Numerical Methods for Elliptic Partial Differential Equations and Optimal Control Problems

Submitted in partial fulfillment of the requirements

of the degree of

Doctor of Philosophy

of the
Indian Institute of Technology Bombay, India
and
Monash University, Australia

by

Devika S

Supervisors:

Supervisor - Professor Neela Nataraj (IIT Bombay)

Supervisor - Professor Jérôme Droniou (Monash University)

The course of study for this award was developed jointly by Monash University, Australia and the Indian Institute of Technology Bombay, India and was given academic recognition by each of them. The programme was administrated by The IITB-Monash Research Academy

(Year 2019)

Declaration

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Student Name: Devika S

IITB ID: 154094001

Monash ID: 27346595

Abstract

This thesis studies numerical methods for elliptic partial differential equations and optimal control problems. Second order and fourth order elliptic partial differential equations arise in various applications, in structural engineering, image processing, thin plates, thin beams, biharmonic problem, the Stokes problem in stream function and vorticity formulation, and so on.

A unified convergence analysis framework, known as the Hessian discretisation method (HDM), for fourth order elliptic equations is designed and analysed in the first part of the dissertation. The principle of the HDM is inspired by the gradient discretisation method for second order problems. The HDM framework, introduced in Chapter 2, covers many different numerical methods such as conforming and nonconforming finite element methods, finite volume methods and methods based on gradient recovery operators. It is established that three properties, namely coercivity, consistency and limit-conformity, are sufficient to prove the convergence of HDM for linear elliptic problems. An additional property of compactness helps to analyse fourth order semi-linear problems with trilinear nonlinearity. This in particular applies to the stream function vorticity formulation of the incompressible 2D Navier–Stokes problem and the von Kármán equations. For these non-linear models, convergence is proved using two different approaches: by compactness techniques, that does not require any additional smoothness or structural assumption on the continuous solution and by error estimates, under some smoothness assumption on the solution. The framework of Hessian schemes enables us to develop one study that encompasses numerous classical methods. Numerical results are presented to support the theoretical estimates.

Optimal control problems have found applications in many different fields, including aerospace, process control, robotics, bioengineering, economics, finance, and management science, and it continues to be an active research area within control theory. The gradient discretisation method (GDM) is a generic framework for the convergence analysis of numerical methods such as conforming and nonconforming finite element methods, finite volume methods and mimetic finite difference methods for diffusion equations. In the second part of this thesis, the numerical approximation of optimal control problem governed by diffusion equation (resp. fourth order linear elliptic equations) with Dirichlet and Neumann boundary conditions (resp. clamped boundary conditions) using the GDM (resp. HDM) have been studied. The pure Neumann control problem is numerically analysed for the first time, even for standard finite element methods. Error estimates of two kinds are derived for the state, adjoint and control variables. Firstly, basic error estimates in a very generic setting are established. Secondly, considering slightly more restrictive assumptions on the admissible control set, super-convergence results for all three variables are derived. The theoretical results are substantiated by the output of numerical experiments.

Contents

Chapter 1

Introduction

This introductory chapter deals with a subsection on motivation, literature survey, chapter-wise description, notations and standard results which are frequently used throughout the thesis.

1.1 Motivation

The theory of partial differential equations is one of the main research areas in mathematics and has applications in various fields, mainly in physics and engineering. The purpose of this thesis is to study the convergence analysis of numerical methods for elliptic problems and optimal control problems. Consider the simple model of the diffusion equation with homogeneous Dirichlet boundary conditions defined by

$$-\mathrm{div}(A\nabla\overline{u}) = f \quad \text{in } \Omega, \tag{1.1.1a}$$

$$\overline{u} = 0 \quad \text{on } \partial\Omega, \tag{1.1.1b}$$

where $\Omega \subsetneq \mathbb{R}^d$ $(d \geq 1)$ is a bounded domain with boundary $\partial\Omega$ and $f \in L^2(\Omega)$. It is assumed that $A : \Omega \to S_d(\mathbb{R})$ is a measurable, coercive and bounded function with values in the space of $d \times d$ symmetric matrices. The above equation arises in various frameworks such as image processing and reservoir engineering (e.g. petroleum simulation). The flow models involve diffusion operators such as in (1.1.1).
The variational formulation of (1.1.1) is given by

$$\text{find } \overline{u} \in H_0^1(\Omega) \text{ such that, for all } v \in H_0^1(\Omega),\ \int_\Omega A\nabla\overline{u}\cdot\nabla v\,\mathrm{d}\boldsymbol{x} = \int_\Omega fv\,\mathrm{d}\boldsymbol{x}.$$

Existence and uniqueness of a weak solution $\overline{u}$ is a straightforward consequence of the Lax-Milgram theorem. If Ω is a convex polygonal domain and A is Lipschitz continuous, then $\overline{u}$ belongs to $H_0^1(\Omega)\cap H^2(\Omega)$. There are a wide variety of numerical methods to approximate the solution of this problem, such as finite element methods, discontinuous Galerkin methods and finite volume methods.

The gradient discretisation method (GDM) is a generic framework for the convergence analysis for diffusion equations of different kinds: linear or non-linear, steady-state or time-dependent. The GDM [48] covers a wide range of numerical methods such as finite element methods, mixed finite element elements, finite volume methods and mimetic methods. The GDM consists in replacing the continuous space and operators by discrete ones in the weak formulation of the partial differential equation (PDE). The set of discrete elements thus chosen is called a gradient discretisation (GD), and the scheme obtained by using these elements is a gradient scheme (GS). The variety of possible choices of GDs result in as many different GSs. Only a few core properties are needed to ensure the convergence of a GDM. For linear problems, a GD must satisfy three core properties; namely coercivity, consistency and limit-conformity, to give rise to a convergent GS. The compactness and piecewise constant reconstruction are the additional properties required to establish the convergence analysis for the non-linear equations.

The optimal control problem [109] consists of seeking a control function that minimizes a cost functional subject to a boundary value problem. Some of its applications lie in aviation and space technology, engineering, the life sciences, robotics and movement sequences in sports. If the control acts in a subdomain of Ω rather than the boundary, the problem is known as a distributed control problem, whereas a boundary control problem is obtained when the control acts through a boundary condition. The optimal control problem governed by diffusion equation with Dirichlet boundary condition originates from the optimal stationary heating, for example, with controlled heat source on a bounded domain. Problem of this kind arises if the body Ω is heated by electromagnetic induction or by microwaves. Assume that the boundary temperature vanishes. Then the corresponding model problem will be a second order elliptic control problem with distributed control and homogeneous Dirichlet boundary condition. Here the control acts as the heat source in the domain Ω and state variable represents the temperature.

The pure Neumann control problems have wider application potential in optimisation problems involving an integral constraint. For example, in the model of [34, 100] describing the miscible displacement of one fluid by another in a porous medium, the pressure is subjected to an elliptic equation with homogeneous Neumann boundary conditions. In this equation, the source terms model the injection and production wells, and are typically the only quantities that engineers can adjust (to some extent). Hence, considering these source terms as controls of the pressure may lead to optimal control problems governed by pure Neumann diffusion equation, with homogeneous boundary conditions, zero average constraints on the state and integral constraints on the control terms.

Fourth order elliptic partial differential equations arise in various applications, such as structural engineering, thin plate theories of elasticity, thin beams, biharmonic problems, the Stokes problem in stream function and vorticity formulation, image processing, etc. Consider the following

fourth order linear model problem with clamped boundary conditions (BC).

$$\sum_{i,j,k,l=1}^{d} \partial_{kl}(a_{ijkl}\partial_{ij}\overline{u}) = f \quad \text{in } \Omega, \tag{1.1.2a}$$

$$\overline{u} = \frac{\partial \overline{u}}{\partial n} = 0 \quad \text{on } \partial\Omega, \tag{1.1.2b}$$

where $\Omega \subsetneq \mathbb{R}^d$ is a bounded domain with boundary $\partial\Omega$, $f \in L^2(\Omega)$ and n is the unit outer normal to Ω.

Some important examples of fourth order linear elliptic problems are the biharmonic problem and the plate problems [41]. The biharmonic equation arises in areas of continuum mechanics, including linear elasticity theory and the solution of Stokes flows. As the Laplace problem with Dirichlet BC models the displacement of a membrane fixed along the boundary and acted upon by a force, the biharmonic problem with clamped BC describes the bending of a thin elastic plate which is clamped along the boundary and acted upon by a force. If we wish to consider a plate which is simply supported on the boundary and is fixed along it, then the boundary condition $\overline{u} = 0$ will be retained but the condition $\frac{\partial \overline{u}}{\partial n}$ will be replaced by some other boundary condition.

Fourth order semi-linear problems with linear biharmonic operator as the leading term and quadratic lower order contributions appear in various domains of mechanics. They model for example 2D incompressible flows through the stream function vorticity approach of the Navier–Stokes equations [19] and the very thin plates deformations of the von Kármán equations [42]. There are advantages in using the stream function vorticity formulation of the incompressible Navier–Stokes equations to compute 2D flows: the continuity equation is automatically satisfied, only one (vorticity equation) transport equation has to be solved, the streamlines of the flow are given by level curves of the stream function, and the vorticity is a conserved quantity. The two-dimensional von Kármán equations for nonlinearly elastic plates were proposed by von Kármán to describe the transverse displacement of the middle surface of the plate and the Airy stress function. Many of the major advances in steady-state bifurcation theory were stimulated and illustrated by studies of buckling of plates described by these equations [42, 43].

1.2 Literature Review

The GDM is a generic framework which contains a wide class of numerical methods (finite elements, mixed finite elements, finite volume, mimetic finite difference methods, etc.) for linear and non-linear elliptic and parabolic diffusion equations (including degenerate equations), the Navier–Stokes equations, variational inequalities, Darcy flows in fractured media, etc. See for example [1, 50–52, 61], and the monograph [48] for a complete presentation of the GDM for various boundary conditions and models. GDM allows a complete convergence analysis for families of numerical methods through a small number of properties depending on the considered model: coercivity, consistency, limit-conformity, compactness and piecewise constant reconstruction.

For linear problems, a GD must satisfy three core properties, coercivity, consistency and limit-conformity, for the convergence analysis to hold. Note that the convergence is established by means of error estimates. Stability can be obtained through the coercivity of the discrete operators that ensure a discrete Poincaré inequality. The consistency is nothing but the interpolation error in the finite element framework. The limit-conformity measures the defect in the discrete Stokes formula and should tend to zero if the underlying mesh size tends to zero. The compactness and piecewise constant reconstruction properties are useful when dealing with non-linearities in the PDE. To deal with low-order non-linearities (e.g. semi-linear equations), compactness property is required that ensures a discrete Rellich theorem. The piecewise constant reconstruction is used to control nonlinearities in the quasi-linear equations. For non-linear models, the convergence of approximate solutions can be proved by compactness techniques. This argument does not require any regularity of the solution and is thus of particular interest.

Numerical methods for second-order optimal control problems governed by Dirichlet boundary condition have been studied in various articles (see, e.g., [3, 33, 77, 95, 96] for distributed control, [4, 94] for boundary control, and references therein). For conforming and mixed finite element methods, superconvergence result of control has been derived in [37, 38, 96] where the control is approximated by piecewise constant functions. Even though the approximation of discrete solution is of $\mathcal{O}(h)$, a postprocessing step improves the convergence rate to $\mathcal{O}(h^2)$ [96]. However, to the best of our knowledge, the improved error estimate which is known in literature as the super-convergence result has not been studied for non-conforming finite element methods. One of the consequences of our generic analysis is to establish superconvergence results for several conforming and non-conforming numerical methods covered by gradient schemes – in particular, the classical Crouzeix-Raviart finite element method (FEM) and the mixed-hybrid mimetic finite difference schemes [2].

Several works cover optimal control for second order Neumann boundary value problems, albeit with an additional (linear or non-linear) reaction term which makes the state equation naturally well- posed, without zero average constraint, see [5, 6, 30, 78, 93]. In [27], error estimates are obtained with order $\mathcal{O}(h^{3/2})$ for Neumann control problems in a two-dimensional domain under the assumptions similar to that in [96]. To the best of our knowledge, the numerical analysis of pure Neumann control problems, without reaction term and thus with the integral constraint, is considered for the first time even for finite element methods. Being established in the GDM framework, our results for this model cover a range of numerical methods, including conforming Galerkin methods, non-conforming finite elements, and mimetic finite differences. Although done on the simple problem, the analysis uncovers some properties of general interest, such as the specific relation formula between the adjoint and control variables and a modified active set algorithm used to compute the solution of the numerical scheme.

Let us also mention that for results on optimal control problems governed by second order non-linear elliptic equations, many references are available, see for example [26, 28–32]. Piecewise linear finite elements are used to approximate the control as well as the state for semilinear equations in [31]. Error estimates for optimal controls in L^∞ norm are investigated in [7, 97]. Note that [7] is concerned with the discretisation of the control with piecewise linear functions, whereas

the control is discretised by piecewise constant functions in [97]. In both papers, error estimates of order $\mathcal{O}(h)$ were proved in L^∞ norm.

Many numerical methods, most of them finite elements, have been developed over the years to approximate the solutions of fourth order models. Each method comes with its own convergence analysis carried out using ad-hoc techniques. Finite element method [41] is one of the popular and classical numerical technique for solving fourth order elliptic boundary value problems. When conforming finite elements are used, the approximation space must be a subspace of $H_0^2(\Omega)$. The corresponding strong continuity requirement of the function and its derivatives makes it difficult to construct such a finite element, and leads to schemes with a large number of unknowns [16, 41, 46, 101, 102]. The classical examples of conforming finite elements are the Argyris finite elements with 21 degrees of freedom in a triangle, and the Bogner-Fox-Schmit rectangle with 16 degrees of freedom in a rectangle [41]. Contrary to these two elements, the Hsieh-Clough-Tocher element is a macro conforming finite element, i.e. it is composed of subelements with piecewise polynomial functions. The nonconforming finite element method (ncFEM) relaxes the continuity requirement and hence employ fewer degrees of freedom, which has a great impact on the resulting scheme. For the fourth order problems, two interesting nonconforming elements are the Adini rectangle and the Morley triangle [41]. The functions in Adini finite element space are continuous on Ω, but not continuously differentiable. The advantage of Morley FEM is that it uses piecewise quadratic polynomials for the approximation and hence is simpler to implement. The finite element methods are well-developed for the fourth order partial differential equations with variable constant coefficients, biharmonic problem and the bending problem, see [8, 12, 62, 63, 82, 85, 86, 88, 99, 103, 111]. Under regularity assumption on the solution $\overline{u}$, it is well-known that the conforming and nonconforming finite element methods give a linear rate of convergence in the energy norm and quadratic rate of convergence (or better) in the L^2 and H^1 norms. The convergence analysis for the conforming finite element methods can be found in [41]. Error estimates for some nonconforming finite elements for the thin and very thin plate bending problems were studied in [85, 87, 92] and references therein.

In [83], a finite element method for the biharmonic equation is presented; this method is based on gradient recovery (GR) operator, where the basis functions of the two involved spaces satisfy a condition of biorthogonality. The main idea is to use the gradient recovery operator to lift the non-differentiable, piecewise-constant gradient of $\mathbb{P}_1$ finite element functions into the $\mathbb{P}_1$ finite element space itself; the lifted functions are thus differentiable, and can be used to compute some kind of Hessian matrix of $\mathbb{P}_1$ finite element functions, see [81–83] for more details. Ensuring the coercivity of the method in [83] on generic triangular/tetrahedral meshes however requires the addition of a stabilisation term. We also refer to [35] for the application of the gradient recovery operator to fourth order eigenvalue problems. Under the regularity assumptions, error estimate for the gradient recovery method between the gradient and the approximation of the gradient of a solution were investigated in [75, 112] and a quadratic order of convergence is established if the mesh is regular.

A cell-centered finite volume scheme for the approximation of a biharmonic problem with Dirichlet boundary conditions was proposed and analyzed in [59], first on grids which satisfy an orthog-

onality condition known as Δ-adapted discretisations, and then on general meshes. The interest of the method in [59] is that it is easy to implement, computationally cheap and requires only one unknown per cell. These methods are designed on the principle that it preserves the flux balance and conservative equations. If the solution $\overline{u}$ belongs to $C^4(\overline{\Omega}) \cap H_0^2(\Omega)$, [59, Theorem 4.3] provides an $\mathcal{O}(h^{1/5})$ estimate for the finite volume method (FVM) based on Δ-adapted discretisations. Our analysis slightly improves this result in the HDM framework. Error estimates are obtained using only three properties of HD. If the solution $\overline{u}$ belongs to $H^4(\Omega) \cap H_0^2(\Omega)$, then $\mathcal{O}(h^{1/4}|\ln(h)|)$ error estimate is obtained for the Hessian scheme based on the HD. However, an $\mathcal{O}(h^2)$ superconvergence rate in L^2 norm has been numerically observed.

The von Kármán equations [42] is a system of fourth order semi-linear elliptic equations that describes the bending of very thin elastic plates. The numerical analysis of von Kármán equations has been studied using conforming finite element methods in [18, 91], Morley nonconforming finite element method in [92], mixed finite element methods in [19, 98], C^0 interior penalty method in [15] and discontinuous Galerkin method in [24]. To the best of our knowledge, the Adini nonconforming finite element method and the method based on gradient recovery operator have not been studied in literature. For the stream function vorticity formulation of the incompressible 2D Navier–Stokes equation, we refer to [24] and the references therein. The HDM framework provides a unified framework for the convergence analysis of several numerical methods, such as, the conforming and non-conforming finite element methods and methods based on gradient recovery operators in an abstract setting. Four properties namely, the coercivity, consistency, limit-conformity and compactness establish the convergence analysis in HDM framework. In addition, a companion operator yields the error estimates and examples of these operators are provided for the finite element methods.

In [23, 64], mixed finite element methods have been proposed and analyzed for a distributed optimal control problem governed by the biharmonic equation with clamped boundary conditions while a C^0 interior penalty method is analyzed in [73] for biharmonic optimal control problem. In [39], an energy space based approach for Dirichlet boundary control problem governed by biharmonic equation has been investigated. An abstract framework for the error analysis of discontinuous finite element methods applied to control constrained optimal control problems has been developed in [40]. Error analysis for a stable C^0 interior penalty method is derived for general fourth order problems on polygonal domains under minimal regularity assumptions on the exact solution in [72]. The last part of this dissertation focusses on the control problem governed by fourth order linear elliptic equations with clamped boundary condition in the HDM framework and thus applicable to several numerical methods, including the conforming FEMs, the Adini and Morley non-conforming FEMs, the gradient recovery methods and the FVMs. A generic error estimate and superconvergence result are established for state, adjoint and control variables.

1.3 Organization and Contributions of the Thesis

Chapter 1 deals with motivations for this study, literature survey, notations and some standard results. In Chapter 2, a unified convergence analysis framework known as the Hessian discreti-

sation method has been designed [53]. The principle of the HDM is inspired by the GDM for second order problems. The HDM is based on four discrete elements (called altogether a Hessian discretisation) and a few intrinsic indicators of accuracy, independent of the considered model. These elements are then substituted, in the weak formulation of the model, to the corresponding continuous space and operators, giving rise to a numerical scheme; this scheme is called a Hessian scheme (HS). In the GDM, the definition of a gradient discretisation is independent of the differential operator. Here, the definition of Hessian discretisation depends on a fourth order tensor B, that appears in the differential operator (see Definition 2.3.1). This is justified by the fact that some methods (such as the finite volume method presented in Section 2.3.1) are not built on an approximation of the entire Hessian of the functions, but only on some of their derivatives (such as the Laplacian of the functions). An error estimate is obtained (Theorem 2.4.4), using only a few intrinsic indicators, namely, coercivity, consistency and limit-conformity, when the HDM framework is applied to linear fourth order problems. It is shown that HDM covers a large number of numerical methods for fourth order elliptic problems: conforming and non-conforming finite element methods as well as finite volume methods. We also use the HDM to design a novel method, based on conforming $\mathbb{P}_1$ finite element space and gradient recovery operators. Further, improved L^2 and H^1-like error estimates (Theorems 2.5.1 and 2.6.2) are established in the framework of HDM and applied it to various schemes [104]. Results of numerical experiments are presented for the novel scheme based on gradient recovery operator and for finite volume schemes.

Chapter 3 deals with the HDM for fourth order semi-linear elliptic equations with a trilinear nonlinearity in an abstract setting. This abstract result applies to the incompressible 2D Navier–Stokes equations in vorticity formulation and the von Kármán equations of plate bending. Some examples of HDM are presented, such as the finite element methods (conforming and non-conforming) and methods based on gradient recovery operators. The convergence analysis is established in HDM with the help of four properties, namely coercivity, consistency, limit-conformity and compactness, associated with the HD. For linear models, limit-conformity defect measures the error in the discrete Stokes formula between the reconstructed Hessian and the reconstructed function and this limit-conformity is sufficient to analyse the convergence of the HDM for linear models. However, the non-linear model involves the gradient, and hence limit-conformity measure between the reconstructed gradient and the reconstructed function is necessary to identify during the convergence analysis along with the limit-conforming measure considered in the linear case. The convergence analysis is first proved by compactness techniques using these four properties without assuming any regularity of the continuous solution (Theorem 3.5.1). Then, upon assuming some structural properties of the continuous solution, an error estimate is obtained for the HDM approximation of the considered non-linear models (Theorem 3.5.12). Results of numerical experiments are presented for the Morley FEM and gradient recovery method, a specific scheme that fits into the framework of the HDM and that is designed based on cheap, local reconstructions of higher-order derivatives for piecewise linear functions.

In Chapter 4, optimal control problems governed by diffusion equations with Dirichlet and Neumann boundary conditions are investigated in the framework of the gradient discretisation method

[55]. Here, the state equation considered in the Neumann control problem has a reaction term. Gradient schemes are defined for the optimality system of the control problem. Basic error estimates that provide a linear convergence rate for all the three variables (control, state, and adjoint) for low order schemes under standard regularity assumptions are established (Theorem 4.3.2). Given that the optimal control is approximated by piecewise constant functions, the convergence rates are optimal. An improved error estimate has been proved for optimal controls, state and adjoint variables with the help of a post-processing step (Theorems 4.3.6 and 4.3.7). These super-convergence results are shown to apply to non-conforming $\mathbb{P}_1$ finite elements, and to the mixed/hybrid mimetic finite differences. Results of numerical experiments are demonstrated for the conforming, non-conforming and mixed-hybrid mimetic finite difference schemes.

Chapter 5 discusses the GDM for distributed optimal control problems governed by diffusion equation with pure Neumann boundary condition and zero average constraint [56]. Optimal order error estimates for the state, adjoint and control variables for low order schemes are derived under standard regularity assumptions (Theorem 5.4.1). For the pure Neumann control problem, without reaction term, one of the objectives is to establish a projection relation between control and adjoint variables (Lemma 5.4.9). This relation, which is non-standard since it has to account for the zero average constraints, is the key to prove the super-convergence result for all three variables (Theorem 5.4.5). A modified active set strategy algorithm for GDM that is adapted to this non-standard projection relation has been designed. Numerical experiments performed using a modified active set strategy algorithm for conforming, nonconforming and mimetic finite difference methods confirm the theoretical rates of convergence.

Chapter 6 deals with the optimal control problems governed by fourth order linear elliptic equations with clamped boundary conditions in the framework of the HDM. Basic error estimates and superconvergence results for the state, adjoint and control variables are established in the HDM framework (Theorems 6.3.2 and 6.3.5). Since HDM covers the conforming FEMs, the Adini and Morley ncFEMs, the GR methods and the FVMs, the basic error estimates is valid for these methods. The superconvergence result for control is established under the superconvergence assumption for the state equation and a few other assumptions. These assumptions are verified for the conforming FEMs, the Adini and Morley ncFEMs and the GR methods. Numerical experiments are implemented for the gradient recovery method and the finite volume method to substantiate the theoretical results.

Chapter 7 presents the summary and conclusion of the present work and the possible extension of our work along with the future plan of work. An appendix is presented that proves some technical results used in the thesis, a general construction to get a companion operator for any HDM. The L^∞ error estimate and bound for the mixed hybrid mimetic schemes in the GDM framework are also derived.

1.4 Preliminaries

Standard notion applies to Lebesgue and Sobolev spaces [16, 57] in this dissertation. Let Ω be a bounded domain in $\mathbb{R}^d$ $(d \geq 1)$ with boundary $\partial\Omega$, where d is the dimension. The norm in

$L^2(\Omega)$, $L^2(\Omega)^d$ for vector-valued functions, and $L^2(\Omega;\mathbb{R}^{d\times d})$ for matrix-valued functions, are denoted by $\|\cdot\|$. For $r>0$, the norm in $L^4(\Omega)^r$ is denoted by $\|\cdot\|_{L^4}$. Let $\mathcal{S}_d(\mathbb{R})$ be the set of symmetric matrices. The Euclidean norm on $\mathbb{R}^d$ is denoted by $|\cdot|$, as is the induced norm on $\mathcal{S}_d(\mathbb{R})$. The Lebesgue measure of a measurable set $E\subset\mathbb{R}^d$ is denoted by $|E|$ (note that the nature of the argument of $|\cdot|$, a vector, a matrix or a set, makes it clear if we talk about the Euclidean norm or the Lebesgue measure). For real-valued functions f,g defined on domain Ω,

$$(f,g) := \int_\Omega fg \,\mathrm{d}\boldsymbol{x}.$$

For non-negative integers m and $1\le p<\infty$, let $W^{m,p}(\Omega)$ denote the Sobolev space

$$W^{m,p}(\Omega) := \{f\in L^p(\Omega) : D^\alpha f\in L^p(\Omega),\ |\alpha|\le m\},$$

with the norm

$$\|f\|_{m,p} := \|f\|_{W^{m,p}(\Omega)} := \left(\sum_{|\alpha|\le m}\|D^\alpha f\|^p_{L^P(\Omega)}\right)^{\frac{1}{p}},$$

and for $p=\infty$,

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

where $\alpha=(\alpha_1,\cdots,\alpha_d)$ is a multi-index and the length of α is given by $|\alpha| := \sum_{i=1}^d \alpha_i$. The Hilbert spaces $W^{m,2}(\Omega)$ are denoted by $H^m(\Omega)$. The semi-norm on $W^{m,p}(\Omega)$ is denoted by $|\cdot|_{m,p,\Omega}$ and is defined by $|\varphi|_{m,p,\Omega} := |\varphi|_{W^{m,p}(\Omega)} := \left(\sum_{|\alpha|=m}|D^\alpha\varphi|^p_{L^P(\Omega)}\right)^{\frac{1}{p}}$. When $p=2$, the semi-norm is denoted by $|\cdot|_{m,\Omega}$.

Let $C(\Omega)$ represent the space of the continuous functions on Ω. Let k be a non-negative integer and $C^k(\Omega)$ denote the space of k times continuously differentiable functions on Ω. The set of all infinitely differentiable functions defined on Ω with compact support in Ω is denoted by $D(\Omega)$. Define the space $W_0^{m,p}(\Omega)$ as the closure of $D(\Omega)$ in $W^{m,p}(\Omega)$. Denote $H^{-m}(\Omega)$ to be the dual space of $H_0^m(\Omega)$ equipped with the norm

$$\|f\|_{H^{-m}(\Omega)} := \sup\left\{\frac{\langle f,g\rangle}{\|g\|_{H^m(\Omega)}} : g\in H_0^m(\Omega),\ \|g\|_{H^m(\Omega)}\neq 0\right\},$$

where $\langle\cdot,\cdot\rangle$ denotes duality pairing between $H_0^m(\Omega)$ and $H^{-m}(\Omega)$.

A fourth order symmetric *tensor* P is a linear map $\mathcal{S}_d(\mathbb{R})\to\mathcal{S}_d(\mathbb{R})$,where p_{ijkl} denote the indices of the fourth order tensor P in the canonical basis of $\mathcal{S}_d(\mathbb{R})$. For simplicity, we follow the Einstein summation convention unless otherwise stated, i.e, if an index is repeated in a product, summation is implied over the repeated index. For $\xi\in\mathcal{S}_d(\mathbb{R})$, using the definition of symmetric tensor, one has $P\xi\in\mathcal{S}_d(\mathbb{R})$ and $p_{ijkl}=p_{jikl}=p_{ijlk}$. The scalar product on $\mathcal{S}_d(\mathbb{R})$ is defined by $\xi:\phi=\xi_{ij}\phi_{ij}$. For a function $\xi:\Omega\to\mathcal{S}_d(\mathbb{R})$, denoting the differential operator by $\mathcal{H}$ we set $\mathcal{H}:\xi=\partial_{ij}\xi_{ij}$. Finally, the transpose P^τ of P is given by $P^\tau=(p_{klij})$, if $P=(p_{ijkl})$. Note that $P^\tau\xi:\phi=\xi:P\phi$. The tensor product $a\otimes b$ of two vectors $a,b\in\mathbb{R}^d$ is the 2-tensor with coefficients a_ib_j.

Definition 1.4.1 (Polytopal mesh [48, Definition 7.2])**.** *Let Ω be a bounded polytopal open subset of $\mathbb{R}^d$ ($d \geq 1$). A polytopal mesh of Ω is $\mathcal{T} = (\mathcal{M}, \mathcal{F}, \mathcal{P})$, where:*

1. *$\mathcal{M}$ is a finite family of non empty connected polytopal open disjoint subsets of Ω (the cells) such that $\overline{\Omega} = \cup_{K\in\mathcal{M}} \overline{K}$. For any $K \in \mathcal{M}$, $|K| > 0$ is the measure of K, h_K denotes the diameter of K, $\bar{\boldsymbol{x}}_K$ is the center of mass of K, and $\boldsymbol{n}_K$ is the outer unit normal to K.*

2. *$\mathcal{F}$ is a finite family of disjoint subsets of $\overline{\Omega}$ (the edges of the mesh in 2D, the faces in 3D), such that any $\sigma \in \mathcal{F}$ is a non empty open subset of a hyperplane of $\mathbb{R}^d$ and $\sigma \subset \overline{\Omega}$. Assume that for all $K \in \mathcal{M}$ there exists a subset $\mathcal{F}_K$ of $\mathcal{F}$ such that the boundary of K is $\bigcup_{\sigma\in\mathcal{F}_K} \overline{\sigma}$. We then set $\mathcal{M}_\sigma = \{K \in \mathcal{M}\,;\, \sigma \in \mathcal{F}_K\}$ and assume that, for all $\sigma \in \mathcal{F}$, $\mathcal{M}_\sigma$ has exactly one element and $\sigma \subset \partial\Omega$, or $\mathcal{M}_\sigma$ has two elements and $\sigma \subset \Omega$. Let $\mathcal{F}_{\rm int}$ be the set of all interior faces, i.e. $\sigma \in \mathcal{F}$ such that $\sigma \subset \Omega$, and $\mathcal{F}_{\rm ext}$ the set of boundary faces, i.e. $\sigma \in \mathcal{F}$ such that $\sigma \subset \partial\Omega$. The $(d-1)$-dimensional measure of $\sigma \in \mathcal{F}$ is $|\sigma|$, and its centre of mass is $\bar{\boldsymbol{x}}_\sigma$.*

3. *$\mathcal{P} = (\boldsymbol{x}_K)_{K\in\mathcal{M}}$ is a family of points of Ω indexed by $\mathcal{M}$ and such that, for all $K \in \mathcal{M}$, $\boldsymbol{x}_K \in K$. Assume that any cell $K \in \mathcal{M}$ is strictly $\boldsymbol{x}_K$-star-shaped, meaning that if $\boldsymbol{x} \in \overline{K}$ then the line segment $[\boldsymbol{x}_K, \boldsymbol{x})$ is included in K.*

The diameter of such a polytopal mesh is $h = \max_{K\in\mathcal{M}} h_K$.

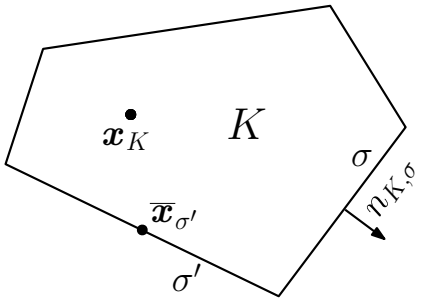

Figure 1.1: A cell K of a polytopal mesh

Figure 1.1 illustrates an example of a cell K of a polytopal mesh $\mathcal{T}$. For all $K \in \mathcal{M}$, set $\mathcal{F}_{K,\rm int} = \mathcal{F}_K \cap \mathcal{F}_{\rm int}$ and $\mathcal{F}_{K,\rm ext} = \mathcal{F}_K \cap \mathcal{F}_{\rm ext}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{F}_K$, let $\boldsymbol{n}_{K,\sigma}$ be the unit vector normal to σ outward to K. For all $\sigma \in \mathcal{F}$, choose an orientation (that is, a cell K such that $\sigma \in \mathcal{F}_K$) and then set $\boldsymbol{n}_\sigma = \boldsymbol{n}_{K,\sigma}$.
We assume that the mesh $\mathcal{M}$ satisfies minimal regularity assumptions, namely, $\bar{\boldsymbol{x}}_K \in K$ and denoting by $\rho_K = \max\{r > 0\,;\, B(\bar{\boldsymbol{x}}_K, r) \subset K\}$ the maximal radius of balls centred at $\bar{\boldsymbol{x}}_K$ and included in K, then there exists $\eta > 0$ (independent of h) such that

$$\forall K \in \mathcal{M}\,,\ \eta \geq \frac{h_K}{\rho_K}. \tag{1.4.1}$$

The set of internal vertices of $\mathcal{M}$ (resp. vertices on the boundary) is denoted by $\mathcal{V}_{\text{int}}$ (resp. $\mathcal{V}_{\text{ext}}$). Let h_σ denotes the diameter of $\sigma \in \mathcal{F}$. Let $k \geq 0$ be an integer and $K \in \mathcal{M}$. Let the space of polynomials of degree at most k in K be denoted by $\mathbb{P}_k(K)$ and $\mathbb{P}_k(\mathcal{M})$ be the broken polynomial space defined by $\mathbb{P}_k(\mathcal{M}) := \{v \in L^\infty(\Omega) : v_{|K} \in \mathbb{P}_k(K)\ \forall K \in \mathcal{M}\}$.

Some standard results

We state some results which are very frequently used in this dissertation.

- **Young's Inequality.** For all non-negative real numbers a, b and positive ε,

$$ab \leq \frac{\varepsilon a^2}{2} + \frac{b^2}{2\varepsilon}.$$

- **Discrete Hölder's inequality.** Let $1 \leq p, q < \infty$ be such that $1/p + 1/q = 1$. Suppose that $\{a_i\}_{i=1}^N$ and $\{b_i\}_{i=1}^N$ are positive numbers. Then

$$\sum_{i=1}^N a_i b_i \leq \left(\sum_{i=1}^N a_i^p\right)^{1/p} \left(\sum_{i=1}^N b_i^q\right)^{1/q}.$$

- **Hölder's inequality.** [16, p. 24] For $1 \leq p, q \leq \infty$ such that $1 = 1/p + 1/q$, if $\phi \in L^p(\Omega)$ and $\psi \in L^q(\Omega)$, then $\phi\psi \in L^1(\Omega)$ and

$$\|\phi\psi\|_{L^1(\Omega)} \leq \|\phi\|_{L^p(\Omega)} \|\psi\|_{L^q(\Omega)}.$$

- **Generalized Hölder's inequality.** Let $1 \leq p, q, r < \infty$ be such that $1/p + 1/q + 1/r = 1$. Suppose that $\phi \in L^p(\Omega)$, $\psi \in L^q(\Omega)$ and $\chi \in L^r(\Omega)$. Then

$$\|\phi\psi\chi\|_{L^1(\Omega)} \leq \|\phi\|_{L^p(\Omega)} \|\psi\|_{L^q(\Omega)} \|\chi\|_{L^r(\Omega)}.$$

- **Schauder's fixed point theorem.** [57, p. 502] Let X be a real Banach space. Suppose $K \subset X$ is compact and convex, and assume that $S : K \to K$ is continuous. Then S has a fixed point in K.

- **Divergence-free property.** [57, p. 440] Let $\xi : \Omega \to \mathbb{R}^2$ be a smooth vector-valued function. Then the cofactor matrix $\text{cof}(D\xi)$ of the gradient matrix $D\xi$ of ξ satisfies the following divergence-free row property:

$$\text{div}(\text{cof}(D\xi))_i = \sum_{j=1}^2 \frac{\partial}{\partial x_j} (\text{cof}(D\xi))_{ij} = 0 \text{ for } i = 1, 2,$$

where $(\text{cof}(D\xi))_i$ and $(\text{cof}(D\xi))_{ij}$ denote the i-th row and the (i, j)-th entry of $\text{cof}(D\xi)$, respectively.

- **Poincaré Inequality.** For all $u \in W_0^{1,p}(\Omega)$, we have

$$\|u\|_{L^p(\Omega)} \leq \operatorname{diam}(\Omega)\|\nabla u\|_{L^p(\Omega)^d}.$$

- **The Trace Inequality.** [110, pp. 87] The following discrete and trace inequalities hold for triangles (resp. tetrahedra) and rectangles (resp. hexahedra) in 2D (resp. 3D) under the regularity assumption (1.4.1). For every element K, every edge $\sigma \in \mathcal{F}_K$, and every function $v \in H^1(K)$, the following trace inequality holds:

$$\|v\|_{L^2(\sigma)} \leq C\big(h_K^{-1/2}\|v\|_{L^2(K)} + h_K^{1/2}\|\nabla v\|_{L^2(K)}\big),$$

  where $C > 0$ is a constant independent of h.

- **Discrete Inverse Inequality.** [45, pp. 26] For all $v_h \in \mathbb{P}_k(\mathcal{M})$ and all $K \in \mathcal{M}$, there exists a constant $C > 0$ independent of h such that

$$\|\nabla v_h\|_{L^2(K)^d} \leq C h_K^{-1}\|v_h\|_{L^2(K)}.$$

- **Discrete Trace Inequality.** [45, pp. 27] For all $v_h \in \mathbb{P}_k(\mathcal{M})$, all $K \in \mathcal{M}$ and all $\sigma \in \mathcal{F}_K$, there exists a constant $C > 0$ independent of h such that

$$\|v_h\|_{L^2(\sigma)} \leq C h_K^{-1/2}\|v_h\|_{L^2(K)}.$$

- **Existence and uniqueness of the solution to optimal control problems.** [57, Theorem 2.14] Let $\{U, \|\cdot\|_U\}$ and $\{H, \|\cdot\|_H\}$ denote real Hilbert spaces and let a nonempty, closed and convex set $\mathcal{U}_{\text{ad}} \subset U$, as well as some $y_d \in H$ and constant $\alpha > 0$ be given. Moreover, $S : U \to H$ be a continuous linear operator. Then the quadratic Hilbert space optimization problem

$$\min_{u \in \mathcal{U}_{\text{ad}}} f(u) := \frac{1}{2}\|Su - y_d\|_H^2 + \frac{\alpha}{2}\|u\|_U^2$$

  admits a unique optimal solution $\overline{u}$.

Chapter 2

The Hessian discretisation method for fourth order linear elliptic equations

This chapter is devoted to the study of Hessian discretisation method (HDM) which covers several numerical schemes and establishes convergence analysis in an abstract framework for fourth order linear elliptic partial differential equations[1].

2.1 Introduction

In this chapter, a generic analysis framework, the Hessian discretisation method (HDM) is designed. It applies for fourth order linear elliptic equations and is based on three abstract properties namely coercivity, consistency, and limit-conformity that ensure the scheme's convergence. Some examples that fit in this approach include conforming and non-conforming finite element methods, finite volume methods and a novel scheme based on conforming $\mathbb{P}_1$ finite element space and gradient recovery operators.

The principle of the HDM is to describe a numerical method using a set of four discrete objects, together called a Hessian discretisation (HD): the space of unknowns, and three operators reconstructing respectively a function, a gradient and a Hessian. Each choice of HD corresponds to a specific numerical scheme. The beauty of the HDM framework is that it identifies the three aforementioned model-independent properties on an HD that ensure that the corresponding scheme converges for a variety of linear models.

Note that the interest of the HDM is that it extends the analysis beyond the setting of FEMs. In particular, it covers situations where the second Strang lemma [106, 107] cannot be applied either because the continuous bilinear form cannot be extended to the space of discrete functions, and

[1] Some of the results in this chapter are published in *Jérôme Droniou, Bishnu. P. Lamichhanne and Devika Shylaja. The Hessian discretisation method for fourth order linear elliptic equations. Journal of Scientific Computing, 32p, 2018. DOI: 10. 1007/s10915-018-0814-7. URL: https://arxiv.org/abs/1803.06985* and the remaining results are communicated in *Devika Shylaja. Improved L^2 and H^1 error estimates for the Hessian discretisation method, 2019. URL: https://arxiv.org/abs/1811.05429.*

match with the discrete bilinear form, or even because the discrete space used in the scheme is not a space of functions (and the sum of the continuous and discrete spaces does not make sense). The HDM is an extension to fourth-order equations of the gradient discretisation method [48], developed for linear and non-linear second-order elliptic and parabolic problems.

Finite element methods have been studied extensively for fourth order linear problems. Conforming finite element (for example, the Argyris triangle, the Bogner–Fox–Schmit rectangle) methods for fourth order elliptic equations requires the approximation space to be a subspace of $H_0^2(\Omega)$, which results in C^1 finite elements that is cumbersome for implementations [41, 46, 101]. The nonconforming Morley elements which are based on piecewise quadratic polynomials are simpler to use and have fewer degrees of freedom (6 degrees of freedom in a triangle). The Adini element is a well-known nonconforming finite element on rectangular meshes with 12 degrees of freedom in a rectangle. For an analysis of finite element approximation by a mixed method, see [20, 63]. In [83], a finite element approximation based on gradient recovery (GR) operator for a biharmonic problem using biorthogonal system has been studied, where the approximation properties of the GR operator ensure the optimality of the finite element approach. The GR operator maps an L^2 function to a piecewise linear globally continuous H^1 function, which enables to define a Hessian matrix starting from $\mathbb{P}_1$ functions, see [81–83] for more details. A cell-centered finite volume scheme for the approximation of a biharmonic problem has been proposed and analyzed in [59], first on grids which satisfy an orthogonality condition, and then on general meshes. This scheme is based on approximations by piecewise constant functions and is thus easy to implement and computationally cheap.

This chapter is organised as follows. In Section 2.2, the model problem is introduced and some important examples of fourth order linear problems are listed. The Hessian discretisation method is presented in Section 2.3, together with some examples of HDM. In Section 2.4, a generic error estimate is established in L^2, H^1 and H^2-like norms using only three measures of accuracy in the HDM framework and applied it to various schemes. It is shown that the error estimate established in the HDM slightly improves the estimates found in [59], see Remark 2.4.13 below. Section 2.5 deals with improved L^2 error estimates for the HDM and the improved H^1-like error estimate is presented in Section 2.6. The Aubin–Nitsche duality arguments are applied to establish the improved L^2 estimate and these have higher order of convergence compared to the estimate in the energy norm in the abstract framework. However, for the H^1-like error estimate, this is not straightforward. Under the assumption that there exists a companion operator that lifts the discrete space to the continuous space with certain properties, an improved H^1-like error estimate is proved in the abstract setting. These estimates are illustrated for some schemes contained in the HDM framework. Since an improved L^2 estimate is not true in general for FVM even in the case of second order problems ([54] and references therein), a modified FVM is designed in which only the right-hand side in the Hessian scheme is modified and an L^2 superconvergence result is proved for this modified scheme. Numerical results are presented to substantiate the theoretical convergence rate established in the HDM for the gradient recovery method, finite volume method and modified finite volume method in Section 2.7.

2.2 Model problem

Let $\Omega \subset \mathbb{R}^d (d \geq 1)$ be a bounded domain with boundary $\partial\Omega$ and consider the following fourth order model problem with clamped boundary conditions.

$$\sum_{i,j,k,l=1}^{d} \partial_{kl}(a_{ijkl}\partial_{ij}\overline{u}) = f \quad \text{in } \Omega, \tag{2.2.1a}$$

$$\overline{u} = \frac{\partial \overline{u}}{\partial n} = 0 \quad \text{on } \partial\Omega, \tag{2.2.1b}$$

where $f \in L^2(\Omega)$, n is the unit outer normal to Ω, $\boldsymbol{x} = (x_1, x_2, ..., x_d) \in \Omega$ and the coefficients $a_{ijkl}(\boldsymbol{x})$ are measurable bounded functions which satisfy the conditions $a_{ijkl} = a_{jikl} = a_{ijlk} = a_{klij}$ for $i, j, k, l = 1, \cdots, d$. For all $\xi, \phi \in \mathcal{S}_d(\mathbb{R})$, assume the existence of a fourth order tensor B such that $A\xi : \phi = B\xi : B\phi$, where A is the four-tensor with indices a_{ijkl}. Note that $B\xi : B\phi = B^\tau B\xi : \phi$, so that $A = B^\tau B$.

Setting

$$V = H_0^2(\Omega) = \left\{ v \in H^2(\Omega); v = \frac{\partial v}{\partial n} = 0 \text{ on } \partial\Omega \right\} = \left\{ v \in H^2(\Omega); v = |\nabla v| = 0 \text{ on } \partial\Omega \right\},$$

the weak formulation of (2.2.1) is

$$\text{Find } \overline{u} \in V \text{ such that } \forall v \in V, \quad a(\overline{u}, v) = \int_\Omega f v \, \mathrm{d}\boldsymbol{x}, \tag{2.2.2}$$

where $a(\overline{u}, v) = \int_\Omega \mathcal{H}^B \overline{u} : \mathcal{H}^B v \, \mathrm{d}\boldsymbol{x}$ with $\mathcal{H}^B v = B\mathcal{H}v$. Note that $\int_\Omega \mathcal{H}^B \overline{u} : \mathcal{H}^B v \, \mathrm{d}\boldsymbol{x} = \int_\Omega A\mathcal{H}\overline{u} : \mathcal{H}v \, \mathrm{d}\boldsymbol{x}$, since $A = B^\tau B$. We assume in the following that B is constant over Ω, and that the following coercivity property holds:

$$\exists \rho > 0 \text{ such that } \|\mathcal{H}^B v\| \geq \rho \|v\|_{H^2(\Omega)} \quad \forall v \in H_0^2(\Omega). \tag{2.2.3}$$

Hence, the weak formulation (2.2.2) has a unique solution by the Lax–Milgram lemma.

Remark 2.2.1. *Adapting the analysis of Section 2.3 to B dependent on $\boldsymbol{x} \in \Omega$ is easy, provided the entries of B belong to $W^{2,\infty}(\Omega)$.*

2.2.1 Examples

Two specific examples of the abstract problem (2.2.1) are given now.

Biharmonic problem

Given $f \in L^2(\Omega)$, the biharmonic problem seeks u such that

$$\Delta^2 u = f \text{ in } \Omega, \qquad u = \frac{\partial u}{\partial n} = 0 \text{ in } \partial\Omega \tag{2.2.4}$$

where the biharmonic operator Δ^2 is defined by $\Delta^2\phi = \phi_{xxxx} + \phi_{yyyy} + 2\phi_{xxyy}$. The weak formulation of this model is given by (2.2.2) provided that B is chosen to satisfy

$$\int_\Omega \mathcal{H}^B u : \mathcal{H}^B v \,\mathrm{d}\boldsymbol{x} = \int_\Omega \Delta u \Delta v \,\mathrm{d}\boldsymbol{x}.$$

One possible choice of B is therefore to set $B\xi = \frac{\mathrm{tr}(\xi)}{\sqrt{d}}\mathrm{Id}$ for $\xi \in \mathcal{S}_d(\mathbb{R})$ (where Id is the identity matrix), in which case $\mathcal{H}^B = \Delta$. When Ω is convex, this choice of B satisfies (2.2.3). Since $\int_\Omega \Delta u \Delta v \,\mathrm{d}\boldsymbol{x} = \int_\Omega \mathcal{H}u : \mathcal{H}v \,\mathrm{d}\boldsymbol{x}$, another possibility is to set B as the identity tensor ($B\xi = \xi$), in which case $\mathcal{H}^B = \mathcal{H}$. By the Poincaré inequality, (2.2.3) is satisfied.

Plate problem

The clamped plate problem [41, Chapter 6] corresponds to (2.2.2) with $d = 2$ and left-hand side $a(u,v)$ is given by

$$\int_\Omega \Delta u \Delta v + (1-\gamma)(2\partial_{12}u\partial_{12}v - \partial_{11}u\partial_{22}v - \partial_{22}u\partial_{11}v)\,\mathrm{d}\boldsymbol{x}. \tag{2.2.5}$$

Here, the constant γ is the Poisson's ratio which lies in the interval $(0, \frac{1}{2})$. Note that (2.2.5) is equal to $\int_\Omega A\mathcal{H}u : \mathcal{H}v \,\mathrm{d}\boldsymbol{x}$, where the fourth order tensor A has non-zero indices $a_{1111} = 1$, $a_{2222} = 1$, $a_{1212} = (1-\gamma)$, $a_{2121} = (1-\gamma)$, $a_{1122} = \gamma$ and $a_{2211} = \gamma$. Its 'square root' can be defined as the tensor B with non-zero indices $b_{1111} = b_{2222} = \sqrt{\frac{1+\sqrt{1-\gamma^2}}{2}}$, $b_{1122} = b_{2211} = \sqrt{\frac{1-\sqrt{1-\gamma^2}}{2}}$ and $b_{1212} = b_{2121} = \sqrt{1-\gamma}$. It can be checked that (2.2.3) holds since, for some $\rho > 0$, $A\boldsymbol{\xi} : \boldsymbol{\xi} \geq \rho^2|\boldsymbol{\xi}|^2$ for all $\boldsymbol{\xi} \in \mathcal{S}_d(\mathbb{R})$.

2.3 The Hessian discretisation method

In this section, the Hessian discretisation method and some examples are presented.

Definition 2.3.1 (B–Hessian discretisation)**.** *A B–Hessian discretisation for clamped boundary conditions is a quadruplet $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_\mathcal{D}, \nabla_\mathcal{D}, \mathcal{H}^B_\mathcal{D})$ such that*

- *$X_{\mathcal{D},0}$ is a finite-dimensional space encoding the unknowns of the method,*
- *$\Pi_\mathcal{D} : X_{\mathcal{D},0} \to L^2(\Omega)$ is a linear mapping that reconstructs a function from the unknowns,*
- *$\nabla_\mathcal{D} : X_{\mathcal{D},0} \to L^2(\Omega)^d$ is a linear mapping that reconstructs a gradient from the unknowns,*
- *$\mathcal{H}^B_\mathcal{D} : X_{\mathcal{D},0} \to L^2(\Omega;\mathbb{R}^{d\times d})$ is a linear mapping that reconstructs a discrete version of $\mathcal{H}^B := B\mathcal{H}$ from the unknowns. It must be chosen such that $\|\cdot\|_\mathcal{D} := \|\mathcal{H}^B_\mathcal{D}\cdot\|$ is a norm on $X_{\mathcal{D},0}$.*

Remark 2.3.2 (Dependence of the Hessian discretisation on B)**.** *In the (2nd order) gradient discretisation method, the definition of a gradient discretisation is independent of the differential*

operator. Here, the definition of Hessian discretisation depends on B, that appears in the differential operator. This is justified by the fact that some methods (such as the finite volume method presented in Section 2.3.1) are not built on an approximation of the entire Hessian of the functions, but only on some of their derivatives (such as the Laplacian of the functions). Although it might be possible to enrich these methods by adding approximations of the 'missing' second order derivatives (as done in [47] in the context of the GDM), it does not seem to be the most natural way to proceed, and it leads to additional technicality in the analysis. Making the definition of HD dependent on the considered model through B enables us to more naturally embed some known methods into the HDM.
Note however that a number of FEMs provide approximations of the entire Hessian of the functions (see Sections 2.3.1). For those methods, a B-Hessian discretisation is built from an Id*-Hessian discretisation (that is independent of the model) by setting* $\mathcal{H}_{\mathcal{D}}^{B} = B\mathcal{H}_{\mathcal{D}}^{\mathrm{Id}}$.

If $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}}^{B})$ is a B–Hessian discretisation, the corresponding scheme for (2.2.1), called Hessian scheme (HS), is given by

$$\begin{aligned} &\text{Find } u_{\mathcal{D}} \in X_{\mathcal{D},0} \text{ such that for any } v_{\mathcal{D}} \in X_{\mathcal{D},0}, \\ &a_{\mathcal{D}}(u_{\mathcal{D}}, v_{\mathcal{D}}) = \int_{\Omega} f \Pi_{\mathcal{D}} v_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x}, \end{aligned} \tag{2.3.1}$$

where $a_{\mathcal{D}}(u_{\mathcal{D}}, v_{\mathcal{D}}) = \int_{\Omega} \mathcal{H}_{\mathcal{D}}^{B} u_{\mathcal{D}} : \mathcal{H}_{\mathcal{D}}^{B} v_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x}$. This HS is obtained by replacing, in the weak formulation (2.2.2), the continuous space V by $X_{\mathcal{D},0}$, and by using the reconstructions $\Pi_{\mathcal{D}}$ and $\mathcal{H}_{\mathcal{D}}^{B}$ in lieu of the function and its Hessian.

2.3.1 Examples

This subsection presents particular HDMs. The first set of examples discuss conforming and non-conforming finite element methods that fit into the HDM framework. Next is a novel scheme that is based on gradient recovery operators, that are constructed using biorthogonal basis. Then, we show that a finite volume method is an example of HDM.

Classical FEMs fitting into the HDM

It is shown that well-known finite element schemes fit into the Hessian discretisation method with $d = 2$, that is, they are Hessian schemes for particular choices of Hessian discretisations.

CONFORMING METHODS:
For conforming finite elements, the finite element space V_h is a subspace of the underlying Hilbert space $H_0^2(\Omega)$. The B–Hessian discretisation is defined by $X_{\mathcal{D},0} = V_h$ and, for $v_{\mathcal{D}} \in X_{\mathcal{D},0}$, $\Pi_{\mathcal{D}} v_{\mathcal{D}} = v_{\mathcal{D}}$, $\nabla_{\mathcal{D}} v_{\mathcal{D}} = \nabla v_{\mathcal{D}}$ and $\mathcal{H}_{\mathcal{D}}^{B} v_{\mathcal{D}} = \mathcal{H}^{B} v_{\mathcal{D}}$.
Recall the polytopal mesh defined in Chapter 1 (Definition 1.4.1). Three finite elements that meet this requirement are the Argyris, Hsieh-Clough-Toucher and Bogner-Fox-Schmit finite elements.

• THE ARGYRIS TRIANGLE [41]: The Argyris triangle (see Figure 2.1) is a triplet $(K, \mathbb{P}_K, \Sigma_K)$ where K is a triangle with vertices a_1, a_2, a_3 and $a_{ij} = \frac{1}{2}(a_i + a_j)$, $1 \le i < j \le 3$ denote the

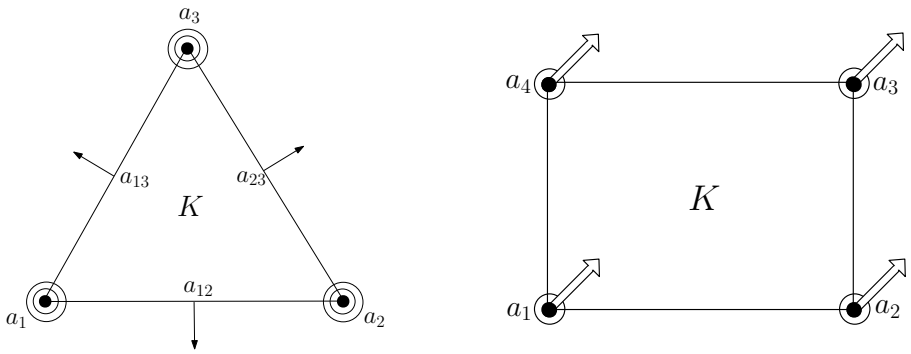

Figure 2.1: Argyris triangle and Bogner-Fox-Schmit rectangle

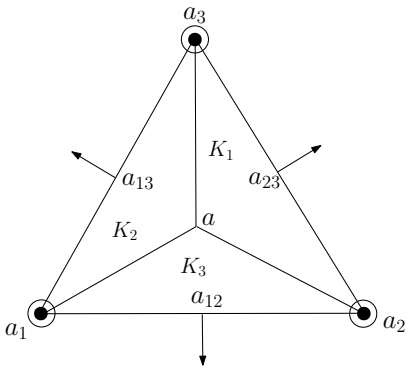

Figure 2.2: Hsieh-Clough-Toucher triangle

midpoints of the edges of K, $\mathbb{P}_K = \mathbb{P}_5(K)$, space of all polynomials of total degree ≤ 5 in two variables defined on K (dim $\mathbb{P}_K = 21$), and Σ_K denote the degrees of freedom given by: for $p \in \mathbb{P}_K$,

$$\Sigma_K = \left\{ p(a_i),\, \partial_1 p(a_i), \partial_2 p(a_i),\, \partial_{11} p(a_i),\, \partial_{12} p(a_i),\, \partial_{22} p(a_i),\, 1 \leq i \leq 3; \frac{\partial p}{\partial n}(a_{ij}),\, 1 \leq i < j \leq 3 \right\}.$$

A modification to the Argyris triangle is the Bell's element [41] which suppresses the values of the normal derivatives at the three midpoint sides, and reduces the number of degrees of freedom to 18 per element.

• BOGNER-FOX-SCHMIT RECTANGLE [41]: The Bogner-Fox-Schmit rectangle (see Figure 2.1) is a triplet $(K, \mathbb{P}_K, \Sigma_K)$ where K is a rectangle with vertices a_i, $1 \leq i \leq 4$, $\mathbb{P}_K = Q_3(K)$, the polynomials of degree ≤ 3 in both variables (dim $\mathbb{P}_K = 16$), and Σ_K is given by:

$$\Sigma_K = \left\{ p(a_i),\, \partial_1 p(a_i),\, \partial_2 p(a_i),\, \partial_{12} p(a_i),\, 1 \leq i \leq 4 \right\}.$$

• HSIEH-CLOUGH-TOUCHER TRIANGLE [41]: The Hsieh-Clough-Tocher (HCT) triangle is an example of composite finite element (also known as macroelement) of class C^1 (see Figure 2.2).

In the HCT triangle, the triangle $K \in \mathcal{M}$ is first decomposed into three triangles by connecting its barycenter with each of its vertices. On each of the subtriangles a cubic polynomial is constructed so that the resulting function is C^1 on the original triangle. There are a total of 12 degrees of freedom per triangle, which consist of the function values and first partial derivatives at the three vertices of the original triangle in addition to the normal derivative at the midpoints of the edges of the original triangle.

NON-CONFORMING METHODS:

Two well-known nonconforming finite elements [41] , the Adini element and the Morley element, are discussed below.

• THE ADINI RECTANGLE: Assume that Ω can be covered by mesh $\mathcal{M}$ made up of rectangles (we restrict the presentation to $d = 2$ for simplicity). The element K consists of a rectangle with vertices $\{a_i, 1 \le i \le 4\}$ (see Figure 2.3, left); the space $\mathbb{P}_K$ is given by $\mathbb{P}_K = \mathbb{P}_3(K) \oplus \{x_1 x_2^3\} \oplus \{x_1^3 x_2\}$, by which we mean polynomials of degree ≤ 4 whose only fourth-degree terms are those involving $x_1 x_2^3$ and $x_1^3 x_2$. Thus $\mathbb{P}_3 \subset \mathbb{P}_K$. The set of degrees of freedom in each cell is

$$\Sigma_K = \big\{p(a_i), \partial_1 p(a_i), \partial_2 p(a_i),\, 1 \le i \le 4\big\}.$$

The global approximation space is then given by

$$\begin{aligned} V_h =: \{v_h \in L^2(\Omega);\ & v_h|_K \in \mathbb{P}_K \,\forall K \in \mathcal{M}, v_h \text{ and } \nabla v_h \text{ are continuous at} \\ & \text{the vertices of elements in } \mathcal{M}, v_h \text{ and } \nabla v_h \text{ vanish at vertices on } \partial\Omega\}. \end{aligned}$$

Note that $V_h \subset H_0^1(\Omega) \cap C^0(\overline{\Omega})$. We define the broken B–Hessian $\mathcal{H}^B_{\mathcal{M}} : V_h \to L^2(\Omega)^{d\times d}$ by

$$\forall v_h \in V_h,\ \forall K \in \mathcal{M},\ \forall \boldsymbol{x} \in K,\ \mathcal{H}^B_{\mathcal{M}} v_h(\boldsymbol{x}) = \mathcal{H}^B(v_h|_K).$$

Definition 2.3.3 (Hessian discretisation for the Adini rectangle)**.** *Each $v_{\mathcal{D}} \in X_{\mathcal{D},0}$ is a vector of three values at each vertex of the mesh (with zero values at boundary vertices), corresponding to function and gradient values, $\Pi_{\mathcal{D}} v_{\mathcal{D}}$ is the function such that the values of $(\Pi_{\mathcal{D}} v_{\mathcal{D}})_{|K} \in \mathbb{P}_K$ and its gradient at the vertices are dictated by $v_{\mathcal{D}}$, $\nabla_{\mathcal{D}} v_{\mathcal{D}} = \nabla(\Pi_{\mathcal{D}} v_{\mathcal{D}})$, and $\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}} = \mathcal{H}^B_{\mathcal{M}}(\Pi_{\mathcal{D}} v_{\mathcal{D}})$ is the broken $\mathcal{H}^B$ of $\Pi_{\mathcal{D}} v_{\mathcal{D}}$.*

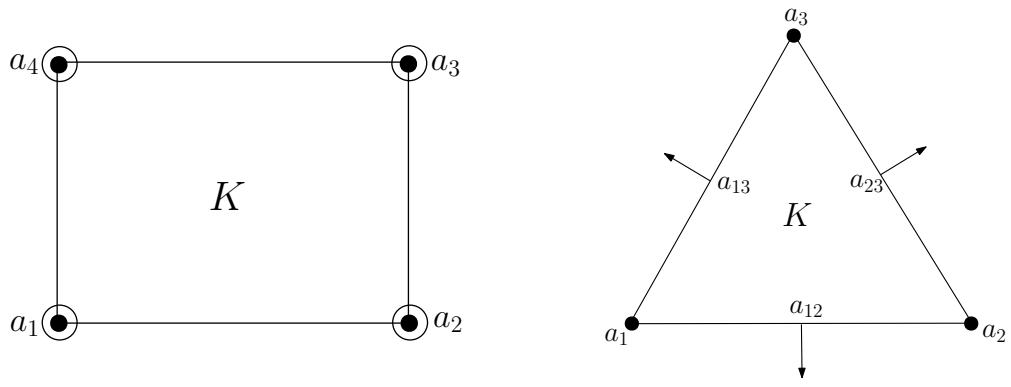

Figure 2.3: Adini rectangle and Morley triangle

• THE MORLEY ELEMENT: For $d = 2$, the Morley FEM is an example of Hessian discretisation method. For a triangle $K \in \mathcal{M}$ with vertices a_1, a_2, a_3, let a_{12}, a_{23} and a_{13} denote the midpoint of the edges opposite to the vertices a_3, a_1 and a_2, respectively (see Figure 2.3, right). The Morley finite element is a triplet $(K, \mathbb{P}_K, \Sigma_K)$ where K is a triangle in $\mathcal{M}$, $\mathbb{P}_K = \mathbb{P}_2(K)$, space of all polynomials of degree ≤ 2 in two variables defined on K (dim $\mathbb{P}_K = 6$) and Σ_K denote the degrees of freedom given by:

$$\Sigma_K = \left\{ p(a_i), 1 \leq i \leq 3; \frac{\partial p}{\partial n}(a_{ij}), 1 \leq i < j \leq 3 \right\}.$$

Then the nonconforming Morley element space associated with the mesh $\mathcal{M}$ is defined by

$$V_h =: \left\{ \phi \in \mathbb{P}_2(\mathcal{M}) | \phi \text{ is continuous at } \mathcal{V}_{\text{int}} \text{ and vanishes at } \mathcal{V}_{\text{ext}}, \right.$$
$$\left. \forall \sigma \in \mathcal{F}_{\text{int}}, \int_\sigma \left[\!\left[\frac{\partial \phi}{\partial n} \right]\!\right] ds = 0; \forall \sigma \in \mathcal{F}_{\text{ext}}, \int_\sigma \frac{\partial \phi}{\partial n} ds = 0 \right\}.$$

Definition 2.3.4 (Hessian discretisation for the Morley triangle)**.** *Each $v_\mathcal{D} \in X_{\mathcal{D},0}$ is a vector of degrees of freedom at the vertices of the mesh (with zero values at boundary vertices) and at the midpoint of the edges opposite to these vertices (with zero values at midpoint of the boundary edges). The function $\Pi_\mathcal{D} v_\mathcal{D}$ is such that $(\Pi_\mathcal{D} v_\mathcal{D})_{|K} \in \mathbb{P}_K$ with $\Pi_\mathcal{D} v_\mathcal{D}$ (resp. its normal derivatives) takes the values at the vertices (resp. at the edge midpoints) dictated by $v_\mathcal{D}$, $\nabla_\mathcal{D} v_\mathcal{D} = \nabla_\mathcal{M}(\Pi_\mathcal{D} v_\mathcal{D})$ is the broken gradient of $\Pi_\mathcal{D} v_\mathcal{D}$ and $\mathcal{H}_\mathcal{D} v_\mathcal{D} = \mathcal{H}_\mathcal{M}(\Pi_\mathcal{D} v_\mathcal{D})$ is the broken Hessian of $\Pi_\mathcal{D} v_\mathcal{D}$.*

Method based on Gradient Recovery Operators

Let V_h be an H^1_0-conforming finite element space that contains the piecewise linear functions with underlying mesh $\mathcal{M} = \mathcal{M}_h$. The gradient ∇u of $u \in V_h$ is well defined, but its second derivative $\nabla\nabla u$ is not. In order to compute some sort of second derivatives, consider a projector $Q_h : L^2(\Omega) \to V_h$, which is extended to $L^2(\Omega)^d$ component-wise. Then ∇u can be projected onto V_h^d, and the resulting function $Q_h \nabla u \in V_h^d$ is differentiable. Therefore, $\nabla(Q_h \nabla u)$ can be considered as a sort of Hessian of u. However, it not necessarily clear, for some interesting choices of practically computable Q_h (see Remark 2.3.6), that this reconstructed Hessian has proper coercivity properties (2.4.1). We therefore also consider a function $\mathfrak{S}_h$ whose role is to stabilise this reconstructed Hessian.
Let $(V_h, Q_h, I_h, \mathfrak{S}_h)$ be a quadruplet of a finite element space $V_h \subset H^1_0(\Omega)$, a reconstruction operator $Q_h : L^2(\Omega) \to V_h$ that is a projector onto V_h (that is, $Q_h = \mathrm{Id}$ on V_h), an interpolation operator $I_h : H^2_0(\Omega) \to V_h$ and a stabilisation function $\mathfrak{S}_h \in L^\infty(\Omega)^d$ such that the following properties are satisfied, with constants C not depending on h.

(P0) [*Structure of V_h and I_h*] The inverse estimate $\|\nabla z\| \leq C h^{-1} \|z\|$ holds for all $z \in V_h$ and, for $\varphi \in H^2_0(\Omega)$, we have $\|\nabla I_h \varphi - \nabla \varphi\| \leq C h \|\varphi\|_{H^2(\Omega)}$.

(P1) [*Stability of* Q_h] For $\phi \in L^2(\Omega)$, we have $\|Q_h\phi\| \leq C\|\phi\|$.

(P2) [$Q_h\nabla I_h$ *approximates* ∇] For some space W densely embedded in $H^3(\Omega)\cap H^2_0(\Omega)$ and for all $\psi \in W$, we have $\|Q_h\nabla I_h\psi - \nabla\psi\| \leq Ch^2\|\psi\|_W$.

(P3) [H^1 *approximation property of* Q_h] For $w \in H^2(\Omega)\cap H^1_0(\Omega)$, we have $\|\nabla Q_h w - \nabla w\| \leq Ch\|w\|_{H^2(\Omega)}$.

(P4) [*Asymptotic density of* $[(Q_h\nabla - \nabla)(V_h)]^\perp$] Setting $N_h = [(Q_h\nabla - \nabla)(V_h)]^\perp$, where the orthogonality is considered for the $L^2(\Omega)^d$-inner product, the following approximation property holds:

$$\inf_{\mu_h\in N_h} \|\mu_h - \varphi\| \leq Ch\|\varphi\|_{H^1(\Omega)^d},\ \forall\varphi \in H^1(\Omega)^d,$$

(P5) [*Stabilisation function*] $1 \leq |\mathfrak{S}_h| \leq C$ and, for all $K \in \mathcal{M}$, denoting by $V_h(K) = \{v_{|K}\,;\, v \in V_h\,,\ K \in \mathcal{M}\}$ the local finite element space,

$$\left[\mathfrak{S}_{h|K} \otimes (Q_h\nabla - \nabla)(V_h(K))\right] \perp \nabla V_h(K)^d,$$

where the orthogonality is understood in $L^2(K)^{d\times d}$ with the inner product induced by ":".

To construct an HD based on such a quadruplet, assume the following stronger form of (2.2.3):

$$\exists C_B > 0 : |B\boldsymbol{\xi}| \geq C_B|\boldsymbol{\xi}|\,, \quad \forall\boldsymbol{\xi} \in \mathcal{S}_d(\mathbb{R}). \tag{2.3.2}$$

Definition 2.3.5 (B–Hessian discretisation using gradient recovery)**.** *Under Assumption* (2.3.2)*, the B-Hessian discretisation based on a quadruplet* $(V_h, Q_h, I_h, \mathfrak{S}_h)$ *satisfying* **(P0)–(P5)** *is defined by:* $X_{\mathcal{D},0} = V_h$ *and, for* $u \in X_{\mathcal{D},0}$*,*

$$\Pi_{\mathcal{D}}u = u\,,\ \nabla_{\mathcal{D}}u = Q_h\nabla u \text{ and } \mathcal{H}^B_{\mathcal{D}}u = B\left[\nabla(Q_h\nabla u) + \mathfrak{S}_h \otimes (Q_h\nabla u - \nabla u)\right].$$

Remark 2.3.6. *A classical operator* Q_h *that satisfies these assumptions, for standard finite element spaces* V_h*, is the* L^2*-orthogonal projector on* V_h*. This operator is however non-local and complicated to compute. We present below a much more efficient construction of* Q_h*, local and based on biorthogonal bases.*

A GRADIENT RECOVERY OPERATOR BASED ON BIORTHOGONAL SYSTEMS:

A particular case of a method based on a gradient recovery operator is presented here, using biorthogonal systems as in [83]. V_h is the conforming $\mathbb{P}_1$ finite element space on a mesh of simplices, and I_h is the Lagrange interpolation with respect to vertices of $\mathcal{M}$. We will build a locally computable projector Q_h, that is, such that determining $Q_h f$ on a cell K only requires the knowledge of f on K and its neighbouring cells.

Let $\mathcal{B}_1 := \{\phi_1, \cdots, \phi_n\}$ be the set of basis functions of V_h associated with the inner vertices in $\mathcal{M}$. Let the set $\mathcal{B}_2 := \{\psi_1, \cdots, \psi_n\}$ be the set of discontinuous piecewise linear functions biorthogonal

to $\mathcal{B}_1$ also associated with the inner vertices of $\mathcal{M}$, so that elements of $\mathcal{B}_1$ and $\mathcal{B}_2$ satisfy the biorthogonality relation

$$\int_\Omega \psi_i \phi_j \, \mathrm{d}\boldsymbol{x} = c_j \delta_{ij}, \; c_j \neq 0, \; 1 \leq i, j \leq n, \tag{2.3.3}$$

where δ_{ij} is the Kronecker symbol and $c_j = \int_\Omega \psi_j \phi_j \, \mathrm{d}\boldsymbol{x}$. Let $M_h := \mathrm{span}\{\mathcal{B}_2\}$. Such biorthogonal systems have been constructed in the context of mortar finite elements, and later extended to gradient recovery operators [76, 82, 83]. The basis functions of M_h can be defined on a reference element. For example, for the reference triangle $\widehat{K}$,

$$\widehat{\psi}_1(\boldsymbol{x}) := 3 - 4x_1 - 4x_2, \quad \widehat{\psi}_2(\boldsymbol{x}) := 4x_1 - 1, \quad \text{and} \quad \widehat{\psi}_3(\boldsymbol{x}) := 4x_2 - 1,$$

associated with its three vertices $(0,0)$, $(1,0)$ and $(0,1)$, respectively. Associated with these vertices, the basis functions of V_h on $\widehat{K}$ is given by

$$\widehat{\phi}_1 = 1 - x_1 - x_2, \quad \widehat{\phi}_2 = x_1, \quad \widehat{\phi}_3 = x_2.$$

A direct calculation yields

$$\int_{\widehat{K}} \widehat{\psi}_i \, \mathrm{d}\boldsymbol{x} = \int_{\widehat{K}} \widehat{\phi}_i \, \mathrm{d}\boldsymbol{x} = \frac{1}{6}, \quad i = 1,2,3 \;\text{ and }\; \int_{\widehat{K}} \widehat{\psi}_i \, \widehat{\phi}_j \, \mathrm{d}\boldsymbol{x} = \frac{\delta_{ij}}{6}, \quad 1 \leq i, j \leq 3.$$

For the reference tetrahedron,

$$\begin{aligned} &\widehat{\psi}_1(\boldsymbol{x}) := 4 - 5x_1 - 5x_2 - 5x_3, \quad \widehat{\psi}_2(\boldsymbol{x}) := 5x_1 - 1, \\ &\widehat{\psi}_3(\boldsymbol{x}) := 5x_2 - 1, \quad \text{and} \quad \widehat{\psi}_4(\boldsymbol{x}) := 5x_3 - 1, \end{aligned}$$

associated with its four vertices $(0,0,0)$, $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$, respectively. These basis functions satisfy

$$\sum_{i=1}^{d+1} \widehat{\psi}_i = 1. \tag{2.3.4}$$

The projection operator $Q_h : L^2(\Omega) \to V_h$ is the oblique projector onto V_h defined as: for $f \in L^2(\Omega)$, $Q_h f \in V_h$ satisfies

$$\int_\Omega (Q_h f) \, \psi_h \, \mathrm{d}\boldsymbol{x} = \int_\Omega f \, \psi_h \, \mathrm{d}\boldsymbol{x}, \quad \forall \psi_h \in M_h. \tag{2.3.5}$$

Due to the biorthogonality relation (2.3.3), Q_h is well-defined and has the explicit representation

$$Q_h f = \sum_{i=1}^{n} \frac{\int_\Omega \psi_i f \, \mathrm{d}\boldsymbol{x}}{c_i} \phi_i. \tag{2.3.6}$$

The relation (2.3.5) shows $M_h \subset [(Q_h - I)(L^2(\Omega))]^\perp$. Hence, if M_h satisfies the approximation property

$$\inf_{\alpha_h \in M_h} \|\alpha_h - \psi\| \leq Ch \|\psi\|_{H^1(\Omega)}, \quad \forall \psi \in H^1(\Omega),$$

(P4) holds. In order to get this approximation property it is sufficient that the basis functions of M_h reproduce constant functions. Let $K \in \mathcal{M}$ be an interior element not touching any boundary vertex. Due to the property (2.3.4)

$$\sum_{i=1}^{d+1} \psi_{\mathsf{v}_i} = 1 \quad \text{on} \quad K,$$

where $\{\psi_{\mathsf{v}_i}\}_{i=1}^{d+1}$ are basis functions of M_h associated with the vertices $(\mathsf{v}_1, \ldots, \mathsf{v}_{d+1})$ of K. However, this property does not hold on $K \in \mathcal{M}$ if K has one or more vertices on the boundary. The piecewise linear basis functions of M_h needs to be modified to guarantee the approximation property [80, 84]. Let $W_h \subset H^1(\Omega)$ be the lowest order finite element space including the basis functions on the boundary vertices of $\mathcal{M}$, and let $\widetilde{M}_h$ the space spanned by the discontinuous basis functions biorthogonal to the basis functions of W_h. M_h is then obtained as a modification of $\widetilde{M}_h$, by moving all vertex basis functions of this latter space to nearby internal vertices using the following three steps.

1. For a basis function $\widetilde{\psi}_k$ of $\widetilde{M}_h$ associated with a vertex v_k on the boundary we find a closest *internal* triangle or tetrahedron $K \in \mathcal{M}$ (that is, K does not have a boundary vertex).

2. Compute the barycentric coordinates $\{\alpha_{K,i}\}_{i=1}^{d+1}$ of v_k with respect to the vertices of K, and modify all the basis functions $\{\widetilde{\psi}_{K,i}\}_{i=1}^{d+1}$ of $\widetilde{M}_h$ associated with K into $\psi_{K,i} = \widetilde{\psi}_{K,i} + \alpha_{K,i}\widetilde{\psi}_k$ for $i = 1, \cdots, d+1$.

3. Remove $\widetilde{\psi}_k$ from the basis of $\widetilde{M}_h$.

An alternative way is to modify the basis functions of all triangles or tetrahedra having one or more boundary vertices as proposed in [76].

1. If all vertices $\{\mathsf{v}_i\}_{i=1}^{d+1}$ of an element $K \in \mathcal{M}$ are inner vertices, then the linear basis functions $\{\psi_{\mathsf{v}_i}\}_{i=1}^{d+1}$ of M_h on K are defined using the biorthogonal relationship (2.3.3) with the basis functions $\{\phi_{\mathsf{v}_i}\}_{i=1}^{d+1}$ of V_h.

2. If an element $K \in \mathcal{M}$ has all boundary vertices, then we find a neighbouring element $\widetilde{K}$, which has at least one inner vertex v, and we extend the support of the basis function $\psi_{\mathsf{v}} \in M_h$ associated with v to the element K by defining $\psi_{\mathsf{v}} = 1$ on K.

3. If an element $K \in \mathcal{M}$ has only one inner vertex v and other boundary vertices, then the basis function $\psi_{\mathsf{v}} \in M_h$ associated with the inner vertex v is defined as $\psi_{\mathsf{v}} = 1$ on K.

4. If an element K has two inner vertices v_1 and v_2 and other boundary vertices, then the basis functions ψ_{v_1}, $\psi_{\mathsf{v}_2} \in M_h$ associated with these points are chosen to satisfy the biorthogonal relationship (2.3.3) with ϕ_{v_1}, $\phi_{\mathsf{v}_2} \in V_h$, as well as the property $\psi_{\mathsf{v}_1} + \psi_{\mathsf{v}_2} = 1$ on K.

5. In the three-dimensional case, we can have an element K with three inner vertices $\{\mathsf{v}_i\}_{i=1}^{3}$ and one boundary vertex. In this case we define three basis functions $\{\psi_{\mathsf{v}_i}\}_{i=1}^{3}$ to satisfy the biorthogonal relationship (2.3.3) with $\{\phi_{\mathsf{v}_i}\}_{i=1}^{3}$ as well as the condition $\sum_{i=1}^{3} \psi_{\mathsf{v}_i} = 1$ on K.

The projection Q_h is stable in L^2 and H^1-norms [82], and hence assumption **(P1)** follows. To establish **(P2)**, the following mesh assumption is needed.

(M) For any vertex v, denoting by $\mathcal{M}_\mathsf{v}$ the set of cells having v as a vertex,

$$\sum_{K\in\mathcal{M}_\mathsf{v}} \frac{|K|}{|S_\mathsf{v}|}(\bar{\boldsymbol{x}}_K - \mathsf{v}) = \mathcal{O}(h^2),$$

where S_v is the support of the basis function ϕ_v of V_h associated with v.

The assumption is required as we need to use some sort of Taylor series expansion to get the error estimate, see [112] for more details. This assumption is satisfied if the triangles of the mesh can be paired in sets of two that share a common edge and form an $\mathcal{O}(h^2)$-parallelogram, that is, the lengths of any two opposite edges differ only by $\mathcal{O}(h^2)$. In three dimensions, **(M)** is satisfied if the lengths of each pair of opposite edges of a given element are allowed to differ only by $\mathcal{O}(h^2)$ [36]. The following theorem establishes **(P2)** with $W = W^{3,\infty}(\Omega)\cap H_0^2(\Omega)$ and can be proved as in [82, 112].

Theorem 2.3.7. *Let $u \in W^{3,\infty}(\Omega)\cap H_0^2(\Omega)$. Assume that the triangulation satisfies the assumption* **(M)**. *Then*

$$\|Q_h\nabla I_h u - \nabla u\| \le Ch^2\|u\|_{W^{3,\infty}(\Omega)}.$$

Since Q_h is a projection onto V_h, $Q_hI_h = I_h$. Hence, for $w \in H^2(\Omega)\cap H_0^1(\Omega)$, introducing $Q_hI_hw = I_hw$ and invoking the H^1-stability property of Q_h [84, Lemma 1.8] leads to

$$\|\nabla Q_h w - \nabla w\| \le \|\nabla Q_h(w - I_h w)\| + \|\nabla I_h w - \nabla w\| \le C\|\nabla I_h w - \nabla w\|.$$

The standard approximation properties of V_h then guarantee **(P3)**. The Assumption **(P4)** is satisfied since $M_h \subset N_h$ (M_h is obtained by combining functions in $\widetilde{M}_h$, that satisfies this property) and the basis functions of M_h locally reproduce constant functions. To build $\mathfrak{S}_h$ that satisfies **(P5)**, divide each triangle $K \in \mathcal{M}$ into four equal triangles K_i ($i = 1,2,3,4$) using the mid-points of each side. Let K_i, $i = 1,3,4$ be the three subtriangles constructed around the vertices of K. Let $\mathfrak{S}_{h|K} = (\alpha,\beta) \in L^\infty(\Omega)^2$ and $\mathfrak{S}_{h|K_i} = (\alpha_i,\beta_i)$, $i = 1,2,3,4$. Property **(P5)** simplifies to

$$\int_K (\alpha,\beta)\cdot(p_1,p_2)\,\mathrm{d}\boldsymbol{x} = 0,$$

where $p_1\,, p_2 \in V_h$. This gives

$$\sum_{i=1}^{4}\int_{K_i} (\alpha_i,\beta_i)\cdot(p_1,p_2)\,\mathrm{d}\boldsymbol{x} = 0.$$

A use of three-point Gaussian quadrature formula in each K_i in the above estimate yields several values of $\mathfrak{S}_h$ and one such value as described in Figure 2.4 is given by

$$\mathfrak{S}_{h|K} = \begin{cases} (1,1) & \text{on } K_i,\ i = 1,3,4 \\ (-3,-3) & \text{on } K_2. \end{cases}$$

A similar construction also works on tetrahedra (in which case $\mathfrak{S}_{h|K}$ is equal to 1 on the four sub-tetrahedra constructed around the vertices of K, and -4 in the rest of K).

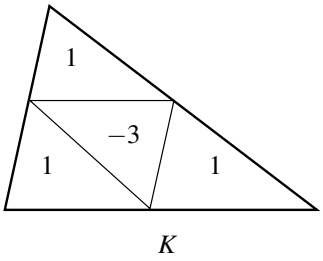

Figure 2.4: Values of the stabilisation function $\mathfrak{S}_h$ inside a cell K.

Finite volume method based on Δ-adapted discretizations

This section deals with the finite volume (FV) scheme from [59] for the biharmonic problem (2.2.4) on Δ-adapted meshes, that is, the meshes that satisfy an orthogonality property as depicted in Figure 2.5 for the two dimensional case.

Definition 2.3.8 (Δ-adapted FV mesh)**.** *A mesh $\mathcal{T} = (\mathcal{M}, \mathcal{F}, \mathcal{P})$ in the sense of Definition 1.4.1 is Δ-adapted if*

1. *for all $\sigma \in \mathcal{F}_{\text{int}}$, denoting by $K, L \in \mathcal{M}$ the cells such that $\mathcal{M}_\sigma = \{K, L\}$, the straight line $(\boldsymbol{x}_K, \boldsymbol{x}_L)$ intersects and is orthogonal to σ,*
2. *for all $\sigma \in \mathcal{F}_{\text{ext}}$ with $\mathcal{M}_\sigma = \{K\}$, the line orthogonal to σ going through $\boldsymbol{x}_K$ intersects σ.*

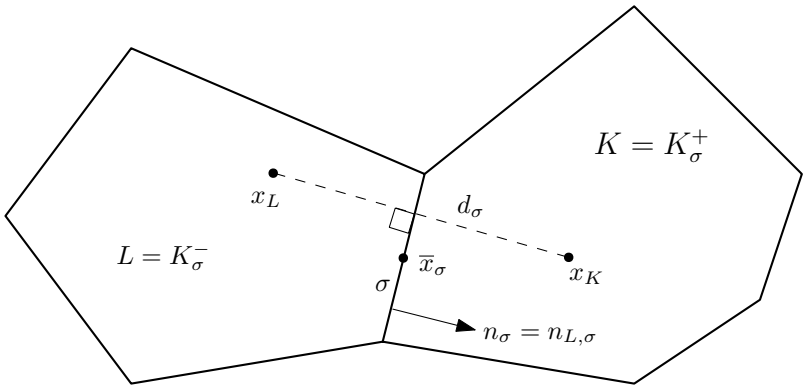

Figure 2.5: Notations for Δ-adapted discretisation

For such a mesh, we let $D_{K,\sigma}$ be the cone with vertex $\boldsymbol{x}_K$ and basis σ, and $D_\sigma = \bigcup_{K \in \mathcal{M}_\sigma} D_{K,\sigma}$. For each $\sigma \in \mathcal{F}_{\text{int}}$, an orientation is chosen by defining one of the two unit normal vectors n_σ, and let the two adjacent control volumes be denoted by K_σ^- and K_σ^+ such that n_σ is oriented from K_σ^- to K_σ^+. For all $\sigma \in \mathcal{F}_{\text{ext}}$, denote the control volume $K \in \mathcal{M}$ such that $\sigma \in \mathcal{F}_K$ by K_σ and define n_σ by $n_{K,\sigma}$. Set

$$d_\sigma = \begin{cases} \text{dist}(\boldsymbol{x}_{K_\sigma^-}, \sigma) + \text{dist}(\boldsymbol{x}_{K_\sigma^+}, \sigma) & \forall \sigma \in \mathcal{F}_{\text{int}} \\ \text{dist}(\boldsymbol{x}_K, \sigma) & \forall \sigma \in \mathcal{F}_{\text{ext}} \end{cases} \tag{2.3.7}$$

where $\mathrm{dist}(\boldsymbol{x}_K, \sigma)$ denotes the distance between $\boldsymbol{x}_K$ and σ. Finally, define the mesh regularity factor by

$$\theta_{\mathcal{T}} = \max\left\{ \max\left(\frac{\mathrm{diam}(K)}{\mathrm{dist}(\boldsymbol{x}_K, \sigma)}, \frac{d_\sigma}{\mathrm{dist}(\boldsymbol{x}_K, \sigma)} \right) ; K \in \mathcal{M}, \sigma \in \mathcal{F}_K \right\}.$$

We now define a notion of B–Hessian discretisation for $B = \frac{\mathrm{tr}(\cdot)}{\sqrt{d}}\mathrm{Id}$, in which case (2.2.2) corresponds to the biharmonic problem (2.2.4), for which the coercivity property (2.2.3) holds.

Definition 2.3.9 (B–Hessian discretisation based on Δ-adapted discretisation)**.** *Let $B = \frac{\mathrm{tr}(\cdot)}{\sqrt{d}}\mathrm{Id}$ and $\mathcal{T}$ be a Δ-adapted mesh. A B–Hessian discretisation is given by $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}}, \mathcal{H}^B_{\mathcal{D}})$ where*

- *$X_{\mathcal{D},0}$ is the space of all real families $u_{\mathcal{D}} = (u_K)_{K\in\mathcal{M}}$, such that $u_K = 0$ for all $K \in \mathcal{M}$ with $\mathcal{F}_{K,\mathrm{ext}} \neq \emptyset$.*
- *For $u_{\mathcal{D}} \in X_{\mathcal{D},0}$, $\Pi_{\mathcal{D}} u_{\mathcal{D}}$ is the piecewise constant function equal to u_K on the cell K.*
- *The discrete gradient $\nabla_{\mathcal{D}} u_{\mathcal{D}}$ is defined by its constant values on the cells:*

$$\nabla_K u_{\mathcal{D}} = \frac{1}{|K|} \sum_{\sigma\in\mathcal{F}_K} \frac{|\sigma|(\delta_{K,\sigma} u_{\mathcal{D}})(\overline{\boldsymbol{x}}_\sigma - \boldsymbol{x}_K)}{d_\sigma}, \tag{2.3.8}$$

  where

$$\delta_{K,\sigma} u_{\mathcal{D}} = \begin{cases} u_L - u_K & \forall \sigma \in \mathcal{F}_{K,\mathrm{int}}\,,\ \mathcal{M}_\sigma = \{K, L\} \\ 0 & \forall \sigma \in \mathcal{F}_{K,\mathrm{ext}}. \end{cases} \tag{2.3.9}$$

- *The discrete Laplace operator $\Delta_{\mathcal{D}}$ is defined by its constant values on the cells:*

$$\Delta_K u_{\mathcal{D}} = \frac{1}{|K|} \sum_{\sigma\in\mathcal{F}_K} \frac{|\sigma|\delta_{K,\sigma} u_{\mathcal{D}}}{d_\sigma}. \tag{2.3.10}$$

  We then set $\mathcal{H}^B_{\mathcal{D}} u_{\mathcal{D}} = \frac{\Delta_{\mathcal{D}} u_{\mathcal{D}}}{\sqrt{d}}\mathrm{Id}$.

For $u_{\mathcal{D}}, v_{\mathcal{D}} \in X_{\mathcal{D},0}$,

$$[u_{\mathcal{D}}, v_{\mathcal{D}}] = \sum_{\sigma\in\mathcal{F}} \frac{|\sigma|\delta_\sigma u_{\mathcal{D}} \delta_\sigma v_{\mathcal{D}}}{d_\sigma} \tag{2.3.11}$$

defines an inner product on $X_{\mathcal{D},0}$, whose associated norm is denoted by $\|u_{\mathcal{D}}\|_{1,\mathcal{D}}$. Here δ_σ is given by

$$\delta_\sigma u_{\mathcal{D}} = \begin{cases} u_{K^+_\sigma} - u_{K^-_\sigma} & \forall \sigma \in \mathcal{F}_{\mathrm{int}} \\ 0 & \forall \sigma \in \mathcal{F}_{\mathrm{ext}}. \end{cases} \tag{2.3.12}$$

It can easily be checked that, with this Hessian discretisation, the Hessian scheme (2.2.2) is the scheme of [59] for the biharmonic equation.

2.4 Basic error estimates

The properties that are required for the convergence analysis of the Hessian scheme are listed in the first part of this section. It is shown that the accuracy of the HS (basic error estimates) can be evaluated using only three measures, all intrinsic to the Hessian discretisation. This estimate is then applied to various schemes mentioned in Section 2.3.1.
The first one is a constant, $C_{\mathcal{D}}^{B}$, which controls the norm of the linear mappings $\Pi_{\mathcal{D}}$ and $\nabla_{\mathcal{D}}$.

$$C_{\mathcal{D}}^{B} = \max_{w_{\mathcal{D}} \in X_{\mathcal{D},0} \setminus \{0\}} \left(\frac{\|\Pi_{\mathcal{D}} w_{\mathcal{D}}\|}{\|\mathcal{H}_{\mathcal{D}}^{B} w_{\mathcal{D}}\|}, \frac{\|\nabla_{\mathcal{D}} w_{\mathcal{D}}\|}{\|\mathcal{H}_{\mathcal{D}}^{B} w_{\mathcal{D}}\|} \right). \tag{2.4.1}$$

The second measure of accuracy is the interpolation error $S_{\mathcal{D}}^{B}$ defined by

$$\begin{aligned} &\forall\, \varphi \in H_0^2(\Omega)\,, \\ &S_{\mathcal{D}}^{B}(\varphi) = \min_{w_{\mathcal{D}} \in X_{\mathcal{D},0}} \Big(\|\Pi_{\mathcal{D}} w_{\mathcal{D}} - \varphi\| + \|\nabla_{\mathcal{D}} w_{\mathcal{D}} - \nabla \varphi\| + \|\mathcal{H}_{\mathcal{D}}^{B} w_{\mathcal{D}} - \mathcal{H}^{B} \varphi\| \Big). \end{aligned} \tag{2.4.2}$$

Finally, the third quantity is a measure of limit-conformity of the HD, that is, how well a discrete integration by parts formula is verified by the discrete operators:

$$\forall\, \xi \in H^{B}(\Omega)\,, W_{\mathcal{D}}^{B}(\xi) = \max_{w_{\mathcal{D}} \in X_{\mathcal{D},0} \setminus \{0\}} \frac{\left| \mathcal{W}_{\mathcal{D}}^{B}(\xi, w_{\mathcal{D}}) \right|}{\|\mathcal{H}_{\mathcal{D}}^{B} w_{\mathcal{D}}\|}, \tag{2.4.3}$$

where $H^{B}(\Omega) = \{\xi \in L^2(\Omega)^{d\times d}\,;\, \mathcal{H} : B^{\tau} B\xi \in L^2(\Omega)\}$ and

$$\mathcal{W}_{\mathcal{D}}^{B}(\xi, w_{\mathcal{D}}) = \int_{\Omega} \Big((\mathcal{H} : B^{\tau} B\xi) \Pi_{\mathcal{D}} w_{\mathcal{D}} - B\xi : \mathcal{H}_{\mathcal{D}}^{B} w_{\mathcal{D}} \Big)\, \mathrm{d}\boldsymbol{x}. \tag{2.4.4}$$

Note that if $\xi \in H^{B}(\Omega)$ and $\phi \in H_0^2(\Omega)$, integration by parts show that $\int_{\Omega} (\mathcal{H} : B^{\tau} B\xi)\phi = \int_{\Omega} B\xi : \mathcal{H}^{B}\phi$. Hence, the quantity in the right-hand side of (2.4.3) measures a defect of discrete integration-by-parts between $\Pi_{\mathcal{D}}$ and $\mathcal{H}_{\mathcal{D}}^{B}$.
Closely associated to the three measures above are the notions of coercivity, consistency and limit-conformity of a sequence of Hessian discretisations.

