^2–projection operator on a given face of an element κ in the finite element mesh $\mathcal{T}_h$; cf. also [69, 144] for analogous results on simplices. Due to the lack of analogous *hp*–approximation results for the local L^2–projection operator on polytopic elements, it is not possible to directly generalise the analysis from [125] to meshes consisting of such elements. To address this issue we prove an inf-sup condition for the inconsistent bilinear form $\tilde{B}(\cdot,\cdot)$, with respect to the following streamline diffusion DGFEM–norm.

Definition 5.1. The streamline diffusion DGFEM–norm is defined by:

$$|||v|||_{\mathrm{s}}^2 := |||v|||_{\mathrm{DG}}^2 + \sum_{\kappa\in\mathcal{T}_h} \tau_\kappa \|\mathbf{b}\cdot\nabla v\|_{L^2(\kappa)}^2, \tag{5.18}$$

where

$$\tau_\kappa := \min\left\{\frac{1}{\|\mathbf{b}\|_{L^\infty(\kappa)}}, \frac{1}{\tilde{\sigma}_\kappa}\right\}\frac{h_\kappa^\perp}{p_\kappa^2} \quad \forall\kappa\in\mathcal{T}_h, \tag{5.19}$$

for $p_\kappa \geq 1$, and $\tilde{\sigma}_\kappa$ is given by

$$\tilde{\sigma}_\kappa \;:=\; C_\sigma \max_{F\subset\partial\kappa}\Big\{ \max_{\substack{\tilde{\kappa}\in\{\kappa,\kappa'\}\\ F\subset\partial\kappa\cap\partial\kappa'}} \Big\{C_{\mathrm{inv},4}\frac{\bar{a}_{\tilde{\kappa}}p_{\tilde{\kappa}}^2}{h_{\tilde{\kappa}}^\perp}d\Big\}\Big\} \quad \forall\kappa\subset\mathcal{T}_h,\quad d=2,3, \tag{5.20}$$

where $C_{\mathrm{inv},4}$ is defined as in Lemma 3.5. The constant $\tilde{\sigma}_\kappa$ may be zero locally where $\bar{a}_\kappa = 0$; in this case it is understood that τ_κ takes the value of the first term in (5.19). Further, the mesh parameter $h_\kappa^\perp$ is defined as follows:

$$h_\kappa^\perp := \min_{F\subset\partial\kappa}\frac{\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|}{|F|}d \qquad \forall\kappa\in\mathcal{T}_h,\quad d=2,3, \tag{5.21}$$

with $\kappa_\flat^F$ as in Definition 3.2. We further deduce the relation

$$h_\kappa^\perp \leq h_\kappa. \tag{5.22}$$

Remark 5.2. We recall from Definition 3.2 that $\kappa_\flat^F$ denotes the family of simplices contained in κ and sharing a face F with κ. From the geometrical property of d–dimensional simplices, it is easy to see that $h_\kappa^\perp$ is the minimum over all faces F, $F\subset\partial\kappa$, of the maximum of the set of all heights of the d–dimensional simplices $\kappa_\flat^F$ sharing a $(d-1)$–dimensional face F with κ.

Remark 5.3. We note that τ_κ can be viewed as an indicator function for each element $k\in\mathcal{T}_h$, which measures the length scale of convection and diffusion over each element. If κ is in the advection dominated regime, then τ_κ takes the first term in the bracket. On the other hand, κ is in the diffusion dominated regime if τ_κ takes the second term in the bracket. By using this choice of τ_κ, the resulting inf-sup stability condition holds in both regimes.

Remark 5.4. With no loss of generality, the case $p_\kappa = 0$, relevant to the hyperbolic regime, is excluded from Definition 5.1 and throughout this chapter. However, if the underlying problem is strictly hyperbolic and $p_\kappa = 0$ is selected for all $\kappa\in\mathcal{T}_h$, then the streamline diffusion DGFEM–norm reduces to the advection-reaction

DGFEM–norm $|||\cdot|||_{\mathrm{ar}}$ defined in (5.15); in this setting, the proceeding analysis is trivial.

By employing the definition of $h_\kappa^\perp$, together with an upper bound on the constant $C_{\mathrm{INV}}(p,\kappa,F)$ defined in Lemma 3.5, the inverse estimate (3.13) can be written in the following manner. For each $v \in \mathcal{P}_p(\kappa)$, $F \subset \partial\kappa$, we have

$$\begin{aligned} \|v\|^2_{L^2(F)} &\leq C_{\mathrm{INV}}(p,\kappa,F)\frac{p^2|F|}{|\kappa|}\|v\|^2_{L^2(\kappa)} \\ &\leq C_{\mathrm{inv},4}\frac{|\kappa|}{\sup_{\kappa_\flat^F\subset\kappa}|\kappa_\flat^F|}\frac{p^2|F|}{|\kappa|}\|v\|^2_{L^2(\kappa)} \leq C_{\mathrm{inv},4}\frac{p^2}{h_\kappa^\perp}d\|v\|^2_{L^2(\kappa)}. \end{aligned} \tag{5.23}$$

Further, from the definition of $\sigma|_F$ given in (4.1), in conjunction with the definition of $h_\kappa^\perp$, cf. (5.21), we deduce the following bound

$$\tilde{\sigma}_\kappa \geq \sigma|_F, \quad F \subset \partial\kappa \ \ \forall\kappa \in \mathcal{T}_h. \tag{5.24}$$

For the reminder of this work we assume the following condition on $\mathbf{b}$:

$$\mathbf{b}\cdot\nabla_h\xi \in S^{\mathbf{p}}_{\mathcal{T}_h} \quad \forall\xi \in S^{\mathbf{p}}_{\mathcal{T}_h}, \tag{5.25}$$

cf. [125]. Under the above assumption, we prove the inf-sup condition for the bilinear form $\tilde{B}(\cdot,\cdot)$, with respect to the streamline diffusion DGFEM–norm (5.18).

Theorem 5.5. *Given Assumptions 3.1.1, 3.2.1, and 3.2.2 hold, there exists a positive constant Λ_s, independent of the mesh size h and the polynomial degree p, such that:*

$$\inf_{\nu\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\sup_{\mu\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{\tilde{B}(\nu,\mu)}{|||\nu|||_{\mathrm{s}}|||\mu|||_{\mathrm{s}}} \geq \Lambda_s, \tag{5.26}$$

where the discontinuity-penalization function σ is as defined in (4.1).

Proof. For all $\nu \in S^{\mathbf{p}}_{\mathcal{T}_h}$, we select $\mu := \nu + \alpha\nu_s$, $\nu_s|_\kappa = \tau_\kappa\mathbf{b}\cdot\nabla\nu$ for all $\kappa \in \mathcal{T}_h$, where α is a positive real number, chosen sufficiently small, cf. (5.41) below. By (5.25), we note that $\mu \in S^{\mathbf{p}}_{\mathcal{T}_h}$; the theorem now follows from the two bounds:

$$|||\mu|||_{\mathrm{s}} \leq C^*|||\nu|||_{\mathrm{s}}, \tag{5.27}$$

and

$$\tilde{B}(\nu,\mu) \geq C_*|||\nu|||^2_{\mathrm{s}}, \tag{5.28}$$

with $\Lambda_s = C_*/C^*$, where C^* and C_* are positive constants, independent of h and p.

We begin by proving (5.27). We first bound each term arising in the norm $|||\cdot|||_{\rm ar}$ of ν_s, where $\nu_s|_\kappa = \tau_\kappa \mathbf{b}\cdot\nabla\nu$, $\kappa\in\mathcal{T}_h$. Employing Lemma 3.7 together with (5.19), the lower bound on c_0 given in (5.7), and inequality (5.22), gives

$$\begin{aligned}
\sum_{\kappa\in\mathcal{T}_h}\|c_0\nu_s\|^2_{L^2(\kappa)} &\le \|c_0\|^2_{L^\infty(\Omega)}\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa^2\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\\
&\le \|c_0\|^2_{L^\infty(\Omega)}\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa^2\|\mathbf{b}\|^2_{L^\infty(\kappa)}\|\nabla\nu\|^2_{L^2(\kappa)}\\
&\le \|c_0\|^2_{L^\infty(\Omega)}C_{\rm inv,5}\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa^2\frac{p_\kappa^4\|\mathbf{b}\|^2_{L^\infty(\kappa)}}{h_\kappa^2}\|\nu\|^2_{L^2(\kappa)}\\
&\le \|c_0\|^2_{L^\infty(\Omega)}\frac{C_{\rm inv,5}}{\gamma_0}\sum_{\kappa\in\mathcal{T}_h}\|c_0\nu\|^2_{L^2(\kappa)} \le C_1|||\nu|||^2_{\rm s}. \qquad (5.29)
\end{aligned}$$

Using the inverse estimate (5.23), we deduce that

$$\begin{aligned}
&\sum_{\kappa\in\mathcal{T}_h}\Big(\frac{1}{2}\|\nu_s^+\|^2_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}+\frac{1}{2}\|\nu_s^+-\nu_s^-\|^2_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}+\frac{1}{2}\|\nu_s^+\|^2_{\partial_+\kappa\cap\mathcal{F}_h^{\mathcal{B}}}\Big)\\
&\qquad\le\sum_{\kappa\in\mathcal{T}_h}\|\mathbf{b}\|_{L^\infty(\kappa)}\tau_\kappa^2\sum_{F\subset\partial\kappa}\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(F)}\\
&\qquad\le C_FC_{\rm inv,4}d\sum_{\kappa\in\mathcal{T}_h}\|\mathbf{b}\|_{L^\infty(\kappa)}\frac{p_\kappa^2}{h_\kappa^\perp}\tau_\kappa^2\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\\
&\qquad\le C_FC_{\rm inv,4}d\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\frac{p_\kappa^2\|\mathbf{b}\|_{L^\infty(\kappa)}}{h_\kappa^\perp}\Big(\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big)\le C_2|||\nu|||^2_{\rm s}. \qquad (5.30)
\end{aligned}$$

Similarly, employing relation (5.25) together with Lemma 3.7, the streamline diffusion term, cf. (5.18), can be bounded as follows:

$$\begin{aligned}
\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\|\mathbf{b}\cdot\nabla\nu_s\|^2_{L^2(\kappa)} &\le \sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\|\mathbf{b}\|^2_{L^\infty(\kappa)}\Big(\tau_\kappa^2\|\nabla(\mathbf{b}\cdot\nabla\nu)\|^2_{L^2(\kappa)}\Big)\\
&\le \sum_{\kappa\in\mathcal{T}_h}C_{\rm inv,5}\tau_\kappa^2\frac{p_\kappa^4\|\mathbf{b}\|^2_{L^\infty(\kappa)}}{h_\kappa^2}\Big(\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big)\\
&\le \sum_{\kappa\in\mathcal{T}_h}C_{\rm inv,5}\Big(\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big)\le C_3|||\nu|||^2_{\rm s}; \qquad (5.31)
\end{aligned}$$

here, we have again exploited a bound on τ_κ, $\kappa\in\mathcal{T}_h$, and (5.22), cf. above.

Secondly, we consider the diffusion component $|||\cdot|||_{\rm d}$ of the streamline diffusion DGFEM–norm of ν_s. This time, the second term on the right hand side of (5.19)

is used as an upper bound on τ_κ, $\kappa \in \mathcal{T}_h$. This, employing Lemma 3.7, the definition of $\tilde{\sigma}_\kappa$ in (5.20), and (5.22), we get

$$\begin{aligned}
\sum_{\kappa\in\mathcal{T}_h} \|\sqrt{a}\nabla\nu_s\|^2_{L^2(\kappa)} &\leq \sum_{\kappa\in\mathcal{T}_h} \bar{a}_\kappa\tau_\kappa^2\|\nabla(\mathbf{b}\cdot\nabla\nu)\|^2_{L^2(\kappa)} \\
&\leq \sum_{\kappa\in\mathcal{T}_h} C_{\mathrm{inv},5}\tau_\kappa\frac{\bar{a}_\kappa p_\kappa^4}{h_\kappa^2}\Big(\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big) \\
&\leq \sum_{\kappa\in\mathcal{T}_h} C_{\mathrm{inv},5}\frac{\bar{a}_\kappa p_\kappa^2}{\tilde{\sigma}_\kappa h_\kappa}\Big(\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big) \\
&\leq \frac{C_{\mathrm{inv},5}}{C_\sigma C_{\mathrm{inv},4} d}\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)} \\
&\equiv C_4\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)} \leq C_4|||\nu|||^2_{\mathrm{s}}. \qquad (5.32)
\end{aligned}$$

Finally, employing (5.23) and noting that $\sigma|_F \leq \tilde{\sigma}_\kappa$ for $F \subset \partial\kappa$, $\kappa \in \mathcal{T}_h$, gives

$$\begin{aligned}
\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F \sigma|[\![\nu_s]\!]|^2\,\mathrm{d}s &\leq 2\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa^2\sum_{F\subset\partial\kappa\cap(\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}})}\sigma\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(F)} \\
&\leq 2C_F C_{\mathrm{inv},4} d\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\frac{\tilde{\sigma}_\kappa p_\kappa^2}{h_\kappa^\perp}\Big(\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big) \\
&\leq C_5\sum_{\kappa\in\mathcal{T}_h}\Big(\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big) \leq C_5|||\nu|||^2_{\mathrm{s}}. \qquad (5.33)
\end{aligned}$$

Combining the above bounds, we deduce that

$$|||\nu_s|||_{\mathrm{s}} \leq \hat{C}|||\nu|||_{\mathrm{s}}, \qquad (5.34)$$

where $\hat{C} = \sqrt{C_1 + C_2 + C_3 + C_4 + C_5}$. Exploiting the triangle inequality, we have that

$$|||\mu|||_{\mathrm{s}} \leq |||\nu|||_{\mathrm{s}} + \alpha|||\nu_s|||_{\mathrm{s}} \leq (1+\alpha\hat{C})|||\nu|||_{\mathrm{s}} \equiv C^*(\alpha)|||\nu|||_{\mathrm{s}}, \qquad (5.35)$$

which gives the desired bound stated in (5.27).

Next we prove (5.28). To this end, we observe that since $\mu := \nu + \alpha\nu_s$, $\tilde{B}(\nu,\mu) = \tilde{B}(\nu,\nu) + \alpha\tilde{B}(\nu,\nu_s)$. Considering the second term $\tilde{B}(\nu,\nu_s)$ first, we note that the advection-reaction part of the bilinear form $B_{\mathrm{ar}}(\nu,\nu_s)$ is given by

$$\begin{aligned}
B_{\mathrm{ar}}(\nu,\nu_s) = \sum_{\kappa\in\mathcal{T}_h}\int_\kappa \tau_\kappa(\mathbf{b}\cdot\nabla\nu)^2 + c\nu(\tau_\kappa\mathbf{b}\cdot\nabla\nu)\,\mathrm{d}\mathbf{x} - \int_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}(\mathbf{b}\cdot\mathbf{n})\lfloor\nu\rfloor(\tau_\kappa\mathbf{b}\cdot\nabla\nu)^+\,\mathrm{d}s \\
- \int_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}(\mathbf{b}\cdot\mathbf{n})\nu^+(\tau_\kappa\mathbf{b}\cdot\nabla\nu)^+\,\mathrm{d}s. \qquad (5.36)
\end{aligned}$$

Employing Lemma 3.7, together with the lower bound on c_0^2 given in (5.7), the second term in (5.36) may be bounded as follows:

$$\begin{aligned}
|\sum_{\kappa\in\mathcal{T}_h}\int_\kappa c\nu(\tau_\kappa\mathbf{b}\cdot\nabla\nu)\,\mathrm{d}\mathbf{x}| &\leq \sum_{\kappa\in\mathcal{T}_h}\|c\|_{L^\infty(\Omega)}\|\nu\|_{L^2(\kappa)}\|\tau_\kappa\mathbf{b}\cdot\nabla\nu\|_{L^2(\kappa)}\\
&\leq \sum_{\kappa\in\mathcal{T}_h}\|c\|_{L^\infty(\Omega)}\|\nu\|_{L^2(\kappa)}\Big(C_{\mathrm{inv},5}^{1/2}\tau_\kappa\frac{p_\kappa^2\|\mathbf{b}\|_{L^\infty(\kappa)}}{h_\kappa}\|\nu\|_{L^2(\kappa)}\Big)\\
&\leq \sum_{\kappa\in\mathcal{T}_h}\frac{C_{\mathrm{inv},5}^{1/2}\|c\|_{L^\infty(\Omega)}}{\gamma_0}\|c_0\nu\|_{L^2(\kappa)}^2. \qquad (5.37)
\end{aligned}$$

To estimate the boundary terms present in (5.36), we exploit the inverse estimate (5.23), the definition of τ_κ given in (5.19), together with the Cauchy-Schwarz inequality. Then, we get

$$\begin{aligned}
&|\sum_{\kappa\in\mathcal{T}_h}\Big(\int_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}(\mathbf{b}\cdot\mathbf{n})\lfloor\nu\rfloor(\tau_\kappa\mathbf{b}\cdot\nabla\nu)^+\,\mathrm{d}s+\int_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}(\mathbf{b}\cdot\mathbf{n})\nu^+(\tau_\kappa\mathbf{b}\cdot\nabla\nu)^+\,\mathrm{d}s\Big)|\\
&\quad\leq\sum_{\kappa\in\mathcal{T}_h}\|\nu^+-\nu^-\|_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}\Big(\sum_{F\subset\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}\|\mathbf{b}\|_{L^\infty(\kappa)}^{\frac{1}{2}}\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|_{L^2(F)}\Big)\\
&\quad+\sum_{\kappa\in\mathcal{T}_h}\|\nu^+\|_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}\Big(\sum_{F\subset\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}\|\mathbf{b}\|_{L^\infty(\kappa)}^{\frac{1}{2}}\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|_{L^2(F)}\Big)\\
&\quad\leq C_F^2C_{\mathrm{inv},4}d\Big(\sum_{\kappa\in\mathcal{T}_h}\|\nu^+-\nu^-\|_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}^2+\|\nu^+\|_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}^2\Big)+\sum_{\kappa\in\mathcal{T}_h}\frac{\tau_\kappa}{4}\|\mathbf{b}\cdot\nabla\nu\|_{L^2(\kappa)}^2\\
&\quad\leq C_F^2C_{\mathrm{inv},4}d\sum_{\kappa\in\mathcal{T}_h}\Big(\|\nu^+-\nu^-\|_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}^2+\|\nu^+\|_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}^2+\|\nu^+\|_{\partial_+\kappa\cap\mathcal{F}_h^{\mathcal{B}}}^2\Big)\\
&\quad+\sum_{\kappa\in\mathcal{T}_h}\frac{\tau_\kappa}{4}\|\mathbf{b}\cdot\nabla\nu\|_{L^2(\kappa)}^2. \qquad (5.38)
\end{aligned}$$

Using (5.17), together with the bounds (5.37) and (5.38), we deduce that

$$\begin{aligned}
B_{\mathrm{ar}}(\nu,\mu)\geq\Big(1-\frac{\alpha C_{\mathrm{inv},5}^{1/2}\|c\|_{L^\infty(\Omega)}}{\gamma_0}\Big)\sum_{\kappa\in\mathcal{T}_h}\|c_0\nu\|_{L^2(\kappa)}^2+\alpha\sum_{\kappa\in\mathcal{T}_h}\Big(\tau_\kappa-\frac{\tau_\kappa}{4}\Big)\|\mathbf{b}\cdot\nabla\nu\|_{L^2(\kappa)}^2\\
+\Big(\frac{1}{2}-\alpha C_F^2C_{\mathrm{inv},4}d\Big)\sum_{\kappa\in\mathcal{T}_h}\Big(\|\nu^+-\nu^-\|_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}^2\\
+\|\nu^+\|_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}^2+\|\nu^+\|_{\partial_+\kappa\cap\mathcal{F}_h^{\mathcal{B}}}^2\Big). \qquad (5.39)
\end{aligned}$$

Next, we consider the diffusion part of the bilinear form, i.e., $\tilde{B}_{\mathrm{d}}(\nu,\nu_s)$. From the continuity of $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$ stated in (4.12), together with the bounds given in (5.32) and

(5.33), we get

$$\begin{aligned}\tilde{B}_{\mathrm{d}}(\nu,\nu_s) &\leq C_{\mathrm{cont}}|||\nu|||_{\mathrm{d}}|||\nu_s|||_{\mathrm{d}} \leq C_{\mathrm{cont}}|||\nu|||_{\mathrm{d}}\sqrt{C_4+C_5}\Big(\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}\Big)^{\frac{1}{2}}\\ &\leq (C_{\mathrm{cont}})^2(C_4+C_5)|||\nu|||^2_{\mathrm{d}}+\sum_{\kappa\in\mathcal{T}_h}\frac{\tau_\kappa}{4}\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}.\end{aligned}$$

Exploiting the coercivity of the bilinear form $\tilde{B}_d(\cdot,\cdot)$, cf. (4.11), gives

$$\tilde{B}_{\mathrm{d}}(\nu,\mu)\geq\Big(C_{\mathrm{coer}}-\alpha(C_{\mathrm{cont}})^2(C_4+C_5)\Big)|||\nu|||^2_{\mathrm{d}}-\alpha\sum_{\kappa\in\mathcal{T}_h}\frac{\tau_\kappa}{4}\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}. \quad (5.40)$$

Finally, combining (5.39) and (5.40), the following bound holds:

$$\begin{aligned}\tilde{B}(\nu,\mu) &= B_{\mathrm{ar}}(\nu,\mu)+\tilde{B}_{\mathrm{d}}(\nu,\mu)\\ &\geq\Big(1-\frac{\alpha C^{1/2}_{\mathrm{inv},5}\|c\|_{L^\infty(\Omega)}}{\gamma_0}\Big)\sum_{\kappa\in\mathcal{T}_h}\|c_0\nu\|^2_{L^2(\kappa)}+\alpha\sum_{\kappa\in\mathcal{T}_h}\Big(\tau_\kappa-\frac{\tau_\kappa}{2}\Big)\|\mathbf{b}\cdot\nabla\nu\|^2_{L^2(\kappa)}.\\ &+\Big(\frac{1}{2}-\alpha C_F^2C_{\mathrm{inv},4}d\Big)\sum_{\kappa\in\mathcal{T}_h}\Big(\|\nu^+-\nu^-\|^2_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}+\|\nu^+\|^2_{\partial_-\kappa\cap(\mathcal{F}_h^{\mathcal{D}}\cup\mathcal{F}_h^-)}+\|\nu^+\|^2_{\partial_+\kappa\cap\mathcal{F}_h^{\mathcal{B}}}\Big).\\ &+\Big(C_{\mathrm{coer}}-\alpha(C_{\mathrm{cont}})^2(C_4+C_5)\Big)\Big(\sum_{\kappa\in\mathcal{T}_h}\|\sqrt{a}\nabla\nu\|^2_{L^2(\kappa)}+\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F\sigma|[\![\nu]\!]|^2\,\mathrm{d}s\Big).\end{aligned}$$

The coefficients in front of the norms arising on the right hand side of the above bound are all positive for sufficient small α, namely if

$$\alpha<\min\left\{\frac{\gamma_0}{C^{1/2}_{\mathrm{inv},5}\|c\|_{L^\infty(\Omega)}},\frac{1}{2C_F^2C_{\mathrm{inv},4}d},\frac{C_{\mathrm{coer}}}{(C_{\mathrm{cont}})^2(C_4+C_5)}\right\}. \quad (5.41)$$

Since the constants in (5.41) are independent of the discretization parameters, we conclude that (5.28) holds as long as α is chosen according to (5.41). □

Remark 5.6. Theorem 5.5 extends the analogous result derived for DGFEMs on meshes comprising of simplices presented in [49, 57, 18] and monograph [84, Chapter 2], to general polytopic elements. It also improves those results in the sense that here the inf-sup constant Λ_s is also independent of the polynomial degree p.

Remark 5.7. The above inf-sup condition has been derived under the assumption that (5.25) holds, hence limiting the validity of the present analysis to problems with piecewise linear convection fields $\mathbf{b}$. However, an analogous inf-sup condition still holds for general $\mathbf{b}$, if we replace the test space $S^{\mathbf{p}}_{\mathcal{T}_h}$ by $W^{\mathbf{p}}_{\mathcal{T}}:=$

$\text{span}\{v+\alpha v_s,\ \ v_s|_\kappa = \tau_\kappa \mathbf{\Pi}_2(\mathbf{b}\cdot\nabla v),\ \ \kappa\in\mathcal{T}_h,\ \ v\in S^{\mathbf{p}}_{\mathcal{T}_h}\}$, endowed with the streamline diffusion DGFEM–norm $|||v|||_{\mathrm{s}}^2 := |||v|||_{\mathrm{DG}}^2 + \sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\|\mathbf{\Pi}_2(\mathbf{b}\cdot\nabla v)\|^2_{L^2(\kappa)}$. This approach, though, results in suboptimal, with respect to the polynomial degree p, *a priori* error bounds, cf. Remark 5.12 below.

5.2.2 A priori error analysis

In this section, we derive an *a priori* error bound for the IP DGFEM (5.14). First, we point out that Galerkin orthogonality does not hold due to the inconsistency of $\tilde{B}(\cdot,\cdot)$. Thereby, we derive the following abstract error bound in the spirit of Strang's second lemma.

Lemma 5.8. *Let u be the analytical solution of* (5.1), (5.5)*, and u_h be the IP DGFEM solution satisfying* (5.14)*. Assuming the inf-sup condition derived in Theorem 5.5 holds, we have that*

$$\begin{aligned} |||u-u_h|||_{\mathrm{s}} \ &\leq\ |||u-\tilde{\Pi}u|||_{\mathrm{s}} + \frac{1}{\Lambda_s}\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{|\tilde{B}(\tilde{\Pi}u-u,\omega_h)|}{|||\omega_h|||_{\mathrm{s}}} \\ &+\ \frac{1}{\Lambda_s}\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{|\tilde{B}(u,\omega_h)-\tilde{l}(\omega_h)|}{|||\omega_h|||_{\mathrm{s}}}, \end{aligned} \tag{5.42}$$

where $\tilde{\Pi}$ is the operator defined in Lemma 3.14.

Proof. The result follows in a standard manner, based on Strang's second lemma. We first use the triangle inequality,

$$|||u-u_h|||_{\mathrm{s}} \ \leq\ |||u-\tilde{\Pi}u|||_{\mathrm{s}} + |||\tilde{\Pi}u-u_h|||_{\mathrm{s}}.$$

Then we use fact that the second term in the above inequality if in $S^{\mathbf{p}}_{\mathcal{T}_h}$ together with relation, (5.26)

$$\begin{aligned} |||\tilde{\Pi}u-u_h|||_{\mathrm{s}} \ &\leq\ \frac{1}{\Lambda_s}\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{\tilde{B}(\tilde{\Pi}u-u_h,\omega_h)}{|||\omega_h|||_{\mathrm{s}}} \\ &\leq\ \frac{1}{\Lambda_s}\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{|\tilde{B}(\tilde{\Pi}u-u,\omega_h)|}{|||\omega_h|||_{\mathrm{s}}} + \frac{1}{\Lambda_s}\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{|\tilde{B}(u-u_h,\omega_h)|}{|||\omega_h|||_{\mathrm{s}}} \\ &=\ \frac{1}{\Lambda_s}\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{|\tilde{B}(\tilde{\Pi}u-u,\omega_h)|}{|||\omega_h|||_{\mathrm{s}}} + \frac{1}{\Lambda_s}\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{|\tilde{B}(u,\omega_h)-\tilde{l}(\omega_h)|}{|||\omega_h|||_{\mathrm{s}}}. \end{aligned}$$

Then the proof is complete. □

The abstract error bound of Lemma 5.8 is used to derive convergence results for the method at hand. These depend on the availability of the *hp*–version approximation estimates of Lemma 3.11. Assume that the mesh $\mathcal{T}_h$ admits a shape regular covering $\mathcal{T}_h^\sharp = \{\mathcal{K}\}$, cf. Definition 3.9, satisfying Assumption 3.3.1. Further assume that $u|_\kappa \in H^{l_\kappa}(\kappa)$, for some $l_\kappa > 1 + d/2$, for each $\kappa \in \mathcal{T}_h$, so that, by Theorem 3.12, $\mathfrak{E}u|_\mathcal{K} \in H^{l_\kappa}(\mathcal{K})$, where $\mathcal{K} \in \mathcal{T}_h^\sharp$ with $\kappa \subset \mathcal{K}$. To bound the first term on the right-hand side of the abstract bound above, we employ the approximation estimates of Lemma 3.14 together with Assumption 3.1.1 give

$$\begin{aligned}
|||u - \tilde{\Pi}u|||_{\mathrm{s}}^2 \;\leq\; & C \sum_{\kappa\in\mathcal{T}_h} \frac{h_\kappa^{2s_\kappa}}{p_\kappa^{2l_\kappa}} \Big(\|c_0\|_{L^\infty(\kappa)}^2 + \tau_\kappa \|\mathbf{b}\|_{L^\infty(\kappa)}^2 \frac{h_\kappa^{-2}}{p_\kappa^{-2}} + \bar{a}_\kappa \frac{h_\kappa^{-2}}{p_\kappa^{-2}} \\
& + \|\mathbf{b}\|_{L^\infty(\kappa)} \frac{h_\kappa^{-d}}{p_\kappa^{-1}} \sum_{F\subset\partial\kappa} C_m(p_\kappa,\kappa,F)|F| \\
& + \frac{h_\kappa^{-d}}{p_\kappa^{-1}} \sum_{F\subset\partial\kappa\cap(\mathcal{F}_h^\mathcal{I}\cup\mathcal{F}_h^\mathcal{D})} C_m(p_\kappa,\kappa,F)\sigma|F| \Big) \|\mathfrak{E}u\|_{H^{l_\kappa}(\mathcal{K})}^2. \qquad (5.43)
\end{aligned}$$

Next, we define $\eta = u - \tilde{\Pi}u$ and embark on bounding the second term on right-hand side of (5.42). Exploiting element-wise integration by parts, the advection-reaction bilinear form $B_{\mathrm{ar}}(\cdot,\cdot)$, cf. (5.9), can be written as:

$$\begin{aligned}
B_{\mathrm{ar}}(\eta,\omega_h) \;=\; & \sum_{\kappa\in\mathcal{T}_h} \Big(\int_\kappa (c - \nabla\cdot\mathbf{b})\omega_h\eta \,\mathrm{d}\mathbf{x} - \int_\kappa (\mathbf{b}\cdot\nabla\omega_h)\eta \,\mathrm{d}\mathbf{x} \\
& + \int_{\partial_-\kappa\setminus\mathcal{F}_h^\mathcal{B}} (\mathbf{b}\cdot\mathbf{n})\lfloor\omega_h\rfloor\eta^- \,\mathrm{d}s + \int_{\partial_+\kappa\cap\mathcal{F}_h^\mathcal{B}} (\mathbf{b}\cdot\mathbf{n})\omega_h^+\eta^+ \,\mathrm{d}s \Big).
\end{aligned}$$

Then, by using Cauchy-Schwarz inequality, we have the following bound:

$$\begin{aligned}
|B_{\mathrm{ar}}(\eta,\omega_h)| \;\leq\; & \sum_{\kappa\in\mathcal{T}_h} \Big(\|c_0\omega_h\|_{L^2(\kappa)} \|c_1\eta\|_{L^2(\kappa)} + \|\tau_\kappa^{\frac{1}{2}}\mathbf{b}\cdot\nabla\omega_h\|_{L^2(\kappa)} \|\tau_\kappa^{-\frac{1}{2}}\eta\|_{L^2(\kappa)} \\
& + \|\omega_h^+ - \omega_h^-\|_{\partial_-\kappa\setminus\mathcal{F}_h^\mathcal{B}} \|\eta^-\|_{\partial_-\kappa\setminus\mathcal{F}_h^\mathcal{B}} + \|\omega_h^+\|_{\partial_+\kappa\cap\mathcal{F}_h^\mathcal{B}} \|\eta^+\|_{\partial_+\kappa\cap\mathcal{F}_h^\mathcal{B}} \Big) \\
\leq\; & \Big(\sum_{\kappa\in\mathcal{T}_h} \|c_1\eta\|_{L^2(\kappa)}^2 + \sum_{\kappa\in\mathcal{T}_h} \tau_\kappa^{-1}\|\eta\|_{L^2(\kappa)}^2 + 2\sum_{\kappa\in\mathcal{T}_h} \|\eta^-\|_{\partial_-\kappa\setminus\mathcal{F}_h^\mathcal{B}}^2 \\
& + 2\sum_{\kappa\in\mathcal{T}_h} \|\eta^+\|_{\partial_+\kappa\cap\mathcal{F}_h^\mathcal{B}}^2 \Big)^{\frac{1}{2}} \times \Big(|||\omega_h|||_{\mathrm{ar}}^2 + \sum_{\kappa\in\mathcal{T}_h} \tau_\kappa \|\mathbf{b}\cdot\nabla\omega_h\|_{L^2(\kappa)}^2 \Big)^{\frac{1}{2}}.
\end{aligned}$$

We now derive a bound for $\tilde{B}(\eta,\omega_h)$ by employing the above result in conjunction with the continuity of $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$. Then, we get

$$\begin{aligned}
|\tilde{B}(\eta,\omega_h)| &= |B_{\mathrm{ar}}(\eta,\omega_h)+\tilde{B}_{\mathrm{d}}(\eta,\omega_h)| \\
&\leq \Big(\sum_{\kappa\in\mathcal{T}_h}\|c_1\eta\|^2_{L^2(\kappa)}+\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa^{-1}\|\eta\|^2_{L^2(\kappa)}+2\sum_{\kappa\in\mathcal{T}_h}\|\eta^-\|^2_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}} \\
&+ 2\sum_{\kappa\in\mathcal{T}_h}\|\eta^+\|^2_{\partial_+\kappa\cap\mathcal{F}_h^{\mathcal{B}}}\Big)^{\frac{1}{2}}\Big(|||\omega_h|||^2_{\mathrm{ar}}+\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa\|\mathbf{b}\cdot\nabla\omega_h\|^2_{L^2(\kappa)}\Big)^{\frac{1}{2}} \\
&+ C_{\mathrm{cont}}|||\eta|||_{\mathrm{d}}|||\omega_h|||_{\mathrm{d}} \\
&\leq \Big(\sum_{\kappa\in\mathcal{T}_h}\gamma_\kappa^2\|\eta\|^2_{L^2(\kappa)}+\sum_{\kappa\in\mathcal{T}_h}\tau_\kappa^{-1}\|\eta\|^2_{L^2(\kappa)} \\
&+ 2\sum_{\kappa\in\mathcal{T}_h}\|\eta^-\|^2_{\partial_-\kappa\setminus\mathcal{F}_h^{\mathcal{B}}}+2\|\eta^+\|^2_{\partial_+\kappa\cap\mathcal{F}_h^{\mathcal{B}}} \\
&+ (C_{\mathrm{cont}})^2\sum_{\kappa\in\mathcal{T}_h}\|\sqrt{a}\nabla\eta\|^2_{L^2(\kappa)}+(C_{\mathrm{cont}})^2\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F\sigma|[\![\eta]\!]|^2\,\mathrm{d}s\Big)^{\frac{1}{2}}|||\omega_h|||_{\mathrm{s}}.
\end{aligned}$$

Hence, by applying the approximation results in Lemma 3.14, we have the following bound:

$$\begin{aligned}
\sup_{\omega_h\in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}}\frac{|\tilde{B}(\tilde{\Pi}u-u,\omega_h)|}{|||\omega_h|||_{\mathrm{s}}} \leq C\Bigg(&\sum_{\kappa\in\mathcal{T}_h}\frac{h_\kappa^{2s_\kappa}}{p_\kappa^{2l_\kappa}}\Big(\gamma_\kappa^2+\tau_\kappa^{-1}+\bar{a}_\kappa\frac{h_\kappa^{-2}}{p_\kappa^{-2}} \\
&+\|\mathbf{b}\|_{L^\infty(\kappa)}\frac{h_\kappa^{-d}}{p_\kappa^{-1}}\sum_{F\subset\partial\kappa}C_m(p_\kappa,\kappa,F)|F| \\
&+\frac{h_\kappa^{-d}}{p_\kappa^{-1}}\sum_{F\subset\partial\kappa\cap(\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}})}C_m(p_\kappa,\kappa,F)\sigma|F|\Big)\|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}\Bigg)^{\frac{1}{2}}. \qquad (5.44)
\end{aligned}$$

Finally, we consider the residual due to the inconsistent formulation given by the third term in (5.42). From the definition of the original and inconsistent bilinear forms given by (5.10) and (5.13), respectively, we deduce that

$$\begin{aligned}
\tilde{B}(u,\omega_h)-\tilde{l}(\omega_h) &= \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F\{\!\{a(\nabla u-\mathbf{\Pi}_2(\nabla u))\}\!\}\cdot[\![\omega_h]\!]\,\mathrm{d}s \\
&\leq \Big(\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}}\int_F\sigma^{-1}|\{\!\{a(\nabla u-\mathbf{\Pi}_2(\nabla u))\}\!\}|^2\,\mathrm{d}s\Big)^{1/2}|||w_h|||_{\mathrm{d}},
\end{aligned}$$

where $\mathbf{\Pi}_2$ denotes the vector-valued L^2–projection onto the finite element space $[S^{\mathbf{p}}_{\mathcal{T}_h}]^d$. Employing the Cauchy Schwarz inequality gives

$$\sup_{\omega_h \in S^{\mathbf{p}}_{\mathcal{T}_h}\setminus\{0\}} \frac{|\tilde{B}(u,\omega_h)-\tilde{l}(\omega_h)|}{|||\omega_h|||_{\mathrm{s}}} \leq \Big(\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \sigma^{-1} |\{\!\{a(\nabla u - \mathbf{\Pi}_2(\nabla u))\}\!\}|^2 \,\mathrm{d}s\Big)^{\frac{1}{2}}.$$

Let $\tilde{\mathbf{\Pi}}$ denote the vector-valued hp–projection operator obtained by applying componentwise the operator $\tilde{\Pi}_{p_\kappa}$ given in (3.27). Adding and subtracting $\tilde{\mathbf{\Pi}}(\nabla u)$, we obtain

$$\begin{aligned}
&\sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}} \int_F \sigma^{-1} |\{\!\{a(\nabla u - \mathbf{\Pi}_2(\nabla u))\}\!\}|^2 \,\mathrm{d}s \\
&\quad\leq \sum_{F\in\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}}} \int_F 2\sigma^{-1} (|\{\!\{a(\nabla u - \tilde{\mathbf{\Pi}}(\nabla u))\}\!\}|^2 + |\{\!\{a(\mathbf{\Pi}_2(\tilde{\mathbf{\Pi}}(\nabla u) - \nabla u))\}\!\}|^2) \,\mathrm{d}s. \\
&\quad\equiv \mathrm{I} + \mathrm{II}.
\end{aligned}$$

Using, as above, the approximation estimate (3.29) yields:

$$\mathrm{I} \leq C \sum_{\kappa\in\mathcal{T}_h} \bar{a}_\kappa^2 \frac{h_\kappa^{2(s_\kappa-1)}}{p_\kappa^{2(l_\kappa-1)}} \frac{h_\kappa^{-d}}{p_\kappa^{-1}} \left(\sum_{F\subset\partial\kappa\cap(\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}})} C_m(p_\kappa,\kappa,F)\sigma^{-1}|F| \right) \|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}.$$

Similarly, the inverse inequality (3.13), the L^2-stability of the projector $\mathbf{\Pi}_2$, and the approximation estimate (3.28), yield:

$$\mathrm{II} \leq C \sum_{\kappa\in\mathcal{T}_h} \bar{a}_\kappa^2 \frac{h_\kappa^{2(s_\kappa-1)}}{p_\kappa^{2(l_\kappa-1)}} \frac{|\kappa|^{-1}}{p_\kappa^{-2}} \left(\sum_{F\subset\partial\kappa\cap(\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}})} C_{\mathrm{INV}}(p_\kappa,\kappa,F)\sigma^{-1}|F| \right) \|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}.$$

Combining the above developments we arrive to the following bound of the residual term:

$$\begin{aligned}
&\sup_{w_h\in S^{\mathbf{p}}_{\mathcal{T}_h}} \frac{|\tilde{B}(u,w_h)-\tilde{\ell}(u,w_h)|}{|||w_h|||_{\mathrm{s}}} \leq \Big(\mathrm{I}+\mathrm{II}\Big)^{1/2} \\
&\leq C \Bigg(\sum_{\kappa\in\mathcal{T}_h} \bar{a}_\kappa^2 \frac{h_\kappa^{2(s_\kappa-1)}}{p_\kappa^{2(l_\kappa-1)}} \\
&\qquad \times \left(\sum_{F\subset\partial\kappa\cap(\mathcal{F}_h^{\mathcal{I}}\cup\mathcal{F}_h^{\mathcal{D}})} \left(C_m(p_\kappa,\kappa,F)\frac{h_\kappa^{-d}}{p_\kappa^{-1}} + C_{\mathrm{INV}}(p_\kappa,\kappa,F)\frac{|\kappa|^{-1}}{p_\kappa^{-2}} \right) \sigma^{-1}|F| \right) \\
&\qquad \times \|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})} \Bigg)^{1/2}. \qquad (5.45)
\end{aligned}$$

Finally, combining the approximation bound (5.43), (5.44), and residual bound (5.45) together with with Lemma 5.8 yield the following DGFEM convergence result.

Theorem 5.9. *Let $\mathcal{T}_h = \{\kappa\}$ be a subdivision of $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, consisting of general polygonal/polyhedral elements satisfying Assumption 3.1.1, Assumption 3.2.1 and Assumptions 3.2.2 with $\mathcal{T}_h^\sharp = \{\mathcal{K}\}$ an associated covering of $\mathcal{T}_h$ consisting of shape-regular d–simplexes, cf., Definition 3.9. Let $u_h \in S^{\mathbf{p}}_{\mathcal{T}_h}$, with $p_\kappa \geq 1$ for all $\kappa \in \mathcal{T}_h$, be the corresponding DGFEM solution defined by (4.5) with the discontinuity-penalization functions given by (4.7). If the exact solution $u \in H^1(\Omega)$ to (4.1)–(4.3) satisfies $u|_\kappa \in H^{l_\kappa}(\kappa)$, $l_\kappa > 1 + d/2$, for each $\kappa \in \mathcal{T}_h$, such that $\mathfrak{E}u|_\mathcal{K} \in H^{l_\kappa}(\mathcal{K})$, where $\mathcal{K} \in \mathcal{T}_h^\sharp$ with $\kappa \subset \mathcal{K}$, then*

$$|||u - u_h|||_{\mathrm{s}}^2 \leq C \sum_{\kappa \in \mathcal{T}_h} \frac{h_\kappa^{2s_\kappa}}{p_\kappa^{2l_\kappa}} \Big(\mathcal{G}_\kappa(F, C_m, p_\kappa, \tau_\kappa) + \mathcal{D}_\kappa(F, C_{\mathrm{INV}}, C_m, p_\kappa) \Big) \|\mathfrak{E}u\|^2_{H^{l_\kappa}(\mathcal{K})}, \tag{5.46}$$

where

$$\begin{aligned} \mathcal{G}_\kappa(F, C_m, p_\kappa, \tau_\kappa) = \|c_0\|^2_{L^\infty(\kappa)} + \gamma_\kappa^2 + \tau_\kappa^{-1} + \tau_\kappa \beta_\kappa^2 p_\kappa^2 h_\kappa^{-2} + \bar{a}_\kappa p_\kappa^2 h_\kappa^{-2} \\ + \beta_\kappa p_\kappa h_\kappa^{-d} \sum_{F \subset \partial\kappa} C_m(p_\kappa, \kappa, F)|F| \\ + p_\kappa h_\kappa^{-d} \sum_{F \subset \partial\kappa \cap (\mathcal{F}_h^\mathcal{I} \cup \mathcal{F}_h^\mathcal{D})} C_m(p_\kappa, \kappa, F)\sigma|F|, \end{aligned} \tag{5.47}$$

and

$$\begin{aligned} \mathcal{D}_\kappa(F, C_{\mathrm{INV}}, C_m, p_\kappa) = \bar{a}_\kappa^2 \Big(p_\kappa^3 h_\kappa^{-d-2} \sum_{F \subset \partial\kappa \cap (\mathcal{F}_h^\mathcal{I} \cup \mathcal{F}_h^\mathcal{D})} C_m(p_\kappa, \kappa, F)\sigma^{-1}|F| \\ + p_\kappa^4 |\kappa|^{-1} h_\kappa^{-2} \sum_{F \subset \partial\kappa \cap (\mathcal{F}_h^\mathcal{I} \cup \mathcal{F}_h^\mathcal{D})} C_{\mathrm{INV}}(p_\kappa, \kappa, F)\sigma^{-1}|F| \Big), \end{aligned} \tag{5.48}$$

with $s_\kappa = \min\{p_\kappa + 1, l_\kappa\}$ and $p_\kappa \geq 1$. Here, $\gamma_\kappa = \|c_1\|_{L^\infty(\kappa)}$, with $c_1(x) := (c(x) - \nabla \cdot \mathbf{b}(x))/(c_0(x))$, c_0 as in (5.7), and $\beta_\kappa = \|\mathbf{b}\|_{L^\infty(\kappa)}$. The positive constant C is independent of the discretization parameters.

