UNIVERSITÀ DEGLI STUDI DI MILANO

Dipartimento di Matematica ”F. Enriques”

Ph. D. Program in Mathematical Sciences

# QUASI-OPTIMAL NONCONFORMING METHODS FOR SYMMETRIC ELLIPTIC PROBLEMS

MAT/08 NUMERICAL ANALYSIS

Candidate

Pietro Zanotti

Advisor

Prof. Andreas Veeser

Ph. D. Program Coordinator

Prof. Vieri Mastropietro

Academic Year 2016/2017

# Contents

# Introduction

In this Ph.D. thesis we characterize quasi-optimal nonconforming methods for certain linear variational problems and investigate their structure. The abstract analysis is complemented by various applications and numerical tests in the finite element framework. In order to provide an easily accessible and self-contained illustration of the subject, we shall restrict our attention to the symmetric elliptic case, although various generalizations are possible. The material collected here substantially results from [63, 64, 65, 62].

A symmetric elliptic linear variational problem can be written as

$$\text{given } \ell \in V', \text{ find } u \in V \text{ such that} \qquad \forall v \in V \quad a(u,v) = \langle \ell, v \rangle \tag{0.1}$$

where $V$ is a Hilbert space, $a$ is a scalar product on $V$ and $\langle \cdot, \cdot \rangle$ denotes the pairing of $V$ and its dual $V'$. Let $S$ be a finite-dimensional linear space and assume that $\| \cdot \|$ extends the energy norm $\sqrt{a(\cdot,\cdot)}$ to $V + S$. We say that an approximation method $u \mapsto U \in S$ is quasi-optimal in the norm $\| \cdot \|$ if there is $C_{\mathrm{qopt}} \geq 1$ such that

$$\forall u \in V \qquad \|u - U\| \leq C_{\mathrm{qopt}} \inf_{s \in S} \|u - s\|. \tag{0.2}$$

Quasi-optimality in the norm $\| \cdot \|$ is a non-asymptotic notion and does not require additional regularity beyond $V$. It is a necessary condition for a near-best balance between quality and cost and, consequently, ensures the effective use of all the degrees of freedom in $S$. It is also a convenient departure point to derive sharp error estimates for regular solutions of the model problem.

The conforming Galerkin method for (0.1)

$$\text{given } \ell \in V', \text{ find } U \in S \text{ such that} \qquad \forall \sigma \in S \quad a(U,\sigma) = \langle \ell, \sigma \rangle \tag{0.3}$$

with discrete space $S \subseteq V$, is quasi-optimal in the energy norm induced by $a$ with $C_{\mathrm{qopt}} = 1$, according to the classical Céa lemma [28]. More generally, Babuška [6] established the quasi-optimality of (0.3) without assuming that $a$ is a scalar product and, in the same setting, Tantardini and Veeser [58] have recently shown that the best constant is

$$C_{\mathrm{qopt}} = \sup_{\sigma \in S} \frac{\sup_{\|v\|=1} a(v,\sigma)}{\sup_{\|s\|=1} a(s,\sigma)}, \tag{0.4}$$

where $v$ varies in $V$ and $s, \sigma$ vary in $S$.

This provides a rather general but still very strong result when the discrete space $S$ is conforming, i.e. it is a subspace of its continuous counterpart $V$. However, methods with nonconforming discrete spaces are of interest because the 'rigidity' of their conforming counterparts may cause problems in approximation, see, e.g., de Boor/DeVore [36] and Babuška/Suri [7], or in stability, see Scott/Vogelius [55].

Popular nonconforming methods for (0.1) are classical nonconforming finite element methods (NCFEM) like the Crouzeix-Raviart or the Morley method and Discontinuous Galerkin (DG) methods. Here the so-called second Strang lemma [10] or variants serve as a replacement for Céa's lemma. Such results are then used to derive error estimates which differ from (0.2), in that they involve extra regularity,

- either of the solution $u$, which has to be taken from a strict compact subset of $V$, see, e.g., Brenner/Scott [21] and Di Pietro/Ern [37],

- or, in the medius analysis initiated by Gudi [43], of the load term $\ell$, which has to be taken from a strict compact subset of $V'$; see also Brenner [16].

In both cases, the extra regularity involved then obstructs a further bound by the best approximation error with respect to the norm $\|\cdot\|$ in order to conclude quasi-optimality.

This simple observation suggests that nonconforming methods may not be quasi-optimal if they are not carefully designed for this purpose. Nonetheless, it is our purpose to show that the possibility of designing quasi-optimal and computationally feasible methods for (0.1) is not generally ruled out by the nonconformity.

The material is organized within the thesis as follows. In Chapter 1, we introduce a rather large class of methods, which mimic the variational formulation of the model problem as follows:

$$\text{(0.5)} \qquad \text{given } \ell \in D(L), \text{ find } U \in S \text{ such that} \qquad \forall \sigma \in S \; b(U, \sigma) = \langle L\ell, \sigma \rangle$$

where $b$ is a nondegenerate bilinear form on $S$ and $L : D(L) \subseteq V' \to S'$ is a densely-defined linear operator. We characterize the quasi-optimality of such method in terms of suitable notions of stability and consistency. Then, as a consequence, we show that the method (0.5) is quasi-optimal if and only if it can be rewritten in the form

$$\text{(0.6)} \qquad \text{given } \ell \in V', \text{ find } U \in S \text{ such that} \qquad \forall \sigma \in S \quad b(U, \sigma) = \langle \ell, E\sigma \rangle$$

for some linear *smoothing* operator $E : S \to V$ such that

$$\text{(0.7)} \qquad \forall u \in V \cap S, \sigma \in S \qquad a(u, E\sigma) = b(u, \sigma).$$

Similarly to (0.4), we identify also the quasi-optimality constant of (0.6), i.e. the best constant in (0.2), and observe its dependence on stability and consistency. For this purpose and further convenience, we associate a stability constant and two consistency measures with each method.

In the truly nonconforming case $S \nsubseteq V$, the use of a smoothing operator $E$ as in the right-hand side of (0.6) is uncommon but not new, see for instance [4, 23]. Moreover, the validity of (0.7) is often not enforced or not fully exploited, see e.g. [8], which contains also a partial quasi-optimality result. For this reason, Chapters 3 and 4 are devoted to exemplify the application of the above-mentioned results, while Chapter 2 collects some necessary preliminaries and notations. We consider three prototypes for the abstract problem (0.1) and propose various nonconforming finite element methods. Each one of them

- seems to be new but differs from an existing (and non-quasi-optimal) one only in the use of the smoothing operator $E$ in the right-hand side of (0.6);
- is quasi-optimal and its quasi-optimality constant is bounded in terms of the shape parameter of the underlying mesh;
- is computationally feasible, in that $O(\dim(S))$ operations are needed to assemble a linear system from (0.6).

More specifically, in Chapter 3 we introduce a subclass of quasi-optimal methods, relating $S$, $b$ and $E$ through a more restrictive counterpart of (0.7). This provides an improved notion of consistency, called overconsistency, entailing that the quasi-optimality constant coincides with the stability constant of (0.6). We obtain overconsistency with

- the Crouzeix-Raviart element for the Poisson problem,
- Crouzeix-Raviart-like elements of arbitrary fixed order for the Poisson problem,
- the Morley element for the biharmonic problem.

Furthermore, in Chapter 4 we design other quasi-optimal (but not overconsistent) methods with

- discontinuous elements for the Poisson problem,
- the lowest-order Crouzeix-Raviart element for linear elasticity,
- a second-order continuous element for the biharmonic problem.

Finally, in Chapter 5 we restrict our attention to the two-dimensional Poisson problem and observe the performance of some quasi-optimal, first-order methods on various benchmarks. The purpose of these tests is the following. First, we aim at assessing the actual size of the constants involved in our analysis. Second, we highlight the importance of full stability, full algebraic consistency and quasi-optimality when rough loads come into play. Third, we also compare different nonconforming methods with their conforming counterpart. All the numerical experiments have been implemented with the help of finite element toolbox ALBERTA [45, 54].

## Acknowledgments

I would like to express my deepest gratitude to Andreas Veeser, for having been much more than my supervisor and for his precious collaboration on all aspects of this project.

I wish to thank also all the mathematicians who contributed to the development of this material during the last three years. In particular, Christian Kreuzer brought the important reference [8] to my attention, while Francesca Tantardini and Rüdiger Verfürth collaborated to the design of the method in Section 3.3.1. Alfred Schmidt generously assisted me in the development of the code for the numerical experiments in Chapter 5. Alexandre Ern, Guido Kanschat and Charalambos Makridakis gave me the possibility of carrying over part of my research at their institutions and contributed with many insightful discussions.

A special acknowledgment is due to Luca Pavarino, who kindly encouraged me to apply for a PhD fellowship.

Finally, I would like to thank the referees who read the thesis and helped me to produce an improved version by their valuable comments.

# Chapter 1

# Abstract Theory

This chapter, which essentially results from [63], is devoted to the abstract analysis of nonconforming methods for symmetric elliptic problems. Before entering into the details, let us begin with a short overview of the main notions and results.

## 1.1 Overview

We shall consider boundary value problems and nonconforming methods which can be cast in the form (0.1) and (0.5), respectively.

Our first main result states that quasi-optimality in the norm $\|\cdot\|$ is equivalent to full algebraic consistency and full stability. Full algebraic consistency means that, whenever the exact solution happens to be in the discrete space, it is also the discrete solution. Notice that this is a quite weak property if the conforming part $S \cap V$ of the discrete space is small. Full stability means that the discrete problem is $\|\cdot\|$-stable for all loads, irrespective of their regularity. Moreover, we show that full stability holds if and only if the discrete problem can be rewritten as in (0.6). Notice that, usually, nonconforming methods are not in this form.

As a second main result, we generalize (0.4) and determine the quasi-optimality constant, i.e. the best constant in (0.2), for a quasi-optimal nonconforming method:

$$C_{\mathrm{qopt}} = \sup_{\sigma \in S} \frac{\sup_{\|v+s\|=1} a(v, E\sigma) + b(s,\sigma)}{\sup_{\|s\|=1} b(s,\sigma)}$$

where $v$ varies in $V$ and $s, \sigma$ belong to $S$. Notice that the numerator handles the nonconformity by an extension interweaving data from the continuous and the discrete problem.

These results reduce the construction of quasi-optimal nonconforming methods for (0.1) to devising suitable operators $E$, mapping discrete functions into continuous ones and enjoying (0.7). This is established for various nonconforming finite element spaces in Chapters 3 and 4.

The rest of this chapter is organized as follows. In §1.2 we set up the notations and notions for our analysis, individuating the concepts of stability and consistency that are necessary for quasi-optimality. In §1.3 we characterize quasi-optimality in terms of such concepts and determine the size of the quasi-optimality constant, inspecting it also by means of suitable consistency measures. Then, we apply these results in §1.4 to determine the structure of quasi-optimal methods for (0.1).

## 1.2 Setting, Stability and Consistency

### 1.2.1 Symmetric Elliptic Problems and Nonconforming Methods

We introduce the abstract boundary value problem and then a class of nonconforming methods, sufficiently large to host our discussion.

Let $V$ be an infinite-dimensional Hilbert space with scalar product $a(\cdot,\cdot)$ and *energy norm* $\|\cdot\| = \sqrt{a(\cdot,\cdot)}$. Moreover, let $V'$ be the topological dual space of $V$, denote by $\langle\cdot,\cdot\rangle$ the pairing of $V$ and $V'$ and endow $V'$ with the *dual energy norm* $\|\ell\|_{V'} := \sup_{v\in V, \|v\|=1}\langle \ell, v\rangle$. We consider the following *'continuous' problem*: given $\ell \in V'$, find $u \in V$ such that

$$\forall v \in V \quad a(u,v) = \langle \ell, v\rangle. \tag{1.2.1}$$

In view of the Riesz representation theorem, this problem is well-posed in the sense of Hadamard and well-conditioned. In fact, if $A : V \to V'$, $v \mapsto a(v,\cdot)$ is the Riesz isometry of $V$, we have $u = A^{-1}\ell$ with

$$\|u\| = \|\ell\|_{V'}. \tag{1.2.2}$$

Given a generic functional $\ell \in V'$, we are interested in 'computable' approximations of the solution $u$ in (1.2.1). In other words, we are interested in approximating the linear operator $A^{-1}$ suitably. Since $A^{-1}$ is bounded, one may want to approximate it by linear operators that are bounded, too. However, in order to embed also existing methods in our setting, we consider more general linear operators $M$, possibly unbounded, with finite-dimensional range $R(M)$ and domain $D(M)$ that is dense in $V'$. We say that $M$ is *entire* whenever it can be directly applied to every instance of the continuous problem: $D(M) = V'$.

We shall analyze methods that build upon the variational structure of (1.2.1) in the following manner. Let $S$ be a nontrivial, finite-dimensional linear space, which will play the role of $V$. We write $\langle\cdot,\cdot\rangle$ also for the pairing of $S$ and $S'$. Notice that we do not require $S \subseteq V$. As a consequence, $\langle \ell, \sigma\rangle$ and $a(s,\sigma)$ may be not defined for some $\ell \in V'$ and $s, \sigma \in S$. We therefore introduce an operator $L : D(L) \subseteq V' \to S'$ and a counterpart $b : S \times S \to \mathbb{R}$ of $a$ and require:

- $L$ is linear, (possibly) unbounded, and densely defined,
- $b$ is bilinear and nondegenerate in that, for any $s \in S$, the property $b(s, \sigma) = 0$ for all $\sigma \in S$ entails $s = 0$.

A method $M$ with domain $D(M) = D(L)$ is then defined by the following *discrete problem*: given $\ell \in D(M)$, find $M\ell \in S$ such that

$$\forall \sigma \in S \quad b(M\ell, \sigma) = \langle L\ell, \sigma \rangle. \tag{1.2.3}$$

*Remark* 1.2.1 (Computing discrete solutions). If $\varphi_1, \dots, \varphi_n$ is some basis of $S$, problem (1.2.3) can be reformulated as a uniquely solvable linear system for the coefficients of $M\ell$ with respect to $\varphi_1, \dots, \varphi_n$. Consequently, $M\ell$ is computable whenever $b(\varphi_j, \varphi_i)$ and $\langle L\ell, \varphi_i \rangle$ can be evaluated for all indices $i, j = 1, \dots, n$. Of course, it is desirable that the number of operations to compute $M\ell$ is of optimal order $O(n)$. A necessary condition for this is that the total number of operations for the aforementioned evaluations is of order $O(n)$.

Methods $M$ with the discrete problem (1.2.3) are given by the triplet $(S, b, L)$, whence we shall write also $M = (S, b, L)$. They may be called *nonconforming linear variational methods* or, shortly, *nonconforming methods*. An important subclass are the *conforming* ones, where the discrete space is contained in the continuous one: $S \subseteq V$. (As for the common usage of 'unbounded' and 'bounded' in operator theory, our usage of 'nonconforming' and 'conforming' is slightly inconsistent in that a conforming method is also nonconforming.)

Conformity allows choosing $b$ and $L$ by means of simple restriction:

$$b = a_{|S \times S} \quad \text{and} \quad \forall \ell \in V' \;\; L\ell = \ell_{|S}. \tag{1.2.4}$$

In this case (1.2.3) is a (conforming) *Galerkin method.* Truly nonconforming examples are DG methods and classical NCFEM.

Introducing the invertible map $B : S \to S'$, $s \mapsto b(s, \cdot)$, the method $M$ is represented by the composition

$$M = B^{-1}L. \tag{1.2.5}$$

Although the target function $u$ is usually unknown, the approximation operator

$$P := MA = B^{-1}LA \tag{1.2.6}$$

with domain $D(P) := A^{-1}D(M)$ in $V$ will turn out to be a useful tool. Figure 1.1 illustrates our setting in a commutative diagram for the special case of an entire method.

$$\begin{tikzcd}
V' \arrow[r, "A^{-1}"] \arrow[d, "L"] \arrow[dr, "M"'] & V \arrow[d, dashed, "P"] \\
S' \arrow[r, "B^{-1}"] & S
\end{tikzcd}$$

Figure 1.1: Commutative diagram with solution operator $A^{-1}$, entire nonconforming variational method $M$ given by $S$, $B$ and $L$, as well as induced approximation operator $P$.

*Remark* 1.2.2 ($S$ and surjectivity of $L$). If $L$ is a linear, unbounded, densely defined operator from $V'$ to $S'$, we have $R(M) \subseteq S$, with equality if and only if $L$ is surjective. In addition, if $R(M)$ is a proper subset of $S$, elementary linear algebra allows to reformulate $M$ as a method over $R(M)$. Consequently, there is some ambiguity in the choice of $S$ if $L$ is not surjective and a slight abuse of notation in writing $M = (S, b, L)$.

### 1.2.2 Defining Quasi-Optimality, Stability and Consistency

We now define the key notions of our analysis for nonconforming methods.

For each $\ell \in V'$, a nonconforming variational method $M = (S, b, L)$ chooses an element of $S$ in order to approximate $u = A^{-1}\ell$. To assess the quality of this choice, we assume that $a$ can be extended to a scalar product $\widetilde{a}$ on $\widetilde{V} := V + S$ and consider the *extended energy norm*

$$\|\cdot\| := \sqrt{\widetilde{a}(\cdot,\cdot)} \quad \text{on } \widetilde{V},$$

with the same notation as for the original one. Observe that $V$ and $S$ are closed subspaces of $\widetilde{V}$.

The best approximation error within $S$ to some function $v \in V$ is then given by $\inf_{s\in S} \|v - s\|$. Of course, it is desirable that a method is uniformly close to this benchmark, i.e. there holds an inequality essentially reversing

$$\forall u \in D(P) \qquad \inf_{s\in S} \|u - s\| \leq \|u - Pu\|.$$

**Definition 1.2.3** (Quasi-optimality). *A nonconforming variational method $M$ with discrete space $S$ and approximation operator $P$ is* quasi-optimal *whenever there exists a constant $C \geq 1$ such that*

$$\forall u \in D(P) \qquad \|u - Pu\| \leq C \inf_{s\in S} \|u - s\|.$$

*The quasi-optimality constant $C_{\mathrm{qopt}}$ of $M$ is then the smallest constant with this property.*

Céa's lemma [28] shows that conforming Galerkin methods for (1.2.1) are quasi-optimal with $C_{\text{qopt}} = 1$ and that the associated approximation operator $P = MA$ is the bounded linear $a$-orthogonal projection (or idempotent) onto $S$: in fact, we have the celebrated Galerkin orthogonality

$$\forall u \in V, \sigma \in S \subseteq V \qquad a(u - Pu, \sigma) = 0. \tag{1.2.7}$$

Before analyzing which of these properties still hold in the general case, let us discuss some necessary conditions for quasi-optimality and their consequences.

*Remark* 1.2.4 (Quasi-optimal needs entire). Let $P$ be the approximation operator of a quasi-optimal method $M$. The application $v \mapsto \inf_{s\in S} \|v - s\|$ is Lipschitz continuous on $V$. Therefore, quasi-optimality implies that also $\mathrm{Id}_V - P$ and $P$ are Lipschitz continuous. Since $D(P)$ is dense in $V$ and $S$ complete, the operator $P$ thus extends to $V$ in a continuous and unique manner. As a consequence, $M$ extends to $V'$ in a continuous and unique manner. In other words: ignoring the aspect of computability, only entire methods can be quasi-optimal.

Notice that most classical NCFEM and DG methods are not defined as entire. Consequently, the simple observation in Remark 1.2.4 questions that these methods can be quasi-optimal. This doubt will be confirmed in Remark 1.4.9 below. See also §5.3.2 for a concrete counter-example.

Generally speaking, stability is associated with the property that small input perturbations result in small output perturbations. The form of the discrete problem (1.2.3) suggests adopting the viewpoint that input is taken from a subset of $V'$. Since (1.2.3) is linear, stability then amounts to some operator norm of $M$. Notice that this differs from the common viewpoint that stability is connected solely with an operator norm of $B^{-1}$, i.e. taking input from $S'$. In the following definition, we consider perturbations and measure them as suggested by the setting of the continuous problem.

**Definition 1.2.5** (Full stability). *We say that $M$ is* fully stable *whenever $D(M) = V'$ and, for some constant $C \geq 0$, we have*

$$\forall \ell \in V' \qquad \|M\ell\| \leq C\|\ell\|_{V'}.$$

*The smallest such constant is the stability constant $C_{\text{stab}}$ of $M$.*

Full stability may go beyond the need for practical computations, but it relates to the previous notions in the following manner.

*Remark* 1.2.6 (Fully stable, quasi-optimal and entire). The approximation operator $P$ of a quasi-optimal method satisfies

$$\|Pu\| \leq \|u\| + \|Pu - u\| \leq (1 + C_{\text{qopt}})\|u\| = (1 + C_{\text{qopt}})\|Au\|_{V'}$$

for all $u \in V$, using $0 \in S$, (1.2.2) and Remark 1.2.4. In view of (1.2.6), full stability is thus necessary for quasi-optimality. Furthermore, full stability itself requires that the method is entire in the vein of Remark 1.2.4.

Roughly speaking, consistency should measure to what extent the exact solution verifies the discrete problem. To this end, one usually substitutes in the discrete problem the discrete solution by the exact one and investigates a possible defect. Here nonconformity entails that the forms $b$ and $L$ cannot be defined by simple restriction and so creates the following issues concerning trial and test space:

- In which sense can we plug the solution $u$ of the continuous problem into the discrete problem? Does this require an extension of $b$ or a representative of $u$ in $S$?
- How do we relate the condition associated with a nonconforming test function $\sigma \in S \setminus V$ in (1.2.3) to the conditions given by the continuous test functions in (1.2.1)?

These issues are usually tackled with the help of regularity assumptions on the exact solution, see, e.g., Arnold et al. [5], or only on data, see Gudi [43]. The following definition takes a different approach within our non-asymptotic setting.

**Definition 1.2.7** (Full algebraic consistency). *The method $M$ is* fully algebraically consistent *whenever $D(M) = V'$ and*

$$\forall u \in V \cap S, \sigma \in S \quad b(u, \sigma) = \langle LAu, \sigma \rangle. \tag{1.2.8}$$

Conforming Galerkin (1.2.4) methods are fully algebraically consistent. Let us discuss further aspects of this property.

*Remark* 1.2.8 (Full algebraic consistency and approximation operator). In view of the discrete problem (1.2.3) and the definition (1.2.6) of the approximation operator, (1.2.8) is equivalent to $b(u - Pu, \sigma) = 0$ for all $u \in V \cap S, \sigma \in S$. Since $b$ is nondegenerate, the consistency condition (1.2.8) is therefore equivalent to

$$\forall u \in V \cap S \quad Pu = u. \tag{1.2.9}$$

In other words: full algebraic consistency means that whenever the exact solution is discrete, it is the discrete solution. The advantage of (1.2.8) is that it is directly formulated in terms of the originally given data $A$, $S$, $b$ and $L$. In Lemma 1.2.10 and Theorem 1.4.14 below, we will present further equivalent formulations.

*Remark* 1.2.9 (Quasi-optimal needs fully algebraically consistent). In light of Remark 1.2.4, a quasi-optimal method $M$ is entire and so its approximation operator $P$ is defined on all $V$. For any $u \in V \cap S$, the best error in $S$ vanishes and so $Pu = u$. Consequently, $M$ is fully algebraically consistent.

Definition 1.2.7 involves only exact solutions from the discrete space $S$, which may be a quite small set. Indeed, for example, when applying the Morley method to the biharmonic problem, the intersection $S \cap V$ has poor approximation properties for certain mesh families; see [36, Theorem 3] and Remark 3.3.12. Other consistency notions of algebraic type involving more exact solutions may thus appear stronger than Definition 1.2.7. The following lemma sheds a different light on this.

**Lemma 1.2.10** (Full algebraic consistency with extension)**.** *Let the method $M$ be fully algebraically consistent and set $\widetilde{V} := V + S$. Then there exists a unique bilinear form $\widetilde{b}$ that extends $b$ as well as $\langle LA\cdot, \cdot\rangle$ on $\widetilde{V} \times S$.*

*Proof.* Observe that the left-hand side of (1.2.8) is defined for all $u \in S$, while its right-hand side is defined in particular for all $u \in V$. We exploit this in order to extend $b$. Given $\widetilde{v} \in \widetilde{V}$ and $\sigma \in S$, we write $\widetilde{v} = v + s$ with $v \in V$ and $s \in S$ and set

$$\widetilde{b}(\widetilde{v}, \sigma) := \langle LAv, \sigma\rangle + b(s, \sigma). \tag{1.2.10}$$

Thanks to (1.2.8), $\widetilde{b}$ is well-defined. Indeed, if $v_1 + s_1 = v_2 + s_2$ with $v_1, v_2 \in V$ and $s_1, s_2 \in S$, we have $v_1 - v_2 = s_2 - s_1 \in V \cap S$ and therefore (1.2.8) yields $\langle LA(v_1 - v_2), \sigma\rangle = -b(s_1 - s_2, \sigma)$, which in turn ensures

$$\langle LAv_1, \sigma\rangle + b(s_1, \sigma) = \langle LAv_2, \sigma\rangle + b(s_2, \sigma).$$

To show uniqueness of the extension, let $\widetilde{\beta}$ be another common extension of $b$ and $\langle LA\cdot, \cdot\rangle$. Given $\widetilde{v} \in \widetilde{V}$ and $\sigma \in S$, we write $\widetilde{v} = v + s$ with $v \in V$ and $s \in S$ as before and infer

$$\widetilde{\beta}(\widetilde{v}, \sigma) = \widetilde{\beta}(v, \sigma) + \widetilde{\beta}(s, \sigma) = \langle LAv, \sigma\rangle + b(s, \sigma) = \widetilde{b}(\widetilde{v}, \sigma)$$

and the proof is complete. □

Notice that full algebraic consistency differs from the usual consistency, as, e.g. in Arnold [3] also for the following aspects: on the one hand, it is stronger in that it requires an algebraic identity instead of a limit. On the other hand, it does not involve approximation properties of the underlying discrete space. In fact, our purpose here is to identify the part of consistency that is necessary for quasi-optimality. As a consequence, algebraic consistency and stability alone are not sufficient for convergence, which hinges on the approximation properties of the discrete space $S$.

Let us conclude this section by introducing a subclass of natural candidates for fully algebraically consistent methods. A method $M = (S, b, L)$ is a *nonconforming Galerkin method* whenever

$$b_{|S_C \times S_C} = a_{|S_C \times S_C} \quad \text{and} \quad \forall \ell \in D(L) \;\; L\ell_{|S_C} = \ell_{|S_C}, \tag{1.2.11}$$

where $S_C = S \cap V$ is the conforming subspace of the discrete space $S$. Thus, a nonconforming Galerkin method is constrained by restriction where applicable. Notice that:

- In contrast to conforming Galerkin methods, nonconforming ones are not completely determined by the continuous problem and the choice of the discrete space.
- The condition (1.2.11) readily yields

  $$\forall u, \sigma \in S \cap V \quad b(u, \sigma) = \langle LAu, \sigma \rangle,$$

  which is weaker than full algebraic consistency in that less test functions are involved.

For example, classical NCFEM, DG and $C^0$ interior penalty methods are nonconforming Galerkin methods.

## 1.3 Characterizing Quasi-Optimality

The purpose of this section is twofold. First, we show that full algebraic consistency and full stability are not only necessary but also sufficient for quasi-optimality. Second, we assess the possible impact of nonconformity on the quasi-optimality constant.

### 1.3.1 Quasi-Optimality and Extended Approximation Operator

To show that the combination of full algebraic consistency and full stability implies quasi-optimality, we start with the following short proof of a 'partial' quasi-optimality, which motivates a new tool for the analysis of nonconforming methods.

Assume that $P$ is the approximation operator of a fully algebraically consistent and fully stable method. Rewriting (1.2.9) as

$$(1.3.1) \qquad \forall v \in V, s \in S \cap V \quad v - Pv = (\mathrm{Id}_V - P)(v - s)$$

and exploiting that full stability entails the boundedness of $P$, we can deduce quasi-optimality with respect to the conforming part $S \cap V$ of the discrete space $S$:

$$\|v - Pv\| \leq \|\mathrm{Id}_V - P\|_{\mathcal{L}(V,\widetilde{V})} \inf_{s \in S \cap V} \|v - s\|.$$

Notice that we do not obtain quasi-optimality with respect to the whole discrete space, just because $Ps = s$ is not available for general $s \in S$. In particular, $Ps$ is not defined for general $s \in S$. We therefore explore an appropriate extension of $P$.

For this purpose, we use the following facts on linear projections; cf., e.g., Buckholtz [25]. Let $K$ and $R$ be subspaces of a Hilbert space $H$ with scalar product $(\cdot,\cdot)_H$ and induced norm $\|\cdot\|_H$. The spaces $K$ and $R$ provide a direct decomposition of $H$, i.e. $H = K \oplus R$, if and only if there exists a unique linear projection $Q$ on $H$ with kernel $N(Q) = K$ and range $R(Q) = R$. Then $\mathrm{Id}_H - Q$ is the linear projection with kernel $R$ and range $K$. As a consequence of the closed graph theorem, $R$ and $K$ are closed if and only if $Q$ is bounded if and only if $\mathrm{Id}_H - Q$ is bounded.

**Lemma 1.3.1** (Extended approximation operator)**.** *Assume that the approximation operator $P$ verifies $P|_{S\cap V} = \mathrm{Id}_{S\cap V}$ and is bounded. Then there exists a unique bounded linear projection $\widetilde{P}$ from $\widetilde{V}$ onto $S$ satisfying $\widetilde{P}_{|V} = P$.*

*Proof.* First, we observe that $\widetilde{P}$ has to satisfy

$$\widetilde{P} : \widetilde{V} \to S \text{ linear}, \quad \widetilde{P}_{|V} = P \quad \text{and} \quad \widetilde{P}_{|S} = \mathrm{Id}_S. \tag{1.3.2}$$

Since $\widetilde{V} = V + S$, linear extension entails that there is at most one operator satisfying (1.3.2) and we are thus led to consider the following definition: given $\widetilde{v} \in \widetilde{V}$, choose $v \in V$ and $s \in S$ such that $\widetilde{v} = v + s$ and set

$$\widetilde{P}\widetilde{v} := Pv + s. \tag{1.3.3}$$

The assumption $P_{|S\cap V} = \mathrm{Id}_{S\cap V}$ means that the two identities in (1.3.2) are compatible and so guarantees that $\widetilde{P}$ is well-defined; compare with the definition of $\widetilde{b}$ in the proof of Lemma 1.2.10.

In order to show the boundedness of $\widetilde{P}$, we represent it in terms of $P$ and the following operators, corresponding to an appropriate choice of $v$ and $s$ in (1.3.3). Let $\Pi_Y$ be the $\widetilde{a}$-orthogonal projection onto $Y := (S \cap V)^\perp$ and let $Q$ be the linear projection on $Y$ with range $V \cap Y$ and kernel $S \cap Y$. We then have

$$\widetilde{P} = PQ\Pi_Y + (\mathrm{Id}_Y - Q)\Pi_Y + (\mathrm{Id}_{\widetilde{V}} - \Pi_Y) = PQ\Pi_Y + \mathrm{Id}_{\widetilde{V}} - Q\Pi_Y.$$

Since the subspaces $S$, $V$, and $Y$ are closed, the projections $\Pi_Y$ and $Q$ are bounded. Consequently, the boundedness of $P$ implies the boundedness of its extension $\widetilde{P}$. □

Using the extended approximation operator $\widetilde{P}$, the proof of the announced characterization of quasi-optimality is quite simple. Notice also that the quantitative aspect of our first main result highlights the importance of $\widetilde{P}$.

**Theorem 1.3.2** (Characterization of quasi-optimality)**.** *A nonconforming method is quasi-optimal if and only if it is fully algebraically consistent and fully stable.*

*Moreover, for any quasi-optimal method, we have*

$$C_{\mathrm{qopt}} = \|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})}$$

*where* $\widetilde{P}$ *is the extended approximation operator from Lemma 1.3.1.*

*Proof.* Remarks 1.2.6 and 1.2.9 show that quasi-optimality implies full algebraic consistency and full stability.

To show the converse, consider any fully algebraically consistent and fully stable nonconforming method. We simply follow the lines of the corresponding part of the proof of Tantardini/Veeser [58, Theorem 2.1], replacing $P$ by $\widetilde{P}$ and exploiting the following generalization of (1.3.1):

$$\forall v \in V, s \in S \quad (\mathrm{Id}_{\widetilde{V}} - \widetilde{P})(v - s) = (\mathrm{Id}_V - P)v. \tag{1.3.4}$$

Given arbitrary $v \in V$ and $s \in S$, we thus derive

$$\|v - Pv\| = \|(v - s) - \widetilde{P}(v - s)\| \leq \|\mathrm{Id}_{\widetilde{V}} - \widetilde{P}\|_{\mathcal{L}(\widetilde{V})}\|v - s\|.$$

Taking the infimum over all $s \in S$ and then the supremum over all $v \in V$, we obtain

$$C_{\mathrm{qopt}} \leq \|\mathrm{Id}_{\widetilde{V}} - \widetilde{P}\|_{\mathcal{L}(\widetilde{V})} \tag{1.3.5}$$

and see that $M$ is quasi-optimal because $\widetilde{P}$ is bounded.

To verify, the identity for $C_{\mathrm{qopt}}$, let us first see that (1.3.5) is actually an equality. In fact, for $v \in V$ and $s \in S$, we derive

$$\|(\mathrm{Id}_{\widetilde{V}} - \widetilde{P})(v + s)\| = \|v - Pv\| \leq C_{\mathrm{qopt}} \inf_{\hat{s} \in S} \|v - \hat{s}\| \leq C_{\mathrm{qopt}}\|v + s\|$$

using (1.3.4) again. We thus obtain the converse to (1.3.5) by taking the supremum over all $v \in V$ and $s \in S$.

Moreover, since $\{0\} \subsetneq S \subsetneq \widetilde{V}$, the extended approximation operator $\widetilde{P}$ is a bounded linear idempotent with $0 \neq \widetilde{P} = \widetilde{P}^2 \neq \mathrm{Id}_{\widetilde{V}}$ on the Hilbert space $\widetilde{V}$. We therefore apply Buckholtz [25, Theorem 2] or Xu/Zikatanov [68, Lemma 5] and conclude

$$C_{\mathrm{qopt}} = \|\mathrm{Id}_{\widetilde{V}} - \widetilde{P}\|_{\mathcal{L}(\widetilde{V})} = \|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})}. \tag{1.3.6}$$ □

Formula (1.3.6) allows for the following geometric interpretation of the quasi-optimality constant.

*Remark* 1.3.3 (Geometry of the quasi-optimality constant). Buckholtz [25] shows that the operator norm of a bounded projection $Q$ on a Hilbert space $H$ satisfies

$$\|Q\|_{\mathcal{L}(H)} = \frac{1}{\sin\theta} = \|\mathrm{Id}_H - Q\|_{\mathcal{L}(H)},$$

where $\theta$ is the angle between $K = N(Q)$ and $R = R(Q)$, that is, $\theta \in (0, \pi/2]$ and its cosine equals $\sup\{|\langle k, r\rangle_H| \mid k \in K, r \in R, \|k\|_H = 1, \|r\|_H = 1\}$. Notice that $N(\widetilde{P}) = R(\mathrm{Id}_{\widetilde{V}} - \widetilde{P}) = R(\mathrm{Id}_V - P)$, where the last identity follows from (1.3.4). Combining these two facts, we deduce

$$C_{\mathrm{qopt}} = \|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} = \frac{1}{\sin\alpha} \tag{1.3.7}$$

where $\alpha$ is the angle between the discrete space $S$ and the range $R(\mathrm{Id}_V - P)$.

Theorem 1.3.2 reveals that the possibly weak full algebraic consistency is still enough consistency to ensure, together with stability, quasi-optimality. However, it does not control the size of the quasi-optimality constant.

### 1.3.2 The Quasi-Optimality Constant and Two Consistency Measures

Let $P$ be the approximation operator of a quasi-optimal method. The fact that $\widetilde{P}$ is an extension of $P$ readily yields

$$C_{\mathrm{qopt}} = \|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} \geq \|P\|_{\mathcal{L}(V,S)} = C_{\mathrm{stab}},$$

where the last identity is due to isometry (1.2.2) of $A$. The possible enlargement of $C_{\mathrm{qopt}}$ with respect to $C_{\mathrm{stab}}$ is a new feature triggered by nonconformity. It is the purpose of the section to quantify this phenomenon.

Our key tool will be the following elementary lemma.

**Lemma 1.3.4** (Operator norm and restrictions). *Assume that $T \in \mathcal{L}(H)$ is a bounded linear operator on a Hilbert space $H$ with scalar product $\langle\cdot,\cdot\rangle_H$ and induced norm $\|\cdot\|_H$. If $Y$ is a linear closed subspace of $H$ and $Y^\perp$ is its orthogonal complement, we have*

$$\max\{C, \delta\} \leq \|T\|_{\mathcal{L}(H)} \leq \sqrt{C^2 + \delta^2}$$

*with*

$$C = \|T_{|Y}\|_{\mathcal{L}(Y,H)} \quad \textit{and} \quad \delta = \|T_{|Y^\perp}\|_{\mathcal{L}(Y^\perp,H)}.$$

*Proof.* The lower bound immediately follows from the definition of the operator norm $\|T\|_{\mathcal{L}(H)} = \sup_{\|x\|_H=1} \|Tx\|_H$. To verify the upper bound, let $x \in H$ be arbitrary and denote by $\pi_Y$ the orthogonal projection onto $Y$. We have

$$\begin{aligned} \|Tx\|_H^2 &= \|T\pi_Y x\|_H^2 + 2\big\langle T\pi_Y x, T(x - \pi_Y x)\big\rangle_H + \|T(x - \pi_Y x)\|_H^2 \\ &\leq C^2\|\pi_Y x\|_H^2 + 2C\delta\|\pi_Y x\|_H\|x - \pi_Y x\|_H + \delta^2\|x - \pi_Y x\|_H^2 \end{aligned} \tag{1.3.8}$$

in view of the bilinearity of the scalar product, the Cauchy-Schwarz inequality and the definitions of $C$ and $\delta$. Notice that

$$\|\pi_Y x\|_H^2 + \|x - \pi_Y x\|_H^2 = \|x\|_H^2$$

thanks to the orthogonality of $\pi_Y$. Thus, if we write $\alpha = \|\pi_Y x\|$, the bound in (1.3.8) becomes

$$\|Tx\|_H^2 \leq h(\alpha)^2 \quad \text{with} \quad h(\alpha) := C\alpha + \delta\sqrt{1-\alpha^2},$$

which implies

$$\|T\|_{\mathcal{L}(H)} \leq \max_{[0,1]} h.$$

A straight-forward discussion of the function $h$ yields $\max_{[0,1]} h = \sqrt{C^2+\delta^2}$ and the upper bound is established, too. □

*Remark* 1.3.5 (Sharpness of bounds via restrictions). Since

$$\max\{C,\delta\} \leq \sqrt{C^2+\delta^2} \leq \sqrt{2}\max\{C,\delta\},$$

the bounds in Lemma 1.3.4 miss an equality at most by the factor $\sqrt{2}$. Let us see with two simple examples that, without additional information on $T$ and $Y$, we cannot improve on this.

First, consider $H = \mathbb{R}^2$, $T_1 = \mathrm{Id}_{\mathbb{R}^2}$ and let $Y$ be any 1-dimensional subspace of $\mathbb{R}^2$. We have $\|T_1\|_{\mathcal{L}(H)} = \|T_{1|Y}\|_{\mathcal{L}(Y,H)} = \|T_{1|Y^\perp}\|_{\mathcal{L}(Y^\perp,H)} = 1$ and so the lower bound becomes an equality, while the upper bound is strict.

Second, consider $H = \mathbb{R}^2$ and let $T_2$ be the linear operator which is represented in the canonical basis of $\mathbb{R}^2$ by the Matlab matrix `1/2*ones(2)`. The operator $T_2$ is the orthogonal projection onto the diagonal $\{(t,t) \mid t \in \mathbb{R}\}$, whence $\|T_2\|_{\mathcal{L}(H)} = 1$. Finally, let $Y = \{(0,t) \mid t \in \mathbb{R}\}$ be the ordinate. Then the operator norms of $T_2$ restricted to $Y$ and $Y^\perp$ correspond to the Euclidean norms of the columns of the aforementioned matrix: $\|T_{2|Y}\|_{\mathcal{L}(Y,H)} = \|T_{2|Y^\perp}\|_{\mathcal{L}(Y^\perp,H)} = 1/\sqrt{2}$. Consequently, here the upper bound is an equality, while the lower bound is strict.

The fact that the extended approximation operator $\widetilde{P}$ is given on $S$ by the identity and on $V$ by $P$ suggests to apply Lemma 1.3.4 with either $Y = S$ or $Y = V$. We start with the first option, which leads to a consistency measure in the spirit of the second Strang lemma.

**Proposition 1.3.6** (Consistency mixed with stability). *Let $\Pi_S$ denote the $\widetilde{a}$-orthogonal projection onto $S$ and $\delta_V \geq 0$ be the smallest constant such that*

$$\forall v \in V \quad \|\Pi_S v - Pv\| \leq \delta_V \|v - \Pi_S v\|.$$

*Then the quasi-optimality constant is given by $C_{\mathrm{qopt}} = \sqrt{1+\delta_V^2}$.*

*Proof.* Owing to Theorem 1.3.2, we may show the claimed identity by verifying $\|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} = \sqrt{1+\delta_V^2}$. Applying Lemma 1.3.4 with $H = \widetilde{V}$, $T = \widetilde{P}$ and $Y = S$, we obtain

$$\|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} \leq \sqrt{1+\delta^2}$$

with $\delta = \|\widetilde{P}\|_{\mathcal{L}(S^\perp, \widetilde{V})}$. Given $s^\perp \in S^\perp$, we write $s^\perp = v + s$ with $v \in V$ and $s \in S$ and observe that

$$s^\perp = s^\perp - \Pi_S s^\perp = v - \Pi_S v \quad \text{and} \quad \widetilde{P} s^\perp = Pv - \Pi_S v.$$

Hence $\delta = \delta_V$ and

$$\|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} \leq \sqrt{1 + \delta_V^2}. \tag{1.3.9}$$

To show that this is actually an equality, note that, for any $v \in V$,

$$\|v - \Pi_S v\|^2 + \|\Pi_S v - Pv\|^2 = \|v - Pv\|^2 \leq \|\widetilde{P}\|^2_{\mathcal{L}(\widetilde{V})} \|v - \Pi_S v\|^2, \tag{1.3.10}$$

where we first combined the orthogonality of $\Pi_S$ with $\Pi_S v - Pv \in S$ and then used Theorem 1.3.2. Rearranging terms, we see that $\delta_V^2 \leq \|\widetilde{P}\|^2_{\mathcal{L}(\widetilde{V})} - 1$, yielding the desired inequality $\sqrt{1 + \delta_V^2} \leq \|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})}$. □

The following two remarks discuss the nature of $\delta_V$.

*Remark* 1.3.7 ($\delta_V$ and (non)conforming consistency). In the conforming case $S \subseteq V$, without assuming the quasi-optimality of the underlying method, the existence of $\delta_V$ is equivalent to full algebraic consistency. Therefore, $\delta_V$ can be seen as a quantitative generalization of full algebraic consistency to the nonconforming case. It measures, in relative manner, how much the method deviates from the best approximation $\Pi_S$. Thus, Proposition 1.3.6 is a specification of the second Strang lemma, where the exploitation of the nonconforming direction is compared with the best approximation error. Let us illustrate this in the purely nonconforming case $V \cap S = \{0\}$. The best case corresponds to $P = \Pi_S$, yielding $\delta_V = 0$ and $C_{\text{qopt}} = 1$. Instead, $P = 0$ is quasi-optimal with $\delta_V = (\inf_{\|s\|=1} \inf_{\Pi_S v = s} \|s - v\|)^{-1}$, which becomes infinity as the distance between $S$ and $V$ tends to 0.

*Remark* 1.3.8 ($\delta_V$ and stability). The size of $\delta_V$ is in general affected by stability. Indeed, using (1.3.9), we readily derive

$$\delta_V \geq \sqrt{\|\widetilde{P}\|^2_{\mathcal{L}(\widetilde{V})} - 1} \geq \sqrt{\|P\|^2_{\mathcal{L}(V,S)} - 1} = \sqrt{C^2_{\text{stab}} - 1}$$

and notice in particular that, if a sequence of methods becomes unstable, the corresponding $\delta_V$'s become unbounded.

We now turn to the second option of applying Lemma 1.3.4. Interestingly, this provides an alternative consistency measure which is essentially independent of stability.

**Proposition 1.3.9** (Consistency without stability). *Let us denote by* $\Pi_V$ *the* $\widetilde{a}$*-orthogonal projection onto* $V$ *and let* $\delta_S \geq 0$ *be the smallest constant such that*

$$\forall s \in S \quad \|s - P\Pi_V s\| \leq \delta_S \|s - \Pi_V s\|.$$

*Then the quasi-optimality constant satisfies*

$$\max\{C_{\mathrm{stab}}, \delta_S\} \leq C_{\mathrm{qopt}} \leq \sqrt{C_{\mathrm{stab}}^2 + \delta_S^2}. \tag{1.3.11}$$

*Proof.* Thanks to Theorem 1.3.2, it is sufficient to apply Lemma 1.3.4 with $H = \widetilde{V}$, $T = \widetilde{P}$ and $Y = V$ and to observe the following identities: given $v^\perp \in V^\perp$, $v \in V$, $s \in S$ such that $v^\perp = v + s$, we have

$$v^\perp = v^\perp - \Pi_V v^\perp = s - \Pi_V s \quad \text{and} \quad \widetilde{P} v^\perp = s - P\Pi_V s. \qquad \square$$

We now discuss also the nature of $\delta_S$, elaborating its differences from the first consistency measure $\delta_V$.

*Remark* 1.3.10 ($\delta_S$ and (non)conforming consistency)*.* As for $\delta_V$, the existence of $\delta_S$ is equivalent to full algebraic consistency in the conforming case $S \subseteq V$. Correspondingly, it can be seen as an alternative, quantitative generalization of full algebraic consistency to the nonconforming case. The alternative $\delta_S$ is however not comparing with the best approximation $\Pi_S$. In particular, we have that $\delta_S = 0$ implies

$$C_{\mathrm{qopt}} = \|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} = \|P\|_{\mathcal{L}(V,S)} = C_{\mathrm{stab}},$$

which is an interesting property not involving the best approximation $\Pi_S$. Let us illustrate how the difference is expressed in measuring the exploitation of the nonconforming directions by considering, as in Remark 1.3.7, the purely nonconforming case $V \cap S = \{0\}$. Here the best choice $P = \Pi_S$ leads to $\delta_S < 1$, while $P = 0$ gives $\delta_S = (\inf_{\|s\|=1} \|s - \Pi_V s\|)^{-1}$. In the latter case, $\delta_S$ like $\delta_V$ becomes infinity as the distance between $S$ and $V$ tends to 0, although in a (possibly) other manner.

