Alexandre Ern
Jean-Luc Guermond

Finite Elements I

Approximation and Interpolation

Springer

Texts in Applied Mathematics

Volume 72

The mathematization of all sciences, the fading of traditional scientific boundaries, the impact of computer technology, the growing importance of computer modelling and the necessity of scientific planning all create the need both in education and research for books that are introductory to and abreast of these developments. The aim of this series is to provide such textbooks in applied mathematics for the student scientist. Books should be well illustrated and have clear exposition and sound pedagogy. Large number of examples and exercises at varying levels are recommended. TAM publishes textbooks suitable for advanced undergraduate and beginning graduate courses, and complements the Applied Mathematical Sciences (AMS) series, which focuses on advanced textbooks and research-level monographs.

More information about this series at http://www.springer.com/series/1214

Alexandre Ern · Jean-Luc Guermond

Finite Elements I

Approximation and Interpolation

Alexandre Ern
CERMICS
Ecole des Ponts and INRIA Paris
Marne la Vallée and Paris, France

Jean-Luc Guermond
Department of Mathematics
Texas A&M University
College Station, TX, USA

Preface

Although the roots of the "Finite Element Method" can be found in the work of Courant [84], the method really took off in the 1950's when engineers started to solve numerically structural mechanics problems in civil engineering and in aeronautics. Since then, finite elements have become ubiquitous in computational sciences and engineering. Numerous academic toolbox and commercial codes based on the finite element method have been developed over the years and are now available to a large public. Numerous books, textbooks, and myriads of technical papers, articles, and conference proceedings have been written on the topic.

We have contributed to this flow in 2004 by publishing *Theory and Practice of Finite Elements*, in the Applied Mathematical Sciences series, volume 159. The approach we adopted at that time was first to present the finite element method as an interpolation tool, then to illustrate the idea that finite elements can be efficiently used to approximate partial differential equations other than the Laplace equation and in particular problems for which the Lax–Milgram lemma is not the ultimate paradigm. One objective of *Theory and Practice of Finite Elements* was to put the emphasis on the *infsup conditions* developed by Babuška in 1970 in the context of finite element methods [14] and stated in a theoretical work by Nečas in 1962 [150]. These inf-sup conditions are necessary and sufficient conditions for the well-posedness of any linear problem set in Banach spaces. From the functional analysis point of view, the inf-sup conditions are a rephrasing of two fundamental theorems by Banach: the closed range theorem and the open mapping theorem. For this reason, we called these conditions the *Banach NečasBabuška* (BNB) theorem. The idea we followed in *Theory and Practice of Finite Elements* was to expose fundamental concepts while staying connected with practical topics such as applications to several PDEs and implementation aspects of the finite element method.

The present work, called *Finite Elements* and organized in three volumes, started as a second edition of *Theory and Practice of Finite Elements* at the invitation of Springer editors, but as we progressed in the rewriting and the reorganizing of the material, an entirely new project emerged. We tried to preserve the spirit of *Theory and Practice of Finite Elements* by covering fundamental aspects in approximation theory and by thoroughly exploring applications and implementation details, but **Finite Elements** is definitely not a re-edition of *Theory and Practice of Finite Elements*. This new book is meant to be used as a graduate textbook and as a reference for researchers and engineers.

The book is divided into three volumes. Volume I focuses on fundamental ideas regarding the construction of finite elements and their approximation properties. We have decided to start Volume I with four chapters on functional analysis which we think could be useful to readers who may not be familiar with Lebesgue integration and weak derivatives. The purpose of these chapters is not to go through arduous technical details, but to familiarize the reader with the functional analysis language. These four chapters are packed with examples and counterexamples which we think should convince the reader of the relevance of the material. Volume I also reviews important implementation details that must be taken care of when either developing or using a finite element toolbox, like the orientation of meshes, and the enumeration of the geometric entities (vertices, edges, faces, cells) or the enumeration of the degrees of freedom. Volume I contains two appendices highlighting basic facts on Banach and Hilbert spaces and on differential calculus.

Volume II starts with fundamental results on well-posed weak formulations and their approximation by the Galerkin method. Key results are the BNB Theorem, Céa's and Strang's lemmas (and their variants) for the error analysis, and the duality argument by Aubin and Nitsche. Important implementation aspects regarding quadratures, linear algebra, and assembling are also covered. The rest of Volume II focuses on applications to PDEs where a coercivity property is available. Various conforming and nonconforming approximation techniques are exposed (Galerkin, boundary penalty, Crouzeix–Raviart, discontinuous Galerkin, hybrid high-order methods). The applications considered are elliptic PDEs (diffusion, elasticity, the Helmholtz problem, Maxwell's equations), eigenvalue problems for elliptic PDEs, and PDEs in mixed form (Darcy and Stokes flows). Volume II contains one appendix collecting fundamental results on the surjectivity, bijectivity, and coercivity of linear operators in Banach spaces.

Volume III develops more advanced topics. The first quarter of the volume focuses on symmetric positive systems of first-order PDEs called Friedrichs' systems. Examples include advection and advection-diffusion equations and various PDEs written in mixed form (Darcy and Stokes flows, Maxwell's equations). One salient aspect of this first part of the volume is the comprehensive and unified treatment of many stabilization techniques from the

literature. The remaining of Volume III deals with time-dependent problems: parabolic equations (such as the heat equation), evolution equations without coercivity (Stokes flows, Friedrichs' systems), and nonlinear hyperbolic equations (scalar conservation equations, hyperbolic systems).

The book is organized into 83 chapters, most of them composed of 10–16 pages, and each chapter is accompanied by exercises. The three volumes contain altogether over 500 exercises with all the solutions available online. For researchers and engineers, the division in short chapters is meant to isolate the key ideas and the most important results. The chapters are relatively independent from each other and the book is not meant to be read linearly. Each volume is supplemented with a long list of references. In order to help the reader, we try to pinpoint the exact chapter, section or theorem each time we refer to a book. Short literature reviews are also included in most of the chapters as well.

When used as a textbook, the division in short chapters is meant to be an aid to teachers and students. The objective is that one chapter can be covered in time units of 1h to 1h15. The salient ideas can be developed and exposed in class, while the rest of the chapter can be assigned as reading material. The exercises are important complements, and teachers are encouraged to use some of the exercises in class. Whenever possible the exercises have been divided into elementary steps with enough hints to be doable by reasonably assiduous students. The book is well adapted to graduate flipped classes as well. A significant portion of the material presented in the book has been taught in graduate classes at Texas A&M, École nationale des ponts et chaussées, École polytechnique, and Institut Henri Poincaré. About one third of the material has actually been taught by the students themselves in flipped classes. The book can be used in many teaching contexts. Among various possibilities it can be used to teach the mathematical bases of finite elements at an introductory level, it can also be used to teach practical implementation aspects (mesh generation, enumeration, orientation, quadratures, assembling), and it can be used to teach sophisticated approximation techniques over a wide range of problems (elliptic PDEs, mixed PDEs, first-order PDEs, eigenvalue problems, parabolic PDEs, hyperbolic conservation equations).

A good part of the material is quite standard, but we have also inserted concepts and ideas which, without being entirely new, will possibly convey some flavor of novelty to the reader. For instance, we have developed in some details and provided examples on how to orient meshes and on how the usual differential operators, as well as normal and tangent vectors, are transformed by geometric mappings. We have developed a step-by-step construction of the usual conforming finite element subspaces by means of the notion of connectivity classes, and we have illustrated this notion by numerous examples. Furthermore, we have included two chapters on quasi-interpolation where we tried to develop a fresh and unifying viewpoint on the construction of quasi-interpolation operators for all the scalar-valued and vector-valued finite elements considered in the book. We have also made an effort to work as much

as possible with dimensionally consistent expressions. Although this may lead to slightly more complex statements for norms and error estimates, we believe that the present choice is important to understand the various physically relevant regimes in the model problems.

Some of the techniques that are used to prove stability and error estimates, without being entirely new, are, in our opinion, not standard in the literature at the time of this writing. In particular, the techniques that we use are essentially designed to invoke as little a priori regularity from the solution as possible. One salient example is the analysis of nonconforming approximation techniques for diffusion problems with contrasted coefficients, and another one concerns Maxwell's equations also in materials with contrasted properties. Moreover, we give a unified analysis of first-order PDEs by means of Friedrichs' systems, and we show that a large class of stabilization methods proposed in the literature so far are more or less equivalent, whether the approximation is continuous or discontinuous. We conclude Volume III by a series of chapters on time-evolution problems, which give a somewhat new perspective on the analysis of well-known time-stepping methods. The last five chapters on hyperbolic equations, we hope, should convince the reader that continuous finite elements are good candidates to solve this class of problems where finite volumes have so far taken the lion's share.

Although the reference list is quite long (about 200, 400, and 300 bibliographic entries in each volume, respectively), the finite element literature is so prolific that we have not been able to cite all the relevant contributions. Anyway, our objective was not to be exhaustive and to write complete reviews of the topics at hand but to isolate the key principles and ideas and to refer the reader to the references we are the most familiar with at the time of the writing.

Paris, France   *Alexandre Ern*
College Station, TX, USA   *Jean-Luc Guermond*
June 2020

Contents

Part IV. Finite element spaces

Part I
Elements of functional analysis

Part I, Chapter 1

Lebesgue spaces

The objective of the four chapters composing Part I is to recall (or gently introduce) some elements of functional analysis that will be used throughout the book: Lebesgue integration, weak derivatives, and Sobolev spaces. We focus in this chapter on Lebesgue integration and Lebesgue spaces. Most of the results are stated without proof, but we include various examples. We refer the reader to Adams and Fournier [3], Bartle [16], Brezis [48], Demengel and Demengel [88], Evans [99], Grisvard [110], Malý and Ziemer [138], Rudin [169, Chap. 11], Rudin [170], Sobolev [180], Tartar [189], Yosida [202].

In this book, d is the space dimension, and D denotes a nonempty subset of $\mathbb{R}^d$. Vectors in $\mathbb{R}^d$, $d \geq 2$, and vector-valued functions are denoted in bold font. We abuse the notation by denoting position vectors in $\mathbb{R}^d$ in bold font as well. Moreover, $\|\cdot\|_{\ell^2(\mathbb{R}^d)}$ denotes the Euclidean norm in $\mathbb{R}^d$ (we write $\|\cdot\|_{\ell^2}$ when the context is unambiguous), and $\boldsymbol{a}\cdot\boldsymbol{b}$ denotes the Euclidean inner product between two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$. For every pair of integers $m \leq n$, we use the notation $\{m{:}n\} := \{p \in \mathbb{N} \mid m \leq p \leq n\}$.

1.1 Heuristic motivation

If one restricts oneself to computational considerations, the Riemann integral is the only notion of integration that is needed in numerical analysis, since the objects that one manipulates in practice are piecewise smooth functions (e.g., polynomials) defined on meshes. However, the Riemann integral becomes useless when one starts to investigate questions like passage to the limit. For instance, assume that one has an interval $I := (a, b)$, a sequence of finite partitions of this interval, say $(I_h)_{h \in \mathcal{H}}$, and a sequence of real-valued functions $(v_h)_{h \in \mathcal{H}}$ defined on I such that v_h is smooth on each subinterval of I_h for all $h \in \mathcal{H}$. Here, $\mathcal{H}$ is a countable set with 0 as unique accumulation point. In the context of finite elements, the index h refers to the size of the mesh that is used to construct the function v_h. Assume also that one can a priori prove

that the sequence $(v_h)_{h\in\mathcal{H}}$ is Cauchy in the following sense: for every $\epsilon > 0$, there is $h(\epsilon)$ such that $\int_a^b |v_{h_1}(x) - v_{h_2}(x)|\,\mathrm{d}x \leq \epsilon$ for all $h_1, h_2 \in \mathcal{H} \cap (0, h(\epsilon))$. One may then wonder whether v_h converges to some object with interesting properties when $h \to 0$. The answer to this question becomes very intricate if one restricts oneself to the Riemann integral, but it becomes simple if one adopts Lebesgue's point of view. Since the above question arises constantly in this book, we now take some time to recall the key ingredients of Lebesgue's theory.

1.2 Lebesgue measure

To define the Lebesgue integral of a function defined on a subset D of $\mathbb{R}^d$, one needs to measure the volume of sets in $\mathbb{R}^d$. For every bounded rectangular parallelepiped $R := [a_1, b_1] \times \cdots \times [a_d, b_d]$, with $a_i \leq b_i$ for all $i \in \{1{:}d\}$, we define the Lebesgue (outer) measure of R to be its volume, i.e., we set $|R| := \prod_{i\in\{1{:}d\}}(b_i - a_i)$.

Definition 1.1 (Lebesgue's outer measure). *Let $\mathcal{R}(\mathbb{R}^d)$ be the set of all the rectangular parallelepipeds in $\mathbb{R}^d$. Let E be a set in $\mathbb{R}^d$. The* Lebesgue's outer measure *of E is defined as*

$$|E|^* := \inf\Big\{ \sum_{i\in\mathbb{N}} |R_i| \mid E \subset \bigcup_{i\in\mathbb{N}} R_i,\ R_i \in \mathcal{R}(\mathbb{R}^d) \Big\}. \tag{1.1}$$

We expect $|E|^*$ to be a reasonable estimate of the volume of E if E is a reasonable set. The outer Lebesgue measure has the following properties: (i) $|\emptyset|^* = 0$; (ii) If $E \subset F$, then $|E|^* \leq |F|^*$; (iii) If $\{E_i\}_{i\in\mathbb{N}}$ is a countable collection of subsets of $\mathbb{R}^d$, then $|\bigcup_{i\in\mathbb{N}} E_i|^* \leq \sum_{i\in\mathbb{N}} |E_i|^*$ (countable subadditivity property; see [169, Thm. 11.8]).

Example 1.2 (Countable sets). The outer Lebesgue measure of a countable set $A := \bigcup_{k\in\mathbb{N}}\{\boldsymbol{x}_k\}$ is zero. Let indeed $\epsilon > 0$. We have $\{\boldsymbol{x}_k\} \subset R(\boldsymbol{x}_k, \epsilon^{\frac{1}{d}})$, where $R(\boldsymbol{z}, r)$ is the cube of side r centered at $\boldsymbol{z}$. Hence, $|\{\boldsymbol{x}_k\}|^* \leq \epsilon$, i.e., $|\{\boldsymbol{x}_k\}|^* = 0$ since $\epsilon > 0$ is arbitrary. Invoking subadditivity yields $|A|^* = 0$. For example, this implies that the outer measure of the set of the rational numbers is zero, i.e., $|\mathbb{Q}|^* = 0$. $\square$

Definition 1.3 (Lebesgue's measure of a set). *A set $E \subset \mathbb{R}^d$ is said to be* Lebesgue-measurable *if $|S|^* = |S \cap E|^* + |S \cap E^c|^*$ for every subset $S \subset \mathbb{R}^d$, where E^c is the complement of E in $\mathbb{R}^d$.*

It turns out that not all the sets of $\mathbb{R}^d$ are Lebesgue-measurable, but the class of Lebesgue-measurable sets (in short, measurable sets) of $\mathbb{R}^d$, say $\mathcal{L}(\mathbb{R}^d)$, is sufficiently vast that we will only encounter measurable sets in this book. In particular, (i) If E is measurable, then E^c is also measurable; (ii) Open sets

of $\mathbb{R}^d$ and closed sets of $\mathbb{R}^d$ are measurable (so that all the usual geometric objects, e.g., parallelepipeds or balls, are measurable); (iii) Countable unions and countable intersections of measurable sets are measurable.

Henceforth, the map $|\cdot| : \mathcal{L}(\mathbb{R}^d) \to [0, \infty]$ such that $|E| := |E|^*$ for all $E \in \mathcal{L}(\mathbb{R}^d)$ is called (d-dimensional) *Lebesgue measure*. Since the action of the Lebesgue measure on measurable sets is simply the outer Lebesgue measure, we infer that (i) $|\emptyset| = 0$; (ii) If $A, B \in \mathcal{L}(\mathbb{R}^d)$ and $A \subset B$, then $|A| \leq |B|$; (iii) The countable subadditivity property holds true on countable collections of measurable sets. By restricting our attention to measurable sets, the property we have gained is that $|A_1 \cup A_2| = |A_1| + |A_2|$ for disjoint measurable sets (since $(A_1 \cup A_2) \cap A_1 = A_1$ and $(A_1 \cup A_2) \cap A_1^c = A_2$). Moreover, if $\{A_k\}_{k \in \mathbb{N}}$ is a countable family of measurable disjoint sets, the union $\bigcup_{k \in \mathbb{N}} A_k$ is measurable and $|\bigcup_{k \in \mathbb{N}} A_k| = \sum_{k \in \mathbb{N}} |A_k|$; see [169, Thm. 11.10].

Example 1.4 (Null sets). Let $A \subset \mathbb{R}^d$. If $|A|^* = 0$, then A is measurable. Let indeed $S \subset \mathbb{R}^d$. Then $|A \cap S|^* \leq |A|^* = 0$, i.e., $|A \cap S|^* = 0$. Moreover, $|S|^* \geq |S \cap A^c|^* = |S \cap A^c|^* + |S \cap A|^*$, and the subadditivity property implies that $|S|^* \leq |S \cap A|^* + |S \cap A^c|^*$, whence the result. $\qquad\square$

Example 1.5 (Cantor set). To define the Cantor ternary set, one starts with the interval $[0, 1]$, then one deletes the open middle third from $[0, 1]$, leaving two line segments: $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Next the open middle third of each of the two remaining segments is deleted, leaving four line segments: $[0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. This process is continued ad infinitum. Setting $C_0 := [0, 1]$ and $C_n := \frac{1}{3}C_{n-1} \cup (\frac{2}{3} + \frac{1}{3}C_{n-1})$, the Cantor ternary set is defined by $C_\infty := \{x \in [0, 1] \mid x \in C_k, \forall k \in \mathbb{N}\}$. Then C_∞ is measurable (as the complement of a countable union of measurable sets), $|C_\infty| \leq |C_k|$ for all $k \in \mathbb{N}$, so that $|C_\infty| = 0$, but it can be shown that C_∞ is not countable. $\quad\square$

Definition 1.6 (Equality a.e.). *Let $D \subset \mathbb{R}^d$ be a measurable set, i.e., $D \in \mathcal{L}(\mathbb{R}^d)$. Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ be two functions. We say that f and g are* equal almost everywhere *if $|\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \neq g(\boldsymbol{x})\}| = 0$. Henceforth, we write $f(\boldsymbol{x}) = g(\boldsymbol{x})$ for a.e. $\boldsymbol{x} \in D$, or $f = g$ a.e. in D.*

Definition 1.7 (ess sup, ess inf). *Let $D \subset \mathbb{R}^d$ be a measurable set and let $f : D \to \mathbb{R}$ be a function. We define*

$$\operatorname*{ess\,sup}_{\boldsymbol{x} \in D} f(\boldsymbol{x}) := \inf\{M \in \mathbb{R} \mid f(\boldsymbol{x}) \leq M \text{ for a.e. } \boldsymbol{x} \in D\}, \qquad (1.2a)$$

$$\operatorname*{ess\,inf}_{\boldsymbol{x} \in D} f(\boldsymbol{x}) := \sup\{m \in \mathbb{R} \mid f(\boldsymbol{x}) \geq m \text{ for a.e. } \boldsymbol{x} \in D\}. \qquad (1.2b)$$

Definition 1.8 (Measurable function). *Let $D \subset \mathbb{R}^d$ be a measurable set. A function $f : D \to \mathbb{R}$ is said to be* measurable *if $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) > r\}$ is measurable for all $r \in \mathbb{R}$.*

The meaning of the above definition is that a function is measurable if all its upper level sets are (Lebesgue) measurable; see also [169, Def. 11.13].

Lemma 1.9 (Characterization). *Let $D \subset \mathbb{R}^d$ be a measurable set. Let $f : D \to \mathbb{R}$. The function f is measurable iff any of the following statements holds true:*

(i) *For all $r \in \mathbb{R}$, the set $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) > r\}$ is measurable.*
(ii) *For all $r \in \mathbb{R}$, the set $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \geq r\}$ is measurable.*
(iii) *For all $r \in \mathbb{R}$, the set $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) < r\}$ is measurable.*
(iv) *For all $r \in \mathbb{R}$, the set $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \leq r\}$ is measurable.*

Proof. Item (i) is the definition of the measurability of f. The identity $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \geq r\} = \bigcap_{n \in \mathbb{N}} \{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) > r - \frac{1}{n+1}\}$ proves that (i) implies (ii). $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) < r\} = D \cap \{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \geq r\}^c$ proves that (ii) implies (iii). $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \leq r\} = \bigcap_{n \in \mathbb{N}} \{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) < r - \frac{1}{n+1}\}$ proves that (iii) implies (iv), and $\{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) > r\} = D \cap \{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \leq r\}^c$ proves that (iv) implies the measurability of f. (See also [169, Thm. 11.15].) $\square$

For every subset $A \subset \mathbb{R}$, let us denote by $f^{-1}(A) := \{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \in A\}$ the inverse image of A by f. Since every open set in $\mathbb{R}$ is a countable union of open intervals, the above result shows that f is measurable if and only if $f^{-1}(U) = \{\boldsymbol{x} \in D \mid f(\boldsymbol{x}) \in U\}$ is measurable for every open set U of $\mathbb{R}$.

Example 1.10 (Measurable functions). Functions that are piecewise continuous and more generally all the functions that are integrable in the Riemann sense are measurable. $\square$

Corollary 1.11 (Measurability and equality a.e.). *Let $D \subset \mathbb{R}^d$ be a measurable set. Let $f : D \to \mathbb{R}$ be a measurable function. Let $g : D \to \mathbb{R}$ be a function. If $f = g$ a.e. in D, then g is measurable.*

Proof. See Exercise 1.2. $\square$

Theorem 1.12 (Pointwise limit of measurable functions). *Let D be a measurable set in $\mathbb{R}^d$. Let $f_n : D \to \mathbb{R}$ for all $n \in \mathbb{N}$ be real-valued measurable functions. Then*

(i) $\limsup_{n \in \mathbb{N}} f_n$ *and* $\liminf_{n \in \mathbb{N}} f_n$ *are both measurable.*
(ii) *Let $f : D \to \mathbb{R}$. Assume that $f_n(\boldsymbol{x}) \to f(\boldsymbol{x})$ for a.e. $\boldsymbol{x} \in D$. Then f is measurable.*

Proof. See Exercise 1.5. $\square$

Example 1.13 (Measurability). Let $D := (0, 1)$. Let $f : D \to \mathbb{R}$ be defined by $f(x) := x$. Let C_∞ be the Cantor set (see Example 1.5). Let $g : D \to \mathbb{R}$ be defined by $g(x) := -2x$ if $x \in C_\infty$, and $g(x) := x$ if $x \notin C_\infty$. The function f is measurable since it is continuous. Recalling that $|C_\infty| = 0$, g is also measurable by virtue of Corollary 1.11 since $f = g$ a.e. in D. $\square$

Theorem 1.14 (Composite functions). *Let D be a measurable set in $\mathbb{R}^d$. Let $g : D \to \mathbb{R}$ be a measurable function. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Then $f \circ g : D \to \mathbb{R}$ is measurable.*

Proof. For every subset $A \subset \mathbb{R}$, we have $(f \circ g)^{-1}(A) = g^{-1}(f^{-1}(A))$. Let U be an open set in $\mathbb{R}$. Then $(f \circ g)^{-1}(U) = g^{-1}(f^{-1}(U))$. But $f^{-1}(U)$ is an open set since f is continuous. Hence, $g^{-1}(f^{-1}(U))$ is measurable since $f^{-1}(U)$ is open and g is measurable. As a result, $(f \circ g)^{-1}(U)$ is measurable. $\square$

Example 1.15 (Composite functions). Let $g : D \to \mathbb{R}$ be a measurable function. Then by virtue of Theorem 1.14, the functions $|g|$, $g + |g|$, $g - |g|$, $|g|^p$ for every $p > 0$, e^g, $\cos(g)$, $\sin(g)$ are also measurable. $\square$

Theorem 1.16 (Operations on measurable functions). *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ be two measurable functions and let $\lambda \in \mathbb{R}$. Then the functions λf, $f + g$, $|f|$ and fg are measurable.*

Proof. See Exercise 1.6. $\square$

1.3 Lebesgue integral

We say that $g : D \to \mathbb{R}$ is a simple nonnegative function if there exist $m \in \mathbb{N}$, a collection of disjoint measurable sets $\{A_k\}_{k \in \{1:m\}}$ in D, and a collection of nonnegative numbers $\{v_k\}_{k \in \{1:m\}}$ such that $g = \sum_{k \in \{1:m\}} v_k \mathbb{1}_{A_k}$ (where $\mathbb{1}_{A_k}(\boldsymbol{x}) := 1$ if $\boldsymbol{x} \in A_k$ and $\mathbb{1}_{A_k}(\boldsymbol{x}) := 0$ otherwise). The Lebesgue integral of g over D is defined by $\int_D g(\boldsymbol{x}) \, \mathrm{d}x := \sum_{k \in \{1:m\}} v_k |A_k|$.

Theorem 1.17 (Simple functions). *Let $D \in \mathcal{L}(\mathbb{R}^d)$. Let $f : D \to [0, \infty]$ be a nonnegative measurable function. Then there exist simple functions $\{g_k\}_{k \in \mathbb{N}}$ s.t. $0 \le g_1 \le g_2 \ldots \le f$ and $\lim_{k \to \infty} g_k(x) = f(x)$ for all $x \in D$.*

Proof. See [170, Thm. 1.17]. $\square$

Definition 1.18 (Lebesgue integral). *Let f be a nonnegative measurable function. The* Lebesgue integral *of f over D is defined in $[0, \infty]$ as follows:*

$$\int_D f(\boldsymbol{x}) \, \mathrm{d}x := \sup \left\{ \int_D g(\boldsymbol{x}) \, \mathrm{d}x \mid g \text{ is simple nonnegative and } g \le f \right\}.$$

Let f be measurable but not necessarily nonnegative. If either $\int_D f^+(\boldsymbol{x}) \, \mathrm{d}x$ or $\int_D f^-(\boldsymbol{x}) \, \mathrm{d}x$ is finite, where $f^{\pm} := \max(\pm f, 0)$, the Lebesgue integral of f is defined by

$$\int_D f(\boldsymbol{x}) \, \mathrm{d}x := \int_D f^+(\boldsymbol{x}) \, \mathrm{d}x - \int_D f^-(\boldsymbol{x}) \, \mathrm{d}x. \tag{1.3}$$

We say that f is (Lebesgue-)integrable on D if both terms in (1.3) are finite.

This definition agrees with the Riemann integral of f if f is Riemann-integrable. Moreover, since $\int_D |f(\boldsymbol{x})| \, \mathrm{d}x = \int_D f^+(\boldsymbol{x}) \, \mathrm{d}x + \int_D f^-(\boldsymbol{x}) \, \mathrm{d}x$, we have by construction $\left| \int_D f(\boldsymbol{x}) \, \mathrm{d}x \right| \le \int_D |f(\boldsymbol{x})| \, \mathrm{d}x$.

An important property of the Lebesgue integral is that if f is integrable on D, then $\int_D |f(\boldsymbol{x})|\,\mathrm{d}x = 0$ if and only if f vanishes everywhere on D up to a set of zero measure. This leads us to introduce a notion of equivalence classes. Two functions are said to belong to the same class if they coincide *almost everywhere* (henceforth, a.e.), i.e., everywhere but on a set of zero Lebesgue measure. Elements of Lebesgue spaces are, strictly speaking, equivalence classes, although we refer to them simply as functions that are defined almost everywhere. For instance, the function $\phi : (0,1) \to \{0,1\}$ that is 1 on the rational numbers and is zero otherwise is in the same equivalence class as the zero function. Hence, $\phi = 0$ a.e. on $(0,1)$. Integrals are always understood in the Lebesgue sense throughout this book. Whenever the context is unambiguous, we simply write $\int_D f\,\mathrm{d}x$ instead of $\int_D f(\boldsymbol{x})\,\mathrm{d}x$. We refer the reader to [170, Chap. 1] for more elaborate notions on the measure theory.

Example 1.19 (Cantor set). Let $f : [0,1] \to \mathbb{R}$ be such that $f(x) := 1$ if x is in C_∞ (see Example 1.5) and $f(x) := 0$ otherwise. Then f is measurable (see Corollary 1.11) and $\int_0^1 f(x)\,\mathrm{d}x = 0$. $\hfill\square$

Remark 1.20 (Literature). It is reported in Denjoy et al. [89, p. 15] that Lebesgue explained his approach to integration as follows: "I have to pay a certain sum, which I have collected in my pocket. I take the bills and coins out of my pocket and give them to the creditor in the order I find them until I have reached the total sum. This is the Riemann integral. But I can proceed differently. After I have taken all the money out of my pocket, I order the bills and coins according to identical values and then I pay the several heaps one after the other to the creditor. This is my integral." To get a clearer connection with the integration process, one could say that Lebesgue went to a grocery store every day in a month, bought items, and asked for credit until the end of the month. His debt at the end of a 30-day month is $\int_0^{30} f(t)\,\mathrm{d}t$, where $f(t)$ is the amount of money he owes per day. What Lebesgue has described above are two different ways to compute $\int_0^{30} f(t)\,\mathrm{d}t$. $\hfill\square$

1.4 Lebesgue spaces

This section introduces the Lebesgue spaces and reviews their key properties.

1.4.1 Lebesgue space $L^1(D)$

Definition 1.21 (Space L^1). *Let D be an open set in $\mathbb{R}^d$. $L^1(D)$ is the vector space composed of all the real-valued measurable functions that are Lebesgue-integrable on D, and we equip $L^1(D)$ with the norm $\|f\|_{L^1(D)} := \int_D |f|\,\mathrm{d}x$ to make it a normed space.*

Theorem 1.22 (Monotone convergence, Beppo Levi). *Let D be an open set in $\mathbb{R}^d$. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions in $L^1(D)$ such that*

$0 \leq f_0 \leq f_1 \leq \ldots \leq f_n \leq f_{n+1} \leq \ldots$ *a.e. on D and* $\sup_{n \in \mathbb{N}} \int_D f_n \, \mathrm{d}x < \infty$. *Then $f_n(x)$ converges to a finite limit for a.e. x in D. Denoting by $f(x)$ the limit in question, f is in $L^1(D)$ and $\lim_{n \in \mathbb{N}} \|f_n - f\|_{L^1(D)} = 0$.*

Proof. See [48, Thm. 4.1] or [170, Thm. 1.26]. $\qquad\square$

Theorem 1.23 (Lebesgue's dominated convergence). *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of functions in $L^1(D)$ such that*

(i) $f_n(\boldsymbol{x}) \to f(\boldsymbol{x})$ *a.e. in D.*
(ii) *There is $g \in L^1(D)$ such that $|f_n(\boldsymbol{x})| \leq g(\boldsymbol{x})$ a.e. in D for all $n \in \mathbb{N}$.*

Then $f \in L^1(D)$ and $f_n \to f$ in $L^1(D)$.

Proof. See [16, p. 123], [48, Thm. 4.2], [170, Thm. 1.34]. $\qquad\square$

Example 1.24 (Application). Let $f_n : D := (0,1) \to \mathbb{R}$, $n \in \mathbb{N}$, with $f_n(x) := 1$ if $x < \frac{1}{n}$ and $f_n(x) := x$ otherwise. We have $f_n(x) \to x$ a.e. in D and $f_n(x) \leq g := 1$ a.e. in D. Hence, $f_n \to x$ in $L^1(D)$. $\qquad\square$

Theorem 1.25 (Fischer–Riesz). *$L^1(D)$ equipped with the L^1-norm from Definition 1.33 is a Banach space.*

Proof. See [3, Thm. 2.16], [16, p. 142], [48, Thm. 4.8], [170, Thm. 3.11]. $\quad\square$

Remark 1.26 (Lebesgue vs. Riemann). The two key results the notion of Lebesgue integration gave us that were missing in the Riemann integration are Lebesgue's dominated convergence theorem and the fact that $L^1(D)$ is now complete, i.e., it is a Banach space. This answers the question raised in §1.1. $\qquad\square$

Theorem 1.27 (Pointwise convergence). *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence in $L^1(D)$ and assume that $f \in L^1(D)$ is such that $\|f_n - f\|_{L^1(D)} \to 0$. Then there exist a subsequence $(f_{n_k})_{k \in \mathbb{N}}$ and a function $g \in L^1(D)$ such that $f_{n_k}(\boldsymbol{x}) \to f(\boldsymbol{x})$ a.e. in D and $|f_{n_k}(\boldsymbol{x})| \leq g(\boldsymbol{x})$ a.e. in D for all $k \in \mathbb{N}$.*

Proof. See [48, Thm. 4.9], [170, Thm. 3.12]. $\qquad\square$

Example 1.28 (Dirac mass). The assumption that there exists some $g \in L^1(D)$ s.t. $|f_n(\boldsymbol{x})| \leq g(\boldsymbol{x})$ a.e. in D for all $n \in \mathbb{N}$, is crucial to apply Lebesgue's dominated convergence theorem. For instance, consider the sequence of functions in $L^1(\mathbb{R})$ s.t. $f_n(x) := 0$ if $|x| > \frac{1}{n}$ and $f_n(x) := \frac{n}{2}$ otherwise. We have $f_n(x) \to 0$ for a.e. x in $\mathbb{R}$ and $\int_{\mathbb{R}} |f_n(x)| \, \mathrm{d}x = 1$, but f_n does not converge in $L^1(\mathbb{R})$. Reasoning by contradiction, let us assume that $f_n \to f$ in $L^1(\mathbb{R})$. Theorem 1.27 implies that there is a subsequence $(f_{n_k})_{k \in \mathbb{N}}$ s.t. $f_{n_k}(x) \to f(x)$ for a.e. x in $\mathbb{R}$. For all $x \neq 0$, we have $f_{n_k}(x) = 0$ for all n_k such that $n_k > \frac{1}{|x|}$. This implies that $f(x) = 0$ for a.e. x in $\mathbb{R}$. This argument shows that $\int_{\mathbb{R}} |f(x)| \, \mathrm{d}x = 0$, but since we assumed that $f_n \to f$ in $L^1(\mathbb{R})$, we also have $\int_{\mathbb{R}} |f(x)| \, \mathrm{d}x = 1$, which is a contradiction. Actually $(f_n)_{\in \mathbb{N}}$ converges to the Dirac mass at 0 in the distribution sense; see Example 4.3. $\square$

Definition 1.29 (Space $L^1_{\text{loc}}(D)$). *Let D be an open set in $\mathbb{R}^d$. The elements of the following space are called* locally integrable *functions:*

$$L^1_{\text{loc}}(D) := \{v \text{ measurable} \mid \forall \text{ compact } K \subset D, v_{|K} \in L^1(K)\}. \qquad (1.4)$$

Definition 1.30 (Support). *Let D be a measurable set in $\mathbb{R}^d$. The* support *in D of a function $\varphi : D \to \mathbb{R}$, henceforth denoted by $\text{supp}(\varphi)$, is defined to be the closure in D of the subset $\{\boldsymbol{x} \in D \mid \varphi(\boldsymbol{x}) \neq 0\}$.*

Definition 1.31 (Space $C^\infty_0(D)$). *We denote by $C^\infty_0(D)$ the space composed of the functions from D to $\mathbb{R}$ that are C^∞ and whose support in D is compact. The members of $C^\infty_0(D)$ are called* test functions.

Theorem 1.32 (Vanishing integral). *Let D be an open set in $\mathbb{R}^d$. Let $v \in L^1_{\text{loc}}(D)$. Then $\int_D v\varphi \, dx = 0$ for all $\varphi \in C^\infty_0(D)$ iff $v = 0$ a.e. in D.*

Proof. See [48, Cor. 4.24], [138, p. 6]. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

1.4.2 Lebesgue spaces $L^p(D)$ and $L^\infty(D)$

Definition 1.33 (L^p spaces). *Let D be an open set in $\mathbb{R}^d$. For all $p \in [1, \infty]$, let $L^p(D) := \{f \text{ measurable} \mid \|f\|_{L^p(D)} < \infty\}$, where*

$$\|f\|_{L^p(D)} := \left(\int_D |f|^p \, dx \right)^{\frac{1}{p}}, \quad \text{if } p \in [1, \infty), \qquad (1.5a)$$

$$\|f\|_{L^\infty(D)} := \operatorname*{ess\,sup}_{\boldsymbol{x} \in D} |f(\boldsymbol{x})| := \inf\{M \in \mathbb{R} \mid |f(\boldsymbol{x})| \leq M \text{ a.e. } \boldsymbol{x} \in D\}. \quad (1.5b)$$

We write $L^p(D; \mathbb{R}^q)$, $q \geq 1$, for the space composed of $\mathbb{R}^q$-valued functions whose components are all in $L^p(D)$, and we use the Euclidean norm in $\mathbb{R}^q$, $\|f\|_{\ell^2(\mathbb{R}^q)}$, instead of $|f|$, to evaluate the norms in (1.5). When $q = d$, we write $\boldsymbol{L}^p(D) := L^p(D; \mathbb{R}^d)$.

Lebesgue's dominated convergence theorem extends to all the L^p spaces, $p \in [1, \infty)$, i.e., if the dominating function g is in $L^p(D)$, the convergence of f_n to f occurs in $L^p(D)$.

Theorem 1.34 (Pointwise convergence). *Let $p \in [1, \infty]$. Let $(f_n)_{n \in \mathbb{N}}$ be a sequence in $L^p(D)$ and let $f \in L^p(D)$ such that $\|f_n - f\|_{L^p(D)} \to 0$. Then there exist a subsequence $(f_{n_k})_{k \in \mathbb{N}}$ and a function $g \in L^p(D)$ such that $f_{n_k}(\boldsymbol{x}) \to f(\boldsymbol{x})$ a.e. in D and $|f_{n_k}(\boldsymbol{x})| \leq g(\boldsymbol{x})$ a.e. in D for all $k \in \mathbb{N}$.*

Proof. See [48, Thm. 4.9], [170, Thm. 3.12]. $\qquad\qquad\qquad\qquad\qquad\square$

Theorem 1.35 (Fischer–Riesz). *For all $p \in [1, \infty]$, $L^p(D)$ equipped with the L^p-norm from Definition 1.33 is a Banach space.*

Proof. See [3, Thm. 2.16], [16, p. 142], [48, Thm. 4.8], [170, Thm. 3.11]. $\quad\square$

Among all the Lebesgue spaces, $L^2(D)$ plays a particular role owing to the following important consequence of the Fischer–Riesz theorem.

Theorem 1.36 (L^2 space). *$L^2(D; \mathbb{R})$ is a Hilbert space when equipped with the inner product $(f,g)_{L^2(D)} := \int_D fg \, \mathrm{d}x$. Similarly, $L^2(D; \mathbb{C})$ is a Hilbert space when equipped with the inner product $(f,g)_{L^2(D)} := \int_D f\overline{g} \, \mathrm{d}x$.*

Remark 1.37 (Continuous embedding on bounded sets). Assume that D is bounded. For all $p, q \in [1, \infty]$ with $p \le q$, Hölder's inequality implies that

$$\|f\|_{L^p(D)} \le |D|^{\frac{1}{p} - \frac{1}{q}} \|f\|_{L^q(D)}, \qquad \forall f \in L^q(D), \tag{1.6}$$

meaning that $L^q(D) \hookrightarrow L^p(D)$ (this notation means that $L^q(D)$ is continuously embedded into $L^p(D)$). One can show that $\lim_{p \to \infty} \|f\|_{L^p(D)} = \|f\|_{L^\infty(D)}$ for all $f \in L^\infty(D)$. Moreover, if $f \in L^p(D)$ for all $p \in [1, \infty)$ and if there is c, uniform w.r.t. p, s.t. $\|f\|_{L^p(D)} \le c$, then $f \in L^\infty(D)$ and $\|f\|_{L^\infty(D)} \le c$; see [3, Thm. 2.14]. $\qquad\square$

Theorem 1.38 (Density of $C_0^\infty(D)$). *Let D be an open set in $\mathbb{R}^d$. Then $C_0^\infty(D)$ is dense in $L^p(D)$ for all $p \in [1, \infty)$.*

Proof. See [170, Thm. 3.14]. $\qquad\square$

Remark 1.39 (The case of $L^\infty(D)$). $C_0^\infty(D)$ is not dense in $L^\infty(D)$. If D is bounded, the completion of $C^\infty(D)$ in $L^\infty(D)$ is $C^0(D)$, and the completion of $C_0^\infty(D)$ is $\{v \in C^0(D) \mid v_{|\partial D} = 0\}$. $\qquad\square$

1.4.3 Duality

Lemma 1.40 (Conjugate, Hölder's inequality). *Let $p \in [1, \infty]$ be a real number. The real number $p' \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{p'} = 1$, with the convention that $p' := 1$ if $p = \infty$ and $p' := \infty$ if $p = 1$, is called* conjugate *of p. Let $f \in L^p(D)$ and $g \in L^{p'}(D)$. Then $fg \in L^1(D)$ and*

$$\int_D |fg| \, \mathrm{d}x \le \|f\|_{L^p(D)} \|g\|_{L^{p'}(D)}. \tag{1.7}$$

Proof. See [3, Thm. 2.4], [16, p. 404], [48, Thm. 4.6], [170, Thm. 3.8]. $\qquad\square$

For $p = p' = 2$, Hölder's inequality becomes the *Cauchy–Schwarz inequality* in $L^2(D)$: $\int_D |fg| \, \mathrm{d}x \le \|f\|_{L^2(D)} \|g\|_{L^2(D)}$ for all $f, g \in L^2(D)$. This inequality is useful to bound $|(f,g)_{L^2(D)}|$ since $|(f,g)_{L^2(D)}| \le \int_D |fg| \, \mathrm{d}x$.

Theorem 1.41 (Riesz–Fréchet). *Let $p \in [1, \infty)$. The dual space of $L^p(D)$ can be identified with $L^{p'}(D)$.*

Proof. See [3, pp. 45–49], [48, Thm. 4.11&4.14], [170, Thm. 6.16]. $\qquad\square$

Remark 1.42 ($L^\infty(D)$). Theorem 1.41 fails for $p = \infty$. Indeed, the dual of $L^\infty(D)$ strictly contains $L^1(D)$ (see [48, p. 102]). □

Corollary 1.43 (Interpolation inequality). *Let $p, q \in [1, \infty]$ with $p \le q$. For all $r \in [p, q]$, letting $\theta \in [0, 1]$ be s.t. $\frac{1}{r} := \frac{\theta}{p} + \frac{1-\theta}{q}$, we have*

$$\|f\|_{L^r(D)} \le \|f\|_{L^p(D)}^\theta \|f\|_{L^q(D)}^{1-\theta}, \qquad \forall f \in L^p(D) \cap L^q(D). \qquad (1.8)$$

Recall from §A.2 that for two Banach spaces V and W, $\mathcal{L}(V; W)$ is composed of the linear operators that map V boundedly to W, and that the norm $\|\cdot\|_{\mathcal{L}(V;W)}$ is defined in (A.2).

Theorem 1.44 (Riesz–Thorin). *Let p_0, p_1, q_0, q_1 be four real numbers such that $1 \le p_0 \le p_1 \le \infty$, $1 \le q_0 \le q_1 \le \infty$. Let $T : L^{p_0}(D) + L^{p_1}(D) \longrightarrow L^{q_0}(D) + L^{q_1}(D)$ be a linear operator that maps $L^{p_0}(D)$ and $L^{p_1}(D)$ boundedly to $L^{q_0}(D)$ and $L^{q_1}(D)$, respectively. Then the operator T maps $L^{p_\theta}(D)$ boundedly to $L^{q_\theta}(D)$ for all $\theta \in (0, 1)$, where p_θ and q_θ are defined by $\frac{1}{p_\theta} := \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$, $\frac{1}{q_\theta} := \frac{1-\theta}{q_0} + \frac{\theta}{q_1}$. Moreover, $\|T\|_{\mathcal{L}(L^{p_\theta}; L^{q_\theta})} \le \|T\|_{\mathcal{L}(L^{p_0}; L^{q_0})}^\theta \|T\|_{\mathcal{L}(L^{p_1}; L^{q_1})}^{1-\theta}$.*

Proof. See [189, Thm. 21.2], Bergh and Löfström [18, Chap. 1]. □

Remark 1.45 (Interpolation). Corollary 1.43 and Theorem 1.44 are related to the interpolation theory between Banach spaces (see §A.5). For instance, $L^p(D)$ can be defined for all $p \in (1, \infty)$, up to equivalent norm, by interpolating between $L^1(D)$ and $L^\infty(D)$, i.e., $L^p(D) = [L^1(D), L^\infty(D)]_{\frac{1}{p'}, p}$; see Tartar [189, p. 111]. □

1.4.4 Multivariate functions

The following results on multivariate functions are useful in many situations.

Theorem 1.46 (Tonelli). *Let $f : D_1 \times D_2 \to \mathbb{R}$ be a measurable function such that the function $D_1 \ni \boldsymbol{x}_1 \mapsto \int_{D_2} |f(\boldsymbol{x}_1, \boldsymbol{x}_2)| \, \mathrm{d}x_2$ is finite a.e. in D_1 and is in $L^1(D_1)$. Then $f \in L^1(D_1 \times D_2)$.*

Proof. See [48, Thm. 4.4]. □

Theorem 1.47 (Fubini). *Let $f \in L^1(D_1 \times D_2)$. Then the function $D_2 \ni \boldsymbol{x}_2 \mapsto f(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is in $L^1(D_2)$ for a.e. $\boldsymbol{x}_1 \in D_1$, and the function $D_1 \ni \boldsymbol{x}_1 \mapsto \int_{D_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \, \mathrm{d}x_2$ is in $L^1(D_1)$. Similarly, the function $D_1 \ni \boldsymbol{x}_1 \mapsto f(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is in $L^1(D_1)$ for a.e. $\boldsymbol{x}_2 \in D_2$, and the function $D_2 \ni \boldsymbol{x}_2 \mapsto \int_{D_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \, \mathrm{d}x_1$ is in $L^1(D_2)$. Moreover, we have*

$$\int_{D_1} \left(\int_{D_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \, \mathrm{d}x_2 \right) \mathrm{d}x_1 = \int_{D_2} \left(\int_{D_1} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \, \mathrm{d}x_1 \right) \mathrm{d}x_2, \qquad (1.9)$$

and both quantities are equal to $\int_{D_1 \times D_2} f(\boldsymbol{x}_1, \boldsymbol{x}_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2$, where $\mathrm{d}x_1 \, \mathrm{d}x_2$ is the product measure on the Cartesian product $D_1 \times D_2$.

Exercises

Exercise 1.1 (Measurability). Let W be a nonmeasurable subset of $D :=$ $(0,1)$. Let $f : W \to \mathbb{R}$ be defined by $f(x) := 1$ if $x \in D\backslash W$ and $f(x) := 0$ if $x \in W$. (i) Is f measurable? (ii) Assume that there is a measurable subset $V \subset W$ s.t. $|V| > 0$. Compute $\sup_{x \in D} f(x)$, ess $\sup_{x \in D} f(x)$, $\inf_{x \in D} f(x)$, ess $\inf_{x \in D} f(x)$. (iii) Is f a member of $L^{\infty}(D)$? (iv) Assume now that W has zero measure (hence W is measurable). Compute $\inf_{x \in D} f(x)$ and ess $\inf_{x \in D} f(x)$.

Exercise 1.2 (Measurability and equality a.e.). Prove Corollary 1.11. (*Hint*: consider the sets $A_r := \{x \in D \mid f(x) > r\}$ and $B_r := \{x \in D \mid g(x) > r\}$ for all $r \in \mathbb{R}$, and show that $B_r = (A_r \cap (A_r\backslash B_r)^c) \cup (B_r \backslash A_r)$.)

Exercise 1.3 (Lebesgue's theorem). Let $D := (-1,1)$. Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of functions in $L^1(D)$ and let $g \in L^1(D)$. Assume that $f_n \to f$ a.e. in D. Propose a counterexample to show that the assumption "$|f_n| \le g$ a.e. for all $n \in \mathbb{N}$" cannot be replaced by "$f_n \le g$ a.e. for all $n \in \mathbb{N}$" in Lebesgue's dominated convergence theorem.

Exercise 1.4 (Compact support). Let $D := (0,1)$ and $f(x) := 1$ for all $x \in D$. What is the support of f in D? Is the support compact?

Exercise 1.5 (Pointwise limit of measurable functions). Let D be a measurable set in $\mathbb{R}^d$. Let $f_n : D \to \mathbb{R}$ for all $n \in \mathbb{N}$ be real-valued measurable functions. (i) Show that $\limsup_{n \in \mathbb{N}} f_n$ and $\liminf_{n \in \mathbb{N}} f_n$ are both measurable. (*Hint*: recall that $\limsup_{n \in \mathbb{N}} f_n(x) := \inf_{n \in \mathbb{N}} \sup_{k \ge n} f_k(x)$ and $\liminf_{n \in \mathbb{N}} f_n(x) := \sup_{n \in \mathbb{N}} \inf_{k \ge n} f_k(x)$ for all $x \in D$). (ii) Let $f : D \to \mathbb{R}$. Assume that $f_n(x) \to f(x)$ for every $x \in D$. Show that f is measurable. (iii) Let $f : D \to \mathbb{R}$. Assume that $f_n(x) \to f(x)$ for a.e. $x \in D$. Show that f is measurable.

Exercise 1.6 (Operations on measurable functions). The objective of this exercise is to prove Theorem 1.6. Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ be two measurable functions and let $\lambda \in \mathbb{R}$. (i) Show that λf is measurable. (*Hint*: use Lemma 1.9). (ii) Idem for $|f|$. (iii) Idem for $f + g$. (iv) Idem for fg. (*Hint*: observe that $fg = \frac{1}{2}(f+g)^2 - \frac{1}{2}(f-g)^2$.)

Part I, Chapter 2

Weak derivatives and Sobolev spaces

We investigate in this chapter the notion of differentiation for Lebesgue integrable functions. We introduce an extension of the classical concept of derivative and partial derivative which is called weak derivative. This notion will be used throughout the book. It is particularly useful when one tries to differentiate finite element functions that are continuous and piecewise polynomial. In that case, one does not need to bother about the points where the classical derivative is multivalued to define the weak derivative. We also introduce the concept of Sobolev spaces. These spaces are useful to study the well-posedness of partial differential equations and their approximation using finite elements.

2.1 Differentiation

We study here the concept of differentiation for Lebesgue integrable functions.

2.1.1 Lebesgue points

Theorem 2.1 (Lebesgue points). *Let $f \in L^1(D)$. Let $B(\boldsymbol{x}, h)$ be the ball of radius $h > 0$ centered at $\boldsymbol{x} \in D$. The following holds true for a.e. $\boldsymbol{x} \in D$:*

$$\lim_{h \downarrow 0} \frac{1}{|B(\boldsymbol{x}, h)|} \int_{B(\boldsymbol{x}, h)} |f(\boldsymbol{y}) - f(\boldsymbol{x})| \, \mathrm{d}y = 0. \tag{2.1}$$

Points $\boldsymbol{x} \in D$ where (2.1) holds true are called Lebesgue points *of f.*

Proof. See, e.g., Rudin [170, Thm. 7.6]. □

This result says that for a.e. $\boldsymbol{x} \in D$, the averages of $|f(\cdot) - f(\boldsymbol{x})|$ are small over small balls centered at $\boldsymbol{x}$, i.e., f does not oscillate too much in the neighborhood of $\boldsymbol{x}$. Notice that if the function f is continuous at $\boldsymbol{x}$, then $\boldsymbol{x}$ is a Lebesgue point of f (recall that a continuous function is uniformly continuous over compact sets).

Let $\mathcal{H} \subset \mathbb{R}$ be is a countable set with 0 as unique accumulation point (the sign of the members of $\mathcal{H}$ is unspecified). Let $F : \mathbb{R} \to \mathbb{R}$. We say that F is strongly differentiable at x if the sequence $(\frac{F(x+h)-F(x)}{h})_{h \in \mathcal{H}}$ converges.

Theorem 2.2 (Lebesgue's differentiation). *Let $f \in L^1(\mathbb{R})$. Let $F(x) := \int_{-\infty}^{x} f(t)\,\mathrm{d}t$. Then F is strongly differentiable at every Lebesgue point x of f, and at these points we have $F'(x) = f(x)$.*

Proof. See Exercise 2.2.   □

In the above theorem, we have $F'(x) = f(x)$ for a.e. x in $\mathbb{R}$. Thus, it is tempting to move away from the classical sense of differentiation and view F' as a function in $L^1(\mathbb{R})$. If we could make sense of F' in $L^1(\mathbb{R})$, then $F(x) = \int_{-\infty}^{x} F'(t)\,\mathrm{d}t$ would be an extension of the fundamental theorem of calculus in Lebesgue spaces. As an example of this possibility, let $f := \mathbb{1}_{[0,\infty)}$ be the Heaviside function (i.e., $f(x) := 1$ if $x \geq 0$ and $f(x) := 0$ otherwise). Notice that $f \notin L^1(\mathbb{R})$ but $f \in L^1_{loc}(\mathbb{R})$ (see Definition 1.29), and $F(x) := \int_{-\infty}^{x} f(t)\,\mathrm{d}t$ is well defined. Then $F(x) = 0$ if $x < 0$ and $F(x) = x$ if $x > 0$ (notice that 0 is not a Lebesgue point of f; see Exercise 2.1). We would like to say that $F' = f$ in $L^1_{loc}(\mathbb{R})$. The objective of the rest of this section is to make sense of the above argument.

2.1.2 Weak derivatives

Definition 2.3 (Weak derivative). *Let D be an open set in $\mathbb{R}^d$. Let $u, v \in L^1_{loc}(D)$. Let $i \in \{1{:}d\}$. We say that v is the weak partial derivative of u in the direction i if*

$$\int_D u\partial_i \varphi \,\mathrm{d}x = -\int_D v\varphi \,\mathrm{d}x, \qquad \forall \varphi \in C_0^\infty(D), \qquad (2.2)$$

and we write $\partial_i u := v$. Let $\alpha \in \mathbb{N}^d$ be a multi-index. We say that v is the weak α-th partial derivative of u and we write $\partial_1^{\alpha_1} \ldots \partial_d^{\alpha_d} u := v$ if

$$\int_D u\partial_1^{\alpha_1} \ldots \partial_d^{\alpha_d} \varphi \,\mathrm{d}x = (-1)^{|\alpha|} \int_D v\varphi \,\mathrm{d}x, \qquad \forall \varphi \in C_0^\infty(D), \qquad (2.3)$$

where $|\alpha| := \alpha_1 + \ldots + \alpha_d$. Finally, we write $\partial^\alpha u := \partial_1^{\alpha_1} \ldots \partial_d^{\alpha_d} u$, and we set $\partial^{(0,\ldots,0)} u := u$.

Lemma 2.4 (Uniqueness). *Let $u \in L^1_{loc}(D)$. If u has a weak α-th partial derivative, then it is uniquely defined.*

Proof. Let $v_1, v_2 \in L^1_{loc}(D)$ be two weak α-th derivatives of u. We have

$$\int_D v_1 \varphi \,\mathrm{d}x = (-1)^{|\alpha|} \int_D u\partial^\alpha \varphi \,\mathrm{d}x = \int_D v_2 \varphi \,\mathrm{d}x, \qquad \forall \varphi \in C_0^\infty(D).$$

Hence, $\int_D (v_1 - v_2)\varphi \, dx = 0$. The vanishing integral theorem (Theorem 1.32) implies that $v_1 = v_2$ a.e. in D. $\square$

If $u \in C^{|\alpha|}(D)$, then the usual and the weak α-th partial derivatives are identical. Moreover, it can be shown that if $\alpha, \beta \in \mathbb{N}^d$ are multi-indices such that $\alpha_i \geq \beta_i$ for all $i \in \{1{:}d\}$, then if the α-th weak derivative of u exists in $L^1_{\mathrm{loc}}(D)$, so does the β-th weak derivative. For instance, with $d = 1$ (writing ∂_x instead of ∂_1), if $\partial_{xx} u$ exists in $L^1_{\mathrm{loc}}(D)$, so does $\partial_x u$; see Exercise 2.4.

Example 2.5 (1D). Let us revisit the heuristic argument at the end of §2.1.1. Let $D := (-1, 1)$. (i) Let us first consider a continuous function $u \in C^0(D; \mathbb{R})$, e.g., $u(x) := 0$ if $x < 0$ and $u(x) := x$ otherwise. Then u has a weak derivative. Indeed, let $v \in L^1(D)$ be s.t. $v(x) := 0$ if $x < 0$ and $v(x) := 1$ otherwise. Let $\varphi \in C_0^\infty(D)$. We have

$$\int_{-1}^{1} u \partial_x \varphi \, dx = \int_0^1 x \partial_x \varphi \, dx = -\int_0^1 \varphi \, dx = -\int_{-1}^1 v \varphi \, dx.$$

Hence, v is the weak derivative of u. (Notice that $\tilde{v}$ defined by $\tilde{v}(x) := 0$ if $x < 0$, $\tilde{v}(0) := \frac{1}{2}$ and $\tilde{v}(x) := 1$ if $x > 0$ is also a weak derivative of u, but $v = \tilde{v}$ a.e. in D, i.e., v and $\tilde{v}$ coincide in the Lebesgue sense.) (ii) Let us now consider a function $u \in L^1(D; \mathbb{R})$ that is piecewise smooth but exhibits a jump at $x = 0$, e.g., $u(x) := -1$ if $x < 0$ and $u(x) := x$ otherwise. Then u does not have a weak derivative. Let us prove this statement by contradiction. Assume that there is $v \in L^1_{\mathrm{loc}}(D)$ s.t. $\partial_x u = v$. We have

$$\int_{-1}^1 v \varphi \, dx = -\int_{-1}^1 u \partial_x \varphi \, dx = \int_{-1}^0 \partial_x \varphi \, dx - \int_0^1 x \partial_x \varphi \, dx = \varphi(0) + \int_0^1 \varphi \, dx,$$

for all $\varphi \in C_0^\infty(D)$. Let $\{\varphi_n\}_{n \in \mathbb{N}}$ be a sequence of functions in $C_0^\infty(D)$ s.t. $0 \leq \varphi_n(x) \leq 1$ for all $x \in D$, $\varphi_n(0) = 1$, and $\varphi_n \to 0$ a.e. in D. Lebesgue's dominated convergence theorem implies that $1 = \lim_{n \to \infty}(\int_{-1}^1 v \varphi_n \, dx - \int_0^1 \varphi_n \, dx) = 0$, which is a contradiction. $\square$

Lemma 2.6 (Passing to the limit). *Let $\{v_n\}_{n \in \mathbb{N}}$ be a sequence in $L^p(D)$, $p \in [1, \infty]$, with weak α-th partial derivatives $\{\partial^\alpha v_n\}_{n \in \mathbb{N}}$ in $L^p(D)$. Assume that $v_n \to v$ in $L^p(D)$ and $\partial^\alpha v_n \to g_\alpha$ in $L^p(D)$. Then v has a weak α-th partial derivative and $\partial^\alpha v = g_\alpha$.*

Proof. The assumptions imply that $\lim_{n \to \infty} \int_D \partial^\alpha v_n \varphi \, dx = \int_D g_\alpha \varphi \, dx$ and

$$\lim_{n \to \infty} \int_D \partial^\alpha v_n \varphi \, dx = (-1)^{|\alpha|} \lim_{n \to \infty} \int_D v_n \partial^\alpha \varphi \, dx = (-1)^{|\alpha|} \int_D v \partial^\alpha \varphi \, dx,$$

for all $\varphi \in C_0^\infty(D)$. The conclusion follows readily. $\square$

A function $v \in L^1_{\mathrm{loc}}(D)$ is said to be *locally Lipschitz* in D if for all $\boldsymbol{x} \in D$, there is a neighborhood $\mathcal{N}_{\boldsymbol{x}}$ of $\boldsymbol{x}$ in D and a constant $L_{\boldsymbol{x}}$ such that $|v(\boldsymbol{z}) - v(\boldsymbol{y})| \leq L_{\boldsymbol{x}} \|\boldsymbol{z} - \boldsymbol{y}\|_{\ell^2(\mathbb{R}^d)}$ for all $\boldsymbol{y}, \boldsymbol{z} \in \mathcal{N}_{\boldsymbol{x}}$.

Theorem 2.7 (Rademacher).   *Let D be an open set in $\mathbb{R}^d$. Let f be a locally Lipschitz function in D. Then f is differentiable in the classical sense a.e. in D. The function f is also weakly differentiable, and the classical and weak derivatives of f coincide a.e. in D.*

Proof. See [99, p. 280], [138, p. 44]. □

2.2 Sobolev spaces

In this section, we introduce integer-order and fractional-order Sobolev spaces. The scale of Sobolev spaces plays a central role in the finite element error analysis to quantify the decay rate of the approximation error.

2.2.1 Integer-order spaces

Definition 2.8 ($W^{m,p}(D)$).   *Let $m \in \mathbb{N}$ and $p \in [1,\infty]$. Let D be an open set in $\mathbb{R}^d$. We define the Sobolev space*

$$W^{m,p}(D) := \{v \in L^1_{\mathrm{loc}}(D) \mid \partial^\alpha v \in L^p(D), \, \forall \alpha \in \mathbb{N}^d \text{ s.t. } |\alpha| \leq m\}, \quad (2.4)$$

where the derivatives are weak partial derivatives. We write $W^{m,p}(D;\mathbb{R}^q)$, $q \geq 1$, for the space composed of $\mathbb{R}^q$-valued functions whose components are all in $W^{m,p}(D)$, and we write $\boldsymbol{W}^{m,p}(D)$ whenever $q = d$.

Whenever it is possible to identify a length scale ℓ_D associated with D, e.g., its diameter $\ell_D := \mathrm{diam}(D)$ if D is bounded, we equip $W^{m,p}(D)$ with the following norm and seminorm: If $p \in [1,\infty)$, we set

$$\|v\|^p_{W^{m,p}(D)} := \sum_{|\alpha| \leq m} \ell_D^{|\alpha|p} \|\partial^\alpha v\|^p_{L^p(D)}, \qquad |v|^p_{W^{m,p}(D)} := \sum_{|\alpha|=m} \|\partial^\alpha v\|^p_{L^p(D)},$$

and if $p = \infty$, we set

$$\|v\|_{W^{m,\infty}(D)} := \max_{|\alpha| \leq m} \ell_D^{|\alpha|} \|\partial^\alpha v\|_{L^\infty(D)}, \qquad |v|_{W^{m,\infty}(D)} := \max_{|\alpha|=m} \|\partial^\alpha v\|_{L^\infty(D)},$$

where the sums and the maxima run over multi-indices $\alpha \in \mathbb{N}^d$. The advantage of using the factor ℓ_D is that all the terms in the sums or maxima have the same dimension (note that $\|\cdot\|_{W^{m,p}(D)}$ and $|\cdot|_{W^{m,p}(D)}$ have a different scaling w.r.t. ℓ_D). If there is no length scale available or if one works with dimensionless space variables, one sets $\ell_D := 1$ in the above definitions.

Proposition 2.9 (Banach space).   *$W^{m,p}(D)$ equipped with the $\|\cdot\|_{W^{m,p}(D)}$-norm is a Banach space. For $p = 2$, the space*

$$H^m(D) := W^{m,2}(D) \qquad (2.5)$$

is a real Hilbert space when equipped with the inner product $(v,w)_{H^m(D)} := \sum_{|\alpha| \le m} \int_D \partial^\alpha v \, \partial^\alpha w \, dx$. Similarly, $H^m(D;\mathbb{C})$ is a complex Hilbert space when equipped with the inner product $(v,w)_{H^m(D;\mathbb{C})} := \sum_{|\alpha| \le m} \int_D \partial^\alpha v \, \overline{\partial^\alpha w} \, dx$.

Proof. We are going to do the proof for $m = 1$. See e.g., [3, Thm. 3.3], [99, p. 249], or [189, Lem. 5.2] for the general case. Let $\{v_n\}_{n \in \mathbb{N}}$ be a Cauchy sequence in $W^{1,p}(D)$. Then $\{v_n\}_{n \in \mathbb{N}}$ is a Cauchy sequence in $L^p(D)$ and the sequences of weak partial derivatives $\{\partial_i v_n\}_{n \in \mathbb{N}}$ are also Cauchy sequences in $L^p(D)$. Hence, there is $v \in L^p(D)$ and there are $g_1, \dots, g_d \in L^p(D)$ such that $v_n \to v$ in $L^p(D)$ and $\partial_i v_n \to g_i$ in $L^p(D)$. We conclude by invoking Lemma 2.6. $\qquad\square$

Example 2.10 ($H^1(D)$)**.** Taking $m := 1$ and $p := 2$ we have

$$H^1(D) := \{v \in L^2(D) \mid \partial_i v \in L^2(D), \, \forall i \in \{1{:}d\}\},$$

(notice that $L^2(D) \subset L^1_{\mathrm{loc}}(D)$) and

$$\|v\|_{H^1(D)} := \left(\|v\|^2_{L^2(D)} + \ell^2_D |v|^2_{H^1(D)}\right)^{\frac{1}{2}}, \quad |v|^2_{H^1(D)} := \sum_{i \in \{1{:}d\}} \|\partial_i v\|^2_{L^2(D)}.$$

Let ∇v be the column vector in $\mathbb{R}^d$ whose components are the directional weak derivatives $\partial_i v$ of v. Then a more compact notation is $H^1(D) := \{v \in L^2(D) \mid \nabla v \in \boldsymbol{L}^2(D)\}$ and $|v|_{H^1(D)} := \|\nabla v\|_{\boldsymbol{L}^2(D)}$. $\qquad\square$

Lemma 2.11 (Kernel of ∇). *Let D be open and connected set in $\mathbb{R}^d$. Let $v \in W^{1,p}(D)$, $p \in [1,\infty]$. Then $\nabla v = \boldsymbol{0}$ a.e. on D iff v is constant.*

Proof. We prove the result for $D := (-1,1)$ and we refer the reader to [138, p. 24], [189, Lem. 6.4] for the general case. Let $u \in L^1_{\mathrm{loc}}(D)$ be such that $\partial_x u = 0$. Fix a function $\rho \in C_0^\infty(D)$ such that $\int_D \rho \, dx = 1$ and set $c_\rho := \int_D u\rho \, dx$. Let now $\varphi \in C_0^\infty(D)$ and set $c_\varphi := \int_D \varphi \, dx$. Then the function $\psi(x) := \int_{-1}^x (\varphi(y) - c_\varphi \rho(y)) \, dy$ is by construction in $C_0^\infty(D)$, and we have $\partial_x \psi(x) = \varphi(x) - c_\varphi \rho(x)$. Since $\int_D u \partial_x \psi \, dx = -\int_D (\partial_x u)\psi \, dx = 0$ by assumption on $\partial_x u$, we infer that

$$\int_D u\varphi \, dx = \int_D u(\partial_x \psi + c_\varphi \rho) \, dx = c_\varphi \int_D u\rho \, dx = c_\rho \int_D \varphi \, dx,$$

for all $\varphi \in C_0^\infty(D)$. Theorem 1.32 shows that $u = c_\rho$. $\qquad\square$

Remark 2.12 (Lipschitz functions). Let D be an open set in $\mathbb{R}^d$. The space of Lipschitz functions $C^{0,1}(D)$ is closely related to the Sobolev space $W^{1,\infty}(D)$. Indeed, $C^{0,1}(D) \cap L^\infty(D)$ is continuously embedded into $W^{1,\infty}(D)$. Conversely, if $v \in W^{1,\infty}(D)$, then $|v(\boldsymbol{y}) - v(\boldsymbol{z})| \le d_D(\boldsymbol{y}, \boldsymbol{z}) \|\nabla v\|_{\boldsymbol{L}^\infty(D)}$ for all $\boldsymbol{y}, \boldsymbol{z} \in D$, where $d_D(\boldsymbol{y}, \boldsymbol{z})$ denotes the geodesic distance of $\boldsymbol{y}$ to $\boldsymbol{z}$ in D, i.e., the shortest length of a smooth path connecting $\boldsymbol{y}$ to $\boldsymbol{z}$ in D (if D is convex,

$d_D(\boldsymbol{y}, \boldsymbol{z}) = \|\boldsymbol{y} - \boldsymbol{z}\|_{\ell^2}$); see [189, Lem. 7.8]. A set $D \subset \mathbb{R}^d$ is said to be quasiconvex if there exists $C \geq 1$ s.t. every pair of points $\boldsymbol{x}, \boldsymbol{y} \in D$ can be joined by a curve γ in D with $\text{length}(\gamma) \leq C\|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2}$. If D is a quasiconvex open set, then $W^{1,\infty}(D) = C^{0,1}(D) \cap L^{\infty}(D)$, and if D is also bounded, then $W^{1,\infty}(D) = C^{0,1}(D)$; see Heinonen [113, Thm. 4.1]. $\qquad\square$

Remark 2.13 (Broken seminorms). Let $D \subset \mathbb{R}^d$ be an open set and let $\{D_i\}_{i \in \{1:I\}}$ be a partition of D, i.e., all the subsets D_i are open, mutually disjoint, and $D \setminus \bigcup_{i \in \{1:I\}} D_i$ has zero Lebesgue measure. Let $v \in W^{1,p}(D)$ and $p \in [1, \infty)$. Then one can write $|v|_{W^{1,p}(D)}^p = \sum_{i \in \{1:I\}} \|(\nabla v)_{|D_i}\|_{\boldsymbol{L}^p(D_i)}^p$. In this book, we are going to abuse the notation by writing $|v|_{W^{1,p}(D)}^p = \sum_{i \in \{1:I\}} \|\nabla v\|_{\boldsymbol{L}^p(D_i)}^p$. This abuse is justified by observing that $(\nabla v)_{|D_i} = \nabla(v_{|D_i})$ for all $v \in W^{1,1}_{\text{loc}}(D)$. We stress that it is important that the weak derivative of v exists to make sense of the above identities. For instance, letting H be the Heaviside function, we have $\|\nabla(H_{|(-1,0)})\|_{\boldsymbol{L}^p(-1,0)}^p + \|\nabla(H_{|(0,1)})\|_{\boldsymbol{L}^p(0,1)}^p = 0$, but $H \notin W^{1,p}(D)$; see Exercise 2.8. $\qquad\square$

2.2.2 Fractional-order spaces

Definition 2.14 ($W^{s,p}(D)$). Let $s \in (0,1)$ and $p \in [1,\infty]$. Let D be an open set in $\mathbb{R}^d$. We define $W^{s,p}(D) := \left\{ v \in L^p(D) \mid |v|_{W^{s,p}(D)} < \infty \right\}$, where

$$|v|_{W^{s,p}(D)} := \left(\int_D \int_D \frac{|v(\boldsymbol{x}) - v(\boldsymbol{y})|^p}{\|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2}^{sp+d}} \, \mathrm{d}x \, \mathrm{d}y \right)^{\frac{1}{p}}, \quad p < \infty, \qquad (2.6)$$

and $|v|_{W^{s,\infty}(D)} := \mathrm{ess\,sup}_{\boldsymbol{x},\boldsymbol{y} \in D} \frac{|v(\boldsymbol{x}) - v(\boldsymbol{y})|}{\|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2}^s}$. Letting now $s > 1$, we define

$$W^{s,p}(D) := \{ v \in W^{m,p}(D) \mid \partial^\alpha v \in W^{\sigma,p}(D), \forall \alpha, |\alpha| = m \}, \qquad (2.7)$$

where $m := \lfloor s \rfloor$ and $\sigma := s - m$. Finally, we denote $H^s(D) := W^{s,2}(D)$. We write $W^{s,p}(D; \mathbb{R}^q)$, $q \geq 1$, for the space composed of $\mathbb{R}^q$-valued functions whose components are all in $W^{s,p}(D)$, and we write $\boldsymbol{W}^{s,p}(D)$ whenever $q = d$.

Definition 2.15 (Sobolev–Slobodeckij norm). Let $s = m + \sigma$ with $m := \lfloor s \rfloor$ and $\sigma := s - m \in (0,1)$. For all $p \in [1,\infty)$ and all $v \in W^{s,p}(D)$, we set $\|v\|_{W^{s,p}(D)}^p := \|v\|_{W^{m,p}(D)}^p + \ell_D^{sp}|v|_{W^{s,p}(D)}^p$ with seminorm $|v|_{W^{s,p}(D)}^p := \sum_{|\alpha|=m} |\partial^\alpha v|_{W^{\sigma,p}(D)}^p$. We also set

$$\|v\|_{W^{s,\infty}(D)} := \max(\|v\|_{W^{m,\infty}(D)}, \ell_D^s |v|_{W^{s,\infty}(D)}),$$

with seminorm $|v|_{W^{s,\infty}(D)} := \max_{|\alpha|=m} |\partial^\alpha v|_{W^{\sigma,\infty}(D)}$. Equipped with this norm $W^{s,p}(D)$ is a Banach space (and a Hilbert space if $p = 2$).

Example 2.16 (Power functions). Let $D := (0,1)$ and consider the function $v(x) := x^\alpha$ with $\alpha \in \mathbb{R}$. One can verify that $v \in L^2(D)$ if $\alpha > -\frac{1}{2}$, $v \in H^1(D)$ if $\alpha > \frac{1}{2}$, and, more generally $v \in H^s(D)$ if $\alpha > s - \frac{1}{2}$. $\qquad\square$

Example 2.17 (Hölder functions). If D is bounded and $p \in [1, \infty)$, then $C^{0,\alpha}(D) \hookrightarrow W^{s,p}(D)$ provided $0 \le s < \alpha \le 1$; see Exercise 2.9. $\qquad\square$

Example 2.18 (Step function). Let $D := (-1, 1)$ and consider $v(x) := 0$ if $x < 0$ and $v(x) := 1$ if $x \ge 0$. Then $v \in W^{s,p}(D)$ iff $sp < 1$ as shown by the following computation (notice that $sp > 0$):

$$|v|^p_{W^{s,p}} = 2 \int_{-1}^0 \int_0^1 \frac{1}{|y-x|^{sp+1}} \, \mathrm{d}x \, \mathrm{d}y = 2 \int_{-1}^0 -\frac{1}{sp}\left(\frac{1}{(1-x)^{sp}} - \frac{1}{|x|^{sp}} \right) \mathrm{d}x.$$

The integral $\int_{-1}^0 \frac{1}{|x|^{sp}} \, \mathrm{d}x$ is convergent if and only if $sp < 1$. $\qquad\square$

Remark 2.19 (Limits $s{\downarrow}0$ and $s{\uparrow}1$). The expression (2.6), which is usually adopted in the literature to define $|v|_{W^{s,p}(D)}$, gives $|v|_{W^{s,p}(D)} \to \infty$ as $s{\uparrow}1$ even if $v \in W^{1,p}(D)$. A remedy to this deficiency has been proposed in Bourgain et al. [38], Maz'ya and Shaposhnikova [140]. It is shown in [38] that by redefining $|v|^*_{W^{s,p}} := (1-s)^{\frac{1}{p}}|v|_{W^{s,p}}$ for all $s \in (0,1)$, and setting $|v|^*_{W^{1,p}(D)} := |v|_{W^{1,p}(D)}$, there exists c, s.t. for all σ, s with $0 < \sigma < s \le 1$ and all $v \in W^{\sigma,p}(D)$, one has $|v|^*_{W^{\sigma,p}(D)} \le c|v|^*_{W^{s,p}(D)}$ and $\lim_{s\to1} |v|^*_{W^{s,p}(D)} = |v|_{W^{1,p}(D)}$ (see Borthagaray and Ciarlet [34, Rmk. 2.3]). It has been proposed [140, Thm. 3] to redefine $|v|^*_{W^{s,p}(D)} := (s(1-s))^{\frac{1}{p}}|v|_{W^{s,p}(D)}$ to improve also the behavior of the seminorm when $s{\downarrow}0$. It is shown therein that if there is $\sigma > 0$ s.t. $v \in \overline{C_0^\infty(\mathbb{R}^d)}^{W^{\sigma,p}}$, then $\lim_{s\downarrow0} s|v|^p_{W^{s,p}(\mathbb{R}^d)} = 2p^{-1}|S^{d-1}|\|v\|^p_{L^p(\mathbb{R}^d)}$, where $|S^{d-1}|$ is the measure of the unit sphere in $\mathbb{R}^d$. $\qquad\square$

Remark 2.20 (Definition by interpolation). Fractional-order Sobolev spaces can also be defined by means of the interpolation theory between Banach spaces (see §A.5). Let $p \in [1, \infty)$ and $s \in (0,1)$. Then we have

$$W^{s,p}(D) = [L^p(D), W^{1,p}(D)]_{s,p},$$

and more generally $W^{m+s,p}(D) = [W^{m,p}(D), W^{m+1,p}(D)]_{s,p}$ for all $m \in \mathbb{N}$, with equivalent norms in all the cases; see Tartar [189, Lem. 36.1]. Using the interpolation theory may not be convenient in finite element analysis if one is interested in local approximation properties. Unless specified otherwise we use the Sobolev–Slobodeckij norm in the book. $\qquad\square$

2.3 Key properties: density and embedding

This section reviews some key properties of Sobolev spaces: the density of smooth functions and the (compact) embedding into Lebesgue spaces or into spaces composed of Hölder continuous functions.

2.3.1 Density of smooth functions

Theorem 2.21 (Meyers–Serrin). *Let D be an open set in $\mathbb{R}^d$. Let $s \geq 0$ and $p \in [1, \infty)$. Then $C^\infty(D) \cap W^{s,p}(D)$ is dense in $W^{s,p}(D)$.*

Proof. See Meyers and Serrin [143] and Adams and Fournier [3, Thm. 3.17]; see also Evans [99, p. 251] for bounded D. $\qquad\square$

Remark 2.22 ($p = \infty$). Let $m \in \mathbb{N}$. The closure of $C^\infty(D) \cap W^{m,\infty}(D)$ with respect to the Sobolev norm $\|\cdot\|_{W^{m,\infty}(D)}$ differs from $W^{m,\infty}(D)$ since it is composed of functions whose derivatives up to order m are continuous and bounded on D. $\qquad\square$

The density of smooth functions in Sobolev spaces allows one to derive many useful results. We list here some of the most important ones.

Corollary 2.23 (Differentiation of a product). *Let D be an open subset of $\mathbb{R}^d$. Then we have $uv \in W^{1,p}(D) \cap L^\infty(D)$ and $\nabla(uv) = v\nabla u + u\nabla v$ for all $u, v \in W^{1,p}(D) \cap L^\infty(D)$ and all $p \in [1, \infty]$.*

Proof. See, e.g., [48, Prop. 9.4, p. 269]. $\qquad\square$

Corollary 2.24 (Differentiation of a composition). *Let $D \subset \mathbb{R}^d$ be an open set. Let $G \in C^1(\mathbb{R})$. Assume that $G(0) = 0$ and there is $M < \infty$ such that $|G'(t)| \leq M$ for all $t \in \mathbb{R}$. Then we have $G(u) \in W^{1,p}(D)$ and $\nabla(G(u)) = G'(u)\nabla u$ for all $u \in W^{1,p}(D)$ and all $p \in [1, \infty]$.*

Proof. See, e.g., [48, Prop. 9.5, p. 270]. $\qquad\square$

Corollary 2.25 (Change of variable). *Let D, D' be two open subsets of $\mathbb{R}^d$. Assume that there exists a bijection $\boldsymbol{T} : D' \to D$ s.t. $\boldsymbol{T} \in C^1(D'; D)$, $\boldsymbol{T}^{-1} \in C^1(D; D')$, $D\boldsymbol{T} \in L^\infty(D'; \mathbb{R}^{d \times d})$, and $D\boldsymbol{T}^{-1} \in L^\infty(D; \mathbb{R}^{d \times d})$, where $D\boldsymbol{T}$ and $D\boldsymbol{T}^{-1}$ are the Jacobian matrices of $\boldsymbol{T}$ and $\boldsymbol{T}^{-1}$, respectively. Then we have $u \circ T \in W^{1,p}(D')$ for all $u \in W^{1,p}(D)$ and all $p \in [1, \infty]$, and $\partial_{x_i'}(u \circ \boldsymbol{T})(\boldsymbol{x}') = \sum_{j \in \{1:d\}} \partial_{x_j} u(T(\boldsymbol{x}))\partial_{x_i'}\boldsymbol{T}(\boldsymbol{x}')$ for all $i \in \{1:d\}$ and $\boldsymbol{x}' \in D'$.*

Proof. See, e.g., [48, Prop. 9.6, p. 270]. $\qquad\square$

2.3.2 Embedding

We use the notation $V \hookrightarrow W$ to mean that the embedding of V into W is continuous, i.e., there is c such that $\|v\|_W \leq c\|v\|_V$ for all $v \in V$ (see §A.2). The main idea of the results in this section is that functions in the Sobolev space $W^{s,p}(D)$ with differentiability index $s > 0$ do have an integrability index larger than p (i.e., they belong to some Lebesgue space $L^q(D)$ with $q > p$), and if s is sufficiently large, for all $u \in W^{s,p}(D)$ (recall that u is actually a class of functions that coincide almost everywhere in D), there is a representative of u that is continuous (or even Hölder continuous). How large s must

be for these properties to hold true depends on the space dimension. The case $d = 1$ is particularly simple since $W^{1,1}(\mathbb{R}) \hookrightarrow C^0(\mathbb{R})$ and $W^{1,1}(D) \hookrightarrow C^0(\overline{D})$ for every bounded interval D; see [189, Lem. 8.5] (see also Exercise 5.7). In the rest of this section, we assume that $d \geq 2$. We first consider the case where $D := \mathbb{R}^d$.

Theorem 2.26 (Embedding of $W^{1,p}(\mathbb{R}^d)$). *Let $d \geq 2$ and let $p \in [1, \infty]$. The following holds true:*

(i) *(Gagliardo–Nirenberg–Sobolev): If $p \in [1, d)$, then*

$$W^{1,p}(\mathbb{R}^d) \hookrightarrow L^q(\mathbb{R}^d), \qquad \forall q \in [p, p^*], \ p^* := \frac{pd}{d-p}. \qquad (2.8)$$

In particular, $\|u\|_{L^{p^*}(\mathbb{R}^d)} \leq \frac{p^*}{1^*}\|\nabla u\|_{L^p(\mathbb{R}^d)}$ with $1^* := \frac{d}{d-1}$ for all $u \in W^{1,p}(\mathbb{R}^d)$. Hence, $W^{1,p}(\mathbb{R}^d) \hookrightarrow L^{p^*}(\mathbb{R}^d)$, and the embedding into $L^q(\mathbb{R}^d)$ for all $q \in [p, p^*)$ follows from Corollary 1.43.

(ii) *If $p = d$, then*

$$W^{1,d}(\mathbb{R}^d) \hookrightarrow L^q(\mathbb{R}^d), \qquad \forall q \in [d, \infty). \qquad (2.9)$$

(iii) *(Morrey): If $p \in (d, \infty]$, then*

$$W^{1,p}(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d) \cap C^{0,\alpha}(\mathbb{R}^d), \qquad \alpha := 1 - \frac{d}{p}. \qquad (2.10)$$

Proof. See [48, Thm. 9.9, Cor. 9.11, Thm. 9.12], [99, p. 263–266], [180, §I.7.4, §I.8.2], [189, Chap. 8–9]. □

Remark 2.27 (Continuous function). The embedding (2.10) means that there is c, only depending on p and d, such that

$$|u(\boldsymbol{x}) - u(\boldsymbol{y})| \leq c\|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2(\mathbb{R}^d)}^\alpha \|\nabla u\|_{\boldsymbol{L}^p(\mathbb{R}^d)}, \quad \text{for a.e. } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d, \qquad (2.11)$$

for all $u \in W^{1,p}(\mathbb{R}^d)$. In other words, there is a continuous function $v \in C^{0,\alpha}(\mathbb{R}^d)$ such that $u = v$ almost everywhere. It is then possible to replace u by its continuous representative v. We will systematically do this replacement in this book when a continuous embedding in a space of continuous functions is invoked. □

The above results extend to Sobolev spaces of arbitrary order.

Theorem 2.28 (Embedding of $W^{s,p}(\mathbb{R}^d)$). *Let $d \geq 2$, $s > 0$, and $p \in [1, \infty]$. The following holds true:*

$$W^{s,p}(\mathbb{R}^d) \hookrightarrow \begin{cases} L^q(\mathbb{R}^d) & \forall q \in [p, \frac{pd}{d-sp}], \text{ if } sp < d, \\ L^q(\mathbb{R}^d) & \forall q \in [p, \infty), \text{ if } sp = d, \\ L^\infty(\mathbb{R}^d) \cap C^{0,\alpha}(\mathbb{R}^d) & \alpha := 1 - \frac{d}{sp}, \text{ if } sp > d. \end{cases} \qquad (2.12)$$

Moreover, $W^{d,1}(\mathbb{R}^d) \hookrightarrow L^\infty(\mathbb{R}^d) \cap C^0(\mathbb{R}^d)$ (case $s = d$ and $p = 1$).

Proof. See [110, Thm. 1.4.4.1], [88, Thm. 4.47] for $p \in (1, \infty)$. For $s = d$ and $p = 1$, see, e.g., Ponce and Van Schaftingen [160] and Campos Pinto [55, Prop. 3.4] (if $d = 2$). $\square$

Our aim is now to generalize Theorem 2.28 to the space $W^{s,p}(D)$, where D is an open set in $\mathbb{R}^d$. A rather generic way to proceed is to use the concept of extension.

Definition 2.29 ((s,p)-extension). *Let $s > 0$ and $p \in [1, \infty]$. Let D be an open set in $\mathbb{R}^d$. The set D is said to have the (s,p)-extension property if there is a bounded linear operator $E : W^{s,p}(D) \to W^{s,p}(\mathbb{R}^d)$ such that $E(u)|_D = u$ for all $u \in W^{s,p}(D)$.*

Theorem 2.28 can be restated by replacing $\mathbb{R}^d$ with any open set D in $\mathbb{R}^d$ that has the (s,p)-extension property. A rather general class of sets that we consider in this book is that of Lipschitz sets in $\mathbb{R}^d$. A precise definition is given in the next chapter. At this stage, it suffices to know that the boundary of a Lipschitz set can be viewed as being composed of a finite collection of epigraphs of Lipschitz functions.

Theorem 2.30 (Extension from Lipschitz sets). *Let $s > 0$ and $p \in [1, \infty]$. Let D be an open, bounded subset of $\mathbb{R}^d$. If D is a Lipschitz set, then it has the (s,p)-extension property.*

Proof. See Calderón [54], Stein [181, p. 181] (for $s \in \mathbb{N}$), [110, Thm. 1.4.3.1] and [88, Prop. 4.43] (for $p \in (1, \infty)$), [141, Thm. A.1&A.4] (for $s \in [0,1]$, $p \in [1, \infty]$ and $s > 0$, $p = 2$) [189, Lem. 12.4] (for $s = 1$). $\square$

Theorem 2.31 (Embedding of $W^{s,p}(D)$). *Let $d \geq 2$, $s > 0$, and $p \in [1, \infty]$. Let D be an open, bounded subset of $\mathbb{R}^d$. If D is a Lipschitz set, then we have*

$$W^{s,p}(D) \hookrightarrow \begin{cases} L^q(D) & \forall q \in [p, \frac{pd}{d-sp}], \text{ if } sp < d, \\ L^q(D) & \forall q \in [p, \infty), \text{ if } sp = d, \\ L^\infty(D) \cap C^{0,\alpha}(D) & \alpha := 1 - \frac{d}{sp}, \text{ if } sp > d. \end{cases} \qquad (2.13)$$

Moreover, $W^{d,1}(D) \hookrightarrow L^\infty(D) \cap C^0(D)$ (case $s = d$ and $p = 1$).

Remark 2.32 (Bounded set). Note that $W^{s,p}(D) \hookrightarrow L^q(D)$ for $sp \leq d$ and all $q \in [1, p]$ since D is bounded. The boundedness of D also implies that $W^{s,p}(D) \hookrightarrow C^{0,\alpha}(\overline{D})$, with $sp > d$ and $\alpha := 1 - \frac{d}{sp}$, and $W^{d,1}(D) \hookrightarrow C^0(\overline{D})$, i.e., there is (Hölder-)continuity up to the boundary. $\square$

Example 2.33 (Embedding into continuous functions). In dimension one, functions in $H^1(D)$ are bounded and continuous, whereas this may not be the case in dimension $d \geq 2$ (see Exercise 2.10). In dimension $d \in \{2,3\}$, Theorem 2.31 says that functions in $H^2(D)$ are bounded and continuous. $\square$

Example 2.34 (Boundary smoothness). Let $\alpha > 1$, $p \in [1,2)$, and $D :=$ $\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \in (0,1),\ x_2 \in (0, x_1^\alpha)\}$. Let $u(x_1, x_2) := x_1^\beta$ with $1 - \frac{1+\alpha}{p} <$ $\beta < 0$ (this is possible since $p < 2 < 1+\alpha$). Then $u \in W^{1,p}(D)$ and $u \in L^q(D)$ for all q such that $1 \le q < p_\alpha$ where $\frac{1}{p_\alpha} := \frac{1}{p} - \frac{1}{1+\alpha}$. Let us set $\frac{1}{p^*} := \frac{1}{p} - \frac{1}{2}$, $\epsilon := \frac{\beta-1}{1+\alpha} + \frac{1}{p} > 0$, and $\frac{1}{p_\beta} := \frac{1}{p_\alpha} - \epsilon$. Notice that $p_\alpha < p^*$ and also $p_\alpha < p_\beta$ since $\epsilon > 0$. Since one can choose β s.t. ϵ is arbitrarily close to zero, we pick β so that $p_\beta < p^*$. Hence, $p_\beta \in (p_\alpha, p^*)$. But $u \notin L^q(D)$ for all $q \in (p_\beta, p^*)$, which would contradict Theorem 2.31 if the Lipschitz property had been omitted. Hence, D cannot be a Lipschitz set in $\mathbb{R}^2$ ($\alpha > 1$ means that D has a cusp at the origin). This counterexample shows that some smoothness assumption on D is needed for Theorem 2.31 to hold true. $\square$

We conclude this section with important compactness results. Recall from §A.4 that the embedding $V \hookrightarrow W$ between two Banach spaces is compact iff from every bounded sequence in V, one can extract a converging subsequence in W.

Theorem 2.35 (Rellich–Kondrachov). *Let $s > 0$ and $p \in [1,\infty]$. Let D be an open, bounded subset of $\mathbb{R}^d$. If D is a Lipschitz set, then the following embeddings are compact:*

(i) *If $sp \le d$, $W^{s,p}(D) \hookrightarrow L^q(D)$ for all $q \in [1, \frac{pd}{d-sp})$.*

(ii) *If $sp > d$, $W^{s,p}(D) \hookrightarrow C^0(\overline{D})$.*

(iii) *$W^{s,p}(D) \hookrightarrow W^{s',p}(D)$ for all $s > s'$.*

Proof. See [3, Thm. 6.3], [48, Thm. 9.16], [99, p. 272], [138, p. 35], [110, Thm. 1.4.3.2]. $\square$

Exercises

Exercise 2.1 (Lebesgue point). Let $a \in \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) := 0$ if $x < 0$, $f(0) := a$, and $f(x) := 1$ if $x > 0$. Show that 0 is not a Lebesgue point of f for all a.

Exercise 2.2 (Lebesgue differentiation). The goal is to prove Theorem 2.2. (i) Let $h \in \mathcal{H}$ (the sign of h is unspecified). Show that $R(x,h) :=$ $\frac{F(x+h)-F(x)}{h} - f(x) = \frac{1}{h}\int_x^{x+h}(f(t) - f(x))\,\mathrm{d}t$. (ii) Conclude.

Exercise 2.3 (Lebesgue measure and weak derivative). Let $D :=$ $(0,1)$. Let C_∞ be the Cantor set (see Example 1.5). Let $f : D \to \mathbb{R}$ be defined by $f(x) := x$ if $x \notin C_\infty$, and $f(x) := 2-5x$ if $x \in C_\infty$. (i) Is f measurable? (*Hint*: use Corollary 1.11.) (ii) Compute $\sup_{x \in D} f(x)$, $\operatorname{ess\,sup}_{x \in D} f(x)$, $\inf_{x \in D} f(x)$, $\operatorname{ess\,inf}_{x \in D} f(x)$, and $\|f\|_{L^\infty(D)}$. (iii) Show that f is weakly differentiable and compute $\partial_x f(x)$. (iv) Compute $f(x) - \int_0^x \partial_t f(t)\,\mathrm{d}t$ for all $x \in D$. (iv) Identify the function $f^c \in C^0(\overline{D})$ that satisfies $f = f^c$ a.e. on D? Compute $f^c(x) - \int_0^x \partial_t f(t)\,\mathrm{d}t$ for all $x \in D$.

Exercise 2.4 (Weak derivative). Let $D := (-1, 1)$. Prove that if $u \in L^1_{\text{loc}}(D)$ has a second-order weak derivative, it also has a first-order weak derivative. (*Hint*: consider $\psi(x) := \int_{-1}^x (\varphi(t) - c_\varphi \rho(t)) \, dt$ for all $\varphi \in C_0^\infty(D)$, with $c_\varphi := \int_D \varphi \, dx$, $\rho \in C_0^\infty(D)$, and $\int_D \rho \, dx = 1$.)

Exercise 2.5 (Clairaut's theorem). Let $v \in L^1_{\text{loc}}(D)$. Let $\alpha, \beta \in \mathbb{N}^d$ and assume that the weak derivatives $\partial^\alpha v$, $\partial^\beta v$ exist and that the weak derivative $\partial^\alpha(\partial^\beta v)$ exists. Prove that $\partial^\beta(\partial^\alpha v)$ exists and $\partial^\alpha(\partial^\beta v) = \partial^\beta(\partial^\alpha v)$.

Exercise 2.6 (Weak and classical derivatives). Let $k \in \mathbb{N}$, $k \geq 1$, and let $v \in C^k(D)$. Prove that, up to the order k, the weak derivatives and the classical derivatives of v coincide.

Exercise 2.7 ($H^1(D)$). (i) Let $D := (-1, 1)$ and $u : D \to \mathbb{R}$ s.t. $u(x) := |x|^{\frac{3}{2}} - 1$. Determine whether u is a member of $H^1(D; \mathbb{R})$. (ii) Let $u_1 \in C^1((-1, 0]; \mathbb{R})$ and $u_2 \in C^1([0, 1); \mathbb{R})$ and assume that $u_1(0) = u_2(0)$. Let u be such that $u_{|(-1,0)} := u_1$ and $u_{|(0,1)} := u_2$. Determine whether u is a member of $H^1(D; \mathbb{R})$. Explain why $u \notin H^1(D; \mathbb{R})$ if $u_1(0) \neq u_2(0)$.

Exercise 2.8 (Broken seminorm). Let D be an open set in $\mathbb{R}^d$. Let $\{D_1, \ldots, D_n\}$ be a partition of D as in Remark 2.13. (i) Show that $(\nabla v)_{|D_i} = \nabla(v_{|D_i})$ for all $i \in \{1{:}n\}$ and all $v \in W^{1,1}_{\text{loc}}(D)$. (ii) Let $p \in [1, \infty)$ and $v \in W^{1,p}(D)$. Show that $\sum_{i \in \{1:n\}} |v_{|D_i}|^p_{W^{1,p}(D_i)} = |v|^p_{W^{1,p}(D)}$. (iii) Let $s \in (0, 1)$, $p \in [1, \infty)$, and $v \in W^{s,p}(D)$. Prove that $\sum_{i \in \{1:n\}} |v_{|D_i}|^p_{W^{s,p}(D_i)} \leq |v|^p_{W^{s,p}(D)}$.

Exercise 2.9 ($W^{s,p}$). Let D be a bounded open set in $\mathbb{R}^d$. Let $\alpha \in (0, 1]$. Show that $C^{0,\alpha}(D; \mathbb{R}) \hookrightarrow W^{s,p}(D; \mathbb{R})$ for all $p \in [1, \infty)$ if $s \in [0, \alpha)$.

Exercise 2.10 (Unbounded function in $H^1(D)$). Let $D := B(\mathbf{0}, \frac{1}{2}) \subset \mathbb{R}^2$ be the ball centered at 0 and of radius $\frac{1}{2}$. (i) Show that the (unbounded) function $u(\boldsymbol{x}) := \ln(-\ln(\|\boldsymbol{x}\|_{\ell^2}))$ has weak partial derivatives. (*Hint*: work on $D \backslash B(\mathbf{0}, \epsilon)$ with $\epsilon \in (0, \frac{1}{2})$, and use Lebesgue's dominated convergence theorem.) (ii) Show that u is in $H^1(D)$.

Exercise 2.11 (Equivalent norm). Let $m \in \mathbb{N}$, $m \geq 2$, and let $p \in [1, \infty)$. Prove that the norm $\|v\| := (\|v\|^p_{L^p} + \ell^{mp}_D |v|^p_{W^{m,p}(D)})^{\frac{1}{p}}$ is equivalent to the canonical norm in $W^{m,p}(D)$. (*Hint*: use the Peetre–Tartar lemma (Lemma A.20) and invoke the compact embeddings from Theorem 2.35.)

Part I, Chapter 3

Traces and Poincaré inequalities

This chapter reviews two types of results on the Sobolev spaces $W^{s,p}(D)$ introduced in the previous chapter. The first one concerns the notion of trace (i.e., loosely speaking, the boundary values) of functions in $W^{s,p}(D)$. The second one is about functional inequalities (due to Poincaré and Steklov) essentially bounding the L^p-norm of a function by that of its gradient. The validity of these results relies on some smoothness properties on the boundary of the set D. In this book, we mainly focus on Lipschitz sets. For any subset $S \subset \mathbb{R}^d$, $d \geq 1$, $\mathrm{int}(S)$ denotes the interior of S and $\overline{S}$ its closure.

3.1 Lipschitz sets and domains

Definition 3.1 (Domain). *Let D be a nonempty subset of $\mathbb{R}^d$. In this book, D is called* domain *if it is open, bounded, and connected.*

For instance, a domain in $\mathbb{R}$ is an open and bounded interval. At many instances in this book we will need to say something on the smoothness of the boundary ∂D of a domain $D \subset \mathbb{R}^d$, $d \geq 2$. To stay simple, we are going to focus our attention on the class of Lipschitz domains. In simple words, a Lipschitz domain D in $\mathbb{R}^d$, $d \geq 2$, is such that at every point $\boldsymbol{x} \in \partial D$, the boundary can be represented in a neighborhood of $\boldsymbol{x}$ as the graph of a Lipschitz function. Equivalently there exists a cone with nonzero aperture angle that can be moved in the neighborhood of $\boldsymbol{x}$ without changing direction and without exiting D. Let us now give some precise definitions.

Definition 3.2 (Lipschitz set and domain). *An open set D in $\mathbb{R}^d$, $d \geq 2$, is said to be* Lipschitz *if for all $\boldsymbol{x} \in \partial D$, there exists a neighborhood $V_{\boldsymbol{x}}$ of $\boldsymbol{x}$ in $\mathbb{R}^d$, a rotation $\boldsymbol{R}_{\boldsymbol{x}} : \mathbb{R}^d \to \mathbb{R}^d$, and two real numbers $\alpha > 0$, $\beta > 0$ (α and β may depend on $\boldsymbol{x}$) s.t. the following holds true:*

(i) $V_{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{R}_{\boldsymbol{x}}(B_\alpha \times I_\beta)$ *with* $B_\alpha := B_{\mathbb{R}^{d-1}}(\boldsymbol{0}, \alpha)$, $I_\beta := (-\beta, \beta)$.

(ii) *There exists a Lipschitz function $\phi_{\boldsymbol{x}} : B_\alpha \to \mathbb{R}$ such that $\phi_{\boldsymbol{x}}(\mathbf{0}) = 0$, $\|\phi_{\boldsymbol{x}}\|_{L^\infty(B_\alpha)} \leq \frac{1}{2}\beta$ and (see Figure 3.1)*

$$D \cap V_{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{R}_{\boldsymbol{x}}(\{(\boldsymbol{y}', y_d) \in B_\alpha \times I_\beta \mid y_d < \phi_{\boldsymbol{x}}(\boldsymbol{y}')\}), \qquad (3.1\text{a})$$

$$\partial D \cap V_{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{R}_{\boldsymbol{x}}(\{(\boldsymbol{y}', y_d) \in B_\alpha \times I_\beta \mid y_d = \phi_{\boldsymbol{x}}(\boldsymbol{y}')\}). \qquad (3.1\text{b})$$

We say that D is a Lipschitz domain *if it is a domain and a Lipschitz set.*
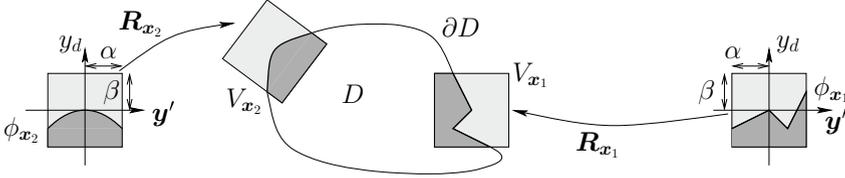
Fig. 3.1 Lipschitz domain and mappings $(\boldsymbol{R}_{\boldsymbol{x}_1}, \phi_{\boldsymbol{x}_1})$, $(\boldsymbol{R}_{\boldsymbol{x}_2}, \phi_{\boldsymbol{x}_2})$.

Definition 3.3 (Cone property). *Let D be an open set in $\mathbb{R}^d$, $d \geq 2$. We say that D has the* uniform cone property *if for all $\boldsymbol{x} \in \partial D$, there exists a neighborhood $V_{\boldsymbol{x}}$ of $\boldsymbol{x}$ in $\mathbb{R}^d$, a rotation $\boldsymbol{R}_{\boldsymbol{x}} : \mathbb{R}^d \to \mathbb{R}^d$, positive real numbers α, β, h, $\theta \in (0, \frac{\pi}{2}]$ (which may depend on $\boldsymbol{x}$) s.t. the following holds true:*

(i) $V_{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{R}_{\boldsymbol{x}}(B_\alpha \times I_\beta)$ *with $B_\alpha := B_{\mathbb{R}^{d-1}}(\mathbf{0}, \alpha)$, $I_\beta := (-\beta, \beta)$.*
(ii) *For all $\boldsymbol{y} \in (\overline{D} \cap V_{\boldsymbol{x}})$, we have $\boldsymbol{y} + \boldsymbol{R}_{\boldsymbol{x}}(C) \subset D$ with the cone $C :=$ $\{(\boldsymbol{y}', y_d) \in \mathbb{R}^{d-1} \times \mathbb{R} \mid -h < y_d < -\cot(\theta)\|\boldsymbol{y}'\|_{\ell^2(\mathbb{R}^{d-1})}\}$.*

Lemma 3.4 (Lipschitz domain and cone property). *A domain in $\mathbb{R}^d$, $d \geq 2$, has the (uniform) cone property iff it is Lipschitz.*

Proof. See Grisvard [110, Thm. 1.2.2.2]. □

Remark 3.5 (Finite covering). Let D be a domain in $\mathbb{R}^d$. Since ∂D is compact, there is a finite set $\mathcal{L} \subset \mathbb{N}$ and a finite covering $\bigcup_{i \in \mathcal{L}} V_{\boldsymbol{x}_i}$ of ∂D with $\boldsymbol{x}_i \in \partial D$ for all $i \in \mathcal{L}$. Definition 3.2 and Definition 3.3 can be equivalently reformulated for the finite set $\{\boldsymbol{x}_i\}_{i \in \mathcal{L}}$ with coefficients $\{\alpha_i, \beta_i, \theta_i, h_i\}_{i \in \mathcal{L}}$ that are bounded from below away from zero (the change of coordinates described by the rotation $\boldsymbol{R}_{\boldsymbol{x}_i}$ being fixed in each $V_{\boldsymbol{x}_i}$). □

Remark 3.6 (Terminology). In the literature, the term "domain" is sometimes defined without requiring D to be bounded. We have incorporated this requirement in our definition since we mostly consider bounded sets in this book. Domains that are Lipschitz in the sense of Definition 3.2 are sometimes called *strongly Lipschitz*. It is also possible to weaken this definition. For instance, some authors say that a domain D in $\mathbb{R}^d$ is *weakly Lipschitz* if for every $\boldsymbol{x} \in \partial D$, there exists a neighborhood $V_{\boldsymbol{x}} \ni \boldsymbol{x}$ in $\mathbb{R}^d$ and a global bilipschitz

mapping $\boldsymbol{M_x} : \mathbb{R}^{d-1} \times \mathbb{R} \to \mathbb{R}^d$ such that $D \cap V_{\boldsymbol{x}} = \boldsymbol{M_x}(\mathbb{R}^{d-1} \times \mathbb{R}_-) \cap V_{\boldsymbol{x}}$ and $\partial D \cap V_{\boldsymbol{x}} = \boldsymbol{M_x}(\mathbb{R}^{d-1} \times \{0\}) \cap V_{\boldsymbol{x}}$. A strongly Lipschitz domain is weakly Lipschitz (using the notation of Definition 3.2, it suffices to set $\boldsymbol{M_x}(\boldsymbol{y'}, y_d) = \boldsymbol{x} + \boldsymbol{R_x}(\boldsymbol{y'}, y_d + \phi_{\boldsymbol{x}}(\boldsymbol{y'})))$, but a weakly Lipschitz domain may not be strongly Lipschitz. For instance, the two-brick domain (see Example 3.7) and the logarithmic spiral $\{re^{i\theta} \mid r > 0, \theta \in \mathbb{R}, a_1 e^{-\theta} < r < a_2 e^{-\theta}\} \subsetneq \mathbb{R}^2$ (with positive real numbers a_1, a_2 s.t. $e^{-2\pi}a_2 < a_1 < a_2$ and $i^2 = -1$) are weakly Lipschitz but are not strongly Lipschitz; see Axelsson and McIntosh [13]. These examples show that the image of a strongly Lipschitz domain by a bilipschitz mapping is not necessarily strongly Lipschitz. A weakly Lipschitz domain is strongly Lipschitz if the mapping $\boldsymbol{M_x}$ is continuously differentiable. The source of the difficulty is that the implicit function theorem does not hold true for Lipschitz functions; see [110, pp. 7–10] for more details. In this book, we only consider strongly Lipschitz domains and, unless explicitly stated otherwise, when we say "let D be a Lipschitz domain" we mean that D is strongly Lipschitz in the sense of Definition 3.2.   □

Lipschitz domains have many important properties:

(i) Outward normal: the outward-pointing unit normal $\boldsymbol{n}$ is well defined a.e. on the boundary of a Lipschitz domain (this follows from Rademacher's theorem (Theorem 2.7)). For an interval in $\mathbb{R}$, the outward unit normal is conventionally set to be -1 at the left endpoint and $+1$ at the right endpoint (in coherence with the conventional orientation of $\mathbb{R}$ from left to right).

(ii) One-sided property: a Lipschitz domain is always located on one side of its boundary, i.e., there cannot be slits or cuts; see Costabel and Dauge [82], Grisvard [110, §1.7] for discussions on domains with cuts.

(iii) Convexity: any Lipschitz domain is quasiconvex (see Remark 2.12). Conversely every convex domain is Lipschitz (see [110, Cor. 1.2.2.3]).
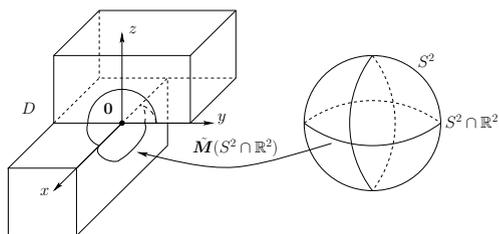
Fig. 3.2 (Surprising) example of non-Lipschitz domain: the two-brick assembly.

Example 3.7 (Two-brick domain). Consider the bricks $B_1 := (-2, 0) \times (-2, 2) \times (0, 2)$ and $B_2 := (-2, 2) \times (-2, 0) \times (-2, 0)$, and the two-brick assembly $D := \mathrm{int}(\overline{B_1} \cup \overline{B_2})$ illustrated in Figure 3.2. Let us show that D is not a Lipschitz domain by using the uniform cone property. For any $\epsilon \in (0, 1)$, let

$V_{\mathbf{0}}$ be the ball of radius 3ϵ centered at $\mathbf{0}$. The points $\boldsymbol{a} := (\epsilon, -\epsilon, 0)$ and $\boldsymbol{a}' := (-\epsilon, \epsilon, 0)$ are both in $V_{\mathbf{0}} \cap \overline{D}$. Let us assume that the uniform cone property holds, and let $\boldsymbol{\zeta} := (\zeta_x, \zeta_y, \zeta_z)^{\mathsf{T}} := R_{\mathbf{0}}((0, 0, -1)^{\mathsf{T}})$. Item (ii) in Definition 3.3 requires that $\boldsymbol{a} + \frac{1}{2}h\boldsymbol{\zeta} \in D$, which in turn implies that $\zeta_z < 0$. But also we must have $\boldsymbol{a}' + \frac{1}{2}h\boldsymbol{\zeta} \in D$, which implies that $\zeta_z > 0$. This contradiction implies that Item (ii) from Definition 3.3 cannot hold true for any neighborhood of $\mathbf{0}$. In other words, one cannot find a coordinate system such that the boundary of D is described by the graph of a Lipschitz function in the neighborhood of the origin. Incidentally, one can show that D is a weakly Lipschitz domain. Letting $\boldsymbol{\psi} : S^2 \to S^2$ be a bilipschitz homeomorphism of the unit sphere in $\mathbb{R}^3$ that maps the circle $S^2 \cap \mathbb{R}^2$ to the curve shown in bold in Figure 3.2, a mapping $\boldsymbol{M_0} : \mathbb{R}^3 \to \mathbb{R}^3$ satisfying the definition from Remark 3.6 can be defined by $\boldsymbol{M_0}(\boldsymbol{x}) := \|\boldsymbol{x}\|_{\ell^2}\boldsymbol{\psi}(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{\ell^2}})$ if $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{M_0}(\mathbf{0}) := \mathbf{0}$. That $\boldsymbol{M_0}$ is a bilipschitz mapping results from the identity $\|r_1\boldsymbol{\omega}_1 - r_2\boldsymbol{\omega}_2\|_{\ell^2}^2 = |r_1 - r_2|^2 + r_1 r_2 \|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2\|_{\ell^2}^2$ with the notation $r_i := \|\boldsymbol{x}_i\|_{\ell^2}$, $\boldsymbol{\omega}_i := \frac{\boldsymbol{x}_i}{\|\boldsymbol{x}_i\|_{\ell^2}}$. $\square$

Remark 3.8 (Stronger smoothness). D is said to be of *class C^m* or *piecewise of class C^m*, $m \geq 1$, if all the local mappings $\phi_{\boldsymbol{x}}$ in the Definition 3.2 are of class C^m or piecewise of class C^m, respectively. In this case, the outward unit normal is well defined for all $\boldsymbol{x} \in \partial D$ and is of class C^{m-1}. $\square$

3.2 Traces as functions at the boundary

Boundary values of functions in $L^p(D)$, $p \in [1, \infty)$, are in general not well defined. For instance, let $D := (0, 1)^2$ and $v(x_1, x_2) := x_1^{-\frac{\alpha}{p}}$ with $\alpha \in (0, 1)$. Then $v \in L^p(D)$ but $v|_{x_1=0} = \infty$. The main idea of this section is to show that it is possible to define the boundary value of a function $v \in W^{s,p}(D)$ if s is large enough. But how large? A first possibility is to invoke Morrey's theorem (see (2.10)): if $sp > d$, one can consider the continuous representative of v to define the boundary value of v. The purpose of the trace theory is to give a meaning to boundary values under the weaker assumption $sp > 1$ (and $s \geq 1$ if $p = 1$) in every space dimension. In what follows, we consider Sobolev spaces defined on ∂D by using the local mappings $\phi_{\boldsymbol{x}}$ from the Definition 3.2: letting $\boldsymbol{\psi}_{\boldsymbol{x}}(\boldsymbol{\xi}) := (\boldsymbol{\xi}, \phi_{\boldsymbol{x}}(\boldsymbol{\xi}))$ for all $\boldsymbol{\xi}$ in the open ball $B(\mathbf{0}, \alpha)$ in $\mathbb{R}^{d-1}$, we say that v is in $W^{s,p}(\partial D)$ if $v \circ \boldsymbol{\psi}_{\boldsymbol{x}} \in W^{s,p}(B(\mathbf{0}, \alpha))$ for all $\boldsymbol{x} \in \partial D$. When D is Lipschitz (resp., of class $C^{1,1}$), this approach defines $W^{s,p}(\partial D)$ up to $s = 1$ (resp., $s = 2$). We refer to Grisvard [110, §1.3.3] for more details.

3.2.1 The spaces $W_0^{s,p}(D)$, $W^{s,p}(D)$ and their traces

Definition 3.9 ($W_0^{s,p}(D)$). *Let $s > 0$ and $p \in [1, \infty)$. Let D be an open set in $\mathbb{R}^d$. We define*

$$W_0^{s,p}(D) := \overline{C_0^\infty(D)}^{W^{s,p}(D)}, \tag{3.2}$$

i.e.,, $W_0^{s,p}(D)$ is the closure of the subspace $C_0^\infty(D)$ in $W^{s,p}(D)$. For $p = 2$, we write $H_0^s(D) := W_0^{s,2}(D)$.

We will see below in Theorem 3.19 that $W^{s,p}(D) = W_0^{s,p}(D)$ if $sp \le 1$ whereas $W_0^{s,p}(D)$ is a proper subspace of $W^{s,p}(D)$ if $sp > 1$ and D is bounded.

Theorem 3.10 (Trace). *Let $p \in [1, \infty)$. Let $s > \frac{1}{p}$ if $p > 1$ or $s \ge 1$ if $p = 1$. Let D be a Lipschitz domain in $\mathbb{R}^d$. There is a bounded linear map $\gamma^g : W^{s,p}(D) \to L^p(\partial D)$ such that*

(i) *$\gamma^g(v) = v_{|\partial D}$ whenever v is smooth, e.g., $v \in C^0(\overline{D})$.*
(ii) *The kernel of γ^g is $W_0^{s,p}(D)$.*
(iii) *If $s = 1$ and $p = 1$, or if $s \in (\frac{1}{2}, \frac{3}{2})$ and $p = 2$, or if $s \in (\frac{1}{p}, 1]$ and $p \notin \{1, 2\}$, then $\gamma^g : W^{s,p}(D) \to W^{s-\frac{1}{p},p}(\partial D)$ is bounded and surjective, that is, there exists C_{γ^g} s.t. for every function $g \in W^{s-\frac{1}{p},p}(\partial D)$, one can find a function $u_g \in W^{s,p}(D)$, called a* lifting *of g, s.t.*

$$\gamma^g(u_g) = g \qquad \text{and} \qquad \|u_g\|_{W^{s,p}(D)} \le C_{\gamma^g} \ell_D^{\frac{1}{p}} \|g\|_{W^{s-\frac{1}{p},p}(\partial D)}, \qquad (3.3)$$

where ℓ_D is a characteristic length of D, e.g., $\ell_D := \operatorname{diam}(D)$.

Proof. See Brezis [48, p. 315] ($s = 1$, $p \in [1, \infty)$), Grisvard [110, Thm. 1.5.1.2 & Cor. 1.5.1.6], McLean [141, Thm. 3.38] ($s \in (\frac{1}{2}, \frac{3}{2})$, $p = 2$); see Gagliardo [104] for the original work with $s = 1$, $p \in [1, \infty)$. □

Remark 3.11 (Notation). The superscript g stands for "gradient" since $\gamma^g(v)$ is meaningful for $v \in W^{1,1}(D)$, i.e., $\gamma^g(v)$ makes sense if the weak gradient of v is integrable. □

Example 3.12 (Elliptic PDEs). Theorem 3.10 (with $s = 1$ and $p = 2$) is crucial in the analysis of elliptic PDEs, where a natural functional setting for the solution is the space $H^1(D)$. Whenever a homogeneous Dirichlet condition is enforced (prescribing to zero the value of the solution at the boundary), Item (ii) shows that the solution is in $H_0^1(D)$. When the boundary condition prescribes a nonzero value, the surjectivity of $\gamma^g : H^1(D) \to H^{\frac{1}{2}}(\partial D)$ is invoked to identify a proper functional setting (see Chapter 31). □

Remark 3.13 ($W^{1,\infty}(D)$). The trace theory in $W^{1,\infty}(D)$ is not trivial since $C^\infty(D)$ is not dense in $L^\infty(D)$; see Remark 1.39. The situation simplifies if D is quasiconvex since $W^{1,\infty}(D) = C^{0,1}(D)$ in this case (see Remark 2.12). □

Remark 3.14 (Trace of gradient). If $v \in W^{s,p}(D)$ with $p \in [1, \infty)$ and $s > 1 + \frac{1}{p}$ if $p > 1$ or $s \ge 2$ if $p = 1$, then $\nabla v \in \boldsymbol{W}^{s-1,p}(D)$, and we can apply Theorem 3.10 componentwise, i.e., $\gamma^g(\nabla v) \in \boldsymbol{W}^{s-1-\frac{1}{p}}(\partial D)$. □

Repeated applications of Theorem 3.10 lead to the following important result to define the domain of various finite element interpolation operators (for simplicity we only consider integrability on the manifold).

Theorem 3.15 (Trace on low-dimensional manifolds). *Let $p \in [1, \infty)$ and let D be a Lipschitz domain in $\mathbb{R}^d$. Let M be a smooth, or polyhedral, manifold of dimension r in $\overline{D}$, $r \in \{0{:}d\}$. Then there is a bounded trace operator from $W^{s,p}(D)$ to $L^p(M)$ provided $sp > d - r$ (or $s \geq d - r$ if $p = 1$).*

When solving boundary value problems, one sometimes has to enforce a Neumann boundary condition which consists of prescribing the value of the normal derivative $\partial_n u := \boldsymbol{n} \cdot \nabla u$ at the boundary; see Chapter 31. Enforcing such a boundary condition is ambiguous if $\boldsymbol{n}$ is discontinuous. For instance, irrespective of the smoothness of the function in question, the normal derivative on polygons and polyhedra cannot be continuous. Let us start to address this problem by considering the simpler case where the boundary enjoys some additional smoothness property.

Theorem 3.16 (Normal derivative). *Let $p \in (1, \infty)$ and $s - \frac{1}{p} \in (1, 2)$. Let D be a domain in $\mathbb{R}^d$ with a boundary of class $C^{k,1}$, with $k := 1$ if $s \leq 2$ and $k := 2$ otherwise. There is a bounded linear map $\gamma^{\partial_n} : W^{s,p}(D) \to W^{s-1-\frac{1}{p},p}(\partial D)$ so that $\gamma^{\partial_n}(v) = (\boldsymbol{n} \cdot \nabla v)_{|\partial D}$ for all $v \in C^1(\overline{D})$, and letting $\gamma_1 := (\gamma^{\mathrm{g}}, \gamma^{\partial_n}) : W^{s,p}(D) \to W^{s-\frac{1}{p},p}(\partial D) \times W^{s-1-\frac{1}{p},p}(\partial D)$,*

(i) *The map γ_1 is bounded and surjective.*
(ii) *The kernel of γ_1 is $W_0^{s,p}(D)$.*

Proof. See Grisvard [110, Thm. 1.5.1.2] for the statement (i) and [110, Cor. 1.5.1.6] for the statement (ii). □

The above theorem can be extended to polygons ($d = 2$) as detailed in [110, Thm. 1.5.2.1]. The situation is more subtle when D is only Lipschitz. An extension of the notion of the normal derivative in this case is introduced in §4.3, and we refer the reader to Example 4.16 where $\boldsymbol{n} \cdot \nabla u$ is defined by duality.

3.2.2 The spaces $\widetilde{W}^{s,p}(D)$

We have seen that a function $v \in W^{s,p}(D)$ has a trace at the boundary ∂D if s is large enough. Another closely related question is whether the *zero-extension* of v to the whole space $\mathbb{R}^d$ belongs to $W^{s,p}(\mathbb{R}^d)$. For instance, the zero-extension to $\mathbb{R}^d$ of a test function $\varphi \in C_0^\infty(D)$ is in $C_0^\infty(\mathbb{R}^d)$. For every function $v \in L^1(D)$, we denote by $\widetilde{v}$ the extension by zero of v to $\mathbb{R}^d$, i.e., $\widetilde{v}(\boldsymbol{x}) := v(\boldsymbol{x})$ if $\boldsymbol{x} \in D$ and $\widetilde{v}(\boldsymbol{x}) := 0$ otherwise.

Definition 3.17 ($\widetilde{W}^{s,p}(D)$). *Let $s > 0$ and $p \in [1, \infty]$. Let D be an open subset of $\mathbb{R}^d$. We define*

$$\widetilde{W}^{s,p}(D) := \{v \in W^{s,p}(D) \mid \widetilde{v} \in W^{s,p}(\mathbb{R}^d)\}. \tag{3.4}$$

For $p = 2$, we write $\widetilde{H}^s(D) := \widetilde{W}^{s,2}(D)$.

Theorem 3.18 (Completion). $\widetilde{W}^{s,p}(D)$ *is a Banach space equipped with the norm* $\|v\|_{\widetilde{W}^{s,p}(D)} := \|\widetilde{v}\|_{W^{s,p}(\mathbb{R}^d)}$. *Moreover,* $\|v\|_{\widetilde{W}^{s,p}(D)} = \|v\|_{W^{s,p}(D)}$ *if* $s \in \mathbb{N}$. *If* $s \notin \mathbb{N}$ *and* D *is a Lipschitz domain in* $\mathbb{R}^d$, $\|v\|_{\widetilde{W}^{s,p}(D)}$ *is equivalent to the norm* $(\|v\|_{W^{s,p}(D)}^p + \ell_D^{sp} \sum_{|\alpha|=m} \int_D (\rho(\boldsymbol{x}))^{-\sigma p} |\partial^\alpha v|^p \, \mathrm{d}x)^{\frac{1}{p}}$, *where* $m := \lfloor s \rfloor$, $\sigma := s - m$, *and* ρ *is the distance to* ∂D, *i.e.,* $\rho(\boldsymbol{x}) := \inf_{\boldsymbol{y} \in \partial D} \|\boldsymbol{x} - \boldsymbol{y}\|_{\ell^2}$.

Proof. See Grisvard [110, Lem. 1.3.2.6], Tartar [189, Lem. 37.1]. □

Theorem 3.19 $\left(W^{s,p}(D), W_0^{s,p}(D), \widetilde{W}^{s,p}(D)\right)$**.** *Let* $s > 0$ *and* $p \in (1, \infty)$. *Let* D *be a Lipschitz domain in* $\mathbb{R}^d$. *The following holds true:*

$$W^{s,p}(D) = W_0^{s,p}(D) = \widetilde{W}^{s,p}(D) \qquad (sp < 1), \qquad (3.5a)$$

$$W^{s,p}(D) = W_0^{s,p}(D) \neq \widetilde{W}^{s,p}(D) \qquad (sp = 1), \qquad (3.5b)$$

$$W^{s,p}(D) \neq W_0^{s,p}(D) = \widetilde{W}^{s,p}(D) \qquad (sp > 1, s - \frac{1}{p} \notin \mathbb{N}). \quad (3.5c)$$

For all $sp > 1$, $W_0^{s,p}(D)$ *is a proper subspace of* $W^{s,p}(D)$. *(The above equalities mean that the sets coincide and the associated norms are equivalent, i.e., the topologies are identical.)*

Proof. See Grisvard [110, Thm. 1.4.2.4, Cor. 1.4.4.5], Tartar [189, Chap. 33], Lions and Magenes [135, Thm. 11.1]; see also Exercise 3.4 for a proof of the fact that $\widetilde{W}^{1,p}(D) \hookrightarrow W_0^{1,p}(D)$. □

Remark 3.20 $(D = \mathbb{R}^d)$**.** We have $W_0^{s,p}(\mathbb{R}^d) = W^{s,p}(\mathbb{R}^d) = \widetilde{W}^{s,p}(\mathbb{R}^d)$ for all $s > 0$ and all $p \in [1, \infty)$; see [110, p. 24], [189, Lem. 6.5]. □

Remark 3.21 (Embedding of $\widetilde{W}^{s,p}(D)$). The same conclusions as in Theorems 2.31 and 2.35 hold true for $\widetilde{W}^{s,p}(D)$ since the (s,p)-extension property is available. □

Remark 3.22 (Density). Let D be a Lipschitz domain in $\mathbb{R}^d$, $s > 0$, $p \in (1, \infty)$. Then $C_0^\infty(D)$ is dense in $\widetilde{W}^{s,p}(D)$; see [110, Thm. 1.4.2.2]. □

Remark 3.23 (Interpolation). Let $p \in [1, \infty)$, $s \in (0, 1)$, and recall that $W^{s,p}(D) = [L^p(D), W^{1,p}(D)]_{s,p}$ with equivalent norms; see Remark 2.20 and [189, Lem. 36.1]. Let us now define

$$W_{00}^{s,p}(D) := [L^p(D; \mathbb{R}^q), W_0^{1,p}(D)]_{s,p}. \qquad (3.6)$$

It is established in Chandler-Wilde et al. [65, Cor. 4.10] that for $p = 2$,

$$\widetilde{H}^s(D) = H_{00}^s(D). \qquad (3.7)$$

(More generally, we conjecture that $\widetilde{W}^{s,p}(D) = W_{00}^{s,p}(D)$.) The equality (3.7) together with Theorem 3.19 implies that $H_{00}^s(D) = H_0^s(D)$ if $s \neq \frac{1}{2}$. □

3.3 Poincaré–Steklov inequalities

We list here a series of functional inequalities that will be used repeatedly in the book; see Remark 3.32 for some historical background and some comments on the terminology.

Lemma 3.24 (Poincaré–Steklov). *Let D be a Lipschitz domain in $\mathbb{R}^d$. Let $\ell_D := \operatorname{diam}(D)$. Let $p \in [1, \infty]$. There is $C_{\mathrm{PS},p}$ (the subscript p is omitted when $p = 2$) s.t.*

$$C_{\mathrm{PS},p} \|v - \underline{v}_D\|_{L^p(D)} \le \ell_D |v|_{W^{1,p}(D)}, \qquad \forall v \in W^{1,p}(D), \qquad (3.8)$$

where $\underline{v}_D := \frac{1}{|D|} \int_D v \, \mathrm{d}x$. The following holds true when D is convex:

$$C_{\mathrm{PS},1} = 2, \quad C_{\mathrm{PS}} := C_{\mathrm{PS},2} = \pi, \quad C_{\mathrm{PS},p} \ge \frac{1}{2} \left(\frac{2}{p} \right)^{\frac{1}{p}}, \ p > 1. \qquad (3.9)$$

Remark 3.25 (Best constant). The values in (3.9) are proved in Acosta and Durán [2] for $p = 1$, in Bebendorf [17] for $p = 2$ (see also Payne and Weinberger [157] for the general idea), and in Chua and Wheeden [72, Thm. 1.2] for general p. The constants given in (3.9) for $p \in \{1, 2\}$ are the best possible. Uniform bounds on the Poincaré–Steklov constant for possibly nonconvex sets are a delicate issue; see Exercise 22.3 and Veeser and Verfürth [194]. □

Lemma 3.26 (Fractional Poincaré–Steklov). *Let $p \in [1, \infty)$ and $s \in (0, 1)$. Let D be a Lipschitz domain in $\mathbb{R}^d$. Let $\ell_D := \operatorname{diam}(D)$. Let us set $\underline{v}_D := \frac{1}{|D|} \int_D v \, \mathrm{d}x$. The following holds true:*

$$\|v - \underline{v}_D\|_{L^p(D)} \le \ell_D^s \left(\frac{\ell_D^d}{|D|} \right)^{\frac{1}{p}} |v|_{W^{s,p}(D)}. \qquad (3.10)$$

We also have $|v - \underline{v}_D|_{W^{r,p}(D)} = |v|_{W^{r,p}(D)} \le \ell_D^{s-r} |v|_{W^{s,p}(D)}$ for all $r \in (0, s]$.

Proof. A direct proof is proposed in Exercise 3.3 following [97, Lem. 7.1]. See also Dupont and Scott [92, Prop. 6.1] and Heuer [116]. The factor $\frac{\ell_D^d}{|D|}$ is often called eccentricity of D. □

Lemma 3.27 (Poincaré–Steklov). *Let $p \in [1, \infty]$ and let D be a Lipschitz domain. Let $\ell_D := \operatorname{diam}(D)$. There is $C_{\mathrm{PS},p} > 0$ (the subscript p is omitted when $p = 2$) such that*

$$C_{\mathrm{PS},p} \|v\|_{L^p(D)} \le \ell_D \|\nabla v\|_{\boldsymbol{L}^p(D)}, \qquad \forall v \in W_0^{1,p}(D). \qquad (3.11)$$

Proof. See Brezis [48, Cor. 9.19], Evans [99, Thm. 3, §5.6]. □

Remark 3.28 (Unit). The Poincaré–Steklov constant $C_{\mathrm{PS},p}$ is a dimensionless number. Its value remains unchanged if D is translated or rotated.

Moreover, assuming $\mathbf{0} \in D$, if $\tilde{D} = \lambda^{-1} D$ with $\lambda > 0$, the two domains D and $\tilde{D}$ have the same Poincaré–Steklov constant. $\qquad\square$

Remark 3.29 (Norm equivalence). The Poincaré–Steklov inequality implies that the seminorm $|\cdot|_{W^{1,p}(D)}$ is a norm equivalent to $\|\cdot\|_{W^{1,p}(D)}$ in $W_0^{1,p}(D)$. For instance, for the H^1-norm $\|v\|_{H^1(D)}^2 = \|v\|_{L^2(D)}^2 + \ell_D^2 |v|_{H^1(D)}^2$ (recall that $|v|_{H^1(D)} = \|\nabla v\|_{\boldsymbol{L}^2(D)}$), we obtain

$$\frac{C_{\mathrm{PS}}}{(1 + C_{\mathrm{PS}}^2)^{\frac{1}{2}}} \|v\|_{H^1(D)} \le \ell_D |v|_{H^1(D)} \le \|v\|_{H^1(D)}, \qquad \forall v \in H_0^1(D). \qquad\square$$

Lemma 3.30 (Extended Poincaré–Steklov). *Let $p \in [1, \infty)$ and let D be a Lipschitz domain in $\mathbb{R}^d$. Let $\ell_D := \mathrm{diam}(D)$. Let f be a bounded linear form on $W^{1,p}(D)$ whose restriction on constant functions is not zero. There is $\check{C}_{\mathrm{PS},p} > 0$ (the subscript p is omitted when $p = 2$) such that*

$$\check{C}_{\mathrm{PS},p} \|v\|_{L^p(D)} \le \ell_D \|\nabla v\|_{\boldsymbol{L}^p(D)} + |f(v)|, \qquad \forall v \in W^{1,p}(D). \qquad (3.12)$$

In particular, letting $\ker(f) := \{v \in W^{1,p}(D) \mid f(v) = 0\}$, we have

$$\check{C}_{\mathrm{PS},p} \|v\|_{L^p(D)} \le \ell_D \|\nabla v\|_{\boldsymbol{L}^p(D)}, \qquad \forall v \in \ker(f). \qquad (3.13)$$

Moreover, if $f(\mathbb{1}_D) = 1$ (where $\mathbb{1}_D$ is the indicator function of D), we have

$$\check{C}_{\mathrm{PS},p} \|v - f(v)\mathbb{1}_D\|_{L^p(D)} \le \ell_D \|\nabla v\|_{\boldsymbol{L}^p(D)}, \qquad \forall v \in W^{1,p}(D). \qquad (3.14)$$

Proof. We use the Peetre–Tartar lemma (Lemma A.20) to prove (3.12). Let $X := W^{1,p}(D)$, $Y := \boldsymbol{L}^p(D) \times \mathbb{R}$, $Z := L^p(D)$, and $A : X \ni v \mapsto (\nabla v, f(v)) \in Y$. Owing to Lemma 2.11 and the hypotheses on f, A is continuous and injective. Moreover, the embedding $X \hookrightarrow Z$ is compact owing to Theorem 2.35. This proves (3.12), and (3.13) is a direct consequence of (3.12). To prove (3.14), we apply (3.12) to the function $\tilde{v} := v - f(v)\mathbb{1}_D$. This function is in $\ker(f)$ since $f(\mathbb{1}_D) = 1$ and it satisfies $\nabla \tilde{v} = \nabla v$. $\qquad\square$

Example 3.31 (Zero mean-value). Lemma 3.30 can be applied with $f(v) := |U|^{-1} \int_U v \, \mathrm{d}x$, where U is a subset of D of nonzero measure (the boundedness of f follows from $|f(v)| \le |U|^{-\frac{1}{p}} \|v\|_{L^p(D)}$ by Hölder's inequality). One can also apply Lemma 3.30 with $f(v) := |\partial D_1|^{-1} \int_{\partial D_1} v \, \mathrm{d}s$, where ∂D_1 is a subset of ∂D of nonzero $(d-1)$-measure (the boundedness of f is a consequence of Theorem 3.10). $\qquad\square$

Remark 3.32 (Terminology). The inequality (3.8) is often called *Poincaré inequality* in the literature, and it is sometimes associated with other names like Wirtinger or Friedrichs. It turns out that Poincaré proved (3.8) for a convex domain in 1890 in [158] (the problem is formulated at the bottom of page 252, and the theorem is given at the bottom of page 258). Poincaré

refined his estimates of C_{PS} in 1894 in [159, p. 76] and gave $C_{\mathrm{PS}} \geq \frac{16}{9}$ for a three-dimensional convex domain. Without invoking the convexity assumption, he has also showed in [158] that the best constant C_{PS}^2 in the inequality

$$C_{\mathrm{PS}}^2 \|v\|_{L^2(D)}^2 \leq \ell_D^2(\alpha \|v\|_{L^2(\partial D)}^2 + |v|_{H^1(D)}^2), \quad \forall v \in H^1(D), \qquad (3.15)$$

is the smallest eigenvalue of the Laplacian supplemented with the Robin boundary condition $(\alpha v + \partial_n v)_{|\partial D} = 0$ (cf. statement in the middle of page 240: "and we must conclude that k_1 is the minimum of the ratio B/A" (in French)). The simplest form of (3.8) on an interval with $p = 2$ can be traced to the work of Steklov (see [184, Lem. 2, p. 156] for the Russian version published in 1897 with $C_{\mathrm{PS}} \geq \sqrt{2}$ and [182, pp. 294–295] for the 1901 French version with $C_{\mathrm{PS}} = \pi$ for functions that are either zero at both ends of the interval or are of zero mean). Steklov makes ample references to the work of Poincaré in each paper. He revisited the work of Poincaré on the spectrum of the Laplacian in [183, 185]. He proved in [183, Thm. VII, p. 66] and in [185, Thm. XIV, p. 107] that C_{PS}^2 in (3.11) is the smallest eigenvalue of the Laplacian supplemented with homogeneous Dirichlet boundary conditions. He reproved that C_{PS}^2 in (3.8) is the smallest eigenvalue of the Laplacian supplemented with homogeneous Neumann boundary conditions in [185, Thm. XV, p. 110]. A detailed survey of the literature on the best constant in (3.8) can be found in Kuznetsov and Nazarov [130]. Note that [183] is cited in [130] for the work of Steklov on the Laplacian with Neumann boundary condition, whereas the paper in question only deals with Dirichlet boundary conditions. For mysterious reasons, the paper by Friedrichs, Eine invariante Formulierung des Newtonschen Gravititationsgesetzes und des Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz. *Math. Ann.* 98 (1927), 566–575, is sometimes cited in the literature in relation to Poincaré's inequalities, including in [130], but the topic of this paper is not even remotely related to the Poincaré inequality. One early work of Friedrichs related to Poincaré's inequalities is a semi-discrete version of (3.15) published in Courant et al. [85, Eq. (13)]. Finally, it seems that the name of Wirtinger has been attached for the first time in 1916 to the inequality $\|f\|_{L^2(0,2\pi)} \leq |f|_{H^1(0,2\pi)}$ for periodic functions by Blaschke in his book [24, p. 105] without any specific reference. A little bit at odd with the rest of the literature, we henceforth adopt the Poincaré–Steklov terminology to refer to inequalities like (3.8) and (3.11). □

Exercises

Exercise 3.1 (Scaling). Let $D \subset \mathbb{R}^d$ be a Lipschitz domain. Let $\lambda > 0$ and $\tilde{D} := \lambda^{-1}D$. (i) Show that D and $\tilde{D}$ have the same Poincaré–Steklov constant in (3.8). (ii) Same question for (3.11).

Exercise 3.2 (Poincaré–Steklov, 1D). Let $D := (0,1)$ and $u \in C^1(D;\mathbb{R})$. Prove the following bounds: (i) $\|u\|_{L^2(D)}^2 \leq \frac{1}{2}\|u'\|_{L^2(D)}^2$ if $u(0) = 0$. (*Hint:* $u(x) = \int_0^x u'(t)\,\mathrm{d}t$.) (ii) $\|u\|_{L^2(D)}^2 \leq \frac{1}{\sqrt{8}}\|u'\|_{L^2(D)}^2$ if $u(0) = u(1) = 0$. (*Hint:* as above, but distinguish whether $x \in (0, \frac{1}{2})$ or $x \in (\frac{1}{2}, 1)$.) (iii) $\|u\|_{L^2(D)}^2 \leq \frac{1}{6}\|u'\|_{L^2(D)}^2 + \underline{u}^2$ with $\underline{u} := \int_0^1 u\,\mathrm{d}x$. (*Hint:* square the identity $u(x) - u(y) = \int_x^y u'(t)\,\mathrm{d}t$.) (iv) $\max_{x\in\overline{D}}|u(x)|^2 \leq 2u(1)^2 + 2\|u'\|_{L^2(D)}^2$. (*Hint:* square $u(x) = u(1) + \int_1^x u'(t)\,\mathrm{d}t$.) (v) $\max_{x\in\overline{D}}|u(x)|^2 \leq 2(\|u\|_{L^2(D)}^2 + \|u'\|_{L^2(D)}^2)$. (*Hint:* prove that $u(x)^2 \leq 2u(y)^2 + 2\|u'\|_{L^2(D)}^2$ and integrate over $y \in D$.)

Exercise 3.3 (Fractional Poincaré–Steklov). (i) Prove (3.10). (*Hint:* write $\int_D |v(\boldsymbol{x}) - \underline{v}_D|^p\mathrm{d}x = \int_D |D|^{-p}\big|\int_D (v(\boldsymbol{x}) - v(\boldsymbol{y}))\,\mathrm{d}y\big|^p\mathrm{d}x$.) (ii) Prove that $|v - \underline{v}_D|_{W^{r,p}(D)} \leq \ell_D^{s-r}|v|_{W^{s,p}(D)}$ for all $r \in (0, s]$ and all $s \in (0, 1)$.

Exercise 3.4 (Zero-extension in $W_0^{1,p}(D)$). Let $p \in [1, \infty)$. Let D be an open set in $\mathbb{R}^d$. Show that $W_0^{1,p}(D) \hookrightarrow \widetilde{W}^{1,p}(D)$ and $\|\tilde{u}\|_{W^{1,p}(\mathbb{R}^d)} \leq \|u\|_{W^{1,p}(D)}$ for all $u \in W_0^{1,p}(D)$.

Exercise 3.5 (Integral representation). Let $v : [0, \infty) \longrightarrow \mathbb{R}$ be a continuous function with bounded derivative, and let $w : [0, \infty) \longrightarrow \mathbb{R}$ be such that $w(x) := \frac{1}{x}\int_0^x (v(t) - v(x))\,\mathrm{d}t$. (i) Show that $|w(x)| \leq \frac{Mx}{2}$ where $M := \sup_{x\in[0,\infty)}|\partial_x v(x)|$. (ii) Estimate $w(0)$. (iii) Show that $\partial_t(tw(t)) = -t\partial_t v(t)$. (iv) Prove that $v(x) - v(0) = -w(x) - \int_0^x \frac{w(t)}{t}\,\mathrm{d}t$. (*Hint:* observe that $v(x) - v(0) = \int_0^x \frac{1}{t}(t\partial_t v(t))\,\mathrm{d}t$, use (iii), and integrate by parts.) (v) Prove the following integral representation formula (see Grisvard [110, pp. 29–30]):

$$v(0) = v(x) + \frac{1}{x}\int_0^x (v(t) - v(x))\,\mathrm{d}t + \int_0^x \frac{1}{y^2}\int_0^y (v(t) - v(y))\,\mathrm{d}t\,\mathrm{d}y.$$

Exercise 3.6 (Trace inequality in $W^{s,p}$, $sp > 1$). Let $s \in (0, 1)$, $p \in [1, \infty)$, and $sp > 1$. Let $a > 0$ and F be an open bounded subset of $\mathbb{R}^{d-1}$. Let $D := F \times (0, a)$. Let $v \in C^1(D) \cap C^0(\overline{D})$. (i) Let $\boldsymbol{y} \in F$. Using the integral representation from Exercise 3.5, show that there are $c_1(s,p)$ and $c_2(s,p)$ such that

$$|v(\boldsymbol{y}, 0)| \leq a^{-\frac{1}{p}}\|v(\boldsymbol{y}, .)\|_{L^p(0,a)} + (c_1(s,p) + c_2(s,p))a^{s-\frac{1}{p}}|v(\boldsymbol{y}, .)|_{W^{s,p}(0,a)}.$$

(ii) Accept as a fact that there is c (depending on s and p) such that

$$\int_F \int_0^a \int_0^a \frac{|v(\boldsymbol{x}_{d-1}, x_d) - v(\boldsymbol{x}_{d-1}, y_d)|^p}{|x_d - y_d|^{sp+1}}\,\mathrm{d}x_1 \ldots \mathrm{d}x_{d-1}\,\mathrm{d}x_d\,\mathrm{d}y_d \leq c\,|v|_{W^{s,p}(D)}.$$

Prove that $\|v(., 0)\|_{L^p(F)} \leq c'\,(a^{-\frac{1}{p}}\|v\|_{L^p(D)} + a^{s-\frac{1}{p}}|v|_{W^{s,p}(D)})$. *Note:* this shows that the trace operator $\gamma^\mathrm{g} : C^1(D) \cap C^0(\overline{D}) \to L^p(F)$ is bounded uniformly w.r.t. the norm of $W^{s,p}(D)$ when $sp > 1$. This means that γ^g can be extended to $W^{s,p}(D)$ since $C^1(D) \cap C^0(\overline{D})$ is dense in $W^{s,p}(D)$.

Part I, Chapter 4

Distributions and duality in Sobolev spaces

The dual space of a Sobolev space is not only composed of functions (defined almost everywhere), but this space also contains more sophisticated objects called distributions, which are defined by their action on smooth functions with compact support. For instance, the function $\frac{1}{x}$ is not in $L^1(0,1)$, but the map $\varphi \mapsto \int_0^1 \frac{1}{x} \varphi(x) \, \mathrm{d}x$ can be given a meaning for every smooth function that vanishes at 0. Dual Sobolev spaces are useful to handle singularities on the right-hand side of PDEs. They are also useful to give a meaning to the tangential and the normal traces of $\mathbb{R}^d$-valued fields that are not in $W^{s,p}(D; \mathbb{R}^d)$ with $sp > 1$. The extension is done in this case by invoking integration by parts formulas involving the curl or the divergence operators.

4.1 Distributions

The notion of distribution is a powerful tool that extends the concept of integrable functions and weak derivatives. In particular, we will see that every distribution is differentiable in some reasonable sense.

Definition 4.1 (Distribution). *Let D be an open set in $\mathbb{R}^d$. A linear map*

$$T : C_0^\infty(D) \ni \varphi \longmapsto \langle T, \varphi \rangle := T(\varphi) \in \mathbb{R} \ or \ \mathbb{C}, \qquad (4.1)$$

is called distribution *in D if for every compact subset K of D, there exist an integer p, called the* order *of T, and a real number c (both can depend on K) s.t. for all $\varphi \in C_0^\infty(D)$ with $supp(\varphi) \subset K$, we have*

$$|\langle T, \varphi \rangle| \leq c \max_{|\alpha| \leq p} \left(\ell_D^{|\alpha|} \|\partial^\alpha \varphi\|_{L^\infty(K)} \right). \qquad (4.2)$$

Let T be distribution of order p. We henceforth abuse the notation by using the symbol T to denote the extension by density of T to $C_0^p(D)$.

Example 4.2 (Locally integrable functions). Every function v in $L^1_{\mathrm{loc}}(D)$ can be identified with the following distribution:

$$T_v : C^\infty_0(D) \ni \varphi \longmapsto \langle T_v, \varphi \rangle := \int_D v\varphi \, \mathrm{d}x.$$

This identification is possible owing to Theorem 1.32, since two functions $v, w \in L^1_{\mathrm{loc}}(D)$ are such that $v = w$ a.e. in D iff $\int_D v\varphi \, \mathrm{d}x = \int_D w\varphi \, \mathrm{d}x$ for all $\varphi \in C^\infty_0(D)$. We will abuse the notation by writing v instead of T_v. Notice that the identification is also compatible with the Riesz–Fréchet theorem (Theorem 1.41) in $L^2(D)$, which allows one to identify $L^2(D)$ with its dual space by means of the L^2-inner product. $\qquad\square$

Example 4.3 (Dirac mass or measure). Let $\boldsymbol{a}$ be a point in D. The *Dirac mass* (or Dirac measure) at $\boldsymbol{a}$ is the distribution defined by $\langle \delta_{\boldsymbol{a}}, \varphi \rangle := \varphi(\boldsymbol{a})$ for all $\varphi \in C^\infty_0(D)$. There is no function $f \in L^1_{\mathrm{loc}}(D)$ such that $\delta_{\boldsymbol{a}} = T_f$. Otherwise, one would have $0 = \int_D f\varphi \, \mathrm{d}x$ for all $\varphi \in C^\infty_0(D\backslash\{\boldsymbol{a}\})$, and owing to Theorem 1.32, this would imply that $f = 0$ a.e. in $D\backslash\{\boldsymbol{a}\}$, i.e., $f = 0$ a.e. in D. Hence, $\delta_{\boldsymbol{a}} \notin T(L^1_{\mathrm{loc}}(D))$. This example shows that there are distributions that cannot be identified with functions in $L^1_{\mathrm{loc}}(D)$. $\qquad\square$

Definition 4.4 (Distributional derivative). *Let T be a distribution in D and let $i \in \{1{:}d\}$. The* distributional derivative $\partial_i T$ *is the distribution in D such that $\langle \partial_i T, \varphi \rangle := -\langle T, \partial_i \varphi \rangle$ for all $\varphi \in C^\infty_0(D)$. More generally, for a multi-index $\alpha \in \mathbb{N}^d$, the distributional derivative $\partial^\alpha T$ is the distribution in D acting as $\langle \partial^\alpha T, \varphi \rangle := (-1)^{|\alpha|}\langle T, \partial^\alpha \varphi \rangle$. We set conventionally $\partial^0 T := T$, and $\nabla T := (\partial_1 T, \ldots, \partial_d T)^{\mathsf{T}}$.*

Example 4.5 (Weak derivative). The notion of distributional derivative extends the notion of weak derivative. Let $v \in L^1_{\mathrm{loc}}(D)$ and assume that v has a weak α-th partial derivative, say $\partial^\alpha v \in L^1_{\mathrm{loc}}(D)$. Just like in Example 4.2, we can identify v and $\partial^\alpha v$ with the distributions T_v and $T_{\partial^\alpha v}$ such that $\langle T_v, \psi \rangle := \int_D v\psi \, \mathrm{d}x$ and $\langle T_{\partial^\alpha v}, \varphi \rangle := (-1)^{|\alpha|} \int_D v\partial^\alpha \varphi \, \mathrm{d}x$. This implies that $\langle T_{\partial^\alpha v}, \varphi \rangle = (-1)^{|\alpha|}\langle T_v, \partial^\alpha \varphi \rangle$, which according to Definition 4.4 shows that $\partial^\alpha T_v = T_{\partial^\alpha v}$, i.e., the distributional derivative of T_v is equal to the distribution associated with the weak derivative of v. $\qquad\square$

Example 4.6 (Step function). Let $D := (-1, 1)$. Let $w \in L^1(D)$ be defined by $w(x) := -1$ if $x < 0$ and $w(x) := 1$ otherwise. For all $\varphi \in C^\infty_0(D)$, we have $-\int_D w\partial_x \varphi \, \mathrm{d}x = \int_{-1}^0 \partial_x \varphi \, \mathrm{d}x - \int_0^1 \partial_x \varphi \, \mathrm{d}x = 2\varphi(0) = 2\langle \delta_0, \varphi \rangle$. This shows that the distributional derivative of w is twice the Dirac mass at 0, i.e., we write $\partial_x w = 2\delta_0$. As established in Example 4.3, δ_0 cannot be identified with any function in $L^1_{\mathrm{loc}}(D)$. Hence, w does not have a weak derivative but w has a distributional derivative. Consider now the function $v(x) := 1 - |x|$ in $L^1(D)$. By proceeding as in Example 2.5, one shows that v has a weak derivative and $\partial_x v(x) = 1$ if $x < 0$, and $\partial_x v(x) = -1$ otherwise. As established in Example 4.5, the distributional derivative of v and its weak

derivative coincide. Notice though that the distributional second derivative of v is $\partial_{xx}v = -2\delta_0$ which is not a weak derivative. $\square$

Example 4.7 (Dirac measure on the unit sphere). (i) Let $\boldsymbol{a} \in \mathbb{R}^d$. Definition 4.4 implies that $\langle \partial^\alpha \delta_{\boldsymbol{a}}, \varphi \rangle = (-1)^{|\alpha|} \partial^\alpha \varphi(\boldsymbol{a})$. (ii) Let $u : \mathbb{R}^d \to \mathbb{R}$ be such that $u(\boldsymbol{x}) := 1$ if $\|\boldsymbol{x}\|_{\ell^2} \leq 1$ and $u(\boldsymbol{x}) := 0$ otherwise. Let $B(\boldsymbol{0}, 1)$ and $S(\boldsymbol{0}, 1)$ be the unit ball and unit sphere in $\mathbb{R}^d$. We define the Dirac measure supported in $S(\boldsymbol{0}, 1)$ by $\langle \delta_{S(\boldsymbol{0},1)}, \varphi \rangle := \int_{S(\boldsymbol{0},1)} \varphi \, \mathrm{d}s$. Let $\boldsymbol{e}_i$ be one of the canonical unit vectors of $\mathbb{R}^d$. Then $\langle \partial_i u, \varphi \rangle = -\int_{B(\boldsymbol{0},1)} \partial_i \varphi \, \mathrm{d}x = -\int_{B(\boldsymbol{0},1)} \nabla \cdot (\varphi \boldsymbol{e}_i) \, \mathrm{d}x$, which proves that $\langle \partial_i u, \varphi \rangle = -\int_{S(\boldsymbol{0},1)} \boldsymbol{n} \cdot \boldsymbol{e}_i \varphi \, \mathrm{d}s$. Hence, $\nabla u = -\delta_{S(\boldsymbol{0},1)} \boldsymbol{n}$. $\square$

Definition 4.8 (Distributional convergence). *Let D be an open set in $\mathbb{R}^d$. We say that a sequence of distributions $\{T_n\}_{n\in\mathbb{N}}$ converges in the distribution sense if one has $\lim_{n\to\infty} \langle T_n, \varphi \rangle = \langle T, \varphi \rangle$ for all $\varphi \in \mathcal{C}_0^\infty(D)$.*

Example 4.9 (Oscillating functions). Let $D := (0, 1)$ and $f_n(x) := \sin(nx)$ for all $n \geq 1$. This sequence does not converge in $L^1(D)$, but $\langle T_{f_n}, \varphi \rangle = \int_0^1 \sin(nx)\varphi \, \mathrm{d}x = \int_0^1 \frac{1}{n} \cos(nx)\varphi' \, \mathrm{d}x$, so that $\lim_{n\to\infty} \langle T_{f_n}, \varphi \rangle = 0$ for all $\varphi \in C_0^\infty(D)$, i.e., $T_{f_n} \to 0$ in the sense of distributions. Up to an abuse of notation we say that f_n converges to 0 in the sense of distributions. Likewise one can show that $\cos(nx) \to 0$ in the sense of distributions. Let us now consider $g_n(x) := \sin^2(nx)$ for all $n \geq 1$. Using the identity $\sin^2(nx) = \frac{1}{2} - \frac{1}{2}\cos(2nx)$ and the above results, we conclude that $g_n \to \frac{1}{2}$ in the sense of distributions. $\square$

4.2 Negative-order Sobolev spaces

Equipped with the notion of distributions we can now define Sobolev spaces of negative order by duality using $W_0^{s,p}(D)$.

Definition 4.10 ($W^{-s,p}(D)$). *Let $s > 0$ and $p \in (1, \infty)$. Let D be an open set in $\mathbb{R}^d$. We define the space $W^{-s,p}(D) := \left(W_0^{s,p'}(D)\right)'$ with $\frac{1}{p} + \frac{1}{p'} = 1$ (for $p = 2$, we write $H^{-s}(D) := W^{-s,2}(D)$), equipped with the norm*

$$\|T\|_{W^{-s,p}(D)} := \sup_{w \in W_0^{s,p'}(D)} \frac{|\langle T, w \rangle|}{\|w\|_{W^{s,p'}(D)}}. \tag{4.3}$$

Identifying $L^p(D)$ with the dual space of $L^{p'}(D)$ (see Theorem 1.41), we infer that $L^p(D) \hookrightarrow W^{-s,p}(D)$ (and both spaces coincide for $s = 0$ since $W_0^{0,p'}(D) = L^{p'}(D)$ by Theorem 1.38). Moreover, any element $T \in W^{-s,p}(D)$ is a distribution since, assuming $s = m \in \mathbb{N}$, we have

$$|\langle T, \varphi \rangle| \leq \|T\|_{W^{-m,p}(D)} |D|^{\frac{1}{p'}} \binom{m+d}{d}^{\frac{1}{p'}} \max_{|\alpha| \leq m} \left(\ell_D^{|\alpha|} \|\partial^\alpha \varphi\|_{L^\infty(K)} \right), \quad (4.4)$$

for all compact subset $K \subsetneq D$ and all $\varphi \in C_0^\infty(D)$ with $\mathrm{supp}(\varphi) \subset K$. The argument can be adapted to the case where $s = m + \sigma$, $\sigma \in (0,1)$.

Example 4.11 (Dirac measure). Some of the objects in $W^{-s,p}(D)$ are not functions but distributions. For instance, the Dirac mass at a point $\boldsymbol{a} \in D$ is in $W^{-s,p}(D)$ if $sp' > d$. $\hfill\square$

Theorem 4.12 ($W^{-1,p}(D)$). *Let $p \in (1,\infty)$. Let D be an open, bounded set in $\mathbb{R}^d$. For all $f \in W^{-1,p}(D)$, there are functions $\{g_i\}_{i \in \{0:d\}}$, all in $L^{p'}(D)$, such that $\|f\|_{W^{-1,p}(D)} = \max_{i \in \{0:d\}} \|g_i\|_{L^{p'}(D)}$ and*

$$\langle f, v \rangle = \int_D g_0 v \, \mathrm{d}x + \sum_{i \in \{1:d\}} \int_D g_i \partial_i v \, \mathrm{d}x, \qquad \forall v \in W_0^{1,p}(D). \quad (4.5)$$

More generally, for all $m \in \mathbb{N}$, one has $v \in W^{-m,p}(D)$ if and only if $v = \sum_{|\alpha| \leq m} \partial^\alpha g_\alpha$ where $g_\alpha \in L^{p'}(D)$.