Remark 5.10. We note that the above *hp*–version *a priori* bound for the IP DGFEM (5.14) holds without the need to impose any assumption concerning the relative size of the faces F, $F \subset \partial\kappa$, of a given polytopic element $\kappa \in \mathcal{T}_h$. If $\mathbf{b} \equiv \mathbf{0}$ and $c \equiv 0$ on Ω, then the streamline diffusion DGFEM–norm degenerates to the diffusion DGFEM–norm $|||\cdot|||_{\mathrm{d}}$ defined in (5.16) and the problem becomes the pure diffusion problem, which is independent of τ_κ with constants β_κ and γ_κ equal to zero. Furthermore, the inf-sup condition is equivalent to the coercivity of

the bilinear form $\tilde{B}_{\rm d}(\cdot,\cdot)$. This can be used to derive an error bound, analogous to the error bound (5.46), which generalises the result presented in [61] for the Poisson equation with constant diffusion. Moreover, in this setting, for uniform orders $p_\kappa = p \geq 1$, $h = \max_{\kappa\in\mathcal{T}_h} h_\kappa$, $s_\kappa = s$, $s = \min\{p+1, l\}$, $l > 1 + d/2$, under the assumption that the diameter of the faces of each element $\kappa \in \mathcal{T}_h$ is of comparable size to the diameter of the corresponding element, i.e., $\mathrm{diam}(F) \sim h_\kappa$, $h_\kappa^\perp \sim h_\kappa$, $F \subset \partial\kappa$, $\kappa \in \mathcal{T}_h$, so that $|F| \sim h_\kappa^{(d-1)}$, the *a priori* error bound of Theorem 5.9 reduces to

$$|||u - u_h|||_{\rm d} \leq C\frac{h^{s-1}}{p^{l-\frac{3}{2}}}\|u\|_{H^l(\Omega)}.$$

This coincides with the analogous result derived in [125] for standard meshes consisting of simplices or tensor-product elements. Here, we have employed Lemma 3.14 and Theorem 3.12, together with Assumption 3.3.1, assuming that for such element domains $C_{\rm INV}(p_\kappa, F) = \mathcal{O}(1)$ and $C_m(p_\kappa, F) = \mathcal{O}(1)$ uniformly for each face $F \subset \partial\kappa$ for all $\kappa \in \mathcal{T}_h$. This error bound is h optimal and p suboptimal by $p^{1/2}$.

Remark 5.11. Consider the purely hyperbolic case when the diffusion tensor $a \equiv 0$. In this case, the constants $\bar{a}_\kappa$ and $\tilde{\sigma}_\kappa$ are identically zero and the inconsistent term $\mathcal{D}_\kappa(F, C_{\rm INV}, C_m, p_\kappa)$ vanishes due to the consistency of the bilinear form $B_{\rm ar}(\cdot,\cdot)$. Then, the streamline diffusion DGFEM–norm is actually stronger than the advection-reaction DGFEM–norm $|||\cdot|||_{\rm ar}$ defined in (5.15) and $\tau_\kappa = \mathcal{O}(\frac{h_\kappa}{p_\kappa^2})$ by (5.19). In this case, for uniform orders, cf. Remark 5.10 above, the *a priori* error bound of Theorem 5.9 yields

$$|||u - u_h|||_{\rm ar} \leq |||u - u_h|||_{\rm s} \leq C\frac{h^{s-\frac{1}{2}}}{p^{l-1}}\|u\|_{H^l(\Omega)}.$$

Hence, the above *hp*–bound is optimal in h and suboptimal in p by $p^{1/2}$. In this case, our bound generalizes the error estimate derived in [125] to general polytopic meshes under the same assumption $\mathbf{b}\cdot\nabla_h\xi \in S^{\mathbf{p}}_{\mathcal{T}_h}$, $\xi \in S^{\mathbf{p}}_{\mathcal{T}_h}$, with a slight loss of p-convergence.

Remark 5.12. As noted in Remark 5.7, the case of general convection fields $\mathbf{b}$ can be treated, based on employing an inf-sup condition with different test and trial spaces. In this setting, the present analysis can easily be adapted to utilize such an inf-sup condition, together with the exploitation of the L^2–projector $\mathbf{\Pi}_2$ onto the polytopic element $\kappa \in \mathcal{T}_h$. However, this yields an error bound in the $|||\cdot|||_{\rm ar}$–norm that is optimal in h but suboptimal in p by $p^{3/2}$ for the purely hyperbolic problem.

We also point out that if we modify the DGFEM by including the streamline-diffusion stabilization term as in [124], then an hp–optimal bound can be derived without the assumption that $\mathbf{b} \cdot \nabla_h \xi \in S^{\mathbf{p}}_{\mathcal{T}_h}$, $\xi \in S^{\mathbf{p}}_{\mathcal{T}_h}$. This is not derived here in detail for brevity.

5.3 Numerical examples

We present a series of computational examples to numerically investigate the asymptotic convergence behaviour of the proposed IP DGFEM on general meshes consisting of polytopic elements. As in [61], the integrals arising in the bilinear and linear forms $B(\cdot,\cdot)$ and $\ell(\cdot)$, respectively, are computed based on employing a quadrature scheme defined on a sub-tessalation of each polytopic element in the underlying finite element mesh. Throughout this section, the IP DGFEM solution u_h defined by (5.8) is computed with the constant C_σ appearing in the discontinuity-penalization parameter σ equal to 10. Given the computations already presented in Chapter 4, here we concentrate on studying the performance of the proposed IP DGFEM in the hyperbolic, mixed parabolic–hyperbolic setting and boundary layer problem. To this end, we first study a pure hyperbolic problem (diffusion matrix $a \equiv 0$) in Section 5.3.1. Secondly, we consider an advection-diffusion-reaction problem with degenerate, anisotropic diffusion matrix a in Section 5.3.2. Within these examples, we employ polygonal meshes generated using the general-purpose mesh generator `PolyMesher`, cf. [179]. Additionally, a classical boundary layer problem is presented in Section 5.3.3 to study the exponential convergence of $\mathcal{P}_p$ basis on anisotropic refined meshes. Finally, in Section 5.3.4, we study the convergence behaviour of the underlying DGFEM for a purely hyperbolic problem in three dimensions on general polytopes generated based on employing agglomeration.

Throughout this section, we compare the performance of employing $\mathcal{P}_p$–polynomial bases on polytopic meshes, with $\mathcal{P}_p$– and $\mathcal{Q}_p$–polynomial bases defined on standard tensor-product meshes.

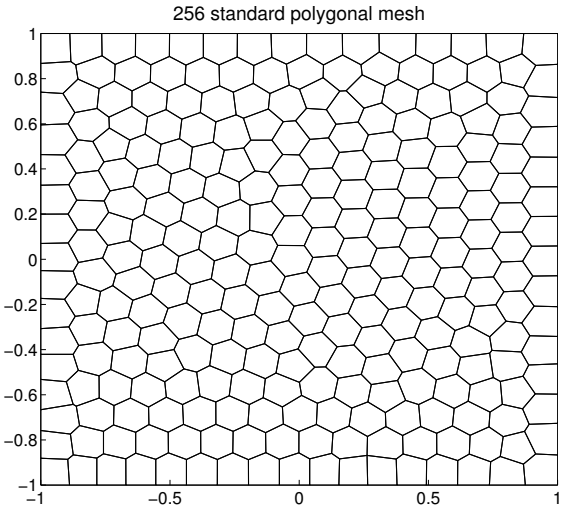


FIGURE 5.2: Example 1: Uniform polygonal mesh, consisting of 256 elements.

5.3.1 Example 1

In this first example, we let Ω be the square domain $(-1,1)^2$, and choose

$$a \equiv 0, \quad \mathbf{b} = (2-y^2, 2-x), \quad c = 1 + (1+x)(1+y)^2; \tag{5.49}$$

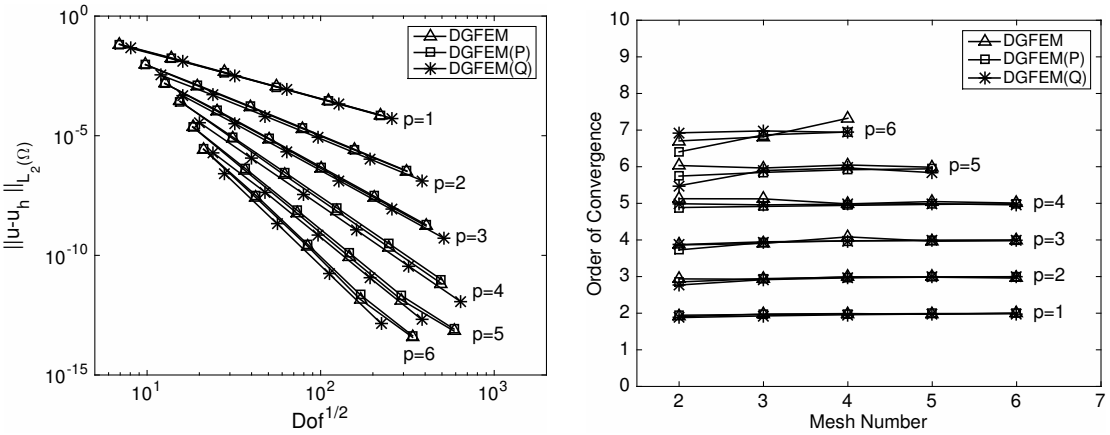
the forcing function f is selected so that the analytical solution to (5.1), (5.5) is given by

$$u(x,y) = 1 + \sin(\pi(1+x)(1+y)^2/8), \tag{5.50}$$

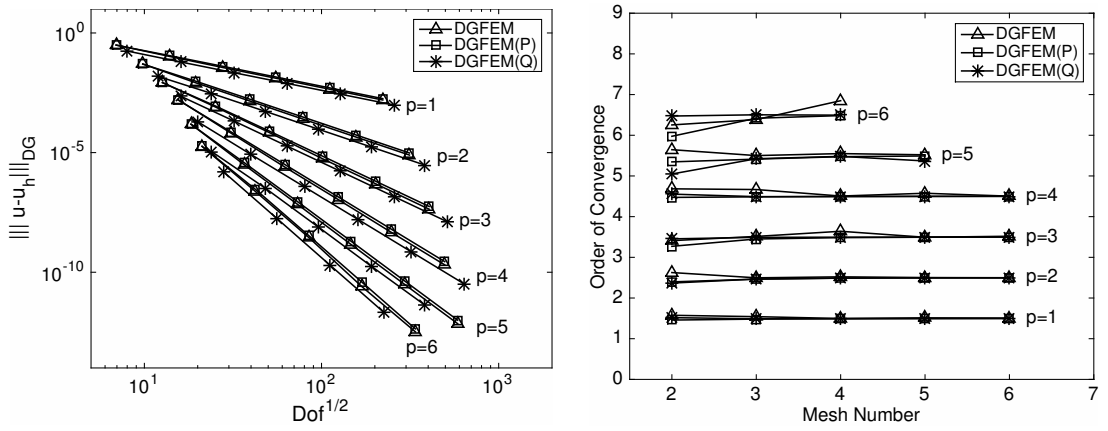
cf. [125].

We investigate the asymptotic behaviour of the *hp*–version DGFEM on a sequence of successively finer polygonal and uniform quadrilateral meshes for different values of the polynomial degree p. Three settings are compared: uniform quadrilateral meshes and local polynomial bases consisting of either $\mathcal{P}_p$ or $\mathcal{Q}_p$ polynomials, and polygonal meshes and local polynomial bases consisting of $\mathcal{P}_p$ polynomials; the three cases are referred to as, respectively, DGFEM(P), DGFEM(Q), and DGFEM. The polygonal meshes used for DGFEM are generated using the Polymesher mesh generator, cf. [179]; a typical mesh, consisting of 256 elements, is depicted in Figure 5.2.

We first examine the convergence behaviour of the three schemes with respect to *h*–refinement, with fixed polynomial p, for $p = 1, \ldots, 6$. In Figure 5.3 we plot the error, measured in terms of both the $L^2(\Omega)$- and DGFEM–norm, against the



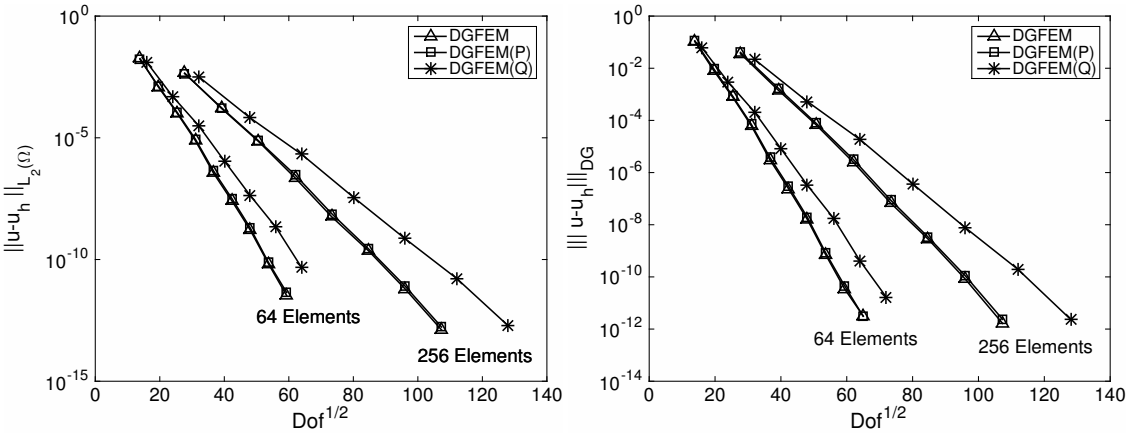
(a)



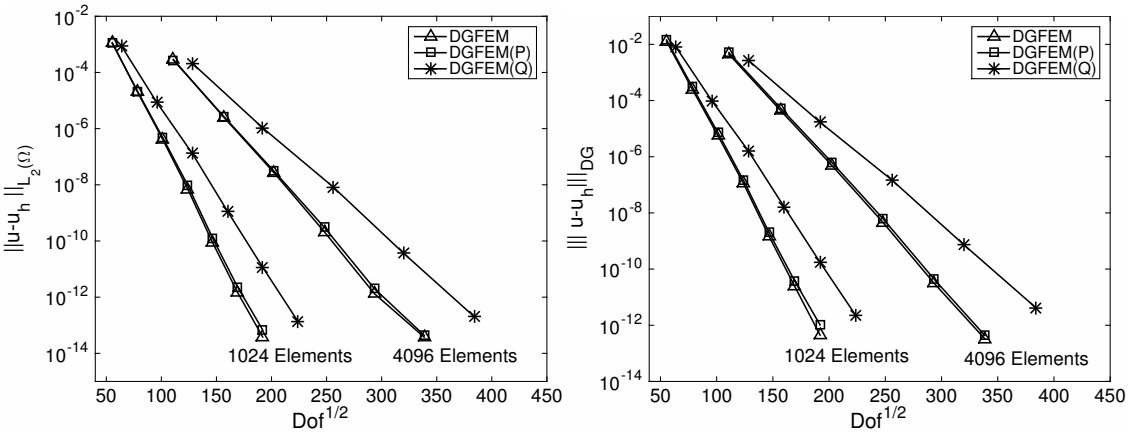
(b)

FIGURE 5.3: Example 1: Convergence of the DGFEM under h–refinement for $p = 1, 2, \ldots, 6$. (a) $\|u - u_h\|_{L^2(\Omega)}$; (b) $|||u - u_h|||_{\mathrm{DG}}$.

square root of the number of degrees of freedom in the underlying finite element space $S^{\mathbf{p}}_{\mathcal{T}_h}$. Here, we clearly observe that $\|u - u_h\|_{L^2(\Omega)}$ and $|||u - u_h|||_{\mathrm{DG}}$ converge to zero at the optimal rates $\mathcal{O}(h^{p+1})$ and $\mathcal{O}(h^{p+\frac{1}{2}})$, respectively, as the mesh size h tends to zero for each fixed p. The latter set of results confirm the optimality of Theorem 5.9, cf. Remark 5.11, in the case when polygonal elements are employed. We point out that the (optimal) convergence rate observed when the error is measured in terms of the $L^2(\Omega)$–norm is not guaranteed on general meshes, cf. [154] (optimal convergence of $\|u - u_h\|_{L^2(\Omega)}$ has been established in [74, 75], but only for special classes of triangular elements.) From Figure 5.3, we also observe that polygonal and square meshes deliver almost identical results given the same number of degrees of freedom, when $\mathcal{P}_p$ elements are used (cf. the errors attained by DGFEM and DGFEM(P)). By comparison, the use of tensor-product polynomials, i.e. the DGFEM(Q) scheme, leads to a marginal decrease in both error



(a)



(b)

FIGURE 5.4: Example 1: Convergence of the DGFEM under p–refinement. Left: $\|u - u_h\|_{L^2(\Omega)}$; Right: $|||u - u_h|||_{\mathrm{DG}}$; (a) Meshes consisting of 64 and 256 elements; (b) Meshes consisting of 1024 and 4096 elements.

quantities.

Finally, in Figure 5.4 we investigate the convergence behaviour of the three schemes under p–refinement, for fixed h. Here, uniform polygonal and square meshes consisting of 64, 256, 1024, and 4096 elements are employed. For each mesh, we plot $\|u - u_h\|_{L^2(\Omega)}$ and $|||u - u_h|||_{\mathrm{DG}}$ against the square root of the number of degrees of freedom in $S^{\mathbf{p}}_{\mathcal{T}_h}$. In each case we clearly observe exponential convergence. We observe that, under p–refinement, the efficiency of employing local $\mathcal{P}_p$ polynomials is apparent. Indeed, both the DGFEM and DGFEM(P) schemes lead to a significant reduction in the error, when measured in terms of both the $L^2(\Omega)$- and DGFEM–norms, for a fixed number of degrees of freedom, when compared with the DGFEM(Q) scheme, cf. [61]. As before, the DGFEM and DGFEM(P) schemes give almost identical results in terms of the size of the discretization error,

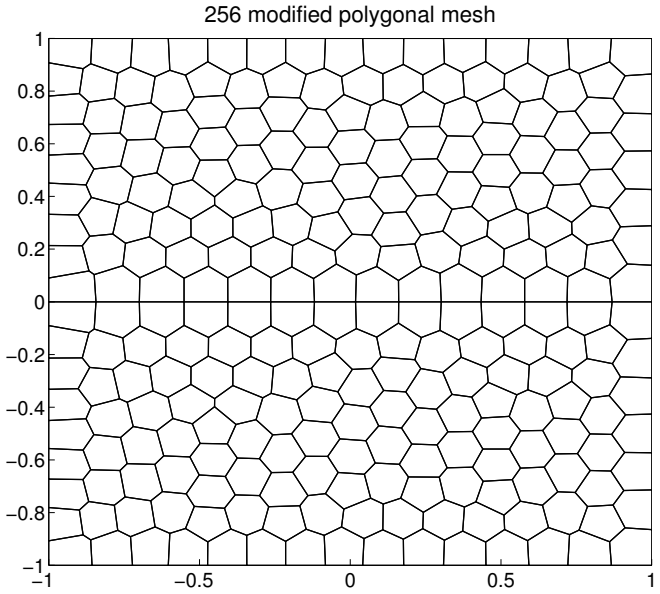


FIGURE 5.5: Example 2: Modified uniform polygonal mesh, consisting of 256 elements

for a fixed number of degrees of freedom, though in some instances, the former scheme is slightly more accurate.

5.3.2 Example 2

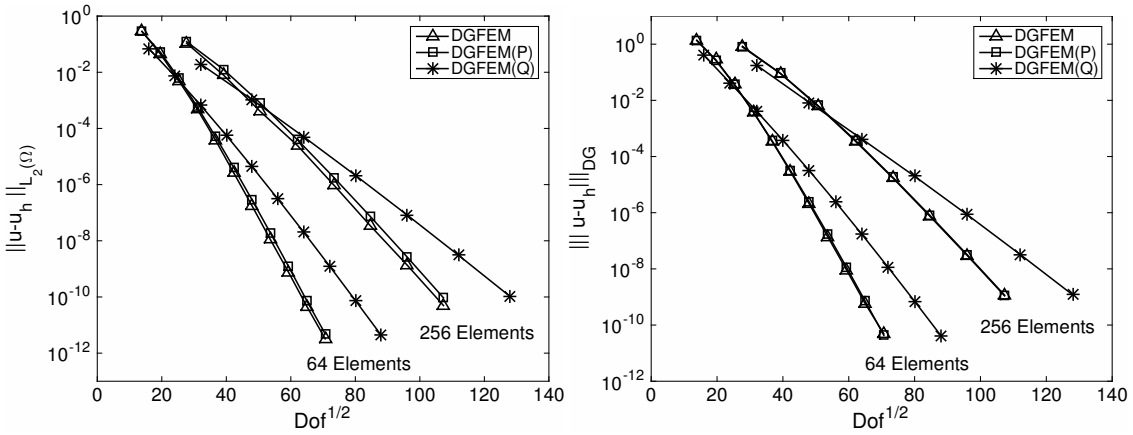
In this second example, we consider a partial differential equation with nonnegative characteristic form of mixed type. To this end, we let $\Omega = (-1, 1)^2$, and consider the PDE problem:

$$\begin{cases} -x^2 u_{yy} + u_x + u = 0, & \text{for } -1 \leq x \leq 1, y > 0, \\ u_x + u = 0, & \text{for } -1 \leq x \leq 1, y \leq 0, \end{cases} \tag{5.51}$$

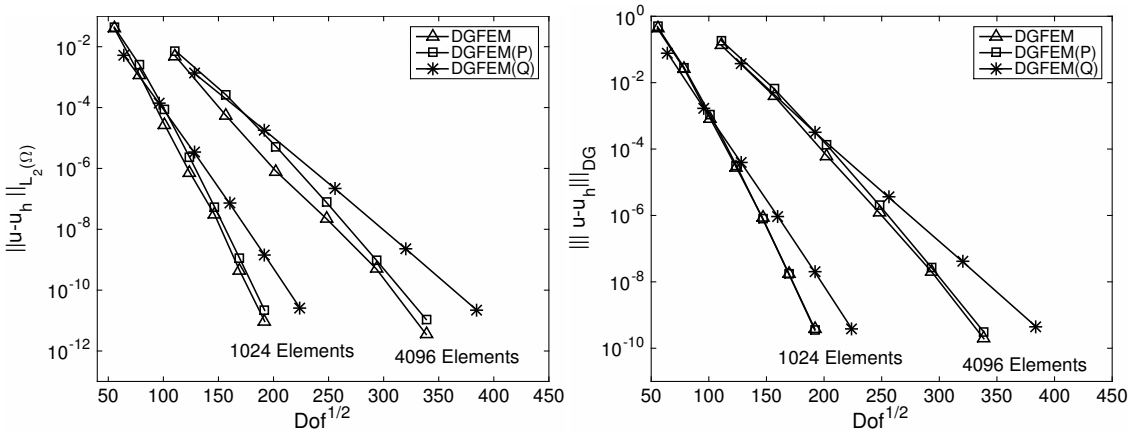
with analytical solution:

$$u(x,y) = \begin{cases} \sin(\frac{1}{2}\pi(1+y)) \exp(-(x + \frac{\pi^2 x^3}{12})), & \text{for } -1 \leq x \leq 1, y > 0, \\ \sin(\frac{1}{2}\pi(1+y)) \exp(-x), & \text{for } -1 \leq x \leq 1, y \leq 0, \end{cases} \tag{5.52}$$

cf. [103]. This problem is hyperbolic in the region $y \leq 0$ and parabolic for $y > 0$. In order to ensure continuity of the normal flux across $y = 0$, where the partial



(a)



(b)

FIGURE 5.6: Example 2: Convergence of the DGFEM under p–refinement. Left: $\|u - u_h\|_{L^2(\Omega)}$; Right: $|||u - u_h|||_{\mathrm{DG}}$; (a) Meshes consisting of 64 and 256 elements; (b) Meshes consisting of 1024 and 4096 elements.

differential equation changes type, the analytical solution has a discontinuity across the line $y = 0$, cf. [125].

To highlight one of the advantages of employing finite element methods with discontinuous piecewise polynomial spaces, we consider a special class of quadrilateral and polygonal meshes for which the discontinuity in the analytical solution lies on element interfaces only; for the case when polygonal elements are employed, a typical mesh is shown in Figure 5.5. In this setting, following [125], we modify the discontinuity-penalization parameter σ, so that σ vanishes on edges which form part of the interface $y = 0$; this ensures that the (physical) discontinuity present in the analytical solution is not penalized within the underlying scheme.

In this case, the *hp*–DGFEM behaves as if the analytical solution were smooth, in the sense that exponential rates of convergence are observed for both the $L^2(\Omega)$-

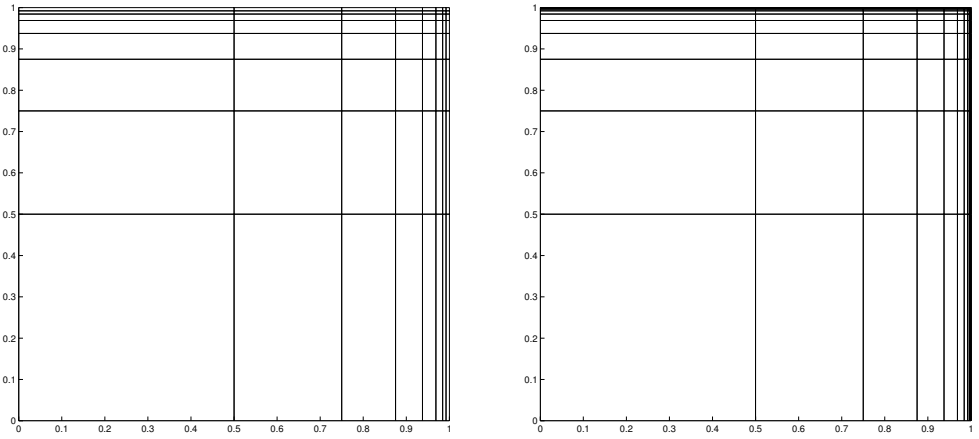


FIGURE 5.7: Example 3: Anisotropically refined meshes. 64 elements (Left); 196 elements (Right).

and DGFEM–norm of the error under p–refinement, cf. Figure 5.6. As in the previous example, we again observe that the slope of the convergence curves for both the DGFEM and DGFEM(P) schemes are steeper than the corresponding convergence curve obtained when local polynomial bases consisting of tensor-product polynomials ($\mathcal{Q}_p$ basis) are employed, cf. the numerical results presented for the DGFEM(Q) scheme. The DGFEM and DGFEM(P) schemes give once more very similar results in terms of the size of the computed error for a given number of degrees of freedom. Nevertheless, we notice more clearly that the use of polygonal elements leads to a slight improvement when considering $\|u - u_h\|_{L^2(\Omega)}$. As noted in [61], cf. also [125], the improvement in the $L^2(\Omega)$–norm when polygons are employed, in comparison with square elements, is attributed to the increase in interelement communication.

5.3.3 Example 3

In the this example, we consider a singularly perturbed advection-diffusion problem equation

$$-\epsilon\Delta u + u_x + u_y = f,$$

with $\Omega := (0,1)^2$, where $0 < \epsilon \ll 1$ and f is chosen so that

$$u(x,y) = x + y(1-x) + \frac{[e^{-1/\epsilon} - e^{-(1-x)(1-y)/\epsilon}]}{[1 - e^{-1/\epsilon}]}. \tag{5.53}$$

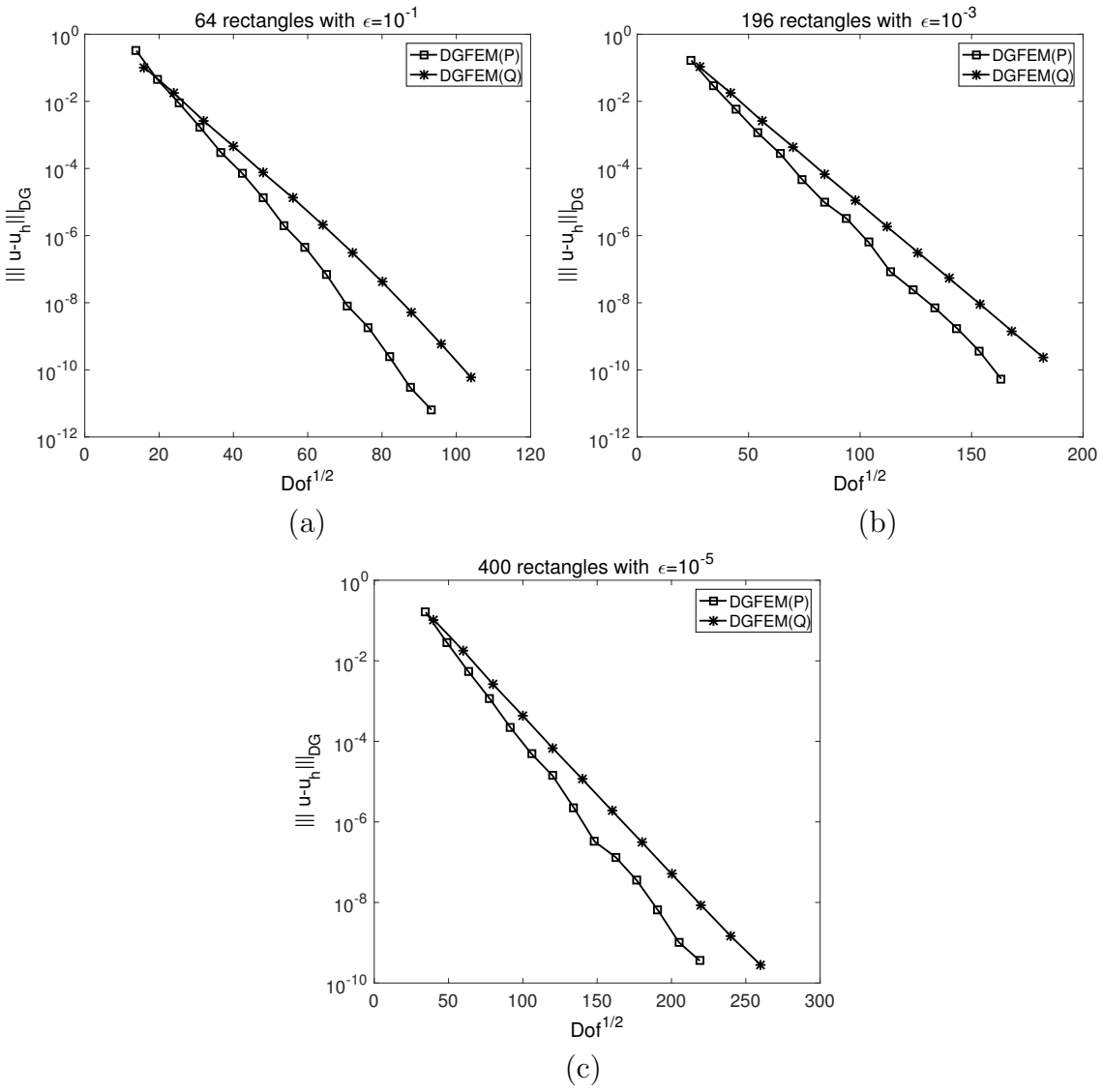


FIGURE 5.8: Example 3: Convergence of the DGFEM under p-refinement (a) $\epsilon = 10^{-1}$ with 64 elements; (b) $\epsilon = 10^{-3}$ with 196 elements; (c) $\epsilon = 10^{-5}$ with 400 elements.

This example is taken from [125]. For $0 < \epsilon \ll 1$, (5.53) has boundary layers along $x = 1$ and $y = 1$. Here, we use anisotropically refined meshes for resolving the boundary layer.

In this numerical experiment we test the robustness of the DGFEM(P) and DGFEM-(Q) on highly stretched anisotropic quadrilateral meshes as the physical diffusion ϵ decreases. The meshes are constructed by geometrical refinement into the boundary layers along $x = 1$ and $y = 1$ and are parameterized by n_ϵ which denotes the number of points in the x and y directions. In Figure 5.7 we show a typical mesh for $n_\epsilon = 9$ and $n_\epsilon = 15$. Figure 5.8 shows a plot of the DG–norm of the error

under p-refinement for $\epsilon = 10^{-1}, 10^{-3}, 10^{-5}$ on geometrically refined quadrilateral meshes with $n_\epsilon = 9, 15, 21$, respectively. It is easy to see that in each case we observe robust exponential convergence as the polynomial degree is increased for both DGFEM(P) and DGFEM(Q) schemes, and DGFEM(P) still have a steeper convergence in all cases.

5.3.4 Example 4

In this final example, we investigate the performance of the proposed DGFEM on sequences of polyhedral meshes in three dimensions for a purely hyperbolic problem. To this end, we consider a three–dimensional variant of the two–dimensional problem considered in Section 5.3.1. In particular, we let Ω be the unit cube $(0,1)^3$ and set

$$a \equiv 0, \quad \mathbf{b} = (-y, z, x), \quad c = xy^2z;$$

f is then selected so that the analytical solution to (5.1), (5.5) is

$$u(x,y) = 1 + \sin(\pi xy^2z/8). \tag{5.54}$$

In this section the DGFEM solution is computed on general polyhedral meshes, stemming from the agglomeration of a given (fixed) fine mesh $\mathcal{T}_f$. More precisely, we employ a fine mesh consisting of approximately 1M tetrahedral elements (1019674 elements, to be precise). cf. Figure 5.9 (a). The coarse agglomerated mesh $\mathcal{T}_h$ is then constructed based on exploiting the graph partitioning package METIS [136]. In order for METIS to partition the mesh $\mathcal{T}_f$, the logical structure of the mesh is first stored in the form of a graph, where each node represents an element domain of $\mathcal{T}_f$, and each link between two nodes represents a face shared by the two elements represented by the graph nodes. The partition of $\mathcal{T}_f$ constructed by METIS is produced with the objective of minimizing the number of neighbours among each of the resulting partitions. In Figure 5.9, we show (the surface mesh of) the polyhedral meshes generated by METIS, which consist of 64, 512, 4096, and 32768 elements.

In Figure 5.10 we investigate the h–version convergence behaviour of the DGFEM on both the polyhedral meshes depicted in Figure 5.9 and uniform hexahedral meshes, using local $\mathcal{P}_p$ polynomial bases; denoted by DGFEM and DGFEM(P), respectively. As already noted in Section 5.3.1, we again observe that $\|u-u_h\|_{L^2(\Omega)}$

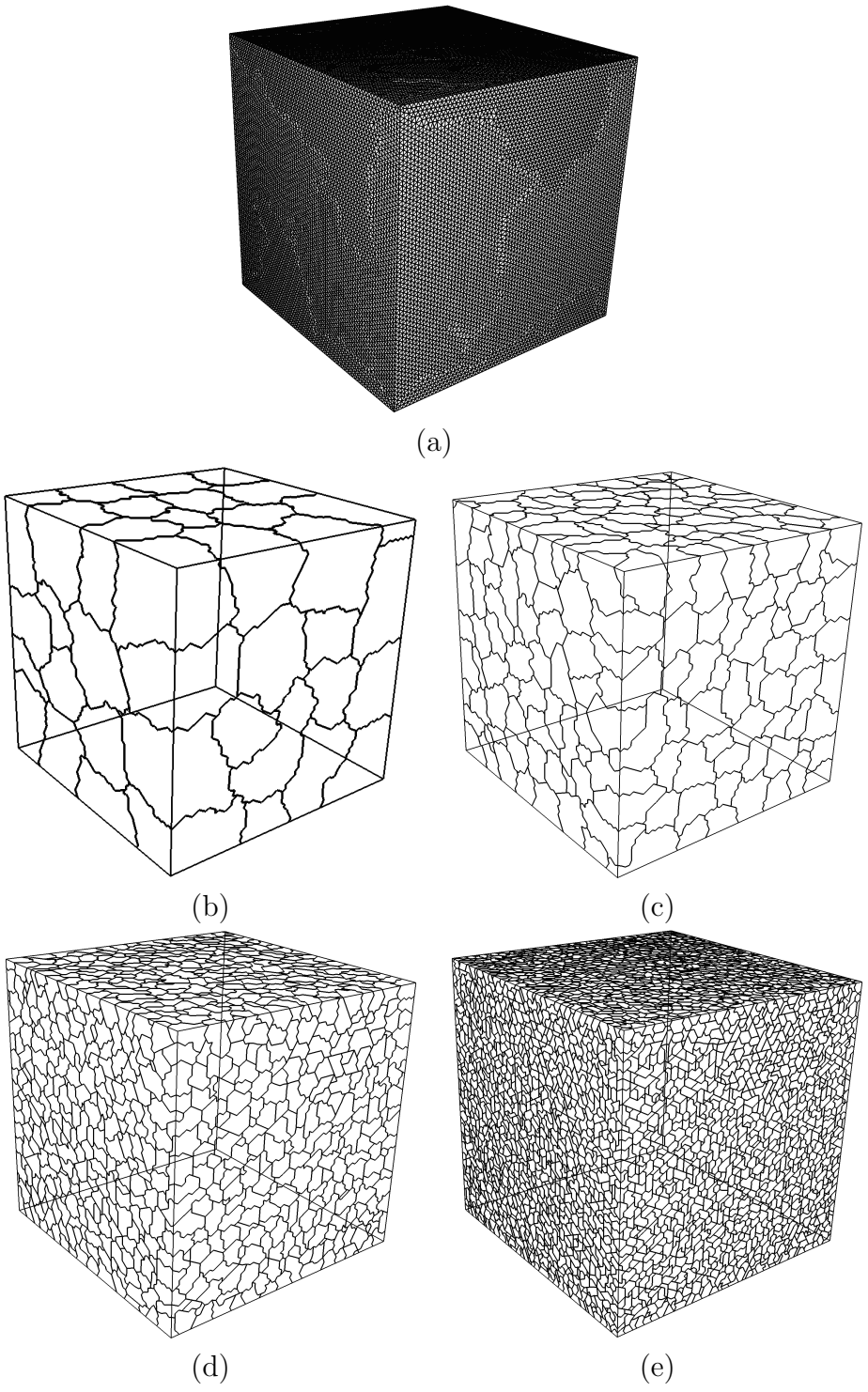
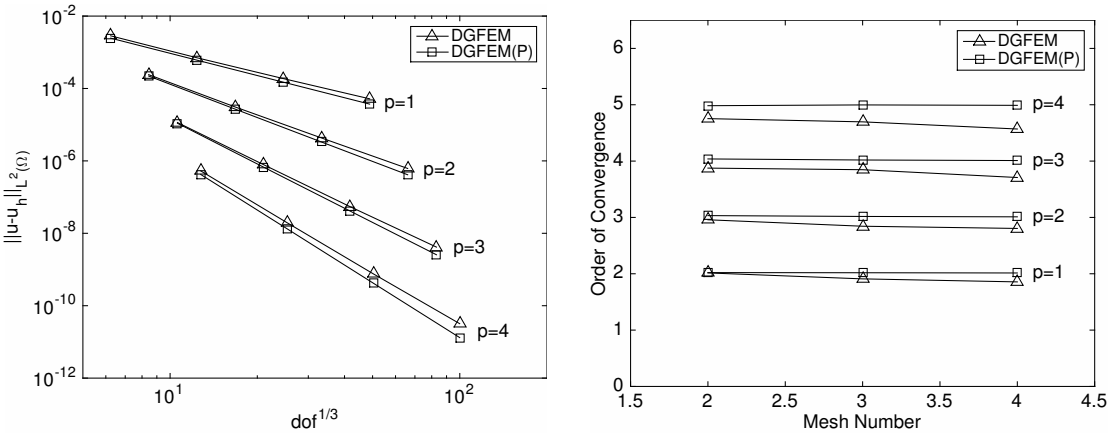
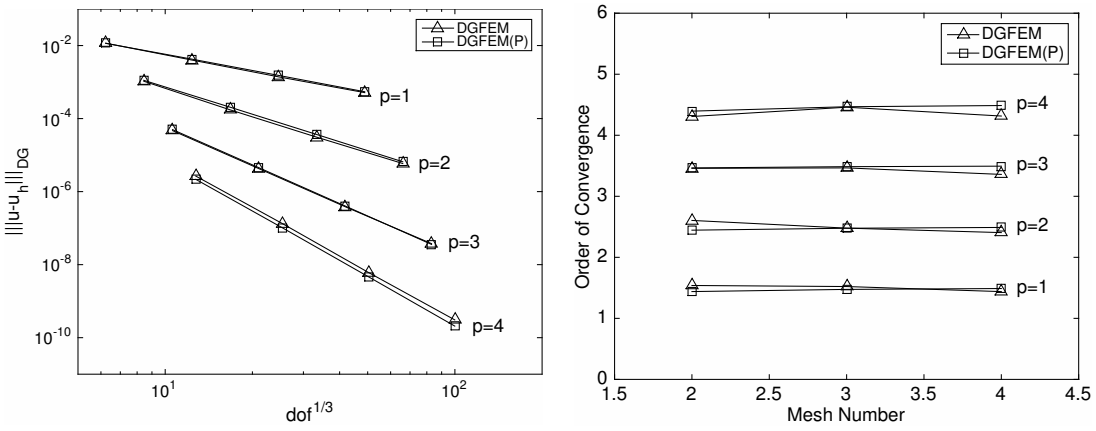


FIGURE 5.9: Example 4: (a). Initial fine mesh, consisting of approximately 1M tetrahedral elements. Agglomerated meshes. (b) 64 elements; (c) 512 elements; (d) 4096 elements; (e) 32768 elements.

and $|||u - u_h|||_{\mathrm{DG}}$ converge to zero at the optimal rates $\mathcal{O}(h^{p+1})$ and $\mathcal{O}(h^{p+\frac{1}{2}})$, respectively, as the mesh size h tends to zero for each fixed p when the DGFEM(P) scheme is employed on uniform tensor-product elements. Moreover, we observe

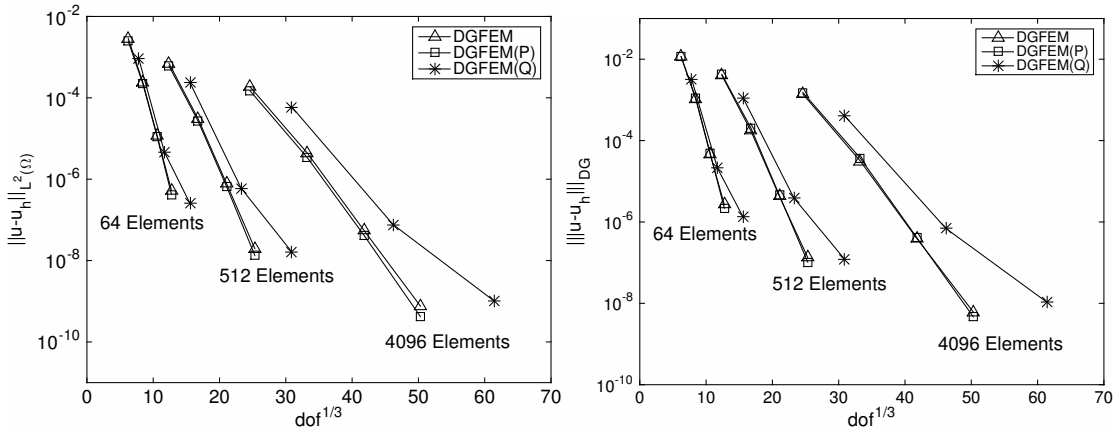


(a)



(b)

FIGURE 5.10: Example 4: Convergence of the DGFEM under h–refinement for $p = 1, 2, 3, 4$. (a) $\|u - u_h\|_{L^2(\Omega)}$; (b) $|||u - u_h|||_{\mathrm{DG}}$.



