Adaptive Spline Finite Element Methods
for fourth order elliptic problems

Ibrahim Harib Al Balushi

Department of Mathematics and Statistics
McGill University, Montreal

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy (August 2020)

This thesis is dedicated to my mother, to my sister and to my late friend Daniel Smyth

Thesis abstract

This thesis concentrates on the error analysis of B-spline based finite-element methods for three fourth-order elliptic partial differential equations subject to essential boundary conditions. The first being the *biharmonic equation* with square-integrable right-hand side and the second and third are models for *quasi-geostrophic equations* (QGE) simulating large-scale wind-driven oceanic currents.

The large scale ocean currents transport heat around the globe which is important in understanding the climate system due to their influence on the temperature variations across many of the Earth's regions. Understanding the dynamics of matter and heat transport resulting from global wind patterns is a non-trivial computational task that can be tackled using finite-element methods. Due to the Earth's axial rotation, annual wind patterns are westward near the equator and eastward at the mid-latitudes. These wind patterns drive strong western intensification of oceanic circulations as demonstrated in the Gulf Stream in the Atlantic ocean, the Kuroshio in the Pacific ocean, the Brazil Current in the south Atlantic ocean, and the Agulhas current off the east coast of Africa. As a consequence, the most striking features of oceanic currents are strong western boundary currents along with weak interior flows and weak eastern boundary currents characterized by the wind forcing and the effects of rotation. The *quasi-geostrophic equations* (QGE) are one of the most popular mathematical models to predict the wind-driven ocean circulation at mid-latitudes. Although the QGE allow for efficient computational simulations, the solutions of the QGE can lead to spurious oscillations and poor resolution due to the presence of boundary layers. The accuracy of the solutions can be improved with refined meshes in the regions where the boundary layer arises. The QGE are derived from the Navier-Stokes equations where the velocity vector field is expressed in terms of steam-function potential. The importance of studying this model is because it is the standard test problem in the geophysical fluid dynamics literature. In this thesis we address adaptive h-refinement finite-element methods for the stationary quasi- geostrophic equation and the Stommel–Munk model.

In contrast to standard Lagrange finite-element methods, B-spline based finite-element approximations can easily obtain high order continuity at a relatively low computational cost. Moreover, complex geometries can be represented accurately using B-splines and NURBS basis functions. In fact, Hughes, Cottrell, and Bazilevs in 2005 introduced an *isogeometric* framework for finite element analysis utilizing the aforementioned geometric capabilities. However, B-splines are non-interpolatory, and imposing even simple boundary conditions can be problematic. Furthermore, for problems where the formation of boundary layers is an important feature, the imposition of boundary conditions in a strong manner may not be appropriate all-together since it may induce artificial oscillations in the solution and it reduces the accuracy of the underlying numerical method as pointed out by Bazilevs and Hughes in 2017. Instead, one can impose the boundary conditions weakly using a classical method by Nitsche dating back to 1971. The so-called *Nitsche's method* has been successfully applied to impose boundary conditions for the second- and fourth-order partial differential equations in a 2010 publication by Embar, Dolbow and Harari. The essence of the method is rooted in the method of Lagrange multipliers with the addition of stabilization penalty terms. The first analysis of this kind was done for the Poisson problem by Babuška in 1973, then generalized by Barbosa and Hughes in 1991 to general Hilbert space setting and later brought to the spotlight by Sternberg in 1995. Kim and Jiang introduced Nitsche-type variational formulations for the stream-function formulation of the *stationary* quasi-geostrophic equation (SQGE) and its simplified linear version the *Stommel–Munk model.* These Nitsche-type formulations can be readily applied for non-interpolatory basis functions such as B-splines and embedded geometries, where the domain can be implicitly defined via a level-set function. The purpose of this thesis, among other things, is to analyze the performance of adaptivity for these Nitsche-type methods and supplement the analyses with benchmark numerical simulations.

The goal of this thesis is two-fold. On one hand, we derive and analyze error estimators for the purpose of adaptive h-refinement. The earliest effort was concerned with the linear Stommel-Munk. We note that a second-order treatment has been done in 2009 by Juntunen and Stenberg where the analysis hinges on a so-called *saturation assumption* to relate the numerical error with the discrete error between two refinements. We carry out a similar analysis for the fourth-order PDE. In the nonlinear SQGE we perform the error analysis *without* a saturation assumption making this work novel in two ways: The treatment requires dealing with the nonlinear convective term and the reliability proofs are saturation-assumption free.

The second goal of this thesis is concerned with the convergence and optimality of Nitsche-type adaptive methods for the biharmonic equation. Such a study for general second order elliptic order equations has been extensively studied when essential boundary conditions are prescribed into the discrete space. The first convergence proof for the Poisson problem was

given by Dörfler in 1996 and improved on by Morin, Nochetto, and Siebert in 2000 where some stringent conditions on the domain partitions were removed. Those ideas were soon to be extended to general second order linear elliptic problems by Mekchay and Nochetto, and finally a convergence analysis in a Hilbert space setting was given by Morin, Siebert and Veeser. The first analysis of convergence rates and quasi-optimality for the Poisson problem is pioneered by Binev, Dahmen and DeVore in 2004 and also by Stevenson where he removed an artificial coarsening step. Those ideas were applied to symmetric second order linear elliptic problems by Cascón, Kreuzer, Nochetto and Siebert and further generalized by Feischl, Führer and Praetorius to non-symmetric linear problems as well as to strongly monotone nonlinear operators. We add that all aforementioned literature consider boundary condition *conforming* finite-element spaces in that those discrete spaces satisfy the boundary conditions. For completeness, we do the same for the biharmonic problem. As far as non-conforming methods are concerned, to the best of our knowledge, no such study has been made for Nitsche's method before the appearance of our work, not even for the Poisson problem. The closest situation we have is that of discontinuous Galerkin methods for symmetric second order elliptic problems which we draw our inspiration from. The convergence and quasi-optimality of discontinuous Galerkin methods was studied by Bonito, Andrea and Nochetto in 2010.

Résumé de la thèse

Cette thèse se concentre sur l'analyse des erreurs des méthodes aux éléments finis basées sur la spline B pour trois équations différentielles partielles (EDP) elliptiques du quatrième ordre soumises à des conditions limites essentielles. La première de ces EDP est l'équation biharmonique avec le côté droit intégré au carré, et a deuxième et troisième EDP sont des modèles pour les équations quasi-géostrophiques (EQG) simulant les courants océaniques à grande échelle poussés par le vent.

Les courants océaniques à grande échelle transportent la chaleur autour du globe. Ceci est important pour comprendre le système climatique en raison de leur influence sur les variations de température dans de nombreuses régions de la Terre. La compréhension de la dynamique du transport de la matière et de la chaleur résultant de la configuration des vents à l'échelle planétaire est une tâche de calcul non triviale qui peut être abordée à l'aide de méthodes par éléments finis. En raison de la rotation axiale de la Terre, les vents annuels se dirigent vers l'ouest près de l'équateur et vers l'est aux latitudes moyennes. Ces régimes de vent entraînent une forte intensification des circulations océaniques vers l'ouest, comme le montrent le Gulf Stream dans l'océan Atlantique, le Kuroshio dans l'océan Pacifique, le courant du Brésil dans l'océan Atlantique sud et le courant des Aiguilles au large de la côte est de l'Afrique. En conséquence, les caractéristiques les plus frappantes des courants océaniques sont les forts courants de l'ouest ainsi que les faibles flux intérieurs et les faibles courants de frontière caractérisés par le forcage du vent et les effets de rotation. Les EQG sont l'un des modèles mathématiques les plus populaires pour prédire la circulation océanique due au vent aux latitudes moyennes. Bien que les EQG permettent des simulations informatiques efficaces, les solutions des EQG peuvent produire à des oscillations parasites et une mauvaise résolution en raison de la présence de couches limites. La précision des solutions peut être améliorée grâce à des maillages raffinés dans les régions où la couche limite est présente. Les EQG sont dérivées des équations de Navier-Stokes où le champ du vecteur vitesse est exprimé

en termes du potentiel de la fonction vapeur. Ce modeèle est important vu qu'il s'agit du problème de test standard dans la littérature sur la dynamique des fluides géophysiques. Dans cette thèse, nous abordons les méthodes adaptatives d'éléments finis de raffinement H pour l'équation stationnaire quasi-géostrophique et le modèle de Stommel-Munk.

Contrairement aux méthodes standard d'éléments finis de Lagrange, les approximations d'éléments finis basées sur la spline B peuvent facilement obtenir une continuité d'ordre élevé à un coût de calcul relativement faible. De plus, les géométries complexes peuvent être représentées avec précision à l'aide de B-splines et de fonctions de base NURBS. En fait, Hughes, Cottrell et Bazilevs ont introduit en 2005 un cadre isogéométrique pour l'analyse par éléments finis en utilisant les capacités géométriques mentionnées ci-dessus. Cependant, les B-splines sont non interprétables et l'imposition de conditions limites, même simples, peut être problématique. En outre, pour les problèmes où la formation de couches limites est une caractéristique importante, l'imposition de conditions limites de manière forte peut ne pas être appropriée dans son ensemble car elle peut induire des oscillations artificielles dans la solution et elle réduit la précision de la méthode numérique sous-jacente comme l'ont souligné Bazilevs et Hughes en 2017. Au lieu de cela, on peut imposer les conditions limites de manière faible en utilisant une méthode classique de Nitsche datant de 1971. La méthode dite de Nitsche a été appliquée avec succès pour imposer des conditions limites pour les EDP du deuxième et du quatrième ordre dans une publication de Embar, Dolbow et Harari en 2010. L'essence de la méthode est ancrée dans la méthode des multiplicateurs de Lagrange avec l'ajout de termes de pénalité de stabilisation. La première analyse de ce type a été réalisée pour le problème de Poisson par Babuška en 1973, puis généralisée par Barbosa et Hughes en 1991 à l'ensemble de l'espace de Hilbert, et enfin mise en lumière par Sternberg en 1995. Kim et Jiang ont introduit des formulations variationnelles de type Nitsche pour la formulation de la fonction de flux de l'équation quasi-géostrophique stationnaire (EQGS) et sa version linéaire simplifiée, le modèle de Stommel-Munk. Ces formulations de type Nitsche peuvent être facilement appliquées pour des fonctions de base non interpolatoires telles que les splines B et les géométries intégrées, où le domaine peut être implicitement défini par une fonction de niveau.

L'objectif de cette thèse est, entre autres, d'analyser les performances de l'adaptabilité pour ces méthodes de type Nitsche et de compléter les analyses par des simulations numériques de référence. D'une part, nous dérivons et analysons les estimateurs d'erreur pour le raffinement H adaptatif. Le premier effort a porté sur le Stommel-Munk linéaire. Nous notons qu'un traitement de second ordre a été effectué en 2009 par Juntunen et Stenberg où l'analyse repose sur une hypothèse dite de saturation pour relier l'erreur numérique à l'erreur discrète entre deux raffinements. Nous effectuons une analyse similaire pour l'EDP du quatrième ordre. Dans l'QSE non linéaire, nous effectuons l'analyse d'erreur sans hypothèse de saturation, ce qui rend ce travail inédit de deux manières : le traitement nécessite de traiter le terme convectif non linéaire et les preuves de fiabilité sont sans hypothèse de saturation.

D'autre part cette thèse concerne la convergence et l'optimalité des méthodes adaptatives de type Nitsche pour l'équation biharmonique. Une telle approche pour les EDP elliptiques générales second ordre a été largement étudiée lorsque des conditions aux limites essentielles sont prescrites dans l'espace discret. La première preuve de convergence pour le problème de Poisson a été donnée par Dörfler en 1996 et améliorée par Morin, Nochetto, et Siebert en 2000 où certaines conditions strictes sur les partitions de domaine ont été supprimées. Ces idées ont bientôt été étendues à des problèmes elliptiques linéaires généraux du second ordre par Mekchay et Nochetto, et enfin une analyse de convergence dans un cadre spatial de Hilbert a été donnée par Morin, Siebert et Veeser. La première analyse des taux de convergence et de la quasi-optimalité pour le problème de Poisson est lancée par Binev, Dahmen et DeVore en 2004 et également par Stevenson où il supprime une étape de grossir artificielle. Ces idées ont été appliquées à des problèmes elliptiques linéaires symétriques du second ordre par Casc'on, Kreuzer, Nochetto et Siebert, puis généralisées par Feischl, Führer et Praetorius à des problèmes linéaires non symétriques ainsi qu'à des opérateurs non linéaires fortement monotones. Nous ajoutons que toute la littérature susmentionnée considère les espaces à éléments finis conformes aux conditions limites en ce sens que ces espaces discrets satisfont aux conditions limites. Par souci d'exhaustivité, nous faisons de même pour le problème biharmonique. En ce qui concerne les méthodes non conformes, à notre connaissance, aucune étude de ce type n'a été réalisée pour la méthode de Nitsche avant l'apparition de nos travaux, même pas pour le problème de Poisson. La situation la plus proche que nous ayons est celle des méthodes discontinues de Galerkin pour les problèmes elliptiques symétriques du second ordre dont nous nous inspirons. La convergence et la quasi-optimalité des méthodes Galerkin discontinues ont été étudiées par Bonito, Andrea et Nochetto en 2010.

Acknowledgements

Forth and foremost I thank my academic advisor Dr. Gantumur Tsogtgerel for his mentorship, his extraordinary patience, unfettered support, and above all for being a friend. I have learned a lot from him over the years and if there is anything which I benefited the most it is patience, persistence and focus.

I would also like to thank my undergraduate mentor Dr. Galia Dafni who unintentionally inspired my switch from engineering to mathematical analysis. She always showed genuine enthusiasm for my professional advancement and keeping contact with her is always a pleasure. I must add that I might have drawn some of my teaching style from her.

I have learnt a lot from taking graduate classes with Dr. Rustum Choksi, Dr. Paul Koosis, Dr. Jean Christophe Nave and Dr.Adam Oberman. I give my special thanks Dr. Axel Hunderman for always being there for me while I developed my teaching skills.

Throughout my academic years I have met some of the greatest and dynamic colleagues among whom were Andy, Geoff, Chris, Tiago, Bilal and Manuela. I give a special thanks to Geoff's Wednesday running club.

Furthermore, I would not have succeeded without the moral support of my immediate family and friends. In particular my mother, my sister Rasha, my cousin Samar, my friend Francois, my friend Liz, my friend Laurent, my friend Shawn, my friend Brandon and my late friend Danial who were all with me every step of the way with all the painful moments and frustrations.

Finally, I thank Emmanuil Georgoulis for his invaluable advice on discontinuous Galerkin methods.

Ibrahim Al Balushi
Montreal, August 2020

Contents

Notations and acronyms

notation	meaning
$\mathbb{N}$	the natural numbers $1, 2, 3, \ldots$
$\mathbb{Z}$, $\mathbb{R}$	integers and real numbers, respectively
Ω, $\partial\Omega$	bounded Lipschitz domain in $\mathbb{R}^d$, and its boundary
$\mathcal{C}^k(\Omega)$	the space of functions with continuous k-order derivatives
$L^p(\Omega)$, L^p	the space of functions on Ω for which $\int_\Omega \|f\|^p$ is finite
$W^s_p(\Omega)$, W^s_p	the Sobolev space with smoothness s measured in L^p
$H^s(\Omega)$, H^s	equal to $W^s_2(\Omega)$
$H^s_0(\Omega)$, H^s_0	the closure of $C^\infty_0(\Omega)$ in $H^s(\Omega)$
$\mathscr{B}^s_q(L^p(\Omega))$, $\mathscr{B}^s_q(L^p)$, $\mathscr{B}^s_{p,q}$	Besov space with smoothness s measured in L^p and secondary index q
$\mathcal{L}(X, Y)$	bounded linear operators between two normed spaces X and Y
$\mathbb{V}$, $\mathbb{W}$	Banach or Hilbert spaces
$\mathbb{V}'$	the dual of $\mathbb{V}$
$\\|\cdot\\|_{\mathbb{V}}$	the norm associated to $\mathbb{V}$
$u, v, w, \ldots$	elements of $\mathbb{V}$ or $\mathbb{W}$
$\langle\cdot,\cdot\rangle$	the duality product on $\mathbb{V} \times \mathbb{V}'$
$\\|\cdot\\|$	the standard norm on L^2
$(\cdot,\cdot)$	the standard inner product in L^2
P, τ	a partitioning of domain Ω and a generic member cell
$\mathcal{E}_P$, $\mathcal{G}_P$, σ	interior edges, boundary edges and a generic member edge
$\mathbb{X}_P$	the B-spline based finite element space consisting of C^1 or C^2 piecewise polynomial splines

notation	meaning
$\|\cdot\|_{s,P}$	the mesh-dependent semi-norm for functions on Ω defined by weighted L^2 integrals along the boundary $\partial\Omega$; see (4.10)
$f \lesssim g$	$f \leq C{\cdot}g$ with a constant $C > 0$ that may depend only on *fixed* constants under consideration
$f \approx g$	$f \lesssim g$ and $g \lesssim f$
$\oslash$	end of example, definition, or long remark
■	end of proof

Chapter 1

Thesis overview

Let Ω be a two-dimensional bounded domain with a polygonal boundary Γ. Given a forcing function $f \in L^2(\Omega)$ and a non-dimensional parameter $\alpha > 0$, we consider the solution u of a fourth-order partial differential equation with homogeneous Dirichlet boundary conditions

$$\begin{cases} \alpha\Delta^2 u + F_0(\cdot, u, \nabla u, \Delta u) + \operatorname{div}\boldsymbol{F}(\cdot, u, \nabla u, \Delta u) = f & \text{in } \Omega, \\ u = \frac{\partial u}{\partial \boldsymbol{n}} = 0 & \text{on } \Gamma, \end{cases} \tag{1.1}$$

for functions $F_i = F_i(x, u, \nabla u, \Delta u)$, $i = 0, 1, 2$, and $\boldsymbol{F} = (F_1, F_2)$ taking three forms:

1. $F_i \equiv 0$ for all i's and $\alpha = 1$ corresponding to the biharmonic operator equation with the right-hand side f. The convergence and quasi-optimality analyses of an h-refinement adaptive method for this differential equation will be addressed. Two approaches will be considered: the first, being the content of Chapter 3, uses polynomial spline-based finite elements which enforces the essential boundary conditions in a strong manner. Specifically, the boundary conditions are encoded in the discrete space. The second approach, which is the subject of discussion of Chapter 6, encodes the boundary conditions indirectly through Nitsche's penalty terms.

2. Here $F_0 = -\varepsilon_{\mathrm{s}}\Delta u - \frac{\partial u}{\partial x}$, $F_1 = F_2 \equiv 0$ and $\alpha = \varepsilon_{\mathrm{m}}$ corresponds to the Stommel–Munk model (Vallis [66]) for wind-driven ocean circulation in an enclosed midlatitude basin, where u and f can be the velocity streamfunction and the wind forcing, respectively. The parameters ε_{s} and ε_{m} are the nondimensional *Stommel* and *Munk numbers*, respectively, which are defined by

$$\varepsilon_{\mathrm{s}} = \frac{\gamma}{\beta L} \quad \text{and} \quad \varepsilon_{\mathrm{m}} = \frac{A}{\beta L^3}, \tag{1.2}$$

where γ is the coefficient of the linear drag (or the Rayleigh friction) as might be generated by a bottom Ekman layer, β is the coefficient multiplying the y-coordinate in the β-plane approximation, A is the eddy viscosity parametrization, and L is the characteristic length scale. In addition to the Laplacian and biharmonic terms, the model involves the rotation term $\frac{\partial u}{\partial x}$ which is also added to introduce an asymmetry in the east-west direction. This is the subject of discussion in Chapter 4.

3. Finally, $F_0 = -\mathrm{Ro}^{-1}\frac{\partial u}{\partial x}$, $\boldsymbol{F} = \Delta u \nabla^{\perp} u$ where $\nabla^{\perp} u = (\frac{\partial u}{\partial y}, -\frac{\partial u}{\partial x})$, and $\alpha = \mathrm{Re}^{-1}$ corresponding to the streamfunction formulation of the one-layer SQGE (Vallis [66] and Foster [37]) for a given velocity streamfunction u and a wind forcing f with $\mathrm{Ro}^{-1}f$ on the right-hand side. The Re and Ro are the *Reynolds* and *Rossby numbers* defined by
$$\mathrm{Re} = \frac{UL}{A} \quad \text{and} \quad \mathrm{Ro} = \frac{U}{\beta L^2}, \tag{1.3}$$
respectively. The Rossby number is a measure of the significance of earth's rotation. Here, U is the characteristic velocity scale. For large scale oceanic flows, the Reynolds number Re is large and the Rossby number Ro is small, indicating small diffusion and large rotation. Thus, the SQGE is dominated by convective terms with the large wind forcing.

The weak formulation for the general problem (1.1) in the Hilbert space $H_0^2(\Omega)$ reads
$$\text{Find } u \in H_0^2(\Omega) \quad \text{such that} \quad a(u,v) + b(u;v) = \ell_f(v) \quad \forall v \in H_0^2(\Omega), \tag{1.4}$$
where the principal part $a : H_0^2(\Omega) \times H_0^2(\Omega) \to \mathbb{R}$ is given by $a(u,v) = \alpha \int_\Omega \Delta u \Delta v$ and $b : H_0^2(\Omega) \times H_0^2(\Omega) \to \mathbb{R}$ is either bilinear or nonlinear in both arguments but linear in the second argument. In both cases the form b is non-symmetric and given by
$$b(u;v) = \int_\Omega F_0(\cdot, u, \nabla u, \Delta u) v - \int_\Omega \boldsymbol{F}(\cdot, u, \nabla u, \Delta u) \cdot \nabla v. \tag{1.5}$$
Finally, $\ell_f \in H^{-2}(\Omega)$ and reads $\ell_f(v) = \int_\Omega fv$. In Chapter 4 the essential boundary conditions are prescribed into the B-spline polynomial space $\mathbb{X}_P$ leading to the discretization:
$$\text{Find } U \in \mathbb{X}_P \quad \text{such that} \quad a(U,V) + b(U;V) = \ell_f(V) \quad \forall V \in \mathbb{X}_P. \tag{1.6}$$
For Nitsche-type methods we initially modify the principal part a to define a new mesh-dependent principal part $a_P : \mathbb{X}_P \times \mathbb{X}_P \to \mathbb{R}$
$$\begin{aligned} a_P(U,V) = \alpha \int_\Omega \Delta U \Delta V + \alpha \int_\Gamma \left(\tfrac{\partial \Delta U}{\partial \boldsymbol{n}} V + U \tfrac{\partial \Delta V}{\partial \boldsymbol{n}}\right) \\ - \alpha \int_\Gamma \left(\Delta U \tfrac{\partial V}{\partial \boldsymbol{n}} + \tfrac{\partial U}{\partial \boldsymbol{n}} \Delta U\right) + \gamma_1 \int_\Gamma h_P^{-3} UV + \gamma_2 \int_\Gamma h_P^{-1} \tfrac{\partial U}{\partial \boldsymbol{n}} \tfrac{\partial V}{\partial \boldsymbol{n}}, \end{aligned} \tag{1.8}$$

where γ_1 and γ_2 are positive stabilization parameters. The lower order terms of form b remains unchanged. Actually in our initial work (Chapter 4) on the linear Stommel-Munk, we did modify b as well, but later it became clear that this was not necessary; compare with Chapter 5. Therefore in hindsight we leave b unchanged here. We solve:

$$\text{Find } U \in \mathbb{X}_P \quad \text{such that} \quad a_P(U,V) + b(U;V) = \ell_f(V) \quad \forall V \in \mathbb{X}_P. \tag{1.9}$$

While (1.9) produces the right solution to (1.6), the *a posteriori* estimation will only be possible using a saturation assumption as a price to pay for having a_P be well-defned for only $\mathbb{X}_P$. The use of this saturation assumption is justifiable for the Poisson problem since a discrete lower bound can be derived. This has been done when the finite-element discretization satisfies the boundary conditions (Dörfler [53] and Morin et al. [54]). However to the best of our knowledge no such estimate exists for the fourth-order problem or when Nitsche's method is employed even for the Poisson problem. We do not attempt to derive a discrete lower bound because it is shown to be unessential for convergence or quasi-optimality (Cascón et al. [61], Feischl et al. [62] and Siebert et al. [58]). We show that the dominance of the *a posteriori* error estimator over the numerical error is realized *without* a saturation assumption but at the cost of the discrete methods consistency with (1.4). Nevertheless, the resulting inconsistency is shown to weaken with refinement. Here we modify a_P to be

$$\begin{aligned} a_P(u,v) := \alpha \int_\Omega \Delta u \Delta v - \alpha \int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} v + u \tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}} \right) \\ + \alpha \int_\Gamma \left(\Pi_P(\Delta u) \tfrac{\partial v}{\partial \boldsymbol{n}} + \tfrac{\partial u}{\partial \boldsymbol{n}} \Pi_P(\Delta v) \right) + \gamma_1 \int_\Gamma h_P^{-3} uv + \gamma_2 \int_\Gamma h_P^{-1} \tfrac{\partial u}{\partial \boldsymbol{n}} \tfrac{\partial v}{\partial \boldsymbol{n}}, \end{aligned} \tag{1.11}$$

with Π_P a suitable projection operator and γ_1 and γ_2 retain the same meaning as in (1.8).

The outline of the thesis is as follows. In Chapter 2 (*Basic principles*), we go over the necessary analytic tools needed to assess the performance of adaptive methods. We cover various elements of approximation theory and smoothness spaces. The primary goal of this section is to familiarize the reader with the underlying theory making some of the analysis more intuitive as proofs of this type can get quite technical.

In Chapter 3 (*Conforming method for the biharmonic equation*), we derive an *a posteriori* error estimator and prove convergence and quasi-optimality of the method, and conclude with the characterization of approximation classes.

In Chapters 4 and 5 (*Wind-driven ocean circulation models*), *a posteriori* error estimators are derived for the linear Stommel-Munk model and the nonlinear SQGE model. In addition, we provide a complete *a priori* analysis for the nonlinear SQGE model. The *a priori* estimate enables us to predict the expected rate of convergence using techniques of approximation theory in Chapter 2.

In Chapter 6 (*Nitsche's method for the biharmonic equation*), a complete analysis analogous to that in Chapter 3 is carried out for the principal part of the models above in the context of Nitsche's discretization.

Chapter 7 (*Conclusion*) closes the thesis with a summary and discussion of the presented research topics, as well as with some suggestions for future research.

Each of Chapters 4 and 5 appeared separately in publications. Chapter 4 is based on [78]:

> Al Balushi, Ibrahim and Jiang, Wen and Tsogtgerel, Gantumur and Kim, Tae-Yeon, *Adaptivity of a B-spline based finite-element method for modeling wind-driven ocean circulation*, Computer Methods in Applied Mechanics and Engineering, Volume 332, pp.1–24, Elsevier, (2018).

Chapter 5 is based on [79]:

> Al Balushi, Ibrahim and Jiang, Wen and Tsogtgerel, Gantumur and Kim, Tae-Yeon, *A posteriori analysis of a B-spline based finite-element method for the stationary quasi-geostrophic equations of the ocean*, Computer Methods in Applied Mechanics and Engineering, Volume 371, pp.113317, Elsevier, (2020).

Moreover, they will include numerical results for benchmark problems using square and L-shaped domains. All mathematical analysis was carried out by the primary first author and all numerical simulations were produced by Wen Jiang.

1.1 Notational conventions

While many notations are summarized in the table on page xv, we would like to highlight some specific ones that appear frequently throughout the thesis. In any case, their definitions appear at the first place where they are introduced.

In this thesis, blackboard bold letters (e.g, $\mathbb{V}$) are used to denote general Banach or Hilbert spaces and capital letters (e.g., T, I and Q) and sometimes the capital Greek letter Π are used to denote operators between those space, and often with subscripts indicating some dependencies.

We will encounter function spaces $\mathcal{C}^k(\Omega)$, $L^p(\Omega)$, $W^s_p(\Omega)$ and $\mathscr{B}^s_q(L^p)$ etc., with Ω being bounded and open domains with a polygonal boundary $\partial\Omega$. Usually we denote the domain boundary by Γ. Elements of those spaces are indicated by lowercase letters (e.g., u and v).

A large portion of the thesis concerns with partitions (or meshes) P of domain Ω consisting of non-overlapping square cells that completely covers Ω. Elements of P will be denoted by the Greek letter τ. It will be useful for us to also consider the edges making up the partitions: $\mathcal{E}_P$, referred to sometime as the skeleton, consist of all the interior edges, and analogously,

$\mathcal{G}_P$ consist all mesh edges along the boundary Γ. Interior and boundary edges will always be denoted by σ. Let h_τ be the diameter of a cell and h_σ be the length of an edge.

In addition to mesh objects, we will encounter numerous piecewise polynomial spline spaces which we mostly denote by $\mathcal{P}_P$, $\mathbb{S}_P$ or $\mathbb{X}_P$. Typically the pieces of these spaces is determined by a partitioning P of domain Ω. The degree of the polynomial spaces will always depend on the letter r interchangeably being the degree or *order*. When r is the order of the space, the degree of the space will be $r - 1$. In the absence of any confusion, we use capital letters (e.g., S, V and W) for elements of polynomial spaces, otherwise we use the Greek letters χ or π.

Finally, in order to avoid the repeated use of generic but unspecified constants, by $f \preceq g$ we mean that $f \leq C{\cdot}g$ with a constant $C > 0$ that may depend only on *fixed* constants under consideration.

Chapter 2

Basic principles

2.1 Preliminaries

In this chapter we briefly summarize all the essential theoretical machinery that drives the analyses of this thesis. The results in this chapter are mostly not new. The only result that has some new elements are Lemma 2.43 and Theorem 2.46.

2.1.1 Sobolev spaces

We recall some elementary definitions and results from (Adams [1] and Grisvard [2]).

Definition 2.1 (Sobolev spaces). For any open subset $\Omega \subset \mathbb{R}^d$, integer $m \geq 0$ and $p \in (0, \infty]$, let

$$W_p^m(\Omega) = \{u \in \mathscr{D}'(\Omega) : \partial^\alpha u \in L^p(\Omega),\ |\alpha| \leq m\}\,, \tag{2.1}$$

with ∂^α understood in the sense of distributions and $\mathscr{D}'(\Omega)$ is the space of distributions. ⊘

In this thesis we consider L^p norms for all positive p values. Integrability range $0 < p < 1$ lacks the attractive topological structure of the $p \geq 1$ counterpart, but their role is indispensable to nonlinear approximation. The functional $\|\cdot\|_{L^p}$ defines a norm for $1 \leq p \leq \infty$ whereas for $0 < p < 1$ it is only a quasi-norm with reversed Minkowski and Holder inequalities. Properties of functions in Sobolev spaces depend strongly on the properties of the domain boundary $\Gamma = \partial\Omega$. For all our intents and purposes we constrain ourselves with open and bounded domains with Lipschitz boundary. All bounded, open and convex subsets Ω of $\mathbb{R}^d$ have Lipschitz boundary including closed polygonal boundaries.

The relationship between different Sobolev spaces represent themselves as embdeddings (of continuous or compact nature) and is entirely dictated by how their smoothness and integrability indices relate. A simple way to express those relationships is via the *Sobolev number* which is defined by $\mathrm{Sob}(W_p^m) = m - \frac{d}{p}$. Higher Sobolev numbers correspond to more regularity. Embedding theorems are initially derived in Fourier space if $\Omega = \mathbb{R}^n$ and extends to open domains Ω with Lipschitz boundary using stable extension operators.

Theorem 2.2 (Calderon extension). *Let $\Omega \subset \mathbb{R}^d$ be bounded and open with a Lipschitz boundary. Then for every integer $m \geq 0$ and every $p \in (1, \infty)$ there exists an extension operator $E \in \mathcal{L}\left(W_p^m(\Omega), W_p^m(\mathbb{R}^d)\right)$ such that $Eu(x) = u(x)$ a.e $x \in \Omega$.*

The embeddings are summarized for integer valued smoothness indices:

Theorem 2.3 (Sobolev Embeddings). *Let $\Omega \subset \mathbb{R}^d$ be open and bounded with Lipschitz boundary. The inclusion*

$$W_q^k(\Omega) \hookrightarrow W_p^m(\Omega), \tag{2.2}$$

is continuous if either $\mathrm{Sob}(W_p^m) \geq \mathrm{Sob}(W_q^k)$*, or, if $m \leq k$ and $p \geq q$ whenever* $\mathrm{Sob}(W_p^m) = \mathrm{Sob}(W_q^k)$*. Moreover, we have*

$$W_p^{m+k}(\Omega) \hookrightarrow \mathcal{C}_b^k(\Omega) = \mathcal{C}^k(\Omega) \cap W_\infty^k(\Omega), \tag{2.3}$$

whenever $\mathrm{Sob}(W_p^m) > 0$*.*

Remark 2.4. A complement to the direction of (2.3) is $\mathcal{C}_b^k(\Omega) \hookrightarrow W_p^k(\Omega)$, but the reverse isn't true. For example in two-dimensions, $\mathrm{Sob}(H^1) = 0$, meaning that the members of $H^1(\Omega)$ are not necessarily continuous. For example $u(r) = \ln(-2 \ln r)$ on the unit open in $\mathbb{R}^2$ centered at the origin. Analogously, from the embedding above we also see that functions in H^2 are continuous but not necessarily C_b^1.

It is of interest to consider non-integer smoothness indices. In particular, fractional order Sobolev spaces are needed when working with boundary value problems since they arise as images of trace operators. Moreover, the performance of approximation methods boils down to measuring smoothness on a continuous scale. The following definition extends the smoothness index to real positive numbers using the Slobodeckij semi-norm (2.4):

Definition 2.5 (Fractional Sobolev spaces). Let $\Omega \subset \mathbb{R}^d$ and let $s = m + \sigma$, $m \in \mathbb{N}$ and $\sigma \in (0, 1)$. Then we define

$$W_p^s(\Omega) = \left\{ u \in W_p^m(\Omega) : \frac{\partial^\alpha u(x) - \partial^\alpha u(y)}{|x-y|^{\sigma + d/p}} \in L^p(\Omega \times \Omega),\ \ |\alpha| \leq m,\ \alpha \in \mathbb{N}^d \right\}. \tag{2.4}$$

$\oslash$

Definition 2.5 extends m to real values by requiring the difference $D^m u(r^n)$ to decay at a rate $+\sigma p$ units faster than just the $\mathcal{O}(r^n)$ required by the standard definition of differentiation. An alternative way to Definition 2.5 is to carry fractional differentiation in Fourier space which is relatively easy. However, this only applies to unbounded or to periodic domains. In the context of this thesis we will need to define fractional order Sobolev spaces as interpolation spaces.

We conclude this section with a trace theorem:

Theorem 2.6 (Sobolev Trace). *Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain with boundary Γ, and let $s > 1/2$. Then there exists a unique trace mapping $\cdot|_\Gamma \in \mathcal{L}(H^s(\Omega), H^{s-1/2}(\Gamma))$.*

A more useful form is:

Theorem 2.7 (General Trace Inequlaity). *Let $\Omega \subset \mathbb{R}^d$ be a bounded and open domain with Lipschitz boundary Γ. Then there exists a constant $C_T > 0$ such that*

$$\|u\|_{L^p(\Gamma)} \leq C_T\Bigg(\varepsilon^{1-1/p}\|\nabla u\|_{L^p(\Omega)} + \varepsilon^{-1/p}\|u\|_{L^p(\Omega)}\Bigg), \tag{2.5}$$

for all $u \in W^1_p(\Omega)$ and $\varepsilon \in (0,1)$.

2.1.2 (θ, q)-quasi-norms and sequence spaces

The quasi-norms are a tool for quantifying the rate of decay of sequences with great precision, making them instrumental in creating different sub-spaces within a larger topological space by how well they can be approximated by smoother objects.

Definition 2.8 ($(\boldsymbol{\theta}, \boldsymbol{q})$-quasi-norms). For $0 < \theta < 1$ and $0 < q \leq \infty$, the following functional

$$\|w\|_{\theta,q} = \begin{cases} \left(\displaystyle\int_0^\infty [t^{-\theta}w(t)]^q \tfrac{dt}{t}\right)^{1/q}, & q < \infty, \\ \text{ess sup}_{t>0}(t^{-\theta}w(t)), & q = \infty, \end{cases} \tag{2.6}$$

defines a quasi-norm for all non-negative real-valued Lebesque measurable functions $w : \mathbb{R}_+ \to \mathbb{R}_+$. ⊘

When the (θ, q)-quasi-norm is finite, we can say something about the behavior of $w(t)$ as $t \to 0$ and as $t \to \infty$. When θ is closer to 1, the function w is expected to decay to zero faster at the origin.

Remark 2.9. Taking $w(t) = t^\alpha$, for $\alpha > 0$, the (θ, q)-quasi-norm is finite if and only if $\alpha > \theta$ irrespective of parameter q making it a secondary parameter in comparison to θ. In this way parameter q creates finer spaces for each θ value.

In what follows, the function $w(t)$ will actually be a functional $w(u;t)$, parameterized by $t > 0$, that measures certain behaviors of a target function u belonging to some function class X. In this way the (θ, q)-quasi-norm can separate X into various nested sub-spaces $X_{\theta,q}$ determined by the measurements carried by $w(f;t)$. If $X_{\theta,q}$ denotes the space of all non-negative functions for which the quasi-norm (2.6) is finite, we will have the embedding:

$$X_{\alpha,q} \subset X_{\beta,p} \quad \text{if } \alpha > \beta \text{ regardless of } q\text{, or, if } \alpha = \beta \text{ and } q < p. \tag{2.7}$$

We consider three relevant functionals $w(f;t)$.

Definition 2.10 (Moduli of smoothness). For $\Omega \subset \mathbb{R}^d$ open, $u \in L^p(\Omega)$, $p \in (0, \infty]$, we denote the modulus of smoothness of order $r \geq 1$ of u by

$$\omega_r(u,t)_p = \sup_{|h| \leq t} \|\Delta_h^r(u, \cdot)\|_{L^p(\Omega_{r,h})}, \quad \Omega_{r,h} = \{x \mid (x, x + rh) \subset \Omega\}, \tag{2.8}$$

where Δ_h^r is the rth order difference with step $h \in \mathbb{R}^d$. ⊘

Some properties follow immediately from the definition of $\omega_r(u,t)_p$. It is clear that $\omega_r(u,t)_p \to 0$ monotonically as $t \to 0$. We expect $\omega_r(u,t)_p \to 0$ faster for smoother functions u. In particular, if $u \in W_p^k(\Omega)$, with $1 \leq p < \infty$, then $\omega_k(u,t)_p \leq t^k |u|_{W_p^k(\Omega)}$.

Definition (2.8) measures smoothness without requiring the existence of weak derivatives, but as long as its rate of decay of $\omega_r(f,t)_p \to 0$ as $t \to 0$ is fast enough, one can infer additional smoothness. Precisely, if $t^k \omega_r(f,t)_p \to 0$, $0 \leq k < r$ then $f \in W_p^k(\Omega)$. We also have a useful result due to (Graham [13]: if $u(x) = |x|^{\alpha - d/p}$, $x \in \mathbb{R}^d$, $1 \leq p \leq \infty$ then

$$\omega_r(u,t)_p = \begin{cases} \mathcal{O}(t^\alpha), & \alpha > r, \\ \mathcal{O}(t^r), & \alpha < r. \end{cases} \tag{2.9}$$

Remark 2.11. If $t^{-r}\omega_r(u,t)_p \to 0$ as $t \to 0$ then u is necessarily an $r-1$ degree polynomial. In essence this means there is a limit to how well approximation can be with respect to a polynomial spaces and results of this kind are said to be *saturation* results and they extends to more complicated settings such as polynomial spline spaces.

A modified modulus is needed when we are in a polynomial spline setting since the addition of several sub-intervals is expected. We define the *averaged* modulus of smoothness to be

$$w_r(u,t)_p^p = \int_{[-t,t]^d} \int_{\Omega_{r,t}} |\Delta_s^r(u,x)|^p \, dxds, \tag{2.10}$$

which is equivalent to (2.8) with proportionality constants depending on r, p and d.

The second form the functional $w(f;t)$ takes is the so-called K-functional which concerns Peetre's real interpolation of spaces.

Definition 2.12 (K-functional). Let $\mathbb{V}$ and $\mathbb{W}$ be quasi-normed spaces such that the embedding $\mathbb{W} \hookrightarrow \mathbb{V}$ is continuous. Let $t > 0$ and $u \in \mathbb{V}$. Then we define

$$K(u,t) = K(u,t;\mathbb{V},\mathbb{W}) = \inf_{w \in \mathbb{W}} \left(\|u - w\|_{\mathbb{V}} + t|w|_{\mathbb{W}} \right). \tag{2.11}$$

$\oslash$

Topological considerations of the spaces $\mathbb{V}$ and $\mathbb{W}$ can be tightened or relaxed, but here the choice is such that the treatment of Sobolev spaces with $0 < p < 1$ is possible. The functional $K(u,t)$ is increasing, concave (therefore monotone) and continuous on $t \geq 0$. The definition says that having $K(u,t) < \varepsilon$ for some $t > 0$ implies that $u \in \mathbb{V}$ can be approximated with error $\|u - w\|_{\mathbb{V}} < \varepsilon$ by some $w \in \mathbb{W}$ with a reasonably sized $|w|_{\mathbb{W}} < \varepsilon t^{-1}$ illustrating that the K-functional can quantify the smoothness of u by how well u can be approximated by smoother objects. In fact, (Johnen and Scherer [10]) show that for any $p \in (0, \infty]$ and integer $r \geq 1$,

$$\omega_r(u,t)_p \sim K(u,t;L^p(\Omega), W_p^r(\Omega)) \quad \forall u \in L^p(\Omega),\ \forall t > 0, \tag{2.12}$$

meaning that if $u \in L^p$ is well-approximated by functions with r-order weak derivatives then the smoothness modulus decay's reasonably fast and that u has to be somewhere in between L^p and W_p^r.

In general, when $w(f;t)$ is monotone, discretization of the (θ,q)-quasi-norms is possible with $\|w\|_{\theta,q} \sim \|(2^{\theta k} w(2^{-k}))_{k\in\mathbb{Z}}\|_{\ell^q}$ relating $\|\cdot\|_{\theta,q}$ with Lorentz sequence spaces.

Definition 2.13 (Lorentz sequence spaces). Let $(a_k)_{k\in\mathbb{Z}}$ be a positive sequence and let $\alpha > 0$,

$$\|(a_k)\|_{\ell_q^\alpha} = \|(2^{\alpha k} a_k)\|_{\ell^q} = \begin{cases} \left(\sum_{k\in\mathbb{Z}} \left(2^{k\alpha} a_k\right)^q \right)^{1/q}, & 0 < q < \infty, \\ \sup_{k\in\mathbb{Z}} 2^{k\alpha} a_k, & q = \infty. \end{cases} \tag{2.13}$$

$\oslash$

The last choice for $w(f;t)$ concerns approximation error.

Definition 2.14 (Error functional). Let $\mathbb{V}$ be a quasi-normed space and let $\mathbb{X}$ be a finite dimensional subspace of $\mathbb{V}$. Then we define

$$E(u,\mathbb{X})_{\mathbb{V}} = \inf_{\chi\in\mathbb{X}} \|u-\chi\|_{\mathbb{V}}. \tag{2.14}$$

$\oslash$

Remark 2.15. If a sequence $(a_n)_{n\in\mathbb{N}}$ is monotone and decreasing, we have $\|(a_{2^{\alpha n}})_{n\in\mathbb{N}}\|_{\ell_q^\alpha}^q \sim \sum_{n\in\mathbb{N}}[n^\alpha a_n]^q\frac{1}{n}$. It is clear that the inclusion $\ell_{q_1}^\alpha \subset \ell_{q_2}^\alpha$ holds whenever $q_2 \leq q_1$. As an application, if $a_n = E(u,\mathbb{X}_n)_{\mathbb{V}}$ for some sequence of approximating spaces $\{\mathbb{X}_n\}_{n\geq 1}$, then $\|(a_n)\|_{\ell_q^s} < \infty$ is equivalent to $E(u,\mathbb{X}_n)_{\mathbb{V}} \preceq n^{-s}$ for any q.

2.1.3 Interpolation spaces

Definition 2.16 (Interpolation spaces). Let $\mathbb{V}$ and $\mathbb{W}$ be quasi-normed spaces such that the embedding $\mathbb{W} \hookrightarrow \mathbb{V}$ is continuous. For $0<\theta<1$ and $0<q\leq\infty$, we let

$$(\mathbb{V},\mathbb{W})_{\theta,q} = \left\{u\in\mathbb{V} \mid |u|_{(\mathbb{V},\mathbb{W})_{\theta,q}} < \infty\right\}, \tag{2.15}$$

where $|u|_{(\mathbb{V},\mathbb{W})_{\theta,q}} = \|K(u,\cdot\,;\mathbb{V},\mathbb{W})\|_{\theta,q}$. $\oslash$

The spaces $X_{\theta,q} = (\mathbb{V},\mathbb{W})_{\theta,q}$ are topologically quasi-normed and will clearly satisfy the embeddings (2.7).

The following theorem asserts that if one interpolates between two interpolation spaces, one does not obtain anything new:

Theorem 2.17 (Reiteration Theorem). *Let* $\mathbb{V}' = (\mathbb{V},\mathbb{W})_{\alpha_1,q_1}$ *and let* $\mathbb{W}' = (\mathbb{V},\mathbb{W})_{\alpha_2,q_2}$ *with* $\alpha_1<\alpha_2$. *Then for any* $0<\theta<1$ *and* $0<q\leq\infty$,

$$(\mathbb{V}',\mathbb{W}')_{\theta,q} = (\mathbb{V},\mathbb{W})_{\alpha,q}, \quad \alpha = (1-\theta)\alpha_1+\theta\alpha_2. \tag{2.16}$$

Interpolation spaces serve a number of relevant applications. First of which is the definition of fractional Sobolev spaces thereby extending Definition 2.1 to continuous smoothness parameter values without resorting to the unpopular Slobodeckij semi-norm (2.4): For $k\in\mathbb{N}$, $0<\sigma<1$ and domain Ω with Lipschitz boundary, the interpolation space

$$W_p^{k+\sigma}(\Omega) = \left(W_p^k(\Omega), W_p^{k+1}(\Omega)\right)_{\sigma,p}, \tag{2.17}$$

has a norm equivalent to that of Definition 2.5. More generally, we define a generalization Sobolev to *Besov Spaces* by interpolation as well:

$$\mathscr{B}_q^\alpha(L^p(\Omega)) = (L^p(\Omega), W_p^r(\Omega))_{\alpha/r,q}, \quad 0<\alpha<r,\ 0<q\leq\infty, \tag{2.18}$$

where $X_{\theta,q} = \mathscr{B}_q^\theta(L^p(\Omega))$ inherits all the (θ,q)-embeddings of (2.7). Moreover, in view of the equivalence (2.12) one obtains an equivalent definition for $\mathscr{B}_q^s(L^p(\Omega))$ using the modulus of smoothness (2.8). We discuss Besov in more detail in the next subsection.

2.1.4 Besov spaces

Definition 2.18 (Besov Spaces). Let $\Omega \subset \mathbb{R}^d$ and α be a positive real number. Let $0 < p, q \leq \infty$, r be any positive integer greater than α. Then we have

$$\mathscr{B}^{\alpha}_{q;r}(L_p(\Omega)) = \left\{ u \in L^p(\Omega) \mid |u|_{\mathscr{B}^{\alpha}_{p,q;r}} < \infty \right\}, \tag{2.19}$$

where $|u|_{\mathscr{B}^{\alpha}_{p,q;r}} = \|\omega_r(u,\cdot)_p\|_{\alpha,q}$. ⊘

Remark 2.19. Recall that $\|\cdot\|_{\mathscr{B}^{\alpha}_{p,q;r}} = \|\cdot\|_{L^p} + |\cdot|_{\mathscr{B}^{\alpha}_{p,q;r}}$ is a quasi-norm for $p < 1$ and a norm otherwise. If $\alpha > r - 1 + \max\{1, \frac{1}{p}\}$ then the space is just the polynomial space $\mathbb{P}_{r-1}$ making the definition (2.19) trivial. Indeed, it can be shown that $\omega_r(f,t)_p = \mathcal{O}(t^{r-1+\max\{1,\frac{1}{p}\}})$ and the rest follows from Remark 2.11. Instead, the r should be chosen strictly greater than $\alpha + 1 - \max\{1, \frac{1}{p}\}$. With such r in hand, any positive integer $r' > r$ will yield a quasi-norm $\mathscr{B}^{\alpha}_{p,q;r'}$ equivalent to $\mathscr{B}^{\alpha}_{p,q;r}$. This Besov space independence of r justifies dropping the r from the definition's symbol. See (DeVore and Lorentz [6]) for details. When the relation is equality, there will be dependence on q (Tsogtgerel [22]). We do not discuss this.

Reiteration Theorem 2.17 asserts if one interpolates two fractional Sobolev spaces, the result is another fractional order Sobolev space with the order resulting from the convex combination of (2.16). The interpolation of Besov spaces is also a Besov space (DeVore and Popov [8]):

Theorem 2.20 (DeVore and Popov [8]). *If $\alpha_i \in (0,\infty)$ and $p_i, q_i \in (0,\infty]$ for $i = 0,1$, then for each $\theta \in (0,1)$, and for $\frac{1}{q} = \theta\frac{1}{q_0} + (1-\theta)\frac{1}{q_1}$ and $\frac{1}{p} = \theta\frac{1}{p_0} + (1-\theta)\frac{1}{p_1}$, we have*

$$\left(\mathscr{B}^{\alpha_0}_{q_0}(L_{p_0}), \mathscr{B}^{\alpha_1}_{q_1}(L_{p_1})\right)_{\theta,q} = \mathscr{B}^{\alpha}_{q}(L_p) \quad \text{with } \alpha = (1-\theta)\alpha_0 + \theta\alpha_1. \tag{2.20}$$

There are some inconsistencies with Besov and Sobolev spaces; the former is not a strict extension of the latter. For example for $p \geq 1$ and $s > 0$, the inclusion $\mathscr{B}^{s}_{q}(L^p) \subset W^s_p$ is strict when $q > 2$, and is flipped when $q < 2$. In the special case of $q = p = 2$ the spaces coincide with equivalent norms for all $s > 0$.

The embeddings of Besov spaces from (2.7) immediately give $\mathscr{B}^{\alpha}_{q_1}(L^p) \subset \mathscr{B}^{\beta}_{q_2}(L^p)$ whenever $\alpha > \beta$ and $0 < q_1, q_2 \leq \infty$, and $\mathscr{B}^{\alpha}_{q_1}(L^p) \subset \mathscr{B}^{\alpha}_{q_2}(L^p)$ whenever $q_1 < q_2$. In addition to the inherited embeddings of Besov spaces from the (θ, q)-quasi-norm structure, we have $\mathscr{B}^{\alpha}_{p_1}(L^{p_1}) \subset \mathscr{B}^{\alpha}_{p_2}(L^{p_2})$ whenever $p_1 < p_2$. More interestingly, we have the continuous embedding

$$\mathscr{B}^{\alpha_1}_{q_1}(L^{p_1}) \subset \mathscr{B}^{\alpha_2}_{q_2}(L^{p_2}), \tag{2.21}$$

whenever $\alpha_1 - \alpha_2 \geq d(\frac{1}{p_1} - \frac{1}{p_2}) > 0$ given any pair $0 < q_1, q_2 \leq \infty$, or, $\alpha_1 - \alpha_2 = d(\frac{1}{p_1} - \frac{1}{p_2}) > 0$ whenever $0 < q_2 \leq q_1 \leq \infty$. In particular, the embedding (2.21) is consistent with that of

Sobolev embedding (2.2) thereby extending the latter relation to fractional order spaces when $q_i = p_i$. The relation between $\mathscr{B}^s_q(L^p)$ spaces, irrespective of the tertiary index q, is best illustrated by the so-called DeVore diagram. Each point $(\frac{1}{p}, s)$ on the diagram indicates the position of Besov space $\mathscr{B}^s_q(L^p)$, where the Sobolev embedding line emanates in the position direction from $(\frac{1}{p}, s)$ is indicated in thick has slope d. In Figure 2.1 where we use $\mathscr{B}^\alpha_2(L^2) = H^2(\Omega)$ an example which corresponds to point $(\frac{1}{p}, s) = (\frac{1}{2}, 2)$.

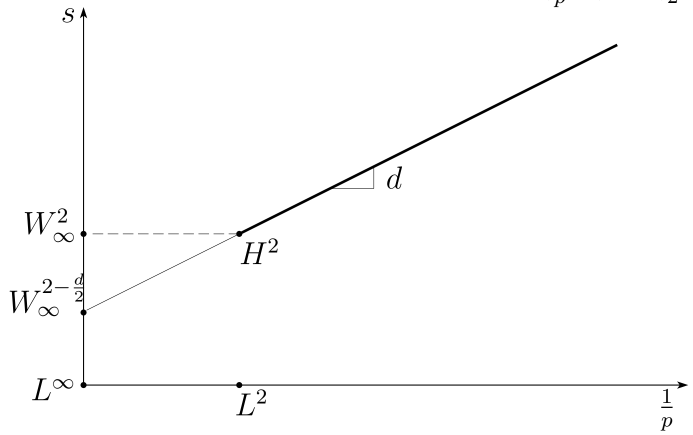


Figure 2.1: *In this so-called DeVore diagram the point* $(\frac{1}{p}, s)$ *represents the Besov space* $\mathscr{B}^s_q(L^p)$*, irrespective of* $0 < q \leq \infty$*. Specific to our interest when p = q, when Besov spaces are equivalent to fractional Sobolev spaces, the thick line has slope d and corresponds to the Sobolev embedding line* $\mathrm{Sob}(W^s_p) = \mathrm{Sob}(H^2)$ *for which* $p \leq 2$ *and* $s \geq 2$*. All spaces on this line and above it and those on and above the demarcated line embeds into* H^2*.*

2.1.5 Approximation classes

There are various types of adaptive approximation procedure in the literature, all with the mission to increase approximation resolution through focusing computational resources to where it is most needed. Adaptivity through mesh refinement requires a suitable class of approximating finite dimensional spaces, an initial mesh configuration, a subdivision strategy that preserves desirable mesh characteristics and derivable computable quantities serving the role of local approximation errors to guide refinement. These so-called guides are *error indicators* usually defined locally, with *error estimators* as the global counterpart. The performance of any adaptive procedure needs to be determined, and because such procedure is generally more difficult to implement than conventional ones, it is always necessary to know whether the effort is worthwhile.

Assume that $\{\mathbb{X}_n\}_{n\geq 1}$ is a "suitable" sequence or family of finite-dimensional approximating spaces for $\mathbb{V}$. The meaning of "suitable" will be made precise later. We can classify all functions $u \in \mathbb{V}$ by how well they can be approximated by the family $\{\mathbb{X}_n\}_{n\geq 1}$, that is, by the decay rate of approximation error.

Definition 2.21 (Approximation class). For $0 < q \leq \infty$ and $s > 0$, let

$$\mathscr{A}_q^s = \mathscr{A}_q^s(\mathbb{V}, \{\mathbb{X}_n\}_{n\geq 1}) = \left\{ u \in \mathbb{V} \mid |u|_{\mathscr{A}_q^s} < \infty \right\}, \tag{2.22}$$

where $|u|_{\mathscr{A}_q^s} = \|(E(u, \mathbb{X}_n)_{\mathbb{V}})_{n\geq 1}\|_{\ell_q^s}$. $\oslash$

The embeddings in (2.7) hold true with $X_{\theta,q} = \mathscr{A}_q^\theta$. The $q = \infty$ case, which merits its own symbol $\mathscr{A}^s = \mathscr{A}_\infty^s$, is of primary interest since it is the largest approximation class where functions are approximated with rate s. Fortunately, it is also the simplest of all the possible $q's$. This is consistent with Remark 2.15.

The following result is instrumental in expressing smoothness in terms of how well it can be approximated:

Theorem 2.22 (Hardy's inequality). *If $(a_k)_{k\in\mathbb{Z}}$ and $(b_k)_{k\in\mathbb{Z}}$ are positive sequences such that for some $\mu, \lambda > 0$,*

$$b_k \preceq 2^{-k\lambda}\left(\sum_{j=-\infty}^{k} \left(2^{j\lambda} a_j\right)^{\mu}\right)^{1/\mu}, \tag{2.23}$$

then $\|(b_k)\|_{\ell_q^\theta} \preceq \|(a_k)\|_{\ell_q^\theta}$ for all $0 < q \leq \infty$ provided that $0 < \theta < \lambda$.

2.2 Approximation theory

The subject of approximation theory is quite technical and wields a sophisticated theoretical machinery. A brief survey in a concrete setting is developed here for the unfamiliar reader. The purpose of this section is to provide the reader with a clear idea of how one can anticipate the upper efficiency limits of an approximating algorithm in terms of the smoothness of a target function f when f itself is accessible. When solving for a numerical solution for a differential equation, one does not have the target function in hand making the analysis harder; a thorough study of solution regularity is needed. Once the solution regularity is known, the rest follows naturally as in the former case up to some minor additional technicalities. In numerical differential equations, as demonstrated in (Binev et al. [59], Stevenson [60] and Cascón et al. [61]) the right-hand side and boundary conditions can interfere with the rate because of *data oscillation*; (see Chapters 3 and 6).

2.2.1 Fundamental ideas through examples

Example 2.23. Here we approximate a target function $f \in C^\infty(I) \cap W_p^1(I)$ with $1 < p < \infty$ and $I = (0,1)$ using piecewise constant functions with error measured in L^p. Consider a grid $\mathcal{G}_N$ given by breakpoints $0 = x_0 < x_1 < \cdots < x_k < \cdots < x_N = 1$ giving rise to N subintervals $J_k = (x_{k-1}, x_k)$. In the absence of confusion we drop subscripts k and N. We can approximate the target function f with a $f_I = \sum_{J \in \mathcal{G}} c_J \mathbb{1}_J$ with $c_J \in \mathbb{R}$ taken to be f at the midpoint of J and a local point-wise error estimation reads

$$|f(x) - c_J| \leq \int_J |f'(t)|\, dt \leq |J|^{1-1/p} |f|_{W_p^1(J)} \quad \forall x \in J, \tag{2.24}$$

from which a local error estimate reads

$$\|f - f_I\|_{L^p}^p \leq |J|^p |f|_{W_p^1(J)}^p, \tag{2.25}$$

and summing over all subintervals to obtain a global error estimate:

$$\|f - f_I\|_{L^p(I)} = \left(\sum_{J \in \mathcal{G}} \|f - c_J\|_{L^p(J)}^p \right)^{1/p} \leq \left(\sum_{J \in \mathcal{G}} |J|^p |f|_{W_p^1(J)}^p \right)^{1/p}. \tag{2.26}$$

A uniform dyadic partition of level $n \geq 1$ is generated by setting $x_k = 2^{-k}$, for $k = 0, ..., 2^n$ giving rise to a dyadic partition $\mathcal{G}$ consisting of $N = 2^n$ sub-intervals J all with $|J| = 2^{-n}$, and the total error converges with

$$\|f - f_I\|_{L^p(I)} \leq 2^{-n} \left(\sum_{J \in \mathcal{G}} |f|_{W_p^1(J)}^p \right)^{1/p} = N^{-1} |f|_{W_p^1(I)}. \tag{2.27}$$

In fact, one can show that if f_I is any near-best piecewise constant approximation in L^p for a sufficiently smooth function f, we will expect error convergence with rate $\mathcal{O}(N^{-1})$. ⊘

There is no lack of abundance in literature extending the estimation idea above to multi-dimensions, higher-order polynomial approximation or to target functions belonging to more general Sobolev spaces. In what follows we summarize the main results and ideas, and give credit to those concerned. If $\mathbb{S} = \mathbb{S}(\Omega)$ any polynomial space on Ω a Euclidean domain, the best-error is given by

$$E(f, \mathbb{S})_p = \inf_{S \in \mathbb{S}} \|f - S\|_{L^p(\Omega)}, \quad f \in L^p(\Omega). \tag{2.28}$$

If $\mathcal{Q}$ a cube in d-dimensions, $\mathcal{Q}$ with side length $l_\mathcal{Q}$ and $\mathcal{P}_{r-1}$ is polynomial spline space of *order* r; that is, polynomial of degree $r-1$, the classical Whitney's theorem (Whitney [9]) asserts that

$$E(f, \mathcal{P}_{r-1})_p \preceq \omega_r(f, l_\mathcal{Q})_p \quad \forall f \in L^p(\mathcal{Q}), \tag{2.29}$$

where $\preceq$ depends on order r and dimension d. The transference of estimate (2.29) to polynomial spline setting is done by the advent of quasi-interpolation operators with effort owed to (De Boor, Carl and Fix [11]); see also (Lyche, Tom and Schumaker [12]) for explicit constructions.

More recently, (DeVore and Popov [8]) establish that if $\mathcal{G}_n$ is a uniform dyadic partition on a d-dimensional unit cube $\mathcal{Q}$ for which sub-cubes have length 2^{-n} which defines $\mathbb{S}_n^r = \mathbb{S}_r(\mathcal{G}_n)$ a polynomial spline space of degree $r-1$ with $r-2$ continuous derivatives and bounded $r-1$ derivatives. Then for every $f \in L^p(\mathcal{Q})$,

$$E(f, \mathbb{S}_n^r)_p \preceq \omega_r(f, 2^{-n})_p. \tag{2.30}$$

Furthermore, noting that the number of subcubes making up $\mathcal{Q}$ is $N = 2^{nd}$, and if $f \in W_p^r(\mathcal{Q})$ then $\omega_r(f, 2^{-n})_p \leq (2^{-n})^r |f|_{W_p^r(\mathcal{Q})}$ making for an approximation rate of $\mathcal{O}(N^{-r/d})$.

The direct estimates above (Eqs. (2.30) and (2.29)) carry over in substance from $p \geq 1$ to $0 < p < 1$, but at the cost of some technical issues which were resolved in (Storozhenko and Oswald [14]). As metioned previously, the importance of the range $0 < p < 1$ is pronounced for nonlinear approximation methods.

It is sometimes possible to determine the smoothness of a function by how well it can be approximated. Assume that $p > 0$, if $\mathbb{S}_n^r$ retains its meaning and $\lambda = \min(r, r-1+1/p)$, then for every $f \in L^p(\mathcal{Q})$ and $\mu \leq \min(1, p)$,

$$\omega_r(f, 2^{-n})_p \preceq 2^{-n\lambda} \left(\sum_{k=0}^{n} 2^{k\lambda\mu} E(f, \mathbb{S}_k^r)_p^\mu \right)^{1/\mu}. \tag{2.31}$$

In the following example we illustrate this idea which requires Hardy's inequality of Theorem 2.22.