Proof. See Brezis [48, Prop. 9.20] for the case $m = 1$ and Adams and Fournier [3, Thm. 3.9]. $\hfill\square$

Example 4.13 (Gradient). Let $s \in (0,1)$, $p \in (1,\infty)$, and $sp \neq 1$. If D is a Lipschitz domain in $\mathbb{R}^d$, then the linear operator ∇ maps $W^{s,p}(D)$ boundedly to $\boldsymbol{W}^{s-1,p}(D)$, i.e., we have $\nabla \in \mathcal{L}(W^{s,p}(D); \boldsymbol{W}^{s-1,p}(D))$; see Grisvard [110, Thm. 1.4.4.6]. $\hfill\square$

Remark 4.14 (Interpolation). Assuming that D is a Lipschitz domain, an alternative definition of negative-order spaces relies on the interpolation theory between Banach spaces (see §A.5). Let $p \in (1,\infty)$ and $s \in (0,1)$. Recalling the space $W^{-1,p}(D)$ from Definition 4.10, let us set

$$\check{W}^{-s,p}(D) := [W^{-1,p}(D), L^p(D)]_{1-s,p}.$$

Theorem A.30 and the definition (3.6) of $W_{00}^{s,p'}(D)$ imply that

$$\check{W}^{-s,p}(D) = [L^{p'}(D), W_0^{1,p'}(D)]'_{s,p'} = \left(W_{00}^{s,p'}(D) \right)'.$$

The arguments from Remark 3.23 imply that $\check{H}^{-s}(D) = H^{-s}(D)$ if $s \neq \frac{1}{2}$ since $H_{00}^s(D) = H_0^s(D)$ in this case (see (3.7)). (One can also infer that $\check{W}^{-s,p}(D) = W^{-s,p}(D)$ for $sp \neq 1$, if $\widetilde{W}^{s,p}(D) = W_{00}^{s,p}(D)$, as conjectured in Remark 3.23.) $\hfill\square$

4.3 Normal and tangential traces

The goal of this section is to give a meaning to the normal or tangential component of $\mathbb{R}^d$-valued fields for which we only have integrability properties on the divergence or the curl, respectively, but not on the whole gradient. The underlying idea is quite general and consists of defining the traces in a Sobolev space of negative order at the boundary by extending a suitable integration by parts formula valid for smooth functions. Recall that for any field $\boldsymbol{v} = (v_i)_{i \in \{1:d\}} \in \boldsymbol{L}^1_{\text{loc}}(D) := L^1_{\text{loc}}(D; \mathbb{R}^d)$, the divergence is defined by

$$\nabla{\cdot}\boldsymbol{v} := \sum_{i \in \{1:d\}} \partial_i v_i, \tag{4.6}$$

and for $d = 3$, the curl $\nabla{\times}\boldsymbol{v}$ is the column vector in $\mathbb{R}^3$ with components $(\nabla{\times}\boldsymbol{v})_i := \sum_{j,k \in \{1:3\}} \varepsilon_{ijk} \partial_j v_k$ for all $i \in \{1:3\}$, where ε_{ijk} denotes the Levi-Civita symbol ($\varepsilon_{ijk} := 0$ if at least two indices take the same value, $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} := 1$ (i.e., for even permutations), and $\varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} := -1$ (i.e., for odd permutations)). In component form, we have

$$\nabla{\times}\boldsymbol{v} := (\partial_2 v_3 - \partial_3 v_2, \partial_3 v_1 - \partial_1 v_3, \partial_1 v_2 - \partial_2 v_1)^{\mathsf{T}}. \tag{4.7}$$

Recall also that the following integration by parts formulas hold true for all $\boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{C}^1(\overline{D})$ and all $q \in C^1(\overline{D})$:

$$\int_{\partial D} (\boldsymbol{v}{\times}\boldsymbol{n}){\cdot}\boldsymbol{w} \, \mathrm{d}s = \int_D \boldsymbol{v}{\cdot}\nabla{\times}\boldsymbol{w} \, \mathrm{d}x - \int_D (\nabla{\times}\boldsymbol{v}){\cdot}\boldsymbol{w} \, \mathrm{d}x, \tag{4.8a}$$

$$\int_{\partial D} (\boldsymbol{v}{\cdot}\boldsymbol{n})q \, \mathrm{d}s = \int_D \boldsymbol{v}{\cdot}\nabla q \, \mathrm{d}x + \int_D (\nabla{\cdot}\boldsymbol{v})q \, \mathrm{d}x. \tag{4.8b}$$

Let $p \in (1, \infty)$ and let us consider the following Banach spaces:

$$\boldsymbol{Z}^{\mathrm{c},p}(D) := \{\boldsymbol{v} \in \boldsymbol{L}^p(D) \mid \nabla{\times}\boldsymbol{v} \in \boldsymbol{L}^p(D)\}, \tag{4.9a}$$

$$\boldsymbol{Z}^{\mathrm{d},p}(D) := \{\boldsymbol{v} \in \boldsymbol{L}^p(D) \mid \nabla{\cdot}\boldsymbol{v} \in L^p(D)\}. \tag{4.9b}$$

For $p = 2$, we write

$$\boldsymbol{H}(\mathrm{curl}; D) := \boldsymbol{Z}^{\mathrm{c},2}(D), \qquad \boldsymbol{H}(\mathrm{div}; D) := \boldsymbol{Z}^{\mathrm{d},2}(D). \tag{4.10}$$

Let $\langle \cdot, \cdot \rangle_{\partial D}$ denote the duality pairing between $\boldsymbol{W}^{-\frac{1}{p},p}(\partial D)$ and $\boldsymbol{W}^{\frac{1}{p},p'}(\partial D)$. The trace operator $\gamma^{\mathrm{g}} : W^{1,p'}(D) \longrightarrow W^{\frac{1}{p},p'}(\partial D)$ being surjective (see Theorem 3.10), we infer that there is $c_{\gamma^{\mathrm{c}}}$ such that for all $\boldsymbol{l} \in \boldsymbol{W}^{\frac{1}{p},p'}(\partial D)$, there is $\boldsymbol{w}(\boldsymbol{l}) \in \boldsymbol{W}^{1,p'}(D)$ s.t. $\gamma^{\mathrm{g}}(\boldsymbol{w}(\boldsymbol{l})) = \boldsymbol{l}$ and $\|\boldsymbol{w}(\boldsymbol{l})\|_{\boldsymbol{W}^{1,p'}(D)} \leq c_{\gamma^{\mathrm{c}}} \|\boldsymbol{l}\|_{\boldsymbol{W}^{\frac{1}{p},p'}(\partial D)}$. We then define the linear map $\gamma^{\mathrm{c}} : \boldsymbol{Z}^{\mathrm{c},p}(D) \to \boldsymbol{W}^{-\frac{1}{p},p}(\partial D)$ by

$$\langle \gamma^{\mathrm{c}}(\boldsymbol{v}), \boldsymbol{l}\rangle_{\partial D} := \int_D \boldsymbol{v}\cdot\nabla\times\boldsymbol{w}(\boldsymbol{l})\,\mathrm{d}x - \int_D (\nabla\times\boldsymbol{v})\cdot\boldsymbol{w}(\boldsymbol{l})\,\mathrm{d}x, \tag{4.11}$$

for all $\boldsymbol{v} \in \boldsymbol{Z}^{\mathrm{c},p}(D)$ and all $\boldsymbol{l} \in \boldsymbol{W}^{\frac{1}{p},p'}(\partial D)$. Note that (4.8a) shows that $\gamma^{\mathrm{c}}(\boldsymbol{v}) = \boldsymbol{v}_{|\partial D}\times\boldsymbol{n}$ when $\boldsymbol{v}$ is smooth. A direct verification invoking Hölder's inequality shows that the map γ^{c} is bounded. Moreover, the definition (4.11) is independent of the choice of $\boldsymbol{w}(\boldsymbol{l})$; see Exercise 4.5.

We also define the linear map $\gamma^{\mathrm{d}} : \boldsymbol{Z}^{\mathrm{d},p}(D) \to W^{-\frac{1}{p},p}(\partial D)$ by

$$\langle \gamma^{\mathrm{d}}(\boldsymbol{v}), l\rangle_{\partial D} := \int_D \boldsymbol{v}\cdot\nabla q(l)\,\mathrm{d}x + \int_D (\nabla\cdot\boldsymbol{v})q(l)\,\mathrm{d}x, \tag{4.12}$$

for all $\boldsymbol{v} \in \boldsymbol{Z}^{\mathrm{d},p}(D)$ and all $l \in W^{\frac{1}{p},p'}(\partial D)$, where $q(l) \in W^{1,p'}(D)$ is such that $\gamma^{\mathrm{g}}(q(l)) = l$, and $\langle\cdot,\cdot\rangle_{\partial D}$ now denotes the duality pairing between $W^{-\frac{1}{p},p}(\partial D)$ and $W^{\frac{1}{p},p'}(\partial D)$. Reasoning as above, one can verify that: $\gamma^{\mathrm{d}}(\boldsymbol{v}) = \boldsymbol{v}_{|\partial D}\cdot\boldsymbol{n}$ when $\boldsymbol{v}$ is smooth; the map γ^{d} is bounded; the definition (4.12) is independent of the choice of $q(l)$.

Theorem 4.15 (Normal/tangential component).   *Let $p \in (1,\infty)$. Let D be a Lipschitz domain in $\mathbb{R}^d$. Let $\gamma^{\mathrm{c}} : \boldsymbol{Z}^{\mathrm{c},p}(D) \to \boldsymbol{W}^{-\frac{1}{p},p}(\partial D)$ and $\gamma^{\mathrm{d}} : \boldsymbol{Z}^{\mathrm{d},p}(D) \to W^{-\frac{1}{p},p}(\partial D)$ be defined in (4.11) and (4.12), respectively. The following holds true:*

(i) $\gamma^{\mathrm{c}}(\boldsymbol{v}) = \boldsymbol{v}_{|\partial D}\times\boldsymbol{n}$ *and* $\gamma^{\mathrm{d}}(\boldsymbol{v}) = \boldsymbol{v}_{|\partial D}\cdot\boldsymbol{n}$ *whenever $\boldsymbol{v}$ is smooth.*

(ii) γ^{d} *is surjective.*

(iii) *Density: setting $\boldsymbol{Z}_0^{\mathrm{c},p}(D) := \overline{\boldsymbol{C}_0^\infty(D)}^{\boldsymbol{Z}^{\mathrm{c},p}(D)}$, $\boldsymbol{Z}_0^{\mathrm{d},p}(D) := \overline{\boldsymbol{C}_0^\infty(D)}^{\boldsymbol{Z}^{\mathrm{d},p}(D)}$, we have*

$$\boldsymbol{Z}_0^{\mathrm{c},p}(D) = \ker(\gamma^{\mathrm{c}}), \qquad \boldsymbol{Z}_0^{\mathrm{d},p}(D) = \ker(\gamma^{\mathrm{d}}). \tag{4.13}$$

Proof. Item (i) is a simple consequence of the definition of γ^{c} and γ^{d}. See Tartar [189, Lem. 20.2] for item (ii) when $p = 2$. See [96, Thm. 4.7] for item (iii) (see also Exercise 23.9). □

Example 4.16 (Normal derivative). In the context of elliptic PDEs, one often deals with functions $v \in H^1(D)$ such that $\nabla\cdot(\nabla v) \in L^2(D)$. For these functions we have $\nabla v \in \boldsymbol{H}(\mathrm{div};D)$. Owing to Theorem 4.15 with $p = 2$, one can then give a meaning to the normal derivative of v at the boundary as $\gamma^{\mathrm{d}}(\nabla v) \in H^{-\frac{1}{2}}(\partial D)$. Assuming more smoothness on v, e.g., $v \in H^s(D)$, $s > \frac{3}{2}$, and some smoothness of ∂D, one can instead invoke Theorem 3.16 to infer that $\gamma^{\partial_n}(v) \in H^{s-\frac{3}{2}}(\partial D) \hookrightarrow L^2(\partial D)$, i.e., the normal derivative is integrable. However, this smoothness assumption is often too strong for elliptic PDEs, and one has to use $\gamma^{\mathrm{d}}(\nabla v)$ to define the normal derivative. □

Example 4.17 (Whitney's paradox). Let us show by a counterexample (see [199, p. 100]) that the normal component of a vector field with integrable divergence over D may not be integrable over ∂D. The two-dimensional field

$\boldsymbol{v}(x_1, x_2) := (\frac{-x_2}{x_1^2 + x_2^2}, \frac{x_1}{x_1^2 + x_2^2})^\mathsf{T}$ in $D := (0,1)^2$ satisfies $\|\boldsymbol{v}(\boldsymbol{x})\|_{\ell^2} = \|\boldsymbol{x}\|_{\ell^2}^{-1}$, $\boldsymbol{v} \in \boldsymbol{L}^p(D)$ for all $p \in [1, 2)$, and $\nabla \cdot \boldsymbol{v} = 0$. However, $\boldsymbol{v} \cdot \boldsymbol{n}$ is not integrable, i.e., $\boldsymbol{v} \cdot \boldsymbol{n} \notin L^1(\partial D)$. $\qquad \square$

Remark 4.18 (2D). In dimension two ($d = 2$), the tangential component is defined using the linear map $\gamma^\mathrm{c} : \boldsymbol{Z}^{\mathrm{c},p}(D) \to W^{-\frac{1}{p}, p}(\partial D)$ as follows:

$$\langle \gamma^\mathrm{c}(\boldsymbol{v}), l \rangle_{\partial D} := \int_D \boldsymbol{v} \cdot \nabla^\perp w(l) \, \mathrm{d}x + \int_D (\nabla \times \boldsymbol{v}) w(l) \, \mathrm{d}x,$$

for all $\boldsymbol{v} \in \boldsymbol{Z}^{\mathrm{c},p}(D)$ and all $l \in W^{\frac{1}{p}, p'}(\partial D)$, where $w(l) \in W^{1,p'}(D)$ is such that $\gamma^\mathrm{g}(w(l)) = l$. Here, $\nabla^\perp v := (-\partial_2 v, \partial_1 v)^\mathsf{T}$ and $\nabla \times \boldsymbol{v} := \partial_1 v_2 - \partial_2 v_1$. Note that $\nabla^\perp v = \boldsymbol{R}_{\frac{\pi}{2}}(\nabla v)$ and $\nabla \times \boldsymbol{v} = -\nabla \cdot (\boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{v}))$, where $\boldsymbol{R}_{\frac{\pi}{2}}$ is the rotation of angle $\frac{\pi}{2}$ in $\mathbb{R}^2$ (i.e., the matrix of $\boldsymbol{R}_{\frac{\pi}{2}}$ relative to the canonical basis of $\mathbb{R}^2$ is $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$). Whenever $\boldsymbol{v}$ is smooth, we have $\gamma^\mathrm{c}(\boldsymbol{v}) = \boldsymbol{v}_{|\partial D} \cdot \boldsymbol{t}$ where $\boldsymbol{t} := \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{n})$ is a unit tangent vector to ∂D. $\qquad \square$

Exercises

Exercise 4.1 (Distributions). Let D be an open set in $\mathbb{R}^d$. Let v be a distribution in D. (i) Let $\psi \in C^\infty(D)$. Show that the map $C_0^\infty(D) \ni \varphi \mapsto \langle v, \psi\varphi \rangle$ defines a distribution in D (this distribution is usually denoted by ψv). (ii) Let $\alpha, \beta \in \mathbb{N}^d$. Prove that $\partial^\alpha(\partial^\beta v) = \partial^\beta(\partial^\alpha v)$ in the distribution sense.

Exercise 4.2 (Dirac measure on a manifold). Let D be a smooth bounded and open set in $\mathbb{R}^d$. Let $u \in C^2(D; \mathbb{R})$ and assume that $u_{|\partial D} = 0$. Let $\tilde{u}$ be the extension by zero of u over $\mathbb{R}^d$. Compute $\nabla \cdot (\nabla \tilde{u}) = \partial_{11} u + \ldots + \partial_{dd} u$ in the distribution sense.

Exercise 4.3 (P.V. $\frac{1}{x}$). Let $D := (-1, 1)$. Prove that the linear map $T : C_0^\infty(D) \to \mathbb{R}$ defined by $\langle T, \varphi \rangle := \lim_{\epsilon \to 0} \int_{|x| > |\epsilon|} \frac{1}{x} \varphi(x) \, \mathrm{d}x$ is a distribution.

Exercise 4.4 (Integration by parts). Prove the two identities in (4.8) by using the divergence formula $\int_D \nabla \cdot \boldsymbol{\phi} \, \mathrm{d}x = \int_{\partial D} (\boldsymbol{\phi} \cdot \boldsymbol{n}) \, \mathrm{d}s$ for all $\boldsymbol{\phi} \in \boldsymbol{C}^1(\overline{D})$.

Exercise 4.5 (Definition (4.11)). Verify that the right-hand side of (4.11) is independent of the choice of $\boldsymbol{w}(\boldsymbol{l})$. (*Hint*: consider two functions $\boldsymbol{w}_1, \boldsymbol{w}_2 \in \boldsymbol{W}^{1,p'}(D)$ s.t. $\gamma^\mathrm{g}(\boldsymbol{w}_1) = \gamma^\mathrm{g}(\boldsymbol{w}_2) = \boldsymbol{l}$ and use the density of $\boldsymbol{C}_0^\infty(D)$ in $\boldsymbol{W}_0^{1,p'}(D)$.)

Part II
Introduction to finite elements

Part II, Chapter 5

Main ideas and definitions

The goal of the three chapters composing Part II is to introduce the main concepts behind finite elements and to present various examples. This chapter introduces key notions such as degrees of freedom, shape functions, and interpolation operator. These notions are illustrated on Lagrange finite elements and modal finite elements, for which the degrees of freedom are values at specific nodes and moments against specific test functions, respectively.

5.1 Introductory example

This section introduces the notion of finite element in dimension one. Let $K := [-1, 1]$ and consider a continuous function $v \in C^0(K)$. Our objective is to devise an interpolation operator that approximates v in a finite-dimensional functional space, say P. For simplicity, we assume that $P = \mathbb{P}_k$ for some integer $k \geq 0$, where $\mathbb{P}_k$ is the real vector space composed of univariate polynomial functions of degree at most k, i.e., $p \in \mathbb{P}_k$ if $p(t) = \sum_{i \in \{0:k\}} \alpha_i t^i$ for all $t \in \mathbb{R}$, with $\alpha_i \in \mathbb{R}$ for every integer $i \in \{0:k\}$.

Let us consider $(k+1)$ distinct points $\{a_i\}_{i \in \{0:k\}}$ in K, which we call *nodes*. We want to construct an operator $\mathcal{I}_K : C^0(K) \to \mathbb{P}_k$ s.t. $\mathcal{I}_K(v)$ verifies

$$\mathcal{I}_K(v) \in \mathbb{P}_k, \qquad \mathcal{I}_K(v)(a_i) := v(a_i), \ \forall i \in \{0:k\}, \tag{5.1}$$

for every function $v \in C^0(K)$. These conditions uniquely determine $\mathcal{I}_K(v)$ since a polynomial in $\mathbb{P}_k$ is uniquely determined by the value it takes at $(k+1)$ distinct points in $\mathbb{R}$. For the same reason $\mathbb{P}_k$ is pointwise invariant under $\mathcal{I}_K$, i.e., $\mathcal{I}_K(p) = p$ for all $p \in \mathbb{P}_k$. To obtain an explicit representation of $\mathcal{I}_K(v)$, we introduce the Lagrange interpolation polynomials defined as follows:

$$\mathcal{L}_i^{[a]}(t) := \frac{\prod_{j \in \{0:k\}\setminus\{i\}}(t - a_j)}{\prod_{j \in \{0:k\}\setminus\{i\}}(a_i - a_j)}, \qquad \forall t \in \mathbb{R}, \qquad \forall i \in \{0:k\}. \tag{5.2}$$

We set $\mathcal{L}_0^{[a]} := 1$ if $k = 0$. By construction, the Lagrange interpolation polynomials satisfy $\mathcal{L}_i^{[a]}(a_i) = 1$ and $\mathcal{L}_i^{[a]}(a_j) = 0$ for all $j \neq i$, which we write concisely as

$$\mathcal{L}_i^{[a]}(a_j) = \delta_{ij}, \qquad \forall i, j \in \{0{:}k\}, \tag{5.3}$$

where δ_{ij} is the Kronecker symbol, i.e., $\delta_{ij} := 1$ if $i = j$ and $\delta_{ij} := 0$ otherwise. The Lagrange interpolation polynomials of degree $k \in \{1, 2, 3\}$ using equidistant nodes in K (including both endpoints) are shown in Figure 5.1. Let us show that the family $\{\mathcal{L}_i^{[a]}\}_{i \in \{0{:}k\}}$ forms a basis of $\mathbb{P}_k$. Since $\dim(\mathbb{P}_k) = k+1$, we only need to show linear independence. Assume that $\sum_{i \in \{0{:}k\}} \alpha_i \mathcal{L}_i^{[a]} = 0$. Evaluating this linear combination at the nodes $\{a_i\}_{i \in \{0{:}k\}}$ yields $\alpha_i = 0$ for all $i \in \{1{:}d\}$, which proves the assertion. In conclusion, the polynomial function $\mathcal{I}_K(v)$ defined in (5.1) is $\mathcal{I}_K(v)(t) := \sum_{i \in \{0{:}k\}} v(a_i) \mathcal{L}_i^{[a]}(t)$.
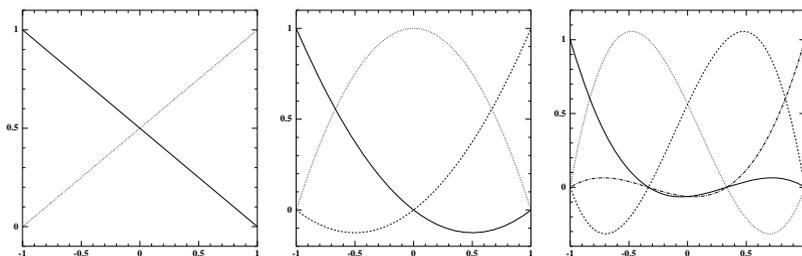
Fig. 5.1 Lagrange interpolation polynomials with equidistant nodes in the interval $K :=$ $[-1, 1]$ of degree $k = 1$ (left), 2 (center), and 3 (right).

Remark 5.1 (Key concepts). To sum up, we used three important ingredients to build the interpolation operator $\mathcal{I}_K$: the interval $K := [-1, 1]$, the finite-dimensional space $P := \mathbb{P}_k$, and a set of degrees of freedom, i.e., linear maps $\{\sigma_i\}_{i \in \{0{:}k\}}$ acting on continuous functions, which consist of evaluations at the nodes $\{a_i\}_{i \in \{0{:}k\}}$, i.e., $\sigma_i(v) := v(a_i)$. A key observation concerning the degrees of freedom is that they uniquely determine functions in P. $\square$

5.2 Finite element as a triple

A *polyhedron* (also called *polytope*) in $\mathbb{R}^d$ is a compact interval if $d = 1$ and if $d \geq 2$, it is a compact, connected subset of $\mathbb{R}^d$ with nonempty interior such that its boundary is a finite union of images by affine mappings of polyhedra in $\mathbb{R}^{d-1}$. In $\mathbb{R}^2$, a polyhedron is also called polygon. Simple examples are presented in Figure 5.2 in dimensions two and three. A polyhedron in $\mathbb{R}^2$ (resp., $\mathbb{R}^3$) can always be described as a finite union of triangles (resp., tetrahedra).
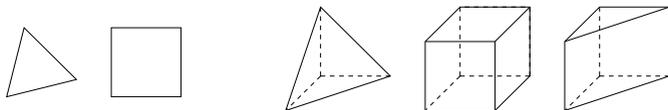
Fig. 5.2 Examples of polyhedra in $\mathbb{R}^2$ and $\mathbb{R}^3$. The hidden edges are shown with dashed lines in $\mathbb{R}^3$. From left to right: triangle, square, tetrahedron, hexahedron, prism.

The following definition of a finite element is due to Ciarlet [76, p. 93].

Definition 5.2 (Finite element). *Let $d \geq 1$, an integer $n_{\mathrm{sh}} \geq 1$, and the set $\mathcal{N} := \{1{:}n_{\mathrm{sh}}\}$. A* finite element *consists of a triple (K, P, Σ) where*

(i) *K is a* polyhedron *in $\mathbb{R}^d$ or the image of a polyhedron in $\mathbb{R}^d$ by some smooth diffeomorphism. More generally, K could be the closure of a Lipschitz domain in $\mathbb{R}^d$ (see §3.1). K is nontrivial, i.e., $\mathrm{int}(K) \neq \emptyset$.*

(ii) *P is a finite-dimensional vector space of functions $p : K \to \mathbb{R}^q$ for some integer $q \geq 1$ (typically $q \in \{1, d\}$). P is nontrivial, i.e., $P \neq \{0\}$. The members of P are polynomial functions, possibly composed with some smooth diffeomorphism.*

(iii) *Σ is a set of n_{sh} linear forms from P to $\mathbb{R}$, say $\Sigma := \{\sigma_i\}_{i \in \mathcal{N}}$, such that the map $\Phi_\Sigma : P \to \mathbb{R}^{n_{\mathrm{sh}}}$ defined by $\Phi_\Sigma(p) := \big(\sigma_i(p)\big)_{i \in \mathcal{N}}$ is an isomorphism. The linear forms σ_i are called* degrees of freedom *(in short* dofs*), and the bijectivity of the map Φ_Σ is referred to as* unisolvence.

Remark 5.3 (Proving unisolvence). To prove unisolvence, it suffices to show that $\dim P \geq n_{\mathrm{sh}} = \mathrm{card}\,\Sigma$ and that Φ_Σ is injective, i.e.,

$$[\,\sigma_i(p) = 0,\ \forall i \in \mathcal{N}\,] \implies [\,p = 0\,], \qquad \forall p \in P. \tag{5.4}$$

Owing to the rank nullity theorem, Φ_Σ is then bijective and $\dim P = n_{\mathrm{sh}}$. $\square$

Remark 5.4 ($\mathcal{L}(P; \mathbb{R})$). Σ is a basis of the space of the linear forms over P, i.e., $\mathcal{L}(P; \mathbb{R})$. Indeed, $\dim(\mathcal{L}(P; \mathbb{R})) = \dim(P) = n_{\mathrm{sh}}$. Moreover, if the vector $X = (X_i)_{i \in \mathcal{N}} \in \mathbb{R}^{n_{\mathrm{sh}}}$ is s.t. $\sum_{i \in \mathcal{N}} X_i \sigma_i(p) = 0$ for all $p \in P$, taking $p := \Phi_\Sigma^{-1}(X)$ yields $\sum_{i \in \mathcal{N}} X_i^2 = 0$. Hence, $X_i = 0$ for all $i \in \mathcal{N}$. $\square$

Proposition 5.5 (Shape functions). (i) *There is a basis $\{\theta_i\}_{i \in \mathcal{N}}$ of P s.t.*

$$\sigma_i(\theta_j) := \delta_{ij}, \qquad \forall i, j \in \mathcal{N}. \tag{5.5}$$

The functions θ_i are called shape functions. (ii) *Let $\{\phi_i\}_{i \in \mathcal{N}}$ be a basis of P. Then defining the* generalized Vandermonde matrix *$\mathcal{V} \in \mathbb{R}^{n_{\mathrm{sh}} \times n_{\mathrm{sh}}}$ with entries $\mathcal{V}_{ij} := \sigma_j(\phi_i)$ for all $i, j \in \mathcal{N}$, the shape functions are given by*

$$\theta_i = \sum_{j \in \mathcal{N}} (\mathcal{V}^{-1})_{ij} \phi_j, \qquad \forall i \in \mathcal{N}. \tag{5.6}$$

Proof. (i) The shape functions are given by $\theta_i = \Phi_\Sigma^{-1}(e_i)$ for all $i \in \mathcal{N}$, where $(e_i)_{i \in \mathcal{N}}$ is the canonical basis of $\mathbb{R}^{n_{\mathrm{sh}}}$. (ii) To show that the matrix $\mathcal{V}$ is invertible, we consider $X \in \mathbb{R}^{n_{\mathrm{sh}}}$ s.t. $X^\mathsf{T} \mathcal{V} = 0$ and set $p := \sum_{i \in \mathcal{N}} X_i \phi_i$. Then $X^\mathsf{T} \mathcal{V} = 0$ implies that $\sigma_j(p) = 0$ for all $j \in \mathcal{N}$, and (5.4) in turn implies that $p = 0$. Hence, $X = 0$ since $\{\phi_i\}_{i \in \mathcal{N}}$ is a basis of P. Finally, $\sigma_k(\sum_{j \in \mathcal{N}} (\mathcal{V}^{-1})_{ij} \phi_j) = \sum_{j \in \mathcal{N}} (\mathcal{V}^{-1})_{ij} \sigma_k(\phi_j) = \sum_{j \in \mathcal{N}} (\mathcal{V}^{-1})_{ij} \mathcal{V}_{jk} = \delta_{ik}$ for all $k \in \mathcal{N}$. This proves that $\theta_i = \sum_{j \in \mathcal{N}} (\mathcal{V}^{-1})_{ij} \phi_j$. $\square$

Proposition 5.5 gives a practical recipe to build the shape functions. One first chooses a basis of P and evaluates the associated Vandermonde matrix $\mathcal{V}$ and its inverse. The components of the shape function θ_i in the chosen basis are then $((\mathcal{V}^{-1})_{ij})_{j \in \mathcal{N}}$ for all $i \in \mathcal{N}$. One must be careful in choosing the basis $\{\phi_i\}_{i \in \mathcal{N}}$ when working with high-order polynomials, since the matrix $\mathcal{V}$ may become ill-conditioned if the basis is not chosen properly. The computation of the shape functions can be affected by roundoff errors if $\mathcal{V}$ is ill-conditioned.

Remark 5.6 (Vandermonde matrix). For $d = 1$, if one uses the monomial basis $\{x^i\}_{i \in \mathcal{N}}$ with the dofs $\sigma_i(p) := p(a_i)$, then $\mathcal{V}$ is a classical Vandermonde matrix with entries $\mathcal{V}_{ij} = a_j^i$ for all $i, j \in \mathcal{N}$. $\square$

5.3 Interpolation: finite element as a quadruple

The notion of interpolation operator is central to the finite element theory. The term "interpolation" is used here in a broad sense, since the degrees of freedom (dofs) are not necessarily point evaluations. For the interpolation operator to be useful, one needs to extend the domain of the linear forms in Σ so that they can act on functions in a space larger than P, which we denote by $V(K)$. The space $V(K)$ is the fourth ingredient defining a finite element.

Definition 5.7 (Interpolation operator). *Let (K, P, Σ) be a finite element. Assume that there exists a Banach space $V(K) \subset L^1(K; \mathbb{R}^q)$ s.t.:*

(i) *$P \subset V(K)$.*
(ii) *The linear forms $\{\sigma_i\}_{i \in \mathcal{N}}$ can be extended to $\mathcal{L}(V(K); \mathbb{R})$, i.e., there exist $\{\tilde{\sigma}_i\}_{i \in \mathcal{N}}$ and c_Σ such that $\tilde{\sigma}_i(p) = \sigma_i(p)$ for all $p \in P$, and $|\tilde{\sigma}_i(v)| \leq c_\Sigma \|v\|_{V(K)}$ for all $v \in V(K)$ and all $i \in \mathcal{N}$. We henceforth abuse the notation and use the symbol σ_i instead of $\tilde{\sigma}_i$.*

We define the interpolation operator *$\mathcal{I}_K : V(K) \to P$ by setting*

$$\mathcal{I}_K(v)(\boldsymbol{x}) := \sum_{i \in \mathcal{N}} \sigma_i(v) \theta_i(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K, \tag{5.7}$$

for all $v \in V(K)$. $V(K)$ is the domain of $\mathcal{I}_K$, and P is its codomain.

Proposition 5.8 (Boundedness). $\mathcal{I}_K$ *belongs to* $\mathcal{L}(V(K); P)$.

Proof. Let $\|\cdot\|_P$ be a norm in P (all the norms are equivalent in the finite-dimensional space P). The triangle inequality and Definition 5.7(ii) imply that $\|\mathcal{I}_K(v)\|_P \le (c_\Sigma \sum_{i\in\mathcal{N}} \|\theta_i\|_P)\|v\|_{V(K)}$ for all $v \in V(K)$. $\qquad\square$

Proposition 5.9 (P-**invariance**). *P is pointwise invariant under $\mathcal{I}_K$, i.e., $\mathcal{I}_K(p) = p$ for all $p \in P$. As a result, $\mathcal{I}_K$ is a projection, i.e., $\mathcal{I}_K{\circ}\mathcal{I}_K = \mathcal{I}_K$.*

Proof. Letting $p = \sum_{j\in\mathcal{N}} \alpha_j \theta_j$ yields $\mathcal{I}_K(p) = \sum_{i,j\in\mathcal{N}} \alpha_j \sigma_i(\theta_j)\theta_i = p$ owing to (5.5). This shows that P is pointwise invariant under $\mathcal{I}_K$, and it immediately follows that $\mathcal{I}_K$ is a projection. $\qquad\square$

Example 5.10 ($V(K)$). If one builds $\mathcal{I}_K(v)$ by using values of the function v at some points in K, like we did in §5.1, then it is natural to set $V(K) := C^0(K; \mathbb{R}^q)$ (recall that K is a closed set in $\mathbb{R}^d$, so that functions in $C^0(K; \mathbb{R}^q)$ are continuous up to the boundary). Another possibility is to set $V(K) := W^{s,p}(K; \mathbb{R}^q)$ for some real numbers $s \ge 0$ and $p \in [1, \infty]$ such that $sp > d$ (or $s \ge d$ if $p = 1$); see Theorem 2.31. If $\mathcal{I}_K(v)$ involves integrals over the faces of K, then one can take $V(K) := W^{s,p}(K; \mathbb{R}^q)$ with $sp > 1$ (or $s \ge 1$ if $p = 1$). More generally, if $\mathcal{I}_K(v)$ involves integrals over manifolds of codimension d', then it is legitimate to set $V(K) := W^{s,p}(K; \mathbb{R}^q)$ with $sp > d'$ (or $s \ge d'$ if $p = 1$). We abuse the notation since we should write $W^{s,p}(\mathrm{int}(K); \mathbb{R}^q)$, where $\mathrm{int}(K)$ denotes the interior of the set K in $\mathbb{R}^d$. $\qquad\square$

5.4 Basic examples

5.4.1 Lagrange (nodal) finite elements

The dofs of scalar-valued Lagrange (or nodal) finite elements are point values. The extension to vector-valued Lagrange elements is done by proceeding componentwise.

Definition 5.11 (Lagrange finite element). *Let (K, P, Σ) be a scalar-valued finite element ($q := 1$ in Definition 5.2). If there is a set of points $\{\boldsymbol{a}_i\}_{i\in\mathcal{N}}$ in K such that for all $i \in \mathcal{N}$,*

$$\sigma_i(p) := p(\boldsymbol{a}_i), \qquad \forall p \in P, \tag{5.8}$$

the triple (K, P, Σ) is called Lagrange *finite element. The points $\{\boldsymbol{a}_i\}_{i\in\mathcal{N}}$ are called* nodes, *and the shape functions $\{\theta_i\}_{i\in\mathcal{N}}$, which are s.t.*

$$\theta_i(\boldsymbol{a}_j) := \delta_{ij}, \qquad \forall i, j \in \mathcal{N}, \tag{5.9}$$

form the nodal basis *of P associated with the nodes $\{\boldsymbol{a}_i\}_{i\in\mathcal{N}}$.* $\qquad\square$

Examples are presented in Chapters 6 and 7. Following Definition 5.7, the *Lagrange interpolation operator* $\mathcal{I}_K^{\mathrm{L}}$ acts as follows:

$$\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{x}) := \sum_{i \in \mathcal{N}} v(\boldsymbol{a}_i)\theta_i(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K. \tag{5.10}$$

By construction, $\mathcal{I}_K^{\mathrm{L}}(v)$ matches the values of v at all the Lagrange nodes, i.e., $\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{a}_j) = v(\boldsymbol{a}_j)$ for all $j \in \mathcal{N}$. The domain of $\mathcal{I}_K^{\mathrm{L}}$ can be $V(K) := C^0(K)$ or $V(K) := W^{s,p}(K)$ with $p \in [1,\infty]$ and $ps > d$ ($s \geq d$ if $p = 1$).

5.4.2 Modal finite elements

The dofs of modal finite elements are moments against test functions using some measure over K. For simplicity, we consider the uniform measure and work in $L^2(K;\mathbb{R}^q)$ with $q \geq 1$. We are going to use the notation $(v,w)_{L^2(K;\mathbb{R}^q)} := \int_K (v,w)_{\ell^2(\mathbb{R}^q)}\, \mathrm{d}x$.

Proposition 5.12 (Modal finite element). *Let K be as in Definition 5.2. Let P be a finite-dimensional subspace of $L^2(K;\mathbb{R}^q)$ and let $\{\zeta_i\}_{i\in\mathcal{N}}$ be a basis of P. Let $\Sigma := \{\sigma_i\}_{i\in\mathcal{N}}$ be composed of the following linear forms $\sigma_i : P \to \mathbb{R}$:*

$$\sigma_i(p) := |K|^{-1}(\zeta_i, p)_{L^2(K;\mathbb{R}^q)}, \qquad \forall p \in P, \; \forall i \in \mathcal{N}. \tag{5.11}$$

(The factor $|K|^{-1}$ is meant to make σ_i independent of the size of K.) Then the triple (K, P, Σ) is a finite element called modal finite element.

Proof. We use Remark 5.3. By definition, $\dim(P) = \mathrm{card}(\Sigma)$. Let $p \in P$ be such that $\sigma_i(p) = 0$ for all $i \in \mathcal{N}$. Writing $p = \sum_{i\in\mathcal{N}} \alpha_i \zeta_i$, we infer that $|K|^{-1}\|p\|_{L^2(K;\mathbb{R}^q)}^2 = \sum_{j\in\mathcal{N}} \alpha_j \sigma_j(p) = 0$, so that $p = 0$. $\qquad\square$

Examples of modal finite elements are presented in Chapter 6. Let us introduce the *mass matrix* $\mathcal{M}$ of order n_{sh} with entries

$$\mathcal{M}_{ij} := |K|^{-1}(\zeta_i, \zeta_j)_{L^2(K;\mathbb{R}^q)}, \qquad \forall i,j \in \mathcal{N}. \tag{5.12}$$

By construction, $\mathcal{M}$ is symmetric, and since

$$(\mathcal{M}X, X)_{\ell^2(\mathbb{R}^{n_{\mathrm{sh}}})} = \sum_{i,j\in\mathcal{N}} \mathcal{M}_{ij} X_i X_j = |K|^{-1}\|\xi\|_{L^2(K;\mathbb{R}^q)}^2,$$

for all $X \in \mathbb{R}^{n_{\mathrm{sh}}}$ with $\xi = \sum_{j\in\mathcal{N}} X_j \zeta_j$, we infer that $(\mathcal{M}X, X)_{\ell^2(\mathbb{R}^{n_{\mathrm{sh}}})} \geq 0$. Moreover, $(\mathcal{M}X, X)_{\ell^2(\mathbb{R}^{n_{\mathrm{sh}}})} = 0$ implies $\xi = 0$, i.e., $X = 0$ since $\{\zeta_i\}_{i\in\mathcal{N}}$ is a basis of P. In conclusion, $\mathcal{M}$ is symmetric positive definite. Furthermore, one readily sees that $\mathcal{M} = \mathcal{V}$, where the Vandermonde matrix $\mathcal{V}$ is defined in Proposition 5.5. Hence, $\theta_i = \sum_{j\in\mathcal{N}}(\mathcal{M}^{-1})_{ij}\zeta_j$ for all $i \in \mathcal{N}$. Following Definition 5.7, the *modal interpolation operator* $\mathcal{I}_K^{\mathrm{m}}$ acts as follows:

$$\mathcal{I}_K^{\mathrm{m}}(v)(\boldsymbol{x}) := \sum_{i\in\mathcal{N}} \left(\frac{1}{|K|}(\zeta_i, v)_{L^2(K;\mathbb{R}^q)} \right) \theta_i(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K. \tag{5.13}$$

The domain of $\mathcal{I}_K^{\mathrm{m}}$ can be defined to be $V(K) := L^2(K;\mathbb{R}^q)$, or even $V(K) := L^1(K;\mathbb{R}^q)$ if $P \subset L^\infty(K;\mathbb{R}^q)$. One can verify that $\mathcal{I}_K^{\mathrm{m}}$ is the L^2-orthogonal projection onto P; see Exercise 5.2. Finally, if the basis $\{\zeta_i\}_{i\in\mathcal{N}}$ is L^2-orthogonal, the mass matrix is diagonal, and in that case the shape functions are given by $\theta_i := (|K|/\|\zeta_i\|_{L^2(K;\mathbb{R}^q)}^2)\zeta_i$ for all $i \in \mathcal{N}$.

5.5 The Lebesgue constant

Recall from Definition 5.7 that the interpolation operator $\mathcal{I}_K$ is in $\mathcal{L}(V(K); P)$. Since $P \subset V(K)$, we can equip P with the norm of $V(K)$ and view $\mathcal{I}_K$ a member of $\mathcal{L}(V(K))$. In this section, we study the quantity

$$\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} := \sup_{v\in V(K)} \frac{\|\mathcal{I}_K(v)\|_{V(K)}}{\|v\|_{V(K)}}, \tag{5.14}$$

which is called the *Lebesgue constant* for $\mathcal{I}_K$. We abuse the notation by writing the supremum over $v \in V(K)$ instead of $v \in V(K) \setminus \{0\}$.

Lemma 5.13 (Lower bound). $\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \geq 1$.

Proof. Since P is nontrivial (i.e., $P \neq \{0\}$) and since $\mathcal{I}_K(p) = p$ for all $p \in P$ owing to Proposition 5.9, we infer that

$$\sup_{v\in V(K)} \frac{\|\mathcal{I}_K(v)\|_{V(K)}}{\|v\|_{V(K)}} \geq \sup_{p\in P} \frac{\|\mathcal{I}_K(p)\|_{V(K)}}{\|p\|_{V(K)}} = 1. \qquad \square$$

The Lebesgue constant arises naturally in the estimate of the interpolation error in terms of the *best-approximation error* of a function $v \in V(K)$ by a function in P, that is, $\inf_{p\in P} \|v - p\|_{V(K)}$. In particular, the next result shows that a large value of the Lebesgue constant is associated with poor approximation properties of $\mathcal{I}_K$.

Theorem 5.14 (Interpolation error). *For all $v \in V(K)$, we have*

$$\|v - \mathcal{I}_K(v)\|_{V(K)} \leq (1 + \|\mathcal{I}_K\|_{\mathcal{L}(V(K))}) \inf_{p\in P} \|v - p\|_{V(K)}, \tag{5.15}$$

and $\|v - \mathcal{I}_K(v)\|_{V(K)} \leq \|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \inf_{p\in P} \|v - p\|_{V(K)}$ if $V(K)$ is a Hilbert space.

Proof. Since $\mathcal{I}_K(p) = p$ for all $p \in P$, we infer that $v - \mathcal{I}_K(v) = (I - \mathcal{I}_K)(v) = (I - \mathcal{I}_K)(v - p)$, where I is the identity operator in $V(K)$, so that

$$\|v - \mathcal{I}_K(v)\|_{V(K)} \leq \|(I - \mathcal{I}_K)(v - p)\|_{V(K)} \leq (1 + \|\mathcal{I}_K\|_{\mathcal{L}(V(K))})\|v - p\|_{V(K)},$$

where we used the triangle inequality. We obtain (5.15) by taking the infimum over $p \in P$. Assume now that $V(K)$ is a Hilbert space. We use the fact that

in any Hilbert space H, any operator $T \in \mathcal{L}(H)$ such that $0 \neq T \circ T = T \neq I$ satisfies $\|T\|_{\mathcal{L}(H)} = \|I - T\|_{\mathcal{L}(H)}$; see Kato [124], Xu and Zikatanov [201, Lem. 5], Szyld [188]. We can apply this result with $H := V(K)$ and $T := \mathcal{I}_K$. Indeed, $\mathcal{I}_K \neq 0$ since P is nontrivial, $\mathcal{I}_K \neq I$ since P is a proper subset of $V(K)$, and $\mathcal{I}_K \circ \mathcal{I}_K = \mathcal{I}_K$ owing to Proposition 5.9. We infer that

$$\|v - \mathcal{I}_K(v)\|_{V(K)} \leq \|I - \mathcal{I}_K\|_{\mathcal{L}(V(K))} \|v - p\|_{V(K)} = \|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \|v - p\|_{V(K)},$$

and we conclude by taking the infimum over $p \in P$. $\square$

Example 5.15 (Lagrange elements). The Lebesgue constant for the Lagrange interpolation operator $\mathcal{I}_K^{\mathrm{L}}$ with nodes $\{\boldsymbol{a}_i\}_{i \in \mathcal{N}}$ and space $V(K) := C^0(K)$ is denoted by $\Lambda^{\mathcal{N}} := \|\mathcal{I}_K^{\mathrm{L}}\|_{\mathcal{L}(C^0(K))}$. Owing to Theorem 5.14, we have $\|v - \mathcal{I}_K^{\mathrm{L}}(v)\|_{C^0(K)} \leq (1 + \Lambda^{\mathcal{N}}) \inf_{p \in P} \|v - p\|_{C^0(K)}$. One can verify (see Exercise 5.6) that $\Lambda^{\mathcal{N}} = \|\lambda^{\mathcal{N}}\|_{C^0(K)}$ with the Lebesgue function $\lambda^{\mathcal{N}}(\boldsymbol{x}) := \sum_{i \in \mathcal{N}} |\theta_i(\boldsymbol{x})|$ for all $\boldsymbol{x} \in K$. $\square$

Example 5.16 (Modal elements). Consider a modal finite element with $V(K) := L^2(K; \mathbb{R}^q)$ (see Proposition 5.12). Since $\mathcal{I}_K^{\mathrm{m}}$ is the L^2-orthogonal projection from $L^2(K; \mathbb{R}^q)$ onto P, the Pythagorean identity $\|v\|_{L^2(K;\mathbb{R}^q)}^2 = \|\mathcal{I}_K^{\mathrm{m}}(v)\|_{L^2(K;\mathbb{R}^q)}^2 + \|v - \mathcal{I}_K^{\mathrm{m}}(v)\|_{L^2(K;\mathbb{R}^q)}^2$ implies that $\|\mathcal{I}_K^{\mathrm{m}}\|_{\mathcal{L}(L^2(K;\mathbb{R}^q))} \leq 1$, which in turn gives $\|\mathcal{I}_K^{\mathrm{m}}\|_{\mathcal{L}(L^2(K;\mathbb{R}^q))} = 1$ owing to Lemma 5.13. $\square$

Let assume that $V(K)$ is a Hilbert space with inner product $(\cdot, \cdot)_{V(K)}$. Following ideas developed in Maday et al. [137], we now show that the Lebesgue constant can be related to the stability of an oblique projection. Owing to Theorem A.16 (or Exercise 5.9), we introduce the functions $q_i \in V(K)$ for all $i \in \mathcal{N}$ s.t. $(q_i, v)_{V(K)} = \sigma_i(v)$ for all $v \in V(K)$. Let us set $Q := \mathrm{span}\{q_i\}_{i \in \mathcal{N}}$, and let $Q^\perp$ be the orthogonal to Q in $V(K)$ for the inner product $(\cdot, \cdot)_{V(K)}$.

Lemma 5.17 (Oblique projection). *Let $\mathcal{I}_K$ be defined in* (5.7). *Then $\mathcal{I}_K$ is the oblique projection onto P along $Q^\perp$, and the Lebesgue constant is* $\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} = \alpha_{PQ}^{-1}$ *with* $\alpha_{PQ} := \inf_{p \in P} \sup_{q \in Q} \frac{(p,q)_{V(K)}}{\|p\|_{V(K)} \|q\|_{V(K)}}$.

Proof. (1) Unisolvence implies that $P \cap Q^\perp = \{0\}$. Indeed, if $p \in P \cap Q^\perp$, then $p \in P$ and $\sigma_i(p) = 0$ for all $i \in \mathcal{N}$, so that $p = 0$. Let now $v \in V(K)$. We observe that $\mathcal{I}_K(v) \in P$ and

$$(q_i, \mathcal{I}_K(v) - v)_{V(K)} = \sigma_i(\mathcal{I}_K(v)) - \sigma_i(v) = 0, \qquad \forall i \in \mathcal{N}.$$

Hence, $\mathcal{I}_K(v) - v \in Q^\perp$. From the decomposition $v = \mathcal{I}_K(v) + (v - \mathcal{I}_K(v))$, we infer that $V(K) = P + Q^\perp$. Therefore, the sum is direct, and $\mathcal{I}_K(v)$ is the oblique projection of v onto P along $Q^\perp$.
(2) We have

$$\alpha_{PQ} \|\mathcal{I}_K(v)\|_{V(K)} \leq \sup_{q \in Q} \frac{(\mathcal{I}_K(v), q)_{V(K)}}{\|q\|_{V(K)}} = \sup_{q \in Q} \frac{(v, q)_{V(K)}}{\|q\|_{V(K)}} \leq \|v\|_{V(K)},$$

for all $v \in V(K)$, showing that $\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \leq \alpha_{PQ}^{-1}$. To prove the lower bound, let us first show that $\mathcal{I}_K(\Pi_Q(p)) = p$ for all $p \in P$, where Π_Q is the $V(K)$-orthogonal projection onto Q. We first observe that

$$(\mathcal{I}_K(\Pi_Q(p)), q)_{V(K)} = (\Pi_Q(p), q)_{V(K)} = (p, q)_{V(K)},$$

for all $q \in Q$, where we used the fact that both $\mathcal{I}_K$ and Π_Q are projections along $Q^\perp$. The above identity implies that $\mathcal{I}_K(\Pi_Q(p)) - p \in P \cap Q^\perp = \{0\}$. Hence, $\mathcal{I}_K(\Pi_Q(p)) = p$. Since P is a finite-dimensional space, a compactness argument shows that there is $p^* \in P$ with $\|p^*\|_{V(K)} = 1$ such that $\alpha_{PQ} = \sup_{q \in Q} \frac{(p^*, q)_{V(K)}}{\|q\|_{V(K)}}$. Since $(p^*, q)_{V(K)} = (\Pi_Q(p^*), q)_{V(K)}$, we infer that $\alpha_{PQ} = \sup_{q \in Q} \frac{(\Pi_Q(p^*), q)_{V(K)}}{\|q\|_{V(K)}} = \|\Pi_Q(p^*)\|_{V(K)}$. We conclude that

$$\|\mathcal{I}_K\|_{\mathcal{L}(V(K))} \geq \frac{\|\mathcal{I}_K(\Pi_Q(p^*))\|_{V(K)}}{\|\Pi_Q(p^*)\|_{V(K)}} = \frac{\|p^*\|_{V(K)}}{\|\Pi_Q(p^*)\|_{V(K)}} = \frac{1}{\alpha_{PQ}}. \qquad \square$$

Further results on the Lebesgue constant for one-dimensional Lagrange elements can be found in §6.3.1.

Exercises

Exercise 5.1 (Linear combination). Let $\mathcal{S} \in \mathbb{R}^{n_{\mathrm{sh}} \times n_{\mathrm{sh}}}$ be an invertible matrix. Let (K, P, Σ) be a finite element. Let $\tilde{\Sigma} := \{\tilde{\sigma}_i\}_{i \in \mathcal{N}}$ with dofs $\tilde{\sigma}_i := \sum_{i' \in \mathcal{N}} \mathcal{S}_{ii'} \sigma_{i'}$ for all $i \in \mathcal{N}$. Prove that $(K, P, \tilde{\Sigma})$ is a finite element. Write the shape functions $\{\tilde{\theta}_j\}_{j \in \mathcal{N}}$ and verify that the interpolation operator does not depend on $\mathcal{S}$, i.e., $\tilde{\mathcal{I}}_K(v)(\boldsymbol{x}) = \mathcal{I}_K(v)(\boldsymbol{x})$ for all $v \in V(K)$ and all $\boldsymbol{x} \in K$.

Exercise 5.2 (Modal finite element). (i) Let (K, P, Σ) and $(K, P, \widetilde{\Sigma})$ be two modal finite elements. Let $\{\zeta_i\}_{i \in \mathcal{N}}$, $\{\tilde{\zeta}_i\}_{i \in \mathcal{N}}$, be the two bases of P s.t. the dofs in Σ and $\widetilde{\Sigma}$ are given by $\sigma_i(p) := |K|^{-1}(\zeta_i, p)_{L^2(K;\mathbb{R}^q)}$ and $\tilde{\sigma}_i(p) := |K|^{-1}(\tilde{\zeta}_i, p)_{L^2(K;\mathbb{R}^q)}$ for all $i \in \mathcal{N}$. Prove that the interpolation operators $\mathcal{I}_K^{\mathrm{m}}$ and $\widetilde{\mathcal{I}}_K^{\mathrm{m}}$ are identical. (ii) Prove that $(p, \mathcal{I}_K^{\mathrm{m}}(v) - v)_{L^2(K;\mathbb{R}^q)} = 0$ for all $p \in P$. (iii) Let $\mathcal{M}$ be defined by (5.12), and let $\mathcal{M}_{ij}^\theta := |K|^{-1}(\theta_i, \theta_j)_{L^2(K;\mathbb{R}^q)}$ for all $i, j \in \mathcal{N}$, where $\{\theta_i\}_{i \in \mathcal{N}}$ are the shape functions associated with (K, P, Σ). Prove that $\mathcal{M}^\theta = \mathcal{M}^{-1}$.

Exercise 5.3 (Variation on $\mathbb{P}_2$). Let $K := [0, 1]$, $P := \mathbb{P}_2$ and $\Sigma := \{\sigma_1, \sigma_2, \sigma_3\}$ be the linear forms on P s.t. $\sigma_1(p) := p(0)$, $\sigma_2(p) := 2p(\frac{1}{2}) - p(0) - p(1)$, $\sigma_3(p) := p(1)$ for all $p \in P$. Show that (K, P, Σ) is a finite element, compute the shape functions, and indicate possible choices for $V(K)$.

Exercise 5.4 (Hermite). Let $K := [0, 1]$, $P := \mathbb{P}_3$, and $\Sigma := \{\sigma_1, \sigma_2, \sigma_3, \sigma_4\}$ be the linear forms on P s.t. $\sigma_1(p) := p(0)$, $\sigma_2(p) := p'(0)$, $\sigma_3(p) := p(1)$,

$\sigma_4(p) := p'(1)$ for all $p \in P$. Show that (K, P, Σ) is a finite element, compute the shape functions, and indicate possible choices for $V(K)$.

Exercise 5.5 (Powell–Sabin). Consider $K := [0, 1]$ and let P be composed of the functions that are piecewise quadratic over the intervals $[0, \frac{1}{2}] \cup [\frac{1}{2}, 1]$ and are of class C^1 over K, i.e., functions in P and their first derivatives are continuous. Let $\Sigma := \{\sigma_1, \dots, \sigma_4\}$ be the linear forms on P s.t. $\sigma_1(p) := p(0)$, $\sigma_2(p) := p'(0)$, $\sigma_3(p) := p(1)$, $\sigma_4(p) := p'(1)$. Prove that the triple (K, P, Σ) is a finite element. Verify that the first two shape functions are

$$\theta_1(t) = \begin{cases} 1 - 2t^2 & \text{if } t \in [0, \frac{1}{2}], \\ 2(1-t)^2 & \text{if } t \in [\frac{1}{2}, 1], \end{cases} \qquad \theta_2(t) = \begin{cases} t(1 - \frac{3}{2}t) & \text{if } t \in [0, \frac{1}{2}], \\ \frac{1}{2}(1-t)^2 & \text{if } t \in [\frac{1}{2}, 1], \end{cases}$$

and compute the other two shape functions. *Note*: a two-dimensional version of this finite element on triangles has been developed in [161].

Exercise 5.6 (Lebesgue constant for Lagrange element). Prove that the Lebesgue constant $\Lambda^{\mathcal{N}}$ defined in Example 5.15 is equal to $\|\mathcal{I}_K^{\mathrm{L}}\|_{\mathcal{L}(C^0(K))}$. (*Hint*: to prove $\|\mathcal{I}_K^{\mathrm{L}}\|_{\mathcal{L}(C^0(K))} \geq \Lambda^{\mathcal{N}}$, consider functions $\{\psi_i\}_{i \in \mathcal{N}}$ taking values in $[0, 1]$ s.t. $\sum_{i \in \mathcal{N}} \psi_i = 1$ in K and $\psi_i(\boldsymbol{a}_j) = \delta_{ij}$ for all $i, j \in \mathcal{N}$.)

Exercise 5.7 (Lagrange interpolation). Let $K := [a, b]$ and let $p \in [1, \infty)$. (i) Prove that $\|v\|_{L^\infty(K)} \leq (b-a)^{-\frac{1}{p}} \|v\|_{L^p(K)} + (b-a)^{1-\frac{1}{p}} \|v'\|_{L^p(K)}$ for all $v \in W^{1,p}(K)$ (*Hint*: use $v(x) - v(y) = \int_x^y v'(t) \, dt$ for all $v \in C^1(K)$, where $|v(y)| := \min_{z \in K} |v(z)|$, then use the density of $C^1(K)$ in $W^{1,p}(K)$.) (ii) Prove that $W^{1,p}(K)$ embeds continuously in $C^0(K)$. (iii) Let $\mathcal{I}_K^{\mathrm{L}}$ be the interpolation operator based on the linear Lagrange finite element using the nodes a and b. Determine the two shape functions and prove that $\mathcal{I}_K^{\mathrm{L}}$ can be extended to $W^{1,p}(K)$. (iv) Assuming that $w \in W^{1,p}(K)$ is zero at some point in K, show that $\|w\|_{L^p(K)} \leq (b-a)\|w'\|_{L^p(K)}$. (v) Prove the following estimates: $\|(v - \mathcal{I}_K^{\mathrm{L}}(v))'\|_{L^p(K)} \leq (b-a)\|v''\|_{L^p(K)}$, $\|v - \mathcal{I}_K^{\mathrm{L}}(v)\|_{L^p(K)} \leq (b-a)\|(v - \mathcal{I}_K^{\mathrm{L}}(v))'\|_{L^p(K)}$, $\|(\mathcal{I}_K^{\mathrm{L}}(v))'\|_{L^p(K)} \leq \|v'\|_{L^p(K)}$, for all $p \in (1, \infty]$ and all $v \in W^{2,p}(K)$.

Exercise 5.8 (Cross approximation). Let X, Y be nonempty subsets of $\mathbb{R}$ and $f : X \times Y \to \mathbb{R}$ be a bivariate function. Let $\mathcal{N} := \{1 : n_{\mathrm{sh}}\}$ with $n_{\mathrm{sh}} \geq 1$, and consider n_{sh} points $\{x_i\}_{i \in \mathcal{N}}$ in X and n_{sh} points $\{y_j\}_{j \in \mathcal{N}}$ in Y. Assume that the matrix $\mathcal{F} \in \mathbb{R}^{n_{\mathrm{sh}} \times n_{\mathrm{sh}}}$ with entries $\mathcal{F}_{ij} := f(x_i, y_j)$ is invertible. Let $\mathcal{I}^{\mathrm{CA}}(f) : X \times Y \to \mathbb{R}$ be s.t. $\mathcal{I}^{\mathrm{CA}}(f)(x, y) := \sum_{i,j \in \mathcal{N}} (\mathcal{F}^{-\mathsf{T}})_{ij} f(x, y_j) f(x_i, y)$. Prove that $\mathcal{I}^{\mathrm{CA}}(f)(x, y_k) = f(x, y_k)$ for all $x \in X$ and all $k \in \mathcal{N}$, and that $\mathcal{I}^{\mathrm{CA}}(f)(x_k, y) = f(x_k, y)$ for all $y \in Y$ and all $k \in \mathcal{N}$.

Exercise 5.9 (Riesz–Fréchet in finite dimension). Let V be a finite-dimensional complex Hilbert space. Show that for every antilinear form $A \in V'$, there is a unique $v \in V$ s.t. $(v, w)_V = \langle A, w \rangle_{V', V}$ for all $w \in V$, with $\|v\|_V = \|A\|_{V'}$.

Part II, Chapter 6

One-dimensional finite elements and tensorization

This chapter presents important examples of finite elements, first in dimension one, then in multiple dimensions using tensor-product techniques. Important computational issues related to the manipulation of high-order polynomial bases are addressed. We also show how to approximate integrals over intervals using the roots of the Legendre and Jacobi polynomials.

6.1 Legendre and Jacobi polynomials

Legendre and related polynomials are useful tools to design high-order finite elements. Their roots are also important to construct nodal finite element bases and to devise approximate integration rules called quadratures.

Definition 6.1 (Legendre polynomials). *The* Legendre polynomials *are univariate polynomial functions* $\mathbb{R} \to \mathbb{R}$ *defined for every integer* $m \geq 0$ *by*

$$L_m(t) := \frac{(-1)^m}{2^m m!} \frac{\mathrm{d}^m}{\mathrm{d}t^m} \big((1-t)^m (1+t)^m \big). \tag{6.1}$$

Proposition 6.2 (L^2-orthogonality). *The Legendre polynomials are L^2-orthogonal over the interval $(-1,1)$, and the following holds true:*

$$\int_{-1}^{1} L_m(t) L_n(t) \, \mathrm{d}t = \frac{2}{2m+1} \delta_{mn}, \qquad \forall m, n \geq 0. \tag{6.2}$$

Legendre polynomials satisfy many useful properties. The most important ones are that L_m is a polynomial of degree m, L_m is an even function if m is even and an odd function if m is odd, $L_m(-1) = (-1)^m$ and $L_m(1) = 1$, $L_m'(-1) = \frac{1}{2}(-1)^{m+1} m(m+1)$, and $L_m'(1) = \frac{1}{2} m(m+1)$. We also have for all $m \geq 1$,

$$\frac{1}{m}(t^2 - 1)L'_m(t) = tL_m(t) - L_{m-1}(t), \tag{6.3a}$$

$$L'_m(t) = mL_{m-1}(t) + tL'_{m-1}(t), \tag{6.3b}$$

$$(1 - t^2)L''_m(t) - 2tL'_m(t) + m(m + 1)L_m(t) = 0, \tag{6.3c}$$

and finally $(m+1)L_{m+1}(t) = (2m+1)tL_m(t) - mL_{m-1}(t)$ (Bonnet's recursion formula). The first four Legendre polynomials are the following.

$$\begin{cases} L_0(t) = 1 \\ L_1(t) = t \\ L_2(t) = \frac{1}{2}(3t^2 - 1) \\ L_3(t) = \frac{1}{2}(5t^3 - 3t) \end{cases}$$
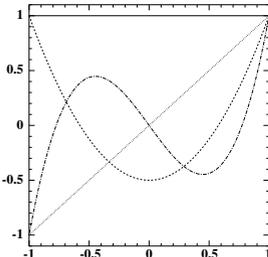
Definition 6.3 (Jacobi polynomials). *Let $\alpha, \beta \in \mathbb{R}$ be such that $\alpha > -1$ and $\beta > -1$. The Jacobi polynomials are univariate polynomial functions $\mathbb{R} \to \mathbb{R}$ defined for every integer $m \geq 0$ by*

$$J_m^{\alpha,\beta}(t) := \frac{(-1)^m}{2^m m!}(1 - t)^{-\alpha}(1 + t)^{-\beta} \frac{\mathrm{d}^m}{\mathrm{d}t^m}\left((1 - t)^{\alpha+m}(1 + t)^{\beta+m}\right). \tag{6.4}$$

The Jacobi polynomials are orthogonal w.r.t. to the L^2-inner product in the interval $(-1, 1)$ weighted by the function $(1 - t)^\alpha(1 + t)^\beta$:

$$\int_{-1}^{1} (1 - t)^\alpha(1 + t)^\beta J_m^{\alpha,\beta}(t)J_n^{\alpha,\beta}(t)\,\mathrm{d}t = c_{m,\alpha,\beta}\delta_{mn}, \qquad \forall m, n \geq 0, \tag{6.5}$$

with $c_{m,\alpha,\beta} := \frac{2^{\alpha+\beta+1}}{2m+\alpha+\beta+1}\frac{\Gamma(m+\alpha+1)\Gamma(m+\beta+1)}{m!\Gamma(m+\alpha+\beta+1)}$, where Γ is the Gamma function (s.t. $\Gamma(n+1) = n!$ for every natural number n). The Jacobi polynomials satisfy the following recursion formula for all $m \geq 1$:

$$2(m + 1)(m + \alpha + \beta + 1)(2m + \alpha + \beta)J_{m+1}^{\alpha,\beta}(t) =$$
$$(2m + \alpha + \beta + 1)((2m + \alpha + \beta + 2)(2m + \alpha + \beta)t + \alpha^2 - \beta^2)J_m^{\alpha,\beta}(t)$$
$$- 2(m + \alpha)(m + \beta)(2m + \alpha + \beta + 2)J_{m-1}^{\alpha,\beta}(t).$$

$J_m^{\alpha,\beta}$ is a polynomial of degree m and $J_m^{\alpha,\beta}(-1) = (-1)^m \binom{m+\beta}{m}$, $J_m^{\alpha,\beta}(1) = \binom{m+\alpha}{m}$. The Legendre polynomials are Jacobi polynomials with parameters $\alpha = \beta = 0$, i.e., $L_m(t) = J_m^{0,0}(t)$ for all $m \geq 0$. The first three Jacobi polynomials corresponding to the parameters $\alpha = \beta = 1$ are

$$J_0^{1,1}(t) = 1, \qquad J_1^{1,1}(t) = 2t, \qquad J_2^{1,1}(t) = \frac{3}{4}(5t^2 - 1).$$

The Jacobi polynomials $J_m^{1,1}$ are related to the integrated Legendre polynomials as follows (see Exercise 6.1):

$$\int_{-1}^{t} L_m(s)\,\mathrm{d}s = -\frac{1}{2m}(1-t^2)J_{m-1}^{1,1}(t), \qquad \forall m \geq 1. \tag{6.6}$$

We refer the reader to Abramowitz and Stegun [1, Chap. 22] for further results on the Legendre and Jacobi polynomials.

6.2 One-dimensional Gauss quadrature

A quadrature formula on, say, the reference interval $K := [-1, 1]$ allows one to approximate the integral of functions ϕ in $C^0(K)$ as follows:

$$\int_{-1}^{1} \phi(t)\,\mathrm{d}t \approx \sum_{l\in\{1:m\}} \omega_l\phi(\xi_l), \tag{6.7}$$

for some integer $m \geq 1$. The points $\{\xi_l\}_{l\in\{1:m\}}$ are called *quadrature nodes*, are all in K, and are all distinct. The real numbers $\{\omega_l\}_{l\in\{1:m\}}$ are called *quadrature weights.* By a change of variables, the quadrature (6.7) can be used on any interval $[a, b]$. Letting $c := \frac{1}{2}(a+b)$ and $\delta := b - a$, (6.7) implies

$$\int_{a}^{b} \phi(t)\,\mathrm{d}t \approx \sum_{l\in\{1:m\}} \tfrac{1}{2}\delta\omega_l\phi(c + \tfrac{1}{2}\delta\xi_l). \tag{6.8}$$

The largest integer $k_{\mathcal{Q}}$ such that equality holds true in (6.7) for every polynomial in $\mathbb{P}_{k_{\mathcal{Q}}}$ is called *quadrature order*, that is, we have

$$\int_{-1}^{1} p(t)\,\mathrm{d}t = \sum_{l\in\{1:m\}} \omega_l p(\xi_l), \qquad \forall p \in \mathbb{P}_{k_{\mathcal{Q}}}, \tag{6.9}$$

and there is $q \in \mathbb{P}_{k_{\mathcal{Q}}+1}$ s.t. $\int_{-1}^{1} q(t)\,\mathrm{d}t \neq \sum_{l\in\{1:m\}} \omega_l q(\xi_l)$. At this stage, it suffices to know that the higher the quadrature order, the more accurate the quadrature (6.7). We refer the reader to Chapter 30 for estimates on the quadrature error and for quadratures in multiple dimensions.

Lemma 6.4 (Quadrature order). *Let $\{\xi_l\}_{l\in\{1:m\}}$ be m distinct points in K. Let $\{\mathcal{L}_l\}_{l\in\{1:m\}}$ be the associated Lagrange interpolation polynomials, i.e., $\mathcal{L}_l(\xi_j) = \delta_{lj}$ for all $l, j \in \{1:m\}$. Set $\omega_l := \int_{-1}^{1} \mathcal{L}_l(t)\,\mathrm{d}t$ for all $l \in \{1:m\}$. Then the quadrature (6.7) is at least of order $(m-1)$ and at most of order $(2m-1)$, i.e., $m - 1 \leq k_{\mathcal{Q}} \leq 2m - 1$.*

Proof. Let $p \in \mathbb{P}_{m-1}$. Since the m quadrature nodes are all distinct, the Lagrange interpolation polynomials $\{\mathcal{L}_l\}_{l\in\{1:m\}}$ form a basis of $\mathbb{P}_{m-1}$. Thus,

we can write $p(t) = \sum_{l \in \{1:m\}} p(\xi_l) \mathcal{L}_l(t)$, whence we infer that

$$\int_{-1}^{1} p(t) \, \mathrm{d}t = \sum_{l \in \{1:m\}} p(\xi_l) \int_{-1}^{1} \mathcal{L}_l(t) \, \mathrm{d}t = \sum_{l \in \{1:m\}} \omega_l p(\xi_l),$$

owing to the linearity of the integral and the definition of the weights. Hence, $k_{\mathcal{Q}} \geq m - 1$. Moreover, the polynomial $q(t) := \prod_{l \in \{1:m\}} (t - \xi_l)^2$ is of degree $2m$ and is not integrated exactly by the quadrature (which approximates its integral by zero). Hence, $k_{\mathcal{Q}} \leq 2m - 1$. $\square$

For all $m \geq 1$, one can show that the m roots of the Legendre polynomial L_m are distinct and are all in the open interval $(-1, 1)$. The most important example of quadrature is the one based on these roots, which we henceforth call *Gauss–Legendre nodes*.

Proposition 6.5 (Gauss–Legendre). *Let $m \geq 1$. Let $\{\xi_l\}_{l \in \{1:m\}}$ be the m roots of the Legendre polynomial $L_m(t)$ (all distinct and in $(-1, 1)$). Let the weights be defined as in Lemma* 6.4. *Then the quadrature* (6.7) *is of order $k_{\mathcal{Q}} = 2m - 1$. Moreover, all the weights are positive and are given by*

$$\omega_l = \frac{2}{(1 - \xi_l^2) L_m'(\xi_l)^2}, \qquad \forall l \in \{1:m\}. \tag{6.10}$$

Proof. (i) Order of the quadrature. We already know from Lemma 6.4 that $m - 1 \leq k_{\mathcal{Q}} \leq 2m - 1$. Consider a polynomial $p \in \mathbb{P}_{2m-1}$. The Euclidean division of polynomials shows that there are $p_1, p_2 \in \mathbb{P}_{m-1}$ such that $p = p_1 L_m + p_2$. Using that $k_{\mathcal{Q}} \geq m - 1$, the L^2-orthogonality of Legendre polynomials, and the identity $p(\xi_l) = p_2(\xi_l)$ (since $L_m(\xi_l) = 0$ for all $l \in \{1:m\}$), we infer that

$$\int_{-1}^{1} p(t) \, \mathrm{d}t = \int_{-1}^{1} p_2(t) \, \mathrm{d}t = \sum_{l \in \{1:m\}} \omega_l p_2(\xi_l) = \sum_{l \in \{1:m\}} \omega_l p(\xi_l).$$

This shows that $k_{\mathcal{Q}} \geq 2m - 1$. Hence, $k_{\mathcal{Q}} = 2m - 1$.
(ii) Let us prove (6.10) for all $l \in \{1:m\}$. Let $\mathcal{L}_l \in \mathbb{P}_{m-1}$ be the Lagrange interpolation polynomial associated with the node ξ_l, i.e., $\mathcal{L}_l(\xi_j) = \delta_{lj}$ for all $l, j \in \{1:m\}$. Since the polynomial $\mathcal{L}_l(t)(1-t)L_m'(t)$ is of degree $(2m-1)$, it is integrated exactly by the quadrature. Hence, we have

$$\int_{-1}^{1} \mathcal{L}_l(t)(1-t)L_m'(t) \, \mathrm{d}t = \omega_l(1 - \xi_l)L_m'(\xi_l).$$

Moreover, integrating by parts and since $\int_{-1}^{1} \big(\mathcal{L}_l(t)(1-t) \big)' L_m(t) \, \mathrm{d}t = 0$ owing to the L^2-orthogonality of the Legendre polynomials, we obtain

$$\int_{-1}^{1} \mathcal{L}_l(t)(1-t)L'_m(t)\,\mathrm{d}t = -2\mathcal{L}_l(-1)L_m(-1).$$

Next, we observe that $\mathcal{L}_l(t) = \frac{L_m(t)}{t-\xi_l}\frac{1}{L'_m(\xi_l)}$ since both functions are polynomials in $\mathbb{P}_{m-1}$ having the same roots and taking the same value 1 at ξ_l owing to l'Hôpital's rule. Thus, $\mathcal{L}_l(-1) = -\frac{L_m(-1)}{1+\xi_l}\frac{1}{L'_m(\xi_l)}$. Combining the above identities leads to

$$\omega_l(1-\xi_l)L'_m(\xi_l) = 2\frac{L_m(-1)^2}{1+\xi_l}\frac{1}{L'_m(\xi_l)},$$

which proves the claim since $L_m(-1)^2 = 1$. $\qquad\square$

In some situations, it is interesting to use quadratures with nodes including one or the two endpoints of the interval $[-1,1]$. The Gauss–Lobatto quadrature corresponds to the case where both endpoints are included. The nodes of this quadrature for $m \geq 3$ are $\{-1,1\}$ plus the $(m-2)$ roots of the polynomial $L'_{m-1}(t)$, which can be shown to be all distinct and contained in the open interval $(-1,1)$.

Proposition 6.6 (Gauss–Lobatto). *Let $m \geq 2$. Let $\{\xi_l\}_{l\in\{1:m\}}$ be the Gauss–Lobatto nodes, i.e., the m roots of the polynomial $(1-t^2)L'_{m-1}(t)$ (they are all distinct and in $[-1,1]$). Let the weights be defined as in Lemma 6.4. Then the quadrature (6.7) is of order $k_{\mathcal{Q}} = 2m-3$. Moreover, all the weights are positive and are given by*

$$\omega_l = \frac{2}{m(m-1)}\frac{1}{L_{m-1}(\xi_l)^2}, \qquad \forall l \in \{1:m\}. \tag{6.11}$$

In particular, we have $\omega_1 = \omega_m = \frac{2}{m(m-1)}$.

Proof. See Exercise 6.2. $\qquad\square$

The case where one keeps only one of the two endpoints leads to the Gauss–Radau quadrature. For brevity, we focus on the right-sided version which keeps the right endpoint $\xi_m = 1$. The left-sided version keeping the left endpoint $\xi_1 = -1$ can be derived from symmetry arguments. The nodes of the right-sided quadrature are the m roots of the polynomial $L_m(t)-L_{m-1}(t)$, which can be shown to be all distinct and contained in $(-1,1]$ (notice that 1 is a root of this polynomial).

Proposition 6.7 (Gauss–Radau, right-sided). *Let $m\geq1$. Let $\{\xi_l\}_{l\in\{1:m\}}$ be the Gauss–Radau nodes, i.e., the m roots of the polynomial $L_m(t) - L_{m-1}(t)$ (they are all distinct and in $(-1,1]$). Let the weights be defined as in Lemma 6.4. Then the quadrature (6.7) is of order $k_{\mathcal{Q}} = 2m-2$. Moreover, all the weights are positive and are given by*

$$\omega_l = \frac{1}{(1+\xi_l)L'_{m-1}(\xi_l)^2}, \quad \forall l \in \{1:m-1\}, \qquad \omega_m = \frac{2}{m^2}. \tag{6.12}$$

Proof. See Exercise 6.3. □

Examples of quadratures on the reference interval $[-1,1]$ are presented in Table 6.1. The Gauss–Legendre quadrature of order 1 is called *midpoint rule*, the Gauss–Lobatto quadrature of order 1 is called *trapezoidal rule* and that of order 3 *Simpson's rule*. For quadratures of higher order, we refer the reader, e.g., to Karniadakis and Sherwin [123, §B.2].

type	order	nodes	weights
G-Rad	0	1	2
G-Leg	1	0	2
G-Lob	1	$-1, 1$	$1, 1$
G-Rad	2	$-\frac{1}{3}, 1$	$\frac{3}{2}, \frac{1}{2}$
G-Leg	3	$-\frac{\sqrt{3}}{3}, \frac{\sqrt{3}}{3}$	$1, 1$
G-Lob	3	$-1, 0, 1$	$\frac{1}{3}, \frac{4}{3}, \frac{1}{3}$
G-Rad	4	$\frac{-1-\sqrt{6}}{5}, \frac{-1+\sqrt{6}}{5}, 1$	$\frac{16+\sqrt{6}}{18}, \frac{16-\sqrt{6}}{18}, \frac{2}{9}$
G-Leg	5	$-\frac{\sqrt{15}}{5}, 0, \frac{\sqrt{15}}{5}$	$\frac{5}{9}, \frac{8}{9}, \frac{5}{9}$
G-Lob	5	$-1, -\frac{\sqrt{5}}{5}, \frac{\sqrt{5}}{5}, 1$	$\frac{1}{6}, \frac{5}{6}, \frac{5}{6}, \frac{1}{6}$

Table 6.1 One-dimensional quadratures on the reference interval $[-1,1]$. G-Leg: Gauss–Legendre, G-Rad: Gauss–Radau, G-Lob: Gauss–Lobatto.

6.3 One-dimensional finite elements

In this section, we present important examples of one-dimensional finite elements. Recall that $\mathbb{P}_k$, $k \geq 0$, is the real vector space composed of univariate polynomial functions of degree at most k. For convenience, degrees of freedom (dofs) and shape functions of one-dimensional finite elements using the polynomial space $\mathbb{P}_k$ are numbered from 0 to k.

6.3.1 Lagrange (nodal) finite elements

Following Definition 5.11, the dofs for Lagrange finite elements are chosen as the values at some set of nodes.

Proposition 6.8 (Lagrange finite element). *Let $k \geq 0$. Let K be a compact interval with nonempty interior, and let $P := \mathbb{P}_k$. Consider a set of $n_{\mathrm{sh}} := k + 1$ distinct nodes $\{a_l\}_{l \in \{0:k\}}$ in K. Let $\Sigma := \{\sigma_l\}_{l \in \{0:k\}}$ be the linear forms on P such that $\sigma_l(p) := p(a_l)$ for all $l \in \{0:k\}$. Then (K, P, Σ) is a Lagrange finite element.*

Proof. We use Remark 5.3. We observe that $\dim P = k + 1 = n_{\text{sh}} = \operatorname{card} \Sigma$. Moreover, let $p \in P$ be such that $\sigma_l(p) = p(a_l) = 0$ for all $l \in \{0{:}k\}$. Then $p = 0$ since p is of degree at most k and has $(k + 1)$ distinct roots. $\qquad\square$

The shape functions of a one-dimensional Lagrange finite element are the Lagrange interpolation polynomials $\{\mathcal{L}_l^{[a]}\}_{i \in \{0{:}k\}}$ defined as in (5.2). Following (5.10), the Lagrange interpolation operator acts as follows:

$$\mathcal{I}_K^{\mathrm{L}}(v)(t) := \sum_{l \in \{0{:}k\}} v(a_l) \mathcal{L}_l^{[a]}(t), \qquad \forall t \in K, \qquad (6.13)$$

and possible choices for the domain of $\mathcal{I}_K^{\mathrm{L}}$ are $V(K) := C^0(K)$ or $V(K) := W^{1,1}(K)$; see Exercise 5.7.

The Lagrange interpolation polynomials based on $(k+1)$ equidistant nodes (including both endpoints) in the interval $K := [-1, 1]$ are henceforth denoted by $\{\mathcal{L}_l^k\}_{l \in \{0{:}k\}}$. The graphs of these polynomials are illustrated in Figure 5.1 for $k \in \{1, 2, 3\}$. Explicit expressions are as follows:

$$\mathcal{L}_0^1(t) := \tfrac{1}{2}(1 - t), \quad \mathcal{L}_0^2(t) := \tfrac{1}{2}t(t - 1), \qquad \mathcal{L}_0^3(t) := \tfrac{9}{16}(t + \tfrac{1}{3})(t - \tfrac{1}{3})(1 - t),$$
$$\mathcal{L}_1^1(t) := \tfrac{1}{2}(1 + t), \quad \mathcal{L}_1^2(t) := (t + 1)(1 - t), \quad \mathcal{L}_1^3(t) := \tfrac{27}{16}(t + 1)(t - \tfrac{1}{3})(t - 1),$$
$$\mathcal{L}_2^2(t) := \tfrac{1}{2}(t + 1)t, \qquad \mathcal{L}_2^3(t) := \tfrac{27}{16}(t + 1)(t + \tfrac{1}{3})(1 - t),$$
$$\mathcal{L}_3^3(t) := \tfrac{9}{16}(t + 1)(t + \tfrac{1}{3})(t - \tfrac{1}{3}).$$

Although the choice of equidistant nodes appears somewhat natural, it is appropriate only when working with low-degree polynomials; see §6.3.5. An alternative choice is to consider the Gauss–Lobatto nodes. The corresponding Lagrange interpolation polynomials for $k = 3$ (four nodes) are illustrated in the left panel of Figure 6.1.

6.3.2 Modal finite elements

Let us illustrate the construction of §5.4.2 in the one-dimensional setting.

Proposition 6.9 (Legendre finite element). *Let $k \geq 0$. Let $K := [-1, 1]$, $P := \mathbb{P}_k$, and $\Sigma := \{\sigma_l\}_{l \in \{0{:}k\}}$ be the $n_{\text{sh}} := k + 1$ linear forms on P s.t.*

$$\sigma_l(p) := \frac{2l + 1}{2} \int_{-1}^{1} L_l(t)p(t)\,\mathrm{d}t, \quad \forall l \in \{0{:}k\}, \qquad (6.14)$$

where L_l is the Legendre polynomial of order l. Then (K, P, Σ) is a finite element, and the shape functions are $\theta_l := L_l$ for all $l \in \{0{:}k\}$.

Proof. Use (6.2) and Proposition 5.12. $\qquad\square$

Following (5.13), the Legendre interpolation operator acts as follows: $\mathcal{I}_K^{\mathrm{m}}(v)(t) := \sum_{l \in \{0{:}k\}} \left(\frac{2l+1}{2} \int_{-1}^{1} L_l(s)v(s)\,\mathrm{d}s \right) L_l(t)$ for all $t \in K$ and all $v \in V(K) := L^1(K)$.
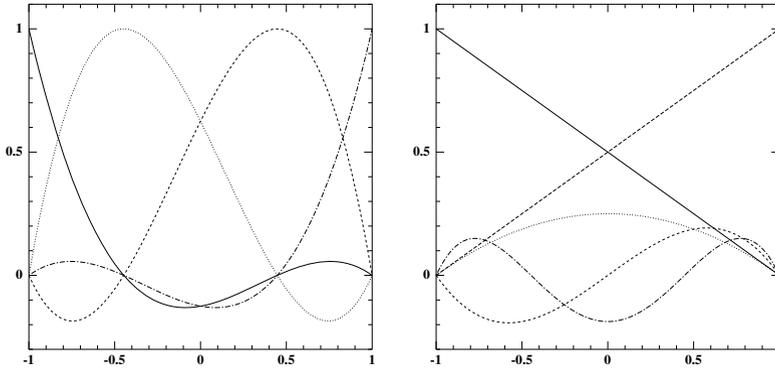
Fig. 6.1 Left: Lagrange interpolation polynomials for Gauss–Lobatto nodes and $k = 3$. Right: Hybrid nodal/modal shape functions for $k = 4$ (see §6.3.3).

6.3.3 Canonical hybrid finite element

Hybrid finite elements mix nodal and modal dofs. When constructing H^1-conforming approximation spaces (see Chapter 19), it is convenient that all the basis functions but one vanish at each endpoint of the interval, say $K := [-1, 1]$. This calls for using the values at ± 1 as nodal dofs on $\mathbb{P}_k$. For $k \geq 2$, some or all of the remaining dofs can be taken to be of modal type. Taking all of them to be moments against polynomials in $\mathbb{P}_{k-2}$ gives a finite element called *canonical hybrid finite element*. A multidimensional extension is presented in §7.6.

Proposition 6.10 (Canonical hybrid finite element). *Let $k \geq 1$. Set $K := [-1, 1]$ and $P := \mathbb{P}_k$. Define $\sigma_0(p) := p(-1)$, $\sigma_k(p) := p(1)$, and, if $k \geq 2$, let $\{\mu_l\}_{l \in \{1:k-1\}}$ be a basis of $\mathbb{P}_{k-2}$ and define $\sigma_l(p) := \int_K p\mu_l \, ds$ for all $l \in \{1:k-1\}$. Set $\Sigma := \{\sigma_l\}_{l \in \{0:k\}}$. Then (K, P, Σ) is a finite element.*

Proof. See Exercise 6.6. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The corresponding interpolation operator is denoted by $\mathcal{I}_K^{\mathrm{g}}$ (the superscript is consistent with the notation introduced in §16.2 where the letter "g" refers to the gradient operator). Its action on functions $v \in V(K) := W^{1,1}(K)$ is s.t. $\mathcal{I}_K^{\mathrm{g}}(v)(\pm 1) = v(\pm 1)$ and $\int_{-1}^1 (\mathcal{I}_K^{\mathrm{g}}(v) - v)q \, ds = 0$ for all $q \in \mathbb{P}_{k-2}$.

Proposition 6.11 (Commuting with derivative). *Let $k \geq 0$. Let $\mathcal{I}_K^{\mathrm{g}}$ be the interpolation operator built from the canonical hybrid finite element of order $(k + 1)$. Let $\mathcal{I}_K^{\mathrm{m}}$ be the interpolation operator built from any modal finite element of order k. Then $\mathcal{I}_K^{\mathrm{g}}(v)' = \mathcal{I}_K^{\mathrm{m}}(v')$ for all $v \in W^{1,1}(K)$.*

Proof. Integrating by parts, using the properties of $\mathcal{I}_K^{\mathrm{g}}$ (i.e., $\int_{-1}^1 (\mathcal{I}_K^{\mathrm{g}}(v) - v)r \, ds = 0$ for all $r \in \mathbb{P}_{k-1}$ and $\mathcal{I}_K^{\mathrm{g}}(v)(\pm 1) = v(\pm 1)$) and recalling that $\mathcal{I}_K^{\mathrm{m}}$ is

the L^2-orthogonal projection onto $\mathbb{P}_k$ (see §5.4.2 and Exercise 5.2), we infer that for all $q \in \mathbb{P}_k$,

$$\int_{-1}^{1} \mathcal{I}_K^{\mathrm{g}}(v)'q \, \mathrm{d}t = -\int_{-1}^{1} \mathcal{I}_K^{\mathrm{g}}(v)q' \, \mathrm{d}t + \left[\mathcal{I}_K^{\mathrm{g}}(v)q\right]_{-1}^{1}$$

$$= -\int_{-1}^{1} vq' \, \mathrm{d}t + \left[vq\right]_{-1}^{1} = \int_{-1}^{1} v'q \, \mathrm{d}t = \int_{-1}^{1} \mathcal{I}_K^{\mathrm{m}}(v')q \, \mathrm{d}t.$$

This proves that $\mathcal{I}_K^{\mathrm{g}}(v)' = \mathcal{I}_K^{\mathrm{m}}(v')$ since both functions are in $\mathbb{P}_k$. $\qquad\square$

The shape functions of the canonical hybrid finite element can be computed explicitly once a choice for the basis functions $\{\mu_l\}_{l \in \{1:k-1\}}$ of $\mathbb{P}_{k-2}$ is made. An example is proposed in Exercise 6.6 using Jacobi polynomials (see the right panel of Figure 6.1 for an illustration).

6.3.4 Hierarchical bases

The notion of hierarchical polynomial bases is important when working with high-order polynomials. It is particularly convenient in simulations where the degree k varies from one element to the other.

Definition 6.12 (Hierarchical basis). *A sequence of polynomials* $(P_k)_{k\in\mathbb{N}}$ *is said to be a* hierarchical polynomial basis *if the set* $\{P_l\}_{l\in\{0:k\}}$ *is a basis of* $\mathbb{P}_k$ *for all* $k \in \mathbb{N}$.

The monomial basis (i.e., $P_k(t) = t^k$) is the simplest example of hierarchical polynomial basis. Another example is the Jacobi polynomials introduced in §6.1. They form a hierarchical basis with the additional property to be L^2-orthogonal with respect to the weight $(1-t)^\alpha(1+t)^\beta$.

The Lagrange shape functions do not form a hierarchical basis, i.e., increasing k to $(k+1)$ requires to recompute the whole basis of shape functions. The shape functions of modal elements form, by construction, a hierarchical basis. Finally, the shape functions of the canonical hybrid finite element do not form in general a hierarchical basis. One can obtain a hierarchical basis though by slightly modifying the dofs. For instance, the following shape functions form a hierarchical basis:

$$\theta_0(t) := \tfrac{1}{2}(1-t), \tag{6.15a}$$

$$\theta_l(t) := \tfrac{1}{4}(1-t)(1+t)J_{l-1}^{1,1}(t), \quad \forall l \in \{1:k-1\}, \quad k \geq 2, \tag{6.15b}$$

$$\theta_k(t) := \tfrac{1}{2}(1+t). \tag{6.15c}$$

The corresponding dofs are $\sigma_0(p) := p(-1)$, $\sigma_k(p) := p(1)$, and $\sigma_l(p) := \alpha_l \int_{-1}^{1} pJ_{l-1}^{1,1} \, \mathrm{d}t + \beta_l^- p(-1) + \beta_l^+ p(1)$ for all $l \in \{1:k-1\}$, where $\alpha_l := 4c_{l-1,1,1}^{-1}$, $\beta_l^\pm := -2c_{l-1,1,1}^{-1} \int_{-1}^{1} (1 \pm t)J_{l-1}^{1,1} \, \mathrm{d}t$, and $c_{l-1,1,1}$ defined in (6.5).

6.3.5 High-order Lagrange elements

The Lagrange polynomials oscillate more and more as the number of interpolation nodes grows. This phenomenon is often referred to as the Runge phenomenon [171] (see also Meray [142]). A classic example illustrating this phenomenon consists of considering the function $f(x) := (1+x^2)^{-1}$, $x \in [-5,5]$. The Lagrange interpolation polynomial of f using n equidistant points over $[-5,5]$ converges uniformly to f in the interval $(-x_c, x_c)$ with $x_c \sim 3.63$ and diverges outside this interval.

The approximation quality in the maximum norm of the one-dimensional Lagrange interpolation operator using $n_{\mathrm{sh}} := k+1$ distinct nodes is quantified by the Lebesgue constant; see Theorem 5.14. Since the Lebesgue constant is invariant under any linear transformation of the interval K, we henceforth restrict the discussion to $K := [-1,1]$. It can be shown that the Lebesgue constant for equidistant nodes grows exponentially with k. More precisely, $\Lambda_k \sim \frac{2^k}{ek(\ln(k)+\gamma)}$ as $k \to \infty$ where $\gamma := 0.5772\cdots$ is the Euler constant; see Trefethen and Weideman [191]. A lower bound for the Lebesgue constant for every set of points is $\frac{2}{\pi} \ln(k) - C$ for some positive C; see Erdös [94]. If the Chebyshev nodes $\left\{ a_l := \cos\left(\frac{(2l-1)\pi}{2n_{\mathrm{sh}}}\right) \right\}_{l \in \{1:n_{\mathrm{sh}}\}}$ are used instead of the equidistant nodes, the Lebesgue constant behaves as $\frac{2}{\pi} \ln(k) + C + \alpha_k$, with $C := \frac{2}{\pi}(\gamma + \ln(\frac{8}{\pi})) = 0.9625\cdots$ and $\alpha_k \to 0$ as $k \to \infty$, showing that this choice is asymptotically optimal; see Luttmann and Rivlin [136] and Rivlin [167, Chap. 4]. The Gauss–Lobatto nodes, which include the two endpoints, lead to an asymptotically optimal Lebesgue constant with upper bound $\frac{2}{\pi} \ln(k) + C$, with $C \sim 0.685$; see Hesthaven [114, Conj. 3.2] and Hesthaven et al. [115, p. 106]. Note that the Lagrange polynomial bases using the Gauss–Lobatto nodes are not hierarchical since the set of n_{sh} Gauss–Lobatto nodes is not included in the set of $(n_{\mathrm{sh}} + 1)$ nodes.

Another important class of sets of nodes is that consisting of the *Fekete points*. These points are defined from a maximization problem. Let $\{a_i\}_{i \in \mathcal{N}}$ be a set of nodes in $K := [-1,1]$ and let $\{\phi_i\}_{i \in \mathcal{N}}$ be a basis of $\mathbb{P}_k$ (recall that $\mathcal{N} := \{1:n_{\mathrm{sh}}\}$). Recall the (generalized) $n_{\mathrm{sh}} \times n_{\mathrm{sh}}$ Vandermonde matrix $\mathcal{V}$ with entries $\mathcal{V}_{ij} := \phi_i(a_j)$ for all $i,j \in \mathcal{N}$. Since the Lagrange polynomials can be expressed as $\mathcal{L}_i^{[a]}(t) = \sum_{j \in \mathcal{N}} (\mathcal{V}^{-1})_{ij} \phi_j(t)$ (see Proposition 5.5), a reasonable criterion for selecting the interpolation nodes is to maximize the determinant of $\mathcal{V}$ with respect to $\{a_i\}_{i \in \mathcal{N}}$ (observe that the solution to this problem does not depend on the chosen basis, since a change of basis only multiplies the determinant by a factor equal to the determinant of the change of basis matrix). It is shown in Fejér [101] that the Fekete points and the Gauss–Lobatto nodes coincide on any interval. The notion of Fekete points extends naturally to any dimension, but the concept of Gauss–Lobatto nodes can be extended to higher dimension only by invoking tensor-product techniques as we show in the next section.

6.4 Multidimensional tensor-product elements

We show in this section that the one-dimensional finite elements presented in §6.3 can be extended to higher dimension by using tensor-product techniques when $K \subset \mathbb{R}^d$ is a *cuboid*, $d \geq 2$, i.e., when K has the Cartesian product structure $K := \prod_{i=1}^d [z_i^-, z_i^+]$ where $z_i^\pm \in \mathbb{R}$ are such that $z_i^- < z_i^+$ for all $i \in \{1{:}d\}$.

6.4.1 The polynomial space $\mathbb{Q}_{k,d}$

Tensor-product finite elements in $\mathbb{R}^d$ make use of the polynomial space

$$\mathbb{Q}_{k,d} := \underbrace{\mathbb{P}_k \otimes \ldots \otimes \mathbb{P}_k}_{d \text{ times}}. \tag{6.16}$$

This space is composed of d-variate polynomial functions $q : \mathbb{R}^d \to \mathbb{R}$ of partial degree at most k with respect to each variable. Thus, we have

$$\mathbb{Q}_{k,d} = \text{span} \left\{ x_1^{\beta_1} \ldots x_d^{\beta_d}, \ 0 \leq \beta_1, \ldots, \beta_d \leq k \right\}. \tag{6.17}$$

We omit the subscript d and simply write $\mathbb{Q}_k$ when the context is unambiguous. Let $\beta := (\beta_1, \ldots, \beta_d) \in \mathbb{N}^d$ be a multi-index, define $\|\beta\|_{\ell^\infty} := \max_{i \in \{1:d\}} \beta_i$, and consider the multi-index set $\mathcal{B}_{k,d} := \{\beta \in \mathbb{N}^d \mid \|\beta\|_{\ell^\infty} \leq k\}$. Polynomial functions $q \in \mathbb{Q}_{k,d}$ can be written in the generic form

$$q(\boldsymbol{x}) = \sum_{\beta \in \mathcal{B}_{k,d}} a_\beta \boldsymbol{x}^\beta, \qquad \boldsymbol{x}^\beta := x_1^{\beta_1} \ldots x_d^{\beta_d}, \tag{6.18}$$

with real numbers a_β. Note that $\text{card}(\mathcal{B}_{k,d}) = \dim(\mathbb{Q}_{k,d}) = (k+1)^d$.

A direct verification leads to the following useful characterization of the trace of polynomials in $\mathbb{Q}_{k,d}$.

Lemma 6.13 (Trace space). *Let H be an affine subspace of $\mathbb{R}^d$ of co-dimension $l \in \{1{:}d{-}1\}$. Let $\boldsymbol{T}_H : \mathbb{R}^{d-l} \to H$ be an affine bijective mapping. Then $q_{|H} \circ \boldsymbol{T}_H \in \mathbb{Q}_{k,d-l}$ for all $q \in \mathbb{Q}_{k,d}$.*

6.4.2 Tensor-product construction of finite elements

We begin with tensor-product Lagrange finite elements with nodes obtained by invoking the tensor product of nodes along each Cartesian direction. This leads to the following construction.

Proposition 6.14 (Tensor-product Lagrange). *Let $K := \prod_{i=1}^d [z_i^-, z_i^+]$ be a cuboid in $\mathbb{R}^d$. Let $P := \mathbb{Q}_{k,d}$ for some integer $k \geq 1$. For all $i \in \{1{:}d\}$, consider $(k+1)$ distinct nodes $\{a_{i,l}\}_{l \in \{0:k\}}$ in $[z_i^-, z_i^+]$. For every multi-index $\beta := (\beta_1, \ldots, \beta_d) \in \mathcal{B}_{k,d}$, let $\boldsymbol{a}_\beta$ be the node in K with Cartesian components*

$(a_{i,\beta_i})_{i\in\{1:d\}}$. Let $\Sigma := \{\sigma_\beta\}_{\beta\in\mathcal{B}_{k,d}}$ be the degrees of freedom (dofs) on P s.t. $\sigma_\beta(p) := p(\boldsymbol{a}_\beta)$ for all $\beta \in \mathcal{B}_{k,d}$. Then (K, P, Σ) is a Lagrange finite element.

Proof. See Exercise 6.7. $\square$

The following property (see Exercise 6.9) is important for the construction of H^1-conforming finite elements spaces using tensor-product finite elements.

Lemma 6.15 (Face unisolvence). *Consider the cuboid $K := \prod_{i=1}^d [z_i^-, z_i^+]$. Assume that $a_{i,0} = z_i^-$ and $a_{i,k} = z_i^+$ for all $i \in \{1:d\}$. Let F be one of the faces of K. Let $\mathcal{N}_F$ be the collection of the indices of the Lagrange nodes on F. The following holds true for all $p \in \mathbb{Q}_{k,d}$:*

$$[\,\sigma_j(p) = 0, \ \forall j \in \mathcal{N}_F\,] \iff [\,p_{|F} = 0\,]. \tag{6.19}$$
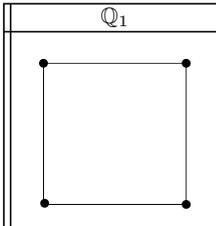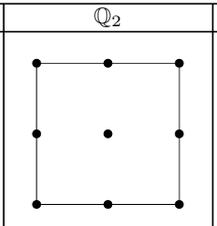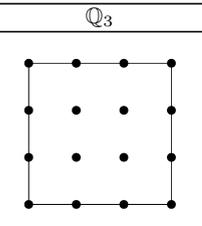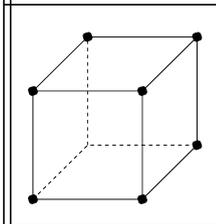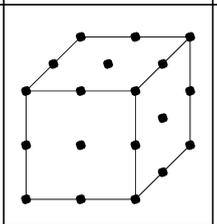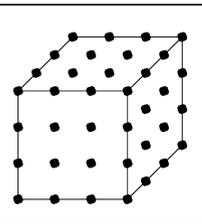
Table 6.2 Two- and three-dimensional Lagrange finite elements $\mathbb{Q}_1$, $\mathbb{Q}_2$, and $\mathbb{Q}_3$. Only visible degrees of freedom are shown in dimension three.

Table 6.2 presents examples for $k \in \{1, 2, 3\}$ in dimensions $d \in \{2, 3\}$ with equidistant nodes in each Cartesian direction. The bullets conventionally indicate the location of the nodes. It is useful to use the tensor product of Gauss–Lobatto nodes when k is large, since it can be shown that the Fekete points in cuboids are the tensor products of the one-dimensional Gauss–Lobatto nodes; see Bos et al. [35].

The shape functions of a tensor-product Lagrange finite element are products of the one-dimensional Lagrange polynomials defined in (5.2):

$$\theta_\beta(\boldsymbol{x}) := \prod_{i\in\{1:d\}} \mathcal{L}_{\beta_i}^{[a_i]}(x_i), \qquad \forall \boldsymbol{x} \in K, \tag{6.20}$$

for all $\beta \in \mathcal{B}_{k,d}$. The Lagrange interpolation operator acts as follows:

$$\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{x}) := \sum_{\beta \in \mathcal{B}_{k,d}} v(\boldsymbol{a}_\beta)\theta_\beta(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K, \tag{6.21}$$

and possible choices for its domain are $V(K) := C^0(K)$ or $V(K) := W^{s,p}(K)$ with real numbers $p \in [1, \infty]$ and $sp > d$ (or $s \geq d$ if $p = 1$). Note that in general $\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{x})$ cannot be factored as a product of univariate functions, except when the function v has the separated form $v(\boldsymbol{x}) = \prod_{i \in \{1:d\}} v_i(x_i)$ with $v_i \in C^0([x_i^-, x_i^+])$, in which case $\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{x}) = \prod_{i \in \{1:d\}} \mathcal{I}_{[x_i^-, x_i^+]}^{\mathrm{L}}(v_i)(x_i)$.

The tensor-product technique can also be used to build modal and hybrid nodal/modal finite elements in cuboids; see [35]. Finite element methods based on nodal bases using tensor products are often referred to as *spectral element methods*; see Patera [156].

6.4.3 Serendipity finite elements

It is possible to remove some nodes inside the cuboid while maintaining the approximation properties of the full tensor product. This is the idea of the serendipity finite elements. The corresponding polynomial space, $\mathbb{S}_k$, is then a proper subspace of $\mathbb{Q}_k$. The main motivation is to reduce computational costs without sacrificing the possibility to build H^1-conforming finite element spaces and without sacrificing the accuracy of the interpolation operator. Classical two-dimensional examples consist of using the 8 boundary nodes if $k = 2$ and the 12 boundary nodes if $k = 3$ (see Table 6.3). If $k = 4$ one uses the 16 boundary nodes plus the barycenter of K. A systematic construction of the serendipity finite elements for all dimensions and all polynomial degrees is devised in Arnold and Awanou [10].
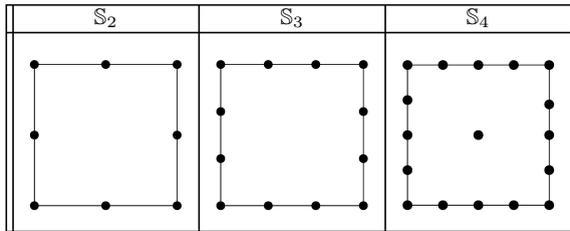
Table 6.3 Two-dimensional serendipity Lagrange finite elements $\mathbb{S}_2$, $\mathbb{S}_3$, and $\mathbb{S}_4$.

Exercises

Exercise 6.1 (Integrated Legendre polynomials). Let $k \geq 2$ and set $\mathbb{P}_k^{(0)} := \{p \in \mathbb{P}_k \mid p(\pm 1) = 0\}$. Show that a basis for $\mathbb{P}_k^{(0)}$ are the integrated

Legendre polynomials $\{\int_{-1}^{t} L_l(s)\,\mathrm{d}s\}_{l\in\{1:\,k-1\}}$. Prove (6.6). (*Hint*: consider moments against polynomials in $\mathbb{P}_{m-2}$ and the derivative at $t=1$.)

Exercise 6.2 (Gauss–Lobatto). The goal of this exercise is to prove Proposition 6.6. (i) Prove that $k_{\mathcal{Q}} = 2m-3$. (*Hint*: for all $p \in \mathbb{P}_{2m-3}$, $m \geq 3$, write $p = p_1(1-t^2)L'_{m-1} + p_2$ with $p_1 \in \mathbb{P}_{m-3}$ and $p_2 \in \mathbb{P}_{m-1}$.) (ii) Prove that $\omega_1 = \omega_m = \frac{2}{m(m-1)}$. (*Hint*: compute $\int_{-1}^{1} L'_{m-1}(t)(1+t)L'_{m-1}(t)\,\mathrm{d}t$ using the quadrature and by integrating by parts.) (iii) Assume $m \geq 3$ and let $l \in \{2:m-1\}$. Prove that $L'_{m-2}(\xi_l) = (1-m)L_{m-1}(\xi_l)$ and $(1-\xi_l^2)L''_{m-1}(\xi_l) + m(m-1)L_{m-1}(\xi_l) = 0$. (*Hint*: use (6.3).) Let $\mathcal{L}_l \in \mathbb{P}_{m-3}$ be the Lagrange interpolation polynomial s.t. $\mathcal{L}_l(\xi_j) = \delta_{lj}$, for all $l,j \in \{2:m-1\}$ (i.e., ξ_1 and ξ_m are excluded). Prove that $\mathcal{L}_l(t) = \frac{L'_{m-1}(t)}{t-\xi_l}\frac{1}{L''_{m-1}(\xi_l)}$. (*Hint*: compare the degree of the polynomials, their roots, and their value at ξ_l.) Finally, prove (6.11). (*Hint*: integrate the polynomial $\mathcal{L}_l(t)(1-t)L'_{m-2}(t)$.) (iv) Let $\|p\|_\xi^2 := \sum_{l\in\{1:m\}} \omega_l p(\xi_l)^2$. Verify that $\|\cdot\|_\xi$ defines a norm on $\mathbb{P}_k$ with $k := m-1$, and prove that $\|p\|_{L^2(K)} \leq \|p\|_\xi \leq (\frac{2k+1}{k})^{\frac{1}{2}}\|p\|_{L^2(K)}$ for all $p \in \mathbb{P}_k$, with $K := (-1,1)$. (*Hint*: write $p = p_{k-1} + \lambda L_k$ with $p_{k-1} \in \mathbb{P}_{k-1}$ and $\lambda \in \mathbb{R}$, and compute $\|p\|_{L^2(K)}^2$ and $\|p\|_\xi^2$.)

Exercise 6.3 (Gauss–Radau). The goal is to prove Proposition 6.7. (i) Prove that $k_{\mathcal{Q}} = 2m-2$. (*Hint*: for all $p \in \mathbb{P}_{2m-2}$, write $p = p_1(L_m - L_{m-1}) + p_2$ with $p_1 \in \mathbb{P}_{m-2}$ and $p_2 \in \mathbb{P}_{m-1}$.) (ii) Prove that $\omega_m = \frac{2}{m^2}$. (*Hint*: integrate the polynomial $\frac{L_m(t)-L_{m-1}(t)}{t-1}L'_{m-1}(t)$.) (iii) Assume $m \geq 2$ and let $l \in \{1:m-1\}$. Prove that $L'_m(\xi_l) = -L'_{m-1}(\xi_l)$. (*Hint*: use (6.3a) and (6.3b).) Let $\mathcal{L}_l \in \mathbb{P}_{m-2}$ be the Lagrange interpolation polynomial s.t. $\mathcal{L}_l(\xi_j) = \delta_{lj}$ for all $l,j \in \{1:m-1\}$ (i.e., ξ_m is excluded). Prove that $\mathcal{L}_l(t) = \frac{L_m(t)-L_{m-1}(t)}{(1-t)(t-\xi_l)}\frac{1-\xi_l}{-2L'_{m-1}(\xi_l)}$. (*Hint*: compare the degree of the polynomials, their roots, and their value at ξ_l.) Finally, prove (6.12). (*Hint*: integrate the polynomial $\mathcal{L}_l(t)(1-t)L'_{m-1}(t)$.)

Exercise 6.4 (Inverse trace inequality). Let $K := [-1,1]^d$. Let $m \geq 3$ and let $\{\xi_l\}_{l\in\{1:m\}}$ be the Gauss–Lobatto (GL) nodes in $[-1,1]$. Set $I_{m,d} := \{1\ldots m\}^d$ and $I_{m,d}^0 := \{2:(m-1)\}^d$. For any $\alpha \in I_{m,d}$, let $\boldsymbol{a}_\alpha \in K$ be the node with Cartesian coordinates $(a_\alpha)_i := \xi_{\alpha_i}$ for all $i \in \{1:d\}$. The set $(\boldsymbol{a}_\alpha)_{\alpha\in I_{m,d}}$ consists of the tensorized GL nodes in K. Let $k := m-1$ and define the polynomial space $\mathbb{Q}_{k,d}^0 := \{q \in \mathbb{Q}_{k,d} \mid q(\boldsymbol{a}_\alpha) = 0, \forall\alpha \in I_{m,d}^0\}$, i.e., polynomials in $\mathbb{Q}_{k,d}^0$ vanish at all the tensorized GL nodes that are located inside K. Prove that

$$\|v\|_{L^2(K)} \leq \left(\frac{2d}{k(k+1)}(2+\tfrac{1}{k})^{d-1}\frac{|K|}{|\partial K|}\right)^{\frac{1}{2}}\|v\|_{L^2(\partial K)},$$

for all $v \in \mathbb{Q}_{k,d}^0$. (*Hint*: use Exercise 6.2.)

Exercise 6.5 (Lagrange mass matrix). Let $\mathcal{M} \in \mathbb{R}^{n_{\mathrm{sh}} \times n_{\mathrm{sh}}}$ be the mass matrix with entries $\mathcal{M}_{ij} := \int_{-1}^{1} \mathcal{L}_{i-1}^{[a]}(t) \mathcal{L}_{j-1}^{[a]}(t) \, dt$ for all $i, j \in \mathcal{N}$. Prove that $\mathcal{M} = (\mathcal{V}^{\mathsf{T}} \mathcal{V})^{-1}$, where $\mathcal{V} \in \mathbb{R}^{n_{\mathrm{sh}} \times n_{\mathrm{sh}}}$ is the (generalized) Vandermonde matrix with entries $\mathcal{V}_{ij} := (\frac{2i-1}{2})^{\frac{1}{2}} L_{i-1}(a_j)$. (*Hint*: see Proposition 5.5.)

Exercise 6.6 (Canonical hybrid element). Prove Proposition 6.10. (*Hint*: use Remark 5.3.) Compute the shape functions when $\mu_l := J_{l-1}^{1,1}$ for all $l \in \{1{:}k{-}1\}$. (*Hint*: consider the polynomials $J_{k-1}^{1,0}$, $J_{l-1}^{1,1}$ for all $l \in \{1{:}k{-}1\}$, and $J_{k-1}^{0,1}$.)

Exercise 6.7 ($\mathbb{Q}_{k,d}$ Lagrange). Prove Proposition 6.14. (*Hint*: observe that any polynomial $q \in \mathbb{Q}_{k,d}$ is such that $q(\boldsymbol{x}) = \sum_{i_d \in \{0:k\}} q_{i_d}(x_1, \ldots, x_{d-1}) x_d^{i_d}$ and use induction on d.)

Exercise 6.8 (Bicubic Hermite). Let K be a rectangle with vertices $\{\boldsymbol{z}_i\}_{1 \le i \le 4}$, $P := \mathbb{Q}_{3,2}$, and $\Sigma := \{p(\boldsymbol{z}_i), \partial_{x_1} p(\boldsymbol{z}_i), \partial_{x_2} p(\boldsymbol{z}_i), \partial_{x_1 x_2}^2 p(\boldsymbol{z}_i)\}_{1 \le i \le 4}$. Show that (K, P, Σ) is a finite element. (*Hint*: write $p \in \mathbb{Q}_{3,2}$ in the form $p(\boldsymbol{x}) = \sum_{i,j \in \{1:4\}} \gamma_{ij} \theta_i(x_1) \theta_j(x_2)$, where $\{\theta_1, \ldots, \theta_4\}$ are the shape functions of the one-dimensional Hermite finite element; see Exercise 5.4.)

Exercise 6.9 (Face unisolvence). Prove Lemma 6.15. (*Hint*: use the hint from Exercise 6.7.)

Part II, Chapter 7

Simplicial finite elements

This chapter deals with finite elements (K, P, Σ) where K is a triangle in $\mathbb{R}^2$, a tetrahedron in $\mathbb{R}^3$, and more generally a simplex in $\mathbb{R}^d$, $d \geq 2$. The degrees of freedom (dofs) Σ are either nodal values in K or integrals over the faces or the edges of K, and P is the space $\mathbb{P}_{k,d}$ composed of multivariate polynomials of total degree at most $k \geq 0$. We focus our attention on scalar-valued finite elements. The results extend to the vector-valued case by reasoning componentwise.

7.1 Simplices

Definition 7.1 (Simplex, vertices, normal). *Let $d \geq 1$. Let $\{\boldsymbol{z}_i\}_{i \in \{0:d\}}$ be a set of points in $\mathbb{R}^d$ such that the vectors $\{\boldsymbol{z}_1 - \boldsymbol{z}_0, \ldots, \boldsymbol{z}_d - \boldsymbol{z}_0\}$ are linearly independent. The convex hull of these points is called* simplex *in $\mathbb{R}^d$, say $K := \mathrm{conv}(\{\boldsymbol{z}_i\}_{i \in \{0:d\}})$. By definition, K is a closed set. The points $\{\boldsymbol{z}_i\}_{i \in \{0:d\}}$ are called* vertices *of K. The outward unit normal vector on ∂K is denoted by $\boldsymbol{n}_K$.*

Example 7.2 ($d \in \{1, 2, 3\}$). A simplex is a compact interval if $d = 1$, a triangle if $d = 2$, and a tetrahedron if $d = 3$ (see Figure 5.2). □

Example 7.3 (Unit simplex). The unit simplex in $\mathbb{R}^d$ is the set $\{\boldsymbol{x} \in \mathbb{R}^d \mid 0 \leq x_i \leq 1, \forall i \in \{1:d\}, \sum_{i \in \{0:d\}} x_i \leq 1\}$. This corresponds to setting $\boldsymbol{z}_0 := \boldsymbol{0}$ and $\boldsymbol{z}_i - \boldsymbol{z}_0 := \boldsymbol{e}_i$ for all $i \in \{1:d\}$, where $\{\boldsymbol{e}_i\}_{i \in \{1:d\}}$ is the canonical Cartesian basis of $\mathbb{R}^d$. The unit simplex has volume $\frac{1}{d!}$. □

Definition 7.4 (Faces, edges). *The convex hull of the set $\{\boldsymbol{z}_0, \ldots, \boldsymbol{z}_d\} \backslash \{\boldsymbol{z}_i\}$ is denoted by F_i for all $i \in \{0:d\}$ and is called the* face *of K opposite to the vertex $\boldsymbol{z}_i$. For all $l \in \{0:d-1\}$, an* l-face *of K is the convex hull of a subset*

of $\{z_i\}_{i \in \{0:d\}}$ of cardinality $(l+1)$ (i.e., usual faces are $(d-1)$-faces). By definition, l-faces are closed sets and are subsets of an affine subspace of $\mathbb{R}^d$ of codimension $(d-l)$. The 0-faces of K are the vertices of K. The 1-faces of K are called edges. In dimension $d = 2$, the notions of edge and face coincide. In dimension $d = 1$, the notions of vertex, edge, and face coincide.

Example 7.5 (Number of faces and edges). The number of l-faces in a simplex in $\mathbb{R}^d$ is equal to $\binom{d+1}{l+1}$, e.g., there are $(d+1)$ faces and vertices, and for $d \geq 2$, there are $\frac{d(d+1)}{2}$ edges.   □

Remark 7.6 (Geometric identities). Let $n_{K|F_i}$ be the value of n_K on F_i for all $i \in \{0:d\}$. Then $\{n_{K|F_i}\}_{i \in \{1:d\}}$ is a basis of $\mathbb{R}^d$. Let c_{F_i} be the barycenter of F_i, c_K that of K, and $\mathbb{I}_d$ the identity matrix in $\mathbb{R}^{d \times d}$. We have

$$\sum_{i \in \{0:d\}} |F_i| n_{K|F_i} = \mathbf{0}, \qquad \sum_{i \in \{0:d\}} |F_i| n_{K|F_i} \otimes (c_{F_i} - c_K) = |K| \mathbb{I}_d. \quad (7.1)$$

See Exercise 7.2. These identities hold true for any polyhedron in $\mathbb{R}^d$.   □

7.2 Barycentric coordinates, geometric mappings

Let K be a simplex in $\mathbb{R}^d$ with vertices $\{z_i\}_{i \in \{0:d\}}$. For all $x \in \mathbb{R}^d$ and all $i \in \{1:d\}$, we denote by $\lambda_i(x)$ the components of the vector $x - z_0$ in the basis $(z_1 - z_0, \ldots, z_d - z_0)$, i.e.,

$$x - z_0 = \sum_{i \in \{1:d\}} \lambda_i(x)(z_i - z_0). \quad (7.2)$$

Differentiating (7.2) twice, we infer that $\sum_{i \in \{0:d\}} D^2 \lambda_i(x)(h_1, h_2)(z_i - z_0) = 0$ for all $h_1, h_2 \in \mathbb{R}^d$. The vectors $\{z_i - z_0\}_{i \in \{1:d\}}$ being linearly independent, this implies that $D^2 \lambda_1(x)(h_1, h_2) = \ldots = D^2 \lambda_d(x)(h_1, h_2) = 0$. Hence, λ_i is an affine function of x, i.e., there exist $\gamma_i \in \mathbb{R}$ and $g_i \in \mathbb{R}^d$ such that $\lambda_i(x) = \gamma_i + g_i \cdot x$ for all $x \in \mathbb{R}^d$, where $a \cdot b$ denotes the inner product in $\mathbb{R}^d$. Note that $D\lambda_i$ is independent of x and $D\lambda_i(h) = g_i \cdot h$ for all $h \in \mathbb{R}^d$. In other words, we have $\nabla \lambda_i = g_i$.

To allow all the vertices of K to play a symmetric role, we introduce the additional function $\lambda_0(x) := 1 - \sum_{i \in \{1:d\}} \lambda_i(x)$. Then we have

$$\sum_{i \in \{0:d\}} \lambda_i(x) = 1 \qquad \text{and} \qquad x = \sum_{i \in \{0:d\}} \lambda_i(x) z_i, \quad (7.3)$$

for all $x \in \mathbb{R}^d$. A consequence of the above definitions is that $\lambda_i(z_j) = \delta_{ij}$ for all $i, j \in \{0:d\}$. This implies that the functions $\{\lambda_i\}_{i \in \{0:d\}}$ are linearly independent: if the linear combination $\sum_{i \in \{0:d\}} \beta_i \lambda_i(x)$ vanishes identically,

evaluating it at the vertex $\boldsymbol{z}_j$ yields $\beta_j = 0$ for all $j \in \{0{:}d\}$. Moreover, since K is the convex hull of $\{\boldsymbol{z}_i\}_{i \in \{0{:}d\}}$, we infer that $0 \leq \lambda_i(\boldsymbol{x}) \leq 1$ for all $\boldsymbol{x} \in K$ and all $i \in \{0{:}d\}$.

Definition 7.7 (Barycentric coordinates). *The functions* $\{\lambda_i\}_{i \in \{0{:}d\}}$ *are called* barycentric coordinates *in K.*

It is shown below that the barycentric coordinates are also the shape functions of the $\mathbb{P}_{1,d}$ Lagrange finite element.

Example 7.8 (Unit simplex). Since $\boldsymbol{x} = \sum_{i \in \{1{:}d\}} x_i \boldsymbol{e}_i$, (7.2) shows that the barycentric coordinates in the unit simplex of $\mathbb{R}^d$ are $\lambda_0(\boldsymbol{x}) := 1 - \sum_{i \in \{1{:}d\}} x_i$ and $\lambda_i(\boldsymbol{x}) := x_i$ for all $i \in \{1{:}d\}$. $\square$

The following construction plays an important role in the rest of the book. Let $\widehat{S}^l := \mathrm{conv}(\{\widehat{\boldsymbol{z}}_j\}_{j \in \{0{:}l\}})$ be the unit simplex in $\mathbb{R}^l$ with barycentric coordinates $\{\widehat{\lambda}_j\}_{j \in \{0{:}l\}}$ (see Example 7.8).

Proposition 7.9 (Geometric mapping). *Let K be a simplex in $\mathbb{R}^d$, let $l \in \{1{:}d\}$, and let $\sigma : \{0{:}l\} \to \{0{:}d\}$ be an injective map, i.e., σ chooses $(l+1)$ distinct integers in $\{0{:}d\}$. Let $S := \mathrm{conv}(\{\boldsymbol{z}_{\sigma(j)}\}_{j \in \{0{:}l\}})$ be an l-face of K or K itself if $l = d$. Let $\boldsymbol{T}_S : \widehat{S}^l \to \mathbb{R}^d$ be the geometric mapping s.t. $\boldsymbol{T}_S(\widehat{\boldsymbol{x}}) = \sum_{j \in \{0{:}l\}} \widehat{\lambda}_j(\widehat{\boldsymbol{x}})\boldsymbol{z}_{\sigma(j)}$ for all $\widehat{\boldsymbol{x}} \in \widehat{S}^l$. Then $S = \boldsymbol{T}_S(\widehat{S}^l)$, and the mapping $\boldsymbol{T}_S$ is a smooth diffeomorphism.*

Proof. We first notice that $\boldsymbol{T}_S(\widehat{\boldsymbol{z}}_j) = \boldsymbol{z}_{\sigma(j)}$ for all $j \in \{0{:}l\}$ and that $\boldsymbol{T}_S$ is an affine mapping since $\boldsymbol{T}_S(\widehat{\boldsymbol{x}}) = \boldsymbol{z}_{i_0} + \sum_{j \in \{1{:}l\}} \widehat{x}_j (\boldsymbol{z}_{\sigma(j)} - \boldsymbol{z}_{i_0})$. Let $\{\theta_j\}_{j \in \{0{:}l\}}$ be any nonnegative numbers s.t. $\sum_{j \in \{0{:}l\}} \theta_j = 1$. We have

$$\sum_{j \in \{0{:}l\}} \theta_j \boldsymbol{z}_{\sigma(j)} = \sum_{j \in \{0{:}l\}} \theta_j \boldsymbol{T}_S(\widehat{\boldsymbol{z}}_j) = \boldsymbol{T}_S\left(\sum_{j \in \{0{:}l\}} \theta_j \widehat{\boldsymbol{z}}_j\right).$$

Since $S = \mathrm{conv}(\{\boldsymbol{z}_{\sigma(j)}\}_{j \in \{0{:}l\}})$ and $\widehat{S}^l = \mathrm{conv}(\{\widehat{\boldsymbol{z}}_j\}_{j \in \{0{:}l\}})$, this proves that $S = \boldsymbol{T}_S(\widehat{S}^l)$. Moreover, the mapping $\boldsymbol{T}_S$ is of class C^∞ since it is linear. We now show that the linear mapping $D\boldsymbol{T}_S : \mathbb{R}^l \to \mathbb{R}^l$ is invertible by verifying the injectivity. Let $\widehat{\boldsymbol{h}} \in \mathbb{R}^l$ be such that $D\boldsymbol{T}_S(\widehat{\boldsymbol{h}}) = \boldsymbol{0}$. Writing $\widehat{\boldsymbol{h}} = \sum_{j \in \{1{:}l\}} \widehat{h}_j(\widehat{\boldsymbol{z}}_j - \widehat{\boldsymbol{z}}_0)$ and since $D\boldsymbol{T}_S(\widehat{\boldsymbol{z}}_j - \widehat{\boldsymbol{z}}_0) = \boldsymbol{T}_S(\widehat{\boldsymbol{z}}_j) - \boldsymbol{T}_S(\widehat{\boldsymbol{z}}_0) = \boldsymbol{z}_{\sigma(j)} - \boldsymbol{z}_{\sigma(0)}$, we infer that $\boldsymbol{0} = \sum_{j \in \{1{:}l\}} \widehat{h}_j(\boldsymbol{z}_{\sigma(j)} - \boldsymbol{z}_{\sigma(0)})$, implying that $\boldsymbol{h} = \boldsymbol{0}$. $\square$

Fig. 7.1 Geometric mapping $\boldsymbol{T}_S$ ($d = 3$, $l = 2$). The face S of K is highlighted in gray, and the vertices of both $\widehat{S}^2$ and S are indicated by bullets.
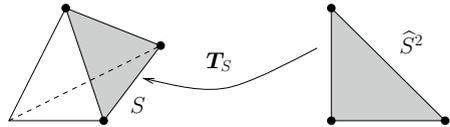
7.3 The polynomial space $\mathbb{P}_{k,d}$

The real vector space $\mathbb{P}_{k,d}$ is composed of d-variate polynomial functions $p : \mathbb{R}^d \to \mathbb{R}$ of total degree at most k. Thus, we have

$$\mathbb{P}_{k,d} := \text{span} \{x_1^{\alpha_1} \ldots x_d^{\alpha_d}, \ 0 \le \alpha_1, \ldots, \alpha_d \le k, \ \alpha_1 + \ldots + \alpha_d \le k\}. \quad (7.4)$$

The importance of the polynomial space $\mathbb{P}_{k,d}$ is rooted in the fact that the Taylor expansion of order k of any d-variate function belongs to $\mathbb{P}_{k,d}$. Another important fact is that for every smooth function $v : \mathbb{R}^d \to \mathbb{R}$,

$$[\,v \in \mathbb{P}_{k,d}\,] \iff [\,D^{k+1}v(\boldsymbol{x}) = 0, \ \forall \boldsymbol{x} \in \mathbb{R}^d\,]. \quad (7.5)$$

The vector space $\mathbb{P}_{k,d}$ has dimension (see Exercise 7.4)

$$\dim \mathbb{P}_{k,d} = \binom{k+d}{d} = \begin{cases} k+1 & \text{if } d = 1, \\ \frac{1}{2}(k+1)(k+2) & \text{if } d = 2, \\ \frac{1}{6}(k+1)(k+2)(k+3) & \text{if } d = 3. \end{cases} \quad (7.6)$$

We omit the subscript d and write $\mathbb{P}_k$ when the context is unambiguous.

An element $\alpha := (\alpha_1, \ldots, \alpha_d)$ of $\mathbb{N}^d$ is called *multi-index*, and its length is defined as $|\alpha| := \alpha_1 + \ldots + \alpha_d$. We define the multi-index set $\mathcal{A}_{k,d} := \{\alpha \in \mathbb{N}^d \mid |\alpha| \le k\}$. Note that $\text{card}(\mathcal{A}_{k,d}) = \dim(\mathbb{P}_{k,d}) = \binom{k+d}{d}$. Any polynomial function $p \in \mathbb{P}_{k,d}$ can be written in the form

$$p(\boldsymbol{x}) = \sum_{\alpha \in \mathcal{A}_{k,d}} a_\alpha \boldsymbol{x}^\alpha, \quad \text{with } \boldsymbol{x}^\alpha := x_1^{\alpha_1} \ldots x_d^{\alpha_d} \text{ and } a_\alpha \in \mathbb{R}. \quad (7.7)$$

Let H be an affine subspace in $\mathbb{R}^d$ of dimension $l \in \{1{:}d{-}1\}$. Given a polynomial $p \in \mathbb{P}_{k,d}$, the following result gives a characterization of the trace of p on H which will be used repeatedly in the book.

Lemma 7.10 (Trace space). *Let H be an affine subspace in $\mathbb{R}^d$ of dimension $l \in \{1{:}d{-}1\}$. Then $p_{|H} \circ \boldsymbol{T}_H \in \mathbb{P}_{k,l}$ for all $p \in \mathbb{P}_{k,d}$ and every affine bijective mapping $\boldsymbol{T}_H : \mathbb{R}^l \to H$. Moreover, $q \circ \boldsymbol{T}_{\mathbb{R}^l} \in \mathbb{P}_{k,d}$ for all $q \in \mathbb{P}_{k,l}$ and every affine mapping $\boldsymbol{T}_{\mathbb{R}^l} : \mathbb{R}^d \to \mathbb{R}^l$.*

Proof. We observe that $D^{k+1}(p_{|H} \circ \boldsymbol{T}_H)(\boldsymbol{y}) = 0$ for all $\boldsymbol{y} \in \mathbb{R}^l$ by using the chain rule and the fact that $\boldsymbol{T}_H$ is affine. Then we apply (7.5). The second statement is proved similarly. $\qquad \square$

7.4 Lagrange (nodal) finite elements

We begin with a simple example where we set $k := 1$; see Table 7.1.

Proposition 7.11 (Simplicial Lagrange, $k := 1$). *Let K be a simplex in $\mathbb{R}^d$ with vertices $\{z_i\}_{i \in \{0:d\}}$. Let $P := \mathbb{P}_{1,d}$. Let $\Sigma := \{\sigma_i\}_{i \in \{0:d\}}$ be the linear forms on P such that $\sigma_i(p) := p(z_i)$ for all $i \in \{0:d\}$. Then (K, P, Σ) is a Lagrange finite element and the shape functions are $\theta_i := \lambda_i$.*

Proof. Let $p \in P$. We use (7.2), i.e., $\boldsymbol{x} - \boldsymbol{z}_0 = \sum_{i \in \{1:d\}} \lambda_i(\boldsymbol{x})(\boldsymbol{z}_i - \boldsymbol{z}_0)$, that p is affine, the linearity of Dp, and the first identity in (7.3) to infer that

$$p(\boldsymbol{x}) = p(\boldsymbol{z}_0) + Dp(\boldsymbol{x} - \boldsymbol{z}_0) = p(\boldsymbol{z}_0) + \sum_{i \in \{1:d\}} \lambda_i(\boldsymbol{x}) Dp(\boldsymbol{z}_i - \boldsymbol{z}_0)$$

$$= \sum_{i \in \{0:d\}} \big(\lambda_i(\boldsymbol{x})(p(\boldsymbol{z}_0) + Dp(\boldsymbol{z}_i - \boldsymbol{z}_0)) \big) = \sum_{i \in \{0:d\}} \lambda_i(\boldsymbol{x}) p(\boldsymbol{z}_i),$$

for all $\boldsymbol{x} \in \mathbb{R}^d$. Now we use Remark 5.3. We have $\dim P = d + 1 = \operatorname{card} \Sigma$, and the above identity shows that any polynomial in P vanishing at the $(d+1)$ vertices of K vanishes identically. Hence, (K, P, Σ) is a finite element. Finally, owing to the above identity applied with $p := \theta_j$, we have $\theta_j(\boldsymbol{x}) = \sum_{i \in \{0:d\}} \lambda_i(\boldsymbol{x}) \theta_j(\boldsymbol{z}_i) = \sum_{i \in \{0:d\}} \lambda_i(\boldsymbol{x}) \delta_{ij} = \lambda_j(\boldsymbol{x})$ for all $\boldsymbol{x} \in K$. This proves that $\theta_j = \lambda_j$ for all $j \in \{0:d\}$. $\qquad \square$

We now extend the above construction to any polynomial order $k \geq 1$ using equidistributed nodes in the simplex K. Other choices are discussed in Remark 7.14.

Proposition 7.12 (Simplicial Lagrange). *Let K be a simplex in $\mathbb{R}^d$. Let $k \geq 1$, $P := \mathbb{P}_{k,d}$, and $\mathcal{A}_{k,d} := \{\alpha \in \mathbb{N}^d \mid |\alpha| \leq k\}$. Set $n_{\mathrm{sh}} := \binom{k+d}{d}$ and consider the set of nodes $\{\boldsymbol{a}_\alpha\}_{\alpha \in \mathcal{A}_{k,d}}$ s.t. $\boldsymbol{a}_\alpha - \boldsymbol{z}_0 := \sum_{i \in \{1:d\}} \frac{\alpha_i}{k}(\boldsymbol{z}_i - \boldsymbol{z}_0)$. Let $\Sigma := \{\sigma_\alpha\}_{\alpha \in \mathcal{A}_{k,d}}$ be the linear forms on P s.t. $\sigma_\alpha(p) := p(\boldsymbol{a}_\alpha)$ for all $\alpha \in \mathcal{A}_{k,d}$. Then (K, P, Σ) is a Lagrange finite element.*

Proof. We use Remark 5.3. Since $\operatorname{card} \Sigma = \operatorname{card} \mathcal{A}_{k,d} = \binom{k+d}{d} = \dim \mathbb{P}_{k,d}$, we need to prove the following property which we call $[\mathcal{P}_{k,d}]$: Any polynomial $p \in \mathbb{P}_{k,d}$ vanishing at all the Lagrange nodes $\{\boldsymbol{a}_\alpha\}_{\alpha \in \mathcal{A}_{k,d}}$ of any simplex in $\mathbb{R}^d$ vanishes identically. Property $[\mathcal{P}_{k,1}]$ holds true for all $k \geq 1$ owing to Proposition 6.8. Assume now that $d \geq 2$ and that $[\mathcal{P}_{k,d-1}]$ holds true for all $k \geq 1$ and let us prove that $[\mathcal{P}_{k,d}]$ holds true for all $k \geq 1$. Assume that $p \in \mathbb{P}_{k,d}$ vanishes at all the Lagrange nodes of a simplex K. Let F_0 be the face of K opposite to the vertex $\boldsymbol{z}_0$ and consider an affine bijective mapping $\boldsymbol{T}_{H_0} : \mathbb{R}^{d-1} \to H_0$, where H_0 is the affine hyperplane supporting F_0. Then $p_0 := p \circ \boldsymbol{T}_{H_0}$ is in $\mathbb{P}_{k,d-1}$ owing to Lemma 7.10, and by assumption, $p_0(\boldsymbol{T}_{H_0}^{-1}(\boldsymbol{a}_\alpha)) = p(\boldsymbol{a}_\alpha) = 0$ for all $\boldsymbol{a}_\alpha \in F_0$. Moreover, $\boldsymbol{a}_\alpha \in F_0$ iff $|\alpha| = k$. Let us set $\tilde{\beta} := (k - |\beta|, \beta_1, \ldots, \beta_{d-1})$ for all $\beta \in \mathcal{A}_{k,d-1}$, so that $\tilde{\beta} \in \mathcal{A}_{k,d}$ and $|\tilde{\beta}| = k$. Setting $\boldsymbol{b}_\beta := \boldsymbol{T}_{H_0}^{-1}(\boldsymbol{a}_{\tilde{\beta}})$ for all $\beta \in \mathcal{A}_{k,d-1}$, we obtain all the Lagrange nodes of the simplex $\boldsymbol{T}_{H_0}^{-1}(F)$ in $\mathbb{R}^{d-1}$. Since $p_0(\boldsymbol{b}_\beta) = p(\boldsymbol{a}_{\tilde{\beta}}) = 0$ for all $\beta \in \mathcal{A}_{k,d-1}$, we infer owing to $[\mathcal{P}_{k,d-1}]$ that $p_0 = 0$. Since $\boldsymbol{T}_{H_0}$ is bijective, we

obtain $p_{|F_0} = 0$. Denoting by $\lambda_0 \in \mathbb{P}_{1,d}$ the barycentric coordinate associated with $\boldsymbol{z}_0$, this implies that there is $q \in \mathbb{P}_{k-1,d}$ s.t. $p = \lambda_0 q$ (see Exercise 7.4(iv)). Let us prove by induction on k that $q = 0$. For $k = 1$, we have already proved $[\mathcal{P}_{1,d}]$ in Proposition 7.11. Let us now assume that $[\mathcal{P}_{k-1,d}]$ holds true for $k \geq 2$. Since $k \geq 2$, q vanishes at all the Lagrange nodes $\boldsymbol{a}_\alpha$ s.t. $|\alpha| < k$ (since $\lambda_0(\boldsymbol{a}_\alpha) \neq 0$ at these nodes), i.e., $|\alpha| \leq k - 1$. Hence, q vanishes at all the Lagrange nodes $\boldsymbol{a}_\alpha$, $\alpha \in \mathcal{A}_{k-1,d}$. Since these nodes belong again to a simplex, $[\mathcal{P}_{k-1,d}]$ implies $q = 0$.   □

We have established the following result in the proof of Proposition 7.12.

Lemma 7.13 (Face unisolvence). *Let F be one of the $(d+1)$ faces of the simplex $K \subset \mathbb{R}^d$. Let $\mathcal{N}_F$ be the collection of the indices of the Lagrange nodes on F. The following holds true for all $p \in \mathbb{P}_{k,d}$:*

$$[\,\sigma_j(p) = 0, \ \forall j \in \mathcal{N}_F\,] \iff [\,p_{|F} = 0\,]. \tag{7.8}$$

$\mathbb{P}_1$	$\mathbb{P}_2$	$\mathbb{P}_3$
λ_i	$\lambda_i(2\lambda_i - 1)$ $4\lambda_i\lambda_j$	$\frac{1}{2}\lambda_i(3\lambda_i - 1)(3\lambda_i - 2)$ $\pm\frac{9}{2}\lambda_i(3\lambda_i - \frac{3}{2} \pm \frac{1}{2})\lambda_j$ $27\lambda_i\lambda_j\lambda_k$

Table 7.1 Two- and three-dimensional $\mathbb{P}_1$, $\mathbb{P}_2$, and $\mathbb{P}_3$ Lagrange elements. Visible degrees of freedom are shown in black, hidden degrees of freedom are in white, and hidden edges are represented with dashed lines. The shape functions are expressed in terms of the barycentric coordinates. The first, second, and third lines list shape functions associated with the vertices ($i \in \{0{:}d\}$), the edges ($i, j \in \{0{:}d\}$, $i < j$), and the faces ($i, j, k \in \{0{:}d\}$, $i < j < k$).

Table 7.1 presents examples of node locations and shape functions for $k \in \{1, 2, 3\}$ in dimension $d \in \{2, 3\}$. The bullets conventionally indicate the location of the nodes; see Exercise 7.5 for some properties of these nodes.

Possible choices for the domain of the interpolation operator are $V(K) :=$ $C^0(K)$ or $V(K) := W^{s,p}(K)$ with $p \in [1, \infty]$ and $sp > d$ (or $s \geq d$ if $p = 1$); see §5.4.1.

Remark 7.14 (High-order). Other sets of Lagrange nodes can be used. For instance, the Fekete points from §6.3.5 can be extended to simplices, although finding Fekete points on simplices for high polynomial degrees is a difficult problem. We refer the reader to Chen and Babuška [66] and Taylor et al. [190] for results on triangles with degrees up to $k = 13$ and $k = 19$, respectively; see also Canuto et al. [58, p. 112]. A comparison of various nodal sets on triangles and tetrahedra can be found in Blyth et al. [26]. □

Remark 7.15 (Modal and hybrid simplicial elements). A hierarchical basis of $\mathbb{P}_{k,d}$ can be built by combining a hierarchical univariate basis of $\mathbb{P}_{k,1}$ with the barycentric coordinates; see Ainsworth and Coyle [6] and Exercise 7.6. One can also introduce a nonlinear transformation mapping the simplex to a cuboid and use tensor products of one-dimensional basis functions in the cuboid; see Proriol [162], Dubiner [91], Owens [154], Karniadakis and Sherwin [123, §3.2]. Another possibility is to use *Bernstein polynomials*, i.e., the basis $\{\binom{p}{m} t^m (1-t)^{p-m}\}_{m \in \{0:p\}}$ if $d = 1$; see Ainsworth et al. [7], Kirby [125] for scalar-valued polynomials and Kirby [126] for the extension to the de Rham complex (see also §16.3). □
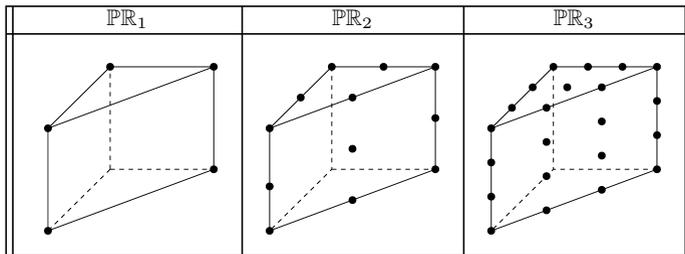
Table 7.2 Nodes for prismatic Lagrange finite elements of degree 1, 2, and 3. The bullets indicate the location of the nodes. Only visible nodes are shown.

Remark 7.16 (Prismatic Lagrange elements). Let $d \geq 3$ and set $\boldsymbol{x}' :=$ $(x_1, \ldots, x_{d-1})$ for all $\boldsymbol{x} \in \mathbb{R}^d$. Let K' be a simplex in $\mathbb{R}^{d-1}$ and $[z_d^-, z_d^+]$ be an interval with $z_d^- < z_d^+$. The set $K := \{\boldsymbol{x} \in \mathbb{R}^d \mid \boldsymbol{x}' \in K', \; x_d \in [z_d^-, z_d^+]\}$ is called *prism* in $\mathbb{R}^d$. Let $k \geq 1$ and let $\mathbb{PR}_k := \text{span}\{p(\boldsymbol{x}) = p_1(\boldsymbol{x}') p_2(x_d) \mid p_1 \in \mathbb{P}_{k,d-1}, \; p_2 \in \mathbb{P}_{k,1}\}$. Examples of prismatic Lagrange elements based on K and $\mathbb{PR}_k$ with equidistributed nodes are shown in Table 7.2 for $k \in \{1, 2, 3\}$. □

7.5 Crouzeix–Raviart finite element

The Crouzeix–Raviart finite element is based on the polynomial space $\mathbb{P}_{1,d}$. It has been introduced in [86] to approximate the Stokes equations. Let K be a simplex in $\mathbb{R}^d$ with vertices $\{z_i\}_{i\in\{0:d\}}$. Recall that the face of K opposite to z_i is denoted by F_i.

Proposition 7.17 (Finite element). *Let K be a simplex in $\mathbb{R}^d$, set $P := \mathbb{P}_{1,d}$, and define the following dofs on P:*

$$\sigma_i^{\mathrm{CR}}(p) := \frac{1}{|F_i|} \int_{F_i} p \, \mathrm{d}s, \qquad \forall i \in \{0{:}d\}. \tag{7.9}$$

Set $\Sigma := \{\sigma_i^{\mathrm{CR}}\}_{i\in\{0:d\}}$. Then (K, P, Σ) is a finite element.

Proof. Since $\operatorname{card}\Sigma = \dim P = d+1$, it suffices to verify that any polynomial p in P satisfying $\sigma_i^{\mathrm{CR}}(p) = \frac{1}{|F_i|} \int_{F_i} p \, \mathrm{d}s = 0$ for all $i \in \{0{:}d\}$ vanishes identically. Since $p \in \mathbb{P}_{1,d}$, we have $p = \sum_{j\in\{0:d\}} p(z_j)\lambda_j$, where $\{\lambda_j\}_{j\in\{0:d\}}$ are the barycentric coordinates in K. Owing to Exercise 7.3(iii), we infer that $\sigma_i^{\mathrm{CR}}(p) = \sum_{j\in\{0:d\}} p(z_j)\sigma_i^{\mathrm{CR}}(\lambda_j) = \frac{1}{d}\sum_{j\neq i} p(z_j)$ since $\sigma_i^{\mathrm{CR}}(\lambda_i) = 0$ and $\sigma_i^{\mathrm{CR}}(\lambda_j) = \frac{1}{d}|F_i|$ for all $j \neq i$. Hence, $\sum_{j\neq i} p(z_j) = 0$ for all $i \in \{0{:}d\}$. This implies that $0 = \sum_{j\neq i} p(z_j) - \sum_{j\neq i'} p(z_j) = p(z_i) - p(z_{i'})$ for every pair (i, i') such that $i \neq i'$. Hence, p takes a constant value at all the vertices of K, and this value must be zero since, say, $\sum_{j\neq 0} p(z_j) = 0$. $\qquad \square$

Using the barycentric coordinates $\{\lambda_i\}_{i\in\{0:d\}}$ in K, one can verify that the shape functions are $\theta_i^{\mathrm{CR}}(\boldsymbol{x}) := 1 - d\lambda_i(\boldsymbol{x})$ for all $i \in \{0{:}d\}$ and all $\boldsymbol{x} \in K$. Note that $\theta_{i|F_i}^{\mathrm{CR}} = 1$ and $\theta_i^{\mathrm{CR}}(z_i) = 1 - d$. The Crouzeix–Raviart interpolation operator acts as follows:

$$\mathcal{I}_K^{\mathrm{CR}}(v)(\boldsymbol{x}) := \sum_{i\in\{0:d\}} \sigma_{K,i}^{\mathrm{CR}}(v)\theta_{K,i}^{\mathrm{CR}}(\boldsymbol{x}) = \sum_{i\in\{0:d\}} \left(\frac{1}{|F_i|} \int_{F_i} v \, \mathrm{d}s \right) \theta_i^{\mathrm{CR}}(\boldsymbol{x}), \tag{7.10}$$

for all $\boldsymbol{x} \in K$. A possible choice for the domain of $\mathcal{I}_K^{\mathrm{CR}}$ is $V(K) := W^{1,1}(K)$ since the trace theorem (Theorem 3.10) applied with $p := 1$ implies that any function in $W^{1,1}(K)$ has a trace in $L^1(\partial K)$. The two- and three-dimensional Crouzeix–Raviart elements are shown in Table 7.3.

Remark 7.18 (Definition as a Lagrange element). The mean-value over a face of a polynomial in $\mathbb{P}_{1,d}$ is equal to the value this polynomial takes at the barycenter of the face. Another possible choice for the dofs is therefore to take the values at the barycenter of all the faces. The resulting finite element is a Lagrange finite element (see Definition 5.11), and $W^{1,1}(K)$ is no longer a legitimate domain for the interpolation operator. One possible choice is the smaller space $V(K) := C^0(K)$. $\qquad \square$

Table 7.3 $\mathbb{P}_1$ Crouzeix–Raviart elements in dimensions two and three. Visible degrees of freedom are shown in black, hidden degrees of freedom are in white, and hidden edges are represented with dashed lines. The shape functions are expressed in terms of the barycentric coordinates.
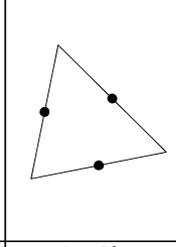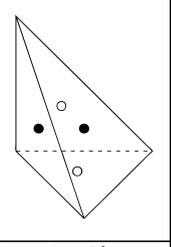
| $1-2\lambda_i$ | $1-3\lambda_i$ |

7.6 Canonical hybrid finite element

We now present a finite element based on the polynomial space $\mathbb{P}_{k,d}$ whose dofs combine values at the vertices of the simplex K with integrals over the l-faces of K for $l \geq 1$ (hence the name hybrid). It is a useful alternative to Lagrange elements that has interesting commuting properties, which will be invoked in §16.3 in the context of the discrete de Rham complex (hence the name canonical).

Let K be a tetrahedron in $\mathbb{R}^3$. Let $\mathcal{V}_K$, $\mathcal{E}_K$, and $\mathcal{F}_K$ be the collections of the vertices, edges, and faces of K, respectively. Let $\boldsymbol{T}_E : \widehat{S}^1 \to E$ for all $E \in \mathcal{E}_K$, and $\boldsymbol{T}_F : \widehat{S}^2 \to F$ for all $F \in \mathcal{F}_K$, be affine bijective mappings (see Proposition 7.9), where $\widehat{S}^1$ and $\widehat{S}^2$ are the unit simplices in $\mathbb{R}$ and $\mathbb{R}^2$. Let $k \geq 1$ be the polynomial degree. The canonical hybrid finite element involves vertex dofs, edge dofs if $k \geq 2$, surface (or face) dofs if $k \geq 3$, and volume (or cell) dofs if $k \geq 4$. We consider the following dofs:

$$\sigma_{\boldsymbol{z}}^{\mathrm{v}}(p) := p(\boldsymbol{z}), \qquad\qquad \boldsymbol{z} \in \mathcal{V}_K, \qquad\qquad (7.11a)$$

$$\sigma_{E,m}^{\mathrm{e}}(p) := \frac{1}{|E|} \int_E (\mu_m \circ \boldsymbol{T}_E^{-1}) p \,\mathrm{d}l, \qquad E \in \mathcal{E}_K,\ m \in \{1{:}n_{\mathrm{sh}}^{\mathrm{e}}\}, \quad (7.11b)$$

$$\sigma_{F,m}^{\mathrm{f}}(p) := \frac{1}{|F|} \int_F (\zeta_m \circ \boldsymbol{T}_F^{-1}) p \,\mathrm{d}s, \qquad F \in \mathcal{F}_K,\ m \in \{1{:}n_{\mathrm{sh}}^{\mathrm{f}}\}, \quad (7.11c)$$

$$\sigma_m^{\mathrm{c}}(p) := \frac{1}{|K|} \int_K \psi_m p \,\mathrm{d}x, \qquad\qquad m \in \{1{:}n_{\mathrm{sh}}^{\mathrm{c}}\}, \qquad (7.11d)$$

where $\{\mu_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$ is a basis of $\mathbb{P}_{k-2,1}$ with $n_{\mathrm{sh}}^{\mathrm{e}} := \binom{k-1}{1}$ if $k \geq 2$, $\{\zeta_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ is a basis of $\mathbb{P}_{k-3,2}$ with $n_{\mathrm{sh}}^{\mathrm{f}} := \binom{k-1}{2}$ if $k \geq 3$, and $\{\psi_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ is a basis of $\mathbb{P}_{k-4,3}$ with $n_{\mathrm{sh}}^{\mathrm{c}} := \binom{k-1}{3}$ if $k \geq 4$. The above construction is possible in any dimension. If $d = 2$ for instance, the vertex dofs are defined in (7.11a), the edge (face) ones in (7.11b) if $k \geq 2$, and the cell ones in (7.11d), where $\{\psi_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ is a basis of $\mathbb{P}_{k-3,2}$ with $n_{\mathrm{sh}}^{\mathrm{c}} := \binom{k-1}{2}$ if $k \geq 3$.

Proposition 7.19 (Canonical hybrid finite element). *Let $k \geq 1$. Let K be a simplex in $\mathbb{R}^d$, let $P := \mathbb{P}_{k,d}$, and let $\Sigma := \{\sigma_i\}_{i \in \mathcal{N}}$ be the collection of all the dofs defined in* (7.11). *Then (K, P, Σ) is a finite element.*

Proof. We use Remark 5.3. Since we use polynomials in $\mathbb{P}_{k-l-1,l}$ to define the dofs of the l-faces, and the number of l-faces is $\binom{d+1}{l+1} = \binom{d+1}{d-l}$ (see Example 7.5), the total number of dofs for all the l-faces is $\binom{k-1}{l}\binom{d+1}{d-l}$. Vandermonde's convolution identity implies that

$$ n_{\mathrm{sh}} = \sum_{j \in \{0 : d\}} \binom{k-1}{j}\binom{d+1}{d-j} = \binom{k+d}{d} = \dim(\mathbb{P}_{k,d}). $$

It remains to prove that if $p \in \mathbb{P}_{k,d}$ is such that $\sigma_i(p) = 0$ for all $i \in \mathcal{N}$, then p vanishes identically. First, p vanishes at all the vertices of K. If $k = 1$, this concludes the proof. If $k \geq 2$, fix an edge E of K. Since $p \circ \boldsymbol{T}_E$ vanishes at the two endpoints of E, $p \circ \boldsymbol{T}_E = \lambda_0 \lambda_1 q$, where $\lambda_0, \lambda_1 \in \mathbb{P}_{1,1}$ are the local barycentric coordinates over $\widehat{S}^1$ and $q \in \mathbb{P}_{k-2,1}$. Since the dofs of p attached to E vanish, we infer that $\int_{\widehat{S}^1} \lambda_0 \lambda_1 q^2 \, \mathrm{d}l = 0$, which implies that $q = 0$. Hence, p is identically zero on all edges of K. If $k = 2$, this completes the proof since all the Lagrange nodes for $k = 2$ are located at the edges of K. If $k \geq 3$, we proceed similarly by fixing a face F of K and showing that p is identically zero on all faces of K. If $k = 3$, this completes the proof since all the Lagrange nodes for $k = 3$ are located at the faces of K. For $k \geq 4$, we finally infer that $p = \lambda_0 \ldots \lambda_d q_K$ where $\{\lambda_i\}_{i \in \{0 : d\}}$ are the barycentric coordinates of K and $q_K \in \mathbb{P}_{k-4,d}$. Since the dofs of p attached to K vanish, we infer that $\int_K \lambda_0 \ldots \lambda_d q_K^2 \, \mathrm{d}x = 0$, which implies that $q_K = 0$, i.e., $p = 0$. $\qquad\square$

The shape functions associated with the vertices, the edges, the faces, and K are denoted by $\{\tilde{\xi}_{\boldsymbol{z}}\}_{\boldsymbol{z} \in \mathcal{V}_K}$, $\{\tilde{\mu}_{E,m}\}_{E \in \mathcal{E}_K, m \in \{1 : n_{\mathrm{sh}}^{\mathrm{e}}\}}$, $\{\tilde{\zeta}_{F,m}\}_{F \in \mathcal{F}_K, m \in \{1 : n_{\mathrm{sh}}^{\mathrm{f}}\}}$, and $\{\tilde{\psi}_m\}_{m \in \{1 : n_{\mathrm{sh}}^{\mathrm{c}}\}}$, respectively. All these functions are in $\mathbb{P}_{k,d}$ and form a basis thereof. Recalling Proposition 5.5 the shape functions are computed by inverting the generalized Vandermonde matrix $\mathcal{V}$ after choosing a basis of $\mathbb{P}_{k,d}$. A basis of $\mathbb{P}_{k,d}$ with a structure close to that of the above shape functions can be found in Fuentes et al. [103, §7.1]. The proposed basis can be organized into functions attached to the vertices of K, the edges of K, the faces of K, and to K itself, and the associated generalized Vandermonde matrix $\mathcal{V}$ is block-triangular. The interpolation operator has domain $V(K) := C^0(K)$ (or $V(K) := W^{s,p}(K)$ with $sp > d$, $p \in [1, \infty]$ or $s \geq d$, $p = 1$) and it acts as follows:

$$ \mathcal{I}_K^{\mathrm{g}}(v)(\boldsymbol{x}) := \sum_{\boldsymbol{z} \in \mathcal{V}_K} \sigma_{\boldsymbol{z}}^{\mathrm{v}}(v) \tilde{\xi}_{\boldsymbol{z}}(\boldsymbol{x}) + \sum_{E \in \mathcal{E}_K} \sum_{m \in \{1 : n_{\mathrm{sh}}^{\mathrm{e}}\}} \sigma_{E,m}^{\mathrm{e}}(v) \tilde{\mu}_{E,m}(\boldsymbol{x}) $$

$$ + \sum_{F \in \mathcal{F}_K} \sum_{m \in \{1 : n_{\mathrm{sh}}^{\mathrm{f}}\}} \sigma_{F,m}^{\mathrm{f}}(v) \tilde{\zeta}_{F,m}(\boldsymbol{x}) + \sum_{m \in \{1 : n_{\mathrm{sh}}^{\mathrm{c}}\}} \sigma_m^{\mathrm{c}}(v) \tilde{\psi}_m(\boldsymbol{x}). $$

Remark 7.20 (Dofs). The interpolation operator $\mathcal{I}_K^{\mathrm{g}}$ is independent of the bases $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$, $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$, and $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ (this follows from Exercise 5.2). It is also independent of the choice of the mappings $\boldsymbol{T}_E$ and $\boldsymbol{T}_F$. Let for instance $\boldsymbol{T}_F$ and $\tilde{\boldsymbol{T}}_F$ be two geometric mappings associated with the face F. Then $\boldsymbol{T}_F^{-1}\circ\tilde{\boldsymbol{T}}_F$ is affine and bijective from $\mathbb{R}^2$ to $\mathbb{R}^2$. Hence, $\zeta_m\circ(\boldsymbol{T}_F^{-1}\circ\tilde{\boldsymbol{T}}_F)\in\mathbb{P}_{k,2}$ for all $m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}$, so that $\zeta_m\circ(\boldsymbol{T}_F^{-1}\circ\tilde{\boldsymbol{T}}_F)=\sum_{n\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}\mathcal{S}_{mn}\zeta_n$ for some real numbers $\mathcal{S}_{mn}$, i.e.,

$$\zeta_m\circ\boldsymbol{T}_F^{-1}=\sum_{n\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}\mathcal{S}_{mn}(\zeta_n\circ\tilde{\boldsymbol{T}}_F^{-1}).$$

Since the mappings $\boldsymbol{T}_F$ and $\tilde{\boldsymbol{T}}_F$ are bijective, the matrix $\mathcal{S}\in\mathbb{R}^{n_{\mathrm{sh}}^{\mathrm{f}}\times n_{\mathrm{sh}}^{\mathrm{f}}}$ is invertible, and we use again Exercise 5.2 to conclude. $\qquad\square$

Exercises

Exercise 7.1 (Lagrange interpolation). Let $\mathcal{I}_K$ be the $\mathbb{P}_1$ Lagrange interpolation operator on a simplex K. Prove that $\|\mathcal{I}_K(v)\|_{C^0(K)}\leq\|v\|_{C^0(K)}$ for all $v\in C^0(K)$. (*Hint*: use the convexity of K and recall that K is closed.) Does this property hold true for $\mathbb{P}_2$ Lagrange elements?

Exercise 7.2 (Geometric identities). Prove the statements in Remark 7.6. (*Hint*: use the divergence theorem to prove (7.1).)

Exercise 7.3 (Barycentric coordinates). Let K be a simplex in $\mathbb{R}^d$. (i) Prove that $\lambda_i(\boldsymbol{x})=1-\frac{|F_i|}{d|K|}\boldsymbol{n}_{K|F_i}\cdot(\boldsymbol{x}-\boldsymbol{z}_i)$ for all $\boldsymbol{x}\in K$ and all $i\in\{0:d\}$, and that $\nabla\lambda_i=-\frac{|F_i|}{d|K|}\boldsymbol{n}_{K|F_i}$. (ii) For all $\boldsymbol{x}\in K$, let $K_i(\boldsymbol{x})$ be the simplex obtained by joining $\boldsymbol{x}$ to the d vertices $\boldsymbol{z}_j$ with $j\neq i$. Show that $\lambda_i(\boldsymbol{x})=\frac{|K_i(\boldsymbol{x})|}{|K|}$. (iii) Prove that $\int_K\lambda_i\,\mathrm{d}x=\frac{1}{d+1}|K|$ for all $i\in\{0:d\}$, and that $\int_{F_j}\lambda_i\,\mathrm{d}s=\frac{1}{d}|F_j|$ for all $j\in\{0:d\}$ with $j\neq i$, and $\int_{F_i}\lambda_i\,\mathrm{d}s=0$. (*Hint*: consider an affine mapping from K to the unit simplex.) (iv) Prove that if $\boldsymbol{h}\in\mathbb{R}^d$ satisfies $D\lambda_i(\boldsymbol{h})=0$ for all $i\in\{1:d\}$, then $\boldsymbol{h}=\boldsymbol{0}$.