Definition 2.4.1 (Coercivity, consistency and limit-conformity)**.** *Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of B–Hessian discretisations in the sense of Definition 2.3.1. We say that*

1. *$(\mathcal{D}_m)_{m\in\mathbb{N}}$ is* coercive *if there exists $C_P \in \mathbb{R}^+$ such that $C_{\mathcal{D}_m}^{B} \le C_P$ for all $m \in \mathbb{N}$.*
2. *$(\mathcal{D}_m)_{m\in\mathbb{N}}$ is* consistent*, if*

$$\forall \varphi \in H_0^2(\Omega)\,, \lim_{m\to\infty} S_{\mathcal{D}_m}^{B}(\varphi) = 0. \tag{2.4.5}$$

3. *$(\mathcal{D}_m)_{m\in\mathbb{N}}$ is* limit-conforming*, if*

$$\forall \xi \in H^{B}(\Omega)\,, \lim_{m\to\infty} W_{\mathcal{D}_m}^{B}(\xi) = 0. \tag{2.4.6}$$

Remark 2.4.2. *As for the (2nd order) gradient discretisation method, see [48, Lemmas 2.16 and 2.17], it is easily proved that, for coercive sequences of HDs, the consistency and limit-conformity properties* (2.4.5) *and* (2.4.6) *only need to be tested for functions in dense subsets of $H_0^2(\Omega)$ and $H^B(\Omega)$, respectively.*

Remark 2.4.3. *If $B = \mathrm{Id}$, we write $\mathcal{H}_\mathcal{D}$ (resp. $C_\mathcal{D}$, $S_\mathcal{D}$ and $W_\mathcal{D}$) instead of $\mathcal{H}_\mathcal{D}^{\mathrm{Id}}$ (resp. $C_\mathcal{D}^{\mathrm{Id}}$, $S_\mathcal{D}^{\mathrm{Id}}$ and $W_\mathcal{D}^{\mathrm{Id}}$).*

The next theorem establishes basic error estimates for the HDM.

Theorem 2.4.4 (Error estimate for Hessian schemes)**.** *Under Assumption* (2.2.3)*, let $\overline{u}$ be the solution to* (2.2.2)*. Let $\mathcal{D}$ be a B–Hessian discretisation and $u_\mathcal{D}$ be the solution to the corresponding Hessian scheme* (2.3.1)*. Then the following error estimates hold true:*

$$\|\Pi_\mathcal{D} u_\mathcal{D} - \overline{u}\| \le C_\mathcal{D}^B W_\mathcal{D}^B(\mathcal{H}\overline{u}) + (C_\mathcal{D}^B + 1) S_\mathcal{D}^B(\overline{u}), \tag{2.4.7}$$

$$\|\nabla_\mathcal{D} u_\mathcal{D} - \nabla\overline{u}\| \le C_\mathcal{D}^B W_\mathcal{D}^B(\mathcal{H}\overline{u}) + (C_\mathcal{D}^B + 1) S_\mathcal{D}^B(\overline{u}), \tag{2.4.8}$$

$$\|\mathcal{H}_\mathcal{D}^B u_\mathcal{D} - \mathcal{H}^B\overline{u}\| \le W_\mathcal{D}^B(\mathcal{H}\overline{u}) + 2S_\mathcal{D}^B(\overline{u}). \tag{2.4.9}$$

(Note that $\mathcal{H}\overline{u} \in H^B(\Omega)$ because $\mathcal{H}\overline{u} \in L^2(\Omega)^{d\times d}$ and $\mathcal{H} : B^\tau B\mathcal{H}\overline{u} = \mathcal{H} : A\mathcal{H}\overline{u} = f \in L^2(\Omega)$.)

The following convergence result is a trivial consequence of the error estimates above.

Corollary 2.4.5 (Convergence)**.** *Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of B–Hessian discretisations that is coercive, consistent and limit-conforming. Then, as $m \to \infty$, $\Pi_{\mathcal{D}_m} u_{\mathcal{D}_m} \to \overline{u}$ in $L^2(\Omega)$, $\nabla_{\mathcal{D}_m} u_{\mathcal{D}_m} \to \nabla\overline{u}$ in $L^2(\Omega)^d$ and $\mathcal{H}_{\mathcal{D}_m}^B u_{\mathcal{D}_m} \to \mathcal{H}^B\overline{u}$ in $L^2(\Omega)^{d\times d}$.*

Let us now prove Theorem 2.4.4.

Proof of Theorem 2.4.4. For all $v_\mathcal{D} \in X_{\mathcal{D},0}$, the equation (2.2.1a) taken in the sense of distributions shows that $f = \mathcal{H} : A\mathcal{H}\overline{u}$, and thus, by the Hessian scheme (2.3.1),

$$\int_\Omega \mathcal{H}_\mathcal{D}^B u_\mathcal{D} : \mathcal{H}_\mathcal{D}^B v_\mathcal{D}\,\mathrm{d}\boldsymbol{x} = \int_\Omega f\Pi_\mathcal{D} v_\mathcal{D}\,\mathrm{d}\boldsymbol{x} = \int_\Omega (\mathcal{H} : B^\tau B\mathcal{H}\overline{u})\Pi_\mathcal{D} v_\mathcal{D}\,\mathrm{d}\boldsymbol{x}.$$

The definition (2.4.3) of $W_\mathcal{D}^B$ implies that

$$\int_\Omega \left(\mathcal{H}^B\overline{u} - \mathcal{H}_\mathcal{D}^B u_\mathcal{D}\right) : \mathcal{H}_\mathcal{D}^B v_\mathcal{D}\,\mathrm{d}\boldsymbol{x} \le W_\mathcal{D}^B(\mathcal{H}\overline{u})\|\mathcal{H}_\mathcal{D}^B v_\mathcal{D}\|. \tag{2.4.10}$$

Define the interpolant $\mathcal{P}_\mathcal{D} : H_0^2(\Omega) \to X_{\mathcal{D},0}$ by

$$\mathcal{P}_\mathcal{D}\overline{u} = \operatorname*{argmin}_{w_\mathcal{D}\in X_{\mathcal{D},0}} \left(\|\Pi_\mathcal{D} w_\mathcal{D} - \overline{u}\| + \|\nabla_\mathcal{D} w_\mathcal{D} - \nabla\overline{u}\| + \|\mathcal{H}_\mathcal{D}^B w_\mathcal{D} - \mathcal{H}^B\overline{u}\|\right)$$

and from (2.4.2), it follows that

$$\|\Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\overline{u} - \overline{u}\| + \|\nabla_\mathcal{D}\mathcal{P}_\mathcal{D}\overline{u} - \nabla\overline{u}\| + \|\mathcal{H}_\mathcal{D}^B\mathcal{P}_\mathcal{D}\overline{u} - \mathcal{H}^B\overline{u}\| \le S_\mathcal{D}^B(\overline{u}). \tag{2.4.11}$$

An introduction of intermediate term $\mathcal{H}^B\overline{u}$ and (2.4.10) leads to

$$\begin{aligned}\int_\Omega \left(\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}}\right) : \mathcal{H}^B_{\mathcal{D}}v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}&\\ = \int_\Omega \left(\mathcal{H}^B\overline{u}-\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}}\right) : \mathcal{H}^B_{\mathcal{D}}v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} + \int_\Omega \left(\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\right) : \mathcal{H}^B_{\mathcal{D}}v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}&\\ \le W^B_{\mathcal{D}}(\mathcal{H}\overline{u})\|\mathcal{H}^B_{\mathcal{D}}v_{\mathcal{D}}\| + \|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\|\,\|\mathcal{H}^B_{\mathcal{D}}v_{\mathcal{D}}\|.&\end{aligned}$$

Choose $v_{\mathcal{D}} = P_{\mathcal{D}}\overline{u} - u_{\mathcal{D}}$ in the above estimate to obtain

$$\|\mathcal{H}^B_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}-u_{\mathcal{D}})\|^2 \le W^B_{\mathcal{D}}(\mathcal{H}\overline{u})\|\mathcal{H}^B_{\mathcal{D}}(P_{\mathcal{D}}\overline{u}-u_{\mathcal{D}})\| + \|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\|\,\|\mathcal{H}^B_{\mathcal{D}}(P_{\mathcal{D}}\overline{u}-u_{\mathcal{D}})\|.$$

This and (2.4.11) imply that

$$\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}}\| \le W^B_{\mathcal{D}}(\mathcal{H}\overline{u}) + S^B_{\mathcal{D}}(\overline{u}). \tag{2.4.12}$$

A use of triangle inequality, (2.4.11) and (2.4.12) yields

$$\|\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}}-\mathcal{H}^B\overline{u}\| \le \|\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}}-\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + \|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\| \le W^B_{\mathcal{D}}(\mathcal{H}\overline{u}) + 2S^B_{\mathcal{D}}(\overline{u}),$$

which is (2.4.9). The definition of $C^B_{\mathcal{D}}$ given by (2.4.1), (2.4.11) and (2.4.12) leads to

$$\begin{aligned}\|\Pi_{\mathcal{D}}u_{\mathcal{D}}-\overline{u}\| &\le \|\Pi_{\mathcal{D}}u_{\mathcal{D}}-\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + \|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\overline{u}\|\\ &\le C^B_{\mathcal{D}}\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}}\| + S^B_{\mathcal{D}}(\overline{u}) \le C^B_{\mathcal{D}}W^B_{\mathcal{D}}(\mathcal{H}\overline{u}) + (C^B_{\mathcal{D}}+1)S^B_{\mathcal{D}}(\overline{u})\end{aligned}$$

and

$$\begin{aligned}\|\nabla_{\mathcal{D}}u_{\mathcal{D}}-\nabla\overline{u}\| &\le \|\nabla_{\mathcal{D}}u_{\mathcal{D}}-\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + \|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\nabla\overline{u}\|\\ &\le C^B_{\mathcal{D}}\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}}\| + S^B_{\mathcal{D}}(\overline{u}) \le C^B_{\mathcal{D}}W^B_{\mathcal{D}}(\mathcal{H}\overline{u}) + (C^B_{\mathcal{D}}+1)S^B_{\mathcal{D}}(\overline{u}).\end{aligned}$$

Hence, (2.4.7) and (2.4.8) are established. □

The particular HDMs given in Section 2.3.1 satisfy the required three properties for the convergence analysis to hold and are discussed below.

2.4.1 Classical FEMs

Conforming FEMs

For conforming FEMs, the estimates on the accuracy measures $C^B_{\mathcal{D}}$, $S^B_{\mathcal{D}}$ and $W^B_{\mathcal{D}}$ easily follow:

- $C^B_{\mathcal{D}}$ is bounded by the constant of the continuous Poincaré inequality in $H^2_0(\Omega)$. That is,

$$C^B_{\mathcal{D}} \le \rho^{-1}\max(\mathrm{diam}(\Omega), \mathrm{diam}(\Omega)^2).$$

- Standard interpolation properties (see, e.g., [41]) yield estimates on the consistency measure $S^B_{\mathcal{D}}$. For $\psi \in H^3(\Omega) \cap H^2_0(\Omega)$, the classical interpolant $\mathcal{P}_{\mathcal{D}}$ satisfy

$$\|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi - \psi\| \leq Ch^3\|\psi\|_{3,\Omega}, \quad \|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi - \nabla\psi\| \leq Ch^2\|\psi\|_{3,\Omega},$$
$$\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi - \mathcal{H}^B\psi\| \leq Ch\|\psi\|_{3,\Omega},$$

  which shows that $S^B_{\mathcal{D}}(\phi) \leq Ch\|\psi\|_{3,\Omega}$, where $C > 0$ is a constant independent of h.

- Integration by parts twice in $H^2_0(\Omega)$ shows that $W^B_{\mathcal{D}}(\xi) = 0$ for all $\xi \in H^B(\Omega)$.

Non-conforming FEMs

The properties of HDM that ensure the covergence analysis of the Adini rectangle and the Morley triangle are proved below. In the sequel, the positive constants C appearing in the inequalities denote generic constants, which will take different values at different places but will always be independent of the mesh size h.

THE ADINI RECTANGLE:
The following theorem talks about the three measures of accuracy of HDM for the Adini rectangle. These help in establishing the convergence of the scheme as described in Theorem 2.4.4.

Theorem 2.4.6. *Let $\mathcal{D}$ be a B–Hessian discretisation in the sense of Definition 2.3.3 with B satisfying the coercive property. Then, there exists a constant C, not depending on $\mathcal{D}$, such that*

- $C^B_{\mathcal{D}} \leq C$,
- $\forall \varphi \in H^3(\Omega) \cap H^2_0(\Omega)$, $S^B_{\mathcal{D}}(\varphi) \leq Ch\|\varphi\|_{H^3(\Omega)}$,
- $\forall \xi \in H^2(\Omega)^{d\times d}$, $W^B_{\mathcal{D}}(\xi) \leq Ch\|\xi\|_{H^2(\Omega)^{d\times d}}$.

The properties of Hessian discretisations built on the Adini rectangle follow from this theorem and Remark 2.4.2.

Corollary 2.4.7. *Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of B–Hessian discretisations built on the Adini rectangle, such that B is coercive and the underlying sequence of meshes are regular and have a size that goes to 0 as $m \to \infty$. Then the sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is coercive, consistent and limit-conforming.*

Proof of Theorem 2.4.6.
• COERCIVITY: Since the approximation space $V_h \subset H^1_0(\Omega)$, for $v \in X_{\mathcal{D},0}$, the Poincaré inequality yields $\|\Pi_{\mathcal{D}}v\| \leq \operatorname{diam}(\Omega)\|\nabla_{\mathcal{D}}v\|$, which gives us part of the estimate on $C^B_{\mathcal{D}}$. Define the *broken Sobolev* space

$$H^1(\mathcal{M}) = \left\{v \in L^2(\Omega)\,;\, \forall K \in \mathcal{M}, v_{|K} \in H^1(K)\right\}$$

and endow it with the dG norm defined by

$$\|w\|_{dG,\mathcal{M}}^2 := \|\nabla_{\mathcal{M}} w\|^2 + \sum_{\sigma\in\mathcal{F}} \frac{1}{h_\sigma} \|[\![w]\!]\|_{L^2(\sigma)}^2. \tag{2.4.13}$$

If $[\![w]\!] = 0$ at the vertices of σ then, by the Poincaré inequality in $H_0^1(\sigma)$ given by Lemma A.1.1,

$$\|[\![w]\!]\|_{L^2(\sigma)} \le h_\sigma \|\nabla_{\mathcal{M}} [\![w]\!]\|_{L^2(\sigma)^d}. \tag{2.4.14}$$

If $\sigma \in \mathcal{F}_{\rm int}$ with $\mathcal{M}_\sigma = \{K,L\}$ then $[\![w]\!] = 0$ at the vertices of σ, and (2.4.14) combined with the trace inequality [45, Lemma 1.46] therefore give

$$\begin{aligned}\|[\![w]\!]\|_{L^2(\sigma)} &\le h_\sigma (\|\nabla_{\mathcal{M}} w_{|K}\|_{L^2(\sigma)^d} + \|\nabla_{\mathcal{M}} w_{|L}\|_{L^2(\sigma)^d}) \\ &\le C_{\rm tr} h_\sigma (h_K^{-1/2} \|\nabla_{\mathcal{M}} w\|_{L^2(K)^d} + h_L^{-1/2} \|\nabla_{\mathcal{M}} w\|_{L^2(L)^d}),\end{aligned} \tag{2.4.15}$$

where $C_{\rm tr}$ depends only on d and the mesh regularity parameter η. Take $v_{\mathcal{D}} \in X_{\mathcal{D},0}$. Since $\nabla_{\mathcal{D}} v_{\mathcal{D}}$ is continuous at the vertices of elements in $\mathcal{M}$ and $\nabla_{\mathcal{D}} v_{\mathcal{D}}$ vanish at vertices along $\partial\Omega$, choose $w = \nabla_{\mathcal{D}} v_{\mathcal{D}}$ in (2.4.14) and (2.4.15) to obtain

$$\|[\![\nabla_{\mathcal{D}} v_{\mathcal{D}}]\!]\|_{L^2(\sigma)^d} \le C_{\rm tr} h_\sigma \Big(h_K^{-1/2} \|\nabla_{\mathcal{M}} (\nabla_{\mathcal{D}} v_{\mathcal{D}})\|_{L^2(K)^{d\times d}} + h_L^{-1/2} \|\nabla_{\mathcal{M}} (\nabla_{\mathcal{D}} v_{\mathcal{D}})\|_{L^2(L)^{d\times d}} \Big).$$

The definition (2.4.13) of the dG norm, the above inequality and the coercivity property of B implies that

$$\begin{aligned}\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\|_{dG,\mathcal{M}}^2 &\le \|\nabla_{\mathcal{M}} (\nabla_{\mathcal{D}} v_{\mathcal{D}})\|^2 + 2C_{\rm tr} \sum_{\sigma\in\mathcal{F}} h_\sigma \Big(h_K^{-1} \|\nabla_{\mathcal{M}} (\nabla_{\mathcal{D}} v_{\mathcal{D}})\|_{L^2(K)^{d\times d}}^2 \\ &\qquad + h_L^{-1} \|\nabla_{\mathcal{M}} (\nabla_{\mathcal{D}} v_{\mathcal{D}})\|_{L^2(L)^{d\times d}}^2 \Big) \\ &\le \|\nabla_{\mathcal{M}} (\nabla_{\mathcal{D}} v_{\mathcal{D}})\|^2 + C \sum_{K\in\mathcal{M}} \|\nabla_{\mathcal{M}} (\nabla_{\mathcal{D}} v_{\mathcal{D}})\|_{L^2(K)^{d\times d}}^2 \\ &\le C \|\mathcal{H}_{\mathcal{M}} (\Pi_{\mathcal{D}} v_{\mathcal{D}})\|^2 \le C\rho^{-2} \|\mathcal{H}_{\mathcal{M}}^B (\Pi_{\mathcal{D}} v_{\mathcal{D}})\|^2 = C\rho^{-2} \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|^2.\end{aligned}$$

Use the fact that $\|w\| \le C\|w\|_{dG,\mathcal{M}}$ whenever w is a broken polynomial on $\mathcal{M}$ (see [45, Theorem 5.3]) to deduce $\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\| \le C\rho^{-1} \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|$, which concludes the estimate on $C_{\mathcal{D}}^B$.

• CONSISTENCY: Consistency follows from the interpolation properties of the family of Adini rectangles. A use of [41, Theorem 3.1.5] leads to, for $\varphi \in H^3(\Omega) \cap H_0^2(\Omega)$,

$$\inf_{w_{\mathcal{D}}\in X_{\mathcal{D},0}} \|\mathcal{H}_{\mathcal{D}}^B w_{\mathcal{D}} - \mathcal{H}^B \varphi\| \le Ch\|\phi\|_{3,\Omega}, \quad \inf_{w_{\mathcal{D}}\in X_{\mathcal{D},0}} \|\nabla_{\mathcal{D}} w_{\mathcal{D}} - \nabla\varphi\| \le Ch^2 \|\phi\|_{3,\Omega}$$
$$\text{and} \quad \inf_{w_{\mathcal{D}}\in X_{\mathcal{D},0}} \|\Pi_{\mathcal{D}} w_{\mathcal{D}} - \varphi\| \le Ch^3 \|\phi\|_{3,\Omega},$$

which implies $S_{\mathcal{D}}^B(\varphi) \le Ch\|\phi\|_{3,\Omega}$.

• LIMIT-CONFORMITY: for $\xi \in H^2(\Omega)^{d\times d}$ and $v_{\mathcal{D}} \in X_{\mathcal{D},0}$, cellwise integration by parts (see Lemma A.1.2) yields

$$\begin{aligned}\int_\Omega (\mathcal{H} : B^\tau B\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} &= \sum_{K\in\mathcal{M}} \int_K (\mathcal{H} : A\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}\\ &= \int_\Omega A\xi : \mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} - \sum_{K\in\mathcal{M}} \int_{\partial K} (A\xi n_K)\cdot \nabla_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x})\\ &\quad + \sum_{K\in\mathcal{M}} \int_{\partial K} (\mathrm{div}(A\xi)\cdot n_K)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x}).\end{aligned}$$

This implies

$$\begin{aligned}&\int_\Omega (\mathcal{H} : A\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} - \int_\Omega A\xi : \mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}\\ &\qquad = -\sum_{\sigma\in\mathcal{F}} \int_\sigma (A\xi n_\sigma)\cdot [\![\nabla_{\mathcal{D}} v_{\mathcal{D}}]\!]\,\mathrm{d}s(\boldsymbol{x}) + \sum_{\sigma\in\mathcal{F}} \int_\sigma (\mathrm{div}(A\xi)\cdot n_\sigma)[\![\Pi_{\mathcal{D}} v_{\mathcal{D}}]\!]\,\mathrm{d}s(\boldsymbol{x}). \quad (2.4.16)\end{aligned}$$

Since $\Pi_{\mathcal{D}} v_{\mathcal{D}} \in H^1_0(\Omega)\cap C(\overline{\Omega})$, $[\![\Pi_{\mathcal{D}} v_{\mathcal{D}}]\!] = 0$. Let Λ_K denote the Q_1 interpolation operator associated with the values at the four vertices of K, and Λ_h be the patched interpolator such that $(\Lambda_h)_{|K} = \Lambda_K$ for all K. $\Lambda_h(\nabla_{\mathcal{D}} v_{\mathcal{D}})$ takes the values of $\nabla_{\mathcal{D}} v_{\mathcal{D}}$ at the vertices, so it is continuous at internal vertices and vanishes at the boundary vertices. Hence, for any $\sigma\in\mathcal{F}$, $[\![\Lambda_h(\nabla_{\mathcal{D}} v_{\mathcal{D}})]\!]$ vanishes on σ since it is linear on this edge and vanishes at its vertices. As a consequence,

$$\begin{aligned}\int_\Omega (\mathcal{H} : A\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} - \int_\Omega A\xi : \mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} &= -\sum_{\sigma\in\mathcal{F}} \int_\sigma (A\xi n_\sigma)\cdot [\![\nabla_{\mathcal{D}} v_{\mathcal{D}} - \Lambda_h(\nabla_{\mathcal{D}} v_{\mathcal{D}})]\!]\,\mathrm{d}s(\boldsymbol{x})\\ &= -\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{F}_K} \int_\sigma A\xi n_{K,\sigma}\cdot \Big(\nabla_{\mathcal{D}} v_{\mathcal{D}} - \Lambda_K(\nabla_{\mathcal{D}} v_{\mathcal{D}})\Big)\,\mathrm{d}s(\boldsymbol{x}). \quad (2.4.17)\end{aligned}$$

Set $\varphi = A\xi n_{K,\sigma}$ and $w = \nabla_{\mathcal{D}} v_{\mathcal{D}}$. A change of variables yields

$$\int_{\sigma\in\mathcal{F}_K} \varphi\cdot\big(w - \Lambda_K(w)\big)\,\mathrm{d}s(\boldsymbol{x}) = |\sigma| \int_{\widehat{\sigma}\in\mathcal{F}_{\widehat{K}}} \widehat{\varphi}\cdot\big(\widehat{w} - \Lambda_{\widehat{K}}(\widehat{w})\big)\,\mathrm{d}s(\boldsymbol{x}), \quad (2.4.18)$$

where $\widehat{K}$ is the reference finite element. Let $\mathcal{F}_K = \{\sigma_1', \sigma_2', \sigma_1'', \sigma_2''\}$ such that $|\sigma_1'| = |\sigma_1''| = h_1$ and $|\sigma_2'| = |\sigma_2''| = h_2$. Consider

$$\delta_{1,K}(\phi, v) = \int_{\sigma_1'} \phi\big(v - \Lambda_K(v)\big)\,\mathrm{d}s(\boldsymbol{x}) - \int_{\sigma_1''} \phi\big(v - \Lambda_K(v)\big)\,\mathrm{d}s(\boldsymbol{x}), \quad (2.4.19)$$

for $\phi \in H^1(K)$ and $v \in \partial_1\mathbb{P}_K$:=$\{\partial_1 p : p \in \mathbb{P}_K\}$. The steps in [41, Theorem 6.2.3] show that $\delta_{1,K}(\phi, v) \le Ch|\phi|_{1,K}|v|_{1,K}$. For the sake of completeness, let us briefly recall the argument. A use of changes of variables leads to $\delta_{1,K}(\phi, v) = h_1\delta_{1,\widehat{K}}(\widehat{\phi}, \widehat{v})$. Since $\mathbb{P}_0 \subset Q_1$, which is preserved by Λ_K, for all $\widehat{v}\in\mathbb{P}_0$ and $\widehat{\phi}\in H^1(\widehat{K})$, $\delta_{1,\widehat{K}}(\widehat{\phi}, \widehat{v}) = 0$ (first polynomial invariance). Let us now

prove that the same relation holds if $\widehat{\phi} \in \mathbb{P}_0$ and $\widehat{v} \in \partial_1 \mathbb{P}_{\widehat{K}}$. Since $\widehat{\phi} \in \mathbb{P}_0$, its value on $\widehat{K}$ is a constant, say, equal to a_0. Since $\widehat{v} \in \partial_1 \mathbb{P}_{\widehat{K}}$, the definition of polynomial space for the Adini element implies that

$$\widehat{v} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_1 x_2 + b_5 x_2^2 + b_6 x_1^2 x_2 + b_7 x_2^3.$$

Take the values at the four vertices to obtain

$$\Lambda_{\widehat{K}} \widehat{v} = b_0 + (b_1 + b_3) x_1 + (b_2 + b_5 + b_7) x_2 + (b_4 + b_6) x_1 x_2.$$

Assume without loss of generality that $\sigma_1^{'}$ is the line $x_1 = 1$ and $\sigma_1^{''}$ is the line $x_1 = 0$. Then

$$(\widehat{v} - \Lambda_{\widehat{K}} \widehat{v})_{|x_1=0} = -(b_5 + b_7) x_2 + b_5 x_2^2 + b_7 x_2^3,$$

$$(\widehat{v} - \Lambda_{\widehat{K}} \widehat{v})_{|x_1=1} = -(b_5 + b_7) x_2 + b_5 x_2^2 + b_7 x_2^3.$$

The relation $\delta_{1,\widehat{K}}(\widehat{\phi}, \widehat{v}) = 0$ (second polynomial invariance) then follows from

$$\int_{\sigma_1^{'}} \widehat{\phi}(\widehat{v} - \Lambda_{\widehat{K}} \widehat{v}) \, \mathrm{d}s(\boldsymbol{x}) = \int_0^1 a_0(-(b_5 + b_7) x_2 + b_5 x_2^2 + b_7 x_2^3) \, \mathrm{d}x_2 = \int_{\sigma_1^{''}} \widehat{\phi}(\widehat{v} - \Lambda_{\widehat{K}} \widehat{v}) \, \mathrm{d}s(\boldsymbol{x}).$$

The bilinear form $\delta_{1,\widehat{K}}(\widehat{\phi}, \widehat{v})$ is continuous over the space $H^1(\widehat{K}) \times \partial_1 \mathbb{P}_{\widehat{K}}$ by the trace theorem. Use the bilinear lemma [41, Theorem 4.2.5] to deduce from the two polynomial invariances the existence of a constant C such that $|\delta_{1,\widehat{K}}(\widehat{\phi}, \widehat{v})| \leq C |\widehat{\phi}|_{1,\widehat{K}} |\widehat{v}|_{1,\widehat{K}}$ for all $\widehat{\phi} \in H^1(\widehat{K})$, $\widehat{v} \in \partial_1 \mathbb{P}_{\widehat{K}}$. A direct change of variables [41, Theorem 3.1.2] shows that

$$|\widehat{\phi}|_{1,\widehat{K}} \leq C |\phi|_{1,K} \quad \text{and} \quad |\widehat{v}|_{1,\widehat{K}} \leq C |v|_{1,K}.$$

Since $\delta_{1,K}(\phi, v) = h_1 \delta_{1,\widehat{K}}(\widehat{\phi}, \widehat{v})$, a use of the above estimates leads to $\delta_{1,K}(\phi, v) \leq Ch |\phi|_{1,K} |v|_{1,K}$. Similarly, $\delta_{2,K}(\phi, v) \leq Ch |\phi|_{1,K} |v|_{1,K}$ (considering integrals over $\sigma_2^{'}$ and $\sigma_2^{''}$). Hence, from (2.4.17), (2.4.18) and (2.4.19),

$$\Big| \int_\Omega (\mathcal{H} : A\xi) \Pi_{\mathcal{D}} v_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} - \int_\Omega A\xi : \mathcal{H}_{\mathcal{D}} v_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \Big| \leq C \|\xi\|_{H^2(\Omega)^{d \times d}} h \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|.$$

The proof of the estimate on $W_{\mathcal{D}}^B(\xi)$ is complete. □

THE MORLEY ELEMENT:

The following theorem verifies the properties of the B-Hessian discretisation (2.4.1)–(2.4.3) for the Morley element.

Theorem 2.4.8. *Let $\mathcal{D}$ be a B-Hessian discretisation for the Morley element in the sense of Definition 2.3.4 with B satisfying the coercive property. Then, there exists a constant C, not depending on $\mathcal{D}$, such that*

- $C_{\mathcal{D}}^B \leq C$,

- $\forall\, \varphi \in H^3(\Omega)\cap H_0^2(\Omega), \quad S_{\mathcal{D}}^B(\varphi) \le Ch\|\varphi\|_{H^3(\Omega)}$,
- $\forall\, \xi \in H^2(\Omega)^{d\times d},\ W_{\mathcal{D}}^B(\xi) \le Ch\|\xi\|_{H^2(\Omega)^{d\times d}}$.

Proof. • COERCIVITY: Let $v_{\mathcal{D}} \in X_{\mathcal{D},0}$. Since $[\![\Pi_{\mathcal{D}} v_{\mathcal{D}}]\!] = 0$ at the face vertices for any $v_{\mathcal{D}} \in X_{\mathcal{D},0}$ and $[\![\nabla_{\mathcal{D}} v_{\mathcal{D}}]\!] = 0$ at the edge midpoints, use Lemma A.1.3 twice and the coercivity property of B given by (2.2.3) to obtain

$$\|\Pi_{\mathcal{D}} v_{\mathcal{D}}\| \le C\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\| \le C\|\mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\| \le C\rho^{-1}\|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|.$$

This with (2.4.1) concludes the estimate on $C_{\mathcal{D}}^B$.
• CONSISTENCY: Consistency follows from the interpolation property of the family of Morley element [41, Chapter 6]. For $\varphi \in H^3(\Omega)\cap H_0^2(\Omega)$,

$$\inf_{w_{\mathcal{D}}\in X_{\mathcal{D},0}} \|\Pi_{\mathcal{D}} w_{\mathcal{D}} - \varphi\| \le Ch^3\|\varphi\|_{H^3(\Omega)}, \quad \inf_{w_{\mathcal{D}}\in X_{\mathcal{D},0}} \|\nabla_{\mathcal{D}} w_{\mathcal{D}} - \nabla\varphi\| \le Ch^2\|\varphi\|_{H^3(\Omega)},$$
$$\inf_{w_{\mathcal{D}}\in X_{\mathcal{D},0}} \|\mathcal{H}_{\mathcal{D}}^B w_{\mathcal{D}} - \mathcal{H}^B\varphi\| \le Ch\|\varphi\|_{H^3(\Omega)}.$$

This concludes that $S_{\mathcal{D}}^B(\varphi) \le Ch\|\phi\|_{3,\Omega}$.
• LIMIT–CONFORMITY: For any $\xi \in H^2(\Omega)^{d\times d}$ and $v_{\mathcal{D}} \in X_{\mathcal{D},0}$, cellwise integration by parts yields

$$\begin{aligned}\int_\Omega (\mathcal{H} : A\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} &= \sum_{K\in\mathcal{M}}\int_K (\mathcal{H} : A\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}\\ &= \int_\Omega A\xi : \mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} - \sum_{K\in\mathcal{M}}\int_{\partial K}(A\xi n_K)\cdot\nabla_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x})\\ &\quad + \sum_{K\in\mathcal{M}}\int_{\partial K}(\mathrm{div}(A\xi)\cdot n_K)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x}).\end{aligned}$$

This gives

$$\begin{aligned}\int_\Omega (\mathcal{H} : A\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} - \int_\Omega A\xi : \mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} &= -\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{F}_K}\int_\sigma (A\xi n_{K,\sigma})\cdot\nabla_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x})\\ &\quad + \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{F}_K}\int_\sigma (\mathrm{div}(A\xi)\cdot n_{K,\sigma})\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x}).\end{aligned} \tag{2.4.20}$$

The remaining arguments follow from the proof of [85, Lemma 3.5] with appropriate modifications. However, for the sake of completeness, we provide a proof. The first term on the right-hand side of (2.4.20) is estimated now. For any function v defined on an edge $\sigma \in \mathcal{F}$, define its mean value $\Pi_0 v$ by $\Pi_0 v = \frac{1}{|\sigma|}\int_\sigma v\,\mathrm{d}s(\boldsymbol{x})$. Introduce $\Pi_0(\nabla_{\mathcal{D}} v_{\mathcal{D}})$ and use the fact that these mean values are same on both sides of edges $\sigma \in \mathcal{F}_{\mathrm{int}}$ and they are zero along $\sigma \in \mathcal{F}_{\mathrm{ext}}$ [85, Lemma 3.1] to deduce

$$\sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{F}_K}\int_\sigma (A\xi n_{K,\sigma})\cdot\nabla_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x}) = \sum_{K\in\mathcal{M}}\sum_{\sigma\in\mathcal{F}_K}\int_\sigma (A\xi n_{K,\sigma})\cdot(\nabla_{\mathcal{D}} v_{\mathcal{D}} - \Pi_0(\nabla_{\mathcal{D}} v_{\mathcal{D}}))\,\mathrm{d}s(\boldsymbol{x}). \tag{2.4.21}$$

Consider for $\sigma \in \mathcal{F}_K$, $\int_\sigma (A\xi n_{K,\sigma}) \cdot (\nabla_{\mathcal{D}} v_{\mathcal{D}} - \Pi_0(\nabla_{\mathcal{D}} v_{\mathcal{D}}))\, \mathrm{d}s(\boldsymbol{x})$. Let $g = A\xi n_{K,\sigma}$ and $v = \nabla_{\mathcal{D}} v_{\mathcal{D}}$. A change of variables leads to

$$\int_{\sigma\in\mathcal{F}_K} g\cdot(v-\Pi_0 v)\,\mathrm{d}s(\boldsymbol{x}) = |\sigma| \int_{\widehat{\sigma}\in\mathcal{F}_{\widehat{K}}} \widehat{g}\cdot(\widehat{v}-\widehat{\Pi_0}\widehat{v})\,\mathrm{d}s(\boldsymbol{x}) := |\sigma| F(\widehat{g},\widehat{v}), \tag{2.4.22}$$

where $\widehat{K}$ is the reference finite element. For all $\widehat{g} \in \mathbb{P}_0$ and $\widehat{v} \in H^1(\widehat{K})$, $F(\widehat{g},\widehat{v}) = 0$. Also, for all $\widehat{g} \in H^1(\widehat{K})$ and $\widehat{v} \in \mathbb{P}_0$, $F(\widehat{g},\widehat{v}) = 0$. Hence use the polynomial invariance result ([85, Lemma 2.1] with $l = k = 0$) to obtain $|F(\widehat{g},\widehat{v})| \le C|\widehat{g}|_{1,\widehat{K}}|\widehat{v}|_{1,\widehat{K}}$, where C depends only on Ω. A substitution of this estimate in (2.4.22) along with $|\widehat{g}|_{1,\widehat{K}} \le C|g|_{1,K}$ and $|\widehat{v}|_{1,\widehat{K}} \le C|v|_{1,K}$ [41, Theorem 3.1.2], summing over all the edges and a use of (2.4.21) and (2.2.3) yields

$$\left| \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{F}_K} \int_\sigma (A\xi n_{K,\sigma})\cdot\nabla_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x}) \right| \le C\rho^{-1} h \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\| \,\|\xi\|_{H^1(\Omega)^{d\times d}}. \tag{2.4.23}$$

Consider the second term on the right-hand side of (2.4.20). Let V_1 be the space of all globally continuous piecewise linear functions and let $\Pi_1 : V_h \to V_1$ be the interpolation operator such that $\Pi_1 v_h$ equal to v_h at the vertices of all triangle K, $v_h \in V_h$, where V_h is the Morley finite element space. Then

$$\sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{F}_K} \int_\sigma (\mathrm{div}(A\xi)\cdot n_{K,\sigma}) \Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x}) = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{F}_K} \int_\sigma P(v-\Pi_1 v)\,\mathrm{d}s(\boldsymbol{x}), \tag{2.4.24}$$

where $P = \mathrm{div}(A\xi)\cdot n_{K,\sigma}$ and $v = \Pi_{\mathcal{D}} v_{\mathcal{D}}$. Note that

$$\int_{\sigma\in\mathcal{F}_K} P(v-\Pi_1 v)\,\mathrm{d}s(\boldsymbol{x}) = |\sigma| \int_{\widehat{\sigma}\in\mathcal{F}_{\widehat{K}}} \widehat{P}(\widehat{v}-\widehat{\Pi_1}\widehat{v})\,\mathrm{d}s(\boldsymbol{x}). \tag{2.4.25}$$

A use of the continuous trace inequality [45], the discrete trace inequality [45, Lemma 1.46], an interpolation estimate [41] and Young's inequality leads to

$$\left| \int_{\widehat{\sigma}\in\mathcal{F}_{\widehat{K}}} \widehat{P}(\widehat{v}-\Pi_1\widehat{v})\,\mathrm{d}s(\boldsymbol{x}) \right| \le C\big(|\widehat{P}|_{0,\widehat{K}} + |\widehat{P}|_{1,\widehat{K}}\big)|\widehat{v}|_{2,\widehat{K}}.$$

Substitute the above displayed estimate in (2.4.25), use [41, Theorem 3.1.2] to go back to K, sum over all the edges, (2.4.24) and (2.2.3) to deduce

$$\left| \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{F}_K} \int_\sigma (\mathrm{div}(A\xi)\cdot n_{K,\sigma}) \Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}s(\boldsymbol{x}) \right| \le C\rho^{-1}(h\|\xi\|_{H^1(\Omega)^{d\times d}} + h^2\|\xi\|_{H^2(\Omega)^{d\times d}})\|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\|. \tag{2.4.26}$$

A substitution of (2.4.23) and (2.4.26) in (2.4.20) yields

$$\left| \int_\Omega (\mathcal{H} : A\xi)\Pi_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} - \int_\Omega A\xi : \mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} \right| \le C\rho^{-1} h \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\| \,\|\xi\|_{H^2(\Omega)^{d\times d}}$$

and this leads to the desired estimate on $W^B_{\mathcal{D}}$. □

Corollary 2.4.9. *Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of B–Hessian discretisations built on the Morley triangle, such that B is coercive and the underlying sequence of meshes are regular and have a size that goes to 0 as $m \to \infty$. Then the sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is coercive, consistent and limit-conforming.*

2.4.2 Gradient Recovery Method

The theorem below gives an estimate on the accuracy measures $C_{\mathcal{D}}^B$, $S_{\mathcal{D}}^B$ and $W_{\mathcal{D}}^B$ associated with an HD $\mathcal{D}$ using gradient recovery method. Incidentally, the estimate on $C_{\mathcal{D}}^B$ also establishes that $\|\mathcal{H}_{\mathcal{D}}^B \cdot\|$ is a norm on $X_{\mathcal{D},0}$.

Theorem 2.4.10 (Estimates for Hessian discretisations based on gradient recovery)**.**
Let $\mathcal{D}$ be a B–Hessian discretisation in the sense of Definition 2.3.5, with B satisfying Estimate (2.3.2) *and $(V_h, I_h, Q_h, \mathfrak{S}_h)$ satisfying* **(P0)–(P5)**. *Then, there exists a constant C, not depending on h, such that*

- $C_{\mathcal{D}}^B \le C$,
- $\forall\, \varphi \in W$, $S_{\mathcal{D}}^B(\varphi) \le Ch\|\varphi\|_W$,
- $\forall\, \xi \in H^2(\Omega)^{d\times d}$, $W_{\mathcal{D}}^B(\xi) \le Ch\|\xi\|_{H^2(\Omega)^{d\times d}}$.

Before proving this theorem, let us note the following straightforward consequence of Remark 2.4.2.

Corollary 2.4.11 (Properties of Hessian discretisation based on gradient recovery)**.**
Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of B–Hessian discretisations, with B satisfying Estimate (2.3.2) *and each $\mathcal{D}_m$ associated with $(V_{h_m}, Q_{h_m}, I_{h_m}, \mathfrak{S}_{h_m})$ satisfying* **(P0)–(P5)** *uniformly with respect to m. Assume that $h_m \to 0$ as $m \to \infty$. Then the sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is coercive, consistent and limit-conforming.*

Proof of Theorem 2.4.10.

• COERCIVITY: Let $v \in X_{\mathcal{D},0}$. Note that $|a\otimes b| = |a||b|$ for any two vectors a and b. The definition of $\mathcal{H}_{\mathcal{D}}^B$, Property (2.3.2) of B and $|\mathfrak{S}| \ge 1$ imply that

$$\begin{aligned}
\|\mathcal{H}_{\mathcal{D}}^B v\|^2 &\ge C_B^2 \int_\Omega |\nabla(Q_h\nabla v) + \mathfrak{S}_h \otimes (Q_h\nabla v - \nabla v)|^2 \,\mathrm{d}\boldsymbol{x} \\
&= C_B^2 \int_\Omega |\nabla(Q_h\nabla v)|^2 \,\mathrm{d}\boldsymbol{x} + C_B^2 \int_\Omega |\mathfrak{S}_h \otimes (Q_h\nabla v - \nabla v)|^2 \,\mathrm{d}\boldsymbol{x} \\
&\quad + 2C_B^2 \int_\Omega \nabla(Q_h\nabla v) : \mathfrak{S}_h \otimes (Q_h\nabla v - \nabla v) \,\mathrm{d}\boldsymbol{x} \\
&\ge C_B^2 \left(\|\nabla(Q_h\nabla v)\|^2 + \|Q_h\nabla v - \nabla v\|^2\right) \\
&\quad + 2C_B^2 \sum_{K\in\mathcal{M}} \int_K \nabla(Q_h\nabla v) : \mathfrak{S}_h \otimes (Q_h\nabla v - \nabla v) \,\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Since $\nabla(Q_h\nabla v)_{|K} \in \nabla V_h(K)^d$, a use of property **(P5)** shows that the last term vanishes, and thus

$$\|\mathcal{H}_{\mathcal{D}}^B v\|^2 \ge C_B^2 \left(\|\nabla(Q_h\nabla v)\|^2 + \|Q_h\nabla v - \nabla v\|^2\right), \tag{2.4.27}$$

which implies

$$C_B^{-1}\sqrt{2}\|\mathcal{H}_{\mathcal{D}}^B v\| \ge \|\nabla(Q_h\nabla v)\| + \|Q_h\nabla v - \nabla v\|. \tag{2.4.28}$$

Apply now the Poincaré inequality twice, the triangle inequality and (2.4.28) to obtain

$$\begin{aligned}\|\Pi_{\mathcal{D}} v\| = \|v\| &\leq \operatorname{diam}(\Omega)\|\nabla v\| \\ &\leq \operatorname{diam}(\Omega)\|\nabla v - Q_h \nabla v\| + \operatorname{diam}(\Omega)\|Q_h \nabla v\| \\ &\leq \operatorname{diam}(\Omega)\|\nabla v - Q_h \nabla v\| + \operatorname{diam}(\Omega)^2 \|\nabla(Q_h \nabla v)\| \\ &\leq C_B^{-1}\sqrt{2}\max(\operatorname{diam}(\Omega), \operatorname{diam}(\Omega)^2)\|\mathcal{H}_{\mathcal{D}}^B v\|. \end{aligned} \tag{2.4.29}$$

From (2.4.27) and the Poincaré inequality,

$$\|\nabla_{\mathcal{D}} v\| = \|Q_h \nabla v\| \leq \operatorname{diam}(\Omega)\|\nabla(Q_h \nabla v)\| \leq \operatorname{diam}(\Omega) C_B^{-1} \|\mathcal{H}_{\mathcal{D}}^B v\|. \tag{2.4.30}$$

Estimates (2.4.29) and (2.4.30) show that $C_{\mathcal{D}}^B \leq C_B^{-1}\sqrt{2}\max(\operatorname{diam}(\Omega), \operatorname{diam}(\Omega)^2)$.

• CONSISTENCY: let $\varphi \in W \subset H^3(\Omega) \cap H_0^2(\Omega)$ and choose $v = I_h \varphi \in X_{\mathcal{D},0}$. Use the properties **(P0)** (which implies $\|I_h \varphi - \varphi\| \leq Ch\|\varphi\|_{H^2(\Omega)}$ by the Poincaré inequality) and **(P2)** to obtain

$$\|\Pi_{\mathcal{D}} v - \varphi\| = \|I_h \varphi - \varphi\| \leq Ch\|\varphi\|_{H^2(\Omega)} \tag{2.4.31}$$

and

$$\|\nabla_{\mathcal{D}} v - \nabla\varphi\| = \|Q_h \nabla I_h \varphi - \nabla\varphi\| \leq Ch^2 \|\varphi\|_W. \tag{2.4.32}$$

Let us now turn to $\|\mathcal{H}_{\mathcal{D}}^B v - \mathcal{H}^B \varphi\|$. Observe that $\nabla\nabla$ is another notation for $\mathcal{H}$. A use of a triangle inequality, the boundedness of B and $\mathfrak{S}_h$ leads to

$$\begin{aligned}\|\mathcal{H}_{\mathcal{D}}^B v - \mathcal{H}^B \varphi\| &= \|B\left[\nabla(Q_h \nabla v) + \mathfrak{S}_h \otimes (Q_h \nabla v - \nabla v)\right] - B\mathcal{H}\varphi\| \\ &\leq \|B\left[\nabla(Q_h \nabla v) - \nabla\nabla\varphi\right]\| + \|B\mathfrak{S}_h \otimes (Q_h \nabla v - \nabla v)\| \\ &\leq C\underbrace{\|\nabla(Q_h \nabla v) - \nabla\nabla\varphi\|}_{A_1} + C\underbrace{\|Q_h \nabla v - \nabla v\|}_{A_2}. \end{aligned} \tag{2.4.33}$$

An introduction of the term $\nabla(Q_h \nabla\varphi)$ and a use of the triangle inequality in sequence, the inverse inequality in **(P0)**, **(P3)**, the projection property of Q_h, **(P1)** and **(P2)** yield

$$\begin{aligned} A_1 &\leq \|\nabla[Q_h \nabla v - Q_h \nabla\varphi]\| + \|\nabla(Q_h \nabla\varphi) - \nabla\nabla\varphi\| \\ &\leq Ch^{-1}\|Q_h \nabla v - Q_h \nabla\varphi\| + Ch\|\nabla\varphi\|_{H^2(\Omega)} \\ &\leq Ch^{-1}\|Q_h\left(Q_h \nabla v - \nabla\varphi\right)\| + Ch\|\nabla\varphi\|_{H^2(\Omega)} \\ &\leq Ch^{-1}\|Q_h \nabla I_h \varphi - \nabla\varphi\| + Ch\|\nabla\varphi\|_{H^2(\Omega)} \leq Ch\|\varphi\|_W. \end{aligned} \tag{2.4.34}$$

To estimate A_2, use the properties **(P2)** and **(P0)**:

$$A_2 \leq \|Q_h \nabla v - \nabla\varphi\| + \|\nabla\varphi - \nabla v\| \leq Ch^2\|\varphi\|_W + Ch\|\varphi\|_{H^2(\Omega)}. \tag{2.4.35}$$

The estimate on $S_{\mathcal{D}}^B(\varphi)$ follows from (2.4.31)–(2.4.35).

• LIMIT-CONFORMITY: for $\xi \in H^2(\Omega)^{d\times d}$ and $v \in X_{\mathcal{D},0}$,

$$\int_\Omega \Big((\mathcal{H} : B^\tau B\xi)\Pi_{\mathcal{D}} v - B\xi : \mathcal{H}^B_{\mathcal{D}} v\Big)\,\mathrm{d}\boldsymbol{x} = \underbrace{\int_\Omega \Big((\mathcal{H} : B^\tau B\xi)\Pi_{\mathcal{D}} v - B\xi : B\nabla(Q_h\nabla v)\Big)\,\mathrm{d}\boldsymbol{x}}_{B_1} - \underbrace{\int_\Omega B\xi : B\mathfrak{S}_h \otimes (Q_h\nabla v - \nabla v)\,\mathrm{d}\boldsymbol{x}}_{B_2}. \tag{2.4.36}$$

Recall that $v = \Pi_{\mathcal{D}} v$ and $A = B^\tau B$. Since $Q_h\nabla v \in H^1_0(\Omega)$, Lemma A.1.2 applied to $(\mathcal{H} : A\xi)v$ and an integration by parts on $B\xi : B\nabla(Q_h\nabla v) = A\xi : \nabla(Q_h\nabla v)$ show that, for any $\mu_h \in N_h = [(Q_h\nabla - \nabla)(V_h)]^\perp$,

$$\begin{aligned} |B_1| &= \Big|\int_\Omega (\mathcal{H} : A\xi)v\,\mathrm{d}\boldsymbol{x} + \int_\Omega Q_h\nabla v \cdot \mathrm{div}(A\xi)\,\mathrm{d}\boldsymbol{x}\Big| = \Big|\int_\Omega (Q_h\nabla v - \nabla v)\cdot \mathrm{div}(A\xi)\,\mathrm{d}\boldsymbol{x}\Big| \\ &= \Big|\int_\Omega (Q_h\nabla v - \nabla v)\cdot(\mathrm{div}(A\xi) - \mu_h)\,\mathrm{d}\boldsymbol{x}\Big| \le \|Q_h\nabla v - \nabla v\|\,\|\mathrm{div}(A\xi) - \mu_h\|. \end{aligned} \tag{2.4.37}$$

Take the infimum over all $\mu_h \in N_h$. Estimate (2.4.28) and Property **(P4)** yield

$$|B_1| \le Ch\|\mathcal{H}^B_{\mathcal{D}} v\|\,\|\mathrm{div}(A\xi)\|_{H^1(\Omega)^d}. \tag{2.4.38}$$

Let ξ_K denote the average of ξ over $K \in \mathcal{M}$. By the mesh regularity assumption, $\|\xi - \xi_K\|_{L^2(K)^{d\times d}} \le Ch\|\xi\|_{H^1(K)^{d\times d}}$ (see, e.g., [48, Lemma B.6]). Moreover, since V_h contains the piecewise constant functions, $\nabla V_h(K)$ contains the constant vector-valued functions on K and thus, by the orthogonality condition in **(P5)**, the Cauchy–Schwarz inequality, the boundedness of B and $\mathfrak{S}_h$, and (2.4.28),

$$\begin{aligned} |B_2| &= \Big|\sum_{K\in\mathcal{M}}\int_K B^\tau B\xi : \mathfrak{S}_h \otimes (Q_h\nabla v - \nabla v)\,\mathrm{d}\boldsymbol{x}\Big| \\ &= \Big|\sum_{K\in\mathcal{M}}\int_K (B^\tau B\xi - B^\tau B\xi_K) : \mathfrak{S}_h \otimes (Q_h\nabla v - \nabla v)\,\mathrm{d}\boldsymbol{x}\Big| \\ &\le C\sum_{K\in\mathcal{M}} \|\xi - \xi_K\|_{L^2(K)}\|Q_h\nabla v - \nabla v\|_{L^2(K)} \le Ch\|\xi\|_{H^1(\Omega)^{d\times d}}\|\mathcal{H}^B_{\mathcal{D}} v\|. \end{aligned} \tag{2.4.39}$$

Plug (2.4.38) and (2.4.39) into (2.4.36) to arrive at

$$\Big|\int_\Omega \Big((\mathcal{H} : B^\tau B\xi)\Pi_{\mathcal{D}} v - B\xi : \mathcal{H}^B_{\mathcal{D}} v\Big)\,\mathrm{d}\boldsymbol{x}\Big| \le Ch\Big(\|\mathrm{div}(A\xi)\|_{H^1(\Omega)^d} + \|\xi\|_{H^1(\Omega)^{d\times d}}\Big)\|\mathcal{H}^B_{\mathcal{D}} v\|.$$

This and the definition (2.4.3) of $W^B_{\mathcal{D}}(\xi)$ concludes the proof of the estimate on this quantity. □

2.4.3 Finite Volume Methods

This section deals with the properties of HDM associated with FVM and shows that the generic error estimate established in the HDM slightly improves the estimates found in [59], see Remark 2.4.13 below.

Theorem 2.4.12. *Let $\mathcal{D}$ be a B–Hessian discretisation in the sense of Definition 2.3.9. Then there exists a constant C, depending only on on $\theta \geq \theta_{\mathcal{T}}$, such that*

- $C_{\mathcal{D}}^B \leq C$,
- *If $\varphi \in C_c^2(\Omega)$, $\Delta\varphi \in H^1(\Omega)$ and $a > 0$ is such that* $\mathrm{supp}(\varphi) \subset \{x \in \Omega\,;\, \mathrm{dist}(\boldsymbol{x}, \partial\Omega) > a\}$, *then*

$$S_{\mathcal{D}}^B(\varphi) \leq Ch\|\Delta\varphi\|_{H^1(\Omega)} + Ch\|\varphi\|_{C^2(\overline{\Omega})} \times \begin{cases} |\ln(a)|a^{-3/2} & \text{if } d = 2, \\ a^{-5/3} & \text{if } d = 3. \end{cases} \tag{2.4.40}$$

- *If $\varphi \in H_0^2(\Omega) \cap C^2(\overline{\Omega})$ with $\Delta\varphi \in H^1(\Omega)$, then*

$$S_{\mathcal{D}}^B(\varphi) \leq Ch\|\Delta\varphi\|_{H^1(\Omega)} + C\|\varphi\|_{C^2(\overline{\Omega})} \times \begin{cases} h^{1/4}|\ln(h)| & \text{if } d = 2, \\ h^{3/13} & \text{if } d = 3. \end{cases} \tag{2.4.41}$$

- $\forall \xi \in H^2(\Omega)^{d\times d}$, $W_{\mathcal{D}}^B(\xi) \leq Ch\|\mathrm{tr}(\xi)\|_{H^2(\Omega)}$.

Remark 2.4.13. *If the solution $\overline{u}$ to* (2.2.4) *belongs to $H^4(\Omega) \cap H_0^2(\Omega)$, then $\overline{u} \in C^2(\overline{\Omega})$ and $\Delta\overline{u} \in H^2(\Omega)$. In that case, Theorems 2.4.4 and 2.4.12 provide an $\mathcal{O}(h^{1/4}|\ln(h)|)$ (in dimension $d = 2$) or $\mathcal{O}(h^{3/13})$ (in dimension $d = 3$) error estimate for the Hessian scheme based on the HD from Definition 2.3.9. This slightly improves the result of [59, Theorem 4.3], in which an $\mathcal{O}(h^{1/5})$ estimate is obtained if $\overline{u} \in C^4(\overline{\Omega}) \cap H_0^2(\Omega)$.*

As for the method based on gradient recovery operators and finite element schemes, the properties of the Hessian discretisation follow from the estimates in Theorem 2.4.12 and from Remark 2.4.2.

Corollary 2.4.14. *Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of B–Hessian discretisations in the sense of Definition 2.3.9, associated to meshes such that $h_m \to 0$ and $(\theta_{\mathcal{T}_m})_{m\in\mathbb{N}}$ is bounded. Then the sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is coercive, consistent and limit-conforming.*
Proof of Theorem 2.4.12.

• COERCIVITY: the discrete Poincaré inequality of [58] states that

$$\|\Pi_{\mathcal{D}} v_{\mathcal{D}}\| \leq \mathrm{diam}(\Omega)\|v_{\mathcal{D}}\|_{1,\mathcal{D}}, \quad \forall v_{\mathcal{D}} \in X_{\mathcal{D},0}, \tag{2.4.42}$$

where $\|\cdot\|_{1,\mathcal{D}}$ is the discrete H^1 norm on $X_{\mathcal{D},0}$ (the norm associated with the inner product given by (2.3.11)). Let us first prove that

$$-\int_\Omega \Pi_{\mathcal{D}} u_{\mathcal{D}} \Delta_{\mathcal{D}} v_{\mathcal{D}} dx = [u_{\mathcal{D}}, v_{\mathcal{D}}]_{\mathcal{D}}, \quad u_{\mathcal{D}}, v_{\mathcal{D}} \in X_{\mathcal{D},0}. \tag{2.4.43}$$

The definitions of $\Pi_{\mathcal{D}}$ and $\Delta_{\mathcal{D}}$ yield

$$-\int_\Omega \Pi_{\mathcal{D}} u_{\mathcal{D}} \Delta_{\mathcal{D}} v_{\mathcal{D}} dx = \sum_{K\in\mathcal{M}} -|K| u_K \Delta_K v_{\mathcal{D}} = -\sum_{K\in\mathcal{M}} u_K \sum_{\sigma\in\mathcal{F}_K} \frac{|\sigma|\delta_{K,\sigma} v_{\mathcal{D}}}{d_\sigma}.$$

For $\sigma \in \mathcal{F}_{\rm ext}$, $\delta_{K,\sigma} v_{\mathcal{D}} = 0$. Gather the sums by edges and use (2.3.9) and (2.3.12) to obtain

$$-\int_\Omega \Pi_{\mathcal{D}} u_{\mathcal{D}} \Delta_{\mathcal{D}} v_{\mathcal{D}} dx = \sum_{K\in\mathcal{M}} u_K \sum_{\sigma\in\mathcal{F}_{K,\rm int}} \frac{|\sigma|(v_K - v_L)}{d_\sigma} = \sum_{\sigma\in\mathcal{F}_{\rm int}} \frac{|\sigma|\delta_\sigma u_{\mathcal{D}} \delta_\sigma v_{\mathcal{D}}}{d_\sigma},$$

which establishes (2.4.43). Choose $v_{\mathcal{D}} = u_{\mathcal{D}}$, apply the Cauchy–Schwarz inequality and use (2.4.42) to deduce

$$\|u_{\mathcal{D}}\|_{1,\mathcal{D}}^2 \le \|\Pi_{\mathcal{D}} u_{\mathcal{D}}\| \|\Delta_{\mathcal{D}} u_{\mathcal{D}}\| \le \operatorname{diam}(\Omega) \|u_{\mathcal{D}}\|_{1,\mathcal{D}} \|\Delta_{\mathcal{D}} u_{\mathcal{D}}\|.$$

Thus,

$$\|u_{\mathcal{D}}\|_{1,\mathcal{D}} \le \operatorname{diam}(\Omega) \|\Delta_{\mathcal{D}} u_{\mathcal{D}}\|. \tag{2.4.44}$$

A combination of (2.4.42) and (2.4.44) leads to

$$\|\Pi_{\mathcal{D}} v_{\mathcal{D}}\| \le \operatorname{diam}(\Omega)^2 \|\Delta_{\mathcal{D}} u_{\mathcal{D}}\|. \tag{2.4.45}$$

The stability of the discrete gradient [59, Lemma 4.1] yields

$$\|\nabla_{\mathcal{D}} u_{\mathcal{D}}\| \le \theta\sqrt{d} \|u_{\mathcal{D}}\|_{\mathcal{D}} \quad \forall u_{\mathcal{D}} \in X_{\mathcal{D},0}.$$

Estimate (2.4.44) then shows that $\|\nabla_{\mathcal{D}} u_{\mathcal{D}}\| \le \operatorname{diam}(\Omega)\theta\sqrt{d}\|\Delta_{\mathcal{D}} u_{\mathcal{D}}\|$, which, together with (2.4.45), concludes the proof of the estimate on $C_{\mathcal{D}}^B$.

• CONSISTENCY – COMPACT SUPPORT: The proof utilises the ideas of [59], with a few improvements of the estimates. For $s > 0$, let $\Omega_s = \{x \in \Omega ; \operatorname{dist}(\boldsymbol{x}, \partial\Omega) > s\}$. In this proof, $A \lesssim B$ means that $A \le CB$ for some constant C depending only on θ.

First consider the case where $\varphi \in C_c^2(\Omega)$ and $\Delta\varphi \in H^1(\Omega)$, with support at distance from $\partial\Omega$ equal to or greater than a. As in [59, Proof of Lemma 4.4], let $\psi^a \in C_c^\infty(\Omega)$, equal to 1 on $\Omega_{3a/4}$, that vanishes on $\Omega\backslash\Omega_{a/4}$, and such that, for all $\alpha \in \mathbb{N}^d$, with $|\alpha| = \sum_{i=1}^d \alpha_i$,

$$\|\partial^\alpha \psi^a\|_{L^\infty(\Omega)} \lesssim a^{-|\alpha|}. \tag{2.4.46}$$

Letting $\psi_{\mathcal{D}}^a = (\psi^a(\boldsymbol{x}_K))_{K\in\mathcal{M}}$, $|\Delta_{\mathcal{D}} \psi_{\mathcal{D}}^a| \lesssim a^{-2}$. Hence, for all $r \in [1,\infty]$, since $\Omega\backslash\Omega_{2a}$ has measure $\lesssim a$,

$$\|\Delta_{\mathcal{D}} \psi_{\mathcal{D}}^a\|_{L^r(\Omega)} \lesssim a^{-2+\frac{1}{r}}. \tag{2.4.47}$$

Let $\widetilde{v} = (\widetilde{v}_K)_{K\in\mathcal{M}}$ be the solution of the two-point flux approximation finite volume scheme with homogeneous Dirichlet boundary conditions and source term $-\Delta\varphi$. Then by [58], with $\varphi_{\mathcal{D}} = (\varphi(\boldsymbol{x}_K))_{K\in\mathcal{M}}$,

$$\left(\sum_{\sigma\in\mathcal{F}} \frac{|\sigma|}{d_\sigma} (\delta_\sigma(\widetilde{v} - \varphi_{\mathcal{D}}))^2\right)^{1/2} \lesssim h\|\varphi\|_{C^2(\overline{\Omega})} \tag{2.4.48}$$

and, for $q \in [1,+\infty)$ if $d = 2$, $q \in [1,6]$ if $d = 3$,

$$\left(\sum_{K\in\mathcal{M}} |K| \, |\widetilde{v}_K - \varphi(\boldsymbol{x}_K)|^q\right)^{1/q} \lesssim qh\|\varphi\|_{C^2(\overline{\Omega})}. \tag{2.4.49}$$

Set $w = (\psi^a(\boldsymbol{x}_K)\widetilde{v}_K)_{K\in\mathcal{M}}$, that belongs to $X_{\mathcal{D},0}$ if $h \le a/4$. It is proved in [59, Proof of Lemma 4.4, p. 2032] that, with $[\Delta\varphi]_K = \frac{1}{|K|}\int_K \Delta\varphi\,dx$,

$$\begin{aligned}\Delta_K w - [\Delta\varphi]_K &= (\widetilde{v}_K - \varphi(\boldsymbol{x}_K))\Delta_K\psi^a_{\mathcal{D}} + \frac{1}{|K|}\sum_{\sigma\in\mathcal{F}_K}\frac{|\sigma|}{d_\sigma}(\delta_{K,\sigma}\psi^a_{\mathcal{D}})\delta_{K,\sigma}(\widetilde{v}-\varphi_{\mathcal{D}}),\\ &= T_{1,K} + T_{2,K}.\end{aligned} \tag{2.4.50}$$

Use Hölder's inequality with exponents $(q, \frac{2q}{q-2})$, for some $q > 2$ admissible in (2.4.49), and recall (2.4.47) to obtain

$$\left(\sum_{K\in\mathcal{M}}|K|\,|T_{1,K}|^2\right)^{1/2} \lesssim qha^{-2+\frac{q-2}{2q}}\|\varphi\|_{C^2(\overline{\Omega})}. \tag{2.4.51}$$

On the other hand, we have $|\delta_{K,\sigma}\psi^a_{\mathcal{D}}| \lesssim d_\sigma a^{-1}$ (see [59, Proof of Lemma 4.4]). Hence, a use of the Cauchy–Schwarz inequality on the sum over the faces, and the estimate $\sum_{\sigma\in\mathcal{F}_K}|\sigma|d_\sigma \lesssim |K|$ yields

$$|T_{2,K}|^2 \lesssim \frac{a^{-2}}{|K|^2}\left(\sum_{\sigma\in\mathcal{F}_K}|\sigma|\,|\delta_{K,\sigma}(\widetilde{v}-\varphi_{\mathcal{D}})|\right)^2 \lesssim \frac{a^{-2}}{|K|}\sum_{\sigma\in\mathcal{F}_K}\frac{|\sigma|}{d_\sigma}(\delta_{K,\sigma}(\widetilde{v}-\varphi_{\mathcal{D}}))^2.$$

Estimate (2.4.48) thus leads to

$$\left(\sum_{K\in\mathcal{M}}|K|\,|T_{2,K}|^2\right)^{1/2} \lesssim a^{-1}h\|\varphi\|_{C^2(\overline{\Omega})}. \tag{2.4.52}$$

Denote by $[\Delta\varphi]_{\mathcal{D}}$ the piecewise constant function equal to $[\Delta\varphi]_K$ on $K\in\mathcal{M}$. Take the L^2 norm of (2.4.50) and use (2.4.51) and (2.4.52) to arrive at, since $a^{-1} \lesssim a^{-\frac{3}{2}-\frac{1}{q}}$,

$$\|\Delta_{\mathcal{D}}w - [\Delta\varphi]_{\mathcal{D}}\|_{L^2(\Omega)} \lesssim qha^{-\frac{3}{2}-\frac{1}{q}}\|\varphi\|_{C^2(\overline{\Omega})}.$$

Taking $q = |\ln(a)|$ if $d = 2$ or $q = 6$ if $d = 3$ shows that

$$\|\Delta_{\mathcal{D}}w - [\Delta\varphi]_{\mathcal{D}}\|_{L^2(\Omega)} \lesssim h\|\varphi\|_{C^2(\overline{\Omega})} \times \begin{cases} |\ln(a)|a^{-3/2} & \text{if } d=2,\\ a^{-5/3} & \text{if } d=3.\end{cases} \tag{2.4.53}$$

A classical estimate [48, Lemma B.6] gives

$$\|[\Delta\varphi]_{\mathcal{D}} - \Delta\varphi\|_{L^2(\Omega)} \lesssim h\|\Delta\varphi\|_{H^1(\Omega)}, \tag{2.4.54}$$

which shows that $\|\Delta_{\mathcal{D}}w - \Delta\varphi\|_{L^2(\Omega)}$ is bounded above by the right-hand side of (2.4.40). The estimates on $\nabla_{\mathcal{D}}w - \nabla\varphi$ and on $\Pi_{\mathcal{D}}w - \varphi$ follow as in [59, Lemma 4.4].

• CONSISTENCY – GENERAL CASE: Consider now the case $\varphi \in H^2_0(\Omega)\cap C^2(\overline{\Omega})$, and take ψ^a as above. The boundary conditions on φ show that $|\varphi(\boldsymbol{x})| \lesssim \|\varphi\|_{C^2(\overline{\Omega})}\mathrm{dist}(\boldsymbol{x},\partial\Omega)^2$ and $|\nabla\varphi(\boldsymbol{x})| \lesssim$

$\|\varphi\|_{C^2(\overline{\Omega})}\mathrm{dist}(\boldsymbol{x},\partial\Omega)$. Hence, a use of (2.4.46), $|\Omega\backslash\Omega_a|\lesssim a$ and the fact that $1-\psi^a=0$ in Ω_a leads to, for all $\alpha\in\mathbb{N}^d$ with $|\alpha|\le 2$,

$$\|\partial^\alpha\varphi-\partial^\alpha(\psi^a\varphi)\|_{L^2(\Omega)}\lesssim a^{1/2}\|\varphi\|_{C^2(\overline{\Omega})}. \tag{2.4.55}$$

Since $\Delta=\sum_{i=1}^2\partial_i^2$, the above estimate applies to Δ instead of ∂^α and, as a consequence,

$$\|[\Delta\varphi]_{\mathcal{D}}-[\Delta(\psi^a\varphi)]_{\mathcal{D}}\|_{L^2(\Omega)}\le\|\Delta\varphi-\Delta(\psi^a\varphi)\|_{L^2(\Omega)}\lesssim a^{1/2}\|\varphi\|_{C^2(\overline{\Omega})}. \tag{2.4.56}$$

Consider now the interpolant $w\in X_{\mathcal{D},0}$ for $\psi^a\varphi\in C_c^2(\Omega)$ constructed above. Apply (2.4.53) to $\psi^a\varphi$ instead of φ, note that $\|\psi^a\varphi\|_{C^2(\overline{\Omega})}\lesssim\|\varphi\|_{C^2(\overline{\Omega})}$ (consequence of (2.4.55)), and use (2.4.56) to obtain

$$\|\Delta_{\mathcal{D}}w-[\Delta\varphi]_{\mathcal{D}}\|_{L^2(\Omega)}\lesssim a^{1/2}\|\varphi\|_{C^2(\overline{\Omega})}+h\|\varphi\|_{C^2(\overline{\Omega})}\times\begin{cases}|\ln(a)|a^{-3/2} & \text{if } d=2,\\ a^{-5/3} & \text{if } d=3.\end{cases}$$

Taking $a=h^{1/2}$ if $d=2$ or $a=h^{6/13}$ if $d=3$ leads to

$$\|\Delta_{\mathcal{D}}w-[\Delta\varphi]_{\mathcal{D}}\|_{L^2(\Omega)}\lesssim\|\varphi\|_{C^2(\overline{\Omega})}\times\begin{cases}h^{1/4}|\ln(h)| & \text{if } d=2,\\ h^{3/13} & \text{if } d=3.\end{cases}$$

Combined with (2.4.54) this shows that $\|\Delta_{\mathcal{D}}w-\Delta\varphi\|_{L^2(\Omega)}$ is bounded above by the right-hand side of (2.4.41). The estimates on $\Pi_{\mathcal{D}}w-\varphi$ and $\nabla_{\mathcal{D}}w-\nabla\varphi$ follow in a similar way.

• LIMIT-CONFORMITY: For $\xi\in H^B(\Omega)$ and $v_{\mathcal{D}}\in X_{\mathcal{D},0}$, $B=\frac{\mathrm{tr}(\cdot)}{\sqrt{d}}\mathrm{Id}$ implies

$$\int_\Omega(\mathcal{H}:B^\tau B\xi)\Pi_{\mathcal{D}}v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}=\int_\Omega(B\mathcal{H}:B\xi)\Pi_{\mathcal{D}}v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}=\int_\Omega\Delta\phi\Pi_{\mathcal{D}}v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x},$$

where $\phi=\mathrm{tr}(\xi)$. Also, by the definition of $\mathcal{H}_{\mathcal{D}}^B$,

$$\int_\Omega B\xi:\mathcal{H}_{\mathcal{D}}^Bv_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}=\int_\Omega\phi\Delta_{\mathcal{D}}v_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}.$$

Thus, (2.4.3) can be rewritten as

$$W_{\mathcal{D}}^B(\xi)=\max_{v_{\mathcal{D}}\in X_{\mathcal{D},0}\backslash\{0\}}\frac{1}{\|\mathcal{H}_{\mathcal{D}}^Bv_{\mathcal{D}}\|}\left|\int_\Omega\Big(\Delta\phi\Pi_{\mathcal{D}}v_{\mathcal{D}}-\phi\Delta_{\mathcal{D}}v_{\mathcal{D}}\Big)\,\mathrm{d}\boldsymbol{x}\right|, \tag{2.4.57}$$

where $\phi=\mathrm{tr}(\xi)$. Define

$$\widehat{\delta}_\sigma\phi=\begin{cases}\phi(\boldsymbol{x}_{K_\sigma^+})-\phi(\boldsymbol{x}_{K_\sigma^-}) & \forall\sigma\in\mathcal{F}_{\mathrm{int}}\\ \phi(\boldsymbol{z}_\sigma)-\phi(\boldsymbol{x}_{K_\sigma}) & \forall\sigma\in\mathcal{F}_{\mathrm{ext}},\end{cases} \tag{2.4.58}$$

where $\boldsymbol{z}_\sigma$ is the orthogonal projection of $\boldsymbol{x}_K$ on the hyperplane which contains σ. For $\xi \in H^2(\Omega)^{d\times d}$, the divergence theorem implies that

$$\int_\Omega \Delta\phi \Pi_\mathcal{D} v_\mathcal{D} \,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{M}} \int_K \Delta\phi \Pi_\mathcal{D} v_\mathcal{D} \,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{F}_K} v_K \int_\sigma \nabla\phi\cdot n_{K,\sigma} \,\mathrm{d}s(\boldsymbol{x}).$$

Gather over the edges and use the definition of δ_σ to obtain

$$\begin{aligned}
\int_\Omega \Delta\phi \Pi_\mathcal{D} v_\mathcal{D} \,\mathrm{d}\boldsymbol{x} &= -\sum_{\sigma\in\mathcal{F}} \delta_\sigma v_\mathcal{D} \int_\sigma \nabla\phi\cdot n_\sigma \,\mathrm{d}s(\boldsymbol{x}) \\
&= -\sum_{\sigma\in\mathcal{F}} \delta_\sigma v_\mathcal{D} \int_\sigma \Big(\frac{\widehat{\delta}_\sigma\phi}{d_\sigma} + \nabla\phi\cdot n_\sigma - \frac{\widehat{\delta}_\sigma\phi}{d_\sigma}\Big)\,\mathrm{d}s(\boldsymbol{x}) \\
&= -\sum_{\sigma\in\mathcal{F}} \delta_\sigma v_\mathcal{D} \frac{\widehat{\delta}_\sigma\phi|\sigma|}{d_\sigma} + \sum_{\sigma\in\mathcal{F}} \delta_\sigma v_\mathcal{D} \int_\sigma \Big(\frac{\widehat{\delta}_\sigma\phi}{d_\sigma} - \nabla\phi\cdot n_\sigma\Big)\,\mathrm{d}s(\boldsymbol{x}).
\end{aligned} \tag{2.4.59}$$

Since $\delta_\sigma v_\mathcal{D} = 0$ for any $\sigma\in\mathcal{F}_{\rm ext}$, a use of (2.4.58), (2.3.9) and (2.3.10) imply

$$\begin{aligned}
-\sum_{\sigma\in\mathcal{F}} \delta_\sigma v_\mathcal{D} \frac{\widehat{\delta}_\sigma\phi|\sigma|}{d_\sigma} &= -\sum_{\sigma\in\mathcal{F}_{\rm int}} \frac{|\sigma|}{d_\sigma} \delta_\sigma v_\mathcal{D} \Big(\phi(\boldsymbol{x}_{K_\sigma}^+) - \phi(\boldsymbol{x}_{K_\sigma}^-)\Big) \\
&= \sum_{K\in\mathcal{M}} \phi(\boldsymbol{x}_K) \sum_{\sigma\in\mathcal{F}_K} \frac{|\sigma|}{d_\sigma} \delta_{K,\sigma} v_\mathcal{D} = \sum_{K\in\mathcal{M}} |K| \phi(\boldsymbol{x}_K) \Delta_K v_\mathcal{D}.
\end{aligned}$$

A substitution of this in (2.4.59) leads to

$$\int_\Omega \Delta\phi \Pi_\mathcal{D} v_\mathcal{D} \,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{M}} |K| \phi(\boldsymbol{x}_K) \Delta_K v_\mathcal{D} + \sum_{\sigma\in\mathcal{F}} \delta_\sigma v_\mathcal{D} \int_\sigma \Big(\frac{\widehat{\delta}_\sigma\phi}{d_\sigma} - \nabla\phi\cdot n_\sigma\Big)\,\mathrm{d}s(\boldsymbol{x}). \tag{2.4.60}$$

To deal with the first term, we first combine the two estimates in [48, Lemma 7.61] to see that

$$|\phi(\boldsymbol{x}_K) - \phi(\boldsymbol{y})| \le Ch|K|^{-1/2} \|\phi\|_{H^2(K)}, \qquad \forall \boldsymbol{y}\in K.$$

Hence, from the Cauchy–Schwarz inequality,

$$\begin{aligned}
\left| \sum_{K\in\mathcal{M}} |K| \phi(\boldsymbol{x}_K) \Delta_K v_\mathcal{D} - \int_\Omega \phi \Delta_\mathcal{D} v_\mathcal{D} \,\mathrm{d}\boldsymbol{x} \right| &= \left| \sum_{K\in\mathcal{M}} |K| \left(\phi(\boldsymbol{x}_K) - \frac{1}{|K|} \int_K \phi(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \right) \Delta_K v_\mathcal{D} \right| \\
&\le Ch \|\phi\|_{H^2(\Omega)} \left(\sum_{K\in\mathcal{M}} |K| |\Delta_K v_\mathcal{D}|^2 \right)^{1/2} = Ch \|\phi\|_{H^2(\Omega)} \|\Delta_\mathcal{D} v_\mathcal{D}\|.
\end{aligned} \tag{2.4.61}$$

Consider the second term in the right-hand side of (2.4.60). Note that the estimate on the terms $R_{K,\sigma}$ in [58, Proof of Theorem 3.4] show that

$$\left| \frac{\widehat{\delta}_\sigma\phi}{d_\sigma} - \nabla\phi\cdot n_\sigma \right| \le Ch \frac{\sqrt{|\sigma|}}{\sqrt{d_\sigma}} \|\mathcal{H}\phi\|_{L^2(\cup_{L\in\mathcal{M}_\sigma} L)^{d\times d}}.$$

A use of the Cauchy–Schwarz inequality yields

$$\left|\sum_{\sigma\in\mathcal{F}}\delta_\sigma v_\mathcal{D}\int_\sigma\left(\frac{\widehat{\delta}_\sigma\phi}{d_\sigma}-\nabla\phi\cdot n_\sigma\right)\mathrm{d}s(\boldsymbol{x})\right|\le Ch\|\mathcal{H}\phi\|\left(\sum_{\sigma\in\mathcal{F}}\frac{|\sigma|}{d_\sigma}(\delta_\sigma v_\mathcal{D})^2\right)^{1/2}$$
$$=Ch\|\phi\|_{H^2(\Omega)}\|v_\mathcal{D}\|_\mathcal{D}\le Ch\mathrm{diam}(\Omega)\|\phi\|_{H^2(\Omega)}\|\Delta_\mathcal{D}v_\mathcal{D}\|, \tag{2.4.62}$$

where (2.4.44) is used in the last line. Plug (2.4.61) and (2.4.62) into (2.4.60) to obtain

$$\left|\int_\Omega\Delta\phi\Pi_\mathcal{D}v_\mathcal{D}\,\mathrm{d}\boldsymbol{x}-\int_\Omega\phi\Delta_\mathcal{D}v_\mathcal{D}\,\mathrm{d}\boldsymbol{x}\right|\le Ch\|\phi\|_{H^2(\Omega)}\|\Delta_\mathcal{D}v_\mathcal{D}\|,$$

and the estimate on $W_\mathcal{D}(\xi)$ then follows from (2.4.57), recalling that $\phi=\mathrm{tr}(\xi)$. □

The following remark is a consequence of the results obtained in this section.

Remark 2.4.15 (Rates of convergence)**.** *Under regularity assumption $\overline{u}\in H^4(\Omega)\cap H^2_0(\Omega)$, for lower order conforming FEMs, Adini and Morley ncFEMs and the gradient recovery methods based on meshes with mesh parameter "h", $\mathcal{O}(h)$ estimates can be obtained for $W^B_\mathcal{D}(\mathcal{H}\overline{u})$ and $S^B_\mathcal{D}(\overline{u})$. Theorem 2.4.4 then gives a linear rate of convergence for these methods. For the finite volume method, if $\overline{u}\in H^4(\Omega)\cap H^2_0(\Omega)$, Theorem 2.4.4 provides an $\mathcal{O}(h^{1/4}|\ln(h)|)$ (in dimension $d=2$) or $\mathcal{O}(h^{3/13})$ (in dimension $d=3$) error estimate for the Hessian scheme based on the Hessian discretisation.*

2.5 Improved L^2 error estimates

The improved L^2 error estimate for HDM is presented in this section. This estimate is then applied to several methods, that is, FEMs, method based on GR operators and a slightly modified FVM (see Definition 2.5.7). The modified FVM has the same matrix as the original FVM, since only the quadrature of the source term is modified, but enjoys a super-convergence result while the standard FVM fails to super-converge. For establishing the lower order L^2 estimates, consider the adjoint problem corresponding to (2.2.2), and its Hessian scheme approximation.

The weak formulation for the dual problem with source term $g\in L^2(\Omega)$ seeks $\varphi_g\in V$ such that

$$a(w,\varphi_g)=(g,w)\text{ for all }w\in V. \tag{2.5.1}$$

The Hessian scheme corresponding to (2.5.1) seeks $\varphi_{g,\mathcal{D}}\in X_{\mathcal{D},0}$ such that

$$a_\mathcal{D}(w_\mathcal{D},\varphi_{g,\mathcal{D}})=(g,\Pi_\mathcal{D}w_\mathcal{D})\text{ for all }w_\mathcal{D}\in X_{\mathcal{D},0}. \tag{2.5.2}$$

The notation $X\lesssim Y$ means that $X\le CY$ for some C depending only on Ω and an upper bound of $C^B_\mathcal{D}$. For $\phi\in H^2(\Omega)$ with $\mathcal{H}\phi\in H^B(\Omega)$, set

$$\mathrm{WS}^B_\mathcal{D}(\phi):=W^B_\mathcal{D}(\mathcal{H}\phi)+S^B_\mathcal{D}(\phi). \tag{2.5.3}$$

Theorem 2.5.1 (Improved L^2 error estimate for Hessian schemes)**.**
Let $\overline{u}$ be the solution to (2.2.2)*. Let $\mathcal{D}$ be a B−Hessian discretisation in the sense of Definition 2.3.1, and let $u_\mathcal{D}$ be the solution to the Hessian scheme* (2.3.1)*. Define*

$$g = \frac{\overline{u} - \Pi_\mathcal{D} u_\mathcal{D}}{\|\overline{u} - \Pi_\mathcal{D} u_\mathcal{D}\|} \in L^2(\Omega)$$

and let φ_g be the solution to (2.5.1)*. Choose $\mathcal{P}_\mathcal{D}\overline{u}, \mathcal{P}_\mathcal{D}\varphi_g \in X_{\mathcal{D},0}$, where $\mathcal{P}_\mathcal{D}$ is a mapping from $H_0^2(\Omega)$ to $X_{\mathcal{D},0}$. Then*

$$\begin{aligned}\|\Pi_\mathcal{D} u_\mathcal{D} - \overline{u}\| \lesssim \left(\|\mathcal{H}_\mathcal{D}^B \mathcal{P}_\mathcal{D}\overline{u} - \mathcal{H}^B\overline{u}\| + \mathrm{WS}_\mathcal{D}^B(\overline{u})\right)\left(\|\mathcal{H}_\mathcal{D}^B \mathcal{P}_\mathcal{D}\varphi_g - \mathcal{H}^B\varphi_g\| + \mathrm{WS}_\mathcal{D}^B(\varphi_g)\right)\\ + \|\Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\overline{u} - \overline{u}\| + \|f\|\|\Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\varphi_g - \varphi_g\| + \left|\mathcal{W}_\mathcal{D}^B(\mathcal{H}\overline{u}, \mathcal{P}_\mathcal{D}\varphi_g)\right| + \left|\mathcal{W}_\mathcal{D}^B(\mathcal{H}\varphi_g, \mathcal{P}_\mathcal{D}\overline{u})\right|,\end{aligned}$$

where $\mathrm{WS}_\mathcal{D}^B$ is defined by (2.5.3)*, and $\mathcal{W}_\mathcal{D}^B$ is defined by* (2.4.4)*.*

Remark 2.5.2 (Dominating terms)**.** *Following Remark 2.4.15, for FEMs and GR methods, it is expected that $\mathrm{WS}_D^B(\overline{u}) = \mathcal{O}(h)$ if $\overline{u} \in H^4(\Omega) \cap H_0^2(\Omega)$. Hence, Theorem 2.5.1 provides an improved result if we can find a mapping $\mathcal{P}_\mathcal{D}$ (usually an interpolant) such that $\|\mathcal{H}_\mathcal{D}^B\mathcal{P}_\mathcal{D}\phi - \mathcal{H}^B\phi\| = \mathcal{O}(h)$, $\|\Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\phi - \phi\| = \mathcal{O}(h^2)$, $\mathcal{W}_\mathcal{D}^B(\xi, \mathcal{P}_\mathcal{D}\phi) = \mathcal{O}(h^2)$ for all $\phi \in H^4(\Omega) \cap H_0^2(\Omega)$ and all $\xi \in H^2(\Omega)^{d\times d}$.*

To prove Theorem 2.5.1, we shall make use of the following Lemma, which estimates the error associated with the continuous bilinear form $a(\cdot,\cdot)$ and discrete bilinear form $a_\mathcal{D}(\cdot,\cdot)$.

Lemma 2.5.3. *Let $\psi, \phi \in H_0^2(\Omega)$ be such that $\mathcal{H} : A\mathcal{H}\psi \in L^2(\Omega)$ and $\mathcal{H} : A\mathcal{H}\phi \in L^2(\Omega)$. Then, for any $\psi_\mathcal{D}, \phi_\mathcal{D} \in X_{\mathcal{D},0}$, the following holds:*

$$|a(\psi,\phi) - a_\mathcal{D}(\psi_\mathcal{D}, \phi_\mathcal{D})| \le \mathrm{E}_\mathcal{D}(\psi, \phi, \psi_\mathcal{D}, \phi_\mathcal{D}), \tag{2.5.4}$$

where

$$\begin{aligned}\mathrm{E}_\mathcal{D}(\psi,\phi,\psi_\mathcal{D},\phi_\mathcal{D}) = |\mathcal{W}_\mathcal{D}^B(\mathcal{H}\psi, \phi_\mathcal{D})| + |\mathcal{W}_\mathcal{D}^B(\mathcal{H}\phi, \psi_\mathcal{D})| + \|\Pi_\mathcal{D}\psi_\mathcal{D} - \psi\|\,\|\mathcal{H} : A\mathcal{H}\phi\|\\ + \|\Pi_\mathcal{D}\phi_\mathcal{D} - \phi\|\,\|\mathcal{H} : A\mathcal{H}\psi\| + \|\mathcal{H}_\mathcal{D}^B\psi_\mathcal{D} - \mathcal{H}^B\psi\|\,\|\mathcal{H}_\mathcal{D}^B\phi_\mathcal{D} - \mathcal{H}^B\phi\|.\end{aligned} \tag{2.5.5}$$

Proof. Use the definitions of $a(\cdot,\cdot)$ and $a_\mathcal{D}(\cdot,\cdot)$ and perform elementary manipulations to obtain

$$\begin{aligned}a(\psi,\phi) - a_\mathcal{D}(\psi_\mathcal{D},\phi_\mathcal{D}) &= \int_\Omega \mathcal{H}^B\psi : \mathcal{H}^B\phi \,\mathrm{d}\boldsymbol{x} - \int_\Omega \mathcal{H}_\mathcal{D}^B\psi_\mathcal{D} : \mathcal{H}_\mathcal{D}^B\phi_\mathcal{D}\,\mathrm{d}\boldsymbol{x}\\ &= \int_\Omega (\mathcal{H}^B\psi - \mathcal{H}_\mathcal{D}^B\psi_\mathcal{D}) : \mathcal{H}^B\phi\,\mathrm{d}\boldsymbol{x}\\ &\quad + \int_\Omega (\mathcal{H}_\mathcal{D}^B\psi_\mathcal{D} - \mathcal{H}^B\psi) : (\mathcal{H}^B\phi - \mathcal{H}_\mathcal{D}^B\phi_\mathcal{D})\,\mathrm{d}\boldsymbol{x}\\ &\quad + \int_\Omega \mathcal{H}^B\psi : (\mathcal{H}^B\phi - \mathcal{H}_\mathcal{D}^B\phi_\mathcal{D})\,\mathrm{d}\boldsymbol{x} =: T_1 + T_2 + T_3.\end{aligned} \tag{2.5.6}$$

T_1 can be estimated using integration by parts twice and (2.4.4).

$$\begin{aligned} T_1 &= \int_\Omega \mathcal{H}^B\psi : \mathcal{H}^B\phi \,\mathrm{d}\boldsymbol{x} - \int_\Omega \mathcal{H}^B_{\mathcal{D}}\psi_{\mathcal{D}} : \mathcal{H}^B\phi \,\mathrm{d}\boldsymbol{x} \\ &= \int_\Omega \psi(\mathcal{H} : A\mathcal{H}\phi)\,\mathrm{d}\boldsymbol{x} + \mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\phi, \psi_{\mathcal{D}}) - \int_\Omega (\mathcal{H} : A\mathcal{H}\phi)\Pi_{\mathcal{D}}\psi_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}. \end{aligned}$$

A use of the Cauchy–Schwarz inequality leads to

$$|T_1| \le |\mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\phi, \psi_{\mathcal{D}})| + \|\mathcal{H} : A\mathcal{H}\phi\|\,\|\psi - \Pi_{\mathcal{D}}\psi_{\mathcal{D}}\|. \tag{2.5.7}$$

A use of the Cauchy–Schwarz inequality leads to an upper bound for the term T_2 as

$$|T_2| \le \|\mathcal{H}^B\psi - \mathcal{H}^B_{\mathcal{D}}\psi_{\mathcal{D}}\|\,\|\mathcal{H}^B\phi - \mathcal{H}^B_{\mathcal{D}}\phi_{\mathcal{D}}\|. \tag{2.5.8}$$

The term T_3 is estimated exactly as T_1 interchanging the roles of $(\psi, \psi_{\mathcal{D}})$ and $(\phi, \phi_{\mathcal{D}})$, which leads to

$$|T_3| \le |\mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\psi, \phi_{\mathcal{D}})| + \|\mathcal{H} : A\mathcal{H}\psi\|\,\|\phi - \Pi_{\mathcal{D}}\phi_{\mathcal{D}}\|. \tag{2.5.9}$$

A substitution of the estimates (2.5.7)–(2.5.9) into (2.5.6) leads to (2.5.4). □

We now prove the main result given by Theorem 2.5.1. Note that the proof is obtained by modification of the arguments of [54, Theorem 3.1] in the GDM framework to that of HDM.

Proof of Theorem 2.5.1. Choose $w = \overline{u}$ in (2.5.1) and $w_{\mathcal{D}} = u_{\mathcal{D}}$ in (2.5.2),

$$\|\overline{u} - \Pi_{\mathcal{D}}u_{\mathcal{D}}\| = (g, \overline{u} - \Pi_{\mathcal{D}}u_{\mathcal{D}}) = a(\overline{u}, \varphi_g) - a_{\mathcal{D}}(u_{\mathcal{D}}, \varphi_{g,\mathcal{D}}). \tag{2.5.10}$$

Since $\overline{u}$ and φ_g both belong to $H^2_0(\Omega)$ with $\mathcal{H} : A\mathcal{H}\overline{u} = f \in L^2(\Omega)$ and $\mathcal{H} : A\mathcal{H}\varphi_g = g \in L^2(\Omega)$, a use of (2.5.4) in (2.5.10) with some manipulations lead to

$$\begin{aligned} \|\overline{u} - \Pi_{\mathcal{D}}u_{\mathcal{D}}\| &= a(\overline{u}, \varphi_g) - a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) + a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) - a_{\mathcal{D}}(u_{\mathcal{D}}, \varphi_{g,\mathcal{D}}) \\ &\le \mathrm{E}_{\mathcal{D}}(\overline{u}, \varphi_g, \mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) + a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) - a_{\mathcal{D}}(u_{\mathcal{D}}, \varphi_{g,\mathcal{D}}) \\ &= a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g - \varphi_{g,\mathcal{D}}) + a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u} - u_{\mathcal{D}}, \varphi_{g,\mathcal{D}}) \\ &\quad + \mathrm{E}_{\mathcal{D}}(\overline{u}, \varphi_g, \mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) =: T_1 + T_2 + \mathrm{E}_{\mathcal{D}}(\overline{u}, \varphi_g, \mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g). \end{aligned} \tag{2.5.11}$$

An introduction of $a(\overline{u}, \varphi_g)$, a use of the triangle inequality, (2.5.4), (2.5.1) with $w = \overline{u}$, (2.5.2) with $w_{\mathcal{D}} = \mathcal{P}_{\mathcal{D}}\overline{u}$ and the Cauchy–Schwarz inequality yields

$$\begin{aligned} |T_1| &\le |a(\overline{u}, \varphi_g) - a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}, \varphi_{g,\mathcal{D}})| + |a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) - a(\overline{u}, \varphi_g)| \\ &\le |a(\overline{u}, \varphi_g) - a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}, \varphi_{g,\mathcal{D}})| + \mathrm{E}_{\mathcal{D}}(\overline{u}, \varphi_g, \mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) \\ &\le |(g, \overline{u} - \Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u})| + \mathrm{E}_{\mathcal{D}}(\overline{u}, \varphi_g, \mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g) \\ &\le \|g\|\,\|\overline{u} - \Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + \mathrm{E}_{\mathcal{D}}(\overline{u}, \varphi_g, \mathcal{P}_{\mathcal{D}}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_g). \end{aligned} \tag{2.5.12}$$

We now turn to T_2. Introduce the terms $a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u},\mathcal{P}_{\mathcal{D}}\varphi_g)$, $a_{\mathcal{D}}(u_{\mathcal{D}},\mathcal{P}_{\mathcal{D}}\varphi_g)$ and choose $v_{\mathcal{D}}=\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}}$ in (2.3.1) to deduce

$$\begin{aligned}T_2 &= -a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u},\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}})+a_{\mathcal{D}}(u_{\mathcal{D}},\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}})+a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}-u_{\mathcal{D}},\mathcal{P}_{\mathcal{D}}\varphi_g)\\ &= -\left[a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u},\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}})-(f,\Pi_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}}))\right]+a_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\overline{u}-u_{\mathcal{D}},\mathcal{P}_{\mathcal{D}}\varphi_g)\\ &= -T_{2,1}+T_{2,2}.\end{aligned} \tag{2.5.13}$$

Since $\mathcal{H}:A\mathcal{H}\overline{u}=f$, (2.4.4) yields

$$\begin{aligned}T_{2,1} &= \int_\Omega \left(\mathcal{H}^B\overline{u}:\mathcal{H}^B_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}})-f\Pi_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}})\right)\mathrm{d}\boldsymbol{x}\\ &\quad+\int_\Omega(\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}):\mathcal{H}^B_{\mathcal{D}}(\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}})\,\mathrm{d}\boldsymbol{x}\\ &= \int_\Omega(\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}):(\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g-\mathcal{H}^B_{\mathcal{D}}\varphi_{g,\mathcal{D}})\,\mathrm{d}\boldsymbol{x}-\mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\overline{u},\mathcal{P}_{\mathcal{D}}\varphi_g-\varphi_{g,\mathcal{D}}).\end{aligned}$$

Therefore, apply (2.4.3), the Cauchy–Schwarz inequality, (2.5.3), a triangle inequality and (2.4.4) to obtain

$$\begin{aligned}|T_{2,1}| &\le W^B_{\mathcal{D}}(\mathcal{H}\overline{u})\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g-\mathcal{H}^B_{\mathcal{D}}\varphi_{g,\mathcal{D}}\|+\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\|\,\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g-\mathcal{H}^B_{\mathcal{D}}\varphi_{g,\mathcal{D}}\|\\ &\lesssim \|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g-\mathcal{H}^B_{\mathcal{D}}\varphi_{g,\mathcal{D}}\|\left(\mathrm{WS}^B_{\mathcal{D}}(\overline{u})+\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\|\right)\\ &\lesssim \left(\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g-\mathcal{H}^B\varphi_g\|+\mathrm{WS}^B_{\mathcal{D}}(\varphi_g)\right)\left(\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\|+\mathrm{WS}^B_{\mathcal{D}}(\overline{u})\right).\end{aligned} \tag{2.5.14}$$

The term $T_{2,2}$ is similar to T_1, upon swapping the primal and dual problems, $(f,\overline{u},u_{\mathcal{D}},g,\varphi_g,\varphi_{g,\mathcal{D}})$ $\leftrightarrow(g,\varphi_g,\varphi_{g,\mathcal{D}},f,\overline{u},u_{\mathcal{D}})$. Hence, from (2.5.12),

$$|T_{2,2}|\le\|f\|\,\|\varphi_g-\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g\|+\mathrm{E}_{\mathcal{D}}(\overline{u},\varphi_g,\mathcal{P}_{\mathcal{D}}\overline{u},\mathcal{P}_{\mathcal{D}}\varphi_g). \tag{2.5.15}$$

Combine the estimates (2.5.13), (2.5.14) and (2.5.15) to obtain

$$\begin{aligned}|T_2| \lesssim{}& \left(\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\|+\mathrm{WS}^B_{\mathcal{D}}(\overline{u})\right)\left(\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g-\mathcal{H}^B\varphi_g\|+\mathrm{WS}^B_{\mathcal{D}}(\varphi_g)\right)\\ &+\|f\|\,\|\varphi_g-\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g\|+\mathrm{E}_{\mathcal{D}}(\overline{u},\varphi_g,\mathcal{P}_{\mathcal{D}}\overline{u},\mathcal{P}_{\mathcal{D}}\varphi_g).\end{aligned} \tag{2.5.16}$$

A substitution of (2.5.12) and (2.5.16) in (2.5.11) leads to

$$\begin{aligned}\|\overline{u}-\Pi_{\mathcal{D}}u_{\mathcal{D}}\| \lesssim{}& \left(\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}-\mathcal{H}^B\overline{u}\|+\mathrm{WS}^B_{\mathcal{D}}(\overline{u})\right)\left(\|\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g-\mathcal{H}^B\varphi_g\|+\mathrm{WS}^B_{\mathcal{D}}(\varphi_g)\right)\\ &+\|\overline{u}-\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\|+\|f\|\,\|\varphi_g-\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_g\|+\mathrm{E}_{\mathcal{D}}(\overline{u},\varphi_g,\mathcal{P}_{\mathcal{D}}\overline{u},\mathcal{P}_{\mathcal{D}}\varphi_g),\end{aligned}$$

where we have used the fact that $\|g\|=1$. Finally, the proof is complete by using the definition (2.5.5) of $\mathrm{E}_{\mathcal{D}}$ and noticing that $\mathcal{H}:A\mathcal{H}\overline{u}=f\in L^2(\Omega)$ and $\mathcal{H}:A\mathcal{H}\varphi_g=g\in L^2(\Omega)$. □

The application of the above theorem to various schemes is discussed below.