Example 2.24. If $f \in L^p(\mathcal{Q})$, with $p > 1$, can be approximated with $E(f, \mathbb{S}_n^r)_p = \mathcal{O}(N^{-s/d})$, N being the number of subcubes, and $s < \lambda$ then in view of Hardy's inequality (2.23) with $a_k = \omega_r(f, 2^{-k})_p$ and $b_k = E(f, \mathbb{S}_k^r)_p$ and indices $\mu = 1$, $q = \infty$ and $\theta = s$ we have making $(\omega_r(f, 2^{-k})_p)_k \in \ell_\infty^s$ and therefore $f \in \mathscr{B}_\infty^s(L^p)$ in view of definitions (2.19). If on the other hand the approximation rate exceeds r/d, then it would mean $\omega_r(f,t)_p = o(t^r)$ which necessarily means $f \in \mathbb{S}_n^r$. The first saturation result in $L^p(0,1)$ space is owed to (Butler and Richarts [15]) Applying inverse estimates such as (2.31) usually follows this aforementioned manner. Direct and inverse estimates provide the following characterization: For $0 < p \leq \infty$, provided $0 < s < \lambda$ we have for all $0 < q \leq \infty$

$$f \in L^p(\mathcal{Q}) \ s.t \ \|\{E(f, \mathbb{S}_k^r)_p\}_{k\geq 0}\|_{\ell_q^s} < \infty \iff f \in \mathscr{B}_q^s(L^p). \tag{2.32}$$

$\oslash$

(Ciesielski [16]) established the inverse estimate (2.31) in univariate $p \geq 1$ but later it was extended to all integration indices $0 < p \leq \infty$ in multi-dimensions by (DeVore and Popov [8]).

Example 2.25. The radial function $f(x) = |x|^\gamma$ with $\gamma \in (0,1)$ acts as a prototype of singularities in the derivatives arising in solutions to differential equations. For simplicity will discuss approximation power of splines in the context of the univariate case on the unit open interval I using piecewise constants splines with error measured in L^p with $p \geq 1$. When $\gamma > 1 - 1/p$, relation (2.27) holds immediately since $x^\gamma \in W^1_p(I)$. On the other hand if $\gamma \leq 1 - \frac{1}{p}$, the target function $x^\gamma \in W^1_\sigma(I)$, for any $\sigma < \frac{1}{1-\gamma}$ which would also be less than p, unsurprisingly since p-integrability is too strong. In this specific example taking $\sigma = 1$ is possible making $f \in W^1_1(I)$. Because $\sigma, p \geq 1$ hold, Holder applies and we arrive at an estimate for the global error

$$\begin{aligned} \|f - f_I\|^p_{L^p(I)} &\leq \sum_{J \in \mathcal{G}} |J|^{p-p/\sigma+1} \|f'\|^p_{L^\sigma(J)}, \\ &= (2^{-n})^{p-p/\sigma+1} \sum_{J \in G} \|f'\|^p_{L^\sigma(J)} \leq (N^{-1})^{p-p/\sigma+1} \|f'\|^p_{L^\sigma(I)}, \end{aligned} \tag{2.34}$$

after employment of uniform dyadic partition. In the last step we used $(\sum_{k \leq N} a_k^p)^{1/p} \leq (\sum_{k \leq N} a_k^\sigma)^{1/\sigma}$ since $\sigma < p$. We conclude that we have convergence of error $\|f - f_I\|_{L^p(I)}$ with rate $N^{-1+1/\sigma-1/p}$. The best possible rate is given by the largest possible $\sigma < p$ which is still suboptimal. For example consider $x^{1/3}$ and error measured in L^2 imposes the restriction that $\sigma < 3/2$.

Using predetermined partitions may limit the performance. We turn our attention to free partitions that depends on f and show that we can recover any lost convergence order. The idea is to optimize the interaction between the norm quantity of f local error $|J|^{p-p/\sigma+1}|f|^p_{W^1_\sigma(J)}$ in (2.34). The quantity σ will be crucial, but for now we will take it to be any $\sigma < \frac{1}{1-\gamma}$. We choose a partition $\mathcal{G}$ such that the global quantity $|f|_{W^1_\sigma(I)}$ is equally distributed over all sub-intervals, i.e find $J \in \mathcal{G}$ such that $|f|^\sigma_{W^1_\sigma(J)} \leq N^{-1}|f|^\sigma_{W^1_\sigma(I)}$ which makes

$$\|f - f_I\|_{L^p(J)} \preceq |f|_{W^1_\sigma(J)} \leq N^{-1/\sigma}|f|_{W^1_\sigma(I)}, \tag{2.35}$$

then, after summing

$$\|f - f_I\|^p_{L^p(I)} \preceq \sum_{J \in \mathcal{G}} N^{-p/\sigma}|f|^p_{W^1_\sigma(I)} = N^{-p/\sigma+1}|f|^p_{W^1_\sigma(I)}. \tag{2.36}$$

It turns out that if $\sigma = (1 + 1/p)^{-1}$, which is both $< p$ and $< \frac{1}{1-\gamma}$, we obtain the optimal convergence rate proportional to $N^{-p/\sigma+1} = N^{-1}$. For such σ, the fractional Sobolev space

/ Besov space $W^1_\sigma(I)$ would be the largest space with smoothness 1 that is embedding in $L^p(I)$. In other words, we can recover lost convergence order by constructing partitions that equidistributes the largest Besov semi-norm of f embedded in L^p. $\oslash$

The largest space $\mathscr{B}^s_\sigma(L^\sigma)$ continuously embedded in L^p is obtained by space interpolation:

Theorem 2.26 (DeVore and Popov [8]). *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain, let $0 < p < \infty$ and let $\alpha > 0$. If $\sigma = (\frac{\alpha}{d} + \frac{1}{p})^{-1}$ then the Besov space $\mathscr{B}^\alpha_\sigma(L^\sigma(\Omega))$ is continuously embedded in $L^p(\Omega)$.*

Remark 2.27. Actually, the statement given in [8] concerns $\mathscr{B}^\alpha_p(L^\sigma(\Omega))$ with $\sigma < p$. It is of more interest to us to have $\mathscr{B}^\alpha_\sigma(L^\sigma(\Omega))$, which is valid because of (2.7), as $\mathscr{B}^\alpha_\sigma(L^\sigma(\Omega))$ corresponds to fractional Sobolev spaces.

2.2.2 Characterization of approximation classes

We move on to a more general setting. Let $\mathbb{V}$ be a quasi-normed space and let $\{\mathbb{X}_n\}_{n\geq 1}$ be a sequence of finite dimensional subspaces of $\mathbb{V}$ which serve as approximating spaces to $\mathbb{V}$. The index n can reflect the dimension number of $\mathbb{X}$ or a related quantity with a one-to-one relationship with the dimension of $\mathbb{X}_n$. The approximation takes place in the topology of $\mathbb{V}$. We define the approximation functional on $\mathbb{V}$

$$E(u, \mathbb{X}_n)_{\mathbb{V}} = \inf_{\chi \in \mathbb{X}_n} \|u - \chi\|_{\mathbb{V}}, \quad (u \in \mathbb{V}). \tag{2.37}$$

The finite dimensional subspaces $\{\mathbb{X}_n\}$ are assumed to be nested and whose union $\cup_{n\geq 1}\mathbb{X}_n$ must be dense in $\mathbb{V}$. Furthermore, $\mathbb{X}_n$ is assumed to have a near-best approximant S for any target function $u \in \mathbb{V}$ with respect to the functional $E(u, \mathbb{X}_n)_{\mathbb{V}}$. A final condition, albeit of strong relevance to this study is that $\mathbb{X}_n + \mathbb{X}_n \subset \mathbb{X}_{cn}$ for some $c \geq 1$. If $c = 1$ then process is said to be a *linear approximation* and if not it is said to be a *nonlinear approximation*. In adaptivity, approximation will always be nonlinear; there are numerous mesh configurations that produce two different approximating spaces with the same number of degrees of freedom.

Let $\mathbb{W}$ be a linear space equipped with semi-quasi-norm $|\cdot|_{\mathbb{W}}$, with μ that recovers triangle inequality ($|f+g|^\mu_{\mathbb{W}} \leq |f|^\mu_{\mathbb{W}} + |g|^\mu_{\mathbb{W}}$), such that the embedding $\mathbb{W} \hookrightarrow \mathbb{V}$ is continuous.

Definition 2.28 (Jackson and Bernstein inequalities). Let $\mathbb{V}$, $\mathbb{W}$ and $\{\mathbb{X}_n\}$ be as above. It is said that the *Jackson inequality* holds for a real number $s > 0$ and a constant $C > 0$ if

$$E(u, \mathbb{X}_n)_{\mathbb{V}} \leq C n^{-s} |u|_{\mathbb{W}} \quad \forall u \in \mathbb{W}, \quad (n \in \mathbb{N}). \tag{2.38}$$

Moreover, it is said that the *Bernstein inequality* holds for a real number $s > 0$ and a constant $C > 0$ if

$$|\chi|_{\mathbb{W}} \leq C n^s \|\chi\|_{\mathbb{V}}, \quad \forall \chi \in \mathbb{X}_n, \quad (n \in \mathbb{N}). \tag{2.39}$$

$\oslash$

If the Jackson inequality holds,

$$E(u, \mathbb{X}_n)_{\mathbb{V}} \leq CK(u, n^{-s}; \mathbb{V}, \mathbb{W}), \tag{2.40}$$

which is exactly the direct estimate (2.30) for dyadic splines when the error is measured in L^p. On the other hand if the Bernstein inequality holds,

$$K(u, n^{-s}; \mathbb{V}, \mathbb{W}) \leq Cn^{-s}\left(\sum_{k=1}^{n}[k^s E(u, \mathbb{X}_n)_{\mathbb{V}}]^{\mu}\frac{1}{n}\right)^{1/\mu}. \tag{2.41}$$

Bernstein estimate is precisely the inverse estimate (2.31) for dyadic splines, meaning that Bernstein's inequality is sufficient to determine the smoothness of target functions in $\mathbb{V}$ in terms of how well they are approximated by X_n. Once both inequalities hold one can draw comparison between $E(u, \mathbb{X}_n)_{\mathbb{V}}$ and $K(u, n^{-r}, \mathbb{V}, \mathbb{W})$ allowing for the characterization of approximation classes as interpolation spaces (DeVore and Popov [18])

Theorem 2.29 (DeVore and Popov [18]). *If the Jackson and Bernstein inequalities are valid, then for each $0 < s < r$ and $0 < q \leq \infty$ the following relations hold between approximation spaces and interpolation spaces*

$$\mathscr{A}_q^s(\mathbb{V}, \{\mathbb{X}_n\}_{n\geq 1}) = (\mathbb{V}, \mathbb{W})_{s/r,q}, \tag{2.42}$$

with equivalent quasi-norms.

The final step is to characterize the interpolation spaces $(\mathbb{V}, \mathbb{W})_{s/r,q}$ in terms of known smoothness spaces and the value of (DeVore and Popov [8]) Theorem 2.20 in approximation theory becomes evident:

Theorem 2.30. *For $0 < p < \infty$, $0 < s < \alpha$ and $\sigma = (\alpha + \frac{1}{p})^{-1}$ we have*

$$(L^p, \mathscr{B}_\sigma^\alpha(L^\sigma))_{s/\alpha,\sigma} = \mathscr{B}_\sigma^s(L^\sigma). \tag{2.43}$$

We have approximation class characterization for approximation with dyadic splines in multidimensions. Let $\mathbb{S}_n^r$ be the space piecewise polynomials of degree $r-1$ with $r-2$ continuous derivatives and bounded $r-1$ derivatives defined on all possible n piece dyadic partitions.

Theorem 2.31 (DeVore and Popov [8]). *Let $\mathcal{Q}$ be a d-dimensional cube. Let $0 < p \leq \infty$, let r be a positive integer. If $s < \lambda = \min(r, r-1+\frac{1}{p})$, Jackson and Bernstein hold with $\mathbb{V} = L^p(\mathcal{Q})$, $\mathbb{W} = \mathscr{B}_p^s(L^p)$ and $\mathbb{X} = \mathbb{S}_n^r$.*

Actually DeVore and Popov only showed that $E(f,\mathbb{S}_k^r)_p \preceq \omega_r(f,2^{-k})_p$. However, this immediately implies $\|\{\omega_r(f,2^{-k})_p\}_{k\geq 0}\|_{\ell_q^\alpha} = |f|_{\mathscr{B}_q^\alpha(L^p)}$, and by the embedding $\ell^q \subset \ell^\infty$ gives Jackson inequality

$$E(f,\mathbb{S}_n^r)_p \preceq 2^{-n\alpha}|f|_{\mathscr{B}_q^\alpha(L^p)} = N^{-\alpha/d}|f|_{\mathscr{B}_q^\alpha(L^p)}. \tag{2.44}$$

When $q=p$ the Besov space $\mathscr{B}_q^\alpha(L^p)$ is equivalent to fractional Sobolev spaces which is more suited to our study. Let $\mathbb{S}_n^r$ be the space of dyadic spline space on d-dimensional cube $\mathcal{Q}$ belonging to $W_\infty^r(\mathcal{Q})$. Given $0<p\leq\infty$ and λ. If $s<r$ then

$$E(f,\mathbb{S}_n^r)_p \preceq N^{-s/d}|f|_{W_p^s} \quad \forall f\in\mathscr{B}_p^s(L^p), \tag{2.45}$$

and if furthermore $s<r-1+\frac{1}{p}$,

$$\|S\|_{\mathscr{B}_p^s(L^p)} \preceq N^{s/d}\|S\|_{L^p} \quad \forall S\in\mathbb{S}_n^r. \tag{2.46}$$

giving us the characterization for any $0<q\leq\infty$

$$\mathscr{A}_q^{s/d}(L^p,\{\mathbb{S}_n^r\}) = (L^p,\mathscr{B}_p^\lambda(L^p))_{s/\lambda,q} = \mathscr{B}_q^s(L^p). \tag{2.47}$$

By setting $q=p$, we obtain a characterization of the approximation class $\mathscr{A}_p^{s/d}$ in terms of the fractional Sobolev space $W_p^s(\mathcal{Q})$.

Moreover, we have approximation class characterization for approximation with free-knot splines in one-dimension. Let Σ_n^r be the space of all possible n piecewise polynomials of degree $r-1$ with $r-2$ continuous derivatives and bounded $r-1$ derivatives. .

Theorem 2.32 (Petrushev [17]). *Let I be an interval. Let $0<p<\infty$, let r be a positive integer. If $\sigma=(r+\frac{1}{p})^{-1}$, Jackson and Bernstein hold with $\mathbb{V}=L^p(I)$, $\mathbb{W}=\mathscr{B}_\sigma^r(L^\sigma)$ and $\mathbb{X}=\Sigma_n^r$. Therefore for all $0<s<r+1$ and $0<q\leq\infty$, if $\sigma=(r+\frac{1}{p})^{-1}$*

$$\mathscr{A}_q^s(L^p,\{\Sigma_n^r\}) = (L^p,\mathscr{B}_\sigma^r(L^\sigma))_{s/r,q}. \tag{2.48}$$

The interpolation space $(L^p,\mathscr{B}_\sigma^r(L^\sigma))_{s/r,q}$ is a Besov space when $q=(s+\frac{1}{\sigma})^{-1}$ making approximation class $\mathscr{A}_\sigma^s(L^p(I),\{\Sigma_n^r\})$ identifiable with fractional Sobolev space $W_\sigma^s(I)$.

Example 2.33. We conclude by applying the characterization theorems on x^γ. Using linear approximation using dyadic piecewise constant polynomials $\mathbb{S}_n^1$ with error measured in L^p with $p\geq 1$. We find Besov norm of x^γ and use characterization Theorem 2.31 to show membership of x^γ in approximation class $\mathscr{A}^s=\mathscr{A}_\infty^s(L^p(\Omega),\{\mathbb{S}_n^1\})$. If $\gamma>1-\frac{1}{p}$ then in view of (2.9) with $r=1$, we have $\omega_1(x^\gamma,t)_p=\mathcal{O}(t)$ and

$$|(x^\gamma)|_{\mathscr{B}_p^s(L^p)}^p \preceq \int_0^1 \left(t^{-s}t\right)^p \tfrac{dt}{t} < \infty \quad \forall s<1, \tag{2.49}$$

making $x^\gamma \in \mathscr{A}^1$ in view of Theorem 2.31 which is consistent with the observation made in (2.27). On the other hand, if $\gamma < 1 - \frac{1}{p}$, we seek to see for which $0 < s < 1$ does the target function x^γ belong to Besov space $\mathscr{B}^s_p(L^p)$. Since $\gamma + \frac{1}{p} < 1$, we have $\omega_1(x^\gamma, t)_p = \mathcal{O}(t^{\gamma+1/p})$ and

$$|(x^\gamma)|^p_{\mathscr{B}^s_p(L^p)} \preceq \int_0^1 \left(t^{-s} t^{\gamma+1/p}\right)^p \tfrac{dt}{t} < \infty, \tag{2.50}$$

holding for every $s < \gamma + \frac{1}{p}$ and with the assertion of (2.32) we would expect convergence with rate $\mathcal{O}(N^{-\gamma-1/p})$. The compromise is clearer when using a change of variables by defining $\gamma = 1 - \frac{1}{p} - \varepsilon$ for some choice $\varepsilon \in (0, 1-\frac{1}{p})$; we have $x^\gamma \in \mathscr{A}^{1-\varepsilon}$. As for nonlinear approximation, recall that embedding $W^1_\sigma(I) \hookrightarrow L^p(I)$ is continuous when $\sigma = (1+\frac{1}{p})^{-1}$ and $|(x^\gamma)|_{W^1_\sigma(I)} < \infty$ for any positive γ, $(\gamma > -\frac{1}{p})$, and as a result optimal convergence is realized; that is, $x^\gamma \in \mathscr{A}^1$. The DeVore diagram depicting the discrepancy between linear and nonlinear approximation powers is given in Figure 2.2. $\oslash$

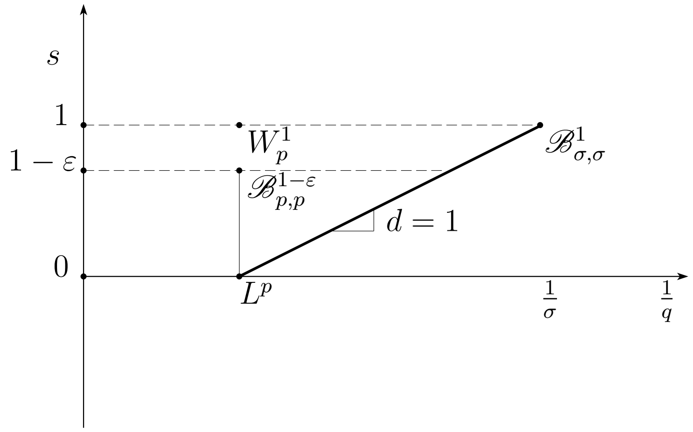


Figure 2.2: *In the DeVore diagram one see the gap between linear and nonlinear approximation when $\gamma = 1 - \frac{1}{p} - \varepsilon$. Here $\sigma = (1+\frac{1}{p})^{-1}$.*

2.2.3 Approximation with polynomials

We describe the construction of the spline spaces used in this thesis in two-dimensional setting and restrict ourselves with the unit square domain Ω only as the idea extends to rectangular and L-shaped domains with immediate facility. In fact, the idea extends immediately to higher dimensions as well due to the tensor-product structure. More complicated geometries can be obtained using the IGA framework (Hughes et al. [67]). The construction of stable adaptive basis spline functions which ensures sharp estimation is not possible on arbitrary partitions. We discuss the mesh structure criteria needed and how to achieve them, concluding with error estimates.

2.2.4 Polynomial splines

Assessing the rate at which finite element methods converge hinges on how well the solution itself can be approximated by polynomial bases. In this section we lay out the tools and idea behind polynomial approximation of functions in Besov spaces. In particular we discuss quasi-interpolation operators in the context of the basis functions required by our adaptive spline basis. This work is a translation of (Scott and Zang [19]) to hierarchical spline setting and is credited (Giannelli et al. [26], Speleers et al. [27] and [28]).

Given $u \in \mathcal{C}^r(\omega)$ defined on a Lipschitz domain ω, we denote by $T^r_y u$ the multi-dimensional Taylor polynomial of degree r centered at $y \in \omega$. If $u \in W^r_p(\Omega)$ then $D^\alpha u$ may not be understood in the point-wise sense. We go around the issue using a mollifier $\varphi_B \in \mathcal{C}^\infty_c(B)$ on a compact ball $B \subset \omega$ with φ_B scaled so that $\int_\omega \varphi_B = 1$. The averaged Taylor polynomial degree r is defined as

$$\mathscr{T}^r_B u(x) = \int_B T^r_y u(x) \varphi_B(y)\, dy. \tag{2.51}$$

The definition can be extended to $u \in L^1(B)$ in a consistent manner by re-writing (2.51) via partial integration. Here $\mathscr{T}^r_B u$ is an r degree polynomial and $\mathscr{T}^r_B$ defines a projection operator on $L^1(\omega)$ into $\mathbb{P}_r(\omega)$.

Theorem 2.34 (Averaged Taylor Polynomial). *Let ω be a convex body in $\mathbb{R}^n$ and let B_ω be a largest ball contained in ω with radius r_ω. For any $u \in W^{r+1}_p(\omega)$ and $1 \le p \le \infty$, we have*

$$\|D^\alpha(u - \mathscr{T}^r_B u)\|_{L^p(\omega)} \le C_{\mathscr{T}} \mathrm{diam}\,(\omega)^{r+1-|\alpha|} |u|_{W^{r+1}_p(\omega)}, \quad 0 \le |\alpha| \le r, \tag{2.52}$$

where $C_{\mathscr{T}}$ is a constant function independent of u but depends on the ratio $\frac{\mathrm{diam}\,(\omega)}{r_\omega}$.

This results in a staple result for piecewise polynomial approximation on a collection of subdomains:

Theorem 2.35 (Bramble-Hilbert Lemma). *Let $\omega \subset \mathbb{R}^n$ be convex with a largest inscribed ball in ω having radius $r_\omega > 0$. There is a $C_{\mathrm{HB}} > 0$, depending only on the ratio $\frac{\mathrm{diam}\,(\omega)}{r_\omega}$ and polynomial degree r, such that for all $0 \le k \le m \le r+1$,*

$$\inf_{P \in \mathbb{P}_r} |u - P|_{W^k_p(\omega)} \le C_{\mathrm{HB}} \mathrm{diam}\,(\omega)^{m-k} |u|_{W^m_p(\omega)} \quad \forall u \in W^k_p(\omega). \tag{2.53}$$

The ratio $\frac{\operatorname{diam}(\omega)}{r_\omega}$ is relevant when ω is triangular or general quadrilateral. If ω is a square then this ratio is fixed. Finally we need some inverse estimates which follow from finite dimensional normed space theory.

Lemma 2.36 (Polynomial inverse estimates). *Let $\tau \in P$. Then, for constants $c_{\mathrm{Inv}}, c_{\mathrm{dTr}} > 0$ and for any integers $1 \leq p, q < \infty$ and $0 \leq s \leq t \leq r+1$,*

$$|V|_{W_q^t(\tau)} \leq c_{\mathrm{Inv}} h_\tau^{s-t+2/q-2/p} |V|_{W_p^s(\tau)} \quad \forall V \in \mathbb{P}_r(\tau), \tag{2.54}$$

and if $\sigma \subset \partial\tau$, for a constant $c_{\mathrm{dTr}} > 0$,

$$\|V\|_{L^2(\sigma)} \leq c_{\mathrm{dTr}} h_\sigma^{-1/2} \|V\|_{L^2(\tau)} \quad \forall V \in \mathbb{P}_r(\tau), \tag{2.55}$$

in which c_{Inv} and c_{dTr} depend only on the polynomial degree r and $|\cdot|_{W_p^s}$ denotes the conventional Sobolev semi-norms.

The following discussion makes the analysis of dyadic splines discussed in the previous section much more constructive.

Univariate B-splines

This discussion closely follows (Schumaker [23]). Let $\mathcal{G}$ be a partitioning of $I = (0,1)$ obtained from the set of breakpoints $Z = \{0 < z_1 < \cdots < z_n < 1\}$ and let $\mathcal{P}_r(\mathcal{G})$ be the space of piecewise polynomials defined on partition $\mathcal{G}$:

$$\mathbb{S}(\mathbb{P}_r, \mathcal{G}, \mathfrak{M}) = \{S \in \mathcal{P}_r(\mathcal{G}) \mid S \in \mathcal{C}^{r-m_i}(\{z_i\}),\ i = 1:n\}, \tag{2.56}$$

where the set $\mathfrak{M} = \{m_i\}_{1\leq i\leq n}$ consisting of integers $1 \leq m_k \leq r+1$ is said to be the *multiplicity vector* of $\mathcal{G}$ and controls the number of continuous derivatives on each break point z_i. The dimension of $\mathbb{S}(\mathbb{P}_r, \mathcal{G}, \mathfrak{M})$ is obtained by subtracting the number of continuity conditions from $\dim \mathcal{P}_r(\mathcal{G})$ and

$$\dim \mathbb{S}(\mathbb{P}_r, \mathcal{G}, \mathfrak{M}) = (r+1)(n+1) - \sum_{i=1}^{n} (r+1-m_i) = M + r + 1, \tag{2.57}$$

where $M = \sum_{i=1}^n m_i$. A simple basis for (2.56) can be formed from translations of truncated polynomials

$$\left\{ \frac{(x-z_i)_+^{r+1-p}}{(r+1-p)!} \,\middle|\, p = 1 : m_i,\ i = 1 : n \right\}, \tag{2.58}$$

but this basis is not suited for numerical methods for a number of reasons, most notably, their support influence is unbounded with refinement making them numerically unstable.

Fortunately, given $\mathcal{G}$ and $\mathfrak{M}$ nontrivial locally supported spline functions exist provided that $M > r - 1$ by looking at the Vandermonde matrix of a locally supported spline in terms of basis (2.58) meaning that the spline space of a specified degree needs to have sufficient flexibility to have locally supported basis functions. One can't ask for both high smoothness and localized support; there will be a trade-off. The proof, which can be found in (Schumaker [23]) is constructive and provides a strategy to obtain the improved basis expressed in terms of divided difference of basis functions $(x-\xi)_+^r$ with ξ take over multiple repetitions of points in $\mathcal{G}$ according to $\mathfrak{M}$:

$$B_j(x) = (-1)^{r+1}(z_{i+r+1} - z_i)[z_i, ..., z_{i+r+1}](x-\xi)_+^r \cdot \mathbb{1}_{[z_i, z_{i+r+1}]}(x), \tag{2.59}$$

running over indices $j = 1 : M + r + 1$. This basis, we denote by $\mathbb{B}_r(\mathcal{G}, \mathfrak{M})$, is said to be the $r+1$ order *B-spline* basis on I and each basis function B_i is strictly positive on its support (z_i, z_{i+r+1}), moreover, the basis is locally linearly independent and forms a partition of unity on I allowing for stable and localized sharp error estimation. The local linear independence is a result of a *finite intersection property* where there is a maximum number of basis function support cell overlap. The study of B-splines has been a subject of great analysis going back to the 1940's.

For integers $r \geq 1$ and hierarchical level $\ell \geq 0$, the grid $\mathcal{G}_\ell$ is obtained from a set of $n_\ell = 2^\ell(r+1) - 1$ distinct *interior break points*

$$Z_\ell = \left\{ z_k^\ell = 2^{-\ell}\tfrac{k}{r+1} \mid k = 1 : n_\ell \right\} \tag{2.60}$$

with the convention that $z_0 = 0$ and $z_{n_\ell+1} = 1$ making the grid $\mathcal{G}_\ell = \left\{ (z_k^\ell, z_{k+1}^\ell) \mid k = 0 : n_\ell \right\}$ which partitions $(0,1)$ and gives rise to a sequence $\{\mathcal{G}_\ell\}_{0 \leq \ell \leq L-1}$ of dyadic partitions. It is immediate that the sequence $\{\mathcal{G}_\ell\}_{0 \leq \ell \leq L-1}$ forms a nested hierarchy of partitions; i.e, $\mathcal{G}_\ell \subset \mathcal{G}_{\ell+1}$. In view of the discussion leading up to the existence of (2.59), we consider the *open/extended knot vector* consisting of repetitions of break points Z_ℓ according to desired smoothness

$$\begin{aligned} \Xi_\ell = \{\xi_1^\ell \leq \cdots \leq \xi_{r+1}^\ell = 0 < \xi_{r+2}^\ell \leq \cdots \leq \xi_{M_\ell+r+1}^\ell < 1 \leq \xi_{M_\ell+r+2}^\ell \\ \leq \cdots \leq \xi_{M_\ell+2r+2}^\ell\}, \end{aligned} \tag{2.62}$$

and the subset $\{\xi_i^\ell\}_{i=r+2}^{M_\ell+r+1}$ is the *interior knot vector*. The presence of $r+1$ repetitions lying outside I are necessary to define spline function and derivative values on the boundary points. Moreover, in view of (2.59) there are $(M_\ell + r + 1)$-tupples $\Xi_i^\ell = \{\xi_i^\ell, ..., \xi_{i+r+1}^\ell\}$ called *local knot vectors* which makes up the support of each spline basis function, given level ℓ. A recursive procedure for the evaluation of (2.59) is given by the deBoor-Cox recursive Algorithm (2.2.4) $B_i^\ell(x) = \mathbf{BOOR}\,[\Xi_i^\ell, x]$: The set $\mathbb{B}_r(\Xi_\ell) = \{B_i^\ell \mid i = 1 : M_\ell + r + 1\}$ forms a basis for the spline space

$$\mathbb{S}_r(\Xi_\ell) = \text{span}(\mathbb{B}_r(\Xi_\ell)) \equiv \mathbb{S}(\mathbb{P}_r, \mathcal{G}_\ell, \mathfrak{M}_\ell), \tag{2.64}$$

Algorithm 2.2.1 Evaluate univariate B-spline basis $\mathbf{BOOR}\,[\Xi^\ell_i, x] \to B^\ell_i(x)$

```
for $j = i : i + r + 1$ do
    $B_{j,0}(x) = \mathbb{1}_{[\xi^\ell_j, \xi^\ell_{j+1})}(x)$
end for
for $k = 1 : r$ do
    for $j = i : i + r - k$ do
        $B^\ell_{j,k}(x) = \frac{x - \xi^\ell_j}{\xi^\ell_{j+k} - \xi^\ell_j} B_{j,k-1}(x) + \frac{\xi^\ell_{j+k+1} - x}{\xi^\ell_{j+k+1} - \xi^\ell_{j+1}} B_{j+1,k-1}(x) \qquad (2.63)$
    end for
end for
```

of dimension $M_\ell + r + 1$ uniquely determined by the inner knot vector. We turn our attention to refinement. In one-dimensions local refinement is done with facility through *knot insertion* by which a new break point, and corresponding knots, are introduced to the extended knot vector. In the context of our dyadic structure, by introducing m_* times repeated breakpoints $z^* = \frac{z^\ell_k + z^\ell_{k+1}}{2}$ in Ξ^ℓ to obtain Ξ', only the local knot vectors $\{\Xi^\ell_i \mid i = k - r : k\}$ will be effected; assuming that $z^\ell_k = \xi^\ell_k < \xi^\ell_{k+1}$,

$$\dots, \underbrace{\xi^\ell_{k-r-1}, \dots, \xi^\ell_k}_{\Xi_{k-r-1}}, \xi^*_1, \dots, \xi^*_{m_*}, \underbrace{\xi^\ell_{k+1}, \dots, \xi^\ell_{k+r+2}}_{\Xi_{k+1}}, \cdots \quad . \tag{2.65}$$

In the refinement process local vectors $\{\Xi^\ell_i \mid i = k - r : k\}$ are removed and replaced with suitable refined local knot vectors $\{\Xi'_i \mid i = k - r : k + 1\}$. The basis is effected locally with only $r + 1$ spline function require modification with all functions $\{B^\ell_i \mid i \neq k - r : k\}$ are not effected. The old basis functions $\{B^\ell_i \mid i = k - r : k\}$ will belong to the span of $\{\mathbf{BOOR}\,[\Xi'_i, \zeta] \mid i = k - r : k + 1\}$ giving us the nesting $\mathbb{S}_r(\Xi^\ell) \subset \mathbb{S}_r(\Xi^*)$ making it possible to express B^ℓ_k on Ξ^* as a linear combination of $\left\{B^{\ell+1}_i \mid i = 2k - 1 : 2k + r\right\}$ via the two-scale relation

$$B^\ell_k = \sum_{j=2k-1}^{2k+r} c^{\ell+1}_j B^{\ell+1}_j \quad \forall k = 1 : n_\ell, \tag{2.66}$$

for some sequence $c^{\ell+1}_j$.

Hierarchical B-spline (HB) spaces

Unfortunately local refinement by knot insertion is inefficient due to the propagation of knot insertions in multiple dimensions due to the tensor product structure. This can be overcome by *hierarchical refinement* (Vuong et al. [24]) through the exploitation of a multi-dimensional analog of (2.66); each to-be refined basis B_k^ℓ would be replaced with $\{B_i^{\ell+1} \mid i = 2k{-}1 : 2k{+}r\}$ leaving behind $\{B_i^\ell \mid i = k-1 : k-r, k+1 : k+r\}$ in the refined basis thus producing a *hierarchical B-spline basis*. The resulting basis would however not necessarily form a partition of unity and there are scenarios where the support of left-out basis functions may include an uncontrollable number of cells compromising finite intersection property. Fortunately both of those outcomes can be rectified with the addition of mesh structural constraints which will be the subject of discussion in the section below.

We now describe hierarchical partitions P in two dimensions and define the hierarchical spline obtained from a hierarchy of multilevel dyadic spline bases $\mathcal{B}_\ell = \mathbb{S}_r(\Xi_\ell)$ for $\ell = 0 : L{-}1$. A cell $\tau \in \mathcal{G}_\ell$ is said to be a cell of level ℓ. A cell τ of level ℓ is said to *active* if $\tau \in \mathcal{G}_\ell$ and appears in P. The hierarchical partition P satisfies the following properties: all cells τ in P are disjoint, and, the interior of the closure of the union $\cup\{\tau : \tau \in P\}$ is equal to Ω. A subdomain Ω^ℓ of $\overline{\Omega}$ is defined as the closure of the union of active cells τ of level ℓ or higher making $\Omega^\ell \supseteq \Omega^{\ell+1}$ for $\ell = 0 : L-1$ with $\operatorname{int}(\Omega^0) = \Omega$ and $\Omega^L = \varnothing$. In view of the two-scale relation (2.66) can always express a spline $S \in \mathbb{S}_\ell$ in terms of $\mathcal{B}_{\ell+1}$:

$$S = \sum_{\beta \in \mathcal{B}_{\ell+1}} c_\beta^{\ell+1}(S)\beta, \tag{2.67}$$

for some coefficients $\{c_\beta^{\ell+1}(S) \mid \beta \in \mathcal{B}_{\ell+1}\}$. A *Hierarchical B-spline* (HB-spline) basis $\mathcal{H}_P$ with respect to hierarchical partition P is defined as

$$\mathcal{H}_P = \left\{\beta \in \mathcal{B}^\ell \mid \operatorname{supp}\beta \subseteq \Omega^\ell \ \wedge\ \operatorname{supp}\beta \nsubseteq \Omega^{\ell+1}\right\}. \tag{2.68}$$

A recursive definition is given in (Speleers et al. [27]). A basis function β of level ℓ is said to *active* if $\beta \in \mathcal{B}_\ell \cap \mathcal{H}_P$, otherwise it is *passive*. Figures 2.3 and 2.4 depict hierarchical refinements in one dimension.

Admissible partitions

For local and stable approximation we need to control the influence of each basis function in order to ensure stable and sharp estimation. With additional restrictions on the structure of partitions P we can guarantee that the number of basis functions acting on any point is bounded and that the diameter of the support of a basis function is comparable to any cell in its support. A partition P is said to be *admissible* if the basis functions in $\mathcal{H}_P \cap \mathcal{B}_\ell$ is supported on cells belonging to at most two levels successive levels, namely, levels ℓ and $\ell + 1$. The procedure required to achieve this is given in **REFINE module** in Section 2.5 executed in a recursive manner. The support extension of a cell $\tau \in \mathcal{G}_\ell$ with respect to level

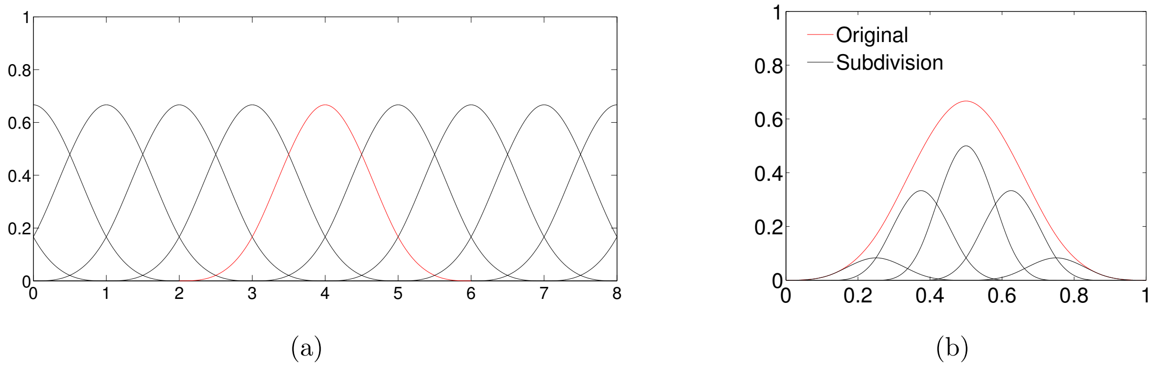


Figure 2.3: *(a) One-dimensional cubic B-spline basis functions. (b) Subdivision of an original uniform cubic B-spline into five contracted B-splines of half the knot span width.*

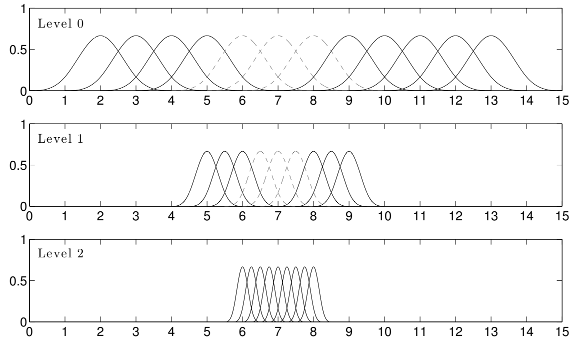


Figure 2.4: *Hierarchical refinement of a set of B-spline basis functions (active basis functions are indicated with solid lines).*

$k \leq \ell$ is defined as

$$S(\tau, k) = \left\{ \tau' \in G^k \mid \exists \beta \in \mathcal{B}^k \ s.t \operatorname{supp} \beta \cap \tau' \neq \varnothing \ \wedge \ \operatorname{supp} \beta \cap \tau \neq \varnothing \right\} \tag{2.69}$$

Note that the support extension consist of cells from the tensor-product mesh $\mathcal{G}^k$. To assess the locality of the basis; i.e, the influence of basis functions have on active cells, it is useful to consider a support extension consisting of all active cells belonging to its support regardless of level. For $\tau \in P$ define

$$\omega_\tau = \bigcup_{\ell=0}^{L-1} S(\tau,\ell) \cap P \equiv \{\tau' \in P \,|\, \operatorname{supp}\beta \cap \tau' \neq \varnothing \implies \operatorname{supp}\beta \cap \tau \neq \varnothing\}, \tag{2.70}$$

indicating the collection of all supports for basis function β's whose supports intersect τ. Analogously, we denote the support extension for an edge $\sigma \in \mathcal{E}_P \cup \mathcal{G}_P$ by

$$\omega_\sigma = \{\tau \in P \,|\, \operatorname{supp}\beta \cap \tau \neq \varnothing \implies \operatorname{supp}\beta \cap \tau \neq \varnothing,\ \sigma \subset \partial\tau\}. \tag{2.71}$$

The following auxiliary subdomain provides a way to ensure mesh admissibility

$$\mathcal{U}^\ell = \bigcup \left\{\overline{\tau} \,|\, \tau \in \mathcal{G}_\ell \ \wedge\ S(\tau,\ell) \subseteq \Omega^\ell\right\} \tag{2.72}$$

Theorem 2.37. *If $\Omega^\ell \subseteq \mathcal{U}^{\ell-1}$ for $\ell = 2 : L-1$, then P is an admissible partition.*

Proof. See (Buffa et al. [64]). ∎

In other words, $\mathcal{U}^\ell$ represents the biggest subset of Ω^ℓ so that the set of B-splines in $\mathcal{B}_\ell$ whose support is contained in Ω^ℓ spans the restriction of $\mathbb{S}^\ell$ to $\mathcal{U}^\ell$. This would ensure that each basis function will have support cells belonging to a maximum of two levels. Finally, partition of unity can be ensured by appropriate scaling of the basis functions; see (Vuong et al. [24]). In all of our numerical experiments such procedure did not seem of necessary.

Resulting partition has the following *shape-regularity* characteristics:

$$\sup_{P\in\mathscr{P}} \max_{\tau\in P} \{\tau' \in P \,|\, \tau' \subset \omega_\tau\} \leq N_{\mathscr{P}} < \infty$$

(locally quasi-uniform)

$$\sup_{P\in\mathscr{P}} \max_{\beta\in\mathcal{H}_P} \#\{\beta' \in \mathcal{H}_P \,|\, \operatorname{supp}\beta' \cap \operatorname{supp}\beta \neq \varnothing\} \leq M_{\mathscr{P}} < \infty$$

(finite intersection property)

from which one can derive the following:

$$\sup_{P\in\mathscr{P}} \max_{\tau\in P} \frac{\operatorname{diam}(\omega_\tau)}{\operatorname{diam}(\tau)} \leq c_{\text{shape}} < \infty, \tag{2.73}$$

and

$$\sum_{\tau\in P} |u|^p_{W^s_p(\omega_\tau)} \leq c^p_{\text{shape}} |u|^p_{W^s_p(\Omega)}. \tag{2.74}$$

An anologous statement to (2.73),(2.74) will hold for edges as well. Finally, for any two partitions $P_1, P_2 \in \mathscr{P}$ there exists a common admissible partition in $\mathscr{P}$, called the *overlay* and denoted by $P_1 \oplus P_2$, such that

$$\#(P_1 \oplus P_2) \leq \#P_1 + \#P_2 - \#P_0. \tag{2.75}$$

2.2.5 Approximation in HB spline spaces

Measuring the performance of piece-wise polynomial approximation will hinge on the results of polynomial approximation of Section 2.2.4, most notably Bramble-Hilbert Lemma and polynomial inverse estimates, and defining a suitable quasi-interpolation operator. Let $\Omega \subset \mathbb{R}^d$ and let $\mathbb{V}(\Omega) \subseteq L^p(\Omega)$, assuming that $\mathbb{S}_P \subset \mathbb{V}(\Omega)$ on a partition P is a spline basis spanned by basis functions $\mathcal{B}_P$ which are locally supported, non-negative and form a partition of unity over Ω,

$$I_P u = \sum_{\beta \in \mathcal{B}_P} \lambda_\beta(u)\beta, \tag{2.76}$$

where the $\lambda_\beta(u)$, for $\beta \in \mathcal{B}_P$, are coefficients which could be chosen in various ways. The set $\Lambda_P = \{\lambda_\beta \mid \beta \in \mathcal{B}_P\}$ can views as functionals on $\mathbb{V}(\Omega)$. Popular choices for coefficients $\lambda_\beta(u)$ could be point-values of u or its derivatives and local integrals of u. In general I_P is not a projection on $\mathbb{S}_P$. A special choice for the functionals λ_β is when Λ_P form a dual-basis for $\mathcal{B}_P$ automatically making I_P a projection operator on $\mathbb{S}_P$. As long as the dual-basis is such that

$$\|\lambda_\beta\|_{\mathbb{V}(\Omega)'} \preceq \operatorname{diam}(\operatorname{supp}\beta)^{-d/p}, \tag{2.77}$$

then, if $\mathcal{B}_\tau = \{\beta \in \mathcal{B}_P \mid \operatorname{supp}\beta \subseteq \omega_\tau\}$ and noting that admissible partition shape-regularity $\operatorname{diam}(\omega_\tau) \preceq \operatorname{diam}(\operatorname{supp}\beta)$ and β forming partition of unity

$$\|I_P u\|_{\mathbb{V}(\tau)} \leq \max_{\beta \in \mathcal{B}_\tau} |\lambda_\beta(u)| \left\| \textstyle\sum_{\beta \in \mathcal{B}_\tau} \beta \right\|_{\mathbb{V}(\omega_\tau)} \preceq \|u\|_{\mathbb{V}(\omega_\tau)}, \tag{2.78}$$

makes $I_P : \mathbb{V}(\Omega) \to \mathbb{S}_P$ a locally stable projection in $\mathbb{V}$ with $I_P u$ serving as a local near-best approximation:

$$\forall u \in \mathbb{V}(\Omega), \quad \|u - I_P u\|_{\mathbb{V}(\tau)} \leq (1 + \|I_P\|) \inf_{S \in \mathbb{S}_P} \|u - \chi\|_{\mathbb{V}(\omega_\tau)} \quad \forall \tau \in P. \tag{2.79}$$

Construction of a projection I_P for hierarchical spline space in multi-dimensions is achieved using the hierarchy of uniform dyadic spline spaces as a building blocks thanks to the nested partition structure. Specifically, as long as we have a suitable quasi-interpolation operator for each level, the hierarchical projection can be constructed with facility due to the work of (Speleers et al. [27]) with only two requirements: each level specific quasi-interpolation operator is a projection and the associated dual-basis functionals $\lambda \in \Lambda_\ell$ are supported in $\Omega_\ell \backslash \Omega_{\ell+1}$ in the sense that if $f|_{\Omega_\ell \backslash \Omega_{\ell+1}} = 0$ then we have $\lambda(f) = 0$

Theorem 2.38 (Speleers [27]). *Let $I_\ell : \mathbb{V}(\Omega) \to \mathbb{S}_\ell$ be a sequence of quasi-interpolation projections such that all of $\lambda \in \Lambda_\ell$ are supported in $\Omega_\ell \backslash \Omega_{\ell+1}$. Then $I_P u : \mathbb{V}(\Omega) \to \mathbb{S}_P$ defined by*

$$I_P u = \sum_{\ell=0}^{L-1} \sum_{\beta \in \mathcal{B}_\ell \cap \mathcal{H}_P} \lambda_\beta(u)\beta, \tag{2.80}$$

is also a projection.

The tensor-product structure will also make the construction of the multi-level quasi-interpolation operators follow from the one-dimensional case in an almost immediate fashion. The one-dimensional construction was studied in great detail by de Boor and Fix in 1973.

Lemma 2.39. *Let $\mathcal{G}_\ell$ be dyadic partition of level ℓ of a unit interval I, let $\mathfrak{M}_\ell$ be a multiplicity vector and let Ξ_ℓ be an ℓ-level extended knot vector and let $\mathbb{B}(\Xi_\ell) = \{\beta_i\}_i$ be the spline basis that generates $\mathbb{S}(\mathbb{P}_r, \mathcal{G}_\ell, \mathfrak{M}_\ell)$. There exists a dual-basis $\Lambda(\Xi_\ell)$ of linear functionals λ_β with*

$$|\lambda^\ell_\beta(f)| \leq (2r+3)9^r \mathrm{diam}\,(\mathrm{supp}\,\beta)^{-1/p} \|f\|_{L^p(\mathrm{supp}\,\beta)}, \quad 1 \leq p \leq \infty. \tag{2.81}$$

The linear functionals λ_β in the lemma have the integral form

$$\lambda_\beta(f) = \int_{\mathrm{supp}\,\beta} f D^{r+1} u_\beta \, dx, \quad \beta \in \mathbb{B}(\Xi_\ell), \tag{2.82}$$

where u_β is defined explicitly in (Schumaker [23]) with

$$\|D^{r+1} u_\beta\|_{L^\infty(I)} \leq (2r+3)9^r \mathrm{diam}\,(\mathrm{supp}\,\beta)^{-1}. \tag{2.83}$$

Together with a tensorization of bases $\{\mathbb{B}(\Xi_\ell) \mid \ell = 0 : L-1\}$ and corresponding dual basis $\{\Lambda(\Xi_\ell) \mid \ell = 0 : L-1\}$ admitting a two-dimensional form of (2.82) we consider the quasi-interpolation operator defined by (2.80). We have a local approximation estimate when $\mathbb{V}(\Omega)$ is a Sobolev space and $\mathbb{S}_P$ has a fixed number of continuous derivatives that is less than r in each direction:

Theorem 2.40 (Speleers [28]). *Let $1 \leq p \leq \infty$, let $1 \leq k \leq r+1$ and let $0 \leq l \leq k-1$. The linear projection $I_P : W^k_p(\Omega) \to \mathbb{S}_P$ is such that*

$$|u - I_P u|_{W^l_p(\tau)} \leq C_Q \mathrm{diam}(\tau)^{k-l} |u|_{W^k_p(\omega_\tau)} \quad \forall v \in W^k_p(\Omega), \quad (\tau \in P), \tag{2.84}$$

and

$$|u - I_P u|_{W^l_p(\sigma)} \leq C_Q \mathrm{diam}(\tau)^{k-l-\frac{1}{2}} |u|_{W^k_p(\omega_\sigma)} \quad \forall v \in W^k_p(\Omega), \quad (\sigma \in \mathcal{E}_P), \tag{2.85}$$

for constant $C_Q > 0$ depends only on r and p.

The result holds by controlling $|u-S|_{W^l_p(\tau)}$ and $|I_P(u-S)|_{W^l_p(\tau)}$ given any spline $S \in \mathbb{S}_P$ where the latter expression is owed to I_P being a projection. The first one follows immediately from Bramble-Hilbert lemma of Theorem 2.35. The second requires more delicate treatment aided with the inverse estimates in Theorem 2.36, local stability of I_P (2.78), and finally Bramble-Hilbert to obtain the rate with respect to cell diameter. In the proof of Lemma 2.43 we carry a similar argument.

Smoothness indices k and l in Theorem 2.40 can be extended to real numbers by the following interpolation result.

Theorem 2.41. *Suppose $\theta \in (0,1)$ and $q \in [1,\infty]$. If $T \in \mathcal{L}(X_i, Y_i)$ with norm M_i, for $i = 0,1$. Then $T \in \mathcal{L}\left((X_0, X_1)_{\theta,q}, (X_1, Y_1)_{\theta,q}\right)$ with norm $M \leq M_0^{\theta} M_1^{1-\theta}$.*

2.2.6 Projection operators

In this thesis we will be using a number of different approximating projection operators all of which will depend on whether the HB spline space $\mathbb{X}_P$ satisfies the Dirichlet boundary condition or not. In the standard setting where $\mathbb{X}_P \subset H^2_0(\Omega)$ (in Chapter 3) we just need the first projection Q_P. The remaining I_P and Π_P are relevant when the boundary conditions are prescribed using Nitsche's penalty terms. Projector I_P is used in Chapters 4 and 5 where convergence is not studied. When we look at convergence we will use Q_P instead for the reason that we would get a better upper bound (see (6.10) and compare with (4.23)) making the convergence proof possible in Nitsche's setting, otherwise the extra powers of γ_1, γ_2 pollutes the contraction result of Theorem 6.9. Finally, the orthogonal projection L^2 is used to avoid the saturation assumption of Chapter 4.

Quasi-interpolation projection Q_P

We assume that the HB spline space $\mathbb{X}_P \subset H^2_0(\Omega)$.

Lemma 2.42 (Quasi-interpolation). *The quasi-interpolation projection operator $Q_P : H^2_0(\Omega) \to \mathbb{X}_P$ is such that for a constant $c_{\text{shape}} > 0$,*

$$h_\tau^{2k-4} |u - Q_P u|^2_{H^k(\tau)} \leq c^2_{\text{shape}} |u|^2_{H^2(\omega_\tau)}, \quad (\tau \in P), \tag{2.86}$$

for $k = 0, 1, 2$. Moreover,

$$h_\sigma^{-3} \|u - Q_P u\|^2_{L^2(\sigma)} \leq c^2_{\text{shape}} |u|^2_{H^2(\omega_\sigma)}, \quad (\sigma \in \mathcal{E}_P), \tag{2.87}$$

and

$$h_\sigma^{-1} \left\| \frac{\partial(u - Q_P u)}{\partial \boldsymbol{n}_\sigma} \right\|^2_{L^2(\sigma)} \leq c^2_{\text{shape}} |u|^2_{H^2(\omega_\sigma)}, \quad (\sigma \in \mathcal{E}_P), \tag{2.88}$$

holding for every $u \in H^2_0(\Omega)$.

Proof. The proof is in essence the same as proof of Theorem 2.40. ■

Quasi-interpolation projection I_P

The HB spline space $\mathbb{X}_P \subset H^2(\Omega)$ violates the boundary conditions. The statements of Lemma 2.42 hold true for $I_P : H^2 \to \mathbb{X}_P$.

L^2-orthogonal projection Π_P

Let $\mathcal{P}^r_P(\Omega)$ be the space of piece-wise polynomials of degree $r \geq 0$ defined on P. Let $\Pi^r_P : H^2(\Omega) \to \mathcal{P}^r_P(\Omega)$ be given by standard L^2-orthogonal projection. We define an auxillary subdomain of Ω which will be of use to us in Chapters 5 and 6:

$$D^\Gamma_P = \overline{\bigcup \{\tau \in P : \tau \text{ adjacent to } \Gamma\}}. \tag{2.89}$$

Lemma 2.43 (L^2-orthogonal projection). *Let P be an admissible partition, let $\sigma \in \mathcal{G}_P$ with $\sigma \subset \partial\tau_\sigma$ for a boundary adjacent cell $\tau_\sigma \in P$. For a constant $c_\Pi > 0$ depending only on polynomial degree r, if $u \in H^s(\Omega)$ with $s \geq 2$, then*

$$\left(\sum_{\sigma\in\mathcal{G}_P} h_\sigma \|u - \Pi^r_P u\|^2_{L^2(\sigma)}\right)^{1/2} \leq c_\Pi |h^s_P u|_{H^s(D^h_\Gamma)}, \tag{2.90}$$

and

$$\left(\sum_{\sigma\in\mathcal{G}_P} h^3_\sigma \left\|\frac{\partial\big(u-\Pi^r_P u\big)}{\partial \boldsymbol{n}_\sigma}\right\|^2_{L^2(\sigma)}\right)^{1/2} \leq c_\Pi |h^s_P u|_{H^s(D^h_\Gamma)}. \tag{2.91}$$

Remark 2.44. It is known that the L^2-projection is stable, i.e., $\|\Pi^r_P u\|_{L^2(\Omega)} \leq c_\Pi \|u\|_{L^2(\Omega)}$. We use c_Π as the stability constant, i.e., $\|\Pi^r_P u\|_{L^2(\Omega)} \leq c_\Pi \|u\|_{L^2(\Omega)}$.

Proof. In what follows, any proportionality relation $\preceq$ depends only on the polynomial degree $r \geq 0$. The general trace inequality gives

$$\|u - \Pi^r_P u\|^2_{L^2(\sigma)} \preceq h^{-1}_\sigma \|u - \Pi^r_P u\|^2_{L^2(\tau_\sigma)} + h_\sigma |u - \Pi^r_P u|^2_{H^1(\tau_\sigma)}. \tag{2.92}$$

The norms of $u - \Pi^r_P u$ on τ_σ is estimated using the following standard argument and an application of Hilbert Bramble lemma of Theorem 2.35. Let $p \in \mathbb{P}_r(\tau_\sigma)$. Then, we have

$$\|u - \Pi^r_P u\|_{L^2(\tau_\sigma)} \leq \|u - p\|_{L^2(\tau_\sigma)} + \|\Pi^r_P(p-u)\|_{L^2(\tau_\sigma)} \leq 2\|u - p\|_{L^2(\tau_\sigma)} \tag{2.93}$$

and

$$\begin{aligned}|u-\Pi_P^r u|_{H^1(\tau_\sigma)} &\le |u-p|_{H^1(\tau_\sigma)} + |\Pi_P^r(p-u)|_{H^1(\tau_\sigma)}\\ &\le |u-p|_{H^1(\tau_\sigma)} + c_{\mathrm{Inv}} h_{\tau_\sigma}^{-1}\|\Pi_P^r(p-u)\|_{L^2(\tau_\sigma)}\\ &\le |u-p|_{H^1(\tau_\sigma)} + c_{\mathrm{Inv}} h_{\tau_\sigma}^{-1}\|u-p\|_{L^2(\tau_\sigma)}\end{aligned} \tag{2.95}$$

where the constant c_{Inv} comes from Lemma 2.36. With any $s_1 \ge 0$, the classical Bramble-Hilbert lemma gives

$$\inf_{p\in\mathbb{P}_r(\tau_\sigma)} \|u-p\|_{L^2(\tau_\sigma)} \preceq h_{\tau_\sigma}^{s_1}|u|_{H^{s_1}(\tau_\sigma)}, \tag{2.96}$$

and similarly, for $s_2 \ge 1$,

$$\inf_{p\in\mathbb{P}_r(\tau_\sigma)} |u-p|_{H^1(\tau_\sigma)} \preceq h_{\tau_\sigma}^{s_2}|u|_{H^{s_2}(\tau_\sigma)}. \tag{2.97}$$

Taking $s = \max\{s_1, s_2\}$ yields

$$\|u-\Pi_P^r u\|_{L^2(\sigma)}^2 \preceq h_\sigma^{-1} h_{\tau_\sigma}^{2s}|u|_{H^s(\tau_\sigma)}^2 + h_\sigma h_{\tau_\sigma}^{2s-2}|u|_{H^s(\tau_\sigma)}^2 \preceq h_\sigma^{2s-1}|u|_{H^s(\tau_\sigma)}^2, \tag{2.98}$$

using $h_{\tau_\sigma} \approx h_\sigma$ because σ is an edge of τ_σ. As a consequence, we obtain

$$\|u-\Pi_P^r u\|_{L^2(\sigma)} \le c_\Pi h_\sigma^{s-1/2}|u|_{H^s(\tau_\sigma)}, \tag{2.99}$$

for a constant $c_\Pi > 0$. Squaring both sides and summing over all boundary edges $\sigma \in \mathcal{G}_P$ result in

$$\sum_{\sigma\in\mathcal{G}_P} h_\sigma \|u-\Pi_P^r u\|_{L^2(\sigma)}^2 \le c_\Pi^2 \sum_{\sigma\in\mathcal{G}_P} h_\sigma^{2s}|u|_{H^s(\tau_\sigma)}^2 = c_\Pi^2 |h_\Gamma^s u|_{H^s(D_\Gamma^h)}^2. \tag{2.100}$$

This completes the proof of (2.90). Similarly, we can prove (2.91) by starting with the following general trace inequality

$$\begin{aligned}\left\|\tfrac{\partial(u-\Pi_P^r u)}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}^2 &\preceq h_\sigma^{-1}\left\|\tfrac{\partial(u-\Pi_P^r u)}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\tau_\sigma)}^2 + h_\sigma\left|\tfrac{\partial(u-\Pi_P^r u)}{\partial \boldsymbol{n}_\sigma}\right|_{H^1(\tau_\sigma)}^2\\ &\preceq h_\sigma^{-1}|u-\Pi_P^r u|_{H^1(\tau_\sigma)}^2 + h_\sigma|u-\Pi_P^r u|_{H^2(\tau_\sigma)}^2\end{aligned} \tag{2.102}$$

and the rest follows similarly as above when using $s \ge 2$ in the Hilbert-Bramble lemma. ■

2.3 Adaptive spline approximation

Designing approximation methods by exploiting *a priori* local error estimates is not possible when direct access to a target function is not available. Nonlinear approximation of solutions to a boundary value problems on partitions must instead be done with the aid of *a posteriori* estimates. We assume that we have at our disposal an estimate $\eta(u,\tau)$ for the local approximation error whose sum $e(u,P) = \sum_{\tau\in P}\eta(u,\tau)$ is a reasonably good upper bound to $E(u,\mathbb{S}_P)_{\mathbb{V}}$. If $\mathscr{P}_n$ is the family of n-piece partitions obtained by subdividing an initial partition P_0, an *optimal* n-piece partition $P \in \mathscr{P}_n$ will realize $e(u,P) = \inf_{P'\in\mathscr{P}_n} E(u,\mathbb{S}_{P'})_{\mathbb{V}}$ however it is a daunting task to commit to this search since it has exponential complexity. Instead, cost efficient *near-optimal* partitions are obtained by adaptive tree approximation (Binev et al. [51]) where we seek in an iterative manner an n-piece partition P such that for absolute constants $c,\ C > 0$, with $c < 1$ depending on the number of new cells produced by each subdivision,

$$e(u,P) \le C \inf_{P'\in\mathscr{P}_{cn}} e(u,P'). \tag{2.103}$$

This is done by subdividing the cell τ corresponding to the highest error functional $\eta(u,\tau)$. The resulting partition is generated from an initial partition P_0. When solving differential equations more sophisticated selection procedures are necessary to achieve optimal efficiency; see Section 2.5. The tree adaptive approximation class is defined by:

$$\mathscr{A}^s(\mathbb{V},\{\mathbb{S}_P\}_{P\in\mathbb{F}}) = \left\{u \in \mathbb{V}(\Omega) \mid |u|_{\mathscr{A}^s} = \sup_{n\ge 1} n^s \inf_{P\in\mathscr{P}_n} e(u,P) < \infty\right\}, \tag{2.104}$$

where $\mathbb{F} = \{\mathscr{P}_n \mid n \ge 1\}$ is the master forest. In practice, methods would terminate after reaching a desired error threshold; for $\varepsilon > 0$, the adaptive method produces a partition P_ε would satisfy $e(u,P_\varepsilon) \le \varepsilon$ and the complexity of $\#P_\varepsilon$ is given by $\mathcal{O}(\varepsilon^{-1/s})$. The adaptive tree algorithm (Binev et al. [51]) maintains that P_ε is near-optimal in the sense that all other partitions obtained by a sequence of subdivisions from the initial mesh P_0 and come close within the error tolerance cannot be have significantly more favorable mesh complexity.