Exercise 7.4 (Space $\mathbb{P}_{k,d}$). (i) Give a basis for $\mathbb{P}_{2,d}$ for $d\in\{1,2,3\}$. (ii) Show that any polynomial $p\in\mathbb{P}_{k,d}$ can be written in the form $p(x_1,\dots,x_d)=r(x_1,\dots,x_{d-1})+x_dq(x_1,\dots,x_d)$, with unique polynomials $r\in\mathbb{P}_{k,d-1}$ and $q\in\mathbb{P}_{k-1,d}$. (iii) Determine the dimension of $\mathbb{P}_{k,d}$. (*Hint*: by induction on d.) (iv) Let K be a simplex in $\mathbb{R}^d$. Let F_0 be the face of K opposite to the vertex $\boldsymbol{z}_0$. Prove that if $p\in\mathbb{P}_{k,d}$ satisfies $p_{|F_0}=0$, then there is $q\in\mathbb{P}_{k-1,d}$ s.t. $p=\lambda_0q$. (*Hint*: write the Taylor expansion of p at $\boldsymbol{z}_d$ and use (7.2) with $\boldsymbol{z}_d$ playing the role of $\boldsymbol{z}_0$.) (v) Prove that $\{\lambda_0^{\beta_0}\dots\lambda_d^{\beta_d}\mid\beta_0+\dots+\beta_d=k\}$ is a basis of $\mathbb{P}_{k,d}$.

Exercise 7.5 (Nodes of simplicial Lagrange FE). Let K be a simplex in $\mathbb{R}^d$, and consider the set of nodes $\{a_i\}_{i\in\mathcal{N}}$ with barycentric coordinates $\left(\frac{i_0}{k},\ldots,\frac{i_d}{k}\right)$, $\forall i_0,\ldots,i_d \in \{0{:}k\}$ with $i_0 + \ldots + i_d = k$. (i) Prove that the number of nodes located on any one-dimensional edge of K is $(k+1)$ in any dimension $d \geq 2$. (ii) Prove that the number of nodes located on any $(d-1)$-dimensional face of K is the dimension of $\mathbb{P}_{k,d-1}$. (iii) Prove that if $k \leq d$, all the nodes are located on the boundary of K.

Exercise 7.6 (Hierarchical basis). Let $k \geq 1$ and let $\{\theta_0,\ldots,\theta_k\}$ be a hierarchical basis of $\mathbb{P}_{k,1}$. Let $\{\lambda_0,\ldots,\lambda_d\}$ be a basis of $\mathbb{P}_{1,d}$ and assume that $\lambda_i : \mathbb{R}^d \to \mathbb{R}$ is surjective for all $i \in \{0{:}d\}$ (i.e., λ_i is not constant). (i) Show that the functions (mapping $\mathbb{R}^d$ to $\mathbb{R}$) $\{\theta_0(\lambda_i),\ldots,\theta_k(\lambda_i)\}$ are linearly independent for all $i \in \{0{:}d\}$. (*Hint*: consider a linear combination $\sum_{l\in\{0:k\}} \alpha_l \theta_l(\lambda_i) \in \mathbb{P}_{k,d}$ and prove that the polynomial $\sum_{l\in\{0:k\}} \alpha_l \theta_l \in \mathbb{P}_{k,1}$ vanishes at $(k+1)$ distinct points.) (ii) Show that the functions (mapping $\mathbb{R}^d$ to $\mathbb{R}$) from the set $S_{k,d} := \{\theta_{\alpha_1}(\lambda_1)\ldots\theta_{\alpha_d}(\lambda_d) \mid (\alpha_1,\ldots\alpha_d) \in \mathbb{N}^d,\ |\alpha| \leq k\}$ are linearly independent. (*Hint*: by induction on d.) (iii) Show that $(S_{k,d})_{k\geq 0}$ is a hierarchical polynomial basis, i.e., $S_{k,d} \subset S_{k+1,d}$ and $S_{k,d}$ is basis of $\mathbb{P}_{k,d}$. (*Note*: the $(d+1)$ vertices of K do not play here the same role.)

Exercise 7.7 (Cubic Hermite triangle). Let K be a triangle with vertices $\{z_0,z_1,z_2\}$. Set $\Sigma := \{p(z_i),\partial_{x_1}p(z_i),\partial_{x_2}p(z_i)\}_{0\leq i\leq 2} \cup \{p(a_K)\}$, where a_K is a point inside K. Show that $(K,\mathbb{P}_{3,2},\Sigma)$ is a finite element. (*Hint*: show that any $p \in \mathbb{P}_{3,2}$ for which all the dofs vanish is identically zero on the three edges of K and infer that $p = c\lambda_0\lambda_1\lambda_2$ for some $c \in \mathbb{R}$.)

Exercise 7.8 ($\mathbb{P}_{2,d}$ canonical hybrid FE). Compute the shape functions of the $\mathbb{P}_{2,d}$ canonical hybrid finite element for the unit simplex for $d = 1$ and $d = 2$ (provide an expression using the Cartesian coordinates and another one using the barycentric coordinates).

Exercise 7.9 ($\mathbb{P}_{4,2}$ Lagrange). Using the Lagrange nodes defined as in Proposition 7.11, give the expression of the $\mathbb{P}_{4,2}$ Lagrange shape functions in terms of the barycentric coordinates.

Exercise 7.10 (Quadratic Crouzeix–Raviart). Let K be the unit simplex. Let $\alpha \in (0,1)$. Let $g_1 := (\alpha,0)$, $g_2 := (1-\alpha,0)$, $g_3 := (1-\alpha,\alpha)$, $g_4 := (\alpha,1-\alpha)$, $g_5 := (0,1-\alpha)$, $g_6 := (0,\alpha)$. (i) Compute $\lambda_0(g_j)^2 + \lambda_1(g_j)^2 + \lambda_2(g_j)^2$ for all $j \in \{1{:}6\}$, where $\lambda_0,\lambda_1,\lambda_2$ are the barycentric coordinates of K. (ii) Let $\sigma_j \in \mathcal{L}(\mathbb{P}_{2,2};\mathbb{R})$ be defined by $\sigma_j(p) := p(g_j)$ for all $p \in \mathbb{P}_{2,2}$ and $j \in \{1{:}6\}$. Let $\Sigma := \{\sigma_j\}_{j\in\{1:6\}}$. Is the triple $(K,\mathbb{P}_{2,2},\Sigma)$ a finite element? (iii) Let F_i, $i \in \{0{:}2\}$, be one of the three faces of K. Let $T_{F_i} : [-1,1] \to F_i$ be one of the two affine mappings that realize a bijection between $[-1,1]$ and F_i. Let $\{q_0,q_1\}$ be a basis of $\mathbb{P}_{1,1}$. Let $\varpi_{2i+k} \in \mathcal{L}(\mathbb{P}_{2,2};\mathbb{R})$, $i \in \{0{:}2\}$, $k \in \{0{:}1\}$, be defined by $\varpi_{2i+k}(p) := \frac{1}{|F_i|}\int_{F_i}(q_k{\circ}T_{F_i}^{-1})p\,\mathrm{d}s$ for all $p \in \mathbb{P}_{2,2}$. Let $\Sigma := \{\varpi_j\}_{j\in\{0:5\}}$. Is the triple $(K,\mathbb{P}_{2,2},\Sigma)$ a finite element? (*Hint*: consider the points $T_{F_i}(\xi_k)$, $i \in \{0{:}2\}$, $k \in \{0{:}1\}$, where ξ_0,ξ_1 are the two nodes of the Gauss–Legendre quadrature of order 3, then use Step (ii).)

Part III
Finite element interpolation

Part III, Chapter 8

Meshes

In Part III, composed of Chapters 8 to 17, we introduce the notion of meshes, show how to generate a finite element on each cell composing the mesh, and estimate the interpolation error in each mesh cell. We also derive important discrete inverse and functional inequalities in each mesh cell. Moreover, we discuss in some detail finite elements in $\boldsymbol{H}(\mathrm{div})$ and $\boldsymbol{H}(\mathrm{curl})$. In the present chapter, we study how to build a mesh of a bounded subset $D \subsetneq \mathbb{R}^d$, i.e., a finite collection of cells forming a partition of D. This is indeed the first important task to realize when one wants to approximate some PDEs posed in D. The viewpoint we adopt in this book is that each mesh cell is the image of a reference cell by some smooth diffeomorphism that we call geometric mapping. We show how to construct the geometric mapping and we present various important notions concerning meshes. We also discuss mesh-related data structures and mesh generators.

8.1 The geometric mapping

Let $\widehat{K}$ be a polyhedron in $\mathbb{R}^d$, called *reference cell*. We want to build a smooth diffeomorphism (i.e., an invertible mapping) $\boldsymbol{T}_K$ from $\widehat{K}$ to $K := \boldsymbol{T}_K(\widehat{K})$ using a set of *geometric nodes* $\{\boldsymbol{g}_i\}_{i \in \mathcal{N}_{\mathrm{geo}}}$ in K with $\mathcal{N}_{\mathrm{geo}} := \{1{:}n_{\mathrm{geo}}\}$ for some integer n_{geo}. In practice, these nodes are provided by a mesh generator. The key idea to build $\boldsymbol{T}_K$ is to use a Lagrange finite element in $\widehat{K}$, say $(\widehat{K}, \widehat{P}_{\mathrm{geo}}, \widehat{\Sigma}_{\mathrm{geo}})$, with reference Lagrange nodes $\{\widehat{\boldsymbol{g}}_i\}_{i \in \mathcal{N}_{\mathrm{geo}}}$ in $\widehat{K}$. This finite element is called *geometric finite element*. It is standard to assume that $\widehat{P}_{\mathrm{geo}}$ is a space of d-variate polynomials and that there is an integer $k_{\mathrm{geo}} \geq 1$ s.t.

$$\mathbb{P}_{k_{\mathrm{geo}},d} \subset \widehat{P}_{\mathrm{geo}} \subset C^\infty(\widehat{K}). \tag{8.1}$$

Notice that $n_{\mathrm{geo}} \geq d+1$ since $k_{\mathrm{geo}} \geq 1$. Let $\{\widehat{\psi}_i\}_{i \in \mathcal{N}_{\mathrm{geo}}}$ be the shape functions of the geometric finite element.

Definition 8.1 (Geometric mapping). *The* geometric mapping $\boldsymbol{T}_K$: $\widehat{K} \to K$ *is defined by*

$$\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := \sum_{i \in \mathcal{N}_{\text{geo}}} \widehat{\psi}_i(\widehat{\boldsymbol{x}}) \boldsymbol{g}_i, \qquad \forall \widehat{\boldsymbol{x}} \in \widehat{K}. \tag{8.2}$$

Since $\widehat{\psi}_i(\widehat{\boldsymbol{g}}_j) = \delta_{ij}$ for all $i, j \in \mathcal{N}_{\text{geo}}$, we have $\boldsymbol{T}_K(\widehat{\boldsymbol{g}}_j) = \boldsymbol{g}_j$. Notice that this construction implies that $\boldsymbol{T}_K$ is of class C^∞. We henceforth assume that $\boldsymbol{T}_K$ is a C^∞ diffeomorphism. Some care has to be taken when choosing the geometric nodes $\{\boldsymbol{g}_i\}_{i \in \mathcal{N}_{\text{geo}}}$ to ensure that $\boldsymbol{T}_K$ is indeed bijective when $\boldsymbol{T}_K$ is not affine. Some counterexamples are shown in Figures 8.1 and 8.2.
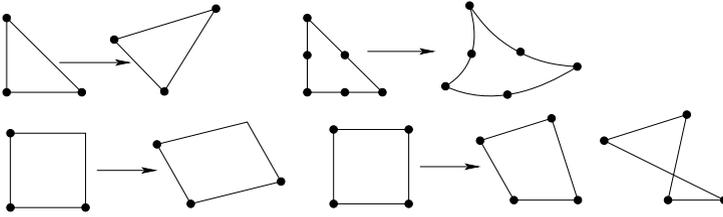
Fig. 8.1 $\mathbb{P}_1$-based generation of a triangle (top left), $\mathbb{P}_2$-based generation of a curved triangle (top right), $\mathbb{P}_1$-based generation of a parallelogram (bottom left), $\mathbb{Q}_1$-based generation of two quadrangles, the second one with a nonbijective mapping (bottom right).

We adopt the usual convention that consists of identifying vectors in $\mathbb{R}^d$ with column vectors. This allows us to identify $\boldsymbol{T}_K$ with the column vector with entries $(\boldsymbol{T}_K)_i$ for all $i \in \{1{:}d\}$ and the Jacobian of $\boldsymbol{T}_K$ with the matrix with entries

$$(\mathbb{J}_K)_{ij} := \partial_j (\boldsymbol{T}_K)_i, \qquad \forall i, j \in \{1{:}d\}, \tag{8.3}$$

where i is the row index and j the column index. The field $\mathbb{J}_K$ is $\mathbb{R}^{d \times d}$-valued and it is constant over $\widehat{K}$ if $\boldsymbol{T}_K$ is affine. Notice that the sign of $\det(\mathbb{J}_K)$ is necessarily constant over $\widehat{K}$ since we assumed that $\det(\mathbb{J}_K)(\widehat{\boldsymbol{x}}) \neq 0$ for all $\widehat{\boldsymbol{x}} \in \widehat{K}$ (this is indeed a necessary condition for $\boldsymbol{T}_K$ to be bijective). Contrary to what is done sometimes in the literature, we do not require that $\det(\mathbb{J}_K)$ has any particular sign.

Example 8.2 (Simplex generation). Let $\widehat{K}$ be the unit simplex in $\mathbb{R}^d$ with barycentric coordinates $\{\widehat{\lambda}_i\}_{i \in \{0:d\}}$ ($\widehat{\lambda}_0(\widehat{\boldsymbol{x}}) := 1 - \sum_{i \in \{1:d\}} \widehat{x}_i$ and $\widehat{\lambda}_i(\widehat{\boldsymbol{x}}) := \widehat{x}_i$ for all $i \in \{1{:}d\}$). Let K be a simplex in $\mathbb{R}^d$. Taking $\widehat{P}_{\text{geo}} := \mathbb{P}_{1,d}$ and the $n_{\text{geo}} := (d+1)$ vertices of K as geometric nodes, the geometric mapping $\boldsymbol{T}_K : \widehat{K} \to K$ is s.t. $\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := \sum_{i \in \{0:d\}} \widehat{\lambda}_i(\widehat{\boldsymbol{x}}) \boldsymbol{z}_i$ for all $\widehat{\boldsymbol{x}} \in \widehat{K}$. In dimension two, taking $\widehat{P}_{\text{geo}} := \mathbb{P}_{2,2}$, i.e., $n_{\text{geo}} := 6$, we can prescribe six geometric nodes in K and build a triangle with curved faces. See Figure 8.1 (top row) for illustrations. When using high-order elements, some care must

be taken to ensure that the geometric mapping $\boldsymbol{T}_K$ is indeed invertible. Figure 8.2 presents two examples where the mapping $\boldsymbol{T}_K$ is not invertible. For the one shown on the left, the enumerations chosen for the geometric nodes of $\widehat{K}$ and K_1 are not compatible. The example shown on the right is slightly more subtle since the singularity comes from the fact that the shape functions of the $\mathbb{P}_{2,2}$ Lagrange finite element can take negative values and that some geometric nodes of K_2 are too close. □
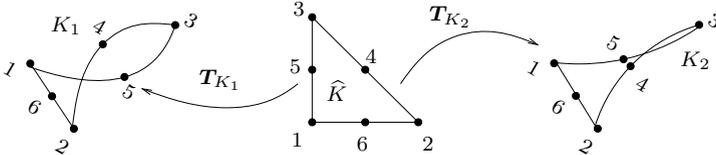
Fig. 8.2 Left: incompatible enumeration of the geometric nodes. Right: compatible enumeration, but some geometric nodes are too close.

Example 8.3 (Quadrangle generation). Let $\widehat{K} := (0,1)^2$ be the unit square in $\mathbb{R}^2$. Let us set $\widehat{\boldsymbol{z}}_0 := (0,0)$, $\widehat{\boldsymbol{z}}_1 := (1,0)$, $\widehat{\boldsymbol{z}}_2 := (0,1)$, and $\widehat{\boldsymbol{z}}_3 := (1,1)$. Taking $\widehat{P}_{\mathrm{geo}} := \mathbb{P}_{1,2}$, so that $n_{\mathrm{geo}} = 3$, we can prescribe three geometric nodes in K to build a smooth diffeomorphism. Let $\boldsymbol{z}_0$ be one vertex of K and let $\boldsymbol{z}_1, \boldsymbol{z}_2$ be the other two vertices of K sharing an edge with $\boldsymbol{z}_0$. Let $\boldsymbol{z}_3$ be the fourth vertex of K. Upon setting $\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := (1-\widehat{x}_1-\widehat{x}_2)\boldsymbol{z}_0 + \widehat{x}_1\boldsymbol{z}_1 + \widehat{x}_2\boldsymbol{z}_2$, we observe that K is a parallelogram. In particular, $\boldsymbol{z}_3 = \boldsymbol{T}_K(\widehat{\boldsymbol{z}}_3) = -\boldsymbol{z}_0 + \boldsymbol{z}_1 + \boldsymbol{z}_2$, i.e., $\boldsymbol{z}_0 + \boldsymbol{z}_3 = \boldsymbol{z}_1 + \boldsymbol{z}_2$. To generate a more general quadrangle, we can take $\widehat{P}_{\mathrm{geo}} := \mathbb{Q}_{1,2}$, so that $n_{\mathrm{geo}} = 4$, and use the four vertices of K as geometric nodes. In this case, $\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) = (1-\widehat{x}_1)(1-\widehat{x}_2)\boldsymbol{z}_0 + \widehat{x}_1(1-\widehat{x}_2)\boldsymbol{z}_1 + (1-\widehat{x}_1)\widehat{x}_2\boldsymbol{z}_2 + \widehat{x}_1\widehat{x}_2\boldsymbol{z}_3$. The mapping $\boldsymbol{T}_K$ is a smooth diffeomorphism whenever the nodes of K are properly enumerated. See the bottom row of Figure 8.1 for illustrations. In the rightmost example, $\boldsymbol{T}_K$ is not invertible because the nodes are not properly enumerated. □

8.2 Main definitions related to meshes

Definition 8.4 (Mesh). *Let D be a Lipschitz domain in $\mathbb{R}^d$. We say that $\mathcal{T}_h$ is a* mesh *of D if $\mathcal{T}_h$ is a finite collection of closed subsets of D, called* mesh cells *(or* mesh elements*), such that* (i) *the interiors of the mesh cells are all nonempty Lipschitz domains in $\mathbb{R}^d$ that are mutually disjoint and* (ii) *all the mesh cells cover $\overline{D}$ exactly, i.e.,*

$$\overline{D} = \bigcup_{K \in \mathcal{T}_h} K. \tag{8.4}$$

The subscript h refers to a level of refinement. It is common in the literature to set $h := \max_{K \in \mathcal{T}_h} h_K$ with $h_K := diam(K) := \max_{\boldsymbol{x}_1, \boldsymbol{x}_2 \in K} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\ell^2}$, where $\|\cdot\|_{\ell^2}$ is the Euclidean norm in $\mathbb{R}^d$, and to call h the meshsize.
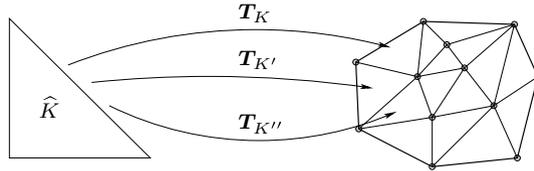
Fig. 8.3 Reference cell $\widehat{K}$ (left), mesh (right). The three arrows indicate the action of the geometric mapping for the three mesh cells K, K', K''.

The mesh cells have often a simple shape. For simplicity, we assume in this book that all the mesh cells have been generated from a fixed reference polyhedron $\widehat{K} \in \mathbb{R}^d$ (see §8.1) so that there is a smooth diffeomorphism $\boldsymbol{T}_K : \widehat{K} \to K$ for all $K \in \mathcal{T}_h$. Figure 8.3 presents an illustration using $\mathbb{P}_1$ geometric mappings to generate triangular cells. More generally, it is possible to consider a finite set of reference polyhedra to generate the mesh cells. One can for instance build meshes mixing triangles and quadrangles in dimension two, etc.

Remark 8.5 (Approximation of D). It happens sometimes that generating meshes that partition D exactly is too complicated, or that it is only possible to construct meshes of approximations of D. For instance, this situation arises when the boundary of D is curved; see §13.1 for examples. Unless specified otherwise, meshes are assumed to partition D exactly.   □

Definition 8.6 (Simplicial/affine mesh).   *The mesh $\mathcal{T}_h$ is said to be* simplicial *when the reference cell $\widehat{K}$ is a simplex, and the mesh $\mathcal{T}_h$ is said to be* affine *when all the geometric mappings $\{\boldsymbol{T}_K\}_{K \in \mathcal{T}_h}$ are affine.*

In this book, we often consider simplicial affine meshes, and we speak of triangulations when $d = 2$. An example is shown in Figure 8.4.
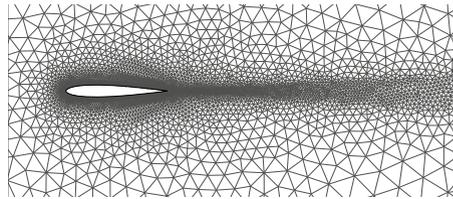
Fig. 8.4 Part of a triangulation around a two-dimensional NACA0012 airfoil profile.

Definition 8.7 (Faces, edges, and vertices of a cell). *Let $K \in \mathcal{T}_h$ be a cell. Assuming $d = 3$, the faces, edges, and vertices of K are defined to be the images by $\boldsymbol{T}_K$ of the faces, edges, and vertices of the reference polyhedron $\widehat{K}$, and these geometric entities are collected in the sets $\mathcal{F}_K$, $\mathcal{E}_K$, and $\mathcal{V}_K$, respectively. The same definition is valid in dimension $d = 2$ with the exception that the notions of edge and face coincide. The same definition is valid in dimension $d = 1$, with the exception that the notions of vertex, edge, and face coincide. We assume in the entire book that we have either $F \subset \partial D$ or $int(F) \subset D$ for all $K \in \mathcal{T}_h$ and all $F \in \mathcal{F}_K$.*

Remark 8.8 (Geometric nodes). The notion of geometric nodes introduced in §8.1 and the notion of vertices are different. In general, the vertices of a cell form a subset of its geometric nodes. These two sets coincide if the geometric element is a $\mathbb{P}_{1,d}$ or $\mathbb{Q}_{1,d}$ Lagrange element. □

Definition 8.9 (Mesh faces, edges, and vertices). *Let $\mathcal{T}_h$ be a mesh. Assume $d = 3$. We say that a closed two-dimensional manifold $F \subset \overline{D}$ is a mesh face if there is a mesh cell $K \in \mathcal{T}_h$ s.t. F is a face of K, i.e., $F \in \mathcal{F}_K$. Similarly, a closed one-dimensional manifold $E \subset \overline{D}$ is a mesh edge if there is a mesh cell $K \in \mathcal{T}_h$ s.t. $E \in \mathcal{E}_K$, and a point $\boldsymbol{z} \in \overline{D}$ is a mesh vertex if there is a mesh cell $K \in \mathcal{T}_h$ s.t. $\boldsymbol{z} \in \mathcal{V}_K$.*

Another important notion is that of interfaces and boundary faces.

Definition 8.10 (Interfaces, boundary faces). *A subset $F \subset \overline{D}$ is an interface if F has positive $(d-1)$-dimensional measure and there are two distinct mesh cells $K_l, K_r \in \mathcal{T}_h$ such that $F := \partial K_l \cap \partial K_r$ and F is a subset of a face of K_l and of a face of K_r. A subset $F \subset \overline{D}$ is a boundary face if F has positive $(d-1)$-dimensional measure and if there is a mesh cell $K_l \in \mathcal{T}_h$ such that $F := \partial K_l \cap \partial D$ and F is a face of K_l. All the interfaces are collected in the set $\mathcal{F}_h^\circ$, all the boundary faces are collected in the set $\mathcal{F}_h^\partial$, and we define*

$$\mathcal{F}_h := \mathcal{F}_h^\circ \cup \mathcal{F}_h^\partial. \tag{8.5}$$

The subscripts $\{l, r\}$ in the definition $F := \partial K_l \cap \partial K_r$ refer to the left cell and to the right cell. The notion of left and right cell will be unambiguously defined later by orienting all the interfaces. Distinguishing the left from the right cell will be important when defining jumps across interfaces (see Definition 18.2). In addition, we also have $F = K_l \cap K_r$ since the mesh cells have mutually disjoint interiors by assumption. Furthermore, we observe that a boundary face is always a mesh face, but an interface is not necessarily a mesh face since the notion of interface depends on the way adjacent mesh cells come into contact. An illustration is presented in Figure 8.5. For the mesh shown in the left panel, we have $\mathcal{F}_h = \bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K$. For that shown in the central panel, we have $\mathcal{F}_h \subset \bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K$ but $\bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K \not\subset \mathcal{F}_h$. For that shown in the right panel, we have $\mathcal{F}_h \not\subset \bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K$ and $\bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K \not\subset \mathcal{F}_h$.

The meshes shown in Figure 8.3, in Figure 8.4, and in the left panel of Figure 8.5 fall into the important class of matching meshes. Matching meshes
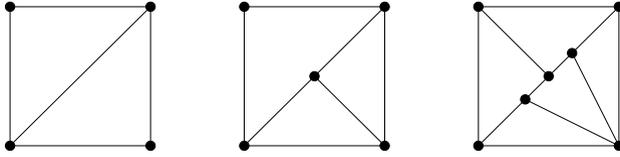
Fig. 8.5 Three examples of a triangulation of a square. Left panel: the mesh is composed of 2 cells and there is one interface. Central panel: the mesh is composed of 3 cells and there are 3 interfaces. Right panel: the mesh is composed of 5 cells and there are 7 interfaces. The three meshes contain 4 boundary faces.

play a central role in this book since they facilitate the construction of discrete spaces composed of piecewise smooth functions having an integrable gradient, curl or divergence (see Chapter 19 and onwards).

Definition 8.11 (Matching mesh). *A mesh $\mathcal{T}_h$ is said to be* matching *if for all cells $K, K' \in \mathcal{T}_h$ s.t. $K \cap K'$ is a manifold of dimension $(d-1)$, then $K \cap K'$ is an entire face of K and an entire face of K'.*

Proposition 8.12 (Mesh faces). *Let $\mathcal{T}_h$ be a matching mesh. Then,*

$$\mathcal{F}_h = \bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K. \tag{8.6}$$

Proof. Let $F \in \mathcal{F}_h$. If $F \in \mathcal{F}_h^\partial$, we infer from Definition 8.10 that $F \in \mathcal{F}_{K_l}$, whence $F \in \bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K$. If $F \in \mathcal{F}_h^\circ$, we have $F := \partial K_l \cap \partial K_r = K_l \cap K_r$, and we infer from Definition 8.11 that $F \in \mathcal{F}_{K_l} \cap \mathcal{F}_{K_r}$, whence $F \in \bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K$. We have thus shown that $\mathcal{F}_h \subset \bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K$. Conversely, let $K \in \mathcal{T}_h$ and $F \in \mathcal{F}_K$. If $F \subset \partial D$, we infer that $F \in \mathcal{F}_h^\partial$. Otherwise, our assumption on the faces of a mesh cell in Definition 8.7 implies that $\text{int}(F) \subset D$, and since the mesh cells form a partition of D, we infer that there is a mesh cell $K' \neq K$ s.t. $K \cap K' \subset F$ and $K \cap K'$ is a manifold of dimension $(d-1)$. Since the mesh is matching, $K \cap K'$ is a full face of both K and K' so that $F = K \cap K'$, which proves that $F \in \mathcal{F}_h^\circ$. We have thus shown that $\bigcup_{K \in \mathcal{T}_h} \mathcal{F}_K \subset \mathcal{F}_h$, and this completes the proof. $\qquad\square$

One can verify that Definition 8.11 implies that if $K \cap K' \neq \emptyset$ and $K \neq K'$, then the set $K \cap K'$ is a face, an edge (if $d = 3$), or a vertex that is common to K and K'. For matching meshes we denote the collection of the mesh edges (if $d = 3$) and the collection of the mesh vertices as follows:

$$\mathcal{E}_h := \bigcup_{K \in \mathcal{T}_h} \mathcal{E}_K, \qquad \mathcal{V}_h := \bigcup_{K \in \mathcal{T}_h} \mathcal{V}_K. \tag{8.7}$$

Remark 8.13 (Euler relations). Let $\mathcal{T}_h$ be a matching mesh of a polyhedron D in $\mathbb{R}^d$. If $d = 2$, let I be the degree of multiple-connectedness of D (i.e., the number of holes in D). Let N_c, N_e, N_v, N_e^∂, N_v^∂ be the number of mesh cells, edges, vertices, boundary edges, and boundary vertices, respectively. Then we have

$$N_{\mathrm{c}} - N_{\mathrm{e}} + N_{\mathrm{v}} = 1 - I, \qquad N_{\mathrm{v}}^{\partial} - N_{\mathrm{e}}^{\partial} = 0. \tag{8.8}$$

If $d = 3$, let additionally J be the number of connected components of the boundary of D, and let N_{f}, $N_{\mathrm{f}}^{\partial}$ be the number of mesh faces and boundary faces, respectively. Then we have

$$N_{\mathrm{c}} - N_{\mathrm{f}} + N_{\mathrm{e}} - N_{\mathrm{v}} = -1 + I - J, \qquad N_{\mathrm{f}}^{\partial} - N_{\mathrm{e}}^{\partial} + N_{\mathrm{v}}^{\partial} = 2(J - I). \qquad \square$$

8.3 Data structure

A mesh is a data structure produced by a mesh generator. This data structure consists of a cloud of points, called *geometric nodes*, that are numbered and connected. There are many ways to construct this data structure. Let us give an example. We start by enumerating the geometric nodes $\{\boldsymbol{g}_1, \dots, \boldsymbol{g}_{N_{\mathrm{geo}}}\}$ where N_{geo} is the number of geometric nodes. This enumeration is said to be global. The geometric nodes are defined by their coordinates in $\mathbb{R}^d$. These quantities are stored in a two-dimensional array of size $d \times N_{\mathrm{geo}}$, which we denote by

$$\mathtt{coord}(1{:}d, 1{:}N_{\mathrm{geo}}), \tag{8.9}$$

and we say that $\mathtt{coord}$ is the *coordinate array* of the mesh. For all $k \in \{1{:}d\}$ and all $n \in \{1{:}N_{\mathrm{geo}}\}$, $\mathtt{coord}(k, n)$ is the k-th coordinate of $\boldsymbol{g}_n$.

The geometric nodes are organized into mesh cells by means of a *connectivity array*, in such a way that every mesh cell is assigned n_{geo} geometric nodes. Let us enumerate the mesh cells as $\{K_1, \dots, K_{N_{\mathrm{c}}}\}$ where N_{c} is the number of mesh cells. The geometric nodes associated with any mesh cell can be recovered from a two-dimensional array of size $N_{\mathrm{c}} \times n_{\mathrm{geo}}$, which we denote by

$$\mathtt{j\_geo}(1{:}N_{\mathrm{c}}, 1{:}n_{\mathrm{geo}}). \tag{8.10}$$

For all $m \in \{1{:}N_{\mathrm{c}}\}$ and all $n \in \mathcal{N}_{\mathrm{geo}}$ (recall that $\mathcal{N}_{\mathrm{geo}} := \{1{:}n_{\mathrm{geo}}\}$), the integer $\mathtt{j\_geo}(m, n)$ is the global index of the n-th node in the m-th cell. The second index in the array $\mathtt{j\_geo}$ provides the *local enumeration* of the geometric nodes for each mesh cell. Using the connectivity array and the coordinate array, it is possible to rewrite the geometric mapping $\boldsymbol{T}_K$ from Definition 8.1 as follows:

$$(\boldsymbol{T}_{K_m}(\widehat{\boldsymbol{x}}))_i = \sum_{n \in \mathcal{N}_{\mathrm{geo}}} \widehat{\psi}_n(\widehat{\boldsymbol{x}}) \, \mathtt{coord}(i, \mathtt{j\_geo}(m, n)), \tag{8.11}$$

for all $\widehat{\boldsymbol{x}} \in \widehat{K}$, all $m \in \{1{:}N_{\mathrm{c}}\}$, and all $i \in \{1{:}d\}$.

Example 8.14 (Enumeration in a simplex). Figure 8.6 shows an example of local and global enumerations. Here, the geometric reference element is the two-dimensional $\mathbb{P}_1$ Lagrange element, i.e., $n_{\mathrm{geo}} = 3$. We consider three

mesh cells with global indices 56, 213, and 315. The values of the connec-
tivity array are $\text{j\_geo}(315,1) = 13$, $\text{j\_geo}(315,2) = 37$, $\text{j\_geo}(315,3) = 250$,
$\text{j\_geo}(56,1) = 13$, $\text{j\_geo}(56,2) = 37$, $\text{j\_geo}(56,3) = 53$, etc. We have adopted
the convention that for any m, the value of $\text{j\_geo}(m,n)$ increases with n. This
choice will be instrumental in Chapter 10 when orienting the mesh. Note that
the sign of $\det(\mathbb{J}_K)$ is different in the cells 315 and 56.   □
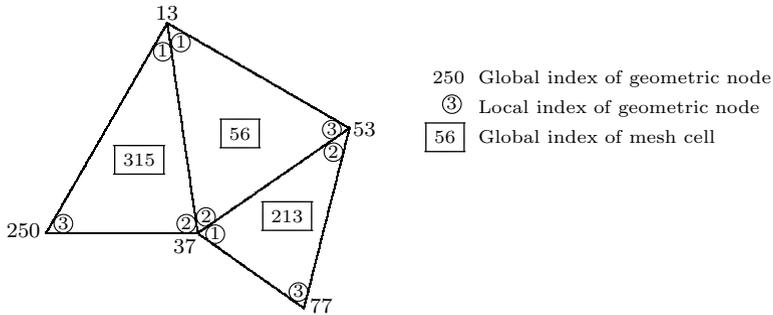
Fig. 8.6 Example of local and global enumerations of geometric nodes for three triangular
mesh cells.

In many situations, it is useful to have two-dimensional arrays providing
the global indices of the faces, edges, and vertices of any mesh cell. The reason
is that finite element matrices are assembled by means of a loop over the mesh
cells (see §29.2.3), and that these arrays are instrumental to identify degrees
of freedom attached to the mesh faces, edges, and vertices. Let us focus on
matching meshes and let us enumerate the mesh faces, edges, and vertices in
$\mathcal{F}_h$, $\mathcal{E}_h$, and $\mathcal{V}_h$ from 1 to N_f, N_e, and N_v, respectively, i.e.,

$$\mathcal{F}_h = \{F_j\}_{j \in \{1:N_\mathrm{f}\}}, \qquad \mathcal{E}_h = \{E_j\}_{j \in \{1:N_\mathrm{e}\}}, \qquad \mathcal{V}_h = \{z_j\}_{j \in \{1:N_\mathrm{v}\}}.$$

Let n_cf, n_ce, and n_cv be, respectively, the number of faces, edges, and vertices
of a mesh cell. For instance, $n_\mathrm{cf} = 4$, $n_\mathrm{ce} = 6$, and $n_\mathrm{cv} = 4$ for a tetrahedron.
We introduce the following two-dimensional arrays:

$$\text{j\_cf}(1{:}N_\mathrm{c}, 1{:}n_\mathrm{cf}), \qquad \text{j\_ce}(1{:}N_\mathrm{c}, 1{:}n_\mathrm{ce}), \qquad \text{j\_cv}(1{:}N_\mathrm{c}, 1{:}n_\mathrm{cv}). \qquad (8.12)$$

For all $m \in \{1{:}N_\mathrm{c}\}$ and all $n \in \{1{:}n_\mathrm{cf}\}$, the integer $\text{j\_cf}(m,n)$ is the global
index of the n-th face in the m-th cell, and similarly for j_ce and j_cv. In
other words we have

$$\boldsymbol{T}_{K_m}(\widehat{F}_n) = F_{\text{j\_cf}(m,n)}, \quad \boldsymbol{T}_{K_m}(\widehat{E}_n) = E_{\text{j\_ce}(m,n)}, \quad \boldsymbol{T}_{K_m}(\widehat{z}_n) = z_{\text{j\_cv}(m,n)}.$$

Notice that the arrays j_cv and j_geo are different in general, just like the
vertices and the geometric nodes may be different objects.

Remark 8.15 (Alternative data structure). Another choice is to consider the two-dimensional arrays $\mathtt{j\_cf}(1{:}N_c, 1{:}n_{cf})$ (as above) together with the two-dimensional arrays $\mathtt{j\_fe}(1{:}N_f, 1{:}n_{fe})$ (providing the global indices of the edges of a given mesh face, where n_{fe} is the number of edges of a face, assuming that this number is face-independent), and $\mathtt{j\_ev}(1{:}N_e, 1{:}2)$ (providing the global indices of the two vertices of a mesh edge). The information stored in the array $\mathtt{j\_ce}$ (resp., $\mathtt{j\_cv}$) can then be recovered from the arrays $\mathtt{j\_cf}$ and $\mathtt{j\_fe}$ (resp., $\mathtt{j\_cf}$, $\mathtt{j\_fe}$, and $\mathtt{j\_ev}$). The reader must be aware that all these compositions involve memory accesses that may be time consuming.

□

8.4 Mesh generation

Mesh generation is a basic ingredient of finite element methods. Generating a mesh is often a time-consuming task, especially for complex three-dimensional configurations. Mesh generators involve two types of tasks: (1) representing geometrically the boundary of the domain by using suitable mappings parameterizing paths or surfaces; (2) meshing the lines, surfaces, and volumes that have been identified in the first task. This section briefly describes how to organize the above two tasks. The material is meant to provide some basic understanding of the process.

8.4.1 Two-dimensional case

Let us consider a two-dimensional domain D and let us think about how D can be geometrically represented.

1. D is entirely defined by its one-dimensional boundary, ∂D.
2. The boundary ∂D can be decomposed into its connected components.
3. Each connected component can be partitioned into a union of paths.
4. Each path can be assigned two extremities (possibly by cutting the paths that are closed). These points are referred to as the vertices of ∂D.
5. Each path can be mapped to the interval $[0, 1]$.

As an illustration, consider the domain shown in Figure 8.7. Its boundary is composed of two connected components. The external component is the union of the three paths PQ, QR, and RP. The internal boundary is transformed into a path that is homeomorphic to a segment by cutting it at S. In conclusion, the boundary of D is decomposed into the union of four paths: $\partial D_1 := PQ$, $\partial D_2 := QR$, $\partial D_3 := RP$, and $\partial D_4 := SS$.

A general algorithm for a two-dimensional mesh generator is obtained by reading in reverse order the above list:
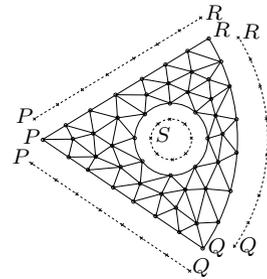
Fig. 8.7 Meshing a two-dimensional
domain.

1. Locate the vertices of ∂D and partition $\partial D = \bigcup_{n \in \{1 : N_{\mathrm{p}}^{\partial}\}} \partial D_n$ so that
   each elementary path ∂D_n is limited by two vertices (possibly identical).
   Here, $N_{\mathrm{p}}^{\partial}$ denotes the total number of elementary paths.
2. Connect the two vertices of ∂D_n for all $n \in \{1 : N_{\mathrm{p}}^{\partial}\}$ by a parameterized
   path $\gamma_n : [0, 1] \to \partial D_n$.
3. Letting $\bigcup_{i \in \{1 : I_n\}} [x_{n,i-1}, x_{n,i}]$ be a partition of $[0, 1]$ into I_n small seg-
   ments, the *boundary mesh* on ∂D is $\bigcup_{n \in \{1 : N_{\mathrm{p}}^{\partial}\}} \bigcup_{i \in \{1 : I_n\}} \gamma_n([x_{n,i-1}, x_{n,i}])$.
4. Finally, mesh the interior of D by extending the boundary mesh. This
   last step usually involves an advancing front method where mesh vertices
   are progressively inserted inside the domain and connected to the other
   vertices to form new triangles (Fig. 8.8); see, e.g., Rebay [166] and the
   references therein.
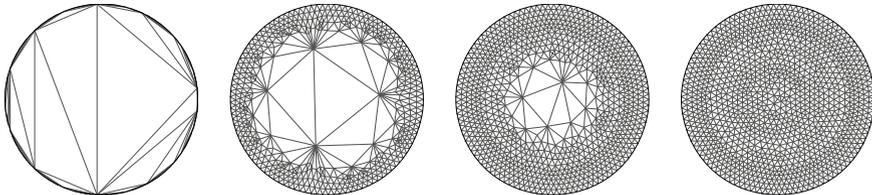
Fig. 8.8 Triangulation of a circle by an advancing front method. Various stages of the
mesh generation process are illustrated.

8.4.2 Three-dimensional case

The above algorithm extends to dimension three. As in dimension two, the
algorithm is deduced from the geometric description of three-dimensional
domains. Let D be a three-dimensional domain.

1. D is entirely defined by its two-dimensional boundary, ∂D.
2. The boundary ∂D can be decomposed into its connected components.
3. Each connected component can be decomposed into a union of orientable
   surfaces with edges, say $\partial D = \bigcup_{n \in \{1 : N_{\mathrm{s}}^{\partial}\}} \partial D_n$ ($N_{\mathrm{s}}^{\partial}$ is the total number

of these surfaces). For instance, a sphere can be decomposed into two hemispheres. The orientation of the connected components of ∂D says on which side of ∂D the interior of D is.

4. Each orientable surface ∂D_n can be mapped to a two-dimensional domain $\partial D_n^{2D} \subset \mathbb{R}^2$ by a mapping $\gamma_n : \partial D_n^{2D} \to \partial D_n$.
5. Each two-dimensional domain ∂D_n^{2D} for all $n \in \{1{:}N_s^{\partial}\}$ can be described by means of the algorithm from §8.4.1.

An illustration is presented in Figure 8.9. The domain is a cone. Since the boundary of the cone is connected but has no edges, it is decomposed into two simpler surfaces by separating the base and the lateral surface. The base is homeomorphic to a disk, ∂D_1^{2D}. The lateral surface is further transformed by cutting it along the segment PQ. The surface thus created is homeomorphic to a triangle, ∂D_2^{2D}. When meshing the two sides of the triangle associated with the segment PQ, one must make sure that the two one-dimensional meshes coincide.
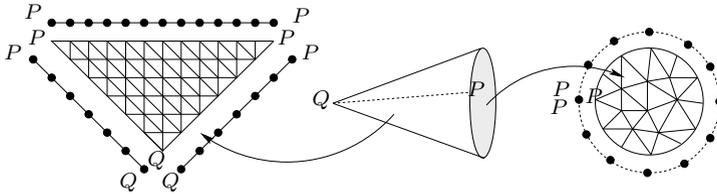
Fig. 8.9 Geometric representation of a three-dimensional domain.

An algorithm to mesh a three-dimensional domain is obtained by reading the above list from bottom to top:

1. Construct a mesh $\mathcal{T}_{h,n}^{2D}$ of each two-dimensional domain ∂D_n^{2D} for all $n \in \{1{:}N_s^{\partial}\}$ by applying the algorithm from §8.4.1.
2. A mesh for ∂D_n is defined to be $\mathcal{T}_{h,n}^{\partial} := \gamma_n(\mathcal{T}_{h,n}^{2D})$ for all $n \in \{1{:}N_s^{\partial}\}$.
3. The union $\bigcup_{n \in \{1{:}N_s^{\partial}\}} \mathcal{T}_{h,n}^{\partial}$ is the boundary mesh.
4. Finally, mesh the interior of D by extending the boundary mesh.

Remark 8.16 (Extruded meshes). Some applications use either cylinders or domains that are homeomorphic to cylinders. A possible strategy to mesh the interior of domains of this type consists of meshing first its right section, which can have any two-dimensional shape, then extruding the mesh of the right section along the generatrix. Depending on the elements chosen to mesh the right section, the volume mesh is typically composed of prisms of triangular or quadrangular base. These prisms can be further decomposed into tetrahedra if needed. □

Exercises

Exercise 8.1 (Curved triangle). Consider the $\mathbb{P}_2$ transformation of a triangle shown in the upper right panel of Figure 8.1. Consider a geometric node of K that is the image of the midpoint of an edge of $\widehat{K}$. Show that the tangent vector to the curved boundary at this node is collinear to the vector formed by the two vertices of the corresponding curved edge. (*Hint*: use the properties of the Lagrange $\mathbb{P}_2$ shape functions.)

Exercise 8.2 (Euler relations). Let $\mathcal{T}_h$ be a matching mesh in $\mathbb{R}^2$ composed of polygons all having ν vertices. (i) Show that $2N_e - N_e^{\partial} = \nu N_c$. (ii) Combine this result with the Euler relations to show that $N_c \sim \frac{2}{\nu-2} N_v$ and $N_e \sim \frac{\nu}{\nu-2} N_v$ for fine enough meshes where $N_v^{\partial} = N_e^{\partial} \ll \min(N_v, N_e, N_c)$.

Exercise 8.3 (Connectivity arrays j_cv, j_ce). Write admissible connectivity arrays j_cv and j_ce for the following mesh where the face enumeration is identified with large circles and the cell enumeration with squares.
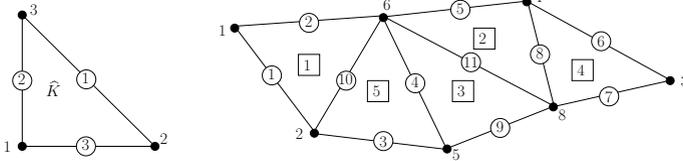
Exercise 8.4 (Connectivity array j_geo). Define a connectivity array j_geo for the following mesh such that the determinant of the Jacobian matrix of $\boldsymbol{T}_K$ is positive for all the cells.
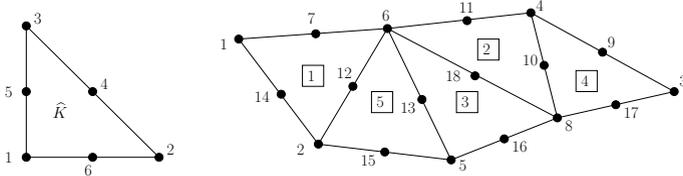
Exercise 8.5 (Geometric mapping). Let $\boldsymbol{z}_1 := (0,0)$, $\boldsymbol{z}_2 := (1,0)$, $\boldsymbol{z}_3 := (0,1)$, $\boldsymbol{z}_4 := (\frac{1}{3}, \frac{1}{3})$. Consider the triangles $K_1 := \operatorname{conv}(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_4)$, $K_2 := \operatorname{conv}(\boldsymbol{z}_2, \boldsymbol{z}_3, \boldsymbol{z}_4)$, and $K_3 := \operatorname{conv}(\boldsymbol{z}_3, \boldsymbol{z}_1, \boldsymbol{z}_4)$. (i) Construct the affine geometric mappings $\boldsymbol{T}_{K_2} : K_1 \to K_2$ and $\boldsymbol{T}_{K_3} : K_1 \to K_3$ s.t. $\boldsymbol{T}_{K_2}(\boldsymbol{z}_1) = \boldsymbol{z}_2$, $\boldsymbol{T}_{K_2}(\boldsymbol{z}_4) = \boldsymbol{z}_4$, and $\boldsymbol{T}_{K_3}(\boldsymbol{z}_1) = \boldsymbol{z}_3$, $\boldsymbol{T}_{K_3}(\boldsymbol{z}_4) = \boldsymbol{z}_4$. (*Hint*: $\boldsymbol{T}_{K_2}$ is of the form $\boldsymbol{T}_{K_2}(\boldsymbol{x}) = \boldsymbol{z}_2 + \mathbb{J}_{K_2}(\boldsymbol{x} - \boldsymbol{z}_1)$.) (ii) Compute $\det(\mathbb{J}_{K_2})\mathbb{J}_{K_2}^{-1}$ and $\det(\mathbb{J}_{K_3})\mathbb{J}_{K_3}^{-1}$. *Note*: the transformation $\boldsymbol{v} \mapsto \det(\mathbb{J}_K)\mathbb{J}_K^{-1}\boldsymbol{v} \circ \boldsymbol{T}_K$ is called contravariant Piola transformation; see (9.9c).

Part III, Chapter 9

Finite element generation

In the previous chapter, we have seen how to generate a mesh from a reference cell and a collection of geometric mappings. We now show how to generate a finite element in each mesh cell from a reference finite element. To this purpose, we need one new concept in addition to the geometric mapping: a functional transformation that maps functions defined on the current mesh cell to functions defined on the reference cell. Key examples of such transformations are the Piola transformations. These transformations arise naturally in the chain rule when one investigates how the standard differential operators (gradient, curl, divergence) are transformed by the geometric mapping. The construction presented in this chapter provides the cornerstone for the analysis of the finite element interpolation error to be performed in Chapter 11. Recall that $\|\cdot\|_{\ell^2}$ is the Euclidean norm in $\mathbb{R}^d$ and $\boldsymbol{a}\cdot\boldsymbol{b}$ denotes the corresponding inner product.

9.1 Main ideas

Let $\mathcal{T}_h$ be a mesh generated as described in Chapter 8. This means that we have at hand a reference cell $\widehat{K}$ (recall that $\widehat{K}$ is a polyhedron) and a geometric mapping $\boldsymbol{T}_K : \widehat{K} \to K$ for every mesh cell $K \in \mathcal{T}_h$. Given an integer $q \geq 1$, our goal is now to define a finite element in K composed of $\mathbb{R}^q$-valued functions defined on K. To this purpose, we assume that we have at hand a fixed finite element $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$, where $\widehat{P}$ is composed of $\mathbb{R}^q$-valued functions defined on $\widehat{K}$, and $\widehat{\Sigma}$ is the collection of the degrees of freedom (dofs) for these functions.

The triple $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ should not be confused with the geometric finite element $(\widehat{K}, \widehat{P}_{\mathrm{geo}}, \widehat{\Sigma}_{\mathrm{geo}})$ whose only use is to define K, whereas $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ is used to interpolate $\mathbb{R}^q$-valued functions. The interpolation is said to be *isopara-*

metric whenever $[\widehat{P}_{\text{geo}}]^q = \widehat{P}$ and *subparametric* whenever $[\widehat{P}_{\text{geo}}]^q \subsetneq \widehat{P}$. The most common example of subparametric interpolation consists of using affine geometric mappings together with shape functions that are quadratic or of higher polynomial order.

Definition 9.1 (Reference element). $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ *is called* reference finite element, *and with obvious notation* $\{\widehat{\sigma}_i\}_{i \in \mathcal{N}}$ *and* $\{\widehat{\theta}_i\}_{i \in \mathcal{N}}$ *are called* reference dofs *and* reference shape functions, *respectively.*

Recalling Definition 5.7, we also assume that we have at hand a Banach space $V(\widehat{K}) \subset L^1(\widehat{K}; \mathbb{R}^q)$ such that $\widehat{P} \subset V(\widehat{K})$ and such that the linear forms $\{\widehat{\sigma}_i\}_{i \in \mathcal{N}}$ can be extended to $\mathcal{L}(V(\widehat{K}); \mathbb{R})$ (we use the same symbol $\widehat{\sigma}_i$ for simplicity). The interpolation operator $\mathcal{I}_{\widehat{K}} : V(\widehat{K}) \to \widehat{P}$ associated with $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ is defined as follows (see (5.7)):

$$\mathcal{I}_{\widehat{K}}(\widehat{v})(\widehat{\boldsymbol{x}}) := \sum_{i \in \mathcal{N}} \widehat{\sigma}_i(\widehat{v})\widehat{\theta}_i(\widehat{\boldsymbol{x}}), \qquad \forall \widehat{\boldsymbol{x}} \in \widehat{K}. \tag{9.1}$$

The operator $\mathcal{I}_{\widehat{K}}$ is called *reference interpolation operator.*

Since our goal is to generate a finite element on K and to build an interpolation operator $\mathcal{I}_K$ acting on functions defined on K, we introduce a counterpart of the space $V(\widehat{K})$ for those functions, say $V(K)$. The new ingredient we need for the construction is a transformation

$$\psi_K : V(K) \to V(\widehat{K}), \tag{9.2}$$

which we assume to be a bounded linear isomorphism. A simple definition of ψ_K is the pullback by the geometric mapping, i.e.,

$$\psi_K(v) := v \circ \boldsymbol{T}_K, \qquad \forall v \in V(K). \tag{9.3}$$

We will see that this definition is well-suited to nodal and modal finite elements. However, we will also see that this definition is not adequate when considering vector-valued functions for which the tangential or the normal component at the boundary of K plays specific roles. This is the reason why we use a general notation for the functional transformation ψ_K.

Proposition 9.2 (Finite element generation). *Let* $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ *be the reference element with extended dofs* $\{\widehat{\sigma}_i\}_{i \in \mathcal{N}} \subset \mathcal{L}(V(\widehat{K}); \mathbb{R})$. *Let* $K \in \mathcal{T}_h$ *be a mesh cell. Assume that we have at hand a Banach space* $V(K)$ *and a bounded linear isomorphism* $\psi_K \in \mathcal{L}(V(K); V(\widehat{K}))$. *Then the triple* (K, P_K, Σ_K) *s.t.*

$$P_K := \psi_K^{-1}(\widehat{P}) = \{p := \psi_K^{-1}(\widehat{p}) \mid \widehat{p} \in \widehat{P}\}, \tag{9.4a}$$

$$\Sigma_K := \widehat{\Sigma} \circ \psi_K = \{\sigma_{K,i} := \widehat{\sigma}_{i|\widehat{P}} \circ \psi_K\}_{i \in \mathcal{N}} \subset \mathcal{L}(P_K; \mathbb{R}), \tag{9.4b}$$

is a finite element. The dofs in Σ_K *can be extended to* $\mathcal{L}(V(K); \mathbb{R})$ *by setting* $\sigma_{K,i} := \widehat{\sigma}_i \circ \psi_K$ *for all* $i \in \mathcal{N}$.

Proof. We apply Remark 5.3 to prove that (K, P_K, Σ_K) is a finite element. Since ψ_K is bijective, we have $\dim(P) = \dim(\widehat{P}) = n_{\text{sh}}$. Let $p \in P_K$ be s.t. $\sigma_{K,i}(p) = 0$ for all $i \in \mathcal{N}$. Then $\widehat{\sigma}_i(\psi_K(p)) = 0$ for all $i \in \mathcal{N}$, so that $\psi_K(p) = 0$ by the unisolvence property of $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$. This implies that $p = 0$ since ψ_K is an isomorphism. Finally, since $(\widehat{\sigma}_i \circ \psi_K)_{|P} = \widehat{\sigma}_{i|\widehat{P}} \circ \psi_K$, the linear map $\widehat{\sigma}_i \circ \psi_K : V(K) \to \mathbb{R}$ is an extension of $\sigma_{K,i} : P_K \to \mathbb{R}$ to $V(K)$ (we use the same notation for simplicity), and we have $\sigma_{K,i} \in \mathcal{L}(V(K); \mathbb{R})$ since $|\sigma_{K,i}(v)| \le \|\widehat{\sigma}_i\|_{\mathcal{L}(V(\widehat{K}); \mathbb{R})} \|\psi_K\|_{\mathcal{L}(V(K); V(\widehat{K}))} \|v\|_{V(K)}$ for all $v \in V(K)$. $\square$

The linear forms $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$ are called *local dofs*. The following functions, called *local shape functions*:

$$\theta_{K,i} := \psi_K^{-1}(\widehat{\theta}_i), \qquad \forall i \in \mathcal{N}, \tag{9.5}$$

satisfy $\sigma_{K,i}(\theta_{K,j}) = \widehat{\sigma}_i(\psi_K(\theta_{K,j})) = \widehat{\sigma}_i(\widehat{\theta}_j) = \delta_{ij}$ for all $i, j \in \mathcal{N}$. The *local interpolation operator* $\mathcal{I}_K : V(K) \to P_K$ acts as follows:

$$\mathcal{I}_K(v)(\boldsymbol{x}) := \sum_{i \in \mathcal{N}} \sigma_{K,i}(v) \theta_{K,i}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K. \tag{9.6}$$

The following result plays a key role in the analysis of the interpolation error.

Proposition 9.3 (Commuting diagram). *We have $\mathcal{I}_K = \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}} \circ \psi_K$, i.e., the following diagram commutes:*

$$
\begin{array}{ccc}
V(K) & \xrightarrow{\ \psi_K\ } & V(\widehat{K}) \\
\Big\downarrow{\scriptstyle \mathcal{I}_K} & & \Big\downarrow{\scriptstyle \mathcal{I}_{\widehat{K}}} \\
P_K & \xrightarrow[\ \psi_K\ ]{} & \widehat{P}
\end{array}
$$

i.e., P_K is pointwise invariant under $\mathcal{I}_K$, that is, $\mathcal{I}_K(p) = p$ for all $p \in P_K$.

Proof. Let v in $V(K)$. The definition (9.4) of (K, P_K, Σ_K) implies that

$$\mathcal{I}_{\widehat{K}}(\psi_K(v)) = \sum_{i \in \mathcal{N}} \widehat{\sigma}_i(\psi_K(v)) \, \widehat{\theta}_i = \sum_{i \in \mathcal{N}} \sigma_{K,i}(v) \, \psi_K(\theta_{K,i}) = \psi_K(\mathcal{I}_K(v)),$$

owing to the linearity of ψ_K. Hence, the above diagram commutes. Let now $p \in P_K$. We have $\mathcal{I}_K(p) = \psi_K^{-1}(\mathcal{I}_{\widehat{K}}(\psi_K(p))) = \psi_K^{-1}(\psi_K(p))$ since $\psi_K(p) \in \widehat{P}$ and $\widehat{P}$ is pointwise invariant under $\mathcal{I}_{\widehat{K}}$. Hence, $\mathcal{I}_K(p) = p$. $\square$

Example 9.4 (Lagrange elements). Let $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ be a Lagrange finite element with nodes $\{\widehat{\boldsymbol{a}}_i\}_{i \in \mathcal{N}}$ and $V(\widehat{K}) := C^0(\widehat{K})$; see §5.4.1. Set $V(K) := C^0(K)$. The map $\psi_K : V(K) \to V(\widehat{K})$ defined in (9.3) is an isomorphism in $\mathcal{L}(V(K); V(\widehat{K}))$. The finite element (K, P_K, Σ_K) constructed in Proposition 9.2 using ψ_K is also a Lagrange finite element. Indeed, we have $\sigma_{K,i}(p) := \widehat{\sigma}_i(\psi_K(p)) := \psi_K(p)(\widehat{\boldsymbol{a}}_i) = (p \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{a}}_i)$ for all $p \in P_K$. Setting

$$\boldsymbol{a}_{K,i} := \boldsymbol{T}_K(\widehat{\boldsymbol{a}}_i), \qquad \forall i \in \mathcal{N},$$

we infer that $\{\boldsymbol{a}_{K,i}\}_{i\in\mathcal{N}}$ are the Lagrange nodes of (K, P_K, Σ_K). The Lagrange interpolation operator $\mathcal{I}_K^{\mathrm{L}}$ acts as follows:

$$\mathcal{I}_K^{\mathrm{L}}(v)(\boldsymbol{x}) := \sum_{i\in\mathcal{N}} v(\boldsymbol{a}_{K,i})\theta_{K,i}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K. \tag{9.7}$$

Note that even if $\widehat{P}$ is a polynomial space, $P_K := \{\widehat{p} \circ \boldsymbol{T}_K^{-1}, \ \widehat{p} \in \widehat{P}\}$ is not necessarily a polynomial space unless $\boldsymbol{T}_K$ is affine. □

Example 9.5 (Modal elements). Let $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ be a modal finite element with dofs $\widehat{\sigma}_i(\widehat{p}) := \frac{1}{|\widehat{K}|}\int_{\widehat{K}} \widehat{\zeta}_i\widehat{p}\,\mathrm{d}\widehat{x}$ for all $\widehat{p} \in \widehat{P}$ and all $i \in \mathcal{N}$, where $\{\widehat{\zeta}_i\}_{i\in\mathcal{N}}$ is a basis of $\widehat{P}$, and let $V(\widehat{K}) := L^1(\widehat{K})$; see §5.4.2 Set $V(K) := L^1(K)$. The map $\psi_K : V(K) \to V(\widehat{K})$ defined in (9.3) is an isomorphism in $\mathcal{L}(V(K); V(\widehat{K}))$. The finite element (K, P_K, Σ_K) constructed in Proposition 9.2 using ψ_K is also a modal finite element. Indeed, we have for all $p \in P_K$,

$$\sigma_{K,i}(p) := \widehat{\sigma}_i(\psi_K(p)) := \frac{1}{|\widehat{K}|} \int_{\widehat{K}} \widehat{\zeta}_i(\widehat{\boldsymbol{x}})(p \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{x}$$

$$= \frac{1}{|\widehat{K}|} \int_{\widehat{K}} \frac{1}{\alpha_K}(\zeta_{K,i} \circ \boldsymbol{T}_K)(p \circ \boldsymbol{T}_K)\,\mathrm{d}\widehat{x} = \frac{1}{|K|} \int_K \zeta_{K,i}p\,\mathrm{d}x,$$

with $\zeta_{K,i} := \alpha_K\widehat{\zeta}_i \circ \boldsymbol{T}_K^{-1}$, $\alpha_K := |\det(\mathbb{J}_K)|^{-1}\frac{|K|}{|\widehat{K}|}$, and $\mathbb{J}_K$ is the Jacobian matrix of $\boldsymbol{T}_K$ defined in (8.3) ($\alpha_K = 1$ if $\boldsymbol{T}_K$ is affine). □

9.2 Differential calculus and geometry

In this section, we present basic identities from differential calculus and geometry showing how the usual differential operators (gradient, curl, and divergence) and normal and tangent vectors are transformed by the geometric mapping. We refer the reader to (4.6) for the definition of the divergence operator and to (4.7) for the definition of the curl operator with $d = 3$ (the material can be adapted to the case $d = 2$ by proceeding as in Remark 4.18).

9.2.1 Transformation of differential operators

Let $\widehat{K}$ be the reference polyhedron in $\mathbb{R}^d$ and let $K \in \mathcal{T}_h$ be a mesh cell. Let $\boldsymbol{T}_K : \widehat{K} \to K$ be the geometric mapping and let $\mathbb{J}_K$ be the Jacobian matrix of $\boldsymbol{T}_K$ (see (8.3)). Recall that we use boldface notation for $\mathbb{R}^d$-valued functions and for functional spaces composed of $\mathbb{R}^d$-valued functions. For instance, we write $\boldsymbol{C}^l(K) := C^l(K; \mathbb{R}^d)$ for all $l \in \mathbb{N}$. The following result is of fundamental importance.

Lemma 9.6 (Differential operators). *Let $v \in C^1(K)$ and $\boldsymbol{v} \in \boldsymbol{C}^1(K)$. The following holds true for all $\widehat{\boldsymbol{x}} \in \widehat{K}$:*

$$\nabla(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}}) = \mathbb{J}_K(\widehat{\boldsymbol{x}})^\mathsf{T}(\nabla v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}})), \tag{9.8a}$$

$$\nabla\times(\mathbb{J}_K^\mathsf{T}(\boldsymbol{v} \circ \boldsymbol{T}_K))(\widehat{\boldsymbol{x}}) = \det(\mathbb{J}_K(\widehat{\boldsymbol{x}}))\mathbb{J}_K^{-1}(\widehat{\boldsymbol{x}})(\nabla\times\boldsymbol{v})(\boldsymbol{T}_K(\widehat{\boldsymbol{x}})), \tag{9.8b}$$

$$\nabla\cdot(\det(\mathbb{J}_K)\mathbb{J}_K^{-1}(\boldsymbol{v} \circ \boldsymbol{T}_K))(\widehat{\boldsymbol{x}}) = \det(\mathbb{J}_K(\widehat{\boldsymbol{x}}))(\nabla\cdot\boldsymbol{v})(\boldsymbol{T}_K(\widehat{\boldsymbol{x}})). \tag{9.8c}$$

Proof. (1) Proof of (9.8a). Since the link between the Jacobian matrix of $\boldsymbol{T}_K$ and its Fréchet derivative (see Definition B.1) is that $D\boldsymbol{T}_K(\widehat{\boldsymbol{x}})(\boldsymbol{h}) = \mathbb{J}_K(\widehat{\boldsymbol{x}})\boldsymbol{h}$ for all $\boldsymbol{h} \in \mathbb{R}^d$, we can use Lemma B.4 (chain rule) with $n := 1$ to infer that

$$D(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})(\boldsymbol{h}) = Dv(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))(D\boldsymbol{T}_K(\boldsymbol{x})(\boldsymbol{h})) = Dv(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))(\mathbb{J}_K(\widehat{\boldsymbol{x}})\boldsymbol{h}).$$

Using the gradient to represent the Fréchet derivative yields (9.8a) since

$$\nabla(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})\cdot\boldsymbol{h} = D(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})(\boldsymbol{h}) = Dv(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))(\mathbb{J}_K(\widehat{\boldsymbol{x}})\boldsymbol{h})$$
$$= (\nabla v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\cdot(\mathbb{J}_K(\widehat{\boldsymbol{x}})\boldsymbol{h}) = (\mathbb{J}_K(\widehat{\boldsymbol{x}})^\mathsf{T}(\nabla v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}})))\cdot\boldsymbol{h}.$$

(2) Proof of (9.8c). This identity is deduced from (9.8a) by integrating by parts. Since $\boldsymbol{T}_K$ is bijective, the ratio $\epsilon_K := \frac{\det(\mathbb{J}_K)}{|\det(\mathbb{J}_K)|}$ is constant over $\widehat{K}$ and is either equal to -1 or 1. Moreover, the volume measure in K at $\boldsymbol{x}$ and in $\widehat{K}$ at $\widehat{\boldsymbol{x}}$ are s.t. $\mathrm{d}x = |\det(\mathbb{J}_K(\widehat{\boldsymbol{x}}))|\,\mathrm{d}\widehat{x}$. Let $q \in C_0^\infty(K)$ be a smooth scalar-valued function compactly supported in K. Integrating by parts and using (9.8a), we infer that

$$\int_{\widehat{K}} (\nabla\cdot\boldsymbol{v})(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))q(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\det(\mathbb{J}_K(\widehat{\boldsymbol{x}}))\,\mathrm{d}\widehat{x} = \epsilon_K \int_K (\nabla\cdot\boldsymbol{v})(\boldsymbol{x})q(\boldsymbol{x})\,\mathrm{d}x$$

$$= -\epsilon_K \int_K (\boldsymbol{v}\cdot\nabla q)(\boldsymbol{x})\,\mathrm{d}x = -\epsilon_K \int_{\widehat{K}} (\boldsymbol{v}\cdot\nabla q)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\,|\det(\mathbb{J}_K(\widehat{\boldsymbol{x}}))|\,\mathrm{d}\widehat{x}$$

$$= -\int_{\widehat{K}} ((\boldsymbol{v} \circ \boldsymbol{T}_K)\cdot(\mathbb{J}_K^{-\mathsf{T}}\nabla(q \circ \boldsymbol{T}_K)))(\widehat{\boldsymbol{x}})\det(\mathbb{J}_K(\widehat{\boldsymbol{x}}))\,\mathrm{d}\widehat{x}$$

$$= -\int_{\widehat{K}} ((\det(\mathbb{J}_K)\mathbb{J}_K^{-1}(\boldsymbol{v} \circ \boldsymbol{T}_K))\cdot\nabla(q \circ \boldsymbol{T}_K))(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{x}$$

$$= \int_{\widehat{K}} \nabla\cdot(\det(\mathbb{J}_K)\mathbb{J}_K^{-1}(\boldsymbol{v} \circ \boldsymbol{T}_K))(\widehat{\boldsymbol{x}})q(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\,\mathrm{d}\widehat{x},$$

which proves (9.8c) since q is arbitrary.

(3) Proof of (9.8b) in $\mathbb{R}^3$. Let ε be the Levi-Civita symbol ($\varepsilon_{ijk} := 0$ if at least two indices take the same value, $\varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} := 1$, and $\varepsilon_{132} = \varepsilon_{213} = \varepsilon_{321} := -1$). Recall that $\det(\mathbb{J}_K) = \varepsilon_{ijk}(\mathbb{J}_K)_{1i}(\mathbb{J}_K)_{2j}(\mathbb{J}_K)_{3k} = \varepsilon_{ijk}(\mathbb{J}_K)_{i1}(\mathbb{J}_K)_{j2}(\mathbb{J}_K)_{k3}$ and $(\nabla\times\boldsymbol{v})_i = \varepsilon_{ijk}\partial_j v_k$, with the Einstein convention on the summation of repeated indices. For all $i \in \{1{:}d\}$, we have

$$(\mathbb{J}_K\nabla\times(\mathbb{J}_K^\mathsf{T}(\boldsymbol{v}\circ\boldsymbol{T}_K)))_i = (\mathbb{J}_K)_{ij}\varepsilon_{jkl}\partial_k(\mathbb{J}_K^\mathsf{T}(\boldsymbol{v}\circ\boldsymbol{T}_K))_l$$
$$= (\mathbb{J}_K)_{ij}\varepsilon_{jkl}\partial_k((\mathbb{J}_K^\mathsf{T})_{lm}(v_m\circ\boldsymbol{T}_K))$$
$$= (\mathbb{J}_K)_{ij}\varepsilon_{jkl}\left(\partial_k(\mathbb{J}_K)_{ml}(v_m\circ\boldsymbol{T}_K) + (\mathbb{J}_K)_{ml}\partial_k(v_m\circ\boldsymbol{T}_K)\right).$$

Let $\mathfrak{T}_1$ and $\mathfrak{T}_2$ be the two terms on the right-hand side of the above equality. Since $\partial_k(\mathbb{J}_K)_{ml} = \partial_{kl}(\boldsymbol{T}_K)_m = \partial_{lk}(\boldsymbol{T}_K)_m = \partial_l(\mathbb{J}_K)_{mk}$, we infer that $\mathfrak{T}_1 = (\mathbb{J}_K)_{ij}\frac{1}{2}(\varepsilon_{jkl} + \varepsilon_{jlk})\partial_k(\mathbb{J}_K)_{ml}(v_m\circ\boldsymbol{T}_K) = 0$. Moreover, since $\varepsilon_{jkl}(\mathbb{J}_K)_{ij}(\mathbb{J}_K)_{nk}(\mathbb{J}_K)_{ml} = \varepsilon_{inm}\det(\mathbb{J}_K)$, we infer that

$$\mathfrak{T}_2 = (\mathbb{J}_K)_{ij}\varepsilon_{jkl}(\mathbb{J}_K)_{ml}((\partial_n v_m)\circ\boldsymbol{T}_K)(\mathbb{J}_K)_{nk}$$
$$= \varepsilon_{jkl}(\mathbb{J}_K)_{ij}(\mathbb{J}_K)_{nk}(\mathbb{J}_K)_{ml}((\partial_n v_m)\circ\boldsymbol{T}_K)$$
$$= \varepsilon_{inm}\det(\mathbb{J}_K)((\partial_n v_m)\circ\boldsymbol{T}_K) = \det(\mathbb{J}_K)((\nabla\times\boldsymbol{v})\circ\boldsymbol{T}_K)_i. \qquad\square$$

Remark 9.7 (Literature). See Marsden and Hughes [139, pp. 116–119], Ciarlet [75, p. 39], Monk [145, §3.9], Rognes et al. [168, p. 4134]. $\qquad\square$

Definition 9.8 (Piola transformations). *Let $v \in C^0(K)$ and $\boldsymbol{v} \in \boldsymbol{C}^0(K)$. The* Piola transformations *are defined as follows:*

$$\psi_K^{\mathrm{g}}(v) := v\circ\boldsymbol{T}_K, \tag{9.9a}$$
$$\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) := \mathbb{J}_K^\mathsf{T}(\boldsymbol{v}\circ\boldsymbol{T}_K), \tag{9.9b}$$
$$\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}(\boldsymbol{v}\circ\boldsymbol{T}_K), \tag{9.9c}$$
$$\psi_K^{\mathrm{b}}(v) := \det(\mathbb{J}_K)(v\circ\boldsymbol{T}_K). \tag{9.9d}$$

ψ_K^{g} *is called* pullback by the geometric mapping, $\boldsymbol{\psi}_K^{\mathrm{c}}$ *is called* covariant Piola transformation, *and* $\boldsymbol{\psi}_K^{\mathrm{d}}$ *is called* contravariant Piola transformation.

Corollary 9.9 (Commuting properties). *The Piola transformations are such that for all $v \in C^1(K)$ and all $\boldsymbol{v} \in \boldsymbol{C}^1(K)$,*

$$\nabla(\psi_K^{\mathrm{g}}(v)) = \boldsymbol{\psi}_K^{\mathrm{c}}(\nabla v), \quad \nabla\times(\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})) = \boldsymbol{\psi}_K^{\mathrm{d}}(\nabla\times\boldsymbol{v}), \quad \nabla\cdot(\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})) = \psi_K^{\mathrm{b}}(\nabla\cdot\boldsymbol{v}).$$

Proof. Apply Lemma 9.6. $\qquad\square$

The superscript g (resp., c, d) refers to the fact that the map ψ_K^{g} (resp., $\boldsymbol{\psi}_K^{\mathrm{c}}, \boldsymbol{\psi}_K^{\mathrm{d}}$) is used when integrability properties on the gradient (resp., curl, divergence) are required. The superscript b for "broken" means that no integrability with respect to any differential operator is invoked.

9.2.2 Normal and tangent vectors

Another important property of the Piola transformations is that $\boldsymbol{\psi}_K^{\mathrm{c}}$ (resp., $\boldsymbol{\psi}_K^{\mathrm{d}}$) preserves the moments of the tangential (resp., normal) components of fields at the edges (resp., the faces) of the mesh cell K. Let us first motivate this claim by a simple example.

Example 9.10 (Piola transformation vs. pullback). Referring to Figure 9.1, let $\widehat{K}$ be the triangle with vertices $(0,0)$, $(1,0)$, and $(0,1)$. Let K be the image of $\widehat{K}$ by the geometric mapping $\boldsymbol{T}_K$ defined as the rotation of center $(0,0)$ and of angle $\frac{\pi}{2}$. Let $\widehat{F}_1$ (resp., $\widehat{F}_2$) be the edge of $\widehat{K}$ corresponding to $x_2 = 0$ (resp., $x_1 = 0$), and let F_1 and F_2 be the images of $\widehat{F}_1$ and $\widehat{F}_2$ by $\boldsymbol{T}_K$, respectively. Consider the constant field $\boldsymbol{v}(\boldsymbol{x}) := (1,0)^\mathsf{T}$. Note that $\psi_K^{\mathrm{g}}(\boldsymbol{v}) = \boldsymbol{v}$ since $\boldsymbol{v}$ is invariant under the pullback by $\boldsymbol{T}_K$ (applied componentwise). Hence, $\boldsymbol{v}$ is tangent to F_2, whereas $\psi_K^{\mathrm{g}}(\boldsymbol{v})$ is normal to $\widehat{F}_2$. Moreover, $\boldsymbol{v}$ is normal to F_1, whereas $\psi_K^{\mathrm{g}}(\boldsymbol{v})$ is tangent to $\widehat{F}_1$. But $\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) = (0,-1)^\mathsf{T}$ is tangent to $\widehat{F}_2$ and $\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) = (0,-1)^\mathsf{T}$ is normal to $\widehat{F}_1$. □
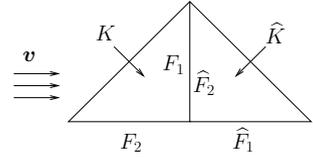
Fig. 9.1 Illustration of Example 9.10.

Our first result identifies how the geometric mapping $\boldsymbol{T}_K : \widehat{K} \to K$ transforms normal and tangent vectors on $\partial\widehat{K}$.

Lemma 9.11 (Normal and tangent).   (i) *Let $\widehat{\boldsymbol{n}}_{\widehat{K}}$ be the outward unit normal to $\partial\widehat{K}$ and let $\boldsymbol{n}_K$ be the outward unit normal to ∂K. Let $\widehat{F}$ be a face of $\widehat{K}$ and let $F := \boldsymbol{T}_K(\widehat{F})$ be the corresponding face of K. Let $\widehat{\boldsymbol{x}} \in int(\widehat{F})$ so that $\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}(\widehat{\boldsymbol{x}})$ is well defined, and let $\boldsymbol{x} := \boldsymbol{T}_K(\widehat{\boldsymbol{x}}) \in int(F)$. Then we have*

$$\boldsymbol{n}_{K|F}(\boldsymbol{x}) = \frac{1}{\|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}})(\widehat{\boldsymbol{x}})\|_{\ell^2}}(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}})(\widehat{\boldsymbol{x}}). \tag{9.10}$$

(ii) *Let $\widehat{E}$ be an edge of $\widehat{K}$ and let $E := \boldsymbol{T}_K(\widehat{E})$ be the corresponding edge of K. Let $\widehat{\boldsymbol{x}} \in int(\widehat{E})$, let $\widehat{\boldsymbol{\tau}}_{\widehat{E}}$ be a unit tangent vector to $\widehat{E}$ at $\widehat{\boldsymbol{x}}$, and let $\boldsymbol{x} := \boldsymbol{T}_K(\widehat{\boldsymbol{x}}) \in E$. Then the vector*

$$\boldsymbol{\tau}_E(\boldsymbol{x}) := \frac{1}{\|(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}})\|_{\ell^2}}(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}}) \tag{9.11}$$

is a unit tangent vector to E at $\boldsymbol{x}$.

Proof. (1) Let $\widehat{\psi}$ be the signed distance function to $\widehat{F}$, assumed to be negative inside $\widehat{K}$. Then $\nabla\widehat{\psi}(\widehat{\boldsymbol{x}}) = \widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}(\widehat{\boldsymbol{x}})$. Defining $\psi := \widehat{\psi}\circ\boldsymbol{T}_K^{-1}$ and using (9.8a), we have $\nabla\psi(\boldsymbol{x}) = \mathbb{J}_K^{-\mathsf{T}}(\widehat{\boldsymbol{x}})\nabla\widehat{\psi}(\widehat{\boldsymbol{x}}) = \mathbb{J}_K^{-\mathsf{T}}(\widehat{\boldsymbol{x}})\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}(\widehat{\boldsymbol{x}})$. Since ψ is constant (equal to zero) over $int(F)$ and takes negative values inside K, the vector $\nabla\psi(\boldsymbol{x})$ is normal to F and points toward the inside of K. This proves (9.10).
(2) Consider an edge $\widehat{E} := \widehat{F}_1 \cap \widehat{F}_2$ of $\widehat{K}$ and let $\widehat{\boldsymbol{x}} \in \widehat{E}$. Since

$$(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}_i})(\widehat{\boldsymbol{x}})\cdot(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}}(\widehat{\boldsymbol{x}})) = \widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}_i}(\widehat{\boldsymbol{x}})\cdot\widehat{\boldsymbol{\tau}}_{\widehat{E}}(\widehat{\boldsymbol{x}}) = 0,$$

we infer from Step (1) that $\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}}(\widehat{\boldsymbol{x}})$ is tangent to $F_i := \boldsymbol{T}_K(\widehat{F}_i)$ for all $i \in \{1,2\}$. Hence, $(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}})$ is tangent to $F_i \cap F_j = E = T_K(\widehat{E})$. $\qquad\qquad\square$

Our next step is to identify how surface and line measures are transformed by the geometric mapping $\boldsymbol{T}_K$. Observe that the unit of $\mathbb{J}_K$ is a length scale and that the unit of $\det(\mathbb{J}_K)$ is a volume. The identity (9.12a) is sometimes called *Nanson's formula* in the continuum mechanics literature; see [149, p. 184] and Truesdell and Toupin [192, p. 249, Eq. (20.8)].

Lemma 9.12 (Surface and line measures). *The surface measures on $\widehat{F}$ at $\widehat{\boldsymbol{x}}$ and on $F := \boldsymbol{T}_K(\widehat{F})$ at $\boldsymbol{x} := \boldsymbol{T}_K(\widehat{\boldsymbol{x}})$ are such that*

$$\mathrm{d}s = |\det(\mathbb{J}_K)(\widehat{\boldsymbol{x}})| \, \|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}})(\widehat{\boldsymbol{x}})\|_{\ell^2} \, \mathrm{d}\widehat{s}, \qquad\qquad (9.12a)$$

$$\mathrm{d}\widehat{s} = |\det(\mathbb{J}_K^{-1})(\boldsymbol{x})| \, \|(\mathbb{J}_K^{\mathsf{T}}\boldsymbol{n}_{K|F})(\boldsymbol{x})\|_{\ell^2} \, \mathrm{d}s. \qquad\qquad (9.12b)$$

The line measures on $\widehat{E}$ at $\widehat{\boldsymbol{x}}$ and on $E := \boldsymbol{T}_K(\widehat{E})$ at $\boldsymbol{x} := \boldsymbol{T}_K(\widehat{\boldsymbol{x}})$ are such that

$$\mathrm{d}l = \|(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\boldsymbol{E}})(\widehat{\boldsymbol{x}})\|_{\ell^2} \, \mathrm{d}\widehat{l}, \qquad \mathrm{d}\widehat{l} = \|(\mathbb{J}_K^{-1}\boldsymbol{\tau}_E)(\boldsymbol{x})\|_{\ell^2} \, \mathrm{d}l. \qquad (9.13)$$

Proof. Let $q \in C_0^\infty(F)$ and let $\boldsymbol{v} \in \boldsymbol{C}^\infty(K)$ be s.t. $\boldsymbol{v}\cdot\boldsymbol{n}_{K|F} = q$ and $\boldsymbol{v}\cdot\boldsymbol{n}_{K|\partial K \setminus F} = 0$ (this construction is possible since q is compactly supported in F and so vanishes near ∂F where $\boldsymbol{n}_K$ is multivalued). Recall that $\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}(\boldsymbol{v} \circ \boldsymbol{T}_K)$ and that $\epsilon_K := \frac{\det(\mathbb{J}_K)}{|\det(\mathbb{J}_K)|} = \pm 1$. Using (9.8c) and (9.10), we infer that

$$\begin{aligned}
\int_F q(\boldsymbol{x})\,\mathrm{d}s &= \int_{\partial K}(\boldsymbol{v}\cdot\boldsymbol{n}_K)(\boldsymbol{x})\,\mathrm{d}s = \int_K(\nabla\cdot\boldsymbol{v})(\boldsymbol{x})\,\mathrm{d}x \\
&= \epsilon_K \int_{\widehat{K}}\nabla\cdot\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{x} = \epsilon_K \int_{\partial\widehat{K}}(\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{n}}_{\widehat{K}})(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{s} \\
&= \int_{\partial\widehat{K}}(\mathbb{J}_K^{-1}\boldsymbol{v})\cdot(\mathbb{J}_K^{\mathsf{T}}\boldsymbol{n}_K)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}})(\widehat{\boldsymbol{x}})\|_{\ell^2}|\det(\mathbb{J}_K)(\widehat{\boldsymbol{x}})|\,\mathrm{d}\widehat{s} \\
&= \int_{\partial\widehat{K}}(\boldsymbol{v}\cdot\boldsymbol{n}_K)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}})(\widehat{\boldsymbol{x}})\|_{\ell^2}|\det(\mathbb{J}_K)(\widehat{\boldsymbol{x}})|\,\mathrm{d}\widehat{s} \\
&= \int_{\widehat{F}}q(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}})(\widehat{\boldsymbol{x}})\|_{\ell^2}|\det(\mathbb{J}_K)(\widehat{\boldsymbol{x}})|\,\mathrm{d}\widehat{s}.
\end{aligned}$$

This yields (9.12a). To prove (9.12b), we use the following identity:

$$\|(\mathbb{J}_K^{\mathsf{T}}\boldsymbol{n}_{K|F})(\boldsymbol{x})\|_{\ell^2} = \|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}})(\widehat{\boldsymbol{x}})\|_{\ell^2}^{-1},$$

which follows from (9.10) and the fact that $\boldsymbol{n}_K$ and $\widehat{\boldsymbol{n}}_{\widehat{K}}$ are unit vectors. We refer the reader to Exercise 9.2 for the transformation of line measures. $\quad\square$

We can now state the main result of this section showing that the Piola transformations $\boldsymbol{\psi}_K^{\mathrm{c}}$ and $\boldsymbol{\psi}_K^{\mathrm{d}}$ are tailored to preserve the moments of the

tangential components of fields over edges and the moments of the normal components of fields over faces, respectively. Let $\widehat{F}$ be a face of $\widehat{K}$ and let $\widehat{E}$ be an edge of $\widehat{K}$. Let $F := \boldsymbol{T}_K(\widehat{F})$ and $E := \boldsymbol{T}_K(\widehat{E})$ be the corresponding face and edge of K. Let $\widehat{\boldsymbol{n}}_{\widehat{F}}$ be a unit vector normal to $\widehat{F}$ and let $\widehat{\boldsymbol{\tau}}_{\widehat{E}}$ be a unit vector tangent to $\widehat{E}$. Note that $\widehat{\boldsymbol{n}}_{\widehat{F}}$ can point either toward the inside of $\widehat{K}$ or the outside of $\widehat{K}$, i.e., we only have $\widehat{\boldsymbol{n}}_{\widehat{F}} = \pm \widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}$. Recall that $\epsilon_K := \frac{\det(\mathbb{J}_K)}{|\det(\mathbb{J}_K)|} = \pm 1$. Lemma 9.11 shows that the following unit vectors:

$$\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}})(\boldsymbol{x}) := \epsilon_K \frac{1}{\|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}})\|_{\ell^2}} (\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}}), \qquad (9.14a)$$

$$\boldsymbol{\Phi}_K^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\boldsymbol{x}) := \frac{1}{\|(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}})\|_{\ell^2}} (\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}}), \qquad (9.14b)$$

are, respectively, normal to F and tangent to E at $\boldsymbol{x} := \boldsymbol{T}_K(\widehat{\boldsymbol{x}})$. The definitions in (9.14) are motivated by the following result.

Lemma 9.13 (Preservation of moments of normal and tangential components). *Let $\boldsymbol{v} \in \boldsymbol{C}^0(K)$ and $q \in C^0(K)$. The following holds true:*

$$\int_F (\boldsymbol{v}\cdot\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}}))(\boldsymbol{x})q(\boldsymbol{x})\,\mathrm{d}s = \int_{\widehat{F}} (\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}})\psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{s}, \qquad (9.15a)$$

$$\int_E (\boldsymbol{v}\cdot\boldsymbol{\Phi}_K^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}}))(\boldsymbol{x})q(\boldsymbol{x})\,\mathrm{d}l = \int_{\widehat{E}} (\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}})\psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{l}. \qquad (9.15b)$$

Proof. To prove (9.15a), we use the transformation of surface measures from Lemma 9.12 followed by the definition (9.14a) of $\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}})$ and the definition of the maps $\boldsymbol{\psi}_K^{\mathrm{d}}$ and ψ_K^{g} (see (9.9)) to obtain

$$\int_F (\boldsymbol{v}\cdot\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}}))(\boldsymbol{x})q(\boldsymbol{x})\,\mathrm{d}s$$

$$= \int_{\widehat{F}} (\boldsymbol{v}\cdot\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}}))(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}})|\det(\mathbb{J}_K)(\widehat{\boldsymbol{x}})|\,\|\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}}(\widehat{\boldsymbol{x}})\|_{\ell^2}\,\mathrm{d}\widehat{s}$$

$$= \int_{\widehat{F}} ((\boldsymbol{v}\circ\boldsymbol{T}_K)\cdot(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}}))(\widehat{\boldsymbol{x}})\psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}})\det(\mathbb{J}_K)(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{s}$$

$$= \int_{\widehat{F}} (\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}})\psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{s}.$$

The proof of (9.15b) uses similar arguments and is left as an exercise. $\qquad\square$

Remark 9.14 (Sign of $\det(\mathbb{J}_K)$). The factor $\epsilon_K = \pm 1$ in the definition (9.14a) is due to the fact that the contravariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{d}}$ may transform an outward-pointing field into an inward-pointing field. The definition (9.14a) is such that the sign of $\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{n}_K)(\widehat{\boldsymbol{x}})\cdot\boldsymbol{n}_{\widehat{F}}(\widehat{\boldsymbol{x}})$ and the sign of $\boldsymbol{n}_K(\boldsymbol{x})\cdot\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}})(\boldsymbol{x})$ are identical. Note that $\epsilon_K = 1$ if $\det(\mathbb{J}_K) > 0$. $\square$

Exercises

Exercise 9.1 (Canonical hybrid element). Consider an affine geometric mapping $\boldsymbol{T}_K$ and the pullback by $\boldsymbol{T}_K$ for ψ_K. Let $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ be the canonical hybrid element of §7.6. Verify that Proposition 9.2 generates the canonical hybrid element in K. Write the dofs.

Exercise 9.2 (Line measure). (i) Prove Lemma 9.12 for line measures. (*Hint*: the change in line measure is $\frac{\mathrm{d}l}{\mathrm{d}\widehat{l}}(\boldsymbol{x}) = \lim_{h\to 0} \frac{\|\boldsymbol{T}_K(\widehat{\boldsymbol{x}}+h\widehat{\boldsymbol{\tau}})-\boldsymbol{T}_K(\widehat{\boldsymbol{x}})\|_{\ell^2}}{\|h\widehat{\boldsymbol{\tau}}\|_{\ell^2}}$.) (ii) Assume that $d = 2$. Show that $|\det(\mathbb{J}_K)| \|\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}\|_{\ell^2(\mathbb{R}^2)} = \|\mathbb{J}_K\widehat{\boldsymbol{\tau}}\|_{\ell^2(\mathbb{R}^2)}$ for any pair of unit vectors $(\widehat{\boldsymbol{n}}, \widehat{\boldsymbol{\tau}})$ that are orthogonal.

Exercise 9.3 (Surface measure). (i) Let $\boldsymbol{T}_F := \boldsymbol{T}_{K|\widehat{F}} : \widehat{F} \to F$ and $\widehat{\boldsymbol{x}} \in \widehat{F}$. Let $\mathbb{J}_F(\widehat{\boldsymbol{x}}) \in \mathbb{R}^{d\times(d-1)}$ be the Jacobian matrix representing the (Fréchet) derivative $D\boldsymbol{T}_F(\widehat{\boldsymbol{x}})$. Let $\mathrm{g}_F(\widehat{\boldsymbol{x}}) = (\mathbb{J}_F(\widehat{\boldsymbol{x}}))^{\mathsf{T}}\mathbb{J}_F(\widehat{\boldsymbol{x}}) \in \mathbb{R}^{(d-1)\times(d-1)}$ be the surface metric tensor at $\widehat{\boldsymbol{x}}$. Prove that $\sqrt{\det(\mathrm{g}_F(\widehat{\boldsymbol{x}}))} = |\det(\mathbb{J}_K)| \|\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}\|_{\ell^2}$. (*Hint*: use that $\mathrm{d}s = \sqrt{\det(\mathrm{g}_F(\widehat{\boldsymbol{x}}))}\,\mathrm{d}\widehat{s}$.) (ii) Let $\widehat{K} := \{(\widehat{x}_1,\widehat{x}_2,\widehat{x}_3) \in \mathbb{R}^3 \mid 0 \le \widehat{x}_1,\widehat{x}_2,\widehat{x}_3,\ \widehat{x}_1 + \widehat{x}_2 + \widehat{x}_3 \le 1\}$ be the unit simplex in $\mathbb{R}^3$. Let $\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := (\widehat{x}_1,\widehat{x}_2,\widehat{x}_1^2+\widehat{x}_2^2-\widehat{x}_3)^{\mathsf{T}}$. Let $\widehat{F}$ be the face $\{\widehat{x}_3 = 0\}$ and $F := \boldsymbol{T}_K(\widehat{F})$. Compute $\mathbb{J}_F$, $\mathbb{J}_K$, g_F and verify the identity proved in Step (i).

Exercise 9.4 (Sobolev spaces). Prove that ψ_K^{g} is a bounded isomorphism from $H^1(K)$ to $H^1(\widehat{K})$, that $\boldsymbol{\psi}_K^{\mathrm{c}}$ is a bounded isomorphism from $\boldsymbol{H}(\mathrm{curl}; K)$ to $\boldsymbol{H}(\mathrm{curl}; \widehat{K})$, and that $\boldsymbol{\psi}_K^{\mathrm{d}}$ is a bounded isomorphism from $\boldsymbol{H}(\mathrm{div}; K)$ to $\boldsymbol{H}(\mathrm{div}; \widehat{K})$.

Exercise 9.5 (Transformation of cross products). Let $\mathbb{A}$ be a 3×3 invertible matrix. Prove that $\mathbb{A}^{-\mathsf{T}}(\boldsymbol{x}\times\boldsymbol{y}) = \det(\mathbb{A})^{-1}(\mathbb{A}\boldsymbol{x}\times\mathbb{A}\boldsymbol{y})$ for any vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$.

Exercise 9.6 ((9.15b)). Prove (9.15b).

Part III, Chapter 10

Mesh orientation

Orienting the edges and the faces of a mesh is crucial when working with finite elements whose degrees of freedom invoke normal or tangential components of vector fields. This notion is important also when working with high-order scalar-valued finite elements to enumerate consistently all the degrees of freedom in each mesh cell sharing the edge or the face in question. In this chapter, we focus on matching meshes (see Definition 8.11), and we assume that the meshes are affine. We first explain how to orient meshes. Then we introduce the important notion of generation-compatible orientation. Finally, we study whether simplicial, quadrangular, and hexahedral meshes can be equipped with a generation-compatible orientation.

10.1 How to orient a mesh

Let us consider a three-dimensional matching mesh. The geometric entities to be oriented are the mesh edges $E \in \mathcal{E}_h$ and the mesh faces $F \in \mathcal{F}_h$ (one can also orient the vertices and the cells of the mesh, but for simplicity, we will not introduce these notions here). The edges of the mesh are oriented by specifying how to circulate along them. This is done by fixing one unit vector tangent to each edge. The faces of the mesh are oriented by specifying how to cross them. This is done by fixing one unit normal vector on each face. Orienting the mesh thus means that we fix once and for all the following collections of unit vectors:

$$\{\boldsymbol{\tau}_E\}_{E \in \mathcal{E}_h}, \qquad \{\boldsymbol{n}_F\}_{F \in \mathcal{F}_h}. \tag{10.1}$$

Since the mesh is affine, the mesh edges are straight and the mesh faces are planar. Hence, one single tangent vector is enough to orient each edge and one normal vector is enough to orient each face.

Let us now consider a two-dimensional mesh. Then the mesh edges and the mesh faces are identical one-dimensional manifolds in $\mathbb{R}^2$, but they are

oriented differently. The orientation of the mesh edges is done as in the three-dimensional case by fixing once and for all a unit tangent vector along the edge, whereas the mesh faces are oriented by rotating the unit tangent vectors anti-clockwise, i.e., for every edge E oriented by the vector $\boldsymbol{\tau}_E$, we set

$$\boldsymbol{n}_E := \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{\tau}_E), \tag{10.2}$$

where the matrix of $\boldsymbol{R}_{\frac{\pi}{2}}$ relative to the canonical basis of $\mathbb{R}^2$ is $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

It is useful to define the following subsets: For every mesh edge $E \in \mathcal{E}_h$ and for every mesh face $F \in \mathcal{F}_h$,

$$\mathcal{T}_E := \{K \in \mathcal{T}_h \mid E \subset K\}, \qquad \mathcal{T}_F := \{K \in \mathcal{T}_h \mid F \subset K\}, \tag{10.3}$$

are the collection of the mesh cells sharing E and F, respectively. The cardinality of the subset $\mathcal{T}_E$ cannot be ascertained a priori, whereas we have $\mathcal{T}_F = \{K_l, K_r\}$ for every interface $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$ and $\mathcal{T}_F = \{K_l\}$ for every boundary face $F := \partial K_l \cap \partial D \in \mathcal{F}_h^\partial$; see Definition 8.10.

Remark 10.1 (Face orientation in 3D). The faces of cells in three-dimensional meshes have connected boundaries. Hence, instead of assigning a normal vector to each face, one can also orient the faces by specifying how to circulate along their boundary. The two ways of orienting faces are equivalent once an orientation for the ambient space $\mathbb{R}^3$ has been fixed (by using the right-hand convention, for example). The boundary-based orientation is more intrinsic since it does not require to embed the faces into $\mathbb{R}^3$. In this book, we adopt the normal-based orientation introduced in (10.1) since it is more convenient to use with finite elements. □

Remark 10.2 (Incidence matrices). Consider a three-dimensional mesh where the vertices, edges, faces, and cells have been enumerated from 1 to N_v, N_e, N_f, and N_c, respectively. Assume that the mesh has been oriented. Incidence matrices can then be defined as follows. The matrix $\mathcal{M}^{\mathrm{ev}} \in \mathbb{R}^{N_e \times N_v}$ is s.t. $\mathcal{M}^{\mathrm{ev}}_{ml} := 1$ if $\boldsymbol{z}_l$ is a vertex of E_m and $\boldsymbol{\tau}_{E_m}$ points toward $\boldsymbol{z}_l$, $\mathcal{M}^{\mathrm{ev}}_{ml} := -1$ if $\boldsymbol{\tau}_{E_m}$ points in the opposite direction, and $\mathcal{M}^{\mathrm{ev}}_{ml} := 0$ if $\boldsymbol{z}_l$ is not a vertex of E_m. The matrix $\mathcal{M}^{\mathrm{fe}} \in \mathbb{R}^{N_f \times N_e}$ is s.t. $\mathcal{M}^{\mathrm{fe}}_{ml} := 1$ if E_l is an edge of F_m and the orientation of E_l prescribed by $\boldsymbol{\tau}_{E_l}$ and that induced by $\boldsymbol{n}_{F_m}$ on $E_l \subset \partial F_m$ using the right-hand convention are the same, $\mathcal{M}^{\mathrm{fe}}_{ml} := -1$ if these orientation are opposite, and $\mathcal{M}^{\mathrm{fe}}_{ml} := 0$ if E_l is not an edge of F_m. The matrix $\mathcal{M}^{\mathrm{cf}} \in \mathbb{R}^{N_c \times N_f}$ is s.t. $\mathcal{M}^{\mathrm{cf}}_{ml} := 1$ if F_l is a face of K_m and $\boldsymbol{n}_F$ points toward the outside of K_m, $\mathcal{M}^{\mathrm{cf}}_{ml} := -1$ if $\boldsymbol{n}_F$ points toward the inside, and $\mathcal{M}^{\mathrm{cf}}_{ml} := 0$ if F_l is not a face of K_m. The incidence matrices $\mathcal{M}^{\mathrm{ev}}$, $\mathcal{M}^{\mathrm{fe}}$, and $\mathcal{M}^{\mathrm{cf}}$ can be viewed as discrete counterparts of the gradient, curl, and divergence operators, respectively. In particular, we have $\mathcal{M}^{\mathrm{fe}}\mathcal{M}^{\mathrm{ev}} = 0_{\mathbb{R}^{N_f \times N_v}}$ and $\mathcal{M}^{\mathrm{cf}}\mathcal{M}^{\mathrm{fe}} = 0_{\mathbb{R}^{N_c \times N_e}}$. We refer the reader to Bossavit [37], Bochev and Hyman [27], Bonelle and Ern [32], Gerritsma [106], and the references therein for further insight into this topic. □

10.2 Generation-compatible orientation

Let $\mathcal{T}_h$ be an oriented mesh and let $K \in \mathcal{T}_h$ be a mesh cell. Recall that the cell K is generated using a geometric mapping $\boldsymbol{T}_K : \widehat{K} \to K$. One of the key results from the previous chapter, Lemma 9.13, deals with the preservation of the moments of the normal and tangential components of fields defined on K. Let $\widehat{F}$ be a face of $\widehat{K}$ and let $\widehat{E}$ be an edge of $\widehat{K}$. Let $F := \boldsymbol{T}_K(\widehat{F})$ and $E := \boldsymbol{T}_K(\widehat{E})$ be the corresponding face and edge of K. Let $\widehat{\boldsymbol{n}}_{\widehat{F}}$ be a unit vector normal to $\widehat{F}$ and let $\widehat{\boldsymbol{\tau}}_{\widehat{E}}$ be a unit vector tangent to $\widehat{E}$. Recall from (9.14) that $\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}})(\boldsymbol{x}) := \epsilon_K \|(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}})\|_{\ell^2}^{-1}(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}})$ is a unit vector normal to F and that $\boldsymbol{\Phi}_K^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\boldsymbol{x}) := \|(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}})\|_{\ell^2}^{-1}(\mathbb{J}_K\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}})$ is a unit vector tangent to E, where $\mathbb{J}_K$ is the Jacobian matrix of $\boldsymbol{T}_K$, $\epsilon_K := \frac{\det(\mathbb{J}_K)}{|\det(\mathbb{J}_K)|} = \pm 1$, and $\boldsymbol{x} := \boldsymbol{T}_K(\widehat{\boldsymbol{x}})$. With the Piola transformations ψ_K^{g}, $\boldsymbol{\psi}_K^{\mathrm{c}}$, and $\boldsymbol{\psi}_K^{\mathrm{d}}$ defined in Definition 9.8, Lemma 9.13 states that the following holds true for all $\boldsymbol{v} \in \boldsymbol{C}^0(K)$ and all $q \in C^0(K)$:

$$\int_F (\boldsymbol{v}\cdot\boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}}))(\boldsymbol{x})q(\boldsymbol{x})\,\mathrm{d}s = \int_{\widehat{F}} (\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}})\psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{s}, \qquad (10.4a)$$

$$\int_E (\boldsymbol{v}\cdot\boldsymbol{\Phi}_K^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}}))(\boldsymbol{x})q(\boldsymbol{x})\,\mathrm{d}l = \int_{\widehat{E}} (\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}})\psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}})\,\mathrm{d}\widehat{l}. \qquad (10.4b)$$

Since we are going to define face and edge dofs for vector-valued finite elements by using the right-hand sides in (10.4), we want to make sure that the results do not depend on the mapping $\boldsymbol{T}_K : \widehat{K} \to K$. For instance, let $F \in \mathcal{F}_h$ be an interface, i.e., $F := \partial K_l \cap \partial K_r$ so that $\mathcal{T}_F = \{K_l, K_r\}$. The way to ascertain that the right-hand side of (10.4a) gives the same results on both sides of F consists of requiring that

$$\boldsymbol{n}_F = \boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}}), \quad \forall K \in \mathcal{T}_F, \text{ with } \widehat{F} := \boldsymbol{T}_K^{-1}(F), \qquad (10.5)$$

that is, letting $\widehat{F}_l := \boldsymbol{T}_{K_l}^{-1}(F)$ and $\widehat{F}_r := \boldsymbol{T}_{K_r}^{-1}(F)$, we would like that $\boldsymbol{n}_F = \boldsymbol{\Phi}_{K_l}^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}_l}) = \boldsymbol{\Phi}_{K_r}^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}_r})$. This idea is illustrated in Figure 10.1.
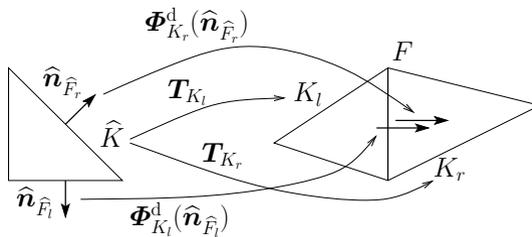
Fig. 10.1 Orientation transfer for face normals.

Similarly, given a mesh edge $E \in \mathcal{E}_h$ oriented by the fixed unit tangent vector $\boldsymbol{\tau}_E$, we want to ascertain that for every mesh cell K of which E is an edge, i.e., for all $K \in \mathcal{T}_E$ (see (10.3)), we have $\boldsymbol{\tau}_E = \boldsymbol{\Phi}_K^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}})$ where $\widehat{E} := \boldsymbol{T}_K^{-1}(E)$. This leads to the following important notion.

Definition 10.3 (Generation-compatible orientation). *Let $\mathcal{T}_h$ be an oriented mesh specified by the collections of unit tangent vectors $\{\boldsymbol{\tau}_E\}_{E \in \mathcal{E}_h}$ and unit normal vectors $\{\boldsymbol{n}_F\}_{F \in \mathcal{F}_h}$ as in (10.1). We say that this orientation is generation-compatible if there is an orientation of the reference cell $\widehat{K}$ specified by the unit tangent vectors $\{\widehat{\boldsymbol{\tau}}_{\widehat{E}}\}_{\widehat{E} \in \mathcal{E}_{\widehat{K}}}$ and the unit normal vectors $\{\widehat{\boldsymbol{n}}_{\widehat{F}}\}_{\widehat{F} \in \mathcal{F}_{\widehat{K}}}$ and a collection of geometric mappings $\{\boldsymbol{T}_K\}_{K \in \mathcal{T}_h}$ such that for all $E \in \mathcal{E}_h$ and all $F \in \mathcal{F}_h$,*

$$\boldsymbol{\tau}_E = \boldsymbol{\Phi}_K^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}}), \qquad \forall K \in \mathcal{T}_E, \ \widehat{E} := \boldsymbol{T}_K^{-1}(E), \tag{10.6a}$$

$$\boldsymbol{n}_F = \boldsymbol{\Phi}_K^{\mathrm{d}}(\widehat{\boldsymbol{n}}_{\widehat{F}}), \qquad \forall K \in \mathcal{T}_F, \ \widehat{F} := \boldsymbol{T}_K^{-1}(F). \tag{10.6b}$$

The key consequence of the notion of generation-compatible mesh is the following result which says that the moments of the normal and tangential components of vector fields are preserved by the transformations $\psi_K^{\mathrm{g}}, \boldsymbol{\psi}_K^{\mathrm{c}}, \boldsymbol{\psi}_K^{\mathrm{d}}$.

Lemma 10.4 (Preservation of moments of normal and tangential components). *Assume that the orientation of $\mathcal{T}_h$ is generation-compatible and let $\boldsymbol{\tau}_E$, $\boldsymbol{n}_F$ be defined in (10.6). The following holds true for all $\boldsymbol{v} \in \boldsymbol{C}^0(K)$ and all $q \in C^0(K)$:*

$$\int_F (\boldsymbol{v} \cdot \boldsymbol{n}_F)(\boldsymbol{x}) q(\boldsymbol{x}) \, \mathrm{d}s = \int_{\widehat{F}} (\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) \cdot \widehat{\boldsymbol{n}}_{\widehat{F}})(\widehat{\boldsymbol{x}}) \psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}}) \, \mathrm{d}\widehat{s}, \tag{10.7a}$$

$$\int_E (\boldsymbol{v} \cdot \boldsymbol{\tau}_E)(\boldsymbol{x}) q(\boldsymbol{x}) \, \mathrm{d}l = \int_{\widehat{E}} (\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) \cdot \widehat{\boldsymbol{\tau}}_{\widehat{E}})(\widehat{\boldsymbol{x}}) \psi_K^{\mathrm{g}}(q)(\widehat{\boldsymbol{x}}) \, \mathrm{d}\widehat{l}. \tag{10.7b}$$

Proof. Apply Lemma 9.13.   □

Whether it is possible to orient a mesh in a generation-compatible way is not guaranteed for general meshes. However, we will see in the following sections that this is indeed possible for simplicial meshes in any dimension, for quadrangular meshes, and for hexahedral meshes (possibly up to an additional subdivision of the cells). The key idea to achieve this is the increasing vertex-index enumeration technique introduced in the next section.

Remark 10.5 (Faces in 2D). Recall that the mesh edges and faces are identical one-dimensional manifolds in $\mathbb{R}^2$, and that we have adopted the convention that once the edges are oriented, the faces are oriented by rotating the unit tangent vectors anti-clockwise; see (10.2). It is proved in Exercise 10.1 that $\boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{\Phi}_K^{\mathrm{c}}(\boldsymbol{z})) = \boldsymbol{\Phi}_K^{\mathrm{d}}(\boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{z}))$ for all $\boldsymbol{z} \in \mathbb{R}^2$. Hence, if (10.6a) holds true,

then (10.6b) holds true as well, because in this case, $\boldsymbol{n}_E := \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{\tau}_E) = \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{\Phi}_K^c(\widehat{\boldsymbol{\tau}}_{\widehat{E}})) = \boldsymbol{\Phi}_K^d(\boldsymbol{R}_{\frac{\pi}{2}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}})) =: \boldsymbol{\Phi}_K^d(\widehat{\boldsymbol{n}}_{\widehat{E}})$. In conclusion, one only needs to prove (10.6a) in dimension two. $\square$

10.3 Increasing vertex-index enumeration

The increasing vertex-index enumeration technique described in this section is the key tool to orient meshes in a generation-compatible way. The technique is illustrated for various types of meshes in §10.4 and §10.5.

Let us enumerate the edges and the faces of $\widehat{K}$ from 1 to n_{ce} and from 1 to n_{cf}, respectively. Orienting the reference cell $\widehat{K}$ consists of prescribing the following unit vectors:

$$\{\widehat{\boldsymbol{\tau}}_{\widehat{E}_n}\}_{n \in \{1:n_{\mathrm{ce}}\}}, \qquad \{\widehat{\boldsymbol{n}}_{\widehat{F}_n}\}_{n \in \{1:n_{\mathrm{cf}}\}}.$$

Recalling the connectivity arrays j_ce and j_cf defined in (8.12), any mesh edge E_l for all $l \in \{1:N_{\mathrm{e}}\}$ satisfies $E_l = \boldsymbol{T}_{K_m}(\widehat{E}_n)$ with $(m,n) \in \{1:N_{\mathrm{c}}\} \times \{1:n_{\mathrm{ce}}\}$ s.t. j_ce$(m,n) = l$. Similarly, any mesh face F_l for all $l \in \{1:N_{\mathrm{f}}\}$ satisfies $F_l = \boldsymbol{T}_{K_m}(\widehat{F}_n)$ with $(m,n) \in \{1:N_{\mathrm{c}}\} \times \{1:n_{\mathrm{cf}}\}$ s.t. j_cf$(m,n) = l$.

Definition 10.6 (Increasing vertex-index enumeration). *A mesh $\mathcal{T}_h$ is said to be oriented according to the* increasing vertex-index convention *if:*

(i) *Every edge E_n with vertices $\boldsymbol{z}_p, \boldsymbol{z}_q$, $p < q$, is oriented by the vector $\boldsymbol{\tau}_{E_n} := \|\boldsymbol{t}_{p,q}\|_{\ell^2}^{-1}\boldsymbol{t}_{p,q}$ with $\boldsymbol{t}_{p,q} := \boldsymbol{z}_q - \boldsymbol{z}_p$;*

(ii) *Every face F_n in dimension two is oriented by the vector $\boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{\tau}_{F_n})$ (here F_n is viewed as an edge, and $\boldsymbol{R}_{\frac{\pi}{2}}$ is the rotation of angle $\frac{\pi}{2}$ in $\mathbb{R}^2$ as in (10.2)), and every face F_n in dimension three is oriented by the vector $\boldsymbol{n}_{F_n} := \|\boldsymbol{t}_{p,q} \times \boldsymbol{t}_{p,r}\|_{\ell^2}^{-1}(\boldsymbol{t}_{p,q} \times \boldsymbol{t}_{p,r})$, where $p < q < r$ are the three global indices of the vertices of F_n.*

The increasing vertex-index enumeration is illustrated in Figure 10.2 for the unit simplex and the unit cuboid in dimension two and dimension three.
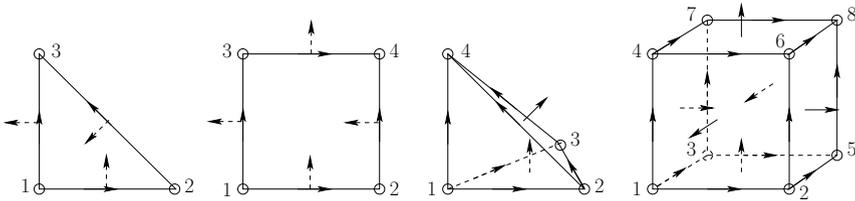
Fig. 10.2 Enumeration of the vertices and orientation of the edges and faces in the reference simplex and the reference cuboid in dimensions two and three.

2D triangle	$\widehat{z}_1 = (0,0), \widehat{z}_2 = (1,0), \widehat{z}_3 = (0,1)$
3D tetrahedron	$\widehat{z}_1 = (0,0,0), \widehat{z}_2 = (1,0,0), \widehat{z}_3 = (0,1,0), \widehat{z}_4 = (0,0,1)$
2D square	$\widehat{z}_1 = (0,0), \widehat{z}_2 = (0,1), \widehat{z}_3 = (1,0), \widehat{z}_4 = (1,1)$
3D cube	$\widehat{z}_1 = (0,0,0), \widehat{z}_2 = (1,0,0), \widehat{z}_3 = (0,1,0), \widehat{z}_4 = (0,0,1)$ $\widehat{z}_5 = (1,1,0), \widehat{z}_6 = (1,0,1), \widehat{z}_7 = (0,1,1), \widehat{z}_8 = (1,1,1)$

Table 10.1 Enumeration of the vertices in the reference simplex and in the reference cuboid in dimensions two and three.

Unless specified otherwise, we enumerate the vertices of the reference element $\widehat{K}$ by using the convention described in Table 10.1. Moreover, $\widehat{K}$ is oriented by using the convention of the increasing vertex-index enumeration as in Figure 10.2.

10.4 Simplicial meshes

Recall that the reference simplex $\widehat{K}$ is oriented by using the increasing vertex-index technique. Let us show that it is possible to find a generation-compatible orientation for every three-dimensional affine mesh $\mathcal{T}_h$ composed of simplices (the construction proposed thereafter is actually independent of the space dimension). The key idea is to orient $\mathcal{T}_h$ by using the increasing vertex-index enumeration. More precisely, let $\{z_n\}_{n \in \{1: N_v\}}$ be the mesh vertices. For every edge E_l with end vertices z_p, z_q, where $p < q$, we orient E_l by introducing $t_{p,q} := z_q - z_p$ and by setting

$$\tau_{E_l} := \|t_{p,q}\|_{\ell^2}^{-1} t_{p,q}. \tag{10.8}$$

For every face F_l defined by its three vertices, say z_p, z_q, z_r with $p < q < r$, we orient F_l by introducing $t_{p,q} := z_q - z_p$, $t_{p,r} := z_r - z_p$ and by setting

$$n_{F_l} := \|t_{p,q} \times t_{p,r}\|_{\ell^2}^{-1} (t_{p,q} \times t_{p,r}). \tag{10.9}$$

Let us now construct the geometric mapping T_K for all $K \in \mathcal{T}_h$. Let z_p, z_q, z_r, z_s be the four vertices of K ordered by *increasing vertex-index*, i.e., $p < q < r < s$. We define T_K by setting

$$T_K(\widehat{z}_1) := z_p, \quad T_K(\widehat{z}_2) := z_q, \quad T_K(\widehat{z}_3) := z_r, \quad T_K(\widehat{z}_4) := z_s. \tag{10.10}$$

Hence, the global index of the mesh vertex $T_K(\widehat{z}_n)$ increases with n. Using the connectivity array j_cv defined by (8.12), we have j_cv$(m,1) = p$, j_cv$(m,2) = q$, j_cv$(m,3) = r$, and j_cv$(m,4) = s$, where m is the global enumeration index of the mesh cell K. Notice that (10.10) is sufficient to define T_K entirely since we assumed that the mesh is affine. We emphasize

that, in the present construction, the mapping $\boldsymbol{T}_K$ is invertible, but its Jacobian determinant can be positive or negative.

Example 10.7 (Orienting a tetrahedron). Consider a tetrahedron whose vertices have global indices 35, 42, 67, and 89 shown in Figure 10.3. The orientation of the (five visible) edges is materialized by dark arrows. The unit normal vector $\boldsymbol{n}_F$ defined by the increasing-vertex enumeration points toward the outside of the tetrahedron for the face defined by the indices $\{35, 42, 67\}$, and it points toward the inside of the tetrahedron for the face defined by the indices $\{42, 67, 89\}$, etc.   □
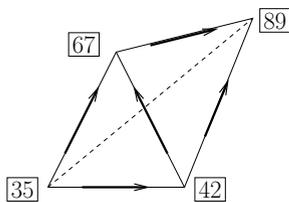
Fig. 10.3 Illustration of Example 10.7.

Theorem 10.8 (Simplicial mesh orientation). *Let $\mathcal{T}_h$ be a simplicial mesh. Let $\widehat{K}$ be oriented by using the increasing vertex-index enumeration. For all $K \in \mathcal{T}_h$, let $\boldsymbol{T}_K$ be defined by the increasing vertex-index convention* (10.10). *Then the orientation of $\mathcal{T}_h$ based on the increasing vertex-index enumeration is generation-compatible.*

Proof. (1) Let us prove (10.6a). Let E_l be an edge with vertices $\boldsymbol{z}_p, \boldsymbol{z}_q$, $p < q$. Let (m,n) be s.t. $E_l = \boldsymbol{T}_{K_m}(\widehat{E}_n)$, i.e., $\texttt{j\_ce}(m,n) = l$. Let $\widehat{\boldsymbol{z}}_i, \widehat{\boldsymbol{z}}_j$ with $i < j$ be the vertices of the edge $\widehat{E}_n$ of $\widehat{K}$. The increasing vertex-index convention (10.10) for the geometric mappings implies that $\boldsymbol{T}_{K_m}(\widehat{\boldsymbol{z}}_i) = \boldsymbol{z}_p$ and $\boldsymbol{T}_{K_m}(\widehat{\boldsymbol{z}}_j) = \boldsymbol{z}_q$. Moreover, the orientation for $\widehat{K}$ implies that $\widehat{\boldsymbol{\tau}}_{\widehat{E}_n} = \|\widehat{\boldsymbol{t}}_{i,j}\|_{\ell^2}^{-1} \widehat{\boldsymbol{t}}_{i,j}$ with $\widehat{\boldsymbol{t}}_{i,j} := \widehat{\boldsymbol{z}}_j - \widehat{\boldsymbol{z}}_i$, so that $\boldsymbol{\Phi}_{K_m}^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}_n}) = \|\mathbb{J}_{K_m}\widehat{\boldsymbol{\tau}}_{\widehat{E}_n}\|_{\ell^2}^{-1} \mathbb{J}_{K_m}\widehat{\boldsymbol{\tau}}_{\widehat{E}_n} = \|\mathbb{J}_{K_m}\widehat{\boldsymbol{t}}_{i,j}\|_{\ell^2}^{-1} \mathbb{J}_{K_m}\widehat{\boldsymbol{t}}_{i,j}$. Since $\boldsymbol{T}_{K_m}$ is affine, we have

$$\mathbb{J}_{K_m}\widehat{\boldsymbol{t}}_{i,j} = \boldsymbol{T}_{K_m}(\widehat{\boldsymbol{z}}_j) - \boldsymbol{T}_{K_m}(\widehat{\boldsymbol{z}}_i) = \boldsymbol{z}_q - \boldsymbol{z}_p = \boldsymbol{t}_{p,q},$$

and we conclude that $\boldsymbol{\Phi}_{K_m}^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}_n}) = \|\boldsymbol{t}_{p,q}\|_{\ell^2}^{-1}\boldsymbol{t}_{p,q} = \boldsymbol{\tau}_{E_l}$.
(2) Let us prove (10.6b) in dimension three. Let F_l be a face with vertices $\boldsymbol{z}_p, \boldsymbol{z}_q, \boldsymbol{z}_r$, $p < q < r$. Let (m,n) be s.t. $F_l = \boldsymbol{T}_{K_m}(\widehat{F}_n)$, i.e., $\texttt{j\_cf}(m,n) = l$. Let $\widehat{\boldsymbol{z}}_i, \widehat{\boldsymbol{z}}_j, \widehat{\boldsymbol{z}}_k$ with $i < j < k$ be the vertices of the face $\widehat{F}_n$ of $\widehat{K}$. Reasoning as above, we have $\mathbb{J}_{K_m}\widehat{\boldsymbol{t}}_{i,j} = \boldsymbol{t}_{p,q}$ and $\mathbb{J}_{K_m}\widehat{\boldsymbol{t}}_{i,k} = \boldsymbol{t}_{p,r}$. Using the identity $\mathbb{A}^{-\mathsf{T}}(\boldsymbol{x}{\times}\boldsymbol{y}) = \det(\mathbb{A})^{-1}(\mathbb{A}\boldsymbol{x}{\times}\mathbb{A}\boldsymbol{y})$ for every $3{\times}3$ invertible matrix $\mathbb{A}$ and all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ (see Exercise 9.5), we have

$$\mathbb{J}_{K_m}^{-\mathsf{T}}(\widehat{\boldsymbol{t}}_{i,j}{\times}\widehat{\boldsymbol{t}}_{i,k}) = \det(\mathbb{J}_{K_m})^{-1}(\boldsymbol{t}_{p,q}{\times}\boldsymbol{t}_{p,r}).$$

Moreover, since $\widehat{\boldsymbol{n}}_{\widehat{F}_n}$ and $\widehat{\boldsymbol{t}}_{i,j}\times\widehat{\boldsymbol{t}}_{i,k}$ are collinear and point in the same direction, the definition (9.14a) implies that

$$\boldsymbol{\Phi}^{\mathrm{d}}_{K_m}(\widehat{\boldsymbol{n}}_{\widehat{F}_n}) = \epsilon_{K_m}\|\mathbb{J}^{-\mathsf{T}}_{K_m}(\widehat{\boldsymbol{t}}_{i,j}\times\widehat{\boldsymbol{t}}_{i,k})\|^{-1}_{\ell^2}\mathbb{J}^{-\mathsf{T}}_{K_m}(\widehat{\boldsymbol{t}}_{i,j}\times\widehat{\boldsymbol{t}}_{i,k}).$$

Since $\|\mathbb{J}^{-\mathsf{T}}_{K_m}(\widehat{\boldsymbol{t}}_{i,j}\times\widehat{\boldsymbol{t}}_{i,k})\|_{\ell^2} = |\det(\mathbb{J}_{K_m})|^{-1}\|\boldsymbol{t}_{p,q}\times\boldsymbol{t}_{p,r}\|_{\ell^2}$, we conclude that

$$\begin{aligned}\boldsymbol{\Phi}^{\mathrm{d}}_{K_m}(\widehat{\boldsymbol{n}}_{\widehat{F}_n}) &= \epsilon_{K_m}|\det(\mathbb{J}_{K_m})|\|\boldsymbol{t}_{p,q}\times\boldsymbol{t}_{p,r}\|^{-1}_{\ell^2}\det(\mathbb{J}_{K_m})^{-1}(\boldsymbol{t}_{p,q}\times\boldsymbol{t}_{p,r}) \\ &= \|\boldsymbol{t}_{p,q}\times\boldsymbol{t}_{p,r}\|^{-1}_{\ell^2}(\boldsymbol{t}_{p,q}\times\boldsymbol{t}_{p,r}) = \boldsymbol{n}_{F_l}.\end{aligned}$$

(3) Finally, by Remark 10.5, the argument in Step (1) implies that (10.6b) holds true in dimension two.   □

Remark 10.9 (Positive Jacobian determinant). If one insists on building geometric mappings such that $\det(\mathbb{J}_K) > 0$, the above orientation of the edges and the faces of the mesh is still generation-compatible if one uses two reference tetrahedra; see Ainsworth and Coyle [6].   □

10.5 Quadrangular and hexahedral meshes

We state without proof a result by Agelek et al. [4] on quadrangular and hexahedral meshes.

Theorem 10.10 (Quad/Hex mesh orientation). *Let the reference square or cube be oriented using the increasing vertex-index enumeration technique. (i) Let $\mathcal{T}_h$ be a quadrangular mesh. It is possible to orient the mesh to make it generation-compatible. (ii) Let now $\mathcal{T}_h$ be a hexahedral mesh and let $\mathcal{T}_{\frac{h}{2}}$ be obtained from $\mathcal{T}_h$ by cutting each hexahedron into eight smaller hexahedra. It is possible to orient $\mathcal{T}_{\frac{h}{2}}$ to make it generation-compatible.*

Let us provide some further insight into this result. Let us start with the faces since orientating the faces is simple and independent of the space dimension. Consider the undirected graph whose vertices are the mesh faces and the edges are the mesh cells. We say that two mesh faces F_1, F_2 are connected through K iff F_1, F_2 are faces of K that are $\boldsymbol{T}_K$-parallel (i.e., images by $\boldsymbol{T}_K$ of faces of $\widehat{K}$ that are parallel). Since each face is connected to either one (boundary face) or two cells (interface), all the connected components of the graph thus constructed are either closed loops or chains whose extremities are boundary faces. In either case, the connected components of the graph realize a partition of the faces of $\mathcal{T}_h$. We then assign the same orientation to all the faces in the same connected component of the graph.

Let us now orient the edges. For quadrangular meshes, the edges are oriented by rotating clockwise the unit normal vector; see the second panel in

Figure 10.2 and the left panel of Figure 10.4 where the dashed lines connect the edges/faces that are in the same equivalence class. For hexahedral meshes, we further need to devise a specific orientation of the edges. Let $\mathcal{E}_h$ be the collection of the mesh edges. We say that two edges of a cell K are $\boldsymbol{T}_K$-parallel if they are images by $\boldsymbol{T}_K$ of edges in $\widehat{K}$ that are parallel. We then define a binary relation $\mathcal{R}$ on $\mathcal{E}_h$. Let $E, E' \in \mathcal{E}_h$ be two mesh edges. We say that $E \mathcal{R} E'$ if either E and E' belong to the same cell K and are $\boldsymbol{T}_K$-parallel or there is a collection of cells $K_1, \ldots, K_L$, all different, and a collection of edges $E =: E_1, \ldots, E_{L+1} := E'$ such that E_l and E_{l+1} are both edges of K_l, $l \in \{1{:}L\}$, and E_l, E_{l+1} are $\boldsymbol{T}_{K_l}$-parallel. This defines an equivalence relation over the edges which in turn generates a partition of $\mathcal{E}_h$. Unfortunately, it is not always possible to give the same orientation to all the edges belonging to the same equivalence class since in dimension three edges in the same equivalence class may actually be sitting on a Möbius strip. An example of nonorientable mesh (in the sense defined above) composed of hexahedra is shown in the right panel of Figure 10.4. Theorem 10.10 then says that after subdivision, this mesh becomes orientable in a generation-compatible way, and more generally, every mesh composed of hexahedra is orientable after one subdivision.
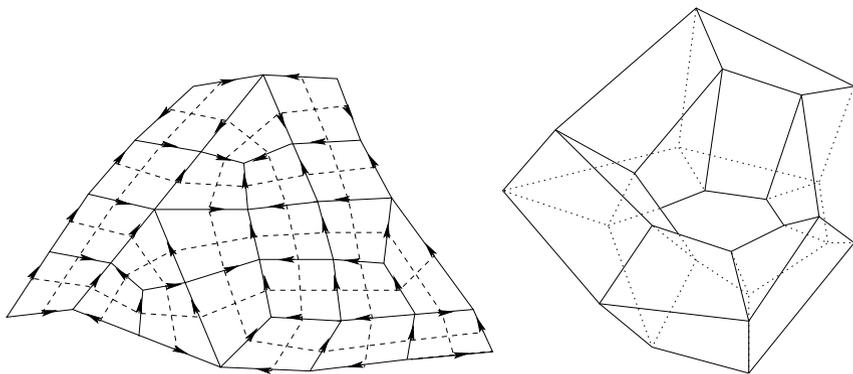
Fig. 10.4 Orientation of the edges in a mesh composed of quadrangles (left). Nonorientable three-dimensional mesh composed of hexahedra (right).

Assuming that the mesh edges have been oriented as discussed above, it is now possible to build the geometric mappings $\boldsymbol{T}_K$ such that the above mesh orientation is generation-compatible. The idea is that for each mesh cell K, there is only one vertex such that all the edges sharing it are oriented away from it. This vertex is called origin of the cell. Then we choose $\boldsymbol{T}_K$ such that $\boldsymbol{T}_K$ maps $\widehat{\boldsymbol{z}}_1$ to the origin of K (recall that $\widehat{\boldsymbol{z}}_1$ is the only vertex of $\widehat{K}$ such that all the edges sharing it are oriented away from it; see Figure 10.2). This choice implies that the image by $\boldsymbol{T}_K$ of $\widehat{\boldsymbol{z}}_4$ (if $d = 2$) and of $\widehat{\boldsymbol{z}}_8$ (if $d = 3$) is the vertex of K opposite to the origin. Finally, the image by $\boldsymbol{T}_K$ of the remaining

two (if $d = 2$) or six (if $d = 3$) vertices can be chosen arbitrarily. One criterion to limit the choices can be to fix a sign for $\det(\mathbb{J}_K)$. In dimension two, one choice gives a positive sign and the other gives a negative sign, whereas in dimension three, three choices give a positive sign and three choices give a negative sign.

Exercises

Exercise 10.1 (Faces in 2D). Let $\boldsymbol{R}_{\frac{\pi}{2}}$ be the rotation of angle $\frac{\pi}{2}$ in $\mathbb{R}^2$. (i) Let $\mathbb{A}$ be an inversible 2×2 matrix. Prove that $\mathbb{A}^{-\mathsf{T}}\boldsymbol{R}_{\frac{\pi}{2}} = \frac{1}{\det(\mathbb{A})}\boldsymbol{R}_{\frac{\pi}{2}}\mathbb{A}$. (ii) Prove that $\boldsymbol{\Phi}_K^{\mathrm{d}}(\boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{z})) = \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{\Phi}_K^{\mathrm{c}}(\boldsymbol{z}))$ for all $\boldsymbol{z} \in \mathbb{R}^2$.

Exercise 10.2 (Connectivity arrays j_cv, j_ce). Consider the mesh shown in Figure 10.5, where the face enumeration is identified with large circles and the cell enumeration is identified with squares. (i) Write the connectivity arrays j_cv and j_ce based on increasing vertex-index enumeration. (ii) Give the sign of the determinant of the Jacobian matrix of $\boldsymbol{T}_K$ for each triangle.
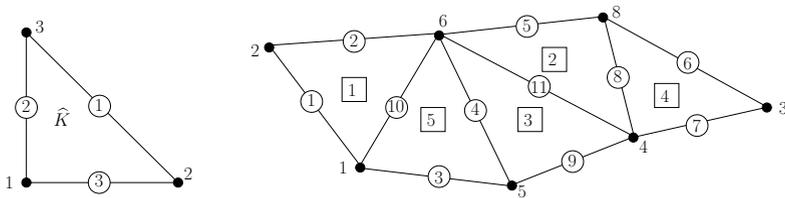
Fig. 10.5 Illustration for Exercise 10.2.

Exercise 10.3 (Connectivity array j_geo). Consider the mesh shown in Figure 10.6 and based on the $\mathbb{P}_{2,2}$ geometric Lagrange element.
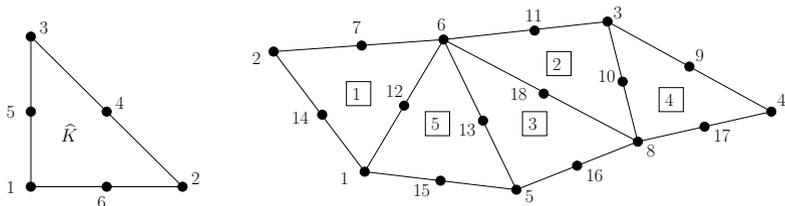
Fig. 10.6 Illustration for Exercise 10.3.

(i) Write the connectivity array j_geo based on increasing vertex-index enumeration. (ii) Give the sign of the determinant of the Jacobian matrix of T_K for each triangle.

Exercise 10.4 (Orientation of quadrangular mesh). (i) Using the enumeration and the orientation conventions proposed in this chapter, orient the mesh shown in Figure 10.7, where the cell enumeration is identified with shaded rectangles. (ii) Give the connectivity array j_geo so that the mesh orientation is generation-compatible and the determinant of the Jacobian matrix of T_K is positive for even quadrangles and negative for odd quadrangles.
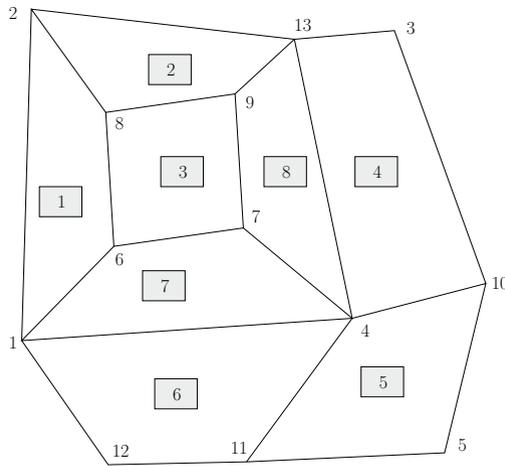
Fig. 10.7 Illustration for Exercise 10.4.

Exercise 10.5 (Mesh extrusion). (i) Let K be a triangular prism. Denote by e_3 the unit vector in the vertical direction. Let z_1, z_2, z_3 be the three vertices of the bottom triangular face of K, and let z_4, z_5, z_6 be the three vertices of its top triangular face, so that the segments $[z_p, z_{p+3}]$ are parallel to e_3 for every $p \in \{1, 2, 3\}$. Propose a way to cut K into three tetrahedra. (ii) Let $\mathcal{T}_h$ be a two-dimensional oriented mesh composed of triangles. Let $\mathcal{T}_h'$ be a copy of $\mathcal{T}_h$ obtained by translating $\mathcal{T}_h$ in the third direction e_3, say $\mathcal{T}_h' := \mathcal{T}_h + e_3$. Propose a way to cut all the prisms thus formed to make a matching mesh composed of tetrahedra.

Part III, Chapter 11

Local interpolation on affine meshes

Check for updates

We have seen in the previous chapter how to build finite elements and local interpolation operators in each cell K of a mesh $\mathcal{T}_h$. In this chapter, we analyze the local interpolation error for smooth $\mathbb{R}^q$-valued functions, $q \geq 1$. We restrict the material to affine meshes and to transformations ψ_K s.t.

$$\psi_K(v) = \mathbb{A}_K(v \circ \boldsymbol{T}_K), \tag{11.1}$$

where $\mathbb{A}_K$ is a matrix in $\mathbb{R}^{q \times q}$. Nonaffine meshes are treated in Chapter 13. We introduce the notion of shape-regular families of affine meshes, we study the transformation of Sobolev norms using (11.1), and we present important approximation results collectively known as the Bramble–Hilbert lemmas. We finally prove the main result of this chapter, which is an upper bound on the local interpolation error over each mesh cell for smooth functions.

11.1 Shape-regularity for affine meshes

Let $\mathcal{T}_h$ be an affine mesh. Let $K \in \mathcal{T}_h$. Since the geometric mapping $\boldsymbol{T}_K$ is affine, its Jacobian matrix $\mathbb{J}_K \in \mathbb{R}^{d \times d}$ defined in (8.3) is such that

$$\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) - \boldsymbol{T}_K(\widehat{\boldsymbol{y}}) = \mathbb{J}_K(\widehat{\boldsymbol{x}} - \widehat{\boldsymbol{y}}), \qquad \forall \widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}} \in \widehat{K}. \tag{11.2}$$

The matrix $\mathbb{J}_K$ is invertible since the mapping $\boldsymbol{T}_K$ is bijective. Moreover, the (Fréchet) derivative of the geometric mapping is such that $D\boldsymbol{T}_K(\widehat{\boldsymbol{x}})(\widehat{\boldsymbol{h}}) = \mathbb{J}_K \widehat{\boldsymbol{h}}$ for all $\widehat{\boldsymbol{h}} \in \mathbb{R}^d$ (see Appendix B). We denote the Euclidean norm in $\mathbb{R}^d$ by $\|\cdot\|_{\ell^2(\mathbb{R}^d)}$, or $\|\cdot\|_{\ell^2}$ when the context is unambiguous. We abuse the notation by using the same symbol for the induced matrix norm.

Lemma 11.1 (Bound on $\mathbb{J}_K$). *Let $\mathcal{T}_h$ be an affine mesh and let $K \in \mathcal{T}_h$. Let ρ_K be the diameter of the largest ball that can be inscribed in K and let*

h_K be the diameter of K, as shown in Figure 11.1. Let $\widehat{\rho}_{\widehat{K}}$ and $\widehat{h}_{\widehat{K}}$ be defined similarly. The following holds true:

$$|\det(\mathbb{J}_K)| = \frac{|K|}{|\widehat{K}|}, \qquad \|\mathbb{J}_K\|_{\ell^2} \leq \frac{h_K}{\rho_{\widehat{K}}}, \qquad \|\mathbb{J}_K^{-1}\|_{\ell^2} \leq \frac{h_{\widehat{K}}}{\rho_K}. \tag{11.3}$$

Proof. The first equality results from the fact that

$$|K| = \int_K \mathrm{d}x = \int_{\widehat{K}} |\det(\mathbb{J}_K)| \, \mathrm{d}\widehat{x} = |\det(\mathbb{J}_K)| \, |\widehat{K}|.$$

Regarding the bound on $\|\mathbb{J}_K\|_{\ell^2}$, we observe that

$$\|\mathbb{J}_K\|_{\ell^2} = \sup_{\widehat{\boldsymbol{h}} \neq 0} \frac{\|\mathbb{J}_K \widehat{\boldsymbol{h}}\|_{\ell^2}}{\|\widehat{\boldsymbol{h}}\|_{\ell^2}} = \frac{1}{\rho_{\widehat{K}}} \sup_{\|\widehat{\boldsymbol{h}}\|_{\ell^2} = \rho_{\widehat{K}}} \|\mathbb{J}_K \widehat{\boldsymbol{h}}\|_{\ell^2}.$$

Any $\widehat{\boldsymbol{h}} \in \mathbb{R}^d$ such that $\|\widehat{\boldsymbol{h}}\|_{\ell^2} = \rho_{\widehat{K}}$ can be written as $\widehat{\boldsymbol{h}} = \widehat{\boldsymbol{x}}_1 - \widehat{\boldsymbol{x}}_2$ with $\widehat{\boldsymbol{x}}_1, \widehat{\boldsymbol{x}}_2 \in \widehat{K}$. We infer that $\mathbb{J}_K \widehat{\boldsymbol{h}} = \boldsymbol{T}_K(\widehat{\boldsymbol{x}}_1) - \boldsymbol{T}_K(\widehat{\boldsymbol{x}}_2) = \boldsymbol{x}_1 - \boldsymbol{x}_2$, which in turn proves that $\|\mathbb{J}_K \widehat{\boldsymbol{h}}\|_{\ell^2} \leq h_K$. This establishes the bound on $\|\mathbb{J}_K\|_{\ell^2}$. The bound on $\|\mathbb{J}_K^{-1}\|_{\ell^2}$ is obtained by exchanging the roles of K and $\widehat{K}$. $\qquad\square$
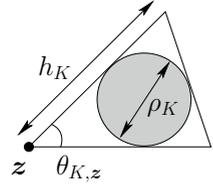
Fig. 11.1 Triangular cell K with vertex z, angle $\theta_{K,z}$, and largest inscribed ball.

Since the analysis of the interpolation error (implicitly) invokes sequences of successively refined meshes, we henceforth denote by $(\mathcal{T}_h)_{h \in \mathcal{H}}$ a sequence of meshes discretizing a domain D in $\mathbb{R}^d$, where the index h takes values in a countable set $\mathcal{H}$ having zero as the only accumulation point.

Definition 11.2 (Shape-regularity). *A sequence of affine meshes* $(\mathcal{T}_h)_{h \in \mathcal{H}}$ *is said to be* shape-regular *if there is* $\sigma_\sharp$ *such that*

$$\sigma_K := \frac{h_K}{\rho_K} \leq \sigma_\sharp, \qquad \forall K \in \mathcal{T}_h, \ \forall h \in \mathcal{H}. \tag{11.4}$$

Occasionally, when the context is unambiguous, we will say that $(\mathcal{T}_h)_{h \in \mathcal{H}}$ is regular instead of shape-regular. Owing to Lemma 11.1, a shape-regular sequence of affine meshes satisfies

$$\|\mathbb{J}_K\|_{\ell^2} \|\mathbb{J}_K^{-1}\|_{\ell^2} \leq \sigma_\sharp \sigma_{\widehat{K}}, \qquad \forall K \in \mathcal{T}_h, \ \forall h \in \mathcal{H}. \tag{11.5}$$

Example 11.3 (Dimension 1). Every sequence of one-dimensional meshes is shape-regular since $h_K = \rho_K$ when $d = 1$.   □

Example 11.4 (Triangulations). A shape-regular sequence of affine triangulations can be obtained from an initial triangulation by connecting all the edge midpoints and repeating this procedure as many times as needed.   □

Remark 11.5 (Angles). Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular sequence of affine simplicial meshes. Assume that $d = 2$, let K be a triangle in $\mathcal{T}_h$ and let $\boldsymbol{z}$ be a vertex of K. Then the angle $\theta_{K,z} \in (0, 2\pi)$ formed by the two edges of K sharing $\boldsymbol{z}$ is uniformly bounded away from zero. Indeed, the angular sector centered at $\boldsymbol{z}$ of angle $\theta_{K,z}$ and radius h_K covers the ball of diameter ρ_K that is inscribed in K (see Figure 11.1). Hence, $\frac{1}{2}h_K^2 \theta_{K,z} \geq \frac{1}{4}\pi \rho_K^2$, which in turn implies that $\theta_{K,z} \geq \frac{1}{2}\pi \sigma_\sharp^{-2}$. Assume now that $d = 3$, let K be a tetrahedron, and let $\boldsymbol{z}$ be a vertex of K. Then the solid angle $\omega_{K,z} \in (0, 4\pi)$ formed by the three faces of K sharing $\boldsymbol{z}$ is uniformly bounded away from zero. Reasoning as above, with volumes instead of surfaces, leads to $\frac{1}{3}h_K^3 \omega_{K,z} \geq \frac{1}{6}\pi \rho_K^3$, so that $\omega_{K,z} \geq \frac{1}{2}\pi \sigma_\sharp^{-3}$.   □

We close this section with a useful result on matching meshes. Recall from §8.2 the notion of mesh faces, edges, and vertices in a matching mesh (assuming $d = 3$). For every mesh vertex $\boldsymbol{z} \in \mathcal{V}_h$, we denote

$$\mathcal{T}_{\boldsymbol{z}} := \{K \in \mathcal{T}_h \mid \boldsymbol{z} \in K\} \tag{11.6}$$

the collection of the mesh cells sharing $\boldsymbol{z}$. Similarly, recall from (10.3) that for every mesh edge $E \in \mathcal{E}_h$ and every mesh face $F \in \mathcal{F}_h$, $\mathcal{T}_E := \{K \in \mathcal{T}_h \mid E \subset K\}$ and $\mathcal{T}_F := \{K \in \mathcal{T}_h \mid F \subset K\}$ are the collection of the mesh cells sharing E and F, respectively.

Proposition 11.6 (Neighboring cells). *Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular sequence of matching affine meshes. Then the cardinality of the set $\mathcal{T}_{\boldsymbol{z}}$ is uniformly bounded for all $\boldsymbol{z} \in \mathcal{V}_h$ and all $h \in \mathcal{H}$, and the sizes of all the cells in $\mathcal{T}_{\boldsymbol{z}}$ are uniformly equivalent w.r.t. $h \in \mathcal{H}$. The same assertion holds true for the sets $\mathcal{T}_E$ and $\mathcal{T}_F$.*

Proof. It suffices to prove the assertions for $\mathcal{T}_{\boldsymbol{z}}$. The bound on card($\mathcal{T}_{\boldsymbol{z}}$) follows from Remark 11.5. Concerning the sizes of the cells in $\mathcal{T}_{\boldsymbol{z}}$, we first observe that if $K', K'' \in \mathcal{T}_{\boldsymbol{z}}$, $K' \neq K''$, share a common face, say F with diameter h_F, then $h_{K'} \leq \sigma_\sharp \rho_{K'} \leq \sigma_\sharp h_F \leq \sigma_\sharp h_{K''}$, and similarly, $h_{K''} \leq \sigma_\sharp h_{K'}$. This shows that the sizes of K' and K'' are uniformly equivalent. Now, for all K' and K'' in $\mathcal{T}_{\boldsymbol{z}}$, there is a finite path of cells linking K' to K'' s.t. any two consecutive mesh cells in the path share a common face. The number of cells composing the path cannot exceed card($\mathcal{T}_{\boldsymbol{z}}$), so that it is uniformly bounded. Hence, the sizes of K' and K'' are uniformly equivalent.   □

11.2 Transformation of Sobolev seminorms

The question we investigate now is the following: given a function $v \in W^{m,p}(K;\mathbb{R}^q)$, how does the seminorm of $\psi_K(v)$ in $W^{m,p}(\widehat{K};\mathbb{R}^q)$ compare to that of v in $W^{m,p}(K;\mathbb{R}^q)$ with ψ_K is defined in (11.1)?

Lemma 11.7 (Norm scaling by ψ_K). *Let $\mathcal{T}_h$ be an affine mesh. Let $s \in [0,\infty)$ and $p \in [1,\infty]$ (with $z^{\pm\frac{1}{p}} := 1$, $\forall z > 0$ if $p = \infty$). There exists c, depending only on s and d, such that the following bounds hold true for all $v \in W^{s,p}(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$:*

$$|\psi_K(v)|_{W^{s,p}(\widehat{K};\mathbb{R}^q)} \leq c\,\gamma_K^{\frac{1}{p}}\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{J}_K\|_{\ell^2}^s\,|\mathrm{det}(\mathbb{J}_K)|^{-\frac{1}{p}}\,|v|_{W^{s,p}(K;\mathbb{R}^q)}, \quad (11.7a)$$

$$|v|_{W^{s,p}(K;\mathbb{R}^q)} \leq c\,\delta_K^{\frac{1}{p}}\|\mathbb{A}_K^{-1}\|_{\ell^2}\|\mathbb{J}_K^{-1}\|_{\ell^2}^s\,|\mathrm{det}(\mathbb{J}_K)|^{\frac{1}{p}}\,|\psi_K(v)|_{W^{s,p}(\widehat{K};\mathbb{R}^q)}, \quad (11.7b)$$

where $\gamma_K = \delta_K := 1$ if $s \in \mathbb{N}$ and $\gamma_K := |\mathrm{det}(\mathbb{J}_K)|^{-1}\|\mathbb{J}_K\|_{\ell^2}^d$, $\delta_K := |\mathrm{det}(\mathbb{J}_K)|\,\|\mathbb{J}_K^{-1}\|_{\ell^2}^d$ otherwise (the real numbers γ_K and δ_K are uniformly bounded w.r.t. $h \in \mathcal{H}$ on shape-regular mesh sequences).

Proof. We start by assuming $s = m \in \mathbb{N}$. The bounds are obvious for $m = 0$. For $m \geq 1$, let α be a multi-index with length $|\alpha| = m$, i.e., $\alpha := (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$ with $\alpha_1 + \ldots + \alpha_d = m$. Let $\widehat{\boldsymbol{x}} \in \widehat{K}$. Owing to (B.6), we infer that

$$\partial^\alpha(\psi_K(v))(\widehat{\boldsymbol{x}}) = \mathbb{A}_K D^m(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})(\underbrace{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_1}_{\alpha_1 \text{times}}, \ldots, \underbrace{\boldsymbol{e}_d, \ldots, \boldsymbol{e}_d}_{\alpha_d \text{times}}),$$

where $D^m(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})$ is the m-th Fréchet derivative of $v \circ \boldsymbol{T}_K$ at $\widehat{\boldsymbol{x}}$ and $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ is the canonical Cartesian basis of $\mathbb{R}^d$. We now apply the chain rule (see Lemma B.4) to $v \circ \boldsymbol{T}_K$. Since $\boldsymbol{T}_K$ is affine, the Fréchet derivative of $\boldsymbol{T}_K$ is independent of $\widehat{\boldsymbol{x}}$ and its higher-order Fréchet derivatives vanish. Hence, we have

$$D^m(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})(\boldsymbol{h}_1, \ldots, \boldsymbol{h}_m) =$$
$$\sum_{\sigma \in \mathcal{S}_m} \frac{1}{m!}(D^m v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))(D\boldsymbol{T}_K(\boldsymbol{h}_{\sigma(1)}), \ldots, D\boldsymbol{T}_K(\boldsymbol{h}_{\sigma(m)})),$$

for all $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_m \in \mathbb{R}^d$, where $\mathcal{S}_m$ is the set of the permutations of $\{1{:}m\}$. Since $D\boldsymbol{T}_K(\boldsymbol{h}) = \mathbb{J}_K\boldsymbol{h}$ for all $\boldsymbol{h} \in \mathbb{R}^d$ owing to (11.2), we infer that

$$|\partial^\alpha(v \circ \boldsymbol{T}_K)(\widehat{\boldsymbol{x}})| \leq \|\mathbb{J}_K\|_{\ell^2}^m\,\|(D^m v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\|_{\mathcal{M}_m(\mathbb{R}^d,\ldots,\mathbb{R}^d;\mathbb{R}^q)},$$

with $\|A\|_{\mathcal{M}_m(\mathbb{R}^d,\ldots,\mathbb{R}^d;\mathbb{R}^q)} := \sup_{(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) \in \mathbb{R}^d \times \ldots \times \mathbb{R}^d} \frac{\|A(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m)\|_{\ell^2}}{\|\boldsymbol{y}_1\|_{\ell^2}\ldots\|\boldsymbol{y}_m\|_{\ell^2}}$ for every multilinear map $A \in \mathcal{M}_m(\mathbb{R}^d, \ldots, \mathbb{R}^d; \mathbb{R}^q)$. Owing to the multilinearity of $D^m v$ and using again (B.6), we infer that (see Exercise 11.1)

$$\|(D^m v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\|_{\mathcal{M}_m(\mathbb{R}^d,\dots,\mathbb{R}^d;\mathbb{R}^q)} \leq c \sum_{|\beta|=m} \|(\partial^\beta v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\|_{\ell^2},$$

where c only depends on m and d. As a result, we have

$$\|\partial^\alpha(\psi_K(v))(\widehat{\boldsymbol{x}})\|_{\ell^2} \leq c\,\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{J}_K\|_{\ell^2}^m \sum_{|\beta|=m} \|(\partial^\beta v)(\boldsymbol{T}_K(\widehat{\boldsymbol{x}}))\|_{\ell^2},$$

and (11.7a) follows by taking the $L^p(\widehat{K})$-norm on both sides of the inequality. The proof of (11.7b) is similar. We refer to Exercise 11.7 when $s \notin \mathbb{N}$. $\square$

Remark 11.8 (Seminorms). The upper bounds in (11.7a) and (11.7b) involve only seminorms because the geometric mappings are affine. $\square$

11.3 Bramble–Hilbert lemmas

This section contains an important result for the analysis of the approximation properties of finite elements. We consider scalar-valued functions. The result extends to vector-valued functions by reasoning componentwise.

Lemma 11.9 ($\mathbb{P}_k$-Bramble–Hilbert/Deny–Lions). *Let S be a Lipschitz domain in $\mathbb{R}^d$. Let $p \in [1, \infty]$. Let $k \in \mathbb{N}$. There is c (depending on k, p, S) s.t. for all $v \in W^{k+1,p}(S)$,*

$$\inf_{q \in \mathbb{P}_{k,d}} \|v - q\|_{W^{k+1,p}(S)} \leq c\,|v|_{W^{k+1,p}(S)}. \tag{11.8}$$

Proof. (1) Consider the bounded linear forms $f_\alpha : W^{k+1,p}(S) \to \mathbb{R}$ s.t.

$$f_\alpha(v) := \ell_S^{|\alpha|-d} \int_S \partial^\alpha v \, \mathrm{d}x, \qquad \forall \alpha \in \mathcal{A}_{k,d},$$

where $\mathcal{A}_{k,d} := \{\alpha := (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d \mid |\alpha| \leq k\}$ and $\ell_S := \mathrm{diam}(S)$ (the factor $\ell_S^{|\alpha|-d}$ is introduced for dimensional consistency). Let us set $N_{k,d} := \mathrm{card}(\mathcal{A}_{k,d}) = \binom{k+d}{d}$. Let us consider the map $\Phi_{k,d} : W^{k+1,p}(S) \to \mathbb{R}^{N_{k,d}}$ s.t.

$$\Phi_{k,d}(q) := (f_\alpha(q))_{\alpha \in \mathcal{A}_{k,d}},$$

and let us prove that the restriction of this map to $\mathbb{P}_{k,d}$ is an isomorphism. To prove this, we observe that $\dim(\mathbb{P}_{k,d}) = N_{k,d}$, so that it is sufficient to prove that $\Phi_{k,d}$ is injective, which we do by induction on k. For $k = 0$, if $q \in \mathbb{P}_0$ satisfies $\Phi_{0,d}(q) = 0$, then $\int_S q \, \mathrm{d}x = q|S| = 0$ so that $q = 0$. Let us assume now that $k \geq 1$ and let $q \in \mathbb{P}_{k,d}$ be such that $\Phi_{k,d}(q) = 0$. Let us write $q(\boldsymbol{x}) = \sum_{\alpha \in \mathcal{A}_{k,d}} a_\alpha \boldsymbol{x}^\alpha$. Whenever $|\alpha| = k$, we obtain $\partial^\alpha q(\boldsymbol{x}) = a_\alpha \alpha_1! \dots \alpha_d!$ so that $f_\alpha(q) = 0$ implies that $a_\alpha = 0$. Since this property is satisfied for all α

such that $|\alpha| = k$, we infer that $q \in \mathbb{P}_{k-1,d}$ and conclude from the induction assumption that $q = 0$.

(2) Let us prove that there is $c > 0$, depending on S, k, and p, such that

$$c\, \|v\|_{W^{k+1,p}(S)} \le \ell_S^{k+1} |v|_{W^{k+1,p}(S)} + \ell_S^{\frac{d}{p}} \|\Phi_{k,d}(v)\|_{\ell^1(\mathbb{R}^{N_{k,d}})}, \tag{11.9}$$

for all $v \in W^{k+1,p}(S)$, with $\|f\|_{\ell^1(\mathbb{R}^{N_{k,d}})} := \sum_{\alpha \in N_{k,d}} |f_\alpha|$. Reasoning by contradiction, let $(v_n)_{n \in \mathbb{N}}$ be a sequence s.t.

$$\|v_n\|_{W^{k+1,p}(S)} = 1, \quad \lim_{n \to \infty} |v_n|_{W^{k+1,p}(S)} = 0, \quad \lim_{n \to \infty} \Phi_{k,d}(v_n) = 0. \tag{11.10}$$

Owing to the Rellich–Kondrachov theorem (Theorem 2.35), we infer that, up to a subsequence (not renumbered for simplicity), the sequence $(v_n)_{n \in \mathbb{N}}$ converges strongly to a function v in $W^{k,p}(S)$. Moreover, $(v_n)_{n \in \mathbb{N}}$ is a Cauchy sequence in $W^{k+1,p}(S)$ since

$$\|v_n - v_m\|_{W^{k+1,p}(S)} \le \|v_n - v_m\|_{W^{k,p}(S)} + \ell_S^{k+1} |v_n - v_m|_{W^{k+1,p}(S)},$$

and $|v_n - v_m|_{W^{k+1,p}(S)} \to 0$ by assumption. Hence, $(v_n)_{n \in \mathbb{N}}$ converges to v strongly in $W^{k+1,p}(S)$ (that the limit is indeed v comes from the uniqueness of the limit in $W^{k,p}(S)$). Owing to (11.10), we infer that $\|v\|_{W^{k+1,p}(S)} = 1$, $|v|_{W^{k+1,p}(S)} = 0$, and $\Phi_{k,d}(v) = 0$. Repeated applications of Lemma 2.11 (stating that in an open connected set S, $\nabla v = 0$ implies that v is constant on S) allow us to infer from $|v|_{W^{k+1,p}(S)} = 0$ that $v \in \mathbb{P}_{k,d}$. Since we have established in Step (1) that the restriction of $\Phi_{k,d}$ to $\mathbb{P}_{k,d}$ is an isomorphism, this yields $v = 0$, which contradicts $\|v\|_{W^{k+1,p}(S)} = 1$.