Proposition 2.5.4. *Let $\overline{u}\in H^4(\Omega)\cap H^2_0(\Omega)$ be the solution to* (2.2.2) *and $u_{\mathcal{D}}$ be the solution to the Hessian scheme* (2.3.1)*. Then, for low-order conforming FEMs, Adini and Morley ncFEMs, and gradient recovery methods, there exists a constant $C>0$ not depending on h such that*

$$\|\Pi_{\mathcal{D}}u_{\mathcal{D}}-\overline{u}\|\le Ch^2.$$

The following Lemmas 2.5.5-2.5.6 helps to prove Proposition 2.5.4. We start by a preliminary result that states the approximation properties of the classical interpolant $\mathcal{P}_{\mathcal{D}}$ for various methods.

Lemma 2.5.5 (Interpolation [41]). *Let $\psi \in H^3(\Omega) \cap H_0^2(\Omega)$ and $\phi \in H^4(\Omega) \cap H_0^2(\Omega)$. The classical interpolant $\mathcal{P}_{\mathcal{D}}$ satisfies*

(i) For conforming FEMs and Morley ncFEM,

$$\|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi - \psi\| \leq Ch^3, \|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi - \nabla\psi\| \leq Ch^2 \text{ and } \|\mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\psi - \mathcal{H}^B\psi\| \leq Ch.$$

(ii) For Adini ncFEM,

$$\|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi - \phi\| \leq Ch^4, \|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi - \nabla\phi\| \leq Ch^3 \text{ and } \|\mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\phi - \mathcal{H}^B\phi\| \leq Ch^2.$$

The next lemma establishes an estimate on the limit–conformity measure $\mathcal{W}_{\mathcal{D}}^B$ given by (2.4.4) for various schemes.

Lemma 2.5.6. *Let $\xi \in H^2(\Omega)^{d\times d}$, $\psi \in H^3(\Omega) \cap H_0^2(\Omega)$ and $\phi \in H^4(\Omega) \cap H_0^2(\Omega)$.*

(i) For conforming FEMs, we have $\mathcal{W}_{\mathcal{D}}^B(\xi, \mathcal{P}_{\mathcal{D}}\psi) = 0$.

(ii) For Adini ncFEM, $\mathcal{W}_{\mathcal{D}}^B(\xi, \mathcal{P}_{\mathcal{D}}\phi) = \mathcal{O}(h^2)$.

(iii) For Morley ncFEM and gradient recovery methods, $\mathcal{W}_{\mathcal{D}}^B(\xi, \mathcal{P}_{\mathcal{D}}\psi) = \mathcal{O}(h^2)$.

Proof. (i) CONFORMING FEMS. Since $X_{\mathcal{D},0} \subseteq H_0^2(\Omega)$, using integration by parts twice, the limit-conformity measure vanishes, that is, $\mathcal{W}_{\mathcal{D}}^B = 0$.

(ii) NONCONFORMING FEM: THE ADINI RECTANGLE. Let $\phi \in H^4(\Omega) \cap H_0^2(\Omega)$ and $\xi \in H^2(\Omega)^{d\times d}$. Introduce the term $(\mathcal{H} : A\xi)\phi$ in (2.4.4), use the Cauchy–Schwarz inequality and Lemma 2.5.5 to obtain

$$\begin{aligned}
\left|\mathcal{W}_{\mathcal{D}}^B(\xi, \mathcal{P}_{\mathcal{D}}\phi)\right| &\leq \left| \int_\Omega \left((\mathcal{H} : A\xi)\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi - (\mathcal{H} : A\xi)\phi\right) \mathrm{d}\boldsymbol{x} \right| \\
&\qquad + \left| \int_\Omega \left((\mathcal{H} : A\xi)\phi - B\xi : \mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\phi\right) \mathrm{d}\boldsymbol{x} \right| \\
&\leq \|\mathcal{H} : A\xi\| \|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi - \phi\| + \left| \int_\Omega \left((\mathcal{H} : A\xi)\phi - B\xi : \mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\phi\right) \mathrm{d}\boldsymbol{x} \right| \\
&\leq Ch^4 + \left| \int_\Omega \left((\mathcal{H} : A\xi)\phi - B\xi : \mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\phi\right) \mathrm{d}\boldsymbol{x} \right|.
\end{aligned}$$

Apply integration by parts twice to deduce

$$\left|\mathcal{W}_{\mathcal{D}}^B(\xi, P_{\mathcal{D}}\phi)\right| \leq Ch^4 + \left| \int_\Omega \left(B\xi : \mathcal{H}^B\phi - B\xi : \mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\phi\right) \mathrm{d}\boldsymbol{x} \right|. \tag{2.5.17}$$

A use of the Cauchy–Schwarz inequality and Lemma 2.5.5 leads to

$$\left|\mathcal{W}_{\mathcal{D}}^B(\xi, P_{\mathcal{D}}\phi)\right| \leq Ch^4 + \|B\xi\| \|\mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\phi - \mathcal{H}^B\phi\| \leq Ch^2.$$

(iii)(a) NONCONFORMING FEM: THE MORLEY TRIANGLE. Proceed as in the proof of limit conformity $\mathcal{W}_{\mathcal{D}}^{B}(\xi,\mathcal{P}_{\mathcal{D}}\psi)$ for the Adini's rectangle (with $\|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi-\psi\|\le Ch^3$) and use (2.5.17) to arrive at

$$\mathcal{W}_{\mathcal{D}}^{B}(\xi,\mathcal{P}_{\mathcal{D}}\psi)\le Ch^3+\left|\int_\Omega\left(B\xi:\mathcal{H}^B\psi-B\xi:\mathcal{H}_{\mathcal{D}}^{B}\mathcal{P}_{\mathcal{D}}\psi\right)\mathrm{d}\boldsymbol{x}\right|. \tag{2.5.18}$$

Let ξ_K be the average value of ξ on the cell $K\in\mathcal{M}$. By the mesh regularity assumption, $\|\xi-\xi_K\|_{L^2(K)^{d\times d}}\le Ch\|\xi\|_{H^1(K)^{d\times d}}$ (see, e.g., [48, Lemma B.6]). An introduction of $B\xi_K$ in the above inequality and a use of the Cauchy–Schwarz inequality and Lemma 2.5.5 yield

$$\begin{aligned}\left|\mathcal{W}_{\mathcal{D}}^{B}(\xi,\mathcal{P}_{\mathcal{D}}\psi)\right|&\le Ch^3+\sum_{K\in\mathcal{M}}\|B\xi-B\xi_K\|_{L^2(K)^{d\times d}}\|\mathcal{H}_{\mathcal{D}}^{B}\psi-\mathcal{H}_{\mathcal{D}}^{B}\mathcal{P}_{\mathcal{D}}\psi\|_{L^2(K)^{d\times d}}\\&\quad+\left|\sum_{K\in\mathcal{M}}\int_K B\xi_K:(\mathcal{H}^B\psi-\mathcal{H}_{\mathcal{D}}^{B}\mathcal{P}_{\mathcal{D}}\psi)\,\mathrm{d}\boldsymbol{x}\right|\\&\le Ch^2+\left|\sum_{K\in\mathcal{M}}\int_K B\xi_K:(\mathcal{H}^B\psi-\mathcal{H}_{\mathcal{D}}^{B}\mathcal{P}_{\mathcal{D}}\psi)\,\mathrm{d}\boldsymbol{x}\right|.\end{aligned}$$

For $K\in\mathcal{M}$, we have [66]

$$\int_K\mathcal{H}_{\mathcal{D}}^{B}\mathcal{P}_{\mathcal{D}}\psi\,\mathrm{d}\boldsymbol{x}=\int_K\mathcal{H}^B\psi\,\mathrm{d}\boldsymbol{x}. \tag{2.5.19}$$

Hence, $\mathcal{W}_{\mathcal{D}}^{B}(\xi,\mathcal{P}_{\mathcal{D}}\psi)=\mathcal{O}(h^2)$.

(iii)(b) GRADIENT RECOVERY METHOD. Note that for the GR method, $\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi=\mathcal{P}_{\mathcal{D}}\psi\in V_h$, an H_0^1-conforming finite element space which contains the piecewise linear functions. From Theorem 2.4.10,

$$\|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi-\psi\|\le Ch^2,\|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi-\nabla\psi\|\le Ch^2\text{ and }\|\mathcal{H}_{\mathcal{D}}^{B}\mathcal{P}_{\mathcal{D}}\psi-\mathcal{H}^B\psi\|\le Ch. \tag{2.5.20}$$

Also, $\|\nabla\mathcal{P}_{\mathcal{D}}\psi-\nabla\psi\|\le Ch$. Let us consider $\mathcal{W}_{\mathcal{D}}^{B}(\xi,\mathcal{P}_{\mathcal{D}}\psi)$. Reproduce the same steps as in the proof for Adini's rectangle (with $\|\Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi-\psi\|\le Ch^2$), use (2.5.17) and the definition of reconstructed Hessian $\mathcal{H}_{\mathcal{D}}^{B}$ to obtain

$$\begin{aligned}\left|\mathcal{W}_{\mathcal{D}}^{B}(\xi,\mathcal{P}_{\mathcal{D}}\psi)\right|&\le Ch^2+\left|\int_\Omega\left(A\xi:\mathcal{H}\psi-A\xi:\nabla Q_h\nabla\mathcal{P}_{\mathcal{D}}\psi\right)\mathrm{d}\boldsymbol{x}\right|\\&\quad+\left|\int_\Omega A\xi:(\mathfrak{S}_h\otimes(Q_h\nabla\mathcal{P}_{\mathcal{D}}\psi-\nabla\mathcal{P}_{\mathcal{D}}\psi))\,\mathrm{d}\boldsymbol{x}\right|=:Ch^2+A_1+A_2.\end{aligned}$$

Since $Q_h\nabla\mathcal{P}_{\mathcal{D}}\psi\in H_0^1(\Omega)$, an integration by parts, the Cauchy–Schwarz inequality and the approximation property of $\mathcal{P}_{\mathcal{D}}$ given by Lemma 2.5.5 show that

$$\begin{aligned}|A_1|&=\left|-\int_\Omega\nabla\psi\cdot\mathrm{div}(A\xi)\,\mathrm{d}\boldsymbol{x}+\int_\Omega Q_h\nabla\mathcal{P}_{\mathcal{D}}\psi\cdot\mathrm{div}(A\xi)\,\mathrm{d}\boldsymbol{x}\right|\\&\le\|Q_h\nabla\mathcal{P}_{\mathcal{D}}\psi-\nabla\psi\|\,\|\mathrm{div}(A\xi)\|=\|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi-\nabla\psi\|\,\|\mathrm{div}(A\xi)\|\le Ch^2.\end{aligned}$$

Let ξ_K denote the average of ξ over $K \in \mathcal{M}$. Since the finite dimensional space V_h contains the piecewise linear functions, $\nabla V_h(K)$ contains the constant vector-valued functions on K, a use of the orthogonality property of the stabilisation function given by **(P5)**, the Cauchy–Schwarz inequality, the boundedness of $\mathfrak{S}_h$, the triangle inequality and the approximation properties of the interpolant leads to

$$\begin{aligned}
|A_2| &= \Big| \sum_{K\in\mathcal{M}} \int_K (A\xi - A\xi_K) : \mathfrak{S}_h \otimes (Q_h \nabla \mathcal{P}_\mathcal{D} \psi - \nabla \mathcal{P}_\mathcal{D} \psi) \, \mathrm{d}\boldsymbol{x} \Big| \\
&\leq C \sum_{K\in\mathcal{M}} \|\xi - \xi_K\|_{L^2(K)^{d\times d}} \|Q_h \nabla \mathcal{P}_\mathcal{D} \psi - \nabla \mathcal{P}_\mathcal{D} \psi\|_{L^2(K)^d} \\
&\leq Ch \|\nabla_\mathcal{D} \mathcal{P}_\mathcal{D} \psi - \nabla \mathcal{P}_\mathcal{D} \psi\| \\
&\leq Ch \Big(\|\nabla_\mathcal{D} \mathcal{P}_\mathcal{D} \psi - \nabla \psi\| + \|\nabla \psi - \nabla \mathcal{P}_\mathcal{D} \psi\| \Big) \leq Ch^2.
\end{aligned}$$

Therefore, $\mathcal{W}^B_\mathcal{D}(\xi, \mathcal{P}_\mathcal{D}\psi) = \mathcal{O}(h^2)$. □

Proof of Proposition 2.5.4. The proof of Proposition 2.5.4 follows from Theorem 2.4.4, Remark 2.4.15, Lemma 2.5.5, (2.5.20) and Lemma 2.5.6. □

Since the super-convergence is not known in general for two point flux approximation (TPFA) for second order problems, it is expected that the same issue occurs for the FVM mentioned in Section 2.3.1. In order to obtain an improved result, ideas developed in [54, Section 4] for GDM is appropriately modified for the HDM. For that, set

$$v_\sigma = \begin{cases} \frac{\mathrm{dist}(\boldsymbol{x}_K,\sigma) v_L + \mathrm{dist}(\boldsymbol{x}_L,\sigma) v_K}{d_\sigma} & \forall \sigma \in \mathcal{F}_{\mathrm{int}}, \ \mathcal{M}_\sigma = \{K, L\} \\ 0 & \forall \sigma \in \mathcal{F}_{\mathrm{ext}}. \end{cases} \tag{2.5.21}$$

We now define a slightly modified HDM for FVM based on Δ-adapted discretisations.

Definition 2.5.7 (Modified FVM $B-$HD)**.** *Let $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_\mathcal{D}, \nabla_\mathcal{D}, \mathcal{H}^B_\mathcal{D})$ be a $B-$ Hessian discretisation in the sense of Definition 2.3.1 for FVM. The modified FVM $B-$Hessian discretisation is $\mathcal{D}^* = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}^*}, \nabla_\mathcal{D}, \mathcal{H}^B_\mathcal{D})$, where the reconstruction function $\Pi_{\mathcal{D}^*}$ is defined by*

$$\forall v_\mathcal{D} \in X_{\mathcal{D},0}, \ \forall K \in \mathcal{M}, \ \forall \boldsymbol{x} \in K, \Pi_{\mathcal{D}^*} v_\mathcal{D}(\boldsymbol{x}) = \Pi_\mathcal{D} v_\mathcal{D}(\boldsymbol{x}) + \widetilde{\nabla}_K v_\mathcal{D} \cdot (\boldsymbol{x} - \boldsymbol{x}_K) \tag{2.5.22}$$

with

$$\widetilde{\nabla}_K v_\mathcal{D} = \frac{1}{|K|} \sum_{\sigma\in\mathcal{F}_K} |\sigma| v_\sigma \, \boldsymbol{n}_{K,\sigma}. \tag{2.5.23}$$

The Hessian scheme corresponding to the modified FVM $B-$HD $\mathcal{D}^*$ in the sense of Definition 2.5.7 is given by (2.3.1), in which only the right-hand side is modified. Thus, the modified FVM has the same matrix as the original FVM.

Consider now a super-admissible mesh in the sense of [48, Lemma 13.20], i.e. for $\sigma \in \mathcal{F}_{\mathrm{int}}$ with $\mathcal{M}_\sigma = \{K, L\}$, the straight line $(\boldsymbol{x}_K, \boldsymbol{x}_L)$ intersects σ at $\overline{\boldsymbol{x}}_\sigma$ (similarly on the boundary). This super-admissibility condition is satisfied by rectangles (with $\boldsymbol{x}_K$ the centre of mass of K) and acute triangles (with $\boldsymbol{x}_K$ the circumcenter of K).

Proposition 2.5.8 (Superconvergence for modified FVM HD)**.** *Let $\overline{u} \in H^4(\Omega) \cap H_0^2(\Omega)$ be the solution to* (2.2.2)*. Let $u_{\mathcal{D}^*}$ be the solution of the Hessian scheme* (2.3.1) *for the modified FVM B−HD $\mathcal{D}^*$ in the sense of Definition 2.5.7 on a super-admissible mesh. Then for the modified FVM based on Δ-adapted discretisations, there exist a constant $C > 0$ independent of h such that*

$$\|\Pi_{\mathcal{D}^*} u_{\mathcal{D}^*} - \overline{u}\| \leq C \begin{cases} h^{1/2} |\ln(h)|^2 & \text{if } d = 2 \\ h^{6/13} & \text{if } d = 3. \end{cases}$$

Recalling Remark 2.4.15, we see that these rates are an improvement over the rates in H^2 norm. Precisely, L^2 error estimate decays as the square of the H^2 error estimate.

Proof of Proposition 2.5.8. As a consequence of Stokes' formula, for $K \in \mathcal{M}$, $\sum_{\sigma \in \mathcal{F}_K} |\sigma| \, n_{K,\sigma} = 0$ (see the proof of [48, Lemma B.3]). A use of (2.5.21) and the superadmissible mesh condition $n_{K,\sigma} = \frac{\overline{\boldsymbol{x}}_\sigma - \boldsymbol{x}_K}{d_{K,\sigma}}$ leads to

$$\widetilde{\nabla}_K v_{\mathcal{D}} = \frac{1}{|K|} \sum_{\sigma \in \mathcal{F}_K} |\sigma| (v_\sigma - v_K) \, n_{K,\sigma} = \nabla_K v_{\mathcal{D}},$$

where $(\nabla_{\mathcal{D}} v_{\mathcal{D}})_K = \nabla_K v_{\mathcal{D}}$. Hence,

$$\int_K \nabla_{\mathcal{D}} v_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} = \int_K \nabla_K v_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} = |K| \widetilde{\nabla}_K v_{\mathcal{D}}.$$

Use the definition of D^*, the above relation between $\widetilde{\nabla}_K$ and $\nabla_{\mathcal{D}}$, and (5.3.6) to obtain

$$\forall v_{\mathcal{D}} \in X_{\mathcal{D},0}, \ \|\Pi_{\mathcal{D}} v_{\mathcal{D}} - \Pi_{\mathcal{D}^*} v_{\mathcal{D}}\|_{L^2(\Omega)} \lesssim h \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|.$$

Therefore, following the proof of [48, Remark 7.51], the same estimates on $C_{\mathcal{D}^*}^B$, $S_{\mathcal{D}^*}^B$ and $W_{\mathcal{D}^*}^B$ can be obtained for $\mathcal{D}^*$ as that for the original FVM HD $\mathcal{D}$. Thus, from Remark 4.2.3, under regularity assumption, an $\mathcal{O}(h^{1/4}|\ln(h)|)$ (in $d = 2$) or $\mathcal{O}(h^{3/13})$ (in $d = 3$) error estimate can be obtained for the Hessian scheme based on modified FVM HD $\mathcal{D}^*$. Note that to prove the error estimates for original FVM, the interpolation $P_{\mathcal{D}}$ is constructed by solving a TPFA scheme for second order problem, i.e, by considering $|K| \Delta_K \mathcal{P}_{\mathcal{D}} \phi = \int_K \Delta \phi \, \mathrm{d}\boldsymbol{x}$ for ϕ smooth enough and $K \in \mathcal{M}$. To preserve a superconvergence for this modified FVM, the idea is to construct $\mathcal{P}_{\mathcal{D}^*} \phi$ by solving the modified TPFA scheme, where $\Pi_{\mathcal{D}}$ is replaced by $\Pi_{\mathcal{D}^*}$. Since TPFA and Hybrid Mimetic Mixed (HMM) schemes are the same on superadmissible meshes, from [54, Theorem 4.6],

$$\|\Pi_{\mathcal{D}^*} \mathcal{P}_{\mathcal{D}^*} \phi - \phi\| \lesssim h^2 \|\phi\|_{H^2(\Omega)}. \tag{2.5.24}$$

To estimate $\mathcal{W}_{\mathcal{D}^*}^B(\xi, \mathcal{P}_{\mathcal{D}^*} \phi)$, for $\phi \in H^4(\Omega) \cap H_0^2(\Omega)$ and $\xi \in H^2(\Omega)^{d \times d}$, consider (2.4.4) with $\mathcal{D} = \mathcal{D}^*$. Introduce $(\mathcal{H} : A\xi)\phi$, use the Cauchy–Schwarz inequality, (2.5.24) and integration by

parts twice to obtain

$$\begin{aligned}\left|\mathcal{W}^B_{\mathcal{D}^*}(\xi,\mathcal{P}_{\mathcal{D}^*}\phi)\right| &\le \left|\int_\Omega \Big((\mathcal{H}:A\xi)(\Pi_{\mathcal{D}^*}\mathcal{P}_{\mathcal{D}^*}\phi-\phi)\Big)\,\mathrm{d}\boldsymbol{x}\right| \\ &\quad + \left|\int_\Omega \Big((\mathcal{H}:A\xi)\phi - B\xi : \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}^*}\phi\Big)\,\mathrm{d}\boldsymbol{x}\right| \\ &\le Ch^2 + \left|\int_\Omega B\xi : (\mathcal{H}^B\phi - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}^*}\phi)\,\mathrm{d}\boldsymbol{x}\right|.\end{aligned}$$

The second term on the right-hand side of the above inequality can be estimated by considering the projection of $B\xi$ on piecewise constant functions on the mesh $\mathcal{M}$. Let $B\xi_K$ be the projection of $B\xi$ on $K\in\mathcal{M}$. Since $\Delta_{\mathcal{D}}\mathcal{P}_{\mathcal{D}^*}\phi$ is the projection of $\Delta\phi$ on piecewise constant functions on $\mathcal{M}$ (that is, $|K|\Delta_K\mathcal{P}_{\mathcal{D}^*}\phi=\int_K\Delta\phi\,\mathrm{d}\boldsymbol{x}$), a use of the orthogonality property of the projection operator, the Cauchy–Schwarz inequality and the approximation property leads to

$$\left|\mathcal{W}^B_{\mathcal{D}^*}(\xi,\mathcal{P}_{\mathcal{D}^*}\phi)\right| \le Ch^2 + \left|\sum_{K\in\mathcal{M}}\int_{\mathcal{M}}(B\xi - B\xi_K) : (\mathcal{H}^B\phi - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}^*}\phi)\,\mathrm{d}\boldsymbol{x}\right| \le Ch^2. \tag{2.5.25}$$

A substitution of (2.5.24), (2.5.25) and estimates given by Remark 2.4.15 in Theorem 2.5.1 with $\mathcal{D}=\mathcal{D}^*$ yields

$$\|\Pi_{\mathcal{D}^*}u_{\mathcal{D}^*}-\overline{u}\| \le C\begin{cases} h^{1/2}|\ln(h)|^2 & \text{if } d=2,\\ h^{6/13} & \text{if } d=3.\end{cases}$$

Hence the proof of superconvergence result for the modified FVM is complete. □

2.6 Improved H^1-like error estimate

To establish an improved H^1-like error estimate, consider the following dual problem of (2.2.2). The weak formulation for the dual problem with source term $q\in H^{-1}(\Omega)$ seeks $\varphi_q\in V$ such that

$$a(w,\varphi_q) = (q,w) \text{ for all } w\in V. \tag{2.6.1}$$

Moreover, when Ω is convex, $\varphi_q\in H^3(\Omega)\cap H^2_0(\Omega)$ with a priori bound $\|\varphi_q\|_{H^3(\Omega)}\le\|q\|_{H^{-1}(\Omega)}$ [13]. In order to state the H^1-like error estimate, we need to consider the limit-conformity measure between the reconstructed Hessian $\mathcal{H}^B_{\mathcal{D}}$ and reconstructed gradient $\nabla_{\mathcal{D}}$. Define

$$\forall\chi\in H^B_{\mathrm{div}}(\Omega)^d,\ \widetilde{W}^B_{\mathcal{D}}(\chi) = \max_{w_{\mathcal{D}}\in X_{\mathcal{D},0}\setminus\{0\}}\frac{\left|\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\chi,w_{\mathcal{D}})\right|}{\|\mathcal{H}^B_{\mathcal{D}}w_{\mathcal{D}}\|}, \tag{2.6.2}$$

where $H^B_{\mathrm{div}}(\Omega)^d=\{\chi\in L^2(\Omega)^{d\times d}:\mathrm{div}(B^\tau B\chi)\in L^2(\Omega)^d\}$ and

$$\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\chi,w_{\mathcal{D}}) := \int_\Omega\Big(B\chi : \mathcal{H}^B_{\mathcal{D}}w_{\mathcal{D}} + \mathrm{div}(B^\tau B\chi)\cdot\nabla_{\mathcal{D}}w_{\mathcal{D}}\Big)\,\mathrm{d}\boldsymbol{x}. \tag{2.6.3}$$

Assume the existence of an operator $E_{\mathcal{D}}$ which maps the discrete unknowns to the continuous space of functions. This operator plays a central role in the H^1-like error estimate analysis for HDM.

Assumption 2.6.1 (Companion operator)**.** *Let $\mathcal{D}$ be a B–Hessian discretisation in the sense of Definition 2.3.1. There exists a linear map $E_{\mathcal{D}} : X_{\mathcal{D},0} \to H_0^2(\Omega)$ called the companion operator. We define*

$$\omega(E_{\mathcal{D}}) := \sup_{\psi_{\mathcal{D}} \in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}\|}{\|\mathcal{H}_{\mathcal{D}}^B\psi_{\mathcal{D}}\|}. \tag{2.6.4}$$

Along a sequence of Hessian discretisations $(\mathcal{D}_m)_{m\in\mathbb{N}}$, it is expected that the companion operators are defined such that $\omega(E_{\mathcal{D}_m}) \to 0$ as $m \to \infty$. For example, an explicit companion operator is well-known for the Morley element with $\omega(E_{\mathcal{D}}) = \mathcal{O}(h)$ [17].

Theorem 2.6.2 (Improved H^1-like error estimate for Hessian schemes)**.**
Let $\overline{u}$ be the solution to (2.2.2)*. Let $\mathcal{D}$ be a Hessian discretisation in the sense of Definition 2.3.1 and $u_{\mathcal{D}}$ be the solution to the Hessian scheme* (2.3.1)*. Assume that the solution to* (2.6.1) *satisfies $\varphi_q \in H^3(\Omega)\cap H_0^2(\Omega)$ and choose $\mathcal{P}_{\mathcal{D}}\overline{u}$, $\mathcal{P}_{\mathcal{D}}\varphi_q \in X_{\mathcal{D},0}$, where $\mathcal{P}_{\mathcal{D}} : H_0^2(\Omega) \to X_{\mathcal{D},0}$. Assume that there exists a companion operator $E_{\mathcal{D}}$ in the sense of Assumption 2.6.1 and define*

$$q = \frac{-\Delta E_{\mathcal{D}}(u_{\mathcal{D}} - \mathcal{P}_{\mathcal{D}}\overline{u})}{\|\nabla E_{\mathcal{D}}(u_{\mathcal{D}} - \mathcal{P}_{\mathcal{D}}\overline{u})\|} \in H^{-1}(\Omega).$$

Then

$$\begin{aligned}\|\nabla_{\mathcal{D}}u_{\mathcal{D}} - \nabla\overline{u}\| \lesssim{}& \big(\omega(E_{\mathcal{D}}) + \widetilde{W}_{\mathcal{D}}^B(\mathcal{H}\varphi_q)\big)\big(\mathrm{WS}_{\mathcal{D}}^B(\overline{u}) + \|\mathcal{H}^B\overline{u} - \mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\overline{u}\|\big)\\ &+ \|\nabla\overline{u} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + |\widetilde{\mathcal{W}}_{\mathcal{D}}^B(\mathcal{H}\varphi_q, \mathcal{P}_{\mathcal{D}}\overline{u})|\\ &+ \mathrm{WS}_{\mathcal{D}}^B(\overline{u})\|\mathcal{H}^B\varphi_q - \mathcal{H}_{\mathcal{D}}^B\mathcal{P}_{\mathcal{D}}\varphi_q\| + |\mathcal{W}_{\mathcal{D}}^B(\mathcal{H}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_q)|,\end{aligned}$$

where $\omega(E_{\mathcal{D}})$ is defined by (2.6.4)*, $\mathrm{WS}_{\mathcal{D}}^B$ is defined by* (2.5.3)*, $\mathcal{W}_{\mathcal{D}}^B$ is defined by* (2.4.4)*, $\widetilde{W}_{\mathcal{D}}^B$ is defined by* (2.6.2)*, and $\widetilde{\mathcal{W}}_{\mathcal{D}}^B$ is defined by* (2.6.3)*.*

Remark 2.6.3. *Following Remark 2.5.2, Theorem 2.6.2 gives an improved error estimate in H^1-like norm if we can find $\mathcal{P}_{\mathcal{D}}$ and $E_{\mathcal{D}}$ such that $\|\nabla\phi - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi\| = \mathcal{O}(h^2)$, $\widetilde{\mathcal{W}}_{\mathcal{D}}^B(\chi, \mathcal{P}_{\mathcal{D}}\phi) = \mathcal{O}(h^2)$, $\omega(E_{\mathcal{D}}) = \mathcal{O}(h)$ and $\widetilde{W}_{\mathcal{D}}^B(\chi) = \mathcal{O}(h)$ for all $\phi \in H^4(\Omega)\cap H_0^2(\Omega)$ and all $\chi \in H^1(\Omega)^{d\times d}$.*

Remark 2.6.4. *The companion operators actually come with estimates on function, gradient given by* (2.6.4) *and Hessian (see e.g., [17]). The estimates on function and Hessian are not needed in the error analysis and hence we leave them undefined.*

Proof of Theorem 2.6.2. A use of the triangle inequality leads to

$$\|\nabla_{\mathcal{D}}u_{\mathcal{D}} - \nabla\overline{u}\| \le \|\nabla_{\mathcal{D}}u_{\mathcal{D}} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + \|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u} - \nabla\overline{u}\|. \tag{2.6.5}$$

Let us estimate $\|\nabla_{\mathcal{D}} u_{\mathcal{D}} - \nabla_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \overline{u}\|$. Set $v_{\mathcal{D}} = u_{\mathcal{D}} - \mathcal{P}_{\mathcal{D}} \overline{u} \in X_{\mathcal{D},0}$. Introduce $\nabla E_{\mathcal{D}} v_{\mathcal{D}}$ and $\mathcal{H}^B \overline{u}$, and use triangle inequalities, (2.6.4) and Theorem 2.4.4 to deduce

$$\begin{aligned}\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\| &\le \|\nabla_{\mathcal{D}} v_{\mathcal{D}} - \nabla E_{\mathcal{D}} v_{\mathcal{D}}\| + \|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\| \le \omega(E_{\mathcal{D}}) \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\| + \|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\| \\ &\le \omega(E_{\mathcal{D}}) \big(\|\mathcal{H}^B_{\mathcal{D}} u_{\mathcal{D}} - \mathcal{H}^B \overline{u}\| + \|\mathcal{H}^B \overline{u} - \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \overline{u}\| \big) + \|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\| \\ &\lesssim \omega(E_{\mathcal{D}}) \big(\mathrm{WS}^B_{\mathcal{D}}(\overline{u}) + \|\mathcal{H}^B \overline{u} - \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \overline{u}\| \big) + \|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\|. \end{aligned} \tag{2.6.6}$$

Consider $\|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\|$. From (2.6.1) with $w = E_{\mathcal{D}} v_{\mathcal{D}}$,

$$\begin{aligned}\|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\| = a(E_{\mathcal{D}} v_{\mathcal{D}}, \varphi_q) &= \int_\Omega (\mathcal{H}^B E_{\mathcal{D}} v_{\mathcal{D}} - \mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}) : \mathcal{H}^B \varphi_q \, \mathrm{d}\boldsymbol{x} \\ &\quad + \int_\Omega \mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}} : \mathcal{H}^B \varphi_q \, \mathrm{d}\boldsymbol{x} =: T_1 + T_2. \end{aligned} \tag{2.6.7}$$

A use of integration by parts, (2.6.3), (2.6.2), the Cauchy–Schwarz inequality, (2.6.4), the triangle inequality and Theorem 2.4.4 yields

$$\begin{aligned}|T_1| &\le \int_\Omega |\mathrm{div}(A \mathcal{H} \varphi_q) \cdot (\nabla_{\mathcal{D}} v_{\mathcal{D}} - \nabla E_{\mathcal{D}} v_{\mathcal{D}})| \, \mathrm{d}\boldsymbol{x} + \widetilde{W}^B_{\mathcal{D}}(\mathcal{H} \varphi_q) \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\| \\ &\le \omega(E_{\mathcal{D}}) \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\| \|\mathrm{div}(A \mathcal{H} \varphi_q)\| + \widetilde{W}^B_{\mathcal{D}}(\mathcal{H} \varphi_q) \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\| \\ &\lesssim \big(\omega(E_{\mathcal{D}}) \|\mathrm{div}(A \mathcal{H} \varphi_q)\| + \widetilde{W}^B_{\mathcal{D}}(\mathcal{H} \varphi_q) \big) \big(\mathrm{WS}^B_{\mathcal{D}}(\overline{u}) + \|\mathcal{H}^B \overline{u} - \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \overline{u}\| \big). \end{aligned} \tag{2.6.8}$$

Simple manipulations leads to

$$\begin{aligned}T_2 &= \int_\Omega (\mathcal{H}^B \overline{u} - \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \overline{u}) : \mathcal{H}^B \varphi_q \, \mathrm{d}\boldsymbol{x} + \int_\Omega (\mathcal{H}^B_{\mathcal{D}} u_{\mathcal{D}} - \mathcal{H}^B \overline{u}) : (\mathcal{H}^B \varphi_q - \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q) \, \mathrm{d}\boldsymbol{x} \\ &\quad + \int_\Omega (\mathcal{H}^B_{\mathcal{D}} u_{\mathcal{D}} - \mathcal{H}^B \overline{u}) : \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q \, \mathrm{d}\boldsymbol{x} =: T_{2,1} + T_{2,2} + T_{2,3}. \end{aligned} \tag{2.6.9}$$

An integration by parts, (2.6.3) and the Cauchy–Schwarz inequality leads to

$$|T_{2,1}| \le \|\mathrm{div}(A \mathcal{H} \varphi_q)\| \|\nabla_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \overline{u} - \nabla \overline{u}\| + |\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\mathcal{H} \varphi_q, \mathcal{P}_{\mathcal{D}} \overline{u})|. \tag{2.6.10}$$

$T_{2,2}$ can be estimated using the Cauchy–Scharwz inequality and Theorem 2.4.4 as

$$|T_{2,2}| \le \|\mathcal{H}^B \overline{u} - \mathcal{H}^B_{\mathcal{D}} u_{\mathcal{D}}\| \|\mathcal{H}^B \varphi_q - \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q\| \lesssim \mathrm{WS}^B_{\mathcal{D}}(\overline{u}) \|\mathcal{H}^B \varphi_q - \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q\|. \tag{2.6.11}$$

Since $\mathcal{H} : A \mathcal{H} \overline{u} = f$, by (2.4.4) and (2.3.1) with $v_{\mathcal{D}} = \mathcal{P}_{\mathcal{D}} \varphi_q$, the term $T_{2,3}$ can be estimated as

$$\begin{aligned}T_{2,3} &\le -\int_\Omega (\mathcal{H} : A \mathcal{H} \overline{u}) \Pi_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q \, \mathrm{d}\boldsymbol{x} + \mathcal{W}^B_{\mathcal{D}}(\mathcal{H} \overline{u}, \mathcal{P}_{\mathcal{D}} \varphi_q) + \int_\Omega \mathcal{H}^B_{\mathcal{D}} u_{\mathcal{D}} : \mathcal{H}^B_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q \, \mathrm{d}\boldsymbol{x} \\ &= -\int_\Omega (\mathcal{H} : A \mathcal{H} \overline{u}) \Pi_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q \, \mathrm{d}\boldsymbol{x} + \mathcal{W}^B_{\mathcal{D}}(\mathcal{H} \overline{u}, \mathcal{P}_{\mathcal{D}} \varphi_q) + \int_\Omega f \Pi_{\mathcal{D}} \mathcal{P}_{\mathcal{D}} \varphi_q \, \mathrm{d}\boldsymbol{x} \\ &= \mathcal{W}^B_{\mathcal{D}}(\mathcal{H} \overline{u}, \mathcal{P}_{\mathcal{D}} \varphi_q). \end{aligned} \tag{2.6.12}$$

A substitution of (2.6.10)–(2.6.12) in (2.6.9) leads to

$$\begin{aligned}|T_2| \lesssim \|\mathrm{div}(A\mathcal{H}\varphi_q)\|\|\nabla\overline{u} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + |\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\mathcal{H}\varphi_q, \mathcal{P}_{\mathcal{D}}\overline{u})| \\ + \mathrm{WS}^B_{\mathcal{D}}(\overline{u})\|\mathcal{H}^B\varphi_q - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_q\| + |\mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_q)|. \end{aligned} \tag{2.6.13}$$

Plug (2.6.8) and (2.6.13) in (2.6.7) to obtain an estimate for $\|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\|$.

$$\begin{aligned}\|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\| \lesssim \big(\omega(E_{\mathcal{D}})\|\mathrm{div}(A\mathcal{H}\varphi_q)\| + \widetilde{W}^B_{\mathcal{D}}(\mathcal{H}\varphi_q)\big)\big(\mathrm{WS}^B_{\mathcal{D}}(\overline{u}) + \|\mathcal{H}^B\overline{u} - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\|\big) \\ + \|\mathrm{div}(A\mathcal{H}\varphi_q)\|\|\nabla\overline{u} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + |\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\mathcal{H}\varphi_q, \mathcal{P}_{\mathcal{D}}\overline{u})| \\ + \mathrm{WS}^B_{\mathcal{D}}(\overline{u})\|\mathcal{H}^B\varphi_q - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_q\| + |\mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_q)|. \end{aligned} \tag{2.6.14}$$

A use of the apriori bound for the dual problem $\|\varphi_q\|_{H^3(\Omega)} \lesssim 1$ yields

$$\begin{aligned}\|\nabla E_{\mathcal{D}} v_{\mathcal{D}}\| \lesssim \big(\omega(E_{\mathcal{D}}) + \widetilde{W}^B_{\mathcal{D}}(\mathcal{H}\varphi_q)\big)\big(\mathrm{WS}^B_{\mathcal{D}}(\overline{u}) + \|\mathcal{H}^B\overline{u} - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\|\big) \\ + \|\nabla\overline{u} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + |\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\mathcal{H}\varphi_q, \mathcal{P}_{\mathcal{D}}\overline{u})| \\ + \mathrm{WS}^B_{\mathcal{D}}(\overline{u})\|\mathcal{H}^B\varphi_q - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_q\| + |\mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_q)|. \end{aligned} \tag{2.6.15}$$

A substitution of (2.6.15) in (2.6.6) yields an estimate on $\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\|$ (with $v_{\mathcal{D}} = u_{\mathcal{D}} - \mathcal{P}_{\mathcal{D}}\overline{u} \in X_{\mathcal{D},0}$) which when plugged on (2.6.5) gives

$$\begin{aligned}\|\nabla_{\mathcal{D}} u_{\mathcal{D}} - \nabla\overline{u}\| \lesssim \big(\omega(E_{\mathcal{D}}) + \widetilde{W}^B_{\mathcal{D}}(\mathcal{H}\varphi_q)\big)\big(\mathrm{WS}^B_{\mathcal{D}}(\overline{u}) + \|\mathcal{H}^B\overline{u} - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\|\big) \\ + \|\nabla\overline{u} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\overline{u}\| + |\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\mathcal{H}\varphi_q, \mathcal{P}_{\mathcal{D}}\overline{u})| \\ + \mathrm{WS}^B_{\mathcal{D}}(\overline{u})\|\mathcal{H}^B\varphi_q - \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\varphi_q\| + |\mathcal{W}^B_{\mathcal{D}}(\mathcal{H}\overline{u}, \mathcal{P}_{\mathcal{D}}\varphi_q)| \end{aligned}$$

and this completes the proof. □

The following proposition talks about the H^1-like error estimate for lower order conforming and non-conforming FEMs.

Proposition 2.6.5. *Let $\overline{u} \in H^4(\Omega) \cap H^2_0(\Omega)$ be the solution to* (2.2.2) *and $u_{\mathcal{D}}$ be the solution to the Hessian scheme* (2.3.1)*. Then, for low-order conforming and non-conforming (Adini and Morley) FEMs, there exists a constant C, not depending on h, such that*

$$\|\nabla_{\mathcal{D}} u_{\mathcal{D}} - \nabla\overline{u}\| \leq Ch^2.$$

Proof. • CONFORMING FEMS. Let $\psi \in H^3(\Omega) \cap H^2_0(\Omega)$. Since $X_{\mathcal{D},0} \subseteq H^2_0(\Omega)$, by applying integration by parts, the limit-conformity measure $\widetilde{W}^B_{\mathcal{D}}$ vanishes. Also, companion operator $E_{\mathcal{D}}$ is nothing but the identity operator which implies $\omega(E_{\mathcal{D}}) = 0$. Hence, under regularity assumption on $\overline{u}$, combine these estimates along with Remark 2.4.15, Lemma 2.5.5 and Lemma 2.5.6 in Theorem 2.6.2 to obtain $\|\nabla_{\mathcal{D}} u_{\mathcal{D}} - \nabla\overline{u}\| \leq Ch^2$.

• NON-CONFORMING FEM: THE ADINI RECTANGLE. The estimate $\omega(E_{\mathcal{D}}) = \mathcal{O}(h)$ for a companion operator which maps the Adini rectangle to the Bogner–Fox–Schmit rectangle [41] has been done in [14]. For $\chi \in H^B_{\mathrm{div}}(\Omega)^d$ and $v_{\mathcal{D}} \in X_{\mathcal{D},0}$, cellwise integration by parts yields

$$\int_\Omega \big(B\chi : \mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}} + \mathrm{div}(A\chi)\cdot\nabla_{\mathcal{D}} v_{\mathcal{D}}\big)\,\mathrm{d}\boldsymbol{x} = \sum_{\sigma\in\mathcal{F}}\int_\sigma (A\chi n_\sigma)\cdot[\![\nabla_{\mathcal{D}} v_{\mathcal{D}}]\!]\,\mathrm{d}s(\boldsymbol{x}).$$

A use of Theorem 2.4.6 and (2.6.2) leads to $\widetilde{W}^B_{\mathcal{D}}(\chi) = \mathcal{O}(h)$. Let $\phi \in H^4(\Omega) \cap H^2_0(\Omega)$. Introduce $\mathrm{div}(A\chi) \cdot \nabla\phi$ in (2.6.3), use an integration by parts, the Cauchy–Schwarz inequality and Lemma 2.5.5 to obtain

$$\begin{aligned}\left|\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\chi, \mathcal{P}_{\mathcal{D}}\phi)\right| &\leq \left| \int_\Omega \left(B\chi : \mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi + \mathrm{div}(A\chi)\cdot\nabla\phi\right) \mathrm{d}\boldsymbol{x} \right| + \left| \int_\Omega \mathrm{div}(A\chi)\cdot(\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi - \nabla\phi)\,\mathrm{d}\boldsymbol{x} \right| \\ &= \left| \int_\Omega B\chi : (\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi - \mathcal{H}^B\phi)\,\mathrm{d}\boldsymbol{x} \right| + \left| \int_\Omega \mathrm{div}(A\chi)\cdot(\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\phi - \nabla\phi)\,\mathrm{d}\boldsymbol{x} \right| \leq Ch^2.\end{aligned}$$

The proof is complete by invoking Remark 2.4.15, Lemmas 2.5.5-2.5.6 and Theorem 2.6.2.

• NON-CONFORMING FEM: THE MORLEY TRIANGLE. For the Morley element, there exists a companion operator such that $\omega(E_{\mathcal{D}}) = \mathcal{O}(h)$, see [17] for more details. To estimate $\widetilde{W}^B_{\mathcal{D}}(\chi)$, where $\chi \in H^B_{\mathrm{div}}(\Omega)^d$, let $v_{\mathcal{D}} \in X_{\mathcal{D},0}$. An integration of parts yields

$$\int_\Omega \left(B\chi : \mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}} + \mathrm{div}(A\chi)\cdot\nabla_{\mathcal{D}} v_{\mathcal{D}}\right) = \sum_{\sigma\in\mathcal{F}} \int_\sigma (A\chi n_\sigma)\cdot[\![\nabla_{\mathcal{D}} v_{\mathcal{D}}]\!]\,\mathrm{d}s(\boldsymbol{x}). \tag{2.6.16}$$

From (2.4.20) and (2.6.2), $\widetilde{W}^B_{\mathcal{D}}(\chi) = \mathcal{O}(h)$. Let $\psi \in H^3(\Omega) \cap H^2_0(\Omega)$. In order to evaluate $\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\chi, \mathcal{P}_{\mathcal{D}}\psi)$, introduce $\mathrm{div}(A\chi)\cdot\nabla\psi$ in (2.6.3), use an integration by parts and the Morley interpolation property given by Lemma 2.5.5. Hence,

$$\left|\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\chi, \mathcal{P}_{\mathcal{D}}\psi)\right| \leq Ch^2 + \left| \int_\Omega B\chi : (\mathcal{H}^B_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\psi - \mathcal{H}^B\psi)\,\mathrm{d}\boldsymbol{x} \right|.$$

Now, reproduce the same steps as in the limit-conformity $\mathcal{W}^B_{\mathcal{D}}(\xi, \mathcal{P}_{\mathcal{D}}\psi)$ proof for the Morley triangle (with $\xi = \chi$) and thus from (2.5.18)–(2.5.19), $\widetilde{\mathcal{W}}^B_{\mathcal{D}}(\chi, \mathcal{P}_{\mathcal{D}}\psi) = \mathcal{O}(h^2)$.
As a consequence, for the Morley triangle, if $\overline{u} \in H^4(\Omega) \cap H^2_0(\Omega)$, combine the above estimates, Theorem 2.4.4, Remark 2.4.15, Lemmas 2.5.5–2.5.6 and Theorem 2.6.2 to obtain the required result. □

Remark 2.6.6. *The construction of a companion operator $E_{\mathcal{D}}$ for the method based on gradient recovery operators with $\omega_{\mathcal{D}}$ small enough is an open problem. Though there is a difficulty of constructing a proper companion operator and hence improved H^1 theoretical rate of convergence are not obtained, we observe that the numerical rates in H^1-like norm are better (see Table 2.2, Section 2.7.1). In numerical test for FVM, the H^2 and H^1 estimated rates of convergences appear to be both of order 1 (see Section 2.7.2). This seems to indicate that we cannot expect an improved estimate in H^1-like norm compared to the estimate in energy norm. Hence, the FVM method is probably not amenable to an application of Theorem 2.6.2 (which is an indication that there might not exist, for this method, a proper companion operator).*

2.7 Numerical results

In this section, the results of some numerical experiments for the gradient recovery method, finite volume method and modified finite volume method are presented. Numerical results for FEMs

are available in literature [25, 65]. All these tests are conducted on the biharmonic problem $\Delta^2\overline{u} = f$ with clamped boundary conditions and for various exact solutions $\overline{u}$.

2.7.1 Numerical results for Gradient Recovery Method

A few examples are presented to illustrate the theoretical estimates of Theorem 2.4.4 and Theorem 2.5.1 (Proposition 2.5.4) on the Hessian discretisation for GR method described in Section 2.3.1. The considered finite element space V_h is the conforming $\mathbb{P}_1$ space, and the implementation was done following the ideas in [82]. The following relative errors, and related orders of convergence, in $L^2(\Omega)$, $H^1(\Omega)$ and $H^2(\Omega)$ norms are presented:

$$\mathsf{err}_{\mathcal{D}}(\overline{u}) := \frac{\|\Pi_{\mathcal{D}}u_{\mathcal{D}} - \overline{u}\|}{\|\overline{u}\|}, \quad \mathsf{err}_{\mathcal{D}}(\nabla\overline{u}) := \frac{\|\nabla_{\mathcal{D}}u_{\mathcal{D}} - \nabla\overline{u}\|}{\|\nabla\overline{u}\|} = \frac{\|Q_h\nabla u_{\mathcal{D}} - \nabla\overline{u}\|}{\|\nabla\overline{u}\|},$$
$$\mathsf{err}(\nabla\overline{u}) := \frac{\|\nabla u_{\mathcal{D}} - \nabla\overline{u}\|}{\|\nabla\overline{u}\|}, \quad \mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u}) := \frac{\|\mathcal{H}^B_{\mathcal{D}}u_{\mathcal{D}} - \mathcal{H}\overline{u}\|}{\|\mathcal{H}\overline{u}\|} = \frac{\|\nabla(Q_h\nabla u_{\mathcal{D}}) - \mathcal{H}\overline{u}\|}{\|\mathcal{H}\overline{u}\|},$$

where $u_{\mathcal{D}}$ is the solution to the Hessian scheme (2.3.1). Figure 2.6 shows the initial triangulation of a square domain and its uniform refinement.

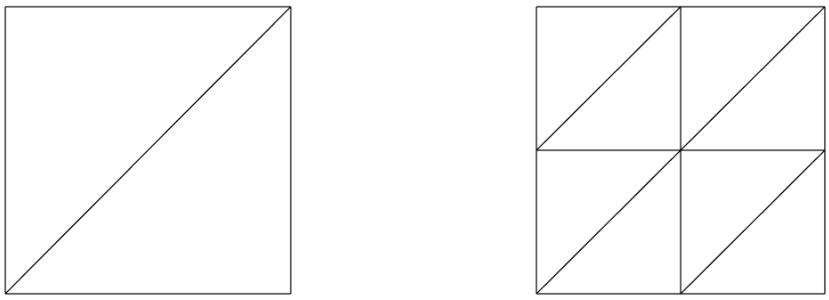

Figure 2.6: Initial triangulation and uniform refinement of square domain

The mesh data for the first four examples are in Table 2.1: mesh sizes h, numbers of unknowns (that is, the number of internal vertices) **nu**, and numbers of non-zero terms **nnz** in the square matrix of the system.

Example 1

The exact solution is chosen to be $\overline{u}(x,y) = x^2(x-1)^2y^2(y-1)^2$. To assess the effect of the stabilisation function $\mathfrak{S}_h$ on the results, we multiply it by a factor τ that takes the values 0.1, 1, 10, and 100.

The errors and orders of convergence for the mesh data for the first three examples numerical approximation to $\overline{u}$ are shown in Tables 2.2–2.5. It can be seen that the rate of convergence is quadratic in L^2-norm, which agrees with the theoretical result in Proposition 2.5.4, and linear in

Table 2.1: (GR) Mesh size, number of unknowns and number of non-zero terms in the square matrix

h	**nu**	**nnz**
0.353553	9	79
0.176777	49	1203
0.088388	225	7011
0.044194	961	32835
0.022097	3969	141315
0.011049	16129	585603

H^1-norm (see $\mathsf{err}(\nabla\overline{u})$). However, using gradient recovery operator, a quadratic order of convergence in H^1 norm is recovered (see $\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$). The rate of convergence in energy norm is linear (see $\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$), as expected by plugging the estimates of Theorem 2.4.10 into Theorem 2.4.4. We also notice a very small effect of τ on the relative errors and rates.

Table 2.2: (GR) Convergence results for the relative errors, Example 1, $\tau = 0.1$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	9.274702	-	31.591906	-	0.568338	-	0.595635	-
49	0.220095	5.3971	0.682922	5.5317	0.164105	1.7921	0.266927	1.1580
225	0.066997	1.7160	0.201282	1.7625	0.049395	1.7322	0.128410	1.0557
961	0.019135	1.8079	0.088805	1.1805	0.013697	1.8505	0.062164	1.0466
3969	0.005133	1.8983	0.040845	1.1205	0.003623	1.9185	0.030457	1.0293
16129	0.001331	1.9474	0.019422	1.0724	0.000933	1.9568	0.015059	1.0161

Table 2.3: (GR) Convergence results for the relative errors, Example 1, $\tau = 1$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	1.050930	-	3.254044	-	0.567670	-	0.582647	-
49	0.214195	2.2947	0.482686	2.7531	0.167145	1.7640	0.267188	1.1248
225	0.067498	1.6660	0.200108	1.2703	0.049952	1.7425	0.128511	1.0560
961	0.019240	1.8107	0.088667	1.1743	0.013806	1.8553	0.062184	1.0473
3969	0.005156	1.8999	0.040835	1.1186	0.003646	1.9209	0.030460	1.0296
16129	0.001336	1.9482	0.019421	1.0722	0.000938	1.9581	0.015060	1.0162

Example 2

Consider here the transcendental exact solution $\overline{u} = x^2(x-1)^2y^2(y-1)^2(\cos(2\pi x) + \sin(2\pi y))$, and $\tau = 0.1, 1$ and 10. Tables 2.6–2.8 presents the numerical results. The same comments as in

Table 2.4: (GR) Convergence results for the relative errors, Example 1, $\tau = 10$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	0.661894	-	0.778521	-	0.583641	-	0.586174	-
49	0.236529	1.4846	0.449484	0.7925	0.195127	1.5807	0.274030	1.0970
225	0.072610	1.7038	0.197892	1.1836	0.055493	1.8140	0.129911	1.0768
961	0.020303	1.8385	0.088413	1.1624	0.014907	1.8963	0.062418	1.0575
3969	0.005382	1.9154	0.040804	1.1156	0.003877	1.9429	0.030494	1.0335
16129	0.001387	1.9564	0.019417	1.0714	0.000990	1.9695	0.015064	1.0174

Table 2.5: (GR) Convergence results for the relative errors, Example 1, $\tau = 100$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	0.784444	-	0.805690	-	0.701021	-	0.695247	-
49	0.409420	0.9381	0.456340	0.8201	0.386868	0.8576	0.408281	0.7680
225	0.123166	1.7330	0.199370	1.1947	0.108498	1.8342	0.157333	1.3757
961	0.031509	1.9667	0.088447	1.1726	0.026358	2.0414	0.066443	1.2436
3969	0.007812	2.0121	0.040790	1.1166	0.006356	2.0521	0.031019	1.0990
16129	0.001934	2.0139	0.019414	1.0711	0.001552	2.0340	0.015130	1.0357

Example 1 can be made about the rates of convergence. Past the coarsest meshes, as in Example 1, τ only has a small impact on the relative errors.

Table 2.6: (GR) Convergence results for the relative errors, Example 2, $\tau = 0.1$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	89.040689	-	183.461721	-	1.211097	-	1.614525	-
49	0.825060	6.7538	3.401374	5.7532	0.235295	2.3638	0.501568	1.6866
225	0.076841	3.4246	0.337917	3.3314	0.050832	2.2107	0.172310	1.5414
961	0.017830	2.1076	0.114315	1.5637	0.013579	1.9044	0.079638	1.1135
3969	0.004565	1.9655	0.052228	1.1301	0.003638	1.9002	0.039166	1.0239
16129	0.001168	1.9662	0.025518	1.0333	0.000949	1.9391	0.019457	1.0093

Example 3

Here, $\overline{u}(x,y) = x^3y^3(1-x)^3(1-y)^3(e^x\sin(2\pi x) + \cos(2\pi x))$ and $\tau = 0.1, 1$ and 10. The results presented in Tables 2.9–2.11 are similar to those obtained for Examples 1 and 2.

Table 2.7: (GR) Convergence results for the relative errors, Example 2, $\tau = 1$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	10.222667	-	19.376883	-	1.058048	-	1.333720	-
49	0.475973	4.4247	1.467316	3.7231	0.229176	2.2069	0.473233	1.4948
225	0.074399	2.6775	0.313397	2.2271	0.050755	2.1748	0.170477	1.4730
961	0.017711	2.0706	0.112806	1.4742	0.013591	1.9009	0.079552	1.0996
3969	0.004547	1.9615	0.052162	1.1128	0.003640	1.9006	0.039162	1.0224
16129	0.001164	1.9657	0.025515	1.0317	0.000949	1.9393	0.019456	1.0092

Table 2.8: (GR) Convergence results for the relative errors, Example 2, $\tau = 10$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	1.413122	-	2.541143	-	0.845365	-	0.894504	-
49	0.313425	2.1727	0.878752	1.5319	0.225247	1.9081	0.396725	1.1729
225	0.066842	2.2293	0.262354	1.7439	0.051757	2.1217	0.165546	1.2609
961	0.016897	1.9840	0.109794	1.2567	0.013783	1.9089	0.079311	1.0616
3969	0.004376	1.9492	0.052012	1.0779	0.003675	1.9072	0.039149	1.0185
16129	0.001123	1.9621	0.025506	1.0280	0.000956	1.9425	0.019455	1.0088

Table 2.9: (GR) Convergence results for the relative errors, Example 3, $\tau = 0.1$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	81.804173	-	164.358300	-	1.068682	-	1.155266	-
49	0.677743	6.9153	2.358209	6.1230	0.232374	2.2013	0.517095	1.1597
225	0.093340	2.8602	0.447143	2.3989	0.048701	2.2544	0.207642	1.3163
961	0.017130	2.4459	0.125296	1.8354	0.010361	2.2328	0.084719	1.2933
3969	0.003975	2.1074	0.053941	1.2159	0.002643	1.9711	0.041197	1.0401
16129	0.000982	2.0167	0.026457	1.0278	0.000692	1.9341	0.020529	1.0049

Table 2.10: (GR) Convergence results for the relative errors, Example 3, $\tau = 1$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	8.708395	-	16.990965	-	0.950590	-	0.990455	-
49	0.516904	4.0744	1.490046	3.5113	0.224877	2.0797	0.492555	1.0078
225	0.089332	2.5326	0.414243	1.8468	0.048056	2.2263	0.203301	1.2767
961	0.016920	2.4005	0.122315	1.7599	0.010349	2.2153	0.084441	1.2676
3969	0.003953	2.0975	0.053813	1.1846	0.002646	1.9678	0.041186	1.0358
16129	0.000978	2.0153	0.026452	1.0246	0.000693	1.9337	0.020528	1.0045

Table 2.11: (GR) Convergence results for the relative errors, Example 3, $\tau = 10$

nu	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
9	1.097695	-	2.068091	-	0.809189	-	0.792818	-
49	0.351280	1.6438	0.969172	1.0935	0.205661	1.9762	0.409436	0.9533
225	0.073936	2.2483	0.306858	1.6592	0.046151	2.1558	0.186959	1.1309
961	0.015689	2.2365	0.113622	1.4333	0.010414	2.1478	0.083455	1.1637
3969	0.003756	2.0624	0.053444	1.0882	0.002689	1.9535	0.041142	1.0204
16129	0.000935	2.0068	0.026437	1.0155	0.000705	1.9309	0.020526	1.0032

Example 4

In this example, choose the right-hand side load function f such that the exact solution is given by $\overline{u}(x,y) = \sin^2(\pi x)\sin^2(\pi y)$. The computed errors and orders of convergence in the energy, H^1 and L^2 norms with $\tau = 1$ are shown in Table 2.12.

Table 2.12: (GR) Convergence results for the relative errors, Example 4, $\tau = 1$

h	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	3.124409	-	0.721457	-	0.855054	-
0.176777	0.145381	4.4257	0.099974	2.8513	0.246640	1.7936
0.088388	0.036224	2.0048	0.023098	2.1138	0.116470	1.0824
0.044194	0.009068	1.9982	0.005552	2.0566	0.057308	1.0232
0.022097	0.002261	2.0037	0.001363	2.0266	0.028470	1.0093
0.011049	0.000564	2.0032	0.000338	2.0116	0.014198	1.0037

Example 5

In this example, consider the non-convex L-shaped domain given by $\Omega = (-1,1)^2 \setminus \big([0,1)\times(-1,0]\big)$. Figure 2.7 shows the initial triangulation of a L-shaped domain and its uniform refinement. The source term f is chosen such that the model problem has the following exact singular solution [71]:

$$\bar{u} = (r^2\cos^2\theta - 1)^2(r^2\sin^2\theta - 1)^2 r^{1+\gamma} g_{\gamma,\omega}(\theta),$$

where (r,θ) denote the polar coordinates, $\gamma \approx 0.5444837367$ is a non-characteristic root of $\sin^2(\gamma\omega) = \gamma^2\sin^2(\omega)$, $\omega = \frac{3\pi}{2}$, and

$$\begin{aligned} g_{\gamma,\omega}(\theta) = &\left[\frac{1}{\gamma-1}\sin\big((\gamma-1)\omega\big) - \frac{1}{\gamma+1}\sin\big((\gamma+1)\omega\big)\right]\left[\cos\big((\gamma-1)\theta\big) - \cos\big((\gamma+1)\theta\big)\right] \\ &- \left[\frac{1}{\gamma-1}\sin\big((\gamma-1)\theta\big) - \frac{1}{\gamma+1}\sin\big((\gamma+1)\theta\big)\right]\left[\cos\big((\gamma-1)\omega\big) - \cos\big((\gamma+1)\omega\big)\right]. \end{aligned}$$

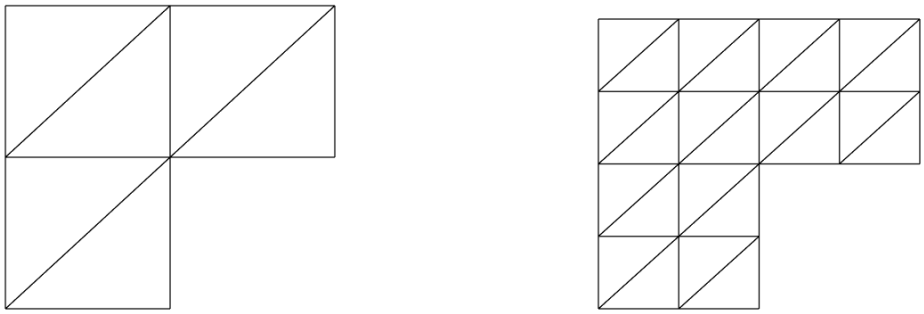

Figure 2.7: Initial triangulation and uniform refinement of L-shaped domain

This example is particularly interesting since the solution is less regular due to the corner singularity. The errors and rates of convergence with $\tau = 0.001$, 1 and 10 are reported in Tables 2.13–2.15 respectively. The domain Ω being nonconvex, we expect only suboptimal orders of convergence in the energy, H^1 and L^2 norms, and this can be clearly seen from the tables.

Table 2.13: (GR) Convergence results for the relative errors, Example 5, $\tau = 0.001$

h	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	1.488937	-	0.394870	-	0.504144	-
0.176777	0.185753	3.0028	0.139904	1.4969	0.218736	1.2046
0.088388	0.058874	1.6577	0.045530	1.6196	0.116520	0.9086
0.044194	0.018039	1.7065	0.013756	1.7267	0.065220	0.8372
0.022097	0.005400	1.7401	0.004197	1.7128	0.038827	0.7483
0.011049	0.001681	1.6835	0.001396	1.5882	0.024390	0.6707
0.005524	0.000570	1.5617	0.000526	1.4085	0.015899	0.6174

Table 2.14: (GR) Convergence results for the relative errors, Example 5, $\tau = 1$

h	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	0.447227	-	0.377554	-	0.441034	-
0.176777	0.177626	1.3322	0.142208	1.4087	0.217792	1.0180
0.088388	0.059387	1.5806	0.046087	1.6256	0.115943	0.9095
0.044194	0.018023	1.7203	0.013886	1.7307	0.064817	0.8390
0.022097	0.005360	1.7496	0.004231	1.7147	0.038615	0.7472
0.011049	0.001661	1.6897	0.001406	1.5894	0.024290	0.6688
0.005524	0.000562	1.5629	0.000529	1.4100	0.015854	0.6156

Table 2.15: (GR) Convergence results for the relative errors, Example 5, $\tau = 10$

h	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	0.488271	-	0.422393	-	0.472514	-
0.176777	0.197355	1.3069	0.162455	1.3785	0.226725	1.0594
0.088388	0.064165	1.6209	0.050639	1.6817	0.116820	0.9567
0.044194	0.019077	1.7500	0.014842	1.7706	0.064360	0.8601
0.022097	0.005598	1.7688	0.0044406	1.7408	0.038226	0.7516
0.011049	0.001718	1.7041	0.001455	1.6102	0.024090	0.6662
0.005524	0.000576	1.5759	0.000541	1.4277	0.015763	0.6119

Numerical tests that do not satisfy the assumption (M)

Here, two type of meshes that do not satisfy the assumption **(M)** in a sub-domain on the unit square domain $\Omega = (0,1)^2$ and $\overline{u}(x,y) = x^2y^2(1-x)^2(1-y)^2$ are considered. The source term can be computed using $f = \Delta^2 u$. The GR scheme was first tested on a series of uniform refinement meshes and then on a random version of redrefine meshes. Let **m** denote the number of internal vertices that do not satisfy the assumption **(M)**.

TEST 1: In this test, we consider the uniform mesh red-refinement process (Figure 2.8).

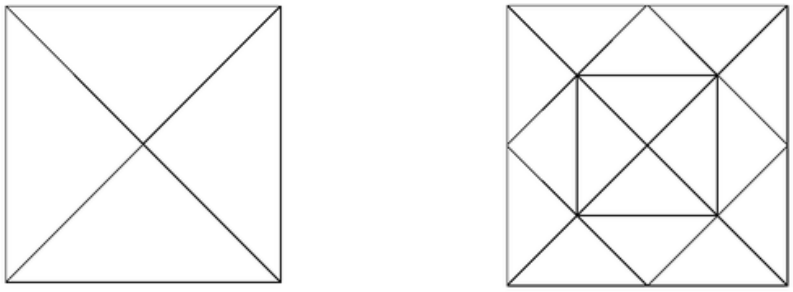

Figure 2.8: Initial triangulation and its uniform refinement

Table 2.16: (GR) mesh data, convergence results, Test 1, $\tau = 1$

h	**nu**	**m**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.250000	25	0	0.263383	-	0.125314	-	0.207352	-
0.125000	113	28	0.053456	2.3007	0.040531	1.6285	0.110634	0.9063
0.062500	481	60	0.015441	1.7915	0.013523	1.5836	0.063824	0.7936
0.031250	1985	124	0.004523	1.7715	0.004579	1.5624	0.040142	0.6690
0.015625	8065	252	0.001393	1.6988	0.001638	1.4831	0.026732	0.5865
0.007813	32513	508	0.000487	1.5170	0.000627	1.3859	0.018353	0.5426

Table 2.16 shows the mesh data and the errors in the L^2, H^1 and energy norms together with their orders of convergence. It can be seen that the GR method displays a loss of optimal rate in L^2, H^1 and energy norms.

TEST 2: Test 2 focuses on a series of random version of the redrefine meshes given by Figure 2.8 by moving each point by a random vector of magnitude $h/4$. Table 2.17 shows the results of the numerical experiment. The same comments as in Test 1 can be made about the rates of convergence.

Table 2.17: (GR) mesh data, convergence results, Test 2, $\tau = 1$

h	**nu**	**m**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.260180	25	0	0.280877	-	0.1414817	-	0.225532	-
0.131095	113	23	0.060605	2.2124	0.048725	1.5379	0.123086	0.8737
0.067403	481	61	0.018869	1.6834	0.017132	1.5079	0.072820	0.7573
0.035053	1985	435	0.006133	1.6213	0.006138	1.4810	0.048684	0.5809
0.018286	8065	4429	0.002254	1.4443	0.002495	1.2988	0.037045	0.3943
0.009588	32513	26499	0.001191	0.9197	0.001330	0.9074	0.031477	0.2350

2.7.2 Numerical results for FVM

In this section, numerical results based on the finite volume (FV) method are presented. As noticed, this scheme requires only one unknown per cell, and is therefore easy to implement and computationally cheap. The schemes were first tested on a series of regular triangular meshes (`mesh1` family) and then on square meshes (`mesh2` family), both taken from [74]. To ensure the correct orthogonality property (see Definition 2.3.8), the point $\boldsymbol{x}_K \in K$ is chosen as the circumcenter of K if K is a triangle, or the center of mass of K if K is a rectangle. As a result, for triangular meshes, the L^2 error, $\mathsf{err}_{\mathcal{D}}(\overline{u})$, is calculated using a skewed midpoint rule, where the circumcenter of each cell is considered instead of its center of mass. Let the relative H^2 error be denoted by

$$\mathsf{err}_{\mathcal{D}}(\Delta\overline{u}) := \frac{\|\Delta_{\mathcal{D}}\overline{u}_{\mathcal{D}} - \Delta\overline{u}\|}{\|\Delta\overline{u}\|}.$$

The H^1 and H^2 errors ($\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$ and $\mathsf{err}_{\mathcal{D}}(\Delta\overline{u})$) are computed using the usual midpoint rule. For comparsion with the gradient recovery method (see Table 2.1), the details of mesh size h, number of unknowns **nu** and the number of non-zero terms in the system square matrix **nnz** for the finite volume method are also provided in the following tables.

Example 1

In the first example, choose the right-hand side load function f such that the exact solution is given by $\overline{u}(x,y) = x^2y^2(1-x)^2(1-y)^2$. Tables 2.18 and 2.19 show the relative errors and order of convergence rates for the variable $\overline{u}_{\mathcal{D}}$ on triangular and square grids. As seen in the table, we

obtain linear (in H^1-like norm) and sub-linear convergence rates (in H^2-like norm) for triangular grids, and quadratic order of convergence for square grids. This behaviour has already been observed in [59]. With respect to L^2 norm, quadratic (or slightly better) order of convergence is obtained. These numerical order of convergence are better than the orders of convergences from the theoretical analysis, see Remark 2.4.13. This is somehow expected as, due to the difficulty of finding a proper interpolant for this very low-order method [59], the theoretical rates are much below than the actual rates.

Table 2.18: (FV) Convergence results, Example 1, triangular grids (`mesh1` family)

h	**nu**	**nnz**	$\mathrm{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\Delta\overline{u})$	Order
0.250000	56	392	0.137345	-	0.256342	-	0.162222	-
0.125000	224	1896	0.031705	2.1150	0.131915	0.9585	0.071457	1.1828
0.062500	896	8264	0.007400	2.0991	0.066136	0.9961	0.038596	0.8886
0.031250	3584	34440	0.001691	2.1297	0.033067	1.0000	0.022662	0.7682
0.015625	14336	140552	0.000352	2.2644	0.016528	1.0005	0.014158	0.6786
0.007813	57344	567816	0.000056	2.6449	0.008262	1.0004	0.009281	0.6092

Table 2.19: (FV) Convergence results, Example 1, square grids (`mesh2` family)

h	**nu**	**nnz**	$\mathrm{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\Delta\overline{u})$	Order
0.353553	16	56	0.328639	-	0.417244	-	0.260189	-
0.176777	64	472	0.081325	2.0147	0.107484	1.9568	0.062624	2.0548
0.088388	256	2552	0.020161	2.0121	0.026808	2.0034	0.015430	2.0210
0.044194	1024	11704	0.005028	2.0035	0.006694	2.0018	0.003842	2.0057
0.022097	4096	49976	0.001256	2.0009	0.001673	2.0005	0.000960	2.0015
0.011049	16384	206392	0.000314	2.0002	0.000418	2.0001	0.000240	2.0004

Example 2

In this example, the numerical experiment is performed for the exact solution given by $\overline{u}(x,y) = x^2y^2(1-x)^2(1-y)^2(\cos(2\pi x)+\sin(2\pi y))$. The errors in the energy norm, H^1 norm and the L^2 norm, together with their orders of convergence, are presented in Tables 2.20 and 2.21. The results are similar to those for Example 1.

Example 3

The numerical results obtained for $\overline{u}(x,y) = x^3y^3(1-x)^3(1-y)^3(\exp(x)\sin(2\pi x)+\cos(2\pi x))$ are shown in Tables 2.22 and 2.23 respectively. As in Examples 1 and 2, the theoretical rates of convergence are confirmed by these numerical outputs, except that on this test a real linear order of convergence is attained in the H^2-like norm.

Table 2.20: (FV) Convergence results, Example 2, triangular grids (`mesh1` family)

h	**nu**	**nnz**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\overline{u})$	Order
0.250000	56	392	0.418276	-	0.533799	-	0.274105	-
0.125000	224	1896	0.075761	2.4649	0.204870	1.3816	0.101375	1.4350
0.062500	896	8264	0.013663	2.4712	0.093729	1.1281	0.044254	1.1958
0.031250	3584	34440	0.003218	2.0862	0.046056	1.0251	0.021933	1.0127
0.015625	14336	140552	0.000784	2.0365	0.022932	1.0060	0.011500	0.9315
0.007813	57344	567816	0.000191	2.0414	0.011454	1.0015	0.006323	0.8630

Table 2.21: (FV) Convergence results, Example 2, square grids (`mesh2` family)

h	**nu**	**nnz**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\overline{u})$	Order
0.353553	16	56	1.333981	-	0.745194	-	0.773521	-
0.176777	64	472	0.223384	2.5781	0.135128	2.4633	0.175192	2.1425
0.088388	256	2552	0.050527	2.1444	0.030239	2.1599	0.042123	2.0563
0.044194	1024	11704	0.012331	2.0347	0.007339	2.0427	0.010416	2.0158
0.022097	4096	49976	0.003065	2.0086	0.001821	2.0109	0.002597	2.0041
0.011049	16384	206392	0.000765	2.0021	0.000454	2.0027	0.000649	2.0010

Table 2.22: (FV) Convergence results, Example 3, triangular grids (`mesh1` family)

h	**nu**	**nnz**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\overline{u})$	Order
0.250000	56	392	0.637895	-	0.825992	-	0.423933	-
0.125000	224	1896	0.050763	3.6515	0.220328	1.9065	0.096604	2.1337
0.062500	896	8264	0.013330	1.9291	0.097939	1.1697	0.045854	1.0750
0.031250	3584	34440	0.003160	2.0765	0.047945	1.0305	0.021417	1.0983
0.015625	14336	140552	0.000786	2.0084	0.023857	1.0070	0.010550	1.0215
0.007813	57344	567816	0.000196	2.0016	0.011914	1.0017	0.005257	1.0049

Table 2.23: (FV) Convergence results, Example 3, square grids (`mesh2` family)

h	**nu**	**nnz**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\overline{u})$	Order
0.353553	16	56	2.478402	-	1.405462	-	1.140625	-
0.176777	64	472	0.242959	3.3506	0.113945	3.6246	0.196693	2.5358
0.088388	256	2552	0.050784	2.2583	0.022495	2.3406	0.049149	2.0007
0.044194	1024	11704	0.012212	2.0561	0.005577	2.0120	0.012217	2.0083
0.022097	4096	49976	0.003025	2.0133	0.001396	1.9982	0.003049	2.0026
0.011049	16384	206392	0.000755	2.0033	0.000349	1.9993	0.000762	2.0007

Comparing Table 2.1 and the Tables for FV, we see that the GR method based on biorthogonal reconstruction has only few unknowns (number of internal vertices) but leads to a large stencil for each of them whereas the FV has more unknowns (number of cells) but produces a much sparser matrix. Looking for example at the finest GR mesh and the finest triangular FV mesh, we notice that the meshes have similar sizes h and the matrices have similar complexity **nnz**, but the FV accuracy in L^2- and H^2-like norms is much better than the GR method; this is expected since the FV method has a number of unknowns **nu** more than 3.5 times larger than that of GR. However, the super-convergence property of the gradient reconstruction gives a clear advantage to GR for the H^1-like norm. For a similar number of unknowns **nu** (which means a matrix that is much cheaper to solve for the FV method than the GR method, due to a reduced **nnz**), the FV method still has a clear advantage in the L^2 norm over the GR method, but similar accuracy in the H^2-like norm (compare the results for the 5th mesh in the `mesh1` family with the finest mesh used for the GR method); the GR method however still preserves a clear lead on the H^1-like norm error.

2.7.3 Numerical results for Modified FVM

In this section, three numerical experiments that justify the theoretical result in Proposition 2.5.8 for modified FVM are presented. We conduct the test on a series of regular triangular meshes (`mesh1` family) taken from [74] over the unit square $\Omega = (0,1)^2$. The orthogonality property is satisfied with the point $\boldsymbol{x}_K \in K$ chosen as the circumcenter of K. Let the relative errors in $L^2(\Omega)$, $H^1(\Omega)$ and $H^2(\Omega)$ norms be denoted by

$$\mathsf{err}_{\mathcal{D}^*}(\overline{u}) := \frac{\|\Pi_{\mathcal{D}^*} u_{\mathcal{D}^*} - \overline{u}\|}{\|\overline{u}\|}, \quad \mathsf{err}_{\mathcal{D}^*}(\nabla\overline{u}) := \frac{\|\nabla_{\mathcal{D}} u_{\mathcal{D}^*} - \nabla\overline{u}\|}{\|\nabla\overline{u}\|}, \quad \mathsf{err}_{\mathcal{D}^*}(\Delta\overline{u}) := \frac{\|\Delta_{\mathcal{D}} u_{\mathcal{D}^*} - \Delta\overline{u}\|}{\|\Delta\overline{u}\|},$$

where $u_{\mathcal{D}^*}$ is the solution to the HS (2.3.1) corresponding to the HD $\mathcal{D}^*$ given by Definition 2.5.7.

Example 1

In the first example, choose the solution to be $\overline{u}(x,y) = x^2y^2(1-x)^2(1-y)^2$. The error estimates and convergence rates in the energy, H^1 and H^2 norms are presented in Table 2.24. We obtain a quadratic (or slightly better) rate of convergence in L^2 norm, linear rate of convergence is H^1 norm and sub-linear rate of convergence in H^2 norm. Note that the numerical test provides better result compared to the theoretical result, see Proposition 2.5.8. The numerical results for modified FVM are similar to those for the FVM.

Table 2.24: (Modified FV) Convergence results, Example 1

h	$\mathsf{err}_{\mathcal{D}^*}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}^*}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}^*}(\Delta\overline{u})$	Order
0.250000	0.095132	-	0.236554	-	0.134417	-
0.125000	0.024787	1.9403	0.130595	0.8571	0.068112	0.9807
0.062500	0.005981	2.0511	0.066013	0.9843	0.038204	0.8342
0.031250	0.001353	2.1442	0.033053	0.9979	0.022618	0.7562
0.015625	0.000267	2.3415	0.016526	1.0000	0.014154	0.6763
0.007813	0.000035	2.9347	0.008262	1.0003	0.009281	0.6089

Example 2

In this case, we consider $\overline{u}(x,y) = x^2y^2(1-x)^2(1-y)^2(\cos(2\pi x)+\sin(2\pi y))$. The numerical results, presented in Table 2.25, are similar to those obtained for Example 1.

Table 2.25: (Modified FV) Convergence results, Example 2

h	$\mathsf{err}_{\mathcal{D}^*}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}^*}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}^*}(\Delta\overline{u})$	Order
0.250000	0.230644	-	0.458624	-	0.190768	-
0.125000	0.046952	2.2964	0.193505	1.2449	0.078850	1.2746
0.062500	0.009022	2.3797	0.092859	1.0593	0.041327	0.9320
0.031250	0.002089	2.1105	0.045960	1.0147	0.021572	0.9379
0.015625	0.000502	2.0562	0.022921	1.0037	0.011457	0.9130
0.007813	0.000120	2.0643	0.011453	1.0010	0.006318	0.8587

Example 3

The exact solution is chosen to be $\overline{u}(x,y) = x^3y^3(1-x)^3(1-y)^3(\exp(x)\sin(2\pi x)+\cos(2\pi x))$. The convergence results are presented in Table 2.26. In this example, an $\mathcal{O}(h)$ convergence rate is obtained in H^2 norm. Since there is no improvement of the rates from H^2 to H^1, as mentioned in Remark 2.6.6, we cannot expect an improved H^1-like estimate for FVM.

Table 2.26: (Modified FV) Convergence results, Example 3

h	$\mathsf{err}_{\mathcal{D}^*}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}^*}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}^*}(\Delta\overline{u})$	Order
0.250000	0.410550	-	0.704301	-	0.295782	-
0.125000	0.029103	3.8183	0.212960	1.7256	0.084328	1.8104
0.062500	0.008773	1.7301	0.096846	1.1368	0.041288	1.0303
0.031250	0.002041	2.1037	0.047833	1.0177	0.020896	0.9825
0.015625	0.000503	2.0203	0.023843	1.0044	0.010486	0.9947
0.007813	0.000125	2.0048	0.011913	1.0011	0.005249	0.9984

Remark 2.7.1. *For rectangular meshes, in order that the orthogonality property is satisfied,* $\boldsymbol{x}_K \in K$ *is chosen as the center of mass of* K*. From [54, Theorem 5.3], it follows that the difference between the source term of modified FVM and FVM is of* $\mathcal{O}(h^2)$*. Therefore similar rate of convergence is obtained for modified FVM, since we see an* $\mathcal{O}(h^2)$ *convergence rate in* L^2 *and* H^1 *norms for FVM, see Section 2.7.2.*

Chapter 3

The Hessian discretisation method for fourth order semi-linear elliptic equations

This chapter deals with the Hessian discretisation method for fourth order semi-linear elliptic equations with a trilinear nonlinearity in an abstract setting [1].

3.1 Introduction

The HDM for fourth order linear elliptic equations and some of its applications are discussed in Chapter 2. In this chapter, the HDM for an abstract setting of semi-linear fourth order elliptic problems with trilinear nonlinearity and clamped boundary conditions is proposed. This in particular applies to the stream function vorticity formulation of the incompressible 2D Navier–Stokes problem [19, 68] and the von Kármán equations [42]. A complete convergence analysis is carried out based on minor adjustments of the three properties associated with linear HD plus an additional compactness assumption on the HD. It is shown that conforming FEMs, Adini and Morley non-conforming FEMs, and methods based on gradient recovery (GR) operator are valid examples of HDM for this non-linear model.

Two different approaches are employed to study the convergence analysis: the first one is based on compactness techniques and the second one using error estimates. The first approach does not rely on any smoothness or structural assumption on the continuous solution. In this approach, the solution to the problem in the weak formulation is obtained as the limit of a sequence of solutions to the approximate problem; the existence of solution for the continuous model is therefore established as a consequence of this convergence analysis. On the contrary, the analysis via error estimates considers a regular solution to the PDE (in the sense that the linearised problem around this solution is well-posed with H^3 regularity), and provides orders of convergence. The two approaches are complementary and, to the best of our knowledge, only the second approach has been considered in literature, for von Kármán equations.

[1] The results of this chapter are communicated in *Jérôme Droniou, Neela Nataraj and Devika Shylaja. Hessian discretisation method for fourth order semi-linear elliptic equations, 2019.*

The contributions of this chapter are summarized as follows:

- *A unified framework* provided by HDM for fourth order semi-linear elliptic equations with a trilinear nonlinearity, in an abstract set-up that applies to several numerical methods.
- *Convergence analysis* by compactness techniques that employs only four properties, namely, the coercivity, consistency, limit-conformity and compactness.
- *Error estimates* under the assumption on the existence of a companion operator that maps the discrete space to the continuous space.
- *Applications* to the stream function vorticity formulation of 2D Navier–Stokes equation and the von Kármán equations using the examples of HDM, namely, conforming FEMs, Adini and Morley ncFEMs, and GR methods.
- *Numerical experiments* on the approximation of non-linear models using the GR method and Morley FEM.

The chapter is organised as follows. The abstract problem with its applications is presented in Section 3.2. Section 3.3 deals with the Hessian discretisation method for fourth order non-linear problems. The four properties that are needed for the convergence analysis of HDM are described in this section. Section 3.4 deals with examples of HDM. In Section 3.5, two different approaches for the analysis are discussed: convergence by compactness, that does not require any additional regularity on the solution, and error estimates, for smooth enough solutions. Results of numerical experiments for the GR method and Morley FEM are provided in Section 3.6.

3.2 Model problem

The abstract setting of weak formulation of semi-linear fourth order elliptic problems with a trilinear nonlinearity and clamped boundary conditions is presented in this section.
Given $k \geq 1$, the continuous abstract problem seeks $\Psi \in \boldsymbol{X} := H_0^2(\Omega)^k$ such that

$$\mathcal{A}(\mathcal{H}\Psi, \mathcal{H}\Phi) + \mathcal{B}(\mathcal{H}\Psi, \nabla\Psi, \nabla\Phi) = \mathcal{L}(\Phi) \quad \forall \Phi \in \boldsymbol{X}, \tag{3.2.1}$$

where $\mathcal{H}\Psi$ and $\nabla\Psi$ are to be understood component-wise, that is: for $\Psi = (\psi_1, \cdots, \psi_k)$, $\mathcal{H}\Psi = (\mathcal{H}\psi_1, \cdots, \mathcal{H}\psi_k)$ and $\nabla\Psi = (\nabla\psi_1, \cdots, \nabla\psi_k)$. Let the following assumptions hold:

(A1) $\mathcal{A}(\cdot,\cdot)$ is a continuous and coercive bilinear form on $L^2(\Omega;\mathbb{R}^{d\times d})^k \times L^2(\Omega;\mathbb{R}^{d\times d})^k$,

(A2) $\mathcal{B}(\cdot,\cdot,\cdot)$ is a continuous trilinear form on $L^2(\Omega;\mathbb{R}^{d\times d})^k \times L^4(\Omega;\mathbb{R}^d)^k \times L^4(\Omega;\mathbb{R}^d)^k$,

(A3) $\mathcal{B}(\Xi,\Theta,\Theta) = 0$ for all $\Xi \in L^2(\Omega;\mathbb{R}^{d\times d})^k$ and $\Theta \in L^4(\Omega;\mathbb{R}^d)^k$.

(A4) $\mathcal{L}(\cdot)$ is a continuous linear form on $L^2(\Omega)^k$.

3.2.1 Examples

We show here that the abstract formulation (3.2.1) covers the stream function vorticity formulation of the incompressible 2D Navier–Stokes problem, and von Kármán equations.

Navier–Stokes problem [19, 89]:

For $f \in L^2(\Omega)$ and viscosity $\nu > 0$, let u solve

$$\nu\Delta^2 u + \frac{\partial}{\partial x_1}\left((-\Delta u)\frac{\partial u}{\partial x_2}\right) - \frac{\partial}{\partial x_2}\left((-\Delta u)\frac{\partial u}{\partial x_1}\right) = f \text{ in } \Omega \tag{3.2.2a}$$

$$u = \frac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega. \tag{3.2.2b}$$

Here n denotes the unit outward normal to the boundary $\partial\Omega$ and the biharmonic operator Δ^2 is defined by $\Delta^2\phi = \phi_{xxxx} + \phi_{yyyy} + 2\phi_{xxyy}$. The weak formulation to (3.2.2) seeks $u \in H_0^2(\Omega)$ such that

$$\mathcal{A}(\mathcal{H}u, \mathcal{H}v) + \mathcal{B}(\mathcal{H}u, \nabla u, \nabla v) = \mathcal{L}(v) \;\; \forall v \in H_0^2(\Omega), \tag{3.2.3}$$

where for all ξ, $\chi \in L^2(\Omega; \mathbb{R}^{2\times 2})$ and ϕ, $\theta \in L^2(\Omega; \mathbb{R}^2)$,

$$\mathcal{A}(\xi, \chi) = \nu \int_\Omega \xi : \chi \, \mathrm{d}\boldsymbol{x}, \quad \mathcal{B}(\xi, \phi, \theta) = \int_\Omega \mathrm{tr}(\xi)\phi \cdot \mathrm{rot}_{\pi/2}(\theta) \, \mathrm{d}\boldsymbol{x}, \quad \mathcal{L}(v) = \int_\Omega f v \, \mathrm{d}\boldsymbol{x}.$$

Note that $\mathrm{tr}(\xi)$ means the trace of the matrix ξ and, for $\theta = (\theta_1, \theta_2)$, $\mathrm{rot}_{\pi/2}(\theta) = \big(-\theta_2, \theta_1\big)^t$. It is easy to check that $\mathcal{A}(\cdot,\cdot)$, $\mathcal{B}(\cdot,\cdot,\cdot)$ and $\mathcal{L}(\cdot)$ satisfy **(A1)-(A4)** with $k = 1$. The continuity of $\mathcal{B}(\cdot,\cdot,\cdot)$ follows using the generalised Hölder's inequality.

The von Kármán equations [42]:

Given $f \in L^2(\Omega)$, seek the vertical displacement u and the Airy stress function v such that

$$\Delta^2 u = [u, v] + f \text{ in } \Omega, \tag{3.2.4a}$$

$$\Delta^2 v = -\frac{1}{2}[u, u] \text{ in } \Omega, \tag{3.2.4b}$$

with clamped boundary conditions

$$u = \frac{\partial u}{\partial n} = v = \frac{\partial v}{\partial n} = 0 \text{ on } \partial\Omega. \tag{3.2.5}$$

The von Kármán bracket $[\cdot,\cdot]$ is defined by $[\xi, \chi] = \xi_{xx}\chi_{yy} + \xi_{yy}\chi_{xx} - 2\xi_{xy}\chi_{xy} = \mathrm{cof}(\mathcal{H}\xi) : \mathcal{H}\chi$, where $\mathrm{cof}(\mathcal{H}\xi)$ denotes the co-factor matrix of $\mathcal{H}\xi$. Then a weak formulation corresponding to (3.2.4) seeks $u, v \in H_0^2(\Omega)$ such that

$$a(u, \phi_1) + 2b(u, \phi_1, v) = (f, \phi_1) \quad \forall \phi_1 \in H_0^2(\Omega), \tag{3.2.6a}$$

$$2a(v, \phi_2) - 2b(u, u, \phi_2) = 0 \quad \forall \phi_2 \in H_0^2(\Omega), \tag{3.2.6b}$$

where for all $\xi, \chi \in H_0^2(\Omega)$,

$$a(\xi,\chi) := \int_\Omega \mathcal{H}\xi : \mathcal{H}\chi \,\mathrm{d}\boldsymbol{x},\ b(\xi,\chi,\phi) := \frac{1}{2}\int_\Omega \mathrm{cof}(\mathcal{H}\xi)\nabla\chi\cdot\nabla\phi \,\mathrm{d}\boldsymbol{x} = -\frac{1}{2}\int_\Omega [\xi,\chi]\phi \,\mathrm{d}\boldsymbol{x}.$$

Note that $b(\cdot,\cdot,\cdot)$ is derived using the divergence-free rows property [57] and is symmetric with respect to all variables. Summing together (3.2.6a) and (3.2.6b), we obtain an equivalent formulation in the vector form (3.2.1) (with $k=2$) that seeks $\Psi = (u,v) \in H_0^2(\Omega)^2$ such that

$$\mathcal{A}(\mathcal{H}\Psi,\mathcal{H}\Phi) + \mathcal{B}(\mathcal{H}\Psi,\nabla\Psi,\nabla\Phi) = \mathcal{L}(\Phi) \quad \forall \Phi \in H_0^2(\Omega)^2, \tag{3.2.7}$$

where for all $\Phi = (\phi_1,\phi_2)$, $\Lambda = (\lambda_1,\lambda_2)$, $\Gamma = (\gamma_1,\gamma_2)$, $\Theta = (\theta_1,\theta_2)$ and $\Xi = (\xi_1,\xi_2)$ with $\Lambda,\Gamma \in L^2(\Omega;\mathbb{R}^{2\times 2})^2$ and Ξ, $\Theta \in L^2(\Omega;\mathbb{R}^2)^2$,

$$\mathcal{A}(\Lambda,\Gamma) := \int_\Omega \lambda_1 : \gamma_1 \,\mathrm{d}\boldsymbol{x} + 2\int_\Omega \lambda_2 : \gamma_2 \,\mathrm{d}\boldsymbol{x}, \tag{3.2.8a}$$

$$\mathcal{B}(\Lambda,\Xi,\Theta) := \int_\Omega \mathrm{cof}(\lambda_1)\theta_1\cdot\xi_2 \,\mathrm{d}\boldsymbol{x} - \int_\Omega \mathrm{cof}(\lambda_1)\xi_1\cdot\theta_2 \,\mathrm{d}\boldsymbol{x} \text{ and} \tag{3.2.8b}$$

$$\mathcal{L}(\Phi) := (f,\phi_1). \tag{3.2.8c}$$

The assumptions **(A1)**-**(A4)** are easy to verify for this example.