(a) (b)

FIGURE 5.11: Example 4: Convergence of the DGFEM under p–refinement. (a) $\|u - u_h\|_{L^2(\Omega)}$; (b) $|||u - u_h|||_{\mathrm{DG}}$.

that the DGFEM–norm of the error, when general polyhedral elements are employed, is very similar to the corresponding quantity computed for the DGFEM(P) scheme. However, we observe a slight degradation of $\|u - u_h\|_{L^2(\Omega)}$, when the DGFEM scheme is employed, when compared to the case when uniform hexahedral elements are exploited. For brevity, the corresponding results for the DGFEM(Q) are omitted, though, we note again that, for fixed p, this approach is more efficient as the mesh is uniformly refined.

Finally, we study the performance of the DGFEM, DGFEM(P), and DGFEM(Q) schemes under p–refinement, for a given fixed mesh. To this end, in Figure 5.11 we plot both $\|u - u_h\|_{L^2(\Omega)}$ and $|||u - u_h|||_{\mathrm{DG}}$ against the third root of the number of degrees of freedom in $S^{\mathbf{p}}_{\mathcal{T}_h}$. As in the previous numerical examples, we again observe the superiority of employing local polynomial bases of total degree p in comparison with full tensor-product bases of degree p in each coordinate direction.

Chapter 6

DGFEMs for Time-Dependent PDEs on Prismatic Meshes

In Chapter 5, we presented a detailed analysis on IP-DGFEM for PDEs with non-negative characteristic form. In this chapter, we will study more in detail space-time DGFEMs for time-dependent parabolic PDEs, which is an important class of PDEs with non-negative characteristic form. The analysis presented here is based on that for IP-DGFEM scheme for pure diffusion problem in Section 4.3 Chapter 4, and also the analysis in Chapter 5, following the arbitrary number of faces per element mesh assumption 4.3.1. We present *a priori* bounds for the IP-DGFEM in $L^2(H^1)$– and $L^2(L^2)$–norms applied to the underlying time-dependent parabolic PDE. We begin by introducing the model problem, thereby extending the findings of Chapter 5 for this important case. The work contained in this chapter is drawn from [58].

6.1 Model problem

Let Ω be a bounded open polyhedral domain in $\mathbb{R}^d$, $d = 2, 3$, and let $J := (0, T)$ a time interval with $T > 0$. We consider the linear parabolic problem:

$$\begin{aligned} \partial_t u - \nabla \cdot (\mathbf{a} \nabla u) &= f \quad \text{in } J \times \Omega, \\ u|_{t=0} = u_0 \quad \text{on } \Omega, \quad &\text{and} \quad u = g_{\mathrm{D}} \quad \text{on } J \times \partial\Omega, \end{aligned} \tag{6.1}$$

for $f \in L^2(J; L^2(\Omega))$ and $\mathbf{a} \in L^\infty(J \times \Omega)^{d\times d}$, symmetric with

$$\boldsymbol{\xi}^\top \mathbf{a}(t, \mathbf{x})\boldsymbol{\xi} \geq \theta|\boldsymbol{\xi}|^2 > 0 \quad \forall \boldsymbol{\xi} \in \mathbb{R}^d, \quad \text{a.e.} \quad (t, x) \in J \times \Omega, \tag{6.2}$$

for some constant $\theta > 0$, $\mathbf{a}$ is allowed to depends on time t, which is different from diffusion tensor a in Chapter 4 and Chapter 5. Note that the differential operator $\nabla := (\partial_1, \partial_2, \cdots, \partial_d)$, i.e., is applied to the spatial variables only. For $u_0 \in L^2(\Omega)$ and $g_{\mathrm{D}} = 0$ the problem (6.1) is well-posed and there exists a unique solution $u \in L^2(J; H_0^1(\Omega))$ with $u \in C(\bar{J}; L^2(\Omega))$ and $\partial_t u \in L^2(J; H^{-1}(\Omega))$, see [139, 142].

6.2 Space-time DGFEMs for parabolic PDEs

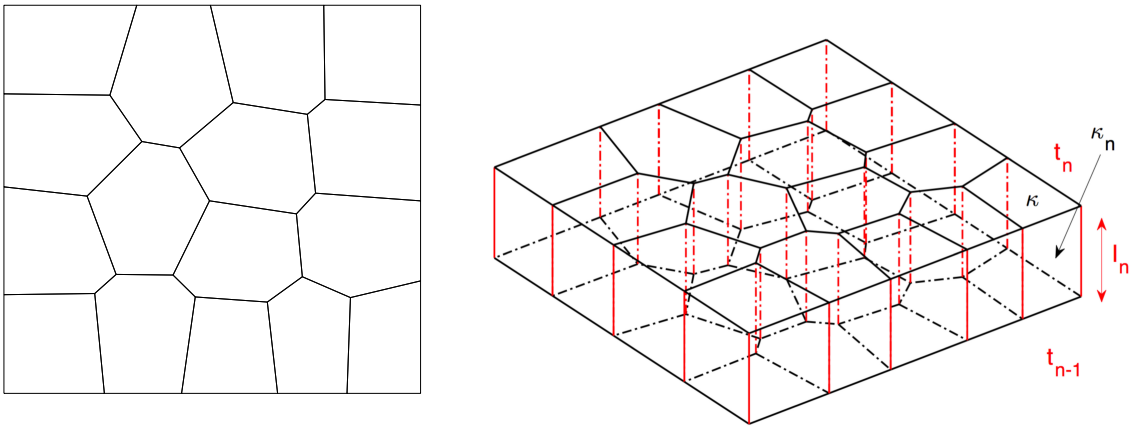
For notational consistency, d denotes the dimension of the spatial domain Ω. So the above parabolic problem (6.1) can be regarded as a $(d+1)$-dimensional PDE with non-negative characteristic form, with hyperbolicity along time and strong ellipticity over the spatial domain. We emphasize that the mesh Assumption 4.3.1 will be used through this chapter. It is possible, however, to repeat the analysis using Assumption 3.1.1, but this is not done here for brevity.

For the sake of simplicity, we consider PDEs with Dirichlet boundary condition $\partial\Omega = \partial\Omega_{\mathrm{D}}$, which implies $\mathcal{F}_h^{\mathcal{B}} = \mathcal{F}_h^{\mathcal{D}}$ and also $\mathcal{F}_h = \mathcal{F}_h^{\mathcal{I}} \cup \mathcal{F}_h^{\mathcal{D}}$. Spatial meshes $\kappa \in \mathcal{T}_h$ are defined in the same way as in the previous chapters.

Next, we introduce the temporal discretisation. Let $\mathcal{U}_h$ be a partition of the time interval J into N_t time steps $\{I_n\}_{n=1}^{N_t}$, with $I_n = (t_{n-1}, t_n)$ with respective set of nodes $\{t_n\}_{n=0}^{N_t}$ defined so that $0 := t_0 < t_1 < \cdots < t_{N_t} := T$. Set also $\lambda_n := t_n - t_{n-1}$, the length of I_n. For every time interval $I_n \in \mathcal{U}_h$ and every space element $\kappa \in \mathcal{T}_h$, we define the $(d+1)$-dimensional space-time *prismatic* element $\kappa_n := I_n \times \kappa$; see Figure 6.1 for an illustration. Let p_{κ_n} denote the (positive) *polynomial degree* of the space-time element κ_n, and collect p_{κ_n} in the vector $\mathbf{p} := (p_{\kappa_n} : \kappa_n \in \mathcal{U}_h \times \mathcal{T}_h)$. We define the *space-time finite element space* with respect to time interval I_n, subdivision $\mathcal{T}$, and $\mathbf{p}$ by

$$V^{\mathbf{p}}(I_n; \mathcal{T}_h) := \{u \in L^2(I_n \times \Omega) : u|_{\kappa_n} \in \mathcal{P}_{p_{\kappa_n}}(\kappa_n), \kappa_n \in I_n \times \mathcal{T}_h\},$$

where $\mathcal{P}_{p_{\kappa_n}}(\kappa)$ denotes the space of polynomials of *total degree* p_{κ_n} on κ_n. The space-time finite element space $S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)$ with respect to $\mathcal{U}_h$, $\mathcal{T}_h$, and $\mathbf{p}$ is defined as $S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h) = \bigoplus_{n=1}^{N_t} V^{\mathbf{p}}(I_n; \mathcal{T}_h)$. Note that the local elemental polynomial



(a) (b)

FIGURE 6.1: (a). 16 polygonal spatial elements over the spatial domain $\Omega = (0,1)^2$; (b) 16 space-time elements over $I_n \times \Omega$.

spaces employed in the definition of $S^{\mathbf{p}}(\mathcal{U}_h;\mathcal{T}_h)$ are defined in the *physical coordinate system*, without the need to map from a given reference/canonical frame; cf. [61]. This setting is crucial to retain full approximation of the finite element space, independently of the element shape. Note that $S^{\mathbf{p}}(\mathcal{U}_h;\mathcal{T}_h)$ employs fewer degrees of freedom per space-time element compared to the standard tensor-product polynomial bases of the usual space-time DGFEMs.

We shall also make use of the broken Sobolev space $H^{\mathbf{l}}(J\times\Omega,\mathcal{U}_h;\mathcal{T}_h)$, up to composite order $\mathbf{l} := (l_{\kappa_n} : \kappa_n \in \mathcal{U}_h \times \mathcal{T}_h)$ defined by

$$H^{\mathbf{l}}(J\times\Omega,\mathcal{U}_h;\mathcal{T}_h) = \{u \in L^2(J\times\Omega) : u|_{\kappa_n} \in H^{l_{\kappa_n}}(\kappa_n), \kappa_n \in \mathcal{U}_h\times\mathcal{T}_h\}. \tag{6.3}$$

For $u \in H^1(\Omega,\mathcal{T})$, we define the broken spatial gradient $(\nabla_h u)|_\kappa = \nabla(u|_\kappa)$, $\kappa \in \mathcal{T}$. Finally, let h_{κ_n} denote the diameter of the space-time element κ_n; for convenience, we collect the h_{κ_n} in the vector $\mathbf{h} := (h_{\kappa_n} : \kappa_n \in \mathcal{U}_h \times \mathcal{T}_h)$.

Remark 6.1. The main reason to introduce the space-time mesh diameter h_{κ_n} is that the proposed DGFEM is using space-time $\mathcal{P}_p$ basis on each element κ_n, $\forall\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$. So the appropriate function space for error analysis is the space-time Sobolev space rather than the Bochner space.

In order to work on the $(d+1)$-dimensional space-time elements $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, we introduce $\tilde{F}_t$ a generic d-dimensional face of a space-time element $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, which should be distinguished from the $(d-1)$-dimensional face F of the spatial element $\kappa \in \mathcal{T}_h$. For any space-time element $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, we define $\partial\kappa_n$ to be

the union of all d-dimensional open faces $\tilde{F}_t$ of κ_n. For convenience, we further subdivide $\tilde{F}_t$ into two disjoint subsets

$$\tilde{F}_t^{\parallel} := \tilde{F}_t \subset J \times \mathcal{F}_h, \quad \text{and} \quad \tilde{F}_t^{\perp} := \tilde{F}_t \subset \{t_n\}_{n=0}^{N_t} \times \Omega, \tag{6.4}$$

i.e., the parallel and perpendicular to the time direction boundaries, respectively. Hence, for each κ_n, there exist exactly two d-dimensional faces $\tilde{F}_t^{\perp}$ and the number of d-dimensional faces $\tilde{F}_t^{\parallel}$ is equal to the number of $(d-1)$-dimensional spatial faces F of the spatial element κ.

Next, we extend the definition of trace operators defined in Chapter 2 over the space-time element $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$. Let κ_n^1 and κ_n^2 be two adjacent space-time elements sharing a face $\tilde{F}_t^{\parallel} = \partial\kappa_n^1 \cap \partial\kappa_n^2$ and $(t,x) \in \tilde{F}_t^{\parallel} \subset J \times \mathcal{F}_h^{\mathcal{I}}$; let also $\bar{\mathbf{n}}_{\kappa_n^1}$ and $\bar{\mathbf{n}}_{\kappa_n^2}$ denote the outward unit normal vectors on $\tilde{F}_t^{\parallel}$, relative to $\partial\kappa_n^1$ and $\partial\kappa_n^2$, respectively. Then, for v and $\mathbf{q}$, scalar- and vector-valued functions, respectively, smooth enough for their traces on $\tilde{F}_t^{\parallel}$ to be well defined, we define the *averages* $\{\!\{v\}\!\} := \frac{1}{2}(v|_{\kappa_n^1} + v|_{\kappa_n^2})$, $\{\!\{\mathbf{q}\}\!\} := \frac{1}{2}(\mathbf{q}|_{\kappa_n^1} + \mathbf{q}|_{\kappa_n^2})$, and the *jumps* $[\![v]\!] := v_{\kappa_n^1}\,\bar{\mathbf{n}}_{\kappa_n^1} + v_{\kappa_n^2}\,\bar{\mathbf{n}}_{\kappa_n^2}$, $[\![\mathbf{q}]\!] := \mathbf{q}_{\kappa_n^1} \cdot \bar{\mathbf{n}}_{\kappa_n^1} + \mathbf{q}_{\kappa_n^2} \cdot \bar{\mathbf{n}}_{\kappa_n^2}$, respectively. On a boundary face $\tilde{F}_t^{\parallel} \subset J \times \mathcal{F}_h^{\mathcal{B}}$ and $\tilde{F}_t^{\parallel} \subset \partial\kappa_n$, we set $\{\!\{v\}\!\} = v|_{\kappa_n}$, $\{\!\{\mathbf{q}\}\!\} = \mathbf{q}|_{\kappa_n}$, $[\![v]\!] = v|_{\kappa_n}\bar{\mathbf{n}}_{\kappa_n}$, $[\![\mathbf{q}]\!] = \mathbf{q}|_{\kappa_n} \cdot \bar{\mathbf{n}}_{\kappa_n}$, with $\mathbf{n}_{\kappa_n}$ denoting the unit outward normal vector on the boundary. Upon defining

$$u_n^+ := \lim_{s \to 0^+} u(t_n + s),\ 0 \le n \le N_t - 1, \quad u_n^- := \lim_{s \to 0^+} u(t_n - s),\ 1 \le n \le N_t,$$

the *time-jump* across t_n, $n = 1, \ldots, N_t - 1$ is given by $\lfloor u \rfloor_n := u_n^+ - u_n^-$.

Remark 6.2. The above *time-jump* across different time nodes is exactly the same *upwind-jump*, due to the fact that the hyperbolicity of parabolic problem is only along the time direction.

Equipped with the above notation, we can now describe the space-time discontinuous Galerkin method for the problem (6.1), reading: find $u_h \in S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)$ such that

$$B(u_h, v_h) = \ell(v_h), \qquad \text{for all } v_h \in S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h), \tag{6.5}$$

where $B : S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h) \times S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h) \to \mathbb{R}$ is defined as

$$B(u,v) := \sum_{n=1}^{N_t} \int_{I_n} \big((\partial_t u, v) + B_{\mathrm{d}}(u,v) \big)\, \mathrm{d}t + \sum_{n=2}^{N_t} (\lfloor u \rfloor_{n-1}, v_{n-1}^+) + (u_0^+, v_0^+), \tag{6.6}$$

with the spatial IP-DGFEM bilinear form $B_{\rm d}(\cdot,\cdot)$ given by

$$B_{\rm d}(u,v) := \sum_{\kappa\in\mathcal{T}_h}\int_\kappa \mathbf{a}\nabla u\cdot\nabla v\,{\rm d}\mathbf{x} - \sum_{F\in\mathcal{F}_h}\int_F \Big(\{\!\{\mathbf{a}\nabla u\}\!\}\cdot[\![v]\!]+\{\!\{\mathbf{a}\nabla v\}\!\}\cdot[\![u]\!]-\sigma[\![u]\!]\cdot[\![v]\!]\Big)\,{\rm d}s,$$

and the linear functional $\ell : S^{\mathbf{p}}(\mathcal{U}_h;\mathcal{T}_h)\to\mathbb{R}$ given by

$$\ell(v) := \sum_{n=1}^{N_t}\int_{I_n}\Big((f,v)-\sum_{F\in\mathcal{F}_h^{\mathcal{D}}}\int_F g_{\rm D}\Big((\mathbf{a}\nabla_h v)\cdot\mathbf{n}-\sigma v\Big)\,{\rm d}s\Big)\,{\rm d}t + (u_0,v_0^+).$$

The nonnegative function $\sigma\in L^\infty(J\times\mathcal{F}_h)$ appearing in $B_{\rm d}$ and ℓ above is again referred to as the *discontinuity-penalization function*; its precise definition, depending on the diffusion tensor $\mathbf{a}$ and on the discretization parameters, will be given in Lemma 6.5 in next section.

The use of prismatic meshes is key in that it permits us to solve for each time-step separately: for each time interval $I_n\in\mathcal{U}_h$, $n=2,\ldots,N_t$, the solution $U_n = u_h|_{I_n}\in V^{\mathbf{p}}(I_n;\mathcal{T}_h)$ is given by:

$$\begin{aligned}&\int_{I_n}(\partial_t U_n,V_n)+B_{\rm d}(U_n,V_n)\,{\rm d}t + (U_{n-1}^+,V_{n-1}^+)\\ &=\int_{I_n}\Big((f,V_n)-\sum_{F\in\mathcal{F}_h^{\mathcal{D}}}\int_F g_{\rm D}\big((\mathbf{a}\nabla_h V_n)\cdot\mathbf{n}-\sigma V_n\big)\,{\rm d}s\Big)\,{\rm d}t + (U_{n-1}^-,V_{n-1}^+),\end{aligned}\tag{6.7}$$

for all $V_n\in V^{\mathbf{p}}(I_n;\mathcal{T}_h)$, with U_{n-1}^- serving as the initial datum at time step I_n; for $n=1$, we set $U_0^- = u_0$.

6.2.1 Inf-sup Stability of space-time DGFEMs

We shall establish the unconditional stability of the above space-time DGFEMs, via the derivation of an inf-sup condition for arbitrary aspect ratio between the time-step and the local spatial mesh-size. The proof circumvents the global shape-regularity assumption, required in the respective result in Theorem 5.5 in Chapter 5 for the case of parabolic problems, since hyperbolicity is only imposed along time.

Lemma 6.3. *Let $v\in\mathcal{P}_{p_{\kappa_n}}(\kappa_n)$, $\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h$ and $\Theta\in\{\kappa,F\}$. Then, there exist positive constants $C_{\rm inv,6}$ and $C_{\rm inv,7}$, independent of v, κ_n, λ_n and p_{κ_n}, such that*

$$\|v\|^2_{L^2(\tilde{F}_t^\perp)}\le C_{\rm inv,6}\frac{p_{\kappa_n}^2}{\lambda_n}\|v\|^2_{L^2(\kappa_n)},\tag{6.8}$$

$$\|\partial_t v\|^2_{L^2(I_n;L^2(\Theta))} \leq C_{\text{inv},7}\frac{p^4_{\kappa_n}}{\lambda_n^2}\|v\|^2_{L^2(I_n;L^2(\Theta))}. \tag{6.9}$$

Proof. We note that $(d+1)$-dimensional spatial element $\kappa_n := I_n \times \kappa$. So we start the proof over the reference time interval $\hat{I} := (-1,1)$ and then the general result can be derived by using the scaling augment. We start with (6.8). For $v \in \mathcal{P}_{p_{\kappa_n}}(\hat{I}\times\kappa)$, we can rewrite v into the Legendre polynomial $L_n(\hat{t})$ up to order p_{κ_n} with respect to variable $\hat{t}$ over reference temporal interval $\hat{I} := (-1,1)$, such that

$$v(\hat{t},\mathbf{x}) = \sum_{n=0}^{p_{\kappa_n}} a_n(\mathbf{x})L_n(\hat{t}) \quad \text{with} \quad a_n(\mathbf{x}) = \frac{2n+1}{2}\int_{\hat{I}} v(\hat{t},\mathbf{x})L_n(\hat{t})\,\mathrm{d}t, \tag{6.10}$$

with L^2–norm

$$\|v(\hat{t},\mathbf{x})\|^2_{L^2(\hat{I}\times\kappa)} = \int_\kappa \sum_{n=0}^{p_{\kappa_n}}\sum_{m=0}^{p_{\kappa_n}} a_n(\mathbf{x})a_m(\mathbf{x})\frac{2\delta_{mn}}{2n+1}\,\mathrm{d}\mathbf{x} = \sum_{n=0}^{p_{\kappa_n}}\|a_n\|^2_{L^2(\kappa)}\frac{2}{2n+1}. \tag{6.11}$$

Here, we have used the orthogonality of Legendre polynomials. The coefficient $a_n(\mathbf{x})$ is a function of the spatial variable $\mathbf{x}$ only. To be more precise, they are polynomial of total degree up to $p_{\kappa_n} - n$ in spatial variables. Next, we have the following result:

$$\|v(\pm 1,\mathbf{x})\|_{L^2(\kappa)} \;\leq\; \sum_{n=0}^{p_{\kappa_n}}\|a_n(\mathbf{x})\|_{L^2(\kappa)}|L_n(\pm 1)| \leq \frac{(p_{\kappa_n}+1)}{\sqrt{2}}\|v(\hat{t},\mathbf{x})\|_{L^2(\kappa_n)}. \tag{6.12}$$

Here, we used the Cauchy-Schwartz inequality and the fact that $|L_n(\pm 1)| = 1$. By using the scaling argument, then (6.8) is proved. Next, we prove (6.9). Here, we introduce set Θ to denote d-dimensional spatial element κ or $(d-1)$-dimensional spatial face F. We first introduce the following result:

$$\|\partial_t v\|_{L^2(\hat{I})} \leq \sum_{n=0}^{p_{\kappa n}}|a_n(\mathbf{x})|\|L_n'(\hat{t})\|_{L^2(\hat{I})} \leq \sum_{n=0}^{p_{\kappa n}}|a_n(\mathbf{x})|(n(n+1))^{1/2} \leq \sqrt{3}p^2_{\kappa_n}\|v\|_{L^2(\hat{I})}. \tag{6.13}$$

Here, we have used result $\|L_n'(\hat{t})\|^2_{L^2(\hat{I})} = n(n+1)$; see [167] for detail. Then we use above result together with Fubini's theorem to derive the following result:

$$\|\partial_t v\|^2_{L^2(\hat{I};L^2(\Theta))} \leq 3p^4_{\kappa_n}\int_\Theta \|v\|^2_{L^2(\hat{I})}\,\mathrm{d}\Theta = 3p^4_{\kappa_n}\|v\|^2_{L^2(\hat{I};L^2(\Theta))}. \tag{6.14}$$

Finally, we can use the scaling argument to derive (6.9) . □

Remark 6.4. The proof of Lemma 6.3 can be viewed as an extension of anisotropic tensor product elements with anisotropic tensor product $\mathcal{Q}_p$ polynomial basis [103, 163, 106]. In our scheme, the spatial mesh is a general polytopic mesh and the polynomial basis here is total degree basis $\mathcal{P}_p$. Due to the fact that the space time element $\kappa_n = I_n \times \kappa$ is constructed by tensor product of spatial and temporal meshes. All inverse estimates related to time variable can be treated as one dimensional inverse estimation problems with respect to time variable. The resulting inverse estimate is sharp in the sense that it only depends on temporal mesh size λ_n.

For the forthcoming stability analysis, we introduce an inconsistent bilinear form $B_{\mathrm{d}}(\cdot,\cdot)$: for $u, v \in \mathcal{S} := L^2(J; H^1(\Omega)) \cap H^1(J; H^{-1}(\Omega)) + S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)$, we set

$$\tilde{B}(u,v) := \sum_{n=1}^{N_t} \int_{I_n} \big((\partial_t u, v) + \tilde{B}_{\mathrm{d}}(u,v)\big)\,\mathrm{d}t + \sum_{n=2}^{N_t} (\lfloor u \rfloor_{n-1}, v_{n-1}^+) + (u_0^+, v_0^+), \tag{6.15}$$

where

$$\begin{aligned} \tilde{B}_{\mathrm{d}}(u,v) \;&:=\; \sum_{\kappa\in\mathcal{T}_h} \int_\kappa \mathbf{a}\nabla u \cdot \nabla v \,\mathrm{d}\mathbf{x} \\ &-\; \sum_{F\in\mathcal{F}_h} \int_F \Big(\{\!\{ \mathbf{a}\boldsymbol{\Pi}_2(\nabla u) \}\!\} \cdot [\![v]\!] + \{\!\{ \mathbf{a}\boldsymbol{\Pi}_2(\nabla v) \}\!\} \cdot [\![u]\!] - \sigma [\![u]\!] \cdot [\![v]\!] \Big)\,\mathrm{d}s, \end{aligned}$$

and a modified linear functional $\tilde{\ell} : \mathcal{S} \to \mathbb{R}$, given by

$$\tilde{\ell}(v) := \sum_{n=1}^{N_t} \int_{I_n} \Big((f,v) - \sum_{F\in\mathcal{F}_h^{\mathcal{D}}} \int_F g_{\mathrm{D}} \Big(\mathbf{a}\boldsymbol{\Pi}_2(\nabla_h v)\cdot\mathbf{n} - \sigma v \Big)\,\mathrm{d}s \Big)\,\mathrm{d}t + (u_0, v_0^+).$$

Here, $\boldsymbol{\Pi}_2 : [L^2(J; L^2(\Omega))]^d \to [S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)]^d$ denotes the vector-valued L^2–projection onto $[S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)]^d$. It is immediately clear, therefore, that $B(u_h, v_h) = \tilde{B}(u_h, v_h)$ and that $l(v_h) = \tilde{l}(v_h)$, for all $v_h \in S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)$.

By recalling the definition of $\sqrt{\mathbf{a}}$ be the square root of $\mathbf{a}$ and set $\bar{\mathbf{a}}_{\kappa_n} = |\sqrt{\mathbf{a}}|_2^2|_{\kappa_n}$, for $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, with $|\cdot|_2$ denoting the matrix l_2–norm. We introduce the DG–norm $|||\cdot|||_{\mathrm{DG}}$:

$$|||v|||_{\mathrm{DG}} := \Big(\int_J |||v|||_{\mathrm{d}}^2 \,\mathrm{d}t + \frac{1}{2}\|v_0^+\|^2 + \sum_{n=1}^{N_t-1} \frac{1}{2}\|\lfloor v \rfloor_n\|^2 + \frac{1}{2}\|v_{N_t}^-\|^2 \Big)^{1/2}, \tag{6.16}$$

with

$$|||v|||_{\mathrm{d}}^2 := \sum_{\kappa\in\mathcal{T}_h} \|\sqrt{a}\nabla v\|_{L^2(\kappa)}^2 + \sum_{F\in\mathcal{F}_h} \int_F \sigma|[\![v]\!]|^2 \,\mathrm{d}s. \tag{6.17}$$

The continuity and coercivity of the inconsistent diffusion bilinear form $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$ with respect to the diffusion DG–norm $|||\cdot|||_{\mathrm{d}}$ is established in following lemma.

Lemma 6.5. *Let Assumption 4.3.1 holds and let $\sigma : J \times \mathcal{F}_h \to \mathbb{R}_+$ be defined face-wise over all $\tilde{F}_t^{\|}$ by*

$$\sigma(t,x) := \begin{cases} C_\sigma \displaystyle\max_{\kappa_n:\tilde{F}_t^{\|}\cap\bar{\kappa}_n\neq\emptyset} \Big\{\frac{\bar{\mathbf{a}}_{\kappa_n}^2(p_{\kappa_n}+1)(p_{\kappa_n}+d)}{h_\kappa}\Big\}, & \tilde{F}_t^{\|}\subset J\times\mathcal{F}_h^{\mathcal{I}}, \\ C_\sigma \displaystyle\max_{\kappa_n:\tilde{F}_t^{\|}\cap\bar{\kappa}_n\neq\emptyset} \frac{\bar{\mathbf{a}}_{\kappa_n}^2(p_{\kappa_n}+1)(p_{\kappa_n}+d)}{h_\kappa}, & \tilde{F}_t^{\|}\subset J\times\mathcal{F}_h^{\mathcal{D}}, \end{cases} \tag{6.18}$$

with $C_\sigma > 0$ sufficiently large, independent of discretization parameters and the number of faces per element. Then, for all $v \in \mathcal{S}$, we have

$$\int_J \tilde{B}_{\mathrm{d}}(v,v)\,\mathrm{d}t \geq C_{\mathrm{coer}} \int_J |||v|||_{\mathrm{d}}^2 \,\mathrm{d}t, \tag{6.19}$$

$$\int_J \tilde{B}_{\mathrm{d}}(w,v)\,\mathrm{d}t \leq C_{\mathrm{cont}} \int_J |||w|||_{\mathrm{d}}\,|||v|||_{\mathrm{d}}\,\mathrm{d}t, \tag{6.20}$$

$$\tilde{B}(v,v) \geq \bar{C}|||v|||_{\mathrm{DG}}^2, \tag{6.21}$$

for all $v \in \mathcal{S}$, with the positive constants C_{coer}, C_{cont} and $\bar{C}$, independent of the discretization parameters, the number of faces per element, and of v.

Proof. The proof of coercivity in relation (6.19) and continuity in relation (6.20) under the mesh Assumption 4.3.1 are exactly the same as in Lemma 4.12 in Section 4.3. Here, C_{coer} depends on the shape regularity constant C_s and also the uniform ellipticity constant θ. Hence, the bilinear form $\tilde{B}_{\mathrm{d}}(\cdot,\cdot)$ is coercive over $\mathcal{S}\times\mathcal{S}$ for $\epsilon > 1/2$ and $C_\sigma > 2C_s\epsilon/\theta$. C_σ depends on constant C_s, but is independent of the number of faces per element.

For (6.21), integration by parts on the first term on the right-hand side of (6.15) along with (6.19) yield

$$\begin{aligned}\tilde{B}(v,v) &\geq C_{\text{coer}} \int_J |||v|||_d^2 \,\mathrm{d}t + \frac{1}{2}\|v_0^+\|^2 + \sum_{n=1}^{N_t-1} \frac{1}{2}\|\lfloor v \rfloor_n\|^2 + \frac{1}{2}\|v_{N_t}^-\|^2 \\ &\geq \bar{C}|||v|||_{\text{DG}}^2,\end{aligned}$$

with $\bar{C} = \min\{1, C_{\text{coer}}\}$. □

Remark 6.6. Our approach is dictated by the shape regularity Assumption 4.3.1 allowing for an *arbitrary* number of faces per element. In contrast, if mesh Assumption 3.1.1 is employed, no shape regularity was explicitly assumed at the expense of imposing a uniform bound on the number of faces per element. Clearly, the two approaches can be combined to produce admissible discretisations on even more general mesh settings; we refrain from doing so here in the interest of brevity and we refer to the forthcoming [60] for the complete treatment.

Moreover, the coercivity constant may depend on the shape regularity constant C_s and on the uniform ellipticity constant θ. To avoid the dependence on the latter, it is possible to combine the present developments with the DGFEM proposed in [105]; we refrain from doing so here, in the interest of simplicity of the presentation.

Before we prove the inf-sup condition, we briefly talk about the reasons why the inf-sup condition is essential for the proposed space-time DGFEMs. The classical *a priori* error analysis for DG time-stepping scheme depends highly on utilising the tensor product structure of the space time basis. The optimal error bound in various norms depends on using special temporal projections introduced by Thomée [180] and elliptic projection introduced by Wheeler [186] over spatial domains, see also [159] for the *hp*–version *a priori* error analysis. Due to the lack of the space-time tensor product structure of the basis, we can not use the classical techniques to do error analysis. To address this issue we prove an inf-sup condition for the inconsistent bilinear form $\tilde{B}(\cdot,\cdot)$, with respect to the following streamline diffusion DGFEM–norm.

Definition 6.7. The streamline diffusion DGFEM–norm is defined by:

$$|||v|||_{\text{s}}^2 := |||v|||_{\text{DG}}^2 + \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n} \|\partial_t v\|_{L^2(\kappa_n)}^2, \tag{6.22}$$

where

$$\tau_{\kappa_n} := \frac{\lambda_n}{\hat{p}^2_{\kappa_n}}, \quad \forall \kappa_n \in \mathcal{U}_h \times \mathcal{T}_h,$$

for $p_{\kappa_n} \geq 1$ and $\hat{p}_{\kappa_n}$ defined as

$$\hat{p}_{\kappa_n} := \max_{\tilde{F}_t^{\parallel} \subset \partial\kappa_n} \left\{ \max_{\substack{\tilde{\kappa}_n \in \{\kappa_n, \kappa_n'\} \\ \tilde{F}_t^{\parallel} \subset \partial\kappa_n \cap \partial\kappa_n'}} \left\{ p_{\tilde{\kappa}_n} \right\} \right\}, \quad \forall \kappa_n \in \mathcal{U}_h \times \mathcal{T}_h; \tag{6.23}$$

$\hat{p}_{\kappa_n}$ is the largest polynomial order among each element $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$ and their spatial neighbouring elements.

Theorem 6.8. *Given Assumption 4.3.1, there exists a constant $\Lambda_s > 0$, independent of the temporal and spatial mesh sizes λ_n, h_κ, of the polynomial degree p_{κ_n} and of the number of faces per element, such that:*

$$\inf_{\nu \in S^{\mathbf{p}}(\mathcal{U}_h;\mathcal{T}_h)\setminus\{0\}} \sup_{\mu \in S^{\mathbf{p}}(\mathcal{U}_h;\mathcal{T}_h)\setminus\{0\}} \frac{\tilde{B}(\nu,\mu)}{|||\nu|||_{\mathrm{s}} |||\mu|||_{\mathrm{s}}} \geq \Lambda_s. \tag{6.24}$$

Proof. For $\nu \in S^{\mathbf{p}}(\mathcal{U}_h;\mathcal{T}_h)$, we select $\mu := \nu + \alpha\nu_s$, with $\nu_s|_{\kappa_n} := \tau_{\kappa_n}\partial_t\nu$, $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, with $0 < \alpha \in \mathbb{R}$, at our disposal. Then, (6.24) follows if both the following:

$$|||\mu|||_{\mathrm{s}} \leq C^* |||\nu|||_{\mathrm{s}}, \tag{6.25}$$

and

$$\tilde{B}(\nu,\mu) \geq C_* |||\nu|||^2_{\mathrm{s}}, \tag{6.26}$$

hold, with $C^* > 0$ and $C_* > 0$ constants independent of h_κ, λ_n, p_{κ_n}, the number of faces per element, and $\Lambda_s = C_*/C^*$.

To show (6.25), we start by considering the jump terms at time nodes $\{t_n\}_{n=0}^{N_t}$. Employing (6.8), we have

$$\begin{aligned}
&\frac{1}{2}\|(\nu_s^+)_0\|^2 + \sum_{n=1}^{N_t-1} \frac{1}{2}\|\lfloor \nu_s \rfloor_n\|^2 + \frac{1}{2}\|(\nu_s^-)_{N_t}\|^2 \\
&\quad \leq \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau^2_{\kappa_n} \sum_{\tilde{F}_t^{\perp} \subset \partial\kappa_n} \|\partial_t\nu\|^2_{L^2(\tilde{F}_t^{\perp})} \\
&\quad \leq \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} 2C_{\mathrm{inv},6} \frac{\tau_{\kappa_n} p^2_{\kappa_n}}{\lambda_n} \Big(\tau_{\kappa_n} \|\partial_t\nu\|^2_{L^2(\kappa_n)} \Big) \leq C_1 |||\nu|||^2_{\mathrm{s}}.
\end{aligned} \tag{6.27}$$

Using (6.9) with $\Theta = \kappa$ and relation (6.23), the second term on the right-hand side of (6.22) is estimated by

$$\sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n} \|\partial_t \nu_s\|^2_{L^2(\kappa_n)} \leq \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} C_{\mathrm{inv},7} \frac{\tau^2_{\kappa_n} p^4_{\kappa_n}}{\lambda^2_n} \Big(\tau_{\kappa_n} \|\partial_t \nu\|^2_{L^2(\kappa_n)} \Big) \leq C_2 |||\nu|||^2_{\mathrm{s}}. \tag{6.28}$$

Next, for the first term on the right-hand side of (6.22), employing (6.9) with $\Theta = \kappa$, the uniform ellipticity condition (6.2), together with Fubini's theorem, we have

$$\begin{aligned} \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \|\sqrt{\mathbf{a}} \nabla \nu_s\|^2_{L^2(\kappa_n)} &\leq \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \bar{\mathbf{a}}_{\kappa_n} \tau^2_{\kappa_n} \|\partial_t (\nabla \nu)\|^2_{L^2(\kappa_n)} \qquad (6.29) \\ &\leq \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \bar{\mathbf{a}}_{\kappa_n} C_{\mathrm{inv},7} \frac{\tau^2_{\kappa_n} p^4_{\kappa_n}}{\lambda^2_n} \|\nabla \nu\|^2_{L^2(\kappa_n)} \\ &\leq \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} C_{\mathrm{inv},7} \frac{\bar{\mathbf{a}}_{\kappa_n}}{\theta} \frac{\tau^2_{\kappa_n} p^4_{\kappa_n}}{\lambda^2_n} \|\sqrt{\mathbf{a}} \nabla \nu\|^2_{L^2(\kappa_n)} \leq C_3 |||\nu|||^2_{\mathrm{s}}. \end{aligned}$$

Finally, employing (6.9) with $\Theta = F$ and (6.23), we have

$$\begin{aligned} \sum_{F \in \mathcal{F}_h} \int_J \int_F \sigma |[\![\nu_s]\!]|^2 \,\mathrm{d}s \,\mathrm{d}t &= \sum_{\tilde{F}^{\|}_t \subset J \times \mathcal{F}_h} \sigma \tau^2_{\kappa_n} \|\partial_t [\![\nu]\!]\|^2_{L^2(\tilde{F}^{\|}_t)} \\ &\leq \sum_{\tilde{F}^{\|}_t \subset J \times \mathcal{F}_h} \sigma C_{\mathrm{inv},7} \frac{\tau^2_{\kappa_n}}{\lambda^2_n} \Big(\max_{\substack{\tilde{\kappa}_n \in \{\kappa_n, \kappa'_n\} \\ \tilde{F}^{\|}_t \subset \partial\kappa_n \cap \partial\kappa'_n}} \{p_{\tilde{\kappa}_n}\} \Big)^4 \|[\![\nu]\!]\|^2_{L^2(\tilde{F}^{\|}_t)} \\ &\leq \sum_{\tilde{F}^{\|}_t \subset J \times \mathcal{F}_h} \sigma C_{\mathrm{inv},7} \|[\![\nu]\!]\|^2_{L^2(\tilde{F}^{\|}_t)} \leq C_4 |||\nu|||^2_{\mathrm{s}}. \qquad (6.30) \end{aligned}$$

Combining the above, we have $|||\nu_s|||_{\mathrm{s}} \leq \hat{C} |||\nu|||_{\mathrm{s}}$, where $\hat{C} = \sqrt{\sum_{i=1}^4 C_i}$, or

$$|||\mu|||_{\mathrm{s}} \leq |||\nu|||_{\mathrm{s}} + \alpha |||\nu_s|||_{\mathrm{s}} \leq (1 + \alpha \hat{C}) |||\nu|||_{\mathrm{s}} \equiv C^*(\alpha) |||\nu|||_{\mathrm{s}}. \tag{6.31}$$

For (6.26), we start by noting that $\tilde{B}(\nu, \mu) = \tilde{B}(\nu, \nu) + \alpha \tilde{B}(\nu, \nu_s)$. Also

$$\begin{aligned} \tilde{B}(\nu, \nu_s) &= \sum_{n=1}^{N_t} \Big(\sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n} \|\partial_t v\|^2_{L^2(\kappa_n)} + \int_{I_n} \tilde{B}_{\mathrm{d}}(\nu, \nu_s) \,\mathrm{d}t \Big) \\ &+ \sum_{n=2}^{N_t} (\lfloor \nu \rfloor_{n-1}, (\nu_s)^+_{n-1}) + (\nu^+_0, (\nu_s)^+_0). \end{aligned}$$

Further, using (6.8), we have

$$\begin{aligned}
&\sum_{n=2}^{N_t}(\lfloor\nu\rfloor_{n-1},(\nu_s)^+_{n-1})+(\nu_0^+,(\nu_s)_0^+)\\
&\leq \sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}\|\lfloor\nu\rfloor\|_{L^2(\tilde{F}_t^\perp\subset\partial\kappa_n)}\Big(\tau_{\kappa_n}\sum_{\tilde{F}_t^\perp\subset\partial\kappa_n}\|\partial_t\nu\|_{L^2(\tilde{F}_t)}\Big)\\
&\leq 4C_{\mathrm{inv},6}\Big(\|\nu_0^+\|^2+\|\lfloor\nu\rfloor_n\|^2+\|\nu_{N_t}^-\|^2\Big)+\sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}\frac{\tau_{\kappa_n}}{4}\|\partial_t\nu\|^2_{L^2(\kappa_n)},
\end{aligned}\tag{6.32}$$

where, with slight abuse of notation, we have extended the definition of the time jump $\lfloor\nu\rfloor$ to time boundary faces. Next, from (6.20), together with (6.29) and (6.30), we get

$$\begin{aligned}
\sum_{n=1}^{N_t}\int_{I_n}\tilde{B}_{\mathrm{d}}(\nu,\nu_s)\,\mathrm{d}t &\leq \sum_{n=1}^{N_t}\int_{I_n}C_{\mathrm{cont}}|||\nu|||_{\mathrm{d}}|||\nu_s|||_{\mathrm{d}}\,\mathrm{d}t\\
&\leq \frac{(C_{\mathrm{cont}})^2}{2}\int_J|||\nu|||^2_{\mathrm{d}}\,\mathrm{d}t+\frac{1}{2}\int_J|||\nu_s|||^2_{\mathrm{d}}\,\mathrm{d}t\\
&\leq \frac{(C_{\mathrm{cont}})^2+C_3+C_4}{2}\int_J|||\nu|||^2_{\mathrm{d}}\,\mathrm{d}t.
\end{aligned}\tag{6.33}$$

Combining (6.21) with (6.32) and (6.33), we arrive at

$$\begin{aligned}
\tilde{B}(\nu,\mu) &= \tilde{B}(\nu,\nu)+\alpha\tilde{B}(\nu,\nu_s)\\
&\geq \Big(\frac{1}{2}-4\alpha C_{\mathrm{inv},6}\Big)\Big(\|\nu_0^+\|^2_{L^2(\Omega)}+\sum_{n=1}^{N_t-1}\|\lfloor\nu\rfloor_n\|^2_{L^2(\Omega)}+\|\nu_{N_t}^-\|^2_{L^2(\Omega)}\Big)\\
&+ \Big(C_{\mathrm{coer}}-\alpha\frac{(C_{\mathrm{cont}})^2+C_3+C_4}{2}\Big)\int_J|||\nu|||^2_{\mathrm{d}}\,\mathrm{d}t\\
&+ \sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}\alpha\Big(\tau_{\kappa_n}-\frac{\tau_{\kappa_n}}{4}\Big)\|\partial_t\nu\|^2_{L^2(\kappa_n)}.
\end{aligned}$$

The coefficients in front of the norms arising on the right hand side of the above bound are all positive if

$$\alpha<\min\Big\{\frac{1}{(8C_{\mathrm{inv},6})},\frac{2C_{\mathrm{coer}}}{((C_{\mathrm{cont}})^2+C_3+C_4)}\Big\},$$

with the latter independent of the discretization parameters and the number of faces per element. □

The above result shows that the space-time DGFEM based on the reduced *total-degree-p* space-time basis is well posed. It extends the stability proof from [59]

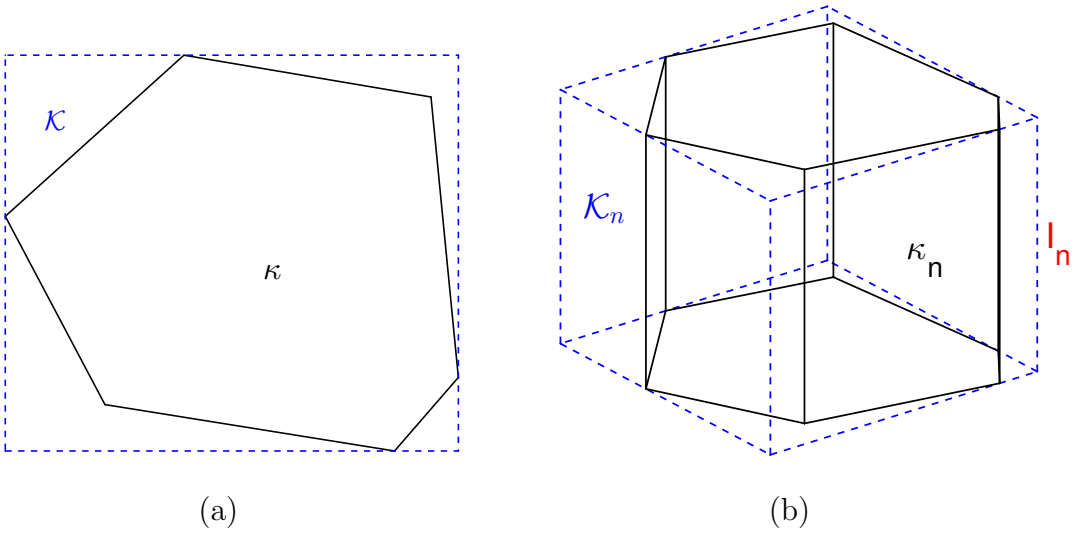


FIGURE 6.2: (a). Polygonal spatial element κ and covering $\mathcal{K}$; (b) space-time element $\kappa_n = I_n \times \kappa$ and covering $\mathcal{K}_n := I_n \times \mathcal{K}$.

to space-time elements with arbitrarily large aspect ratio between the time-step λ_n and local mesh-size h_κ for parabolic problems. Moreover, the inf-sup stability result holds *without any assumptions on the number of faces per spatial mesh*, too. Therefore, the scheme is shown to be stable for extremely general, possibly anisotropic, space-time meshes.