(3) Let $v \in W^{k+1,p}(S)$ and define $\pi(v) \in \mathbb{P}_{k,d}$ such that $\Phi_{k,d}(\pi(v)) = \Phi_{k,d}(v)$. This is possible since the restriction of $\Phi_{k,d}$ to $\mathbb{P}_{k,d}$ is an isomorphism. Then

$$c \inf_{q \in \mathbb{P}_{k,d}} \|v - q\|_{W^{k+1,p}(S)} \le c\|v - \pi(v)\|_{W^{k+1,p}(S)}$$

$$\le \ell_S^{k+1} |v - \pi(v)|_{W^{k+1,p}(S)} + \|\Phi_{k,d}(v - \pi(v))\|_{\ell^1(\mathbb{R}^{N_{k,d}})} = \ell_S^{k+1} |v|_{W^{k+1,p}(S)},$$

since $\partial^\alpha \pi(v) = 0$ for all $\alpha \in \mathbb{N}^d$ such that $|\alpha| = k + 1$. $\qquad \square$

Remark 11.10 (Peetre–Tartar lemma). Step (2) in the above proof is similar to the Peetre–Tartar lemma (Lemma A.20). Define $X := W^{k+1,p}(S)$, $Y := [L^p(D)]^{N_{k+1,d} - N_{k,d}} \times \mathbb{R}^{N_{k,d}}$, $Z := W^{k,p}(S)$, and the operator

$$A : X \ni v \longmapsto ((\partial^\alpha v)_{|\alpha|=k+1}, \Phi_{k,d}(v)) \in Y.$$

Since A is bounded and injective, and the embedding $X \hookrightarrow Z$ is compact, the property (11.9) results from the Peetre–Tartar lemma. $\qquad \square$

Corollary 11.11 ($\mathbb{P}_k$-Bramble–Hilbert for linear functionals). *Under the hypotheses of Lemma 11.9, there is c s.t. the following holds true for all*

$f \in (W^{k+1,p}(S))' := \mathcal{L}(W^{k+1,p}(S); \mathbb{R})$ *vanishing on* $\mathbb{P}_{k,d}$,

$$|f(v)| \le c \, \|f\|_{(W^{k+1,p}(S))'} \ell_S^{k+1} |v|_{W^{k+1,p}(S)}, \qquad \forall v \in W^{k+1,p}(S). \quad (11.11)$$

Proof. Left as an exercise.   □

Remark 11.12 (Literature). The estimate (11.8) is proved in Bramble and Hilbert [40, Thm. 1] and in Ciarlet and Raviart [79, Lem. 7]; see also Deny and Lions [90]. The estimate (11.11) is proved in Bramble and Hilbert [40, Thm. 2] and in Ciarlet and Raviart [79, Lem. 6]. There is some variability in the literature regarding the terminology for these results. For instance, Lemma 11.9 is called Bramble–Hilbert lemma in Brenner and Scott [47, Lem. 4.3.8] and Ciarlet and Raviart [78, p. 219], whereas it is called Deny–Lions lemma in Ciarlet [77, p. 111], and it is not given any name in Braess [39, p. 77]. Corollary 11.11 is called Bramble–Hilbert lemma in Ciarlet [77, p. 192] and Braess [39, p. 78]. Incidentally, there are two additional results that are the counterparts of Lemma 11.9 and Corollary 11.11 for $\mathbb{Q}_{k,d}$ polynomials; see Lemma 13.8 and Corollary 13.9.   □

11.4 Local finite element interpolation

This section contains our main result on local finite element interpolation. Recall the construction of §9.1 to generate a finite element and a local interpolation operator in each mesh cell $K \in \mathcal{T}_h$. Our goal is now to estimate the interpolation error $v - \mathcal{I}_K(v)$ for every smooth function v. The key point is that we want this bound to depend on K only through its size h_K under the assumption that the mesh sequence is shape-regular. The Bramble–Hilbert/Deny–Lions lemma cannot be used directly on K since this would give a constant depending on the shape of K. The crucial idea is then to use the fact that $\mathcal{I}_K = \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}} \circ \psi_K$ owing to Proposition 9.3 and to apply Lemma 11.9 on the fixed reference cell $\widehat{K}$.

Theorem 11.13 (Local interpolation). *Let $\widehat{P}$ be finite-dimensional, $\mathcal{I}_{\widehat{K}} \in \mathcal{L}(V(\widehat{K}); \widehat{P})$, $p \in [1, \infty]$, $k, l \in \mathbb{N}$, and assume that the following holds true:*

(i) *$[\mathbb{P}_{k,d}]^q \subset \widehat{P} \subset W^{k+1,p}(\widehat{K}; \mathbb{R}^q)$.*
(ii) *$[\mathbb{P}_{k,d}]^q$ is pointwise invariant under $\mathcal{I}_{\widehat{K}}$.*
(iii) *$W^{l,p}(\widehat{K}; \mathbb{R}^q) \hookrightarrow V(\widehat{K})$, i.e., $\|\widehat{v}\|_{V(\widehat{K})} \le \widehat{c} \, \|\widehat{v}\|_{W^{l,p}(\widehat{K}; \mathbb{R}^q)}$ for all $\widehat{v} \in V(\widehat{K})$.*

Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular sequence of affine meshes, let the transformation ψ_K be defined in (11.1) for all $K \in \mathcal{T}_h$, and assume that there is γ s.t. for all $K \in \mathcal{T}_h$ and all $h \in \mathcal{H}$,

$$\|\mathbb{A}_K\|_{\ell^2} \|\mathbb{A}_K^{-1}\|_{\ell^2} \le \gamma \, \|\mathbb{J}_K\|_{\ell^2} \|\mathbb{J}_K^{-1}\|_{\ell^2}. \quad (11.12)$$

Define the operator

$$\mathcal{I}_K := \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}} \circ \psi_K. \tag{11.13}$$

There is c s.t. the following local interpolation error estimates hold true:
(i) *If $l \le k+1$, then for every integers $r \in \{l{:}k+1\}$ and $m \in \{0{:}r\}$, all $v \in W^{r,p}(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K;\mathbb{R}^q)} \le c\, h_K^{r-m} |v|_{W^{r,p}(K;\mathbb{R}^q)}. \tag{11.14}$$

(ii) *If $l > k+1$, then for every integer $m \in \{0{:}k+1\}$, all $v \in W^{l,p}(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K;\mathbb{R}^q)} \le c \sum_{n \in \{k+1{:}l\}} h_K^{n-m} |v|_{W^{n,p}(K;\mathbb{R}^q)}. \tag{11.15}$$

Proof. We present a unified proof of (11.14) and (11.15). Let

$$\bar{r} \in \{l{:} \max(l, k+1)\}, \quad \underline{r} = \min(\bar{r}, k+1), \quad m \in \{0{:}\underline{r}\}.$$

If $l \le k+1$, then $\bar{r} \in \{l{:}k+1\}$, $\underline{r} = \bar{r} =: r$, $m \in \{0{:}r\}$, whereas if $l > k+1$, then $\bar{r} = l$, $\underline{r} = k+1$, $m \in \{0{:}k+1\}$. Thus, proving (11.14) and (11.15) is equivalent to prove that for all $v \in W^{\bar{r},p}(K;\mathbb{R}^q)$,

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K;\mathbb{R}^q)} \le c \sum_{n \in \{\underline{r}{:}\bar{r}\}} h_K^{n-m} |v|_{W^{n,p}(K;\mathbb{R}^q)}.$$

Let c denote a generic constant whose value can change at each occurrence as long as it is independent of v, K, and h. We take $\ell_{\widehat{K}} := 1$.
(1) For all $\widehat{v} \in W^{\bar{r},p}(\widehat{K};\mathbb{R}^q)$, we set $\mathcal{G}(\widehat{v}) := \widehat{v} - \mathcal{I}_{\widehat{K}}(\widehat{v})$. Since all the norms are equivalent in $\widehat{P}$, there is a constant $c_{\widehat{P}}$ such that

$$\|\widehat{p}\|_{W^{m,p}(\widehat{K};\mathbb{R}^q)} \le c_{\widehat{P}} \|\widehat{p}\|_{V(\widehat{K})}, \qquad \forall \widehat{p} \in \widehat{P}.$$

Using $m \le \underline{r} \le \bar{r}$, the above bound applied to $\widehat{p} := \mathcal{I}_{\widehat{K}}(\widehat{v})$, $\mathcal{I}_{\widehat{K}} \in \mathcal{L}(V(\widehat{K}))$, and Assumption (iii), we infer that

$$\begin{aligned}
\|\mathcal{G}(\widehat{v})\|_{W^{m,p}(\widehat{K};\mathbb{R}^q)} &\le \|\widehat{v}\|_{W^{m,p}(\widehat{K};\mathbb{R}^q)} + \|\mathcal{I}_{\widehat{K}}(\widehat{v})\|_{W^{m,p}(\widehat{K};\mathbb{R}^q)} \\
&\le \|\widehat{v}\|_{W^{\bar{r},p}(\widehat{K};\mathbb{R}^q)} + c_{\widehat{P}} \|\mathcal{I}_{\widehat{K}}(\widehat{v})\|_{V(\widehat{K})} \\
&\le \|\widehat{v}\|_{W^{\bar{r},p}(\widehat{K};\mathbb{R}^q)} + c_{\widehat{P}} \|\mathcal{I}_{\widehat{K}}\|_{\mathcal{L}(V(\widehat{K}))} \|\widehat{v}\|_{V(\widehat{K})} \\
&\le \|\widehat{v}\|_{W^{\bar{r},p}(\widehat{K};\mathbb{R}^q)} + c_{\widehat{P}} \|\mathcal{I}_{\widehat{K}}\|_{\mathcal{L}(V(\widehat{K}))} \widehat{c}\, \|\widehat{v}\|_{W^{l,p}(\widehat{K};\mathbb{R}^q)}.
\end{aligned}$$

Since $l \le \bar{r}$, this shows that $\mathcal{G} \in \mathcal{L}(W^{\bar{r},p}(\widehat{K};\mathbb{R}^q); W^{m,p}(\widehat{K};\mathbb{R}^q))$.
(2) Let us prove that

$$|\widehat{v} - \mathcal{I}_{\widehat{K}}(\widehat{v})|_{W^{m,p}(\widehat{K};\mathbb{R}^q)} \le c \left(|\widehat{v}|_{W^{\underline{r},p}(\widehat{K};\mathbb{R}^q)} + \ldots + |\widehat{v}|_{W^{\bar{r},p}(\widehat{K};\mathbb{R}^q)} \right). \tag{11.16}$$

The estimate is trivial if $\underline{r} = 0$. Assume now that $\underline{r} \geq 1$. Then $0 \leq \underline{r} - 1 \leq k$, so that $\mathbb{P}_{\underline{r}-1,d} \subset \mathbb{P}_{k,d}$, which implies that $[\mathbb{P}_{\underline{r}-1,d}]^q$ is pointwise invariant under $\mathcal{I}_{\widehat{K}}$. Hence, the operator $\mathcal{G}$ vanishes on $\mathbb{P}_{\underline{r}-1,d}$. We then infer that

$$|\widehat{v} - \mathcal{I}_{\widehat{K}}(\widehat{v})|_{W^{m,p}(\widehat{K};\mathbb{R}^q)} = |\mathcal{G}(\widehat{v})|_{W^{m,p}(\widehat{K};\mathbb{R}^q)} = \inf_{\widehat{p} \in [\mathbb{P}_{\underline{r}-1,d}]^q} |\mathcal{G}(\widehat{v} - \widehat{p})|_{W^{m,p}(\widehat{K};\mathbb{R}^q)}$$

$$\leq \|\mathcal{G}\|_{\mathcal{L}(W^{\overline{r},p}(\widehat{K};\mathbb{R}^q);W^{m,p}(\widehat{K};\mathbb{R}^q))} \inf_{\widehat{p} \in [\mathbb{P}_{\underline{r}-1,d}]^q} \|\widehat{v} - \widehat{p}\|_{W^{\overline{r},p}(\widehat{K};\mathbb{R}^q)}$$

$$\leq c \inf_{\widehat{p} \in [\mathbb{P}_{\underline{r}-1,d}]^q} \|\widehat{v} - \widehat{p}\|_{W^{\overline{r},p}(\widehat{K};\mathbb{R}^q)} \leq c \left(|\widehat{v}|_{W^{\underline{r},p}(\widehat{K};\mathbb{R}^q)} + \ldots + |\widehat{v}|_{W^{\overline{r},p}(\widehat{K};\mathbb{R}^q)} \right),$$

since $\|\widehat{v} - \widehat{p}\|^p_{W^{\overline{r},p}(\widehat{K};\mathbb{R}^q)} = \|\widehat{v} - \widehat{p}\|^p_{W^{\underline{r},p}(\widehat{K};\mathbb{R}^q)} + \sum_{n \in \{\underline{r}+1:\overline{r}\}} |\widehat{v}|^p_{W^{n,p}(\widehat{K};\mathbb{R}^q)}$ for $\underline{r} < \overline{r}$, and owing to the estimate (11.8) from the Bramble–Hilbert/Deny–Lions lemma applied componentwise to $\|\widehat{v} - \widehat{p}\|^p_{W^{\underline{r},p}(\widehat{K};\mathbb{R}^q)}$. This proves (11.16).

(3) Finally, let $v \in W^{\overline{r},p}(K;\mathbb{R}^q)$. We infer that

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K;\mathbb{R}^q)}$$

$$\leq c \|\mathbb{A}_K^{-1}\|_{\ell^2} \|\mathbb{J}_K^{-1}\|_{\ell^2}^m |\det(\mathbb{J}_K)|^{\frac{1}{p}} |\psi_K(v - \mathcal{I}_K(v))|_{W^{m,p}(\widehat{K};\mathbb{R}^q)}$$

$$\leq c \|\mathbb{A}_K^{-1}\|_{\ell^2} \|\mathbb{J}_K^{-1}\|_{\ell^2}^m |\det(\mathbb{J}_K)|^{\frac{1}{p}} |\psi_K(v) - \mathcal{I}_{\widehat{K}}(\psi_K(v))|_{W^{m,p}(\widehat{K};\mathbb{R}^q)}$$

$$\leq c \|\mathbb{A}_K^{-1}\|_{\ell^2} \|\mathbb{J}_K^{-1}\|_{\ell^2}^m |\det(\mathbb{J}_K)|^{\frac{1}{p}} \left(|\psi_K(v)|_{W^{\underline{r},p}(\widehat{K};\mathbb{R}^q)} + \ldots + |\psi_K(v)|_{W^{\overline{r},p}(\widehat{K};\mathbb{R}^q)} \right)$$

$$\leq c \|\mathbb{J}_K^{-1}\|_{\ell^2}^m \left(\|\mathbb{J}_K\|_{\ell^2}^{\underline{r}} |v|_{W^{\underline{r},p}(K;\mathbb{R}^q)} + \ldots + \|\mathbb{J}_K\|_{\ell^2}^{\overline{r}} |v|_{W^{\overline{r},p}(K;\mathbb{R}^q)} \right),$$

where we used the bound (11.7b) in the first line, the linearity of ψ_K and $\mathcal{I}_K = \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}} \circ \psi_K$ in the second line, the bound (11.16) in the third line, and the bound (11.7a) together with (11.12) in the fourth line. The expected error estimate follows by using (11.3) and the fact that $\sigma_K := \frac{h_K}{\rho_K}$ is uniformly bounded owing to the shape-regularity of the mesh sequence. □

Definition 11.14 (Degree of a finite element). *The largest integer k such that $[\mathbb{P}_{k,d}]^q \subset \widehat{P} \subset W^{k+1,p}(\widehat{K};\mathbb{R}^q)$ is called* degree *of the finite element.*

Remark 11.15 (Assumptions). The assumption (i) in Theorem 11.13 is easy to satisfy for finite elements since $\widehat{P}$ is in general composed of polynomial functions. If $\widehat{P}$ and $\mathcal{I}_{\widehat{K}}$ are generated from a finite element construction, then the assumption (ii) follows from (i) since $\widehat{P}$ is then pointwise invariant under $\mathcal{I}_{\widehat{K}}$. The assumption (iii) requires a bit more care since it amounts to finding an integer l s.t. $\mathcal{I}_{\widehat{K}} : W^{l,p}(\widehat{K};\mathbb{R}^q) \to \widehat{P}$ is bounded, i.e., the assumption (iii) is a stability property of the reference interpolation operator. □

Remark 11.16 (Fractional order). For simplicity, the interpolation error estimates from Theorem 11.13 are derived for functions in Sobolev spaces of integer order. We refer the reader to Chapter 22 for interpolation error estimates in Sobolev spaces of fractional order. □

11.5 Some examples

In this section, we present some examples of the application of Theorem 11.13 where ψ_K is the pullback by the geometric mapping. We refer the reader to Chapter 16 for applications of Theorem 11.13 to vector-valued finite elements when ψ_K is one of the Piola transformations from Definition 9.8.

11.5.1 Lagrange elements

Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular sequence of affine meshes. For Lagrange elements, we have seen in Example 9.4 that the transformation ψ_K is the pullback by the geometric mapping, i.e., $\psi_K(v) := \psi_K^{\mathrm{g}}(v) := v \circ \boldsymbol{T}_K$ (see (9.9a)). Hence, the choice (11.1) with $\mathbb{A}_K := 1$ for ψ_K is legitimate, and (11.12) trivially holds true (with $\gamma := 1$). Proposition 9.3 shows that $\mathcal{I}_K^{\mathrm{L}} = \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}}^{\mathrm{L}} \circ \psi_K$, where $\mathcal{I}_{\widehat{K}}^{\mathrm{L}}$ and $\mathcal{I}_K^{\mathrm{L}}$ are, respectively, the Lagrange interpolation operator in the reference cell $\widehat{K}$ and in a generic mesh cell $K \in \mathcal{T}_h$. Hence, (11.13) holds true. Furthermore, Assumption (i) in Theorem 11.13 holds true with k being the degree of the Lagrange element. Assumption (ii) also holds true since $\widehat{P}$ is pointwise invariant under $\mathcal{I}_{\widehat{K}}$. It remains to verify Assumption (iii). This assumption is satisfied if we take l to be the smallest integer such that $l > \frac{d}{p}$ (or $l \geq d$ if $p = 1$). This indeed implies that $W^{l,p}(\widehat{K}) \hookrightarrow C^0(\widehat{K})$ owing to Theorem 2.35. Assuming that $k + 1 > \frac{d}{p}$ (so that $k + 1 \geq l$), the estimate (11.14) implies that there is c s.t. for every integers $r \in \{l{:}k+1\}$ and $m \in \{0{:}r\}$, all $v \in W^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,

$$|v - \mathcal{I}_K^{\mathrm{L}}(v)|_{W^{m,p}(K)} \leq c\, h_K^{r-m} |v|_{W^{r,p}(K)}. \tag{11.17}$$

If $k + 1 \leq \frac{d}{p}$, the more general estimate (11.15) has to be used. For instance, assume that $k = 1$, $d = 3$, and $p \in [1, \frac{3}{2}]$, so that $k + 1 = 2 \leq \frac{3}{p}$. In the range $p \in [1, \frac{3}{2}]$, we can take $l = 3$ in Assumption (iii) (since either $3 > \frac{3}{p}$ for $p > 1$ or $3 \geq \frac{3}{1}$). For $m = 0$, we get

$$\|v - \mathcal{I}_K^{\mathrm{L}}(v)\|_{L^p(K)} \leq c\, h_K^2 (|v|_{W^{2,p}(K)} + h_K |v|_{W^{3,p}(K)}). \tag{11.18}$$

Remark 11.17 (Quadrangular meshes). When working on quadrangular (or hexahedral meshes), the geometric mapping is affine if and only if all the cells are parallelograms (or parallelotopes). If one wants to work with more general meshes, nonaffine geometric mappings need to be considered. This case is treated in §13.5. □

11.5.2 Modal elements

Consider now a modal finite element of degree k and let $\mathcal{I}_{\widehat{K}}^{\mathrm{m}}$ and $\mathcal{I}_K^{\mathrm{m}}$ be the modal interpolation operators in the reference cell $\widehat{K}$ and in a generic

mesh cell $K \in \mathcal{T}_h$, respectively. We have seen in Example 9.5 that the choice $\psi_K(v) := \psi_K^{\text{g}}(v) := v \circ \boldsymbol{T}_K$ is legitimate, that is, we take $\mathbb{A}_K := 1$ in (11.1) to define ψ_K, so that (11.12) trivially holds true (with $\gamma := 1$). Proposition 9.3 shows that $\mathcal{I}_K^{\text{m}} = \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}}^{\text{m}} \circ \psi_K$. Hence, (11.13) holds true. As for Lagrange elements, Assumptions (i) and (ii) in Theorem 11.13 are easy to verify. Concerning Assumption (iii), it is legitimate to take $l = 0$ since $V(\widehat{K}) = L^1(\widehat{K}; \mathbb{R}^q)$. Hence, the estimate (11.14) can always be used, i.e., there is c s.t. for every integers $r \in \{0{:}k{+}1\}$ and $m \in \{0{:}r\}$, all $v \in W^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,

$$|v - \mathcal{I}_K^{\text{m}}(v)|_{W^{m,p}(K)} \le c\, h_K^{r-m} |v|_{W^{r,p}(K)}. \tag{11.19}$$

11.5.3 L^2-orthogonal projection

Let $\widehat{P}$ be a finite-dimensional space such that $\mathbb{P}_{k,d} \subset \widehat{P} \subset W^{k+1,\infty}(\widehat{K})$. The L^2-orthogonal projection onto $\widehat{P}$ is the linear operator $\mathcal{I}_{\widehat{K}}^{\text{b}} : L^1(\widehat{K}) \to \widehat{P}$ such that for all $\widehat{v} \in L^1(\widehat{K})$, $\mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v})$ is the unique element in $\widehat{P}$ s.t.

$$\int_{\widehat{K}} (\mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v}) - \widehat{v})\widehat{q}\, \mathrm{d}\widehat{x} = 0, \qquad \forall \widehat{q} \in \widehat{P}. \tag{11.20}$$

Since $\widehat{v} - \mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v})$ and $\mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v}) - \widehat{q}$ are L^2-orthogonal for all $\widehat{q} \in \widehat{P}$, the Pythagorean identity gives

$$\|\widehat{v} - \widehat{q}\|_{L^2(\widehat{K})}^2 = \|\widehat{v} - \mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v})\|_{L^2(\widehat{K})}^2 + \|\mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v}) - \widehat{q}\|_{L^2(\widehat{K})}^2. \tag{11.21}$$

This implies that

$$\mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v}) = \arg\min_{\widehat{q} \in \widehat{P}} \|\widehat{v} - \widehat{q}\|_{L^2(\widehat{K})}. \tag{11.22}$$

Hence, $\mathcal{I}_{\widehat{K}}^{\text{b}}(\widehat{v})$ is the element in $\widehat{P}$ that is the closest to $\widehat{v}$ in the L^2-norm, and $\widehat{P}$ is pointwise invariant under $\mathcal{I}_{\widehat{K}}^{\text{b}}$.

Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular sequence of affine meshes. Let $K \in \mathcal{T}_h$. Let ψ_K^{g} be the pullback by the geometric mapping $\boldsymbol{T}_K$, i.e., $\psi_K^{\text{g}}(v) := v \circ \boldsymbol{T}_K$, and set $P_K := (\psi_K^{\text{g}})^{-1}(\widehat{P})$. The L^2-orthogonal projection onto P_K is the linear operator $\mathcal{I}_K^{\text{b}} : L^1(K) \to P_K$ such that for all $v \in L^1(K)$, $\mathcal{I}_K^{\text{b}}(v)$ is the unique element in P_K s.t.

$$\int_K (\mathcal{I}_K^{\text{b}}(v) - v)q\, \mathrm{d}x = 0, \qquad \forall q \in P_K. \tag{11.23}$$

As above, $\mathcal{I}_K^{\text{b}}(v)$ is the element in P_K that is the closest to v in the L^2-norm, and P_K is pointwise invariant under $\mathcal{I}_K^{\text{b}}$.

Lemma 11.18 (L^2-projection). *Let $p \in [1, \infty]$. There is c s.t. for every integers $r \in \{0{:}k+1\}$ and $m \in \{0{:}r\}$, all $v \in W^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$|v - \mathcal{I}_K^{\mathrm{b}}(v)|_{W^{m,p}(K)} \leq c\, h_K^{r-m} |v|_{W^{r,p}(K)}. \tag{11.24}$$

Proof. We apply Theorem 11.13. Recall from (9.9d) the Piola transformation $\psi_K^{\mathrm{b}}(v) := \det(\mathbb{J}_K)(v \circ \boldsymbol{T}_K)$. Observe that $(\psi_K^{\mathrm{b}})^{-1}(\widehat{P}) = P_K$. The map ψ_K^{b} is of the general form (11.1), i.e., $\psi_K^{\mathrm{b}}(v) = \mathbb{A}_K(v \circ \boldsymbol{T}_K)$ where $\mathbb{A}_K := \det(\mathbb{J}_K)$ is a 1×1 matrix (i.e., a real number) that trivially satisfies (11.12) (with $\gamma := 1$). For all $q \in P_K$ with $q = \widehat{q} \circ \boldsymbol{T}_K^{-1}$, we have

$$\int_K (\psi_K^{\mathrm{b}})^{-1}(\mathcal{I}_{\widehat{K}}^{\mathrm{b}}(\psi_K^{\mathrm{b}}(v)))q\,\mathrm{d}x = \int_K \det(\mathbb{J}_K)^{-1}(\mathcal{I}_{\widehat{K}}^{\mathrm{b}}(\psi_K^{\mathrm{b}}(v)) \circ \boldsymbol{T}_K^{-1})q\,\mathrm{d}x$$

$$= \int_{\widehat{K}} \epsilon_K \mathcal{I}_{\widehat{K}}^{\mathrm{b}}(\psi_K^{\mathrm{b}}(v))\widehat{q}\,\mathrm{d}\widehat{x} = \int_{\widehat{K}} \epsilon_K \psi_K^{\mathrm{b}}(v)\widehat{q}\,\mathrm{d}\widehat{x} = \int_K vq\,\mathrm{d}x,$$

with $\epsilon_K := \frac{\det(\mathbb{J}_K)}{|\det(\mathbb{J}_K)|}$, which proves that $\mathcal{I}_K^{\mathrm{b}} = (\psi_K^{\mathrm{b}})^{-1} \circ \mathcal{I}_{\widehat{K}}^{\mathrm{b}} \circ \psi_K^{\mathrm{b}}$ since $(\psi_K^{\mathrm{b}})^{-1}(\widehat{P}) = P_K$, i.e., (11.13) holds true with $\psi_K := \psi_K^{\mathrm{b}}$. It remains to verify the assumptions (i), (ii), and (iii). Assumption (i) follows from our assumption on $\widehat{P}$. Assumption (ii) follows from $\widehat{P}$ being pointwise invariant under $\mathcal{I}_{\widehat{K}}^{\mathrm{b}}$. Finally, Assumption (iii) holds true with $l := 0$. Since $l \leq k+1$, we can apply the estimate (11.14), which is nothing but (11.24). □

Remark 11.19 (Beyond finite elements). The above example shows that Theorem 11.13 can be understood more generally as an approximation result for the operator $\mathcal{I}_K$ defined by $\mathcal{I}_K := \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}} \circ \psi_K$ without directly invoking any finite element structure to build the operator $\mathcal{I}_{\widehat{K}}$. Given the affine geometric mapping $\boldsymbol{T}_K : \widehat{K} \to K$ and the transformation $\psi_K(v) := \mathbb{A}_K(v \circ \boldsymbol{T}_K)$, the key requirements are that $\mathcal{I}_{\widehat{K}} : W^{l,p}(\widehat{K}; \mathbb{R}^q) \to \widehat{P}$ is bounded, $\mathbb{P}_{k,d}$ is pointwise invariant under $\mathcal{I}_{\widehat{K}}$, and ψ_K is such that $\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2} \leq \gamma\,\|\mathbb{J}_K\|_{\ell^2}\|\mathbb{J}_K^{-1}\|_{\ell^2}$. In conclusion, the finite element construction of §9.1 is sufficient to apply Theorem 11.13 but not necessary. □

Exercises

Exercise 11.1 (High-order derivative). Let two integers $m, d \geq 2$. Consider the map $\Phi : \{1{:}d\}^m \ni \mathbf{j} \longmapsto (\Phi_1(\mathbf{j}), \ldots, \Phi_d(\mathbf{j})) \in \mathbb{N}^d$, where $\Phi_i(\mathbf{j}) := \mathrm{card}\{k \in \{1{:}m\} \mid \mathbf{j}_k = i\}$ for all $i \in \{1{:}d\}$, so that $|\Phi(\mathbf{j})| = m$ by construction. Let $C_{m,d} := \max_{\alpha \in \mathbb{N}^d, |\alpha|=m} \mathrm{card}\{\mathbf{j} \in \{1{:}d\}^m \mid \Phi(\mathbf{j}) = \alpha\}$. Let v be a smooth (scalar-valued) function. (i) Show that $\|D^m v\|_{\mathcal{M}_m(\mathbb{R}^d,\ldots,\mathbb{R}^d;\mathbb{R})} \leq C_{m,d}^{\frac{1}{2}}\left(\sum_{\alpha \in \mathbb{N}^d, |\alpha|=m} |\partial^\alpha v|^2\right)^{\frac{1}{2}}$. (ii) Show that $C_{m,2} = \max_{0 \leq l \leq m} \binom{m}{l} = 2^m$. (iii) Evaluate $C_{m,3}$ and $m \in \{2,3\}$. (iv) Show that $\sum_{\alpha \in \mathbb{N}^d, |\alpha|=m} |\partial^\alpha v| \leq \binom{d+m-1}{d-1}\|D^m v\|_{\mathcal{M}_m(\mathbb{R}^d,\ldots,\mathbb{R}^d;\mathbb{R})}$.

Exercise 11.2 (Flat triangle). Let K be a triangle with vertices $(0,0)$, $(1,0)$ and $(-1,\epsilon)$ with $0 < \epsilon \ll 1$. Consider the function $v(x_1, x_2) := x_1^2$.

Evaluate the $\mathbb{P}_1$ Lagrange interpolant $\mathcal{I}_K^L(v)$ (see (9.7)) and show that $|v - \mathcal{I}_K^L(v)|_{H^1(K)} \geq \epsilon^{-1} |v|_{H^2(K)}$. (*Hint*: use a direct calculation of $\mathcal{I}_K^L(v)$.)

Exercise 11.3 (Barycentric coordinate). Let K be a simplex with barycentric coordinates $\{\lambda_i\}_{i \in \{0:d\}}$. Prove that $|\lambda_i|_{W^{1,\infty}(K)} \leq \rho_K^{-1}$ for all $i \in \{0:d\}$.

Exercise 11.4 (Bramble–Hilbert). Prove Corollary 11.11. (*Hint*: use the Bramble–Hilbert/Deny–Lions lemma.)

Exercise 11.5 (Taylor polynomial). Let K be a convex cell. Consider a Lagrange finite element of degree $k \geq 1$ with nodes $\{a_i\}_{i \in \mathcal{N}}$ and associated shape functions $\{\theta_i\}_{i \in \mathcal{N}}$. Consider a sufficiently smooth function v. For all $x, y \in K$, consider the Taylor polynomial of order k and the exact remainder defined as follows:

$$\mathbb{T}_k(x, y) := v(x) + Dv(x)(y - x) + \ldots + \frac{1}{k!} D^k v(x) \underbrace{(y - x, \ldots, y - x)}_{k \text{ times}},$$

$$R_k(v)(x, y) := \frac{1}{(k+1)!} D^{k+1} v(\eta x + (1 - \eta) y) \underbrace{(y - x, \ldots, y - x)}_{(k+1) \text{ times}},$$

so that $v(y) = \mathbb{T}_k(x, y) + R_k(v)(x, y)$ for some $\eta \in [0, 1]$. (i) Prove that $v(x) = \mathcal{I}_K^L(v)(x) - \sum_{i \in \mathcal{N}} R_k(v)(x, a_i) \theta_i(x)$, where $\mathcal{I}_K^L$ is the Lagrange interpolant defined in (9.7). (*Hint*: interpolate with respect to y.) (ii) Prove that $D^m v(x) = D^m (\mathcal{I}_K^L(v))(x) - \sum_{i \in \mathcal{N}} R_k(v)(x, a_i) D^m \theta_i(x)$ for all $m \leq k$. (*Hint*: proceed as in (i), take m derivatives with respect to y at x, and observe that $v(x) = \mathbb{T}_k(x, x)$.) (iii) Deduce that $|v - \mathcal{I}_K^L(v)|_{W^{m,\infty}(K)} \leq c \sigma_K^m h_K^{k+1-m} |v|_{W^{k+1,\infty}(K)}$ with $c := \frac{1}{(k+1)!} c_* h_{\widehat{K}}^m \sum_{i \in \mathcal{N}} |\widehat{\theta}_i|_{W^{m,\infty}(\widehat{K})}$, where c_* comes from (11.7b) with $s = m$ and $p = \infty$.

Exercise 11.6 (L^p-stability of Lagrange interpolant). Let $\alpha \in (0, 1)$. Consider the Lagrange $\mathbb{P}_1$ shape functions $\theta_1(x) := 1 - x$ and $\theta_2(x) := x$. Consider the sequence of continuous functions $\{u_n\}_{n \in \mathbb{N} \setminus \{0\}}$ defined over the interval $K := [0, 1]$ as $u_n(x) := n^\alpha - 1$ if $0 \leq x \leq \frac{1}{n}$ and $u_n(x) := x^{-\alpha} - 1$ otherwise. (i) Prove that the sequence is uniformly bounded in $L^p(0, 1)$ for all p such that $p\alpha < 1$. (ii) Compute $\mathcal{I}_K^L(u_n)$. Is the operator $\mathcal{I}_K^L$ stable in the L^p-norm? (iii) Is the operator $\mathcal{I}_K^L$ stable in any L^r-norm with $r \in [1, \infty)$?

Exercise 11.7 (Norm scaling, $s \notin \mathbb{N}$). Complete the proof of Lemma 11.7 for the case $s \notin \mathbb{N}$. (*Hint*: use (2.6) with $s = m + \sigma$, $m := \lfloor s \rfloor$, $\sigma := s - m \in (0, 1)$.)

Exercise 11.8 (Morrey's polynomial). Let U be a nonempty open set in $\mathbb{R}^d$. Let $k \in \mathbb{N}$ and $p \in [1, \infty]$. Let $u \in W^{k,p}(U)$. Show that there is a unique polynomial $q \in \mathbb{P}_{k,d}$ s.t. $\int_U \partial^\alpha (u - q) \, dx = 0$ for all $\alpha \in \mathbb{N}^d$ of length at most k. (*Hint*: see the proof of Lemma 11.9 and also Morrey [148, Thm. 3.6.10].)

Exercise 11.9 (Fractional Sobolev norm). Let $r \in (0, 1)$. Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be an shape-regular affine mesh sequence and let $\widehat{K}$ be the reference element. Let K be an affine cell in $\mathcal{T}_h$. Using the notation $\widehat{v} := v \circ \boldsymbol{T}_K$, show that there is c such that $\|\widehat{v}\|_{H^r(\widehat{K})} \leq c h_K^{r - \frac{d}{2}} |v|_{H^r(K)}$ for all $v \in H^r(K)$ such that $\int_K v \, dx = 0$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$. (*Hint*: use Lemma 3.26.)

Part III, Chapter 12

Local inverse and functional inequalities

Inverse inequalities rely on the fact that all the norms are equivalent in finite-dimensional normed vector spaces, e.g., in the local (polynomial) space P_K generated from the reference finite element. The term 'inverse' refers to the fact that high-order Sobolev (semi)norms are bounded by lower-order (semi)norms, but the constants involved in these estimates either tend to zero or to infinity as the meshsize goes to zero. Our purpose is then to study how the norm-equivalence constants depend on the local meshsize and the polynomial degree of the reference finite element. We also derive some local functional inequalities valid in infinite-dimensional spaces. All of these inequalities are regularly invoked in this book. In the whole chapter, we consider the same setting as in Chapter 11, i.e., $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ is the reference finite element, $(\mathcal{T}_h)_{h \in \mathcal{H}}$ is a shape-regular sequence of affine meshes, $\boldsymbol{T}_K : \widehat{K} \to K$ is the geometric mapping for every mesh cell $K \in \mathcal{T}_h$, and the local finite element (K, P_K, Σ_K) is generated by using the transformation $\psi_K(v) := \mathbb{A}_K(v \circ \boldsymbol{T}_K)$ with $\mathbb{A}_K \in \mathbb{R}^{q \times q}$ s.t. $\|\mathbb{A}_K\|_{\ell^2} \|\mathbb{A}_K^{-1}\|_{\ell^2} \leq c$ (which follows from (11.12) and the regularity of the mesh sequence).

12.1 Inverse inequalities in cells

Lemma 12.1 (Bound on Sobolev seminorm). *Let $l \in \mathbb{N}$ be s.t. $\widehat{P} \subset W^{l,\infty}(\widehat{K};\mathbb{R}^q)$. There is c s.t. for every integer $m \in \{0{:}l\}$, all $p, r \in [1,\infty]$, all $v \in P_K$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$, the following holds true:*

$$|v|_{W^{l,p}(K;\mathbb{R}^q)} \leq c\, h_K^{m-l+d(\frac{1}{p}-\frac{1}{r})} |v|_{W^{m,r}(K;\mathbb{R}^q)}. \tag{12.1}$$

Proof. (1) Since all the norms in the finite-dimensional space $\widehat{P}$ are equivalent, there exists $\widehat{c}$, only depending on $\widehat{K}$, l, and q, such that $\|\widehat{v}\|_{W^{l,\infty}(\widehat{K};\mathbb{R}^q)} \leq \widehat{c}\|\widehat{v}\|_{L^1(\widehat{K};\mathbb{R}^q)}$ for all $\widehat{v} \in \widehat{P}$, which in turn means that for all $p, r \in [1,\infty]$,

$$\|\widehat{v}\|_{W^{l,p}(\widehat{K};\mathbb{R}^q)} \le \widehat{c}\,\|\widehat{v}\|_{L^r(\widehat{K};\mathbb{R}^q)}, \qquad \forall \widehat{v} \in \widehat{P}. \tag{12.2}$$

(2) Let now $v \in P_K$. Since $P_K := \psi_K^{-1}(\widehat{P})$, $\widehat{v} := \psi_K(v)$ is in $\widehat{P}$. Let $j \in \{0{:}l\}$. Using Lemma 11.7, (12.2), the assumption $\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2} \le c$, and the regularity of the mesh sequence implies that (the value of c changes at each occurrence)

$$\begin{aligned}
|v|_{W^{j,p}(K;\mathbb{R}^q)} &\le c\,\|\mathbb{A}_K^{-1}\|_{\ell^2}\|\mathbb{J}_K^{-1}\|_{\ell^2}^j|\det(\mathbb{J}_K)|^{\frac{1}{p}}\|\widehat{v}\|_{W^{j,p}(\widehat{K};\mathbb{R}^q)} \\
&\le c\,\|\mathbb{A}_K^{-1}\|_{\ell^2}\|\mathbb{J}_K^{-1}\|_{\ell^2}^j|\det(\mathbb{J}_K)|^{\frac{1}{p}}\|\widehat{v}\|_{L^r(\widehat{K};\mathbb{R}^q)} \\
&\le c\,\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2}\|\mathbb{J}_K^{-1}\|_{\ell^2}^j|\det(\mathbb{J}_K)|^{\frac{1}{p}-\frac{1}{r}}\|v\|_{L^r(K;\mathbb{R}^q)} \\
&\le c\,h_K^{-j+d(\frac{1}{p}-\frac{1}{r})}\|v\|_{L^r(K;\mathbb{R}^q)}.
\end{aligned}$$

Taking $j = l$ proves (12.1) for $m = 0$.

(3) Let now $m \in \{0{:}l\}$. Let α be a multi-index of length l, i.e., $|\alpha| = l$. One can find two multi-indices β and γ such that $\alpha = \beta + \gamma$ with $|\gamma| = m$ and $|\beta| = l - m$. It follows from Step (2) that

$$\begin{aligned}
\|\partial^\alpha v\|_{L^p(K;\mathbb{R}^q)} &= \|\partial^\beta(\partial^\gamma v)\|_{L^p(K;\mathbb{R}^q)} \le |\partial^\gamma v|_{W^{l-m,p}(K;\mathbb{R}^q)} \\
&\le c\,h_K^{m-l+d(\frac{1}{p}-\frac{1}{r})}\|\partial^\gamma v\|_{L^r(K;\mathbb{R}^q)} \le c\,h_K^{m-l+d(\frac{1}{p}-\frac{1}{r})}|v|_{W^{m,r}(K;\mathbb{R}^q)},
\end{aligned}$$

which proves (12.1) for every integer $m \in \{0{:}l\}$. $\qquad\square$

Remark 12.2 (Scale invariance). Inverse inequalities are invariant when K is dilated by any factor $\lambda > 0$. Indeed, the left-hand side of (12.1) scales as $\lambda^{-l+\frac{d}{p}}$ and the right-hand side as $\lambda^{m-l+d(\frac{1}{p}-\frac{1}{q})}\lambda^{-m+\frac{d}{q}} = \lambda^{-l+\frac{d}{p}}$. This fact is useful to verify the correctness of the exponent of h_K in (12.1). $\qquad\square$

Example 12.3 (Bound on gradient). Lemma 12.1 with $l := 1$, $m := 0$ yields

$$\|\nabla v\|_{\boldsymbol{L}^p(K;\mathbb{R}^q)} \le c\,h_K^{-1}\|v\|_{L^p(K;\mathbb{R}^q)},$$

for all $p \in [1,\infty]$, all $v \in P_K$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$. $\qquad\square$

Example 12.4 (L^p vs. L^q-norms). Lemma 12.1 with $m := 0$, $l := 0$ yields

$$\|v\|_{L^p(K;\mathbb{R}^q)} \le c\,h_K^{d(\frac{1}{p}-\frac{1}{r})}\|v\|_{L^r(K;\mathbb{R}^q)}, \tag{12.3}$$

for all $p,r \in [1,\infty]$, all $v \in P_K$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$. $\qquad\square$

Proposition 12.5 (dof-based norm). *There is c s.t.*

$$c\,\|v\|_{L^p(K;\mathbb{R}^q)} \le |K|^{\frac{1}{p}}\|\mathbb{A}_K^{-1}\|_{\ell^2}\Big(\max_{i \in \mathcal{N}}|\sigma_{K,i}(v)|\Big) \le c^{-1}\,\|v\|_{L^p(K;\mathbb{R}^q)}, \tag{12.4}$$

for all $p \in [1,\infty]$, all $v \in P_K$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. See Exercise 12.3. □

Example 12.6 (dof-based norm). For any Lagrange finite element with nodes $(\boldsymbol{a}_{K,i})_{i \in \mathcal{N}}$, $\|v\|_{L^p(K;\mathbb{R}^q)}$ is uniformly equivalent to

$$h_K^{\frac{d}{p}} \max_{i \in \mathcal{N}} \|v(\boldsymbol{a}_{K,i})\|_{\ell^2(\mathbb{R}^q)},$$

where $|K|^{\frac{1}{p}}$ has been replaced by $h_K^{\frac{d}{p}}$ owing to regularity of the mesh sequence. For the Raviart–Thomas $\mathbb{RT}_{k,d}$ element (see Chapter 14), inspection of the dofs shows that $\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}$ is uniformly equivalent to

$$h_K^{\frac{1}{p}} \max_{F \in \mathcal{F}_K} \|\boldsymbol{v}\cdot\boldsymbol{n}_F\|_{L^p(F)} + \|\Pi_K^{k-1}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)},$$

where $\boldsymbol{n}_F$ is the unit normal vector orienting the face F of K, and Π_K^{k-1} is the $\boldsymbol{L}^2(K)$-orthogonal projection onto $\mathbb{P}_{k-1,d}$ ($k \geq 1$). For the Nédélec $\mathbb{N}_{k,d}$ element (see Chapter 15), $\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}$ is uniformly equivalent to

$$h_K^{\frac{2}{p}} \max_{E \in \mathcal{E}_K} \|\boldsymbol{v}\cdot\boldsymbol{\tau}_E\|_{L^p(E)} + h_K^{\frac{1}{p}} \max_{F \in \mathcal{F}_K} \|\Pi_K^{k-1}(\boldsymbol{v})\times\boldsymbol{n}_F\|_{L^p(F)} + \|\Pi_K^{k-2}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)},$$

where $\boldsymbol{\tau}_E$ is the unit tangent vector orienting the edge E of K and Π_K^{k-2} is the $\boldsymbol{L}^2(K)$-orthogonal projection onto $\mathbb{P}_{k-2,d}$ ($k \geq 2$). □

Sharp estimates of the constant c appearing in the above inverse inequalities can be important in various contexts. For instance, the *hp*-finite element analysis requires to know how c behaves with respect to the polynomial degree; see, e.g., Schwab [177]. It turns out that estimating c in terms of the polynomial degree can be done in some particular cases. One of the earliest known inverse inequalities with a sharp estimate on c is the Markov inequality proved in the 1890s by Andrey Markov and Vladimir Markov for univariate polynomials over the interval $[-1,1]$.

Lemma 12.7 (Markov inequality). *Let $k,l \in \mathbb{N}$ with $l \leq k$ and $k \geq 1$. The following holds true for every univariate polynomial $v \in \mathbb{P}_{k,1}$:*

$$\|v^{(l)}\|_{L^\infty(-1,1)} \leq C_{\infty,k,l}\|v\|_{L^\infty(-1,1)}, \tag{12.5}$$

with $C_{\infty,k,l} := \frac{k^2(k^2-1^2)\dots(k^2-(l-1)^2)}{1\cdot3\dots(2l-1)}$.

Setting $l := 1$ in (12.5) gives $\|v'\|_{L^\infty(-1,1)} \leq C_{\infty,k}\|v\|_{L^\infty(-1,1)}$ with $C_{\infty,k} := k^2$. This type of result can be extended to the multivariate case in any dimension. In particular, it is shown in Wilhelmsen [200] that

$$\|\nabla v\|_{\boldsymbol{L}^\infty(\widehat{K})} \leq \frac{4k^2}{\text{width}(\widehat{K})}\|v\|_{L^\infty(\widehat{K})}, \qquad \forall v \in \mathbb{P}_{k,d}, \tag{12.6}$$

for all compact convex sets $\widehat{K}$ in $\mathbb{R}^d$ with nonempty interior, where width$(\widehat{K})$ is the width of $\widehat{K}$, i.e., the minimal distance between two parallel supporting hyperplanes of $\widehat{K}$; see also Kroó and Révész [129].

Results are also available for the L^2-Markov inequality in the univariate and multivariate cases; see Harari and Hughes [112], Schwab [177], Kroó [128], Özişik et al. [155]. For instance, it is shown in [177, Thm. 4.76] that

$$\|v'\|_{L^2(-1,1)} \leq C_{2,k}\|v\|_{L^2(-1,1)}, \qquad \forall v \in \mathbb{P}_{k,1}, \tag{12.7}$$

with $C_{2,k} := k((k+1)(k+\frac{1}{2}))^{\frac{1}{2}}$. Sharp estimates of the constant $C_{2,k}$ can be derived by computing the largest eigenvalue of the stiffness matrix $\mathcal{A}$ of order $(k+1)$ with entries $\mathcal{A}_{mn} := \int_{-1}^{1}(\tilde{L}_m)'(t)(\tilde{L}_n)'(t)\,\mathrm{d}t$ for all $m,n \in \{0{:}k\}$, where $\tilde{L}_m := \left(\frac{2m+1}{2}\right)^{\frac{1}{2}} L_m$, L_m being the Legendre polynomial from Definition 6.1, i.e., $\{\tilde{L}_m\}_{m\in\{0:k\}}$ is an L^2-orthonormal basis of $\mathbb{P}_{k,1}$. For instance, it is found in [155] that $C_{2,1} = 3$, $C_{2,2} = 15$, $C_{2,3} = \frac{45+\sqrt{1605}}{2}$, and $C_{2,4} = \frac{105+3\sqrt{805}}{2}$. The multivariate situation is slightly more complicated, but when $\widehat{K}$ is the unit triangle or the unit square, it is shown in [177] that

$$\|\nabla v\|_{\boldsymbol{L}^2(\widehat{K})} \leq c\,k^2\|v\|_{L^2(\widehat{K})}, \qquad \forall v \in \mathbb{P}_{k,2}, \tag{12.8}$$

where c is uniform with respect to k. By numerically evaluating the largest eigenvalue of the stiffness matrix assembled from an L^2-orthonormal basis of $\mathbb{P}_{k,2}$ on the reference triangle $\widehat{K}$, it is shown in [155] that

$$\|\nabla v\|_{\boldsymbol{L}^2(K)} \leq C_{2,k}\frac{|\partial K|}{|K|}\|v\|_{L^2(K)}, \qquad k \in \{1,2,3,4\}, \tag{12.9}$$

for every triangle K, with $C_{2,1} := \sqrt{6} \sim 2.449$, $C_{2,2} := 3\sqrt{\frac{5}{2}} \sim 4.743$, $C_{2,3} \sim 7.542$, and $C_{2,4} \sim 10.946$. Values of $C_{2,k}$ for tetrahedra with $k \in \{1{:}4\}$ are also given in [155].

12.2 Inverse inequalities on faces

Let $\mathcal{F}_K$ be the collection of the faces of a mesh cell $K \in \mathcal{T}_h$.

Lemma 12.8 (Discrete trace inequality). *Assume that $\widehat{P} \subset L^\infty(\widehat{K};\mathbb{R}^q)$. There is c s.t. the following holds true:*

$$\|v\|_{L^p(F;\mathbb{R}^q)} \leq c\,h_K^{-\frac{1}{p}+d(\frac{1}{p}-\frac{1}{r})}\|v\|_{L^r(K;\mathbb{R}^q)}, \tag{12.10}$$

for all $p,r \in [1,\infty]$, all $v \in P_K$, all $K \in \mathcal{T}_h$, all $F \in \mathcal{F}_K$, and all $h \in \mathcal{H}$.

Proof. Let $\widehat{v} := \psi_K(v)$. Then $\|v\|_{L^p(F;\mathbb{R}^q)} \leq \|\mathbb{A}_K^{-1}\|_{\ell^2} \left(\frac{|F|}{|\widehat{F}|}\right)^{\frac{1}{p}} \|\widehat{v}\|_{L^p(\widehat{F};\mathbb{R}^q)}$. Using norm equivalence in $\widehat{P}$, we infer that $\|\widehat{v}\|_{L^p(\widehat{F};\mathbb{R}^q)} \leq \widehat{c}\|\widehat{v}\|_{L^p(\widehat{K};\mathbb{R}^q)}$. Hence, $\|v\|_{L^p(F;\mathbb{R}^q)} \leq c'\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2} \left(\frac{|F|}{|\widehat{F}|}\frac{|\widehat{K}|}{|K|}\right)^{\frac{1}{p}} \|v\|_{L^p(K;\mathbb{R}^q)}$. The regularity of the mesh sequence yields (12.10) if $p = r$. The result for $r \neq p$ follows from (12.3). $\qquad\square$

Again, it can be important to have an accurate estimate of the constant c appearing in the discrete trace inequality (12.10). For instance, this constant is invoked to determine a minimal threshold on the stability parameter that is used to enforce boundary conditions weakly in the boundary penalty method and the discontinuous Galerkin method for elliptic PDEs; see Chapters 37 and 38. It is indeed possible to estimate c in the Hilbertian setting (with $p = q = 2$), when K is a simplex or a cuboid. We start with the case of the cuboid; see Canuto and Quarteroni [57], Bernardi and Maday [22].

Lemma 12.9 (Discrete trace inequality in cuboid). *Let K be a cuboid in $\mathbb{R}^d$ and let $F \in \mathcal{F}_K$. The following holds true for all $v \in \mathbb{Q}_{k,d}$:*

$$\|v\|_{L^2(F)} \leq (k+1)|F|^{\frac{1}{2}}|K|^{-\frac{1}{2}}\|v\|_{L^2(K)}. \tag{12.11}$$

Proof. We first consider the reference hypercube $\widehat{K} := [-1,1]^d$ and the face $\widehat{F} := \{\widehat{x}_d = -1\}$. Recall the rescaled Legendre polynomials $\tilde{L}_m := \left(\frac{2m+1}{2}\right)^{\frac{1}{2}} L_m$, i.e., $\{\tilde{L}_m\}_{m\in\{0:k\}}$ is an L^2-orthonormal basis of $\mathbb{Q}_{k,1} = \mathbb{P}_{k,1}$. An L^2-orthonormal basis of $\mathbb{Q}_{k,d}$ is obtained by constructing the tensor product of this one-dimensional basis. Let $\widehat{v} \in \mathbb{Q}_{k,d}$ and write

$$\widehat{v}(\widehat{\boldsymbol{x}}) = \sum_{i_1\in\{0:k\}} \cdots \sum_{i_d\in\{0:k\}} \widehat{v}_{i_1\ldots i_d}\tilde{L}_{i_1}(\widehat{x}_1)\ldots\tilde{L}_{i_d}(\widehat{x}_d).$$

Let $V \in \mathbb{R}^{(k+1)^d}$ be the coordinate vector of $\widehat{v}$ in this tensor-product basis. Using orthonormality, we infer that

$$\int_{\widehat{F}} \widehat{v}(\widehat{\boldsymbol{x}})^2 \, d\widehat{s} = V^\mathsf{T}\mathcal{T}V,$$

where the $(k+1)^d\times(k+1)^d$ symmetric matrix $\mathcal{T}$ is block-diagonal with $(k+1)^{d-1}$ diagonal blocks all equal to the rank-one matrix $\mathcal{U} := UU^\mathsf{T}$, where $U = (\tilde{L}_0(-1),\ldots,\tilde{L}_k(-1))^\mathsf{T}$. As a result, the largest eigenvalue of $\mathcal{T}$ is

$$\lambda_{\max}(\mathcal{T}) = \lambda_{\max}(\mathcal{U}) = \|U\|_{\ell^2(\mathbb{R}^{k+1})}^2 = \sum_{m\in\{0:k\}} \frac{2m+1}{2} = \frac{(k+1)^2}{2}.$$

Since $V^\mathsf{T}V = \|\widehat{v}\|_{L^2(\widehat{K})}^2$ by orthonormality of the basis, we infer that

$$\|\widehat{v}\|^2_{L^2(\widehat{F})} \le \lambda_{\max}(\mathcal{T})\|\widehat{v}\|^2_{L^2(\widehat{K})} = \tfrac{1}{2}(k+1)^2\|\widehat{v}\|^2_{L^2(\widehat{K})}.$$

Finally, we obtain (12.11) by mapping the above estimate back to the cuboid K and by observing that $|\widehat{K}| = 2|\widehat{F}|$. □

Lemma 12.10 (Discrete trace inequality in simplices). *Let K be a simplex in $\mathbb{R}^d$ and let $F \in \mathcal{F}_K$. The following holds true for all $v \in \mathbb{P}_{k,d}$:*

$$\|v\|_{L^2(F)} \le \left((k+1)(k+d)d^{-1}\right)^{\frac{1}{2}} |F|^{\frac{1}{2}}|K|^{-\frac{1}{2}}\|v\|_{L^2(K)}. \tag{12.12}$$

Proof. See Warburton and Hesthaven [196]. □

12.3 Functional inequalities in meshes

This section presents two important functional inequalities: the Poincaré–Steklov inequality for functions having zero mean-value over a mesh cell and the multiplicative trace inequality for functions having a trace at the boundary of a mesh cell.

12.3.1 Poincaré–Steklov inequality in cells

Lemma 12.11 (Poincaré–Steklov). *Let $K \in \mathcal{T}_h$ and assume that K is a convex set. Then for all $v \in H^1(K)$ with $\underline{v}_K := \frac{1}{|K|}\int_K v\,\mathrm{d}x$, we have*

$$\|v - \underline{v}_K\|_{L^2(K)} \le \pi^{-1}h_K|v|_{H^1(K)}. \tag{12.13}$$

Proof. This is a paraphrase of Lemma 3.24. □

Lemma 12.12 (Fractional Poincaré–Steklov). *Let $p \in [1, \infty)$, $r \in (0, 1)$, and let $K \in \mathcal{T}_h$. Then for all $v \in W^{r,p}(K)$ with $\underline{v}_K := \frac{1}{|K|}\int_K v\,\mathrm{d}x$, we have*

$$\|v - \underline{v}_K\|_{L^p(K)} \le h_K^s \left(\frac{h_K^d}{|K|}\right)^{\frac{1}{p}} |v|_{W^{r,p}(K)}. \tag{12.14}$$

Proof. This is a paraphrase of Lemma 3.26. □

Corollary 12.13 (Polynomial approximation). *Assume that the mesh sequence $(\mathcal{T}_h)_{h \in \mathcal{H}}$ is shape-regular. Let $k \in \mathbb{N}$. There is c s.t. for every real numbers $r \in [0, k+1]$ and $p \in [1, \infty)$ if $r \notin \mathbb{N}$ and $p \in [1, \infty]$ if $r \in \mathbb{N}$, every integer $m \in \{0:\lfloor r \rfloor\}$ (where $\lfloor r \rfloor$ denotes the largest integer $n \in \mathbb{N}$ s.t. $n \le r$), all $v \in W^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$\inf_{q \in \mathbb{P}_{k,d}} |v - q|_{W^{m,p}(K)} \le c\, h_K^{r-m}|v|_{W^{r,p}(K)}, \tag{12.15}$$

where the mesh cells are supposed to be convex sets if $r \ge 1$.

Proof. If $m = r$, there is nothing to prove, so let us assume that $m < r$. If $r \in (0, 1)$, we have $m = 0$, and (12.15) follows from the fractional Poincaré–Steklov (12.14) and the regularity of the mesh sequence. If $r = 1$, we only need to consider the case $m = 0$ (since otherwise $m = 1 = r$), and (12.15) follows from the Poincaré–Steklov inequality (12.13) and the convexity of K. If $k = 0$, the proof is complete. Otherwise, $k \geq 1$ and let us assume now that $r > 1$. Let $\ell \in \mathbb{N}$ be s.t. $\ell := \lceil r \rceil - 1$ (where $\lceil r \rceil$ denotes the smallest integer $n \in \mathbb{N}$ s.t. $n \geq r$). Notice that we have $m \leq \ell \leq k$ and $1 \leq \ell$. The key idea is to take $q := \pi_\ell(v) \in \mathbb{P}_{\ell,d} \subset \mathbb{P}_{k,d}$ since $\ell \leq k$, where $\pi_\ell(v)$ is defined by $\int_K \partial^\alpha(v - \pi_\ell(v)) \, dx = 0$ for all $\alpha \in \mathbb{N}^d$ of length at most ℓ (see Exercise 11.8), and then to invoke the above Poincaré–Steklov inequalities in K. Since $\partial^\alpha(v - \pi_\ell(v))$ has zero mean-value on K for every multi-index $\alpha \in \mathbb{N}^d$ of length m with $0 \leq m \leq \ell - 1$, repeated applications of the Poincaré–Steklov inequality (12.13) (and the convexity of K) imply that

$$|v - \pi_\ell(v)|_{W^{m,p}(K)} \leq c\, h_K^{\ell - m}|v - \pi_\ell(v)|_{W^{\ell,p}(K)}.$$

Since $\partial^\alpha(v - \pi_\ell(v))$ has zero mean-value on K for any multi-index $\alpha \in \mathbb{N}^d$ of length ℓ as well, we can apply one more time either (12.13) or (12.14) to the right-hand side. If $r \in \mathbb{N}$, we invoke the convexity of K and apply (12.13) to obtain (12.15). If $r \notin \mathbb{N}$, we apply (12.14) and invoke the regularity of the mesh sequence to obtain (12.15). □

Remark 12.14 (Comparison). The estimate (12.15) is similar in spirit to the Bramble–Hilbert lemma (Lemma 11.9), except that in Lemma 11.9 it is not known how the constant c depends on K. This difficulty was circumvented in Theorem 11.13 by using that all the mesh cells are generated from a fixed reference cell. This assumption is not used in the proof of (12.15), which instead assumes the mesh cells to be convex sets. The estimate (12.15) can be extended to (connected) cells that can be partitioned into a uniformly finite number of convex subsets (e.g., simplices). The key point to establish this result is that the Poincaré–Steklov inequality (12.13) can be generalized to such sets; see Remark 22.11. □

12.3.2 Multiplicative trace inequality

Let $K \in \mathcal{T}_h$ and let $F \in \mathcal{F}_K$ be a face of K. Consider a function $v \in W^{1,p}(K)$. Then v has a trace in $L^p(F)$ (see Theorem 3.10). The following result gives an estimate of $\|v\|_{L^p(F)}$ in terms of powers of $\|v\|_{L^p(K)}$ and $\|\nabla v\|_{\boldsymbol{L}^p(K)}$ (hence the name multiplicative).

Lemma 12.15 (Multiplicative trace inequality). *Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular sequence of affine simplicial meshes in $\mathbb{R}^d$. There is c s.t. for all $p \in [1, \infty]$, all $v \in W^{1,p}(K)$, all $K \in \mathcal{T}_h$, all $F \in \mathcal{F}_K$, and all $h \in \mathcal{H}$,*

$$\|v\|_{L^p(F)} \leq c\, \|v\|_{L^p(K)}^{1 - \frac{1}{p}} \left(h_K^{-\frac{1}{p}} \|v\|_{L^p(K)}^{\frac{1}{p}} + \|\nabla v\|_{\boldsymbol{L}^p(K)}^{\frac{1}{p}} \right). \tag{12.16}$$

Proof. Let $K \in \mathcal{T}_h$ and $v \in W^{1,p}(K)$. Assume first that $p \in [1, \infty)$. Let F be a face of K and let $\boldsymbol{z}_F$ be the vertex of K opposite to F. Consider the Raviart–Thomas function $\boldsymbol{\theta}_F(\boldsymbol{x}) := \frac{|F|}{d|K|}(\boldsymbol{x} - \boldsymbol{z}_F)$ (see §14.1). One can verify that the normal component of $\boldsymbol{\theta}_F$ is equal to 1 on F and 0 on the other faces of K. Since $\nabla \cdot \boldsymbol{\theta}_F = \frac{|F|}{|K|}$, we infer using the divergence theorem that

$$
\begin{aligned}
\|v\|_{L^p(F)}^p &= \int_{\partial K} |v|^p (\boldsymbol{\theta}_F \cdot \boldsymbol{n}) \, \mathrm{d}s = \int_K \nabla \cdot (|v|^p \boldsymbol{\theta}_F) \, \mathrm{d}x \\
&= \int_K \left(|v|^p \nabla \cdot \boldsymbol{\theta}_F + pv|v|^{p-2} \boldsymbol{\theta}_F \cdot \nabla v \right) \mathrm{d}x \\
&= \frac{|F|}{|K|} \|v\|_{L^p(K)}^p + \frac{p}{d} \frac{|F|}{|K|} \int_K v|v|^{p-2}(\boldsymbol{x} - \boldsymbol{z}_F) \cdot \nabla v \, \mathrm{d}x.
\end{aligned}
$$

Using Hölder's inequality and introducing the length $\ell_F^\perp$ defined as the largest length of an edge of K having $\boldsymbol{z}_F$ as an endpoint, we infer that

$$
\|v\|_{L^p(F)}^p \le \frac{|F|}{|K|} \|v\|_{L^p(K)}^p + \frac{p}{d} \frac{|F|\ell_F^\perp}{|K|} \|v\|_{L^p(K)}^{p-1} \|\nabla v\|_{\boldsymbol{L}^p(K)},
$$

which implies the bound (12.16) using the regularity of the mesh sequence and the fact that $p^{\frac{1}{p}} \le e^{\frac{1}{e}} < \frac{3}{2}$. Finally, the bound for $p = \infty$ is obtained by passing to the limit $p \to \infty$ in (12.16) since c is uniform w.r.t. p and since $\lim_{p \to \infty} \|\cdot\|_{L^p(K)} = \|\cdot\|_{L^\infty(K)}$. $\qquad\square$

Remark 12.16 (Literature). The idea of using a Raviart–Thomas function to prove (12.16) can be traced to Monk and Süli [146, App. B] and Carstensen and Funken [62, Thm. 4.1]. See also Ainsworth [5, Lem. 10] and Veeser and Verfürth [193, Prop. 4.2]. $\qquad\square$

Remark 12.17 (Application). Let $\mathcal{I}_K : V(K) \to P_K$ be an interpolation operator s.t. $W^{1,p}(K) \hookrightarrow V(K)$, $p \in [1, \infty]$, and $|v - \mathcal{I}_K(v)|_{W^{m,p}(K)} \le ch_K^{r-m}|v|_{W^{r,p}(K)}$ for all $r \in \{1{:}k{+}1\}$, $k \ge 0$, $m \in \{0, 1\}$, all $v \in W^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$. The multiplicative trace inequality (12.16) can then be used to estimate the approximation properties of $\mathcal{I}_K$ in $L^p(F)$. Combining (12.16) with the above estimate on $|v - \mathcal{I}_K(v)|_{W^{m,p}(K)}$ gives

$$
\|v - \mathcal{I}_K(v)\|_{L^p(F)} \le c \, h_K^{r-\frac{1}{p}} |v|_{W^{r,p}(K)}.
$$

When $\mathcal{I}_K := \mathcal{I}_K^{\mathrm{b}}$ is the L^2-orthogonal projection built using $\mathbb{P}_{k,d}$ (see §11.5.3), it is shown in Chernov [67, Thm. 1.1] that c decays like $k^{-r+\frac{1}{2}}$ for $p = 2$. $\quad\square$

Remark 12.18 (Nonsimplicial cells). Lemma 12.15 can be extended to nonsimplicial cells s.t. one can find a vector-valued function $\boldsymbol{\theta}_F$ with normal component equal to 1 on F and 0 on the other faces, and satisfying $h_K \|\nabla \cdot \boldsymbol{\theta}_F\|_{L^\infty(K)} + \|\boldsymbol{\theta}_F\|_{\boldsymbol{L}^\infty(K)} \le c$ uniformly w.r.t. F, K, and h. $\qquad\square$

Remark 12.19 (Fractional trace inequality). The multiplicative trace inequality from Lemma 12.15 can be extended to functions in fractional Sobolev spaces. Let $p \in (1, \infty)$ and $s \in (\frac{1}{p}, 1)$ (we exclude the case $s = 1$ since it is already covered by Lemma 12.15). Functions in $W^{s,p}(K)$ have traces in $L^p(F)$ for every face F of K (see Theorem 3.10). Then one can show (see Exercise 12.6 or Ciarlet [73, Prop. 3.1] and the work by the authors [97, Lem. 7.2]) that there is c s.t. for all $v \in W^{s,p}(K)$, all $K \in \mathcal{T}_h$, all $F \in \mathcal{F}_K$, and all $h \in \mathcal{H}$,

$$\|v\|_{L^p(F)} \leq c \left(h_K^{-\frac{1}{p}} \|v\|_{L^p(K)} + h_K^{s-\frac{1}{p}} |v|_{W^{s,p}(K)} \right). \tag{12.17}$$

The constant c is uniform w.r.t. s and p as long as sp is bounded from below away from 1, but c can grow unboundedly as $sp \downarrow 1$. □

Exercises

Exercise 12.1 (ℓ^p vs. ℓ^r). Let p, r be two nonnegative real numbers. Let $\{a_i\}_{i \in I}$ be a finite sequence of nonnegative numbers. Set $\|a\|_{\ell^p(\mathbb{R}^I)} := (\sum_{i \in I} a_i^p)^{\frac{1}{p}}$ and $\|a\|_{\ell^r(\mathbb{R}^I)} := (\sum_{i \in I} a_i^r)^{\frac{1}{r}}$. (i) Prove that $\|a\|_{\ell^p(\mathbb{R}^I)} \leq \|a\|_{\ell^r(\mathbb{R}^I)}$ for $r \leq p$. (*Hint*: set $\theta_i := a_i^r / \|a\|_{\ell^r(\mathbb{R}^I)}^r$.) (ii) Prove that $\|a\|_{\ell^p(\mathbb{R}^I)} \leq \operatorname{card}(I)^{\frac{r-p}{pr}} \|a\|_{\ell^r(\mathbb{R}^I)}$ for $r > p$.

Exercise 12.2 (L^p-norm of shape functions). Let $\theta_{K,i}$, $i \in \mathcal{N}$, be a local shape function. Let $p \in [1, \infty]$. Assume that $(\mathcal{T}_h)_{h \in \mathcal{H}}$ is shape-regular. Prove that $\|\theta_{K,i}\|_{L^p(K)}$ is equivalent to $h_K^{d/p}$ uniformly w.r.t. $K \in \mathcal{T}_h$ and $h \in \mathcal{H}$.

Exercise 12.3 (dof norm). Prove Proposition 12.5. (*Hint*: use Lemma 11.7.)

Exercise 12.4 (Inverse inequality). (i) Let $k \geq 1$, $p \in [1, \infty]$, let $\widehat{K} := \{(\widehat{x}_1, \dots, \widehat{x}_d) \in (0, 1)^d \mid \sum_{i \in \{1:d\}} \widehat{x}_i \leq 1\}$, and set $\widehat{c}_{k,p} := \sup_{\widehat{v} \in \mathbb{P}_{k,d}} \frac{\|\nabla \widehat{v}\|_{\boldsymbol{L}^p(\widehat{K})}}{\|\widehat{v}\|_{L^p(\widehat{K})}}$. Explain why $\widehat{c}_{k,p}$ is finite. (ii) Let K be a simplex in $\mathbb{R}^d$ and let ρ_K denote the diameter of its largest inscribed ball. Show that $\|\nabla v\|_{\boldsymbol{L}^p(K)} \leq \widehat{c}_{k,p} \frac{\sqrt{2}}{\rho_K} \|v\|_{L^p(K)}$ for all $v \in \mathbb{P}_{k,d} \circ \boldsymbol{T}_K$, where $\boldsymbol{T}_K : \widehat{K} \to K$ is the geometric mapping. (*Hint*: use (9.8a) and Lemma 11.1.)

Exercise 12.5 (Markov inequality). (i) Justify that the constant $C_{2,k}$ in the Markov inequality (12.7) can be determined as the largest eigenvalue of the stiffness matrix $\mathcal{A}$. (ii) Compute numerically the constant $C_{2,k}$ for $k \in \{1, 2, 3\}$.

Exercise 12.6 (Fractional trace inequality). Prove (12.17). (*Hint*: use a trace inequality in $W^{s,p}(\widehat{K})$.)

Exercise 12.7 (Mapped polynomial approximation). Let $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ be a reference finite element such $\mathbb{P}_{k,d} \subset \widehat{P}$, $k \in \mathbb{N}$. Let $\mathcal{T}_h$ be a member of a shape-regular mesh sequence. Let $\boldsymbol{T}_K(\widehat{K}) = K \in \mathcal{T}_h$ and let (K, P_K, Σ_K) be the finite element generated by the geometric mapping $\boldsymbol{T}_K$ and the functional transformation $\psi_K(v) := \mathbb{A}_K(v \circ \boldsymbol{T}_K)$. Recall that $P_K = \psi_K^{-1}(\widehat{P})$. Show that there is c s.t.

$$\inf_{q \in P_K} |v - q|_{W^{m,p}(K)} \leq c\, h_K^{r-m} |v|_{W^{r,p}(K)}, \tag{12.18}$$

for all $r \in [0, k+1]$, all $p \in [1, \infty)$ if $r \notin \mathbb{N}$ or all $p \in [1, \infty]$ if $r \in \mathbb{N}$, every integer $m \in \{0 : \lfloor r \rfloor\}$, all $v \in W^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$, where the mesh cells are supposed to be convex sets if $r \geq 1$. (*Hint*: use Lemma 11.7 and Corollary 12.13.)

Exercise 12.8 (Trace inequality). Let U be a Lipschitz domain in $\mathbb{R}^d$. Prove that there are $c_1(U)$ and $c_2(U)$ such that $\|v\|_{L^p(\partial U)} \leq c_1(U)\|v\|_{L^p(U)} + c_2(U)\|\nabla v\|_{\boldsymbol{L}^p(U)}^{\frac{1}{p}}\|v\|_{L^p(U)}^{1-\frac{1}{p}})$ for all $p \in [1, \infty)$ and all $v \in W^{1,p}(U)$. (*Hint*: accept as a fact that there exists a smooth vector field $\boldsymbol{N} \in \boldsymbol{C}^1(\overline{U})$ and $c_0(U) > 0$ such that $(\boldsymbol{N} \cdot \boldsymbol{n})_{|\partial U} \geq c_0(U)$ and $\|\boldsymbol{N}(\boldsymbol{x})\|_{\ell^2(\mathbb{R}^d)} = 1$ for all $\boldsymbol{x} \in U$.)

Exercise 12.9 (Weighted inverse inequalities). Let $k \in \mathbb{N}$. (i) Prove that $\|(1-t^2)^{\frac{1}{2}}v'\|_{L^2(-1,1)} \leq \big(k(k+1)\big)^{\frac{1}{2}}\|v\|_{L^2(-1,1)}$ for all $v \in \mathbb{P}_{k,1}$. (*Hint*: let $\tilde{L}_m := \big(\frac{2m+1}{2}\big)^{1/2} L_m$, L_m being the Legendre polynomial from Definition 6.1, and prove that $\int_{-1}^{1}(1-t^2)(\tilde{L}_m)'(t)(\tilde{L}_n)'(t)\,\mathrm{d}t = \delta_{mn}m(m+1)$ for every integers $m, n \in \{0 : k\}$.) (ii) Prove that $\|v\|_{L^2(-1,1)} \leq (k+2)\|(1-t^2)^{\frac{1}{2}}v\|_{L^2(-1,1)}$ for all $v \in \mathbb{P}_{k,1}$. (*Hint*: consider a Gauss–Legendre quadrature with $l_{\mathcal{Q}} := k+2$ and use the fact that the rightmost Gauss–Legendre node satisfies $\xi_{l_{\mathcal{Q}}} \leq \cos(\frac{\pi}{2l_{\mathcal{Q}}})$.) *Note*: see also Verfürth [195].

Part III, Chapter 13

Local interpolation on nonaffine meshes

In this chapter, we extend the results of Chapter 11 to nonaffine meshes. For simplicity, the transformation ψ_K is the pullback by the geometric mapping, but this mapping is now nonaffine. The first difficulty consists of proving a counterpart of Lemma 11.7 to compare Sobolev norms. This is not a trivial task since the chain rule involves higher-order derivatives of the geometric mapping. The second difficulty is to define a notion of shape-regularity for mesh sequences built using nonaffine geometric mappings. We show how to do this using a perturbation theory, and we present various examples.

13.1 Introductory example on curved simplices

If one wants to approximate a problem posed in a domain D with a curved boundary ∂D using a finite element of degree $k \geq 2$, it is often necessary to use nonaffine cells since otherwise the geometric error in the representation of the boundary can dominate the approximation error. A relatively straightforward way to generate nonaffine cells is as follows: (i) Construct a mesh $(\widetilde{\mathcal{T}}_h)_{h \in \mathcal{H}}$ composed of affine cells with all the vertices lying on the curved boundary ∂D. (ii) For each affine cell $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ having a nonempty intersection with ∂D, design a new geometric mapping (of degree larger than 1) that approximates the boundary more accurately than $\widetilde{K}$. Then replace $\widetilde{K}$ by the new cell thus created.

Example 13.1 (Simple construction). An example relying on $\mathbb{P}_{2,2}$ or $\mathbb{Q}_{2,2}$ Lagrange elements in $\mathbb{R}^2$ to build the geometric mapping (see Figure 13.1) is as follows. (i) Let $\widetilde{K}$ be a triangle or a quadrangle having an edge whose vertices lie on ∂D. Let $\{\widetilde{\boldsymbol{a}}_i\}_{i \in \mathcal{N}_{\mathrm{geo}}}$ be the geometric nodes of $\widetilde{K}$ with $\mathcal{N}_{\mathrm{geo}} := \{1{:}n_{\mathrm{geo}}\}$ ($n_{\mathrm{geo}} := 6$ for a triangle and $n_{\mathrm{geo}} := 9$ for a quadrangle). (ii) For all $i \in \mathcal{N}_{\mathrm{geo}}$, construct from $\widetilde{\boldsymbol{a}}_i$ a new node $\boldsymbol{a}_i$ as follows: If $\widetilde{\boldsymbol{a}}_i$ is located at the middle of an edge whose vertices lie on ∂D, $\boldsymbol{a}_i$ is defined as
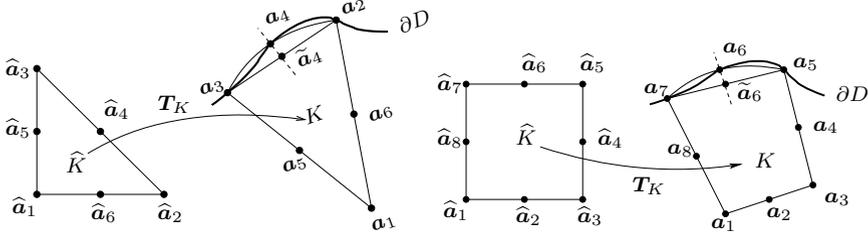
Fig. 13.1 Construction of a curved triangle (left) and a curved quadrangle (right).

the intersection with ∂D of the line normal to the corresponding edge and passing through the node $\widetilde{\boldsymbol{a}}_i$. Otherwise, set $\boldsymbol{a}_i := \widetilde{\boldsymbol{a}}_i$. (iii) Replace $\widetilde{K}$ by the curved triangle K by defining the mapping $\boldsymbol{T}_K : \widehat{K} \to K$ such that $\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := \sum_{i\in\mathcal{N}_{\mathrm{geo}}} \widehat{\psi}_i(\widehat{\boldsymbol{x}})\boldsymbol{a}_i$ for all $\widehat{\boldsymbol{x}} \in \widehat{K}$, where $\{\widehat{\psi}_i\}_{i\in\mathcal{N}_{\mathrm{geo}}}$ are the reference $\mathbb{P}_{2,2}$ or $\mathbb{Q}_{2,2}$ Lagrange shape functions.   □

13.2 A perturbation theory

This section presents a perturbation theory introduced by Ciarlet and Raviart [78] to analyze the finite element interpolation error on nonaffine cells.

13.2.1 Setting and notation

Let $(\widehat{K}, \widehat{P}_{\mathrm{geo}}, \widehat{\Sigma}_{\mathrm{geo}})$ be a reference geometric Lagrange finite element with the nodes $\{\widehat{\boldsymbol{a}}_i\}_{i\in\mathcal{N}_{\mathrm{geo}}}$ and the shape functions $\{\widehat{\psi}_i\}_{i\in\mathcal{N}_{\mathrm{geo}}}$. Let us now consider two sets of points in $\mathbb{R}^d$,

$$\{\widetilde{\boldsymbol{a}}_i\}_{i\in\mathcal{N}_{\mathrm{geo}}}, \qquad \{\boldsymbol{a}_i\}_{i\in\mathcal{N}_{\mathrm{geo}}}. \tag{13.1}$$

Let $\widetilde{\boldsymbol{T}} : \widehat{K} \to \mathbb{R}^d$ and $\boldsymbol{T} : \widehat{K} \to \mathbb{R}^d$ be the mappings defined as follows:

$$\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}) := \sum_{i\in\mathcal{N}_{\mathrm{geo}}} \widehat{\psi}_i(\widehat{\boldsymbol{x}})\widetilde{\boldsymbol{a}}_i, \qquad \boldsymbol{T}(\widehat{\boldsymbol{x}}) := \widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}) + \sum_{i\in\mathcal{N}_{\mathrm{geo}}} \widehat{\psi}_i(\widehat{\boldsymbol{x}})(\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i). \tag{13.2}$$

Notice that $\widetilde{\boldsymbol{T}}, \boldsymbol{T} \in [\widehat{P}_{\mathrm{geo}}]^d$. Let us set $\widetilde{K} := \widetilde{\boldsymbol{T}}(\widehat{K})$ and $K := \boldsymbol{T}(\widehat{K})$. The subscripts K and $\widetilde{K}$ are henceforth omitted for the geometric mappings $\widetilde{\boldsymbol{T}}$ and $\boldsymbol{T}$ to simplify the notation.

The setting we consider is as follows. We assume that we have at hand a mesh sequence $(\widetilde{\mathcal{T}}_h)_{h\in\mathcal{H}}$ such that every cell $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ is generated from the reference cell $\widehat{K}$ using the geometric mapping $\widetilde{\boldsymbol{T}}$, which we assume to be a well-behaved diffeomorphism. By using the geometric mapping $\boldsymbol{T}$ defined in (13.2), one constructs a new mesh $\mathcal{T}_h$ composed of (possibly curved) cells

K that are perturbations of the cells $\widetilde{K}$. Our goal is to estimate the finite element interpolation error when working with the geometric mapping $\boldsymbol{T}$.

Assuming that $\widetilde{\boldsymbol{T}} : \widehat{K} \to \widetilde{K}$ is a reasonable diffeomorphism, our first step is to ascertain that $\boldsymbol{T} : \widehat{K} \to K$ is also a diffeomorphism with reasonable smoothness properties. This is done by making sure that K is close to $\widetilde{K}$, i.e., that $\max_{i \in \mathcal{N}_{\mathrm{geo}}} \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2(\mathbb{R}^d)}$ is small enough. The small parameter that comes into play is the nondimensional ratio h/ℓ_D, where h is the diameter of the mesh cell $\widetilde{K}$ or K and ℓ_D is the diameter of the domain D.

Example 13.2 (1D). Let us consider the $\mathbb{P}_{2,1}$ Lagrange element ($n_{\mathrm{geo}} := 3$) with $\widehat{K} := [0,1]$. Consider the two cells $\widetilde{K} = K := [0,h]$ with the assumption $h \ll \ell_D$. Consider the Lagrange nodes $\widetilde{\boldsymbol{a}}_1 := 0$, $\widetilde{\boldsymbol{a}}_3 := \frac{1}{2}h$, $\widetilde{\boldsymbol{a}}_2 := h$, and $\boldsymbol{a}_1 := 0$, $\boldsymbol{a}_3 := \frac{1}{2}h + \frac{1}{4}\frac{h^2}{\ell_D}$, $\widetilde{\boldsymbol{a}}_2 := h$. Then $\widetilde{\boldsymbol{T}}(\widehat{x}) = \widehat{x}h$ and $\boldsymbol{T}(\widehat{x}) = \widehat{x}h + \widehat{x}(1-\widehat{x})\frac{h^2}{\ell_D}$. Notice that in this example $\widetilde{\boldsymbol{T}}$ is affine ($\|D\widetilde{\boldsymbol{T}}\| = h$, $\|D^2\widetilde{\boldsymbol{T}}\| = 0$) and $\boldsymbol{T}$ is quadratic ($\|D\boldsymbol{T}(\widehat{x})\| = h(1 + \frac{h}{\ell_D} - 2\widehat{x}\frac{h}{\ell_D})$, $\|D^2\boldsymbol{T}(\widehat{x})\| = 2h\frac{h}{\ell_D}$, and $\|D^3\boldsymbol{T}(\widehat{x})\| = 0$), and $\boldsymbol{T}$ converges to $\widetilde{\boldsymbol{T}}$ as $\frac{h}{\ell_D} \to 0$. □

13.2.2 Bounds on the derivatives of T and T^{-1}

The (Fréchet) derivatives of $\boldsymbol{T}$ and $\widetilde{\boldsymbol{T}}$ of order $m \geq 1$ at a point $\widehat{\boldsymbol{x}} \in \widehat{K}$ are denoted by $D^m\boldsymbol{T}(\widehat{\boldsymbol{x}})$ and $D^m\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}})$, respectively (the superscript is omitted if $m = 1$). Recall from Appendix B that $D^m\boldsymbol{T}$ and $D^m\widetilde{\boldsymbol{T}}$ are members of $\mathcal{M}_m(\mathbb{R}^d, \ldots, \mathbb{R}^d; \mathbb{R}^d)$, i.e., they are multilinear maps from $\mathbb{R}^d \times \ldots \times \mathbb{R}^d$ to $\mathbb{R}^d$ (i.e., $D\boldsymbol{T}$ and $D\widetilde{\boldsymbol{T}}$ are linear maps in $\mathcal{L}(\mathbb{R}^d; \mathbb{R}^d)$, $D^2\boldsymbol{T}$ and $D^2\widetilde{\boldsymbol{T}}$ are bilinear maps in $\mathcal{M}_2(\mathbb{R}^d, \mathbb{R}^d; \mathbb{R}^d)$, etc.). For every map $A \in C^m(\widehat{K}; \mathbb{R}^q)$, $q \geq 1$, and all $\widehat{\boldsymbol{x}} \in \widehat{K}$, we set

$$|D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{P}} := \max_{\widehat{\boldsymbol{h}}_1, \ldots, \widehat{\boldsymbol{h}}_m \in \mathbb{R}^d} \frac{\|D^m A(\widehat{\boldsymbol{x}})(\widehat{\boldsymbol{h}}_1, \ldots, \widehat{\boldsymbol{h}}_m)\|_{\ell^2(\mathbb{R}^q)}}{\|\widehat{\boldsymbol{h}}_1\|_{\ell^2(\mathbb{R}^d)} \ldots \|\widehat{\boldsymbol{h}}_m\|_{\ell^2(\mathbb{R}^d)}}. \tag{13.3}$$

The notation for the subscript $\mathbb{P}$ is motivated by the fact that for all $k \geq 0$, $|D^{k+1}A(\widehat{\boldsymbol{x}})|_{\mathbb{P}} = 0$ for all $\widehat{\boldsymbol{x}} \in K$ if and only if A is $[\mathbb{P}_{k,d}]^q$-valued. Note that the right-hand side of (13.3) is the canonical norm in $\mathcal{M}_m(\mathbb{R}^d, \ldots, \mathbb{R}^d; \mathbb{R}^q)$. We use the shorthand notation $\|D^m A\| := \| |D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{P}}\|_{L^\infty(\widehat{K})}$ for all $m \geq 1$.

Let us first identify a condition ensuring that $\boldsymbol{T}$ is a C^1-diffeomorphism with reasonable bounds on $D\boldsymbol{T}$, $D(\boldsymbol{T}^{-1})$, and $\det(D\boldsymbol{T})$.

Lemma 13.3 (Bound on $D\boldsymbol{T}$, $D(\boldsymbol{T}^{-1})$). *Let $\widetilde{\boldsymbol{T}}, \boldsymbol{T}$ be defined in (13.2). Assume that $\widetilde{\boldsymbol{T}}$ is a C^1-diffeomorphism, $\widehat{P}_{\mathrm{geo}} \subset C^1(\widehat{K}; \mathbb{R})$, and there is $c_1 \in [0,1)$ s.t.*

$$\|(D\widetilde{\boldsymbol{T}})^{-1}\| \sum_{i \in \mathcal{N}_{\mathrm{geo}}} \|D\widehat{\psi}_i\| \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2(\mathbb{R}^d)} \leq c_1. \tag{13.4}$$

Then $\boldsymbol{T}$ is a C^1-diffeomorphism and the following holds true for all $\widehat{\boldsymbol{x}} \in \widehat{K}$:

$$\|D\boldsymbol{T}\| \leq (1+c_1)\|D\widetilde{\boldsymbol{T}}\|, \tag{13.5}$$

$$\|D(\boldsymbol{T}^{-1})\| \leq (1-c_1)^{-1}\|(D\widetilde{\boldsymbol{T}})^{-1}\| \tag{13.6}$$

$$(1-c_1)^d|\det(D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}))| \leq |\det(D\boldsymbol{T}(\widehat{\boldsymbol{x}}))| \leq (1+c_1)^d|\det(D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}))|. \tag{13.7}$$

Proof. This is Theorem 3 in [78]. The definition of $\boldsymbol{T}$ in (13.2) implies that

$$D\boldsymbol{T}(\widehat{\boldsymbol{x}}) = D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}}) + \boldsymbol{E}(\widehat{\boldsymbol{x}}) = D\widetilde{\boldsymbol{T}}(\widehat{\boldsymbol{x}})(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}(\widehat{\boldsymbol{x}})\boldsymbol{E}(\widehat{\boldsymbol{x}})),$$

with $\boldsymbol{E}(\widehat{\boldsymbol{x}})(\boldsymbol{\xi}) := \sum_{i \in \mathcal{N}_{\text{geo}}} D\widehat{\psi}_i(\widehat{\boldsymbol{x}})(\boldsymbol{\xi})(\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i)$. Owing to the assumption (13.4), we infer that $\|(D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E}\| \leq c_1 < 1$. This immediately implies that the mapping $\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}(\widehat{\boldsymbol{x}})\boldsymbol{E}(\widehat{\boldsymbol{x}})$ is invertible, i.e., $D\boldsymbol{T}(\widehat{\boldsymbol{x}})$ is invertible and

$$\|D\boldsymbol{T}\| = \|D\widetilde{\boldsymbol{T}}(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})\| \leq (1+c_1)\|D\widetilde{\boldsymbol{T}}\|,$$

$$\|(D\boldsymbol{T})^{-1}\| = \|(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})^{-1}(D\widetilde{\boldsymbol{T}})^{-1}\| \leq (1-c_1)^{-1}\|(D\widetilde{\boldsymbol{T}})^{-1}\|.$$

Since $\det(D\boldsymbol{T}) = \det(D\widetilde{\boldsymbol{T}})\det(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})$ and $\|\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E}\| \leq 1 + c_1$, the upper bound in (13.7) results from

$$|\det(D\boldsymbol{T})| = |\det(D\widetilde{\boldsymbol{T}})| \times |\det(\boldsymbol{I} + (D\widetilde{\boldsymbol{T}})^{-1}\boldsymbol{E})| \leq |\det(D\widetilde{\boldsymbol{T}})|(1+c_1)^d,$$

where we used that $|\det(\boldsymbol{A})| \leq \|\boldsymbol{A}\|^d$ for any endomorphism in $\mathbb{R}^d$. The lower bound is shown similarly. $\square$

Remark 13.4 (Regularity of $\widehat{P}_{\text{geo}}$). In practice, the smoothness assumption $\widehat{P}_{\text{geo}} \subset C^1(\widehat{K};\mathbb{R})$ is satisfied since $\widehat{P}_{\text{geo}}$ is usually composed of smooth (polynomial) functions; see (8.1). $\square$

We now bound the higher-order derivatives of $\boldsymbol{T}$ and $\boldsymbol{T}^{-1}$.

Lemma 13.5 (Higher-order derivatives). *Assume* (13.4). *Assume that there is an integer $k \geq 1$ s.t. $\widetilde{\boldsymbol{T}}$ is C^{k+1}-diffeomorphism and that $\widehat{P}_{\text{geo}} \subset C^{k+1}(\widehat{K};\mathbb{R})$. Then $\boldsymbol{T}$ is a C^{k+1}-diffeomorphism. Moreover, assume that there are real numbers $c_2, \ldots, c_{k+1}$ s.t.*

$$\|D^m\boldsymbol{T}\| \leq c_m\|D\widetilde{\boldsymbol{T}}\|, \qquad \forall m \in \{2{:}k{+}1\}. \tag{13.8}$$

Let $\kappa := \|D\widetilde{\boldsymbol{T}}\|\|D(\widetilde{\boldsymbol{T}}^{-1})\|$ and c_1 be defined in (13.4). *Then for every integer $m \in \{2{:}k{+}1\}$, there is c_{-m} depending on $\kappa, c_1, \ldots, c_m$ s.t.*

$$\|D^m(\boldsymbol{T}^{-1})\| \leq c_{-m}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m. \tag{13.9}$$

Proof. This is Theorem 4 in [78]. The assumption $\widehat{P}_{\text{geo}} \subset C^{k+1}(\widehat{K};\mathbb{R})$ implies that $\boldsymbol{T}$ is of class C^{k+1}, and it has already been established in Lemma 13.3 that $\boldsymbol{T}$ is a diffeomorphism since (13.4) holds true. Let us prove (13.9) for $m = 2$. Using the chain rule (see Lemma B.4) and the identity $\boldsymbol{T}^{-1}(\boldsymbol{T}(\widehat{\boldsymbol{x}})) =$

$\widehat{\boldsymbol{x}}$, we infer that

$$D^2(\boldsymbol{T}^{-1})(\boldsymbol{h}_1, \boldsymbol{h}_2) = -D(\boldsymbol{T}^{-1})(D^2\boldsymbol{T}((D\boldsymbol{T})^{-1}(\boldsymbol{h}_1), (D\boldsymbol{T})^{-1}(\boldsymbol{h}_2))),$$

for all $\boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^d$. Using that $\|D(\boldsymbol{T}^{-1})\| = \|(D\boldsymbol{T})^{-1}\|$, this implies that

$$\|D^2(\boldsymbol{T}^{-1})\| \leq \|D(\boldsymbol{T}^{-1})\|\|D^2\boldsymbol{T}\|\|(D\boldsymbol{T})^{-1}\|^2 = \|D^2\boldsymbol{T}\|\|D(\boldsymbol{T}^{-1})\|^3.$$

Owing to (13.6) and (13.8), we infer that $\|D^2(\boldsymbol{T}^{-1})\| \leq c_{-2}\kappa\|D(\widetilde{\boldsymbol{T}}^{-1})\|^2$ with $c_{-2} = c_2(1 - c_1)^{-3}$. The rest of the proof is left as an exercise. $\qquad\square$

13.3 Interpolation error on nonaffine meshes

The goal of this section is to establish approximation properties of the finite element defined in Proposition 9.2 using the transformation $\psi_K(v) := v \circ \boldsymbol{T}$, where the nonaffine geometric mapping $\boldsymbol{T}$ is defined in (13.2) as a perturbation of the geometric mapping $\widetilde{\boldsymbol{T}}$. We adopt the same notation and definitions as in §13.2. We proceed as in Chapter 11: we first study how the Sobolev norms are transformed by $\boldsymbol{T}$ and then we apply these results to bound the interpolation error.

13.3.1 Transformation of Sobolev norms

We first state a result on the comparison of Sobolev norms. The main difference with respect to the affine case (see Lemma 11.7) is that now the full Sobolev norm, and not only the seminorm, appears on the right-hand side. To be dimensionally consistent, we consider the Sobolev norm $\|v\|_{W^{l,p}(K)} := \left(\sum_{m \in \{0:l\}} \ell_D^{mp} |v|_{W^{m,p}(K)}^p\right)^{\frac{1}{p}}$ (recall that $\ell_D := \text{diam}(D)$). Recall that the usual seminorm in $W^{m,p}(K)$ involves all the derivatives, i.e., $|v|_{W^{m,p}(K)} := \left(\sum_{|\alpha|=m} \|\partial^\alpha v\|_{L^p(K)}^p\right)^{\frac{1}{p}}$, and that this seminorm is equivalent to $\||D^m v|_{\mathbb{P}}\|_{L^p(K)}$, where $|D^m v|_{\mathbb{P}}$ is defined in (13.3). When analyzing geometric mappings based on $\mathbb{Q}_{k,d}$ Lagrange elements, it is useful to consider a different seminorm which does not include the mixed derivatives. For every map $A \in C^m(\widehat{K}; \mathbb{R}^q)$, $q \geq 1$, we define the following seminorms:

$$|D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} := \max_{\widehat{\boldsymbol{e}} \in \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}} \|D^m A(\widehat{\boldsymbol{x}})(\widehat{\boldsymbol{e}}, \ldots, \widehat{\boldsymbol{e}})\|_{\ell^2(\mathbb{R}^q)}, \qquad (13.10)$$

where $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ is the canonical Cartesian basis of $\mathbb{R}^d$. The notation for the subscript is motivated by the fact that for all $k \geq 0$, $|D^{k+1} A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} = 0$ for all $\widehat{\boldsymbol{x}} \in \widehat{K}$ if and only if A is $[\mathbb{Q}_{k,d}]^q$-valued (see Exercise 13.2). Notice that (13.10) defines only a seminorm and that $|D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \leq |D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{P}}$ for all $\widehat{\boldsymbol{x}} \in \widehat{K}$. We then introduce the associated seminorm over $W^{m,p}(\widehat{K}; \mathbb{R}^q)$,

$$\llbracket A \rrbracket_{W^{m,p}(\widehat{K})} := \| \, |D^m A(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \|_{L^p(\widehat{K})}. \tag{13.11}$$

The seminorm $\llbracket A \rrbracket_{W^{m,p}(\widehat{K})}$ involves only the pure partial derivatives of A of order m, i.e., the mixed derivatives are not involved.

Lemma 13.6 (Norm scaling by pullback). *Let $\widetilde{\boldsymbol{T}}, \boldsymbol{T}$ be defined in* (13.2). *Let the integer $k \geq 1$ satisfy the assumptions of Lemma* 13.5, *i.e.,* $\widetilde{\boldsymbol{T}}$ *is a C^{k+1}-diffeomorphism, $\widehat{P}_{\mathrm{geo}} \subset C^{k+1}(\widehat{K}; \mathbb{R})$, and* (13.8) *holds true. Assume that there are constants $c'_2, \ldots, c'_{k+1}$ s.t.*

$$\text{either} \quad \| D^m \boldsymbol{T} \| \leq c'_m \ell_D^{1-m} \| D\widetilde{\boldsymbol{T}} \|^m, \quad \forall m \in \{2{:}k{+}1\}, \tag{13.12}$$

$$\text{or} \qquad \llbracket D^m \boldsymbol{T} \rrbracket \leq c'_m \ell_D^{1-m} \| D\widetilde{\boldsymbol{T}} \|^m, \quad \forall m \in \{2{:}k{+}1\}, \tag{13.13}$$

recalling the shorthand notation $\| D^m \boldsymbol{T} \| := \| \, |D^m \boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{P}} \|_{L^\infty(\widehat{K})}$ and defining similarly $\llbracket D^m \boldsymbol{T} \rrbracket := \| \, |D^m \boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \|_{L^\infty(\widehat{K})}$. Then for every integer $l \geq 0$ and all $p \in [1, \infty]$, there is c, depending only on κ, $c_1, \ldots, c_{k+1}$, $c'_2, \ldots, c'_{k+1}$, p, and $\widehat{K}$, s.t. the following holds true for all $v \in W^{l,p}(K)$ with $K := \boldsymbol{T}(\widehat{K})$:

$$\text{either } \ell_D^l |v \circ \boldsymbol{T}|_{W^{l,p}(\widehat{K})} \leq c \, \| \det(D\widetilde{\boldsymbol{T}})^{-1} \|_{L^\infty(\widehat{K})}^{\frac{1}{p}} \| D\widetilde{\boldsymbol{T}} \|^l \| v \|_{W^{l,p}(K)}, \tag{13.14}$$

$$\text{or} \quad \ell_D^l \llbracket v \circ \boldsymbol{T} \rrbracket_{W^{l,p}(\widehat{K})} \leq c \, \| \det(D\widetilde{\boldsymbol{T}})^{-1} \|_{L^\infty(\widehat{K})}^{\frac{1}{p}} \| D\widetilde{\boldsymbol{T}} \|^l \| v \|_{W^{l,p}(K)}, \tag{13.15}$$

and

$$|v|_{W^{l,p}(K)} \leq c \, \| \det(D\widetilde{\boldsymbol{T}}) \|_{L^\infty(\widehat{K})}^{\frac{1}{p}} \| D(\widetilde{\boldsymbol{T}}^{-1}) \|^l \| v \circ \boldsymbol{T} \|_{W^{l,p}(\widehat{K})}. \tag{13.16}$$

Proof. Proof of (13.14). Assume first that $l \geq 2$. Using the chain rule (see Lemma B.4) together with the assumption (13.12), we infer that

$$|D^l(v \circ \boldsymbol{T})(\widehat{\boldsymbol{x}})|_{\mathbb{P}} \leq c \sum_{m \in \{1{:}l\}} |(D^m v)(\boldsymbol{T}(\widehat{\boldsymbol{x}}))|_{\mathbb{P}} \sum_{|r|=l} |D^{r_1}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{P}} \ldots |D^{r_m}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{P}}$$
$$\leq c \, \ell_D^{-l} \| D\widetilde{\boldsymbol{T}} \|^l \sum_{m \in \{1{:}l\}} \ell_D^m |(D^m v)(\boldsymbol{T}(\widehat{\boldsymbol{x}}))|_{\mathbb{P}},$$

for all $\widehat{\boldsymbol{x}} \in \widehat{K}$, where $|r| := r_1 + \ldots + r_m$ and c is generic constant having the same dependencies as in the assertion. Raising to the power p, integrating over $\widehat{K}$, changing variables on the right-hand side so as to integrate over K, and since $|v \circ \boldsymbol{T}|_{W^{l,p}(\widehat{K})} \leq c \| \, |D^l(v \circ \boldsymbol{T})(\widehat{\boldsymbol{x}})|_{\mathbb{P}} \|_{L^p(K)}$, we infer that

$$|v \circ \boldsymbol{T}|_{W^{l,p}(\widehat{K})}^p \leq c \, \ell_D^{-pl} \| D\widetilde{\boldsymbol{T}} \|^{pl} \sum_{m \in \{1{:}l\}} \ell_D^{pm} \int_K |D^m v(\boldsymbol{x})|_{\mathbb{P}}^p |\det(D\boldsymbol{T}^{-1}(\boldsymbol{x}))| \, \mathrm{d}x.$$

We conclude the proof of (13.14) for $l \geq 2$ using the estimate (13.7) on the determinant. The proof for $l = 0$ is evident. The proof for $l = 1$ can be done

as above by using (13.5) instead of (13.12). The proof of (13.15) is similar once one realizes that the chain rule preserves the pure derivatives of $\boldsymbol{T}$, i.e.,

$$|D^l(v \circ \boldsymbol{T})(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \le c \sum_{m \in \{1:l\}} |(D^m v)(\boldsymbol{T}(\widehat{\boldsymbol{x}}))|_{\mathbb{P}} \sum_{|r|=l} |D^{r_1}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{Q}} \ldots |D^{r_m}\boldsymbol{T}(\widehat{\boldsymbol{x}})|_{\mathbb{Q}}.$$

The estimate (13.16) is derived similarly by using the bound (13.9). $\square$

Remark 13.7 (Assumption (13.8)**).** Since $\|D\widetilde{\boldsymbol{T}}\|$ is proportional to the diameter of the cell generated by the nodes $\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_{n_{\text{geo}}}\}$, it is reasonable to assume that $(\ell_D^{-1}\|D\widetilde{\boldsymbol{T}}\|)^{m-1} \le 1$ if the diameter of K is small enough. In this case, the assumption (13.12) implies (13.8). In conclusion, the bound (13.8) from Lemma 13.5 has to be included in the assumptions of Lemma 13.6 only when invoking the assumption (13.13) (pure-derivatives case). $\square$

13.3.2 Bramble–Hilbert lemmas in $\mathbb{Q}_{k,d}$

We now formulate the Bramble–Hilbert lemma for polynomials in $\mathbb{Q}_{k,d}$ (this is the counterpart of Lemma 11.9 stated for polynomials in $\mathbb{P}_{k,d}$). It is at this stage that the seminorm based on pure derivatives enters the analysis.

Lemma 13.8 ($\mathbb{Q}_k$-Bramble–Hilbert). *Let S be a Lipschitz domain in $\mathbb{R}^d$. Let $p \in [1, \infty]$. Let $k \ge 0$ be an integer. There is c such that*

$$\inf_{q \in \mathbb{Q}_{k,d}} \|v + q\|_{W^{k+1,p}(S)} \le c \, [\![v]\!]_{W^{k+1,p}(S)}. \tag{13.17}$$

Corollary 13.9 ($\mathbb{Q}_k$-Bramble–Hilbert for linear functionals). *Under the hypotheses of Lemma 13.8, there is c such that the following holds true for all $g \in W^{k+1,p}(S)'$ vanishing on $\mathbb{Q}_{k,d}$:*

$$|g(v)| \le c \, \|g\|_{(W^{k+1,p}(S))'} [\![v]\!]_{W^{k+1,p}(S)}, \qquad \forall v \in W^{k+1,p}(S). \tag{13.18}$$

Proof. The estimate (13.17) is proved in Bramble and Hilbert [41, Thm. 1]. The estimate (13.18) is proved in [41, Thm. 2]. $\square$

13.3.3 Interpolation error estimates

We are now in the position to present the main result of this section.

Theorem 13.10 (Local interpolation). *Let $\widetilde{\boldsymbol{T}}, \boldsymbol{T}$ be defined in (13.2) and let $K := \boldsymbol{T}(\widehat{K})$. Let $p \in [1, \infty]$. Let the integer $k \ge 1$ satisfy the assumptions of Lemma 13.5. Assume that*

$$\text{either} \quad (13.12) \text{ holds and } \mathbb{P}_{k,d} \subset \widehat{P} \subset W^{k+1,p}(\widehat{K}) \hookrightarrow V(\widehat{K}), \tag{13.19}$$

$$\text{or} \quad (13.13) \text{ holds and } \mathbb{Q}_{k,d} \subset \widehat{P} \subset W^{k+1,p}(\widehat{K}) \hookrightarrow V(\widehat{K}). \tag{13.20}$$

(Recall that (13.12)–(13.13) are the hypotheses of Lemma 13.6.) Let $\mathcal{I}_K$ be the interpolation operator defined in (9.6). Let $l \in \{1{:}k+1\}$ be an integer s.t. $W^{l,p}(\widehat{K}) \hookrightarrow V(\widehat{K})$. Let

$$\lambda := \|\det(D\widetilde{\boldsymbol{T}})\|_{L^\infty(\widehat{K})}\|\det(D\widetilde{\boldsymbol{T}})^{-1}\|_{L^\infty(\widehat{K})}, \qquad \kappa := \|D\widetilde{\boldsymbol{T}}\|\|D(\widetilde{\boldsymbol{T}}^{-1})\|.$$

There is c, only depending on κ, $c_1, \ldots, c_{k+1}$, $c_2', \ldots, c_{k+1}'$, p, and $\widehat{K}$, s.t. for all $v \in W^{l,p}(K)$ and all $m \in \{0{:}l\}$,

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K)} \le c\,\lambda^{\frac{1}{p}}\kappa^m \ell_D^{-l}\|D\widetilde{\boldsymbol{T}}\|^{l-m}\|v\|_{W^{l,p}(K)}. \tag{13.21}$$

Proof. (1) Let us prove the assertion assuming (13.19). Using (13.16) from Lemma 13.6 and the commuting property $\mathcal{I}_K(w) \circ \boldsymbol{T} = \mathcal{I}_{\widehat{K}}(w \circ \boldsymbol{T})$ (see Proposition 9.3), we infer that

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K)} \le c\|\det(D\widetilde{\boldsymbol{T}})\|_{L^\infty(\widehat{K})}^{\frac{1}{p}}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m\|\widehat{v} - \mathcal{I}_{\widehat{K}}(\widehat{v})\|_{W^{m,p}(\widehat{K})},$$

with $\widehat{v} := v \circ \boldsymbol{T}$. Just like in the proof of Theorem 11.13, the assumptions (13.19) imply that there is c s.t. $\|\widehat{v} - \mathcal{I}_{\widehat{K}}(\widehat{v})\|_{W^{m,p}(\widehat{K})} \le c\,|\widehat{v}|_{W^{l,p}(\widehat{K})}$. This, together with (13.14), proves the claim since

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K)} \le c\,\|\det(D\widetilde{\boldsymbol{T}})\|_{L^\infty(\widehat{K})}^{\frac{1}{p}}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m|\widehat{v}|_{W^{l,p}(\widehat{K})}$$

$$\le c\,\|\det(D\widetilde{\boldsymbol{T}})\|_{L^\infty(\widehat{K})}^{\frac{1}{p}}\|\det(D\widetilde{\boldsymbol{T}})^{-1}\|_{L^\infty(\widehat{K})}^{\frac{1}{p}}\|D(\widetilde{\boldsymbol{T}}^{-1})\|^m\|D\widetilde{\boldsymbol{T}}\|^l\ell_D^{-l}\|v\|_{W^{l,p}(K)}.$$

(2) The only change in the above argument when proving (13.21) assuming (13.20) is that $\|\widehat{v} - \mathcal{I}_{\widehat{K}}(\widehat{v})\|_{W^{m,p}(\widehat{K})} \le c\,[\![\widehat{v}]\!]_{W^{l,p}(\widehat{K})}$ owing to the (13.17) from the Bramble–Hilbert lemma. We then conclude using (13.15). $\qquad\square$

Remark 13.11 (Key assumptions). The key assumptions to be verified for Theorem 13.10 to hold are either (13.4) and (13.12) for $\mathbb{P}_k$-based geometric mappings or (13.4), (13.8), and (13.13) for $\mathbb{Q}_k$-based geometric mappings. Of course, the above theory makes sense only for meshes for which the numbers λ, κ, $c_1, \ldots, c_{k+1}$, $c_2', \ldots, c_{k+1}'$ are uniformly bounded with respect to $K \in \mathcal{T}_h$ and $h \in \mathcal{H}$. To ensure that these numbers are uniformly bounded, it is necessary to make assumptions on the geometric mappings $\widetilde{\boldsymbol{T}}$ and on how far the geometric nodes $\boldsymbol{a}_i$ lie from $\widetilde{\boldsymbol{a}}_i$. Examples are given in the forthcoming sections. Notice that the theory can be applied with $\boldsymbol{T} = \widetilde{\boldsymbol{T}}$ as will be demonstrated in §13.5 for $\mathbb{Q}_1$-quadrangles. In this case, (13.4) is trivial to verify. $\qquad\square$

Remark 13.12 (Extensions). Generalizations of the above ideas can be found in Bernardi [20], Brenner and Scott [47, §4.7], Ciarlet [76, §4.3-4.4], Ciarlet [76], Lenoir [132], and Zlámal [203, 204]. $\qquad\square$

13.4 Curved simplices

Let us now describe how the above technique can be applied with curved $\mathbb{P}_2$-simplices, i.e., we set $k := 2$. Let us assume for the time being that we have at hand a mesh $\widetilde{\mathcal{T}}_h$ composed of affine simplices. Let $\widetilde{K} \in \widetilde{\mathcal{T}}_h$. Lemma 11.1 implies that

$$\frac{\rho_{\widetilde{K}}}{h_{\widehat{K}}} \leq \|D\widetilde{\boldsymbol{T}}\| \leq \frac{h_{\widetilde{K}}}{\rho_{\widehat{K}}}, \qquad \frac{\rho_{\widehat{K}}}{h_{\widetilde{K}}} \leq \|D(\widetilde{\boldsymbol{T}}^{-1})\| \leq \frac{h_{\widehat{K}}}{\rho_{\widetilde{K}}}. \tag{13.22}$$

Let us define $\rho_K := \rho_{\widetilde{K}}$ and $h_K := h_{\widetilde{K}}$. Assume that the simplicial mesh sequence $(\widetilde{\mathcal{T}}_h)_{h \in \mathcal{H}}$ is shape-regular (see Definition 11.2), i.e., there is $\sigma_\sharp$ s.t. $\sigma_{\widetilde{K}} = \frac{h_{\widetilde{K}}}{\rho_{\widetilde{K}}} \leq \sigma_\sharp$ for all $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ and all $h \in \mathcal{H}$. Let us consider one element $\widetilde{K} \in \widetilde{\mathcal{T}}_h$, and let $\{\widetilde{\boldsymbol{a}}_i\}_{i \in \mathcal{N}_{\text{geo}}}$ be its geometric nodes. Assume now that by means of some algorithm (see, e.g., Example 13.1), we construct the points $\{\boldsymbol{a}_i\}_{i \in \mathcal{N}_{\text{geo}}}$ from the set $\{\widetilde{\boldsymbol{a}}_i\}_{i \in \mathcal{N}_{\text{geo}}}$ and define the corresponding cell K using (13.2); see Figure 13.1 in dimension $d = 2$ where $n_{\text{geo}} := 6$. Let $\ell_D := \text{diam}(D)$. Assume that this construction is done so that there is a constant c_a such that for all $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ and all $h \in \mathcal{H}$,

$$\max_{i \in \mathcal{N}_{\text{geo}}} \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2(\mathbb{R}^d)} \leq c_a \ell_D^{-1} h_K^2. \tag{13.23}$$

This assumption is reasonable if the midpoint on each edge is constructed as explained in Example 13.1.

The key assumptions to be verified for Theorem 13.10 to hold true are (13.4) and (13.12). Using $\|D(\widetilde{\boldsymbol{T}}^{-1})\| \leq h_{\widehat{K}} \rho_K^{-1} \leq \sigma_\sharp h_{\widehat{K}} h_{\widetilde{K}}^{-1} = \sigma_\sharp h_{\widehat{K}} h_K^{-1}$, we observe that the left-hand side of (13.4) can be bounded by the factor $(c_a h_{\widehat{K}} \sum_{i \in \mathcal{N}_{\text{geo}}} \|D\widehat{\psi}_i\|)\sigma_\sharp \frac{h_K}{\ell_D}$ which is less than 1 for $\frac{h_K}{\ell_D}$ small enough, i.e., (13.4) holds true for $\frac{h_K}{\ell_D}$ small enough. Using that $D^2\widetilde{\boldsymbol{T}} = 0$ and $h_K \leq \sigma_\sharp h_{\widehat{K}} \|D\widetilde{\boldsymbol{T}}\|$ owing to (13.22), we infer that

$$\|D^2\boldsymbol{T}\| \leq \left(\sum_{i \in \mathcal{N}_{\text{geo}}} \|D^2\widehat{\psi}_i\| \right) \max_{i \in \mathcal{N}_{\text{geo}}} \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2}$$

$$\leq \left(c_a h_{\widehat{K}}^2 \sigma_\sharp^2 \sum_{i \in \mathcal{N}_{\text{geo}}} \|D^2\widehat{\psi}_i\| \right) \ell_D^{-1} \|D\widetilde{\boldsymbol{T}}\|^2.$$

Notice also that $\|D^3\boldsymbol{T}\| = 0$. Hence, (13.12) holds true for all $k \geq 1$. Moreover, since $\widetilde{\boldsymbol{T}}$ is affine, $\lambda = 1$ and $\kappa \leq \sigma_\sharp \frac{h_{\widehat{K}}}{\rho_{\widehat{K}}}$. In conclusion, assuming that the mesh sequence $(\widetilde{\mathcal{T}}_h)_{h \in \mathcal{H}}$ is shape-regular, there exists $h_0 > 0$ so that Theorem 13.10 implies that there is c (depending on the shape-regularity parameter $\sigma_\sharp$) such that for all $p \in [1, \infty]$, all $K \in \mathcal{T}_h$, all $h \in \mathcal{H} \cap (0, h_0)$, every integer $l \in \{0{:}k+1\}$ s.t. $W^{l,p}(K) \hookrightarrow V(K)$ (i.e., $W^{l,p}(K)$ is in the domain of $\mathcal{I}_K$),

all $v \in W^{l,p}(K)$, and every integer $m \in \{0{:}l\}$,

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K)} \le c\,\ell_D^{-l}\,h_K^{l-m}\|v\|_{W^{l,p}(K)}. \qquad (13.24)$$

Remark 13.13 (Extensions). An algorithm that constructs $\mathbb{P}_3$-simplices in dimension two is described in Ciarlet and Raviart [78, p. 240], Ciarlet [77, §4.3], Ciarlet [76, p. 247]. An algorithm that constructs curved simplices of any order in any dimension and that satisfies the assumptions of the perturbation theory in §13.2 is described in Lenoir [132]. It is a recursive technique based on the following principle: the construction of curved $\mathbb{P}_{m+1}$ simplices that approximate the boundary with $\mathcal{O}(h^{m+2})$ accuracy relies on the existence of a construction technique of curved $\mathbb{P}_m$ simplices that approximate the boundary with $\mathcal{O}(h^{m+1})$ accuracy, $m \ge 1$. $\square$

13.5 $\mathbb{Q}_1$-quadrangles

Let us now consider a mesh where all the cells are nondegenerate convex quadrangles in $\mathbb{R}^2$. All the cells can be generated from the unit square $\widehat{K} := [0,1]^2$ using geometric mappings $\boldsymbol{T} \in [\mathbb{Q}_1(\widehat{K})]^2$; see Figure 13.2. $\boldsymbol{T}$ maps the edges of $\widehat{K}$ to the edges of K, but unless K is a parallelogram, $\boldsymbol{T}$ is not affine. We are going to apply the theory from §13.2 with $K = \widetilde{K}$, $\boldsymbol{T} = \widetilde{\boldsymbol{T}}$, and $\boldsymbol{a}_i = \widetilde{\boldsymbol{a}}_i$, for all $i \in \{1{:}4\}$.
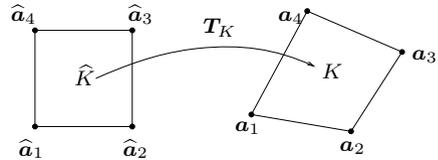
Fig. 13.2 Nonaffine mapping from the unit square to a quadrangle.

Upon identifying the points $\boldsymbol{a}_{i \in \{1:4\}}$ with column vectors and $D\boldsymbol{T}$ with the Jacobian matrix, a simple computation shows that

$$D\boldsymbol{T}(\widehat{\boldsymbol{x}})$$
$$= (\boldsymbol{a}_2 - \boldsymbol{a}_1 + \widehat{x}_2(\boldsymbol{a}_3 - \boldsymbol{a}_4 + \boldsymbol{a}_1 - \boldsymbol{a}_2), \boldsymbol{a}_4 - \boldsymbol{a}_1 + \widehat{x}_1(\boldsymbol{a}_3 - \boldsymbol{a}_4 + \boldsymbol{a}_1 - \boldsymbol{a}_2))$$
$$= ((1 - \widehat{x}_2)(\boldsymbol{a}_2 - \boldsymbol{a}_1) + \widehat{x}_2(\boldsymbol{a}_3 - \boldsymbol{a}_4), (1 - \widehat{x}_1)(\boldsymbol{a}_4 - \boldsymbol{a}_1) + \widehat{x}_1(\boldsymbol{a}_3 - \boldsymbol{a}_2)),$$

for all $\widehat{\boldsymbol{x}} := (\widehat{x}_1, \widehat{x}_2) \in \widehat{K}$. It follows from the first equality that $\det(D\boldsymbol{T}(\widehat{\boldsymbol{x}}))$ is in $\mathbb{P}_1$, implying that $\max_{\widehat{\boldsymbol{x}} \in \widehat{K}} |\det(D\boldsymbol{T}(\widehat{\boldsymbol{x}}))| = \max_{i \in \{1:4\}} |\det(D\boldsymbol{T}(\widehat{\boldsymbol{a}}_i))|$ since we assumed that K is convex. Let P_i be the parallelogram formed by $\boldsymbol{a}_{i-1}$, $\boldsymbol{a}_i$, $\boldsymbol{a}_{i+1}$ (with the convention $\boldsymbol{a}_0 := \boldsymbol{a}_4$ and $\boldsymbol{a}_5 := \boldsymbol{a}_1$). It can be verified that $\det(D\boldsymbol{T}(\widehat{\boldsymbol{a}}_i)) = |P_i|$. As a result, letting $S_{\min} := \min_{i \in \{1:4\}} |P_i|$, $S_{\max} := \max_{i \in \{1:4\}} |P_i|$, we infer that

$$\|\det(D\boldsymbol{T})\|_{L^\infty(\widehat{K})} \leq S_{\max}, \qquad \|\det(D(\boldsymbol{T}^{-1}))\|_{L^\infty(\widehat{K})} \leq \frac{1}{S_{\min}}. \qquad (13.25)$$

Let $\boldsymbol{d}_1(\widehat{\boldsymbol{x}})$, $\boldsymbol{d}_2(\widehat{\boldsymbol{x}})$ be the columns of $D\boldsymbol{T}(\widehat{\boldsymbol{x}})$ and $\theta(\widehat{\boldsymbol{x}})$ be the angle formed by these two vectors. The vector $\boldsymbol{d}_1(\widehat{\boldsymbol{x}})$ is a convex combination of the sides $(\boldsymbol{a}_2 - \boldsymbol{a}_1)$ and $(\boldsymbol{a}_3 - \boldsymbol{a}_4)$, whereas the vector $\boldsymbol{d}_2(\widehat{\boldsymbol{x}})$ is a convex combination of the sides $(\boldsymbol{a}_4 - \boldsymbol{a}_1)$ and $(\boldsymbol{a}_3 - \boldsymbol{a}_2)$. The angle $\theta(\widehat{\boldsymbol{x}})$ takes its extreme values at the vertices of K, say $\theta_1, \ldots, \theta_4$. Let $\boldsymbol{h} = (h_1, h_2) \in \mathbb{R}^2$ with $\|\boldsymbol{h}\|_{\ell^2} = 1$. Then,

$$\begin{aligned}
\|D\boldsymbol{T}(\widehat{\boldsymbol{x}})(\boldsymbol{h})\|_{\ell^2}^2 &= h_1^2 \|\boldsymbol{d}_1\|_{\ell^2}^2 + h_2^2 \|\boldsymbol{d}_2\|_{\ell^2}^2 + 2h_1 h_2 \boldsymbol{d}_1 \cdot \boldsymbol{d}_2 \\
&\geq h_1^2 \|\boldsymbol{d}_1\|_{\ell^2}^2 + h_2^2 \|\boldsymbol{d}_2\|_{\ell^2}^2 - 2|h_1||h_2| \|\boldsymbol{d}_1\|_{\ell^2} \|\boldsymbol{d}_2\|_{\ell^2} |\cos(\theta)| \\
&\geq h_1^2 \|\boldsymbol{d}_1\|_{\ell^2}^2 (1 - |\cos(\theta)|) + (1 - h_1^2) \|\boldsymbol{d}_2\|_{\ell^2}^2 (1 - |\cos(\theta)|) \\
&\geq \min(\|\boldsymbol{d}_1\|_{\ell^2}^2, \|\boldsymbol{d}_2\|_{\ell^2}^2)(1 - |\cos(\theta)|),
\end{aligned}$$

where dependencies of $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, and θ on $\widehat{\boldsymbol{x}}$ have been omitted. Denoting by $h_{\min}$ the length of the smallest side of K and $\gamma := \max_{i \in \{1:4\}} |\cos(\theta_i)|$, we infer that $\|D\boldsymbol{T}(\widehat{\boldsymbol{x}})(\boldsymbol{y})\|_{\ell^2} \geq h_{\min}(1-\gamma)$ for all $\boldsymbol{y}$ with $\|\boldsymbol{y}\|_{\ell^2} = 1$ and all $\widehat{\boldsymbol{x}} \in \widehat{K}$, implying that $\|D(\boldsymbol{T}^{-1})\| = \|(D\boldsymbol{T})^{-1}\| \leq (h_{\min}(1 - \gamma))^{-1}$. By proceeding similarly, we also obtain that $\|D\boldsymbol{T}\| \leq 2h_{\max}$ and $[\![D^m\boldsymbol{T}]\!] = 0$ for $m \geq 2$, where $h_{\max}$ is the length of the largest side of K. In conclusion, we have

$$\|D(\boldsymbol{T}^{-1})\| \leq \frac{1}{h_{\min}(1 - \gamma)}, \quad \|D\boldsymbol{T}\| \leq 2h_{\max}, \quad [\![D^l\boldsymbol{T}]\!] = 0, \ \forall l \geq 2. \quad (13.26)$$

The key assumptions to be verified for Theorem 13.10 to hold true are (13.4), (13.8), and (13.13). Assumption (13.4) trivially holds since $\boldsymbol{T} = \widetilde{\boldsymbol{T}}$. Assumption (13.8) trivially holds for $m > 2$, and $\|D^2\boldsymbol{T}\| \leq \|\boldsymbol{a}_1 - \boldsymbol{a}_2 + \boldsymbol{a}_3 - \boldsymbol{a}_4\|_{\ell^2} \leq 2h_{\max} \leq \frac{2h_{\max}}{(1-\gamma)h_{\min}} \|D\boldsymbol{T}\|$. Assumption (13.13) holds since (13.26) implies that $[\![D^m\boldsymbol{T}]\!] = 0 \leq c'_m \ell_D^{1-m} \|D\boldsymbol{T}\|^m$ for all $m \geq 2$. Furthermore, owing to (13.25) we have $\lambda \leq \frac{S_{\max}}{S_{\min}}$, and owing to (13.26) we have $\kappa \leq \frac{2h_{\max}}{h_{\min}(1-\gamma)}$. These bounds show that a reasonable notion of shape-regularity for $\mathbb{Q}_1$-quadrangular meshes is to assume that there is $\sigma_\sharp$ such that for all $K \in \mathcal{T}_h$ and all $h \in \mathcal{H}$,

$$\max\left(\frac{S_{\max}(K)}{S_{\min}(K)}, \frac{2h_{\max}(K)}{h_{\min}(K)(1 - \gamma(K))}\right) \leq \sigma_\sharp. \qquad (13.27)$$

Then, if the $\mathbb{Q}_1$-quadrangular mesh sequence is shape-regular in the above sense, Theorem 13.10 implies that there is c (depending on the shape-regularity parameter $\sigma_\sharp$) such that for all $p \in [1, \infty]$, all $K \in \mathcal{T}_h$, all $h \in \mathcal{H}$, every integer $l \in \{0{:}k + 1\}$ s.t. $W^{l,p}(K) \hookrightarrow V(K)$ (i.e., $W^{l,p}(K)$ is in the domain of $\mathcal{I}_K$), all $v \in W^{l,p}(K)$, and every integer $m \in \{0{:}l\}$,

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K)} \leq c\, \ell_D^{-l} h_K^{l-m} \|v\|_{W^{l,p}(K)}. \qquad (13.28)$$

Remark 13.14 (Pure derivatives). Here, the critical assumption (13.13) hinges on the property $[\![D^2\boldsymbol{T}]\!] = 0$. This assumption would not have been true if we had used the full seminorm (involving the mixed derivative), since a simple computation shows that $[\![D^2\boldsymbol{T}]\!] = \|(\boldsymbol{a}_3 - \boldsymbol{a}_4) + (\boldsymbol{a}_1 - \boldsymbol{a}_2)\|_{\ell^2}$, yielding $\sqrt{2(1 + \cos(\theta_1 + \theta_4))h_{\min}^2 + (h_{\max} - h_{\min})^2} \leq [\![D^2\boldsymbol{T}]\!] \leq 2h_{\max}$, thereby showing that $\|D^2\boldsymbol{T}\| \sim \|D\boldsymbol{T}\|$ (unless K is a parallelogram). The term h_K^{l-m} in (13.28) would then be replaced by h_K^{1-m}, which would not give any convergence for $m = 1$. The reader is referred to Ciarlet and Raviart [79, pp. 245–247] and Girault and Raviart [107, p. 104] for more details.   □

13.6 $\mathbb{Q}_2$-curved quadrangles

We now describe how to construct meshes composed of $\mathbb{Q}_2$-curved quadrangles and how to bound the interpolation error on such meshes. Assume that we have at hand a sequence of $\mathbb{Q}_1$-quadrangular meshes $(\widetilde{\mathcal{T}}_h)_{h\in\mathcal{H}}$ that is shape-regular in the sense of §13.5, i.e., (13.27) holds true for all $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ and all $h \in \mathcal{H}$. Let $\widetilde{K} \in \widetilde{\mathcal{T}}_h$, let $\{\widetilde{\boldsymbol{a}}_1, \ldots, \widetilde{\boldsymbol{a}}_4\}$ be the vertices of $\widetilde{K}$, and let $\{\widetilde{\boldsymbol{a}}_5, \ldots, \widetilde{\boldsymbol{a}}_8\}$ be the midpoints of the four edges of $\widetilde{K}$; see Figure 13.1. Assume that the curved cell K is constructed by means of the technique explained in Example 13.1, i.e., we assume that the new points $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_8\}$ are positioned so that the following criterion is satisfied for all $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ and all $h \in \mathcal{H}$:

$$\max_{i \in \{1:8\}} \|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2(\mathbb{R}^2)} \leq c\, \ell_D^{-1} h_{\max}^2. \tag{13.29}$$

Let $\mathcal{T}_h$ be the mesh thus constructed for all $h \in \mathcal{H}$. Let us verify that the assumptions (13.4), (13.8), and (13.13) hold true for all $K \in \mathcal{T}_h$ and $h \in \mathcal{H}$. Starting with (13.4), we observe that $\|(D\widetilde{\boldsymbol{T}})^{-1}\| \leq (h_{\min}(1-\gamma))^{-1} \leq \frac{\sigma_\sharp}{2}h_{\max}^{-1}$ owing to (13.27), so that

$$\|(D\widetilde{\boldsymbol{T}})^{-1}\| \sum_{i \in \{1:8\}} \|D\widehat{\psi}_i\|\|\boldsymbol{a}_i - \widetilde{\boldsymbol{a}}_i\|_{\ell^2(\mathbb{R}^2)} \leq c\,\sigma_\sharp \ell_D^{-1} h_{\max}, \tag{13.30}$$

which is less than 1 provided the cells are small enough. Moreover, using the estimates $\|D^2\widetilde{\boldsymbol{T}}\| \leq 2h_{\max}$ (see Remark 13.14) and $\|D\widetilde{\boldsymbol{T}}\| \geq h_{\min}(1-\gamma)$, we infer that $\|D^2\boldsymbol{T}\| \leq \|D^2\widetilde{\boldsymbol{T}}\| + c\ell_D^{-1}h_{\max}^2 \leq c'h_{\max} \leq c'\sigma_\sharp\|D\widetilde{\boldsymbol{T}}\|$, which proves (13.8) for $m = 2$. Moreover, $\|D^m\boldsymbol{T}\| \leq c\ell_D^{-1}h_{\max}^2 \leq c'h_{\max} \leq c'\sigma_\sharp\|D\widetilde{\boldsymbol{T}}\|$ since $D^m\widetilde{\boldsymbol{T}} = 0$ for all $m \geq 3$, and this proves (13.8) for all $m \geq 3$. Furthermore, $[\![D^2\boldsymbol{T}]\!] \leq c\ell_D^{-1}h_{\max}^2 \leq c\ell_D^{-1}\sigma_\sharp^2\|D\widetilde{\boldsymbol{T}}\|^2$ since $[\![D^2\widetilde{\boldsymbol{T}}]\!] = 0$, and $[\![D^m\boldsymbol{T}]\!] = 0$ for all $m \geq 3$, and this proves (13.13). Finally, we have already seen that the quantities $\kappa := \|D\widetilde{\boldsymbol{T}}\|\|D(\widetilde{\boldsymbol{T}}^{-1})\|$ and $\lambda := \|\det(D\widetilde{\boldsymbol{T}})\|_{L^\infty(\widehat{K})}\|\det(D\widetilde{\boldsymbol{T}})^{-1}\|_{L^\infty(\widehat{K})}$ are bounded owing to the regularity of the mesh sequence. In conclusion, if the $\mathbb{Q}_1$-quadrangular mesh sequence $(\widetilde{\mathcal{T}}_h)_{h\in\mathcal{H}}$ is shape-regular, there exists

$h_0 > 0$ so that Theorem 13.10 implies that there is c (depending on the shape-regularity parameter $\sigma_\sharp$) such that for all $p \in [1, \infty]$, all $K \in \mathcal{T}_h$, all $h \in \mathcal{H} \cap (0, h_0)$, every integer $l \in \{0{:}k+1\}$ s.t. $W^{l,p}(K) \hookrightarrow V(K)$ (i.e., $W^{l,p}(K)$ is in the domain of $\mathcal{I}_K$), all $v \in W^{l,p}(K)$, and every integer $m \in \{0{:}l\}$,

$$|v - \mathcal{I}_K(v)|_{W^{m,p}(K)} \leq c\, \ell_D^{-l} h_K^{l-m} \|v\|_{W^{l,p}(K)}. \tag{13.31}$$

Exercises

Exercise 13.1 (Chain rule). Let $f \in \mathcal{C}^3(U; W_1)$ and $g \in \mathcal{C}^3(W_1; W_2)$, where V, W_1, W_2 are Banach spaces and U is an open set in V. (i) Evaluate the pure derivatives $D^2(g \circ f)(x)(h, h)$ and $D^3(g \circ f)(x)(h, h, h)$ for $x \in U$ and $h \in V$. (ii) Rewrite these expressions when f and g map from $\mathbb{R}$ to $\mathbb{R}$.

Exercise 13.2 (Pure derivatives, $\mathbb{Q}_{k,d}$-polynomials). Let $\{e_i\}_{i \in \{1:d\}}$ be the canonical Cartesian basis of $\mathbb{R}^d$. Let $k \geq 1$. Verify that $D^{k+1}q(e_i, \dots, e_i) = 0$ for all $i \in \{1{:}d\}$ if and only if $q \in \mathbb{Q}_{k,d}$. (*Hint*: by induction on d.) What is instead the characterization of polynomials in $\mathbb{P}_{k,d}$ in terms of $D^{k+1}q$?

Exercise 13.3 (Lemma 13.5). Complete the proof of Lemma 13.5 by proving (13.9) for all $m \leq k + 1$. (*Hint*: use induction on m and the chain rule formula (B.4) applied to $\boldsymbol{T}^{-1}(\boldsymbol{T}(\widehat{\boldsymbol{x}}))$.)

Exercise 13.4 (Tensor-product transformation). Assume the transformation $\boldsymbol{T}$ has the tensor-product form $\boldsymbol{T}(\widehat{\boldsymbol{x}}) = \sum_{j \in \{1:d\}} t_j(\widehat{\boldsymbol{x}}_j) e_j$ for some univariate function t_j, for all $j \in \{1{:}d\}$, where $\{e_j\}_{j \in \{1:d\}}$ is the canonical Cartesian basis of $\mathbb{R}^d$. (i) Show that (13.15) can be sharpened as $\llbracket w \circ \boldsymbol{T} \rrbracket_{W^{l,p}(\widehat{K})} \leq c \|\det(D\widetilde{\boldsymbol{T}})^{-1}\|_{L^\infty(\widehat{K})}^{\frac{1}{p}} \|D\widetilde{\boldsymbol{T}}\|^l \llbracket w \rrbracket_{W^{l,p}(K)}$. (*Hint*: recall that $\llbracket w \rrbracket_{W^{l,p}(K)}$ is a seminorm and there exists a uniform constant c so that $\ell_D^l \llbracket w \rrbracket_{W^{l,p}(K)} \leq c \|w\|_{W^{l,p}(K)}$.) (ii) What is the consequence of this new bound on the error estimate (13.21) under the assumption (13.20)?

Exercise 13.5 ($\mathbb{Q}_1$-quadrangles). Prove that $\det(D\boldsymbol{T}(\widehat{\boldsymbol{a}}_i)) = |P_i|$, where P_i is the parallelogram formed by $\boldsymbol{a}_{i-1}$, $\boldsymbol{a}_i$, $\boldsymbol{a}_{i+1}$ (with $\boldsymbol{a}_0 := \boldsymbol{a}_4$ and $\boldsymbol{a}_5 := \boldsymbol{a}_1$). (*Hint*: see §13.5.)

Exercise 13.6 (Butterfly subdivision algorithm). Consider a mesh composed of four triangles with the connectivity array such that j_geo$(1, 1{:}3) := (3, 4, 5)$, j_geo$(2, 1{:}3) := (0, 4, 5)$, j_geo$(3, 1{:}3) := (1, 3, 5)$, j_geo$(4, 1{:}3) := (2, 3, 4)$. Let $\boldsymbol{m}$ be the midpoint of the edge $(\boldsymbol{z}_3, \boldsymbol{z}_4)$. Let $\widehat{\boldsymbol{z}}_0 := (0, 0)$, $\widehat{\boldsymbol{z}}_1 := (1, 0)$, $\widehat{\boldsymbol{z}}_2 := (0, 1)$, $\widehat{\boldsymbol{z}}_3 := (\frac{1}{2}, \frac{1}{2})$, $\widehat{\boldsymbol{z}}_4 := (0, \frac{1}{2})$, $\widehat{\boldsymbol{z}}_5 := (\frac{1}{2}, 0)$. Consider now the curved triangle given by the $\mathbb{P}_2$ geometric mapping $\boldsymbol{T}$ that transforms $\widehat{\boldsymbol{z}}_i$ to $\boldsymbol{z}_i$ for all $i \in \{0{:}5\}$. Let $\{f_0, \dots, f_7\} \in \mathbb{R}$. Let $\widehat{p} \in \mathbb{P}_{2,2}$ be the polynomial defined by $\widehat{p}(\widehat{\boldsymbol{z}}_i) := f_i$ for all $i \in \{0{:}5\}$. (i) Compute $\widehat{p}(\boldsymbol{T}^{-1}(\boldsymbol{m}))$.

(ii) Consider two additional points $\boldsymbol{z}_6$, $\boldsymbol{z}_7$ and two more triangles given by $\mathtt{j\_geo}(5, 1{:}3) := (2, 3, 6)$, $\mathtt{j\_geo}(6, 1{:}3) := (2, 4, 7)$. Let $\boldsymbol{T}'$ be the $\mathbb{P}_2$ geometric mapping that transforms $\widehat{\boldsymbol{z}}_i$ to $\boldsymbol{z}_i$ for all $i \in \{2{:}7\}$. Let $\widehat{p}' \in \mathbb{P}_{2,2}$ be defined by $\widehat{p}'(\widehat{\boldsymbol{z}}_i) := f_i$ for all $i \in \{2{:}7\}$. Compute $\frac{1}{2}\big(\widehat{p}(\boldsymbol{T}^{-1}(\boldsymbol{m})) + \widehat{p}'((\boldsymbol{T}')^{-1}(\boldsymbol{m}))\big)$. *Note*: the name of the algorithm comes from the shape of the generic configuration. The algorithm is used for three-dimensional computer graphics. It allows the representation of smooth surfaces via the specification of coarser piecewise linear polygonal meshes. Given an initial polygonal mesh, a smooth surface is obtained by recursively applying the butterfly subdivision algorithm to the Cartesian coordinates of the vertices; see Dyn et al. [93].

Part III, Chapter 14

$\boldsymbol{H}(\mathrm{div})$ finite elements

The goal of this chapter is to construct $\mathbb{R}^d$-valued finite elements $(K, \boldsymbol{P}, \Sigma)$ with $d \geq 2$ such that (i) $\mathbb{P}_{k,d} := [\mathbb{P}_{k,d}]^d \subset \boldsymbol{P}$ for some $k \geq 0$ and (ii) the degrees of freedom (dofs) in Σ fully determine the normal components of the polynomials in $\boldsymbol{P}$ on all the faces of K. The first requirement is key for proving convergence rates on the interpolation error. The second one is key for constructing $\boldsymbol{H}(\mathrm{div})$-conforming finite element spaces (see Chapter 19). The finite elements introduced in this chapter are used, e.g., in Chapter 51 to approximate Darcy's equations which constitute a fundamental model for porous media flows. The focus here is on defining a reference element and generating finite elements on the mesh cells. The estimation of the interpolation error is done in Chapters 16 and 17. We detail the construction for the simplicial Raviart–Thomas finite elements. Some alternative elements are outlined at the end of the chapter.

14.1 The lowest-order case

We start by considering the lowest-order *Raviart–Thomas finite element*. Let $d \geq 2$ be the space dimension, and define the polynomial space

$$\mathbb{RT}_{0,d} := \mathbb{P}_{0,d} \oplus \boldsymbol{x}\, \mathbb{P}_{0,d}. \tag{14.1}$$

Since the above sum is indeed direct, $\mathbb{RT}_{0,d}$ is a vector space of dimension $\dim(\mathbb{RT}_{0,d}) = d + 1$. A basis of $\mathbb{RT}_{0,2}$ is $\left\{ \binom{1}{0}, \binom{0}{1}, \binom{x_1}{x_2} \right\}$. The space $\mathbb{RT}_{0,d}$ has several interesting properties. (a) One has $\mathbb{P}_{0,d} \subset \mathbb{RT}_{0,d}$ in agreement with the first requirement stated above. (b) If $\boldsymbol{v} \in \mathbb{RT}_{0,d}$ is divergence-free, then $\boldsymbol{v}$ is constant. (c) If H is an affine hyperplane of $\mathbb{R}^d$ with normal vector $\boldsymbol{\nu}_H$, then the function $\boldsymbol{v} \cdot \boldsymbol{\nu}_H$ is constant on H for all $\boldsymbol{v} \in \mathbb{RT}_{0,d}$. Writing $\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{a} + b\boldsymbol{x}$ with $\boldsymbol{a} \in \mathbb{R}^d$ and $b \in \mathbb{R}$, we indeed have $(\boldsymbol{v}(\boldsymbol{x}_1) - \boldsymbol{v}(\boldsymbol{x}_2)) \cdot \boldsymbol{\nu}_H = b(\boldsymbol{x}_1 - \boldsymbol{x}_2) \cdot \boldsymbol{\nu}_H = 0$ for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in H$.

Let K be a simplex in $\mathbb{R}^d$ and let $\mathcal{F}_K$ be the collection of the faces of K. Each face $F \in \mathcal{F}_K$ is oriented by a fixed unit normal vector $\boldsymbol{n}_F$, and we set $\boldsymbol{\nu}_F := |F|\boldsymbol{n}_F$. Let Σ be the collection of the following linear forms acting on $\mathbb{RT}_{0,d}$:

$$\sigma_F^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{\nu}_F)\, \mathrm{d}s, \qquad \forall F \in \mathcal{F}_K. \tag{14.2}$$

Since $\boldsymbol{v} \cdot \boldsymbol{\nu}_F$ is constant on F, $\sigma_F^{\mathrm{f}}(\boldsymbol{v}) = 0$ implies that $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F = 0$ in agreement with the second requirement stated above. Note that we could have written more simply $\sigma_F^{\mathrm{f}}(\boldsymbol{v}) := \int_F (\boldsymbol{v} \cdot \boldsymbol{n}_F)\, \mathrm{d}s$, but the expression (14.2) is introduced to be consistent with later notation. In any case, the unit of $\sigma_F^{\mathrm{f}}(\boldsymbol{v})$ is a surface times the dimension of $\boldsymbol{v}$. A graphic representation of the dofs is shown in Figure 14.1.
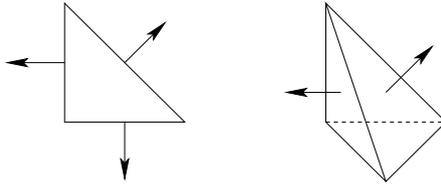
Fig. 14.1 $\mathbb{RT}_{0,d}$ finite element in dimensions two (left) and three (right). Only visible degrees of freedom are shown in dimension three. (The arrows have been drawn outward under the assumption that the vectors $\boldsymbol{\nu}_F$ point outward. The orientation of the arrows must be changed if some vectors $\boldsymbol{\nu}_F$ point inward.)

Proposition 14.1 (Finite element). $(K, \mathbb{RT}_{0,d}, \Sigma)$ *is a finite element.*

Proof. Since $\dim(\mathbb{RT}_{0,d}) = \operatorname{card}(\Sigma) = d + 1$, we just need to prove that the only function $\boldsymbol{v} \in \mathbb{RT}_{0,d}$ that annihilates the dofs in Σ is zero. Since $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F$ is constant and has zero mean-value on F, we have $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F = 0$ for all $F \in \mathcal{F}_K$. Moreover, the divergence theorem implies that $\int_K (\nabla \cdot \boldsymbol{v})\, \mathrm{d}x = \sum_{F \in \mathcal{F}_K} \int_F (\boldsymbol{v} \cdot \boldsymbol{n}_F)\, \mathrm{d}s = 0$. Since $\nabla \cdot \boldsymbol{v} \in \mathbb{P}_{0,d}$, we infer that $\nabla \cdot \boldsymbol{v}$ is zero, so that $\boldsymbol{v} \in \mathbb{P}_{0,d}$. Hence, $\boldsymbol{v} \cdot \boldsymbol{\nu}_F$ vanishes identically in K for all $F \in \mathcal{F}_K$. Since $\operatorname{span}\{\boldsymbol{\nu}_F\}_{F \in \mathcal{F}_K} = \mathbb{R}^d$ (see Exercise 7.3(iv)), we conclude that $\boldsymbol{v} = \boldsymbol{0}$. $\square$

Since the volume of a simplex is $|K| = \frac{1}{d}|F|h_F^\perp$ for all $F \in \mathcal{F}_K$ where $h_F^\perp$ is the height of K measured from the vertex $\boldsymbol{z}_F$ opposite to F, one readily verifies that the shape functions are

$$\boldsymbol{\theta}_F^{\mathrm{f}}(\boldsymbol{x}) := \frac{\iota_{F,K}}{d|K|}(\boldsymbol{x} - \boldsymbol{z}_F), \qquad \forall \boldsymbol{x} \in \mathbb{R}^d,\ \forall F \in \mathcal{F}_K, \tag{14.3}$$

where $\iota_{F,K} := 1$ if $\boldsymbol{\nu}_F$ points outward and $\iota_{F,K} := -1$ otherwise (i.e., $\iota_{F,K} = \boldsymbol{n}_F \cdot \boldsymbol{n}_K$ where $\boldsymbol{n}_K$ is the outward unit normal to K). The normal component of $\boldsymbol{\theta}_F^{\mathrm{f}}$ is constant on each of the $(d+1)$ faces of K (as expected), it is equal to

1 on F and to 0 on the other faces. See Exercise 14.1 for additional properties of the $\mathbf{RT}_{0,d}$ shape functions.

14.2 The polynomial space $\mathbf{RT}_{k,d}$

We now generalize the construction of §14.1 to an arbitrary polynomial order $k \in \mathbb{N}$. Let $d \geq 2$ be the space dimension. Recall from §7.3 the multi-index set $\mathcal{A}_{k,d} := \{\alpha \in \mathbb{N}^d \mid |\alpha| \leq k\}$ where $|\alpha| := \alpha_1 + \ldots + \alpha_d$. We additionally introduce the subset $\mathcal{A}_{k,d}^H := \{\alpha \in \mathcal{A}_{k,d} \mid |\alpha| = k\}$. For instance, $\mathcal{A}_{1,2} = \{(0,0),(1,0),(0,1)\}$ and $\mathcal{A}_{1,2}^H = \{(1,0),(0,1)\}$.

Definition 14.2 (Homogeneous polynomials). *A polynomial $p \in \mathbb{P}_{k,d}$ is said to be* homogeneous *of degree k if $p(\boldsymbol{x}) = \sum_{\alpha \in \mathcal{A}_{k,d}^H} a_\alpha \boldsymbol{x}^\alpha$ with real coefficients a_α. The real vector space composed of homogeneous polynomials is denoted by $\mathbb{P}_{k,d}^H$ or $\mathbb{P}_k^H$ when the context is unambiguous.*

Lemma 14.3 (Properties of $\mathbb{P}_{k,d}^H$). *We have $\boldsymbol{x}{\cdot}\nabla q = kq$ (Euler's identity) and $\nabla{\cdot}(\boldsymbol{x}q) = (k+d)q$ for all $q \in \mathbb{P}_{k,d}^H$.*

Proof. By linearity, it suffices to verify the assertion with $q(\boldsymbol{x}) := \boldsymbol{x}^\alpha$ for all $\alpha \in \mathcal{A}_{k,d}^H$. We have $\boldsymbol{x}{\cdot}\nabla q = \sum_{i \in \{1:d\}} \alpha_i x_i x_1^{\alpha_1} \ldots x_i^{\alpha_i - 1} \ldots x_d^{\alpha_d} = (\sum_{i \in \{1:d\}} \alpha_i)q = kq$. Moreover, the assertion for $\nabla{\cdot}(\boldsymbol{x}q)$ follows from the observation that $\nabla{\cdot}\boldsymbol{x} = d$ and $\nabla{\cdot}(\boldsymbol{x}q) = q\nabla{\cdot}\boldsymbol{x} + \boldsymbol{x}{\cdot}\nabla q$. $\qquad\square$

Definition 14.4 ($\mathbf{RT}_{k,d}$). *Let $k \in \mathbb{N}$ and let $d \geq 2$. We define the following real vector space of $\mathbb{R}^d$-valued polynomials:*

$$\mathbf{RT}_{k,d} := \mathbb{P}_{k,d} \oplus \boldsymbol{x}\,\mathbb{P}_{k,d}^H. \tag{14.4}$$

The above sum is direct since polynomials in $\boldsymbol{x}\,\mathbb{P}_{k,d}^H$ are members of $\mathbb{P}_{k+1,d}^H$, whereas the degree of any polynomial in $\mathbb{P}_{k,d}$ does not exceed k.

Example 14.5 ($k=1$, $d=2$). $\dim(\mathbf{RT}_{1,2}) = 8$ and $\{ \binom{1}{0}, \binom{x_1}{0}, \binom{x_2}{0}, \binom{0}{1}, \binom{0}{x_1}, \binom{0}{x_2}, \binom{x_1^2}{x_1 x_2}, \binom{x_1 x_2}{x_2^2} \}$ is a basis of $\mathbf{RT}_{1,2}$. $\qquad\square$

Lemma 14.6 (Dimension of $\mathbf{RT}_{k,d}$). $\dim(\mathbf{RT}_{k,d}) = (k+d+1)\binom{k+d-1}{k}$, *in particular* $\dim(\mathbf{RT}_{k,2}) = (k+1)(k+3)$ *and* $\dim(\mathbf{RT}_{k,3}) = \frac{1}{2}(k+1)(k+2)(k+4)$.

Proof. Since $\dim(\mathbb{P}_{k,d}) = \binom{k+d}{k}$, $\dim(\mathbb{P}_{k,d}^H) = \binom{k+d-1}{k}$, and the sum in (14.4) is direct, $\dim(\mathbf{RT}_{k,d}) = d\binom{k+d}{k} + \binom{k+d-1}{k} = (k+d+1)\binom{k+d-1}{k}$. $\qquad\square$

Lemma 14.7 (Trace space). *Let H be an affine hyperplane in $\mathbb{R}^d$ with normal vector $\boldsymbol{n}_H$, and let $\boldsymbol{T}_H : \mathbb{R}^{d-1} \to H$ be an affine bijective mapping. Then $\boldsymbol{v}_{|H}{\cdot}\boldsymbol{n}_H \in \mathbb{P}_{k,d-1} \circ \boldsymbol{T}_H^{-1}$ for all $\boldsymbol{v} \in \mathbf{RT}_{k,d}$.*

Proof. Let $\boldsymbol{v} \in \mathbb{RT}_{k,d}$ with $\boldsymbol{v} = \boldsymbol{p} + \boldsymbol{x}q$, $\boldsymbol{p} \in \mathbb{P}_{k,d}$, and $q \in \mathbb{P}_{k,d}^{\mathrm{H}}$. Let $\boldsymbol{x} \in H$ and set $\boldsymbol{y} := \boldsymbol{T}_H^{-1}(\boldsymbol{x})$. Since the quantity $\boldsymbol{x} \cdot \boldsymbol{n}_H$ is constant, say $\boldsymbol{x} \cdot \boldsymbol{n}_H =: c_H$, we infer that $(\boldsymbol{v}_{|H} \cdot \boldsymbol{n}_H)(\boldsymbol{x}) = (\boldsymbol{p}_{|H} \cdot \boldsymbol{n}_H)(\boldsymbol{x}) + (\boldsymbol{x} \cdot \boldsymbol{n}_H) q(\boldsymbol{x}) = ((\boldsymbol{p} \circ \boldsymbol{T}_H) \cdot \boldsymbol{n}_H)(\boldsymbol{y}) + c_H(q \circ \boldsymbol{T}_H)(\boldsymbol{y})$. Hence, $(\boldsymbol{v}_{|H} \cdot \boldsymbol{n}_H) \circ \boldsymbol{T}_H = (\boldsymbol{p} \circ \boldsymbol{T}_H) \cdot \boldsymbol{n}_H + c_H(q \circ \boldsymbol{T}_H)$, and both terms in the sum are in $\mathbb{P}_{k,d-1}$ by virtue of Lemma 7.10. □

Remark 14.8 ($\boldsymbol{T}_H$). Consider a second affine bijective mapping $\tilde{\boldsymbol{T}}_H : \mathbb{R}^{d-1} \to H$. Since $\boldsymbol{S} := \boldsymbol{T}_H^{-1} \circ \tilde{\boldsymbol{T}}_H$ is an affine bijective mapping from $\mathbb{R}^{d-1}$ onto itself, we have $\mathbb{P}_{k,d-1} \circ \boldsymbol{S} = \mathbb{P}_{k,d-1}$. Hence, $\mathbb{P}_{k,d-1} \circ \boldsymbol{T}_H^{-1} = \mathbb{P}_{k,d-1} \circ \boldsymbol{S} \circ \tilde{\boldsymbol{T}}_H^{-1} = \mathbb{P}_{k,d-1} \circ \tilde{\boldsymbol{T}}_H^{-1}$. This proves that the assertion of Lemma 14.7 is independent of the mapping $\boldsymbol{T}_H$. □

Lemma 14.9 (Divergence). *$\nabla \cdot \boldsymbol{v} \in \mathbb{P}_{k,d}$ for all $\boldsymbol{v} \in \mathbb{RT}_{k,d}$, and if the function $\boldsymbol{v}$ is divergence-free, then $\boldsymbol{v} \in \mathbb{P}_{k,d}$.*

Proof. That $\nabla \cdot \boldsymbol{v} \in \mathbb{P}_{k,d}$ follows from $v_i \in \mathbb{P}_{k+1,d}$ for all $i \in \{1{:}d\}$. Let $\boldsymbol{v} \in \mathbb{RT}_{k,d}$ be divergence-free. Since $\boldsymbol{v} \in \mathbb{RT}_{k,d}$, there are $\boldsymbol{p} \in \mathbb{P}_{k,d}$ and $q \in \mathbb{P}_{k,d}^{\mathrm{H}}$ such that $\boldsymbol{v} = \boldsymbol{p} + \boldsymbol{x}q$. Owing to Lemma 14.3, we infer that $\nabla \cdot \boldsymbol{p} + (k+d)q = 0$, which implies that $q = 0$ since $\mathbb{P}_{k,d}^{\mathrm{H}} \cap \mathbb{P}_{k-1,d} = \{0\}$ if $k \geq 1$. The argument for $k = 0$ is trivial. Hence, $\boldsymbol{v} = \boldsymbol{p} \in \mathbb{P}_{k,d}$. □

14.3 Simplicial Raviart–Thomas elements

Let $k \in \mathbb{N}$ and let $d \geq 2$. Let K be a simplex in $\mathbb{R}^d$. Each face $F \in \mathcal{F}_K$ of K is oriented by the normal vector $\boldsymbol{\nu}_F := |F| \boldsymbol{n}_F$ (so that $\|\boldsymbol{\nu}_F\|_{\ell^2} = |F|$). The simplex K itself is oriented by the d vectors $\{\boldsymbol{\nu}_{K,j} := |F_j| \boldsymbol{n}_{F_j}\}_{j \in \{1:d\}}$ where $\{F_j\}_{j \in \{1:d\}}$ are the d faces of K sharing the vertex with the lowest index. Note that $\{\boldsymbol{\nu}_{K,j}\}_{j \in \{1:d\}}$ is a basis of $\mathbb{R}^d$ (see Exercise 7.3(iv)), and this basis coincides with the canonical Cartesian basis of $\mathbb{R}^d$ when K is the unit simplex. The dofs of the $\mathbb{RT}_{k,d}$ finite element involve integrals over the faces of K or over K itself (for $k \geq 1$). Since the face dofs require to evaluate moments against $(d-1)$-variate polynomials, we introduce an affine bijective mapping $\boldsymbol{T}_F : \widehat{S}^{d-1} \to F$ for all $F \in \mathcal{F}_K$, where $\widehat{S}^{d-1}$ is the unit simplex of $\mathbb{R}^{d-1}$; see Figure 14.2. For instance, after enumerating the d vertices of $\widehat{S}^{d-1}$ and the $(d+1)$ vertices of K, we can define $\boldsymbol{T}_F$ such that the d vertices of $\widehat{S}^{d-1}$ are mapped to the d vertices of F with increasing indices.