Remark 3.2.1. *The more commonly used equivalent weak formulation of the von Kármán model [18, 91, 92]* (3.2.4) *seeks* $(u,v) \in H_0^2(\Omega)^2$ *such that*

$$a(u,\phi_1) + 2b(u,v,\phi_1) = (f,\phi_1) \quad \forall \phi_1 \in H_0^2(\Omega) \tag{3.2.9a}$$

$$a(v,\phi_2) - b(u,u,\phi_2) = 0 \quad \forall \phi_2 \in H_0^2(\Omega). \tag{3.2.9b}$$

An advantage of (3.2.6) *is that it ensures the proper cancellation in the trilinear term, in a purely algebraic way (corresponding to* **(A3)***) without further integration-by-parts. As a consequence, this cancellation, which is at the core of* a priori *estimates on the solution, directly transfers to the discrete level – on which integration-by-parts would not be possible for non-conforming methods. This formulation of the non-linear term is similar in spirit to what is usually done for finite element discretisations of the Navier–Stokes equations, see [108].*

3.3 The Hessian discretisation method

The HDM for linear problems is presented in Section 2.3. This section is devoted to the presentation of the HDM for fourth order non-linear elliptic equations, design of which is adapted from the HDM for linear problems (see Remark 3.3.5 below). A Hessian discretisation (HD) is based on a set of four elements, namely, a discrete space and three reconstructed operators. Once a HD is selected, the HDM consists in expressing the numerical scheme known as Hessian scheme (HS) by replacing the space and the continuous operators in the weak formulation (3.2.1) with these discrete components. The four quantities associated with HD to establish the convergence analysis is also discussed in this section.

Definition 3.3.1 (Hessian discretisation)**.** *A Hessian discretisation for fourth order non-linear elliptic equations with clamped boundary conditions is a quadruplet* $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}})$ *such that*

- $X_{\mathcal{D},0}$ *is a finite dimensional real vector space,*
- $\Pi_{\mathcal{D}} : X_{\mathcal{D},0} \to L^2(\Omega)$ *is a linear mapping that reconstructs functions from vectors in* $X_{\mathcal{D},0}$,
- $\nabla_{\mathcal{D}} : X_{\mathcal{D},0} \to L^4(\Omega)^d$ *is a linear mapping that reconstructs gradient from vectors in* $X_{\mathcal{D},0}$,
- $\mathcal{H}_{\mathcal{D}} : X_{\mathcal{D},0} \to L^2(\Omega;\mathbb{R}^{d\times d})$ *is a linear mapping that reconstructs a discrete version of Hessian from* $X_{\mathcal{D},0}$*. It must be chosen such that* $\|\cdot\|_{\mathcal{D}} =: \|\mathcal{H}_{\mathcal{D}}\cdot\|$ *is a norm on* $X_{\mathcal{D},0}$.

In order to approximate (3.2.1) by the Hessian discretisation method, consider a Hessian discretisation $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}})$ in the sense of Definition 3.3.1. The associated Hessian scheme for (3.2.1) seeks $\Psi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0} := X_{\mathcal{D},0}^k$ such that

$$\mathcal{A}(\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}, \nabla_{\mathcal{D}}\Psi_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) = \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}) \quad \forall \Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}, \tag{3.3.1}$$

where $\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}$, $\nabla_{\mathcal{D}}\Phi_{\mathcal{D}}$ and $\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}$ act component-wise in the sense that if $\Phi_{\mathcal{D}} = (\phi_{\mathcal{D},1}, \cdots, \phi_{\mathcal{D},k})$ and $F_{\mathcal{D}} \in \{\Pi_{\mathcal{D}}, \nabla_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}}\}$, then $F_{\mathcal{D}}\Phi_{\mathcal{D}} = (F_{\mathcal{D}}\phi_{\mathcal{D},1}, \cdots, F_{\mathcal{D}}\phi_{\mathcal{D},k})$. The convergence analysis of a Hessian scheme is based on four quantities and associated notions, measuring the stability and accuracy of the chosen Hessian discretisation (see Theorems 3.5.1 and 3.5.12).
The first quantity is a constant, $C_{\mathcal{D}}$, that controls the norm of $\Pi_{\mathcal{D}}$ and $\nabla_{\mathcal{D}}$. It is defined by

$$C_{\mathcal{D}} = \max_{w\in X_{\mathcal{D},0}\setminus\{0\}} \left(\frac{\|\Pi_{\mathcal{D}} w\|}{\|\mathcal{H}_{\mathcal{D}} w\|}, \frac{\|\nabla_{\mathcal{D}} w\|_{L^4}}{\|\mathcal{H}_{\mathcal{D}} w\|} \right). \tag{3.3.2}$$

Definition 3.3.2 (Coercivity)**.** *A sequence* $(\mathcal{D}_m)_{m\in\mathbb{N}}$ *of Hessian discretisations in the sense of Definition 3.3.1 is* coercive *if there exists* $C_P \in \mathbb{R}^+$ *such that* $C_{\mathcal{D}_m} \le C_P$ *for all* $m \in \mathbb{N}$.

The second quantity is the interpolation error $S_{\mathcal{D}}$ defined by: for all $\varphi \in H_0^2(\Omega)$,

$$S_{\mathcal{D}}(\varphi) = \min_{w\in X_{\mathcal{D},0}} \Big(\|\Pi_{\mathcal{D}} w - \varphi\| + \|\nabla_{\mathcal{D}} w - \nabla\varphi\|_{L^4} + \|\mathcal{H}_{\mathcal{D}} w - \mathcal{H}\varphi\| \Big). \tag{3.3.3}$$

Definition 3.3.3 (Consistency)**.** *A sequence* $(\mathcal{D}_m)_{m\in\mathbb{N}}$ *of Hessian discretisations in the sense of Definition 3.3.1 is* consistent*, if*

$$\forall \varphi \in H_0^2(\Omega), \lim_{m\to\infty} S_{\mathcal{D}_m}(\varphi) = 0.$$

To define the limit-conformity measure for the HS, introduce $H(\Omega) = \{\xi \in L^2(\Omega)^{d\times d} ; \mathcal{H} : \xi \in L^2(\Omega)\}$ and $H_{\mathrm{div}}(\Omega) = \{\phi \in L^2(\Omega)^d : \mathrm{div}\phi \in L^2(\Omega)\}$. For all $\xi \in H(\Omega)$ and $\phi \in H_{\mathrm{div}}(\Omega)$, set

$$W_{\mathcal{D}}(\xi) = \max_{w\in X_{\mathcal{D},0}\setminus\{0\}} \frac{1}{\|\mathcal{H}_{\mathcal{D}} w\|} \Big| \int_\Omega \Big((\mathcal{H} : \xi)\Pi_{\mathcal{D}} w - \xi : \mathcal{H}_{\mathcal{D}} w \Big) \,\mathrm{d}\boldsymbol{x} \Big|, \tag{3.3.4}$$

$$\widehat{W}_{\mathcal{D}}(\phi) = \max_{w \in X_{\mathcal{D},0}\setminus\{0\}} \frac{1}{\|\mathcal{H}_{\mathcal{D}} w\|} \left| \int_\Omega \left(\nabla_{\mathcal{D}} w \cdot \phi + \Pi_{\mathcal{D}} w \, \mathrm{div}\phi \right) \mathrm{d}\boldsymbol{x} \right|. \tag{3.3.5}$$

Here $W_{\mathcal{D}}$ measures the defect of a double integration by parts (Chapter 2) and is the limit-conformity measure between the reconstructed Hessian and reconstructed function. $\widehat{W}_{\mathcal{D}}$ measures the defect of a Stokes formula between the reconstructed gradient and function.

Definition 3.3.4 (Limit-conformity)**.** *A sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ of Hessian discretisations in the sense of Definition 3.3.1 is* limit-conforming *if*

$$\forall\, \xi \in H(\Omega), \forall \phi \in H_{\mathrm{div}}(\Omega), \quad \lim_{m\to\infty} \left(W_{\mathcal{D}_m}(\xi) + \widehat{W}_{\mathcal{D}_m}(\phi) \right) = 0.$$

In the sequel, we also need

$$\widetilde{W}_{\mathcal{D}}(\xi) = \max_{w \in X_{\mathcal{D},0}\setminus\{0\}} \frac{1}{\|\mathcal{H}_{\mathcal{D}} w\|} \left| \int_\Omega \left(\xi : \mathcal{H}_{\mathcal{D}} w + (\mathrm{div}\xi) \cdot \nabla_{\mathcal{D}} w \right) \mathrm{d}\boldsymbol{x} \right|, \tag{3.3.6}$$

for all $\xi \in H_{\mathrm{div}}(\Omega)^d$, where $H_{\mathrm{div}}(\Omega)^d = \{\phi \in L^2(\Omega)^{d\times d} : \mathrm{div}\phi \in L^2(\Omega)^d\}$. Note that $\widetilde{W}_{\mathcal{D}}$ measures the error in the discrete Stokes formula between the reconstructed Hessian and the reconstructed gradient. It is easy to show that, for all $\xi \in H_{\mathrm{div}}(\Omega)^d$ with $\mathcal{H} : \xi \in L^2(\Omega)$, it holds $\widetilde{W}_{\mathcal{D}}(\xi) \le W_{\mathcal{D}}(\xi) + \widehat{W}_{\mathcal{D}}(\mathrm{div}\xi)$ by noticing

$$\begin{aligned}\int_\Omega \left(\xi : \mathcal{H}_{\mathcal{D}} w + (\mathrm{div}\xi) \cdot \nabla_{\mathcal{D}} w \right) \mathrm{d}\boldsymbol{x} &= \int_\Omega \left(\xi : \mathcal{H}_{\mathcal{D}} w - (\mathcal{H} : \xi) \Pi_{\mathcal{D}} w \right) \mathrm{d}\boldsymbol{x} \\ &\quad + \int_\Omega \left((\mathcal{H} : \xi) \Pi_{\mathcal{D}} w + (\mathrm{div}\xi) \cdot \nabla_{\mathcal{D}} w \right) \mathrm{d}\boldsymbol{x}\end{aligned}$$

and $\mathrm{div}(\mathrm{div}\xi) = \mathcal{H} : \xi$.

Remark 3.3.5 (Comparison with the linear setting)**.** *For linear equations, $C_{\mathcal{D}}$ ((2.4.1)) and $S_{\mathcal{D}}$ ((2.4.2)) are defined using the L^2-norms of the gradients. Dealing with the trilinear non-linearity requires higher integrability properties, and thus the use of the L^4-norms of gradients in the definitions of $C_{\mathcal{D}}$ ((3.3.2)) and $S_{\mathcal{D}}$ ((3.3.3)).*
Another difference in comparison to the linear setting is the introduction of $\widehat{W}_{\mathcal{D}}$. The limit-conformity defect $W_{\mathcal{D}_m}$ is sufficient to analyze the convergence of the HDM for linear models. Here, however, the non-linear model (3.2.7) *involves the gradient, and accounting for $\widehat{W}_{\mathcal{D}}$ in the definition of limit-conformity is necessary to identify the limit of the reconstructed gradients during the convergence analysis.*

Definition 3.3.6 (Compactness)**.** *A sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ of Hessian discretisations in the sense of Definition 3.3.1 is* compact *if for any sequence $u_m \in X_{\mathcal{D}_m,0}$ such that $(\|u_m\|_{\mathcal{D}_m})_{m\in\mathbb{N}}$ is bounded, $(\Pi_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ is relatively compact in $L^2(\Omega)$, and $(\nabla_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ is relatively compact in $L^4(\Omega)^d$.*

Remark 3.3.7. *In most cases, by the continuous Sobolev embedding (which is often also valid at the discrete level [48, Appendix B]) we actually expect $(\Pi_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ and $(\nabla_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ to be compact in L^p for all $p < 2^*$, where 2^* is a Sobolev exponent associated with 2.*

3.4 Examples of Hessian discretisation method

This section discusses some known schemes that fit into the Hessian discretisation method for fourth order semi-linear equations. For a detailed discussion on the methods (FEMs and GR methods), see Section 2.3 in Chapter 2. Recall the polytopal mesh defined in Chapter 1 (Definition 1.4.1).

3.4.1 Conforming FEMs

As in Chapter 2, for conforming FEMs, a Hessian discretisation is defined by $X_{\mathcal{D},0} =: V_h$, a finite dimensional subspace of the space $H_0^2(\Omega)$ and, for $v_\mathcal{D} \in X_{\mathcal{D},0}$, $\Pi_\mathcal{D} v_\mathcal{D} = v_\mathcal{D}$, $\nabla_\mathcal{D} v_\mathcal{D} = \nabla v_\mathcal{D}$ and $\mathcal{H}_\mathcal{D} v_\mathcal{D} = \mathcal{H} v_\mathcal{D}$. The estimates on $C_\mathcal{D}$, $S_\mathcal{D}$, $W_\mathcal{D}$, $\widehat{W}_\mathcal{D}$ and the compactness property easily follow:

- $C_\mathcal{D}$ is bounded by the maximum of the constants of the continuous Poincaré inequality in $H_0^2(\Omega)$ and the continuous Sobolev imbedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$.
- Standard interpolation properties (see, e.g., [41]) and the continuous Sobolev imbedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$ yield an $\mathcal{O}(h)$ estimate on $S_\mathcal{D}(\varphi)$, provided that $\varphi \in H^3(\Omega) \cap H_0^2(\Omega)$; the proof that $\lim_{h\to 0} S_\mathcal{D}(\varphi) \to 0$ for all $\varphi \in H_0^2(\Omega)$ can be done by a density argument as in [48, Lemma 2.16] .
- Integration by parts in $H_0^2(\Omega)$ shows that $W_\mathcal{D}(\xi) = 0$ for all $\xi \in H(\Omega)$ and $\widehat{W}_\mathcal{D}(\phi) = 0$ for all $\phi \in H_{\rm div}(\Omega)$.
- The compactness of $(\mathcal{D}_m)_{m\in\mathbb{N}}$ follows from the Rellich and Sobolev imbedding theorems.

Classical C^1 elements that are used for the approximation the solution of fourth order elliptic problems are the Argyris triangle and Bogner-Fox-Schmit rectangle, see Chapter 2 for more details.

3.4.2 Non-conforming FEMs

We show here that two non-conforming finite element methods in dimension $d = 2$, namely the Morley FEM and the Adini FEM, fit into the framework of Hessian discretisation method. It has been proved in Chapter 2 that the Adini rectangle and the Morley triangle satisfies the properties (3.3.2)-(3.3.4) of a Hessian discretisation method for fourth order linear elliptic problems (that is, with L^2 norms for the gradient terms, see Remark 3.3.5). In this section, the four measures ((3.3.2)–(3.3.5) and Definition 3.3.6) associated with the HD using the Morley and the Adini FEMs for non-linear problems are estimated.

The auxiliary results discussed below are useful to prove the convergence of the Adini and Morley HDM for non-linear equations. Recall $\|\cdot\|_{dG,\mathcal{M}}$ given by (2.4.13): For all $w \in H^1(\mathcal{M})$,

$$\|w\|_{dG,\mathcal{M}}^2 := \|\nabla_\mathcal{M} w\|^2 + \sum_{\sigma\in\mathcal{F}} \frac{1}{h_\sigma} \|[\![w]\!]\|_{L^2(\sigma)}^2.$$

Lemma 3.4.1. *[45, Theorems 5.3, 5.6] It holds*

(i) [Discrete Sobolev embedding] For all $v_h \in \mathbb{P}_\ell(\mathcal{M})$, $\|v_h\|_{L^4} \leq C\|v_h\|_{dG,\mathcal{M}}$.

(ii) [Discrete Rellich theorem] Let $(\mathcal{M}_{h_m})_{m\in\mathbb{N}}$ *be sequence of regular triangular or rectangular meshes, whose diameter* h_m *tend to* 0 *as* $m \to \infty$. *For all* $m \in \mathbb{N}$, *let* $v_m \in \mathbb{P}_\ell(\mathcal{M}_{h_m})$. *If* $(\|v_m\|_{dG,\mathcal{M}_{h_m}})_{m\in\mathbb{N}}$ *is bounded, then, for all* $1 \leq q < 2^*$ *(where* 2^* *is a Sobolev exponent of* 2*), the sequence* $(v_m)_{m\in\mathbb{N}}$ *is relatively compact in* $L^q(\Omega)$.

The next theorem provides estimates on the quantities given by (3.3.2)–(3.3.5) and shows that, along refined meshes, the HD corresponding to the Morely and Adini element satisfy the coercivity, consistency, limit-conformity and compactness properties. These properties are essential to apply Theorems 3.5.1 and 3.5.12.

Theorem 3.4.2. *Let* $\mathcal{D}$ *be a Hessian discretisation for the Morley (resp. Adini) ncFEM in the sense of Definition 2.3.4 (resp. Definition 2.3.3). Then, there exists a constant C, not depending on* $\mathcal{M}$, *such that*

(i) $C_\mathcal{D} \leq C$,

(ii) $\forall\, \varphi \in H^3(\Omega) \cap H^2_0(\Omega)$, $S_\mathcal{D}(\varphi) \leq Ch\|\varphi\|_{H^3(\Omega)}$,

(iii) $\forall\, \xi \in H^2(\Omega)^{2\times 2}, \forall \phi \in H^1(\Omega)^2$,

$$W_\mathcal{D}(\xi) + \widehat{W}_\mathcal{D}(\phi) \leq Ch\big(\|\xi\|_{H^2(\Omega)^{2\times 2}} + \|\phi\|_{H^1(\Omega)}\big),$$

(iv) For a sequence of meshes $(\mathcal{M}_{h_m})_{m\in\mathbb{N}}$ *with* $h_m \to 0$, *denoting the HD constructed on* $\mathcal{M}_{h_m}$ *as above by* $\mathcal{D}_m$, *the sequence* $(\mathcal{D}_m)_{m\in\mathbb{N}}$ *is compact.*

Proof. (I)THE MORLEY ELEMENT:
(i) Coercivity: Let $v_\mathcal{D} \in X_{\mathcal{D},0}$. Since $[\![\nabla_\mathcal{D} v_\mathcal{D}]\!] = 0$ at the edge midpoints, a use of Lemma 3.4.1(i) and (A.1.4) given by Lemma A.1.3 leads to

$$\|\nabla_\mathcal{D} v_\mathcal{D}\|_{L^4} \leq C\|\nabla_\mathcal{D} v_\mathcal{D}\|_{dG,\mathcal{M}} \leq C\|\mathcal{H}_\mathcal{D} v_\mathcal{D}\|. \tag{3.4.1}$$

The estimate (3.4.1) and Theorem 2.4.8 concludes the proof of the estimate on $C_\mathcal{D}$.
(ii) Consistency: Let $\varphi \in H^3(\Omega) \cap H^2_0(\Omega)$. By [41], the standard interpolant I_h satisfies

$$\begin{aligned} &\|I_h\varphi - \varphi\| \lesssim h^3\|\varphi\|_{H^3(\Omega)},\ \|\nabla_\mathcal{M} I_h\varphi - \nabla\varphi\|_{L^4} \lesssim h^{3/2}\|\varphi\|_{H^3(\Omega)} \\ &\text{and } \|\mathcal{H}_\mathcal{M} I_h\varphi - \mathcal{H}\varphi\| \lesssim h\|\varphi\|_{H^3(\Omega)}. \end{aligned} \tag{3.4.2}$$

Hence, $w \in X_{\mathcal{D},0}$ corresponding to the degrees of freedom of $I_h\varphi$ in the definition (3.3.3) of $S_\mathcal{D}(\varphi)$ yields the result.
(iii) Limit-conformity: Let $\phi \in H_{\mathrm{div}}(\Omega)$. A use of integration by parts leads to

$$\left|\int_\Omega \Big(\nabla_\mathcal{D} v_\mathcal{D} \cdot \phi + \Pi_\mathcal{D} v_\mathcal{D} \mathrm{div}\phi\Big)\,\mathrm{d}\boldsymbol{x}\right| = \left|\sum_{\sigma\in\mathcal{F}} \int_\sigma (\phi \cdot n_\sigma)[\![\Pi_\mathcal{D} v_\mathcal{D}]\!]\,\mathrm{d}s(\boldsymbol{x})\right|. \tag{3.4.3}$$

Proceed the same steps as in the proof of limit-conformity in Theorem 2.4.8 (with $\mathrm{div}(A\xi)$ replaced by ϕ in (2.4.24)-(2.4.26)) to obtain

$$\left|\int_\Omega \left(\nabla_{\mathcal{D}} v_{\mathcal{D}} \cdot \phi + \Pi_{\mathcal{D}} v_{\mathcal{D}} \mathrm{div}\phi\right) \mathrm{d}\boldsymbol{x}\right| \leq C(h\|\phi\| + h^2\|\nabla\phi\|)\|\mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\|.$$

Therefore, the above estimate together with Theorem 2.4.8 leads to the required estimate on $\widehat{W}_{\mathcal{D}}$ and $W_{\mathcal{D}}$.

(iv) Compactness: Let a sequence $u_m \in X_{\mathcal{D}_m,0}$ be such that $(\|u_m\|_{\mathcal{D}_m})_{m\in\mathbb{N}}$ is bounded. Since $[\![\Pi_{\mathcal{D}_m} u_m]\!] = 0$ at the edge vertices, a use of (A.1.4) given by Lemma A.1.3 and (3.4.1) yields

$$\|\Pi_{\mathcal{D}_m} u_m\|_{dG,\mathcal{M}_m} \leq C\|\nabla_{\mathcal{D}_m} u_m\| \leq C\|\nabla_{\mathcal{D}_m} u_m\|_{L^4} \leq C\|\mathcal{H}_{\mathcal{D}_m} u_m\|.$$

Since $[\![\nabla_{\mathcal{D}_m} u_m]\!] = 0$ at the edge midpoints, choose $w = \nabla_{\mathcal{D}_m} u_m$ in (A.1.4) given by Lemma A.1.3 to obtain

$$\|\nabla_{\mathcal{D}_m} u_m\|_{dG,\mathcal{M}_m} \leq C\|\mathcal{H}_{\mathcal{D}_m} u_m\|.$$

Use the fact that $(\|u_m\|_{\mathcal{D}_m})_{m\in\mathbb{N}}$ is bounded to deduce $(\Pi_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ and $(\nabla_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ are bounded in the $\|\cdot\|_{dG,\mathcal{M}_m}$ norm. Lemma 3.4.1(ii) then gives the relatively compactness of $(\Pi_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ in $L^2(\Omega)$, and of $(\nabla_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ in $L^4(\Omega)^d$.

(II)THE ADINI ELEMENT:

(i) Coercivity: Since $\nabla_{\mathcal{D}} v_{\mathcal{D}}$ is continuous at the vertices of elements in $\mathcal{M}$ and $\nabla_{\mathcal{D}} v_{\mathcal{D}}$ vanish at vertices along $\partial\Omega$, $[\![\nabla_{\mathcal{D}} v_{\mathcal{D}}]\!] = 0$ at the vertices. Therefore, a use of Lemma 3.4.1(i) and (A.1.4) given by Lemma A.1.3 leads to

$$\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\|_{L^4} \leq C\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\|_{dG,\mathcal{M}} \leq C\|\mathcal{H}_{\mathcal{D}} v_{\mathcal{D}}\|.$$

Use the above estimate and Theorem 2.4.6 to conclude that $C_{\mathcal{D}} \leq C$.

(ii) Consistency: The standard interpolant satisfies (3.4.2) and hence yields the desired estimate on $S_{\mathcal{D}}$.

(iii) Limit-conformity: Apply integrations by parts in each cell to obtain

$$\left|\int_\Omega \left(\nabla_{\mathcal{D}} v_{\mathcal{D}} \cdot \phi + \Pi_{\mathcal{D}} v_{\mathcal{D}} \mathrm{div}\phi\right) \mathrm{d}\boldsymbol{x}\right| = \left|\sum_{\sigma\in\mathcal{F}} \int_\sigma (\phi \cdot n_\sigma) [\![\Pi_{\mathcal{D}} v_{\mathcal{D}}]\!] \,\mathrm{d}s(\boldsymbol{x})\right|.$$

Since $\Pi_{\mathcal{D}} v_{\mathcal{D}} \in H_0^1(\Omega) \cap C(\overline{\Omega})$, $[\![\Pi_{\mathcal{D}} v_{\mathcal{D}}]\!] = 0$, which implies $\widehat{W}_{\mathcal{D}}(\phi) = 0$. This and Theorem 2.4.6 yields an estimate on $\widehat{W}_{\mathcal{D}}$ and $W_{\mathcal{D}}$.

(iv) Compactness: The proof follows as for the Morley element using the fact that $[\![\Pi_{\mathcal{D}} v_{\mathcal{D}}]\!] = 0$ and $[\![\nabla_{\mathcal{D}} v_{\mathcal{D}}]\!] = 0$ at the vertices. □

3.4.3 Method based on Gradient Recovery Operators

Let $(V_h, Q_h, I_h, \mathfrak{S}_h)$ be a quadruplet of a finite element space $V_h \subset H_0^1(\Omega)$, a projector $Q_h : L^2(\Omega) \to V_h$, an interpolant $I_h : H_0^2(\Omega) \to V_h$ and a stabilisation function $\mathfrak{S}_h \in L^\infty(\Omega)^2$.

The next theorem gives an estimate on the accuracy measures associated with an HD $\mathcal{D}$ using gradient recovery.

Theorem 3.4.3 (Estimates for Hessian discretisations based on gradient recovery)**.** *Let $\mathcal{D}$ be a Hessian discretisation in the sense of Definition 2.3.5 and $(V_h, I_h, Q_h, \mathfrak{S}_h)$ satisfying* **(P0)**–**(P5)**. *Then,*

(i) $C_{\mathcal{D}} \leq C$,

(ii) $\forall\, \varphi \in W$, $S_{\mathcal{D}}(\varphi) \leq Ch\|\varphi\|_W$,

(iii) $\forall\, \xi \in H^2(\Omega)^{d\times d}$, $\forall \phi \in H_{\mathrm{div}}(\Omega)$, $\quad W_{\mathcal{D}}(\xi) \leq Ch\|\xi\|_{H^2(\Omega)^{d\times d}}$, $\widehat{W}_{\mathcal{D}}(\phi) = 0$,

(iv) If $(\mathcal{M}_m)_{m\in\mathbb{N}}$ is a sequence of meshes and $\mathcal{D}_m$ is an gradient recovery HD based on $\mathcal{M}_m$ for discrete elements satisfying **(P0)**–**(P5)** *uniformly with respect to m, then $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is compact.*

Proof. From Theorem 2.4.10,

$$\sqrt{2}\|\mathcal{H}_{\mathcal{D}} v\| \geq \|\nabla(Q_h \nabla v)\| + \|Q_h \nabla v - \nabla v\|. \tag{3.4.4}$$

(i) COERCIVITY: For $v \in X_{\mathcal{D},0}$, the Sobolev embedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$ and (3.4.4) yield

$$\|\nabla_{\mathcal{D}} v\|_{L^4} = \|Q_h \nabla v\|_{L^4} \leq \|\nabla(Q_h \nabla v)\| \leq \sqrt{2}\|\mathcal{H}_{\mathcal{D}} v\|.$$

The above estimate along with Theorem 2.4.10 show that $C_{\mathcal{D}} \leq C$.

(ii) CONSISTENCY: Let $\varphi \in W \subset H^3(\Omega) \cap H_0^2(\Omega)$ and choose $v = I_h \varphi \in X_{\mathcal{D},0}$. A use of Sobolev embedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$ and Theorem 2.4.10 (see (2.4.34)) leads to

$$\|\nabla_{\mathcal{D}} v - \nabla\varphi\|_{L^4} \leq \|\nabla(Q_h \nabla v) - \nabla\nabla\varphi\| \leq Ch\|\varphi\|_W. \tag{3.4.5}$$

Thus, the estimate on $S_{\mathcal{D}}(\varphi)$ follows from (3.4.5) and Theorem 2.4.10.

(iii) LIMIT-CONFORMITY: For $\xi \in H^2(\Omega)^{d\times d}$, Theorem 2.4.10 yields $W_{\mathcal{D}}(\xi) \leq Ch\|\xi\|_{H^2(\Omega)^{d\times d}}$. Let $\phi \in H_{\mathrm{div}}(\Omega)$. The fact that $\widehat{W}_{\mathcal{D}} \equiv 0$ follows from an integration by parts, valid since $\Pi_{\mathcal{D}} v_{\mathcal{D}} \in H_0^1(\Omega)$ for all $v \in X_{\mathcal{D},0}$.

(iv) COMPACTNESS: Let a sequence $u_m \in X_{\mathcal{D}_m,0}$ be such that $(\|u_m\|_{\mathcal{D}_m})_{m\in\mathbb{N}}$ is bounded. Since $\Pi_{\mathcal{D}_m} u_m \in H_0^1(\Omega)$, a use of triangle inequality, the Poincaré inequality and (3.4.4) leads to

$$\begin{aligned}\|\nabla(\Pi_{\mathcal{D}_m} u_m)\| = \|\nabla u_m\| &\leq \|Q_{h_m}\nabla u_m\| + \|Q_{h_m}\nabla u_m - \nabla u_m\| \\ &\leq C\|\nabla Q_{h_m}\nabla u_m\| + \|Q_{h_m}\nabla u_m - \nabla u_m\| \leq C\|\mathcal{H}_{\mathcal{D}_m} u_m\|.\end{aligned}$$

Since $(\|u_m\|_{\mathcal{D}_m})_{m\in\mathbb{N}}$ is bounded, it follows that $(\nabla(\Pi_{\mathcal{D}_m} u_m))_{m\in\mathbb{N}}$ is bounded in $L^2(\Omega)^d$ and hence the standard Rellich theorem shows that $(\Pi_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ is relatively compact in $L^2(\Omega)$. Note that $\nabla_{\mathcal{D}_m} u_m = Q_{h_m}\nabla u_m \in H_0^1(\Omega)$. From (3.4.4), $\|\nabla Q_{h_m}\nabla u_m\| \leq C\|\mathcal{H}_{\mathcal{D}_m} u_m\|$. Thus, a use of the Rellich and Sobolev imbedding theorems yields the required compactness property of $(\nabla_{\mathcal{D}_m} u_m)_{m\in\mathbb{N}}$ in $L^4(\Omega)^d$. □

3.5 Convergence analysis

In this section, the main results of this chapter that uses two different approaches for convergence analysis of the Hessian discretisation method are presented. The first one (Theorem 3.5.1) relies on compactness arguments whereas the second one (Theorem 3.5.12) is based on error estimates.

3.5.1 Convergence by compactness

The convergence of the Hessian scheme is established in this section, provided that the underlying sequences of HDs satisfy the properties in Definitions 3.3.2–3.3.6. This convergence is proved without any extra-regularity assumption on the exact solution, or the assumption that the linearized problem around this solution is well-posed.

Theorem 3.5.1 (Convergence and existence of solution)**.** *Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of Hessian discretisations in the sense of Definition 3.3.1; that it is coercive, consistent, limit-conforming and compact. Then, for any $m \in \mathbb{N}$, there exists at least one solution $\Psi_{\mathcal{D}_m}$ to* (3.3.1)*, with $\mathcal{D} = \mathcal{D}_m$. Moreover, as $m \to \infty$, there exist a subsequence of $(\mathcal{D}_m)_{m\in\mathbb{N}}$ (denoted using the same notation $(\mathcal{D}_m)_{m\in\mathbb{N}}$), and a solution Ψ of the abstract problem* (3.2.1) *such that $\Pi_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} \to \Psi$ in $L^2(\Omega)^k$, $\nabla_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} \to \nabla\Psi$ in $L^4(\Omega;\mathbb{R}^d)^k$ and $\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} \to \mathcal{H}\Psi$ in $L^2(\Omega;\mathbb{R}^{d\times d})^k$.*

The following lemma helps to establish the result in Theorem 3.5.1.

Lemma 3.5.2 (Regularity of the limit)**.** *Let $(\mathcal{D}_m)_{m\in\mathbb{N}}$ be a sequence of Hessian discretisations in the sense of Definition 3.3.1 that is coercive and limit-conforming in the sense of Definitions 3.3.2 and 3.3.4. Let $u_m \in X_{\mathcal{D}_m,0}$ be such that $\|u_m\|_{\mathcal{D}_m}$ remains bounded. Then there exists a subsequence of $(\mathcal{D}_m, u_m)_{m\in\mathbb{N}}$ (denoted using the same notation) and $u \in H_0^2(\Omega)$ such that $\Pi_{\mathcal{D}_m}u_m$ converges weakly to u in $L^2(\Omega)$, $\nabla_{\mathcal{D}_m}u_m$ converges weakly to ∇u in $L^4(\Omega)^d$ and $\mathcal{H}_{\mathcal{D}_m}u_m$ converges weakly to $\mathcal{H}u$ in $L^2(\Omega)^{d\times d}$.*

Proof. By coercivity of $(\mathcal{D}_m)_{m\in\mathbb{N}}$, the bound on $\|u_m\|_{\mathcal{D}_m}$ shows that $(\Pi_{\mathcal{D}_m}u_m)_m$ and $(\nabla_{\mathcal{D}_m}u_m)_m$ are bounded in $L^2(\Omega)$ and $L^4(\Omega)^d$, respectively. Therefore, there exists a subsequence of $(\mathcal{D}_m, u_m)_{m\in\mathbb{N}}$ and $u \in L^2(\Omega), v \in L^4(\Omega)^d$ and $w \in L^2(\Omega)^{d\times d}$ such that $\Pi_{\mathcal{D}_m}u_m$ converges weakly in $L^2(\Omega)$ to u, $\nabla_{\mathcal{D}_m}u_m$ converges weakly in $L^4(\Omega)^d$ to v, and $\mathcal{H}_{\mathcal{D}_m}u_m$ converges weakly in $L^2(\Omega)^{d\times d}$ to w. It remains to prove that $v = \nabla u$, $w = \mathcal{H}u$ and $u \in H_0^2(\Omega)$. We extend $\Pi_{\mathcal{D}_m}u_m, u, \nabla_{\mathcal{D}_m}u_m, v, \mathcal{H}_{\mathcal{D}_m}u_m$ and w by 0 outside Ω, and the same convergence results hold, respectively, in $L^2(\mathbb{R}^d)$, $L^4(\mathbb{R}^d)^d$ and $L^2(\mathbb{R}^d)^{d\times d}$. Using the limit-conformity of $(\mathcal{D}_m)_{m\in\mathbb{N}}$ and the bound on $\|u_m\|_{\mathcal{D}_m}$, passing to the limit in (3.3.4)-(3.3.5) gives

$$\forall\, \xi \in H(\mathbb{R}^d),\ \int_{\mathbb{R}^d} \big((\mathcal{H} : \xi)u - \xi : w\big)\, \mathrm{d}\boldsymbol{x} = 0 \tag{3.5.1}$$

$$\text{and } \forall\, \phi \in H_{\mathrm{div}}(\mathbb{R}^d),\ \int_{\mathbb{R}^d} \big(v\cdot\phi + u\mathrm{div}\phi\big)\, \mathrm{d}\boldsymbol{x} = 0. \tag{3.5.2}$$

For $\phi \in C_c^\infty(\mathbb{R}^d)^d$ and $\xi \in C_c^\infty(\mathbb{R}^d;\mathbb{R}^{d\times d})$, (3.5.1) and (3.5.2) show that $w = \mathcal{H}u$ and $v = \nabla u$, in the sense of distributions on $\mathbb{R}^d$. This implies $u \in H^2(\mathbb{R}^d)$ and, since $u = 0$ outside the domain Ω, that $u \in H_0^2(\Omega)$. □

We now prove Theorem 3.5.1.

Proof of Theorem 3.5.1. The proof is divided into four steps.
Step 1: existence of a solution to the scheme
For any Hessian discretisation $\mathcal{D}$, let $\overline{\Psi}_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ be given and $\Psi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ be such that, for all $\Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$,

$$\mathcal{A}_{\overline{\Psi}_{\mathcal{D}}}(\Psi_{\mathcal{D}}, \Phi_{\mathcal{D}}) := \mathcal{A}(\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}, \nabla_{\mathcal{D}}\Psi_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) = \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}). \quad (3.5.3)$$

Since $\mathcal{A}(\cdot,\cdot)$ is bilinear, $\mathcal{B}(\cdot,\cdot,\cdot)$ is trilinear and $\overline{\Psi}_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ is fixed, $\mathcal{A}_{\overline{\Psi}_{\mathcal{D}}}(\cdot,\cdot)$ is bilinear. Therefore, $\Psi_{\mathcal{D}}$ is sought as a solution to the bilinear system $\mathcal{A}_{\overline{\Psi}_{\mathcal{D}}}(\Psi_{\mathcal{D}}, \Phi_{\mathcal{D}}) = \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}})$. Since $\boldsymbol{X}_{\mathcal{D},0}$ is finite-dimensional and $\mathcal{L}(\Pi_{\mathcal{D}}\cdot)$ is linear, $\mathcal{L}(\Pi_{\mathcal{D}}\cdot)$ is a continuous linear functional on $\boldsymbol{X}_{\mathcal{D},0}$. Use the fact that $\mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}, \nabla_{\mathcal{D}}\Psi_{\mathcal{D}}, \nabla_{\mathcal{D}}\Psi_{\mathcal{D}}) = 0$ and $\mathcal{A}(\cdot,\cdot)$ is coercive to infer that

$$\mathcal{A}_{\overline{\Psi}_{\mathcal{D}}}(\Psi_{\mathcal{D}}, \Psi_{\mathcal{D}}) = \mathcal{A}(\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}) \geq \alpha \|\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}\|^2 = \alpha \|\Psi_{\mathcal{D}}\|_{\mathcal{D}}^2, \quad (3.5.4)$$

where α is the coercivity constant of $\mathcal{A}(\cdot,\cdot)$. Thus, $\mathcal{A}_{\overline{\Psi}_{\mathcal{D}}}(\cdot,\cdot)$ is coercive. The Lax Milgram Lemma implies the existence and uniqueness of solution $\Psi_{\mathcal{D}}$ satisfying (3.5.3). Define $F : \boldsymbol{X}_{\mathcal{D},0} \to \boldsymbol{X}_{\mathcal{D},0}$ by $F(\overline{\Psi}_{\mathcal{D}}) = \Psi_{\mathcal{D}}$, where $\Psi_{\mathcal{D}}$ is the solution to (3.5.3). To prove the continuity of F, consider $\overline{\Psi}_{\mathcal{D}}^n \to \overline{\Psi}_{\mathcal{D}}$ in $\boldsymbol{X}_{\mathcal{D},0}$ as $n \to \infty$. Let $F(\overline{\Psi}_{\mathcal{D}}^n) = \Psi_{\mathcal{D}}^n$ and $F(\overline{\Psi}_{\mathcal{D}}) = \Psi_{\mathcal{D}}$. From (3.5.5), $\|\Psi_{\mathcal{D}}^n\|_{\mathcal{D}}$ is bounded and thus, this space being finite dimensional, up to a subsequence we can assume $\Psi_{\mathcal{D}}^n \to \chi_{\mathcal{D}}$ for the $\boldsymbol{X}_{\mathcal{D},0}$ norm. It remains to prove that $\chi_{\mathcal{D}} = \Psi_{\mathcal{D}} = F(\overline{\Psi}_{\mathcal{D}})$. For that, consider the weak formulation of (3.5.3) for $\Psi_{\mathcal{D}}^n$: for all $\Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$, seeks $\Psi_{\mathcal{D}}^n \in \boldsymbol{X}_{\mathcal{D},0}$ such that

$$\mathcal{A}(\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}^n, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}^n, \nabla_{\mathcal{D}}\Psi_{\mathcal{D}}^n, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) = \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}).$$

$\Psi_{\mathcal{D}}^n \to \chi_{\mathcal{D}}$ in $\boldsymbol{X}_{\mathcal{D},0}$ shows that $\mathcal{A}(\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}^n, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}) \to \mathcal{A}(\mathcal{H}_{\mathcal{D}}\chi_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}})$ since $\mathcal{A}(\cdot,\cdot)$ is continuous w.r.t its first variable. Use the fact that $\overline{\Psi}_{\mathcal{D}}^n \to \overline{\Psi}_{\mathcal{D}}$ in $\boldsymbol{X}_{\mathcal{D},0}$, $\Psi_{\mathcal{D}}^n \to \chi_{\mathcal{D}}$ in $\boldsymbol{X}_{\mathcal{D},0}$, apply the coercivity property (3.3.2) and Lemma A.1.5 to pass to the limit in $\mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}^n, \nabla_{\mathcal{D}}\Psi_{\mathcal{D}}^n, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}})$ to obtain

$$\mathcal{A}(\mathcal{H}_{\mathcal{D}}\chi_{\mathcal{D}}, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}, \nabla_{\mathcal{D}}\chi_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) = \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}).$$

This precisely proves that $\chi_{\mathcal{D}} = F(\overline{\Psi}_{\mathcal{D}})$ and concludes the proof of the continuity of F. Moreover, (3.5.4) and (3.5.3) imply,

$$\alpha \|\Psi_{\mathcal{D}}\|_{\mathcal{D}}^2 \leq \mathcal{A}_{\overline{\Psi}_{\mathcal{D}}}(\Psi_{\mathcal{D}}, \Psi_{\mathcal{D}}) = \mathcal{L}(\Pi_{\mathcal{D}}\Psi_{\mathcal{D}}) \leq \|\mathcal{L}\| \|\Pi_{\mathcal{D}}\Psi_{\mathcal{D}}\| \leq C_{\mathcal{D}} \|\mathcal{L}\| \|\Psi_{\mathcal{D}}\|_{\mathcal{D}},$$

where $C_{\mathcal{D}}$ is defined by (3.3.2). Hence,

$$\|\Psi_{\mathcal{D}}\|_{\mathcal{D}} \leq \alpha^{-1} C_{\mathcal{D}} \|\mathcal{L}\| := R_{\mathcal{D}}. \quad (3.5.5)$$

This shows that F maps $\boldsymbol{X}_{\mathcal{D},0}$ into the closed ball $B_{R_{\mathcal{D}}}$ of center 0 and radius $R_{\mathcal{D}}$ with respect to $\|\cdot\|_{\mathcal{D}}$. Therefore, the Brouwer fixed point theorem proves that F has at least one fixed point $\Psi_{\mathcal{D}}$ in this ball. Recalling the problem (3.5.3) shows that this fixed point is a solution to (3.3.1).
From here onwards, let $\Psi_{\mathcal{D}_m} \in X_{\mathcal{D}_m,0}^k$ denote such a solution for $\mathcal{D} = \mathcal{D}_m$.

Step 2: strong convergence of $\Pi_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$ and $\nabla_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$, and weak convergence of $\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$.
From (3.5.5), $\alpha\|\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}\| = \alpha\|\Psi_{\mathcal{D}_m}\|_{\mathcal{D}_m} \leq C_{\mathcal{D}_m}\|\mathcal{L}\|$. Thus, $\|\Psi_{\mathcal{D}_m}\|_{\mathcal{D}_m}$ is bounded and Lemma 3.5.2 gives a subsequence of $(\mathcal{D}_m, \Psi_{\mathcal{D}_m})_{m\in\mathbb{N}}$, and $\Psi \in \boldsymbol{X}$, such that $\Pi_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$ converges weakly to Ψ in $L^2(\Omega)^k$, $\nabla_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$ converges weakly to $\nabla\Psi$ in $L^4(\Omega;\mathbb{R}^d)^k$, and $\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$ converges weakly to $\mathcal{H}\Psi$ in $L^2(\Omega;\mathcal{S}_d)^k$. The compactness hypothesis then shows that $\Pi_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$ converges strongly to Ψ in $L^2(\Omega)^k$ and $\nabla_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$ converges strongly to $\nabla\Psi$ in $L^4(\Omega;\mathbb{R}^d)^k$.

Step 3: Ψ is a solution to Problem (3.2.1).
Define $\mathcal{P}_{\mathcal{D}} : \boldsymbol{X} \to \boldsymbol{X}_{\mathcal{D},0}$ by

$$\mathcal{P}_{\mathcal{D}}\Psi = \operatorname*{argmin}_{w\in\boldsymbol{X}_{\mathcal{D},0}} \Big(\|\Pi_{\mathcal{D}}w - \Psi\| + \|\nabla_{\mathcal{D}}w - \nabla\Psi\|_{L^4} + \|\mathcal{H}_{\mathcal{D}}w - \mathcal{H}\Psi\| \Big). \tag{3.5.6}$$

The consistency of $(\mathcal{D}_m)_{m\in\mathbb{N}}$ implies $\Pi_{\mathcal{D}_m}\mathcal{P}_{\mathcal{D}_m}\Phi \to \Phi$ in $L^2(\Omega)^k$, $\nabla_{\mathcal{D}_m}\mathcal{P}_{\mathcal{D}_m}\Phi \to \nabla\Phi$ in $L^4(\Omega;\mathbb{R}^d)^k$ and $\mathcal{H}_{\mathcal{D}_m}\mathcal{P}_{\mathcal{D}_m}\Phi \to \mathcal{H}\Phi$ in $L^2(\Omega;\mathcal{S}_d)^k$ as $m \to \infty$. Using Lemma A.1.5 and the bilinearity and continuity of $\mathcal{A}$, as $m \to \infty$,

$$\begin{aligned}\mathcal{A}(\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}, \mathcal{H}_{\mathcal{D}_m}\mathcal{P}_{\mathcal{D}_m}\Phi) + B_{\mathcal{D}}(\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}, \nabla_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}, \nabla_{\mathcal{D}_m}\mathcal{P}_{\mathcal{D}_m}\Phi) \\ \to \mathcal{A}(\mathcal{H}\Psi, \mathcal{H}\Phi) + \mathcal{B}(\mathcal{H}\Psi, \nabla\Psi, \nabla\Phi).\end{aligned} \tag{3.5.7}$$

Moreover, since $\Pi_{\mathcal{D}_m}\mathcal{P}_{\mathcal{D}_m}\Phi \to \Phi$ in $L^2(\Omega)^k$ as $m \to \infty$,

$$\mathcal{L}(\Pi_{\mathcal{D}_m}\mathcal{P}_{\mathcal{D}_m}\Phi) \to \mathcal{L}(\Phi) \text{ as } m \to \infty. \tag{3.5.8}$$

Letting $\Phi_{\mathcal{D}_m} = \mathcal{P}_{\mathcal{D}_m}\Phi$ in (3.3.1) for $\mathcal{D} = \mathcal{D}_m$, use (3.5.7) and (3.5.8) to pass to the limit and conclude that Ψ is a solution to (3.2.1).

Step 4: strong convergence of $\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$.
From (3.3.1) for $\mathcal{D} = \mathcal{D}_m$, using $\mathcal{B}(\Xi,\Theta,\Theta) = 0$ for all $\Xi \in L^2(\Omega;\mathcal{S}_d)^k$ and $\Theta \in L^4(\Omega;\mathbb{R}^d)^k$, and passing to the limit, we obtain

$$\lim_{m\to\infty} \mathcal{A}(\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}_m}, \mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}_m}) = \lim_{m\to\infty} \mathcal{L}(\Pi_{\mathcal{D}}\Psi_{\mathcal{D}_m}) = \mathcal{L}(\Psi) = \mathcal{A}(\mathcal{H}\Psi, \mathcal{H}\Psi),$$

since Ψ is a solution to (3.2.1). By bilinearity of $\mathcal{A}$, we also have

$$\begin{aligned}&\mathcal{A}(\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} - \mathcal{H}\Psi, \mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} - \mathcal{H}\Psi) \\ &\qquad = \mathcal{A}(\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}, \mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}) - \mathcal{A}(\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}, \mathcal{H}\Psi) - \mathcal{A}(\mathcal{H}\Psi, \mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} - \mathcal{H}\Psi).\end{aligned}$$

The coercivity of $\mathcal{A}$ and weak convergence of $\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m}$ therefore lead to

$$\limsup_{m\to\infty} \alpha\|\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} - \mathcal{H}\Psi\|^2 \leq \limsup_{m\to\infty} \mathcal{A}(\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} - \mathcal{H}\Psi, \mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} - \mathcal{H}\Psi) = 0.$$

This shows that $\|\mathcal{H}_{\mathcal{D}_m}\Psi_{\mathcal{D}_m} - \mathcal{H}\Psi\| \to 0$ as $m \to \infty$. □

Remark 3.5.3. *As seen in Section 3.4.1, it is easy to construct a coercive, consistent, limit-conforming and compact sequence of HDs. A consequence of this and Theorem 3.5.1 is the existence of a solution to the abstract problem* (3.2.1).

3.5.2 Error estimates

The results that enable the proofs of local existence of discrete solution, uniqueness and error estimates (with respect to a regular solution) for the Hessian scheme are discussed in this section. Here, we assume that $\mathcal{A}(\cdot,\cdot)$ is the L^2 inner product on tensors. That is,

$$\forall \Phi, \forall \Theta \in L^2(\Omega;\mathbb{R}^{d\times d})^k, \quad \mathcal{A}(\Phi,\Theta) = \int_\Omega \Phi : \Theta \,\mathrm{d}\boldsymbol{x}. \tag{3.5.9}$$

Also assume that for all $\Phi \in L^2(\Omega;\mathbb{R}^{d\times d})^k$, $\Theta, \Xi \in L^4(\Omega;\mathbb{R}^d)^k$,

$$\mathcal{B}(\Phi,\Theta,\Xi) = \int_\Omega \Phi : h(\Theta,\Xi)\,\mathrm{d}\boldsymbol{x}, \tag{3.5.10}$$

where $h(\cdot,\cdot)$ is bilinear on $\mathbb{R}^{dk}\times\mathbb{R}^{dk}$. It can be verified that the bilinear and trilinear forms corresponding to the Navier–Stokes and von Kàrman equations (see Section 3.2.1) satisfy (3.5.9) and (3.5.10), respectively (up to a trivial scaling).

Remark 3.5.4. *For $\eta_\mathcal{D} \in X_{\mathcal{D},0}, \psi, \zeta \in H_0^2(\Omega)$,*

(i) for Navier–Stokes equation,

$$\mathcal{B}(\mathcal{H}_\mathcal{D}\eta_\mathcal{D}, \nabla\psi, \nabla\zeta) = \sum_{K\in\mathcal{M}} \int_K \mathrm{tr}(\mathcal{H}_\mathcal{D}\eta_\mathcal{D})\nabla\psi\cdot \mathrm{rot}_{\pi/2}(\nabla\zeta)\,\mathrm{d}\boldsymbol{x} = \sum_{\mathrm{K}\in\mathcal{M}} \int_\mathrm{K} \mathcal{H}_\mathcal{D}\eta_\mathcal{D} : \mathrm{P}_{\psi,\zeta},$$

where

$$P_{\psi,\zeta} = \begin{bmatrix} \partial_y\psi\partial_x\zeta - \partial_x\psi\partial_y\zeta & 0 \\ 0 & \partial_y\psi\partial_x\zeta - \partial_x\psi\partial_y\zeta \end{bmatrix}.$$

(ii) for von Kármán equations,

$$b_\mathcal{D}(\eta_\mathcal{D},\psi,\zeta) := \frac{1}{2}\sum_{K\in\mathcal{M}} \int_K \mathrm{cof}(\mathcal{H}_\mathcal{D}\eta_\mathcal{D})\nabla\psi\cdot\nabla\zeta = \frac{1}{2}\sum_{\mathrm{K}\in\mathcal{M}} \int_\mathrm{K} \mathcal{H}_\mathcal{D}\eta_\mathcal{D} : \mathrm{P}_{\psi,\zeta},$$

where

$$P_{\psi,\zeta} = \begin{bmatrix} \partial_y\psi\partial_y\zeta & -\partial_x\psi\partial_y\zeta \\ -\partial_y\psi\partial_x\zeta & \partial_x\psi\partial_x\zeta \end{bmatrix}.$$

In the following, we assume that Ω is convex. Then when the load function belongs to $H^{-1}(\Omega)^k$ [13, Theorem 7], the exact solution Ψ belongs to $H^3(\Omega)^k$. We note that, by Sobolev embeddings, this smoothness implies $\nabla\Psi \in (L^\infty(\Omega)^d)^k$ and $\Psi \in W^{2,4}(\Omega)^k$.

Fixing $\Psi \in \boldsymbol{X}$, a linearization of (3.2.1) around Ψ in the direction of Θ is given by

$$\mathbb{A}_\Psi(\Theta,\Phi) := \mathcal{A}(\mathcal{H}\Theta,\mathcal{H}\Phi) + \mathcal{B}(\mathcal{H}\Psi,\nabla\Theta,\nabla\Phi) + \mathcal{B}(\mathcal{H}\Theta,\nabla\Psi,\nabla\Phi).$$

Definition 3.5.5 (Regular solution [18]). *The solution Ψ of* (3.2.1) *is said to be regular if the linearization $\mathbb{A}_\Psi(\cdot,\cdot)$ is wellposed; that is, for a given $G \in L^2(\Omega)^k$,*

$$\mathbb{A}_\Psi(\Theta,\Phi) = (G,\Phi) \quad \forall \Phi \in \boldsymbol{X} \tag{3.5.11}$$

has a unique solution $\Theta \in \boldsymbol{X}$, and this solution satisfies $\|\Theta\|_{\boldsymbol{X}} \leq C\|G\|$, where C is independent of G.

As proved in [18], Ψ is a regular solution (that is, the bilinear form $\mathbb{A}_\Psi(\cdot,\cdot)$ is non-singular on $\boldsymbol{X}\times\boldsymbol{X}$) iff there exists a constant $\beta > 0$ such that

$$\beta\|\mathcal{H}\Theta\| \leq \sup_{\|\mathcal{H}\Phi\|=1} \mathbb{A}_\Psi(\Theta,\Phi);\ \ \beta\|\mathcal{H}\Phi\| \leq \sup_{\|\mathcal{H}\Theta\|=1} \mathbb{A}_\Psi(\Theta,\Phi). \tag{3.5.12}$$

The next lemma talks about the wellposedness of the dual problem. The result follows easily under the assumption that Ψ is a regular solution of (3.2.1).

Lemma 3.5.6 (Wellposedness of the dual problem [91]). *If Ω be a convex domain and Ψ is a regular solution of* (3.2.1)*, then the dual problem defined by: given $Q \in H^{-1}(\Omega)^k$, find $\zeta \in \boldsymbol{X}$ such that*

$$\mathbb{A}_\Psi(\Phi,\zeta) = (Q,\Phi), \quad \forall \Phi \in \boldsymbol{X}, \tag{3.5.13}$$

is well posed and satisfies the a priori bounds:

$$\|\zeta\|_{H^3(\Omega)^k} \leq C\|Q\|_{H^{-1}(\Omega)^k}. \tag{3.5.14}$$

The Hessian scheme that corresponds to the linearized problem (3.5.11) seeks $\Theta_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ such that

$$\mathbb{A}_{\mathcal{D},\Psi}(\Theta_{\mathcal{D}},\Phi_{\mathcal{D}}) = (G,\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}), \quad \forall \Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}, \tag{3.5.15}$$

where

$$\mathbb{A}_{\mathcal{D},\Psi}(\Theta_{\mathcal{D}},\Phi_{\mathcal{D}}) = \mathcal{A}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}},\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}\Psi,\nabla_{\mathcal{D}}\Theta_{\mathcal{D}},\nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}},\nabla\Psi,\nabla_{\mathcal{D}}\Phi_{\mathcal{D}}). \tag{3.5.16}$$

The wellposedness of the discrete linearized problem given by (3.5.15) can be proved if there exists a *companion operator* that maps the discrete space to the continuous space of functions with certain properties stated below.

(A5) (Companion operator) *There exists a linear map $E_{\mathcal{D}} : X_{\mathcal{D},0} \to H^2_0(\Omega)$ called the companion operator. We then define*

$$\delta(E_{\mathcal{D}}) := \sup_{\psi_{\mathcal{D}} \in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\|}{\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|}, \tag{3.5.17a}$$

$$\omega(E_{\mathcal{D}}) := \sup_{\psi_{\mathcal{D}} \in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}\|_{L^4}}{\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|}, \tag{3.5.17b}$$

$$\Gamma(E_{\mathcal{D}}) := \sup_{\psi_{\mathcal{D}} \in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\mathcal{H}E_{\mathcal{D}}\psi_{\mathcal{D}}\|}{\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|}. \tag{3.5.17c}$$

Remark 3.5.7 (Asymptotic behaviour)**.** *To establish error estimates using the companion operator, it will be expected that, along the considered sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ of HDs, the corresponding companion operators will be such that $\delta(E_{\mathcal{D}_m})\to 0$, $\omega(E_{\mathcal{D}_m})\to 0$ and $\Gamma(E_{\mathcal{D}_m})$ remains bounded. In Appendix A.2 we give an abstract generic construction of a companion operator that satisfies these properties if $(\mathcal{D}_m)_{m\in\mathbb{N}}$ is coercive, consistent, limit-conforming and compact. An explicit companion operator is well known for the Morley finite element, with $\delta(E_{\mathcal{D}})=h^2$, $\omega(E_{\mathcal{D}})=h^{1/2}$ and $\Gamma(E_{\mathcal{D}})\le C$, where C is independent of h, see [17].*

Inspired by the notion of space size for (2nd order) gradient discretisations [48, Definition 2.22], set,

$$\alpha_{\mathcal{D}}=\sup_{\phi\in(H^3(\Omega)\cap H_0^2(\Omega))\setminus\{0\}}\frac{S_{\mathcal{D}}(\phi)}{\|\phi\|_{H^3(\Omega)}},\ \gamma_{\mathcal{D}}=\sup_{\xi\in H_{\mathrm{div}}(\Omega)^d\setminus\{0\}}\frac{\widetilde{W}_{\mathcal{D}}(\xi)}{\|\xi\|_{H^1(\Omega)}}. \tag{3.5.18}$$

Remark 3.5.8. *Based on estimates that one can establish on $S_{\mathcal{D}}$ and $\widetilde{W}_{\mathcal{D}}$, it is expected that $\alpha_{\mathcal{D}}$ and $\gamma_{\mathcal{D}}$ will be small for HDs based on small meshes, see for example Theorem 3.4.2.*

Theorem 3.5.9 (Wellposedness of the discrete linearized problem)**.** *Let Ω be a convex domain and Ψ be a regular solution of* (3.2.1)*. For any $\Gamma\ge 0$, there exists $\rho>0$ such that if $C_{\mathcal{D}}\le\Gamma$, $\Gamma(E_{\mathcal{D}})\le\Gamma$, $\omega(E_{\mathcal{D}})\le\rho$, $\alpha_{\mathcal{D}}\le\rho$ and $\gamma_{\mathcal{D}}\le\rho$, then the discrete linearized problem* (3.5.15) *is well posed, and satisfied the inf-sup condition with a constant $\overline{\beta}$ that only depends on Γ,ρ and the inf-sup constant β given in* (3.5.12) *for the continuous problem.*

Proof. Since $\boldsymbol{X}_{\mathcal{D},0}$ is finite dimensional and (3.5.15) is linear, the existence of *a priori* bound implies that the problem has a unique solution. Let us therefore focus on establishing these *a priori* bounds. Let $\Phi_{\mathcal{D}}\in\boldsymbol{X}_{\mathcal{D},0}$. Then the definitions (3.5.16), (3.5.9) and (3.5.10) of $\mathbb{A}_{\mathcal{D},\Psi}$, $\mathcal{A}$ and $\mathcal{B}$, the generalised Hölder inequality and the definition (3.3.2) of $C_{\mathcal{D}}$ leads to the following Gårdings-type inequality:

$$\begin{aligned}\mathbb{A}_{\mathcal{D},\Psi}(\Phi_{\mathcal{D}},\Phi_{\mathcal{D}})&=\mathcal{A}(\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}},\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}})+\mathcal{B}(\mathcal{H}\Psi,\nabla_{\mathcal{D}}\Phi_{\mathcal{D}},\nabla_{\mathcal{D}}\Phi_{\mathcal{D}})+\mathcal{B}(\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}},\nabla\Psi,\nabla_{\mathcal{D}}\Phi_{\mathcal{D}})\\&\ge\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\|^2-CC_{\mathcal{D}}\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\|\|\nabla_{\mathcal{D}}\Phi_{\mathcal{D}}\|\|\mathcal{H}\Psi\|_{L^4}\\&\quad-C\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\|\|\nabla_{\mathcal{D}}\Phi_{\mathcal{D}}\|\|\nabla\Psi\|_{L^\infty(\Omega)},\end{aligned}$$

where $C>0$ is independent of $\mathcal{D}$. Substitute $\Phi_{\mathcal{D}}=\Theta_{\mathcal{D}}$ in (3.5.15), and use the above inequality and (3.3.2) to obtain

$$\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|\le C\big(C_{\mathcal{D}}\|\mathcal{H}\Psi\|_{L^4}+\|\nabla\Psi\|_{L^\infty(\Omega)}\big)\|\nabla_{\mathcal{D}}\Theta_{\mathcal{D}}\|+C_{\mathcal{D}}\|G\|. \tag{3.5.19}$$

A use of triangle inequality and (3.5.17b) leads to an estimate for $\|\nabla_{\mathcal{D}}\Theta_{\mathcal{D}}\|$ in the above expression as

$$\|\nabla_{\mathcal{D}}\Theta_{\mathcal{D}}\|\le\|\nabla_{\mathcal{D}}\Theta_{\mathcal{D}}-\nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}\|+\|\nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}\|\le\omega(E_{\mathcal{D}})\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|+\|\nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}\|. \tag{3.5.20}$$

To estimate $\|\nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}\|$, choose $Q = -\Delta E_{\mathcal{D}}\Theta_{\mathcal{D}}$ and $\Phi = E_{\mathcal{D}}\Theta_{\mathcal{D}}$ in (3.5.13). Introduce the terms $\pm\mathcal{B}(\mathcal{H}\Psi, \nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta)$ and $\pm\mathcal{B}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta)$, and use (3.5.15) with $\Phi_{\mathcal{D}} = \mathcal{P}_{\mathcal{D}}\zeta$ to obtain

$$\begin{aligned}\|\nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}\|^2 &= \mathbb{A}_{\Psi}(E_{\mathcal{D}}\Theta_{\mathcal{D}}, \zeta)\\ &= \mathcal{A}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}}, \mathcal{H}\zeta) + \mathcal{B}(\mathcal{H}\Psi, \nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\zeta - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta)\\ &\quad + \mathcal{B}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi, \nabla\zeta - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta) + \mathcal{B}(\mathcal{H}\Psi, \nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta)\\ &\quad + \mathcal{B}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta) - \mathbb{A}_{\mathcal{D},\Psi}(\Theta_{\mathcal{D}}, \mathcal{P}_{\mathcal{D}}\zeta) + (G, \Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta).\end{aligned}$$

An introduction of $\pm\mathcal{A}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \mathcal{H}\zeta)$ and $\pm\mathcal{B}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi, \nabla\zeta)$, leads to

$$\begin{aligned}\|\nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}\|^2 &= \mathcal{A}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \mathcal{H}\zeta) + \mathcal{A}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \mathcal{H}\zeta - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta)\\ &\quad + \mathcal{B}(\mathcal{H}\Psi, \nabla E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\zeta - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta) + \mathcal{B}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi, \nabla\zeta - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta)\\ &\quad + \mathcal{B}(\mathcal{H}\Psi, \nabla E_{\mathcal{D}}\Theta_{\mathcal{D}} - \nabla_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta) + \mathcal{B}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi, \nabla\zeta)\\ &\quad + \mathcal{B}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta - \nabla\zeta) + (G, \Pi_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta) =: \sum_{i=1}^{8} T_i. \qquad (3.5.21)\end{aligned}$$

We now estimate each T_i for $i = 1, \cdots, 8$. Use (3.5.9), an integration by parts, (3.3.6), Cauchy–Schwarz inequality, (3.5.17b) and (3.5.18) to obtain

$$\begin{aligned}T_1 = \mathcal{A}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \mathcal{H}\zeta) &= \int_{\Omega}(\mathcal{H}E_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}) : \mathcal{H}\zeta \,\mathrm{d}x\\ &\leq -\int_{\Omega}\nabla E_{\mathcal{D}}\Theta_{\mathcal{D}} \cdot \mathrm{div}(\mathcal{H}\zeta)\,\mathrm{d}x + \int_{\Omega}\nabla_{\mathcal{D}}\Theta_{\mathcal{D}} \cdot \mathrm{div}(\mathcal{H}\zeta)\,\mathrm{d}x + \widetilde{W}_{\mathcal{D}}(\mathcal{H}\zeta)\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|\\ &\leq \big(\omega(E_{\mathcal{D}})\|\mathrm{div}(\mathcal{H}\zeta)\| + \gamma_{\mathcal{D}}\|\mathcal{H}\zeta\|_{H^1(\Omega)}\big)\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|. \qquad (3.5.22)\end{aligned}$$

A use of (3.5.9), Cauchy–Schwarz inequality, (3.3.3) and (3.5.18) yields

$$T_2 = \mathcal{A}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \mathcal{H}\zeta - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta) \leq \|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|S_{\mathcal{D}}(\zeta) \leq \alpha_{\mathcal{D}}\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|\,\|\zeta\|_{H^3(\Omega)}. \qquad (3.5.23)$$

By the generalised Hölder's inequality, Sobolev imbedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$, (3.3.3), (3.5.17c) and (3.5.18), we have

$$\begin{aligned}T_3 &\leq C\Gamma(E_{\mathcal{D}})\|\mathcal{H}\Psi\|\,\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|S_{\mathcal{D}}(\zeta)\\ &\leq C\alpha_{\mathcal{D}}\Gamma(E_{\mathcal{D}})\|\mathcal{H}\Psi\|\,\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|\,\|\zeta\|_{H^3(\Omega)}, \qquad (3.5.24)\end{aligned}$$

$$\begin{aligned}T_4 &\leq C\Gamma(E_{\mathcal{D}})\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|S_{\mathcal{D}}(\zeta)\|\nabla\Psi\|_{L^4}\\ &\leq C\alpha_{\mathcal{D}}\Gamma(E_{\mathcal{D}})\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|\,\|\Psi\|_{H^2(\Omega)}\|\zeta\|_{H^3(\Omega)}, \qquad (3.5.25)\end{aligned}$$

and since $\|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\zeta\|_{L^4} \leq S_{\mathcal{D}}(\zeta) + \|\nabla\zeta\|_{L^4}$, from (3.5.17b) and (3.5.18),

$$\begin{aligned}T_5 &\leq C\omega(E_{\mathcal{D}})\|\mathcal{H}\Psi\|\,\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|\big(S_{\mathcal{D}}(\zeta) + \|\nabla\zeta\|_{L^4}\big)\\ &\leq C\omega(E_{\mathcal{D}})\big(\alpha_{\mathcal{D}} + 1\big)\|\mathcal{H}\Psi\|\,\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|\,\|\zeta\|_{H^3(\Omega)}. \qquad (3.5.26)\end{aligned}$$

A use of (3.5.10), integration by parts, (3.3.6), Cauchy–Schwarz inequality, (3.5.17b) and (3.5.18) leads to

$$\begin{aligned} T_6 &= \int_\Omega (\mathcal{H}E_\mathcal{D}\Theta_\mathcal{D} - \mathcal{H}_\mathcal{D}\Theta_\mathcal{D}) : h(\nabla\Psi, \nabla\zeta)\,\mathrm{d}x \\ &\leq \int_\Omega (\nabla_\mathcal{D}\Theta_\mathcal{D} - \nabla E_\mathcal{D}\Theta_\mathcal{D}) \cdot \mathrm{div}(h(\nabla\Psi, \nabla\zeta))\,\mathrm{d}x + \widetilde{W}_\mathcal{D}(h(\nabla\Psi, \nabla\zeta))\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\| \\ &\leq \big(\omega(E_\mathcal{D})\|\mathrm{div}(h(\nabla\Psi, \nabla\zeta))\| + \widetilde{W}_\mathcal{D}(h(\nabla\Psi, \nabla\zeta))\big)\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\| \\ &\leq (\omega(E_\mathcal{D}) + \gamma_\mathcal{D})\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\|\,\|\Psi\|_{H^3(\Omega)}\|\zeta\|_{H^3(\Omega)}. \end{aligned} \tag{3.5.27}$$

Apply the generalised Hölder's inequality, (3.5.17c), (3.3.3), (3.5.18) and Sobolev imbedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$ to obtain

$$\begin{aligned} T_7 &\leq C(\Gamma(E_\mathcal{D}) + 1)\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\|S_\mathcal{D}(\zeta)\|\nabla\Psi\|_{L^4} \\ &\leq C\alpha_\mathcal{D}(\Gamma(E_\mathcal{D}) + 1)\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\|\,\|\Psi\|_{H^2(\Omega)}\|\zeta\|_{H^3(\Omega)}. \end{aligned} \tag{3.5.28}$$

A use of Cauchy–Schwarz inequality, (3.3.3) and (3.5.18) leads to

$$T_8 = (G, \Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\zeta) \leq \|G\|(S_\mathcal{D}(\zeta) + \|\zeta\|) \leq \|G\|(\alpha_\mathcal{D}\|\zeta\|_{H^3(\Omega)} + \|\zeta\|). \tag{3.5.29}$$

Plug in (3.5.22)–(3.5.29) into (3.5.21), use the Sobolev imbedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$, (3.5.14) and $\|Q\|_{H^{-1}(\Omega)} \leq \|\nabla E_\mathcal{D}\Theta_\mathcal{D}\|$ to obtain

$$\|\nabla E_\mathcal{D}\Theta_\mathcal{D}\| \leq C\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\|\big(\omega(E_\mathcal{D}) + \gamma_\mathcal{D} + \alpha_\mathcal{D}(1 + \Gamma(E_\mathcal{D}) + \omega(E_\mathcal{D}))\big) + \|G\|(\alpha_\mathcal{D} + 1), \tag{3.5.30}$$

where $C > 0$ is independent of $\mathcal{D}$, but depends on Ψ. Now, (3.5.19), (3.5.20) and (3.5.30) yield C independent of $\mathcal{D}$ such that

$$\begin{aligned} \|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\| \leq C(C_\mathcal{D} + 1)\big[\omega(E_\mathcal{D}) + \gamma_\mathcal{D} + \alpha_\mathcal{D}(1 + \Gamma(E_\mathcal{D}) + \omega(E_\mathcal{D}))\big]\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\| \\ + (\alpha_\mathcal{D} + 1 + C_\mathcal{D})\|G\|. \end{aligned} \tag{3.5.31}$$

For $C_\mathcal{D} \leq \Gamma$ and $\Gamma(E_\mathcal{D}) \leq \Gamma$, choose ρ such that

$$C(\Gamma + 1)(3\rho + \rho(\Gamma + \rho)) \leq 1/2.$$

If $\omega(E_\mathcal{D}) \leq \rho$, $\alpha_\mathcal{D} \leq \rho$ and $\gamma_\mathcal{D} \leq \rho$, the estimate (3.5.31) gives $\|\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}\| \leq 2C(\rho + 1 + \Gamma)\|G\|$, which is the sought *a priori* estimate on the solution to the linearized problem (3.5.15). □

Lemma 3.5.10 (Non-singularity of perturbed bilinear form)**.** *Under the assumptions of Theorem 3.5.9, and reducing perhaps ρ, the perturbed bilinear form defined by*

$$\begin{aligned} \mathfrak{A}_{\mathcal{D},\Psi}(\Theta_\mathcal{D}, \Phi_\mathcal{D}) = \mathcal{A}(\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}, \mathcal{H}_\mathcal{D}\Phi_\mathcal{D}) + \mathcal{B}(\mathcal{H}_\mathcal{D}\mathcal{P}_\mathcal{D}\Psi, \nabla_\mathcal{D}\Theta_\mathcal{D}, \nabla_\mathcal{D}\Phi_\mathcal{D}) \\ + \mathcal{B}(\mathcal{H}_\mathcal{D}\Theta_\mathcal{D}, \nabla_\mathcal{D}\mathcal{P}_\mathcal{D}\Psi, \nabla_\mathcal{D}\Phi_\mathcal{D}) \end{aligned} \tag{3.5.32}$$

is non-singular on $\boldsymbol{X}_{\mathcal{D},0} \times \boldsymbol{X}_{\mathcal{D},0}$.

Proof. From Theorem 3.5.9, for $\Theta_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$, there exists $\Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ such that $\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\| = 1$ and $\overline{\beta}\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\| \leq \mathbb{A}_{\mathcal{D},\Psi}(\Theta_{\mathcal{D}},\Phi_{\mathcal{D}})$. Then, from (3.5.32), (3.5.16), (3.3.2) and (3.3.3), for some C independent of $\mathcal{D}$,

$$\begin{aligned}\mathfrak{A}_{\mathcal{D},\Psi}(\Theta_{\mathcal{D}},\Phi_{\mathcal{D}}) &= \mathbb{A}_{\mathcal{D},\Psi}(\Theta_{\mathcal{D}},\Phi_{\mathcal{D}}) - \mathcal{B}(\mathcal{H}\Psi - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\qquad - \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla\Psi - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\geq \overline{\beta}\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\| - CC_{\mathcal{D}}(1+C_{\mathcal{D}})S_{\mathcal{D}}(\Psi)\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\| \geq \frac{\overline{\beta}}{2}\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}\|,\end{aligned}$$

provided $S_{\mathcal{D}}(\Psi) \leq \frac{\overline{\beta}}{2CC_{\mathcal{D}}(1+C_{\mathcal{D}})}$. Hence the required result, provided that ρ is as in Theorem 3.5.9 and further satisfies $\rho\|\Psi\|_{H^3(\Omega)} \leq \frac{\overline{\beta}}{2CC_{\mathcal{D}}(1+C_{\mathcal{D}})}$. □

Existence and error estimates: Under the assumptions of Lemma 3.5.10, we can define the nonlinear operator $\mu : \boldsymbol{X}_{\mathcal{D},0} \to \boldsymbol{X}_{\mathcal{D},0}$ such that, for $\Theta_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$, $\mu(\Theta_{\mathcal{D}})$ is the unique solution to:

$$\begin{aligned}\mathfrak{A}_{\mathcal{D},\Psi}(\mu(\Theta_{\mathcal{D}}),\Phi_{\mathcal{D}}) &= \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\quad + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) - \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}),\end{aligned} \tag{3.5.33}$$

for all $\Phi_{\mathcal{D}} \in X_{\mathcal{D},0}$. Observe that any fixed point of μ is a solution to (3.3.1) and vice-versa. Hence, in order to establish the existence of a solution to (3.3.1), we will prove that the mapping μ has a fixed point. For $R > 0$, define

$$B_R(\mathcal{P}_{\mathcal{D}}\Psi) = \{\Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0} : \|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq R\}.$$

Theorem 3.5.11. *(Mapping of ball into ball) Let Ω be a convex domain and Ψ be a regular solution of* (3.2.1)*. For any $\Gamma \geq 0$, there exists $\rho > 0$ such that if $C_{\mathcal{D}} \leq \Gamma$, $\gamma_{\mathcal{D}} \leq \rho$, $\omega(E_{\mathcal{D}}) \leq \rho$, $\alpha_{\mathcal{D}} \leq \rho$ and $\delta(E_{\mathcal{D}}) \leq \rho$, then there exists $\mathcal{K} > 0$ not depending on ρ or $\mathcal{D}$ such that, setting $R = \mathcal{K}\rho$, μ maps $B_R(\mathcal{P}_{\mathcal{D}}\Psi)$ into itself.*

Proof. Since $\mathfrak{A}_{\mathcal{D},\Psi}(\cdot,\cdot)$ is non-singular, by Theorem 3.5.9, there exists $\Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ such that $\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\| = 1$ and

$$\overline{\beta}\|\mathcal{H}_{\mathcal{D}}\mu(\Theta_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq \mathfrak{A}_{\mathcal{D},\Psi}(\mu(\Theta_{\mathcal{D}}) - \mathcal{P}_{\mathcal{D}}\Psi, \Phi_{\mathcal{D}}). \tag{3.5.34}$$

A use of (3.5.33), (3.5.32) and (3.2.1) with $\Phi = E_{\mathcal{D}}\Phi_{\mathcal{D}}$ yields

$$\begin{aligned}&\mathfrak{A}_{\mathcal{D},\Psi}(\mu(\Theta_{\mathcal{D}}) - \mathcal{P}_{\mathcal{D}}\Psi, \Phi_{\mathcal{D}}) \\ &\quad= \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\qquad - \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) - \mathcal{A}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\qquad - 2\mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\quad= \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}} - E_{\mathcal{D}}\Phi_{\mathcal{D}}) + [\mathcal{A}(\mathcal{H}\Psi, \mathcal{H}E_{\mathcal{D}}\Phi_{\mathcal{D}}) - \mathcal{A}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}})] \\ &\qquad + [\mathcal{B}(\mathcal{H}\Psi, \nabla\Psi, \nabla E_{\mathcal{D}}\Phi_{\mathcal{D}}) - \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}})] \\ &\qquad + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi - \mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}}, \nabla_{\mathcal{D}}\Theta_{\mathcal{D}} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) =: \sum_{i=1}^{4} B_i.\end{aligned} \tag{3.5.35}$$

Use the continuity of $\mathcal{L}(\cdot)$ and (3.5.17a) to estimate B_1 as

$$B_1 \lesssim \|\Pi_{\mathcal{D}}\Phi_{\mathcal{D}} - E_{\mathcal{D}}\Phi_{\mathcal{D}}\| \lesssim \delta(E_{\mathcal{D}})\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\|. \tag{3.5.36}$$

Introduce the term $\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}$ and use (3.5.22) with $(\Theta_{\mathcal{D}}, \zeta) = (\Phi_{\mathcal{D}}, \Psi)$, (3.3.3) and (3.5.18) to obtain

$$\begin{aligned} B_2 &\leq \big|\mathcal{A}(\mathcal{H}\Psi, \mathcal{H}E_{\mathcal{D}}\Phi_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}})\big| + \big|\mathcal{A}(\mathcal{H}\Psi - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}})\big| \\ &\leq \big(\omega(E_{\mathcal{D}})\|\mathrm{div}(\mathcal{H}\Psi)\| + \gamma_{\mathcal{D}}\|\mathcal{H}\Psi\|_{H^1(\Omega)} + \alpha_{\mathcal{D}}\|\Psi\|_{H^3(\Omega)}\big)\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\|. \end{aligned} \tag{3.5.37}$$

A use of the continuity of $\mathcal{B}(\cdot,\cdot,\cdot)$, (3.5.17b), (3.3.2), (3.3.3), the Sobolev imbedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$ and (3.5.18) leads to

$$\begin{aligned} B_3 &\leq \big|\mathcal{B}(\mathcal{H}\Psi, \nabla\Psi, \nabla E_{\mathcal{D}}\Phi_{\mathcal{D}} - \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) - \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi - \mathcal{H}\Psi, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}})\big| \\ &\quad + \big|\mathcal{B}(\mathcal{H}\Psi, \nabla\Psi - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}})\big| \\ &\lesssim \big[\omega(E_{\mathcal{D}})\|\mathcal{H}\Psi\|\|\nabla\Psi\|_{L^4} + \big(\|\nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\|_{L^4} + \|\mathcal{H}\Psi\|\big)C_{\mathcal{D}}S_{\mathcal{D}}(\Psi)\big]\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\| \\ &\lesssim \big[\omega(E_{\mathcal{D}})\|\mathcal{H}\Psi\|^2 + \big(\alpha_{\mathcal{D}}\|\Psi\|_{H^3(\Omega)} + \|\mathcal{H}\Psi\|\big)C_{\mathcal{D}}\alpha_{\mathcal{D}}\|\Psi\|_{H^3(\Omega)}\big]\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\|. \end{aligned} \tag{3.5.38}$$

Finally, use (3.3.2) to estimate B_4 as

$$B_4 \lesssim C_{\mathcal{D}}^2\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\|^2\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\|. \tag{3.5.39}$$

A substitution of (3.5.36)-(3.5.39) in (3.5.35) and then in (3.5.34) leads to

$$\begin{aligned} \|\mathcal{H}_{\mathcal{D}}\mu(\Theta_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq \mathcal{K}_1\big(&\gamma_{\mathcal{D}} + \alpha_{\mathcal{D}}(C_{\mathcal{D}} + 1 + C_{\mathcal{D}}\alpha_{\mathcal{D}}) + \delta(E_{\mathcal{D}}) \\ &+ \omega(E_{\mathcal{D}}) + C_{\mathcal{D}}^2\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\|^2\big), \end{aligned}$$

where $\mathcal{K}_1 > 0$ is independent of $\mathcal{D}$, but depends on $\overline{\beta}$ and $\|\Psi\|_{H^3(\Omega)}$. Let $\Gamma > 0$ and $\rho > 0$ be such that $C_{\mathcal{D}} \leq \Gamma$, $\gamma_{\mathcal{D}} \leq \rho$, $\omega(E_{\mathcal{D}}) \leq \rho$, $\alpha_{\mathcal{D}} \leq \rho$ and $\delta(E_{\mathcal{D}}) \leq \rho$. Then,

$$\|\mathcal{H}_{\mathcal{D}}\mu(\Theta_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq \mathcal{K}_1\big(\rho(\Gamma + 4 + \Gamma\rho) + \Gamma^2\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\|^2\big).$$

Choose $\rho > 0$ such that $4\mathcal{K}_1^2\Gamma^2\rho(\Gamma + 4 + \Gamma\rho) \leq 1$. Setting $R = 2\mathcal{K}_1\rho(\Gamma + 4 + \Gamma\rho)$ and assuming that $\|\mathcal{H}_{\mathcal{D}}\Theta_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq R$ leads to

$$\|\mathcal{H}_{\mathcal{D}}\mu(\Theta_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq \mathcal{K}_1\rho(\Gamma + 4 + \Gamma\rho)\big(1 + 4\mathcal{K}_1^2\Gamma^2\rho(\Gamma + 4 + \Gamma\rho)\big) \leq R.$$

This completes the proof. □

Theorem 3.5.12 (Existence of discrete solution and error estimates)**.** *Let Ω be a convex domain and Ψ be a regular solution of* (3.2.1)*. For any $\Gamma \geq 0$, there exists $\rho > 0$ such that if $C_{\mathcal{D}} \leq \Gamma$, $\Gamma(E_{\mathcal{D}}) \leq \Gamma$, $\omega(E_{\mathcal{D}}) \leq \rho$, $\delta(E_{\mathcal{D}}) \leq \rho$, $\alpha_{\mathcal{D}} \leq \rho$ and $\gamma_{\mathcal{D}} \leq \rho$, then there exists a solution $\Psi_{\mathcal{D}}$ of* (3.3.1) *that satisfies $\|\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq C\rho$, as well as the following estimates:*

$$\|\mathcal{H}\Psi - \mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}\| \leq C\rho, \|\nabla\Psi - \nabla_{\mathcal{D}}\Psi_{\mathcal{D}}\|_{L^4} \leq C\rho, \|\Psi - \Pi_{\mathcal{D}}\Psi_{\mathcal{D}}\| \leq C\rho, \tag{3.5.40}$$

where $C > 0$ depends on $\overline{\beta}, \Gamma, \Psi$ but not ρ or $\mathcal{D}$.

Proof. In order to prove the continuity of μ, take $\overline{\Psi}^m_{\mathcal{D}} \to \overline{\Psi}_{\mathcal{D}}$ in $\boldsymbol{X}_{\mathcal{D},0}$ as $m \to \infty$, let $\mu(\overline{\Psi}^m_{\mathcal{D}}) = \Psi^m_{\mathcal{D}}$ and $\mu(\overline{\Psi}_{\mathcal{D}}) = \Psi_{\mathcal{D}}$. The definition of the non-singularity of $\mathfrak{A}_{\mathcal{D},\Psi}(\cdot,\cdot)$ and Theorem 3.5.9 imply that there exists $\Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ such that $\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\| = 1$ and

$$\begin{aligned}\overline{\beta}\|\mathcal{H}_{\mathcal{D}}\Psi^m_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}\| =& \overline{\beta}\|\mathcal{H}_{\mathcal{D}}\mu(\overline{\Psi}^m_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mu(\overline{\Psi}_{\mathcal{D}})\| \leq \mathfrak{A}_{\mathcal{D},\Psi}(\mu(\overline{\Psi}^m_{\mathcal{D}}) - \mu(\overline{\Psi}_{\mathcal{D}}), \Phi_{\mathcal{D}})\\ =& \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\overline{\Psi}^m_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}^m_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}})\\ &- \mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}^m_{\mathcal{D}}, \nabla_{\mathcal{D}}\overline{\Psi}^m_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) - \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}})\\ &- \mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}, \nabla_{\mathcal{D}}\overline{\Psi}_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}).\end{aligned}$$

Since $\overline{\Psi}^m_{\mathcal{D}} \to \overline{\Psi}_{\mathcal{D}}$ in $\boldsymbol{X}_{\mathcal{D},0}$ as $m \to \infty$, we conclude $\Psi^m_{\mathcal{D}} \to \Psi_{\mathcal{D}}$ in $\boldsymbol{X}_{\mathcal{D},0}$ (the finite dimensional property of $\boldsymbol{X}_{\mathcal{D},0}$ and the linearity of $\mathcal{B}(\cdot,\cdot,\cdot)$ show that $\mathcal{B}(\cdot,\cdot,\cdot)$ is a continuous trilinear function on $\boldsymbol{X}_{\mathcal{D},0}$). Since μ maps the ball $B_R(\mathcal{P}_{\mathcal{D}}\Psi)$ to itself from Theorem 3.5.11, the Brouwer fixed point theorem yields that the mapping μ has a fixed point, say $\Psi_{\mathcal{D}}$. Hence, $\Psi_{\mathcal{D}}$ is a solution of (3.3.1) that satisfies $\|\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \leq R$, where $R = \mathcal{K}\rho$. This proves the existence part in Theorem 3.5.12, as well as the estimate on $\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi$.
Let us now estimate $\|\mathcal{H}\Psi - \mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}\|$. Introduce the term $\mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}$ and use the definition of $S_{\mathcal{D}}$ and (3.5.18) to obtain

$$\|\mathcal{H}\Psi - \mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}\| \leq \|\mathcal{H}\Psi - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| + \|\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi - \mathcal{H}_{\mathcal{D}}\Psi_{\mathcal{D}}\| \leq (\mathcal{K}+1)\rho.$$

The remaining two estimates in (3.5.40) follows in a similar way using triangle inequality, (3.3.3) and (3.3.2). □

The following lemma establishes the local uniqueness of the solution of (3.3.1).

Lemma 3.5.13 (Local uniqueness)**.** *Let Ω be a convex domain and Ψ be a regular solution of* (3.2.1)*. For $\Theta^1_{\mathcal{D}}, \Theta^2_{\mathcal{D}} \in B_R(\mathcal{P}_{\mathcal{D}}\Psi)$ with R as defined as in Theorem 3.5.11, the following result holds true:*

$$\|\mathcal{H}_{\mathcal{D}}\mu(\Theta^1_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mu(\Theta^2_{\mathcal{D}})\| \leq CR\|\mathcal{H}_{\mathcal{D}}\Theta^1_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta^2_{\mathcal{D}}\|,$$

for some positive constant C independent of $\mathcal{D}$, but depends on Γ and $\overline{\beta}$. Hence, if ρ is small enough, μ is a contraction on $B_R(\mathcal{P}_{\mathcal{D}}\Psi)$ and the solution to (3.3.1) *in this ball is unique.*

Proof. For $\Theta^1_{\mathcal{D}}, \Theta^2_{\mathcal{D}} \in B_R(\mathcal{P}_{\mathcal{D}}\Psi)$, let $\mu(\Theta^i_{\mathcal{D}}), i = 1,2$ be the solutions of:

$$\begin{aligned}\mathfrak{A}_{\mathcal{D},\Psi}(\mu(\Theta^i_{\mathcal{D}}), \Phi_{\mathcal{D}}) = \mathcal{L}(\Pi_{\mathcal{D}}\Phi_{\mathcal{D}}) + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Theta^i_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}})\\ + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta^i_{\mathcal{D}}, \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) - \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta^i_{\mathcal{D}}, \nabla_{\mathcal{D}}\Theta^i_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}). \quad (3.5.41)\end{aligned}$$

The non-singularity of $\mathfrak{A}_{\mathcal{D},\Psi}(\cdot,\cdot)$, Theorem 3.5.9, (3.5.41), the continuity of $\mathcal{B}(\cdot,\cdot,\cdot)$ and (3.3.2)

leads to the existence of $\Phi_{\mathcal{D}} \in \boldsymbol{X}_{\mathcal{D},0}$ such that $\|\mathcal{H}_{\mathcal{D}}\Phi_{\mathcal{D}}\| = 1$ and

$$\begin{aligned}\overline{\beta}\|\mathcal{H}_{\mathcal{D}}\mu(\Theta^1_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mu(\Theta^2_{\mathcal{D}})\| &\leq \mathfrak{A}_{\mathcal{D},\Psi}(\mu(\Theta^1_{\mathcal{D}}) - \mu(\Theta^2_{\mathcal{D}}), \Phi_{\mathcal{D}}) \\ &= \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta^2_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta^1_{\mathcal{D}}, \nabla_{\mathcal{D}}\Theta^1_{\mathcal{D}} - \nabla_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\quad + \mathcal{B}(\mathcal{H}_{\mathcal{D}}\Theta^2_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi, \nabla_{\mathcal{D}}\Theta^2_{\mathcal{D}} - \nabla_{\mathcal{D}}\Theta^1_{\mathcal{D}}, \nabla_{\mathcal{D}}\Phi_{\mathcal{D}}) \\ &\leq CC^2_{\mathcal{D}}\|\mathcal{H}_{\mathcal{D}}\Theta^2_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta^1_{\mathcal{D}}\| \Big(\|\mathcal{H}_{\mathcal{D}}\Theta^1_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \\ &\qquad + \|\mathcal{H}_{\mathcal{D}}\Theta^2_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\mathcal{P}_{\mathcal{D}}\Psi\| \Big).\end{aligned}$$

Since $\Theta^1_{\mathcal{D}}, \Theta^2_{\mathcal{D}} \in B_R(\mathcal{P}_{\mathcal{D}}\Psi)$ and $C_{\mathcal{D}} \leq \Gamma$, for a choice of R as in the proof of Theorem 3.5.11, we obtain

$$\|\mathcal{H}_{\mathcal{D}}\mu(\Theta^1_{\mathcal{D}}) - \mathcal{H}_{\mathcal{D}}\mu(\Theta^2_{\mathcal{D}})\| \leq CR\|\mathcal{H}_{\mathcal{D}}\Theta^1_{\mathcal{D}} - \mathcal{H}_{\mathcal{D}}\Theta^2_{\mathcal{D}}\|,$$

where C depends only on Γ and $\overline{\beta}$. This completes the proof. □

Remark 3.5.14 (Companion operators)**.** *For conforming FEMs, companion operator $E_{\mathcal{D}}$ is nothing but the identity operator which implies $\delta(E_{\mathcal{D}}) = 0, \omega(E_{\mathcal{D}}) = 0$ and $\Gamma(E_{\mathcal{D}}) = 1$. An explicit companion operator is well known for the Morley ncFEM, with $\delta(E_{\mathcal{D}}) = \mathcal{O}(h^2)$, $\omega(E_{\mathcal{D}}) = \mathcal{O}(h^{1/2})$ and $\Gamma(E_{\mathcal{D}}) \leq C$, where C is independent of h, see [17]. The estimates $\delta(E_{\mathcal{D}}) = \mathcal{O}(h^2)$, $\omega(E_{\mathcal{D}}) = \mathcal{O}(h^{1/2})$ and $\Gamma(E_{\mathcal{D}}) \leq C$ for a companion operator which maps the Adini rectangle to the Bogner–Fox–Schmit rectangle has been done in [14]. The construction of a companion operator $E_{\mathcal{D}}$ for the method based on gradient recovery operators with $\omega(E_{\mathcal{D}})$ and $\delta(E_{\mathcal{D}})$ small enough and $\Gamma(E_{\mathcal{D}})$ bounded is an open problem.*

3.6 Numerical results

In this section, numerical results for the Navier–Stokes (NS) equation and the von Kármán (vK) equations using the gradient recovery method and the Morley FEM are performed. Let the relative errors in $L^2(\Omega)$, $H^1(\Omega)$ and $H^2(\Omega)$ norms be denoted by

$$\mathsf{err}_{\mathcal{D}}(\overline{u}) := \frac{\|\Pi_{\mathcal{D}}u_{\mathcal{D}} - \overline{u}\|}{\|\overline{u}\|}, \quad \mathsf{err}_{\mathcal{D}}(\nabla\overline{u}) := \frac{\|\nabla_{\mathcal{D}}u_{\mathcal{D}} - \nabla\overline{u}\|}{\|\nabla\overline{u}\|}, \quad \mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u}) := \frac{\|\mathcal{H}_{\mathcal{D}}u_{\mathcal{D}} - \mathcal{H}\overline{u}\|}{\|\mathcal{H}\overline{u}\|},$$

where $\overline{u}$ is the continuous solution and $u_{\mathcal{D}}$ is the corresponding HS solution. Here, h denotes the mesh sizes and **nu** be the numbers of unknowns. The model problem is constructed in such a way that the exact solution is known. The discrete problem is solved using Newton's method.

3.6.1 Numerical results for Gradient Recovery Method

In this section, we consider the unit square domain $\Omega = (0,1)^2$ and the stabilisation parameter τ is chosen to be 1 that corresponds to the stabilisation function $\mathfrak{S}_h$. The finite dimensional space V_h is the conforming $\mathbb{P}_1$ space.

Navier Stokes equation

For the numerical experiments for the Navier Stokes equation, the viscosity ν is chosen equal to 0.01 or 1.

Example 1:

The exact solution $\overline{u}$ is given by $\overline{u}(x,y) = x^2y^2(1-x)^2(1-y)^2$. The source term f can be computed using (3.2.2). The errors and orders of convergence for the numerical approximation of $\overline{u}$ are shown in Tables 3.1 and 3.2. It can be seen that the rate of convergence is close to quadratic in L^2 and H^1 norm and is linear in H^2 norm when $\nu = 1$. Though there is a difficulty of constructing a proper companion operator and hence theoretical rate of convergence are not obtained for GR method (Remark 3.5.14), we observe expected rate of convergence numerically (see Section 2.7.1 for the numerical results for GR method conducted on the biharmonic problem). However, there is a loss of rate in the L^2 and H^1 norms when $\nu = 0.01$.

Table 3.1: (GR) Convergence results, NS, Example 1, $\nu = 0.01$

h	**nu**	$\mathrm{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	9	1.051258	-	0.568001	-	0.583070	-
0.176777	49	0.214196	2.2951	0.167145	1.7648	0.267212	1.1257
0.088388	225	0.067513	1.6657	0.050018	1.7406	0.128592	1.0552
0.044194	961	0.019298	1.8067	0.014060	1.8308	0.062380	1.0436
0.022097	3969	0.005363	1.8473	0.004567	1.6223	0.030876	1.0146
0.011049	16129	0.001992	1.4291	0.002924	0.6432	0.015897	0.9577

Table 3.2: (GR) Convergence results, NS, Example 1, $\nu = 1$

h	**nu**	$\mathrm{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathrm{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	9	1.050933	-	0.567673	-	0.582651	-
0.176777	49	0.214195	2.2947	0.167145	1.7640	0.267188	1.1248
0.088388	225	0.067498	1.6660	0.049952	1.7425	0.128511	1.0560
0.044194	961	0.019240	1.8107	0.013806	1.8552	0.062184	1.0473
0.022097	3969	0.005156	1.8999	0.003646	1.9209	0.030460	1.0296
0.011049	16129	0.001336	1.9482	0.000939	1.9575	0.015060	1.0162

Example 2:

In this example, $\overline{u}(x,y) = x^3y^3(1-x)^3(1-y)^3(\exp(x)\sin(2\pi x) + \cos(2\pi x))$. The errors together with their order of convergences are presented in Tables 3.3 and 3.4. We observe on this example similar rates of convergence to those obtained in Example 1.