*Remark* 1.3.11 ($\delta_S$ and stability)*.* We illustrate that the quantities $\delta_S$ and $C_{\mathrm{stab}}$ are essentially independent. In order to make sure that this is not affected by a possible lack of approximability, we consider the following setting with a sequence of discrete spaces:

$$\widetilde{V} = \ell_2(\mathbb{R}) \text{ with canonical basis } (e_i)_{i=0}^\infty, \quad \widetilde{a}(v,w) = \sum_{i=0}^\infty v_i w_i,$$

where we identify $v = \sum_{i=0}^\infty v_i e_i$ with $(v_i)_{i=0}^\infty$, etc., and

$$V = \overline{\operatorname{span}\{e_i \mid i \geq 1\}},$$
$$S_n = \operatorname{span}\{e_i \mid i = 1, \dots, n-1\} + \operatorname{span}\{\alpha_n e_0 + e_n\},$$

where $n \geq 1$ and $(\alpha_n)_n \subseteq \mathbb{R}_+$ is some sequence of positive reals. Here only $\alpha_n e_0 + e_n$ is nonconforming and thus not involved in full algebraic consistency. If $\lim_{n\to\infty} \alpha_n = 0$, this direction becomes a new conforming direction, while for $\lim_{n\to\infty} \alpha_n = \infty$, it gets orthogonal to $V$. In any case, we have

$$S_n \cap V = \operatorname{span}\{e_i \mid i = 1, \dots, n-1\} \quad \text{and} \quad V = \overline{\bigcup_{n\geq 1} S_n}.$$

Moreover, straight-forward computations reveal that the orthogonal projections onto $S_n$ and $V$ are given by

$$\Pi_{S_n} v = \sum_{i=1}^{n-1} v_i e_i + \frac{v_n}{1+\alpha_n^2}(\alpha_n e_0 + e_n) \qquad \text{for } v \in V$$

$$\Pi_V s = \sum_{i=1}^{n} s_i e_i \qquad \text{for } s \in S.$$

One possibility to deal with the nonconforming direction $\alpha_n e_0 + e_n$ is to ignore it, e.g., by choosing methods with the approximation operators

$$P_{1,n} v = \sum_{i=1}^{n-1} v_i e_i \quad \text{for} \quad v \in V.$$

Each approximation operator $P_{1,n}$ is fully algebraically consistent and fully stable with $\|P_{1,n}\|_{\mathcal{L}(V,S)} = 1$. Moreover, $\Pi_V(\alpha_n e_0 + e_n) = e_n$ and $P_{1,n} e_n = 0$ yield

$$\delta_{S_n} \geq \frac{\|\overline{s}_n - P_{1,n}\Pi_V \overline{s}_n\|}{\|\overline{s}_n - \Pi_V \overline{s}_n\|} = \frac{\|\overline{s}_n\|}{\alpha_n \|e_0\|} = \frac{\sqrt{1+\alpha_n^2}}{\alpha_n} \geq \frac{1}{\alpha_n}.$$

with $\overline{s}_n := \alpha_n e_0 + e_n$. Consequently, letting $\alpha_n \to 0$ shows that $\delta_S$ can become arbitrarily large, while the stability constant attains its minimal value for the case $S_n \cap V \neq \{0\}$.

Given a sequence $(\beta_n)_n \subseteq \mathbb{R}_+$ of positive reals, the approximation operators

$$P_{2,n} v := \sum_{i=1}^{n-1} v_i e_i + \left(v_n + \frac{\beta_n}{1+\alpha_n^2} v_{n+1}\right)(\alpha_n e_0 + e_n) \quad \text{for} \quad v \in V$$

exploit the nonconforming direction $\alpha_n e_0 + e_n$. Again, each $P_{2,n}$ is fully algebraically consistent and fully stable. Here, since $P_{2,n}\Pi_V s = s$ for all $s \in S$, we have that $\delta_S = 0$, while

$$\|P_{2,n}\|_{\mathcal{L}(V,S)} \geq \frac{\|P_{2,n} e_{n+1}\|}{\|e_{n+1}\|} \geq \frac{\beta_n}{\sqrt{1+\alpha_n^2}}.$$

Thus, $\beta_n / \sqrt{1+\alpha_n^2} \to \infty$ shows that the stability constant can become arbitrarily large, while $\delta_S$ attains its minimal value 0.

*Remark* 1.3.12 (Asymptotic consistency). The preceding remark exemplifies that the exploitation of the nonconforming direction measured by $\delta_V$ and $\delta_S$ is relevant also 'in the limit' for sequences of discrete spaces and can be controlled via the uniform boundedness of the consistency measures.

We conclude this section with slight generalizations of Propositions 1.3.6 and 1.3.9.

*Remark* 1.3.13 (Consistency measures and non-quasi-optimality). Whenever the method underlying $P$ is not quasi-optimal, we may set $C_{\mathrm{qopt}} = \infty$. Similarly, if $\delta_V$ (or $\delta_S$) does not exist, we set $\delta_V = \infty$ (or $\delta_S = \infty$). Then

$$\delta_V = \infty \iff C_{\mathrm{qopt}} = \infty \qquad \text{and} \qquad \delta_S = \infty \implies C_{\mathrm{qopt}} = \infty$$

and, using standard conventions for $\infty$, the formulas in Propositions 1.3.6 and 1.3.9 hold irrespective of quasi-optimality.

## 1.4 The Structure of Quasi-Optimal Methods

The task of this section is to determine the structure of nonconforming methods that are quasi-optimal. In light of Theorem 1.3.2, this reduces to determine the structure of full stability and full algebraic consistency.

### 1.4.1 Extended Approximation Operator and Extended Bilinear Form

Our analysis of quasi-optimality in the previous section has been centered around the extended approximation operator $\widetilde{P}$. In this subsection we relate this key tool to the extended bilinear form $\widetilde{b}$ from Lemma 1.2.10 and, thus, more closely to the data $a$ and $(S, b, L)$ defining problem and method.

**Lemma 1.4.1** (Extensions of approximation operator and bilinear forms). *The approximation operator $P$ extends to a bounded linear projection $\widetilde{P}$ from $\widetilde{V}$ onto $S$ if and only if there exists a bounded common extension $\widetilde{b}$ of $b$ and $\langle LA\cdot, \cdot\rangle$ to $\widetilde{V} \times S$.*

*If one of the two extensions exists, we have the following generalization of the Galerkin orthogonality:*

$$\forall \widetilde{v} \in \widetilde{V}, \sigma \in S \quad \widetilde{b}(\widetilde{v} - \widetilde{P}\widetilde{v}, \sigma) = 0.$$

*Proof.* Assume $\widetilde{P}$ is a bounded linear projection from $\widetilde{V}$ onto $S$ extending $P$. Then

$$\widetilde{b}(\widetilde{v}, \sigma) := b(\widetilde{P}\widetilde{v}, \sigma) \tag{1.4.1}$$

defines a bounded bilinear form on $\widetilde{V} \times S$. Since $\widetilde{P}$ is a projection onto $S$, $\widetilde{b}$ is an extension of $b$. Furthermore, if $v \in V$ and $\sigma \in S$, then $\widetilde{P}_{|V} = P$ yields $\widetilde{b}(v, \sigma) = b(Pv, \sigma) = \langle LAv, \sigma\rangle$. Consequently, $\widetilde{b}$ extends also $\langle LA\cdot, \cdot\rangle$.

Conversely, assume that $\widetilde{b}$ is a bounded common extension of $b$ and $\langle LA\cdot, \cdot\rangle$ on $\widetilde{V} \times S$. Given $\widetilde{v} \in \widetilde{V}$, define $\widetilde{P}\widetilde{v}$ by

$$\widetilde{P}\widetilde{v} \in S \quad \text{such that} \quad \forall \sigma \in S \;\; b(\widetilde{P}\widetilde{v}, \sigma) = \widetilde{b}(\widetilde{v}, \sigma). \tag{1.4.2}$$

Since $b$ is a nondegenerate bilinear form on $S \times S$, the element $\widetilde{P}\widetilde{v}$ exists, is unique and depends on $\widetilde{v}$ linearly. The uniqueness and $\widetilde{b} = b$ on $S \times S$ give $\widetilde{P}_{|S} = \mathrm{Id}_S$. Using $\widetilde{b} = \langle LA\cdot, \cdot\rangle = b(P\cdot, \cdot)$ on $V \times S$, we obtain $\widetilde{P}_{|V} = P$. Finally, the boundedness of $\widetilde{b}$ entails the boundedness of $\widetilde{P}$ and the claimed equivalence is verified.

It remains to verify the generalized Galerkin orthogonality. If one of the two extensions exists, then the other one is given either by (1.4.1) or by (1.4.2), which both just restate the claimed generalization. □

The close relationship between the two extensions $\widetilde{P}$ and $\widetilde{b}$ suggests that the operator norm $\|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})}$ can be reformulated in terms of $\widetilde{b}$. To this end, the following lemma will be very useful, which in turn exploits the following fact from linear functional analysis; see, e.g., Brezis [24]. If $X$ and $Y$ are normed linear spaces, $T : X \to Y$ is a linear operator and $T^\star$ stands for its adjoint, then

$$T \text{ is bounded} \implies D(T^\star) = Y' \text{ with } \|T^\star\|_{\mathcal{L}(Y',X')} = \|T\|_{\mathcal{L}(X,Y)}. \tag{1.4.3}$$

**Lemma 1.4.2** ($b$-duality for energy norm on $S$)**.** *The nondegenerate bilinear form $b$ induces a norm on $S$ by*

$$\|\sigma\|_b := \|b(\cdot,\sigma)\|_{S'} = \sup_{s\in S, \|s\|=1} b(s,\sigma), \qquad \sigma \in S,$$

*satisfying*

$$\|s\| = \sup_{\sigma \in S} \frac{b(s,\sigma)}{\|\sigma\|_b}.$$

*Proof.* Obviously, $\|\cdot\|_b$ is a seminorm and definite thanks to the nondegeneracy of $b$. To verify the claimed identity, we observe

$$\sup_{s\in S}\sup_{\sigma\in S} \frac{b(s,\sigma)}{\|s\|\|\sigma\|_b} = \sup_{\sigma\in S}\sup_{s\in S} \frac{b(s,\sigma)}{\|s\|\|\sigma\|_b} = 1 \tag{1.4.4}$$

and

$$\inf_{s\in S}\sup_{\sigma\in S} \frac{b(s,\sigma)}{\|s\|\|\sigma\|_b} = \inf_{\sigma\in S}\sup_{s\in S} \frac{b(s,\sigma)}{\|s\|\|\sigma\|_b} = 1, \tag{1.4.5}$$

where the rightmost identities follow from the definition of $\|\cdot\|_b$ and the first equality in (1.4.5) follows from (1.4.3) applied to the inverse of $B$, the linear operator representing $b$. Combining (1.4.4) and (1.4.5), we see that

$$\sup_{\sigma\in S} \frac{b(s,\sigma)}{\|s\|\|\sigma\|_b} = 1$$

for all $s \in S$ and the claimed identity is verified. □

**Lemma 1.4.3** (Norms of extensions). *If one of the extensions in Lemma 1.4.1 exists, we have*

$$\|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} = \sup_{\sigma\in S} \frac{\|\widetilde{b}(\cdot,\sigma)\|_{\widetilde{V}'}}{\|b(\cdot,\sigma)\|_{S'}}$$

*with the 'extended' dual norm* $\|\ell\|_{\widetilde{V}'} := \sup_{\widetilde{v}\in\widetilde{V},\|\widetilde{v}\|=1}\langle \ell, \widetilde{v}\rangle$.

*Proof.* Applying Lemma 1.4.2, the generalized Galerkin orthogonality of Lemma 1.4.1 and the definition of the extended dual norm, we infer

$$\begin{aligned}\|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} &= \sup_{\widetilde{v}\in\widetilde{V}} \frac{\|\widetilde{P}\widetilde{v}\|}{\|\widetilde{v}\|} = \sup_{\widetilde{v}\in\widetilde{V},\sigma\in S} \frac{b(\widetilde{P}\widetilde{v},\sigma)}{\|\widetilde{v}\|\|\sigma\|_b} = \sup_{\widetilde{v}\in\widetilde{V},\sigma\in S} \frac{\widetilde{b}(\widetilde{v},\sigma)}{\|\widetilde{v}\|\|\sigma\|_b} \\ &= \sup_{\sigma\in S} \frac{\|\widetilde{b}(\cdot,\sigma)\|_{\widetilde{V}'}}{\|\sigma\|_b} = \sup_{\sigma\in S} \frac{\|\widetilde{b}(\cdot,\sigma)\|_{\widetilde{V}'}}{\|b(\cdot,\sigma)\|_{S'}}.\end{aligned}$$ □

Before closing this subsection, two remarks are in order.

*Remark* 1.4.4 (Alternative proof and formula). The proof of Lemma 1.4.3 may be alternatively based on a continuous counterpart of the norm $\|\cdot\|_b$ from Lemma 1.4.2; see Tantardini and Veeser [58, Theorem 2.1]. Using that approach, one derives also

$$\|\widetilde{P}\|_{\mathcal{L}(\widetilde{V})} = \sup_{s\in S,\|s\|=1} \inf_{\sigma\in S} \frac{\|\widetilde{b}(\cdot,\sigma)\|_{\widetilde{V}'}}{|b(s,\sigma)|}.$$

by duality.

*Remark* 1.4.5 (Reformulations of quasi-optimality). A suitable combination of Remarks 1.2.6 and 1.2.9, Lemmas 1.3.1 and 1.4.1 as well as Theorem 1.3.2 shows that the following statements are equivalent reformulations of quasi-optimality for a nonconforming method $M = (S, b, L)$ with approximation operator $P$:

(1.4.6a) $M$ is fully algebraically consistent and fully stable.

(1.4.6b) $Ps = s$ for all $s \in S \cap V$ and $P$ is bounded.

(1.4.6c) $P$ extends to a linear projection $\widetilde{P}$ from $\widetilde{V}$ onto $S$ that is bounded.

(1.4.6d) $b$ and $\langle LA\cdot,\cdot\rangle$ have a common extension $\widetilde{b}$ that is bounded.

(1.4.6e) $P$ is bounded and $b, P$ have extensions $\widetilde{b}, \widetilde{P}$ such that $\widetilde{b}(\widetilde{v} - \widetilde{P}\widetilde{v}, \sigma) = 0$ for all $\widetilde{v} \in \widetilde{V}$ and $\sigma \in S$.

It is worth observing that no additional regularity beyond the natural one in (1.2.1) is involved. All this illustrates that extensions, as developed in our approach, are a well-tuned tool in the analysis of the quasi-optimality of nonconforming methods.

### 1.4.2 The Structure of Full Stability

In this subsection we determine the structure of nonconforming methods that are fully stable.

To this end, (1.4.3) and the following facts of linear functional analysis will be basic: if $X$ and $Y$ are normed linear spaces and $T : X \to Y$ linear, then

$$\dim X < \infty \iff \text{all linear operators } X \to Y \text{ are bounded}, \tag{1.4.7}$$

$$\text{if } \dim X < \infty, \text{ then } T^{\star} \text{ surjective} \iff T \text{ injective}, \tag{1.4.8}$$

see, e.g., [24] and [25, p. 1418].

Let $M = (S, b, L)$ be a nonconforming method and recall that $M$ is fully stable if and only if the operator $M : V' \to S$ is bounded, where $V'$ and $S$ are equipped, respectively, with the dual and extended energy norm.

We claim that the full stability of $M$ hinges on the boundedness of $L$. In light of Remark 1.2.6, we may assume that $D(M) = D(L) = V'$. The equivalence (1.4.7) yields the following two consequences. First, the boundedness of $M : V' \to S$ is a true requirement, because its domain $V'$ has infinite dimension. Second, the critical operator in the composition $M = B^{-1}L$ from (1.2.5) is $L$. In fact, its domain $V'$ has infinite dimension, while the domain $S'$ of $B^{-1}$ has finite dimension. Consequently, a method $M$ is fully stable if and only if it is entire and the operator $L : V' \to S'$ is bounded.

Next, we characterize the class of bounded linear operators from $V'$ to $S'$ and derive first a necessary condition. Let $L : V' \to S'$ be linear and bounded. Owing to (1.4.3), its adjoint $L^{\star}$ is a bounded linear operator from $S''$ to $V''$. Since the spaces $S$ and $V$ are reflexive, we thus deduce the existence of a linear operator $E : S \to V$ such that

$$\forall \ell \in V', \sigma \in S \quad \langle L\ell, \sigma \rangle = \langle \ell, E\sigma \rangle \, . \tag{1.4.9}$$

Conversely, if $E : S \to V$ is a linear operator satisfying (1.4.9), then $L$ is bounded on $V'$ with $\|L\|_{\mathcal{L}(V',S')} = \|E\|_{\mathcal{L}(S,V)}$ by (1.4.3) and (1.4.7).

*Remark* 1.4.6 (Smoothing of $E$). Usually, the nonconformity $S \not\subseteq V$ arises from a lack of smoothness, e.g., across interelement boundaries in the case of finite element methods. The operator $E : S \to V$ may then be viewed as a smoothing operator.

The above observations prepare the following result, which is our first step towards the structure of quasi-optimal methods.

**Theorem 1.4.7** (Full stability and smoothing)**.** *A nonconforming method* $M = (S, b, L)$ *for* (1.2.1) *is fully stable if and only if* $L$ *is the adjoint of a linear smoothing operator* $E : S \to V$.

*The discrete problem for $\ell \in V'$ then reads*

$$\forall \sigma \in S \quad b(M\ell, \sigma) = \langle \ell, E\sigma \rangle \tag{1.4.10}$$

*and the stability constant satisfies*

$$C_{\text{stab}} = \|M\|_{\mathcal{L}(V',S)} = \sup_{\sigma \in S} \frac{\|E\sigma\|}{\|b(\cdot,\sigma)\|_{S'}}. \tag{1.4.11}$$

*Moreover, the range of $M$ is $S$ if and only if $E$ is injective.*

*Proof.* The observations preceding Theorem 1.4.7 show that $M$ is fully stable if and only if $L$ is the adjoint of a linear smoothing operator $E : S \to V$. Moreover, they provide the claimed form of the discrete problem via (1.4.9). The second equivalence readily follows from (1.4.8) and Remark 1.2.2.

To verify (1.4.11), we combine Lemma 1.4.2 with the following identity $\|v\| = \sup_{\ell \in V', \|\ell\|_{V'}=1} \langle \ell, v \rangle$, see, e.g., Brezis [24, Corollary 1.4]:

$$\begin{aligned} C_{\text{stab}} = \|M\|_{\mathcal{L}(V',S)} &= \sup_{\ell \in V'} \frac{\|M\ell\|}{\|\ell\|_{V'}} = \sup_{\ell \in V', \sigma \in S} \frac{b(M\ell,\sigma)}{\|\ell\|_{V'}\|\sigma\|_b} \\ &= \sup_{\sigma \in S, \ell \in V'} \frac{\langle \ell, E\sigma \rangle}{\|\ell\|_{V'}\|\sigma\|_b} = \sup_{\sigma \in S} \frac{\|E\sigma\|}{\|\sigma\|_b} = \sup_{\sigma \in S} \frac{\|E\sigma\|}{\|b(\cdot,\sigma)\|_{S'}}. \end{aligned}$$ □

Let us start the discussion of this result by considering a canonical choice for the smoother $E$.

*Remark* 1.4.8 (Trivial smoothing for conforming methods). Assume that the discrete space $S \subseteq V$ is conforming and consider the simplest choice $E = \mathrm{Id}_S$. For this classical case, (1.4.11) reduces to the well-known identity

$$C_{\text{stab}} = \sup_{\sigma \in S} \frac{\|\sigma\|}{\|b(\cdot,\sigma)\|_{S'}} = \left( \inf_{\sigma \in S} \sup_{s \in S} \frac{b(s,\sigma)}{\|s\|\|\sigma\|} \right)^{-1} = \left( \inf_{s \in S} \sup_{\sigma \in S} \frac{b(s,\sigma)}{\|s\|\|\sigma\|} \right)^{-1}.$$

*Remark* 1.4.9 (Failure of $\mathrm{Id}_S$). Let $S$ be a nonconforming discrete space with $S \not\subseteq V$. Then the choice $E = \mathrm{Id}_S$ is not compatible with full stability and so, in view of Theorem 1.3.2, not with quasi-optimality. Indeed, Theorem 1.4.7 shows that $E(S) \subseteq V$ is necessary for full stability. Consequently, the condition $Es = s$ entails $s \in S \cap V$ and thus produces a contradiction for any $s \in S \setminus V$. We therefore need to define $Es$ for $s \in S \setminus V$ differently, which, in view of the nature of $S$ and $V$ in applications, typically amounts to some kind of smoothing.

Most of the classical NCFEM and DG methods rely on the simple choice $E = \mathrm{Id}_S$, requiring that the load term $\ell$ in (1.2.1) has some additional regularity. Remark 1.4.9 implies that these methods are not fully stable and so, in view of Theorem 1.3.2, not quasi-optimal. This provides an alternative to falsify quasi-optimality with Remark 1.2.4.

We end this subsection by considering first alternatives to $E = \mathrm{Id}_S$ and illustrating that the choice of $E$ is in general a delicate matter.

*Remark* 1.4.10 (Previous uses of smoothing)*.* Advantages offered by suitable smoothing have been previously observed. An obvious one is that the method can be made entire and this has been pointed out, e.g., in the DG context by Di Pietro and Ern [37].

While comparing the Hellan-Hermann-Johnson method with the Morley method, Arnold and Brezzi [4] showed that a particular smoothing in the Morley method leads to an a priori error estimate requiring less regularity of the underlying load term. This corresponds to an increased stability thanks to the employed smoothing.

Also in the context of fourth order problems, Brenner and Sung [23] proposed $C^0$ interior penalty methods and proved a priori error estimates also for nonsmooth loads. Furthermore, the involved regularity is minimal from the viewpoint of approximation.

Finally, Badia et al. [8] used a rather involved smoother, which is related to our construction in Chapters 3 and 4, to show a partial quasi-optimality result for the Stokes problem.

*Remark* 1.4.11 (Smoothers into $S \cap V$)*.* It may look natural to use smoothers $E$ that map into the conforming part $S \cap V$ of the discrete space. In view of Remark 1.2.2, the range $R(M)$ of the corresponding method is a proper subspace of $S$, whenever $S \setminus V \neq \emptyset$. Quasi-optimality is then not ruled out, but it hinges on the validity of results like [60, Corollary 3.1] by Veeser and requires in particular that $S \cap V$ is not small.

*Remark* 1.4.12 (Optimal smoothing)*.* The structure of full stability does not principally exclude methods that are optimal from the viewpoint of approximation. Consequently, the variational crime of nonconformity does not necessarily result in some consistency error. To see this, consider the discrete bilinear form $b = \widetilde{a}_{|S \times S}$. Since

$$\forall v \in V, \sigma \in S \quad \widetilde{a}(Pv - v, \sigma) = \widetilde{a}(v, E\sigma - \sigma),$$

we have

$$P = \Pi_S \iff E = \Pi_V.$$

In other words: a nonconforming method $(S, \widetilde{a}_{S \times S}, E^\star)$ provides the best approximation if and only if the smoother $E$ is the $\widetilde{a}$-orthogonal projection onto $V$. However, this smoother is likely not feasible in the sense of the following remark.

*Remark* 1.4.13 (Feasible smoothing)*.* Adopt the notation of Remark 1.2.1 and let $\varphi_1, \dots, \varphi_n$ be a computionally convenient basis for the discrete bilinear form $b$. In order to compute $M\ell$ by (1.4.10) with optimal complexity, the total number of operations for evaluating $\langle \ell, E\varphi_i \rangle$ for all $i = 1, \dots, n$ has to be of order $O(n)$. A sufficient condition for this is that, for each $i = 1, \dots, n$, the function $E\varphi_i$ is locally supported so that $\langle \ell, E\varphi_i \rangle$ can be evaluated at cost $O(1)$.

### 1.4.3 The Structure of Quasi-Optimality

We are finally ready for the main results of our abstract analysis about the quasi-optimality of nonconforming methods.

**Theorem 1.4.14** (Quasi-optimality and smoothing). *A nonconforming method $M = (S, b, L)$ for* (1.2.1) *is quasi-optimal if and only if there exists a linear smoothing operator $E : S \to V$ such that the discrete problem reads*

$$\forall \sigma \in S \quad b(M\ell, \sigma) = \langle \ell, E\sigma \rangle$$

*for any $\ell \in V'$ and*

$$\forall u \in S \cap V, \sigma \in S \quad b(u, \sigma) = a(u, E\sigma). \tag{1.4.12}$$

*Its quasi-optimality constant is given by*

$$C_{\text{qopt}} = \sup_{\sigma \in S} \frac{\sup_{\|v+s\|=1} a(v, E\sigma) + b(s, \sigma)}{\sup_{\|s\|=1} b(s, \sigma)}, \tag{1.4.13}$$

*where $v$ varies in $V$ and $s$ in $S$.*

*Proof.* We first check the claimed equivalence. The form of the discrete problem means that $L$ is the adjoint of $E$ and, in view of Theorem 1.4.7, that $M$ is fully stable. Moreover, since

$$\langle LAu, \sigma \rangle = \langle Au, E\sigma \rangle = a(u, E\sigma) \tag{1.4.14}$$

for all $u \in V$ and $\sigma \in S$, (1.4.12) is equivalent to (1.2.8), i.e. full algebraic consistency. Consequently, the claimed equivalence follows from Theorem 1.3.2.

To show the identity for the quasi-optimality constant, we observe that the extension $\widetilde{b}$ exists and satisfies, for $\widetilde{v} \in \widetilde{V}$, $v \in V$, $s, \sigma \in S$ such that $\widetilde{v} = v + s$,

$$\widetilde{b}(\widetilde{v}, \sigma) = \langle LAv, \sigma \rangle + b(s, \sigma) = a(v, E\sigma) + b(s, \sigma)$$

thanks to (1.4.14). Therefore, formula (1.4.13) for $C_{\text{qopt}}$ follows from Theorem 1.3.2 and Lemma 1.4.3. □

We start the discussion about Theorem 1.4.14 by a remark about the notion of Galerkin methods.

*Remark* 1.4.15 (Galerkin methods). Assume first that the discrete space $S \subseteq V$ is conforming. Then trivial smoothing $E = \text{Id}_S$ in (1.4.12) yields $b = a_{|S \times S}$. In other words: conforming Galerkin methods are the only quasi-optimal methods with the simplest choice $E = \text{Id}_S$ for smoothing.

Next, consider a general nonconforming discrete space $S$, together with the simplest choice for smoothing in the conforming part $S \cap V$, i.e. with

$E_{|S\cap V} = \mathrm{Id}_{S\cap V}$. Here (1.4.12) yields $b_{|S_C\times S_C} = a_{|S_C\times S_C}$ where $S_C$ is an abbreviation for $S\cap V$. Thus, nonconforming Galerkin methods are the only candidates for quasi-optimal methods with $E_{|S\cap V} = \mathrm{Id}_{S\cap V}$. In this context, the following observation is useful in constructing $E$ with $E_{|S\cap V} = \mathrm{Id}_{S\cap V}$. If $E$ maps some $s \in S\setminus V$ in $S\cap V$, then the injectivity of $E$ is broken and, in view of Theorem 1.4.7, the range of the method is a strict subspace of $S$.

*Remark* 1.4.16 (Comparison with second Strang lemma). For conforming Galerkin methods, Theorem 1.4.14 reduces to the well-known Céa lemma, with $C_{\mathrm{qopt}} = 1$. Céa's lemma is a basic building block in the analysis of the energy norm error for conforming methods. In the context of nonconforming methods, the second Strang lemma is often used as a replacement. Theorem 1.4.14 provides a specialization revealing the structure of quasi-optimal methods and so lays the groundwork for their design.

*Remark* 1.4.17 (Comparison with conforming Petrov-Galerkin methods). Our setting of §1.2.1 includes the application of Petrov-Galerkin methods to (1.2.1). It is therefore of interest to compare formula (1.4.13) with its conforming counterpart in [58, Theorem 2.1] by Tantardini and Veeser:

$$C_{\mathrm{qopt}} = \sup_{\sigma\in S} \frac{\sup_{\|v\|=1} b(v,\sigma)}{\sup_{\|s\|=1} b(s,\sigma)},$$

where here $b$ stands for the continuous (and discrete) bilinear form, $v$, $s$, and $\sigma$ vary, respectively, in the continuous trial space, in the discrete trial space and in the discrete test space. We see that (1.4.13) generalizes this formula, replacing the continuous bilinear form by the extended one, which interweaves discrete and continuous problems.

*Remark* 1.4.18 ('Classical' bound for quasi-optimality constant). A consequence of the formula for the quasi-optimality constant in Theorem 1.4.14 and (1.4.3) is the following upper bound:

$$C_{\mathrm{qopt}} \le \frac{C_{\widetilde{b}}}{\beta} \tag{1.4.15}$$

with the continuity and inf-sup constants

$$C_{\widetilde{b}} := \sup_{\|v+s\|=1, \|\sigma\|=1} a(v, E\sigma) + b(s,\sigma), \qquad \beta := \inf_{\|s\|=1} \sup_{\|\sigma\|=1} b(s,\sigma),$$

where $v$ varies in $V$ and $s$ and $\sigma$ in $S$. This upper bound has the classical form of constants appearing in quasi-optimality results, apart from the slight difference that the continuity constant of the numerator involves the extended bilinear form; see also Remark 1.4.17.

It is worth mentioning that the right-hand side of (1.4.15) can become arbitrarily large, while its left-hand side remains bounded, as we point out in Remark 3.2.7.

Let us now assess what determines the size of the quasi-optimality constant.

**Theorem 1.4.19** (Size of quasi-optimality constant)**.** *Assume $M = (S, b, L)$ is a quasi-optimal nonconforming method with linear smoother $E : S \to V$ and stability constant $C_{\text{stab}}$. The consistency measure $\delta_V$ introduced in Proposition 1.3.6 is finite and is*

$$\delta_V = \sup_{v\in V,\, \Pi_S v\neq v} \ \sup_{\sigma\in S} \frac{b(\Pi_S v, \sigma) - a(v, E\sigma)}{\|\Pi_S v - v\| \|b(\cdot, \sigma)\|_{S'}}. \tag{1.4.16}$$

*Similarly, the consistency measure $\delta_S$ introduced in Proposition 1.3.9 is finite and the smallest positive constant such that*

$$\forall s \in S \quad \sup_{\sigma\in S} \frac{b(s, \sigma) - a(\Pi_V s, E\sigma)}{\|b(\cdot, \sigma)\|_{S'}} \le \delta_S \|s - \Pi_V s\|.$$

*Then the quasi-optimality constant of $M$ satisfies*

$$\max\{C_{\text{stab}}, \delta_S\} \le C_{\text{qopt}} = \sqrt{1 + \delta_V^2} \le \sqrt{C_{\text{stab}}^2 + \delta_S^2}.$$

*Proof.* Lemma 1.4.2 readily yields the identities

$$\|\Pi_S v - Pv\| = \sup_{\sigma\in S} \frac{b(\Pi_S v - Pv, \sigma)}{\|\sigma\|_b}$$

$$\|s - P\Pi_V s\| = \sup_{\sigma\in S} \frac{b(s - P\Pi_V s, \sigma)}{\|\sigma\|_b}.$$

We have also the identities $b(Pv, \sigma) = b(MAv, \sigma) = \langle LAv, \sigma\rangle = a(v, E\sigma)$ and $\|\sigma\|_b = \|b(\cdot, \sigma)\|_{S'}$ for $v \in V$ and $\sigma \in S$ as well as $V \setminus S \neq \emptyset$. Therefore, $\delta_V$ and $\delta_S$ coincide with the corresponding quantities in Propositions 1.3.6 and 1.3.9 and Theorem 1.4.19 just restates their conclusions. □

We refer to §1.3.2 for a discussion of the relationship between $C_{\text{qopt}}$ and $C_{\text{stab}}$ and in particular the consistency measures $\delta_V$ and $\delta_S$. Let us further connect the expression of $\delta_V$ in this theorem with classical consistency.

*Remark* 1.4.20 ($\delta_V$ and classical consistency error). It is worth noticing that the numerator of (1.4.16) represents the action of a linear functional on $S$,

$$b(\Pi_S v, \sigma) - a(v, E\sigma) = \langle B\Pi_S v - LAv, \sigma\rangle =: \langle \rho, \sigma\rangle .$$

Let us recall that $LAv$ is the discrete load associated with $v$ in problem (1.2.3) and $B\Pi_S v$ is the linear functional obtained from the representative $\Pi_S v$ of $v$ in $S$, through the isomorphism $B$. Introducing the norm $\|\cdot\|_{S',b} := \sup_{\|b(\cdot,\sigma)\|_{S'}=1} \langle\cdot, \sigma\rangle$, the quantity $\|\rho\|_{S',b}$ is a consistency error in the sense of Arnold [3]. The measure $\delta_V$ compares this quantity with the natural benchmark in the context of quasi-optimality, i.e. the best error $\|v - \Pi_S v\|$.

# Chapter 2

# Some Preliminaries to Finite Elements

The goal of this chapter is to provide the necessary preliminaries to the application of the abstract results from §1 to the design and analysis of nonconforming finite element methods. For this purpose, we devote the first part of Section 2.1 to fix our assumptions and notations concerning the spatial domain and the mesh. Then, we introduce spaces of piecewise regular and piecewise polynomial functions and recall some of their basic properties. In Section 2.2 we propose nodal averaging operators for discontinuous piecewise polynomials.

## 2.1 Simplicial Meshes and (Broken) Spaces

We indicate Lebesgue and Sobolev spaces as usual, see, e.g., [21], and adopt the following notations.

For $n \in \{0, \dots, d\}$, an *$n$-simplex* $C \subseteq \mathbb{R}^d$ is the convex hull of $n+1$ points $z_1, \dots, z_{n+1} \in \mathbb{R}^d$ spanning an $n$-dimensional affine space. The uniquely determined points $z_1, \dots, z_{n+1}$ are the vertices of $C$ and form the set $\mathcal{L}_1(C)$. If $n \geq 1$, we let $\mathcal{F}_C$ denote the $(n-1)$-dimensional faces of $C$, which are the $(n-1)$-simplices arising by picking $n$ distinct vertices from $\mathcal{L}_1(C)$. Given a vertex $z \in \mathcal{L}_1(C)$, its barycentric coordinate $\lambda_z^C$ is the unique first order polynomial on $C$ such that $\lambda_z^C(y) = \delta_{zy}$ for all $y \in \mathcal{L}_1(C)$. Then $0 \leq \lambda_z^C \leq 1$ and $\sum_{z \in \mathcal{L}_1(C)} \lambda_z^C = 1$ in $C$ and, if $\alpha = (\alpha_z)_{z \in \mathcal{L}_1(C)} \in \mathbb{N}_0^{n+1}$ is multi-index,

$$\int_C \prod_{z \in \mathcal{L}_1(C)} (\lambda_z^C)^{\alpha_z} = \frac{n!\alpha!}{(n+|\alpha|)!} |C| , \tag{2.1.1}$$

where $|C|$ stands also for the $n$-dimensional Hausdorff measure in $\mathbb{R}^d$. We write $h_C := \operatorname{diam}(C)$ for the diameter of $C$, $\rho_C$ for the diameter of its largest inscribed $n$-dimensional ball, and $\gamma_C$ for its shape coefficient $\gamma_C := h_C/\rho_C$.

Let $\mathcal{M}$ be a simplicial, face-to-face *mesh* of some open, bounded, connected and polyhedral set $\Omega \subset \mathbb{R}^d$ with Lipschitz boundary $\partial\Omega$. More precisely, $\mathcal{M}$ is a finite collection of $d$-simplices in $\mathbb{R}^d$ such that $\overline{\Omega} = \bigcup_{K\in\mathcal{M}} K$ and the intersection of two arbitrary elements $K_1, K_2 \in \mathcal{M}$ is either empty or an $n$-simplex with $n \in \{0\dots,d\}$ and $\mathcal{L}_1(K_1 \cap K_2) = \mathcal{L}_1(K_1) \cap \mathcal{L}_1(K_2)$. We let $\mathcal{F} := \bigcup_{K\in\mathcal{M}} \mathcal{F}_K$ denote the $(d-1)$-dimensional faces of $\mathcal{M}$ and distinguish between boundary faces $\mathcal{F}^b := \{F \in \mathcal{F} \mid F \subseteq \partial\Omega\}$ and interior faces $\mathcal{F}^i := \mathcal{F} \setminus \mathcal{F}^b$. Moreover, let $\Sigma := \cup_{F\in\mathcal{F}} F$ be the skeleton of $\mathcal{M}$ and, fixing a unit normal $n_F$ for each interior face $F \in \mathcal{F}^i$, extend the outer normal $n$ of $\partial\Omega$ to $\Sigma$ by $n_{|F} = n_F$ for $F \in \mathcal{F}^i$. The ambiguity in the orientation of $n_F$ is insignificant to our discussion. The meshsize $h$ on $\Sigma$ is given by $h_{|F} = h_F$ for all $F \in \mathcal{F}$ and the shape coefficient of $\mathcal{M}$ is

$$\gamma_{\mathcal{M}} := \max_{K\in\mathcal{M}} \gamma_K.$$

For $k \in \mathbb{N}$, the broken Sobolev space of order $k$ is

$$H^k(\mathcal{M}) := \{v \in L^2(\Omega) \mid \forall K \in \mathcal{M}\ v_{|K} \in H^k(K)\}.$$

If $v \in H^k(\mathcal{M})$, we use the subscript $\mathcal{M}$ to indicate the piecewise variant of a differential operator. For instance, $\nabla_{\mathcal{M}} v$ is given by $(\nabla_{\mathcal{M}} v)_{|K} := \nabla(v_{|K})$ for all $K \in \mathcal{M}$. Jumps and averages are defined as follows. Given an interior face $F \in \mathcal{F}^i$, let $K_1, K_2 \in \mathcal{M}$ be the two elements such that $F = K_1 \cap K_2$ and the outer normal of $K_1$ coincides with $n$. Set

$$[\![v]\!] := v_{|K_1} - v_{|K_2}, \quad \{\!\{v\}\!\} := \frac{1}{2}\left(v_{|K_1} + v_{|K_2}\right) \quad \text{on } F. \tag{2.1.2a}$$

Again, the fact that the signs of $[\![v]\!]$ depends on the ordering of $K_1$ and $K_2$ will be insignificant to our discussion. Instead, it will be convenient to extend these definitions on $\partial\Omega$. Given $F \in \mathcal{F}^b$, let $K \in \mathcal{M}$ be the element such that $F = K \cap \partial\Omega$ and set

$$[\![v]\!] := \{\!\{v\}\!\} := v_{|K} \quad \text{on } F. \tag{2.1.2b}$$

In this notation, piecewise integration by parts reads as follows: if we have $v, w \in H^1(\mathcal{M})$ and $j \in \{1, \dots, d\}$, then

$$\begin{aligned}\int_\Omega (\partial_{j,\mathcal{M}} v) w - \int_{\Sigma\setminus\partial\Omega} [\![v]\!]\, \{\!\{w\}\!\}\, n \cdot e_j &= \\ = -\int_\Omega v(\partial_{j,\mathcal{M}} w) + \int_{\Sigma\setminus\partial\Omega} \{\!\{v\}\!\}\, [\![w]\!]\, n \cdot e_j &+ \int_{\partial\Omega} vw\, n \cdot e_j.\end{aligned} \tag{2.1.3}$$

Notice that the surface integrals are independent of the orientation of $n$ and that, e.g., the singular part of the distributional derivative $\partial_j v$ is represented by means of the negative jumps $-[\![v]\!]$, $F \in \mathcal{F}^i$.

Given $p \in \mathbb{N}_0$, we write $\mathbb{P}_p(C)$ for the linear space of *polynomials* on the $n$-simplex $C$ with (total) degree $\leq p$. Consider $p \in \mathbb{N}$, excluding the trivial case $p = 0$. A polynomial in $\mathbb{P}_p(C)$ is determined by its point values at the Lagrange nodes $\mathcal{L}_p(C)$ of order $p$, which, for $p \geq 2$, are given by

$$\mathcal{L}_p(C) := \left\{ x \in C \mid \forall z \in \mathcal{L}_1(C)\ p\lambda_z^C(x) \in \mathbb{N}_0 \right\}$$

We let $\Psi^p_{C,z}$, $z \in \mathcal{L}_p(C)$, denote the associated nodal basis in $\mathbb{P}_p(C)$ given by $\Psi^p_{C,z}(y) = \delta_{zy}$ for all $y, z \in \mathcal{L}_p(C)$. The Lagrange nodes are nested in that $\mathcal{L}_p(F) = \mathcal{L}_p(C) \cap F$ for any face $F \in \mathcal{F}_C$. Thus, the restriction $P_{|F}$ of $P \in \mathbb{P}_p(C)$ is determined by the 'restriction' $\mathcal{L}_p(C) \cap F$ of the Lagrange nodes and we have $\Psi^p_{C,z|F} = \Psi^p_{F,z}$ for all $z \in \mathcal{L}_p(F)$.

Given $k, p \in \mathbb{N}_0$, the space of functions that are *piecewise polynomial* with degree $\leq p$ and are in $H^k_0(\Omega)$ is

$$S^k_p := \Big\{ s \in H^k_0(\Omega) \mid \forall K \in \mathcal{M}\ s_{|K} \in \mathbb{P}_p(K) \Big\}. \tag{2.1.4}$$

The cases $p \in \mathbb{N}$ with $k \in \{0, 1\}$ are of particular interest.

Consider first $S^0_p$ with $p \in \mathbb{N}$ and extend each $\Psi^p_{K,z}$ outside of $K \in \mathcal{M}$ by 0. The functions $\{\Psi^p_{K,z}\}_{K \in \mathcal{M}, z \in \mathcal{L}_p(K)}$ form a basis of $S^0_p$ with $\Psi_{K,z|K'}(z') = \delta_{K,K'}\delta_{z,z'}$ for $K, K' \in \mathcal{M}$ and $z \in \mathcal{L}_p(K)$, $z' \in \mathcal{L}_p(K')$, which amounts to distinguishing Lagrange nodes from different elements.

The construction of a basis of $S^1_p$ is a little more involved. Here, identifying coinciding Lagrange nodes, we set $\mathcal{L}_p := \cup_{K \in \mathcal{M}} \mathcal{L}_p(K)$ as well as $\mathcal{L}^i_p := \mathcal{L}_p \setminus \partial\Omega$, and write $\Phi^p_z$, $z \in \mathcal{L}_p$, for the function given piecewise by $\Phi^p_{z|K} := \Psi^p_{K,z}$ if $z \in K$ and $\Phi^p_{z|K} := 0$ otherwise. Then the nestedness of Lagrange nodes implies: $\{\Phi^p_z\}_{z \in \mathcal{L}^i_p}$ is a basis of $S^1_p$ satisfying $\Phi^p_z(y) = \delta_{zy}$ for all $y, z \in \mathcal{L}^i_p$. In connection with these basis functions, the following subdomains are useful. Let $\omega_z := \bigcup_{K' \ni z} K'$ be the star around $z \in \mathcal{L}_p$ and let $\omega_K := \bigcup_{K' \cap K \neq \emptyset} K'$ be the patch around $K \in \mathcal{M}$. Since $\partial\Omega$ is Lipschitz, stars are face-connected in the sense of [60]: given $z \in \mathcal{L}_p$ and any pair $K, K' \in \mathcal{M}$ with $z \in K \cap K'$, there exists a path $\{K_i\}_{i=1}^n \subset \mathcal{M}$ of elements containing $z$ such that $K_1 = K$, $K_n = K'$, and each $K_i \cap K_{i+1} \in \mathcal{F}^i$.

If not specified differently, $C_*$ stands for a function which is not necessarily the same at each occurrence and depends on a subset $*$ of $\{d, \gamma_\mathcal{M}, p\}$, increasing in $\gamma_\mathcal{M}$ and $p$ if present. For instance, we have, for $K, K' \in \mathcal{M}$,

$$K \cap K' \neq \emptyset \quad \Longrightarrow \quad |K| \leq C_{\gamma_\mathcal{M}} \left|K'\right| \text{ and } h_K \leq C_{\gamma_\mathcal{M}} \rho_{K'} \tag{2.1.5}$$

and, for $p \in \mathbb{N}$, $K \in \mathcal{M}$, and $z \in \mathcal{L}_p(K)$,

$$c_{d,p}|K|^{\frac{1}{2}} h_K^{-1} \leq \|\nabla \Psi^p_{K,z}\|_{L^2(K)} \leq C_{d,p}|K|^{\frac{1}{2}} \rho_K^{-1}. \tag{2.1.6}$$

If there is no danger of confusion, $A \leq C_* B$ may be abbreviated as $A \lesssim B$.

## 2.2 Averaging Operators

Let $\Omega \subseteq \mathbb{R}^d$ and $\mathcal{M}$ be as in the previous section. For all $p \geq 1$, the spaces $S^0_p$ and $S^1_p$ are connected by the following *simplified averaging operator* $A_p : S^0_p \to S^1_p$, based upon evaluating at Lagrange nodes. For every interior node $z \in \mathcal{L}^i_p$, fix some element $K_z \in \mathcal{M}$ containing $z$ and set

$$A_p\sigma := \textstyle\sum_{z\in\mathcal{L}^i_p} \sigma_{|K_z}(z)\Phi^p_z, \qquad \sigma \in S^0_p. \tag{2.2.1}$$

Clearly, $A_p\sigma(z) = \sigma(z)$ whenever $\sigma$ is continuous at $z \in \mathcal{L}^i_p$ and so $A_p$ is a projection onto $S^1_p$. On the one hand, the operator $A_p$ is a restriction of Scott-Zhang interpolation [56] defined for broken $H^1$-functions and, on the other hand, it is a simplified variant of the standard nodal averaging in (2.2.8) below, in that it requires only one evaluation per degree of freedom. Standard nodal averaging has been used in various nonconforming contexts, see, e.g., Brenner [13], Karakashian/Pascal [46], Oswald [51]. Our interest in the operator $A_p$ is motivated by the application of Theorems 1.4.7 and 1.4.14 to second-order elliptic problems. Indeed, we have $A_p\sigma \in H^1_0(\Omega)$ for all $\sigma \in S^0_p$, in view of (2.1.4), as well as the stability bounds below. All results in this section are tailored to our subsequent use and essentially known in the literature, so that proofs are only intended for completeness. More details about averaging operators can be found, for instance, in [26], [39] and references therein.

**Lemma 2.2.1** (Simplified nodal averaging and $L^2$-norms of jumps)**.** *Let $p \geq 1$, $\sigma \in S^0_p$ piecewise polynomial, $K \in \mathcal{M}$, and $z \in \mathcal{L}_p(K)$ be a Lagrange node. If $z \notin \partial K$, then $A_p\sigma(z) = \sigma_{|K}(z)$, else*

$$\left|\sigma_{|K}(z) - A_p\sigma(z)\right| \leq C_{d,p} \sum_{F\in\mathcal{F}:F\ni z} \frac{1}{|F|^{\frac{1}{2}}} \| \llbracket\sigma\rrbracket \|_{L^2(F)}. \tag{2.2.2}$$

*Proof.* The 'then'-part of the claim readily follows from the non-overlapping of elements in $\mathcal{M}$. In order to show the 'else'-part, we start by claiming that, for any $z \in \partial K$,

$$\left|\sigma_{|K}(z) - A_p\sigma(z)\right| \leq \sum_{F\ni z} \left|\llbracket\sigma\rrbracket(z)\right| \tag{2.2.3}$$

where $F$ varies in $\mathcal{F}$. To verify this, we shall exploit that $\mathcal{M}$ has face-connected stars in the sense of [60], distinguishing the cases $z \in \Omega$ and $z \in \partial\Omega$. If $z \in \Omega$ is an interior node, we choose a path $(K'_j)^n_{j=0}$ in $\omega_z$ such that $K'_0 = K$, $K'_n = K_z$ and $K'_{j-1} \cap K'_j =: F_j \in \mathcal{F}^i$ for $j = 1, \dots n$. Then we bound the telescopic sum $\sigma_{|K}(z) - A_p(z) = \sum^n_{j=1} \sigma_{|K_{j-1}}(z) - \sigma_{|K_j}(z)$ with the triangle inequality, independently of the choice of the path and $K_z$ and obtain (2.2.3). If $z \in \partial\Omega$ is a boundary node, we proceed similarly but terminate the path with an element $K_b \in \mathcal{M}$ that has a boundary face

$F \in \mathcal{F}^b$ and use the identity $\sigma_{|K_b}(z) - A_p(z) = \sigma_{|K_b}(z) = [\![\sigma]\!](z)$. Finally, we apply the inverse inequality $\|\cdot\|_{L^\infty(F)} \leq C_{d,p}\,|F|^{-\frac{1}{2}}\,\|\cdot\|_{L^2(F)}$ in $\mathbb{P}_p(F)$ to (2.2.3) conclude the proof. □

The estimate established in this lemma is particularly convenient for the applications to interior penalty methods discussed in Chapter 4. Instead, the definition of the nonconforming elements in Chapter 3, like e.g. the Crouzeix-Raviart element, suggests to derive an alternative bound for the left-hand side of (2.2.2), not involving the scaled $L^2-$norms of jumps on the mesh faces. To this end, let us preliminarily recall the following trace identity, see e.g. [61, Proposition 4.2]. For all elements $K \in \mathcal{M}$ and faces $F \in \mathcal{F}_K$, we have

$$\forall v \in H^1(K) \qquad \frac{1}{|F|}\int_F v = \frac{1}{|K|}\int_K v + \frac{1}{|K|}\int_K \nabla v \cdot \xi_F \tag{2.2.4}$$

where $\xi_F(x) := (x - z_F)/d$ and $z_F \in \mathcal{L}_1(K)$ denotes the vertex of $K$ not belonging to $F$.