Definition 14.10 (dofs). *We denote by Σ the collection of the following linear forms acting on $\mathbb{RT}_{k,d}$:*

$$\sigma_{F,m}^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{\nu}_F)(\zeta_m \circ \boldsymbol{T}_F^{-1}) \, \mathrm{d}s, \qquad \forall F \in \mathcal{F}_K, \qquad (14.5\mathrm{a})$$

$$\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) := \frac{1}{|K|} \int_K (\boldsymbol{v} \cdot \boldsymbol{\nu}_{K,j}) \psi_m \, \mathrm{d}x, \qquad \forall j \in \{1{:}d\}, \qquad (14.5\mathrm{b})$$
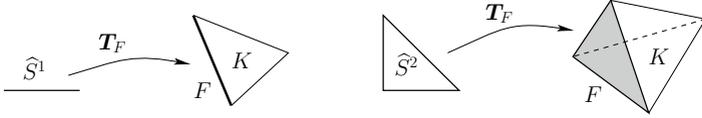
Fig. 14.2 Reference face $\widehat{S}^{d-1}$ and mapping $\boldsymbol{T}_F$ for $d = 2$ (left, the face F is indicated in bold) and $d = 3$ (right, the face F is highlighted in gray).

where $\{\zeta_m\}_{m\in\{1:n_{\text{sh}}^{\text{f}}\}}$ is a basis of $\mathbb{P}_{k,d-1}$ with $n_{\text{sh}}^{\text{f}} := \dim(\mathbb{P}_{k,d-1}) = \binom{d+k-1}{k}$ and $\{\psi_m\}_{m\in\{1:n_{\text{sh}}^{\text{c}}\}}$ is a basis of $\mathbb{P}_{k-1,d}$ with $n_{\text{sh}}^{\text{c}} := \dim(\mathbb{P}_{k-1,d}) = \binom{d+k-1}{k-1}$ if $k \geq 1$. We regroup the dofs as follows:

$$\Sigma_F^{\text{f}} := \{\sigma_{F,m}^{\text{f}}\}_{m\in\{1:n_{\text{sh}}^{\text{f}}\}}, \qquad\qquad \forall F \in \mathcal{F}_K, \qquad (14.6a)$$

$$\Sigma^{\text{c}} := \{\sigma_{j,m}^{\text{c}}\}_{(j,m)\in\{1:d\}\times\{1:n_{\text{sh}}^{\text{c}}\}}. \qquad\qquad\qquad (14.6b)$$

Remark 14.11 (dofs). The unit of all the dofs is a surface times the dimension of $\boldsymbol{v}$. We could also have written $\sigma_{j,m}^{\text{c}}(\boldsymbol{v}) := \ell_K^{-1} \int_K (\boldsymbol{v}\cdot\boldsymbol{e}_j)\psi_m \, \mathrm{d}x$ for the cell dofs, where ℓ_K is a length scale of K and $\{\boldsymbol{e}_j\}_{j\in\{1:d\}}$ is the canonical Cartesian basis of $\mathbb{R}^d$. We will see that the definition (14.5b) is more natural when using the contravariant Piola transformation to generate other finite elements. The dofs are defined here on $\mathbb{RT}_{k,d}$. Their extension to some larger space $\boldsymbol{V}(K)$ is addressed in Chapters 16 and 17. $\qquad\square$

Lemma 14.12 (Invariance w.r.t. $\boldsymbol{T}_F$). *Assume that every affine bijective mapping $\boldsymbol{S} : \widehat{S}^{d-1} \to \widehat{S}^{d-1}$ leaves the basis $\{\zeta_m\}_{m\in\{1:n_{\text{sh}}^{\text{f}}\}}$ globally invariant, i.e., $\{\zeta_m\}_{m\in\{1:n_{\text{sh}}^{\text{f}}\}} = \{\zeta_m \circ \boldsymbol{S}\}_{m\in\{1:n_{\text{sh}}^{\text{f}}\}}$. Then for all $F \in \mathcal{F}_K$, the set Σ_F^{f} is independent of the affine bijective mapping $\boldsymbol{T}_F$.*

Proof. Let $\boldsymbol{T}_F, \tilde{\boldsymbol{T}}_F$ be two affine bijective mappings from $\widehat{S}^{d-1}$ to F. Then $\boldsymbol{S} := \boldsymbol{T}_F^{-1} \circ \tilde{\boldsymbol{T}}_F$ is an affine bijective mapping from $\widehat{S}^{d-1}$ to $\widehat{S}^{d-1}$. Let $m \in \{1:n_{\text{sh}}^{\text{f}}\}$. The invariance assumption implies that there exists ζ_n, $n \in \{1:n_{\text{sh}}^{\text{f}}\}$, s.t. $\zeta_m \circ \boldsymbol{S} = \zeta_n$. Hence, with obvious notation we have

$$|F|\sigma_{F,m}^{\text{f}}(\boldsymbol{v}) = \int_F (\boldsymbol{v}\cdot\boldsymbol{\nu}_F)(\zeta_m \circ \boldsymbol{T}_F^{-1}) \, \mathrm{d}s$$

$$= \int_F (\boldsymbol{v}\cdot\boldsymbol{\nu}_F)((\zeta_m\circ\boldsymbol{S})\circ\tilde{\boldsymbol{T}}_F^{-1}) \, \mathrm{d}s = \int_F (\boldsymbol{v}\cdot\boldsymbol{\nu}_F)(\zeta_n\circ\tilde{\boldsymbol{T}}_F^{-1}) \, \mathrm{d}s = |F|\tilde{\sigma}_{F,n}^{\text{f}}(\boldsymbol{v}). \qquad \square$$

Example 14.13 (Vertex permutation). For every affine bijective mapping $\boldsymbol{S} : \widehat{S}^{d-1} \to \widehat{S}^{d-1}$, there is a unique permutation σ of the set $\{0:d-1\}$ s.t. $\boldsymbol{S}(\widehat{\boldsymbol{z}}_i) := \widehat{\boldsymbol{z}}_{\sigma(i)}$ for all $i \in \{0:d-1\}$, where $\{\widehat{\boldsymbol{z}}_i\}_{i\in\{0:d-1\}}$ are the vertices of $\widehat{S}^{d-1}$. Then the above invariance holds true holds true iff all the vertices of $\widehat{S}^{d-1}$ play symmetric roles when defining the basis functions $\{\zeta_m\}_{m\in\{1:n_{\text{sh}}^{\text{f}}\}}$. For instance, for $d := 2$, $\widehat{S}^1 := [0,1]$, and $k := 1$, the basis $\{1, s\}$ of $\mathbb{P}_{1,1}$ is not invariant w.r.t. vertex permutation, but the basis $\{1 - s, s\}$ is. $\qquad\square$

A graphic representation of the dofs is shown in Figure 14.3. The number of arrows on a face counts the number of moments of the normal component considered over the face. The number of pairs of gray circles inside the triangle counts the number of moments inside the cell (one circle for the component along $\boldsymbol{\nu}_{K,1}$ and one for the component along $\boldsymbol{\nu}_{K,2}$).
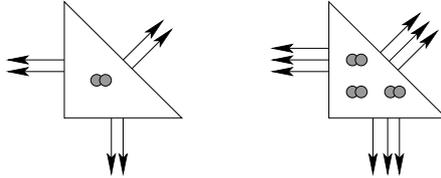
Fig. 14.3 Degrees of freedom of $\mathbb{R}\mathbb{T}_{k,d}$ finite elements for $d = 2$ and $k = 1$ (left) or $k = 2$ (right) (assuming that all the normals point outward).

Lemma 14.14 (Face unisolvence). *For all $\boldsymbol{v} \in \mathbb{R}\mathbb{T}_{k,d}$ and all $F \in \mathcal{F}_K$,*

$$[\, \sigma(\boldsymbol{v}) = 0, \ \forall \sigma \in \Sigma_F^{\mathrm{f}} \,] \iff [\, \boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F = 0 \,]. \tag{14.7}$$

Proof. The condition $\sigma(\boldsymbol{v}) = 0$ for all $\sigma \in \Sigma_F^{\mathrm{f}}$ means that $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F$ is orthogonal to $\mathbb{P}_{k,d-1} \circ \boldsymbol{T}_F^{-1}$. Since Lemma 14.7 implies that $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F \in \mathbb{P}_{k,d-1} \circ \boldsymbol{T}_F^{-1}$, we infer that $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F = 0$. $\qquad\square$

Proposition 14.15 (Finite element). $(K, \mathbb{R}\mathbb{T}_{k,d}, \Sigma)$ *is a finite element.*

Proof. We have already established the assertion for $k = 0$. Let us consider $k \geq 1$. Observe first that the cardinality of Σ can be evaluated as follows:

$$\mathrm{card}(\Sigma) = d\, n_{\mathrm{sh}}^{\mathrm{c}} + (d+1) n_{\mathrm{sh}}^{\mathrm{f}} = d \binom{d+k-1}{k-1} + (d+1) \binom{d+k-1}{k}$$
$$= \frac{(d+k-1)!}{(d-1)!(k-1)!} \left(1 + \frac{d+1}{k} \right) = \dim(\mathbb{R}\mathbb{T}_{k,d}).$$

Hence, the statement will be proved once it is established that zero is the only function in $\mathbb{R}\mathbb{T}_{k,d}$ that annihilates the dofs in Σ. Let $\boldsymbol{v} \in \mathbb{R}\mathbb{T}_{k,d}$ be such that $\sigma(\boldsymbol{v}) = 0$ for all $\sigma \in \Sigma$. Owing to Lemma 14.14, we infer that $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F = 0$ for all $F \in \mathcal{F}_K$. This in turn implies that $\int_K \boldsymbol{v} \cdot (\nabla \nabla \cdot \boldsymbol{v}) \, \mathrm{d}x = -\int_K (\nabla \cdot \boldsymbol{v})^2 \, \mathrm{d}x$. Observing that $\nabla \nabla \cdot \boldsymbol{v}$ is in $\mathbb{P}_{k-1,d}$ (recall that $\nabla \cdot \boldsymbol{v} \in \mathbb{P}_{k,d}$ from Lemma 14.9), the assumption that $\sigma(\boldsymbol{v}) = 0$ for all $\sigma \in \Sigma^{\mathrm{c}}$ (i.e., $\boldsymbol{v}$ is orthogonal to $\mathbb{P}_{k-1,d}$), together with the above identity imply that $\nabla \cdot \boldsymbol{v} = 0$. Using Lemma 14.9, we conclude that $\boldsymbol{v} \in \mathbb{P}_{k,d}$ and $\boldsymbol{v}_{|F} \cdot \boldsymbol{\nu}_F = 0$ for all $F \in \mathcal{F}_K$. Let $j \in \{1{:}d\}$. Since $\boldsymbol{\nu}_{K,j} = \boldsymbol{\nu}_{F_j} = |F_j| \boldsymbol{n}_{F_j}$ for some face $F_j \in \mathcal{F}_K$, we infer that $\boldsymbol{v}(\boldsymbol{x}) \cdot \boldsymbol{\nu}_{K,j} = \lambda_j(\boldsymbol{x}) r_j(\boldsymbol{x})$ for all $\boldsymbol{x} \in K$, where λ_j is the barycentric coordinate of K associated with the vertex opposite to F_j (i.e., λ_j vanishes

on F_j) and $r_j \in \mathbb{P}_{k-1,d}$; see Exercise 7.4(iv). The condition $\sigma(\boldsymbol{v}) = 0$ for all $\sigma \in \Sigma^c$ implies that $\int_K (\boldsymbol{v} \cdot \boldsymbol{\nu}_{K,j}) r_j \, dx = 0$, which in turn means that $0 = \int_K (\boldsymbol{v} \cdot \boldsymbol{\nu}_{K,j}) r_j \, dx = \int_K \lambda_j r_j^2 \, dx$, thereby proving that $r_j = 0$ since λ_j is positive in the interior of K. Hence, $\boldsymbol{v} \cdot \boldsymbol{\nu}_{K,j}$ vanishes identically for all $j \in \{1{:}d\}$. This proves that $\boldsymbol{v} = \boldsymbol{0}$ since $\{\boldsymbol{\nu}_{K,j}\}_{j \in \{1:d\}}$ is a basis of $\mathbb{R}^d$. $\qquad\square$

The shape functions $\{\theta_i\}_{i \in \mathcal{N}}$ associated with the dofs $\{\sigma_i\}_{i \in \mathcal{N}}$ defined in (14.5) can be constructed by choosing a basis $\{\phi_i\}_{i \in \mathcal{N}}$ of the polynomial space $\mathbb{RT}_{k,d}$ and by inverting the corresponding generalized Vandermonde matrix $\mathcal{V}$ as explained in Proposition 5.5. Recall that this matrix has entries $\mathcal{V}_{ij} = \sigma_j(\phi_i)$ and that the i-th line of $\mathcal{V}^{-1}$ gives the components of the shape function θ_i in the basis $\{\phi_i\}_{i \in \mathcal{N}}$. The basis $\{\phi_i\}_{i \in \mathcal{N}}$ chosen in Bonazzoli and Rapetti [31] (built by dividing the simplex into smaller sub-simplices following the ideas in Rapetti and Bossavit [163], Christiansen and Rapetti [70]) is particularly interesting since the entries of $\mathcal{V}^{-1}$ are integers. One could also choose $\{\phi_i\}_{i \in \mathcal{N}}$ to be the hierarchical basis of $\mathbb{RT}_{k,d}$ constructed in Fuentes et al. [103, §7.3]. This basis can be organized into functions attached to the faces of K and to K itself in such a way that the generalized Vandermonde matrix $\mathcal{V}$ is block-triangular (notice though that this matrix is not block-diagonal).

Remark 14.16 (Dof independence). As in Remark 7.20, we infer from Exercise 5.2 that the interpolation operator $\mathcal{I}_K^d$ associated with the $\mathbb{RT}_{k,d}$ element is independent of the bases $\{\zeta_m\}_{m \in \{1:n_{sh}^f\}}$ and $\{\psi_m\}_{m \in \{1:n_{sh}^c\}}$ used to define the dofs in (14.5). This operator is also independent of the mappings $\boldsymbol{T}_F$ and of the orientation vectors $\{\boldsymbol{\nu}_F\}_{F \in \mathcal{F}_K}$ and $\{\boldsymbol{\nu}_{K,j}\}_{j \in \{1:d\}}$. $\qquad\square$

Remark 14.17 (Literature). The $\mathbb{RT}_{k,d}$ finite element has been introduced in Raviart and Thomas [164, 165] for $d = 2$; see also Weil [198, p. 127], Whitney [199, Eq. (12), p. 139] for $k = 0$. The generalization to $d \geq 3$ is due to Nédélec [151]. The reading of [151] is highly recommended; see also Boffi et al. [29, §2.3.1], Hiptmair [117], Monk [145, pp. 118–126]. The name Raviart–Thomas seems to be an accepted practice in the literature. $\qquad\square$

14.4 Generation of Raviart–Thomas elements

Let $\widehat{K}$ be the reference simplex in $\mathbb{R}^d$. Let $\mathcal{T}_h$ be an affine simplicial mesh. Let $K = \boldsymbol{T}_K(\widehat{K})$ be a mesh cell, where $\boldsymbol{T}_K : \widehat{K} \to K$ is the geometric mapping, and let $\mathbb{J}_K$ be the Jacobian matrix of $\boldsymbol{T}_K$. Let $F \in \mathcal{F}_K$ be a face of K. We have $F = \boldsymbol{T}_K(\widehat{F})$ for some face $\widehat{F} \in \mathcal{F}_{\widehat{K}}$. Owing to Theorem 10.8, it is possible (using the increasing vertex-index enumeration) to orient the faces F and $\widehat{F}$ in a way that is compatible with the geometric mapping $\boldsymbol{T}_K$. This means that the unit normal vectors $\boldsymbol{n}_F$ and $\widehat{\boldsymbol{n}}_{\widehat{F}}$ satisfy (10.6b), i.e., $\boldsymbol{n}_F = \boldsymbol{\Phi}_K^d(\widehat{\boldsymbol{n}}_{\widehat{F}})$ with $\boldsymbol{\Phi}_K^d$ defined in (9.14a). In other words, we have

$$\boldsymbol{n}_F \circ \boldsymbol{T}_{K|\widehat{F}} = \epsilon_K \frac{1}{\|\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}}\|_{\ell^2}} \mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}}, \qquad\qquad (14.8)$$

where $\epsilon_K := \frac{\det(\mathbb{J}_K)}{|\det(\mathbb{J}_K)|} = \pm 1$. Recalling that $\boldsymbol{\nu}_F := |F|\boldsymbol{n}_F$, $\widehat{\boldsymbol{\nu}}_{\widehat{F}} := |\widehat{F}|\widehat{\boldsymbol{n}}_{\widehat{F}}$ and that $|F| = |\det(\mathbb{J}_K)| \|\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{n}}_{\widehat{F}}\|_{\ell^2}|\widehat{F}|$ owing to Lemma 9.12, we infer that

$$\boldsymbol{\nu}_F \circ \boldsymbol{T}_{K|\widehat{F}} = \det(\mathbb{J}_K)\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{\nu}}_{\widehat{F}}. \qquad\qquad (14.9)$$

Due to the role played by the normal component of vector fields on the faces of K, we are going to use in Proposition 9.2 the contravariant Piola transformation

$$\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}(\boldsymbol{v} \circ \boldsymbol{T}_K) \qquad\qquad (14.10)$$

to define Raviart–Thomas elements on K from a reference Raviart–Thomas element defined on $\widehat{K}$. For scalar fields, we consider the pullback by the geometric mapping, i.e., $\psi_K^{\mathrm{g}}(q) := q \circ \boldsymbol{T}_K$. Finally, we orient K and $\widehat{K}$ with the d vectors $\{\boldsymbol{\nu}_{K,j} := |F_j|\boldsymbol{n}_{F_j}\}_{j\in\{1:d\}}$ and $\{\widehat{\boldsymbol{\nu}}_{\widehat{K},j} := |\widehat{F}_j|\widehat{\boldsymbol{n}}_{F_j}\}_{j\in\{1:d\}}$ associated with the d faces of K and $\widehat{K}$ that share the vertex with the lowest index, i.e., we have $F_j = \boldsymbol{T}_K(\widehat{F}_j)$ for all $j \in \{1:d\}$. The above considerations show that $\boldsymbol{\nu}_{K,j} \circ \boldsymbol{T}_K = \det(\mathbb{J}_K)\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{\nu}}_{\widehat{K},j}$ for all $j \in \{1:d\}$.

Lemma 14.18 (Transformation of dofs). *Let $\boldsymbol{v} \in \boldsymbol{C}^0(K)$ and let $q \in C^0(K)$. The following holds true:*

$$\frac{1}{|F|}\int_F (\boldsymbol{v}\cdot\boldsymbol{\nu}_F)q\,\mathrm{d}s = \frac{1}{|\widehat{F}|}\int_{\widehat{F}} (\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{\nu}}_{\widehat{F}})\psi_K^{\mathrm{g}}(q)\,\mathrm{d}\widehat{s}, \qquad \forall F \in \mathcal{F}_K, \quad (14.11\mathrm{a})$$

$$\frac{1}{|K|}\int_K (\boldsymbol{v}\cdot\boldsymbol{\nu}_{K,j})q\,\mathrm{d}x = \frac{1}{|\widehat{K}|}\int_{\widehat{K}} (\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{\nu}}_{\widehat{K},j})\psi_K^{\mathrm{g}}(q)\,\mathrm{d}\widehat{x}, \quad \forall j \in \{1:d\}. \quad (14.11\mathrm{b})$$

Proof. The identity (14.11a) is nothing but (10.7a) from Lemma 10.4, which itself is a reformulation of (9.15a) from Lemma 9.13 (the fact that $\boldsymbol{T}_K$ is affine is not used here). The proof of (14.11b) is similar since

$$\int_K (\boldsymbol{v}\cdot\boldsymbol{\nu}_{K,j})q\,\mathrm{d}x = \int_{\widehat{K}} (\boldsymbol{v} \circ \boldsymbol{T}_K)\cdot(\boldsymbol{\nu}_{K,j} \circ \boldsymbol{T}_K)\psi_K^{\mathrm{g}}(q)|\det(\mathbb{J}_K)|\,\mathrm{d}\widehat{x}$$

$$= \int_{\widehat{K}} (\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{\nu}}_{\widehat{K},j})\psi_K^{\mathrm{g}}(q)|\det(\mathbb{J}_K)|\,\mathrm{d}\widehat{x},$$

and since $\boldsymbol{T}_K$ is affine, we have $|K| = |\det(\mathbb{J}_K)|\,|\widehat{K}|$. $\qquad\square$

Proposition 14.19 (Generation). *Let $(\widehat{K}, \widehat{\boldsymbol{P}}, \widehat{\Sigma})$ be a simplicial $\mathbb{RT}_{k,d}$ element with face and cell dofs defined using the polynomial bases $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ and $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ (if $k \geq 1$) of $\mathbb{P}_{k,d-1}$ and $\mathbb{P}_{k-1,d}$, respectively, as in (14.5). Assume that the geometric mapping $\boldsymbol{T}_K$ is affine and that (14.9) holds true. Then the finite element $(K, \boldsymbol{P}_K, \Sigma_K)$ generated using Proposition 9.2 with*

the contravariant Piola transformation (14.10) *is a simplicial* $\mathbb{RT}_{k,d}$ *finite element with dofs*

$$\sigma^{\mathrm{f}}_{F,m}(\boldsymbol{v}) = \frac{1}{|F|} \int_F (\boldsymbol{v}{\cdot}\boldsymbol{\nu}_F)(\zeta_m \circ \boldsymbol{T}_{K,F}^{-1})\,\mathrm{d}s, \qquad \forall F \in \mathcal{F}_K, \qquad (14.12\mathrm{a})$$

$$\sigma^{\mathrm{c}}_{j,m}(\boldsymbol{v}) = \frac{1}{|K|} \int_K (\boldsymbol{v}{\cdot}\boldsymbol{\nu}_{K,j})(\psi_m \circ \boldsymbol{T}_K^{-1})\,\mathrm{d}x, \qquad \forall j \in \{1{:}d\}, \qquad (14.12\mathrm{b})$$

where $\boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}}$ *is the affine bijective mapping from* $\widehat{S}^{d-1}$ *to* F *that maps the* d *vertices of* $\widehat{S}^{d-1}$ *to the* d *vertices of* F *with increasing indices.*

Proof. See Exercise 14.4 for the proof that $\boldsymbol{P}_K = \mathbb{RT}_{k,d}$. Use Lemma 14.18 to prove (14.12a)–(14.12b). □

Remark 14.20 (Unit). Given some length unit L, the shape functions scale as L^{1-d} since the unit of all the dofs is L^{d-1}. □

Remark 14.21 (Nonaffine meshes). Proposition 9.2 together with the map $\boldsymbol{\psi}_K^{\mathrm{d}}$ defined in (14.10) can still be used to generate a finite element $(K, \boldsymbol{P}_K, \Sigma_K)$ if the geometric mapping $\boldsymbol{T}_K$ is nonaffine. The function space $\boldsymbol{P}_K$ and the dofs in Σ_K then differ from those of the $\mathbb{RT}_{k,d}$ element. □

14.5 Other $H(\mathrm{div})$ finite elements

14.5.1 Brezzi–Douglas–Marini elements

Brezzi–Douglas–Marini (BDM) elements [49, 50] offer an interesting alternative to Raviart–Thomas elements since in this case the polynomial space is $\boldsymbol{P} := \mathbb{P}_{k,d} \subsetneq \mathbb{RT}_{k,d}$, $k \geq 1$. This space is optimal from the approximation viewpoint. The price to pay for this simplification is that the divergence operator $\nabla{\cdot}$ is surjective from $\mathbb{P}_{k,d}$ onto $\mathbb{P}_{k-1,d}$ only. This is not a limitation if the functions one wants to interpolate are divergence-free (or have a divergence that belongs to $\mathbb{P}_{k-1,d}$).

Let K be a simplex in $\mathbb{R}^d$. The dofs of BDM elements are attached to the $(d+1)$ faces of K and to K itself (for $k \geq 2$). The face dofs are the same as for Raviart–Thomas elements, i.e., the linear forms $\sigma^{\mathrm{f}}_{F,m}$ defined in (14.5a) for all $F \in \mathcal{F}_K$ and every $m \in \{1{:}n^{\mathrm{f}}_{\mathrm{sh}}\}$ with $n^{\mathrm{f}}_{\mathrm{sh}} := \dim(\mathbb{P}_{k,d-1})$. Note that the cell dofs for Raviart–Thomas elements are moments against a set of basis functions of $\mathbb{P}_{k-1,d}$, whereas those for BDM elements are moments against a set of basis functions of the Nédélec polynomial space $\mathbb{N}_{k-2,d}$ introduced in the next chapter (see §15.2). At this stage, it is sufficient to know that $\mathbb{P}_{k-2,d} \subsetneq \mathbb{N}_{k-2,d} \subsetneq \mathbb{P}_{k-1,d}$ and that $\dim(\mathbb{N}_{k-2,2}) = (k-1)(k+1)$ and $\dim(\mathbb{N}_{k-2,3}) = \frac{1}{2}(k-1)(k+1)(k+2)$ (see Lemma 15.7). We define

$$\tilde{\sigma}_m^{\mathrm{c}}(\boldsymbol{v}) := \int_K \boldsymbol{v} \cdot \tilde{\boldsymbol{\psi}}_m \, \mathrm{d}x, \qquad \forall m \in \{1 : \tilde{n}_{\mathrm{sh}}^{\mathrm{c}}\}, \tag{14.13}$$

where $\{\tilde{\boldsymbol{\psi}}_m\}_{m \in \{1 : \tilde{n}_{\mathrm{sh}}^{\mathrm{c}}\}}$ is a basis of $\mathbb{N}_{k-2,d}$ and $\tilde{n}_{\mathrm{sh}}^{\mathrm{c}} := \dim(\mathbb{N}_{k-2,d})$. Let us set $\Sigma := \{\sigma_{F,m}^{\mathrm{f}}\}_{F \in \mathcal{F}_K, m \in \{1 : n_{\mathrm{sh}}^{\mathrm{f}}\}} \cup \{\tilde{\sigma}_m^{\mathrm{c}}\}_{m \in \{1 : \tilde{n}_{\mathrm{sh}}^{\mathrm{c}}\}}$.

Proposition 14.22 (Finite element). $(K, \mathbb{P}_{k,d}, \Sigma)$ *is a finite element.*

Proof. See Boffi et al. [29, p. 88]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Hierarchical basis functions for the BDM element are constructed in Ainsworth and Coyle [6], Schöberl and Zaglmayr [176].

Remark 14.23 (Generation). Generating BDM elements also involves the covariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{w}) := \mathbb{J}_K^{\mathsf{T}}(\boldsymbol{w} \circ \boldsymbol{T}_K)$ defined in (9.9b), so that $\int_K \boldsymbol{v} \cdot \tilde{\boldsymbol{\psi}}_m \, \mathrm{d}x = \epsilon_K \int_{\widehat{K}} \boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) \cdot \boldsymbol{\psi}_K^{\mathrm{c}}(\tilde{\boldsymbol{\psi}}_m) \, \mathrm{d}\widehat{x}$ with $\epsilon_K := \frac{\det(\mathbb{J}_K)}{|\det(\mathbb{J}_K)|} = \pm 1$. $\quad\square$

14.5.2 Cartesian Raviart–Thomas elements

Let us briefly review the Cartesian Raviart–Thomas finite elements. We refer the reader to Exercise 14.6 for the proofs. For a multi-index $\alpha \in \mathbb{N}^d$, we define the (anisotropic) polynomial space $\mathbb{Q}_{\alpha_1, \ldots, \alpha_d}$ composed of d-variate polynomials whose degree with respect to x_i is at most α_i for all $i \in \{1 : d\}$. Let $k \in \mathbb{N}$ and define

$$\mathbb{R}\mathbb{T}_{k,d}^{\square} := \mathbb{Q}_{k+1,k,\ldots,k} \times \ldots \times \mathbb{Q}_{k,\ldots,k,k+1}. \tag{14.14}$$

Since $\dim(\mathbb{Q}_{k+1,k,\ldots,k}) = \ldots = \dim(\mathbb{Q}_{k,\ldots,k,k+1}) = (k+2)(k+1)^{d-1}$, we have $\dim(\mathbb{R}\mathbb{T}_{k,d}^{\square}) = d(k+2)(k+1)^{d-1}$. Moreover, one can verify that

$$\nabla \cdot \boldsymbol{v} \in \mathbb{Q}_{k,d}, \qquad \boldsymbol{v}_{|H} \cdot \boldsymbol{\nu}_H \in \mathbb{Q}_{k,d-1} \circ \boldsymbol{T}_H^{-1}, \tag{14.15}$$

for all $\boldsymbol{v} \in \mathbb{R}\mathbb{T}_{k,d}^{\square}$ and every affine hyperplane H in $\mathbb{R}^d$ with normal vector $\boldsymbol{\nu}_H$ parallel to one of the vectors of the canonical basis of $\mathbb{R}^d$ and where $\boldsymbol{T}_H : \mathbb{R}^{d-1} \to H$ is any affine bijective mapping.

Let K be a cuboid in $\mathbb{R}^d$. Each face $F \in \mathcal{F}_K$ of K is oriented by the normal vector $\boldsymbol{\nu}_F$ with $\|\boldsymbol{\nu}_F\|_{\ell^2} = |F|$. Let $\boldsymbol{T}_F$ be an affine bijective mapping from $[0,1]^{d-1}$ onto F. Let us orient K using $\boldsymbol{\nu}_{K,j} := |F_j| \boldsymbol{e}_j$ for all $j \in \{1 : d\}$, where $\{\boldsymbol{e}_j\}_{j \in \{1 : d\}}$ is the canonical basis of $\mathbb{R}^d$ and $|F_j|$ is the measure of any of the two faces of K supported in a hyperplane perpendicular to $\boldsymbol{e}_j$. Let Σ be the set composed of the following linear forms:

$$\sigma_{F,m}^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{\nu}_F)(\zeta_m \circ \boldsymbol{T}_F^{-1}) \, \mathrm{d}s, \qquad \forall F \in \mathcal{F}_K, \tag{14.16a}$$

$$\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) := \frac{1}{|K|} \int_K (\boldsymbol{v} \cdot \boldsymbol{\nu}_{K,j}) \psi_{j,m} \, \mathrm{d}x, \qquad \forall j \in \{1 : d\}, \tag{14.16b}$$

where $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ is a basis of $\mathbb{Q}_{k,d-1}$ with $n_{\mathrm{sh}}^{\mathrm{f}} := (k+1)^{d-1}$, and $\{\psi_{j,m}\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ is a basis of $\mathbb{Q}_{k,\ldots,k,k-1,k,\ldots,k}$ with $n_{\mathrm{sh}}^{\mathrm{c}} := k(k+1)^{d-1}$ if $k \geq 1$, with the index $(k-1)$ at the j-th position for all $j \in \{1:d\}$.

Proposition 14.24 (Finite element). $(K, \mathbb{RT}_{k,d}^{\square}, \Sigma)$ *is a finite element.*

Cartesian Raviart–Thomas elements can be generated for all the mesh cells of an affine mesh composed of parallelotopes by using affine geometric mappings and the contravariant Piola transformation (recall, however, that orienting such meshes and making the orientation generation-compatible requires some care; see Theorem 10.10).

Example 14.25 (Shape functions and dofs for $\mathbb{RT}_{0,d}^{\square}$). Let $K := [0,1]^d$. Let F_i and F_{d+i} be the faces defined by $x_i = 0$ and $x_i = 1$, respectively, for all $i \in \{1:d\}$. Using the basis function $\zeta_1 := 1$ for $\mathbb{Q}_{0,d-1}$, the $2d$ dofs are the mean-value of the normal component over each face of K, and the shape functions are $\boldsymbol{\theta}_i^{\mathrm{f}}(\boldsymbol{x}) := (1-x_i)\boldsymbol{n}_{F_i}$ and $\boldsymbol{\theta}_{d+i}^{\mathrm{f}}(\boldsymbol{x}) := x_i\boldsymbol{n}_{F_i}$ for all $i \in \{1:d\}$. The dofs are illustrated in Figure 14.4. □
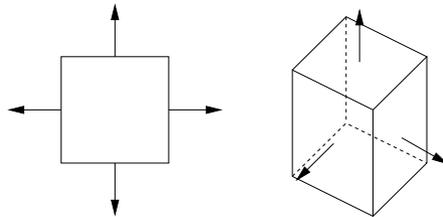
Fig. 14.4 Degrees of freedom of the lowest-order Cartesian Raviart–Thomas element $\mathbb{RT}_{0,d}^{\square}$ in dimensions two (left) and three (right, only visible dofs are shown).

Remark 14.26 (Other elements). Alternative elements are the Cartesian Brezzi–Douglas–Marini elements in dimension two, the Brezzi–Douglas–Durán–Fortin elements in dimension three (see [49, 50]), and their reduced versions by Brezzi–Douglas–Fortin–Marini [51]. □

Exercises

Exercise 14.1 ($\mathbb{RT}_{0,d}$). (i) Prove that $\int_K \iota_{F,K}\boldsymbol{\theta}_F^{\mathrm{f}}\,\mathrm{d}x = \boldsymbol{c}_F - \boldsymbol{c}_K$, where $\boldsymbol{\theta}_F^{\mathrm{f}}$ is defined in (14.3), and $\boldsymbol{c}_F$, $\boldsymbol{c}_K$ are the barycenters of F and K, respectively. (*Hint*: use (14.3) and $\int_F \boldsymbol{x}\,\mathrm{d}s = |F|\boldsymbol{c}_F$.) Provide a second proof without using (14.3). (*Hint*: fix $\boldsymbol{e} \in \mathbb{R}^d$, define $\phi(\boldsymbol{x}) = (\boldsymbol{x}-\boldsymbol{c}_F)\cdot\boldsymbol{e}$, observe that $\nabla\phi = \boldsymbol{e}$, and compute $\boldsymbol{e}\cdot\int_K \boldsymbol{\theta}_F^{\mathrm{f}}\,\mathrm{d}x$.) (ii) Prove that $\sum_{F\in\mathcal{F}_K}|F|\boldsymbol{\theta}_F^{\mathrm{f}}(\boldsymbol{x})\otimes\boldsymbol{n}_F = \mathbb{I}_d$ for

all $\boldsymbol{x} \in K$. (*Hint*: use (7.1).) (iii) Prove that $\boldsymbol{v}(\boldsymbol{x}) = \langle \boldsymbol{v} \rangle_K + \frac{1}{d}(\nabla{\cdot}\boldsymbol{v})(\boldsymbol{x} - \boldsymbol{c}_K)$ for all $\boldsymbol{v} \in \mathbb{RT}_{0,d}$, where $\langle \boldsymbol{v} \rangle_K := \frac{1}{|K|}\int_K \boldsymbol{v}\,\mathrm{d}x$ is the mean-value of $\boldsymbol{v}$ on K.

Exercise 14.2 ($\mathbb{RT}_{0,d}$ in 3D). Let $d = 3$. Let F_i, $i \in \{0{:}3\}$, be a face of K with vertices $\{\boldsymbol{a}_r, \boldsymbol{a}_p, \boldsymbol{a}_q\}$ s.t. $\big((\boldsymbol{z}_q - \boldsymbol{z}_r){\times}(\boldsymbol{z}_p - \boldsymbol{z}_r)\big){\cdot}\boldsymbol{n}_{K|F_i} > 0$. (i) Prove that $\nabla\lambda_p{\times}\nabla\lambda_q = \frac{\boldsymbol{z}_r - \boldsymbol{z}_i}{6|K|}$ and prove similar formulas for $\nabla\lambda_q{\times}\nabla\lambda_r$ and $\nabla\lambda_r{\times}\nabla\lambda_p$. (*Hint*: prove the formula in the reference simplex, then use Exercise 9.5.) (ii) Prove that $\boldsymbol{\theta}_i^{\mathrm{f}} = -2(\lambda_p\nabla\lambda_q{\times}\nabla\lambda_r + \lambda_q\nabla\lambda_r{\times}\nabla\lambda_p + \lambda_r\nabla\lambda_p{\times}\nabla\lambda_q)$. Find the counterpart of this formula if $d = 2$.

Exercise 14.3 (Piola transformation). (i) Let $\boldsymbol{v} \in \boldsymbol{C}^1(K)$ and $q \in C^0(K)$. Prove that $\int_K q\,\nabla{\cdot}\boldsymbol{v}\,\mathrm{d}x = \int_{\widehat{K}} \psi_K^{\mathrm{g}}(q)\nabla{\cdot}\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\,\mathrm{d}\widehat{x}$. (ii) Show that $\int_K \boldsymbol{v}{\cdot}\boldsymbol{\theta}\,\mathrm{d}x = \epsilon_K \int_{\widehat{K}} \boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}){\cdot}\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{\theta})\,\mathrm{d}\widehat{x}$ for all $\boldsymbol{\theta} \in \boldsymbol{C}^1(K)$.

Exercise 14.4 (Generating $\mathbb{RT}_{k,d}$). (i) Let $\boldsymbol{c} \in \mathbb{R}^d$, $q \in \mathbb{P}_{k,d}^{\mathrm{H}}$, and $\mathbb{A} \in \mathbb{R}^{d \times d'}$. Show that there is $r \in \mathbb{P}_{k-1,d'}$ such that $q(\mathbb{A}\boldsymbol{y} + \boldsymbol{c}) = q(\mathbb{A}\boldsymbol{y}) + r(\boldsymbol{y})$. (ii) Defining $s(\boldsymbol{y}) := q(\mathbb{A}\boldsymbol{y})$, show that $s \in \mathbb{P}_{k,d'}^{\mathrm{H}}$. (iii) Prove that $(\boldsymbol{\psi}_K^{\mathrm{d}})^{-1}(\mathbb{RT}_{k,d}) \subset \mathbb{RT}_{k,d}$. (iv) Prove the converse inclusion.

Exercise 14.5 (BDM). Verify that $\mathrm{card}(\Sigma) = \dim(\mathbb{P}_{k,d})$ for $d \in \{2, 3\}$.

Exercise 14.6 (Cartesian Raviart–Thomas element). (i) Propose a basis for $\mathbb{RT}_{0,2}^{\square}$ and for $\mathbb{RT}_{0,3}^{\square}$ in $K := [0,1]^d$. (ii) Prove (14.15). (iii) Prove Proposition 14.24.

Part III, Chapter 15

$H(\mathrm{curl})$ finite elements

The goal of this chapter is to construct $\mathbb{R}^d$-valued finite elements $(K, \boldsymbol{P}, \Sigma)$ with $d \in \{2,3\}$ such that (i) $\mathbb{P}_{k,d} \subset \boldsymbol{P}$ for some $k \geq 0$ and (ii) the degrees of freedom (dofs) in Σ fully determine the tangential components of the polynomials in $\boldsymbol{P}$ on all the faces of K. The first requirement is key for proving convergence rates on the interpolation error. The second one is key for constructing $\boldsymbol{H}(\mathrm{curl})$-conforming finite element spaces (see Chapter 19). The finite elements introduced in this chapter are used, e.g., in Chapter 43 to approximate (simplified forms of) Maxwell's equations which constitute a fundamental model in electromagnetism. The focus here is on defining a reference element and generating finite elements on the mesh cells. The interpolation error analysis is done in Chapters 16 and 17. We detail the construction for the simplicial Nédélec finite elements of the first kind. Some alternative elements are outlined at the end of the chapter.

15.1 The lowest-order case

Let us consider the lowest-order *Nédélec finite element*. Let $d \in \{2,3\}$ be the space dimension, and define the polynomial space

$$\mathbb{N}_{0,d} := \mathbb{P}_{0,d} \oplus \mathbb{S}_{1,d}, \tag{15.1}$$

where $\mathbb{S}_{1,d} := \{\boldsymbol{q} \in \mathbb{P}_{1,d}^{\mathrm{H}} \mid \boldsymbol{q}(\boldsymbol{x}){\cdot}\boldsymbol{x} = 0\}$, i.e.,

$$\mathbb{S}_{1,2} := \operatorname{span}\left\{ \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \right\}, \qquad \mathbb{S}_{1,3} := \operatorname{span}\left\{ \begin{pmatrix} 0 \\ -x_3 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_3 \\ 0 \\ -x_1 \end{pmatrix}, \begin{pmatrix} -x_2 \\ x_1 \\ 0 \end{pmatrix} \right\}. \tag{15.2}$$

The sum in (15.1) is indeed direct, so that $\dim(\mathbb{N}_{0,d}) = \frac{d(d+1)}{2} =: d'$ (i.e., $d' = 3$ if $d = 2$ and $d' = 6$ if $d = 3$). Note that d' is the number of edges of a simplex in $\mathbb{R}^d$. The space $\mathbb{N}_{0,d}$ has several interesting properties. (a) One has $\mathbb{P}_{0,d} \subset \mathbb{N}_{0,d}$ in agreement with the first requirement stated above. (b)

The gradient of $\boldsymbol{v} \in \mathbb{N}_{0,d}$ is skew-symmetric. Indeed, only the component $\boldsymbol{q} \in \mathbb{S}_{1,d}$ contributes to the gradient, and the identity $\partial_{x_i x_j}(\boldsymbol{q}(\boldsymbol{x}) \cdot \boldsymbol{x}) = 0$, $i \neq j$, yields $\partial_i q_j + \partial_j q_i = 0$. (c) If $\boldsymbol{v} \in \mathbb{N}_{0,d}$ is curl-free, then $\boldsymbol{v}$ is constant. Indeed, $\boldsymbol{v}$ being curl-free means that $\nabla \boldsymbol{v}$ is symmetric, which implies $\nabla \boldsymbol{v} = \boldsymbol{0}$ owing to (b). (d) The tangential component of $\boldsymbol{v} \in \mathbb{N}_{0,d}$ along an affine line in $\mathbb{R}^d$ is constant along that line. Let indeed $\boldsymbol{x}, \boldsymbol{y}$ be two distinct points on the line, say L, with tangent vector $\boldsymbol{t}_L$. Then there is $\lambda \in \mathbb{R}$ such that $\boldsymbol{t}_L = \lambda(\boldsymbol{x} - \boldsymbol{y})$ and since $\boldsymbol{v} = \boldsymbol{r} + \boldsymbol{q}$ with $\boldsymbol{r} \in \mathbb{P}_{0,d}$ and $\boldsymbol{q} \in \mathbb{S}_{1,d}$, we infer that $\boldsymbol{v}(\boldsymbol{x}) \cdot \boldsymbol{t}_L - \boldsymbol{v}(\boldsymbol{y}) \cdot \boldsymbol{t}_L = (\boldsymbol{q}(\boldsymbol{x}) - \boldsymbol{q}(\boldsymbol{y})) \cdot \boldsymbol{t}_L = \lambda \boldsymbol{q}(\boldsymbol{x} - \boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y}) = 0$.

Let K be a simplex in $\mathbb{R}^d$ and let $\mathcal{E}_K$ collect the edges of K. Any edge $E \in \mathcal{E}_K$ is oriented by fixing an edge vector $\boldsymbol{t}_E$ s.t. $\|\boldsymbol{t}_E\|_{\ell^2} = |E|$. Conventionally, we set $\boldsymbol{t}_E := \boldsymbol{z}_q - \boldsymbol{z}_p$, where $\boldsymbol{z}_p, \boldsymbol{z}_q$ are the two endpoints of E with $p < q$. We denote by Σ the collection of the following linear forms acting on $\mathbb{N}_{0,d}$:

$$\sigma_E^{\mathrm{e}}(\boldsymbol{v}) := \frac{1}{|E|} \int_E (\boldsymbol{v} \cdot \boldsymbol{t}_E) \, \mathrm{d}l, \qquad \forall E \in \mathcal{E}_K. \tag{15.3}$$

Note that the unit of $\sigma_E^{\mathrm{e}}(\boldsymbol{v})$ is a length times the dimension of $\boldsymbol{v}$. A graphic representation of the dofs is shown in Figure 15.1. Each arrow indicates the orientation of the corresponding edge.
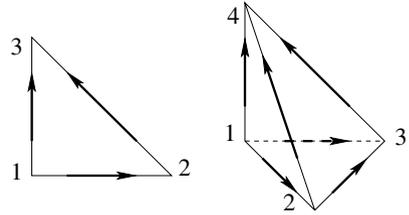
Fig. 15.1 Degrees of freedom of the $\mathbb{N}_{0,d}$ finite element in dimensions two (left) and dimension three (right).

Proposition 15.1 (Face (edge) unisolvence, $d = 2$). *Let $\boldsymbol{v} \in \mathbb{N}_{0,2}$. Let $E \in \mathcal{E}_K$ be an edge of K. Then $\sigma_E^{\mathrm{e}}(\boldsymbol{v}) = 0$ implies that $\boldsymbol{v}_{|E} \cdot \boldsymbol{t}_E = 0$.*

Proof. Since we have established above that $\boldsymbol{v}_{|E} \cdot \boldsymbol{t}_E$ is constant, the assertion follows readily. □

Proposition 15.2 (Finite element, 2D). $(K, \mathbb{N}_{0,2}, \Sigma)$ *is a finite element.*

Proof. Since $\dim(\mathbb{N}_{0,2}) = \mathrm{card}(\Sigma) = 3$, we just need to verify that the only function $\boldsymbol{v} \in \mathbb{N}_{0,2}$ that annihilates the three dofs in Σ is zero. This follows from Proposition 15.1 since $\mathrm{span}\{\boldsymbol{t}_E\}_{E \in \mathcal{E}_K} = \mathbb{R}^2$. □

The above results hold also true if $d = 3$, but the proofs are more intricate since the tangential component on an affine hyperplane of a function in $\mathbb{N}_{0,3}$ is not necessarily constant. Let $F \in \mathcal{F}_K$ be a face of K and let us fix a unit vector $\boldsymbol{n}_F$ normal to F. There are two ways to define the tangential component

of a function $\boldsymbol{v}$ on F: one can define it either as $\boldsymbol{v}\times\boldsymbol{n}_F$ or as $\Pi_F(\boldsymbol{v}) := \boldsymbol{v} - (\boldsymbol{v}\cdot\boldsymbol{n}_F)\boldsymbol{n}_F$. We will use both definitions. The first one is convenient when working with the $\nabla\times$ operator. The second one is more geometric. The two definitions produce ℓ^2-orthogonal vectors since $(\boldsymbol{v}\times\boldsymbol{n}_F)\cdot\Pi_F(\boldsymbol{v}) = 0$ as shown in Figure 15.2.
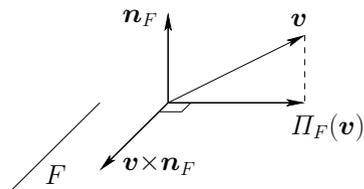
Fig. 15.2 Two possible definitions of the tangential component of a vector.

Proposition 15.3 (Face unisolvence, 3D). *Let $\boldsymbol{v} \in \mathbb{N}_{0,3}$. Let $F \in \mathcal{F}_K$ be a face of K and let $\mathcal{E}_F$ be the collection of the three edges of K forming the boundary of F. Then $\sigma^{\mathrm{e}}_E(\boldsymbol{v}) = 0$ for all $E \in \mathcal{E}_F$ implies that $\boldsymbol{v}_{|F}\times\boldsymbol{n}_F = \boldsymbol{0}$.*

Proof. Let $\widehat{S}^2$ be the unit simplex in $\mathbb{R}^2$. Let $\boldsymbol{T}_F : \widehat{S}^2 \to F$ be defined by $\boldsymbol{T}_F(0,0) := \boldsymbol{z}_p$, $\boldsymbol{T}_F(1,0) := \boldsymbol{z}_q$, $\boldsymbol{T}_F(0,1) := \boldsymbol{z}_r$, where $\boldsymbol{z}_p, \boldsymbol{z}_q, \boldsymbol{z}_r$ are the three vertices of F enumerated by increasing vertex-index. Let $\mathbb{J}_F$ be the 3×2 Jacobian matrix of $\boldsymbol{T}_F$. Note that for all $\widehat{\boldsymbol{y}} \in \mathbb{R}^2$ the vector $\mathbb{J}_F\widehat{\boldsymbol{y}}$ is parallel to F and $\boldsymbol{T}_F(\widehat{\boldsymbol{y}}) - \boldsymbol{z}_p = \mathbb{J}_F\widehat{\boldsymbol{y}}$. Let $\boldsymbol{v} = \boldsymbol{r} + \boldsymbol{q}$ with $\boldsymbol{r} \in \mathbb{P}_{0,3}$ and $\boldsymbol{q} \in \mathbb{S}_{1,3}$. Let us set $\widehat{\boldsymbol{v}} := \mathbb{J}_F^\mathsf{T}\Pi_F(\boldsymbol{v} \circ \boldsymbol{T}_F)$ and let us show that $\widehat{\boldsymbol{v}} \in \mathbb{N}_{0,2}$. For all $\widehat{\boldsymbol{y}} \in \mathbb{R}^2$, we have

$$
\begin{aligned}
\widehat{\boldsymbol{y}}\cdot\widehat{\boldsymbol{v}}(\widehat{\boldsymbol{y}}) &= \widehat{\boldsymbol{y}}\cdot\left(\mathbb{J}_F^\mathsf{T}\Pi_F(\boldsymbol{v}(\boldsymbol{T}_F(\widehat{\boldsymbol{y}})))\right) = \widehat{\boldsymbol{y}}\cdot\left(\mathbb{J}_F^\mathsf{T}\Pi_F(\boldsymbol{r} + \boldsymbol{q}(\boldsymbol{T}_F(\widehat{\boldsymbol{y}})))\right) \\
&= \widehat{\boldsymbol{y}}\cdot\left(\mathbb{J}_F^\mathsf{T}\Pi_F(\boldsymbol{r} + \boldsymbol{q}(\boldsymbol{z}_p) + \boldsymbol{q}(\mathbb{J}_F\widehat{\boldsymbol{y}}))\right) \\
&= \widehat{\boldsymbol{y}}\cdot(\mathbb{J}_F^\mathsf{T}\Pi_F(\boldsymbol{r} + \boldsymbol{q}(\boldsymbol{z}_p))) + (\mathbb{J}_F\widehat{\boldsymbol{y}})\cdot(\boldsymbol{q}(\mathbb{J}_F\widehat{\boldsymbol{y}})).
\end{aligned}
$$

Setting $\widehat{\boldsymbol{c}} := \mathbb{J}_F^\mathsf{T}\Pi_F(\boldsymbol{r} + \boldsymbol{q}(\boldsymbol{z}_p)) \in \mathbb{R}^2$ and using that $\boldsymbol{q} \in \mathbb{S}_{1,3}$, we infer that $\widehat{\boldsymbol{y}}\cdot\widehat{\boldsymbol{v}}(\widehat{\boldsymbol{y}}) = \widehat{\boldsymbol{y}}\cdot\widehat{\boldsymbol{c}}$. Since $\widehat{\boldsymbol{v}} \in \mathbb{P}_{1,2}$, we have $\widehat{\boldsymbol{v}} = \widehat{\boldsymbol{r}} + \widehat{\boldsymbol{q}}$ where $\widehat{\boldsymbol{r}} \in \mathbb{P}_{0,2}$ and $\widehat{\boldsymbol{q}} \in \mathbb{P}^{\mathrm{H}}_{1,2}$. Then $\widehat{\boldsymbol{y}}\cdot\widehat{\boldsymbol{r}} + \widehat{\boldsymbol{y}}\cdot\widehat{\boldsymbol{q}}(\widehat{\boldsymbol{y}}) = \widehat{\boldsymbol{y}}\cdot\widehat{\boldsymbol{c}}$ for all $\widehat{\boldsymbol{y}} \in \mathbb{R}^2$. This implies that the quadratic form $\widehat{\boldsymbol{y}}\cdot\widehat{\boldsymbol{q}}(\widehat{\boldsymbol{y}})$ is zero. Hence, $\widehat{\boldsymbol{v}} \in \mathbb{N}_{0,2}$. Let now $\widehat{E}$ be any of the three edges of $\widehat{S}^2$. Then $E := \boldsymbol{T}_F(\widehat{E})$ is one of the three edges of F. We obtain that

$$
\begin{aligned}
\int_{\widehat{E}}(\widehat{\boldsymbol{v}}\cdot\boldsymbol{t}_{\widehat{E}})\,\mathrm{d}\widehat{l} &= \int_{\widehat{E}}(\mathbb{J}_F^\mathsf{T}\Pi_F(\boldsymbol{v}\circ\boldsymbol{T}_F))\cdot\boldsymbol{t}_{\widehat{E}}\,\mathrm{d}\widehat{l} \\
&= \int_{\widehat{E}}(\boldsymbol{v}\circ\boldsymbol{T}_F)\cdot\boldsymbol{t}_E\,\mathrm{d}\widehat{l} = \frac{|\widehat{E}|}{|E|}\int_E \boldsymbol{v}\cdot\boldsymbol{t}_E\,\mathrm{d}l = |\widehat{E}|\sigma^{\mathrm{e}}_E(\boldsymbol{v}) = 0.
\end{aligned}
$$

Since $\widehat{\boldsymbol{v}} \in \mathbb{N}_{0,2}$ annihilates the three edge dofs in $\widehat{S}^2$, Proposition 15.2 implies that $\widehat{\boldsymbol{v}} = \boldsymbol{0}$. After observing that $\mathrm{im}(\Pi_F)$ is orthogonal to $\ker(\mathbb{J}_F^\mathsf{T})$, we conclude that the tangential component of $\boldsymbol{v}$ on F is zero. $\qquad\square$

Proposition 15.4 (Finite element, 3D). $(K, \mathbb{N}_{0,3}, \Sigma)$ *is a finite element.*

Proof. Since $\dim(\mathbb{N}_{0,3}) = \operatorname{card}(\Sigma) = 6$, we just need to verify that the only function $\boldsymbol{v} \in \mathbb{N}_{0,3}$ that annihilates the six dofs in Σ is zero. Face unisolvence implies that $\boldsymbol{v}_{|F} \times \boldsymbol{n}_F = \boldsymbol{0}$ for all $F \in \mathcal{F}_K$. Let $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ be the canonical basis of $\mathbb{R}^3$. Using (4.11), we infer that $\int_K (\nabla \times \boldsymbol{v}) \cdot \boldsymbol{e}_i \, dx = -\int_{\partial K} (\boldsymbol{v} \times \boldsymbol{n}_K) \cdot \boldsymbol{e}_i \, ds = 0$, where $\boldsymbol{n}_K$ is the outward unit normal to K. Since $\nabla \times \boldsymbol{v}$ is actually constant on K, we have $\nabla \times \boldsymbol{v} = \boldsymbol{0}$, and we have seen that this implies that $\boldsymbol{v} \in \mathbb{P}_{0,3}$, i.e., $\boldsymbol{v} = \nabla p$ for some $p \in \mathbb{P}_{1,3}$. Integrating ∇p along the edges of K, we infer that p takes the same value at all the vertices of K. Hence, p is constant, which in turn implies that $\boldsymbol{v}$ is zero. $\qquad \square$

One can verify that the shape functions are such that

$$\boldsymbol{\theta}_E^{\mathrm{e}}(\boldsymbol{x}) = \lambda_p(\boldsymbol{x}) \nabla \lambda_q - \lambda_q(\boldsymbol{x}) \nabla \lambda_p, \qquad \forall E \in \mathcal{E}_K, \tag{15.4}$$

for all $\boldsymbol{x} \in K$, with $\boldsymbol{t}_E := \boldsymbol{z}_q - \boldsymbol{z}_p$. For every $E' \in \mathcal{E}_K$, we have $\boldsymbol{\theta}_E^{\mathrm{e}} \cdot \boldsymbol{t}_{E'} = \delta_{EE'}$. We refer the reader to Exercise 15.3 for additional properties of the $\mathbb{N}_{0,3}$ shape functions.

15.2 The polynomial space $\mathbb{N}_{k,d}$

Let $k \in \mathbb{N}$ and let $d \in \{2, 3\}$ (the material of this section extends to any dimension $d \geq 2$). Let $\mathbb{P}_{k,d}^{\mathrm{H}}$ be the space of the homogeneous polynomials of degree k (see Definition 14.2). Set $\mathbb{P}_{k,d}^{\mathrm{H}} := [\mathbb{P}_{k,d}^{\mathrm{H}}]^d$ and $\mathbb{P}_{k,d} := [\mathbb{P}_{k,d}]^d$.

Definition 15.5 ($\mathbb{N}_{k,d}$). *We define the following real vector space of $\mathbb{R}^d$-valued polynomials:*

$$\mathbb{N}_{k,d} := \mathbb{P}_{k,d} \oplus \mathbb{S}_{k+1,d}, \quad \text{with} \quad \mathbb{S}_{k+1,d} := \{\boldsymbol{q} \in \mathbb{P}_{k+1,d}^{\mathrm{H}} \mid \boldsymbol{q}(\boldsymbol{x}) \cdot \boldsymbol{x} = 0\}. \tag{15.5}$$

Note that the above sum is direct since $\mathbb{P}_{k,d} \cap \mathbb{S}_{k+1,d} \subset \mathbb{P}_{k,d} \cap \mathbb{P}_{k+1,d}^{\mathrm{H}} = \{\boldsymbol{0}\}$.

Example 15.6 (Space $\mathbb{S}_{2,d}$). The set $\{(-x_2^2, x_1 x_2)^{\mathsf{T}}, (x_1 x_2, -x_1^2)^{\mathsf{T}}\}$ is a basis of $\mathbb{S}_{2,2}$, and the set $\{(-x_2^2, x_1 x_2, 0)^{\mathsf{T}}, (-x_3^2, 0, x_1 x_3)^{\mathsf{T}}, (x_1 x_2, -x_1^2, 0)^{\mathsf{T}}, (0, -x_3^2, x_2 x_3)^{\mathsf{T}}, (x_1 x_3, 0, -x_1^2)^{\mathsf{T}}, (0, x_2 x_3, -x_2^2)^{\mathsf{T}}, (x_2 x_3, -x_1 x_3, 0)^{\mathsf{T}}, (0, x_1 x_3, -x_1 x_2)^{\mathsf{T}}\}$ is a basis of $\mathbb{S}_{2,3}$. Note that $\dim(\mathbb{S}_{2,2}) = 2$ and $\dim(\mathbb{S}_{2,3}) = 8$. $\quad \square$

Lemma 15.7 (Dimension of $\mathbb{N}_{k,d}$). *Let $k \in \mathbb{N}$ and $d \geq 2$. We have*

$$\dim(\mathbb{N}_{k,d}) = \frac{(k+d+1)!}{k!(d-1)!(k+2)}. \tag{15.6}$$

Hence, $\dim(\mathbb{N}_{k,2}) = (k+1)(k+3)$ *and* $\dim(\mathbb{N}_{k,3}) = \frac{1}{2}(k+1)(k+3)(k+4)$.

Proof. (1) Let us first prove that the map $\phi : \mathbb{P}_{k,d}^{\mathrm{H}} \ni \boldsymbol{p} \mapsto \boldsymbol{x} \cdot \boldsymbol{p} \in \mathbb{P}_{k+1,d}^{\mathrm{H}}$ is surjective. By linearity, it suffices to prove that for each monomial $q \in \mathbb{P}_{k+1,d}^{\mathrm{H}}$ s.t. $q(\boldsymbol{x}) := \boldsymbol{x}^{\alpha}$ with $|\alpha| := k+1$, there is $\boldsymbol{r} \in \mathbb{P}_{k,d}^{\mathrm{H}}$ such that $q(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{r}(\boldsymbol{x})$. Let $\{\boldsymbol{e}_i\}_{i \in \{1:d\}}$ be the canonical Cartesian basis of $\mathbb{R}^d$. Since $|\alpha| = k+1 \geq 1$, there exists $i \in \{1:d\}$ s.t. $\alpha_i \geq 1$. Setting $\boldsymbol{r}(\boldsymbol{x}) := x_1^{\alpha_1} \dots x_i^{\alpha_i - 1} \dots x_d^{\alpha_d} \boldsymbol{e}_i$, we have $\boldsymbol{r} \in \mathbb{P}_{k,d}^{\mathrm{H}}$ and $q(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{r}(\boldsymbol{x})$.
(2) Observing that $\ker(\phi) = \mathbb{S}_{k,d}$, the rank nullity theorem implies that $\dim(\mathbb{S}_{k,d}) + \dim(\operatorname{im}(\phi)) = \dim \mathbb{P}_{k,d}^{\mathrm{H}}$, i.e., $\dim(\mathbb{S}_{k,d}) = d \dim \mathbb{P}_{k,d}^{\mathrm{H}} - \dim \mathbb{P}_{k+1,d}^{\mathrm{H}} = d \binom{k+d-1}{k} - \binom{k+d}{k+1} = \binom{k+d-1}{k}(d - \frac{k+d}{k+1}) = k \frac{(k+d-1)!}{(k+1)!(d-2)!}$. The sum in (15.5) being direct, we conclude that

$$\dim(\mathbb{N}_{k,d}) = d \dim(\mathbb{P}_{k,d}) + \dim(\mathbb{S}_{k+1,d})$$
$$= \frac{(k+d)!}{k!(d-1)!} + (k+1)\frac{(k+d)!}{(k+2)!(d-2)!} = \frac{(k+d+1)!}{k!(d-1)!(k+2)}. \qquad \square$$

Lemma 15.8 (Trace space). *Let H be an affine hyperplane in $\mathbb{R}^d$, let $\boldsymbol{n}_H$ be a unit normal vector to H, and let $\boldsymbol{T}_H : \mathbb{R}^{d-1} \to H$ be an affine bijective mapping with Jacobian matrix $\mathbb{J}_H$. Let $\Pi_H(\boldsymbol{v}) := \boldsymbol{v} - (\boldsymbol{v} \cdot \boldsymbol{n}_H)\boldsymbol{n}_H$ be the ℓ^2-orthogonal projection of $\boldsymbol{v}$ onto the tangent space to H (i.e., the linear hyperplane in $\mathbb{R}^d$ parallel to H). Then $\mathbb{J}_H^{\mathsf{T}} \Pi_H(\boldsymbol{v}_{|H}) \in \mathbb{N}_{k,d-1} \circ \boldsymbol{T}_H^{-1}$ for all $\boldsymbol{v} \in \mathbb{N}_{k,d}$.*

Proof. Identical to the proof of Proposition 15.3. $\qquad \square$

Lemma 15.9 ($d = 2$). $\mathbb{N}_{k,2} = \boldsymbol{R}_{\frac{\pi}{2}}(\mathbb{RT}_{k,2})$, *where $\boldsymbol{R}_{\frac{\pi}{2}}$ is the rotation of angle $\frac{\pi}{2}$ in $\mathbb{R}^2$, i.e., $\boldsymbol{R}_{\frac{\pi}{2}}\boldsymbol{x} = (-x_2, x_1)^{\mathsf{T}}$ for all $\boldsymbol{x} = (x_1, x_2)^{\mathsf{T}} \in \mathbb{R}^2$.*

Proof. See Exercise 15.4. $\qquad \square$

Lemma 15.10 (Curl). *Assume $d \in \{2,3\}$. Then $\nabla \times \boldsymbol{v} \in \mathbb{P}_{k,d}$ for all $\boldsymbol{v} \in \mathbb{N}_{k,d}$, and if $\nabla \times \boldsymbol{v} = \boldsymbol{0}$, there is $p \in \mathbb{P}_{k+1,d}$ such that $\boldsymbol{v} = \nabla p$ (that is, $\boldsymbol{v} \in \mathbb{P}_{k,d}$).*

Proof. That $\nabla \times \boldsymbol{v} \in \mathbb{P}_{k,d}$ results from $\mathbb{N}_{k,d} \subset \mathbb{P}_{k+1,d}$. The condition $\nabla \times \boldsymbol{v} = \boldsymbol{0}$ together with $\boldsymbol{v} \in \mathbb{N}_{k,d} \subset \mathbb{P}_{k+1,d}$ implies that there is $p \in \mathbb{P}_{k+2,d}$ such that $\boldsymbol{v} = \nabla p$. The definition of $\mathbb{N}_{k,d}$ implies that $\boldsymbol{v} = \nabla p_1 + \nabla p_2$ with $p_1 \in \mathbb{P}_{k+1,d}$ and $\nabla p_2 \in \mathbb{S}_{k+1,d}$. We infer that $p_2(\boldsymbol{x}) - p_2(\boldsymbol{0}) = \int_0^1 \nabla p_2(t\boldsymbol{x}) \cdot (t\boldsymbol{x}) t^{-1} \, dt = 0$, which means that p_2 is constant. Hence, $\boldsymbol{v} = \nabla p_1$ with $p_1 \in \mathbb{P}_{k+1,d}$. $\qquad \square$

15.3 Simplicial Nédélec elements

Let $k \in \mathbb{N}$ and let $d \in \{2,3\}$. Let K be a simplex in $\mathbb{R}^d$. In this section, we define the dofs in order to make the triple $(K, \mathbb{N}_{k,d}, \Sigma)$ a finite element. The construction can be generalized to any dimension.

15.3.1 Two-dimensional case

Let us orient the three edges $E \in \mathcal{E}_K$ of K with the edge vectors $\boldsymbol{t}_E$. Let us orient K with the two vectors $\{\boldsymbol{t}_{K,j}\}_{j \in \{1,2\}}$ which are the edge vectors for the two edges of K sharing the vertex with the lowest enumeration index. Note that $\{\boldsymbol{t}_{K,j}\}_{j \in \{1,2\}}$ is a basis of $\mathbb{R}^2$. Let $\boldsymbol{T}_E$ be an affine bijective mapping from the unit simplex $\widehat{S}^1 := [0,1]$ in $\mathbb{R}$ onto E. We define the dofs of the two-dimensional Nédélec element $(K, \mathbb{N}_{k,2}, \Sigma)$ as follows:

$$\sigma^{\mathrm{e}}_{E,m}(\boldsymbol{v}) := \frac{1}{|E|} \int_E (\boldsymbol{v} \cdot \boldsymbol{t}_E)(\mu_m \circ \boldsymbol{T}_E^{-1}) \, \mathrm{d}l, \qquad \forall E \in \mathcal{E}_K, \qquad (15.7a)$$

$$\sigma^{\mathrm{c}}_{j,m}(\boldsymbol{v}) := \frac{1}{|K|} \int_K (\boldsymbol{v} \cdot \boldsymbol{t}_{K,j}) \psi_m \, \mathrm{d}x, \qquad \forall j \in \{1{:}2\}, \qquad (15.7b)$$

where $\{\mu_m\}_{m \in \{1 : n^{\mathrm{e}}_{\mathrm{sh}}\}}$ is a basis of $\mathbb{P}_{k,1}$ with $n^{\mathrm{e}}_{\mathrm{sh}} := \dim(\mathbb{P}_{k,1}) = k+1$, and $\{\psi_m\}_{m \in \{1 : n^{\mathrm{c}}_{\mathrm{sh}}\}}$ is a basis of $\mathbb{P}_{k-1,2}$ with $n^{\mathrm{c}}_{\mathrm{sh}} := \dim(\mathbb{P}_{k-1,2}) = \frac{1}{2}k(k+1)$ if $k \geq 1$. Since $\mathbb{N}_{k,2} = \boldsymbol{R}_{\frac{\pi}{2}}(\mathbb{RT}_{k,2})$ owing to Lemma 15.9 and since the above dofs are those of the $\mathbb{RT}_{k,2}$ finite element once the edges (faces) are oriented by the vectors $\boldsymbol{\nu}_E := \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{t}_E)$ and K is oriented by the vectors $\boldsymbol{\nu}_{K,j} := \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{t}_{K,j})$, it follows from Proposition 14.15 that the triple $(K, \mathbb{N}_{k,2}, \Sigma)$ is a finite element for all $k \geq 0$. The unit of all the above dofs is a length times the dimension of $\boldsymbol{v}$.

Remark 15.11 (2D Piola transformations). Owing to the identity $\mathbb{A}^{\mathsf{T}} = \det(\mathbb{A}) \boldsymbol{R}_{\frac{\pi}{2}}^{-1} \mathbb{A}^{-1} \boldsymbol{R}_{\frac{\pi}{2}}$ for all $\mathbb{A} \in \mathbb{R}^{2 \times 2}$, the two-dimensional contravariant and covariant Piola transformations satisfy $\boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{\psi}^{\mathrm{c}}_K(\boldsymbol{v})) = \boldsymbol{\psi}^{\mathrm{d}}_K(\boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{v}))$. $\qquad\square$

15.3.2 Three-dimensional case

Let K be a simplex (tetrahedron) in $\mathbb{R}^3$. Let $\mathcal{E}_K$ be the collection of the six edges of K and let $\mathcal{F}_K$ be the collection of the four faces of K. Each edge $E \in \mathcal{E}_K$ is oriented by the edge vector $\boldsymbol{t}_E := \boldsymbol{z}_q - \boldsymbol{z}_p$, where $\boldsymbol{z}_p, \boldsymbol{z}_q$ are the two vertices of E with $p < q$ (note that $\|\boldsymbol{t}_E\|_{\ell^2} = |E|$). Each face $F \in \mathcal{F}_K$ is oriented by the two edge vectors $\{\boldsymbol{t}_{F,j}\}_{j \in \{1,2\}}$ with $\boldsymbol{t}_{F,1} := \boldsymbol{z}_q - \boldsymbol{z}_p$, $\boldsymbol{t}_{F,2} := \boldsymbol{z}_r - \boldsymbol{z}_p$, where $\boldsymbol{z}_p, \boldsymbol{z}_q, \boldsymbol{z}_r$ are the three vertices of F with $p < q < r$. Note that the unit normal vector $\boldsymbol{n}_F$ is then defined as $\boldsymbol{t}_{F,1} \times \boldsymbol{t}_{F,2} / \|\boldsymbol{t}_{F,1} \times \boldsymbol{t}_{F,2}\|_{\ell^2}$; see for instance (10.9). Note also that $\{\boldsymbol{t}_{F,j}\}_{j \in \{1,2\}}$ is a basis of the tangent space of the affine hyperplane supporting F. Finally, the cell K is oriented by the three edge vectors $\{\boldsymbol{t}_{K,j}\}_{j \in \{1:3\}}$ with $\boldsymbol{t}_{K,1} := \boldsymbol{z}_q - \boldsymbol{z}_p$, $\boldsymbol{t}_{K,2} := \boldsymbol{z}_r - \boldsymbol{z}_p$, $\boldsymbol{t}_{K,3} := \boldsymbol{z}_s - \boldsymbol{z}_p$, where $\boldsymbol{z}_p, \boldsymbol{z}_q, \boldsymbol{z}_r, \boldsymbol{z}_s$ are the four vertices of K with $p < q < r < s$. Note that $\{\boldsymbol{t}_{K,j}\}_{j \in \{1:3\}}$ is a basis of $\mathbb{R}^3$. In order to define dofs using moments on the edges and moments on the faces of K, we introduce affine bijective mappings $\boldsymbol{T}_F : \widehat{S}^2 \to F$ and $\boldsymbol{T}_E : \widehat{S}^1 \to E$, where $\widehat{S}^2$ is the unit simplex in $\mathbb{R}^2$ and $\widehat{S}^1$ is the unit simplex in $\mathbb{R}$; see Figure 15.3. For instance, after enumerating the vertices of $\widehat{S}^1$, $\widehat{S}^2$, these mappings can be constructed by using the increasing vertex-index enumeration technique of §10.2.
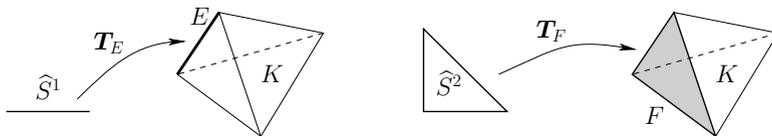
Fig. 15.3 Reference edge $\widehat{S}^1$ and reference face $\widehat{S}^2$ with the corresponding mappings.

Definition 15.12 (dofs). *The set Σ is defined to be the collection of the following linear forms acting on $\mathbb{N}_{k,3}$:*

$$\sigma_{E,m}^{\mathrm{e}}(\boldsymbol{v}) := \frac{1}{|E|} \int_E (\boldsymbol{v}\cdot\boldsymbol{t}_E)(\mu_m \circ \boldsymbol{T}_E^{-1})\,\mathrm{d}l, \quad \forall E \in \mathcal{E}_K \tag{15.8a}$$

$$\sigma_{F,j,m}^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v}\cdot\boldsymbol{t}_{F,j})(\zeta_m \circ \boldsymbol{T}_F^{-1})\,\mathrm{d}s, \quad \forall F \in \mathcal{F}_K, \forall j \in \{1,2\}, \tag{15.8b}$$

$$\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) := \frac{1}{|K|} \int_K (\boldsymbol{v}\cdot\boldsymbol{t}_{K,j})\psi_m\,\mathrm{d}x, \qquad \forall j \in \{1,2,3\}, \tag{15.8c}$$

where $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$ is a basis of $\mathbb{P}_{k,1}$ with $n_{\mathrm{sh}}^{\mathrm{e}} := k+1$, $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ is a basis of $\mathbb{P}_{k-1,2}$ with $n_{\mathrm{sh}}^{\mathrm{f}} := \frac{1}{2}(k+1)k$ if $k \geq 1$, and $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ is a basis of $\mathbb{P}_{k-2,3}$ with $n_{\mathrm{sh}}^{\mathrm{c}} := \frac{1}{6}(k+1)k(k-1)$ if $k \geq 2$. We regroup the dofs as follows:

$$\Sigma_E^{\mathrm{e}} := \{\sigma_{E,m}^{\mathrm{e}}\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}, \qquad\qquad \forall E \in \mathcal{E}_K, \tag{15.9a}$$

$$\Sigma_F^{\mathrm{f}} := \{\sigma_{F,j,m}^{\mathrm{f}}\}_{(j,m)\in\{1,2\}\times\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}, \qquad \forall F \in \mathcal{F}_K, \tag{15.9b}$$

$$\Sigma^{\mathrm{c}} := \{\sigma_{j,m}^{\mathrm{c}}\}_{(j,m)\in\{1:3\}\times\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}. \tag{15.9c}$$

Remark 15.13 (dofs). The unit of all the dofs is a length times the dimension of $\boldsymbol{v}$. For the cell dofs, we could also have written $\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) := \ell_K^{-2} \int_K (\boldsymbol{v}\cdot\boldsymbol{e}_j)\psi_m\,\mathrm{d}x$, where ℓ_K is a length scale associated with K and $\{\boldsymbol{e}_j\}_{j\in\{1:d\}}$ is the canonical Cartesian basis of $\mathbb{R}^d$. We will see that the definition (15.8c) is more natural when using the covariant Piola transformation to generate Nédélec finite elements. The dofs are defined here on $\mathbb{N}_{k,d}$. Their extension to some larger space $\boldsymbol{V}(K)$ is addressed in Chapters 16 and 17. □

Lemma 15.14 (Invariance). *Assume that every affine bijective mapping $\boldsymbol{S} : \widehat{S}^1 \to \widehat{S}^1$ (resp., $\boldsymbol{S} : \widehat{S}^2 \to \widehat{S}^2$) leaves the basis $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$ (resp., $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$) globally invariant. Then for all $E \in \mathcal{E}_K$ and all $F \in \mathcal{F}_K$, the set Σ_E^{e} and Σ_F^{f} are independent of the affine bijective mapping $\boldsymbol{T}_E$ and $\boldsymbol{T}_F$, respectively.*

Proof. Similar to that of Lemma 14.12; see also Example 14.13 for the invariance w.r.t. vertex permutation.   □

The following result is important in view of $\boldsymbol{H}(\mathrm{curl})$-conformity.

Lemma 15.15 (Face unisolvence).   *Let* $\boldsymbol{v} \in \mathbb{N}_{k,3}$ *and let* $F \in \mathcal{F}_K$ *be a face of* K. *Let* $\mathcal{E}_F$ *be the collection of the (three) edges forming the boundary of* F, *let* $\Sigma_F^{\mathrm{e}} := \bigcup_{E \in \mathcal{E}_F} \Sigma_E^{\mathrm{e}}$, *and let* $\boldsymbol{n}_F$ *be a unit normal to* F. *We have the following equivalence:*

$$[\, \sigma(\boldsymbol{v}) = 0, \quad \forall \sigma \in \Sigma_F^{\mathrm{f}} \cup \Sigma_F^{\mathrm{e}} \,] \iff [\, \boldsymbol{v}_{|F} {\times} \boldsymbol{n}_F = \boldsymbol{0} \,]. \tag{15.10}$$

Moreover, both assertions in (15.10) *imply that* $(\nabla {\times} \boldsymbol{v})_{|F} {\cdot} \boldsymbol{n}_F = 0$.

Proof. We only need to prove the implication in (15.10) since the converse is evident. The proof is an extension of that of Proposition 15.3 accounting for the richer structure of the dofs. We introduce $\widehat{\boldsymbol{v}} := \mathbb{J}_F^\mathsf{T} \Pi_F (\boldsymbol{v} \circ \boldsymbol{T}_F)$. It can be shown that $\widehat{\boldsymbol{v}} \in \mathbb{N}_{k,2}$; see Exercise 15.6. The unit simplex $\widehat{S}^2$ is oriented by the two edge vectors $\{\widehat{\boldsymbol{t}}_j\}_{j \in \{1,2\}}$ s.t. $\mathbb{J}_F \widehat{\boldsymbol{t}}_j = \boldsymbol{t}_{F,j} \circ \boldsymbol{T}_F$ for all $j \in \{1,2\}$. For the face dofs, we have

$$\begin{aligned}
\frac{1}{|\widehat{S}^2|} \int_{\widehat{S}^2} (\widehat{\boldsymbol{v}} {\cdot} \widehat{\boldsymbol{t}}_j) \zeta_m \, \mathrm{d}\widehat{s} &= \frac{1}{|\widehat{S}^2|} \int_{\widehat{S}^2} ((\mathbb{J}_F^\mathsf{T} (\boldsymbol{v} - (\boldsymbol{v} {\cdot} \boldsymbol{n}_F) \boldsymbol{n}_F) \circ \boldsymbol{T}_F) {\cdot} \widehat{\boldsymbol{t}}_j) \zeta_m \, \mathrm{d}\widehat{s} \\
&= \frac{1}{|\widehat{S}^2|} \int_{\widehat{S}^2} (((\boldsymbol{v} - (\boldsymbol{v} {\cdot} \boldsymbol{n}_F) \boldsymbol{n}_F) {\cdot} \boldsymbol{t}_{F,j}) \circ \boldsymbol{T}_F) \zeta_m \, \mathrm{d}\widehat{s} \\
&= \frac{1}{|\widehat{S}^2|} \int_{\widehat{S}^2} ((\boldsymbol{v} {\cdot} \boldsymbol{t}_{F,j}) \circ \boldsymbol{T}_F) \zeta_m \, \mathrm{d}\widehat{s} \\
&= \frac{1}{|F|} \int_F (\boldsymbol{v} {\cdot} \boldsymbol{t}_{F,j})(\zeta_m \circ \boldsymbol{T}_F^{-1}) \, \mathrm{d}s = \sigma_{F,j,m}^{\mathrm{f}}(\boldsymbol{v}) = 0.
\end{aligned}$$

One proves similarly that the edge dofs vanish. This proves that $\widehat{\boldsymbol{v}} = \boldsymbol{0}$ because $\widehat{\boldsymbol{v}} \in \mathbb{N}_{k,2}$. Since $\mathbb{J}_F^\mathsf{T}$ has full rank, we infer that $\Pi_F(\boldsymbol{v}_{|F}) = \boldsymbol{0}$, which implies that $\boldsymbol{v}_{|F} {\times} \boldsymbol{n}_F = \Pi_F(\boldsymbol{v}_{|F}) {\times} \boldsymbol{n}_F = \boldsymbol{0}$. Finally, $(\nabla {\times} \boldsymbol{v})_{|F} {\cdot} \boldsymbol{n}_F = 0$ immediately follows from $\boldsymbol{v}_{|F} {\times} \boldsymbol{n}_F = \boldsymbol{0}$. $\qquad\square$

Proposition 15.16 (Finite element). $(K, \mathbb{N}_{k,3}, \Sigma)$ *is a finite element.*

Proof. Observe first that the cardinality of Σ can be evaluated as follows:

$$\mathrm{card}(\Sigma) = 3n_{\mathrm{sh}}^{\mathrm{c}} + 2 \times 4n_{\mathrm{sh}}^{\mathrm{f}} + 6n_{\mathrm{sh}}^{\mathrm{e}} = 3\binom{k+1}{3} + 8\binom{k+1}{2} + 6(k+1)$$

$$= \frac{1}{2}(k+1)(k+3)(k+4) = \dim(\mathbb{N}_{k,3}).$$

Hence, the assertion will be proved once it is established that zero is the only function in $\mathbb{N}_{k,3}$ that annihilates all the dofs in Σ. Let $\boldsymbol{v} \in \mathbb{N}_{k,3}$ be such that $\sigma(\boldsymbol{v}) = 0$ for all σ in Σ. We are going to show that $\boldsymbol{v} = \boldsymbol{0}$. Owing to Lemma 15.15, $\boldsymbol{v}_{|F} {\times} \boldsymbol{n}_F = \boldsymbol{0}$ and $(\nabla {\times} \boldsymbol{v})_{|F} {\cdot} \boldsymbol{n}_F = 0$ for every face $F \in \mathcal{F}_K$.
(1) Let us prove that $\boldsymbol{w} := \nabla {\times} \boldsymbol{v} = \boldsymbol{0}$. Since $\boldsymbol{w} \in \mathbb{P}_{k,3} \subset \mathbb{RT}_{k,3}$, it suffices to prove that $\boldsymbol{w}$ annihilates all the dofs of the $\mathbb{RT}_{k,3}$ element. Since $\boldsymbol{w}_{|F} {\cdot} \boldsymbol{n}_F = 0$, $\boldsymbol{w}$ annihilates all the dofs associated with the faces of K. In addition, if $k \geq 1$,

we observe that $\int_K \boldsymbol{w}\cdot\boldsymbol{q}\,\mathrm{d}x = \int_K \nabla\times\boldsymbol{v}\cdot\boldsymbol{q}\,\mathrm{d}x = \int_K \boldsymbol{v}\cdot\nabla\times\boldsymbol{q}\,\mathrm{d}x$ for all $\boldsymbol{q} \in \mathbb{P}_{k-1,3}$, since $\boldsymbol{v}\times\boldsymbol{n}_K = \boldsymbol{0}$ on ∂K, where $\boldsymbol{n}_K$ is the outward unit normal to K. This in turn implies that $\int_K \boldsymbol{w}\cdot\boldsymbol{q}\,\mathrm{d}x = 0$ since $\nabla\times\boldsymbol{q} \in \mathbb{P}_{k-2,3}$ and $\sigma(\boldsymbol{v}) = 0$ for all $\sigma \in \varSigma^{\mathrm{c}}$ if $k \geq 2$. The statement is obvious if $k = 1$. In conclusion, $\nabla\times\boldsymbol{v} = \boldsymbol{w} = \boldsymbol{0}$.

(2) Using Lemma 15.10, we infer that there is $p \in \mathbb{P}_{k+1,3}$ such that $\boldsymbol{v} = \nabla p$. The condition $\boldsymbol{v}\times\boldsymbol{n}_K = \boldsymbol{0}$ on ∂K implies that p is constant on ∂K. Without loss of generality, we take this constant equal to zero. This in turn implies that $p = 0$ if $k \leq 2$ (see Exercise 7.5(iii)), so that it remains to consider the case $k \geq 3$. In this case, we infer that $p = \lambda_0 \ldots \lambda_3 r$ where λ_i, $i \in \{0{:}3\}$, are the barycentric coordinates in K and $r \in \mathbb{P}_{k-3,3}$. Writing this polynomial in the form $r(\boldsymbol{x}) = \sum_{|\alpha|\leq k-3} a_\alpha \boldsymbol{x}^\alpha$, we consider the field $\boldsymbol{q}(\boldsymbol{x}) := \sum_{|\alpha|\leq k-3} \frac{1}{\alpha_1+1} a_\alpha x_1 \boldsymbol{x}^\alpha \boldsymbol{e}_1$, where $\boldsymbol{e}_1$ is the first vector of the canonical Cartesian basis of $\mathbb{R}^3$. Since $\boldsymbol{q} \in \mathbb{P}_{k-2,3}$, the fact that $\sigma(\boldsymbol{v}) = 0$ for all $\sigma \in \varSigma^{\mathrm{c}}$ implies that $\int_K \boldsymbol{v}\cdot\boldsymbol{q}\,\mathrm{d}x = 0$. Integration by parts and the fact that $p_{|\partial K} = 0$ yield $0 = \int_K \boldsymbol{v}\cdot\boldsymbol{q}\,\mathrm{d}x = -\int_K p\nabla\cdot\boldsymbol{q}\,\mathrm{d}x = -\int_K \lambda_0 \ldots \lambda_3 r^2 \,\mathrm{d}x$. In conclusion, $r = 0$, so that $\boldsymbol{v} = \nabla p = \boldsymbol{0}$. $\qquad\square$

The shape functions $\{\theta_i\}_{i\in\mathcal{N}}$ associated with the dofs $\{\sigma_i\}_{i\in\mathcal{N}}$ defined in (15.8) can be constructed by choosing a basis $\{\phi_i\}_{i\in\mathcal{N}}$ of the polynomial space $\mathbb{N}_{k,3}$ and by inverting the corresponding generalized Vandermonde matrix as explained in Proposition 5.5. Recall that this matrix has entries $\mathcal{V}_{ij} = \sigma_j(\phi_i)$ and that the i-th line of $\mathcal{V}^{-1}$ gives the components of the shape function θ_i in the basis $\{\phi_i\}_{i\in\mathcal{N}}$. The basis $\{\phi_i\}_{i\in\mathcal{N}}$ chosen in Bonazzoli and Rapetti [31] (built by dividing the simplex into smaller sub-simplices following the ideas in Rapetti and Bossavit [163], Christiansen and Rapetti [70]) is particularly interesting since the entries of $\mathcal{V}^{-1}$ are integers. One could also choose $\{\phi_i\}_{i\in\mathcal{N}}$ to be the hierarchical basis of $\mathbb{N}_{k,d}$ constructed in Fuentes et al. [103, §7.2]. This basis can be organized into functions attached to the the edges of K, the faces of K, and to K itself, in such a way that the generalized Vandermonde matrix $\mathcal{V}$ is block-triangular (notice though that this matrix is not block-diagonal). For earlier work on shape functions and basis functions for the $\mathbb{N}_{k,3}$ element, see Webb [197], Gopalakrishnan et al. [109].

Remark 15.17 (Dof independence). As in Remark 14.16, the results from Exercise 5.2 imply that the interpolation operator $\mathcal{I}_K^{\mathrm{c}}$ associated with the $\mathbb{N}_{k,3}$ element is independent of the bases $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$, $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$, and $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ that are used to define the dofs in (15.8). The interpolation operator is also independent of the mappings $\boldsymbol{T}_E$, $\boldsymbol{T}_F$ and of the orientation vectors $\{\boldsymbol{t}_E\}_{E\in\mathcal{E}_K}$, $\{\boldsymbol{t}_{F,j}\}_{F\in\mathcal{F}_K, j\in\{1,2\}}$, and $\{\boldsymbol{t}_{K,j}\}_{j\in\{1,2,3\}}$. $\qquad\square$

Remark 15.18 (Literature). The $\mathbb{N}_{k,d}$ finite element has been introduced by Nédélec [151]; see also Weil [198], Whitney [199] for $k = 0$. It is an accepted practice in the literature to call this element *edge element* or *Nédélec element*. See also Bossavit [36, Chap. 3], Hiptmair [117], Monk [145, Chap. 5]. $\qquad\square$

15.4 Generation of Nédélec elements

Let $\widehat{K}$ be the reference simplex in $\mathbb{R}^3$. Let $\mathcal{T}_h$ be an affine simplicial mesh. Let $K = \boldsymbol{T}_K(\widehat{K})$ be a mesh cell where $\boldsymbol{T}_K : \widehat{K} \to K$ is the geometric mapping, and let $\mathbb{J}_K$ be the Jacobian matrix of $\boldsymbol{T}_K$. Let $F \in \mathcal{F}_K$ be a face of K. We have $F = \boldsymbol{T}_K(\widehat{F})$, where $\widehat{F} \in \mathcal{F}_{\widehat{K}}$ is a face of $\widehat{K}$. Similarly, let $E \in \mathcal{E}_K$ be an edge of K. We have $E = \boldsymbol{T}_K(\widehat{E})$ where $\widehat{E} \in \mathcal{E}_{\widehat{K}}$ is an edge of $\widehat{K}$. Using the increasing vertex-index enumeration, Theorem 10.8 shows that it is possible to orient the edges E and $\widehat{E}$ in a way that is compatible with the geometric mapping $\boldsymbol{T}_K$. This means that the unit tangent vectors $\boldsymbol{\tau}_E$ and $\widehat{\boldsymbol{\tau}}_{\widehat{E}}$ satisfy (10.6a), i.e., $\boldsymbol{\tau}_E = \boldsymbol{\Phi}_K^{\mathrm{c}}(\widehat{\boldsymbol{\tau}}_{\widehat{E}})$ with $\boldsymbol{\Phi}_K^{\mathrm{c}}$ defined in (9.14b). In other words, we have

$$\boldsymbol{\tau}_E \circ \boldsymbol{T}_{K|\widehat{E}} = \frac{1}{\|\mathbb{J}_K \widehat{\boldsymbol{\tau}}_{\widehat{E}}\|_{\ell^2}} \mathbb{J}_K \widehat{\boldsymbol{\tau}}_{\widehat{E}}. \tag{15.11}$$

Since $\boldsymbol{t}_E := |E|\boldsymbol{\tau}_E$, $\widehat{\boldsymbol{\tau}}_{\widehat{E}} := |\widehat{E}|\widehat{\boldsymbol{t}}_{\widehat{E}}$ and since $|E| = \|\mathbb{J}_K \widehat{\boldsymbol{\tau}}_{\widehat{E}}\|_{\ell^2}|\widehat{E}|$ owing to Lemma 9.12, we infer that

$$\boldsymbol{t}_E \circ \boldsymbol{T}_{K|\widehat{E}} = \mathbb{J}_K \widehat{\boldsymbol{t}}_{\widehat{E}}. \tag{15.12}$$

We also orient the faces of K by using the two edge vectors originating from the vertex with the lowest index in each face. We finally orient K by using the three edge vectors originating from the vertex with the lowest index in K. Reasoning as above, we infer that

$$\boldsymbol{t}_{F,j} \circ \boldsymbol{T}_{K|\widehat{F}} = \mathbb{J}_K \widehat{\boldsymbol{t}}_{\widehat{F},j}, \ \forall j \in \{1,2\} \quad \boldsymbol{t}_{K,j} \circ \boldsymbol{T}_K = \mathbb{J}_K \widehat{\boldsymbol{t}}_{\widehat{K},j}, \ \forall j \in \{1{:}3\}. \tag{15.13}$$

Recall the covariant Piola transformation introduced in (9.9b) such that

$$\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) := \mathbb{J}_K^{\mathsf{T}}(\boldsymbol{v} \circ \boldsymbol{T}_K), \tag{15.14}$$

and the pullback by the geometric mapping such that $\psi_K^{\mathrm{g}}(q) := q \circ \boldsymbol{T}_K$.

Lemma 15.19 (Transformation of dofs). *Let $\boldsymbol{v} \in \boldsymbol{C}^0(K)$ and let $q \in C^0(K)$. The following holds true:*

$$\frac{1}{|E|} \int_E (\boldsymbol{v} \cdot \boldsymbol{t}_E) q \, \mathrm{d}l = \frac{1}{|\widehat{E}|} \int_{\widehat{E}} (\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) \cdot \widehat{\boldsymbol{t}}_{\widehat{E}}) \psi_K^{\mathrm{g}}(q) \, \mathrm{d}\widehat{l}, \qquad \forall E \in \mathcal{E}_K,$$

$$\frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{t}_{F,j}) q \, \mathrm{d}s = \frac{1}{|\widehat{F}|} \int_{\widehat{F}} (\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) \cdot \widehat{\boldsymbol{t}}_{\widehat{F},j}) \psi_K^{\mathrm{g}}(q) \, \mathrm{d}\widehat{s}, \quad \forall F \in \mathcal{F}_K, \ j \in \{1,2\},$$

$$\frac{1}{|K|} \int_K (\boldsymbol{v} \cdot \boldsymbol{t}_{K,j}) q \, \mathrm{d}x = \frac{1}{|\widehat{K}|} \int_{\widehat{K}} (\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) \cdot \widehat{\boldsymbol{t}}_{\widehat{K},j}) \psi_K^{\mathrm{g}}(q) \, \mathrm{d}\widehat{x}, \quad \forall j \in \{1{:}3\}.$$

Proof. The first identity is nothing but (10.7b) from Lemma 10.4, which itself is a reformulation of (9.15b) from Lemma 9.13 (the fact that $\boldsymbol{T}_K$ is affine is not used here). The proof of the other two identities is similar to (9.15b)

using (15.13) and the fact that $\mathrm{d}s = \frac{|F|}{|\widehat{F}|}\,\mathrm{d}\widehat{s}$, $\mathrm{d}x = \frac{|K|}{|\widehat{K}|}\,\mathrm{d}\widehat{x}$ since $\boldsymbol{T}_K$ is affine. For instance, we have

$$
\frac{1}{|F|}\int_F (\boldsymbol{v}\cdot\boldsymbol{t}_{F,j})q\,\mathrm{d}s = \frac{1}{|F|}\int_{\widehat{F}} (\boldsymbol{v}\cdot\boldsymbol{t}_{F,j})\circ\boldsymbol{T}_{K|\widehat{F}}(q\circ\boldsymbol{T}_{K|\widehat{F}})\frac{|F|}{|\widehat{F}|}\,\mathrm{d}\widehat{s}
$$

$$
= \frac{1}{|\widehat{F}|}\int_{\widehat{F}} ((\mathbb{J}_K^{\mathsf{T}}\boldsymbol{v})\cdot(\mathbb{J}_K^{-1}\boldsymbol{t}_{F,j}))\circ\boldsymbol{T}_{K|\widehat{F}}\psi_K^{\mathrm{g}}(q)\,\mathrm{d}\widehat{s}
$$

$$
= \frac{1}{|\widehat{F}|}\int_{\widehat{F}} (\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\cdot\widehat{\boldsymbol{t}}_{\widehat{F},j})\psi_K^{\mathrm{g}}(q)\,\mathrm{d}\widehat{s}. \qquad \square
$$

Proposition 15.20 (Generation). *Let $(\widehat{K},\widehat{\boldsymbol{P}},\widehat{\Sigma})$ be a simplicial Nédélec element with edge, face, and cell dofs defined by using the polynomial bases $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$, $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ (if $k\geq 1$), and $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ (if $k\geq 2$) of $\mathbb{P}_{k,1}$, $\mathbb{P}_{k-1,2}$, and $\mathbb{P}_{k-2,3}$, respectively, as in (15.8). Assume that the geometric mapping $\boldsymbol{T}_K$ is affine and that (15.12)–(15.13) hold true. Then the finite element $(K,\boldsymbol{P}_K,\Sigma_K)$ generated using Proposition 9.2 with the covariant Piola transformation (15.14) is a simplicial Nédélec element with dofs*

$$
\sigma_{E,m}^{\mathrm{e}}(\boldsymbol{v}) = \frac{1}{|E|}\int_E (\boldsymbol{v}\cdot\boldsymbol{t}_E)(\mu_m\circ\boldsymbol{T}_{K,E}^{-1})\,\mathrm{d}l, \qquad \forall E\in\mathcal{E}_K, \tag{15.15a}
$$

$$
\sigma_{F,j,m}^{\mathrm{f}}(\boldsymbol{v}) = \frac{1}{|F|}\int_F (\boldsymbol{v}\cdot\boldsymbol{t}_{F,j})(\zeta_m\circ\boldsymbol{T}_{K,F}^{-1})\,\mathrm{d}s, \qquad \forall F\in\mathcal{F}_K,\ \forall j\in\{1,2\}, \tag{15.15b}
$$

$$
\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) = \frac{1}{|K|}\int_K (\boldsymbol{v}\cdot\boldsymbol{t}_{K,j})(\psi_m\circ\boldsymbol{T}_K^{-1})\,\mathrm{d}x, \qquad \forall j\in\{1,2,3\}, \tag{15.15c}
$$

where $\boldsymbol{T}_{K,E} := \boldsymbol{T}_{K|\widehat{E}}\circ\boldsymbol{T}_{\widehat{E}} : \widehat{S}^1 \to E$ and $\boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}}\circ\boldsymbol{T}_{\widehat{F}} : \widehat{S}^2 \to F$ are the affine bijective mappings that map vertices with increasing indices.

Proof. Let us first prove that $\boldsymbol{P}_K = \mathbb{N}_{k,3}$. We can write $\boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := \mathbb{J}_K\widehat{\boldsymbol{x}} + \boldsymbol{b}_K$ with $\mathbb{J}_K\in\mathbb{R}^{3\times 3}$ and $\boldsymbol{b}_K\in\mathbb{R}^3$. Let $\boldsymbol{v}$ be a member of $\boldsymbol{P}_K$. Then $\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) = \widehat{\boldsymbol{p}}+\widehat{\boldsymbol{q}}$ with $\widehat{\boldsymbol{p}}\in\mathbb{P}_{k,3}$ and $\widehat{\boldsymbol{q}}\in\mathbb{S}_{k+1,3}$, yielding $\boldsymbol{v} = \mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{p}}\circ\boldsymbol{T}_K^{-1} + \mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{q}}\circ\boldsymbol{T}_K^{-1}$. Since each component of $\widehat{\boldsymbol{q}}$ is in $\mathbb{P}_{k+1,3}^{\mathrm{H}}$, we infer that $\widehat{\boldsymbol{q}}\circ\boldsymbol{T}_K^{-1}(\boldsymbol{x}) = \widehat{\boldsymbol{q}}(\mathbb{J}_K^{-1}\boldsymbol{x} - \mathbb{J}_K^{-1}\boldsymbol{b}_K) = \widehat{\boldsymbol{q}}(\mathbb{J}_K^{-1}\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})$, where $\boldsymbol{r}\in\mathbb{P}_{k,3}$; see Exercise 14.4. As a result, $\boldsymbol{v} = (\boldsymbol{p}+\boldsymbol{r}) + \boldsymbol{q}$, where $\boldsymbol{p} = \mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{p}}\circ\boldsymbol{T}_K^{-1} \in \mathbb{P}_{k,3}$ and $\boldsymbol{q} = \mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{q}}\circ\mathbb{J}_K^{-1}$. Note that $\boldsymbol{p}+\boldsymbol{r}\in\mathbb{P}_{k,3}$ and $\widehat{\boldsymbol{q}}\circ\mathbb{J}_K^{-1}$ is a member of $\mathbb{P}_{k+1,3}^{\mathrm{H}}$, which implies that $\boldsymbol{q}$ is also in $\mathbb{P}_{k+1,3}^{\mathrm{H}}$. Moreover, $\boldsymbol{x}\cdot(\mathbb{J}_K^{-\mathsf{T}}\widehat{\boldsymbol{q}}(\mathbb{J}_K^{-1}\boldsymbol{x})) = (\mathbb{J}_K^{-1}\boldsymbol{x})\cdot\widehat{\boldsymbol{q}}(\mathbb{J}_K^{-1}\boldsymbol{x}) = 0$ which in turn implies that $\boldsymbol{q}\in\mathbb{S}_{k+1,3}$. In conclusion, $\boldsymbol{v}\in\mathbb{N}_{k,3}$, meaning that $\boldsymbol{P}_K \subset \mathbb{N}_{k,3}$. The converse statement follows from a dimension argument. Finally, the definition of the dofs results from Lemma 15.19, and the properties of the mappings $\boldsymbol{T}_{K,E}$ and $\boldsymbol{T}_{K,F}$ from those of $\boldsymbol{T}_K$, $\boldsymbol{T}_{\widehat{E}}$, and $\boldsymbol{T}_{\widehat{F}}$. $\qquad\square$

Remark 15.21 (Unit). The shape functions scale like the reciprocal of a length unit. $\qquad\square$

Remark 15.22 (Nonaffine meshes). Proposition 9.2 together with the map (15.14) can still be used to generate a finite element $(K, \boldsymbol{P}_K, \Sigma_K)$ if the geometric mapping $\boldsymbol{T}_K$ is nonaffine. The function space $\boldsymbol{P}_K$ and the dofs in Σ_K then differ from those of the $\mathbb{N}_{k,3}$ element.   □

15.5 Other $\boldsymbol{H}$(curl) finite elements

15.5.1 Nédélec elements of the second kind

Nédélec elements of the second kind [152] offer an interesting alternative to those investigated in §15.3 (and often called Nédélec elements of the first kind) since in this case the polynomial space is $\boldsymbol{P} := \mathbb{P}_{k,d} \subsetneq \mathbb{N}_{k,d}$, $k \geq 1$. This space is optimal from the approximation viewpoint. The price to pay for this simplification is that the curl operator maps onto $\mathbb{P}_{k-1,d}$. This is not a limitation if the functions to be interpolated are curl-free.

Let K be a simplex in $\mathbb{R}^3$. The dofs are attached to the edges of K, its faces (for $k \geq 2$), and to K itself (for $k \geq 3$). The edge dofs are defined in (15.8a) as for the elements of the first kind, whereas the face dofs are moments on each face of K of the tangential component against a set of basis functions of $\mathbb{RT}_{k-2,2}$ up to a contravariant Piola transformation (instead of basis functions of $\mathbb{P}_{k-1,2}$ for the elements of the first kind), and the cell dofs are moments against a set of basis functions of $\mathbb{RT}_{k-3,3}$ (instead of basis functions of $\mathbb{P}_{k-2,3}$ for the elements of the first kind). It is shown in [152] that the triple $(K, \boldsymbol{P}, \Sigma)$ is a finite element. Hierarchical basis functions for the Nédélec element of the second kind are constructed in Ainsworth and Coyle [6], Schöberl and Zaglmayr [176].

15.5.2 Cartesian Nédélec elements

The Cartesian version of Nédélec elements have been introduced in Nédélec [151, pp. 330–333]. Let us briefly review these elements (see Exercise 15.8 for the proofs). We focus on the case $d = 3$, since two-dimensional Cartesian Nédélec elements can be built by a rotation of the two-dimensional Cartesian Raviart–Thomas elements from §14.5.2. Let $k \in \mathbb{N}$ and define

$$\mathbb{N}_{k,3}^{\square} := \mathbb{Q}_{k,k+1,k+1} \times \mathbb{Q}_{k+1,k,k+1} \times \mathbb{Q}_{k+1,k+1,k}, \qquad (15.16)$$

where the anisotropic polynomial spaces $\mathbb{Q}_{\alpha_1,\alpha_2,\alpha_3}$ are defined in §14.5.2. Since the three anisotropic spaces in (15.16) have dimension $(k+1)(k+2)^2$, we have $\dim(\mathbb{N}_{k,3}^{\square}) = 3(k+1)(k+2)^2$.

Let $K := (0,1)^3$ be the unit cube in $\mathbb{R}^3$. Let $\mathcal{F}_K$ collect the six faces of K, and let $\mathcal{E}_K$ collect the twelve edges of K. Let $\boldsymbol{T}_F$, $F \in \mathcal{F}_K$ (resp., $\boldsymbol{T}_E$, $E \in \mathcal{E}_K$) be an affine geometric mapping from $[0,1]^2$ onto F (resp., $[0,1]$ onto E). Let $\widehat{t}^{\mathrm{e}} := 1$ be the canonical basis of $\mathbb{R}$. We orient $E \in \mathcal{E}_K$ using $\boldsymbol{t}_E := \mathbb{J}_E \widehat{t}^{\mathrm{e}}$,

where $\mathbb{J}_E$ is the Jacobian matrix of $\boldsymbol{T}_E$. Let $\{\boldsymbol{t}_j^{\mathrm{f}}\}_{j\in\{1,2\}}$ be the canonical basis of $\mathbb{R}^2$. We orient $F \in \mathcal{F}_K$ by using $\boldsymbol{t}_{F,j} := \mathbb{J}_F \widehat{\boldsymbol{t}}_j^{\mathrm{f}}$ for all $j \in \{1,2\}$, where $\mathbb{J}_F$ is the Jacobian matrix of $\boldsymbol{T}_F$. We orient K by using the canonical basis $\{\boldsymbol{t}_{K,j} := \boldsymbol{e}_j\}_{j\in\{1:3\}}$ of $\mathbb{R}^3$. Let Σ be the set composed of the following linear forms:

$$\sigma_{E,m}^{\mathrm{e}}(\boldsymbol{v}) := \frac{1}{|E|} \int_E (\boldsymbol{v}\cdot\boldsymbol{t}_E)(\mu_m \circ \boldsymbol{T}_E^{-1})\,\mathrm{d}l, \qquad \forall E \in \mathcal{E}_K, \qquad (15.17\mathrm{a})$$

$$\sigma_{F,j,m}^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v}\cdot\boldsymbol{t}_{F,j})(\zeta_{j,m} \circ \boldsymbol{T}_F^{-1})\,\mathrm{d}s, \qquad \forall F \in \mathcal{F}_K, \,\forall j \in \{1,2\}, \qquad (15.17\mathrm{b})$$

$$\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) := \frac{1}{|K|} \int_K (\boldsymbol{v}\cdot\boldsymbol{t}_{K,j})\psi_{j,m}\,\mathrm{d}x, \qquad \forall j \in \{1,2,3\}, \qquad (15.17\mathrm{c})$$

where $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$ is a basis of $\mathbb{P}_{k,1}$ with $n_{\mathrm{sh}}^{\mathrm{e}} := k + 1$, $\{\zeta_{j,m}\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ is a basis of the space $\mathbb{Q}_{k,k-1}$ if $j = 1$ and $\mathbb{Q}_{k-1,k}$ if $j = 2$, with $n_{\mathrm{sh}}^{\mathrm{f}} := (k+1)k$ (if $k \geq 1$), and $\{\psi_{j,m}\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ is a basis of the space $\mathbb{Q}_{k,k-1,k-1}$ if $j = 1$, $\mathbb{Q}_{k-1,k,k-1}$ if $j = 2$, and $\mathbb{Q}_{k-1,k-1,k}$ if $j = 3$, with $n_{\mathrm{sh}}^{\mathrm{c}} := (k+1)k^2$ (if $k \geq 1$).

Proposition 15.23 (Finite element). $(K, \mathbb{N}_{k,3}^{\square}, \Sigma)$ *is a finite element.*

Cartesian Nédélec elements can be generated for all the mesh cells of an affine mesh composed of parallelotopes by using affine geometric mappings and the covariant Piola transformation. Recall however that orienting such meshes requires some care; see Theorem 10.10.

Exercises

Exercise 15.1 ($\mathbb{S}_{1,d}$). (i) Prove that for all $\boldsymbol{q} \in \mathbb{S}_{1,d}$, there is a unique skew-symmetric matrix $\mathbb{Q}$ s.t. $\boldsymbol{q}(\boldsymbol{x}) = \mathbb{Q}\boldsymbol{x}$. (ii) Propose a basis of $\mathbb{S}_{1,d}$. (iii) Show that $\boldsymbol{q} \in \mathbb{S}_{1,3}$ if and only if there is $\boldsymbol{b} \in \mathbb{R}^3$ such that $\boldsymbol{q}(\boldsymbol{x}) = \boldsymbol{b}\times\boldsymbol{x}$.

Exercise 15.2 (Cross product). (i) Prove that $(\mathbb{A}\boldsymbol{b})\times(\mathbb{A}\boldsymbol{c}) = \mathbb{A}(\boldsymbol{b}\times\boldsymbol{c})$ for every rotation matrix $\mathbb{A} \in \mathbb{R}^{3\times3}$ and all $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$. (*Hint*: use Exercise 9.5.) (ii) Show that $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c} = (\boldsymbol{a}\cdot\boldsymbol{c})\boldsymbol{b} - (\boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{a}$. (*Hint*: $(\boldsymbol{a}\times\boldsymbol{b})_k = \varepsilon_{ikj}a_ib_j$ with Levi-Civita tensor ε_{ikj}; see also the proof of Lemma 9.6.) (iii) Prove that $-(\boldsymbol{b}\times\boldsymbol{n})\times\boldsymbol{n} + (\boldsymbol{b}\cdot\boldsymbol{n})\boldsymbol{n} = \boldsymbol{b}$ if $\boldsymbol{n}$ is a unit vector.

Exercise 15.3 ($\mathbb{N}_{0,3}$). (i) Prove (15.4). (*Hint*: verify that $\boldsymbol{t}_E\cdot\nabla\lambda_q = 1$ and $\boldsymbol{t}_E\cdot\nabla\lambda_p = -1$.) (ii) Prove that $\boldsymbol{v} = \langle\boldsymbol{v}\rangle_K + \frac{1}{2}(\nabla\times\boldsymbol{v})\times(\boldsymbol{x}-\boldsymbol{c}_K)$ for all $\boldsymbol{v} \in \mathbb{N}_{0,3}$, where $\langle\boldsymbol{v}\rangle_K$ is the mean-value of $\boldsymbol{v}$ on K and $\boldsymbol{c}_K$ is the barycenter of K. (*Hint*: $\nabla\times(\boldsymbol{b}\times\boldsymbol{x}) = 2\boldsymbol{b}$ for $\boldsymbol{b} \in \mathbb{R}^3$.) (iii) Let $\boldsymbol{\theta}_E^{\mathrm{e}}$ be the shape function associated with the edge $E \in \mathcal{E}_K$. Let $F \in \mathcal{F}_K$ with unit normal $\boldsymbol{n}_{K|F}$ pointing outward K. Prove that $(\boldsymbol{\theta}_E^{\mathrm{e}})_{|F}\times\boldsymbol{n}_{K|F} = \boldsymbol{0}$ if E is not an edge of F, and $\int_F \boldsymbol{\theta}_E^{\mathrm{e}}\times\boldsymbol{n}_{K|F}\,\mathrm{d}s = $

$\iota_{E,F}(\boldsymbol{c}_E - \boldsymbol{c}_F)$ otherwise, where $\boldsymbol{c}_E$ is the barycenter of E, $\boldsymbol{c}_F$ that of F, and $\iota_{E,F} = -1$ if $\boldsymbol{n}_{K|F} \times \boldsymbol{t}_E$ points outward F, $\iota_{E,F} = 1$ otherwise. (*Hint*: use Lemma 15.15 and Exercise 14.1(ii).) (iv) Let $\mathcal{F}_E$ collect the two faces sharing $E \in \mathcal{E}_K$. Prove that $\int_K \boldsymbol{\theta}_E^e \, \mathrm{d}x = \frac{1}{2} \sum_{F \in \mathcal{F}_E} \iota_{E,F}(\boldsymbol{c}_F - \boldsymbol{c}_K) \times (\boldsymbol{c}_E - \boldsymbol{c}_F)$. (*Hint*: take the inner product with an arbitrary vector $\boldsymbol{e} \in \mathbb{R}^3$ and introduce the function $\boldsymbol{\psi}(\boldsymbol{x}) := \frac{1}{2}\boldsymbol{e} \times (\boldsymbol{x} - \boldsymbol{c}_K)$.)

Exercise 15.4 (Rotated $\mathbb{RT}_{k,2}$). Prove Lemma 15.9. (*Hint*: observe that $\boldsymbol{R}_{\frac{\pi}{2}}(\mathbb{P}_{k,2}) = \mathbb{P}_{k,2}$ and $\mathbb{S}_{k+1,2} = \boldsymbol{R}_{\frac{\pi}{2}}(\boldsymbol{x})\mathbb{P}_{k,2}^{\mathrm{H}}$.)

Exercise 15.5 (Hodge decomposition). Prove that for all $k \in \mathbb{N}$,

$$\mathbb{P}_{k+1,d} = \mathbb{N}_{k,d} \oplus \nabla \mathbb{P}_{k+2,d}^{\mathrm{H}}.$$

(*Hint*: compute $\mathbb{N}_{k,d} \cap \nabla \mathbb{P}_{k+2,d}^{\mathrm{H}}$, and use a dimension argument.)

Exercise 15.6 (Face element). We use the notation from the proof of Lemma 15.15. Let $F \in \mathcal{F}_K$. Let $\boldsymbol{T}_F : \widehat{S}^2 \to F$ be an affine bijective mapping. Let $\mathbb{J}_F$ be the Jacobian matrix of $\boldsymbol{T}_F$. Let $\boldsymbol{v} \in \mathbb{N}_{k,3}$ and let $\widehat{\boldsymbol{v}} := \mathbb{J}_F^{\mathsf{T}}(\mathbb{I}_3 - \boldsymbol{n}_F \otimes \boldsymbol{n}_F)(\boldsymbol{v} \circ \boldsymbol{T}_F)$. Show that $\widehat{\boldsymbol{v}} \in \mathbb{N}_{k,2}$. (*Hint*: compute $\widehat{\boldsymbol{y}}^{\mathsf{T}} \widehat{\boldsymbol{v}}(\widehat{\boldsymbol{y}})$ and apply the result from Exercise 14.4.)

Exercise 15.7 (Geometric mapping T_A). Let A be an affine subspace of $\mathbb{R}^d$ of dimension $l \in \{1{:}d-1\}$, $d \geq 2$. Let $\boldsymbol{a} \in A$ and let $\boldsymbol{P}_A(\boldsymbol{x}) := \boldsymbol{a} + \Pi_A(\boldsymbol{x} - \boldsymbol{a})$ be the orthogonal projection onto A, where $\Pi_A \in \mathbb{R}^{d \times d}$. (i) Let $\boldsymbol{n} \in \mathbb{R}^d$ be such that $\boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{y}) = 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in A$ (we say that $\boldsymbol{n}$ is normal to A). Show that $\Pi_A \boldsymbol{n} = \boldsymbol{0}$. Let $\boldsymbol{t} \in \mathbb{R}^d$ be such that $\boldsymbol{a} + \boldsymbol{t} \in A$ (we say that $\boldsymbol{t}$ is tangent to A). Show that $\Pi_A(\boldsymbol{t}) = \boldsymbol{t}$. (ii) Let $q \in \mathbb{P}_{k,l}$ and let $\tilde{q}(\boldsymbol{x}) := q(\boldsymbol{T}_A^{-1} \circ \boldsymbol{P}_A(\boldsymbol{x}))$. Compute $\nabla \tilde{q}$. (iii) Show that there are $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_l$ tangent vectors and $q_1, \ldots, q_l$ polynomials in $\mathbb{P}_{k,l}$ such that $\nabla \tilde{q}(\boldsymbol{x}) = \sum_{s \in \{1{:}l\}} q_s(\boldsymbol{T}_A^{-1}(\boldsymbol{x}))\boldsymbol{t}_s$ for all $\boldsymbol{x} \in A$. (iv) Let $\boldsymbol{t}$ be a tangent vector. Show that there is $\mu \in \mathbb{P}_{k,l}$ such that $\boldsymbol{t} \cdot \nabla \tilde{q}(\boldsymbol{x}) = \mu(\boldsymbol{T}_A^{-1}(\boldsymbol{x}))$.

Exercise 15.8 (Cartesian Nédélec element). (i) Propose a basis for $\mathbb{N}_{0,3}^{\square}$. (ii) Prove Proposition 15.23. (*Hint*: accept as a fact that any field $\boldsymbol{v} \in \mathbb{N}_{k,3}^{\square}$ annihiliating all the edge and faces dofs defined in (15.17) satisfies $\boldsymbol{v}_{|F} \times \boldsymbol{n}_F = \boldsymbol{0}$ for all $F \in \mathcal{F}_K$; then adapt the proof of Lemma 15.16 by using the $\mathbb{RT}_{k,3}^{\square}$ finite element defined in §14.5.2.)

Part III, Chapter 16

Local interpolation in $\boldsymbol{H}(\mathrm{div})$ and $\boldsymbol{H}(\mathrm{curl})$ (I)

In this chapter and the next one, we study the interpolation operators associated with the finite elements introduced in Chapters 14 and 15. We consider a shape-regular sequence $(\mathcal{T}_h)_{h \in \mathcal{H}}$ of affine simplicial meshes with a generation-compatible orientation (this is possible owing to Theorem 10.8). In the present chapter, we show how the degrees of freedom (dofs) attached to the faces and the edges can be extended by using the scale of the Sobolev spaces. On the way, we discover fundamental commuting properties of the interpolation operators embodied in the de Rham complex. In the next chapter, we study a different way of extending the dofs attached to the faces and the edges by requiring some integrability of the divergence or the curl.

16.1 Local interpolation in $\boldsymbol{H}(\mathrm{div})$

The goal of this section is to extend the dofs of the $\mathbb{RT}_{k,d}$ finite element introduced in Chapter 14 and to study the properties of the resulting interpolation operator.

16.1.1 Extending the dofs

Let $K \in \mathcal{T}_h$ be a simplex in $\mathbb{R}^d$ with $d \geq 2$. We generate a $\mathbb{RT}_{k,d}$ finite element in K from the $\mathbb{RT}_{k,d}$ finite element in the reference cell $\widehat{K}$ by using Proposition 14.19. Hence, the dofs in K consist of the following face dofs and cell dofs (if $k \geq 1$): For all $\boldsymbol{v} \in \mathbb{RT}_{k,d}$,

$$\sigma_{F,m}^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{\nu}_F)(\zeta_m \circ \boldsymbol{T}_{K,F}^{-1}) \, \mathrm{d}s, \qquad \forall F \in \mathcal{F}_K, \qquad (16.1a)$$

$$\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) := \frac{1}{|K|} \int_K (\boldsymbol{v} \cdot \boldsymbol{\nu}_{K,j})(\psi_m \circ \boldsymbol{T}_K^{-1}) \, \mathrm{d}x, \qquad \forall j \in \{1{:}d\}, \qquad (16.1b)$$

where $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$, $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ are bases of $\mathbb{P}_{k,d-1}$, $\mathbb{P}_{k-1,d}$ $(k \geq 1)$, respectively, $\boldsymbol{\nu}_F$ is the normal vector orienting F, $\{\boldsymbol{\nu}_{K,j} := |F_j|\boldsymbol{n}_{F_j}\}_{j\in\{1:d\}}$ are the vectors orienting K, and $\boldsymbol{T}_{K,F} : \widehat{S}^{d-1} \to F$, $\boldsymbol{T}_K : \widehat{K} \to K$ are geometric mappings. The local dofs in K are collectively denoted by $\{\sigma_{K,i}\}_{i\in\mathcal{N}}$.

We are going to extend the above dofs to the following functional space:

$$\boldsymbol{V}^{\mathrm{d}}(K) := \boldsymbol{W}^{s,p}(K), \quad sp > 1, p \in (1,\infty) \text{ or } s = 1, p = 1, \qquad (16.2)$$

recalling that $\boldsymbol{W}^{s,p}(K) := W^{s,p}(K;\mathbb{R}^d)$. The idea behind (16.2) is to invoke a trace theorem (Theorem 3.15) to give a meaning to the face dofs. Fixing the real number p in (16.2), one wants to take s as small as possible to make the space $\boldsymbol{V}^{\mathrm{d}}(K)$ as large as possible. Thus, we can assume without loss of generality that $s \leq 1$. We can also take $p = \infty$ and $s = 1$ in (16.2).

Proposition 16.1 (Extended dofs). *Let $\boldsymbol{V}^{\mathrm{d}}(K)$ be defined in (16.2). Let $\boldsymbol{V}^{\mathrm{d}}(\widehat{K})$ be defined similarly. Then the contravariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{d}}$ is in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K); \boldsymbol{V}^{\mathrm{d}}(\widehat{K}))$. Moreover, the local dofs are in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K);\mathbb{R})$ and there is c s.t. for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$\max_{i\in\mathcal{N}} |\sigma_{K,i}(\boldsymbol{v})| \leq c\, h_K^{d-1-\frac{d}{p}} \big(\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K^s|\boldsymbol{v}|_{\boldsymbol{W}^{s,p}(K)}\big). \qquad (16.3)$$

Proof. (1) Let $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$. Since the mesh is affine and $\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) := \mathbb{A}_K^{\mathrm{d}}(\boldsymbol{v}\circ\boldsymbol{T}_K)$ with $\mathbb{A}_K^{\mathrm{d}} := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}$, we can apply Lemma 11.7 to obtain

$$\|\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} \leq c\,\|\mathbb{A}_K^{\mathrm{d}}\|_{\ell^2}|\det(\mathbb{J}_K)|^{-\frac{1}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} \leq c'\,h_K^{d-1-\frac{d}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)},$$

where the second bound follows from the regularity of the mesh sequence. Moreover, letting $\gamma_K := |\det(\mathbb{J}_K)|^{-1}\|\mathbb{J}_K\|_{\ell^2}^d$ if $s < 1$ and $\gamma_K := 1$ if $s = 1$, as in Lemma 11.7, we obtain

$$|\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})|_{\boldsymbol{W}^{s,p}(\widehat{K})} \leq c\,\gamma_K^{\frac{1}{p}}\|\mathbb{A}_K^{\mathrm{d}}\|_{\ell^2}\|\mathbb{J}_K\|_{\ell^2}^s|\det(\mathbb{J}_K)|^{-\frac{1}{p}}|\boldsymbol{v}|_{\boldsymbol{W}^{s,p}(K)}$$
$$\leq c'\,h_K^{d-1-\frac{d}{p}+s}|\boldsymbol{v}|_{\boldsymbol{W}^{s,p}(K)},$$

where the second bound follows from the regularity of the mesh sequence. The above bounds show that $\boldsymbol{\psi}_K^{\mathrm{d}} \in \mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K);\boldsymbol{V}^{\mathrm{d}}(\widehat{K}))$ with

$$\|\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} + \ell_{\widehat{K}}^s|\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})|_{\boldsymbol{W}^{s,p}(\widehat{K})} \leq c\,h_K^{d-1-\frac{d}{p}}\big(\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K^s|\boldsymbol{v}|_{\boldsymbol{W}^{s,p}(K)}\big),$$

where $\ell_{\widehat{K}} := 1$ is a length scale associated with the reference cell $\widehat{K}$.
(2) Since the local dofs in K are s.t. $\sigma_{K,i} := \widehat{\sigma}_i \circ \boldsymbol{\psi}_K^{\mathrm{d}}$ for all $i \in \mathcal{N}$, we need to bound the reference dofs $\{\widehat{\sigma}_i\}_{i\in\mathcal{N}}$. Let $\widehat{\boldsymbol{v}} \in \boldsymbol{V}^{\mathrm{d}}(\widehat{K})$. If $\widehat{\sigma}_i$ is a cell dof, we have $|\widehat{\sigma}_i(\widehat{\boldsymbol{v}})| \leq \widehat{c}\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})}$, whereas if $\widehat{\sigma}_i$ is a face dof, we have $|\widehat{\sigma}_i(\widehat{\boldsymbol{v}})| \leq \widehat{c}(\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})} + \ell_{\widehat{K}}^s|\widehat{\boldsymbol{v}}|_{\boldsymbol{W}^{s,p}(\widehat{K})})$ owing to Theorem 3.15 since $sp > 1$ if $p \in (1,\infty)$

and $s = 1$ if $p = 1$. The above bound on $\boldsymbol{\psi}_K^{\mathrm{d}}$ shows that the local dofs in K are in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K); \mathbb{R})$ and that (16.3) holds true. $\qquad\square$

16.1.2 Commuting and approximation properties

In this section, we study the properties of the local Raviart–Thomas interpolation operator

$$\mathcal{I}_K^{\mathrm{d}} : \boldsymbol{V}^{\mathrm{d}}(K) \to \mathbb{R}\mathbb{T}_{k,d} \tag{16.4}$$

with $\boldsymbol{V}^{\mathrm{d}}(K)$ defined in (16.2). Recall that for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$, $\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})$ is defined as the unique polynomial in $\mathbb{R}\mathbb{T}_{k,d}$ s.t. the function $(\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v}) - \boldsymbol{v})$ annihilates all the $\mathbb{R}\mathbb{T}_{k,d}$ dofs. Let us start with an important commuting property. Let $\mathcal{I}_K^{\mathrm{b}} : V^{\mathrm{b}}(K) := L^1(K) \to \mathbb{P}_{k,d}$ be the L^2-orthogonal projection onto $\mathbb{P}_{k,d}$, i.e., $\int_K (\mathcal{I}_K^{\mathrm{b}}(\phi) - \phi)q\,\mathrm{d}x = 0$ for all $\phi \in L^1(K)$ and all $q \in \mathbb{P}_{k,d}$; see §11.5.3.

Lemma 16.2 (Commuting with $\nabla\cdot$). *The following diagram commutes:*

$$
\begin{array}{ccc}
\check{\boldsymbol{V}}^{\mathrm{d}}(K) & \xrightarrow{\ \nabla\cdot\ } & V^{\mathrm{b}}(K) \\
{\scriptstyle \mathcal{I}_K^{\mathrm{d}}}\Big\downarrow & & \Big\downarrow{\scriptstyle \mathcal{I}_K^{\mathrm{b}}} \\
\mathbb{R}\mathbb{T}_{k,d} & \xrightarrow{\ \nabla\cdot\ } & \mathbb{P}_{k,d}
\end{array}
$$

where $\check{\boldsymbol{V}}^{\mathrm{d}}(K) := \{\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K) \mid \nabla\cdot\boldsymbol{v} \in V^{\mathrm{b}}(K)\}$. In other words, we have

$$\nabla\cdot(\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})) = \mathcal{I}_K^{\mathrm{b}}(\nabla\cdot\boldsymbol{v}), \qquad \forall \boldsymbol{v} \in \check{\boldsymbol{V}}^{\mathrm{d}}(K). \tag{16.5}$$

Proof. Let $\boldsymbol{v} \in \check{\boldsymbol{V}}^{\mathrm{d}}(K)$. Since the divergence operator maps $\mathbb{R}\mathbb{T}_{k,d}$ to $\mathbb{P}_{k,d}$ by Lemma 14.9, we have $\nabla\cdot(\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})) \in \mathbb{P}_{k,d}$. Therefore, it suffices to show that $\int_K (\mathcal{I}_K^{\mathrm{b}}(\nabla\cdot\boldsymbol{v}) - \nabla\cdot(\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})))q\,\mathrm{d}x = 0$ for all $q \in \mathbb{P}_{k,d}$, and by definition of $\mathcal{I}_K^{\mathrm{b}}$, this amounts to $\int_K (\nabla\cdot\boldsymbol{\zeta})q\,\mathrm{d}x = 0$ for all $q \in \mathbb{P}_{k,d}$ where $\boldsymbol{\zeta} := \boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})$. Note that by definition $\boldsymbol{\zeta}$ annihilates all the dofs of the $\mathbb{R}\mathbb{T}_{k,d}$ element in K. Integrating by parts and decomposing the boundary integral over the faces in $\mathcal{F}_K$, we infer that

$$\int_K (\nabla\cdot\boldsymbol{\zeta})q\,\mathrm{d}x = -\int_K \boldsymbol{\zeta}\cdot\nabla q\,\mathrm{d}x + \sum_{F \in \mathcal{F}_K} \int_F \boldsymbol{\zeta}\cdot\boldsymbol{n}_{K|F}q_{|F}\,\mathrm{d}s,$$

where $\boldsymbol{n}_K$ is the outward unit normal to K. If $k \geq 1$, we use that $\{\boldsymbol{\nu}_{K,j}\}_{j\in\{1:d\}}$ is a basis of $\mathbb{R}^d$ and $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ is a basis of $\mathbb{P}_{k-1,d}$ to infer that there are real numbers $\alpha_{j,m}$ s.t. $\nabla q = \sum_{j\in\{1:d\}}\sum_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}} \alpha_{j,m}\boldsymbol{\nu}_{K,j}(\psi_m \circ \boldsymbol{T}_K^{-1})$. Recalling that $\boldsymbol{\zeta}$ annihilates all the cell dofs, we obtain

$$\int_K \boldsymbol{\zeta}\cdot\nabla q\,\mathrm{d}x = 0.$$

If $k = 0$, this equality is trivial. Let us now consider the integrals over the faces of K. For all $F \in \mathcal{F}_K$, we use that $\boldsymbol{\nu}_F = |F|\boldsymbol{n}_F$ and $\boldsymbol{n}_F = \pm\boldsymbol{n}_{K|F}$, $q_{|F} \circ \boldsymbol{T}_{K,F}^{-1} \in \mathbb{P}_{k,d-1}$ owing to Lemma 7.10, and that $\boldsymbol{\zeta}$ annihilates all the face dofs attached to F to infer that

$$\int_F \boldsymbol{\zeta}{\cdot}\boldsymbol{n}_{K|F} q_{|F} \,\mathrm{d}s = 0.$$

This concludes the proof.   □

Example 16.3 (Gradient interpolation). Let us set $s = p := 1$ in (16.2). Let $\phi \in W^{2,1}(K)$. Then $\nabla\phi \in \boldsymbol{W}^{1,1}(K) = \boldsymbol{V}^{\mathrm{d}}(K)$, and since $\nabla{\cdot}(\nabla\phi) \in L^1(K)$, we have $\nabla\phi \in \check{\boldsymbol{V}}^{\mathrm{d}}(K)$. Lemma 16.2 implies that $\nabla{\cdot}\mathcal{I}_K^{\mathrm{d}}(\nabla\phi) = \mathcal{I}_K^{\mathrm{b}}(\Delta\phi)$.   □

Theorem 16.4 (Approximation, $r \geq 1$). *Let $\mathcal{I}_K^{\mathrm{d}}$ be the $\mathbb{RT}_{k,d}$ interpolation operator in K. There is c s.t. for every integers $r \in \{1{:}k+1\}$ and $m \in \{0{:}r\}$, all $p \in [1,\infty]$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})|_{\boldsymbol{W}^{m,p}(K)} \leq c\, h_K^{r-m}|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}. \tag{16.6}$$

Moreover, for every integers $r \in \{0{:}k+1\}$ and $m \in \{0{:}r\}$, all $p \in [1,\infty]$, all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$ such that $\nabla{\cdot}\boldsymbol{v} \in W^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$, we have

$$|\nabla{\cdot}(\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v}))|_{W^{m,p}(K)} \leq c\, h_K^{r-m}|\nabla{\cdot}\boldsymbol{v}|_{W^{r,p}(K)}. \tag{16.7}$$

Proof. Let us start with (16.6). We apply Theorem 11.13. The contravariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{d}}$ is of the form (11.1) with $\mathbb{A}_K^{\mathrm{d}} := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}$, which satisfies the bound (11.12) with $\gamma := 1$. Moreover, we can take $l := 1$ in Theorem 11.13 since $\boldsymbol{W}^{1,p}(\widehat{K}) \hookrightarrow \boldsymbol{V}^{\mathrm{d}}(\widehat{K})$. Since $l \leq k+1$, we can apply the estimate (11.14), which is nothing but (16.6). Finally, to prove (16.7), we use Lemma 16.2 to infer that $\nabla{\cdot}(\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})) = \nabla{\cdot}\boldsymbol{v} - \mathcal{I}_K^{\mathrm{b}}(\nabla{\cdot}\boldsymbol{v})$, and we conclude using Lemma 11.18 ($P_K = \mathbb{P}_{k,d}$ since the mesh is affine).   □

Remark 16.5 (Error on the divergence). It is remarkable that the bound on $\nabla{\cdot}(\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v}))$ only depends on the smoothness of $\nabla{\cdot}\boldsymbol{v}$. This is a direct consequence of the commuting property stated in Lemma 16.2.   □

Theorem 16.6 (Approximation, $r > \frac{1}{p}$). *The estimate (16.6) holds true for all $r \in (\frac{1}{p},1)$, $m = 0$, all $p \in (1,\infty)$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$, and c can grow unboundedly as $r \downarrow \frac{1}{p}$.*

Proof. We first prove the following stability property:

$$\|\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)} \leq c\big(\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K^r|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}\big), \tag{16.8}$$

for all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$ (notice that $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$ since $rp > 1$). The triangle inequality, Proposition 12.5, and the regularity of the mesh sequence imply that

$$\|\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)} \leq \sum_{i\in\mathcal{N}} |\sigma_{K,i}(\boldsymbol{v})| \, \|\boldsymbol{\theta}_{K,i}\|_{\boldsymbol{L}^p(K)} \leq c \, h_K^{\frac{d}{p}+1-d} \sum_{i\in\mathcal{N}} |\sigma_{K,i}(\boldsymbol{v})|.$$

Hence, (16.8) follows from the bound (16.3) on the local dofs in K. Since $\mathbb{P}_{0,d} \subset \boldsymbol{R}\boldsymbol{T}_{k,d}$ is pointwise invariant under $\mathcal{I}_K^{\mathrm{d}}$, we infer that

$$\begin{aligned}
\|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)} &\leq \inf_{\boldsymbol{q}\in\mathbb{P}_{0,d}} \left(\|\boldsymbol{v} - \boldsymbol{q}\|_{\boldsymbol{L}^p(K)} + \|\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v} - \boldsymbol{q})\|_{\boldsymbol{L}^p(K)} \right) \\
&\leq c \inf_{\boldsymbol{q}\in\mathbb{P}_{0,d}} \left(\|\boldsymbol{v} - \boldsymbol{q}\|_{\boldsymbol{L}^p(K)} + h_K^r |\boldsymbol{v} - \boldsymbol{q}|_{\boldsymbol{W}^{r,p}(K)} \right) \\
&\leq c' h_K^r |\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)},
\end{aligned}$$

where we used (16.8), $|\boldsymbol{v} - \boldsymbol{q}|_{\boldsymbol{W}^{r,p}(K)} = |\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}$ since $\boldsymbol{q}$ is constant on K, and the fractional Poincaré–Steklov inequality (12.14) in K. $\qquad\square$

16.2 Local interpolation in $\boldsymbol{H}(\mathrm{curl})$

The goal of this section is to extend the dofs of the $\mathbb{N}_{k,d}$ finite element introduced in Chapter 15 for $d = 3$ and to study the properties of the resulting interpolation operator.

16.2.1 Extending the dofs

Let K be a simplex in $\mathbb{R}^d$ with $d = 3$. We generate a $\mathbb{N}_{k,d}$ finite element in K from the $\mathbb{N}_{k,d}$ finite element in the reference cell $\widehat{K}$ by using Proposition 15.20. Hence, the dofs in K consist of the following edge dofs, face dofs (if $k \geq 1$), and cell dofs (if $k \geq 2$): For all $\boldsymbol{v} \in \mathbb{N}_{k,d}$,

$$\sigma_{E,m}^{\mathrm{e}}(\boldsymbol{v}) := \frac{1}{|E|} \int_E (\boldsymbol{v}\cdot\boldsymbol{t}_E)(\mu_m \circ \boldsymbol{T}_{K,E}^{-1}) \, \mathrm{d}l, \quad \forall E \in \mathcal{E}_K, \tag{16.9a}$$

$$\sigma_{F,j,m}^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v}\cdot\boldsymbol{t}_{F,j})(\zeta_m \circ \boldsymbol{T}_{K,F}^{-1}) \, \mathrm{d}s, \quad \forall F \in \mathcal{F}_K, \; \forall j \in \{1,2\}, \tag{16.9b}$$

$$\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) := \frac{1}{|K|} \int_K (\boldsymbol{v}\cdot\boldsymbol{t}_{K,j})(\psi_m \circ \boldsymbol{T}_K^{-1}) \, \mathrm{d}x, \; \forall j \in \{1,2,3\}, \tag{16.9c}$$

where $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$, $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$, and $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ are bases of $\mathbb{P}_{k,1}$, $\mathbb{P}_{k-1,2}$ $(k \geq 1)$, and $\mathbb{P}_{k-2,3}$ $(k \geq 2)$, respectively, $\boldsymbol{t}_E$ is the tangent vector orienting E, $\{\boldsymbol{t}_{F,j}\}_{j\in\{1,2\}}$ the two tangent vectors orienting F, and $\{\boldsymbol{t}_{K,j}\}_{j\in\{1,2,3\}}$ the three vectors orienting K, and $\boldsymbol{T}_{K,E} : \widehat{S}^1 \to E$, $\boldsymbol{T}_{K,F} : \widehat{S}^2 \to F$, and $\boldsymbol{T}_K : \widehat{K} \to K$ are geometric mappings. The local dofs in K are collectively denoted by $\{\sigma_{K,i}\}_{i\in\mathcal{N}}$.

We are going to extend the above dofs to the following functional space:

$$\boldsymbol{V}^{\mathrm{c}}(K) := \boldsymbol{W}^{s,p}(K), \quad sp > 2, p \in (1,\infty) \text{ or } s = 2, p = 1, \tag{16.10}$$

The idea behind (16.10) is again to use a trace theorem (Theorem 3.15) to give a meaning to the edge (and face) dofs. Fixing the real number p in (16.10), we want to take s as small as possible to make the space $\boldsymbol{V}^{\mathrm{c}}(K)$ as large as possible. Thus, we can assume without loss of generality that $s \leq 1$ if $p \in (2, \infty)$ and $s \leq 2$ if $p \in [1, 2]$. We can also take $p = \infty$ and $s = 1$ in (16.10). We consider the norm $\|\cdot\|_{\boldsymbol{W}^{\tilde{s},p}(K)}$ defined as follows: If $s \in (0, 1]$ (i.e., if $p \in (2, \infty]$), we set

$$\widetilde{s} := 0, \qquad \|\boldsymbol{v}\|_{\boldsymbol{W}^{\tilde{s},p}(K)} := \|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}, \tag{16.11a}$$

whereas if $s \in (1, 2]$ (i.e., if $p \in [1, 2]$), we set

$$\widetilde{s} := 1, \qquad \|\boldsymbol{v}\|_{\boldsymbol{W}^{\tilde{s},p}(K)} := \|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K |\boldsymbol{v}|_{\boldsymbol{W}^{1,p}(K)}. \tag{16.11b}$$

Proposition 16.7 (Extended dofs). *Let $\boldsymbol{V}^{\mathrm{c}}(K)$ be defined in (16.10). Let $\boldsymbol{V}^{\mathrm{c}}(\widehat{K})$ be defined similarly. Then the covariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{c}}$ is in $\mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(K); \boldsymbol{V}^{\mathrm{c}}(\widehat{K}))$. Moreover, the local dofs are in $\mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(K); \mathbb{R})$, and there is c s.t. for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{c}}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$\max_{i \in \mathcal{N}} |\sigma_{K,i}(\boldsymbol{v})| \leq c\, h_K^{1-\frac{d}{p}} \left(\|\boldsymbol{v}\|_{\boldsymbol{W}^{\tilde{s},p}(K)} + h_K^s |\boldsymbol{v}|_{\boldsymbol{W}^{s,p}(K)} \right). \tag{16.12}$$

Proof. (1) Let $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{c}}(K)$. Since the mesh is affine and $\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) := \mathbb{A}_K^{\mathrm{c}}(\boldsymbol{v} \circ \boldsymbol{T}_K)$ with $\mathbb{A}_K^{\mathrm{c}} := \mathbb{J}_K^{\mathsf{T}}$, we can proceed as in the proof of Proposition 16.1 and invoke Lemma 11.7 to show that $\boldsymbol{\psi}_K^{\mathrm{c}} \in \mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(K); \boldsymbol{V}^{\mathrm{c}}(\widehat{K}))$ with $\|\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{W}^{\tilde{s},p}(\widehat{K})} + \ell_{\widehat{K}}^s |\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})|_{\boldsymbol{W}^{s,p}(\widehat{K})} \leq c\, h_K^{1-\frac{d}{p}} \left(\|\boldsymbol{v}\|_{\boldsymbol{W}^{\tilde{s},p}(K)} + h_K^s |\boldsymbol{v}|_{\boldsymbol{W}^{s,p}(K)} \right)$, where the norm $\|\cdot\|_{\boldsymbol{W}^{\tilde{s},p}(\widehat{K})}$ is defined similarly to $\|\cdot\|_{\boldsymbol{W}^{\tilde{s},p}(K)}$ using $\ell_{\widehat{K}} := 1$.
(2) To bound the local dofs, we invoke Theorem 3.15 and proceed again as in the proof of Proposition 16.1. $\qquad \square$

16.2.2 Commuting and approximation properties

In this section, we study the properties of the local Nédélec interpolation operator

$$\mathcal{I}_K^{\mathrm{c}} : \boldsymbol{V}^{\mathrm{c}}(K) \to \mathbb{N}_{k,d} \tag{16.13}$$

with $\boldsymbol{V}^{\mathrm{c}}(K)$ defined in (16.10). Recall that for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{c}}(K)$, $\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})$ is defined as the unique polynomial in $\mathbb{N}_{k,d}$ such that the function $(\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v}) - \boldsymbol{v})$ annihilates all the $\mathbb{N}_{k,d}$ dofs.

Lemma 16.8 (Commuting with $\nabla \times$). *The following diagram commutes:*

$$
\begin{array}{ccc}
\check{\boldsymbol{V}}^{\mathrm{c}}(K) & \xrightarrow{\nabla \times} & \boldsymbol{V}^{\mathrm{d}}(K) \\
\downarrow{\scriptstyle \mathcal{I}_K^{\mathrm{c}}} & & \downarrow{\scriptstyle \mathcal{I}_K^{\mathrm{d}}} \\
\mathbb{N}_{k,d} & \xrightarrow{\nabla \times} & \mathbb{RT}_{k,d}
\end{array}
$$

where $\check{\boldsymbol{V}}^{\mathrm{c}}(K) := \{\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{c}}(K) \mid \nabla{\times}\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)\}$. *In other words, we have*

$$\nabla{\times}(\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})) = \mathcal{I}_K^{\mathrm{d}}(\nabla{\times}\boldsymbol{v}), \qquad \forall \boldsymbol{v} \in \check{\boldsymbol{V}}^{\mathrm{c}}(K). \tag{16.14}$$

Proof. Let us first observe that $\nabla{\times}\mathbb{N}_{k,d} \subset \mathbb{P}_{k,d} \subset \mathbb{RT}_{k,d}$ (see Lemma 15.10), which implies that $\nabla{\times}$ maps $\mathbb{N}_{k,d}$ to $\mathbb{RT}_{k,d}$. Note also that $\nabla{\times}$ maps $\check{\boldsymbol{V}}^{\mathrm{c}}(K)$ to $\boldsymbol{V}^{\mathrm{d}}(K)$ by definition of these spaces. Let $\boldsymbol{v} \in \check{\boldsymbol{V}}^{\mathrm{c}}(K)$. The proof of (16.14) consists of showing that $\boldsymbol{\delta} := \nabla{\times}(\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})) - \mathcal{I}_K^{\mathrm{d}}(\nabla{\times}\boldsymbol{v}) \in \mathbb{RT}_{k,d}$ annihilates all the dofs of the $\mathbb{RT}_{k,d}$ finite element in K. Let us set $\boldsymbol{\zeta} := \boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})$ and $\boldsymbol{\xi} := \nabla{\times}\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\nabla{\times}\boldsymbol{v})$, so that we have

$$\boldsymbol{\delta} = \nabla{\times}(\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})) - \nabla{\times}\boldsymbol{v} + \nabla{\times}\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\nabla{\times}\boldsymbol{v}) = \boldsymbol{\xi} - \nabla{\times}\boldsymbol{\zeta}.$$

(1) Let us consider first the dofs attached to K for $k \geq 1$. Let $\boldsymbol{e}$ be a unit vector in $\mathbb{R}^d$ and let $\psi \in \mathbb{P}_{k-1,d}$. We want to show that $\int_K \boldsymbol{\delta}{\cdot}\boldsymbol{e}\psi \,\mathrm{d}x = 0$. Since $\boldsymbol{\xi}$ annihilates all the cell dofs of the $\mathbb{RT}_{k,d}$ element, we have $\int_K \boldsymbol{\xi}{\cdot}\boldsymbol{e}\psi \,\mathrm{d}x = 0$, so that $\int_K \boldsymbol{\delta}{\cdot}\boldsymbol{e}\psi \,\mathrm{d}x = - \int_K (\nabla{\times}\boldsymbol{\zeta}){\cdot}\boldsymbol{e}\psi \,\mathrm{d}x$. Using the integration by parts formula (4.8a), we have

$$\int_K (\nabla{\times}\boldsymbol{\zeta}){\cdot}\boldsymbol{e}\psi \,\mathrm{d}x = \int_K \boldsymbol{\zeta}{\cdot}\nabla{\times}(\boldsymbol{e}\psi) - \sum_{F \in \mathcal{F}_K} \int_F \boldsymbol{\zeta}{\cdot}(\boldsymbol{n}_{K|F}{\times}\boldsymbol{e})\psi \,\mathrm{d}s.$$

If $k \geq 2$, we use that $\boldsymbol{\zeta}$ annihilates the cell dofs of the $\mathbb{N}_{k,d}$ element to infer that $\int_K \boldsymbol{\zeta}{\cdot}\nabla{\times}(\boldsymbol{e}\psi) = 0$. If $k = 1$, this equality is obvious. Moreover, since $\boldsymbol{\zeta}$ also annihilates the face dofs of the $\mathbb{N}_{k,d}$ element and since the vector $(\boldsymbol{n}_{K|F}{\times}\boldsymbol{e})$ is tangent to F, we infer that $\int_F \boldsymbol{\zeta}{\cdot}(\boldsymbol{n}_{K|F}{\times}\boldsymbol{e})\psi \,\mathrm{d}s = 0$ for all $F \in \mathcal{F}_K$. In conclusion, $\int_K (\nabla{\times}\boldsymbol{\zeta}){\cdot}\boldsymbol{e}\psi \,\mathrm{d}x = 0$, so that $\int_K \boldsymbol{\delta}{\cdot}\boldsymbol{e}\psi \,\mathrm{d}x = 0$.
(2) Let us now consider the dofs attached to a face $F \in \mathcal{F}_K$. We want to show that $\int_F \boldsymbol{\delta}{\cdot}\boldsymbol{n}_F\psi \,\mathrm{d}s = 0$ for all $\psi \in \mathbb{P}_{k,d}$. This is a sufficient condition to annihilate the $\mathbb{RT}_{k,d}$ dofs attached to F, since for all $q \in \mathbb{P}_{k,d-1}$, there exists $\psi \in \mathbb{P}_{k,d}$ such that $\psi_{|F} = q \circ \boldsymbol{T}_{K,F}^{-1}$ owing to Lemma 7.10. Since $\boldsymbol{\xi}$ annihilates the face dofs of the $\mathbb{RT}_{k,d}$ element, we have $\int_F \boldsymbol{\delta}{\cdot}\boldsymbol{n}_F\psi \,\mathrm{d}s = - \int_F (\nabla{\times}\boldsymbol{\zeta}){\cdot}\boldsymbol{n}_F\psi \,\mathrm{d}s$. Moreover, since $\nabla{\times}(\psi\boldsymbol{\zeta}) = \nabla\psi{\times}\boldsymbol{\zeta} + \psi\nabla{\times}\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}$ annihilates the face dofs of the $\mathbb{N}_{k,d}$ element, we infer that

$$\int_F (\nabla{\times}\boldsymbol{\zeta}){\cdot}\boldsymbol{n}_F\psi \,\mathrm{d}s = \int_F \nabla{\times}(\psi\boldsymbol{\zeta}){\cdot}\boldsymbol{n}_F \,\mathrm{d}s - \int_F \boldsymbol{\zeta}{\cdot}(\boldsymbol{n}_F{\times}\nabla\psi) \,\mathrm{d}s$$

$$= \int_F \nabla{\times}(\psi\boldsymbol{\zeta}){\cdot}\boldsymbol{n}_F \,\mathrm{d}s = \int_{\partial F} (\psi\boldsymbol{\zeta}){\cdot}\boldsymbol{\tau}_F \,\mathrm{d}l = \sum_{E \in \mathcal{E}_F} \int_E \boldsymbol{\zeta}{\cdot}(\boldsymbol{\tau}_{F|E}\psi) \,\mathrm{d}l,$$

where we used the Kelvin–Stokes formula (16.15) with $\boldsymbol{\tau}_F$ being the unit vector tangent to ∂F whose orientation is compatible with that of $\boldsymbol{n}_F$, and where we decomposed the integral over ∂F into the integrals over the edges composing F. Since $\boldsymbol{\tau}_{F|E}$ is tangent to the edge E and $\boldsymbol{\zeta}$ annihilates the edge

dofs of the $\mathbb{N}_{k,d}$ element, we obtain $\int_F (\nabla\times\boldsymbol{\zeta})\cdot\boldsymbol{n}_F\psi\,\mathrm{d}s = 0$. Hence, we have $\int_F \boldsymbol{\delta}\cdot\boldsymbol{n}_F\psi\,\mathrm{d}s = 0$, and this concludes the proof. $\square$

Lemma 16.9 (Kelvin–Stokes). *Let K be a simplex in $\mathbb{R}^3$. Let F be a face of K with orientation defined by $\boldsymbol{n}_F$ and with boundary ∂F. Let $\boldsymbol{\tau}_F$ be the unit vector tangent to ∂F whose orientation is compatible with that of $\boldsymbol{n}_F$, i.e., for all $\boldsymbol{x} \in \partial F$, the vector $\boldsymbol{\tau}_F(\boldsymbol{x})\times\boldsymbol{n}_F(\boldsymbol{x})$ points outside of F. The following holds true for all $\boldsymbol{w} \in \check{\boldsymbol{V}}^{\mathrm{c}}(K)$:*

$$\int_F (\nabla\times\boldsymbol{w})\cdot\boldsymbol{n}_F\,\mathrm{d}s = \int_{\partial F} \boldsymbol{w}\cdot\boldsymbol{\tau}_F\,\mathrm{d}l. \qquad (16.15)$$

Theorem 16.10 (Approximation, $r \geq 1$ or $r \geq 2$). *Let $\mathcal{I}_K^{\mathrm{c}}$ be the local $\mathbb{N}_{k,d}$ interpolation operator. There is c s.t. the following holds true:*
(i) If $p \in (2,\infty]$, then we have

$$|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})|_{\boldsymbol{W}^{m,p}(K)} \leq c\,h_K^{r-m}|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}, \qquad (16.16)$$

for every integers $r \in \{1{:}k{+}1\}$ and $m \in \{0{:}r\}$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.
(ii) If $p \in [1,2]$, the estimate (16.16) holds true if $k \geq 1$ for every integers $r \in \{2{:}k+1\}$ and $m \in \{0{:}r\}$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$, whereas if $k = 0$, we have

$$|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})|_{\boldsymbol{W}^{m,p}(K)} \leq c\left(h_K^{1-m}|\boldsymbol{v}|_{\boldsymbol{W}^{1,p}(K)} + h_K^{2-m}|\boldsymbol{v}|_{\boldsymbol{W}^{2,p}(K)}\right), \qquad (16.17)$$

for all $m \in \{0,1\}$, all $\boldsymbol{v} \in \boldsymbol{W}^{2,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.
(iii) Finally, we have

$$|\nabla\times(\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v}))|_{\boldsymbol{W}^{m,p}(K)} \leq c\,h_K^{r-m}|\nabla\times\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}, \qquad (16.18)$$

for every integers $r \in \{1{:}k{+}1\}$ and $m \in \{0{:}r\}$, all $p \in [1,\infty]$, all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{c}}(K)$ such that $\nabla\times\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. Let us start with (16.16) and (16.17). We apply Theorem 11.13. The covariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{c}}$ is of the form (11.1) with $\mathbb{A}_K^{\mathrm{c}} := \mathbb{J}_K^{\mathsf{T}}$, which satisfies the bound (11.12) with $\gamma := 1$. Moreover, we can take $l := 2$ if $p \in [1,2]$ and $l := 1$ if $p \in (2,\infty]$ since in both cases we have $\boldsymbol{W}^{l,p}(\widehat{K}) \hookrightarrow \boldsymbol{V}^{\mathrm{c}}(\widehat{K})$. If $p \in (2,\infty]$ or if $p \in [1,2]$ and $k \geq 1$, we have $l \leq k+1$, so that we can apply the estimate (11.14), which is nothing but (16.16). In the case where $p \in [1,2]$ and $k = 0$, we apply (11.15), which is nothing but (16.17). Finally, to prove (16.18), we use Lemma 16.8 to infer that $\nabla\times(\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})) = \nabla\times\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\nabla\times\boldsymbol{v})$, and we conclude using Theorem 16.4. $\square$

Remark 16.11 (Error on the curl). It is remarkable that the bound on $\nabla\times(\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v}))$ only depends on the smoothness of $\nabla\times\boldsymbol{v}$. This is a direct consequence of the commuting property stated in Lemma 16.8. $\square$

Theorem 16.12 (Approximation, $r > \frac{2}{p}$). *There is c, unbounded as $r \downarrow \frac{2}{p}$, such that:*
(i) *If $p \in (2, \infty)$, the estimate (16.16) holds true for all $r \in (\frac{2}{p}, 1)$, $m = 0$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.*
(ii) *If $p \in (1, 2]$, the estimate (16.16) holds true if $k \geq 1$ for all $r \in (\frac{2}{p}, 2)$, all $m \in \{0, 1\}$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$, whereas if $k = 0$, we have*

$$|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})|_{\boldsymbol{W}^{m,p}(K)} \leq c\left(h_K^{1-m}|\boldsymbol{v}|_{\boldsymbol{W}^{1,p}(K)} + h_K^{r-m}|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}\right), \quad (16.19)$$

for all $r \in (\frac{2}{p}, 2)$, all $m \in \{0, 1\}$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. Let us set $l := 2$ if $p \in (1, 2]$ and $l := 1$ if $p \in (2, \infty)$. Let $r \in (\frac{2}{p}, l)$, so that $\boldsymbol{W}^{r,p}(K) \hookrightarrow \boldsymbol{V}^{\mathrm{c}}(K)$. Combining the bound from Proposition 12.5, the regularity of the mesh sequence, and the estimate (16.12) on the local dofs, we infer the stability estimate

$$\|\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)} \leq c\left(\|\boldsymbol{v}\|_{\boldsymbol{W}^{\tilde{r},p}(K)} + h_K^r|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}\right),$$

with $\tilde{r} := 0$ if $r \in (0, 1]$ and $\tilde{r} := 1$ if $r \in (1, 2)$.
(i) Assume that $p \in (2, \infty)$. Then $r < 1$ so that $\|\boldsymbol{v}\|_{\boldsymbol{W}^{\tilde{r},p}(K)} = \|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}$. Since $\mathbb{P}_{0,d} \subset \mathbb{N}_{k,d}$, we infer that

$$\|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)} \leq c \inf_{\boldsymbol{q} \in \mathbb{P}_{0,d}} \left(\|\boldsymbol{v} - \boldsymbol{q}\|_{\boldsymbol{L}^p(K)} + |\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v} - \boldsymbol{q})|_{\boldsymbol{L}^p(K)}\right)$$
$$\leq c \left(\inf_{\boldsymbol{q} \in \mathbb{P}_{0,d}} \|\boldsymbol{v} - \boldsymbol{q}\|_{\boldsymbol{L}^p(K)} + h_K^r|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}\right),$$

where we used that $|\boldsymbol{v} - \boldsymbol{q}|_{\boldsymbol{W}^{r,p}(K)} = |\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}$. The estimate (16.16) with $m = 0$ follows from the fractional Poincaré–Steklov inequality (see Lemma 12.12).
(ii) Assume that $p \in (1, 2)$. Then $r \in (1, 2)$ so that $\|\boldsymbol{v}\|_{\boldsymbol{W}^{\tilde{r},p}(K)} = \|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K|\boldsymbol{v}|_{\boldsymbol{W}^{1,p}(K)}$. Let $n := \min(1, k)$. Since $n \leq k$ and $n \leq 1 < r$, proceeding as above, we infer that

$$\|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(K)} \leq c \left(\inf_{\boldsymbol{q} \in \mathbb{P}_{n,d}} \phi_K(\boldsymbol{v} - \boldsymbol{q}) + h_K^r|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}\right),$$

with $\phi_K(\boldsymbol{v} - \boldsymbol{q}) := \|\boldsymbol{v} - \boldsymbol{q}\|_{\boldsymbol{L}^p(K)} + h_K|\boldsymbol{v} - \boldsymbol{q}|_{\boldsymbol{W}^{1,p}(K)}$ Using the inverse inequality $|\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v} - \boldsymbol{q})|_{\boldsymbol{W}^{1,p}(K)} \leq c h_K^{-1}\|\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v} - \boldsymbol{q})\|_{\boldsymbol{L}^p(K)}$ (see Lemma 12.1) and proceeding again as above, we infer that

$$|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})|_{\boldsymbol{W}^{1,p}(K)} \leq c \left(\inf_{\boldsymbol{q} \in \mathbb{P}_{n,d}} h_K^{-1}\phi_K(\boldsymbol{v} - \boldsymbol{q}) + h_K^{r-1}|\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}\right).$$

If $k \geq 1$, we have $n = 1$, and the estimate (16.16) follows from Corollary 12.13 for all $m \in \{0, 1\}$, whereas if $k = 0$, we have $n = 0$, and the estimate (16.19) for all $m \in \{0, 1\}$ follows from the fractional Poincaré–Steklov inequality. $\square$

16.3 The de Rham complex

In this section, we introduce the notion of de Rham complex, and we reinterpret the previous commuting properties from Lemma 16.2 and Lemma 16.8 in this context. We assume that $d = 3$; see Remark 16.17 below to adapt the material when $d = 2$.

Definition 16.13 (Exact cochain complex). *Let $I \geq 2$ be an integer. A cochain complex is composed of a sequence of Banach spaces $(V_i)_{i \in \{0:I\}}$ and a sequence of linear operators $(d_i)_{i \in \{1:I\}}$ between these spaces*

$$V_0 \xrightarrow{d_1} V_1 \dots V_{i-1} \xrightarrow{d_i} V_i \xrightarrow{d_{i+1}} V_{i+1} \dots V_{I-1} \xrightarrow{d_I} V_I, \qquad (16.20)$$

such that for all $i \in \{1:I\}$, $\mathrm{im}(d_i)$ is closed in V_i and if $i < I$, $\mathrm{im}(d_i) \subseteq \ker(d_{i+1})$ (this means that $d_{i+1} \circ d_i = 0$). The cochain complex is said to be exact if $\mathrm{im}(d_i) = \ker(d_{i+1})$ for all $i \in \{1:I-1\}$.

The exactness of a cochain complex is useful since it gives a simple way of knowing whether an element $v_i \in V_i$ is in $\mathrm{im}(d_i)$ by checking whether $d_{i+1}(v_i) = 0$. In this book, we focus on one fundamental example of cochain complex, namely the de Rham complex which involves the gradient, curl, and divergence operators.

Proposition 16.14 (de Rham complex). *Let D be a Lipschitz domain in $\mathbb{R}^3$. Assume that D is simply connected and that ∂D is connected. The following cochain complex, called* de Rham complex, *is exact:*

$$\mathbb{R} \xrightarrow{i} H^1(D) \xrightarrow{\nabla} \boldsymbol{H}(\mathit{curl}; D) \xrightarrow{\nabla \times} \boldsymbol{H}(\mathrm{div}; D) \xrightarrow{\nabla \cdot} L^2(D) \xrightarrow{o} \{0\}, \quad (16.21)$$

where i maps a real number to a constant function and o is the zero map.