Table 3.3: (GR) Convergence results, NS, Example 2, $\nu = 0.01$

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	9	8.711003	-	0.951241	-	0.991303	-
0.176777	49	0.517173	4.0741	0.224828	2.0810	0.492709	1.0086
0.088388	225	0.089445	2.5316	0.047787	2.2341	0.203366	1.2767
0.044194	961	0.017152	2.3826	0.010256	2.2201	0.084506	1.2669
0.022097	3969	0.004427	1.9541	0.003024	1.7619	0.041260	1.0343
0.011049	16129	0.002039	1.1186	0.002032	0.5736	0.020654	0.9983

Table 3.4: (GR) Convergence results, NS, Example 2, $\nu = 1$

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	9	8.708421	-	0.950596	-	0.990463	-
0.176777	49	0.516906	4.0744	0.224876	2.0797	0.492556	1.0078
0.088388	225	0.089333	2.5326	0.048053	2.2264	0.203302	1.2767
0.044194	961	0.016922	2.4003	0.010346	2.2156	0.084441	1.2676
0.022097	3969	0.003955	2.0971	0.002642	1.9692	0.041186	1.0358
0.011049	16129	0.000980	2.0134	0.000690	1.9378	0.020528	1.0045

The von Kármán equations

In this example, choose the right-hand side load functions such that $\overline{u} = \overline{v} = x^2y^2(1-x)^2(1-y)^2$. The load functions are computed by $f = \Delta^2\overline{u} - [\overline{u},\overline{v}]$ and $g = \Delta^2\overline{v} + \frac{1}{2}[\overline{u},\overline{u}]$. Tables 3.5 and 3.6 show the relative errors and orders of convergence for the variable $u_{\mathcal{D}}$ and $v_{\mathcal{D}}$. The tables provide once again linear rate of convergences in the energy norm for both the variables. Observe that qudratic rate of convergences are obtained in L^2 and H^1 norms for $v_{\mathcal{D}}$ whereas a loss of quadratic rate is noticed for $u_{\mathcal{D}}$.

Table 3.5: (GR) Convergence results for the relative errors of $\overline{u}$, vK

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
0.353553	9	1.049207	-	0.567835	-	0.582623	-
0.176777	49	0.214594	2.2896	0.167636	1.7601	0.267284	1.1242
0.088388	225	0.067946	1.6591	0.050446	1.7325	0.128565	1.0559
0.044194	961	0.019702	1.7860	0.014295	1.8192	0.062217	1.0471
0.022097	3969	0.005632	1.8068	0.004146	1.7858	0.030483	1.0293
0.011049	16129	0.001844	1.6109	0.001483	1.4828	0.015082	1.0152

Table 3.6: (GR) Convergence results for the relative errors of $\overline{v}$, vK

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{v})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{v})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{v})$	Order
0.353553	9	1.051793	-	0.567587	-	0.582660	-
0.176777	49	0.213996	2.2972	0.166900	1.7659	0.267141	1.1251
0.088388	225	0.067275	1.6694	0.049707	1.7475	0.128485	1.0560
0.044194	961	0.0190112	1.8232	0.013567	1.8733	0.062171	1.0473
0.022097	3969	0.004929	1.9474	0.003417	1.9894	0.030454	1.0296
0.011049	16129	0.001124	2.1325	0.000742	2.2036	0.015059	1.0160

3.6.2 Numerical results for Morley FEM

The results of numerical experiments for the Morley nonconforming FEM for the von Kármán equations are presented in this section, as the formulation in this article is different from that in [18, 91, 92] (see Remark 3.2.1).

Example 1

Let the computational domain be $\Omega = (0,1)^2$ and the model problem is constructed in such a way that the exact solution is known and is given by $\overline{u} = x^2y^2(1-x)^2(1-y)^2$ and $\overline{v} = \sin^2(\pi x)\sin^2(\pi y)$. Then the right-hand side load functions are computed by $f = \Delta^2\overline{u} - [\overline{u},\overline{v}]$ and $g = \Delta^2\overline{v} + \frac{1}{2}[\overline{u},\overline{u}]$.

Table 3.7: (Morley) Convergence results for the relative errors of $\overline{u}$, vK, Example 1

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
1.00000	5	7.871553	-	3.261532	-	2.394697	-
0.50000	25	2.443023	1.6880	1.319491	1.3056	1.541304	0.6357
0.25000	113	0.580661	2.0729	0.332025	1.9906	0.713766	1.1106
0.12500	481	0.156130	1.8949	0.094203	1.8174	0.367990	0.9558
0.06250	1985	0.039992	1.9650	0.024552	1.9399	0.185982	0.9845
0.03125	8065	0.010065	1.9904	0.006208	1.9836	0.093262	0.9958

As seen in the Tables 3.7-3.8, we obtain linear order of convergence in the energy norm and quadratic orders of convergence in H^1 and L^2 norm for the displacement and Airy stress functions. Note that $\omega(E_{\mathcal{D}}) = \mathcal{O}(h^{1/2})$ for Morley FEM. Thus, the numerical tests show a better convergence rate than the one given by Theorem 3.5.12 in the HDM framework.

Example 2

In this example, we consider the non-convex L-shaped domain given by $\Omega = (-1,1)^2 \setminus \big([0,1)\times(-1,0]\big)$. Choose the right-hand functions such that the exact singular solution [71] in polar

Table 3.8: (Morley) Convergence results for the relative errors of $\overline{v}$, vK, Example 1

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{v})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{v})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{v})$	Order
1.00000	5	2.042170	-	0.830880	-	0.999850	-
0.50000	25	2.055671	-0.0095	1.112735	-0.4214	1.380217	-0.4651
0.25000	113	0.474397	2.1154	0.296274	1.9091	0.680937	1.0193
0.12500	481	0.128741	1.8816	0.084071	1.8173	0.362159	0.9109
0.06250	1985	0.033048	1.9618	0.021912	1.9399	0.184550	0.9726
0.03125	8065	0.008322	1.9896	0.005542	1.9832	0.092744	0.9927

coordinates is given by

$$\overline{u} = \overline{v} = (r^2\cos^2\theta - 1)^2(r^2\sin^2\theta - 1)^2 r^{1+\gamma} g_{\gamma,\omega}(\theta),$$

where $\gamma \approx 0.5444837367$ is a non-characteristic root of $\sin^2(\gamma\omega) = \gamma^2\sin^2(\omega)$, $\omega = \frac{3\pi}{2}$, and $g_{\gamma,\omega}(\theta) = (\frac{1}{\gamma-1}\sin((\gamma-1)\omega) - \frac{1}{\gamma+1}\sin((\gamma+1)\omega))(\cos((\gamma-1)\theta) - \cos((\gamma+1)\theta)) - (\frac{1}{\gamma-1}\sin((\gamma-1)\theta) - \frac{1}{\gamma+1}\sin((\gamma+1)\theta))(\cos((\gamma-1)\omega) - \cos((\gamma+1)\omega))$.

Table 3.9: (Morley) Convergence results for the relative errors of $\overline{u}$, vK, Example 2

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{u})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{u})$	Order
1.414214	5	1.075148	-	1.557125	-	1.252892	-
0.707107	33	2.826994	-1.3947	1.985957	-0.3509	1.758240	-0.4889
0.353553	161	0.874885	1.6921	0.623930	1.6704	0.984743	0.8363
0.176777	705	0.250204	1.8060	0.181811	1.7789	0.524270	0.9094
0.088388	2945	0.071856	1.7999	0.053249	1.7716	0.273319	0.9397
0.044194	12033	0.022050	1.7044	0.017351	1.6178	0.143736	0.9272
0.022097	48641	0.007491	1.5575	0.006560	1.4033	0.077744	0.8866

Table 3.10: (Morley) Convergence results for the relative errors of $\overline{v}$, vK, Example 2

h	**nu**	$\mathsf{err}_{\mathcal{D}}(\overline{v})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{v})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\overline{v})$	Order
1.414214	5	1.076538	-	1.469632	-	1.214306	-
0.707107	33	1.910146	-0.8273	1.293881	0.1837	1.351562	-0.1545
0.353553	161	0.794724	1.2652	0.569137	1.1849	0.966468	0.4838
0.176777	705	0.229244	1.7936	0.167686	1.7630	0.527682	0.8731
0.088388	2945	0.064624	1.8267	0.047896	1.8078	0.275565	0.9373
0.044194	12033	0.019339	1.7406	0.015209	1.6550	0.144849	0.9278
0.022097	48641	0.006411	1.5929	0.005694	1.4175	0.078259	0.8882

Since Ω is non-convex, we expect only sub-optimal order of convergences in the energy, H^1 and L^2 norms. Tables 3.9-3.10 confirms these estimates numerically.

Comparing the Tables for GR and Morley methods for the vK equations (Tables 3.5-3.8), we see that for a given mesh, GR has much less number of unknowns than Morley. For a similar meshes

of mesh size $h = 0.044194$ for GR and $h = 0.03125$ for Morley, GR method has only 961 degrees of freedom whereas Morley ncFEM is with 8065 degrees of freedom. For these meshes, similar convergence rates are obtained in the energy norm, but the Morley accuracy in L^2 and H^1 norms is much better than the GR method.

Chapter 4

The gradient discretisation method for optimal control problems

In this chapter, optimal control problems governed by diffusion equations are investigated in the framework of the gradient discretisation method[1].

4.1 Introduction

This chapter deals with numerical schemes for second order distributed optimal control problems governed by diffusion equations with Dirichlet and Neumann boundary conditions (BC). When considering Neumann BC, the model has a reaction term to ensure its full coercivity. The case of pure Neumann BC, without reaction term, is covered in Chapter 5. We present basic convergence results and super-convergence results for the state, adjoint and control variables. The results cover various numerical methods, that include conforming Galerkin methods, non-conforming finite elements, and mimetic finite differences. This is achieved by using the framework of the gradient discretisation method.

The gradient discretisation method (GDM) is a generic framework for the convergence analysis of numerical methods for diffusion equations [48]. Note that only linear control problems are considered here, but the GDM has been designed to also deal with non-linear models. For non-linear state equations that are amenable to error estimates (e.g. the p-Laplace equation [48, Theorem 3.28]), an adaptation of the results presented here is conceivable.

Basic error estimates that provide a linear convergence rate for all the three variables (control, state, and adjoint) for low order schemes under standard regularity assumptions are established in this chapter. Given that the control is approximated by piecewise constant functions, the convergence rates are optimal. An improved error estimate is also proved for optimal controls, state

[1]The results of this chapter are published in *Jérôme Droniou, Neela Nataraj and Devika Shylaja. The gradient discretisation method for optimal control problems, with super-convergence for non-conforming finite elements and mixed-hybrid mimetic finite differences. SIAM J. Control Optim. 55 (6), pp. 3640-3672, 2017. DOI: 10.1137/17M1117768. URL: https://arxiv.org/abs/1608.01726.*

and adjoint variables with the help of a post-processing step. In the numerical implementation procedure, the discrete problem is solved using the primal-dual active set strategy [109].

The numerical analysis of optimal control problem governed by second order elliptic equations has been discussed in [96] for conforming FEM. To the best of our knowledge, super-convergence has not been studied for non-conforming methods in literature. In this chapter, following the ideas in [96], superconvergence results are established in the gradient discretisation framework, which covers a wide variety of numerical methods – in particular, the classical Crouzeix-Raviart finite element method and the mixed-hybrid mimetic finite difference schemes. The superconvergence for the control variable is obtained under a superconvergence assumption on the underlying numerical scheme for the state and adjoint equations. If not already known, such superconvergence can be checked for various gradient schemes by using the improved L^2 estimate of [54].

The chapter is organised as follows. Subsection 4.1.1 defines the distributed optimal control problem governed by the diffusion equation with homogeneous Dirichlet BC. Two particular cases of main results are stated in Subsection 4.1.2; these cases cover non-conforming finite element methods and mixed-hybrid mimetic finite difference schemes. In Section 4.2, the GDM is introduced, the concept of gradient discretisation (GD) is defined and the properties on the spaces and mappings that are important for the convergence analysis of the resulting GS are stated. Some classical examples of GDM are presented in Subsection 4.2.1. The basic error estimates and superconvergence results for the GDM applied to the control problems are presented in Section 4.3. In Section 4.4, the distributed and boundary optimal control problems with Neumann BC in the presence of a reaction term is presented. The GDM for Neumann BC is discussed in this section and the core properties that the GDs must satisfy to provide a proper approximation of given problem are highlighted. The results of some numerical experiments are presented in Section 4.5.

4.1.1 The optimal control problem for homogeneous Dirichlet BC

Consider the distributed optimal control problem governed by the diffusion equation defined by

$$\min_{u\in\mathcal{U}_{\rm ad}} J(u) \quad \text{subject to} \tag{4.1.1a}$$

$$-\mathrm{div}(A\nabla y(u)) = f+u \quad \text{in } \Omega, \tag{4.1.1b}$$

$$y(u) = 0 \quad \text{on } \partial\Omega, \tag{4.1.1c}$$

where $\Omega \subsetneq \mathbb{R}^d$ $(d \geq 2)$ is a bounded domain with boundary $\partial\Omega$; $y(u)$ is the state variable, and u is the control variable;

$$J(u) := \frac{1}{2}\|y(u)-\overline{y}_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u-\overline{u}_d\|^2_{L^2(\Omega)} \tag{4.1.2}$$

is the cost functional, $\alpha > 0$ is a fixed regularization parameter, $\overline{y}_d \in L^2(\Omega)$ is the desired state variable and $\overline{u}_d \in L^2(\Omega)$ is the desired control variable; $A : \Omega \to M_d(\mathbb{R})$ is a measurable, bounded

and uniformly elliptic matrix-valued function such that (for simplicity purposes) $A(\boldsymbol{x})$ is symmetric for a.e. $\boldsymbol{x} \in \Omega$; $f \in L^2(\Omega)$; $\mathcal{U}_{\rm ad} \subset L^2(\Omega)$ is the non-empty, convex and closed set of admissible controls.

It is well known that given $u \in \mathcal{U}_{\rm ad}$, there exists a unique weak solution $y(u) \in H^1_0(\Omega) := \{w \in H^1(\Omega) : w = 0 \text{ on } \partial\Omega\}$ of (4.1.1b)-(4.1.1c), that is, $y(u) \in H^1_0(\Omega)$ is such that for all $w \in H^1_0(\Omega)$,

$$a(y(u), w) = \int_\Omega (f+u) w \,\mathrm{d}\boldsymbol{x}, \tag{4.1.3}$$

where $a(z,w) = \int_\Omega A\nabla z \cdot \nabla w \,\mathrm{d}\boldsymbol{x}$. The term $y = y(u)$ is the state associated with the control u. Express the state y by means of the state-to-control mapping S in the form $y = S(u)$. Then a variational inequality can be derived that is simplified by the introduction of the co-state p. The co-state is the Lagrange multiplier associated with the state equation [109].

The convex control problem (4.1.1) has a unique weak solution $(\overline{y}, \overline{u}) \in H^1_0(\Omega) \times \mathcal{U}_{\rm ad}$. Also there exists a co-state $\overline{p} \in H^1_0(\Omega)$ such that the triplet $(\overline{y}, \overline{p}, \overline{u}) \in H^1_0(\Omega) \times H^1_0(\Omega) \times \mathcal{U}_{\rm ad}$ satisfies the Karush-Kuhn-Tucker (KKT) optimality conditions [90, Theorem 1.3]:

$$a(\overline{y}, w) = (f + \overline{u}, w) \qquad \forall\, w \in H^1_0(\Omega), \tag{4.1.4a}$$

$$a(w, \overline{p}) = (\overline{y} - \overline{y}_d, w) \qquad \forall\, w \in H^1_0(\Omega), \tag{4.1.4b}$$

$$(\overline{p} + \alpha(\overline{u} - \overline{u}_d), v - \overline{u}) \geq 0 \qquad \forall\, v \in \mathcal{U}_{\rm ad}. \tag{4.1.4c}$$

4.1.2 Two particular cases of main results

The analysis of numerical methods for (4.1.4) is based on the abstract framework of the gradient discretisation method. To give an idea of the extent of the main results, let us consider two particular schemes, based on a mesh $\mathcal{M}$ of Ω in the sense of Definition 1.4.1. Assume that $\mathcal{U}_{\rm ad} = \{v \in L^2(\Omega) : a \leq v \leq b \text{ a.e.}\}$ for some constants a, b (possibly infinite) and, to simplify the presentation, that $\overline{u}_d = 0$. Set $P_{[a,b]}(s) = \min(b, \max(a, s))$. Let $\widetilde{u}$ be a post-processed control, whose scheme-dependent definition is given below.

- nc$\mathbb{P}_1$/$\mathbb{P}_0$: $\mathcal{M}$ is a conforming triangular/tetrahedral mesh, the state and adjoint unknowns $(\overline{y}, \overline{p})$ are approximated using non-conforming $\mathbb{P}_1$ finite elements, and the control $\overline{u}$ is approximated using piecewise constant functions on $\mathcal{M}$.

  Let the post-processed continuous control be $\widetilde{u} = \overline{u}$.

- hMFD [2]: $\mathcal{M}$ is a polygonal/polyhedral mesh, the state and adjoint unknowns $(\overline{y}, \overline{p})$ are approximated using mixed-hybrid mimetic finite differences (hMFD), and the control $\overline{u}$ is approximated using piecewise constant functions on $\mathcal{M}$. The hMFD schemes form a sub-class of the hybrid mimetic mixed (HMM) methods [49, 50] presented in Section 4.2.1 below.

  Define a post-processed continuous control $\widetilde{u}$ by

$$\widetilde{u}_{|K} = P_{[a,b]}(-\alpha^{-1}\overline{p}(\overline{\boldsymbol{x}}_K)) \quad \text{ for all } K \in \mathcal{M},$$

where $\overline{\boldsymbol{x}}_K$ denotes the centroid of the cell K.

In either case, the post-processed discrete control is $\widetilde{u}_h = P_{[a,b]}(-\alpha^{-1}\overline{p}_h)$, where $\overline{p}_h$ denotes the discrete co-state. One of the consequences of the first main theorem (Theorem 4.3.6) is the following super-convergence result on the control, under standard regularity assumptions on the mesh and the data: there exists C that depends only on Ω, A, α, a, b, $\overline{u}$, and the shape regularity of $\mathcal{M}$, such that

$$\|\widetilde{u} - \widetilde{u}_h\| \leq Ch^r(1 + \|\overline{y}_d\|_{H^1(\Omega)} + \|f\|_{H^1(\Omega)} + \|\overline{u}_d\|_{H^2(\Omega)}), \tag{4.1.5}$$

where $r = 2 - \varepsilon$ (for any $\varepsilon > 0$) if $d = 2$, and $r = \frac{11}{6}$ if $d = 3$. This estimate is also valid for conforming $\mathbb{P}_1$ finite elements.

Under the additional assumption of quasi-uniformity of the mesh (that is, each cell has a measure comparable to h^d), the second main theorem (Theorem 4.3.7) shows that (4.1.5) can be improved into a full quadratic rate of convergence:

$$\|\widetilde{u} - \widetilde{u}_h\| \leq Ch^2(1 + \|\overline{y}_d\|_{H^1(\Omega)} + \|f\|_{H^1(\Omega)} + \|\overline{u}_d\|_{H^2(\Omega)}). \tag{4.1.6}$$

The quasi-uniformity assumption prevents from considering local mesh refinement, so (4.1.6) is not ensured in these cases. On the contrary, the rate (4.1.5) still holds true for locally refined meshes. Moreover, in dimension $d = 2$, the $h^{2-\varepsilon}$ rate in (4.1.5) is numerically indistinguishable from a full super-convergence h^2 rate. If $d = 3$, the $h^{\frac{11}{6}}$ rate of convergence remains very close to h^2. To compare, for $h = 10^{-6}$ (which is well below the usual mesh sizes in 3D computational tests) we have $h^2/h^{\frac{11}{6}} = 10^{-1}$.

Precise statements of the assumptions and the proofs of (4.1.5) and (4.1.6) are given in Corollary 4.3.10.

4.2 The gradient discretisation method for the control problem

The gradient discretisation method consists in writing numerical schemes, called gradient schemes (GS), by replacing in the weak formulation of the problem the continuous space and operators by discrete ones [48, 50, 60]. These discrete space and operators are given by a GD.

Definition 4.2.1 (Gradient discretisation for homogeneous Dirichlet BC)**.** *A gradient discretisation for homogeneous Dirichlet BC is given by $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}})$ such that*

- *the set of discrete unknowns (degrees of freedom) $X_{\mathcal{D},0}$ is a finite dimensional real vector space,*
- *$\Pi_{\mathcal{D}} : X_{\mathcal{D},0} \to L^2(\Omega)$ is a linear mapping that reconstructs a function from the degrees of freedom,*

- $\nabla_{\mathcal{D}} : X_{\mathcal{D},0} \to L^2(\Omega)^d$ *is a linear mapping that reconstructs a gradient from the degrees of freedom. It must be chosen such that* $\|\nabla_{\mathcal{D}} \cdot \|$ *is a norm on* $X_{\mathcal{D},0}$.

Let $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}})$ be a GD in the sense of the above definition. If $F \in L^2(\Omega)$, then the related gradient scheme for a linear elliptic problem

$$\begin{cases} -\mathrm{div}(A\nabla\psi) = F & \text{in } \Omega, \\ \psi = 0 & \text{on } \partial\Omega \end{cases} \tag{4.2.1}$$

is obtained by writing the weak formulation of (4.2.1) with the continuous spaces, function and gradient replaced with their discrete counterparts:

$$\text{Find } \psi_{\mathcal{D}} \in X_{\mathcal{D},0} \text{ such that, for all } w_{\mathcal{D}} \in X_{\mathcal{D},0},\ a_{\mathcal{D}}(\psi_{\mathcal{D}}, w_{\mathcal{D}}) = (F, \Pi_{\mathcal{D}} w_{\mathcal{D}}), \tag{4.2.2}$$

where $a_{\mathcal{D}}(\psi_{\mathcal{D}}, w_{\mathcal{D}}) = \int_\Omega A\nabla_{\mathcal{D}}\psi_{\mathcal{D}} \cdot \nabla_{\mathcal{D}} w_{\mathcal{D}}\, \mathrm{d}\boldsymbol{x}$.

The flexibility of the GDM analysis framework comes from the wide possible range of choices for $(X_{\mathcal{D},0}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}})$. Each of these choices correspond to a particular numerical scheme (see Subsection 4.2.1 for a few examples). The space $X_{\mathcal{D},0}$ represents the degrees of freedom (unknowns) of the method; a vector in $X_{\mathcal{D},0}$ gathers values for such unknowns. The operators $\Pi_{\mathcal{D}}$ and $\nabla_{\mathcal{D}}$ reconstruct, from a set of values of these unknowns, a scalar (resp. vector) function on the entire set Ω. The scalar function is supposed to play the role of the solution/test functions itself in the weak formulation of the PDE; the vector-function, reconstructed "gradient", is used in lieu of the gradients of these solution/test functions. Performing these substitutions in the weak formulation leads to a finite-dimensional system of equations (on the unknowns), which is referred to as the gradient scheme corresponding to the gradient discretisation $\mathcal{D}$.

Let $\mathcal{U}_h$ be a finite-dimensional subspace of $L^2(\Omega)$, and $\mathcal{U}_{\mathrm{ad},h} = \mathcal{U}_{\mathrm{ad}} \cap \mathcal{U}_h$. A gradient discretisation $\mathcal{D}$ being given, the corresponding GS for (4.1.4) consists in seeking $(\overline{y}_{\mathcal{D}}, \overline{p}_{\mathcal{D}}, \overline{u}_h) \in X_{\mathcal{D},0} \times X_{\mathcal{D},0} \times \mathcal{U}_{\mathrm{ad},h}$ such that

$$a_{\mathcal{D}}(\overline{y}_{\mathcal{D}}, w_{\mathcal{D}}) = (f + \overline{u}_h, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall\, w_{\mathcal{D}} \in X_{\mathcal{D},0}, \tag{4.2.3a}$$

$$a_{\mathcal{D}}(w_{\mathcal{D}}, \overline{p}_{\mathcal{D}}) = (\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \overline{y}_d, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall\, w_{\mathcal{D}} \in X_{\mathcal{D},0}, \tag{4.2.3b}$$

$$(\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} + \alpha(\overline{u}_h - \overline{u}_d), v_h - \overline{u}_h) \geq 0 \qquad \forall\, v_h \in \mathcal{U}_{\mathrm{ad},h}. \tag{4.2.3c}$$

Arguing as in [109, Theorem 2.25], it is straightforward to see that (4.2.3) is equivalent to the following minimisation problem:

$$\begin{aligned} &\min_{u_h \in \mathcal{U}_{\mathrm{ad},h}} \frac{1}{2}\|\Pi_{\mathcal{D}} y_{\mathcal{D}} - \overline{y}_d\|^2 + \frac{\alpha}{2}\|u_h - \overline{u}_d\|^2 \quad \text{subject to} \\ &y_{\mathcal{D}} \in X_{\mathcal{D},0} \text{ and, for all } w_{\mathcal{D}} \in X_{\mathcal{D},0},\ a_{\mathcal{D}}(y_{\mathcal{D}}, w_{\mathcal{D}}) = (f + u_h, \Pi_{\mathcal{D}} w_{\mathcal{D}}). \end{aligned} \tag{4.2.4}$$

Existence and uniqueness of a solution to (4.2.4), and thus to (4.2.3), follows from standard variational theorems.

4.2.1 Examples of gradient discretisations

A few examples based on known numerical methods are briefly presented in this section. See [48] for a detailed analysis of these methods, and more examples of gradient discretisations. As demonstrated by these examples, the GDM cover a wide range of different numerical methods. This means that the analysis carried out in the GDM framework for the control problem in (4.1.1) readily applies to all these methods. In particular, this makes the control problem accessible to numerical schemes not usually considered but relevant to diffusion models, such as schemes applicables on generic meshes (not just triangular/quadrangular meshes) as encountered for example in reservoir engineering applications.

Consider a mesh $\mathcal{M}$ of Ω in the sense of Definition 1.4.1.

Conforming $\mathbb{P}_1$ finite elements. The simplest gradient discretisation is perhaps obtained by considering conforming $\mathbb{P}_1$ finite elements. The mesh is made of triangles (in 2D) or tetrahedra (in 3D), with no hanging nodes. Each $v_\mathcal{D} \in X_{\mathcal{D},0}$ is a vector of values at the internal vertices of the mesh (the standard unknowns of conforming $\mathbb{P}_1$ finite elements). $\Pi_\mathcal{D} v_\mathcal{D}$ is the continuous piecewise linear function on the mesh which takes these values at the vertices, and $\nabla_\mathcal{D} v_\mathcal{D} = \nabla(\Pi_\mathcal{D} v_\mathcal{D})$. Then (4.2.2) is the standard $\mathbb{P}_1$ finite element scheme for (4.2.1).

Non-conforming $\mathbb{P}_1$ finite elements. As above, the mesh is made of conforming triangles or tetrahedra. Each $v_\mathcal{D} \in X_{\mathcal{D},0}$ is a vector of values at the centers of mass of the internal edges/faces, $\Pi_\mathcal{D} v_\mathcal{D}$ is the piecewise linear function on the mesh which takes these values at these centers of mass, and $\nabla_\mathcal{D} v_\mathcal{D} = \nabla_\mathcal{M}(\Pi_\mathcal{D} v_\mathcal{D})$ is the broken gradient of $\Pi_\mathcal{D} v_\mathcal{D}$. In that case, (4.2.2) gives the non-conforming $\mathbb{P}_1$ finite element approximation of (4.2.1).

Mass-lumped non-conforming $\mathbb{P}_1$ finite elements. Still considering a conforming triangular/tetrahedral mesh, the space $X_{\mathcal{D},0}$ and gradient reconstruction $\nabla_\mathcal{D}$ are identical to those of the non-conforming $\mathbb{P}_1$ finite elements described above, but the function reconstruction is modified to be piecewise constant. For each edge/face σ of the mesh, consider the diamond D_σ around σ constructed from the edge/face and the one or two cell centers on each side (see Figure 4.1). Then, for $v = (v_\sigma)_{\sigma \in \mathcal{F}_{\rm int}} \in X_{\mathcal{D},0}$, the reconstructed function $\Pi_\mathcal{D} v$ is the piecewise constant function on the diamonds, equal to v_σ on D_σ for all internal edge/face σ.

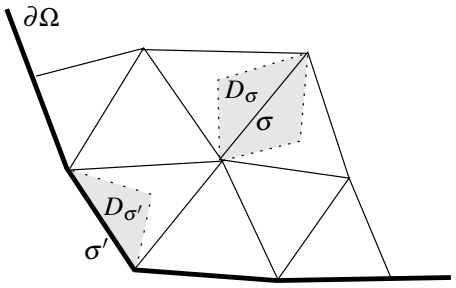

Figure 4.1: Diamonds for the definition of the mass-lumped non-conforming $\mathbb{P}_1$ finite elements

Hybrid mixed mimetic method (HMM). Consider a generic mesh $\mathcal{M}$ (not necessarily triangular/tetrahedral) with one point $\boldsymbol{x}_K$ chosen in each cell $K \in \mathcal{M}$ such that K is strictly star-shaped with respect to $\boldsymbol{x}_K$; see Figure 4.2 for some notations. A vector $v \in X_{\mathcal{D},0}$ is made of cell $(v_K)_K$ and face $(v_\sigma)_{\sigma\in\mathcal{F}_{\rm int}}$ values, and the operator $\Pi_\mathcal{D}$ reconstructs a piecewise constant function from the cell values: for any cell K, $\Pi_\mathcal{D} v = v_K$ on K. The gradient reconstruction $\nabla_\mathcal{D}$ is built in two pieces: a consistent gradient $\overline{\nabla}_K$ constant over the cell and stabilisation terms constant over the half-diamonds $D_{K,\sigma}$ (and akin to the remainders of first-order Taylor expansions between the cell and face values). For any cell K and any face σ of K, set

$$\nabla_\mathcal{D} v = \overline{\nabla}_K v + \frac{\sqrt{d}}{d_{K,\sigma}} \Big(v_\sigma - v_K - \overline{\nabla}_K v \cdot (\overline{\boldsymbol{x}}_\sigma - \boldsymbol{x}_K) \Big) n_{K,\sigma} \quad \text{on } K,$$

where $d_{K,\sigma}$ is the orthogonal distance between $\boldsymbol{x}_K$ and σ, $\overline{\boldsymbol{x}}_\sigma$ is the center of mass of σ, $n_{K,\sigma}$ is the outer normal to K on σ and, denoting by $\mathcal{F}_K$ the set of faces of K,

$$\overline{\nabla}_K v = \frac{1}{|K|} \sum_{\sigma\in\mathcal{F}_K} |\sigma| v_\sigma n_{K,\sigma}.$$

Once used in the gradient scheme (4.2.2), this HMM gradient discretisation gives rise to a numerical method that can be applied on very general meshes (including with hanging nodes, non-convex cells, etc.). This scheme can also be re-interpreted as a finite volume method [48, Section 13.3].

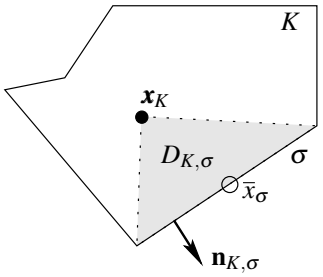

Figure 4.2: Notations for the construction of the HMM gradient discretisation

4.2.2 Results on the GDM for elliptic PDEs

We recall the basic notions and known results on the GDM for elliptic PDEs [48].
The accuracy of a GS (4.2.2) is measured by three quantities. The first one, which ensures the *coercivity* of the method, controls the norm of $\Pi_\mathcal{D}$.

$$C_\mathcal{D} := \max_{w\in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\Pi_\mathcal{D} w\|}{\|\nabla_\mathcal{D} w\|}. \tag{4.2.5}$$

The second measure involves an estimate of the interpolation error, called the GD-*consistency* (or consistency, for short) in the framework of the GDM. It corresponds to the interpolation error in the finite element nomenclature.

$$\forall \varphi \in H_0^1(\Omega),\ S_{\mathcal{D}}(\varphi) = \min_{w \in X_{\mathcal{D},0}} \left(\|\Pi_{\mathcal{D}} w - \varphi\| + \|\nabla_{\mathcal{D}} w - \nabla \varphi\| \right). \tag{4.2.6}$$

Finally, the *limit-conformity* of a GD is measured by defining

$$\forall \boldsymbol{\varphi} \in H_{\mathrm{div}}(\Omega),\ W_{\mathcal{D}}(\boldsymbol{\varphi}) = \max_{w \in X_{\mathcal{D},0} \setminus \{0\}} \frac{1}{\|\nabla_{\mathcal{D}} w\|} \left| \widetilde{W}_{\mathcal{D}}(\boldsymbol{\varphi}, w) \right|, \tag{4.2.7}$$

where $H_{\mathrm{div}}(\Omega) = \{\boldsymbol{\varphi} \in L^2(\Omega)^d : \mathrm{div}\boldsymbol{\varphi} \in L^2(\Omega)\}$ and

$$\widetilde{W}_{\mathcal{D}}(\boldsymbol{\varphi}, w) = \int_\Omega (\Pi_{\mathcal{D}} w \, \mathrm{div}\boldsymbol{\varphi} + \nabla_{\mathcal{D}} w \cdot \boldsymbol{\varphi}) \, \mathrm{d}\boldsymbol{x}. \tag{4.2.8}$$

Recall that in Chapter 2, the notation $X \lesssim Y$ means $X \leq CY$ for some C depending only on Ω and an upper bound of $C_{\mathcal{D}}^{\mathcal{B}}$ defined by 2.4.1. Here,

$$\begin{aligned} &X \lesssim Y \text{ means that } X \leq CY \text{ for some } C \text{ depending} \\ &\text{only on } \Omega, A \text{ and an upper bound of } C_{\mathcal{D}} \text{ defined by (4.2.5).} \end{aligned} \tag{4.2.9}$$

The following basic error estimate on GSs is standard, see [48, Theorem 3.2].

Theorem 4.2.2. *Let $\mathcal{D}$ be a GD in the sense of Definition 4.2.1, ψ be the solution to* (4.2.1)*, and $\psi_{\mathcal{D}}$ be the solution to* (4.2.2)*. Then*

$$\|\Pi_{\mathcal{D}} \psi_{\mathcal{D}} - \psi\| + \|\nabla_{\mathcal{D}} \psi_{\mathcal{D}} - \nabla \psi\| \lesssim \mathrm{WS}_{\mathcal{D}}(\psi), \tag{4.2.10}$$

where

$$\mathrm{WS}_{\mathcal{D}}(\psi) = W_{\mathcal{D}}(A\nabla\psi) + S_{\mathcal{D}}(\psi) \tag{4.2.11}$$

$S_{\mathcal{D}}$ is defined by (4.2.6) *and $W_{\mathcal{D}}$ is defined by* (4.2.7)*.*

Remark 4.2.3 (Rates of convergence for the PDE)**.** *For all classical first-order methods based on meshes with mesh parameter "h", $\mathcal{O}(h)$ estimates can be obtained for $W_{\mathcal{D}}(A\nabla\psi)$ and $S_{\mathcal{D}}(\psi)$, if A is Lipschitz continuous and $\psi \in H^2(\Omega)$ (see [48, Chapter 8]). Theorem 4.2.2 then gives a linear rate of convergence for these methods.*

4.3 Basic error estimate and super-convergence

In this section, the main contributions are presented and the assumptions are discussed in details. The basic error estimates for control, state and adjoint variables are established in the HDM framework. Superconvergence results are proved using a post-processing step. Let us start with a straightforward stability result, which will be useful for the analysis.

Proposition 4.3.1 (Stability of gradient schemes)**.** *Let $\underline{a}$ be a coercivity constant of A. If $\psi_{\mathcal{D}}$ is the solution to the gradient scheme* (4.2.2)*, then*

$$\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}}\| \le \frac{C_{\mathcal{D}}}{\underline{a}}\|F\| \quad \text{and} \quad \|\Pi_{\mathcal{D}}\psi_{\mathcal{D}}\| \le \frac{C_{\mathcal{D}}^2}{\underline{a}}\|F\|. \tag{4.3.1}$$

Proof. Choose $w_{\mathcal{D}} = \psi_{\mathcal{D}}$ in (4.2.2) and use the definition of $C_{\mathcal{D}}$ to write

$$\underline{a}\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}}\|^2 \le \|F\|\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}}\| \le C_{\mathcal{D}}\|F\|\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}}\|.$$

The proof of first inequality in (4.3.1) is complete. The second estimate follows from the definition of $C_{\mathcal{D}}$. □

4.3.1 Basic error estimate for the GDM for the control problem

To state the error estimates, let $\mathrm{Pr}_h : L^2(\Omega) \to \mathcal{U}_h$ be the L^2 orthogonal projector on $\mathcal{U}_h$ for the standard scalar product.

Theorem 4.3.2 (Control estimate)**.** *Let $\mathcal{D}$ be a GD, $(\overline{y}, \overline{p}, \overline{u})$ be the solution to* (4.1.4) *and $(\overline{y}_{\mathcal{D}}, \overline{p}_{\mathcal{D}}, \overline{u}_h)$ be the solution to* (4.2.3)*. Assume that*

$$\mathrm{Pr}_h(\mathcal{U}_{\mathrm{ad}}) \subset \mathcal{U}_{\mathrm{ad},h}. \tag{4.3.2}$$

Then,

$$\begin{aligned}\sqrt{\alpha}\|\overline{u} - \overline{u}_h\| \lesssim \sqrt{\alpha}\|\alpha^{-1}\overline{p} - \mathrm{Pr}_h(\alpha^{-1}\overline{p})\| + (\sqrt{\alpha}+1)\|\overline{u} - \mathrm{Pr}_h\overline{u}\| \\ + \sqrt{\alpha}\|\overline{u}_d - \mathrm{Pr}_h\overline{u}_d\| + \frac{1}{\sqrt{\alpha}}\mathrm{WS}_{\mathcal{D}}(\overline{p}) + \mathrm{WS}_{\mathcal{D}}(\overline{y}).\end{aligned} \tag{4.3.3}$$

The technique used here is an adaptation of classical ideas used (e.g., for the error-analysis of finite-element based discretisations) to the gradient discretisation method.
Define the scaled norm $|||\cdot|||$ and projection error E_h by

$$\forall W \in L^2(\Omega),\ |||W||| = \sqrt{\alpha}\|W\| \text{ and } E_h(W) = |||W - \mathrm{Pr}_h W|||. \tag{4.3.4}$$

To establish the error estimates, we need the following auxiliary discrete problem:

$$\begin{aligned}&\text{seek } (y_{\mathcal{D}}(\overline{u}), p_{\mathcal{D}}(\overline{u})) \in X_{\mathcal{D},0} \times X_{\mathcal{D},0} \text{ such that} \\ &a_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}), w_{\mathcal{D}}) = (f + \overline{u}, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall\, w_{\mathcal{D}} \in X_{\mathcal{D},0}, \qquad (4.3.5a) \\ &a_{\mathcal{D}}(w_{\mathcal{D}}, p_{\mathcal{D}}(\overline{u})) = (\overline{y} - \overline{y}_d, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall\, w_{\mathcal{D}} \in X_{\mathcal{D},0}. \qquad (4.3.5b)\end{aligned}$$

Proof of Theorem 4.3.2. Let $P_{\mathcal{D},\alpha}(\overline{u}) = \alpha^{-1}\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u})$, $\overline{P}_{\mathcal{D},\alpha} = \alpha^{-1}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}$, and $\overline{P}_\alpha = \alpha^{-1}\overline{p}$. Since $\overline{u}_h \in \mathcal{U}_{\mathrm{ad},h} \subset \mathcal{U}_{\mathrm{ad}}$, from the optimality condition (4.1.4c),

$$-\alpha(\overline{P}_\alpha + \overline{u} - \overline{u}_d, \overline{u} - \overline{u}_h) \ge 0. \tag{4.3.6}$$

By (4.3.2), $\mathrm{Pr}_h\overline{u} \in \mathcal{U}_{\mathrm{ad},h}$ and therefore a use of the discrete optimality condition (see (4.2.3c)) yields

$$\begin{aligned}\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h-\overline{u}_d,\overline{u}-\overline{u}_h) &= \alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h-\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u})\\ &\quad+\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h-\overline{u}_d,\mathrm{Pr}_h\overline{u}-\overline{u}_h)\\ &\geq \alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h-\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u}). \end{aligned} \tag{4.3.7}$$

An addition of (4.3.6) and (4.3.7) yields

$$\begin{aligned}|||\overline{u}-\overline{u}_h|||^2 \leq & -\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h-\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u})+\alpha(\overline{P}_{\mathcal{D},\alpha}-\overline{P}_\alpha,\overline{u}-\overline{u}_h)\\ = & -\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h-\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u})-\alpha(\overline{P}_\alpha-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h)\\ & +\alpha(\overline{P}_{\mathcal{D},\alpha}-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h). \end{aligned} \tag{4.3.8}$$

The first term in the right-hand side of (4.3.8) is recast now. By orthogonality property of Pr_h, $(\overline{u}_h-\mathrm{Pr}_h\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u})=0$ and $(\mathrm{Pr}_h\overline{P}_\alpha,\overline{u}-\mathrm{Pr}_h\overline{u})=0$. Therefore,

$$\begin{aligned}&-\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h-\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u})\\ &= -\alpha(\overline{P}_\alpha,\overline{u}-\mathrm{Pr}_h\overline{u})+\alpha(\overline{P}_\alpha-\overline{P}_{\mathcal{D},\alpha},\overline{u}-\mathrm{Pr}_h\overline{u})-\alpha(\mathrm{Pr}_h\overline{u}_d-\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u})\\ &= -\alpha(\overline{P}_\alpha-\mathrm{Pr}_h\overline{P}_\alpha,\overline{u}-\mathrm{Pr}_h\overline{u})+\alpha(\overline{P}_\alpha-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\mathrm{Pr}_h\overline{u})\\ &\quad+\alpha(P_{\mathcal{D},\alpha}(\overline{u})-\overline{P}_{\mathcal{D},\alpha},\overline{u}-\mathrm{Pr}_h\overline{u})+\alpha(\overline{u}_d-\mathrm{Pr}_h\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u}). \end{aligned} \tag{4.3.9}$$

Let us turn to the third term in the right-hand side of (4.3.8). From (4.2.3b) and (4.3.5b), for all $w_{\mathcal{D}} \in X_{\mathcal{D},0}$,

$$a_{\mathcal{D}}(w_{\mathcal{D}},\overline{p}_{\mathcal{D}}-p_{\mathcal{D}}(\overline{u}))=(\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\overline{y},\Pi_{\mathcal{D}}w_{\mathcal{D}}). \tag{4.3.10}$$

Also, from (4.2.3a) and (4.3.5a),

$$a_{\mathcal{D}}(\overline{y}_{\mathcal{D}}-y_{\mathcal{D}}(\overline{u}),w_{\mathcal{D}})=(\overline{u}_h-\overline{u},\Pi_{\mathcal{D}}w_{\mathcal{D}}). \tag{4.3.11}$$

A use of symmetry of $a_{\mathcal{D}}$, a choice of $w_{\mathcal{D}}=\overline{y}_{\mathcal{D}}-y_{\mathcal{D}}(\overline{u})$ in (4.3.10) and $w_{\mathcal{D}}=\overline{p}_{\mathcal{D}}-p_{\mathcal{D}}(\overline{u})$ in (4.3.11) gives an expression for the third term on the righthand side of (4.3.8) as

$$\begin{aligned}\alpha(\overline{P}_{\mathcal{D},\alpha}-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h) &= -(\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\overline{y},\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}))\\ &= (\overline{y}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}),\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}))\\ &\quad-\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|^2. \end{aligned} \tag{4.3.12}$$

A substitution of (4.3.9) and (4.3.12) in (4.3.8) yields

$$\begin{aligned}&|||\overline{u}-\overline{u}_h|||^2+\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|^2\\ &\quad\leq -\alpha(\overline{P}_\alpha-\mathrm{Pr}_h\overline{P}_\alpha,\overline{u}-\mathrm{Pr}_h\overline{u})+\alpha(\overline{P}_\alpha-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\mathrm{Pr}_h\overline{u})\\ &\qquad+\alpha(P_{\mathcal{D},\alpha}(\overline{u})-\overline{P}_{\mathcal{D},\alpha},\overline{u}-\mathrm{Pr}_h\overline{u})+\alpha(\overline{u}_d-\mathrm{Pr}_h\overline{u}_d,\overline{u}-\mathrm{Pr}_h\overline{u})\\ &\qquad-\alpha(\overline{P}_\alpha-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h)+(\overline{y}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}),\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}))\\ &\quad=: T_1+T_2+T_3+T_4+T_5+T_6. \end{aligned} \tag{4.3.13}$$

The terms T_i, $i = 1, \dots, 6$ are estimated now. By the Cauchy–Schwarz inequality,

$$T_1 \leq E_h(\overline{P}_\alpha) E_h(\overline{u}). \tag{4.3.14}$$

Equation (4.3.5b) shows that $p_{\mathcal{D}}(\overline{u})$ is the solution of the GS corresponding to the adjoint problem (4.1.4b), whose solution is $\overline{p}$. Therefore, by Theorem 4.2.2,

$$\left|\left|\left| \overline{P}_\alpha - P_{\mathcal{D},\alpha}(\overline{u}) \right|\right|\right| = \frac{1}{\sqrt{\alpha}} \|\overline{p} - \Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u})\| \lesssim \frac{1}{\sqrt{\alpha}} \mathrm{WS}_{\mathcal{D}}(\overline{p}). \tag{4.3.15}$$

Hence, a use of the Cauchy–Schwarz inequality leads to

$$T_2 \lesssim \frac{1}{\sqrt{\alpha}} E_h(\overline{u}) \mathrm{WS}_{\mathcal{D}}(\overline{p}). \tag{4.3.16}$$

By writing the difference of (4.3.5b) and (4.2.3b) we see that $p_{\mathcal{D}}(\overline{u}) - \overline{p}_{\mathcal{D}}$ is the solution to the GS (4.2.2) with source term $F = \overline{y} - \Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}}$. Hence, use Proposition 4.3.1 to deduce

$$\begin{aligned}
\left|\left|\left| P_{\mathcal{D},\alpha}(\overline{u}) - \overline{P}_{\mathcal{D},\alpha} \right|\right|\right| &= \frac{1}{\sqrt{\alpha}} \|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} \overline{p}_{\mathcal{D}}\| \\
&\lesssim \frac{1}{\sqrt{\alpha}} \|\overline{y} - \Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}}\| \\
&\lesssim \frac{1}{\sqrt{\alpha}} \|\overline{y} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| + \frac{1}{\sqrt{\alpha}} \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}}\|.
\end{aligned}$$

A use of Theorem 4.2.2 with $\psi = \overline{y}$ to bound the first term in the above expression yields, by Young's inequality,

$$T_3 \leq \frac{C_1}{\sqrt{\alpha}} E_h(\overline{u}) \mathrm{WS}_{\mathcal{D}}(\overline{y}) + \frac{C_1}{\alpha} E_h(\overline{u})^2 + \frac{1}{4} \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}}\|^2, \tag{4.3.17}$$

where C_1 only depends on Ω, A and an upper bound of $C_{\mathcal{D}}$.
A use of the Cauchy–Schwarz inequality and the Young inequality leads to

$$T_4 \leq E_h(\overline{u}) E_h(\overline{u}_d) \leq \frac{1}{2} E_h(\overline{u})^2 + \frac{1}{2} E_h(\overline{u}_d)^2. \tag{4.3.18}$$

The term T_5 can be estimated using (4.3.15) and Young's inequality:

$$T_5 \leq \frac{1}{2} |||\overline{u} - \overline{u}_h|||^2 + \frac{C_2}{\alpha} \mathrm{WS}_{\mathcal{D}}(\overline{p})^2, \tag{4.3.19}$$

where C_2 only depends on Ω, A and an upper bound of $C_{\mathcal{D}}$. Finally, to estimate T_6, by Theorem 4.2.2 with $\psi = \overline{y}$,

$$\begin{aligned}
T_6 &\leq C_3 \mathrm{WS}_{\mathcal{D}}(\overline{y}) \|\Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| \\
&\leq C_3^2 \mathrm{WS}_{\mathcal{D}}(\overline{y})^2 + \frac{1}{4} \|\Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\|^2,
\end{aligned} \tag{4.3.20}$$

with C_3 only depending on Ω, A and an upper bound of $C_{\mathcal{D}}$.
We then plug (4.3.14), (4.3.16), (4.3.17), (4.3.18), (4.3.19) and (4.3.20) into (4.3.13). A use of Young's inequality and $\sqrt{\sum_i a_i^2} \leq \sum_i a_i$ concludes the proof. □

Remark 4.3.3. *The proof of Theorem 4.3.2 follows from the continuous and discrete KKT optimality conditions given by* (4.1.4c) *and* (4.2.3c)*, Theorem 4.2.2 and Proposition 4.3.1. In particular, the result holds true for any bilinear form $a(\cdot,\cdot)$ defined on a subspace X of $L^2(\Omega)$ and $a_{\mathcal{D}}(\cdot,\cdot)$ defined on a discrete space $X_{\mathcal{D},0}$ (along with the continuous and discrete KKT optimality conditions), provided the following holds. If ψ is the solution to $a(\psi,\phi) = (g,\phi)$ for all $\phi \in X$ and $\psi_{\mathcal{D}}$ is the corresponding solution to $a_{\mathcal{D}}(\psi_{\mathcal{D}},\phi_{\mathcal{D}}) = (f,\Pi_{\mathcal{D}}\phi_{\mathcal{D}})$ for all $\phi_{\mathcal{D}} \in X_{\mathcal{D},0}$, then the following stability and error estimate hold true:*

$$\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}}\| \le C\|g\| \text{ and } \|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - \psi\| \le C\mathrm{WS}_{\mathcal{D}}(\psi).$$

Proposition 4.3.4 (State and adjoint error estimates)**.** *Let $\mathcal{D}$ be a GD, $(\overline{y},\overline{p},\overline{u})$ be the solution to* (4.1.4) *and $(\overline{y}_{\mathcal{D}},\overline{p}_{\mathcal{D}},\overline{u}_h)$ be the solution to* (4.2.3)*. Then the following error estimates hold:*

$$\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \overline{y}\| + \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla\overline{y}\| \lesssim \|\overline{u} - \overline{u}_h\| + \mathrm{WS}_{\mathcal{D}}(\overline{y}), \tag{4.3.21}$$

$$\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \overline{p}\| + \|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \nabla\overline{p}\| \lesssim \|\overline{u} - \overline{u}_h\| + \mathrm{WS}_{\mathcal{D}}(\overline{y}) + \mathrm{WS}_{\mathcal{D}}(\overline{p}). \tag{4.3.22}$$

Proof. A use of the triangle inequality twice leads to

$$\begin{aligned}\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \overline{y}\| + \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla\overline{y}\| \le{}& \|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\| + \|\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}) - \overline{y}\| \\ &+ \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\| + \|\nabla_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}) - \nabla\overline{y}\|.\end{aligned}$$

The second and last terms on the right hand side of the above inequality are estimated using Theorem 4.2.2 as

$$\|\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}) - \overline{y}\| + \|\nabla_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}) - \nabla\overline{y}\| \lesssim \mathrm{WS}_{\mathcal{D}}(\overline{y}).$$

Subtract (4.2.3a) and (4.3.5a), and use the stability property of GSs (Proposition 4.3.1) to obtain

$$\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\| + \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\| \lesssim \|\overline{u} - \overline{u}_h\|.$$

A combination of the above two results yields the error estimates (4.3.21) for the state variable. The error estimate for the adjoint variable can be obtained similarly. □

Remark 4.3.5 (Rates of convergence for the control problem)**.** *Owing to Remark 4.2.3, under sufficient smoothness assumption on $\overline{u}_d$, if A is Lipschitz continuous and $(\overline{y},\overline{p},\overline{u}) \in H^2(\Omega)^2 \times H^1(\Omega)$ then* (4.3.3)*,* (4.3.21) *and* (4.3.22) *give linear rates of convergence for all classical first-order methods.*

4.3.2 Super-convergence for post-processed controls

We consider here the case $d \le 3$, and the standard situation where admissible controls are those bounded above and below by appropriate constants a and b, that is

$$\mathcal{U}_{\mathrm{ad}} = \{u \in L^2(\Omega) : a \le u \le b \text{ a.e.}\}. \tag{4.3.23}$$

Consider a mesh $\mathcal{M}$ of Ω, that is, a finite partition of Ω into polygonal/polyhedral cells (Definition 1.4.1) such that each cell $K \in \mathcal{M}$ is star-shaped with respect to its centroid $\overline{\boldsymbol{x}}_K$. The discrete space $\mathcal{U}_h$ is then defined as the space of piecewise constant functions on this partition:

$$\mathcal{U}_h = \{v : \Omega \to \mathbb{R} : \forall K \in \mathcal{M}\,,\ v_{|K} \text{ is a constant}\}. \tag{4.3.24}$$

These choices (4.3.23) and (4.3.24) of $\mathcal{U}_{\rm ad}$ and $\mathcal{U}_h$ satisfy (4.3.2). Owing to Remark 4.3.5, for low-order methods such as conforming and non-conforming FEMs or MFD schemes, under standard regularity assumptions the estimate (4.3.3) provides an $\mathcal{O}(h)$ convergence rate on $\|\overline{u} - \overline{u}_h\|$. Given that $\overline{u}_h$ is piecewise constant, this is optimal. However, using post-processed controls and following the ideas of [96], a super-convergence result for the control can be obtained.

The projection operators $\boldsymbol{\mathcal{P}}_{\mathcal{M}} : L^1(\Omega) \to \mathcal{U}_h$ (orthogonal projection on piecewise constant functions on $\mathcal{M}$) and $P_{[a,b]} : \mathbb{R} \to [a,b]$ are defined as

$$\forall v \in L^1(\Omega)\,,\ \forall K \in \mathcal{M}\,, \quad (\boldsymbol{\mathcal{P}}_{\mathcal{M}} v)_{|K} := \fint_K v \,\mathrm{d}\boldsymbol{x}$$

and

$$\forall s \in \mathbb{R}\,, \quad P_{[a,b]}(s) := \min(b, \max(a,s)).$$

To prove the superconvergence result, the following assumptions are made which are discussed, along with the post-processing, in Section 4.3.2.

(A1) [*Approximation and interpolation errors*] For each $w \in H^2(\Omega)$, there exists $w_{\mathcal{M}} \in L^2(\Omega)$ such that:
i) If $w \in H^2(\Omega) \cap H^1_0(\Omega)$ solves $-\mathrm{div}(A\nabla w) = g \in H^1(\Omega)$, and $w_{\mathcal{D}}$ is the solution to the corresponding GS, then

$$\|\Pi_{\mathcal{D}} w_{\mathcal{D}} - w_{\mathcal{M}}\| \lesssim h^2 \|g\|_{H^1(\Omega)}. \tag{4.3.25}$$

ii) For any $w \in H^2(\Omega)$, it holds

$$\forall v \in X_{\mathcal{D},0}\,,\ \left|(w - w_{\mathcal{M}}, \Pi_{\mathcal{D}} v_{\mathcal{D}})\right| \lesssim h^2 \|w\|_{H^2(\Omega)} \|\Pi_{\mathcal{D}} v_{\mathcal{D}}\| \tag{4.3.26}$$

and

$$\|\boldsymbol{\mathcal{P}}_{\mathcal{M}}(w - w_{\mathcal{M}})\| \lesssim h^2 \|w\|_{H^2(\Omega)}. \tag{4.3.27}$$

(A2) The estimate $\|\Pi_{\mathcal{D}} v_{\mathcal{D}} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}} v_{\mathcal{D}})\| \lesssim h \|\nabla_{\mathcal{D}} v_{\mathcal{D}}\|$ holds for any $v_{\mathcal{D}} \in X_{\mathcal{D},0}$.

(A3) [*Discrete Sobolev imbedding*] For all $v \in X_{\mathcal{D},0}$, it holds

$$\|\Pi_{\mathcal{D}} v_{\mathcal{D}}\|_{L^{2^*}(\Omega)} \lesssim \|\nabla_{\mathcal{D}} v_{\mathcal{D}}\|,$$

where 2^* is a Sobolev exponent of 2, that is, $2^* \in [2, \infty)$ if $d = 2$, and $2^* = \frac{2d}{d-2}$ if $d \geq 3$.

Let

$$\mathcal{M}_2 = \{K \in \mathcal{M} : \overline{u} = a \text{ a.e. on } K, \text{ or } \overline{u} = b \text{ a.e. on } K, \text{ or } a < \overline{u} < b \text{ a.e. on } K\}$$

be the set of fully active or fully inactive cells, and $\mathcal{M}_1 = \mathcal{M} \setminus \mathcal{M}_2$ be the set of cells where $\overline{u}$ takes on the value a (resp. b) as well as values greater than a (resp. lower than b). For $i = 1, 2$, let $\Omega_{i,\mathcal{M}} = \mathrm{int}(\cup_{K \in \mathcal{M}_i} \overline{K})$. The space $W^{1,\infty}(\mathcal{M}_1)$ is the usual broken Sobolev space, endowed with its broken norm. The last assumption is:

(A4) $|\Omega_{1,\mathcal{M}}| \lesssim h$ and $\overline{u}_{|\Omega_{1,\mathcal{M}}} \in W^{1,\infty}(\mathcal{M}_1)$, where $|\cdot|$ denotes the Lebesgue measure in $\mathbb{R}^d$.

From (4.1.4c) and (4.2.3c), following the reasoning in [109, Theorem 2.28], the following pointwise relations can be obtained: for a.e. $x \in \Omega$,

$$\begin{aligned} \overline{u}(\boldsymbol{x}) &= P_{[a,b]}\left(\overline{u}_d(\boldsymbol{x}) - \frac{1}{\alpha}\overline{p}(\boldsymbol{x})\right), \\ \overline{u}_h(\boldsymbol{x}) &= P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\left(\overline{u}_d(\boldsymbol{x}) - \frac{1}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}(\boldsymbol{x})\right)\right). \end{aligned} \tag{4.3.28}$$

Assuming $\overline{p} \in H^2(\Omega)$ (see Theorem 4.3.6) and letting $\overline{p}_{\mathcal{M}}$ be defined as in **(A1)**, the post-processed continuous and discrete controls are then defined by

$$\begin{aligned} \widetilde{u}(\boldsymbol{x}) &= P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d(\boldsymbol{x}) - \frac{1}{\alpha}\overline{p}_{\mathcal{M}}(\boldsymbol{x})\right), \\ \widetilde{u}_h(\boldsymbol{x}) &= P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d(\boldsymbol{x}) - \frac{1}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}(\boldsymbol{x})\right). \end{aligned} \tag{4.3.29}$$

Recall the mesh regularity assumption given by (1.4.1). In this section, the following extension of the notation (4.2.9) is used:

$$\begin{aligned} &X \lesssim_\eta Y \text{ means that } X \le CY \text{ for some } C \text{ depending} \\ &\text{only on } \Omega, A, \text{ an upper bound of } C_{\mathcal{D}}, \text{ and } \eta. \end{aligned}$$

Discussion on (A1)–(A4) and post-processings

To discuss **(A1)**, **(A2)** and the post-processing choices (4.3.29), let us consider two situations depending on the nature of $\Pi_{\mathcal{D}}$. This nature drives the choices of $w_{\mathcal{M}}$, to ensure that the super-convergence result (4.3.25) holds.

$\Pi_{\mathcal{D}}$ IS A PIECEWISE LINEAR RECONSTRUCTION. Consider here the case where $\Pi_{\mathcal{D}} v_{\mathcal{D}}$ is piecewise linear on $\mathcal{M}$ for all $v_{\mathcal{D}} \in X_{\mathcal{D},0}$. Then a super-convergence result (4.3.25) usually holds with $w_{\mathcal{M}} = w$ (and even $\|g\|$ instead of $\|g\|_{H^1(\Omega)}$). This is for example well-known for conforming and non-conforming $\mathbb{P}_1$ FEM. In that case, (4.3.26) and (4.3.27) are trivially satisfied.
Assumption **(A2)** then follows from a simple Taylor expansion if $\nabla_{\mathcal{D}} v_{\mathcal{D}}$ is the classical broken gradient (i.e. the gradient of $\Pi_{\mathcal{D}} v_{\mathcal{D}}$ in each cell). This is again the case for conforming and non-conforming $\mathbb{P}_1$ FE.
The post-processing (4.3.29) of $\overline{u}$ then solely consists in projecting $\overline{u}_d$ on piecewise constant functions. In particular, if $\overline{u}_d$ is already piecewise constant on the mesh, then $\widetilde{u} = \overline{u}$.

$\Pi_{\mathcal{D}}$ IS A PIECEWISE CONSTANT RECONSTRUCTION. Consider $\Pi_{\mathcal{D}} v_{\mathcal{D}}$ as piecewise constants on $\mathcal{M}$ for all $v_{\mathcal{D}} \in X_{\mathcal{D},0}$. Then the super-convergence (4.3.25) requires to project the exact solution on piecewise constant functions on the mesh. This is usually done by setting $w_{\mathcal{M}}(\boldsymbol{x}) = w(\overline{\boldsymbol{x}}_K)$ for all $\boldsymbol{x} \in K$ and all $K \in \mathcal{M}$. This super-convergence result is well-known for hMFD and nodal MFD schemes (see [9, 54]).
In that case, Property (4.3.27) follows (with $\lesssim$ replaced with $\lesssim_\eta$) from the classical approximation result (4.3.33). The estimate (4.3.27) along with the orthogonality property of $\boldsymbol{\mathcal{P}}_{\mathcal{M}}$ and (4.3.26) lead to

$$\begin{aligned} |(w - w_{\mathcal{M}}, \Pi_{\mathcal{D}} v_{\mathcal{D}})| = |(\boldsymbol{\mathcal{P}}_{\mathcal{M}}(w - w_{\mathcal{M}}), \Pi_{\mathcal{D}} v_{\mathcal{D}})| &\leq \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}(w - w_{\mathcal{M}})\| \|\Pi_{\mathcal{D}} v_{\mathcal{D}}\| \\ &\lesssim_\eta h^2 \|w\|_{H^2(\Omega)} \|\Pi_{\mathcal{D}} v_{\mathcal{D}}\|. \end{aligned}$$

For a piecewise constant reconstruction, **(A2)** is trivial since $\Pi_{\mathcal{D}} v_{\mathcal{D}} = \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}} v_{\mathcal{D}})$.

ASSUMPTIONS **(A3)** AND **(A4)**. Using the discrete functional analysis tools of [48, Chapter 8] the discrete Sobolev embedding **(A3)** is rather straightforward for all methods that fit in the GDM. This includes conforming and non-conforming $\mathbb{P}_1$ schemes as well as MFD schemes.
Assumption **(A4)** is identical to the assumption (A3) in [96]. Let R be the region where the bounds a and b pass from active to inactive, i.e. where $\overline{u}_d - \alpha^{-1}\overline{p}$ crosses these bounds. If R is of co-dimension 1, which is a rather natural situation, then the condition $|\Omega_{1,\mathcal{M}}| \lesssim h$ holds.
The $W^{1,\infty}$ regularity on $\overline{u}$ mentioned in **(A4)** can be established in a number of situations. It holds, for example, if Ω is a bounded open subset of class $C^{1,1}$, the coefficients of A belong to $C^{0,1}(\bar{\Omega})$, $\overline{u}_d \in W^{1,\infty}(\Omega)$ and $\overline{y}_d \in L^q(\Omega)$ for some $q > d$. Indeed, under these assumptions, [70, Theorem 2.4.2.5] ensures that the state and adjoint equations admit unique solutions in $H_0^1(\Omega) \cap W^{2,q}(\Omega) \subset W^{1,\infty}(\Omega)$. The projection formula (4.3.28) then shows that $\overline{u}$ inherits this Lipschitz continuity property over Ω. This also holds if Ω has corners but adequate symmetries (that preserve the $W^{2,q}(\Omega)$ regularity).
Assumption **(A4)** actually does not require the full $W^{1,\infty}$ regularity of $\overline{u}$, only this regularity on a neighbourhood of R. Considering a generic open set Ω with Lipschitz (but not necessarily smooth) boundary, [105, Theorem 7.3] ensures that $\overline{p}$ is continuous. If $\overline{u}_d$ is continuous and $a < (\overline{u}_d)_{|\partial\Omega} < b$, then $\overline{u}_d - \alpha^{-1}\overline{p}$ does not cross the levels a and b close to $\partial\Omega$, which means that R is a compact set inside Ω. The Lipschitz regularity of $\overline{u}$ then follows, under the same assumptions on A, $\overline{u}_d$ and $\overline{y}_d$ as above, from local regularity results (internal to Ω), without assuming that the boundary of Ω is $C^{1,1}$.
In all these cases, notice that, although the mesh $\mathcal{M}$ depends on h, the norm $\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)}$ remains bounded independently on $h \to 0$. Indeed, this norm is bounded by a Lipschitz constant of $\overline{u}$ on a neighbourhood of R.
Two main super-convergence results are established below.

Theorem 4.3.6 (Super-convergence for post-processed controls I)**.** *Let $\mathcal{D}$ be a GD and $\mathcal{M}$ be a mesh. Assume that*

- $\mathcal{U}_{\rm ad}$ *and* $\mathcal{U}_h$ *are given by* (4.3.23) *and* (4.3.24)*,*

- **(A1)–(A4)** *hold,*
- $\overline{u}_d$, $\overline{y}$ *and* $\overline{p}$ *belong to* $H^2(\Omega)$,
- $\overline{y}_d$ *and* f *belong to* $H^1(\Omega)$,

and let $\widetilde{u}$, $\widetilde{u}_h$ *be the post-processed controls defined by* (4.3.29)*. Then there exists* C *depending only on* α *such that*

$$\|\widetilde{u}-\widetilde{u}_h\| \lesssim_\eta Ch^{2-\frac{1}{2^*}}\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + Ch^2\mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p}), \tag{4.3.30}$$

where

$$\begin{aligned}\mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p}) = {}& \operatorname{minmod}(a,b) + \|\overline{y}_d\|_{H^1(\Omega)} + \|\overline{u}_d\|_{H^2(\Omega)} + \|f\|_{H^1(\Omega)} \\ & + \|\overline{y}\|_{H^2(\Omega)} + \|\overline{p}\|_{H^2(\Omega)}\end{aligned}$$

with $\operatorname{minmod}(a,b) = 0$ *if* $ab \le 0$ *and* $\operatorname{minmod}(a,b) = \min(|a|,|b|)$ *otherwise.*

The following auxiliary problem will be useful to prove the superconvergence of the control. For $g \in L^2(\Omega)$, let $p_{\mathcal{D}}^*(g) \in X_{\mathcal{D},0}$ solve

$$a_{\mathcal{D}}(w_{\mathcal{D}}, p_{\mathcal{D}}^*(g)) = (\Pi_{\mathcal{D}} y_{\mathcal{D}}(g) - \overline{y}_d, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \quad \forall w_{\mathcal{D}} \in X_{\mathcal{D},0}, \tag{4.3.31}$$

where $y_{\mathcal{D}}(g)$ is given by (4.3.5a) with $\overline{u}$ replaced by g.
Let us recall two approximation properties of $\boldsymbol{\mathcal{P}}_{\mathcal{M}}$. As proved in [48, Lemma 8.10],

$$\forall \phi \in H^1(\Omega)\,,\ \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\phi - \phi\| \lesssim_\eta h\|\phi\|_{H^1(\Omega)}. \tag{4.3.32}$$

For $K \in \mathcal{M}$, let $\overline{\boldsymbol{x}}_K$ be the centroid (centre of gravity) of K. The standard approximation property (see e.g. [54, Lemma 7.7] with $w_K \equiv 1$) yields

$$\forall K \in \mathcal{M}\,,\ \forall \phi \in H^2(K)\,,\ \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\phi - \phi(\overline{\boldsymbol{x}}_K)\|_{L^2(K)} \lesssim_\eta \operatorname{diam}(K)^2\|\phi\|_{H^2(K)}. \tag{4.3.33}$$

Proof of Theorem 4.3.6. Define $\widehat{u}$, $\widehat{p}$ and $\widehat{u}_d$ a.e. on Ω by: for all $K \in \mathcal{M}$ and all $\boldsymbol{x} \in K$, $\widehat{u}(\boldsymbol{x}) = \overline{u}(\overline{\boldsymbol{x}}_K)$, $\widehat{p}(\boldsymbol{x}) = \overline{p}(\overline{\boldsymbol{x}}_K)$ and $\widehat{u}_d(\boldsymbol{x}) = \overline{u}_d(\overline{\boldsymbol{x}}_K)$. From (4.3.29) and the Lipschitz continuity of $P_{[a,b]}$, it follows that

$$\begin{aligned}\|\widetilde{u}-\widetilde{u}_h\| &\le \alpha^{-1}\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \overline{p}_{\mathcal{M}}\| \\ &\le \alpha^{-1}\|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}p_{\mathcal{D}}^*(\overline{u})\| + \alpha^{-1}\|\Pi_{\mathcal{D}}p_{\mathcal{D}}^*(\overline{u}) - \Pi_{\mathcal{D}}p_{\mathcal{D}}^*(\widehat{u})\| \\ &\quad + \alpha^{-1}\|\Pi_{\mathcal{D}}p_{\mathcal{D}}^*(\widehat{u}) - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| \\ &=: \alpha^{-1}A_1 + \alpha^{-1}A_2 + \alpha^{-1}A_3.\end{aligned} \tag{4.3.34}$$

Step 1: estimate of A_1.

Recalling the equations (4.1.4b) and (4.3.5b) on $\overline{p}$ and $p_{\mathcal{D}}(\overline{u})$, a use of triangle inequality and **(A1)**-i) yields

$$\begin{aligned} A_1 &\leq \|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u})\| + \|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u})\| \\ &\lesssim h^2 \|\overline{y} - \overline{y}_d\|_{H^1(\Omega)} + \|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u})\|. \end{aligned} \tag{4.3.35}$$

The last term in this inequality is estimated now. Subtract (4.3.31) with $g = \overline{u}$ from (4.3.5b), substitute $w_{\mathcal{D}} = p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u})$, use the Cauchy–Schwarz inequality and property (4.3.26) in **(A1)**-ii) to obtain

$$\begin{aligned} \|\nabla_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\|^2 &\lesssim a_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}), p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u})) \\ &= (\overline{y} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}), \Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))) \\ &= (\overline{y} - \overline{y}_{\mathcal{M}}, \Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))) \\ &\quad + (\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}), \Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))) \\ &\lesssim h^2 \|\overline{y}\|_{H^2(\Omega)} \|\Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\| \\ &\quad + \|\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| \|\Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\|. \end{aligned}$$

The definition of $C_{\mathcal{D}}$ and **(A1)**-i) lead to $\|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u})\| \lesssim h^2 \|\overline{y}\|_{H^2(\Omega)} + h^2 \|f + \overline{u}\|_{H^1(\Omega)}$. Plugged into (4.3.35), this estimate yields

$$A_1 \lesssim h^2 (\|\overline{y} - \overline{y}_d\|_{H^1(\Omega)} + \|\overline{y}\|_{H^2(\Omega)} + \|f + \overline{u}\|_{H^1(\Omega)}). \tag{4.3.36}$$

Step 2: estimate of A_2.
Subtract the equations (4.3.31) satisfied by $p^*_{\mathcal{D}}(\overline{u})$ and $p^*_{\mathcal{D}}(\widehat{u})$ to obtain, for all $v_{\mathcal{D}} \in X_{\mathcal{D},0}$,

$$a_{\mathcal{D}}(v_{\mathcal{D}}, p^*_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\widehat{u})) = (\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u}), \Pi_{\mathcal{D}} v_{\mathcal{D}}). \tag{4.3.37}$$

As a consequence of (4.3.37) and Proposition 4.3.1,

$$A_2 = \|\Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u})\| \lesssim \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u})\|. \tag{4.3.38}$$

Choose $v_{\mathcal{D}} = y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u})$ in (4.3.37), set $w_{\mathcal{D}} = p^*_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\widehat{u})$, subtract the equations (4.3.5a) satisfied by $y_{\mathcal{D}}(\overline{u})$ and $y_{\mathcal{D}}(\widehat{u})$, use the orthogonality property of the projection operator $\boldsymbol{\mathcal{P}}_{\mathcal{M}}$, and invoke (4.3.32) and **(A2)** to deduce

$$\begin{aligned} \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|^2 &= (\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u}), \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u})) \\ &= a_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}), p^*_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\widehat{u})) \\ &= (\overline{u} - \widehat{u}, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \\ &= (\overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}} \overline{u}, \Pi_{\mathcal{D}} w_{\mathcal{D}} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}} w_{\mathcal{D}})) + (\boldsymbol{\mathcal{P}}_{\mathcal{M}} \overline{u} - \widehat{u}, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \\ &\lesssim_\eta h \|\overline{u}\|_{H^1(\Omega)} h \|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| \\ &\quad + \underbrace{\int_{\Omega_{1,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}} \overline{u} - \widehat{u}) \Pi_{\mathcal{D}} w_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x}}_{A_{21}} + \underbrace{\int_{\Omega_{2,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}} \overline{u} - \widehat{u}) \Pi_{\mathcal{D}} w_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x}}_{A_{22}}. \end{aligned} \tag{4.3.39}$$

Equation (4.3.37) and Proposition 4.3.1 show that

$$\|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| = \|\nabla_{\mathcal{D}}(p_{\mathcal{D}}^*(\overline{u}) - p_{\mathcal{D}}^*(\widehat{u}))\| \lesssim \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|. \tag{4.3.40}$$

A substitution of this estimate in (4.3.39) yields

$$\|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|^2 \lesssim_\eta h^2 \|\overline{u}\|_{H^1(\Omega)} \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\| + A_{21} + A_{22}. \tag{4.3.41}$$

A use of the Hölder's inequality, **(A4)**, **(A3)** and (4.3.40) leads to

$$\begin{aligned} A_{21} &\le \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}\|_{L^2(\Omega_{1,\mathcal{M}})} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\|_{L^2(\Omega_{1,\mathcal{M}})} \\ &\le h \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} |\Omega_{1,\mathcal{M}}|^{\frac{1}{2}} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\|_{L^{2^*}(\Omega)} |\Omega_{1,\mathcal{M}}|^{\frac{1}{2}-\frac{1}{2^*}} \\ &\lesssim h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| \\ &\lesssim h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|. \end{aligned} \tag{4.3.42}$$

Consider now A_{22}. For any $K \in \mathcal{M}_2$, $\overline{u} = a$ on K, $\overline{u} = b$ on K, or, by (4.3.28), $\overline{u} = \overline{u}_d - \alpha^{-1}\overline{p}$. Hence, $\overline{u} \in H^2(K)$ and, use (4.3.33), the definition of $C_{\mathcal{D}}$ and (4.3.40) to obtain

$$\begin{aligned} A_{22} &\le \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}\|_{L^2(\Omega_{2,\mathcal{M}})} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\| \\ &\lesssim_\eta h^2 \|\overline{u}\|_{H^2(\Omega_{2,\mathcal{M}})} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\| \\ &\lesssim_\eta h^2 \left(\|\overline{u}_d\|_{H^2(\Omega_{2,\mathcal{M}})} + \alpha^{-1} \|\overline{p}\|_{H^2(\Omega_{2,\mathcal{M}})} \right) \|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| \\ &\lesssim_\eta h^2 \left(\|\overline{u}_d\|_{H^2(\Omega_{2,\mathcal{M}})} + \alpha^{-1} \|\overline{p}\|_{H^2(\Omega_{2,\mathcal{M}})} \right) \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|. \end{aligned} \tag{4.3.43}$$

A substitution of (4.3.42) and (4.3.43) into (4.3.41) yields

$$\begin{aligned} \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\| \lesssim_\eta h^2 \|\overline{u}\|_{H^1(\Omega)} + h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \\ + h^2 \left(\|\overline{u}_d\|_{H^2(\Omega_{2,\mathcal{M}})} + \alpha^{-1} \|\overline{p}\|_{H^2(\Omega_{2,\mathcal{M}})} \right). \end{aligned} \tag{4.3.44}$$

Hence, use this in (4.3.38) to infer

$$A_2 \lesssim_\eta h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + h^2 (\|\overline{u}\|_{H^1(\Omega)} + \alpha^{-1} \|\overline{p}\|_{H^2(\Omega)} + \|\overline{u}_d\|_{H^2(\Omega)}). \tag{4.3.45}$$

Step 3: estimate of A_3.
Apply twice the stability result of Proposition 4.3.1 (first on the equation satisfied by $p_{\mathcal{D}}^*(\widehat{u}) - \overline{p}_{\mathcal{D}}$, and then on $y_{\mathcal{D}}(\widehat{u}) - \overline{y}_{\mathcal{D}}$) to write

$$A_3 = \|\Pi_{\mathcal{D}} p_{\mathcal{D}}^*(\widehat{u}) - \Pi_{\mathcal{D}} \overline{p}_{\mathcal{D}}\| \lesssim \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u}) - \Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}}\| \lesssim \|\widehat{u} - \overline{u}_h\|. \tag{4.3.46}$$

A use of the continuous optimality condition (4.1.4c), as in the proof of [96, Lemma 3.5] leads to, for a.e. $\boldsymbol{x} \in \Omega$,

$$\left[\overline{p}(\boldsymbol{x}) + \alpha(\overline{u}(\boldsymbol{x}) - \overline{u}_d(\boldsymbol{x}))\right] \left[v(\boldsymbol{x}) - \overline{u}(\boldsymbol{x})\right] \ge 0 \text{ for all } v \in \mathcal{U}_{\text{ad}}.$$

Since $\overline{u}$, $\overline{p}$ and $\overline{u}_h$ are continuous at the centroid $\overline{\boldsymbol{x}}_K$, choose $\boldsymbol{x} = \overline{\boldsymbol{x}}_K$ and $v(\overline{\boldsymbol{x}}_K) = \overline{u}_h(\overline{\boldsymbol{x}}_K)(= \overline{u}_h$ on $K)$. All the involved functions being constants over K, this gives

$$(\widehat{p} + \alpha(\widehat{u} - \widehat{u}_d))\,(\overline{u}_h - \widehat{u}) \geq 0 \text{ on } K, \text{ for all } K \in \mathcal{M}.$$

An integration over K and sum over $K \in \mathcal{M}$ yields

$$(\widehat{p} + \alpha\,(\widehat{u} - \widehat{u}_d)\,, \overline{u}_h - \widehat{u}) \geq 0.$$

Choose $v_h = \widehat{u}$ in the discrete optimality condition (4.2.3c) to obtain

$$(\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} + \alpha(\overline{u}_h - \overline{u}_d), \widehat{u} - \overline{u}_h) \geq 0.$$

An addition of the above two inequalities yields

$$(\widehat{p} - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} + \alpha(\widehat{u} - \overline{u}_h) + \alpha(\overline{u}_d - \widehat{u}_d), \overline{u}_h - \widehat{u}) \geq 0$$

and thus

$$\begin{aligned}
\alpha\|\widehat{u} - \overline{u}_h\|^2 &\leq (\widehat{p} - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}, \overline{u}_h - \widehat{u}) + \alpha(\overline{u}_d - \widehat{u}_d, \overline{u}_h - \widehat{u}) \\
&= (\widehat{p} - \overline{p}_{\mathcal{M}}, \overline{u}_h - \widehat{u}) + (\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\widehat{u}), \overline{u}_h - \widehat{u}) \\
&\quad + (\Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\widehat{u}) - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}, \overline{u}_h - \widehat{u}) + \alpha(\overline{u}_d - \widehat{u}_d, \overline{u}_h - \widehat{u}) \\
&=: M_1 + M_2 + M_3 + M_4. \qquad (4.3.47)
\end{aligned}$$

Since $\overline{u}_h - \widehat{u}$ is piecewise constant on $\mathcal{M}$, the orthogonality property of $\boldsymbol{\mathcal{P}}_{\mathcal{M}}$, (4.3.33) and (4.3.27) in **(A1)**-ii) lead to

$$\begin{aligned}
M_1 &= (\widehat{p} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{p}_{\mathcal{M}}, \overline{u}_h - \widehat{u}) \\
&= (\widehat{p} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{p}, \overline{u}_h - \widehat{u}) + (\boldsymbol{\mathcal{P}}_{\mathcal{M}}(\overline{p} - \overline{p}_{\mathcal{M}}), \overline{u}_h - \widehat{u}) \\
&\leq \|\widehat{p} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{p}\|\,\|\overline{u}_h - \widehat{u}\| + \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}(\overline{p} - \overline{p}_{\mathcal{M}})\|\,\|\overline{u}_h - \widehat{u}\| \\
&\lesssim_\eta h^2\|\overline{p}\|_{H^2(\Omega)}\|\overline{u}_h - \widehat{u}\|. \qquad (4.3.48)
\end{aligned}$$

A use of the Cauchy–Schwarz inequality, triangle inequality and the definitions of A_1 and A_2 yields

$$M_2 \leq \|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\widehat{u})\|\,\|\overline{u}_h - \widehat{u}\| \lesssim (A_1 + A_2)\|\overline{u}_h - \widehat{u}\|. \qquad (4.3.49)$$

Subtract the equations (4.2.3a) and (4.3.5a) (with $\widehat{u}$ instead of $\overline{u}$) satisfied by $\overline{y}_{\mathcal{D}}$ and $y_{\mathcal{D}}(\widehat{u})$, choose $w_{\mathcal{D}} = p^*_{\mathcal{D}}(\widehat{u}) - \overline{p}_{\mathcal{D}}$, and use the equations (4.2.3b) and (4.3.31) on $\overline{p}_{\mathcal{D}}$ and $p^*_{\mathcal{D}}(\widehat{u})$ to deduce

$$\begin{aligned}
M_3 &= (\Pi_{\mathcal{D}}(p^*_{\mathcal{D}}(\widehat{u}) - \overline{p}_{\mathcal{D}}), \overline{u}_h - \widehat{u}) \\
&= a_{\mathcal{D}}(\overline{y}_{\mathcal{D}} - y_{\mathcal{D}}(\widehat{u}), p^*_{\mathcal{D}}(\widehat{u}) - \overline{p}_{\mathcal{D}}) \\
&= (\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\widehat{u}) - \overline{y}_{\mathcal{D}}), \Pi_{\mathcal{D}}(\overline{y}_{\mathcal{D}} - y_{\mathcal{D}}(\widehat{u})) \leq 0. \qquad (4.3.50)
\end{aligned}$$

A use of the orthogonality property of $\mathcal{P}_{\mathcal{M}}$, (4.3.33) yields

$$\begin{aligned} M_4 = \alpha(\overline{u}_d - \widehat{u}_d, \overline{u}_h - \widehat{u}) &= \alpha(\mathcal{P}_{\mathcal{M}}\overline{u}_d - \widehat{u}_d, \overline{u}_h - \widehat{u}) \\ &\lesssim_\eta \alpha \|\mathcal{P}_{\mathcal{M}}\overline{u}_d - \widehat{u}_d\| \, \|\overline{u}_h - \widehat{u}\| \\ &\lesssim_\eta \alpha h^2 \|\overline{u}_d\|_{H^2(\Omega)} \|\overline{u}_h - \widehat{u}\|. \end{aligned} \tag{4.3.51}$$

A substitution of (4.3.48), (4.3.49) (together with the estimates (4.3.36) and (4.3.45) on A_1 and A_2), (4.3.50) and (4.3.51) into (4.3.47) yields an estimate on $\|\overline{u}_h - \widehat{u}\|$ which, when plugged into (4.3.46), gives

$$\begin{aligned} A_3 &\lesssim \|\overline{u}_h - \widehat{u}\| \\ &\lesssim_\eta \alpha^{-1} h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \\ &\quad + \alpha^{-1} h^2 \big[\|\overline{y} - \overline{y}_d\|_{H^1(\Omega)} + \|\overline{y}\|_{H^2(\Omega)} + (1+\alpha^{-1}) \|\overline{p}\|_{H^2(\Omega)} \\ &\qquad + \|f + \overline{u}\|_{H^1(\Omega)} + \|\overline{u}\|_{H^1(\Omega)} + (1+\alpha) \|\overline{u}_d\|_{H^2(\Omega)} \big]. \end{aligned} \tag{4.3.52}$$

Step 4: conclusion.
It is easy to check that $|P_{[a,b]}(s)| \leq \operatorname{minmod}(a,b) + |s|$, where minmod is defined in Theorem 4.3.6. Hence, by (4.3.28) and Lipschitz continuity of $P_{[a,b]}$,

$$\begin{aligned} \|\overline{u}\|_{H^1(\Omega)} &\leq \|P_{[a,b]}\left(\overline{u}_d - \alpha^{-1}\overline{p}\right)\|_{L^2(\Omega)} + \|\nabla\left(P_{[a,b]}\left(\overline{u}_d - \alpha^{-1}\overline{p}\right)\right)\|_{L^2(\Omega)^n} \\ &\leq \operatorname{minmod}(a,b)|\Omega|^{1/2} + 2\|\overline{u}_d - \alpha^{-1}\overline{p}\|_{H^1(\Omega)} \\ &\leq \operatorname{minmod}(a,b)|\Omega|^{1/2} + 2\|\overline{u}_d\|_{H^1(\Omega)} + 2\alpha^{-1}\|\overline{p}\|_{H^1(\Omega)}. \end{aligned} \tag{4.3.53}$$

Use this inequality and insert (4.3.36), (4.3.45) and (4.3.52) in (4.3.34) to conclude the proof of Theorem 4.3.6. □

Theorem 4.3.7 (Super-convergence for post-processed controls II)**.** *Let the assumptions and notations of Theorem 4.3.6 hold, except* **(A3)** *which is replaced by:*

$$\begin{aligned} &\text{there exists } \delta > 0 \text{ such that, for any } F \in L^2(\Omega), \\ &\text{the solution } \psi_D \text{ to (4.2.2) satisfies } \|\Pi_D \psi_D\|_{L^\infty(\Omega)} \leq \delta \|F\|. \end{aligned} \tag{4.3.54}$$

Then there exists C depending only on α and δ such that

$$\|\widetilde{u} - \widetilde{u}_h\| \lesssim_\eta Ch^2 \left[\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + \mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p}) \right]. \tag{4.3.55}$$

Proof. The proof of this theorem is identical to the proof of Theorem 4.3.6, except for the estimate of A_{21}. This estimate is the only source of the $2-\frac{1}{2^*}$ power (instead of 2), and the only place where we used Assumption **(A3)**, here replaced by (4.3.54). The estimate of A_{21} using this L^∞-bound assumption is actually rather simple. Recall **(A4)** and use (4.3.54) on the equation

(4.3.37) satisfied by $p^*_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\widehat{u})$ to obtain

$$\begin{aligned} A_{21} &= \int_{\Omega_{1,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u})\big(\Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u})\big)\,\mathrm{d}\boldsymbol{x} \\ &\lesssim \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}\|_{L^\infty(\Omega_{1,\mathcal{M}})} \|\Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u})\|_{L^\infty(\Omega_{1,\mathcal{M}})} |\Omega_{1,\mathcal{M}}| \\ &\lesssim h^2 \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \|\Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u})\|_{L^\infty(\Omega)} \\ &\lesssim h^2 \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \delta \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u})\|. \end{aligned} \tag{4.3.56}$$

The rest of the proof follows from this estimate. □

Remark 4.3.8. *The estimates* (4.3.30) *and* (4.3.55) *also hold if the two terms* $\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d$ *in* (4.3.29) *are replaced with* $\overline{u}_d$.

The super-convergence of the state and adjoint variables follow easily.

Corollary 4.3.9 (Super-convergence for the state and adjoint variables)**.** *Let* $(\overline{y},\overline{p})$ *and* $(\overline{y}_{\mathcal{D}},\overline{p}_{\mathcal{D}})$ *be the solutions to* (4.1.4a)–(4.1.4b) *and* (4.2.3a)–(4.2.3b)*. Under the assumptions of Theorem 4.3.6, the following error estimates hold, with C depending only on* α*:*

$$\|\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| \lesssim_\eta Ch^r \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + Ch^2 \mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p}), \tag{4.3.57}$$

$$\|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| \lesssim_\eta Ch^r \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + Ch^2 \mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p}), \tag{4.3.58}$$

where $\overline{y}_{\mathcal{M}}$ *and* $\overline{p}_{\mathcal{M}}$ *are defined as in* **(A1)***, and* $r = 2 - \frac{1}{2^*}$.
Under the assumptions of Theorem 4.3.7, (4.3.57) *and* (4.3.58) *hold with* $r = 2$ *and C depending only* α *and* δ.

Proof. A use of triangle inequality leads to

$$\|\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| \le \|\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| + \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u})\| + \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u}) - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\|. \tag{4.3.59}$$

Consider the first term on the right hand side of (4.3.59). Use the Assumption **(A1)**-i) to obtain

$$\|\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| \lesssim h^2 \|\overline{u} + f\|_{H^1(\Omega)}. \tag{4.3.60}$$

Under the assumptions of Theorem 4.3.6, the second term on the right hand side of (4.3.59) is estimated by (4.3.44), and the third term is estimated by using (4.3.46) and (4.3.52). Plug these estimates alongside (4.3.60) into (4.3.59), and use (4.3.53) to conclude the proof of (4.3.57). The result for the adjoint variable can be derived similarly.
The full h^2 estimates are obtained, under the assumptions of Theorem 4.3.7, by following the same reasoning and using the improved estimate (4.3.56) on A_{21} (which leads to improved estimates (4.3.44) and (4.3.52)).

□

Even for classical schemes, the L^∞ estimate (4.3.54) is only known under restrictive assumptions on the mesh. For example, for conforming and non-conforming $\mathbb{P}_1$ finite elements, it requires the quasi-uniformity of the mesh [67], which prevents considering local refinements widely used in practical applications. The scope of Theorem 4.3.7 is therefore limited in that sense, but it nonetheless extends to various methods (see e.g. Corollary 4.3.10) the super-convergence established in [96] for conforming $\mathbb{P}_1$ finite elements.
On the contrary, Theorem 4.3.6 holds under much less restrictive assumptions (see below for a discussion of **(A1)–(A4)**), and applies seamlessly to locally refined meshes, for essentially all numerical methods currently covered by the GDM. It is also useful to notice that Theorem 4.3.6 *nearly* provides an h^2 convergence rate. If $d = 2$, the Sobolev exponent 2^* can be any finite number. In that case, (4.3.30), (4.3.57) and (4.3.58) are $\mathcal{O}(h^{2-\varepsilon})$ estimates, for any $\varepsilon > 0$. If $d = 3$, the estimates are of order $\mathcal{O}(h^{11/6})$. In each case, as noticed in Section 4.1.2, these rates are numerically very close to a full $\mathcal{O}(h^2)$ convergence rate.

Application to non-conforming $\mathbb{P}_1$ and hMFD

The generic results on the GDM apply to all methods covered by this framework. In particular, as mentioned in Section 4.1.2, to non-conforming $\mathbb{P}_1$ finite elements and hMFD methods. We state here a corollary of the super-convergence results on the control (Theorems 4.3.6 and 4.3.7) for these two methods. We could as easily state obvious consequence for these two schemes of Theorem 4.3.2, Proposition 4.3.4 and Corollary 4.3.9.

Corollary 4.3.10 (Super-convergence of the control for nc$\mathbb{P}_1$ and hMFD schemes)**.**
Assume that Ω is convex and A is Lipschiz-continuous. Let $\mathcal{M}$ be a mesh in the sense of [51, Definition 2.21], with centers at the centers of mass of the cells. Assume $\mathcal{U}_{\rm ad}$ and $\mathcal{U}_h$ are given by (4.3.23) *and* (4.3.24)*,* **(A4)** *holds, $\overline{u}_d \in H^2(\Omega)$ and that $(\overline{y}_d, f) \in H^1(\Omega)$.*
Consider either one of the following schemes, as described in Section 4.1.2, with associated post-processed controls (here, $(\overline{y}_h, \overline{p}_h, \overline{u}_h)$ is the solution to the scheme for the control problem):

- nc$\mathbb{P}_1/\mathbb{P}_0$ *scheme: η satisfies* (1.4.1)*, $\widetilde{u} = P_{[a,b]}(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d - \alpha^{-1}\overline{p})$, and $\widetilde{u}_h = P_{[a,b]}(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d - \alpha^{-1}\overline{p}_h)$.*
- hMFD *schemes: η is an upper bound of $\theta_{\mathcal{M}}$ defined by [51, Eq. (2.27)] and, for all $K \in \mathcal{M}$,*

$$\widetilde{u}_{|K} = P_{[a,b]}\left(\fint_K \overline{u}_d - \alpha^{-1}\overline{p}(\overline{\boldsymbol{x}}_K)\right) \text{ and } (\widetilde{u}_h)_{|K} = P_{[a,b]}\left(\fint_K \overline{u}_d - \alpha^{-1}(\overline{p}_h)_K\right).$$

Then there exists C depending only on Ω, A, α, a, b, $\overline{u}$, $\overline{u}_d$, $\overline{y}_d$, f and η such that

$$\|\widetilde{u} - \widetilde{u}_h\| \leq Ch^{2-\frac{1}{2^*}}. \tag{4.3.61}$$

Moreover, if $\chi \geq \max_{K\in\mathcal{M}} \frac{h^d}{|K|}$, then there exists C depending only on Ω, A, α, a, b, $\overline{u}$, $\overline{u}_d$, $\overline{y}_d$, f, η and χ such that

$$\|\widetilde{u} - \widetilde{u}_h\| \leq Ch^2. \tag{4.3.62}$$

Remark 4.3.11. *The conforming $\mathbb{P}_1$ FEM is a GDM for the gradient discretisation defined by $\mathcal{D} = (V_h, \mathrm{Id}, \nabla)$, where V_h is the conforming $\mathbb{P}_1$ space on the considered mesh. Then, $W_\mathcal{D} \equiv 0$ and $S_\mathcal{D}$ is bounded above by the interpolation error of the $\mathbb{P}_1$ method. For this gradient discretisation method, Theorems 4.3.2 and 4.3.7 provide, respectively, $\mathcal{O}(h)$ error estimates on the control and $\mathcal{O}(h^2)$ error estimates on the post-processed controls (under a quasi-uniformity assumption on the sequence of meshes). These rates are the same that are already proved in [96]. For* nc$\mathbb{P}_1$ *FEM, the estimate* (4.3.62) *provides quadratic rate of convergence in a similar way as for conforming $\mathbb{P}_1$ method.*

Proof of Corollary 4.3.10. [51, Sections 3.2.1 and 3.6.1] presents a description of the GDs corresponding to the nc$\mathbb{P}_1$ and hMFD schemes (the latter is seen as a GS through its identification as a hybrid mimetic mixed method, see [49, 50]; the corresponding GD is recalled in Section A.3, Appendix).
Using these gradient discretisations, (4.3.61) follows from Theorem 4.3.6 if we can prove that **(A1)**–**(A3)** hold, for a proper choice of operator $w \mapsto w_\mathcal{M}$.
Note that the assumptions on Ω and A ensure that the state (and thus adjoint) equations satisfy the elliptic regularity: if the source terms are in $L^2(\Omega)$ then the solutions belong to $H^2(\Omega)$.
For the nc$\mathbb{P}_1$ scheme, recall that $w_\mathcal{M} = w$ and the superconvergence result (4.3.25) is known under the elliptic regularity. Also, $\Pi_\mathcal{D} w_\mathcal{D}$ is simply the solution w_h to the scheme. Properties (4.3.26) and (4.3.27) are obvious since $w - w_\mathcal{M} = 0$. This proves **(A1)**. Assumption **(A2)** follows easily from a Taylor expansion since $\nabla_\mathcal{D} v_\mathcal{D}$ is the broken gradient of $\Pi_\mathcal{D} v_\mathcal{D}$. Assumption **(A3)** follows from [44, Proposition 5.4], by noticing that for piecewise polynomial functions that match at the face centroids, the discrete $\|\cdot\|_{1,2,h}$ norm in [44] boils down to the $L^2(\Omega)^d$ norm of the broken gradient.
Now consider the hMFD scheme, for which let $(w_\mathcal{M})_{|K} = w(\bar{\boldsymbol{x}}_K)$ for all $K \in \mathcal{M}$. The superconvergence result of **(A1)**-i) is proved in, e.g., [21, 54]. As mentioned in Section 4.3.2, Properties (4.3.26) and (4.3.27) follow from (4.3.33); **(A2)** is trivially true, and **(A3)** follows from the discrete functional analysis results of [48, Lemma 8.15 and Lemma 13.11].