**Lemma 2.2.2** (Simplified nodal averaging and integrals of jumps). *Let $p \geq 1$, $\sigma \in S_p^0$ piecewise polynomial, $K \in \mathcal{M}$, and $z \in \mathcal{L}_p(K)$ a Lagrange node. If $z \notin \partial K$, then $A_p\sigma(z) = \sigma_{|K}(z)$, else*

$$\left|\sigma_{|K}(z) - A_p\sigma(z)\right| \leq \sum_{F \ni z} \frac{1}{|F|}\left|\int_F [\![\sigma]\!]\right| + C_{d,p}\sum_{K' \ni z} \frac{h_{K'}}{|K'|^{\frac{1}{2}}}\|\nabla\sigma\|_{L^2(K')},$$

*where $F$ and $K'$ vary in $\mathcal{F}$ and $\mathcal{M}$, respectively.*

*Proof.* As before, the 'then'-part of the claim readily follows from the non-overlapping of elements in $\mathcal{M}$. In order to show the 'else'-part, we bound each jump in (2.2.3) suitably. To this end, we consider two cases, $F \in \mathcal{F}^i$ and $F \in \mathcal{F}^b$, and start with the first one. Let $K_1, K_2 \in \mathcal{M}$ be the two elements such that $F = K_1 \cap K_2$. We insert the face means $f_j := |F|^{-1}\int_F \sigma_{|K_j}$ as well as the element means $k_j := |K_j|^{-1}\int_{K_j}\sigma$, $j = 1, 2$. Using an inverse estimate in $\mathbb{P}_p(F)$, we deduce

$$\left|[\![\sigma]\!](z)\right| \leq \frac{1}{|F|}\left|\int_F [\![\sigma]\!]\right| + \sum_{j=1,2}\left(|f_j - k_j| + \frac{C_{d,p}}{|F|^{\frac{1}{2}}}\|\sigma_{|K_j} - k_j\|_{L^2(F)}\right). \tag{2.2.5}$$

For $j = 1, 2$, the trace identity (2.2.4) gives

$$|f_j - k_j| \leq \frac{h_{K_j}}{d\,|K_j|}\|\nabla\sigma\|_{L^1(K_j)} \leq \frac{h_{K_j}}{d\,|K_j|^{\frac{1}{2}}}\|\nabla\sigma\|_{L^2(K_j)},$$

while [60, Lemma 3], which is a combination of the trace identity and the Poincaré inequality, provides

$$|F|^{-\frac{1}{2}} \|\sigma_{|K_j} - k_j\|_{L^2(F)} \leq \sqrt{\frac{1}{\pi^2} + \frac{2}{\pi d}} \frac{h_{K_j}}{|K_j|^{\frac{1}{2}}} \|\nabla\sigma\|_{L^2(K_j)}.$$

Inserting the last two inequalities in (2.2.5), we arrive at

$$|\llbracket\sigma\rrbracket(z)| \leq \frac{1}{|F|} \left| \int_F \llbracket\sigma\rrbracket \right| + C_{d,p} \sum_{j=1}^{2} \frac{h_{K_j}}{|K_j|^{\frac{1}{2}}} \|\nabla\sigma\|_{L^2(K_j)} \tag{2.2.6a}$$

in this case. If, instead, $F \in \mathcal{F}^b$, we denote by $K \in \mathcal{M}$ the element with $F = K \cap \partial\Omega$ and, using the means $f := |F|^{-1} \int_F \sigma_{|K}$ and $k := |K|^{-1} \int_K \sigma$ similarly as before, we obtain

$$|\llbracket\sigma\rrbracket(z)| \leq \frac{1}{|F|} \left| \int_F \llbracket\sigma\rrbracket \right| + C_{d,p} \frac{h_K}{|K|^{\frac{1}{2}}} \|\nabla\sigma\|_{L^2(K)}. \tag{2.2.6b}$$

Inserting (2.2.6) into (2.2.3) then finishes the proof. □

Assume that $p \geq 2$. The first-order simplified nodal averaging $A_1$ is naturally defined also on $S_p^0$ and is cheaper to evaluate than $A_p$. Since $A_1$ is not a projection onto $S_p^1$, it cannot fulfill an estimate like (2.2.2) in Lemma 2.2.1. Nevertheless, the following counterpart of Lemma 2.2.2 holds.

**Lemma 2.2.3** (First-order averaging on $S_p^0$)**.** *Let $p \geq 2$, $\sigma \in S_p^0$ piecewise polynomial, $K \in \mathcal{M}$, and $F \in \mathcal{F}_K$. For all Lagrange nodes $z \in \mathcal{L}_p(K) \cap F$ we have*

$$\begin{aligned} &\left|\sigma_{|K}(z) - A_1\sigma(z)\right| \leq \\ &\qquad \leq C_{d,p} \left( \sum_{F' \cap F \neq \emptyset} \frac{1}{|F'|} \left| \int_{F'} \llbracket\sigma\rrbracket \right| + \sum_{K' \cap F \neq \emptyset} \frac{h_{K'}}{|K'|^{\frac{1}{2}}} \|\nabla\sigma\|_{L^2(K')} \right) \end{aligned}$$

*where $F'$ and $K'$ vary in $\mathcal{F}$ and $\mathcal{M}$, respectively.*

*Proof.* We distinguish two cases, depending whether or not $z$ is a vertex.

*Case 1:* $z \in \mathcal{L}_1(K)$. Then we have $A_1\sigma(z) = A_p\sigma(z)$ and the claimed estimate follows from Lemma 2.2.2.

*Case 2:* $z \in \mathcal{L}_p(K)\backslash\mathcal{L}_1(K)$. Since $A_1\sigma_{|F} \in \mathbb{P}_1(F)$ and $\sum_{y\in\mathcal{L}_1(F)} \lambda_y^K = 1$, we may write

$$|\sigma_{|K}(z) - A_1\sigma(z)| \leq \textstyle\sum_{y\in\mathcal{L}_1(F)} \left|\sigma_{|K}(z) - A_1\sigma(y)\right| \lambda_y^K(z) \tag{2.2.7}$$

and, for any $y \in \mathcal{L}_1(F)$,

$$\left|\sigma_{|K}(z) - A_1\sigma(y)\right| \leq \left|\sigma_{|K}(z) - \sigma_{|K}(y)\right| + \left|\sigma_{|K}(y) - A_1\sigma(y)\right|.$$

As the second term of the right-hand side is already bounded in Case 1, it remains to bound the first term. Writing $c$ for the mean value of $\sigma$ in $K$, we deduce

$$\begin{aligned}\left|\sigma_{|K}(z) - \sigma_{|K}(y)\right| &\leq \left|\sigma_{|K}(z) - c\right| + \left|\sigma_{|K}(y) - c\right| \\ &\lesssim |F|^{-\frac{1}{2}} \left\|\sigma_{|K} - c\right\|_{L^2(F)} \lesssim h_K |K|^{-\frac{1}{2}} \|\nabla\sigma\|_{L^2(K)}\end{aligned}$$

with the help of an inverse estimate in $\mathbb{P}_p(F)$ and [60, Lemma 3]. □

*Remark* 2.2.4 (Evaluation of nodal averaging operators). It is worth noticing that the previous lemmas hold also if we replace $A_p$ with the standard nodal averaging operator $\widetilde{A_p} : S^0_p \to S^1_p$ given by

$$\widetilde{A_p}\sigma := \sum_{z \in \mathcal{L}_p} \left( \frac{1}{\#\omega_z} \sum_{K \ni z} \sigma_{|K}(z) \right) \Phi^p_z, \qquad \sigma \in S^0_p \tag{2.2.8}$$

where $p \geq 1$, $K$ varies in $\mathcal{M}$ and $\#\omega_z$ denotes the number of mesh elements in the star $\omega_z$. Nonetheless, the fact that evaluating $A_p\sigma$ is cheaper that $\widetilde{A_p}\sigma$ may be attractive when the averaging is employed as building-block in the construction of smoothing operators for quasi-optimal methods; see Chapters 3 and 4 below. In this case, the use of $A_p$ instead of $\widetilde{A_p}$ makes no difference in the abstract analysis but can save a considerable number of operations when assembling the right-hand side of the discrete problem (1.4.10). For the same reason, we may even consider the possibility of using $A_1$ in place of $A_p$ for $p \geq 2$. Possible effects of these variants are briefly discussed in Remarks 3.3.6 and 3.3.11, Theorem 4.2.12 and §5.3.1.

Averaging operators into $H^2_0$−conforming spaces are also of interest, when dealing with fourth-order problems. Unlike the previous case, we do not base our construction on the spaces $H^2(\Omega) \cap S^0_p$, due to their complicated structure [49]. Also, we only consider piecewise quadratic polynomials in the two-dimensional case $d = 2$. The Hsieh-Clough-Tocher (HCT) space with boundary conditions is [33]

$$\begin{aligned} HCT := \{s \in C^1(\overline{\Omega}) \mid \forall K \in \mathcal{M} \quad & s_{|K} \in C^1(K) \cap \mathbb{P}_3(\mathcal{M}_K), \\ & s = \partial_n s = 0 \text{ on } \partial\Omega \ \}, \end{aligned} \tag{2.2.9}$$

where $\mathcal{M}_K$ stands for the triangulation obtained by connecting each vertex of the triangle $K$ with its barycenter $m_K$. Notice that $HCT \subseteq H^2_0(\Omega)$ and every element $s \in HCT$ is uniquely determined by the values $s(z)$, $\nabla s(z)$ at the Lagrange nodes $z \in \mathcal{L}^i_1$ and $\nabla s(m_F) \cdot n_F$ at the midpoints $m_F$ of the interior edges $F \in \mathcal{F}^i$; see [21]. Then, for each vertex $z \in \mathcal{L}^i_1$ and edge $F \in \mathcal{F}^i$, we pick elements $K_z, K_F \in \mathcal{M}$ containing $z$ or $F$, respectively, and define $A_{HCT} : S_2^0 \to HCT$ as

$$\begin{aligned} A_{HCT}\sigma := \sum_{z\in\mathcal{L}_1^i} \left( \sigma_{|K_z}(z)\Upsilon_z^0 + \sum_{j=1}^{2} \partial_j\big(\sigma_{|K_z}\big)(z)\Upsilon_z^j \right) + \\ + \sum_{F\in\mathcal{F}^i} \frac{\partial\big(\sigma_{|K_F}\big)}{\partial n}(m_F)\Upsilon_F, \end{aligned} \tag{2.2.10}$$

where $\Upsilon_z^j$ with $z \in \mathcal{L}_1^i$, $j \in \{0,1,2\}$ and $\Upsilon_F$ with $F \in \mathcal{F}^i$ form the nodal basis of $HCT$. In the next lemma, jumps of vector-valued maps are intended componentwise. Consequently, if $v \in H^2(\mathcal{M})$, then $[\![\nabla v]\!] \cdot n$ indicates the jump of the normal derivative of $v$ on the skeleton $\Sigma$.

**Lemma 2.2.5** (Simplified nodal averaging into HCT)**.** *Let $\sigma \in S_2^0$ be a piecewise quadratic polynomial and $K \in \mathcal{M}$. For all vertices $z \in \mathcal{L}_1(K)$ and edges $F \in \mathcal{F}_K$, we have*

$$\big|\nabla\sigma_{|K}(z) - \nabla A_{HCT}\sigma(z)\big| \leq C \sum_{F'\ni z} \frac{1}{|F'|^{\frac{1}{2}}} \| \, [\![\nabla\sigma]\!] \, \|_{L^2(F')} \tag{2.2.11a}$$

$$\begin{aligned} &\big|\nabla\sigma_{|K}(z) - \nabla A_{HCT}\sigma(z)\big| \leq \\ &\qquad \leq \sum_{F'\ni z} \frac{1}{|F'|} \left| \int_{F'} [\![\nabla\sigma]\!] \right| + C \sum_{K'\ni z} \frac{h_{K'}}{|K'|^{\frac{1}{2}}} \| \, \mathrm{D}^2\, \sigma \|_{L^2(K')} \end{aligned} \tag{2.2.11b}$$

$$\big|\nabla\sigma_{|K}(m_F)\cdot n - \nabla A_{HCT}\sigma(m_F)\cdot n\big| \leq \frac{1}{|F|} \left| \int_F [\![\nabla\sigma]\!] \cdot n \right| \tag{2.2.11c}$$

*where $F'$ and $K'$ vary in $\mathcal{F}$ and $\mathcal{M}$, respectively.*

*Proof.* We have $\nabla_{\mathcal{M}}\,\sigma \in (S_1^0)^2$ and $\nabla A_{HCT}\sigma(z) = A_1\,\nabla_{\mathcal{M}}\,\sigma(z)$ for all vertices $z \in \mathcal{L}_1(K)$. Hence (2.2.11a) and (2.2.11b) easily follow by applying Lemmas 2.2.1 and 2.2.2 to $\nabla_{\mathcal{M}}\,\sigma$.

Next, for all edges $F \in \mathcal{F}$, we have

$$\big|\nabla\sigma_{|K}(m_F)\cdot n - \nabla A_{HCT}\sigma(m_F)\cdot n\big| \leq | \, [\![\nabla\sigma(z)]\!] \, (m_F)\cdot n \, | \, .$$

Indeed, the left-hand side either vanishes or coincides with the right-hand side, because $m_F$ is shared by at most two triangles of $\mathcal{M}$. Since $\nabla_{\mathcal{M}}\,\sigma$ is piecewise affine on $\mathcal{M}$, this inequality directly provides (2.2.11c). □

*Remark* 2.2.6 (Alternative averaging operators into $H_0^2(\Omega)$)*.* Standard (not simplified) nodal averaging into $HCT$ is considered, for instance, in [17]. Nodal averaging operators into $H_0^2$-conforming could also be defined with the help of the Argyris element, the reduced Argyris element or the reduced $HCT$ element [29], see also Remark 3.3.17.

In view of the applications in the next chapters, it is useful to complement Lemma 2.2.5 with a counterpart of (2.1.6) for some of the HCT basis functions. To this end, observe that (2.1.6) is derived by means of affine equivalence, while HCT elements are not affine equivalent.

**Lemma 2.2.7** (Scalings of averaged HCT basis functions)**.** *For any element $K \in \mathcal{M}$, vertex $z \in \mathcal{L}_1(K)$, edges $F, F' \in \mathcal{F}_K$ and $j \in \{1,2\}$, we have*

$$\left|\int_{F'} \nabla \Upsilon_F \cdot n_{F'}\right| \leq C\left|F'\right| \quad \text{and} \quad \left|\int_{F'} \nabla \Upsilon_z^j \cdot n_{F'}\right| \leq C\left|F'\right| \tag{2.2.12a}$$

$$\|\operatorname{D}^2 \Upsilon_F\|_{L^2(K)} \leq C\gamma_K \frac{|K|^{\frac{1}{2}}}{\rho_K} \quad \text{and} \quad \|\operatorname{D}^2 \Upsilon_z^j\|_{L^2(K)} \leq C\gamma_K \frac{|K|^{\frac{1}{2}}}{\rho_K}. \tag{2.2.12b}$$

*Proof.* We can compute the integrals in (2.2.12a) by the Simpson's formula, because both $\nabla \Upsilon_F \cdot n_{F'}$ and $\nabla \Upsilon_z^j \cdot n_{F'}$ are in $\mathbb{P}_2(F')$. Owing to the duality

$$\Upsilon_F(y) = 0, \quad \nabla \Upsilon_F(y) = 0, \quad \nabla \Upsilon_F(m_{F'}) \cdot n_{F'} = \delta_{FF'} \tag{2.2.13}$$

for all $y \in \mathcal{L}_1(K)$ and $F' \in \mathcal{F}_K$, we derive

$$\int_{F'} \nabla \Upsilon_F \cdot n_{F'} = \frac{2}{3}\left|F'\right| \delta_{FF'}.$$

Similarly, the duality of $\Upsilon_z^j$

$$\Upsilon_z^j(y) = 0, \quad \nabla \Upsilon_z^j(y) = \delta_{yz} e_j, \quad \nabla \Upsilon_z^j(m_{F'}) \cdot n_{F'} = 0 \tag{2.2.14}$$

reveals $\nabla \Upsilon_z^j \cdot n_{F'} \equiv 0$ on $F'$ for $z \notin F'$ and

$$\int_{F'} \nabla \Upsilon_z^j \cdot n_{F'} = \frac{e_j \cdot n_{F'}}{6}\left|F'\right|$$

otherwise. Next, to check the validity of (2.2.12b), we argue as suggested by Ciarlet [31, Theorem 46.2]. More precisely, we employ an auxiliary finite element that is given by the following 12 functionals on $C^1(K) \cap \mathbb{P}_3(\mathcal{M}_K)$: $P(z)$ for $z \in \mathcal{L}_1(K)$ as well as $\nabla P(z) \cdot (y - z)$ for $y, z \in \mathcal{L}_1(K)$ with $y \neq z$ and $\nabla P(m_F) \cdot (m_K - m_F)$ for $F \in \mathcal{F}_K$. We denote the corresponding nodal basis on $K$ by $\widetilde{\Upsilon}_z$, $\widetilde{\Upsilon}_z^y$, $z, y \in \mathcal{L}_1(K)$ with $y \neq z$, and $\widetilde{\Upsilon}_F$, $F \in \mathcal{F}^i$. Since this element is affine equivalent, a comparison with a reference element yields, for every of its nodal basis function $\widetilde{\Upsilon}$ on $K$,

$$\|\operatorname{D}^2 \widetilde{\Upsilon}\|_{L^2(K)} \leq C\rho_K^{-2}\left|K\right|^{\frac{1}{2}}. \tag{2.2.15}$$

In view of (2.2.13)-(2.2.14), we obtain the following representations in terms of affine equivalent basis function:

$$\Upsilon_F = (m_K - m_F) \cdot n_F \widetilde{\Upsilon}_F$$

and

$$\Upsilon_z^j = \sum_{y\in\mathcal{L}_1(K)\setminus\{z\}} (y-z)\cdot e_j \widetilde{\Upsilon}_z^y - \frac{1}{4} \sum_{F\in\mathcal{F}_K : F\ni z} t_F \cdot e_j \, (m_K - m_F)\cdot t_F \widetilde{\Upsilon}_F.$$

Combining these identities with (2.2.15) completes the proof. □

# Chapter 3

# Overconsistency and Classical Nonconforming Elements

This chapter collects the material from [64] and is devoted to exemplify the abstract results from Chapter 1 by the construction of quasi-optimal methods with classical nonconforming elements. The Crouzeix-Raviart element [35] approximating the Poisson problem may be viewed as a prototypical example of such methods. Thus, we begin with an overview of our motivation and main results in this case.

## 3.1 Overview

Let $\mathcal{M}$ be a simplicial mesh of a domain $\Omega \subseteq \mathbb{R}^d$, $d \geq 2$, with faces $\mathcal{F}$. Furthermore, let $CR$ be the discrete space of real-valued functions on $\Omega$ that are piecewise affine, continuous in the midpoints of the internal faces of $\mathcal{M}$ and vanish at the midpoints of boundary faces. Since such functions can be discontinuous or nonzero in other points of the faces, $CR$ is not a subspace of the Sobolev space $H^1_0(\Omega)$. However, the Crouzeix-Raviart interpolant $\Pi_{CR} : H^1_0(\Omega) \to CR$, given by

$$\forall F \in \mathcal{F} \quad \int_F \Pi_{CR} u = \int_F u, \tag{3.1.1}$$

reveals remarkable approximation properties: for any function $u \in H^1_0(\Omega)$, we have

$$\begin{aligned} \inf_{s \in CR} \| \nabla_{\mathcal{M}}(u - s)\|_{L^2(\Omega)} &= \| \nabla_{\mathcal{M}}(u - \Pi_{CR} u)\|_{L^2(\Omega)} \\ &= \left( \sum_{K \in \mathcal{M}} \inf_{p \in \mathbb{P}_1(K)} \|\nabla(u - p)\|^2_{L^2(K)} \right)^{\frac{1}{2}}. \end{aligned} \tag{3.1.2}$$

We see that, although the global best error of the Crouzeix-Raviart space is coupled or constrained at the midpoints of the faces, it is locally computable and exploits optimally the approximation capabilities of its shape functions. The latter improves on the space of continuous piecewise affine functions, which exploits the shape functions only in a quasi-optimal manner, depending on the shape coefficient of $\mathcal{M}$; cf. Veeser [60].

The space $CR$ is used in the homonymous method for the Poisson problem,

$$U \in CR \quad \text{such that} \quad \forall \sigma \in CR \quad \int_\Omega \nabla_{\mathcal{M}} U \cdot \nabla_{\mathcal{M}} \sigma = \int_\Omega f\sigma, \tag{3.1.3}$$

where we suppose $f \in L^2(\Omega)$. This is a nonconforming Galerkin method in the sense of (1.2.11), because the underlying bilinear and linear forms on the conforming part $CR \cap H_0^1(\Omega)$ of the discrete space arise by simple restriction of their infinite-dimensional counterparts.

The question arises how much of the aforementioned remarkable approximation properties of the Crouzeix-Raviart space $CR$ are exploited in the method (3.1.3). Remark 1.4.9 reveals that the error $\| \nabla_{\mathcal{M}}(u-U)\|_{L^2(\Omega)}$ cannot be bounded only in terms of the best error $\inf_{s\in CR} \| \nabla_{\mathcal{M}}(u-s)\|_{L^2(\Omega)}$. The reason for this lies in the fact that (3.1.3) applies nonconforming functions to the load $f$. Thus, the classical Crouzeix-Raviart method is not quasi-optimal with respect to $\| \nabla_{\mathcal{M}} \cdot\|_{L^2(\Omega)}$ and so does not always fully exploit the approximation properties of its underlying space $CR$, see also Proposition 5.3.2.

In order to remedy, we may exploit Theorem 1.4.14 and consider the following two variants of the original Crouzeix-Raviart method:

$$U_E \in CR \quad \text{such that} \quad \forall \sigma \in CR \quad \int_\Omega \nabla_{\mathcal{M}} U_E \cdot \nabla_{\mathcal{M}} \sigma = \langle f, E\sigma\rangle, \tag{3.1.4a}$$

$$\bar{U}_E \in CR \quad \text{such that} \quad \forall \sigma \in CR \quad \int_\Omega \nabla_{\mathcal{M}} \bar{U}_E \cdot \nabla E\sigma = \langle f, E\sigma\rangle \tag{3.1.4b}$$

for a bounded linear smoothing operator $E : CR \to H_0^1(\Omega)$ to be specified. Both variants are well-defined for arbitrary $f \in H^{-1}(\Omega) = H_0^1(\Omega)'$ and each one has attractive features: the bilinear form of (3.1.4a) is symmetric, while the error of (3.1.4b) is orthogonal to the range of $E$. Analyzing an abstract version of (3.1.4b) with the tools from Chapter 1, we find that its quasi-optimality constant depends only on the range of $E$ and that, for a fixed range, the energy norm condition number of its bilinear form becomes minimal, if $E$ is a right inverse of the best approximation operator onto $CR$. Notably, the two variants also coincide under this condition.

Combining (3.1.1) and (3.1.2), we see that $E$ is a right-inverse of the best approximation operator onto $CR$ if and only if

$$\forall \sigma \in CR, F \in \mathcal{F} \quad \int_F E\sigma = \int_F \sigma. \tag{3.1.5}$$

Exploiting this local characterization, we construct a computationally feasible operator $E$ such that (3.1.4b), or equivalently (3.1.4a), is quasi-optimal. More precisely, we have

$$\| \nabla_{\mathcal{M}}(u - U_E)\| \leq \|E\|_{\mathcal{L}(S,V)} \inf_{s \in CR} \| \nabla_{\mathcal{M}}(u - s)\|_{L^2(\Omega)},$$

where $\|E\|_{\mathcal{L}(S,V)}$ is the best constant and equals the stability constant of resulting method. The construction of $E$, which is inspired by the one in Badia et al. [8], also ensures that $\|E\|_{\mathcal{L}(S,V)}$ can be bounded in terms of the shape coefficient of the mesh $\mathcal{M}$. It is also instrumental for designing quasi-optimal DG and other interior penalty methods in the next chapter.

The rest of this chapter is organized as follows. In §3.2 we analyze well-posedness, conditioning and quasi-optimality of the abstract counterpart of (3.1.4b). In §3.3 we then construct the aforementioned smoothing operator $E$, as well as similar operators when approximating the Poisson problem with Crouzeix-Raviart-like elements of arbitrary fixed order and the biharmonic problem with the Morley element.

In the discussion of the examples, we restrict ourselves to polyhedral Lipschitz domains and homogeneous essential boundary conditions. More general settings will be discussed elsewhere.

## 3.2 Overconsistency

In this chapter we shall look for quasi-optimal methods in a suitable subclass of nonconforming linear variational methods (1.2.3). First of all, taking into account Theorems 1.3.2 and 1.4.7 and the fact that we do not require the inclusion $S \subseteq V$, we let $E : S \to V$ be a linear smoothing operator and restrict our attention to methods $M : V' \to S$ given by the discrete problem

$$\forall \sigma \in S \quad b(M\ell, \sigma) = \langle \ell, E\sigma \rangle, \tag{3.2.1}$$

corresponding to the triplet $(S, b, E^\star)$. Then, for further convenience, let us collect in the next theorem the other relevant results from §1 that will be instrumental to our discussion.

**Theorem 3.2.1** (Stability, consistency, and quasi-optimality). *Any nonconforming method $M = (S, b, E^\star)$ for (1.2.1) satisfies:*

*(i) $M$ is fully stable, with*

$$C_{\text{stab}} := \|M\|_{\mathcal{L}(V',S)} = \sup_{\sigma \in S} \frac{\|E\sigma\|}{\sup_{s \in S, \|s\|=1} b(s, \sigma)}.$$

*(ii) $M$ is quasi-optimal if and only if it is fully algebraically consistent*

$$\forall u \in S \cap V, \sigma \in S \quad b(u, \sigma) = a(u, E\sigma).$$

*(iii) If $M$ is quasi-optimal, then its quasi-optimality constant is*

$$C_{\mathrm{qopt}} = \sup_{\sigma \in S} \frac{\sup_{v \in V, s \in S, \|v+s\|=1} a(v, E\sigma) + b(s,\sigma)}{\sup_{s \in S, \|s\|=1} b(s,\sigma)}$$

*and satisfies*

$$\max\{C_{\mathrm{stab}}, \delta_S\} \leq C_{\mathrm{qopt}} \leq \sqrt{C_{\mathrm{stab}}^2 + \delta_S^2},$$

*where $\delta_S \in [0,\infty)$ is the consistency measure given by the smallest constant in*

$$\forall s, \sigma \in S \quad |b(s,\sigma) - \widetilde{a}(s, E\sigma)| \leq \delta_S \sup_{\hat{s} \in S, \|\hat{s}\|=1} b(\hat{s}, \sigma) \inf_{v \in V} \|s - v\|.$$

*Proof.* This is a partial restatement of Theorems 1.4.7, 1.4.14 and 1.4.19. □

We say that a method $M = (S, b, E^\star)$ is *(algebraically) overconsistent* whenever its consistency measure $\delta_S$ vanishes. Hereby, we restrict our attention to this subclass of quasi-optimal methods. According to item (iii) of Theorem 3.2.1, the condition $\delta_S = 0$ immediately yields the identity $C_{\mathrm{qopt}} = C_{\mathrm{stab}}$. Also, notice that conforming Galerkin methods (1.2.4) are always overconsistent.

While overconsistency somehow aims at minimizing the effect of (nonconforming) consistency on the size of the quasi-optimality constant, it prescribes also a certain rigidity in the structure of a method $M$. To see this, assume that we are given $V$ and $a$ of the continuous problem (1.2.1) and a discrete space $S$, along with an extended scalar product $\widetilde{a}$. Then, the design of an overconsistent quasi-optimal method on $S$ reduces to the task of finding a smoothing operator $E$ and a bilinear form $b$ such that

$$\forall s, \sigma \in S \qquad b(s,\sigma) = \widetilde{a}(s, E\sigma). \tag{3.2.2}$$

It is worth mentioning that, in contrast to full algebraic consistency, overconsistency hinges on the specific extension $\widetilde{a}$ of $a$ at hand.

Three possibilities to define the form $b$ in terms of $\widetilde{a}$ and a smoother $E$ are the following:

$$\widetilde{a}(\cdot,\cdot), \qquad \widetilde{a}(\cdot, E\cdot), \qquad \text{and} \qquad \widetilde{a}(E\cdot, E\cdot).$$

Since the third option corresponds to a conforming Galerkin method on the range $T = R(E)$ of $E$ also when $S \not\subseteq V$, it is covered by standard theory. We therefore do not consider it here. The first two, truly nonconforming options separate the advantages of a conforming Galerkin method for (1.2.1): the first one is a symmetric bilinear form, while the second one corresponds to overconsistency in view of (3.2.2). Interestingly, the two options coincide and unify their advantages if and only if the smoothing operator $E$ is a right inverse for the $\widetilde{a}$-orthogonal projection $\Pi$ from $\widetilde{V}$ onto $S$ because of the identity $\widetilde{a}(s, E\sigma) = \widetilde{a}(s, \Pi E\sigma)$ for all $s, \sigma \in S$.

Here we investigate the second option, which shall partially bring us back to the first one, and set

$$b_E(s, \sigma) := \widetilde{a}(s, E\sigma), \qquad s, \sigma \in S, \tag{3.2.3}$$

Writing $M_E$ as an abbreviation for $(S, b_E, E^\star)$, the resulting discrete problem reads as follows: given any $\ell \in V'$, find $M_E\ell \in S$ such that

$$\forall \sigma \in S \quad \widetilde{a}(M_E\ell, E\sigma) = \langle \ell, E\sigma \rangle. \tag{3.2.4}$$

Since the test function $\sigma$ enters only via $E\sigma$, such a method can be viewed as a Petrov-Galerkin method over $S \times T$ with the conforming test space $T := R(E)$. In other words, (3.2.4) is equivalent to

$$\forall \tau \in T \quad \widetilde{a}(M_E\ell, \tau) = \langle \ell, \tau \rangle.$$

Consequently, properties of the map $M_E$ depend on $E$ only through its range $T = R(E)$. In what follows, we underline this aspect whenever applicable. Let us start by examining the solvability and related properties of (3.2.4).

*Remark* 3.2.2 (Injectivity of smoothing). In view of $M_E = B^{-1}E^\star$, the injectivity of the smoothing operator $E$ is equivalent to the surjectivity of $M_E$. In connection with a bilinear form $b_E$, it becomes also a necessary condition for the well-posedness of (3.2.4).

**Lemma 3.2.3** (Nondegeneracy of $b_E$)**.** *For any injective linear operator $E : S \to V$ with range $T = R(E)$, the following statements are equivalent:*

$$b_E \text{ is nondegenerate on } S \times S, \tag{3.2.5a}$$

$$\widetilde{a}(\cdot, \cdot) \text{ is nondegenerate on } S \times T, \tag{3.2.5b}$$

$$\Pi_{|T} \text{ is invertible}, \tag{3.2.5c}$$

$$S \cap T^\perp = \{0\}, \tag{3.2.5d}$$

*where $\Pi$ stands for the $\widetilde{a}$-orthogonal projection from $\widetilde{V}$ onto $S$. If $b_E$ is nondegenerate, then its energy norm condition number is given by*

$$\operatorname{cond}(b_E) = \|(\Pi E)^{-1}\|_{\mathcal{L}(T)} \|\Pi E\|_{\mathcal{L}(S)} \geq 1, \tag{3.2.6}$$

*which is minimized by $E = (\Pi_{|T})^{-1}$.*

*Proof.* The claimed equivalences are essentially a special case of the inf-sup theory; we provide the details of their proofs for the sake of completeness.

We first observe that $E$ is a linear isomorphism from $S$ to $T$, which implies $\dim S = \dim T$ as well as (3.2.5a) $\iff$ (3.2.5b).

Next, we verify (3.2.5b) $\implies$ (3.2.5c) and let $\tau \in T$ with $\Pi\tau = 0$. This yields $0 = \widetilde{a}(s, \Pi\tau) = \widetilde{a}(s, \tau)$ for all $s \in S$ and so, using (3.2.5b), we see that $\tau = 0$. Consequently, the kernel of $\Pi_{|T}$ is trivial and the rank-nullity theorem yields that $\Pi_{|T}$ is a linear isomorphism from $T$ to $S$.

To show (3.2.5c) $\implies$ (3.2.5d), consider any $s \in S \cap T^\perp$. Then $\tau := (\Pi_{|T})^{-1}s \in T$ thanks to (3.2.5c) and $0 = \widetilde{a}(s, \tau) = \widetilde{a}(s, (\Pi_{|T})^{-1}s) = \widetilde{a}(s, \Pi(\Pi_{|T})^{-1}s) = \widetilde{a}(s, s)$ gives $s = 0$. Hence we have $S \cap T^\perp = \{0\}$.

We complete the proof of the claimed equivalences by showing that (3.2.5d) $\implies$ (3.2.5b). Since $\dim S = \dim T$, it suffices to check the nondegeneracy for the first argument of $\widetilde{a}$, that is, given $s \in S$, $\widetilde{a}(s, \tau) = 0$ for all $\tau \in T$ implies $s = 0$. This condition is just a reformulation of (3.2.5d), so that the desired implication is verified.

Finally, assuming that $b_E$ is nondegenerate, we turn to (3.2.6) and recall that the energy norm condition number of $b_E$ is given by $\operatorname{cond}(b_E) = C_E/\beta_E$, where

$$C_E := \sup_{s,\sigma \in S} \frac{b_E(s,\sigma)}{\|s\|\|\sigma\|} \geq \inf_{s\in S} \sup_{\sigma \in S} \frac{b_E(s,\sigma)}{\|s\|\|\sigma\|} = \inf_{\sigma\in S} \sup_{s \in S} \frac{b_E(s,\sigma)}{\|s\|\|\sigma\|} =: \beta_E > 0.$$

We claim that, for any $\sigma \in S$,

$$\sup_{s\in S} \frac{b_E(s,\sigma)}{\|s\|} = \|\Pi E\sigma\|. \tag{3.2.7}$$

Indeed, if $s \in S$, the properties of $\Pi$ and the Cauchy-Schwarz inequality yield $b_E(s,\sigma) = \widetilde{a}(s, E\sigma) = \widetilde{a}(s, \Pi E\sigma) \leq \|s\|\|\Pi E\sigma\|$, with equality for $s = \Pi E\sigma$. Exploiting (3.2.7) in the definition of $C_E$ and the second expression for $\beta_E$, we conclude

$$\operatorname{cond}(b_E) = \frac{\sup_{\sigma\in S, \|\sigma\|=1} \|\Pi E\sigma\|}{\inf_{\sigma\in S, \|\sigma\|=1} \|\Pi E\sigma\|} = \|(\Pi E)^{-1}\|_{\mathcal{L}(S)} \|\Pi E\|_{\mathcal{L}(S)}. \qquad \square$$

Next, ignoring computational feasibility, we characterize the existence of at least one smoothing operator $E$ giving rise to a nondegenerate bilinear form $b_E$. Condition (3.2.8c) below reveals that the search for right inverses is not restrictive. Moreover, we shall use this characterization in Chapter 4 to observe that we cannot obtain overconsistency in certain settings.

**Lemma 3.2.4** (Existence of nondegenerate $b_E$)**.** *For any discrete space $S$ and extended scalar product $\widetilde{a}$, the following statements are equivalent:*

(3.2.8a) *there is an injective $E : S \to V$ such that $b_E$ is nondegenerate,*

(3.2.8b) $S \cap V^\perp = \{0\}$,

(3.2.8c) $\Pi_{|V}$ *admits a right inverse.*

*Proof.* First, we verify (3.2.8a) $\Longrightarrow$ (3.2.8b). Assume $E : S \to V$ is injective and such that $b_E$ is nondegenerate. Using the previous lemma, we infer $S \cap T^\perp = \{0\}$ for $T = R(E)$. Since $T \subseteq V$, we have $V^\perp \subseteq T^\perp$ and $S \cap V^\perp \subseteq S \cap T^\perp = \{0\}$, whence $S \cap V^\perp = \{0\}$.

To show the implication (3.2.8b) $\Longrightarrow$ (3.2.8c), we assume $S \cap V^\perp = \{0\}$ and observe $s \in S \cap V^\perp \iff s \in S \cap \Pi(V)^\perp$ with the help of the identity $\widetilde{a}(v, s) = \widetilde{a}(\Pi v, s)$ for all $v \in V$ and $s \in S$. We thus infer $\Pi(V) = S$ and can apply [24, Theorem 2.12] to obtain: $\Pi_{|V}$ admits a right inverse if and only if $N(\Pi_{|V})$ admits a complement in $V$. Since $\Pi$ is $\widetilde{a}$-orthogonal, we have $N(\Pi_{|V}) = S^\perp \cap V$, which has the complement $S \cap V$ in $V$. Hence (3.2.8c) holds.

The missing implication (3.2.8c) $\Longrightarrow$ (3.2.8a) is straight-forward. Let $E : S \to V$ be a right inverse of $\Pi_{|V}$ and observe that $E$ and $\Pi_{|R(E)}$ have to be injective. Thus, Lemma 3.2.3 provides (3.2.8a). □

Let us now turn to stability and quasi-optimality of overconsistent methods.

**Theorem 3.2.5** (Overconsistent quasi-optimality). *Let $E : S \to V$ be any injective smoothing operator with range $T = R(E)$. If $S \cap T^\perp = \{0\}$, then the method $M_E = (S, b_E, E)$ is quasi-optimal with*

$$C_{\text{qopt}} = \|(\Pi_{|T})^{-1}\|_{\mathcal{L}(S,V)} = C_{\text{stab}}.$$

*Proof.* Since $S \cap T^\perp = \{0\}$, Lemma 3.2.3 ensures that $b_E$ is nondegenerate. Furthermore, $M_E$ is fully stable and overconsistent by construction and so Theorem 3.2.1 shows that $M_E$ is quasi-optimal with $C_{\text{qopt}} = C_{\text{stab}}$. We conclude by deriving

$$\begin{aligned} C_{\text{stab}} &= \sup_{\sigma \in S} \frac{\|E\sigma\|}{\|\Pi E\sigma\|} = \sup_{\tau \in T} \frac{\|\tau\|}{\|\Pi\tau\|} = \\ &= \sup_{\sigma \in S} \frac{\|(\Pi_{|T})^{-1}\sigma\|}{\|\sigma\|} = \|(\Pi_{|T})^{-1}\|_{\mathcal{L}(S,V)} \end{aligned} \tag{3.2.9}$$

which follows by inserting (3.2.7) into Theorem 3.2.1 (i) and exploiting that $E : S \to T$ and $\Pi_{|T}$ are bijective. □

*Remark* 3.2.6 (Overconsistency and increasing nonconformity). For overconsistent methods, the constants $C_{\text{qopt}} = C_{\text{stab}}$ grow with increasing nonconformity. To see this, let $\sigma \in S \setminus V$ with $\|\sigma\| = 1$ be a nonconforming direction and recall that $V$ is a closed subspace of $\widetilde{V}$. The angle of $\sigma$ and $V$ is then $\alpha \in [0, \pi/2)$, given by $\cos\alpha = \sup_{v \in V, \|v\|=1} |\widetilde{a}(v, \sigma)| > 0$. Since $T = R(E) \subseteq V$, the angle between $\sigma \in S$ and $(\Pi_{|T})^{-1}\sigma$ is bigger than $\alpha$. Hence $\widetilde{a}(\sigma, (\Pi_{|T})^{-1}\sigma) = \|\sigma\|^2 = 1$ yields $C_{\text{qopt}} \geq \|(\Pi_{|T})^{-1}\sigma\| \geq (\cos\alpha)^{-1}$.

*Remark* 3.2.7 (Possible overestimation by classical upper bounds). The first identity in (3.2.9) and $\|E\|_{\mathcal{L}(S,V)} = \sup_{\|\sigma\|=1} \sup_{\|\widetilde{v}\|=1} \widetilde{a}(\widetilde{v}, E\sigma) =: \widetilde{C}_E$ yield

$$C_{\text{qopt}} \leq \|(\Pi E)^{-1}\|_{\mathcal{L}(S)} \|E\|_{\mathcal{L}(S,V)} = \frac{\widetilde{C}_E}{\beta_E}, losed$$

where the right-hand side admits the classical form of an upper bound for the quasi-optimality constant. Notably, this bound depends on $E$ not only through its range $T = R(E)$ and, closely related, may be pessimistic if $E$ has singular values of different size.

Neglecting the issue of computational feasibility, our analysis of overconsistent methods does not reveal any disadvantage of restricting the search of smoothing operators to right inverses for the $\widetilde{a}$-orthogonal projection $\Pi$. On the contrary, the bilinear form is given by simple restriction of $\widetilde{a}$, thus symmetric, and minimizes its energy norm condition number within smoothing operators of the same range. We therefore aim at invoking the following special case of Theorem 3.2.5.

**Corollary 3.2.8** (Smoothing with right inverses)*. Let $E : S \to V$ be a right inverse for the $\widetilde{a}$-orthogonal projection $\Pi$ from $\widetilde{V}$ onto $S$. Then, we have $M_E = (S, \widetilde{a}_{|S\times S}, E^\star)$ and this is a nonconforming Galerkin method if and only if $E_{|S\cap V} = \mathrm{Id}_{S\cap V}$. Moreover, $M_E$ is quasi-optimal with*

$$C_{\text{qopt}} = C_{\text{stab}} = \|E\|_{\mathcal{L}(S,V)}.$$

## 3.3 Applications to Classical Nonconforming Methods

In light of Corollary 3.2.8, the key step for quasi-optimality is to find a right inverse $E$ for the projection $\Pi$ that provides $V$-smoothing, is suitably bounded and *computationally feasible* in the sense of Remark 1.4.13. In the context of finite element methods, the latter is given if, for the finite element basis $\varphi_1, \dots, \varphi_n$ at hand, the evaluations $\langle \ell, E\varphi_i \rangle$, $i = 1, \dots, n$, can be implemented with $O(n)$ operations. In this section, we construct such right inverses not only for the setting considered in the introduction §3.1, but also for nonconforming elements of arbitrary fixed order and for fourth order problems.

### 3.3.1 A Quasi-Optimal Crouzeix-Raviart Method for the Poisson Problem

In order to prove the results illustrated in the overview §3.1, we consider the approximation with Crouzeix-Raviart elements of the Poisson problem

$$-\Delta u = f \text{ in } \Omega, \qquad u = 0 \text{ on } \partial\Omega, \tag{3.3.1}$$

where $\Omega$ and $\mathcal{M}$ are as in §2.1, with $d \geq 2$ and $\#\mathcal{M} > 1$. Introducing the bilinear form $a_{\mathcal{M}} : H^1(\mathcal{M}) \times H^1(\mathcal{M}) \to \mathbb{R}$ by

$$a_{\mathcal{M}}(w_1, w_2) := \int_\Omega \nabla_{\mathcal{M}} w_1 \cdot \nabla_{\mathcal{M}} w_2, \tag{3.3.2}$$

we want to apply Corollary 3.2.8 with the following setting:

$$\begin{gathered} V = H^1_0(\Omega), \quad S = CR = \left\{ s \in S^0_1 \mid \forall F \in \mathcal{F} \int_F [\![s]\!] = 0 \right\}, \\ \widetilde{a} = a_{\mathcal{M}|\widetilde{V}\times\widetilde{V}} \text{ with } \widetilde{V} = H^1_0(\Omega) + CR, \end{gathered} \tag{3.3.3}$$

where $\widetilde{a}_{|V\times V}$ provides a weak formulation of $-\Delta$. Before embarking on the construction of the smoothing operator $E$, let us recall some relevant properties of $CR$; see, e.g., [21]. The characterization of $CR$ in terms of jumps is a consequence of the midpoint rule: whenever $s \in CR$ and $F \in \mathcal{F}_K$, then $\fint_F s_{|K} = s(m_F)$, where $m_F$ is the midpoint of $F$. Hence, for all $s \in CR$, the integral mean value $\fint_F s$, $F \in \mathcal{F}$, is well-defined and vanishes if $F \in \mathcal{F}^b$. The bilinear form $a_{\mathcal{M}}$ is therefore a scalar product and induces the norm $\|\cdot\| = \|\nabla_{\mathcal{M}} \cdot\|_{L^2(\Omega)}$ on $CR$. Moreover, the functionals $s \mapsto \fint_F s$, $F \in \mathcal{F}^i$, form a set of degrees of freedom for $CR$. We write $\Psi_F$, $F \in \mathcal{F}^i$, for the associated nodal basis satisfying $\fint_{F'} \Psi_F = \delta_{F,F'}$ for all $F, F' \in \mathcal{F}^i$. The support of each basis function $\Psi_F$ is the union $\omega_F$ of the two elements sharing $F$. Finally, we have $CR \cap H^1_0(\Omega) = S^1_1$, which is a strict subspace of $CR$ as $\#\mathcal{M} > 1$.

The next lemma characterizes the right inverses of the Crouzeix-Raviart projection $\Pi_{CR}$, i.e. the $a_{\mathcal{M}}$-orthogonal projection of $\widetilde{V}$ onto $CR$. It also motivates our use of the same notation as for the Crouzeix-Raviart interpolant in (3.1.1).