Proof. That $\ker(\nabla) = \mathbb{R}$, $\ker(\nabla \times) = \mathrm{im}(\nabla)$, and $\ker(\nabla \cdot) = \mathrm{im}(\nabla \times)$ are well-known facts from calculus since D is, respectively, connected, simply connected, and has a connected boundary. Finally, that $\mathrm{im}(\nabla \cdot) = L^2(D)$ is proved in Lemma 51.2. $\qquad \square$

Proposition 16.15 (Discrete de Rham complex). *Let $\kappa \in \mathbb{N}$. The following cochain complex, called* discrete de Rham complex, *is exact:*

$$\mathbb{R} \xrightarrow{i} \mathbb{P}_{\kappa+1,3} \xrightarrow{\nabla} \boldsymbol{\mathbb{N}}_{\kappa,3} \xrightarrow{\nabla \times} \boldsymbol{\mathbb{R}\mathbb{T}}_{\kappa,3} \xrightarrow{\nabla \cdot} \mathbb{P}_{\kappa,3} \xrightarrow{o} \{0\}. \qquad (16.22)$$

Proof. $\ker(\nabla) = \mathrm{im}(i)$ is obvious, and $\ker(\nabla \times) = \mathrm{im}(\nabla)$ follows from Lemma 15.10. For $\ker(\nabla \cdot) = \mathrm{im}(\nabla \times)$, $\ker(o) = \mathrm{im}(\nabla \cdot)$; see Exercise 16.6. $\qquad \square$

We now connect the above two de Rham complexes by means of interpolation operators. Let K be a simplex in $\mathbb{R}^d$, $d = 3$. Let $p \in [1, \infty)$ and let s be such that $sp > 3$ if $p > 1$ or $s = 3$ if $p = 1$. Recall the following functional spaces where $V^{\mathrm{b}}(K) := L^1(K)$:

$$\check{V}^{\mathrm{g}}(K) := \{f \in W^{s,p}(K) \mid \nabla f \in \boldsymbol{W}^{s-\frac{1}{p},p}(K)\}, \tag{16.23a}$$

$$\check{\boldsymbol{V}}^{\mathrm{c}}(K) := \{\boldsymbol{g} \in \boldsymbol{W}^{s-\frac{1}{p},p}(K) \mid \nabla{\times}\boldsymbol{g} \in \boldsymbol{W}^{s-\frac{2}{p},p}(K)\}, \tag{16.23b}$$

$$\check{\boldsymbol{V}}^{\mathrm{d}}(K) := \{\boldsymbol{g} \in \boldsymbol{W}^{s-\frac{2}{p},p}(K) \mid \nabla{\cdot}\boldsymbol{g} \in V^{\mathrm{b}}(K)\}. \tag{16.23c}$$

Lemma 16.16 (Commuting diagrams). *Let* $\kappa \in \mathbb{N}$*. Let* K *be a simplex in* $\mathbb{R}^d$*,* $d = 3$*. Let* $\mathcal{I}^{\mathrm{g}}_{\kappa+1,K}$ *be the interpolation operator associated with the canonical hybrid element of degree* $(\kappa + 1)$ *defined in* §7.6*. Let* $\mathcal{I}^{\mathrm{c}}_{\kappa,K}$ *be the* $\mathbb{N}_{\kappa,3}$ *interpolation operator, let* $\mathcal{I}^{\mathrm{d}}_{\kappa,K}$ *be the* $\mathbb{R}\mathbb{T}_{\kappa,d}$ *interpolation operator, and let* $\mathcal{I}^{\mathrm{b}}_{\kappa,K}$ *be the* L^2*-orthogonal projection onto* $\mathbb{P}_{\kappa,d}$*. The following diagrams commute:*

$$
\begin{array}{ccccccc}
\check{V}^{\mathrm{g}}(K) & \xrightarrow{\ \nabla\ } & \check{\boldsymbol{V}}^{\mathrm{c}}(K) & \xrightarrow{\ \nabla\times\ } & \check{\boldsymbol{V}}^{\mathrm{d}}(K) & \xrightarrow{\ \nabla\cdot\ } & V^{\mathrm{b}}(K) \\
\Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{g}}_{\kappa+1,K}} & & \Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{c}}_{\kappa,K}} & & \Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{d}}_{\kappa,K}} & & \Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{b}}_{\kappa,K}} \\
\mathbb{P}_{\kappa+1,d} & \xrightarrow{\ \nabla\ } & \mathbb{N}_{\kappa,d} & \xrightarrow{\ \nabla\times\ } & \mathbb{R}\mathbb{T}_{\kappa,d} & \xrightarrow{\ \nabla\cdot\ } & \mathbb{P}_{\kappa,d}
\end{array}
$$

Proof. Recalling Lemma 16.2 and Lemma 16.8, it only remains to prove that the leftmost diagram commutes. This is done in Exercise 16.3. $\qquad\square$

Remark 16.17 (2D). There are two possible versions of Lemma 16.16 if $d = 2$, using either the operator $\nabla{\times}\boldsymbol{f} := \partial_1 f_2 - \partial_2 f_1$ or the operator $\nabla^{\perp} f := (-\partial_2 f, \partial_1 f)^{\mathsf{T}}$. One can show that the following two diagrams commute:

$$
\begin{array}{ccccc}
\check{V}^{\mathrm{g}}(K) & \xrightarrow{\ \nabla^{\perp}\ } & \check{\boldsymbol{V}}^{\mathrm{d}}(K) & \xrightarrow{\ \nabla\cdot\ } & V^{\mathrm{b}}(K) \\
\Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{g}}_{\kappa+1,K}} & & \Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{d}}_{\kappa,K}} & & \Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{b}}_{\kappa,K}} \\
\mathbb{P}_{\kappa+1,d} & \xrightarrow{\ \nabla^{\perp}\ } & \mathbb{R}\mathbb{T}_{\kappa,d} & \xrightarrow{\ \nabla\cdot\ } & \mathbb{P}_{\kappa,d}
\end{array}
\qquad
\begin{array}{ccccc}
\check{V}^{\mathrm{g}}(K) & \xrightarrow{\ \nabla\ } & \check{\boldsymbol{V}}^{\mathrm{c}}(K) & \xrightarrow{\ \nabla\times\ } & V^{\mathrm{b}}(K) \\
\Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{g}}_{\kappa+1,K}} & & \Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{c}}_{\kappa,K}} & & \Big\downarrow{\scriptstyle\mathcal{I}^{\mathrm{b}}_{\kappa,K}} \\
\mathbb{P}_{\kappa+1,d} & \xrightarrow{\ \nabla\ } & \mathbb{N}_{\kappa,d} & \xrightarrow{\ \nabla\times\ } & \mathbb{P}_{\kappa,d}
\end{array}
$$

with $\check{V}^{\mathrm{g}}(K)$ defined in (16.23a) with $sp > 2$ if $p \in (1,\infty)$ or $s = 2$ if $p = 1$, $\check{\boldsymbol{V}}^{\mathrm{c}}(K) := \{\boldsymbol{g} \in \boldsymbol{W}^{s-\frac{1}{p},p}(K) \mid \nabla{\times}\boldsymbol{g} \in L^1(K)\}$, and $\check{\boldsymbol{V}}^{\mathrm{d}}(K) := \boldsymbol{R}_{\frac{\pi}{2}}(\check{\boldsymbol{V}}^{\mathrm{c}}(K)) = \{\boldsymbol{g} \in \boldsymbol{W}^{s-\frac{1}{p},p}(K) \mid \nabla{\cdot}\boldsymbol{g} \in V^{\mathrm{b}}(K)\}$, where $\boldsymbol{R}_{\frac{\pi}{2}}$ is the rotation matrix of angle $\frac{\pi}{2}$ in $\mathbb{R}^2$. $\qquad\square$

Remark 16.18 (Cuboids). The commuting diagrams from Lemma 16.16 can be adapted when K is a cuboid by using the Cartesian Raviart–Thomas and Nédélec spaces from §14.5.2 and §15.5.2. $\qquad\square$

Remark 16.19 (Literature). The construction and analysis of finite elements leading to discrete de Rham complexes has witnessed significant progresses since the early 2000s and has lead to the notion of finite element exterior calculus; see Arnold et al. [11, 12]. Regularity estimates in Sobolev (and other) norms for right inverse operators of the gradient, curl, and divergence can be found in Costabel and McIntosh [83]. $\qquad\square$

Exercises

Exercise 16.1 ($\check{V}^{\mathrm{d}}(K)$). Show that $\boldsymbol{V}^{\mathrm{d}}(K)$ defined in (16.2) can be used in the commuting diagram of Lemma 16.2 after replacing $L^1(K)$ by $W^{s-1,p}(K)$. (*Hint*: use Theorem 3.19.)

Exercise 16.2 ($\mathcal{I}_K^{\mathrm{d}}$). Prove that the estimate (16.6) holds true for all $r \in [1, k+1]$, $r \notin \mathbb{N}$, every integer $m \in \{0:\lfloor r \rfloor\}$, and all $p \in [1, \infty)$. Prove that (16.7) holds true for all $r \in [0, k+1]$, $r \notin \mathbb{N}$, every integer $m \in \{0:\lfloor r \rfloor\}$, and all $p \in [1, \infty)$. (*Hint*: combine $W^{m,p}$-stability with Corollary 12.13.)

Exercise 16.3 (de Rham).   Prove that the leftmost diagram in Lemma 16.16 commutes. (*Hint*: verify that $\nabla \mathcal{I}_K^{\mathrm{g}}(v) - \mathcal{I}_K^{\mathrm{c}}(\nabla v)$ annihilates all dofs in $\boldsymbol{\mathsf{N}}_{k,d}$.)

Exercise 16.4 (Poincaré operators). Assume that K is star-shaped with respect to a point $\boldsymbol{a} \in K$. Let f and $\boldsymbol{g}$ be smooth functions on K. Define $P^{\mathrm{g}}(\boldsymbol{g})(\boldsymbol{x}) := (\boldsymbol{x} - \boldsymbol{a}) \cdot \int_0^1 \boldsymbol{g}(\boldsymbol{a} + t(\boldsymbol{x} - \boldsymbol{a})) \, \mathrm{d}t$, $\boldsymbol{P}^{\mathrm{c}}(\boldsymbol{g})(\boldsymbol{x}) := -(\boldsymbol{x} - \boldsymbol{a}) \times \int_0^1 \boldsymbol{g}(\boldsymbol{a} + t(\boldsymbol{x} - \boldsymbol{a})) \, \mathrm{d}t$ (if $d = 3$), and $\boldsymbol{P}^{\mathrm{d}}(f)(\boldsymbol{x}) := (\boldsymbol{x} - \boldsymbol{a}) \int_0^1 f(\boldsymbol{a} + t(\boldsymbol{x} - \boldsymbol{a})) t^{d-1} \, \mathrm{d}t$. Verify that (i) $\nabla P^{\mathrm{g}}(\boldsymbol{g}) = \boldsymbol{g}$ if $\partial_i g_j = \partial_j g_i$ for all $i, j \in \{1:d\}$; (ii) $\nabla \times \boldsymbol{P}^{\mathrm{c}}(\boldsymbol{g}) = \boldsymbol{g}$ if $\nabla \cdot \boldsymbol{g} = 0$; (iii) $\nabla \cdot \boldsymbol{P}^{\mathrm{d}}(f) = f$.

Exercise 16.5 (Koszul operator). (i) Let $\boldsymbol{v} \in \mathbb{P}_{k,d}^{\mathrm{H}}$ with $d = 3$. Prove that $\nabla(\boldsymbol{x} \cdot \boldsymbol{v}) - \boldsymbol{x} \times (\nabla \times \boldsymbol{v}) = (k+1)\boldsymbol{v}$ and $-\nabla \times (\boldsymbol{x} \times \boldsymbol{v}) + \boldsymbol{x}(\nabla \cdot \boldsymbol{v}) = (k+2)\boldsymbol{v}$. (*Hint*: use Euler's identity from Lemma 14.3.) (ii) Prove that $\mathbb{P}_{k,d} = \nabla \mathbb{P}_{k+1,d} \oplus (\boldsymbol{x} \times \mathbb{P}_{k-1,d}) = \nabla \times \mathbb{P}_{k+1,d} \oplus (\boldsymbol{x} \mathbb{P}_{k-1,d})$. (*Hint*: establish first these identities for homogeneous polynomials.) *Note*: defining the Koszul operators $\kappa^{\mathrm{g}}(\boldsymbol{v}) := \boldsymbol{x} \cdot \boldsymbol{v}$ and $\boldsymbol{\kappa}^{\mathrm{c}}(\boldsymbol{v}) := -\boldsymbol{x} \times \boldsymbol{v}$ for vector fields and $\boldsymbol{\kappa}^{\mathrm{d}}(v) := \boldsymbol{x} v$ for scalar fields, one has $\kappa^{\mathrm{g}}(\nabla q) = kq$ (Euler's identity) and $\nabla \cdot (\boldsymbol{\kappa}^{\mathrm{d}}(q)) = (k+d)q$ for all $q \in \mathbb{P}_{k,d}^{\mathrm{H}}$, and $\nabla(\kappa^{\mathrm{g}}(\boldsymbol{q})) + \boldsymbol{\kappa}^{\mathrm{c}}(\nabla \times \boldsymbol{q}) = (k+1)\boldsymbol{q}$ and $\nabla \times (\boldsymbol{\kappa}^{\mathrm{c}}(\boldsymbol{q})) + \boldsymbol{\kappa}^{\mathrm{d}}(\nabla \cdot \boldsymbol{q}) = (k+2)\boldsymbol{q}$ for all $\boldsymbol{q} \in \mathbb{P}_{k,d}^{\mathrm{H}}$; see [11, Sec. 3.2].

Exercise 16.6 ($\nabla \cdot \boldsymbol{\mathsf{RT}}_{k,d}$ and $\nabla \times \boldsymbol{\mathsf{N}}_{k,3}$). (i) Prove that $\nabla \cdot \boldsymbol{\mathsf{RT}}_{k,d} = \mathbb{P}_{k,d}$. (*Hint*: prove that $\nabla \cdot : \boldsymbol{x} \mathbb{P}_{k,d} \to \mathbb{P}_{k,d}$ is injective using Lemma 14.3.) (ii) Let us set $\boldsymbol{\mathsf{RT}}_{k,d}^{\mathrm{div}=0} := \{\boldsymbol{v} \in \boldsymbol{\mathsf{RT}}_{k,d} \mid \nabla \cdot \boldsymbol{v} = 0\}$. Determine $\dim(\boldsymbol{\mathsf{RT}}_{k,d}^{\mathrm{div}=0})$ for $d \in \{2, 3\}$. (iii) Show that $\boldsymbol{\mathsf{RT}}_{k,3}^{\mathrm{div}=0} = \nabla \times \mathbb{P}_{k+1,3}$. (*Hint*: use Lemma 14.9.) (iv) Prove that $\boldsymbol{\mathsf{RT}}_{k,3}^{\mathrm{div}=0} = \nabla \times \boldsymbol{\mathsf{N}}_{k,3}$. (*Hint*: use the rank nullity theorem.)

Exercise 16.7 ($\nabla \mathbb{P}_{k+1,d}$ and $\nabla \times \mathbb{P}_{k+1,3}$). Let $k \in \mathbb{N}$. (i) Set $\mathbb{P}_{k,d}^{\mathrm{c}} := \nabla \mathbb{P}_{k+1,d}$. Show that $\dim(\mathbb{P}_{k,d}^{\mathrm{c}}) = \binom{k+d+1}{d} - 1$. (ii) Assume $d = 3$. Set $\mathbb{P}_{k,3}^{\mathrm{d}} := \nabla \times \mathbb{P}_{k+1,3}$. Show that $\dim(\mathbb{P}_{k,3}^{\mathrm{d}}) = 3\binom{k+4}{3} - \binom{k+5}{3} + 1 = 3\binom{k+3}{3} - \binom{k+2}{3}$ (with the convention that $\binom{2}{3} = 0$). (*Hint*: use the exact cochain complex $\mathbb{P}_{0,d} \xrightarrow{i} \mathbb{P}_{k+2,d} \xrightarrow{\nabla} \mathbb{P}_{k+1,d} \xrightarrow{\nabla \times} \mathbb{P}_{k,d} \xrightarrow{\nabla \cdot} \mathbb{P}_{k-1,d} \xrightarrow{o} \{0\}$.)

Part III, Chapter 17

Local interpolation in $\boldsymbol{H}(\mathrm{div})$ and $\boldsymbol{H}(\mathrm{curl})$ (II)

In this chapter, we continue our investigation of the interpolation operators associated with $\boldsymbol{H}(\mathrm{div})$ and $\boldsymbol{H}(\mathrm{curl})$ finite elements. As before, we consider a shape-regular sequence $(\mathcal{T}_h)_{h\in\mathcal{H}}$ of affine simplicial meshes with a generation-compatible orientation. The key idea here is to extend the degrees of freedom (dofs) on the faces and the edges by requiring some integrability of the divergence or the curl of the function to be interpolated. This approach is useful when such integrability properties can be extracted from a PDE solved by the function in question, as it is often the case in applications (see, e.g., Chapter 51 for Darcy's equations and Chapters 43–44 for Maxwell's equations). The crucial advantage of the present approach over that from the previous chapter based only on the scale of Sobolev spaces is that interpolation error estimates with lower smoothness requirements can be obtained. On the way, we also devise a face-to-cell lifting operator that will be useful in the analysis of nonconforming approximations of elliptic problems in Chapters 40–41.

17.1 Face-to-cell lifting operator

Let us first motivate our approach informally. Let $K \in \mathcal{T}_h$ be a mesh cell and let $F \in \mathcal{F}_K$ be a face of K. Let $\boldsymbol{v}$ be a vector field defined on K. We are looking for (mild) smoothness requirements on the field $\boldsymbol{v}$ to give a meaning to the quantity $\int_F (\boldsymbol{v}{\cdot}\boldsymbol{n}_K)\phi \,\mathrm{d}s$, where ϕ is a given smooth function on F (e.g., a polynomial function) and $\boldsymbol{n}_K$ is the outward unit normal vector on ∂K. We have seen in §4.3 that it is possible to give a weak meaning in $H^{-\frac{1}{2}}(\partial K)$ to the normal trace of $\boldsymbol{v}$ on ∂K by means of an integration by parts formula if $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}; K) := \{\boldsymbol{v} \in \boldsymbol{L}^2(K) \mid \nabla{\cdot}\boldsymbol{v} \in L^2(K)\}$. In this situation, one can define the normal trace $\gamma^{\mathrm{d}}_{\partial K}(\boldsymbol{v}) \in H^{-\frac{1}{2}}(\partial K)$ by setting

$$\langle \gamma^{\mathrm{d}}_{\partial K}(\boldsymbol{v}), \psi \rangle_{\partial K} := \int_K \Big(\boldsymbol{v}{\cdot}\nabla w(\psi) + (\nabla{\cdot}\boldsymbol{v})w(\psi) \Big) \,\mathrm{d}x, \qquad (17.1)$$

for all $\psi \in H^{\frac{1}{2}}(\partial K)$, where $w(\psi) \in H^1(K)$ is a lifting of ψ, i.e., $\gamma^{\mathrm{g}}_{\partial K}(w(\psi)) = \psi$, where $\gamma^{\mathrm{g}}_{\partial K} : H^1(K) \to H^{\frac{1}{2}}(\partial K)$ is the trace map. Then one has $\gamma^{\mathrm{d}}_{\partial K}(\boldsymbol{v}) = \boldsymbol{v}_{|\partial K} \cdot \boldsymbol{n}_K$ whenever $\boldsymbol{v}$ is smooth, e.g., if $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}; K) \cap \boldsymbol{C}^0(K)$. However, the above meaning is too weak for our purpose because we need to localize the action of the normal trace to functions ψ only defined on a face F, i.e., ψ may not be defined on the whole boundary ∂K. The key to achieve this is to extend ψ by zero from F to ∂K. This obliges us to change the functional setting since the extended function is no longer in $H^{\frac{1}{2}}(\partial K)$. In what follows, we are going to use the fact that the zero-extension of a smooth function defined on a face F of ∂K is in $W^{1-\frac{1}{t},t}(\partial K)$ if $t < 2$. Let us now present a rigorous construction.

Let p, q be two real numbers such that

$$p > 2, \qquad q > \frac{2d}{2+d}. \tag{17.2}$$

Notice that $q > 1$ since $d \geq 2$. Let $\boldsymbol{v}$ be a vector field on K s.t. $\boldsymbol{v} \in \boldsymbol{L}^p(K)$ and $\nabla \cdot \boldsymbol{v} \in L^q(K)$. Let $\widetilde{p} \in (2, p]$ be such that $q \geq \frac{\widetilde{p}d}{\widetilde{p}+d}$. This is indeed possible since $p > 2$ and the function $z \mapsto \frac{zd}{z+d}$ is increasing over $\mathbb{R}_+$. We are going to construct a lifting operator (see Lemma 17.1 below)

$$L^K_F : W^{\frac{1}{\widetilde{p}}, \widetilde{p}'}(F) \longrightarrow W^{1, \widetilde{p}'}(K), \tag{17.3}$$

with conjugate number $\widetilde{p}'$ s.t. $\frac{1}{\widetilde{p}} + \frac{1}{\widetilde{p}'} = 1$, so that for all $\phi \in W^{\frac{1}{\widetilde{p}}, \widetilde{p}'}(F)$, $L^K_F(\phi)$ is a lifting of the zero-extension of ϕ to ∂K, i.e.,

$$\gamma^{\mathrm{g}}_{\partial K}(L^K_F(\phi))_{|\partial K \setminus F} = 0, \qquad \gamma^{\mathrm{g}}_{\partial K}(L^K_F(\phi))_{|F} = \phi. \tag{17.4}$$

Notice that the domain of L^K_F is $W^{1-\frac{1}{t}, t}(F)$ with $t := \widetilde{p}' < 2$, which is consistent with the above observation regarding the zero-extension to ∂K of functions defined on F. We also observe that

$$L^K_F(\phi) \in W^{1, p'}(K) \cap L^{q'}(K), \tag{17.5}$$

with conjugate numbers p', q' s.t. $\frac{1}{p} + \frac{1}{p'} = 1$, $\frac{1}{q} + \frac{1}{q'} = 1$. Indeed, $L^K_F(\phi) \in W^{1, p'}(K)$ just follows from $p' \leq \widetilde{p}'$ (i.e., $\widetilde{p} \leq p$), whereas $L^K_F(\phi) \in L^{q'}(K)$ follows from $W^{1, \widetilde{p}'}(K) \hookrightarrow L^{q'}(K)$ owing to the Sobolev embedding theorem (Theorem 2.31) (since $q' \leq \frac{\widetilde{p}'d}{d-\widetilde{p}'}$ as can be verified from $d \geq 2 > \widetilde{p}'$ and $\frac{1}{\widetilde{p}'} - \frac{1}{d} = 1 - (\frac{1}{\widetilde{p}} + \frac{1}{d}) \leq 1 - \frac{1}{q} = \frac{1}{q'}$ because $q \geq \frac{\widetilde{p}d}{\widetilde{p}+d}$).

With the lifting operator L^K_F in hand and fixing $\phi \in W^{\frac{1}{\widetilde{p}}, \widetilde{p}'}(F)$, we define the linear form $\widetilde{\sigma}_\phi$ on $\boldsymbol{V}^{\mathrm{d}}(K) := \{\boldsymbol{v} \in \boldsymbol{L}^p(K) \mid \nabla \cdot \boldsymbol{v} \in L^q(K)\}$ s.t.

$$\widetilde{\sigma}_\phi(\boldsymbol{v}) := \int_K \left(\boldsymbol{v} \cdot \nabla L^K_F(\phi) + (\nabla \cdot \boldsymbol{v}) L^K_F(\phi) \right) \mathrm{d}x. \tag{17.6}$$

The right-hand side of (17.6) is well defined owing to Hölder's inequality, and whenever the field $\boldsymbol{v}$ is smooth, we have

$$\widetilde{\sigma}_\phi(\boldsymbol{v}) = \int_{\partial K} (\boldsymbol{v}\cdot\boldsymbol{n}_K)\gamma^{\mathrm{g}}_{\partial K}(L_F^K(\phi))\,\mathrm{d}s = \int_F (\boldsymbol{v}\cdot\boldsymbol{n}_K)\phi\,\mathrm{d}s, \qquad (17.7)$$

where the second equality follows from (17.4). Thus, the linear form $\boldsymbol{v} \mapsto \widetilde{\sigma}_\phi(\boldsymbol{v})$ is an extension of the linear form $\boldsymbol{v} \mapsto \int_F (\boldsymbol{v}\cdot\boldsymbol{n}_K)\phi\,\mathrm{d}s$, which is meaningful for smooth fields $\boldsymbol{v} \in \boldsymbol{C}^0(K)$. This extension is bounded for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$, i.e., $\boldsymbol{v} \in \boldsymbol{L}^p(K)$, $\nabla\cdot\boldsymbol{v} \in L^q(K)$ with $p > 2$, $q > \frac{2d}{2+d}$ (recall that the function ϕ is fixed here).

Let us now turn our attention to the construction of the operator L_F^K.

Lemma 17.1 (Face-to-cell lifting). *Let $p > 2$ and $q > \frac{2d}{2+d}$. Let $\widetilde{p} \in (2,p]$ be such that $q \geq \frac{\widetilde{p}d}{\widetilde{p}+d}$. Let $K \in \mathcal{T}_h$ be a mesh cell and let $F \in \mathcal{F}_K$ be a face of K. There exists a lifting operator $L_F^K : W^{\frac{1}{p},\widetilde{p}'}(F) \to W^{1,\widetilde{p}'}(K)$ satisfying (17.4), and there exists c s.t. the following holds true:*

$$h_K^{\frac{d}{p}}|L_F^K(\phi)|_{W^{1,p'}(K)} + h_K^{-1+\frac{d}{q}}\|L_F^K(\phi)\|_{L^{q'}(K)} \leq c\,h_K^{-\frac{1}{\widetilde{p}}+\frac{d}{p}}\|\phi\|_{W^{\frac{1}{p},\widetilde{p}'}(F)}, \quad (17.8)$$

for all $\phi \in W^{\frac{1}{p},\widetilde{p}'}(F)$, all $K \in \mathcal{T}_h$, all $F \in \mathcal{F}_K$, and all $h \in \mathcal{H}$, with the norm $\|\phi\|_{W^{\frac{1}{p},\widetilde{p}'}(F)} := \|\phi\|_{L^{\widetilde{p}'}(F)} + h_F^{\frac{1}{\widetilde{p}}}|\phi|_{W^{\frac{1}{p},\widetilde{p}'}(F)}.$

Proof. (1) The face-to-cell lifting operator L_F^K is constructed from a lifting operator $L_{\widehat{F}}^{\widehat{K}}$ on the reference cell. Let $\widehat{K}$ be the reference cell and let $\widehat{F}$ be one of its faces. Let us define the operator $L_{\widehat{F}}^{\widehat{K}} : W^{\frac{1}{p},\widetilde{p}'}(\widehat{F}) \to W^{1,\widetilde{p}'}(\widehat{K})$. For every function $\psi \in W^{\frac{1}{p},\widetilde{p}'}(\widehat{F})$, let $\widetilde{\psi}$ denote the zero-extension of ψ to $\partial\widehat{K}$. Owing to Theorem 3.19, $\widetilde{\psi}$ is in $W^{\frac{1}{p},\widetilde{p}'}(\partial\widehat{K})$ since $\frac{\widetilde{p}'}{\widetilde{p}}=\frac{1}{\widetilde{p}-1}<1$ (i.e., $\widetilde{p} > 2$), and we have $\|\widetilde{\psi}\|_{W^{\frac{1}{p},\widetilde{p}'}(\partial\widehat{K})} \leq \widehat{c}_1\|\psi\|_{W^{\frac{1}{p},\widetilde{p}'}(\widehat{F})}$ with the norm $\|\psi\|_{W^{\frac{1}{p},\widetilde{p}'}(\widehat{F})} := \|\psi\|_{L^{\widetilde{p}'}(\widehat{F})} + \ell_{\widehat{K}}^{\frac{1}{\widetilde{p}}}|\psi|_{W^{\frac{1}{p},\widetilde{p}'}(\widehat{F})}$ and $\ell_{\widehat{K}} := 1$ is a length scale associated with $\widehat{K}$. Then we use the surjectivity of the trace map $\gamma^{\mathrm{g}}_{\widehat{K}} : W^{1,\widetilde{p}'}(\widehat{K}) \to W^{\frac{1}{p},\widetilde{p}'}(\partial\widehat{K})$ (see Theorem 3.10) to define $L_{\widehat{F}}^{\widehat{K}}(\psi) \in W^{1,\widetilde{p}'}(\widehat{K})$ such that $\gamma^{\mathrm{g}}_{\widehat{K}}(L_{\widehat{F}}^{\widehat{K}}(\psi)) = \widetilde{\psi}$ and $\|L_{\widehat{F}}^{\widehat{K}}(\psi)\|_{W^{1,\widetilde{p}'}(\widehat{K})} \leq \widehat{c}_2\|\widetilde{\psi}\|_{W^{\frac{1}{p},\widetilde{p}'}(\partial\widehat{K})}$, i.e., $\|L_{\widehat{F}}^{\widehat{K}}(\psi)\|_{W^{1,\widetilde{p}'}(\widehat{K})} \leq \widehat{c}\|\psi\|_{W^{\frac{1}{p},\widetilde{p}'}(\widehat{F})}$, with $\widehat{c} := \widehat{c}_1\widehat{c}_2$. By construction, we have $\gamma^{\mathrm{g}}_{\partial\widehat{K}}(L_{\widehat{F}}^{\widehat{K}}(\psi))_{|\widehat{F}} = \psi$ and $\gamma^{\mathrm{g}}_{\partial\widehat{K}}(L_{\widehat{F}}^{\widehat{K}}(\psi))_{|\partial\widehat{K}\setminus\widehat{F}} = 0$.

(2) We define the lifting operator $L_F^K : W^{\frac{1}{p},\widetilde{p}'}(F) \to W^{1,\widetilde{p}'}(K)$ by setting

$$L_F^K(\phi)(\boldsymbol{x}) := L_{\widehat{F}}^{\widehat{K}}(\phi \circ \boldsymbol{T}_{K|\widehat{F}})(\boldsymbol{T}_K^{-1}(\boldsymbol{x})), \quad \forall\boldsymbol{x} \in K, \quad \forall\phi \in W^{\frac{1}{p},\widetilde{p}'}(F), \quad (17.9)$$

where $\boldsymbol{T}_K : \widehat{K} \to K$ is the geometric mapping and $\widehat{F} := \boldsymbol{T}_K^{-1}(F)$. By definition, if $\boldsymbol{x} \in F$, then $\widehat{\boldsymbol{x}} := \boldsymbol{T}_K^{-1}(\boldsymbol{x}) \in \widehat{F}$ and $\boldsymbol{T}_{K|\widehat{F}}(\widehat{\boldsymbol{x}}) = \boldsymbol{x}$, so that

$$\gamma_{\partial K}^{\mathrm{g}}(L_F^K(\phi))(\boldsymbol{x}) = \gamma_{\partial \widehat{K}}^{\mathrm{g}}(L_{\widehat{F}}^{\widehat{K}}(\phi \circ \boldsymbol{T}_{K|\widehat{F}}))(\widehat{\boldsymbol{x}}) = \phi(\boldsymbol{T}_{K|\widehat{F}}(\widehat{\boldsymbol{x}})) = \phi(\boldsymbol{x}),$$

whereas if $\boldsymbol{x} \in \partial K \setminus F$, then $\widehat{\boldsymbol{x}} \in \partial \widehat{K} \setminus \widehat{F}$, so that $\gamma_{\partial \widehat{K}}^{\mathrm{g}}(L_{\widehat{F}}^{\widehat{K}}(\phi \circ \boldsymbol{T}_{K|\widehat{F}}))(\widehat{\boldsymbol{x}}) = 0$.
The above argument shows that (17.4) holds true.
(3) It remains to prove (17.8). Let us first bound $|L_F^K(\phi)|_{W^{1,p'}(K)}$. Notice that
the definition of L_K^F is equivalent to $L_F^K(\phi) \circ \boldsymbol{T}_K(\widehat{\boldsymbol{x}}) := L_{\widehat{F}}^{\widehat{K}}(\phi \circ \boldsymbol{T}_{K|\widehat{F}})(\widehat{\boldsymbol{x}})$, that
is, $\psi_K^{\mathrm{g}}(L_F^K(\phi)) := L_{\widehat{F}}^{\widehat{K}}(\psi_F^{\mathrm{g}}(\phi))$, where ψ_K^{g} is the pullback by $\boldsymbol{T}_K$, and ψ_F^{g} is
the pullback by $\boldsymbol{T}_{K|\widehat{F}}$. We infer that

$$\begin{aligned}
|L_F^K(\phi)|_{W^{1,p'}(K)} &\le c\,\|\mathbb{J}_K^{-1}\|_{\ell^2}|\det(\mathbb{J}_K)|^{\frac{1}{p'}}|L_{\widehat{F}}^{\widehat{K}}(\psi_F^{\mathrm{g}}(\phi))|_{W^{1,p'}(\widehat{K})} \\
&\le c'\,\|\mathbb{J}_K^{-1}\|_{\ell^2}|\det(\mathbb{J}_K)|^{\frac{1}{p'}}|L_{\widehat{F}}^{\widehat{K}}(\psi_F^{\mathrm{g}}(\phi))|_{W^{1,\widetilde{p}'}(\widehat{K})} \\
&\le c''\,\|\mathbb{J}_K^{-1}\|_{\ell^2}|\det(\mathbb{J}_K)|^{\frac{1}{p'}}\|\psi_F^{\mathrm{g}}(\phi)\|_{W^{\frac{1}{\widetilde{p}},\widetilde{p}'}(\widehat{F})},
\end{aligned}$$

where the first inequality follows from (11.7b) in Lemma 11.7 (transformation of Sobolev seminorms by a pullback), the second is a consequence of $\widetilde{p}' \ge p'$ (since $\widetilde{p} \le p$), and the third follows from the stability of the reference lifting operator $L_{\widehat{F}}^{\widehat{K}}$. Using now the estimate (11.7a) in Lemma 11.7 and the regularity of the mesh sequence, we infer that $\|\psi_F^{\mathrm{g}}(\phi)\|_{W^{\frac{1}{\widetilde{p}},\widetilde{p}'}(\widehat{F})} \le c|\det(\mathbb{J}_F)|^{-\frac{1}{\widetilde{p}'}}\|\phi\|_{W^{\frac{1}{\widetilde{p}},\widetilde{p}'}(F)}$, where $\mathbb{J}_F$ is the Jacobian of the mapping $\boldsymbol{T}_{K|\widehat{F}} : \widehat{F} \to F$. Combining these bounds, we obtain

$$\begin{aligned}
|L_F^K(\phi)|_{W^{1,p'}(K)} &\le c\,\|\mathbb{J}_K^{-1}\|_{\ell^2}|\det(\mathbb{J}_K)|^{\frac{1}{p'}}|\det(\mathbb{J}_F)|^{-\frac{1}{\widetilde{p}'}}\|\phi\|_{W^{\frac{1}{\widetilde{p}},\widetilde{p}'}(F)} \\
&\le c'\,h_K^{-\frac{1}{\widetilde{p}}+d(\frac{1}{\widetilde{p}}-\frac{1}{p})}\|\phi\|_{W^{\frac{1}{\widetilde{p}},\widetilde{p}'}(F)},
\end{aligned}$$

where the second bound follows from the regularity of the mesh sequence. This proves the bound on $|L_F^K(\phi)|_{W^{1,p'}(K)}$ in (17.8). The proof of the bound on $\|L_F^K(\phi)\|_{L^{q'}(K)}$ uses similar arguments together with $W^{1,\widetilde{p}'}(\widehat{K}) \hookrightarrow L^{q'}(\widehat{K})$ owing to the Sobolev embedding theorem and $q' \le \frac{\widetilde{p}'d}{d-\widetilde{p}'}$ (as already shown above). $\square$

17.2 Local interpolation in $\boldsymbol{H}(\mathrm{div})$ using liftings

Let $K \in \mathcal{T}_h$ be a mesh cell. Our goal is to show that one can extend the local dofs $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$ of the $\mathbb{RT}_{k,d}$ element to the functional space

$$\boldsymbol{V}^{\mathrm{d}}(K) := \{\boldsymbol{v} \in \boldsymbol{L}^p(K) \mid \nabla{\cdot}\boldsymbol{v} \in L^q(K)\}, \quad p > 2, q > \tfrac{2d}{2+d}. \tag{17.10}$$

Notice that $q = 2$ is always legitimate in (17.10). (Generally, one wants to take p and q as small as possible in order to make the space $\boldsymbol{V}^{\mathrm{d}}(K)$ as large as possible.) We are going to proceed as follows: we first show that the reference dofs $\{\widehat{\sigma}_i\}_{i \in \mathcal{N}}$ of the $\mathbb{RT}_{k,d}$ element can be extended to $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(\widehat{K}); \mathbb{R})$ (we use the same notation for the extended dofs for simplicity), where $\boldsymbol{V}^{\mathrm{d}}(\widehat{K})$ is defined as in (17.10). Then we establish that the contravariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{d}}$ is in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K); \boldsymbol{V}^{\mathrm{d}}(\widehat{K}))$. Owing to Proposition 9.2, we are then going to conclude that the local dofs $\{\sigma_{K,i} := \widehat{\sigma}_i \circ \boldsymbol{\psi}_K^{\mathrm{d}}\}_{i \in \mathcal{N}}$ are in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K); \mathbb{R})$. Recall that the reference dofs are defined as follows: For all $\widehat{\boldsymbol{v}} \in \mathbb{RT}_{k,d}$,

$$\widehat{\sigma}_{\widehat{F},m}^{\mathrm{f}}(\widehat{\boldsymbol{v}}) := \frac{1}{|\widehat{F}|} \int_{\widehat{F}} (\widehat{\boldsymbol{v}}{\cdot}\widehat{\boldsymbol{\nu}}_{\widehat{F}})(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) \, \mathrm{d}\widehat{s}, \qquad \forall \widehat{F} \in \mathcal{F}_{\widehat{K}}, \tag{17.11a}$$

$$\widehat{\sigma}_{j,m}^{\mathrm{c}}(\widehat{\boldsymbol{v}}) := \frac{1}{|\widehat{K}|} \int_{\widehat{K}} (\widehat{\boldsymbol{v}}{\cdot}\widehat{\boldsymbol{\nu}}_{\widehat{K},j})\psi_m \, \mathrm{d}\widehat{x}, \qquad \forall j \in \{1{:}d\}, \tag{17.11b}$$

where $\{\zeta_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$, $\{\psi_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ are bases of $\mathbb{P}_{k,d-1}$, $\mathbb{P}_{k-1,d}$ ($k \geq 1$), respectively, $\{\widehat{\boldsymbol{\nu}}_{\widehat{F}} := |\widehat{F}|\widehat{\boldsymbol{n}}_{\widehat{F}}\}_{\widehat{F} \in \mathcal{F}_{\widehat{K}}}$ and $\{\widehat{\boldsymbol{\nu}}_{\widehat{K},j} := |\widehat{F}_j|\boldsymbol{n}_{F_j}\}_{j \in \{1:d\}}$ are the normal vectors orienting the faces of $\widehat{K}$ and $\widehat{K}$ itself, respectively, and $\boldsymbol{T}_{\widehat{F}} : \widehat{S}^{d-1} \to \widehat{F}$ is an affine geometric mapping.

Let $\widehat{\boldsymbol{v}} \in \boldsymbol{V}^{\mathrm{d}}(\widehat{K})$. For the reference face dofs, inspired by §17.1, we set

$$\widehat{\sigma}_{\widehat{F},m}^{\mathrm{f}}(\widehat{\boldsymbol{v}}) := \epsilon_{\widehat{K},\widehat{F}} \int_{\widehat{K}} \left(\widehat{\boldsymbol{v}}{\cdot}\nabla L_{\widehat{F}}^{\widehat{K}}(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) + (\nabla{\cdot}\widehat{\boldsymbol{v}})L_{\widehat{F}}^{\widehat{K}}(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) \right) \mathrm{d}\widehat{x}, \tag{17.12}$$

where $\epsilon_{\widehat{K},\widehat{F}} := \widehat{\boldsymbol{n}}_{\widehat{F}}{\cdot}\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}} = \pm 1$, $\widehat{\boldsymbol{n}}_{\widehat{K}}$ is the outward unit normal to $\widehat{K}$, and $L_{\widehat{F}}^{\widehat{K}}$ is the face-to-cell lifting operator on the reference element $\widehat{K}$. For the reference cell dofs, we still use (17.11b).

Lemma 17.2 (Extended reference dofs). (i) *The definitions* (17.11a) *and* (17.12) *coincide on* $\mathbb{RT}_{k,d}$. (ii) *The extended reference dofs are s.t.* $\{\widehat{\sigma}_i\}_{i \in \mathcal{N}} \subset \mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(\widehat{K}); \mathbb{R})$.

Proof. (i) For all $\widehat{\boldsymbol{v}} \in \mathbb{RT}_{k,d}$, the divergence formula and (17.4) imply that

$$\widehat{\sigma}_{\widehat{F},m}^{\mathrm{f}}(\widehat{\boldsymbol{v}}) = \epsilon_{\widehat{K},\widehat{F}} \int_{\partial\widehat{K}} (\widehat{\boldsymbol{v}}{\cdot}\widehat{\boldsymbol{n}}_K)L_{\widehat{F}}^{\widehat{K}}(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) \, \mathrm{d}\widehat{s} = \int_{\widehat{F}} (\widehat{\boldsymbol{v}}{\cdot}\widehat{\boldsymbol{n}}_{\widehat{F}})(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) \, \mathrm{d}\widehat{s},$$

showing that (17.11a) and (17.12) coincide on $\mathbb{RT}_{k,d}$.
(ii) Since $\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}$ is a smooth function on $\widehat{F}$, $L_{\widehat{F}}^{\widehat{K}}(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) \in W^{1,p'}(\widehat{K}) \cap L^{q'}(\widehat{K})$, where p', q' are the conjugate numbers of p, q. Hence, Hölder's inequality implies that

$$|\sigma_{\widehat{F},m}^{\mathrm{f}}(\widehat{\boldsymbol{v}})| \leq \widehat{c}\,\ell_{\widehat{K}}^{d-1}(\ell_{\widehat{K}}^{-\frac{d}{p}}\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})} + \ell_{\widehat{K}}^{1-\frac{d}{q}}\|\nabla{\cdot}\widehat{\boldsymbol{v}}\|_{L^q(\widehat{K})}), \tag{17.13}$$

where $\ell_{\widehat{K}} := 1$ is a reference length for $\widehat{K}$. This shows that the extended reference face dofs are in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(\widehat{K});\mathbb{R})$. For the extended reference cell dofs, we simply have

$$|\widehat{\sigma}_{j,m}^{\mathrm{c}}(\widehat{\boldsymbol{v}})| \leq \widehat{c}\,\ell_{\widehat{K}}^{d-1-\frac{d}{p}} \|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})}, \tag{17.14}$$

since $\frac{1}{|\widehat{K}|}\|\widehat{\boldsymbol{\nu}}_{\widehat{K},j}\|_{\ell^2} \leq \widehat{c}\,\ell_{\widehat{K}}^{-1}$, $\{\psi_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{c}}\}} \subset L^\infty(\widehat{K})$, and $\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^1(\widehat{K})} \leq \ell_{\widehat{K}}^{d-\frac{d}{p}}\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})}$ owing to Hölder's inequality. $\qquad\square$

Proposition 17.3 (Extended local dofs). *Let $K \in \mathcal{T}_h$ be a mesh cell. Let $\boldsymbol{V}^{\mathrm{d}}(K)$ be defined in (17.10). Let $\boldsymbol{V}^{\mathrm{d}}(\widehat{K})$ be defined similarly.* (i) *The contravariant Piola transformation $\boldsymbol{\psi}_K^{\mathrm{d}}$ is in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K);\boldsymbol{V}^{\mathrm{d}}(\widehat{K}))$.* (ii) *The extended local dofs*

$$\sigma_{K,i} := \widehat{\sigma}_i \circ \boldsymbol{\psi}_K^{\mathrm{d}} : \boldsymbol{V}^{\mathrm{d}}(K) \to \mathbb{R}, \qquad \forall i \in \mathcal{N}, \tag{17.15}$$

are in $\mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K);\mathbb{R})$, and there is c s.t.

$$\max_{i\in\mathcal{N}} |\sigma_{K,i}(\boldsymbol{v})| \leq c\,h_K^{d-1}\Big(h_K^{-\frac{d}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K^{1-\frac{d}{q}}\|\nabla{\cdot}\boldsymbol{v}\|_{L^q(K)}\Big), \tag{17.16}$$

for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. (1) Let $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$. Since the mesh is affine and $\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) := \mathbb{A}_K^{\mathrm{d}}(\boldsymbol{v}\circ\boldsymbol{T}_K)$ with $\mathbb{A}_K^{\mathrm{d}} := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}$, we infer, as in the proof of Proposition 16.1, that

$$\|\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} \leq c\,\|\mathbb{A}_K^{\mathrm{d}}\|_{\ell^2}|\det(\mathbb{J}_K)|^{-\frac{1}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} \leq c'\,h_K^{d-1-\frac{d}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}.$$

Moreover, since $\nabla{\cdot}\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) = \det(\mathbb{J}_K)(\nabla{\cdot}\boldsymbol{v})\circ\boldsymbol{T}_K$ owing to (9.8c), we infer that

$$\|\nabla{\cdot}\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\|_{L^q(\widehat{K})} = |\det(\mathbb{J}_K)|^{1-\frac{1}{q}}\|\nabla{\cdot}\boldsymbol{v}\|_{L^q(K)} \leq c\,h_K^{d(1-\frac{1}{q})}\|\nabla{\cdot}\boldsymbol{v}\|_{L^q(K)},$$

where we used the regularity of the mesh sequence. The above bounds show that $\boldsymbol{\psi}_K^{\mathrm{d}} \in \mathcal{L}(\boldsymbol{V}^{\mathrm{d}}(K);\boldsymbol{V}^{\mathrm{d}}(\widehat{K}))$ with

$$\ell_{\widehat{K}}^{-\frac{d}{p}}\|\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} + \ell_{\widehat{K}}^{1-\frac{d}{q}}\|\nabla{\cdot}\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v})\|_{L^q(\widehat{K})} \leq$$
$$c\,h_K^{d-1}\Big(h_K^{-\frac{d}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K^{1-\frac{d}{q}}\|\nabla{\cdot}\boldsymbol{v}\|_{L^q(K)}\Big).$$

(2) The assertion on the extended local dofs follows from the above bound on $\boldsymbol{\psi}_K^{\mathrm{d}}$ and the bounds (17.13)–(17.14) on the extended reference dofs. $\qquad\square$

Remark 17.4 (Extended local dofs). The reader is invited to verify (see Exercise 17.2) that for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$, all $F \in \mathcal{F}_K$, and every integer $m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}$,

$$\sigma^{\mathrm{f}}_{F,m}(\boldsymbol{v}) = \epsilon_{K,F} \int_K \left(\boldsymbol{v}\cdot\nabla L^K_F(\zeta_m \circ \boldsymbol{T}^{-1}_{K,F}) + (\nabla\cdot\boldsymbol{v})L^K_F(\zeta_m \circ \boldsymbol{T}^{-1}_{K,F}) \right) \mathrm{d}x, \quad (17.17)$$

with $\boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}}$, $\epsilon_{K,F} := \boldsymbol{n}_F\cdot\boldsymbol{n}_K = \pm 1$, and L^K_F defined in (17.9). For all $\boldsymbol{v} \in \mathbb{RT}_{k,d}$, and more generally, for all $\boldsymbol{v} \in \boldsymbol{W}^{s,p}(K)\cap\boldsymbol{V}^{\mathrm{d}}(K)$ with $sp > 1$, $p \in (1,\infty)$ or $s = 1$, $p = 1$, we recover that $\sigma^{\mathrm{f}}_{F,m}(\boldsymbol{v}) = \frac{1}{|F|}\int_F (\boldsymbol{v}\cdot\boldsymbol{\nu}_F)(\zeta_m \circ \boldsymbol{T}^{-1}_{K,F})\,\mathrm{d}s$, where $\boldsymbol{\nu}_F = |F|\boldsymbol{n}_F$ is the normal vector orienting F. Concerning the extended local cell dofs, it follows from Lemma 14.18 that $\sigma^{\mathrm{c}}_{j,m}(\boldsymbol{v}) = \frac{1}{|K|}\int_K (\boldsymbol{v}\cdot\boldsymbol{\nu}_{K,j})(\psi_m \circ \boldsymbol{T}^{-1}_K)\,\mathrm{d}x$ for every integers $m \in \{1{:}n^{\mathrm{c}}_{\mathrm{sh}}\}$ and $j \in \{1{:}d\}$, where $\{\boldsymbol{\nu}_{K,j}\}_{j\in\{1:d\}}$ are the d normal vectors orienting K. $\qquad\square$

Using the extended dofs to define $\mathcal{I}^{\mathrm{d}}_K : \boldsymbol{V}^{\mathrm{d}}(K) \to \mathbb{RT}_{k,d}$, we can now derive an estimate on $\|\boldsymbol{v} - \mathcal{I}^{\mathrm{d}}_K(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)}$ for all $\boldsymbol{v} \in \boldsymbol{H}^r(K)$ with $r \in (0,1]$. This result complements Theorem 16.6 which is valid only for $r \in (\frac{1}{2},1]$.

Theorem 17.5 (Approximation, $r\in(0,1]$). *For all $r \in (0,1]$ and all $q > \frac{2d}{2+d}$, there is c, unbounded as either $r \downarrow 0$ or $q \downarrow \frac{2d}{2+d}$, such that*

$$\|\boldsymbol{v} - \mathcal{I}^{\mathrm{d}}_K(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} \le c \left(h^r_K |\boldsymbol{v}|_{\boldsymbol{H}^r(K)} + h^{1-d(\frac{1}{q}-\frac{1}{2})}_K \|\nabla\cdot\boldsymbol{v}\|_{L^q(K)} \right), \quad (17.18)$$

for all $\boldsymbol{v} \in \boldsymbol{H}^r(K)$ with $\nabla\cdot\boldsymbol{v} \in L^q(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. Let $\boldsymbol{v} \in \boldsymbol{H}^r(K)$ be s.t. $\nabla\cdot\boldsymbol{v} \in L^q(K)$. If $2r \ge d$ (i.e., if $r = 1$ and $d = 2$), let p be any real number larger than 2. If $2r < d$, let $p := \frac{2d}{d-2r}$ (note that $p > 2$ since $r > 0$). Owing to the Sobolev embedding theorem (Theorem 2.31), we have $\boldsymbol{H}^r(K) \hookrightarrow \boldsymbol{L}^p(K)$ which implies that $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{d}}(K)$, so that $\mathcal{I}^{\mathrm{d}}_K(\boldsymbol{v})$ is well defined. Moreover, since $\boldsymbol{H}^r(\widehat{K}) \hookrightarrow \boldsymbol{L}^p(\widehat{K})$, we have $\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})} \le \widehat{c}(\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^2(\widehat{K})} + |\widehat{\boldsymbol{v}}|_{\boldsymbol{H}^r(\widehat{K})})$ for all $\widehat{\boldsymbol{v}} \in \boldsymbol{H}^r(\widehat{K})$. Taking $\widehat{\boldsymbol{v}} := \boldsymbol{v} \circ \boldsymbol{T}_K$ and using the regularity of the mesh sequence, this gives

$$h^{-\frac{d}{p}}_K \|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} \le c\, h^{-\frac{d}{2}}_K \left(\|\boldsymbol{v}\|_{\boldsymbol{L}^2(K)} + h^r_K |\boldsymbol{v}|_{\boldsymbol{H}^r(K)} \right). \quad (17.19)$$

Using again the regularity of the mesh sequence, Proposition 12.5, and the bound (17.16) on the local dofs, we infer that

$$h^{-\frac{d}{2}}_K \|\mathcal{I}^{\mathrm{d}}_K(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} \le c\, h^{1-d}_K \max_{i\in\mathcal{N}} |\sigma_{K,i}(\boldsymbol{v})|$$
$$\le c' \left(h^{-\frac{d}{p}}_K \|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h^{1-\frac{d}{q}}_K \|\nabla\cdot\boldsymbol{v}\|_{L^q(K)} \right). \quad (17.20)$$

Combining (17.19) with (17.20) leads to

$$\|\mathcal{I}^{\mathrm{d}}_K(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} \le c \left(\|\boldsymbol{v}\|_{\boldsymbol{L}^2(K)} + h^r_K |\boldsymbol{v}|_{\boldsymbol{H}^r(K)} + h^{1-d(\frac{1}{q}-\frac{1}{2})}_K \|\nabla\cdot\boldsymbol{v}\|_{L^q(K)} \right).$$

Since $\mathbb{P}_{0,d} \subset \mathbb{RT}_{k,d}$ is pointwise invariant under $\mathcal{I}^{\mathrm{d}}_K$, we infer that

$$\|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} = \inf_{\boldsymbol{q} \in \mathbb{P}_{0,d}} \|\boldsymbol{v} - \boldsymbol{q} + \mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v} - \boldsymbol{q})\|_{\boldsymbol{L}^2(K)}$$

$$\leq c\, h_K^r |\boldsymbol{v}|_{\boldsymbol{H}^r(K)} + \inf_{\boldsymbol{q} \in \mathbb{P}_{0,d}} \|\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v} - \boldsymbol{q})\|_{\boldsymbol{L}^2(K)},$$

where we used the fractional Poincaré–Steklov inequality (12.14) in K (and $r \leq 1$). Moreover, the above bound on $\mathcal{I}_K^{\mathrm{d}}$ together with $|\boldsymbol{v} - \boldsymbol{q}|_{\boldsymbol{H}^r(K)} = |\boldsymbol{v}|_{\boldsymbol{H}^r(K)}$ and $\nabla \cdot \boldsymbol{q} = 0$ for all $\boldsymbol{q} \in \mathbb{P}_{0,d}$ implies that

$$\|\mathcal{I}_K^{\mathrm{d}}(\boldsymbol{v} - \boldsymbol{q})\|_{\boldsymbol{L}^2(K)} \leq c\Big(\|\boldsymbol{v} - \boldsymbol{q}\|_{\boldsymbol{L}^2(K)} + h_K^r |\boldsymbol{v}|_{\boldsymbol{H}^r(K)} + h_K^{1 - d(\frac{1}{q} - \frac{1}{2})} \|\nabla \cdot \boldsymbol{v}\|_{L^q(K)} \Big).$$

Taking the infimum over $\boldsymbol{q} \in \mathbb{P}_{0,d}$ and invoking again the fractional Poincaré–Steklov inequality (12.14) leads to the bound (17.18). □

Remark 17.6 (Quasi-interpolation). We construct in Chapter 22 a quasi-interpolation operator $\mathcal{I}_h^{\mathrm{d,av}}$ s.t. $\|\boldsymbol{v} - \mathcal{I}_h^{\mathrm{d,av}}(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} \leq c h_K^r |\boldsymbol{v}|_{\boldsymbol{H}^r(D_K)}$ for all $\boldsymbol{v} \in \boldsymbol{H}^r(D_K)$, $r \in (0, 1]$, where D_K is a local neighborhood of K (see Theorem 22.6). Thus, contrary to the canonical interpolation operator $\mathcal{I}_h^{\mathrm{d}}$, the quasi-interpolation operator $\mathcal{I}_h^{\mathrm{d,av}}$ gives an optimal error estimate for all $r > 0$ without making any assumption on the divergence of the vector field that is approximated. We are going to invoke $\mathcal{I}_h^{\mathrm{d,av}}$ instead of $\mathcal{I}_h^{\mathrm{d}}$ most of the time in the rest of the book. □

17.3 Local interpolation in $\boldsymbol{H}(\mathrm{curl})$ using liftings

In this section, we assume that $d = 3$. Let $K \in \mathcal{T}_h$ be a mesh cell with outward unit normal $\boldsymbol{n}_K$. Our goal is to show that one can extend the local dofs $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$ of the $\mathbb{N}_{k,d}$ element to the functional space

$$\boldsymbol{V}^{\mathrm{c}}(K) := \{\boldsymbol{v} \in \boldsymbol{L}^p(K) \mid \nabla \times \boldsymbol{v} \in \boldsymbol{L}^p(K),\, \boldsymbol{v} \times \boldsymbol{n}_K \in \boldsymbol{L}^p(\partial K)\}, \qquad (17.21)$$

where $p > 2$. We proceed as in §17.2: the idea is to use face-to-cell lifting operators to give a (weak) meaning to the face dofs, together with additional edge-to-face lifting operators to give a (weak) meaning to the edge dofs.

Let us start with the reference dofs $\{\widehat{\sigma}_i\}_{i \in \mathcal{N}}$ of the $\mathbb{N}_{k,d}$ element and let us show that these dofs can be extended to $\mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(\widehat{K}); \mathbb{R})$ (we use the same notation for the extended dofs for simplicity), where $\boldsymbol{V}^{\mathrm{c}}(\widehat{K})$ is defined similarly to (17.21). Recall that the reference dofs are defined as follows: For all $\widehat{\boldsymbol{v}} \in \mathbb{N}_{k,d}$:

$$\widehat{\sigma}_{\widehat{E},m}^{\mathrm{e}}(\widehat{\boldsymbol{v}}) := \frac{1}{|\widehat{E}|} \int_{\widehat{E}} (\widehat{\boldsymbol{v}} \cdot \widehat{\boldsymbol{t}}_{\widehat{E}})(\mu_m \circ \boldsymbol{T}_{\widehat{E}}^{-1}) \, \mathrm{d}\widehat{l}, \quad \forall \widehat{E} \in \mathcal{E}_{\widehat{K}}, \tag{17.22a}$$

$$\widehat{\sigma}_{\widehat{F},j,m}^{\mathrm{f}}(\widehat{\boldsymbol{v}}) := \frac{1}{|\widehat{F}|} \int_{\widehat{F}} (\widehat{\boldsymbol{v}} \cdot \widehat{\boldsymbol{t}}_{\widehat{F},j})(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) \, \mathrm{d}\widehat{s}, \quad \forall \widehat{F} \in \mathcal{F}_{\widehat{K}}, \forall j \in \{1,2\}, \tag{17.22b}$$

$$\widehat{\sigma}_{j,m}^{\mathrm{c}}(\widehat{\boldsymbol{v}}) := \frac{1}{|\widehat{K}|} \int_{\widehat{K}} (\widehat{\boldsymbol{v}} \cdot \widehat{\boldsymbol{t}}_{\widehat{K},j}) \psi_m \, \mathrm{d}\widehat{x}, \qquad \forall j \in \{1{:}3\}, \tag{17.22c}$$

where $\{\mu_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$, $\{\zeta_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$, and $\{\psi_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{c}}\}}$ are bases of $\mathbb{P}_{k,1}$, $\mathbb{P}_{k-1,2}$ ($k \geq 1$), and $\mathbb{P}_{k-2,3}$ ($k \geq 2$), respectively, the tangent vectors $\{\widehat{\boldsymbol{t}}_{\widehat{E}}\}_{\widehat{E} \in \mathcal{E}_{\widehat{K}}}$, $\{\widehat{\boldsymbol{t}}_{\widehat{F},j}\}_{\widehat{F} \in \mathcal{F}_{\widehat{K}}, j \in \{1,2\}}$, and $\{\widehat{\boldsymbol{t}}_{\widehat{K},j}\}_{j \in \{1,2,3\}}$ orient the edges of $\widehat{K}$, the faces of $\widehat{K}$, and $\widehat{K}$ itself, respectively, and $\boldsymbol{T}_{\widehat{E}} : \widehat{S}^1 \to \widehat{E}$ and $\boldsymbol{T}_{\widehat{F}} : \widehat{S}^2 \to \widehat{F}$ are affine geometric mappings.

Let $\widehat{\boldsymbol{v}} \in \boldsymbol{V}^{\mathrm{c}}(\widehat{K})$. One does not need to change the definition of the cell and face dofs, i.e., we still use (17.22b)–(17.22c). The difficulty consists of extending the edge dofs defined in (17.22a). Let $\widehat{E} \in \mathcal{E}_{\widehat{K}}$ be an edge of $\widehat{K}$ and assume that $\widehat{E}$ is an edge of a face $\widehat{F} \in \mathcal{F}_{\widehat{K}}$. By proceeding as in §17.1 (we take $\widetilde{p} = p > 2$ and $q = p$, noticing that $q \geq \frac{\widetilde{p}d}{\widetilde{p}+d}$ and $q \geq \frac{\widetilde{p}(d-1)}{\widetilde{p}+d-1}$), we can define an edge-to-face lifting operator $L_{\widehat{E}}^{\widehat{F}} : W^{\frac{1}{p},p'}(\widehat{E}) \to W^{1,p'}(\widehat{F})$. Then $L_{\widehat{E}}^{\widehat{K}} := L_{\widehat{F}}^{\widehat{K}} \circ L_{\widehat{E}}^{\widehat{F}} : W^{\frac{1}{p},p'}(\widehat{E}) \to W^{1,p'}(\widehat{K})$ is a bounded edge-to-cell lifting operator since $W^{1,p'}(\widehat{F}) \hookrightarrow W^{\frac{1}{p},p'}(\widehat{F})$. We extend the reference edge dofs as follows:

$$\widehat{\sigma}_{\widehat{E},m}^{\mathrm{e}}(\widehat{\boldsymbol{v}}) := \epsilon_{\widehat{K},\widehat{F},\widehat{E}} \bigg(\int_{\widehat{K}} (\nabla \times \widehat{\boldsymbol{v}}) \cdot \nabla L_{\widehat{E}}^{\widehat{K}}(\mu_m \circ \boldsymbol{T}_{\widehat{E}}^{-1}) \, \mathrm{d}\widehat{x}$$

$$+ \int_{\widehat{F}} (\widehat{\boldsymbol{v}} \times \widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}) \cdot \nabla L_{\widehat{E}}^{\widehat{F}}(\mu_m \circ \boldsymbol{T}_{\widehat{E}}^{-1}) \, \mathrm{d}\widehat{s} \bigg), \tag{17.23}$$

with $\epsilon_{\widehat{K},\widehat{F},\widehat{E}} := \widehat{\boldsymbol{\tau}}_{\widehat{K},\widehat{F}|\widehat{E}} \cdot \widehat{\boldsymbol{\tau}}_{\widehat{E}}$, where $\widehat{\boldsymbol{\tau}}_{\widehat{K},\widehat{F}}$ is the unit vector tangent to $\partial \widehat{F}$ with the (unique) orientation that is compatible with $\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}$, i.e., $\widehat{\boldsymbol{\tau}}_{\widehat{K},\widehat{F}} \times \widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}$ points outward $\widehat{F}$ a.e. on $\partial \widehat{F}$, $\widehat{\boldsymbol{n}}_{\widehat{K}}$ being the outward unit normal to $\widehat{K}$.

Lemma 17.7 (Extended reference dofs). (i) *The definitions* (17.22a) *and* (17.23) *coincide on* $\mathbf{N}_{k,d}$. (ii) *The extended reference dofs are s.t.* $\{\widehat{\sigma}_i\}_{i \in \mathcal{N}} \subset \mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(\widehat{K}); \mathbb{R})$.

Proof. (i) For all $\widehat{\boldsymbol{v}} \in \mathbf{N}_{k,d}$, using the notation $\mu_{\widehat{E},m} := \mu_m \circ \boldsymbol{T}_{\widehat{E}}^{-1}$, we have

$$\epsilon_{\widehat{K},\widehat{F},\widehat{E}}\widehat{\sigma}^{\mathrm{e}}_{\widehat{E},m}(\widehat{\boldsymbol{v}})$$

$$= \int_{\widehat{K}}(\nabla\times\widehat{\boldsymbol{v}})\cdot\nabla L^{\widehat{K}}_{\widehat{E}}(\mu_{\widehat{E},m})\,\mathrm{d}\widehat{x} + \int_{\widehat{F}}(\widehat{\boldsymbol{v}}\times\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}})\cdot\nabla L^{\widehat{F}}_{\widehat{E}}(\mu_{\widehat{E},m})\,\mathrm{d}\widehat{s}$$

$$= \int_{\partial\widehat{K}}((\nabla\times\widehat{\boldsymbol{v}})L^{\widehat{K}}_{\widehat{E}}(\mu_{\widehat{E},m}))\cdot\widehat{\boldsymbol{n}}_{K}\,\mathrm{d}\widehat{s} + \int_{\widehat{F}}(\widehat{\boldsymbol{v}}\times\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}})\cdot\nabla L^{\widehat{F}}_{\widehat{E}}(\mu_{\widehat{E},m})\,\mathrm{d}\widehat{s}$$

$$= \int_{\widehat{F}}((\nabla\times\widehat{\boldsymbol{v}})L^{\widehat{F}}_{\widehat{E}}(\mu_{\widehat{E},m}))\cdot\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}\,\mathrm{d}\widehat{s} + \int_{\widehat{F}}(\nabla L^{\widehat{F}}_{\widehat{E}}(\mu_{\widehat{E},m})\times\widehat{\boldsymbol{v}})\cdot\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}\,\mathrm{d}\widehat{s}$$

$$= \int_{\widehat{F}}\nabla\times(\widehat{\boldsymbol{v}}L^{\widehat{F}}_{\widehat{E}}(\mu_{\widehat{E},m}))\cdot\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}\,\mathrm{d}\widehat{s}$$

$$= \int_{\partial\widehat{F}}(\widehat{\boldsymbol{v}}L^{\widehat{F}}_{\widehat{E}}(\mu_{\widehat{E},m}))\cdot\widehat{\boldsymbol{\tau}}_{\widehat{K},\widehat{F}}\,\mathrm{d}\widehat{l} = \epsilon_{\widehat{K},\widehat{F},\widehat{E}}\int_{\widehat{E}}(\widehat{\boldsymbol{v}}\cdot\widehat{\boldsymbol{\tau}}_{\widehat{E}})\mu_{\widehat{E},m}\,\mathrm{d}\widehat{l},$$

where we used the definition (17.23) in the first line, the divergence theorem in $\widehat{K}$ and $\nabla\cdot(\nabla\times)=0$ in the second line, the definition of $L^{\widehat{K}}_{\widehat{E}}$ and the fact that $L^{\widehat{K}}_{\widehat{F}}$ vanishes outside $\widehat{F}$ in the third line, an algebraic identity on the curl operator in the fourth line, and the Kelvin–Stokes formula (16.15) together with the fact that $L^{\widehat{F}}_{\widehat{E}}$ vanishes outside $\widehat{E}$ and the definition of $\epsilon_{\widehat{K},\widehat{F},\widehat{E}}$ in the last line. This proves the first assertion.

(ii) Owing to Hölder's inequality, we bound the reference edge dofs as follows:

$$|\widehat{\sigma}^{\mathrm{e}}_{\widehat{E},m}(\widehat{\boldsymbol{v}})| \leq \widehat{c}\,\ell_{\widehat{K}}\big(\ell_{\widehat{K}}^{1-\frac{3}{p}}\|\nabla\times\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^{p}(\widehat{K})} + \ell_{\widehat{K}}^{-\frac{2}{p}}\|\widehat{\boldsymbol{v}}\times\widehat{\boldsymbol{n}}_{\widehat{K}}\|_{\boldsymbol{L}^{p}(\partial\widehat{K})}\big),$$

with the reference length scale $\ell_{\widehat{K}} := 1$. Similarly, for the reference face dofs and the reference cell dofs, we obtain

$$|\widehat{\sigma}^{\mathrm{f}}_{\widehat{F},j,m}(\widehat{\boldsymbol{v}})| \leq \widehat{c}\,\ell_{\widehat{K}}\ell_{\widehat{K}}^{-\frac{2}{p}}\|\widehat{\boldsymbol{v}}\times\widehat{\boldsymbol{n}}_{\widehat{K}}\|_{\boldsymbol{L}^{p}(\partial\widehat{K})}, \quad |\widehat{\sigma}^{\mathrm{c}}_{j,m}(\widehat{\boldsymbol{v}})| \leq \widehat{c}\,\ell_{\widehat{K}}\ell_{\widehat{K}}^{-\frac{3}{p}}\|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^{p}(\widehat{K})},$$

since $|\widehat{\boldsymbol{v}}\cdot\widehat{\boldsymbol{t}}_{\widehat{F},j}| \leq \|\widehat{\boldsymbol{v}}\times\widehat{\boldsymbol{n}}_{\widehat{K}|\widehat{F}}\|_{\ell^{2}}$, $\forall j\in\{1,2\}$, and $|\widehat{\boldsymbol{v}}\cdot\widehat{\boldsymbol{t}}_{\widehat{K},j}| \leq \|\widehat{\boldsymbol{v}}\|_{\ell^{2}}$, $\forall j\in\{1{:}3\}$. This proves that $\{\widehat{\sigma}_{i}\}_{i\in\mathcal{N}} \subset \mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(\widehat{K});\mathbb{R})$. $\qquad\square$

Remark 17.8 (Edge dofs). The proof of Step (i) shows that $\widehat{\sigma}^{\mathrm{e}}_{\widehat{E},m}(\widehat{\boldsymbol{v}})$ is independent of the face $\widehat{F}\in\mathcal{F}_{\widehat{K}}$ containing $\widehat{E}$ if $\widehat{\boldsymbol{v}}$ is smooth enough. $\qquad\square$

Proposition 17.9 (Extended local dofs). *Let $K\in\mathcal{T}_{h}$ be a mesh cell. Let $\boldsymbol{V}^{\mathrm{c}}(K)$ be defined in (17.21). Let $\boldsymbol{V}^{\mathrm{c}}(\widehat{K})$ be defined similarly.* (i) *The covariant Piola transformation $\boldsymbol{\psi}^{\mathrm{c}}_{K}$ is in $\mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(K);\boldsymbol{V}^{\mathrm{c}}(\widehat{K}))$.* (ii) *The extended local dofs*

$$\sigma_{K,i} := \widehat{\sigma}_{i}\circ\boldsymbol{\psi}^{\mathrm{c}}_{K} : \boldsymbol{V}^{\mathrm{c}}(K)\to\mathbb{R}, \qquad \forall i\in\mathcal{N}, \qquad\qquad (17.24)$$

are in $\mathcal{L}(\boldsymbol{V}^{\mathrm{c}}(K);\mathbb{R})$, and there is c s.t., for all $\boldsymbol{v}\in\boldsymbol{V}^{\mathrm{c}}(K)$, all $K\in\mathcal{T}_{h}$, and all $h\in\mathcal{H}$,

$$\max_{i\in\mathcal{N}}|\sigma_{K,i}(\boldsymbol{v})| \le c\,h_K\Big(h_K^{-\frac{3}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K^{1-\frac{3}{p}}\|\nabla{\times}\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}$$
$$+\, h_K^{-\frac{2}{p}}\|\boldsymbol{v}{\times}\boldsymbol{n}_K\|_{\boldsymbol{L}^p(\partial K)}\Big). \quad (17.25)$$

Proof. (i) Corollary 9.9 gives $\nabla{\times}\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) = \boldsymbol{\psi}_K^{\mathrm{d}}(\nabla{\times}\boldsymbol{v})$. Owing to Lemma 11.7, we infer that

$$\|\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} \le |\mathrm{det}(\mathbb{J}_K)|^{-\frac{1}{p}}\|\mathbb{J}_K^{\mathsf{T}}\|_{\ell^2}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)},$$
$$\|\nabla{\times}\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} = \|\boldsymbol{\psi}_K^{\mathrm{d}}(\nabla{\times}\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} \le |\mathrm{det}(\mathbb{J}_K)|^{1-\frac{1}{p}}\|\mathbb{J}_K^{-1}\|_{\ell^2}\|\nabla{\times}\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}.$$

Invoking the regularity of the mesh sequence, we obtain

$$\ell_{\widehat{K}}^{\frac{3}{p}}\|\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} + \ell_{\widehat{K}}^{1-\frac{3}{p}}\|\nabla{\times}\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^p(\widehat{K})} \le$$
$$c\,h_K\Big(h_K^{-\frac{3}{p}}\|\boldsymbol{v}\|_{\boldsymbol{L}^p(K)} + h_K^{1-\frac{3}{p}}\|\nabla{\times}\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}\Big). \quad (17.26)$$

Moreover, since (9.10) implies that $\widehat{\boldsymbol{n}}_{\widehat{K}} = \|\mathbb{J}_K^{\mathsf{T}}(\boldsymbol{n}_K\circ\boldsymbol{T}_K)\|_{\ell^2}^{-1}\mathbb{J}_K^{\mathsf{T}}(\boldsymbol{n}_K\circ\boldsymbol{T}_K)$ a.e. on $\partial\widehat{K}$ (that is, for all $\widehat{\boldsymbol{x}}\in\partial\widehat{K}$ s.t. $\widehat{\boldsymbol{x}}$ lies in the interior of a face of $\widehat{K}$), we have

$$\|\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}){\times}\widehat{\boldsymbol{n}}_{\widehat{K}}\|_{\boldsymbol{L}^p(\partial\widehat{K})}^p = \int_{\partial\widehat{K}}\big\|(\mathbb{J}_K^{\mathsf{T}}(\boldsymbol{v}\circ\boldsymbol{T}_K)){\times}\widehat{\boldsymbol{n}}_{\widehat{K}}\big\|_{\ell^2}^p\,\mathrm{d}\widehat{s}$$
$$= \int_{\partial\widehat{K}}\|\mathbb{J}_K^{\mathsf{T}}(\boldsymbol{n}_K\circ\boldsymbol{T}_K)\|_{\ell^2}^{-p}\big\|(\mathbb{J}_K^{\mathsf{T}}(\boldsymbol{v}\circ\boldsymbol{T}_K)){\times}(\mathbb{J}_K^{\mathsf{T}}(\boldsymbol{n}_K\circ\boldsymbol{T}_K))\big\|_{\ell^2}^p\,\mathrm{d}\widehat{s}.$$

Using the result from Exercise 9.5, we infer that

$$\|\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}){\times}\widehat{\boldsymbol{n}}_{\widehat{K}}\|_{\boldsymbol{L}^p(\partial\widehat{K})}^p =$$
$$\int_{\partial\widehat{K}}\|\mathbb{J}_K^{\mathsf{T}}(\boldsymbol{n}_K\circ\boldsymbol{T}_K)\|_{\ell^2}^{-p}|\mathrm{det}(\mathbb{J}_K^{\mathsf{T}})|^p\big\|\mathbb{J}_K^{-1}((\boldsymbol{v}{\times}\boldsymbol{n}_K)\circ\boldsymbol{T}_K)\big\|_{\ell^2}^p\,\mathrm{d}\widehat{s}.$$

We now use the transformation of the surface measure from Lemma 9.12, $\mathrm{d}\widehat{s} = |\mathrm{det}(\mathbb{J}_K)|^{-1}\|\mathbb{J}_K^{\mathsf{T}}\boldsymbol{n}_K\|_{\ell^2}\,\mathrm{d}s$, to obtain

$$\|\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}){\times}\widehat{\boldsymbol{n}}_{\widehat{K}}\|_{\boldsymbol{L}^p(\partial\widehat{K})}^p = \int_{\partial K}\|\mathbb{J}_K^{\mathsf{T}}\boldsymbol{n}_K\|_{\ell^2}^{1-p}|\mathrm{det}(\mathbb{J}_K)|^{p-1}\big\|\mathbb{J}_K^{-1}(\boldsymbol{v}{\times}\boldsymbol{n}_K)\big\|_{\ell^2}^p\,\mathrm{d}s$$
$$\le \int_{\partial K}\|\mathbb{J}_K\|_{\ell^2}^{1-p}\|\mathbb{J}_K^{-1}\|_{\ell^2}^p|\mathrm{det}(\mathbb{J}_K)|^{p-1}\|\boldsymbol{v}{\times}\boldsymbol{n}_K\|_{\ell^2}^p\,\mathrm{d}s.$$

Using the regularity of the mesh sequence, we conclude that

$$\ell_{\widehat{K}}^{1-\frac{2}{p}}\|\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}){\times}\widehat{\boldsymbol{n}}_{\widehat{K}}\|_{\boldsymbol{L}^p(\partial\widehat{K})} \le c\,h_K^{1-\frac{2}{p}}\|\boldsymbol{v}{\times}\boldsymbol{n}_K\|_{\boldsymbol{L}^p(\partial K)}. \quad (17.27)$$

(ii) We have shown in the proof of Lemma 17.7 that

$$\max_{i \in \mathcal{N}} |\widehat{\sigma}_i(\widehat{\boldsymbol{v}})| \leq \widehat{c}\, \ell_{\widehat{K}} \Big(\ell_{\widehat{K}}^{-\frac{3}{p}} \|\widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})} $$
$$+ \ell_{\widehat{K}}^{1-\frac{3}{p}} \|\nabla \times \widehat{\boldsymbol{v}}\|_{\boldsymbol{L}^p(\widehat{K})} + \ell_{\widehat{K}}^{-\frac{2}{p}} \|\widehat{\boldsymbol{v}} \times \widehat{\boldsymbol{n}}_{\widehat{K}}\|_{\boldsymbol{L}^p(\partial \widehat{K})} \Big).$$

Then the assertion on the extended local dofs follows from the above bound together with (17.26)–(17.27). □

Remark 17.10 (Extended local dofs). Let $E \in \mathcal{E}_K$ be an edge of K oriented by the unit tangent vector $\boldsymbol{\tau}_E$ and let $F \in \mathcal{F}_K$ be a face of K s.t. $E \subset \partial F$. Let $\epsilon_{K,F,E} := \boldsymbol{\tau}_{K,F|E} \cdot \boldsymbol{\tau}_E$, where $\boldsymbol{\tau}_{K,F}$ is the unit vector tangent to ∂F with the (unique) orientation that is compatible with $\boldsymbol{n}_{K|F}$ (i.e., $\boldsymbol{\tau}_{K,F} \times \boldsymbol{n}_{K|F}$ points outward F). The reader is invited to verify (see Exercise 17.4) that for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{c}}(K)$ and every integer $m \in \{1:n_{\mathrm{sh}}^{\mathrm{e}}\}$,

$$\sigma_{E,m}^{\mathrm{e}}(\boldsymbol{v}) = \epsilon_{K,F,E} \bigg(\int_K (\nabla \times \boldsymbol{v}) \cdot \nabla L_E^K(\mu_m \circ \boldsymbol{T}_{K,E}^{-1}) \, \mathrm{d}x$$
$$+ \int_F (\boldsymbol{v} \times \boldsymbol{n}_{K|F}) \cdot \nabla L_E^F(\mu_m \circ \boldsymbol{T}_{K,E}^{-1}) \, \mathrm{d}s \bigg), \qquad (17.28)$$

where $L_E^F(\phi)(\boldsymbol{y}) := L_{\widehat{E}}^{\widehat{F}}(\phi \circ \boldsymbol{T}_{K|\widehat{E}})(\boldsymbol{T}_{K|F}^{-1}(\boldsymbol{y}))$ for all $\phi \in W^{\frac{1}{p},p'}(E)$ and $\boldsymbol{y} \in F$, $L_E^K(\phi)(\boldsymbol{x}) := L_{\widehat{E}}^{\widehat{K}}(\phi \circ \boldsymbol{T}_{K|\widehat{E}})(\boldsymbol{T}_K^{-1}(\boldsymbol{x}))$ for all $\boldsymbol{x} \in K$, and $\boldsymbol{T}_{K,E} := \boldsymbol{T}_{K|\widehat{E}} \circ \boldsymbol{T}_{\widehat{E}}$. If $\boldsymbol{v} \in \mathbb{N}_{k,d}$, and more generally, if $\boldsymbol{v} \in \boldsymbol{W}^{s,p}(K) \cap \boldsymbol{V}^{\mathrm{c}}(K)$ with $sp > 2$, $p \in (1, \infty)$ or $s = 2$, $p = 1$, we recover that $\sigma_{E,m}^{\mathrm{e}}(\boldsymbol{v}) = \frac{1}{|E|} \int_E (\boldsymbol{v} \cdot \boldsymbol{t}_E)(\mu_m \circ \boldsymbol{T}_{K,E}^{-1}) \, \mathrm{d}l$, where $\boldsymbol{t}_E := |E|\boldsymbol{\tau}_E$ is the tangent vector orienting E. Concerning the extended local face and cell dofs, it follows from Lemma 15.19 that $\sigma_{F,j,m}^{\mathrm{f}}(\boldsymbol{v}) = \frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{t}_{F,j})(\zeta_m \circ \boldsymbol{T}_{K,F}^{-1}) \, \mathrm{d}s$ for all $F \in \mathcal{F}_K$ and every integers $j \in \{1,2\}$ and $m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}$, where $\{\boldsymbol{t}_{F,j}\}_{j \in \{1,2\}}$ are the two tangent vectors orienting F and $\boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}}$, and that $\sigma_{j,m}^{\mathrm{c}}(\boldsymbol{v}) = \frac{1}{|K|} \int_K (\boldsymbol{v} \cdot \boldsymbol{t}_{K,j})(\psi_m \circ \boldsymbol{T}_K^{-1}) \, \mathrm{d}x$ for every integers $j \in \{1:3\}$ and $m \in \{1:n_{\mathrm{sh}}^{\mathrm{c}}\}$, where $\{\boldsymbol{t}_{K,j}\}_{j \in \{1:3\}}$ are the three tangent vectors orienting K. □

Using the extended dofs to define $\mathcal{I}_K^{\mathrm{c}} : \boldsymbol{V}^{\mathrm{c}}(K) \to \mathbb{N}_{k,d}$, we can now derive an estimate of $\|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)}$ for $\boldsymbol{v} \in \boldsymbol{H}^r(K)$ with $r \in (\frac{1}{2}, 1]$. This result complements Theorem 16.12 which is valid for $r > 1$.

Theorem 17.11 (Approximation, $r \in (\frac{1}{2}, 1]$). Let $r \in (\frac{1}{2}, 1]$ and let $p \in (2, \frac{4}{3-2r}]$. There is c, unbounded as $r \downarrow \frac{1}{2}$ (i.e., $p \downarrow 2$), such that

$$\|\boldsymbol{v} - \mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} \leq c \Big(h_K^r |\boldsymbol{v}|_{\boldsymbol{H}^r(K)} + h_K^{1+3(\frac{1}{2}-\frac{1}{p})} \|\nabla \times \boldsymbol{v}\|_{\boldsymbol{L}^p(K)} \Big), \quad (17.29)$$

for all $\boldsymbol{v} \in \boldsymbol{H}^r(K)$ with $\nabla \times \boldsymbol{v} \in \boldsymbol{L}^p(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. See Exercise 17.3. □

Remark 17.12 (Literature). The space $\boldsymbol{V}^{\mathrm{c}}(K)$ defined in (17.21) has been introduced in Amrouche et al. [9, Lem. 4.7], and Theorem 17.11 is established in Boffi and Gastaldi [28]. One can also extend the dofs of the $\mathbb{N}_{k,d}$ finite element to $\{\boldsymbol{v} \in \boldsymbol{H}^r(K) \mid \nabla{\times}\boldsymbol{v} \in \boldsymbol{H}^s(K)\}$ with suitable smoothness indices r, s. See, e.g., Monk [144, Lem. 2.3] for $r = s = 1$, Alonso and Valli [8, §5] and Ciarlet and Zou [74, §3] for $r = s \in (\frac{1}{2}, 1]$, and Bermúdez et al. [19, Lem. 5.1] for $r \in (\frac{1}{2}, 1]$, $s \in (0, 1]$. $\qquad\square$

Remark 17.13 (Quasi-interpolation). We construct in Chapter 22 a quasi-interpolation operator $\mathcal{I}_h^{\mathrm{c,av}}$ s.t. $\|\boldsymbol{v} - \mathcal{I}_h^{\mathrm{c,av}}(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} \leq ch_K^r |\boldsymbol{v}|_{\boldsymbol{H}^r(D_K)}$ for all $\boldsymbol{v} \in \boldsymbol{H}^r(D_K)$, $r \in (0, 1]$, where D_K is a local neighborhood of K (see Theorem 22.6). Thus, contrary to the canonical interpolation operator $\mathcal{I}_h^{\mathrm{c}}$, the quasi-interpolation operator $\mathcal{I}_h^{\mathrm{c,av}}$ gives an optimal error estimate for all $r > 0$ without making any assumption on the curl of the vector field that is approximated. We are going to invoke $\mathcal{I}_h^{\mathrm{c,av}}$ instead of $\mathcal{I}_h^{\mathrm{c}}$ most of the time in the rest of the book. $\qquad\square$

Exercises

Exercise 17.1 (Lifting). Let $D := (0,1)^2$. Let $\boldsymbol{x} := (x_1, x_2)^{\mathsf{T}}$ and consider the function $\phi(\boldsymbol{x}) := \frac{x_1}{\sqrt{x_1^2 + x_2^2}}$. (i) Compute $\lim_{x_1 \downarrow 0} \phi(\boldsymbol{x})$ and $\lim_{x_2 \downarrow 0} \phi(\boldsymbol{x})$. (ii) Without invoking a trace argument, prove directly that $\phi \notin H^1(D)$. (iii) Construct a function $\psi \in C^\infty(D; [0,1])$ s.t. $\lim_{x_1 \downarrow 0} \psi(\boldsymbol{x}) = 0$, $\lim_{x_2 \uparrow 1} \psi(\boldsymbol{x}) = 0$, $\lim_{x_1 \uparrow 1} \psi(\boldsymbol{x}) = 0$, and $\lim_{x_2 \downarrow 0} \psi(\boldsymbol{x}) = 1$.

Exercise 17.2 (Extended face dofs for $\mathbb{RT}_{k,d}$). (i) Let $\epsilon_{K,F} := \boldsymbol{n}_F{\cdot}\boldsymbol{n}_{K|F}$, $\epsilon_{\widehat{K},\widehat{F}} := \boldsymbol{n}_{\widehat{F}}{\cdot}\boldsymbol{n}_{\widehat{K}|\widehat{F}}$, and $\epsilon_K := \det(\mathbb{J}_K)/|\det(\mathbb{J}_K)|$. Prove that $\epsilon_{K,F} = \epsilon_{\widehat{K},\widehat{F}}\epsilon_K$. (ii) Prove (17.17). (*Hint*: show that $L_F^K(\zeta_m \circ \boldsymbol{T}_{K,F}^{-1}) = L_{\widehat{F}}^{\widehat{K}}(\zeta_m \circ \boldsymbol{T}_{\widehat{F}}^{-1}) \circ \boldsymbol{T}_K^{-1}$ and use (9.8a).)

Exercise 17.3 ($\mathcal{I}_K^{\mathrm{c}}$). (i) Let $r > \frac{1}{2}$ and $p \in (2, \frac{4}{3-2r}]$. Prove the stability estimate $\|\mathcal{I}_K^{\mathrm{c}}(\boldsymbol{v})\|_{\boldsymbol{L}^2(K)} \leq c\left(\|\boldsymbol{v}\|_{\boldsymbol{L}^2(K)} + h_K^r|\boldsymbol{v}|_{\boldsymbol{H}^r(K)} + h_K^{1+3(\frac{1}{2}-\frac{1}{p})}\|\nabla{\times}\boldsymbol{v}\|_{\boldsymbol{L}^p(K)}\right)$ for all $\boldsymbol{v} \in \boldsymbol{V}^{\mathrm{c}}(K)$. (*Hint*: use the trace theorem (Theorem 3.10), the Sobolev embedding theorem (Theorem 2.31), and reason as in the proof of Theorem 17.5.) (ii) Prove Theorem 17.11. (*Hint*: proceed as in the proof of Theorem 17.5.)

Exercise 17.4 (Extended edge dofs for $\mathbb{N}_{k,d}$). Use the notation from Remark 17.10. (i) Let $\boldsymbol{w} \in \boldsymbol{C}^1(K)$ be a smooth function. Prove that $\epsilon_{K,F,E} = \epsilon_K \epsilon_{\widehat{K},\widehat{F},\widehat{E}}$ where $\epsilon_K := \det(\mathbb{J}_K)/|\det(\mathbb{J}_K)|$. (*Hint*: apply the Kelvin–Stokes formula (16.15) to the shape function of the lowest-order Nédélec element associated with E). (ii) Prove (17.28). (*Hint*: proceed as in Exercise 17.2(ii) and use (9.8b).)

Part IV
Finite element spaces

Part IV, Chapter 18

From broken to conforming spaces

In Parts II and III, we have introduced many examples of finite elements and devised techniques to generate finite elements in each cell of a mesh. In Part IV, composed of Chapters 18 to 23, we show how these methods can be used to build finite-dimensional spaces composed of piecewise smooth functions whose gradient, curl, or divergence is integrable. We also devise quasi-interpolation operators enjoying fundamental stability, approximation, and commutation properties. These spaces and operators will be used repeatedly in Volumes II and III to approximate various PDEs and estimate the approximation error. In the present chapter, we introduce broken Sobolev spaces and broken finite element spaces based on a mesh from a family of meshes $(\mathcal{T}_h)_{h\in\mathcal{H}}$ covering exactly a domain $D \subsetneq \mathbb{R}^d$. Then we identify jump conditions across the mesh interfaces that are necessary and sufficient for every function in some broken Sobolev space to have an integrable gradient, curl, or divergence. These conditions lead to the notion of conforming finite element spaces. Finally, we show how to construct L^1-stable (local) interpolation operators in the broken finite element space with optimal local approximation properties.

18.1 Broken spaces and jumps

In this section, we are only concerned with broken Sobolev spaces and with broken finite element spaces. Membership to broken spaces is defined by requiring that some property be satisfied in each mesh cell without requiring any continuity across the mesh interfaces.

18.1.1 Broken Sobolev spaces and jumps

The notions introduced hereafter will be used repeatedly in this book. We consider $\mathbb{R}^q$-valued functions for some integer $q \geq 1$.

Definition 18.1 (Broken Sobolev space). *Let $p \in [1, \infty]$ and $s > 0$ be a positive real number. The space defined by*

$$W^{s,p}(\mathcal{T}_h; \mathbb{R}^q) := \{v \in L^p(D; \mathbb{R}^q) \mid v_{|K} \in W^{s,p}(K; \mathbb{R}^q), \forall K \in \mathcal{T}_h\}, \quad (18.1)$$

is called broken Sobolev space. *This space is equipped with the norm*

$$\|v\|_{W^{s,p}(\mathcal{T}_h; \mathbb{R}^q)}^p := \sum_{K \in \mathcal{T}_h} \|v\|_{W^{s,p}(K; \mathbb{R}^q)}^p, \quad (18.2)$$

if $p \in [1, \infty)$ and $\|v\|_{W^{s,\infty}(\mathcal{T}_h; \mathbb{R}^q)} := \max_{K \in \mathcal{T}_h} \|v\|_{W^{s,\infty}(K; \mathbb{R}^q)}$ if $p = \infty$. We write $W^{s,p}(\mathcal{T}_h) := W^{s,p}(\mathcal{T}_h; \mathbb{R})$ when $q = 1$.

An important notion in broken Sobolev spaces is the jump of functions across mesh interfaces (see Figure 18.1). Recall from the Definition 8.10 that the collection of the mesh interfaces is denoted by $\mathcal{F}_h^\circ$ and that for all $F \in \mathcal{F}_h^\circ$, there are two distinct mesh cells $K_l, K_r \in \mathcal{T}_h$ such that $F = \partial K_l \cap \partial K_r$. The interface F is oriented by means of the unit normal vector $\boldsymbol{n}_F$ pointing from K_l to K_r.

Definition 18.2 (Jump). *Let $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$ be a mesh interface. Let $v \in W^{s,p}(\mathcal{T}_h; \mathbb{R}^q)$ with $s > \frac{1}{p}$ if $p \in (1, \infty)$ or $s \geq 1$ if $p = 1$ (notice that $(v_{|K_l})_{|F} \in L^1(F)$ and $(v_{|K_r})_{|F} \in L^1(F)$). The* jump *of v across F is defined as follows a.e. in F:*

$$[\![v]\!]_F := v_{|K_l} - v_{|K_r}. \quad (18.3)$$

The subscript F is dropped when the context is unambiguous.
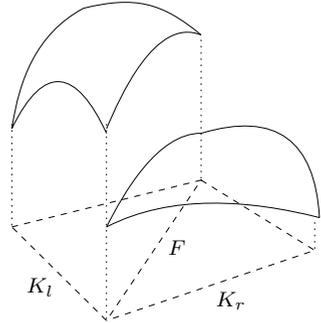
Fig. 18.1 Jump of a piecewise smooth function across the interface $F := \partial K_l \cap \partial K_r$.

Remark 18.3 (Alternative definition). Another definition of the jump where K_l, K_r play symmetric roles consists of setting $[\![v]\!]_F^* := v_{|K_l} \otimes \boldsymbol{n}_{K_l|F} + v_{|K_r} \otimes \boldsymbol{n}_{K_r|F}$, where $\boldsymbol{n}_{K_i|F}$, $i \in \{l, r\}$, is the unit normal to F pointing away from K_i, i.e., $[\![v]\!]_F \otimes \boldsymbol{n}_F = [\![v]\!]_F^*$. The advantage of (18.3) over this definition is that the jump $[\![v]\!]_F$ is $\mathbb{R}^q$-valued instead of being $\mathbb{R}^{q \times d}$-valued. Both definitions are commonly used in the literature. $\square$

Remark 18.4 (Zero-jumps in $W^{s,p}$). Let $p \in (1,\infty)$ and $s > \frac{1}{p}$, or $p = 1$ and $s \geq \frac{1}{p}$. Owing to Theorem 2.21, smooth functions are dense in $W^{s,p}(D)$. Let $v \in W^{s,p}(D)$ and let $(v_n)_{n \in \mathbb{N}}$ be a sequence in $C^\infty(D) \cap W^{s,p}(D)$ converging to v in $W^{s,p}(D)$. Let $F \in \mathcal{F}_h^\circ$ be a mesh interface. Then $0 = [\![v_n]\!]_F \to [\![v]\!]_F$ as $n \to \infty$ since the trace map is bounded on $W^{s,p}(D)$. Hence, $0 = [\![v]\!]_F$ for all $F \in \mathcal{F}_h^\circ$. This shows that functions in $W^{s,p}(D)$ have a single-valued trace in $L^1(F)$ for all $F \in \mathcal{F}_h^\circ$. $\qquad\square$

18.1.2 Broken finite element spaces

Let $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ be the reference finite element of degree $k \geq 0$, where $\widehat{P}$ is composed of $\mathbb{R}^q$-valued functions for some integer $q \geq 1$. We assume that $\widehat{P} \subset L^\infty(\widehat{K}; \mathbb{R}^q)$ (this is a mild assumption since in general $\widehat{P}$ is composed of polynomial functions). Consider a $\mathcal{T}_h$-based family of finite elements $\{(K, P_K, \Sigma_K)\}_{K \in \mathcal{T}_h}$ constructed as in Proposition 9.2 by using the geometric mappings $\boldsymbol{T}_K : \widehat{K} \to K$ and the transformations $\psi_K : V(K) \to V(\widehat{K})$ for all $K \in \mathcal{T}_h$. We assume henceforth that $\psi_K \in \mathcal{L}(L^\infty(K; \mathbb{R}^q), L^\infty(\widehat{K}; \mathbb{R}^q))$. Recall that we denote by $\{\theta_{K,i}\}_{i \in \mathcal{N}}$ the local shape functions in K and by $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$ the local degrees of freedom (dofs).

Definition 18.5 (Broken finite element space). *The* broken finite element space *is defined as follows:*

$$P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q) := \{v_h \in L^\infty(D; \mathbb{R}^q) \mid \psi_K(v_{h|K}) \in \widehat{P}, \forall K \in \mathcal{T}_h\}. \qquad (18.4)$$

We simply write $P_k^{\mathrm{b}}(\mathcal{T}_h)$ whenever $q = 1$.

Recalling that $P_K := \psi_K^{-1}(\widehat{P})$ (see (9.4a)), we have $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$ iff $v_{h|K} \in P_K$ for all $K \in \mathcal{T}_h$. The above assumptions on $\widehat{P}$ and ψ_K imply that $P_K \subset L^\infty(K; \mathbb{R}^q)$, which in turn means that $P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$ is indeed a subspace of $L^\infty(D; \mathbb{R}^q)$. Moreover, since functions in $P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$ can be defined independently in each mesh cell, we have

$$\dim(P^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)) = \mathrm{card}(\mathcal{N}) \times \mathrm{card}(\mathcal{T}_h) =: n_{\mathrm{sh}} \times N_{\mathrm{c}}, \qquad (18.5)$$

where n_{sh} is the number of dofs in $\widehat{\Sigma}$ (i.e., the cardinality of the set $\mathcal{N}$), and N_{c} is the number of mesh cells in $\mathcal{T}_h$. Then the set $\{\tilde{\theta}_{K,i}\}_{(K,i) \in \mathcal{T}_h \times \mathcal{N}}$, where $\tilde{\theta}_{K,i}$ is the zero-extension of $\theta_{K,i}$ to D, is a basis of $P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$. The functions $\tilde{\theta}_{K,i}$ are called *global shape functions* in $P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$.

Example 18.6 (Piecewise polynomials). On affine meshes the choice $\widehat{P} := \mathbb{P}_{k,d}$ (resp., $\widehat{P} := \mathbb{Q}_{k,d}$) together with $\psi_K(v) := v \circ \boldsymbol{T}_K$ and $q := 1$ (i.e., scalar-valued functions) leads to $P_k^{\mathrm{b}}(\mathcal{T}_h) = \{v_h \in L^\infty(D) \mid v_{h|K} \in \mathbb{P}_{k,d}, \forall K \in \mathcal{T}_h\}$ (resp., $\{v_h \in L^\infty(D) \mid v_{h|K} \in \mathbb{Q}_{k,d}, \forall K \in \mathcal{T}_h\}$) since $v_{h|K} \in \mathbb{P}_{k,d}$ iff $v_h \circ \boldsymbol{T}_K \in \mathbb{P}_{k,d}$ (resp., $\mathbb{Q}_{k,d}$). $\qquad\square$

Remark 18.7 (Connectivity array). In practice, the global shape functions are enumerated, say from 1 to I. For the broken finite element space, we have $P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q) = \mathrm{span}\{\varphi_1, \ldots, \varphi_I\}$ with $I = n_{\mathrm{sh}} N_{\mathrm{c}}$. The connection between the local and the global shape functions is materialized by a connectivity array $\mathtt{j\_dof} : \{1 : N_{\mathrm{c}}\} \times \mathcal{N} \to \{1 : I\}$ defined such that $\varphi_{\mathtt{j\_dof}(m,n)|_{K_m}} := \theta_{K_m, i}$ for all $m \in \{1 : N_{\mathrm{c}}\}$ and all $n \in \mathcal{N}$. The most common approach to define $\mathtt{j\_dof}$ consists of enumerating first the dofs in the first cell, then in the second cell, and so on, leading to $\mathtt{j\_dof}(m,n) := (m-1)n_{\mathrm{sh}} + n$. $\square$

18.2 Conforming finite element subspaces

Given a piecewise smooth function on the mesh $\mathcal{T}_h$, either scalar- or vector-valued, depending on the context, we want to find necessary and sufficient conditions for this function to be in $H^1(D)$, $\boldsymbol{H}(\mathrm{curl}; D)$, or $\boldsymbol{H}(\mathrm{div}; D)$. It turns out that the answer to this question hinges on the continuity properties of the function, its normal component, or its tangential component across the mesh interfaces.

18.2.1 Membership in H^1

The global integrability of the gradient of a piecewise smooth function is characterized by the following result.

Theorem 18.8 (Integrability of ∇). *Let $v \in W^{1,p}(\mathcal{T}_h; \mathbb{R}^q)$ with $p \in [1, \infty]$. Then $\nabla v \in \boldsymbol{L}^p(D)$ iff $[\![v]\!]_F = 0$ a.e. on all $F \in \mathcal{F}_h^{\circ}$.*

Proof. We prove the assertion for $q = 1$. The general case is treated by working componentwise. Let $v \in W^{1,p}(\mathcal{T}_h)$ and let $C_0^{\infty}(D)$ be the set of the smooth functions compactly supported in D. For all $\boldsymbol{\Phi} \in \boldsymbol{C}_0^{\infty}(D)$, we have

$$
\int_D v \nabla \cdot \boldsymbol{\Phi} \, \mathrm{d}x = \sum_{K \in \mathcal{T}_h} \int_K v_{|K} \nabla \cdot \boldsymbol{\Phi} \, \mathrm{d}x
$$

$$
= - \sum_{K \in \mathcal{T}_h} \int_K \nabla(v_{|K}) \cdot \boldsymbol{\Phi} \, \mathrm{d}x + \sum_{K \in \mathcal{T}_h} \int_{\partial K} v_{|K} \boldsymbol{n}_K \cdot \boldsymbol{\Phi} \, \mathrm{d}s
$$

$$
= - \sum_{K \in \mathcal{T}_h} \int_K \nabla(v_{|K}) \cdot \boldsymbol{\Phi} \, \mathrm{d}x + \sum_{F \in \mathcal{F}_h^{\circ}} \int_F [\![v]\!]_F \boldsymbol{n}_F \cdot \boldsymbol{\Phi} \, \mathrm{d}s,
$$

where $\boldsymbol{n}_K$ is the outward unit normal to K and $\boldsymbol{n}_F$ is the unit vector defining the orientation of F.
(i) If $[\![v]\!]_F = 0$ a.e. on all $F \in \mathcal{F}_h^{\circ}$, we infer from the above identity that $\int_D v \nabla \cdot \boldsymbol{\Phi} \, \mathrm{d}x = - \sum_{K \in \mathcal{T}_h} \int_K \boldsymbol{\Phi} \cdot \nabla(v_{|K}) \, \mathrm{d}x$, which shows that v has a weak gradient in $\boldsymbol{L}^p(D)$ s.t. $(\nabla v)_{|K} = \nabla(v_{|K})$ for all $K \in \mathcal{T}_h$. Hence, $v \in W^{1,p}(D)$.
(ii) Conversely let $v \in W^{1,p}(D)$. We can conclude by invoking Remark 18.4.

Let us give a more direct proof. Owing to Lemma 18.9 below, we infer that $(\nabla v)_{|K} = \nabla(v_{|K})$ for all $K \in \mathcal{T}_h$. Hence, the above identity implies that $\sum_{F \in \mathcal{F}_h^\circ} \int_F [\![v]\!]_F \boldsymbol{n}_F \cdot \boldsymbol{\Phi} \, \mathrm{d}s = 0$ for all $\boldsymbol{\Phi} \in \boldsymbol{C}_0^\infty(D)$. Let $F \in \mathcal{F}_h^\circ$ be an arbitrary interface. After localizing the support of $\boldsymbol{\Phi}$ in such a way that it intersects F and no other interface in $\mathcal{F}_h^\circ$, it follows from the vanishing integral theorem (Theorem 1.32) that $[\![v]\!]_F = 0$, since $\boldsymbol{\Phi}_{|F} \cdot \boldsymbol{n}_F$ can be arbitrarily chosen, and $[\![v]\!]_F \in L^1(F)$ because the trace map is bounded on $W^{1,p}(D)$. $\qquad\square$

Lemma 18.9 (Local weak derivative). *Let $p \in [1, \infty]$ and let $v \in W^{1,p}(D)$. Then $\nabla(v_{|K}) = (\nabla v)_{|K}$ a.e. in K for all $K \in \mathcal{T}_h$.*

Proof. Let $K \in \mathcal{T}_h$ and let $\phi \in \boldsymbol{C}_0^\infty(K)$. Let $\widetilde{\phi} \in \boldsymbol{C}_0^\infty(D)$ be the zero-extension of ϕ to D. For all $v \in W^{1,p}(D)$, we infer that

$$
\int_K \nabla(v_{|K}) \cdot \boldsymbol{\phi} \, \mathrm{d}x = -\int_K v_{|K} \nabla \cdot \boldsymbol{\phi} \, \mathrm{d}x
$$
$$
= -\int_D v \nabla \cdot \widetilde{\boldsymbol{\phi}} \, \mathrm{d}x = \int_D \nabla v \cdot \widetilde{\boldsymbol{\phi}} \, \mathrm{d}x = \int_K (\nabla v)_{|K} \cdot \boldsymbol{\phi} \, \mathrm{d}x.
$$

The assertion follows from Theorem 1.32 since ϕ is arbitrary in $\boldsymbol{C}_0^\infty(K)$. $\quad\square$

Figure 18.2 illustrates Theorem 18.8 in dimension one.
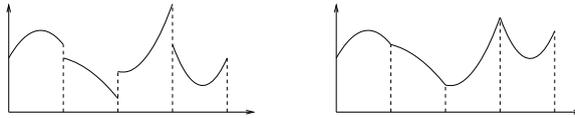
Fig. 18.2 One-dimensional example with two piecewise quadratic functions. The one on the left is not in H^1, the one on the right is.

18.2.2 Membership in $\boldsymbol{H}(\mathrm{curl})$ and $\boldsymbol{H}(\mathrm{div})$

Let us now consider the integrability of the curl or the divergence of vector-valued piecewise smooth functions. Let $\boldsymbol{v} \in \boldsymbol{W}^{1,p}(\mathcal{T}_h) := W^{1,p}(\mathcal{T}_h; \mathbb{R}^d)$, $p \in [1, \infty)$. We also use the notation $\boldsymbol{W}^{s,p}(\mathcal{T}_h) := W^{s,p}(\mathcal{T}_h; \mathbb{R}^d)$, $s > 0$. The jump of the tangential component of $\boldsymbol{v}$ (if $d = 3$) and the jump of its normal component across a mesh interface $F \in \mathcal{F}_h^\circ$, with $F := \partial K_l \cap \partial K_r$, are defined as follows a.e. in F:

$$
[\![\boldsymbol{v} \times \boldsymbol{n}]\!]_F := (\boldsymbol{v}_{|K_l} \times \boldsymbol{n}_F) - (\boldsymbol{v}_{|K_r} \times \boldsymbol{n}_F) = [\![\boldsymbol{v}]\!]_F \times \boldsymbol{n}_F, \tag{18.6a}
$$
$$
[\![\boldsymbol{v} \cdot \boldsymbol{n}]\!]_F := (\boldsymbol{v}_{|K_l} \cdot \boldsymbol{n}_F) - (\boldsymbol{v}_{|K_r} \cdot \boldsymbol{n}_F) = [\![\boldsymbol{v}]\!]_F \cdot \boldsymbol{n}_F, \tag{18.6b}
$$

where $[\![\boldsymbol{v}]\!]_F$ is the componentwise jump of $\boldsymbol{v}$ across F from Definition 18.2. The subscript F is dropped when the context is unambiguous.

Theorem 18.10 (Integrability of $\nabla\times$ and $\nabla\cdot$). *Let $\boldsymbol{v} \in \boldsymbol{W}^{1,p}(\mathcal{T}_h)$ with $p \in [1,\infty]$. (i) If $d = 3$, $\nabla\times\boldsymbol{v} \in \boldsymbol{L}^p(D)$ if and only if $[\![\boldsymbol{v}\times\boldsymbol{n}]\!]_F = 0$ a.e. on all $F \in \mathcal{F}_h^\circ$. (ii) $\nabla\cdot\boldsymbol{v} \in L^p(D)$ if and only if $[\![\boldsymbol{v}\cdot\boldsymbol{n}]\!]_F = 0$ a.e. on all $F \in \mathcal{F}_h^\circ$.*

Proof. Proceed as in the proof of Theorem 18.8. See Exercise 18.1. □

Remark 18.11 (Extension). The statement of Theorem 18.10 can be extended to functions $\boldsymbol{v} \in \boldsymbol{W}^{s,p}(\mathcal{T}_h)$ with $s > \frac{1}{p}$ if $p \in (1,\infty)$ or $s \geq 1$ if $p = 1$. The following holds true: (i) If $d = 3$ and $\nabla\times(\boldsymbol{v}_{|K}) \in \boldsymbol{L}^p(K)$ for all $K \in \mathcal{T}_h$, then $\nabla\times\boldsymbol{v} \in \boldsymbol{L}^p(D)$ iff $[\![\boldsymbol{v}\times\boldsymbol{n}]\!]_F = 0$ for all $F \in \mathcal{F}_h^\circ$. (ii) If $\nabla\cdot(\boldsymbol{v}_{|K}) \in L^p(K)$ for all $K \in \mathcal{T}_h$, then $\nabla\cdot\boldsymbol{v} \in L^p(D)$ iff $[\![\boldsymbol{v}\cdot\boldsymbol{n}]\!]_F = 0$ for all $F \in \mathcal{F}_h^\circ$. □

18.2.3 Unified notation for conforming subspaces

To allow for a unified treatment of H^1-, $\boldsymbol{H}(\mathrm{curl})$-, and $\boldsymbol{H}(\mathrm{div})$-conformity, we use the superscript $\mathrm{x} \in \{\mathrm{g}, \mathrm{c}, \mathrm{d}\}$ (referring to the gradient, curl, and divergence operators), and we consider $\mathbb{R}^q$-valued functions with $q := 1$ if $\mathrm{x} = \mathrm{g}$, $q = d = 3$ if $\mathrm{x} = \mathrm{c}$, and $q = d$ if $\mathrm{x} = \mathrm{d}$. Let $p \in [1,\infty)$ and let $s > \frac{1}{p}$ if $p > 1$ or $s \geq 1$ if $p = 1$. Let $K \in \mathcal{T}_h$ be a mesh cell and let $F \in \mathcal{F}_K$ be a face of K. We define the local trace operators $\gamma_{K,F}^{\mathrm{x}} : W^{s,p}(K;\mathbb{R}^q) \to L^1(F;\mathbb{R}^t)$ s.t.

$$\gamma_{K,F}^{\mathrm{g}}(v) := v_{|F} \qquad (q = t = 1), \tag{18.7a}$$

$$\gamma_{K,F}^{\mathrm{c}}(\boldsymbol{v}) := \boldsymbol{v}_{|F}\times\boldsymbol{n}_F \quad (q = t = d = 3), \tag{18.7b}$$

$$\gamma_{K,F}^{\mathrm{d}}(\boldsymbol{v}) := \boldsymbol{v}_{|F}\cdot\boldsymbol{n}_F \quad (q = d,\ t = 1). \tag{18.7c}$$

This leads to the following notion of γ-jump: For all $v \in W^{s,p}(\mathcal{T}_h;\mathbb{R}^q)$,

$$[\![v]\!]_F^{\mathrm{x}}(\boldsymbol{x}) := \gamma_{K_l,F}^{\mathrm{x}}(v_{|K_l})(\boldsymbol{x}) - \gamma_{K_r,F}^{\mathrm{x}}(v_{|K_r})(\boldsymbol{x}) \qquad \text{a.e. on } F. \tag{18.8}$$

Let $(\widehat{K}, \widehat{P}^{\mathrm{g}}, \Sigma^{\mathrm{g}})$ be one of the Lagrange elements or the canonical hybrid element introduced in Chapters 6 and 7. Let $k \geq 1$ be the degree of the finite element. The corresponding broken finite element space is

$$P_k^{\mathrm{g,b}}(\mathcal{T}_h) := \{v_h \in L^\infty(D) \mid \psi_K^{\mathrm{g}}(v_{h|K}) \in \widehat{P}^{\mathrm{g}}, \forall K \in \mathcal{T}_h\}, \tag{18.9}$$

where $\psi_K^{\mathrm{g}}(v) := v \circ \boldsymbol{T}_K$ is the pullback by the geometric mapping $\boldsymbol{T}_K$. The H^1-conforming finite element subspace is defined as follows:

$$P_k^{\mathrm{g}}(\mathcal{T}_h) := P_k^{\mathrm{g,b}}(\mathcal{T}_h) \cap H^1(D). \tag{18.10}$$

Similarly, let $(\widehat{K}, \widehat{\boldsymbol{P}}^{\mathrm{c}}, \Sigma^{\mathrm{c}})$ be one of the Nédélec elements introduced in Chapter 15, and let $(\widehat{K}, \widehat{\boldsymbol{P}}^{\mathrm{d}}, \Sigma^{\mathrm{d}})$ be one of the Raviart–Thomas elements introduced in Chapter 14. Let $k \geq 0$ be the degree of the finite element. The corresponding broken finite element spaces are

$$\boldsymbol{P}_k^{\mathrm{c,b}}(\mathcal{T}_h) := \{\boldsymbol{v}_h \in \boldsymbol{L}^\infty(D) \mid \boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}_{h|K}) \in \widehat{\boldsymbol{P}}^{\mathrm{c}}, \ \forall K \in \mathcal{T}_h\}, \qquad (18.11\mathrm{a})$$

$$\boldsymbol{P}_k^{\mathrm{d,b}}(\mathcal{T}_h) := \{\boldsymbol{v}_h \in \boldsymbol{L}^\infty(D) \mid \boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}_{h|K}) \in \widehat{\boldsymbol{P}}^{\mathrm{d}}, \ \forall K \in \mathcal{T}_h\}, \qquad (18.11\mathrm{b})$$

where $\boldsymbol{\psi}_K^{\mathrm{c}}(\boldsymbol{v}) := \mathbb{J}_K^{\mathsf{T}}(\boldsymbol{v} \circ \boldsymbol{T}_K)$ is the covariant Piola transformation and $\boldsymbol{\psi}_K^{\mathrm{d}}(\boldsymbol{v}) := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}(\boldsymbol{v} \circ \boldsymbol{T}_K)$ is the contravariant Piola transformation. The corresponding $\boldsymbol{H}(\mathrm{curl})$- and $\boldsymbol{H}(\mathrm{div})$-conforming finite element subspaces are defined as follows:

$$\boldsymbol{P}_k^{\mathrm{c}}(\mathcal{T}_h) := \boldsymbol{P}_k^{\mathrm{c,b}}(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{curl}; D), \qquad (18.12\mathrm{a})$$

$$\boldsymbol{P}_k^{\mathrm{d}}(\mathcal{T}_h) := \boldsymbol{P}_k^{\mathrm{d,b}}(\mathcal{T}_h) \cap \boldsymbol{H}(\mathrm{div}; D). \qquad (18.12\mathrm{b})$$

The zero-jump conditions from Theorem 18.8 and Theorem 18.10 imply that

$$P_k^{\mathrm{g}}(\mathcal{T}_h) = \{v_h \in P_k^{\mathrm{g,b}}(\mathcal{T}_h) \mid [\![v_h]\!]_F^{\mathrm{g}} = 0, \ \forall F \in \mathcal{F}_h^\circ\}, \qquad (18.13\mathrm{a})$$

$$\boldsymbol{P}_k^{\mathrm{c}}(\mathcal{T}_h) = \{\boldsymbol{v}_h \in \boldsymbol{P}_k^{\mathrm{c,b}}(\mathcal{T}_h) \mid [\![\boldsymbol{v}_h]\!]_F^{\mathrm{c}} = \boldsymbol{0}, \ \forall F \in \mathcal{F}_h^\circ\}, \qquad (18.13\mathrm{b})$$

$$\boldsymbol{P}_k^{\mathrm{d}}(\mathcal{T}_h) = \{\boldsymbol{v}_h \in \boldsymbol{P}_k^{\mathrm{d,b}}(\mathcal{T}_h) \mid [\![\boldsymbol{v}_h]\!]_F^{\mathrm{d}} = 0, \ \forall F \in \mathcal{F}_h^\circ\}. \qquad (18.13\mathrm{c})$$

In the next chapters, we study the construction and the interpolation properties of the above conforming finite element subspaces. To stay general, we employ the following unified notation with $\mathrm{x} \in \{\mathrm{g,c,d}\}$:

$$P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q) := \{v_h \in \boldsymbol{P}_k^{\mathrm{x,b}}(\mathcal{T}_h; \mathbb{R}^q) \mid [\![v_h]\!]_F^{\mathrm{x}} = 0, \ \forall F \in \mathcal{F}_h^\circ\}, \qquad (18.14)$$

where $\boldsymbol{P}_k^{\mathrm{x,b}}(\mathcal{T}_h; \mathbb{R}^q)$ is one of the broken finite element spaces defined above.

Remark 18.12 (2D discrete Sobolev inequality). We have $P_k^{\mathrm{g}}(\mathcal{T}_h) \subset L^\infty(D) \cap H^1(D)$ by construction, but as shown in Example 2.33, if $d \geq 2$, there exist functions in $H^1(D)$ that are unbounded. It turns out that in dimension two, it is possible to derive a bound on the $\|\cdot\|_{L^\infty}$-norm of functions in $P_k^{\mathrm{g}}(\mathcal{T}_h)$ that blows up very mildly w.r.t. the meshsize. This bound involves a global length scale associated with D, say δ_D. More precisely, since D is Lipschitz, one can show that there exist a length scale $\delta_D > 0$ and an angle $\omega \in (0, 2\pi)$ such that any point $\boldsymbol{x} \in D$ is the vertex of a cone $\mathfrak{C}(\boldsymbol{x}) \subset D$, where $\mathfrak{C}(\boldsymbol{x})$ is the image by a translation and rotation of the cone $\mathfrak{C} := \{(r, \theta) \mid r \in (0, \delta_D), \theta \in (0, \omega)\}$ defined in polar coordinates; see Lemma 3.4. Then assuming $d := 2$, one can show (see Exercise 18.2 and Bramble et al. [42]) the following inverse inequality, called *discrete Sobolev inequality*: There is $c > 0$ s.t.

$$c\,\delta_D^{-\frac{1}{2}} \|v_h\|_{L^\infty(K)} \leq \delta_D^{-1} \|v_h\|_{L^2(D)} + \ln\left(\frac{\delta_D}{h_K}\right)^{\frac{1}{2}} \|\nabla v_h\|_{\boldsymbol{L}^2(D)}, \qquad (18.15)$$

for all $v_h \in P_k^{\mathrm{g}}(\mathcal{T}_h)$, all $K \in \mathcal{T}_h$ such that $h_K \leq \frac{1}{2}\delta_D$, and all $h \in \mathcal{H}$. $\qquad \square$

18.3 L^1-stable local interpolation

In this section, we devise a local interpolation operator that is L^1-stable and maps $L^1(D)$ onto the broken finite element space $P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$ defined in (18.4). The construction is local in each mesh cell. The key idea is to extend the dofs of the reference finite element so as to be able to interpolate boundedly all the functions that are in $L^1(D)$.

We assume that the geometric mappings $\boldsymbol{T}_K$ are affine for all $K \in \mathcal{T}_h$, and that all the transformations ψ_K are of the form $\psi_K(v) := \mathbb{A}_K(v \circ \boldsymbol{T}_K)$ (see (11.1)) where $\mathbb{A}_K \in \mathbb{R}^{q \times q}$ satisfies (see (11.12))

$$\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2} \le c\, \|\mathbb{J}_K\|_{\ell^2}\|\mathbb{J}_K^{-1}\|_{\ell^2}, \qquad (18.16)$$

with c uniform w.r.t. $K \in \mathcal{T}_h$ and $h \in \mathcal{H}$, where $\mathbb{J}_K$ is the Jacobian matrix of $\boldsymbol{T}_K$. Let us define the adjoint transformation $\phi_K(w) := \mathbb{B}_K(w \circ \boldsymbol{T}_K)$ where $\mathbb{B}_K := |\det(\mathbb{J}_K)|\mathbb{A}_K^{-\mathsf{T}}$. The terminology is motivated by the following identity:

$$(w, v)_{L^2(K;\mathbb{R}^q)} = (\phi_K(w), \psi_K(v))_{L^2(\widehat{K};\mathbb{R}^q)}, \qquad (18.17)$$

for all $v \in L^p(K; \mathbb{R}^q)$, all $w \in L^{p'}(K; \mathbb{R}^q)$, and all $p \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{p'} = 1$. Indeed, we have

$$
(\phi_K(w), \psi_K(v))_{L^2(\widehat{K};\mathbb{R}^q)} = \int_{\widehat{K}} |\det(\mathbb{J}_K)| (\mathbb{A}_K^{-\mathsf{T}}(w \circ \boldsymbol{T}_K), \mathbb{A}_K(v \circ \boldsymbol{T}_K))_{\ell^2(\mathbb{R}^q)}\, \mathrm{d}\widehat{x}
$$
$$
= \int_K (w \circ \boldsymbol{T}_K, v \circ \boldsymbol{T}_K)_{\ell^2(\mathbb{R}^q)}\, \mathrm{d}x = (w, v)_{L^2(K;\mathbb{R}^q)}.
$$

Moreover, we have $\|\mathbb{B}_K\|_{\ell^2}\|\mathbb{B}_K^{-1}\|_{\ell^2} = \|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2}$ since $\|\mathbb{A}_K^{\mathsf{T}}\|_{\ell^2} = \|\mathbb{A}_K\|_{\ell^2}$.

We first extend the dofs of the reference finite element. Let $\widehat{\rho}_i \in \widehat{P}$ for all $i \in \mathcal{N}$ be such that

$$\frac{1}{|\widehat{K}|}(\widehat{\rho}_i, \widehat{p})_{L^2(\widehat{K};\mathbb{R}^q)} := \widehat{\sigma}_i(\widehat{p}), \qquad \forall \widehat{p} \in \widehat{P}. \qquad (18.18)$$

The function $\widehat{\rho}_i$ is well defined owing to the Riesz–Fréchet theorem (see either Exercise 5.9 or Theorem A.16 applied here in the finite-dimensional space $\widehat{P}$ equipped with the L^2-inner product weighted by $|\widehat{K}|^{-1}$). This leads us to define the extended dofs as follows:

$$\widehat{\sigma}_i^{\sharp}(\widehat{v}) := \frac{1}{|\widehat{K}|}(\widehat{\rho}_i, \widehat{v})_{L^2(\widehat{K};\mathbb{R}^q)}, \qquad \forall \widehat{v} \in L^1(\widehat{K}; \mathbb{R}^q). \qquad (18.19)$$

We then define the interpolation operator s.t. for all $\widehat{\boldsymbol{x}} \in \widehat{K}$,

$$\mathcal{I}_{\widehat{K}}^{\sharp}(\widehat{v})(\widehat{\boldsymbol{x}}) := \sum_{i \in \mathcal{N}} \widehat{\sigma}_i^{\sharp}(\widehat{v})\widehat{\theta}_i(\widehat{\boldsymbol{x}}), \qquad \forall \widehat{v} \in L^1(\widehat{K}; \mathbb{R}^q). \qquad (18.20)$$

We can take $V(\widehat{K}) := L^1(\widehat{K}; \mathbb{R}^q)$ for the domain of $\mathcal{I}_{\widehat{K}}^\sharp$. One can show that $\mathcal{I}_{\widehat{K}}^\sharp$ is actually the L^2-orthogonal projection onto $\widehat{P}$; see Exercise 18.3.

Lemma 18.13 (Invariance and stability). *Let $\mathcal{I}_{\widehat{K}}^\sharp$ be defined in* (18.20). *(i) $\widehat{P}$ is pointwise invariant under $\mathcal{I}_{\widehat{K}}^\sharp$. (ii) $\mathcal{I}_{\widehat{K}}^\sharp$ is L^p-stable for all $p \in [1, \infty]$, i.e., there is $\widehat{c}$ s.t.*

$$\|\mathcal{I}_{\widehat{K}}^\sharp(\widehat{v})\|_{L^p(\widehat{K}; \mathbb{R}^q)} \le \widehat{c} \|\widehat{v}\|_{L^p(\widehat{K}; \mathbb{R}^q)}, \qquad \forall \widehat{v} \in L^p(\widehat{K}; \mathbb{R}^q). \qquad (18.21)$$

Proof. (i) Since $\widehat{\sigma}_i^\sharp(\widehat{p}) = \widehat{\sigma}_i(\widehat{p})$ for all $\widehat{p} \in \widehat{P}$ and all $i \in \mathcal{N}$, we obtain $\mathcal{I}_{\widehat{K}}^\sharp(\widehat{p}) = \sum_{i \in \mathcal{N}} \widehat{\sigma}_i(\widehat{p})\widehat{\theta}_i = \widehat{p}$. (ii) Since $\widehat{P} \subset L^\infty(\widehat{K}; \mathbb{R}^q)$, we have $\widehat{\rho}_i \in L^\infty(\widehat{K}; \mathbb{R}^q)$. Hölder's inequality implies that

$$|\widehat{\sigma}_i^\sharp(\widehat{v})| \le |\widehat{K}|^{-\frac{1}{p}} \|\widehat{\rho}_i\|_{L^\infty(\widehat{K}; \mathbb{R}^q)} \|\widehat{v}\|_{L^p(\widehat{K}; \mathbb{R}^q)},$$

for all $\widehat{v} \in L^p(\widehat{K}; \mathbb{R}^q)$. We conclude that (18.21) holds true with $\widehat{c} := \sum_{i \in \mathcal{N}} |\widehat{K}|^{-\frac{1}{p}} \|\widehat{\rho}_i\|_{L^\infty(\widehat{K}; \mathbb{R}^q)} \|\widehat{\theta}_i\|_{L^p(\widehat{K}; \mathbb{R}^q)}$. $\qquad\square$

Consider now a mesh cell $K \in \mathcal{T}_h$ from a shape-regular mesh sequence $(\mathcal{T}_h)_{h \in \mathcal{H}}$ and let (K, P_K, Σ_K) be the finite element generated in K using the transformation ψ_K (see Proposition 9.2). The assumption $\psi_K(v) = \mathbb{A}_K(v \circ \boldsymbol{T}_K)$ implies that $\psi_K^{-1}(L^1(\widehat{K}; \mathbb{R}^q)) = L^1(K; \mathbb{R}^q)$. We extend the dofs in Σ_K to $L^1(K; \mathbb{R}^q)$ by setting $\sigma_{K,i}^\sharp(v) := \widehat{\sigma}_i^\sharp(\psi_K(v))$, i.e., owing to (18.17),

$$\sigma_{K,i}^\sharp(v) = \frac{1}{|\widehat{K}|} (\widehat{\rho}_i, \psi_K(v))_{L^2(\widehat{K}; \mathbb{R}^q)} = \frac{1}{|\widehat{K}|} (\phi_K^{-1}(\widehat{\rho}_i), v)_{L^2(K; \mathbb{R}^q)}, \qquad (18.22)$$

and we define the local interpolation operator in K s.t. for all $\boldsymbol{x} \in K$,

$$\mathcal{I}_K^\sharp(v)(\boldsymbol{x}) := \sum_{i \in \mathcal{N}} \sigma_{K,i}^\sharp(v)\theta_{K,i}(\boldsymbol{x}), \qquad \forall v \in V(K) := L^1(K; \mathbb{R}^q), \quad (18.23)$$

recalling that the local shape functions are given by $\theta_{K,i} := \psi_K^{-1}(\widehat{\theta}_i)$ for all $i \in \mathcal{N}$. The linearity of ψ_K implies that

$$\psi_K\left(\mathcal{I}_K^\sharp(v)\right) := \psi_K\left(\sum_{i \in \mathcal{N}} \sigma_{K,i}^\sharp(v)\psi_K^{-1}(\widehat{\theta}_i)\right) = \sum_{i \in \mathcal{N}} \widehat{\sigma}_i^\sharp(\psi_K(v))\widehat{\theta}_i = \mathcal{I}_{\widehat{K}}^\sharp(\psi_K(v)).$$

In other words, the following key relation holds true:

$$\mathcal{I}_K^\sharp = \psi_K^{-1} \circ \mathcal{I}_{\widehat{K}}^\sharp \circ \psi_K. \qquad (18.24)$$

One can show that $\mathcal{I}_K^\sharp$ is the oblique projection onto $P_K = \psi_K^{-1}(\widehat{P})$ parallel to $Q_K^\perp$ with $Q_K := \Phi_K^{-1}(\widehat{P})$. Note that $\mathcal{I}_K^\sharp$ is L^2-orthogonal whenever the matrix $\mathbb{A}_K$ is unitary; see Exercise 18.3.

Theorem 18.14 (Local approximation). *Let $\mathcal{I}_K^\sharp$ be defined by* (18.23). *Let k be the degree of the finite element, i.e., $[\mathbb{P}_{k,d}]^q \subset \widehat{P} \subset W^{k+1,p}(\widehat{K};\mathbb{R}^q)$. (i) P_K is pointwise invariant under $\mathcal{I}_K^\sharp$. (ii) Assuming that the mesh sequence is shape-regular, there is c s.t. for all $r \in [0, k+1]$, all $p \in [1, \infty)$ if $r \notin \mathbb{N}$ or all $p \in [1, \infty]$ if $r \in \mathbb{N}$, every integer $m \in \{0 : \lfloor r \rfloor\}$, all $v \in W^{r,p}(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$|v - \mathcal{I}_K^\sharp(v)|_{W^{m,p}(K;\mathbb{R}^q)} \leq c\, h_K^{r-m} |v|_{W^{r,p}(K;\mathbb{R}^q)}. \tag{18.25}$$

Proof. The property (i) follows from (18.24). The property (ii) for $r \in \mathbb{N}$ follows from Theorem 11.13 with $l := 0$ since $\mathcal{I}_{\widehat{K}}^\sharp$ is stable in L^p owing to Lemma 18.13. Taking $m := r$ in (18.25) implies the $W^{m,p}$-stability of $\mathcal{I}_K^\sharp$ for every integer $m \in \{0 : k+1\}$, i.e.,

$$|\mathcal{I}_K^\sharp(w)|_{W^{m,p}(K;\mathbb{R}^q)} \leq c\, |w|_{W^{m,p}(K;\mathbb{R}^q)}, \qquad \forall w \in W^{m,p}(K;\mathbb{R}^q). \tag{18.26}$$

Since $\mathcal{I}_K^\sharp(g) = g$ for all $g \in P_K$, (18.26) and the triangle inequality yield

$$\begin{aligned}
|v - \mathcal{I}_K^\sharp(v)|_{W^{m,p}(K;\mathbb{R}^q)} &= \inf_{q \in P_K} |v - g - \mathcal{I}_K^\sharp(v - g)|_{W^{m,p}(K;\mathbb{R}^q)} \\
&\leq c \inf_{q \in P_K} |v - g|_{W^{m,p}(K;\mathbb{R}^q)}.
\end{aligned}$$

Invoking the bound (12.18) on $\inf_{q \in P_K} |v - g|_{W^{m,p}(K;\mathbb{R}^q)}$, we infer that the property (ii) holds true for all $r \notin \mathbb{N}$ as well. $\qquad\square$

Corollary 18.15 (Approximation on faces). (i) *Let $p \in [1, \infty)$ and $r \in (\frac{1}{p}, k+1]$ if $p > 1$ or $r \in [1, k+1]$ if $p = 1$. There is c s.t.*

$$\|v - \mathcal{I}_K^\sharp(v)\|_{L^p(F;\mathbb{R}^q)} \leq c\, h_K^{r-\frac{1}{p}} |v|_{W^{r,p}(K;\mathbb{R}^q)}, \tag{18.27}$$

for all $v \in W^{r,p}(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, all $F \in \mathcal{F}_K$, and all $h \in \mathcal{H}$, where the constant c grows unboundedly as $rp \downarrow 1$ if $p > 1$. (ii) Assume $k \geq 1$. Let $p \in [1, \infty)$ and $r \in (\frac{1}{p}, k]$ if $p > 1$ or $r \in [1, k]$ if $p = 1$. There is c s.t.

$$\|\nabla(v - \mathcal{I}_K^\sharp(v))\|_{\boldsymbol{L}^p(F;\mathbb{R}^q)} \leq c\, h_K^{r-\frac{1}{p}} |v|_{W^{1+r,p}(K;\mathbb{R}^q)}, \tag{18.28}$$

for all $v \in W^{1+r,p}(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$, where the constant c grows unboundedly as $rp \downarrow 1$ if $p > 1$.

Proof. For simplicity, we assume that $q = 1$. The general case is treated by reasoning componentwise. Let us prove (18.27). Assume first that $r \in$

$[1, k + 1]$. Owing to the multiplicative trace inequality (12.16), we infer that, with $\eta := v - \mathcal{I}_K^\sharp(v)$,

$$\|\eta\|_{L^p(F)} \le c \left(h_K^{-\frac{1}{p}} \|\eta\|_{L^p(K)} + \|\eta\|_{L^p(K)}^{1-\frac{1}{p}} \|\nabla\eta\|_{\boldsymbol{L}^p(K)}^{\frac{1}{p}} \right).$$

Invoking (18.25) with $m \in \{0, 1\}$ (note that $m \le \lfloor r \rfloor$) shows that (18.27) holds true in this case. Let us now assume that $r \in (\frac{1}{p}, 1)$ with $p > 1$. Let $q_0 \in \psi_K^{-1}(\mathbb{P}_{0,d}) = \mathbb{P}_{0,d}$ be arbitrary. We have

$$\begin{aligned}
h_K^{\frac{1}{p}} \|\eta\|_{L^p(F)} &\le h_K^{\frac{1}{p}} \|v - q_0\|_{L^p(F)} + h_K^{\frac{1}{p}} \|\mathcal{I}_K^\sharp(v) - q_0\|_{L^p(F)} \\
&\le c \left(\|v - q_0\|_{L^p(K)} + h_K^r |v|_{W^{r,p}(K)} + \|\mathcal{I}_K^\sharp(v) - q_0\|_{L^p(K)} \right) \\
&\le c \left(\|v - q_0\|_{L^p(K)} + h_K^r |v|_{W^{r,p}(K)} + \|v - \mathcal{I}_K^\sharp(v)\|_{L^p(K)} \right),
\end{aligned}$$

where we used the triangle inequality in the first line, the fractional trace inequality (12.17), the discrete trace inequality (12.10) and $q_0 \in \mathbb{P}_{0,d}$ in the second line, and the triangle inequality in the third line. Invoking the best-approximation estimate (12.15) from Corollary 12.13 (observe that q_0 is arbitrary in $\mathbb{P}_{0,d}$) and (18.25) with $m = 0$ leads again to (18.27). Finally, the proof of (18.28) is similar and is left as an exercise. $\qquad\square$

We define $\mathcal{I}_h^\sharp : L^1(D; \mathbb{R}^q) \to P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$ s.t. for all $v \in L^1(D; \mathbb{R}^q)$,

$$\mathcal{I}_h^\sharp(v)_{|K} := \mathcal{I}_K^\sharp(v_{|K}), \qquad \forall K \in \mathcal{T}_h. \tag{18.29}$$

The approximation properties of $\mathcal{I}_h^\sharp$ readily follow from Theorem 18.14.

18.4 Broken L^2-orthogonal projection

Let $K \in \mathcal{T}_h$ be a mesh cell. The L^2-orthogonal projection $\mathcal{I}_K^{\mathrm{b}} : L^1(K; \mathbb{R}^q) \to P_K$ is defined s.t. for all $v \in L^1(K; \mathbb{R}^q)$,

$$(\mathcal{I}_K^{\mathrm{b}}(v) - v, q)_{L^2(K;\mathbb{R}^q)} = 0, \qquad \forall q \in P_K, \tag{18.30}$$

where $P_K := \psi_K^{-1}(\widehat{P})$ and $\psi_K(v) := \mathbb{A}_K(v \circ \boldsymbol{T}_K)$. Since (18.30) implies that

$$\|v - q\|_{L^2(K;\mathbb{R}^q)}^2 = \|v - \mathcal{I}_K^{\mathrm{b}}(v)\|_{L^2(K;\mathbb{R}^q)}^2 + \|\mathcal{I}_K^{\mathrm{b}}(v) - q\|_{L^2(K;\mathbb{R}^q)}^2, \tag{18.31}$$

we have the optimality property

$$\mathcal{I}_K^{\mathrm{b}}(v) = \arg\min_{q \in P_K} \|v - q\|_{L^2(K;\mathbb{R}^q)}. \tag{18.32}$$

The stability and approximation properties of $\mathcal{I}_K^b$ can be analyzed by using the L^1-stable interpolation operator $\mathcal{I}_K^\sharp$ introduced in the previous section.

Theorem 18.16 (Stability and local approximation). *Let $\mathcal{I}_K^b$ be defined by* (18.30). *Let k be the degree of the finite element, i.e., $[\mathbb{P}_{k,d}]^q \subset \widehat{P} \subset W^{k+1,p}(\widehat{K};\mathbb{R}^q)$. Assume that the mesh sequence is shape-regular.* (i) P_K *is pointwise invariant under $\mathcal{I}_K^b$.* (ii) $\mathcal{I}_K^b$ *is L^p-stable for all $p \in [1,\infty]$, i.e., there is c s.t. $\|\mathcal{I}_K^b(v)\|_{L^p(K;\mathbb{R}^q)} \leq c\|v\|_{L^p(K;\mathbb{R}^q)}$ for all $v \in L^p(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.* (iii) *There is c s.t.*

$$|v - \mathcal{I}_K^b(v)|_{W^{m,p}(K;\mathbb{R}^q)} \leq c\, h_K^{r-m}|v|_{W^{r,p}(K;\mathbb{R}^q)}, \qquad (18.33)$$

for all $r \in [0, k+1]$, all $p \in [1,\infty)$ if $r \notin \mathbb{N}$ or all $p \in [1,\infty]$ if $r \in \mathbb{N}$, every integer $m \in \{0{:}\lfloor r \rfloor\}$, all $v \in W^{r,p}(K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. (i) The pointwise invariance of P_K under $\mathcal{I}_K^b$ follows from (18.30).
(ii) Stability. Let $v \in L^p(K;\mathbb{R}^q)$. We observe that

$$
\begin{aligned}
\|\mathcal{I}_K^b(v)\|_{L^p(K;\mathbb{R}^q)}^2 &\leq c\, h_K^{d(\frac{2}{p}-1)}\|\mathcal{I}_K^b(v)\|_{L^2(K;\mathbb{R}^q)}^2 = c\, h_K^{d(\frac{2}{p}-1)}(v, \mathcal{I}_K^b(v))_{L^2(K;\mathbb{R}^q)} \\
&\leq c\, h_K^{d(\frac{2}{p}-1)}\|v\|_{L^p(K;\mathbb{R}^q)}\|\mathcal{I}_K^b(v)\|_{L^{p'}(K;\mathbb{R}^q)} \\
&\leq c'\, h_K^{d(\frac{2}{p}-1+\frac{1}{p'}-\frac{1}{p})}\|v\|_{L^p(K;\mathbb{R}^q)}\|\mathcal{I}_K^b(v)\|_{L^p(K;\mathbb{R}^q)} \\
&= c'\, \|v\|_{L^p(K;\mathbb{R}^q)}\|\mathcal{I}_K^b(v)\|_{L^p(K;\mathbb{R}^q)},
\end{aligned}
$$

where we used the inverse inequality (12.3) (between L^p and L^2), (18.30) with $q := \mathcal{I}_K^b(v)$, Hölder's inequality (with $\frac{1}{p} + \frac{1}{p'} = 1$), and again the inverse inequality (12.3) (between $L^{p'}$ and L^p). This proves the L^p-stability of $\mathcal{I}_K^b$.
(iii) Local approximation. Since $\mathcal{I}_K^\sharp(v) \in P_K$ and P_K is left pointwise invariant by $\mathcal{I}_K^b$, we have

$$
\begin{aligned}
|v - \mathcal{I}_K^b(v)|_{W^{m,p}(K;\mathbb{R}^q)} &\leq |v - \mathcal{I}_K^\sharp(v)|_{W^{m,p}(K;\mathbb{R}^q)} + |\mathcal{I}_K^b(v - \mathcal{I}_K^\sharp(v))|_{W^{m,p}(K;\mathbb{R}^q)} \\
&\leq |v - \mathcal{I}_K^\sharp(v)|_{W^{m,p}(K;\mathbb{R}^q)} + ch_K^{-m}\|\mathcal{I}_K^b(v - \mathcal{I}_K^\sharp(v))\|_{L^p(K;\mathbb{R}^q)} \\
&\leq |v - \mathcal{I}_K^\sharp(v)|_{W^{m,p}(K;\mathbb{R}^q)} + c'h_K^{-m}\|v - \mathcal{I}_K^\sharp(v)\|_{L^p(K;\mathbb{R}^q)} \\
&\leq c''h_K^{r-m}|v|_{W^{r,p}(K;\mathbb{R}^q)},
\end{aligned}
$$

where we used the triangle inequality, the inverse inequality from Lemma 12.1, the L^p-stability of $\mathcal{I}_K^b$, and the approximation property (18.25) of $\mathcal{I}_K^\sharp$. $\qquad\square$

We define $\mathcal{I}_h^b : L^1(D;\mathbb{R}^q) \rightarrow P_k^b(\mathcal{T}_h;\mathbb{R}^q)$ s.t. for all $v \in L^1(D;\mathbb{R}^q)$, $\mathcal{I}_h^b(v)_{|K} := \mathcal{I}_K^b(v_{|K})$ for all $K \in \mathcal{T}_h$. One readily verifies that $\mathcal{I}_h^b$ is the L^2-orthogonal projection onto $P_k^b(\mathcal{T}_h;\mathbb{R}^q)$. The stability and approximation properties of $\mathcal{I}_h^b$ follow from Theorem 18.16.

Remark 18.17 (Approximation on faces). A result similar to Corollary 18.15 holds true for $\mathcal{I}_K^b$ on the mesh faces. $\qquad\square$

Remark 18.18 (Pullback). One cannot investigate the approximation properties of $\mathcal{I}_K^{\text{b}}$ by introducing the L^2-orthogonal projection onto $\widehat{P}$ (i.e., the operator $\mathcal{I}_{\widehat{K}}^{\sharp}$) and using Theorem 11.13, since we have seen that $\psi_K^{-1} \circ \mathcal{I}_{\widehat{K}}^{\sharp} \circ \psi_K$ is actually the oblique projection $\mathcal{I}_K^{\sharp}$ and not the L^2-orthogonal projection $\mathcal{I}_K^{\text{b}}$. The two projections $\mathcal{I}_K^{\sharp}$ and $\mathcal{I}_K^{\text{b}}$ coincide when the matrix $\mathbb{A}_K$ is unitary (see Exercise 18.3). This happens when ψ_K is the pullback by the geometric mapping $\boldsymbol{T}_K$, i.e., when $\mathbb{A}_K$ is the identity as is the case for scalar-valued elements. In this situation, Theorem 18.16 has already been established in Lemma 11.18 (at least for $r \in \{0{:}k+1\}$). □

Remark 18.19 (Algebraic realization). To evaluate the L^2-orthogonal projection $\mathcal{I}_K^{\text{b}}(v)$ of a function v, one has to solve the linear system $\mathcal{M}_K X = Y$, where the local mass matrix has entries $\mathcal{M}_{K,mn} := \int_K (\theta_{K,m}, \theta_{K,n})_{\ell^2(\mathbb{R}^q)} \,\mathrm{d}x$ for all $m, n \in \mathcal{N}$, and the right-hand side vector Y has components $Y_n := \int_K (v, \theta_{K,n})_{\ell^2(\mathbb{R}^q)} \,\mathrm{d}x$. Then we have $\mathcal{I}_K^{\text{b}}(v) = \sum_{n \in \mathcal{N}} X_n \theta_{K,n}$; see §5.4.2. □

Exercises

Exercise 18.1 ($\boldsymbol{H}(\text{div}), \boldsymbol{H}(\text{curl})$). Prove Theorem 18.10. (*Hint*: use (4.8).)

Exercise 18.2 (Discrete Sobolev inequality). (i) Assume $d \geq 3$. Prove that $\|v_h\|_{L^\infty(K)} \leq c h_K^{1-\frac{d}{2}} \|\nabla v_h\|_{L^2(K)}$ for all $v_h \in P_k^{\text{g,b}}(\mathcal{T}_h)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$. (*Hint*: use Theorem 2.31.) (ii) Assume $d = 2$. Prove (18.15). (*Hint*: let $K \in \mathcal{T}_h$ with $h_K \leq \frac{\delta_D}{2}$, let $\boldsymbol{x} \in K$ and let $\boldsymbol{y}$ have polar coordinates (r, θ) with respect to $\boldsymbol{x}$ with $r \geq \frac{\delta_D}{2}$ and $\theta \in (0, \omega)$, use that $v_h(\boldsymbol{x}) = v_h(\boldsymbol{y}) - \int_0^r \partial_\rho v_h(\rho, \theta) \,\mathrm{d}\rho$, decompose the integral as $\int_0^r \cdot \,\mathrm{d}\rho = \int_0^{h_K} \cdot \,\mathrm{d}\rho + \int_{h_K}^r \cdot \,\mathrm{d}\rho$, and bound the two addends.)

Exercise 18.3 (Orthogonal and oblique projections). (i) Show that $\mathcal{I}_{\widehat{K}}^{\sharp}$ is the L^2-orthogonal projection onto $\widehat{P}$. (*Hint*: observe that $(\widehat{\rho}_i, \widehat{\theta}_j)_{L^2(\widehat{K};\mathbb{R}^q)} = |\widehat{K}| \delta_{ij}$ for all $i, j \in \mathcal{N}$.) (ii) Prove that $\mathcal{I}_K^{\sharp}$ is the oblique projection onto $P_K = \psi_K^{-1}(\widehat{P})$ parallel to $Q_K^\perp$ with $Q_K := \Phi_K^{-1}(\widehat{P})$. (*Hint*: use (18.17).) (iii) Show that $P_K = Q_K$ if the matrix $\mathbb{A}_K$ is unitary, i.e., $\mathbb{A}_K^{\mathsf{T}} \mathbb{A}_K = \mathbb{A}_K \mathbb{A}_K^{\mathsf{T}} = \mathbb{I}_q$.

Exercise 18.4 (Approximation on faces). Prove (18.28).

Part IV, Chapter 19

Main properties of the conforming subspaces

In this chapter, we continue the study of the interpolation properties of the conforming finite element subspaces introduced in the previous chapter. Recall that

$$P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q) := \{ v_h \in \boldsymbol{P}_k^{\mathrm{x,b}}(\mathcal{T}_h; \mathbb{R}^q) \mid [\![v_h]\!]_F^{\mathrm{x}} = 0, \ \forall F \in \mathcal{F}_h^{\circ} \},$$

where $\boldsymbol{P}_k^{\mathrm{x,b}}(\mathcal{T}_h; \mathbb{R}^q)$ is a broken finite element space, with $q \in \{1, d\}$ depending on the superscript $\mathrm{x} \in \{\mathrm{g, c, d}\}$, and the jump operator $[\![\cdot]\!]_F^{\mathrm{x}}$ is defined in (18.7). Recall that the H^1-conforming subspace $P_k^{\mathrm{g}}(\mathcal{T}_h)$ ($q = 1$) is built using a Lagrange element or a canonical hybrid element of degree $k \geq 1$, the $\boldsymbol{H}(\mathrm{curl})$-conforming subspace $\boldsymbol{P}_k^{\mathrm{c}}(\mathcal{T}_h)$ ($q = d = 3$) is built using a Nédélec element of degree $k \geq 0$, and the $\boldsymbol{H}(\mathrm{div})$-conforming subspace $\boldsymbol{P}_k^{\mathrm{d}}(\mathcal{T}_h)$ ($q = d$) is built using a Raviart–Thomas element of degree $k \geq 0$. The cornerstone of the construction, which is presented in a unified way for $\mathrm{x} \in \{\mathrm{g, c, d}\}$, is a connectivity array with ad hoc clustering properties of the local degrees of freedom (dofs). In the present chapter, we postulate the existence of the connectivity array and show how it allows us to build global shape functions and a global interpolation operator in $P_k^{\mathrm{g}}(\mathcal{T}_h)$. The actual construction of this mapping is undertaken in Chapters 20 and 21. In this book, we shall implicitly assume that the mesh $\mathcal{T}_h$ is matching (see Definition 8.11) when the conforming space $P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$ is invoked.

19.1 Global shape functions and dofs

For all $K \in \mathcal{T}_h$, the local dofs are $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$, and the local shape functions are $\{\theta_{K,i}\}_{i \in \mathcal{N}}$. Recall that $\{\theta_{K,i}\}_{i \in \mathcal{N}}$ is a basis of P_K and that $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$ is a basis of $\mathcal{L}(P_K; \mathbb{R})$. We start by organizing all the dofs and shape functions

$$\{\sigma_{K,i}\}_{(K,i) \in \mathcal{T}_h \times \mathcal{N}}, \qquad \{\theta_{K,i}\}_{(K,i) \in \mathcal{T}_h \times \mathcal{N}},$$

by grouping them into clusters, which we are going to call *connectivity classes.*
We assume that we have at hand a nonzero natural number I and a *connectivity array*

$$\texttt{j\_dof} : \mathcal{T}_h \times \mathcal{N} \to \mathcal{A}_h := \{1{:}I\}. \tag{19.1}$$

Without loss of generality we assume that the mapping $\texttt{j\_dof}$ is surjective,
i.e., for every connectivity class $a \in \mathcal{A}_h$, there exists $(K, i) \in \mathcal{T}_h \times \mathcal{N}$ s.t.
$\texttt{j\_dof}(K, i) = a$. This hypothesis is nonessential and can always be satisfied
by rearranging the codomain of $\texttt{j\_dof}$.

Definition 19.1 (Connectivity class). *Two pairs* $(K, i), (K', i') \in \mathcal{T}_h \times \mathcal{N}$
are said to be in the same connectivity class if $\texttt{j\_dof}(K, i) = \texttt{j\_dof}(K', i')$.

We require that the mapping $\texttt{j\_dof}$ satisfies two key properties.
(1) The first one is that for all $v_h \in P_k^{\mathrm{x,b}}(\mathcal{T}_h)$,

$$[\, v_h \in P_k^{\mathrm{x}}(\mathcal{T}_h) \,] \iff \left[\begin{array}{l} \text{For all } (K, i), (K', i') \text{ in the same} \\ \text{connectivity class, we have} \\ \sigma_{K,i}(v_{h|K}) = \sigma_{K',i'}(v_{h|K'}) \end{array} \right]. \tag{19.2}$$

Thus, (19.2) means that for every function v_h in the broken finite element
space $P_k^{\mathrm{x,b}}(\mathcal{T}_h)$, a *necessary and sufficient* condition for v_h to be a member of the conforming subspace $P_k^{\mathrm{x}}(\mathcal{T}_h)$ is that for all $a \in \mathcal{A}_h$, the quantity
$\sigma_{K,i}(v_{h|K})$ is independent of the choice of the pair (K, i) in the preimage
$\texttt{j\_dof}^{-1}(a) := \{(K', i') \in \mathcal{T}_h \times \mathcal{N} \mid \texttt{j\_dof}(K', i') = a\}$.
(2) The second key property is that

$$\forall K \in \mathcal{T}_h, \qquad \texttt{j\_dof}(K, \cdot) : \mathcal{N} \to \mathcal{A}_h \text{ is injective}, \tag{19.3}$$

i.e., if (K, i) and (K, i') are in the same connectivity class, then $i = i'$.
We now construct global dofs and shape functions in $P_k^{\mathrm{x}}(\mathcal{T}_h)$. Since for all
$a \in \mathcal{A}_h$ and all $v_h \in P_k^{\mathrm{x}}(\mathcal{T}_h)$, (19.2) implies that the value of $\sigma_{K,i}(v_{h|K})$ is
independent of the choice of the pair (K, i) in the connectivity class a, it is
legitimate to introduce the following definition: For all $a \in \mathcal{A}_h$, we define the
linear form $\sigma_a : P_k^{\mathrm{x}}(\mathcal{T}_h) \to \mathbb{R}$ s.t. for all $v_h \in P_h^{\mathrm{x}}(\mathcal{T}_h)$,

$$\sigma_a(v_h) := \sigma_{K,i}(v_{h|K}), \qquad \forall (K, i) \in \texttt{j\_dof}^{-1}(a), \tag{19.4}$$

i.e., $\sigma_a(v_h) := \sigma_{K,i}(v_{h|K})$ for every pair (K, i) in the connectivity class a.
Observe that $\sigma_a \in \mathcal{L}(P_k^{\mathrm{x}}(\mathcal{T}_h); \mathbb{R})$. We now define the function $\varphi_a : \overline{D} \to \mathbb{R}^q$
for all $a \in \mathcal{A}_h$ by

$$\varphi_{a|K} := \begin{cases} \theta_{K,i} & \text{if there exists } i \in \mathcal{N} \text{ s.t. } (K, i) \in \texttt{j\_dof}^{-1}(a), \\ 0 & \text{otherwise.} \end{cases} \tag{19.5}$$

This definition makes sense since if $(K, i) \in \texttt{j\_dof}^{-1}(a)$ and $(K, i') \in \texttt{j\_dof}^{-1}(a)$,
then $i = i'$ owing to (19.3).

Definition 19.2 (Global shape functions and dofs). *The functions φ_a are called* global shape functions, *and the linear forms σ_a are called* global degrees of freedom (dofs).

For all $a \in \mathcal{A}_h$, let us introduce the following collection of cells:

$$\mathcal{T}_a := \{K \in \mathcal{T}_h \mid \exists i \in \mathcal{N}, (K, i) \in \mathtt{j\_dof}^{-1}(a)\}, \tag{19.6}$$

i.e., $\mathcal{T}_a = \{K \in \mathcal{T}_h \mid a \in \mathtt{j\_dof}(K, \mathcal{N})\}$. A direct consequence of the definition (19.5) is that

$$\mathrm{supp}(\varphi_a) = \bigcup_{K \in \mathcal{T}_a} K. \tag{19.7}$$

Lemma 19.3 (Conformity). *For all a in $\mathcal{A}_h$, $\varphi_a \in P_k^{\mathrm{x}}(\mathcal{T}_h)$ and*

$$\sigma_a(\varphi_{a'}) = \delta_{aa'}, \qquad \forall a' \in \mathcal{A}_h. \tag{19.8}$$

Proof. Let $a \in \mathcal{A}_h$ and let us prove that $\varphi_a \in P_k^{\mathrm{x}}(\mathcal{T}_h)$. Since $\varphi_a \in P_k^{\mathrm{x,b}}(\mathcal{T}_h)$, we prove the assertion by checking that the property on the right-hand side of (19.2) holds true. Let a' be arbitrary in $\mathcal{A}_h$. We need to show that the quantity $\sigma_{K,i}(\varphi_{a|K})$ is independent of the pair $(K, i) \in \mathtt{j\_dof}^{-1}(a')$.
(1) Assume first that $a' = a$. Let (K, i) be an arbitrary pair in $\mathtt{j\_dof}^{-1}(a')$. Then $\mathtt{j\_dof}(K, i) = a' = a$, and the definition of φ_a implies that $\varphi_{a|K} = \theta_{K,i}$. Hence, $\sigma_{K,i}(\varphi_{a|K}) = \sigma_{K,i}(\theta_{K,i}) = 1$ for all $(K, i) \in \mathtt{j\_dof}^{-1}(a')$.
(2) Assume now that $a' \neq a$. Let (K, i) be an arbitrary pair in $\mathtt{j\_dof}^{-1}(a')$. If there exists $j \in \mathcal{N}$ s.t. $\mathtt{j\_dof}(K, j) = a$, then $\varphi_{a|K} = \theta_{K,j}$. Notice that $j \neq i$ owing to (19.3), since $\mathtt{j\_dof}(K, j) = a \neq a' = \mathtt{j\_dof}(K, i)$. We infer in this case that $\sigma_{K,i}(\varphi_{a|K}) = \sigma_{K,i}(\theta_{K,j}) = 0$ since $j \neq i$. If there is no $j \in \mathcal{N}$ s.t. $\mathtt{j\_dof}(K, j) = a$, then $\varphi_{a|K} = 0$ and again $\sigma_{K,i}(\varphi_{a|K}) = 0$. To sum up, $\sigma_{K,i}(\varphi_{a|K}) = 0$ for all $(K, i) \in \mathtt{j\_dof}^{-1}(a')$.
(3) In conclusion, the above argument shows that $\sigma_a(\varphi_a) = 1$ and $\sigma_{a'}(\varphi_a) = 0$ if $a' \neq a$, i.e., i $\sigma_{K,i}(\varphi_{a|K})$ is independent of the pair $(K, i) \in \mathtt{j\_dof}^{-1}(a')$ for all $a' \in \mathcal{A}_h$, and (19.8) holds true. $\square$

Proposition 19.4 (Basis). $\{\varphi_a\}_{a \in \mathcal{A}_h}$ *is a basis of $P_k^{\mathrm{x}}(\mathcal{T}_h)$, and $\{\sigma_a\}_{a \in \mathcal{A}_h}$ is a basis of $\mathcal{L}(P_k^{\mathrm{x}}(\mathcal{T}_h); \mathbb{R})$.*

Proof. Assume that $\sum_{a \in \mathcal{A}_h} \lambda_a \varphi_a$ vanishes identically on $\overline{D}$ for some real numbers $\{\lambda_a\}_{a \in \mathcal{A}_h}$. Using the linearity of σ_a and (19.8) yields

$$0 = \sigma_{a'}(0) = \sigma_{a'}\left(\sum_{a \in \mathcal{A}_h} \lambda_a \varphi_a \right) = \sum_{a \in \mathcal{A}_h} \lambda_a \sigma_{a'}(\varphi_a) = \lambda_{a'}.$$

Hence, $\lambda_{a'} = 0$ for all $a' \in \mathcal{A}_h$, i.e., $\{\varphi_a\}_{a \in \mathcal{A}_h}$ is linearly independent. To show that $\{\varphi_a\}_{a \in \mathcal{A}_h}$ is a spanning set of $P_k^{\mathrm{x}}(\mathcal{T}_h)$, let $v_h \in P_k^{\mathrm{x}}(\mathcal{T}_h)$ and let us set $\delta_h := v_h - \sum_{a' \in \mathcal{A}_h} \sigma_{a'}(v_h) \varphi_{a'}$. We are going to prove that $\delta_{h|K} = 0$ for all $K \in \mathcal{T}_h$, and since $\delta_{h|K} \in P_K$, we do so by showing that $\delta_{h|K}$ annihilates all

the local dofs in K, i.e., $\sigma_{K,i}(\delta_{h|K}) = 0$ for all $i \in \mathcal{N}$. Let K be an arbitrary cell in $\mathcal{T}_h$, let i be an arbitrary index in $\mathcal{N}$, and let $a := \mathtt{j\_dof}(K, i)$. Then

$$\sigma_{K,i}(\delta_{h|K}) = \sigma_a(\delta_h) = \sigma_a(v_h) - \sigma_a(v_h) = 0,$$

where the first equality follows from the fact that $\delta_h \in P_k^{\mathrm{x}}(\mathcal{T}_h)$ and the second one from (19.8). We have thus proved that $\delta_{h|K} = 0$ for all $K \in \mathcal{T}_h$, and hence that δ_h vanishes identically because K is arbitrary. In conclusion, $\{\varphi_a\}_{a \in \mathcal{A}_h}$ is a basis of $P_k^{\mathrm{x}}(\mathcal{T}_h)$. Since $\{\varphi_a\}_{a \in \mathcal{A}_h}$ is a basis of $P_k^{\mathrm{x}}(\mathcal{T}_h)$, the identity (19.8) implies that $\{\sigma_a\}_{a \in \mathcal{A}_h}$ is a basis of $\mathcal{L}(P_k^{\mathrm{x}}(\mathcal{T}_h); \mathbb{R})$. $\square$

To sum up, we have shown that provided we have at hand a connectivity array $\mathtt{j\_dof} : \mathcal{T}_h \times \mathcal{N} \to \mathcal{A}_h$ satisfying the properties (19.2) and (19.3), we can build in a simple manner the global basis functions and the global dofs in the conforming finite element subspace $P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$. The actual construction of the mapping $\mathtt{j\_dof}$ will be undertaken in the following two chapters.