The full super-convergence result (4.3.62) follows from Theorem 4.3.7 if the L^∞ bound (4.3.54) can be established under the assumption that χ is bounded – i.e. the mesh is quasi-uniform. This L^∞ bound is known for the nc$\mathbb{P}_1$ finite element method [67], and is proved in Theorem A.3.1 for the HMM method. □

4.4 The case of Neumann BC, with distributed and boundary control

4.4.1 Model problem

Consider the distributed and boundary optimal control problem governed by elliptic equations with Neumann BC given by:

$$\min_{U\in\mathcal{U}_{\rm ad}} J(U) \quad \text{subject to} \tag{4.4.1a}$$

$$-\mathrm{div}(A\nabla y(U)) + c_0 y(U) = f + u \quad \text{in } \Omega, \tag{4.4.1b}$$

$$A\nabla y(U)\cdot n = f_b + u_b \quad \text{on } \partial\Omega, \tag{4.4.1c}$$

where Ω, A and f are as in Section 4.1.1, $f_b \in L^2(\partial\Omega)$, $c_0 > 0$ is a positive constant, n is the outer unit normal to Ω, u, u_b are the control variables, $U = (u, u_b)$ and $y(U)$ is the state variable. The cost functional is

$$J(U) := \frac{1}{2}\|y(U) - \overline{y}_d\|^2 + \frac{\alpha}{2}\|u\|^2 + \frac{\beta}{2}\|u_b\|^2_{L^2(\partial\Omega)}$$

with $\alpha > 0$ and $\beta > 0$ being fixed regularization parameters and $\overline{y}_d \in L^2(\Omega)$ being the desired state variable. The set of admissible controls $\mathcal{U}_{\rm ad} \subset L^2(\Omega)\times L^2(\partial\Omega)$ is a non-empty, convex and closed set. For a general element $V \in L^2(\Omega)\times L^2(\partial\Omega)$, v and v_b denote its components in $L^2(\Omega)$ and $L^2(\partial\Omega)$, that is, $V = (v, v_b)$.
It is well known that given $U \in \mathcal{U}_{\rm ad}$, there exists a unique weak solution $y(U) \in H^1(\Omega)$ of (4.4.1b)-(4.4.1c). That is, $y = y(U) \in H^1(\Omega)$ such that, for all $w \in H^1(\Omega)$,

$$a(y(U), w) = \int_\Omega (f+u)w\,\mathrm{d}\boldsymbol{x} + \int_{\partial\Omega}(f_b+u_b)\gamma(w)\,\mathrm{d}s(\boldsymbol{x}), \tag{4.4.2}$$

where $a(z,w) = \int_\Omega (A\nabla z\cdot\nabla w + c_0 zw)\,\mathrm{d}\boldsymbol{x}$ and $\gamma : H^1(\Omega)\to L^2(\partial\Omega)$ is the trace operator.

Here and throughout, $\|\cdot\|_\partial$ and $(\cdot,\cdot)_\partial$ denote the norm and scalar product in $L^2(\partial\Omega)$. Also denote $[\![\cdot\,|\,\cdot]\!]$ as the scalar product on $L^2(\Omega)\times L^2(\partial\Omega)$ defined by

$$\forall U, V \in L^2(\Omega)\times L^2(\partial\Omega), \quad [\![U\,|\,V]\!] = \alpha(u,v) + \beta(u_b, v_b)_\partial.$$

The convex control problem (4.4.1) has a unique solution $(\overline{y}, \overline{U}) \in H^1(\Omega)\times\mathcal{U}_{\rm ad}$ and there exists a co-state $\overline{p} \in H^1(\Omega)$ such that the triplet $(\overline{y}, \overline{p}, \overline{U}) \in H^1(\Omega)\times H^1(\Omega)\times\mathcal{U}_{\rm ad}$ satisfies the Karush-Kuhn-Tucker (KKT) optimality conditions [90]:

$$a(\overline{y}, w) = (f+\overline{u}, w) + (f_b + \overline{u}_b, \gamma(w))_\partial \qquad \forall\, w \in H^1(\Omega), \tag{4.4.3a}$$

$$a(w, \overline{p}) = (\overline{y} - \overline{y}_d, w) \qquad \forall\, w \in H^1(\Omega), \tag{4.4.3b}$$

$$[\![\overline{U} + \overline{P}_{\alpha,\beta}\,|\,V - \overline{U}]\!] \geq 0 \qquad \forall V \in \mathcal{U}_{\rm ad}, \tag{4.4.3c}$$

where $\overline{P}_{\alpha,\beta} = (\alpha^{-1}\overline{p}, \beta^{-1}\gamma(\overline{p}))$.

4.4.2 The GDM for elliptic equations with Neumann BC

Definition 4.4.1 (GD for Neumann BC with reaction)**.** *A gradient discretisation for Neumann BC is a quadruplet $\mathcal{D} = (X_\mathcal{D}, \Pi_\mathcal{D}, \mathbb{T}_\mathcal{D}, \nabla_\mathcal{D})$ such that*

- *$X_\mathcal{D}$ is a finite dimensional space of degrees of freedom,*
- *$\Pi_\mathcal{D} : X_\mathcal{D} \to L^2(\Omega)$ is a linear mapping that reconstructs a function from the degrees of freedom,*
- *$\mathbb{T}_\mathcal{D} : X_\mathcal{D} \to L^2(\partial\Omega)$ is a linear mapping that reconstructs a trace from the degrees of freedom,*
- *$\nabla_\mathcal{D} : X_\mathcal{D} \to L^2(\Omega)^d$ is a linear mapping that reconstructs a gradient from the degrees of freedom.*
- *The following quantity is a norm on $X_\mathcal{D}$:*

$$\|w\|_\mathcal{D} := \|\nabla_\mathcal{D} w\| + \|\Pi_\mathcal{D} w\|. \tag{4.4.4}$$

If $F \in L^2(\Omega)$ and $G \in L^2(\partial\Omega)$, a GS for a linear elliptic problem

$$\begin{cases} -\mathrm{div}(A\nabla\psi) + c_0\psi = F \text{ in } \Omega, \\ A\nabla\psi \cdot n = G \text{ on } \partial\Omega \end{cases} \tag{4.4.5}$$

is then obtained from a GD $\mathcal{D}$ by writing:

$$\begin{aligned} &\text{Find } \psi_\mathcal{D} \in X_\mathcal{D} \text{ such that, for all } w_\mathcal{D} \in X_\mathcal{D}, \\ &a_\mathcal{D}(\psi_\mathcal{D}, w_\mathcal{D}) = \int_\Omega F\Pi_\mathcal{D} w_\mathcal{D}\,\mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} G\mathbb{T}_\mathcal{D} w_\mathcal{D}\,\mathrm{d}s(\boldsymbol{x}), \end{aligned} \tag{4.4.6}$$

where $a_\mathcal{D}(\psi_\mathcal{D}, w_\mathcal{D}) = \int_\Omega (A\nabla_\mathcal{D}\psi_\mathcal{D} \cdot \nabla_\mathcal{D} w_\mathcal{D} + c_0\Pi_\mathcal{D}\psi_\mathcal{D}\Pi_\mathcal{D} w_\mathcal{D})\,\mathrm{d}\boldsymbol{x}$.
For Neumann boundary value problems, the quantities $C_\mathcal{D}$, $S_\mathcal{D}$ and $W_\mathcal{D}$ measuring the accuracy of the GS are defined as follows.

$$C_\mathcal{D} := \max_{w \in X_\mathcal{D}\setminus\{0\}} \left(\frac{\|\mathbb{T}_\mathcal{D} w\|_\partial}{\|w\|_\mathcal{D}}, \frac{\|\Pi_\mathcal{D} w\|}{\|w\|_\mathcal{D}} \right). \tag{4.4.7}$$

$$\forall \varphi \in H^1(\Omega),\ S_\mathcal{D}(\varphi) = \min_{w \in X_\mathcal{D}} \Big(\|\Pi_\mathcal{D} w - \varphi\| + \|\mathbb{T}_\mathcal{D} w - \gamma(\varphi)\|_\partial + \|\nabla_\mathcal{D} w - \nabla\varphi\| \Big). \tag{4.4.8}$$

$$\forall \boldsymbol{\varphi} \in H_{\mathrm{div},\partial}(\Omega),$$

$$W_\mathcal{D}(\boldsymbol{\varphi}) = \max_{w \in X_\mathcal{D}\setminus\{0\}} \frac{1}{\|w\|_\mathcal{D}} \left| \int_\Omega \Pi_\mathcal{D} w \mathrm{div}\boldsymbol{\varphi} + \nabla_\mathcal{D} w \cdot \boldsymbol{\varphi}\,\mathrm{d}\boldsymbol{x} - \int_{\partial\Omega} \mathbb{T}_\mathcal{D} w \gamma_n(\boldsymbol{\varphi})\,\mathrm{d}s(\boldsymbol{x}) \right|, \tag{4.4.9}$$

where γ_n is the normal trace on $\partial\Omega$, and $H_{\mathrm{div},\partial}(\Omega) = \{\boldsymbol{\varphi} \in L^2(\Omega)^d : \mathrm{div}\boldsymbol{\varphi} \in L^2(\Omega),\ \gamma_n(\boldsymbol{\varphi}) \in L^2(\partial\Omega)\}$.
Using these quantities, define $\mathrm{WS}_\mathcal{D}$ as in (4.2.11) and the following error estimate can be established.

Theorem 4.4.2 (Error estimate for the PDE with Neumann BC)**.** *Let $\mathcal{D}$ be a GD in the sense of Definition 4.4.1, let ψ be the solution in $H^1(\Omega)$ to* (4.4.5)*, and let $\psi_\mathcal{D}$ be the solution to* (4.4.6)*. Then*

$$\|\Pi_\mathcal{D}\psi_\mathcal{D} - \psi\| + \|\nabla_\mathcal{D}\psi_\mathcal{D} - \nabla\psi\| + \|\mathbb{T}_\mathcal{D}\psi_\mathcal{D} - \gamma(\psi)\|_\partial \lesssim \mathrm{WS}_\mathcal{D}(\psi).$$

Proof. The estimate

$$\|\Pi_\mathcal{D}\psi_\mathcal{D} - \psi\| + \|\nabla_\mathcal{D}\psi_\mathcal{D} - \nabla\psi\| \lesssim \mathrm{WS}_\mathcal{D}(\psi) \tag{4.4.10}$$

is standard, and can be established as for homogeneous Dirichlet BC (see, e.g., [48, Theorem 3.11] for the pure Neumann problem). The estimate on the traces is less standard, and hence we detail it now. Introduce an interpolant

$$\mathcal{P}_\mathcal{D}\psi \in \operatorname*{argmin}_{w\in X_\mathcal{D}} \Big(\|\Pi_\mathcal{D} w - \psi\| + \|\mathbb{T}_\mathcal{D} w - \gamma(\psi)\|_\partial + \|\nabla_\mathcal{D} w - \nabla\psi\| \Big)$$

and notice that

$$\|\Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\psi - \psi\| + \|\mathbb{T}_\mathcal{D}\mathcal{P}_\mathcal{D}\psi - \gamma(\psi)\|_\partial + \|\nabla_\mathcal{D}\mathcal{P}_\mathcal{D}\psi - \nabla\psi\| \le S_\mathcal{D}(\psi). \tag{4.4.11}$$

By definition of $C_\mathcal{D}$ and of the norm $\|\cdot\|_\mathcal{D}$, for all $v \in X_\mathcal{D}$,

$$\|\mathbb{T}_\mathcal{D} v\|_\partial \le C_\mathcal{D} \left(\|\Pi_\mathcal{D} v\| + \|\nabla_\mathcal{D} v\| \right).$$

Substituting $v = \psi_\mathcal{D} - \mathcal{P}_\mathcal{D}\psi$, a triangle inequality and (4.4.11) therefore lead to

$$\begin{aligned}
&\|\mathbb{T}_\mathcal{D}\psi_\mathcal{D} - \gamma(\psi)\|_\partial \\
&\quad \le \|\mathbb{T}_\mathcal{D}(\psi_\mathcal{D} - \mathcal{P}_\mathcal{D}\psi)\|_\partial + \|\mathbb{T}_\mathcal{D}\mathcal{P}_\mathcal{D}\psi - \gamma(\psi)\|_\partial \\
&\quad \le C_\mathcal{D} \left(\|\Pi_\mathcal{D}\psi_\mathcal{D} - \Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\psi\| + \|\nabla_\mathcal{D}\psi_\mathcal{D} - \nabla_\mathcal{D}\mathcal{P}_\mathcal{D}\psi\| \right) + S_\mathcal{D}(\psi).
\end{aligned} \tag{4.4.12}$$

Use the triangle inequality again and the estimates (4.4.10) and (4.4.11) to write

$$\begin{aligned}
&\|\Pi_\mathcal{D}\psi_\mathcal{D} - \Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\psi\| + \|\nabla_\mathcal{D}\psi_\mathcal{D} - \nabla_\mathcal{D}\mathcal{P}_\mathcal{D}\psi\| \\
&\quad \le \|\Pi_\mathcal{D}\psi_\mathcal{D} - \psi\| + \|\psi - \Pi_\mathcal{D}\mathcal{P}_\mathcal{D}\psi\| + \|\nabla_\mathcal{D}\psi_\mathcal{D} - \nabla\psi\| + \|\nabla\psi - \nabla_\mathcal{D}\mathcal{P}_\mathcal{D}\psi\| \\
&\quad \lesssim \mathrm{WS}_\mathcal{D}(\psi).
\end{aligned}$$

The proof is complete by plugging this result in (4.4.12). □

4.4.3 The GDM for the Neumann control problem

Let $\mathcal{D}$ be a GD as in Definition 4.4.1, $\mathcal{U}_h$ be a finite dimensional space of $L^2(\Omega)$, and set $\mathcal{U}_{\mathrm{ad},h} = \mathcal{U}_\mathrm{ad} \cap \mathcal{U}_h$. A GS for (4.4.3) consists in seeking $(\overline{y}_\mathcal{D}, \overline{p}_\mathcal{D}, \overline{U}_h) \in X_\mathcal{D} \times X_\mathcal{D} \times \mathcal{U}_{\mathrm{ad},h}$, with $\overline{U}_h = (\overline{u}_h, \overline{u}_{b,h})$, such that

$$a_\mathcal{D}(\overline{y}_\mathcal{D}, w_\mathcal{D}) = (f + \overline{u}_h, \Pi_\mathcal{D} w_\mathcal{D}) + (f_b + \overline{u}_{b,h}, \mathbb{T}_\mathcal{D} w_\mathcal{D})_\partial \qquad \forall\, w_\mathcal{D} \in X_\mathcal{D}, \tag{4.4.13a}$$

$$a_\mathcal{D}(w_\mathcal{D}, \overline{p}_\mathcal{D}) = (\Pi_\mathcal{D}\overline{y}_\mathcal{D} - \overline{y}_d, \Pi_\mathcal{D} w_\mathcal{D}) \qquad \forall\, w_\mathcal{D} \in X_\mathcal{D}, \tag{4.4.13b}$$

$$\left[\!\!\left[\overline{U}_h + \overline{P}_{\mathcal{D},\alpha,\beta} \,\middle|\, V_h - \overline{U}_h \right]\!\!\right] \ge 0 \qquad \forall\, V_h \in \mathcal{U}_{\mathrm{ad},h}, \tag{4.4.13c}$$

where $\overline{P}_{\mathcal{D},\alpha,\beta} = (\alpha^{-1}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}, \beta^{-1}\mathbb{T}_{\mathcal{D}}\bar{p}_{\mathcal{D}})$.
Let $\mathrm{Pr}_h : L^2(\Omega)\times L^2(\partial\Omega) \to \mathcal{U}_h$ be the L^2 orthogonal projection on $\mathcal{U}_h$ for the scalar product $[\![\cdot\,|\,\cdot]\!]$. Denote the norm on $L^2(\Omega)\times L^2(\partial\Omega)$ associated to $[\![\cdot\,|\,\cdot]\!]$ by $|||\cdot|||$, so that $|||V||| = \sqrt{\alpha\|v\|^2 + \beta\|v_b\|^2}$. If $W \in L^2(\Omega)\times L^2(\partial\Omega)$, define

$$E_h(W) = |||W - \mathrm{Pr}_h W|||.$$

Theorem 4.4.3 (Control estimate)**.** *Let $\mathcal{D}$ be a GD in the sense of Definition 4.4.1, $\overline{U}$ be the optimal control for* (4.4.3) *and $\overline{U}_h$ be the optimal control for the GS* (4.4.13)*. Assume that*

$$\mathrm{Pr}_h(\mathcal{U}_{\mathrm{ad}}) \subset \mathcal{U}_{\mathrm{ad},h}. \tag{4.4.14}$$

Then there exists C only depending on Ω, A, α, β and an upper bound of $C_{\mathcal{D}}$ such that

$$|||\overline{U} - \overline{U}_h||| \le C\Big(E_h(\overline{P}_{\alpha,\beta}) + E_h(\overline{U}) + \mathrm{WS}_{\mathcal{D}}(\overline{p}) + \mathrm{WS}_{\mathcal{D}}(\overline{y})\Big).$$

Proof. The proof is identical to the proof of Theorem 4.3.2 (taking $\overline{u}_d = 0$), with obvious substitutions (e.g. $\overline{P}_{\mathcal{D},\alpha} \rightsquigarrow \overline{P}_{\mathcal{D},\alpha,\beta}$ and $\overline{u}_h \rightsquigarrow \overline{U}_h$) and the L^2 inner products $(\cdot,\cdot)$ replaced by $[\![\cdot\,|\,\cdot]\!]$ whenever they involve $\overline{P}_{\mathcal{D},\alpha}$ or $\overline{u}_h$. □

Remark 4.4.4 (Super-convergence of the control for Neumann problems)**.** *Using the same technique as in the proof of Theorem 4.3.6, and extending the assumptions* **(A1)–(A4)** *to boundary terms in a natural way (based on trace inequalities and Sobolev embedding of $H^{1/2}(\partial\Omega)$), an $\mathcal{O}(h^{3/2})$ super-convergence result can be proved on post-processed controls for Neumann BC.*

Remark 4.4.5. *Consider the distributed optimal control problem governed by elliptic equations with Neumann BC given by:*

$$\min_{u\in\mathcal{U}_{\mathrm{ad}}} J(u) \quad \textit{subject to} \tag{4.4.15a}$$

$$-\mathrm{div}(A\nabla y(u)) = u \quad \textit{in } \Omega, \tag{4.4.15b}$$

$$A\nabla y(u)\cdot n = 0 \quad \textit{on } \partial\Omega, \quad \int_\Omega y(u)\,\mathrm{d}\boldsymbol{x} = 0, \tag{4.4.15c}$$

where Ω and A are as in Section 4.1.1. The cost functional is (4.1.2) *with $\overline{u}_d = 0$ and $\overline{y}_d \in L^2(\Omega)$ is such that $\int_\Omega \overline{y}_d\,\mathrm{d}\boldsymbol{x} = 0$. Fixing $a < 0 < b$, the admissible set of controls is chosen as*

$$\mathcal{U}_{\mathrm{ad}} = \left\{u \in L^2(\Omega) \,:\, a \le u \le b \text{ a.e. and } \int_\Omega u\,\mathrm{d}\boldsymbol{x} = 0\right\}.$$

For a given $u \in \mathcal{U}_{\mathrm{ad}}$, there exists a unique weak solution $y = y(u) \in H^1_\star(\Omega) := \{w \in H^1(\Omega) : \int_\Omega w\,\mathrm{d}\boldsymbol{x} = 0\}$ of (4.4.15b)–(4.4.15c)*.*

The convex control problem (4.4.15) *has a unique solution* $(\overline{y},\overline{u}) \in H^1_\star(\Omega)\times\mathcal{U}_{\rm ad}$ *and there exists a co-state* $\overline{p}\in H^1_\star(\Omega)$ *such that the triplet* $(\overline{y},\overline{p},\overline{u}) \in H^1_\star(\Omega)\times H^1_\star(\Omega)\times\mathcal{U}_{\rm ad}$ *satisfies the Karush-Kuhn-Tucker optimality conditions [90]:*

$$a(\overline{y},w) = (\overline{u},w) \qquad \forall\, w\in H^1(\Omega), \tag{4.4.16a}$$

$$a(w,\overline{p}) = (\overline{y}-\overline{y}_d,w) \qquad \forall\, w\in H^1(\Omega), \tag{4.4.16b}$$

$$(\overline{p}+\alpha\overline{u},v-\overline{u}) \geq 0 \qquad \forall\, v\in\mathcal{U}_{\rm ad}, \tag{4.4.16c}$$

where $a(z,w) = \int_\Omega A\nabla z\cdot\nabla w\,\mathrm{d}\boldsymbol{x}$.
The adaptation of the theoretical analysis and numerical algorithms for this problem is presented in next chapter.

4.5 Numerical results

In this section, numerical results to support the theoretical estimates obtained in the previous sections are presented. Three specific schemes are used for the state and adjoint variables: conforming finite element method, non-conforming finite element method, and hybrid mimetic mixed (HMM) method (a family that contains, the hMFD schemes analyzed for example in [21], owing to the results in [49]). See [51] for the description of the GDs corresponding to these methods (see also Section A.3, Appendix for the HMM GD). The control variable is discretised using piecewise constant functions. The discrete solution is computed by using the primal-dual active set algorithm, see [109, Section 2.12.4].
Let the relative errors be denoted by

$$\mathsf{err}_{\mathcal{D}}(\overline{y}) := \frac{\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\overline{y}_{\mathcal{M}}\|}{\|\overline{y}_{\mathcal{M}}\|}, \quad \mathsf{err}_{\mathcal{D}}(\nabla\overline{y}) := \frac{\|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\nabla\overline{y}\|}{\|\nabla\overline{y}\|}$$

$$\mathsf{err}_{\mathcal{D}}(\overline{p}) := \frac{\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}-\overline{p}_{\mathcal{M}}\|}{\|\overline{p}_{\mathcal{M}}\|}, \quad \mathsf{err}_{\mathcal{D}}(\nabla\overline{p}) := \frac{\|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}-\nabla\overline{p}\|}{\|\nabla\overline{p}\|}$$

$$\mathsf{err}(\overline{u}) := \frac{\|\overline{u}_h-\overline{u}\|}{\|\overline{u}\|} \quad \text{and} \quad \mathsf{err}(\widetilde{u}) := \frac{\|\widetilde{u}_h-\widetilde{u}\|}{\|\overline{u}\|}.$$

Here, the definitions of $\widetilde{u}$ and $\widetilde{u}_h$ follow from (4.3.29).

- For FEMs,

$$\widetilde{u} = P_{[a,b]}(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d - \alpha^{-1}\overline{p}) \text{ and } \widetilde{u}_h = P_{[a,b]}(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d - \alpha^{-1}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}).$$

- For HMM methods,

$$\widetilde{u}_{|K} = P_{[a,b]}(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d - \alpha^{-1}\overline{p}(\overline{\boldsymbol{x}}_K)) \text{ for all } K\in\mathcal{M}, \text{ and}$$
$$\widetilde{u}_h = P_{[a,b]}(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}_d - \alpha^{-1}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}) = \overline{u}_h.$$

The L^2 errors of state and adjoint variables corresponding to the FEMs are computed using a seven point Gaussian quadrature formula, and the energy norms are calculated using midpoint rule. In the case of HMM, both the energy and L^2 norms are computed using the midpoint rule. The L^2 errors of control variable is computed using a three point Gaussian quadrature formula. The post-processed control corresponding to the FEMs is evaluated using a seven point Gaussian quadrature formula, whereas for the HMM methods, the post-processed control is computed using midpoint rule. For HMM methods, simpler quadrature rules can be used owing to the fact that the reconstructed functions are piecewise constants. These errors are plotted against the mesh parameter h in the log-log scale.

4.5.1 Dirichlet BC

The model problem is constructed in such a way that the exact solution is known.

Example 1

This example is taken from [2]. In this experiment, the computational domain Ω is taken to be the unit square $(0,1)^2$. The data in the optimal distributed control problem (4.1.1a)–(4.1.1c) are chosen as follows:

$$\begin{aligned}
&\bar{y} = \sin(\pi x)\sin(\pi y), \quad \bar{p} = \sin(\pi x)\sin(\pi y),\\
&\bar{u}_d = 1 - \sin(\pi x/2) - \sin(\pi y/2), \quad \alpha = 1,\\
&\mathcal{U}_{\text{ad}} = [0,\infty), \quad \bar{u} = \max(\bar{u}_d - \bar{p}, 0).
\end{aligned}$$

The source term f and the desired state $\bar{y}_d$ are the computed using

$$f = -\Delta\bar{y} - \bar{u}, \quad \bar{y}_d = \bar{y} + \Delta\bar{p}.$$

Figure 4.3 shows the initial triangulation of a square domain and its uniform refinement.

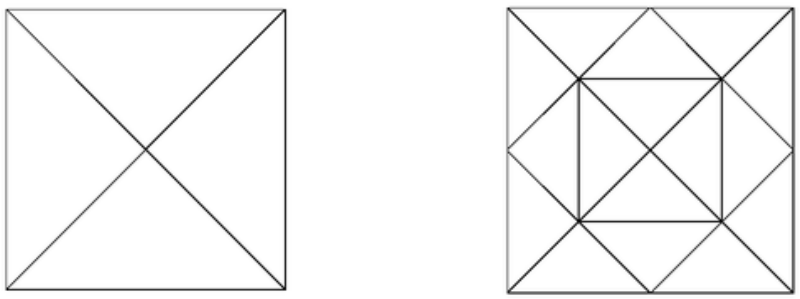

Figure 4.3: Initial triangulation and its uniform refinement

Since Ω is convex, Theorems 4.3.2 and 4.3.6 (see also the discussion before Section 4.3.2), Proposition 4.3.4 and Corollary 4.3.9 predict linear order of convergence for the state and adjoint

variable in the energy norm, nearly quadratic order of convergence for state and adjoint variables in L^2 norm, linear order of convergence for the control variable in L^2 norm, and a nearly quadratic order of convergence for the post-processed control. These nearly-quadratic convergence properties only occur in case of a super-convergence result for the state equation (i.e. Estimate (4.3.25)), which is always true for the FEMs but depends on some choice of points for the HMM scheme (see [54], and below).

Conforming FEM: The discrete solution is computed on several uniform grids with mesh sizes $h = \frac{1}{2^i}, i = 2, \ldots, 6$. The error estimates and the convergence rates of the control, the state and the adjoint variables are calculated. The post-processed control is also computed. Figure 4.4 displays the convergence history of the error on uniform meshes. As noticed in Remark 4.3.11 and as already seen in [96], we obtain linear order of convergence for the control and quadratic convergence for the post-processed control. The theoretical rates of convergence are confirmed by these numerical outputs.

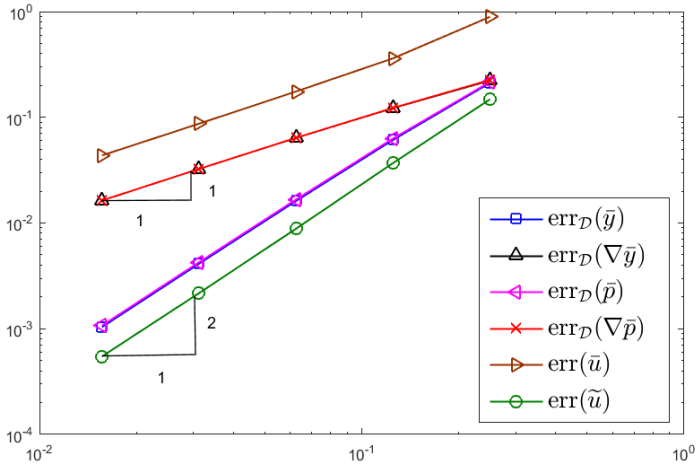

Figure 4.4: Dirichlet BC, example 1, conforming FEM

Non-Conforming FEM: For comparison, the solutions of the nc$\mathbb{P}_1$ finite element method on the same grids are computed. The errors of the numerical approximations to state, adjoint and control variables on uniform meshes are evaluated. The convergence behaviour of state, adjoint and control variables is illustrated in Figure 4.5. Here also, these outputs confirm the theoretical rates of convergence.

HMM scheme: In this section, the schemes were first tested on a series of regular triangle meshes from [74] (see Figure 4.6, left) where the points $\mathcal{P}$ (see [51, Definition 2.21]) are located at the

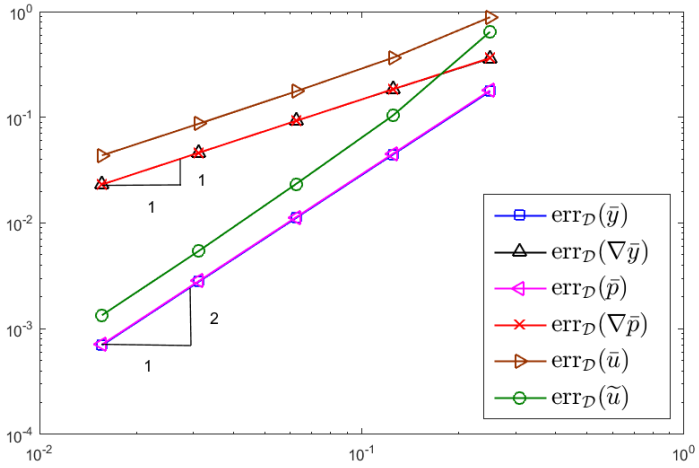

Figure 4.5: Dirichlet BC, example 1, non-conforming FEM

center of gravity of the cells (**Test1**). For such meshes, the state and adjoint equations enjoy a super-convergence property in L^2 norm [22, 54] and thus, as expected, so does the scheme for the entire control problem after projection of the exact control. In Figure 4.7, the graph of the relative errors corresponding to control, state and adjoint variables against the discretisation parameter is plotted in the loglog scale. **Test 2** focuses on a cartesian grid where the points $\mathcal{P}$ are shifted away from the centre of gravity (see Figure 4.6, right). For such a sequence of meshes, it has been observed in [54] that the HMM method can display a loss of superconvergence for the state equation. It is therefore expected that the same loss occurs, for all variables, for the control problem. This can be clearly seen in Figure 4.8.

Example 2

In this example, the results of numerical tests carried out for the L-shaped domain $\Omega = (-1,1)^2 \setminus ([0,1)\times(-1,0])$ are reported. The exact solutions are chosen as follows, and correspond to $\bar{u}_d = 0$.

$$\bar{y}(r,\theta) = \left(r^2\cos^2\theta - 1\right)\left(r^2\sin^2\theta - 1\right) r^{2/3} g(\theta), \quad \mathcal{U}_{\mathrm{ad}} = [-600,-50],$$
$$\alpha = 10^{-3}, \quad \bar{u} = P_{[-600,-50]}\left(-\frac{1}{\alpha}\bar{p}\right)$$

where $g(\theta) = (1-\cos\theta)(1+\sin\theta)$ and (r,θ) are the polar coordinates. The source term f and the desired state $\bar{y}_d$ can be determined using the above functions. The interest of this test-case is

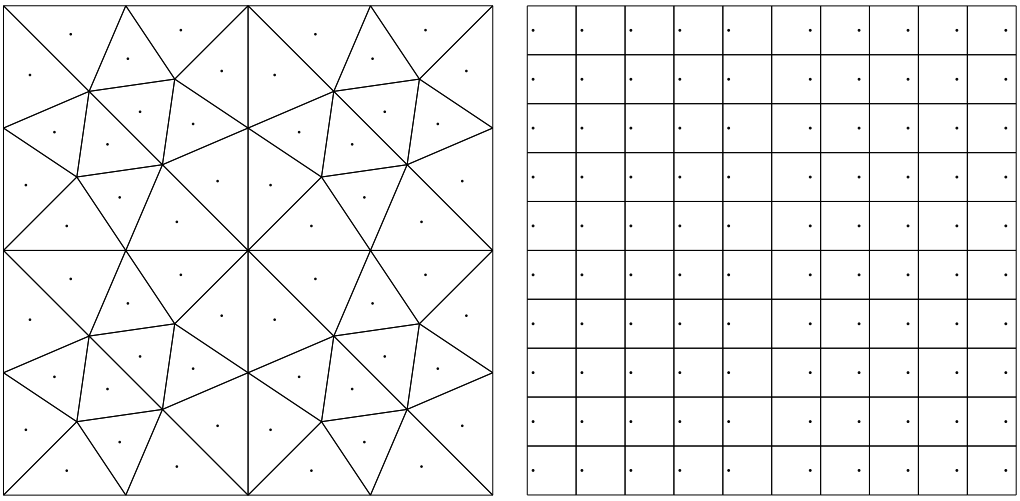

Figure 4.6: Mesh patterns for the tests using the HMM method (left: Test 1; right: Test 2).

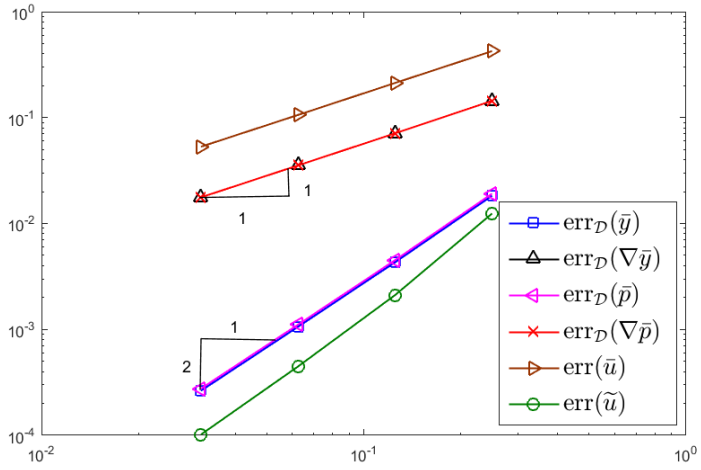

Figure 4.7: Dirichlet BC, example 1, HMM (**Test1**)

the loss of H^2-regularity property for the state and adjoint equations. Figure 4.9 shows the initial triangulation of a L-shape domain and its uniform refinement.
Conforming FEM: The errors in the energy norm and the L^2 norm together with their orders of convergence are evaluated. These numerical order of convergence clearly match the expected order of convergence, given the regularity property of the exact solutions. The convergence rates are plotted in the log-log scale in Figure 4.10.

Non-Conforming FEM: The errors between the true and computed solutions are computed for different mesh sizes. In Figure 4.11, the L^2-norm and H^1 norm of the error against the mesh

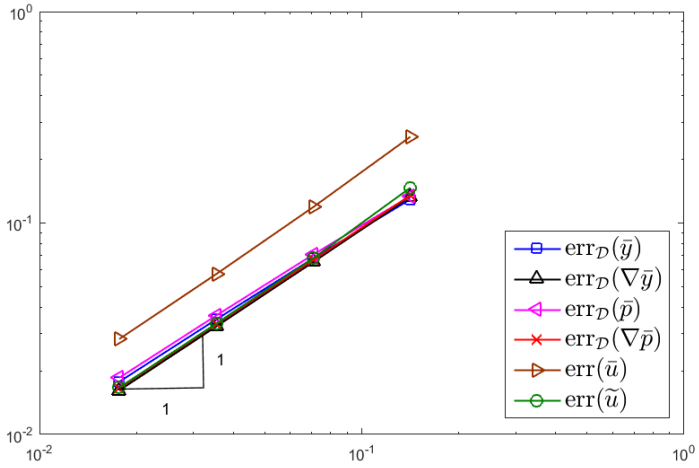

Figure 4.8: Dirichlet BC, example 1, HMM (**Test2**)

Figure 4.9: Initial triangulation and its uniform refinement

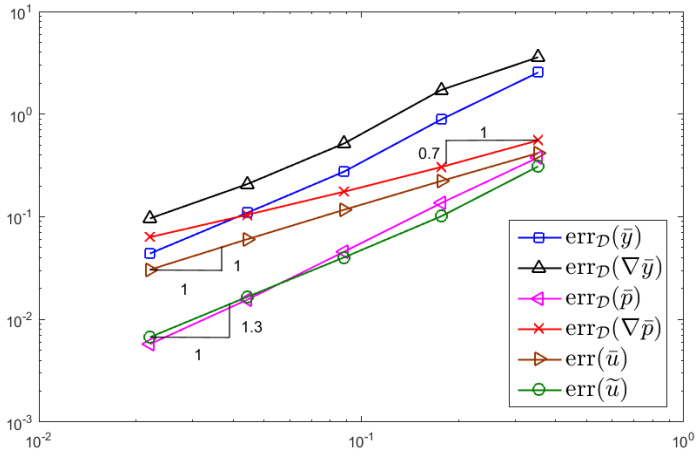

Figure 4.10: Dirichlet BC, example 2 (L-shaped domain), conforming FEM

parameter h are plotted.

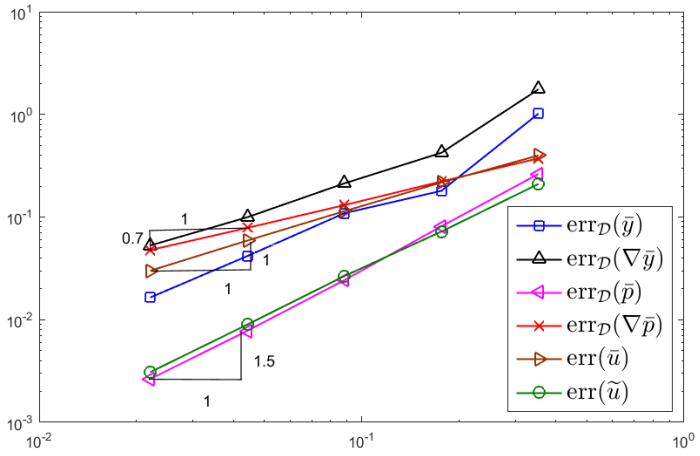

Figure 4.11: Dirichlet BC, example 2 (L-shaped domain), non-conforming FEM

HMM method: The errors corresponding to control, adjoint and state variables are computed using HMM (**Test 1**). In Figure 4.12, the graph of the errors are plotted against the mesh size h in the log-log scale.

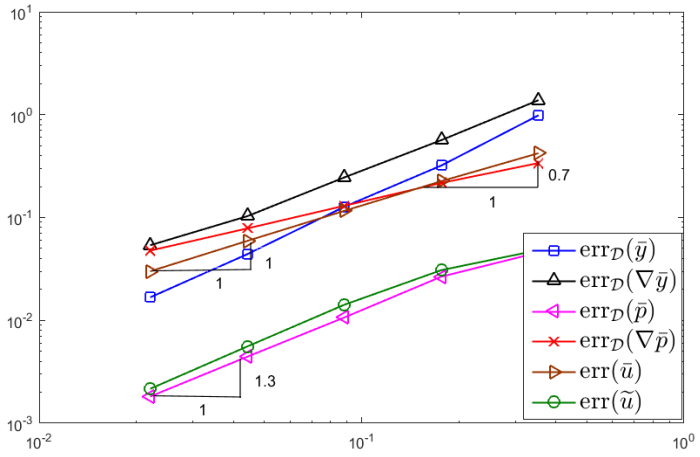

Figure 4.12: Dirichlet BC, example 2 (L-shaped domain), HMM

Since Ω is non-convex, we obtain only suboptimal orders of convergence for the state and adjoint variables in the energy norms and L^2 norms. Also we observe suboptimal order of convergence for the post processed control. However, the control converges at the optimal rate of h.

4.5.2 Neumann BC

In this example, consider the optimal control problem defined by (4.4.1) with $\Omega = (0,1)^2$ and $c_0 = 1$. Choose the exact state variable $\bar{y}$ and the adjoint variable $\bar{p}$ as

$$\begin{aligned} &\bar{y} = \frac{-1}{\pi}(\cos(\pi x) + \cos(\pi y)), \quad \bar{p} = \frac{-1}{\pi}(\cos(\pi x) + \cos(\pi y)), \\ &\mathcal{U}_{\text{ad}} = [-750, -50], \quad \alpha = 10^{-3}, \quad \bar{u}(\boldsymbol{x}) = P_{[-750,-50]}\left(-\frac{1}{\alpha}\bar{p}(\boldsymbol{x})\right). \end{aligned} \tag{4.5.1}$$

We therefore have $\bar{u}_d = 0$. The source term f and the observation $\bar{y}_d$ can be computed using

$$f = -\Delta\bar{y} + \bar{y} - \bar{u}, \quad \bar{y}_d = \bar{y} + \Delta\bar{p} - \bar{p}.$$

Conforming FEM: The errors and the orders of convergence for the control, state and adjoint variables are calculated for different mesh parameter h. The numerical errors are plotted against the discretisation parameter in the log-log scale in Figure 4.13.

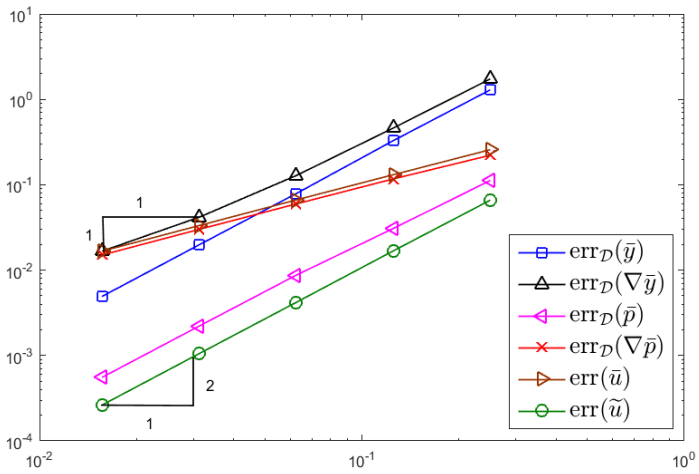

Figure 4.13: Neumann BC, test case corresponding to (4.5.1), conforming FEM

Non-Conforming FEM: The error estimates and the convergence rates of the control, the state and the adjoint variables are evaluated. The post-processed control is also computed. Figure 4.14 displays the convergence history of the error on uniform meshes.

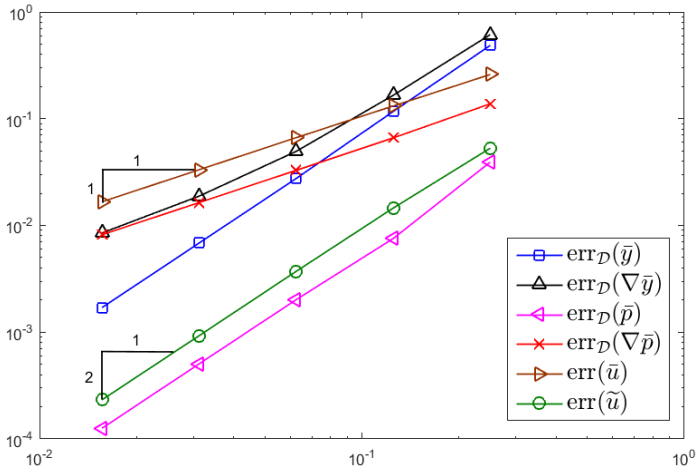

Figure 4.14: Neumann BC, test case corresponding to (4.5.1), non-conforming FEM

The observed orders of convergences agree with the predicted ones as seen in the figures.

Chapter 5

Approximation of pure Neumann control problems using the gradient discretisation method

The chapter discusses the GDM for distributed optimal control problems governed by diffusion equation with pure Neumann boundary condition[1]. Contrary to the control problem considered in Section 4.4, the state equation does not have a reaction term here. As a consequence, its wellposedness requires the imposition of an average condition on the solution, which in turns impacts the admissible controls and the relation between control and co-state variables.

5.1 Introduction

Consider the following distributed optimal control problem governed by the diffusion equation with Neumann boundary condition:

$$\min_{u\in\mathcal{U}_{\rm ad}} J(u) \ \text{ subject to} \tag{5.1.1a}$$

$$-\mathrm{div}(A\nabla y(u)) = u + f \ \text{ in } \Omega, \tag{5.1.1b}$$

$$A\nabla y(u)\cdot n = 0 \ \text{ on } \partial\Omega, \qquad \fint_\Omega y(u)\,\mathrm{d}\boldsymbol{x} = 0. \tag{5.1.1c}$$

Here, $\Omega \subset \mathbb{R}^d$ $(d \leq 3)$ is a bounded domain with boundary $\partial\Omega$ and n is the outer unit normal to Ω. The cost functional, dependent on the control variable u and the state variable $y(u)$, is given by

$$J(u) := \frac{1}{2}\|y(u) - \overline{y}_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \tag{5.1.2}$$

[1]The results of this chapter are published in *Jérôme Droniou, Neela Nataraj and Devika Shylaja. Numerical analysis for the pure Neumann control problem using the gradient discretisation method. Comput. Meth. Appl. Math. 18 (4), pp. 609-637, 2018. DOI: 10.1515/cmam-2017-0054. URL: https://arxiv.org/abs/ 1705.03256.*

with $\alpha > 0$ and $\fint_\Omega y(u)\,\mathrm{d}\boldsymbol{x} := \frac{1}{|\Omega|}\int_\Omega y(u)\,\mathrm{d}\boldsymbol{x}$ denotes the average value of the function $y(u)$ over Ω. The desired state variable $\bar{y}_d \in L^2(\Omega)$ is chosen to satisfy $\fint_\Omega \bar{y}_d\,\mathrm{d}\boldsymbol{x} = 0$. The source term $f \in L^2(\Omega)$ also satisfies the zero average condition $\fint_\Omega f\,\mathrm{d}\boldsymbol{x} = 0$. The diffusion matrix $A : \Omega \to \mathcal{M}_d(\mathbb{R})$ is a measurable, bounded and uniformly elliptic matrix-valued function such that $A(\boldsymbol{x})$ is symmetric for a.e. $\boldsymbol{x} \in \Omega$. Finally, the admissible set of controls $\mathcal{U}_{\text{ad}}$ is the non-empty convex set defined by

$$\mathcal{U}_{\text{ad}} = \left\{ u \in L^2(\Omega) \,:\, a \le u \le b \text{ and } \fint_\Omega u\,\mathrm{d}\boldsymbol{x} = 0 \right\}, \tag{5.1.3}$$

where a and b are constants in $[-\infty, +\infty]$ with $a < 0 < b$ (this condition is necessary to ensure that $\mathcal{U}_{\text{ad}}$ is not empty or reduced to $\{0\}$).

For Dirichlet BC, the super-convergence of post-processed controls for conforming finite element methods has been investigated in [96]. This result was extended to the GDM framework in the previous chapter for Dirichlet BC and Neumann BC with reaction term. For second order Neumann boundary value problems with reaction term (and hence without zero average constraint), see [5, 6, 30, 78, 93]. This chapter covers the more challenging case of pure Neumann BC without the reaction term.

One of the objectives in this chapter is to establish a projection relation between control and adjoint variables. This relation, which is non-standard since it has to account for the zero average constraints, is the key to prove the super-convergence result for all three variables. A modified active set strategy algorithm for GDM that is adapted to this non-standard projection relation is designed.

The chapter is organised as follows. Section 5.2 deals with the optimality conditions for (5.1.1). Section 5.3 recalls the GDM for elliptic problems with Neumann BC and the properties needed to prove its convergence. Section 5.4 deals with the GDM for the optimal control problem (5.1.1). The basic error estimates and super-convergence results are presented in Subsections 5.4.2 and 5.4.3. Discussions on post-processed controls and the projection relation between control and proper adjoint are presented in Subsection 5.4.4. The active set strategy is an algorithm to solve the non-linear Karush-Kuhn-Tucker (KKT) formulation of the optimal control problem [109]. Subsection 5.5.1 presents a modification of this algorithm that accounts for the zero average constraint on the control. This modified active set algorithm also automatically selects the proper discrete adjoint whose projection provides the discrete control variable. In Subsection 5.5.2, the results of some numerical experiments are presented.

5.2 Continuous control problem

The optimality conditions for (5.1.1) is discussed in this section. For a given $u \in \mathcal{U}_{\text{ad}}$, there exists a unique weak solution $y(u) \in H^1_\star(\Omega) := \{w \in H^1(\Omega) : \fint_\Omega w\,\mathrm{d}\boldsymbol{x} = 0\}$ of (5.1.1b)–(5.1.1c). That is, for $u \in \mathcal{U}_{\text{ad}}$, there exists a unique $y = y(u) \in H^1_\star(\Omega)$ such that for all $w \in H^1_\star(\Omega)$,

$$a(y, w) = \int_\Omega uw\,\mathrm{d}\boldsymbol{x}, \tag{5.2.1}$$

where $a(\phi,\psi) = \int_\Omega A\nabla\phi\cdot\nabla\psi\,\mathrm{d}\boldsymbol{x}$ for all $\phi,\psi\in H^1(\Omega)$. The term $y(u)$ is the state associated with the control u.
The convex control problem (5.1.1) has a unique solution $(\overline{y},\overline{u})\in H^1_\star(\Omega)\times\mathcal{U}_{\mathrm{ad}}$ and there exists a co-state $\overline{p}\in H^1(\Omega)$ such that the triplet $(\overline{y},\overline{p},\overline{u})\in H^1_\star(\Omega)\times H^1(\Omega)\times\mathcal{U}_{\mathrm{ad}}$ satisfies the KKT optimality conditions [90, Chapter 2]:

$$a(\overline{y},w) = (\overline{u}+f,w) \qquad \forall\, w\in H^1_\star(\Omega), \tag{5.2.2a}$$
$$a(z,\overline{p}) = (\overline{y}-\overline{y}_d,z) \qquad \forall\, z\in H^1(\Omega), \tag{5.2.2b}$$
$$(\overline{p}+\alpha\overline{u},v-\overline{u})\geq 0 \qquad \forall\, v\in\mathcal{U}_{\mathrm{ad}}. \tag{5.2.2c}$$

Several co-states satisfy the optimality conditions (5.2.2), as $\overline{p}$ is only determined up to an additive constant by (5.2.2). The same will be true for the discrete co-state, solution to a discrete version of these KKT equations. Establishing error estimates require the continuous and discrete co-states to have the same average. The usual choice is to fix this average as zero. However, for the control problem with pure Neumann conditions, this is not the best choice. Indeed, as seen in Lemma 5.4.9, establishing a proper relation between the control and co-state requires a certain zero average of a *non-linear* function of this co-state. A more efficient approach, that we will adopt, to fix the proper co-states is thus the following:

1. Design an algorithm (the modified active set algorithm of Subsection 5.5.1) that computes a discrete co-state with the proper condition, so that the discrete control can be easily obtained in terms of this discrete co-state,

2. Fix the average of the continuous co-state $\overline{p}$ to be the same as the average of the discrete co-state obtained above.

As we will see, an algebraic relation between this $\overline{p}$ and the continuous control $\overline{u}$ can still be written, upon selecting a proper (but non-explicit) translation of $\overline{p}$.

Remark 5.2.1 (Zero average constraint on the source term and desired state)**.**

(i) If we consider (5.1.1) *without the constraint* $\fint_\Omega f\,\mathrm{d}\boldsymbol{x}=0$ *on the source term, the set of admissible controls needs to be modified into*

$$\mathcal{U}_{\mathrm{ad}} = \left\{ u\in L^2(\Omega)\,:\, a\leq u\leq b \text{ and } \fint_\Omega (u+f)\,\mathrm{d}\boldsymbol{x}=0\right\}.$$

In this case, a simple transformation can bring us back to the case of a source term with zero average. Rewrite the state equation (5.1.1b) *as* $-\mathrm{div}(A\nabla y) = u^\star + f^\star$ *with* $u^\star = u + \fint_\Omega f\,\mathrm{d}\boldsymbol{x}$ *and* $f^\star = f - \fint_\Omega f\,\mathrm{d}\boldsymbol{x}$*. Then,* $\fint_\Omega f^\star\,\mathrm{d}\boldsymbol{x}=0$ *and* $u^\star\in\mathcal{U}^\star_{\mathrm{ad}}$ *where*

$$\mathcal{U}^\star_{\mathrm{ad}} = \left\{ u^\star\in L^2(\Omega)\,:\, a^\star\leq u^\star\leq b^\star \text{ and } \fint_\Omega u^\star\,\mathrm{d}\boldsymbol{x}=0\right\}$$

with $a^\star = a+\fint_\Omega f\,\mathrm{d}\boldsymbol{x}$ *and* $b^\star = b+\fint_\Omega f\,\mathrm{d}\boldsymbol{x}$*.*

(ii) If the desired state $\overline{y}_d \in L^2(\Omega)$ is such that $\fint_\Omega \overline{y}_d \,\mathrm{d}\boldsymbol{x} =: m \neq 0$, then it is natural to select states y in (5.1.1) *with the same average m (since the average of these states can be freely fixed, and the choice made in* (5.1.1c) *is arbitrary). This ensures the best possible approximation of the desired state $\overline{y}_d$. In that case, working with $y - m$ and $\overline{y}_d - m$ instead of y and $\overline{y}_d$ brings back to the original formulation* (5.1.1) *with a desired state $\overline{y}_d - m$ having a zero average.*

Remark 5.2.2 (Non-homogeneous BCs)**.** *The study of second order distributed control problem* (5.1.1) *with non-homogeneous boundary conditions $A\nabla y \cdot n = g$ on $\partial\Omega$ (with $g \in L^2(\partial\Omega)$) follows in a similar way. In this case, the source terms and boundary condition are supposed to satisfy the compatibility condition*

$$\int_\Omega f \,\mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} g \,\mathrm{d}s(\boldsymbol{x}) = 0.$$

The controls are still taken in $\mathcal{U}_{\mathrm{ad}}$ defined by (5.1.3) *and the KKT optimality condition is [90]: Seek $(\overline{y}, \overline{p}, \overline{u}) \in H^1_\star(\Omega) \times H^1(\Omega) \times \mathcal{U}_{\mathrm{ad}}$ such that*

$$\begin{aligned}
&a(\overline{y}, w) = (\overline{u} + f, w) + (g, \gamma(w))_\partial && \forall\, w \in H^1_\star(\Omega),\\
&a(z, \overline{p}) = (\overline{y} - \overline{y}_d, z) && \forall\, z \in H^1(\Omega),\\
&(\overline{p} + \alpha\overline{u}, v - \overline{u}) \geq 0 && \forall\, v \in \mathcal{U}_{\mathrm{ad}},
\end{aligned}$$

where $\gamma : H^1(\Omega) \to L^2(\partial\Omega)$ is the trace operator and $(\cdot,\cdot)_\partial$ is the inner product in $L^2(\partial\Omega)$.

5.3 GDM for elliptic PDE with Neumann BC

We presented the GDM for homogeneous Dirichlet BC in Section 4.2, and for Neumann BC with reaction term in Section 4.4.2. Here, it is shown that how the GDM is adapted to pure Neumann BC without reaction term.

5.3.1 Gradient discretisation and gradient scheme

A notion of gradient discretisation for Neumann BC is given in [48, Definition 3.1]. The following extends this definition by demanding the existence of the element $1_{\mathcal{D}}$ and is always satisfied in practical applications. This existence ensures that the zero average condition can be put in the discretisation space or in the bilinear form as for the continuous formulation, see Remark 5.3.2.

Definition 5.3.1 (Gradient discretisation for Neumann boundary conditions)**.** *A gradient discretisation (GD) for homogeneous Neumann boundary conditions is given by $\mathcal{D} = (X_{\mathcal{D}}, \Pi_{\mathcal{D}}, \nabla_{\mathcal{D}})$ such that*

- *$X_{\mathcal{D}}$ is a finite dimensional vector space on $\mathbb{R}$.*
- *$\Pi_{\mathcal{D}} : X_{\mathcal{D}} \to L^2(\Omega)$ and $\nabla_{\mathcal{D}} : X_{\mathcal{D}} \to L^2(\Omega)^d$ are linear mappings.*

- *The quantity*

$$\|w\|_{\mathcal{D}}^2 := \|\nabla_{\mathcal{D}} w\|^2 + \left| \fint_\Omega \Pi_{\mathcal{D}} w \,\mathrm{d}\boldsymbol{x} \right|^2 \tag{5.3.1}$$

  is a norm on $X_{\mathcal{D}}$.

- *There exists* $1_{\mathcal{D}} \in X_{\mathcal{D}}$ *such that* $\Pi_{\mathcal{D}} 1_{\mathcal{D}} = 1$ *on* Ω *and* $\nabla_{\mathcal{D}} 1_{\mathcal{D}} = 0$ *on* Ω.

If $F \in L^2(\Omega)$ is such that $\fint_\Omega F \,\mathrm{d}\boldsymbol{x} = 0$, the weak formulation of the Neumann boundary value problem

$$\begin{cases} -\mathrm{div}(A\nabla\psi) = F \text{ in } \Omega, \\ A\nabla\psi \cdot n = 0 \text{ on } \partial\Omega \end{cases} \tag{5.3.2}$$

is given by

$$\text{Find } \psi \in H^1_\star(\Omega) \text{ such that, for all } w \in H^1_\star(\Omega),\ a(\psi, w) = \int_\Omega F w \,\mathrm{d}\boldsymbol{x}. \tag{5.3.3}$$

As explained in Chapter 4, a gradient scheme for (5.3.2) is then obtained from a GD $\mathcal{D}$ by writing the weak formulation (5.3.3) with the continuous spaces, functions and gradients replaced with their discrete counterparts:

$$\begin{aligned} &\text{Find } \psi_{\mathcal{D}} \in X_{\mathcal{D},\star} \text{ such that, for all } w_{\mathcal{D}} \in X_{\mathcal{D},\star}, \\ &a_{\mathcal{D}}(\psi_{\mathcal{D}}, w_{\mathcal{D}}) = \int_\Omega F \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x}, \end{aligned} \tag{5.3.4}$$

where $a_{\mathcal{D}}(\phi_{\mathcal{D}}, z_{\mathcal{D}}) = \displaystyle\int_\Omega A\nabla_{\mathcal{D}}\phi_{\mathcal{D}} \cdot \nabla_{\mathcal{D}} z_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x}$, for all $\phi_{\mathcal{D}}, z_{\mathcal{D}} \in X_{\mathcal{D}}$, and

$$X_{\mathcal{D},\star} = \{ w_{\mathcal{D}} \in X_{\mathcal{D}} : \fint_\Omega \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x} = 0 \}.$$

Remark 5.3.2. *Owing to* $\fint_\Omega F \,\mathrm{d}\boldsymbol{x} = 0$*, the continuous formulation* (5.3.3) *is equivalent to*

$$\begin{aligned} &\textit{Find } \psi \in H^1(\Omega) \textit{ such that, for all } w \in H^1(\Omega), \\ &a(\psi, w) + \rho \left(\fint_\Omega \psi \,\mathrm{d}\boldsymbol{x} \right) \left(\fint_\Omega w \,\mathrm{d}\boldsymbol{x} \right) = \int_\Omega F w \,\mathrm{d}\boldsymbol{x} \end{aligned}$$

for any $\rho > 0$*. As for the continuous formulation, using the element* $1_{\mathcal{D}} \in X_{\mathcal{D}}$ *actually enables us to consider in* (5.3.4) *test functions* $w_{\mathcal{D}}$ *in* $X_{\mathcal{D}}$*, rather than just* $X_{\mathcal{D},\star}$*. The simplest technique to achieve this is to use a quadratic penalty method [69, Chapter 11]. Problem* (5.3.4) *can be shown equivalent to*

$$\begin{aligned} &\textit{Find } \psi_{\mathcal{D}} \in X_{\mathcal{D}} \textit{ such that, for all } w_{\mathcal{D}} \in X_{\mathcal{D}}, \\ &a_{\mathcal{D}}(\psi_{\mathcal{D}}, w_{\mathcal{D}}) + \rho \left(\fint_\Omega \Pi_{\mathcal{D}} \psi_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x} \right) \left(\fint_\Omega \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x} \right) = \int_\Omega F \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x}. \end{aligned} \tag{5.3.5}$$

Indeed, considering $w_{\mathcal{D}} = 1_{\mathcal{D}}$ *in* (5.3.5) *and recalling that* $\fint_\Omega F \,\mathrm{d}\boldsymbol{x} = 0$ *shows that the solution to this problem belongs to* $X_{\mathcal{D},\star}$ *and is therefore a solution to* (5.3.4)*. The converse is straightforward.*

5.3.2 Error estimates for the GDM for the Neumann problem

As for Dirichlet BC, the accuracy of a gradient scheme (5.3.4) is measured by three quantities. The first one is a discrete Poincaré–Wirtinger constant $C_{\mathcal{D}}$, which ensures the *coercivity* of the method.

$$C_{\mathcal{D}} := \max_{w \in X_{\mathcal{D}} \setminus \{0\}} \frac{\|\Pi_{\mathcal{D}} w\|}{\|w\|_{\mathcal{D}}}. \tag{5.3.6}$$

The second quantity is the interpolation error $S_{\mathcal{D}}$, which measures what is called, in the GDM framework, the *GD-consistency* of $\mathcal{D}$.

$$\forall \varphi \in H^1(\Omega),\ S_{\mathcal{D}}(\varphi) = \min_{w \in X_{\mathcal{D}}} \left(\|\Pi_{\mathcal{D}} w - \varphi\| + \|\nabla_{\mathcal{D}} w - \nabla \varphi\| \right). \tag{5.3.7}$$

Finally, the *limit-conformity* of a GD is measured by defining

$$\forall \boldsymbol{\varphi} \in H_0^{\mathrm{div}}(\Omega),\ W_{\mathcal{D}}(\boldsymbol{\varphi}) = \max_{w \in X_{\mathcal{D}} \setminus \{0\}} \frac{1}{\|w\|_{\mathcal{D}}} \left| \int_\Omega (\Pi_{\mathcal{D}} w \mathrm{div} \boldsymbol{\varphi} + \nabla_{\mathcal{D}} w \cdot \boldsymbol{\varphi}) \, \mathrm{d}\boldsymbol{x} \right|, \tag{5.3.8}$$

where $H_0^{\mathrm{div}}(\Omega) = \{\boldsymbol{\varphi} \in L^2(\Omega)^d : \mathrm{div}\boldsymbol{\varphi} \in L^2(\Omega),\ \gamma_n(\boldsymbol{\varphi}) = 0\}$ with γ_n being the normal trace of $\boldsymbol{\varphi}$ on $\partial\Omega$.

Using these quantities, an error estimate can be established for the GS (5.3.4). Recall the notation (4.2.9) from Chapter 4. Here

$$\begin{aligned} &X \lesssim Y \text{ means that } X \leq CY \text{ for some } C \text{ depending} \\ &\text{only on } \Omega, A \text{ and an upper bound of } C_{\mathcal{D}} \text{ defined by (5.3.6).} \end{aligned} \tag{5.3.9}$$

Theorem 5.3.3 (Error estimate for the GDM [48])**.** *Let $\mathcal{D}$ be a GD in the sense of Definition 5.3.1, let ψ be the solution to* (5.3.3)*, and let $\psi_{\mathcal{D}}$ be the solution to* (5.3.4)*. Then*

$$\|\Pi_{\mathcal{D}} \psi_{\mathcal{D}} - \psi\| + \|\nabla_{\mathcal{D}} \psi_{\mathcal{D}} - \nabla \psi\| \lesssim \mathrm{WS}_{\mathcal{D}}(\psi), \tag{5.3.10}$$

where

$$\mathrm{WS}_{\mathcal{D}}(\psi) = W_{\mathcal{D}}(A\nabla\psi) + S_{\mathcal{D}}(\psi). \tag{5.3.11}$$

Remark 5.3.4 (Rates of convergence)**.** *For all classical low order methods based on meshes (such as $\mathbb{P}_1$ conforming and non-conforming finite element methods, finite volume methods, etc.), if A is Lipschitz continuous and $\psi \in H^2(\Omega)$ then $\mathcal{O}(h)$ estimates can be obtained for $W_{\mathcal{D}}(A\nabla\psi)$ and $S_{\mathcal{D}}(\psi)$ [48]. Theorem 5.3.3 then gives a linear rate of convergence for these methods.*

Remark 5.3.5. *Note that Theorem 5.3.3 also holds if the zero average condition on ψ and $\Pi_{\mathcal{D}}\psi_{\mathcal{D}}$ is replaced with $\int_\Omega \Pi_{\mathcal{D}}\psi_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} = \int_\Omega \psi \, \mathrm{d}\boldsymbol{x}$. In this case, the estimate* (5.3.10) *can be obtained by considering the translation of $\psi_{\mathcal{D}}$ and ψ. Set $\widetilde{\psi}_{\mathcal{D}} = \psi_{\mathcal{D}} - c1_{\mathcal{D}}$ and $\widetilde{\psi} = \psi - c1$, where $c = \int_\Omega \Pi_{\mathcal{D}}\psi_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} = \int_\Omega \psi \, \mathrm{d}\boldsymbol{x}$ and 1 is the constant function. Use Definition 5.3.1 to obtain $\Pi_{\mathcal{D}}\widetilde{\psi}_{\mathcal{D}} = \Pi_{\mathcal{D}}\psi_{\mathcal{D}} - c$, $\nabla_{\mathcal{D}}\widetilde{\psi}_{\mathcal{D}} = \nabla_{\mathcal{D}}\psi_{\mathcal{D}}$ and $\nabla\widetilde{\psi} = \nabla\psi$. This gives $\int_\Omega \Pi_{\mathcal{D}}\widetilde{\psi}_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} = \int_\Omega \widetilde{\psi} \, \mathrm{d}\boldsymbol{x} = 0$. Applying Theorem 5.3.3,*

$$\|\Pi_{\mathcal{D}}\widetilde{\psi}_{\mathcal{D}} - \widetilde{\psi}\| + \|\nabla_{\mathcal{D}}\widetilde{\psi}_{\mathcal{D}} - \nabla\widetilde{\psi}\| \lesssim \mathrm{WS}_{\mathcal{D}}(\widetilde{\psi})$$

which implies

$$\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - \psi\| + \|\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla\psi\| \lesssim \mathrm{WS}_{\mathcal{D}}(\widetilde{\psi}) = \mathrm{WS}_{\mathcal{D}}(\psi).$$

The following stability result, useful to the analysis, is straightforward.

Proposition 5.3.6 (Stability of the GDM)**.** *Let $\underline{a}$ be a coercivity constant of A. If $\psi_{\mathcal{D}}$ is the solution to the gradient scheme* (5.3.4)*, then*

$$\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}}\| \le \frac{C_{\mathcal{D}}}{\underline{a}}\|F\| \quad \text{and} \quad \|\Pi_{\mathcal{D}}\psi_{\mathcal{D}}\| \le \frac{C_{\mathcal{D}}^2}{\underline{a}}\|F\|. \tag{5.3.12}$$

Proof. Choose $w_{\mathcal{D}} = \psi_{\mathcal{D}}$ in (5.3.4) and use the definition of $C_{\mathcal{D}}$ to write

$$\underline{a}\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}}\|^2 \le \|F\|\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}}\| \le C_{\mathcal{D}}\|F\|\|\psi_{\mathcal{D}}\|_{\mathcal{D}}.$$

Since $\int_\Omega \Pi_{\mathcal{D}}\psi_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} = 0$, recalling the Definition (5.3.1) of $\|\cdot\|_{\mathcal{D}}$ shows that $\|\psi_{\mathcal{D}}\|_{\mathcal{D}} = \|\nabla_{\mathcal{D}}\psi_{\mathcal{D}}\|$ and the proof of first estimate is complete. The second estimate follows from the definition of $C_{\mathcal{D}}$. □

5.4 GDM for the control problem and main results

This section starts with a description of GDM for the optimal control problem and is followed by the basic error estimates and super-convergence results in Subsections 5.4.2 and 5.4.3. A discussion on post-processed controls and the projection relation between control and proper adjoint is presented in Subsection 5.4.4.

5.4.1 GDM for the optimal control problem

Let $\mathcal{D}$ be a GD as in Definition 5.3.1. The space $\mathcal{U}_h$ is defined as the space of piecewise constant functions on a mesh $\mathcal{M}$ of Ω. The space $\mathcal{U}_{\mathrm{ad},h} = \mathcal{U}_{\mathrm{ad}} \cap \mathcal{U}_h$ is a finite dimensional subset of $\mathcal{U}_{\mathrm{ad}}$. A gradient scheme for (5.2.2) consists in seeking $(\overline{y}_{\mathcal{D}}, \overline{p}_{\mathcal{D}}, \overline{u}_h) \in X_{\mathcal{D},\star} \times X_{\mathcal{D}} \times \mathcal{U}_{\mathrm{ad},h}$, such that

$$a_{\mathcal{D}}(\overline{y}_{\mathcal{D}}, w_{\mathcal{D}}) = (\overline{u}_h + f, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall\, w_{\mathcal{D}} \in X_{\mathcal{D},\star}, \tag{5.4.1a}$$

$$a_{\mathcal{D}}(z_{\mathcal{D}}, \overline{p}_{\mathcal{D}}) = (\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \overline{y}_d, \Pi_{\mathcal{D}} z_{\mathcal{D}}) \qquad \forall\, z_{\mathcal{D}} \in X_{\mathcal{D}}, \tag{5.4.1b}$$

$$(\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} + \alpha\overline{u}_h, v_h - \overline{u}_h) \ge 0 \qquad \forall\, v_h \in \mathcal{U}_{\mathrm{ad},h}. \tag{5.4.1c}$$

As in the continuous KKT conditions (5.2.2), these equations do not define $\overline{p}_{\mathcal{D}}$ uniquely. One possible constraint that fixes a unique $\overline{p}_{\mathcal{D}}$ is described in Lemma 5.4.9. This particular choice ensures a simple projection relation between $\overline{p}_{\mathcal{D}}$ and $\overline{u}_h$.

As in Chapter 4, two projection operators play a major role throughout this chapter: the orthogonal projection on piecewise constant functions on $\mathcal{M}$, namely $\mathcal{P}_{\mathcal{M}} : L^1(\Omega) \to \mathcal{U}_h$ and the cut-off function $P_{[a,b]} : \mathbb{R} \to [a,b]$. Recall from Section 4.3.2 that

$$\forall v \in L^1(\Omega), \forall K \in \mathcal{M}, \quad (\mathcal{P}_{\mathcal{M}} v)_{|K} := \fint_K v\,\mathrm{d}\boldsymbol{x},$$

$$\forall s \in \mathbb{R}, \quad P_{[a,b]}(s) := \min(b, \max(a,s)). \tag{5.4.2}$$

5.4.2 Basic error estimate for the GDM for the control problem

The proofs of the basic error estimates follow by adapting the corresponding proofs in Chapter 4 to account for the pure Neumann boundary conditions and integral constraints. For the sake of completeness and readability, we provide here detailed proofs, highlighting in chosen places where modifications are required due to the pure Neumann boundary conditions (which mostly amount to making sure that certain averages have been properly fixed).

Theorem 5.4.1 (Control estimate)**.** *Let $\mathcal{D}$ be a GD, $(\overline{y},\overline{p},\overline{u})$ be a solution to* (5.2.2) *and $(\overline{y}_\mathcal{D},\overline{p}_\mathcal{D},\overline{u}_h)$ be a solution to* (5.4.1) *such that $\fint_\Omega \Pi_\mathcal{D}\overline{p}_\mathcal{D}\,\mathrm{d}\boldsymbol{x} = \fint_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$. Then, recalling* (5.3.9)*,* (5.3.11) *and* (4.3.4)*, there exists a constant C depending only on α such that*

$$\|\overline{u}-\overline{u}_h\| \lesssim C\left(E_h(\overline{p})+E_h(\overline{u})+\mathrm{WS}_\mathcal{D}(\overline{p})+\mathrm{WS}_\mathcal{D}(\overline{y})\right), \tag{5.4.3}$$

where the projection error E_h is defined by (4.3.4)*.*

Proof. Define the following auxiliary discrete problem:

$$\begin{aligned}
&\text{Seek } (y_\mathcal{D}(\overline{u}),p_\mathcal{D}(\overline{u}))\in X_{\mathcal{D},\star}\times X_\mathcal{D} \text{ such that}\\
&a_\mathcal{D}(y_\mathcal{D}(\overline{u}),w_\mathcal{D}) = (f+\overline{u},\Pi_\mathcal{D}w_\mathcal{D}) \qquad \forall\, w_\mathcal{D}\in X_{\mathcal{D},\star}, &\text{(5.4.4a)}\\
&a_\mathcal{D}(z_\mathcal{D},p_\mathcal{D}(\overline{u})) = (\overline{y}-\overline{y}_d,\Pi_\mathcal{D}z_\mathcal{D}) \qquad \forall\, z_\mathcal{D}\in X_\mathcal{D}, &\text{(5.4.4b)}
\end{aligned}$$

where the co-state $p_\mathcal{D}(\overline{u})$ is chosen such that $\fint_\Omega \Pi_\mathcal{D}p_\mathcal{D}(\overline{u})\,\mathrm{d}\boldsymbol{x} = \fint_\Omega \Pi_\mathcal{D}\overline{p}_\mathcal{D}\,\mathrm{d}\boldsymbol{x} = \fint_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$. For Neumann boundary conditions, this particular choice is essential as it ensures that $p_\mathcal{D}(\overline{u})-\overline{p}_\mathcal{D}\in X_{\mathcal{D},\star}$ can be used as a test function $w_\mathcal{D}$ in (5.4.1a) and (5.4.4a). Recall that $\boldsymbol{\mathcal{P}}_\mathcal{M}$ is the orthogonal projection on piecewise constant functions on $\mathcal{M}$. This gives $\boldsymbol{\mathcal{P}}_\mathcal{M}(\mathcal{U}_\mathrm{ad})\subset\mathcal{U}_h$. Also, for $u\in\mathcal{U}_\mathrm{ad}$ and $K\in\mathcal{M}$, $\boldsymbol{\mathcal{P}}_\mathcal{M}u_{|K} = \fint_K u\,\mathrm{d}\boldsymbol{x}\in[a,b]$ and, using (5.1.3) we also see that

$$\fint_\Omega \boldsymbol{\mathcal{P}}_\mathcal{M}u\,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{M}}\fint_K \boldsymbol{\mathcal{P}}_\mathcal{M}u\,\mathrm{d}\boldsymbol{x} = \sum_{K\in\mathcal{M}}\fint_K u\,\mathrm{d}\boldsymbol{x} = \fint_\Omega u\,\mathrm{d}\boldsymbol{x} = 0.$$

Hence, $\boldsymbol{\mathcal{P}}_\mathcal{M}(\mathcal{U}_\mathrm{ad})\subset\mathcal{U}_{\mathrm{ad},h}$.
Set $P_{\mathcal{D},\alpha}(\overline{u}) = \alpha^{-1}\Pi_\mathcal{D}p_\mathcal{D}(\overline{u})$, $\overline{P}_{\mathcal{D},\alpha} = \alpha^{-1}\Pi_\mathcal{D}\overline{p}_\mathcal{D}$ and $\overline{P}_\alpha = \alpha^{-1}\overline{p}$. Since $\overline{u}_h\in\mathcal{U}_{\mathrm{ad},h}\subset\mathcal{U}_\mathrm{ad}$ and $\boldsymbol{\mathcal{P}}_\mathcal{M}\overline{u}\in\mathcal{U}_{\mathrm{ad},h}$, from the optimality conditions ((5.2.2c) and (5.4.1c)),

$$-\alpha(\overline{P}_\alpha+\overline{u},\overline{u}-\overline{u}_h)\geq 0, \quad \alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h,\overline{u}-\overline{u}_h)\geq\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h,\overline{u}-\boldsymbol{\mathcal{P}}_\mathcal{M}\overline{u}).$$

Add these two inequalities to obtain

$$\begin{aligned}
\alpha\|\overline{u}-\overline{u}_h\|^2 \leq\ & -\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h,\overline{u}-\boldsymbol{\mathcal{P}}_\mathcal{M}(\overline{u}))+\alpha(\overline{P}_{\mathcal{D},\alpha}-\overline{P}_\alpha,\overline{u}-\overline{u}_h)\\
=\ & -\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h,\overline{u}-\boldsymbol{\mathcal{P}}_\mathcal{M}\overline{u})+\alpha(\overline{P}_{\mathcal{D},\alpha}-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h)\\
& -\alpha(\overline{P}_\alpha-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h).
\end{aligned}\tag{5.4.5}$$

By orthogonality property of $\boldsymbol{\mathcal{P}}_{\mathcal{M}}$, $(\overline{u}_h, \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) = 0$ and $(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{P}_\alpha, \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) = 0$. Therefore, the first term in the right-hand side of (5.4.5) can be re-cast as

$$\begin{aligned} -\alpha(\overline{P}_{\mathcal{D},\alpha} + \overline{u}_h, \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) = & -\alpha(\overline{P}_\alpha - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{P}_\alpha, \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) + \alpha(\overline{P}_\alpha - P_{\mathcal{D},\alpha}(\overline{u}), \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) \\ & + \alpha(P_{\mathcal{D},\alpha}(\overline{u}) - \overline{P}_{\mathcal{D},\alpha}, \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}). \end{aligned} \tag{5.4.6}$$

By the Cauchy–Schwarz inequality, the first term on the right hand side of (5.4.6) is estimated as

$$-\alpha(\overline{P}_\alpha - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{P}_\alpha, \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) \leq E_h(\overline{p})E_h(\overline{u}). \tag{5.4.7}$$

Equation (5.4.4b) shows that $p_{\mathcal{D}}(\overline{u})$ is the solution of the GS corresponding to the adjoint problem (5.2.2b), whose solution is $\overline{p}$. Therefore, use the fact that $\fint_\Omega \Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u})\,\mathrm{d}\boldsymbol{x} = \fint_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$ (note that the specific relation between the continuous and discrete co-states is essential here) and Theorem 5.3.3 to deduce

$$\|\overline{P}_\alpha - P_{\mathcal{D},\alpha}(\overline{u})\| = \alpha^{-1}\|\overline{p} - \Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u})\| \lesssim \alpha^{-1}\mathrm{WS}_{\mathcal{D}}(\overline{p}). \tag{5.4.8}$$

Hence, a use of the Cauchy–Schwarz inequality implies

$$\alpha(\overline{P}_\alpha - P_{\mathcal{D},\alpha}(\overline{u}), \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) \lesssim \mathrm{WS}_{\mathcal{D}}(\overline{p})E_h(\overline{u}). \tag{5.4.9}$$

Use the definitions of $C_{\mathcal{D}}$, $\|\cdot\|_{\mathcal{D}}$ and the fact that $p_{\mathcal{D}}(\overline{u}) - \overline{p}_{\mathcal{D}} \in X_{\mathcal{D},\star}$ to write

$$\|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\|^2 \lesssim \|p_{\mathcal{D}}(\overline{u}) - \overline{p}_{\mathcal{D}}\|_{\mathcal{D}}^2 = \|\nabla_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}\|^2. \tag{5.4.10}$$

By writing the difference of (5.4.4b) and (5.4.1b) we see that $p_{\mathcal{D}}(\overline{u}) - \overline{p}_{\mathcal{D}}$ is the solution to the GS (5.3.4) with source term $F = \overline{y} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}$. i.e, for all $z_{\mathcal{D}} \in X_{\mathcal{D}}$

$$a_{\mathcal{D}}(z_{\mathcal{D}}, p_{\mathcal{D}}(\overline{u}) - \overline{p}_{\mathcal{D}}) = (\overline{y} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}, \Pi_{\mathcal{D}} z_{\mathcal{D}}).$$

Choose $z_{\mathcal{D}} = p_{\mathcal{D}}(\overline{u}) - \overline{p}_{\mathcal{D}}$ in the above equality and use it in (5.4.10) to obtain

$$\|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\|^2 \lesssim \|\nabla_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}\|^2 \lesssim \|\overline{y} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\|\,\|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\|.$$

As a consequence,

$$\begin{aligned} \|P_{\mathcal{D},\alpha}(\overline{u}) - \overline{P}_{\mathcal{D},\alpha}\| &= \alpha^{-1}\|\Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| \lesssim \alpha^{-1}\|\overline{y} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| \\ &\lesssim \alpha^{-1}\|\overline{y} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| + \alpha^{-1}\|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\|. \end{aligned}$$

Use Theorem 5.3.3 with $\psi = \overline{y}$ to bound the first term in the above expression. This along with an application of Young's inequality yields an estimate for the last term in (5.4.6) as

$$\alpha(P_{\mathcal{D},\alpha}(\overline{u}) - \overline{P}_{\mathcal{D},\alpha}, \overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) \leq C_1 E_h(\overline{u})\mathrm{WS}_{\mathcal{D}}(\overline{y}) + C_1 E_h(\overline{u})^2 + \frac{1}{4}\|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\|^2, \tag{5.4.11}$$

where C_1 depends only on Ω, A and an upper bound of $C_{\mathcal{D}}$. A substitution of (5.4.7), (5.4.9) and (5.4.11) in (5.4.6) yields

$$-\alpha(\overline{P}_{\mathcal{D},\alpha}+\overline{u}_h,\overline{u}-\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}) \le E_h(\overline{p})E_h(\overline{u})+C_2E_h(\overline{u})\mathrm{WS}_{\mathcal{D}}(\overline{p})+C_1E_h(\overline{u})\mathrm{WS}_{\mathcal{D}}(\overline{y}) \\ +C_1E_h(\overline{u})^2+\frac{1}{4}\|\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})-\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\|^2, \tag{5.4.12}$$

where C_2 is the hidden constant in (5.4.9). Let us turn to the second term in the right-hand side of (5.4.5). From (5.4.1b) and (5.4.4b), for all $z_{\mathcal{D}} \in X_{\mathcal{D}}$,

$$a_{\mathcal{D}}(z_{\mathcal{D}},\overline{p}_{\mathcal{D}}-p_{\mathcal{D}}(\overline{u})) = (\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\overline{y},\Pi_{\mathcal{D}}z_{\mathcal{D}}). \tag{5.4.13}$$

Also, from (5.4.1a) and (5.4.4a), for all $w_{\mathcal{D}} \in X_{\mathcal{D},\star}$,

$$a_{\mathcal{D}}(\overline{y}_{\mathcal{D}}-y_{\mathcal{D}}(\overline{u}),w_{\mathcal{D}}) =(\overline{u}_h-\overline{u},\Pi_{\mathcal{D}}w_{\mathcal{D}}). \tag{5.4.14}$$

Choose $z_{\mathcal{D}}=\overline{y}_{\mathcal{D}}-y_{\mathcal{D}}(\overline{u}) \in X_{\mathcal{D}}$ in (5.4.13), $w_{\mathcal{D}}=\overline{p}_{\mathcal{D}}-p_{\mathcal{D}}(\overline{u}) \in X_{\mathcal{D},\star}$ in (5.4.14), use the symmetry of $a_{\mathcal{D}}(\cdot,\cdot)$, Theorem 5.3.3 with $\psi=\overline{y}$ and Young's inequality to obtain

$$\begin{aligned}\alpha(\overline{P}_{\mathcal{D},\alpha}-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h) =& -(\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\overline{y},\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}))\\ &= (\overline{y}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}),\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}))-\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|^2\\ &\lesssim \mathrm{WS}_{\mathcal{D}}(\overline{y})\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|-\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|^2\\ &\le C_3\mathrm{WS}_{\mathcal{D}}(\overline{y})^2+\frac{1}{4}\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|^2-\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|^2,\end{aligned} \tag{5.4.15}$$

where C_3 only depends on Ω, A and an upper bound of $C_{\mathcal{D}}$. The last term in the right hand side of (5.4.5) can be estimated using (5.4.8) and Young's inequality:

$$-\alpha(\overline{P}_{\alpha}-P_{\mathcal{D},\alpha}(\overline{u}),\overline{u}-\overline{u}_h) \le \frac{\alpha}{2}\|\overline{u}-\overline{u}_h\|^2+C_4\mathrm{WS}_{\mathcal{D}}(\overline{p})^2, \tag{5.4.16}$$

where C_4 only depends on Ω, A, α and an upper bound of $C_{\mathcal{D}}$. Substitute (5.4.12), (5.4.15) and (5.4.16) into (5.4.5), apply the Young's inequality and $\sqrt{\sum_i a_i^2} \le \sum_i a_i$ to complete the proof. □

Proposition 5.4.2 (State and adjoint error estimates)**.** *Let $\mathcal{D}$ be a GD, $(\overline{y},\overline{p},\overline{u})$ be a solution to* (5.2.2) *and $(\overline{y}_{\mathcal{D}},\overline{p}_{\mathcal{D}},\overline{u}_h)$ be a solution to* (5.4.1)*. Assume that $\int_\Omega \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} = \int_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$. Then the following error estimates hold:*

$$\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\overline{y}\|+\|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\nabla\overline{y}\| \lesssim \|\overline{u}-\overline{u}_h\|+\mathrm{WS}_{\mathcal{D}}(\overline{y}), \tag{5.4.17}$$

$$\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}-\overline{p}\|+\|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}-\nabla\overline{p}\| \lesssim \|\overline{u}-\overline{u}_h\|+\mathrm{WS}_{\mathcal{D}}(\overline{y})+\mathrm{WS}_{\mathcal{D}}(\overline{p}). \tag{5.4.18}$$

Proof. The triangle inequality leads to

$$\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\overline{y}\|+\|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\nabla\overline{y}\| \le \|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|+\|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}}-\nabla_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\| \\ +\|\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})-\overline{y}\|+\|\nabla_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})-\nabla\overline{y}\|. \tag{5.4.19}$$

Subtract (5.4.1a) and (5.4.4a), and use the stability property of GS (Proposition 5.3.6) to estimate the first two terms in the right hand side of the above inequality as

$$\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| + \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla_{\mathcal{D}} y_{\mathcal{D}}(\overline{u})\| \lesssim \|\overline{u} - \overline{u}_h\|.$$

The last two terms on the right hand side of the (5.4.19) are estimated using Theorem 5.3.3 as

$$\|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \overline{y}\| + \|\nabla_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \nabla\overline{y}\| \lesssim \mathrm{WS}_{\mathcal{D}}(\overline{y}).$$

A combination of the above two results yields the error estimates (5.4.17) for the state variable. A use of $\int_\Omega \Pi_{\mathcal{D}} p_{\mathcal{D}}(\overline{u})\,\mathrm{d}\boldsymbol{x} = \int_\Omega \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}$ in Proposition 5.3.6 leads to the error estimates for the adjoint variable in a similar way. □

Remark 5.4.3 (Rates of convergence for the control problem)**.** *As in Remark 5.3.4, if A is Lipschitz continuous and $(\overline{y},\overline{p},\overline{u}) \in H^2(\Omega)^2 \times H^1(\Omega)$ then* (5.4.3)*,* (5.4.17) *and* (5.4.18) *give linear rates of convergence for all classical first-order methods.*

5.4.3 Super-convergence for post-processed controls

In this subsection, the post-processed continuous and discrete controls (see (5.4.23)) are defined and super-convergence results are established.
We make here the following assumptions, similar to the assumptions in Section 4.3.2, taking into account for the pure Neumann BC and zero average constraint.

(A1) [*Interpolation operator*] For each $w \in H^2(\Omega)$, there exists $w_{\mathcal{M}} \in L^2(\Omega)$ such that:
i) If $w \in H^2(\Omega)$ solves $-\mathrm{div}(A\nabla w) = g \in H^1(\Omega)$, and $w_{\mathcal{D}}$ is the solution to the corresponding GS with $\int_\Omega \Pi_{\mathcal{D}} w_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} = \int_\Omega w\,\mathrm{d}\boldsymbol{x}$, then

$$\|\Pi_{\mathcal{D}} w_{\mathcal{D}} - w_{\mathcal{M}}\| \lesssim h^2 \|g\|_{H^1(\Omega)}. \tag{5.4.20}$$

ii) For any $w \in H^2(\Omega)$, it holds

$$\forall v_{\mathcal{D}} \in X_{\mathcal{D}},\ \left|(w - w_{\mathcal{M}}, \Pi_{\mathcal{D}} v_{\mathcal{D}})\right| \lesssim h^2 \|w\|_{H^2(\Omega)} \|\Pi_{\mathcal{D}} v_{\mathcal{D}}\|, \tag{5.4.21}$$

$$\|\boldsymbol{\mathcal{P}}_{\mathcal{M}}(w - w_{\mathcal{M}})\| \lesssim h^2 \|w\|_{H^2(\Omega)}. \tag{5.4.22}$$

(A2) The estimate $\|\Pi_{\mathcal{D}} v_{\mathcal{D}} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}} v_{\mathcal{D}})\| \lesssim h\|\nabla_{\mathcal{D}} v_{\mathcal{D}}\|$ holds for any $v_{\mathcal{D}} \in X_{\mathcal{D}}$.

(A3) [*Discrete Sobolev imbedding*] For all $v_{\mathcal{D}} \in X_{\mathcal{D}}$, it holds

$$\|\Pi_{\mathcal{D}} v_{\mathcal{D}}\|_{L^{2^*}(\Omega)} \lesssim \|v_{\mathcal{D}}\|_{\mathcal{D}},$$

where 2^* is a Sobolev exponent of 2, that is, $2^* \in [2,\infty)$ if $d = 2$, and $2^* = 6$ if $d = 3$.

Let

$$\mathcal{M}_2 = \{K \in \mathcal{M} : \overline{u} = a \text{ a.e. on } K, \text{ or } \overline{u} = b \text{ a.e. on } K, \text{ or } a < \overline{u} < b \text{ a.e. on } K\},$$

and $\mathcal{M}_1 = \mathcal{M} \setminus \mathcal{M}_2$. That is, $\mathcal{M}_1$ is the set of cells where $\overline{u}$ crosses at least one constraint a or b. For $i = 1, 2$, let $\Omega_{i,\mathcal{M}} = \mathrm{int}(\cup_{K \in \mathcal{M}_i} \overline{K})$. The space $W^{1,\infty}(\mathcal{M}_1)$ is the usual broken Sobolev space, endowed with its broken norm. The last assumption is:

(A4) $|\Omega_{1,\mathcal{M}}| \lesssim h$ and $\overline{u}_{|\Omega_{1,\mathcal{M}}} \in W^{1,\infty}(\mathcal{M}_1)$.

Note that the assumptions **(A1)–(A4)** are similar to that in Chapter 4 with $X_{\mathcal{D},0}$ substituted by $X_{\mathcal{D}}$, and an additional average condition in **(A1)**. See Chapter 4 for a detailed discussion on **(A1)–(A4)**.

Assuming $\overline{p} \in H^2(\Omega)$ (see Theorem 5.4.4) and letting $\overline{p}_{\mathcal{M}}$ be defined as in **(A1)**, the post-processed continuous and discrete controls are given by

$$\widetilde{u}(\boldsymbol{x}) = P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p}_{\mathcal{M}}(\boldsymbol{x})\right) \quad \text{and} \quad \widetilde{u}_h(\boldsymbol{x}) = P_{[a,b]}\left(-\frac{1}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}(\boldsymbol{x})\right). \tag{5.4.23}$$

For a detailed discussion on the post-processed controls, we refer the reader to Subsection 5.4.4. We use the following extension of the notation (5.3.9):

$$\begin{aligned} &X \lesssim_\eta Y \text{ means that } X \le CY \text{ for some } C \text{ depending} \\ &\text{only on } \Omega, A, \text{ an upper bound of } C_{\mathcal{D}}, \text{ and } \eta. \end{aligned}$$

Theorem 5.4.4 (Super-convergence for post-processed controls I)**.** *Let $\mathcal{D}$ be a GD and $\mathcal{M}$ be a mesh. Assume that*

- **(A1)–(A4)** *hold,*
- $\overline{y}$ *and* $\overline{p}$ *belong to* $H^2(\Omega)$*,*
- $\overline{y}_d$ *and* f *belong to* $H^1(\Omega)$*,*

and let $\widetilde{u}$, $\widetilde{u}_h$ be the post-processed controls defined by (5.4.23) *where $\overline{p}$ and $\overline{p}_{\mathcal{D}}$ are chosen such that $\int_\Omega \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} = \int_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$. Then there exists C depending only on α in* (5.1.2) *such that*

$$\|\widetilde{u} - \widetilde{u}_h\| \lesssim_\eta Ch^{2-\frac{1}{2^*}}\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + Ch^2\mathcal{F}(\overline{y}_d, f, \overline{y}, \overline{p}), \tag{5.4.24}$$

$$\mathcal{F}(\overline{y}_d, f, \overline{y}, \overline{p}) = \|\overline{y}_d\|_{H^1(\Omega)} + \|f\|_{H^1(\Omega)} + \|\overline{y}\|_{H^2(\Omega)} + \|\overline{p}\|_{H^2(\Omega)}.$$

Proof. Consider the auxiliary problem defined by: For $g \in L^2(\Omega)$, let $p^*_{\mathcal{D}}(g) \in X_{\mathcal{D}}$ solve

$$a_{\mathcal{D}}(z_{\mathcal{D}}, p^*_{\mathcal{D}}(g)) = (\Pi_{\mathcal{D}} y_{\mathcal{D}}(g) - \overline{y}_d, \Pi_{\mathcal{D}} z_{\mathcal{D}}) \qquad \forall z_{\mathcal{D}} \in X_{\mathcal{D}}, \tag{5.4.25}$$

where $y_{\mathcal{D}}(g)$ is given by (5.4.4a) with $\overline{u}$ replaced by g. Fix $p^*_{\mathcal{D}}(g)$ by imposing $\int_\Omega \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(g)\,\mathrm{d}\boldsymbol{x} = \int_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$. This choice is dictated by the pure Neumann boundary condition and will be essential.