**Lemma 3.3.1** (Right inverses of CR projection). *Let $E : CR \to H^1_0(\Omega)$ be a linear operator. Then we have*

$$\Pi_{CR} E = \mathrm{Id}_{CR} \quad \iff \quad \forall \sigma \in CR, F \in \mathcal{F}^i \int_F E\sigma = \int_F \sigma.$$

*Proof.* For any $v \in H^1_0(\Omega)$ and $s \in CR$, the $a_{\mathcal{M}}$-orthogonality of $\Pi_{CR}$ and piecewise integration by parts yields

$$\begin{aligned} 0 = a_{\mathcal{M}}(s, v - \Pi_{CR} v) &= \sum_{K\in\mathcal{M}} \int_{\partial K} \frac{\partial s}{\partial n_K}(v - \Pi_{CR} v) \\ &= \sum_{F\in\mathcal{F}^i} [\![\nabla s]\!] \cdot n \int_F (v - \Pi_{CR} v) \end{aligned}$$

thanks to the fact that $\nabla_{\mathcal{M}} s$ is piecewise constant and $\int_F v = 0 = \int_F \Pi_{CR} v$ for every $F \in \mathcal{F}^b$. Since the orthogonal projection $\Pi_{CR} v$ is unique and the averages over interior faces are degrees of freedom for $CR$, we obtain that

$$\forall F \in \mathcal{F}^i \qquad \int_F \Pi_{CR} v = \int_F v \tag{3.3.4}$$

uniquely determines $\Pi_{CR} v$. This characterization readily implies the claimed equivalence. □

Thus, we turn to the construction of a linear right inverse of $\Pi_{CR}$. The normalized face bubbles

$$\bar{\Phi}_F := \frac{(2d)!}{d!\,|F|}\Phi_F \quad \text{with} \quad \Phi_F := \prod_{z \in \mathcal{L}_1(F)} \Phi_z^1 = \frac{1}{d^d}\Phi_{m_F}^d, \qquad F \in \mathcal{F}^i, \tag{3.3.5}$$

may be viewed as $H_0^1(\Omega)$-counterparts of the basis functions $\Psi_F$, $F \in \mathcal{F}^i$. Indeed, they satisfy $\bar{\Phi}_F \in H_0^1(\Omega)$ and $\int_{F'} \bar{\Phi}_F = \delta_{F,F'}$ for all $F' \in \mathcal{F}^i$ due to (2.1.1). We thus readily see that the bubble smoother $B_1 : CR \to H_0^1(\Omega)$ given by

$$B_1 \sigma := \sum_{F \in \mathcal{F}^i} \left( \int_F \sigma \right) \bar{\Phi}_F \tag{3.3.6}$$

is well-defined and a right inverse of the Crouzeix-Raviart projection $\Pi_{CR}$. Unfortunately, the operator $B_1$ is not uniformly stable under refinement; see Remark 3.3.4 below. We therefore introduce the following variant that is stabilized with simplified nodal averaging.

**Proposition 3.3.2** (Stable right inverse of CR projection)**.** *The linear operator $E_1 : CR \to H_0^1(\Omega)$ given by*

$$E_1 \sigma := A_1 \sigma + B_1(\sigma - A_1 \sigma), \tag{3.3.7}$$

*is invariant on $S_1^1$, a right inverse of the Crouzeix-Raviart projection $\Pi_{CR}$, and $H_0^1(\Omega)$-stable with stability constant $\leq C_{d,\gamma_{\mathcal{M}}}$.*

*Proof.* The linear operator $E_1$ is well-defined owing to $R(A_1) = S_1^1 \subseteq CR$ and provides $H_0^1(\Omega)$-smoothing, because $\Phi_z^1 \in H_0^1(\Omega)$ for $z \in \mathcal{L}_1^i$ and we have $\Phi_F \in H_0^1(\Omega)$ for $F \in \mathcal{F}^i$. Owing to $A_{1|S_1^1} = \mathrm{Id}_{S_1^1}$, it holds $E_{1|S_1^1} = \mathrm{Id}_{S_1^1}$ on the conforming part $S_1^1 = CR \cap H_0^1(\Omega)$ of the Crouzeix-Raviart space. Furthermore, $E_1$ is a right inverse of the Crouzeix-Raviart projection in view of Lemma 3.3.1. Indeed, by rearranging terms and since $B_1$ preserves face means, we find

$$\int_F E_1 \sigma = \int_F B_1 \sigma + \underbrace{\int_F (A_1 \sigma - B_1 A_1 \sigma)}_{=0} = \int_F \sigma. \tag{3.3.8}$$

It remains to bound $\|E_1\|_{\mathcal{L}(CR,H^1_0(\Omega))}$. Given $\sigma \in CR$, we may write

$$\|\nabla E_1\sigma\|_{L^2(\Omega)} \leq \| \nabla_{\mathcal{M}} \sigma\|_{L^2(\Omega)} + \| \nabla_{\mathcal{M}}(\sigma - A_1\sigma)\|_{L^2(\Omega)} + \|\nabla B_1(\sigma - A_1\sigma)\|_{L^2(\Omega)}$$

so that we have to bound the second and third term of the right-hand side by the first one. In both cases, we first establish a local bound for $K \in \mathcal{M}$. For the second term, we combine (2.1.5) and (2.1.6) with Lemma 2.2.2, deriving

$$\begin{aligned}\|\nabla(\sigma - A_1\sigma)\|_{L^2(K)} &\leq \sum_{z\in\mathcal{L}_1(K)} \left|\sigma_{|K}(z) - A_1\sigma(z)\right| \|\nabla\Phi^1_z\|_{L^2(K)} \\ &\leq C_d \sum_{z\in\mathcal{L}_1(K)} \sum_{K'\in\mathcal{M},K'\ni z} \frac{h_{K'}}{\rho_K} \frac{|K|^{\frac{1}{2}}}{|K'|^{\frac{1}{2}}} \|\nabla\sigma\|_{L^2(K')} \lesssim \| \nabla_{\mathcal{M}} \sigma\|_{L^2(\omega_K)}.\end{aligned} \tag{3.3.9}$$

For the third term, inserting $\int_F \Phi^1_z = d^{-1}|F|$ and (3.3.5) into (3.3.6) yields

$$B_1(\sigma - A_1\sigma)_{|K} = \frac{(2d)!}{d!\, d^{d+1}} \sum_{F\in\mathcal{F}_K} \sum_{z\in\mathcal{L}_1(F)} \left[\sigma_{|K}(z) - A_1\sigma(z)\right] \Phi^d_{m_F}.$$

Hence, another combination of (2.1.5) and (2.1.6) with Lemma 2.2.2 leads to

$$\|\nabla B_1(\sigma - A_1\sigma)\|_{L^2(K)} \lesssim \| \nabla_{\mathcal{M}} \sigma\|_{L^2(\omega_K)}. \tag{3.3.10}$$

We conclude by summing (3.3.9) and (3.3.10) over all elements $K \in \mathcal{M}$, observing that the number of elements in each star $\omega_K$ is $\leq C_{d,\gamma_{\mathcal{M}}}$. □

The technique used here to stabilize the bubble smoother $B_1$ is not new. For instance, it is applied to the construction of Fortin operators for the Stokes problem in [11, Section 8.4.1].

Setting $E = E_1$ in (3.1.4a), we obtain a *new Crouzeix-Raviart method*, $M_{CR} = (CR, a_{\mathcal{M}}, E^{\star}_{CR})$. Notice that the assembling of its load vector is computationally feasible in the following sense:

- it suffices to know the evaluations $\langle f, \Phi^1_z\rangle$, $z \in \mathcal{L}^i_1$, and $\langle f, \Phi_F\rangle$, $F \in \mathcal{F}^i$,
- it is local in that $\operatorname{supp} E_1\Psi_F \subseteq \omega_{K_1} \cup \omega_{K_2}$, where $K_1, K_2 \in \mathcal{M}$ are the two elements containing the interior face $F \in \mathcal{F}^i$.

The method $M_{CR}$ distinguishes from the classical Crouzeix-Raviart method by the following property.

**Theorem 3.3.3** (Quasi-optimality of $M_{CR}$)**.** *The method $M_{CR}$ is a nonconforming Galerkin method for* (3.3.1) *and it is $\| \nabla_{\mathcal{M}} \cdot\|$-quasi-optimal with $C_{\mathrm{qopt}} \leq C_{d,\gamma_{\mathcal{M}}}$.*

*Proof.* Notice that $M_{CR} = (CR, b, E^{\star}_1)$, where $b$ is the restriction of $a_{\mathcal{M}}$ in (3.3.3) to $CR \times CR$. Thus, the claim follows by using Proposition 3.3.2 in Corollary 3.2.8. □

The following two remarks clarify that the single ingredients for $E_1$ are not suitable smoothing operators for quasi-optimality, thereby underlining their complementary roles.

*Remark* 3.3.4 (Instability of bubble smoothing). The right inverse $B_1$ is not uniformly $H^1_0(\Omega)$-stable under refinement. To see this, let $\mathcal{M}$ be a mesh of $\Omega = (0,1)^2$ the elements of which have diameter $h > 0$ and consider the function $\sigma := \sum_{F\in\mathcal{F}^i} \Psi_F$. Then $\sigma = 1$ in all elements except those touching $\partial\Omega$, while $B_1\sigma$ oscillates between 0 and 1 in all elements. Accordingly, $\bar{\Phi}_F = d^{-d}\Phi^d_{m_F}$, (2.1.6), and $h^{-1} \gtrsim |\nabla\Psi_F|$ give

$$\|\nabla B_1\sigma\|_{L^2(\Omega)} \gtrsim \#\mathcal{M} \gtrsim h^{-1}\#\{K \in \mathcal{M} \mid K \cap \partial\Omega \neq \emptyset\} \gtrsim h^{-1}\| \nabla_{\mathcal{M}}\, \sigma\|_{L^2(\Omega)}.$$

*Remark* 3.3.5 (Inconsistency of (simplified) nodal averaging). The use of smoothing operator $A_{1|CR}$ in (3.1.4a) does not lead to full algebraic consistency and so in particular not to quasi-optimality. As $\dim CR > \dim S^1_1$, the kernel $N(A_{1|CR})$ is non-trivial. Moreover, since $A_{1|CR}$ is not $a_{\mathcal{M}}$-orthogonal, $N(A_{1|CR})$ and $S^1_1$ are not $\widetilde{a}$-orthogonal. Consequently, we can find $\sigma \in CR$ which is $\widetilde{a}$-orthogonal to $S^1_1$ and such that $s := A_1\sigma \neq 0$. Then, we have $s \in S^1_1 = CR \cap H^1_0(\Omega)$ and $b(s,\sigma) = 0 \neq a(s, A_1\sigma)$, which contradicts full algebraic consistency. We give also numerical evidence of this observation in §5.3.3

We end the discussion on the method $M_{CR}$ with a comment on the use of the operator $A_1$.

*Remark* 3.3.6 (Standard and simplified nodal averaging). The simplified nodal averaging operator $A_1$ contributes to the proof of Proposition 3.3.2 via the estimate in Lemma 2.2.2. As already mentioned in Remark 2.2.4, such lemma still holds if $A_1$ is replaced with the averaging operator $\widetilde{A_1}$ from (2.2.8). Thus, while the evaluation of $A_1\sigma$, $\sigma \in CR$, is generally less expensive than that one of $\widetilde{A_1}\sigma$, our analysis does not reveal any disadvantage of employing $A_1$ instead of $\widetilde{A_1}$ in the definition of the smoother $E_1$. Furthermore, we propose a numerical comparison of the two options in §5.3.1.

### 3.3.2 Quasi-Optimal Crouzeix-Raviart Like Methods of Arbitrary Order for the Poisson Problem

In this section we generalize the quasi-optimal Crouzeix-Raviart method $M_{CR}$ of §3.3.1 to arbitrary fixed order $p \geq 2$. To this end, let $\Omega$ and $\mathcal{M}$ be as in §2.1, with $d \geq 2$ and $\#\mathcal{M} > 1$. This time, we want to apply Corollary 3.2.8 with

$$V = H^1_0(\Omega),$$

$$S^1_p \subseteq S \subseteq CR_p := \left\{ s \in S^0_p \mid \forall F \in \mathcal{F},\, q \in \mathbb{P}_{p-1}(F)\ \int_F [\![s]\!]\, q = 0 \right\}, \tag{3.3.11}$$

$$\widetilde{a} = a_{\mathcal{M}|\widetilde{V}\times\widetilde{V}} \text{ with } \widetilde{V} = V + S$$

and $a_{\mathcal{M}}$ as in (3.3.2). For any $d \geq 2$, the space $CR_1$ coincides with the Crouzeix-Raviart space $CR$ from §3.3.1. If $d = 2$, then $CR_p$ is the Fortin-Soulie space [42] for $p = 2$, the Crouzeix-Falk space [34] for $p = 3$, and, in general, the Gauss-Legendre space of Baran and Stoyan [57] of order $p$. The last reference provides a finite element basis of the Gauss-Legendre spaces, distinguishing odd and even polynomial degree $p$. For $d = 3$, Fortin [41] for $p = 2$ and Ciarlet et al. [32] in general construct finite element bases for nonconforming subspaces of $CR_p$, strict in certain situations. In order to cover also these Crouzeix-Raviart-like spaces, we require in (3.3.11) only $S \subseteq CR_p$.

Independently of the choice of $S$, we have that, for every $s \in S$, the moment $\int_F sq$ is well-defined for all $F \in \mathcal{F}$ and all $q \in \mathbb{P}_{p-1}(F)$ and vanishes whenever $F \in \mathcal{F}^b$. As a consequence, $\|\cdot\| = \|\nabla_{\mathcal{M}} \cdot\|_{L^2(\Omega)}$, which is induced by $a_{\mathcal{M}}$, is a norm on $\widetilde{V}$.

Let $\Pi_S$ denote the $a_{\mathcal{M}}$-orthogonal projection of $\widetilde{V}$ onto $S \subseteq CR_p$. Some right inverses thereof can be construct as follows.

**Lemma 3.3.7** (Right inverses of CR-like projections)**.** *Let $S \subseteq CR_p$ with $p \geq 2$ and $E : S \to H^1_0(\Omega)$ be a linear operator. If we have*

$$\int_F (E\sigma)q = \int_F \sigma q, \qquad \int_K (E\sigma)r = \int_K \sigma r \tag{3.3.12}$$

*for all $\sigma \in S$, $F \in \mathcal{F}^i, q \in \mathbb{P}_{p-1}(F)$ and $K \in \mathcal{M}, r \in \mathbb{P}_{p-2}(K)$, then $\Pi_S E = \mathrm{Id}_S$.*

*Proof.* Given $s, \sigma \in S \subseteq CR_p$, we integrate piecewise by parts and obtain

$$\begin{aligned} a_{\mathcal{M}}(s, \sigma - E\sigma) &= \sum_{K\in\mathcal{M}} \left( \int_{\partial K} \frac{\partial s}{\partial n_K}(\sigma - E\sigma) - \int_K \triangle s(\sigma - E\sigma) \right) \\ &= \sum_{F\in\mathcal{F}^i} \int_F [\![\nabla_{\mathcal{M}}\, s \cdot n]\!]\,(\sigma - E\sigma) - \sum_{K\in\mathcal{M}} \int_K \triangle s(\sigma - E\sigma) = 0 \end{aligned}$$

thanks to the hypotheses on $E$. Hence, $0 = \Pi_S(\sigma - E\sigma) = \sigma - \Pi_S E\sigma$. □

Let us construct such a smoothing operator by following the lines of the construction of $E_1$ in §3.3.1. In order to define a higher order bubble smoother, we employ local weighted $L^2$-projections associated to faces and elements. For every interior face $F \in \mathcal{F}^i$, let $Q_F : L^2(F) \to \mathbb{P}_{p-1}(F)$ be given by

$$\forall q \in \mathbb{P}_{p-1}(F) \qquad \int_F (Q_F v)q\,\Phi_F = \int_F vq, \tag{3.3.13}$$

where $\Phi_F \in S^1_d$ is the face bubble function of (3.3.5) with $\operatorname{supp}\Phi_F = \omega_F$, and, for every mesh element $K \in \mathcal{M}$, let $Q_K : L^2(K) \to \mathbb{P}_{p-2}(K)$ be given by

$$\forall r \in \mathbb{P}_{p-2}(K) \quad \int_K (Q_K v) r \, \Phi_K = \int_K vr, \tag{3.3.14}$$

where $\Phi_K := \prod_{z \in \mathcal{L}_1(K)} \Phi_z^1 \in S^1_{d+1}$ is the element bubble function with $\operatorname{supp} \Phi_K = K$. This leads to the global bubble operators

$$B_{\mathcal{M},p} v := \sum_{K \in \mathcal{M}} (Q_K v) \Phi_K, \quad B_{\mathcal{F},p} v := \sum_{F \in \mathcal{F}^i} \sum_{z \in \mathcal{L}_{p-1}(F)} (Q_F v)(z) \Phi_z^{p-1} \Phi_F,$$

where $B_{\mathcal{F},p}$ incorporates an extension by means of Lagrange basis functions, since $Q_F v = \sum_{z \in \mathcal{L}_{p-1}(F)} (Q_F v)(z) \Phi_z^{p-1}|_F$. The combination of $B_{\mathcal{M},p}$ and $B_{\mathcal{F},p}$ provides a right inverse of $\Pi_S$.

**Lemma 3.3.8** (Higher order bubble smoother)**.** *For any $p \geq 2$, the linear operator $B_p : CR_p \to H^1_0(\Omega)$ defined by*

$$B_p \sigma := B_{\mathcal{F},p} \sigma + B_{\mathcal{M},p} (\sigma - B_{\mathcal{F},p} \sigma) \tag{3.3.15}$$

*satisfies* (3.3.12) *and the local stability estimate*

$$\|\nabla B_p \sigma\|_{L^2(K)} \leq \frac{C_{d,p}}{\rho_K} \left( \sup_{r \neq 0} \frac{\int_K \sigma r}{\|r\|_{L^2(K)}} + \sum_{F \in \mathcal{F}_K} \frac{|K|^{\frac{1}{2}}}{|F|^{\frac{1}{2}}} \sup_{q \neq 0} \frac{\int_F \sigma q}{\|q\|_{L^2(F)}} \right)$$

*where $r$ and $q$ vary in $\mathbb{P}_{p-2}(K)$ and $\mathbb{P}_{p-1}(F)$, respectively.*

*Proof.* The operator $B_p$ is well-defined, because in particular the right-hand sides of (3.3.13) are well-defined moments of any $\sigma \in CR_p$. Moreover, it maps into $H^1_0(\Omega)$, since $\Phi_F \in H^1_0(\Omega)$ for $F \in \mathcal{F}^i$ and $\Phi_K \in H^1_0(\Omega)$ for $K \in \mathcal{M}$.

In order to verify (3.3.12), let $\sigma \in S$ and consider, first, an interior face $F \in \mathcal{F}^i$ and $q \in \mathbb{P}_{p-1}(F)$. In view of $\Phi_{K'|F} = 0$ for $K' \in \mathcal{M}$ and $\Phi_{F'|F} = 0$ for $F' \neq F$, (3.3.13) gives

$$\int_F (B_p \sigma) q = \int_F (Q_F \sigma) \Phi_F q = \int_F \sigma q.$$

Second, let $K \in \mathcal{M}$ and $r \in \mathbb{P}_{p-2}(K)$. Here, thanks to $\Phi_{K'|K} = 0$ for $K' \neq K$, (3.3.14) leads to

$$\int_K (B_p \sigma) r = \int_K (B_{\mathcal{F},p} \sigma) r + \int_K Q_K (\sigma - B_{\mathcal{F},p} \sigma) \Phi_K r = \int_K \sigma r.$$

Finally, let us verify the stability estimate. Employing inverse estimates in $\mathbb{P}_{p+d-1}(K)$ and $\mathbb{P}_{p-1}(F)$ as well as $0 \leq \Phi_K \leq 1$ and (2.1.1), we derive

$$\begin{aligned} \|\nabla B_p \sigma\|_{L^2(K)} &\leq C_{d,p} \rho_K^{-1} \|B_p \sigma\|_{L^2(K)} \\ &\leq C_{d,p} \frac{|K|^{\frac{1}{2}}}{\rho_K |F|^{\frac{1}{2}}} \|Q_F \sigma\|_{L^2(F)} + \frac{C_{d,p}}{\rho_K} \|Q_K \sigma\|_{L^2(K)}. \end{aligned} \tag{3.3.16}$$

Moreover, another inverse estimate in every $\mathbb{P}_{p-2}(K)$ yields

$$\|Q_K\sigma\|_{L^2(K)}^2 \leq C_{d,p} \int_K |Q_K\sigma|^2 \Phi_K = C_{d,p} \int_K \sigma Q_K \sigma,$$

whence

$$\|Q_K\sigma\|_{L^2(K)} \leq C_{d,p} \sup_{r\in\mathbb{P}_{p-2}(K)} \frac{\int_K \sigma r}{\|r\|_{L^2(K)}}. \tag{3.3.17}$$

A similar argument in every $\mathbb{P}_{p-1}(F)$ gives

$$\|Q_F\sigma\|_{L^2(F)} \leq C_{d,p} \sup_{q\in\mathbb{P}_{p-1}(F)} \frac{\int_F \sigma q}{\|q\|_{L^2(F)}}. \tag{3.3.18}$$

We then obtain the stability estimate by inserting (3.3.17) and (3.3.18) into (3.3.16). □

Stabilizing the bubble smoother $B_p$ with simplified nodal averaging $A_p$, we obtain a smoothing operator with the desired properties.

**Proposition 3.3.9** (Stable right inverses of CR-like projections)**.** *Let $p \geq 2$ and $S_p^1 \subseteq S \subseteq CR_p$. The linear operator $E_p : S \to H_0^1(\Omega)$ given by*

$$E_p\sigma := A_p\sigma + B_p(\sigma - A_p\sigma) \tag{3.3.19}$$

*is invariant on $S_p^1$, a right inverse of the Crouzeix-Raviart-like projection $\Pi_S$, and $H_0^1(\Omega)$-stable with stability constant $\leq C_{d,p,\gamma_\mathcal{M}}$.*

*Proof.* We follow the lines of the proof of Proposition 3.3.2 and easily check that $E_p$ is well-defined, provides $H_0^1(\Omega)$-smoothing and is invariant on $S_p^1$. Arguing as in (3.3.8) for any $F \in \mathcal{F}^i$ and any $q \in \mathbb{P}_{p-1}(F)$ as well as for mesh element $K \in \mathcal{M}$ and $r \in \mathbb{P}_{p-2}(K)$, we find that that $E_p$ is a right inverse of $\Pi_S$ onto $S$.

It remains to bound $\|E_p\|_{\mathcal{L}(S,H_0^1(\Omega))}$ appropriately. We let $\sigma \in S$ and write

$$\|\nabla E_p\sigma\|_{L^2(\Omega)} \leq \| \nabla_\mathcal{M} \sigma\|_{L^2(\Omega)} + \| \nabla_\mathcal{M}(\sigma - A_p\sigma)\|_{L^2(\Omega)} + \\ + \|\nabla B_p(\sigma - A_p\sigma)\|_{L^2(\Omega)}.$$

To bound the second and third term, fix a mesh element $K \in \mathcal{M}$. For the second term, we argue as in (3.3.9), with the polynomial degree 1 replaced by $p$, and obtain

$$\|\nabla(\sigma - A_p\sigma)\|_{L^2(K)} \leq C_{d,p,\gamma_\mathcal{M}} \| \nabla_\mathcal{M} \sigma\|_{L^2(\omega_K)}. \tag{3.3.20}$$

Regarding the third term, (2.1.1) gives

$$\sup_{r \in \mathbb{P}_{p-2}(K)} \frac{\int_K (\sigma - A_p\sigma) r}{\|r\|_{L^2(K)}} \leq C_{d,p} \, |K|^{\frac{1}{2}} \sum_{z \in \mathcal{L}_p(\partial K)} \left|\sigma_{|K}(z) - A_p\sigma(z)\right|$$

and, for every $F \in \mathcal{F}_K$,

$$\sup_{q \in \mathbb{P}_{p-1}(F)} \frac{\int_F (\sigma - A_p\sigma) q}{\|q\|_{L^2(F)}} \leq C_{d,p} \, |F|^{\frac{1}{2}} \sum_{z \in \mathcal{L}_p(F)} \left|\sigma_{|K}(z) - A_p\sigma(z)\right|.$$

Employing the stability estimate of Lemma 3.3.8, the last two inequalities and then Lemma 2.2.2, we derive

$$\|\nabla B_1(\sigma - A_p\sigma)\|_{L^2(K)} \lesssim \| \nabla_{\mathcal{M}} \sigma\|_{L^2(\omega_K)}. \tag{3.3.21}$$

Then summing (3.3.20) and (3.3.21) over all mesh elements $K \in \mathcal{M}$ finishes the proof, as for Proposition 3.3.2. □

We let $M_S$ denote the *new Crouzeix-Raviart-like method of arbitrary fixed order* combining the setting (3.3.11) with the smoothing operator $E_p$ in Proposition 3.3.9. We have $M_S = (S, a_{\mathcal{M}}, E_p^{\star})$ and its discrete problem with load $f \in H^{-1}(\Omega)$ reads:

$$U_S \in S \quad \text{such that} \quad \forall \sigma \in S \; \int_\Omega \nabla_{\mathcal{M}} U_S \cdot \nabla_{\mathcal{M}} \sigma = \langle f, E_p\sigma \rangle. \tag{3.3.22}$$

Concerning the computational feasibility of $E_p$, notice that

- it suffices to know the values $\langle f, \Phi_z^p \rangle$ for $z \in \mathcal{L}_p^i$ as well as $\langle f, \Phi_z^{p-1}\Phi_F \rangle$ for $F \in \mathcal{F}^i$, $z \in \mathcal{L}_{p-1}(F)$, and $\langle f, \Phi_z^{p-2}\Phi_K \rangle$ for $K \in \mathcal{M}$, $z \in \mathcal{L}_{p-2}(K)$,
- $E_p$ is local in that, if $\omega$ is the support of a basis function $\Phi$ from references [32, 41, 57], then $\omega$ is a mesh element, a pair or a star of elements and $\operatorname{supp} E\Phi \subset \cup_{K \subset \omega} \omega_K$,
- the operators $Q_F$ and $Q_K$ in (3.3.13) and (3.3.14) can be implemented by means of matrices which are precalculated on a reference element and, for $d = 2$ and $Q_F$, can be diagonalized with the help of Legendre polynomials.

In contrast to the methods in [32, 41, 57], method $M_S$ enjoys the following property.

**Theorem 3.3.10** (Quasi-optimality of $M_S$)**.** *For any $p \geq 2$ and any subspace $S$ with $S_p^1 \subseteq S \subseteq CR_p$, the method $M_S$ is a $\| \nabla_{\mathcal{M}} \cdot \|$-quasi-optimal nonconforming Galerkin method for the Poisson problem* (3.3.1) *with quasi-optimality constant $\leq C_{d,p,\gamma_{\mathcal{M}}}$.*

*Proof.* Use Proposition 3.3.9 in Corollary 3.2.8. □

*Remark* 3.3.11 (Alternative smoothing operator). The evaluation of the right-hand side of problem (3.3.22) is less expensive if we replace the smoothing operator $E_p$ with $\widetilde{E}_p : S \to H^1_0(\Omega)$ defined as

$$\widetilde{E}_p\sigma := A_1\sigma + B_p(\sigma - A_1\sigma).$$

Indeed, the lowest-order averaging $A_1$ is also defined on $S$ and less expensive to evaluate than $A_p$. According to Lemmas 3.3.8 and 2.2.3, $\widetilde{E}_p$ is also a right inverse of the orthogonal projection $\Pi_S$ and $H^1_0(\Omega)-$stable with stability constant $\leq C_{d,p,\gamma_\mathcal{M}}$. On the other hand, unlike $E_p$, it is not invariant on the conforming subspace $S^1_p$ of $S$. Thus, from the viewpoint of the abstract theory, $(S, a_\mathcal{M}, \widetilde{E}^\star_p)$ differs from $(S, a_\mathcal{M}, E^\star_p)$ only in that it is not a nonconforming Galerkin method. We consider the possibility of stabilizing higher-order elements with the lowest-order averaging also for DG methods in Proposition 4.2.11. Remarkably, in that context, the fact that we cannot obtain nonconforming Galerkin methods by means of $A_1$ results in a more pessimistic bound for the quasi-optimality constant.

### 3.3.3 A Quasi-Optimal Morley Method for the Biharmonic Problem

This section constructs a quasi-optimal Morley method for the 'biharmonic equation' with clamped boundary conditions,

$$\Delta^2 u = f \text{ in } \Omega, \quad u = 0 \text{ and } \partial_n u = 0 \text{ on } \partial\Omega, \tag{3.3.23}$$

where $\Omega$ and $\mathcal{M}$ are as in §2.1, $d = 2$, and $\#\mathcal{M} > 1$. We set

$$a_\mathcal{M}(w_1, w_2) := \int_\Omega \mathrm{D}^2_\mathcal{M}\, w_1 : \mathrm{D}^2_\mathcal{M}\, w_2, \qquad w_1, w_2 \in H^2(\mathcal{M}),$$

and aim at applying Corollary 3.2.8 with the following setting:

$$V = H^2_0(\Omega) \qquad \text{and} \qquad S = MR,$$

$$MR := \left\{ s \in S^0_2 \mid s \text{ is cont. in } \mathcal{L}_1,\ s_{|\mathcal{L}^b_1} = 0,\ \forall F \in \mathcal{F}\ \int_F [\![\nabla s]\!] \cdot n = 0 \right\},$$

$$\widetilde{a} = a_{\mathcal{M}|\widetilde{V}\times\widetilde{V}} \quad \text{with} \quad \widetilde{V} := H^2_0(\Omega) + MR,$$

where $a_{\mathcal{M}|V\times V}$ provides a weak formulation of $\Delta^2$ and $MR$ is the Morley space [50] over $\mathcal{M}$. In order to recall some useful properties of $MR$, let $n_F$ and $t_F$ be normal and tangent unit vectors for every edge $F \in \mathcal{F}$, with arbitrary but fixed orientation. The functionals $s \mapsto s(z)$, $z \in \mathcal{L}^i_1$, and $s \mapsto \int_F \nabla s \cdot n_F$, $F \in \mathcal{F}^i$, are well-defined for any $s \in MR$ and determine it. Furthermore, the integrals $\int_F \nabla s \cdot t_F$ and so also $\int_F \nabla s = |F| \nabla s(m_F)$ are well-defined for all $F$ and vanish if $F \in \mathcal{F}^b$. Hence, $a_\mathcal{M}$ induces the norm $\| \mathrm{D}^2_\mathcal{M} \cdot \|_{L^2(\Omega)}$ on $\widetilde{V}$.

*Remark* 3.3.12 (Poor conforming part). The conforming part $MR \cap H_0^2(\Omega)$ of the Morley space can be quite small, thereby providing only poor approximation properties; cf. de Boor and DeVore [36, Theorem 3]. We illustrate this with an extreme example. Given any $n \in \mathbb{N}$, subdivide $\Omega = (0,1)^2$ into $n^2$ squares of equal size and obtain $\mathcal{M}$ by inserting in each square the diagonal parallel to the line $\{(x,x) \mid x \in \mathbb{R}\}$. Then $MR \cap H_0^2(\Omega) = \{0\}$.

We refer to the $a_{\mathcal{M}}$-orthogonal projection of $\widetilde{V}$ onto $MR$ as the Morley projection $\Pi_{MR}$. As before, the first step is to describe right inverses thereof.

**Lemma 3.3.13** (Right inverses of Morley projection)**.** *Given any linear operator $E : MR \to H_0^2(\Omega)$, we have $\Pi_{MR}E = \mathrm{Id}_{MR}$ if and only if, for all $\sigma \in MR$,*

$$\forall z \in \mathcal{L}_1^i \; E\sigma(z) = \sigma(z) \quad \textit{and} \quad \forall F \in \mathcal{F}^i \; \int_F \nabla E\sigma \cdot n_F = \int_F \nabla\sigma \cdot n_F. \tag{3.3.24}$$

*Proof.* Let us first characterize $\Pi_{MR}v$ for any $v \in H_0^2(\Omega)$. Defining $\sigma \in MR$ by

$$\forall z \in \mathcal{L}_1^i \; \sigma(z) = v(z) \quad \text{and} \quad \forall F \in \mathcal{F}^i \; \int_F \nabla\sigma \cdot n_F = \int_F \nabla v \cdot n_F, \tag{3.3.25}$$

we have $\int_F \nabla v = \int_F \nabla\sigma$. Thus, integrating piecewise by parts, we infer

$$\forall s \in MR \quad a_{\mathcal{M}}(s, \sigma - v) = \sum_{K \in \mathcal{M}} \sum_{F \in \mathcal{F}_K} \int_F \mathrm{D}^2(s_{|K}) n_K \cdot \nabla(\sigma - v) = 0$$

because $\mathrm{D}^2_{\mathcal{M}} s$ is piecewise constant on $\mathcal{M}$. Since the Morley projection of $v$ is unique, we derive that $\sigma = \Pi_{MR}v$ and (3.3.25) characterizes $\Pi_{MR}v$. This characterization readily yields the claimed equivalence. □

In order to construct a right inverse of $\Pi_{MR}$ that is stable under mesh refinement, we again mimic the approach of §3.3.1. Technical difficulties arise from the stronger regularity requirement $E\sigma \in H_0^2(\Omega)$; in particular, neither $A_2$ nor $B_2$ fulfill such condition. In order to replace the former, we employ the HCT space (2.2.9) and restrict to $MR$ the simplified averaging operator $A_{HCT} : S_2^0 \to HCT$ defined in (2.2.10).

The operator $A_{HCT}$ incidentally fulfills the first part of (3.3.24). Aiming at a right inverse of the form $A_{HCT} + B_{\partial_n}(\mathrm{Id}_{MR} - A_{HCT})$, we thus only need to adjust the means of the normal derivative across interior faces by a suitable $H_0^2(\Omega)$-bubble smoother $B_{\partial_n}$. To this end, we replace the face bubbles in the bubble smoother $B_1$ of §3.3.1 by the following ones inspired by Verfürth [66]. Given any interior edge $F \in \mathcal{F}^i$, let $K_1, K_2 \in \mathcal{M}$ be the two elements such that $F = K_1 \cap K_2$ and consider their barycentric coordinates

$(\lambda_z^{K_i})_{z\in\mathcal{L}_1(K_i)}$, $i=1,2$, as first-order polynomials on $\mathbb{R}^2$. Then

$$\bar\phi_F := \frac{30}{|F|}\phi_F \quad\text{with}\quad \phi_F := \begin{cases} \prod\limits_{z\in\mathcal{L}_1(F)} \big(\lambda_z^{K_1}\lambda_z^{K_2}\big)^2 & \text{in } K_1\cup K_2,\\ 0 & \text{in } \Omega\setminus(K_1\cup K_2)\end{cases}$$

is an $H_0^2(\Omega)$-counterpart of the normalized face bubble $\bar\Phi_F$ from (3.3.5) and

$$\bar\Phi_{n_F} := \zeta_F\bar\phi_F \quad\text{with}\quad \zeta_F(x) := (x-m_F)\cdot n_F,\ x\in\mathbb{R}^2, \tag{3.3.26}$$

is in $H_0^2(\Omega)$ and satisfies $\int_{F'}\nabla\bar\Phi_{n_F}\cdot n_{F'} = \int_{F'} n_F\cdot n_{F'}\bar\phi_F = \delta_{F,F'}$ for all $F'\in\mathcal{F}^i$ thanks to (2.1.1). Hence, the operator $B_{\partial_n} : MR + HCT \to H_0^2(\Omega)$ given by

$$B_{\partial_n}\sigma := \sum_{F\in\mathcal{F}^i}\left(\int_F \nabla\sigma\cdot n_F\right)\bar\Phi_{n_F}$$

provides $H_0^2(\Omega)$-smoothing with

$$\forall F\in\mathcal{F}^i \qquad \int_F \nabla(B_{\partial_n}\sigma)\cdot n_F = \int_F \nabla\sigma\cdot n_F. \tag{3.3.27}$$

We have the following scaling of the bubbles $\bar\Phi_{n_F}$.

**Lemma 3.3.14** (Scaling of $H^2$ face bubbles). *If $K, K'\in\mathcal{M}$ are the two elements containing the interior edge $F\in\mathcal{F}^i$, we have*

$$\|\,\mathrm{D}^2\,\bar\Phi_{n_F}\|_{L^2(K)} \le C\gamma_K^3\gamma_{K'}^2\frac{|K|^{\frac12}}{\rho_K|F|}.$$

*Proof.* If we use an inverse inequality in $\mathbb{P}_9(K)$ and $|\zeta_F|\le h_K$ on $K$, we obtain

$$\|\,\mathrm{D}^2\,\bar\Phi_{n_F}\|_{L^2(K)} \le C\rho_K^{-2}\|\bar\Phi_{n_F}\|_{L^2(K)} \le C\rho_K^{-2}h_K|F|^{-1}\|\phi_F\|_{L^2(K)}.$$

Moreover, we have $|\lambda_z^{K'}|\le h_K|\nabla\lambda_z^{K'}|\le \gamma_K\,|F|\,\rho_{K'}^{-1}\le\gamma_K\gamma_{K'}$ in $K$, for any $z\in\mathcal{L}_1(F)$. Using this and (2.1.1), we finish the proof with

$$\|\phi_F\|_{L^2(K)} \le \gamma_K^2\gamma_{K'}^2\,\|\prod_{z\in\mathcal{L}_1(F)}(\lambda_z^K)^2\|_{L^2(K)} = \gamma_K^2\gamma_{K'}^2\frac{|K|^{\frac12}}{\sqrt{180}}. \qquad \square$$

Owing to this auxiliary result and the scaling of the HCT basis functions in Lemma 2.2.7, we obtain a stable right inverse of $\Pi_{MR}$, combining $A_{HCT}$ and $B_{\partial_n}$ as before.

**Proposition 3.3.15** (Stable right inverse of Morley projection). *The linear operator $E_{MR} : MR\to H_0^2(\Omega)$ given by*

$$E_{MR}\sigma := A_{HCT}\sigma + B_{\partial_n}(\sigma - A_{HCT}\sigma)$$

*is invariant on $MR\cap H_0^2(\Omega)$, a right inverse of the Morley projection $\Pi_{MR}$, and $H_0^2(\Omega)$-stable with stability constant $\le C_{\gamma_\mathcal{M}}$.*

*Proof.* The operator $E_{MR}$ is invariant on $MR \cap H^2_0(\Omega)$, because $A_{HCT}$ is invariant on $MR \cap HCT = MR \cap H^2_0(\Omega)$. In order to check that $E_{MR}$ is a right inverse of $\Pi_{MR}$, we verify condition (3.3.24) in Lemma 3.3.13 and let $\sigma \in MR$. First, given a Lagrange node $z \in \mathcal{L}^i_1$, we have $A_{HCT}\sigma(z) = \sigma(z)$ and so $E_{MR}\sigma(z) = \sigma(z)$, because each bubble $\bar{\Phi}_{n_F}$, $F \in \mathcal{F}^i$, vanishes in $\mathcal{L}_1$. Second, given an interior edge $F \in \mathcal{F}^i$, we derive $\int_F \nabla E_{MR}\sigma \cdot n_F = \int_F \nabla \sigma \cdot n_F$ as in (3.3.8) by means of (3.3.27).

We may finish the proof by bounding $\|\mathrm{Id}_{MR} - E_{MR}\|_{\mathcal{L}(MR, H^2_0(\Omega))}$ appropriately. To this end, let $\sigma \in MR$, fix a mesh element $K \in \mathcal{M}$, and write

$$\begin{aligned}\| \mathrm{D}^2(E_{MR}\sigma - \sigma)\|_{L^2(K)} \leq & \| \mathrm{D}^2(\sigma - A_{HCT}\sigma)\|_{L^2(K)} + \\ & + \| \mathrm{D}^2 B_{\partial_n}(\sigma - A_{HCT}\sigma)\|_{L^2(K)}.\end{aligned}$$

For the first term on the right-hand side, we proceed as in (3.3.9). Combining inequality (2.2.11b) from Lemma 2.2.5 with the scaling of the HCT basis functions (2.2.12b) in Lemma 2.2.7, we obtain

$$\tag{3.3.28}\begin{aligned}& \| \mathrm{D}^2(\sigma - A_{HCT}\sigma)\|_{L^2(K)} \leq \\ & \leq \sum_{z \in \mathcal{L}_1(K)} \sum_{j=1}^{2} \left|\partial_j(\sigma_{|K})(z) - (\partial_j A_{HCT}\sigma)(z)\right| \| \mathrm{D}^2 \Upsilon^j_z\|_{L^2(K)} \\ & \leq C \sum_{z \in \mathcal{L}_1(K)} \sum_{K' \in \mathcal{M}, K' \ni z} \frac{h_{K'}}{\rho_K} \frac{|K|^{\frac{1}{2}}}{|K'|^{\frac{1}{2}}} \| \mathrm{D}^2 \sigma\|_{L^2(K')} \\ & \leq C_{\gamma_{\mathcal{M}}} \| \mathrm{D}^2_{\mathcal{M}} \sigma\|_{L^2(\omega_K)}.\end{aligned}$$

For the second term, we expand again $(\sigma - A_{HCT}\sigma)_{|K}$ and obtain

$$\begin{aligned}& B_{\partial_n}(\sigma - A_{HCT}\sigma) = \\ & \sum_{F \in \mathcal{F}_K \cap \mathcal{F}^i} \sum_{z \in \mathcal{L}_1(F)} \sum_{j=1}^{2} \left[\partial_j(\sigma_{|K})(z) - (\partial_j A_{HCT}\sigma)(z)\right] \left(\int_F \nabla \Upsilon^j_z \cdot n_F\right) \bar{\Phi}_{n_F}\end{aligned}$$

in $K$. Consequently, we may argue as for the previous term, with the help of (2.2.11b), (2.2.12a) and Lemma 3.3.14

$$\| \mathrm{D}^2 B_{\partial_n}(\sigma - A_{HCT}\sigma)\|_{L^2(K)} \lesssim \| \mathrm{D}^2_{\mathcal{M}} \sigma\|_{L^2(\omega_K)}.$$

We can finish the proof as for Proposition 3.3.2, by summing over all mesh elements $K \in \mathcal{M}$, and observing that the number of elements in each star $\omega_K$ is $\leq C_{\gamma_{\mathcal{M}}}$. □

Let $M_{MR}$ denote the *new Morley method* for the biharmonic problem (3.3.23), with the setting of this section and the smoothing operator $E_{MR}$

from Proposition 3.3.15. Then $M_{MR} = (MR, a_{\mathcal{M}}, E^{\star}_{MR})$ and its discrete problem for $f \in H^{-2}(\Omega)$ reads

$$U_{MR} \in MR \text{ such that } \forall \sigma \in MR \int_{\Omega} \mathrm{D}^2_{\mathcal{M}} U_{MR} : \mathrm{D}^2_{\mathcal{M}} \sigma = \langle f, E_{MR}\sigma \rangle.$$

The smoother $E_{MR}$ is computationally feasible in that

- it suffices to know the evaluations $\langle f, \Upsilon^j_z \rangle$ for $z \in \mathcal{L}^i_1$, $j \in \{0,1,2\}$, and $\langle f, \Upsilon_F \rangle$ for $F \in \mathcal{F}^i$, as well as $\langle f, \bar{\Phi}_{n_F} \rangle$ for $F \in \mathcal{F}^i$,
- $E_{MR}$ is local: if $\omega$ is the support of a Morley basis function, then $\omega$ is a pair or a star of elements and $\operatorname{supp} E_{MR}\Phi \subset \cup_{K \subset \omega} \omega_K$.

The approximation properties of $M_{MR}$ are superior to the original Morley method in the following sense.

**Theorem 3.3.16** (Quasi-optimality of $M_{MR}$). *We have that the method $M_{MR}$ is a $\| \mathrm{D}^2_{\mathcal{M}} \cdot \|$-quasi-optimal nonconforming Galerkin method for the biharmonic problem* (3.3.23) *with quasi-optimality constant $\leq C_{\gamma_{\mathcal{M}}}$.*

*Proof.* Use Proposition 3.3.15 in Corollary 3.2.8. □

*Remark* 3.3.17 (Alternative simplified nodal averaging into rHCT). One obtains a variant of $M_{MR}$ by replacing in $E_{MR}$ the simplified nodal averaging $A_{HCT}$ from (2.2.10) by

$$A_{rHCT}\sigma := \sum_{z \in \mathcal{L}^i_1} \left( \sigma(z)\Theta^0_z + \sum_{j=1}^{2} \partial_j(\sigma_{|K_z})(z)\Theta^j_z \right),$$

where $\Theta^j_z$, $z \in \mathcal{L}^i_1$, $j \in \{0,1,2\}$, are the nodal basis functions of the reduced HCT space from Ciarlet [30]. As Lemmas 2.2.5 and 2.2.7 carry over to the new basis and the reduced HCT space contains $MR \cap H^2_0(\Omega)$, this modification of $M_{MR}$ is also a quasi-optimal nonconforming Galerkin method for (3.3.23) with quasi-optimality constant $\leq C_{\gamma_{\mathcal{M}}}$.

# Chapter 4

# DG and Other Interior Penalty Methods

This chapter follows the same line as [65]. We apply the framework developed in Chapter 1 to design and analyze quasi-optimal finite element methods with interior penalty. In contrast to the previous examples, overconsistency cannot be achieved in the applications proposed here and, consequently, it cannot be the guiding principle of our construction. To give an overview, let us illustrate the setting and main results in the case of approximating the Poisson problem with discontinuous linear elements via the symmetric interior penalty (SIP) method. Interior penalty methods were first studied by Baker [9], Wheeler [67] and Arnold [1].

## 4.1 Overview

Let $u \in H_0^1(\Omega)$ be the weak solution of the Poisson problem (3.3.1) and let $\mathcal{M}$ be a mesh of the domain $\Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$. In the notation of §2.1, the SIP approximation $U \in S_1^0$ solves the discrete problem

$$\forall \sigma \in S_1^0 \qquad b(U, \sigma) = \int_\Omega f\sigma \tag{4.1.1}$$

where $f \in L^2(\Omega)$, the bilinear form $b := b_1 + b_2$ is given by

$$\begin{aligned} b_1(s,\sigma) &:= \int_\Omega \nabla_{\mathcal{M}} s \cdot \nabla_{\mathcal{M}} \sigma - \int_\Sigma \{\!\{\nabla s\}\!\} \cdot n \, [\![\sigma]\!] , \\ b_2(s,\sigma) &:= \int_\Sigma \frac{\eta}{h} [\![s]\!] \, [\![\sigma]\!] - \int_\Sigma [\![s]\!] \, \{\!\{\nabla \sigma\}\!\} \cdot n, \end{aligned}$$

and the penalty parameter $\eta > 0$ is so large that $b$ is coercive. Replacing $s$ by $u \in H^1_0(\Omega)$, we see that

$$u \in H^2(\Omega) \implies \forall \sigma \in S^0_1 \;\; b_1(u, \sigma) = \int_\Omega f\sigma, \tag{4.1.2a}$$

$$\forall \sigma \in S^0_1 \;\; b_2(u, \sigma) = 0. \tag{4.1.2b}$$

Hence, $b_2$ establishes symmetry and coercivity, without impairing the *consistency* provided by $b_1$. These properties can be used to derive convergence up to optimal order in the norm

$$|v|^2_{1;\eta} := \int_\Omega |\nabla_{\mathcal{M}} v|^2 + \int_\Sigma \frac{\eta}{h} | [\![ v ]\!] |^2, \quad v \in H^1_0(\Omega) + S^0_1,$$

cf. Di Pietro and Ern [37, Theorem 4.17] and Gudi [43, §3.2]. However, the extension of $b_1$ underlying (4.1.2a) and the right-hand side in the discrete problem (4.1.1) are not defined for general $f \in H^{-1}(\Omega)$. This observation and the abstract argument in Remark 1.4.9 entail that the SIP method (4.1.1) is not $|\cdot|_{1;\eta}$-quasi-optimal and so does not always fully exploit the approximation potential offered by its discrete space $S^0_1$.