In our setting, the complexity of ensuring admissible partitions must be taken into account. Typically in each iteration the mesh undergoes further refinements to maintain conditions of Theorem 2.37. This process is said to be a *completion* step. We define the set of *marked* elements $\#\mathscr{M}_k$ to be the number of cells needed to be refined to achieve sufficient reduction in error prior to mesh completion. If $\bar{P} = \mathbf{REFINE}(P,\mathscr{M})$ where P is admissible, $\mathscr{M}$ obtained from a suitable estimator, and $P_* = \mathbf{CONF}(P)$ generates an admissible partition from $\bar{P}$, The result of (Binev et al. [51]) enables establishing an equivalence between approximation classes defined on all possible partitions of fixed complexity and those limited to admissible partition $\mathscr{P}_n^a$:

$$\#P_* \le C_\Lambda \#\bar{P}. \tag{2.105}$$

In other words, any possible domino effect is not detrimental to approximation performance and the class (2.104) will be equivalent to

$$\mathscr{A}^s(\mathbb{V},\{\mathbb{S}_P \mid P \in \mathscr{P}_n^a\}_{n\ge1}). \tag{2.106}$$

A final point to make is that if we are at a refinement iteration k it would not be possible to estimate the difference of two subsequent admissible partitions $P_{k+1} - \#P_k$ in terms of $\mathscr{M}_k$ as it is possible that the refinement of one marked element may result in arbitrarily large number of additional refinements due to completion resulting in a domino effect. Fortunately, we have that the order of $\#P_n$ is controlled by the sequence of marked elements only up to iteration k:

$$\#P_n \leq \#P_0 + C_\Lambda(\#\mathscr{M}_0 + \cdots + \#\mathscr{M}_{n-1}). \tag{2.107}$$

See, e.g., (Buffa [65] and Binev et al. [59])

Algorithm 2.3.1 Theoretical adaptive procedure **ADAPTIVE** $[P, \varepsilon] \to P_*$

```
$\mathscr{M} = \{\tau \in P \mid e(\tau, P) > \varepsilon\}$
while $\mathscr{M} \neq \varnothing$ do
    $\bar{P}$ = REFINE$(P, \mathscr{M})$
    $P_*$ = CONF$(\bar{P})$
    $\mathscr{M} = \{\tau \in P_* \mid e(\tau, P_*) > \varepsilon\}$
    $P = P_*$
end while
```

Following result is due to (Binev et al. [20]) which we include for completeness.

Lemma 2.45. *Let $v \in \mathscr{B}^\alpha_{p,p}(\Omega)$ for $\alpha \geq 0$, $0 < p < \infty$ and let $\delta > 0$.*

$$e(\tau, P) = |\tau|^\delta |v|_{\mathscr{B}^\alpha_{p,p}(\omega)}, \quad \omega = \tau \text{ or } \omega_\tau. \tag{2.108}$$

Given any $\varepsilon > 0$, the adaptive Algorithm 2.3 we will terminates in finite steps and produces an admissible partition $P \in \mathscr{P}$ for which

$$\sum_{\tau \in P} e(\tau, P)^2 \preceq \#P\varepsilon^2 \quad \text{and} \quad \#P - \#P_0 \preceq |v|^{p/(1+\delta p)}_{\mathscr{B}^\alpha_{p,p}(\Omega)} \varepsilon^{-p/(1+\delta p)}. \tag{2.109}$$

Proof. With each refinement step, foe error quantities $e(\tau_*, P_\ell)$ exceeding $\varepsilon > 0$, $|\tau|$ will reduce by a factor $1/4$ and $e(\text{child}(\tau_*), P_{\ell+1}) \leq 4^{-\delta} e(\tau_*, P_{\ell+1})$. We will have $\mathscr{M}_\ell = \varnothing$ after a finite number of steps L; set $P = P_L$ we obtain the first relation in (2.109).

We estimate the cardinality of the resulting partition P. Let $R_\ell \subset P_\ell$ be the set of refined cells and put $\mathcal{R} = \cup_{\ell=0}^L R_\ell$. Let $\Gamma_j = \{\tau \in \mathcal{R} : 2^{-j-1} \leq |\tau| \leq 2^{-j}\}$ and let $m_j = \#\Lambda_j$.

First of all, there can be at most $2^{j+1}|\Omega|$ disjoint τ of size $> 2^{-j-1}$ which makes $m_j \leq 2^{j+1}|\Omega|$ which gives us one upper bound on m_j.

We obtain a second upper bound in the following manner. Let $\tau \in \Gamma_j$, then

$$e(\tau, P) = |\tau|^\delta |v|_{\mathscr{B}^\alpha_{p,p}(\omega)} < 2^{-j\delta} |v|_{\mathscr{B}^\alpha_{p,p}(\omega)}, \tag{2.110}$$

and

$$m_j\varepsilon^p < \sum_{\tau\in\Gamma_j} e(\tau,P)^p < 2^{-jp\delta}\sum_{\tau\in\Gamma_j}|v|^p_{\mathscr{B}^\alpha_{p,p}(\omega_\tau)} \preceq 2^{-jp\delta}|v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}, \tag{2.111}$$

by shape-regularity. We therefore obtain $m_j \preceq 2^{-jp\delta}|v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-p}$.

Let j_0 be the smallest integer for which $|\Omega| < 2^{j_0}$. Then if $\mathscr{M} = \cup_{\ell=0}^L \mathscr{M}_\ell$

$$\#\mathscr{M} \le \sum_{j=-j_0}^{\infty}\#m_j \preceq \sum_{j=-j_0}^{\infty}\min\{2^j|\Omega|, 2^{-jp\delta}|v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-p}\}. \tag{2.112}$$

If k is biggest integer for which $2^k|\Omega| \le 2^{-kp\delta}|v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-p}$, then

$$\sum_{j=-j_0}^{\infty}\min\{2^j|\Omega|, 2^{-jp\delta}|v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-p}\} = |\Omega|\sum_{j=-j_0}^{k}2^j + |v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-p}\sum_{j=k+1}^{\infty}2^{-jp\delta}. \tag{2.113}$$

Observe that

$$\sum_{j=-j_0}^{k}2^j \preceq 2^k, \quad \sum_{j=k+1}^{\infty}2^{-jp\delta} \preceq 2^{-kp\delta} \quad \text{and} \quad 2^{k(1+p\delta)} \le |\Omega|^{-1}|v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-p} \tag{2.114}$$

which makes

$$\#P - \#P_0 \preceq \#\mathscr{M} \preceq 2^{-kp\delta}|v|^p_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-p} \le \left(|\Omega|^\delta|v|_{\mathscr{B}^\alpha_{p,p}(\Omega)}\varepsilon^{-1}\right)^{p/(1+\delta p)}, \tag{2.116}$$

■

where we invoked (2.107).

We have the following one-sided characterization for (2.104) in terms of Besov spaces, with error measured in $\mathbb{V} = H^2(\Omega)$ and hierarchical B-spline spaces $\mathbb{S}_P \subset \mathcal{P}_r \cap C^1(\Omega)$, with $r \ge 2$, defined on admissible partitions :

Theorem 2.46. *We have $\mathscr{B}^{s+2}_{p,p} \hookrightarrow \mathscr{A}^{s/2}$ for $s < r - 2 + \max\{1, \frac{1}{p}\}$ with $0 < \frac{1}{p} \le \frac{s+1}{2}$.*

Remark 2.47. If s exceeds the first aforementioned condition in the result above, then any such function will necessarily belong to $\mathbb{S}_P$ which is a saturation result. We illustrate the statement of this result by a DeVore diagram (see Figure 2.5).

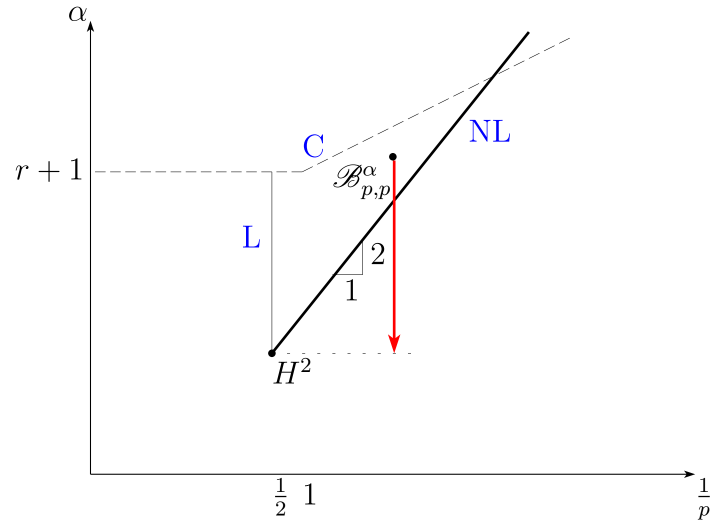


Figure 2.5: *DeVore diagram depicting the embedding of Besov spaces into the approximation $\mathscr{A}^s$. The demarcated Line C constraining α comes from avoiding non-trivial Besov spaces; Besov spaces above the line are r-degree polynomials. Line L corresponds to approximation using uniform refinement and is limited by basis functions polynomial order only. Line NL corresponds to adaptive refinement and coincides with the Sobolev embedding line. If $\mathscr{B}^{\alpha}_{p,p}$ is positioned between lines L, NL and C, then $\mathscr{B}^{\alpha}_{p,p} \hookrightarrow \mathscr{A}^s$ with $s = \frac{\alpha}{2} - 1$. The length of the red vector is equal to s.*

Proof. Let $\pi \in \mathbb{P}_r(\omega_\tau)$.

$$\begin{aligned} |v - I_P v|_{H^2(\tau)} &\leq |v-\pi|_{H^2(\tau)} + |I_P(\pi - v)|_{H^2(\tau)}, \\ &\leq |v-\pi|_{H^2(\tau)} + c_{\text{shape}}|\pi - v|_{H^2(\omega_\tau)} \preceq c_{\text{shape}}|v-\pi|_{H^2(\omega_\tau)}. \end{aligned} \tag{2.118}$$

Let $\omega_\tau = T(G)$ and $\hat{v} = v \circ T$. For $s < r - 2 + \max\{1, \frac{1}{p}\}$ we have nontrivial Besov spaces $\mathscr{B}^{s+2}_{p,p}(G)$ when defined with $\omega_{r+1}(\cdot, t)_p$ (see Remark 2.19). Moreover, if $\frac{1}{p} \leq \frac{s+1}{2}$ we have the continuous embedding $\mathscr{B}^{s+2}_{p,p}(G) \hookrightarrow H^2(G)$. Together with the facts $|\hat{v}|_{\mathscr{B}^{s+2}_{p,p}(G)} \approx h_\tau^{2+s-2/p}|v|_{\mathscr{B}^{s+2}_{p,p}(\omega_\tau)}$ and $|\hat{\pi}|_{\mathscr{B}^{s+2}_{p,p}(G)} = 0$ we arrive at

$$h_\tau |v - \pi|_{H^2(\omega_\tau)} \approx |\hat{v} - \hat{\pi}|_{H^2(G)} \preceq \|\hat{v} - \hat{\pi}\|_{L^p(G)} + |\hat{v}|_{\mathscr{B}^{s+2}_{p,p}(G)}. \tag{2.119}$$

Invoking Whitney's estimate (2.29),

$$\inf_{\pi \in \mathbb{P}_r(\omega_\tau)} h_\tau |v-\pi|_{H^2(\omega_\tau)} \preceq |\hat{v}|_{\mathscr{B}^{s+2}_{p,p}(G)} \approx h_\tau^{2+s-2/p}|v|_{\mathscr{B}^{s+2}_{p,p}(\omega_\tau)}, \tag{2.120}$$

from we obtain

$$\inf_{\pi \in \mathbb{P}_r(\omega_\tau)} |v-\pi|_{H^2(\omega_\tau)} \preceq |\tau|^{\delta}|v|_{\mathscr{B}^{s+2}_{p,p}(\omega_\tau)}, \tag{2.121}$$

with $\delta = \frac{s+1}{2} - \frac{1}{p} > 0$. We have the local estimate

$$\forall \tau \in P, \quad |v - I_P v|_{H^2(\tau)} \preceq c_{\text{shape}} |\tau|^{\delta} |v|_{\mathscr{B}^{s+2}_{p,p}(\omega_\tau)}, \tag{2.122}$$

and therefore the global error is

$$|v - I_P v|^2_{H^2(\Omega)} = \sum_{\tau \in P} |v - I_P v|^2_{H^2(\tau)} \preceq \sum_{\tau \in P} e(\tau, P)^2. \tag{2.124}$$

In view of Lemma 2.45 with $\omega = \omega_\tau$, there exists an admissible mesh $P \in \mathscr{P}$ such that

$$|v - I_P v|^2_{H^2(\Omega)} \preceq \#P\varepsilon^2 \quad \text{with} \quad \#P - \#P_0 \preceq |v|^{p/(1+\delta p)}_{\mathscr{B}^{2+\alpha}_{p,p}} \varepsilon^{-p/(1+\delta p)}. \tag{2.125}$$

Using the definition of δ, we determine that $p/(1+\delta p) = 2/(s+1)$. Let $N = \#P$ and let $\varepsilon = N^{-(s+1)/2} |v|_{\mathscr{B}^{s+2}_{p,p}(\Omega)}$ then

$$|v - I_P v|_{H^2(\Omega)} \preceq |v|_{\mathscr{B}^{s+2}_{p,p}(\Omega)} N^{-s/2} \quad \text{and} \quad \#P - \#P_0 \preceq N. \tag{2.126}$$

Therefore,

$$\begin{aligned} |v|_{\mathscr{A}^{s/2}} &= \sup_{N>0} N^{s/2} \inf_{P \in \mathscr{P}_N} \inf_{V \in \mathbb{X}_P} |u - V|_{H^2(\Omega)}, \\ &\leq \sup_{N>0} N^{s/2} |v - I_P v|_{H^2(\Omega)} \preceq |v|_{\mathscr{B}^{s+2}_{p,p}(\Omega)} < \infty. \end{aligned} \tag{2.128}$$

■

The direct estimate above cannot be paired with an inverse estimate hindering the application of Characterization Theorem 2.29. This is primarily to the smoothness limitation of the space $\mathbb{S}_P$. Instead generalized Besov spaces which are characterized by multi-scale decomposition are used. We do not proceed with this and refer to (Binev et al. [20], Gaspoz and Morin [21] and Tsogtgerel [22]) for detailed discussion.

Remark 2.48. The parameter s in Theorem 2.46 should not be confused with the Besov norm smoothness parameter in $|\cdot|_{\mathscr{B}^s_{p,p}}$. Here it means the additional regularity exceeding its minimal possible smoothness of $\alpha = 2$; only in excess of $\alpha = 2$ can approximation be possible. Interchangeably, we can always write $\mathscr{B}^{s+2}_{p,p} \hookrightarrow \mathscr{A}^{s/2}$ as $\mathscr{B}^{\alpha}_{p,p} \hookrightarrow \mathscr{A}^{s}$ whenever $s = \frac{\alpha}{2} - 1$.

2.4 Regularity results

The fourth order problem requires Green's identity as a tool for analysis, and since full $H^4(\Omega)$ regularity of solutions cannot be expected, we present here the identity under weaker assumptions. Only having $u \in H^2(\Omega)$, the Green identity no longer holds in the usual sense since $\frac{\partial \Delta u}{\partial \boldsymbol{n}}|_\Gamma$ and $\Delta u|_\Gamma$ lose their meaning. We reiterate the result of (Bhattacharyya et al. [4], Blum, Rannacher and Leis [3]) and briefly discuss the main ideas used, namely that if u belongs to the subspace $H^2(\Omega)$ of functions with $\Delta^2 u \in L^2(\Omega)$ then the traces of Δu and $\frac{\partial \Delta u}{\partial \boldsymbol{n}}$ are understood as members of $H^{-1/2}(\Gamma)$ and $H^{-3/2}(\Gamma)$, respectively, and it holds that

$$\langle \tfrac{\partial \Delta u}{\partial \boldsymbol{n}}, v\rangle_{H^{-3/2}(\Gamma)} - \langle \Delta u, \tfrac{\partial v}{\partial \boldsymbol{n}}\rangle_{H^{-1/2}(\Gamma)} = \int_\Omega \Delta^2 uv - \int_\Omega \Delta u \Delta v. \tag{2.129}$$

The regularity of weak solution u of (1.1) near boundary corners ν is dictated by the interior angle ω_ν and is determined by the principal linear part $\Delta^2 u$. When the boundary corners are not large, the weak solution admits full regularity. More precisely, let $k = 0$ or 1 be fixed. If the maximum angle ω for each boundary corner ν satisfies (i) $\omega < 126.28...^\circ$, if $k = 0$, or (ii) $F_i \equiv 0$ $\omega < 180^\circ$, if $k = 1$, then for $f \in H^{-k}(\Omega)$, each weak solution u belongs to $H^{4-k}(\Omega)$ with

$$\|u\|_{H^{4-k}(\Omega)} \leq C\left(\|u\|_{H^2(\Omega)} + \|f\|_{H^{-k}(\Omega)}\right). \tag{2.130}$$

As a result, the operator $\Delta^2 : H^2_0(\Omega) \cap H^3(\Omega) \to H^{-1}(\Omega)$ is invertible for any convex polygonal domain Ω. In the event the angle condition is violated, regularity will be lost near problematic vertices. In fact, without loss of generality, if ν is the only vertex with interior angle violating the assumption above, and if $\Omega_\nu \subset \Omega$ is a radial neighborhood of ν, then $u \in H^{4-k}(\Omega \backslash \Omega_\nu)$ and admits the form

$$u(r,\phi) = \tilde{u}(r,\phi) + \sum_{z_\eta \in Z} \sum_{\mu=1}^{m_\eta} a_{\eta\mu} r^{z_\eta} \ln^{(\mu-1)} r \psi_{\eta\mu}(\phi), \quad (r,\phi) \in \Omega_\nu \tag{2.131}$$

on Ω_ν where $\tilde{u} \in H^{4-k}(\Omega_\nu)$ is the regular part and the terms under the summation constitute the singular part. The terms $a_{\nu\mu}$ are complex coefficients that depend continuously on $\|u\|_{H^2(\Omega)}$ and $\|f\|_{H^{-k}(\Omega_\nu)}$ whereas the $\psi_{\nu\mu}$ are angular algebraic functions. More importantly, the powers z_η are complex and are the poles, with multiplicities m_ν, of certain resolvent operator and are distributed over the strip

$$Z = \{z \in \mathbb{C} \,|\, 1 < \operatorname{Re} z < 3 - k\},$$

and determine the regularity. The real parts $\operatorname{Re} z_\eta : [0, 2\pi] \to (3/2, \infty)$ are continuous functions of interior vertex angle ω_ν and monotonically decrease in a manner that the regularity deteriorates over $128.28...^\circ < \omega_\nu < 2\pi$ when $k = 0$ and $180^\circ < \omega_\nu < 2\pi$ when $k = -1$.

2.5 Adaptive h-refinement finite element methods

The standard adaptive mesh-refining algorithm which incorporates the ideas of Section 2.3 is an iteration of the following operations

$$\boxed{\textbf{SOLVE}} \longrightarrow \boxed{\textbf{ESTIMATE}} \longrightarrow \boxed{\textbf{MARK}} \longrightarrow \boxed{\textbf{REFINE}} \tag{2.132}$$

The module **SOLVE** computes a piecewise polynomial B-spline finite-element approximation U of u with respect to a given hierarchical mesh P. For the module **ESTIMATE**, we use a residual-based error estimator η_P derived from a posteriori analysis. The module **MARK** follows the Dörlfer marking criterion dictated by the error estimator (Dörfler [53]). Finally, the module **REFINE** consists of two steps. The first step is to obtain local refined meshes by splitting the marked elements $\mathscr{M}$ into four new cells producing a new mesh satisfying the shape regularity constraints (2.73). The second step is to refine the spline basis via B-spline subdivision wherever mesh refinement took place.

2.5.1 The AFEM modules

In what follows we discuss the modules **SOLVE**, **ESTIMATE**, **MARK** and **REFINE** in detail.

The module SOLVE

The module **SOLVE**$[P, f]$ produces an approximation U to u of problems (1.6) and (1.9). We quickly recap the the discretizations treated in this thesis. Let $\mathbb{X}_P^0 \subset H_0^2(\Omega)$ and $\mathbb{X}_P \subset H^2(\Omega)$ be the discrete spline spaces concerned in this thesis. In Chapter 3, $U \in \mathbb{X}_P^0$ is given by

$$a(U, V) = \ell_f(V), \quad V \in \mathbb{X}_P^0, \tag{2.133}$$

with a being the principal part given in Section 1 and $\alpha = 1$. In Chapter 4 $U \in \mathbb{X}_P$ is given by

$$a_P(U, V) + b(u, v) = \ell_f(V), \quad V \in \mathbb{X}_P, \tag{2.134}$$

where a_P satisfies (1.8) with α and b are given by the Stommel-Munk model. In Chapter 5, $U \in \mathbb{X}_P$ is given by (2.134) where a_P is taken to be (1.11) with α and b given by the SQGE. Finally, in Chapter 6, $U \in \mathbb{X}_P$ is given by

$$a_P(U, V) = \ell_f(V), \quad V \in \mathbb{X}_P, \tag{2.135}$$

with a_P given by (1.8) with $\alpha = 1$.

All discrete systems produced are numerically stable by coercivity. In the conforming formulation the discrete formulation (1.6) is consistent with (1.4) resulting in the Galerkin orthogonality:

$$a(u - U, V) = 0 \quad \forall V \in \mathbb{X}_P^0, \tag{2.136}$$

which is contrary to discretizations of Chapters 5 and 6, where

$$a_P(u - U, V) = \langle \mathscr{E}_P, V \rangle \quad \forall V \in \mathbb{X}_P, \tag{2.137}$$

for a suitable inconsistency functional $\mathscr{E}_P$. In all our discritizations a Cea-type estimate can be realized, where the most challenging of all was in the treatment of the SQGE model of Chapter 5. For this reason we include a complete *a priori* error analysis. For the remaining chapters similar estimates hold by standard arguments which we omit.

The module ESTIMATE

The module **ESTIMATE**$[U, P]$ produces a collection of *error indicators* η_τ for each element $\tau \in P$, given by

$$\eta_P^2(V,\tau) = h_\tau^4 \|R_\tau\|_{L^2(\tau)}^2 + \sum_{\sigma\subset\partial\tau} \left(h_\sigma^3 \,\|J_{\sigma,1}\|_{L^2(\sigma)}^2 + h_\sigma \|J_{\sigma,2}\|_{L^2(\sigma)}^2 \right), \tag{2.138}$$

where interior residual quantity

$$R_\tau = (f - \mathcal{L}V\,)\,|_\tau, \quad (\tau \in P), \tag{2.139}$$

edge jump terms

$$J_{\sigma,1} = \left| \left[\!\!\left[\frac{\partial \Delta V}{\partial \boldsymbol{n}_\sigma} \right]\!\!\right] \right|_\sigma \quad \text{and} \quad J_{\sigma_2} = [\![\Delta V]\!]_\sigma\,, \quad (\sigma \in \mathcal{E}_P), \tag{2.140}$$

and $\mathcal{L}V = F_0(\cdot, V, \nabla V, \Delta V) + \mathrm{div}\boldsymbol{F}(\cdot, V, \nabla V, \Delta V)$. Here, $[\![V]\!]_\sigma$ evaluates the jump of V across interface σ:

$$[\![V]\!]_\sigma\,(x) = \lim_{t\to 0}[V(x + t\boldsymbol{n}_\sigma) - V(x - t\boldsymbol{n}_\sigma)], \quad x \in \sigma. \tag{2.141}$$

The *error estimator* over any subset $\omega \subset \Omega$ is given by

$$\eta_P^2(V,\omega) = \sum_{\tau\in P:\tau\subset\omega} \eta_P^2(V,\tau). \tag{2.142}$$

We define the *data oscillation* term over a subset $\omega \subseteq \Omega$ by

$$\mathrm{osc}_P^2(f,\omega) = \sum_{\tau\in P:\tau\subset\omega} h_\tau^4 \|(\mathrm{id} - \Pi_P^m) f\|_{L^2(\tau)}^2. \tag{2.143}$$

The module MARK

A number of ways of selecting cells to refine have been explored, among the most notable is the *Maximum strategy* selecting all cells with error exceeding $\theta \times 100\%$ of the largest cell error, and the *Modified equidistribution strategy* aiming to replicate the equidistribution principle realized in (Babuška and Rheinboldt [47]). The *Dörlfer marking strategy* (Dörfler [53]) is shown to be an optimal marking strategy (Stevenson [60]). For prescribed $0 < \theta \leq 1$, the strategy reads

$$\text{Find smallest subset } \mathscr{M} \subseteq P : \quad \sum_{\tau \in \mathscr{M}} \eta_P^2(U, \tau) \geq \theta \sum_{\tau \in P} \eta_P^2(U, \tau). \tag{2.144}$$

This method suppresses excessive iterations and over-refinement by fixing the ratio between good and bad cells throughout all iterations.

To ensure minimal cardinality of $\mathcal{M}$ in the marking strategy one typically undergoes QuickSort which has an average complexity of $\mathcal{O}(n \log n)$ to produce the indexing set J. On the other hand, the log-factor can be avoided using a bucket sort, placing indicators in error range bins; see Stevenson [60] for more details. When $\theta = 1$ the refinement is uniform.

The module REFINE

The refinement is designed to recursively extend the marked cells $\mathscr{M}$ obtained from module **MARK** to a set $\omega_{R_{P \to P_*}}$ for which the new mesh P_* is admissible. We define the *neighbourhood* of $\tau \in P \cap \mathcal{G}_\ell$ as

$$\mathcal{N}(P, \tau) = \{\tau' \in P \cap \mathcal{G}_{\ell-1} : \exists \tau'' \in S(\tau, \ell),\ \tau'' \subseteq \tau'\}, \tag{2.145}$$

when $\ell - 1 > 0$, and $\mathcal{N}(P, \tau) = \varnothing$ otherwise. To put in concrete terms, the neighbourhood $\mathcal{N}(P, \tau)$ of an active cell in $\mathcal{G}_\ell$ consist of active cells τ' of level $\ell - 1$ overlapping the support extension of τ with respect to level ℓ. Procedure **REFINE** will ensure that for a constant $c_{\text{shape}} > 0$, depending only on the polynomial degree of the spline space, all considered partitions therefore will satisfy the shape-regularity constraints (2.73).

Algorithm 2.5.1 Recursive refinement **recursive_refine** $[P, \tau] \to P_*$

```
for all $\tau' \in \mathcal{N}(P, \tau)$ do
    $P \leftarrow$ recursive_refine $[P, \tau']$
end for
if $\tau \in P$ then
    $\{\tau_j\}_{j=1}^4 \leftarrow$ dyadic-refine $\tau$
    $P_* \leftarrow (P \backslash \tau) \cup \{\tau_j\}_{j=1}^4$
end if
```

Algorithm 2.5.2 Carry admissible mesh refinement **REFINE** $[P, \mathscr{M}] \to P_*$

```
for $\tau \in \mathscr{M}$ do
    $P \leftarrow$ recursive_refine $[P, \tau]$
end for
$P_* \leftarrow P$
```

2.5.2 Performance analysis

We conclude this section by applying Theorem 2.46 to various scenarios reflecting subsequent numerical experiments. We assume in all cases that $f \in L^2(\Omega)$ so that any singularity that compromises solution's regularity is entirely determined by the boundary's vertex ν with largest interior angle ω_ν. We consider splines with degrees $r \geq 2$ and explore the anticipated convergence rates when using linear and nonlinear approximation methods on a rectangular domain ($\omega_\nu = \frac{\pi}{2}$), where full smoothness is ensured, and, an L-shaped domain ($\omega_\nu = \frac{3\pi}{2}$) where a singularity in higher-order derivatives form at the central corner. From the regularity results of [3] we expect that $u \in H^4(\Omega)$ in the rectangular domain. Determining the Sobolev regularity for u in the L-shaped domain can be obtained from the following heuristic argument. Proceeding in polar coordinates, the singular part of u in (2.131) behaves approximately like $\rho^{1.6}$ near the obtuse vertex; see in [3] that $\mathrm{Re} z \approx 1.6$. We estimate $|u|_{H^\alpha(\Omega_\nu)}$ using $\int_0^1 \rho^{2(1.6-\alpha)} \rho \, d\rho$. The integral is finite whenever $\alpha < 2.6$ and therefore we expect that $u \in H^{2.6}(\Omega)$. We now look at the anticipated rates. If u is a solution on a rectangular mesh and we carry approximation using uniform refinement, then the linear method is expected to converge such that

$$\|u - U\|_{H^2(\Omega)} \preceq |h_P^2 u|_{H^4(\Omega)} \preceq N^{-1} |u|_{H^4(\Omega)}. \tag{2.146}$$

On the other hand, if $u \in H^{2.6}(\Omega)$, being a solution on the L-shape geometry, we expect that

$$\|u - U\|_{H^2(\Omega)} \preceq |h_P^{0.6} u|_{H^{2.6}(\Omega)} \preceq N^{-0.3} |u|_{H^{2.6}(\Omega)}. \tag{2.147}$$

These expectations are consistent with Theorem 2.46. Indeed, $\mathscr{B}^4_{2,2} \hookrightarrow \mathscr{A}^1$ whereas $\mathscr{B}^{2.6}_{2,2} \hookrightarrow \mathscr{A}^{0.3}$ which is consistent with the anticipated rates one can expect from linear methods. On the other hand, when a nonlinear approach is adopted via h-refinement adaptivity, we first note that $\mathscr{B}^4_{p,p}(\Omega) \hookrightarrow H^2(\Omega)$ when $p = \frac{2}{3}$. In view of Theorem 2.46 $u \in \mathscr{B}^4_{p,p}(\Omega) \hookrightarrow \mathscr{A}^1$. In the language discussed in Example 2.25, constructing a mesh that equidistributes the $\mathscr{B}^4_{\frac{2}{3},\frac{2}{3}}(\Omega)$-norm of u over all partition cells yields

$$\|u - U\|_{H^2(\Omega)} \preceq N^{-1} |u|_{\mathscr{B}^4_{\frac{2}{3},\frac{2}{3}}(\Omega)}. \tag{2.148}$$

We depict those realizations in a DeVore diagram of Figure 2.6.

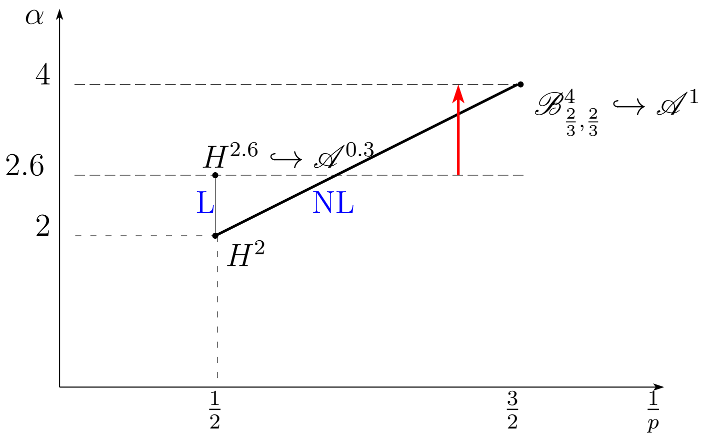

α
4
2.6
2
$H^{2.6} \hookrightarrow \mathscr{A}^{0.3}$
$\mathscr{B}^4_{\frac{2}{3},\frac{2}{3}} \hookrightarrow \mathscr{A}^1$
L
NL
H^2
$\frac{1}{2}$
$\frac{3}{2}$
$\frac{1}{p}$

Figure 2.6: *DeVore diagram indicating the embedding of various solutions, given their Sobolev norms, into approximation classes. With abuse of notation kept in mind, $\mathscr{A}^{0.3}$ and $\mathscr{A}^1$ correspond to approximation using uniform and adaptive refinement, respectively. The length of the red arrow indicates gain in convergence rate owed to adaptive refinement.*

Chapter 3

A quasi-optimal boundary-condition conforming adaptive spline-based finite element method for the bi-Laplace operator

3.1 Introduction

We address the performance of an adaptive finite element method for the biharmonic problem using boundary condition conforming B-splines. In this chapter, we prove that the adaptive procedure (2.132) guided by the *a posteriori* error estimator converges with optimal rate.

In the pioneering work of (Babuška and Rheinboldt [47]) a mathematical theory is developed for a class of *a posteriori* error estimates for solutions of finite element methods providing a blue-print for h-adaptive mesh generation strategies. Under very general assumptions, the numerical error is estimated in terms of localized quantities which can be computed approximately. While the existence of optimal partitions was not addressed until the early 2000's by Binev, DeVore and Dahmen, their work lead to a heuristic characterization of optimal meshes as ones that equally distributes the numerical error over all the partition elements. The error estimator class of Babuška and Rheinboldt were sharp in that they formed *upper* and *lower bounds* up to fixed proportionality constants, but their calculation require solving local Dirichlet-type problems over every cell, in every iteration. Simultaneously, Babuška and Rheinboldt derived an explicitly defined estimator in [48] based on weighted combinations of suitable *residual* quantities which approximate the loss of smoothness brought by singularities in the derivatives. Later in (Verfürth [50]), when treating various boundary value problems for the Poisson equation, the idea of residual-based estimation devised by Babuška and Rheinboldt was made systematic using integration by parts. This will be the framework used in this thesis. The avoidance of local problems had an impact on the sharpness of these residual-based error indicators; compare equations (4.8) and (4.9) of Proposition 4.2 in (Verfürth [50]) with equation (3.3a) of Theorem 3.2 in (Babuška and Rheinboldt [47]). The *global upper bound* (see for example Lemma 3.2) and *local lower bound* (see for example Eq. (3.58)) were compromised by initial data. Later in (Morin et al [54]), when studying convergence, it was discovered that the sacrifice in sharpness is intrinsic and cannot be overcome.

All convergence results in AFEM literature are realized as a contraction of some numerical error quantity with each iteration. The first Adaptive Finite Element Method (AFEM) convergence result was given in (Babuška [49]) when treating general second order symmetric elliptic ODEs. Convergence of AFEM for the Poisson equation for general Dirichlet boundary conditions in two-dimensions was given in seminal work of (Dörfler [53]) with the advent of the novel Dörfler marking strategy and an initial-mesh *fineness* assumption placed to ensure problem datum, the right-hand side source function and boundary condition values to be sufficiently resolved before executing the adaptive loop. As a result, the initial mesh assumption captures the finer features of the initial data thus guaranteeing sharp *a posteriori* local estimation in all subsequent iterations and strict numerical error reduction is achieved with each iteration. However, it did not exclude the possibility of an over-refined initial mesh. In the case of homogeneous boundary conditions, the initial-mesh finess assumption read $\|h_P f\|_{L^2(\Omega)}$.

Later in (Morin et al [54]; see also Morin et al [55]), when treating the more general second order elliptic operator $-\mathrm{div}(\boldsymbol{A}\nabla u)$ with coefficient matrix $\boldsymbol{A}$ taken to be piecewise constant with respect to the initial mesh, the authors came to the realization that the discretization of the prescribed data results in averaging of finer initial data features which interferes with the numerical error reduction irrespective of quadrature. This averaging error manifests into a weighted L^2-error projection error $\|h_P(f-\bar{f})\|_{L^2(\Omega)}$, where $\bar{f}$ is the L^2 average of f, initially observed in (Verfürth [50]). In contrast to (Dörfler [53]), it is the *oscillation* of f from $\bar{f}$ and not the size of $\|h_P f\|_{L^2(\Omega)}$ that's important. As a result, the initial mesh assumption was removed and replaced with a Dörfler-inspired *separate marking* strategy that ensures data oscillation to be sufficiently small before entering a new iteration. In addition to controlling the data oscillation, the subdivision of each marked element needed to be performed multiple times so as to produce an interior node amounting to a total of three bisections in addition to those produced by the completion step. This refinement rule gives rise to a so-called *interior node property* and provides a *discrete* counterpart to the local lower bound to (Verfürth [50]). The combination of separate marking and the interior node property in each refinement makes up the Morin-Nochetto-Sierbert (MNS) algorithm and successfully recovers strict error reduction with every iteration while circumventing an overly-refined initial mesh.

At the heart of all modern AFEM convergence results is a Pythagoras-type relation relating the numerical error between two refinement levels. The geometric relation is primarily a consequence of finite element space nesting and the discrete formulation consistency with the weak problem. When the bilinear form is symmetric, the relation is an equality (see for example Lemma 3.11), otherwise it is not. This was first addressed in (Mekchay et al. [56]) where convergence results was extended to general non-symmetric second order elliptic problems with variable coefficient matrix $\boldsymbol{A} : \Omega \to \mathbb{R}^{d\times d}$ Lipschitz, symmertic positive-definite with bounded eigenvalues. Mekchay et al. brought to light new ideas to the field. Firstly, the Galerkin-Pythagoras equality broken by the asymmetry still fulfilled its purpose at the price of some initial-mesh condition (vaguely analogous to that of (Dörfler [53])) that depends on the non-symmetric term. On the other hand, the variable nature of the coefficient matrix resulted in the incorporation of the coefficient matrix into the oscillation term. As a result, the numerical error and oscillation term can no longer be treated separately and resulted in a novel new convergence argument.

The rate of convergence of the MNS algorithm was studied in the landmark publication of (Binev, DeVore and Dahmen [59]). Their discovery of controlled complexity of conforming refinement (2.107) and their introduction of a *coarsening step* laid the foundations for estimating the rate of convergence in terms of the number of partition elements. Another fact that they brought to light is that the approximation class containing the solution u was not enough into ensure the rate of convergence; the oscillation term should also be taken to account to obtain a complete picture of the AFEM's optimality. The coarsening step was intended to mitigate the cost that came from enforcing the interior node property. Later however, the coarsening step was shown to be artificial in the important paper by (Stevenson [60]). Significant fundamental changes resulted from (Stevenson [60]) in addition to the removal of the coarsening. The Dörfler marking was shown to be optimal, in virtue of a *discrete upper bound* estimate (see for example Lemma 3.7), making the mesh complexity analysis significantly simpler than in (Binev, DeVore and Dahmen [59]).

For the better part of the 2000's the MNS algorithm stood as state of the art for adaptive finite element methods for linear elliptic problems. It was in seminal paper of (Cascón et al. [61]) where convergence was achieved without a need for the discrete lower bound. The discrete lower bound was replaced with a weaker *estimator error reduction* estimate (see for example Lemma 3.9) that neither included an oscillation term nor did it require an interior node property resulting in the ultimate removal of the costly refinement condition as well as separate marking for oscillation. Subsequently, convergence is shown to hold in a *quasi-error* norm (see (3.95)). The total- and quasi- error norms are equivalent in the asymptotic regime, and the convergence in quasi-norm is sufficient to achieve the desired result: quasi-optimality of AFEM in total-error.

We now begin with describing the set-up of this chapter. The energy norm $\|\!|\!| \cdot \|\!|\!| =$

$\|\Delta \cdot \|_{L^2(\Omega)}$, owing to the Poincaré inequality is a norm on $\mathbb{V} = H_0^2(\Omega)$. The B-spline based finite element space is given by

$$\mathbb{X}_P = \mathbb{S}_P^r \cap H_0^2(\Omega) \subset \mathcal{C}^1(\Omega), \quad r \geq 2. \tag{3.1}$$

The *a priori* error estimation is well-known for the boundary-condition conforming discretization and follows from a standard derivation of Céa's lemma. We limit ourselves with mentioning it here: If u and U are the weak and discrete solutions, respectively, then

$$\|\!|u - U|\!\| \leq \frac{C_{\text{cont}}}{C_{\text{coer}}} \inf_{V \in \mathbb{X}_P} \|\!|u - V|\!\|, \tag{3.2}$$

and if furthermore $u \in H^s(\Omega)$ with $2 < s < r + 1$ then,

$$\|\!|u - U|\!\| \leq \frac{c_{\text{shape}} C_{\text{cont}}}{C_{\text{coer}}} |u|_{H^s(\Omega)}. \tag{3.3}$$

The constants $C_{\text{cont}} > 0$ and $C_{\text{coer}} > 0$ are the bilinear form's continuity and coercivity coefficients, and c_{shape} is the partition shape-regularity coefficient.

3.2 *A posteriori* error estimation

We define the residual quantity $\mathscr{R}_P \in \mathbb{V}'$ by

$$\langle \mathscr{R}_P, v \rangle = a(u - U, v), \quad v \in \mathbb{V}. \tag{3.4}$$

In view of continuity and coercivity of the bilinear form we have

$$C_{\text{cont}}^{-1} \|\mathscr{R}_P\|_{\mathbb{V}'} \leq \|\!|u - U|\!\| \leq C_{\text{coer}}^{-1} \|\mathscr{R}_P\|_{\mathbb{V}'}. \tag{3.5}$$

The quantity $\|\mathscr{R}_P\|_{\mathbb{V}'}$ is computable since it only depends on available discrete approximation of solution u. We follow the techniques devised in (Verfürth [50]) to estimate $\|\mathscr{R}_P\|_{\mathbb{V}'}$.

Estimator reliability

In the following, we estimate the residual $\|\mathscr{R}_P\|_{\mathbb{V}'}$ first by deriving and L^2-representation for residual quantity $\mathscr{R}_P$ then prove the reliability of the proposed error estimator.

Lemma 3.1 (Residual L^2-representation). *The functional $\mathscr{R}_P \in \mathbb{V}'$ admits the L^2-representation as*

$$\langle \mathscr{R}_P, v \rangle = \sum_{\tau \in P} \int_\tau R_\tau v - \sum_{\sigma \in \mathcal{E}_P} \left(\int_\sigma J_{\sigma,1} v - \int_\sigma J_{\sigma 2} \tfrac{\partial v}{\partial \boldsymbol{n}_\sigma} \right) \tag{3.6}$$

where R_τ and $J_{\sigma,i}$ are defined in (5.89) *and* (5.90)*.*

Proof. Let $v \in \mathbb{V}$. Performing partial integration on each cell $\tau \in P$ yields

$$\langle \mathscr{R}_P, v\rangle = \sum_{\tau \in P} \int_\tau (f - \mathcal{L}U)\, v - \alpha \sum_{\tau\in P} \left(\oint_{\partial\tau} \tfrac{\partial \Delta U}{\partial \boldsymbol{n}_\tau} v - \oint_{\partial\tau} \Delta U \tfrac{\partial v}{\partial \boldsymbol{n}_\tau} \right). \tag{3.7}$$

Let τ_1 and τ_2 be cells sharing an interior edge σ with corresponding outward unit normal vectors $\boldsymbol{n}_{\tau_1}$ and $\boldsymbol{n}_{\tau_2}$ and $\boldsymbol{n}_{\tau_1} = -\boldsymbol{n}_{\tau_2}$. By setting $\boldsymbol{n}_\sigma = \boldsymbol{n}_{\tau_1}$, we have

$$\begin{aligned} \int_{\partial\tau_1\cap\sigma} \Delta U \frac{\partial v}{\partial \boldsymbol{n}_{\tau_1}} &+ \int_{\partial\tau_2\cap\sigma} \Delta U \frac{\partial v}{\partial \boldsymbol{n}_{\tau_2}} \\ &= \int_\sigma \left[(\Delta U)_{\tau_1} (\nabla v \cdot \boldsymbol{n}_\sigma) - (\Delta U)_{\tau_2} (\nabla v \cdot \boldsymbol{n}_\sigma) \right] \\ &= \int_\sigma \left[(\Delta U)_{\tau_1} - (\Delta U)_{\tau_2} \right] (\nabla v \cdot \boldsymbol{n}_\sigma) = \int_\sigma [\![\Delta U]\!]_\sigma \nabla v \cdot \boldsymbol{n}_\sigma. \end{aligned} \tag{3.9}$$

Similarly, we obtain

$$\int_{\partial\tau_1\cap\sigma} \frac{\partial \Delta U}{\partial \boldsymbol{n}_{\tau_1}} v + \int_{\partial\tau_2\cap\sigma} \frac{\partial \Delta U}{\partial \boldsymbol{n}_{\tau_2}} v = \int_\sigma \left[\!\left[\frac{\partial \Delta U}{\partial \boldsymbol{n}_\sigma} \right]\!\right]_\sigma v. \tag{3.10}$$

Upon summation of (3.9) and (3.10) over all cells τ we can write

$$\sum_{\tau\in P} \oint_{\partial\tau} \Delta U \frac{\partial v}{\partial \boldsymbol{n}_\tau} = \sum_{\sigma\in\mathcal{E}_P} \int_\sigma [\![\Delta U]\!]_\sigma \frac{\partial v}{\partial \boldsymbol{n}}, \tag{3.11}$$

and

$$\sum_{\tau\in P} \oint_{\partial\tau} \frac{\partial \Delta U}{\partial \boldsymbol{n}_\tau} v = \sum_{\sigma\in\mathcal{E}_P} \int_\sigma \left[\!\left[\frac{\partial \Delta U}{\partial \boldsymbol{n}_\sigma} \right]\!\right]_\sigma v. \tag{3.12}$$

By applying (5.107) and (5.108) to (5.106), we obtain

$$\alpha \sum_{\tau\in P} \left(\oint_{\partial\tau} \frac{\partial \Delta U}{\partial \boldsymbol{n}_\tau} v - \oint_{\partial\tau} \Delta U \frac{\partial v}{\partial \boldsymbol{n}_\tau} \right) = \sum_{\sigma\in\mathcal{E}_P} \left(\int_\sigma J_{\sigma,1} v - \int_\sigma J_{\sigma,2} \frac{\partial v}{\partial \boldsymbol{n}_\sigma} \right). \tag{3.13}$$

■

We now prove that the estimator is reliable in that it forms a global upper bound to the numerical error.

Lemma 3.2 (Global upper bound). *Let P be an admissible partition of Ω. The module* **ESTIMATE** *produces an* a posteriori *error estimate η_P for the discrete error such that for a constants $C_U > 0$,*

$$|\!|\!| u - U |\!|\!|^2 \leq C_U \eta_P^2(U, \Omega), \tag{3.15}$$

with constants depending only on c_{shape}, C_{cont} and C_{coer}.

Proof. We have from the previous Lemma

$$\begin{aligned}|\langle \mathscr{R}_P, v\rangle| \le \sum_{\tau\in P} \|R_\tau\|_{L^2(\tau)}\|v - Q_P v\|_{L^2(\tau)} &+ \sum_{\sigma\in\mathcal{E}_P} \|J_{\sigma,1}\|_{L^2(\sigma)}\, \|(v - Q_P v)\|_{L^2(\sigma)} \\ + \sum_{\sigma\in\mathcal{E}_P} \|J_{\sigma,2}\|_{L^2(\sigma)} \left\| \tfrac{\partial}{\partial \boldsymbol{n}_\sigma}(v - Q_P v)\right\|_{L^2(\sigma)}&\end{aligned} \tag{3.17}$$

We use the approximation results from Lemma 2.42 to estimate interior residual terms

$$\begin{aligned}\sum_{\tau\in P} \|R_\tau\|_{L^2(\tau)}\|v - Q_P v\|_{L^2(\tau)} &\le \sum_{\tau\in P} \|R_\tau\|_{L^2(\tau)} c_{\text{shape}} h_\tau^2 |v|_{H^2(\omega_\tau)}, \\ &\le c_{\text{shape}}\left(\sum_{\tau\in P} h_\tau^4 \|R_\tau\|_{L^2(\tau)}^2\right)^{1/2}\left(\sum_{\tau\in P} |v|_{H^2(\omega_\tau)}^2\right)^{1/2}, \\ &= c_{\text{shape}}\left(\sum_{\tau\in P} h_\tau^4 \|R_\tau\|_{L^2(\tau)}^2\right)^{1/2} |v|_{H^2(\Omega)}.\end{aligned} \tag{3.19}$$

where in the last step we used (2.74). As for the interior edge jump terms,

$$\begin{aligned}\sum_{\sigma\in\mathcal{E}_P} &\|J_{\sigma,1}\|_{L^2(\sigma)}\, \|v - Q_P v\|_{L^2(\sigma)} \\ &\le \sum_{\sigma\in\mathcal{E}_P} \|J_{\sigma,1}\|_{L^2(\sigma)}\, c_{\text{shape}} h_\sigma^{3/2} |v|_{H^2(\omega_\sigma)}, \\ &\le c_{\text{shape}}\left(\sum_{\sigma\in\mathcal{E}_P} h_\sigma^3 \|J_{\sigma,1}\|_{L^2(\sigma)}^2\right)^{1/2} |v|_{H^2(\Omega)},\end{aligned} \tag{3.21}$$

and similarly,

$$\sum_{\sigma\in\mathcal{E}_P} \|J_{\sigma,2}\|_{L^2(\sigma)} \left\|\tfrac{\partial}{\partial \boldsymbol{n}_\sigma}(v - Q_P v)\right\|_{L^2(\sigma)} \le c_{\text{shape}}\left(\sum_{\sigma\in\mathcal{E}_P} h_\sigma \|J_{\sigma,2}\|_{L^2(\sigma)}^2\right)^{1/2} |v|_{H^2(\Omega)}. \tag{3.22}$$

Summing up we arrive at

$$\begin{aligned}\frac{|\langle \mathscr{R}_P, v\rangle|}{\|v\|_{\mathbb{V}}} \le c_{\text{shape}} \Bigg\{ &\left(\sum_{\tau\in P} h_\tau^4 \|R_\tau\|_{L^2(\tau)}^2\right)^{1/2} + \left(\sum_{\sigma\in\mathcal{E}_P} h_\sigma^3\, \|J_{\sigma,1}\|_{L^2(\sigma)}^2\right)^{1/2} \\ &+ \left(\sum_{\sigma\in\mathcal{E}_P} h_\sigma \|J_{\sigma,2}\|_{L^2(\sigma)}^2\right)^{1/2} \Bigg\},\end{aligned} \tag{3.24}$$

and the desired result follows from the equivalence (3.5). ■

Estimator efficiency

The efficiency of the estimator, which is the L^2 form of the lower bound in (3.5), ensures that the estimator is sharp and equivalent to the numerical error up to the oscillation term. Arriving at this requires using certain bubble functions that allows us to isolate the residual quantities from one another.

Lemma 3.3 (Bubble functions). *Let $\tau \in P$ and let $\sigma = \partial\tau_1 \cap \partial\tau_2$ for $\tau_i \in P$ and let $D_\sigma = \overline{\tau_1 \cup \tau_2}$. There exists piecewise polynomial functions ψ_τ be any smooth cut-off such that*

$$\operatorname{supp}\psi_\tau \subseteq \tau, \quad \psi_\tau \geq 0, \quad \max_{x\in\tau}\psi_\tau(x) \leq 1, \tag{3.25}$$

and two C^1 cut-off functions $\psi_\sigma \geq 0$ and χ_σ, both supported on D_σ, such that

$$\tfrac{\partial\psi_\sigma}{\partial \boldsymbol{n}_\sigma} \equiv 0, \quad \chi_\sigma \equiv 0 \quad \textit{and} \quad d_1 h_\sigma^{-1}\psi_\sigma \leq \tfrac{\partial\chi_\sigma}{\partial \boldsymbol{n}_\sigma} \leq d_2 h_\sigma^{-1}\psi_\sigma \quad \textit{along } \sigma. \tag{3.26}$$

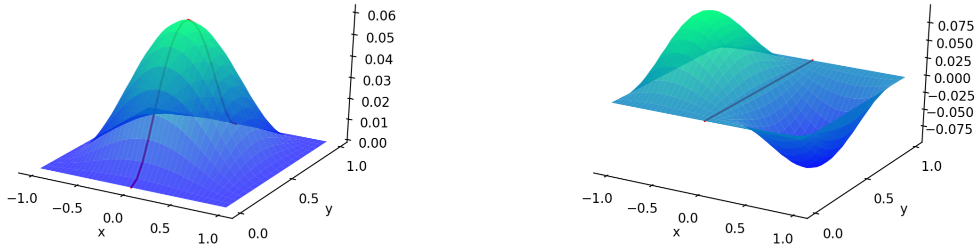
Proof. The construction of the first bubble function ψ_τ is easy to see. We focus on the edge-based ones. Let $H(x,y) = y^2(1-y)^2(1-x^2)^2$ and define $\hat{\psi}$ and $\hat{\chi}$ by

$$\hat{\psi}(x,y) = H(x,y)\mathbb{1}_{x\geq 0} + H(-x,y)\mathbb{1}_{x\leq 0} \quad \text{and} \quad \hat{\chi} = \frac{\partial\hat{\psi}(x,y)}{\partial x} \quad (x,y) \in \hat{D} := [-1,1]\times[0,1]. \tag{3.27}$$

See Figure 3.1 for illustrations of the bubble functions. We focus on $\hat{\tau} = (0,1)\times(0,1)$. It is also easy to verify that $\hat{\chi}|_{\hat{\tau}} = \frac{\partial H}{\partial x} = -\frac{4x}{1-x^2}H(x,y)$ which means $\hat{\chi}|_{\hat{\sigma}} = 0$, where $\hat{\sigma} = \{0\}\times(0,1)$. Moreover, we also have $\frac{\partial(\hat{\chi}|_{\hat{\tau}})}{\partial x}|_{\hat{\sigma}} = -4H(0,y) = -4\hat{\psi}|_{\hat{\sigma}}$. Now let $\hat{\tau}_1 = \hat{\tau}$, let $\hat{\tau}_2 = (-1,0)\times(0,1)$ and let $\boldsymbol{n} = (1,0)$ be the unit normal vector associated with $\hat{\sigma}$. Finally, let F_σ be the affine transformation that maps $\hat{D}$ onto D_σ and define

$$\psi_\sigma = \hat{\psi}\circ F_\sigma^{-1}, \quad \chi_\sigma = \hat{\chi}\circ F_\sigma^{-1}. \tag{3.28}$$

■



(a) The bubble function $\hat{\psi}$

(b) The bubble function $\hat{\chi}$

Figure 3.1: *The bubble functions on the reference subdomain $\hat{D}$. The values of $\hat{\psi}$ ad $\hat{\chi}$ along the edge $\hat{\sigma}$ adjacent to two reference cells $\hat{\tau}_1$ and $\hat{\tau}_2$ are shown in red.*

Lemma 3.4 (Localizing estimates). *Let τ be a cell in partition P. For a constant c_m depending only on polynomial degree m,*

$$\|q\|^2_{L^2(\tau)} \leq c_m \int_\tau \psi_\tau q^2 \quad \forall q \in \mathbb{P}_m(\tau). \tag{3.29}$$

Let σ be an edge in $\mathcal{E}_P$ for which $\sigma \subseteq \partial\tau$. We also have

$$\|q\|^2_{L^2(\sigma)} \leq c_m \int_\sigma \psi_\sigma q^2 \tag{3.30}$$

and

$$\|\psi_\sigma^{1/2} E_\sigma q\|_{L^2(\tau)} \leq c_m h_\sigma^{1/2} \|q\|_{L^2(\sigma)} \tag{3.31}$$

holding for every $q \in \mathbb{P}_m(\sigma)$.

Proof. Relations (3.29) and (3.30) are proven in the same fashion as in (Verfürth [50]). We focus on (3.31). We will define extension operators $E_\sigma : C(\sigma) \to C(\tau)$ for all edges σ with $\sigma \subset \partial\tau$. Let $\hat{\tau} = [0,1] \times [0,1]$ and $\hat{\sigma} = \{0\} \times [0,1]$. Let $F_\tau : \mathbb{R}^2 \to \mathbb{R}^2$ be the affine transformation comprising of translation and scaling mapping $\hat{\tau}$ onto τ and $\hat{\sigma}$ onto σ. Define $\hat{E} : C(\hat{\sigma}) \to C(\hat{\tau})$ via

$$\hat{E}v(x,y) = v(x) \quad \forall x \in \hat{\sigma}, \quad (x,y) \in \hat{\tau}, \quad v \in C(\hat{\sigma}). \tag{3.32}$$

To this end, let σ be an edge of a cell $\tau \in P$, then define $E_\sigma : C(\sigma) \to C(\tau)$ via

$$E_\sigma v = [\hat{E}(v \circ F_\tau)] \circ F_\tau^{-1}. \tag{3.33}$$

In other words extending the values of v from σ into τ along inward $\boldsymbol{n}_\sigma$. We now prove that $q \mapsto \|\psi_\sigma^{1/2} E_\sigma q\|_{L^2(\tau)}$ is a norm on $\mathbb{P}_m(\sigma)$.

$$\|E_\sigma q\|_{L^2(\tau)} = |\tau|^{1/2} \|\hat{E}(q \circ F_\tau)\|_{L^2(\hat{\tau})}. \tag{3.35}$$

It is clear that $\hat{q} \in \mathbb{P}_m(\hat{\sigma})$ is identically zero if and only if its extension $\hat{E}\hat{q}$ is identically zero on $\hat{\tau}$. So $\hat{q} \mapsto \|\hat{E}\hat{q}\|_{L^2(\hat{\tau})}$ is an equivalent norm on $\mathbb{P}_m(\hat{\sigma})$ we have

$$\|\hat{E}(q \circ F_\tau)\|_{L^2(\hat{\tau})} \preceq \|q \circ F_\tau\|_{L^2 \hat{\sigma}} = h_\sigma^{-1/2} \|q\|_{L^2(\sigma)} \tag{3.36}$$

so with $|\tau|^{1/2} h_\sigma^{1/2} \preceq h_\sigma^{1/2}$

$$\|\psi_\sigma^{1/2} E_\sigma q\|_{L^2(\tau)} \preceq h_\sigma^{1/2} \|q\|_{L^2(\sigma)}. \tag{3.37}$$

■

We now have all the tools we need to prove the estimator efficiency estimate.

Remark 3.5. If $\overline{R_\tau} = \Pi_P^m R_\tau$ for $m \leq r-4$,

$$R_\tau - \overline{R_\tau} = (\mathrm{id} - \Pi_P^m) f - (\mathrm{id} - \Pi_P^m) \mathcal{L} V = (\mathrm{id} - \Pi_P^m) f. \tag{3.38}$$

Note that $\mathcal{L}V|_\tau \in \mathbb{P}_{r-4}$.

Lemma 3.6 (Global Lower Bound). *Let P be an admissible partition of Ω. The module* **ESTIMATE** *produces an* a posteriori *error estimate of the discrete solution error such that*

$$C_L \eta_P^2(U, \Omega) \leq \|\!|u - U|\!\|^2 + \mathrm{osc}_P^2(f, \Omega), \tag{3.39}$$

with constant C_L depending only on c_{shape} and polynomial degree r.