The above inf-sup condition will be instrumental in the proof of the a priori error bounds below, as the total-degree-p space-time basis does not allow for classical space-time tensor-product arguments [180] to be employed.

6.2.2 A priori error analysis in $L^2(H^1)$–norm

In view of using known approximation results, we shall require a shape-regularity assumption for the space-time elements.

Assumption 6.2.1. *We assume the existence of a constant $c_{reg} > 0$ such that*

$$c_{reg}^{-1} \leq h_\kappa/\lambda_n \leq c_{reg},$$

uniformly for all $\kappa_n \in \mathcal{U} \times \mathcal{T}$, i.e., the space-time elements are also shape-regular.

In this section, we need to slightly modify the Definition 3.9 for the spatial mesh coverings.

Definition 6.9. A *covering* $\mathcal{T}_h^\sharp = \{\mathcal{K}\}$ related to the polytopic mesh $\mathcal{T}_h$ is a set of shape-regular d–simplices or hypercubes $\mathcal{K}$, such that for each $\kappa \in \mathcal{T}_h$, there exists a $\mathcal{T}_h^\sharp$, with $\kappa \subset \mathcal{K}$. We refer to Figure 6.2(a) for an illustration. Given $\mathcal{T}_h^\sharp$, we denote by $\Omega_\sharp$ the *covering domain* given by $\Omega_\sharp := \left(\cup_{\mathcal{K}\in\mathcal{T}_\sharp}\bar{\mathcal{K}}\right)^\circ$, with D° denoting the interior of a set $D \subset \mathbb{R}^d$.

The covering $\mathcal{K}$ satisfies the Assumption 3.3.1. As a consequence, we have $\mathrm{diam}(\mathcal{K}) \leq C_{\mathrm{diam}} h_\kappa$, for each pair $\kappa \in \mathcal{T}_h$, $\mathcal{K} \in \mathcal{T}_h^\sharp$, with $\kappa \subset \mathcal{K}$, for a constant $C_{\mathrm{diam}} > 0$, uniformly with respect to the mesh size.

Theorem 6.10 (Stein)**.** *Let Ω be a domain with a Lipschitz boundary. Then, there exists a linear extension operator $\mathfrak{E} : H^s(\Omega) \to H^s(\mathbb{R}^d)$, $s \in \mathbb{N}_0$, such that $\mathfrak{E}v|_\Omega = v$ and $\|\mathfrak{E}v\|_{H^s(\mathbb{R}^d)} \leq C\|v\|_{H^s(\Omega)}$, with $C > 0$ constant depending only on s and Ω.*

Moreover, we shall also denote by $\mathfrak{E}v$ the (trivial) space-time extension $\mathfrak{E}v : L^2(J; H^s(\Omega)) \to L^2(J; H^s(\mathbb{R}^d))$ defined as the spatial extension above, for every $t \in J$. Next, we present the *hp*-approximation results in next lemma.

Lemma 6.11. *Let $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, $\tilde{F}_t \subset \partial\kappa_n$ a face, and $\mathcal{K} \in \mathcal{T}_h^\sharp$ as in Definition 6.9 and let $\mathcal{K}_n = I_n \times \mathcal{K}$ (see Figure 6.2(b) for an illustration). Let $v \in L^2(J \times \Omega)$, such that $\mathfrak{E}v|_{\mathcal{K}_n} \in H^{l_{\kappa_n}}(\mathcal{K}_n)$, for some $l_{\kappa_n} \geq 0$. Suppose also that Assumptions 6.2.1 and 3.3.1 hold. Then, there exists $\tilde{\Pi}v|_{\kappa_n} \in \mathcal{P}_{p_{\kappa_n}}(\kappa_n)$, such that*

$$\|v - \tilde{\Pi}v\|_{H^q(\kappa_n)} \leq C\frac{h_{\kappa_n}^{s_{\kappa_n}-q}}{p_{\kappa_n}^{l_{\kappa_n}-q}}\|\mathfrak{E}v\|_{H^{l_{\kappa_n}}(\mathcal{K}_n)}, \quad l_{\kappa_n} \geq 0, \tag{6.34}$$

for $0 \leq q \leq l_{\kappa_n}$,

$$\|v - \tilde{\Pi}v\|_{L^2(\partial\kappa_n \cap \tilde{F}_t^\perp)} \leq C\frac{h_{\kappa_n}^{s_{\kappa_n}-1/2}}{p_{\kappa_n}^{l_{\kappa_n}-1/2}}\|\mathfrak{E}v\|_{H^{l_{\kappa_n}}(\mathcal{K}_n)}, \quad l_{\kappa_n} > 1/2, \tag{6.35}$$

and

$$\|v - \tilde{\Pi}v\|_{L^2(\partial\kappa_n \cap \tilde{F}_t^\parallel)} \leq C\frac{h_{\kappa_n}^{s_{\kappa_n}-1/2}}{p_{\kappa_n}^{l_{\kappa_n}-1/2}}\|\mathfrak{E}v\|_{H^{l_{\kappa_n}}(\mathcal{K}_n)}, \quad l_{\kappa_n} > 1/2, \tag{6.36}$$

with $s_{\kappa_n} = \min\{p_{\kappa_n}+1, l_{\kappa_n}\}$, and $C > 0$ constant, depending on the shape-regularity of $\mathcal{K}_n$, but independent of v, h_{κ_n}, p_{κ_n} and the number of faces per element.

Proof. The bound (6.34) can be proved in completely analogous fashion to the bounds appearing in relation (3.28) in Section 3.3. The proof of (6.35) also follows using an anisotropic version of the classical trace inequality (see, e.g., [103]) and

(6.34) for $q = 0, 1$. The proof for (6.36) follows the same proof as Lemma 4.10 in Section 4.3. Here, the constant C depends on the constant from the trace inequality, but is independent of the discretization parameters and number of faces per element. □

We first give an a priori error bound for the space-time DGFEM in the $|||\cdot|||_{\mathrm{s}}-$norm, before using this bound to prove a respective $L^2(L^2)$–norm a priori error bound.

Theorem 6.12. *Let Assumptions 4.3.1, 6.2.1 and 3.3.1 hold, and let $u_h \in S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)$ be the space-time DGFEM approximation to the exact solution $u \in L^2(J; H^1(\Omega)) \cap H^1(J; H^{-1}(\Omega))$, with the discontinuity-penalization function given by (6.18), and suppose that $u|_{\kappa_n} \in H^{l_{\kappa_n}}(\kappa_n)$, $l_{\kappa_n} \geq 1$, for each $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, such that $\mathfrak{E}u|_{\mathcal{K}_n} \in H^{l_{\kappa_n}}(\mathcal{K}_n)$. Then, the following error bound holds:*

$$|||u - u_h|||_{\mathrm{s}}^2 \leq C \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \frac{h_{\kappa_n}^{2s_{\kappa_n}}}{p_{\kappa_n}^{2l_{\kappa_n}}} \big(\mathcal{G}_{\kappa_n}(h_{\kappa_n}, p_{\kappa_n}) + \mathcal{D}_{\kappa_n}(h_{\kappa_n}, p_{\kappa_n})\big) \|\mathfrak{E}u\|_{H^{l_{\kappa_n}}(\mathcal{K}_n)}^2, \quad (6.37)$$

where

$$\mathcal{G}_{\kappa_n}(h_{\kappa_n}, p_{\kappa_n}) = \tau_{\kappa_n}^{-1} + \tau_{\kappa_n} p_{\kappa_n}^2 h_{\kappa_n}^{-2} + p_{\kappa_n} h_{\kappa_n}^{-1} + \bar{\mathbf{a}}_{\kappa_n} p_{\kappa_n}^2 h_{\kappa_n}^{-2} + p_{\kappa_n} h_{\kappa_n}^{-1} \max_{\tilde{F}_t^{\parallel} \subset \partial\kappa_n} \sigma,$$

and

$$\mathcal{D}_{\kappa_n}(h_{\kappa_n}, p_{\kappa_n}) \;=\; \bar{\mathbf{a}}_{\kappa_n}^2 \big(p_{\kappa_n}^3 h_{\kappa_n}^{-3} \max_{\tilde{F}_t^{\parallel} \subset \partial\kappa_n} \sigma^{-1} + p_{\kappa_n}^4 h_{\kappa_n}^{-3} \max_{\tilde{F}_t^{\parallel} \subset \partial\kappa_n} \sigma^{-1}\big), \quad (6.38)$$

with $s_\kappa = \min\{p_\kappa + 1, l_\kappa\}$ and $p_\kappa \geq 1$. Here, the positive constant C is independent of the discretization parameters, number of faces per element and u.

Proof. After noting that $\lambda_n \leq c_{reg} h_\kappa$ by Assumption 6.2.1, an a priori bound can be derived following a similar approach as Theorem 5.9 where an a priori bound for general second order linear problems is presented. However, we point out that here a different treatment of the trace terms to take advantages of the mesh Assumption 4.3.1 used here by employing the Lemma 6.11. □

Remark 6.13. The above a priori bound holds *without* any assumptions on the relative size of the spatial faces F, $F \subset \partial\kappa$, and number of faces of a given spatial polytopic element $\kappa \in \mathcal{T}_h$, i.e., elements with arbitrarily small faces and/or arbitrary number of faces are permitted, as long as they satisfy Assumption 4.3.1.

For later reference, we note that $\mathcal{D}_{\kappa_n}(h_{\kappa_n}, p_{\kappa_n})$ given in (6.38), estimates the inconsistency part of the error; and it is identical to the term appeared in (5.48) in Chapter 5.

Remark 6.14. The proposed method uses space-time $\mathcal{P}_p$ basis on each element κ_n, $\forall \kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$, instead of the tensor-product basis used by standard DG-time stepping schemes. Consequently, the above a priori bound (6.37) requires a space-time Sobolev regularity which is stronger than the natural regularity of the parabolic problem at hand. This extra regularity has to be assumed.

Corollary 6.15. *Assume the hypotheses of Theorem 6.37 and consider uniform elemental polynomial degrees $p_{\kappa_n} = p \geq 1$. Assume also that $h = \max_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} h_{\kappa_n}$, $s_{\kappa_n} = s$ and $s = \min\{p+1, l\}$, $l \geq 1$. Then, we have the bound*

$$\|u - u_h\|_{L^2(J;H^1(\Omega))} \leq C \frac{h^{s-1}}{p^{l-3/2}} \|u\|_{H^l(J\times\Omega)},$$

for $C > 0$ constant, independent of u, u_h, number of faces per element, and of the mesh parameters.

Proof. We begin by observing the bounds

$$\theta \|v\|^2_{L^2(J;H^1(\Omega))} \leq |||v|||^2_{\mathrm{DG}} \leq |||v|||^2_{\mathrm{s}}. \tag{6.39}$$

Theorem 6.10, together with Assumption 3.3.1, implies that

$$\|\mathfrak{E}u\|_{H^l(J\times\Omega_\sharp)} \leq C \|u\|_{H^l(J\times\Omega)},$$

and the result follows. □

The above bound is, therefore, $h-$optimal and $p-$suboptimal by $p^{1/2}$.

6.2.3 A priori error analysis in $L^2(L^2)-$norm

In this section, we derive an error bound in the $L^2(J; L^2(\Omega))-$norm using a parabolic duality argument. To this end, the backward adjoint problem of (6.1) is defined by

$$\begin{aligned} -\partial_t z - \nabla \cdot (\mathbf{a} \nabla z) &= \phi \quad \text{in } J \times \Omega, \\ z|_{t=T} &= g \quad \text{on } \Omega, \quad \text{and} \quad u = 0 \quad \text{on } J \times \partial\Omega. \end{aligned} \tag{6.40}$$

Assume that $g \in H_0^1(\Omega)$ and $\phi \in L^2(J; L^2(\Omega))$. Then we have

$$z \in L^2(J; H^2(\Omega)) \cap L^\infty(J; H_0^1(\Omega)), \quad \partial_t z \in L^2(J; L^2(\Omega)), \tag{6.41}$$

We assume that Ω is convex and $\mathbf{a}$ is smooth such that the parabolic regularity estimate

$$\begin{aligned} \|z\|_{L^\infty(J;H_0^1(\Omega))} + \|z\|_{L^2(J;H^2(\Omega))} \quad + \quad \|z\|_{H^1(J;L^2(\Omega))} \leq \\ C_r(\|\phi\|_{L^2(J;L^2(\Omega))} + \|g\|_{H_0^1(\Omega)}), \end{aligned} \tag{6.42}$$

holds with the constant $C_r > 0$ depending only on Ω, T and $\mathbf{a}$; cf. [96, p.360] for smooth domains, and the parabolic regularity results can be extended to convex domains by using results in [113, Chapter 3].

For the sake of simplicity, we make the following local bounded variation assumption.

Assumption 6.2.2. *For any two d-dimensional spatial elements κ, $\kappa' \in \mathcal{T}$ sharing the same $(d-1)-$face, we have:*

$$\max(h_\kappa, h_{\kappa'}) \leq c_h \min(h_\kappa, h_{\kappa'}), \quad \max(p_{\kappa_n}, p_{\kappa'_n}) \leq c_p \min(p_{\kappa_n}, p_{\kappa'_n}), \tag{6.43}$$

for $n = 1, \ldots, N_t$, $c_h > 0$, $c_p > 0$ constants, independent of discretization parameters.

Before deriving the main results in this section, we introduce some approximation results in the following lemma.

Lemma 6.16. *For $v \in H^1(I_n)$, $I_n \in \mathcal{U}_h$ with ∂I_n denotes the end points of the interval I_n, let π_p^t denote the L^2 orthogonal projection onto the polynomial space $\mathcal{P}_p(I_n)$, $p \geq 0$. Then the following relation holds*

$$\|v - \pi_p^t v\|_{L^2(I_n)} \leq C \frac{\lambda_n}{p+1} \|\partial_t v\|_{L^2(I_n)}, \tag{6.44}$$

and

$$\|v - \pi_p^t v\|_{L^2(\partial I_n)} \leq C (\frac{\lambda_n}{p+1})^{1/2} \|\partial_t v\|_{L^2(I_n)}. \tag{6.45}$$

Here, $\partial I_n = \{t_{n-1}, t_n\}$. We also have

$$\|v - \pi_p^t v\|_{L^2(I_n)} \leq C \lambda_n^{1/2} \|v\|_{L^\infty(I_n)}, \tag{6.46}$$

here, $C > 0$ constant is independent of v, λ_n, p.

Proof. Bounds (6.44) and (6.45) can be proved by using Legendre polynomial expansion (see, e.g. [125]). Bound (6.46) can be proved by using the stability of L^2 projector and Holder's inequality. □

Theorem 6.17. *Consider the setting of Theorem 6.12, and assume the parabolic regularity estimate* (6.42) *holds along with Assumption 6.2.2. Then, we have the bound*

$$\begin{aligned}\|u-u_h\|^2_{L^2(J;L^2(\Omega))} \;\leq\; C \max_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h} h_{\kappa_n} \sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h} \frac{h_{\kappa_n}^{2s_{\kappa_n}}}{p_{\kappa_n}^{2l_{\kappa_n}}} \big(\mathcal{G}_{\kappa_n}(h_{\kappa_n},p_{\kappa_n}) \\ +\mathcal{D}_{\kappa_n}(h_{\kappa_n},p_{\kappa_n})\big)\|\mathfrak{E}u\|^2_{H^{l_{\kappa_n}}(\mathcal{K}_n)},\end{aligned}$$

with the constant $C > 0$, independent of u, u_h, of the discretization parameters and of number of faces per element.

Proof. We set $g = 0$ and $\phi = u - u_h$ in (6.40). After integration by parts, we have,

$$\begin{aligned}\|u-u_h\|^2_{L^2(J;L^2(\Omega))} &= \sum_{n=1}^{N_t}\int_{I_n} -(\partial_t z, u-u_h) + B_{\mathrm{d}}(z,u-u_h)\,\mathrm{d}t \qquad (6.47)\\ &- \sum_{n=1}^{N_t-1}(\lfloor z\rfloor_n,(u-u_h)^-_n) + (z^-_{N_t},(u-u_h)^-_{N_t}) = B(u-u_h,z),\end{aligned}$$

with z the solution to (6.40); cf. [180]. Now, using the inconsistent formulation, we have

$$\|u-u_h\|^2_{L^2(J;L^2(\Omega))} = \tilde{B}(u-u_h,z) - R(z,u-u_h),$$

with

$$R(v,\omega) := \sum_{F\in\mathcal{F}_h}\int_J\int_F \{\!\{\mathbf{a}(\nabla v - \mathbf{\Pi}_2(\nabla v))\}\!\}\cdot[\![\omega]\!]\,\mathrm{d}s\,\mathrm{d}t.$$

Here, we point out that if $\omega \in H^1(\Omega)$, then above inconsistent term is zero. Further, for any $z_h \in S^{\mathbf{p}}(\mathcal{U}_h;\mathcal{T}_h)$, we have

$$\tilde{B}(u-u_h,z_h) \;=\; \tilde{B}(u-u_h,z_h) - B(u-u_h,z_h) = R(u,z_h),$$

and also $R(u,z_h) = -R(u,z-z_h)$ since $R(u,z) = 0$ by relation (6.42). The above imply that

$$\|u-u_h\|^2_{L^2(J;L^2(\Omega))} = \tilde{B}(u-u_h,z-z_h) - R(z,u-u_h) - R(u,z-z_h). \qquad (6.48)$$

For brevity, we set $e := u - u_h$ and $\eta := z - z_h$. Let $z_h \in S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)$ defined on each element $\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h$ by

$$z_h|_{\kappa_n} := \begin{cases} \pi_{\bar{p}}^t \tilde{\Pi}_{\bar{p}} z & \text{, for } p_{\kappa_n} \text{ even;} \\ \pi_{\bar{p}}^t \tilde{\Pi}_{\bar{p}+1} z & \text{, for } p_{\kappa_n} \text{ odd,} \end{cases}$$

for $\bar{p} := \lfloor \frac{p_{\kappa_n}}{2} \rfloor$, with π_q^t denoting the L^2-orthogonal projection onto polynomials of degree q with respect to the time variable defined in Lemma 6.16, and $\tilde{\Pi}_q$ is the projector defined in Lemma 3.14 over d-dimensional spatial variables. Note that this choice ensures that $z_h \in S^{\mathbf{p}}(\mathcal{U}_h; \mathcal{T}_h)$.

For the first term on the right-hand side of (6.48), using (6.20) together with the Cauchy-Schwarz inequality we have

$$\begin{aligned} \tilde{B}(e,\eta) &= \sum_{n=1}^{N_t} \int_{I_n} (\partial_t e, \eta) + \tilde{B}_{\mathrm{d}}(e,\eta)\,\mathrm{d}t + \sum_{n=1}^{N_t-1} (\lfloor e \rfloor_n, \eta_n^+) + (e_0^+, \eta_0^+) \\ &\leq \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \|\tau_{\kappa_n}^{1/2} \partial_t e\|_{L^2(\kappa_n)} \|\tau_{\kappa_n}^{-1/2} \eta\|_{L^2(\kappa_n)} + \sum_{n=1}^{N_t} \int_{I_n} C_{\mathrm{cont}} |||e|||_{\mathrm{d}} |||\eta|||_{\mathrm{d}}\,\mathrm{d}t \\ &+ \sum_{n=1}^{N_t-1} \|\lfloor e \rfloor_n\| \|\eta_n^+\| + \|e_0^+\| \|\eta_0^+\| \\ &\leq \Big(\sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n}^{-1} \|\eta\|_{L^2(\kappa_n)}^2 + (C_{\mathrm{cont}})^2 \sum_{n=1}^{N_t} \int_{I_n} |||\eta|||_{\mathrm{d}}^2\,\mathrm{d}t \\ &+ 2 \sum_{n=0}^{N_t-1} \|\eta_n^+\|^2 \Big)^{\frac{1}{2}} |||e|||_{\mathrm{s}}. \end{aligned} \tag{6.49}$$

We shall now estimate the terms involving η on the right-hand side of (6.49). Recalling standard *hp*-approximation bounds in Lemma 6.16, we have for $r \in \{\bar{p}, \bar{p}+1\}$,

$$\begin{aligned} \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n}^{-1} \|\eta\|_{L^2(\kappa_n)}^2 &= \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n}^{-1} \|z - \pi_{\bar{p}}^t \tilde{\Pi}_{\bar{p}} z\|_{L^2(\kappa_n)}^2 \\ &\leq 2 \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n}^{-1} \Big(\|z - \pi_{\bar{p}}^t z\|_{L^2(\kappa_n)}^2 + \|\pi_{\bar{p}}^t z - \pi_{\bar{p}}^t \tilde{\Pi}_r z\|_{L^2(\kappa_n)}^2 \Big) \\ &\leq C \sum_{\kappa_n \in \mathcal{U}_h \times \mathcal{T}_h} \tau_{\kappa_n}^{-1} \Big(\frac{\lambda_n^2}{p_{\kappa_n}^2} \|\partial_t z\|_{L^2(\kappa_n)}^2 + \frac{h_\kappa^4}{p_{\kappa_n}^4} \|\mathfrak{E} z\|_{L^2(I_n; H^2(\mathcal{K}))}^2 \Big) \\ &\leq C \max_{\kappa_n} h_{\kappa_n} \Big(\|z\|_{H^1(J;L^2(\Omega))}^2 + \max_{\kappa_n} \frac{h_{\kappa_n}^2}{p_{\kappa_n}^2} \|z\|_{L^2(J;H^2(\Omega))}^2 \Big), \end{aligned} \tag{6.50}$$

using the triangle inequality, the stability of L^2-projection, Assumptions 6.2.1, 6.2.2, Lemma 6.11, and Theorem 6.10, respectively. Next, we have

$$\begin{aligned}
\sum_{n=0}^{N_t-1}\|\eta_n^+\|^2 &\leq 2\sum_{n=0}^{N_t-1}\sum_{\kappa\in\mathcal{T}_h}\Big(\|(z-\pi_{\bar{p}}^t z)_n^+\|^2_{L^2(\kappa)}+\|(\pi_{\bar{p}}^t z-\pi_{\bar{p}}^t\tilde{\Pi}_r z)_n^+\|^2_{L^2(\kappa)}\Big)\\
&\leq C\sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}\Big(\frac{\lambda_n}{p_{\kappa_n}}\|\partial_t z\|^2_{L^2(\kappa_n)}+\frac{p^2_{\kappa_n}}{\lambda_n}\|\pi_{\bar{p}}^t(z-\tilde{\Pi}_r z)\|^2_{L^2(\kappa_n)}\Big)\\
&\leq C\sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}\Big(\frac{\lambda_n}{p_{\kappa_n}}\|\partial_t z\|^2_{L^2(\kappa_n)}+\frac{h^4_{\kappa}}{\lambda_n p^2_{\kappa_n}}\|\mathfrak{E}z\|^2_{L^2(J;H^2(\mathcal{K}))}\Big)\\
&\leq C\max_{\kappa_n}\frac{h_{\kappa_n}}{p_{\kappa_n}}\Big(\|z\|^2_{H^1(J;L^2(\Omega))}+\max_{\kappa_n}\frac{h^2_{\kappa_n}}{p_{\kappa_n}}\|z\|^2_{L^2(J;H^2(\Omega))}\Big), \qquad (6.51)
\end{aligned}$$

using an *hp*-version inverse estimate over time variable and working as before. Next, we have

$$\begin{aligned}
\int_J\sum_{\kappa\in\mathcal{T}_h}\|\nabla\eta\|^2_{L^2(\kappa)}\,\mathrm{d}t &= \sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}\|\nabla(z-\pi_{\bar{p}}^t\tilde{\Pi}_{\bar{p}}^s z)\|^2_{L^2(\kappa_n)}\\
&\leq \sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}2\Big(\|\nabla(z-\pi_{\bar{p}}^t z)\|^2_{L^2(\kappa_n)}+\|\nabla(\pi_{\bar{p}}^t z-\pi_{\bar{p}}^t\tilde{\Pi}_r z)\|^2_{L^2(\kappa_n)}\Big)\\
&\leq C\sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}\Big(\lambda_n\|\nabla z\|^2_{L^\infty(I_n;L^2(\kappa))}+\frac{h^2_{\kappa}}{p^2_{\kappa_n}}\|\mathfrak{E}z\|^2_{L^2(I_n;H^2(\mathcal{K}))}\Big)\\
&\leq C\max_{\kappa_n}h_{\kappa_n}\Big(\|z\|^2_{L^\infty(J;H^1_0(\Omega))}+\max_{\kappa_n}\frac{h_{\kappa_n}}{p^2_{\kappa_n}}\|z\|^2_{L^2(J;H^2(\Omega))}\Big). \qquad (6.52)
\end{aligned}$$

Using similar arguments as before. Also, since $[\![z]\!]=0=[\![\pi_{\bar{p}}^t z]\!]$, due to relation (6.42), we have $[\![z-\pi_{\bar{p}}^t\tilde{\Pi}_r z]\!]=\pi_{\bar{p}}^t[\![z-\tilde{\Pi}_r z]\!]$, thus,

$$\begin{aligned}
\sum_{F\in\mathcal{F}_h}\int_J\int_F\sigma|[\![\eta]\!]|^2\,\mathrm{d}s\,\mathrm{d}t &= \sum_{\tilde{F}_t^{\|}\subset J\times\mathcal{F}_h}\sigma\|[\![z-\pi_{\bar{p}}^t\tilde{\Pi}_{\bar{p}}^s z]\!]\|^2_{L^2(\tilde{F}_t^{\|})}\\
&\leq 2\sum_{\tilde{F}_t^{\|}\subset J\times\mathcal{F}_h}\sigma\|[\![z-\tilde{\Pi}_r z]\!]\|^2_{L^2(\tilde{F}_t^{\|})}\\
&\leq C\sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h}(\max_{\tilde{F}_t^{\|}\subset\partial\kappa_n}\sigma)\frac{h^3_{\kappa_n}}{p^3_{\kappa_n}}\|\mathfrak{E}z\|^2_{L^2(J;H^2(\mathcal{K}))}\\
&\leq C\max_{\kappa_n}\frac{h^2_{\kappa_n}}{p_{\kappa_n}}\|z\|^2_{L^2(J;H^2(\Omega))}, \qquad (6.53)
\end{aligned}$$

by Assumption 6.2.2.

Substituting (6.50), (6.51), (6.52), (6.53) into (6.49), along with (6.42), leads to

$$\tilde{B}(e,\eta) \leq CC_r \max_{\kappa_n} h_{\kappa_n}^{1/2} |||e|||_{\mathrm{s}} \|e\|_{L^2(J;L^2(\Omega))}. \tag{6.54}$$

Moving on to the second term on the right-hand side of (6.48), we have

$$\begin{aligned} R(z,e) &= \sum_{F\in\mathcal{F}_h} \int_J \int_F \{\!\{\mathbf{a}(\nabla z - \boldsymbol{\Pi}_2(\nabla z))\}\!\} \cdot [\![e]\!] \,\mathrm{d}s\,\mathrm{d}t \\ &\leq \Big(\sum_{F\in\mathcal{F}_h} \int_J \int_F \sigma^{-1} |\{\!\{\mathbf{a}(\nabla z - \boldsymbol{\Pi}_2(\nabla z))\}\!\}|^2 \,\mathrm{d}s\,\mathrm{d}t \Big)^{\frac{1}{2}} |||e|||_{\mathrm{s}}. \end{aligned}$$

To bound further $R(z,e)$, it is sufficient to bound $\mathrm{I} + \mathrm{II}$ instead, where

$$\begin{aligned} \mathrm{I} &:= \sum_{F\in\mathcal{F}_h} \int_J \int_F 2\sigma^{-1} |\{\!\{\mathbf{a}(\nabla z - \boldsymbol{\pi}_{\bar{\boldsymbol{p}}}^{\boldsymbol{t}} \tilde{\boldsymbol{\Pi}}_{\boldsymbol{r}}(\nabla z))\}\!\}|^2 \,\mathrm{d}s\,\mathrm{d}t, \\ \mathrm{II} &:= \sum_{F\in\mathcal{F}_h} \int_J \int_F 2\sigma^{-1} |\{\!\{\mathbf{a}\boldsymbol{\Pi}_2(\boldsymbol{\pi}_{\bar{\boldsymbol{p}}}^{\boldsymbol{t}} \tilde{\boldsymbol{\Pi}}_{\boldsymbol{r}}(\nabla z) - \nabla z)\}\!\}|^2 \,\mathrm{d}s\,\mathrm{d}t. \end{aligned}$$

Here, $\boldsymbol{\pi}_{\bar{\boldsymbol{p}}}^{\boldsymbol{t}} \tilde{\boldsymbol{\Pi}}_{\boldsymbol{r}}$ denotes the vector valued projector $\pi_{\bar{p}}^t \tilde{\Pi}_r$. To bound the term I, using Lemma 6.11 and working as before gives

$$\mathrm{I} \leq C \max_{\kappa_n} \frac{h_{\kappa_n}^{3/2}}{p_{\kappa_n}^2} \Big(\|z\|_{L^\infty(J;H_0^1(\Omega))}^2 + \|z\|_{L^2(J;H^2(\Omega)}^2 \Big). \tag{6.55}$$

By using the inverse estimation Lemma 4.9 and stability of $\boldsymbol{\Pi}_2$, and working as above, we also have

$$\mathrm{II} \leq C \max_{\kappa_n} h_{\kappa_n} \Big(\|z\|_{L^\infty(J;H_0^1(\Omega))}^2 + \|z\|_{L^2(J;H^2(\Omega)}^2 \Big). \tag{6.56}$$

Therefore, (6.55) and (6.56), together with (6.42) give

$$R(z,e) \leq CC_r \max_{\kappa_n} h_{\kappa_n}^{1/2} |||e|||_{\mathrm{s}} \|e\|_{L^2(J;L^2(\Omega))}. \tag{6.57}$$

Next, we bound the last term on the right-hand side of (6.48), which is given by

$$R(u,\eta) = \sum_{n=1}^{N_t} \int_{I_n} \int_\Gamma \{\!\{\mathbf{a}(\nabla_h u - \boldsymbol{\Pi}_2(\nabla_h u))\}\!\} \cdot [\![\eta]\!] \,\mathrm{d}s\,\mathrm{d}t.$$

The Cauchy-Schwarz inequality together with (6.53), result in

$$\begin{aligned} R(u,\eta) &\leq \Big(\sum_{F\in\mathcal{F}_h} \int_J \int_F \sigma^{-1} |\{\!\{ \mathbf{a}(\nabla_h u - \mathbf{\Pi}_2(\nabla_h u)) \}\!\}|^2 \,\mathrm{d}s\,\mathrm{d}t \Big)^{\frac{1}{2}} \\ &\quad \times \Big(\sum_{F\in\mathcal{F}_h} \int_J \int_F \sigma [\![\eta]\!]^2 \,\mathrm{d}s\,\mathrm{d}t \Big)^{\frac{1}{2}} \\ &\leq CC_r \max_{\kappa_n} \frac{h_{\kappa_n}}{p_{\kappa_n}^{1/2}} \|e\|_{L^2(J;L^2(\Omega))} \\ &\quad \times \Big(\sum_{\kappa_n\in\mathcal{U}_h\times\mathcal{T}_h} \frac{h_{\kappa_n}^{2s_{\kappa_n}}}{p_{\kappa_n}^{2l_{\kappa_n}}} \mathcal{D}_{\kappa_n}(h_{\kappa_n},p_{\kappa_n}) \|\mathfrak{E}u\|^2_{H^{l_{\kappa_n}}(\mathcal{K}_n)} \Big)^{1/2}. \end{aligned} \tag{6.58}$$

Here, $\mathcal{D}_{\kappa_n}(h_{\kappa_n}, p_{\kappa_n})$ is defined in (6.38), which measures the inconsistency error. Finally, combining (6.54), (6.57) and (6.58) with (6.48), the result follows. □

Remark 6.18. If we use the same assumptions as in Corollary 6.15, then we can see that the $L^2(J;L^2(\Omega))$–norm error bound in Theorem 6.17 can be simplified to

$$\|u-u_h\|_{L^2(J;L^2(\Omega))} \leq C \frac{h^{s-1/2}}{p^{l-3/2}} \|u\|_{H^l(J\times\Omega)},$$

with $s = \min\{p+1, l\}$, which is suboptimal with respect to the meshsize h by half an order of h, and sub-optimal in p by 3/2 orders. (The respective space-time tensor-product basis DGFEMs, using the same approach can be shown to be $h-$optimal and $p-$suboptimal by one order of p.) The numerical experiments in the next section confirm the suboptimality in h for the proposed method, but at the same time highlight its competitiveness with respect to standard (optimal) methods.

An interesting further development would be the use of different polynomial degrees in space and in time as done, e.g., in [173, 182] in the context of *total degree* space-time basis. The exploration of a number of index sets for space-time polynomial basis, including this case, will be discussed elsewhere. Nevertheless, the above proof of the $L^2(J;L^2(\Omega))$–norm error bound would carry through with minor modifications only for various choices of space-time basis function index sets.

6.3 Numerical examples

We shall present a series of numerical experiments to investigate the asymptotic convergence behavior of the proposed space-time DGFEMs. We shall also make

comparisons with known methods on space-time hexahedral meshes, such as the tensor-product space-time DGFEM and the DG time-stepping scheme combined with conforming finite elements in space. Furthermore, an implementation using prismatic space-time meshes with polygonal bases is presented and its convergence is assessed. In all experiments we choose $C_\sigma = 10$.

6.3.1 Example 1

We begin by considering a smooth problem for which u_0 and f are chosen such that the exact solution u of (6.1) is given by:

$$u(x, y, t) = \sin(20\pi t)e^{-5((x-0.5)^2+(y-0.5)^2)} \quad \text{in } J \times \Omega, \tag{6.59}$$

for $J = (0, 1)$ and $\Omega = (0, 1)^2$, and $\mathbf{a}(x, y, t)$ is an identity matrix. Notice that the solution oscillates in time. To asses the convergence rate with respect to the space-time mesh diameter h_{κ_n} on (quasi)uniform meshes, we fix the ratio between the spatial and temporal mesh sizes to be $h_{\kappa_n}/\lambda_n = 10$.

The convergence rate with respect to decreasing space-time mesh size h_{κ_n} in three different norms is given in Figure 6.3 for space-time prismatic elements with rectangular bases (standard hexahedral space-time elements) and for prismatic meshes with quasi-uniform polygonal bases: all computations are performed over 16, 64, 256, 1024, 4096 spatial rectangular or polygonal elements and for 40, 80, 160, 320, 640 time-steps, respectively.

The left three plots in Figure 6.3, show the rate of convergence for the proposed DGFEM using the $\mathcal{P}_p$ basis, for $p = 1, 2, \ldots, 6$, on each 3-dimensional space-time element, against the total space-time degrees of freedom (Dof). This will be referred to as 'DG(P)' for short, with 'rect' meaning spatial rectangular elements and 'poly' referring to general polygonal spatial elements in the legends. The observed rates of convergence are also given in the legends. The error appears to decay at essentially the same rate for both rectangular and polygonal spatial meshes, with very similar constants. Indeed, the DG(P) scheme appears to converge at an optimal rate $\mathcal{O}(h^p)$ in the $L^2(J; H^1(\Omega))$–norm for $p = 1, 2, \ldots, 6$ (cf. Corollary 6.15), while the convergence appears to be slightly sub-optimal, $\mathcal{O}(h^{p+1/2})$, in the $L^2(J; L^2(\Omega))$- and $L^\infty(J; L^2(\Omega))$–norms. Again, the observed $L^2(J; L^2(\Omega))$–norm convergence rate is in accordance with the theory, cf. the a priori bound of Theorem 6.17.

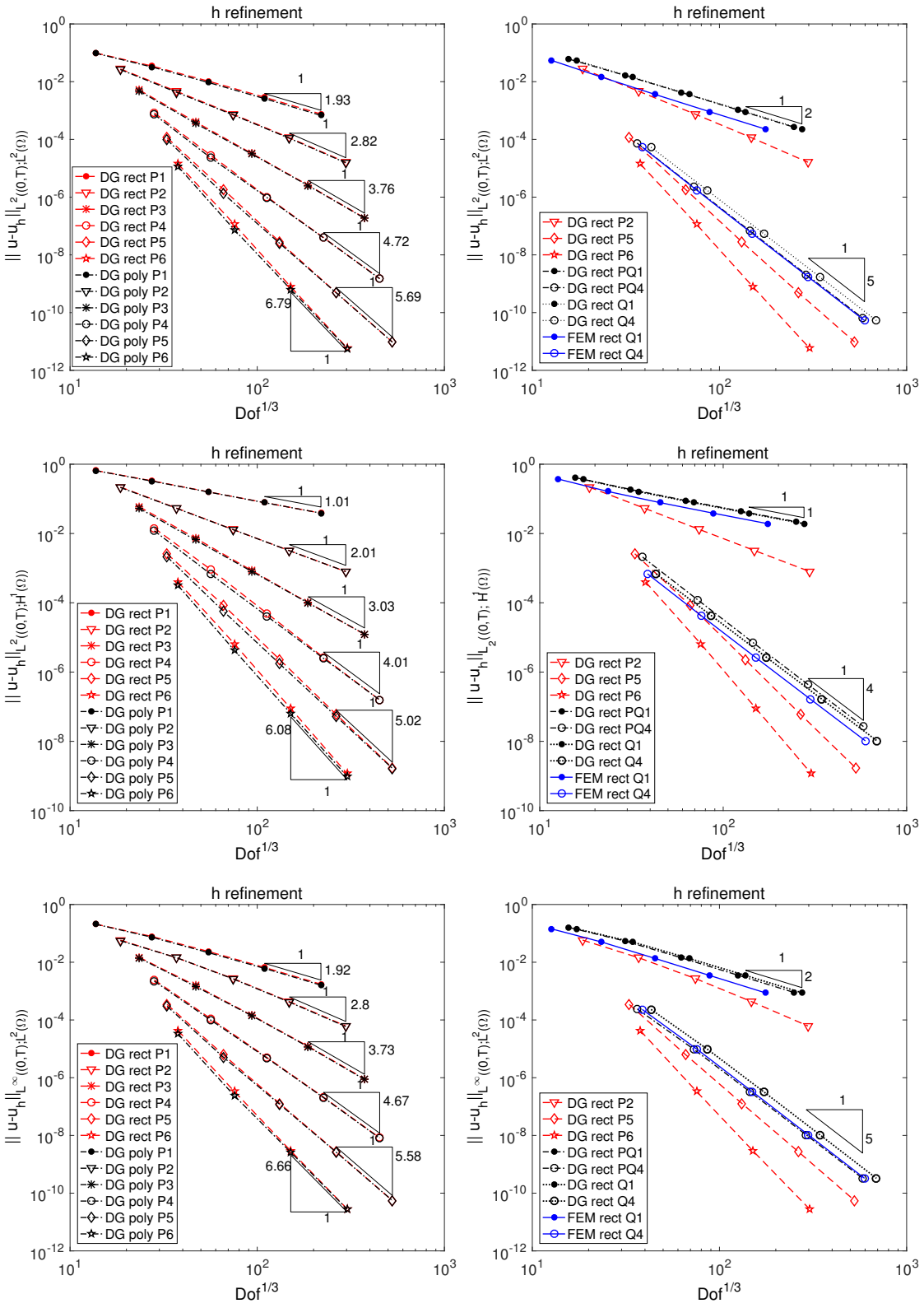


FIGURE 6.3: Example 1. DG(P) under h–refinement (left) and comparison with other methods (right) for three different norms.