Remark 19.5 (Connectivity class). Another way to formalize the grouping of the dofs consists of introducing the equivalence relation $\mathcal{R}$ in $\mathcal{T}_h \times \mathcal{N}$ defined by $(K, i) \,\mathcal{R}\, (K', i')$ iff $\mathtt{j\_dof}(K, i) = \mathtt{j\_dof}(K', i')$. One can then redefine $\mathcal{A}_h$ to be the set of the equivalence classes for $\mathcal{R}$. The elements of $\mathcal{A}_h$ are then sets and are called connectivity classes. In this case, we write $(K, i) \in a$ instead of $\mathtt{j\_dof}(K, i) = a$. We are going to adopt this equivalent viewpoint from Chapter 20 onward. $\square$

19.2 Examples

In this section, we illustrate the concepts developed in §19.1 for the spaces $P_k^{\mathrm{g}}(\mathcal{T}_h)$, $\boldsymbol{P}_k^{\mathrm{c}}(\mathcal{T}_h)$, and $\boldsymbol{P}_k^{\mathrm{d}}(\mathcal{T}_h)$.

19.2.1 H^1-conforming subspace $P_k^{\mathrm{g}}(\mathcal{T}_h)$

Let $(\widehat{K}, \widehat{P}^{\mathrm{g}}, \widehat{\Sigma}^{\mathrm{g}})$ be one of the scalar-valued Lagrange elements of degree $k \geq 1$ introduced in §6.4 or §7.4, or one of the canonical hybrid finite elements of degree $k \geq 1$ introduced in §7.6. The broken finite element space is

$$P_k^{\mathrm{g,b}}(\mathcal{T}_h) := \{\boldsymbol{v}_h \in L^\infty(D) \mid \psi_K^{\mathrm{g}}(v_h) \in \widehat{P}^{\mathrm{g}}, \ \forall K \in \mathcal{T}_h\}, \tag{19.9}$$

where $\psi_K^{\mathrm{g}}(v) := v \circ \boldsymbol{T}_K$ is the pullback by the geometric mapping, and the corresponding H^1-conforming subspace is

$$P_k^{\mathrm{g}}(\mathcal{T}_h) := \{v_h \in P_k^{\mathrm{g,b}}(\mathcal{T}_h) \mid \llbracket v_h \rrbracket_F = 0, \ \forall F \in \mathcal{F}_h^\circ\}. \tag{19.10}$$

We have $P_k^{\mathrm{g}}(\mathcal{T}_h) \subset Z^{\mathrm{g},p}(D) := W^{1,p}(D) = \{v \in L^p(D) \mid \nabla v \in \boldsymbol{L}^p(D)\}$ for all $p \in [1, \infty]$ (note that $Z^{\mathrm{g},2}(D) := H^1(D)$). We show in Figure 19.1 the

connectivity classes generated by j_dof on a mesh composed of four triangles with $\mathbb{P}_{2,2}$ Lagrange elements.
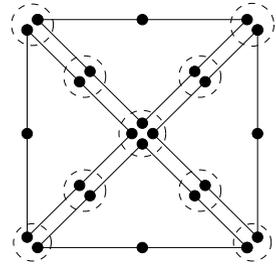
Fig. 19.1 $\mathbb{P}_{2,2}$ Lagrange nodes in the same connectivity class for a mesh composed of four triangles (drawn slightly apart).

The Lagrange and the canonical hybrid finite elements of the same degree generate the same space $P_k^{\mathrm{g}}(\mathcal{T}_h)$, but the shape functions and dofs differ for $k \geq 2$. Some global shape functions in $\mathbb{P}_1^{\mathrm{g}}(\mathcal{T}_h)$ and $\mathbb{P}_2^{\mathrm{g}}(\mathcal{T}_h)$ in dimension 2 are shown in Figure 19.2 for Lagrange elements. The function shown in the left panel is continuous and piecewise affine, and it takes the value 1 at one mesh vertex and the value 0 at all the other mesh vertices. Because its graph is reminiscent of a hat, this function is often called *hat basis function* (and sometimes also *Courant basis functions* [84]). The functions shown in the central and right panels are continuous and piecewise quadratic. The function on the central panel takes the value 1 at one mesh vertex and the value 0 at all the other mesh vertices, and it takes the value 0 at all the edge midpoints. The function in the right panel takes the value 0 at all the mesh vertices, and it takes the value 1 at one edge midpoint and the value 0 at the midpoint of all the other edges.
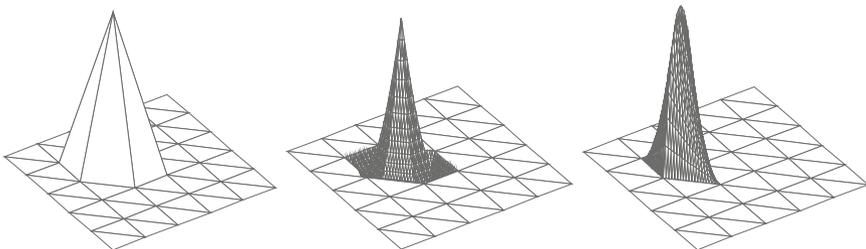
Fig. 19.2 Global shape functions in dimension 2: $\mathbb{P}_{1,2}$ (left) and $\mathbb{P}_{2,2}$ (center and right) Lagrange finite elements.

Let N_{v}, N_{e}, N_{f}, N_{c} be the number of vertices, edges, faces, and cells in the mesh $\mathcal{T}_h$ (recall that $\mathcal{T}_h$ is assumed to be a matching mesh). For a simplicial Lagrange element, the number of Lagrange nodes per edge that are not located at the extremities of the edge is $\binom{k-1}{1}$ (if $k \geq 2$), the number

of Lagrange nodes per face that are not located at the boundary of the face is $\binom{k-1}{2}$ (if $k \geq 3$), and the number of Lagrange nodes per cell that not located at the boundary of the cell is $\binom{k-1}{3}$ (if $k \geq 4$). These numbers are the same for the canonical hybrid finite element. We will establish in Chapter 21 that

$$\dim(P_k^g(\mathcal{T}_h)) = N_v + \binom{k-1}{1}N_e + \binom{k-1}{2}N_f + \binom{k-1}{3}N_c \quad \text{if } d = 3, \quad (19.11a)$$

$$\dim(P_k^g(\mathcal{T}_h)) = N_v + \binom{k-1}{1}N_e + \binom{k-1}{2}N_c \qquad \text{if } d = 2, \quad (19.11b)$$

with the convention that for natural numbers n, m, $\binom{n}{m} := 0$ if $n < m$. In the lowest-order case ($k = 1$), we have $\dim(P_1^g(\mathcal{T}_h)) = N_v$, and the connectivity array j_dof coincides with the double-entry array j_cv defined in §8.3.

19.2.2 $H(\mathrm{curl})$-conforming subspace $P_k^c(\mathcal{T}_h)$

Let $(\widehat{K}, \widehat{P}^c, \widehat{\Sigma}^c)$ be one of the Nédélec finite elements of degree $k \geq 0$ described in Chapter 15. The broken finite element space is

$$P_k^{c,b}(\mathcal{T}_h) := \{v_h \in L^\infty(D) \mid \psi_K^c(v_{h|K}) \in \widehat{P}^c, \ \forall K \in \mathcal{T}_h\}, \qquad (19.12)$$

with the covariant Piola transformation $\psi_K^c(v) := \mathbb{J}_K^\mathsf{T}(v \circ T_K)$, and the corresponding $H(\mathrm{curl})$-conforming subspace is

$$P_k^c(\mathcal{T}_h) := \{v_h \in P_k^{c,b}(\mathcal{T}_h) \mid [\![v_h]\!]_F \times n_F = 0, \ \forall F \in \mathcal{F}_h^\circ\}. \qquad (19.13)$$

We have $P_k^c(\mathcal{T}_h) \subset Z^{c,p}(D) := \{v \in L^p(D) \mid \nabla \times v \in L^p(D)\}$ for all $p \in [1, \infty]$ (note that $Z^{c,2}(D) := H(\mathrm{curl}; D)$). A global shape function attached to an edge is shown in the left panel of Figure 19.3 for the $\mathbb{N}_{0,2}$ element. Notice that the tangential component is continuous across the interface, but the normal component is not.
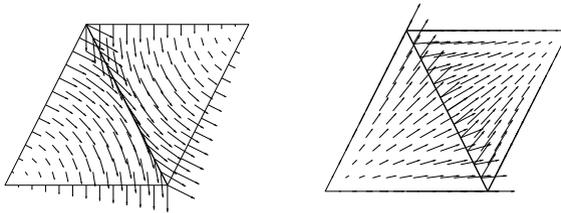
Fig. 19.3 Global shape functions for the lowest-order Nédélec (left) and Raviart–Thomas (right) elements in dimension 2.

Let N_e, N_f, N_c be the number of edges, faces, and cells in $\mathcal{T}_h$. We will show in Chapter 21 that

$$\dim(\boldsymbol{P}_k^c(\mathcal{T}_h)) = \binom{k+1}{1}N_e + 2\binom{k+1}{2}N_f + 3\binom{k+1}{3}N_c, \qquad \text{if } d = 3,$$

$$\dim(\boldsymbol{P}_k^c(\mathcal{T}_h)) = \binom{k+1}{1}N_e + 2\binom{k+1}{2}N_c, \qquad \text{if } d = 2,$$

with the convention that $\binom{n}{m} := 0$ if $n < m$. In the lowest-order case $(k = 0)$, we have $\dim(\boldsymbol{P}_0^c(\mathcal{T}_h)) = N_e$, and the connectivity array $\mathtt{j\_dof}$ coincides with the double-entry array $\mathtt{j\_ce}$ defined in §8.3.

19.2.3 $H(\mathrm{div})$-conforming subspace $\boldsymbol{P}_k^d(\mathcal{T}_h)$

Let $(\widehat{K}, \widehat{\boldsymbol{P}}^d, \widehat{\Sigma}^d)$ be one of the Raviart–Thomas finite elements of degree $k \geq 0$ introduced in Chapter 14. The broken finite element space is

$$\boldsymbol{P}_k^{d,b}(\mathcal{T}_h) := \{\boldsymbol{v}_h \in \boldsymbol{L}^1(D) \mid \psi_K^d(\boldsymbol{v}_{h|K}) \in \widehat{\boldsymbol{P}}^d, \ \forall K \in \mathcal{T}_h\}, \qquad (19.15)$$

with the contravariant Piola transformation $\psi_K^d(\boldsymbol{v}) := \det(\mathbb{J}_K)\, \mathbb{J}_K^{-1}(\boldsymbol{v} \circ \boldsymbol{T}_K)$. The corresponding $\boldsymbol{H}(\mathrm{div})$-conforming subspace is

$$\boldsymbol{P}_k^d(\mathcal{T}_h) := \{\boldsymbol{v}_h \in \boldsymbol{P}_k^{d,b}(\mathcal{T}_h) \mid [\![\boldsymbol{v}_h]\!]_F \cdot \boldsymbol{n}_F = 0, \ \forall F \in \mathcal{F}_h^\circ\}. \qquad (19.16)$$

We have $\boldsymbol{P}_k^d(\mathcal{T}_h) \subset \boldsymbol{Z}^{d,p}(D) := \{\boldsymbol{v} \in \boldsymbol{L}^p(D) \mid \nabla \cdot \boldsymbol{v} \in L^p(D)\}$ for all $p \in [1, \infty]$ (note that $\boldsymbol{Z}^{d,2}(D) := \boldsymbol{H}(\mathrm{div}; D)$). A global shape function attached to a face is shown in the right panel of Figure 19.3 for the $\mathbb{RT}_{0,2}$ element (the normal component is continuous across the interface, but the tangential component is not). We will establish in Chapter 21 that

$$\dim(\boldsymbol{P}_k^d(\mathcal{T}_h)) = \binom{k+2}{2}N_f + 3\binom{k+2}{3}N_c, \qquad \text{if } d = 3, \qquad (19.17a)$$

$$\dim(\boldsymbol{P}_k^d(\mathcal{T}_h)) = \binom{k+2}{1}N_f + 2\binom{k+2}{2}N_c, \qquad \text{if } d = 2, \qquad (19.17b)$$

with the convention that $\binom{n}{m} := 0$ if $n < m$. Notice that the spaces $\boldsymbol{P}_k^c(\mathcal{T}_h)$ and $\boldsymbol{P}_k^d(\mathcal{T}_h)$ have the same dimension when $d = 2$. In the lowest-order case $(k = 0)$, we have $\dim(\boldsymbol{P}_0^d(\mathcal{T}_h)) = N_f$, and the connectivity array $\mathtt{j\_dof}$ coincides with the double-entry array $\mathtt{j\_cf}$ defined in §8.3.

19.3 Global interpolation operators

The goal of this section is to study the commuting and approximation properties of the global interpolation operators in the conforming finite element subspaces $P_k^x(\mathcal{T}_h; \mathbb{R}^q)$ with $x \in \{g, c, d\}$. Recall that $q = 1$ if $x = g$ and $q = d$ if $x \in \{c, d\}$ (and $d = 3$ if $x = c$). We start by introducing the global spaces

$$V^{x,b}(D) := \{v \in L^1(D; \mathbb{R}^q) \mid v_{|K} \in V^x(K), \ \forall K \in \mathcal{T}_h\}, \qquad (19.18a)$$

$$V^x(D) := \{v \in V^{x,b}(D) \mid [\![v]\!]_F^x = 0, \ \forall F \in \mathcal{F}_h^\circ\}, \qquad (19.18b)$$

where $V^{\mathrm{x}}(K)$ is the domain of the local interpolation operator $\mathcal{I}_K^{\mathrm{x}}$ (see Definition 5.7). For instance, owing to Theorem 18.8 and Theorem 18.10 and letting $p \in [1, \infty)$, admissible choices for these spaces are as follows:

$$V^{\mathrm{g}}(D) := W^{s,p}(D), \quad \text{with } s > \tfrac{d}{p} \text{ if } p > 1 \text{ or } s = d \text{ if } p = 1, \qquad (19.19\mathrm{a})$$

$$\boldsymbol{V}^{\mathrm{c}}(D) := \boldsymbol{W}^{s,p}(D), \quad \text{with } s > \tfrac{2}{p} \text{ if } p > 1 \text{ or } s = 2 \text{ if } p = 1, \qquad (19.19\mathrm{b})$$

$$\boldsymbol{V}^{\mathrm{d}}(D) := \boldsymbol{W}^{s,p}(D), \quad \text{with } s > \tfrac{1}{p} \text{ if } p > 1 \text{ or } s = 1 \text{ if } p = 1. \qquad (19.19\mathrm{c})$$

Recall that since Chapter 5 we have abused the notation regarding the definition of the dofs. In particular, we have used the same symbols to denote the dofs in $\mathcal{L}(P_K; \mathbb{R})$ and the extended dofs in $\mathcal{L}(V(K); \mathbb{R})$. We are going to be a little bit more careful in this chapter and in Chapters 20 and 21. More precisely, we are going to use the symbol $\sigma_{K,i}$ to denote dofs acting on functions in P_K and the symbol $\widetilde{\sigma}_{K,i}$ to denote the extension of $\sigma_{K,i}$ acting on functions in $V^{\mathrm{x}}(K)$. This means that the local interpolation operator $\mathcal{I}_K : V^{\mathrm{x}}(K) \to P_K$ is s.t.

$$\mathcal{I}_K(v)(\boldsymbol{x}) := \sum_{i \in \mathcal{N}} \widetilde{\sigma}_{K,i}(v) \theta_{K,i}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in K. \qquad (19.20)$$

We assume that the extension of the dofs is done in such a way that the following property holds true (compare with (19.2)): For all $v \in V^{\mathrm{x},\mathrm{b}}(D)$,

$$[\, v \in V^{\mathrm{x}}(D) \,] \implies \left[\begin{array}{l} \text{For all } (K, i), (K', i') \text{ in the same} \\ \text{connectivity class, we have} \\ \widetilde{\sigma}_{K,i}(v_{|K}) = \widetilde{\sigma}_{K',i'}(v_{|K'}) \end{array} \right]. \qquad (19.21)$$

In other words, for every function v in $V^{\mathrm{x},\mathrm{b}}(D)$, a *necessary* condition for v to be a member of the subspace $V^{\mathrm{x}}(D)$ is that, for all $a \in \mathcal{A}_h$, the quantity $\widetilde{\sigma}_{K,i}(v_{|K})$ is independent of the choice of the pair (K, i) in $\mathtt{j\_dof}^{-1}(a)$. (This condition is not sufficient since the knowledge of the values of $\{\widetilde{\sigma}_{K,i}(v_{|K})\}_{i \in \mathcal{N}}$ does not uniquely determine the function $v_{|K}$.) We then define the global interpolation operator $\mathcal{I}_h^{\mathrm{x}} : V^{\mathrm{x}}(D) \to P_k^{\mathrm{x}}(\mathcal{T}_h)$ s.t.

$$\mathcal{I}_h^{\mathrm{x}}(v)(\boldsymbol{x}) := \sum_{a \in \mathcal{A}_h} \widetilde{\sigma}_a(v) \varphi_a(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \overline{D}, \qquad (19.22)$$

where $\widetilde{\sigma}_a(v)$ is defined by setting $\widetilde{\sigma}_a(v) := \widetilde{\sigma}_{K,i}(v_{|K})$ for all (K, i) in the connectivity class a, i.e., $\mathtt{j\_dof}(K, i) = a$, which makes sense owing to (19.21). The definitions of $\widetilde{\sigma}_a$ and φ_a imply that

$$\mathcal{I}_h^{\mathrm{x}}(v)_{|K} = \sum_{i \in \mathcal{N}} \widetilde{\sigma}_{K,i}(v_{|K}) \theta_{K,i} = \mathcal{I}_K^{\mathrm{x}}(v_{|K}), \qquad \forall K \in \mathcal{T}_h. \qquad (19.23)$$

The above construction leads to the global interpolation operators:

$$\mathcal{I}_{k,h}^{\mathrm{L}} : V^{\mathrm{g}}(D) \to P_k^{\mathrm{g}}(\mathcal{T}_h), \qquad \mathcal{I}_{k,h}^{\mathrm{g}} : V^{\mathrm{g}}(D) \to P_k^{\mathrm{g}}(\mathcal{T}_h), \qquad \text{(19.24a)}$$

$$\mathcal{I}_{k,h}^{\mathrm{c}} : \boldsymbol{V}^{\mathrm{c}}(D) \to \boldsymbol{P}_k^{\mathrm{c}}(\mathcal{T}_h), \qquad \mathcal{I}_{k,h}^{\mathrm{d}} : \boldsymbol{V}^{\mathrm{d}}(D) \to \boldsymbol{P}_k^{\mathrm{d}}(\mathcal{T}_h), \qquad \text{(19.24b)}$$

for Lagrange, canonical hybrid, Nédélec, and Raviart–Thomas elements, respectively. We indicate explicitly the degree of the underlying finite element in the notation to avoid ambiguities. (Recall that $k \geq 1$ in (19.24a) and $k \geq 0$ in (19.24b).) Let us consider for $k \geq 0$ the L^2-orthogonal projection

$$\mathcal{I}_{k,h}^{\mathrm{b}} : V^{\mathrm{b}}(D) \to P_k^{\mathrm{b}}(\mathcal{T}_h) := \{ v_h \in L^\infty(D) \mid \psi_K^{\mathrm{b}}(v_{h|K}) \in \widehat{P}^{\mathrm{b}}, \, \forall K \in \mathcal{T}_h \},$$
$$\text{(19.25)}$$

where $V^{\mathrm{b}}(D) := L^1(D)$, $\psi_K^{\mathrm{b}}(v) := \det(\mathbb{J}_K)(v \circ \boldsymbol{T}_K)$, and $\widehat{P}^{\mathrm{b}} := \mathbb{P}_{k,d}$ if $\widehat{K}$ is a simplex and $\widehat{P}^{\mathrm{b}} := \mathbb{Q}_{k,d}$ is $\widehat{K}$ is a cuboid. Note that since the mesh is affine, the factor $\det(\mathbb{J}_K)$ is irrelevant in the definition of $P_k^{\mathrm{b}}(\mathcal{T}_h)$.

Lemma 19.6 (de Rham complex).   *Let us set*

$$\check{V}^{\mathrm{g}}(D) := \{ f \in V^{\mathrm{g}}(D) \mid \nabla f \in \boldsymbol{V}^{\mathrm{c}}(D) \}, \qquad \text{(19.26a)}$$

$$\check{\boldsymbol{V}}^{\mathrm{c}}(D) := \{ \boldsymbol{g} \in \boldsymbol{V}^{\mathrm{c}}(D) \mid \nabla \times \boldsymbol{g} \in \boldsymbol{V}^{\mathrm{d}}(D) \}, \qquad \text{(19.26b)}$$

$$\check{\boldsymbol{V}}^{\mathrm{d}}(D) := \{ \boldsymbol{g} \in \boldsymbol{V}^{\mathrm{d}}(D) \mid \nabla \cdot \boldsymbol{g} \in V^{\mathrm{b}}(D) \}. \qquad \text{(19.26c)}$$

Let $\kappa \in \mathbb{N}$. The following diagrams commute:

$$
\begin{array}{ccccccc}
\check{V}^{\mathrm{g}}(D) & \xrightarrow{\ \nabla\ } & \check{\boldsymbol{V}}^{\mathrm{c}}(D) & \xrightarrow{\ \nabla\times\ } & \check{\boldsymbol{V}}^{\mathrm{d}}(D) & \xrightarrow{\ \nabla\cdot\ } & V^{\mathrm{b}}(D) \\
\downarrow{\scriptstyle \mathcal{I}_{\kappa+1,h}^{\mathrm{g}}} & & \downarrow{\scriptstyle \mathcal{I}_{\kappa,h}^{\mathrm{c}}} & & \downarrow{\scriptstyle \mathcal{I}_{\kappa,h}^{\mathrm{d}}} & & \downarrow{\scriptstyle \mathcal{I}_{\kappa,h}^{\mathrm{b}}} \\
P_{\kappa+1}^{\mathrm{g}}(\mathcal{T}_h) & \xrightarrow{\ \nabla\ } & \boldsymbol{P}_\kappa^{\mathrm{c}}(\mathcal{T}_h) & \xrightarrow{\ \nabla\times\ } & \boldsymbol{P}_\kappa^{\mathrm{d}}(\mathcal{T}_h) & \xrightarrow{\ \nabla\cdot\ } & P_\kappa^{\mathrm{b}}(\mathcal{T}_h)
\end{array}
\qquad \text{(19.27)}
$$

Proof. Combine Lemma 16.16 (and Remark 16.18) with (19.23). □

Remark 19.7 (Interpolation with extended domain). The commuting diagram (19.27) shows that we can extend the domain of $\mathcal{I}_{\kappa,h}^{\mathrm{c}}$ to $\widetilde{\boldsymbol{V}}^{\mathrm{c}}(D) := \boldsymbol{V}^{\mathrm{c}}(D) + \nabla V^{\mathrm{g}}(D)$, that of $\mathcal{I}_{\kappa,h}^{\mathrm{d}}$ to $\widetilde{\boldsymbol{V}}^{\mathrm{d}}(D) := \boldsymbol{V}^{\mathrm{d}}(D) + \nabla \times \boldsymbol{V}^{\mathrm{c}}(D)$, and that of $\mathcal{I}_{\kappa,h}^{\mathrm{b}}$ to $\widetilde{V}^{\mathrm{b}}(D) := \boldsymbol{V}^{\mathrm{d}}(D) + \nabla \cdot V^{\mathrm{b}}(D)$. Keeping the same notation for the differential operators, this leads to the following commuting diagrams:

$$
\begin{array}{ccccccc}
V^{\mathrm{g}}(D) & \xrightarrow{\ \nabla\ } & \widetilde{\boldsymbol{V}}^{\mathrm{c}}(D) & \xrightarrow{\ \nabla\times\ } & \widetilde{\boldsymbol{V}}^{\mathrm{d}}(D) & \xrightarrow{\ \nabla\cdot\ } & \widetilde{V}^{\mathrm{b}}(D) \\
\downarrow{\scriptstyle \mathcal{I}_{\kappa+1,h}^{\mathrm{g}}} & & \downarrow{\scriptstyle \mathcal{I}_{\kappa,h}^{\mathrm{c}}} & & \downarrow{\scriptstyle \mathcal{I}_{\kappa,h}^{\mathrm{d}}} & & \downarrow{\scriptstyle \mathcal{I}_{\kappa,h}^{\mathrm{b}}} \\
P_{\kappa+1}^{\mathrm{g}}(\mathcal{T}_h) & \xrightarrow{\ \nabla\ } & \boldsymbol{P}_\kappa^{\mathrm{c}}(\mathcal{T}_h) & \xrightarrow{\ \nabla\times\ } & \boldsymbol{P}_\kappa^{\mathrm{d}}(\mathcal{T}_h) & \xrightarrow{\ \nabla\cdot\ } & P_\kappa^{\mathrm{b}}(\mathcal{T}_h)
\end{array}
\qquad \text{(19.28)}
$$

For instance, for all $\boldsymbol{v} = \boldsymbol{w} + \nabla\psi \in \widetilde{\boldsymbol{V}}^{\mathrm{c}}(D)$ with $\boldsymbol{w} \in \boldsymbol{V}^{\mathrm{c}}(D)$ and $\psi \in V^{\mathrm{g}}(D)$, we set $\mathcal{I}_{\kappa,h}^{\mathrm{c}}(\boldsymbol{v}) := \mathcal{I}_{\kappa,h}^{\mathrm{c}}(\boldsymbol{w}) + \nabla \mathcal{I}_{\kappa+1,h}^{\mathrm{g}}(\psi)$. To verify that $\mathcal{I}_{\kappa,h}^{\mathrm{c}}(\boldsymbol{v})$ is well defined, we observe that if $\boldsymbol{v} = \boldsymbol{w}_1 + \nabla\psi_1 = \boldsymbol{w}_2 + \nabla\psi_2$, then $\psi_1 - \psi_2 \in \check{V}^{\mathrm{g}}(D)$ so

that $\nabla(\mathcal{I}^{\mathrm{g}}_{\kappa+1,h}(\psi_1 - \psi_2)) = \mathcal{I}^{\mathrm{c}}_{\kappa,h}(\nabla(\psi_1 - \psi_2)) = \mathcal{I}^{\mathrm{c}}_{\kappa,h}(\boldsymbol{w}_2 - \boldsymbol{w}_1)$. Thus, we have $\mathcal{I}^{\mathrm{c}}_{\kappa,h}(\boldsymbol{w}_1) + \nabla\mathcal{I}^{\mathrm{g}}_{\kappa+1,h}(\psi_1) = \mathcal{I}^{\mathrm{c}}_{\kappa,h}(\boldsymbol{w}_2) + \nabla\mathcal{I}^{\mathrm{g}}_{\kappa+1,h}(\psi_2)$. $\qquad\square$

Let us now turn to the approximation properties of the global interpolation operators defined in (19.24). Henceforth, the subscript k is omitted when the context is unambiguous. The following results follow from the localization property (19.23) combined with the corresponding local interpolation results, and from Lemma 19.6 for the approximation properties on the divergence and the curl.

Corollary 19.8 (H^1-conforming interpolation). *Let $(\mathcal{T}_h)_{h\in\mathcal{H}}$ be a shape-regular sequence of affine matching meshes. Let $p \in [1,\infty]$. Let l be the smallest integer such that $l > \frac{d}{p}$ if $p > 1$ or $l := d$ if $p = 1$. The following estimates hold true, uniformly w.r.t. p, with either $\mathcal{I}_h = \mathcal{I}^{\mathrm{g}}_h : V^{\mathrm{g}}(D) \to P^{\mathrm{g}}_k(\mathcal{T}_h)$ or $\mathcal{I}_h = \mathcal{I}^{\mathrm{L}}_h : V^{\mathrm{g}}(D) \to P^{\mathrm{g}}_k(\mathcal{T}_h)$, $k \geq 1$:*
(i) If $l \leq k + 1$, then for every integers $r \in \{l{:}k + 1\}$ and $m \in \{0{:}r\}$, all $v \in W^{r,p}(D)$, and all $h \in \mathcal{H}$,

$$|v - \mathcal{I}_h(v)|_{W^{m,p}(\mathcal{T}_h)} \leq c \left(\sum_{K\in\mathcal{T}_h} h_K^{p(r-m)} |v|^p_{W^{r,p}(K)} \right)^{\frac{1}{p}}, \qquad (19.29)$$

for $p < \infty$, and $|v - \mathcal{I}_h(v)|_{W^{m,\infty}(\mathcal{T}_h)} \leq c \max_{K\in\mathcal{T}_h} h_K^{r-m} |v|_{W^{r,\infty}(K)}$.
(ii) If $l > k + 1$, then for every integer $m \in \{0{:}k + 1\}$, all $v \in W^{l,p}(D)$, and all $h \in \mathcal{H}$,

$$|v - \mathcal{I}_h(v)|_{W^{m,p}(\mathcal{T}_h)} \leq c \left(\sum_{K\in\mathcal{T}_h} \sum_{n\in\{k+1:l\}} h_K^{p(n-m)} |v|^p_{W^{n,p}(K)} \right)^{\frac{1}{p}}, \qquad (19.30)$$

for $p < \infty$, and $|v - \mathcal{I}_h(v)|_{W^{m,\infty}(\mathcal{T}_h)} \leq c \max_{K\in\mathcal{T}_h, n\in\{k+1:l\}} h_K^{n-m} |v|_{W^{n,\infty}(K)}$.

Corollary 19.9 ($\boldsymbol{H}(\mathrm{curl})$-conforming interpolation). *Let $(\mathcal{T}_h)_{h\in\mathcal{H}}$ be a shape-regular sequence of affine matching meshes. Let $p \in [1,\infty]$ and let $l := 1$ if $p > 2$ and $l := 2$ if $p \in [1,2]$. The following holds true, uniformly w.r.t. p, with $\mathcal{I}^{\mathrm{c}}_h : \boldsymbol{V}^{\mathrm{c}}(D) \to \boldsymbol{P}^{\mathrm{c}}_k(\mathcal{T}_h)$, $k \geq 0$:*
(i) If $p > 2$ or if $p \in [1,2]$ and $k \geq 1$, then for every integers $r \in \{l{:}k + 1\}$ and $m \in \{0{:}r\}$, all $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(D)$, and all $h \in \mathcal{H}$,

$$|\boldsymbol{v} - \mathcal{I}^{\mathrm{c}}_h(\boldsymbol{v})|_{\boldsymbol{W}^{m,p}(\mathcal{T}_h)} \leq c \left(\sum_{K\in\mathcal{T}_h} h_K^{p(r-m)} |\boldsymbol{v}|^p_{\boldsymbol{W}^{r,p}(K)} \right)^{\frac{1}{p}}, \qquad (19.31)$$

for $p < \infty$, and $|\boldsymbol{v} - \mathcal{I}^{\mathrm{c}}_h(\boldsymbol{v})|_{\boldsymbol{W}^{m,\infty}(\mathcal{T}_h)} \leq c \max_{K\in\mathcal{T}_h} h_K^{r-m} |\boldsymbol{v}|_{\boldsymbol{W}^{r,\infty}(K)}$.
(ii) If $p \in [1,2]$ and $k = 0$, then for every integer $m \in \{0{:}1\}$, all $\boldsymbol{v} \in \boldsymbol{W}^{2,p}(D)$, and all $h \in \mathcal{H}$,

$$|\boldsymbol{v} - \mathcal{I}_h^{\rm c}(\boldsymbol{v})|_{\boldsymbol{W}^{m,p}(\mathcal{T}_h)} \le c \left(\sum_{K \in \mathcal{T}_h} \sum_{n \in \{1,2\}} h_K^{p(n-m)} |\boldsymbol{v}|_{\boldsymbol{W}^{n,p}(K)}^p \right)^{\frac{1}{p}}, \qquad (19.32)$$

for $p < \infty$, and $|\boldsymbol{v} - \mathcal{I}_h^{\rm c}(\boldsymbol{v})|_{\boldsymbol{W}^{m,\infty}(\mathcal{T}_h)} \le c \max_{K \in \mathcal{T}_h, n \in \{1,2\}} h_K^{n-m} |\boldsymbol{v}|_{\boldsymbol{W}^{n,\infty}(K)}$.
(iii) *For every integers* $r \in \{1{:}k+1\}$ *and* $m \in \{0{:}r\}$, *all* $\boldsymbol{v} \in \boldsymbol{V}^{\rm c}(D)$ *with* $\nabla \times \boldsymbol{v} \in \boldsymbol{W}^{r,p}(D)$, *and all* $h \in \mathcal{H}$,

$$|\nabla \times (\boldsymbol{v} - \mathcal{I}_h^{\rm c}(\boldsymbol{v}))|_{\boldsymbol{W}^{m,p}(\mathcal{T}_h)} \le c \left(\sum_{K \in \mathcal{T}_h} h_K^{p(r-m)} |\nabla \times \boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}^p \right)^{\frac{1}{p}}, \qquad (19.33)$$

for $p < \infty$, and $|\nabla \times (\boldsymbol{v} - \mathcal{I}_h^{\rm c}(\boldsymbol{v}))|_{\boldsymbol{W}^{m,\infty}(\mathcal{T}_h)} \le c \max_{K \in \mathcal{T}_h} h_K^{r-m} |\nabla \times \boldsymbol{v}|_{\boldsymbol{W}^{r,\infty}(K)}$.

Corollary 19.10 ($\boldsymbol{H}$(div)-conforming interpolation). *Let* $(\mathcal{T}_h)_{h \in \mathcal{H}}$ *be a shape-regular sequence of affine matching meshes. Let* $p \in [1,\infty]$. *The following holds true, uniformly w.r.t.* p, *with* $\mathcal{I}_h^{\rm d} : \boldsymbol{V}^{\rm d}(D) \to \boldsymbol{P}_k^{\rm d}(\mathcal{T}_h)$, $k \ge 0$:
(i) *For every integers* $r \in \{1{:}k+1\}$ *and* $m \in \{0{:}r\}$, *all* $\boldsymbol{v} \in \boldsymbol{W}^{r,p}(D)$, *and all* $h \in \mathcal{H}$,

$$|\boldsymbol{v} - \mathcal{I}_h^{\rm d}(\boldsymbol{v})|_{\boldsymbol{W}^{m,p}(\mathcal{T}_h)} \le c \left(\sum_{K \in \mathcal{T}_h} h_K^{p(r-m)} |\boldsymbol{v}|_{\boldsymbol{W}^{r,p}(K)}^p \right)^{\frac{1}{p}}, \qquad (19.34)$$

for $p < \infty$, and $|\boldsymbol{v} - \mathcal{I}_h^{\rm d}(\boldsymbol{v})|_{\boldsymbol{W}^{m,\infty}(\mathcal{T}_h)} \le c \max_{K \in \mathcal{T}_h} h_K^{r-m} |\boldsymbol{v}|_{\boldsymbol{W}^{r,\infty}(K)}$.
(ii) *For every integers* $r \in \{0{:}k+1\}$ *and* $m \in \{0{:}r\}$, *all* $\boldsymbol{v} \in \boldsymbol{V}^{\rm d}(D)$ *with* $\nabla \cdot \boldsymbol{v} \in W^{r,p}(D)$, *and all* $h \in \mathcal{H}$,

$$|\nabla \cdot (\boldsymbol{v} - \mathcal{I}_h^{\rm d}(\boldsymbol{v}))|_{W^{m,p}(\mathcal{T}_h)} \le c \left(\sum_{K \in \mathcal{T}_h} h_K^{p(r-m)} |\nabla \cdot \boldsymbol{v}|_{W^{r,p}(K)}^p \right)^{\frac{1}{p}}, \qquad (19.35)$$

for $p < \infty$, and $|\nabla \cdot (\boldsymbol{v} - \mathcal{I}_h^{\rm d}(\boldsymbol{v}))|_{W^{m,\infty}(\mathcal{T}_h)} \le c \max_{K \in \mathcal{T}_h} h_K^{r-m} |\nabla \cdot \boldsymbol{v}|_{W^{\infty,p}(K)}$.

19.4 Subspaces with zero boundary trace

In this section, we briefly review the main changes to be applied when one wishes to enforce homogeneous boundary conditions to the functions in $P_k^{\rm x}(\mathcal{T}_h)$. Let $p \in [1,\infty)$ and let $s > \frac{d}{p}$ if $p > 1$ and $s = d$ if $p = 1$. We consider the trace operator $\gamma^{\rm x} : W^{s,p}(\mathcal{T}_h; \mathbb{R}^q) \longrightarrow L^1(\partial D; \mathbb{R}^t)$ defined by

$$\gamma^{\rm g}(v) := v_{|\partial D} \qquad (q = t = 1), \qquad (19.36a)$$

$$\gamma^{\rm c}(\boldsymbol{v}) := \boldsymbol{v}_{|\partial D} \times \boldsymbol{n} \quad (q = t = d = 3), \qquad (19.36b)$$

$$\gamma^{\rm d}(\boldsymbol{v}) := \boldsymbol{v}_{|\partial D} \cdot \boldsymbol{n} \qquad (q = d, t = 1), \qquad (19.36c)$$

where $\boldsymbol{n}$ is the outward unit normal to D. Notice that $\gamma^{\mathrm{x}}(v)_{|F} = \gamma^{\mathrm{x}}_{K_l,F}(v_{|K_l})$ for all $F \in \mathcal{F}_h^\partial$ with $F := \partial K_l \cap \partial D$ and $\gamma^{\mathrm{x}}_{K_l,F}$ is the operator defined in (18.7) for the mesh cell K_l. We are interested in the following subspace of $P_k^{\mathrm{x}}(\mathcal{T}_h)$:

$$P_{k,0}^{\mathrm{x}}(\mathcal{T}_h) := \{v_h \in P_k^{\mathrm{x}}(\mathcal{T}_h) \mid \gamma^{\mathrm{x}}(v) = 0\}. \qquad (19.37)$$

Definition 19.11 (Boundary & internal classes). *We say that a connectivity class $a \in \mathcal{A}_h$ is a* boundary connectivity class *if and only if $\sigma_a(v) = 0$ for all $v \in P_{k,0}^{\mathrm{x}}(\mathcal{T}_h)$. The collection of boundary connectivity classes is denoted by $\mathcal{A}_h^\partial$. The classes in $\mathcal{A}_h^\circ := \mathcal{A}_h \setminus \mathcal{A}_h^\partial$ are called* internal connectivity classes.

We assume that the following properties hold true:

$$\forall v_h \in P_k^{\mathrm{x}}(\mathcal{T}_h), \qquad [\,\gamma^{\mathrm{x}}(v_h) = 0\,] \iff [\,\sigma_a(v_h) = 0,\ \forall a \in \mathcal{A}_h^\partial\,], \quad (19.38\mathrm{a})$$

$$\forall v \in V^{\mathrm{x}}(D), \qquad [\,\gamma^{\mathrm{x}}(v) = 0\,] \implies [\,\widetilde{\sigma}_a(v) = 0,\ \forall a \in \mathcal{A}_h^\partial\,]. \qquad (19.38\mathrm{b})$$

We are going to show in Chapters 20 and 21 that these properties are indeed satisfied by most of the finite elements considered in this book.

Example 19.12 ($\mathcal{A}_h^\partial$). For Lagrange elements, $a \in \mathcal{A}_h^\partial$ iff σ_a is an evaluation at a node located on ∂D. For canonical hybrid elements, $a \in \mathcal{A}_h^\partial$ iff σ_a is an evaluation at a vertex located on ∂D, or σ_a is an integral over an edge or a face located on ∂D. For Nédélec elements, $a \in \mathcal{A}_h^\partial$ iff σ_a is an integral over an edge or a face located on ∂D, and for Raviart–Thomas elements, $a \in \mathcal{A}_h^\partial$ iff σ_a is an integral over a face located on ∂D. $\qquad\square$

Proposition 19.13 (Basis). *$\{\varphi_a\}_{a \in \mathcal{A}_h^\circ}$ is a basis of $P_{k,0}^{\mathrm{x}}(\mathcal{T}_h)$, and $\{\sigma_a\}_{a \in \mathcal{A}_h^\circ}$ is a basis of $\mathcal{L}(P_{k,0}^{\mathrm{x}}(\mathcal{T}_h); \mathbb{R})$.*

Proof. See Exercise 19.3. $\qquad\square$

Let $V^{\mathrm{x}}(D)$ be defined in (19.19). Since functions in $V^{\mathrm{x}}(D)$ have a γ^{x}-trace on ∂D, it is legitimate to set

$$V_0^{\mathrm{x}}(D) := \{v \in V^{\mathrm{x}}(D) \mid \gamma^{\mathrm{x}}(v) = 0\}. \qquad (19.39)$$

The interpolation operator with prescribed boundary conditions $\mathcal{I}_{h0}^{\mathrm{x}} : V_0^{\mathrm{x}}(D) \to P_{k,0}^{\mathrm{x}}(\mathcal{T}_h)$ acts as follows:

$$\mathcal{I}_{h0}^{\mathrm{x}}(v)(\boldsymbol{x}) := \sum_{a \in \mathcal{A}_h^\circ} \widetilde{\sigma}_a(v)\varphi_a(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \overline{D}, \qquad (19.40)$$

and (19.38b) implies that

$$\mathcal{I}_{h0}^{\mathrm{x}}(v) = \mathcal{I}_h^{\mathrm{x}}(v), \qquad \forall v \in V_0^{\mathrm{x}}(D). \qquad (19.41)$$

Hence, the approximation properties of $\mathcal{I}_{h0}^{\mathrm{x}}$ are identical to those of the restriction of $\mathcal{I}_h^{\mathrm{x}}$ to $V_0^{\mathrm{x}}(D)$. Moreover, we have the following commuting properties.

Lemma 19.14 (de Rham complex with boundary prescription). *Let* $\check{V}_0^x(D) := \{v \in \check{V}^x(D) \mid \gamma^x(v) = 0\}$ *with* $\check{V}^x(D)$ *defined in* (19.26), *and*

$$V_0^b(D) := \{v \in V^b(D) := L^1(D) \mid (v, 1)_{L^2(D)} = 0\}, \tag{19.42a}$$

$$P_{\kappa,0}^b(\mathcal{T}_h) := \{v_h \in P_\kappa^b(\mathcal{T}_h) \mid (v_h, 1)_{L^2(D)} = 0\}. \tag{19.42b}$$

Let $\kappa \in \mathbb{N}$. *The following diagrams commute:*

$$
\begin{array}{ccccccc}
\check{V}_0^g(D) & \xrightarrow{\nabla} & \check{\boldsymbol{V}}_0^c(D) & \xrightarrow{\nabla\times} & \check{\boldsymbol{V}}_0^d(D) & \xrightarrow{\nabla\cdot} & V_0^b(D) \\
\downarrow{\scriptstyle\mathcal{I}_{\kappa+1,h0}^g} & & \downarrow{\scriptstyle\mathcal{I}_{\kappa,h0}^c} & & \downarrow{\scriptstyle\mathcal{I}_{\kappa,h0}^d} & & \downarrow{\scriptstyle\mathcal{I}_h^b} \\
P_{\kappa+1,0}^g(\mathcal{T}_h) & \xrightarrow{\nabla} & \boldsymbol{P}_{\kappa,0}^c(\mathcal{T}_h) & \xrightarrow{\nabla\times} & \boldsymbol{P}_{\kappa,0}^d(\mathcal{T}_h) & \xrightarrow{\nabla\cdot} & P_{\kappa,0}^b(\mathcal{T}_h)
\end{array}
\tag{19.43}
$$

Proof. Observe that the tangential boundary trace of ∇f is zero if $\gamma^g(f) = 0$ and that the normal trace of $\nabla\times\boldsymbol{g}$ is zero if $\gamma^c(\boldsymbol{g}) = \boldsymbol{0}$. $\qquad\square$

Remark 19.15 (Extensions). The above argumentation can be adapted to enforce a zero trace on a part of the boundary that corresponds to a strict subset of the boundary faces in $\mathcal{F}_h^\partial$. The details are left to the reader. Furthermore, the commuting diagram (19.43) can be rewritten by using the spaces $V_0^g(D)$, $\boldsymbol{V}_0^c(D)+\nabla V_0^g(D)$, $\boldsymbol{V}_0^d(D)+\nabla\times\boldsymbol{V}_0^c(D)$, and $\boldsymbol{V}_0^b(D)+\nabla\cdot\boldsymbol{V}_0^d(D)$ instead of $\check{V}_0^g(D)$, $\check{\boldsymbol{V}}_0^c(D)$, $\check{\boldsymbol{V}}_0^d(D)$, $V_0^b(D)$. $\qquad\square$

Exercises

Exercise 19.1 (Connectivity classes). Consider the mesh shown in Figure 19.4 and let $P_2^g(\mathcal{T}_h)$ be the associated finite element space composed of continuous Lagrange $\mathbb{P}_2$ finite elements. Assume that the enumeration of the Lagrange nodes has been done with the increasing vertex-index technique (see (10.10)). (i) What is the domain and the codomain of j_dof? (ii) Identify j_dof^{-1}(8) and j_dof^{-1}(13). (iii) Identify $\mathcal{T}_6$ and $\mathcal{T}_{10}$.
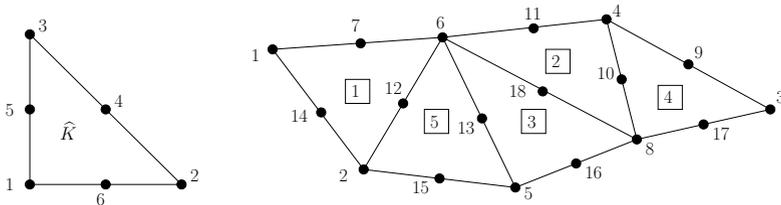
Fig. 19.4 Illustration for Exercise 19.1.

Exercise 19.2 (Stiffness, mass, incidence matrices). Let $\{\lambda_n\}_{n\in\{1:N_{\mathrm{v}}\}}$ be the global shape functions in $P_1^{\mathrm{g}}(\mathcal{T}_h)$. Let $\{\boldsymbol{\theta}_m\}_{m\in\{1:N_{\mathrm{e}}\}}$ be the global shape functions in $\boldsymbol{P}_0^{\mathrm{c}}(\mathcal{T}_h)$. (i) Recall the incidence matrix $\mathcal{M}^{\mathrm{ev}} \in \mathbb{R}^{N_{\mathrm{e}}\times N_{\mathrm{v}}}$ defined in Remark 10.2. Prove that $\nabla\lambda_n = \sum_{m\in\{1:N_{\mathrm{e}}\}} \mathcal{M}^{\mathrm{ev}}_{mn}\boldsymbol{\theta}_m$ for all $n \in \{1:N_{\mathrm{v}}\}$. (*Hint*: compute $\sigma_m^{\mathrm{e}}(\nabla\lambda_n)$ where $\{\sigma_m^{\mathrm{e}}\}_{m\in\{1:N_{\mathrm{e}}\}}$ is the dual basis of $\{\boldsymbol{\theta}_m\}_{m\in\{1:N_{\mathrm{e}}\}}$, i.e., the associated dofs.) (ii) Let $\mathcal{A} \in \mathbb{R}^{N_{\mathrm{v}}\times N_{\mathrm{v}}}$ be the Courant stiffness matrix with entries $\mathcal{A}_{nn'} := \int_D \nabla\lambda_n\cdot\nabla\lambda_{n'}\,\mathrm{d}x$ for all $n, n' \in \{1:N_{\mathrm{v}}\}$, and let $\mathcal{N} \in \mathbb{R}^{N_{\mathrm{e}}\times N_{\mathrm{e}}}$ be the Nédélec mass matrix with entries $\mathcal{N}_{mm'} := \int_D \boldsymbol{\theta}_m\cdot\boldsymbol{\theta}_{m'}\,\mathrm{d}x$ for all $m, m' \in \{1:N_{\mathrm{e}}\}$. Prove that $\mathcal{A} = (\mathcal{M}^{\mathrm{ev}})^{\mathsf{T}}\mathcal{N}\mathcal{M}^{\mathrm{ev}}$.

Exercise 19.3 (Zero trace). (i) Show that $\varphi_a \in P_{k,0}^{\mathrm{x}}(\mathcal{T}_h)$ for all $a \in \mathcal{A}_h^{\circ}$. (ii) Prove Proposition 19.13.

Exercise 19.4 (Approximability in L^p). Let $p \in [1,\infty)$. Prove that $\lim_{h\downarrow 0} \inf_{v_h\in P_k^{\mathrm{g}}(\mathcal{T}_h)} \|v - v_h\|_{L^p(D)} = 0$ for all $v \in L^p(D)$. (*Hint*: by density.)

Exercise 19.5 (Hermite). Let $\mathcal{T}_h := \{[x_i, x_{i+1}]\}_{i\in\{0:I\}}$ be a mesh of the interval $D := (a, b)$. Recall the Hermite finite element from Exercise 5.4. Specify global shape functions $\{\varphi_{i,0}, \varphi_{i,1}\}_{i\in\{0:I+1\}}$ in $H_h := \{v_h \in C^1(\overline{D}) \mid \forall i \in \{0:I\}, v_{h|[x_i,x_{i+1}]} \in \mathbb{P}_3\}$. (*Hint*: consider values of the function or of its derivative at the mesh nodes.) Can the bicubic Hermite rectangular finite element from Exercise 6.8 be used to enforce C^1-continuity for $d = 2$?

Part IV, Chapter 20

Face gluing

The goal of this chapter and the following one is to construct the connectivity array j_dof introduced in the previous chapter so that the two structural properties (19.2) and (19.3) hold true. In the present chapter, we focus on (19.2), and more specifically we are going to see how we can enforce the zero-jump condition $[\![v_h]\!]_F^{\mathrm{x}} = 0$ by means of the degrees of freedom (dofs) on the two mesh cells sharing the interface $F \in \mathcal{F}_h^\circ$ for v_h in the broken finite element space $P_k^{\mathrm{x,b}}(\mathcal{T}_h)$. In particular, we identify two key structural assumptions on the dofs of the finite element making this construction possible. The first assumption is called face unisolvence (see Assumption 20.1), and the second one is called face matching (see Assumption 20.3). We first introduce these ideas with Lagrange elements to make the argumentation easier to understand. Then we generalize the concepts to the Nédélec and the Raviart–Thomas finite elements in a unified setting that encompasses all the finite elements considered in the book. The two main results of this chapter are Lemma 20.4 for Lagrange elements and Lemma 20.15 for the general situation. In the entire chapter, D is a polyhedron in $\mathbb{R}^d$ and $\mathcal{T}_h$ is an oriented matching mesh covering D exactly (see Chapter 10 on mesh orientation).

20.1 The two gluing assumptions (Lagrange)

For Lagrange elements our aim is to construct the H^1-conforming subspace

$$P_k^{\mathrm{g}}(\mathcal{T}_h) := \{v_h \in P_k^{\mathrm{g,b}}(\mathcal{T}_h) \mid [\![v_h]\!]_F^{\mathrm{g}} = 0, \ \forall F \in \mathcal{F}_h^\circ\}, \tag{20.1}$$

where $P_k^{\mathrm{g,b}}(\mathcal{T}_h)$ is a broken finite element space and $[\![\cdot]\!]_F^{\mathrm{g}} := [\![\cdot]\!]_F$ is the jump operator across the mesh interfaces introduced in Definition 8.10. Recall that we have $P_k^{\mathrm{g}}(\mathcal{T}_h) = P_k^{\mathrm{g,b}}(\mathcal{T}_h) \cap H^1(D)$.

The Lagrange nodes of the reference cell $\widehat{K}$ are denoted by $\{\widehat{\boldsymbol{a}}_i\}_{i\in\mathcal{N}}$ so that the dofs $\widehat{\Sigma} := \{\widehat{\sigma}_i\}_{i\in\mathcal{N}}$ are s.t. $\widehat{\sigma}_i(\widehat{p}) := \widehat{p}(\widehat{\boldsymbol{a}}_i)$ for all $i \in \mathcal{N}$ and all $\widehat{p} \in \widehat{P}$. The

Lagrange nodes of $K \in \mathcal{T}_h$ are denoted by $\{\boldsymbol{a}_{K,i} := \boldsymbol{T}_K(\widehat{\boldsymbol{a}}_i)\}_{i \in \mathcal{N}}$, where $\boldsymbol{T}_K :$ $\widehat{K} \to K$ is the geometric mapping. The dofs in K are s.t. $\sigma_{K,i}(p) = p(\boldsymbol{a}_{K,i})$ for all $i \in \mathcal{N}$ and all $p \in P_K$ with $P_K := (\psi_K^{\mathrm{g}})^{-1}(\widehat{P})$, where $\psi_K^{\mathrm{g}}(v) := v \circ \boldsymbol{T}_K$ is the pullback by the geometric mapping. We do not assume in this section that the geometric mapping $\boldsymbol{T}_K$ is affine.

We now formalize the structure of the reference element that will allow us to enforce the zero-jump condition in (20.1). We make two assumptions which we will show hold true in the next section for the simplicial and the tensor-product Lagrange elements. Our first key assumption is the following.

Assumption 20.1 (Face unisolvence). *Let $\widehat{F}$ be a face of $\widehat{K}$, i.e., $\widehat{F} \in \mathcal{F}_{\widehat{K}}$, and let $\mathcal{N}_{\widehat{K},\widehat{F}} \subset \mathcal{N}$ be the collection of the indices of the Lagrange nodes in $\widehat{K}$ located on $\widehat{F}$. We assume that*

$$\forall \widehat{p} \in \widehat{P}, \qquad [\,\widehat{\sigma}_i(\widehat{p}) = 0, \,\forall i \in \mathcal{N}_{\widehat{K},\widehat{F}}\,] \iff [\,\widehat{p}_{|\widehat{F}} = 0\,]. \qquad (20.2)$$

Let K be a mesh cell and let F be a face of $K.$, i.e., $F \in \mathcal{F}_K$. Let $\widehat{F}$ be the face of $\widehat{K}$ s.t. $\widehat{F} := \boldsymbol{T}_K^{-1}(F)$. Let $\mathcal{N}_{K,F} \subset \mathcal{N}$ be the collection of the indices of the Lagrange nodes in K located on F. The above definitions imply that

$$[\,i \in \mathcal{N}_{K,F}\,] \iff [\,\boldsymbol{a}_{K,i} \in F\,] \iff [\,\widehat{\boldsymbol{a}}_i \in \widehat{F}\,] \iff [\,i \in \mathcal{N}_{\widehat{K},\widehat{F}}\,], \qquad (20.3)$$

that is, we have

$$\mathcal{N}_{K,F} = \mathcal{N}_{\widehat{K},\widehat{F}} = \mathcal{N}_{\widehat{K},\boldsymbol{T}_K^{-1}(F)}, \qquad \forall K \in \mathcal{T}_h, \,\forall F \in \mathcal{F}_K. \qquad (20.4)$$

We define the trace space $P_{K,F} := \operatorname{span}\{\theta_{K,i|F}\}_{i \in \mathcal{N}_{K,F}}$, so that $P_{K,F} = \gamma_{K,F}^{\mathrm{g}}(P_K)$, where we recall that the trace map $\gamma_{K,F}^{\mathrm{g}}$ is defined by setting $\gamma_{K,F}^{\mathrm{g}}(v) := v_{|F}$ for all $v \in P_K$. We define the set of the dofs associated with the Lagrange nodes located on F, $\Sigma_{K,F} := \{\sigma_{K,F,i}\}_{i \in \mathcal{N}_{K,F}}$, by setting $\sigma_{K,F,i}(q) := q(\boldsymbol{a}_{K,i})$ for all $i \in \mathcal{N}_{K,F}$ and all $q \in P_{K,F}$. Notice that $\sigma_{K,F,i}$ acts on functions in $P_{K,F}$ (i.e., functions defined on F), whereas $\sigma_{K,i}$ acts on functions in P_K (i.e., functions defined on K).

Let us state an important consequence of Assumption 20.1.

Lemma 20.2 (Face element). *Let $K \in \mathcal{T}_h$ and $F \in \mathcal{F}_K$. Under Assumption 20.1, the triple $(F, P_{K,F}, \Sigma_{K,F})$ is a finite element.*

Proof. We use Remark 5.3 to prove unisolvence. Since we have

$$\sigma_{K,F,j}(\theta_{K,i|F}) = \theta_{K,i|F}(\boldsymbol{a}_{K,j}) = \theta_{K,i}(\boldsymbol{a}_{K,j}) = \delta_{ij},$$

for all $i, j \in \mathcal{N}_{K,F}$, we infer that the family $\{\theta_{K,i|F}\}_{i \in \mathcal{N}_{K,F}}$ is linearly independent, which implies that $\dim(P_{K,F}) = \operatorname{card}(\Sigma_{K,F})$. Let now $q \in P_{K,F}$ be s.t. $\sigma_{K,F,i}(q) = 0$ for all $i \in \mathcal{N}_{K,F}$. By definition of $P_{K,F}$ and P_K, there is $\widehat{p} \in \widehat{P}$ s.t. $q = (\widehat{p} \circ \boldsymbol{T}_K^{-1})_{|F}$. Hence, for all $i \in \mathcal{N}_{\widehat{K},\widehat{F}} = \mathcal{N}_{K,F}$, we have $\boldsymbol{a}_{K,i} \in F$ and

$$\widehat{\sigma}_i(\widehat{p}) = \widehat{p}(\widehat{\boldsymbol{a}}_i) = (\widehat{p} \circ T_K^{-1})(\boldsymbol{a}_{K,i}) = (\widehat{p} \circ T_K^{-1})_{|F}(\boldsymbol{a}_{K,i})$$
$$= q(\boldsymbol{a}_{K,i}) = \sigma_{K,F,i}(q) = 0.$$

Assumption 20.1 (face unisolvence) implies that $\widehat{p}_{|\widehat{F}} = 0$, so that $q = 0$. □

Recall that since the mesh is matching, any interface $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$ is a face of K_l and a face of K_r, i.e., $F \in \mathcal{F}_{K_l} \cap \mathcal{F}_{K_r}$. Our second key assumption is formulated as follows.

Assumption 20.3 (Face matching). *For all $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$, we have* (i) $P_{K_l,F} = P_{K_r,F} =: P_F$ *and* (ii) $\Sigma_{K_l,F} = \Sigma_{K_r,F} =: \Sigma_F$, *i.e., there is a bijective map* $\chi_{lr} : \mathcal{N}_{K_l,F} \to \mathcal{N}_{K_r,F}$ *s.t.* $\boldsymbol{a}_{K_l,i} = \boldsymbol{a}_{K_r,\chi_{lr}(i)}$ *for all* $i \in \mathcal{N}_{K_l,F}$.

We are now in a position to state the main result of this section.

Lemma 20.4 (Zero-jump). *Let* $v_h \in P_k^{\mathrm{g,b}}(\mathcal{T}_h)$ *and* $F \in \mathcal{F}_h^\circ$. *Under Assumptions* 20.1 *and* 20.3, *the following equivalence holds true:*

$$[\,[\![v_h]\!]_F = 0\,] \iff [\,v_{h|K_l}(\boldsymbol{a}_{K_l,i}) = v_{h|K_r}(\boldsymbol{a}_{K_r,\chi_{lr}(i)}), \forall i \in \mathcal{N}_{K_l,F}\,]. \quad (20.5)$$

Proof. Let $v_h \in P_k^{\mathrm{g,b}}(\mathcal{T}_h)$ and $F \in \mathcal{F}_h^\circ$. Let v_l be the restriction of $v_{h|K_l}$ to F, and let v_r be the restriction of $v_{h|K_r}$ to F. Since $v_h \in P_k^{\mathrm{g,b}}(\mathcal{T}_h)$, we have $v_l \in P_{K_l,F}$ and $v_r \in P_{K_r,F}$. Owing to Assumption 20.3, we also have $v_r \in P_{K_l,F}$, i.e., $[\![v_h]\!]_F = v_l - v_r \in P_{K_l,F}$. Since $(F, P_{K_l,F}, \Sigma_{K_l,F})$ is a finite element owing to Lemma 20.2 (which follows from Assumption 20.1), we infer that $[\![v_h]\!]_F = v_l - v_r = 0$ iff $(v_l - v_r)(\boldsymbol{a}_{K_l,i}) = 0$ for all $i \in \mathcal{N}_{K_l,F}$. But $v_l(\boldsymbol{a}_{K_l,i}) = v_{h|K_l}(\boldsymbol{a}_{K_l,i})$ and, owing to Assumption 20.3, we also have $v_r(\boldsymbol{a}_{K_l,i}) = v_{h|K_r}(\boldsymbol{a}_{K_l,i}) = v_{h|K_r}(\boldsymbol{a}_{K_r,\chi_{lr}(i)})$. This proves (20.5). □

20.2 Verification of the assumptions (Lagrange)

In this section, we verify Assumptions 20.1 and 20.3 for Lagrange $\mathbb{P}_{k,d}$ elements when $\widehat{K}$ is a simplex and for Lagrange $\mathbb{Q}_{k,d}$ elements when $\widehat{K}$ is a cuboid. Since these two assumptions trivially hold true when $d = 1$, we assume in this section that $d \geq 2$. We do not assume that the geometric mapping $\boldsymbol{T}_K : \widehat{K} \to K$ is affine.

20.2.1 Face unisolvence

Assumption 20.1 has been proved in Lemma 6.15 for Lagrange $\mathbb{Q}_{k,d}$ elements and in Lemma 7.13 for Lagrange $\mathbb{P}_{k,d}$ elements. Note that the face unisolvence assumption is not met for the Crouzeix–Raviart element.

20.2.2 The space $P_{K,F}$

Let us now identify the space $P_{K,F}$ for all $K \in \mathcal{T}_h$ and all $F \in \mathcal{F}_K$. Let us set $\widehat{F} := \boldsymbol{T}_K^{-1}(F)$. Then $\widehat{F} \in \mathcal{F}_{\widehat{K}}$, i.e., $\widehat{F}$ is a face of the reference cell $\widehat{K}$. Let $\widehat{F}^{d-1}$ be the unit simplex in $\mathbb{R}^{d-1}$ if $\widehat{K}$ is the unit simplex of $\mathbb{R}^d$ or let $\widehat{F}^{d-1}$ be the unit cuboid of $\mathbb{R}^{d-1}$ if $\widehat{K}$ is the unit cuboid of $\mathbb{R}^d$. Since both $\widehat{F}^{d-1}$ and $\widehat{F}$ are either $(d-1)$-dimensional simplices or cuboids, it is always possible to construct an affine bijective mapping $\boldsymbol{T}_{\widehat{F}}$ from $\widehat{F}^{d-1}$ to $\widehat{F}$. Let us denote

$$\boldsymbol{T}_{\widehat{F}} : \widehat{F}^{d-1} \to \widehat{F}, \qquad \boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}} : \widehat{F}^{d-1} \to F. \qquad (20.6)$$

Lemma 20.5 (Characterization of $P_{K,F}$). *Let $\widehat{K}$ be either a simplex or a cuboid. Then $P_{K,F} = \widehat{P}_k^{d-1} \circ \boldsymbol{T}_{K,F}^{-1}$ where $\widehat{P}_k^{d-1} := \mathbb{P}_{k,d-1}$ if $\widehat{K}$ is a simplex and $\widehat{P}_k^{d-1} := \mathbb{Q}_{k,d-1}$ if $\widehat{K}$ is a cuboid.*

Proof. Let $q \in P_{K,F}$. By definition of $P_{K,F}$, there is $\widehat{p} \in \widehat{P}$ s.t.

$$q = (\widehat{p} \circ \boldsymbol{T}_K^{-1})_{|F} = \widehat{p}_{|\widehat{F}} \circ \boldsymbol{T}_{K|F}^{-1} = (\widehat{p}_{|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}}) \circ (\boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}})^{-1} = (\widehat{p}_{|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}}) \circ \boldsymbol{T}_{K,F}^{-1}.$$

Since $\widehat{p}_{|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}} \in \widehat{P}_k^{d-1}$ (see Lemma 6.13 or Lemma 7.10 depending on the nature of $\widehat{F}$), we conclude that $q \in \widehat{P}_k^{d-1} \circ \boldsymbol{T}_{K,F}^{-1}$. This shows that $P_{K,F} \subset \widehat{P}_k^{d-1} \circ \boldsymbol{T}_{K,F}^{-1}$. The converse inclusion is proved by similar arguments. $\qquad \square$

20.2.3 Face matching

We now establish that $P_{K_l,F} = P_{K_r,F}$ and $\Sigma_{K_l,F} = \Sigma_{K_r,F}$.

Lemma 20.6 (Face matching, (i)). *Assume that $\widehat{K}$ is either a simplex or a cuboid. Let $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$. Then $P_{K_l,F} = P_{K_r,F}$.*

Proof. Let us set $\widehat{F}_l := \boldsymbol{T}_{K_l}^{-1}(F)$ and $\widehat{F}_r := \boldsymbol{T}_{K_r}^{-1}(F)$. Since the mesh is matching, $\widehat{F}_l$ and $\widehat{F}_r$ are faces of $\widehat{K}$. By construction, the mapping

$$\boldsymbol{T}_{K_l|F}^{-1} \circ \boldsymbol{T}_{K_r|\widehat{F}_r} : \widehat{F}_r \to \widehat{F}_l$$

is bijective, and turns out to be affine even when the mappings $\boldsymbol{T}_{K_l}$ and $\boldsymbol{T}_{K_r}$ are nonaffine as shown in Exercise 20.1. Then the mapping $\boldsymbol{S}_{rl} : \widehat{F}^{d-1} \to \widehat{F}^{d-1}$ s.t.

$$\boldsymbol{S}_{rl} := \boldsymbol{T}_{K_l,F}^{-1} \circ \boldsymbol{T}_{K_r,F} = \boldsymbol{T}_{\widehat{F}_l}^{-1} \circ \boldsymbol{T}_{K_l|F}^{-1} \circ \boldsymbol{T}_{K_r|\widehat{F}_r} \circ \boldsymbol{T}_{\widehat{F}_r}$$

is affine (because the mappings $\boldsymbol{T}_{\widehat{F}_l}^{-1}$, $\boldsymbol{T}_{K_l|F}^{-1} \circ \boldsymbol{T}_{K_r|\widehat{F}_r}$, and $\boldsymbol{T}_{\widehat{F}_r}$ are affine) and bijective; see Figure 20.1. Since $\widehat{P}_k^{d-1} = \mathbb{P}_{k,d-1}$ or $\widehat{P}_k^{d-1} = \mathbb{Q}_{k,d-1}$ depending

on the nature of $\widehat{K}$, we infer that $\widehat{P}_k^{d-1}$ is invariant under $\boldsymbol{S}_{rl}$, i.e., $\widehat{P}_k^{d-1} \circ \boldsymbol{S}_{rl} = \widehat{P}_k^{d-1}$. Using this property together with the identity $P_{K,F} = \widehat{P}_k^{d-1} \circ \boldsymbol{T}_{K,F}^{-1}$ proved in Lemma 20.5, we infer that

$$P_{K_l,F} = \widehat{P}_k^{d-1} \circ \boldsymbol{T}_{K_l,F}^{-1} = \widehat{P}_k^{d-1} \circ \boldsymbol{S}_{rl} \circ \boldsymbol{T}_{K_r,F}^{-1} = \widehat{P}_k^{d-1} \circ \boldsymbol{T}_{K_r,F}^{-1} = P_{K_r,F}. \qquad \square$$
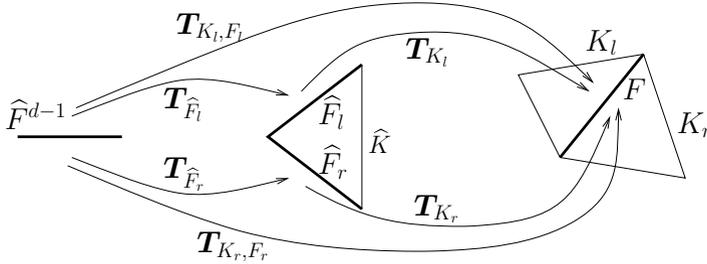
Fig. 20.1 Two-dimensional example ($d = 2$): geometric mappings associated with an interface F, the reference faces $\widehat{F}_l$ and $\widehat{F}_r$, and the unit segment $\widehat{F}^{d-1}$.

To establish that $\Sigma_{K_l,F} = \Sigma_{K_r,F}$ for a general set of Lagrange nodes in $\widehat{K}$, we formulate a symmetry assumption on the Lagrange nodes located on the faces of $\widehat{K}$. This assumption turns out to be sufficient in order to establish that $\Sigma_{K_l,F} = \Sigma_{K_r,F}$. Combined with the result from Lemma 20.6, this allows us to conclude that Assumption 20.3 (face matching) is indeed satisfied.

Assumption 20.7 (Invariance by vertex permutation). *We assume that there is a set $\{\widehat{\boldsymbol{s}}_m\}_{m \in \mathcal{N}_{\widehat{F}^{d-1}}}$ of Lagrange nodes in $\widehat{F}^{d-1}$, with $\mathcal{N}_{\widehat{F}^{d-1}} :=$ $\{1{:}n^{\mathrm{f}}\}$ for some integer $n^{\mathrm{f}} \geq 1$, s.t. the following holds true:* (i) *The set $\{\widehat{\boldsymbol{s}}_m\}_{m \in \mathcal{N}_{\widehat{F}^{d-1}}}$ is invariant under any vertex permutation of $\widehat{F}^{d-1}$.* (ii) *For every face $\widehat{F}$ of $\widehat{K}$, $\{\boldsymbol{T}_{\widehat{F}}(\widehat{\boldsymbol{s}}_m)\}_{m \in \mathcal{N}_{\widehat{F}^{d-1}}}$ are the Lagrange nodes on $\widehat{F}$.*

Assumption 20.7(i) means that for every affine bijective mapping $\boldsymbol{S}$: $\widehat{F}^{d-1} \to \widehat{F}^{d-1}$, there is a permutation $\chi_{\boldsymbol{S}}$ of $\mathcal{N}_{\widehat{F}^{d-1}}$ such that $\boldsymbol{S}(\widehat{\boldsymbol{s}}_m) = \widehat{\boldsymbol{s}}_{\chi_{\boldsymbol{S}}(m)}$ for all $m \in \mathcal{N}_{\widehat{F}^{d-1}}$. Assumption 20.7(ii) means that $\mathrm{card}(\mathcal{N}_{\widehat{K},\widehat{F}}) = n^{\mathrm{f}}$ is independent of the face $\widehat{F}$ of $\widehat{K}$ and that, for every $\widehat{F} \in \mathcal{F}_{\widehat{K}}$, there is a bijective map $\mathtt{j}_{\widehat{F}}^{\mathtt{fc}} : \mathcal{N}_{\widehat{F}^{d-1}} \to \mathcal{N}_{\widehat{K},\widehat{F}}$ such that (see Figure 20.2)

$$\boldsymbol{T}_{\widehat{F}}(\widehat{\boldsymbol{s}}_m) = \widehat{\boldsymbol{a}}_{\mathtt{j}_{\widehat{F}}^{\mathtt{fc}}(m)}, \qquad \forall m \in \mathcal{N}_{\widehat{F}^{d-1}}. \tag{20.7}$$

Example 20.8 ($\mathbb{Q}_{k,d}$ Lagrange elements). After inspection of Proposition 6.14 on the reference cuboid $\widehat{K} := [0,1]^d$, we realize that Assumption 20.7 holds true for tensor-product Lagrange elements provided that for every $i \in \{1{:}d\}$, the set of points $\{a_{i,l}\}_{l \in \{0:k\}}$ is such that $a_{i,l} = \alpha_l$ for
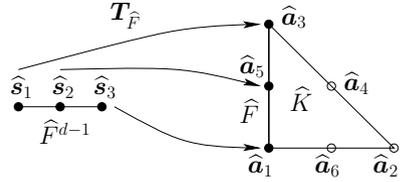
every $l \in \{0{:}k\}$, where the points $0 = \alpha_0 < \ldots < \alpha_k = 1$ are all distinct
in the interval $[0,1]$ and satisfy the symmetry property $\alpha_l = 1 - \alpha_{k-l}$ for
all $l \in \{0{:}\lfloor \frac{k}{2} \rfloor\}$. The Gauss–Lobatto nodes satisfy these assumptions (up to
rescaling from $[-1,1]$ to $[0,1)$; see §6.2.   □

Example 20.9 ($\mathbb{P}_{k,d}$ Lagrange elements). The simplicial Lagrange element described in Proposition 7.12 also satisfies the assumption on invariance by vertex permutation. In dimension two, for instance, the edge nodes are invariant under symmetry about the midpoint as shown in the left panel of Figure 20.3 (for $k = 2$). Note that it is possible to use a set of Lagrange nodes that is different from the one introduced in Proposition 7.12 provided the vertex permutation assumption holds true (in addition to the face unisolvence). For instance, one can use the Fekete points mentioned in Remark 7.14.   □
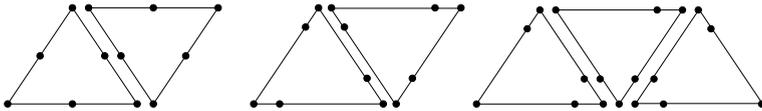
Fig. 20.3 $\mathbb{P}_{2,2}$ Lagrange element: two-dimensional example (left) and counterexample (center) for Assumption 20.3 (the triangles K_l and K_r are drawn slightly apart). In the rightmost panel, Assumption 20.3 is satisfied but not Assumption 20.7. This illustrates the fact that Assumption 20.7 is not needed to establish Assumption 20.3 if one enforces extra constraints on the way adjacent mesh cells come into contact.

Lemma 20.10 (Face matching, (ii)). *Assume that $\widehat{K}$ is either a simplex or a cuboid. Let $F := \partial K_l \cap \partial K_r \in \mathcal{F}^\circ_h$. Let Assumption 20.7 on invariance by vertex permutation be fulfilled. Then $\Sigma_{K_l,F} = \Sigma_{K_r,F}$.*

Proof. Let $i \in \mathcal{N}_{K_r,F} = \mathcal{N}_{\widehat{K},\widehat{F}_r}$ and let $\boldsymbol{a}_{K_r,i}$ be the corresponding Lagrange node of K_r located on F. Then $\boldsymbol{T}_{K_r}^{-1}(\boldsymbol{a}_{K_r,i}) = \widehat{\boldsymbol{a}}_i$ is a Lagrange node on $\widehat{F}_r$. Let $m \in \mathcal{N}_{\widehat{F}^{d-1}}$ be such that $i = \mathtt{j}^{\mathtt{fc}}_{\widehat{F}_r}(m)$, that is, $\widehat{\boldsymbol{a}}_i = \boldsymbol{T}_{\widehat{F}_r}(\widehat{\boldsymbol{s}}_m)$. Since we have established above that the mapping $\boldsymbol{S}_{rl} := \boldsymbol{T}_{\widehat{F}_l}^{-1} \circ \boldsymbol{T}_{K_l|F}^{-1} \circ \boldsymbol{T}_{K_r|F} \circ \boldsymbol{T}_{\widehat{F}_r}$ is affine, there is a permutation $\chi_{\boldsymbol{S}_{rl}} : \mathcal{N}_{\widehat{F}^{d-1}} \to \mathcal{N}_{\widehat{F}^{d-1}}$ such that $\boldsymbol{S}_{rl}(\widehat{\boldsymbol{s}}_m) = \widehat{\boldsymbol{s}}_{\chi_{\boldsymbol{S}_{rl}}(m)}$ for all $m \in \mathcal{N}_{\widehat{F}^{d-1}}$. Then the identity $\boldsymbol{S}_{rl}(\widehat{\boldsymbol{s}}_m) = \widehat{\boldsymbol{s}}_{\chi_{\boldsymbol{S}_{rl}}(m)}$ means that

$(\boldsymbol{T}_{K_r|\widehat{F}_r} \circ \boldsymbol{T}_{\widehat{F}_r})(\widehat{\boldsymbol{s}}_m) = (\boldsymbol{T}_{K_l|\widehat{F}_l} \circ \boldsymbol{T}_{\widehat{F}_l})(\widehat{\boldsymbol{s}}_{\chi_{\boldsymbol{s}_{rl}}(m)})$, which can also be rewritten as $\boldsymbol{a}_{K_r, \mathsf{j}^{\mathsf{fc}}_{\widehat{F}_r}(m)} = \boldsymbol{a}_{K_l, \mathsf{j}^{\mathsf{fc}}_{\widehat{F}_l}(\chi_{\boldsymbol{s}_{rl}}(m))}$. Hence, we have

$$\sigma_{K_r,F,\mathsf{j}^{\mathsf{fc}}_{\widehat{F}_r}(m)}(q) = q(\boldsymbol{a}_{K_r,\mathsf{j}^{\mathsf{fc}}_{\widehat{F}_r}(m)}) = q(\boldsymbol{a}_{K_l,\mathsf{j}^{\mathsf{fc}}_{\widehat{F}_l}(\chi_{\boldsymbol{s}_{rl}}(m))}) = \sigma_{K_l,F,\mathsf{j}^{\mathsf{fc}}_{\widehat{F}_l}(\chi_{\boldsymbol{s}_{rl}}(m))}(q),$$

for all $q \in P_F$ and all $m \in \mathcal{N}_{\widehat{F}^{d-1}}$. This proves that $\Sigma_{K_l,F} = \Sigma_{K_r,F}$ since $\mathsf{j}^{\mathsf{fc}}_{\widehat{F}_l} \circ \chi_{\boldsymbol{s}_{rl}} \circ (\mathsf{j}^{\mathsf{fc}}_{\widehat{F}_r})^{-1}$ is bijective.   □

Remark 20.11 (Serendipity and prismatic elements). The reader is invited to verify that the face unisolvence assumption 20.1 holds true also for the serendipity elements described in §6.4.3 and for the prismatic elements described in Remark 7.16. The face matching assumption 20.3 holds true for the serendipity elements since the face dofs are the same as those of the corresponding $\mathbb{Q}_{k,d}$ element. The assumption 20.3 can also be shown to hold true for the prismatic elements provided the Lagrange nodes on the triangular faces and the Lagrange nodes on the quadrangular faces each satisfy the vertex permutation assumption.   □

20.3 Generalization of the two gluing assumptions

In this section, we generalize the theory developed in §20.1 to enforce the jump condition $[\![v_h]\!]^{\mathrm{x}}_F = 0$ across all the mesh interfaces $F \in \mathcal{F}^{\circ}_h$ for $\mathrm{x} \in \{\mathrm{g,c,d}\}$ and $v_h \in P^{\mathrm{x,b}}_k(\mathcal{T}_h; \mathbb{R}^q)$. We are going to rephrase §20.1 in a slightly more abstract language. Recall from (18.8) that $[\![v_h]\!]^{\mathrm{x}}_F := \gamma^{\mathrm{x}}_{K_l,F}(v_{h|K_l}) - \gamma^{\mathrm{x}}_{K_r,F}(v_{h|K_r})$ with $F := \partial K_l \cap \partial K_r$ and the trace operator $\gamma^{\mathrm{x}}_{K,F}$ defined in (18.7) for every mesh cell $K \in \mathcal{T}_h$ and every face $F \in \mathcal{F}_K$ of K. We drop the superscript x whenever the context is unambiguous.

We start by identifying two structural properties of the finite element which we will call face unisolvence and face matching assumptions. We proceed in two steps. First, given a mesh cell $K \in \mathcal{T}_h$, we use the local finite element (K, P_K, Σ_K) with local shape functions $\{\theta_{K,i}\}_{i \in \mathcal{N}}$ and local dofs $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$, and invoke the face unisolvence assumption to construct a finite element attached to each face $F \in \mathcal{F}_K$. Then for every mesh interface $F := \partial K_l \cap \partial K_r \in \mathcal{F}^{\circ}_h$, we invoke the face matching assumption to make sure that the two face elements built on F from K_l and from K_r are identical (note that $F \in \mathcal{F}_{K_l} \cap \mathcal{F}_{K_r}$ since the mesh is matching). The theory is illustrated with various examples in §20.4. In this section (and the next one), we restrict the maps $\{\sigma_{K,i}\}_{i \in \mathcal{N}}$ and $\gamma_{K,F}$ to P_K, so that the kernels of these maps are to be understood as subspaces of P_K (for simplicity, we keep the same notation for the restrictions). Our first key assumption is the following.

Assumption 20.12 (Face unisolvence). *For all $K \in \mathcal{T}_h$ and all $F \in \mathcal{F}_K$, there is a nonempty subset $\mathcal{N}_{K,F} \subset \mathcal{N}$ s.t. $\ker(\gamma_{K,F}) = \bigcap_{i \in \mathcal{N}_{K,F}} \ker(\sigma_{K,i})$, i.e., for all $p \in P_K$,*

$$[\,\sigma_{K,i}(p) = 0,\, \forall i \in \mathcal{N}_{K,F}\,] \iff [\,\gamma_{K,F}(p) = 0\,]. \tag{20.8}$$

Equivalently, we have $\ker(\gamma_{K,F}) = \mathrm{span}\{\theta_{K,i}\}_{i \notin \mathcal{N}_{K,F}}$.

Let $\mathcal{N}_{K,F} \subset \mathcal{N}$ be defined according to Assumption 20.12. Let us define the corresponding trace space $P_{K,F}$ by setting

$$P_{K,F} := \gamma_{K,F}(P_K) = \mathrm{span}\{\gamma_{K,F}(\theta_{K,i})\}_{i \in \mathcal{N}_{K,F}}. \tag{20.9}$$

Notice that $\gamma_{K,F}(\theta_{K,i}) \neq 0$ for all $i \in \mathcal{N}_{K,F}$ by construction. The inclusion $\ker(\gamma_{K,F}) \subset \ker(\sigma_{K,i})$ for all $i \in \mathcal{N}_{K,F}$ (which follows from Assumption 20.12) implies that there is a unique linear map $\sigma_{K,F,i} : P_{K,F} \to \mathbb{R}$ s.t. $\sigma_{K,i} = \sigma_{K,F,i} \circ \gamma_{K,F}$ (see Exercise 20.2). Finally, let us set

$$\Sigma_{K,F} := \{\sigma_{K,F,i}\}_{i \in \mathcal{N}_{K,F}}. \tag{20.10}$$

We can now state an important consequence of Assumption 20.12.

Lemma 20.13 (Face element). *Let $K \in \mathcal{T}_h$ and $F \in \mathcal{F}_K$. Under Assumption 20.12, the triple $(F, P_{K,F}, \Sigma_{K,F})$ is a finite element.*

Proof. We use Remark 5.3 to prove unisolvence. Since Assumption 20.12 means that $\ker(\gamma_{K,F}) = \mathrm{span}\{\theta_{K,i}\}_{i \notin \mathcal{N}_{K,F}}$, we infer that $\dim(\ker(\gamma_{K,F})) = \mathrm{card}(\mathcal{N}) - \mathrm{card}(\mathcal{N}_{K,F})$. The rank nullity theorem implies that

$$\dim(P_{K,F}) = \dim(P_K) - \dim(\ker(\gamma_{K,F})) = \mathrm{card}(\mathcal{N}_{K,F}) = \mathrm{card}(\Sigma_{K,F}).$$

Let now $q \in P_{K,F}$ be s.t. $\sigma_{K,F,i}(q) = 0$ for all $i \in \mathcal{N}_{K,F}$. The definition of $P_{K,F}$ implies that there is $p \in P_K$ s.t. $q = \gamma_{K,F}(p)$. Hence, $\sigma_{K,i}(p) = \sigma_{K,F,i}(q) = 0$ for all $i \in \mathcal{N}_{K,F}$. In other words, $p \in \bigcap_{i \in \mathcal{N}_{K,F}} \ker(\sigma_{K,i})$. Hence, $p \in \ker(\gamma_{K,F})$. We conclude that $q = \gamma_{K,F}(p) = 0$. $\qquad\square$

Let $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ be the reference element and let ψ_K be the functional transformation that has been used to generate (K, P_K, Σ_K). Let $F \in \mathcal{F}_K$ and consider the face $\widehat{F} := \boldsymbol{T}_K^{-1}(F)$ of $\widehat{K}$. We are going to assume that for all $p \in P_K$, $\gamma_{K,F}(p) = 0$ iff $\gamma_{\widehat{K},\widehat{F}}(\widehat{p}) = 0$ with $\widehat{p} := \psi_K(p)$, i.e., we assume that

$$\ker(\gamma_{K,F}) = \ker(\gamma_{\widehat{K},\widehat{F}} \circ \psi_K). \tag{20.11}$$

This assumption holds true if ψ_K is the pullback by the geometric mapping $\boldsymbol{T}_K$ or one of the Piola transformations. Then Assumption 20.12 can be formulated on the reference element, and this assumption amounts to requiring that there exists a nonempty subset $\mathcal{N}_{\widehat{K},\widehat{F}} \subset \mathcal{N}$ s.t. $\bigcap_{i \in \mathcal{N}_{\widehat{K},\widehat{F}}} \ker(\widehat{\sigma}_i) = \ker(\gamma_{\widehat{K},\widehat{F}})$. Then we have

$$\mathcal{N}_{K,F} = \mathcal{N}_{\widehat{K},\widehat{F}} = \mathcal{N}_{\widehat{K},\boldsymbol{T}_K^{-1}(F)}, \qquad \forall K \in \mathcal{T}_h,\ \forall F \in \mathcal{F}_K. \tag{20.12}$$

Our second key assumption is the following.

Assumption 20.14 (Face matching). *For all $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$, we have (i) $P_{K_l,F} = P_{K_r,F} =: P_F$ and (ii) $\Sigma_{K_l,F} = \Sigma_{K_r,F} =: \Sigma_F$, i.e., there is a bijective map $\chi_{lr} : \mathcal{N}_{K_l,F} \to \mathcal{N}_{K_r,F}$ s.t. $\sigma_{K_l,F,i} = \sigma_{K_r,F,\chi_{lr}(i)}$ for all $i \in \mathcal{N}_{K_l,F}$.*

We are now in a position to state the main result of this section.

Lemma 20.15 (Zero γ-jump). *Let $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$ and $F \in \mathcal{F}_h^\circ$. Under Assumptions 20.12 and 20.14, the following equivalence holds true:*

$$[\![v_h]\!]_F = 0 \iff [\sigma_{K_l,i}(v_{h|K_l}) = \sigma_{K_r,\chi_{lr}(i)}(v_{h|K_r}), \; \forall i \in \mathcal{N}_{K_l,F}]. \quad (20.13)$$

Proof. Since $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$, we have $v_{h|K_l} \in P_{K_l}$ and $v_{h|K_r} \in P_{K_r}$. Set $v_l := \gamma_{K_l,F}(v_{h|K_l})$ and $v_r := \gamma_{K_r,F}(v_{h|K_r})$, so that $[\![v_h]\!]_F = v_l - v_r$. Note that $v_l \in \gamma_{K_l,F}(P_{K_l}) = P_{K_l,F}$. Similarly, $v_r \in P_{K_r,F}$, and Assumption 20.14 implies that $v_r \in P_{K_l,F}$, i.e., $v_l - v_r \in P_{K_l,F}$. Since $(F, P_{K_l,F}, \Sigma_{K_l,F})$ is a finite element owing to Lemma 20.13 (which follows from Assumption 20.12), we infer that $[\![v_h]\!]_F = v_l - v_r = 0$ iff $\sigma_{K_l,F,i}(v_l - v_r) = 0$ for all $i \in \mathcal{N}_{K_l,F}$. To conclude the proof, we need to show that $\sigma_{K_l,F,i}(v_l - v_r) = \sigma_{K_l,i}(v_{h|K_l}) - \sigma_{K_r,\chi_{lr}(i)}(v_{h|K_r})$. On the one hand we have $\sigma_{K_l,F,i}(v_l) = \sigma_{K_l,F,i}(\gamma_{K_l,F}(v_{h|K_l})) = \sigma_{K_l,i}(v_{h|K_l})$, and on the other hand Assumption 20.14 implies that $\sigma_{K_l,F,i}(v_r) = \sigma_{K_r,F,\chi_{lr}(i)}(v_r) = \sigma_{K_r,F,\chi_{lr}(i)}(\gamma_{K_r,F}(v_{h|K_r})) = \sigma_{K_r,\chi_{lr}(i)}(v_{h|K_r})$. $\qquad\square$

20.4 Verification of the two gluing assumptions

We now present examples of finite elements satisfying the two structural assumptions of §20.3. These assumptions have already been shown in §20.2 to hold true for Lagrange elements. In the present section, we focus on affine simplicial matching meshes and assume that the mesh is oriented in a generation-compatible way (see §10.2). We invite the reader to verify that these examples can be adapted to affine Cartesian meshes.

20.4.1 Raviart–Thomas elements

Let $k \geq 0$ and let us show that the $\mathbb{RT}_{k,d}$ Raviart–Thomas elements introduced in §14.3 can be used to build discrete functions with integrable divergence. Let $K \in \mathcal{T}_h$ and $F \in \mathcal{F}_K$. We consider the γ^{d}-trace defined by (18.7c), i.e., $\gamma_{K,F}^{\mathrm{d}}(\boldsymbol{v}) := \boldsymbol{v}_{|F} \cdot \boldsymbol{n}_F$ where $\boldsymbol{n}_F$ is the unit normal vector orienting F. Following §14.4, consider the face dofs $\sigma_{F,m}^{\mathrm{f}}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{\nu}_F)(\zeta_m \circ \boldsymbol{T}_{K,F}^{-1}) \, \mathrm{d}s$, where $\boldsymbol{\nu}_F := |F| \boldsymbol{n}_F$, $\boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}} : \widehat{S}^{d-1} \to F$, $\boldsymbol{T}_{\widehat{F}} : \widehat{S}^{d-1} \to \widehat{F}$ is an affine bijective mapping, $\{\zeta_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ is a fixed basis of $\mathbb{P}_{k,d-1}$, and $n_{\mathrm{sh}}^{\mathrm{f}} := \dim(\mathbb{P}_{k,d-1})$ (see (14.12a)).

Lemma 20.16 (Face unisolvence). *Assumption* 20.12 *holds true with*

$$\mathcal{N}_{K,F} := \{i \in \mathcal{N} \mid \exists m(i) \in \{1{:}n_{\mathrm{sh}}^{\mathrm{f}}\}, \ \sigma_{K,i} = \sigma_{F,m(i)}^{\mathrm{f}}\}, \tag{20.14}$$

i.e., $\mathcal{N}_{K,F}$ collects all the indices of the dofs involving an integral over F.

Proof. We first observe that the subset $\mathcal{N}_{K,F}$ is nonempty. Since $\gamma_{K,F}^{\mathrm{d}}(\boldsymbol{v}) = 0$ implies that $\boldsymbol{v}_{|F}{\cdot}\boldsymbol{n}_F = 0$ and since $\boldsymbol{n}_F$ and $\boldsymbol{\nu}_F$ are collinear, we infer that $\sigma_{K,i}(\boldsymbol{v}) = 0$ for all $i \in \mathcal{N}_{K,F}$ and all $\boldsymbol{v} \in \ker(\gamma_{K,F}^{\mathrm{d}})$, i.e., $\ker(\gamma_{K,F}^{\mathrm{d}}) \subset \bigcap_{i\in\mathcal{N}_{K,F}} \ker(\sigma_{K,i})$. The converse inclusion results from Lemma 14.14. Hence, Assumption 20.12 holds true. $\qquad\square$

Lemma 20.17 ($P_{K,F}^{\mathrm{d}}$). *We have $P_{K,F}^{\mathrm{d}} := \gamma_{K,F}^{\mathrm{d}}(\mathbb{RT}_{k,d}) = \mathbb{P}_{k,d-1} \circ \boldsymbol{T}_{K,F}^{-1}$.*

Proof. We have $P_{K,F}^{\mathrm{d}} \subset \mathbb{P}_{k,d-1}\circ\boldsymbol{T}_{K,F}^{-1}$ owing to Lemma 14.7, and the equality follows by observing that $\dim(P_{K,F}^{\mathrm{d}}) = n_{\mathrm{sh}}^{\mathrm{f}} = \dim(\mathbb{P}_{k,d-1})$. $\qquad\square$

Let us set $\mathcal{N}_{\widehat{S}^{d-1}} := \{1{:}n_{\mathrm{sh}}^{\mathrm{f}}\}$ and for all $\widehat{F} \in \mathcal{F}_{\widehat{K}}$, let us introduce the bijective map $\mathtt{j}_{\widehat{F}}^{\mathtt{sf}} : \mathcal{N}_{\widehat{S}^{d-1}} \to \mathcal{N}_{\widehat{K},\widehat{F}}$ defined by setting $\mathtt{j}_{\widehat{F}}^{\mathtt{sf}}(m) := i$ for all $m \in \mathcal{N}_{\widehat{S}^{d-1}}$, where i is s.t. $\widehat{\sigma}_i = \sigma_{\widehat{F},m}^{\mathrm{f}}$. Then Lemma 20.16 applied on the reference element means that $\mathcal{N}_{\widehat{K},\widehat{F}} = \mathtt{j}_{\widehat{F}}^{\mathtt{sf}}(\mathcal{N}_{\widehat{S}^{d-1}})$. Owing to (20.12), we infer that we have for all $K \in \mathcal{T}_h$ and all $F \in \mathcal{F}_K$,

$$\mathcal{N}_{K,F} = \mathcal{N}_{\widehat{K},\boldsymbol{T}_K^{-1}(F)} = \mathtt{j}_{\boldsymbol{T}_K^{-1}(F)}^{\mathtt{sf}}(\mathcal{N}_{\widehat{S}^{d-1}}). \tag{20.15}$$

Lemma 20.18 (Face matching). (i) *$(F, P_{K,F}^{\mathrm{d}}, \Sigma_{K,F}^{\mathrm{d}})$ is a modal scalar-valued finite element with $\sigma_{K,F,i}^{\mathrm{d}}(\phi) := \int_F (\zeta_m \circ \boldsymbol{T}_{K,F}^{-1})\phi \, \mathrm{d}s$, $i := \mathtt{j}_{\boldsymbol{T}_K^{-1}(F)}^{\mathtt{sf}}(m)$, for all $\phi \in P_{K,F}^{\mathrm{d}}$ and all $m \in \mathcal{N}_{\widehat{S}^{d-1}}$.* (ii) *For all $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$, we have $P_{K_l,F}^{\mathrm{d}} = P_{K_r,F}^{\mathrm{d}} =: P_F^{\mathrm{d}}$.* (iii) *$\Sigma_{K_l,F}^{\mathrm{d}} = \Sigma_{K_r,F}^{\mathrm{d}} =: \Sigma_F^{\mathrm{d}}$ if the basis $\{\zeta_m\}_{m\in\mathcal{N}_{\widehat{S}^{d-1}}}$ of $\mathbb{P}_{k,d-1}$ is invariant under any* vertex permutation *of $\widehat{S}^{d-1}$, i.e., for every affine bijective mapping $\boldsymbol{S} : \widehat{S}^{d-1} \to \widehat{S}^{d-1}$, there exists a permutation $\chi_{\boldsymbol{S}}$ of $\mathcal{N}_{\widehat{S}^{d-1}}$ such that $\zeta_m \circ \boldsymbol{S} = \zeta_{\chi_{\boldsymbol{S}}(m)}$ for all $m \in \mathcal{N}_{\widehat{S}^{d-1}}$.*

Proof. (i) The first claim is a consequence of Lemma 20.17 and of the definition of the face dofs of the $\mathbb{RT}_{k,d}$ element.
(ii) Let $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$, and set $\widehat{F}_l := \boldsymbol{T}_{K_l}^{-1}(F)$ and $\widehat{F}_r := \boldsymbol{T}_{K_r}^{-1}(F)$. Recalling that the mapping $\boldsymbol{S}_{rl} = \boldsymbol{T}_{K_l,F}^{-1} \circ \boldsymbol{T}_{K_r,F}$ is affine, as shown in Figure 20.1, we observe that

$$P_{K_l,F}^{\mathrm{d}} = \mathbb{P}_{k,d-1} \circ \boldsymbol{T}_{K_l,F}^{-1} = (\mathbb{P}_{k,d-1} \circ \boldsymbol{S}_{rl}) \circ \boldsymbol{T}_{K_r,F}^{-1} = \mathbb{P}_{k,d-1} \circ \boldsymbol{T}_{K_r,F}^{-1} = P_{K_r,F}^{\mathrm{d}},$$

as in the proof of Lemma 20.6.
(iii) Letting $\chi_{\boldsymbol{S}_{rl}}$ be the index permutation associated with the mapping $\boldsymbol{S}_{rl}$, the following holds true for all $m \in \mathcal{N}_{\widehat{S}^{d-1}}$:

$$\sigma^{\mathrm{d}}_{K_l,F,\mathrm{j}^{\mathrm{sf}}_{\widehat{F}_l}(m)}(\phi) = \int_F (\zeta_m \circ \boldsymbol{T}^{-1}_{K_l,F})\phi \, \mathrm{d}s = \int_F (\zeta_m \circ \boldsymbol{S}_{rl} \circ \boldsymbol{T}^{-1}_{K_r,F})\phi \, \mathrm{d}s$$

$$= \int_F (\zeta_{\chi s_{rl}(m)} \circ \boldsymbol{T}^{-1}_{K_r,F})\phi \, \mathrm{d}s = \sigma^{\mathrm{d}}_{K_r,F,\mathrm{j}^{\mathrm{sf}}_{\widehat{F}_r}(\chi s_{rl}(m))}(\phi),$$

i.e., any dof $\sigma^{\mathrm{d}}_{K_l,F,i}$ in $\Sigma^{\mathrm{d}}_{K_l,F}$ is also in $\Sigma^{\mathrm{d}}_{K_r,F}$, and conversely. $\qquad\square$

Remark 20.19 (Basis). Let us give two examples of a permutation-invariant basis of $\mathbb{P}_{k,d-1}$. Let $\{\widehat{\boldsymbol{s}}_0, \ldots, \widehat{\boldsymbol{s}}_{d-1}\}$ be the vertices of $\widehat{S}^{d-1}$. Let $\mathcal{A}_{k,d-1} := \{\alpha \in \mathbb{N}^{d-1} \mid |\alpha| \leq k\}$ and consider the Lagrange nodes $\{\widehat{\boldsymbol{a}}_\alpha\}_{\alpha \in \mathcal{A}_{k,d-1}}$ defined by $\widehat{\boldsymbol{a}}_\alpha := \widehat{\boldsymbol{s}}_0 + \sum_{i \in \{1:d-1\}} \frac{\alpha_i}{k}(\widehat{\boldsymbol{s}}_i - \widehat{\boldsymbol{s}}_0)$. Then the Lagrange polynomials associated with $\{\widehat{\boldsymbol{a}}_\alpha\}_{\alpha \in \mathcal{A}_{k,d-1}}$ form a permutation-invariant basis of $\mathbb{P}_{k,d-1}$. Likewise the modal basis $\{\widehat{\lambda}_0^{\beta_0} \ldots \widehat{\lambda}_{d-1}^{\beta_{d-1}}, \beta_0 + \ldots + \beta_{d-1} = k\}$, where $(\widehat{\lambda}_0, \ldots, \widehat{\lambda}_{d-1})$ are the barycentric coordinates in $\widehat{S}^{d-1}$, is also a permutation-invariant basis of $\mathbb{P}_{k,d-1}$ (see Exercise 7.4(v)). $\qquad\square$

20.4.2 Nédélec elements

Let $k \geq 0$ and let us show that the $\mathbf{N}_{k,d}$ Nédélec elements introduced in §15.3 can be used to build discrete functions with integrable curl. We assume that $d = 3$ (the construction is analogous but simpler for $d = 2$). Let $K \in \mathcal{T}_h$ and $F \in \mathcal{F}_K$. We consider the γ^{c}-trace defined in (18.7b), i.e., $\gamma^{\mathrm{c}}_{K,F}(\boldsymbol{v}) := \boldsymbol{v}_{|F} \times \boldsymbol{n}_F$ where $\boldsymbol{n}_F$ is the unit normal vector orienting F. Proceeding as in §15.4, we consider the edge dofs $\sigma^{\mathrm{e}}_{E,m}(\boldsymbol{v}) := \frac{1}{|E|} \int_E (\boldsymbol{v} \cdot \boldsymbol{t}_E)(\mu_m \circ \boldsymbol{T}^{-1}_{K,E}) \, \mathrm{d}l$, where $\boldsymbol{T}_{K,E} := \boldsymbol{T}_{K|\widehat{E}} \circ \boldsymbol{T}_{\widehat{E}} : \widehat{S}^1 \to E$, $\boldsymbol{T}_{\widehat{E}} : \widehat{S}^1 \to \widehat{E}$ is an affine bijective mapping, $\boldsymbol{t}_E$ is the edge vector orienting E, $\{\mu_m\}_{m \in \{1:n^{\mathrm{e}}_{\mathrm{sh}}\}}$ is a fixed basis of $\mathbb{P}_{k,1}$, and $n^{\mathrm{e}}_{\mathrm{sh}} := \dim(\mathbb{P}_{k,1})$. If $k \geq 1$, we also consider the face dofs $\sigma^{\mathrm{f}}_{F,j,m}(\boldsymbol{v}) := \frac{1}{|F|} \int_F (\boldsymbol{v} \cdot \boldsymbol{t}_{F,j})(\zeta_m \circ \boldsymbol{T}^{-1}_{K,F}) \, \mathrm{d}s$, where $\boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}} : \widehat{S}^2 \to F$, $\boldsymbol{T}_{\widehat{F}} : \widehat{S}^2 \to \widehat{F}$ is an affine bijective mapping, $\{\boldsymbol{t}_{F,j}\}_{j \in \{1,2\}}$ are the two edge vectors orienting F, $\{\zeta_m\}_{m \in \{1:n^{\mathrm{f}}_{\mathrm{sh}}\}}$ is a fixed basis of $\mathbb{P}_{k-1,2}$, and $n^{\mathrm{f}}_{\mathrm{sh}} := \dim(\mathbb{P}_{k-1,2})$. For all $F \in \mathcal{F}_K$, let $\mathcal{E}_F$ be the collection of the three edges composing the boundary of F. Let

$$\mathcal{N}^{\mathrm{e}}_{K,F} := \{i \in \mathcal{N} \mid \exists (E(i), m(i)) \in \mathcal{E}_F \times \{1:n^{\mathrm{e}}_{\mathrm{sh}}\}, \ \sigma_{K,i} = \sigma^{\mathrm{e}}_{E(i),m(i)}\}$$

be the collection of the indices of the edge dofs associated with F and

$$\mathcal{N}^{\mathrm{f}}_{K,F} := \{i \in \mathcal{N} \mid \exists (j(i), m(i)) \in \{1,2\} \times \{1:n^{\mathrm{f}}_{\mathrm{sh}}\}, \sigma_{K,i} = \sigma^{\mathrm{f}}_{F,j(i),m(i)}\}$$

be the collection of the indices of the face dofs associated with F ($k \geq 1$). We adopt the convention that $\mathcal{N}^{\mathrm{f}}_{K,F} := \emptyset$ if $k = 0$.

Lemma 20.20 (Face unisolvence). *Assumption 20.12 holds true with the subset* $\mathcal{N}_{K,F} := \mathcal{N}^{\mathrm{e}}_{K,F} \cup \mathcal{N}^{\mathrm{f}}_{K,F}$.

Proof. We first observe that the subset $\mathcal{N}_{K,F}$ is nonempty. Let $\boldsymbol{v} \in \boldsymbol{P}_K^{\mathrm{c}}$ be such that $\gamma_{K,F}^{\mathrm{c}}(\boldsymbol{v}) = \boldsymbol{0}$, i.e., $\boldsymbol{v}_{|F} \times \boldsymbol{n}_F = \boldsymbol{0}$. Then $\sigma_{K,i}(\boldsymbol{v}) = 0$ for all $i \in \mathcal{N}_{K,F}$, so that $\ker(\gamma_{K,F}) \subset \bigcap_{i \in \mathcal{N}_{K,F}} \ker(\sigma_{K,i})$. The converse inclusion results from Lemma 15.15. $\qquad\square$

Lemma 20.21 ($\boldsymbol{P}_{K,F}^{\mathrm{c}}$). $\boldsymbol{P}_{K,F}^{\mathrm{c}} := \gamma_{K,F}^{\mathrm{c}}(\mathbb{N}_{k,d}) = \mathbb{J}_{K,F}^{-\mathsf{T}}(\mathbb{N}_{k,2} \circ \boldsymbol{T}_{K,F}^{-1}) \times \boldsymbol{n}_F$.

Proof. The inclusion $\boldsymbol{P}_{K,F}^{\mathrm{c}} \subset \mathbb{J}_{K,F}^{-\mathsf{T}}(\mathbb{N}_{k,2} \circ \boldsymbol{T}_{K,F}^{-1}) \times \boldsymbol{n}_F$ is shown as in the proof of Lemma 15.8. Equality follows by invoking a dimension argument, i.e., $\dim(\mathbb{J}_{K,F}^{-\mathsf{T}}(\mathbb{N}_{k,2} \circ \boldsymbol{T}_{K,F}^{-1}) \times \boldsymbol{n}_F) = \dim(\mathbb{N}_{k,2})$ and $\mathrm{card}(\mathcal{N}_{K,F}) = 2\dim(\mathbb{P}_{k-1,2}) + 3\dim(\mathbb{P}_{k,1}) = (k+1)(k+3) = \dim(\mathbb{N}_{k,2})$ owing to Lemma 15.7. $\qquad\square$

Lemma 20.22 (Face matching). (i) *The triple* $(F, \boldsymbol{P}_{K,F}^{\mathrm{c}}, \Sigma_{K,F}^{\mathrm{c}})$ *is a two-dimensional Raviart–Thomas finite element with dofs*

$$\sigma_{K,F,i}^{\mathrm{c}}(\boldsymbol{\phi}) := \frac{1}{|E(i)|} \int_{E(i)} (\boldsymbol{\phi} \cdot \boldsymbol{t}_{E(i)}^{\perp})(\mu_{m(i)} \circ \boldsymbol{T}_{K,E(i)}^{-1}) \,\mathrm{d}l, \quad \forall i \in \mathcal{N}_{K,F}^{\mathrm{e}}, \quad (20.16\mathrm{a})$$

$$\sigma_{K,F,i}^{\mathrm{c}}(\boldsymbol{\phi}) := \frac{1}{|F|} \int_{F} (\boldsymbol{\phi} \cdot \boldsymbol{t}_{F,j(i)}^{\perp})(\zeta_{m(i)} \circ \boldsymbol{T}_{K,F}^{-1}) \,\mathrm{d}s, \quad \forall i \in \mathcal{N}_{K,F}^{\mathrm{f}}, \qquad (20.16\mathrm{b})$$

for all $\boldsymbol{\phi} \in \boldsymbol{P}_{K,F}^{\mathrm{c}}$ *and all* $i \in \mathcal{N}_{K,F}$, *with* $\boldsymbol{t}_{E(i)}^{\perp} := \boldsymbol{t}_{E(i)} \times \boldsymbol{n}_F$ *and* $\boldsymbol{t}_{F,j(i)}^{\perp} := \boldsymbol{t}_{F,j(i)} \times \boldsymbol{n}_F$. (ii) *For all* $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^{\circ}$, *we have* $\boldsymbol{P}_{K_l,F}^{\mathrm{c}} = \boldsymbol{P}_{K_r,F}^{\mathrm{c}} =: \boldsymbol{P}_F^{\mathrm{c}}$. (iii) *We have* $\Sigma_{K_l,F}^{\mathrm{c}} = \Sigma_{K_r,F}^{\mathrm{c}} =: \Sigma_F^{\mathrm{c}}$ *if the chosen bases* $\{\zeta_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ *and* $\{\mu_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$ *are invariant under any vertex permutation of* $\widehat{S}^2$ *and* $\widehat{S}^1$, *respectively.*

Proof. The expressions in (20.16) follow from the definition of the edge and the face dofs of the $\mathbb{N}_{k,d}$ element and from the fact that $(\boldsymbol{n}_F \times (\boldsymbol{h} \times \boldsymbol{n}_F)) \cdot \boldsymbol{t} = \boldsymbol{h} \cdot \boldsymbol{t}$ for all $\boldsymbol{h} \in \mathbb{R}^3$ and every vector $\boldsymbol{t}$ that is tangent to F. The rest of the proof is similar to that of Lemma 20.18. $\qquad\square$

Remark 20.23 (Choice of basis). Examples of permutation-invariant bases of $\mathbb{P}_{k-1,2}$ and $\mathbb{P}_{k,1}$ are the nodal and the modal bases built by using either the Lagrange nodes in $\widehat{S}^2$ and $\widehat{S}^1$ or the barycentric coordinates in $\widehat{S}^2$ and $\widehat{S}^1$ as in Remark 20.19. $\qquad\square$

20.4.3 Canonical hybrid elements

Let $k \geq 1$ and let us show that the canonical hybrid finite element introduced in §7.6 can be used to build discrete functions with integrable gradient. Assume $d = 3$ (the case $d = 2$ is similar). As for the Lagrange elements, we consider the γ^{g}-trace defined in (18.7a), i.e., $\gamma_{K,F}^{\mathrm{g}}(v) := v_{|F}$ for all $F \in \mathcal{F}_K$. Recall that the dofs of the canonical hybrid element are defined in (7.11). Let $\mathcal{N}_{K,F}$ be the collection of the dof indices of the following types: integrals over F of products with functions from the fixed basis $\{\zeta_m\}_{m \in \{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ of $\mathbb{P}_{k-3,2}$ (if $k \geq 3$); integrals over the edges of F of products with functions from the

fixed basis $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$ of $\mathbb{P}_{k-2,1}$ (if $k \geq 2$); evaluation at the vertices of F. Note that $\mathrm{card}(\mathcal{N}_{K,F}) = 3 + 3n_{\mathrm{sh}}^{\mathrm{e}} + n_{\mathrm{sh}}^{\mathrm{f}}$ if $k \geq 3$. Assume that the basis $\{\mu_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{e}}\}}$ is invariant under every permutation of the vertices of the unit simplex $\widehat{S}^1$, and the basis $\{\zeta_m\}_{m\in\{1:n_{\mathrm{sh}}^{\mathrm{f}}\}}$ in invariant under every permutation of the vertices of the unit simplices $\widehat{S}^2$. Then one can prove that the canonical hybrid element satisfies the Assumptions 20.12 and 20.14; see Exercise 20.6.

Exercises

Exercise 20.1 (Affine mapping between faces). Let $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$ and set $\widehat{F}_l := \boldsymbol{T}_{K_l}^{-1}(F)$ and $\widehat{F}_r := \boldsymbol{T}_{K_r}^{-1}(F)$. Prove that the mapping $\boldsymbol{T}_{rl} := \boldsymbol{T}_{K_l}^{-1}\circ\boldsymbol{T}_{K_r|\widehat{F}_r}$ is affine. (*Hint*: let $(\widehat{K}, \widehat{P}_{\mathrm{geo}}, \widehat{\Sigma}_{\mathrm{geo}})$ be the geometric reference Lagrange finite element. Observe that the two face finite elements $(\widehat{F}_l, \widehat{P}_{\mathrm{geo},l}^{\mathrm{g}}, \widehat{\Sigma}_{\mathrm{geo},l}^{\mathrm{g}})$ and $(\widehat{F}_r, \widehat{P}_{\mathrm{geo},r}^{\mathrm{g}}, \widehat{\Sigma}_{\mathrm{geo},r}^{\mathrm{g}})$ can be constructed from the same reference Lagrange finite element $(\widehat{F}^{d-1}, \widehat{P}_{\mathrm{geo}}^{d-1}, \widehat{\Sigma}_{\mathrm{geo}}^{d-1})$.)

Exercise 20.2 (Linear maps). Let E, F, G be finite-dimensional vector spaces, let $A \in \mathcal{L}(E;F)$ and let $T \in \mathcal{L}(E;G)$. Assume that $\ker(T) \subset \ker(A)$. Set $\tilde{G} := T(E)$. (i) Prove that there is $\tilde{A} \in \mathcal{L}(\tilde{G};F)$ s.t. $A = \tilde{A} \circ T$. (*Hint*: build a right inverse of T using a direct sum $E = E_1 \oplus E_2$ with $E_1 := \ker(T)$.) (ii) Show that $\tilde{A}$ is uniquely defined, i.e., does not depend on E_2.

Exercise 20.3 ($\gamma_{K,F}$ and $\mathcal{N}_{K,F}$). (i) Prove that $P_K = \sum_{F\in\mathcal{F}_K} \ker(\gamma_{K,F}^{\mathrm{x}})$ (nondirect sum of vector spaces) if and only if there is $F \in \mathcal{F}_K$ s.t. $i \notin \mathcal{N}_{K,F}$ for all $i \in \mathcal{N}$. (ii) Let the face unisolvence assumption hold true. Let $\mathcal{F}(K,i) := \{F \in \mathcal{F}_K \mid \ker(\gamma_{K,F}) \subset \ker(\sigma_{K,i})\}$. Prove the following statements: (ii.a) $F \in \mathcal{F}(K,i)$ iff $i \in \mathcal{N}_{K,F}$; (ii.b) $F \in \mathcal{F}(K,i)$ iff $\gamma_{K,F}(\theta_{K,i}) \neq 0$ where $\theta_{K,i}$ is the local shape function associated with the dof i.

Exercise 20.4 (Reference face element). Let $\widehat{F}$ be any face of $\widehat{K}$. Let $\widehat{P}^{\mathrm{x}} := \gamma_{\widehat{K},\widehat{F}}^{\mathrm{x}}(\widehat{P})$ and let $\mathcal{N}_{\widehat{K},\widehat{F}}$ be the subset of $\mathcal{N}$ s.t. $\bigcap_{i\in\mathcal{N}_{\widehat{K},\widehat{F}}} \ker(\sigma_{\widehat{K},i}) = \ker(\gamma_{\widehat{K},\widehat{F}})$. Recall that this means that there exists $\widehat{\sigma}_{\widehat{F},i}^{\mathrm{x}} : \widehat{P}_{\widehat{K},\widehat{F}}^{\mathrm{x}} \to \mathbb{R}$ s.t. $\widehat{\sigma}_i = \widehat{\sigma}_{\widehat{F},i}^{\mathrm{x}} \circ \gamma_{\widehat{K},\widehat{F}}^{\mathrm{x}}$ for all $i \in \mathcal{N}_{\widehat{K},\widehat{F}}$. Assume that $\mathcal{N}_{\widehat{K},\widehat{F}}$ is nonempty, that the triple $\{\widehat{F}, \widehat{P}^{\mathrm{x}}, \widehat{\Sigma}^{\mathrm{x}}\}$ with $\widehat{\Sigma}^{\mathrm{x}} := \{\widehat{\sigma}_{\widehat{F},i}^{\mathrm{x}}\}_{i\in\mathcal{N}_{\widehat{K},\widehat{F}}}$ is a finite element, and that there is a linear bijective map $\psi_F : P_{K,F}^{\mathrm{x}} \to \widehat{P}^{\mathrm{x}}$ s.t. $\psi_F^{-1} \circ \gamma_{\widehat{K},\widehat{F}}^{\mathrm{x}} = \gamma_{K,F}^{\mathrm{x}} \circ \psi_K^{-1}$. Prove that Assumption 20.12 holds true and $\mathcal{N}_{K,F} = \mathcal{N}_{\widehat{K},\widehat{F}}$. (*Hint*: show that the finite element $\{F, P_{K,F}^{\mathrm{x}}, \Sigma_{K,F}^{\mathrm{x}}\}$ is generated from $\{\widehat{F}, \widehat{P}^{\mathrm{x}}, \widehat{\Sigma}^{\mathrm{x}}\}$ using the map ψ_F.)

Exercise 20.5 (Permutation invariance). Let $\widehat{S}^1 := [0, 1]$ and consider the bases $\mathfrak{B}_1 := \{\mu_1(s) = 1 - s, \mu_2(s) = s\}$ and $\mathfrak{B}_2 := \{\mu_1(s) = 1, \mu_2(s) = s\}$. Are these bases invariant under permutation of the vertices of $\widehat{S}^1$?

Exercise 20.6 (Canonical hybrid element, $d = 3$). Consider the assumptions made in §20.4.3. (i) Prove the face unisolvence assumption 20.12. (ii) Let $F \in \mathcal{F}_K$. Let $\boldsymbol{T}_{\widehat{F}} : \widehat{S}^2 \to \widehat{F}$ be an affine bijective mapping, and let $\boldsymbol{T}_{K,F} := \boldsymbol{T}_{K|\widehat{F}} \circ \boldsymbol{T}_{\widehat{F}} : \widehat{S}^2 \to F$. Verify that $P_{K,F}^{\mathrm{g}} = \mathbb{P}_{k,d-1} \circ \boldsymbol{T}_{K,F}^{-1}$ and that $\{F, P_{K,F}^{\mathrm{g}}, \Sigma_{K,F}^{\mathrm{g}}\}$ is a two-dimensional canonical hybrid element. (iii) Prove that $P_{K_l,F}^{\mathrm{g}} = P_{K_r,F}^{\mathrm{g}} =: P_F^{\mathrm{g}}$ and $\Sigma_{K_l,F}^{\mathrm{g}} = \Sigma_{K_r,F}^{\mathrm{g}} =: \Sigma_F^{\mathrm{g}}$.

Exercise 20.7 ($P_{K,F}$). Let $\widehat{K}$ be the unit simplex in $\mathbb{R}^2$ and let $\{\widehat{F}_i\}_{i \in \{0:2\}}$ be the faces of $\widehat{K}$. Recall that for $\mathbb{P}_{k,d}$ scalar-valued elements we have $P_{\widehat{K},\widehat{F}_i} := \gamma_{\widehat{K},\widehat{F}_i}^{\mathrm{g}}(\mathbb{P}_{k,d})$. (i) Compute a basis of $P_{\widehat{K},\widehat{F}_i}$ for all $i \in \{0:2\}$ assuming that $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ is the $\mathbb{P}_1$ Lagrange element. Is $(\widehat{F}_i, P_{\widehat{K},\widehat{F}_i}, \Sigma_{\widehat{K},\widehat{F}_i})$ a finite element? (ii) Compute a basis of $P_{\widehat{K},\widehat{F}_i}$ for all $i \in \{0:2\}$ assuming that $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ is the $\mathbb{P}_1$ Crouzeix–Raviart element. Is $(\widehat{F}_i, P_{\widehat{K},\widehat{F}_i}, \Sigma_{\widehat{K},\widehat{F}_i})$ a finite element?

Part IV, Chapter 21

Construction of the connectivity classes

In this chapter, we finish the construction of the connectivity classes which we characterize by means of an equivalence relation on the pairs in $\mathcal{T}_h \times \mathcal{N}$. We show that the resulting equivalence classes verify the two key assumptions (19.2) and (19.3) introduced in Chapter 19. Our starting point is to assume that the finite element at hand satisfies the two fundamental assumptions introduced in Chapter 20: the face unisolvence assumption (Assumption 20.12) and the face matching assumption (Assumption 20.14). These two assumptions turn out to be sufficient to fully characterize the connectivity classes of Raviart–Thomas elements. For the other elements (Lagrange, canonical hybrid, and Nédélec) for which there are degrees of freedom (dofs) attached to geometric entities of smaller dimension, we have to consider two additional abstract assumptions, the M-unisolvence assumption (Assumption 21.9) and the M-matching assumption (Assumption 21.10), which we show hold true for these elements. At the end of the chapter we propose enumeration techniques that facilitate the practical construction of the map χ_{lr} introduced in Assumption 20.14. This map is a key tool for the construction of the connectivity array j_dof. We assume in the entire chapter that the reference cell is either a simplex or a cuboid, we assume that $d = 3$, and we continue to use the notation introduced in Chapters 19 and 20.

21.1 Connectivity classes

In this section, we describe a way to build the connectivity classes that makes the two key assumptions from Chapter 19 hold true. This is done by constructing an equivalence relation on the set $\mathcal{T}_h \times \mathcal{N}$.

21.1.1 Geometric entities and macroelements

We start by introducing the geometric objects to which we will attach the dofs. Let $\mathcal{T}_h$ be a matching mesh and let $\mathcal{V}_h$, $\mathcal{E}_h$, and $\mathcal{F}_h$ be the sets collecting, respectively, the vertices, edges, and faces in the mesh $\mathcal{T}_h$ as defined in §8.2.

Definition 21.1 (Geometric entity). *Let $\mathcal{T}_h$ be a matching mesh. We call M geometric entity if M is a vertex $\boldsymbol{z} \in \mathcal{V}_h$, an edge $E \in \mathcal{E}_h$, a face $F \in \mathcal{F}_h$, or a cell $K \in \mathcal{T}_h$.*

Definition 21.2 (Macroelement). *Let M be a geometric entity. We associate with M the following subsets of $\mathcal{T}_h$ and D:*

$$\mathcal{T}_M := \{K \in \mathcal{T}_h \mid M \subset K\} \subset \mathcal{T}_h, \tag{21.1a}$$
$$D_M := \mathrm{int}(\{\boldsymbol{x} \in D \mid \exists K \in \mathcal{T}_M, \boldsymbol{x} \in K\}) \subset D. \tag{21.1b}$$

The set D_M is called macroelement *associated with the geometric entity M.*

Notice that the notion of macroelement is trivial for a mesh cell since in this case $\mathcal{T}_K := \{K\}$ and $D_K := \mathrm{int}(K)$. This notion is also very simple for a mesh face, since if $F \in \mathcal{F}_h^\circ$, then $\mathcal{T}_F := \{K_l, K_r\}$ where $F := \partial K_l \cap \partial K_r$ (so that $\mathrm{card}(\mathcal{T}_F) = 2$), whereas if $F \in \mathcal{F}_h^\partial$, then $\mathcal{T}_F := \{K_l\}$ where $F := \partial K_l \cap \partial D$ (so that $\mathrm{card}(\mathcal{T}_F) = 1$). For a vertex $\boldsymbol{z} \in \mathcal{V}_h$ or an edge $E \in \mathcal{E}_h$, there are in general more than two cells in $\mathcal{T}_{\boldsymbol{z}}$ and $\mathcal{T}_E$, and $\mathrm{card}(\mathcal{T}_{\boldsymbol{z}})$ and $\mathrm{card}(\mathcal{T}_E)$ are not known a priori. Figure 21.1 illustrates these concepts for a triangular mesh. Notice that if the geometric entity M is s.t. $\mathrm{card}(\mathcal{T}_M) \geq 2$, then M is a face, an edge, or a vertex. Hence, $\mathcal{T}_M$ can also be characterized as follows when $\mathrm{card}(\mathcal{T}_M) \geq 2$:

$$\mathcal{T}_M = \{K \in \mathcal{T}_h \mid M \subset \partial K\} \subset \mathcal{T}_h. \tag{21.2}$$
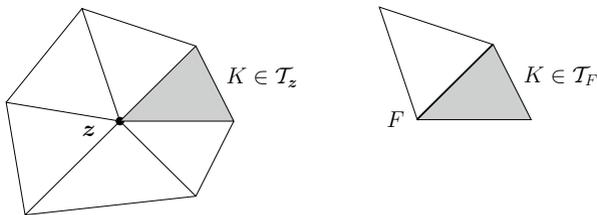
Fig. 21.1 Left: mesh vertex $\boldsymbol{z} \in \mathcal{V}_h$, macroelement $\mathcal{T}_{\boldsymbol{z}}$ composed of six mesh cells with one cell $K \in \mathcal{T}_{\boldsymbol{z}}$ highlighted in gray. Right: mesh face $F \in \mathcal{F}_h$, macroelement $\mathcal{T}_F$ composed of two mesh cells with one cell $K \in \mathcal{T}_F$ highlighted in gray. Note that the subsets $D_{\boldsymbol{z}}$ and D_F are connected.

Definition 21.3 (M-path). *Let M be a geometric entity. A collection of cells $(K_0, \ldots, K_L)$ in $\mathcal{T}_M$ is called M-path if either $L = 0$ or the following*

holds true for all $l \in \{1{:}L\}$: $F_l := \partial K_{l-1} \cap \partial K_l \in \mathcal{F}_h^\circ$. *We say that* L *is the* length *of the* M*-path and that the* M*-path connects* K_0 *with* K_L.

Lemma 21.4 (M-path). *Let* M *be a geometric entity. Assume* $\mathrm{card}(\mathcal{T}_M) \geq 2$. *Then for every pair* (K, K') *of distinct cells in* $\mathcal{T}_M$, *there exists an* M*-path of length* $L \geq 1$ *connecting* K *with* K', *and we have* $M \subset \bigcap_{l \in \{1{:}L\}} F_l$.

Proof. The subset D_M is connected since D is a Lipschitz domain. This implies the existence of the M-path. Finally, since $\mathrm{card}(\mathcal{T}_M) \geq 2$, (21.2) holds true, and since $K_{l-1}, K_l \in \mathcal{T}_M$ for all $l \in \{1{:}L\}$, we have $M \subset \partial K_{l-1}$ and $M \subset \partial K_l$. Hence, $M \subset F_l$ for all $l \in \{1{:}L\}$. $\qquad\square$

It will be useful to describe geometric entities as an intersection of faces.

Lemma 21.5 (Geometric entity as intersection of faces). *Let* $K \in \mathcal{T}_h$ *be a mesh cell. The following holds true:* (i) *Let* $\mathcal{G} \subset \mathcal{F}_K$ *be a nonempty collection of faces of* K. *Then* $M := \bigcap_{F \in \mathcal{G}} F$ *is always a geometric entity when* $M \neq \emptyset$. (ii) *Let* M *be a geometric entity that is not a cell. Then there is a unique subset* $\mathcal{G}_{K,M} \subset \mathcal{F}_K$ *s.t.* $M = \bigcap_{F \in \mathcal{G}_{K,M}} F$.

Proof. (i) $\bigcap_{F \in \mathcal{G}} F$ is always a geometric entity when it is nonempty because K is a polyhedron.
(ii) Whether K is a simplex or a cuboid, if $M := \bigcap_{F \in \mathcal{G}} F$ is nonempty, then M is a vertex, an edge, or a face of K, and there cannot be any other possibility. If M is a vertex, there can only be exactly d faces s.t. $M = \bigcap_{F \in \mathcal{G}} F$. If M is an edge, there can only be exactly 2 faces s.t. $M = \bigcap_{F \in \mathcal{G}} F$. If M is a face, $\mathcal{G}$ contains only one face. $\qquad\square$

Remark 21.6 (Prisms). The proof of Lemma 21.5 shows that for the statement (ii) to hold true when $d = 3$, every vertex has to be shared by exactly d faces. In addition to the tetrahedron and the hexahedron, another polyhedron having this property is the prism with triangular basis. $\qquad\square$

21.1.2 The two key assumptions

Let us briefly motivate what we want to do. Our goal is to partition the set $\mathcal{N}$ according to the nature of the dofs and to use the same partition on every mesh cell. Let $K \in \mathcal{T}_h$. We say that i is an *internal dof* if there is no face $F \in \mathcal{F}_K$ s.t. $i \in \mathcal{N}_{K,F}$, and we write $i \in \mathcal{N}^\circ$. We say that i is a *boundary dof* if there is at least one face $F \in \mathcal{F}_K$ s.t. $i \in \mathcal{N}_{K,F}$, and we write $i \in \mathcal{N}^\partial$. A first natural partition of the dofs is thus $\mathcal{N} = \mathcal{N}^\circ \cup \mathcal{N}^\partial$. If all the subsets $\mathcal{N}_{K,F}$ are mutually disjoint, as it happens for the Raviart–Thomas elements, the collection of boundary dofs is further partitioned as $\mathcal{N}^\partial = \bigcup_{F \in \mathcal{F}_K} \mathcal{N}_{K,F}$. The situation is more intricate when the subsets $\mathcal{N}_{K,F}$ are not mutually disjoint since in this case we need to consider the intersections $\bigcap_{F \in \mathcal{G}} \mathcal{N}_{K,F}$ for the nonempty subsets $\mathcal{G} \subset \mathcal{F}_K$, and we are only interested in the subsets $\mathcal{G} \subset \mathcal{F}_K$ s.t. the above intersection is nonempty. The following lemma shows that for the finite elements considered in this book, the set $\bigcap_{F \in \mathcal{G}} F$ is nonempty if the set $\bigcap_{F \in \mathcal{G}} \mathcal{N}_{K,F}$ is nonempty.

Lemma 21.7 (Intersection of boundary dofs). *Let $K \in \mathcal{T}_h$ be a simplex or a cuboid. If K is a simplex, assume that there is no local shape function that has a nonzero γ-trace on all the faces of K. If K is a cuboid, assume that there is no local shape function that has a nonzero γ-trace on two opposite faces of K. Then for every nonempty set $\mathcal{G} \subset \mathcal{F}_K$, if $\bigcap_{F \in \mathcal{G}} \mathcal{N}_{K,F}$ is nonempty, then $\bigcap_{F \in \mathcal{G}} F$ is nonempty as well.*

Proof. Let us reason by contradiction and assume that $\bigcap_{F \in \mathcal{G}} F = \emptyset$. If K is a simplex, this implies that $\mathcal{G} = \mathcal{F}_K$, whereas if K is a cuboid, this implies that $\mathcal{G}$ contains two opposite faces. Recalling that $i \in \mathcal{N}_{K,F}$ iff $\gamma_{K,F}(\theta_{K,i}) \neq 0$, we infer from our assumption on the shape functions that $\bigcap_{F \in \mathcal{G}} \mathcal{N}_{K,F}$ is empty. This concludes the proof. $\square$

All the simplicial finite elements considered in this book satisfy the assumption of Lemma 21.7 since the γ-trace of every shape function vanishes on at least one face. All the cuboidal finite elements considered in this book also satisfy the assumption of Lemma 21.7 since there is no shape function that has a nonzero γ-trace on two opposite faces.

Lemma 21.7 combined with Lemma 21.5 allows us to identify the geometric entities that are different from K with those nonempty subsets $\mathcal{G} \subset \mathcal{F}_K$ such that $\bigcap_{F \in \mathcal{G}} \mathcal{N}_{K,F}$ is nonempty. This leads to the following definition.

Definition 21.8 ($\mathcal{M}_h$). *We denote by $\mathcal{M}_h$ the collection of all the geometric entities M s.t. for all $K \in \mathcal{T}_M$, the unique nonempty subset $\mathcal{G}_{K,M} \subset \mathcal{F}_K$ satisfying $M = \bigcap_{F \in \mathcal{G}_{K,M}} F$ is s.t.*

$$\mathcal{N}_{K,M} := \bigcap_{F \in \mathcal{G}_{K,M}} \mathcal{N}_{K,F} \neq \emptyset. \tag{21.3}$$

We say that the finite element has face dofs *if $\mathcal{F}_h \subset \mathcal{M}_h$,* edge dofs *if $\mathcal{E}_h \subset \mathcal{M}_h$, and* vertex dofs *if $\mathcal{V}_h \subset \mathcal{M}_h$.*

Since $\mathcal{N}_{K,F}$ is nonempty for all $K \in \mathcal{T}_h$ and all $F \in \mathcal{F}_K$ (see Assumption 20.12 on face unisolvence), all the mesh faces are in $\mathcal{M}_h$, i.e., $\mathcal{F}_h \subset \mathcal{M}_h$. This means that all the finite elements considered in this book have face dofs. We will see in the next section that $\mathcal{M}_h = \mathcal{V}_h \cup \mathcal{E}_h \cup \mathcal{F}_h$ for the Lagrange elements and the canonical hybrid element, $\mathcal{M}_h = \mathcal{E}_h \cup \mathcal{F}_h$ for Nédélec elements, and $\mathcal{M}_h = \mathcal{F}_h$ for Raviart–Thomas elements.

We can now state the two key assumptions regarding the structure of the dofs that will help us identify the connectivity classes.

Assumption 21.9 (M-unisolvence). *For every geometric entity $M \in \mathcal{M}_h$ and every cell $K \in \mathcal{T}_M$ (i.e., $M \subset \partial K$), the following holds true: (i) There is a linear map $\gamma_{K,M}$ s.t. for every face $F \in \mathcal{G}_{K,M}$, we have $\ker(\gamma_{K,F}) \subset \ker(\gamma_{K,M})$. (ii) For all $i \in \mathcal{N}_{K,M}$, there is a linear form $\sigma_{K,M,i}$ s.t. $\sigma_{K,i} = \sigma_{K,M,i} \circ \gamma_{K,M}$. (iii) The triple $(M, P_{K,M}, \Sigma_{K,M})$ is a finite element where $P_{K,M} := \gamma_{K,M}(P_K)$ and $\Sigma_{K,M} := \{\sigma_{K,M,i}\}_{i \in \mathcal{N}_{K,M}}$.*

Assumption 21.10 (M-matching). *The following holds true for every interface $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$ and every geometric entity $M \in \mathcal{M}_h$ s.t. $M \subset F$ (so that $K_l, K_r \in \mathcal{T}_M$ and $F \in \mathcal{G}_{K_l,M} \cap \mathcal{G}_{K_r,M}$): (i) $P_{K_l,M} = P_{K_r,M}$. (ii) The map χ_{lr} introduced in Assumption 20.14 is such that $\chi_{lr}(\mathcal{N}_{K_l,M}) = \mathcal{N}_{K_r,M}$, and the map $\chi_{lr,M} : \mathcal{N}_{K_l,M} \to \mathcal{N}_{K_r,M}$ defined by $\chi_{lr,M} := \chi_{F|\mathcal{N}_{K_l,M}}$ is s.t.*

$$\sigma_{K_l,M,i_l} = \sigma_{K_r,M,\chi_{lr,M}(i_l)}, \qquad \forall i_l \in \mathcal{N}_{K_l,M}, \tag{21.4}$$

i.e., $\Sigma_{K_l,M} = \Sigma_{K_r,M}$ and $\chi_{lr,M} : \mathcal{N}_{K_l,M} \to \mathcal{N}_{K_r,M}$ is bijective.

The definition of $\chi_{lr,M}$ in Assumption 21.10 is meaningful because $\mathcal{N}_{K_l,M} \subset \mathcal{N}_{K_l,F}$ and $\mathcal{N}_{K_r,M} \subset \mathcal{N}_{K_r,F}$ owing to (21.3). When the geometric entity M is a face, Assumption 21.9 and Assumption 21.10 are identical to Assumption 20.12 (face unisolvence) and Assumption 20.14 (face matching).

Given an M-path (see Definition 21.3) of length $L \geq 1$, we define the map $\chi_{F_l}^\epsilon$ for all $l \in \{1{:}L\}$ by setting $\chi_{F_l}^\epsilon := \chi_{F_l}$ if $\boldsymbol{n}_{F_l}$ points from K_{l-1} to K_l and $\chi_{F_l}^\epsilon := \chi_{F_l}^{-1}$ otherwise, where $\boldsymbol{n}_{F_l}$ is the unit normal vector orienting F_l.

Lemma 21.11 (Path independence). *Let $M \in \mathcal{M}_h$. Let K, K' be two cells in $\mathcal{T}_M$ (possibly identical) connected by an M-path of length $L \geq 1$, say $(K =: K_0, \ldots, K_L := K')$. Then for all $i \in \mathcal{N}_{K,M}$, the index $\chi_{F_L}^\epsilon \circ \ldots \circ \chi_{F_1}^\epsilon(i)$ with $F_l := \partial K_{l-1} \cap \partial K_l$, $\forall l \in \{1{:}L\}$, is independent of the M-path.*

Proof. Let $M \in \mathcal{M}_h$ be a geometric entity and let K, K' be two cells in $\mathcal{T}_M$. Let $(K =: K_{\beta,0} \ldots, K_{\beta,L_\beta} := K')$, $\forall \beta \in \{1, 2\}$, be two M-paths in $\mathcal{T}_M$ connecting K to K', with $F_{\beta,l} := \partial K_{\beta,l-1} \cap \partial K_{\beta,l}$ for all $l \in \{1{:}L_\beta\}$. Let $i'_1 := \chi_{F_{1,L_1}}^\epsilon \circ \ldots \circ \chi_{F_{1,1}}^\epsilon(i)$ and $i'_2 := \chi_{F_{2,L_2}}^\epsilon \circ \ldots \circ \chi_{F_{2,1}}^\epsilon(i)$. Assumption 21.10 implies that $\sigma_{K,M,i} = \sigma_{K_{1,1},M,\chi_{F_{1,1}}^\epsilon(i)} = \ldots = \sigma_{K',M,i'_1}$ and $\sigma_{K,M,i} = \sigma_{K_{2,1},M,\chi_{F_{2,1}}^\epsilon(i)} = \ldots = \sigma_{K',M,i'_2}$. Hence, $\sigma_{K',M,i'_1} = \sigma_{K',M,i'_2}$. But by Assumption 21.9 $(M, P_{K',M}, \Sigma_{K',M})$ is a finite element. Hence, $\sigma_{K',M,i'_1} = \sigma_{K',M,i'_2}$ iff $i'_1 = i'_2$. □

21.1.3 Connectivity classes as equivalence classes

For all $(K, i) \in \mathcal{T}_h \times \mathcal{N}$, we introduce the smallest geometric entity associated with the dof $\sigma_{K,i}$. This object is the last brick we need to define the equivalence relation mentioned at the beginning of the chapter.

Lemma 21.12 ($\mathcal{M}_{K,i}$). *Let $K \in \mathcal{T}_h$ and i be a boundary dof. Then the following set is nonempty and is a member of $\mathcal{M}_h$:*

$$\mathcal{M}_{K,i} := \bigcap_{\{M \in \mathcal{M}_h \ | \ i \in \mathcal{N}_{K,M}\}} M. \tag{21.5}$$

Proof. The subset $\mathcal{G}_{K,i} := \{F \in \mathcal{F}_K \mid i \in \mathcal{N}_{K,F}\}$ is nonempty since i is a boundary dof. Then the set $\widetilde{\mathcal{M}}_{K,i} := \bigcap_{F \in \mathcal{G}_{K,i}} F$ is nonempty owing to Lemma 21.7 since $i \in \bigcap_{F \in \mathcal{G}_{K,i}} \mathcal{N}_{K,F}$, and it is a geometric entity owing to

Lemma 21.5. The rest of the proof consists of showing that $\widetilde{\mathcal{M}}_{K,i} = \mathcal{M}_{K,i}$. Since $\mathcal{G}_{K,i} \subset \{M \in \mathcal{M}_h \mid i \in \mathcal{N}_{K,M}\}$, we have $\mathcal{M}_{K,i} \subset \widetilde{\mathcal{M}}_{K,i}$. To prove the converse inclusion, let us consider M in the set $\{M \in \mathcal{M}_h \mid i \in \mathcal{N}_{K,M}\}$. By Lemma 21.5, there is $\emptyset \neq \mathcal{G}_{K,M} \subset \mathcal{F}_K$ s.t. $M = \bigcap_{F \in \mathcal{G}_{K,M}} F$, and the definition (21.3) of $\mathcal{N}_{K,M}$ implies that $i \in \bigcap_{F \in \mathcal{G}_{K,M}} \mathcal{N}_{K,F}$. Hence, for all $F \in \mathcal{G}_{K,M}$, we have $i \in \mathcal{N}_{K,F}$, which means that $\mathcal{G}_{K,M} \subset \mathcal{G}_{K,i}$, and this in turn yields $\widetilde{\mathcal{M}}_{K,i} = \bigcap_{F \in \mathcal{G}_{K,i}} F \subset \bigcap_{F \in \mathcal{G}_{K,M}} F = M$. Since the geometric entity M is arbitrary in $\{M \in \mathcal{M}_h \mid i \in \mathcal{N}_{K,M}\}$, we conclude that $\widetilde{\mathcal{M}}_{K,i} \subset \mathcal{M}_{K,i}$.   □

We now partition the product set $\mathcal{T}_h \times \mathcal{N}$ into equivalence classes.

Definition 21.13 (Binary relation $\mathcal{R}$). *We say that $(K, i) \, \mathcal{R} \, (K', i')$ if and only if either $(K, i) = (K', i')$, or $K \neq K'$, $\mathcal{M}_{K,i} = \mathcal{M}_{K',i'} := M$, and given an M-path connecting K to K' in $\mathcal{T}_M$, say $(K = K_0, \ldots, K_L = K')$, with $F_l := \partial K_{l-1} \cap \partial K_l$, $\forall l \in \{1{:}L\}$, we have $i' = \chi^\epsilon_{F_L} \circ \ldots \circ \chi^\epsilon_{F_1}(i)$.*

This definition makes sense when $K \neq K'$ since in this case M cannot be equal to either K or K', and since $M \subset K \cap K'$, the cells K and K' are in $\mathcal{T}_M$. Owing to Lemma 21.4, K and K' can be connected by an M-path, and owing to Lemma 21.11, the index $\chi^\epsilon_{F_L} \circ \ldots \circ \chi^\epsilon_{F_1}(i)$ is independent of the M-path that is chosen to connect K to K'.

Lemma 21.14 (Equivalence relation). *Let Assumptions 21.9 and 21.10 hold true. Then the binary relation $\mathcal{R}$ is an equivalence relation.*

Proof. $\mathcal{R}$ is by definition reflexive. By enumerating the cells in the M-path in reverse order, we infer that $\mathcal{R}$ is symmetric. Finally, let us prove that $\mathcal{R}$ is transitive. Let $(K, i) \, \mathcal{R} \, (K', i')$ and $(K', i') \, \mathcal{R} \, (K'', i'')$. Then $\mathcal{M}_{K,i} = \mathcal{M}_{K',i'} = \mathcal{M}_{K'',i''} := M$. If $(K, i) = (K', i')$ or $(K', i') = (K'', i'')$, there is nothing to prove. Otherwise, we have $K \neq K'$ and $K' \neq K''$. Let $(K =: K_{1,0} \ldots, K_{1,L_1} := K')$, $(K' =: K_{2,0} \ldots, K_{2,L_2} := K'')$ be two M-paths, respectively, connecting K to K' and K' to K''. Let us set $F_{\beta,l} := \partial K_{\beta,l-1} \cap \partial K_{\beta,l}$ for all $l \in \{1{:}L_\beta\}$ and all $\beta \in \{1, 2\}$. Then $(K =: K_{1,0} \ldots, K_{1,L_1} = K_{2,0} \ldots, K_{2,L_2} := K'')$ is an M-path and $i'' = \chi^\epsilon_{F_{2,L_2}} \circ \ldots \circ \chi^\epsilon_{F_{2,1}}(i') = \chi^\epsilon_{F_{2,L_2}} \circ \ldots \circ \chi^\epsilon_{F_{2,1}} \circ \chi^\epsilon_{F_{1,L_1}} \circ \ldots \circ \chi^\epsilon_{F_{1,1}}(i)$. If $K \neq K''$, this argument proves that $(K, i)\mathcal{R}(K'', i'')$. If $K = K''$, Assumption 21.10 implies that $\sigma_{K,M,i} = \sigma_{K'',M,i''} = \sigma_{K,M,i''}$, which is possible only if $i = i''$ owing to Assumption 21.9. Hence, we have again $(K, i)\mathcal{R}(K'', i'')$.   □

Let $\mathcal{A}_h$ be the set of the equivalence classes induced by $\mathcal{R}$ over $\mathcal{T}_h \times \mathcal{N}$. Let us now consider any map $\mathtt{j\_dof} : \mathcal{T}_h \times \mathcal{N} \to \mathcal{A}_h$ such that

$$[\, \mathtt{j\_dof}(K, i) = \mathtt{j\_dof}(K', i') \,] \iff [\, (K, i)\mathcal{R}(K', i') \,]. \qquad (21.6)$$

Letting I be the cardinality of $\mathcal{A}_h$, there are $I!$ ways to define $\mathtt{j\_dof}$. Whichever choice that is made to define $\mathtt{j\_dof}$, let us now prove that the

two assumptions (19.2) and (19.3) made in Chapter 19 hold true. Recall that these are the two structural conditions that we required from j_dof in Chapter 19 to construct the conforming subspace $P_k^x(\mathcal{T}_h; \mathbb{R}^q)$.

Lemma 21.15 (Equivalence relation at interfaces). *Let $F \in \mathcal{F}_h^\circ$ with $F := \partial K_l \cap \partial K_r$ and let χ_{lr} be the map introduced in Assumption 20.14. The following holds true for all $i_l \in \mathcal{N}_{K_l,F}$: (i) $\mathcal{M}_{K_l,i_l} = \mathcal{M}_{K_r,\chi_{lr}(i_l)}$; (ii) j_dof$(K_l, i_l) = $ j_dof$(K_r, \chi_{lr}(i_l))$.*

Proof. Since $\chi_{lr}(\mathcal{N}_{K_l,M}) = \mathcal{N}_{K_r,M}$ owing to the M-matching assumption, we have

$$\{M \in \mathcal{M}_h \mid \chi_{lr}(i_l) \in \mathcal{N}_{K_r,M}\} = \{M \in \mathcal{M}_h \mid \chi_{lr}(i_l) \in \chi_{lr}(\mathcal{N}_{K_l,M})\}$$
$$= \{M \in \mathcal{M}_h \mid i_l \in \mathcal{N}_{K_l,M}\}.$$

Owing to the identity (21.5), we infer that $\mathcal{M}_{K_l,i_l} = \mathcal{M}_{K_r,\chi_{lr}(i_l)}$. The second claim follows readily because $\mathcal{M}_{K_l,i_l} = \mathcal{M}_{K_r,\chi_{lr}(i_l)}$ and the two distinct cells K_l and K_r can be connected by an M-path of length 1 crossing F in such a way that (trivially) $\chi_{lr}(i_l) = \chi_{lr}(i_l)$. This proves that $(K_l, i_l)\mathcal{R}(K_r, \chi_{lr}(i_l))$, i.e., we have j_dof$(K_l, i_l) = $ j_dof$(K_r, \chi_{lr}(i_l))$ owing to (21.6). □

Let $a \in \mathcal{A}_h$ with representative (K, i). Let us set $M := \mathcal{M}_{K,i}$ and $\chi_{K,K,M}(i) := i$. For all $K' \in \mathcal{T}_M$ such that $K \neq K'$, let us set $\chi_{K,K',M}(i) := \chi_{F_L}^\epsilon \circ \ldots \circ \chi_{F_1}^\epsilon(i)$, where $(K =: K_0, \ldots K_L := K')$ is any M-path connecting K to K'. Lemma 21.11 together with Item (i) from Lemma 21.15 gives the following characterization of the connectivity class a:

$$a = \bigcup_{K' \in \mathcal{T}_M} \{(K', \chi_{K,K',M}(i))\}. \tag{21.7}$$

We conclude by stating the main result of this section.

Theorem 21.16 (Verification of the assumptions from Chapter 19). *Let Assumptions 21.9 and 21.10 hold true. Let j_dof be defined in (21.6). Then Assumptions (19.2) and (19.3) hold true.*

Proof. Let us start with (19.3) which is easier to verify. By definition, we have $(K, i)\mathcal{R}(K, i')$ iff $i = i'$, that is, j_dof$(K, i) = $ j_dof(K, i') implies that $i = i'$. Let us now prove (19.2) for all $v_h \in P_k^{x,b}(\mathcal{T}_h; \mathbb{R}^q)$. Let us start with the implication $\Longrightarrow$ in (19.2), i.e., we assume that $v_h \in P_k^x(\mathcal{T}_h; \mathbb{R}^q)$. Let $(K, i), (K', i')$ be two pairs in the same connectivity class and let $M := \mathcal{M}_{K,i} = \mathcal{M}_{K',i'}$. We want to show that $\sigma_{K,i}(v_{h|K}) = \sigma_{K',i'}(v_{h|K'})$. Since this claim is obvious if $K = K'$, we assume that $K \neq K'$ and we consider an M-path connecting K to K' in $\mathcal{T}_M$, say $(K =: K_0 \ldots, K_L := K')$ and $F_l := \partial K_{l-1} \cap \partial K_l$, $\forall l \in \{1{:}L\}$. Repeated applications of the implication $\Longrightarrow$ from Lemma 20.15 show that since $[\![v_h]\!]_{F_l}^x = 0$ for all $l \in \{1{:}L\}$, we have $\sigma_{K,i}(v_{h|K}) = \sigma_{K',\chi_{F_L}^\epsilon \circ \ldots \circ \chi_{F_1}^\epsilon(i)}(v_{h|K'}) = \sigma_{K',i'}(v_{h|K'})$, which is the

desired result. Let us now prove the other implication $\impliedby$ in (19.2). Let us consider $v_h \in P_k^{x,b}(\mathcal{T}_h; \mathbb{R}^q)$ and let $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$ be a mesh interface. For all $i_l \in \mathcal{N}_{K_l, F}$, we have $\mathtt{j\_dof}(K_l, i_l) = \mathtt{j\_dof}(K_r, \chi_{lr}(i_l))$ owing to Lemma 21.15. By assumption, we also have $\sigma_{K_l, i_l}(v_{h|K_l}) = \sigma_{K_r, \chi_{lr}(i_l)}(v_{h|K_r})$ for all $i_l \in \mathcal{N}_{K_l, F}$. Owing to the implication $\impliedby$ from Lemma 20.15, we infer that $[\![v_h]\!]_F^x = 0$. Since this result holds true for all $F \in \mathcal{F}_h^\circ$, we conclude that $v_h \in P_k^x(\mathcal{T}_h; \mathbb{R}^q)$. $\square$

21.2 Verification of the assumptions

The goal of this section is to verify that Assumptions 21.9 and 21.10 are indeed satisfied by the Lagrange, canonical hybrid, Nédélec, and Raviart–Thomas elements. We assume that $d = 3$.

21.2.1 Lagrange and canonical hybrid elements

For the Lagrange elements there are four types of geometric entities: cells, faces, edges, and vertices. We have to verify Assumptions 21.9 and 21.10 for the vertices and the edges.

Assume first that M is a vertex, say $M := \{z\}$. For all $K \in \mathcal{T}_z$, let $\boldsymbol{a}_{K,i}$ be the unique vertex in K such that $\boldsymbol{a}_{K,i} = z$ and let us set $\gamma_{K,z}(p) := p(\boldsymbol{a}_{K,i})$ for all $p \in P_K$. Clearly $\ker(\gamma_{K,F}) \subset \ker(\gamma_{K,z})$ for all $F \in \mathcal{F}_K$. Then $P_{K,z} := \gamma_{K,z}(P_K) = \mathbb{R}$ because $p(\boldsymbol{a}_{K,i}) = p(z_i)$ spans $\mathbb{R}$ when p spans P_K. Furthermore, setting $\sigma_{K,z,i}(x) := x$ for all $x \in \mathbb{R}$, we have $\sigma_{K,i}(p) = p(\boldsymbol{a}_{K,i}) = \sigma_{K,z,i}(p(\boldsymbol{a}_{K,i})) = (\sigma_{K,z,i} \circ \gamma_{K,z})(p)$. We observe that $P_{K,z}$ and $\Sigma_{K,z} := \{\sigma_{K,z,i}\}$ do not depend on K and that $(z, P_{K,z}, \Sigma_{K,z})$ is a finite element.

Assume now that $M := E$ is an edge of K, and let us set $\widehat{E} := \boldsymbol{T}_K^{-1}(E)$. We define $\gamma_{K,E}(p) := p_{|E}$ for all $p \in P_K$. Hence, $\ker(\gamma_{K,F}) \subset \ker(\gamma_{K,E})$ for all $F \in \mathcal{F}_K$. Moreover, $\gamma_{K,E}(p) = \widehat{p} \circ \boldsymbol{T}_{K|\widehat{E}}^{-1} = \widehat{p} \circ \boldsymbol{T}_{\widehat{E}} \circ \boldsymbol{T}_{\widehat{E}}^{-1} \circ \boldsymbol{T}_{K|\widehat{E}}^{-1}$, where $\boldsymbol{T}_{\widehat{E}} : \widehat{S}^1 \to \widehat{E}$ is any bijective affine mapping between the unit segment in $\mathbb{R}$ and the reference edge $\widehat{E}$. By proceeding as in the proof of Lemma 7.10, we conclude that $P_{K,E} := \gamma_{K,E}(P_K) = \mathbb{P}_{k,1} \circ \boldsymbol{T}_{K,E}^{-1}$ with $\boldsymbol{T}_{K,E} := \boldsymbol{T}_{K|\widehat{E}} \circ \boldsymbol{T}_{\widehat{E}}$. By proceeding as in the proof of Lemma 20.6, we conclude that $P_{K_l,E} = P_{K_r,E}$ for all $K_l, K_r \in \mathcal{T}_E$ with a common interface. For every Lagrange node $\boldsymbol{a}_{K,i}$ located on E, we define $\sigma_{K,E,i}(p) := p(\boldsymbol{a}_{K,i})$ for all $p \in P_{K,E}$, and we denote by $\Sigma_{K,M}$ the collection of these dofs. All the Lagrange finite elements considered in this book are such that $(E, P_{K,E}, \Sigma_{K,E})$ is a finite element.

In conclusion, we have verified that Assumption 21.9 and Item (i) of Assumption 21.10 hold true, whether M is a vertex or an edge. It remains to verify that one can construct a map $\chi_{lr} : \mathcal{N}_{K_l, F} \to \mathcal{N}_{K_r, F}$ s.t. Item (ii) of Assumption 21.10 also holds true. This construction is done in §21.3.

Similar arguments as above can be invoked for the canonical hybrid element. We invite the reader to verify that Assumption 21.9 and Item (i) of Assumption 21.10 hold true for the canonical hybrid element, whether M is a vertex or an edge.

21.2.2 Nédélec elements

We invite the reader to verify that Assumption 21.9 and Item (i) of Assumption 21.10 hold true for the edge dofs of the $\mathbb{N}_{k,d}$. It remains to verify that one can construct a map $\chi_{lr} : \mathcal{N}_{K_l,F} \to \mathcal{N}_{K_r,F}$ s.t. Item(ii) of Assumption 21.10 also holds true. This construction is done in §21.3.

21.2.3 Raviart–Thomas elements

There is nothing to prove for these elements since Assumption 21.9 is identical to Assumption 20.12 and Assumption 21.10 is identical to Assumption 20.14, and we have already verified in §20.4.1 that Assumption 20.12 and Assumption 20.14 are met by the Raviart–Thomas elements.

21.3 Practical construction

In this section, we investigate systematic ways to construct the maps χ_{lr} and j_dof. The construction of χ_{lr} is done in such a way that Item (ii) of Assumption 21.10 holds true. As before, the reference cell $\widehat{K}$ can be either a simplex or a cuboid in $\mathbb{R}^d$, $d \in \{2, 3\}$.

21.3.1 Enumeration of the geometric entities in $\widehat{K}$

The construction of χ_{lr} is greatly simplified by adopting reasonable enumeration conventions on the reference cell $\widehat{K}$ and by using the orientation of the mesh. We start by enumerating the geometric entities in $\widehat{K}$. We first enumerate the n_{cv} vertices, say from 1 to n_{cv}, as in Table 10.1 in §10.2. We start with the origin of $\widehat{K}$, say $\widehat{z}_1 := \mathbf{0}$, then we enumerate d vertices in such a way that the orientation of the basis $(\widehat{z}_2 - \widehat{z}_1, \ldots, \widehat{z}_{d+1} - \widehat{z}_1)$ is the same as that of the ambient space $\mathbb{R}^d$ (assumed to be based on the right-hand rule). There is no other vertex to enumerate if $\widehat{K}$ is the unit simplex. If $\widehat{K}$ is the unit square, the last vertex is assigned number 4, and if $\widehat{K}$ is the unit cube, the last vertex of the face containing $\{\widehat{z}_1, \widehat{z}_2, \widehat{z}_3\}$ is assigned number 5, then we set $\widehat{z}_6 := \widehat{z}_2 + \boldsymbol{e}_z$, $\widehat{z}_7 := \widehat{z}_3 + \boldsymbol{e}_z$, and $\widehat{z}_8 := \widehat{z}_5 + \boldsymbol{e}_z$; see Figure 21.2 and Figure 10.2.