Proof. The proof is carried out by localizing the error contributions coming from the cells residuals R_τ and edge jumps $J_{\sigma,1}$ and $J_{\sigma,2}$. For $\tau \in P$ let $\psi_\tau \in H_0^2(\tau)$ be as in (3.25) and let $\overline{R_\tau}$ be a polynomial approximation of R_τ by means of the L^2-orthogonal projector Π_P^m with $m \leq r-4$. Using the norm-equivalence relation (3.29) of Lemma 3.4

$$\|\overline{R_\tau}\|_{L^2(\tau)}^2 \leq c_m \int_\tau \overline{R_\tau} (\overline{R_\tau} \psi_\tau) = c_m \|\overline{R_\tau}\|_{H^{-2}(\tau)} \|\overline{R_\tau} \psi_\tau\|_{H^2(\tau)}. \tag{3.40}$$

From Lemma 2.36 and equation (3.25), $\|\overline{R_\tau}\psi_\tau\|_{H^2(\tau)} \leq c_m h_\tau^{-2} \|\overline{R_\tau}\|_{L^2(\tau)}$ and expression (3.40) now reads $h_\tau^2 \|\overline{R_\tau}\|_{L^2(\tau)} \leq c_m \|\overline{R_\tau}\|_{H^{-2}(\tau)}$. It follows that

$$\begin{aligned} h_\tau^2 \|R_\tau\|_{L^2(\tau)} &\leq h_\tau^2 \|\overline{R_\tau}\|_{L^2(\tau)} + h_\tau^2 \|R_\tau - \overline{R_\tau}\|_{L^2(\tau)}, \\ &\leq c_m \|\overline{R_\tau}\|_{H^{-2}(\tau)} + h_\tau^2 \|(\mathrm{id} - \Pi_P^m) f\|_{L^2(\tau)}, \\ &\leq c_m \left(\|R_\tau\|_{H^{-2}(\tau)} + \|(\mathrm{id} - \Pi_P^m) f\|_{H^{-2}(\tau)} \right) + h_\tau^2 \|(\mathrm{id} - \Pi_P^m) f\|_{L^2(\tau)}. \end{aligned} \tag{3.42}$$

We know that embedding $L^2(\tau) \hookrightarrow H^{-2}(\tau)$ is continuous. Moreover, we can view R_τ as a restriction of $\mathscr{R}_P$ to τ and therefore in view of definition (2.143) we may right

$$h_\tau^2 \|R_\tau\|_{L^2(\tau)} \preceq c_m \|\mathscr{R}_P\|_{H^{-2}(\tau)} + \operatorname{osc}_P(f,\tau). \tag{3.43}$$

We turn our attention to the jump terms across the interior edges. We begin with the edge residual $J_{\sigma,1}$. Let an edge $\sigma \in \mathcal{E}_P$ and cells $\tau_1, \tau_2 \in P$ be such that $\sigma \subset \partial\tau_1 \cap \partial\tau_2$ and denote $D_\sigma = \overline{\tau_1 \cup \tau_2}$. If $v \in H_0^2(D_\sigma)$ then

$$\langle \mathscr{R}_P, v\rangle = \int_{D_\sigma} R_\tau v + \int_\sigma J_{\sigma,1} v - \int_\sigma J_{\sigma,2} \tfrac{\partial v}{\partial \boldsymbol{n}_\sigma}. \tag{3.44}$$

Let ψ_σ be the bubble function (3.28) and extend the values of $J_{\sigma,1}$ in directions $\pm\boldsymbol{n}_\sigma$; i.e, into each of τ_i, and set $v_\sigma = \psi_\sigma J_{\sigma,1}$. Then (3.44) reads

$$c_m \|J_{\sigma,1}\|^2_{L^2(\sigma)} \preceq \int_\sigma \psi_\sigma J_{\sigma,1}^2 = \langle \mathscr{R}_P, v_\sigma\rangle - \int_{D_\sigma} R_\tau v_\sigma. \tag{3.45}$$

From Lemma 2.36 and (3.28) we have the estimates

$$\|\psi_\sigma E_\sigma J_{\sigma,1}\|_{H^2(D_\sigma)} \le c_{\mathrm{Inv}} h_\sigma^{-2} \|E_\sigma J_{\sigma,1}\|_{L^2(D_\sigma)} \tag{3.46}$$

and $\|\psi_\sigma E_\sigma J_{\sigma,1}\|_{L^2(D_\sigma)} \le c_m h_\sigma^{1/2} \|J_{\sigma,1}\|_{L^2(\sigma)}$ which we apply to (3.45) to obtain

$$\begin{aligned} c_m \|J_{\sigma,1}\|^2_{L^2(\sigma)} &\preceq \left(c_{\mathrm{Inv}} h_\sigma^{-2} \|\mathscr{R}_P\|_{H^{-2}(D_\sigma)} + \|R_\tau\|_{L^2(D_\sigma)}\right) \|J_{\sigma,1}\|_{L^2(D_\sigma)}, \\ &\le c_{\mathrm{dTr}} h_\sigma^{1/2} \left(c_{\mathrm{Inv}} h_\sigma^{-2} \|\mathscr{R}_P\|_{H^{-2}(D_\sigma)} + \|R_\tau\|_{L^2(D_\sigma)}\right) \|J_{\sigma,1}\|_{L^2(\sigma)}, \end{aligned} \tag{3.48}$$

where the the last line follows from (3.31). Now let χ_σ be the function (3.30), extend the values of $J_{\sigma,2}$ into D_σ and set $w_\sigma = \chi_\sigma J_{\sigma,2}$. We then have

$$\langle \mathscr{R}_P, w_\sigma\rangle = \int_{D_\sigma} R_\tau w_\sigma - h_\sigma^{-1} \int_\sigma \psi_\sigma J_{\sigma,2}^2. \tag{3.49}$$

Similarly, we obtain

$$c_m h_\sigma^{-1} \|J_{\sigma,2}\|^2_{L^2(\sigma)} \le c_{\mathrm{dTr}} h_\sigma^{1/2} \left(c_{\mathrm{Inv}} h_\sigma^{-2} \|\mathscr{R}_P\|_{H^{-2}(D_\sigma)} + \|R_\tau\|_{L^2(D_\sigma)}\right) \|J_{\sigma,2}\|_{L^2(\sigma)}. \tag{3.50}$$

We have from (3.48) and (3.50)

$$h_\sigma^3 \|J_{\sigma,1}\|^2_{L^2(\sigma)} + h_\sigma \|J_{\sigma,2}\|^2_{L^2(\sigma)} \le \tfrac{c_{\mathrm{Inv}} c_{\mathrm{dTr}}}{c_m} \|\mathscr{R}_P\|^2_{H^{-2}(D_\sigma)} + \tfrac{c_{\mathrm{dTr}}}{c_m} h_\sigma^4 \|R_\tau\|^2_{L^2(D_\sigma)}. \tag{3.52}$$

Summing up we have

$$\begin{aligned}\eta_P^2(V,\tau) &= h_\tau^4\|R_\tau\|^2_{L^2(\tau)} + \sum_{\sigma\in\partial\tau}\left(h_\sigma^3\|J_{\sigma,1}\|^2_{L^2(\sigma)} + h_\sigma\|J_{\sigma,2}\|^2_{L^2(\sigma)}\right),\\ &\le h_\tau^4\|R_\tau\|^2_{L^2(\tau)} + \tfrac{c_{\mathrm{dTr}}}{c_r}\sum_{\sigma\in\mathcal{E}_P}\left(c_{\mathrm{Inv}}\|\mathscr{R}_P\|^2_{H^{-2}(D_\sigma)} + h_\sigma^4\|R_\tau\|^2_{L^2(D_\sigma)}\right), \qquad (3.54)\\ &\le (1+c_{\mathrm{shape}})h_\tau^4\|R_\tau\|^2_{L^2(\omega_\tau)} + \tfrac{c_{\mathrm{Inv}}c_{\mathrm{dTr}}}{c_m}\sum_{\sigma\in\partial\tau}\|\mathscr{R}_P\|^2_{H^{-2}(D_\sigma)},\end{aligned}$$

we arrive at

$$\eta_P^2(V,\tau) \le C\left(\|\mathscr{R}_P\|^2_{H^{-2}(\tau)} + \sum_{\sigma\in\partial\tau}\|\mathscr{R}\|^2_{H^{-2}(D_\sigma)} + \mathrm{osc}_P^2(f,\omega_\tau)\right), \tag{3.56}$$

for a generic constant C that depends on c_{shape} and polynomial degree r. We now estimate $\|\mathscr{R}_P\|_{H^{-2}(\omega)}$, for $\omega\in\{\tau, D_\sigma\}$, in terms of the numerical error:

$$\|\mathscr{R}_P\|_{H^{-2}(\omega)} = \sup_{v\in H_0^2(\omega)}\frac{\langle\mathscr{R}_P, v\rangle}{|v|_{H_0^2(\omega)}} = \sup_{v\in H_0^2(\omega)}\frac{a(u-U,v)}{|v|_{H_0^2(\omega)}} \le C_{\mathrm{cont}}\|\Delta(u-U)\|^2_{L^2(\omega)}, \tag{3.57}$$

which makes $\|\mathscr{R}_P\|^2_{H^{-2}(\tau)} + \sum_{\sigma\in\partial\tau}\|\mathscr{R}\|^2_{H^{-2}(D_\sigma)} \preceq |u-U|^2_{H^2(\omega_\tau)}$ and we have a *continuous local lower bound*:

$$\eta_P^2(U,\tau) \le C\left(|u-U|^2_{H^2(\omega_\tau)} + \mathrm{osc}_P^2(f,\omega_\tau)\right). \tag{3.58}$$

The desired conclusion follows from shape-regularity and summing (3.58) over all cells $\tau\in P$.■

It is clear now from relation (3.43) that the right-hand side f is the culprit for weakening the equivalence (3.5). This is the price to pay for having a simple L^2 representation of the residual instead of having to solve local-problems as done in (Babuška and Rheinboldt [47]). However, the price is not too high for we would expect $h_\tau^2\|R_\tau - \overline{R}_\tau\|_{L^2(\tau)}$ to decay faster than the interior residual quantity $h_\tau^2\|R_\tau\|_{L^2(\tau)}$; the oscillation term may pollute the error estimation at first, but will eventually vanish faster than the estimator. As a result, we have estimator dominance over oscillation $\mathrm{osc}_P^2(U,\tau) \le \eta_P^2(U,\tau)$ in the asymptotic regime. The combination of Lemmata 3.2 and 3.6 provides an L^2 version of equivalence (3.5); i.e, the size of the residual when measured in L^2 is equivalent to the numerical error up-to the oscillation term:

$$C_L\eta^2(U,\Omega) - \mathrm{osc}_P^2(f,\Omega) \le |||u-U|||^2 \le C_U\eta_P^2(U,\Omega) \tag{3.59}$$

motivates measuring the decay rate of the *total-error*

$$\rho_P(v,V,g) = \left(|||v-V|||^2 + \mathrm{osc}_P^2(g,\Omega)\right)^{1/2}. \tag{3.60}$$

Discrete upper bound

The following result is not used for convergence but it is instrumental in quasi-optimality. In particular, the optimality of Dörlfer marking follows from it as will be seen in Lemma 6.13.

Lemma 3.7 (Discrete upper bound). *Let P be an admissible partition of Ω and let $P_* =$ $\mathbf{REFINE}\,[P, \mathscr{M}]$ for some marked set $\mathscr{M} \subseteq P$. Let U and U_* are the discrete solutions on P and P_*, respectively. Then for a constants $C_{dU,1}, C_{dU,2} > 0$, depending only on c_{shape},*

$$|\!|\!|U_* - U|\!|\!|^2 \leq C_{dU,1}\eta_P^2(U, \omega_{R_{P\to P_*}}), \tag{3.62}$$

where $\omega_{R_{P\to P_*}}$ is understood as the union of support extensions of refined cells from P to obtain P_*.

Proof. Let $e_* = U_* - U$. First note that if $V \in \mathbb{X}_P$ then in view of the nesting $\mathbb{X}_P \subset \mathbb{X}_{P_*}$,

$$a(U_* - U, e_*) = a(U_* - U, e_* - V). \tag{3.63}$$

To localize, we form disconnected subdomains $\Omega_i \subseteq \Omega$, $i \in J$, each formed from the interiors of connected components of $\Omega_* = \cup_{\tau \in R_{P\to P_*}} \overline{\tau}$. Then to each subdomain Ω_i we form a partition $P_i = \{\tau \in P : \tau \subset \Omega_i\}$, interior edges $\mathcal{E}_i = \{\sigma \in \mathcal{E}_P : \sigma \subset \partial\tau,\ \tau \in P_i\}$, and a corresponding finite-element space $\mathbb{X}_i$. Let $I_i : H^2(\Omega_i) \to \mathbb{X}_i$. Let $V \in \mathbb{X}_P$ be an approximation of e_* be given by

$$V = e_* \mathbb{1}_{\Omega\backslash\Omega_*} + \sum_{i\in J}(I_i e_*) \cdot \mathbb{1}_{\Omega_i}. \tag{3.64}$$

Then $e_* - V \equiv 0$ on $\Omega\backslash\Omega_*$ and performing integration by parts will yield

$$\begin{aligned} a(U_* - U, e_* - V) = \sum_{i\in J}\Bigg[&\sum_{\tau\in P_i}\langle R, e_* - I_P e_*\rangle_\tau \\ &+ \sum_{\sigma\in\mathcal{E}_i}\{\langle J_1, e_* - I_i e_*\rangle_\sigma + \langle J_2, e_* - I_i e_*\rangle_\sigma\}\Bigg], \end{aligned} \tag{3.66}$$

Following the same procedure carried in Lemma 3.2 we have

$$\begin{aligned} &\sum_{\tau\in P_i}\langle R, e_* - I_i e_*\rangle_\tau + \sum_{\sigma\in\mathcal{E}_i}\{\langle J_1, e_* - I_i e_*\rangle_\sigma + \langle J_2, e_* - I_i e_*\rangle_\sigma\} \\ &\qquad\qquad \leq c_\Pi\bigg(\sum_{\tau\in P_i}\eta_P^2(U,\tau)\bigg)^{1/2}\bigg(\sum_{\tau\in P_i}|e_*|^2_{H^2(\omega_\tau)}\bigg)^{1/2}, \end{aligned} \tag{3.68}$$

Set $\omega_{R_{P\to P_*}} = \cup\{\omega_\tau : \tau \in R_{P\to P_*}\}$, and therefore we have $|\!|\!|U_* - U|\!|\!|^2 \leq C_{dU}\eta_P(U, \omega_{R_{P\to P_*}})|e_*|_{H^2(\Omega)}$ as desired. ■

3.3 Convergence of AFEM

In this section we show that the derived computable estimator (2.142) when used to direct refinement will result in decreased error. This will hinge on the estimator Lipschitz property of Lemma 3.8. To show that procedure (2.132) exhibits convergence we must be able to relate the errors of consecutive discrete solutions. The symmetry of the bilinear form, consistency of the formulation and finite-element spline space nesting will readily provide that via Galerkin Pythagoras in Lemma 3.11.

Error reduction

The result of Lemma 3.9 says that the estimator will reduce in value with refinement by an amount proportional to the difference between two consecutive discrete solutions. This is an improvement from (Mekchay [56]) in that reduction takes place irrespective of oscillation. The result follows from well-behaved perturbation of the error indicator on different partitions.

Lemma 3.8 (Lipschitz property of estimator). *Let P be an admissible partition of Ω. There exists a constant $C_{\text{lip}} > 0$, depending only on c_{shape}, such that for any cell $\tau \in P$ we have*

$$|\eta_P(V,\tau) - \eta_P(W,\tau)| \leq C_{\text{lip}} |V - W|_{H^2(\omega_\tau)}, \tag{3.69}$$

holding for every pair of finite-element splines V and W in $\mathbb{X}_P$.

Proof. Let V and W be finite-element splines in $\mathbb{X}_P$ and let τ be a cell in partition P.

$$\begin{aligned} \eta_P(V,\tau) - \eta_P(W,\tau) = h_\tau^2 &\left(\|f - \mathcal{L}V\|_{L^2(\tau)} - \|f - \mathcal{L}W\|_{L^2(\tau)} \right) \\ &+ \sum_{\sigma \subset \partial\tau} h_\sigma^{1/2} \left(\| [\![\Delta V]\!]_\sigma \|_{L^2(\sigma)} - \| [\![\Delta W]\!]_\sigma \|_{L^2(\sigma)} \right) \\ &+ \sum_{\sigma \subset \partial\tau} h_\sigma^{3/2} \left(\| [\![\partial_\sigma \Delta V]\!]_\sigma \|_{L^2(\sigma)} - \| [\![\partial_\sigma \Delta W]\!]_\sigma \|_{L^2(\sigma)} \right). \end{aligned} \tag{3.71}$$

Treating the interior term,

$$\|f - \mathcal{L}V\|_{L^2(\tau)} - \|f - \mathcal{L}W\|_{L^2(\tau)} \leq |V - W|_{H^4(\tau)} \leq c_{\text{Inv}} h_\tau^{-2} |V - W|_{H^2(\tau)}. \tag{3.73}$$

Treating the edge terms we have

$$\| [\![\Delta V]\!]_\sigma \|_{L^2(\sigma)} - \| [\![\Delta W]\!]_\sigma \|_{L^2(\sigma)} \leq \| [\![\Delta V - \Delta W]\!]_\sigma \|_{L^2(\sigma)}. \tag{3.74}$$

Let τ' from P be a cell that shares the edge σ, i.e τ' is an adjacent cell to τ. For any finite-element spline $V \in \mathbb{X}_P$ we have

$$\| [\![V]\!]_\sigma \|_\sigma \leq c_{\text{dTr}} \left(h_\sigma^{-1/2} \|V\|_\tau + h_\sigma^{-1/2} \|V\|_{\tau'} \right) \leq c_{\text{dTr}} h_\sigma^{-1/2} \|V\|_{\omega_\tau}. \tag{3.75}$$

Replacing V with $\Delta V - \Delta W$ gives

$$h_\sigma^{1/2} \| \llbracket \Delta V - \Delta W \rrbracket_\sigma \|_\sigma \leq c_{\text{Inv}} c_{\text{dTr}} |V - W|_{H^2(\omega_\tau)}. \tag{3.76}$$

Similarly, we have

$$h_\sigma^{3/2} \left(\| \llbracket \partial_\sigma \Delta V \rrbracket_\sigma \|_{L^2(\sigma)} - \| \llbracket \partial_\sigma \Delta W \rrbracket_\sigma \|_{L^2(\sigma)} \right) \leq c_{\text{Inv}} c_{\text{dTr}} |V - W|_{H^2(\omega_\tau)}. \tag{3.77}$$

It then follows from (3.71)

$$\begin{aligned} |\eta_P(V,\tau) - \eta_P(W,\tau)| &\leq c_{\text{Inv}}(|V - W|_{H^2(\tau)} + 2c_{\text{dTr}}|V - W|_{H^2(\omega_\tau)}), \\ &\leq c_{\text{Inv}}(1 + 2c_{\text{dTr}})|V - W|_{H^2(\omega_\tau)}. \end{aligned} \tag{3.79}$$

■

Lemma 3.9 (Estimator error reduction). *Let P be an admissible partition of Ω, let $\mathscr{M} \subseteq P$ and let $P_* =$ **REFINE** $[P, \mathscr{M}]$. There exists constants $\lambda \in (0,1)$ and $C_{\text{est}} > 0$, depending only on c_{shape}, such that for any $\delta > 0$ it holds that for any pair of finite-element splines $V \in \mathbb{X}_P$ and $V_* \in \mathbb{X}_{P_*}$ we have*

$$\eta_{P_*}^2(V_*, \Omega) \leq (1 + \delta) \left\{ \eta_P^2(V, \Omega) - \tfrac{1}{2}\eta_P^2(V, \mathscr{M}) \right\} + c_{\text{shape}}(1 + \tfrac{1}{\delta}) \|\!| V - V_* |\!\|^2. \tag{3.80}$$

Proof. Let $\mathscr{M} \subseteq P$ be a set of marked elements from partition P and let $P_* =$ **REFINE** $[P, \mathscr{M}]$. For notational simplicity we denote $\mathbb{X}_{P_*}$ and η_{P_*} by $\mathbb{X}_*$ and η_*, respectively. Let V and V_* be the respective finite-element splines from $\mathbb{X}_P$ and $\mathbb{X}_*$. Let τ be a cell from partition P_*. In view of the Lipschitz property of the estimator (Lemma 3.8) and the nesting $\mathbb{X}_P \subseteq \mathbb{X}_*$,

$$\eta_*^2(V_*, \tau) \preceq \eta_*^2(V, \tau) + |V - V_*|_{H^2(\omega_\tau)}^2 + 2\eta_*(V_*, \tau)|V - V_*|_{H^2(\omega_\tau)}. \tag{3.81}$$

Given any $\delta > 0$, an application of Young's inequality on the last term gives

$$2\eta_*(V_*, \tau)|V - V_*|_{H^2(\omega_\tau)} \leq \delta\eta_*^2(V_*, \tau) + \tfrac{1}{\delta}|V - V_*|_{H^2(\omega_\tau)}^2. \tag{3.82}$$

We now have

$$\eta_*^2(V_*, \tau) \preceq (1 + \delta)\eta_*^2(V, \tau) + (1 + \tfrac{1}{\delta})|V - V_*|_{H^2(\omega_\tau)}^2. \tag{3.83}$$

Recalling that the partition cell are disjoint with uniformly bounded support extensions, we may sum over all the cells $\tau \in P_*$ to obtain

$$\eta_*^2(V_*, P_*) \leq (1 + \delta)\eta_*^2(V, P_*) + c_{\text{shape}}(1 + \tfrac{1}{\delta})\|\!| V - V_* |\!\|^2. \tag{3.84}$$

It remains to estimate $\eta_*^2(V, P_*)$. Let $|\mathscr{M}|$ be the sum areas of all cells in $\mathscr{M}$. For every marked element $\tau \in \mathscr{M}$ define $P_{*,\mathscr{M}} = \{\text{child}(\tau) : \tau \in \mathscr{M}\}$. Let $b > 0$ denote the number of bisections required to obtain the conforming partition P_* from P. Let τ_* be a child of a cell $\tau \in \mathscr{M}$. Then $h_{\tau_*} \leq 2^{-1} h_\tau$. Noting that $V \in \mathbb{X}_P$ we have no jumps within τ

$$\eta_*^2(V, \tau_*) = h_{\tau_*}^4 \|f - \mathcal{L}V\|_{\tau_*}^2 \leq (2^{-1} h_\tau)^4 \|f - \mathcal{L}V\|_{\tau_*}^2, \tag{3.85}$$

summing over all children

$$\sum_{\tau_* \in \text{children}(\tau)} \eta_*^2(V, \tau_*) \leq 2^{-1} \eta_P^2(V, \tau), \tag{3.86}$$

and we obtain by disjointness of partitions an estimate on the error reduction

$$\sum_{\tau_* \in P_{*,M}} \eta_*^2(V, \tau_*) \leq 2^{-1} \eta_P^2(V, \mathscr{M}). \tag{3.87}$$

For the remaining cells $T \in P \backslash \mathscr{M}$, the estimator monotonicity implies $\eta_{P_*}(V, T) \leq \eta_P(V, T)$. Decompose the partition P as union of marked cells in $\mathscr{M}$ and their complement $P \backslash \mathscr{M}$ to conclude the total error reduction obtained by **REFINE** and the choice of Dörfler parameter θ

$$\eta_{P_*}^2(V, \Omega) \leq \eta_P^2(V, \Omega \backslash \mathscr{M}) + 2^{-1} \eta_P^2(V, \mathscr{M}) = \eta_P^2(V, \Omega) - \tfrac{1}{2} \eta_P^2(V, \mathscr{M}). \tag{3.88}$$

■

Corollary 3.10. *There exists constants* $q_{\text{est}} \in (0,1)$ *and* $C_{\text{est}} > 0$ *such that*

$$\eta_{P_*}^2(U_*, \Omega) \leq q_{\text{est}} \eta_P^2(U, \Omega) + C_{\text{est}}^{-1} \|U_* - U\|^2. \tag{3.89}$$

Proof. Define constants

$$q_{\text{est}}(\theta, \delta) = (1 + \delta) \left(1 - \tfrac{\theta^2}{2}\right) \quad \text{and} \quad C_{\text{est}}^{-1}(\delta) = c_{\text{shape}} \left(1 + \tfrac{1}{\delta}\right),$$

so that in view of Dörlfer $-\eta_P^2(U, \mathscr{M}) \leq -\theta^2 \eta_P^2(U, \Omega)$ we have

$$(1 + \delta) \left\{\eta_P^2(U, \Omega) - \tfrac{1}{2} \eta_P^2(U, \mathscr{M})\right\} \leq q_{\text{est}} \eta_P(U, \Omega)^2.$$

The choice $\delta < \frac{\theta^2}{2 - \theta^2}$ ensures that $0 < q_{\text{est}} < 1$ as desired. ■

The following result allows one to relate the numerical error between two iterations.

Lemma 3.11 (Galerkin Pythagoras). *Let* P *and* P_* *be an admissible partitions of* Ω *with* $P_* \geq P$ *and let* $U \in \mathbb{X}_P$ *and* $U_* \in \mathbb{X}_{P_*}$ *be discrete solutions. Then*

$$\|u - U_*\|^2 = \|u - U_*\|^2 - \|U_* - U\|^2. \tag{3.90}$$

Proof. At first we express

$$\begin{aligned} a(u-U_*,u-U_*) &= a(u-U,u-U) - a(U_*-U,U_*-U) \\ &\quad + a(U-U_*,u-U_*) + a(u-U_*,U-U_*). \end{aligned} \tag{3.92}$$

Recognizing that $a(u-U_*,U-U_*) = a(U-U_*,u-U) = 0$, we arrive at

$$a(u-U_*,u-U_*) = a(u-U,u-U) - a(U_*-U,U_*-U). \tag{3.93}$$

■

Remark 3.12. The Pythagoras relation of Lemma 3.11 is instrumental in achieving convergence and the expression is really the best case scenario. In particular, the symmetry gives us the equality. When symmetry is broken due to a linear advection term, a weaker version can derived (Mekchay et al. [56] and Feischl et al. [62]). The so-called quasi-Pythagoras relation reads:

$$|\!|\!| u-U_* |\!|\!|^2 \le \Lambda |\!|\!| u-U |\!|\!|^2 - |\!|\!| U_*-U |\!|\!|^2, \tag{3.94}$$

for a constant $\Lambda > 1$ that can be made arbitrarily close to 1 with refinement. In order to make the convergence proof valid for non-symmetric bilinear forms, we consider (3.94) instead of (3.90). For the convergence of (2.132) in this general setting, we will need P_0 to be fine enough such that (3.94) holds with $\Lambda = 1+\varepsilon$ with $0 < \varepsilon < (1-q_{\text{est}})\frac{C_{\text{est}}}{C_U}$. In our symmetric setting, no initial-mesh assumption is needed; we automatically have $\Lambda = 1$.

Contraction of quasi-error

We define the *quasi-error* by

$$|\!|\!| u-U |\!|\!|^2 + C_{\text{est}}\eta_P^2(U,\Omega). \tag{3.95}$$

The constant C_{est} will be chosen below.

Theorem 3.13 (Convergence of conforming AFEM). *For a contraction factor $\alpha \in (0, 1/\Lambda)$, there exists a suitable choice for $C_{\text{est}} > 0$ such that given any consecutive admissible mesh partitions P and P_*, $f \in L^2(\Omega)$ and Dörlfer parameter $\theta \in (0,1]$, the adaptive procedure* **AFEM** $[P,f,\theta]$ *will produce two successive discrete solutions $U \in \mathbb{X}_P$ and $U_* \in \mathbb{X}_{P_*}$ for which*

$$|\!|\!| u-U_* |\!|\!|^2 + C_{\text{est}}\eta_{P_*}^2(U_*,\Omega) \le \alpha\Lambda\big(|\!|\!| u-U |\!|\!|^2 + C_{\text{est}}\eta_P^2(U,\Omega)\big). \tag{3.96}$$

Proof. Adopt the following abbreviations:

$$\begin{gathered} e_P = u-U, \qquad e_{P_*} = u-U_*, \qquad \varepsilon_P = U_*-U, \\ \eta_P = \eta_P(U,\Omega), \quad \eta_{P_*} = \eta_{P_*}(U_*,\Omega). \end{gathered}$$

The objective is to show that the quasi-error (3.95) is contracts with consecutive refinement; that is, we will show that there exists a factor $0 < \alpha\Lambda < 1$ such that (3.96) holds true. Applying Galerkin quasi-orthogonality (3.94) and estimator error reduction in the form expressed in Corollary 3.10 to the quasi-error

$$\begin{aligned} \|e_{P_*}\|^2 + C_{\text{est}}\eta_{P_*}^2 &\leq (\Lambda\|e_P\|^2 - \|\varepsilon_P\|^2) + C_{\text{est}}(q_{\text{est}}\eta_P^2 + C_{\text{est}}^{-1}\|\varepsilon_P\|^2), \\ &= \Lambda\|e_P\|^2 + q_{\text{est}}C_{\text{est}}\eta_P^2. \end{aligned} \tag{3.98}$$

thus removing the perturbation term $\|\varepsilon_P\|^2$ by our parameter choice C_{est} in the quasi-norm definition.

Let $\alpha \in (0,1)$, to be chose later, and write $\|e_P\|^2 = \alpha\|e_P\|^2 + (1-\alpha)\|e_P\|^2$. Now by Global Upper Bound of Lemma 3.2:

$$\|e_{P_*}\|^2 + C_{\text{est}}\eta_{P_*}^2 \leq \alpha\Lambda\|e_P\|^2 + C_{\text{est}}\left((1-\alpha)\Lambda\frac{C_U}{C_{\text{est}}} + q_{\text{est}}\right)\eta_P^2. \tag{3.99}$$

The expression $(1-\alpha)\Lambda\frac{C_U}{C_{\text{est}}} + q_{\text{est}} = \alpha\Lambda$ holds when α is chosen to be

$$\alpha = \frac{\Lambda C_U + q_{\text{est}}C_{\text{est}}}{\Lambda(C_U + C_{\text{est}})}. \tag{3.100}$$

We show that the product $0 < \alpha\Lambda < 1$ holds. In view Remark 3.12,

$$\alpha\Lambda = \frac{(1+\varepsilon)C_U + q_{\text{est}}C_{\text{est}}}{C_U + C_{\text{est}}} < \frac{C_U + (1-q_{\text{est}})C_{\text{est}} + q_{\text{est}}C_{\text{est}}}{C_U + C_{\text{est}}} = 1. \tag{3.101}$$

■

Remark 3.14. The rest of the analysis resumes with $\Lambda = 1$; symmetry will be invoked.

3.4 Approximation classes

The aim is to show that the proposed AFEM (2.132) will generate a sequence of partitions $\{P_\ell\}_{\ell\geq 1}$ for which $\rho_{P_\ell}^2(u, U_\ell, f)$ decays with order $(\#P_\ell)^{-s}$. The right-hand side function f is directly given by $\mathcal{L}u$ which justifies looking at an AFEM approximation class described by the total-error norm (3.60). For $s > 0$ define

$$\mathbb{A}^s = \left\{v \in H_0^2(\Omega) : |v|_{\mathbb{A}^s} = \sup_{N>0} N^s \inf_{P\in\mathscr{P}_N} \inf_{V\in\mathbb{X}_P} \rho_P(v, V, \mathcal{L}v) < \infty\right\} \tag{3.102}$$

The AFEM approximation class $\mathbb{A}^s$ is not standard, however, we will express it in terms of following standard approximation classes: for $s > 0$ define

$$\mathscr{A}^s = \mathscr{A}_\infty^s(H_0^2(\Omega), \{\mathbb{X}_P\}_{P\in\mathscr{P}}), \tag{3.103}$$

and

$$\mathscr{O}^s = \mathscr{A}^s_\infty(H^{-2}(\Omega), \{\mathcal{P}^{r-4}_P\}_{P\in\mathscr{P}}). \tag{3.104}$$

Remark 3.15. We take the definition of $\mathscr{O}^s$ to be a subspace of $L^2(\Omega)$ only and measure the error in $H^{-2}(\Omega)$. Therefore, the approximation error norm $E(f, \mathcal{P}^{r-4}_P)_{H^{-2}(\Omega)}$ characterizing the approximation class $\mathscr{O}^s$ is equivalent to $\inf_{S\in\mathcal{P}^{r-4}_P} \|h^2_P(f-S)\|_{L^2(\Omega)}$ and since L^2-orthogonal projections yields optimal error in $L^2(\Omega)$, we may just take $E(f, \mathcal{P}^{r-4}_P)_{H^{-2}(\Omega)} = \operatorname{osc}_P(f, \Omega)$.

Remark 3.16. We will have to restrict values of s so as not to yield trivial spaces; spaces consisting of spline polynomials only. For $\mathscr{A}^s$ we have already done that; see Theorem 2.46. As for $\mathscr{O}^s$, see Theorem 3.22 below.

It is immediate that any weak solution $u \in \mathbb{A}^s$ implies that $(u, \mathcal{L}u) \in \mathscr{A}^s \times \mathscr{O}^s$. The other direction merits proving:

Lemma 3.17 (Equivalence of classes). *Let u be the weak solution. If $u \in \mathscr{A}^s$ and $\mathcal{L}u \in \mathscr{O}^s$, then $u \in \mathbb{A}^s$.*

Proof. By definition we have two admissible partitions $P_1, P_2 \in \mathscr{P}_N$ and a finite-element spline $V \in \mathbb{X}_{P_1}$ such that $|\!|\!| u - V |\!|\!| \preceq N^{-s}$ and $\operatorname{osc}_{P_2}(f) \preceq N^{-s}$. Invoking Mesh Overlay (2.75) we obtain an admissible partition $P = P_1 \oplus P_2$ for which $\#P \preceq 2N$ and because of spline space nesting we have

$$|\!|\!| u - V |\!|\!|^2 + \operatorname{osc}^2_P(f) \preceq N^{-2s}. \tag{3.105}$$

■

3.5 Quasi-optimality

The contraction achieved in the convergence proof is ensured by the Dörfler marking strategy. However the relationship between the Dörfler strategy and error reduction in the total-error norm goes deeper than asserted in Theorem 3.13. In the following lemma we show that if $R_{P\to P_*}$ is a set of refined elements resulting in a reduction of error in contractive sense, then necessarily the Dörfler property holds for the set $\omega_{R_{P\to P_*}}$. The fact will be instrumental in proving that the cardinality of marked cells will keep the partition cardinality at each refinement step proprtional to the optimal quantity dictated by nononlinear approximation.

In what follows we show that Cea's lemma holds for the total-error norm, that is, that the finite element solution U is an optimal choice from $\mathbb{X}_P$ in total-error norm.

Lemma 3.18 (Optimality of total error). *Let u be the weak solution and let $U \in \mathbb{X}_P$ be the discrete solution. Then,*

$$\rho_P^2(u, U, f) \le \inf_{V \in \mathbb{X}_P} \rho_P^2(u, V, f). \tag{3.106}$$

Proof. In view of Galerkin orthogonality and the symmetry of the bilinear form $a(u-U, U-V) = a(U-V, u-U) = 0$ we have

$$a(u-V, u-V_\varepsilon) = a(u-U, u-U) + a(U-V, U-V), \tag{3.108}$$

and we have $|||u-V|||^2 = |||u-U|||^2 + |||U-V|||^2$. Therefore,

$$\begin{aligned} \rho_P^2(u, U, f) &\le |||u-U|||^2 + \operatorname{osc}_P^2(f, \Omega) + |||U-V|||^2, \\ &= |||u-V|||^2 + \operatorname{osc}_P^2(f, \Omega) = \rho_P^2(u, V, f). \end{aligned} \tag{3.110}$$

■

Lemma 3.19 (Optimal Marking). *Let $U = \mathbf{SOLVE}\,[P, f]$, let P_* be any refinement of P and let $U_* = \mathbf{SOLVE}\,[P_*, f]$. If for some positive $\mu < 1$*

$$|||u-U_*|||^2 + \operatorname{osc}_*^2(f, \Omega) \le \mu\big(|||u-U|||^2 + \operatorname{osc}_P^2(f, \Omega)\big), \tag{3.111}$$

and $R_{P\to P_*}$ denotes collection of all elements in P requiring refinement to obtain P_* from P, then for $\theta \in (0, \theta_*)$ we have

$$\eta_P(U, \omega_{R_{P\to P_*}}) \ge \theta \eta_P(U, \Omega). \tag{3.112}$$

Proof. Let $\theta < \theta_*$, the parameter θ_* to be specified later, such that the linear contraction of the total error holds for $\mu = 1 - \frac{\theta^2}{\theta_*^2} > 0$. The Efficiency Estimate (3.39) together with the assumption (6.94)

$$\begin{aligned} (1-\mu) C_L \eta_P^2(U, \Omega) &\le (1-\mu)\rho_P^2(u, U, f), \\ &= \rho_P^2(u, U, f) - \rho_*^2(u_*, U_*, f), \\ &= |||u-U|||^2 - |||u-U_*|||^2 + \operatorname{osc}_P^2(f, \Omega) - \operatorname{osc}_{P_*}^2(f, \Omega). \end{aligned} \tag{3.114}$$

In view of Galerkin pythagorus gives $|||u-U|||^2 - |||u-U_*|||^2 = |||U-U_*|||^2$. $R_{P\to P_*} \subset P$ so $\operatorname{osc}_P^2(f, \Omega) - \operatorname{osc}_{P_*}^2(f, \Omega) \le \operatorname{osc}_P^2(f, \omega_{R_{P\to P_*}})$. Estimator asymptotic dominance over oscillation $\operatorname{osc}_P^2(U, \tau) \le \eta_P^2(U, \tau)$ and Discrete Upper Bound (3.62)

$$(1-\mu) C_L \eta_P^2(U, P) \le (1 + C_{dU}) \eta_P^2(U, \omega_{R_{P\to P_*}}). \tag{3.115}$$

By definition $\theta^2 = (1-\mu)\theta_*^2 < \theta_*^2$ we arrive at $\theta^2 \eta_P^2(U, \Omega) \le \eta_P^2(U, \omega_{R_{P\to P_*}})$ for $\theta^2 < \frac{C_L}{1+C_{dU}} =: \theta_*^2$.

■

Lemma 3.20 (Cardinality of Marked Cells). *Let $\{(P_\ell, \mathbb{X}_\ell, U_\ell)\}_{\ell\geq 0}$ be sequence generated by* **AFEM** $(P_0, f; \varepsilon, \theta)$ *for admissible P_0 and the pair $u \in \mathbb{A}^s$ for some $s > 0$ then*

$$\#\mathscr{M}_\ell \preceq \left(1 - \frac{\theta^2}{\theta_*^2}\right)^{-\frac{1}{2s}} |u|_{\mathbb{A}_s}^{-\frac{1}{s}} \left\{ |\!|\!| u - U_\ell |\!|\!|^2 + \operatorname{osc}_\ell^2(f, P_\ell) \right\}^{-\frac{1}{2s}}. \tag{3.116}$$

Proof. Assume that the marking parameter satisfies the hypothesis of Theorem 6.13 and suppose that $u \in \mathbb{A}_s$ for some $s > 0$. Set $\mu = 1 - \frac{\theta^2}{\theta_*^2}$ and let $\varepsilon = \mu \rho_\ell(u, U_\ell, f)^2$. Then by definition of $\mathbb{A}_s$ there exists an admissible partition P_ε and a spline $V_\varepsilon \in \mathbb{X}_\varepsilon$ for which

$$\rho_\varepsilon^2(u, V_\varepsilon, f) \leq \varepsilon^2 \quad \text{with} \quad \#P_\varepsilon - \#P_0 \preceq |u|_{\mathbb{A}_s}^{1/s} \varepsilon^{-1/s}. \tag{3.117}$$

Let $P_* = P_\varepsilon \oplus P_\ell$ be the overlay partition of P_ε and P_ℓ, $\ell \geq 0$, and let $U_* \in \mathbb{X}_*$ be the corresponding spline solution. In view of Optimality of Total Error in Lemma 3.18 and the fact $P_* \geq P_\varepsilon$ makes $\mathbb{X}_* \supseteq \mathbb{X}_\varepsilon$ and

$$\rho_*^2(u, U_*, f) \leq \rho_\varepsilon^2(u, V_\varepsilon, f) \leq \varepsilon^2 = \mu \rho_\ell^2(u, U_\ell, f). \tag{3.118}$$

From Optimal Marking of Lemma 6.13 we have $R_{P_\ell \to P_*} \subset P_\ell$ satisfying Dörfler property for $\theta < \theta_*$.

$$\#\mathscr{M}_\ell \leq \#R_{P_\ell \to P_*} \leq \#P_* - \#P_\ell. \tag{3.119}$$

In view of overlay property $\#P_* \leq P_\varepsilon + \#P_\ell - \#P_0$ in (2.75) and definition of ε we arrive at

$$\#\mathscr{M}_\ell \leq \#P_\varepsilon - \#P_0 \preceq \mu^{-1/2s} |u|_{\mathbb{A}_s}^{1/s} \rho_\ell(u, U_\ell, f)^{-1/s}. \tag{3.120}$$

■

Theorem 3.21 (Quasi-optimality of conforming AFEM). *If $u \in \mathbb{A}^s$ and P_0 is admissible, then the call* **AFEM** $[P_0, f, \varepsilon, \theta]$ *generates a sequence $\{(P_\ell, \mathbb{X}_\ell, U_\ell)\}_{\ell\geq 0}$ of strictly admissible partitions P_ℓ, conforming finite-element spline spaces $\mathbb{X}_\ell$ and discrete solutions U_ℓ satisfying*

$$\rho_\ell(u, U_\ell, f) \preceq \Phi(, \theta) |(u, f)|_{\mathbb{A}_s} (\#P - \#P_0)^{-s}, \tag{3.121}$$

with $\Phi(, \theta) = (1 - \theta^2/\theta_*^2)^{-1/2}$

Proof. Let $\theta < \theta_*$ be given and assume that $u \in \mathbb{A}^s(\rho)$. We will show that the adaptive procedure **AFEM** will produce a sequence $\{(P_\ell, \mathbb{X}_\ell, U_\ell)\}_{\ell\geq 0}$ such that $\rho_\ell \preceq (\#P_\ell - \#P_0)^{-s}$. Let $A(\theta, s) = (1 - \theta^2/\theta_*^2)^{-1/2s} |u|_{\mathbb{A}^s}^{-1/s}$ Cardinality of Marked Cells (6.109) and (2.107) yields

$$\#P_\ell - \#P_0 \preceq A(\theta, s) \sum_{j=0}^{\ell-1} \rho_j^{-1/s}. \tag{3.122}$$

In view of Convergence Theorem 3.13, we have for a factor $C_{\text{est}} > 0$ and a contractive factor $\alpha \in (0, 1)$

$$e_\ell^2 + C_{\text{est}}\eta_\ell^2 \le \alpha^{2(\ell-j)}\left(e_j^2 + C_{\text{est}}\eta_j^2\right), \quad j = 1, .., \ell-1, \tag{3.123}$$

holding for any iteration $\ell \ge 0$. At each intermediate step, the Efficiency Estimate (3.39) makes $e_j^2 + \gamma\eta_j^2 \le (1 + C_{\text{est}}/C_{\text{eff}})\,\rho_j^2$ so we may write

$$\rho_j^{-\frac{1}{s}} \le \alpha^{\frac{\ell-j}{s}}\left(1 + \tfrac{C_{\text{est}}}{C_{\text{eff}}}\right)^{\frac{1}{2s}}\left(e_\ell^2 + C_{\text{est}}\eta_\ell^2\right)^{-\frac{1}{2s}}. \tag{3.124}$$

We sum (3.124) over $j = 0 : \ell - 1$ and we recover the total-error from the quasi-error using estimator domination over oscillation,

$$\sum_{j=0}^{\ell-1}\rho_j^{-\frac{1}{s}} \le \sum_{j=0}^{\ell-1}\alpha^{\frac{\ell-j}{s}}\left(1 + \tfrac{C_{\text{est}}}{C_L}\right)^{\frac{1}{2s}}\left(e_\ell^2 + C_{\text{est}}\text{osc}_\ell^2\right)^{-\frac{1}{2s}}. \tag{3.125}$$

We obtain

$$\#P_\ell - \#P_0 \preceq M(\theta, s)\left(e_\ell^2 + C_{\text{est}}\text{osc}_\ell^2\right)^{-\frac{1}{2s}}\sum_{j=1}^{\ell}\alpha^{\frac{j}{s}}, \tag{3.126}$$

where $M(\theta, s) = A(\theta, s)\left(1 + \frac{C_{\text{est}}}{C_L}\right)^{\frac{1}{2s}}$ and $\sum_{j=1}^{\ell}\alpha^{\frac{j}{s}} \le \alpha^{1/s}(1 - \alpha^{1/s})^{-1} =: S(\theta, s)$ for any $\ell \ge 1$.

$$\#P_\ell - \#P_0 \preceq S(\theta, s)M(\theta, s)\rho_\ell(u, U_\ell, f)^{-\frac{1}{s}}. \tag{3.127}$$

■

Finally, we have the following characterizations:

Theorem 3.22. *We have* $\mathscr{B}^{s-2}_{p,p}(\Omega) \hookrightarrow \mathscr{O}^{s/2}$ *whenever* $s \ge \frac{s}{p} + 1$ *and* $p \in (0, \infty)$.

Proof. The proof mirrors the argument used to prove Theorem 2.46 so we just highlight the main steps. Let $\pi \in \mathbb{P}_{r-4}(\omega_\tau)$, and choose $p \in (0, \infty)$ to be such that the embedding $\mathscr{B}^{s-2}_{p,p}(\Omega) \hookrightarrow L^2(\Omega)$ is continuous. We will have by (2.29),

$$\inf_{\pi \in \mathbb{P}_{r-4}(\tau)}\|f - \pi\|_{L^2(\tau)} \preceq h_\tau^{s-1-\frac{2}{p}}|f|_{\mathscr{B}^{s-2}_{p,p}(\tau)}. \tag{3.128}$$

Moreover,

$$\text{osc}_P^2(f) = \sum_{\tau \in P} h_\tau^4\|f - \Pi_P^{r-4}f\|_{L^2(\tau)}^2 \preceq \sum_{\tau \in P} h_\tau^{2(s+1-\frac{2}{p})}|f|^2_{\mathscr{B}^{s-2}_{p,p}(\tau)}. \tag{3.130}$$

Now let $e(\tau, P) = |\tau|^{\delta} |f|_{\mathscr{B}^s_{p,p}(\tau)}$ with $\delta = \frac{s+1}{2} - \frac{1}{p}$ and apply Lemma 2.45 with $\omega = \tau$ to obtain an admissible mesh $P \in \mathscr{P}$ such that

$$\operatorname{osc}^2_P(f) \preceq \#P\varepsilon^2 \quad \text{with} \quad \#P - \#P_0 \preceq |v|^{p/(1+\delta p)}_{\mathscr{B}^{s-2}_{p,p}(\Omega)} \varepsilon^{-p/(1+\delta p)}. \tag{3.131}$$

Using the definition of δ, we determine that $p/(1+\delta p) = 2/(s+1)$. Let $N = \#P$ and let $\varepsilon = N^{-(s+1)/2} |v|_{\mathscr{B}^{s-2}_{p,p}(\Omega)}$ from which we conclude

$$\operatorname{osc}_P(f) \preceq |f|_{\mathscr{B}^{s-2}_{p,p}(\Omega)} N^{-s/2} \quad \text{and} \quad \#P - \#P_0 \preceq N. \tag{3.132}$$

■

Figure 3.2 illustrates the results of Theorem 3.22 in terms of a DeVore diagram.

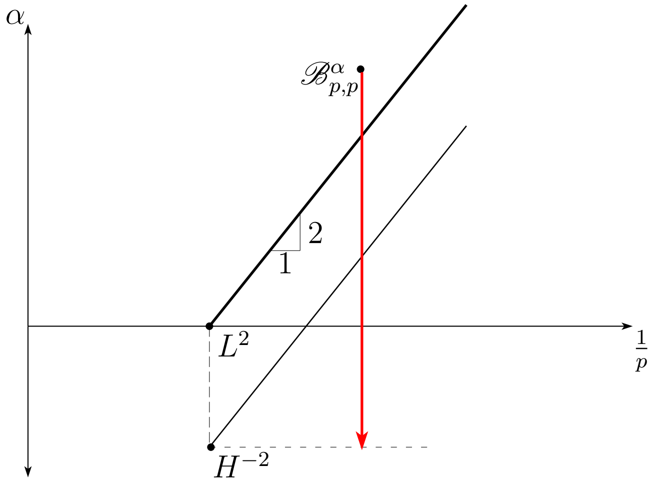


Figure 3.2: *If $\mathscr{B}^{\alpha}_{p,p}$ is positioned on or above the Sobolev embedding line emanating from L^2, depicted by the thick line, the embedding $\mathscr{B}^{\alpha}_{p,p} \hookrightarrow \mathscr{A}^s$ holds with $s = \frac{\alpha}{2} + 1$ and the length of the red arrow is equal to s. It is worth noting that even though we look at $\mathcal{L}u$ as a member of L^2, we are in fact measuring the approximation of $\mathcal{L}u$ in H^{-2}. In particular, the height from the demarcated horizontal line until the $s = 0$ axis manifests in the h^2_P factor in* osc*; see also Remark 3.15.*

The previous two results in combination with Lemma 3.17 and Theorem 2.46 yields a one-sided characterization of the AFEM approximation class $\mathbb{A}^s$ in terms of smoothness spaces:

Corollary 3.23 (One-sided characterization for $\mathbb{A}^s$). *Let u be the weak solution. If $u \in \mathscr{B}^{s+2}_{p,p;r+1}(\Omega) \cap H^2_0(\Omega)$ with $\frac{2}{p} - 1 \leq s < r - 2 + \max\{1, \frac{1}{p}\}$, for some $p \in (0, \infty)$ and $\mathcal{L}u \in \mathscr{B}^{s-2}_{q,q}(\Omega) \cap L^2(\Omega)$ with $s \geq \frac{2}{q} + 1$ for some $q \in (0, \infty)$, then we have $u \in \mathbb{A}^{s/2}$.*

Chapter 4

Adaptivity of a B-spline based finite-element method for modeling wind-driven ocean circulation

4.1 Introduction

Recently, (Kim et al. [36]) introduced Nitsche-type variational formulations for the stream-function formulation of the stationary QGE, the Stommel model, and the Stommel–Munk model. These Nitsche-type formulations can be readily applied for non-interpolatory basis functions such as B-splines and embedded geometries (Jiang and Kim [39]), where the domain can be implicitly defined via a level-set function. A distinguishing feature of these formulations is the employment of Nitsche's method (Nitsche [33]) to weakly impose the Drichlet boundary conditions and stabilization. Nitsche's method has been successfully applied to impose boundary conditions for the second- and fourth-order partial differential equations (Embar et al. [69], Kim et al. [70] and Jiang et al. [71]). Moreover, Nitsche-type non-conforming formulations for fourth-order partial differential equations using C^0-elements have been developed for a second-gradient theory (Kim et al. [40–42] and Kim and Dolbow [43]) and the stationary QGE (Kim et al. [44]).

Following the previous work of Kim et al. [36], Kim with his colleagues performed a priori error estimate for the Nitsche-type formulation of the Stommel–Munk model (see Rotunda et al. [45]). In this paper, we perform a posteriori error analysis for mesh-refinement in Section 3.3 and verify the analysis via numerical tests. This analysis gives rise to the *a posterior error* indicators (4.20) for the efficient mesh-refinement in an automatic manner. The capability of the a posterior error indicator is then verified via several benchmark problems using cubic

B-splines. In particular, our approach is a hierarchical B-spline refinement technique (Jiang and Dolbow [46], Vuong et al. [73], Schillinger et al. [74] and Bornemann et al. [75]) relying on the a posteriori error indicator to create meshes well adapted to the solution.

The remainder of this Chapter is organized as follows. In Section 4.2, we present the Stommel–Munk model and recall the Nitsche-type variational formulation and its discretization for the Stommel–Munk model from Kim et al. [36]. Following this, in Section 4.3, the *a posteriori* error estimation is performed to derive the error indicators. Finally, in Section 4.4, numerical studies with two benchmark problems in rectangular and L-shape domains are provided to show the performance of the analysis in Section 4.3.

4.2 The Stommel–Munk model

We consider an open set $\Omega \subset \mathbb{R}^2$ with polygonal boundary Γ. The Stommel–Munk model (Vallis [66]) is given

$$\begin{aligned} -\varepsilon_{\mathrm{s}}\Delta u + \varepsilon_{\mathrm{m}}\Delta^2 u - \frac{\partial u}{\partial x} &= f \quad \text{in} \quad \Omega, \\ u = 0, \quad \nabla u \cdot \boldsymbol{n} &= 0 \quad \text{on} \quad \Gamma. \end{aligned} \tag{4.2}$$

For the wind-driven ocean circulation in an enclosed midlatitude basin, u and f can be the velocity streamfunction and the wind forcing, respectively. The parameters ε_{s} and ε_{m} are the nondimensional Stommel and Munk numbers, respectively, which were already defined in (1.2).

We reiterate the Nitsche-type weak formulation for the Stommel–Munk model introduced by Kim et al. [36] given by (1.9). For a spline finite element space $\mathbb{X}_P \subset H^2(\Omega)$ such that

$$\mathbb{X}_P \subset \mathcal{P}_P^r(\Omega) \cap C^2(\Omega), \quad (r \geq 3), \tag{4.3}$$

the discrete problem reads:

$$\text{Find } U \in \mathbb{X}_P \text{ such that} \quad a_P(U,V) + b_P(U,V) = \ell_f(V) \quad \text{for all } V \in \mathbb{X}_P, \tag{4.4}$$

where a_P is as in (1.8) with $\alpha = \varepsilon_{\mathrm{m}}$,

$$\begin{aligned} a_P(U,V) = \varepsilon_{\mathrm{m}} \int_\Omega \Delta U \Delta V + \varepsilon_{\mathrm{m}} \int_\Gamma \left(\tfrac{\partial \Delta U}{\partial \boldsymbol{n}} V + U \tfrac{\partial \Delta V}{\partial \boldsymbol{n}}\right) \\ - \varepsilon_{\mathrm{m}} \int_\Gamma \left(\Delta U \tfrac{\partial V}{\partial \boldsymbol{n}} + \tfrac{\partial U}{\partial \boldsymbol{n}} \Delta U\right) + \gamma_1 \int_\Gamma h_P^{-3} UV + \gamma_2 \int_\Gamma h_P^{-1} \tfrac{\partial U}{\partial \boldsymbol{n}} \tfrac{\partial V}{\partial \boldsymbol{n}}, \end{aligned} \tag{4.6}$$

and lower-order form b is non-symmetric, bilinear and is given by

$$b(U,V) = \varepsilon_{\mathrm{s}} \int_{\Omega} \nabla U \cdot \nabla V - \int_{\Omega} \tfrac{\partial U}{\partial x} V - \varepsilon_{\mathrm{s}} \int_{\Gamma} \left(\tfrac{\partial U}{\partial \boldsymbol{n}} V + U \tfrac{\partial V}{\partial \boldsymbol{n}} \right). \tag{4.7}$$

We emphasis that the addition of the boundary terms $\int_{\Gamma} \left(\frac{\partial U}{\partial \boldsymbol{n}} V + U \frac{\partial V}{\partial \boldsymbol{n}} \right)$ arising from the Laplacian can be neglected. This is because we have observed that the values of U and $\frac{\partial U}{\partial \boldsymbol{n}}$ along the boundary decay very rapidly and therefore their contribution to the numerical linear system is insignificant. We leave them here so that our analysis remains consistent with the rest of the PhD. For brevity, we take $\mathscr{B}_P : \mathbb{X}_P \times \mathbb{X}_P \to \mathbb{R}$ is given by

$$\mathscr{B}_P(U,V) = a_P(U,V) + b(U;V). \tag{4.8}$$

In the error analysis, we use the following mesh-dependent norms:

$$|||u|||_P^2 = \|\nabla u\|_{L^2(\Omega)}^2 + \|\Delta u\|_{L^2(\Omega)}^2 + \gamma_1 \|u\|_{3/2,P}^2 + \gamma_2 \left\| \tfrac{\partial u}{\partial \boldsymbol{n}} \right\|_{1/2,P}^2, \tag{4.9}$$

where

$$\|u\|_{s,P}^2 = \sum_{\sigma \in \mathcal{G}_P} h_\sigma^{-2s} \|u\|_{L^2(\sigma)}^2, \tag{4.10}$$

the $\mathcal{G}_P$ denotes all mesh edges along Γ and the parameters γ_1 and γ_2 are chosen such that the bilinear form $\mathscr{B}_P$ is coercive; see Remark 4.3 below. Moreover, the norm is a full norm equivalent to the standard $H^2(\Omega)$. We justify this fact in Remark 5.2 when treating the nonlinear model in the coming chapter.

The following corollary is a consequence of Lemma 2.42 and it is instrumental to the analyses of this Chapter.

Corollary 4.1. *Let P and P_* be admissible partitions with $P_* \geq P$, let $V \in \mathbb{X}_P$ and let $V_* \in \mathbb{X}_{P_*}$. Then,*

$$\begin{aligned} \sum_{\tau \in P_*} h_\tau^{-4} \|V_* - I_P V_*\|_{L^2(\tau)}^2 &+ \sum_{\sigma \in \mathcal{I}_{P_*} \cup \mathcal{G}_{P_*}} h_\sigma^{-3} \|V_* - I_P V_*\|_{L^2(\sigma)}^2 \\ &+ \sum_{\sigma \in \mathcal{I}_{P_*} \cup \mathcal{G}_{P_*}} h_\sigma^{-1} \left\| \tfrac{\partial}{\partial \boldsymbol{n}_\sigma} (V_* - I_P V_*) \right\|_{L^2(\sigma)}^2 \leq c_{\mathrm{shape}} |V_*|_{H^2(\Omega)}^2, \end{aligned} \tag{4.12}$$

and

$$\sum_{\sigma \in \mathcal{G}_{P_*}} \left(h_\sigma \|\Delta(V_* - I_P V_*)\|_{L^2(\sigma)}^2 + h_\sigma^3 \left\| \tfrac{\partial \Delta}{\partial \boldsymbol{n}_\sigma} (V_* - I_P V_*) \right\|_{L^2(\sigma)}^2 \right) \leq c_{\mathrm{shape}} |V_*|_{H^2(\Omega)}^2. \tag{4.13}$$

Proof. Since $\mathbb{X}_{P_*} \subset H^2(\Omega)$, applying Lemma 2.42 gives

$$\begin{aligned} h_\tau^{-4}\|V_* - I_P V_*\|^2_{L^2(\tau)} + \sum_{\sigma\subset\partial\tau} h_\sigma^{-3}\|V_* - I_P V_*\|^2_{L^2(\sigma)} \\ + \sum_{\sigma\subset\partial\tau} h_\sigma^{-1}\left\|\tfrac{\partial}{\partial \boldsymbol{n}_\sigma}(V_* - I_P V_*)\right\|^2_{L^2(\sigma)} \leq c_{\text{shape}}|V_*|^2_{H^2(\omega_\tau)}, \end{aligned} \tag{4.15}$$

where an edge $\sigma \subset \partial\tau$ has a support extension $\omega_\sigma \subset \omega_\tau$. By summing over $\tau \in P_*$ and using the shape regularity we obtain the estimate (4.12).

For a boundary edge $\sigma \in \mathcal{G}_{P_*}$, $(V_* - I_P V_*)|_{\omega_\sigma}$ belongs to $\mathbb{X}_P$. Hence, in view of the inverse estimates of Lemma 2.36, we can obtain inequalities $\|\Delta(V_* - I_P V_*)\|_{L^2(\sigma)} \preceq h_\sigma^{-2}\|V_* - I_P V_*\|_{L^2(\sigma)}$ and $\|\frac{\partial\Delta}{\partial \boldsymbol{n}_\sigma}(V_* - I_P V_*)\|_{L^2(\sigma)} \preceq h_\sigma^{-3}\,\|V_* - I_P V_*\|_{L^2(\sigma)}$. Using Lemma 2.42, we obtain $\|V_* - I_P V_*\|_{L^2(\sigma)} \preceq h_\sigma^{3/2}|V|_{H^2(\omega_\sigma)}$.

$$h_\sigma\|\Delta(V_* - I_P V_*)\|^2_{L^2(\sigma)} + h_\sigma^3\left\|\tfrac{\partial\Delta}{\partial \boldsymbol{n}_\sigma}(V_* - I_P V_*)\right\|^2_{L^2(\sigma)} \leq c_{\text{shape}}|V_*|^2_{H^2(\omega_\sigma)}. \tag{4.16}$$

We arrive the estimate (4.13) by summing over the boundary edges and invoking shape-regularity. ■

4.3 *A posteriori* error analysis

In this section, we derive an *a posteriori* error estimator and prove that the estimator is reliable with respect to the mesh-dependent norm. We first begin by deriving an upper bound on the error measured in $\|\!|\cdot|\!\|_P$. Then, we demonstrate that the stabilization terms can be dominated by the interior error indicators for sufficiently large stabilization parameters.

Our study is based on an analysis introduced in (Juntunen and Stenberg [72]) for the Poisson problem. Controlling the boundary terms follows techniques inspired by the work (Bonito et al. [63]) in the treatment of discontinuous Galerkin methods.

First we recall some results that are of use to us. The bilinear form is automatically bounded since it is defined on a finite-dimensional space. Although the bilinear form $\mathscr{B}_P$ is initially defined on $\mathbb{X}_P \times \mathbb{X}_P$, it can be extended to $H^s(\Omega) \times \mathbb{X}_P$ with $s > 7/2$ using the definition and it is bounded: For a constant $C_{\text{cont}} > 0$,

$$|\mathscr{B}_P(u,\chi)| \leq C_{\text{cont}}\|\!|u|\!\|_\ell\|\chi\|_\ell \quad \text{for all } (u,\chi) \in H^s(\Omega)\times\mathbb{X}_P,\ s > \tfrac{7}{2}. \tag{4.17}$$

Theorem 4.2 (Coercivity of $\mathscr{B}_P$). *There exists stabilization parameters $\gamma_1 > 0$ and $\gamma_2 > 0$ large enough such*

$$\mathscr{B}_P(V,V) \geq C_{\text{coer}}\|\!|V|\!\|_P^2 \quad \textit{for every } V \in \mathbb{X}_P \tag{4.18}$$

for a constant $C_{\text{coer}} > 0$ that depends only on the stabilization parameters γ_1 and γ_2.

Remark 4.3. The choice for the γs will depend on the non-dimensional constants ε_{s} and ε_{m}, as well as, the proportionality constants of Lemma 2.36. See (Kim et al. [36]) for a detailed discussion. We note that we develop a detailed coercivity argument in the next chapter when treating the nonlinear SQGE.

Lemma 4.4 (Consistency). *Let u be a smooth solution to* (4.2)*. Then,*

$$\mathscr{B}_P(u, V) = \ell_f(V) \quad \textit{for every } V \in \mathbb{X}_P. \tag{4.19}$$

Proof. Detail proof is in (Kim et al. [36]). We carry our own partial-consistency proof for the nonlinear SQGE in the next chapter. ∎

Theorem 4.5. *For all cells $\tau \in P$ and interior edges $\sigma \in \mathcal{I}^\ell$, let*

$$R_\tau = \left(f + \varepsilon_s \Delta U - \varepsilon_m \Delta^2 U + \tfrac{\partial U}{\partial x}\right)\Big|_\tau \quad \textit{and} \quad J_\sigma = \left|\left[\tfrac{\partial \Delta U}{\partial \boldsymbol{n}_\sigma}\right]\right|_\sigma. \tag{4.20}$$

If

$$\eta_P^2(\Omega) = \sum_{\tau \in P} h_\tau^4 \|R_\tau\|^2_{L^2(\tau)} + \sum_{\sigma \in \mathcal{I}_P} h_\sigma^3 \|J_\sigma\|^2_{L^2(\sigma)}, \tag{4.22}$$

then

$$|\!|\!| u - U |\!|\!|_P \preceq C_{\mathrm{rel}} \left(\eta_P(\Omega) + \gamma_1 \|U\|_{3/2,P} + \gamma_2 \left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}\right), \tag{4.23}$$

for a constant $C_{\mathrm{rel}} > 0$.

Remark 4.6. It is worth nothing that the power of γ_1 and γ_2 is half an order higher than defined in (4.9). This will remain true for the nonlinear SQGE in the following chapter. It will be clear from Corollary 4.9 below that the estimate (4.23) is not sharp. In the convergence analysis of Chapter 6 we will need a sharper estimate. For practical purposes the sub-optimal estimation is not of much relevance.