In order to achieve quasi-optimality, Theorem 1.4.14 suggests to consider the following variant of the discrete problem (4.1.1): find $U_E \in S^0_1$ such that

$$\forall \sigma \in S^0_1 \quad b(U_E, \sigma) = \langle f, E\sigma \rangle, \tag{4.1.3}$$

where the smoother $E : S^0_1 \to H^1_0(\Omega)$, to be specified, enables $f \in H^{-1}(\Omega)$. If we extend (3.1.5) and require that the means on internal faces are conserved, as in Badia et al. [8],

$$\forall \sigma \in S^0_1,\, F \in \mathcal{F}^i \quad \int_F E\sigma = \int_F \{\!\{ \sigma \}\!\}, \tag{4.1.4}$$

then piecewise integrating by parts with (2.1.3) shows

$$\forall s, \sigma \in S^0_1 \quad b_1(s, \sigma) = \int_\Sigma [\![ \nabla s ]\!] \cdot n \, \{\!\{ \sigma \}\!\} = \int_\Omega \nabla_{\mathcal{M}} s \cdot \nabla (E\sigma).$$

Interestingly, the right-hand side provides a new extension $\widetilde{b}_1$ of $b_1$ onto $H^1_0(\Omega)$ which improves upon (4.1.2a) in that

$$\forall u \in H^1_0(\Omega),\, \sigma \in S^0_1 \quad \widetilde{b}_1(u, \sigma) = \langle f, E\sigma \rangle.$$

In order to define a smoothing operator $E$ that satisfies (4.1.4) and is computionally feasible, we simply extend the construction proposed in §3.3.1 and ensure that the operator norm $\|E\|_{\mathcal{L}(S^0_1, H^1_0(\Omega))}$ is bounded in terms of the shape coefficient $\gamma_{\mathcal{M}}$ of $\mathcal{M}$ and the space dimension $d$.

Exploiting the full stability and full algebraic consistency delivered by (4.1.3) and (4.1.4), the abstract theory of Chapter 1 then yields

$$|u - U_E|_{1;\eta} \leq \left(1 + C\eta^{-1}\right)^{\frac{1}{2}} \inf_{s \in S_1^0} |u - s|_{1;\eta} \,,$$

where $C$ depends on $d$ and $\gamma_{\mathcal{M}}$ and $\eta$ is sufficiently large. Notably, as $\eta \to \infty$, the discontinuous space $S_1^0$ is replaced by $S_1^1$ and we end up exactly in Céa's lemma for the conforming Galerkin method with $S_1^1$.

The rest of this chapter is organized as follows. We first summarize in a convenient form the relevant result from Chapter 1 to be used here. Then, we introduce new variants of various interior penalty methods and prove their quasi-optimality. Firstly, we design quasi-optimal DG methods of arbitrary fixed order for the Poisson problem, covering also the setting illustrated in this introduction. Secondly, we devise a quasi-optimal Crouzeix-Raviart method with jump penalty for linear elasticity and establish a robust bound for its error in the nearly-incompressible regime. Lastly, we conclude with a quasi-optimal variant of the quadratic $C^0$-interior penalty method for the biharmonic problem.

As before, we consider polyhedral domains with Lipschitz boundaries and homogeneous essential boundary conditions.

## 4.2 Applications to Interior Penalty Methods

As in the previous chapter, we design nonconforming methods $M$ with discrete problem

$$\forall \sigma \in S \quad b(M\ell, \sigma) = \langle \ell, E\sigma \rangle, \tag{4.2.1}$$

corresponding to the triplet $(S, b, E^\star)$, where $E : S \to V$ is a linear smoothing operator. For each example we shall refer to the following result, which differ from Theorem 3.2.1 only in that we exploit the consistency measure $\delta_V$ from Proposition 1.3.6 instead of $\delta_S$. According to Remark 1.4.20, this approach to nonconforming consistency is closely related to the so-called second Strang lemma [10].

**Theorem 4.2.1** (Stability, consistency, and quasi-optimality)**.** *Given a nonconforming method $M = (S, b, E^\star)$ for* (1.2.1) *and an extended scalar product $\widetilde{a}$ on $\widetilde{V} = V + S$, introduce the bilinear form $d : \widetilde{V} \times S \to \mathbb{R}$ by*

$$d(\widetilde{v}, \sigma) := b(\Pi_S \widetilde{v}, \sigma) - \widetilde{a}(\widetilde{v}, E\sigma),$$

*where $\Pi_S$ denotes the $\widetilde{a}$-orthogonal projection onto $S$. Then:*

*(i) $M$ is fully stable, with*

$$C_{\text{stab}} := \|M\|_{\mathcal{L}(V', S)} = \sup_{\sigma \in S} \frac{\|E\sigma\|}{\sup_{s \in S, \|s\|=1} b(s, \sigma)}.$$

*(ii) $M$ is quasi-optimal if and only if it is fully algebraically consistent, in that*

$$\forall u \in S \cap V, \sigma \in S \quad 0 = d(u, \sigma) = b(u, \sigma) - a(u, E\sigma).$$

*(iii) If $M$ is quasi-optimal, then its quasi-optimality constant satisfies*

$$C_{\mathrm{stab}} \leq C_{\mathrm{qopt}} = \sqrt{1 + \delta_V^2},$$

*where $\delta_V \in [0, \infty)$ is the consistency measure given by the smallest constant in*

$$\forall v \in V, \sigma \in S \quad |d(v, \sigma)| \leq \delta_V \sup_{\hat{s} \in S, \|\hat{s}\|=1} b(\hat{s}, \sigma) \inf_{s \in S} \|v - s\|.$$

*Proof.* Item (i) follows from Theorem 1.4.7, while (ii) is a consequence of Theorem 1.4.14 and (i). Finally, the first part of Theorem 1.4.19 guarantees the validity of item (iii). □

This theorem is formulated with the following viewpoint. The discrete bilinear form decomposes as $b = \widetilde{a}(\cdot, E\cdot) + d(\cdot, \cdot)$ on $S$ and $d(\cdot, \cdot) = 0$ corresponds to an overconsistent method, in the vein of §3.2. According to item (ii) of Theorem 4.2.1, we can achieve quasi-optimality also if $b \neq \widetilde{a}(\cdot, E\cdot)$, provided the perturbation introduced by $d(\cdot, \cdot)$ is compatible with full algebraic consistency. As we pointed out in Lemma 3.2.4, setting $d(\cdot, \cdot) \neq 0$ is sometimes necessary to enforce the nondegeneracy of $b$.

The effect of the perturbation induced by $d(\cdot, \cdot)$ on the size of the quasi-optimality constant could be quantified by both the consistency measures devised in §1.3.2. More specifically, the size of the constant $\delta_S$ would somehow quantify the distance of the method $(S, b, E^\star)$ from overconsistency, which indeed corresponds to $\delta_S = 0$. However, we shall proceed here as suggested by item (iii) of Theorem 4.2.1 and access the quasi-optimality constant through the other consistency measure $\delta_V$. In fact, this approach seems to provide slightly better estimates, in particular for large values of the penalty parameters used in this chapter to enforce the coercivity of the discrete bilinear forms; cf. Remark 4.2.8.

It is therefore of interest to bound the stability constant $C_{\mathrm{stab}}$ and $\delta_V$, connecting them to another well-known and important, but not yet mentioned constant.

*Remark* 4.2.2 (Stability, consistency and inf-sup constants). Consider any nonconforming method $M = (S, b, E^\star)$. As $S$ is finite-dimensional, the nondegeneracy of $b$ entails that the corresponding inf-sup constant is positive:

$$\alpha := \inf_{\sigma \in S, \|\sigma\|=1} \sup_{s \in S, \|s\|=1} b(s, \sigma) > 0.$$

Then the definitions of $C_{\mathrm{stab}}$ and $\delta_V$ readily yield

$$C_{\mathrm{stab}} \leq \frac{\|E\|_{\mathcal{L}(S,V)}}{\alpha} \quad \text{and} \quad \delta_V \leq \frac{\gamma}{\alpha} \tag{4.2.2}$$

where $\gamma \geq 0$ verifies $|d(v,\sigma)| \leq \gamma \inf_{s \in S} \|v - s\| \|\sigma\|$ for all $v \in V$ and $\sigma \in S$. Hence, up to the inverse of the inf-sup constant $\alpha$, the constants $C_{\mathrm{stab}}$ and $\delta_V$ depend, respectively, only on the smoothing operator $E$ and the bilinear form $d$. It is worth noting that these bounds may be pessimistic, as it is pointed out in Remark 3.2.7 for $C_{\mathrm{stab}}$.

In view of Theorem 4.2.1, we may achieve quasi-optimality by the following steps: given a continuous problem in the form of (1.2.1) and a nonconforming finite element space $S$,

- extend the scalar product $a$ to the sum $\widetilde{V} = V + S$,
- find a computationally feasible smoothing operator $E : S \to V$, possibly with $E_{|V \cap S} = \mathrm{Id}_{V \cap S}$,
- if necessary, use the bilinear form $d$ to arrange that $b = \widetilde{a}(\cdot, E\cdot) + d$ is nondegenerate and has other optional properties like symmetry.

Recall also that the condition $E_{|V \cap S} = \mathrm{Id}_{V \cap S}$ is not necessary for quasi-optimality but characterizes the subclass of nonconforming Galerkin methods from (1.2.11). In this chapter, we shall propose methods with and without this property.

We shall carry out the aforementioned steps for three different settings, involving vector and fourth order problems as well as various couplings between elements (completely discontinuous, Crouzeix-Raviart, continuous). In each case the nondegeneracy of $b$ will be obtained by means of interior penalties.

### 4.2.1 Quasi-Optimal Discontinuous Galerkin Methods for the Poisson Problem

In this subsection we devise quasi-optimal DG methods for the Poisson problem, covering the results illustrated in the overview §4.1.

Let $\Omega$ and $\mathcal{M}$ be as in §2.1 and, with $\eta \geq 0$, define

$$(v,w)_{1;\eta} := \int_\Omega \nabla_{\mathcal{M}} v \cdot \nabla_{\mathcal{M}} w + \sum_{F \in \mathcal{F}} \frac{\eta}{h_F} \int_F [\![v]\!] \, [\![w]\!] \,, \qquad |v|_{1;\eta} := (v,v)_{1;\eta}^{\frac{1}{2}}$$

on $H^1(\mathcal{M})$ and abbreviate $(\cdot,\cdot)_{1;0}$ to $(\cdot,\cdot)_1$. We consider the following setting

$$V = H^1_0(\Omega), \qquad S = S^0_p, \qquad \widetilde{a} = (\cdot,\cdot)_{1;\eta} \text{ on } \widetilde{V} = H^1_0(\Omega) + S^0_p \tag{4.2.3}$$

for any fixed $p \in \mathbb{N}$. Then $\widetilde{a}$ is a scalar product for $\eta > 0$, with induced norm $|\cdot|_{1,\eta}$, and the abstract variational problem (1.2.1) provides a weak formulation of the Poisson problem (3.3.1). Our setting has two parameters: the polynomial degree $p$ and the scaling factor $\eta$ of the jumps; the latter will be also the penalty parameter. In order to keep notation simple, we shall sometimes suppress the dependencies on $p$ and $\eta$. The conforming part of $S^0_p$ is the strict subspace $S^0_p \cap H^1_0(\Omega) = S^1_p$. Moreover, we easily see that

$$\emptyset \neq S^0_0 \subseteq S^0_p \cap V^\perp, \tag{4.2.4}$$

which precludes overconsistency in light of Lemma 3.2.4.

In order to obtain hints for a suitable choice of the smoothing operator, we invoke integration by parts element by element and the structure of $S^0_p$. Let $s, \sigma \in S^0_p$ be arbitrary. On the one hand, the integration by parts formula (2.1.3) yields

$$(s, E\sigma)_{1;\eta} = \sum_{K \in \mathcal{M}} \int_K (-\Delta s) E\sigma + \sum_{F \in \mathcal{F}^i} \int_F [\![\nabla s]\!] \cdot n E\sigma$$

due to $E\sigma \in H^1_0(\Omega)$. On the other hand, we want $\int_\Omega \nabla_{\mathcal{M}} s \cdot \nabla_{\mathcal{M}} \sigma = (s, \sigma)_1$ to appear in the discrete bilinear form. For this term, (2.1.2b) and (2.1.3) give

$$(s, \sigma)_1 = \sum_{K \in \mathcal{M}} \int_K (-\Delta s)\sigma + \sum_{F \in \mathcal{F}^i} \int_F [\![\nabla s]\!] \cdot n \{\!\{\sigma\}\!\} \quad + \sum_{F \in \mathcal{F}} \int_F \{\!\{\nabla s\}\!\} \cdot n [\![\sigma]\!] .$$

A comparison of these two identities suggests that the smoothing operator $E$ should conserve certain moments on faces and elements and proves the following lemma. Such moment conservation was already used in Badia et al. [8, §6] to design a DG method for the Stokes problem with a partial quasi-optimality result for the velocity field. It is also an extension of the sufficient condition devised in Lemma 3.3.7 to construct overconsistent Crouzeix-Raviart-like methods of arbitrary fixed order.

**Lemma 4.2.3** (Conservation of moments)**.** *Let $p \in \mathbb{N}$ and, for notational convenience, set $\mathbb{P}_{-1}(K) = \emptyset$ for all $K \in \mathcal{M}$. If a smoothing operator $E : S^0_p \to H^1_0(\Omega)$ satisfies*

$$\int_F q(E\sigma) = \int_F q \{\!\{\sigma\}\!\} \quad \textit{and} \quad \int_K r(E\sigma) = \int_K r\sigma \tag{4.2.5}$$

*for all $F \in \mathcal{F}^i$, $q \in \mathbb{P}_{p-1}(F)$, $K \in \mathcal{M}$, $r \in \mathbb{P}_{p-2}(K)$ and $\sigma \in S^0_p$, then*

$$(s, E\sigma)_{1;\eta} = \int_\Omega \nabla_{\mathcal{M}} s \cdot \nabla_{\mathcal{M}} \sigma - \sum_{F \in \mathcal{F}} \int_F \{\!\{\nabla s\}\!\} \cdot n [\![\sigma]\!] \tag{4.2.6}$$

*for all $s, \sigma \in S^0_p$.*

The analogy with Lemma 3.3.7 suggests to adapt the construction of the smoothing operators in §3.3.2 to the current setting. Since only minor modifications are needed, we do not change the notation. We begin with an extension of the so-called bubble smoother in (3.3.15), with the help of the same weighted $L^2$-projections introduced before.

For every interior face $F \in \mathcal{F}^i$, let us recall the face bubble function $\Phi_F = \prod_{z \in \mathcal{L}_1(F)} \Phi_z^1 \in S_d^1$, which is supported in the two elements containing $F$. Then, let the operator $Q_F : L^2(F) \to \mathbb{P}_{p-1}(F)$ be given by (3.3.13) and define $B_{\mathcal{F},p} : H^1(\mathcal{M}) \to H_0^1(\Omega)$ by

$$B_{\mathcal{F},p} v := \sum_{F \in \mathcal{F}^i} \sum_{z \in \mathcal{L}_{p-1}(F)} \big(Q_F \{\!\!\{v\}\!\!\}\big)(z) \Phi_z^{p-1} \Phi_F.$$

As before, we incorporate an extension by means of Lagrange basis functions, in view of the partition of unity $\sum_{z \in \mathcal{L}_{p-1}(F)} \Phi_z^{p-1} = 1$.

Next, for every mesh element $K \in \mathcal{M}$, set $Q_K = 0$ if $p = 1$, otherwise let the operator $Q_K : L^2(K) \to \mathbb{P}_{p-2}(K)$ be given by (3.3.14). We define $B_{\mathcal{M},p} : H^1(\mathcal{M}) \to H_0^1(\Omega)$

$$B_{\mathcal{M},p} \sigma := \sum_{K \in \mathcal{M}} (Q_K v) \Phi_K,$$

where $\Phi_K := \prod_{z \in \mathcal{L}_1(K)} \Phi_z^1 \in S_{d+1}^1$ is the element bubble function with support $K$.

A suitable combination of $B_{\mathcal{F},p}$ and $B_{\mathcal{M},p}$ provides the desired property and an extension of the operator in (3.3.15).

**Lemma 4.2.4** (Bubble smoother)**.** *For all $p \in \mathbb{N}$, the smoothing operator $B_p : S_p^0 \to H_0^1(\Omega)$ defined by*

$$B_p \sigma := B_{\mathcal{F},p} \sigma + B_{\mathcal{M},p}(\sigma - B_{\mathcal{F},p} \sigma)$$

*satisfies* (4.2.5) *and the local stability estimate*

$$\|\nabla B_p \sigma\|_{L^2(K)} \leq \frac{C_{d,p}}{\rho_K} \left( \sup_{r \neq 0} \frac{\int_K \sigma r}{\|r\|_{L^2(K)}} + \sum_{F \in \mathcal{F}_K} \frac{|K|^{\frac{1}{2}}}{|F|^{\frac{1}{2}}} \sup_{q \neq 0} \frac{\int_F \{\!\!\{\sigma\}\!\!\} q}{\|q\|_{L^2(F)}} \right)$$

*where $r$ and $q$ vary in $\mathbb{P}_{p-2}(K)$ and $\mathbb{P}_{p-1}(F)$, respectively.*

*Proof.* Proceed as in the proof of Lemma 3.3.8. □

The argument in Remark 3.3.4 confirms that $B_p$ is not uniformly stable under refinement for the current setting, as one may suspect in view of the factor $\rho_K^{-1}$ in the stability estimate. However, since our bound involves lower order norms, we have the possibility to stabilize. This can be done with the help of the simplified nodal averaging operator $A_p : S_p^0 \to S_p^1$ introduced in

(2.2.1) or the (standard) averaging $\widetilde{A_p} : S_p^0 \to S_p^1$ from (2.2.8). In analogy with Chapter 3, here we consider only the first option, cf. Remarks 2.2.4 and 3.3.6.

Stabilizing the bubble smoother $B_p$ with the simplified nodal averaging $A_p$, we obtain a smoothing operator with the desired properties, which extends the one introduced in (3.3.19)

**Proposition 4.2.5** (Stable smoothing with moment conservation)**.** *The smoothing operator $E_p : S_p^0 \to H_0^1(\Omega)$ given by*

$$E_p\sigma := A_p\sigma + B_p(\sigma - A_p\sigma)$$

*is invariant on $S_p^1$, satisfies* (4.2.5) *and, for all $\sigma \in S_p^0$,*

$$\| \nabla_{\mathcal{M}}(\sigma - E_p\sigma)\|_{L^2(\Omega)} \le C_{d,\gamma_{\mathcal{M}},p}\|h^{-\frac{1}{2}} [\![\sigma]\!] \|_{L^2(\Sigma)}.$$

*Proof.* We adapt the proof of Propositions 3.3.9 to the current setting with jumps in the extended energy norm. Clearly, the operator $E_p$ is well-defined and maps into $H_0^1(\Omega)$. With $A_p$, also $E_p$ is a projection onto $S_p^1$. Arguing as in (3.3.8), it is also immediate to check that $E_p$ conserves the moments in (4.2.5), with the help of Lemma 4.2.4.

Finally, we turn to the claimed stability bound. Let $\sigma \in S_p^0$ and write

$$\| \nabla_{\mathcal{M}}(\sigma - E_p\sigma)\|_{L^2(\Omega)} \le \| \nabla_{\mathcal{M}}(\sigma - A_p\sigma)\|_{L^2(\Omega)} + \|\nabla B_p(\sigma - A_p\sigma)\|_{L^2(\Omega)}.$$

In order to bound the right-hand side, we fix a mesh element $K \in \mathcal{M}$ and consider the first term. Employing $\Phi^p_{z|K} = \Psi^p_{K,z}$, the scaling (2.1.6) and then Lemma 2.2.1, we obtain

$$\begin{aligned} \|\nabla(\sigma - A_p\sigma)\|_{L^2(K)} &\le \sum_{z\in\mathcal{L}_p(K)} \left|\sigma_{|K}(z) - A_p\sigma(z)\right| \|\nabla\Phi_z^p\|_{L^2(K)} \\ &\le C_{d,\gamma_{\mathcal{M}},p} \sum_{z\in\mathcal{L}_p(K)} \left|\sigma_{|K}(z) - A_p\sigma(z)\right| \frac{|K|^{\frac{1}{2}}}{\rho_K} \qquad (4.2.7) \\ &\le C_{d,\gamma_{\mathcal{M}},p} \sum_{z\in\mathcal{L}_p(K)} \sum_{F'\in\mathcal{F}, F'\ni z} \frac{|K|^{\frac{1}{2}}}{\rho_K\,|F'|^{\frac{1}{2}}} \| [\![\sigma]\!] \|_{L^2(F')} \end{aligned}$$

If $K' \in \mathcal{M}$ contains a face $F'$ of the sum, then (2.1.5) implies

$$\frac{|K|^{\frac{1}{2}}}{\rho_K\,|F'|^{\frac{1}{2}}} \le \frac{h_K}{\rho_K}\left(\frac{h_K^{d-2}}{\rho_{K'}^{d-1}}\right)^{\frac{1}{2}} \lesssim \rho_{K'}^{-\frac{1}{2}} \lesssim h_{F'}^{-\frac{1}{2}}.$$

Consequently, with the help of $\#\{K' \in \mathcal{M} \mid K' \subseteq \omega_K\} \le C_{d,\gamma_{\mathcal{M}}}$, we obtain

$$\|\nabla(\sigma - A_p\sigma)\|_{L^2(K)} \lesssim \left(\sum_{F\in\mathcal{F}, F\cap K\neq\emptyset} h_F^{-1}\| [\![\sigma]\!] \|^2_{L^2(F)}\right)^{\frac{1}{2}}. \qquad (4.2.8)$$

Next, consider the second term and observe that (2.1.1) gives

$$\sup_{r\in\mathbb{P}_{p-2}(K)} \frac{\int_K (\sigma - A_p\sigma) r}{\|r\|_{L^2(K)}} \leq C_{d,p}\,|K|^{\frac{1}{2}} \sum_{z\in\mathcal{L}_p(\partial K)} \left|\sigma_{|K}(z) - A_p\sigma(z)\right|$$

and, for every $F \in \mathcal{F}_K$,

$$\sup_{q\in\mathbb{P}_{p-1}(F)} \frac{\int_F (\{\!\{\sigma\}\!\} - A_p\sigma) q}{\|q\|_{L^2(F)}} \leq C_{d,p}\,|F|^{\frac{1}{2}} \sum_{K'\supset F} \sum_{z\in\mathcal{L}_p(F)} \left|\sigma_{|K'}(z) - A_p\sigma(z)\right|.$$

Inserting these two bounds in the stability estimate of Lemma 4.2.4, we find essentially the bound after the second inequality in (4.2.7) and, proceeding as before, it follows

$$\|\nabla B_p(\sigma - A_p\sigma)\|_{L^2(K)} \lesssim \left( \sum_{F\in\mathcal{F}, F\cap K\neq\emptyset} h_F^{-1} \| \, [\![\sigma]\!] \, \|^2_{L^2(F)} \right)^{\frac{1}{2}}. \tag{4.2.9}$$

We arrive at the claimed inequality by summing (4.2.8) and (4.2.9) over all $K \in \mathcal{M}$, observing that the number of elements touching a given face is $\leq C_{d,\gamma_\mathcal{M}}$. □

The smoothing operator $E_p$ in Proposition 4.2.5 is computationally feasible in the sense of Remark 1.4.13. In fact, it enjoys all the properties of its counterpart in (3.3.19) listed before Theorem 3.3.10.

After having found a suitable smoothing operator, we now choose the bilinear form $d(\cdot,\cdot)$. For this purpose, recall that, due to (4.2.4), the form $(\cdot, E_p\cdot)_{1;\eta}$ is degenerate and so $d(\cdot,\cdot)$ needs to be nontrivial. Several choices are possible; see, e.g., Arnold et al. [5]. Here we shall discuss the interplay between $E_p$ and some of them.

### A quasi-optimal NIP method

One possibility to achieve nondegeneracy is to employ the jump penalization in $(\cdot,\cdot)_{1;\eta}$. Owing to Lemma 4.2.3, we may also neutralize the possible downgrading of coercivity in $(\cdot, E_p\cdot)_{1;\eta}$, due to the term $-\int_\Sigma \{\!\{\nabla s\}\!\} \cdot n \, [\![q]\!]$. This suggests to define $b_{\mathrm{nip}} : S_p^0 \times S_p^0 \to \mathbb{R}$

$$b_{\mathrm{nip}}(s,\sigma) := (s, E_p\sigma)_{1;\eta} + \int_\Sigma \left( [\![s]\!] \, \{\!\{\nabla\sigma\}\!\} \cdot n + \frac{\eta}{h_F} [\![s]\!] \, [\![\sigma]\!] \right). \tag{4.2.10}$$

which just reestablish the bilinear form of the nonsymmetric interior penalty (NIP) method introduced in [53].

The next lemma recalls well-known properties of $b_{\mathrm{nip}}$ that are instrumental to our analysis, in connection with Remark 4.2.2. We also provide a proof for the sake of completeness.

**Lemma 4.2.6** ($b_{\text{nip}}$ and extended energy norm)**.** *For any penalty parameter $\eta > 0$, we have*

$$\forall s, \sigma \in S \quad b_{\text{nip}}(s,s) \geq |s|_{1;\eta}^2 \quad \textit{and} \quad b_{\text{nip}}(s,\sigma) \leq \left(1 + \sqrt{\eta^{-1}\eta_*}\right) |s|_{1;\eta} \, |\sigma|_{1;\eta} \, ,$$

*where $\eta_* > 0$ depends on $d$, $p$, and $\gamma_{\mathcal{M}}$.*

*Proof.* The coercivity bound holds by construction. For the continuity bound, we observe that, if $F \in \mathcal{F}_K$ is a face of any $K \in \mathcal{M}$, we have the inverse estimate $\| \cdot \|_{L^2(F)} \leq C_{d,\gamma_{\mathcal{M}},p} h_F^{-\frac{1}{2}} \| \cdot \|_{L^2(K)}$ in $\mathbb{P}_{p-1}(K)$ and set $\eta_* := (d+1) C^2_{d,\gamma_{\mathcal{M}},p}$. Then

$$\|h^{\frac{1}{2}} \{\!\{\nabla \sigma\}\!\} \|^2_{L^2(\Sigma)} \leq \eta_* \| \nabla_{\mathcal{M}} \sigma \|^2_{L^2(\Omega)} \tag{4.2.11}$$

and the claimed continuity bound follows by standard steps. □

According to this lemma, if the penalty parameter $\eta$ is not too small, we may consider $|\cdot|_{1;\eta}$ with the same $\eta$ to be the discrete energy norm associated with $b_{\text{nip}}$. Remarkably, as $\eta \to \infty$, the coercivity and continuity constants tend to their respective counterparts of the limiting conforming Galerkin method in $S^1_p$.

We thus arrive at $M_{\text{nip}} = (S^0_p, b_{\text{nip}}, E^\star_p)$, a *new variant of the NIP method of order $p$* with the discrete problem

$$U \in S^0_p \quad \text{such that} \quad \forall \sigma \in S^0_p \;\; b_{\text{nip}}(U,\sigma) = \langle f, E_p \sigma \rangle. \tag{4.2.12}$$

Since $b_{\text{nip}} = (\cdot,\cdot)_1$ and $E = \text{Id}$ on $S^1_p$, this is a nonconforming Galerkin method. In contrast to the original NIP method, it applies to any load $f \in H^{-1}(\Omega)$ and has the following property.

**Theorem 4.2.7** (Quasi-optimality of $M_{\text{nip}}$)**.** *For any $\eta > 0$, the method $M_{\text{nip}}$ is $|\cdot|_{1;\eta}$-quasi-optimal for the Poisson problem* (3.3.1) *with constant $\leq \sqrt{1 + C_{d,\gamma_{\mathcal{M}},p}\eta^{-1}}$.*

*Proof.* The quasi-optimality of $M_{\text{nip}}$ follows from Lemma 4.2.3, Proposition 4.2.5 and item (ii) in Theorem 4.2.1. The subsequent item (iii) entails also

$$C_{\text{qopt}} = \sqrt{1 + \left(\delta_{H^1_0(\Omega)}\right)^2}$$

so that it only remains to bound the consistency measure in the right-hand side. To this end, let $\Pi_{\eta,p}$ denote the $(\cdot,\cdot)_{1;\eta}$-orthogonal projection of $H^1(\mathcal{M})$ onto the space $S^0_p$. Following the notation of Theorem 4.2.1, we introduce the bilinear form $d_{\text{nip}} : H^1(\mathcal{M}) \times S^0_p \to \mathbb{R}$

$$d_{\text{nip}}(\widetilde{v}, \sigma) := b_{\text{nip}}(\Pi_{\eta,p}\widetilde{v}, \sigma) - (\widetilde{v}, E_p \sigma)_{1,\eta}.$$

For all $v \in H^1_0(\Omega)$ and $\sigma \in S^0_p$, Lemma 4.2.3, Proposition 4.2.5, the identities $[\![v]\!] = 0 = [\![E_p\sigma]\!]$ and the $(\cdot,\cdot)_{1,\eta}$-orthogonality of $\Pi_{\eta,p}$ imply

$$\begin{aligned} d_{\mathrm{nip}}(v,\sigma) &= (\Pi_{\eta,p}v - v, E_p\sigma)_1 + \int_\Sigma [\![\Pi_{\eta,p}v]\!] \{\!\{\nabla\sigma\}\!\} \cdot n + \int_\Sigma \frac{\eta}{h} [\![\Pi_{\eta,p}v]\!] [\![\sigma]\!] \\ &= (\Pi_{\eta,p}v - v, E_p\sigma - \sigma)_1 + \int_\Sigma [\![\Pi_{\eta,p}v - v]\!] \{\!\{\nabla\sigma\}\!\} \cdot n. \end{aligned}$$

Combining this identity with the upper bound in Proposition 4.2.5 and a standard argument, it follows

$$|d_{\mathrm{nip}}(v,\sigma)| \leq C_{d,\gamma_{\mathcal{M}},p}\eta^{-\frac{1}{2}} |\sigma|_{1;\eta} |\Pi_{\eta,p}v - v|_{1;\eta}. \tag{4.2.13}$$

We thus derive $\delta_{H^1_0(\Omega)} \leq C_{d,\gamma_{\mathcal{M}},p}\eta^{-\frac{1}{2}}$ with the help of Remark 4.2.2 and the coercivity bound in Lemma 4.2.6. □

The following remarks complement Theorem 4.2.7. The first one underlines the importance of Theorem 4.2.1 to obtain a sharp bound of the quasi-optimality constant of $M_{\mathrm{nip}}$. The second one improves on the observation that we recover Cea's lemma in the limit $\eta \to +\infty$.

*Remark* 4.2.8 (Overestimation of $C_{\mathrm{qopt}}$ by Theorem 3.2.1). Denote by $\delta_{S^0_p}$ the consistency measure from Proposition 1.3.9, associated with the method $M_{\mathrm{nip}}$. The orthogonality (4.2.4) readily yields $\delta_{S^0_p} \geq 1$. Moreover, for all $\sigma \in S^1_p$, item (i) of Theorem 4.2.1, Lemma 4.2.6 and $E_p\sigma = \sigma$ reveal

$$C_{\mathrm{stab}} \geq \frac{|\sigma|_{1,\eta}}{\sup_{s\in S^0_p, |s|_{1,\eta}=1} b_{\mathrm{nip}}(s,\sigma)} \geq \left(1 + \sqrt{\eta^{-1}\eta_*}\right)^{-1}.$$

Thus, the upper bound in item (iii) of Theorem 3.2.1 (slightly) overestimates the quasi-optimality constant of $M_{\mathrm{nip}}$ in the limit $\eta \to +\infty$.

*Remark* 4.2.9 (Pointwise convergence of $M_{\mathrm{nip}}$ for $\eta \to +\infty$). Let $u \in H^1_0(\Omega)$ be the weak solution of the Poisson problem with load $f \in H^{-1}(\Omega)$ and denote by $U_\eta \in S^0_p$ the quasi-optimal approximation of $u$ by the method $M_{\mathrm{nip}}$ with parameter $\eta > 0$, i.e. $U_\eta = M_{\mathrm{nip}}f$. Theorem 4.2.7 and $C_{\mathrm{stab}} \leq C_{\mathrm{qopt}}$ entail

$$|U_\eta|_{1,\bar\eta} \leq |U_\eta|_{1,\eta} \leq C_{d,p,\gamma_{\mathcal{M}}} \|f\|_{H^{-1}(\Omega)} \tag{4.2.14}$$

for all $\eta \geq \bar\eta > 0$. Since $S^0_p$ is finite-dimensional, we have

$$U_\eta \to U_\infty \in S^0_p \quad \text{as} \quad \eta \to \infty$$

up to a subsequence, where the convergence is intended in the $\eta$-independent norm $|\cdot|_{1,\bar\eta}$. The second inequality in (4.2.14) reveals

$$\left(\sum_{F\in\mathcal{F}} \frac{\eta}{h_F} \| [\![U_\eta]\!] \|^2_{L^2(F)}\right)^{\frac{1}{2}} \leq C_{d,p,\gamma_{\mathcal{M}}} \|f\|_{H^{-1}(\Omega)}, \tag{4.2.15}$$

which yields $U_\infty \in S^1_p$ by continuity. Then, testing with $\sigma \in S^1_p$ in the discrete problem (4.2.12) for $U_\eta$ and taking the limit $\eta \to +\infty$, we infer

$$\forall \sigma \in S^1_p \qquad (U_\infty, \sigma)_1 = \langle f, \sigma \rangle$$

in view of (4.2.11), (4.2.15) and $E_p\sigma = \sigma$. Since this problem is uniquely solvable, we conclude that $U_\infty = M_{\mathrm{cG}} f$, where $M_{\mathrm{cG}} := (S^1_p, (\cdot,\cdot)_1, \mathrm{Id}^\star_{S^1_p})$ is the conforming Galerkin method in $S^1_p$.

### A quasi-optimal SIP method

The NIP bilinear form $b_{\mathrm{nip}}$ arises in particular by enforcing coercivity. As an alternative, one can achieve symmetry by changing the sign of the first term in $d_{\mathrm{nip}}$. This leads to the SIP bilinear form $b_{\mathrm{sip}}$; cf. (4.1.1). While $b_{\mathrm{sip}}$ verifies the same continuity bound as $b_{\mathrm{nip}}$, the coercivity bound can be replaced as follows. Inequality (4.2.11) implies

$$\left| \int_\Sigma [\![s]\!] \, \{\!\!\{\nabla \sigma\}\!\!\} \cdot n \right| \le \frac{1}{2}\sqrt{\eta_* \eta^{-1}} \left( \eta \| h^{-\frac{1}{2}} [\![s]\!] \|^2_{L^2(\Sigma)} + \| \nabla_{\mathcal{M}} \sigma \|^2_{L^2(\Omega)} \right),$$

from which we get

$$\text{(4.2.16)} \qquad \forall s \in S^0_p \quad b_{\mathrm{sip}}(s,s) \ge \alpha(\eta_* \eta^{-1}) \, |s|^2_{1;\eta} \quad \text{with} \quad \alpha(t) = 1 - \sqrt{t}.$$

Hence, if $\eta > \eta_*$, then the discrete problem

$$\text{(4.2.17)} \qquad U \in S^0_p \quad \text{such that} \quad \forall \sigma \in S^0_p \;\; b_{\mathrm{sip}}(U, \sigma) = \langle f, E_p \sigma \rangle$$

is well-posed and gives rise to a *new variant of the SIP method*, which is a nonconforming Galerkin method and denoted by $M_{\mathrm{sip}}$. The following theorem covers the results illustrated in the introduction §4.1 and is proven like Theorem 4.2.7.

**Theorem 4.2.10** (Quasi-optimality of $M_{\mathrm{sip}}$). *For any $\eta > \eta_*$, the method $M_{\mathrm{sip}}$ is $|\cdot|_{1,\eta}$-quasi-optimal for the Poisson problem* (3.3.1) *with constant* $\le \sqrt{1 + C_{d,\gamma_{\mathcal{M}},p} \big(\alpha(\eta_*/\eta)\eta\big)^{-1}}$.

For $\eta \to \infty$, we again end up in Céa's lemma for the limiting conforming Galerkin method in $S^1_p$. Moreover, the observations in Remarks 4.2.8 and 4.2.9 applies also to $M_{\mathrm{sip}}$.

### High-order smoothing with first-order averaging

Assume that $p \ge 2$. As pointed out in §2.2, the simplified averaging operator $A_1$ is defined also on $S^0_p$ and so we may consider

$$\text{(4.2.18)} \qquad \widetilde{E}_p \sigma := A_1 \sigma + B_p(\sigma - A_1 \sigma), \quad \sigma \in S^0_p,$$

which is cheaper to evaluate than $E_p$, cf. Remark 3.3.11. Using Lemma 2.2.3 in the proof of Proposition 4.2.5, we obtain the following properties of $\widetilde{E}_p$.

**Proposition 4.2.11** (Moment conservation with first-order averaging)**.** *The linear operator* $\widetilde{E}_p$ *from* (4.2.18) *is invariant on* $S_1^1$*, satisfies* (4.2.5) *and, for all* $\sigma \in S_p^0$*,*

$$\| \nabla_{\mathcal{M}}(\sigma - \widetilde{E}_p\sigma)\|_{L^2(\Omega)} \leq C_{d,\gamma_{\mathcal{M}},p} \left( \sum_{F\in\mathcal{F}} h_F^{-d} \left| \int_F [\![\sigma]\!] \right|^2 + \| \nabla_{\mathcal{M}} \sigma\|^2_{L^2(\Omega)} \right)^{\frac{1}{2}}.$$

Notice that a bound solely in terms of jumps is not possible, because $\widetilde{E}_p$ is not invariant on $S_p^1$.

Combining the new smoothing operator $\widetilde{E}_p$ with one of the previous bilinear forms $b_{\mathrm{var}}$, $\mathrm{var} \in \{\mathrm{nip}, \mathrm{sip}\}$, leads to a nonconforming method $\widetilde{M}_{\mathrm{var}}$ with discrete problem

$$U \in S_p^0 \quad \text{such that} \quad \forall \sigma \in S_p^0 \;\; b_{\mathrm{var}}(U,\sigma) = \langle f, \widetilde{E}_p\sigma\rangle, \tag{4.2.19}$$

which is well-posed for all $\eta > \eta_{\mathrm{var}}$. Hereafter

$$\eta_{\mathrm{var}} := \begin{cases} 0, & \text{if var = nip,} \\ \eta_*, & \text{if var = sip,} \end{cases} \quad \text{and} \quad \alpha_{\mathrm{var}}(t) := \begin{cases} 1, & \text{if var = nip,} \\ 1-\sqrt{t}, & \text{if var = sip.} \end{cases}$$

As $\widetilde{E}_p$ is only invariant on the strict subset $S_1^1$ of $H_0^1(\Omega) \cap S_p^0$for $p \geq 2$, the method $\widetilde{M}_{\mathrm{var}}$ is *not* a nonconforming Galerkin method. Nevertheless:

**Theorem 4.2.12** (Quasi-optimality of $\widetilde{M}_{\mathrm{var}}$)**.** *Let* $\mathrm{var} \in \{\mathrm{nip}, \mathrm{sip}\}$ *and assume* $\eta > \eta_{\mathrm{var}}$*. Then, the method* $\widetilde{M}_{\mathrm{var}}$ *is* $|\cdot|_{1;\eta}$*-quasi-optimal for the Poisson problem* (3.3.1) *with constant* $\leq C_{d,\gamma_{\mathcal{M}},p}\sqrt{1 + \left(\alpha_{\mathrm{var}}(\eta_*/\eta)\eta\right)^{-1}}$.

*Proof.* Proceed as in the proof of Theorem 4.2.7 or as indicated for Theorem 4.2.10, replacing $E_p$ by $\widetilde{E}_p$. The only difference is that, in the derivation of the counterpart of (4.2.13), we use

$$\sum_{F\in\mathcal{F}} h_F^{-d} \left| \int_F [\![\sigma]\!] \right|^2 \lesssim \sum_{F\in\mathcal{F}} h_F^{-1} \int_F |[\![\sigma]\!]|^2$$

and obtain only

$$|b_{\mathrm{var}}(\Pi_{\eta,p}v,\sigma) - (v, \widetilde{E}_p\sigma)_1| \leq$$
$$\leq C_{d,\gamma_{\mathcal{M}},p}\sqrt{1 + \left(\alpha_{\mathrm{var}}(\eta_*/\eta)\eta\right)^{-1}}\, |\sigma|_{1;\eta} \, |\Pi_{\eta,p}v - v|_{1;\eta}$$

because the stability bound in Proposition 4.2.11 involves gradient terms. □

Here, due to the use of a smoothing operator which is not invariant on the conforming part of $S_p^0$, we do not recover Cea's lemma in the limit $\eta \to +\infty$. Nevertheless, we have the following counterpart of Remark 4.2.9.

*Remark* 4.2.13 (Pointwise convergence of $\widetilde{M}_{\mathrm{var}}$ for $\eta \to +\infty$). Let $u$ denote the weak solution of of the Poisson problem with load $f \in H^{-1}(\Omega)$ and set $\widetilde{U}_\eta := \widetilde{M}_{\mathrm{var}} f$. Arguing as in Remark 4.2.9, we derive that $\widetilde{U}_\eta$ converges to the solution $\widetilde{U}_\infty \in S^1_p$ of

$$\forall \sigma \in S^1_p \qquad (\widetilde{U}_\infty, \sigma)_1 = \langle f, \widetilde{E}_p \sigma \rangle$$

in the norm $|\cdot|_{1,\bar{\eta}}$, $\bar{\eta} > 0$, as $\eta \to +\infty$. Notice that this is the discrete problem of the quasi-optimal conforming method $\widetilde{M} := (S^1_p, (\cdot,\cdot)_1, \widetilde{E}^\star_p)$. Since the smoother $\widetilde{E}_p$ does not reduce to the identity on $S^1_p$, $\widetilde{M}$ differs from the conforming Galerkin method in $S^1_p$ and its quasi-optimality constant is $> 1$. This observation agrees with the fact that we do not recover Cea's lemma in the limit $\eta \to +\infty$ for the method $\widetilde{M}_{\mathrm{var}}$.

### Quasi-optimal methods with weak interior penalty

The idea of Weak Interior Penalty (WIP) methods and variants, see e.g. [18, 19, 20], is to penalize some projection of the jump $[\![v]\!]$ instead of the jump itself. For instance, if $S = S^0_p$, one could replace the term $\int_\Sigma \eta h^{-1} [\![\widetilde{v}_1]\!] \, [\![\widetilde{v}_1]\!]$ with $\int_\Sigma \eta h^{-1} \Pi_{p-1} [\![\widetilde{v}_1]\!] \, \Pi_{p-1} [\![\widetilde{v}_1]\!]$ both in the extended scalar product and the discrete bilinear form, where $\Pi_{p-1}$ denotes the $L^2$-orthogonal projection onto discontinuous piecewise polynomials of degree $p-1$ on $\Sigma$. Not surprisingly, the procedure illustrated in the previous examples applies to this case as well. In particular, the lowest-order case $p = 1$ is quite instructive, although rather specific. In fact, it is possible to establish counterparts of Theorems 4.2.7-4.2.10 where the quasi-optimality constant does not depend on the employed penalty parameter, cf. Remark 4.2.16.

To illustrate this observation, let $\mu > 0$ be a penalty parameter and assume $d = 2$ for simplicity. We introduce the scalar product

$$(v, w)_{1,\mu} = \int_\Omega \nabla_{\mathcal{M}} v \cdot \nabla_{\mathcal{M}} w + \mu \sum_{F \in \mathcal{F}} \left( \fint_F [\![v]\!] \right) \left( \fint_F [\![w]\!] \right),$$

on $H^1(\mathcal{M})$, denoting by $|\cdot|_{1,\mu}$ the induced norm. As before, we abbreviate $(\cdot,\cdot)_{1,0}$ by $(\cdot,\cdot)_1$. We consider the following variant of the setting (4.2.3)

$$V = H^1_0(\Omega), \qquad S = S^0_1, \qquad \widetilde{a} = (\cdot,\cdot)_{1;\mu} \text{ on } \widetilde{V} = H^1_0(\Omega) + S^0_1. \tag{4.2.20}$$

As before, we have

$$\emptyset \neq S^0_0 \subseteq S^0_1 \cap V^\perp,$$

which precludes overconsistency in light of Lemma 3.2.4. However, the smoothing operator $E_1 : S^0_1 \to H^1_0(\Omega)$ from Proposition 4.2.5 readily satisfies the following counterpart of (4.2.6) in Lemma 4.2.3

$$(s, E_1\sigma)_{1;\mu} = \int_\Omega \nabla_{\mathcal{M}} s \cdot \nabla_{\mathcal{M}} \sigma - \sum_{F \in \mathcal{F}} \int_F \{\!\{\nabla s\}\!\} \cdot n \, [\![\sigma]\!] \tag{4.2.21}$$

for all $s, \sigma \in S_1^0$, irrespective of $\mu$. This suggests that we may mimic the definitions of the SIP and NIP forms, introducing

$$b_{\text{wsip}}(s,\sigma) := (s, E_1\sigma)_{1;\mu} - \int_\Sigma [\![s]\!] \{\!\{\nabla\sigma\}\!\} \cdot n + \mu \sum_{F\in\mathcal{F}} \left( \fint_F [\![v]\!] \right) \left( \fint_F [\![w]\!] \right)$$

$$b_{\text{wnip}}(s,\sigma) := (s, E_1\sigma)_{1;\mu} + \int_\Sigma [\![s]\!] \{\!\{\nabla\sigma\}\!\} \cdot n + \mu \sum_{F\in\mathcal{F}} \left( \fint_F [\![v]\!] \right) \left( \fint_F [\![w]\!] \right)$$

on $S_1^0$. These forms are closely related to those ones in [18, 20].

The nondegeneracy of $b_{\text{wnip}}$ is clear, in view of

$$\forall s \in S_1^0 \qquad b_{\text{wnip}}(s,s) = |s|_{1,\mu}^2 .$$

To check that also $b_{\text{wsip}}$ is $|\cdot|_{1,\mu}$-coercive, we observe that (4.2.11) entails

$$\forall \sigma \in S_1^0 \qquad \| \{\!\{\nabla\sigma\}\!\} \|_{L^1(\Sigma)}^2 \leq \| h^{\frac{1}{2}} \{\!\{\nabla\sigma\}\!\} \|_{L^2(\Sigma)}^2 \leq \mu_* \| \nabla_{\mathcal{M}} \sigma \|_{L^2(\Omega)}^2,$$

for some constant $\mu_* > 0$ depending only on $d$ and $\gamma_{\mathcal{M}}$. This yields

$$\left| \int_\Sigma [\![s]\!] \{\!\{\nabla\sigma\}\!\} \cdot n \right| \leq \frac{1}{2}\sqrt{\mu_*\mu^{-1}} \left( \mu \sum_{F\in\mathcal{F}} \left( \fint_F [\![s]\!] \right)^2 + \| \nabla_{\mathcal{M}} \sigma \|_{L^2(\Omega)}^2 \right)$$

for all $s, \sigma \in S_1^0$, because $\nabla_{\mathcal{M}} \sigma$ is piecewise constant on $\mathcal{M}$. Consequently, if the penalty parameter satisfies $\mu > \mu_*$, the form $b_{\text{wsip}}$ is $|\cdot|_{1,\mu}$-coercive with constant $1 - \sqrt{\mu_*\mu^{-1}}$.

It is worth noticing that the current setting can be viewed as an extension of the one in §3.3.1 with $S = S_1^0$. In fact, $E_1$ reduces to the smoothing operator in Proposition 3.3.2 on the Crouzeix-Raviart space $CR$ and we have $b_{\text{wsip}}(s,\sigma) = b_{\text{wnip}}(s,\sigma) = (s,\sigma)_1$ for $s,\sigma \in CR$. Thus, introducing the nonconforming method $M_{\text{wip}} := (S_1^0, b_{\text{wip}}, E_1)$ for wip $\in \{\text{wsip}, \text{wnip}\}$, we expect that $M_{\text{wip}}$ is $|\cdot|_{1,\mu}$-quasi-optimal and its quasi-optimality constant is closely related to the one of $M_{CR}$. The following theorem confirms this claim.