We now assess whether the deterioration in the h-convergence rates is an acceptable trade-off for the DG(P) method. We present a comparison between 4 different space-time schemes over rectangular space-time meshes in the right plots of Figure 6.3. More specifically, we compare the proposed DG(P) method, against the time-DGFEM with: 1) discontinuous tensor-product space-time bases consisting of $\mathcal{P}_p$-basis in space ('DG(PQ)' for short), 2) full discontinuous tensor-product $\mathcal{Q}_p$ basis in space ('DG(Q)' for short) and, 3) the standard finite element method with conforming tensor-product $\mathcal{Q}_p$ basis in space ('FEM(Q)' for short) [180, 159]. Unlike the proposed DG(P) scheme, the three other methods achieve the optimal h-convergence rate in the three different norms: $\mathcal{O}(h^{p+1})$ in $L^2(J;L^2(\Omega))$– and $L^\infty(J;L^2(\Omega))$–norms and $\mathcal{O}(h^p)$ in $L^2(J;H^1(\Omega))$–norm, respectively. Nevertheless, plotting the error against the total degrees of freedom, a more relevant measure of computational effort, we see, for instance, that DG(P) with $p=2$ use less Dofs compared to the other 3 methods with $p=1$, to achieve the same level of accuracy, at least for relatively large number of space time elements. More pronounced gains are observed when comparing DG(P) with $p=5,6$ with the other methods with $p=4$, across all mesh sizes and error norms. Analogous results hold for DG(P) with $p=3,4$.

Moving on to the p-version, Figure 6.4 shows the error for all four methods in the three different norms for fixed space-time meshsize under p-refinement. The left three plots are with final time $T=1$, for fixed 64 spatial elements and 80 time steps. As expected, exponential convergence is observed since the solution to (6.59) is analytic over the computational domain. However, the convergence slope for DG(P) with both rectangular and polygonal spatial elements appears to be steeper than the other 3 methods. Indeed, DG(P) achieves the same level of accuracy for $p \geq 3$ with less number of Dofs in all 3 different norms.

The right three plots for the same computation run for a longer time interval with final time $T=40$, that is 3200 time-steps. Since DG(P) use less Dofs per space-time element compared to the other three methods, the acceleration of $p-$convergence for the DG(P) is expected to be more pronounced for long time computations. Again DG(P) achieves the same level of accuracy with fewer degrees of freedom for $p \geq 3$. For instance, the total DG(P) Dofs for this problem are about 45 million when $p=9$, compared to about 53 million Dofs with $p=6$ for FEM(Q), while the error for DG(P) is about 100 times smaller than the error of FEM(Q) in all three norms.

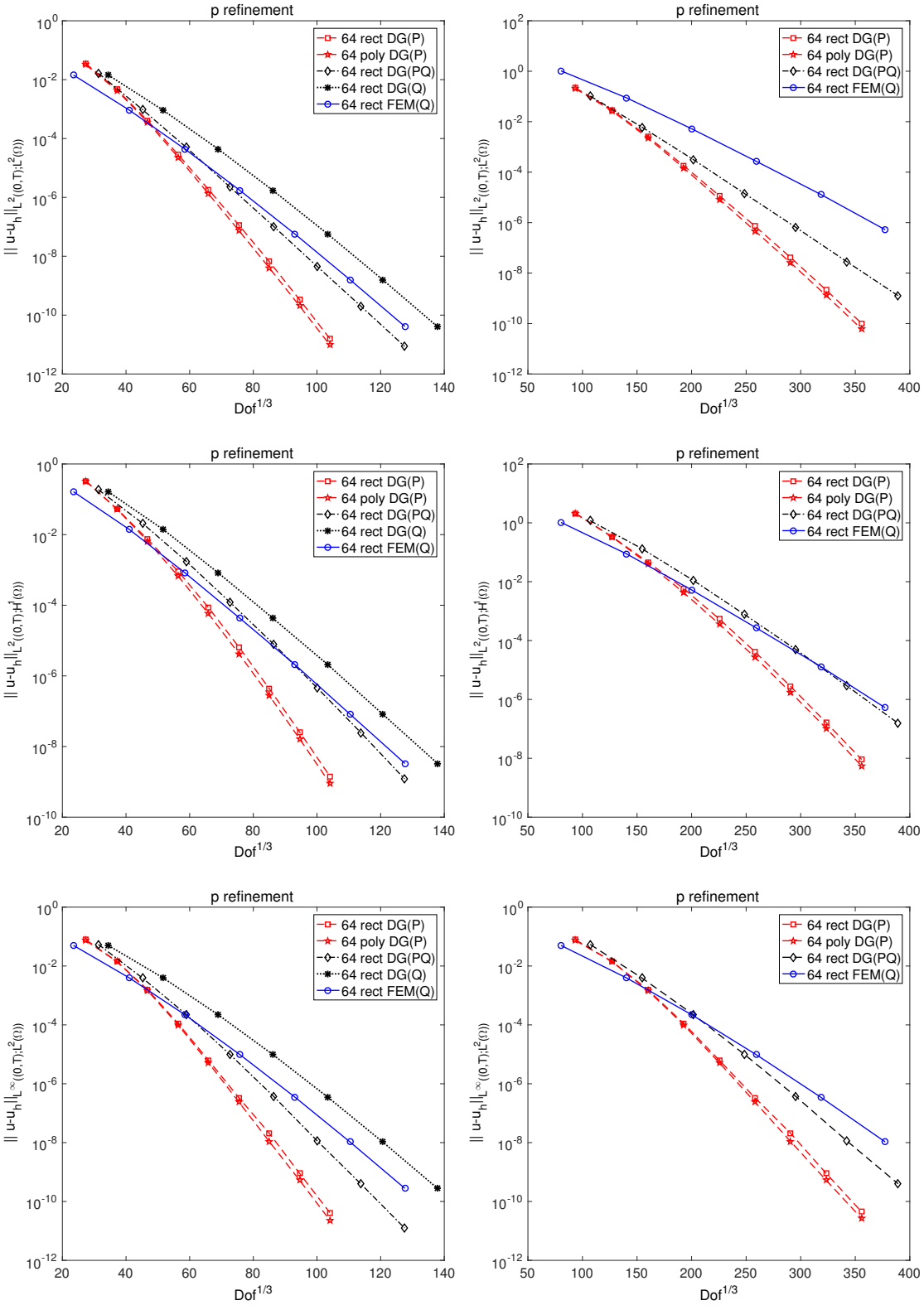


FIGURE 6.4: Example 1. Convergence under p–refinement for $T = 1$ with 80 time steps (left) and for $T = 40$ with 3200 time steps (right) for three different norms.

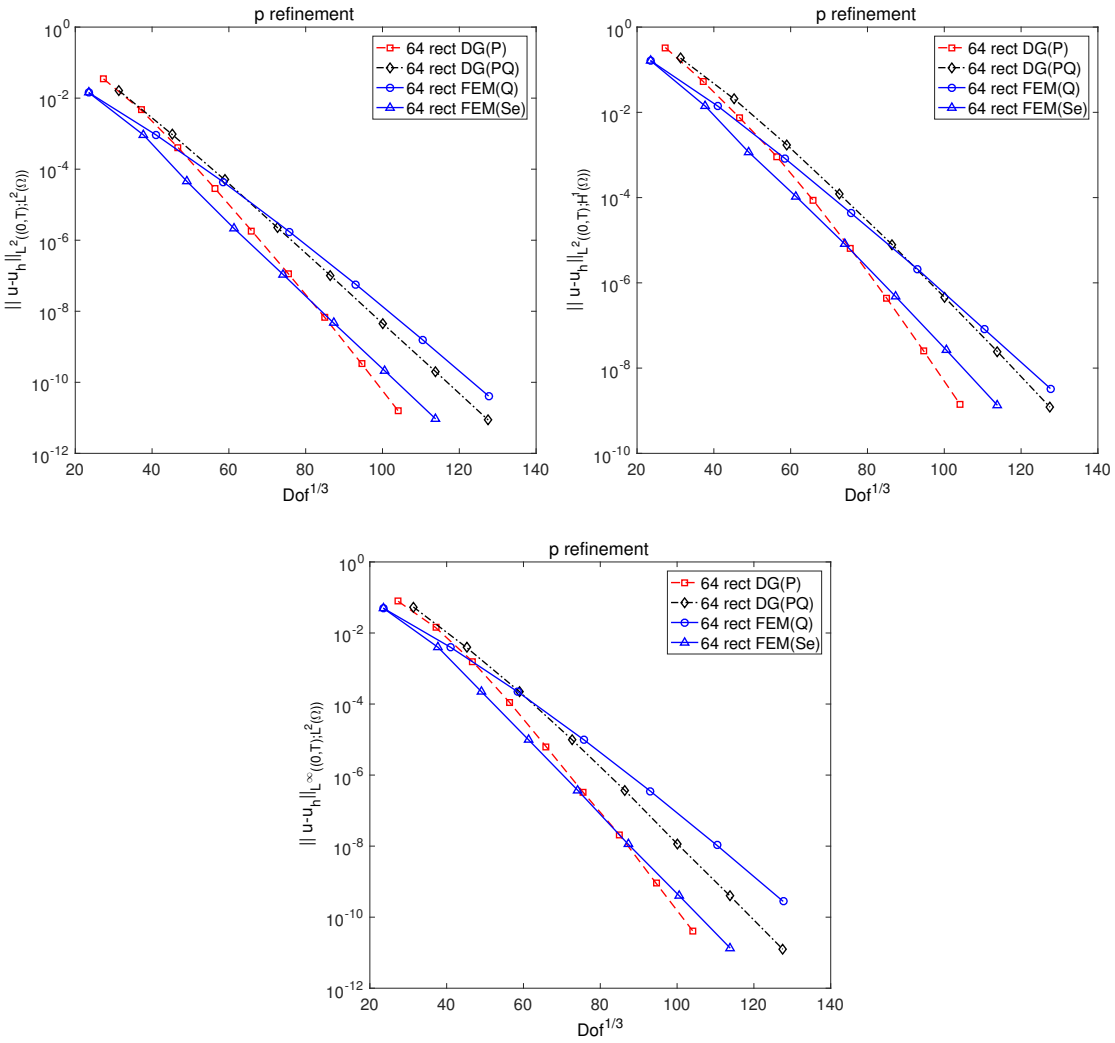


FIGURE 6.5: Example 1. Convergence under p–refinement for $T = 1$ with 80 time steps for three different norms.

Finally, we investigate the convergence performance of the proposed approach against DG time-stepping spatially conforming FEM with the cheaper conforming serendipity elements in space on hexahedral space-time meshes. Numerical results under p-refinement are given in Figure 6.5, with FEM(Se) standing for the latter method. We note that for $d = 2$, the cardinality of the local serendipity space equals the cardinality of $\mathcal{P}_p$–basis plus two more Dofs. We observe that the convergence slope of FEM(Se) is steeper than that of FEM(Q) and almost parallel to DG(PQ), but it is still not steeper than the convergence slope of DG(P). We observe that DG(P) with $p = 7$ gives smaller error against Dofs than FEM(Se) with $p = 6$. Noting that serendipity basis in three dimensions uses considerably more Dofs compared to total degree $\mathcal{P}_p$-basis, it is expected that DG(P) will achieve smaller error for the same Dofs than FEM(Se) with lower order that 7

polynomials for $d = 3$.

6.3.2 Example 2

We shall now assess the performance of the *hp*-version of the proposed method for a problem with an initial layer. Let $\mathbf{a}(x, y, t)$ to be the identity matrix, and u_0 and f chosen so that the exact solution of (6.1) is given by

$$u(x, y, t) = t^{\alpha} \sin(\pi x) \sin(\pi y) \quad \text{in } J \times \Omega, \tag{6.60}$$

with $J = (0, 0.1)$ and $\Omega = (0, 1)^2$. We set $\alpha = 1/2$, so that $u \in H^{1-\epsilon}(J; L^2(\Omega))$, for all $\epsilon > 0$. This problem is analytic over the spatial domain, but has low regularity at $t = 0$. To achieve exponential rates of convergence, we use temporal meshes, geometrically graded towards $t = 0$, in conjunction with temporally varying polynomial degree p, starting from $p = 1$ on the elements belonging to the initial time slab, and linearly increasing p when moving away from $t = 0$; see [167, 159] for details. Following [159], we consider a short time interval with $T = 0.1$. Let $0 < \sigma < 1$ be the mesh grading factor which defines a class of temporal meshes $t_n = \sigma^{N-n} \times 0.1$ for $n = 1, \ldots, N$. Let also μ be the polynomial order increasing factor determining the polynomial order over different time steps by $p_{\kappa_n} := \lfloor \mu n \rfloor$ for for $n = 1, \ldots, N$.

The three left plots in Figure 6.6 show the convergence history for DG(P) and FEM(Q) for this problem. All computations are performed over 256 spatial elements with geometrically graded temporal meshes based on 3 different grading factors $\sigma = 0.1, 0.172, 0.5$ and fixed $\mu = 1.5$. The error for both DG(P) and FEM(Q) appears to decay exponentially under the *hp* refinement strategy described above for all three grading factors considered. The choice of $\sigma = 0.5$, is motivated by the meshes constructed in standard adaptive algorithms; $\sigma = 0.172$, is classical in that it was shown that it is the optimal grading factor for one-dimensional functions with r^{α}-type singularity for elliptic problem in [114], while $\sigma = 0.1$ appears to be a better choice in the current context. We also note that the convergence rate of DG(P) appears to be steeper than FEM(Q) under the same mesh and polynomial distribution. Furthermore, performing the same experiments on general polygonal spatial meshes, we observe that the error decay does not appear to depend on the shape of the spatial elements. This is expected, as the error in the time variable dominates in this example.

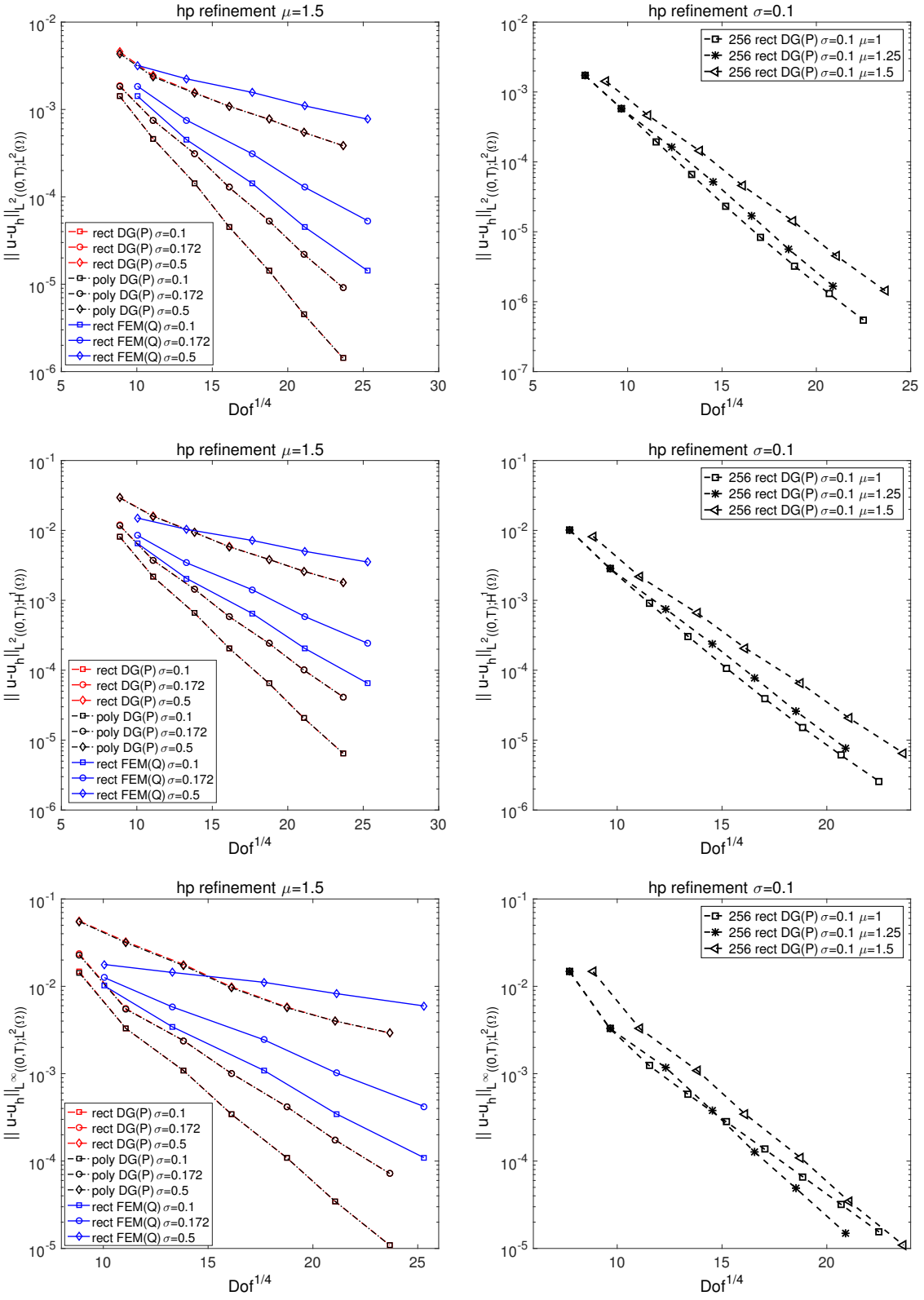


FIGURE 6.6: Example 2: Convergence under *hp*–refinement with fixed $\mu = 1.5$ (left); with fixed $\sigma = 0.1$ (right) for three different norms.

For completeness, we also report on how the choice of the polynomial order increasing factor μ influences the exponential error decay for DG(P) with fixed mesh grading factor $\sigma = 0.1$; these are given in the three right plots in Figure 6.6. For both $L^2(J; L^2(\Omega))$– and $L^2(J; H^1(\Omega))$–norms, the results show that $\mu = 1$ gives the fastest convergence, while $\mu = 1.25$ gives the fastest error decay in the $L^\infty(J; L^2(\Omega))$–norm.

Chapter 7

Exponential Convergence for DGFEMs with $\mathcal{P}_p$ basis

We will present some *hp*-approximation results for the total degree $\mathcal{P}_p$ basis on standard tensor product elements. The new results can be viewed as a natural extension of the classical *hp*-approximation results with the tensor product $\mathcal{Q}_p$ basis on tensor product elements. Here, we will focus on deriving an optimal *hp*-approximation bound for the L^2-orthogonal projector onto the $\mathcal{P}_p$ basis in the L^2–norm, and optimal *hp*-approximation bounds for H^1-projector onto the $\mathcal{S}_p$ basis in the L^2– and H^1–norms. The technique for proving these bounds will be different from the existing techniques for *hp*-approximation with $\mathcal{Q}_p$ basis. The main difficulty is due to the lack of tensor product structure in the $\mathcal{P}_p$ basis and the $\mathcal{S}_p$ basis, thereby hindering the use of tensor product arguments together with 1D stability and approximation results. The main technique used below is the multi-dimensional orthogonal polynomial expansion. The resulting bounds are *hp*-optimal with respect to both Sobolev regularity and polynomial approximation order.

Here, we mention that there are at least two reasons why we need new approximation results with the $\mathcal{P}_p$ and $\mathcal{S}_p$ bases: the first reason is to explain the findings of the numerical experiments in the previous chapters, where we observed that the error compared against number of degrees of freedom for DGFEMs with the $\mathcal{P}_p$ basis has a steeper exponential convergence than for DGFEMs with the $\mathcal{Q}_p$ basis, for sufficiently smooth problems. This situation has been numerically tested on different examples. We also observed that the ratio of the slope of the exponential error decay for the $\mathcal{P}_p$ basis compared to that of the $\mathcal{Q}_p$ basis depends only

on the space dimension. A natural intuition to explain this is that the $\mathcal{Q}_p$ basis contains in a sense "too many" basis functions other than those of $\mathcal{P}_p$. These basis functions do not increase the order in p of the error bound, but instead only reduce the "constant" in the error bound. The same phenomenon is also observed in standard FEM with the $\mathcal{S}_p$ basis.

The second reason is of a theoretical nature. The exponential convergence proofs depend on the *hp*-approximation bound for the L^2-orthogonal projector and the H^1-projector over tensor product elements. In general, the classical *hp*-approximation results are only proved for $\mathcal{Q}_p$ by using the tensor product arguments with the 1D stability and approximation results. The resulting bound is sharp in the sense that it is optimal in both h and p. Typically, bounds for projectors onto $\mathcal{P}_p$ or $\mathcal{S}_p$ are proved using the fact that there exists a $q \leq p$ such that $\mathcal{Q}_q$ is a subspace of $\mathcal{P}_p$ or $\mathcal{S}_p$, together with the help of the approximation results for the $\mathcal{Q}_p$ basis. We emphasise that *by using this technique, the resulting hp-approximation bound is p-optimal for functions with finite Sobolev regularity, but not p-optimal for analytic functions.* So for the above two reasons we derive the new approximation results for projectors onto $\mathcal{P}_p$ and $\mathcal{S}_p$.

We note that the *hp*-approximation results used in the previous chapters can not be used to prove exponential convergence. The key reason is because the proof of the *hp*-bound in previous chapters is based on Babuška & Suri operator in Lemma 3.11, which is the classical tools in *hp*-FEMs [24, 25]. Although the Babuška & Suri operator is a novel tool in *hp*-approximation due to the fact that it is simultaneously optimal in h and p in all Sobolev norms with finite Sobolev indices, it seems not to be useful in proving the exponential convergence of the p-version of the FEM for sufficiently smooth solutions. There are two reasons for this: first, the constant $C_{\mathrm{I},1}$ in Lemma 3.11 blows up as $l \to \infty$, which means we can not take Sobolev indices to infinity; second, even in cases where $C_{\mathrm{I},1}$ is uniformly bounded with respect to l, we can only prove spectral convergence but not exponential convergence.

7.1 Polynomial approximation over tensor product elements with $\mathcal{P}_p$ basis and $\mathcal{S}_p$ basis

In this section, we derive the *hp*-approximation results for the L^2- and H^1-projectors over tensor product elements with the $\mathcal{P}_p$ basis and $\mathcal{S}_p$ basis, respectively. We will employ the approximation results for projectors onto the $\mathcal{Q}_p$ basis from [124, 125] without giving a detailed proof. For the sake of simplicity, we only consider the tensor product elements which can be considered as an affine equivalent family of the reference element $\hat{\kappa} := (-1, 1)^d$.

7.1.1 The L^2-projection onto $\mathcal{P}_p$ over a d-dimensional cube

We start our analysis over the standard reference element $\hat{\kappa} := (-1, 1)^d$, by introducing some necessary notation. We shall employ the multi-index $i = (i_1, i_2, \ldots, i_d)$, and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$. With $|\cdot|$ we denote the l_1–norm of the multi-index i, with $|i| = \sum_{j=1}^{d} |i_k|$. Further, for multi-indices, the relation $\alpha \geq i$ means that $\alpha_k \geq i_k$ for all $k = 1, \ldots, d$.

For the reference element $\hat{\kappa} := (-1, 1)^d$, let

$$W^{\alpha}(\hat{x}) = \prod_{k=1}^{d} W_k(\hat{x}_k)^{\alpha_k}, \tag{7.1}$$

with, for $k = 1, \ldots, d$,

$$W_k(\hat{x}_k) = (1 - \hat{x}_k^2)^{1/2}, \tag{7.2}$$

being a weight function as $\alpha_k > -1$, $\alpha_k \in \mathbb{R}$. This is referred to as the Jacobi weight.

Next, we define the Jacobi-weighted Sobolev spaces $V^l(\hat{\kappa})$ as a closure of $C^{\infty}(\hat{\kappa})$ in the norm with the Jacobi weight

$$\|u\|^2_{V^l(\hat{\kappa})} = \sum_{|\alpha|=0}^{l} \|W^{\alpha} D^{\alpha} u\|^2_{L^2(\hat{\kappa})}. \tag{7.3}$$

By $|\cdot|_{V^l(\hat{\kappa})}$ we denote the seminorm

$$|u|^2_{V^l(\hat{\kappa})} = \sum_{|\alpha|=l} \|W^{\alpha} D^{\alpha} u\|^2_{L^2(\hat{\kappa})}. \tag{7.4}$$

It is easy to see that $|u|_{V^l(\hat{\kappa})} \leq |u|_{H^l(\hat{\kappa})}$, $\forall u \in H^l(\kappa)$. The key reason to introduce the Jacobi-weighted Sobolev spaces is to deal with the loss of orthogonality suffered by orthogonal polynomials in standard Sobolev spaces; the L^2-orthogonality is preserved in Jacobi-weighted Sobolev spaces. As we shall see in the forthcoming analysis, orthogonality plays a key role in deriving optimal error bounds in the polynomial order p.

In order to distinguish the same projectors onto spaces with different polynomial bases, we use superscripts to signify the basis type: we use $\Pi_p^{\mathcal{Q}} := \Pi_p^{(1)}\Pi_p^{(2)}\ldots\Pi_p^{(d)}$ to denote the L^2-projection onto $\mathcal{Q}_p$, which can be constructed by using the tensor product arguments together with 1D L^2-projection. On the other hand, L^2-projector onto $\mathcal{P}_p$ is denoted by $\Pi_p^{\mathcal{P}}$.

First, we take the the following approximation lemma for the L^2-projection $\Pi_p^{\mathcal{Q}}$ from [125].

Lemma 7.1. *Let $\hat{\kappa} = (-1,1)^d$. Suppose that $u|_{\hat{\kappa}} \in H^l(\hat{\kappa})$, for some $l \geq 0$. Let $\Pi_p^{\mathcal{Q}}u$ be the L^2-projection of u onto $\mathcal{Q}_p(\hat{\kappa})$ with $p \geq 0$. Then, for any integer s, with $0 \leq s \leq \min\{p+1, l\}$, and $W_k = W_k(\hat{x}_k)$, we have:*

$$\begin{aligned}\|u - \Pi_p^{\mathcal{Q}}u\|_{L^2(\hat{\kappa})} &\leq \Big(\frac{\Gamma(p-s+2)}{\Gamma(p+s+2)}\Big)^{1/2}\sum_{k=1}^{d}\|W_k^s D_k^s u\|_{L^2(\hat{\kappa})} \\ &\leq d\Big(\frac{\Gamma(p-s+2)}{\Gamma(p+s+2)}\Big)^{1/2}|u|_{V^s(\hat{\kappa})} \\ &\leq C(s)d(p+1)^{-s}|u|_{H^s(\hat{\kappa})}, \end{aligned} \tag{7.5}$$

where Γ is the Gamma function.

We remark on the asymptotic behaviour of the Gamma function. Making use of Stirling's formula, see (9.15) in [98]

$$\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n} \leq \Gamma(n+1) \leq en^{n+\frac{1}{2}}e^{-n}, \quad n \geq 0, \tag{7.6}$$

we can see that,

$$\frac{\Gamma(p-s+2)}{\Gamma(p+s+2)} \leq C(s)(p+1)^{-2s}, \tag{7.7}$$

with $0 \leq s \leq p+1$ and $C(s)$ depending on the generic constant s only.

For $u|_{\hat{\kappa}} \in H^l(\hat{\kappa})$, $l \geq 0$, we introduce its Legendre polynomial expansion over the reference element $\hat{\kappa}$, given by

$$u(\hat{x}) = \sum_{|i|=0}^{\infty} a_i \prod_{k=1}^{d} L_{i_k}(\hat{x}_k), \tag{7.8}$$

where $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_d)$. We use $L_{i_k}(\hat{x}_k)$ to denote the Legendre polynomial with order i_k over the variable $\hat{x}_k$, and a_i is defined by

$$a_i = \int_{\hat{\kappa}} u(\hat{x}) \prod_{k=1}^{d} \Big(\frac{2i_k+1}{2}\Big) L_{i_k}(\hat{x}_k)\, \mathrm{d}\hat{x}. \tag{7.9}$$

The Legendre polynomials have the following orthogonality property:

$$\int_{-1}^{1} L_i(\xi) L_j(\xi)\, \mathrm{d}\xi = \frac{2\delta_{ij}}{2i+1}, \tag{7.10}$$

which implies that

$$\|u\|^2_{L^2(\hat{\kappa})} = \sum_{|i|=0}^{\infty} |a_i|^2 \prod_{k=1}^{d} \frac{2}{2i_k+1}. \tag{7.11}$$

The derivatives of the function u can be expressed as

$$D^{\alpha} u(\hat{x}) = \sum_{i_1=\alpha_1}^{\infty} \sum_{i_2=\alpha_2}^{\infty} \cdots \sum_{i_d=\alpha_d}^{\infty} a_i \prod_{k=1}^{d} L_{i_k}^{(\alpha_k)}(\hat{x}_k). \tag{7.12}$$

By Lemma 3.10 in [167], the derivatives of the Legendre polynomials satisfy the orthogonality property

$$\int_{-1}^{1} (1-\xi^2)^k L_i^{(k)}(\xi) L_j^{(k)}(\xi)\, \mathrm{d}\xi = \frac{2\delta_{ij}}{2i+1} \frac{\Gamma(i+k+1)}{\Gamma(i-k+1)}, \tag{7.13}$$

where δ_{ij} is the Kronecker delta. Identity (7.13) is related to the following property of Legendre polynomials,

$$L_i^{(k)}(x) = \frac{\Gamma(i+k+1)}{2^k \Gamma(i+1)} P_{i-k}(x; k),$$

where $P_{i-k}(x; k)$ is the Jacobi polynomial of degree $i-k$ with weight $(1-x^2)^k$.

By employing (7.13), we have

$$\|W^{\alpha}D^{\alpha}u\|^2_{L^2(\hat{\kappa})} = \sum_{i_1=\alpha_1}^{\infty}\sum_{i_2=\alpha_2}^{\infty}\cdots\sum_{i_d=\alpha_d}^{\infty}|a_i|^2\prod_{k=1}^{d}\frac{2}{2i_k+1}\frac{\Gamma(i_k+\alpha_k+1)}{\Gamma(i_k-\alpha_k+1)}. \tag{7.14}$$

With the help of (7.14), we shall derive an L^2–norm error bound for $\Pi_p^{\mathcal{P}}$. The proof will be split into several steps.

We first solve a constrained optimization problem in the following Lemma 7.2, which plays a key role for deriving the sharp *hp* bounds.

Lemma 7.2. *Let $\xi = (\xi_1, \xi_2, \ldots, \xi_d)$ and $\rho = (\rho_1, \rho_2, \ldots, \rho_d)$ be two non-negative real valued vectors, $\rho \geq \xi$, and $|\rho| = M$, $|\xi| = m$. Then, the function $F(\xi, \rho)$ will have the global upper bound*

$$F(\xi,\rho) = \prod_{k=1}^{d}\frac{\Gamma(\rho_k-\xi_k+1)}{\Gamma(\rho_k+\xi_k+1)} \leq \Big(\frac{\Gamma(\frac{M-m}{d}+1)}{\Gamma(\frac{M+m}{d}+1)}\Big)^d. \tag{7.15}$$

Furthermore, the maximum value of $F(\xi, \rho)$ under the above constraints on ρ and ξ is obtained at $\xi_k = m/d$, $\rho_k = M/d$, $k = 1, \ldots, d$.

Proof. The proof follows the constrained optimization procedure. We introduce the Lagrange multiplier for $F(\xi, \rho)$,

$$L(\xi,\rho,\mu,\lambda) = F(\xi,\rho) + \mu(|\xi| - m) + \lambda(|\rho| - M), \tag{7.16}$$

and we calculate the stationary points. We consider the partial derivative with respect to ξ_j and ρ_j, $j = 1, \ldots, d$,

$$\frac{\partial L}{\partial \xi_j} = -\left(\frac{\Gamma'(\rho_j-\xi_j+1)}{\Gamma(\rho_j-\xi_j+1)} + \frac{\Gamma'(\rho_j+\xi_j+1)}{\Gamma(\rho_j+\xi_j+1)}\right)F(\xi,\rho) + \mu = 0,$$

and

$$\frac{\partial L}{\partial \rho_j} = \left(\frac{\Gamma'(\rho_j-\xi_j+1)}{\Gamma(\rho_j-\xi_j+1)} - \frac{\Gamma'(\rho_j+\xi_j+1)}{\Gamma(\rho_j+\xi_j+1)}\right)F(\xi,\rho) + \lambda = 0,$$

which satisfy the following conditions:

$$\frac{\Gamma'(\rho_j-\xi_j+1)}{\Gamma(\rho_j-\xi_j+1)} = \frac{\mu-\lambda}{2F(\xi,\rho)}, \quad j = 1, \ldots, d, \tag{7.17}$$

and

$$\frac{\Gamma'(\rho_j+\xi_j+1)}{\Gamma(\rho_j+\xi_j+1)}=\frac{\mu+\lambda}{2F(\xi,\rho)},\quad j=1,\ldots,d, \tag{7.18}$$

by using the fact that $F(\xi,\rho)\geq 1$. The right-hand sides of (7.17) and (7.18) are independent of the index j. Moreover, the function $\psi(z)=\Gamma(z)'/\Gamma(z)$ is the so-called Digamma function with the following property (see [3], (6.3.16)):

$$\psi(z+1)=-\gamma+\sum_{n=1}^{\infty}\frac{z}{n(n+z)}=-\gamma+\sum_{n=1}^{\infty}\Big(\frac{1}{n}-\frac{1}{n+z}\Big),\quad z\neq -1,-2,\ldots,$$

where γ is the Euler constant. For $z\geq 0$, the function $\psi(z+1)$ is a continuous monotonically increasing function, which shows that (7.17) and (7.18) under the constraints will have only one solution. This solution is $\tilde{\xi}_j=m/d$ and $\tilde{\rho}_j=M/d$, $j=1,\ldots,d$, and the $F(\xi,\rho)$ will have the extreme value at this point, given by

$$F(\tilde{\xi},\tilde{\rho})=\Big(\frac{\Gamma(\frac{M-m}{d}+1)}{\Gamma(\frac{M+m}{d}+1)}\Big)^d. \tag{7.19}$$

In order to find the global maximum, we need to prove the following asymptotic relationship:

$$\Big(\frac{\Gamma(\frac{M-m}{k}+1)}{\Gamma(\frac{M+m}{k}+1)}\Big)^k\leq\Big(\frac{\Gamma(\frac{M-m}{d}+1)}{\Gamma(\frac{M+m}{d}+1)}\Big)^d,\qquad k=1,\ldots,d-1. \tag{7.20}$$

The proof of this can be split into three steps. We first consider the special case $m=0$. In this case, (7.20) holds trivially because both sides of the inequality are identically 1. Next, we consider the case $m=\delta M$, with $0<\delta<1$. By using the property of Gamma functions (7.6), we have the following bounds:

$$\Big(\frac{\Gamma(\frac{M-m}{k}+1)}{\Gamma(\frac{M+m}{k}+1)}\Big)^k\leq\Big(\frac{e}{\sqrt{2\pi}}\Big)^k(ek)^{2m}\frac{(M-m)^{M-m+\frac{k}{2}}}{(M+m)^{M+m+\frac{k}{2}}},$$

and

$$\Big(\frac{\Gamma(\frac{M-m}{k}+1)}{\Gamma(\frac{M+m}{k}+1)}\Big)^k\geq\Big(\frac{\sqrt{2\pi}}{e}\Big)^k(ek)^{2m}\frac{(M-m)^{M-m+\frac{k}{2}}}{(M+m)^{M+m+\frac{k}{2}}}.$$

By using the upper and lower bounds from the above inequalities, we can derive the lower bound

$$\begin{aligned}\frac{\left(\frac{\Gamma(\frac{M-m}{d}+1)}{\Gamma(\frac{M+m}{d}+1)}\right)^d}{\left(\frac{\Gamma(\frac{M-m}{k}+1)}{\Gamma(\frac{M+m}{k}+1)}\right)^k} &\geq \frac{\left(\frac{\sqrt{2\pi}}{e}\right)^d (ed)^{2m}\left(\frac{(M-m)^{M-m+\frac{d}{2}}}{(M+m)^{M+m+\frac{d}{2}}}\right)}{\left(\frac{e}{\sqrt{2\pi}}\right)^k (ek)^{2m}\left(\frac{(M-m)^{M-m+\frac{k}{2}}}{(M+m)^{M+m+\frac{k}{2}}}\right)} \\ &\geq \left(\frac{\sqrt{2\pi}}{e}\right)^{d+k}\left(\frac{d}{k}\right)^{2\delta M}\left(\frac{1-\delta}{1+\delta}\right)^{\frac{d-k}{2}}. \qquad (7.21)\end{aligned}$$

By recalling that $0 < \delta < 1$ and $k = 1, \ldots, d-1$, we have that $0 < \frac{1-\delta}{1+\delta} < 1$ and the function $(\frac{d}{k})^{2\delta M}$ is monotonically increasing with respect to M. This implies that, for $M \geq \left((d+k)\log(\frac{e}{\sqrt{2\pi}}) + \frac{d-k}{2}\log(\frac{1+\delta}{1-\delta})\right)\left(2\delta\log(\frac{d}{k})\right)^{-1}$, the above quotient formula is greater than 1 and therefore (7.20) holds. The upper bound for the above quotient can also be derived by using similar techniques, producing

$$\frac{\left(\frac{\Gamma(\frac{M-m}{d}+1)}{\Gamma(\frac{M+m}{d}+1)}\right)^d}{\left(\frac{\Gamma(\frac{M-m}{k}+1)}{\Gamma(\frac{M+m}{k}+1)}\right)^k} \leq \left(\frac{e}{\sqrt{2\pi}}\right)^{d+k}\left(\frac{d}{k}\right)^{2\delta M}\left(\frac{1-\delta}{1+\delta}\right)^{\frac{d-k}{2}}. \qquad (7.22)$$

Finally, we consider the case $m = M$. Using the same techniques used to derive (7.21) together with the fact that $\Gamma(1) = 1$, we have

$$\frac{\left(\frac{\Gamma(\frac{M-m}{d}+1)}{\Gamma(\frac{M+m}{d}+1)}\right)^d}{\left(\frac{\Gamma(\frac{M-m}{k}+1)}{\Gamma(\frac{M+m}{k}+1)}\right)^k} = \frac{(\Gamma(\frac{2M}{k}+1))^k}{(\Gamma(\frac{2M}{d}+1))^d} \geq \frac{(\sqrt{2\pi})^k}{e^d}\left(\frac{d}{2M}\right)^{\frac{d-k}{2}}\left(\frac{d}{k}\right)^{2M+\frac{k}{2}}. \qquad (7.23)$$

By using the fact that exponentially increasing functions grow faster than polynomials, we know that for sufficiently large M the right hand side of (7.23) is greater than 1 and therefore (7.20) holds.

Next, we need to show that the extreme value (7.19) is the global maximum value of $F(\xi, \rho)$ under the constraints $|\xi| = m$ and $|\rho| = M$.

First, we can see that the function $F(\xi, \rho)$ is symmetric and continuous with respect to ξ and ρ. The constraints $|\xi| = m$ and $|\rho| = M$ restrict the domain of ξ and ρ to be a $(d-1)$-dimensional simplex, which is convex and compact. So *the maximum value of the function $F(\xi, \rho)$ over the domain will be obtained only at the boundary of the domain or the stationary point of $F(\xi, \rho)$*. We have calculated the function value at the stationary point in (7.19) already, so now we just need to check the function values on the boundary of the domain.

This may be proved by induction. We start with the case $d = 2$: the domain of ξ and ρ satisfying the constrains are two straight lines. Here, the stationary point is the mid-point of each of the two lines $\tilde{\xi} = (m/2, m/2)$, $\tilde{\rho} = (M/2, M/2)$, and the boundary of the domain consist of the points $\xi^b = (0, m)$, $\rho^b = (0, M)$ or $\xi^b = (m, 0)$, $\rho^b = (M, 0)$, due to the constraints $\rho \geq \xi$. Using the symmetry of the function and of the domain, we know that at the two boundary points of the domain, $F(\xi, \rho)$ will attain the same value, with $F(\xi^b, \rho^b) = \frac{\Gamma(M-m+1)}{\Gamma(M+m+1)}$. By using the asymptotic relation (7.20), the following relation holds

$$F(\xi^b, \rho^b) = \frac{\Gamma(M-m+1)}{\Gamma(M+m+1)} \leq \Big(\frac{\Gamma(\frac{M-m}{2}+1)}{\Gamma(\frac{M+m}{2}+1)}\Big)^2 = F(\tilde{\xi}, \tilde{\rho}).$$

The above relation shows that the extreme value (7.19) is the global maximum value under the constraints for $d = 2$.

Next, we consider the case $d = 3$, where the domain of each of ξ and ρ will be a triangle. In this case, the stationary point of $F(\xi, \rho)$ is when ξ and ρ are located at the barycenter of their respective triangle. The boundary of each domain consists of 3 straight lines. We need to calculate the maximum value of $F(\xi, \rho)$ on the boundary of the domain. By using the symmetry of $F(\xi, \rho)$, and that fact that $|\xi| = m$ and $|\rho| = M$, we only need to consider one part of domain boundary where $\xi_3 = 0$ and $\rho_3 = 0$. Then, the maximum of $F(\xi, \rho)$ on the domain boundary can be viewed as exactly the same problem with the same constraints as in the case $d = 2$. Consequently, the maximum value of $F(\xi, \rho)$ along the boundary of the domain is $F(\xi^b, \rho^b) = \Big(\frac{\Gamma(\frac{M-m}{2}+1)}{\Gamma(\frac{M+m}{2}+1)}\Big)^2$. Again, by using the same techniques as for $d = 2$, we deduce that

$$F(\xi^b, \rho^b) = \Big(\frac{\Gamma(\frac{M-m}{2}+1)}{\Gamma(\frac{M+m}{2}+1)}\Big)^2 \leq \Big(\frac{\Gamma(\frac{M-m}{3}+1)}{\Gamma(\frac{M+m}{3}+1)}\Big)^3 = F(\tilde{\xi}, \tilde{\rho}).$$

The above relation shows that the extreme value (7.19) is the global maximum value under the constraints for $d = 3$. For the general d-dimensional case, the proof can be carried out in a similar way. Another key observation is that the maximum value of $F(\xi, \rho)$ on the domain boundary will be at the stationary points of $F(\xi, \rho)$ on the boundary. By using the relation

$$\Big(\frac{\Gamma(\frac{M-m}{d-1}+1)}{\Gamma(\frac{M+m}{d-1}+1)}\Big)^{d-1} \leq \Big(\frac{\Gamma(\frac{M-m}{d}+1)}{\Gamma(\frac{M+m}{d}+1)}\Big)^{d},$$

the proof is complete. □

With the help of Lemma 7.2, we present an approximation result for the L^2-projection operator $\Pi_p^{\mathcal{P}}$.