Proof. The so-called saturation assumption claims that there exists a constant $0 < \rho < 1$ such that if $P_* = \mathbf{REFINE}(P, \mathscr{M})$ we have

$$|\!|\!| u - U_* |\!|\!|_{P_*} \leq \rho |\!|\!| u - U |\!|\!|_P. \tag{4.24}$$

By estimating

$$|\!|\!| U - U_* |\!|\!|_{P_*} \geq |\!|\!| U - u |\!|\!|_{P_*} - |\!|\!| u - U_* |\!|\!|_{P_*} \geq |\!|\!| U - u |\!|\!|_P - \rho |\!|\!| u - U |\!|\!|_P, \tag{4.25}$$

we may write

$$|\!|\!| u - U |\!|\!|_P \leq \frac{1}{1-\rho} |\!|\!| U_* - U |\!|\!|_{P_*}. \tag{4.26}$$

Our proof boils down to estimating the discrete error $|\!|\!| U_* - U |\!|\!|_{P_*}$. It suffices to bound the error $|\!|\!| U - U_* |\!|\!|_{P_*}$ using coercivity of $\mathscr{B}_{P_*}$, i.e.,

$$C_{\mathrm{coer}} |\!|\!| U - U_* |\!|\!|^2_{P_*} \leq \mathscr{B}_{P_*}(U - U_*, U - U_*). \tag{4.27}$$

If we set $V_* = \frac{U-U_*}{|\!|\!| U-U_* |\!|\!|_{P_*}}$, then $|\!|\!| V_* |\!|\!|_{P_*} = 1$ and

$$C_{\mathrm{coer}} |\!|\!| U - U_* |\!|\!|_{P_*} \leq \mathscr{B}_{P_*}(U - U_*, V_*). \tag{4.28}$$

Introducing $-I_P V_* + I_P V_*$ into $\mathscr{B}_{P_*}$ yields

$$\mathscr{B}_{P_*}(U - U_*, V_*) = \mathscr{B}_{P_*}(U - U_*, V_* - I_P V_*) + \mathscr{B}_{P_*}(U - U_*, I_P V_*). \tag{4.29}$$

For convenience, we denote the first term as W_1 and the second term as W_2 of (4.29), i.e.,

$$W_1 = \mathscr{B}_{P_*}(U - U_*, V_* - I_P V_*), \tag{4.30}$$

$$W_2 = \mathscr{B}_{P_*}(U - U_*, I_P V_*). \tag{4.31}$$

We first start by estimating W_1. From (4.4) the equality $\mathscr{B}_{P_*}(U_*, V_* - I_P V_*) = \ell_f(V_* - I_P V_*)$ holds. As a result, (4.30) is written as

$$W_1 = \mathscr{B}_{P_*}(U, V_* - I_P V_*) - \ell_f(V_* - I_P V_*). \tag{4.32}$$

By denoting $E_* = V_* - I_P V_*$, W_1 can be rewritten as

$$\begin{aligned} W_1 = &- \int_\Omega f E_* - \varepsilon_{\mathrm{s}} \nabla U \cdot \nabla E_* - \varepsilon_{\mathrm{m}} \Delta U \Delta E_* + \tfrac{\partial U}{\partial x} E_* - \varepsilon_{\mathrm{s}} \int_\Gamma \left(\tfrac{\partial U}{\partial \boldsymbol{n}} E_* + U \tfrac{\partial E_*}{\partial \boldsymbol{n}} \right) \\ &+ \varepsilon_{\mathrm{m}} \int_\Gamma \left(\tfrac{\partial \Delta U}{\partial \boldsymbol{n}} E_* + U \tfrac{\partial \Delta E_*}{\partial \boldsymbol{n}} \right) - \varepsilon_{\mathrm{m}} \int_\Gamma \left(\Delta U \tfrac{\partial E_*}{\partial \boldsymbol{n}} + \tfrac{\partial U}{\partial \boldsymbol{n}} \Delta E_* \right) \\ &+ \sum_{\sigma \in \mathcal{G}_{P_*}} \left(\gamma_1 h_\sigma^{-3} \int_\sigma U E_* + \gamma_2 h_\sigma^{-1} \int_\sigma \tfrac{\partial U}{\partial \boldsymbol{n}} \tfrac{\partial E_*}{\partial \boldsymbol{n}} \right). \end{aligned} \tag{4.34}$$

If we decompose $\int_\Omega$ in terms of the partition P_*,

$$\int_\Omega \varepsilon_{\mathrm{s}} \nabla U \cdot \nabla E_* + \varepsilon_{\mathrm{m}} \Delta U \Delta E_* - \tfrac{\partial U}{\partial x} E_* = \sum_{\tau \in P_*} \int_\tau \varepsilon_{\mathrm{s}} \nabla U \cdot \nabla E_* + \varepsilon_{\mathrm{m}} \Delta U \Delta E_* - \tfrac{\partial U}{\partial x} E_*. \tag{4.35}$$

For $\tau \in P_*$, applying Green identity yields

$$\begin{aligned}\int_\tau \varepsilon_{\mathrm{s}} \nabla U \cdot \nabla E_* + \varepsilon_{\mathrm{m}} \Delta U \Delta E_* = \int_\tau \left(-\varepsilon_{\mathrm{s}} \Delta U + \varepsilon_{\mathrm{m}} \Delta^2 U\right) E_* + \varepsilon_{\mathrm{s}} \oint_{\partial\tau} (\nabla U \cdot \boldsymbol{n}_\tau) E_* \\ - \varepsilon_{\mathrm{m}} \oint_{\partial\tau} (\nabla \Delta U \cdot \boldsymbol{n}_\tau) E_* + \varepsilon_{\mathrm{m}} \oint_{\partial\tau} \Delta U \nabla E_* \cdot \boldsymbol{n}_\tau\end{aligned} \tag{4.37}$$

where $\boldsymbol{n}_\tau$ denotes a unit outward normal vector on $\partial\tau$. We observe that

$$\oint_{\partial\tau} \chi|_{\partial\tau} = \sum_{\sigma\in\mathcal{I}_\tau} \int_\sigma (\chi|_\tau)|_\sigma \implies \sum_{\tau\in P_*} \oint_{\partial\tau} \chi|_{\partial\tau} = \sum_{\sigma\in\mathcal{I}_{P_*}} \oint \chi|_\sigma \tag{4.38}$$

where a set of interior edges $\mathcal{I}_\tau = \{\sigma \in \mathcal{I}_P : |\sigma \cap \bar{\tau}| > 0\}$ and $\partial\tau = \cup\{\sigma : \sigma \in \mathcal{I}_\tau\}$. If $\sigma \in \mathcal{I}_{P_*}$ shares two cells τ and Q, $\chi|_\sigma = (\chi|_\tau)|_\sigma$ if $\sigma \subset \partial\tau$ and $\chi|_\sigma = (\chi|_Q)|_\sigma$ if $\sigma \subset \partial Q$. Then,

$$\int_\sigma \chi = \int_\sigma (\chi|_\tau)|_\sigma + (\chi|_Q)|_\sigma. \tag{4.39}$$

Using $-\boldsymbol{n}_Q = \boldsymbol{n}_\tau =: n_\sigma$ and (4.39), we have

$$\int_\sigma [\nabla(U|_\tau) \cdot \boldsymbol{n}_\tau]|_\sigma E_* + [\nabla(U|_Q) \cdot \boldsymbol{n}_Q]|_\sigma E_* = \int_\sigma [\nabla(U|_\tau - U|_Q) \cdot \boldsymbol{n}_\tau]|_\sigma E_*. \tag{4.40}$$

Since $[\![\nabla U \cdot \boldsymbol{n}_\sigma]\!]_\sigma = [\nabla(U|_\tau - U|_Q) \cdot \boldsymbol{n}_\tau]|_\sigma$, we have

$$\sum_{\tau\in P_*} \oint_{\partial\tau} (\nabla U \cdot \boldsymbol{n}_\tau) E_* = \sum_{\sigma\in\mathcal{I}_{P_*}} \int [\![\nabla U \cdot \boldsymbol{n}_\sigma]\!]_\sigma E_* = 0 \tag{4.41}$$

using the second equality in (4.38). The jump term $[\![\nabla U \cdot \boldsymbol{n}_\sigma]\!]_\sigma = 0$ because $U \in \mathcal{C}^2(\Omega)$. We apply (4.39) to $\oint_{\partial\tau} \Delta U \nabla E_* \cdot \boldsymbol{n}_\tau$ to conclude that the jump terms $[\![\Delta U]\!]_\sigma$ are also zero. Therefore, the last terms of the first and second lines on the right of (4.37) vanish. On the other hand, the first term in the second line of (4.37) survives due to the presence of jumps in their third derivatives:

$$\sum_{\tau\in P_*} \oint_{\partial\tau} (\nabla \Delta U \cdot \boldsymbol{n}_\tau) E_* = \sum_{\sigma\in\mathcal{I}_{P_*}} \int_\sigma [\![\nabla \Delta U \cdot \boldsymbol{n}_\sigma]\!]_\sigma E_*. \tag{4.42}$$

Applying above results and (4.37) into (4.34) yields

$$\begin{aligned} W_1 = &-\sum_{\tau\in P_*}\int_\tau (f-\mathcal{L}U)\,E_* - \varepsilon_{\mathrm{s}}\int_\Gamma U\nabla E_*\cdot\boldsymbol{n} \\ &+\varepsilon_{\mathrm{m}}\int_\Gamma U\nabla\Delta E_*\cdot\boldsymbol{n} + \sum_{\sigma\in\mathcal{I}_{P_*}}\int_\sigma \varepsilon_{\mathrm{m}}\,[\![\nabla\Delta U\cdot\boldsymbol{n}_\sigma]\!]_\sigma\,E_* - \varepsilon_{\mathrm{m}}\int_\Gamma(\nabla U\cdot\boldsymbol{n})\Delta E_* \\ &+\gamma_1\sum_{\sigma\in\mathcal{G}_{P_*}} h_\sigma^{-3}\int_\sigma UE_* + \gamma_2\sum_{\sigma\in\mathcal{G}_{P_*}}\int_\sigma h_\sigma^{-1}(\nabla U\cdot\boldsymbol{n}_\sigma)(\nabla E_*\cdot\boldsymbol{n}_\sigma). \end{aligned} \tag{4.44}$$

Let

$$W_1 = R_1 + R_2 + R_3, \tag{4.45}$$

where

$$\begin{aligned} R_1 =&\ -\sum_{\tau\in P_*}\int_\tau R_\tau E_* + \varepsilon_{\mathrm{m}}\sum_{\sigma\in\mathcal{I}_{P_*}}\int_\sigma J_\sigma E_*, \\ R_2 =&\ -\varepsilon_{\mathrm{s}}\int_\Gamma U\tfrac{\partial E_*}{\partial\boldsymbol{n}} - \varepsilon_{\mathrm{m}}\int_\Gamma \tfrac{\partial U}{\partial\boldsymbol{n}}\Delta E_* + \varepsilon_{\mathrm{m}}\sum_{\sigma\in\mathcal{I}_{P_*}}\int_\sigma U\tfrac{\partial\Delta E_*}{\partial\boldsymbol{n}}, \\ R_3 =&\ \gamma_1\sum_{\sigma\in\mathcal{G}_{P_*}} h_\sigma^{-3}\int_\sigma UE_* + \gamma_2\sum_{\sigma\in\mathcal{G}_{P_*}} h_\sigma^{-1}\int_\sigma \tfrac{\partial U}{\partial\boldsymbol{n}_\sigma}\tfrac{\partial E_*}{\partial\boldsymbol{n}_\sigma}. \end{aligned}$$

To estimate R_1, we have

$$\begin{aligned} R_1 =&\ -\sum_{\tau\in P_*}\int_\tau R_\tau E_* + \varepsilon_{\mathrm{m}}\sum_{\sigma\in\mathcal{I}_{P_*}}\int_\sigma J_\sigma E_*, \\ \le&\ \sum_{\tau\in P_*}\|R_\tau\|_{L^2(\tau)}\,\|E_*\|_{L^2(\tau)} + \varepsilon_{\mathrm{m}}\sum_{\sigma\in\mathcal{I}_{P_*}}\|J_\sigma\|_{L^2(\sigma)}\,\|E_*\|_{L^2(\sigma)}, \\ \le&\ \Big(\sum_{\tau\in P_*} h_\tau^4\,\|R_\tau\|^2_{L^2(\tau)}\Big)^{1/2}\Big(\sum_{\tau\in P_*} h_\tau^{-4}\|E_*\|^2_{L^2(\tau)}\Big)^{1/2}, \\ &+\varepsilon_{\mathrm{m}}\Big(\sum_{\sigma\in\mathcal{I}_{P_*}} h_\sigma^3\,\|J_\sigma\|^2_{L^2(\sigma)}\Big)^{1/2}\Big(\sum_{\sigma\in\mathcal{I}_{P_*}} h_\sigma^{-3}\|E_*\|^2_{L^2(\sigma)}\Big)^{1/2}, \end{aligned} \tag{4.47}$$

The approximation estimate (4.12) together with the fact $|\!|\!|V_*|\!|\!|^2_{P_*} = 1$ can be rewritten as

$$\sum_{\tau\in P_*} h_\tau^{-4}\|V_* - I_P V_*\|^2_{L^2(\tau)} + \sum_{\sigma\in\mathcal{I}_{P_*}} h_\sigma^{-3}\|V_* - I_P V_*\|^2_{L^2(\sigma)} \preceq c_{\mathrm{shape}}. \tag{4.48}$$

Applying above inequality to (4.47) completes the estimate of R_1 by

$$R_1 \preceq c_{\text{shape}} \left\{ \left(\sum_{\tau \in P_*} h_\tau^4 \left\| R_\tau \right\|_{L^2(\tau)}^2 \right)^{1/2} + \left(\sum_{\sigma \in \mathcal{I}_{P_*}} h_\sigma^3 \left\| J_\sigma \right\|_{L^2(\sigma)}^2 \right)^{1/2} \right\}. \tag{4.50}$$

The simple fact that $a + b \leq \sqrt{2}(a^2 + b^2)^{1/2}$ holds for all positive real numbers will be useful from now and onward to show that

$$\left(\sum_{\tau \in P_*} h_\tau^4 \left\| R_\tau \right\|_{L^2(\tau)}^2 \right)^{1/2} + \left(\sum_{\sigma \in \mathcal{I}_{P_*}} h_\sigma^3 \left\| J_\sigma \right\|_{L^2(\sigma)}^2 \right)^{1/2} \preceq \eta_{P_*}(\Omega). \tag{4.51}$$

Similarly, we estimate R_2

$$\begin{aligned} R_2 \leq \varepsilon_{\text{s}} & \Big(\sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma \|U\|_{L^2(\sigma)}^2 \Big)^{1/2} \Bigg(\sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma^{-1} \left\| \tfrac{\partial}{\partial \boldsymbol{n}_\sigma}(V_* - I_P V_*) \right\|_{L^2(\sigma)}^2 \Bigg)^{1/2} \\ + \varepsilon_{\text{m}} & \Big(\sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma^{-3} \|U\|_{L^2(\sigma)}^2 \Big)^{1/2} \Bigg(\sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma^{3} \left\| \tfrac{\partial \Delta}{\partial \boldsymbol{n}_\sigma}(V_* - I_P V_*) \right\|_{L^2(\sigma)}^2 \Bigg)^{1/2} \\ + \varepsilon_{\text{m}} & \Bigg(\sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma^{-1} \left\| \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)}^2 \Bigg)^{1/2} \Big(\sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma \|\Delta(V_* - I_P V_*)\|_{L^2(\sigma)}^2 \Big)^{1/2}. \end{aligned} \tag{4.53}$$

The final step is to apply Corollary 4.1 on the projection errors. Finally, assume that $h_\sigma \leq h_\sigma^{-3}$, the estimate for R_2 is therefore

$$R_2 \leq c_{\text{shape}} \left(\max\{\varepsilon_{\text{s}}, \varepsilon_{\text{m}}\} \|U\|_{3/2,P_*} + \varepsilon_{\text{m}} \left\| \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \right\|_{1/2,P_*} \right). \tag{4.54}$$

In a similar fashion, we can obtain the estimate of R_3. Applying Cauchy-Schwarz's inequality to R_3 yields

$$\begin{aligned} R_3 \leq \ & \gamma_1 \sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma^{-3/2} \|U\|_{L^2(\sigma)} h_\sigma^{-3/2} \|E_*\|_{L^2(\sigma)} \\ & + \gamma_2 \sum_{\sigma \in \mathcal{G}_{P_*}} h_\sigma^{-1/2} \left\| \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)} h_\sigma^{-1/2} \left\| \tfrac{\partial E_*}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)} \\ \leq \ & \gamma_1 \|U\|_{3/2,P_*} \|E_*\|_{3/2,P_*} + \gamma_2 \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|_{1/2,P_*} \left\| \tfrac{\partial E_*}{\partial \boldsymbol{n}} \right\|_{1/2,P_*}. \end{aligned} \tag{4.56}$$

In view of estimates (4.12) and (4.13) of Corollary 4.1 we write,

$$R_3 \leq c_{\text{shape}} \left(\gamma_1 \|U\|_{3/2,P_*} + \gamma_2 \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|_{1/2,P_*} \right). \tag{4.57}$$

Upon applying (4.50), (4.54), and (4.57) to W_1, we obtain the estimate for W_1 by

$$\begin{aligned} W_1 = R_1 + R_2 + R_3 \preceq c_{\text{shape}}\eta_{P_*}(\Omega) + c_{\text{shape}}\left(\max\{\varepsilon_{\text{s}}, \varepsilon_{\text{m}}\} + \gamma_1\right) \|U\|_{3/2,P_*} \\ + c_{\text{shape}}\left(\varepsilon_{\text{m}} + \gamma_2\right) \left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P_*}. \end{aligned} \tag{4.59}$$

To estimate W_2 in (4.29), we use $\mathscr{B}_{P_*}(U_*, I_P V_*) = \ell_f(I_P V_*)$ due to the nesting $\mathbb{X}_P \subset \mathbb{X}_{P_*}$ and $\mathscr{B}_P(U, I_P V_*) - \ell_f(I_P V_*) = 0$ to obtain

$$\begin{aligned} W_2 &= \mathscr{B}_{P_*}(U, I_P V_*) - f(I_P V_*) + \mathscr{B}_P(U, I_P V_*) - \ell_f(I_P V_*) \\ &= \mathscr{B}_{P_*}(U, I_P V_*) - \mathscr{B}_P(U, I_P V_*). \end{aligned} \tag{4.61}$$

Above result indicates that the only terms involving in W_2 are those without edge diameter factors because both U and $I_P V_*$ are members of $\mathbb{X}_P$. To avoid confusion, we denote edges from $\mathcal{G}_{P_*}$ by σ and edges from Γ^ℓ by E. Then,

$$\begin{aligned} W_2 = &\sum_{\sigma \in \mathcal{G}_{P_*}} \left(\gamma_1 h_\sigma^{-3} \int_\sigma U I_P V_* + \gamma_2 h_\sigma^{-1} \int_\sigma \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \tfrac{\partial (I_P V_*)}{\partial \boldsymbol{n}_\sigma}\right) \\ &- \sum_{E \in \mathcal{G}_P} \left(\gamma_1 h_E^{-3} \int_E U I_P V_* + \gamma_2 h_E^{-1} \int_E \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \tfrac{\partial (I_P V_*)}{\partial \boldsymbol{n}_\sigma}\right). \end{aligned} \tag{4.63}$$

Since $h_E = 2h_\sigma$ for $\sigma \subset E$,

$$W_2 = \sum_{\sigma \in \mathcal{G}_{P_*}} \left\{\gamma_1 \left(1 - \frac{1}{8}\right) h_\sigma^{-3} \int_\sigma U I_P V_* + \gamma_2 \left(1 - \frac{1}{2}\right) h_\sigma^{-1} \int_\sigma \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \tfrac{\partial (I_P V_*)}{\partial \boldsymbol{n}_\sigma}\right\}. \tag{4.64}$$

Applying stability of the projector I_P

$$\sum_{\sigma \in \mathcal{G}_{P_*}} \left\{h_\sigma^{-3} \|I_P V_*\|_{L^2(\sigma)}^2 + h_\sigma^{-1} \left\|\tfrac{\partial (I_P V_*)}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}^2\right\} \leq c_{\text{shape}}, \tag{4.65}$$

completes the estimate of W_2 as

$$W_2 \leq c_{\text{shape}} \left(\gamma_1 \|U\|_{3/2,P_*} + \gamma_2 \left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P_*}\right). \tag{4.66}$$

Summing up (4.59) and (4.66), we obtain the upper bound

$$|\!|\!| U_* - U |\!|\!|_{P_*} \preceq \eta_{P_*}(\Omega) + \gamma_1 \|U\|_{3/2,P_*} + \gamma_2 \left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P_*}, \tag{4.68}$$

where the implicit constant $\preceq$ depends on c_{shape}, ε_{s} and ε_{m}. The right-hand side of the expression is with respect to P_*, but $\eta_P(\Omega) \leq \eta_{P_*}(\Omega)$ whenever $P_* \geq P$, moreover, since P_* is just one iteration after P, $\|\cdot\|_{s,P_*} \preceq \|\cdot\|_{s,P}$. Substituting (4.68) into (4.26) completes the proof of (4.23). Substituting (4.68) into (4.26) completes the proof of (4.23). ■

The remainder of this section is dedicated to control the boundary terms (i.e., last two terms) in (4.68). We prove that, for suitably large stabilization parameters γ_1 and γ_2, the error $|\!|\!| u - U |\!|\!|_P$ is bounded by the error estimator $\eta_P(\Omega)$ only.

Lemma 4.7. *Let $V \in \mathbb{X}_P$. There are finite-element functions $V_0 \in H^2_0(\Omega) \cap \mathbb{X}_P$ and $V_\perp \in \mathbb{X}_P$ such that the decomposition $V = V_0 + V_\perp$ holds with*

$$|V - V_0|^2_{H^s(\omega_\sigma)} \le C_{\mathrm{bdry}} \left(h_\sigma^{-2s+1} \|V_\perp\|^2_{L^2(\sigma)} + h_\sigma^{-2s+3} \left\| \tfrac{\partial V_\perp}{\partial \boldsymbol{n}_\sigma} \right\|^2_{L^2(\sigma)} \right) \quad \forall \sigma \in \mathcal{G}_P, \tag{4.70}$$

with a constant $C_{\mathrm{bdry}} > 0$ depending only on c_{shape}.

Proof. Let $\mathbb{X}_P = \mathrm{span}(N_i)_{i \in J}$ and let $X_0 = \mathbb{X}_P \cap H^2_0(\Omega)$ with $\{N_i : i \in J\}$ being the B-spline basis functions for $\mathbb{X}_P$. Introduce the indexing set $J_{\mathrm{int}} = \{j \in J : \mathrm{supp}\,(N_i) \subset \Omega\}$ and $J_{\mathrm{ext}} = J_\ell \backslash J_{\mathrm{int}}$. Let σ be an edge in $\mathcal{G}_P$. We claim that the semi-norm

$$|\cdot|^2_\sigma = \|\cdot\|^2_{L^2(\sigma)} + \left\| \tfrac{\partial}{\partial \boldsymbol{n}_\sigma} \cdot \right\|^2_{L^2(\sigma)} \tag{4.71}$$

defines a norm on $\mathbb{X}_P^\perp = \mathrm{span}(\{N_i : i \in J_{\mathrm{ext}}\,\text{such that}\,\mathrm{supp}\ N_i \subset \omega_\sigma\})$. Indeed, $|\cdot|_\sigma$ trivially satisfies the triangular inequality and norm homogeneity property. We show that it separates points in $\mathbb{X}_P^\perp$. Let $V \in \mathbb{X}_P^\perp$. If $|V|_\sigma = 0$, the polynomial $V|_\sigma$ is identically zero. Consequently, V is identically zero on ω_σ because every basis function $N_i > 0$ on ω_σ for $i \in J_{\mathrm{ext}} \cap \{j : \mathrm{supp}\ N_j \subset \omega_\sigma\}$.

For $V \in \mathbb{X}_P$ let $V_0 = \sum_{i \in J_{\mathrm{int}}} \lambda_i N_i$ and let $V_\perp = \sum_{i \in J_{\mathrm{ext}}} \lambda_i N_i$. Then, $V_0 \in H^2_0(\Omega)$ and $V = V_0 + V_\perp$. More importantly, $V - V_0 \in \mathbb{X}_P^\perp$. Scale the subdomain ω_σ to $\hat\omega$ with $\mathrm{diam}(\hat\omega) = 1$ via an affine map $A : \hat\omega \ni \hat x \mapsto b + T\hat x \in \omega_\sigma$, where $b \in \mathbb{R}^2$ and T is an invertible matrix, and denote the scaling of S by $\hat\sigma = A(\sigma)$. By the equivalence of norms in finite-dimensions (see Lemma 2.36), we have

$$|\hat V|_{L^2(\hat\sigma)} \preceq |\hat V_\perp|_{H^s(\hat\omega)} \preceq |\hat V|_{L^2(\hat\sigma)} \quad \forall \hat V_\perp \in \mathbb{X}_P^\perp, \tag{4.72}$$

where $\preceq$ depends only on the polynomial degree. We scale back to ω_σ to obtain

$$|T^{-1}|^{-s} |\det(T)|^{-1/2} |V_\perp|_{H^s(\omega_\sigma)} \le |\hat V_\perp|_{H^s(\hat\omega)} \tag{4.73}$$

and

$$|\hat\chi_\ell|_{H^t(\hat\sigma)} \le |(T|_\sigma)|^t |\det(T|_\sigma)|^{-1/2} |V|_{H^t(\sigma)}, \tag{4.74}$$

with $\det(T) = \frac{\mathrm{meas}(\omega_\sigma)}{\mathrm{meas}(\hat\omega)} \le C_{\mathrm{shape}} h_\sigma^2$, $|T^{-1}|_{\mathbb{R}^{2\times 2}} \le \frac{\mathrm{diam}(\hat\omega)}{h_\sigma} \le h_\sigma^{-1}$, $\det(T|_\sigma) = \frac{\mathrm{meas}(\sigma)}{\mathrm{meas}(\hat\sigma)} \preceq h_\sigma$ and $|(T|_\sigma)|_{\mathbb{R}^{1\times 1}} \le \frac{h_\sigma}{\hat\rho} \preceq h_\sigma$. We then arrive at

$$C_{\mathrm{shape}}^{-1} h_\sigma^{2s-2} |V_\perp|^2_{H^s(\omega_\sigma)} \le h_\sigma^{-1} \|V_\perp\|^2_{L^2(\sigma)} + h_\sigma \left\| \tfrac{\partial V_\perp}{\partial \boldsymbol{n}_\sigma} \right\|^2_{L^2(\sigma)}. \tag{4.75}$$

■

Theorem 4.8. *We have*

$$(\gamma_1 - C_{\text{res}})\|U\|_{3/2,P}^2 + (\gamma_2 - C_{\text{res}})\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2 \preceq \frac{\max\{1,\varepsilon_s,\varepsilon_m\}}{C_{\text{coer}}}\eta_P^2(\Omega), \tag{4.76}$$

with constant $C_{\text{res}} = \frac{3\max\{1,\varepsilon_s,\varepsilon_m\}}{2C_{\text{coer}}}C_{\text{bdry}}$.

Proof. Let $\chi \in X_0$

$$\begin{aligned}\gamma_1\|U\|_{3/2,P}^2 + \gamma_2\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2 &\leq C_{\text{coer}}^{-1}\,\mathscr{B}_P(U-V_0,U-V_0)\\ &= C_{\text{coer}}^{-1}\left\{\mathscr{B}_P(U,U-V_0) - \mathscr{B}_P(V_0,U-V_0)\right\}\\ &= C_{\text{coer}}^{-1}\left\{\ell_f(U-V_0) - \mathscr{B}_P(V_0,U-V_0)\right\}.\end{aligned} \tag{4.78}$$

Let $V_0 = U + (V_0 - U)$ and express

$$\mathscr{B}_P(V_0,U-V_0) = \mathscr{B}_P(U,U-V_0) - \mathscr{B}_P(U-V_0,U-V_0). \tag{4.79}$$

Using above equality, we obtain

$$\begin{aligned}\ell_f(U-V_0)-\mathscr{B}_P(V_0,U-V_0) &= \ell_f(U-V_0) - \mathscr{B}_P(U,U-V_0) + \mathscr{B}_P(U-V_0,U-V_0),\\ &= \sum_{\tau\in P}\int_\tau R_\tau(U-V_0) + \varepsilon_{\text{m}}\sum_{\sigma\in\mathcal{I}^\ell}\int_\sigma J_\sigma(U-V_0)\\ &\quad - \varepsilon_{\text{s}}\int_\Gamma U\tfrac{\partial(U-V_0)}{\partial\boldsymbol{n}} - \varepsilon_{\text{m}}\int_\Gamma U\tfrac{\partial\Delta(U-V_0)}{\partial\boldsymbol{n}} + \varepsilon_{\text{m}}\int_\Gamma \tfrac{\partial U}{\partial\boldsymbol{n}}\Delta(U-V_0)\\ &\quad + \sum_{\sigma\in\mathcal{G}_P}\left(\gamma_1 h_\sigma^{-3}\int_\sigma U(U-V_0) + \gamma_2 h_\sigma^{-1}\int_\sigma \tfrac{\partial U}{\partial\boldsymbol{n}_\sigma}\tfrac{\partial(U-V_0)}{\partial\boldsymbol{n}_\sigma}\right)\\ &\quad + \mathscr{B}_P(U-V_0,U-V_0).\end{aligned} \tag{4.81}$$

Then, we have

$$\begin{aligned}\ell_f(U-V_0)-\mathscr{B}_P(V_0,U-V_0) &= \varepsilon_{\text{s}}\|\nabla(U-V_0)\|_\Omega^2 + \varepsilon_{\text{m}}\|\Delta(U-V_0)\|_\Omega^2\\ &\quad + \sum_{\tau\in P}\int_\tau R_\tau(U-V_0) + \varepsilon_{\text{m}}\sum_{\sigma\in\mathcal{I}^\ell}\int_\sigma J_\sigma(U-V_0)\\ &\quad + \varepsilon_{\text{s}}\int_\Gamma U\tfrac{\partial(U-V_0)}{\partial\boldsymbol{n}} - \varepsilon_{\text{m}}\int_\Gamma U\tfrac{\partial\Delta(U-V_0)}{\partial\boldsymbol{n}} + \varepsilon_{\text{m}}\int_\Gamma \tfrac{\partial U}{\partial\boldsymbol{n}}\Delta(U-V_0).\end{aligned} \tag{4.83}$$

We decompose $U = U_0 + U_\perp$ and let $\mathcal{D}_\Gamma = \overline{\bigcup_{\sigma\in\mathcal{G}_P}\omega_\sigma}$. Since the difference $U - U_0$ is supported only on $\mathcal{D}_\Gamma$, $\|\nabla(U-U_0)\|_{L^2(\Omega)}$ and $\|\Delta(U-u_0)\|_{L^2(\Omega)}$ reduce to $\|\nabla(U-U_0)\|_{L^2(\mathcal{D}_\Gamma)}$. Invoking

Lemma 4.7 on every $\sigma \in \Gamma^\ell$ results in

$$\begin{aligned}
&\varepsilon_{\mathrm{s}}\|\nabla(U-U_0)\|^2_{L^2(\Omega)}+\varepsilon_{\mathrm{m}}\|\Delta(U-u_0)\|^2_{L^2(\Omega)} \\
&\qquad\qquad\le \max\{\varepsilon_{\mathrm{s}},\varepsilon_{\mathrm{m}}\}C_{\mathrm{bdry}}\left(\|U_\perp\|^2_{3/2,P}+\left\|\tfrac{\partial U_\perp}{\partial \boldsymbol{n}}\right\|^2_{1/2,P}\right).
\end{aligned}$$

In (4.83), by replacing V_0 with U_0 and substituting above inequality yield

$$\begin{aligned}
\ell_f(U-U_0)-\mathscr{B}_P(U_0,U-U_0) &\le \varepsilon_{\mathrm{s}}\|\nabla(U-U_0)\|^2_{L^2(\mathcal{D}_\Gamma)}+\varepsilon_{\mathrm{m}}\|\Delta(U-U_0)\|^2_{L^2(\mathcal{D}_\Gamma)} \\
&\quad+\sum_{\tau\in P\cap\mathcal{D}_\Gamma}\|R_\tau\|_{L^2(\tau)}\|U_\perp\|_{L^2(\tau)}+\varepsilon_{\mathrm{m}}\sum_{\sigma\in\mathcal{I}^\ell\cap\mathcal{D}_\Gamma}\|J_\sigma\|_{L^2(\sigma)}\|U_\perp\|_{L^2(\sigma)} \\
&\quad+\sum_{\sigma\in\mathcal{G}_P}\Bigg(\varepsilon_{\mathrm{s}}\|U\|_{L^2(\sigma)}\left\|\tfrac{\partial U_\perp}{\partial\boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}+\varepsilon_{\mathrm{m}}\|U\|_{L^2(\sigma)}\left\|\tfrac{\partial \Delta U_\perp}{\partial\boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)} \\
&\quad+\varepsilon_{\mathrm{m}}\left\|\tfrac{\partial U}{\partial\boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}\|\Delta U_\perp\|_\sigma\Bigg) \\
&\preceq \max\{\varepsilon_{\mathrm{s}},\varepsilon_{\mathrm{m}}\}C_{\mathrm{bdry}}\sum_{\sigma\in\mathcal{G}_P}\left(h_\sigma^{-3}\|U_\perp\|^2_{L^2(\sigma)}+h_\sigma^{-1}\left\|\tfrac{\partial U_\perp}{\partial\boldsymbol{n}_\sigma}\right\|^2_{L^2(\sigma)}\right) \\
&\quad+\Bigg(\sum_{\tau\in P\cap\mathcal{D}_\Gamma}h_K^4\|R_\tau\|^2_{L^2(\tau)}\Bigg)^{1/2}\Bigg(\sum_{\tau\in P\cap\mathcal{D}_\Gamma}h_K^{-4}\|U_\perp\|^2_{L^2(\tau)}\Bigg)^{1/2} \\
&\quad+\varepsilon_{\mathrm{m}}\Bigg(\sum_{\sigma\in\mathcal{I}^\ell\cap\mathcal{D}_\Gamma}h_\sigma^3\|J_\sigma\|^2_{L^2(\sigma)}\Bigg)^{1/2}\Bigg(\sum_{\sigma\in\mathcal{I}^\ell\cap\mathcal{D}_\Gamma}h_\sigma^{-3}\|U_\perp\|^2_{L^2(\sigma)}\Bigg)^{1/2} \\
&\quad+\sum_{\sigma\in\mathcal{G}_P}\Bigg(\varepsilon_{\mathrm{s}}\|U\|_{L^2(\sigma)}|U_\perp|_{H^1(\sigma)}+\varepsilon_{\mathrm{m}}\|U\|_{L^2(\sigma)}|U_\perp|_{H^3(\sigma)} \\
&\quad+\varepsilon_{\mathrm{m}}\left\|\tfrac{\partial U}{\partial\boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}|U_\perp|_{H^2(\sigma)}\Bigg).
\end{aligned} \tag{4.85}$$

In view of (4.72) in Lemma 4.7 and a scaling argument,

$$\|U_\perp\|_{L^2(\sigma)}\preceq h_\sigma^{-1/2}\sum_{K\subset\omega_\sigma\cap\mathcal{D}_\Gamma}\|U_\perp\|_{L^2(\tau)}, \tag{4.86}$$

so we have

$$\begin{aligned}
\sum_{\sigma\in\mathcal{I}^\ell\cap\mathcal{D}_\Gamma}h_\sigma^{-3}\|U_\perp\|^2_{L^2(\sigma)} &\preceq \sum_{\tau\in P\cap\mathcal{D}_\Gamma}h_K^{-4}\|U_\perp\|^2_{L^2(\tau)} \\
&\le C_{\mathrm{bdry}}\Bigg(\|U_\perp\|^2_{3/2,P}+\left\|\tfrac{\partial U_\perp}{\partial\boldsymbol{n}}\right\|^2_{1/2,P}\Bigg),
\end{aligned} \tag{4.88}$$

which makes

$$\begin{aligned}\Bigg(\sum_{\tau\in P\cap\mathcal{D}_\Gamma} h_K^4\|R_\tau\|_{L^2(\tau)}^2\Bigg)^{1/2}&\Bigg(\sum_{\tau\in P\cap\mathcal{D}_\Gamma} h_K^{-4}\|U_\perp\|_{L^2(\tau)}^2\Bigg)^{1/2}\\ &+\varepsilon_{\mathrm{m}}\Bigg(\sum_{\sigma\in\mathcal{I}^\ell\cap\mathcal{D}_\Gamma} h_\sigma^3\|J_\sigma\|_{L^2(\sigma)}^2\Bigg)^{1/2}\Bigg(\sum_{\sigma\in\mathcal{I}^\ell\cap\mathcal{D}_\Gamma} h_\sigma^{-3}\|U_\perp\|_{L^2(\sigma)}^2\Bigg)^{1/2},\\ &\preceq\sqrt{C_{\mathrm{bdry}}}\eta_P(\Omega)\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\}^{1/2}.\end{aligned}\tag{4.90}$$

Moreover, we also have $|U_\perp|_{H^k(\sigma)}\preceq h_\sigma^{3/2-k}|U_\perp|_{H^2(\omega_\sigma)}$ thanks to (4.72). This makes

$$\begin{aligned}\sum_{\sigma\in\mathcal{G}_P}\Bigg(\varepsilon_{\mathrm{s}}\|U\|_{L^2(\sigma)}|U_\perp|_{H^1(\sigma)}&+\varepsilon_{\mathrm{m}}\|U\|_{L^2(\sigma)}|U_\perp|_{H^3(\sigma)}+\varepsilon_{\mathrm{m}}\left\|\tfrac{\partial U}{\partial\boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}|U_\perp|_{H^2(\sigma)}\Bigg)\\ &\le\|U\|_{3/2,P}\left\{\varepsilon_{\mathrm{s}}\Bigg(\sum_{\sigma\in\mathcal{G}_P}h_\sigma^3|U_\perp|_{H^1(\sigma)}^2\Bigg)^{1/2}+\varepsilon_{\mathrm{m}}\Bigg(\sum_{\sigma\in\mathcal{G}_P}h_\sigma^3|U_\perp|_{H^3(\sigma)}^2\Bigg)^{1/2}\right\}\\ &\quad+\varepsilon_{\mathrm{m}}\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}\Bigg(\sum_{\sigma\in\mathcal{G}_P}h_\sigma|U_\perp|_{H^2(\sigma)}^2\Bigg)^{1/2},\\ &\preceq\left(\|U\|_{3/2,P}+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}\right)|U_\perp|_{H^2(\mathcal{D}_\Gamma)},\\ &\preceq\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\}^{1/2}|U_\perp|_{H^2(\mathcal{D}_\Gamma)}.\end{aligned}\tag{4.92}$$

Then

$$\begin{aligned}\ell_f(U-U_0)-\mathscr{B}_P(U_0,U-U_0)&\preceq\max\{\varepsilon_{\mathrm{s}},\varepsilon_{\mathrm{m}}\}C_{\mathrm{bdry}}\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\}\\ &\quad+\sqrt{C_{\mathrm{bdry}}}\eta_P(\Omega)\left\{\|U\|_{3/2,P}^2+\varepsilon_{\mathrm{m}}\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\}^{1/2}\\ &\quad+\sqrt{C_{\mathrm{bdry}}}\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\},\\ &\preceq\sqrt{C_{\mathrm{bdry}}}\eta_P(\Omega)\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\}^{1/2}\\ &\quad+C_{\mathrm{bdry}}\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\}.\end{aligned}\tag{4.94}$$

Writing

$$\eta_P(\Omega)\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\}^{1/2}\le\frac{1}{2}\eta_P^2(\Omega)+\frac{1}{2}\left\{\|U\|_{3/2,P}^2+\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}^2\right\},\tag{4.95}$$

makes

$$\begin{aligned}\gamma_1\|U\|^2_{3/2,P} &+ \gamma_2\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|^2_{1/2,P}\\ &\leq \frac{\max\{1,\varepsilon_{\mathrm{s}},\varepsilon_{\mathrm{m}}\}}{C_{\mathrm{coer}}}\left(\frac{3C_{\mathrm{bdry}}}{2}\left\{\|U\|^2_{3/2,P}+\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|^2_{1/2,P}\right\}+\eta^2_P(\Omega)\right),\end{aligned}\tag{4.97}$$

and we arrive at

$$(\gamma_1 - C_{\mathrm{res}})\|U\|^2_{3/2,P} + (\gamma_2 - C_{\mathrm{res}})\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|^2_{1/2,P} \leq \frac{\max\{1,\varepsilon_{\mathrm{s}},\varepsilon_{\mathrm{m}}\}}{C_{\mathrm{coer}}}\,\eta^2_P(\Omega),\tag{4.98}$$

with residual controlling constant $C_{\mathrm{res}} = \frac{3\max\{1,\varepsilon_{\mathrm{s}},\varepsilon_{\mathrm{m}}\}}{2C_{\mathrm{coer}}}C_{\mathrm{bdry}}$. ■

Corollary 4.9. *For sufficiently large γ_1 and γ_2 the estimator* (4.22) *admits*

$$|\!|\!| u - U |\!|\!|^2_P \leq C_{\mathrm{Rel}}\max\{\gamma_1,\gamma_2\}\eta^2_P(\Omega),\tag{4.99}$$

for a constant $C_{\mathrm{Rel}} > 0$ that depends on c_{shape}, ε_s, ε_m and C_{coer}.

Proof. First pick γ_1 and γ_2 be large enough to satisfy the condition of Lemma 4.2 and $\min\{\gamma_1 - C_{\mathrm{res}}, \gamma_2 - C_{\mathrm{res}}\} > 0$. Then in view of Theorem 4.8 we may write

$$\begin{aligned}\gamma_1\|U\|_{3/2,P} &+ \gamma_2\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}\\ &\leq \max\{\gamma_1,\gamma_2\}\sqrt{2}\left(\|U\|^2_{3/2,P}+\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|^2_{1/2,P}\right)^{1/2},\\ &\leq \sqrt{\tfrac{2\max\{1,\varepsilon_{\mathrm{s}},\varepsilon_{\mathrm{m}}\}}{C_{\mathrm{coer}}}}\,\tfrac{\max\{\gamma_1,\gamma_2\}}{\sqrt{\min\{\gamma_1-C_{\mathrm{res}},\gamma_2-C_{\mathrm{res}}\}}}\eta_P(\Omega).\end{aligned}\tag{4.101}$$

We estimate $\frac{\max\{\gamma_1,\gamma_2\}}{\sqrt{\min\{\gamma_1-C_{\mathrm{res}},\gamma_2-C_{\mathrm{res}}\}}}$. Without loss of generality, assume that $\gamma_2 = c\gamma_1$ with $0 < c < 1$, then it follows that

$$\frac{\gamma_1}{\sqrt{c\gamma_1 - C_{\mathrm{res}}}} = \frac{\gamma_1}{\sqrt{c\gamma_1}}\sqrt{\frac{1}{1-\frac{C_{\mathrm{res}}}{c\gamma_2}}} \preceq \sqrt{\gamma_1/c}\sqrt{1+\frac{C_{\mathrm{res}}}{c\gamma_1}} \preceq \sqrt{\gamma_1}.$$

The result is now immediate from Theorem 4.5. ■

4.4 Numerical study

To study the capability of the *a posteriori* error estimator (4.22), we perform numerical studies on several benchmark problems in geophysical fluid dynamics. Specifically, we provide convergence studies using adaptive refinement approach presented in Section 2.5. For this purpose, we define errors $||e||_{L^2}$, $||e||_{H^1}$, and $||e||_{H^2}$ in the L^2-norm, the H^1-semi norm, and the H^2-semi norm by

$$||e||_{L^2} = \frac{||u-U||_{L^2}}{||u||_{L^2}}, \quad ||e||_{H^1} = \frac{|u-U|_{H^1}}{|u|_{H^1}}, \quad ||e||_{H^2} = \frac{|u-U|_{H^2}}{|u|_{H^2}}, \tag{4.102}$$

respectively, where U is the approximation of u. For adaptivity, we choose $\theta = 0.9$ for the Dörfler marking strategy and the maximum refinement levels to be 4. The choices of the stabilization parameters are based on the analysis of Kim et al. [36].

4.4.1 Western boundary layer in a rectangular domain

We first begin by performing a convergence study on a rectangular domain for the test problem with the closed-form solution

$$u(x,y) = [(1-x/3)(1-e^{-20x})\sin(y)]^2 \quad \text{in } \Omega = [0,3]\times[0,1]. \tag{4.103}$$

This problem has previously been used to test a finite-element algorithm for large-scale ocean circulation problems (Foster et al. [37] and Cascón et al. [76]). Notice that the forcing term f is chosen to match that given by the solution (4.103). We consider a rectangular ocean as a computational domain, as shown in Figure 5.7. With the origin of a Cartesian coordinate system at the southwest corner, the x- and y-axis point eastward and northward, respectively, and the boundaries of the computational domain are the shores of the ocean.

In Figure 4.2, we display the convergence rates for the adaptive refinement along with those for the uniform refinement. Notice that the optimal rates of convergence of a finite-element discretization using cubic B-splines are respectively quartic, cubic, and quadratic in the L^2-, H^1-, and H^2-norms (Rotunda et al. [45]). With the uniform refinement, the rates of convergence appeared in the figure are optimal, i.e., 2.07, 1.56, and 1.02 in the L^2-, H^1-, and H^2-norms, respectively, with respect to the mesh-size in a manner that is consistent with (2.146). The results show that the rates of convergence significantly increase by adding the first two levels of refinement although the rates gradually reduce to reach the optimal rates by adding more refinement levels. This means that the proportionality constant of (2.146) is larger in the case of uniform refinement than it is for the adaptive method. Figure 4.3 shows the refinement levels for the simulation of this test problem. These refinements are achieved using the *a posteriori* error indicator (4.22). In Figure 4.3, the numerical solution with the western boundary layer is displayed. Importantly, the mesh is refined near the western boundary layer, indicating the accuracy and efficiency of the proposed error estimator.

In (4.23), the error is bounded by the error estimator $\eta_P(\Omega)$ and the boundary terms with the stabilization parameters γ_1 and γ_2. Theorem 4.8 proved that two boundary terms can be controlled by $\eta_P(\Omega)$. Hence, our study uses the *a posterior* error estimator (4.22) without the boundary terms. In Figure 4.1, we investigate Theorem 4.8 by studying the influence of the boundary terms on the rates of the convergence in all three error norms. The convergence plots with the boundary terms are almost identical to those without the boundary terms. Our numerical study shows that the boundary terms can be controlled by $C\|\eta_P\|^2$ with the choice of $C = \max(\gamma_1, \gamma_2)$.

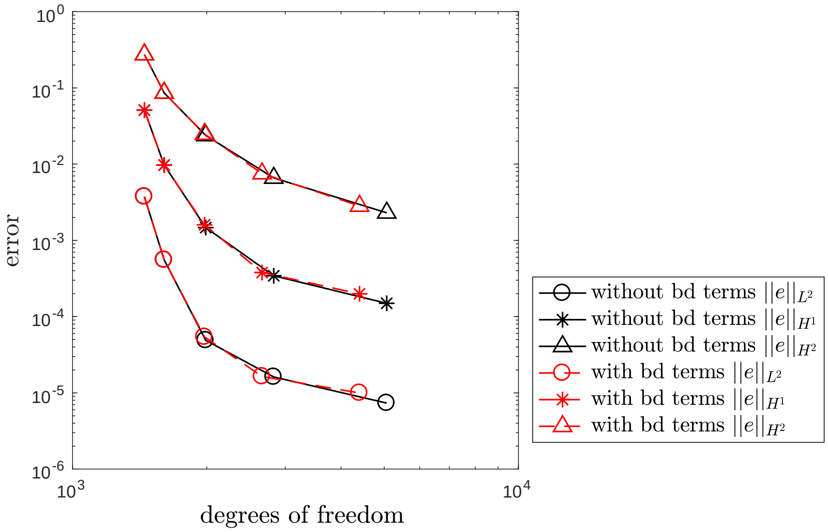


Figure 4.1: *Numerical investigation of Theorem 4.8: The convergence rates with and without boundary terms.*

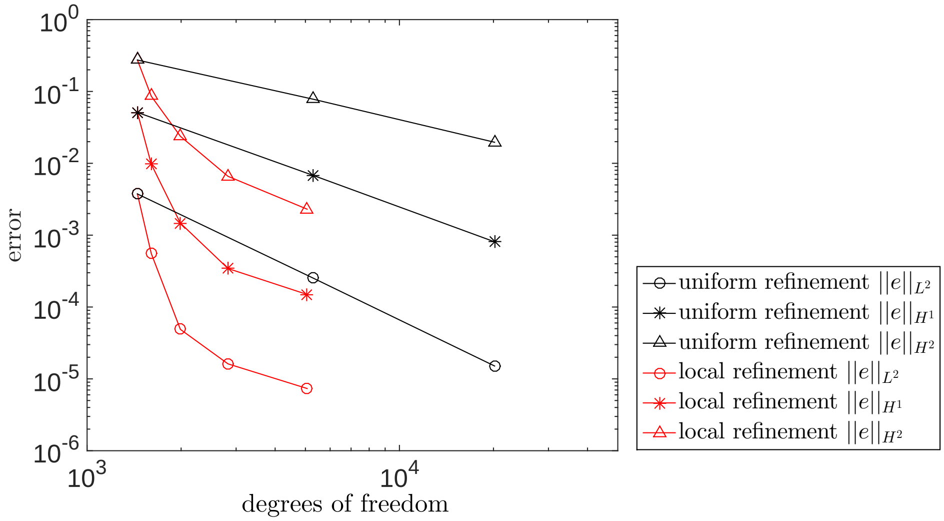
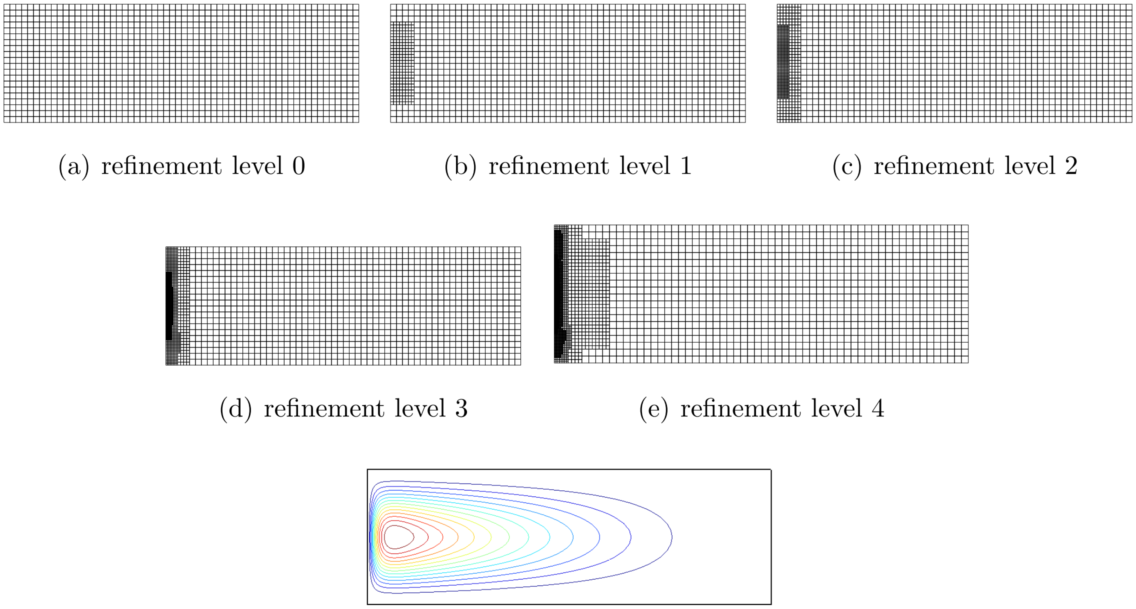


Figure 4.2: *Convergence rates in the L^2-norm, the H^1-norm, and the H^2-norm for the rectangular geometry.*

4.4.2 L-shape geometry

In this section, we further study the capability of the proposed error indicator (4.22) for the test problem on a L-shape geometry which is more suitable for the test of adaptive refinement. We use the forcing term $f = \sin(y)$ derived from the derivative of the wind stress taken from Myers and Weaver [77]. For this test problem, an analytical solution is not available. Therefore, we compute the solution on a sufficiently fine grid with 2,250,000 elements and use it as a standard solution for convergence study. The fine grid has a sufficiently large number of elements that the *a posteriori* error estimator decays with asymptotic regime rate.

Analogous to the previous example, we display the convergence rates for the adaptive refinement along with those for the uniform refinement in Figure 4.4. With the uniform refinement, our numerical study shows the significant reduction of convergence rates, i.e., 0.7, 0.64 and 0.30 in the L^2-norm, the H^1-norm, and the H^2-norm, respectively, with respect to the number of cells in a manner that is consistent with (2.147). We attribute this to the presence of the reentrant corner in the L-shape geometry. Interestingly, the use of the adaptive refinement increases significantly the rates of convergence to 2.89, 2.75 and 1.30 in the L^2-norm, the H^1-norm, and the H^2-norm, respectively, with respect to the number of cells in a manner that is consistent with (2.148). Moreover, the solution from the adaptive refinement is much more accurate at lower resolution than that from the uniform refinement,



(a) refinement level 0

(b) refinement level 1

(c) refinement level 2

(d) refinement level 3

(e) refinement level 4

(f) solution of refinement level 4

Figure 4.3: *Local refinement mesh of rectangular geometry example ($\theta = 0.9$)*

indicating the efficiency of our algorithm. Figure 4.5 shows local refinement meshes and the solution. Importantly, the mesh is refined at the reentrant corner as well as the boundary layers.

Using the L-shape geometry problem, we next examine the influence of the parameter θ of Dörfler marker's strategy on the accuracy of the solution. Figure 4.6 shows the L^2-norm of the error versus the parameter θ. Plots are provided for three cases of $\theta = 0.1, 0.5,$ and 0.9 along with the convergence plot of the uniform refinement. The results show that while the smaller θ leads to slightly better convergence rates, the larger θ results in much smaller value of the L^2-norm error at the same refinement levels. Based on this study, notice that all of our numerical studies are performed using $\theta = 0.9$.

4.4.3 Computational efficiency of the adaptive algorithm

In this section, we examine the efficiency of the proposed adaptive algorithm. This is achieved by comparing the computational time of our local refinement algorithm with that of the uniform refinement. All calculations for this example are carried out on a workstation with

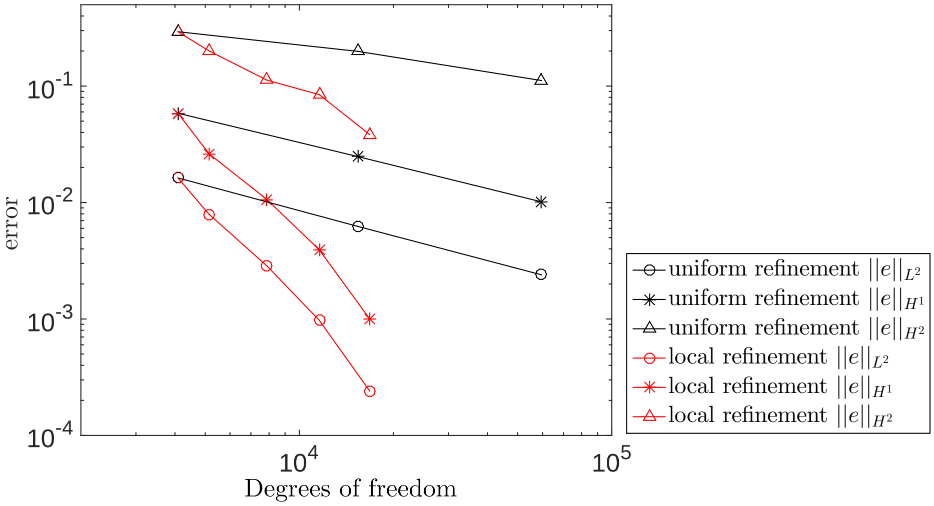
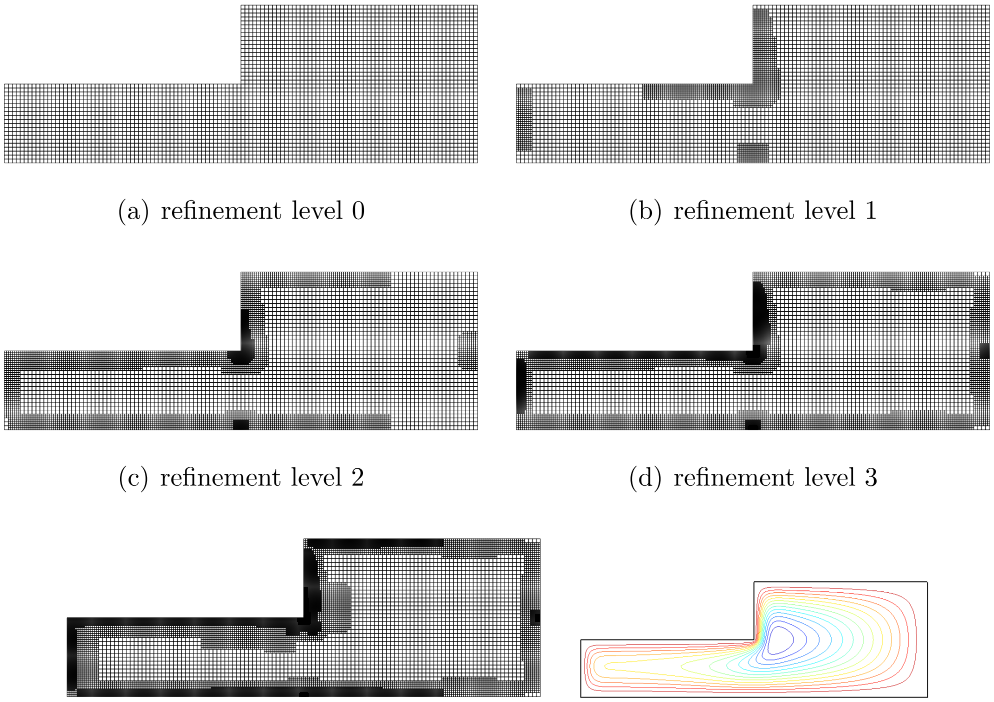


Figure 4.4: *Convergence rates in the L^2-norm, the H^1-norm, and the H^2-norm for L-shape geometry.*

Xeon E5 v3 2637 3.5 GHz CPU and 64GB of memory. A direct LU solver is employed to obtain the solutions of the linear algebraic system of equations. In Figures 4.7(a) and 4.7(b), we provide the CPU time versus all three norms of the error for the rectangular geometry and the L-shape geometry, respectively. Importantly, the computational time is greatly reduced by locally refining the mesh. This study shows that our proposed adaptive mesh algorithm gives rises to accurate results and significant computational savings compared to uniform mesh approaches.



(a) refinement level 0

(b) refinement level 1

(c) refinement level 2

(d) refinement level 3

(e) refinement level 4

(f) solution of refinement level 4

Figure 4.5: *Local refinement mesh of L-shape geometry example* ($\theta = 0.9$)

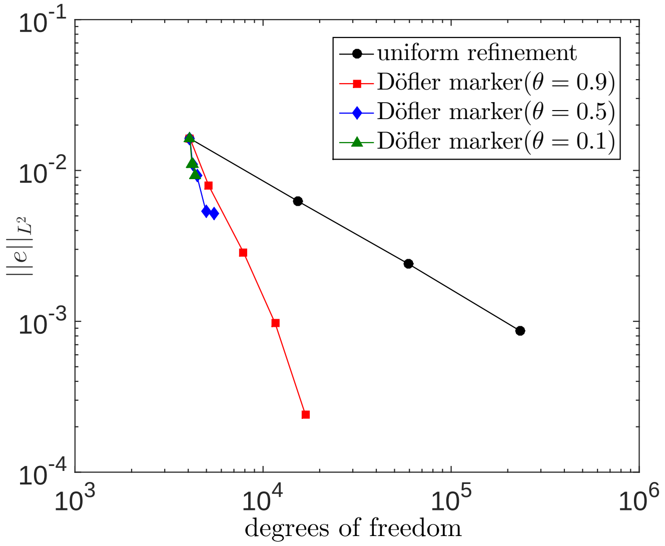
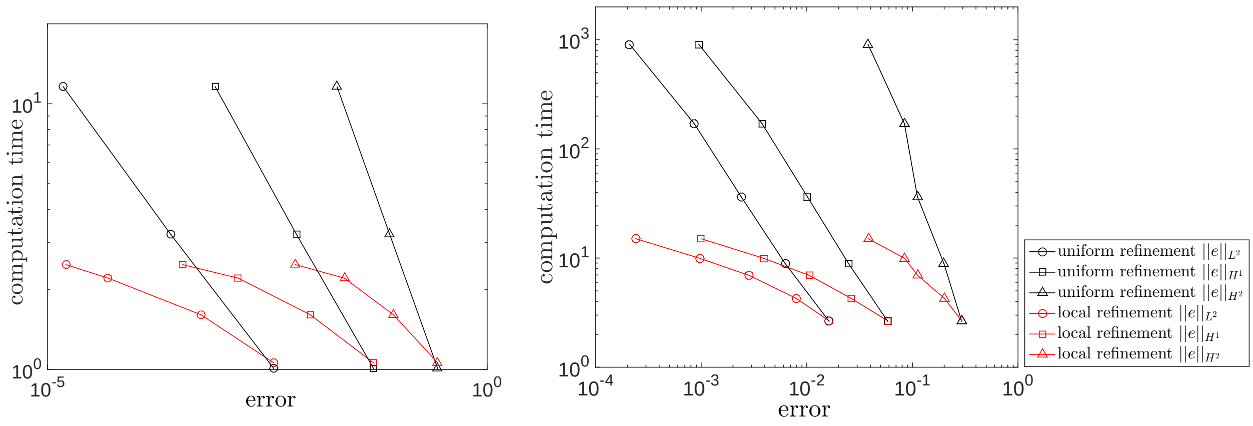


Figure 4.6: *error in L_2 norm of L-shape geometry example with $\theta = 0.1, 0.5$ and 0.9*



(a) rectangular geometry example

(b) *L*-shape geometry example

Figure 4.7: *Computation time versus error in the L^2-norm, the H^1-norm, and the H^2-norm ($\theta = 0.9$)*

Chapter 5

A posterior analysis of a B-spline based finite-element method for the stationary quasi-geostrophic equations of the ocean

5.1 Introduction

This study focuses on an adaptive mesh-refinement algorithm using B-splines for the SQGE. In the linear Stommel–Munk model in Chapter 4, a saturation assumption that enables relating the numerical error $u - U$ with a discrete error $U_* - U$ was employed to circumvent the limitations posed by the definition of the bilinear form. More precisely, the coercivity result in Chapter 4 does not hold for general $u \in H^2$ because the definition of a is limited to the discrete space $\mathbb{X}_P$. Any error estimation, whether it be *a priori* or *a posteriori* cannot be related with the numerical error (through coercivity) unless solution u is sufficiently smooth, which can't be expected near boundary layers or problematic corners in the domain. One can prove the saturation assumption by deriving a local discrete lower bound, but proving such an estimate is very difficult. To the best of our knowledge, no such estimate exists for a fourth-order problem or when Nitsche's method is employed even for the Poisson problem. However, it is shown in various publications that the discrete lower bound is not essential for neither convergence nor quasi-optimality analyses of adaptive finite-element methods.

To achieve this goal, we introduce the weak formulation of the SQGE for B-splines using a standard L^2-orthogonal projection operator. We show the dominance of *a posteriori* error estimator over the numerical error without a saturation assumption, but at the cost of consistency of the weak formulation with the strong form of the SQGE. The resulting inconsistency is however shown to be weaken with refinement. The idea of extending the definition of the bilinear form is inspired by the treatment of adaptive discontinuous finite-element methods (ADFEM) in Bonito et al. [63] highlighting the similarity in nature of both methods, theoretically as well as numerically.

The remainder of the Chapter is organized as follows. In Section 5.2, we present the SQGE along with its weak formulation for B-splines. In Section 5.3, *a priori* estimate and coercivity analysis for the weak formulation are provided. In Section 5.4, *a posteriori* error analysis is performed to derive *a posteriori* error indicator. Numerical studies with benchmark problems on rectangular and L-shape domains are provided to show the performance of the analysis in Section 5.4.