**Theorem 4.2.14** (Quasi-optimality of $M_{\text{wip}}$). *For* wip $\in \{\text{wsip}, \text{wnip}\}$*, assume that* $\mu > \mu_*$ *if* wip = wsip *and* $\mu > 0$ *if* wip = wnip*. Then, the method* $M_{\text{wip}}$ *is* $|\cdot|_{1;\mu}$*-quasi-optimal for the Poisson problem* (3.3.1) *with quasi-optimality constant* $C_{\text{qopt}} = \|E_1\|_{\mathcal{L}(CR, H_0^1(\Omega))}$.

Thus, the quasi-optimality constant of $M_{\text{wip}}$ equals the one of $M_{CR}$.

*Proof.* The quasi-optimality of $M_{\text{wip}}$ is a consequence of item (ii) in Theorem 4.2.1 and the identity

$$\forall u \in S_1^0 \cap H_0^1(\Omega), \sigma \in S_1^0 \qquad b_{\text{wip}}(u,\sigma) = (u, E_1\sigma)_{1,\mu}$$

which easily follows from the definition of $b_{\mathrm{wip}}$.

Let us now turn to the claimed identity for the quasi-optimality constant. For notational convenience we denote by $\delta^{CR} := \delta^{CR}_{H_0^1(\Omega)}$ the consistency measure associated with $M_{CR}$ in item (iii) of Theorem 4.2.1. Since the $a_{\mathcal{M}}$-orthogonal projection $\Pi_{CR}$ of $H_0^1(\Omega)$ onto $CR$ is given by (3.3.4) and $E_1$ is a right inverse of $\Pi_{CR}$, we have

$$(s,\sigma)_1 = (s, E_1\sigma)_1 = (s, \Pi_{CR}E_1\sigma)_1, \tag{4.2.22}$$

for all $s, \sigma \in CR$. Given $u \in H_0^1(\Omega)$ and $\sigma \in CR$ and using the previous identity, we further deduce

$$\begin{aligned} d_{CR}(u,\sigma) &:= (u, E_1\sigma)_1 - (\Pi_{CR}u, \sigma)_1 \\ &= (u - \Pi_{CR}u, E_1\sigma)_1 \\ &= (u - \Pi_{CR}u, E_1\sigma - \Pi_{CR}E_1\sigma)_1. \end{aligned} \tag{4.2.23}$$

Hence, setting $u = E_1\sigma$, we infer

$$\delta^{CR} = \|E_1 - \Pi_{CR}E_1\|_{\mathcal{L}(CR, H_0^1(\Omega))} = \|E_1 - \mathrm{Id}_{CR}\|_{\mathcal{L}(CR, H_0^1(\Omega))}. \tag{4.2.24}$$

Next, we observe that $\Pi_{CR}$ is also the $(\cdot,\cdot)_{1,\mu}$-orthogonal projection of $H_0^1(\Omega)$ onto $S_1^0$, for all $\mu > 0$. To see this, we let $v \in H_0^1(\Omega)$ and $\sigma \in S_1^0$ and integrate by parts with the help of (2.1.3). We obtain

$$(v - \Pi_{CR}v, \sigma)_{1,\mu} = \int_\Sigma (v - \Pi_{CR}v) \, [\![\nabla\sigma]\!] \cdot n = 0$$

because $\nabla_{\mathcal{M}} \sigma$ is piecewise constant on $\mathcal{M}$.

After this preparation, we turn to the WIP methods and denote by $\delta^{\mathrm{wip}} := \delta^{\mathrm{wip}}_{H_0^1(\Omega)}$ the consistency measure from item (iii) of Theorem 4.2.1. Considering again generic $u \in H_0^1(\Omega)$ and $\sigma \in S_1^0$, we proceed as for (4.2.23) and obtain

$$d_{\mathrm{wip}}(u,\sigma) := (u, E_1\sigma)_{1,\mu} - b_{\mathrm{wip}}(\Pi_{CR}u, \sigma) = a_{\mathcal{M}}(u - \Pi_{CR}u, E_1\sigma - \Pi_{CR}E_1\sigma).$$

Choosing again $u = E_1\sigma$, this identity yields

$$\delta^{\mathrm{wip}} = \sup_{\sigma \in S_1^0} \frac{|E_1\sigma - \Pi_{CR}E_1\sigma|_{1,\mu}}{\sup\limits_{\hat{s} \in S_1^0, |\hat{s}|_{1,\mu}=1} b_{\mathrm{wip}}(\hat{s},\sigma)}.$$

We now aim at showing the identity $\delta^{\mathrm{wip}} = \delta^{CR}$. For this purpose, we extend $\Pi_{CR}$ to $S_1^0$ by requiring $\Pi_{CR}\sigma \in CR$ and $\int_F \Pi_{CR}\sigma = \int_F \{\!\{\sigma\}\!\}$ for all $\sigma \in S_1^0$ and $F \in \mathcal{F}^i$. Clearly, we have $\Pi_{CR}E_1\sigma = \Pi_{CR}\sigma$. We combine this commuting property with (4.2.21) and the $(\cdot,\cdot)_{1,\mu}$-orthogonality of $\Pi_{CR}$

$$\forall \hat{s} \in CR \qquad b_{\mathrm{wip}}(\hat{s},\sigma) = (\hat{s}, E_1\sigma)_{1,\mu} = (\hat{s}, \Pi_{CR}\sigma)_{1,\mu}. \tag{4.2.25}$$

Since the norm $|\cdot|_{1,\mu}$ reduces to $\| \nabla_{\mathcal{M}} \cdot\|_{L^2(\Omega)}$ on $CR$, we obtain

$$\sup_{\hat{s}\in CR, |\hat{s}|_{1,\mu}=1} b_{\text{wip}}(\hat{s},\sigma) = \| \nabla_{\mathcal{M}} \Pi_{CR}\sigma\|_{L^2(\Omega)}$$

which provides the upper bound

$$\delta^{\text{wip}} \le \sup_{\sigma\in S_1^0} \frac{\| \nabla_{\mathcal{M}}(E_1\sigma - \Pi_{CR}E_1\sigma)\|_{L^2(\Omega)}}{\| \nabla_{\mathcal{M}} \Pi_{CR}\sigma\|_{L^2(\Omega)}} = \delta^{CR}. \tag{4.2.26a}$$

in view of $E_1\sigma = E_1\Pi_{CR}\sigma$.

To reverse this bound, let $\sigma \in CR$ and assume first wip = wsip. Then, proceeding as in (4.2.25) leads to

$$\sup_{\hat{s}\in S_1^0, |\hat{s}|_{1,\mu}=1} b_{\text{wsip}}(\hat{s},\sigma) = |\Pi_{CR}\sigma|_{1,\mu} = \| \nabla_{\mathcal{M}} \sigma\|_{L^2(\Omega)} \tag{4.2.26b}$$

and, thus,

$$\delta^{CR} = \sup_{\sigma\in CR} \frac{\| \nabla_{\mathcal{M}}(E_1\sigma - \Pi_{CR}E_1\sigma)\|_{L^2(\Omega)}}{\sup\limits_{\hat{s}\in S_1^0, |\hat{s}|_{1,\mu}=1} b_{\text{wsip}}(\hat{s},\sigma)} \le \delta^{\text{wsip}}. \tag{4.2.26c}$$

Next, assume wip = wnip. In the light of

$$\forall \hat{s} \in S_1^0, \sigma \in CR \qquad b_{\text{wnip}}(\hat{s},\sigma) = 2\int_\Omega \nabla_{\mathcal{M}} \hat{s} \cdot \nabla_{\mathcal{M}} \sigma - b_{\text{wsip}}(\hat{s},\sigma),$$

and (4.2.25), we obtain

$$\sup_{\hat{s}\in S_1^0, |\hat{s}|_{1,\mu}=1} b_{\text{wnip}}(\hat{s},\sigma) = |2\sigma - \Pi_{CR}\sigma|_{1,\mu} = \| \nabla_{\mathcal{M}} \sigma\|_{L^2(\Omega)} \tag{4.2.26d}$$

and therefore

$$\delta^{CR} = \sup_{\sigma\in CR} \frac{\| \nabla_{\mathcal{M}}(E_1\sigma - \Pi_{CR}E_1\sigma)\|_{L^2(\Omega)}}{\sup\limits_{\hat{s}\in S_1^0, |\hat{s}|_{1,\mu}=1} b_{\text{wnip}}(\hat{s},\sigma)} \le \delta^{\text{wnip}}. \tag{4.2.26e}$$

Combining inequalities (4.2.26), it follows $\delta^{CR} = \delta^{\text{wsip}} = \delta^{\text{wnip}}$. Invoking Corollary 3.2.8, Proposition 3.3.2 and item (iii) in Theorem 4.2.1, we conclude that the quasi-optimality constant $C_{\text{qopt}}$ of $M_{\text{wip}}$ is given by

$$C_{\text{qopt}}^2 = 1 + (\delta^{\text{wip}})^2 = 1 + (\delta^{CR})^2 = \|E_1\|^2_{\mathcal{L}(CR,H_0^1(\Omega))}$$

for wip $\in \{\text{wsip}, \text{wnip}\}$. □

*Remark* 4.2.15 (Stability and quasi-optimality constants of $M_{\text{wip}}$). Combining item (i) of Theorems 4.2.1 and 4.2.14 with (4.2.26b) and (4.2.26d), we derive the following inequalities, involving the stability and quasi-optimality constants of $M_{\text{wip}}$

$$C_{\text{stab}} \geq \sup_{\sigma \in CR} \frac{\|E_1 \sigma\|_1}{\sup\limits_{\hat{s} \in S_1^0, |\hat{s}|_{1,\mu}=1} b_{\text{wip}}(\hat{s}, \sigma)} = \|E_1\|_{\mathcal{L}(CR, H_0^1(\Omega))} = C_{\text{qopt}}.$$

Hence, for the WIP methods, stability and quasi-optimality constants coincide without overconsistency and the upper bound in item (iii) of Theorem 3.2.1 is an overestimation.

*Remark* 4.2.16 ($C_{\text{qopt}}$ and discrete coercivity). The fact that $C_{\text{qopt}}$ does not depend on the penalty $\mu$ is somehow surprising, because the bilinear form $b_{\text{wip}}$ is degenerate, at least, for $\mu = 0$ and certain choices of $\mathcal{M}$, see [47, Section 3.3]. This indicates that the quasi-optimality constant of a nonconforming method can be of moderate size, even if the coercivity constant of the underlying bilinear form is not. Thus, in particular, the upper bounds in Theorems 4.2.7-4.2.10 possibly provide a pessimistic overestimation.

*Remark* 4.2.17 (Pointwise convergence of $M_{\text{wip}}$ for $\mu \to +\infty$). The technique illustrated in Remark 4.2.9 can be used to check that $M_{\text{wip}} f$ converges to $M_{CR} f$ in the norm $|\cdot|_{1,\bar{\mu}}$, $\bar{\mu} > 0$, for $\mu \to +\infty$ and for all $f \in H^{-1}(\Omega)$.

### 4.2.2 A Quasi-Optimal and Locking-Free Crouzeix-Raviart Method for Linear Elasticity

The goal of this section is to conceive a quasi-optimal and locking-free method for linear elasticity.

Given $\Omega \subseteq \mathbb{R}^d$ as in §2.1, we consider the displacement formulation of the linear elasticity problem with pure displacement boundary conditions: find $u \in H_0^1(\Omega)^d$ such that

$$-\operatorname{div}\big(2\mu\, \varepsilon(u) + \lambda \operatorname{div}(u)\big) = f \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega. \tag{4.2.27}$$

Hereafter $\varepsilon(v) := (\nabla v + \nabla v^T)/2$ is the symmetric gradient and $\mu$, $\lambda > 0$ are the Lamé coefficients. We shall mostly suppress the dependencies on $\mu$ in the notation, while we trace the ones on $\lambda$.

Let $\mathcal{M}$ be a mesh of $\Omega$ as in §2.1 and, for $\eta \geq 0$, define

$$a_{\lambda;\eta}(v, w) := \int_\Omega (2\mu\, \varepsilon_{\mathcal{M}}(v) : \varepsilon_{\mathcal{M}}(w) + \lambda \operatorname{div}_{\mathcal{M}} v \operatorname{div}_{\mathcal{M}} w) + \int_\Sigma \frac{\mu\eta}{h} [\![v]\!]\, [\![w]\!]\,,$$

$$\|v\|_{\lambda;\eta} = \sqrt{a_{\lambda;\eta}(v, v)}$$

for $v, w \in H^1(\mathcal{M})^d$ and abbreviate $a_{\lambda;0}$ to $a_\lambda$. We aim at applying Theorem 4.2.1 with the following setting:

$$V = H_0^1(\Omega)^d, \quad S \subseteq S_1^0, \quad \widetilde{a} = a_{\lambda;\eta} \text{ on } \widetilde{V} = H_0^1(\Omega) + S, \tag{4.2.28}$$

where $S$ will be specified below, $\eta > 0$, and the colon indicates the matrix scalar product $G : H = \sum_{j,\ell=1}^{d} G_{j\ell} H_{j\ell}$. Notice that $a_{\lambda;\eta}$ is then a scalar product and the abstract variational problem (1.2.1) provides a weak formulation of (4.2.27).

We readily deduce the following counterpart of Lemma 4.2.3.

**Lemma 4.2.18** (Moment conservation). *If we are given a linear smoothing operator* $E : (S_1^0)^d \to H_0^1(\Omega)^d$ *which satisfies*

$$\forall \sigma \in (S_1^0)^d, F \in \mathcal{F}^i \qquad \int_F E\sigma = \int_F \{\!\{\sigma\}\!\}, \tag{4.2.29}$$

*then, for all* $s, \sigma \in (S_1^0)^d$,

$$a_{\lambda;\eta}(s, E\sigma) = a_\lambda(s,\sigma) - \int_\Sigma \big( \{\!\{2\mu\, \varepsilon_{\mathcal{M}}(s) + \lambda \operatorname{div}_{\mathcal{M}}(s) I\}\!\} \big)\, n \cdot [\![\sigma]\!] .$$

In the previous section, the impact on coercivity or symmetry of the counterpart of the term $\int_\Sigma (\{\!\{2\mu\, \varepsilon_{\mathcal{M}}(s) + \lambda \operatorname{div}_{\mathcal{M}}(s) I\}\!\}) n \cdot [\![\sigma]\!]$ was compensated by adding suitable terms to the discrete bilinear form. Here we shall handle it with the choice of the discrete space $S$. More precisely, we choose $S = CR^d$, i.e. the Crouzeix-Raviart space from (3.3.3) on each component. Then this term vanishes, because the average $(\{\!\{2\mu\, \varepsilon_{\mathcal{M}}(s) + \lambda \operatorname{div}_{\mathcal{M}}(s) I\}\!\})n$ is a constant on each face $F \in \mathcal{F}$. Furthermore, $\int_F \sigma$, $F \in \mathcal{F}$, is well-defined and equals the right-hand side of (4.2.29).

An important difference between the current setting and the one proposed in §3.3.1 for the Poisson problem is the following. Arnold [2] shows that, for certain choices of $\Omega$ and $\mathcal{M}$, there is a nonzero function

$$s_0 \in CR^2 \setminus \{0\} \quad \text{with} \quad \varepsilon_{\mathcal{M}}(s_0) = 0 \text{ and } \operatorname{div}_{\mathcal{M}} s_0 = 0, \tag{4.2.30}$$

entailing that

$$0 \neq s_0 \in CR^2 \cap (H_0^1(\Omega)^2)^\perp.$$

This observation generally rules out the possibility of designing overconsistent methods with $S = CR^d$, in view of Lemma 3.2.4.

As (4.2.29) reduces on $CR^d$ to the vector version of the condition devised in Lemma 3.3.1, we can take the computionally feasible smoothing operator $E_1$ from Proposition 3.3.2 componentwise. We denote this vector version again by $E_1$. Let us recall that $E_1$ can alternatively be seen as the restriction to the Crouzeix-Raviart space of the smoothing operator introduced in Proposition 4.2.5 for $p = 1$. This allows to apply here useful results from both §3.3.1 and §4.2.1.

Since $a_{\lambda;\eta}(\cdot, E_1\cdot)$ may be degenerate in view of (4.2.30), we enforce positive definiteness by the same jump penalization as in the definition of $a_{\lambda;\eta}$. We obtain $b_{\mathrm{HL}} : CR^d \times CR^d \to \mathbb{R}$

$$b_{\mathrm{HL}}(s,\sigma) := a_{\eta,\lambda}(s, \sigma) + \int_\Sigma \frac{\mu\eta}{h} [\![s]\!]\, [\![\sigma]\!], \quad \eta > 0 \tag{4.2.31}$$

which coincides with the discrete bilinear form proposed by Hansbo and Larson [44, eq. (26)]. We thus introduce a *new penalized Crouzeix-Raviart method* $M_{\mathrm{HL}} = (CR^d, b_{\mathrm{HL}}, E_1^\star)$ given by the following discrete problem: find $U \in CR^d$ such that

$$\forall \sigma \in CR^d \qquad b_{\mathrm{HL}}(U,\sigma) = \langle f, E_1\sigma\rangle. \tag{4.2.32}$$

The method $M_{\mathrm{HL}}$ is a nonconforming Galerkin method. The modification of the right-hand side with respect to [44] allows to apply $H^{-1}(\Omega)$-volume forces with the following property.

**Theorem 4.2.19** (Quasi-optimality of $M_{\mathrm{HL}}$)**.** *For all $\eta > 0$, the method $M_{\mathrm{HL}}$ is $\|\cdot\|_{\lambda;\eta}$-quasi-optimal for* (4.2.27) *with quasi-optimality constant constant $\leq \sqrt{1 + C_{d,\gamma_{\mathcal{M}}}(2\mu+\lambda)\eta^{-1}}$.*

*Proof.* We proceed as in the proof of Theorem 4.2.7. First, we derive the quasi-optimality of $M_{\mathrm{HL}}$ from item (ii) of Theorem 4.2.1, combined with Proposition 3.3.2 and Lemma 4.2.18. Then, it only remains to bound the consistency measure in the identity

$$C_{\mathrm{qopt}} = \sqrt{1 + \left(\delta_{H_0^1(\Omega)^d}\right)^2}$$

provided by item (iii) of Theorem 4.2.1. Let $v \in H_0^1(\Omega)^d$, $\sigma \in CR^d$, and denote by $\Pi_{\lambda;\eta}$ the $a_{\lambda;\eta}$-orthogonal projection onto $CR^d$. Lemma 4.2.18, the definition of $CR$, $[\![v]\!] = 0 = [\![E_1\sigma]\!]$, and the definition of $\Pi_{\lambda;\eta}$ imply

$$b_{\mathrm{HL}}(\Pi_{\lambda;\eta}v,\sigma) - a_\lambda(v, E_1\sigma) = a_\lambda(\Pi_{\lambda;\eta}v - v, E_1\sigma - \sigma).$$

for all $v$ in $H_0^1(\Omega)^d$ and $\sigma$ in $CR^d$. Following the notation of Theorem 4.2.1, we introduce the bilinear form $d_{\mathrm{HL}} : H^1(\mathcal{M}) \times S_p^0 \to \mathbb{R}$

$$d_{\mathrm{HL}}(\widetilde{v},\sigma) := b_{\mathrm{HL}}(\Pi_{\lambda,\eta}\widetilde{v},\sigma) - a_\lambda(\widetilde{v}, E_1\sigma)$$

and observe that the previous identity, combined with Proposition 4.2.5, yields

$$|d_{\mathrm{HL}}(v,\sigma)| \leq C_{d,\gamma_{\mathcal{M}}}\,(\mu\eta)^{-1/2}\sqrt{2\mu+\lambda}\,\|\Pi_{\lambda;\eta}v - v\|_{\lambda;\eta}\|\sigma\|_{\lambda;\eta}.$$

Hence, we have $\delta_{H_0^1(\Omega)^d} \lesssim \eta^{-\frac{1}{2}}\sqrt{2\mu+\lambda}$ and the proof is finished. □

Thus, we recover Cea's lemma in the conforming limit $\eta \to +\infty$, as it is proved for the quasi-optimal NIP and SIP methods in the previous section. The following remarks show that the upper bound of the quasi-optimality constant in Theorem 4.2.19 captures the correct asymptotic behavior also for $\eta \to 0$ and $\lambda \to +\infty$. It is worth mentioning that the second remark is closely connected with the observations in [48] by Linke, about poor mass conservation in the approximation of incompressible flows.

*Remark* 4.2.20 ($C_{\mathrm{qopt}}$ as $\eta \to 0$). The degeneracy of the bilinear form $a_{\lambda;\eta}(\cdot, E_1 \cdot)$ entails $C_{\mathrm{qopt}} \geq C_\lambda \eta^{-\frac{1}{2}}$. To see this, suppose that $s_0$ satisfies (4.2.30) and notice that Lemma 3.3.1 guarantees that $E_1$ is injective. We then have that $C_\lambda = \|E_1 s_0\|_{\lambda;\eta} \neq 0$ and $\|s_0\|_{\lambda;\eta} = C\eta^{\frac{1}{2}}$. Hence, items (i) and (iii) of Theorem 4.2.1 yield $C_{\mathrm{qopt}} \geq C_{\mathrm{stab}} = \|E_1\|_{\mathcal{L}(CR^d, H_0^1(\Omega)^d)} \geq C_\lambda \eta^{-\frac{1}{2}}$.

*Remark* 4.2.21 ($C_{\mathrm{qopt}}$ as $\lambda \to +\infty$). The property

$$E_1\big(\{s \in CR^d \mid \operatorname{div}_{\mathcal{M}} s = 0\}\big) \not\subseteq \{v \in H_0^1(\Omega)^d \mid \operatorname{div} v = 0\} \tag{4.2.33}$$

results in $C_{\mathrm{qopt}} \geq C_\eta \lambda^{\frac{1}{2}}$. Indeed, if $s \in CR^d$ is such that $\operatorname{div}_{\mathcal{M}} s = 0$ and $\operatorname{div}(E_1 s) \neq 0$, we have $C_\eta = \|s\|_{\lambda;\eta} \neq 0$ and $\|E_1 s\|_{\lambda;\eta} \approx C\lambda^{\frac{1}{2}}$ as $\lambda \to \infty$ and so Theorem 4.2.1 implies $C_{\mathrm{qopt}} \geq C_{\mathrm{stab}} \geq C_\eta \lambda^{\frac{1}{2}}$.

In order to verify (4.2.33), fix any face $F \in \mathcal{F}^i$ of a given mesh $\mathcal{M}$. Let $\Psi_F$ the associated basis function in $CR$. Since $\Psi_{F|K} = 0$ whenever $F \notin \mathcal{F}_K$, we can appropriately pick the elements $K_z$ in the definition (2.2.1) of $A_1$ and arrange $A_1 \Psi_F = 0$. This entails $E_1 \Psi_F = \beta \Phi_F$ with some $\beta > 0$ and $\Phi_F$ as in (3.3.5). Consider $\Psi_F t_F \in CR^d$, where $t_F$ is a unit tangent vector of $F$. On the one hand, we have $\operatorname{div}_{\mathcal{M}}(t_F \Psi_F) = t_F \cdot \nabla_{\mathcal{M}} \Psi_F = 0$ and, on the other hand, $\operatorname{div} E_1(t_F \Psi_F) = \beta \operatorname{div}(t_F \Phi_F) = \beta t_F \cdot \nabla \Phi_F \neq 0$.

It is instructive to shed additional light on the performance of $M_{\mathrm{HL}}$ for nearly incompressible materials. One may note that the choice $\eta \approx \lambda$ makes the quasi-optimality constant of $M_{\mathrm{HL}}$ independent of $\lambda$. However, this will not result in a robust approximation for the nearly incompressible limit $\lambda \to +\infty$. In fact, the argument in Remark 4.2.9 can be used to check that the solution $U$ of (4.2.32) approaches the space $S_1^1$ as $\eta \to +\infty$. Moreover, it is known that $S_1^1$ shows locking on a wide family of meshes, due to the poor approximation properties of $\{s \in S_1^1 \mid \operatorname{div} s = 0\}$; see [21].

In contrast, the use of a fixed penalty parameter cannot guarantee a uniform control of the quasi-optimality constant as $\lambda \to +\infty$, but provides robust approximation under standard assumptions. To see this, the following lemma, which is also of interest by its own, will be useful. It quantifies the difference between the original method $\hat{M}_{\mathrm{HL}}$ of Hansbo and Larson, and its new variant $M_{\mathrm{HL}}$. Recall that, if $f \in L^2(\Omega)^d$, the discrete solution $\hat{U} \in CR^d$ of Hansbo and Larson [44] is given by

$$\forall \sigma \in CR^d \quad b_{\mathrm{HL}}(\hat{U}, \sigma) = \int_\Omega f \cdot \sigma. \tag{4.2.34}$$

**Lemma 4.2.22** ($M_{\mathrm{HL}}$ and $\hat{M}_{\mathrm{HL}}$). *Assume $f \in L^2(\Omega)^d$ and let $U, \hat{U} \in CR^d$ solve* (4.2.32) *and* (4.2.34)*, respectively. Then*

$$\|U - \hat{U}\|_{\lambda;\eta} \leq C_{d,\gamma_{\mathcal{M}}} \min\left\{1, \eta^{-\frac{1}{2}}\right\} \left(\sum_{K \in \mathcal{M}} h_K^2 \|f\|^2_{L^2(K)^d}\right)^{\frac{1}{2}}. \tag{4.2.35}$$

*Proof.* The definition of $U$ and $\hat{U}$ immediately gives

$$\|U - \hat{U}\|_{\lambda;\eta} = \sup_{\|\sigma\|_{\lambda;\eta}=1} \left| b_{\mathrm{HL}}(U - \hat{U}, \sigma) \right| = \sup_{\|\sigma\|_{\lambda;\eta}=1} \left| \int_\Omega f \cdot (E_1\sigma - \sigma) \right|,$$

where $\sigma$ varies in $CR^d$. For any element $K \in \mathcal{M}$, we have $\int_{\partial K} E_1\sigma - \sigma = 0$ implying the Poincaré inequality $\|E_1\sigma - \sigma\|_{L^2(K)^d} \lesssim h_K \|\nabla(E_1\sigma - \sigma)\|_{L^2(K)^d}$. Therefore,

$$\left| \int_\Omega f \cdot (E_1\sigma - \sigma) \right| \lesssim \left( \sum_{K \in \mathcal{M}} h_K^2 \|f\|^2_{L^2(K)^d} \right)^{\frac{1}{2}} \| \nabla_{\mathcal{M}} (E_1\sigma - \sigma)\|_{L^2(\Omega)^d}.$$

Hence, an application of Proposition 4.2.5 shows that the best constant in (4.2.35) is $\leq C_{d,\gamma_{\mathcal{M}}}(\mu\eta)^{-\frac{1}{2}}$. To check also the inequality $\leq C_{d,\gamma_{\mathcal{M}}}$, it is sufficient to apply Proposition 3.3.2 and combine it with the piecewise Korn inequality [15, Theorem 3.1] by Brenner. □

We readily see from this proof that the asymptotic closeness of $U$ and $\hat{U}$ could be increased, for regular loads, by requiring that the smoothing operator conserves also element moments. We shall use a similar argument in Lemma 5.3.1 to compare different nonconforming methods for the two-dimensional Poisson problem.

A consequence of Lemma 4.2.22 is the following equivalence concerning the asymptotic error bounds

$$\|u - U\|_{\lambda;\eta} \leq Ch\|f\|_{L^2(\Omega)^d}, \quad \|u - \hat{U}\|_{\lambda;\eta} \leq \hat{C}h\|f\|_{L^2(\Omega)^d}$$

with best constants $C$ and $\hat{C}$ for all $h := \max_{K\in\mathcal{M}} h_K$ and $f \in L^2(\Omega)^d$:

$$C \text{ is independent of } \lambda \iff \hat{C} \text{ is independent of } \lambda. \tag{4.2.36}$$

Therefore, the robustness result [44, Theorem 3.1] for $\hat{M}_{\mathrm{HL}}$, which ensures that $\hat{C}$ is independent of $\lambda$ for polygons $\Omega \subseteq \mathbb{R}^2$, carries over to $M_{\mathrm{HL}}$. In summary, for smooth volume forces, the method $M_{\mathrm{HL}}$ is locking-free. The non-robustness of the quasi-optimality constant is thus due to rough volume forces, including forces for which the locking-free nonconforming methods in Falk [40], Brenner and Sung [22], and Hansbo and Larson [44] are not defined.

Let us conclude this section with few comments on the generalization to order $p \geq 2$, where $CR$ is replaced by its higher order counterpart from $CR_p$ from Baran and Stoyan [57]. This case is of different nature. In fact, the Korn inequalities of Brenner [15] ensure that $\|\cdot\|_{\lambda;\eta}$ is a norm on $H^1_0(\Omega)^d + CR^d_p$ even for $\eta = 0$. This allows to construct overconsistent methods with the help of the smoother $E_p$ from Proposition 4.2.5.

### 4.2.3 A Quasi-Optimal $C^0$ Interior Penalty Method for the Biharmonic Problem

In this subsection, we introduce a new $C^0$ interior penalty method for the biharmonic problem with clamped boundary conditions (3.3.23) and prove its quasi-optimality. We let $\Omega$ and $\mathcal{M}$ be as in §2.1 and fix $d = 2$. Given $\eta \geq 0$, we set

$$(v,w)_{2;\eta} := \int_\Omega \mathrm{D}^2_{\mathcal{M}} v : \mathrm{D}^2_{\mathcal{M}} w + \int_\Sigma \frac{\eta}{h_F} \left( [\![\nabla v]\!] \cdot n \right) \left( [\![\nabla w]\!] \cdot n \right), \qquad |v|_{2;\eta} := \sqrt{(v,v)_{2;\eta}} \tag{4.2.37}$$

for $v, w \in H^2(\mathcal{M})$ and abbreviate $(\cdot,\cdot)_{2;0}$ to $(\cdot,\cdot)_2$. We consider the following setting for Theorem 4.2.1:

$$V = H^2_0(\Omega), \quad S = S^1_2, \quad \widetilde{a} = (\cdot,\cdot)_{2;\eta} \text{ on } \widetilde{V} = H^2_0(\Omega) + S^1_2. \tag{4.2.38}$$

For all $\eta > 0$, the bilinear form $(\cdot,\cdot)_{2;\eta}$ is a scalar product on

$$H^2_0(\Omega) + S^1_2 \subseteq \{ v \in H^2(\mathcal{M}) \mid \forall F \in \mathcal{F} \quad [\![\nabla v]\!]_{|F} \cdot t_F = 0 \} \tag{4.2.39}$$

where $t_F$ is a tangent unit vector to the edge $F$. The abstract problem (1.2.1) with (4.2.38) provides a weak formulation of the biharmonic problem (3.3.23). The conforming part of $S^1_2$ is the strict subspace

$$S^1_2 \cap H^2_0(\Omega) = \{ s \in S^1_2 \mid [\![\nabla s]\!] \cdot n = 0 \}, \tag{4.2.40}$$

which may be even trivial; cf. Remark 3.3.12. Finally, as for the other examples of this chapter, overconsistency is ruled out by the inclusion

$$\{0\} \neq S^1_1 \subseteq S^1_2 \cap H^2_0(\Omega)^\perp \tag{4.2.41}$$

as it is shown in Lemma 3.2.4.

Let us turn to the choice of the smoothing operator. Interestingly, Brenner and Sung [23] propose a $C^0$ interior penalty method $M_{\mathrm{BS}}$ involving a smoothing operator based upon nodal averaging. In contrast to similar methods, $M_{\mathrm{BS}}$ is well-defined for general loads $\ell \in H^{-2}(\Omega)$, fully stable according to Theorem 4.2.1 (i), and, for any $\alpha > 0$ and all $\ell \in H^{-2+\alpha}(\Omega)$, its error in $|\cdot|_{2;\eta}$, with a suitable $\eta$, decays at the optimal rate $\alpha$. Nevertheless, $M_{\mathrm{BS}}$ is not guaranteed to be quasi-optimal with respect to $|\cdot|_{2;\eta}$, because it is not designed to be fully algebraically consistent. The discussion in §5.3.3, although in a simplified setting, is intended to clarify this point.

To devise a method ensuring full algebraic consistency, we proceed as before and derive the following counterpart of Lemma 4.2.3 with the help of the integration by parts (2.1.3).

**Lemma 4.2.23** (Moment conservation)**.** *If we are given a smoothing operator $E : S^1_2 \to H^2_0(\Omega)$ satisfying*

$$\forall \sigma \in S^1_2, F \in \mathcal{F}^i \quad \int_F \nabla E\sigma = \int_F \{\!\{\nabla\sigma\}\!\}\,, \tag{4.2.42}$$

*then*

$$\forall s, \sigma \in S^1_2 \quad (s, E\sigma)_{2;\eta} = \int_\Omega \mathrm{D}^2_{\mathcal{M}}\, s : \mathrm{D}^2_{\mathcal{M}}\, \sigma - \int_\Sigma \left\{\!\!\left\{\frac{\partial^2 s}{\partial n^2}\right\}\!\!\right\} [\![\nabla\sigma]\!] \cdot n.$$

Thanks to $S^1_2 + H^2_0(\Omega) \subseteq C^0(\overline{\Omega})$ and the fundamental theorem of calculus, we may ensure the conservation (4.2.42) of the mean gradients on faces by the conditions

$$\begin{aligned} &\forall z \in \mathcal{L}^i_1 \;\; E\sigma(z) = \sigma(z) \\ &\forall F \in \mathcal{F}^i \;\; \int_F \nabla E\sigma \cdot n = \int_F \{\!\{\nabla\sigma\}\!\} \cdot n. \end{aligned} \tag{4.2.43}$$

The smoothing operator for Morley functions in §3.3.3 verifies similar requirements. We adapt its construction to the current setting, focusing on the modifications only.

For all $F \in \mathcal{F}^i$, recall the face bubble $\bar{\Phi}_{n_F} \in H^2_0(\Omega)$ introduced in (3.3.26). On the one hand, the bubble smoother $B_{\partial_n} : H^2(\mathcal{M}) \to H^2_0(\Omega)$

$$B_{\partial_n}\sigma := \sum_{F\in\mathcal{F}^i} \left(\int_F \{\!\{\nabla\sigma\}\!\} \cdot n\right) \bar{\Phi}_{n_F}$$

verifies $B_{\partial_n}\sigma(z) = 0$ for all $z \in \mathcal{L}^i_1$ and the second part of (4.2.43). On the other hand, the simpified averaging operator $A_{HCT}$ into the HCT space (2.2.10) can be used to stabilize $B_{\partial_n}$ and incidentally fulfills the first part of (4.2.43). The combination of bubble smoother and averaging thus provides, as usual, the desired moment conservation in a stable manner.

**Proposition 4.2.24** (Stable smoothing with moment conservation)**.** *The linear operator $E_{\mathrm{C}^0} : S^1_2 \to H^2_0(\Omega)$ given by*

$$E_{\mathrm{C}^0}\sigma := A_{HCT}\sigma + B_{\partial_n}(\sigma - A_{HCT}\sigma)$$

*is invariant on $S^1_2 \cap H^2_0(\Omega)$, verifies* (4.2.42) *and, for all $\sigma \in S^1_2$,*

$$\| \mathrm{D}^2_{\mathcal{M}}(\sigma - E_{\mathrm{C}^0}\sigma)\|_{L^2(\Omega)} \le C_{\gamma_{\mathcal{M}}} \left(\sum_{F\in\mathcal{F}} h_F^{-1} \| [\![\nabla\sigma]\!] \cdot n\|^2_{L^2(F)}\right)^{\frac{1}{2}}.$$

*Proof.* By construction, the operator $A_{HCT}$ is invariant on the conforming part of $S^1_2$, entailing that $E_{\mathrm{C}^0}\sigma = A_{HCT}\sigma = \sigma$ for all $\sigma \in S^1_2 \cap H^2_0(\Omega)$. The fact that $E_{\mathrm{C}^0}$ verifies (4.2.43) follows from the properties of $B_{\partial_n}$ and $A_{HCT}$ mentioned above and can be checked as in the proof of Proposition 3.3.15. This entails that $E_{\mathrm{C}^0}$ fulfills also (4.2.42).

In order to prove the claimed stability bound, we proceed by standard steps. Let $\sigma \in S_2^1$ and $K \in \mathcal{M}$. The triangle inequality yields

$$\begin{aligned}\| \mathrm{D}^2(E_{\mathrm{C}^0}\sigma - \sigma)\|_{L^2(K)} \leq & \| \mathrm{D}^2(\sigma - A_{HCT}\sigma)\|_{L^2(K)} + \\ & + \| \mathrm{D}^2 B_{\partial_n}(\sigma - A_{HCT}\sigma)\|_{L^2(K)}.\end{aligned}$$

Expanding the first term on the right-hand side, we find

$$\begin{aligned}(\sigma - A_{HCT}\sigma)_{|K} = & \sum_{z\in\mathcal{L}_1(K)} \sum_{j=1}^{2} \left[\partial_j(\sigma_{|K})(z) - \partial_j(A_{HCT}\sigma)(z)\right] \Upsilon_z^j + \\ & + \sum_{F\in\mathcal{F}_K} \left[\nabla(\sigma_{|K})(m_F)\cdot n_F - \nabla(A_{HCT}\sigma)(m_F)\cdot n_F\right] \Upsilon_F.\end{aligned}$$

We thus derive

$$\| \mathrm{D}^2(\sigma - A_{HCT}\sigma)\|_{L^2(K)} \leq C_{\gamma_{\mathcal{M}}} \sum_{F\in\mathcal{F}, F\cap K\neq\emptyset} h_F^{-\frac{1}{2}} \| [\![\nabla\sigma]\!]\cdot n\|_{L^2(F)}$$

with the help of the triangle inequality, Lemmas 2.2.5-2.2.7 and (4.2.40). Similarly, we expand the second term

$$B_{\partial_n}(\sigma - A_{HCT}\sigma)_{|K} = \sum_{F\in\mathcal{F}_K\cap\mathcal{F}^i} \left(\alpha_F \int_F \nabla\Upsilon_z^j \cdot n_F + \beta_F \int_F \nabla\Upsilon_F\cdot n_F\right) \bar{\Phi}_{n_F}$$

with

$$\begin{aligned}\alpha_F &= \sum_{z\in\mathcal{L}_1(F)} \sum_{j=1}^{2} \left[\partial_j(\sigma_{|K})(z) - (\partial_j A_{HCT}\sigma)(z)\right] \\ \beta_F &= \left[\nabla(\sigma_{|K})(m_F)\cdot n_F - \nabla(A_{HCT}\sigma)(m_F)\cdot n_F\right]\end{aligned}$$

The same ingredients used above and Lemma 3.3.14 then yield

$$\| \mathrm{D}^2 B_{\partial_n}(\sigma - A_{HCT}\sigma)\|_{L^2(K)} \leq C_{\gamma_{\mathcal{M}}} \sum_{F\in\mathcal{F}, F\cap K\neq\emptyset} h_F^{-\frac{1}{2}} \| [\![\nabla\sigma]\!]\cdot n\|_{L^2(F)}$$

We conclude by summing this estimate and the previous one over all triangles in $\mathcal{M}$ and recalling that the maximum number of edges touching each triangle is $\leq C_{\gamma_{\mathcal{M}}}$. □

It remains to choose the discrete bilinear form b. In view of (4.2.41), we need to establish nondegeneracy, for example in the vein of the extended energy norm $|\cdot|_{2;\eta}$. Requiring also symmetry then leads to the bilinear form of Brenner and Sung [23]:

$$b_{\mathrm{BS}}(s,\sigma) = (s,\sigma)_{2;\eta} - \int_\Sigma \left( \left\{\!\!\left\{ \frac{\partial^2 s}{\partial n^2} \right\}\!\!\right\} [\![\nabla\sigma]\!] \cdot n + [\![\nabla s]\!] \cdot n \left\{\!\!\left\{ \frac{\partial^2 \sigma}{\partial n^2} \right\}\!\!\right\} \right).$$

Similarly to the SIP bilinear form, there is $\eta_* > 0$, depending on $\gamma_{\mathcal{M}}$, such that

$$\|h^{-\frac{1}{2}} \{\!\!\{ \partial^2\sigma/\partial^2 n \}\!\!\} \|_{L^2(\Sigma)} \leq \eta_* \| \mathrm{D}^2_{\mathcal{M}}\, \sigma \|_{L^2(\Omega)} \tag{4.2.44}$$

and therefore $b_{\mathrm{BS}}$ is $|\cdot|_{2,\eta}$-coercive with constant $\sqrt{\alpha(\eta_*/\eta)}$ whenever $\eta > \eta_*$; cf. (4.2.16) and [23, Lemma 7]. Under this assumption, the discrete problem

$$U \in S^1_2 \text{ such that } \forall \sigma \in S^1_2 \;\; b_{\mathrm{BS}}(U,\sigma) = \langle f, E_{\mathrm{C}^0}\sigma \rangle \tag{4.2.45}$$

is well-posed and introduces a *new $C^0$ interior penalty method* $M_{\mathrm{C}^0}$ for the biharmonic problem (3.3.23). Inspecting $b_{\mathrm{BS}}$, $E_{\mathrm{C}^0}$ and recalling Proposition 4.2.24, we see that $M_{\mathrm{C}^0} = (S^1_2, b_{\mathrm{BS}}, E^\star_{\mathrm{C}^0})$ is a nonconforming Galerkin method with a computationally feasible smoothing operator. It differs from the original method of Brenner and Sung [23] in the choice of the smoother and the following property.

**Theorem 4.2.25** (Quasi-optimality of $M_{\mathrm{C}^0}$). *For any penalty parameter $\eta > \eta_*$, the method $M_{\mathrm{C}^0}$ is $|\cdot|_{2;\eta}$-quasi-optimal for the biharmonic problem (3.3.23) with constant $\leq \sqrt{1 + C_{\gamma_{\mathcal{M}}}\big(\alpha(\eta_*/\eta)\eta\big)^{-1}}$.*

*Proof.* Since $\eta > \eta_*$, the form $b_{\mathrm{BS}}$ is coercive and $M_{\mathrm{C}^0}$ is well-defined. After making use of Lemma 4.2.23, Proposition 4.2.24 and (4.2.40), items (ii) and (iii) of Theorem 4.2.1 reveal that $M_{\mathrm{C}^0}$ is quasi-optimal with

$$C_{\mathrm{qopt}} = \sqrt{1 + \left(\delta_{H^2_0(\Omega)}\right)^2}.$$

To bound the constant $\delta_{H^2_0(\Omega)}$, we denote by $\Pi_\eta$ the $(\cdot,\cdot)_{2;\eta}$-orthogonal projection onto $S^1_2$. For all $v \in H^2_0(\Omega)$ and $\sigma \in S^1_2$, we derive

$$b_{\mathrm{BS}}(\Pi_\eta v, \sigma) - (v, E_{\mathrm{C}^0}\sigma)_2 = (\Pi_\eta v - v, E_{\mathrm{C}^0}\sigma - \sigma)_2 - \int_\Sigma [\![\Pi_\eta v - v]\!] \cdot n \left\{\!\!\left\{ \frac{\partial^2\sigma}{\partial n} \right\}\!\!\right\}$$

with the help of $[\![\nabla E_{\mathrm{C}^0}\sigma]\!] = 0 = [\![\nabla v]\!]$, Lemma 4.2.23 and the orthogonality of $\Pi_\eta$. Introducing the bilinear form $d_{\mathrm{C}^0} : H^2(\mathcal{M}) \times S^1_2 \to \mathbb{R}$

$$d_{\mathrm{C}^0}(\widetilde{v}, \sigma) := b_{\mathrm{BS}}(\Pi_\eta \widetilde{v}, \sigma) - (\widetilde{v}, E_{\mathrm{C}^0}\sigma)_2$$

we infer

$$|d_{\mathrm{BS}}(v,\sigma)| \leq C_{\gamma_{\mathcal{M}}} \eta^{-\frac{1}{2}} |\Pi_\eta v - v|_{2;\eta} |\sigma|_{2;\eta}$$

according to the previous identity, Proposition 4.2.24 and (4.2.44). The coercivity of $b_{\mathrm{BS}}$ thus implies $\delta_{H^2_0(\Omega)} \leq C_{\gamma_{\mathcal{M}}} \big(\alpha(\eta_*/\eta)\eta\big)^{-1/2}$ and the proof is finished. □

The presented approach can be extended to design quasi-optimal $C^0$ interior penalty methods of order $p \geq 3$. Perhaps the simplest manner is to keep the HCT averaging $A_{HCT}$ and to construct a higher order version of the bubble smoother similar to $B_p$ in §4.2.1. This does not result in a nonconforming Galerkin method, but achieves quasi-optimality.

*Remark* 4.2.26 (Locking effect for $\eta \to +\infty$). The argument in Remark 4.2.9 can be used to check that the solution $U \in S_2^1$ of (4.2.45) approaches the conforming subspace $S_2^1 \cap H_0^2(\Omega)$ in the norm $|\cdot|_{2,\bar{\eta}}$, $\bar{\eta} > 0$, for $\eta \to +\infty$. Since such subspace has poor approximation properties for certain combinations of $\Omega$ and $\mathcal{M}$, cf. Remark 3.3.12, the method $M_{\mathrm{C}^0}$ may be affected by locking in the sense of [7] for $\eta \to +\infty$.

# Chapter 5

# Numerical Investigations for the Poisson Problem

In this chapter, which essentially results from [62], we want to investigate numerically some of the new nonconforming methods that have been proposed and proven to be quasi-optimal in the previous Chapters 3 and 4. In this way we confirm and complement the theoretical results therein and substantiate their practical relevance. As usual, we begin with an overview of the main issues that will be discussed here.

## 5.1 Overview

Let $u \in H^1_0(\Omega)$ be the weak solution of the Poisson problem (3.3.1) and let $\mathcal{M}$ be a triangulation of a planar domain $\Omega \subseteq \mathbb{R}^2$. Our numerical tests and comparisons address methods that can be expressed in the form

$$\text{(5.1.1)} \qquad \text{find } U \in S \quad \text{such that} \quad \forall \sigma \in S \;\; b(U, \sigma) = \langle f, E\sigma \rangle,$$

where $S \subseteq S^0_1$ is a subspace of the piecewise affine functions, $b$ is the bilinear form on $S \times S$ of

- the Crouzeix-Raviart method (CR),
- the (non)symmetric interior penalty method (NIP, SIP) or
- the (non)symmetric weak interior penalty method (WNIP, WSIP) penalizing with jump means over edges,

and $E : S \to H^1_0(\Omega)$ is the linear smoothing operator from Proposition 4.2.5. These methods are *variants* of the original ones with the respective bilinear forms, because $E$ reduces to the identity only on the conforming subspace $S^1_1 = H^1_0(\Omega) \cap S^0_1$.

In order to assess the approximation properties of these methods in unified manner, we employ the following extension of the energy norm associated with the Laplacian:

$$\|v\| := \left( \int_\Omega |\nabla_{\mathcal{M}} v|^2 + \int_\Sigma \frac{\mu}{h} |Q_0 [\![v]\!] |^2 \right)^{\frac{1}{2}} \tag{5.1.2}$$

where $\mu > 0$ is a parameter and $Q_0$ is the $L^2(\Sigma)$-orthogonal projection onto the piecewise constant function on $\Sigma$. With this error notion, it can be checked that all methods have comparable approximation properties (cf. Corollary 5.4.2 below). In fact, each one of them is *quasi-optimal with respect to its shape functions*, meaning that

$$\inf_{v \in S_1^0} \|u - v\| \leq \|u - U\| \leq D \inf_{v \in S_1^0} \|u - v\|. \tag{5.1.3}$$

for some constant $D \geq 1$. Hereafter we assume that $D$ always indicates the best constant in the rightmost inequality. Since $S \subseteq S_1^0$, this notion of quasi-optimality slightly improves on that one previously considered.