Theorem 7.3. *Let $\hat{\kappa} = (-1,1)^d$. Suppose that $u|_{\hat{\kappa}} \in H^l(\hat{\kappa})$, for some $l \geq 0$. Let $\Pi_p^{\mathcal{P}} u$ be the $L^2(\hat{\kappa})$ projection of u onto $\mathcal{P}_p(\hat{\kappa})$ with $p \geq 0$. Then for any integer s, $0 \leq s \leq \min\{p+1, l\}$, we have:*

$$\|u - \Pi_p^{\mathcal{P}} u\|^2_{L^2(\hat{\kappa})} \leq \Big(\frac{\Gamma(\frac{p+1-s}{d}+1)}{\Gamma(\frac{p+1+s}{d}+1)}\Big)^d |u|^2_{V^s(\hat{\kappa})} \leq C(s)\Big(\frac{d}{p+1}\Big)^{2s} |u|^2_{H^s(\hat{\kappa})}. \quad (7.24)$$

Proof. Using the definition of $\Pi_p^{\mathcal{P}}$, (7.8), for any integer s, $0 \leq s \leq \min\{p+1, l\}$, we have

$$\begin{aligned}
\|u - \Pi_p^{\mathcal{P}} u\|^2_{L^2(\hat{\kappa})} &= \sum_{|i|=p+1}^{\infty} |a_i|^2 \prod_{k=1}^{d} \frac{2}{2i_k+1} \\
&\leq \sum_{|\alpha|=s} \sum_{|i|=p+1, i\geq\alpha}^{\infty} |a_i|^2 \prod_{k=1}^{d} \frac{2}{2i_k+1} \\
&\leq \sum_{|\alpha|=s} \sum_{|i|=p+1, i\geq\alpha}^{\infty} |a_i|^2 \Big(\prod_{k=1}^{d} \frac{2}{2i_k+1}\frac{\Gamma(i_k+\alpha_k+1)}{\Gamma(i_k-\alpha_k+1)}\Big)\Big(\prod_{k=1}^{d}\frac{\Gamma(i_k-\alpha_k+1)}{\Gamma(i_k+\alpha_k+1)}\Big) \\
&\leq \Big(\frac{\Gamma(\frac{p+1-s}{d}+1)}{\Gamma(\frac{p+1+s}{d}+1)}\Big)^d \sum_{|\alpha|=s} \sum_{|i|=p+1, i\geq\alpha}^{\infty} |a_i|^2 \prod_{k=1}^{d} \frac{2}{2i_k+1}\frac{\Gamma(i+k+1)}{\Gamma(i-k+1)} \\
&\leq \Big(\frac{\Gamma(\frac{p+1-s}{d}+1)}{\Gamma(\frac{p+1+s}{d}+1)}\Big)^d \sum_{|\alpha|=s} \|W^\alpha D^\alpha u\|^2_{L^2(\hat{\kappa})} \\
&= \Big(\frac{\Gamma(\frac{p+1-s}{d}+1)}{\Gamma(\frac{p+1+s}{d}+1)}\Big)^d |u|^2_{V^s(\hat{\kappa})} \leq C(s)\Big(\frac{d}{p+1}\Big)^{2s} |u|^2_{H^s(\hat{\kappa})}. \quad (7.25)
\end{aligned}$$

In step two, the index set is enlarged; indeed, some of the terms with multi-index $|i| \geq p+1$ have been used more than once. In step three, we use Lemma 7.2, taking $\xi_k = i_k \geq 0$, $\rho_k = \alpha_k \geq 0$, $M = p+1$, $m = s$, together with the restriction $0 \leq s \leq \min\{p+1, l\}$. The bound holds by Stirling's formula (7.6). □

Remark 7.4. We make the comparison between the L^2–norm bound (7.5) for the projector $\Pi_p^{\mathcal{Q}}$ and (7.24) for the projector $\Pi_p^{\mathcal{P}}$. Both bounds are p-optimal for functions with finite Sobolev regularity and also for analytic functions. We can also see that the bound in (7.24) will have a larger constant compared to the bound in (7.5), and this constant only depends on the dimension d. This result will play a key role in deriving the exponential convergence for the $\mathcal{P}_p$ basis.

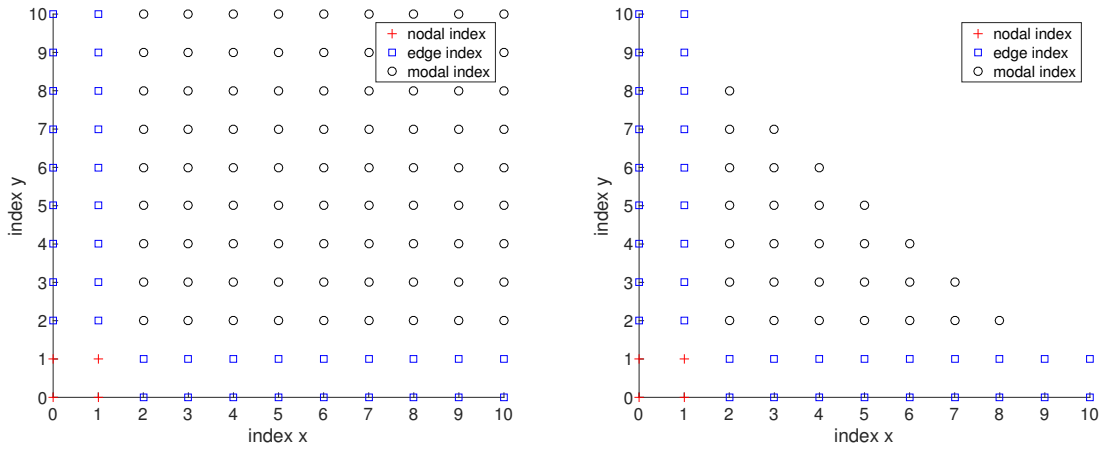


FIGURE 7.1: $\mathcal{Q}_p$ (left) and $\mathcal{S}_p$ (right) with polynomial order 10.

7.1.2 The H^1-projection onto $\mathcal{S}_p$ over the reference square

In this section, we shall consider the H^1-projection over the reference element $\hat{\kappa} := (-1,1)^2$. For the sake of simplicity, we only consider the two-dimensional case. We start by introducing the two-dimensional serendipity finite element space

$$\mathcal{S}_p(\hat{\kappa}) := \mathcal{P}_p(\hat{\kappa}) + \text{span}\{x^p y, y^p x\}. \tag{7.26}$$

Here, we can see in Figure 7.1 that the serendipity space $\mathcal{S}_p$ contains two more basis functions than the $\mathcal{P}_p$ basis for $p \geq 2$. Another way to interpret the serendipity basis is to consider a decomposition of the C^0 finite element space over a rectangle. For polynomial order p, the $\mathcal{S}_p$ basis has the same number of nodal basis functions and edge basis functions as the $\mathcal{Q}_p$ basis, but the $\mathcal{S}_p$ basis only has modal basis functions (those with zero value along the element boundary) whose total degree is less than or equal p. For more details about serendipity FEMs, we refer to [14, 17].

Similarly to the case of the L^2-projection, we use $\mathcal{H}_p^{\mathcal{Q}} := \mathcal{H}_p^{(1)}\mathcal{H}_p^{(2)}$ to denote the H^1-projection onto the $\mathcal{Q}_p$ basis, which can be constructed via a tensor product of one dimensional H^1-projections. Similarly, the H^1-projection onto the $\mathcal{S}_p$ basis is denoted by $\mathcal{H}_p^{\mathcal{S}}$, which is defined in (7.31). Here, we introduce some properties of the one-dimensional H^1-projector $\mathcal{H}_p$ from [167]. To this end, we set $\hat{I} := (-1,1)$.

Then for $u \in H^l(\hat{I})$, $l \geq 1$, the projector $\mathcal{H}_p u \in \mathcal{P}_p(\hat{I})$, $p \geq 1$, is defined by

$$\begin{aligned} \mathcal{H}_p u &= \int_{-1}^{x} \Pi_{p-1} u' \, \mathrm{d}x + u(-1), \\ &= \sum_{j=0}^{p-1} a_j \int_{-1}^{x} L_j(x) \, \mathrm{d}x + u(-1) = \sum_{j=0}^{p-1} a_j \psi_j(x) \, \mathrm{d}x + u(-1), \end{aligned} \tag{7.27}$$

where a_j are as in (7.9), and Π_{p-1} is the L^2-projection. The function $\psi_j(x)$ is the anti-derivative of $L_j(x)$ with degree $j+1$, and satisfies $\psi_j(\pm 1) = 0$ for $j \geq 1$. Moreover, for $j \geq 1$, we have

$$\psi_j(x) = -\frac{1}{j(j+1)}(1-x^2)L_j'(x), \tag{7.28}$$

giving

$$\int_{\hat{I}} \psi_j(x)\psi_k(x)\frac{1}{1-x^2} \, \mathrm{d}x = \frac{2\delta_{jk}}{j(j+1)(2i+1)}. \tag{7.29}$$

The orthogonality property in the weighted L^2–norm will play a key role in the following analysis.

Next, we construct the two-dimensional H^1 projection. First, we consider $\mathcal{H}_p^{\mathcal{Q}} = \mathcal{H}_p^{(1)}\mathcal{H}_p^{(2)}$: for $u \in H^l(\hat{\kappa})$, $l \geq 2$, the projector $\mathcal{H}_p^{\mathcal{Q}} u \in \mathcal{Q}_p(\hat{\kappa})$, $p \geq 1$, is defined by

$$\begin{aligned} \mathcal{H}_p^{\mathcal{Q}} u &:= \int_{-1}^{x_1}\int_{-1}^{x_2} \Pi_{p-1}^{\mathcal{Q}} \partial_1\partial_2 u \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &+ \int_{-1}^{x_1} \Pi_{p-1}^{(1)} \partial_1 u(x_1,-1) \, \mathrm{d}x_1 + \int_{-1}^{x_2} \Pi_{p-1}^{(2)} \partial_2 u(-1,x_2) \, \mathrm{d}x_2 + u(-1,-1) \\ &= \sum_{m=0}^{p-1}\sum_{n=0}^{p-1} a_{mn}\psi_m(x_1)\psi_n(x_2) \\ &+ \sum_{m=0}^{p-1} b_m\psi_m(x_1) + \sum_{n=0}^{p-1} c_n\psi_n(x_2) + u(-1,-1), \end{aligned} \tag{7.30}$$

with a_{nm}, b_m and c_n given by:

$$\begin{aligned} a_{mn} &= \frac{2m+1}{2}\frac{2n+1}{2}\int_{\hat{\kappa}} \partial_1\partial_2 u(x_1,x_2)L_m(x_1)L_n(x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2, \\ b_m &= \frac{2m+1}{2}\int_{-1}^{1} \partial_1 u(x_1,-1)L_m(x_1) \, \mathrm{d}x_1, \\ c_n &= \frac{2n+1}{2}\int_{-1}^{1} \partial_2 u(-1,x_2)L_n(x_2) \, \mathrm{d}x_2. \end{aligned}$$

From the definition of $\mathcal{S}_p$, $\mathcal{H}_p^{\mathcal{S}}$ can be constructed by removing the modal basis functions with order greater than p in $\mathcal{H}_p^{\mathcal{Q}}$. More specifically, for $u \in H^l(\hat{\kappa})$, $l \geq 2$, $\mathcal{H}_p^{\mathcal{S}} u \in \mathcal{S}_p(\hat{\kappa})$, $p \geq 1$, is defined by

$$\begin{aligned}\mathcal{H}_p^{\mathcal{S}} u \;:=\; & \sum_{\substack{m\geq 1, n\geq 1 \\ p-2\geq m+n\geq 2}} a_{mn}\psi_m(x_1)\psi_n(x_2) \\ +\; & \sum_{m=0}^{p-1} a_{m0}\psi_m(x_1)\psi_0(x_2) + \sum_{n=1}^{p-1} a_{0n}\psi_0(x_1)\psi_n(x_2) \\ +\; & \sum_{m=0}^{p-1} b_m\psi_m(x_1) + \sum_{n=0}^{p-1} c_n\psi_n(x_2) + u(-1,-1). \end{aligned} \tag{7.31}$$

Next, we recall the following approximation lemma from [124].

Lemma 7.5. *Let $\hat{\kappa} = (-1,1)^2$. Suppose that $u|_{\hat{\kappa}} \in H^{l+1}(\hat{\kappa})$, for some $l \geq 1$. Let $\mathcal{H}_p^{\mathcal{Q}} u$ be the H^1-projection of u onto $\mathcal{Q}_p(\hat{\kappa})$ with $p \geq 1$. Then, we have*

$$\mathcal{H}_p^{\mathcal{Q}} u = u \quad \textit{at the vertices of } \hat{\kappa}, \tag{7.32}$$

and the following error estimates hold:

$$\begin{aligned}\|u - \mathcal{H}_p^{\mathcal{Q}} u\|^2_{L^2(\hat{\kappa})} \;\leq\; & \frac{2}{p(p+1)}\frac{\Gamma(p-s+1)}{\Gamma(p+s+1)}\Big(\|\partial_1^{s+1}u\|^2_{L^2(\hat{\kappa})} + 2\|\partial_2^{s+1}u\|^2_{L^2(\hat{\kappa})}\Big) \\ +\; & \frac{4}{p^2(p+1)^2}\frac{\Gamma(p-s+2)}{\Gamma(p+s)}\|\partial_1^1\partial_2^s u\|^2_{L^2(\hat{\kappa})}, \end{aligned} \tag{7.33}$$

$$\begin{aligned}\|\nabla(u - \mathcal{H}_p^{\mathcal{Q}} u)\|^2_{L^2(\hat{\kappa})} \leq\; & 2\frac{\Gamma(p-s+1)}{\Gamma(p+s+1)}\Big(\|\partial_1^{s+1}u\|^2_{L^2(\hat{\kappa})} + \|\partial_2^{s+1}u\|^2_{L^2(\hat{\kappa})}\Big) \\ & + \frac{8}{p(p+1)}\frac{\Gamma(p-s+2)}{\Gamma(p+s)}\Big(\|\partial_1^s\partial_2^1 u\|^2_{L^2(\hat{\kappa})} + \|\partial_1^1\partial_2^s u\|^2_{L^2(\hat{\kappa})}\Big), \end{aligned} \tag{7.34}$$

for any integer s, $0 \leq s \leq \min\{p, l\}$.

Now, we derive the L^2–norm error and H^1–norm error bound for the H^1-projection $\mathcal{H}_p^{\mathcal{S}}$.

Theorem 7.6. *Let $\hat{\kappa} = (-1,1)^2$. Suppose that $u|_{\hat{\kappa}} \in H^{l+1}(\hat{\kappa})$, for some $l \geq 1$. Let $\mathcal{H}_p^{\mathcal{S}} u$ be the H^1 projection of u onto $\mathcal{S}_p(\hat{\kappa})$ with $p \geq 1$. Then, we have*

$$\mathcal{H}_p^{\mathcal{S}} u = u \quad \textit{at the vertices of } \hat{\kappa}, \tag{7.35}$$

and for any integer s, $1 \le s \le \min\{p, l\}$, the following error estimates hold:

$$\begin{aligned}
\|u - \mathcal{H}_p^{\mathcal{S}} u\|^2_{L^2(\hat{\kappa})} &\le \frac{4}{p(p+1)} \frac{\Gamma(p-s+1)}{\Gamma(p+s+1)} \Big(\|\partial_1^{s+1} u\|^2_{L^2(\hat{\kappa})} + 2\|\partial_2^{s+1} u\|^2_{L^2(\hat{\kappa})} \Big) \\
&+ \frac{8}{p^2(p+1)^2} \frac{\Gamma(p-s+2)}{\Gamma(p+s)} \|\partial_1^1 \partial_2^s u\|^2_{L^2(\hat{\kappa})} \\
&+ 72 \Big(\frac{\Gamma(\frac{p-s}{2}+1)}{\Gamma(\frac{p+s+2}{2}+1)} \Big)^2 |\partial_1 \partial_2 u|^2_{V^{s-1}(\hat{\kappa})} \\
&\le C(s) \Big(\frac{2}{p+2} \Big)^{2s+2} |u|^2_{H^{s+1}(\hat{\kappa})}. \qquad (7.36)
\end{aligned}$$

$$\begin{aligned}
\|\nabla(u - \mathcal{H}_p^{\mathcal{S}} u)\|^2_{L^2(\hat{\kappa})} &\le 4 \frac{\Gamma(p-s+1)}{\Gamma(p+s+1)} \Big(\|\partial_1^{s+1} u\|^2_{L^2(\hat{\kappa})} + \|\partial_2^{s+1} u\|^2_{L^2(\hat{\kappa})} \Big) \\
&+ \frac{16}{p(p+1)} \frac{\Gamma(p-s+2)}{\Gamma(p+s)} \Big(\|\partial_1^s \partial_2^1 u\|^2_{L^2(\hat{\kappa})} + \|\partial_1^1 \partial_2^s u\|^2_{L^2(\hat{\kappa})} \Big) \\
&+ 24 \Big(\frac{\Gamma(\frac{p-s}{2}+1)}{\Gamma(\frac{p+s}{2}+1)} \Big)^2 |\partial_1 \partial_2 u|^2_{V^{s-1}(\hat{\kappa})} \\
&\le C(s) \Big(\frac{2}{p} \Big)^{2s} |u|^2_{H^{s+1}(\hat{\kappa})}. \qquad (7.37)
\end{aligned}$$

Proof. The key observation is the fact that the serendipity basis $\mathcal{S}_p$ differs from $\mathcal{Q}_p$ only at the modal basis functions which vanish along the boundary of $\hat{\kappa}$. Indeed, using (7.30) and (7.31), we have

$$\mathcal{H}_p^{\mathcal{Q}} u - \mathcal{H}_p^{\mathcal{S}} u = \sum_{\substack{p-1 \ge m \ge 1 \\ p-1 \ge n \ge 1 \\ m+n \ge p-1}} a_{mn} \psi_m(x_1) \psi_n(x_2). \qquad (7.38)$$

Using the fact $\psi_m(\pm 1) = 0$, for $m \ge 1$, we deduce that $(\mathcal{H}_p^{\mathcal{Q}} u - \mathcal{H}_p^{\mathcal{S}} u)|_{\partial\hat{\kappa}} = 0$. Thus, (7.35) is proved.

Next, we derive (7.36). The first step is the use of the triangle inequality,

$$\|u - \mathcal{H}_p^{\mathcal{S}} u\|^2_{L^2(\hat{\kappa})} \le 2\|u - \mathcal{H}_p^{\mathcal{Q}} u\|^2_{L^2(\hat{\kappa})} + 2\|\mathcal{H}_p^{\mathcal{Q}} u - \mathcal{H}_p^{\mathcal{S}} u\|^2_{L^2(\hat{\kappa})}. \qquad (7.39)$$

Thus, we only need to consider the error from the second term in the above bound. By using (7.30), (7.31), (7.28) and (7.29) and the orthogonality of $\psi_j(x)$ for $j \ge 1$,

we have

$$\begin{aligned}
\|\mathcal{H}_p^{\mathcal{S}}u - \mathcal{H}_p^{\mathcal{Q}}u\|^2_{L^2(\hat{\kappa})} &\leq \|(\mathcal{H}_p^{\mathcal{S}}u - \mathcal{H}_p^{\mathcal{Q}}u)W_1^{-1}W_2^{-1}\|^2_{L^2(\hat{\kappa})} \\
&= \sum_{\substack{p-1\geq m\geq 1, p-1\geq n\geq 1 \\ m+n\geq p-1}} |a_{mn}|^2 \frac{2}{2m+1}\frac{2}{2n+1}\frac{1}{m(m+1)}\frac{1}{n(n+1)} \\
&\leq \sum_{|\alpha|=s-1}\sum_{\substack{m\geq\alpha_1, n\geq\alpha_2 \\ m+n\geq p-1}} |a_{mn}|^2 \frac{2}{2m+1}\frac{2}{2n+1}\frac{1}{m(m+1)}\frac{1}{n(n+1)} \\
&\quad \times\Big(\frac{\Gamma(m-\alpha_1+1)}{\Gamma(m+\alpha_1+1)}\frac{\Gamma(n-\alpha_2+1)}{\Gamma(n+\alpha_2+1)}\Big) \\
&\quad \times\Big(\frac{\Gamma(m+\alpha_1+1)}{\Gamma(m-\alpha_1+1)}\frac{\Gamma(n+\alpha_2+1)}{\Gamma(n-\alpha_2+1)}\Big).
\end{aligned}$$

In step three, we enlarge the summation index sets by adding the high order terms.

$$\begin{aligned}
&\|\mathcal{H}_p^{\mathcal{S}}u - \mathcal{H}_p^{\mathcal{Q}}u\|^2_{L^2(\hat{\kappa})} \\
&\quad\leq \sum_{|\alpha|=s-1}\sum_{\substack{m\geq\alpha_1, n\geq\alpha_2 \\ m+n\geq p-1}} |a_{mn}|^2 \frac{2}{2m+1}\frac{2}{2n+1}\Big(\frac{\Gamma(m+\alpha_1+1)}{\Gamma(m-\alpha_1+1)}\frac{\Gamma(n+\alpha_2+1)}{\Gamma(n-\alpha_2+1)}\Big) \\
&\qquad\times\Big(\frac{\Gamma(m-\alpha_1+1)}{\Gamma(m+\alpha_1+1)}\frac{\Gamma(n-\alpha_2+1)}{\Gamma(n+\alpha_2+1)}\frac{1}{m(m+1)}\frac{1}{n(n+1)}\Big) \\
&\quad\leq \sum_{|\alpha|=s-1}\sum_{\substack{m\geq\alpha_1, n\geq\alpha_2 \\ m+n\geq p-1}} |a_{mn}|^2 \frac{2}{2m+1}\frac{2}{2n+1}\Big(\frac{\Gamma(m+\alpha_1+1)}{\Gamma(m-\alpha_1+1)}\frac{\Gamma(n+\alpha_2+1)}{\Gamma(n-\alpha_2+1)}\Big) \\
&\qquad\times 36\Big(\frac{\Gamma(m-\alpha_1+1)}{\Gamma(m+\alpha_1+3)}\frac{\Gamma(n-\alpha_2+1)}{\Gamma(n+\alpha_2+3)}\Big) \\
&\quad\leq 36\Big(\frac{\Gamma(\frac{p-s}{2}+1)}{\Gamma(\frac{p+s+2}{2}+1)}\Big)^2 \sum_{|\alpha|=s-1}\|W^\alpha D^\alpha(\partial_1\partial_2 u)\|^2_{L^2(\hat{\kappa})} \\
&\quad\leq 36\Big(\frac{\Gamma(\frac{p-s}{2}+1)}{\Gamma(\frac{p+s+2}{2}+1)}\Big)^2 |\partial_1\partial_2 u|^2_{V^{s-1}(\hat{\kappa})} \leq C(s)\Big(\frac{2}{p+2}\Big)^{2s+2}|u|^2_{H^{s+1}(\hat{\kappa})}. \qquad (7.40)
\end{aligned}$$

In step two, by employing the relation $m \geq \alpha_1$, we have

$$\begin{aligned}
\frac{1}{m(m+1)} &= \frac{1}{(m+\alpha_1+1)(m+\alpha_1+2)}\frac{(m+\alpha_1+1)(m+\alpha_1+2)}{m(m+1)} \\
&\leq \frac{6}{(m+\alpha_1+1)(m+\alpha_1+2)}.
\end{aligned}$$

In step three, we use Lemma 7.2, with $\xi_1 = \alpha_1 + 1 \geq 0$, $\xi_2 = \alpha_2 + 1 \geq 0$, $\rho_1 = m + 1 \geq 0$, $\rho_2 = n + 1 \geq 0$, $M = p + 1$, and $m = s + 1$, together with the restriction $1 \leq s \leq \min\{p, l-1\}$, and in the last step, we use Stirling's formula (7.6).

Using the same techniques, we can derive the error estimate for the H^1–seminorm. We have

$$\begin{aligned}\|\partial_1(\mathcal{H}_p^{\mathcal{S}}u - \mathcal{H}_p^{\mathcal{Q}}u)\|^2_{L^2(\hat{\kappa})} &\leq \|\partial_1(\mathcal{H}_p^{\mathcal{S}}u - \mathcal{H}_p^{\mathcal{Q}}u)W_2^{-1}\|^2_{L^2(\hat{\kappa})} \\ &\leq \sum_{|\alpha|=s-1}\sum_{\substack{m\geq\alpha_1, n\geq\alpha_2 \\ m+n\geq p-1}} |a_{mn}|^2 \frac{2}{2m+1}\frac{2}{2n+1}\frac{1}{n(n+1)} \\ &\quad\times\Big(\frac{\Gamma(m-\alpha_1+1)}{\Gamma(m+\alpha_1+1)}\frac{\Gamma(n-\alpha_2+1)}{\Gamma(n+\alpha_2+1)}\Big) \\ &\quad\times\Big(\frac{\Gamma(m+\alpha_1+1)}{\Gamma(m-\alpha_1+1)}\frac{\Gamma(n+\alpha_2+1)}{\Gamma(n-\alpha_2+1)}\Big).\end{aligned}$$

In step two, we enlarge the summation index sets by adding the high order terms.

$$\begin{aligned}&\|\partial_1(\mathcal{H}_p^{\mathcal{S}}u - \mathcal{H}_p^{\mathcal{Q}}u)\|^2_{L^2(\hat{\kappa})} \\ &\quad\leq \sum_{|\alpha|=s-1}\sum_{\substack{m\geq\alpha_1, n\geq\alpha_2 \\ m+n\geq p-1}} |a_{mn}|^2 \frac{2}{2m+1}\frac{2}{2n+1}\Big(\frac{\Gamma(m+\alpha_1+1)}{\Gamma(m-\alpha_1+1)}\frac{\Gamma(n+\alpha_2+1)}{\Gamma(n-\alpha_2+1)}\Big) \\ &\qquad\times 6\Big(\frac{\Gamma(m-\alpha_1+1)}{\Gamma(m+\alpha_1+1)}\frac{\Gamma(n-\alpha_2+1)}{\Gamma(n+\alpha_2+3)}\Big) \\ &\quad\leq 6\Big(\frac{\Gamma(\frac{p-s}{2}+1)}{\Gamma(\frac{p+s}{2}+1)}\Big)^2 \sum_{|\alpha|=s-1}\|W^\alpha D^\alpha(\partial_1\partial_2 u)\|^2_{L^2(\hat{\kappa})} \\ &\quad= 6\Big(\frac{\Gamma(\frac{p-s}{2}+1)}{\Gamma(\frac{p+s}{2}+1)}\Big)^2 |\partial_1\partial_2 u|^2_{V^{s-1}(\hat{\kappa})} \leq C(s)\Big(\frac{2}{p}\Big)^{2s}|u|^2_{H^{s+1}(\hat{\kappa})},\end{aligned} \tag{7.41}$$

where in step two we use Lemma 7.2, taking $\xi_1 = \alpha_1 \geq 0$, $\xi_2 = \alpha_2 + 1 \geq 0$, $\rho_1 = m \geq 0$, $\rho_2 = n + 1 \geq 0$, $M = p$, and $m = s$, together with the restriction $1 \leq s \leq \min\{p, l-1\}$.

Therefore, we have the bound

$$\begin{aligned}\|\nabla(\mathcal{H}_p^{\mathcal{S}}u - \mathcal{H}_p^{\mathcal{Q}}u)\|^2_{L^2(\hat{\kappa})} &\leq 12\Big(\frac{\Gamma(\frac{p-s}{2}+1)}{\Gamma(\frac{p+s}{2}+1)}\Big)^2 |\partial_1\partial_2 u|^2_{V^{s-1}(\hat{\kappa})} \\ &\leq C(s)\Big(\frac{2}{p}\Big)^{2s}|u|^2_{H^{s+1}(\hat{\kappa})}.\end{aligned} \tag{7.42}$$

Finally, using (7.40), (7.42) and Lemma 7.5, the bounds (7.36) and (7.37) follow. □

Remark 7.7. We again make the comparison between the bounds in the L^2– and H^1–norms, given in (7.33) and (7.34) respectively for $\mathcal{H}_p^{\mathcal{Q}}$, and (7.36) and (7.37) respectively for $\mathcal{H}_p^{\mathcal{S}}$. Similarly to the comparisons for the L^2-projection onto $\mathcal{P}_p$ and $\mathcal{Q}_p$, both bounds are p-optimal for functions with finite Sobolev regularity

and also for analytic functions. We can also see that the bounds for $\mathcal{H}_p^{\mathcal{S}}$ have a larger constant than those for $\mathcal{H}_p^{\mathcal{Q}}$.

Finally, we present the error bound for $\mathcal{H}_p^{\mathcal{P}}$ which we shall define now. The key observation is that the $\mathcal{P}_p$ basis with polynomial order p contains the $\mathcal{S}_{p+1-d}$ basis for $p \geq d$. Then, we can simply define $\mathcal{H}_p^{\mathcal{P}} = \mathcal{H}_{p-1}^{\mathcal{S}}$ for $d = 2$.

Corollary 7.8. *Let $\hat{\kappa} = (-1, 1)^2$. Suppose that $u|_{\hat{\kappa}} \in H^{l+1}(\hat{\kappa})$, for some $l \geq 1$. Let $\mathcal{H}_p^{\mathcal{P}} u := \mathcal{H}_{p-1}^{\mathcal{S}} u$ be the H^1 projection of u onto $\mathcal{P}_p(\hat{\kappa})$ with $p \geq 2$. Then, we have:*

$$\mathcal{H}_p^{\mathcal{P}} u = u \quad \text{at the vertices of } \hat{\kappa}, \tag{7.43}$$

and the following error estimates hold:

$$\|u - \mathcal{H}_p^{\mathcal{P}} u\|^2_{L^2(\hat{\kappa})} = \|u - \mathcal{H}_{p-1}^{\mathcal{S}} u\|^2_{L^2(\hat{\kappa})} \leq C(s)\Big(\frac{2}{p+1}\Big)^{2s+2} |u|^2_{H^{s+1}(\hat{\kappa})}. \tag{7.44}$$

$$\|\nabla(u - \mathcal{H}_p^{\mathcal{P}} u)\|^2_{L^2(\hat{\kappa})} = \|\nabla(u - \mathcal{H}_{p-1}^{\mathcal{S}} u)\|^2_{L^2(\hat{\kappa})} \leq C(s)\Big(\frac{2}{p-1}\Big)^{2s} |u|^2_{H^{s+1}(\hat{\kappa})}. \tag{7.45}$$

for any integer s, $1 \leq s \leq \min\{p-1, l\}$.

Remark 7.9. We emphasize that *the above error bound for the $\mathcal{H}_p^{\mathcal{P}}$ projector is p-sub-optimal by one order for analytic functions, and p-optimal for functions with finite Sobolev regularity in the case $l \leq p-1$.* However, sub-optimality by one order in p is better than using the $\mathcal{H}_{\lfloor p/2 \rfloor}^{\mathcal{Q}}$ projector, as suggested by [167] (see Corollary 4.52 on p190), which is sub-optimal in p by at least $p/2$ orders. Moreover, the one order sub-optimality in p for analytic functions does not influence the exponential convergence results presented in the next section.

7.2 Exponential convergence for DGFEMs

We shall be concerned with the proof of exponential convergence for DGFEMs with $\mathcal{P}_p$ basis over tensor product elements. For simplicity, we only consider the case when the given problem is piecewise analytic over the whole computational domain. Exponential convergence is then achieved by fixing the computational mesh $\mathcal{T}_h$, and increasing the polynomial order p. Only parallelepiped meshes are considered, which are the affine family obtained from the reference element

$\hat{\kappa} = (-1,1)^d$. The analysis of DGFEMs with a general *hp*-refinement strategy are beyond the scope of this analysis (see [163, 164, 165] for details).

The proof of exponential convergence for DGFEMs depends on proving exponential convergence of L^2- and H^1-projections for piecewise analytic functions under p-refinement, as shown in the previous section. For deriving error bounds for DGFEMs using the L^2- and H^1- projectors onto $\mathcal{Q}_p$, we refer to [124, 125, 103]. *Following similar techniques, we can prove the corresponding hp-bounds for DGFEMs employing the $\mathcal{P}_p$ basis, albeit with sub-optimal rate in p. The sub-optimality in p is due to the fact that the H^1-projector onto $\mathcal{P}_p$ is one order sub-optimal.* As we proved in the previous section, the sub-optimality in p is independent of p, and therefore does not influence the slope of the exponential convergence. Additionally, we point out that the approximation results for the H^1-projector $\mathcal{H}_p^{\mathcal{S}}$ onto $\mathcal{S}_p$ can be directly applied to *hp*-FEMs for elliptic problems with same optimal rate as the H^1 projector $\mathcal{H}_p^{\mathcal{Q}}$, see [167] for details.

For the sake of simplicity, we focus on deriving the exponential convergence for the L^2-projection in the L^2–norm on sufficiently smooth problems under p-refinement. The proof for the H^1-projection can be done analogously.

We shall derive the exponential convergence on general parallelepiped meshes. Let κ be an element of $\mathcal{T}_h$ with diameter $h_\kappa \leq 1$. For a function u having an analytic extension into an open neighbourhood of $\bar{\kappa}$, we have for every $s_\kappa \geq 0$:

$$\exists R_\kappa > 0, \quad C > 0 \quad \forall s_\kappa : |u|_{H^{s_\kappa}(\kappa)} \leq C(R_\kappa)^{s_\kappa}\Gamma(s_\kappa+1)|\kappa|^{1/2}, \tag{7.46}$$

where $|\kappa|$ denotes the measure of element κ, cf. [83, Theorem 1.9.3].

Lemma 7.10. *Let $u : \kappa \to \mathbb{R}$ have an analytic extension to an open neighbourhood of $\bar{\kappa}$. Also let $p_\kappa \geq 0$ and $0 \leq s_\kappa \leq p_\kappa + 1$ be two positive numbers such that $s_\kappa = \epsilon(p_\kappa+1)$, $0 < \epsilon < 1$ and $d = 2, 3$. Then the following bounds hold:*

$$\begin{aligned} \|u - \Pi_{p_\kappa}^{\mathcal{Q}} u\|^2_{L^2(\kappa)} &\leq d^2\Big(\frac{h_\kappa}{2}\Big)^{2s_\kappa}\frac{\Gamma(p_\kappa - s_\kappa + 2)}{\Gamma(p_\kappa + s_\kappa + 2)}|u|^2_{H^{s_\kappa}(\hat{\kappa})} \\ &\leq C(u)(p+1)e^{-2b_\kappa^1(p_\kappa+1)}|\kappa|, \end{aligned} \tag{7.47}$$

and

$$\begin{aligned} \|u - \Pi_{p_\kappa}^{\mathcal{P}} u\|^2_{L^2(\kappa)} &\leq \Big(\frac{h_\kappa}{2}\Big)^{2s_\kappa}\Big(\frac{\Gamma(\frac{p_\kappa+1-s_\kappa}{d}+1)}{\Gamma(\frac{p_\kappa+1+s_\kappa}{d}+1)}\Big)^d|u|^2_{H^{s_\kappa}(\hat{\kappa})} \\ &\leq C(u)(p+1)e^{-2b_\kappa^2(p_\kappa+1)}|\kappa|. \end{aligned} \tag{7.48}$$

Here, $C(u)$ is a positive constant depending on u, $F_1(R_\kappa,\epsilon) = \frac{(1-\epsilon)^{1-\epsilon}}{(1+\epsilon)^{1+\epsilon}}(\epsilon R_k)^{2\epsilon}$, $\epsilon_{\min} = 1/\sqrt{1+R_\kappa^2}$, $b_\kappa^1 := \frac{1}{2}|\log F_1(R_\kappa,\epsilon_{\min})| + \epsilon_{\min}|\log\frac{h_\kappa}{2}|$ and $b_\kappa^2 := b_\kappa^1 - \epsilon_{\min}\log d$.

Proof. Using standard scaling arguments, we have the approximation results for L^2-projection over κ. For brevity, we set $q_\kappa = p_\kappa + 1$. By employing Stirling's formula, we have the bounds:

$$\begin{aligned}\frac{\Gamma(p_\kappa - s_\kappa + 2)}{\Gamma(p_\kappa + s_\kappa + 2)}|u|^2_{H^{s_\kappa}(\hat\kappa)} &\le C(R_\kappa)^{2s_\kappa}\Gamma(s_\kappa+1)^2\frac{\Gamma(q_\kappa - s_\kappa + 1)}{\Gamma(q_\kappa + s_\kappa + 1)}|\kappa| \\ &\le C(R_\kappa)^{2\epsilon q_\kappa}\frac{(\epsilon q_\kappa)^{2\epsilon q_\kappa+1}}{e^{2\epsilon q_\kappa}}\frac{((1-\epsilon)q_\kappa)^{(1-\epsilon)q_\kappa}e^{-(1-\epsilon)q_\kappa}}{((1+\epsilon)q_\kappa)^{(1+\epsilon)q_\kappa}e^{-(1+\epsilon)q_\kappa}}|\kappa| \\ &\le Cq_\kappa(F_1(R_\kappa,\epsilon))^{q_\kappa}|\kappa|,\end{aligned}$$

where

$$F_1(R_\kappa,\epsilon) = \frac{(1-\epsilon)^{1-\epsilon}}{(1+\epsilon)^{1+\epsilon}}(\epsilon R_k)^{2\epsilon}.$$

Recalling (7.46), we have $R_\kappa > 1$,

$$\min_{0<\epsilon<1} F_1(R_\kappa,\epsilon) = F_1(R_\kappa,\epsilon_{\min}) = \left(\frac{R_\kappa}{\sqrt{1+R_\kappa^2}+1}\right)^2 < 1, \quad \epsilon_{\min} = \frac{1}{\sqrt{1+R_\kappa^2}}. \tag{7.49}$$

Thus, we have

$$\frac{\Gamma(p_\kappa - s_\kappa + 2)}{\Gamma(p_\kappa + s_\kappa + 2)}|u|^2_{H^{s_\kappa}(\hat\kappa)} \le Cq_\kappa e^{-|\log F_1(R_\kappa,\epsilon_{\min})|q_\kappa}|\kappa|. \tag{7.50}$$

Therefore, we have the exponential convergence for the L^2-projection $\Pi_p^{\mathcal{Q}}$, via

$$\|u - \Pi_{p_\kappa}^{\mathcal{Q}}u\|^2_{L^2(\kappa)} \le C(p+1)e^{-2b_\kappa^1(p_\kappa+1)}|\kappa|, \tag{7.51}$$

with $b_\kappa^1 := \frac{1}{2}|\log F_1(R_\kappa,\epsilon_{\min})| + \epsilon_{\min}|\log\frac{h_\kappa}{2}|$. Similarly, for the L^2-projection $\Pi_p^{\mathcal{P}}$, Stirling's formula implies

$$\begin{aligned}\Big(\frac{\Gamma(\frac{p_\kappa+1-s_\kappa}{d}+1)}{\Gamma(\frac{p_\kappa+1+s_\kappa}{d}+1)}\Big)^d|u|^2_{H^{s_\kappa}(\hat\kappa)} &\le C(R_\kappa)^{2s_\kappa}\Gamma(s_\kappa+1)^2\Big(\frac{\Gamma(\frac{q_\kappa-s_\kappa}{d}+1)}{\Gamma(\frac{q_\kappa+s_\kappa}{d}+1)}\Big)^d|\kappa| \\ &\le C(R_\kappa)^{2\epsilon q_\kappa}\frac{(\epsilon q_\kappa)^{2\epsilon q_\kappa+1}}{e^{2\epsilon q_\kappa}}\frac{((1-\epsilon)q_\kappa)^{(1-\epsilon)q_\kappa}(ed)^{-(1-\epsilon)q_\kappa}}{((1+\epsilon)q_\kappa)^{(1+\epsilon)q_\kappa}(ed)^{-(1+\epsilon)q_\kappa}}|\kappa| \\ &\le Cq_\kappa(F_2(R_\kappa,\epsilon))^{q_\kappa}|\kappa|,\end{aligned}$$

where,

$$F_2(R_\kappa,\epsilon) = \frac{(1-\epsilon)^{1-\epsilon}}{(1+\epsilon)^{1+\epsilon}}(\epsilon R_k d)^{2\epsilon},$$

with the minimum,

$$\min_{0<\epsilon<1} F_2(R_\kappa, \epsilon) = \left(\frac{R_\kappa d}{\sqrt{1 + (R_\kappa d)^2} + 1} \right)^2 < 1.$$

In order to make comparison with the slope of projector $\Pi_p^{\mathcal{Q}}$, here we will use the same $\epsilon_{\min}$. We have

$$\min_{0<\epsilon<1} F_2(R_\kappa, \epsilon) \leq F_2(R_\kappa, \epsilon_{\min}) = F_1(R_\kappa, \epsilon_{\min}) d^{2\epsilon_{\min}}.$$

Thus, we have

$$\|u - \Pi_{p_\kappa}^{\mathcal{P}} u\|_{L^2(\kappa)}^2 \leq C(p+1) e^{-2b_\kappa^2 (p_\kappa + 1)} |\kappa|, \tag{7.52}$$

with slope $b_\kappa^2 := \frac{1}{2} |\log F_1(R_\kappa, \epsilon_{\min})| + \epsilon_{\min}(|\log \frac{h_\kappa}{2}| - \log d)$. The proof is complete. □

In the above theorem, we can see that the L^2–norm error for both L^2-projections $\Pi_{p_\kappa}^{\mathcal{Q}}$ and $\Pi_{p_\kappa}^{\mathcal{P}}$ decays exponentially for analytic functions under p-refinement. If we measure the error against p, the slope b_κ^1 for the $\mathcal{Q}_p$ basis is greater than the slope b_κ^2 for the $\mathcal{P}_p$ basis by a small factor of $(\log d)/\sqrt{1 + R_\kappa^2}$. From Lemma 7.10 we can also derive the following corollary.

Corollary 7.11. *Let u be an analytic function as defined in Lemma 7.10. Then, the following bounds hold:*

$$\|u - \Pi_{p_\kappa}^{\mathcal{Q}} u\|_{L^2(\kappa)}^2 \leq C(u) e^{-2b_\kappa^1 \sqrt[d]{Dof}} |\kappa|, \tag{7.53}$$

and

$$\|u - \Pi_{p_\kappa}^{\mathcal{P}} u\|_{L^2(\kappa)}^2 \leq C(u) e^{-2(b_\kappa^2 \sqrt[d]{d!}) \sqrt[d]{Dof}} |\kappa|. \tag{7.54}$$

Proof. By recalling the relationship between degrees of freedom and polynomial order p for both $\mathcal{P}_p$ basis and $\mathcal{Q}_p$ basis, we have

$$Dof(\mathcal{Q}_p) = (p+1)^d \quad \text{and} \quad Dof(\mathcal{Q}_p) = \binom{p+d}{d} = \frac{(p+1)^d}{d!} + \mathcal{O}((p+1)^{d-1}).$$

Then, (7.53) and (7.54) follow from Lemma 7.10. □

For $d = 2, 3$, if the following condition

$$\frac{1}{2}|\log F_1(R_\kappa, \epsilon_{\min})| + \epsilon_{\min}|\log \frac{h_\kappa}{2}| \gg \epsilon_{\min} \log d, \tag{7.55}$$

holds, then we have $b_\kappa^2 \approx b_\kappa^1$. By recalling (7.49), we know that for sufficiently small R_κ or sufficiently small mesh size h, the condition (7.55) will be satisfied.

Now, if we consider the error in terms of $\sqrt[d]{Dof}$ for the above bounds, a by a fixed factor of $\sqrt[d]{d!}$. The above exponential convergence for the L^2 projector with each basis type also holds for H^1 projector. It is also possible to prove the same steeper slope in error against degrees of freedom for $\mathcal{H}_p^{\mathcal{P}}$ and $\mathcal{H}_p^{\mathcal{S}}$ with respect to $\mathcal{H}_p^{\mathcal{Q}}$, due to the fact that the number of degrees of freedom in the $\mathcal{P}_p$ basis and the $\mathcal{S}_p$ grow at asymptotically the same rate in p. For brevity, we do not prove this here.

We have observed the better slope in error against $\sqrt[d]{Dof}$ for DGFEMs with $\mathcal{P}_p$. For $d = 2$, this suggests a typical ratio between convergence slopes of DGFEMs with $\mathcal{P}_p$ and $\mathcal{Q}_p$ basis to be $\sqrt{2!} \approx 1.414$. For $d = 3$, this ratio is $\sqrt[3]{3!} \approx 1.817$. The numerical examples show that the ratio is slightly worse than the ideal ratio. For $d = 2$, the computed ratio is approximately between 1.3 and 1.4. and for $d = 3$, the computed ratio is approximately 1.6. The numerical examples in the next section confirm the statements above.

7.3 Numerical examples

We present some numerical examples to confirm the theoretical analysis in this chapter. The comparisons are made between the slope of DGFEMs with $\mathcal{P}_p$ and $\mathcal{Q}_p$ basis over rectangle meshes for $d = 2$ and hexahedral meshes for $d = 3$ under p-refinement. The slopes of the convergence lines are calculated by taking the average of the last two slopes of the line segments of each convergence line.