For $K \in \mathcal{M}$, let $\overline{\boldsymbol{x}}_K$ be the centroid (centre of mass) of K. A standard approximation property (see e.g. [54, Lemma A.7] with $w_K \equiv 1$) yields

$$\forall K \in \mathcal{M}\,,\ \forall \phi \in H^2(K)\,,\ \|\mathcal{P}_{\mathcal{M}}\phi - \phi(\overline{\boldsymbol{x}}_K)\|_{L^2(K)} \lesssim_\eta \operatorname{diam}(K)^2 \|\phi\|_{H^2(K)}. \tag{5.4.26}$$

Define $\widehat{u}$ and $\widehat{p}$ a.e. on Ω by: For all $K \in \mathcal{M}$ and all $\boldsymbol{x} \in K$, $\widehat{u}(\boldsymbol{x}) = \overline{u}(\overline{\boldsymbol{x}}_K)$ and $\widehat{p}(\boldsymbol{x}) = \overline{p}(\overline{\boldsymbol{x}}_K)$. From (5.4.23) and the Lipschitz continuity of $P_{[a,b]}$,

$$\begin{aligned}\|\widetilde{u} - \widetilde{u}_h\| &\le \alpha^{-1}\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \overline{p}_{\mathcal{M}}\| \\ &\le \alpha^{-1}\|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\overline{u})\| + \alpha^{-1}\|\Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\widehat{u})\| \\ &\quad + \alpha^{-1}\|\Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\widehat{u}) - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| =: \alpha^{-1}T_1 + \alpha^{-1}T_2 + \alpha^{-1}T_3.\end{aligned} \tag{5.4.27}$$

Step 1: estimate of T_1.
A use of triangle inequality, (5.2.2b), (5.4.4b) and **(A1)**-i) leads to

$$\begin{aligned}T_1 &\le \|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}p_{\mathcal{D}}(\overline{u})\| + \|\Pi_{\mathcal{D}}p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\overline{u})\| \\ &\lesssim h^2\|\overline{y} - \overline{y}_d\|_{H^1(\Omega)} + \|\Pi_{\mathcal{D}}p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\overline{u})\|.\end{aligned} \tag{5.4.28}$$

The last term in this inequality is estimated now. Use the definitions of $C_{\mathcal{D}}$, $\|\cdot\|_{\mathcal{D}}$ and the fact that $⨍_\Omega \Pi_{\mathcal{D}}p_{\mathcal{D}}(\overline{u})\,\mathrm{d}\boldsymbol{x} = ⨍_\Omega \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\overline{u})\,\mathrm{d}\boldsymbol{x}$ to obtain

$$\|\Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\|^2 \lesssim \|\nabla_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\|^2. \tag{5.4.29}$$

Subtract (5.4.25) with $g = \overline{u}$ from (5.4.4b), substitute $z_{\mathcal{D}} = p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u})$, use property (5.4.21) in **(A1)**-ii) and Cauchy–Schwarz inequality to obtain

$$\begin{aligned}\|\nabla_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\|^2 &\lesssim a_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}), p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u})) \\ &= (\overline{y} - \Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}), \Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))) \\ &= (\overline{y} - \overline{y}_{\mathcal{M}}, \Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))) \\ &\quad + (\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}), \Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))) \\ &\lesssim h^2\|\overline{y}\|_{H^2(\Omega)}\|\Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\| \\ &\quad + \|\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u})\|\,\|\Pi_{\mathcal{D}}(p_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\overline{u}))\|.\end{aligned}$$

A use of (5.4.29) and **(A1)**-i) leads to $\|\Pi_{\mathcal{D}}p_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}p^*_{\mathcal{D}}(\overline{u})\| \lesssim h^2\|\overline{y}\|_{H^2(\Omega)} + h^2\|f + \overline{u}\|_{H^1(\Omega)}$. Plugged into (5.4.28), this estimate yields

$$T_1 \lesssim h^2(\|\overline{y} - \overline{y}_d\|_{H^1(\Omega)} + \|\overline{y}\|_{H^2(\Omega)} + \|f + \overline{u}\|_{H^1(\Omega)}). \tag{5.4.30}$$

Step 2: estimate of T_2.
Subtract the equations (5.4.25) satisfied by $p^*_{\mathcal{D}}(\overline{u})$ and $p^*_{\mathcal{D}}(\widehat{u})$ to obtain, for all $z_{\mathcal{D}} \in X_{\mathcal{D}}$,

$$a_{\mathcal{D}}(z_{\mathcal{D}}, p^*_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\widehat{u})) = (\Pi_{\mathcal{D}}y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}}y_{\mathcal{D}}(\widehat{u}), \Pi_{\mathcal{D}}z_{\mathcal{D}}). \tag{5.4.31}$$

Since $p_{\mathcal{D}}^*(\widehat{u}) - p_{\mathcal{D}}^*(\overline{u}) \in X_{\mathcal{D},\star}$, a use of Proposition 5.3.6 in (5.4.31) yields

$$T_2 = \|\Pi_{\mathcal{D}} p_{\mathcal{D}}^*(\overline{u}) - \Pi_{\mathcal{D}} p_{\mathcal{D}}^*(\widehat{u})\| \lesssim \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u})\|. \tag{5.4.32}$$

Choose $z_{\mathcal{D}} = y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u})$ in (5.4.31), subtract the equations (5.4.4a) satisfied by $y_{\mathcal{D}}(\overline{u})$ and $y_{\mathcal{D}}(\widehat{u})$, since $p_{\mathcal{D}}^*(\overline{u}) - p_{\mathcal{D}}^*(\widehat{u}) \in X_{\mathcal{D},\star}$, to obtain

$$\|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|^2 = a_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}), p_{\mathcal{D}}^*(\overline{u}) - p_{\mathcal{D}}^*(\widehat{u})) = (\overline{u} - \widehat{u}, \Pi_{\mathcal{D}} p_{\mathcal{D}}^*(\overline{u}) - \Pi_{\mathcal{D}} p_{\mathcal{D}}^*(\widehat{u})).$$

Set $w_{\mathcal{D}} = p_{\mathcal{D}}^*(\overline{u}) - p_{\mathcal{D}}^*(\widehat{u})$, use orthogonality of $\boldsymbol{\mathcal{P}}_{\mathcal{M}}$, Cauchy–Schwarz inequality and **(A2)** to infer

$$\begin{aligned}
\|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|^2 &= (\overline{u} - \widehat{u}, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \\
&= (\overline{u} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u}, \Pi_{\mathcal{D}} w_{\mathcal{D}} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}} w_{\mathcal{D}})) + (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \\
&\lesssim_\eta h\|\overline{u}\|_{H^1(\Omega)} h\|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| + \int_{\Omega_{1,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}) \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x} \\
&\quad + \int_{\Omega_{2,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}) \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x}.
\end{aligned} \tag{5.4.33}$$

Equation (5.4.31) and the stability of the GDM (Proposition 5.3.6) show that

$$\|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| = \|\nabla_{\mathcal{D}}(p_{\mathcal{D}}^*(\overline{u}) - p_{\mathcal{D}}^*(\widehat{u}))\| \lesssim \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|. \tag{5.4.34}$$

A use of Holder's inequality, **(A4)**, **(A3)**, the fact that $w_{\mathcal{D}} \in X_{\mathcal{D},\star}$ and (5.4.34) yields an estimate for the second term on the right hand side of (5.4.33) as follows:

$$\begin{aligned}
\int_{\Omega_{1,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}) \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x} &\le \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}\|_{L^2(\Omega_{1,\mathcal{M}})} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\|_{L^2(\Omega_{1,\mathcal{M}})} \\
&\le h\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} |\Omega_{1,\mathcal{M}}|^{\frac{1}{2}} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\|_{L^{2^*}(\Omega)} |\Omega_{1,\mathcal{M}}|^{\frac{1}{2}-\frac{1}{2^*}} \\
&\lesssim h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \|w_{\mathcal{D}}\|_{\mathcal{D}} \\
&= h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| \\
&\lesssim h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|.
\end{aligned} \tag{5.4.35}$$

Consider now the last term on the right hand side of (5.4.33). For any $K \in \mathcal{M}_2$, $\overline{u} = a$ on K, $\overline{u} = b$ on K, or, by (5.4.55), $\overline{u} = -\alpha^{-1}\overline{p} + \overline{c}$ on K. Hence, on K, $\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u} = 0$ or $\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u} = \alpha^{-1}(\widehat{p} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{p})$. This leads to $|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}| \le \alpha^{-1}|\widehat{p} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{p}|$ on $\Omega_{2,\mathcal{M}}$. Use (5.4.26), the definition of $C_{\mathcal{D}}$, the fact that $w_{\mathcal{D}} \in X_{\mathcal{D},\star}$ and (5.4.34) to obtain

$$\begin{aligned}
\int_{\Omega_{2,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}) \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x} &\le \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}\|_{L^2(\Omega_{2,\mathcal{M}})} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\| \\
&\le \alpha^{-1} \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{p} - \widehat{p}\|_{L^2(\Omega_{2,\mathcal{M}})} \|\Pi_{\mathcal{D}} w_{\mathcal{D}}\| \\
&\lesssim_\eta h^2 \alpha^{-1} \|\overline{p}\|_{H^2(\Omega_{2,\mathcal{M}})} \|\nabla_{\mathcal{D}} w_{\mathcal{D}}\| \\
&\lesssim_\eta h^2 \alpha^{-1} \|\overline{p}\|_{H^2(\Omega_{2,\mathcal{M}})} \|\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\overline{u}) - y_{\mathcal{D}}(\widehat{u}))\|.
\end{aligned} \tag{5.4.36}$$

Plug (5.4.34), (5.4.35) and (5.4.36) into (5.4.33) and then in (5.4.32) to get

$$T_2 \lesssim_\eta h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + h^2 \left(\|\overline{u}\|_{H^1(\Omega)} + \alpha^{-1} \|\overline{p}\|_{H^2(\Omega_{2,\mathcal{M}})} \right). \tag{5.4.37}$$

Step 3: estimate of T_3.
Subtract (5.4.1b) from (5.4.25) with $g = \widehat{u}$ and (5.4.1a) from (5.4.4a) with $\widehat{u}$ instead of $\overline{u}$ to obtain for all $z_\mathcal{D} \in X_\mathcal{D}$ and $w_\mathcal{D} \in X_{\mathcal{D},\star}$,

$$a_\mathcal{D}(z_\mathcal{D}, p^*_\mathcal{D}(\widehat{u}) - \overline{p}_\mathcal{D}) = (\Pi_\mathcal{D} y_\mathcal{D}(\widehat{u}) - \Pi_\mathcal{D} \overline{y}_\mathcal{D}, \Pi_\mathcal{D} z_\mathcal{D}), \tag{5.4.38}$$

$$a_\mathcal{D}(y_\mathcal{D}(\widehat{u}) - \overline{y}_\mathcal{D}, w_\mathcal{D}) = (\widehat{u} - \overline{u}_h, \Pi_\mathcal{D} w_\mathcal{D}). \tag{5.4.39}$$

Substitute $z_\mathcal{D} = p^*_\mathcal{D}(\widehat{u}) - \overline{p}_\mathcal{D} \in X_{\mathcal{D},\star}$ in (5.4.38), $w_\mathcal{D} = y_\mathcal{D}(\widehat{u}) - \overline{y}_\mathcal{D} \in X_{\mathcal{D},\star}$ in (5.4.39) and use Proposition 5.3.6 to obtain

$$T_3 = \|\Pi_\mathcal{D} p^*_\mathcal{D}(\widehat{u}) - \Pi_\mathcal{D} \overline{p}_\mathcal{D}\| \lesssim \|\Pi_\mathcal{D} y_\mathcal{D}(\widehat{u}) - \Pi_\mathcal{D} \overline{y}_\mathcal{D}\| \lesssim \|\widehat{u} - \overline{u}_h\|. \tag{5.4.40}$$

The optimality condition (5.2.2c) [96, Lemma 3.5] yields for a.e. $\boldsymbol{x} \in \Omega$,

$$\left(\overline{p}(\boldsymbol{x}) + \alpha \overline{u}(\boldsymbol{x})\right) \left(v(\boldsymbol{x}) - \overline{u}(\boldsymbol{x})\right) \geq 0 \text{ for all } v \in \mathcal{U}_{\text{ad}}.$$

Since $\overline{u}$, $\overline{p}$ and $\overline{u}_h$ are continuous at the centroid $\overline{\boldsymbol{x}}_K$, we can choose $\boldsymbol{x} = \overline{\boldsymbol{x}}_K$ and $v(\overline{\boldsymbol{x}}_K) = \overline{u}_h(\overline{\boldsymbol{x}}_K) (= \overline{u}_h \text{ on } K)$. All the involved functions being constants over K, this gives

$$(\widehat{p} + \alpha \widehat{u}) (\overline{u}_h - \widehat{u}) \geq 0 \text{ on } K, \text{ for all } K \in \mathcal{M}.$$

Integrate over K and sum over $K \in \mathcal{M}$ to deduce

$$(\widehat{p} + \alpha \widehat{u}, \overline{u}_h - \widehat{u}) \geq 0.$$

Choose $v_h = \widehat{u}$ in the discrete optimality condition (5.4.1c) to establish

$$(\Pi_\mathcal{D} \overline{p}_\mathcal{D} + \alpha \overline{u}_h, \widehat{u} - \overline{u}_h) \geq 0.$$

Add the above two inequalities to obtain

$$(\widehat{p} - \Pi_\mathcal{D} \overline{p}_\mathcal{D} + \alpha(\widehat{u} - \overline{u}_h), \overline{u}_h - \widehat{u}) \geq 0,$$

and thus

$$\begin{aligned} \alpha \|\widehat{u} - \overline{u}_h\|^2 &\leq (\widehat{p} - \Pi_\mathcal{D} \overline{p}_\mathcal{D}, \overline{u}_h - \widehat{u}) \\ &= (\widehat{p} - \overline{p}_\mathcal{M}, \overline{u}_h - \widehat{u}) + (\overline{p}_\mathcal{M} - \Pi_\mathcal{D} p^*_\mathcal{D}(\widehat{u}), \overline{u}_h - \widehat{u}) \\ &\quad + (\Pi_\mathcal{D} p^*_\mathcal{D}(\widehat{u}) - \Pi_\mathcal{D} \overline{p}_\mathcal{D}, \overline{u}_h - \widehat{u}) = M_1 + M_2 + M_3. \end{aligned} \tag{5.4.41}$$

Since $\overline{u}_h - \widehat{u}$ is piecewise constant on $\mathcal{M}$, the orthogonality property of $\mathcal{P}_{\mathcal{M}}$, (5.4.26) and (5.4.22) in **(A1)**-ii) lead to

$$\begin{aligned} M_1 &= (\widehat{p} - \mathcal{P}_{\mathcal{M}}\overline{p}_{\mathcal{M}}, \overline{u}_h - \widehat{u}) \\ &= (\widehat{p} - \mathcal{P}_{\mathcal{M}}\overline{p}, \overline{u}_h - \widehat{u}) + (\mathcal{P}_{\mathcal{M}}(\overline{p} - \overline{p}_{\mathcal{M}}), \overline{u}_h - \widehat{u}) \lesssim_\eta h^2 \|\overline{p}\|_{H^2(\Omega)} \|\overline{u}_h - \widehat{u}\|. \end{aligned} \tag{5.4.42}$$

By Cauchy–Schwarz inequality, triangle inequality and the notations in (5.4.27),

$$M_2 \leq \|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u})\| \, \|\overline{u}_h - \widehat{u}\| \lesssim (T_1 + T_2) \|\overline{u}_h - \widehat{u}\|. \tag{5.4.43}$$

Subtract the equations (5.4.1a) and (5.4.4a) (with $\widehat{u}$ instead of $\overline{u}$) satisfied by $\overline{y}_{\mathcal{D}}$ and $y_{\mathcal{D}}(\widehat{u})$, choose $w_{\mathcal{D}} = p^*_{\mathcal{D}}(\widehat{u}) - \overline{p}_{\mathcal{D}}$, and use the equations (5.4.1b) and (5.4.25) on $\overline{p}_{\mathcal{D}}$ and $p^*_{\mathcal{D}}(\widehat{u})$ to arrive at

$$\begin{aligned} M_3 &= (\Pi_{\mathcal{D}}(p^*_{\mathcal{D}}(\widehat{u}) - \overline{p}_{\mathcal{D}}), \overline{u}_h - \widehat{u}) = a_{\mathcal{D}}(\overline{y}_{\mathcal{D}} - y_{\mathcal{D}}(\widehat{u}), p^*_{\mathcal{D}}(\widehat{u}) - \overline{p}_{\mathcal{D}}) \\ &= (\Pi_{\mathcal{D}}(y_{\mathcal{D}}(\widehat{u}) - \overline{y}_{\mathcal{D}}), \Pi_{\mathcal{D}}(\overline{y}_{\mathcal{D}} - y_{\mathcal{D}}(\widehat{u})) \leq 0. \end{aligned} \tag{5.4.44}$$

A substitution of (5.4.42)–(5.4.44) (together with the estimates (5.4.30) and (5.4.37) of T_1 and T_2) into (5.4.41) yields an estimate on $\|\overline{u}_h - \widehat{u}\|$ which, when plugged into (5.4.40), gives

$$\begin{aligned} T_3 \lesssim_\eta \; & \alpha^{-1} h^{2-\frac{1}{2^*}} \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \\ & + \alpha^{-1} h^2 \big(\|\overline{y} - \overline{y}_d\|_{H^1(\Omega)} + \|\overline{y}\|_{H^2(\Omega)} + (1 + \alpha^{-1}) \|\overline{p}\|_{H^2(\Omega)} + \|f + \overline{u}\|_{H^1(\Omega)} + \|\overline{u}\|_{H^1(\Omega)} \big). \end{aligned} \tag{5.4.45}$$

Step 4: conclusion.
A use of (5.1.2) and the fact that $\overline{u}$ is optimal leads to

$$\frac{\alpha}{2} \|\overline{u}\|^2 \leq J(\overline{y}, \overline{u}) \leq J(y(0), 0) = \frac{1}{2} \|y(0) - \overline{y}_d\|^2,$$

where $y(0)$ is the solution to the state equation (5.1.1b) with $u = 0$. Hence,

$$\|\overline{u}\| \lesssim \sqrt{\alpha}^{-1} \left(\|f\| + \|\overline{y}_d\| \right). \tag{5.4.46}$$

From (5.4.55) and (5.4.2),

$$\nabla \overline{u} = \nabla P_{[a,b]}(-\alpha^{-1}\overline{p} + \overline{c}) = \mathbb{I}_{(-\alpha^{-1}\overline{p} + \overline{c}) \in [a,b]} \nabla(-\alpha^{-1}\overline{p} + \overline{c}),$$

where $\mathbb{I}_X$ is the characteristic function of the set X. Note that $|\nabla(-\alpha^{-1}\overline{p} + \overline{c})| = \alpha^{-1}|\nabla \overline{p}|$. Therefore,

$$\|\nabla \overline{u}\|^2 = \int_\Omega |\nabla \overline{u}|^2 \, \mathrm{d}\boldsymbol{x} = \int_\Omega |\mathbb{I}_{(-\alpha^{-1}\overline{p} + \overline{c}) \in [a,b]} \nabla(-\alpha^{-1}\overline{p} + \overline{c})|^2 \, \mathrm{d}\boldsymbol{x} \lesssim \alpha^{-2} \|\nabla \overline{p}\|^2. \tag{5.4.47}$$

Combine (5.4.46) and (5.4.47) to obtain

$$\|\overline{u}\|_{H^1(\Omega)} \lesssim \sqrt{\alpha}^{-1} \left(\|f\| + \|\overline{y}_d\| \right) + \alpha^{-1} \|\nabla \overline{p}\|. \tag{5.4.48}$$

Use (5.4.48) in (5.4.30), (5.4.37) and (5.4.45) and plug the resulting estimates in (5.4.27) to complete the proof. □

Theorem 5.4.5 (Super-convergence for post-processed controls II)**.** *Let the assumptions and notations of Theorem 5.4.4 hold, except* **(A3)** *which is replaced by:*

$$\begin{aligned}&\text{there exists } \delta > 0 \text{ such that, for any } F \in L^2(\Omega),\\ &\text{the solution } \psi_D \text{ to (5.3.4) satisfies } \|\Pi_D \psi_D\|_{L^\infty(\Omega)} \le \delta \|F\|.\end{aligned} \tag{5.4.49}$$

Then there exists C depending only on α and δ such that

$$\|\widetilde{u} - \widetilde{u}_h\| \lesssim_\eta Ch^2 \left[\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + \mathcal{F}(\overline{y}_d, f, \overline{y}, \overline{p}) \right]. \tag{5.4.50}$$

Proof. The proof of this theorem is identical to the proof of Theorem 5.4.4, except for the estimate of T_2. This estimate is the only source of the $2 - \frac{1}{2^*}$ power (instead of 2), and the only place where we used Assumption **(A3)**, here replaced by (5.4.49). Recall **(A4)** and use (5.4.49) in (5.4.31) satisfied by $p^*_{\mathcal{D}}(\overline{u}) - p^*_{\mathcal{D}}(\widehat{u})$ to write

$$\begin{aligned}\int_{\Omega_{1,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}) \Pi_{\mathcal{D}} w_{\mathcal{D}} \,\mathrm{d}\boldsymbol{x} &= \int_{\Omega_{1,\mathcal{M}}} (\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}) \left(\Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u}) \right) \mathrm{d}\boldsymbol{x}\\ &\lesssim \|\boldsymbol{\mathcal{P}}_{\mathcal{M}}\overline{u} - \widehat{u}\|_{L^\infty(\Omega_{1,\mathcal{M}})} \|\Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u})\|_{L^\infty(\Omega_{1,\mathcal{M}})} |\Omega_{1,\mathcal{M}}|\\ &\lesssim h^2 \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \|\Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} p^*_{\mathcal{D}}(\widehat{u})\|_{L^\infty(\Omega)}\\ &\lesssim h^2 \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} \delta \|\Pi_{\mathcal{D}} y_{\mathcal{D}}(\overline{u}) - \Pi_{\mathcal{D}} y_{\mathcal{D}}(\widehat{u})\|.\end{aligned} \tag{5.4.51}$$

The rest of the proof follows from this estimate. □

Remark 5.4.6. *For most methods, assumption* (5.4.49) *is satisfied if the mesh is quasi-uniform (see [67] for conforming and non-conforming $\mathbb{P}_1$ finite element method, and Appendix A.3 for HMM methods for Dirichlet BCs; the adaptation to Neumann BCs is straightforward).*

Corollary 5.4.7 (Super-convergence for the state and adjoint variables)**.** *Under the assumptions of Theorem 5.4.4, the following error estimates hold, with C depending only on α:*

$$\|\overline{y}_{\mathcal{M}} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| \lesssim_\eta Ch^r \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + Ch^2 \mathcal{F}(\overline{y}_d, f, \overline{y}, \overline{p}), \tag{5.4.52}$$

$$\|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| \lesssim_\eta Ch^r \|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)} + Ch^2 \mathcal{F}(\overline{y}_d, f, \overline{y}, \overline{p}), \tag{5.4.53}$$

where $\overline{y}_{\mathcal{M}}$ and $\overline{p}_{\mathcal{M}}$ are defined as in **(A1)***, and $r = 2 - \frac{1}{2^*}$.*
Under the assumptions of Theorem 5.4.5, (5.4.52) *and* (5.4.53) *hold with $r = 2$ and C depending only α and δ.*

Proof. The result for the state and adjoint variables can be derived exactly as in Corollary 4.3.9 □

5.4.4 Discussion on post-processed controls

In this section, a detailed analysis of post-processed controls given by (5.4.23) is presented. This analysis is performed under the assumptions of Section 5.4.3, and by also assuming that $\mathrm{WS}_{\mathcal{D}}(\varphi) \lesssim h$ for all $\varphi \in H^2(\Omega)$ (see Remark 5.3.4). We begin by stating and proving two lemmas which discuss projection relations between control and adjoint variables for the pure Neumann problem, both at the continuous level and at the discrete level. We then show that the post-processed controls remain $\mathcal{O}(h)$ close to their corresponding original controls, see (5.4.59) and (5.4.63). Hence, the super-convergence result makes sense: since $\overline{u}_h$ is piecewise constant, it is impossible to expect more than $\mathcal{O}(h)$ approximation on the controls; but by "moving" these controls by a specific $\mathcal{O}(h)$, computable post-processed controls are obtained that enjoy an $\mathcal{O}(h^2)$ convergence result.

Lemma 5.4.8. *Let $-\infty \le a < 0 < b \le \infty$ and $\phi \in L^1(\Omega)$. Define $\Gamma : \mathbb{R} \to \mathbb{R}$ by*

$$\Gamma(c) = \int_\Omega P_{[a,b]}(\phi + c)\,\mathrm{d}\boldsymbol{x},$$

where $P_{[a,b]}$ is given by (5.4.2). *Set $m = a - \operatorname{ess\,sup}(\phi) \in [-\infty, +\infty)$ and $M = b - \operatorname{ess\,inf}(\phi) \in (-\infty, +\infty]$. Then we have the following.*

1. *Γ is Lipschitz continuous.*
2. *$\lim_{c\to m} \Gamma(c) = a|\Omega|$, $\lim_{c\to M} \Gamma(c) = b|\Omega|$, and there is $c^\star \in (m,M)$ such that $\Gamma(c^\star) = 0$.*
3. *If $\phi \in H^1(\Omega)$, then for any compact interval Q in (m,M), there exists $\rho_Q > 0$ such that if $c, c' \in Q$ with $c < c'$, then*
$$\Gamma(c') - \Gamma(c) \ge \rho_Q(c' - c). \tag{5.4.54}$$
As a consequence, the real number $c^\star$ in Item 2 is unique.

Proof. Item 1 is obvious since $P_{[a,b]}$ is Lipschitz continuous.

Let us now analyze the limits in Item 2. Let (c_n) be a sequence in $\mathbb{R}$ such that $c_n \to M$ as $n \to \infty$. By definition of M, this implies $P_{[a,b]}(\phi + c_n) \to b$ a.e on Ω. Let (c_n) be bounded below by R and note that $\phi + R \in L^1(\Omega)$. Moreover, for $s \in \mathbb{R}$, $a < 0 < b$ implies $P_{[a,b]}(s) \ge \min(s, 0)$ so $P_{[a,b]}(\phi + c_n) \ge \min(\phi + c_n, 0) \ge \min(\phi + R, 0) \in L^1(\Omega)$. By Fatou's Lemma,

$$\int_\Omega b\,\mathrm{d}\boldsymbol{x} \le \liminf_{n\to\infty} \int_\Omega P_{[a,b]}(\phi(\boldsymbol{x}) + c_n)\,\mathrm{d}\boldsymbol{x}$$

which gives $b|\Omega| \le \liminf_{n\to\infty} \Gamma(c_n)$. Since $\Gamma(c_n) \le b|\Omega|$ (because $P_{[a,b]}(s) \le b$), that $\lim_{n\to\infty} \Gamma(c_n) = b|\Omega|$, and thus that $\lim_{c\to M} \Gamma(c) = b|\Omega|$. In a similar way, we deduce that $\lim_{c\to m} \Gamma(c) = a|\Omega|$. The existence of $c^\star$ such that $\Gamma(c^\star) = 0$ then follows from the intermediate value theorem and $\lim_{c\to m} \Gamma(c) = a|\Omega| < 0 < b|\Omega| = \lim_{c\to M} \Gamma(c)$.

Now assume that $\phi \in H^1(\Omega)$ and consider Item 3. For a.e $c \in \mathbb{R}$, $\Gamma'(c) = \int_\Omega \mathbb{I}_{(a,b)}(\phi(\boldsymbol{x}) + c)\,\mathrm{d}\boldsymbol{x}$, Define $\Theta(c) = \int_\Omega \mathbb{I}_{(a,b)}(\phi(\boldsymbol{x}) + c)\,\mathrm{d}\boldsymbol{x}$, for all $c \in \mathbb{R}$. We claim that

- Θ is lower semi-continuous,
- $\forall c \in (m,M)$, $\Theta(c) > 0$.

To prove that Θ is lower semi-continuous, let $c_n \to c$ as $n \to \infty$. Since $\mathbb{I}_{(a,b)}$ is lower semi-continuous on $\mathbb{R}$, for all $\boldsymbol{x} \in \Omega$,

$$\mathbb{I}_{(a,b)}(\phi(\boldsymbol{x}) + c) \leq \liminf_{n\to\infty} \mathbb{I}_{(a,b)}(\phi(\boldsymbol{x}) + c_n).$$

Applying Fatou's Lemma,

$$\Theta(c) \leq \liminf_{n\to\infty} \int_\Omega \mathbb{I}_{(a,b)}(\phi(\boldsymbol{x}) + c_n)\,\mathrm{d}\boldsymbol{x} = \liminf_{n\to\infty} \Theta(c_n).$$

Hence, Θ is lower semi-continuous. We now show that $\Theta > 0$ on (m,M). Let $c \in (m,M)$. Then $I = (a-c, b-c) \cap (\text{ess inf } \phi, \text{ess sup } \phi)$ is an interval of positive length, since $a - c < \text{ess sup } \phi$ and $b - c > \text{ess inf } \phi$. The set $W_{I,c} = \{\boldsymbol{x} : \phi(\boldsymbol{x}) \in I\}$ has a non-zero measure because $\phi \in H^1(\Omega)$ and Ω is connected. To see this, let $\alpha < \beta$ be the endpoints of I and assume that $\phi \in H^1(\Omega)$ takes some values less than α on a non-null set, some values greater than β on a non-null set, but that $W_{I,c}$ is a null set. Then $P_{[\alpha,\beta]}(\phi) \in H^1(\Omega)$ exactly takes the values α and β (outside a set of zero measure). Hence $\nabla P_{[\alpha,\beta]}(\phi) = \mathbb{I}_{[\alpha,\beta]}(\phi)\nabla\phi = 0$ and $P_{[\alpha,\beta]}(\phi)$ should be constant, since Ω is connected, which is a contradiction. Thus, $W_{I,c}$ has a non-zero measure. Since $\{\boldsymbol{x} : \phi(\boldsymbol{x}) + c \in (a,b)\} \supseteq W_{I,c}$, this gives $\Theta(c) \geq |W_{I,c}| > 0$.
Coming back to Item 3, let Q be a compact interval in (m,M). We know that $\Theta > 0$ on Q and Θ is lower semi-continuous. Hence Θ reaches its minimum on Q and $\inf_Q \Theta = \Theta(c_0) > 0$ for some $c_0 \in Q$. Since $\Gamma' = \Theta$ a.e, $\Gamma' \geq \inf_Q \Theta$ a.e on Q and, Γ being Lipschitz and $[c,c'] \subset Q$,

$$\Gamma(c') - \Gamma(c) = \int_c^{c'} \Gamma'(s)\,\mathrm{d}s \geq \left[\inf_Q \Theta\right](c' - c),$$

which establishes (5.4.54). The uniqueness of $c^\star$ such that $\Gamma(c^\star) = 0$ follows from this inequality, which shows that Γ is strictly increasing on (m,M). □

Lemma 5.4.9 (Projection formulas for the controls)**.** *If $\overline{p} \in H^1(\Omega)$ is a co-state and $\overline{c} \in \mathbb{R}$ is such that $⨍_\Omega P_{[a,b]}(-\frac{1}{\alpha}\overline{p}(\boldsymbol{x}) + \overline{c})\,\mathrm{d}\boldsymbol{x} = 0$, then the continuous optimal control $\overline{u}$ in* (5.2.2) *can be expressed in terms of the projection formula*

$$\overline{u}(\boldsymbol{x}) = P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p}(\boldsymbol{x}) + \overline{c}\right). \tag{5.4.55}$$

If $\mathcal{D}$ is a GD and $\overline{p}_\mathcal{D}$ is chosen such that

$$⨍_\Omega P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_\mathcal{M}\left(-\frac{1}{\alpha}\Pi_\mathcal{D}\overline{p}_\mathcal{D}\right)\right)\mathrm{d}\boldsymbol{x} = 0, \tag{5.4.56}$$

then the discrete optimal control in (5.4.1) *is given by*

$$\overline{u}_h(\boldsymbol{x}) = P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_\mathcal{M}\left(-\frac{1}{\alpha}\Pi_\mathcal{D}\overline{p}_\mathcal{D}(\boldsymbol{x})\right)\right). \tag{5.4.57}$$

Proof. Set $\widetilde{p} = \overline{p} - \alpha\overline{c}$. Clearly, $\widetilde{p} \in H^1(\Omega)$. The optimality condition for the control problem (5.2.2c) implies

$$(\widetilde{p} + \alpha\overline{u}, v - \overline{u}) \geq 0 \ \forall v \in \mathcal{U}_{\text{ad}},$$

since $\int_\Omega \overline{u}\,\mathrm{d}\boldsymbol{x} = \int_\Omega v\,\mathrm{d}\boldsymbol{x} = 0$. Set $U = P_{[a,b]}(-\alpha^{-1}\widetilde{p})$ i.e,

$$U = \begin{cases} a & \text{on } \Omega_+ = \{\boldsymbol{x} \in \Omega : \widetilde{p}(\boldsymbol{x}) + \alpha U(\boldsymbol{x}) > 0\} \\ -\alpha^{-1}\widetilde{p} & \text{on } \Omega_0 = \{\boldsymbol{x} \in \Omega : \widetilde{p}(\boldsymbol{x}) + \alpha U(\boldsymbol{x}) = 0\} \\ b & \text{on } \Omega_- = \{\boldsymbol{x} \in \Omega : \widetilde{p}(\boldsymbol{x}) + \alpha U(\boldsymbol{x}) < 0\}. \end{cases}$$

It is then straightforward to see that $U \in \mathcal{U}_{\text{ad}}$, i.e, $U \in [a,b]$ and $\fint_\Omega U\,\mathrm{d}\boldsymbol{x} = 0$ (by choice of $\overline{c}$). Then, using the definitions of Ω_+, Ω_0 and Ω_-, since $v \geq a = U$ on Ω_+ and $v \leq b = U$ on Ω_-,

$$\begin{aligned}(\widetilde{p} + \alpha U, v - U) = &\int_{\Omega_+} (\widetilde{p} + \alpha U)(v - U)\,\mathrm{d}\boldsymbol{x} + \int_{\Omega_0} (\widetilde{p} + \alpha U)(v - U)\,\mathrm{d}\boldsymbol{x} \\ &+ \int_{\Omega_-} (\widetilde{p} + \alpha U)(v - U)\,\mathrm{d}\boldsymbol{x} \geq 0.\end{aligned}$$

Recall that the optimality condition is nothing but a characterisation of the $L^2(\Omega)$ orthogonal projection of $-\alpha^{-1}\widetilde{p}$ on $\mathcal{U}_{\text{ad}}$ and, as such, defines a unique element $\overline{u}$ of $\mathcal{U}_{\text{ad}}$. We just proved that $U = P_{[a,b]}(-\alpha^{-1}\widetilde{p})$ satisfies this optimality condition, which shows that it is equal to $\overline{u}$. The proof of (5.4.55) is complete.
The second relation follows in a similar way by noticing that, since controls are piecewise-constants on $\mathcal{M}$, (5.4.1c) is equivalent to $(\boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}) + \alpha\overline{u}_h, v_h - \overline{u}_h) \geq 0$, for all $v_h \in \mathcal{U}_{\text{ad},h}$. Also notice that, by definition of $\boldsymbol{\mathcal{P}}_{\mathcal{M}}$ and the assumption (5.4.56), $P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\left(-\frac{1}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\right)\right) \in \mathcal{U}_{\text{ad},h}$. □

Remark 5.4.10. *There is at least one adjoint $\overline{p}_{\mathcal{D}}$ such that* (5.4.56) *is satisfied: start from any adjoint $\overline{p}^0_{\mathcal{D}}$ and, by applying Lemma 5.4.8 (Item 2) to $\phi = \boldsymbol{\mathcal{P}}_{\mathcal{M}}(-\alpha^{-1}\Pi_{\mathcal{D}}\overline{p}^0_{\mathcal{D}})$ and by noticing that $\phi + c^\star = \boldsymbol{\mathcal{P}}_{\mathcal{M}}(-\alpha^{-1}\Pi_{\mathcal{D}}\overline{p}^0_{\mathcal{D}}) + c^\star$, find $c^\star$ such that $\overline{p}_{\mathcal{D}} = \overline{p}^0_{\mathcal{D}} - \alpha c^\star 1_{\mathcal{D}}$ satisfies* (5.4.56).
Since the discrete co-state $\overline{p}_{\mathcal{D}}$ is a computable quantity, its average is easier to fix than the average of the non-computable $\overline{p}$. Hence, the projection relation (5.4.57) *is the most natural choice to express the discrete control $\overline{u}_h$ in terms of the discrete adjoint variable. This is the choice made in the modified active set strategy presented in Subsection 5.5.1. Once this choice is made, since $\overline{p}$ must have the same average as $\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}$ for $\widetilde{u}$ defined in* (5.4.23) *to satisfy super-convergence estimates, it is clear that $P_{[a,b]}(-\frac{1}{\alpha}\overline{p})$ will not have a zero average in general. Hence, if we want to express the continuous control in terms of $\overline{p}$, we need to offset this $\overline{p}$ by the correct $\overline{c}$, as stated in the lemma.*

Lemma 5.4.11 (Stability of the discrete states)**.** *Let $\mathcal{D}$ be a GD, $(\overline{y}, \overline{p}, \overline{u})$ be a solution to* (5.2.2) *and $(\overline{y}_{\mathcal{D}}, \overline{p}_{\mathcal{D}}, \overline{u}_h)$ be a solution to* (5.4.1)*. Assume that $\fint_\Omega \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} = \fint_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$. Then*

$$\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| + \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| + \|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| + \|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| \lesssim 1. \tag{5.4.58}$$

Proof. Let $\phi \in H_0^{\mathrm{div}}(\Omega)$. Take $w = 0$ in (5.3.7) to obtain $S_{\mathcal{D}}(\phi) \le \|\phi\| + \|\nabla\phi\|$. By the Cauchy–Schwarz inequality, using (5.3.6) and (5.3.1), for $w \in X_{\mathcal{D}}$,

$$\begin{aligned}\int_\Omega (\Pi_{\mathcal{D}} w \,\mathrm{div}\phi + \nabla_{\mathcal{D}} w \cdot \phi)\,\mathrm{d}\boldsymbol{x} &\le \|\Pi_{\mathcal{D}} w\|\,\|\mathrm{div}\phi\| + \|\nabla_{\mathcal{D}} w\|\,\|\phi\| \\ &\le C_{\mathcal{D}}\|w\|_{\mathcal{D}}\left(\|\mathrm{div}\phi\| + \|\phi\|\right).\end{aligned}$$

With (5.3.8), this implies $\mathrm{WS}_{\mathcal{D}}(\phi) \le C_{\mathcal{D}}\left(\|\mathrm{div}\phi\| + \|\phi\|\right)$.
Therefore, for $A\nabla\psi \in H_0^{\mathrm{div}}(\Omega)$, the definition (5.3.11) of $\mathrm{WS}_{\mathcal{D}}$ leads to

$$\mathrm{WS}_{\mathcal{D}}(\psi) \lesssim \|\psi\|_{H^1(\Omega)} + \|\mathrm{div}(A\nabla\psi)\| + \|A\nabla\psi\| \lesssim \|\psi\|_{H^1(\Omega)} + \|\mathrm{div}(A\nabla\psi)\| \lesssim 1.$$

This along with Proposition 5.4.2 and Theorem 5.4.1 show that

$$\|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| + \|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| \lesssim \|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \overline{y}\| + \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla\overline{y}\| \lesssim 1.$$

The result for the adjoint variable can be derived similarly and hence (5.4.58) follows. □

In the rest of this section, we establish $\mathcal{O}(h)$ estimates between the controls $\overline{u}$, $\overline{u}_h$ and their post-processed versions $\widetilde{u}$, $\widetilde{u}_h$. These estimates justify that the post-processed controls are indeed meaningful quantities.

A use of (5.4.23), (5.4.57), the Lipschitz continuity of $P_{[a,b]}$, **(A2)** and Lemma 5.4.11 leads to the following estimate between $\overline{u}_h$ and $\widetilde{u}_h$:

$$\|\widetilde{u}_h - \overline{u}_h\| \le \alpha^{-1}\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}\left(\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\right)\| \lesssim \alpha^{-1}h\|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| \lesssim \alpha^{-1}h. \tag{5.4.59}$$

Let us now turn to estimating $\overline{u} - \widetilde{u}$. The co-state $\overline{p} \in H^1(\Omega)$ in (5.2.2) is still taken such that $\int_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x} = \int_\Omega \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x}$. From Lemma 5.4.8, it follows that there exists a unique constant $\overline{c} \in (m,M)$ such that $\int_\Omega P_{[a,b]}(-\frac{1}{\alpha}\overline{p} + \overline{c})\,\mathrm{d}\boldsymbol{x} = 0$, where m and M are defined as in Lemma 5.4.8. Using Lemma 5.4.9 and recalling (5.4.23),

$$\overline{u}(\boldsymbol{x}) = P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p}(\boldsymbol{x}) + \overline{c}\right) \quad \text{and} \quad \widetilde{u}(\boldsymbol{x}) = P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p}_{\mathcal{M}}(\boldsymbol{x})\right). \tag{5.4.60}$$

Starting from (5.4.60) and using the Lipschitz continuity of $P_{[a,b]}$, the assumption $\mathrm{WS}_{\mathcal{D}}(\varphi) \lesssim h$, Corollary 5.4.7 and Proposition 5.4.2, we get a constant C depending only on α, f, $\overline{y}_d$, $\overline{p}$, $\overline{y}$ and $\overline{u}$ such that

$$\begin{aligned}\|\overline{u} - \widetilde{u}\| \le \alpha^{-1}\|\overline{p}_{\mathcal{M}} - \overline{p} + \alpha\overline{c}\| &\lesssim \alpha^{-1}\|\overline{p}_{\mathcal{M}} - \overline{p}\| + |\overline{c}| \\ &\lesssim \alpha^{-1}\left(\|\overline{p}_{\mathcal{M}} - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| + \|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \overline{p}\|\right) + |\overline{c}| \\ &\lesssim_\eta Ch + |\overline{c}|. \end{aligned} \tag{5.4.61}$$

To estimate the last term in (5.4.61), recall the definition of $\Gamma(c)$ from Lemma 5.4.8 for $\phi = \alpha^{-1}\overline{p}$.

$$\Gamma(c) = \int_\Omega P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p} + c\right)\,\mathrm{d}\boldsymbol{x}.$$

By choice of $\overline{c}$, $\Gamma(\overline{c}) = 0$. From Lemma 5.4.9, the choice of $\overline{p}_{\mathcal{D}}$ shows that

$$\begin{aligned}\Gamma(0) &= \int_\Omega P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p}\right)\mathrm{d}\boldsymbol{x}\\ &= \int_\Omega \left(P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p}\right) - P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\left(-\frac{1}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\right)\right)\right)\mathrm{d}\boldsymbol{x}.\end{aligned}$$

Let $q_{\mathcal{D}}$ be the solution to (5.4.1b) with source term $\overline{y} - \overline{y}_d$ (that is, the solution to the GS for the equation (5.2.2b) satisfied by $\overline{p}$) such that $\int_\Omega \Pi_{\mathcal{D}} q_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} = \int_\Omega \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\,\mathrm{d}\boldsymbol{x} = \int_\Omega \overline{p}\,\mathrm{d}\boldsymbol{x}$. Use the Lipschitz continuity of $P_{[a,b]}$, the Cauchy–Schwarz inequality, the triangle inequality, Remark 5.3.5, Proposition 5.3.6, **(A2)**, Theorem 5.3.3 and Lemma 5.4.11 to obtain

$$\begin{aligned}|\Gamma(0)| &\le \int_\Omega \left|P_{[a,b]}\left(-\frac{1}{\alpha}\overline{p}\right) - P_{[a,b]}\left(\boldsymbol{\mathcal{P}}_{\mathcal{M}}\left(-\frac{1}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\right)\right)\right|\mathrm{d}\boldsymbol{x}\\ &\lesssim \alpha^{-1}\|\overline{p} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}})\|\\ &\lesssim \alpha^{-1}\|\overline{p} - \Pi_{\mathcal{D}}q_{\mathcal{D}}\| + \alpha^{-1}\|\Pi_{\mathcal{D}}q_{\mathcal{D}} - \Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\| + \alpha^{-1}\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}})\|\\ &\lesssim \alpha^{-1}\left(\mathrm{WS}_{\mathcal{D}}(\overline{p}) + \alpha^{-1}\|\overline{y} - \Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}}\| + h\|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}\|\right)\\ &\lesssim \alpha^{-1}\left(\mathrm{WS}_{\mathcal{D}}(\overline{p}) + \mathrm{WS}_{\mathcal{D}}(\overline{y}) + h\|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}}\|\right) \lesssim \alpha^{-1}h. \end{aligned} \tag{5.4.62}$$

Let m, M be as in Lemma 5.4.8 for $\phi = \alpha^{-1}\overline{p}$. Relation (5.4.62) shows that $a|\Omega| < \Gamma(0) < b|\Omega|$ if h is small enough; hence, in this case, $0 \in (m,M)$. There is therefore a compact interval Q in (m,M) depending only on $\overline{p}$ such that 0 and $\overline{c}$ belong to Q. Without loss of generality, assume that $\overline{c} \ge 0$. A use of Lemma 5.4.8 leads to

$$\Gamma(\overline{c}) - \Gamma(0) \ge \rho_Q \overline{c},$$

where $\rho_Q > 0$. This implies $0 \le \overline{c} \lesssim \alpha^{-1}h/\rho_Q$, using (5.4.62) and the fact that $\Gamma(\overline{c}) = 0$. Combine this with (5.4.61) to deduce

$$\|\widetilde{u} - \overline{u}\| \lesssim_\eta \left(C + \frac{\alpha^{-1}}{\rho_Q}\right)h. \tag{5.4.63}$$

5.5 Numerical experiments

In this section, we first present the modified active set strategy. This is followed by results of numerical experiments for conforming, non-conforming and mimetic finite difference methods.

5.5.1 A modified active set strategy

The interest of choosing an adjoint given by (5.4.56) is highlighted in Lemma 5.4.9: we have the projection relation (5.4.57) between the discrete control and adjoint. Such a relation is at the core of the (standard) active set algorithm [109]. For a detailed analysis of this method, see

[10, 11, 79]. Here, a modified active set algorithm that enforces the proper zero average condition is proposed, and thus the proper relation between discrete adjoint and control.
First notice that, when selecting the $\overline{p}_{\mathcal{D}}$ such that (5.4.56) holds, the KKT optimality conditions (5.4.1) can be rewritten as: Seek $(\overline{y}_{\mathcal{D}}, \overline{p}_{\mathcal{D}}, \overline{u}_h) \in X_{\mathcal{D}} \times X_{\mathcal{D}} \times \mathcal{U}_{\mathrm{ad},h}$, such that

$$\begin{aligned} a_{\mathcal{D}}(\overline{y}_{\mathcal{D}}, w_{\mathcal{D}}) + \rho \left(\fint_{\Omega} \Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) \left(\fint_{\Omega} \Pi_{\mathcal{D}} w_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) \\ = (\overline{u}_h + f, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall\, w_{\mathcal{D}} \in X_{\mathcal{D}}, \end{aligned} \tag{5.5.1a}$$

$$\begin{aligned} a_{\mathcal{D}}(z_{\mathcal{D}}, \overline{p}_{\mathcal{D}}) + \rho \left(\fint_{\Omega} P_{[a,b]} \left[\boldsymbol{\mathcal{P}}_{\mathcal{M}}(-\alpha^{-1} \Pi_{\mathcal{D}} \overline{p}_{\mathcal{D}}) \right] \mathrm{d}\boldsymbol{x} \right) \left(\fint_{\Omega} \Pi_{\mathcal{D}} z_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) \\ = (\Pi_{\mathcal{D}} \overline{y}_{\mathcal{D}} - \overline{y}_d, \Pi_{\mathcal{D}} z_{\mathcal{D}}) \qquad \forall\, z_{\mathcal{D}} \in X_{\mathcal{D}}, \end{aligned} \tag{5.5.1b}$$

$$(\Pi_{\mathcal{D}} \overline{p}_{\mathcal{D}} + \alpha \overline{u}_h, v_h - \overline{u}_h) \geq 0 \qquad \forall\, v_h \in \mathcal{U}_{\mathrm{ad},h}, \tag{5.5.1c}$$

where $\rho > 0$ is constant.
Set $\bar{\mu}_h = -(\alpha^{-1} \Pi_{\mathcal{D}} \overline{p}_{\mathcal{D}} + \overline{u}_h)$. As the original active set strategy [109], the modified active set strategy is an iterative algorithm. As initial guesses, two arbitrary functions, u_h^0, μ_h^0 are chosen. In the nth step of the algorithm, define the set of active and inactive restrictions by

$$A_{a,h}^n(\boldsymbol{x}) = \{\boldsymbol{x} : u_h^{n-1}(\boldsymbol{x}) + \mu_h^{n-1}(\boldsymbol{x}) < a\},\ A_{b,h}^n(\boldsymbol{x}) = \{\boldsymbol{x} : u_h^{n-1}(\boldsymbol{x}) + \mu_h^{n-1}(\boldsymbol{x}) > b\},$$

$$I_h^n = \Omega \setminus (A_{a,h}^n \cup A_{b,h}^n).$$

If

$$\max \left(\frac{\|u_h^n - u_h^{n-1}\|_{L^\infty(\Omega)}}{\|u_h^{n-1}\|_{L^\infty(\Omega)}}, \frac{\|\Pi_{\mathcal{D}} p_{\mathcal{D}}^n - \Pi_{\mathcal{D}} p_{\mathcal{D}}^{n-1}\|_{L^\infty(\Omega)}}{\|\Pi_{\mathcal{D}} p_{\mathcal{D}}^{n-1}\|_{L^\infty(\Omega)}} \right) \leq 10^{-10},$$

then terminate the algorithm. In this case, we notice that the relative L^∞ difference between $\Pi_{\mathcal{D}} y_{\mathcal{D}}^{n-1}$ and $\Pi_{\mathcal{D}} y_{\mathcal{D}}^n$ is less than 10^{-6} for all examples in Section 5.5.2. Else we find $y_{\mathcal{D}}^n, p_{\mathcal{D}}^n$ and u_h^n solution to the system

$$\begin{aligned} a_{\mathcal{D}}(y_{\mathcal{D}}^n, w_{\mathcal{D}}) + \rho \left(\fint_{\Omega} \Pi_{\mathcal{D}} y_{\mathcal{D}}^n \, \mathrm{d}\boldsymbol{x} \right) \left(\fint_{\Omega} \Pi_{\mathcal{D}} w_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) \\ = (u_h^n + f, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall w_{\mathcal{D}} \in X_{\mathcal{D}}, \end{aligned} \tag{5.5.2a}$$

$$\begin{aligned} a_{\mathcal{D}}(z_{\mathcal{D}}, p_{\mathcal{D}}^n) + \rho \left(\fint_{\Omega} P_{[a,b]} \left[\boldsymbol{\mathcal{P}}_{\mathcal{M}}(-\alpha^{-1} \Pi_{\mathcal{D}} p_{\mathcal{D}}^n) \right] \mathrm{d}\boldsymbol{x} \right) \left(\fint_{\Omega} \Pi_{\mathcal{D}} z_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) \\ = (\Pi_{\mathcal{D}} y_{\mathcal{D}}^n - \overline{y}_d, \Pi_{\mathcal{D}} z_{\mathcal{D}}) \qquad \forall z_{\mathcal{D}} \in X_{\mathcal{D}}, \end{aligned} \tag{5.5.2b}$$

$$u_h^n = \begin{cases} a & \text{on } A_{a,h}^n \\ \boldsymbol{\mathcal{P}}_{\mathcal{M}}(-\alpha^{-1} \Pi_{\mathcal{D}} p_{\mathcal{D}}^n) & \text{on } I_h^n \\ b & \text{on } A_{b,h}^n. \end{cases} \tag{5.5.2c}$$

The above algorithm consists of non-linear equations. It can however be approximated by a linearized system in the following way, thus leading to the final modified active set algorithm.

Instead of solving (5.5.2), solve

$$a_{\mathcal{D}}(y_{\mathcal{D}}^n, w_{\mathcal{D}}) + \rho \left(\fint_\Omega \Pi_{\mathcal{D}} y_{\mathcal{D}}^n \, \mathrm{d}\boldsymbol{x} \right) \left(\fint_\Omega \Pi_{\mathcal{D}} w_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) = (u_h^n + f, \Pi_{\mathcal{D}} w_{\mathcal{D}}) \qquad \forall w_{\mathcal{D}} \in X_{\mathcal{D}}, \tag{5.5.3a}$$

$$a_{\mathcal{D}}(z_{\mathcal{D}}, p_{\mathcal{D}}^n) + \rho \left(\fint_\Omega \Pi_{\mathcal{D}} p_{\mathcal{D}}^n \, \mathrm{d}\boldsymbol{x} \right) \left(\fint_\Omega \Pi_{\mathcal{D}} z_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) = (\Pi_{\mathcal{D}} y_{\mathcal{D}}^n - \bar{y}_d, \Pi_{\mathcal{D}} z_{\mathcal{D}}) + \rho S_{\mathcal{D}}^{n-1} \qquad \forall z_{\mathcal{D}} \in X_{\mathcal{D}}, \tag{5.5.3b}$$

$$u_h^n = \begin{cases} a & \text{on } A_{a,h}^n \\ \mathcal{P}_{\mathcal{M}}(-\alpha^{-1} \Pi_{\mathcal{D}} p_{\mathcal{D}}^{n-1}) & \text{on } I_h^n \\ b & \text{on } A_{b,h}^n, \end{cases} \tag{5.5.3c}$$

where

$$S_{\mathcal{D}}^{n-1} = \left(\fint_\Omega \Pi_{\mathcal{D}} z_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \right) \left(\fint_\Omega \left\{ \Pi_{\mathcal{D}} p_{\mathcal{D}}^{n-1} - P_{[a,b]} \left[\mathcal{P}_{\mathcal{M}}(-\alpha^{-1} \Pi_{\mathcal{D}} p_{\mathcal{D}}^{n-1}) \right] \right\} \mathrm{d}\boldsymbol{x} \right).$$

Note that (5.5.3c) can be re-written in the following more commonly used form:

$$u_h^n + \left(1 - \mathbb{I}_{a,h}^n - \mathbb{I}_{b,h}^n \right) \alpha^{-1} \Pi_{\mathcal{D}} p_{\mathcal{D}}^n = \mathbb{I}_{a,h}^n a + \mathbb{I}_{b,h}^n b,$$

where $\mathbb{I}_{a,h}^n$ and $\mathbb{I}_{b,h}^n$ denote the teristic functions of the sets $A_{a,h}^n$ and $A_{b,h}^n$ respecively.

Remark 5.5.1. *The convergence analysis of the proposed algorithm is a plan for future study. However, if* $(\Pi_{\mathcal{D}} y_{\mathcal{D}}^n, \Pi_{\mathcal{D}} p_{\mathcal{D}}^n)$ *converges weakly to* $(\Pi_{\mathcal{D}} \bar{y}_{\mathcal{D}}, \Pi_{\mathcal{D}} \bar{p}_{\mathcal{D}})$ *in* $H^1(\Omega)$ *and* u_h^n *converges to* $\bar{u}_h$ *in* $L^2(\Omega)$*, then the solution to* (5.5.3) *converges to the solution of* (5.5.1) *as* $n \to \infty$*.*

5.5.2 Examples

In this section, examples for the numerical solution of (5.1.1) are illustrated. Three specific schemes are used for the state and adjoint variables: conforming finite element method, non-conforming ($\mathrm{nc}\mathbb{P}_1$) finite element method and hybrid mimetic mixed (HMM) method. All three are GDMs with gradient discretisations with bounds on $C_{\mathcal{D}}$, order h estimate on $\mathrm{WS}_{\mathcal{D}}$, and satisfying assumptions **(A1)–(A4)**, and (5.4.49) on quasi-uniform meshes; see [48], Chapter 4 and Remark 5.3.4.

The control variable is discretised using piecewise constant functions on the corresponding meshes. The discrete solution is computed using the modified active set algorithm mentioned in Subsection 5.5.1 with zero as an initial guess for both u and μ. Here, U_a and Y_a denote the average values of the computed control $\bar{u}_h$ and the reconstructed state solution $\Pi_{\mathcal{D}} \bar{y}_{\mathcal{D}}$ respectively. Let **ni** denote the number of iterations required for the convergence of the modified active set algorithm, and f_a denote the numerical average of the source term f calculated using the same quadrature rule as in the implementation of the schemes, i.e,

$$f_a = \frac{1}{|\Omega|} \sum_{K \in \mathcal{M}} |K| f(\bar{\boldsymbol{x}}_K),$$

where $\overline{\boldsymbol{x}}_K$ denotes the center of mass of the cell K. This numerical average enables us to evaluate the quality of the quadrature rule for each mesh; in particular, since f has a zero average, any quantity of the order of f_a can be considered to be equal to zero, up to quadrature error. The relative errors are denoted by

$$\mathsf{err}_{\mathcal{D}}(\overline{y}) := \frac{\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \overline{y}_{\mathcal{M}}\|}{\|\overline{y}\|}, \quad \mathsf{err}_{\mathcal{D}}(\nabla\overline{y}) := \frac{\|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla\overline{y}\|}{\|\nabla\overline{y}\|},$$

$$\mathsf{err}_{\mathcal{D}}(\overline{p}) := \frac{\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \overline{p}_{\mathcal{M}}\|}{\|\overline{p}\|}, \quad \mathsf{err}_{\mathcal{D}}(\nabla\overline{p}) := \frac{\|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \nabla\overline{p}\|}{\|\nabla\overline{p}\|},$$

$$\mathsf{err}(\overline{u}) := \frac{\|\overline{u}_h - \overline{u}\|}{\|\overline{u}\|}, \quad \text{and} \quad \mathsf{err}(\widetilde{u}) := \frac{\|\widetilde{u}_h - \widetilde{u}\|}{\|\overline{u}\|}.$$

The data in the optimal control problem (5.1.1) are chosen as follows:

$$\overline{y} = 2\cos(\pi x)\cos(\pi y), \quad \overline{p} = 2\cos(\pi x)\cos(\pi y),$$
$$\alpha = 1, \quad \mathcal{U}_{\mathrm{ad}} = [a,b], \quad \overline{u} = P_{[a,b]}\left(-\overline{p} + \overline{c}\right),$$

where $\overline{c}$ is chosen to ensure that $\int_\Omega \overline{u}\,\mathrm{d}\boldsymbol{x} = 0$. The matrix-valued function is given by $A = \mathrm{Id}$ unless otherwise specified. The source term f and the desired state $\overline{y}_d$ are then computed using

$$f = -\Delta\overline{y} - \overline{u}, \quad \overline{y}_d = \overline{y} + \Delta\overline{p}.$$

Example 1 :

$\Omega = (0,1)^2$, $\rho = 10^{-4}$, $a = -1$, $b = 1$.
Consider the computational domain $\Omega = (0,1)^2$. We have $\overline{p}(x,y) = -\overline{p}(1-x,y)$ and, since $P_{[-1,1]}$ is odd, $P_{[-1,1]}(-\overline{p})(1-x,y) = -P_{[-1,1]}(-\overline{p})(x,y)$. Integrating this relation over Ω shows that $P_{[-1,1]}(-\overline{p})$ has a zero average and thus, by Lemma 5.4.9, that $\overline{c} = 0$. Hence, $\overline{u} = P_{[-1,1]}\left(-\overline{p}\right)$.
Conforming FEM: The discrete solution is computed on a family of uniform grids with mesh sizes $h = \frac{1}{2^i}, i = 2,\ldots,6$. Due to the symmetry of the mesh and of the solution, approximate solutions are also symmetric and thus have zero average at an order compatible with the stopping criterion in the active set algorithm (the discrete solutions of (5.4.1) are only approximated by this algorithm), see Table 5.1. As also seen in this table, the number of iterations of the modified active set algorithm remains very small, and independent on the mesh size. The error estimates and the convergence rates of the control, the post-processed control, the state and the adjoint variables are presented in Table 5.2. The numerical results corroborate Theorem 5.4.1, Theorem 5.4.5 and Corollary 5.4.7.
Non-Conforming FEM: For comparison, the solutions of the nc$\mathbb{P}_1$ finite element method on the same grids are computed. As for conforming FEM, the symmetry of the problem ensures that the approximation solutions have a zero average at an order dictated by the stopping criterion used in the active set algorithm. The results in Tables 5.3 and 5.4 are similar to those obtained with the conforming FE.

Table 5.1: Example 1, conforming FEM

h	U_a	f_a	Y_a	**ni**
0.250000	0.002752×10^{-13}	0.20699×10^{-14}	-0.008396×10^{-13}	2
0.125000	-0.008049×10^{-13}	0.20912×10^{-14}	-0.004684×10^{-13}	3
0.062500	-0.001370×10^{-13}	0.20548×10^{-14}	0.010486×10^{-13}	3
0.031250	-0.032432×10^{-13}	0.21299×10^{-14}	0.050725×10^{-13}	3
0.015625	-0.917129×10^{-13}	0.20367×10^{-14}	-0.495753×10^{-13}	3

Table 5.2: Convergence results, Example 1, conforming FEM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.325104	-	0.293424	-	0.333213	-
0.125000	0.086450	1.9110	0.129922	1.1753	0.089153	1.9021
0.062500	0.022176	1.9628	0.064767	1.0043	0.022967	1.9567
0.031250	0.005591	1.9879	0.032578	0.9914	0.005798	1.9860
0.015625	0.001402	1.9960	0.016337	0.9958	0.001453	1.9960

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.300144	-	0.464300	-	0.222006	-
0.125000	0.131070	1.1953	0.254036	0.8700	0.065430	1.7626
0.062500	0.064930	1.0134	0.126358	1.0075	0.016668	1.9728
0.031250	0.032599	0.9941	0.063453	0.9938	0.004226	1.9797
0.015625	0.016339	0.9965	0.031778	0.9977	0.001047	2.0136

Table 5.3: Example 1, nc$\mathbb{P}_1$FEM

h	U_a	f_a	Y_a	**ni**
0.250000	-0.003919×10^{-13}	0.206991×10^{-14}	0.000518×10^{-12}	3
0.125000	-0.067706×10^{-13}	0.209121×10^{-14}	-0.003856×10^{-12}	3
0.062500	0.030900×10^{-13}	0.205478×10^{-14}	0.017217×10^{-12}	3
0.031250	-0.075427×10^{-13}	0.212989×10^{-14}	-0.041154×10^{-12}	3
0.015625	-0.187208×10^{-13}	0.203674×10^{-14}	0.933499×10^{-12}	3

HMM scheme: This scheme was tested on a series of regular triangular meshes from [74] where the points $\mathcal{P}$ (see [51, Definition 2.21]) are located at the center of mass of the cells. These meshes are no longer symmetric and thus the symmetry of the approximate solution is lost. Zero averages are thus obtained up to quadrature error, see Table 5.5. It has been proved in [22, 54] that the state and adjoint equations enjoy a super-convergence property in L^2 norm for such a sequence of meshes; hence, as expected from Theorem 5.4.5, so does the scheme for the entire control problem after post-processing of the control. The errors in the energy norm and the L^2

Table 5.4: Convergence results, Example 1, nc$\mathbb{P}_1$FEM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.148286	-	0.409750	-	0.146306	-
0.125000	0.033274	2.1559	0.189599	1.1118	0.033499	2.1268
0.062500	0.008134	2.0324	0.093105	1.0260	0.008122	2.0443
0.031250	0.002023	2.0077	0.046348	1.0064	0.002025	2.0036
0.015625	0.000505	2.0019	0.023148	1.0016	0.000505	2.0041

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.408120	-	0.473176	-	0.284795	-
0.125000	0.189770	1.1047	0.250457	0.9178	0.071206	1.9999
0.062500	0.093102	1.0274	0.126078	0.9902	0.017716	2.0069
0.031250	0.046349	1.0063	0.063407	0.9916	0.004440	1.9965
0.015625	0.023149	1.0016	0.031770	0.9970	0.001109	2.0007

norm, together with their orders of convergence, are presented in Table 5.6.

Table 5.5: Example 1, HMM

h	U_a	f_a	Y_a	**ni**
0.250000	-0.016326	0.016324	-0.017271	4
0.125000	-0.005300	0.005300	-0.004968	4
0.062500	-0.001503	0.001503	-0.001277	3
0.031250	-0.000352	0.000352	-0.000321	3

Table 5.6: Convergence results, Example 1, HMM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.025586	-	0.143963	-	0.033104	-
0.125000	0.006764	1.9194	0.070970	1.0204	0.010044	1.7207
0.062500	0.001709	1.9847	0.035358	1.0052	0.002443	2.0397
0.031250	0.000429	1.9958	0.017663	1.0013	0.000619	1.9811

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.144012	-	0.214573	-	0.034890	-
0.125000	0.070972	1.0209	0.109352	0.9725	0.009603	1.8613
0.062500	0.035359	1.0052	0.055045	0.9903	0.002403	1.9989
0.031250	0.017663	1.0013	0.027551	0.9985	0.000605	1.9893

For all three methods (conforming $\mathbb{P}_1$ FEM, nc$\mathbb{P}_1$ FEM and HMM), the theoretical rates of convergence are confirmed by the numerical outputs. Without post-processing, an $\mathcal{O}(h)$ conver-

gence rate is obtained on the controls, which validates Theorem 5.4.1. With post-processing of the controls, the order of convergence of Theorem 5.4.5 is recovered. We also notice that the super-convergence on the state and adjoint stated in Corollary 5.4.7 is confirmed, provided that the exact state and adjoint are properly projected (usage of the functions $\overline{y}_{\mathcal{M}}$ and $\overline{p}_{\mathcal{M}}$ in $\mathsf{err}_{\mathcal{D}}(\overline{y})$ and $\mathsf{err}_{\mathcal{D}}(\overline{p})$).

Remark 5.5.2. *As seen in Table 5.5, the modified active set algorithm converges in very few iterations if $\rho = 10^{-4}$. We however found that, if $\rho = 1$, the modified active set algorithm no longer converges. Further work will investigate in more depth the convergence analysis of the modified active set algorithm, to understand better its dependency with respect to ρ.*

Example 2 :

$A = 100Id$, $\Omega = (0,1)^2$, $\rho = 10^{-2}$, $a = -1$, $b = 1$.
In this subsection, some numerical results for the control problem defined on the unit square domain $\Omega = (0,1)^2$ and $A = 100Id$ are presented. As explained in Example 1, $a = -1$ and $b = 1$ imply $\overline{c} = 0$.

Conforming FEM: The details of active set algorithm for the conforming finite element method are provided in Table 5.7. As expected, the symmetries of the problem provide approximate solutions with a nearly perfect average. For such grids, we obtain super-convergence result for the post-processed control. The errors between the true and computed solutions are computed for different mesh sizes and presented in Table 5.8. They still follow the expected theoretical rates, and the number of iterations of the active set algorithm remain small.

Table 5.7: Example 2, conforming FEM

h	U_a	f_a	Y_a	**ni**
0.250000	0.002280×10^{-11}	0.209361×10^{-12}	0.009117×10^{-11}	2
0.125000	0.018065×10^{-11}	0.209375×10^{-12}	0.024496×10^{-11}	3
0.062500	0.027564×10^{-11}	0.209400×10^{-12}	-0.012778×10^{-11}	3
0.031250	-0.103755×10^{-11}	0.209420×10^{-12}	-0.028850×10^{-11}	3
0.015625	-0.168277×10^{-11}	0.209297×10^{-12}	0.624160×10^{-11}	3

Non-Conforming FEM: The results, presented in Tables 5.9 and 5.10, are similar to those for the conforming FEM.
HMM scheme: The results are presented in Tables 5.11 and 5.12. They are qualitatively similar to those for Example 1. As mentioned before, the algorithm is not convergent for $\rho = 1$.

Example 3 :

$\Omega = (0,1)^2$, $\rho = 10^{-4}$, $a = -0.5$, $b = 1$. In this case, since $P_{[a,b]}$ is no longer odd, $P_{[a,b]}(-\overline{p})$ no longer has a zero average and, to compute $\mathsf{err}_{\mathcal{D}}(\overline{u})$, we need to find $\overline{c}$ such that $\int_\Omega P_{[a,b]}(-\overline{p}+$

Table 5.8: Convergence results, Example 2, conforming FEM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.328223	-	0.296014	-	0.328304	-
0.125000	0.087182	1.9126	0.130232	1.1846	0.087209	1.9125
0.062500	0.022409	1.9600	0.064814	1.0067	0.022417	1.9599
0.031250	0.005653	1.9870	0.032584	0.9921	0.005655	1.9870
0.015625	0.001417	1.9963	0.016338	0.9960	0.001417	1.9963

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.296080	-	0.463836	-	0.218080	-
0.125000	0.130243	1.1848	0.253857	0.8696	0.064390	1.7600
0.062500	0.064816	1.0068	0.126333	1.0068	0.016358	1.9768
0.031250	0.032584	0.9922	0.063449	0.9936	0.004145	1.9807
0.015625	0.016338	0.9960	0.031778	0.9976	0.001026	2.0139

Table 5.9: Example 2, nc$\mathbb{P}_1$FEM

h	U_a	f_a	Y_a	**ni**
0.250000	-0.000977$\times10^{-10}$	0.209361$\times10^{-12}$	0.000739 $\times10^{-10}$	3
0.125000	-0.006518$\times10^{-10}$	0.209375$\times10^{-12}$	-0.000960$\times10^{-10}$	3
0.062500	-0.004320$\times10^{-10}$	0.209400$\times10^{-12}$	0.002330$\times10^{-10}$	3
0.031250	-0.007236$\times10^{-10}$	0.209420 $\times10^{-12}$	0.054029 $\times10^{-10}$	3
0.015625	0.346321$\times10^{-10}$	0.209297$\times10^{-12}$	0.276914 $\times10^{-10}$	3

Table 5.10: Convergence results, Example 2, nc$\mathbb{P}_1$FEM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.148286	-	0.409750	-	0.148262	-
0.125000	0.033263	2.1564	0.189599	1.1118	0.033265	2.1561
0.062500	0.008131	2.0324	0.093105	1.0260	0.008131	2.0325
0.031250	0.002022	2.0077	0.046348	1.0064	0.002022	2.0076
0.015625	0.000505	2.0019	0.023148	1.0016	0.000505	2.0019

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.409732	-	0.473136	-	0.286537	-
0.125000	0.189600	1.1117	0.250390	0.9181	0.071004	2.0128
0.062500	0.093105	1.0260	0.126079	0.9899	0.017719	2.0026
0.031250	0.046348	1.0064	0.063407	0.9916	0.004439	1.9970
0.015625	0.023148	1.0016	0.031770	0.9970	0.001109	2.0004

$\overline{c})\,\mathrm{d}\boldsymbol{x} = 0$. This $\overline{c}$ can be found by a bisection method, by computing the averages on a very

Table 5.11: Example 2, HMM

h	U_a	f_a	Y_a	**ni**
0.250000	-1.000000	1.817180	81.718002	-
0.125000	-0.528617	0.523335	-0.528289	6
0.062500	-0.136036	0.134678	-0.135812	5
0.031250	-0.034208	0.033866	-0.034178	5

Table 5.12: Convergence results, Example 2, HMM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	81.717083	-	0.143996	-	392673.3	-
0.125000	0.528572	7.2724	0.070971	1.0207	0.876152	18.7737
0.062500	0.135884	1.9597	0.035359	1.0052	0.216641	2.0159
0.031250	0.034196	1.9905	0.017663	1.0013	0.054238	1.9979

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.143987	-	1.686504	-	1.691346	-
0.125000	0.070970	1.0207	0.878849	0.9403	0.874778	0.9512
0.062500	0.035358	1.0052	0.237129	1.8899	0.230873	1.9218
0.031250	0.017663	1.0013	0.064673	1.8744	0.058612	1.9778

thin mesh and bisecting until we find a proper $\overline{c}$. Using a mesh of size $h = 0.00195$, we find $c \approx -0.24596797$.

Conforming FEM: The numerical results obtained using conforming finite element method are shown in Tables 5.13 and 5.14 respectively. Since there is a loss of symmetry, the approximate solutions have zero averages only up to quadrature error (compare U_a and f_a in Table 5.13). Here, it is observed that the modified active set algorithm converges only when $\rho \leq 10^{-1}$. When it does, though, the number of iterations remain very small. As in Examples 1 and 2, the theoretical rates of convergence are confirmed by these numerical outputs.

Table 5.13: Example 3, conforming FEM

h	U_a	f_a	Y_a	**ni**
0.250000	0.0020160	-0.0020160	0.201602×10^{-6}	4
0.125000	0.0055595	-0.0055595	0.555952×10^{-6}	4
0.062500	-0.0004794	0.0004795	-0.047944×10^{-6}	4
0.031250	0.0001470	-0.0001470	0.014705×10^{-6}	5
0.015625	-0.0000136	0.0000136	-0.001362×10^{-6}	5

Table 5.14: Convergence results, Example 3, conforming FEM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.325266	-	0.293567	-	0.346894	-
0.125000	0.086733	1.9070	0.130041	1.1747	0.097046	1.8378
0.062500	0.022291	1.9601	0.064790	1.0051	0.025081	1.9521
0.031250	0.005624	1.9868	0.032581	0.9917	0.006219	2.0117
0.015625	0.001410	1.9963	0.016337	0.9959	0.001569	1.9865

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.300149	-	0.466701	-	0.234197	-
0.125000	0.131075	1.1953	0.268982	0.7950	0.064265	1.8656
0.062500	0.064931	1.0134	0.138258	0.9602	0.016053	2.0012
0.031250	0.032599	0.9941	0.069620	0.9898	0.003996	2.0064
0.015625	0.016339	0.9965	0.034944	0.9945	0.001002	1.9950

Non-Conforming FEM: The results are similar to those obtained with the conforming FEM (see Tables 5.15 and 5.16).

Table 5.15: Example 3, nc$\mathbb{P}_1$FEM

h	U_a	f_a	Y_a	**ni**
0.250000	0.002016	-0.002016	0.0201601×10^{-5}	4
0.125000	0.005560	-0.005559	-0.1301803×10^{-5}	5
0.062500	-0.000480	0.000479	-0.0017424×10^{-5}	5
0.031250	0.000147	-0.000147	0.0011093×10^{-5}	5
0.015625	-0.000014	0.000014	-0.0001436×10^{-5}	5

HMM scheme: Tables 5.17 and 5.18 show that the HMM scheme behave similarly to the FEMs. Note that, here too, the convergence of the modified active set algorithm is only observed if $\rho \leq 10^{-1}$.

Table 5.16: Convergence results, Example 3, nc$\mathbb{P}_1$FEM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.148286	-	0.409750	-	0.141781	-
0.125000	0.033270	2.1561	0.189600	1.1118	0.032889	2.1080
0.062500	0.008133	2.0324	0.093105	1.0260	0.008041	2.0322
0.031250	0.002022	2.0077	0.046348	1.0064	0.001994	2.0118
0.015625	0.000505	2.0019	0.023148	1.0016	0.000498	2.0015

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.408120	-	0.494425	-	0.269888	-
0.125000	0.189770	1.1047	0.269866	0.8735	0.080165	1.7513
0.062500	0.093102	1.0274	0.138223	0.9652	0.019967	2.0054
0.031250	0.046349	1.0063	0.069625	0.9893	0.005091	1.9715
0.015625	0.023149	1.0016	0.034941	0.9947	0.001283	1.9883

Table 5.17: Example 3, HMM

h	U_a	f_a	Y_a	**ni**
0.250000	-0.019043	0.019041	-0.017271	5
0.125000	-0.005459	0.005459	-0.004968	5
0.062500	-0.001300	0.001300	-0.001277	5
0.031250	-0.000331	0.000331	-0.000321	5

Table 5.18: Convergence results, Example 3, HMM

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order
0.250000	0.026037	-	0.144014	-	0.055044	-
0.125000	0.006841	1.9284	0.070972	1.0209	0.013361	2.0425
0.062500	0.001728	1.9853	0.035359	1.0052	0.003342	1.9995
0.031250	0.000433	1.9956	0.017663	1.0013	0.000843	1.9869

h	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}(\overline{u})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.144013	-	0.237647	-	0.050184	-
0.125000	0.070972	1.0209	0.120112	0.9844	0.013482	1.8962
0.062500	0.035359	1.0052	0.061226	0.9722	0.003468	1.9586
0.031250	0.017663	1.0013	0.030583	1.0014	0.000872	1.9916

Chapter 6

Numerical approximation of optimal control problems using the Hessian discretisation method

6.1 Introduction

This chapter deals with the numerical approximation of optimal control problems governed by fourth order linear elliptic equations with clamped boundary conditions using the Hessian discretisation method. The HDM, an abstract framework that covers several numerical schemes and establishes convergence analysis for fourth order linear and semi-linear elliptic PDEs, is discussed in Chapters 2 and 3.

The HDM for fourth order linear elliptic equations is a generic convergence analysis framework based on a set of four discrete elements known as a Hessian discretisation and three core properties namely, coercivity, consistency and limit-conformity of Hessian discretisation. Some examples of schemes that fit into the HDM framework are the conforming finite element methods, the Adini and Morley non-conforming finite element methods, the finite volume methods and a method based on gradient recovery (GR) operators. A generic error estimate is established in L^2, H^1 and H^2-like norms in the HDM framework in Section 2.4. Also, improved L^2 and H^1-like error estimates compared to that in the energy norm in the abstract setting are derived in Sections 2.5 and 2.6. Under regularity assumption, the improved L^2 estimate provides a quadratic rate of convergence for the FEMs, the Adini and Morley ncFEMs and the GR methods (see Proposition 2.5.4).

The problems described by fourth order linear elliptic equations arise from fluid mechanics and solid mechanics such as bending of elastic plates [42]. In [64], a mixed formulation has been used for the biharmonic control problem where the state variable is discretized in primal mixed form using continuous piecewise biquadratic finite elements. In [73], a C^0 interior penalty method has been analyzed and a discontinuous finite element method has been investigated in [40]. The general fourth-order elliptic problem on polygonal domains with clamped boundary conditions

is considered in [72]. Error analysis for a stable C^0 interior penalty method is derived under minimal regularity assumptions on the exact solution. To the best of our knowledge, the control problem using the method based on gradient recovery operators and the finite volume methods have not been studied in literature.

In this chapter, the optimal control problems governed by fourth order linear elliptic equations are discretised using the Hessian discretisation method. The basic error estimates applied to the control problems are established in a very generic setting with the help of three core properties associated with the Hessian discretisation. As a result, for these problems, all the schemes entering the HDM framework converge, in particular, FEMs, Adini and Morley ncFEMs, FVMs and GR methods. Under regularity assumption, for conforming FEMs, Adini and Morley ncFEMs, and GR methods, the basic error estimate yields $\mathcal{O}(h)$ convergence for the control variable. Given the control is discretised using piecewise constant functions, this rate is optimal. However, using a post-processing step and following the ideas of Chapter 4 taking into account of the additional challenges offered by fourth order problems, convergence rate can be improved to $\mathcal{O}(h^2)$. Numerical experiments are performed for the gradient recovery method and finite volume methods. In the numerical implementation, the discretisation problem is solved using the primal-dual active set strategy [109].

This chapter is organised as follows. Section 6.2 deals with the optimal control problem governed by fourth order linear elliptic equation with clamped boundary conditions and the Hessian discretisation method for the optimal control problem. Section 6.3.1 establishes basic error estimates for the control, state and adjoint variables in the HDM framework. The superconvergence result is obtained in Section 6.3.2 using a post-processing step and under a few generic assumptions on the Hessian discretisation which are discussed in detail for conforming FEMs, Adini and Morley ncFEMs, and GR methods. The numerical results for the gradient recovery method and finite volume method are presented in Section 6.4.

6.2 The optimal control problem

Consider the distributed optimal control problem governed by fourth order linear elliptic equations defined by:

$$\min_{u\in\mathcal{U}_{\rm ad}} J(u) \quad \text{subject to} \tag{6.2.1a}$$

$$\sum_{i,j,k,l=1}^{d} \partial_{kl}(a_{ijkl}\partial_{ij}y(u)) = f + \mathcal{C}u \quad \text{in } \Omega, \tag{6.2.1b}$$

$$y(u) = \frac{\partial y(u)}{\partial n} = 0 \quad \text{on } \partial\Omega, \tag{6.2.1c}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded domain with boundary $\partial\Omega$, $\partial_{kl} = \frac{\partial^2}{\partial x_k \partial x_l}$, u is the control variable and $y(u)$ is the state variable associated with u. The coefficients a_{ijkl} are measurable bounded functions which satisfy the condition, $a_{ijkl} = a_{jikl} = a_{ijlk} = a_{klij}$ for $i,j,k,l = 1,\cdots,d$ and n is the unit

outward normal to the boundary $\partial\Omega$. The load function f belongs to $L^2(\Omega)$, $\mathcal{C} \in \mathcal{L}(L^2(\omega), L^2(\Omega))$ is a localization operator defined by $\mathcal{C}u(\boldsymbol{x}) = u(\boldsymbol{x})\chi_\omega(\boldsymbol{x})$, where χ_ω is the characteristic function of $\omega \subset L^2(\Omega)$.

$$J(u) := \frac{1}{2}\|y(u) - \overline{y}_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u - \overline{u}_d\|^2_{L^2(\omega)}, \tag{6.2.2}$$

is the cost functional, $\alpha > 0$ is a fixed regularization parameter, $\overline{y}_d$ is the desired state variable for $y(u)$, $\overline{u}_d \in L^2(\omega)$ is the desired control variable and $\mathcal{U}_{\text{ad}} \subset L^2(\omega)$ is a non-empty, convex and closed admissible space of controls.

For a given $u \in \mathcal{U}_{\text{ad}}$, the weak formulation of (6.2.1b)-(6.2.1c) seeks $y = y(u) \in H^2_0(\Omega)$ such that $\forall\, w \in H^2_0(\Omega)$,

$$a(y(u), w) = \int_\Omega (f + \mathcal{C}u) w \,\mathrm{d}\boldsymbol{x}, \tag{6.2.3}$$

where

$$a(z, w) = \sum_{i,j,k,l=1}^{d} \int_\Omega a_{ijkl}\partial_{ij} z \partial_{kl} w \,\mathrm{d}\boldsymbol{x} = \int_\Omega \mathcal{H}^B z : \mathcal{H}^B w \,\mathrm{d}\boldsymbol{x} \text{ with } \mathcal{H}^B w = B\mathcal{H}w.$$

As in Chapter 2, assume in the following that B is constant over Ω, and that the following coercivity property holds:

$$\exists \rho > 0 \text{ such that } \|\mathcal{H}^B v\| \geq \rho \|v\|_{H^2(\Omega)} \quad \forall v \in H^2_0(\Omega). \tag{6.2.4}$$

Hence, (6.2.3) has a unique solution by the Lax–Milgram lemma.
The control problem (6.2.1) has a unique weak solution $(\overline{y}, \overline{u}) \in H^2_0(\Omega) \times \mathcal{U}_{\text{ad}}$ and there exists an adjoint state $\overline{p} \in H^2_0(\Omega)$ associated with $(\overline{y}, \overline{u})$ such that the triplet $(\overline{y}, \overline{p}, \overline{u}) \in H^2_0(\Omega) \times H^2_0(\Omega) \times \mathcal{U}_{\text{ad}}$ satisfies the Karush-Kuhn-Tracker (KKT) optimality conditions [90]:

$$a(\overline{y}, w) = (f + \mathcal{C}\overline{u}, w) \qquad \forall\, w \in H^2_0(\Omega), \tag{6.2.5a}$$

$$a(w, \overline{p}) = (\overline{y} - \overline{y}_d, w) \qquad \forall\, w \in H^2_0(\Omega), \tag{6.2.5b}$$

$$(\mathcal{C}\overline{p} + \alpha(\overline{u} - \overline{u}_d), v - \overline{u}) \geq 0 \qquad \forall\, v \in \mathcal{U}_{\text{ad}}. \tag{6.2.5c}$$

Note that the adjoint operator of $\mathcal{C}$ is denoted by $\mathcal{C}^*$ and in this case, $\mathcal{C}^* = \mathcal{C}$. Define $P_{[a,b]} : \mathbb{R} \to [a, b]$ by, for all $s \in \mathbb{R}$, $P_{[a,b]}(s) := \min(b, \max(a, s))$. From the first order optimality condition (6.2.5c), the following pointwise relation hold true for a.e. $\boldsymbol{x} \in \Omega$ [109, Theorem 2.28]:

$$\overline{u}(\boldsymbol{x}) = P_{[a,b]}\left(\overline{u}_d(\boldsymbol{x}) - \frac{\mathcal{C}}{\alpha}\overline{p}\right). \tag{6.2.6}$$

6.2.1 The Hessian discretisation method for the control problem

Let $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_\mathcal{D}, \nabla_\mathcal{D}, \mathcal{H}^B_\mathcal{D})$ be a B–Hessian discretisation for fourth order linear equations in the sense of Definition 2.3.1. Let $\mathcal{U}_{\text{ad},h} = \mathcal{U}_{\text{ad}} \cap \mathcal{U}_h$, where $\mathcal{U}_h$ is a finite-dimensional subspace of

$L^2(\omega)$. Given a B–Hessian discretisation $\mathcal{D}$, the corresponding Hessian scheme for (6.2.5) seeks $(\overline{y}_\mathcal{D}, \overline{p}_\mathcal{D}, \overline{u}_h) \in X_{\mathcal{D},0} \times X_{\mathcal{D},0} \times \mathcal{U}_{\mathrm{ad},h}$ satisfying the KKT optimality conditions [90]:

$$a_\mathcal{D}(\overline{y}_\mathcal{D}, w_\mathcal{D}) = (f + \mathcal{C}\overline{u}_h, \Pi_\mathcal{D} w_\mathcal{D}) \qquad \forall\, w_\mathcal{D} \in X_{\mathcal{D},0}, \tag{6.2.7a}$$

$$a_\mathcal{D}(w_\mathcal{D}, \overline{p}_\mathcal{D}) = (\Pi_\mathcal{D}\overline{y}_\mathcal{D} - \overline{y}_d, \Pi_\mathcal{D} w_\mathcal{D}) \qquad \forall\, w_\mathcal{D} \in X_{\mathcal{D},0}, \tag{6.2.7b}$$

$$(\mathcal{C}\Pi_\mathcal{D}\overline{p}_\mathcal{D} + \alpha(\overline{u}_h - \overline{u}_d), v_h - \overline{u}_h) \geq 0 \qquad \forall\, v_h \in \mathcal{U}_{\mathrm{ad},h}, \tag{6.2.7c}$$

where

$$a_\mathcal{D}(\overline{y}_\mathcal{D}, w_\mathcal{D}) = \int_\Omega \mathcal{H}_\mathcal{D}^B \overline{y}_\mathcal{D} : \mathcal{H}_\mathcal{D}^B w_\mathcal{D} \, \mathrm{d}\boldsymbol{x}.$$

As in the continuous case, existence and uniqueness of a solution to (6.2.7) follows from standard variational theorems [90, 109].
Some examples of Hessian discretisation method are the conforming FEMs, the Adini and Morley nonconforming FEMs, the methods based on gradient recovery operators and the finite volume methods, see Chapter 2 for more details.

6.3 Basic error estimate and super-convergence

This section is devoted to the basic error estimate and super-convergence results for the HDM applied to the control problem. The basic error estimate provides a linear rate of convergence on the control problem for the finite element methods and the methods based on gradient recovery operator. However, the superconvergence result can be obtained under a superconvergence assumption on the state and adjoint equations and some additional assumptions. The proofs of the results stated in this section follow by adapting the corresponding proofs in Chapter 4 to accounts for the Hessian discretisation method.

6.3.1 Basic error estimate for the control problem

Recall the measures associated with the HD from Chapter 2, namely coercivity ($C_\mathcal{D}^B$), consistency ($S_\mathcal{D}^B$) and limit-conformity ($W_\mathcal{D}^B$) defined by (2.4.1)-(2.4.3). Also, recall the definition (2.5.3) of $\mathrm{WS}_\mathcal{D}^B$, that is, for $\phi \in H^2(\Omega)$ with $\mathcal{H}\phi \in H(\Omega)$,

$$\mathrm{WS}_\mathcal{D}^B(\phi) := W_\mathcal{D}^B(\mathcal{H}\phi) + S_\mathcal{D}^B(\phi).$$

The following proposition enables to establish the basic error estimates for the control problem. For that, let $\mathcal{D} = (X_{\mathcal{D},0}, \Pi_\mathcal{D}, \nabla_\mathcal{D})$ be a HD in the sense of the Definiton 2.3.1. If $F \in L^2(\Omega)$, recall the related Hessian scheme for linear elliptic problem

$$\sum_{i,j,k,l=1}^{d} \partial_{kl}(a_{ijkl}\partial_{ij}\psi) = F \quad \text{in } \Omega,$$

$$\psi = \frac{\partial \psi}{\partial n} = 0 \quad \text{on } \partial\Omega,$$

seeks $\psi_{\mathcal{D}} \in X_{\mathcal{D},0}$ such that, for all $w_{\mathcal{D}} \in X_{\mathcal{D},0}$,

$$a_{\mathcal{D}}(\psi_{\mathcal{D}}, w_{\mathcal{D}}) = (F, \Pi_{\mathcal{D}} w_{\mathcal{D}}). \tag{6.3.1}$$

Proposition 6.3.1 (Stability of Hessian schemes)**.** *If $\psi_{\mathcal{D}}$ is the solution to the Hessian scheme* (6.3.1)*, then*

$$\|\mathcal{H}^B_{\mathcal{D}} \psi_{\mathcal{D}}\| \leq C^B_{\mathcal{D}} \|F\|, \quad \|\nabla_{\mathcal{D}} \psi_{\mathcal{D}}\| \leq (C^B_{\mathcal{D}})^2 \|F\| \quad \textit{and} \quad \|\Pi_{\mathcal{D}} \psi_{\mathcal{D}}\| \leq (C^B_{\mathcal{D}})^2 \|F\|. \tag{6.3.2}$$

Proof. A choice of $w_{\mathcal{D}} = \psi_{\mathcal{D}}$ in (6.3.1) and the definition of $C^B_{\mathcal{D}}$ given by (2.4.1) lead to

$$\|\mathcal{H}^B_{\mathcal{D}} \psi_{\mathcal{D}}\|^2 \leq \|F\| \|\Pi_{\mathcal{D}} \psi_{\mathcal{D}}\| \leq C^B_{\mathcal{D}} \|F\| \|\mathcal{H}^B_{\mathcal{D}} \psi_{\mathcal{D}}\|.$$

Hence the first inequality in (6.3.2) follows. The remaining two estimates follows from the definition of $C^B_{\mathcal{D}}$ in (2.4.1). □

The notation $X \lesssim Y$ means that $X \leq CY$ for some C depending only on Ω, ρ and an upper bound of $C^B_{\mathcal{D}}$ defined by (2.4.1).
As in Chapters 4, $\mathrm{Pr}_h : L^2(\Omega) \to \mathcal{U}_h$ denotes the L^2 orthogonal projector on $\mathcal{U}_h$ for the standard scalar product. The following theorem establishes basic error estimates for the control problem in the framework of HDM.

Theorem 6.3.2 (Control estimate)**.** *Let $\mathcal{D}$ be a Hessian discretisation in the sense of Definition 2.3.1, $(\overline{y}, \overline{p}, \overline{u})$ be the solution to* (6.2.5) *and $(\overline{y}_{\mathcal{D}}, \overline{p}_{\mathcal{D}}, \overline{u}_h)$ be the solution to* (6.2.7)*. Assume that*

$$\mathrm{Pr}_h(\mathcal{U}_{\mathrm{ad}}) \subset \mathcal{U}_{\mathrm{ad},h}. \tag{6.3.3}$$

Then,

$$\begin{aligned} \sqrt{\alpha}\|\overline{u} - \overline{u}_h\| \lesssim \sqrt{\alpha}\|\alpha^{-1}\overline{p} - \mathrm{Pr}_h(\alpha^{-1}\overline{p})\| + (\sqrt{\alpha}+1)\|\overline{u} - \mathrm{Pr}_h\overline{u}\| \\ + \sqrt{\alpha}\|\overline{u}_d - \mathrm{Pr}_h\overline{u}_d\| + \frac{1}{\sqrt{\alpha}}\mathrm{WS}^B_{\mathcal{D}}(\overline{p}) + \mathrm{WS}^B_{\mathcal{D}}(\overline{y}). \end{aligned} \tag{6.3.4}$$

Proof. The proof follows by recalling Remark 4.3.3, the KKT optimality system (6.2.5c) and (6.2.7c), Proposition 6.3.1 and Theorem 2.4.4. □

Remark 4.3.3 shows that the basic error estimates for the control problem using the GDM (for second order problems) and HDM (for fourth order problems) do not depend on the order of the PDEs. This is one of the major advantages of carrying out the analysis in the generic GDM and HDM framework.

The following proposition establishes error estimates for the state and adjoint variables. The proof is similar to that of Proposition 4.3.4 and is skipped.