Our numerical investigations have been conducted in ALBERTA 3.0 (see [45, 54]) and split into two groups. The first group in §5.3 aims at clarifying the role of the smoothing operator $E$ and the structural conditions on it. For this purpose, we carry out several numerical tests with the penalization-free CR bilinear form. The results

- illustrate the enhanced stability and consistency due to suitable smoothing and

- show that, for reasonable meshes, the stability constant associated with $E$ is only slightly bigger than 1, the stability constant for the weak formulation of the Poisson problem.

The second group of numerical investigations in §5.4 conducts a comparison of the aforementioned nonconforming methods and the continuous Galerkin (cG) method. Here we obtain numerical evidence of the following statements:

- the quasi-optimality constants of all involved methods are rather similar,

- the trade-off between error and number of degree of freedom only slightly favors the cG method over the considered nonconforming methods.

This suggests that nonconforming methods are a valid alternative to conforming methods in that their greater flexibility and the associated advantages in more involved situations come at a relatively low price.

## 5.2 A Pool of Quasi-Optimal Nonconforming Methods

Let $\Omega$ be a planar polygonal domain, i.e. $\Omega \subset \mathbb{R}^2$ is nonempty, open, connected, bounded and $\partial\Omega$ is polygon. We also assume that $\partial\Omega$ is locally represented by a Lipschitz graph. Although this assumption could be relaxed, it will cover all the examples in this chapter.

Two possible extensions of the norm $\|\nabla \cdot \|_{L^2(\Omega)}$ to $H^1(\mathcal{M}) \supseteq H_0^1(\Omega) + S_1^0$ have been proposed in §4.2.1. The first one, which is used to measure errors for the SIP and NIP methods is

$$(5.2.1) \qquad |v|_\eta = \left( \int_\Omega |\nabla_\mathcal{M} v|^2 + \int_\Sigma \frac{\eta}{h} |\llbracket v \rrbracket|^2 \right)^{\frac{1}{2}}, \quad v \in H^1(\mathcal{M}).$$

where $\eta > 0$ is a penalty parameter. The second one is the norm used with the WSIP and WNIP methods, namely

$$(5.2.2) \qquad \|v\|_\mu = \left( \int_\Omega |\nabla_\mathcal{M} v|^2 + \mu \sum_{F \in \mathcal{F}} \left( \fint_F \llbracket v \rrbracket \right)^2 \right)^{\frac{1}{2}},$$

where $\mu > 0$ is also a penalty parameter. Notice the notations $|\cdot|_{1,\eta}$ and $|\cdot|_{1,\mu}$ from §4.2.1 are replaced here, respectively, by $|\cdot|_\eta$ and $\|\cdot\|_\mu$ for convenience. In the following lemma, we clarify the relationship between the two options.

**Lemma 5.2.1** (Equivalence of extended energy norms). *For any $\mu > 0$ and $v \in H^1(\mathcal{M})$, we have*

$$(5.2.3) \qquad \|v\|_\mu \le |v|_\mu \le C\gamma_\mathcal{M} \max\{1, \mu\}^{\frac{1}{2}} \|v\|_\mu$$

*Proof.* For any $v \in H^1(\mathcal{M})$ and $F \in \mathcal{F}$, the Cauchy-Schwarz inequality yields

$$(5.2.4) \qquad \left( \fint_F \llbracket v \rrbracket \right)^2 \le |F|^{-1} \int_F |\llbracket v \rrbracket|^2 = \int_F h^{-1} |\llbracket v \rrbracket|^2,$$

whence $\|\cdot\|_\mu \le |\cdot|_\mu$.

In order to show the converse inequality, let $F \in \mathcal{F}$. Since the mean value is the best constant when approximating in $L^2$, we have

$$(5.2.5) \qquad |F|^{-1} \|\llbracket v \rrbracket\|_{L^2(F)}^2 = \left( \fint_F \llbracket v \rrbracket \right)^2 + |F|^{-1} \|\llbracket v \rrbracket - \fint_F \llbracket v \rrbracket\|_{L^2(F)}^2$$

and it remains to bound the second term on the right-hand side suitably. To this end, we proceed as in the proof of Lemma 2.2.2 and consider two cases, $F \in \mathcal{F}^i$ and $F \in \mathcal{F}^b$, and start with the first case. Let $K_1, K_2 \in \mathcal{M}$ be the two elements such that $F = K_1 \cap K_2$. Inserting the face means

$f_j := |F|^{-1} \int_F v_{|K_j}$ as well as the element means $k_j := |K_j|^{-1} \int_{K_j} v$, we obtain

$$\begin{aligned} |F|^{-\frac{1}{2}} \| \llbracket v \rrbracket - \fint_F \llbracket v \rrbracket \|_{L^2(F)} &\leq \\ &\leq \sum_{j=1,2} \left( |f_j - k_j| + |F|^{-\frac{1}{2}} \|v_{|K_j} - k_j\|_{L^2(F)} \right). \end{aligned} \tag{5.2.6}$$

Then, the same argument used to derive the first part of (2.2.6) from (2.2.5) applies and we infer

$$|F|^{-\frac{1}{2}} \| \llbracket v \rrbracket - \fint_F \llbracket v \rrbracket \|_{L^2(F)} \leq C\gamma_{\mathcal{M}} \sum_{j=1}^{2} \|\nabla \sigma\|_{L^2(K_j)} \tag{5.2.7a}$$

in this case. If, instead, $F \in \mathcal{F}^b$, we denote by $K \in \mathcal{M}$ the element with $F = K \cap \partial\Omega$ and, similarly, using the means $f := |F|^{-1} \int_F \sigma_{|K}$ and $k := |K|^{-1} \int_K \sigma$, we obtain

$$|F|^{-\frac{1}{2}} \| \llbracket v \rrbracket - \fint_F \llbracket v \rrbracket \|_{L^2(F)} \leq C\gamma_{\mathcal{M}} \|\nabla \sigma\|_{L^2(K)}. \tag{5.2.7b}$$

Inserting (5.2.7) into (5.2.5) the proof is finished. □

In view of this lemma, we shall refer hereafter only to the norm $\| \cdot \|_\mu$. To motivate this choice, we observe that $\| \cdot \|_\mu$ reduces to $\| \nabla_{\mathcal{M}} \cdot \|_{L^2(\Omega)}$ on the sum $H^1_0(\Omega) + CR$, irrespective to the parameter $\mu$. An interesting consequence of this is that its corresponding best error in $S^0_1$ is locally computable and the best approximation does not depend on $\mu$. To see this, let $K \in \mathcal{M}$ and recall that $\mathbb{P}_1(K)$ is determined by the three functionals $v \mapsto \fint_F v$, $F \in \mathcal{F}_K$. Writing $\Psi_{K,F}$, $F \in \mathcal{F}_K$, for the associated nodal basis satisfying $\fint_{F'} \Psi_{K,F} = \delta_{F,F'}$ for all $F, F' \in \mathcal{F}_K$, we define

$$\Pi v := \sum_{K \in \mathcal{M}} \sum_{F \in \mathcal{F}_K} \left( \fint_F v_{|K} \right) \Psi_{K,F}, \quad v \in H^1(\mathcal{M}), \tag{5.2.8}$$

which is an extension of Crouzeix-Raviart interpolation (3.1.1).

**Lemma 5.2.2** (Best approximant). *The operator* $\Pi$ *is the* $(\cdot,\cdot)_\mu$*-orthogonal projection onto* $S^0_1$ *and, for any* $v \in H^1(\mathcal{M})$*, we have*

$$\inf_{s \in S^0_1} \|v - s\|_\mu = \|v - \Pi v\|_\mu.$$

*Proof.* As $\| \cdot \|_\mu$ is induced by the scalar product

$$(v, w)_\mu := \int_\Omega \nabla_{\mathcal{M}} v \cdot \nabla_{\mathcal{M}} w + \mu \sum_{F \in \mathcal{F}} \left( \fint_F \llbracket v \rrbracket \right) \left( \fint_F \llbracket w \rrbracket \right)$$

on $H^1(\mathcal{M})$, the function $\Pi v$ is the best approximant for any $v \in H^1(\mathcal{M})$ if and only if $\Pi$ is the $(\cdot,\cdot)_\mu$-orthogonal projection onto $S_1^0$. In order to verify the latter, we observe that the definition of $\Pi$ implies

$$\forall K \in \mathcal{M}, F \in \mathcal{F}_K \quad \int_F (\Pi v)_{|K} = \int_F v_{|K}, \tag{5.2.9}$$

whence, with the help of integration by parts (2.1.3),

$$\int_K \partial_i(\Pi v) = \int_{\partial K} (\Pi v)_{|K} n_i = \int_{\partial K} v_{|K} n_i = \int_K \partial_i v, \quad i = 1, 2.$$

In other words: for any element $K \in \mathcal{M}$ and edge $F \in \mathcal{F}$, the function $\Pi v$ is a local best approximation in $\mathbb{P}_1(K)$ with respect to $\|\nabla \cdot\|_{L^2(K)}$ and in $\mathbb{P}_0(F)$ with respect to $\|\cdot\|_{L^2(F)}$. We thus conclude the orthogonality $(v - \Pi v, \sigma)_\mu = 0$ for all $\sigma \in S_1^0$. □

*Remark* 5.2.3 (Motivating the Crouzeix-Raviart space). Motivated by the weak formulation of the Poisson problem, we restrict to target functions $u \in H_0^1(\Omega)$. Then we have $⨍_F [\![\Pi u]\!] = 0$ for all $F \in \mathcal{F}$, so that $\Pi u \in CR$ and

$$\inf_{s \in S_1^0} \|u - s\|_\mu = \inf_{s \in CR} \|u - s\|_\mu \tag{5.2.10}$$

In other words, the approximability offered by $S_1^0$ in the norm $\|\cdot\|_\mu$ is the same one provided by the subspace $CR \subsetneqq S_1^0$.

Combining the proof of Lemma 5.2.2 with the Poincaré inequality of Payne and Weinberger [52], we immediately obtain the following bound in terms of the seminorm

$$\left|hD^2 v\right|_{0;\Omega}^2 := \sum_{K \in \mathcal{M}} h_K^2 \sum_{|\alpha|=2} \|\partial^\alpha v\|_{L^2(K)}^2, \quad v \in H^2(\mathcal{M}).$$

**Lemma 5.2.4** (Error bound). *For any $v \in H^2(\mathcal{M})$, we have*

$$\|v - \Pi v\|_\mu \leq \pi^{-1} \left|hD^2 v\right|_{0;\Omega}$$

This bound is optimal in that the induced convergence rate is the maximal one and the employed regularity is the minimal one (within Hilbert spaces).

We shall compare the quasi-optimal first order methods introduced in §3.3.1 and §4.2.1. In order to handle them in a unified fashion, let us define

$$\begin{aligned} b_{\epsilon,\mu,\eta}(s,\sigma) := &\int_\Omega \nabla_\mathcal{M} s \cdot \nabla_\mathcal{M} \sigma - \int_\Sigma \{\!\{\nabla s\}\!\} \cdot n [\![\sigma]\!] \\ &+ \epsilon \int_\Sigma [\![s]\!] \{\!\{\nabla \sigma\}\!\} \cdot n + \mu \sum_{F \in \mathcal{F}} \left(⨍_F [\![s]\!]\right)\left(⨍_F [\![\sigma]\!]\right) + \int_\Sigma \frac{\eta}{h} [\![s]\!] [\![\sigma]\!] \end{aligned}$$

for $s, \sigma \in S^0_1$. The discrete bilinear form of each method can be recovered from $b_{\epsilon,\mu,\eta}$ by a suitable combination of $\epsilon, \mu$ and $\eta$. For instance, penalizing the jumps, i.e. $\eta > 0$ and $\mu = 0$, we obtain the bilinear forms $b_{\mathrm{sip},\eta} := b_{-1,0,\eta}$ and $b_{\mathrm{nip},\eta} := b_{1,0,\eta}$ of the SIP and NIP methods. If, instead, we penalize with the jump means, i.e. $\eta = 0$ and $\mu > 0$, we find the counterparts $b_{\mathrm{wsip},\mu} := b_{-1,\mu,0}$ and $b_{\mathrm{wnip},\mu} := b_{1,\mu,0}$ of the previous forms with weak interior penalty. Finally, the restriction of $b_{\epsilon,\mu,0}$ to the Crouzeix-Raviart space reduces to

$$b_{CR}(s, \sigma) := \int_\Omega \nabla_{\mathcal{M}}\, s \cdot \nabla_{\mathcal{M}}\, \sigma, \quad s, \sigma \in CR.$$

We shall employ the bilinear forms $b_{\mathrm{op}}$ with $\mathrm{op} \in \{\mathrm{sip}, \mathrm{nip}, \mathrm{wsip}, \mathrm{wnip}, CR\}$.

The respective discrete space is $S_{\mathrm{op}} = S^0_1$ for $\mathrm{op} \in \{\mathrm{sip}, \mathrm{nip}, \mathrm{wsip}, \mathrm{wnip}\}$ and $S_{CR} = CR$ otherwise. In any case, its conforming subspace is given by $S^{\mathrm{op}} \cap H^1_0(\Omega) = S^1_1$, which only sees the following common part of the aforementioned bilinear forms:

$$b_0(s, \sigma) := \int_\Omega \nabla_{\mathcal{M}}\, s \cdot \nabla_{\mathcal{M}}\, \sigma - \int_\Sigma \{\!\{\nabla s\}\!\} \cdot n \, [\![\sigma]\!], \quad s, \sigma \in S^0_1; \tag{5.2.11}$$

note that $b_0$ simplifies to $b_{CR}$ on $CR$ because $\{\!\{\nabla s\}\!\}$ is piecewise constant. While $b_0$ is symmetric and $\|\cdot\|_\mu$-coercive on $CR$, it is degenerate on $S^0_1$, see also (4.2.4). Thus, for the bilinear form $b^{\mathrm{op}}$, with $\mathrm{op} \neq CR$, the terms in addition to $b_0$ arrange $\|\cdot\|_\mu$-coercivity and, in some cases, also symmetry.

We also propose two options for the smoothing operator to be employed in the discretization of the right-hand side. Such options differ only in the choice of the averaging operator $A : S^0_1 \to S^1_1$ used to stabilize the bubble smoother $B := B_1 : S^0_1 \to H^1_0(\Omega)$ from Lemma 4.2.4. Indeed, according to Remarks 2.2.4 and 3.3.6, as well as the discussion immediately before Proposition 4.2.5, we may employ either the standard averaging operator from (2.2.8) or the simplified averaging in (2.2.1). Having fixed $p = 1$, we modify the previous notation and write, for further convenience, $A := A_{\mathrm{av}}$ and $A = A_{\mathrm{sz}}$, respectively. The subscript *sz* aims at remarking the close relationship between the latter averaging operator and the Scott-Zhang interpolation [56].

According to Propositions 3.3.2 and 4.2.5, we define

$$E_{\mathrm{smt}}\sigma := A_{\mathrm{smt}}\sigma + B(\sigma - A_{\mathrm{smt}}\sigma) \tag{5.2.12}$$

for $\mathrm{smt} \in \{\mathrm{av}, \mathrm{sz}\}$. Thus, we have five options $\mathrm{op} \in \{\mathrm{sip}, \mathrm{nip}, \mathrm{wsip}, \mathrm{wnip}, CR\}$ to discretize the Laplacian and two options $\mathrm{smt} \in \{\mathrm{av}, \mathrm{sz}\}$ to discretize the right-hand side in the discrete problem

$$\text{find} \quad U^{\mathrm{op}}_{\mathrm{smt}} \in S_{\mathrm{op}} \quad \text{such that} \quad \forall \sigma \in S_{\mathrm{op}} \;\; b_{\mathrm{op}}(U^{\mathrm{op}}_{\mathrm{smt}}, \sigma) = \langle f, E_{\mathrm{smt}}\sigma \rangle$$

with $f \in H^{-1}(\Omega)$. Irrespective of the combination, the corresponding method $M^{\mathrm{op}}_{\mathrm{smt}}$ is a nonconforming Galerkin method in the sense of (1.2.11), because of $E_{\mathrm{smt}|S_1^1} = \mathrm{Id}_{S_1^1}$ and the identity

$$\forall s, \sigma \in S_1^1 \quad b_{\mathrm{op}}(s,\sigma) = b_0(s,\sigma) = \int_\Omega \nabla s \cdot \nabla \sigma.$$

Every method $M_{\mathrm{smt}}^{\mathrm{op}}$ has been implemented with the help of the finite element library ALBERTA 3.0 of [45]. The implementation is standard, except for the assembly of the load vector $(\langle f, E_{\mathrm{smt}}\Psi_i\rangle)_{i\in I}$ where $\{\Psi_i\}_{i\in I}$ is the nodal basis of $S_{\mathrm{op}}$. For this purpose, we first assemble a vector with coefficients $\langle f, \Phi_1^z\rangle$, $z \in \mathcal{L}_1^i$, and $\langle f, \bar{\Phi}_F\rangle$, $F \in \mathcal{F}^i$, and then obtain the desired load vector by multiplication with $R_{\mathrm{smt}}^T$, where $R_{\mathrm{smt}}$ is a rectangular matrix arising from $E_{\mathrm{smt}}$. For both options smt $\in \{\mathrm{av}, \mathrm{sz}\}$, the matrix $R_{\mathrm{smt}}$ is sparse and does not need to be explicitly stored, because it is involved only in one matrix-vector product. Moreover, the sparsity of $R_{\mathrm{av}}$ depends on $\max_{z\in\mathcal{L}_1^i} N_z$ and, in any case, $R_{\mathrm{av}}$ has more nonzero entries as $R_{\mathrm{sz}}$. Consequently, the use of $E_{\mathrm{sz}}$ is convenient from the viewpoint of costs.

## 5.3 Mean-Preserving Smoothing in Action

The smoothing operator distinguishes the methods in §5.2 from most nonconforming methods for the Poisson problem. The goal of this section is to discuss the main features related to the choice of the smoother, to highlight possible advantages of its use, and to clarify the role of the structural conditions on it.

For the sake of simplicity, we restrict to the space $CR$ and the bilinear form $b_{CR}$. This has also the advantage that we do not need to probe for a possible dependence of the numerical data on the choice of the penalty parameter $\mu$ in the method and in the error measure. In fact, the corresponding errors will be then in $H_0^1(\Omega) + CR$ and the value of $\mu > 0$ in the error norm $\|\cdot\|_\mu$ is insignificant.

In order to analyze numerical data, we shall use in particular the so-called experimental order of convergence (EOC) with respect to the number of degree of freedoms (#DOFs). Given two approximations with $N_i$ DOFs and errors $e_i$, $i = 1, 2$, the corresponding EOC 'estimates' the rate of convergence by the ratio

$$\mathrm{EOC} := \frac{\log(e_1/e_2)}{\log(N_1/N_2)}. \tag{5.3.1}$$

### 5.3.1 Regular Forces

Assume that the 'force' in the Poisson problem (3.3.1) verifies

$$f \in L^2(\Omega).$$

Then the classical Crouzeix-Raviart approximation $U^{CR} \in CR$ is well-defined and given by

$$\forall \sigma \in CR \quad \int_\Omega \nabla_{\mathcal{M}} U^{CR} \cdot \nabla_{\mathcal{M}} \sigma = \int_\Omega f\sigma. \tag{5.3.2}$$

Thus, the question arises about the relationship between $U^{CR}$ and the quasi-optimal approximations $U^{CR}_{\mathrm{smt}}$, $\mathrm{smt} \in \{\mathrm{av}, \mathrm{sz}\}$, obtained with smoothing. First, if we assume for the moment that $\Omega$ is convex, then the elliptic regularity, [21, (10.3.11)] and the combination of Theorem 3.3.3 with Lemma 5.2.2 and Lemma 5.2.4 yield

$$\|u - U^{CR}\|_\mu \leq C_{\gamma_{\mathcal{M}}} |hD^2 u|_{0;\Omega} \quad \text{and} \quad \|u - U^{CR}_{\mathrm{smt}}\|_\mu \leq C_{\gamma_{\mathcal{M}}} |hD^2 u|_{0;\Omega}.$$

Thus, the triangle inequality yields the same estimate for the difference $\|U^{CR} - U^{CR}_{\mathrm{smt}}\|_\mu$. Next, the following counterpart of Lemma 4.2.22 slightly improves on the previous assumptions.

**Lemma 5.3.1** (Difference due to smoothing). *If $f \in L^2(\Omega)$, $\mu > 0$ and $\mathrm{smt} \in \{\mathrm{av}, \mathrm{sz}\}$, then*

$$\|U^{CR} - U^{CR}_{\mathrm{smt}}\|_\mu \leq C_{\gamma_{\mathcal{M}}} \left( \sum_{K \in \mathcal{M}} h_K^2 \|f\|^2_{L^2(K)} \right)^{\frac{1}{2}}.$$

*Proof.* The proof is almost the same as for Lemma 4.2.22, despite the different error notion and discrete bilinear form involved. The definitions of $U^{CR}$ and $U^{CR}_{\mathrm{smt}}$ immediately give

$$\begin{aligned} \|U^{CR} - U^{CR}_{\mathrm{smt}}\|_\mu &= \sup_{\|\sigma\|_\mu = 1} \left| \int_\Omega \nabla_{\mathcal{M}} (U^{CR} - U^{CR}_{\mathrm{smt}}) \cdot \nabla_{\mathcal{M}} \sigma) \right| \\ &= \sup_{\|\sigma\|_\mu = 1} \left| \int_\Omega f(\sigma - E_{\mathrm{smt}}\sigma) \right|, \end{aligned}$$

where $\sigma$ varies in $CR$. For all $K \in \mathcal{M}$, we have $\int_{\partial K} E_1\sigma - \sigma = 0$ implying the Poincaré-type inequality $\|E_1\sigma - \sigma\|_{L^2(K)} \lesssim h_K \|\nabla(E_1\sigma - \sigma)\|_{L^2(K)}$. Therefore,

$$\left| \int_\Omega f(E_{\mathrm{smt}}\sigma - \sigma) \right| \leq C_{\gamma_{\mathcal{M}}} \left( \sum_{K \in \mathcal{M}} h_K^2 \|f\|^2_{L^2(K)} \right)^{\frac{1}{2}} \| \nabla_{\mathcal{M}} (\sigma - E_{\mathrm{smt}}\sigma)\|_{L^2(\Omega)}$$

and the $H^1$-stability of $E_{\mathrm{smt}}$ from Proposition 3.3.2 finishes the proof. □

We complement Lemma 5.3.1 with numerical experiments when the solution of the (inhomogeneous) Poisson problem is

$$u^{\mathrm{rsym}}_\rho(x) := |x|^\rho, \quad x \in \Omega = (-1, 1)^2, \tag{5.3.3}$$

for some $\rho > 0$. Note that $-\Delta u_\rho \in L^2(\Omega)$ if and only if $\rho > 1$ and that, for $\rho \notin \mathbb{N}$, we have $u_\rho \in H^s(\Omega)$ if and only if $s < \rho$.

We shall observe several convergence histories. The initial triangulation for (5.3.3) is always given by drawing the diagonals of $\Omega$. Let us first consider uniform refinement, i.e., for each new mesh, two bisections are applied to every triangle. Figures 5.1-5.2 and Tables 5.1-5.2 display data corresponding to the case $\rho = 1.1$. We notice that the errors with smoothing are slightly bigger than without smoothing, more emphasized when using the simplified averaging $A_{\mathrm{sz}}$ to define $E_{\mathrm{sz}}$. The respective EOCs are consistent with Lemma 5.3.1 and the preceding bounds. For stabilization by standard averaging $A_{\mathrm{av}}$, the EOCs associated with the difference $\|U^{CR} - U^{CR}_{\mathrm{av}}\|_\mu$ are even slightly better than predicted by Lemma 5.3.1.

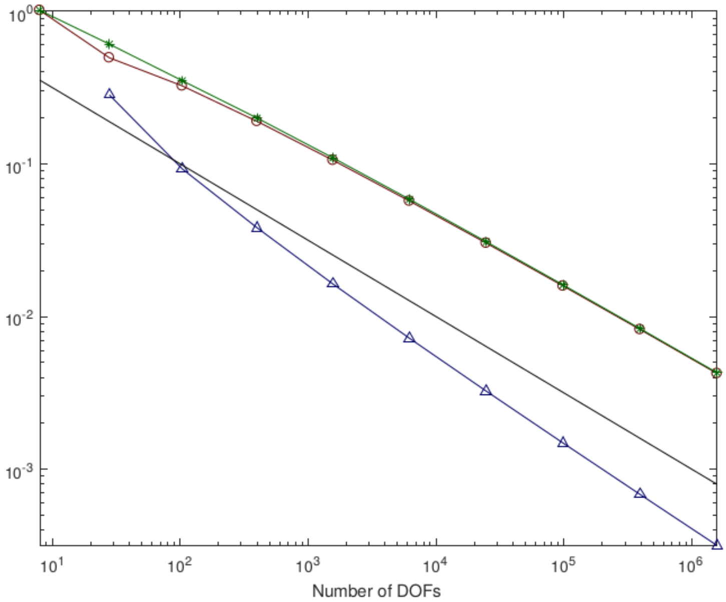


Figure 5.1: Example $u^{\mathrm{rsym}}_{1.1}$: Convergence histories of Crouzeix-Raviart error without smoothing ($\circ$), with smoothing $E_{\mathrm{av}}(*)$, and difference of respective approximations ($\triangle$) for uniform refinement. Plain line indicates decay rate $\#\mathrm{DOFs}^{-0.5}$.

In order to clarify this observation, we consider the case $\rho = 1.9$. Here, Figure 5.3 suggests that, for uniform refinement, the difference between $U^{CR}$ and $U^{CR}_{\mathrm{av}}$ superconverges with maximal rate 0.75, which corresponds to an improved bound with power $\frac{3}{2}$ for $h$ when compared with Lemma 5.3.1. However, this is a delicate effect that seems to hinge on mesh symmetries: the EOCs drop back to 0.5 for random refinement, where triangles are marked for a double bisection with probability $\frac{1}{3}$, see Figure 5.4. In fact, in this case, we do not observe numerical evidence for some superconvergence of the difference between $U^{CR}$ and $U^{CR}_{\mathrm{av}}$. The results of one such test are displayed

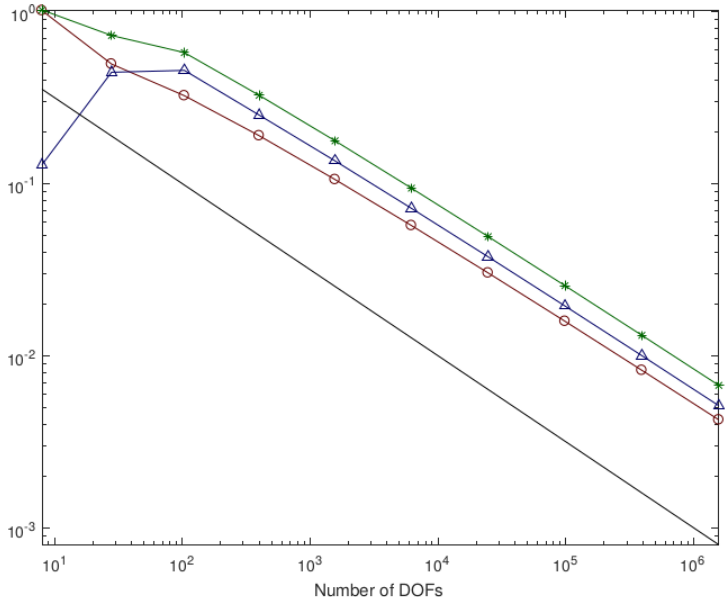


Figure 5.2: Example $u_{1.1}^{\mathrm{rsym}}$: Convergence histories of Crouzeix-Raviart error without smoothing $E_{\mathrm{sz}}(\circ)$, with smoothing $(*)$, and difference of respective approximations $(\triangle)$ for uniform refinement. Plain line indicates decay rate $\#\mathrm{DOFs}^{-0.5}$.

| #DOFs | $\|u - U^{CR}\|_\mu$ | EOC |
|---:|---|---|
| 24 704 | 3.026507e-02 | |
| | | 0.47 |
| 98 560 | 1.585021e-02 | |
| | | 0.47 |
| 393 728 | 8.224799e-03 | |
| | | 0.48 |
| 1 573 888 | 4.238544e-03 | |

Table 5.1: Example $u_{1.1}^{\mathrm{rsym}}$: Crouzeix-Raviart error $\|u - U^{CR}\|_\mu$ without smoothing and its EOCs for the last four triangulations in Figures 5.1-5.2.

| #DOFs | $\|U^{CR} - U_{\mathrm{av}}^{CR}\|_\mu$ | EOC | $\|U^{CR} - U_{\mathrm{sz}}^{CR}\|_\mu$ | EOC |
|---:|---|---|---|---|
| 24 704 | 3.249767e-03 | | 3.754355e-02 | |
| | | 0.57 | | 0.48 |
| 98 560 | 1.485089e-03 | | 1.944739e-02 | |
| | | 0.56 | | 0.48 |
| 393 728 | 6.844091e-04 | | 1.000615e-02 | |
| | | 0.56 | | 0.48 |
| 1 573 888 | 3.170169e-04 | | 5.122284e-03 | |

Table 5.2: Example $u_{1.1}^{\mathrm{rsym}}$: differences $\|U^{CR} - U_{\mathrm{av}}^{CR}\|_\mu$ and $\|U^{CR} - U_{\mathrm{sz}}^{CR}\|_\mu$ and their respective EOCs for the last four triangulations in Figures 5.1-5.2.

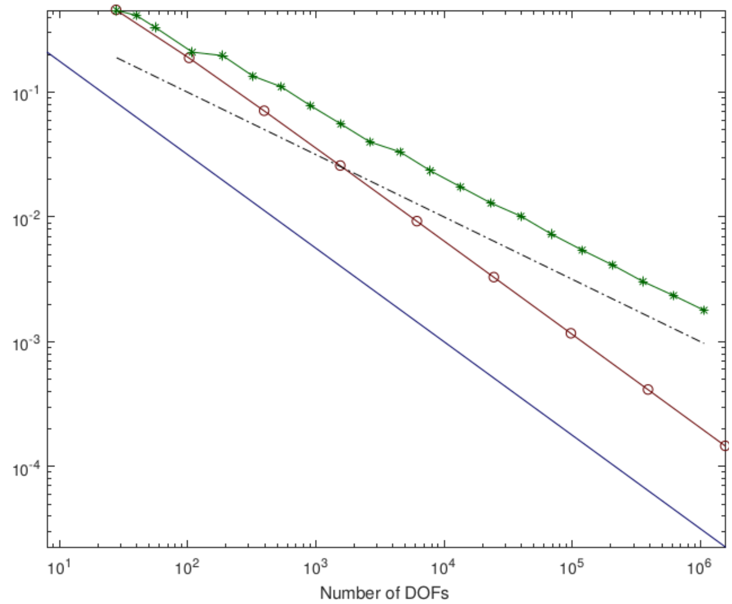


Figure 5.3: Example $u_{1.9}^{\mathrm{rsym}}$: Convergence histories of the difference between $U^{CR}$ and $U_{\mathrm{av}}^{CR}$ for uniform ($\circ$) and random ($*$) refinement. Plain and dash-dot lines indicate, respectively, decay rate $\#\mathrm{DOFs}^{-0.75}$ and $\#\mathrm{DOFs}^{-0.5}$.

in Figure 5.3, while Figure 5.4 shows a triangulation obtained by random refinements.

### 5.3.2 A Numerical Illustration of Full Stability

The weak formulation of the Poisson problem enjoys the following stability property. For any $f \in H^{-1}(\Omega)$, the corresponding solution $u$ verifies

$$\|u\|_{\mu} = \|f\|_{H^{-1}(\Omega)}. \tag{5.3.4}$$

This section investigates up to which degree various methods mimic this property.

In the notation of Chapter 1, the classical Crouzeix-Raviart method (5.3.2) is represented by the triplet $(CR, b_{CR}, L_{CR})$. Identifying $L^2(\Omega)$ with its dual space, the linear operator $L_S : L^2(\Omega) \to CR'$ is given by

$$\langle L_{CR} f, \sigma\rangle := \int_{\Omega} f\sigma, \quad \sigma \in CR.$$

Owing to the piecewise Poincaré-Friedrichs inequality [21, (10.6.12)], we then infer stability with respect to $L^2$-forces

$$\|U^{CR}\|_{\mu} \leq C\|f\|_{L^2(\Omega)}.$$

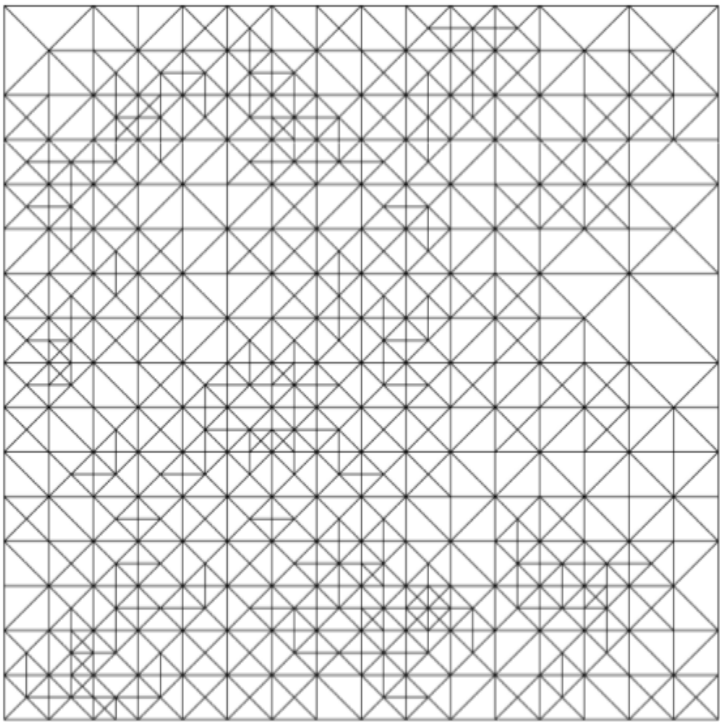


Figure 5.4: A triangulation generated by random refinement.

However, as most nonconforming methods for the Poisson problem, the approximation $U^{CR}$ is not defined for general $f \in H^{-1}(\Omega)$; consequently, and according to Remark 1.4.9, the classical Crouzeix-Raviart method is not fully stable. The following proposition confirms this observation and provides a model argument to check that $L_{CR}$ does not extend to a bounded operator on $H^{-1}(\Omega)$.

**Proposition 5.3.2** (No $H^{-1}$-extension of $L_{CR}$)**.** *The linear operator $L_{CR}$ cannot be boundedly extended to $H^{-1}(\Omega)$.*

*Proof.* First of all, we observe that $\tilde{V} := H_0^1(\Omega) + CR$ is dense in $L^2(\Omega)$ and thus $\tilde{V} \subseteq L^2(\Omega) \subseteq \tilde{V}'$ is a Hilbert triplet. Exploiting the nonconformity, we pick some $s \in CR \setminus H_0^1(\Omega)$. Then the Hahn-Banach theorem guarantees the existence of a linear functional $\ell \in \tilde{V}'$ such that

$$\langle \ell, s \rangle = 1 \quad \text{and} \quad H_0^1(\Omega) \subseteq \ker(\ell).$$

Next, we choose a sequence $(f_k)_{k\in\mathbb{N}} \subseteq L^2(\Omega)$ such that

$$\sup_{w\in\tilde{V}, \|w\|_\mu \leq 1} \left| \langle \ell, w \rangle - \int_\Omega f_k w \right| \to 0 \quad \text{as } k \to \infty$$

and identify each $f_k$ with the functional $H_0^1(\Omega) \ni w \mapsto \int_\Omega f_k w \in \mathbb{R}$. Combining the properties of $\ell$ with the convergence of the sequence $(f_k)_k$, we derive

$$\langle L_{CR} f_k, s\rangle = \int_\Omega f_k s \to 1 \quad \text{and} \quad \|f_k\|_{H^{-1}(\Omega)} \to 0 \quad \text{as} \quad k \to \infty,$$

which yields

$$\frac{\|L_{CR} f_k\|_{CR'}}{\|f_k\|_{H^{-1}(\Omega)}} \geq \frac{\langle L_{CR} f_k, s\rangle}{\|f_k\|_{H^{-1}(\Omega)} \|s\|_\mu} \to +\infty \quad \text{as} \quad k \to +\infty. \qquad \square$$

Proposition 5.3.2 entails that the classical Crouzeix-Raviart method may be affected by undesired numerical artifacts. To illustrate this, we approximate the solution $u_\xi^{\text{lheat}}$ of the Poisson problem, where domain and force are given by

$$\Omega = (0,1)^2, \quad \langle f_\xi, \varphi\rangle = 100 \int_0^1 y\varphi(\xi, y)\, dy \quad \text{with} \quad \xi \in (0,1). \tag{5.3.5}$$

Functionals like $f_\xi$ appear, for example, in the modeling of the production of latent heat by moving phase boundaries. Notice that $f_\xi \in CR'$ and $U^{CR}$ is well-defined for $\xi \neq 0.5$, although $f_\xi$ cannot be identified with the action of a $L^2(\Omega)$-force. However, the low regularity of $f_\xi$ is insignificant to this experiment and serves only to keep the setting in (5.3.5) as simple as possible. Indeed, arguing by density, we could approximate this functional by more regular ones.

In view of

$$\begin{aligned} 10^{-2}\langle f_\xi - f_\nu, \varphi\rangle &\leq \int_0^1 |\varphi(\xi, \cdot) - \varphi(\nu, \cdot)| \\ &\leq \int_{(\xi,\nu)\times(0,1)} |\nabla\varphi| \leq |\xi - \nu|^{\frac{1}{2}} \|\nabla\varphi\|_{L^2(\Omega)} \end{aligned}$$

for $0 \leq \xi \leq \nu \leq 1$, the map

$$(0,1) \ni \xi \mapsto f_\xi \in H^{-1}(\Omega) \tag{5.3.6}$$

is continuous. Figure 5.6 shows that the classical Crouzeix-Raviart approximation $U^{CR}$ in the point $(0.375, 0.375)$ suffers from a jump when passing $\xi = \frac{1}{2}$, where it is not defined, cf. Figure 5.5. Since the point evaluation is a bounded linear operator on the finite-dimensional space $CR$, this means that $U^{CR}$ does not depend on $\xi$ in a continuous manner. In view of (5.3.4) and (5.3.6), this is a numerical artifact.

Notice that, neglecting the requirement of $L^2(\Omega)$-forces, we obtain a concrete example for the sequence in the proof of Proposition 5.3.2 simply by defining $g_\xi := f_\xi - f_{1-\xi}$ for $\xi \in (0,1)$.

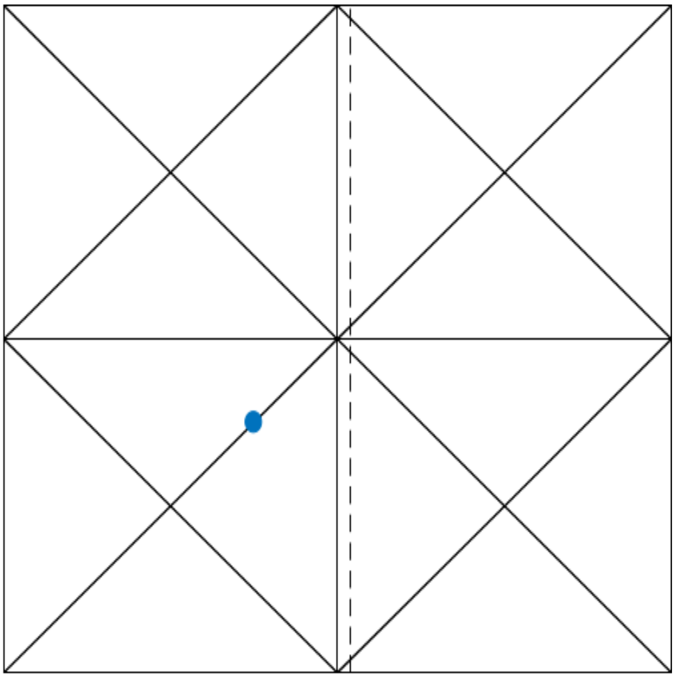


Figure 5.5: Example $u_\xi^{\text{lheat}}$: Mesh with support of $f_{0.52}$ (dashed line) and point of evaluation (•) in (0.375, 0.375).

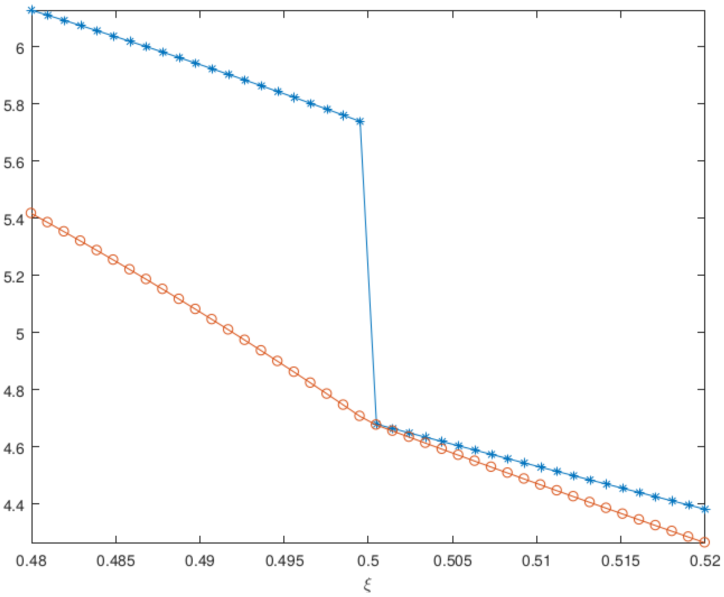


Figure 5.6: Example $u_\xi^{\text{lheat}}$: Behavior of $U^{CR}$ ($*$) and $U_{\text{av}}^{CR}$ ($\circ$) in the indicated point for $0.48 \leq \xi \leq 0.52$.

Figure 5.6 also confirms that $U_{\rm av}^{CR}$ (and, similarly, $U_{\rm sz}^{CR}$) depends continuously on $\xi$. Such continuity hinges on the full stability of the methods $M_{\rm smt}^{CR} := (CR, b_{CR}, E_{\rm smt}^{\star})$, smt $\in \{\rm av, sz\}$, which provides the following discrete counterpart of (5.3.4)

$$\forall f \in H^{-1}(\Omega) \qquad \|M_{\rm smt}^{CR} f\|_{\mu} \leq C_{\rm stab} \|f\|_{H^{-1}(\Omega)}$$

with

$$C_{\rm stab} = \sup_{\sigma \in CR} \frac{\|E_{\rm smt}\sigma\|_{\mu}}{\|\sigma\|_{\mu}} = \|E_{\rm smt}\|_{\mathcal{L}(CR, H_0^1(\Omega))},$$

in view of Corollary 3.2.8 and Proposition 3.3.2.

The last identity entails that we can compute the stability constant of $M_{\rm smt}^{CR}$ by maximizing a generalized Rayleigh quotient. To see this, let $\{\Psi_F\}_{F \in \mathcal{F}^i}$ and $\{\Phi_z^2\}_{z \in \mathcal{L}_2^i}$ denote the nodal bases of the spaces $CR$ and $S_2^1 = E_{\rm smt} CR$, respectively. Introducing the stiffness matrices

$$\mathbb{A}_{FF'} := \int_{\Omega} \nabla_{\mathcal{M}} \Psi_F \cdot \nabla_{\mathcal{M}} \Psi_{F'} \qquad \text{and} \qquad \mathbb{B}_{zz'} := \int_{\Omega} \nabla_{\mathcal{M}} \Phi_z \cdot \nabla_{\mathcal{M}} \Phi_{z'}$$

for $F, F' \in \mathcal{F}^i$ and $z, z' \in \mathcal{L}_2^i$, and denoting by $R_{\rm smt}$ the matrix representing the smoother $E_{\rm smt}$ with respect to these bases, we have

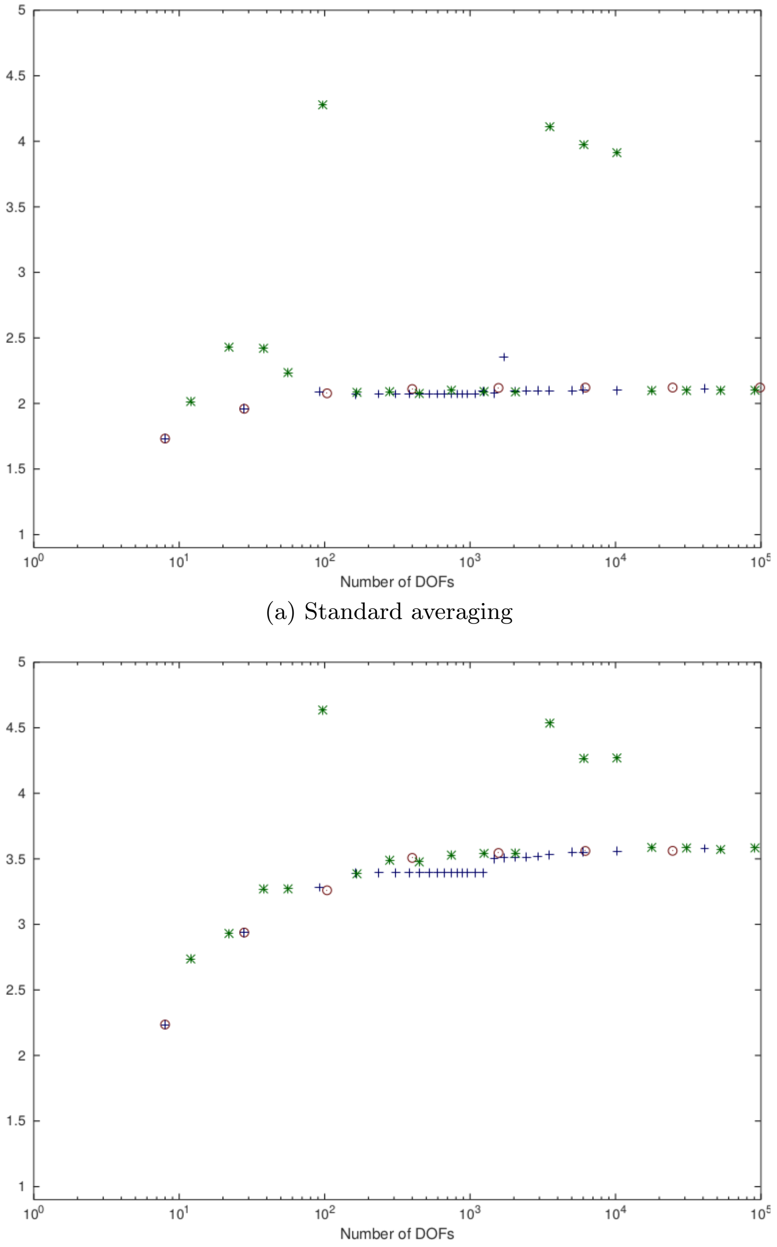
$$C_{\rm stab} = \sup_{x \in \mathbb{R}^{\dim CR}} \frac{x^T (R_{\rm smt} \mathbb{B} R_{\rm smt}^T) x}{x^T \mathbb{A} x}.$$

This formula is particularly attractive, because the quasi-optimality constant of $M_{\rm smt}^{CR}$ coincides with its stability constant, according to Corollary 3.2.8.