5.2 The stationary quasi-geostrophic equations

We consider a domain $\Omega \in \mathbb{R}^2$ with a polygonal boundary Γ. For a given velocity streamfunction u and a wind forcing f, the streamfunction formulation [37] of the one-layer SQGE along with boundary conditions is given by

$$\begin{cases} \mathrm{Re}^{-1}\Delta^2 u + J(u, \Delta u) - \mathrm{Ro}^{-1}\frac{\partial u}{\partial x} = \mathrm{Ro}^{-1} f & \text{in} \quad \Omega, \\ u = 0, \quad \frac{\partial u}{\partial \boldsymbol{n}} = 0 & \text{on} \quad \Gamma. \end{cases} \tag{5.1}$$

Notice that above boundary conditions indicate no normal transport and no-slip on the boundary. Here, $J(\cdot,\cdot)$ is the Jacobian operator given by

$$J(u, v) = \nabla^{\perp} u \cdot \nabla v \quad \text{with} \quad \nabla^{\perp} u = \left(\frac{\partial u}{\partial y}, -\frac{\partial u}{\partial x}\right), \tag{5.2}$$

and Re and Ro are the Reynolds and Rossby numbers defined by (1.3).

5.2.1 Weak formulation

We recall the auxiliary subdomain domain D_P^{Γ}, initially defined in (2.89), and define its support extension ω_P^{Γ}:

$$D_P^{\Gamma} = \overline{\bigcup \{\tau \in P : \tau \text{ adjacent to } \Gamma\}} \quad \text{and} \quad \omega_P^{\Gamma} = \overline{\bigcup_{\sigma \in \mathcal{G}_P} \omega_{\sigma}}. \tag{5.3}$$

For later use, we define the mesh-dependent semi-norm as

$$|\!|\!| u |\!|\!|_P^2 = \mathrm{Re}^{-1} \left\| \Delta u \right\|_{L^2(\Omega)}^2 + \gamma_1 \left\| u \right\|_{3/2,P}^2 + \gamma_2 \left\| \frac{\partial u}{\partial \boldsymbol{n}} \right\|_{1/2,P}^2 . \tag{5.4}$$

It will be value to abbreviate the boundary norms:

$$|u|_P^2 = \|u\|_{3/2,P}^2 + \left\|\tfrac{\partial u}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2. \tag{5.5}$$

The standard conforming weak formulation of the SQGE (5.1) can be stated as:

$$u \in H_0^2(\Omega) \quad \text{such that} \quad \langle \mathscr{F}(u), v\rangle = \ell_f(v), \quad \forall v \in H_0^2(\Omega), \tag{5.6}$$

where $\mathscr{F} : H_0^2(\Omega) \times H_0^2(\Omega) \to \mathbb{R}$ is the nonlinear form

$$\langle \mathscr{F}(u), v\rangle = \mathrm{Re}^{-1} \int_\Omega \Delta u \Delta v + \int_\Omega \Delta u J(u,v) - \mathrm{Ro}^{-1} \int_\Omega \tfrac{\partial u}{\partial x} v, \tag{5.7}$$

and $\ell_f(v) = \mathrm{Ro}^{-1} \int_\Omega fv$. Notice that (5.6) can be valid for any interpolatory C^1-continuous basis functions such as conventional Lagrangian or Hermitian polynomial basis functions. For non-interpolatory basis functions such as B-splines, the Nitsche-type variational formulation was provided in equation (5) of Kim et al. [36]. Dirichlet boundary conditions are weakly imposed along with stabilization. Moreover, the formulation includes additional boundary integral terms on the left-hand side of (5.8) to prove consistency with the SQGE (5.1). Such additional terms make *a posteriori* analysis difficult using the consistent variational formulation. In the following, we discuss this difficulty and motivation of our analysis strategy allowing us to circumvent the employment of the saturation assumption used in Chapter 4 for the Stommel–Munk model.

Recall that the Green identity,

$$\int_\Gamma \tfrac{\partial \Delta u}{\partial \boldsymbol{n}} v - \int_\Gamma \Delta u \tfrac{\partial v}{\partial \boldsymbol{n}} = \int_\Omega \Delta^2 u v - \int_\Omega \Delta u \Delta v \quad \forall v \in H^2(\Omega), \tag{5.8}$$

which is valid for $u \in H^4(\Omega)$. The integrals $\int_\Gamma \Delta u \frac{\partial v}{\partial \boldsymbol{n}}$ and $\int_\Gamma \frac{\partial \Delta u}{\partial \boldsymbol{n}} v$ cannot be understood in the usual way because (5.8) is not valid if u is only a member of $H_0^2(\Omega)$ without additional higher-order derivative integrability. This is because the traces of $\frac{\partial \Delta u}{\partial \boldsymbol{n}}$ and Δu are no longer well-defined on Γ. As a result, the general approach of using (5.8) to obtain a nonlinear weak form defined on $H^2(\Omega) \times H^2(\Omega)$ that is coercive with a completely justifiable *a posteriori* derivation breaks down. To resolve this problem, the terms $\frac{\partial \Delta u}{\partial \boldsymbol{n}}$ and Δu are replaced with suitable $r-2$ degree polynomial approximations on each cell τ using the L^2-orthogonal projection $\Pi_P = \Pi_P^{r-2}$ of Lemma 2.43 to make the boundary integrals in (5.8) valid for $u \in \mathbb{V}_P$. Moreover, we consider the test and trial spaces as

$$\mathbb{V}_P = H_0^2(\Omega) + \mathbb{X}_P, \tag{5.9}$$

where $\mathbb{X}_P$ is the spline space defined in (4.3). Then, any function $u \in \mathbb{V}_P$ can be decomposed into an H_0^2-conforming part u_0 and a spline part U, i.e., $u = u_0 + U$. We propose the weak formulation for the SQGE (5.1):

$$\text{Find } U \in \mathbb{X}_P \text{ such that} \quad \langle \mathscr{F}_P(U), V\rangle = \ell_f(V) \quad \text{for all } V \in \mathbb{X}_P, \tag{5.10}$$

with $\mathscr{F}_P : \mathbb{V}_P \times \mathbb{V}_P \to \mathbb{R}$ defined by

$$\begin{aligned} \langle \mathscr{F}_P(u), v\rangle = \langle \mathscr{F}(u), v\rangle &+ \mathrm{Re}^{-1} \int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} v + u \tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}} \right) \\ &- \mathrm{Re}^{-1} \int_\Gamma \left(\Pi_P(\Delta u) \tfrac{\partial v}{\partial \boldsymbol{n}} + \tfrac{\partial u}{\partial \boldsymbol{n}} \Pi_P(\Delta v) \right) \\ &+ \gamma_1 \int_\Gamma h_P^{-3} u v + \gamma_2 \int_\Gamma h_P^{-1} \tfrac{\partial u}{\partial \boldsymbol{n}} \tfrac{\partial v}{\partial \boldsymbol{n}}, \end{aligned} \tag{5.12}$$

where γ_1 and γ_2 are positive stabilization parameters.

Remark 5.1. Another issue comes from the identity

$$\int_\Omega J(u, \Delta u) v = \int_\Omega \Delta u J(u, v) - \int_\Gamma \Delta u (\nabla^\perp u \cdot \boldsymbol{n}) v, \quad (u, v) \in H^{5/2}(\Omega) \times H^{3/2}(\Omega), \tag{5.13}$$

which requires the presence of a nonlinear boundary integral $\int_\Gamma \Delta u (\nabla^\perp u \cdot \boldsymbol{n}) v$ in our form. This was the case in the last two terms of (6) in Kim et al. [36]. In this work, we forgo this term with the reasoning that any solution $u \in H_0^2(\Omega)$ will maintain that $\nabla^\perp u \cdot \boldsymbol{n}$ is zero along Γ and therefore an approximate solution $U \in \mathbb{X}_P$ is expected to give a small values for $\nabla^\perp U \cdot \boldsymbol{n}$ along Γ. Dropping the nonlinear boundary term does not interfere with the forms consistency with the PDE; see Lemma 5.5 below. In a similar vain, the lower order boundary integral terms coming from the Laplace term (see (4.7)) may also be dropped at no cost to consistency.

Remark 5.2. The function space $\mathbb{V}_P$ is a proper subset of $H^2(\Omega)$ for which (5.4) defines a full norm equivalent to $\|\cdot\|_{H^2(\Omega)}$, i.e.,

$$\|u\|_{H^2(\Omega)} \leq c_E \|u\|_P, \quad \forall u \in \mathbb{V}_P, \tag{5.14}$$

with a constant $c_E > 0$ depending on the stabilization parameters and the mesh shape-regularity. It is immediate that $\|\Delta \cdot\|_{L^2(\Omega)}$ defines a full norm on $H_0^2(\Omega)$. As for the discrete portion $\mathbb{X}_P$ which does not belong to $H_0^2(\Omega)$, the presence of the boundary integrals in (5.4) is enough to ensure the positive-definiteness of the semi-norm (5.4) owing to Lemma 4.7. For completeness, we provide a proof of (5.14)

Proof. Let $u \in \mathbb{V}_P$. In view of decomposition (5.21), let $u = (u_0 + U^0) + U^\perp$ and $u^0 = u_0 + U^0$. By applying (4.70) and Poincaré's inequality with a constant C_P, we obtain

$$\begin{aligned} \|u\|_{H^2(\Omega)} &\leq C_P \|\Delta u^0\|_{L^2(\Omega)} + C_{\mathrm{bdry}}^{1/2} |u|_P \\ &\leq C_P \left(\|\Delta u\|_{L^2(\Omega)} + |U^\perp|_{H^2(\Omega)} \right) + C_{\mathrm{bdry}}^{1/2} |u|_P \\ &\preceq \|\Delta u\|_{L^2(\Omega)} + \max_i \{\gamma_i^{-1}\} \left(\gamma_1 \|u\|_{3/2,P}^2 + \gamma_2 \left\| \tfrac{\partial u}{\partial \boldsymbol{n}} \right\|_{1/2,P}^2 \right)^{1/2}. \end{aligned} \tag{5.16}$$

Notice that the proportionality constant c_E in (5.14) depends on C_P, C_{bdry}, γ_1 and γ_2. ■

Remark 5.3. The weak form (5.12) can be decomposed into the linear part $L_P(u,v)$ and the nonlinear part $N(u,u,v) = \int_\Omega \Delta u J(u,v)$, i.e.,

$$\langle \mathscr{F}_P(u), v \rangle = L_P(u,v) + N(u,u,v). \tag{5.17}$$

Notice that $\langle \mathscr{F}_P(u), u \rangle$ reduces to $L_P(u,u)$ for any $u \in \mathbb{V}_P$ because $J(u,u) = 0$. The nonlinear term $N(u,v,\phi) = \int_\Omega \Delta u J(v,\phi)$ is trilinear in each entry. Moreover, in view of the continuous embedding $H^1(\Omega) \hookrightarrow L^4(\Omega)$, N admits the continuity-type estimate

$$|N(u,v,\phi)| \leq C_{\mathcal{N}} \|\Delta u\|_{L^2(\Omega)} \|v\|_{H^2(\Omega)} \|\phi\|_{H^2(\Omega)}. \tag{5.18}$$

Remark 5.4. The L^2 forcing term in dual-norm can be estimated by

$$\|\ell_f\|_{\mathbb{V}_P'} = \sup_{v \in \mathbb{V}_P} \frac{\mathrm{Ro}^{-1} \int_\Omega f v}{\|\!|v|\!\|_P} \leq c_E \mathrm{Ro}^{-1} \|f\|_{L^2(\Omega)}, \tag{5.19}$$

where $c_E > 0$ is the norm equivalence constant between $\|\cdot\|_{H^2(\Omega)}$ and $|\!|\!|\cdot|\!|\!|_P$.

5.2.2 Formulation inconsistency

The solution u to the conforming weak formulation (5.6) does not satisfy the discrete formulation (5.10) when tested against members of $\mathbb{V}_P \backslash H_0^2(\Omega)$. As a consequence, (5.10) has only partial consistency, i.e.,

$$\langle \mathscr{F}_P(u), v_0 \rangle = \langle \mathscr{F}(u), v_0 \rangle = \ell_f(v_0) \quad \forall v_0 \in H_0^2(\Omega). \tag{5.20}$$

To quantify inconsistency, we extract the part V^0 of $V \in \mathbb{X}_P$ comprising of the spline basis functions belonging to $H_0^2(\Omega)$ and the part $V^\perp = V - V^0$ that does not satisfy Dirichlet boundary conditions on Γ as described in Lemma 4.7. In other words, we decompose $\mathbb{X}_P$ into the conforming part $\mathbb{X}_P^0 = \mathbb{X}_P \cap H_0^2(\Omega)$ and the nonconforming part $\mathbb{X}_P^\perp$, i.e.,

$$\mathbb{X}_P = \mathbb{X}_P^0 + \mathbb{X}_P^\perp. \tag{5.21}$$

Notice that $U^\perp$ is identical to U along boundary edges σ since by definition $U^\perp = U - U^0$ and $U^0 \in H_0^2(\Omega)$. Above inequality is used for *a posteriori* analysis in Lemma 5.15.

Upon using (5.21), we can obtain the following inequality whose proof is given below. If $u \in H_0^2(\Omega)$ is the solution to (5.6) and we let $v = v_0 + V \in \mathbb{V}_P$,

$$|\langle \mathscr{F}_P(u), v\rangle - \ell_f(v)| \preceq |u|_{H^2(\omega_P^\Gamma)} |V|_P. \tag{5.22}$$

Notice that the right-hand-side decays as $\omega_P^\Gamma \to 0$ because of the vanishing integration domain in $|u|_{H^2(\omega_P^\Gamma)}$. Moreover, if $v \in H_0^2(\Omega)$, the right-hand-side vanishes indicating partial consistency.

Proof. If $v \in \mathbb{V}_P$, $v = v^0 + V^\perp$ with $v^0 = v_0 + V^0 \in \mathbb{V}_P \cap H_0^2(\Omega)$ in view of (5.9) and (5.21). By partial consistency, we have

$$\langle \mathscr{F}_P(u), v\rangle = \langle \mathscr{F}_P(u), v^0\rangle + \langle \mathscr{F}_P(u), V^\perp\rangle = \ell_f(v^0) + \langle \mathscr{F}_P(u), V^\perp\rangle. \tag{5.23}$$

Then,

$$\begin{aligned} \langle \mathscr{F}_P(u), v\rangle - \ell_f(v) &= \langle \mathscr{F}_P(u), v^0\rangle - \ell_f(v^0) + \langle \mathscr{F}_P(U), V^\perp\rangle - \ell_f(V^\perp), \\ &= \langle \mathscr{F}_P(u), V^\perp\rangle - \ell_f(V^\perp). \end{aligned} \tag{5.25}$$

Since $V^\perp$ is supported only on ω_P^Γ and $U = U^\perp$ along Γ, we write

$$\begin{aligned} &\langle \mathscr{F}_P(u), V^\perp\rangle - \ell_f(V^\perp) \\ &\qquad = \int_{\omega_P^\Gamma} \left(\mathrm{Re}^{-1}\Delta u \Delta V^\perp + \Delta u J(u, V^\perp) - \mathrm{Ro}^{-1}\tfrac{\partial u}{\partial x} V^\perp - \mathrm{Ro}^{-1} f V^\perp\right) \\ &\qquad - \mathrm{Re}^{-1}\int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} V^\perp - \Pi_P(\Delta u)\tfrac{\partial V^\perp}{\partial \boldsymbol{n}}\right). \end{aligned} \tag{5.27}$$

The difference $\langle \mathscr{F}_P(u), V^\perp\rangle - \ell_f(V^\perp)$ is the source of inconsistency, and we show that it decays when the near-boundary cells are refined. Each term of above is estimated as follows.

$$\left|\int_{\omega_P^\Gamma} \Delta u \Delta V^\perp\right| \le |\Delta u|_{H^2(\omega_P^\Gamma)} |V^\perp|_{H^2(\omega_P^\Gamma)}, \tag{5.28}$$

and

$$\left|\int_{\omega_P^\Gamma} \tfrac{\partial u}{\partial x} V^\perp\right| \le |u|_{H^1(\omega_P^\Gamma)} \|V^\perp\|_{L^2(\omega_P^\Gamma)} \le C_P |u|_{H^2(\omega_P^\Gamma)} \|V^\perp\|_{L^2(\omega_P^\Gamma)}, \tag{5.29}$$

where the last inequality holds by Poincaré's inequality and C_P is taken to be the Poincaré constant. Moreover, in view of (5.18),

$$\left|\int_{\omega_P^\Gamma} \Delta u J(u, V^\perp)\right| \le C_{\mathcal{N}} |u|_{H^2(\omega_P^\Gamma)} \|u\|_{H^2(\omega_P^\Gamma)} \|V^\perp\|_{H^2(\omega_P^\Gamma)}, \tag{5.30}$$

and

$$\left| \int_{\omega_P^\Gamma} f V^\perp \right| \leq \|f\|_{L^2(\omega_P^\Gamma)} \|V^\perp\|_{L^2(\omega_P^\Gamma)}. \tag{5.31}$$

In view of the norm equivalence (5.14), combining above estimates results in

$$\begin{aligned} &\left| \int_{\omega_P^\Gamma} \left(\mathrm{Re}^{-1} \Delta u \Delta V^\perp + \Delta u J(u, V^\perp) - \mathrm{Ro}^{-1} \tfrac{\partial u}{\partial x} V^\perp - \mathrm{Ro}^{-1} f V^\perp \right) \right| \\ &\qquad \leq c_E |u|_{H^2(\omega_P^\Gamma)} \left(\mathrm{Re}^{-1} + C_P + C_{\mathcal{N}} |u|_{H^2(\omega_P^\Gamma)} + \mathrm{Ro}^{-1} \|f\|_{L^2(\omega_P^\Gamma)} \right) |V|_P. \end{aligned} \tag{5.33}$$

It is left to treat the boundary integrals. To avoid including too many calculations at this point, we direct the reader to the proof of Lemma 5.15 and Remark 2.44 for a similar treatment that leads us to the estimate:

$$\left| \mathrm{Re}^{-1} \int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} V^\perp - \Pi_P(\Delta u) \tfrac{\partial V^\perp}{\partial \boldsymbol{n}} \right) \right| \leq c_\Pi \mathrm{Re}^{-1} |u|_{H^2(\omega_P^\Gamma)} |V|_P. \tag{5.34}$$

Incorporating the last two inequalities completes the proof. ∎

We have so far discussed inconsistency of (5.10) when the weak solution does not have any weak derivatives beyond the second order. We show that inconsistency is not strong and will decay with refinement if u solves (5.1) point-wise almost-everywhere and $\Delta u \in H^s(\Omega)$ for $s \geq 0$. In doing so, we define the inconsistency term $\mathscr{E}_P \in \mathbb{V}_P'$ given by

$$\langle \mathscr{E}_P, v \rangle = \int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} - \tfrac{\partial \Delta u}{\partial \boldsymbol{n}} \right) v - \int_\Gamma \left(\Pi_P(\Delta u) - \Delta u \right) \tfrac{\partial v}{\partial \boldsymbol{n}}, \quad v \in \mathbb{V}_P. \tag{5.35}$$

In the following lemma, above inconsistency term is derived from (5.12).

Lemma 5.5 (Inconsistency). *If $u \in H_0^2(\Omega)$ with $\Delta^2 u \in L^2(\Omega)$ is the solution to (5.1), then it holds that*

$$\langle \mathscr{F}_P(u), v \rangle = \ell_f(v) + \langle \mathscr{E}_P, v \rangle, \quad \forall v \in \mathbb{V}_P. \tag{5.36}$$

Proof. By applying integration by parts, we have

$$\begin{aligned} \langle \mathscr{F}_P(u), v \rangle - \ell_f(v) = &\int_\Omega \left(\mathrm{Re}^{-1} \Delta^2 u + J(u, \Delta u) - \mathrm{Ro}^{-1} \tfrac{\partial u}{\partial x} - \mathrm{Ro}^{-1} f \right) v \\ &- \mathrm{Re}^{-1} \int_\Gamma \tfrac{\partial \Delta u}{\partial \boldsymbol{n}} v + \mathrm{Re}^{-1} \int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} v + u \tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}} \right) \\ &+ \mathrm{Re}^{-1} \int_\Gamma \Delta u \tfrac{\partial v}{\partial \boldsymbol{n}} - \mathrm{Re}^{-1} \int_\Gamma \left(\Pi_P(\Delta u) \tfrac{\partial v}{\partial \boldsymbol{n}} + \tfrac{\partial u}{\partial \boldsymbol{n}} \Pi_P(\Delta v) \right) \\ &+ \gamma_1 \int_\Gamma h_P^{-3} u v + \gamma_2 \int_\Gamma h_P^{-1} \tfrac{\partial u}{\partial \boldsymbol{n}} \tfrac{\partial v}{\partial \boldsymbol{n}}. \end{aligned} \tag{5.38}$$

Invoking the assumptions on u, all integrals over Ω and those boundary integrals with $u|_\Gamma$ and $\frac{\partial u}{\partial \boldsymbol{n}}|_\Gamma$ vanish. Thus, we have

$$\langle \mathscr{F}_P(u), v\rangle - \ell_f(v) = \mathrm{Re}^{-1}\int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} - \tfrac{\partial \Delta u}{\partial \boldsymbol{n}}\right) v - \mathrm{Re}^{-1}\int_\Gamma \left(\Pi_P(\Delta u) - \Delta u\right)\tfrac{\partial v}{\partial \boldsymbol{n}}. \tag{5.39}$$

■

The consistency $\mathscr{E}_P$ is not strong and it comes from projection Π_P needed to make the nonlinear form (5.12) well-defined on $\mathbb{V}_P$. In fact, as h_P decreases with refinement, the area of D_P^Γ will decrease with it, weakening the inconsistency. In Lemma 5.6, we show that the inconsistency term (5.35) decays with an order of at least h_P^s for $\Delta u \in H^s(\Omega)$.

Lemma 5.6 (Asymptotic consistency). *If u is a solution to* (5.1) *and $\Delta u \in H^s(\Omega)$, $s \geq 0$, then*

$$\langle \mathscr{E}_P, v\rangle \leq c_\Pi \mathrm{Re}^{-1}\, |h_P^s \Delta u|_{H^s(D_P^\Gamma)} |v|_P \quad \textit{for all } v \in \mathbb{V}_P, \tag{5.40}$$

equivalently,

$$\|\mathscr{E}_P\|_{\mathbb{V}_P'} \leq c_\Pi \mathrm{Re}^{-1}\, |h_P^s \Delta u|_{H^s(D_P^\Gamma)}\,. \tag{5.41}$$

Proof. Let $v \in \mathbb{V}_P$. Then,

$$\begin{aligned}\langle \mathscr{E}_P, v\rangle &= \mathrm{Re}^{-1}\int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} - \tfrac{\partial \Delta u}{\partial \boldsymbol{n}}\right) v - \mathrm{Re}^{-1}\int_\Gamma \left(\Pi_P(\Delta u) - \Delta u\right)\tfrac{\partial v}{\partial \boldsymbol{n}}\\ &\leq \mathrm{Re}^{-1}\sum_{\sigma\in\mathcal{G}_P}\left\|\tfrac{\partial(\Pi_P(\Delta u)-\Delta u)}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}\|v\|_{L^2(\sigma)}\\ &\quad + \mathrm{Re}^{-1}\sum_{\sigma\in\mathcal{G}_P}\|\Pi_P(\Delta u) - \Delta u\|_{L^2(\sigma)}\left\|\tfrac{\partial v}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}.\end{aligned} \tag{5.43}$$

Applying the Cauchy-Schwarz inequality into above yields

$$\begin{aligned}\langle \mathscr{E}_P, v\rangle &\leq \mathrm{Re}^{-1}\Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma^3\left\|\tfrac{\partial(\Pi_P(\Delta u)-\Delta u)}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}^2\Bigg)^{1/2}\|v\|_{3/2,P}\\ &\quad + \mathrm{Re}^{-1}\Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma\|\Pi_P(\Delta u) - \Delta u\|_{L^2(\sigma)}^2\Bigg)^{1/2}\left\|\tfrac{\partial v}{\partial \boldsymbol{n}}\right\|_{1/2,P}.\end{aligned} \tag{5.45}$$

In view of Lemma 2.43 we have

$$\langle \mathscr{E}_P, v\rangle \preceq c_\Pi \mathrm{Re}^{-1}\, |h_P^s \Delta u|_{L^2(D_P^\Gamma)}\, |v|_P. \tag{5.47}$$

■

5.3 Coercivity and *a priori* analysis

In this section, we prove the well-posedness of the weak formulation (5.10).

Lemma 5.7 (Coercivity). *For a constant $C_{\mathrm{Coer}} > 0$,*

$$C_{\mathrm{Coer}} \vert\!\vert\!\vert u \vert\!\vert\!\vert_P^2 \leq \langle \mathscr{F}_P(u), u\rangle, \quad \forall u \in \mathbb{V}_P \tag{5.48}$$

where $\vert\!\vert\!\vert u \vert\!\vert\!\vert_P$ is defined in (5.4).

Proof. Let $u \in \mathbb{V}_P$ and recall that $J(u,u) = 0$. From (5.12), we have

$$\begin{aligned}\langle \mathscr{F}_P(u), u\rangle &= \mathrm{Re}^{-1}\|\Delta u\|_{L^2(\Omega)}^2 - \mathrm{Ro}^{-1}\int_\Omega \tfrac{\partial u}{\partial x} u \\ &\quad + 2\mathrm{Re}^{-1}\int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} u - \Pi_P(\Delta u)\tfrac{\partial u}{\partial \boldsymbol{n}}\right) + \gamma_1\|u\|_{3/2,P}^2 + \gamma_2\left\|\tfrac{\partial u}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2 .\end{aligned} \tag{5.50}$$

The first integral on the boundary Γ in the second line of (5.50) is estimated as

$$\begin{aligned}\mathrm{Re}^{-1}\int_\Gamma \tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} u &\leq \mathrm{Re}^{-1}\left(\int_\Gamma \left(h_P^{3/2}\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}}\right)^2\right)^{1/2} \|u\|_{3/2,P} \\ &\leq c_{\mathrm{Inv}} c_{\mathrm{dTr}} c_\Pi \mathrm{Re}^{-1}\|\Delta u\|_{L^2(\Omega)}\|u\|_{3/2,P} \\ &\leq \left(c_{\mathrm{Inv}} c_{\mathrm{dTr}} c_\Pi \mathrm{Re}^{-1}\right)^2 \tfrac{\delta_1}{2}\|\Delta u\|_{L^2(\Omega)}^2 + \tfrac{1}{2\delta_1}\|u\|_{3/2,P}^2,\end{aligned} \tag{5.52}$$

using

$$\begin{aligned}\int_\Gamma \left(h_P^{3/2}\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}}\right)^2 &= \sum_{\sigma \in \mathcal{G}_P} h_\sigma^3 \left\|\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}^2, \\ &\leq (c_{\mathrm{Inv}} c_{\mathrm{dTr}})^2 \sum_{\sigma \in \mathcal{G}_P} \|\Pi_P(\Delta u)\|_{L^2(\tau_\sigma)}^2, \\ &\leq (c_{\mathrm{Inv}} c_{\mathrm{dTr}})^2 \|\Pi_P(\Delta u)\|_{L^2(\Omega)}^2 \leq (c_{\mathrm{Inv}} c_{\mathrm{dTr}} c_\Pi)^2 \|\Delta u\|_{L^2(\Omega)}^2,\end{aligned} \tag{5.54}$$

and Young's inequality with $\delta_1 > 0$. The above inequality is obtained using

$$\left\|\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)} \leq c_{\mathrm{Inv}} c_{\mathrm{dTr}} h_\sigma^{-3/2}\|\Pi_P(\Delta u)\|_{L^2(\tau_\sigma)}, \tag{5.55}$$

from Lemma 2.36 and the stability of the L^2-projection in Remark 2.44 for the last inequality. Similarly, the second integral along the boundary Γ in (5.50) can be estimated as

$$\mathrm{Re}^{-1}\int_\Gamma \Pi_P(\Delta u)\tfrac{\partial u}{\partial \boldsymbol{n}} \leq \left(c_{\mathrm{dTr}} c_\Pi \mathrm{Re}^{-1}\right)^2 \tfrac{\delta_2}{2}\|\Delta u\|_{L^2(\Omega)}^2 + \tfrac{1}{2\delta_2}\left\|\tfrac{\partial u}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2, \tag{5.56}$$

for $\delta_2 > 0$. Substituting (5.52) and (5.56) into (5.50) results in

$$\begin{aligned}\langle \mathscr{F}_P(u), u\rangle \geq \mathrm{Re}^{-1}\|\Delta u\|^2_{L^2(\Omega)} - \tfrac{\mathrm{Ro}^{-1}}{2}\|u\|^2_{L^2(\Gamma)} - \left(c_{\mathrm{Inv}}c_{\mathrm{dTr}}c_{\Pi}\mathrm{Re}^{-1}\right)^2\delta_1\|\Delta u\|^2_{L^2(\Omega)} \\ - \tfrac{1}{\delta_1}\|u\|^2_{3/2,P} - \left(c_{\mathrm{dTr}}c_{\Pi}\mathrm{Re}^{-1}\right)^2\delta_2\|\Delta u\|^2_{L^2(\Omega)} - \tfrac{1}{\delta_2}\left\|\tfrac{\partial u}{\partial \boldsymbol{n}}\right\|^2_{1/2,P} \\ + \gamma_1\|u\|^2_{3/2,P} + \gamma_2\left\|\tfrac{\partial u}{\partial \boldsymbol{n}}\right\|^2_{1/2,P},\end{aligned} \tag{5.58}$$

with the inequality $-2\int_\Omega \frac{\partial u}{\partial x}u \leq \|u\|^2_{L^2(\Gamma)}$ obtained by applying the integration by parts to the rotational term. By assuming the sufficiently fine mesh, i.e., $h_P < 1$, $\|u\|_{L^2(\Gamma)} < \|u\|_{3/2,P}$ is valid. Applying this inequality to the rotational term, we have

$$\begin{aligned}\langle \mathscr{F}_P(u), u\rangle \geq \left(1 - (c_{\mathrm{Inv}}c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1}\delta_1 - (c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1}\delta_2\right)\mathrm{Re}^{-1}\|\Delta u\|^2_{L^2(\Omega)} \\ + \left(1 - \tfrac{\mathrm{Ro}^{-1}}{2\gamma_1} - \tfrac{1}{\delta_1\gamma_1}\right)\gamma_1\|u\|^2_{3/2,P} + \left(1 - \tfrac{1}{\delta_2\gamma_2}\right)\gamma_2\left\|\tfrac{\partial u}{\partial \boldsymbol{n}}\right\|^2_{1/2,P}.\end{aligned} \tag{5.60}$$

Choosing $\delta_1 > 0$ and $\delta_2 > 0$ small enough such that $\delta_1^{-1} > 2(c_{\mathrm{Inv}}c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1}$ and $\delta_2^{-1} > 2(c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1}$ makes

$$1 - (c_{\mathrm{Inv}}c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1}\delta_1 - (c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1}\delta_2 > 0, \tag{5.61}$$

and setting $\gamma_1 > 0$ and $\gamma_2 > 0$ large enough so that

$$\gamma_1 > \tfrac{\mathrm{Ro}^{-1}}{2} + 2(c_{\mathrm{Inv}}c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1} \quad \text{and} \quad \gamma_2 > 2(c_{\mathrm{dTr}}c_{\Pi})^2\mathrm{Re}^{-1} \tag{5.62}$$

■

ensures $1 - \gamma_1^{-1}\left(\frac{\mathrm{Ro}^{-1}}{2} + \frac{1}{\delta_1}\right) > 0$ and $1 - \gamma_2^{-1}\frac{1}{\delta_2} > 0$ as required.

In the following lemma, we prove the stability of u and U.

Lemma 5.8 (Stability). *Let u and U be the solutions to* (5.1) *and* (5.10), *respectively. Then,*

$$|||u|||_P \leq C_{\mathrm{coer}}^{-1}c_E\mathrm{Ro}^{-1}\|f\|_{L^2(\Omega)} \quad \textit{and} \quad |||U|||_P \leq C_{\mathrm{coer}}^{-1}c_E\mathrm{Ro}^{-1}\|f\|_{L^2(\Omega)}. \tag{5.63}$$

Proof. From coercivity and duality,

$$C_{\mathrm{coer}}|||u|||_P^2 \leq \langle \mathscr{F}_P(u), u\rangle = \ell_f(u) \leq c_E\mathrm{Ro}^{-1}\|f\|_{L^2(\Omega)}|||u|||_P, \tag{5.65}$$

and

$$C_{\mathrm{coer}}|||U|||_P^2 \leq \langle \mathscr{F}_P(U), U\rangle = \ell_f(U) \leq c_E\mathrm{Ro}^{-1}\|f\|_{L^2(\Omega)}|||U|||_P. \tag{5.67}$$

■

Lemma 5.9 (*A priori* error estimate). *Let u and U be the solutions to* (5.1) *and* (5.10), *respectively. Then,*

$$\|u - U\|_P \leq C_{\text{apriori}}\left(\inf_{V\in\mathbb{X}_P} \|u - V\|_P + \|\mathscr{E}_P\|_{\mathbb{V}'_P}\right), \tag{5.68}$$

with C_{apriori} *defined below* (5.87).

Proof. If we let $e_P = u - U$,

$$\|u - U\|_P \leq \|u - V\|_P + \|U - V\|_P. \tag{5.69}$$

Let $W = V - U \equiv (V - u) + e_P$. Then,

$$\begin{aligned} C_{\text{coer}}\|U - V\|_P^2 &\leq \langle \mathscr{F}_P(W), W\rangle = L_P(W, W) = L_P(e_P, W) + L_P(V - u, W), \\ &= -[N(u, u, W) - N(U, U, W)] + \langle \mathscr{E}_P, W\rangle + L_P(V - u, W), \\ &\leq |N(u, u, W) - N(U, U, W)| + \|\mathscr{E}_P\|_{\mathbb{V}'_P}\|W\|_P + C_{\text{cont}}\|u - V\|_P\|W\|_P. \end{aligned} \tag{5.71}$$

Adding and subtracting $N(U, u, W)$ to the nonlinear term yields

$$\begin{aligned} N(u, u, W) - N(U, U, W) &= N(u, u, W) - N(U, u, W) + N(U, u, W) - N(U, U, W), \\ &= N(u - U, u, W) + N(U, u - U, W). \end{aligned} \tag{5.73}$$

Using $e_P = W + (u - \mathbb{V}_P)$ and $N(U, W, W) = 0$, we obtain

$$\begin{aligned} N(u, u, W) - N(U, U, W) =& N(W, u, W) + N(u - V, u, W) \\ &+ N(U, W, W) + N(U, u - V, W), \\ =& N(W, u, W) + N(u - V, u, W) + N(U, u - V, W). \end{aligned} \tag{5.75}$$

Applying (5.18) results in

$$\begin{aligned} &|N(u, u, W) - N(U, U, W)| \\ &\quad\leq C_{\mathcal{N}} \left(\|\Delta W\|_{L^2}\|u\|_{H^2} + \|\Delta(u - V)\|_{L^2}\|u\|_{H^2} + \|\Delta U\|_{L^2}\|u - V\|_{H^2}\right) \|W\|_{H^2}. \end{aligned} \tag{5.77}$$

Going back to the estimation of $\|W\|_P$, we have

$$\begin{aligned} C_{\text{coer}}\|W\|_P^2 \leq& C_{\mathcal{N}} \left(\|u\|_P\|W\|_P + \|u - V\|_P\|u\|_P + \|U\|_P\|u - V\|_P\right) \|W\|_P \\ &+ \|\mathscr{E}_P\|_{\mathbb{V}'_P}\|W\|_P + C_{\text{cont}}\|u - V\|_P\|W\|_P. \end{aligned} \tag{5.79}$$

Applying stability in Lemma 5.8 yields

$$\begin{aligned} C_{\text{coer}}\|W\|_P^2 \leq& C_{\mathcal{N}}C_{\text{coer}}^{-1} \left(c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}\|W\|_P + 2c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}\|u - V\|_P\right) \|W\|_P \\ &+ \|\mathscr{E}_P\|_{\mathbb{V}'_P}\|W\|_P + C_{\text{cont}}\|u - V\|_P\|W\|_P, \end{aligned} \tag{5.81}$$

from which we obtain

$$\begin{aligned}\big(C_{\text{coer}}-C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}\big)\,|\!|\!|W|\!|\!|_P \\ \leq \big(2C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}+C_{\text{cont}}\big)\,|\!|\!|u-V|\!|\!|_P+\|\mathscr{E}_P\|_{\mathbb{V}'_P}.\end{aligned} \tag{5.83}$$

Then, the estimation of $|\!|\!|W|\!|\!|_P$ can be achieved as

$$|\!|\!|W|\!|\!|_P \leq \tfrac{2C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}+C_{\text{cont}}}{C_{\text{coer}}-C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}}|\!|\!|u-V|\!|\!|_P+\tfrac{1}{C_{\text{coer}}-C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}}\|\mathscr{E}_P\|_{\mathbb{V}'_P}. \tag{5.84}$$

Now having estimated $|\!|\!|W|\!|\!|_P$, we go back to (5.69) to arrive at

$$\begin{aligned}|\!|\!|u-U|\!|\!|_P \leq \left(1+\tfrac{C_{\text{cont}}+2C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}}{C_{\text{coer}}-C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}}\right)|\!|\!|u-V|\!|\!|_P \\ +\tfrac{1}{C_{\text{coer}}-C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}}\|\mathscr{E}_P\|_{\mathbb{V}'_P},\end{aligned} \tag{5.86}$$

and the desired estimate is achieved with

$$C_{\text{apriori}}=1+\tfrac{\max\{1,C_{\text{cont}}+2C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}\}}{C_{\text{coer}}-C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)}}. \tag{5.87}$$

■

Remark 5.10. Since the parameter $C_{\text{apriori}}>0$, $\|f\|_{L^2(\Omega)}<C_{\text{coer}}^2/(C_{\mathcal{N}}c_E\text{Ro}^{-1})$. This condition is analogous to the small data condition used to prove the uniqueness for the steady-state two-dimensional Navier–Stokes equation [37]. However, as noted in [37], it is also very restrictive since C_{coer}^{-1} is of order $ReRo^{-1}$ which contains the Rossby number; see [37] for more details.

5.4 *A posteriori* analysis

From *a posteriori* analysis, we will obtain an *a posteriori* error estimator η_P over a subset $\omega\subset\Omega$ given by

$$\eta_P^2(\omega)=\sum_{\tau\in P:\tau\subset\omega}\eta_\tau^2,\quad \eta_\tau^2=h_\tau^2\|R_\tau\|_{L^2(\tau)}^2+\sum_{\sigma\subset\partial\tau}h_\sigma^3\|J_\sigma\|_{L^2(\sigma)}^2, \tag{5.88}$$

where

$$R_\tau=\big(\text{Ro}^{-1}f-\text{Re}^{-1}\Delta^2U-J(U,\Delta U)+\text{Ro}^{-1}\tfrac{\partial U}{\partial x}\big)\,\big|_\tau, \tag{5.89}$$

and

$$J_\sigma=\text{Re}^{-1}\left\|\left[\tfrac{\partial\Delta U}{\partial \boldsymbol{n}_\sigma}\right]\right\|_\sigma, \tag{5.90}$$

with $[\![\cdot]\!]_\sigma$ being the edge jump operator. Notice that all our simulation are preformed by using the error indicator (5.88), and it can be an upper bound for the numerical error $|\!|\!| u - U |\!|\!|_P$ up to a decaying term as shown in Corollary 5.18.

To derive (5.88), we define the residual quantity $\mathscr{R}_P \in \mathbb{V}'_P$ given by

$$\langle \mathscr{R}_P, v\rangle = \ell_f(v) - \langle \mathscr{F}_P(U), v\rangle, \quad \forall v \in \mathbb{V}_P. \tag{5.91}$$

The brief outline of this section is as follows. In Lemma 5.11, we begin by obtaining a bound of $|\!|\!| u - U |\!|\!|_P$ with respect to $\|\mathscr{R}_P\|_{\mathbb{V}'_P}$ and $\|\mathscr{E}_P\|_{\mathbb{V}'_P}$. In Lemma 5.13, we verify that $\|\mathscr{R}_P\|_{\mathbb{V}'_P}$ can be dominated by the error estimator (5.88) and two boundary integral terms including the nonlinearity and the stabilization terms along with their estimations in Lemmas 5.14 and 5.15. Finally, we summarize the estimation of $\|\mathscr{R}_P\|_{\mathbb{V}'_P}$ in Theorem 5.16 and an upper bound of $|\!|\!| u - U |\!|\!|_P$ in Corollary 5.18 upon using the estimation of $|\!|\!| \mathscr{E}_P |\!|\!|_{\mathbb{V}'_P}$ in Lemma 5.6.

Lemma 5.11. *Let u and U be the solutions to* (5.1) *and* (5.10)*, respectively. Then,*

$$|\!|\!| u - U |\!|\!|_P \leq C_U \left(\|\mathscr{R}_P\|_{\mathbb{V}'_P} + \|\mathscr{E}_P\|_{\mathbb{V}'_P} \right), \tag{5.92}$$

with $C_U = 1/(C_{\text{coer}} - 2C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)})$.

Proof. Setting $e_P = u - U$ as both arguments of (5.17) yields

$$\begin{aligned}\langle \mathscr{F}_P(e_P), e_P\rangle &= L_P(e_P, e_P) + N(e_P, e_P, e_P), \\ &= L_P(u, e_P) - L_P(U, e_P) - N(U, U, e_P) + N(U, U, e_P), \\ &= L_P(u, e_P) - \langle \mathscr{F}_P(U), e_P\rangle + N(U, U, e_P).\end{aligned} \tag{5.94}$$

In view of inconsistency in Lemma 5.5,

$$\begin{aligned}\langle \mathscr{F}_P(e_P), e_P\rangle &= \ell_f(e_P) - \langle \mathscr{F}_P(U), e_P\rangle + N(U, U, e_P) - N(u, u, e_P) + \langle \mathscr{E}_P, e_P\rangle, \\ &= \langle \mathscr{R}_P, e_P\rangle + N(U, U, e_P) - N(u, u, e_P) + \langle \mathscr{E}_P, e_P\rangle.\end{aligned} \tag{5.96}$$

To estimate the nonlinear terms in (5.96), we obtain

$$\begin{aligned}N(U, U, e_P) - N(u, u, e_P) &= N(U, U, e_P) - N(u, U, e_P) + N(u, U, e_P) \\ &\quad - N(u, u, e_P), \\ &= -N(e_P, U, e_P) - N(e_P, u, e_P).\end{aligned} \tag{5.98}$$

Invoking (5.18) and stability in Lemma 5.8 gives rise to

$$\begin{aligned}|N(U, U, e_P) - N(u, u, e_P)| &\leq C_{\mathcal{N}} |e_P|_{H^2(\Omega)} \left(\|U\|_{H^2(\Omega)} + \|u\|_{H^2(\Omega)} \right) \|e_P\|_{H^2(\Omega)}, \\ &\leq 2C_{\mathcal{N}}C_{\text{coer}}^{-1}c_E\text{Ro}^{-1}\|f\|_{L^2(\Omega)} |\!|\!| e_P |\!|\!|_P^2.\end{aligned} \tag{5.100}$$

By applying above into (5.96) and using coercivity in Lemma 5.7, we obtain

$$C_{\mathrm{coer}}|\!|\!|e_P|\!|\!|_P^2 \leq \langle \mathscr{F}_P(e_P), e_P\rangle \leq \langle \mathscr{R}_P, e_P\rangle + 2C_{\mathcal{N}}C_{\mathrm{coer}}^{-1}c_E\mathrm{Ro}^{-1}\|f\|_{L^2(\Omega)}|\!|\!|e_P|\!|\!|_P^2 + \langle \mathscr{E}_P, e_P\rangle, \tag{5.101}$$

and

$$\left(C_{\mathrm{coer}} - 2C_{\mathcal{N}}C_{\mathrm{coer}}^{-1}c_E\mathrm{Ro}^{-1}\|f\|_{L^2(\Omega)}\right)|\!|\!|e_P|\!|\!|_P^2 \leq \langle \mathscr{F}_P(e_P), e_P\rangle \leq \langle \mathscr{R}_P, e_P\rangle + \langle \mathscr{E}_P, e_P\rangle. \tag{5.102}$$

$\blacksquare$

Notice that, in (5.92), a small data condition $c_E\|f\|_{L^2(\Omega)} < C_{\mathrm{coer}}^2/(2C_{\mathcal{N}}\mathrm{Ro}^{-1})$ is required. In the following, we estimate $\|\mathscr{R}_P\|_{\mathbb{V}'_P}$ and $\|\mathscr{E}_P\|_{\mathbb{V}'_P}$.

Lemma 5.12 (Residual L^2-representation). *The functional $\mathscr{R}_P \in \mathbb{V}'_P$ admits the L^2-representation as*

$$\begin{aligned}\langle \mathscr{R}_P, v\rangle = &\sum_{\tau\in P}\int_\tau R_\tau v - \sum_{\sigma\in\mathcal{E}_P}\int_\sigma J_\sigma v - \mathrm{Re}^{-1}\int_\Gamma \tfrac{\partial U}{\partial \boldsymbol{n}}\Pi_P(\Delta v) + \mathrm{Re}^{-1}\int_\Gamma U\tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}} \\ &- \int_\Gamma \Delta U(\nabla^\perp U\cdot \boldsymbol{n})v - \gamma_1\int_\Gamma h_P^{-3}Uv - \gamma_2\int_\Gamma h_P^{-1}\tfrac{\partial U}{\partial \boldsymbol{n}}\tfrac{\partial v}{\partial \boldsymbol{n}}\end{aligned} \tag{5.104}$$

where R_τ and J_σ are defined in (5.89) *and* (5.90).

Proof. Let $v \in \mathbb{V}_P$. Taking integration by parts on each cell $\tau \in P$ yields

$$\begin{aligned}\langle \mathscr{R}_P, v\rangle = &\sum_{\tau\in P}\int_\tau \left(\mathrm{Ro}^{-1}f - \mathrm{Re}^{-1}\Delta^2 U - J(U,\Delta U) + \mathrm{Ro}^{-1}\tfrac{\partial U}{\partial x}\right)v \\ &+ \sum_{\tau\in P}\left(\mathrm{Re}^{-1}\oint_{\partial\tau}\Delta U\tfrac{\partial v}{\partial \boldsymbol{n}_\tau} - \mathrm{Re}^{-1}\oint_{\partial\tau}\tfrac{\partial \Delta U}{\partial \boldsymbol{n}_\tau}v - \oint_{\partial\tau}\Delta U(\nabla^\perp U\cdot \boldsymbol{n}_\tau)v\right) \\ &- \mathrm{Re}^{-1}\int_\Gamma\left(\Delta U\tfrac{\partial v}{\partial \boldsymbol{n}} + \tfrac{\partial U}{\partial \boldsymbol{n}}\Pi_P(\Delta v)\right) + \mathrm{Re}^{-1}\int_\Gamma\left(\tfrac{\partial \Delta U}{\partial \boldsymbol{n}}v + U\tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}}\right) \\ &- \gamma_1\int_\Gamma h_P^{-3}Uv - \gamma_2\int_\Gamma h_P^{-1}\tfrac{\partial U}{\partial \boldsymbol{n}}\tfrac{\partial v}{\partial \boldsymbol{n}}.\end{aligned} \tag{5.106}$$

We already know that

$$\sum_{\tau\in P}\oint_{\partial\tau}\Delta U\tfrac{\partial v}{\partial \boldsymbol{n}_\tau} = \int_\Gamma \Delta U\tfrac{\partial v}{\partial \boldsymbol{n}}, \tag{5.107}$$

and

$$\sum_{\tau\in P}\oint_{\partial\tau}\tfrac{\partial \Delta U}{\partial \boldsymbol{n}_\tau}v = \sum_{\sigma\in\mathcal{E}_P}\int_\sigma \left[\!\left[\tfrac{\partial \Delta U}{\partial \boldsymbol{n}_\sigma}\right]\!\right]_\sigma v + \int_\Gamma \tfrac{\partial \Delta U}{\partial \boldsymbol{n}}v. \tag{5.108}$$

Let τ_1 and τ_2 be cells sharing an interior edge σ with corresponding outward unit normal vectors $\boldsymbol{n}_{\tau_1}$ and $\boldsymbol{n}_{\tau_2}$ and $\boldsymbol{n}_{\tau_1} = -\boldsymbol{n}_{\tau_2}$. By setting $\boldsymbol{n}_\sigma = \boldsymbol{n}_{\tau_1}$, we have

$$\begin{aligned}\int_{\partial\tau_1\cap\sigma} \Delta U(\nabla^\perp U \cdot \boldsymbol{n}_{\tau_1})v + \int_{\partial\tau_2\cap\sigma} \Delta U(\nabla^\perp U \cdot \boldsymbol{n}_{\tau_2})v,&\\ = \int_\sigma \left[(\Delta U)|_{\tau_1}(\nabla^\perp U \cdot \boldsymbol{n}_\sigma) - (\Delta U)|_{\tau_2}(\nabla^\perp U \cdot \boldsymbol{n}_\sigma)\right] v,&\\ = \int_\sigma \left[(\Delta U)|_{\tau_1} - (\Delta U)|_{\tau_2}\right](\nabla^\perp U \cdot \boldsymbol{n}_\sigma)v,&\\ = \int_\sigma [\![\Delta U]\!]_\sigma (\nabla^\perp U \cdot \boldsymbol{n}_\sigma)v = 0.&\end{aligned} \tag{5.110}$$

Upon summing (5.110) over all cells τ, we can write

$$\sum_{\tau\in P} \oint_{\partial\tau} \Delta U(\nabla^\perp U \cdot \boldsymbol{n}_\tau)v = \int_\Gamma \Delta U(\nabla^\perp U \cdot \boldsymbol{n})v. \tag{5.111}$$

By applying (5.107), (5.108), (5.111) and (5.90) to the second line of (5.106), we obtain

$$\begin{aligned}\sum_{\tau\in P}\left(\mathrm{Re}^{-1}\oint_{\partial\tau}\Delta U\tfrac{\partial v}{\partial \boldsymbol{n}_\tau} - \mathrm{Re}^{-1}\oint_{\partial\tau}\tfrac{\partial\Delta U}{\partial\boldsymbol{n}_\tau}v - \oint_{\partial\tau}\Delta U(\nabla^\perp U\cdot\boldsymbol{n}_\tau)v\right)&\\ = \mathrm{Re}^{-1}\int_\Gamma \Delta U\tfrac{\partial v}{\partial\boldsymbol{n}} - \sum_{\sigma\in\mathcal{E}_P}\int_\sigma J_\sigma v - \mathrm{Re}^{-1}\int_\Gamma \tfrac{\partial\Delta U}{\partial\boldsymbol{n}}v - \int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})v.&\end{aligned} \tag{5.113}$$

Substituting above equality and (5.89) into (5.106) results in

$$\begin{aligned}\langle\mathscr{R}_P, v\rangle = &\sum_{\tau\in P}\int_\tau R_\tau v - \sum_{\sigma\in\mathcal{E}_P}\int_\sigma J_\sigma v\\ &- \mathrm{Re}^{-1}\int_\Gamma \tfrac{\partial U}{\partial\boldsymbol{n}}\Pi_P(\Delta v) + \mathrm{Re}^{-1}\int_\Gamma U\tfrac{\partial\Pi_P(\Delta v)}{\partial\boldsymbol{n}}\\ &- \int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})v - \gamma_1\int_\Gamma h_P^{-3}Uv - \gamma_2\int_\Gamma h_P^{-1}\tfrac{\partial U}{\partial\boldsymbol{n}}\tfrac{\partial v}{\partial\boldsymbol{n}}.\end{aligned} \tag{5.115}$$

■

Lemma 5.13. *For $v \in \mathbb{V}_P$, the residual $\mathscr{R}_P$ is estimated as*

$$\begin{aligned}|\langle\mathscr{R}_P, v\rangle| \le c_{\text{shape}}^2\big\{&\eta_P(\Omega) + (\gamma_1 + c_{\text{Inv}}c_{\text{dTr}}c_\Pi)\|U\|_{3/2,P}\\ &+ (\gamma_2 + c_{\text{dTr}}c_\Pi)\left\|\tfrac{\partial U}{\partial\boldsymbol{n}}\right\|_{1/2,P}\big\}|v|_{H^2(\Omega)} + \left|\int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})(v - I_P v)\right|.\end{aligned} \tag{5.117}$$

Proof. Since $\langle \mathscr{R}_P, I_P v\rangle = \ell_f(I_P v) - \langle \mathscr{F}_P(U), I_P v\rangle = 0$, we have

$$\langle \mathscr{R}_P, v\rangle = \langle \mathscr{R}_P, v - I_P v\rangle. \tag{5.118}$$

Then,

$$\begin{aligned}|\langle \mathscr{R}_P, v - I_P v\rangle| \le& \sum_{\tau\in P} \|R_\tau\|_{L^2(\tau)}\|v - I_P v\|_{L^2(\tau)} + \sum_{\sigma\in\mathcal{E}_P} \|J_\sigma\|_{L^2(\sigma)}\|v - I_P v\|_{L^2(\sigma)}\\ &+ \mathrm{Re}^{-1}\left|\int_\Gamma \tfrac{\partial U}{\partial \boldsymbol{n}}\Pi_P[\Delta(v - I_P v)]\right| + \mathrm{Re}^{-1}\left|\int_\Gamma U\tfrac{\partial \Pi_P[\Delta(v-I_P v)]}{\partial \boldsymbol{n}}\right|\\ &+ \left|\int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})(v - I_P v)\right| + \gamma_1\left|\int_\Gamma h_P^{-3}U(v - I_P v)\right|\\ &+ \gamma_2\left|\int_\Gamma h_P^{-1}\tfrac{\partial U}{\partial \boldsymbol{n}}\tfrac{\partial(v-I_P v)}{\partial \boldsymbol{n}}\right|.\end{aligned} \tag{5.120}$$

Upon applying (2.86) with $k = 0$ from Theorem 2.42 on the interior terms, we have

$$\begin{aligned}\sum_{\tau\in P}\|R_\tau\|_{L^2(\tau)}\|v - I_P v\|_{L^2(\tau)} &\le \left(\sum_{\tau\in P} h_\tau^4\|R_\tau\|^2_{L^2(\tau)}\right)^{1/2}\left(\sum_{\tau\in P} h_\tau^{-4}\|v - I_P v\|^2_{L^2(\tau)}\right)^{1/2},\\ &\le \left(\sum_{\tau\in P} h_\tau^4\|R_\tau\|^2_{L^2(\tau)}\right)^{1/2} c_{\text{shape}}\left(\sum_{\tau\in P}|v|^2_{H^2(\omega_\tau)}\right)^{1/2},\\ &\le c^2_{\text{shape}}\left(\sum_{\tau\in P} h_\tau^4\|R_\tau\|^2_{L^2(\tau)}\right)^{1/2}|v|_{H^2(\Omega)}.\end{aligned} \tag{5.122}$$

Similarly, the jump terms in (5.120) can be treated as

$$\sum_{\sigma\in\mathcal{E}_P}\|J_\sigma\|_{L^2(\sigma)}\|v - I_P v\|_{L^2(\sigma)} \le c^2_{\text{shape}}\left(\sum_{\sigma\in\mathcal{E}_P} h_\sigma^3\|J_\sigma\|^2_{L^2(\sigma)}\right)^{1/2}|v|_{H^2(\Omega)}. \tag{5.124}$$

From now on, we estimate the domain boundary integral terms in (5.120) as follows. The first term in the second line can be estimated as follows:

$$
\begin{aligned}
\left|\int_\Gamma \tfrac{\partial U}{\partial \boldsymbol{n}}\Pi_P[\Delta(v-I_Pv)]\right| &\le \sum_{\sigma\in\mathcal{G}_h}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)}\|\Pi_P[\Delta(v-I_Pv)]\|_{L^2(\sigma)},\\
&\le \sum_{\sigma\in\mathcal{G}_h}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)} c_{\mathrm{dTr}}h_\sigma^{-1/2}\left\|\Pi_P[\Delta(v-I_Pv)]\right\|_{L^2(\tau_\sigma)},\\
&\le c_{\mathrm{dTr}}\Bigg(\sum_{\sigma\in\mathcal{G}_h}h_\sigma^{-1}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\right\|^2_{L^2(\sigma)}\Bigg)^{1/2}\Bigg(\sum_{\sigma\in\mathcal{G}_h}\|\Pi_P[\Delta(v-I_Pv)]\|^2_{L^2(\tau_\sigma)}\Bigg)^{1/2},\\
&\le c_{\mathrm{dTr}}\Bigg(\sum_{\sigma\in\mathcal{G}_h}h_\sigma^{-1}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\right\|^2_{L^2(\sigma)}\Bigg)^{1/2}\Bigg(\sum_{\tau\in P}\|\Pi_P[\Delta(v-I_Pv)]\|^2_{L^2(\tau)}\Bigg)^{1/2},\\
&\le c_{\mathrm{dTr}}\Bigg(\sum_{\sigma\in\mathcal{G}_h}h_\sigma^{-1}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\right\|^2_{L^2(\sigma)}\Bigg)^{1/2}\left\|\Pi_P[\Delta(v-I_Pv)]\right\|_{L^2(\Omega)},\\
&\le c_{\mathrm{dTr}}c_\Pi c^2_{\mathrm{shape}}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}|v|_{H^2(\Omega)}.
\end{aligned}
\tag{5.126}
$$

For the last inequality, we use the following stability of the L^2-projector Π_P in $L^2(\Omega)$ (see Remark 2.44) and Theorem 2.42, with $k=2$;

$$
\|\Pi_P[\Delta(v-I_Pv)]\|_{L^2(\Omega)} \le c_\Pi^2\|\Delta(v-I_Pv)\|_{L^2(\Omega)} \le c_\Pi^2 c^4_{\mathrm{shape}}|v|^2_{H^2(\Omega)}. \tag{5.127}
$$

In a similar manner, with the aid of a discrete inverse estimate, the second term in the second line of (5.120) can be estimated as

$$
\begin{aligned}
\left|\int_\Gamma U\tfrac{\partial\Pi_P[\Delta(v-I_Pv)]}{\partial \boldsymbol{n}}\right| &\le \sum_{\sigma\in\mathcal{G}_h}\|U\|_{L^2(\sigma)}\left\|\tfrac{\partial}{\partial \boldsymbol{n}_\sigma}\Pi_P[\Delta(v-I_Pv)]\right\|_{L^2(\sigma)},\\
&\le \sum_{\sigma\in\mathcal{G}_h}\|U\|_{L^2(\sigma)}c_{\mathrm{dTr}}h_\sigma^{-1/2}c_{\mathrm{Inv}}h_\sigma^{-1}\|\Pi_P[\Delta(v-I_Pv)]\|_{L^2(\tau_\sigma)},\\
&\le c_{\mathrm{dTr}}\Bigg(\sum_{\sigma\in\mathcal{G}_h}h_\sigma^{-3/2}\|U\|^2_{L^2(\sigma)}\Bigg)^{1/2}\Bigg(\sum_{\sigma\in\mathcal{G}_h}\|\Pi_P[\Delta(v-I_Pv)]\|^2_{L^2(\tau_\sigma)}\Bigg)^{1/2},\\
&\le c_\Pi c_{\mathrm{dTr}}c_{\mathrm{Inv}}c^2_{\mathrm{shape}}\|U\|_{3/2,P}|v|_{H^2(\Omega)}.
\end{aligned}
\tag{5.129}
$$

Moreover, the second term in the third line of (5.120) can be estimated as

$$\begin{aligned}\left|\int_\Gamma h_P^{-3} U(v - I_P v)\right| &\le \sum_{\sigma\in\mathcal{G}_h} h_\sigma^{-3/2}\|U\|_{L^2(\sigma)} h_\sigma^{-3/2}\|v - I_P v\|_{L^2(\sigma)},\\ &\le \left(\sum_{\sigma\in\mathcal{G}_h} h_\sigma^{-3}\|U\|^2_{L^2(\sigma)}\right)^{1/2}\left(\sum_{\sigma\in\mathcal{G}_h} h_\sigma^{-3}\|v - I_P v\|^2_{L^2(\sigma)}\right)^{1/2},\\ &\le c_{\text{shape}}\|U\|_{3/2,P}\left(\sum_{\sigma\in\mathcal{G}_h}|v|^2_{H^2(\omega_\sigma)}\right)^{1/2},\\ &= c^2_{\text{shape}}\|U\|_{3/2,P}|v|_{H^2(\Omega)},\end{aligned} \tag{5.131}$$

and, similarly, the last term of (5.120) can be estimate as

$$\left|\int_\Gamma h_P^{-1}\tfrac{\partial U}{\partial \boldsymbol{n}}\tfrac{\partial}{\partial \boldsymbol{n}}(v - I_P v)\right| \le c^2_{\text{shape}}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}|v|_{H^2(\Omega)}. \tag{5.133}$$

Substituting all above inequalities into (5.120) and (5.118) results in

$$\begin{aligned}|\langle \mathscr{R}_P, v\rangle| \le{}& c^2_{\text{shape}}\left\{\left(\sum_{\tau\in P} h_\tau^4\|R_\tau\|^2_{L^2(\tau)}\right)^{1/2} + \left(\sum_{\sigma\in\mathcal{E}_P} h_\sigma^3\|J_\sigma\|^2_{L^2(\sigma)}\right)^{1/2}\right\}|v|_{H^2(\Omega)}\\ &+ c^2_{\text{shape}}\left\{(c_{\text{Inv}}c_{\text{dTr}}c_\Pi + \gamma_1)\|U\|_{3/2,P} + (c_{\text{dTr}}c_\Pi + \gamma_2)\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}\right\}|v|_{H^2(\Omega)}\\ &+ \left|\int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})(v - I_P v)\right|.\end{aligned} \tag{5.135}$$

■

In the following lemma, we estimate the nonlinear boundary integral term in (5.117).

Lemma 5.14. *Let u and U be the solutions to* (5.1) *and* (5.10)*, respectively. Then,*

$$\left|\int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})(v - I_P v)\right| \le c^2_{\text{shape}}C_N\|U\|_{3/2,P}|v|_{H^2(\omega_P^\Gamma)}, \quad v\in\mathbb{V}_P, \tag{5.136}$$

where $C_N = c_{\text{dTr}}c^2_{\text{Inv}}C^{-1}_{\text{coer}}c_E\|f\|_{L^2(\Omega)}$.