7.3.1 Example 1

Let Ω be the square domain $(-1, 1)^2$, and choose

$$a \equiv 0, \quad \mathbf{b} = (2 - y^2, 2 - x), \quad c = 1 + (1 + x)(1 + y)^2; \tag{7.56}$$

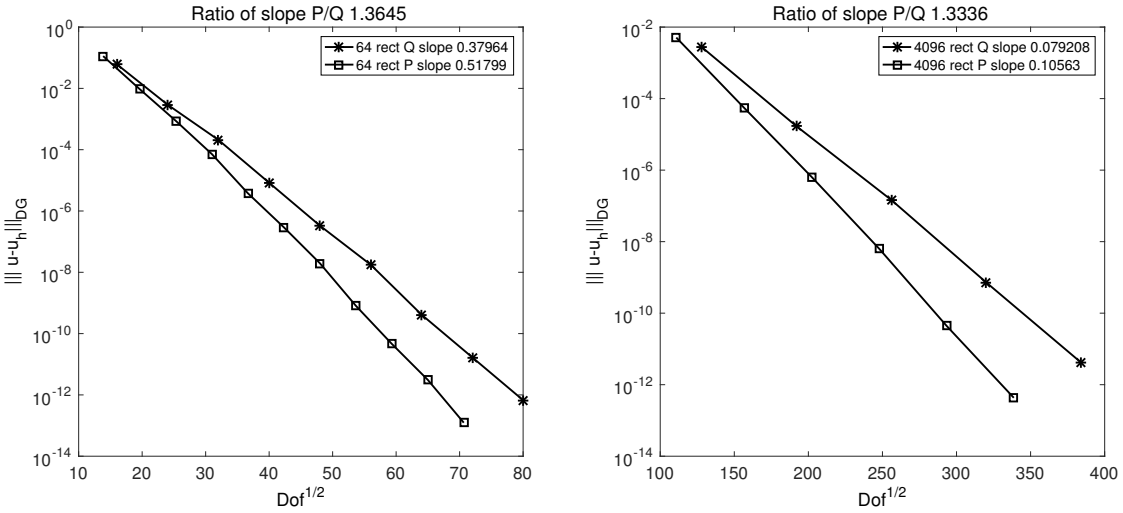


FIGURE 7.2: Example 1: Convergence of the DGFEM under p-refinement. Square meshes with 64 elements (left) and 4096 elements (right).

the forcing function f is selected so that the analytical solution to (5.1), (5.5) is given by

$$u(x,y) = 1 + \sin(\pi(1+x)(1+y)^2/8). \tag{7.57}$$

This example is the one from Section 5.3.1. In Figure 7.2, we can see that the slope of DGFEMs with $\mathcal{P}_p$ basis is greater than the slope of DGFEMs with $\mathcal{Q}_p$ basis in error against $\sqrt[2]{Dof}$. The ratio between the two slopes is about 1.35.

7.3.2 Example 2

Let $\Omega = (-1,1)^2$, and consider the PDE problem:

$$\begin{cases} -x^2 u_{yy} + u_x + u = 0, & \text{for } -1 \le x \le 1, y > 0, \\ u_x + u = 0, & \text{for } -1 \le x \le 1, y \le 0, \end{cases} \tag{7.58}$$

with analytical solution:

$$u(x,y) = \begin{cases} \sin(\frac{1}{2}\pi(1+y))\exp(-(x+\frac{\pi^2 x^3}{12})), & \text{for } -1 \le x \le 1, y > 0, \\ \sin(\frac{1}{2}\pi(1+y))\exp(-x), & \text{for } -1 \le x \le 1, y \le 0. \end{cases} \tag{7.59}$$

This example is the one from Section 5.3.2. In Figure 7.3, we can see that the slope of DGFEMs with $\mathcal{P}_p$ basis is greater than the slope of DGFEMs with $\mathcal{Q}_p$ basis in error against $\sqrt[2]{Dof}$. The ratio between the two slopes is about 1.38.

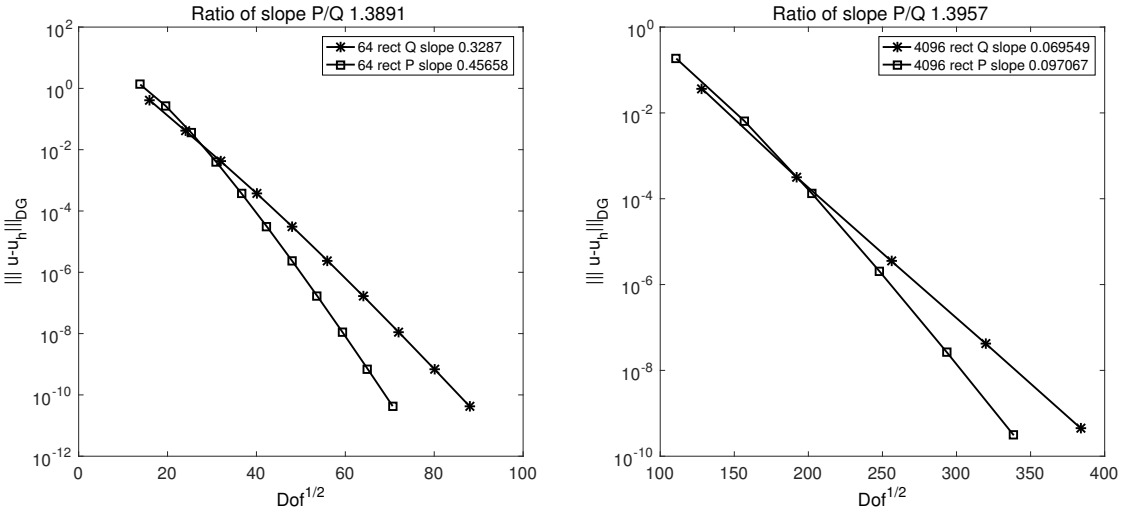


FIGURE 7.3: Example 2: Convergence of the DGFEM under p-refinement. Square meshes with 64 elements (left) and 4096 elements (right).

7.3.3 Example 3

We now consider a singularly perturbed advection-diffusion problem equation

$$-\epsilon\Delta u + u_x + u_y = f,$$

with $\Omega := (0,1)^2$, where $0 < \epsilon \ll 1$ and f is chosen so that

$$u(x,y) = x + y(1-x) + \frac{[e^{-1/\epsilon} - e^{-(1-x)(1-y)/\epsilon}]}{[1-e^{-1/\epsilon}]}. \tag{7.60}$$

From Section 5.3.3. We observe the same behaviour as before over anisotropically refined meshes graded towards to layer with slope ratio 1.31.

7.3.4 Example 4

Moving to three-dimensional problems, we consider

$$-\Delta u = f,$$

over the domain $\Omega = (0,1)^3$. The analytic solution is $u = \sin(\pi x)\sin(\pi y)\sin(\pi z)$.

In Figure 7.5, we observe that the slope of DGFEMs with $\mathcal{P}_p$ basis is greater than the slope of DGFEMs with $\mathcal{Q}_p$ basis in error against $\sqrt[3]{Dof}$. The ratio between the two slopes is about 1.61.

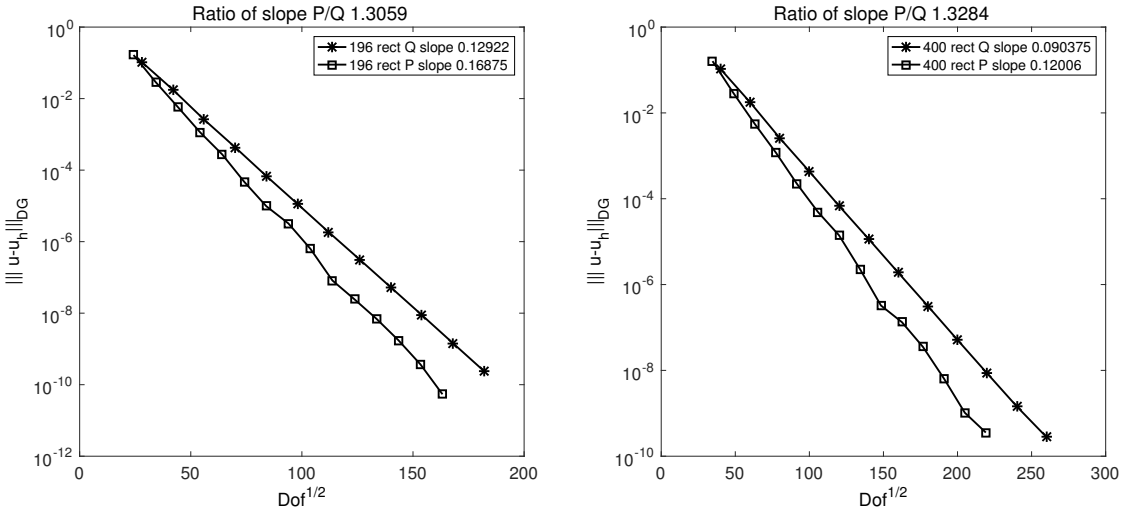


FIGURE 7.4: Example 3: Convergence of the DGFEM under p-refinement. Anisotropically refined meshes with 196 elements (left) and 400 elements (right).

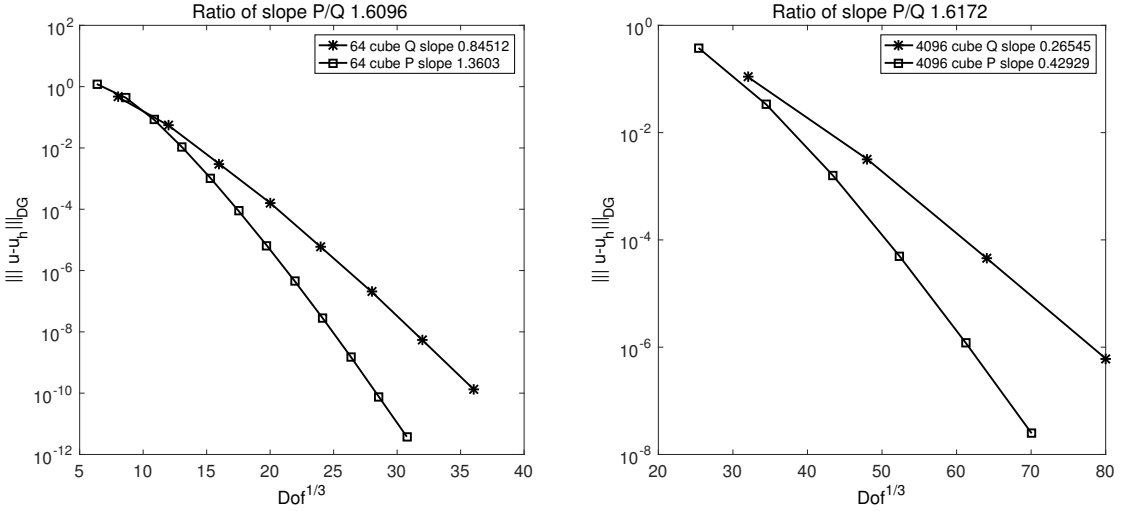


FIGURE 7.5: Example 4: Convergence of the DGFEM under p-refinement. Cube meshes with 64 elements (left) and 4096 elements (right).

7.3.5 Example 5

In the last example, we consider the biharmonic problem

$$\Delta^2 u = f,$$

over the domain $\Omega = (0,1)^3$, the analytic solution is $u = \sin(\pi x)\sin(\pi y)\sin(\pi z)$. Although our analysis does not cover the biharmonic problem, we follow the IP DGFEMs defined in [176, 146, 107] to approximate its solution, starting with $p = 2$.

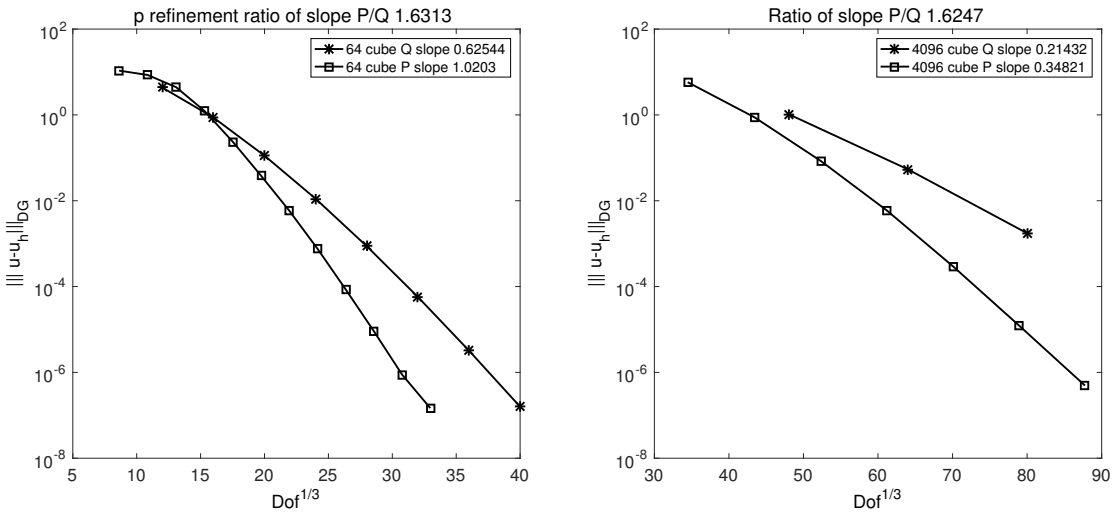


FIGURE 7.6: Example 5: Convergence of the DGFEM under p-refinement. Cube meshes with 64 elements (left) and 4096 elements (right).

We need to emphasize that for biharmonic problems, the minimum polynomial order is 2.

In Figure 7.6, we observe the same behaviour as the previous example with slope ratio 1.62.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

In this work, we presented an *hp*-version interior penalty discontinuous Galerkin finite element method on extremely general classes of meshes consisting of polytopic elements, possibly with arbitrarily small $(d-k)$-dimensional faces, $k = 1, \ldots, d-1$. We applied the proposed DGFEMs to solve partial differential equations with non-negative characteristic form, and with mixed Dirichlet and Neumann boundary conditions. Furthermore, we presented the space-time DGFEMs for solving time-dependent parabolic problems over prismatic meshes as a particular application. The main goal in this work was to derive the *hp*-error bound for DGFEMs over polytopic meshes. For this purpose, new *hp*-version inverse estimate and polynomial approximation results over polytopic elements have been derived. These results are sharp with respect to $(d-k)$-dimensional face degeneration, which has been a key aim of this work.

We presented detailed stability and a priori error analysis for DGFEMs over polytopic elements under two different types of mesh assumptions, which allow both shape irregular polytopic meshes with bounded number of faces per element and shape regular polytopic meshes with arbitrary number of faces. Due to lack of *hp*-approximation theory for the L^2-projection over polytopic elements, we presented a new way for deriving the inf-sup stability and a priori error bound for DGFEMs for solving PDEs with non-negative characteristic forms. Moreover, due to the use of the total degree $\mathcal{P}_p$ basis over space-time prismatic elements, new

stability and a priori error estimates for space-time DGFEMs are derived avoiding the space-time tensor product setting.

A series of numerical experiments have been presented which, not only confirm the theoretical results derived in this work, but also demonstrate the efficiency of employing the total degree polynomial space $\mathcal{P}_p$, defined in the physical coordinate system, compared with the tensor-product polynomial space $\mathcal{Q}_p$, mapped from a given reference or canonical frame, under p-refinement.

Furthermore, we also derived new hp-approximation results for the L^2-projector onto the total degree $\mathcal{P}_p$ basis and the H^1 projector onto the serendipity $\mathcal{S}_p$ basis on the tensor product element. The new results show that the extra basis functions in the tensor-product $\mathcal{Q}_p$ basis other than the total-degree $\mathcal{P}_p$ basis and the serendipity $\mathcal{S}_p$ basis do not increase the convergence rate of p for the L^2-projection and H^1-projection error bound in several norms, but instead only reduce the "constant" in the error bound. The above new approximation results may be of independent interest. One interesting application of these new approximation results is in the proof of exponential convergence for DGFEMs with the $\mathcal{P}_p$ basis. We showed that for fixed tensor product elements, DGFEMs with the $\mathcal{P}_p$ basis converges exponentially to the analytical solution under p-refinement. Moreover, the slope of exponential convergence of DGFEMs with the $\mathcal{P}_p$ basis is steeper than the slope of DGFEMs with the $\mathcal{Q}_p$ basis if we measure the error against number of degrees of freedom. The sharpness of these results was confirmed by a series of numerical examples.

8.2 Future Work

In this section, we outline some future directions of research that naturally arise from this work.

8.2.1 Adaptivity

The first and most important topic for future research is the design of adaptive algorithms for the proposed DGFEMs over general polytopic elements based on a posteriori error estimators. DGFEMs are ideally suited for adaptive algorithms,

as they naturally allow hanging nodes and different polynomial basis between different elements.

The fundamental difficulty in deriving a posteriori error estimators for DGFEMs in energy norm is that the DGFEM solution does not live inside the function space on which the PDEs is defined. The first work to successfully derive rigorous a-posteriori error estimators for DGFEMs in energy norm is by Karakashian & Pascal [134]. Therein, the authors introduced a special type of recovery operator for post-processing the DGFEM solution in order to split the DGFEM solution error into conforming and non-conforming components. Adaptive DGFEMs have been developed in the past 15 years, see [36, 108, 166, 188] for h-version a-posteriori error estimator, and see [123, 190] for hp-version.

All the above mentioned works are concerned with standard meshes. For polytopic meshes, energy norm based a-posteriori error estimators can be found in [111]. However, the theory in [111] does not work for general polytopic elements with arbitrary small $(d-k)$-dimensional faces, $k = 1, \ldots, d-1$. The extension of their results to such meshes is still elusive.

8.2.2 Space-time DGFEMs for problems on evolving domains

Modeling of PDEs over evolving domains is both interesting and challenging. Historically, Jamet [127] was the first to propose the discontinuous Galerkin time-stepping method for solving parabolic PDEs on evolving domains. Due to the discontinuity over different time intervals, the problem can be solved with the high order DGFEMs over each space-time slap with good stability. More recently, the DG time-stepping method for an advection-diffusion model defined on moving domains written in the Arbitrary Lagrangian Eulerian (ALE) framework, has been considered in [42, 43].

Space-time DGFEMs over general prismatic meshes have several advantages compared to the classical DG time-stepping schemes. General shaped prismatic elements will offer great flexibility in practical computations. Using general shaped elements on evolving domains will reduce the computational cost for mesh refinement and coarsening. Also, due to the discontinuous nature of space-time DGFEMs, this approximation has great flexibility in choosing basis functions for each element without considering the conformity of the finite element space. This

is very important in view of reducing the complexity of the space-time DGFEMs with high-order polynomial basis, as we saw in Chapter 6.

8.2.3 Other directions for further research

- **DGFEMs for problems with multiple scales.** Many problems of fundamental and practical importance have multiple-scale solutions, e.g. Composite materials, porous media and turbulent transport in high Reynolds number flows. Multicale-DGFEMs has been very popular in recent years, [1, 2, 90, 91]. To develop multiscale DGFEMs over polytopic meshes for solving problems imposed on complicated domains with multiple scales will be a very interesting project.

- **Extension of *hp*-approximation theory.** The approximation result of the L^2-projection onto the total degree space $\mathcal{P}_p$ in L^2–norm can be easily extended to the Jacobi projector onto the $\mathcal{P}_p$ basis in Jacobi weighted Sobolev norms. Following [21, 22, 20], we can study the new *hp*-approximation results with the $\mathcal{P}_p$ basis in Jacobi-weighted Besov spaces. One application of those results is that we can proof the sharp *hp*-optimal bound for serendipity FEMs for PDEs containing singularity in $r^\gamma \log^\nu r$ type, $\gamma \in \mathbb{R}^+$, $\nu \in \mathbb{N}$. Furthermore, the optimal trace estimates for L^2-projection onto the $\mathcal{P}_p$ basis on simplicial elements has been shown in [69, 144]. It remains, however, an open question their results can be extended to tensor product elements.

- ***hp*-FEMs with serendipity basis.** Serendipity FEMs are popular among engineers. Their mathematical development, however, is relatively recent [14, 17]. There is no sharp theory of serendipity *hp*-FEMs. In this work, we derived some new *hp*-approximation theory for H^1-projection with serendipity $\mathcal{S}_p$ basis in two dimensions. The next step would be to extend their results to three dimensions and construct polynomial trace lifting results for serendipity FEMs. Then we can derive, e.g., exponential convergence for serendipity *hp*-FEMs with *hp*-refinement following [160].

Appendix A

Implementation of hp-Version Discontinuous Galerkin Methods on Polytopic Mehses

We present some of the key ingredients/techniques used in the implementation of the proposed hp-version DGFEMs for general advection-diffusion-reaction boundary value problem and time dependent parabolic problem over polytopic meshes and prismatic meshes.

A.1 DGFEMs for boundary value problems over polytopic meshes

A.1.1 Construction of the finite element basis functions on general polygons/polyhedra

The finite element space $S^{\mathbf{p}}_{\mathcal{T}_h}$ may be constructed in a number of different ways. In the case when the computational mesh $\mathcal{T}_h$ consists of standard affine element domains (simplices, parallelograms, etc), standard polynomial bases on reference elements may simply be mapped from the reference frame to the physical element; indeed, this is the standard approach used within most finite element software packages.

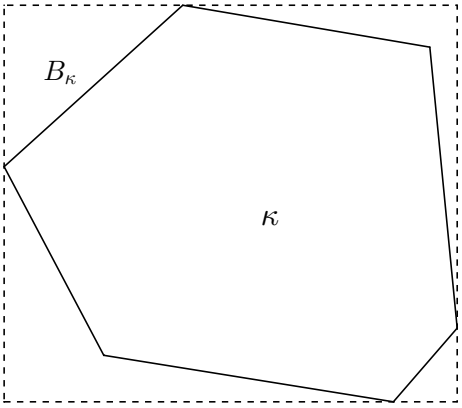


FIGURE A.1: Bounding box B_κ of an element $\kappa \in \mathcal{T}_h$.

Here, we introduce an alternative approach based on employing polynomial spaces defined over the bounding box of each element; cf. [109]. More precisely, given an element $\kappa \in \mathcal{T}_h$, we first construct the Cartesian bounding box B_κ, such that $\bar{\kappa} \subseteq \bar{B}_\kappa$, cf. Figure A.1. On the bounding box B_κ we may define a standard polynomial space $\mathcal{P}_{p_\kappa}(B_\kappa)$ spanned by a set of basis functions $\{\phi_{i,\kappa}\}$, $i = 1, \ldots, \dim(\mathcal{P}_{p_\kappa}(B_\kappa))$. With this in mind, we employ tensor-product (scaled) orthonormal Legendre polynomials; indeed, writing $\hat{I} = (-1, 1)$, we denote the family of $L^2(\hat{I})$-orthonormal (Legendre) polynomials by $\{\tilde{L}_i(x)\}_{i=0}^{\infty}$. Thereby, given a general interval $I_b = (x_1, x_2)$, the corresponding scaled Legendre polynomials may be defined by

$$L_i^{[b]}(x) = (1/h_b)^{1/2} \tilde{L}_i((x - m_b)/h_b),$$

such that

$$\int_{I_b} L_i^{[b]}(x) L_j^{[b]}(x)\, \mathrm{d}x = \delta_{ij},$$

where $h_b = (x_2 - x_1)/2$, $m_b = (x_1 + x_2)/2$ and δ_{ij} is the Kronecker symbol. With this notation, a polynomial basis on B_κ may be defined as follows: writing $B_\kappa = I_1 \times I_2 \times \cdots \times I_d$, where I_j, $j = 1, \ldots, d$, denotes a one–dimensional interval, the space of polynomials $\mathcal{P}_{p_\kappa}(B_\kappa)$ of total degree p_k over B_κ is given by

$$\mathcal{P}_{p_\kappa}(B_\kappa) = \mathrm{span}\{\phi_{i,\kappa}\}_{i=1}^{\dim(\mathcal{P}_{p_\kappa}(B_\kappa))},$$

where

$$\phi_{i,\kappa}(\mathbf{x}) = L_{i_1}^{[1]}(x_1) L_{i_2}^{[2]}(x_2) \cdots L_{i_d}^{[d]}(x_d), \qquad i_1 + i_2 + \ldots + i_d \leq p_\kappa,\ i_k \geq 0,\ k = 1, \ldots, d,$$

and $\mathbf{x} = (x_1, x_2, \ldots, x_d)$. The polynomial basis over the general polygonal/polyhedral element κ may be defined by simply restricting the support of $\{\phi_{i,\kappa}\}$,

$i = 1, \ldots, \dim(\mathcal{P}_{p_\kappa}(B_\kappa))$ to κ; i.e., the polynomial basis defined over κ is given by $\{\phi_{i,\kappa}|_\kappa\}$, $i = 1, \ldots, \dim(\mathcal{P}_{p_\kappa}(B_\kappa))$.

Remark A.1. Notice that, if the underlying polytopic elements are axisparallel tensor product elements, then the resulting mass matrix $\mathcal{M}$ is the identity matrix due to the orthogonality of the basis functions. Moreover, we point out that the basis functions constructed based on bounding box can be defined via diagonal affine transfer from tensor product reference element.

A.1.2 Quadrature rules for polytopic meshes

Following [129], quadrature over general polygonal/polyhedral element domains is undertaken based on first constructing a sub-triangulation, followed by the exploitation of integration schemes introduced in Section A.2. Therefore, given $\kappa \in \mathcal{T}_h$, we first construct a non-overlapping sub-triangulation $\kappa_\mathcal{S} = \{\tau_\kappa\}$ consisting of simplicial elements. As an example, if we consider the local stiffness matrix, restricted to κ, then we compute

$$\begin{aligned}\int_\kappa \nabla u \cdot \nabla v \, \mathrm{d}\mathbf{x} &= \sum_{\tau_\kappa \in \kappa_\mathcal{S}} \int_{\tau_\kappa} \nabla u \cdot \nabla v \, \mathrm{d}\mathbf{x} \\ &\approx \sum_{\tau_\kappa \in \kappa_\mathcal{S}} \sum_{i=1}^{q} \nabla u(F_\kappa(\boldsymbol{\xi}_i)) \cdot \nabla v(F_\kappa(\boldsymbol{\xi}_i)) \det(J_{F_\kappa}(\boldsymbol{\xi}_i)) w_i,\end{aligned}$$

where $F_\kappa : \hat{\kappa} \to \tau_\kappa$ is the mapping from the reference element (simplex) $\hat{\kappa}$ to τ_κ, with Jacobi matrix J_{F_κ}, and $(\boldsymbol{\xi}_i, w_i)_{i=1}^q$ denotes the quadrature rule defined on $\hat{\kappa}$. We point out that the gradient operators are not transformed, as would be the case if the element κ was mapped to a reference frame.

We point out that alternative integration methods which do not require a sub-triangulation of the underlying polygonal/polyhedral element have recently been considered in [148, 37, 40]. For related work we refer to [31, 143] and the references cited therein.

A.2 Quadrature rules over simplices/polytopes

In this section, we give the theoretical and practical background for constructing stable Gauss quadrature rules over a d-dimensional simplex.

A.2.1 Gauss-Jacobi quadrature rules in 1D

We will review some well-known results for the Gauss-Jacobi quadrature rules, following [120, 135, 102, 178].

The classical Jacobi polynomial $P_n^{(\alpha,\beta)}(x)$ of order n is the solution of the singular Sturm-Liouvile eigenvalue problem

$$\frac{d}{dx}(1-x^2)\omega(x)\frac{d}{dx}P_n^{(\alpha,\beta)}(x) + n(n+\alpha+\beta+1)\omega(x)P_n^{(\alpha,\beta)}(x) = 0, \tag{A.1}$$

for $x \in [-1,1]$, with weight function $\omega(x) = (1-x)^\alpha(1+x)^\beta$, for $\alpha, \beta > -1$. The Jacobi polynomials are normalized to be orthonormal:

$$\int_{-1}^{1} \tilde{P}_i^{(\alpha,\beta)}(x)\tilde{P}_j^{(\alpha,\beta)}(x)\omega(x)\,\mathrm{d}x = \delta_{ij}. \tag{A.2}$$

An important property of Jacobi polynomials is [178]:

$$\frac{d}{dx}\tilde{P}_n^{(\alpha,\beta)}(x) = \sqrt{n(n+\alpha+\beta+1)}\tilde{P}_{n-1}^{(\alpha+1,\beta+1)}(x). \tag{A.3}$$

The special case of $\tilde{P}_n^{(0,0)}(x)$, are as the Legendre polynomials $\hat{L}_n(x)$.

A classical way to evaluate the Jacobi polynomials is to use the recurrence relation

$$x\tilde{P}_n^{(\alpha,\beta)}(x) = a_n\tilde{P}_{n-1}^{(\alpha,\beta)}(x) + b_n\tilde{P}_n^{(\alpha,\beta)}(x) + a_{n+1}\tilde{P}_{n+1}^{(\alpha,\beta)}(x), \tag{A.4}$$

where the coefficients are given as

$$a_n = \frac{2}{2n+\alpha+\beta}\sqrt{\frac{n(n+\alpha+\beta)(n+\alpha)(n+\beta)}{(2n+\alpha+\beta-1)(2n+\alpha+\beta+1)}},$$

$$b_n = -\frac{\alpha^2-\beta^2}{(2n+\alpha+\beta)(2n+\alpha+\beta+2)}.$$

To get the recurrence started, we need the initial values

$$\tilde{P}_0^{(\alpha,\beta)}(x) = \sqrt{2^{-1-\alpha-\beta}\frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}},$$

$$\tilde{P}_1^{(\alpha,\beta)}(x) = \frac{1}{2}\tilde{P}_0^{(\alpha,\beta)}(x)\sqrt{\frac{(\alpha+\beta+3)}{(\alpha+1)(\beta+1)}}((\alpha+\beta+2)x + (\alpha-\beta)).$$

There is a close connection between Jacobi polynomials and Gaussian quadratures for the approximation of integrals in the form

$$\int_{-1}^{i} f(x)\omega(x)\,\mathrm{d}x = \sum_{i=0}^{N} f(x_i)\omega_i.$$

Here, (x_i, ω_i) are the quadrature nodes and weights, and $f(x)$ is a polynomial function. It can be shown that if one chooses x_i as the roots of $\tilde{P}_{N+1}^{(\alpha,\beta)}(x)$ and the weights, ω_i, by requiring the integration to be exact for polynomials up to order N, the above summation is in fact exact for f being a polynomial of order $2N+1$. *A key feature of Gauss quadrature rule is that all the quadrature weights are strictly positive, which guarantees the stability of the quadrature, see [66].*

Finding the nodes and weights can be done in several ways. One classical and numerically stable way is based on recurrence, via (A.4). Starting from the three term recurrence, the quadrature rule may be generated by computing the eigenvalues and first component of the orthornormalized eigenvectors of a symmetric tridiagonal matrix; this is the celebrated Golub-Welsch (GW) algorithm, for the details, we refer to [112].

In general, the GW algorithm takes $\mathcal{O}(n^2)$ operations to solve the eigenvalue problem by taking advantage of the structure of the matrix and noting that only the first component of the normalized eigenvector needs to be computed. Moreover, it has been observed in [117] that the GW method leads to an $\mathcal{O}(n)$ error in the Gauss-Legendre nodes and an $\mathcal{O}(n^{3/2})$ error for the relative maximum error in the weights. Here, we will use the alternative approach proposed in [117] which is to solve (A.4) by Newton iterates. It is shown that Newton iterates converge to the zeros of the orthogonal polynomial with total complexity $\mathcal{O}(n^2)$, see [117] for details.

A.2.2 Quadrature rules over triangles

In this section, we present more details about stable and efficient computation of quadratures over a triangle. As illustrated in Figure A.2, the reference square $\mathcal{Q}^2$ and the reference triangle $\mathcal{T}^2$ in Cartesian coordinates (η_1, η_2) and (ξ_1, ξ_2) are represented as:

$$\mathcal{Q}^2 = \{(\eta_1, \eta_2) | -1 \leq \eta_1, \eta_2 \leq 1\},$$

and

$$\mathcal{T}^2 = \{(\xi_1, \xi_2)| -1 \leq \xi_1, \xi_2, \quad \xi_1 + \xi_2 \leq 0\}.$$

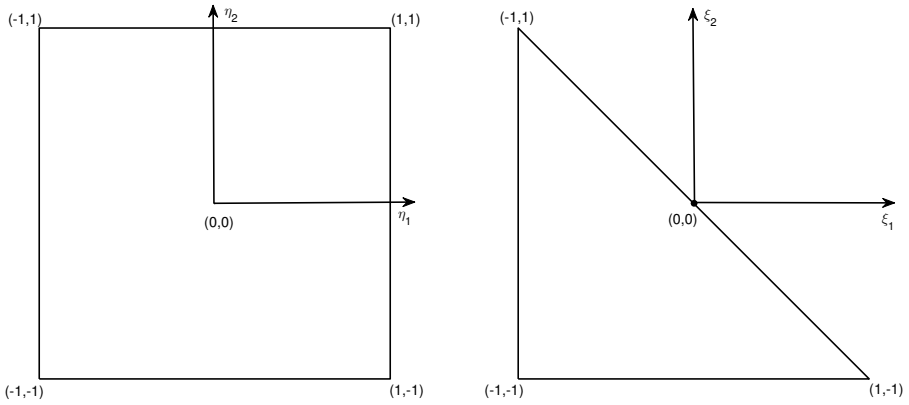


FIGURE A.2: Reference square $\mathcal{Q}^2$ (left); reference triangle $\mathcal{T}^2$ (right).

In order to apply Gauss-quadrature rules over each triangle, we link $\mathcal{T}^2$ and $\mathcal{Q}^2$. The key technique is to use the *Duffy transformation (collapsed transformation)* [135] to link two different coordinate system (η_1, η_2) and (ξ_1, ξ_2) together. The transformation is defined as follows:

$$\eta_1 = 2\frac{1+\xi_1}{1-\xi_2} - 1, \qquad \eta_2 = \xi_2, \tag{A.5}$$

and has inverse transformation

$$\xi_1 = \frac{(1+\eta_1)(1-\eta_2)}{2} - 1, \qquad \xi_2 = \eta_2. \tag{A.6}$$

We emphasize that *for any polynomial functions $u(\xi_1, \xi_2)$ over the the region $\mathcal{T}^2$, under Duffy transformation, $u(\eta_1, \eta_2)$ is still a polynomial function over $\mathcal{Q}^2$.* If we want to compute the integral of polynomials $u(\xi_1, \xi_2)$ over the region $\mathcal{T}^2$, recalling the collapsed system (η_1, η_2), we obtain

$$\begin{aligned} \int_{\mathcal{T}^2} u(\xi_1, \xi_2)\, \mathrm{d}\xi_1\, \mathrm{d}\xi_2 &= \int_{-1}^{1}\int_{-1}^{-\xi_2} u(\xi_1, \xi_2)\, \mathrm{d}\xi_1\, \mathrm{d}\xi_2 \\ &= \int_{-1}^{1}\int_{-1}^{1} u(\eta_1, \eta_2) \left| \frac{\partial(\xi_1, \xi_2)}{\partial(\eta_1, \eta_2)} \right| \mathrm{d}\eta_1\, \mathrm{d}\eta_2, \end{aligned} \tag{A.7}$$

where $\partial(\xi_1, \xi_2)/\partial(\eta_1, \eta_2)$ is the Jacobian of the Cartestian to the local coordinate transformation and can be expressed in terms of η_2 by

$$\frac{\partial(\xi_1, \xi_2)}{\partial(\eta_1, \eta_2)} = \frac{1 - \eta_2}{2}.$$

The last term in (A.7) can be approximated using one-dimensional Gaussian quadrature rules to arrive at

$$\int_{-1}^{1}\int_{-1}^{1} u(\eta_1, \eta_2)\frac{1-\eta_2}{2}\,\mathrm{d}\eta_1\,\mathrm{d}\eta_2 = \sum_{i=0}^{Q_1-1}\omega_i\left\{\sum_{j=0}^{Q_2-1}\omega_j u(\eta_{1i}, \eta_{2j})\frac{1-\eta_{2j}}{2}\right\}, \tag{A.8}$$

where η_{1i} and η_{2j} are quadrature points in the η_1 and η_2 directions, respectively. The weights ω_i used in (A.8) correspond to the standard Gauss-Legendre rule. However, if we take into account that the Jacobian term $\partial(\xi_1, \xi_2)/\partial(\eta_1, \eta_2) = (1 - \eta_2)/2$, which is a singular function appearing inside the general Jacobi polynomial, then we can use Gauss-Jacobi quadrature with $\alpha = 1$ and $\beta = 0$ along η_2 direction. Accordingly, the integration scheme over $\mathcal{T}^2$ becomes

$$\int_{-1}^{1}\int_{-1}^{1} u(\eta_1, \eta_2)\frac{1-\eta_2}{2}\,\mathrm{d}\eta_1\,\mathrm{d}\eta_2 = \sum_{i=0}^{Q_1-1}\omega_i^{0,0}\left\{\sum_{j=0}^{Q_2-1}\hat{\omega}_j^{1,0} u(\eta_{1i}^{0,0}, \eta_{2j}^{1,0})\right\}, \tag{A.9}$$

and

$$\hat{\omega}_j^{1,0} = \frac{\omega_j^{1,0}}{2}, \tag{A.10}$$

where $\omega_j^{1,0}$ and $\eta_{2j}^{1,0}$ are the weights and nodes of the Gauss-Jacobi quadrature with weight $\alpha = 1$ and $\beta = 0$, respectively. The Gauss-Jacobi rule therefore uses fewer quadrature points than the standard Gauss-Legendre quadrature rule to achieve an equivalent accuracy.

Here, as we mentioned in Section A.1.2, the quadratures rules are applied on physical polytopic meshes, so we need to transform the above Gauss-Jacobi quadrature points into the physical meshes.

First we can generate quadrature points on a reference square $\mathcal{Q}^2$. Next, by using the inverse mapping of Duffy transformation, we find the corresponding quadrature points on the reference triangle $\mathcal{T}^2$. The weight function is invariant; see Figure A.3 for an illustration.

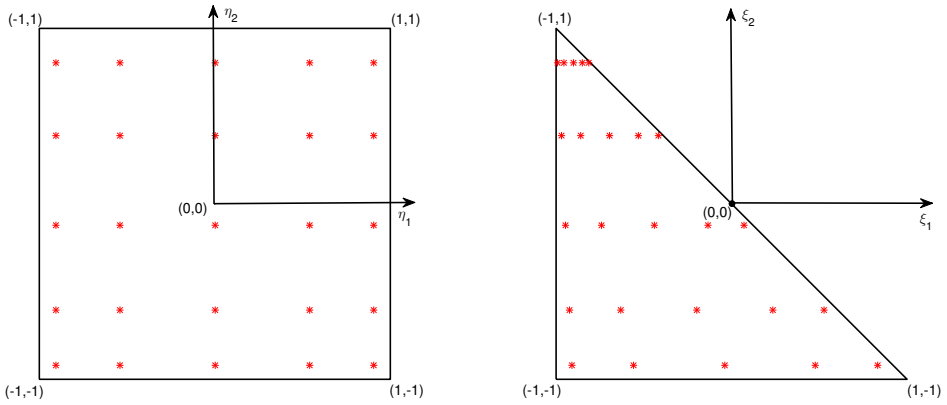


FIGURE A.3: Quadrature points over $\mathcal{Q}^2$ with Gauss-Legendre points along η_1 and Gauss-Jacobi points ($\alpha = 1$ and $\beta = 0$) along η_2 (left); quadrature points over $\mathcal{T}^2$ after transformation (right).

Finally, we use the affine map to transform all the quadrature points from the reference triangle to the physical triangles and glue them up to get the quadrature points over the polygons. During the affine mapping, the weight of quadrature points will change; see Figure A.4 for an illustration.

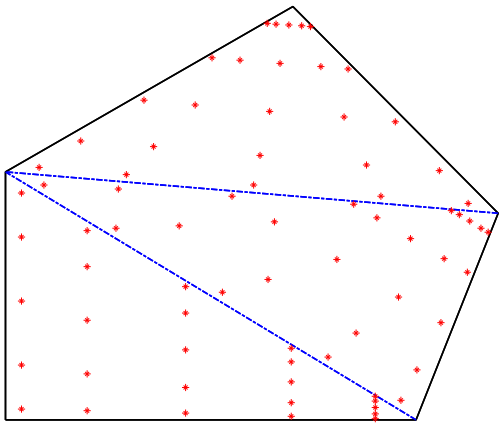


FIGURE A.4: Quadrature points for polygons

A.2.3 Quadrature rules over tetrahedra

In this section, we present quadratures over tetrahedra based on the same technique used in previous sections. As illustrated in Figure A.5, the reference cube $\mathcal{Q}^3$ and the reference triangle $\mathcal{T}^3$ in Cartesian coordinates (η_1, η_2, η_3) and (ξ_1, ξ_2, ξ_3)

are represented as:

$$\mathcal{Q}^3 = \{(\eta_1, \eta_2, \eta_3) | -1 \leq \eta_1, \eta_2, \eta_3 \leq 1\},$$

and

$$\mathcal{T}^3 = \{(\xi_1, \xi_2, \xi_3) | -1 \leq \xi_1, \xi_2, \xi_3, \quad \xi_1 + \xi_2 + \xi_3 \leq 0\}.$$

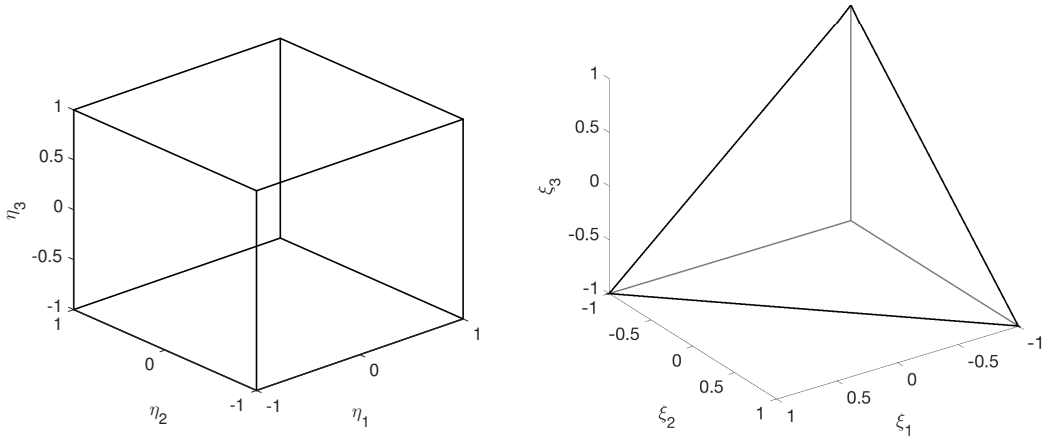


FIGURE A.5: Reference square $\mathcal{Q}^3$ (left); reference tetrahedron $\mathcal{T}^3$(right).