With the help of MATLAB command `eigs`, we computed the stability constant of $M_{\rm smt}^{CR}$ on meshes that are generated by uniform and random refinement of $\Omega = (-1,1)^2$ as in §5.3.1. We also considered graded meshes of $\Omega$, obtained by approximating the exact solution $u_{0.25}^{\rm rsym}$ from (5.3.3) by Dörfler's strategy with parameter 0.9 on the local best errors $\|\nabla(u - \Pi u)\|_{L^2(K)}$, $K \in \mathcal{M}$. For all meshes, the number of nonzero entries of the matrix $R_{\rm av}$ associated with $E_{\rm av}$ is about 3-4 times the corresponding number of $R_{\rm sz}$. This saving is typically confronted by a 1.75 times greater stability constant. This hold for uniform and adaptive meshes. Only for randomly refined meshes, there are 'outliers' suggesting that also here local mesh symmetry plays a role. In any case, the observed values of the stability constants are shown in Figure 5.7. All values are quite moderate in size: the maximum for both stabilizations is around 4.5.

### 5.3.3 A Numerical Illustration of Full Algebraic Consistency

This section is intended to illustrate the importance of full algebraic consistency for quasi-optimality. To this end, let us compare the two Crouzeix-Raviart methods $(CR, b_{CR}, E^{\star})$, where $E = E_{\rm av}$ or $E = A_{\rm av}$. Both methods



(a) Standard averaging

(b) Simplified averaging

Figure 5.7: Stability constants for stabilization by nodal averaging and meshes generated by uniform (○), adaptive (+), and random ($*$) refinement.

| #DOFs | #nonzeros in $R_{\mathrm{av}}$ | #nonzeros in $R_{\mathrm{sz}}$ |
|---|---|---|
| 400 | 7664 | 2162 |
| 1568 | 33920 | 9298 |
| 6208 | 142496 | 38546 |
| 24704 | 583904 | 156946 |

Table 5.3: Nonzero elements in the matrices $R_{\mathrm{av}}$ and $R_{\mathrm{sz}}$ with uniform refinements.

are entire and fully stable. However, the former leads to full algebraic consistency, while the latter does not, as we observed in Remark 3.3.5.

To access the difference between the two methods, we first exploit again Lemma 5.2.2. According to Theorem 3.3.3, the smoothing operator $E = E_{\mathrm{av}}$ (or, equivalently, $E = E_{\mathrm{sz}}$) yields quasi-optimality and we have

$$\|u - U_{\mathrm{av}}^{CR}\|_\mu \leq C_{\gamma_{\mathcal{M}}} \left( \sum_{K \in \mathcal{M}} \inf_{p \in \mathbb{P}_1(K)} \|\nabla(u-p)\|^2_{L^2(K)} \right)^{\frac{1}{2}} \tag{5.3.7}$$

where $u$ denotes the weak solution of the Poisson problem and $U_{\mathrm{av}}^{CR}$ is the corresponding approximation.

In contrast, the averaging operator $E = A_{\mathrm{av}}$ (or, equivalently, $E = A_{\mathrm{sz}}$) enjoys the following consistency property, where $\omega_F$ denotes the union of the two triangles containing the interior edge $F \in \mathcal{F}^i$.

**Proposition 5.3.3** (Consistency with plain smoothing)**.** *The bilinear form $b_0$ from* (5.2.11) *and the averaging operator $A_{\mathrm{av}}$ verify*

$$\sup_{\sigma \in S, \|\sigma\|_\mu = 1} \left| b_0(\Pi u, \sigma) - \int_\Omega \nabla u \cdot \nabla A_{\mathrm{av}} \sigma \right| \leq$$
$$\leq C_{\gamma_{\mathcal{M}}} \left( \sum_{F \in \mathcal{F}^i} \inf_{p \in \mathbb{P}_1(\omega_F)} \|\nabla(u-p)\|^2_{L^2(\omega_F)} \right)^{\frac{1}{2}}.$$

*for all $u \in H^1_0(\Omega)$.*

*Proof.* Exploiting integration by parts (2.1.3), that $\nabla_{\mathcal{M}}(\Pi u)$ and $\nabla A_{\mathrm{av}}\sigma$ are piecewise constant, and the local best approximation properties of $\Pi u$, we deduce

$$b_0(\Pi u, \sigma) - \int_\Omega \nabla u \cdot \nabla A_{\mathrm{av}} \sigma = \int_{\Sigma \setminus \partial\Omega} [\![ \nabla \Pi u ]\!] \cdot n \left( \{\!\{ \sigma \}\!\} - A_{\mathrm{av}} \sigma \right). \tag{5.3.8}$$

Let $F \in \mathcal{F}^i$ and denote $\widetilde{\omega}_F$ the union of all triangle touching $F$. Since $\sigma \in CR$, Lemma 2.2.2 entails

$$\| \{\!\{ \sigma \}\!\} - A_{\mathrm{av}} \sigma \|_{L^2(F)} \leq |F|^{\frac{1}{2}} \| \nabla_{\mathcal{M}} \sigma \|_{L^\infty(F)} \lesssim |F|^{\frac{1}{2}} \| \nabla_{\mathcal{M}} \sigma \|_{L^2(\widetilde{\omega}_F)}.$$

Moreover, given any polynomial $p \in \mathbb{P}_1(\omega_F)$, we derive

$$\begin{aligned}\| [\![\nabla \Pi u]\!] \|_{L^2(F)} = \| [\![\nabla (\Pi u - p)]\!] \|_{L^2(F)} &\lesssim \big(\gamma_{\mathcal{M}} |F|\big)^{-\frac{1}{2}} \|\nabla(\Pi u - p)\|_{L^2(\omega_F)} \\ &\lesssim \big(\gamma_{\mathcal{M}} |F|\big)^{-\frac{1}{2}} \|\nabla(u - p)\|_{L^2(\omega_F)},\end{aligned}$$

where we used $\|\nabla(\Pi u - p)\|^2_{L^2(\omega_F)} = \|\nabla(u-p)\|^2_{L^2(\omega_F)} - \|\nabla(\Pi u - u)\|^2_{L^2(\omega_F)}$ in the last step. Combining this inequalities with the identity (5.3.8) yields then the desired bound. □

Arguing as in the second Strang lemma [10] then leads to the following regularity-free error localization to pairs of elements.

**Theorem 5.3.4** (Error bound with plain smoothing)**.** *Let $u$ be the weak solution of the Poisson problem and denote by $U$ the corresponding approximation generated by the method $(CR, b_{CR}, A^{\star}_{\mathrm{av}})$. We have*

$$\|u - U\|_\mu \le C_{\gamma_{\mathcal{M}}} \left( \sum_{F \in \mathcal{F}^i} \inf_{p \in \mathbb{P}_1(\omega_F)} \|\nabla(u-p)\|^2_{L^2(\omega_F)} \right)^{\frac{1}{2}}.$$

*Proof.* Thanks to the optimality of $\Pi u$, we can decompose the error as follows: $\|u - U\|^2_\mu = \|u - \Pi u\|^2_\mu + \|\Pi u - U\|^2_\mu$. Then, it is sufficient to note $\|\Pi u - U\|_\mu = \sup_{\sigma \in CR, \|\sigma\|_\mu = 1} b_{CR}(\Pi u - U, \sigma)$ and

$$b_{CR}(\Pi u - U, \sigma) = b_0(\Pi u, \sigma) - \int_\Omega \nabla u \cdot \nabla A_{\mathrm{smt}} \sigma.$$

and apply Proposition 5.3.3. □

The difference between this upper bound and that one in (5.3.7) is subtle; see [59, §6.1] for a theoretical discussion in as slightly different setting. Here we illustrate it by approximating the exact solution

$$u^{\mathrm{kink}}(x,y) := \min\{1 - |x|, 1 - |y|\}, \quad (x,y) \in \Omega := (-1,1)^2 \tag{5.3.9}$$

by the aforementioned Crouzeix-Raviart methods, where the initial triangulation $\mathcal{M}_0$ is given by drawing the two diagonals of $\Omega$ and we perform uniform refinements. In view of the kinks, we have $u^{\mathrm{kink}} \in H^2(\mathcal{M}_0)$ and $u^{\mathrm{kink}} \notin H^2(\Omega)$. Hence, we expect that the error $\|u - U^{CR}_{\mathrm{av}}\|_\mu$ decays at the maximum rate, fully exploiting the piecewise regularity of $u^{\mathrm{kink}}$. Instead, the rate of convergence of $\|u - U\|_\mu$ depends on the global regularity of $u^{\mathrm{kink}}$ on $\Omega$ and is not the maximum possible. Figure 5.8 corroborates this.

Notice also that the classical Crouzeix-Raviart method (5.3.2) could not be applied in this case, because the weak Laplacian of $u^{\mathrm{kink}}$ involves first-order moments on the interior edges of $\mathcal{M}_0$.

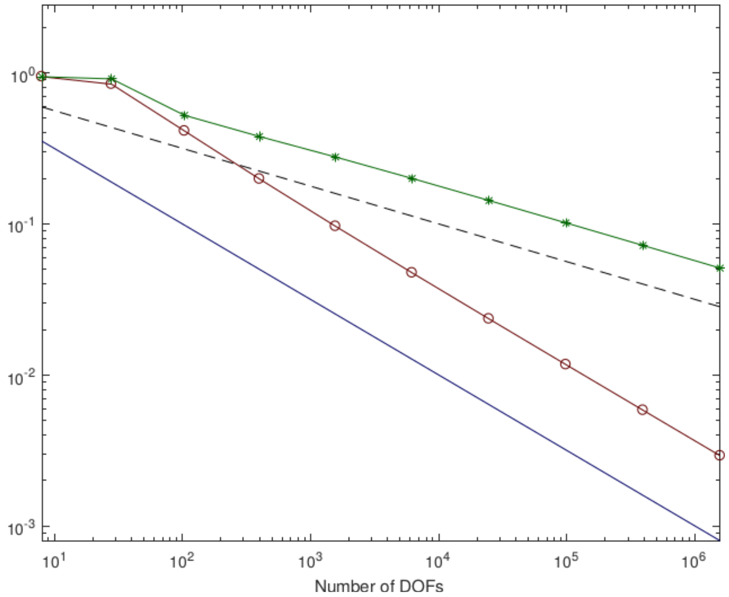


Figure 5.8: Example $u^{\mathrm{kink}}$: Convergence histories of Crouzeix-Raviart error with moment-conserving smoothing $E_{\mathrm{av}}$ ($\circ$) and bare averaging $A_{\mathrm{av}}$ ($*$). Plain and dashed lines indicate, respectively, decay rates #DOFs$^{-0.5}$ and #DOFs$^{-0.25}$.

## 5.4 Combining Discrete Laplacians and Smoothing

In §5.2 we have presented various options to discretize the Laplace operator. The goal of this section is to illustrate their interplay with the mean-preserving smoothing $E_{\mathrm{smt}}$. Doing so, we consider also the conforming Galerkin method $M^{\mathrm{cG}} := (S^{\mathrm{cG}}, b^{\mathrm{cG}}, \mathrm{Id}_{S^{\mathrm{cG}}})$ with discrete problem

$$\text{find} \quad U^{\mathrm{cG}} \in S^{\mathrm{cG}} \quad \text{such that} \quad \forall \sigma \in S_1^1 \quad b^{\mathrm{cG}}(U^{\mathrm{cG}}, \sigma) = \langle f, \sigma \rangle,$$

where $S^{\mathrm{cG}} := S_1^1$ and $b^{\mathrm{cG}} = b_0$. Furthermore, in the numerical experiments, we restrict to the smoothing operator $E_{\mathrm{av}}$ with standard nodal averaging.

The next theorem states that all methods introduced in §5.2 and $M^{\mathrm{cG}}$ are quasi-optimal with respect to their shape functions and provides a quantitative control of the best constant $D$ in (5.1.3).

**Theorem 5.4.1** (Uniform $S_1^0$-quasi-optimality)**.** *Let*

$$\mathrm{op} \in \{\mathrm{cG}, CR, \mathrm{wnip}, \mathrm{wsip}, \mathrm{nip}, \mathrm{sip}\} \quad \textit{and} \quad \mathrm{smt} \in \{\mathrm{av}, \mathrm{sz}\}$$

*and assume that, for* $\mathrm{op} \in \{\mathrm{wsip}, \mathrm{sip}\}$*, the penalty parameter* $\mu$ *is so large that* $b^{\mathrm{op}}$ *is* $\|\cdot\|_\mu$*-coercive. Then the quasi-optimality constant of the method*

$M^{\mathrm{op}}_{\mathrm{smt}}$ *with respect to its shape functions is bounded in terms of* $\gamma_{\mathcal{M}}$ *and the penalty parameter* $\mu$*, if present:*

$$D^{\mathrm{op}}_{\mathrm{smt}} \leq C_{\gamma_{\mathcal{M}},\mu}.$$

*Proof.* We start with the continuous Galerkin method op = cG, where the choice of smt is irrelevant in view of $E_{\mathrm{smt}|S^1_1} = \mathrm{Id}_{S^1_1}$. Cea's lemma and

$$\forall u \in H^1_0(\Omega) \quad \inf_{s\in S^1_1} \|u-s\|_\mu \leq C_{\gamma_{\mathcal{M}}} \inf_{s\in S^0_1} \|u-s\|_\mu,$$

which follows from [60, Lemma 3.1], then readily yields $D^{\mathrm{cG}} \leq C_{\gamma_{\mathcal{M}}}$. Next, Theorems 3.3.3, 4.2.7, 4.2.10 and 4.2.14 entail that

$$\forall u \in H^1_0(\Omega) \qquad \|u - U^{\mathrm{op}}_{\mathrm{smt}}\|_\mu \leq C_{\gamma_{\mathcal{M}}} \inf_{s\in S^{\mathrm{op}}} \|u-s\|_\mu$$

for op $\neq$ cG and smt = sz. A similar estimate can be also obtained for smt = sz in the vein of Remark 2.2.4. Since $S^{\mathrm{op}} = S^0_1$ or $S^{\mathrm{cG}} = CR$, we conclude by invoking (5.2.10). □

Let us take, numerically, a closer look at the relationship between the error of the various approximate solutions and the best error in $S^0_1$, which can be computed via Lemma 5.2.2. For this purpose, we fix

$$\mu = 12, \tag{5.4.1}$$

cf. [38, Remark 12], and introduce the ratio

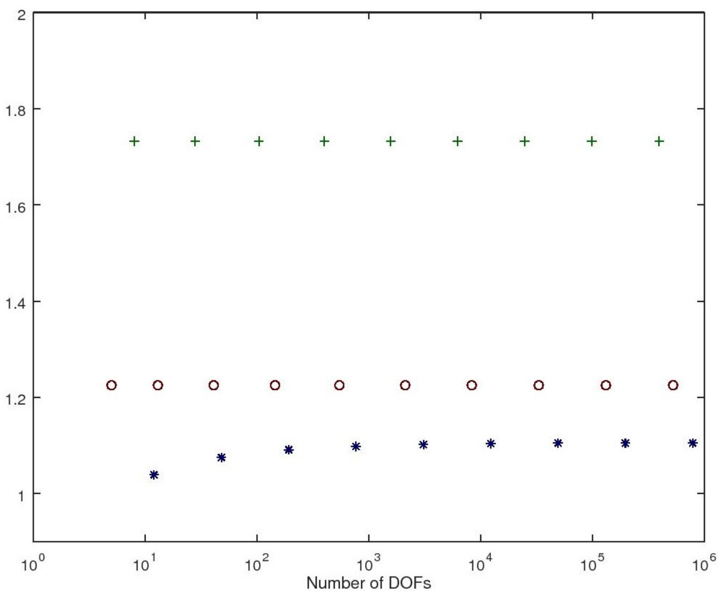
$$q^{\mathrm{op}}_{\mathrm{smt}}(u) := \frac{\|u - U^{\mathrm{op}}_{\mathrm{smt}}\|_\mu}{\inf_{s\in S^0_1} \|u-s\|_\mu} \leq D^{\mathrm{op}}_{\mathrm{smt}}, \tag{5.4.2}$$

where $u \in H^1_0(\Omega) \setminus S^1_1$ is an exact solution of the Poisson problem (3.3.1). Let us consider first the harmonic and regular solution

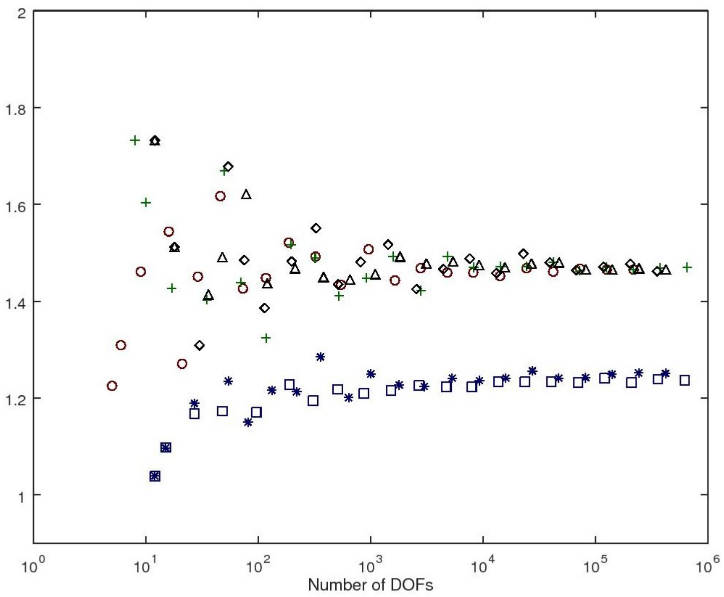
$$u^{\mathrm{hreg}}(x) := r^2 \sin(2\varphi), \quad x = r(\sin\varphi, \cos\varphi) \in \Omega := (0,1)^2.$$

Note that here the smoothing operator $E_{\mathrm{av}}$ is inactive. The ratios displayed in Figure 5.9 favor the SIP and NIP methods. Furthermore, as the stability constant in Figure 5.7, they are affected by local non-symmetries in the mesh, although the size of this effect fades away asymptotically.

To investigate a case with active smoothing operator, we reconsider $u^{\mathrm{rsym}}_\rho$ from (5.3.3) for the regular case $\rho = 1.9$ and the singular case $\rho = 0.25$; see Figures 5.10-5.11. Here the slightly better exploitation of $S^0_1$ by the SIP and NIP methods is no longer present. The asymptotic exploitation of all methods is similar, with a slight advantage for the nonconforming methods in the singular case with adaptive refinement. In order to generate the

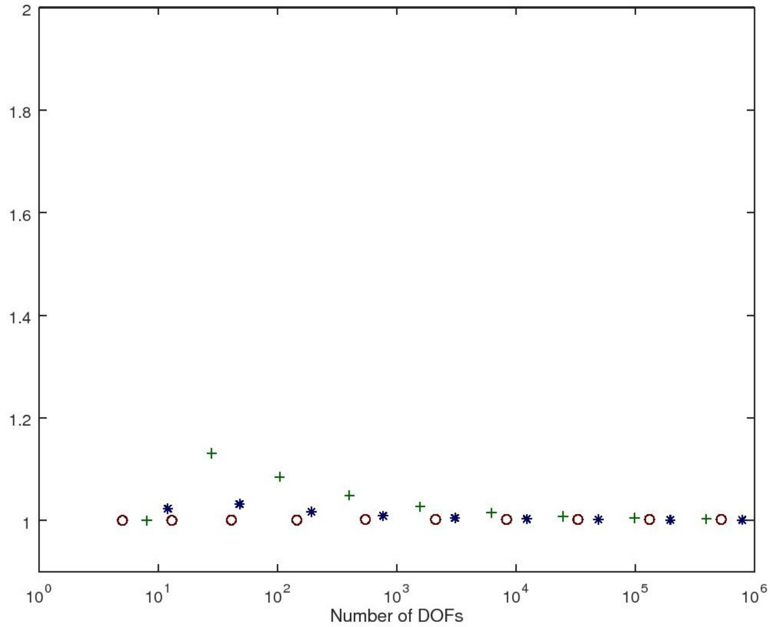


(a) Uniform refinement

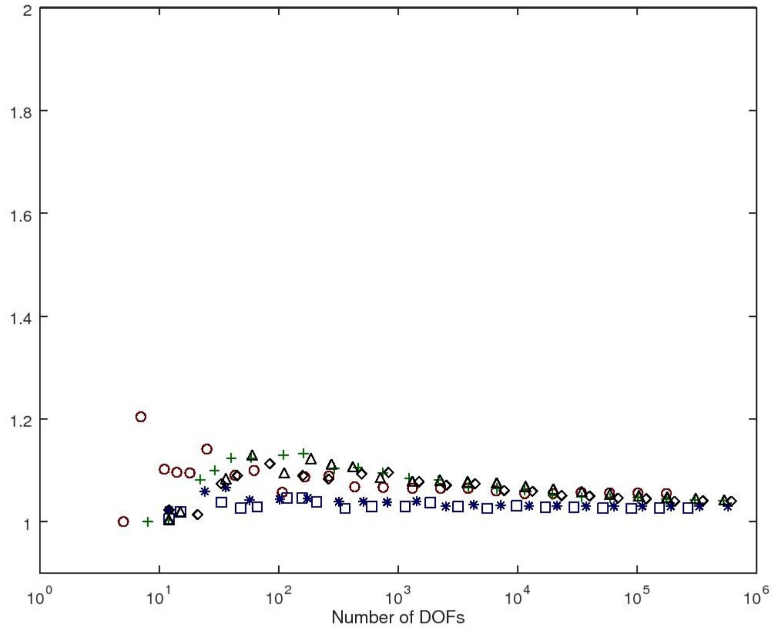


(b) Random refinement

Figure 5.9: Example $u^{\mathrm{hreg}}$: Ratios $q_{\mathrm{av}}^{\mathrm{op}}(u^{\mathrm{hreg}})$ from (5.4.2) versus #DOFs for op =cG ($\circ$), CR ($+$), SIP ($*$), NIP ($\Box$), WNIP ($\triangle$), and WSIP ($\diamond$). For uniform refinement, the WIP and NIP methods are not displayed because they graphically coincide with the CR and SIP methods, respectively.

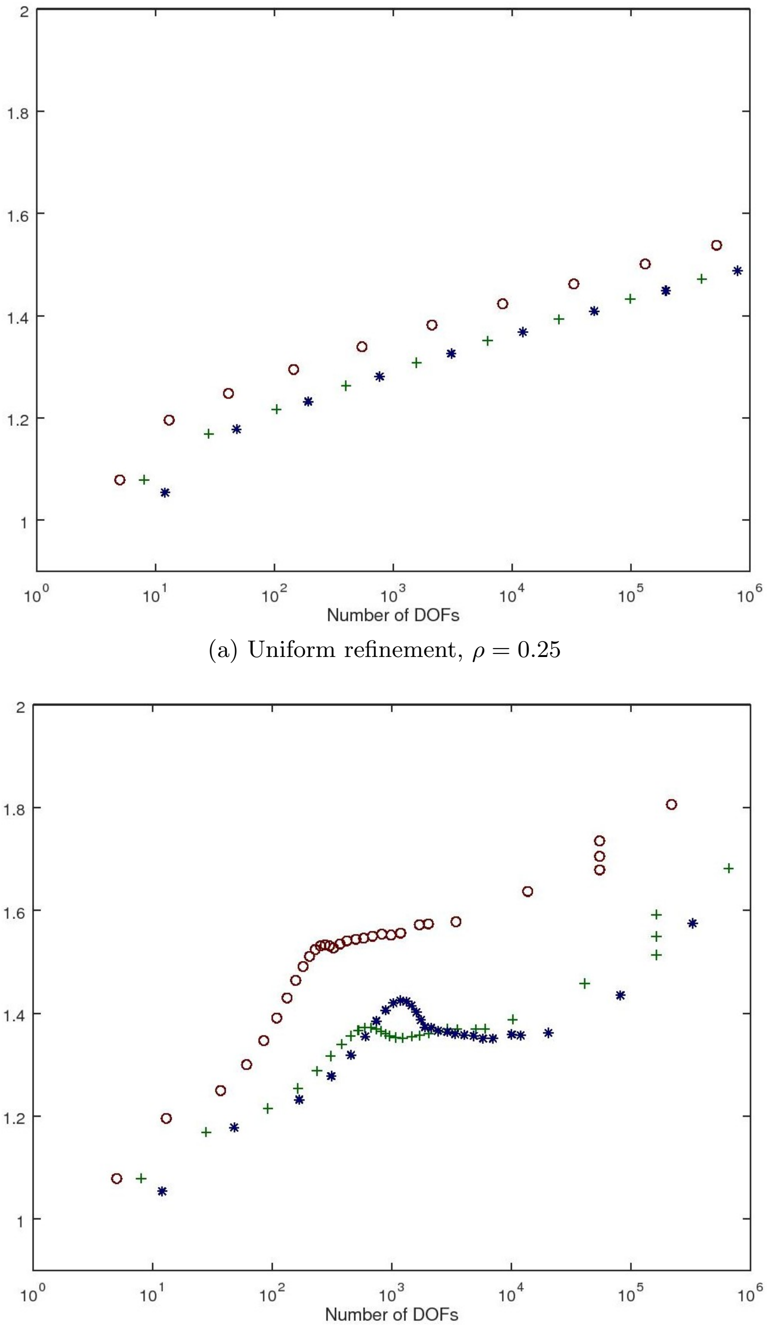


(a) Uniform refinement, $\rho = 1.9$



(b) Random refinement, $\rho = 1.9$

Figure 5.10: Example $u_{1.9}^{\mathrm{rsym}}$: Ratios $q_{\mathrm{av}}^{\mathrm{op}}(u_{1.9}^{\mathrm{rsym}})$ from (5.4.2) versus #DOFs for op =cG ($\circ$), CR ($+$), SIP ($*$), NIP ($\Box$), WNIP ($\triangle$), and WSIP ($\diamond$) with smoother $E_{\mathrm{av}}$. For uniform refinement, the SIP method is representative for the other interior penalty methods.



(a) Uniform refinement, $\rho = 0.25$

(b) Adaptive refinement, $\rho = 0.25$

Figure 5.11: Example $u_{0.25}^{\text{rsym}}$: Ratios $q_{\text{av}}^{\text{op}}(u_{0.25}^{\text{rsym}})$ from (5.4.2) versus #DOFs for op =cG ($\circ$), CR ($+$), SIP ($*$) with smoother $E_{\text{av}}$. The SIP method is representative for the other interior penalty methods.

adaptive meshes, we applied Dörfler's strategy with parameter 0.9 on the local best errors $\|\nabla(u - \Pi u)\|_{L^2(K)}$, $K \in \mathcal{M}$.

A consequence of Theorem 5.4.1 is the following comparison. This result is similar to those ones of Braess [12] and Carstensen et al. [27] for nonconforming methods without smoothing. However, thanks to the use of a smoothing operator in the methods under consideration, no oscillation terms are involved.

**Corollary 5.4.2** (Comparison)**.** *Let*

$$\mathrm{op}_1, \mathrm{op}_2 \in \{\mathrm{cG}, \mathit{CR}, \mathrm{wnip}, \mathrm{wsip}, \mathrm{nip}, \mathrm{sip}\} \quad \textit{and} \quad \mathrm{smt}_1, \mathrm{smt}_2 \in \{\mathrm{av}, \mathrm{sz}\}$$

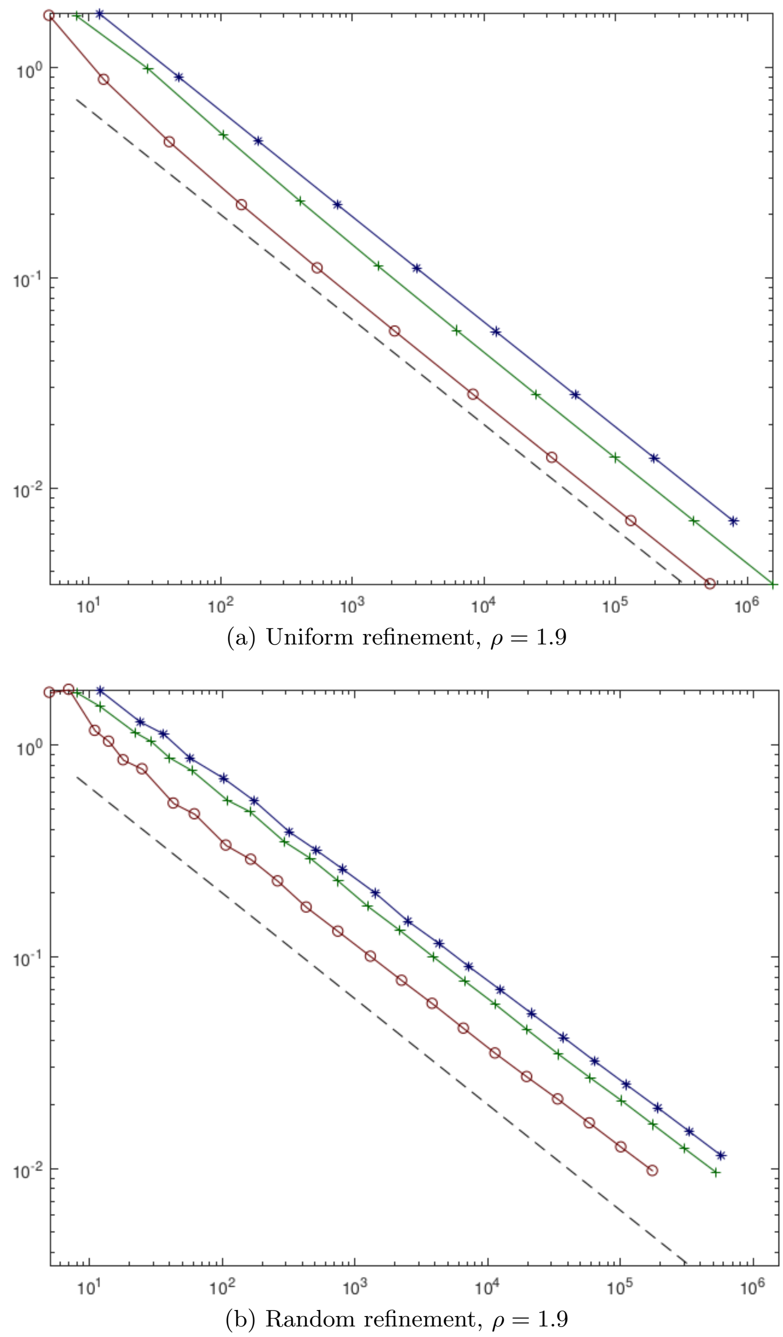
*and assume that $\mu \geq \mu_*$ for some parameter $\mu_* > 0$ so large that the forms $b^{\mathrm{op}_1}$ and $b^{\mathrm{op}_2}$ are $\|\cdot\|_\mu$-coercive. Then*

$$\forall u \in H_0^1(\Omega) \quad \|u - U^{\mathrm{op}_1}_{\mathrm{smt}_1}\|_\mu \approx \|u - U^{\mathrm{op}_2}_{\mathrm{smt}_2}\|_\mu,$$

*where the hidden constants depend only on the shape coefficient $\gamma_{\mathcal{M}}$ of the underlying mesh.*

Of course, this result regards only the error of the considered methods, and not the trade-off between error and cost for obtaining the approximate solution. This trade-off is more important, but also more delicate. In particular, a realistic, implementation-independent measure of the cost is not obvious, because it should take into account the cost for (efficiently) building and solving the corresponding linear systems. Here we shall use the admittedly coarse measure #DOFs for this purpose.

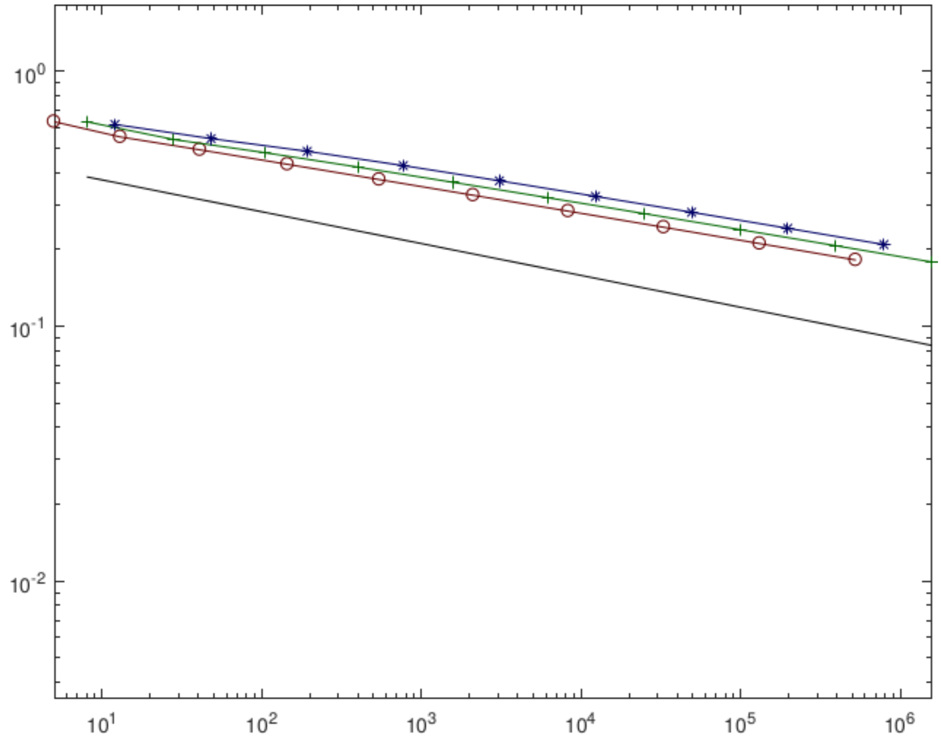
Figures 5.12-5.13 provide the concrete trade-offs for the examples in Figures 5.10-5.11. In the presented cases, increasing nonconformity worsens the trade-off, but in a moderate amount. This indicates that the greater flexibility of nonconforming methods, which is important in more complex problems, comes with a comparable trade-off in terms of #DOFs.



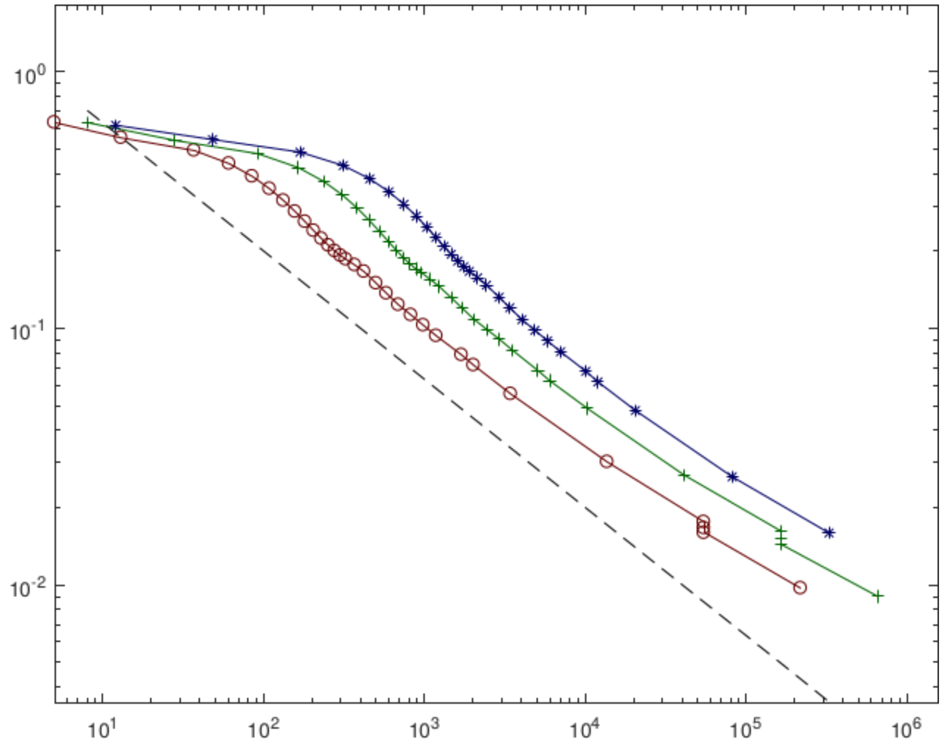
(a) Uniform refinement, $\rho = 1.9$

(b) Random refinement, $\rho = 1.9$

Figure 5.12: Example $u_{1.9}^{\mathrm{rsym}}$: Trade-off between error and #DOFs in log-log scale for cG ($\circ$), CR ($+$), and SIP ($*$) with smoother $E_{\mathrm{av}}$. The SIP method with $E_{\mathrm{av}}$ is representative for the other interior penalty methods. Dashed lines indicate the decay rate $\#\mathrm{DOFs}^{-0.5}$.



(a) Uniform refinement, $\rho = 0.25$



(b) Adaptive refinement, $\rho = 0.25$

Figure 5.13: Example $u_{0.25}^{\mathrm{rsym}}$: Trade-off between error and #DOFs in log-log scale for cG ($\circ$), CR ($+$), and SIP ($*$) with smoother $E_{\mathrm{av}}$. The SIP method with $E_{\mathrm{av}}$ is representative for the other interior penalty methods. Plain and dashed line indicate, respectively, decay rate #DOFs$^{-0.125}$ and #DOFs$^{-0.5}$.

# Bibliography

[1] D. N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19(4):742–760, 1982.

[2] D. N. Arnold. On nonconforming linear-constant elements for some variants of the Stokes equations. *Istit. Lombardo Accad. Sci. Lett. Rend. A*, 127(1), 1993.

[3] D. N. Arnold. Stability, consistency, and convergence of numerical discretizations. *Encyclopedia of Applied and Computational Mathematics*, B. Engquist, ed., Springer:1358–1364, 2015.

[4] D. N. Arnold and F. Brezzi. Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates. *RAIRO Modél. Math. Anal. Numér.*, 19(1):7–32, 1985.

[5] D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5):1749–1779, 2001/02.

[6] I. Babuška. Error-bounds for finite element method. *Numer. Math.*, 16:322–333, 1970/1971.

[7] I. Babuška and M. Suri. On locking and robustness in the finite element method. *SIAM J. Numer. Anal.*, 29(5):1261–1293, 1992.

[8] S. Badia, R. Codina, T. Gudi, and J. Guzmán. Error analysis of discontinuous Galerkin methods for the Stokes problem under minimal regularity. *IMA J. Numer. Anal.*, 34(2):800–819, 2014.

[9] G. A. Baker. Finite element methods for elliptic equations using nonconforming elements. *Mathematics of Computation*, 31(137):45–59, 1977.

[10] A. Berger, R. Scott, and G. Strang. Approximate boundary conditions in the finite element method. *Symposia Mathematica, Vol. X (Convegno di Analisi Numerica, INDAM, Rome)*, pages 295–313, 1972.

[11] D. Boffi, F. Brezzi, and M. Fortin. *Mixed finite element methods and applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2013.

[12] D. Braess. An a posteriori error estimate and a comparison theorem for the nonconforming $P_1$ element. *Calcolo*, 46(2):149–155, 2009.

[13] S. C. Brenner. A two-level additive Schwarz preconditioner for nonconforming plate elements. *Numer. Math.*, 72(4):419–447, 1996.

[14] S. C. Brenner. Poincaré-Friedrichs inequalities for piecewise $H^1$ functions. *SIAM J. Numer. Anal.*, 41(1):306–324, 2003.

[15] S. C. Brenner. Korn's inequalities for piecewise $H^1$ vector fields. *Math. Comp.*, 73(247):1067–1087, 2004.

[16] S. C. Brenner. Forty years of the Crouzeix-Raviart element. *Numer. Methods Partial Differential Equations*, 31(2):367–396, 2015.

[17] S. C. Brenner, T. Gudi, and L.Y. Sung. An a posteriori error estimator for a quadratic $C^0$-interior penalty method for the biharmonic problem. *IMA J. Numer. Anal.*, 30(3):777–798, 2010.

[18] S. C. Brenner and L. Owens. A weakly over-penalized non-symmetric interior penalty method. *JNAIAM J. Numer. Anal. Ind. Appl. Math.*, 2(1-2):35–48, 2007.

[19] S. C. Brenner, L. Owens, and L.-Y. Sung. Higher order weakly over-penalized symmetric interior penalty methods. *Journal of Computational and Applied Mathematics*, 236(11):2883–2894, 2012.

[20] S. C. Brenner, L. Owens, and L.Y. Sung. A weakly over-penalized symmetric interior penalty method. *Electron. Trans. Numer. Anal.*, 30:107–127, 2008.

[21] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[22] S. C. Brenner and L. Y. Sung. Linear finite element methods for planar linear elasticity. *Math. Comp.*, 59(200):321–338, 1992.

[23] S. C. Brenner and L. Y. Sung. $C^0$ interior penalty methods for fourth order elliptic boundary value problems on polygonal domains. *J. Sci. Comput.*, 22/23:83–118, 2005.

[24] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[25] D. Buckholtz. Hilbert space idempotents and involutions. *Proc. Amer. Math. Soc.*, 128(5):1415–1418, 2000.

[26] E. Burman and A. Ern. Continuous interior penalty *hp*-finite element methods for advection and advection-diffusion equations. *Mathematics of computation*, 76(259):1119–1140, 2007.

[27] C. Carstensen, D. Peterseim, and M. Schedensack. Comparison results of finite element methods for the Poisson model problem. *SIAM J. Numer. Anal.*, 50(6):2803–2823, 2012.

[28] J. Céa. Approximation variationnelle des problèmes aux limites. *Ann. Inst. Fourier (Grenoble)*, 14(fasc. 2):345–444, 1964.

[29] P. G. Ciarlet. *The finite element method for elliptic problems.* North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.

[30] P. G. Ciarlet. Interpolation error estimates for the reduced Hsieh-Clough-Tocher triangle. *Math. Comp.*, 32(142):335–344, 1978.

[31] P. G. Ciarlet. Basic error estimates for elliptic problems. In *Handbook of numerical analysis, Vol. II*, Handb. Numer. Anal., II, pages 17–351. North-Holland, Amsterdam, 1991.

[32] P. Ciarlet Jr., C. F. Dunkl, and S.A Sauter. A family of Crouzeix-Raviart finite elements in 3D. *arXiv:1703.03224v1*, 2017.

[33] R. W. Clough and J. L. Tocher. Finite element stiffness matrices for analysis of plates in bending. *Proceedings Conference on Matrix Methods in Structural Mechanics*, Wright Patterson A.F.B. Ohio:515–545, 1965.

[34] M. Crouzeix and R. S. Falk. Nonconforming finite elements for the Stokes problem. *Math. Comp.*, 52(186):437–456, 1989.

[35] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations. I. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, 7(R-3):33–75, 1973.

[36] C. de Boor and R. DeVore. Approximation by smooth multivariate splines. *Trans. Amer. Math. Soc.*, 276(2):775–788, 1983.

[37] D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathématiques & Applications (Berlin) [Mathematics & Applications].* Springer, Heidelberg, 2012.

[38] Y. Epshteyn and B. Rivière. Estimation of penalty parameters for symmetric interior penalty Galerkin methods. *J. Comput. Appl. Math.*, 206(2):843–872, 2007.

[39] A. Ern and J.-L. Guermond. Finite element quasi-interpolation and best approximation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 51(4):1367–1385, 2017.

[40] R. S. Falk. Nonconforming finite element methods for the equations of linear elasticity. *Math. Comp.*, 57(196):529–550, 1991.

[41] M. Fortin. A three-dimensional quadratic nonconforming element. *Numer. Math.*, 46(2):269–279, 1985.

[42] M. Fortin and M. Soulie. A nonconforming piecewise quadratic finite element on triangles. *Internat. J. Numer. Methods Engrg.*, 19(4):505–520, 1983.

[43] T. Gudi. A new error analysis for discontinuous finite element methods for linear elliptic problems. *Math. Comp.*, 79(272):2169–2189, 2010.

[44] P. Hansbo and M. G. Larson. Discontinuous Galerkin and the Crouzeix-Raviart element: application to elasticity. *M2AN Math. Model. Numer. Anal.*, 37(1):63–72, 2003.

[45] K.-J. Heine, D. Köster, O. Kriessl, A. Schmidt, and K. Siebert. Alberta - an adaptive hierachical finite element toolbox. http://www.alberta-fem.de.

[46] O. A. Karakashian and F. Pascal. A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems. *SIAM J. Numer. Anal.*, 41(6):2374–2399, 2003.

[47] M. G. Larson and A. J. Niklasson. Analysis of a nonsymmetric discontinuous galerkin method for elliptic problems: stability and energy error estimates. *SIAM journal on numerical analysis*, 42(1):252–264, 2004.

[48] A. Linke. On the role of the Helmholtz decomposition in mixed methods for incompressible flows and a new variational crime. *Comput. Methods Appl. Mech. Engrg.*, 268:782–800, 2014.

[49] J. Morgan and R. Scott. A nodal basis for $C^1$ piecewise polynomials of degree $n \geq 5$. *Math. Comput.*, 29:736–740, 1975.

[50] L. S. D. Morley. The triangular equilibrium element in the solution of plate bending problems. *Aeronautical Quarterly*, 19(02):149–169, 1968.

[51] P. Oswald. On a BPX-preconditioner for P1 elements. *Computing*, 51(2):125–133, 1993.

[52] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.

[53] B. Rivière, M. F. Wheeler, and V. Girault. A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems. *SIAM J. Numer. Anal.*, 39(3):902–931, 2001.

[54] A. Schmidt and K. G. Siebert. *Design of adaptive finite element software*, volume 42 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2005. The finite element toolbox ALBERTA, With 1 CD-ROM (Unix/Linux).

[55] L. R. Scott and M. Vogelius. Conforming finite element methods for incompressible and nearly incompressible continua. In *Large-scale computations in fluid mechanics, Part 2 (La Jolla, Calif., 1983)*, volume 22 of *Lectures in Appl. Math.*, pages 221–244. Amer. Math. Soc., Providence, RI, 1985.

[56] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[57] G. Stoyan and Á. Baran. Crouzeix-Velte decompositions for higher-order finite elements. *Comput. Math. Appl.*, 51(6-7):967–986, 2006.

[58] F. Tantardini and A. Veeser. The $L^2$-projection and quasi-optimality of Galerkin methods for parabolic equations. *SIAM J. Numer. Anal.*, 54(1):317–340, 2016.

[59] F. Tantardini, A. Veeser, and R. Verfürth. Robust localization of the best error with finite elements in the reaction-diffusion norm. *Constr. Approx.*, 42(2):313–347, 2015.

[60] A. Veeser. Approximating gradients with continuous piecewise polynomial functions. *Found. Comput. Math.*, 16(3):723–750, 2016.

[61] A. Veeser and R. Verfürth. Explicit upper bounds for dual norms of residuals. *SIAM J. Numer. Anal.*, 47(3):2387–2405, 2009.

[62] A. Veeser and P. Zanotti. Applying quasi-optimal nonconforming methods to the Poisson problem. In preparation.

[63] A. Veeser and P. Zanotti. Quasi-optimal nonconforming methods for symmetric elliptic problems. I – Abstract theory. arXiv:1710.03331 [math.NA].

[64] A. Veeser and P. Zanotti. Quasi-optimal nonconforming methods for symmetric elliptic problems. II – Overconsistency and classical nonconforming elements. arXiv:1710.03447 [math.NA].

[65] A. Veeser and P. Zanotti. Quasi-optimal nonconforming methods for symmetric elliptic problems. III – DG and other interior penalty methods. arXiv:1710.03452 [math.NA].

[66] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques*. Adv. Numer. Math., John Wiley & Sons Inc, Chichester, UK, 1996.

[67] M. F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15(1):152–161, 1978.

[68] J. Xu and L. Zikatanov. Some observations on Babuška and Brezzi theories. *Numer. Math.*, 94(1):195–202, 2003.