Proof. Decomposing the integral into edges and applying the Cauchy-Schwarz inequality on each edge integral separately result in

$$\begin{aligned}\left|\int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})(v-I_P v)\right| &= \left|\sum_{\sigma\in\mathcal{G}_P}\int_\sigma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})(v-I_P v)\right|,\\ &\le \sum_{\sigma\in\mathcal{G}_P}\|\Delta U\nabla^\perp U\cdot\boldsymbol{n}\|_{L^2(\sigma)}\|v-I_P v\|_{L^2(\sigma)},\\ &\le \Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma^3\|\Delta U\nabla^\perp U\cdot\boldsymbol{n}\|^2_{L^2(\sigma)}\Bigg)^{1/2}\Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma^{-3}\|v-I_P v\|^2_{L^2(\sigma)}\Bigg)^{1/2},\\ &\le c^2_{\text{shape}}\Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma^3\|\Delta U\nabla^\perp U\cdot\boldsymbol{n}\|^2_{L^2(\sigma)}\Bigg)^{1/2}|v|_{H^2(\omega_P^\Gamma)},\end{aligned}\tag{5.138}$$

using (2.87) in the last line and the shape-regularity $\sum_{\sigma\in\mathcal{G}_P}|u|^2_{H^2(\omega_\sigma)}\le c^2_{\text{shape}}|u|^2_{H^2(\omega_P^\Gamma)}$. Applying the Cauchy–Schwarz inequality to the first norm of the above last line yields

$$\|\Delta U\nabla^\perp U\cdot\boldsymbol{n}\|_{L^2(\sigma)}\le\|\Delta U\|_{L^4(\sigma)}\|\nabla^\perp U\cdot\boldsymbol{n}\|_{L^4(\sigma)}.\tag{5.139}$$

The first term in (5.139) is estimated as

$$\begin{aligned}\|\Delta U\|_{L^4(\sigma)}\le c_{\text{Inv}}h_\sigma^{-1/4}\|\Delta U\|_{L^2(\sigma)} &\le c_{\text{Inv}}c_{\text{dTr}}h_\sigma^{-1/4}h_\sigma^{-1/2}\|\Delta U\|_{L^2(\tau)},\\ &\le c_{\text{Inv}}c_{\text{dTr}}h_\sigma^{-3/4}\|\Delta U\|_{L^2(\tau_\sigma)},\end{aligned}\tag{5.141}$$

where τ_σ is the cell along Γ with σ as one of its edges. To obtain the first line of (5.141) we use (2.54) with $s=t=0$, $q=4$, and $p=2$ for the first inequality and (2.55) for the second inequality.

Next, the second norm in (5.139) is estimated as

$$\|\nabla^\perp U\cdot\boldsymbol{n}\|_{L^4(\sigma)}=\|\nabla^\perp U\cdot\boldsymbol{n}\|_{L^4(\sigma)}\le c_{\text{Inv}}h_\sigma^{-1/4}\|\nabla^\perp U\cdot\boldsymbol{n}\|_{L^2(\sigma)}.\tag{5.142}$$

By combining (5.141) and (5.142) into (5.139) and applying Lemma 5.8 into $\|\Delta U\|_{L^2(\tau_\sigma)}$, we arrive at

$$\begin{aligned}&\Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma^3\|\Delta U\nabla^\perp U\cdot\boldsymbol{n}\|^2_{L^2(\sigma)}\Bigg)^{1/2}\\ &\qquad\le c_{\text{Inv}}c_{\text{dTr}}c_E C_{\text{coer}}^{-1}\|f\|_{L^2(\Omega)}\Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma\|\nabla^\perp U\cdot\boldsymbol{n}\|^2_{L^2(\sigma)}\Bigg)^{1/2}.\end{aligned}\tag{5.144}$$

Applying an inverse estimate $\|\nabla^\perp U \cdot \boldsymbol{n}\|_{L^2(\sigma)} \leq |U|_{H^1(\sigma)} \leq c_{\mathrm{Inv}} h_\sigma^{-1} \|U\|_{L^2(\sigma)}$ to the second norm in the above yields

$$\sum_{\sigma \in \mathcal{G}_P} h_\sigma \|\nabla^\perp U \cdot \boldsymbol{n}\|_{L^2(\sigma)}^2 \leq c_{\mathrm{Inv}}^2 \sum_{\sigma \in \mathcal{G}_P} h_\sigma^{-1} \|U\|_{L^2(\sigma)}^2 \leq c_{\mathrm{Inv}}^2 \|U\|_{3/2,P}^2, \tag{5.145}$$

using $h_\sigma^{-1} \leq h_\sigma^{-3}$ by assuming sufficiently refined mesh (i.e., $h_P \leq 1$). This results in

$$\left(\sum_{\sigma \in \mathcal{G}_P} h_\sigma^3 \|\Delta U \nabla^\perp U \cdot \boldsymbol{n}\|_{L^2(\sigma)}^2 \right)^{1/2} \leq c_{\mathrm{dTr}} c_{\mathrm{Inv}}^2 C_{\mathrm{coer}}^{-1} c_E \|f\|_{L^2(\Omega)} \|U\|_{3/2,P}. \tag{5.146}$$

Substituting the above inequality into (5.138) completes the proof of the lemma. ∎

In the following lemma, we estimate the stabilization terms in (5.117).

Lemma 5.15. *For sufficiently large γ_1 and γ_2 we have*

$$|U|_P^2 \leq \frac{\frac{1}{2} C_{\mathrm{coer}}^{-1}}{\min\{\gamma_1 - C_{\mathrm{coer}}^{-1} C_\Gamma, \gamma_2 - C_{\mathrm{coer}}^{-1} C_\Gamma\}} \eta_P^2(\omega_P^\Gamma), \tag{5.147}$$

with

$$\begin{aligned} C_\Gamma : &= C_{\mathrm{bdry}} \max\left\{\mathrm{Re}^{-1} + C_N + \tfrac{1}{2}, \mathrm{Ro}^{-1}\right\} \\ &\quad + c_{\mathrm{dTr}}^2 c_{\mathrm{shape}}^4 C_{\mathrm{bdry}}^2 + 3\mathrm{Re}^{-1} c_{\mathrm{Inv}} (c_{\mathrm{dTr}} + 1). \end{aligned} \tag{5.149}$$

Proof. Using coercivity in Lemma 5.7, (5.10), and (5.17), for any $V^0 \in \mathbb{X}_P^0$, we obtain

$$\begin{aligned} \gamma_1 \|U\|_{3/2,P}^2 + \gamma_2 \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|_{1/2,P}^2 &\leq C_{\mathrm{coer}}^{-1} \langle \mathscr{F}_P(U - V^0), U - V^0 \rangle, \\ &= C_{\mathrm{coer}}^{-1} L_P(U - V^0, U - V^0), \\ &= C_{\mathrm{coer}}^{-1} \left\{ L_P(U, U - V^0) - L_P(V^0, U - V^0) \right\}, \\ &= C_{\mathrm{coer}}^{-1} \left\{ \ell_f(U - V^0) - N(U, U, U - V^0) \right\} \\ &\quad - C_{\mathrm{coer}}^{-1} L_P(V^0, U - V^0). \end{aligned} \tag{5.151}$$

If we let $V^0 = U + (V^0 - U)$,

$$L_P(V^0, U - V^0) = L_P(U, U - V^0) - L_P(U - V^0, U - V^0). \tag{5.152}$$

The first term on the right can be expressed as

$$L_P(U, U - V^0) = \ell_f(U - V^0) - \langle \mathscr{F}_P(U), U - V^0 \rangle = \langle \mathscr{R}_P, U - V^0 \rangle,$$

using (5.17) and (5.91), and we arrive at

$$\gamma_1\|U\|_{3/2,P}^2+\gamma_2\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2 \le C_{\text{coer}}^{-1}\langle \mathscr{R}_P, U-V^0\rangle + C_{\text{coer}}^{-1}L_P(U-V^0,U-V^0). \tag{5.154}$$

Set $V^0=U^0$ and define $U^\perp = U-U^0$. In view of Lemma 5.12, we have

$$\begin{aligned}\langle \mathscr{R}_P, U^\perp\rangle = &\sum_{\tau\in P\cap\omega_P^\Gamma}\int_\tau R_\tau U^\perp - \sum_{\sigma\in\mathcal{E}_P}\int_\sigma J_\sigma U^\perp - \mathrm{Re}^{-1}\int_\Gamma \tfrac{\partial U}{\partial \boldsymbol{n}}\Delta U^\perp + \mathrm{Re}^{-1}\int_\Gamma U\tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}}\\ &-\int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})U^\perp - \gamma_1\|U^\perp\|_{3/2,P}^2-\gamma_2\left\|\tfrac{\partial U^\perp}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2.\end{aligned} \tag{5.156}$$

From the definition of L_P, we have

$$\begin{aligned}L_P(U^\perp,U^\perp) = &\,\mathrm{Re}^{-1}\|\Delta U^\perp\|_{L^2(\Omega)}^2 - \mathrm{Ro}^{-1}\int_\Omega \frac{\partial U^\perp}{\partial x}U^\perp\\ &+\mathrm{Re}^{-1}\int_\Gamma\left(\tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}}U^\perp + U^\perp\tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}}\right) - \mathrm{Re}^{-1}\int_\Gamma\left(\Delta U^\perp\tfrac{\partial U^\perp}{\partial \boldsymbol{n}}+\tfrac{\partial U^\perp}{\partial \boldsymbol{n}}\Delta U^\perp\right)\\ &+\gamma_1\|U^\perp\|_{3/2,P}^2+\gamma_2\left\|\tfrac{\partial U^\perp}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2.\end{aligned} \tag{5.158}$$

Summing above two equalities result in

$$\begin{aligned}\langle \mathscr{R}_P, U^\perp\rangle + L_P(U^\perp,U^\perp) = &\,\mathrm{Re}^{-1}\|\Delta U^\perp\|_{L^2(\Omega)}^2 - \mathrm{Ro}^{-1}\int_\Omega \tfrac{\partial U^\perp}{\partial x}U^\perp\\ &+\sum_{\tau\in P}\int_\tau R_\tau U^\perp + \sum_{\sigma\in\mathcal{E}_P}\int_\sigma J_\sigma U^\perp - \int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})U^\perp\\ &+2\mathrm{Re}^{-1}\int_\Gamma \tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}}U^\perp - 2\mathrm{Re}^{-1}\int_\Gamma \Delta U^\perp\tfrac{\partial U^\perp}{\partial \boldsymbol{n}}.\end{aligned} \tag{5.160}$$

Following the exact calculation as carried in the proof of Theorem 4.8, we estimate all terms in (5.160). The spline function $U^\perp$ is supported only on ω_P^Γ, and thus $U=U^\perp$ along Γ and all norms and integrals on Ω reduce to ones over ω_P^Γ. We begin by estimating the first two terms in (5.160). By applying Theorem 4.8 with $k=2$, we have

$$\begin{aligned}\|\Delta U^\perp\|_{L^2(\Omega)}^2 = \|\Delta U^\perp\|_{L^2(\omega_P^\Gamma)}^2 &\le |U^\perp|_{H^2(\omega_P^\Gamma)}^2,\\ &=\sum_{\sigma\in\mathcal{G}_P}|U^\perp|_{H^2(\omega_\sigma)}^2 \le C_{\text{bdry}}|U|_P^2,\end{aligned} \tag{5.162}$$

and, provided that $h_P\le 1$,

$$-2\int_\Omega \tfrac{\partial U^\perp}{\partial x}U^\perp \le \|U^\perp\|_{L^2(\Gamma)}^2 = \|U\|_{L^2(\Gamma)}^2 \le \|U\|_{3/2,P}^2. \tag{5.163}$$

By combining above two inequalities, we have

$$\begin{aligned}\mathrm{Re}^{-1}\|\Delta U^{\perp}\|^2_{L^2(\Omega)} - \mathrm{Ro}^{-1}\int_{\Omega}\tfrac{\partial U^{\perp}}{\partial x}U^{\perp}& \\ \leq C_{\mathrm{bdry}}\left\{\left(\mathrm{Re}^{-1}+\frac{1}{2}\right)\|U\|^2_{3/2,P} + \mathrm{Ro}^{-1}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|^2_{1/2,P}\right\}&.\end{aligned} \tag{5.165}$$

Next, the indicators in (5.160) can be estimated as

$$\begin{aligned}\sum_{\tau\in P}\int_{\tau}R_{\tau}U^{\perp} - \sum_{\sigma\in\mathcal{E}_P}\int_{\sigma}J_{\sigma}U^{\perp}& \\ \leq \Bigg(\sum_{\tau\in P\cap\omega_P^{\Gamma}} h_{\tau}^4\|R_{\tau}\|^2_{L^2(\tau)}\Bigg)^{1/2}\Bigg(\sum_{\tau\in P\cap\omega_P^{\Gamma}} h_{\tau}^{-4}\|U^{\perp}\|^2_{L^2(\tau)}\Bigg)^{1/2}& \\ +\Bigg(\sum_{\sigma\in\mathcal{E}_P\cap\omega_P^{\Gamma}} h_{\sigma}^3\|J_{\sigma}\|^2_{L^2(\sigma)}\Bigg)^{1/2}\Bigg(\sum_{\sigma\in\mathcal{E}_P\cap\omega_P^{\Gamma}} h_{\sigma}^{-3}\|U^{\perp}\|^2_{L^2(\sigma)}\Bigg)^{1/2}&.\end{aligned} \tag{5.167}$$

By the shape-regularity of the mesh in (2.73), $\operatorname{diam}(\omega_{\sigma}) \leq c_{\mathrm{shape}}h_{\tau}$ for a cell $\tau\subset\omega_{\sigma}$. This makes the cell diameter h_{τ} comparable to the edge length h_{σ} because $h_{\sigma}\leq\operatorname{diam}(\omega_{\sigma})$. Then, $h_{\tau}^{-1}\leq c_{\mathrm{shape}}h_{\sigma}^{-1}$. Using Theorem 4.8, with $k=0$, the last norm in the first line of (5.167) is treated as

$$\begin{aligned}h_{\tau}^{-4}\|U^{\perp}\|^2_{L^2(\tau)} &\leq C_{\mathrm{bdry}}h_{\tau}^{-4}\left(h_{\sigma}\|U\|^2_{L^2(\sigma)} + h_{\sigma}^3\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_{\sigma}}\right\|^2_{L^2(\sigma)}\right), \\ &\leq C_{\mathrm{bdry}}c_{\mathrm{shape}}^4h_{\sigma}^{-4}\left(h_{\sigma}\|U\|^2_{L^2(\sigma)} + h_{\sigma}^3\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_{\sigma}}\right\|^2_{L^2(\sigma)}\right), \\ &\leq c_{\mathrm{shape}}^4C_{\mathrm{bdry}}|U|^2_P,\end{aligned} \tag{5.169}$$

for $\tau\subset\omega_{\sigma}$ and $\sigma\in\mathcal{G}_P$. Summing over all cells $\tau\in P\cap\omega_P^{\Gamma}$, we have

$$\sum_{\tau\in P\cap\omega_P^{\Gamma}} h_{\tau}^{-4}\|U^{\perp}\|^2_{L^2(\tau)} \leq (c_{\mathrm{dTr}}c_{\mathrm{shape}}^2)^2C_{\mathrm{bdry}}|U|^2_P. \tag{5.171}$$

Applying (2.54) to the second norm in the second line of (5.167) gives rise to $h_{\sigma}^{-3}\|U^{\perp}\|^2_{L^2(\sigma)}\leq c_{\mathrm{dTr}}^2h_{\sigma}^{-4}\|U^{\perp}\|^2_{L^2(\tau)}$. Then, by summing over edges $\sigma\in\mathcal{E}_P\cap\omega_P^{\Gamma}$, we have

$$\sum_{\sigma\in\mathcal{E}_P\cap\omega_P^{\Gamma}} h_{\sigma}^{-3}\|U^{\perp}\|^2_{L^2(\sigma)} \leq c_{\mathrm{dTr}}^2\sum_{\tau\in P\cap\omega_P^{\Gamma}} h_{\sigma}^{-4}\|U^{\perp}\|^2_{L^2(\tau)} \leq (c_{\mathrm{dTr}}c_{\mathrm{shape}}^2)^2C_{\mathrm{bdry}}|U|^2_P. \tag{5.173}$$

Substituting above inequalities into (5.167) and applying Young's inequality yield

$$\begin{aligned}
&\Bigg(\sum_{\tau\in P\cap\omega_P^\Gamma} h_\tau^4 \|R_\tau\|_{L^2(\tau)}^2 \Bigg)^{1/2} \Bigg(\sum_{\tau\in P\cap\omega_P^\Gamma} h_\tau^{-4} \|U^\perp\|_{L^2(\tau)}^2 \Bigg)^{1/2} \\
&\quad + \Bigg(\sum_{\sigma\in\mathcal{E}_P\cap\omega_P^\Gamma} h_\sigma^3 \|J_\sigma\|_{L^2(\sigma)}^2 \Bigg)^{1/2} \Bigg(\sum_{\sigma\in\mathcal{E}_P\cap\omega_P^\Gamma} h_\sigma^{-3} \|U^\perp\|_{L^2(\sigma)}^2 \Bigg)^{1/2}, \\
&\le \frac{1}{2}\Bigg(\sum_{\tau\in P\cap\omega_P^\Gamma} h_\tau^4 \|R_\tau\|_{L^2(\tau)}^2 + \sum_{\sigma\in\mathcal{E}_P\cap\omega_P^\Gamma} h_\sigma^3 \|J_\sigma\|_{L^2(\sigma)}^2 \Bigg) + (c_{\mathrm{dTr}} c_{\mathrm{shape}}^2)^2 C_{\mathrm{bdry}} |U|_P^2.
\end{aligned} \tag{5.175}$$

Finally, we obtain the estimate of the indicators as

$$\sum_{\tau\in P} \int_\tau R_\tau U^\perp - \sum_{\sigma\in\mathcal{E}_P} \int_\sigma J_\sigma U^\perp \le \frac{1}{2}\eta_P^2(\omega_P^\Gamma) + (c_{\mathrm{dTr}} c_{\mathrm{shape}}^2)^2 C_{\mathrm{bdry}} |U|_P^2. \tag{5.176}$$

Next, we begin by estimating the last two boundary edge terms in (5.160) as

$$\begin{aligned}
&\int_\Gamma \tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}} U^\perp - \int_\Gamma \Delta U^\perp \tfrac{\partial U^\perp}{\partial \boldsymbol{n}} \\
&\qquad \le \sum_{\sigma\in\mathcal{G}_P} \left\| \tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)} \|U^\perp\|_{L^2(\sigma)} + \sum_{\sigma\in\mathcal{G}_P} \|\Delta U^\perp\|_{L^2(\sigma)} \left\| \tfrac{\partial U^\perp}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)}.
\end{aligned} \tag{5.178}$$

In view of inverse and trace estimates of Theorem 2.36, we have

$$\begin{aligned}
\left\| \tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)} &\le c_{\mathrm{Inv}} c_{\mathrm{dTr}} h_\sigma^{-3/2} \|\Delta U^\perp\|_{L^2(\tau)} \le c_{\mathrm{Inv}} c_{\mathrm{dTr}} h_\sigma^{-3/2} |U^\perp|_{H^2(\tau)} \\
&\le c_{\mathrm{Inv}} c_{\mathrm{dTr}} C_{\mathrm{bdry}}^{1/2} h_\sigma^{-3/2} \left\{ h_\sigma^{-3} \|U\|_{L^2(\sigma)}^2 + h_\sigma^{-1} \left\| \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)}^2 \right\}^{1/2},
\end{aligned} \tag{5.180}$$

for $\sigma \subset \partial\tau$, using Theorem 4.8 with $k = 2$ in the last inequality. For any non-negative pair a and b, it follows that $a^2 + b^2 \le (a+b)^2$; i.e, $\sqrt{a^2+b^2} \le a + b$, which makes

$$\begin{aligned}
&\left\| \tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)} \|U^\perp\|_{L^2(\sigma)} \\
&\le c_{\mathrm{Inv}} c_{\mathrm{dTr}} C_{\mathrm{bdry}}^{1/2} h_\sigma^{-3/2} \|U^\perp\|_{L^2(\sigma)} \left\{ h_\sigma^{-3/2} \|U\|_{L^2(\sigma)} + h_\sigma^{-1/2} \left\| \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)} \right\}, \\
&\le c_{\mathrm{Inv}} c_{\mathrm{dTr}} C_{\mathrm{bdry}}^{1/2} \left(\frac{h_\sigma^{-3} \|U^\perp\|_{L^2(\sigma)}^2}{2} + \frac{1}{2} \left\{ h_\sigma^{-3/2} \|U\|_{L^2(\sigma)} + h_\sigma^{-1/2} \left\| \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)} \right\}^2 \right), \\
&\le c_{\mathrm{Inv}} c_{\mathrm{dTr}} C_{\mathrm{bdry}}^{1/2} \left(\frac{h_\sigma^{-3} \|U^\perp\|_{L^2(\sigma)}^2}{2} + \left\{ h_\sigma^{-3} \|U\|_{L^2(\sigma)}^2 + h_\sigma^{-1} \left\| \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma} \right\|_{L^2(\sigma)}^2 \right\} \right),
\end{aligned} \tag{5.182}$$

where in the last inequality we used $(a+b)^2 \leq 2(a^2+b^2)$. We arrive at

$$\begin{aligned}\sum_{\sigma\in\mathcal{G}_P} \left\|\tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)} \|U^\perp\|_{L^2(\sigma)} \\ &\leq c_{\mathrm{Inv}} c_{\mathrm{dTr}} C_{\mathrm{bdry}}^{1/2} \sum_{\sigma\in\mathcal{G}_P}\left(\tfrac{3}{2}h_\sigma^{-3}\|U\|^2_{L^2(\sigma)} + h_\sigma^{-1}\left\|\tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\right\|^2_{L^2(\sigma)}\right), \qquad (5.184)\\ &\leq \tfrac{3}{2} c_{\mathrm{Inv}} c_{\mathrm{dTr}} C_{\mathrm{bdry}}^{1/2} |U|_P^2.\end{aligned}$$

Similarly, we have

$$\sum_{\sigma\in\mathcal{G}_P} \|\Delta U^\perp\|_{L^2(\sigma)} \left\|\tfrac{\partial U^\perp}{\partial \boldsymbol{n}_\sigma}\right\|_{L^2(\sigma)} \leq \tfrac{3}{2} c_{\mathrm{dTr}} C_{\mathrm{bdry}}^{1/2} |U|_P^2. \qquad (5.185)$$

As a consequence of above inequalities, we obtain

$$2\mathrm{Re}^{-1}\int_\Gamma \tfrac{\partial \Delta U^\perp}{\partial \boldsymbol{n}} U^\perp - 2\mathrm{Re}^{-1}\int_\Gamma \Delta U^\perp \tfrac{\partial U^\perp}{\partial \boldsymbol{n}} \leq 3\mathrm{Re}^{-1} c_{\mathrm{Inv}}(c_{\mathrm{dTr}}+1) C_{\mathrm{bdry}}^{1/2} |U|_P^2. \qquad (5.187)$$

Finally, the nonlinear boundary integral term in (5.160) is estimated as

$$\begin{aligned}\left|\int_\Gamma \Delta U(\nabla^\perp U\cdot\boldsymbol{n})U^\perp\right| &\leq \sum_{\sigma\in\mathcal{G}_P} h_\sigma^{3/2}\|\Delta U \nabla^\perp U\cdot\boldsymbol{n}\|_{L^2(\sigma)} h_\sigma^{-3/2}\|U^\perp\|_{L^2(\sigma)},\\ &\leq \left(\sum_{\sigma\in\mathcal{G}_P} h_\sigma^3 \|\Delta U\nabla^\perp U\cdot\boldsymbol{n}\|^2_{L^2(\sigma)}\right)^{1/2} \|U^\perp\|_{3/2,P} \qquad (5.189)\\ &\leq C_N \|U^\perp\|^2_{3/2,P}.\end{aligned}$$

by applying (5.146) to the last inequality. Substituting (5.165), (5.176), (5.187) and (5.189) into (5.154) yields

$$\gamma_1\|U\|^2_{3/2,P} + \gamma_2\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|^2_{1/2,P} \leq \tfrac{1}{2}C_{\mathrm{coer}}^{-1}\eta_P^2(\omega_P^\Gamma) + C_{\mathrm{coer}}^{-1}C_\Gamma|U|_P^2, \qquad (5.191)$$

where C_Γ is defined above. Finally, we have

$$\left(\gamma_1 - C_{\mathrm{coer}}^{-1}C_\Gamma\right)\|U\|^2_{3/2,P} + \left(\gamma_2 - C_{\mathrm{coer}}^{-1}C_\Gamma\right)\left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|^2_{1/2,P} \leq \tfrac{1}{2}C_{\mathrm{coer}}^{-1}\eta_P^2(\omega_P^\Gamma), \qquad (5.192)$$

which proves (5.147). ■

Using above lemmas, we finally estimate the residual $\|\mathscr{R}_P\|_{\mathbb{V}_P'}$ in the following theorem.

Theorem 5.16. *Let u and U be the solutions to* (5.1) *and* (5.10)*, respectively. Then,*

$$\|\mathscr{R}_P\|_{\mathbb{V}_P'} \leq D_1 \eta_P(\Omega) + D_2 \eta_P(\omega_P^\Gamma), \tag{5.194}$$

with

$$D_1 = c_E c_{\text{shape}}^2 \quad \text{and} \quad D_2 = \frac{c_E C_{\text{coer}}^{-1/2} \max\{\gamma_1 + C_1, \gamma_2 + C_2\}}{\sqrt{\min\{\gamma_1 - C_{\text{coer}}^{-1} C_\Gamma, \gamma_2 - C_{\text{coer}}^{-1} C_\Gamma\}}}, \tag{5.195}$$

where $C_1 = c_{\text{Inv}} c_{\text{dTr}} c_\Pi + c_{\text{shape}}^2 C_N$ *and* $C_2 = c_{\text{dTr}} c_\Pi$.

Proof. In view of Lemmata 5.13 and 5.14, we write

$$\begin{aligned} |\langle \mathscr{R}_P, v\rangle| \leq{}& c_{\text{shape}}^2 \eta_P(\Omega) |v|_{H^2(\Omega)} + c_{\text{shape}}^2 C_\Gamma \|U\|_{3/2,P} |v|_{H^2(\omega_P^\Gamma)} \\ &+ \left((\gamma_1 + c_{\text{Inv}} c_{\text{dTr}} c_\Pi) \|U\|_{3/2,P} + (\gamma_2 + c_{\text{dTr}} c_\Pi) \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|_{1/2,P} \right) |v|_{H^2(\Omega)}, \end{aligned} \tag{5.197}$$

which, by norm equivalence (5.14), reduces to

$$\begin{aligned} |\langle \mathscr{R}_P, v\rangle| \leq{}& c_E c_{\text{shape}}^2 \eta_P(\Omega) |\!|\!| v |\!|\!|_P + c_E \Bigg((\gamma_1 + c_{\text{Inv}} c_{\text{dTr}} c_\Pi + c_{\text{shape}}^2 C_\Gamma) \|U\|_{3/2,P} \\ &+ (\gamma_2 + c_{\text{dTr}} c_\Pi) \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|_{1/2,P} \Bigg) |\!|\!| v |\!|\!|_P. \end{aligned} \tag{5.199}$$

Let $C_1 = c_{\text{Inv}} c_{\text{dTr}} c_\Pi + c_{\text{shape}}^2 C_\Gamma$ and $C_2 = c_{\text{dTr}} c_\Pi$. As the final step, applying Lemma 5.15 results in

$$(\gamma_1 + C_1) \|U\|_{3/2,P} + (\gamma_2 + C_2) \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|_{1/2,P} \leq \frac{C_{\text{coer}}^{-1/2} \max\{\gamma_1 + C_1, \gamma_2 + C_2\}}{\sqrt{\min\{\gamma_1 - C_{\text{coer}}^{-1} C_\Gamma, \gamma_2 - C_{\text{coer}}^{-1} C_\Gamma\}}} \eta_P(\omega_P^\Gamma). \tag{5.201}$$

■

Remark 5.17. The estimator $\eta_P(\omega_P^\Gamma)$ is always smaller than $\eta_P(\Omega)$ by the inclusion $\omega_P^\Gamma \subset \Omega$, so it is sufficient to reduce $\eta_P(\Omega)$ only.

Finally, we relate the estimator to the numerical error.

Corollary 5.18 (Upper Bound). *Let u and U be the solutions to* (5.1) *and* (5.10)*, respectively. Then,*

$$|\!|\!| u - U |\!|\!|_P \leq C_{\mathscr{U}} \eta_P(\Omega) + 2 c_\Pi C_U \mathrm{Re}^{-1} |h_P^s \Delta u|_{H^s(D_P^\Gamma)}, \tag{5.202}$$

with $C_{\mathscr{U}} = C_U \max\{D_1, D_2\}$.

Proof. The proof is immediate from Lemma 5.6, Lemma 5.11, and Theorem 5.16. ■

Corollary 5.18 shows that the cell-wise indicators (5.88) summed over all cells provide a computable estimate for the global numerical error. The only issue we can encounter comes from the term

$$|h_P^s \Delta u|_{H^s(D_P^\Gamma)} = \left(\sum_{\sigma \in \mathcal{G}_P} h_\sigma^{2s} |\Delta u|^2_{H^s(\tau_\sigma)} \right)^{1/2}, \tag{5.203}$$

as it is not dominated by a computable estimation. On one hand, we expect that on any cells τ_σ adjacent to a boundary edge σ situated away from problematic corners or boundary layers, the solution u will be smooth and the power s in the decay factor h_σ^s is high, and thus diminishing the contribution of $|\Delta u|_{H^s(\tau_\sigma)}$. On the other hand, in regions where singularities might occur, the error indicator will force deeper refinement making the area boundary-adjacent cells τ_σ small which will reduce the contribution of $|\Delta u|_{H^s(\tau_\sigma)}$ even though the effect of decay factor h_σ^s diminishes.

We conclude this section by quantifying the rates at which the inconsistent part of the discrete solution decays. Let $\mathbb{X}_P^0 = \mathbb{X}_P \cap H_0^2(\Omega)$. We characterize an orthogonal complement $\mathbb{X}_P^\perp$ to $\mathbb{X}_P^0$ using a projection operator $\pi_P^0 : \mathbb{X}_P \to \mathbb{X}_P^0$ defined by the linear problem

$$\pi_P^0 V \in \mathbb{X}_P^0 \quad \text{such that} \quad L_P(w^0, V - \pi_P^0 V) = 0 \quad \forall w^0 \in \mathbb{X}_P^0. \tag{5.204}$$

The problem (5.204) is well-posed by the virtue of the Lax-Milgram Lemma. Indeed, let $B : \mathbb{X}_P^0 \times \mathbb{X}_P^0 \to \mathbb{R}$ be given by $B(V^0, W^0) = L_P(W^0, V)$ and for each $V \in \mathbb{X}_P$, define a linear functional l on the subspace $\mathbb{X}_P^0$ by $l(W^0) = L_P(W^0, V)$. The bilinear form is bounded on coercive on any subspace of $H_0^2(\Omega)$ and therefore the linear problem

$$V^0 \in \mathbb{X}_P^0 \quad \text{such that} \quad B(V^0, W^0) = l(W^0) \quad \forall W^0 \in \mathbb{X}_P^0, \tag{5.205}$$

admites a unique solution. By setting $\pi_P^\perp V = V - \pi_P^0 V$ for any $V \in \mathbb{X}_P$, we obtain the following decomposition for every finite-element spline

$$V = \pi_P^0 V + \pi_P^\perp V \in \mathbb{X}_P^0 \oplus \mathbb{X}_P^\perp \equiv \mathbb{X}_P, \tag{5.206}$$

with $L_P(V^0, w^\perp) = 0$ for every pair V and w. We will write $V^0 = \pi_P^0 V$ and $V^\perp = \pi_P^\perp$.

Lemma 5.19. *If U is a solution to* (5.10),

$$|\!|\!| U^\perp |\!|\!|_P \le \frac{C_\perp}{C_{\text{coer}} - C_{\mathcal{N}} C_{\text{coer}}^{-1} c_E \|f\|_{L^2(\Omega)}} \left(\inf_{V \in \mathbb{X}_P^0} |u - V|_{H^2(\Omega)} + |\!|\!| \mathscr{E}_P |\!|\!|_{\mathbb{V}_P'} \right), \tag{5.207}$$

where $C_\perp = \frac{C_{\text{cont}}}{C_{\text{coer}}} \max\left\{1, C_{\text{cont}} + 2C_{\mathcal{N}} C_{\text{coer}}^{-1} c_E \|f\|_{L^2(\Omega)}\right\}$.

Proof. Let $V^0 \in \mathbb{X}_P^0$. By the orthogonal decomposition (5.206), we write

$$\begin{aligned} C_{\text{coer}} \| U^\perp \|_P^2 &\leq L_P(U^\perp, U^\perp) = L_P(U^\perp + U^0 - V^0, U^\perp), \\ &= L_P(U - V^0, U^\perp) \leq C_{\text{cont}} \| U - V^0 \|_P \| U^\perp \|_P. \end{aligned} \tag{5.209}$$

To estimate $\| U - V^0 \|_P$, the calculation is done in the same manner as carried in *a priori* estimation so we omit the calculation and give the final form:

$$C_{\text{coer}} \| U^\perp \|_P \leq \frac{C_{\text{cont}} \max\left\{1, C_{\text{cont}} + 2C_{\mathcal{N}} C_{\text{coer}}^{-1} c_E \|f\|_{L^2(\Omega)}\right\}}{C_{\text{coer}} - C_{\mathcal{N}} C_{\text{coer}}^{-1} c_E \|f\|_{L^2(\Omega)}} \left(|u - V^0|_{H^2(\Omega)} + \|\mathscr{E}_P\|_{\mathbb{V}_P'} \right). \tag{5.210}$$

■

Corollary 5.20. *Let $u \in H^2(\Omega)$ be a solution to* (5.6)*. If $\Delta u \in H^s(\Omega)$ for any $s > 0$,*

$$\| U^\perp \|_P \leq \frac{2C_\perp \max\{c_{\text{shape}}, c_\Pi \text{Re}^{-1}\} c_{\text{shape}}}{C_{\text{coer}} - C_{\mathcal{N}} \text{Re}\text{Ro}^{-1} c_E \|f\|_{L^2(\Omega)}} |h_\Omega^s \Delta u|_{H^s(\Omega)}. \tag{5.211}$$

Proof. We have at our disposal a quasi-interpolant projection from $H_0^2(\Omega)$ into $\mathbb{X}_P^0$ which will estimate

$$\begin{aligned} \inf_{V \in \mathbb{X}_P^0} |u - V|_{H^2(\Omega)} &= \inf_{V \in \mathbb{X}_P^0} \left(\sum_{\tau \in P} \|\Delta(u - V)\|_{L^2(\tau)}^2 \right)^{1/2}, \\ &\leq \left(\sum_{\tau \in P} c_{\text{shape}}^2 h_\tau^{2s} \|\Delta u\|_{H^s(\omega_\tau)}^2 \right)^{1/2} = c_{\text{shape}} |h_P^s \Delta u|_{H^s(\Omega)}. \end{aligned} \tag{5.213}$$

With Lemma 5.6 we arrive from Lemma 5.19,

$$\| U^\perp \|_P \leq \frac{C_\perp}{C_{\text{coer}} - C_{\mathcal{N}} C_{\text{coer}}^{-1} c_E \|f\|_{L^2(\Omega)}} \left(c_{\text{shape}} |h_P^s \Delta u|_{H^s(\Omega)} + c_\Pi \text{Re}^{-1} \left| h_P^s \Delta u \right|_{H^s(D_P^\Gamma)} \right). \tag{5.214}$$

■

Remark 5.21. As a result, if h is the size of largest edge σ on boundary Γ, $\|U^\perp\|_{L^2(\Gamma)} \preceq \gamma_1^{-1} h^{3/2+s}$ and $\left\| \frac{\partial U^\perp}{\partial \boldsymbol{n}} \right\|_{L^2(\Gamma)} \preceq \gamma_2^{-1} h^{1/2+s}$ for some $s > 0$.

5.5 Numerical study

To verify adaptivity of the weak formulation (5.10), we provide convergence studies using cubic B-splines. For this purpose, we define errors $||e||_{L^2}$, $||e||_{H^1}$, and $||e||_{H^2}$ in the L^2-norm, the H^1-semi norm, and the H^2-semi norm by

$$||e||_{L^2} = \frac{||u - U||_{L^2}}{||u||_{L^2}}, \quad ||e||_{H^1} = \frac{|u - U|_{H^1}}{|u|_{H^1}}, \quad ||e||_{H^2} = \frac{|u - U|_{H^2}}{|u|_{H^2}}, \tag{5.215}$$

respectively, where U is the approximation of u. Notice that the optimal convergence rates in the L^2-, H^1-, and H^2-norms are quartic, cubic, and quadratic rates, respectively, for the finite-element discretization using cubic B-splines. Several benchmark problems are used for the test of the present adaptive refinement algorithm. Unless otherwise specified, we take the Reynolds number of $\mathrm{Re} = 1.667$ and the Rossby number of $\mathrm{Ro} = 10^{-4}$. A Newton-Raphson iteration solver is used as a nonlinear solver.

For adaptive refinement, we use a residual-based error indicator η_P in (5.88) based on the Dörlfer marking criterion. In other words, the refinement consists of two steps, the first step is to obtain local refined meshes by splitting the marked elements (or cells) $\mathscr{M}_P$ into four succeeding cells to produce a new mesh satisfying geometric constraints. The second step is to refine the spline basis via B-spline subdivision wherever mesh refinement took place. All of our simulations are performed using $\theta = 0.9$. This choice is based on our previous study (Chapter 4) on the influence of θ of Dörfler marking strategy on the accuracy of the solution.

5.5.1 Convergence study in a rectangular domain

The accuracy and robustness of the weak formulation (5.10) and the convergence of the Newton-Raphson iteration solver are sensitive to the stabilization parameters as shown in Kim et al. [36]. As a result, it is important to check if the stabilization parameters derived from coercivity analysis in Lemma 5.7 give rise to the accurate and robust numerical results. To do so, the convergence studies for benchmark problems are performed by taking the stabilization parameters of $\gamma_1 = \frac{\mathrm{Ro}^{-1}}{2} + 2(c_{\mathrm{Inv}} c_{\mathrm{dTr}} c_{\Pi})^2 \mathrm{Re}^{-1} + 1/2$ and $\gamma_2 = (c_{\mathrm{dTr}} c_{\Pi})^2 \mathrm{Re}^{-1} + 1/2$ from (5.62).

We consider a rectangular ocean as a computational domain as shown in Figure 5.1. With the origin of a Cartesian coordinate system at the southwest corner, the x- and y-axis point eastward and northward, respectively, and the boundaries of the computational domain are the shores of the ocean. We choose the following two problems that were frequently used to test a finite-element algorithm for large scale ocean circulation problems [36, 37]. These two problems have exact solutions

$$u(x, y) = \sin^2(\pi x/3) \sin^2(\pi y) \quad \text{in } \Omega = [0, 3] \times [0, 1] \tag{5.216}$$

and

$$u(x, y) = [(1 - x/3)(1 - e^{-20x}) \sin^2(\pi y)]^2 \quad \text{in } \Omega = [0, 3] \times [0, 1]. \tag{5.217}$$

The forcing term f in the SQGE (5.1) is chosen to match those determined by the above exact solutions. In Figure 5.1, the plot of the numerical solution of (5.216) is provided. As shown in the solution, this test problem does not have any western boundary layer. Figure 5.2 shows the convergence rates in the L^2-, H^1-, and H^2-norms. The optimal rates of convergence are obtained for all three norms.

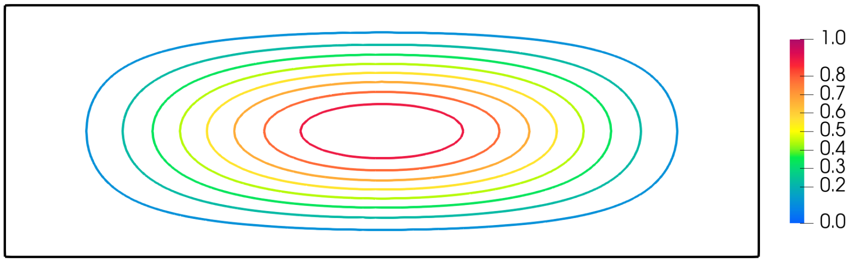


Figure 5.1: *The streamfunction for the test problem* (5.216) *without a western boundary layer.*

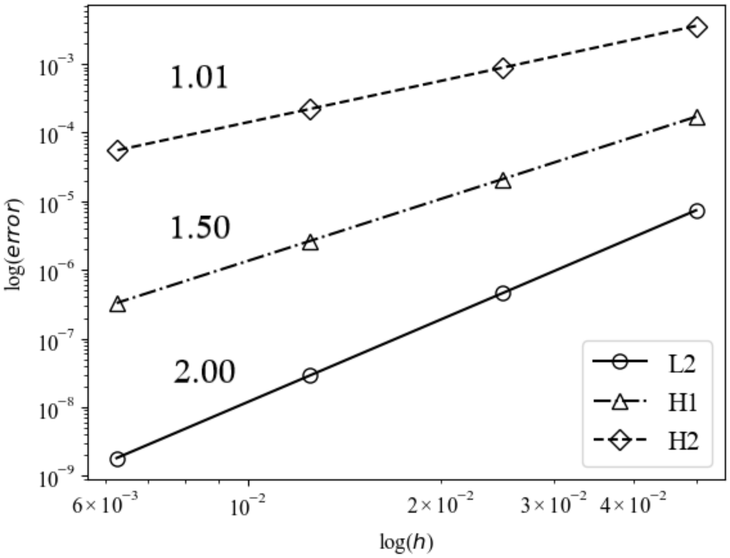


Figure 5.2: *Convergence study for the test problem* (5.216) *without a western boundary layer.*

Figure 5.3 displays the numerical solution of (5.217), with a thin boundary layer, in the vicinity of $x = 0$, corresponding to a western boundary layer. Figure 5.4 shows the convergence rates in the L^2-, H^1-, and H^2-norms. While the convergence rate in the H^1-norm is optimal, the rates of convergence for the L^2- and H^2-norms are slightly suboptimal due to the presence of the western boundary layer.

These results from the two example problems indicate that the robust and accurate numerical solutions can be obtained by choosing the stabilization parameters from (5.62) for the large-scale wind driven ocean circulation.

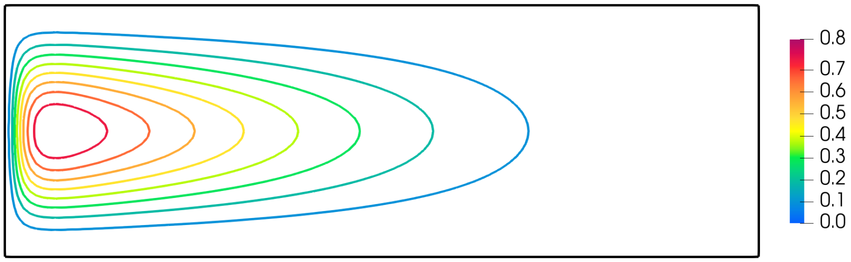


Figure 5.3: *The streamfunction for the test problem* (5.217) *with a western boundary layer.*

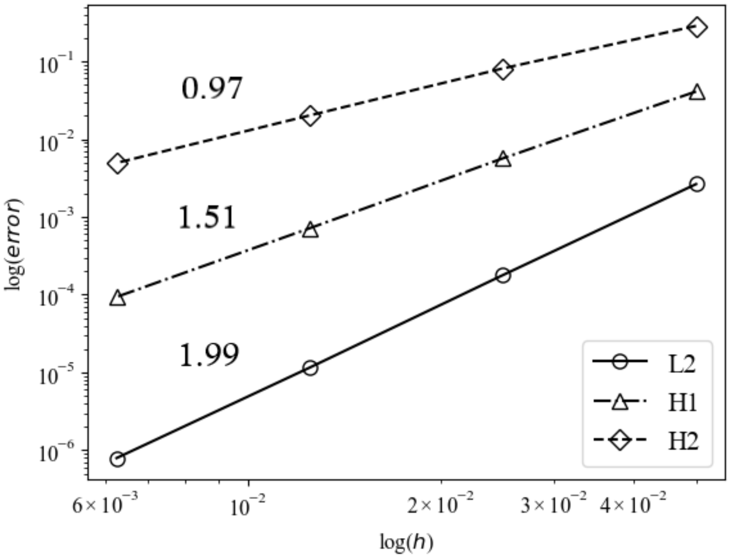
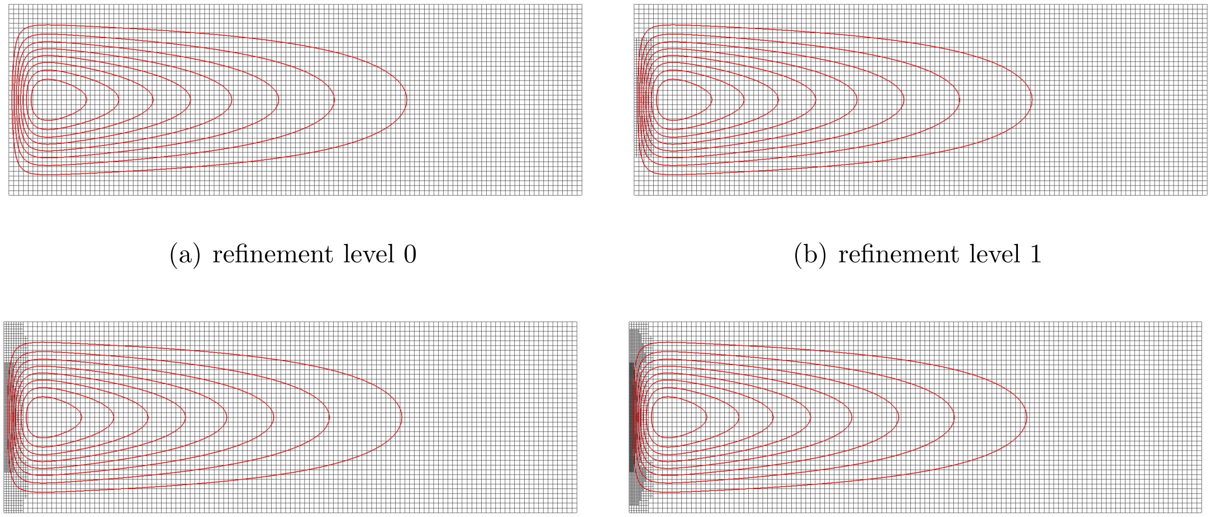


Figure 5.4: *Convergence study for the test problem* (5.217) *with a western boundary layer.*

5.5.2 Adaptivity on rectangular geometry

The performance of the adaptive refinement algorithm using the *a posteriori* error indicator (5.88) is investigated for the test problem (5.217) on a rectangular domain with a strong western boundary layer.

In Figure 5.5, we provide four refinement levels that were obtained using the error indicator (5.88). The results clearly show highly refined mesh in the vicinity of the western boundary layer, indicating the efficiency of the proposed adaptive algorithm. In Figure 5.6, we display the convergence rates for the adaptive refinement along with those for the uniform refinement. The rates of convergence significantly increase by adding the first two levels of refinement although the rates gradually reduce to reach the optimal rates by adding more refinement levels. Moreover, in spite of the presence of the western boundary layer, the solutions with the adaptive refinement are much higher accurate than those with the uniform refinement in all three norms.


(a) refinement level 0

(b) refinement level 1

(c) refinement level 2

(d) refinement level 3

Figure 5.5: *Adaptive refinement levels for rectangular geometry.*

5.5.3 Adaptivity on L-shape geometry

The adaptive refinement algorithm is further examined for the ocean circulation on L-shape geometry which is more suitable for the test of adaptivity. This problem was used for the

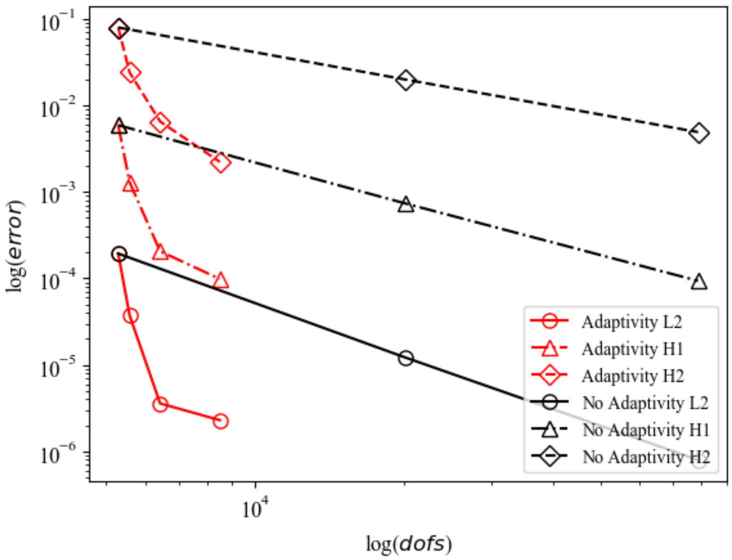
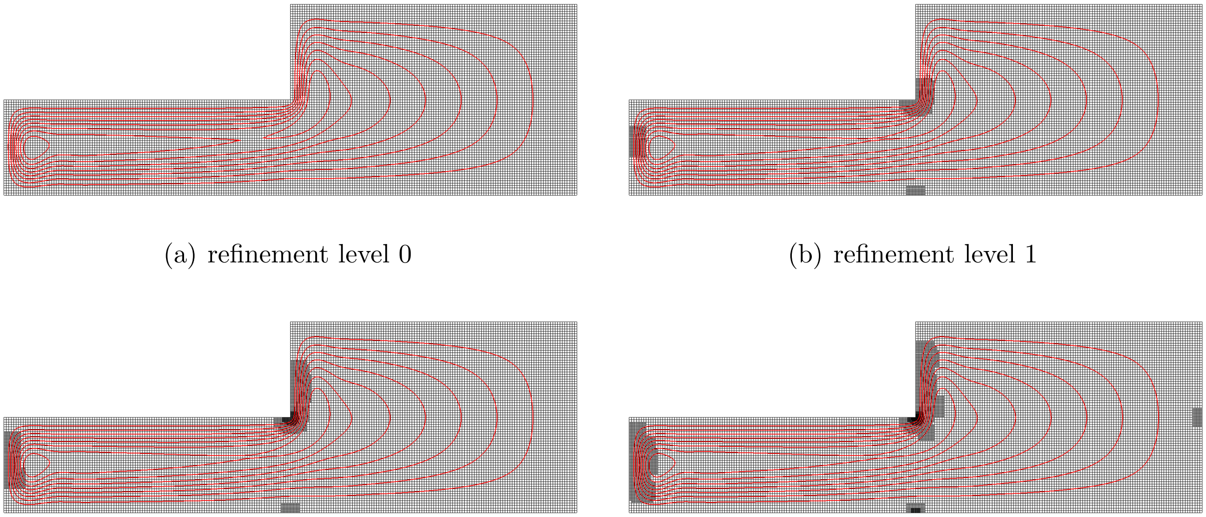


Figure 5.6: *Convergence study for adaptivity on rectangular geometry.*

verification of adaptivity of the linear Stommel–Munk model in Chapter 4. The wind forcing term $f = \sin(y)$ is chosen from the derivative of the wind stress [77]. Due to the complexity of geometry, an analytical solution does not exist for this test problem. As a consequence, the numerical solution obtained from a sufficiently fine grid with 562,500 elements is used as a standard solution for convergence study. The fine grid has a sufficiently large number of elements that the *a posteriori* error estimator decays with asymptotic regime rate.

In Figure 5.7, adaptive refinement meshes obtained using the *a posterior* error indicator (5.88) are displayed along with the numerical solutions with the strong western boundary layers. Importantly, the plots show locally refined meshes in the vicinity of the reentrant corner as well as the western boundary layers, indicating the efficiency of the *a posteriori* error indicator. Figure 5.8 shows the rates of convergence for the adaptive refinement along with those of the uniform refinement. With the uniform refinement, the significant reduction of convergence rates is obtained for all three norms, due to the presence of the western boundary layers and the rectangular corner in L-shape geometry. Importantly, the rates of convergence significantly increase with the adaptive refinement. Moreover, the solution from the adaptive refinement is much more accurate at lower resolution than that from the uniform refinement, verifying the accuracy of our algorithm.



(a) refinement level 0

(b) refinement level 1

(c) refinement level 2

(d) refinement level 3

Figure 5.7: *Adaptive refinement levels for the L-shape geometry.*

5.5.4 Computational efficiency of the adaptive algorithm

In this section, we examine the efficiency of the proposed adaptive algorithm. All calculations for this example are carried out on a workstation with Xeon E5 v3 2637 3.5 GHz CPU and 64GB of memory. A Newton–Raphson iteration method is used as a nonlinear solver. For each nonlinear iteration, a direct LU solver is employed to obtain the solutions of the linear algebraic system of equations. It typically takes four or five nonlinear iterations to satisfy our convergence criterion. In Figures 5.9(a) and 5.9(b), we provide the CPU time versus all three norms of the error for the rectangular geometry and the L-shape geometry, respectively. The computational time is greatly reduced by locally refining the mesh. This study shows that our proposed adaptive mesh algorithm gives rises to accurate results and significant computational savings compared to uniform mesh approaches.

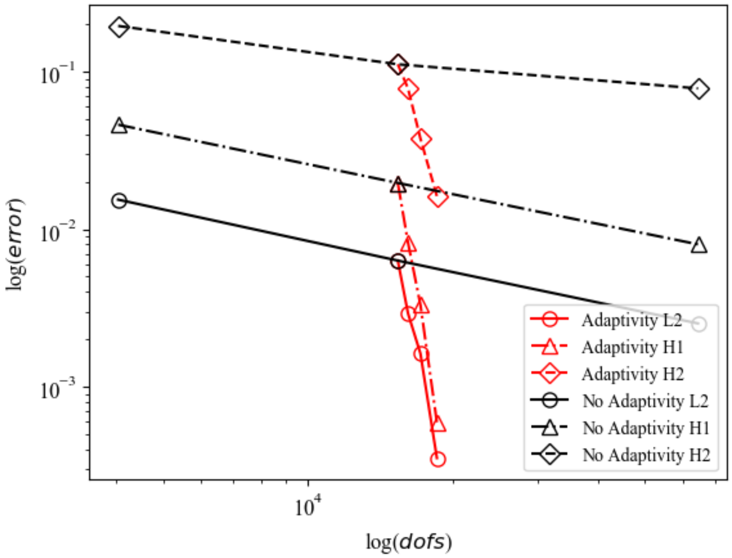
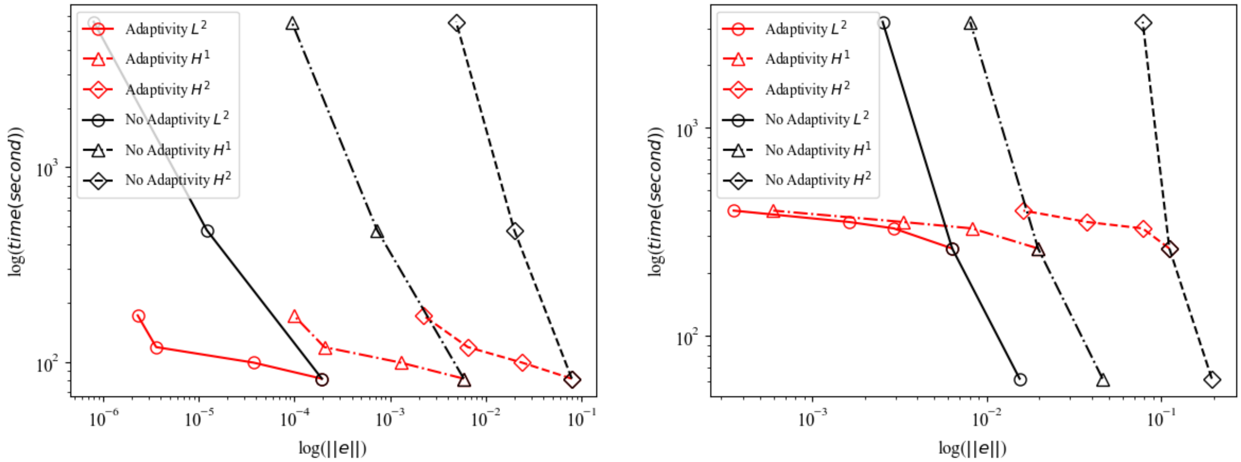


Figure 5.8: *Convergence study for adaptivity on L-shape geometry.*



(a) rectangular geometry example

(b) L-shape geometry example

Figure 5.9: *Computation time versus error in the L^2-norm, the H^1-norm, and the H^2-norm*

Chapter 6

A quasi-optimal adaptive spline-based finite element method for the bi-Laplace operator using Nitsche's method

6.1 Introduction

In this final chapter we study the performance of (2.132) on Nitsche's method for the bi-Laplace operator. The numerical error will be measured in energy norm (5.4), and the L^2-projection of Lemma 2.43 will be used again. The inconsistency term (5.35) and the consequences of Lemmata 5.5 and 5.6 can be derived with facility. The coercivity of the bilinear form in the energy norm and an *a priori* error estimate is obtained in essentially the same manner in Lemma 5.9 and we proceed to *a posteriori* error estimation without re-mention.

In the derivation of the global upper bound the same phenomenon of having to control the boundary terms $|U^{\perp}|_P$ appearing in Chapters 4 and 5 appears here as well. The global lower bound of Lemma 3.6 and estimator error reduction Lemma 3.9 carries over immediately; this is because those estimates are focused on the domain interior and do not take the boundary terms into account.

When analyzing the convergence of (2.132), the approach of the previous chapter will not carry forward immediately. The powers of the stabilization parameters in Lemma 5.13 are too high and will interfere with the convergence analysis. In addition, due to the formulation inconsistency and the dependence of the bilinear form on the mesh, obtaining a good Galerkin Pythagoras identity (as in Lemma 3.11) is not possible. Instead, in Lemma 6.8 we derive a weaker estimate that fulfills the purpose of comparing the numerical error of consecutive solutions.

Let Ω be a bounded domain in $\mathbb{R}^2$ with polygonal boundary Γ. For a source function $f \in L^2(\Omega)$ we consider the following homogenous Dirichlet boundary-valued problem

$$\begin{aligned} \Delta^2 u = f \qquad & \text{in } \Omega, \\ u = 0, \quad \tfrac{\partial u}{\partial \boldsymbol{n}} = 0 \qquad & \text{on } \Gamma. \end{aligned} \tag{6.1}$$

We consider the bilinear form $a_P : \mathbb{V}_P(\Omega) \times \mathbb{V}_P(\Omega) \to \mathbb{R}$

$$\begin{aligned} a_P(u, v) = \int_\Omega \Delta u \Delta v - \int_\Gamma \left(\Pi_P(\Delta u) \tfrac{\partial v}{\partial \boldsymbol{n}} + \tfrac{\partial u}{\partial \boldsymbol{n}} \Pi_P(\Delta v) \right) + \gamma_1 \int_\Gamma h_P^{-3} uv \\ + \int_\Gamma \left(\tfrac{\partial \Pi_P(\Delta u)}{\partial \boldsymbol{n}} v + u \tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}} \right) + \gamma_2 \int_\Gamma h_P^{-1} \tfrac{\partial u}{\partial \boldsymbol{n}} \tfrac{\partial v}{\partial \boldsymbol{n}}. \end{aligned} \tag{6.3}$$

We measure the numerical error using the mesh-dependent norm

$$|\!|\!| u |\!|\!|_P^2 = \|\Delta u\|_{L^2(\Omega)}^2 + \gamma_1 \|u\|_{3/2,P}^2 + \gamma_2 \left\| \tfrac{\partial u}{\partial \boldsymbol{n}} \right\|_{1/2,P}^2, \tag{6.4}$$

and use the abbreviation

$$|U|_P^2 = \|U\|_{3/2,P}^2 + \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|_{1/2,P}^2. \tag{6.5}$$

As previously, $\mathbb{X}_P$ is decomposed into $\mathbb{X}_P^0 = \mathbb{X}_P \cap H_0^2(\Omega)$ and $\mathbb{X}_P^\perp$ in the manner described by Lemma 4.7. The proof of Lemma 4.7 can be extended to the following result:

Lemma 6.1. *Semi-norm $|\cdot|_P$ defines a norm on $\mathbb{X}_P^\perp$. In particular, for a constant $C_\perp > 0$*

$$|\!|\!| V^\perp |\!|\!|_P \le C_\perp |V^\perp|_P \quad \forall V^\perp \in \mathbb{X}_P^\perp. \tag{6.6}$$

Proof. Let $\mathcal{D}_\Gamma = \overline{\bigcup_{\sigma \in \mathcal{G}_P} \omega_\sigma}$. If $|V^\perp|_P = 0$ then $V^\perp = \frac{\partial V^\perp}{\partial \boldsymbol{n}} \equiv 0$ on $\mathcal{D}_\Gamma$ due to the finite-dimensionality of polynomial space $\mathbb{X}_P^\perp$. Necessarily we have $V^\perp \equiv 0$ everywhere; otherwise $V^\perp \in \mathbb{X}_P^0$. ■

We have partial consistency virtue of an argument analogous to Lemma 5.5

$$a_P(U, V^0) = \ell_f(V^0) \quad \forall V^0 \in \mathbb{X}_P^0, \tag{6.7}$$

which gives Partial Galerkin Orthogonality:

$$a_P(u - U, V^0) = 0. \tag{6.8}$$

Remark 6.2. It is possible to obtain (6.8) with weaker regularity assumptions on u, as done in Bonito et al. [63], using decomposition (5.206).

6.2 *A posteriori* error estimation

We derive a sharper Upper Bound than given in Chapters 4—6. This genuine improvement was initially recognized by Bonito et al. [63] when applied to adaptive discontinuous finite-element methods.

Lemma 6.3 (Estimator reliability). *Let P be an admissible partition of Ω. The module* **ESTIMATE** *produces* a posteriori *error estimate η_P for the discrete error such that*

$$a_P(u-U,u-U) \leq C_{U,1}\eta_P^2(U,\Omega) + C_{U,2}\left(\gamma_1 \|U\|_{3/2,P}^2 + \gamma_2 \left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,P}^2\right), \tag{6.10}$$

with constants $C_{U,1}, C_{U,2} > 0$ depending only on c_{shape}.