We introduce the Duffy transformation to link the two different coordinate systems (η_1, η_2, η_3) and (ξ_1, ξ_2, ξ_3). The transformation is defined as:

$$\eta_1 = \frac{2(1+\xi_1)}{-\xi_2 - \xi_3} - 1, \qquad \eta_2 = \frac{2(1+\xi_2)}{1-\xi_3} - 1, \qquad \eta_3 = \xi_3, \tag{A.11}$$

and has the inverse transformation

$$\xi_1 = \frac{(1+\eta_1)(1-\eta_2)(1-\eta_3)}{4} - 1, \qquad \xi_2 = \frac{(1+\eta_2)(1-\eta_3)}{2} - 1, \qquad \xi_3 = \eta_3. \tag{A.12}$$

If we want to compute integral of polynomials $u(\xi_1, \xi_2, \xi_3)$ over the region $\mathcal{T}^3$, recalling the collapsed system (η_1, η_2, η_2), we obtain

$$\int_{\mathcal{T}^3} u(\xi_1, \xi_2, \xi_3)\, \mathrm{d}\xi_1\, \mathrm{d}\xi_2\, \mathrm{d}\xi_3 \;=\; \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} u(\eta_1, \eta_2, \eta_3) J\, \mathrm{d}\eta_1\, \mathrm{d}\eta_2\, \mathrm{d}\eta_3, \tag{A.13}$$

where

$$J = \frac{\partial(\xi_1, \xi_2, \xi_3)}{\partial(\eta_1, \eta_2, \eta_3)} = \frac{1-\eta_2}{2}\Big(\frac{1-\eta_3}{2}\Big)^2.$$

We can include the Jacobian in the quadrature weights by using the Gauss-Jacobi integration rules with $\alpha = 0$, $\beta = 0$ along the η_1 direction (i.e., Gauss-Legendre quadrature), $\alpha = 1$, $\beta = 0$ along the η_2 direction, and $\alpha = 2$, $\beta = 0$ along the η_3

direction. The integration rule over $\mathcal{T}^3$ then becomes

$$\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} u(\eta_1,\eta_2,\eta_3)\frac{1-\eta_2}{2}\Big(\frac{1-\eta_3}{2}\Big)^2 \,\mathrm{d}\eta_1\,\mathrm{d}\eta_2\,\mathrm{d}\eta_3$$
$$= \sum_{i=0}^{Q_1-1}\sum_{i=0}^{Q_2-1}\sum_{i=0}^{Q_3-1} u(\eta_{1i}^{0,0},\eta_{2j}^{1,0},\eta_{3j}^{2,0})\omega_k^{0,0}\hat{\omega}_j^{1,0}\hat{\omega}_k^{2,0}, \tag{A.14}$$

$$\hat{\omega}_j^{1,0} = \frac{\omega_j^{1,0}}{2}, \qquad \hat{\omega}_j^{2,0} = \frac{\omega_j^{2,0}}{4}, \tag{A.15}$$

and Q_1, Q_2 and Q_3 are the number of quadrature points in the η_1, η_2 and η_3 directions, respectively.

During implementation, we generate the quadrature points on reference cube $\mathcal{Q}^3$ and by using the inverse mapping of Duffy transformation, we find the corresponding quadrature points on the reference tetrahedron $\mathcal{T}^3$, keeping the weight function fixed; see Figure A.3 for an illustration. General polyhedra can be sub-triangulated into finite number of tetrahedra, so we can use affine mapping to link the quadrature points from the reference tetrahedron to physical polyhedron.

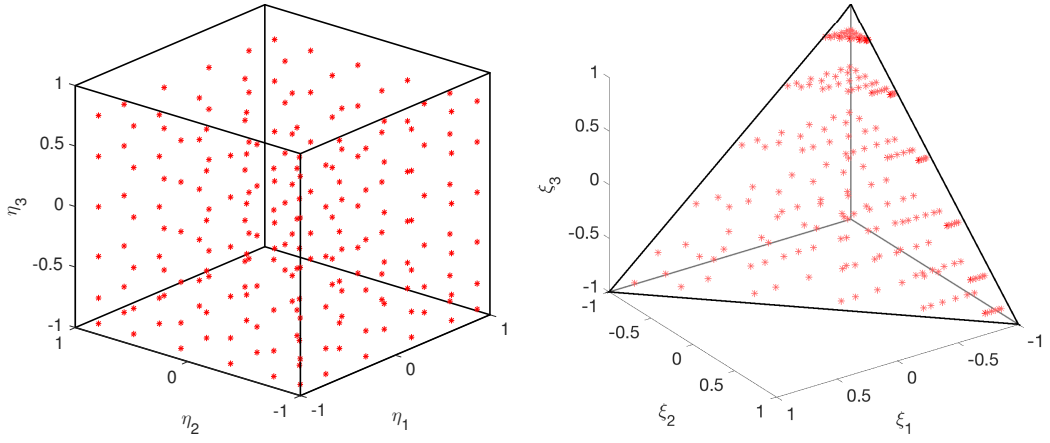


FIGURE A.6: Quadrature points over $\mathcal{Q}^3$ with Gauss-Legendre points along η_1, Gauss-Jacobi points ($\alpha = 1$ and $\beta = 0$) along η_2 and Gauss-Jacobi points ($\alpha = 2$ and $\beta = 0$) along η_3 (left); quadrature points over $\mathcal{T}^3$ after transformation (right).

We emphasize that the above quadrature rules over the simplex may use more quadrature points to deal with polynomials with total degrees basis ($\mathcal{P}_p$-type) compared with the quadrature rules in [170, 89]. The latter quadrature rules, however do not take advantage of tensorial construction of the unstructured basis.

Additionally, the order of these schemes also tends to be restricted by the numerical process of evaluating the quadrature weights. Furthermore, in three dimension, there exist quadrature points outside the underlying tetrahedron with negative weights. *Negative quadrature weights may cause numerical instability in practical computation for large classes of functions, see example 4.1.2 in [70] (page* 202*).* On the other hand, the proposed Gauss quadrature rules do not suffer from these drawbacks.

We point out that alternative integration methods which do not require a sub-triangulation of the underlying polygonal/polyhedral element have recently been considered in [148, 37, 40]. For related work, we refer to [31, 143], and the references cited therein.

A.3 DGFEMs for parabolic problems over prismatic meshes

A.3.1 Construction of finite element basis functions on prismatic meshes

The key point for constructing the space-time basis over the prismatic meshes is to utilize the tensor product structure of a space-time element. The basis functions can be constructed in a similar way as in Section A.1. The spatial basis functions are still constructed based on the bounding box B_κ for κ, and the temporal basis is constructed based on the temporal interval $I_n \subset \mathbb{R}$. We introduce the space-time bounding box B_{κ_n} for each prismatic mesh κ_n; see Figure A.7 for an illustration.

A.3.2 Quadrature rules for prismatic meshes

The quadrature rules over prismatic meshes can be constructed by exploiting their tensor product structure. Assuming that the spatial dimension $d = 2$, we will first use the quadrature rules to get all the quadrature points over the spatial element κ, and then we use the one dimensional Gauss-Legendre quadrature rules along temporal interval I_n. Finally, the quadrature points over the space-time element k_n is constructed by tensor product argument; see Figure A.8 for details.

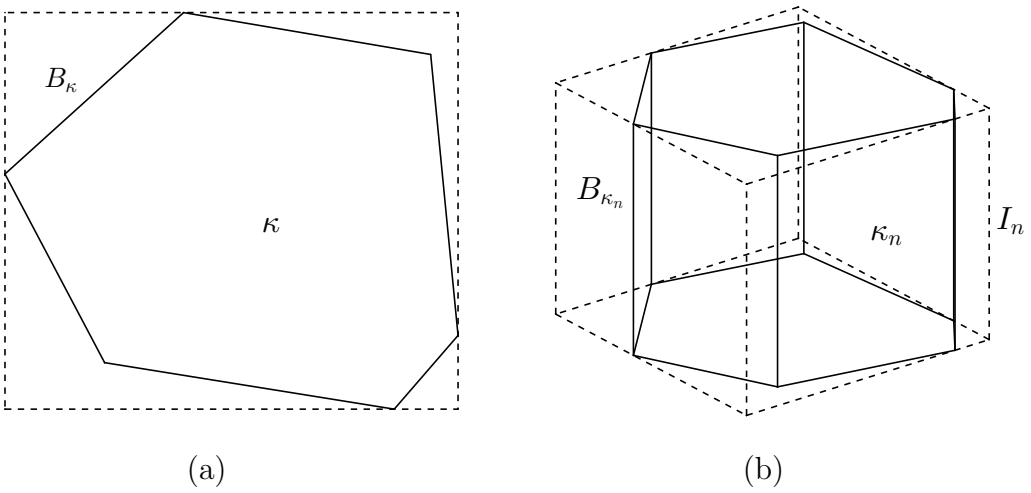


FIGURE A.7: (a). Polygonal spatial element κ and bounding box B_κ; (b). space-time element $\kappa_n = I_n \times \kappa$ and space-time bounding box $B_{\kappa_n} := I_n \times B_\kappa$.

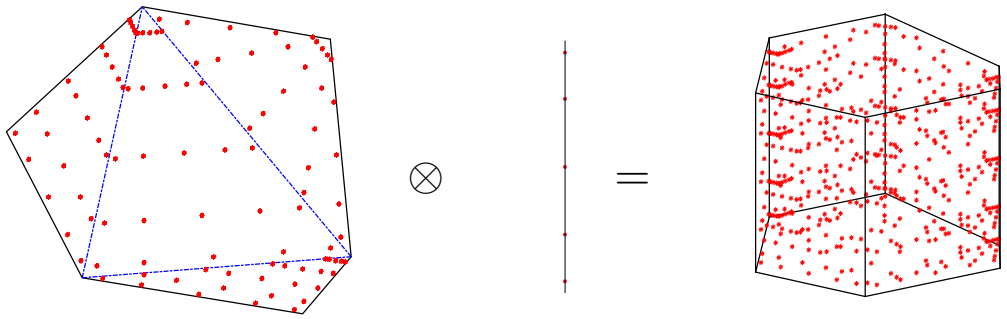


FIGURE A.8: Quadrature points over the space-time element κ_n, with 2D spatial element κ.

Bibliography

[1] A. Abdulle, *Discontinuous Galerkin finite element heterogeneous multiscale method for elliptic problems with multiple scales*, Math. Comp., 81 (2012), pp. 687–713.

[2] A. Abdulle and M. E. Huber, *Discontinuous Galerkin finite element heterogeneous multiscale method for advection–diffusion problems with multiple scales*, Numer. Math., 126 (2014), pp. 589–633.

[3] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55 of National Bureau of Standards Applied Mathematics Series, For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[4] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, vol. 140 of Pure and Applied Mathematics (Amsterdam), Elsevier/Academic Press, Amsterdam, second ed., 2003.

[5] P. Antonietti and B. Ayuso, *Schwarz domain decomposition preconditioners for discontinuous Galerkin approximations of elliptic problems: non-overlapping case*, M2AN Math. Model. Numer. Anal., 41 (2007), pp. 21–54.

[6] ——, *Multiplicative Schwarz methods for discontinuous Galerkin approximations of elliptic problems*, M2AN Math. Model. Numer. Anal., 42 (2008), pp. 443–469.

[7] P. Antonietti, A. Cangiani, J. Collis, Z. Dong, E. Georgoulis, S. Giani, and P. Houston, *Review of discontinuous Galerkin finite element methods for partial differential equations on complicated domains*, Building Bridges: Connections and Challenges in Modern Approaches to Numerical Partial Differential Equations. Lecture Notes in Computational Science and Engineering, Springer Verlag, (2016).

[8] P. Antonietti, S. Giani, and P. Houston, *hp–Version composite discontinuous Galerkin methods for elliptic problems on complicated domains*, SIAM J. Sci. Comput., 35 (2013), pp. A1417–A1439.

[9] P. Antonietti, S. Giani, and P. Houston, *Domain decomposition preconditioners for Discontinuous Galerkin methods for elliptic problems on complicated domains*, J. Sci. Comput., 60 (2014), pp. 203–227.

[10] P. Antonietti and P. Houston, *A class of domain decomposition preconditioners for hp-discontinuous Galerkin finite element methods*, J. Sci. Comp., 46 (2011), pp. 124–149.

[11] P. Antonietti, P. Houston, M. Sarti, and M. Verani, *Multigrid algorithms for hp–version interior penalty discontinuous Galerkin methods on polygonal and polyhedral meshes*, arXiv preprint arXiv:1412.0913, (2014).

[12] P. Antonietti, M. Sarti, and M. Verani, *Multigrid algorithms for hp-Discontinuous Galerkin discretizations of elliptic problems*, SIAM J. Numer. Anal., 53 (2015), pp. 598–618.

[13] D. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal., 19 (1982), pp. 742–760.

[14] D. Arnold, D. Boffi, and R. Falk, *Approximation by quadrilateral finite elements*, Math. Comp., 71 (2002), pp. 909–922.

[15] D. Arnold, D. Boffi, R. Falk, and L. Gastaldi, *Finite element approximation on quadrilateral meshes*, Commun. Numer. Meth. Engrg., 17 (2001), pp. 805–812.

[16] D. Arnold, F. Brezzi, B. Cockburn, and L. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., 39 (2001), pp. 1749–1779.

[17] D. N. Arnold and G. Awanou, *The serendipity family of finite elements*, Found. Comput. Math., 11 (2011), pp. 337–344.

[18] B. Ayuso and L. Marini, *Discontinuous Galerkin methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 47 (2009), pp. 1391–1420.

[19] I. Babuška, C. E. Baumann, and J. T. Oden, *A discontinuous hp finite element method for diffusion problems: 1-D analysis*, Comput. Math. Appl., 37 (1999), pp. 103–122.

[20] I. Babuška and B. Guo, *Optimal estimates for lower and upper bounds of approximation errors in the p-version of the finite element method in two dimensions*, Numer. Math., 85 (2000), pp. 219–255.

[21] I. Babuska and B. Guo, *Direct and inverse approximation theorems for the p-version of the finite element method in the framework of weighted besov spaces. part I: Approximability of functions in the weighted besov spaces*, SIAM J. Numer. Anal., 39 (2002), pp. 1512–1538.

[22] I. Babuška and B. Guo, *Direct and inverse approximation theorems for the p-version of the finite element method in the framework of weighted besov spaces part II: Optimal rate of convergence of the p-version finite element solutions*, Math. Models Methods Appl. Sci., 12 (2002), pp. 689–719.

[23] I. Babuška and J. E. Osborn, *Generalized finite element methods: their performance and their relation to mixed methods*, SIAM J. Numer. Anal., 20 (1983), pp. 510–536.

[24] I. Babuška and M. Suri, *The h-p version of the finite element method with quasi-uniform meshes*, RAIRO Modél. Math. Anal. Numér., 21 (1987), pp. 199–238.

[25] ——, *The optimal convergence rate of the p-version of the finite element method*, SIAM J. Numer. Anal., 24 (1987), pp. 750–776.

[26] I. Babuška, *The finite element method with penalty*, Math. Comp., 27 (1973), pp. 221–228.

[27] G. Baker, *Finite element methods for elliptic equations using nonconforming elements*, Math. Comp., 31 (1977), pp. 45–59.

[28] G. A. Baker, W. N. Jureidini, and O. A. Karakashian, *Piecewise solenoidal vector fields and the stokes problem*, SIAM J. Numer. Anal., 27 (1990), pp. 1466–1485.

[29] R. Bass, *Diffusion and Elliptic Operators*, Spinger–Verlag, New York, 1997.

[30] F. Bassi, L. Botti, and A. Colombo, *Agglomeration-based physical frame dG discretizations: An attempt to be mesh free*, Math. Models Methods Appl. Sci., 24 (2014), pp. 1495–1539.

[31] F. Bassi, L. Botti, A. Colombo, D. Di Pietro, and P. Tesini, *On the flexibility of agglomeration based physical space discontinuous Galerkin discretizations*, J. Comput. Phys., 231 (2012), pp. 45–65.

[32] F. Bassi, L. Botti, A. Colombo, and S. Rebay, *Agglomeration based discontinuous Galerkin discretization of the Euler and Navier-Stokes equations*, Comput. & Fluids, 61 (2012), pp. 77–85.

[33] F. Bassi and S. Rebay, *A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations*, J. Comput. Phys., 131 (1997), pp. 267–279.

[34] F. Bassi and S. Rebay, *Gmres discontinuous Galerkin solution of the compressible navier-stokes equations*, in Discontinuous Galerkin Methods, Springer, 2000, pp. 197–208.

[35] P. Bastian and C. Engwer, *An unfitted finite element method using discontinuous Galerkin*, Internat. J. Numer. Methods Engrg., 79 (2009), pp. 1557–1576.

[36] R. Becker, P. Hansbo, and M. Larson, *Energy norm a posteriori error estimation for discontinuous Galerkin methods*, Comput. Methods Appl. Mech. Engrg., 192 (2003), pp. 723–733.

[37] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. Marini, and A. Russo, *Basic principles of virtual element methods*, Math. Models Methods Appl. Sci., 23 (2013), pp. 199–214.

[38] L. Beirão da Veiga, F. Brezzi, L. Marini, and A. Russo, *The hitchhiker's guide to the virtual element method*, Math. Models Methods Appl. Sci., 24 (2014), pp. 1541–1573.

[39] L. Beirão da Veiga, J. Droniou, and G. Manzini, *A unified approach for handling convection terms in finite volumes and mimetic discretization methods for elliptic problems*, IMA J. Numer. Anal., 31 (2011), pp. 1357–1401.

[40] L. Beirão da Veiga and G. Manzini, *A virtual element method with arbitrary regularity*, IMA J. Numer. Anal., in press (2013).

[41] K. Bey and T. Oden, *hp-version discontinuous Galerkin methods for hyperbolic conservation laws*, Comput. Methods Appl. Mech. Engrg., 133 (1996), pp. 259–286.

[42] A. Bonito, I. Kyza, and R. H. Nochetto, *Time-discrete higher order ale formulations: A priori error analysis*, Numer. Math., 125 (2013), pp. 225–257.

[43] ——, *Time-discrete higher-order ale formulations: Stability*, SIAM J. Numer. Anal., 51 (2013), pp. 577–604.

[44] S. Brenner, J. Cui, and L.-Y. Sung, *Multigrid methods for the symmetric interior penalty method on graded meshes*, Numer. Linear Algebra Appl., 16 (2009), pp. 481–501.

[45] S. Brenner and J. Zhao, *Convergence of multigrid algorithms for interior penalty methods*, Appl. Numer. Anal. Comput. Math., 2 (2005), pp. 3–18.

[46] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.

[47] F. Brezzi, A. Buffa, and K. Lipnikov, *Mimetic finite differences for elliptic problems*, M2AN Math. Model. Numer. Anal., 43 (2009), pp. 277–295.

[48] F. Brezzi, G. Manzini, D. Marini, P. Pietra, and A. Russo, *Discontinuous Galerkin approximations for elliptic problems*, Num. Meth. Part. Diff. Eqs., 16 (2000), pp. 365–378.

[49] A. Buffa, T. Hughes, and G. Sangalli, *Analysis of a multiscale discontinuous Galerkin method for convection-diffusion problems*, SIAM J. Numer. Anal., 44 (2006), pp. 1420–1440.

[50] E. Burman, *A unified analysis for conforming and nonconforming stabilized finite element methods using interior penalty*, SIAM J. Numer. Anal., 43 (2005), pp. 2012–2033 (electronic).

[51] E. Burman, S. Claus, P. Hansbo, M. G. Larson, and A. Massing, *CutFEM: Discretizing geometry and partial differential equations*, Internat. J. Numer. Methods Engrg., 104 (2015), pp. 472–501.

[52] E. Burman and A. Ern, *Continuous interior penalty hp-finite element methods for advection and advection-diffusion equations*, Math. Comp., 76 (2007), p. 1119.

[53] E. Burman and M. A. Fernández, *Continuous interior penalty finite element method for the time-dependent navier–stokes equations: space discretization and convergence*, Numer. Math., 107 (2007), pp. 39–77.

[54] E. Burman, M. A. Fernández, and P. Hansbo, *Continuous interior penalty finite element method for oseen's equations*, SIAM J. Numer. Anal., 44 (2006), p. 1248.

[55] E. Burman, P. Hansbo, M. G. Larson, and A. Massing, *A cut discontinuous Galerkin method for the Laplace–Beltrami operator*, IMA J. Numer. Anal., (2016), p. Published online.

[56] E. Burman, P. Hansbo, M. G. Larson, A. Massing, and S. Zahedi, *Full gradient stabilized cut finite element methods for surface partial differential equations*, Comput. Methods Appl. Mech. Engrg., 310 (2016), pp. 278–296.

[57] A. Cangiani, J. Chapman, E. Georgoulis, and M. Jensen, *On the stability of continuous–discontinuous Galerkin methods for advection–diffusion–reaction problems*, J. Sci. Comput., 57 (2013), pp. 313–330.

[58] A. Cangiani, Z. Dong, and E. Georgoulis, *hp-version space-time discontinuous Galerkin methods for parabolic problems on prismatic meshes*, submitted, (2016).

[59] A. Cangiani, Z. Dong, E. Georgoulis, and P. Houston, *hp-version discontinuous Galerkin methods for advection-diffusion-reaction problems on polytopic meshes*, M2AN Math. Model. Numer. Anal., 50 (2016), pp. 699–725.

[60] A. Cangiani, Z. Dong, E. H. Georgoulis, and P. Houston, *hp-Version discontinuous Galerkin methods on polytopic meshes*, Springer.

[61] A. Cangiani, E. Georgoulis, and P. Houston, *hp–version discontinuous Galerkin methods on polygonal and polyhedral meshes*, Math. Models Methods Appl. Sci., 24 (2014), pp. 2009–2041.

[62] A. Cangiani, E. H. Georgoulis, and M. Jensen, *Discontinuous Galerkin methods for mass transfer through semipermeable membranes*, SIAM J. Numer. Anal., 51 (2013), pp. 2911–2934.

[63] A. Cangiani, E. H. Georgoulis, T. Pryer, and O. J. Sutton, *A posteriori error estimates for the virtual element method*, arXiv preprint arXiv:1603.05855, (2016).

[64] A. Cangiani, G. Manzini, and A. Russo, *Convergence analysis of the mimetic finite difference method for elliptic problems*, SIAM J. Numer. Anal., 47 (2009), pp. 2612–2637.

[65] A. Cangiani, G. Manzini, and O. Sutton, *Conforming and nonconforming virtual element methods for elliptic problems*, IMA J. Numer. Anal., Publish online, (2016).

[66] C. Canuto, M. Y. Hussaini, A. M. Quarteroni, and A. Thomas Jr, *Spectral methods in fluid dynamics*, Springer Science & Business Media, 2012.

[67] C. Carstensen and S. A. Funken, *Constants in Clément-interpolation error and residual based a posteriori error estimates in finite element methods*, East-West J. Numer. Math., 8 (2000), pp. 153–175.

[68] F. Chave, D. A. D. Pietro, F. Marche, and F. Pigeonneau, *A hybrid high-order method for the Cahn–Hilliard problem in mixed form*, SIAM J. Numer. Anal., 54 (2016), pp. 1873–1898.

[69] A. Chernov, *Optimal convergence estimates for the trace of the polynomial L^2-projection operator on a simplex*, Math. Comp., 81 (2012), pp. 765–787.

[70] P. Ciarlet, *The finite element method for elliptic problems*, North-Holland Publishing Co., Amsterdam, 1978. Studies in Mathematics and its Applications, Vol. 4.

[71] B. Cockburn, *An introduction to the discontinuous Galerkin method for convection-dominated problems*, in Advanced numerical approximation of nonlinear hyperbolic equations (Cetraro, 1997), Springer, Berlin, 1998, pp. 151–268.

[72] ——, *Discontinuous Galerkin methods for convection-dominated problems*, in High-order methods for computational physics, Springer, Berlin, 1999, pp. 69–224.

[73] B. Cockburn, D. A. Di Pietro, and A. Ern, *Bridging the hybrid high-order and hybridizable discontinuous galerkin methods*, M2AN Math. Model. Numer. Anal., 50 (2016), pp. 635–650.

[74] B. Cockburn, B. Dong, and J. Guzmán, *Optimal convergence of the original DG method for the transport-reaction equation on special meshes*, SIAM J. Numer. Anal., 46 (2008), pp. 1250–1265.

[75] B. Cockburn, B. Dong, J. Guzmán, and J. Qian, *Optimal convergence of the original DG method on special meshes for variable transport velocity*, SIAM J. Numer. Anal., 48 (2010), pp. 133–146.

[76] B. Cockburn, S. Hou, and C.-W. Shu, *The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV. The multidimensional case*, Math. Comp., 54 (1990), pp. 545–581.

[77] B. Cockburn, G. Karniadakis, and C.-W. Shu, eds., *Discontinuous Galerkin methods*, Springer-Verlag, Berlin, 2000. Theory, computation and applications, Papers from the 1st International Symposium held in Newport, RI, May 24–26, 1999.

[78] B. Cockburn, S. Lin, and C.-W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. III. One-dimensional systems*, J. Comput. Phys., 84 (1989), pp. 90–113.

[79] B. Cockburn and C.-W. Shu, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. II. General framework*, Math. Comp., 52 (1989), pp. 411–435.

[80] ——, *The local discontinuous Galerkin method for time-dependent convection-diffusion systems*, SIAM J. Numer. Anal., 35 (1998), pp. 2440–2463 (electronic).

[81] ——, *The Runge-Kutta discontinuous Galerkin method for conservation laws. V. Multidimensional systems*, J. Comput. Phys., 141 (1998), pp. 199–224.

[82] R. Courant et al., *Variational methods for the solution of problems of equilibrium and vibrations*, Bull. Amer. Math. Soc, 49 (1943), pp. 1–23.

[83] P. J. Davis, *Interpolation and approximation*, Courier Corporation, 1975.

[84] D. Di Pietro and A. Ern, *Mathematical aspects of discontinuous Galerkin methods*, vol. 69 of Mathématiques & Applications (Berlin) [Mathematics & Applications], Springer, Heidelberg, 2012.

[85] D. A. Di Pietro and A. Ern, *A hybrid high-order locking-free method for linear elasticity on general meshes*, Comput. Methods Appl. Mech. Engrg., 283 (2015), pp. 1–21.

[86] ——, *Hybrid high-order methods for variable-diffusion problems on general meshes*, Comptes Rendus Mathématique, 353 (2015), pp. 31–34.

[87] D. A. Di Pietro, A. Ern, and J.-L. Guermond, *Discontinuous Galerkin methods for anisotropic semidefinite diffusion with advection*, SIAM J. Numer. Anal., 46 (2008), pp. 805–831.

[88] D. A. Di Pietro, A. Ern, and S. Lemaire, *An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators*, Computational Methods in Applied Mathematics, 14 (2014), pp. 461–472.

[89] D. Dunavant, *High degree efficient symmetrical gaussian quadrature rules for the triangle*, Internat. J. Numer. Methods Engrg., 21 (1985), pp. 1129–1148.

[90] D. Elfverson, E. H. Georgoulis, and A. Målqvist, *An adaptive discontinuous Galerkin multiscale method for elliptic problems*, Multiscale Model. Simul., 11 (2013), pp. 747–765.

[91] D. Elfverson, E. H. Georgoulis, A. Målqvist, and D. Peterseim, *Convergence of a discontinuous Galerkin multiscale method*, SIAM J. Numer. Anal., 51 (2013), pp. 3351–3372.

[92] A. Ern and J.-L. Guermond, *Discontinuous Galerkin methods for Friedrichs' systems. I. General theory*, SIAM J. Numer. Anal., 44 (2006), pp. 753–778.

[93] A. Ern and J.-L. Guermond, *Discontinuous Galerkin methods for Friedrichs' systems. II. Second-order elliptic PDEs*, SIAM J. Numer. Anal., 44 (2006), pp. 2363–2388.

[94] ——, *Discontinuous Galerkin methods for Friedrichs' systems. III. Multifield theories with partial coercivity*, SIAM J. Numer. Anal., 46 (2008), pp. 776–804.

[95] A. Ern, A. F. Stephansen, and P. Zunino, *A discontinuous Galerkin method with weighted averages for advection-diffusion equations with locally small and anisotropic diffusivity*, IMA J. Numer. Anal., 29 (2009), pp. 235–256.

[96] L. C. Evans, *Partial differential equations*, vol. 19 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, second ed., 2010.

[97] R. S. Falk and G. R. Richter, *Local error estimates for a finite element method for hyperbolic and convection-diffusion equations*, SIAM J. Numer. Anal., 29 (1992), pp. 730–754.

[98] W. Feller, *An Introduction to Probability Theory and its Applications: volume I*, vol. 3, John Wiley & Sons London-New York-Sydney-Toronto, 1968.

[99] X. Feng and O. Karakashian, *Two-level additive Schwarz methods for a discontinuous Galerkin approximation of second order elliptic problems*, SIAM J. Numer. Anal., 39 (2001), pp. 1343–1365 (electronic).

[100] X. Feng and O. Karakashian, *Fully discrete dynamic mesh discontinuous Galerkin methods for the Cahn-Hilliard equation of phase transition*, Math. Comp., 76 (2007), pp. 1093–1117.

[101] T.-P. Fries and T. Belytschko, *The extended/generalized finite element method: an overview of the method and its applications*, Internat. J. Numer. Methods Engrg., 84 (2010), pp. 253–304.

[102] W. Gautschi, *Orthogonal polynomials: computation and approximation*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2004. Oxford Science Publications.

[103] E. Georgoulis, *Discontinuous Galerkin methods on shape-regular and anisotropic meshes*, D.Phil. Thesis, University of Oxford, (2003).

[104] ——, *Inverse-type estimates on hp-finite element spaces and applications*, Math. Comp., 77 (2008), pp. 201–219 (electronic).

[105] E. Georgoulis and A. Lasis, *A note on the design of hp-version interior penalty discontinuous Galerkin finite element methods for degenerate problems.*, IMA J. Numer. Anal., 26 (2006), pp. 381–390.

[106] E. H. Georgoulis, E. Hall, and P. Houston, *Discontinuous Galerkin methods on hp-anisotropic meshes. I. A priori error analysis*, Int. J. Comput. Sci. Math., 1 (2007), pp. 221–244.

[107] E. H. Georgoulis and P. Houston, *Discontinuous Galerkin methods for the biharmonic problem*, IMA J. Numer. Anal., 29 (2009), pp. 573–594.

[108] E. H. Georgoulis, P. Houston, and J. Virtanen, *An a posteriori error indicator for discontinuous Galerkin approximations of fourth-order elliptic problems*, IMA J. Numer. Anal., (2009), p. drp023.

[109] S. Giani and P. Houston, *Domain decomposition preconditioners for discontinuous Galerkin discretizations of compressible fluid flows*, Numer. Math. Theory Methods Appl., 7 (2014).

[110] ——, *hp–Adaptive composite discontinuous Galerkin methods for elliptic problems on complicated domains*, Num. Meth. Part. Diff. Eqs., 30 (2014), pp. 1342–1367.

[111] S. Giani and P. Houston, *hp–adaptive composite discontinuous Galerkin methods for elliptic problems on complicated domains*, Num. Meth. Part. Diff. Eqs., 30 (2014), pp. 1342–1367.

[112] G. H. Golub and J. H. Welsch, *Calculation of gauss quadrature rules*, Math. Comp., 23 (1969), pp. 221–230.

[113] P. Grisvard, *Elliptic problems in nonsmooth domains*, vol. 69, SIAM, 2011.

[114] W. Gui and I. Babuška, *The h, p and h-p versions of the finite element method in* 1 *dimension. I–III.*, Numer. Math., 49 (1986), pp. 577–683.

[115] W. Hackbusch and S. Sauter, *Composite finite elements for problems containing small geometric details. Part II: Implementation and numerical results*, Comput. Visual Sci., 1 (1997), pp. 15–25.

[116] ——, *Composite finite elements for the approximation of PDEs on domains with complicated micro-structures*, Numer. Math., 75 (1997), pp. 447–472.

[117] N. Hale and A. Townsend, *Fast and accurate computation of gauss–legendre and gauss–jacobi quadrature nodes and weights*, SIAM J. Sci. Comput., 35 (2013), pp. A652–A674.

[118] R. Hartmann and P. Houston, *Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws*, SIAM J. Sci. Comput., 24 (2002), pp. 979–1004.

[119] F. Heimann, C. Engwer, O. Ippisch, and P. Bastian, *An unfitted interior penalty discontinuous Galerkin method for incompressible Navier–Stokes two-phase flow*, Internat. J. Numer. Methods Engrg., 71 (2013), pp. 269–293.

[120] J. Hesthaven and T. Warburton, *Nodal discontinuous Galerkin methods*, vol. 54 of Texts in Applied Mathematics, Springer, New York, 2008. Algorithms, analysis, and applications.

[121] J. S. Hesthaven and T. Warburton, *Nodal high-order methods on unstructured grids: I. time-domain solution of maxwell's equations*, J. Comput. Phys., 181 (2002), pp. 186–221.

[122] P. Houston, I. Perugia, and D. Schotzau, *Mixed discontinuous Galerkin approximation of the maxwell operator*, SIAM J. Numer. Anal., 42 (2004), pp. 434–459.

[123] P. Houston, D. Schötzau, and T. Wihler, *Energy norm a posteriori error estimation of hp-adaptive discontinuous Galerkin methods for elliptic problems*, Math. Models Methods Appl. Sci., 17 (2007), pp. 33–62.

[124] P. Houston, C. Schwab, and E. Süli, *Stabilized hp-finite element methods for first-order hyperbolic problems*, SIAM J. Numer. Anal., 37 (2000), pp. 1618–1643 (electronic).

[125] ——, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems*, SIAM J. Numer. Anal., 39 (2002), pp. 2133–2163 (electronic).

[126] P. Houston and E. Süli, *Stabilised hp–finite element approximation of partial differential equations with nonnegative characteristic form*, Computing, 66 (2001), pp. 99–119.

[127] P. Jamet, *Galerkin-type approximations which are discontinuous in time for parabolic equations in a variable domain*, SIAM J. Numer. Anal., 15 (1978), pp. 912–928.

[128] M. Jensen, *Discontinuous Galerkin methods for friedrichs systems*, D.Phil. Thesis, University of Oxford, (2005).

[129] A. Johansson and M. Larson, *A high order discontinuous Galerkin Nitsche method for elliptic problems with fictitious boundary*, Numer. Math., 123 (2013), pp. 607–628.

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

[131] C. Johnson, U. Nävert, and J. Pitkäranta, *Finite element methods for linear hyperbolic problems*, Comput. Methods Appl. Mech. Engrg., 45 (1984), pp. 285–312.

[132] C. Johnson and J. Pitkäranta, *An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation*, Math. Comp., 46 (1986), pp. 1–26.

[133] ——, *An analysis of the discontinuous Galerkin method for a scalar hyperbolic equation*, Math. Comp., 46 (1986), pp. 1–26.

[134] O. Karakashian and F. Pascal, *A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 2374–2399 (electronic).

[135] G. Karniadakis and S. Sherwin, *Spectral/hp element methods for computational fluid dynamics*, Oxford University Press, 2013.

[136] G. Karypis and V. Kumar, *A fast and highly quality multilevel scheme for partitioning irregular graphs*, SIAM J. Sci. Comput., 20 (1999), pp. 359–392.

[137] D. Kay, V. Styles, and E. Süli, *Discontinuous Galerkin finite element approximation of the Cahn-Hilliard equation with convection*, SIAM J. Numer. Anal., 47 (2009), pp. 2660–2685.

[138] D. Kröner, *Numerical Schemes for Conservation Laws*, Wiley-Teubner, 1997.

[139] O. A. Ladyzhenskaia, V. A. Solonnikov, and N. N. Uraltseva, *Linear and quasi-linear equations of parabolic type*, vol. 23, American Mathematical Soc., 1988.

[140] C. Lasser and A. Toselli, *An overlapping domain decomposition preconditioner for a class of discontinuous Galerkin approximations of advection-diffusion problems*, Math. Comp., 72 (2003), pp. 1215–1238 (electronic).

[141] P. Lesaint and P.-A. Raviart, *On a finite element method for solving the neutron transport equation*, in Mathematical aspects of finite elements in partial differential equations (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1974), Math. Res. Center, Univ. of Wisconsin-Madison, Academic Press, New York, 1974, pp. 89–123. Publication No. 33.

[142] J.-L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications. Vol. I*, Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.

[143] A. Massing, *Analysis and implementation of Finite Element Methods on overlapping and Fictitious Domains*, PhD thesis, University of Oslo, 2012.

[144] J. M. Melenk and T. Wurzer, *On the stability of the boundary trace of the polynomial L_2-projection on triangles and tetrahedra*, Comput. Math. Appl., 67 (2014), pp. 944–965.

[145] P. Monk and E. Süli, *The adaptive computation of far-field patterns by a posteriori error estimation of linear functionals*, SIAM J. Numer. Anal., 36 (1998), pp. 251–274.

[146] I. Mozolevski and E. Süli, *A priori error analysis for the hp-version of the discontinuous Galerkin finite element method for the biharmonic equation*, Comput. Methods Appl. Math., 3 (2003), pp. 596–607.

[147] R. Muñoz-Sola, *Polynomial liftings on a tetrahedron and applications to the h-p version of the finite element method in three dimensions*, SIAM J. Numer. Anal., 34 (1997), pp. 282–314.

[148] S. Natarajan, S. Bordas, and D. Mahapatra, *Numerical integration over arbitrary polygonal domains based on Schwarz-Christoffel conformal mapping*, Internat. J. Numer. Methods Engrg., 80 (2009), pp. 103–134.

[149] J. Nitsche, *Über ein Variationsprinzip zur Lösung von Dirichlet Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind*, Abh. Math. Sem. Uni. Hamburg, 36 (1971), pp. 9–15.

[150] J. T. Oden, I. Babuška, and C. E. Baumann, *A discontinuous hp finite element method for diffusion problems*, J. Comput. Phys., 146 (1998), pp. 491–519.

[151] O. Oleinik and E. Radkevič, *Second Order Equations with Nonnegative Characteristic Form*, American Mathematical Society, 1973.

[152] I. Perugia and D. Schötzau, *An hp-analysis of the local discontinuous Galerkin method for diffusion problems*, J. Sci. Comput., 17 (2002), pp. 561–571.

[153] ——, *The hp-local discontinuous Galerkin method for low-frequency time-harmonic maxwell equations*, Math. Comp., 72 (2003), pp. 1179–1214.

[154] T. Peterson, *A note on the convergence of the discontinuous Galerkin method for a scalar hyperbolic equation*, SIAM J. Numer. Anal., 28 (1991), pp. 133–140.

[155] W. Reed and T. Hill, *Triangular mesh methods for the neutron transport equation.*, Technical Report LA-UR-73-479 Los Alamos Scientific Laboratory, (1973).

[156] B. Rivière, M. Wheeler, and V. Girault, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. I*, Comput. Geosci., 3 (1999), pp. 337–360 (2000).

[157] ——, *A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems*, SIAM J. Numer. Anal., 39 (2001), pp. 902–931 (electronic).

[158] S. A. Sauter and R. Warnke, *Extension operators and approximation on domains containing small geometric details*, East-West J. Numer. Math., 7 (1999), pp. 61–77.

[159] D. Schötzau and C. Schwab, *Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method*, SIAM J. Numer. Anal., 38 (2000), pp. 837–875.

[160] D. Schötzau and C. Schwab, *Exponential convergence for hp-version and spectral finite element methods for elliptic problems in polyhedra*, Math. Models Methods Appl. Sci., 25 (2015), pp. 1617–1661.

[161] D. Schötzau, C. Schwab, and A. Toselli, *Mixed hp-DGFEM for incompressible flows*, SIAM J. Numer. Anal., 40 (2002), pp. 2171–2194.

[162] ——, *Mixed hp-DGFEM for incompressible flows II: Geometric edge meshes*, IMA J. Numer. Anal., 24 (2004), pp. 273–308.

[163] D. Schötzau, C. Schwab, and T. P. Wihler, *hp-dGFEM for second-order elliptic problems in polyhedra I: Stability on geometric meshes*, SIAM J. Numer. Anal., 51 (2013), pp. 1610–1633.

[164] ——, *hp-dGFEM for second order elliptic problems in polyhedra II: Exponential convergence*, SIAM J. Numer. Anal., 51 (2013), pp. 2005–2035.

[165] ——, *hp-dGFEM for second-order mixed elliptic problems in polyhedra*, Math. Comp., 85 (2016), pp. 1051–1083.

[166] D. Schötzau and L. Zhu, *A robust a-posteriori error estimator for discontinuous Galerkin methods for convection–diffusion equations*, Appl. Numer. Math., 59 (2009), pp. 2236–2255.

[167] C. Schwab, *p– and hp–Finite element methods: Theory and applications in solid and fluid mechanics*, Oxford University Press: Numerical mathematics and scientific computation, 1998.

[168] I. Smears and E. Süli, *Discontinuous Galerkin finite element approximation of Hamilton–Jacobi–Bellman equations with Cordes coefficients*, SIAM J. Numer. Anal., 52 (2014), pp. 993–1016.

[169] ——, *Discontinuous Galerkin finite element methods for time-dependent Hamilton–Jacobi–Bellman equations with Cordes coefficients*, Numer. Math., 133 (2016), pp. 141–176.

[170] P. Solin, K. Segeth, and I. Dolezel, *Higher-order finite element methods*, Studies in advanced mathematics, Chapman & Hall/CRC, Boca Raton, London, 2004.

[171] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton, University Press, Princeton, N.J., 1970.

[172] G. Strang and G. Fix, *An analysis of the finite element method*, Prentice-Hall Inc., Englewood Cliffs, N. J., 1973. Prentice-Hall Series in Automatic Computation.

[173] J. J. Sudirham, J. J. W. van der Vegt, and R. M. J. van Damme, *Space-time discontinuous Galerkin method for advection-diffusion problems on time-dependent domains*, Appl. Numer. Math., 56 (2006), pp. 1491–1518.

[174] N. Sukumar and A. Tabarraei, *Conforming polygonal finite elements*, Internat. J. Numer. Methods Engrg., 61 (2004), pp. 2045–2066.

[175] E. Süli, P. Houston, and C. Schwab, *Hyperbolic problems*, The Mathematics of Finite Elements and Applications X (MAFELAP 1999), (2000), p. 143.

[176] E. Süli and I. Mozolevski, *hp-version interior penalty DGFEMs for the biharmonic equation*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 1851–1863.

[177] O. Sutton, *The virtual element method in 50 lines of matlab*, arXiv preprint arXiv:1604.06021, (2016).

[178] G. Szegö, *Orthogonal polynomials*, vol. 23, American Mathematical Soc., 1939.

[179] C. Talischi, G. Paulino, A. Pereira, and I. Menezes, *Polymesher: A general-purpose mesh generator for polygonal elements written in Matlab*, Struct. Multidisc. Optim., 45 (2012), pp. 309–328.

[180] V. Thomée, *Galerkin finite element methods for parabolic problems*, vol. 1054 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 1984.

[181] E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics*, Springer, 1997.

[182] J. J. W. van der Vegt and J. J. Sudirham, *A space-time discontinuous Galerkin method for the time-dependent Oseen equations*, Appl. Numer. Math., 58 (2008), pp. 1892–1917.

[183] R. Verfürth, *On the constants in some inverse inequalities for finite element functions.*, Tech. Rep. 257, University of Bochum, 1999.

[184] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Oxford University Press, 2013.

[185] T. Warburton and J. S. Hesthaven, *On the constants in hp-finite element trace inverse inequalities*, Comput. Methods Appl. Mech. Engrg., 192 (2003), pp. 2765–2773.

[186] M. F. Wheeler, *A priori L_2 error estimates for Galerkin approximations to parabolic partial differential equations*, SIAM J. Numer. Anal., 10 (1973), pp. 723–759.

[187] ——, *An elliptic collocation-finite element method with interior penalties*, SIAM J. Numer. Anal., 15 (1978), pp. 152–161.

[188] T. Wihler, *Locking-free adaptive discontinuous Galerkin FEM for linear elasticity problems*, Math. Comp., 75 (2006), pp. 1087–1102.

[189] T. Wihler, P. Frauenfelder, and C. Schwab, *Exponential convergence of the hp-DGFEM for diffusion problems*, Comput. Math. Appl., 46 (2003), pp. 183–205.

[190] L. Zhu, S. Giani, P. Houston, and D. Schötzau, *Energy norm a posteriori error estimation for hp-adaptive discontinuous Galerkin methods for elliptic problems in three dimensions*, Math. Models Methods Appl. Sci., 21 (2011), pp. 267–306.