Proposition 6.3.3 (State and adjoint error estimates)**.** *Let $\mathcal{D}$ be a HD, $(\overline{y}, \overline{p}, \overline{u})$ be the solution to* (6.2.5) *and $(\overline{y}_{\mathcal{D}}, \overline{p}_{\mathcal{D}}, \overline{u}_h)$ be the solution to* (6.2.7)*. Then the following error estimates hold:*

$$\|\Pi_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \overline{y}\| + \|\nabla_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \nabla\overline{y}\| + \|\mathcal{H}^B_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \mathcal{H}^B\overline{y}\| \lesssim \|\overline{u} - \overline{u}_h\| + \mathrm{WS}^B_{\mathcal{D}}(\overline{y}), \tag{6.3.5}$$

$$\|\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \overline{p}\| + \|\nabla_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \nabla\overline{p}\| + \|\mathcal{H}^B_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \mathcal{H}^B\overline{p}\| \lesssim \|\overline{u} - \overline{u}_h\| + \mathrm{WS}^B_{\mathcal{D}}(\overline{y}) + \mathrm{WS}^B_{\mathcal{D}}(\overline{p}). \tag{6.3.6}$$

Remark 6.3.4 (Rates of convergence for the control problem)**.** *Recalling Remark 2.4.15, under sufficient smoothness assumption on $\overline{u}_d$, if $(\overline{y},\overline{p},\overline{u}) \in H^4(\Omega)^2 \times H^1(\Omega)$ and $\mathcal{U}_h$ is made of piecewise constant functions then* (6.3.4)*,* (6.3.5) *and* (6.3.6) *give linear rates of convergence for low-order conforming FEMs, ncFEMs and gradient recovery methods. Also for the finite volume method, Theorem 6.3.2 and Proposition 6.3.3 provide an $\mathcal{O}(h^{1/4}|\ln(h)|)$ (in dimension $d=2$) or $\mathcal{O}(h^{3/13})$ (in dimension $d=3$) error estimate.*

6.3.2 Super-convergence for post-processed controls

In this section, a super-convergence result for the HDM applied to control problem by imposing additional assumptions **(A1)**-**(A4)** is presented. The post-processing step establishes a super-convergence result for the control variable by following the ideas of Chapter 4. These assumptions cover for example, the conforming FEMs, the Adini and Morley ncFEMs, and the GR methods. Let $\mathcal{M}$ be a mesh of Ω, that is a finite partition of Ω into polygonal/polyhedral cells (Definition 1.4.1) such that each cell $K \in \mathcal{M}$ is star-shaped with respect to its centroid $\overline{\boldsymbol{x}}_K$. Assume that ω is a polygonal/polyhedral domain such that $\mathcal{M}_{|\omega}$ yields a mesh for ω. The admissible set of controls $\mathcal{U}_{\rm ad}$ is given by

$$\mathcal{U}_{\rm ad} = \{u \in L^2(\omega) \,:\, a \le u \le b \text{ a.e.}\}. \tag{6.3.7}$$

The control variable is discretised by piecewise constant functions on this partition and is given by

$$\mathcal{U}_h = \{v : \Omega \to \mathbb{R} \,:\, \forall K \in \mathcal{M}\,,\ v_{|K} \text{ is a constant}\}. \tag{6.3.8}$$

Recall the projection operator $\boldsymbol{\mathcal{P}}_{\mathcal{M}} : L^1(\Omega) \to \mathcal{U}_h$ (orthogonal projection on piecewise constant functions on $\mathcal{M}$) associated with the superconvergence result from Chapter 4, that is,

$$\forall v \in L^1(\Omega)\,,\ \forall K \in \mathcal{M}\,, \quad (\boldsymbol{\mathcal{P}}_{\mathcal{M}} v)_{|K} := \fint_K v \,\mathrm{d}\boldsymbol{x}.$$

Let us impose the following assumptions in order to obtain superconvergence result. These are an extension of the assumptions for GDM explained in Chapter 4 (see Section 4.3.2) to HDM. The discussion on the assumptions for HDM are also stated in this section.

(A1) [*Approximation error*] For each $w \in H^2(\Omega)$, there exists $w_{\mathcal{M}} \in L^2(\Omega)$ such that:

i) If $w \in H^4(\Omega) \cap H^2_0(\Omega)$ solves $\sum_{i,j,k,l=1}^d \partial_{kl}(a_{ijkl}\partial_{ij} w) = g \in L^2(\Omega)$, and $w_{\mathcal{D}}$ is the solution to the corresponding HS, then

$$\|\Pi_{\mathcal{D}} w_{\mathcal{D}} - w_{\mathcal{M}}\| \lesssim h^2 \|g\|. \tag{6.3.9}$$

ii) For any $w \in H^2(\Omega)$, it holds

$$\forall v_{\mathcal{D}} \in X_{\mathcal{D},0}\,,\ \left|(w - w_{\mathcal{M}}, \Pi_{\mathcal{D}} v_{\mathcal{D}})\right| \lesssim h^2 \|\Pi_{\mathcal{D}} v_{\mathcal{D}}\| \, \|w\|_{H^2(\Omega)} \tag{6.3.10}$$

and

$$\|\boldsymbol{\mathcal{P}}_{\mathcal{M}}(w - w_{\mathcal{M}})\| \lesssim h^2 \|w\|_{H^2(\Omega)}. \tag{6.3.11}$$

(A2) [*Projection estimate*] The estimate $\|\Pi_{\mathcal{D}} v_{\mathcal{D}} - \mathcal{P}_{\mathcal{M}}(\Pi_{\mathcal{D}} v_{\mathcal{D}})\| \lesssim h \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\|$ holds for any $v_{\mathcal{D}} \in X_{\mathcal{D},0}$.

(A3) [*Discrete Sobolev imbedding*] For all $v_{\mathcal{D}} \in X_{\mathcal{D},0}$, it holds

$$\|\Pi_{\mathcal{D}} v_{\mathcal{D}}\|_{L^\infty(\Omega)} \lesssim \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\|.$$

The last assumption is identical to the assumption **(A4)** in Chapter 4: using the notation $\Omega_{1,\mathcal{M}}$ defined there,

(A4) $|\Omega_{1,\mathcal{M}}| \lesssim h$ and $\overline{u}_{|\Omega_{1,\mathcal{M}}} \in W^{1,\infty}(\mathcal{M}_1)$.

The post-processed continuous and discrete controls are defined by

$$\widetilde{u}(\boldsymbol{x}) = P_{[a,b]}\left(\mathcal{P}_{\mathcal{M}}\overline{u}_d(\boldsymbol{x}) - \frac{\mathcal{C}}{\alpha}\overline{p}_{\mathcal{M}}\right), \quad \widetilde{u}_h(\boldsymbol{x}) = P_{[a,b]}\left(\mathcal{P}_{\mathcal{M}}\overline{u}_d(\boldsymbol{x}) - \frac{\mathcal{C}}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\right) \tag{6.3.12}$$

where $\overline{p}_{\mathcal{M}}$ is defined as in **(A1)**.

The assumptions **(A1)**-**(A3)** for the conforming FEMs, the Adini and Morley ncFEMs, and the method based on GR operators are discussed below. See Chapter 4 for a discussion on **(A4)**.

Conforming FEMs

For the conforming FEMs, super-convergence result (6.3.9) for elliptic equations usually holds with $w_{\mathcal{M}} = w$ (see Proposition 2.5.4). In that case, (6.3.10) and (6.3.11) are trivially satisfied. Assumption **(A2)** then follows from a simple Taylor expansion and using the definition of $C^B_{\mathcal{D}}$. The discrete Sobolev embedding **(A3)** is straightforward for the conforming FEM using the continuous Sobolev embedding $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$.

Nonconforming FEMs

As in the conforming FEMs, the superconvergence result (6.3.9) for the Adini and Morley ncFEMs is satisfied by taking $w_{\mathcal{M}} = w$ (Proposition 2.5.4). Then (6.3.10) and (6.3.11) are trivial. Since $\nabla_{\mathcal{D}} v_{\mathcal{D}}$ is the classical broken gradient (i.e. the gradient of $\Pi_{\mathcal{D}} v_{\mathcal{D}}$ in each cell), a use of Taylor expansion and the definition of $C^B_{\mathcal{D}}$ leads to **(A2)** for both Adini and Morley ncFEMs. Assumption **(A3)** is verified with the help of a companion operator. The companion operator $E_{\mathcal{D}} : X_{\mathcal{D},0} \to H^2_0(\Omega)$ for the Morley nonconforming FEM has been done in [17] and for the Adini ncFEM, $E_{\mathcal{D}}$ has been studied in [14]. In both cases, by recalling the coercivity property (6.2.4) of B, for $v_{\mathcal{D}} \in X_{\mathcal{D},0}$, the companion operator $E_{\mathcal{D}}$ satisfies

$$\|\Pi_{\mathcal{D}} v_{\mathcal{D}} - E_{\mathcal{D}} v_{\mathcal{D}}\| \lesssim h^2 \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\|, \quad \|\mathcal{H} E_{\mathcal{D}} v_{\mathcal{D}}\| \lesssim \|\mathcal{H}^B_{\mathcal{D}} v_{\mathcal{D}}\|.$$

Note that the range of $E_{\mathcal{D}}$ is made of piecewise polynomial functions. An introduction of $E_{\mathcal{D}}v_{\mathcal{D}}$, a use of the triangle inequality, the inverse estimate, the above estimate and the continuous Sobolev embedding $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$ lead to

$$\begin{aligned}\|\Pi_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)} &\leq \|\Pi_{\mathcal{D}}v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)} + \|E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)} \\ &\leq \sum_{K\in\mathcal{M}} h_K^{-1}\|\Pi_{\mathcal{D}}v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^2(K)} + \|E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)} \\ &\lesssim \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\| + \|\mathcal{H}E_{\mathcal{D}}v_{\mathcal{D}}\| \lesssim \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|.\end{aligned}$$

Thus, assumption **(A3)** is satisfied by the Adini and Morley ncFEMs.

Gradient Recovery Method

The superconvergence assumption **(A1)** i) is proved in Proposition 2.5.4 with $w_{\mathcal{M}} = w$. Since $w_{\mathcal{M}} = w$, both the estimates in **(A1)** ii) are trivial. Apply Taylor expansion, the triangle inequality, the Poincaré inequality and (2.4.28) to obtain the estimate in **(A2)** as

$$\begin{aligned}\|\Pi_{\mathcal{D}}v_{\mathcal{D}} - \boldsymbol{\mathcal{P}}_{\mathcal{M}}(\Pi_{\mathcal{D}}v_{\mathcal{D}})\| &\leq h\|\nabla v_{\mathcal{D}}\| \leq h\|\nabla v_{\mathcal{D}} - Q_h\nabla v_{\mathcal{D}}\| + h\|Q_h\nabla v_{\mathcal{D}}\| \\ &\leq h\|\nabla v_{\mathcal{D}} - Q_h\nabla v_{\mathcal{D}}\| + h\mathrm{diam}(\Omega)\|\nabla(Q_h\nabla v_{\mathcal{D}})\| \\ &\leq hC_B^{-1}\sqrt{2}\max(1,\mathrm{diam}(\Omega))\|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\| \lesssim h\|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|.\end{aligned}$$

To check **(A3)** for the gradient recovery method, let X_h be the Hsieh-Clough-Toucher conforming macro finite element (see Section 2.3.1) and construct a companion operator $E_{\mathcal{D}} : X_{\mathcal{D},0} \to X_h$ as follows: Recall that the local degrees of freedom of HCT on triangle $K \in \mathcal{M}$ are the function values and first partial derivatives at the three vertices of K in addition to the normal derivative at the midpoints of the edges of K. For the GR method, $X_{\mathcal{D},0} = V_h$, the conforming $\mathbb{P}_1$ finite element space and $Q_h : L^2(\Omega) \to V_h$. Let the set of vertices of $\mathcal{M}$ be denoted by $\mathcal{V}$ and $\overline{\boldsymbol{x}}_{\boldsymbol{\sigma}}$ be the midpoint of the edge $\boldsymbol{\sigma}$. Define $E_{\mathcal{D}}v_{\mathcal{D}} \in X_h$ by setting the degrees of freedom as follows:

$$\forall p \in \mathcal{V}, \quad E_{\mathcal{D}}v_{\mathcal{D}}(p) = v_{\mathcal{D}}(p) \tag{6.3.13}$$

$$\forall p \in \mathcal{V}, \quad \nabla E_{\mathcal{D}}v_{\mathcal{D}}(p) = Q_h\nabla v_{\mathcal{D}}(p) \tag{6.3.14}$$

$$\forall \boldsymbol{\sigma} \in \mathcal{F}, \quad (\nabla E_{\mathcal{D}}v_{\mathcal{D}}\cdot n_{\boldsymbol{\sigma}})(\overline{\boldsymbol{x}}_{\boldsymbol{\sigma}}) = (Q_h\nabla v_{\mathcal{D}}\cdot n_{\boldsymbol{\sigma}})(\overline{\boldsymbol{x}}_{\boldsymbol{\sigma}}). \tag{6.3.15}$$

Let $K \in \mathcal{M}$ and w be a polynomial function on K. Using the scaling argument [66], we have

$$\|w\|_{L^2(K)}^2 \approx \sum_{N\in\mathcal{N}(K)} (\mathrm{diam}(K))^{2(1+\mathfrak{O}(N))}(N(w))^2,$$

where $\mathcal{N}(K)$ is the set of degrees of freedom and $\mathfrak{O}(N)$ is the order of differentiation in the degrees of freedom. Here, (6.3.13) is of order 0 and (6.3.14) and (6.3.15) are of order 1. Since $v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}} \in \mathbb{P}_3$ on a submesh and $N(v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}) = 0$ if N is of type (6.3.13),

$$\|v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^2(K)}^2 \approx \sum_{N\in\mathcal{N}(K)} h_K^4(N(v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}))^2.$$

This and the definition of $E_{\mathcal{D}}$ imply that

$$h_K^{-4}\|v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^2(K)}^2 \approx \sum_{p\in\mathcal{V}_K} |(\nabla v_{\mathcal{D}} - Q_h\nabla v_{\mathcal{D}})(p)|^2 + \sum_{\sigma\in\mathcal{F}_K} |((\nabla v_{\mathcal{D}} - Q_h\nabla v_{\mathcal{D}})\cdot n_\sigma)(\overline{\boldsymbol{x}}_\sigma)|^2,$$

where $\mathcal{V}_K$ is the set of vertices associated with K. A use of the above estimate, an inverse estimate and (2.4.28) leads to

$$h_K^{-4}\|v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^2(K)}^2 \lesssim \|\nabla v_{\mathcal{D}} - Q_h\nabla v_{\mathcal{D}}\|_{L^\infty(K)^2}^2 \lesssim h_K^{-2}\|\nabla v_{\mathcal{D}} - Q_h\nabla v_{\mathcal{D}}\|_{L^2(K)^2}^2 \lesssim h_K^{-2}\|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|^2.$$

Therefore,

$$\|v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^2(K)} \lesssim h_K\|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|. \tag{6.3.16}$$

A use of the triangle inequality, inverse estimate, (6.3.16) and the continuous Sobolev embedding $H^2(\Omega)\hookrightarrow L^\infty(\Omega)$ leads to

$$\begin{aligned}\|\Pi_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)} = \|v_{\mathcal{D}}\|_{L^\infty(\Omega)} &\le \|v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)} + \|E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)}\\ &\le \sum_{K\in\mathcal{M}} h_K^{-1}\|v_{\mathcal{D}} - E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^2(K)} + \|E_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)}\\ &\lesssim \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\| + \|\mathcal{H}E_{\mathcal{D}}v_{\mathcal{D}}\|. \end{aligned}\tag{6.3.17}$$

Introduce $\nabla Q_h\nabla v_{\mathcal{D}}$, use triangle inequality, inverse estimate [45, Lemma 1.44], the definition of $E_{\mathcal{D}}$ and (2.4.28) to obtain

$$\begin{aligned}\|\mathcal{H}E_{\mathcal{D}}v_{\mathcal{D}}\| &\le \|\nabla\nabla E_{\mathcal{D}}v_{\mathcal{D}} - \nabla Q_h\nabla v_{\mathcal{D}}\| + \|\nabla Q_h\nabla v_{\mathcal{D}}\|\\ &\le h^{-1}\|\nabla E_{\mathcal{D}}v_{\mathcal{D}} - Q_h\nabla v_{\mathcal{D}}\| + \|\nabla Q_h\nabla v_{\mathcal{D}}\| \lesssim \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|.\end{aligned}$$

A substitution of the above estimate in (6.3.17) yields

$$\|\Pi_{\mathcal{D}}v_{\mathcal{D}}\|_{L^\infty(\Omega)} \lesssim \|\mathcal{H}_{\mathcal{D}}^B v_{\mathcal{D}}\|.$$

Thus, assumption **(A3)** follows for the gradient recovery methods.

The notation $X\lesssim_\eta Y$ means that $X\le CY$ for some C depending only on Ω, B, ρ, an upper bound of $C_{\mathcal{D}}$, and η.
The following theorem states the main super-convergence result for post-processed controls. The proof is obtained by modifying the proof of Theorem 4.3.6 for GDM to HDM by adapting the assumptions **(A1)**-**(A4)** discussed above and hence is skipped.

Theorem 6.3.5 (Super-convergence for post-processed controls)**.** *Let $\mathcal{D}$ be a HD and $\mathcal{M}$ be a mesh. Assume that*

- *$\mathcal{U}_{\rm ad}$ and $\mathcal{U}_h$ are given by* (6.3.7) *and* (6.3.8)*,*
- **(A1)**–**(A4)** *hold,*

- $\overline{y}$ *and* $\overline{p}$ *belong to* $H^4(\Omega)$,
- $\overline{u}_d$ *belongs to* $H^2(\Omega)$,

and let $\widetilde{u}$, $\widetilde{u}_h$ *be the post-processed controls defined by* (6.3.12). *Then there exists C depending only on* α *such that*

$$\|\widetilde{u}-\widetilde{u}_h\| \lesssim_\eta Ch^2\left(\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)}+\mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p})\right), \tag{6.3.18}$$

where $\mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p})$ *is defined in Theorem 4.3.6, that is,*

$$\mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p}) = \operatorname{minmod}(a,b)+\|\overline{y}_d\|+\|\overline{u}_d\|_{H^2(\Omega)}+\|f\|+\|\overline{y}\|_{H^4(\Omega)}+\|\overline{p}\|_{H^4(\Omega)}$$

with $\operatorname{minmod}(a,b)=0$ *if* $ab\le 0$ *and* $\operatorname{minmod}(a,b)=\min(|a|,|b|)$ *otherwise.*

The super-convergence of the state and adjoint variables is stated below. The proof is similar to that of Corollary 4.3.9.

Corollary 6.3.6 (Super-convergence for the state and adjoint variables)**.** *Let* $(\overline{y},\overline{p})$ *and* $(\overline{y}_\mathcal{D},\overline{p}_\mathcal{D})$ *be the solutions to* (6.2.5a)–(6.2.5b) *and* (6.2.7a)–(6.2.7b). *Under the assumptions of Theorem 6.3.5, the following error estimates hold, with C depending only on* α*:*

$$\|\overline{y}_\mathcal{M}-\Pi_\mathcal{D}\overline{y}_\mathcal{D}\| \lesssim_\eta Ch^2\left(\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)}+\mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p})\right), \tag{6.3.19}$$

$$\|\overline{p}_\mathcal{M}-\Pi_\mathcal{D}\overline{p}_\mathcal{D}\| \lesssim_\eta Ch^2\left(\|\overline{u}\|_{W^{1,\infty}(\mathcal{M}_1)}+\mathcal{F}(a,b,\overline{y}_d,\overline{u}_d,f,\overline{y},\overline{p})\right), \tag{6.3.20}$$

where $\overline{y}_\mathcal{M}$ *and* $\overline{p}_\mathcal{M}$ *are defined as in* **(A1)**.

The linear control problem considered in this Chapter helps us to extend our analysis to control problem governed by non-linear elliptic equations, which is a plan of future work.

6.4 Numerical results

In this section, the numerical results to support the theoretical estimates obtained in the previous sections are presented. Two specific schemes are used for the state and adjoint variables: gradient recovery method and finite volume method presented in Chapter 2. The control variable is discretised using piecewise constant functions. The discrete solution is computed by using the primal-dual active set algorithm, see [109, Section 2.12.4]. Let the relative errors be denoted by

$$\mathsf{err}_\mathcal{D}(\overline{y}) := \frac{\|\Pi_\mathcal{D}\overline{y}_\mathcal{D}-\overline{y}\|}{\|\overline{y}\|}, \quad \mathsf{err}_\mathcal{D}(\nabla\overline{y}) := \frac{\|\nabla_\mathcal{D}\overline{y}_\mathcal{D}-\nabla\overline{y}\|}{\|\nabla\overline{y}\|}, \quad \mathsf{err}_\mathcal{D}(\mathcal{H}\overline{y}) := \frac{\|\mathcal{H}^B_\mathcal{D}\overline{y}_\mathcal{D}-\mathcal{H}\overline{y}\|}{\|\mathcal{H}\overline{y}\|}$$

$$\mathsf{err}_\mathcal{D}(\overline{p}) := \frac{\|\Pi_\mathcal{D}\overline{p}_\mathcal{D}-\overline{p}\|}{\|\overline{p}\|}, \quad \mathsf{err}_\mathcal{D}(\nabla\overline{p}) := \frac{\|\nabla_\mathcal{D}\overline{p}_\mathcal{D}-\nabla\overline{p}\|}{\|\nabla\overline{p}\|}, \quad \mathsf{err}_\mathcal{D}(\mathcal{H}\overline{p}) := \frac{\|\mathcal{H}^B_\mathcal{D}\overline{p}_\mathcal{D}-\mathcal{H}\overline{p}\|}{\|\mathcal{H}\overline{p}\|}$$

$$\mathsf{err}(\overline{u}) := \frac{\|\overline{u}_h - \overline{u}\|}{\|\overline{u}\|} \quad \text{and} \quad \mathsf{err}(\widetilde{u}) := \frac{\|\widetilde{u}_h - \widetilde{u}\|}{\|\overline{u}\|}.$$

Here, the definitions of $\widetilde{u}$ and $\widetilde{u}_h$ follow from (6.3.12) and $\overline{u}$ is given by (6.2.6).

$$\overline{u}_h(\boldsymbol{x}) = P_{[a,b]}\left(\mathcal{P}_{\mathcal{M}}\left(\overline{u}_d(\boldsymbol{x}) - \frac{\mathcal{C}}{\alpha}\Pi_{\mathcal{D}}\overline{p}_{\mathcal{D}}\right)\right).$$

The model problem is constructed in such a way that the exact solution is known. In the experiment, the computational domain Ω is taken to be the unit square $(0,1)^2$. The data in the optimal distributed control problem are chosen as follows:

$$\overline{y} = \sin^2(\pi x)\sin^2(\pi y), \quad \overline{p} = \sin^2(\pi x)\sin^2(\pi y),$$
$$\overline{u}_d = 0, \quad \alpha = 10^{-3}, \quad \mathcal{U}_{\text{ad}} = [-750, -50], \quad \overline{u} = P_{[-750,-50]}\left(-\frac{1}{\alpha}\overline{p}\right).$$

The source term f and the desired state $\overline{y}_d$ are the computed using

$$f = \Delta^2\overline{y} - \overline{u}, \quad \overline{y}_d = \overline{y} - \Delta^2\overline{p}.$$

6.4.1 Gradient Recovery Method

Here, $X_{\mathcal{D},0}$ is the conforming $\mathbb{P}_1$ finite element space, and the implementation was done following the ideas in [82]. The stabilisation factor τ is chosen to be 1, see Section 2.7.1 for more details. The error estimates and the convergence rates of the control, the post-processed control, the state and the adjoint variables are presented in Table 6.1. As seen in the table, we obtain linear order of convergence for the state and adjoint variable in the energy norm, quadratic order of convergence for state and adjoint variables in L^2 and H^1 norm, linear order of convergence for the control variable in L^2 norm, and a quadratic order of convergence for the post-processed control. They follow the expected theoretical rates given in Theorem 6.3.2, Proposition 6.3.3, Remark 6.3.4, Theorem 6.3.5 and Corollary 6.3.6.

6.4.2 Finite Volume Method

In this method, the schemes were first tested on a series of regular triangular meshes (`mesh1` family) and then on square meshes (`mesh2` family), both taken from [74]. As mentioned in Chapter 2, to ensure the correct orthogonality property, the point $\boldsymbol{x}_K \in K$ is chosen as the circumcenter of K if K is a triangle, or the center of mass of K if K is a rectangle. Denote the relative H^2 error by

$$\mathsf{err}_{\mathcal{D}}(\Delta\overline{y}) := \frac{\|\Delta_{\mathcal{D}}\overline{y}_{\mathcal{D}} - \Delta\overline{y}\|}{\|\Delta\overline{y}\|}, \quad \mathsf{err}_{\mathcal{D}}(\Delta\overline{p}) := \frac{\|\Delta_{\mathcal{D}}\overline{p}_{\mathcal{D}} - \Delta\overline{p}\|}{\|\Delta\overline{p}\|}.$$

The errors of the numerical approximations to state, adjoint and control variables on uniform meshes are shown in Tables 6.2-6.3. In the case of triangular meshes, slightly better quadratic order of convergence for the state and adjoint variables in L^2 norm, linear order of convergence

Table 6.1: (GR) Convergence results for the relative errors

h	$\mathsf{err}_{\mathcal{D}}(\bar{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\bar{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\bar{y})$	Order	$\mathsf{err}(\bar{u})$	Order
0.353553	2.192387	-	0.692406	-	0.817825	-	0.537029	-
0.176777	0.131323	4.0613	0.079054	3.1307	0.245715	1.7348	0.190741	1.4934
0.088388	0.032735	2.0042	0.019531	2.0171	0.116596	1.0755	0.081011	1.2354
0.044194	0.008220	1.9936	0.004757	2.0376	0.057374	1.0230	0.038235	1.0832
0.022097	0.002081	1.9821	0.001215	1.9695	0.028479	1.0105	0.018865	1.0192

h	$\mathsf{err}_{\mathcal{D}}(\bar{p})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\bar{p})$	Order	$\mathsf{err}_{\mathcal{D}}(\mathcal{H}\bar{p})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.353553	3.132234	-	0.721611	-	0.855785	-	0.593791	-
0.176777	0.145384	4.4293	0.099972	2.8516	0.246647	1.7948	0.126971	2.2255
0.088388	0.036226	2.0048	0.023097	2.1138	0.116471	1.0825	0.032031	1.9870
0.044194	0.009068	1.9982	0.005552	2.0567	0.057308	1.0231	0.007716	2.0536
0.022097	0.002261	2.0037	0.001363	2.0266	0.028470	1.0093	0.001874	2.0416

Table 6.2: (FV) Convergence results for the relative errors, triangular mesh

h	$\mathsf{err}_{\mathcal{D}}(\bar{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\bar{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\bar{y})$	Order	$\mathsf{err}(\bar{u})$	Order
0.250000	0.200670	-	0.298405	-	0.165136	-	0.245085	-
0.125000	0.021019	3.2551	0.135346	1.1406	0.057870	1.5128	0.116630	1.0713
0.062500	0.005108	2.0409	0.066054	1.0349	0.030285	0.9342	0.057540	1.0193
0.031250	0.001178	2.1169	0.032808	1.0096	0.016785	0.8514	0.028819	0.9976
0.015625	0.000265	2.1513	0.016374	1.0026	0.009900	0.7617	0.014408	1.0001

h	$\mathsf{err}_{\mathcal{D}}(\bar{p})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\bar{p})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\bar{p})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.250000	0.230914	-	0.316994	-	0.189286	-	0.094040	-
0.125000	0.032775	2.8167	0.136993	1.2104	0.061257	1.6276	0.024703	1.9286
0.062500	0.007282	2.1703	0.066202	1.0492	0.030607	1.0010	0.004857	2.3465
0.031250	0.001693	2.1049	0.032824	1.0121	0.016820	0.8637	0.001200	2.0174
0.015625	0.000380	2.1557	0.016376	1.0032	0.009904	0.7641	0.000260	2.2042

in H^1 norm and sublinear in H^2 norm are obtained. The control converges at the optimal rate of h, whereas the post processed control converges with quadratic rate, which is a superconvergence result.

For the square meshes, we obtain quadratic rate of convergence in L^2, H^1 and H^2 norms for the state and adjoint variables. The superconvergence in H^2 norm is not entirely surprising, since rectangular meshes are extremely regular and symmetric. Without post-processing, an $\mathcal{O}(h)$ convergence rate is obtained on the controls and post-processing step leads to quadratic order of convergence. The numerical results are better than the theoretical rates stated in Theorem 6.3.2, Proposition 6.3.3 and Remark 6.3.4. Also, using a post-processing step, an improved error estimate for the control variable is obtained numerically.

Table 6.3: (FV) Convergence results for the relative errors, square mesh

h	$\mathsf{err}_{\mathcal{D}}(\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{y})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\overline{y})$	Order	$\mathsf{err}(\overline{u})$	Order
0.353553	0.288994	-	0.196325	-	0.270192	-	0.398184	-
0.176777	0.060061	2.2665	0.045562	2.1073	0.056607	2.2549	0.187167	1.0891
0.088388	0.015072	1.9946	0.010322	2.1420	0.014538	1.9612	0.092209	1.0213
0.044194	0.003700	2.0263	0.002590	1.9945	0.003551	2.0334	0.045989	1.0036
0.022097	0.000927	1.9968	0.000642	2.0125	0.000891	1.9941	0.022985	1.0006

h	$\mathsf{err}_{\mathcal{D}}(\overline{p})$	Order	$\mathsf{err}_{\mathcal{D}}(\nabla\overline{p})$	Order	$\mathsf{err}_{\mathcal{D}}(\Delta\overline{p})$	Order	$\mathsf{err}(\widetilde{u})$	Order
0.353553	0.300063	-	0.189326	-	0.285835	-	0.144022	-
0.176777	0.066723	2.1690	0.039945	2.2448	0.065909	2.1166	0.035315	2.0280
0.088388	0.016237	2.0389	0.009482	2.0747	0.016092	2.0342	0.009726	1.8604
0.044194	0.004033	2.0093	0.002337	2.0208	0.003998	2.0091	0.002471	1.9766
0.022097	0.001007	2.0023	0.000582	2.0054	0.000998	2.0023	0.000589	2.0693

Chapter 7

Summary and Future Work

This concluding chapter of the dissertation highlights the main contributions of the present work. Further, it discusses the possible extensions and the scope for future problems.

7.1 Summary

The first part of the thesis considers the Hessian discretisation method for fourth order elliptic partial differential equations. The idea of the HDM is to construct a scheme by replacing the continuous space and operators in the weak formulation by discrete ones given in a HD. In Chapter 2, the HDM for linear elliptic equations is proposed and analyzed. It is shown that some known classical conforming and non-conforming FEMs, a novel scheme based on the $\mathbb{P}_1$ finite element space and a gradient recovery (GR) designed using biorthogonal systems, and the finite volume method in [59] fit into the framework of HDM. A generic error estimate and improved error estimates are proved in the HDM framework using the three core properties of HD (namely, coercivity, consistency and limit-conformity). Since an improved L^2 estimate is not expected in general for finite volume method (FVM), following the ideas in [54] for GDM, a modified FVM is considered by changing the quadrature of the source term and a superconvergence result is proved for this modified FVM. A generic notion of companion operator is defined in the HDM setting. The existence of such an operator is an essential tool to establish improved H^1-like error estimate. Companion operators are known to exist for some non-conforming FEMs, but to this day the existence of one for the gradient recovery method is unknown. The chapter concludes with numerical tests, illustrated for the GR method and the FVM, that confirm the theoretical convergence result for the GR method. For the FVM, the tests show a better convergence rate than the one given by theory. These tests also show that the FVM does not display an improved convergence rate in H^1-like norm compared to that in the energy norm, hinting at the non-existence of a companion operator for this method. On the contrary, improved rates in H^1-like norm are observed for the gradient recovery method, which leaves open the possibility of existence of a companion operator for this method.

Chapter 3 discusses the HDM for fourth order semilinear elliptic equations with trilinear nonlin-

earity in an abstract formulation. The stream function vorticity formulation of the incompressible 2D Navier-Stokes equations and von Karmán equations can be written under this abstract form. To deal with the non-linearity in the model, the three basic properties of the HDM must be slightly adjusted (to yield stronger integrability properties, and measure the limit-conformity between the reconstructed gradient and function). Convergence analysis is proved in two different ways: by compactness techniques (which requires the introduction of a compactness property for sequences of HDs), and by error estimates. The compactness technique does not provide any order of convergence, but convergence is obtained without assuming any regularity of the solution. Conforming FEMs, nonconforming FEMs (Adini rectangle and Morley triangle) and the method based on gradient recovery operators are shown to be some examples of HDM in the nonlinear case. Numerical experiments are performed for the gradient recovery method and the Morley nonconforming FEM using Newton's method. Though the theoretical convergence result for the GR method is proved only by compactness techniques, the numerical results illustrate expected convergence rates.

Chapter 4 deals with optimal control problems governed by second order diffusion equations with Dirichlet boundary conditions and Neumann boundary conditions with reaction term. For these models, the relevant analysis framework is that of the gradient discretisation method (GDM), leading to gradient schemes. A gradient scheme is defined for the optimal control problem by discretising the corresponding optimality system, which involves state, adjoint and control variables. A generic error estimate is established in the GDM framework. Under a few additional assumptions, following the ideas developed in [96], super-convergence results for all three variables are derived in a post-processing step. Note that superconvergence results are twofolded. Under generic assumptions on the GD, which allow for local mesh refinements, an $\mathcal{O}(h^{2-\varepsilon})$ super-convergence is proved with $\varepsilon > 0$ in dimension 2 and $\varepsilon = 1/6$ in dimension 3. Under an L^∞-bound assumption on the solution to the GS, which for most methods requires the quasi-uniformity of the meshes, a full $\mathcal{O}(h^2)$ convergence result is proved. Two particular cases of the main results are considered, non-conforming finite element methods and mixed-hybrid mimetic finite difference schemes, in this chapter. Results of numerical experiments are demonstrated for the conforming, non-conforming and mixed-hybrid mimetic finite difference schemes.

In Chapter 5, basic error estimates are established that provide $\mathcal{O}(h)$ convergence rate for all the three variables (control, state and adjoint) for low order schemes under standard regularity assumptions for the pure Neumann problem. Contrary to the setting covered in Chapter 4, and in most of the literature, the Neumann problems considered here do not include any reaction term. As a consequence, they are only well-posed if the solution's average is fixed (say to zero). This equation appears as an additional constraint on the control problem. Also, super-convergence result is proved for post-processed optimal controls, state and adjoint variables. A projection relation is established between control and adjoint variables. This relation, which is non-standard since it has to account for the zero average constraints, is the key to prove the super-convergence result for all three variables. A modified active set strategy algorithm for GDM that is adapted to this non-standard projection relation is designed. The first super-convergence result provides a nearly quadratic convergence rate for a post-processed control. Under an L^∞ stability assumption of the GDM, the second super-convergence theorem establishes a full quadratic super-

convergence rate. Finally, numerical results that confirm the theoretical rates of convergence for conforming, nonconforming finite element methods and mimetic finite difference methods are performed.

Chapter 6 studies the optimal control problem governed by fourth order linear elliptic equations using the HDM framework. The basic error estimates for control, state and adjoint variables are proved by following the ideas developed in Chapter 4 for GDM for second order problems. This estimate yields $\mathcal{O}(h)$ convergence rate for the conforming FEMs, the Adini and Morley ncFEMs, and the gradient recovery methods. With a post-processing step, an improved error estimate of order $\mathcal{O}(h^2)$ is obtained. This superconvergence result is established under an L^2 superconvergence assumption on the elliptic PDEs and a few assumptions. These assumptions are verified for conforming FEMs, ncFEMs and GR methods. Several numerical experiments are illustrated for the GR method and the FVM. For the finite volume method, superconvergence result is numerically observed.

7.2 Future Work

The results of this thesis could be extended in the following directions:

- The HDM for fourth order elliptic equations covers the conforming FEMs, Adini and Morley nonconforming FEMs and a method based on gradient recovery operator. For the linear models, finite volume method based on Δ-adapted discretizations is also an example of HDM. Future work will be efforts to establish that other numerical methods can be viewed in the HDM framework. The FVM on admissible meshes has been considered in the HDM framework. We could look into some other methods for more generic meshes, for example, an Hybrid mimetic mixed (HMM) scheme or a Vertex approximated gradient (VAG) method.

- The HDM also gives the tools to design new methods. If a scheme satisfies three (resp. four) core properties of HD for linear (resp. non-linear) problems, then the scheme is convergent. Thus, any scheme entering the HDM framework is known to converge. Each choice of HDs corresponds to a particular scheme. It will be interesting to develop new methods such that the reconstructions satisfy the properties of the HDM.

- The companion operator $E_{\mathcal{D}}$ defined by **(A5)** in Section 3.5.2 enables us to prove, in the HDM framework, the improved H^1-like error estimates for linear equations (Theorem 2.6.2) and also to prove the convergence analysis by error estimates for non-linear problems (Theorem 3.5.12). The construction of a companion operator $E_{\mathcal{D}}$ for the method based on gradient recovery operators, with proper features (see Remark 3.5.7), is an interesting avenue to explore as it would show that the aforementioned results are satisfied by the gradient recovery method.

- For the pure Neumann control problem without reaction term and hence with zero average constraint, a modified active set algorithm is designed in Section 5.5.1. As seen in

the numerical experiments provided in Section 5.5.2, this modified algorithm converges numerically. The theoretical convergence analysis of this proposed algorithm is a plan for future study.

- In Chapter 6, we have considered the control problems governed by fourth order linear elliptic equation in the abstract framework of HDM. If the state equation is a semi-linear elliptic equation, then the same analysis cannot be applied. Chapter 3 discusses the HDM for semi-linear fourth order problems in an abstract setting. The analysis can be extended to the control problem governed by the semi-linear elliptic equations with trilinear nonlinearity in the HDM framework. To make more precisely, recall the abstract weak formulation considered in Chapter 3:

  Given $k \geq 1$, the continuous abstract problem seeks $\Psi \in \boldsymbol{X} := H_0^2(\Omega)^k$ such that

  $$\mathcal{A}(\mathcal{H}\Psi, \mathcal{H}\Phi) + \mathcal{B}(\mathcal{H}\Psi, \nabla\Psi, \nabla\Phi) = \mathcal{L}(\Phi) \quad \forall \Phi \in \boldsymbol{X}.$$

  This model problem has application to the stream function vorticity formulation of 2D Navier–Stokes equation for $k = 1$ and the von Kármán equations for $k = 2$. Consider the optimal control problem governed by the weak formulation defined by

  $$\min_{u \in U_{ad}} J(u) \text{ subject to}$$
  $$\mathcal{A}(\mathcal{H}\Psi(u), \mathcal{H}\Phi) + \mathcal{B}(\mathcal{H}\Psi(u), \nabla\Psi(u), \nabla\Phi) = \mathcal{L}(\Phi) + (\mathcal{C}u, \phi_1) \quad \forall \Phi \in \boldsymbol{X}.$$

  The operator $\mathcal{C} \in \mathcal{L}(L^2(\omega), L^2(\Omega))$ is the extension operator defined by $\mathcal{C}u(x) = u(x)\chi_\omega(x)$, where χ_ω is the characteristic function of $\omega \subset L^2(\Omega)$, u is the control variable and $\Psi(u)$ is the state variable associated with the control u. Here

  $$J(u) := \frac{1}{2}\|\Psi(u) - \Psi_d\|^2 + \frac{\alpha}{2}\|u - u_d\|^2_{L^2(\omega)},$$

  is the cost functional, $\alpha > 0$ is a fixed regularization parameter, Ψ_d is the desired state variable for $\Psi(u)$, $u_d \in L^2(\omega)$ is the desired control variable and $U_{ad} \subset L^2(\omega)$ is a non-empty, convex and bounded admissible space of controls.

  Due to the presence of non-linearity in the weak formulation of the governing partial differential equations, the problem becomes non-convex and hence the solution is not unique, which leads to additional technicalities in the analysis.

Appendix

A.1 Technical results

Lemma A.1.1 (Poincaré inequality along an edge in L^2 norm)**.** *Let σ be an edge of a polygonal cell, $w \in H^1(\sigma)$ and assume that w vanishes at a point on the edge $\sigma \in \mathcal{F}$. Then*

$$\|w\|_{L^2(\sigma)} \leq h_\sigma \|\partial w\|_{L^2(\sigma)},$$

where ∂ denotes the derivative along the edge and h_σ is the length of the edge.

Proof. Let m denote the point on the edge σ which satisfies $w(m) = 0$. For $m < \boldsymbol{x}$, we obtain

$$w(\boldsymbol{x}) = w(m) + \int_m^{\boldsymbol{x}} \partial w(y)\,\mathrm{d}y = \int_m^{\boldsymbol{x}} \partial w(y)\,\mathrm{d}y.$$

A use of the Cauchy–Schwarz inequality yields

$$|w(\boldsymbol{x})| \leq |\boldsymbol{x} - m|^{1/2} \left(\int_m^{\boldsymbol{x}} |\nabla w|^2 \,\mathrm{d}y \right)^{1/2} \leq \sqrt{h_\sigma} \left(\int_\sigma |\partial w|^2 \,\mathrm{d}y \right)^{1/2}.$$

Squaring this yields $|w(\boldsymbol{x})|^2 \leq h_\sigma \int_\sigma |\partial w|^2 \,\mathrm{d}y$ and integrating over the edge concludes the proof. □

Lemma A.1.2 (Integration by parts)**.** *Let P be a fourth order tensor. For $\xi \in H^2(\Omega)^{d\times d}$ and $\phi \in H^1(\Omega)$, we have*

$$\int_\Omega (\mathcal{H} : P\xi)\phi = -\int_\Omega \nabla\phi \cdot \operatorname{div}(P\xi) + \int_{\partial\Omega} \operatorname{div}(P\xi \cdot n)\phi.$$

For $\psi \in H^2(\Omega)$,

$$\int_\Omega P\xi : \mathcal{H}\psi = -\int_\Omega \nabla\psi \cdot \operatorname{div}(P\xi) + \int_{\partial\Omega} (\operatorname{div}(P\xi n)) \cdot \nabla\psi.$$

For $\zeta \in H^1(\Omega)^d$,

$$\int_\Omega P\xi : \nabla\zeta = -\int_\Omega \operatorname{div}(P\xi) \cdot \zeta + \int_{\partial\Omega} (\operatorname{div}(P\xi n)) \cdot \zeta.$$

Lemma A.1.3. *Let $w \in P_k(\mathcal{M})$. If for all $\sigma \in \mathcal{F}$ there exists $x_\sigma \in \sigma$ such that $[\![w]\!](x_\sigma) = 0$, then*

$$\|w\| \leq C\|\nabla_{\mathcal{M}} w\|,$$

where $C > 0$ depends only on Ω, k and mesh regularity parameter η.

Proof. Consider the $\|\cdot\|_{dG,\mathcal{M}}$ norm defined by (2.4.13): For all $w \in H^1(\mathcal{M})$,

$$\|w\|_{dG,\mathcal{M}}^2 := \|\nabla_{\mathcal{M}} w\|^2 + \sum_{\sigma\in\mathcal{F}} \frac{1}{h_\sigma} \|[\![w]\!]\|_{L^2(\sigma)}^2. \tag{A.1.1}$$

Since $[\![w]\!](x_\sigma) = 0$ for all $\sigma \in \mathcal{F}$, a use of the Poincaré inequality along an edge in L^2 norm given by Lemma A.1.1 leads to

$$\|[\![w]\!]\|_{L^2(\sigma)} \leq h_\sigma \|\nabla_{\mathcal{M}} [\![w]\!]\|_{L^2(\sigma)^d} \leq h_\sigma \Big(\|\nabla_{\mathcal{M}} w_{|K}\|_{L^2(\sigma)^d} + \|\nabla_{\mathcal{M}} w_{|L}\|_{L^2(\sigma)^d} \Big). \tag{A.1.2}$$

Let $\sigma \in \mathcal{F}_{\rm int}$ be such that $\mathcal{M}_\sigma = \{K, L\}$. Use (A.1.2) and the trace inequality (see [45, Lemma 1.46]) to obtain

$$\begin{aligned}
\|[\![w]\!]\|_{L^2(\sigma)} &\leq h_\sigma \Big(\|\nabla_{\mathcal{M}} w_{|K}\|_{L^2(\sigma)^d} + \|\nabla_{\mathcal{M}} w_{|L}\|_{L^2(\sigma)^d} \Big) \\
&\leq C_{\rm tr} h_\sigma \Big(h_K^{-1/2} \|\nabla_{\mathcal{M}} w\|_{L^2(K)^d} + h_L^{-1/2} \|\nabla_{\mathcal{M}} w\|_{L^2(L)^d} \Big),
\end{aligned} \tag{A.1.3}$$

where $C_{\rm tr}$ depends only on k and mesh regularity parameter η. A substitution of (A.1.3) in (A.1.1) leads to

$$\begin{aligned}
\|w\|_{dG,\mathcal{M}}^2 &\leq \|\nabla_{\mathcal{M}} w\|^2 + 2 \sum_{\sigma\in\mathcal{F}} C h_\sigma \Big(h_K^{-1} \|\nabla_{\mathcal{M}} v\|_{L^2(K)^d}^2 + h_L^{-1} \|\nabla_{\mathcal{M}} v\|_{L^2(L)^d}^2 \Big) \\
&\leq \|\nabla_{\mathcal{M}} w\|^2 + C \sum_{K\in\mathcal{M}} \sum_{\sigma\in\mathcal{F}_K} \|\nabla_{\mathcal{M}} v\|_{L^2(K)^d}^2 \\
&\leq \|\nabla_{\mathcal{M}} w\|^2 + 3C \sum_{K\in\mathcal{M}} \|\nabla_{\mathcal{M}} v\|_{L^2(K)^d}^2 \\
&\leq C \|\nabla_{\mathcal{M}} w\|^2,
\end{aligned} \tag{A.1.4}$$

where $C > 0$ depends only on Ω, k and η. Use the fact that $\|w\| \leq C\|w\|_{dG,\mathcal{M}}$ (see [45, Theorem 5.3]) to deduce $\|w\| \leq C\|\nabla_{\mathcal{M}} w\|$. □

Lemma A.1.4 (Weak-strong convergence)**.** *Let $1 \leq p, q, r \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$. If $f_n \to f$ strongly in $L^p(\Omega)^d$, $g_n \to g$ strongly in $L^q(\Omega)^d$ and $h_n \to h$ weakly in $L^r(\Omega)^d$, then*

$$\int_\Omega f_n g_n h_n \,\mathrm{d}\boldsymbol{x} \to \int_\Omega fgh \,\mathrm{d}\boldsymbol{x}.$$

Proof. By Banach-Steinhaus theorem, (h_n) is bounded in $L^r(\Omega)^d$ and the boundedness of (f_n) and (g_n) follows from the convergence property. We therefore write, using generalized Hölder's inequality,

$$\begin{aligned}
&\Big| \int_\Omega f_n g_n h_n \,\mathrm{d}\boldsymbol{x} - \int_\Omega fgh \,\mathrm{d}\boldsymbol{x} \Big| \\
&\qquad = \Big| \int_\Omega (f_n - f) g_n h_n \,\mathrm{d}\boldsymbol{x} + \int_\Omega f(g_n - g) h_n \,\mathrm{d}\boldsymbol{x} + \int_\Omega fg(h_n - h) \,\mathrm{d}\boldsymbol{x} \Big| \\
&\qquad \leq \|f_n - f\|_p \|g_n\|_q \|h_n\|_r + \|f\|_p \|g_n - g\|_q \|h_n\|_r + \Big| \int_\Omega fg(h_n - h) \,\mathrm{d}\boldsymbol{x} \Big|.
\end{aligned}$$

The first two terms converge to 0 by strong convergence of $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$, and the last term converges to 0 by weak convergence of $(h_n)_{n\in\mathbb{N}}$ and the fact that $fg \in L^{r'}(\Omega)^d$ since $\frac{1}{r'} = 1 - \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. □

Lemma A.1.5. *Let $\Xi_m \to \Xi$ weakly in $L^2(\Omega;\mathbb{R}^{d\times d})^k$, $\Theta_m \to \Theta$ in $L^4(\Omega;\mathbb{R}^d)^k$ and $X_m \to X$ in $L^4(\Omega;\mathbb{R}^d)^k$ as $m \to \infty$. Then with $\mathcal{B}(\cdot,\cdot,\cdot)$ as in Section 3.2, we have*

$$\mathcal{B}(\Xi_m, \Theta_m, X_m) \to \mathcal{B}(\Xi, \Theta, X) \text{ as } m \to \infty.$$

Proof. We write

$$\begin{aligned}
\mathcal{B}(\Xi_m, \Theta_m, X_m) - \mathcal{B}(\Xi, \Theta, X) = \mathcal{B}(\Xi_m, \Theta_m, X_m - X) + \mathcal{B}(\Xi_m, \Theta_m - \Theta, X) \\
+ \mathcal{B}(\Xi_m - \Xi, \Theta, X).
\end{aligned}$$

Set $l(\Xi_m) = \mathcal{B}(\Xi_m, \Theta, X)$. Since $\mathcal{B}(\cdot,\cdot,\cdot)$ is a trilinear continuous function, $l(\cdot)$ is a linear continuous functional on $L^2(\Omega;\mathbb{R}^{d\times d})^k$. The weak convergence property of $(\Xi_m)_{m\in\mathbb{N}}$ then ensures that $l(\Xi_m) \to l(\Xi)$ as $m \to \infty$. The continuity of $\mathcal{B}(\cdot,\cdot,\cdot)$ yields a constant C_b such that

$$\begin{aligned}
\Big| \mathcal{B}(\Xi_m, \Theta_m, X_m) - \mathcal{B}(\Xi, \Theta, X) \Big| &\leq C_b \|\Xi_m\| \, \|\Theta_m\|_{L^4(\Omega;\mathbb{R}^d)^k} \|X_m - X\|_{L^4(\Omega;\mathbb{R}^d)^k} \\
&\quad + \|\Xi_m\| \, \|\Theta_m - \Theta\|_{L^4(\Omega;\mathbb{R}^d)^k} \|X\|_{L^4(\Omega;\mathbb{R}^d)^k} \\
&\quad + \big| l(\Xi_m) - l(\Xi) \big|.
\end{aligned}$$

Since strongly/weakly convergent sequences are bounded, the convergences of $(\Theta_m)_{m\in\mathbb{N}}$, $(X_m)_{m\in\mathbb{N}}$ and $(l(\Xi_m))_{m\in\mathbb{N}}$ conclude the proof. □

A.2 A generic companion operator

We present here a generic companion operator, that can be constructed using solely the notions of HDs (without referring to the specific considered method), and for which we prove that the quantities in (3.5.17) behave appropriately along sequences of coercive, limit-conforming, consistent and compact HDs. For specific choices of HDs, other $E_{\mathcal{D}}$ can be constructed with, perhaps, more precise estimates on $\delta(E_{\mathcal{D}})$, $\omega(E_{\mathcal{D}})$ and $\Gamma(E_{\mathcal{D}})$, for example see [17] for the Morley element.

Companion operator: For $\psi_{\mathcal{D}} \in X_{\mathcal{D},0}$, the companion function of $\psi_{\mathcal{D}}$, denoted by $E_{\mathcal{D}}\psi_{\mathcal{D}}$, is defined as the solution in $H_0^2(\Omega)$ of:

$$\int_\Omega \mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}} : \mathcal{H}\phi \, \mathrm{d}\boldsymbol{x} = \int_\Omega \mathcal{H}(E_{\mathcal{D}}\psi_{\mathcal{D}}) : \mathcal{H}\phi \, \mathrm{d}\boldsymbol{x}, \quad \forall \phi \in H_0^2(\Omega). \tag{A.2.1}$$

By the Riesz representation Theorem, there exists a unique solution to (A.2.1) and it satisfies

$$\|\mathcal{H}E_{\mathcal{D}}\psi_{\mathcal{D}}\| \leq \|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|. \tag{A.2.2}$$

Lemma A.2.1. *For $\psi_{\mathcal{D}} \in X_{\mathcal{D},0}$ the companion function $E_{\mathcal{D}}\psi_{\mathcal{D}}$ satisfies*

$$\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\| \leq W_{\mathcal{D}}(\mathcal{H}\phi)\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|, \tag{A.2.3}$$

where $\phi \in H_0^2(\Omega)$ is such that, if $\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\| \neq 0$,

$$\Delta^2\phi = \frac{\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}}{\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\|}.$$

Also, for $\Gamma \in H_0^1(\Omega)$ and $\omega \in L^2(\Omega)^d$ such that $\mathrm{div}(\omega) = 0$,

$$\int_\Omega (\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}) \cdot (\nabla\Gamma + \omega) \, \mathrm{d}\boldsymbol{x} \leq \|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|(\widetilde{W}_{\mathcal{D}}(\mathcal{H}\chi) + \widehat{W}_{\mathcal{D}}(\omega)), \tag{A.2.4}$$

where $\chi \in H^2(\Omega) \cap H_0^1(\Omega)$ is such that $\Delta\chi = \Gamma$.

Proof. Consider $\phi \in H_0^2(\Omega)$ such that $\mathcal{H} : \mathcal{H}\phi = \Delta^2\phi = \frac{\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}}{\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\|}$, provided $\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\| \neq 0$. A use of integration by parts, (A.2.1) and (3.3.4) leads to

$$\begin{aligned}
\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\| &= \int_\Omega (\mathcal{H} : \mathcal{H}\phi)\Pi_{\mathcal{D}}\psi_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} - \int_\Omega (\mathcal{H} : \mathcal{H}\phi)E_{\mathcal{D}}\psi_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} \\
&= \int_\Omega (\mathcal{H} : \mathcal{H}\phi)\Pi_{\mathcal{D}}\psi_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} - \int_\Omega \mathcal{H}(E_{\mathcal{D}}\psi_{\mathcal{D}}) : \mathcal{H}\phi \, \mathrm{d}\boldsymbol{x} \\
&= \int_\Omega (\mathcal{H} : \mathcal{H}\phi)\Pi_{\mathcal{D}}\psi_{\mathcal{D}} \, \mathrm{d}\boldsymbol{x} - \int_\Omega \mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}} : \mathcal{H}\phi \, \mathrm{d}\boldsymbol{x} \leq W_{\mathcal{D}}(\mathcal{H}\phi)\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|.
\end{aligned}$$

Let us now estimate (A.2.4). Using the divergence free property of ω and (3.3.5),

$$\begin{aligned}
&\int_\Omega (\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}) \cdot (\nabla\Gamma + \omega) \, \mathrm{d}\boldsymbol{x} \\
&\quad = \int_\Omega (\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}) \cdot \nabla\Gamma \, \mathrm{d}\boldsymbol{x} + \int_\Omega \nabla_{\mathcal{D}}\psi_{\mathcal{D}} \cdot \omega \, \mathrm{d}\boldsymbol{x} + \int_\Omega \Pi_{\mathcal{D}}\psi_{\mathcal{D}}\mathrm{div}(\omega) \, \mathrm{d}\boldsymbol{x} \\
&\quad \leq \int_\Omega (\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}) \cdot \nabla\Gamma \, \mathrm{d}\boldsymbol{x} + \widehat{W}_{\mathcal{D}}(\omega)\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|.
\end{aligned}$$

The first term on the right hand side of the above inequality can be estimated as follows. Set χ as in the theorem. Since $\operatorname{div}(\mathcal{H}\chi) = \nabla\Gamma$, using integration by parts, (A.2.1) and (3.3.6), we obtain

$$\begin{aligned}
&\int_\Omega \left(\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}\right)\cdot \operatorname{div}(\mathcal{H}\chi)\,\mathrm{d}\boldsymbol{x} \\
&\qquad \le \int_\Omega \nabla_{\mathcal{D}}\psi_{\mathcal{D}}\cdot\operatorname{div}(\mathcal{H}\chi)\,\mathrm{d}\boldsymbol{x} + \int_\Omega \mathcal{H}(E_{\mathcal{D}}\psi_{\mathcal{D}}) : \mathcal{H}\chi\,\mathrm{d}\boldsymbol{x} \\
&\qquad \le \int_\Omega \nabla_{\mathcal{D}}\psi_{\mathcal{D}}\cdot\operatorname{div}(\mathcal{H}\chi)\,\mathrm{d}\boldsymbol{x} + \int_\Omega \mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}} : \mathcal{H}\chi\,\mathrm{d}\boldsymbol{x} \\
&\qquad \le \|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|\widetilde{W}_{\mathcal{D}}(\mathcal{H}\chi).
\end{aligned}$$

Therefore, we get

$$\left|\int_\Omega (\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}})\cdot(\nabla\Gamma + \omega)\,\mathrm{d}\boldsymbol{x}\right| \le \|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|(\widetilde{W}_{\mathcal{D}}(\mathcal{H}\chi) + \widehat{W}_{\mathcal{D}}(\omega)).$$

□

Theorem A.2.2. *Recall $\delta(E_{\mathcal{D}})$, $\omega(E_{\mathcal{D}})$ and $\Gamma(E_{\mathcal{D}})$ from* (3.5.17)*. That is,*

$$\sup_{\psi_{\mathcal{D}}\in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - E_{\mathcal{D}}\psi_{\mathcal{D}}\|}{\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|} = \delta(E_{\mathcal{D}}), \quad \sup_{\psi_{\mathcal{D}}\in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\mathcal{H}E_{\mathcal{D}}\psi_{\mathcal{D}}\|}{\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|} = \Gamma(E_{\mathcal{D}}),$$

$$\text{and} \quad \sup_{\psi_{\mathcal{D}}\in X_{\mathcal{D},0}\setminus\{0\}} \frac{\|\nabla_{\mathcal{D}}\psi_{\mathcal{D}} - \nabla E_{\mathcal{D}}\psi_{\mathcal{D}}\|_{L^4(\Omega)^d}}{\|\mathcal{H}_{\mathcal{D}}\psi_{\mathcal{D}}\|} = \omega(E_{\mathcal{D}}).$$

Then, for a sequence $(\mathcal{D}_m)_{m\in\mathbb{N}}$ of Hessian discretisation that is coercive, consistent, limit-conforming and compact, it holds $\delta(E_{\mathcal{D}_m}) \to 0$ and $\omega(E_{\mathcal{D}_m}) \to 0$ as $m \to \infty$, and $(\Gamma(E_{\mathcal{D}_m}))_{m\in\mathbb{N}}$ is bounded.

Proof. A use of the estimate (A.2.2) leads to $\Gamma(E_{\mathcal{D}}) = 1$. We prove that $\delta(E_{\mathcal{D}_m})$ and $\omega(E_{\mathcal{D}_m})$ converge to 0 as $m \to \infty$ by way of contradiction. If this does not hold, there exist $\varepsilon_1, \varepsilon_2 \ge 0$ and a subsequence of $(\mathcal{D}_m)_{m\in\mathbb{N}}$, still denoted by $(\mathcal{D}_m)_{m\in\mathbb{N}}$, such that, for some $\psi_{\mathcal{D}_m} \in X_{\mathcal{D}_m,0}\setminus\{0\}$, we have

$$\|\Pi_{\mathcal{D}_m}\psi_{\mathcal{D}_m} - E_{\mathcal{D}_m}\psi_{\mathcal{D}_m}\| \ge \varepsilon_1 \quad \text{or} \quad \|\nabla_{\mathcal{D}_m}\psi_{\mathcal{D}_m} - \nabla E_{\mathcal{D}_m}\psi_{\mathcal{D}_m}\|_{L^4} \ge \varepsilon_2$$

for all $m \in \mathbb{N}$. Without loss of generality, assume $\|\mathcal{H}_{\mathcal{D}_m}\psi_{\mathcal{D}_m}\| = 1$. Thanks to the coercivity, the sequence $(\Pi_{\mathcal{D}_m}\psi_{\mathcal{D}_m})_{m\in\mathbb{N}}$ (resp. $(\nabla_{\mathcal{D}_m}\psi_{\mathcal{D}_m})_{m\in\mathbb{N}}$) remains bounded in $L^2(\Omega)$ (resp. $L^4(\Omega)^d$). Using the compactness hypothesis and Lemma 3.5.2, there exists $\psi \in H_0^2(\Omega)$ such that $\Pi_{\mathcal{D}_m}\psi_{\mathcal{D}_m}$ converges to ψ in $L^2(\Omega)$, $\nabla_{\mathcal{D}_m}\psi_{\mathcal{D}_m}$ converges to $\nabla\psi$ in $L^4(\Omega)^d$ and $\mathcal{H}_{\mathcal{D}_m}\psi_{\mathcal{D}_m}$ converges weakly to $\mathcal{H}\psi$ in $L^2(\Omega)^{d\times d}$. A use of (A.2.2) leads to the fact that $E_{\mathcal{D}}\Psi_{\mathcal{D}}$ is bounded in $H_0^2(\Omega)$. Therefore, there exists $\Gamma \in H_0^2(\Omega)$ such that $E_{\mathcal{D}_m}\psi_{\mathcal{D}_m}$ converges to Γ in $L^2(\Omega)$, $\nabla E_{\mathcal{D}_m}\psi_{\mathcal{D}_m}$ converges to $\nabla\Gamma$ in $L^4(\Omega)^d$ and $\mathcal{H}E_{\mathcal{D}_m}\psi_{\mathcal{D}_m}$ converges weakly to $\mathcal{H}\Gamma$ in $L^2(\Omega)^{d\times d}$. From (A.2.1), we obtain $\mathcal{H}\psi = \mathcal{H}\Gamma$ and, since $\psi - \Gamma \in H_0^2(\Omega)$, we get $\psi = \Gamma$ and $\nabla\psi = \nabla\Gamma$, which gives a contradiction. □

A.3 L^∞ estimates for the HMM method

Let us first briefly recall the GD corresponding to the HMM method [48, 50]. We consider a polytopal mesh $\mathcal{T} = (\mathcal{M}, \mathcal{F}, \mathcal{P})$ in the sense of Definition 1.4.1: $\mathcal{M}$ is the set of cells (generic notation K), $\mathcal{F}$ is the set of faces (generic notation σ) and $\mathcal{P}$ is a set made of one point per cell (notation $\boldsymbol{x}_K$ – this point does not need to be the center of mass of K in general). If $K \in \mathcal{M}$ then $\mathcal{F}_K$ is the set of faces of K. For $\sigma \in \mathcal{F}_K$, $|\sigma|$ is the measure of σ, $\overline{\boldsymbol{x}}_\sigma$ is the center of mass of σ, $d_{K,\sigma} = (\overline{\boldsymbol{x}}_\sigma - \boldsymbol{x}_K) \cdot n_{K,\sigma}$ is the orthogonal distance between $\boldsymbol{x}_K$ and σ, $n_{K,\sigma}$ is the outer normal to K on σ, and $D_{K,\sigma}$ is the convex hull of $\boldsymbol{x}_K$ and σ. An HMM GD is defined the following way.

- The degrees of freedom are made of one value in each cell and one value on each edge, so $X_{\mathcal{D},0} = \{v = ((v_K)_{K\in\mathcal{M}}, (v_\sigma)_{\sigma\in\mathcal{F}_K}) : v_K \in \mathbb{R},\ v_\sigma \in \mathbb{R},\ v_\sigma = 0 \text{ if } \sigma \subset \partial\Omega\}$.
- The reconstructed functions are piecewise constant in the cells: for $v \in X_{\mathcal{D},0}$, $\Pi_\mathcal{D} v \in L^2(\Omega)$ is defined by $(\Pi_\mathcal{D} v)_{|K} = v_K$ for all $K \in \mathcal{M}$.
- The reconstructed gradient is piecewise constant in the sets $(D_{K,\sigma})_{K\in\mathcal{M},\sigma\in\mathcal{F}_K}$: if $v \in X_{\mathcal{D},0}$, then $\nabla_\mathcal{D} v \in L^2(\Omega)^d$ is defined by

$$\forall K \in \mathcal{M},\ \forall \sigma \in \mathcal{F}_K,$$
$$(\nabla_\mathcal{D} v)_{|D_{K,\sigma}} = \overline{\nabla}_K v + \frac{\sqrt{d}}{d_{K,\sigma}} (v_\sigma - v_K - \overline{\nabla}_K v \cdot (\overline{\boldsymbol{x}}_\sigma - \boldsymbol{x}_K)) n_{K,\sigma},$$

  where

$$\overline{\nabla}_K v = \frac{1}{|K|} \sum_{\sigma\in\mathcal{F}_K} |\sigma| v_\sigma n_{K,\sigma}$$

  (this gradient is perhaps the most natural choice; though not the only possible choice within the HMM family; see [48, 50] for a more complete presentation).

Under standard local regularity assumptions on the mesh, [48, Propositions 12.14 and 12.15] yield the following error estimate on $\mathrm{WS}_\mathcal{D}$: if A is Lipschitz-continuous and $\psi \in H^2(\Omega)$, for some C not depending on ψ or $\mathcal{T}$:

$$\mathrm{WS}_\mathcal{D}(\psi) \le Ch \|\psi\|_{H^2(\Omega)}. \tag{A.3.1}$$

The following L^∞ error estimate and bound for the HMM is established under the quasi-uniformity assumption on the mesh.

Theorem A.3.1 (L^∞ estimates for HMM)**.** *Consider the dimension $d = 2$ or 3. Let $\mathcal{T}$ be a polytopal mesh and $\mathcal{D}$ be an HMM gradient discretisation. Take $\rho \ge \theta_\mathcal{T} + \zeta_\mathcal{D} + \chi_\mathcal{T}$, where $\theta_\mathcal{T}$ and $\zeta_\mathcal{D}$ are defined by [48, Eqs. (7.8) and (12.18)], and*

$$\chi_\mathcal{T} = \max_{K\in\mathcal{M}} \frac{h^d}{|K|}.$$

Assume that A is Lipschitz-continuous, that Ω is convex and that $F \in L^2(\Omega)$. There exists then C, depending only on Ω, A and ρ, such that, if ψ solves (5.3.2) *and $\psi_{\mathcal{D}}$ solves* (5.3.4)*,*

$$\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}} - \psi_{\mathcal{M}}\|_{L^\infty(\Omega)} \le C\|F\| \begin{cases} h|\ln(h)| & \text{if } d=2, \\ h^{1/2} & \text{if } d=3, \end{cases} \tag{A.3.2}$$

where $(\psi_{\mathcal{M}})_{|K} = \psi(\boldsymbol{x}_K)$ for all $K \in \mathcal{M}$, and

$$\|\Pi_{\mathcal{D}}\psi_{\mathcal{D}}\|_{L^\infty(\Omega)} \le C\|F\|. \tag{A.3.3}$$

Proof. In this proof, $X \lesssim Y$ means that $X \le MY$ for some M depending only on Ω, A and ρ. The theorem's assumptions ensure that $\psi \in H^2(\Omega) \cap H^1_0(\Omega) \subset C(\overline{\Omega})$.
Let $v = ((\psi(\boldsymbol{x}_K))_{K\in\mathcal{M}}, (\psi(\overline{\boldsymbol{x}}_\sigma))_{\sigma\in\mathcal{F}}) \in X_{\mathcal{D},0}$. By the proof of [48, Proposition A.6] (see also [48, (A.10)]),

$$\|\Pi_{\mathcal{D}}v - \psi\| + \|\nabla_{\mathcal{D}}v - \nabla\psi\| \lesssim h\|\psi\|_{H^2(\Omega)} \lesssim h\|F\|.$$

A use of (A.3.1) and the triangle inequality then gives

$$\|\nabla_{\mathcal{D}}(v - \psi_{\mathcal{D}})\| \lesssim h\|F\|. \tag{A.3.4}$$

[48, Lemma B.12] establishes the following discrete Sobolev embedding, for all $q \in [1,6]$ if $d=3$ and all $q \in [1,+\infty)$ if $d=2$:

$$\forall w \in X_{\mathcal{D},0},\ \|\Pi_{\mathcal{D}}w\|_{L^q(\Omega)} \lesssim q\|\nabla_{\mathcal{D}}w\|. \tag{A.3.5}$$

An inspection of the constants appearing in the proof of [48, Lemma B.12] shows that the inequality $\lesssim$ in (A.3.5) is independent of q. Substitute $w = v - \psi_{\mathcal{D}}$ in (A.3.5) and use (A.3.4) to obtain

$$\|\Pi_{\mathcal{D}}(v - \psi_{\mathcal{D}})\|_{L^q(\Omega)} \lesssim hq\|F\|.$$

Let $K_0 \in \mathcal{M}$ be such that $\|\Pi_{\mathcal{D}}(v - \psi_{\mathcal{D}})\|_{L^\infty(\Omega)} = |\Pi_{\mathcal{D}}(v - \psi_{\mathcal{D}})_{|K_0}|$ and write

$$\begin{aligned} \|\Pi_{\mathcal{D}}(v - \psi_{\mathcal{D}})\|_{L^\infty(\Omega)} &= |K_0|^{-1/q} \left(|K_0|\,|\Pi_{\mathcal{D}}(v - \psi_{\mathcal{D}})_{|K_0}|^q\right)^{1/q} \\ &\le \chi_{\mathcal{T}}^{1/q} h^{-\frac{d}{q}} \|\Pi_{\mathcal{D}}(v - \psi_{\mathcal{D}})\|_{L^q(\Omega)} \lesssim h^{1-\frac{d}{q}} q\|F\|. \end{aligned}$$

Minimising $q \mapsto h^{1-\frac{d}{q}}q$ over $q \in [1,\infty)$ if $d=2$, or taking $q=6$ if $d=3$, yields

$$\|\Pi_{\mathcal{D}}(v - \psi_{\mathcal{D}})\|_{L^\infty(\Omega)} \lesssim \|F\| \begin{cases} h|\ln(h)| & \text{if } d=2, \\ h^{1/2} & \text{if } d=3. \end{cases}$$

This concludes (A.3.2) since $\Pi_{\mathcal{D}}v = \psi_{\mathcal{M}}$. Estimate (A.3.3) follows from (A.3.2) by using the triangle inequality, the estimate $\|\psi_{\mathcal{M}}\|_{L^\infty(\Omega)} \le \|\psi\|_{L^\infty(\Omega)} \lesssim \|F\|$ and the property $\max(h^{1/2}, h|\ln(h)|) \lesssim 1$. □

Bibliography

[1] Yahya Alnashri and Jérôme Droniou, *Gradient schemes for the Signorini and the obstacle problems, and application to hybrid mimetic mixed methods*, Comput. Math. Appl. **72** (2016), no. 11, 2788–2807.

[2] Paola F. Antonietti, Nadia Bigoni, and Marco Verani, *Mimetic discretizations of elliptic control problems*, J. Sci. Comput. **56** (2013), no. 1, 14–27.

[3] Th. Apel, A. Rösch, and G. Winkler, *Discretization error estimates for an optimal control problem in a nonconvex domain*, Numerical Mathematics and Advanced Applications, Springer, Berlin, 2006, pp. 299–307.

[4] Thomas Apel, Mariano Mateos, Johannes Pfefferer, and Arnd Rösch, *Error estimates for Dirichlet control problems in polygonal domains: quasi-uniform meshes*, Math. Control Relat. Fields **8** (2018), no. 1, 217–245.

[5] Thomas Apel, Johannes Pfefferer, and Arnd Rösch, *Finite element error estimates for Neumann boundary control problems on graded meshes*, Comput. Optim. Appl. **52** (2012), no. 1, 3–28.

[6] ———, *Finite element error estimates on the boundary with application to optimal control*, Math. Comp. **84** (2015), no. 291, 33–70.

[7] Nadir Arada, Eduardo Casas, and Fredi Tröltzsch, *Error estimates for the numerical approximation of a semilinear elliptic control problem*, Comput. Optim. Appl. **23** (2002), no. 2, 201–229.

[8] S. Balasundaram and P. K. Bhattacharyya, *A mixed finite element method for fourth order elliptic equations with variable coefficients*, Comput. Math. Appl. **10** (1984), no. 3, 245–256.

[9] Lourenço Beirão da Veiga, Konstantin Lipnikov, and Gianmarco Manzini, *The mimetic finite difference method for elliptic problems*, MS&A. Modeling, Simulation and Applications, vol. 11, Springer, Cham, 2014.

[10] Maïtine Bergounioux, Kazufumi Ito, and Karl Kunisch, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim. **37** (1999), no. 4, 1176–1194.

[11] Maïtine Bergounioux and Karl Kunisch, *Primal-dual strategy for state-constrained optimal control problems*, Comput. Optim. Appl. **22** (2002), no. 2, 193–224.

[12] P. K. Bhattacharyya, *Mixed finite element methods for fourth order elliptic problems with variable coefficients*, Variational methods in engineering (Southampton, 1985), Springer, Berlin, 1985, pp. 2.3–2.12.

[13] H. Blum and R. Rannacher, *On the boundary value problem of the biharmonic operator on domains with angular corners*, Math. Methods Appl. Sci. **2** (1980), no. 4, 556–581.

[14] Susanne C. Brenner, *A two-level additive Schwarz preconditioner for nonconforming plate elements*, Numer. Math. **72** (1996), no. 4, 419–447.

[15] Susanne C. Brenner, Michael Neilan, Armin Reiser, and Li-Yeng Sung, *A C^0 interior penalty method for a von Kármán plate*, Numer. Math. **135** (2017), no. 3, 803–832.

[16] Susanne C. Brenner and L. Ridgway Scott, *The mathematical theory of finite element methods*, third ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008.

[17] Susanne C. Brenner, Li-yeng Sung, Hongchao Zhang, and Yi Zhang, *A Morley finite element method for the displacement obstacle problem of clamped Kirchhoff plates*, J. Comput. Appl. Math. **254** (2013), 31–42.

[18] F. Brezzi, *Finite element approximations of the von Kármán equations*, RAIRO Anal. Numér. **12** (1978), no. 4, 303–312, v.

[19] F. Brezzi, J. Rappaz, and P.-A. Raviart, *Finite-dimensional approximation of nonlinear problems. I. Branches of nonsingular solutions*, Numer. Math. **36** (1980/81), no. 1, 1–25.

[20] F. Brezzi and P.-A. Raviart, *Mixed finite element methods for 4th order elliptic equations, topics in numerical analysis, III (Proc. Roy. Irish Acad. Conf., Trinity Coll., Dublin, 1976)*, Academic Press, London, 1977.

[21] Franco Brezzi, Konstantin Lipnikov, and Mikhail Shashkov, *Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes*, SIAM J. Numer. Anal. **43** (2005), no. 5, 1872–1896.

[22] Franco Brezzi, Konstantin Lipnikov, and Valeria Simoncini, *A family of mimetic finite difference methods on polygonal and polyhedral meshes*, Math. Models Methods Appl. Sci. **15** (2005), no. 10, 1533–1551.

[23] Weidong Cao and Danping Yang, *Ciarlet-Raviart mixed finite element approximation for an optimal control problem governed by the first bi-harmonic equation*, J. Comput. Appl. Math. **233** (2009), no. 2, 372–388.

[24] C. Carstensen, G. Mallik, and N. Nataraj, *A priori and a posteriori error control of discontinuous Galerkin finite element methods for the von Kármán equations*, IMA J. Numer. Anal. **39** (2019), no. 1, 167–200.

[25] Carsten Carstensen and Dietmar Gallistl, *Guaranteed lower eigenvalue bounds for the biharmonic equation*, Numer. Math. **126** (2014), no. 1, 33–51.

[26] Eduardo Casas and Karl Kunisch, *Optimal control of semilinear elliptic equations in measure spaces*, SIAM J. Control Optim. **52** (2014), no. 1, 339–364.

[27] Eduardo Casas and Mariano Mateos, *Error estimates for the numerical approximation of Neumann control problems*, Comput. Optim. Appl. **39** (2008), no. 3, 265–295.

[28] Eduardo Casas, Mariano Mateos, and Jean-Pierre Raymond, *Error estimates for the numerical approximation of a distributed control problem for the steady-state Navier-Stokes equations*, SIAM J. Control Optim. **46** (2007), no. 3, 952–982.

[29] _____, *Penalization of Dirichlet optimal control problems*, ESAIM Control Optim. Calc. Var. **15** (2009), no. 4, 782–809.

[30] Eduardo Casas, Mariano Mateos, and Fredi Tröltzsch, *Error estimates for the numerical approximation of boundary semilinear elliptic control problems*, Comput. Optim. Appl. **31** (2005), no. 2, 193–219.

[31] Eduardo Casas and Jean-Pierre Raymond, *Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations*, SIAM J. Control Optim. **45** (2006), no. 5, 1586–1611.

[32] Eduardo Casas and Fredi Tröltzsch, *A general theorem on error estimates with application to a quasilinear elliptic optimal control problem*, Comput. Optim. Appl. **53** (2012), no. 1, 173–206.

[33] Yanzhen Chang and Danping Yang, *Superconvergence for optimal control problem governed by nonlinear elliptic equations*, Numer. Funct. Anal. Optim. **35** (2014), no. 5, 509–538.

[34] G. Chavent and J. Jaffré, *Mathematical models and finite elements for reservoir simulation*, Studies in Mathematics and its Applications, Vol. 17, North-Holland, Amsterdam, 1986.

[35] Hongtao Chen, Hailong Guo, Zhimin Zhang, and Qingsong Zou, *A C^0 linear finite element method for two fourth-order eigenvalue problems*, IMA J. Numer. Anal. **37** (2017), no. 4, 2120–2138.

[36] Jie Chen, Desheng Wang, and Qiang Du, *Linear finite element superconvergence on simplicial meshes*, Math. Comp. **83** (2014), no. 289, 2161–2185.

[37] Yanping Chen, *Superconvergence of mixed finite element methods for optimal control problems*, Math. Comp. **77** (2008), no. 263, 1269–1291.

[38] Yanping Chen, Yunqing Huang, Wenbin Liu, and Ningning Yan, *Error estimates and superconvergence of mixed finite element methods for convex optimal control problems*, J. Sci. Comput. **42** (2010), no. 3, 382–403.

[39] Sudipto Chowdhury and Thirupathi Gudi, *A C^0 interior penalty method for the Dirichlet control problem governed by biharmonic operator*, J. Comput. Appl. Math. **317** (2017), 290–306.

[40] Sudipto Chowdhury, Thirupathi Gudi, and A. K. Nandakumaran, *A framework for the error analysis of discontinuous finite element methods for elliptic optimal control problems and applications to C^0 IP methods*, Numer. Funct. Anal. Optim. **36** (2015), no. 11, 1388–1419.

[41] Philippe G. Ciarlet, *The finite element method for elliptic problems*, North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978, Studies in Mathematics and its Applications, Vol. 4.

[42] ———, *Mathematical elasticity. Vol. II*, Studies in Mathematics and its Applications, vol. 27, North-Holland Publishing Co., Amsterdam, 1997, Theory of plates.

[43] Philippe G. Ciarlet and Patrick Rabier, *Les équations de von Kármán*, Lecture Notes in Mathematics, vol. 826, Springer, Berlin, 1980.

[44] Daniele A. Di Pietro and Jérôme Droniou, *A hybrid high-order method for Leray-Lions elliptic equations on general meshes*, Math. Comp. **86** (2017), no. 307, 2159–2191.

[45] Daniele Antonio Di Pietro and Alexandre Ern, *Mathematical aspects of discontinuous Galerkin methods*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 69, Springer, Heidelberg, 2012.

[46] Jim Douglas, Jr., Todd Dupont, Peter Percell, and Ridgway Scott, *A family of C^1 finite elements with optimal approximation properties for various Galerkin methods for 2nd and 4th order problems*, RAIRO Anal. Numér. **13** (1979), no. 3, 227–255.

[47] J. Droniou and R. Eymard, *The asymmetric gradient discretisation method*, Finite volumes for complex applications VIII—methods and theoretical aspects, Springer Proc. Math. Stat., vol. 199, Springer, Cham, 2017, pp. 311–319.

[48] Jérôme Droniou, Robert Eymard, Thierry Gallouët, Cindy Guichard, and Raphaèle Herbin, *The gradient discretisation method*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 82, Springer, Cham, 2018.

[49] Jérôme Droniou, Robert Eymard, Thierry Gallouët, and Raphaèle Herbin, *A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods*, Math. Models Methods Appl. Sci. **20** (2010), no. 2, 265–295.

[50] Jérôme Droniou, Robert Eymard, Thierry Gallouet, and Raphaele Herbin, *Gradient schemes: a generic framework for the discretisation of linear, nonlinear and nonlocal elliptic and parabolic equations*, Math. Models Methods Appl. Sci. **23** (2013), no. 13, 2395–2432.

[51] Jérôme Droniou, Robert Eymard, and Raphaèle Herbin, *Gradient schemes: generic tools for the numerical analysis of diffusion equations*, ESAIM Math. Model. Numer. Anal. **50** (2016), no. 3, 749–781.

[52] Jérôme Droniou, Julian Hennicker, and Roland Masson, *Numerical analysis of a two-phase flow discrete fracture matrix model*, Numer. Math. **141** (2019), no. 1, 21–62.

[53] Jérôme Droniou, Bishnu P. Lamichhane, and Devika Shylaja, *The Hessian Discretisation Method for Fourth Order Linear Elliptic Equations*, J. Sci. Comput. **78** (2019), no. 3, 1405–1437.

[54] Jérôme Droniou and Neela Nataraj, *Improved L^2 estimate for gradient schemes and superconvergence of the TPFA finite volume scheme*, IMA J. Numer. Anal. **38** (2018), no. 3, 1254–1293.

[55] Jérôme Droniou, Neela Nataraj, and Devika Shylaja, *The gradient discretization method for optimal control problems, with superconvergence for nonconforming finite elements and mixed-hybrid mimetic finite differences*, SIAM J. Control Optim. **55** (2017), no. 6, 3640–3672.

[56] ———, *Numerical analysis for the pure Neumann control problem using the gradient discretisation method*, Comput. Methods Appl. Math. **18** (2018), no. 4, 609–637.

[57] Lawrence C. Evans, *Partial differential equations*, Graduate Studies in Mathematics, vol. 19, American Mathematical Society, Providence, RI, 1998.

[58] R. Eymard, T. Gallouët, and R. Herbin, *Finite volume methods*, Techniques of Scientific Computing, Part III (P. G. Ciarlet and J.-L. Lions, eds.), Handbook of Numerical Analysis, VII, North-Holland, Amsterdam, 2000, pp. 713–1020.

[59] R. Eymard, T. Gallouët, R. Herbin, and A. Linke, *Finite volume schemes for the biharmonic problem on general meshes*, Math. Comp. **81** (2012), no. 280, 2019–2048.

[60] Robert Eymard, Cindy Guichard, and Raphaèle Herbin, *Small-stencil 3D schemes for diffusive flows in porous media*, ESAIM Math. Model. Numer. Anal. **46** (2012), no. 2, 265–290.

[61] Robert Eymard, Cindy Guichard, Raphaèle Herbin, and Roland Masson, *Gradient schemes for two-phase flow in heterogeneous porous media and Richards equation*, ZAMM Z. Angew. Math. Mech. **94** (2014), no. 7-8, 560–585.

[62] Robert Eymard and Raphaèle Herbin, *Approximation of the biharmonic problem using piecewise linear finite elements*, C. R. Math. Acad. Sci. Paris **348** (2010), no. 23-24, 1283–1286.

[63] Richard S. Falk, *Approximation of the biharmonic equation by a mixed finite element method*, SIAM J. Numer. Anal. **15** (1978), no. 3, 556–567.

[64] S. Frei, R. Rannacher, and W. Wollner, *A priori error estimates for the finite element discretization of optimal distributed control problems governed by the biharmonic operator*, Calcolo **50** (2013), no. 3, 165–193.

[65] D. Gallistl, *Adaptive finite element computation of eigenvalues*, Doctoral dissertation, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2014.

[66] Dietmar Gallistl, *Morley finite element method for the eigenvalues of the biharmonic operator*, IMA J. Numer. Anal. **35** (2015), no. 4, 1779–1811.

[67] Lucia Gastaldi and Ricardo Nochetto, *Optimal L^∞-error estimates for nonconforming and mixed finite element methods of lowest order*, Numer. Math. **50** (1987), no. 5, 587–611.

[68] V. Girault and P.-A. Raviart, *Finite element approximation of the Navier-Stokes equations*, Lecture Notes in Mathematics, vol. 749, Springer-Verlag, Berlin-New York, 1979.

[69] Mark S. Gockenbach, *Understanding and implementing the finite element method*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006.

[70] P. Grisvard, *Elliptic problems in nonsmooth domains*, Monographs and Studies in Mathematics, vol. 24, Pitman (Advanced Publishing Program), Boston, MA, 1985.

[71] ———, *Singularities in boundary value problems*, Recherches en Mathématiques Appliquées [Research in Applied Mathematics], vol. 22, Masson, Paris; Springer-Verlag, Berlin, 1992.

[72] Thirupathi Gudi, Hari Shanker Gupta, and Neela Nataraj, *Analysis of an interior penalty method for fourth order problems on polygonal domains*, J. Sci. Comput. **54** (2013), no. 1, 177–199.

[73] Thirupathi Gudi, Neela Nataraj, and Kamana Porwal, *An interior penalty method for distributed optimal control problems governed by the biharmonic operator*, Comput. Math. Appl. **68** (2014), no. 12, part B, 2205–2221.

[74] Raphaèle Herbin and Florence Hubert, *Benchmark on discretization schemes for anisotropic diffusion problems on general grids*, Finite volumes for complex applications V, ISTE, London, 2008, pp. 659–692.

[75] M Ilyas, B. P. Lamichhane, and M.H. Meylan, *A gradient recovery method based on an oblique projection and boundary modification*, Proceedings of the 18th Biennial Computational Techniques and Applications Conference, CTAC-2016, ANZIAM J., vol. 58, 2017, pp. C34–C45.

[76] Chisup Kim, Raytcho D. Lazarov, Joseph E. Pasciak, and Panayot S. Vassilevski, *Multiplier spaces for the mortar finite element method in three dimensions*, SIAM J. Numer. Anal. **39** (2001), no. 2, 519–538.

[77] Axel Kröner and Boris Vexler, *A priori error estimates for elliptic optimal control problems with a bilinear state equation*, J. Comput. Appl. Math. **230** (2009), no. 2, 781–802.

[78] K. Krumbiegel and J. Pfefferer, *Superconvergence for Neumann boundary control problems governed by semilinear elliptic equations*, Comput. Optim. Appl. **61** (2015), no. 2, 373–408.

[79] K. Kunisch and A. Rösch, *Primal-dual active set strategy for a general class of constrained optimal control problems*, SIAM J. Optim. **13** (2002), no. 2, 321–334.

[80] B. P. Lamichhane, R. P. Stevenson, and B. I. Wohlmuth, *Higher order mortar finite element methods in 3D with dual Lagrange multiplier bases*, Numer. Math. **102** (2005), no. 1, 93–121.

[81] Bishnu P. Lamichhane, *A mixed finite element method for the biharmonic problem using biorthogonal or quasi-biorthogonal systems*, J. Sci. Comput. **46** (2011), no. 3, 379–396.

[82] ———, *A stabilized mixed finite element method for the biharmonic equation based on biorthogonal systems*, J. Comput. Appl. Math. **235** (2011), no. 17, 5188–5197.

[83] ———, *A finite element method for a biharmonic equation based on gradient recovery operators*, BIT **54** (2014), no. 2, 469–484.

[84] B.P. Lamichhane, *Higher Order Mortar Finite Elements with Dual Lagrange Multiplier Spaces and Applications*, Ph.D. thesis, Universität Stuttgart, 2006.

[85] P. Lascaux and P. Lesaint, *Some nonconforming finite elements for the plate bending problem*, Rev. Française Automat. Informat. Recherche Operationnelle Sér. Rouge Anal. Numér. **9** (1975), no. R-1, 9–53.

[86] Jichun Li, *Full-order convergence of a mixed finite element method for fourth-order elliptic equations*, J. Math. Anal. Appl. **230** (1999), no. 2, 329–349.

[87] Mingxia Li, Xiaofei Guan, and Shipeng Mao, *New error estimates of the Morley element for the plate bending problems*, J. Comput. Appl. Math. **263** (2014), 405–416.

[88] Yuan Li, Rong An, and Kaitai Li, *Some optimal error estimates of biharmonic problem using conforming finite element*, Appl. Math. Comput. **194** (2007), no. 2, 298–308.

[89] J.-L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Dunod; Gauthier-Villars, Paris, 1969.

[90] ———, *Optimal control of systems governed by partial differential equations.*, Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170, Springer-Verlag, New York-Berlin, 1971.

[91] Gouranga Mallik and Neela Nataraj, *Conforming finite element methods for the von Kármán equations*, Adv. Comput. Math. **42** (2016), no. 5, 1031–1054.

[92] ———, *A nonconforming finite element approximation for the von Karman equations*, ESAIM Math. Model. Numer. Anal. **50** (2016), no. 2, 433–454.

[93] Mariano Mateos and Arnd Rösch, *On saturation effects in the Neumann boundary control of elliptic optimal control problems*, Comput. Optim. Appl. **49** (2011), no. 2, 359–378.

[94] S. May, R. Rannacher, and B. Vexler, *Error analysis for a finite element approximation of elliptic Dirichlet boundary control problems*, SIAM J. Control Optim. **51** (2013), no. 3, 2585–2611.

[95] Pedro Merino, Fredi Tröltzsch, and Boris Vexler, *Error estimates for the finite element approximation of a semilinear elliptic control problem with state constraints and finite dimensional control space*, M2AN Math. Model. Numer. Anal. **44** (2010), no. 1, 167–188.

[96] C. Meyer and A. Rösch, *Superconvergence properties of optimal control problems*, SIAM J. Control Optim. **43** (2004), no. 3, 970–985 (electronic).

[97] C. Meyer and A. Rösch, *L^∞-estimates for approximated optimal control problems*, SIAM J. Control Optim. **44** (2005), no. 5, 1636–1649.

[98] Tetsuhiko Miyoshi, *A mixed finite element method for the solution of the von Kármán equations*, Numer. Math. **26** (1976), no. 3, 255–269.

[99] Neela Nataraj, P. K. Bhattacharyya, S. Balasundaram, and S. Gopalsamy, *On a mixed-hybrid finite element method for anisotropic plate bending problems*, Internat. J. Numer. Methods Engrg. **39** (1996), no. 23, 4063–4089.

[100] D.W. Peaceman, *Improved treatment of dispersion in numerical calculation of multidimensional miscible displacement*, Soc. Pet. Eng. J. **6** (1966), no. 3, 213–216.

[101] Peter Percell, *On cubic and quartic Clough-Tocher finite elements*, SIAM J. Numer. Anal. **13** (1976), no. 1, 100–103.

[102] M. J. D. Powell and M. A. Sabin, *Piecewise quadratic approximations on triangles*, ACM Trans. Math. Software **3** (1977), no. 4, 316–325.

[103] T. Scapolla, *A mixed finite element method for the biharmonic problem*, RAIRO Anal. Numér. **14** (1980), no. 1, 55–79.

[104] Devika Shylaja, *Improved L^2 and H^1 error estimates for the Hessian discretisation method*, (2019), Submitted. https://arxiv.org/pdf/1811.05429.pdf.

[105] Guido Stampacchia, *Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus*, Ann. Inst. Fourier (Grenoble) **15** (1965), no. fasc. 1, 189–258.

[106] Gilbert Strang, *Variational crimes in the finite element method*, The mathematical foundations of the finite element method with applications to partial differential equations (Proc. Sympos., Univ. Maryland, Baltimore, Md., 1972), Academic Press, New York, 1972, pp. 689–710.

[107] Gilbert Strang and George Fix, *An analysis of the finite element method*, second ed., Wellesley-Cambridge Press, Wellesley, MA, 2008.

[108] Roger Temam, *Navier-Stokes equations. Theory and numerical analysis*, North-Holland Publishing Co., Amsterdam-New York-Oxford, 1977, Studies in Mathematics and its Applications, Vol. 2.

[109] Fredi Tröltzsch, *Optimal control of partial differential equations*, Graduate Studies in Mathematics, vol. 112, American Mathematical Society, Providence, RI, 2010, Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.

[110] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.

[111] Zheng Hui Xie, *Error estimate of nonconforming finite element approximation for a fourth order elliptic variational inequality*, Northeast. Math. J. **8** (1992), no. 3, 329–336.

[112] Jinchao Xu and Zhimin Zhang, *Analysis of recovery type a posteriori error estimators for mildly structured grids*, Math. Comp. **73** (2004), no. 247, 1139–1152.

List of Publications from Thesis

Papers Published/Submitted

1. Jérôme Droniou, Neela Nataraj and Devika Shylaja. The gradient discretisation method for optimal control problems, with super-convergence for non-conforming finite elements and mixed-hybrid mimetic finite differences. *SIAM J. Control Optim.* 55 (6), pp. 3640-3672, 2017. DOI: 10.1137/17M1117768. URL: https://arxiv.org/abs/1608.01726.

2. Jérôme Droniou, Neela Nataraj and Devika Shylaja. Numerical analysis for the pure Neumann control problem using the gradient discretisation method. *Comput. Meth. Appl. Math*. 18 (4), pp. 609-637, 2018. DOI: 10.1515/cmam-2017-0054. URL: https://arxiv.org/abs/1705.03256.

3. Jérôme Droniou, Bishnu. P. Lamichhanne and Devika Shylaja. The Hessian discretisation method for fourth order linear elliptic equations. *Journal of Scientific Computing*, 32p, 2018. DOI: 10. 1007/s10915-018-0814-7. URL: https://arxiv.org/abs/1803.06985.

4. Devika Shylaja. Improved L^2 and H^1 error estimates for the Hessian discretisation method. *Submitted, 2019.* URL: https://arxiv.org/abs/1811.05429.

5. Jérôme Droniou, Neela Nataraj and Devika Shylaja. Hessian discretisation method for fourth order semi-linear elliptic equations. *Submitted, 2019.*

Acknowledgements

Let me take a moment to express my gratitude to all the people who have supported me during various stages of my research.

I am deeply indebted to my supervisors, Prof. Neela Nataraj and Prof. Jérôme Droniou, for their valuable guidance and constant support throughout this study. Their contributions have been vital in providing direction to this thesis and have helped me transform abstract ideas into mathematical results. With their expert knowledge, clarity of thought and gentle words of encouragement they have made sure that I stayed on course during these past four years.

It has been a great experience to work with Prof. Neela Nataraj who I have come to regard as a personal and professional role model. I thank her for always finding time despite a busy schedule to discuss, review and revise my work multiple time thereby ensuring that I was able to achieve her stratospheric standards.

Prof. Jérôme Droniou has been a constant source of insightful advice and constructive feedback and has been instrumental in giving shape to my research. I thank him for his patience and for spending so much time to help me refine and polish my thesis.

I am extremely grateful to Prof. Bishnu. P. Lamichhane, University of Newcastle, for giving me the opportunity to collaborate with him which helped to expand the scope of my research work. A special thanks to my senior Sudipto Chowdhury for his help and mathematical advice during the study.

I truly appreciate the insights and comments provided by my Research Progress Committee members, Prof. K. Suresh Kumar, Prof. Hans De Sterck and Prof. Janosch Rieger, during my annual progress seminar presentations.

I would also like to express my gratitude to the staff members of IITB-Monash Research Academy, IIT Bombay for providing administrative support during the course of my research. I am thankful to the teaching and non-teaching staff members of the Department of Mathematics, IIT Bombay and School of Mathematical Sciences, Monash University for their kind assistance.

I would like to thank my friends for their invaluable friendship, encouragement and moral support. In particular, I would like to mention Akansha, Rakhi Singh, Shashibhushan Biliangadi, Vinod Vijay Kumar and Bankim Mahanta who made my stay at IIT Bombay and Monash University campus memorable and enjoyable. I am also grateful to Ruma Rani Maity, Gopikrishnan C R, Wasim Akram and all my other colleagues from both the institutes for their assistance, understanding and support.

Special thanks to my mother, brother and husband for their love and support. I am also grateful to all the other people who I have not mentioned by name for their support, help and advice during my research.