Proof. Let $e = u - U$ and let $v = u - U^0$ and we may write $e = v - U^\perp$. Since $I_P v \in \mathbb{X}_P^0$, Partial Galerkin orthogonality (6.8) implies $a_P(e, I_P v) = 0$ and we have

$$a_P(e,e) = a_P(e, v - I_P v) - a_P(e, U^\perp). \tag{6.11}$$

The treatment of the term $a_P(e, v - I_P v)$ is essentially the same as in Chapter 5 so we omit a detailed calculation:

$$\begin{aligned} |a_P(e, v - I_P v)| \leq c_\Pi \Bigg\{ &\left(\sum_{\tau \in P} h_\tau^4 \|R_\tau\|_{L^2(\tau)}^2\right)^{1/2} + \left(\sum_{\sigma \in \mathcal{E}_P} h_\sigma^3 \left\|J_{\sigma,1}\right\|_{L^2(\sigma)}^2\right)^{1/2} \\ &+ \left(\sum_{\sigma \in \mathcal{E}_P} h_\sigma \|J_{\sigma,2}\|_{L^2(\sigma)}^2\right)^{1/2} \Bigg\} |v|_{H^2(\Omega)} + C_1 |U^\perp|_P |v|_{H^2(\Omega)}, \end{aligned} \tag{6.13}$$

where $C_1 = c_\Pi c_{\mathrm{dTr}} \max\{c_{\mathrm{Inv}}, 1\}$. To control the inconsistency term $a_P(e, U^\perp)$, we employ Young's inequality and the norm equivalence from Lemma 6.1,

$$\begin{aligned} a_P(e, U^\perp) &\leq C_{\mathrm{cont}} \|e\|_P \|U^\perp\|_P \leq \tfrac{C_{\mathrm{cont}}}{C_{\mathrm{coer}}^{1/2}} a_P(e,e)^{1/2} \|U^\perp\|_P, \\ &\leq \tfrac{a_P(e,e)}{4} + \tfrac{C_{\mathrm{cont}}^2}{C_{\mathrm{coer}}} \|U^\perp\|_P^2 \leq \frac{a_P(e,e)}{4} + \tfrac{C_{\mathrm{cont}}^2}{C_{\mathrm{coer}}} C_\perp^2 |U^\perp|_P^2. \end{aligned} \tag{6.15}$$

Let $C_2 = \frac{C_{\mathrm{cont}}^2}{C_{\mathrm{coer}}} C_\perp^2$. Since $v = e + U^\perp$,

$$\begin{aligned} |v|_{H^2(\Omega)}^2 &\leq C_{\mathrm{coer}}^{-1} a_P(e + U^\perp, e + U^\perp), \\ &= C_{\mathrm{coer}}^{-1} \left(a_P(e,e) + 2a_P(e,U^\perp) + a_P(U^\perp, U^\perp)\right), \\ &\leq C_{\mathrm{coer}}^{-1} \left(2a_P(e,e) + (1 + 2C_2)\,|U^\perp|_P^2\right). \end{aligned} \tag{6.17}$$

Let $C_3^2 = C_{\text{coer}}^{-1} \max\{2, (1+2C_2)\}$. Summing up, applying Young's inequality with $\delta = 1/2$,

$$\begin{aligned} \tfrac{3}{4} a_P(e,e) &\leq c_\Pi \left(\eta_P(\Omega) + C_1 |U^\perp|_P\right) |v|_{H^2(\Omega)} + C_2 |U^\perp|_P^2, \\ &\leq c_\Pi C_3 \left(\eta_P(\Omega) + C_1 |U^\perp|_P\right) \left(a_P(e,e) + |U^\perp|_P^2\right)^{1/2} + C_2 |U^\perp|_P^2, \\ &\leq C_3 \left(\eta_P(\Omega) + C_1|U^\perp|_P\right)^2 + \tfrac{1}{4}\left(a_P(e,e) + |U^\perp|_P^2\right) + C_2 |U^\perp|_P^2, \end{aligned} \tag{6.19}$$

which makes for constants $C_{U,1} > 0$ and $C_{U,2} > 0$ depending on C_1, C_2 and C_3,

$$\tfrac{1}{2} a_P(e,e) \leq \tfrac{C_{U,1}}{2} \eta_P(\Omega) + \tfrac{C_{U,2}}{2} |U^\perp|_P^2. \tag{6.21}$$

■

In the following Lemma we show a local version of Lemma 6.3. While the result is not needed for convergence, it is required for quasi-optimality.

Lemma 6.4 (Estimator discrete reliability). *Let P be an admissible partition of Ω and let $P_* =$ **REFINE** $[P, R]$ for some refined set $R \subseteq P$. If U and U_* are the respective discrete solutions on partitions P and P_*, then for constants $C_{dU,1}, C_{dU,2} > 0$, depending only on c_{shape} and c_Π,*

$$\|U_*^0 - U\|_P^2 \leq C_{dU,1} \eta_P^2(U, \omega_{R_{P\to P_*}}) + C_{dU,2} \left(\gamma_1 \|U\|_{3/2,R}^2 + \gamma_2 \left\|\tfrac{\partial U}{\partial \boldsymbol{n}}\right\|_{1/2,R}^2\right), \tag{6.23}$$

where $\omega_{R_{P\to P_*}}$ is understood as the union of support extensions of refined cells from P to obtain P_*.

Proof. In view of Partial Consistency (6.7) and the nesting of spline spaces, $a_P(U_*^0, V^0) = \ell_f(V^0)$ holds if $V^0 \in \mathbb{X}_P^0$ from which we obtain $a_P(U_*^0 - U, V^0) = 0$ for every $V^0 \in \mathbb{X}_P^0$. Let $E_*^0 = U_*^0 - U_0$ and let $E_* = U_*^0 - U \equiv E_*^0 - U^\perp$. Then for any $V_0 \in \mathbb{X}_P^0$ we write an analogous expression to (6.11)

$$a_P(E_*, E_*) = a_P(E_*, E_*^0 - U^\perp) = a_P(E_*, E_*^0 - V^0) - a_P(E_*, U^\perp), \tag{6.24}$$

which we proceed to control in terms of the estimator. For the first term, we form disconnected subdomains $\Omega_i \subseteq \Omega$, $i \in J$, each formed from the interior of connected union of cell support extensions. Set $\Omega_* = \cup_{\tau \in R_{P\to P_*}} \overline{\omega_\tau}$. Then to each subdomain Ω_i we form a partition $P_i = \{\tau \in P : \tau \subset \Omega_i\}$, interior edges $\mathcal{E}_i = \{\sigma \in \mathcal{E}_P : \sigma \subset \partial\tau,\ \tau \in P_i\}$ and boundary edges $\mathcal{G}_i = \{\sigma \in \mathcal{G}_P : \sigma \subset \partial\tau,\ \tau \in P_i\}$, and a corresponding finite-element space $\mathbb{X}_i$. Let $I_i : L^2(\Omega_i) \to \mathbb{X}_i$ satisfy the local estimates (2.86)— (2.88). Let $V^0 \in \mathbb{X}_P^0$ be an approximation of E_*^0 be given by

$$V^0 = E_*^0 \mathbb{1}_{\Omega \setminus \Omega_*} + \sum_{i \in J} (I_i^0 E_*^0) \cdot \mathbb{1}_{\Omega_i}. \tag{6.25}$$

Then $E_*^0 - V^0 \equiv 0$ on $\Omega\backslash\Omega_*$. To localize the error on $\omega_{R_{P\to P_*}}$ we use intergration by parts to express

$$\begin{aligned} a_P(E_*,E_*^0 - V^0) = \sum_{i\in J}\Bigg[&\sum_{\tau\in P_i}\langle R_\tau, E_*^0 - I_P E_*^0\rangle_\tau \\ &+ \sum_{\sigma\in\mathcal{E}_i}\big\{\langle J_{\sigma,1}, E_*^0 - I_P E_*^0\rangle_\sigma + \langle J_{\sigma,2}, E_*^0 - I_P E_*^0\rangle_\sigma\big\} \\ &+ \sum_{\sigma\in\mathcal{G}_i}\left(\int_\sigma U \tfrac{\partial}{\partial \boldsymbol{n}_\sigma}\left[\Pi_P\Delta(E_*^0 - I_P E_*^0)\right] - \int_\sigma \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\Pi_P\Delta(E_*^0 - I_P E_*^0)\right)\Bigg], \end{aligned} \tag{6.27}$$

$$\begin{aligned} \sum_{\tau\in P_i}&\langle R_\tau, E_*^0 - I_P E_*^0\rangle_\tau + \sum_{\sigma\in\mathcal{E}_i}\big\{\langle J_{\sigma,1}, E_*^0 - I_P E_*^0\rangle_\sigma + \langle J_{\sigma,2}, E_*^0 - I_P E_*^0\rangle_\sigma\big\} \\ &\le c_\Pi\bigg(\sum_{\tau\in P_i}\eta_P^2(U,\tau)\bigg)^{1/2}\bigg(\sum_{\tau\in P_i}\|E_*^0\|^2_{H^2(\omega_\tau)}\bigg)^{1/2}, \\ &\le c_\Pi c_{\text{shape}}\eta_P(U,\Omega_i)\|E_*^0\|_{H^2(\Omega_i)}. \end{aligned} \tag{6.29}$$

The boundary integral terms will be control by the inconsistent part of the spline solution

$$\begin{aligned} \sum_{\sigma\in\mathcal{G}_i}\left(\int_\sigma U \tfrac{\partial}{\partial \boldsymbol{n}_\sigma}\left[\Pi_P\Delta(E_*^0 - I_P E_*^0)\right] - \int_\sigma \tfrac{\partial U}{\partial \boldsymbol{n}_\sigma}\Pi_P\Delta(E_*^0 - I_P E_*^0)\right) \\ \le |U^\perp|_{P_i}\|E_*^0\|_{H^2(\Omega_i)}. \end{aligned} \tag{6.31}$$

Together we arrive at an estimate for the first term in (6.4)

$$a_P(E_*, E_*^0 - V^0) \le c_\Pi c_{\text{shape}}\left(\eta_P(U,\Omega_*) + C_1|U^\perp|_P\right)\|E_*^0\|_{H^2(\Omega_*)}. \tag{6.32}$$

To control the inconsistent term from (6.4), we follow the same reasoning made in (6.15) from Lemma 6.3 to get

$$a_P(E_*, U^\perp) \le \tfrac{1}{2}a_P(E_*,E_*) + \tfrac{C_2}{2}|U^\perp|_P^2, \tag{6.34}$$

where C_2 retains the same meaning as before. Noting that $E_*^0 = E_* + U^\perp$, $\|E_*^0\|_{H^2(\Omega_*)} \le \|E_*\|_{H^2(\Omega_*)} + \|U^\perp\|_{H^2(\Omega_*)}$. Invoking norm equivalence (6.6). Summing up we arrive

$$a_P(E_*,E_*) \le C_{dU,1}\eta_P^2(U,\Omega_*) + C_{dU,2}|U^\perp|_P^2. \tag{6.35}$$

■

The presence of negative powers in $|U^\perp|_P$ on the right-hand side in (6.10) and (6.23) may appear to pose a problem with decreasing mesh-size along the boundary. We now show that contributions from domain boundary integrals are dominated by the those coming from the mesh interior.

Lemma 6.5. *For sufficiently large stabilization terms γ_1 and γ_2,*

$$(\gamma_1 - C_{\mathrm{R}})\|U\|^2_{3/2,P} + (\gamma_2 - C_{\mathrm{R}}) \left\| \tfrac{\partial U}{\partial \boldsymbol{n}} \right\|^2_{1/2,P} \leq C^{-1}_{\mathrm{coer}} \eta^2_P(U,\Omega), \tag{6.36}$$

with $C_{\mathrm{R}} \preceq \frac{c_{\mathrm{shape}}}{C_{\mathrm{coer}}}$.

Corollary 6.6. *Under the assumptions of Lemma 6.3 and lemma 6.4, if* $\gamma = \min\{\gamma_1 - C_R, \gamma_2 - C_R\} > 0$ *then*

$$a_P(u-U, u-U) \leq C_U \eta^2_P(U,\Omega), \tag{6.37}$$

and

$$\|U^0_* - U\|^2_{P_*} \leq C_{dU} \eta^2_P(U, \omega_{R_{P\to P_*}}) + \gamma^{-1} C^{-1}_{\mathrm{coer}} \eta^2_P(U,\Omega). \tag{6.38}$$

6.3 Convergence

To show that procedure (2.132) exhibits convergence we must be able to relate the errors of consecutive discrete solutions. In the conforming discrete method of Chapter 3 the symmetry of the bilinear form, consistency of the formulation and finite-element spline space nesting would readily provide that via Galerkin Pythagoras. This is not the case in Nitsche's method since our formulation is no longer consistent with the weak problem (1.4).

In what follows we establish estimates that allows us to compare two spline solutions on different admissible meshes. This replaces the unavailable Galerkin Pythagoras which the conformning formulation enjoyed.

Lemma 6.7 (Mesh perturbation). *Let P and P_* be successive admissible partitions which are obtained by* **REFINE**. *Then for a constant $C_{\mathrm{comp}} > 0$, depending only on c_{shape}, we have for any $\delta > 0$*

$$a_{P_*}(v,v) \leq (1 + 4\delta C_{\mathrm{coer}}) a_P(v,v) + \frac{C_{\mathrm{comp}}}{\delta} \left(\gamma_1 \|v\|^2_{3/2,P} + \gamma_2 \left\| \tfrac{\partial v}{\partial \boldsymbol{n}} \right\|^2_{1/2,P} \right), \tag{6.39}$$

holding for every function $v \in H^2(\Omega)$.

Proof. Given any $v \in H^2(\Omega)$ we write

$$\begin{aligned} a_{P_*}(v,v) = a_P(v,v) &+ 2\bigg(\int_\Gamma \Pi_P(\Delta v) \tfrac{\partial v}{\partial \boldsymbol{n}} - \int_\Gamma \tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}} v \bigg) \\ &- 2\bigg(\int_\Gamma \Pi_{P_*}(\Delta v) \tfrac{\partial v}{\partial \boldsymbol{n}} - \int_\Gamma \tfrac{\partial \Pi_{P_*}(\Delta v)}{\partial \boldsymbol{n}} v \bigg) \\ &- \gamma_1 \left(\|v\|^2_{P,3/2} - \|v\|^2_{P_*,3/2} \right) - \gamma_2 \left(\left\| \tfrac{\partial v}{\partial \boldsymbol{n}} \right\|^2_{P,1/2} - \left\| \tfrac{\partial v}{\partial \boldsymbol{n}} \right\|^2_{P_*,1/2} \right). \end{aligned} \tag{6.41}$$

Look at the boundary integral terms depending on P. Let $\sigma \in \mathcal{G}_P$ an edge to some cell $\tau \in P$,

$$\int_\sigma \Pi_P(\Delta v)\frac{\partial v}{\partial \boldsymbol{n}_\sigma} \le \|\Pi_P(\Delta v)\|_\sigma \left\|\tfrac{\partial v}{\partial \boldsymbol{n}_\sigma}\right\|_\sigma \le c_{\mathrm{dTr}} c_\Pi h_\sigma^{-1/2}\|\Delta v\|_\tau \left\|\tfrac{\partial v}{\partial \boldsymbol{n}_\sigma}\right\|_\sigma . \tag{6.43}$$

Summing (6.43) over all $\sigma \in \mathcal{G}_P$ and an application of Schwarz's inequality on the summation would give

$$\begin{aligned}\left|\int_\Gamma \Pi_P(\Delta v)\tfrac{\partial v}{\partial \boldsymbol{n}}\right| &\preceq \Bigg(\sum_{\sigma\in\mathcal{G}_P} h_\sigma^{-1}\left\|\tfrac{\partial v}{\partial \boldsymbol{n}_\sigma}\right\|^2_\sigma\Bigg)^{1/2}\Bigg(\sum_{\tau\in P:\partial\tau\cap\Gamma\neq\varnothing}\|\Delta v\|^2_\tau\Bigg)^{1/2}, \\ &\le \left\|\tfrac{\partial v}{\partial \boldsymbol{n}}\right\|_{P,1/2}\|\Delta v\|_{L^2(\Omega)}.\end{aligned} \tag{6.45}$$

Similarly, using the inverse-estimate $\|\frac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}_\sigma}\|_\sigma \le c_{\mathrm{Inv}} h_\sigma^{-1}\|\Pi_P(\Delta v)\|_\sigma$, we obtain

$$\left|\int_\Gamma \tfrac{\partial \Pi_P(\Delta v)}{\partial \boldsymbol{n}} v\right| \le c_{\mathrm{dTr}} c_{\mathrm{Inv}} c_\Pi \|v\|_{P,3/2}\|\Delta v\|_{L^2(\Omega)}. \tag{6.46}$$

We carry the same reasoning for the remaining boundary integral. Employing Young's inequality with $\delta > 0$ we arrive at

$$\begin{aligned} a_{P_*}(v,v) \preceq a_P(v,v) &+ 4\delta\|\Delta v\|^2_{L^2(\Omega)} + \left(\tfrac{1}{\delta}+\gamma_1\right)\|v\|^2_{P,3/2} + \left(\tfrac{1}{\delta}+\gamma_1\right)\|v\|^2_{P_*,3/2} \\ &+ \left(\tfrac{1}{\delta}+\gamma_2\right)\left\|\tfrac{\partial v}{\partial \boldsymbol{n}}\right\|^2_{P,1/2} + \left(\tfrac{1}{\delta}+\gamma_2\right)\left\|\tfrac{\partial v}{\partial \boldsymbol{n}}\right\|^2_{P_*,1/2}.\end{aligned} \tag{6.48}$$

With the fact that $h_\sigma \le c_{\mathrm{shape}} h_{\sigma_*}$, with $\sigma \in \mathcal{G}_P$ and $\sigma_* \in \mathcal{G}_{P_*}$, we infer that $\|v\|_{3/2,P_*} \le c^{-1}_{\mathrm{shape}}\|v\|_{P,3/2}$ and $\|\frac{\partial v}{\partial \boldsymbol{n}}\|_{1/2,P_*} \le c^{-1}_{\mathrm{shape}}\|\frac{\partial v}{\partial \boldsymbol{n}}\|_{1/2,P}$.

$$\left(\tfrac{1}{\delta}+\gamma_1\right)\left(\|v\|^2_{P_*,3/2} + \|v\|^2_{P_*,3/2}\right) \le \tfrac{C_{\mathrm{comp}}\gamma_1}{\delta}\|v\|^2_{P,3/2}, \tag{6.49}$$

where $C_{\mathrm{comp}} > 0$ is an appropriate proportionality parameter that depends on c_{shape}. A similar argument holds for terms including boundary norms of $\frac{\partial v}{\partial \boldsymbol{n}}$. ■

Lemma 6.8 (Comparison of solutions). *Let P and P_* be successive admissible partitions obtained by* **REFINE** *and let $U \in \mathbb{X}_P$ and $U_* \in \mathbb{X}_{P_*}$ be the finite-element spline solutions. Then we have for any $\Lambda > 1$*

$$a_{P_*}(e_{P_*}, e_{P_*}) \le \Lambda a_P(e_P, e_P) - \frac{C_{\mathrm{coer}}}{2}|\!|\!|U_* - U|\!|\!|^2_{P_*} + \frac{C_{\mathrm{Comp}}}{(\Lambda-1)\gamma}\eta_P^2, \tag{6.51}$$

where $\gamma = \min\{\gamma_1 - C_R, \gamma_2 - C_R\}$.

Proof. We follow the following abbreviation. Let $e = u - U$, let $e_* = u - U_*$, let $E_*^0 = U_*^0 - U^0$, and let $E_*^\perp = U_*^\perp - U^\perp$. Partial Galerkin Orthogonality (6.8) implies

$$a_{P_*}(e_*, e_*) = a_{P_*}(e_*, e_* + E_*^0) = a_{P_*}(e_* + E_*^0, e_* + E_*^0) - a_{P_*}(E_*^0, e_* + E_*^0). \tag{6.53}$$

Again, by Partial Galerkin Orthogonality (6.8) and symmetry we have

$$a_{P_*}(e_*, e_*) = a_{P_*}(e_* + E_*^0, e_* + E_*^0) - a_{P_*}(E_*^0, E_*^0). \tag{6.54}$$

Rewriting $U_* - E_*^0 = U - E_*^\perp$ we can express $e_* + E_*^0 = e - E_*^\perp$ and therefore

$$a_{P_*}(e_* + E_*^0, e_* + E_*^0) = a_{P_*}(e, e) - 2a_{P_*}(e, E{*}^\perp) + a_{P_*}(E_*^\perp, E_*^\perp). \tag{6.55}$$

We then have

$$a_{P_*}(e_*, e_*) = a_{P_*}(e, e) - 2a_{P_*}(e, E{*}^\perp) + a_{P_*}(E_*^\perp, E_*^\perp) - a_{P_*}(E_*^0, E_*^0). \tag{6.57}$$

Employ Young's inequality

$$a_{P_*}(e, e) - 2a_{P_*}(e, E_*^\perp) \le (1 + \delta)a_{P_*}(e, e) + \tfrac{C_{\mathrm{cont}}^2}{\delta C_{\mathrm{coer}}}|||E_*^\perp|||_{P_*}^2. \tag{6.58}$$

Writing $E_*^0 = E_* - E_*^\perp$ and with $|||E_*|||_{P_*}^2 \le 2|||E_*^0|||_{P_*}^2 + 2|||E_*|||_{P_*}^2$ makes $|||E_*^0|||_{P_*}^2 \ge \frac{1}{2}|||E_*|||_{P_*}^2 - |||E_*^\perp|||_{P_*}^2$ and

$$\begin{aligned} a_{P_*}(E_*^\perp, E_*^\perp) - a_{P_*}(E_0^*, E_*^0) &\le C_{\mathrm{cont}}|||E_*^\perp|||_{P_*}^2 - C_{\mathrm{coer}}|||E_*^0|||_{P_*}^2, \\ &\le -\tfrac{C_{\mathrm{coer}}}{2}|||E_*|||_{P_*}^2 + C_4|||E_*^\perp|||_{P_*}^2, \end{aligned} \tag{6.60}$$

where $C_4 = C_{\mathrm{coer}} + C_{\mathrm{cont}}$. We therefore have, with $C_5 = \max\{C_4, \frac{C_{\mathrm{cont}}^2}{C_{\mathrm{coer}}}\}$,

$$a_{P_*}(e_*, e_*) \le (1 + \delta)a_{P_*}(e, e) - \tfrac{C_{\mathrm{coer}}}{2}|||E_*|||_{P_*}^2 + C_5\big(1 + \tfrac{1}{\delta}\big)|||E_*^\perp|||_{P_*}^2. \tag{6.62}$$

Using the fact that edge sizes between two consecutive refinement steps are comparable and (6.6)

$$|||E_*^\perp|||_{P_*}^2 \preceq |U_*^\perp|_{P_*}^2 + |U^\perp|_P^2 \preceq \tfrac{C_{\mathrm{coer}}^{-1}}{\gamma}\left(\eta_{P_*}^2(\Omega) + \eta_P^2(\Omega)\right). \tag{6.63}$$

In view of Lemma 6.7, for the same $\delta > 0$ above, and Lemma (6.36)

$$a_{P_*}(e, e) \le (1 + 4\delta C_{\mathrm{coer}})a_P(e, e) + \tfrac{C_{\mathrm{comp}}C_{\mathrm{coer}}^{-1}}{\delta\gamma}\eta_P^2(\Omega). \tag{6.65}$$

Summing up we have

$$a_{P_*}(e_*, e_*) \le (1 + C\delta)a_P(e, e) - \tfrac{C_{\mathrm{coer}}}{2}|||E_*|||_{P_*}^2 + \tfrac{C_{\mathrm{Comp}}}{\delta\gamma}\left(\eta_{P_*}^2(\Omega) + \eta_P^2(\Omega)\right), \tag{6.67}$$

■

where $C > 1$ and depends on C_{coer}. Define $\Lambda(\delta) = 1 + C\delta$.

Theorem 6.9 (Convergence of Nitsche's AFEM). *Given $f \in L^2(\Omega)$ and Dörlfer parameter $\theta \in (0,1]$, there exists $\gamma_C(\theta) > 0$, a contractive factor $\alpha \in (0,1)$ and a constant $C_{\text{est}} > 0$, such that for all $\gamma \geq \gamma_C$ the adaptive procedure* **AFEM** $[P, f, \theta]$ *will produce two successive solutions $U \in \mathbb{X}_P$ and $U_* \in \mathbb{X}_{P_*}$ for which*

$$a_{P_*}(e_{P_*}, e_{P_*}) + \tfrac{C_{\text{coer}}}{2} C_{\text{est}} \eta_{P_*}^2(U_*, \Omega) \leq \alpha \left(a_P(e_P, e_P) + \tfrac{C_{\text{coer}}}{2} C_{\text{est}} \eta_P^2(U, \Omega) \right). \tag{6.68}$$

Proof. The convergence proof for Nitsche's method is somewhat more delicate than the proof of Theorem 3.13, but the core ideas remain the same. We adopte the same abbreviates as in Theorem 3.13.

$$\begin{aligned} a_{P_*}(e_{P_*}, e_{P_*}) + \tfrac{C_{\text{coer}}}{2} C_{\text{est}} \eta_{P_*}^2 \leq & \Lambda a_P(e_P, e_P) - \tfrac{C_{\text{coer}}}{2} \| \varepsilon_P \|_{P_*}^2 + \tfrac{C_{\text{comp}}}{\gamma(\Lambda-1)} \left(\eta_{P_*}^2 + e_P^2 \right) \\ & + \tfrac{C_{\text{coer}}}{2} C_{\text{est}} \left(q_{\text{est}} \eta_P^2 + C_{\text{est}}^{-1} \| \varepsilon_P \|_{P_*}^2 \right). \end{aligned} \tag{6.70}$$

Quasi-norm factor chosen so that $\| \varepsilon_P \|_{P_*}^2$ is removed and reduces to

$$a_{P_*}(e_{P_*}, e_{P_*}) + \left(\tfrac{C_{\text{coer}}}{2} C_{\text{est}} - \tfrac{C_{\text{comp}}}{\gamma(\Lambda-1)} \right) \eta_{P_*}^2 \leq \Lambda a_P(e_P, e_P) + \left(\tfrac{C_{\text{coer}}}{2} C_{\text{est}} q_{\text{est}} + \tfrac{C_{\text{comp}}}{\gamma(\Lambda-1)} \right) \eta_P^2. \tag{6.71}$$

With $a_P(e_P, e_P) = \alpha a_P(e_P, e_P) + (1-\alpha) a_P(e_P, e_P)$ and upper bound $\gamma \geq \gamma_C$:

$$\begin{aligned} & a_{P_*}(e_{P_*}, e_{P_*}) + \tfrac{C_{\text{coer}}}{2} C_{\text{est}} \underbrace{\left(1 - \frac{C_{\text{comp}}}{\frac{C_{\text{coer}}}{2} C_{\text{est}} \gamma(\Lambda-1)} \right)}_{=:A_1} \eta_{P_*}^2 \\ & \leq \alpha \Lambda a_P(e_P, e_P) + \tfrac{C_{\text{coer}}}{2} C_{\text{est}} \underbrace{\left((1-\alpha)\Lambda \frac{C_U}{\frac{C_{\text{coer}}}{2} C_{\text{est}}} + q_{\text{est}} + \frac{C_{\text{comp}}}{\frac{C_{\text{coer}}}{2} C_{\text{est}} \gamma(\Lambda-1)} \right)}_{=:A_2} \eta_P^2. \end{aligned} \tag{6.73}$$

Making notation simpler: $c = \frac{C_U}{\frac{C_{\text{coer}}}{2} C_{\text{est}}}$. If α chosen to be

$$\alpha = \frac{q_{\text{est}} + c\Lambda}{\Lambda(1+c)} + \frac{C_{\text{comp}}/C_U}{\Lambda(1+c)} \gamma^{-1} (\Lambda - 1)^{-1}, \tag{6.74}$$

then it holds that $A_2 = \alpha \Lambda$. Let $\Lambda = 1 + \varepsilon$, with $\varepsilon \leq \frac{1-q_{\text{est}}}{2c}$, write $q_{\text{est}} = 1 - (1 - q_{\text{est}})$, then

$$\frac{q_{\text{est}} + c\Lambda}{1+c} = 1 - \frac{(1-q_{\text{est}}) - \varepsilon c}{1+c} \leq 1 - \tfrac{1}{2(1+c)}(1 - q_{\text{est}}). \tag{6.75}$$

Furthermore, make $\gamma^{-1} < \frac{C_U}{2C_{\text{comp}}}(1 - q_{\text{est}})\varepsilon$ so that

$$A_2 \leq 1 - \tfrac{1}{2(1+c)}(1 - q_{\text{est}}) + \frac{C_{\text{comp}}/C_U}{1+c} \gamma^{-1} \varepsilon^{-1} < 1. \tag{6.76}$$

It is left to make sure that $A_2 < A_1$ so that contraction of quasi-norm is valid. We know that $A_2 < 1$ and the $\frac{C_{\mathrm{comp}}}{\frac{C_{\mathrm{coer}}}{2} C_{\mathrm{est}} \gamma \varepsilon}$ part of A_1 can be made arbitrarily small by increasing γ, so it is possible to make $A_1 = 1 - \frac{C_{\mathrm{comp}}}{\frac{C_{\mathrm{coer}}}{2} C_{\mathrm{est}} \gamma \varepsilon} > A_2$ for sufficiently large γ. This means $A_2 = \beta A_1$ for some $0 < \beta < 1$ and

$$a_{P_*}(e_{P_*}, e_{P_*}) + \tfrac{C_{\mathrm{coer}}}{2} C_{\mathrm{est}} A_1 \eta_{P_*}^2 \leq \max\{A_2, \beta\} \left(a_P(e_P, e_P) + \tfrac{C_{\mathrm{coer}}}{2} C_{\mathrm{est}} A_1 \eta_P^2 \right). \tag{6.77}$$

■

6.4 Quasi-optimlaity of AFEM

The total-error norm is given by

$$\rho_P(v, V, g) = \left(\|\!|v - V|\!\|_P^2 + \operatorname{osc}_P^2(g) \right)^{1/2}. \tag{6.78}$$

The AFEM approximation class defined in Chapter 3 characterized by nonlinear approximation from spline spaces contained in $H_0^2(\Omega)$ will be shown to be equivalent to the AFEM approximation class of this chapter.

We define the approximation class in which approximation using spline spaces $\mathbb{X}_P$ with $H^2(\Omega)$ regularity but not necessarily in $H_0^2(\Omega)$,

$$\mathbb{A}^s = \left\{ v \in H_0^2(\Omega) : \sup_{N>0} N^s \inf_{P \in \mathscr{P}_N} E_P(v) < \infty \right\}, \tag{6.79}$$

where

$$E_P(v) = \inf_{V \in \mathbb{X}_P} \left(|v - V_0|_{H^2(\Omega)}^2 + \operatorname{osc}_P^2(\mathcal{L}v) \right)^{1/2}, \quad v \in H_0^2(\Omega). \tag{6.80}$$

Analogously, we define the approximation class in which approximation comes from boundary conforming spline spaces $\mathbb{X}_P^0$ used in Chapter 3 by

$$\mathbb{A}_0^s = \left\{ v \in H_0^2(\Omega) : \sup_{N>0} N^s \inf_{P \in \mathscr{P}_N} E_P^0(v) < \infty \right\}, \tag{6.81}$$

where

$$E_P^0(v) = \inf_{V_0 \in \mathbb{X}_P^0} \left(|v - V_0|_{H^2(\Omega)}^2 + \operatorname{osc}_P^2(\mathcal{L}v) \right)^{1/2}, \quad v \in H_0^2(\Omega). \tag{6.82}$$

Lemma 6.10 (Equivalence of classes). $\mathbb{A}^s = \mathbb{A}_0^s$

Proof. It is immediate that $\mathbb{A}_0^s \subset \mathbb{A}^s$. Conversely, let $u \in \mathbb{A}_s$, for $s > 0$, let $N > \#P_0$, let $P_* \in \mathscr{P}_N$ and let $V_* \in \mathbb{X}_{P_*}$ be such that

$$\rho_{P_*}(u, V_*, f) = \inf_{P \in \mathscr{P}_N} E_P(u). \tag{6.83}$$

Using the triangle inequality $\|u - V_*^0\|_{P_*} \leq \|u - V_*\|_{P_*} + \|V_* - V_*^0\|_{P_*}$ with the fact that $|V_*|_{P_*} = |u - V_*|_{P_*}$ we have in view of norm equivalence (6.6)

$$\|V_* - V_*^0\|_{P_*} \leq C_\perp |V_*|_{P_*} \preceq \|u - V_*\|_{P_*}, \tag{6.84}$$

from which we obtain

$$\|u - V_*^0\|_{P_*}^2 + \operatorname{osc}_{P_*}^2(f) \preceq \|u - V_*\|_{P_*}^2 + \operatorname{osc}_{P_*}^2(f). \tag{6.85}$$

Upon taking infimum we arrive at

$$\|u - V_*^0\|_{P_*}^2 + \operatorname{osc}_{P_*}^2(f) \preceq E_P^2(u, f) \preceq N^{-2s}. \tag{6.86}$$

■

Remark 6.11. In other words, the one-sided characterization of Corollary 3.23 is valid for approximation from $\mathbb{X}_P$.

Lemma 6.12 (Quasi-optimality of total error). *For constants $C_{\text{QOTE}} > 0$ and $\gamma_Q > 0$ we have for all $\gamma \geq \gamma_Q$*

$$\rho_P^2(u, U, f) \leq C_{\text{QOTE}} \inf_{V \in \mathbb{X}_P} \rho_P^2(u, V, f). \tag{6.87}$$

Proof. Let $e = u - U$. In view of coercivity partial Galerkin orthogonality (6.8) and continuity of bilinear form

$$\begin{aligned} C_{\text{coer}} \|e\|_P^2 &\leq a_P(e, u - U) = a_P(e, u - U^0) - a_P(e, U^\perp), \\ &= a_P(e, u - V_0) + a_P(e, U^\perp), \\ &= a_P(e, u - V) + a_P(e, V^\perp) + a_P(e, U^\perp), \\ &\leq C_{\text{cont}} \|e\|_P \left(\|u - V\|_P + \|V^\perp\|_P + \|U^\perp\|_P \right). \end{aligned} \tag{6.89}$$

Norm equivalence (6.1) makes $\|V^\perp\|_P \leq C_\perp |u - V^\perp|_P \leq \|u - V\|_P$. Nonconforming control (6.36) and Global Lower Bound (3.39) makes $\|U^\perp\|_P \preceq \gamma^{-1/2} \eta_P \leq \gamma^{-1/2} C_L \rho_P(u, U, f)$. From

$$C_{\text{coer}} \|e\|_P \preceq C_{\text{cont}} \left(\|u - V\|_P + \gamma^{-1/2} C_L \rho_P(u, U, f) \right), \tag{6.90}$$

we get

$$|\!|\!|e|\!|\!|_P^2 \preceq \tfrac{C_{\text{cont}}^2}{C_{\text{coer}}^2} \left(|\!|\!|u - V|\!|\!|_P^2 + \gamma^{-1} C_L^2 \rho_P^2(u, U, f) \right). \tag{6.91}$$

Add $\operatorname{osc}_P^2(f)$ to the preceding expression to get

$$\left(1 - \tfrac{C_{\text{cont}}^2 C_L^2}{C_{\text{coer}}^2} \gamma^{-1}\right) \rho_P^2(u, U, f) \preceq \tfrac{C_{\text{cont}}^2 C_L^2}{C_{\text{coer}}^2} \rho_P^2(u, V, f). \tag{6.92}$$

Let $\gamma_Q = \frac{C_{\text{cont}}^2 C_L^2}{C_{\text{coer}}^2}$. ■

From now on we define

$$\theta_*(\gamma) = \left(\frac{C_L - 2C_{dU}\gamma^{-1}}{2(1 + C_{dU})} \right)^{1/2} \quad \text{and} \quad \gamma_*(\theta) = \max\left(\frac{2C_{dU}}{C_L}, \gamma_Q, \gamma_C(\theta) \right). \tag{6.93}$$

Then $\theta_* > 0$ and since $C_L < C_{dU}$, it is sure that $\theta_* < 1$.

Lemma 6.13 (Optimal marking). *Let $U = \mathbf{SOLVE}\,[P, f]$, let P_* be any refinement of P and let $U_* = \mathbf{SOLVE}\,[P_*, f]$. If for some positive $\mu < 1$*

$$|\!|\!|u - U_*^0|\!|\!|_{P_*}^2 + \operatorname{osc}_*^2(f, P_*) \leq \mu\big(|\!|\!|u - U|\!|\!|^2 + \operatorname{osc}_P^2(f, P) \big), \tag{6.94}$$

and $R_{P\to P_*}$ denotes collection of all elements in P requiring refinement to obtain P_* from P, then for $\theta \in (0, \theta_*(\gamma))$ we have

$$\eta_P(U, \omega_{R_{P\to P_*}}) \geq \theta \eta_P(U, \Omega). \tag{6.95}$$

Proof. Let $\theta < \theta_*$, the parameter θ_* to be specified later, such that the linear contraction of the total error holds for

$$\mu(\theta, \gamma) = \tfrac{1}{2}\left(1 - \tfrac{2C_{dU}\gamma^{-1}}{C_L}\right)\left(1 - \tfrac{\theta^2}{\theta_*^2}\right) < \tfrac{1}{2}, \quad (\gamma \geq \gamma_*). \tag{6.96}$$

The efficiency estimate (3.39) together with the assumption (6.94)

$$\begin{aligned} (1 - 2\mu) C_L \eta_P^2(U, P) &\leq (1 - \mu)\rho_P^2(u, U, f), \\ &= \rho_P^2(u, U, f) - \rho_*^2(u_*, U_*^0, f), \\ &= |\!|\!|u - U|\!|\!|_P^2 - 2|\!|\!|u - U_*^0|\!|\!|_{P_*}^2 + \operatorname{osc}_P^2(f, \Omega) - 2\operatorname{osc}_{P_*}^2(f, \Omega). \end{aligned} \tag{6.98}$$

Discrete reliability (6.23)

$$\begin{aligned} |\!|\!|u - U|\!|\!|_P^2 - 2|\!|\!|u - U_*^0|\!|\!|_{P_*}^2 &\leq 2|\!|\!|U_*^0 - U|\!|\!|_P^2, \\ &\leq 2C_{\text{dRel}}\left(\eta_P^2(U, \omega_{R_{P\to P_*}}) + \gamma^{-1}\eta_P^2(U, \Omega)\right). \end{aligned} \tag{6.100}$$

Estimator Dominance over oscillation

$$\mathrm{osc}_P^2(f,\Omega) - 2\mathrm{osc}_{P_*}^2(f,\Omega) \le 2\mathrm{osc}_P^2(f,\omega_{R_{P\to P_*}}) \le 2\eta_P^2(U,\omega_{R_{P\to P_*}}). \tag{6.102}$$

From

$$(1-2\mu)C_L\eta_P^2(U,P) \le 2(1+C_{dU})\eta_P^2(U,\omega_{R_{P\to P_*}}) + 2C_{dU}\gamma^{-1}\eta_P^2(U,\Omega), \tag{6.103}$$

re-write into

$$\left((1-2\mu)C_L + 2C_{dU}\gamma^{-1}\right)\eta_P^2(U,P) \le 2(1+C_{dU})\eta_P^2(U,\omega_{R_{P\to P_*}}). \tag{6.104}$$

For reader's clarity we show that

$$\tfrac{(1-2\mu)C_L - 2C_{dU}\gamma^{-1}}{2(1+C_{dU})} = \theta^2. \tag{6.105}$$

Express

$$(1-2\mu)C_L - 2C_{dU}\gamma^{-1} = \theta^2 2(1+C_{dU}) = \tfrac{\theta^2}{\theta_*^2}(C_L - 2C_{dU}\gamma^{-1}), \tag{6.106}$$

which is same as

$$-2\mu = \tfrac{\theta^2}{\theta_*^2}\left(1 - \tfrac{2C_{dU}\gamma^{-1}}{C_L}\right) + \tfrac{2C_{dU}\gamma^{-1}}{C_L} - 1 = \left(1 - \tfrac{2C_{dU}\gamma^{-1}}{C_L}\right)\left(\tfrac{\theta^2}{\theta_*^2} - 1\right). \tag{6.108}$$

■

Lemma 6.14 (Cardinality of Marked Cells). *Let $\{(P_\ell, \mathbb{X}_\ell, U_\ell)\}_{\ell\ge 0}$ be a sequence generated by* **AFEM** $(P_0, f; \varepsilon, \theta)$ *for admissible P_0 and the pair $u \in \mathbb{A}^s$ for some $s > 0$ then*

$$\#\mathscr{M}_\ell \preceq \left(1 - \frac{\theta^2}{\theta_*^2}\right)^{-\frac{1}{2s}} |u|_{\mathbb{A}_s}^{-\frac{1}{s}} \rho_\ell(u, U_\ell, f)^{-\frac{1}{s}}. \tag{6.109}$$

Proof. Let $(u,f) \in \mathbb{A}_s$ and set $\varepsilon^2 = \mu C_{\mathrm{QOTE}}^{-1}\rho_\ell^2(u, U_\ell, f)$. In view of Lemma 6.10, $u \in \mathbb{A}_s^0$ and there exists an admissible partition P_ε and $V_\varepsilon^0 \in \mathbb{X}_\varepsilon^0$ with $\rho_\varepsilon^2(u, V_\varepsilon^0, f) \le \varepsilon^2$ and $\#P_\varepsilon \preceq |u|_{\mathbb{A}^s}^{1/s}\varepsilon^{-1/s}$. Let P_* be the overlay of meshes P_ℓ and P_ε. From Partial Consistency (6.7)

$$a_{P_*}(U_*^0, W^0) = \ell_f(W^0) \quad \forall W^0 \in \mathbb{X}_{P_*}^0, \tag{6.110}$$

we invoke Lemma 6.12 on U_*^0 and use the fact $P_* \ge P_\varepsilon$ makes $\mathbb{X}_{P_*} \supseteq \mathbb{X}_\varepsilon$ and obtain

$$\rho_*^2(u, U_*^0, f) \le C_{\mathrm{QOTE}}\rho_\varepsilon^2(u, V_\varepsilon^0, f) \le \varepsilon^2 = \mu\rho_\ell^2(u, U_\ell, f). \tag{6.111}$$

We may now invoke Lemma 6.13 and $R_{P_\ell\to P_*}$ satisfies Dörfler property Minimal cardinality of marked cells

$$\#\mathscr{M}_\ell \le \#R_{P_\ell\to P_*} \le \#P_* - \#P_\ell. \tag{6.112}$$

In view of mesh overlay property $\#P_* \le P_\varepsilon + \#P_\ell - \#P_0$ in (2.75) and definition of ε we arrive at

$$\#\mathscr{M}_\ell \le \#P_\varepsilon - \#P_0 \preceq \mu^{-1/2s}|u|_{\mathbb{A}^s}^{1/s}\rho_\ell(u, U_\ell, f)^{-1/s}. \tag{6.113}$$

■

Theorem 6.15 (Quasi-optimality of Nitsche's method). *Let γ_* and θ_* be as above. If $\gamma > \gamma_*$ and $\theta \in (0, \theta_*(\gamma))$, $u \in \mathbb{A}^s$ and P_0 is admissible, then the call* $\mathbf{AFEM}\,[P_0, f, \varepsilon, \theta]$ *generates a sequence $\{(P_\ell, \mathbb{X}_\ell, U_\ell)\}_{\ell \geq 0}$ of strictly admissible partitions P_ℓ, conforming finite-element spline spaces $\mathbb{X}_\ell$ and discrete solutions U_ℓ satisfying*

$$\rho_\ell(u, U_\ell, f) \preceq \Phi(s, \theta)|u|_{\mathbb{A}^s}(\#P - \#P_0)^{-s}, \tag{6.114}$$

with $\Phi(s, \theta) = (1 - \theta^2/\theta_*^2)^{-\frac{1}{2}}$.

Proof. The proof is similar to that of the conforming formulation of Chapter 3. For completeness we outline the analysis. Let $\theta < \theta_*$ be given and assume that $u \in \mathbb{A}^s$. We will show that the adaptive procedure **AFEM** will produce a sequence $\{(P_\ell, \mathbb{X}_\ell, U_\ell)\}_{\ell \geq 0}$ such that $\rho_\ell \preceq (\#P_\ell - \#P_0)^{-s}$. In view of convergence Theorem 6.9, we have for a factor $C_{\text{est}} > 0$ and a contractive factor $\alpha \in (0, 1)$, Efficiency Estimate (3.39) and estimator dominance over oscillation

$$\sum_{j=0}^{\ell-1} \rho_j^{-\frac{1}{s}} \leq \sum_{j=0}^{\ell-1} \alpha^{\frac{\ell-j}{s}} \left(1 + \tfrac{C_{\text{est}}}{C_L}\right)^{\frac{1}{2s}} (e_\ell^2 + C_{\text{est}}\text{osc}_\ell^2)^{-\frac{1}{2s}}\,. \tag{6.115}$$

Cardinality of Marked Cells (6.109) and (2.107) yields

$$\#P_\ell - \#P_0 \preceq |u|_{\mathbb{A}^s}^{-1/s} \left(1 + \tfrac{C_{\text{est}}}{C_L}\right)^{1/2s} \frac{\alpha^{1/s}}{1 - \alpha^{1/s}} \left(1 - \tfrac{\theta^2}{\theta_*^2}\right)^{-1/2s} \rho_\ell(u, U_\ell, f)^{-\frac{1}{s}}. \tag{6.116}$$

■

Chapter 7

Conclusion

7.1 Discussion

In [61], Cascon, Kreuzer, Nochetto and Siebert provided the first complete performance analysis for adaptive h-refinement methods for general symmetric second order elliptic partial differential equations built on the pioneering work of Verfuth, Babuška, Dörfler, Morin, Siebert, Veeser, Binev, Dahmen, and DeVore and Stevenson. An analogous analysis for the biharmonic equation is carried out in Chapter 3. The crucial differences in the treatment of the fourth-order problem are the necessity for an approximation tool suitable for C^1 quasi-interpolation in a basis that is not interpolatory and in the derivation of the continuous lower bound (3.39). All remaining ingredients for convergence and quasi-optimality analyses followed in an almost exact manner.

The remainder of this thesis explores a weak prescription of the essential boundary conditions using Nitsche's method. In Chapters 4 and 5 we performed *a priori* and *a posteriori* error estimations for the Stommel–Munk models and the stationary quasi-geostrophic equation simulating the large scale wind-driven circulation. Through several benchmark examples, we demonstrated that the theoretical analysis predicts the performance of our algorithms on rectangular and L-shaped geometries with a thin western boundary layer. Adaptive refinements were achieved using hierarchical B-splines. Convergence rates of the adaptive methods were compared with those of the uniform refinement methods.

For both rectangular and L-shaped geometries, higher accuracy of the local refinement was observed in comparison to the uniform refinement during the initial iterations. In particular, local refinement resulted in initially stronger convergence on the rectangular domain with a boundary layer but later returned to the expected optimal rate. We believe that this is because the local refinement resolved the finer details in an already smooth function with efficiency. Moreover, the local refinement method successfully recovered the optimal rate of convergence on the L-shaped geometry in contrast to the sub-optimal performance of the uniform refinement indicating the potential advantage of our algorithm on more complex geometries. Moreover, we compared the computational time of the local refinement with the uniform refinement. The results clearly show that the computational time can be significantly reduced with the local refinement to obtain accurate results relative to the uniform refinement, indicating the efficiency of our adaptive algorithm.

We concluded this thesis by adapting the analysis of Bonito and Nochetto [63] to Nitsche's formulation. Ingredients to prove convergence and quasi-optimality had to take into account the dependence of Nitsche's bilinear form on the mesh.

7.2 Future work

In the future, it is possible to extend our study to the Nitsche-type variational formulation of the nonlinear model, i.e., the stationary quasi-geostrophic equation. More complex geometries with arbitrary coastal boundaries can be considered. Finally, it would be interesting to extend this study to the time-dependent QGE.

The convergence and quasi-optimality of the Nitsche's method can be studied for the general class of fourth-order problems addressed in [3] by Blum, Rannacher and Leis.

Bibliography

[1] Adams and Fournier, *Sobolev spaces*, Elsevier, (2003).

[2] Grisvard, *Elliptic problems in nonsmooth domains*, Volume 69, SIAM, (2011).

[3] Blum, Rannacher and Leis, *On the boundary value problem of the biharmonic operator on domains with angular corners*, Mathematical Methods in the Applied Sciences, Volume 2, No. 4, pp. 556–581, Wiley Online Library, (1980)

[4] Bhattacharyya and Gopalsamy, *On existence and uniqueness of solutions of boundary value problems of fourth order elliptic partial differential equations with variable coefficients*, Journal of mathematical analysis and applications, Volume 136, No. 2, pp. 589–608, Elsevier, (1988)

[5] DeVore, *Nonlinear approximation*, Acta numerica, Volume 7, pp. 51–150, Cambridge University Press, (1998)

[6] DeVore, Ronald A and Lorentz, George G, *Constructive approximation*, Volume 303, Springer Science & Business Media, (1993)

[7] DeVore, Leviatan, Dany and Yu, *Polynomial approximation* L^p $(0 < p < 1)$, Constructive Approximation, Volume 8, No. 2, pp. 187–201, Citeseer, (1992)

[8] DeVore and Popov, *Interpolation of Besov spaces*, Transactions of the American Mathematical Society, Volume 305, No. 1, pp. 397–414, (1988)

[9] Whitney, *On functions with bounded nth differences*, Journal de Mathématiques Pures et Appliquées, Volume 36, pp. 67-95, (1957)

[10] Johnen and Scherer, *On the equivalence of the K-functional and moduli of continuity and some applications*, Constructive theory of functions of several variables, pp. 119–140, Springer, (1977)

[11] de Boor and Fix, *Spline approximation by quasi-interpolants*, Journal of Approximation Theory, Volume 8, No. 1, pp. 19–45, Academic Press, (1973)

[12] Lyche and Schumaker, *Local spline approximation methods*, Journal of Approximation Theory, Volume 15, No. 4, pp. 294–325, Academic Press, (1975)

[13] Graham, *Estimates for the modulus of smoothness*, Journal of approximation theory, Volume 44, 2, pp. 95–112, Academic Press, (1985)

[14] Oswald and Storozhenko, *Jackson's theorem in the spaces $L^p(\mathbb{R}^k)$, $0 < p < 1$*, Siberian. Math, Volume 19, pp. 630–639, (1978)

[15] Butler and Richards, *An L p saturation theorem for splines*, Canadian Journal of Mathematics, Volume 24, No. 5, pp. 957–966, Cambridge University Press, (1972)

[16] Ciesielski, *Constructive function theory and spline systems*, Studia Mathematica, Volume 53, pp. 277–302, Instytut Matematyczny Polskiej Akademii Nauk, (1975)

[17] Petrushev, *Direct and converse theorems for best spline approximation with free knots and Besov spaces*, CR Acad. Bulgare Sci, Volume 39, pp. 25–28, (1986)

[18] DeVor and Popov, *Interpolation spaces and non-linear approximation*, Function spaces and applications, 191–205, Springer, (1988)

[19] Scott and Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Mathematics of Computation, Volume 54, pp. 483–493, (1990)

[20] Binev, Dahmen, DeVore and Petrushev, *Approximation classes for adaptive methods*, Serdica Mathematical Journal, Volume 28, number 4, pp3.91–416, (2002)

[21] Gaspoz, Fernando and Morin, *Approximation classes for adaptive higher order finite element approximation*, Mathematics of Computation, Volume 83, No. 289, pp. 2127–2160, (2014)

[22] Tsogtgerel, *Convergence rates of adaptive methods, Besov spaces, and multilevel approximation*, Foundations of Computational Mathematics, Volume 17, No. 4, pp. 917–956, Springer, (2017)

[23] Schumaker, *Spline functions: basic theory*, Cambridge University Press, 2007

[24] Vuong, Giannelli, Jüttler and Simeon, *A hierarchical approach to adaptive local refinement in isogeometric analysis*, Computer Methods in Applied Mechanics and Engineering, Volume 200, pp. 3554–3567, (2011)

[25] Bazilevs, Beirao da Veiga, Cottrell, Hughes and Sangalli, *Isogeometric analysis: approximation, stability and error estimates for h-refined meshes*, Mathematical Models and Methods in Applied Sciences, Volume 16, pp. 1031–1090, (2006)

[26] Giannelli, Jüttler and Speleers, *Strongly stable bases for adaptively refined multilevel spline spaces*, Advances in Computational Mathematics Volume 40, No. 2, pp. 459–490, Springer, (2014)

[27] Speleers and Manni, *Effortless quasi-interpolation in hierarchical spaces*, Numerische Mathematik, Volume 132, number 1, pages 155–184, Springer, (2016)

[28] Speleers, *Hierarchical spline spaces: quasi-interpolants and local approximation estimates*, Advances in Computational Mathematics, Volume 43, No. 2, pp. 235–255, Springer, (2017)

[29] Stenberg, *On some techniques for approximating boundary conditions in the finite element method*, Journal of Computational and Applied Mathematics, Volume 63, pp. 139 –148, (1995)

[30] Juntunen and Stenberg, *Nitsche's method for general boundary conditions*, Mathematics of computation, Volume 78, pp. 1353–1374, (2009)

[31] Brezzi, *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*, Publications mathématiques et informatique de Rennes, No. S4, pp. 1–26, (1974)

[32] Babuška, *The finite element method with Lagrangian multipliers*, Numerische Mathematik, Volume 20, No. 3, pp. 179–192, Springer, (1973)

[33] Nitsche, *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind*, Abhandlungen aus dem mathematischen Seminar der Universität Hamburg, Volume 36, No. 1, pp. 9–15, Springer, (1971)

[34] Embar, Dolbow and Harari *Imposing Dirichlet boundary conditions with Nitsche's method and spline-based finite elements*, International journal for numerical methods in engineering, Volume 83, No. 7, pp. 877–898, Wiley Online Library, (2010)

[35] Barbosa and Hughes *The finite element method with Lagrange multipliers on the boundary: circumventing the Babuvska-Brezzi condition*, Computer Methods in Applied Mechanics and Engineering, Volume 85, No. 1, pp. 109–128, Elsevier, (1991)

[36] Kim, Iliescu and Fried, *B-spline based finite-element method for the stationary quasi-geostrophic equations of the ocean*, Computer Methods in Applied Mechanics and Engineering, Volume 286, pp. 168–191, (2015)

[37] Foster, Iliescu and Wang *A Finite Element Discretization of the Streamfunction Formulation of the Stationary Quasi-Geostrophic Equations of the Ocean*, Computer Methods in Applied Mechanics and Engineering, Volumes 261-262, pp. 105-117, (2013)

[38] Foster, Iliescu and Wells, *A conforming finite element discretization of the streamfunction form of the unsteady quasi-geotrophic equations*, International Journal of Numerical Analysis & Modeling, Volume 13(6), (2016)

[39] Jiang and Kim, *Spline based finite-element method for the stationary quasi-geostrophic equations on arbitrary shaped costal boundaries*, Computer Methods in Applied Mechanics and Engineering, Volume 299, pp. 144–160, (2016)

[40] Kim, Dolbow and Fried, *A numerical method of a second-gradient theory of incompressible fluid flow*, Journal of Computational Physics, Volume 223, pp. 551–570, (2007)

[41] Kim, Dolbow and Fried, *Numerical study of the grain-size dependent Young's modulus and Poisson's ratio of bulk nanocrystalline materials*, International Journal of Solids and Structures, Volume 49, pp. 3942–3952, (2012)

[42] Kim, Neda, Rebholz and Fried, *A numerical study of the Navier–Stokes-$\alpha\beta$ model*, Computer Methods in Applied Mechanics and Engineering, Volume 200, pp. 2891–2902, (2011)

[43] Kim and Dolbow, *An edge-bubble stabilized finite element method for fourth-order parabolic problems*, Finite Elements in Analysis and Design, Volume 45, pp. 485–494, (2009)

[44] Kim, Park and Shin, *A C^0-discontinuous Galerkin method for the stationary quasi-geostrophic equations of the ocean*, Computer Methods in Applied Mechanics and Engineering, Volume 300, pp. 225–244, (2016)

[45] Rotunda, Kim, Jiang, Heltai and Fried, *Error analysis of a B-spline based finite-element method for modeling wind-driven ocean circulation*, Journal of Scientific Computing, pp.1–30, (2016)

[46] Jiang and Dolbow, *Adaptive refinement of hierarchical B-spline finite elements with an efficient data transfer algorithm*, International Journal for Numerical Methods in Engineering, Volume 102, No. 3–4, pp. 233–256, (2015)

[47] Babuška and Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM Journal on Numerical Analysis, Volume 15 , pp. 736–754, (1978)

[48] Babuška and Rheinboldt, *A-posteriori error estimates for the finite element method*, International Journal for Numerical Methods in Engineering, Volume 12, pp. 1597–1615, (1978)

[49] Babuška and Vogelius, *Feedback and adaptive finite element solution of one-dimensional boundary value problems*, Numerische Mathematik, Volume 44, pp. 75–102, (1984)

[50] Verfürth, *A posteriori error estimation and adaptive mesh-refinement techniques*, Journal of Computational and Applied Mathematics, Volume 50, pp. 67–83, (1994)

[51] Binev and DeVore, *Fast computation in adaptive tree approximation*, Numerische Mathematik, Volume 97, No. 2, pp. 193–217, Springer, (2004)

[52] Binev, Dahmen, and DeVore, *Adaptive finite element methods with convergence rates*, Numerische Mathematik, Volume 97, pp. 219–268, (2004)

[53] Dörfler, *A convergent adaptive algorithm for poisson's equation*, SIAM Journal on Numerical Analysis, Volume 33, pp. 1106–1124, (1996)

[54] Morin, Nochetto, and Siebert, *Data oscillation and convergence of adaptive FEM*, SIAM Journal on Numerical Analysis, Volume 38, pp. 466–488, (2000)

[55] Morin, Nochetto and Siebert *Convergence of adaptive finite element methods*, SIAM review, Volume 44, pp. 631–658, (2002)

[56] Mekchay, Khamron and Nochetto, *Convergence of adaptive finite element methods for general second order linear elliptic PDEs*, SIAM Journal on Numerical Analysis, Volume 43, No. 5, pp. 1803–1827, SIAM, (2005)

[57] Morin, Siebert, and Veeser, *A basic convergence result for conforming adaptive finite elements*, Mathematical Models and Methods in Applied Sciences, Volume 18, pp. 707–737, (2008)

[58] Siebert, *A convergence proof for adaptive finite elements without lower bound*, IMA journal of numerical analysis, Volume 31, pp. 947–970, (2010)

[59] Binev, Dahmen, and DeVore, *Adaptive finite element methods with convergence rates*, Numerische Mathematik, Volume 97, pp. 219–268, (2004)

[60] Stevenson, *An optimal adaptive finite element method*, SIAM journal on numerical analysis, Volume 42, pp. 2188–2217, (2005)

[61] Cascón, Kreuzer, Nochetto and Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM Journal on Numerical Analysis, Volume 46, pp. 2524–2550, (2008)

[62] Feischl, Führer and Praetorius, *Adaptive fem with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems*, SIAM Journal on Numerical Analysis, Volume 52, pp. 601–625, (2014)

[63] Bonito and Nochetto, *Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method*, SIAM Journal on Numerical Analysis, Volume 48, No. 2, pp. 734–771, (2010)

[64] Buffa and Giannelli, *Adaptive isogeometric methods with hierarchical splines: error estimator and convergence*, Mathematical Models and Methods in Applied Sciences, Volume 26, No. 1, pp. 1–25, World Scientific, (2016)

[65] Buffa, Giannelli, Morgenstern and Peterseim *Complexity of hierarchical refinement for a class of admissible mesh configurations*, Computer Aided Geometric Design, Volume 47, pp. 83–92, Elsevier, (2016)

[66] Vallis, *Atmosphere and ocean fluid dynamics: fundamentals and large-scale circulation*, Cambridge University Press, (2006).

[67] Hughes, Cottrell and Bazilevs, *Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement*, Computer Methods in Applied Mechanics and Engineering, Volume 194, pp. 4135–4195, (2005)

[68] Bazilevs and Hughes, *Weak imposition of Dirichlet boundary conditions in fluid mechanics*, Computers & Fluids, Volume 36, No. 1, pp. 12–16, (2017)

[69] Embar, Dolbow and Harari, *Imposing Dirichlet boundary conditions with Nitsche's method and spline-based finite elements.* International Journal For Numerical Methods in Engineering, Volume 83, pp. 877–898, (2010)

[70] Kim, Puntel and Fried, *Numerical study of the wrinkling of a stretched thin sheet, International Journal of Solids and Structures*, Volume 49, pp. 771–782, (2012)

[71] Jiang, Annavarapu, Dolbow and Harari, *A robust Nitsche's formulation for interface problems with spline-based finite elements*, International Journal for Numerical Methods in Engineering, Volume 104, No. 7, pp. 676–696, (2015)

[72] Juntunen and Stenberg, *Nitsche's Method for General Boundary Conditions*, Mathematics of Computation, Volume 78, No. 267, pp.1353-1374, (2009)

[73] Vuong, Giannelli, Jüttler and Simeon . *A hierarchical approach to adaptive local refinement in isogeometric analysis*, Computer Methods in Applied Mechanics and Engineering, Volume 200, No. 49–52, pp. 3554–3567, (2011)

[74] Schillinger, Dedè, Scott, Evans, Borden, Rank and Hughes, *An isogeometric design-through-analysis methodology based on adaptive hierarchical refinement of NURBS, immersed boundary methods, and T-spline CAD surfaces*, Computer Methods in Applied Mechanics and Engineering, Volume 249, pp. 116–150, (2012)

[75] Bornemann and Cirak, *A subdivision-based implementation of the hierarchical b-spline finite element method*, Computer Methods in Applied Mechanics and Engineering, Volume 253, pp. 584–598, (2013)

[76] Cascón, Garcia and Rodriguez, *A priori and a posteriori error analysis for a large-scale ocean circulation finite element model*, Computer Methods in Applied Mechanics and Engineering, Volume 192, pp. 5305–5327, (2003)

[77] Myers and Weaver, *A diagnostic barotropic finite-element ocean circulation model*, Journal of Atmospheric and Oceanic Technology, Volume 12, pp. 511–521, (1995)

[78] Al Balushi, Jiang, Tsogtgerel and Kim, *Adaptivity of a B-spline based finite-element method for modeling wind-driven ocean circulation*, Computer Methods in Applied Mechanics and Engineering, Volume 332, pp. 1–24, Elsevier, (2018)

[79] Al Balushi, Jiang, Tsogtgerel and Kim, *A posteriori analysis of a B-spline based finite-element method for the stationary quasi-geostrophic equations of the ocean*, Computer Methods in Applied Mechanics and Engineering, Volume 371, Elsevier, (2020)