We now enumerate the edges of $\widehat{K}$ from 1 to n_{ce} and the faces of $\widehat{K}$ from 1 to n_{cf}. The way the enumeration is done does not really matter for our purpose, but to be complete, we now propose one possible enumeration
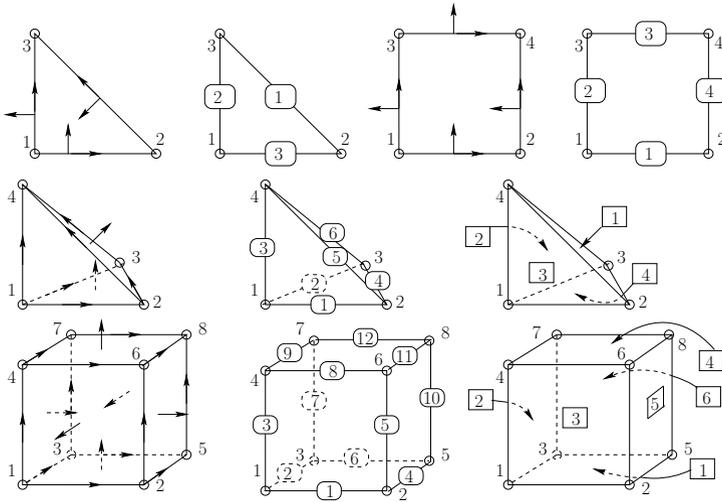
Fig. 21.2 Orientation of the edges and faces and enumeration of the vertices, edges, and faces of the reference cell in dimensions two and three. In dimension two, edges and faces coincide as geometric entities but they are oriented differently: an edge is oriented by a tangent vector and a face by a normal vector.

technique in Figure 21.2 and Table 21.1. The convention adopted in Table 21.1 is that $\widehat{E} = (\widehat{\boldsymbol{z}}_p, \widehat{\boldsymbol{z}}_q)$, $p < q$, means that $\widehat{E}$ passes through the two vertices $\widehat{\boldsymbol{z}}_p$, $\widehat{\boldsymbol{z}}_q$, and the edge is oriented by setting $\widehat{\boldsymbol{\tau}}_E := (\widehat{\boldsymbol{z}}_q - \widehat{\boldsymbol{z}}_p)/\|\widehat{\boldsymbol{z}}_q - \widehat{\boldsymbol{z}}_p\|_{\ell^2}$. The point $\widehat{\boldsymbol{z}}_p$ is called origin of the oriented edge $\widehat{E}$. The notation $\widehat{F} = (\widehat{\boldsymbol{z}}_p, \widehat{\boldsymbol{z}}_q, \widehat{\boldsymbol{z}}_r)$, $p < q < r$, means that $\widehat{F}$ passes though the three vertices $\widehat{\boldsymbol{z}}_p$, $\widehat{\boldsymbol{z}}_q$, $\widehat{\boldsymbol{z}}_r$, and that the unit normal $\widehat{\boldsymbol{n}}_{\widehat{F}}$ orienting $\widehat{F}$ is such that $(\widehat{\boldsymbol{z}}_q - \widehat{\boldsymbol{z}}_p, \widehat{\boldsymbol{z}}_r - \widehat{\boldsymbol{z}}_p, \widehat{\boldsymbol{n}}_{\widehat{F}})$ is a right-hand basis, i.e., $\widehat{\boldsymbol{n}}_{\widehat{F}} = ((\widehat{\boldsymbol{z}}_q - \widehat{\boldsymbol{z}}_p) \times (\widehat{\boldsymbol{z}}_r - \widehat{\boldsymbol{z}}_p))/\|(\widehat{\boldsymbol{z}}_q - \widehat{\boldsymbol{z}}_p) \times (\widehat{\boldsymbol{z}}_r - \widehat{\boldsymbol{z}}_p)\|_{\ell^2}$ (see (10.9)). The vertex $\widehat{\boldsymbol{z}}_p$ is called origin of the oriented face $\widehat{F}$. Note that for both the reference simplex and the reference cuboid, the orientation of the geometric entities is done by using the increasing vertex-index enumeration technique explained in §10.4.

Let now K be a cell in a mesh $\mathcal{T}_h$. Let $\boldsymbol{z}$, E, F be a vertex, an edge, and a face of K, respectively. We are going to say in the rest of this section that the local index of $\boldsymbol{z}$, E, F in K is, respectively, p, q, r if there is a vertex $\widehat{\boldsymbol{z}}_p$, $p \in \{1{:}n_{\mathrm{cv}}\}$, an edge $\widehat{E}_q$, $q \in \{1{:}n_{\mathrm{ce}}\}$, and a face $\widehat{F}_r$, $r \in \{1{:}n_{\mathrm{cf}}\}$, such that $\boldsymbol{z} = \boldsymbol{T}_K(\widehat{\boldsymbol{z}}_p)$, $E = \boldsymbol{T}_K(\widehat{E}_q)$, and $F = \boldsymbol{T}_K(\widehat{F}_r)$.

21.3.2 Example of a construction of χ_{lr} and j_dof

We now present an example of practical construction of the maps χ_{lr} and j_dof. One important advantage of the proposed enumeration is that it can be implemented in parallel since for each cell K of index $m \in \{1{:}N_{\mathrm{c}}\}$, the

2D simplex	V	$\widehat{z}_1 = (0,0),\ \widehat{z}_2 = (1,0),\ \widehat{z}_3 = (0,1)$
	E	$\widehat{E}_1 = (\widehat{z}_2, \widehat{z}_3),\ \widehat{E}_2 = (\widehat{z}_1, \widehat{z}_3),\ \widehat{E}_2 = (\widehat{z}_1, \widehat{z}_2)$
3D simplex	V	$\widehat{z}_1 = (0,0,0),\ \widehat{z}_2 = (1,0,0),\ \widehat{z}_3 = (0,1,0),\ \widehat{z}_4 = (0,0,1)$
	E	$\widehat{E}_1 = (\widehat{z}_1, \widehat{z}_2),\ \widehat{E}_2 = (\widehat{z}_1, \widehat{z}_3),\ \widehat{E}_3 = (\widehat{z}_1, \widehat{z}_4)$
		$\widehat{E}_4 = (\widehat{z}_2, \widehat{z}_3),\ \widehat{E}_5 = (\widehat{z}_2, \widehat{z}_4),\ \widehat{E}_6 = (\widehat{z}_3, \widehat{z}_4)$
	F	$\widehat{F}_1 = (\widehat{z}_2, \widehat{z}_3, \widehat{z}_4),\ \widehat{F}_2 = (\widehat{z}_1, \widehat{z}_3, \widehat{z}_4)$
		$\widehat{F}_3 = (\widehat{z}_1, \widehat{z}_2, \widehat{z}_4),\ \widehat{F}_4 = (\widehat{z}_1, \widehat{z}_2, \widehat{z}_3)$
2D square	V	$\widehat{z}_1 = (0,0),\ \widehat{z}_2 = (1,0),\ \widehat{z}_3 = (0,1),\ \widehat{z}_4 = (1,1)$
	E	$\widehat{E}_1 = (\widehat{z}_1, \widehat{z}_2),\ \widehat{E}_2 = (\widehat{z}_1, \widehat{z}_3),\ \widehat{E}_3 = (\widehat{z}_3, \widehat{z}_4),\ \widehat{E}_4 = (\widehat{z}_2, \widehat{z}_4)$
3D cube	V	$\widehat{z}_1 = (0,0,0),\ \widehat{z}_2 = (1,0,0),\ \widehat{z}_3 = (0,1,0),\ \widehat{z}_4 = (0,0,1)$
		$\widehat{z}_5 = (1,1,0),\ \widehat{z}_6 = (1,0,1),\ \widehat{z}_7 = (0,1,1),\ \widehat{z}_8 = (1,1,1)$
	E	$\widehat{E}_1 = (\widehat{z}_1, \widehat{z}_2),\ \widehat{E}_2 = (\widehat{z}_1, \widehat{z}_3),\ \widehat{E}_3 = (\widehat{z}_1, \widehat{z}_4),\ \widehat{E}_4 = (\widehat{z}_2, \widehat{z}_5)$
		$\widehat{E}_5 = (\widehat{z}_2, \widehat{z}_6),\ \widehat{E}_6 = (\widehat{z}_3, \widehat{z}_5),\ \widehat{E}_7 = (\widehat{z}_3, \widehat{z}_7),\ \widehat{E}_8 = (\widehat{z}_4, \widehat{z}_6)$
		$\widehat{E}_9 = (\widehat{z}_4, \widehat{z}_7),\ \widehat{E}_{10} = (\widehat{z}_5, \widehat{z}_8),\ \widehat{E}_{11} = (\widehat{z}_6, \widehat{z}_8),\ \widehat{E}_{12} = (\widehat{z}_7, \widehat{z}_8)$
	F	$\widehat{F}_1 = (\widehat{z}_1, \widehat{z}_2, \widehat{z}_3),\ \widehat{F}_2 = (\widehat{z}_1, \widehat{z}_3, \widehat{z}_4),\ \widehat{F}_3 = (\widehat{z}_1, \widehat{z}_2, \widehat{z}_4)$
		$\widehat{F}_4 = (\widehat{z}_4, \widehat{z}_6, \widehat{z}_7),\ \widehat{F}_5 = (\widehat{z}_2, \widehat{z}_5, \widehat{z}_6),\ \widehat{F}_6 = (\widehat{z}_3, \widehat{z}_5, \widehat{z}_7)$

Table 21.1 Enumeration and orientation of the vertices, edges, and faces in simplices and cuboids in dimensions two and three.

proposed enumeration technique only requires to have access to local information like $\mathtt{j\_cv}(m, 1{:}n_{\mathrm{cv}})$, $\mathtt{j\_ce}(m, 1{:}n_{\mathrm{ce}})$, $\mathtt{j\_cf}(m, 1{:}n_{\mathrm{cf}})$, which is usually provided by mesh generators. Recall that $\mathtt{j\_cf}(m, i)$ is the global index of the i-th vertex of the m-th cell, $\mathtt{j\_ce}(m, e)$ is the global index of the e-th edge of the m-th cell, and $\mathtt{j\_cf}(m, f)$ is the global index of the f-th face of the m-th cell.

Enumeration of the vertex dofs. Let us assume that there are $n_{\mathrm{sh}}^{\mathrm{v}}$ dofs per vertex. For scalar-valued Lagrange elements or the scalar-valued canonical hybrid element, we have $n_{\mathrm{sh}}^{\mathrm{v}} := 1$. We adopt the convention $n_{\mathrm{sh}}^{\mathrm{v}} := 0$ for $\boldsymbol{H}(\mathrm{curl})$ and $\boldsymbol{H}(\mathrm{div})$ elements. Given a mesh cell K, we enumerate the local dofs in K as follows. Letting $n \in \{1{:}n_{\mathrm{sh}}^{\mathrm{v}}\}$, $v \in \{1{:}n_{\mathrm{cv}}\}$, the n-th dof attached to the v-th vertex is assigned the index $i := (v-1)n_{\mathrm{sh}}^{\mathrm{v}} + n$.

Let us now define $\mathtt{j\_dof}$ and, given an interface $F := \partial K_l \cap \partial K_r$, let us define χ_{lr}. Let $\boldsymbol{z}$ be vertex of the face F. Let $v_l, v_r \in \{1{:}n_{\mathrm{cv}}\}$ be the local index of $\boldsymbol{z}$ in K_l, K_r, respectively, and let m_l, m_r be the indices of K_l, K_r in $\mathcal{T}_h$, respectively. Hence, $\mathtt{j\_cv}(m_l, v_l) = \mathtt{j\_cv}(m_r, v_r)$. Let $i_{l0} := (v_l-1)n_{\mathrm{sh}}^{\mathrm{v}}$ and $i_{r0} := (v_r-1)n_{\mathrm{sh}}^{\mathrm{v}}$. Then upon setting $\chi_{lr}(i_{l0} + i) := i_{r0} + i$ for all $i \in \{1{:}n_{\mathrm{sh}}^{\mathrm{v}}\}$, we observe that χ_{lr} maps a vertex dof of K_l to a vertex dof of K_r and by construction the vertex associated with $i_{l0} + i$ (with index $\mathtt{j\_cv}(m_l, v_l)$) is the same as that associated with $i_{r0} + i$ (with index $\mathtt{j\_cv}(m_r, v_r)$). Finally, $\mathtt{j\_dof}$ is obtained by setting

$$i := (v - 1)n_{\text{sh}}^{\text{v}} + n, \tag{21.8a}$$

$$\texttt{j\_dof}(m, i) := (\texttt{j\_cv}(m, v) - 1)n_{\text{sh}}^{\text{v}} + n, \tag{21.8b}$$

for all $n \in \{1{:}n_{\text{sh}}^{\text{v}}\}$ and all $v \in \{1{:}n_{\text{cv}}\}$. This defines $n_{\text{sh}}^{\text{v}} N_{\text{v}}$ equivalence classes enumerated from 1 to $n_{\text{sh}}^{\text{v}} N_{\text{v}}$.
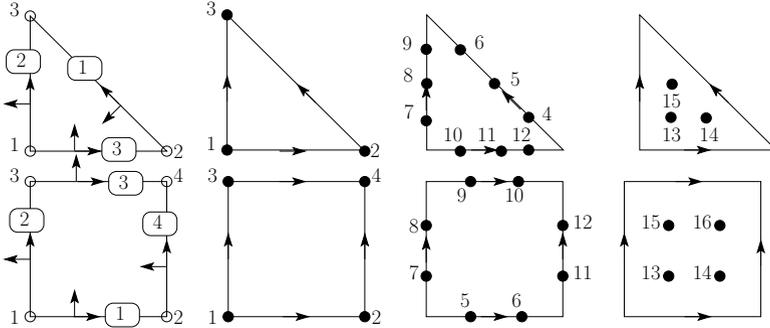
Fig. 21.3 Enumeration of geometric entities and dofs for triangles (top) and squares (bottom). Orientation of edges and faces, enumeration of vertices and faces (leftmost panels), enumeration of vertex dofs (center left panels), enumeration of edge dofs for $\mathbb{P}_{4,2}$ and $\mathbb{Q}_{3,2}$ elements (center right panels), enumeration of volume dofs (rightmost panels).

Enumeration of the edge dofs. Let n_{sh}^{e} be the number of dofs per edge. For $\mathbb{P}_{k+1,d}$ and $\mathbb{Q}_{k+1,d}$ scalar-valued elements (Lagrange or canonical hybrid) and for $\mathbb{N}_{k,d}$ Nédélec elements, we have $n_{\text{sh}}^{\text{e}} = \dim(\mathbb{P}_{k,1})$ with $k \geq 0$. Let us now adopt a strategy to enumerate the edge dofs in K that allows us to generate χ_{lr} with information associated with the edges only. Let $E := (\boldsymbol{z}_p, \boldsymbol{z}_q)$ be an oriented edge of K with origin $\boldsymbol{z}_p$, $p, q \in \{1{:}N_{\text{v}}\}$. Let $e \in \{1{:}n_{\text{ce}}\}$ be the local index of E in K. Setting $i_0 := n_{\text{cv}}n_{\text{sh}}^{\text{v}} + (e-1)n_{\text{sh}}^{\text{e}}$, we enumerate the dofs associated with E from $i_0 + 1$ to $i_0 + n_{\text{sh}}^{\text{e}}$ by moving along E from $\boldsymbol{z}_p$ to $\boldsymbol{z}_q$. Since the orientation of the mesh is generation-compatible (see Definition 10.3), the orientation of the edge is unchanged by the geometric mapping $\boldsymbol{T}_K$ for all $K \in \mathcal{T}_E$. This implies that no matter which edge $\widehat{E}$ of $\widehat{K}$ is mapped to E, the edge dofs $\{\sigma_{K,E,i}\}_{i \in \{1{:}n_{\text{ce}}\}}$ are always listed in the same order as those in $\{\widehat{\sigma}_{\widehat{K},\widehat{E},i}\}_{i \in \{1{:}n_{\text{ce}}\}}$ because the edge dofs are invariant under any vertex permutation (see Assumption 20.7 and Item (iii) in Lemma 20.22). The proposed enumeration is illustrated in the two panels in the third column of Figure 21.3 for the $\mathbb{P}_{4,2}$ and $\mathbb{Q}_{3,2}$ Lagrange elements, in the left panel of Figure 21.4 for the $\mathbb{N}_{2,3}$ Nédélec element, in Figure 21.5 for the $\mathbb{P}_{3,3}$ Lagrange element, and in Figure 21.6 for the $\mathbb{Q}_{3,3}$ Lagrange element.

Let us now define $\texttt{j\_dof}$ and, given an interface $F := \partial K_l \cap \partial K_r$, let us define χ_{lr}. Let E be an edge of the face F. Let $e_l, e_r \in \{1{:}n_{\text{ce}}\}$ be the local index of E in K_l, K_r, respectively, and let m_l, m_r be the index of K_l, K_r in $\mathcal{T}_h$,
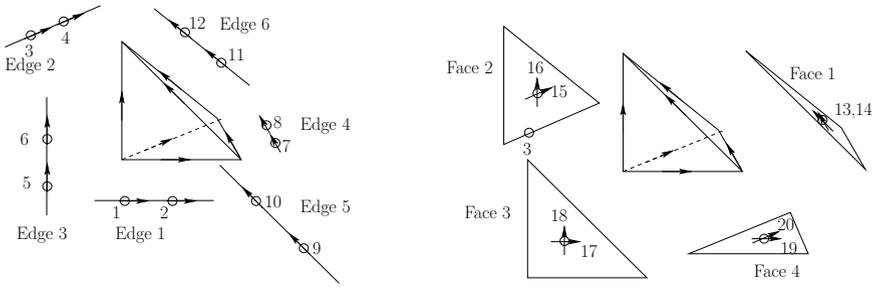
Fig. 21.4 Enumeration of dofs for the $\mathbb{N}_{2,3}$ element. Left: edge dofs. Right: face dofs.

respectively. Hence, $\mathtt{j\_ce}(m_l, e_l) = \mathtt{j\_ce}(m_r, e_r)$. Let $i_{l0} := n_{\mathrm{cv}}n_{\mathrm{sh}}^{\mathrm{v}} + (e_l - 1)n_{\mathrm{sh}}^{\mathrm{e}}$ and $i_{r0} := n_{\mathrm{cv}}n_{\mathrm{sh}}^{\mathrm{v}} + (e_r - 1)n_{\mathrm{sh}}^{\mathrm{e}}$. Then setting $\chi_{lr}(i_{l0} + i) := i_{r0} + i$ for all $i \in \{1{:}n_{\mathrm{sh}}^{\mathrm{e}}\}$, we observe that χ_{lr} maps an edge dof of K_l to an edge dof of K_r and by construction the edge associated with $i_{l0}+i$ (with index $\mathtt{j\_ce}(m_l, e_l)$) is the same as that associated with $i_{r0}+i$ (with index $\mathtt{j\_ce}(m_r, e_r)$). Concerning $\mathtt{j\_dof}$, since all the vertex dofs have already been enumerated using (21.8), we continue with the edge dofs by setting

$$i := n_{\mathrm{cv}}n_{\mathrm{sh}}^{\mathrm{v}} + (e - 1)n_{\mathrm{sh}}^{\mathrm{e}} + n, \tag{21.9a}$$

$$\mathtt{j\_dof}(m, i) := n_{\mathrm{sh}}^{\mathrm{v}}N_{\mathrm{v}} + (\mathtt{j\_ce}(m, e) - 1)n_{\mathrm{sh}}^{\mathrm{e}} + n, \tag{21.9b}$$

for all $n \in \{1{:}n_{\mathrm{sh}}^{\mathrm{e}}\}$ and all $e \in \{1{:}n_{\mathrm{ce}}\}$. This defines $n_{\mathrm{sh}}^{\mathrm{e}}N_{\mathrm{e}}$ equivalence classes enumerated from $n_{\mathrm{sh}}^{\mathrm{v}}N_{\mathrm{v}} + 1$ to $n_{\mathrm{sh}}^{\mathrm{v}}N_{\mathrm{v}} + n_{\mathrm{sh}}^{\mathrm{e}}N_{\mathrm{e}}$.

Enumeration of the face dofs. Let us proceed with the enumeration of the face dofs in dimension 3. Let F be a face of K. Let $\boldsymbol{z}_p$ be the origin of F. Let $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ be the two unit vectors orienting the edges starting from $\boldsymbol{z}_p$ (recall that $\boldsymbol{n}_F$ has been defined s.t. $(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{n}_F)$ has the same orientation as the right-hand basis in $\mathbb{R}^3$ (see (10.9)). Let $f \in \{1{:}n_{\mathrm{cf}}\}$ be the local index of F in K. The face dofs on F are enumerated from $i_0 + 1$ to $i_0 + n_{\mathrm{sh}}^{\mathrm{f}}$, where $i_0 := n_{\mathrm{cv}}n_{\mathrm{sh}}^{\mathrm{v}} + n_{\mathrm{ce}}n_{\mathrm{sh}}^{\mathrm{e}} + (f - 1)n_{\mathrm{sh}}^{\mathrm{f}}$. When the dofs in F are attached to nodes located in F, as for Lagrange elements, one possible enumeration technique is to look at F with the vector $\boldsymbol{\tau}_1$ horizontal, the origin of F on the left, $\boldsymbol{\tau}_2$ pointing upward, and $\boldsymbol{n}_F$ pointing towards us. Then one enumerates the dofs on F by moving across F from left to right and bottom to top. The proposed enumeration is illustrated in Figure 21.5 for the $\mathbb{P}_{3,3}$ Lagrange element (where there is 1 face dof) and in Figure 21.6 for the $\mathbb{Q}_{3,3}$ Lagrange element (where there are 4 face dofs). For the Nédélec and Raviart–Thomas elements, the enumeration of the face dofs can be performed by enumerating the modal basis associated with these dofs just like above. For the Nédélec elements, one has two dofs for each modal basis function, say one associated with $\boldsymbol{\tau}_1$ and one associated with $\boldsymbol{\tau}_2$. One first enumerates the dof associated with $\boldsymbol{\tau}_1$,

then the dof associated with τ_2. An example is shown in the right panel of Figure 21.4.
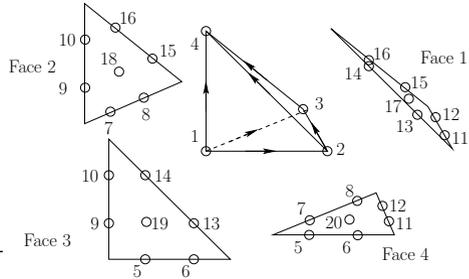
Fig. 21.5 Enumeration of dofs in dimension three for the $\mathbb{P}_{3,3}$ element.
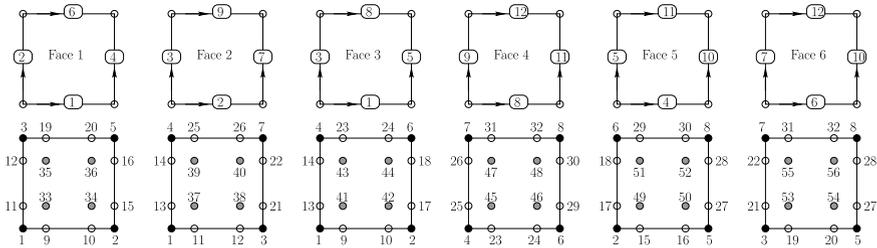
Fig. 21.6 Enumeration of $\mathbb{Q}_{3,3}$ dofs in a cube. The enumeration of the edges and faces is shown in the top panels. The enumeration of the dofs is shown in the bottom panels for the 6 faces of the cube. The vertex dofs are shown in black, the edge dofs are shown in white, and the face dofs are shown in gray. The remaining 8 volume dofs are hidden.

Assume now that $F := \partial K_l \cap \partial K_r$. Let $f_l, f_r \in \{1{:}n_{\mathrm{cf}}\}$ be the index of F in K_l, K_r, and let m_l, m_r be the indices of K_l, K_r in $\mathcal{T}_h$, i.e., $\mathsf{j\_cf}(m_l, f_l) = \mathsf{j\_cf}(m_r, f_r)$. Let $i_{l0} := n_{\mathrm{cv}}n_{\mathrm{sh}}^{\mathrm{v}} + n_{\mathrm{ce}}n_{\mathrm{sh}}^{\mathrm{e}} + (f_l - 1)n_{\mathrm{sh}}^{\mathrm{f}}$ and $i_{r0} := n_{\mathrm{cv}}n_{\mathrm{sh}}^{\mathrm{v}} + n_{\mathrm{ce}}n_{\mathrm{sh}}^{\mathrm{e}} + (f_r - 1)n_{\mathrm{sh}}^{\mathrm{f}}$. Then we set $\chi_{lr}(i_{l0} + i) := i_{r0} + i$ for all $i \in \{1{:}n_{\mathrm{sh}}^{\mathrm{f}}\}$. Concerning $\mathsf{j\_dof}$, since all the vertex and edge dofs have already been enumerated using (21.8) and (21.9), we continue with the face dofs by setting

$$i := n_{\mathrm{cv}}n_{\mathrm{sh}}^{\mathrm{v}} + n_{\mathrm{ce}}n_{\mathrm{sh}}^{\mathrm{e}} + (f - 1)n_{\mathrm{sh}}^{\mathrm{f}} + n, \tag{21.10a}$$

$$\mathsf{j\_dof}(m, i) := n_{\mathrm{sh}}^{\mathrm{v}}N_{\mathrm{v}} + n_{\mathrm{sh}}^{\mathrm{e}}N_{\mathrm{e}} + (\mathsf{j\_cf}(m, f) - 1)n_{\mathrm{sh}}^{\mathrm{f}} + n, \tag{21.10b}$$

for all $n \in \{1{:}n_{\mathrm{sh}}^{\mathrm{f}}\}$ and all $f \in \{1{:}n_{\mathrm{cf}}\}$. This defines $n_{\mathrm{sh}}^{\mathrm{f}}N_{\mathrm{f}}$ equivalence classes enumerated from $n_{\mathrm{sh}}^{\mathrm{v}}N_{\mathrm{v}} + n_{\mathrm{sh}}^{\mathrm{e}}N_{\mathrm{e}} + 1$ to $n_{\mathrm{sh}}^{\mathrm{v}}N_{\mathrm{v}} + n_{\mathrm{sh}}^{\mathrm{e}}N_{\mathrm{e}} + n_{\mathrm{sh}}^{\mathrm{f}}N_{\mathrm{f}}$. An example using the proposed enumeration for the $\mathbb{P}_{3,3}$ element is shown in Figure 21.5. An example of enumeration for the $\mathbb{Q}_{3,3}$ element is shown in Figure 21.6.

Enumeration of the volume dofs. The way the enumeration of the volume dofs is done does not matter, but to be consistent with the above definitions, one can proceed as follows. For Lagrange elements, one starts with the dof that is the closest to the origin of K and traverse the volume dofs by using the orientation of K. In dimension two, for instance, one can proceed as above since K can be viewed as a two-dimensional face, as illustrated in the rightmost panels in Figure 21.3 for the $\mathbb{P}_{4,2}$ and $\mathbb{Q}_{3,2}$ Lagrange elements. In dimension three, one can traverse all the volume dofs by moving first along the direction $\boldsymbol{\tau}_1$, then along the direction $\boldsymbol{\tau}_2$, and finally along the direction $\boldsymbol{\tau}_3$. For Nédélec and Raviart–Thomas elements, one uses the enumeration of the modal basis functions defining the volume dofs. For these elements one has 3 dofs for each modal basis function (in dimension 3), say one associated with each direction $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{\tau}_3$. For each modal function one first enumerates the dof associated with $\boldsymbol{\tau}_1$, then the dof associated with $\boldsymbol{\tau}_2$, and one finishes with the dof associated with $\boldsymbol{\tau}_3$, then one moves to the next modal function. The connectivity array can now be completed by setting

$$i := n_{\mathrm{cv}} n_{\mathrm{sh}}^{\mathrm{v}} + n_{\mathrm{ce}} n_{\mathrm{sh}}^{\mathrm{e}} + n_{\mathrm{cf}} n_{\mathrm{sh}}^{\mathrm{f}} + n, \tag{21.11a}$$

$$\texttt{j\_dof}(m, i) := n_{\mathrm{sh}}^{\mathrm{v}} N_{\mathrm{v}} + n_{\mathrm{sh}}^{\mathrm{e}} N_{\mathrm{e}} + n_{\mathrm{sh}}^{\mathrm{f}} N_{\mathrm{f}} + (m - 1) n_{\mathrm{sh}}^{\mathrm{v}} + n, \tag{21.11b}$$

for all $n \in \{1 : n_{\mathrm{sh}}^{\mathrm{c}}\}$ and all $m \in \{1 : N_{\mathrm{c}}\}$.

Exercises

Exercise 21.1 (Mesh orientation, $\mathcal{N}_{K,F}$, χ_{lr}). Consider the mesh $\mathcal{T}_h$ shown in Exercise 19.1. (i) Orient the mesh by using the increasing vertex-index enumeration technique. (ii) Consider the corresponding space $P_2^{\mathrm{g}}(\mathcal{T}_h)$. Use the enumeration convention adopted in this chapter for the dofs. Find the two cells K_l, K_r for the second face of the cell 5 and for the first face of the cell 3. (iii) Let F be the second face of the cell 5. Identify $\mathcal{N}_{5,F}$, $\texttt{j\_dof}(5, \mathcal{N}_{5,F})$, and the map χ_{lr}. (iv) Let F' be the first face of the cell 3. Identify $\mathcal{N}_{3,F'}$, $\texttt{j\_dof}(3, \mathcal{N}_{3,F'})$, and the map χ_{lr}.

Exercise 21.2 (M-dofs). Let $K \in \mathcal{T}_h$, let $F \in \mathcal{F}_K$, and let $M \in \mathcal{M}_h$ be a geometric entity s.t. $M \subset F$. Prove that $\mathcal{N}_{K,M} \subset \mathcal{N}_{K,F}$.

Exercise 21.3 ($\mathbb{Q}_{k,3}$ dofs). Determine $n_{\mathrm{sh}}^{\mathrm{v}}, n_{\mathrm{sh}}^{\mathrm{e}}, n_{\mathrm{sh}}^{\mathrm{f}}, n_{\mathrm{sh}}^{\mathrm{c}}$ for scalar-valued $\mathbb{Q}_{k,3}$ Lagrange elements.

Part IV, Chapter 22

Quasi-interpolation and best approximation

Let v be any function in $W^{r,p}(D;\mathbb{R}^q)$ with real numbers $r > 0$ and $p \in [1,\infty]$. One of the objectives of this chapter is to estimate the decay rate of the best-approximation errors

$$\inf_{v_h \in P_k^{\mathrm{x}}(\mathcal{T}_h;\mathbb{R}^q)} |v - v_h|_{W^{m,p}(D;\mathbb{R}^q)}, \quad \forall m \in \{0{:}\lfloor r \rfloor\}, \tag{22.1}$$

where $P_k^{\mathrm{x}}(\mathcal{T}_h;\mathbb{R}^q)$ is one of the conforming finite element spaces built in the previous chapters with $\mathrm{x} \in \{\mathrm{g},\mathrm{c},\mathrm{d}\}$ and $\lfloor r \rfloor$ denotes the largest integer $n \in \mathbb{N}$ s.t. $n \le r$. Recall that $k \ge 1$ if $\mathrm{x} = \mathrm{g}$ and $k \ge 0$ otherwise. Whenever the context is unambiguous, we drop the superscript x. The interpolation operators constructed so far do not give a satisfactory answer to the above question when the function v has a low smoothness index r. In this chapter, we introduce the important notion of quasi-interpolation, i.e., we build linear operators

$$\mathcal{I}_h^{\mathrm{av}} : L^1(D;\mathbb{R}^q) \to P_k(\mathcal{T}_h;\mathbb{R}^q) \tag{22.2}$$

that are L^1-stable, are projections onto $P_k(\mathcal{T}_h;\mathbb{R}^q)$, and have optimal local approximation properties. We do this by composing one of the L^1-stable operators $L^1(D;\mathbb{R}^q) \to P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$ introduced in §18.3 with a simple averaging operator $\mathcal{J}_h^{\mathrm{av}} : P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q) \to P_k(\mathcal{T}_h;\mathbb{R}^q)$. We also adapt the construction to enforce zero γ-traces at the boundary, and we study the approximation properties of the L^2-orthogonal projection onto the conforming space $P_k(\mathcal{T}_h;\mathbb{R}^q)$. The material of this chapter is useful to investigate the approximation of solutions to PDEs with low regularity.

22.1 Discrete setting

Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular family of affine meshes s.t. the geometric mappings $\boldsymbol{T}_K$ are affine for all $K \in \mathcal{T}_h$. As before, when dealing with the

conforming spaces $P_k(\mathcal{T}_h; \mathbb{R}^q)$, we assume that the meshes are matching. All the functions v_h in the broken finite element space $P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$ are such that $v_{h|K} \in P_K := \psi_K^{-1}(\widehat{P})$ for all $K \in \mathcal{T}_h$, where $(\widehat{K}, \widehat{P}, \widehat{\Sigma})$ is the reference finite element. Let $k \geq 0$ be the degree of the reference finite element, i.e., $[\mathbb{P}_{k,d}]^q \subset \widehat{P} \subset W^{k+1,p}(\widehat{K}; \mathbb{R}^q)$ for some $p \in [1, \infty]$; see Definition 11.14. We assume that the transformations ψ_K are of the form (11.1), i.e., $\psi_K(v) := \mathbb{A}_K(v \circ \boldsymbol{T}_K)$ for some matrix $\mathbb{A}_K \in \mathbb{R}^{q \times q}$ and there are c, c' s.t. for all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,

$$\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2} \leq c \, \|\mathbb{J}_K\|_{\ell^2}\|\mathbb{J}_K^{-1}\|_{\ell^2} \leq c', \qquad (22.3)$$

where $\mathbb{J}_K$ is the Jacobian matrix of $\boldsymbol{T}_K$ (the second bound follows from the regularity of the mesh sequence). The main examples are $\mathbb{A}_K^{\mathrm{g}} := 1$ ($q = 1$) for Lagrange and canonical hybrid elements, $\mathbb{A}_K^{\mathrm{c}} := \mathbb{J}_K^{\mathsf{T}}$ ($q = d = 3$) for Nédélec elements, and $\mathbb{A}_K^{\mathrm{d}} := \det(\mathbb{J}_K)\mathbb{J}_K^{-1}$ ($q = d$) for Raviart–Thomas elements. We additionally assume that $\widehat{P} \subset W^{1,\infty}(\widehat{K}; \mathbb{R}^q)$, so that $P_K \subset W^{1,\infty}(K; \mathbb{R}^q)$ for all $K \in \mathcal{T}_h$ owing to Lemma 11.7.

Let $\mathtt{j\_dof} : \mathcal{T}_h \times \mathcal{N} \to \mathcal{A}_h$ be the connectivity array introduced in Chapter 19 and let $a \in \mathcal{A}_h$ be a connectivity class. Recall that we write either $\mathtt{j\_dof}(K, i) = a$ or $(K, i) \in a$. The connectivity classes are used to enumerate the global shape functions and dofs in the conforming space $P_k(\mathcal{T}_h; \mathbb{R}^q)$. The subset $\mathcal{T}_a := \{K' \in \mathcal{T}_h \mid \exists i' \in \mathcal{N}, (K', i') \in a\}$ is the collection of the mesh cells of which a is a dof (see (19.6)). For all $K \in \mathcal{T}_h$, we introduce the notation

$$\check{\mathcal{T}}_K := \bigcup_{i \in \mathcal{N}} \mathcal{T}_{\mathtt{j\_dof}(K,i)}, \qquad D_K := \mathrm{int}\left(\bigcup_{K' \in \check{\mathcal{T}}_K} K'\right), \qquad (22.4)$$

i.e., $\check{\mathcal{T}}_K$ is the collection of the cells sharing some dof(s) with K, and D_K is the set of the points composing the cells from the set $\check{\mathcal{T}}_K$. For instance, if the finite element has vertex dofs, any mesh cell K' touching K is in $\check{\mathcal{T}}_K$ since in this case K' shares at least a vertex with K. Notice that $\mathrm{card}(\check{\mathcal{T}}_K)$ is uniformly bounded owing to the regularity of the mesh sequence. We have

$$\check{\mathcal{T}}_K = \{K' \in \mathcal{T}_h \mid \exists i' \in \mathcal{N}, \mathcal{M}_{K',i'} \subset K\}, \qquad (22.5)$$

where $\mathcal{M}_{K',i'}$ is the geometric entity associated with a pair (K', i') (see (21.5)).

Let $K \in \mathcal{T}_h$ and let $F \in \mathcal{F}_K$ be a face of K. We consider the trace operator $\gamma_{K,F} : W^{s,p}(K; \mathbb{R}^q) \to L^1(F; \mathbb{R}^t)$ defined in (18.7), with $sp > 1$ if $p > 1$ or $s \geq 1$ if $p = 1$, i.e., we have $\gamma_{K,F}^{\mathrm{g}}(v) := v_{|F}$ ($q = t = 1$), $\gamma_{K,F}^{\mathrm{c}}(\boldsymbol{v}) := \boldsymbol{v}_{|F} \times \boldsymbol{n}_F$ ($q = t = d = 3$), and $\gamma_{K,F}^{\mathrm{d}}(\boldsymbol{v}) := \boldsymbol{v}_{|F} \cdot \boldsymbol{n}_F$ ($q = d, t = 1$). Using the notation of §20.3, there is a linear map $\sigma_{K,F,i} : P_F := \gamma_{K,F}(P_K) \to \mathbb{R}$ and a subset $\mathcal{N}_{K,F} \subset \mathcal{N}$ s.t. $\sigma_{K,i} = \sigma_{K,F,i} \circ \gamma_{K,F}$ for all $i \in \mathcal{N}_{K,F}$. (Since $P_K \subset W^{1,\infty}(K; \mathbb{R}^q)$, the γ-trace is well defined on P_K and $P_F \subset L^\infty(F; \mathbb{R}^t)$.)

It turns out that the analysis of the quasi-interpolation operators devised in this chapter is based on just one unified assumption on the face dofs of the finite element.

Assumption 22.1 (Boundedness of face dofs). *There is c s.t.*

$$\max_{i \in \mathcal{N}_{K,F}} |\sigma_{K,F,i}(q)| \le c \, \|\mathbb{A}_K\|_{\ell^2} \|q\|_{L^\infty(F;\mathbb{R}^t)}, \tag{22.6}$$

for all $q \in P_F$, all $K \in \mathcal{T}_h$, all $F \in \mathcal{F}_K$, and all $h \in \mathcal{H}$.

Assumption 22.1 is supposed to hold true in the entire chapter. One can verify that this is indeed the case for all the finite elements considered in this book, i.e., Lagrange, canonical hybrid, Nédélec, and Raviart–Thomas elements.

Let us now derive an important consequence of Assumption 22.1 allowing us to control the jump of the dofs across the mesh interfaces. Recall that for all $F \in \mathcal{F}_h^\circ$, there are two distinct mesh cells $K_l, K_r \in \mathcal{T}_h$ s.t. $F := \partial K_l \cap \partial K_r$. Let us define for all $a \in \mathcal{A}_h$ the following collection of interfaces:

$$\mathcal{F}_a^\circ := \{F \in \mathcal{F}_h^\circ \mid \exists i_l, i_r \in \mathcal{N}, \; (K_l, i_l), (K_r, i_r) \in a\}. \tag{22.7}$$

Notice that the set $\mathcal{F}_a^\circ$ is empty if $\mathrm{card}(a) = 1$, i.e., if a is the class of a cell dof. The characterization of classes of face, edge, and vertex dofs shows that the set $\mathcal{F}_a^\circ$ is always nonempty if $\mathrm{card}(a) \ge 2$.

Lemma 22.2 (Bound on dofs jump). *Let $a \in \mathcal{A}_h$ be such that $\mathcal{F}_a^\circ$ is nonempty. There is c s.t.*

$$\left| \sigma_{K_l, i_l}(v_{h|K_l}) - \sigma_{K_r, i_r}(v_{h|K_r}) \right| \le c \min(\|\mathbb{A}_{K_l}\|_{\ell^2}, \|\mathbb{A}_{K_r}\|_{\ell^2}) \|[\![v_h]\!]_F\|_{L^\infty(F;\mathbb{R}^t)}, \tag{22.8}$$

for all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$, all $F \in \mathcal{F}_a^\circ$, all $(K_l, i_l), (K_r, i_r) \in a$, and all $h \in \mathcal{H}$.

Proof. Let $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$. Since we have $i_r = \chi_{lr}(i_l)$ and $\sigma_{K_l,F,i_l} = \sigma_{K_r,F,i_r}$ owing to Assumption 20.14, letting $\delta\sigma_{lr}(v_h) := \sigma_{K_l,i_l}(v_{h|K_l}) - \sigma_{K_r,i_r}(v_{h|K_r})$, we infer that

$$\delta\sigma_{lr}(v_h) = \sigma_{K_l,F,i_l}\big(\gamma_{K_l,F}(v_{h|K_l}) - \gamma_{K_r,F}(v_{h|K_r})\big) = \sigma_{K_l,F,i_l}([\![v_h]\!]_F),$$

recalling that $[\![v_h]\!]_F := \gamma_{K_l,F}(v_{h|K_l}) - \gamma_{K_r,F}(v_{h|K_r})$. Since $P_F = \gamma_{K_l,F}(P_{K_l}) = \gamma_{K_r,F}(P_{K_r})$ owing to Assumption 20.14, we have $[\![v_h]\!]_F \in P_F$. Owing to Assumption 22.1, we infer that $|\delta\sigma_{lr}(v_h)| \le c\|\mathbb{A}_{K_l}\|_{\ell^2}\|[\![v_h]\!]_F\|_{L^\infty(F;\mathbb{R}^t)}$. Finally, (22.8) follows by exchanging the roles of K_l, K_r. $\qquad\square$

22.2 Averaging operator

We define the *averaging operator* $\mathcal{J}_h^{\mathrm{av}} : P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q) \to P_k(\mathcal{T}_h; \mathbb{R}^q)$ by setting

$$\mathcal{J}_h^{\mathrm{av}}(v_h) := \sum_{a \in \mathcal{A}_h} \left(\frac{1}{\mathrm{card}(a)} \sum_{(K,i) \in a} \sigma_{K,i}(v_{h|K}) \right) \varphi_a. \tag{22.9}$$

Since $\sigma_{K,i}(v_{h|K}) = \sigma_a(v_h)$ for all $(K,i) \in a$ and all $v_h \in P_k(\mathcal{T}_h; \mathbb{R}^q)$, we have $\mathcal{J}_h^{\mathrm{av}}(v_h) = \sum_{a \in \mathcal{A}_h} \sigma_a(v_h) \varphi_a = v_h$. Therefore, $P_k(\mathcal{T}_h; \mathbb{R}^q)$ is pointwise invariant under the action of $\mathcal{J}_h^{\mathrm{av}}$. Let now $K \in \mathcal{T}_h$ and let us set

$$\check{\mathcal{F}}_K^\circ := \bigcup_{i \in \mathcal{N}} \mathcal{F}_{\mathtt{j\_dof}(K,i)}^\circ = \{ F \in \mathcal{F}_h^\circ \mid \exists i \in \mathcal{N}, \, \mathcal{M}_{K,i} \subset F \}, \qquad (22.10)$$

i.e., $\check{\mathcal{F}}_K^\circ$ is the collection of all the mesh interfaces sharing some dof(s) with K. For instance, if the finite element has vertex dofs, any interface F touching K is in $\check{\mathcal{F}}_K^\circ$ since in this case $F \cap K$ contains at least a vertex of K. Note that $\mathrm{card}(\check{\mathcal{F}}_K^\circ)$ is uniformly bounded owing to the regularity of the mesh sequence.

Lemma 22.3 (Approximation by averaging). *There exists c s.t. the following holds true:*

$$|v_h - \mathcal{J}_h^{\mathrm{av}}(v_h)|_{W^{m,p}(K;\mathbb{R}^q)} \leq c \, h_K^{d\left(\frac{1}{p} - \frac{1}{r}\right) + \frac{1}{r} - m} \sum_{F \in \check{\mathcal{F}}_K^\circ} \|[\![v_h]\!]_F\|_{L^r(F;\mathbb{R}^t)}, \;\; (22.11)$$

for every integer $m \in \{0{:}k+1\}$, all $p, r \in [1, \infty]$, all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Proof. We only prove the bound for $m = 0$ and $p = r = \infty$, the other cases follow by invoking the inverse inequalities from Lemma 12.1 (with d replaced by $(d-1)$ when working on F). Let $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$, set $e_h := v_h - \mathcal{J}_h^{\mathrm{av}}(v_h)$ and observe that $e_h \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$. Let $K \in \mathcal{T}_h$. Using Proposition 12.5 with $p = \infty$, we infer that

$$\|e_h\|_{L^\infty(K;\mathbb{R}^q)} \leq c \|\mathbb{A}_K^{-1}\|_{\ell^2} \sum_{i \in \mathcal{N}} |\sigma_{K,i}(e_{h|K})|.$$

Owing to the definition (22.9) of $\mathcal{J}_h^{\mathrm{av}}$, we have for all $i \in \mathcal{N}$,

$$\sigma_{K,i}(e_{h|K}) = \frac{1}{\mathrm{card}(a)} \sum_{(K',i') \in a} \left(\sigma_{K,i}(v_{h|K}) - \sigma_{K',i'}(v_{h|K'}) \right),$$

where $a := \mathtt{j\_dof}(K,i)$. Notice that $\sigma_{K,i}(e_{h|K}) = 0$ if $\mathrm{card}(a) = 1$. Let us now assume that $\mathrm{card}(a) \geq 2$. For all $(K',i') \in a$, there is an M-path of mesh cells in $\mathcal{T}_a$ connecting K with K' s.t. any two consecutive mesh cells in the path share a common face $F \in \mathcal{F}_a^\circ$ (see Lemma 21.4). It is possible to assume that each face crossed by the M-path is crossed only once. Since for every consecutive pairs $(K_l, i_l), (K_r, i_r)$ in the M-path, we have $\partial K_l \cap \partial K_r =: F \in \mathcal{F}_a^\circ$, Lemma 22.2 implies that

$$|\sigma_{K_l,i_l}(v_{h|K_l}) - \sigma_{K_r,i_r}(v_{h|K_r})| \leq c \, \min(\|\mathbb{A}_{K_l}\|_{\ell^2}, \|\mathbb{A}_{K_r}\|_{\ell^2}) \|[\![v_h]\!]_F\|_{L^\infty(F;\mathbb{R}^t)}.$$

Since $\mathrm{card}(\mathcal{T}_a)$ is uniformly bounded, we obtain (recall that $a = \mathtt{j\_dof}(K,i)$)

$$\|e_h\|_{L^\infty(K;\mathbb{R}^q)} \le c\,\|\mathbb{A}_K^{-1}\|_{\ell^2}\|\mathbb{A}_K\|_{\ell^2} \sum_{i\in\mathcal{N}}\sum_{F\in\mathcal{F}_a^\circ} \|[\![v_h]\!]_F\|_{L^\infty(F;\mathbb{R}^t)}.$$

The estimate (22.11) follows from (22.3) and the definition of $\check{\mathcal{F}}_K^\circ$. □

Corollary 22.4 (L^p-stability). *There is c s.t. for all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$\|\mathcal{J}_h^{\mathrm{av}}(v_h)\|_{L^p(K;\mathbb{R}^q)} \le c\,\|v_h\|_{L^p(D_K;\mathbb{R}^q)}. \tag{22.12}$$

Proof. Use the inverse inequality $\|[\![v_h]\!]\|_{L^r(F;\mathbb{R}^t)} \le c\sum_{K\in\mathcal{T}_F} h_K^{-\frac{1}{r}}\|v_h\|_{L^r(K;\mathbb{R}^q)}$ where $\mathcal{T}_F := \{K_l, K_r\}$ for all $F := \partial K_l \cap \partial K_r \in \mathcal{F}_h^\circ$, and the bound (22.11). See Exercise 22.2 for a direct proof. □

Remark 22.5 (Literature). Early work on averaging operators in the scalar-valued case (H^1-setting) includes Oswald [153, Eq. (25)–(26)] and Brenner [44, p. 13]. These operators were used to analyze nonconforming finite elements in Brenner [45], Hoppe and Wohlmuth [119], discontinuous Galerkin (dG) methods in Brenner [46], Karakashian and Pascal [122], Ern and Vohralík [98], Gudi [111], Schöberl and Lehrenfeld [175], stabilized finite elements in Burman and Ern [52, 53], and multiscale methods in Kornhuber et al. [127]. The dependence of the constant in (22.11) on the polynomial degree k is studied in Burman and Ern [52], Houston et al. [120]. In the vector-valued case, averaging operators were considered for dG methods in Cockburn et al. [81] ($\boldsymbol{H}$(div)-setting) and Campos Pinto and Sonnendrücker [56] ($\boldsymbol{H}$(curl)-setting). □

22.3 Quasi-interpolation operator

We consider one of the L^1-stable operators $L^1(D;\mathbb{R}^q) \to P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$ introduced in §18.3. To fix the ideas, we work with $\mathcal{I}_h^\sharp$, but the L^2-orthogonal projection could be considered as well (these two operators coincide in the scalar-valued case). Letting $\mathcal{J}_h^{\mathrm{av}} : P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q) \to P_k(\mathcal{T}_h;\mathbb{R}^q)$ be the averaging operator defined in §22.2, we now define a global *quasi-interpolation operator* $\mathcal{I}_h^{\mathrm{av}} : L^1(D;\mathbb{R}^q) \to P_k(\mathcal{T}_h;\mathbb{R}^q)$ by setting

$$\mathcal{I}_h^{\mathrm{av}} := \mathcal{J}_h^{\mathrm{av}} \circ \mathcal{I}_h^\sharp. \tag{22.13}$$

By construction, $P_k(\mathcal{T}_h;\mathbb{R}^q)$ is pointwise invariant under $\mathcal{I}_h^{\mathrm{av}}$, i.e., $\mathcal{I}_h^{\mathrm{av}}$ is a projection. Let now us study the local approximation properties of $\mathcal{I}_h^{\mathrm{av}}$ in integer-order and fractional-order Sobolev spaces, the latter being equipped with the Sobolev–Slobodeckij norm (based on the double integral, as defined in §2.2.2). Recall the Definition 18.1 of broken Sobolev spaces.

Theorem 22.6 (Local approximation). *Let $r \in [0, k+1]$. Let $p \in [1, \infty)$ if $r \notin \mathbb{N}$, or $p \in [1, \infty]$ if $r \in \mathbb{N}$.*
(i) Assume $rp > 1$ and $p > 1$, or $r \geq 1$ and $p = 1$. There is c s.t. for every integer $m \in \{0 : \lfloor r \rfloor\}$, all $v \in W^{r,p}(D_K; \mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,

$$|v - \mathcal{I}_h^{\mathrm{av}}(v)|_{W^{m,p}(K; \mathbb{R}^q)} \leq c \, h_K^{r-m} |v|_{W^{r,p}(\check{\mathcal{T}}_K; \mathbb{R}^q)}. \tag{22.14}$$

The constant c is uniform w.r.t. r and p as long as rp is bounded from below away from 1, but c can grow unboundedly as $rp \downarrow 1$ if $p > 1$.
(ii) Assume $rp \leq 1$ and $p > 1$, or $r < 1$ and $p = 1$. There is c, uniform w.r.t. r and p, s.t. for all $v \in W^{r,p}(D_K; \mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,

$$\|v - \mathcal{I}_h^{\mathrm{av}}(v)\|_{L^p(K; \mathbb{R}^q)} \leq c \, h_K^r |v|_{W^{r,p}(D_K; \mathbb{R}^q)}. \tag{22.15}$$

Proof. Let $K \in \mathcal{T}_h$, $h \in \mathcal{H}$, and $v \in W^{r,p}(D_K; \mathbb{R}^q)$.
(1) Assume that $rp > 1$ and $p > 1$ or that $r \geq 1$ and $p = 1$. Then v has zero γ-jumps across the interfaces (see Remark 18.4). Let us set $v_h := \mathcal{I}_h^\sharp(v) \in P_k^{\mathrm{b}}(\mathcal{T}_h; \mathbb{R}^q)$, so that $\mathcal{I}_h^{\mathrm{av}}(v) = \mathcal{J}_h^{\mathrm{av}}(v_h)$. The triangle inequality gives

$$|v - \mathcal{I}_h^{\mathrm{av}}(v)|_{W^{m,p}(K; \mathbb{R}^q)} \leq |v - v_h|_{W^{m,p}(K; \mathbb{R}^q)} + |v_h - \mathcal{J}_h^{\mathrm{av}}(v_h)|_{W^{m,p}(K; \mathbb{R}^q)}.$$

Let $\mathfrak{T}_1$ and $\mathfrak{T}_2$ be the two terms on the right-hand side. $\mathfrak{T}_1$ is estimated by using Theorem 18.14, which leads to $|\mathfrak{T}_1| \leq c h_K^{r-m} |v|_{W^{r,p}(K; \mathbb{R}^q)}$. $\mathfrak{T}_2$ is estimated by using the approximation properties of $\mathcal{J}_h^{\mathrm{av}}$ established in Lemma 22.3 as follows:

$$h_K^m |\mathfrak{T}_2| \leq c \, h_K^{\frac{1}{p}} \sum_{F \in \check{\mathcal{F}}_K^\circ} \|[\![v_h]\!]_F\|_{L^p(F; \mathbb{R}^t)} = c \, h_K^{\frac{1}{p}} \sum_{F \in \check{\mathcal{F}}_K^\circ} \|[\![v - v_h]\!]_F\|_{L^p(F; \mathbb{R}^t)}$$

$$\leq c \, h_K^{\frac{1}{p}} \sum_{K' \in \check{\mathcal{T}}_K} \sum_{F \subset \partial K' \cap \check{\mathcal{F}}_K^\circ} \|(v - v_h)_{|K'}\|_{L^p(F; \mathbb{R}^q)} \leq c \, h_K^r \sum_{K' \in \check{\mathcal{T}}_K} |v|_{W^{r,p}(K'; \mathbb{R}^q)},$$

where we used the triangle inequality to bound the jump by the values over the two adjacent mesh cells, the multiplicative trace inequality (12.17), the approximation result of Theorem 18.14, and the regularity of the mesh sequence. Combining the bounds on $\mathfrak{T}_1$ and $\mathfrak{T}_2$ gives (22.14).
(2) Assume now that $rp \leq 1$ and $p > 1$ or that $r < 1$ and $p = 1$. In both cases, we have $r < 1$. Combining the L^p-stability of $\mathcal{J}_h^{\mathrm{av}}$ (Corollary 22.4) with that of $\mathcal{I}_h^\sharp$ (see (18.26) with $m := 0$), we infer that

$$\|\mathcal{I}_h^{\mathrm{av}}(v)\|_{L^p(K; \mathbb{R}^q)} = \|\mathcal{J}_h^{\mathrm{av}}(\mathcal{I}_h^\sharp(v))\|_{L^p(K; \mathbb{R}^q)}$$
$$\leq c \, \|\mathcal{I}_h^\sharp(v)\|_{L^p(D_K; \mathbb{R}^q)} \leq c' \, \|v\|_{L^p(D_K; \mathbb{R}^q)}. \tag{22.16}$$

This proves (22.15) if $r = 0$. Let us now consider the case $r \in (0, 1)$. Since $[\mathbb{P}_{0,d}]^q$ is pointwise invariant under $\mathcal{I}_h^{\mathrm{av}}$, the L^p-stability estimate (22.16), and

the triangle inequality imply that

$$\|v - \mathcal{I}_h^{\mathrm{av}}(v)\|_{L^p(K;\mathbb{R}^q)} \leq c \inf_{q\in[\mathbb{P}_{0,d}]^q} \|v - q\|_{L^p(D_K;\mathbb{R}^q)}, \tag{22.17}$$

and we conclude by invoking the fractional Poincaré–Steklov inequality in D_K from Lemma 3.26 where we use that $\ell_{D_K} \leq c_1|D_K|^{\frac{1}{d}} \leq c_2 h_K$ owing to the regularity of the mesh sequence.   □

Remark 22.7 (Seminorm). We use the broken seminorm $|v|_{W^{r,p}(\check{\mathcal{T}}_K;\mathbb{R}^q)}$ in (22.14) and the seminorm $|v|_{W^{r,p}(D_K;\mathbb{R}^q)}$ in (22.15). It is possible to break the seminorm over $\check{\mathcal{T}}_K$ in (22.14) because $rp > 1$, but this is not possible when $rp < 1$. Indeed, there is only one constant at our disposal in (22.17) to minimize the L^p-norm of $(v - q)$ over D_K.   □

Corollary 22.8 ($W^{m,p}$-stability). *There exists c s.t. for all $p \in [1,\infty]$, every integer $m \in \{0{:}k+1\}$, all $v \in W^{m,p}(D_K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$|\mathcal{I}_h^{\mathrm{av}}(v)|_{W^{m,p}(K;\mathbb{R}^q)} \leq c\,|v|_{W^{m,p}(D_K;\mathbb{R}^q)}. \tag{22.18}$$

Proof. For $m = 0$, the stability follows from (22.16). For $m \geq 1$, the stability follows from (22.14) with $r := m$ and the triangle inequality.   □

Corollary 22.9 (Best approximation in L^p). *There exists c s.t. for all $r \in [0, k+1]$, all $p \in [1,\infty)$ if $r \notin \mathbb{N}$ or all $p \in [1,\infty]$ if $r \in \mathbb{N}$, all $v \in W^{r,p}(D;\mathbb{R}^q)$, and all $h \in \mathcal{H}$,*

$$\inf_{w_h\in P_k(\mathcal{T}_h;\mathbb{R}^q)} \|v - w_h\|_{L^p(D;\mathbb{R}^q)} \leq c\,h^r|v|_{W^{r,p}(D;\mathbb{R}^q)}. \tag{22.19}$$

Remark 22.10 (Literature). Quasi-interpolation operators have been developed in the specific case x = g in Clément [80], Scott and Zhang [178], Bernardi and Girault [21], Carstensen [59], Carstensen and Verfürth [63] by performing averages of functions on macroelements attached to vertices. The present construction, introduced in [97], is different since it projects functions onto the broken finite element space before averaging the resulting dofs. Moreover, the present construction handles in a unified way H^1-, $\boldsymbol{H}$(curl)-, and $\boldsymbol{H}$(div)-conforming finite elements.   □

Remark 22.11 (Poincaré–Steklov in D_K). One can show that there is c s.t. for all $p \in [1,\infty]$, all $v \in W^{1,p}(D_K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,

$$\|v - \underline{v}_{D_K}\|_{L^p(D_K)} \leq c\,h_K|v|_{W^{1,p}(D_K)}, \tag{22.20}$$

where $\underline{v}_{D_K}$ denotes the mean-value of v on D_K. See Veeser and Verfürth [194, §2.3], the work by the authors [97, Lem. 5.7], and Exercise 22.3. (We do not invoke the inequality (3.8) with $U := D_K$ since we want to assert uniformity of the constant c w.r.t. K, which is nontrivial if the set D_K is not convex.)   □

22.4 Quasi-interpolation with zero trace

In this section, we revisit the above construction so that the quasi-interpolation operator boundedly maps $L^1(D;\mathbb{R}^q)$ onto the subspace $P_{k,0}(\mathcal{T}_h;\mathbb{R}^q)$ composed by those functions in $P_k(\mathcal{T}_h;\mathbb{R}^q)$ with zero γ-trace at the boundary of D (see §19.4).

22.4.1 Averaging operator revisited

Recall Definition 19.11 for the internal and boundary connectivity classes $\mathcal{A}_h^\circ$ and $\mathcal{A}_h^\partial$. The operator $\mathcal{J}_{h0}^{\mathrm{x,av}} : P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q) \to P_{k,0}(\mathcal{T}_h;\mathbb{R}^q)$ is defined as

$$\mathcal{J}_{h0}^{\mathrm{av}}(v_h) := \sum_{a\in\mathcal{A}_h^\circ} \left(\frac{1}{\mathrm{card}(a)} \sum_{(K,i)\in a} \sigma_{K,i}(v_{h|K}) \right) \varphi_a, \qquad (22.21)$$

for all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$. The difference between $\mathcal{J}_h^{\mathrm{av}}(v_h)$ and $\mathcal{J}_{h0}^{\mathrm{av}}(v_h)$ is that now $\sigma_a(\mathcal{J}_{h0}^{\mathrm{av}}(v_h)) = 0$ for all $a \in \mathcal{A}_h^\partial$, i.e., all the dofs associated with boundary classes are set to zero. By construction, $P_{k,0}(\mathcal{T}_h;\mathbb{R}^q)$ is pointwise invariant under $\mathcal{J}_{h0}^{\mathrm{av}}$. The approximation properties of $\mathcal{J}_{h0}^{\mathrm{av}}$ now depend not only on γ-jumps across interfaces, but also on γ-traces at boundary faces. We abuse the notation by writing $[\![v_h]\!]_F := \gamma_{K_l,F}(v_h)$ for all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$ and all $F := \partial K_l \cap \partial D \in \mathcal{F}_h^\partial$ ($K_l \in \mathcal{T}_h$ is uniquely defined). For all $a \in \mathcal{A}_h^\partial$, we define

$$\mathcal{F}_a^\partial := \{F \in \mathcal{F}_h^\partial \mid \exists i_l \in \mathcal{N}, (K_l,i_l)\in a\}, \qquad (22.22)$$

and we set $\mathcal{F}_a^\partial := \emptyset$ if $a \in \mathcal{A}_h^\circ$. We infer from (22.6) that (compare with (22.8))

$$|\sigma_{K_l,i_l}(v_h)| \le c\, \|\mathbb{A}_{K_l}\|_{\ell^2} \|[\![v_h]\!]_F\|_{L^\infty(F;\mathbb{R}^t)}, \qquad (22.23)$$

for all $F \in \mathcal{F}_a^\partial$, all pairs $(K_l,i_l) \in a$, and all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$. For all $K \in \mathcal{T}_h$, let $\check{\mathcal{F}}_K := \{F \in \mathcal{F}_h \mid \exists i \in \mathcal{N}, \mathcal{M}_{K,i} \subset F\}$ be the collection of the mesh faces (interfaces and boundary faces) sharing some dof(s) with K.

Lemma 22.12 (Approximation by averaging). *There is c s.t.*

$$|v_h - \mathcal{J}_{h0}^{\mathrm{av}}(v_h)|_{W^{m,p}(K;\mathbb{R}^q)} \le c\, h_K^{d\left(\frac{1}{p}-\frac{1}{r}\right)+\frac{1}{r}-m} \sum_{F\in\check{\mathcal{F}}_K} \|[\![v_h]\!]_F\|_{L^r(F;\mathbb{R}^t)}, \quad (22.24)$$

for every integer $m \in \{0{:}k+1\}$, all $p,r \in [1,\infty]$, all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$

Proof. Adapt the proof of Lemma 22.3 using (22.23) and the fact that $\sigma_{K,i}(v_h - \mathcal{J}_{h0}^{\mathrm{av}}(v_h)) = \sigma_{K,i}(v_h)$ if $(K,i) \in \mathcal{A}_h^\partial$. $\qquad\square$

Corollary 22.13 (L^p-stability). *There is c s.t. for all $v_h \in P_k^{\mathrm{b}}(\mathcal{T}_h;\mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,*

$$\|\mathcal{J}_{h0}^{\mathrm{av}}(v_h)\|_{L^p(K;\mathbb{R}^q)} \le c\,\|v_h\|_{L^p(D_K;\mathbb{R}^q)}. \tag{22.25}$$

Proof. The proof is similar to that of Corollary 22.4. $\qquad\square$

22.4.2 Quasi-interpolation operator revisited

We define the operator $\mathcal{I}_{h0}^{\mathrm{av}} : L^1(D;\mathbb{R}^q) \to P_{k,0}(\mathcal{T}_h;\mathbb{R}^q)$ by setting

$$\mathcal{I}_{h0}^{\mathrm{av}} := \mathcal{J}_{h0}^{\mathrm{av}} \circ \mathcal{I}_h^\sharp. \tag{22.26}$$

By construction, $P_{k,0}(\mathcal{T}_h;\mathbb{R}^q)$ is pointwise invariant under $\mathcal{I}_{h0}^{\mathrm{av}}$, i.e., $\mathcal{I}_{h0}^{\mathrm{av}}$ is a projection.

We now study the approximation properties of $\mathcal{I}_{h0}^{\mathrm{av}}$ in $W^{r,p}(D;\mathbb{R}^q)$ with real numbers $r \in [0, k+1]$ and $p \in [1,\infty]$ ($p \in [1,\infty)$ if $r \notin \mathbb{N}$). If $rp > 1$, functions in $W^{r,p}(D;\mathbb{R}^q)$ have traces on ∂D, and it makes sense to define

$$W_{0,\gamma}^{r,p}(D;\mathbb{R}^q) := \{v \in W^{r,p}(D;\mathbb{R}^q) \mid \gamma(v) = 0\}. \tag{22.27}$$

We quote the following results from [97, Thm. 6.4 & Cor. 6.5] where we use the following notation:

$$\mathcal{T}_h^\circ := \{K \in \mathcal{T}_h \mid \forall i \in \mathcal{N},\ \mathtt{j\_dof}(K,i) \in \mathcal{A}_h^\circ\}, \tag{22.28a}$$

$$\mathcal{T}_h^\partial := \mathcal{T}_h \setminus \mathcal{T}_h^\circ = \{K \in \mathcal{T}_h \mid \exists i \in \mathcal{N},\ \mathtt{j\_dof}(K,i) \in \mathcal{A}_h^\partial\}, \tag{22.28b}$$

$$D^\partial := \mathrm{int}\left(\bigcup_{K \in \mathcal{T}_h^\partial} K\right), \tag{22.28c}$$

that is, $\mathcal{T}_h^\circ$ is the collection of the cells whose global dofs are all internal connectivity classes, $\mathcal{T}_h^\partial$ is the collection of the cells having at least one global dof that is a boundary connectivity class, and D^∂ is the interior of the set of points composing the cells in $\mathcal{T}_h^\partial$.

Theorem 22.14 (Approximation). *Let $r \in [0, k+1]$. Let $p \in [1,\infty)$ if $r \notin \mathbb{N}$, or $p \in [1,\infty]$ if $r \in \mathbb{N}$.*
(i) Assume $rp > 1$ and $p > 1$, or $r \ge 1$ and $p = 1$. There is c s.t. for every integer $m \in \{0{:}\lfloor r \rfloor\}$, all $v \in W^{r,p}(D_K;\mathbb{R}^q)$ and $K \in \mathcal{T}_h^\circ$, or all $v \in W_{0,\gamma}^{r,p}(D;\mathbb{R}^q)$ and $K \in \mathcal{T}_h^\partial$, and all $h \in \mathcal{H}$,

$$|v - \mathcal{I}_{h0}^{\mathrm{av}}(v)|_{W^{m,p}(K;\mathbb{R}^q)} \le c\,h_K^{r-m}|v|_{W^{r,p}(\check{\mathcal{T}}_K;\mathbb{R}^q)}. \tag{22.29}$$

The constant c is uniform w.r.t. r and p as long as rp is bounded from below away from 1, but c can grow unboundedly when $rp \downarrow 1$ if $p > 1$.
(ii) Assume $rp \le 1$ and $p > 1$ or $r < 1$ and $p = 1$. There is c, uniform w.r.t. r and p, s.t. for all $v \in W^{r,p}(D_K;\mathbb{R}^q)$, all $K \in \mathcal{T}_h^\circ$, and all $h \in \mathcal{H}$,

$$\|v - \mathcal{I}_{h0}^{\mathrm{av}}(v)\|_{L^p(K;\mathbb{R}^q)} \le c\,h_K^r|v|_{W^{r,p}(D_K;\mathbb{R}^q)}. \tag{22.30}$$

Moreover, there is c s.t. for all $v \in W^{r,p}(D;\mathbb{R}^q)$ and all $h \in \mathcal{H}$,

$$\|v - \mathcal{I}_{h0}^{\mathrm{av}}(v)\|_{L^p(D^\partial;\mathbb{R}^q)} \leq c\,h^r \ell_D^{-r} \|v\|_{W^{r,p}(D;\mathbb{R}^q)}, \qquad (22.31)$$

but c can grow unboundedly as $rp \uparrow 1$.

Proof. See [97, Thm. 6.4]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 22.15 ($W^{m,p}$-stability). *There exists c s.t. for all $p \in [1,\infty]$, every integer $m \in \{0{:}k+1\}$, all $K \in \mathcal{T}_h$, all $h \in \mathcal{H}$, and all $v \in W_{0,\gamma}^{m,p}(D;\mathbb{R}^q)$ if $m \geq 1$, or all $v \in W^{0,p}(D;\mathbb{R}^q) = L^p(D;\mathbb{R}^q)$ if $m = 0$,*

$$|\mathcal{I}_{h0}^{\mathrm{av}}(v)|_{W^{m,p}(K;\mathbb{R}^q)} \leq c\,|v|_{W^{m,p}(D_K;\mathbb{R}^q)}. \qquad (22.32)$$

Proof. Similar to the proof of Corollary 22.8. $\qquad\qquad\qquad\qquad\square$

Corollary 22.16 (Best approximation in L^p). *Let $r \in [0, k+1]$. Let $p \in [1,\infty)$ if $r \notin \mathbb{N}$ or $p \in [1,\infty]$ if $r \in \mathbb{N}$. (i) Assume $rp > 1$. There is c s.t. for all $v \in W_{0,\gamma}^{r,p}(D;\mathbb{R}^q)$ and all $h \in \mathcal{H}$,*

$$\inf_{v_h \in P_{k,0}(\mathcal{T}_h;\mathbb{R}^q)} \|v - v_h\|_{L^p(D;\mathbb{R}^q)} \leq c\,h^r |v|_{W^{r,p}(D;\mathbb{R}^q)}. \qquad (22.33)$$

The constant c is uniform w.r.t. r and p as long as rp is bounded from below away from 1, but c can grow unboundedly as $rp \downarrow 1$ if $p > 1$.
(ii) Assume $rp < 1$. There is c s.t. for all $v \in W^{r,p}(D;\mathbb{R}^q)$ and all $h \in \mathcal{H}$,

$$\inf_{v_h \in P_{k,0}(\mathcal{T}_h;\mathbb{R}^q)} \|v - v_h\|_{L^p(D;\mathbb{R}^q)} \leq c\,h^r \ell_D^{-r} \|v\|_{W^{r,p}(D;\mathbb{R}^q)}. \qquad (22.34)$$

The constant c is uniform w.r.t. r and p as long as rp is bounded from above away from 1, but c can grow unboundedly as $rp \uparrow 1$ if $p > 1$.

Remark 22.17 (rp). An estimate similar to (22.29) for the Scott–Zhang interpolation operator with $rp > 1$ and x = g and can be found in Ciarlet [73]. The estimate (22.31) for $rp < 1$ essentially says that the difference $v - \mathcal{I}_{h0}^{\mathrm{av}}(v)$ does not blow up too fast close to the boundary. A better result cannot be expected since $\mathcal{I}_{h0}^{\mathrm{av}}(v)$ is forced to be zero at ∂D, whereas v can blow up like $\rho^{-s}w$ with $w \in L^p(D;\mathbb{R}^q)$ and ρ is the distance function to the boundary. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 22.18 ($rp = 1$). Let $r \in (0,1)$. Using the notation from the real interpolation theory (see §A.5), one has $W^{r,p}(D) = [L^p(D), W^{1,p}(D)]_{r,p}$ since D is Lipschitz; see Tartar [189, Lem. 36.1]. Let us define

$$W_{00,\gamma}^{r,p}(D;\mathbb{R}^q) := [L^p(D;\mathbb{R}^q), W_{0,\gamma}^{1,p}(D;\mathbb{R}^q)]_{r,p}.$$

Using Theorem 22.14 with $l \in \{0,1\}$ and $m = 0$, the real interpolation theory implies that $\|v - \mathcal{I}_{h0}^{\mathrm{av}}(v)\|_{L^p(D;\mathbb{R}^q)} \leq ch^r \ell_D^{-r} \|v\|_{W_{00,\gamma}^{r,p}(D;\mathbb{R}^q)}$ for all $p \in [1,\infty)$ and all $v \in W_{00,\gamma}^{r,p}(D;\mathbb{R}^q)$. This estimate holds true in particular for $r = \frac{1}{p}$, but it is not fully satisfactory. Using the notation from Definition 3.17, it is

known for x = g and $p = 2$ that $[L^2(D; \mathbb{R}^q), H_0^1(D; \mathbb{R}^q)]_{r,2} = \widetilde{H}^r(D; \mathbb{R}^q)$ (see Lions and Magenes [135, Thm. 11.7], Tartar [189, p. 160], Chandler-Wilde et al. [65, Cor. 4.10]), but to the best of our knowledge, a full characterization of $W_{00,\gamma}^{r,p}(D; \mathbb{R}^q)$ in terms of zero extensions is not yet available in the literature for x = c and x = d. □

22.5 Conforming L^2-orthogonal projections

As an application of independent interest, we study here how the above quasi-interpolation operators can be used to analyze the *conforming L^2-orthogonal projection operator* $\mathcal{P}_h^{\mathrm{x}} : L^1(D; \mathbb{R}^q) \to P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$ s.t.

$$\int_D (\mathcal{P}_h^{\mathrm{x}}(v) - v, w_h)_{\ell^2(\mathbb{R}^q)} \, \mathrm{d}x = 0, \qquad \forall w_h \in P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q). \tag{22.35}$$

$\mathcal{P}_h^{\mathrm{x}}(v)$ is well defined for all $v \in L^1(D; \mathbb{R}^q)$ since $P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q) \subset L^\infty(D; \mathbb{R}^q)$. The Pythagorean identity

$$\|v\|_{L^2(D;\mathbb{R}^q)}^2 = \|\mathcal{P}_h^{\mathrm{x}}(v)\|_{L^2(D;\mathbb{R}^q)}^2 + \|v - \mathcal{P}_h^{\mathrm{x}}(v)\|_{L^2(D;\mathbb{R}^q)}^2$$

implies the L^2-stability property $\|\mathcal{P}_h^{\mathrm{x}}(v)\|_{L^2(D;\mathbb{R}^q)} \le \|v\|_{L^2(D;\mathbb{R}^q)}$ for all $v \in L^2(D; \mathbb{R}^q)$, and the optimality property

$$\mathcal{P}_h^{\mathrm{x}}(v) = \operatorname*{arg\,min}_{w_h \in P_k^{\mathrm{x}}(\mathcal{T}_h;\mathbb{R}^q)} \|v - w_h\|_{L^2(D;\mathbb{R}^q)}. \tag{22.36}$$

The conforming L^2-orthogonal projection $\mathcal{P}_h^{\mathrm{x}} : L^2(D; \mathbb{R}^q) \to P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$, should not be confused with the broken L^2-orthogonal projection from §18.4 which maps $L^2(D; \mathbb{R}^q)$ to the broken finite element space $P_k^{\mathrm{x,b}}(\mathcal{T}_h; \mathbb{R}^q)$.

Proposition 22.19 (Approximation in L^2). *Let $r \in (\frac{1}{2}, k+1]$. There is c s.t.*

$$\|v - \mathcal{P}_h^{\mathrm{x}}(v)\|_{L^2(D;\mathbb{R}^q)} \le c \left(\sum_{K \in \mathcal{T}_h} h_K^{2r} |v|_{H^r(K;\mathbb{R}^q)}^2 \right)^{\frac{1}{2}}, \tag{22.37}$$

for all $v \in H^r(D; \mathbb{R}^q)$ and all $h \in \mathcal{H}$, with c growing unboundedly as $r \downarrow \frac{1}{2}$. The bound (22.37) remains valid for all $r \in (0, \frac{1}{2}]$ with $|v|_{H^r(K;\mathbb{R}^q)}$ replaced by $|v|_{H^r(D_K;\mathbb{R}^q)}$ and c uniform w.r.t. r.

Proof. Combine (22.36) with Theorem 22.6 (with $p := 2$, $m := 0$) and use the regularity of the mesh sequence when $r > \frac{1}{2}$. □

The stability and convergence analysis in $H^1(D; \mathbb{R}^q)$ of the L^2-orthogonal projection in the specific case x = g is delicate owing to the global nature of this operator, which precludes the possibility of using local estimates. The situation can be simplified on quasi-uniform mesh sequences.

Definition 22.20 (Quasi-uniformity). *A mesh sequence* $(\mathcal{T}_h)_{h \in \mathcal{H}}$ *is said to be* quasi-uniform *if it is shape-regular and if there is c s.t.*

$$h_K \geq c\,h, \qquad \forall K \in \mathcal{T}_h, \ \forall h \in \mathcal{H}. \tag{22.38}$$

The main motivation for Definition 22.20 is to use global inverse inequalities. If the mesh sequence $(\mathcal{T}_h)_{h \in \mathcal{H}}$ is quasi-uniform, summing over the mesh cells the inverse inequality (12.1) (with $l := 1$, $m := 0$, $p := r := 2$), we infer that

$$|v_h|_{H^1(D)} \leq c\,h^{-1}\|v_h\|_{L^2(D)}, \qquad \forall v_h \in P_k^{\mathrm{g}}(\mathcal{T}_h). \tag{22.39}$$

Notice that we would only have the bound $|v_h|_{H^1(D)} \leq ch_\flat^{-1}\|v_h\|_{L^2(D)}$ with $h_\flat := \min_{K \in \mathcal{T}_h} h_K$ if the mesh sequence is shape-regular.

Proposition 22.21 (Stability and approximation in H^1). *Assume that the mesh sequence* $(\mathcal{T}_h)_{h \in \mathcal{H}}$ *is quasi-uniform. There is c s.t.*

$$|\mathcal{P}_h^{\mathrm{g}}(v)|_{H^1(D;\mathbb{R}^q)} \leq c\,|v|_{H^1(D;\mathbb{R}^q)}, \tag{22.40}$$

for all $v \in H^1(D;\mathbb{R}^q)$ and all $h \in \mathcal{H}$ Moreover, the following holds true:

$$|v - \mathcal{P}_h^{\mathrm{g}}(v)|_{H^1(D;\mathbb{R}^q)} \leq c\left(\sum_{K \in \mathcal{T}_h} h_K^{2(r-1)}|v|_{H^r(K;\mathbb{R}^q)}^2 \right)^{\frac{1}{2}}, \tag{22.41}$$

for all $r \in [1, k+1]$, all $v \in H^r(D;\mathbb{R}^q)$, and all $h \in \mathcal{H}$.

Proof. We assume $q = 1$ for simplicity. The bound (22.40) results from

$$
\begin{aligned}
|\mathcal{P}_h^{\mathrm{g}}(v)|_{H^1(D)} &\leq |\mathcal{P}_h^{\mathrm{g}}(v) - \mathcal{I}_h^{\mathrm{g,av}}(v)|_{H^1(D)} + |\mathcal{I}_h^{\mathrm{g,av}}(v)|_{H^1(D)} \\
&= |\mathcal{P}_h^{\mathrm{g}}(v - \mathcal{I}_h^{\mathrm{g,av}}(v))|_{H^1(D)} + |\mathcal{I}_h^{\mathrm{g,av}}(v)|_{H^1(D)} \\
&\leq c\,h^{-1}\|\mathcal{P}_h^{\mathrm{g}}(v - \mathcal{I}_h^{\mathrm{g,av}}(v))\|_{L^2(D)} + |\mathcal{I}_h^{\mathrm{g,av}}(v)|_{H^1(D)} \\
&\leq c\,h^{-1}\|v - \mathcal{I}_h^{\mathrm{g,av}}(v)\|_{L^2(D)} + |\mathcal{I}_h^{\mathrm{g,av}}(v)|_{H^1(D)} \leq c'\,|v|_{H^1(D)},
\end{aligned}
$$

where we used the triangle inequality, that $\mathcal{P}_h^{\mathrm{g}}(\mathcal{I}_h^{\mathrm{g,av}}(v)) = \mathcal{I}_h^{\mathrm{g,av}}(v)$, the global inverse inequality (22.39), the L^2-stability of $\mathcal{P}_h^{\mathrm{g}}$, and the H^1-stability and approximation properties of $\mathcal{I}_h^{\mathrm{g,av}}$ from Theorem 22.6. To prove (22.41), we use (22.40) and similar arguments to infer that

$$
\begin{aligned}
|v - \mathcal{P}_h^{\mathrm{g}}(v)|_{H^1(D)} &\leq |v - \mathcal{I}_h^{\mathrm{g,av}}(v)|_{H^1(D)} + |\mathcal{P}_h^{\mathrm{g}}(v - \mathcal{I}_h^{\mathrm{g,av}}(v))|_{H^1(D)} \\
&\leq (1 + c)|v - \mathcal{I}_h^{\mathrm{g,av}}(v)|_{H^1(D)},
\end{aligned}
$$

and we conclude by invoking again Theorem 22.6. $\qquad\square$

Remark 22.22 (Zero trace). The above results can be adapted to the L^2-orthogonal projection $\mathcal{P}_{h0}^{\mathrm{x}} : L^1(D;\mathbb{R}^q) \to P_{k,0}^{\mathrm{x}}(\mathcal{T}_h;\mathbb{R}^q)$. $\qquad\square$

Remark 22.23 (H^1-stability). The H^1-stability of $\mathcal{P}_h^{\mathrm{g}}$ without the quasi-uniformity assumption has been extensively studied in the literature; see Crouzeix and Thomée [87], Eriksson and Johnson [95], Bramble et al. [43], Carstensen [60, 61], Boman [30], Bank and Yserentant [15], Gaspoz et al. [105]. Under a relatively mild grading condition on the mesh (assumed to be composed of simplices), it is shown in [15] that the following inequality holds true uniformly with respect to $h \in \mathcal{H}$ for polynomials up to order twelve if $d = 2$ and up to order seven if $d = 3$:

$$\|\hbar^{-1}\mathcal{P}_h^{\mathrm{g}}(v)\|_{L^2(D)} \le c\|\hbar^{-1}v\|_{L^2(D)}, \qquad \forall v \in L^2(D), \qquad (22.42)$$

where $\hbar$ is the piecewise constant function such that $\hbar_{|K} = h_K$ for all $K \in \mathcal{T}_h$. The H^1-stability of $\mathcal{P}_h^{\mathrm{g}}$ follows from (22.42); see Exercise 22.6. The grading condition on the mesh states that there is an integer-valued level function over $\mathcal{T}_h$, say $\ell : \mathcal{T}_h \longrightarrow \mathbb{N}\backslash\{0\}$, s.t. h_K is uniformly equivalent to $2^{-\ell(K)}$ and $|\ell(K) - \ell(K')| \le 1$ if K and K' share a vertex. An optimal grading estimate is obtained in [105] for adaptive triangulations ($d = 2$) generated by the Newest Vertex Bisection strategy (see Morin et al. [147], Stevenson [186, 187]).   □

Exercises

Exercise 22.1 ($\check{\mathcal{F}}_K^\circ$). Identify the set $\check{\mathcal{F}}_K^\circ$ for the canonical hybrid, Nédélec, and Raviart–Thomas elements.

Exercise 22.2 (L^p-stability). Prove directly, i.e., without using Lemma 22.3, the L^p-stability of $\mathcal{J}_h^{\mathrm{av}}$. (*Hint*: use Proposition 12.5.)

Exercise 22.3 (Poincaré–Steklov in D_K). The goal is to prove (22.20). Let $p \in [1, \infty]$, $K \in \mathcal{T}_h$, and $v \in W^{1,p}(D_K)$ (i) Let $K_l, K_r \in \check{\mathcal{T}}_K$ sharing an interface $F := \partial K_l \cap \partial K_r$. Show that

$$|K|^{\frac{1}{p}}|\underline{v}_{K_l} - \underline{v}_{K_r}| \le c\,h_K|v|_{W^{1,p}(K_l \cup K_r)}.$$

(*Hint*: observe that $|F|^{-\frac{1}{p}}|\underline{v}_{K_l} - \underline{v}_{K_r}| \le \|v_{K_l} - \underline{v}_{K_l}\|_{L^p(F)} + \|v_{K_r} - \underline{v}_{K_r}\|_{L^p(F)}$, then use the trace inequality (12.16).) (ii) Prove (22.20). (*Hint*: use that $\underline{v}_{D_K} - \underline{v}_{K'} = \sum_{K'' \in \check{\mathcal{T}}_K} \frac{|K''|}{|D_K|}(\underline{v}_{K''} - \underline{v}_{K'})$ for all $K' \in \check{\mathcal{T}}_K$.)

Exercise 22.4 (Polynomial approximation in D_K). Prove that there is c s.t. for all $r \in [0, k+1]$, all $p \in [1, \infty)$ if $r \notin \mathbb{N}$ or all $p \in [1, \infty]$ if $r \in \mathbb{N}$, every integer $m \in \{0 : \lfloor r \rfloor\}$, all $v \in W^{r,p}(D_K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$:

$$\inf_{g \in \mathbb{P}_{k,d}} |v - g|_{W^{m,p}(D_K)} \le c\,h_K^{r-m}|v|_{W^{r,p}(D_K)}. \qquad (22.43)$$

(*Hint*: use Morrey's polynomial as in the proof of Corollary 12.13.)

Exercise 22.5 (Approximation on faces). (i) Prove that

$$\|v - \mathcal{I}_h^{\mathrm{av}}(v)\|_{L^p(F)} \leq c\, h_K^{r-\frac{1}{p}} |v|_{W^{r,p}(\check{\mathcal{T}}_K)},$$

for all $p \in [1,\infty)$, all $r \in (\frac{1}{p}, k+1]$ if $p > 1$ or $r \in [1, k+1]$ if $p = 1$, all $v \in W^{r,p}(D_K)$, all $K \in \mathcal{T}_h$, all $F \in \mathcal{F}_K$, and all $h \in \mathcal{H}$ (c can grow unboundedly as $rp \downarrow 1$ if $p > 1$). (*Hint*: use the multiplicative trace inequality (12.16) or its fractional version (12.17).) (ii) Assume $k \geq 1$. Prove that

$$\|\nabla(v - \mathcal{I}_h^{\mathrm{av}}(v))\|_{\boldsymbol{L}^p(F)} \leq c\, h_K^{r-\frac{1}{p}} |v|_{W^{1+r,p}(\check{\mathcal{T}}_K)},$$

for all $r \in (\frac{1}{p}, k]$ if $p > 1$ or $r \in [1, k]$ if $p = 1$, all $v \in W^{1+r,p}(D_K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$.

Exercise 22.6 (L^2-projection). (i) Prove that (22.42) implies the H^1-stability of $\mathcal{P}_h^{\mathrm{g}}$. (*Hint*: adapt the proof of Proposition 22.21.) (ii) Set $\|y\|_{*,r} := \sup_{w \in H^r(D;\mathbb{R}^q)} \frac{(y,w)_{L^2(D;\mathbb{R}^q)}}{\|w\|_{H^r(D;\mathbb{R}^q)}}$ for all $y \in L^2(D;\mathbb{R}^q)$ (this is not the standard norm of the dual space $H^{-r}(D;\mathbb{R}^q) := (H_0^r(D;\mathbb{R}^q))'$). Prove that there is c s.t. for every integer $r \in \{1{:}k+1\}$, all $v \in L^2(D;\mathbb{R}^q)$, and all $h \in \mathcal{H}$,

$$\|v - \mathcal{P}_h(v)\|_{*,r} \leq c\, h^r \|v - \mathcal{P}_h(v)\|_{L^2(D;\mathbb{R}^q)},$$
$$\|v - \mathcal{P}_{h0}(v)\|_{H^{-r}(D;\mathbb{R}^q)} \leq c\, h^r \|v - \mathcal{P}_{h0}(v)\|_{L^2(D;\mathbb{R}^q)}.$$

(*Hint*: use $\mathcal{I}_h^{\mathrm{av}}(v)$.)

Exercise 22.7 (Discrete commutator). Let $(\mathcal{T}_h)_{h \in \mathcal{H}}$ be a shape-regular mesh sequence. The goal is to prove that there is c s.t. for every integers $l \in \{0{:}1\}$ and $m \in \{0{:}l\}$, all $p \in [1,\infty]$, all $v_h \in P_k^{\mathrm{g}}(\mathcal{T}_h)$, all $K \in \mathcal{T}_h$, all $h \in \mathcal{H}$, and all ϕ in $W^{1+l,\infty}(D)$,

$$\|\phi v_h - \mathcal{I}_h^{\mathrm{g,av}}(\phi v_h)\|_{W^{m,p}(K)} \leq c\, h_K^{1+l-m} \|v_h\|_{W^{l,p}(D_K)} \|\phi\|_{W^{1+l,\infty}(D_K)}.$$

This property provides a useful tool to analyze nonlinear problems; see Bertoluzza [23] and Johnson and Szepessy [121]. (i) Fix $K \in \mathcal{T}_h$. Let $\underline{v}_{D_K}$ denote the mean value of v_h in D_K. Prove that

$$\|\phi \underline{v}_{D_K} - \mathcal{I}_h^{\mathrm{g,av}}(\phi \underline{v}_{D_K})\|_{W^{m,p}(K)} \leq c\, h_K^{1+l-m} \|v_h\|_{L^p(D_K)} \|\phi\|_{W^{1+l,\infty}(D_K)}.$$

(*Hint*: use Theorem 22.6 and verify that $\|\underline{v}_{D_K}\|_{L^p(D_K)} \leq \|v_h\|_{L^p(D_K)}$.) (ii) Set $\eta_h := v_h - \underline{v}_{D_K}$. Prove that

$$\|\phi \eta_h - \mathcal{I}_h^{\mathrm{g,av}}(\phi \eta_h)\|_{W^{m,p}(K)} \leq c h_K^{1+l-m} \|v_h\|_{W^{l,p}(D_K)} \|\phi\|_{W^{1,\infty}(D_K)}.$$

(*Hint*: observe that $\phi(\boldsymbol{x}_K)\eta_h = \mathcal{I}_h^{\mathrm{g,av}}(\phi(\boldsymbol{x}_K)\eta_h)$ where $\boldsymbol{x}_K$ is some point in K, e.g., the barycenter of K, then use (22.20) to bound η_h.) Conclude.

Part IV, Chapter 23

Commuting quasi-interpolation

The quasi-interpolation operators $\mathcal{I}_h^{\text{av}}$ and $\mathcal{I}_{h0}^{\text{av}}$ introduced in Chapter 22 are L^1-stable, are projections, and have optimal (local) approximation properties. However, they do not commute with the usual differential operators ∇, $\nabla\times$, and $\nabla\cdot$, which makes them difficult to use to approximate simultaneously a vector-valued function and its curl or its divergence. Since these commuting properties are important in some applications, we introduce in this chapter quasi-interpolation operators that are L^1-stable, are projections, have optimal (global) approximation properties, and have the expected commuting properties. The key idea is to compose the canonical interpolation operators defined in §19.3 with mollification operators, i.e., smoothing operators based on convolution with a smooth kernel.

23.1 Smoothing by mollification

Smoothing by *mollification* (i.e., by convolution with a smooth kernel) is an important tool for the analysis and approximation of PDEs that has been introduced by Leray [133, p. 206], Sobolev [179, p. 487], and Friedrichs [102, pp. 136–139]. The goal of this section is to define mollification operators that commute with the usual differential operators, and that converge optimally when the function to be smoothed is defined on a Lipschitz domain D in $\mathbb{R}^d$. We use a shrinking technique of D (see Bonito et al. [33] and [96]) to avoid the need to extend the function to be smoothed outside D. The starting point is that, since D is a Lipschitz domain, Proposition 2.3 in Hofmann et al. [118] implies the existence of a vector field $\boldsymbol{j} \in \boldsymbol{C}^\infty(\mathbb{R}^d)$ that is globally transversal on ∂D (i.e., there is a real number $\gamma > 0$ such that $\boldsymbol{n}(\boldsymbol{x})\cdot\boldsymbol{j}(\boldsymbol{x}) \geq \gamma$ at a.e. point $\boldsymbol{x}$ on ∂D where $\boldsymbol{n}$ is the unit normal vector pointing outward D) and $\|\boldsymbol{j}(\boldsymbol{x})\|_{\ell^2} = 1$ for all $\boldsymbol{x} \in \partial D$.

Lemma 23.1 (Shrinking mapping). *For all $\delta \in [0,1]$, define the mapping*

$$\varphi_\delta : \mathbb{R}^d \ni \boldsymbol{x} \longmapsto \boldsymbol{x} - \delta \boldsymbol{j}(\boldsymbol{x}) \in \mathbb{R}^d. \qquad (23.1)$$

Then $\varphi_\delta \in C^\infty(\mathbb{R}^d)$ for all $\delta \in [0,1]$, and for every integer $k \in \mathbb{N}$, there is c such that $\max_{\boldsymbol{x} \in D} \|D^k \varphi_\delta(\boldsymbol{x}) - D^k \boldsymbol{x}\|_{\ell^2} \leq c \, \ell_D^{-k} \delta$ for all $\delta \in [0,1]$, with $\ell_D := \mathrm{diam}(D)$. Moreover, there is $r > 0$ such that

$$\varphi_\delta(D) + B(\boldsymbol{0}, \delta r) \subset D, \qquad \forall \delta \in [0,1]. \qquad (23.2)$$

Proof. The smoothness properties are consequences of $\boldsymbol{j}$ being of class C^∞ and D being bounded, whereas (23.2) follows from Proposition 4.15 in [118] and the uniform cone property (see [118, pp. 599–600]).   □

Let us consider the following kernel:

$$\rho(\boldsymbol{y}) := \begin{cases} \eta \, e^{-\frac{1}{1-\|\boldsymbol{y}\|_{\ell^2}^2}} & \text{if } \|\boldsymbol{y}\|_{\ell^2} < 1, \\ 0 & \text{if } \|\boldsymbol{y}\|_{\ell^2} \geq 1, \end{cases} \qquad (23.3)$$

where η is chosen s.t. $\int_{\mathbb{R}^d} \rho(\boldsymbol{y}) \, \mathrm{d}y = \int_{B(\boldsymbol{0},1)} \rho(\boldsymbol{y}) \, \mathrm{d}y = 1$. Let $\delta \in [0,1]$ and let $f \in L^1(D; \mathbb{R}^q)$. Given some smooth field $\mathbb{K}_\delta : D \to \mathbb{R}^{q \times q}$, we define a mollification operator as follows:

$$(\mathcal{K}_\delta(f))(\boldsymbol{x}) := \int_{B(\boldsymbol{0},1)} \rho(\boldsymbol{y}) \mathbb{K}_\delta(\boldsymbol{x}) f(\varphi_\delta(\boldsymbol{x}) + (\delta r)\boldsymbol{y}) \, \mathrm{d}y, \quad \forall \boldsymbol{x} \in D. \qquad (23.4)$$

Note that the definition (23.4) makes sense owing to (23.2). The examples we have in mind for the field $\mathbb{K}_\delta$ (inspired by Schöberl [172, 174]) are

$$\mathbb{K}_\delta^{\mathrm{g}}(\boldsymbol{x}) := 1 \ (q = 1), \qquad\qquad \mathbb{K}_\delta^{\mathrm{c}}(\boldsymbol{x}) := \mathbb{J}_\delta^{\mathsf{T}}(\boldsymbol{x}) \ (q = d = 3), \qquad (23.5\mathrm{a})$$

$$\mathbb{K}_\delta^{\mathrm{d}}(\boldsymbol{x}) := \det(\mathbb{J}_\delta(\boldsymbol{x}))\mathbb{J}_\delta^{-1}(\boldsymbol{x}) \ (q = d), \quad \mathbb{K}_\delta^{\mathrm{b}}(\boldsymbol{x}) := \det(\mathbb{J}_\delta(\boldsymbol{x})) \ (q = 1), \quad (23.5\mathrm{b})$$

where $\mathbb{J}_\delta$ is the Jacobian matrix of φ_δ at $\boldsymbol{x} \in D$. The mollification operator built using the field $\mathbb{K}_\delta^{\mathrm{x}}$ is denoted by $\mathcal{K}_\delta^{\mathrm{x}}$ with $\mathrm{x} \in \{\mathrm{g, c, d, b}\}$.

Let us just state the main properties of the mollification operator $\mathcal{K}_\delta$. We refer the reader to [96] for the proofs.

Lemma 23.2 (Smoothness). *For all $f \in L^1(D; \mathbb{R}^q)$ and all $\delta \in (0,1]$, $\mathcal{K}_\delta(f) \in C^\infty(\overline{D}; \mathbb{R}^q)$, i.e., $\mathcal{K}_\delta(f) \in C^\infty(D; \mathbb{R}^q)$ and $\mathcal{K}_\delta(f)$ as well as all its derivatives admit a continuous extension to $\overline{D}$.*

Let $p \in [1, \infty]$. Assuming $d = 3$, let us set (see (4.9))

$$Z^{\mathrm{g},p}(D) := W^{1,p}(D) = \{f \in L^p(D) \mid \nabla f \in \boldsymbol{L}^p(D)\}, \qquad (23.6\mathrm{a})$$

$$\boldsymbol{Z}^{\mathrm{c},p}(D) := \{\boldsymbol{g} \in \boldsymbol{L}^p(D) \mid \nabla{\times}\boldsymbol{g} \in \boldsymbol{L}^p(D)\}, \qquad (23.6\mathrm{b})$$

$$\boldsymbol{Z}^{\mathrm{d},p}(D) := \{\boldsymbol{g} \in \boldsymbol{L}^p(D) \mid \nabla{\cdot}\boldsymbol{g} \in L^p(D)\}. \qquad (23.6\mathrm{c})$$

Lemma 23.3 (Commuting). *The following holds true:*

(i) $\nabla\mathcal{K}_\delta^{\mathrm{g}}(f) = \mathcal{K}_\delta^{\mathrm{c}}(\nabla f)$ *for all* $f \in Z^{\mathrm{g},p}(D)$,
(ii) $\nabla\times\mathcal{K}_\delta^{\mathrm{c}}(\boldsymbol{g}) = \mathcal{K}_\delta^{\mathrm{d}}(\nabla\times\boldsymbol{g})$ *for all* $\boldsymbol{g} \in \boldsymbol{Z}^{\mathrm{c},p}(D)$,
(iii) $\nabla\cdot\mathcal{K}_\delta^{\mathrm{d}}(\boldsymbol{g}) = \mathcal{K}_\delta^{\mathrm{b}}(\nabla\cdot\boldsymbol{g})$ *for all* $\boldsymbol{g} \in \boldsymbol{Z}^{\mathrm{d},p}(D)$,

i.e., letting $Z^{\mathrm{b},p}(D) := L^p(D)$, *the following diagrams commute:*

$$
\begin{array}{ccccccc}
Z^{\mathrm{g},p}(D) & \xrightarrow{\ \nabla\ } & \boldsymbol{Z}^{\mathrm{c},p}(D) & \xrightarrow{\ \nabla\times\ } & \boldsymbol{Z}^{\mathrm{d},p}(D) & \xrightarrow{\ \nabla\cdot\ } & Z^{\mathrm{b},p}(D) \\
\downarrow{\scriptstyle\mathcal{K}_\delta^{\mathrm{g}}} & & \downarrow{\scriptstyle\mathcal{K}_\delta^{\mathrm{c}}} & & \downarrow{\scriptstyle\mathcal{K}_\delta^{\mathrm{d}}} & & \downarrow{\scriptstyle\mathcal{K}_\delta^{\mathrm{b}}} \\
C^\infty(\overline{D}) & \xrightarrow{\ \nabla\ } & \boldsymbol{C}^\infty(\overline{D}) & \xrightarrow{\ \nabla\times\ } & \boldsymbol{C}^\infty(\overline{D}) & \xrightarrow{\ \nabla\cdot\ } & C^\infty(\overline{D})
\end{array}
$$

Theorem 23.4 (Convergence). *There are* c *and* $\delta_0 > 0$ *s.t.*

$$
\|\mathcal{K}_\delta(f)\|_{L^p(D;\mathbb{R}^q)} \le c\,\|f\|_{L^p(D;\mathbb{R}^q)}, \quad \forall \delta \in [0, \delta_0]. \tag{23.7}
$$

Moreover, we have

$$
\lim_{\delta\to 0}\|\mathcal{K}_\delta(f) - f\|_{L^p(D;\mathbb{R}^q)} = 0, \quad \forall f \in L^p(D;\mathbb{R}^q), \quad \forall p \in [1,\infty). \tag{23.8}
$$

Finally, for all $s \in (0,1]$, *there is* c *s.t. for all* $f \in W^{s,p}(D;\mathbb{R}^q)$, *all* $\delta \in [0,\delta_0]$, *and all* $p \in [1,\infty)$ $(p \in [1,\infty]$ *if* $s = 1)$,

$$
\|\mathcal{K}_\delta(f) - f\|_{L^p(D;\mathbb{R}^q)} \le c\,\delta^s \ell_D^{-s}\|f\|_{W^{s,p}(D;\mathbb{R}^q)}. \tag{23.9}
$$

Corollary 23.5 (Convergence of derivatives). *The following convergence results hold true for all* $p \in [1,\infty)$:

(i) $\lim_{\delta\to 0}\|\nabla(\mathcal{K}_\delta^{\mathrm{g}}(f) - f)\|_{\boldsymbol{L}^p(D)} = 0$, $\forall f \in Z^{\mathrm{g},p}(D)$.
(ii) $\lim_{\delta\to 0}\|\nabla\times(\mathcal{K}_\delta^{\mathrm{c}}(\boldsymbol{g}) - \boldsymbol{g})\|_{\boldsymbol{L}^p(D)} = 0$, $\forall \boldsymbol{g} \in \boldsymbol{Z}^{\mathrm{c},p}(D)$.
(iii) $\lim_{\delta\to 0}\|\nabla\cdot(\mathcal{K}_\delta^{\mathrm{d}}(\boldsymbol{g}) - \boldsymbol{g})\|_{L^p(D)} = 0$, $\forall \boldsymbol{g} \in \boldsymbol{Z}^{\mathrm{d},p}(D)$.

Moreover, for all $s \in (0,1]$ *and all* $p \in [1,\infty)$, *there is* c *s.t. for all* $\delta \in [0,\delta_0]$, *the following holds true under the same smoothness assumptions on* f *and* $\boldsymbol{g}$ *(c does not depend on* $p \in [1,\infty]$ *if* $s = 1$):

(i) *If* $\nabla f \in \boldsymbol{W}^{s,p}(D)$, $\|\nabla(\mathcal{K}_\delta^{\mathrm{g}}(f) - f)\|_{\boldsymbol{L}^p(D)} \le c\,\delta^s \ell_D^{-s}\|\nabla f\|_{\boldsymbol{W}^{s,p}(D)}$.
(ii) *If* $\nabla\times\boldsymbol{g} \in \boldsymbol{W}^{s,p}(D)$, $\|\nabla\times(\mathcal{K}_\delta^{\mathrm{c}}(\boldsymbol{g}) - \boldsymbol{g})\|_{\boldsymbol{L}^p(D)} \le c\,\delta^s \ell_D^{-s}\|\nabla\times\boldsymbol{g}\|_{\boldsymbol{W}^{s,p}(D)}$.
(iii) *If* $\nabla\cdot\boldsymbol{g} \in W^{s,p}(D)$, $\|\nabla\cdot(\mathcal{K}_\delta^{\mathrm{d}}(\boldsymbol{g}) - \boldsymbol{g})\|_{L^p(D)} \le c\,\delta^s \ell_D^{-s}\|\nabla\cdot\boldsymbol{g}\|_{W^{s,p}(D)}$.

Proof. Combine Lemma 23.3 with Theorem 23.4. $\qquad\square$

Remark 23.6 (Convergence in D). Corollary 23.5(i) strengthens the original result by Friedrichs where strong convergence of the gradient only occurs in compact subsets of D (see, e.g., [48, Thm. 9.2]). Note though that Corollary 23.5(i) is valid on Lipschitz domains, whereas the original result by Friedrichs is valid on every open set. $\qquad\square$

Remark 23.7 (Literature). Another possibility to mollify scalar-valued functions over bounded sets without invoking extensions is the convolution-translation technique by Blouza and Le Dret [25] and Girault and Scott [108]. However, it is not clear how to extend this technique to vector-valued functions and at the same time achieve the commuting properties of Lemma 23.3. □

Remark 23.8 (Density). Lemma 23.2 together with Lemma 23.3 and (23.8) implies that $C^\infty(\overline{D}; \mathbb{R}^q)$ is dense in $Z^{\mathrm{x},p}(D)$ for all $\mathrm{x} \in \{\mathrm{g}, \mathrm{c}, \mathrm{d}\}$.   □

Remark 23.9 (Norm vs. seminorm). The estimate (23.9) for $\mathcal{K}_\delta^{\mathrm{g}}$ can be sharpened to $\|\mathcal{K}_\delta^{\mathrm{g}}(f) - f\|_{L^p(D;\mathbb{R}^q)} \leq c\,\delta^s|f|_{W^{s,p}(D;\mathbb{R}^q)}$ since constants are invariant under $\mathcal{K}_\delta^{\mathrm{g}}$. This is not the case for $\mathcal{K}_\delta^{\mathrm{x}}$ with $\mathrm{x} \in \{\mathrm{c}, \mathrm{d}\}$. This is the reason why the error estimates (i), (ii), (iii) in Corollary 23.5 involve norms instead of seminorms on the right-hand side. We refer the reader to Exercise 23.5 where the mollifiers are modified to preserve constants while still commuting with the differential operators.   □

23.2 Mesh-dependent mollification

Our ultimate goal is to compose the canonical interpolation operators defined in §19.3 with the mollification operators from §23.1. To achieve optimal convergence properties, we need to relate the shrinking parameter δ to the meshsize. The difficulty we face is that the parameter δ is so far defined globally over D, whereas the meshsize can change locally. Requiring the parameter δ to be of the order of the meshsize would limit the applications of the method to quasi-uniform mesh sequences. To handle shape-regular mesh sequences, we redefine the mollification operators with a space-dependent shrinking parameter $\delta(\boldsymbol{x})$, $\boldsymbol{x} \in D$.

Let $(\mathcal{T}_h)_{h\in\mathcal{H}}$ be a shape-regular sequence of affine meshes so that each mesh covers D exactly. It is possible to define a meshsize function $\mathfrak{h} \in C^{0,1}(\overline{D}; \mathbb{R})$ so that there are three constants $c, c', c'' > 0$ s.t. for all $K \in \mathcal{T}_h$ and all $h \in \mathcal{H}$,

$$\|\mathfrak{h}\|_{W^{1,\infty}(D;\mathbb{R})} \leq c, \qquad c'h_K \leq \mathfrak{h}(\boldsymbol{x}) \leq c''h_K, \quad \forall \boldsymbol{x} \in K. \qquad (23.10)$$

One possibility to construct $\mathfrak{h}$ consists of applying the averaging operator $\mathcal{J}_h^{\mathrm{g,av}}$ of degree $k = 1$ from §22.2 to the piecewise constant function that is equal to h_K in each mesh cell K. We introduce $\epsilon \in (0,1)$ and define

$$\delta(\boldsymbol{x}) := \epsilon\mathfrak{h}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in D. \qquad (23.11)$$

The actual value of ϵ will be fixed later; see (23.17). Then we define $\boldsymbol{\varphi}_\delta$ by setting $\boldsymbol{\varphi}_\delta(\boldsymbol{x}) := \boldsymbol{x} - \delta(\boldsymbol{x})\boldsymbol{j}(\boldsymbol{x})$ (compare with (23.1)), and we define generic mollifying operators $\mathcal{K}_\delta$ as in (23.4) by setting (we use the same notation for simplicity)

$$(\mathcal{K}_\delta(f))(\boldsymbol{x}) := \int_{B(\boldsymbol{0},1)} \rho(\boldsymbol{y})\mathbb{K}_\delta(\boldsymbol{x},\boldsymbol{y}) f(\boldsymbol{\varphi}_\delta(\boldsymbol{x}) + \delta(\boldsymbol{x})\boldsymbol{y})\,\mathrm{d}y, \qquad (23.12)$$

for all $\boldsymbol{x} \in D$ and all $f \in L^1(D;\mathbb{R}^q)$, where $\mathbb{K}_\delta(\boldsymbol{x},\boldsymbol{y})$ is related to the Jacobian matrix $\mathbb{J}_\delta$ of the mapping $\boldsymbol{x} \mapsto \boldsymbol{\varphi}_\delta(\boldsymbol{x}) + \delta(\boldsymbol{x})\boldsymbol{y}$ at $\boldsymbol{x} \in D$ as above (notice the additional dependence on $\boldsymbol{y}$). Lemma 23.1 holds true for $k \in \{0,1\}$ only, and the smoothness statement in Lemma 23.2 must be replaced by the weaker statement $\mathcal{K}_\delta(f) \in C^1(\overline{D};\mathbb{R}^q)$ for all $f \in L^1(D;\mathbb{R}^q)$, since δ is only Lipschitz. All the other statements in §23.1 remain unchanged. That $\mathcal{K}_\delta(f)$ is well defined will follow from (23.15b) below.

Recall the discrete setting outlined in §22.1. In particular, we consider affine geometric mappings $\boldsymbol{T}_K : \widehat{K} \to K$ and transformations ψ_K of the form $\psi_K(v) := \mathbb{A}_K(v \circ \boldsymbol{T}_K)$ for some matrix $\mathbb{A}_K \in \mathbb{R}^{q \times q}$ satisfying $\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2} \leq c$ uniformly w.r.t. $K \in \mathcal{T}_h$ and $h \in \mathcal{H}$. Let $K \in \mathcal{T}_h$ and recall the sets $\check{\mathcal{T}}_K$ and D_K defined in (22.4). We now sharpen the assumption (22.6) on the dofs by assuming that there is a partition $\mathcal{N} = \mathcal{N}_0 \cup \mathcal{N}'$ s.t. for all $K \in \mathcal{T}_h$, the dofs $\sigma_{K,i}$ is either an evaluation at a point $\boldsymbol{a}_{K,i}$ in K if $i \in \mathcal{N}_0$, or an integral over a (closed) geometric entity $\mathcal{M}_{K,i}$ that can be an edge of K, a face of K, or K itself (with the obvious extension to higher dimension) if $i \in \mathcal{N}'$. This assumption is formalized by assuming that there is c s.t. for all $v \in V^{\mathrm{x}}(K)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$,

$$|\sigma_{K,i}(v)| \leq c\,\|\mathbb{A}_K\|_{\ell^2} \times \begin{cases} \|v(\boldsymbol{a}_{K,i})\|_{\ell^2} & \text{if } i \in \mathcal{N}_0, \\ |\mathcal{M}_{K,i}|^{-1}\|v\|_{L^1(\mathcal{M}_{K,i};\mathbb{R}^q)} & \text{if } i \in \mathcal{N}', \end{cases} \qquad (23.13)$$

where $V^{\mathrm{g}}(K) := W^{s,p}(K)$, $\boldsymbol{V}^{\mathrm{c}}(K) := \boldsymbol{W}^{s-\frac{1}{p},p}(K)$, $\boldsymbol{V}^{\mathrm{d}}(K) := \boldsymbol{W}^{s-\frac{2}{p},p}(K)$ with $sp > d$ and $p > 1$ or $s = d$ and $p = 1$ (see §19.3). Notice that (23.13) sharpens Assumption 22.1 on the face dofs.

For all $i \in \mathcal{N}_0$, since $\boldsymbol{a}_{K,i} = \boldsymbol{T}_K(\widehat{\boldsymbol{a}}_i)$ for some reference point $\widehat{\boldsymbol{a}}_i$ in $\widehat{K}$ and since card$(\mathcal{N})$ is finite, there exists a distance $\widehat{\ell}_0 > 0$ (only depending on the reference element) such that only one of the following four situations occurs: (1) $\widehat{\boldsymbol{a}}_i$ is a vertex of $\widehat{K}$; (2) $\widehat{\boldsymbol{a}}_i$ is in the interior of an edge of $\widehat{K}$ and is at least at distance $\widehat{\ell}_0$ from any vertex; (3) $\widehat{\boldsymbol{a}}_i$ is in the interior of a face of $\widehat{K}$ and is at least at distance $\widehat{\ell}_0$ from any edge; (4) $\widehat{\boldsymbol{a}}_i$ is in the interior of $\widehat{K}$ and is at least at distance $\widehat{\ell}_0$ from any face. The regularity of the mesh sequence implies that there is a constant $c_\flat$ (depending on $\widehat{\ell}_0$ but uniform with respect to i, K, and h) s.t. the open ball $B(\boldsymbol{a}_{K,i}, c_\flat h_K)$ has the following property: For all $K' \in \mathcal{T}_h$ such that $K' \cap B(\boldsymbol{a}_{K,i}, c_\flat h_K) \neq \emptyset$ and all $\boldsymbol{x} \in K' \cap B(\boldsymbol{a}_{K,i}, c_\flat h_K)$, the entire segment $[\boldsymbol{x}, \boldsymbol{a}_{K,i}]$ is in K'. One can always take $c_\flat$ small enough so that the only cells with a nonempty intersection with $B(\boldsymbol{a}_{K,i}, c_\flat h_K)$ belong to the set $\check{\mathcal{T}}_K$. The above observations imply that $B(\boldsymbol{a}_{K,i}, c_\flat h_K) \subset D_K$ and that

$$\|v(\boldsymbol{x}) - v(\boldsymbol{a}_{K,i})\|_{\ell^2} \leq \|\boldsymbol{x} - \boldsymbol{a}_{K,i}\|_{\ell^2}\|\nabla v\|_{L^\infty(K';\mathbb{R}^q)}, \qquad (23.14)$$

for all $\boldsymbol{x} \in K' \cap B(\boldsymbol{a}_{K,i}, c_\flat h_K)$ and all $v \in P_{K'}$. We define $\epsilon_{\max} > 0$ such that the following holds true, uniformly w.r.t. $K \in \mathcal{T}_h$, $h \in \mathcal{H}$, and every function $\delta := \epsilon\mathfrak{h}$ with $\epsilon \in (0, \epsilon_{\max}]$, as illustrated in Figure 23.1:

$$\boldsymbol{\varphi}_\delta(\boldsymbol{a}_{K,i}) + \delta(\boldsymbol{a}_{K,i})B(\boldsymbol{0}, 1) \subset B(\boldsymbol{a}_{K,i}, c_\flat h_K), \qquad \forall i \in \mathcal{N}_0, \qquad (23.15a)$$

$$\boldsymbol{\varphi}_\delta(\boldsymbol{x}) + \delta(\boldsymbol{x})B(\boldsymbol{0}, 1) \subset D_K, \qquad \forall \boldsymbol{x} \in K. \qquad (23.15b)$$
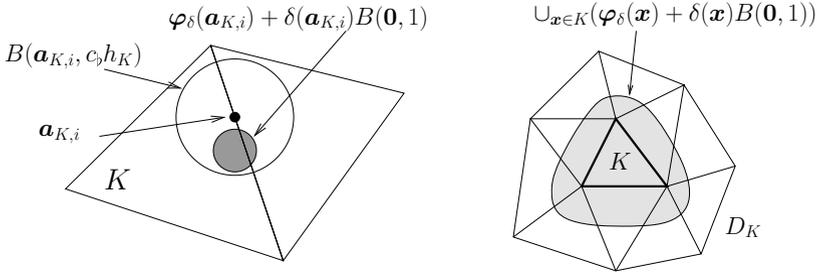
Fig. 23.1 Illustration of properties (23.15a) (left) and (23.15b) (right).

23.3 L^1-stable commuting projection

In this section, we build L^1-stable *commuting projections* $\mathcal{J}_h^{\mathrm{x}} : L^1(D; \mathbb{R}^q) \to P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$ with $\mathrm{x} \in \{\mathrm{g, c, d, b}\}$. Recall that $k \geq 1$ if $\mathrm{x = g}$ and $k \geq 0$ otherwise. The construction proceeds in two steps. First one constructs an L^1-stable commuting operator $\mathcal{I}_h \circ \mathcal{K}_\delta : L^1(D; \mathbb{R}^q) \to P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$ by composing the above mollification operators with the canonical interpolation operators. Then one takes the mollification parameter to be small enough with respect to the local meshsize so that the restriction of $\mathcal{I}_h \circ \mathcal{K}_\delta$ to $P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$ is invertible. The operator obtained by composing $\mathcal{I}_h \circ \mathcal{K}_\delta$ with the inverse of this restriction then leaves $P_k^{\mathrm{x}}(\mathcal{T}_h; \mathbb{R}^q)$ pointwise invariant and enjoys all the required properties. The construction presented in this section hinges on the seminal ideas of Schöberl [173, Lem. 6], Arnold et al. [11], Christiansen [68], and Christiansen and Winther [71, Lem. 4.2].

23.3.1 First step: the operator $\mathcal{I}_h \circ \mathcal{K}_\delta$

Owing to the smoothing properties of the (mesh-dependent) mollification operators, it makes sense to consider the following operators:

$$\mathcal{I}_h^{\mathrm{g}}\circ\mathcal{K}_\delta^{\mathrm{g}}: L^1(D)\to P_k^{\mathrm{g}}(\mathcal{T}_h),\qquad \mathcal{I}_{h0}^{\mathrm{g}}\circ\mathcal{K}_\delta^{\mathrm{g}}: L^1(D)\to P_{k,0}^{\mathrm{g}}(\mathcal{T}_h),\quad (23.16\mathrm{a})$$

$$\mathcal{I}_h^{\mathrm{c}}\circ\mathcal{K}_\delta^{\mathrm{c}}: \boldsymbol{L}^1(D)\to \boldsymbol{P}_k^{\mathrm{c}}(\mathcal{T}_h),\qquad \mathcal{I}_{h0}^{\mathrm{c}}\circ\mathcal{K}_\delta^{\mathrm{c}}: \boldsymbol{L}^1(D)\to \boldsymbol{P}_{k,0}^{\mathrm{c}}(\mathcal{T}_h),\quad (23.16\mathrm{b})$$

$$\mathcal{I}_h^{\mathrm{d}}\circ\mathcal{K}_\delta^{\mathrm{d}}: \boldsymbol{L}^1(D)\to \boldsymbol{P}_k^{\mathrm{d}}(\mathcal{T}_h),\qquad \mathcal{I}_{h0}^{\mathrm{d}}\circ\mathcal{K}_\delta^{\mathrm{d}}: \boldsymbol{L}^1(D)\to \boldsymbol{P}_{k,0}^{\mathrm{d}}(\mathcal{T}_h),\quad (23.16\mathrm{c})$$

as well as $\mathcal{I}_h^{\mathrm{b}}\circ\mathcal{K}_\delta^{\mathrm{b}}: L^1(D)\to P_k^{\mathrm{b}}(\mathcal{T}_h)$. All of these operators can be analyzed in a unified setting. Let $\mathcal{I}_h$ be one of the seven interpolation operators introduced above, and let $\mathcal{K}_\delta$ be the corresponding smoothing operator. Let $P(\mathcal{T}_h)$ denote the generic finite element space (whether homogeneous boundary conditions are enforced or not), i.e., $P(\mathcal{T}_h)$ is either the broken finite element space or one of the conforming finite element spaces from Chapter 19. Let $\mathbb{A}_K$ be the field defining ψ_K, and let $\mathbb{K}_\delta$ be the field defining $\mathcal{K}_\delta$. The symbol $\mathrm{x}\in\{\mathrm{g,c,d,b}\}$ and the indices k and 0 are omitted in the rest of this section. The difficulty we now face is that the finite element space $P(\mathcal{T}_h;\mathbb{R}^q)$ is not pointwise invariant under $\mathcal{I}_h\circ\mathcal{K}_\delta$. A key result to solve this difficulty is that $\mathcal{I}_h\circ\mathcal{K}_\delta$ has ϵ-dependent approximation properties on $P(\mathcal{T}_h;\mathbb{R}^q)$ (recall that the parameter ϵ is used to define the shrinking function δ in (23.11)).

Lemma 23.10 (Discrete L^p-approximation). *Let $\epsilon\in(0,\epsilon_{\max}]$. There is a constant $c_{\mathrm{stab}}>0$ s.t. $\|f_h-(\mathcal{I}_h\circ\mathcal{K}_\delta)(f_h)\|_{L^p(D;\mathbb{R}^q)}\le c_{\mathrm{stab}}\epsilon\|f_h\|_{L^p(D;\mathbb{R}^q)}$ for all $p\in[1,\infty]$, all $f_h\in P(\mathcal{T}_h;\mathbb{R}^q)$, and all $h\in\mathcal{H}$.*

Proof. Let $f_h\in P(\mathcal{T}_h;\mathbb{R}^q)$. Let us set $e:=f_h-\mathcal{K}_\delta(f_h)$ and $e_h:=f_h-\mathcal{I}_h(\mathcal{K}_\delta(f_h))$. Let $K\in\mathcal{T}_h$. The local shape functions satisfy $\|\theta_{K,i}\|_{L^p(K;\mathbb{R}^q)}\le c\,|K|^{\frac{1}{p}}\|\mathbb{A}_K^{-1}\|_{\ell^2}$ for all $i\in\mathcal{N}$ (apply Proposition 12.5 to $v:=\theta_{K,i}$). Since $e_h=\mathcal{I}_h(e)$, we infer that

$$\|e_h\|_{L^p(K;\mathbb{R}^q)}\le\sum_{i\in\mathcal{N}}|\sigma_{K,i}(e)|\|\theta_{K,i}\|_{L^p(K;\mathbb{R}^q)}\le\det(\mathbb{J}_K)^{\frac{1}{p}}\|\mathbb{A}_K^{-1}\|_{\ell^2}\sum_{i\in\mathcal{N}}|\sigma_{K,i}(e)|.$$

The rest of the proof consists of estimating $\sigma_{K,i}(e)$ for all $i\in\mathcal{N}$.
(1) Assume first that $i\in\mathcal{N}_0$. Using (23.13), we infer that $|\sigma_{K,i}(e)|\le c\,\|\mathbb{A}_K\|_{\ell^2}\|e(\boldsymbol{a}_{K,i})\|_{\ell^2}$. Since $e(\boldsymbol{a}_{K,i})=f_h(\boldsymbol{a}_{K,i})-\mathcal{K}_\delta(f_h(\boldsymbol{a}_{K,i}))$, we have

$$e(\boldsymbol{a}_{K,i})=\int_{B(\boldsymbol{0},1)}\rho(\boldsymbol{y})\mathbb{K}_\delta(\boldsymbol{a}_{K,i},\boldsymbol{y})\big(f_h(\boldsymbol{a}_{K,i})-f_h(\boldsymbol{\varphi}_\delta(\boldsymbol{a}_{K,i})+\delta(\boldsymbol{a}_{K,i})\boldsymbol{y})\big)\,\mathrm{d}y$$
$$+\int_{B(\boldsymbol{0},1)}\rho(\boldsymbol{y})(\mathbb{I}-\mathbb{K}_\delta(\boldsymbol{a}_{K,i},\boldsymbol{y}))\,\mathrm{d}yf_h(\boldsymbol{a}_{K,i}).$$

Note that f_h is single-valued at $\boldsymbol{a}_{K,i}$ since $i\in\mathcal{N}_0$, i.e., it makes sense to invoke $f_h(\boldsymbol{a}_{K,i})$. Since $\|\mathbb{K}_\delta\|_{L^\infty(D\times B(\boldsymbol{0},1);\mathbb{R}^{q\times q})}\le c$ and $\|\mathbb{K}_\delta-\mathbb{I}\|_{L^\infty(D\times B(\boldsymbol{0},1);\mathbb{R}^{q\times q})}\le c\epsilon$ (this follows from (23.10)–(23.11)), upon invoking (23.15a) (recall that $\epsilon\le\epsilon_{\max}$) and (23.14), we infer that

$$\|e(\boldsymbol{a}_{K,i})\|_{\ell^2} \le c_1 \max_{\boldsymbol{y}\in B(\boldsymbol{0},1)} \|f_h(\boldsymbol{a}_{K,i}) - f_h(\psi_\delta(\boldsymbol{a}_{K,i},\boldsymbol{y}))\|_{\ell^2} + c_2\epsilon\|f_h(\boldsymbol{a}_{K,i})\|_{\ell^2}$$

$$\le c_1' \,\delta(\boldsymbol{a}_{K,i}) \max_{K'\in\check{\mathcal{T}}_K} \|\nabla f_h\|_{\boldsymbol{L}^\infty(K';\mathbb{R}^q)} + c_2\epsilon\|f_h\|_{L^\infty(K;\mathbb{R}^q)}$$

$$\le c_1'' \,\epsilon h_K \max_{K'\in\check{\mathcal{T}}_K} \|\nabla f_h\|_{\boldsymbol{L}^\infty(K';\mathbb{R}^q)} + c_2\epsilon\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)},$$

with $\boldsymbol{\psi}_\delta(\boldsymbol{x},\boldsymbol{y}) := \boldsymbol{\varphi}_\delta(\boldsymbol{x}) + \delta(\boldsymbol{x})\boldsymbol{y}$. An inverse inequality in D_K (since the mesh sequence is shape-regular) gives $|\sigma_{K,i}(e)| \le c\,\epsilon\|\mathbb{A}_K\|_{\ell^2}\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)}$. In the above argument, we used that f_h is piecewise Lipschitz on each cell in $\check{\mathcal{T}}_K$ and that f_h is continuous at $\boldsymbol{a}_{K,i}$.

(2) Assume now that $i \in \mathcal{N}'$. We define $\check{\mathcal{T}}_{\mathcal{M}_{K,i}} := \{K' \in \check{\mathcal{T}}_K \mid \mathcal{M}_{K,i} \subset K'\}$ and we introduce

$$\mathcal{M}_{K,i}^\circ := \{\boldsymbol{x} \in \mathcal{M}_{K,i} \mid \boldsymbol{\varphi}_\delta(\boldsymbol{x}) + \delta(\boldsymbol{x})B(\boldsymbol{0},1) \subset \check{\mathcal{T}}_{\mathcal{M}_{K,i}}\},$$
$$\mathcal{M}_{K,i}^\partial := \mathcal{M}_{K,i}\backslash\mathcal{M}_{K,i}^\circ.$$

Using (23.13) and (23.15b), we infer that $\int_{\mathcal{M}_{K,i}^\circ}\|e\|_{\ell^2}\,\mathrm{d}s \le \mathfrak{T}_1 + \mathfrak{T}_2$ with

$$\mathfrak{T}_1 := \int_{\mathcal{M}_{K,i}^\circ}\int_{B(\boldsymbol{0},1)}\rho(\boldsymbol{y})\|\mathbb{K}_\delta(\boldsymbol{x},\boldsymbol{y})\|_{\ell_2}\,\|f_h(\boldsymbol{\psi}_\delta(\boldsymbol{x},\boldsymbol{y})) - f_h(\boldsymbol{x})\|_{\ell^2}\,\mathrm{d}y\,\mathrm{d}s$$

$$\le c_1 \int_{\mathcal{M}_{K,i}^\circ}\sum_{K'\in\check{\mathcal{T}}_{\mathcal{M}_{K,i}}}\int_{\substack{\boldsymbol{y}\in B(\boldsymbol{0},1)\\ \boldsymbol{\psi}_\delta(\boldsymbol{x},\boldsymbol{y})\in K'}}\|f_h(\boldsymbol{\psi}_\delta(\boldsymbol{x},\boldsymbol{y})) - f_h(\boldsymbol{x})\|_{\ell^2}\,\mathrm{d}s\,\mathrm{d}y$$

$$\le c_1' \,|\mathcal{M}_{K,i}|\epsilon h_K \sum_{K'\in\check{\mathcal{T}}_{\mathcal{M}_{K,i}}}\|\nabla f_h\|_{\boldsymbol{L}^\infty(K';\mathbb{R}^q)},$$

$$\mathfrak{T}_2 := \int_{\mathcal{M}_{K,i}^\circ}\int_{B(\boldsymbol{0},1)}\rho(\boldsymbol{y})(\mathbb{I} - \mathbb{K}_\delta(\boldsymbol{x},\boldsymbol{y}))\,\mathrm{d}yf_h(\boldsymbol{x})\,\mathrm{d}s$$

$$\le c_2|\mathcal{M}_{K,i}|\epsilon\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)}.$$

Using the regularity of the mesh sequence (i.e., $h_{K'} \le ch_K$) and an inverse inequality we obtain $\int_{\mathcal{M}_{K,i}^\circ}\|e\|_{\ell^2}\,\mathrm{d}s \le c\,|\mathcal{M}_{K,i}|\epsilon\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)}$. Notice again that in the above construction we used that f_h is (a priori) only piecewise Lipschitz. Moreover, if $\boldsymbol{x} \in \mathcal{M}_{K,i}^\partial$, there is $\boldsymbol{y} \in B(\boldsymbol{0},1)$ s.t. $\boldsymbol{z} := \boldsymbol{\varphi}_\delta(\boldsymbol{x}) + \delta(\boldsymbol{x})\boldsymbol{y}$ is not in $\check{\mathcal{T}}_{\mathcal{M}_{K,i}}$. The regularity of the mesh sequence implies that $\|\boldsymbol{z} - \boldsymbol{x}\|_{\ell^2} \ge c\,d(\boldsymbol{x},\partial\mathcal{M}_{K,i})$ and $\|\boldsymbol{z} - \boldsymbol{x}\|_{\ell^2} \le c\delta(\boldsymbol{x}) \le c'\epsilon h_K$. Combining these bounds we obtain that $|\mathcal{M}_{K,i}^\partial| \le c\epsilon h_K h_K^{d-2} \le c'\epsilon|\mathcal{M}_{K,i}|$. As a result, we have

$$\int_{\mathcal{M}_{K,i}^\partial}\|e\|_{\ell^2}\,\mathrm{d}s \le \int_{\mathcal{M}_{K,i}^\partial}(\|f_h\|_{\ell^2} + \|\mathcal{K}_\delta(f_h)\|_{\ell^2})\,\mathrm{d}s$$

$$\le c\,|\mathcal{M}_{K,i}^\partial|\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)} \le c'\,|\mathcal{M}_{K,i}|\epsilon\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)}.$$

The above two estimates yield $|\sigma_{K,i}(e)| \leq c\,\epsilon\|\mathbb{A}_K\|_{\ell^2}\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)}$.
(3) In conclusion, we have established that $|\sigma_{K,i}(e)| \leq c\,\epsilon\|\mathbb{A}_K\|_{\ell^2}\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)}$
for all the dofs. Since $\|\mathbb{A}_K\|_{\ell^2}\|\mathbb{A}_K^{-1}\|_{\ell^2}$ is uniformly bounded and using an
inverse inequality, we infer that

$$
\begin{aligned}
\|f_h - (\mathcal{I}_h\circ\mathcal{K}_\delta)(f_h)\|_{L^p(K;\mathbb{R}^q)} &= \|e_h\|_{L^p(K;\mathbb{R}^q)} \\
&\leq c\,\epsilon\det(\mathbb{J}_K)^{\frac{1}{p}}\|\mathbb{A}_K^{-1}\|_{\ell^2}\|\mathbb{A}_K\|_{\ell^2}\|f_h\|_{L^\infty(D_K;\mathbb{R}^q)} \leq c\,\epsilon\|f_h\|_{L^p(D_K;\mathbb{R}^q)}.
\end{aligned}
$$

We conclude by summing over $K \in \mathcal{T}_h$ and invoking the regularity of the
mesh sequence. $\qquad\square$

23.3.2 Second step: the operator $J_h\circ\mathcal{I}_h\circ\mathcal{K}_\delta$

Lemma 23.10 implies that $\|\mathbb{I} - (\mathcal{I}_h\circ\mathcal{K}_\delta)_{|P(\mathcal{T}_h;\mathbb{R}^q)}\|_{\mathcal{L}(L^p;L^p)} \leq c_{\mathrm{stab}}\epsilon$ for all
$\epsilon \in (0,\epsilon_{\max}]$, where $\mathbb{I}$ denotes the identity operator in $P(\mathcal{T}_h;\mathbb{R}^q)$. From now
on, we choose ϵ once and for all by setting

$$
\epsilon := \epsilon_{\min} := \min(\epsilon_{\max},(2c_{\mathrm{stab}})^{-1}). \tag{23.17}
$$

This choice implies that $\|\mathbb{I} - (\mathcal{I}_h\circ\mathcal{K}_\delta)_{|P(\mathcal{T}_h;\mathbb{R}^q)}\|_{\mathcal{L}(L^p;L^p)} \leq \frac{1}{2}$, which in turn
proves that $(\mathcal{I}_h\circ\mathcal{K}_\delta)_{|P(\mathcal{T}_h;\mathbb{R}^q)}$ is invertible. Let $J_h : P(\mathcal{T}_h;\mathbb{R}^q) \to P(\mathcal{T}_h;\mathbb{R}^q)$ be
the inverse of $(\mathcal{I}_h\circ\mathcal{K}_\delta)_{|P(\mathcal{T}_h;\mathbb{R}^q)}$, i.e.,

$$
J_h\circ(\mathcal{I}_h\circ\mathcal{K}_\delta)_{|P(\mathcal{T}_h;\mathbb{R}^q)} = (\mathcal{I}_h\circ\mathcal{K}_\delta)_{|P(\mathcal{T}_h;\mathbb{R}^q)}\circ J_h = \mathbb{I}. \tag{23.18}
$$

Note that the definition of J_h implies that $\|J_h\|_{\mathcal{L}(L^p;L^p)} \leq 2$. We have the
following important stability result.

Lemma 23.11 (L^p-stability). *Let $\epsilon := \epsilon_{\min}$ be defined in (23.17). There is
$c(\epsilon_{\min})$ s.t. for all $p \in [1,\infty]$ and all $h \in \mathcal{H}$, $\|\mathcal{I}_h\circ\mathcal{K}_\delta\|_{\mathcal{L}(L^p;L^p)} \leq c(\epsilon_{\min})$.*

Proof. Let $f \in L^p(D;\mathbb{R}^q)$ and assume $p < \infty$ (the argument for $p = \infty$ is
similar). Since $\mathcal{I}_h(v)_{|K} = \sum_{i\in\mathcal{N}} \sigma_{K,i}(v)\theta_{K,i}$, we infer that

$$
\|(\mathcal{I}_h\circ\mathcal{K}_\delta)(f)\|_{L^p(D;\mathbb{R}^q)}^p \leq c \sum_{K\in\mathcal{T}_h} \int_K \sum_{i\in\mathcal{N}} |\sigma_{K,i}(\mathcal{K}_\delta(f))|^p \|\theta_{K,i}\|_{\ell^2}^p \,\mathrm{d}x.
$$

Using (23.13), which yields $|\sigma_{K,i}(\mathcal{K}_\delta(f))| \leq c\,\|\mathbb{A}_K\|_{\ell^2}\|\mathcal{K}_\delta(f)\|_{L^\infty(K;\mathbb{R}^q)}$, and
recalling that $\|\theta_{K,i}\|_{L^\infty(K;\mathbb{R}^q)} \leq c\,\|\mathbb{A}_K^{-1}\|_{\ell^2}$, we infer that

$$
\|(\mathcal{I}_h\circ\mathcal{K}_\delta)(f)\|_{L^p(D;\mathbb{R}^q)}^p \leq c \sum_{K\in\mathcal{T}_h} \sum_{i\in\mathcal{N}} \|\mathbb{A}_K\|_{\ell^2}^p \|\mathcal{K}_\delta(f)\|_{L^\infty(K;\mathbb{R}^q)}^p \|\mathbb{A}_K^{-1}\|_{\ell^2}^p |K|.
$$

We conclude by using that there is c s.t.

$$
\|\mathcal{K}_\delta(f)\|_{L^\infty(K;\mathbb{R}^q)} \leq c\epsilon_{\min}^{-d}|K|^{-\frac{1}{p}}\|f\|_{L^p(D_K;\mathbb{R}^q)}, \tag{23.19}
$$

for all $f \in L^p(D; \mathbb{R}^q)$, all $K \in \mathcal{T}_h$, and all $h \in \mathcal{H}$; see Exercise 23.6.   □

23.3.3 Main results

We define the operator

$$\mathcal{J}_h := J_h \circ \mathcal{I}_h \circ \mathcal{K}_\delta, \tag{23.20}$$

that is,

$$\mathcal{J}_h^x := J_h^x \circ \mathcal{I}_h^x \circ \mathcal{K}_\delta^x : L^1(D; \mathbb{R}^q) \to P_k^x(\mathcal{T}_h; \mathbb{R}^q), \qquad \mathrm{x} \in \{\mathrm{g, c, d, b}\},$$
$$\mathcal{J}_{h0}^x := J_{h0}^x \circ \mathcal{I}_{h0}^x \circ \mathcal{K}_\delta^x : L^1(D; \mathbb{R}^q) \to P_{k,0}^x(\mathcal{T}_h; \mathbb{R}^q), \qquad \mathrm{x} \in \{\mathrm{g, c, d}\},$$

with $k \geq 1$ if $\mathrm{x = g}$ and $k \geq 0$ otherwise. We drop the symbol x and the index 0 whenever the context is unambiguous. Recall the spaces $Z^{\mathrm{x},p}(D)$ from §23.1 with $p \in [1, \infty]$. Let us also set $Z_0^{\mathrm{x},p}(D) := \overline{C_0^\infty(D)}^{Z^{\mathrm{x},p}(D)}$. Owing to Theorems 3.19 and 4.15, we also have $Z_0^{\mathrm{x},p}(D) = \{f \in Z^{\mathrm{x},p}(D) \mid \gamma^{\mathrm{x}}(f) = 0\}$ for $p \in (1, \infty)$.

Theorem 23.12 (Properties of $\mathcal{J}_h$). *The following properties hold true:*
(i) *$P_k(\mathcal{T}_h; \mathbb{R}^q)$ is pointwise invariant under $\mathcal{J}_h$.*
(ii) *There is c s.t. for all $p \in [1, \infty]$ and all $h \in \mathcal{H}$, $\|\mathcal{J}_h\|_{\mathcal{L}(L^p; L^p)} \leq c$ and*

$$\|f - \mathcal{J}_h(f)\|_{L^p(D;\mathbb{R}^q)} \leq c \inf_{f_h \in P_k(\mathcal{T}_h;\mathbb{R}^q)} \|f - f_h\|_{L^p(D;\mathbb{R}^q)}, \quad \forall f \in L^p(D;\mathbb{R}^q).$$

(iii) *The following diagrams commute for all $\kappa \in \mathbb{N}$:*

$$
\begin{array}{ccccccc}
Z^{\mathrm{g},p}(D) & \xrightarrow{\nabla} & \boldsymbol{Z}^{\mathrm{c},p}(D) & \xrightarrow{\nabla\times} & \boldsymbol{Z}^{\mathrm{d},p}(D) & \xrightarrow{\nabla\cdot} & Z^{\mathrm{b},p}(D) \\
\downarrow{\scriptstyle\mathcal{J}_h^{\mathrm{g}}} & & \downarrow{\scriptstyle\mathcal{J}_h^{\mathrm{c}}} & & \downarrow{\scriptstyle\mathcal{J}_h^{\mathrm{d}}} & & \downarrow{\scriptstyle\mathcal{J}_h^{\mathrm{b}}} \\
P_{\kappa+1}^{\mathrm{g}}(\mathcal{T}_h) & \xrightarrow{\nabla} & \boldsymbol{P}_\kappa^{\mathrm{c}}(\mathcal{T}_h) & \xrightarrow{\nabla\times} & \boldsymbol{P}_\kappa^{\mathrm{d}}(\mathcal{T}_h) & \xrightarrow{\nabla\cdot} & P_\kappa^{\mathrm{b}}(\mathcal{T}_h)
\end{array}
$$

$$
\begin{array}{ccccccc}
Z_0^{\mathrm{g},p}(D) & \xrightarrow{\nabla} & \boldsymbol{Z}_0^{\mathrm{c},p}(D) & \xrightarrow{\nabla\times} & \boldsymbol{Z}_0^{\mathrm{d},p}(D) & \xrightarrow{\nabla\cdot} & Z^{\mathrm{b},p}(D) \\
\downarrow{\scriptstyle\mathcal{J}_{h0}^{\mathrm{g}}} & & \downarrow{\scriptstyle\mathcal{J}_{h0}^{\mathrm{c}}} & & \downarrow{\scriptstyle\mathcal{J}_{h0}^{\mathrm{d}}} & & \downarrow{\scriptstyle\mathcal{J}_h^{\mathrm{b}}} \\
P_{\kappa+1,0}^{\mathrm{g}}(\mathcal{T}_h) & \xrightarrow{\nabla} & \boldsymbol{P}_{\kappa,0}^{\mathrm{c}}(\mathcal{T}_h) & \xrightarrow{\nabla\times} & \boldsymbol{P}_{\kappa,0}^{\mathrm{d}}(\mathcal{T}_h) & \xrightarrow{\nabla\cdot} & P_\kappa^{\mathrm{b}}(\mathcal{T}_h)
\end{array}
$$

Proof. Item (i) is a consequence of the definition of J_h (see (23.18)). Item (ii) is proved by observing that $\|J_h\|_{\mathcal{L}(L^p; L^p)} \leq 2$ and that $\mathcal{I}_h \circ \mathcal{K}_\delta$ is uniformly bounded since $\epsilon := \epsilon_{\min}$ is now a fixed real number (see Lemma 23.11 and (23.17)). Using that $\mathcal{J}_h(f_h) = f_h$ for all $f_h \in P_k(\mathcal{T}_h; \mathbb{R}^q)$, we infer that

$$\|f - \mathcal{J}_h(f)\|_{L^p(D;\mathbb{R}^q)} = \inf_{f_h \in P_k(\mathcal{T}_h;\mathbb{R}^q)} \|f - f_h - \mathcal{J}_h(f - f_h)\|_{L^p(D;\mathbb{R}^q)}$$

$$\leq (1 + \|\mathcal{J}_h\|_{\mathcal{L}(L^p;L^p)}) \inf_{f_h \in P_k(\mathcal{T}_h;\mathbb{R}^q)} \|f - f_h\|_{L^p(D;\mathbb{R}^q)},$$

which establishes (ii). Let us now prove (iii). We are just going to show that the leftmost top diagram commutes. The proof for the other diagrams is identical, and whether boundary conditions are imposed or not is irrelevant in the argument. Let us first show that $J_h^c(\nabla\phi_h) = \nabla(J_h^g(\phi_h))$ for all $\phi_h \in P_k^g(\mathcal{T}_h)$. Since $\mathbb{I} = (\mathcal{I}_h^g \circ \mathcal{K}_\delta^g)_{|P_k^g(\mathcal{T}_h)} \circ J_h^g$ (see (23.18)) and $(\mathcal{I}_h^g \circ \mathcal{K}_\delta^g)_{|P_k^g(\mathcal{T}_h)} \circ J_h^g = \mathcal{I}_h^g \circ \mathcal{K}_\delta^g \circ J_h^g$ (because the range of J_h^g is in $P_k^g(\mathcal{T}_h)$), we have

$$\nabla\phi_h = \nabla(\mathcal{I}_h^g \circ \mathcal{K}_\delta^g)_{|P_k^g(\mathcal{T}_h)}(J_h^g(\phi_h)) = \nabla\mathcal{I}_h^g(\mathcal{K}_\delta^g(J_h^g(\phi_h)))$$
$$= \mathcal{I}_h^c(\nabla(\mathcal{K}_\delta^g(J_h^g(\phi_h)))) = \mathcal{I}_h^c(\mathcal{K}_\delta^c(\nabla(J_h^g(\phi_h)))),$$

where we used that $\nabla\mathcal{I}_h^g = \mathcal{I}_h^c\nabla$ (see Lemma 19.6) and $\nabla\mathcal{K}_\delta^g = \mathcal{K}_\delta^c\nabla$ (see Lemma 23.3). Since $\nabla(J_h^g(\phi_h)) \in \boldsymbol{P}_k^c(\mathcal{T}_h)$, the above argument together with (23.18) proves that

$$\nabla\phi_h = (\mathcal{I}_h^c \circ \mathcal{K}_\delta^c)_{|\boldsymbol{P}_k^c(\mathcal{T}_h)}\nabla J_h^g(\phi_h) = (J_h^c)^{-1}\nabla(J_h^g(\phi_h)).$$

In conclusion, $J_h^c(\nabla\phi_h) = \nabla J_h^g(\phi_h)$. Now we finish the proof by using an arbitrary function $\phi \in Z^{g,p}(D)$ and infer that

$$\mathcal{J}_h^c(\nabla\phi) = J_h^c(\mathcal{I}_h^c(\mathcal{K}_\delta^c(\nabla\phi))) = J_h^c(\mathcal{I}_h^c(\nabla\mathcal{K}_\delta^g(\phi)))$$
$$= J_h^c(\nabla\mathcal{I}_h^g(\mathcal{K}_\delta^g(\phi))) = \nabla J_h^g(\mathcal{I}_h^g(\mathcal{K}_\delta^g(\phi))),$$

where the last equality results from $J_h^c(\nabla\phi_h) = \nabla J_h^g(\phi_h)$ for all $\phi_h \in P_k^g(\mathcal{T}_h;\mathbb{R}^q)$ (as established above). This proves that $\mathcal{J}_h^c(\nabla\phi) = \nabla\mathcal{J}_h^g(\phi)$. $\quad\square$

Remark 23.13 (Approximation property). The operators $\mathcal{J}_h$ are globally defined owing to the use of the inverse operator J_h. This means that these operators cannot be used to derive local approximation properties (in contrast with the quasi-interpolation operators constructed in the previous chapter). We refer the reader to Christiansen [69] where global approximation properties of the operators $\mathcal{J}_h$ are established in Sobolev spaces of fractional order. $\quad\square$

Remark 23.14 (Variants). A local construction of commuting projections is proposed in Falk and Winther [100], but stability is achieved in the graph space of the appropriate differential operator and not just in L^1. The case with homogeneous boundary conditions enforced on only part of the boundary is studied in Licht [134]. $\quad\square$

23.4 Mollification with extension by zero

We now construct mollifying operators such that the mollified function is compactly supported in D. These operators can be used for instance to identify the kernel of the trace operators; see Exercise 23.9. We omit most of the proofs and refer the reader to Bonito et al. [33] and [96].

Since D is bounded, there are $\boldsymbol{x}_D \in \mathbb{R}^d$ and $r_D > 0$ such that $D \subset B(\boldsymbol{x}_D, r_D)$. Let $\mathcal{O} := B(\boldsymbol{x}_D, r_D) \backslash \overline{D}$ and notice that $\partial D \subset \partial \mathcal{O}$. Since $\mathcal{O}$ is an open, bounded, and Lipschitz set, there exists a vector field $\boldsymbol{k} \in \boldsymbol{C}^\infty(\mathbb{R}^d)$ that is globally transversal on $\mathcal{O}$ (i.e., there is $\gamma < 0$ such that $\boldsymbol{k}(\boldsymbol{x}) \cdot \boldsymbol{n}_\mathcal{O}(\boldsymbol{x}) \le \gamma$ for all $\boldsymbol{x} \in \partial \mathcal{O}$ where $\boldsymbol{n}_\mathcal{O}$ is the outer unit normal on $\partial \mathcal{O}$), and $\|\boldsymbol{k}(\boldsymbol{x})\|_{\ell^2} = 1$ for all $\boldsymbol{x} \in \partial \mathcal{O}$. Note that $\boldsymbol{k}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) \ge \gamma$ for all $\boldsymbol{x} \in \partial D$, where $\boldsymbol{n}$ denotes the outer unit normal on ∂D (which points toward the inside of $\mathcal{O}$). For all $\delta \in [0, 1]$, we define the mapping:

$$\boldsymbol{\vartheta}_\delta : \mathbb{R}^d \ni \boldsymbol{x} \longmapsto \boldsymbol{x} + \delta \boldsymbol{k}(\boldsymbol{x}) \in \mathbb{R}^d. \qquad (23.21)$$

Lemma 23.15 (Properties of $\boldsymbol{\vartheta}_\delta$). *The mapping $\boldsymbol{\vartheta}_\delta$ is of class C^∞ for all $\delta \in [0, 1]$, and there is c s.t. $\max_{\boldsymbol{x} \in D} \|D^k \boldsymbol{\vartheta}_\delta(\boldsymbol{x}) - D^k \boldsymbol{x}\|_{\ell^2} \le c\, \ell_D^{-k} \delta$ for all $\delta \in [0, 1]$ and all $k \in \mathbb{N}$. Moreover, there is $\zeta > 0$ s.t.*

$$\boldsymbol{\vartheta}_\delta(\overline{\mathcal{O}}) + B(\boldsymbol{0}, 2\delta\zeta) \subset \mathcal{O}, \qquad \forall \delta \in (0, 1]. \qquad (23.22)$$

We consider the following operator acting on functions in $L^1(D; \mathbb{R}^q)$:

$$(\mathcal{K}_{\delta,0}(f))(\boldsymbol{x}) := \int_{B(\boldsymbol{0},1)} \rho(\boldsymbol{y}) \mathbb{B}_\delta(\boldsymbol{x}) \widetilde{f}(\boldsymbol{\vartheta}_\delta(\boldsymbol{x}) + (\delta\zeta)\boldsymbol{y})\, \mathrm{d}\boldsymbol{y}, \quad \forall \boldsymbol{x} \in D, \quad (23.23)$$

where $\widetilde{f}$ denotes the zero-extension of f to $\mathbb{R}$. Here, $\mathbb{B}_\delta : D \to \mathbb{R}^{q \times q}$ is a smooth field. The examples we have in mind are $\mathbb{B}_\delta^{\mathrm{g}}(\boldsymbol{x}) := 1$, $\mathbb{B}_\delta^{\mathrm{c}}(\boldsymbol{x}) := \mathbb{K}_\delta^\mathsf{T}(\boldsymbol{x})$, $\mathbb{B}_\delta^{\mathrm{d}}(\boldsymbol{x}) := \det(\mathbb{K}_\delta(\boldsymbol{x})) \mathbb{K}_\delta^{-1}(\boldsymbol{x})$, and $\mathbb{B}_\delta^{\mathrm{b}}(\boldsymbol{x}) := \det(\mathbb{K}_\delta(\boldsymbol{x}))$, where $\mathbb{K}_\delta$ is the Jacobian matrix of $\boldsymbol{\vartheta}_\delta$ at $\boldsymbol{x} \in D$.

Lemma 23.16 (Smoothness, compact support). $\mathcal{K}_{\delta,0}(f) \in C_0^\infty(D; \mathbb{R}^q)$ *for all $f \in L^1(D; \mathbb{R}^q)$ and all $\delta \in (0, 1]$.*

Lemma 23.17 (Commuting). *The following diagrams commute:*

$$
\begin{array}{ccccccc}
\widetilde{Z}^{\mathrm{g},p}(D) & \xrightarrow{\nabla} & \widetilde{\boldsymbol{Z}}^{\mathrm{c},p}(D) & \xrightarrow{\nabla\times} & \widetilde{\boldsymbol{Z}}^{\mathrm{d},p}(D) & \xrightarrow{\nabla\cdot} & Z^{\mathrm{b},p}(D) \\
\downarrow{\mathcal{K}_{\delta,0}^{\mathrm{g}}} & & \downarrow{\mathcal{K}_{\delta,0}^{\mathrm{c}}} & & \downarrow{\mathcal{K}_{\delta,0}^{\mathrm{d}}} & & \downarrow{\mathcal{K}_{\delta,0}^{\mathrm{b}}} \\
C_0^\infty(D) & \xrightarrow{\nabla} & \boldsymbol{C}_0^\infty(D) & \xrightarrow{\nabla\times} & \boldsymbol{C}_0^\infty(D) & \xrightarrow{\nabla\cdot} & C_0^\infty(D)
\end{array}
$$

Theorem 23.18 (Convergence). (i) *There is $\tilde{\delta}_0 > 0$ s.t. for all $p \in [1, \infty]$, $(\mathcal{K}_{\delta,0})_{\delta \in [0, \tilde{\delta}_0]}$ is bounded in $\mathcal{L}(L^p; L^p)$. Moreover, we have*

$$\lim_{\delta \to 0} \|\mathcal{K}_{\delta,0}(f) - f\|_{L^p(D;\mathbb{R}^q)} = 0, \quad \forall f \in L^p(D;\mathbb{R}^q), \quad \forall p \in [1,\infty). \quad (23.24)$$

(ii) *Let $s \in (0,1]$. For all $p \in [1,\infty)$, there is c (c does not depend on $p \in [1,\infty]$ if $s = 1$) s.t. for all $f \in \widetilde{W}^{s,p}(D;\mathbb{R}^q)$, and all $\delta \in [0,\tilde{\delta}_0]$, the following holds true:*

$$\|\mathcal{K}_{\delta,0}(f) - f\|_{L^p(D;\mathbb{R}^q)} \le c\,\delta^s \ell_D^{-s} \|f\|_{\widetilde{W}^{s,p}(D;\mathbb{R}^q)}. \quad (23.25)$$

Corollary 23.19 (Derivatives). *The following holds true for all $p \in [1,\infty)$:*

(i) $\lim_{\delta \to 0} \|\nabla(\mathcal{K}_{\delta,0}^{\mathrm{g}}(f) - f)\|_{\boldsymbol{L}^p(D)} = 0, \ \forall f \in \widetilde{Z}^{\mathrm{g},p}(D)$.

(ii) $\lim_{\delta \to 0} \|\nabla \times (\mathcal{K}_{\delta,0}^{\mathrm{c}}(\boldsymbol{g}) - \boldsymbol{g})\|_{\boldsymbol{L}^p(D)} = 0, \ \forall \boldsymbol{g} \in \widetilde{\boldsymbol{Z}}^{\mathrm{c},p}(D)$.

(iii) $\lim_{\delta \to 0} \|\nabla \cdot (\mathcal{K}_{\delta,0}^{\mathrm{d}}(\boldsymbol{g}) - \boldsymbol{g})\|_{L^p(D)} = 0, \ \forall \boldsymbol{g} \in \widetilde{\boldsymbol{Z}}^{\mathrm{d},p}(D)$.

Remark 23.20 (Convergence rate on derivatives). Let $p \in [1,\infty)$ and $s > 0$. Assume first that $sp < 1$. Using Theorem 3.19 combined with Lemma 23.17 and Theorem 23.18, we infer that

$$\|\nabla(\mathcal{K}_{\delta,0}^{\mathrm{g}}(f) - f)\|_{\boldsymbol{L}^p(D)} \le c\,\delta^s \ell_D^{-s} \|\nabla f\|_{\boldsymbol{W}^{s,p}(D)},$$
$$\|\nabla \times (\mathcal{K}_{\delta,0}^{\mathrm{c}}(\boldsymbol{g}) - \boldsymbol{g})\|_{\boldsymbol{L}^p(D)} \le c\,\delta^s \ell_D^{-s} \|\nabla \times \boldsymbol{g}\|_{\boldsymbol{W}^{s,p}(D)},$$
$$\|\nabla \cdot (\mathcal{K}_{\delta,0}^{\mathrm{d}}(\boldsymbol{g}) - \boldsymbol{g})\|_{L^p(D)} \le c\,\delta^s \ell_D^{-s} \|\nabla \cdot \boldsymbol{g}\|_{W^{s,p}(D)},$$

for all $f \in L^p(D)$ with $\nabla f \in \boldsymbol{W}^{s,p}(D)$, all $\boldsymbol{g} \in \boldsymbol{L}^p(D)$ with $\nabla \times \boldsymbol{g} \in \boldsymbol{W}^{s,p}(D)$, and all $\boldsymbol{g} \in \boldsymbol{L}^p(D)$ with $\nabla \cdot \boldsymbol{g} \in W^{s,p}(D)$, respectively. If $sp > 1$, boundary conditions on the derivatives are needed for the above bounds to hold true, i.e., one needs also to assume that $\nabla f \in \boldsymbol{W}_0^{s,p}(D)$, $\nabla \times \boldsymbol{g} \in \boldsymbol{W}_0^{s,p}(D)$, and $\nabla \cdot \boldsymbol{g} \in W_0^{s,p}(D)$, respectively. □

Exercises

Exercise 23.1 (Star-shaped domain). Assume that $\boldsymbol{0} \in D$ and that D is star-shaped with respect to the ball $B(\boldsymbol{0},r)$ for some $r > 0$. Verify that the mapping $\boldsymbol{\varphi}_\delta : \mathbb{R}^d \to \mathbb{R}^d$ such that $\boldsymbol{\varphi}(\boldsymbol{x}) := (1-\delta)\boldsymbol{x}$ verifies the properties stated in Lemma 23.1.

Exercise 23.2 (Commuting). Prove Lemma 23.3. (*Hint*: use Lemma 9.6.)

Exercise 23.3 (Translation). Let $\lambda_0 > 0$. Assume that $\boldsymbol{\psi}_\lambda : D \to D$ is a diffeomorphism of class C^1 such that $\|\boldsymbol{\psi}_\lambda(\boldsymbol{x}) - \boldsymbol{x}\|_{\ell^2} \le c'\lambda$ and $\|D\boldsymbol{\psi}_\lambda(\boldsymbol{x}) - \mathbb{I}\|_{\ell^2} \le \frac{1}{2}$ for all $\boldsymbol{x} \in D$ and all $\lambda \in [0,\lambda_0]$. Assume also that $\boldsymbol{\mu}_{\lambda,t} : \boldsymbol{x} \mapsto \boldsymbol{x} + t(\boldsymbol{\psi}_\lambda(\boldsymbol{x}) - \boldsymbol{x})$ maps D into D for all $t \in [0,1]$ and all $\lambda \in [0,\lambda_0]$. Show that there is c such that $\|f \circ \boldsymbol{\psi}_\lambda - f\|_{L^p(D)} \le c\,\lambda \|\nabla f\|_{\boldsymbol{L}^p(D)}$ for all $\lambda \in [0,\lambda_0]$, all $f \in W^{1,p}(D)$, and all $p \in [1,\infty]$. (*Hint*: assume first that f is smooth, then use Remark 23.8.)

Exercise 23.4 (Approximation). (i) Prove (23.9) for $\mathcal{K}_\delta^{\mathrm{g}}$ with $s \in (0,1)$, $p \in [1,\infty)$. (ii) Prove the result for $s = 1$, $p \in [1,\infty]$. (*Hint*: use Exercise 23.3.) (iii) Prove (23.9) for $\mathcal{K}_\delta^{\mathrm{x}}$ for $\mathrm{x} \in \{\mathrm{c}, \mathrm{d}, \mathrm{b}\}$. (*Hint*: observe that $\mathcal{K}_\delta^{\mathrm{x}}(f) = \mathbb{K}^{\mathrm{x}}\mathcal{K}_\delta^{\mathrm{g}}(f)$.)

Exercise 23.5 (Preserving constants). Propose a definition of $\mathcal{K}_\delta$ that preserves constants and commutes with the differential operators. (*Hint*: start with $\check{\mathcal{K}}_\delta^{\mathrm{g}}(f) := \mathcal{K}_\delta^{\mathrm{g}}(f - \underline{f} - \underline{\nabla f}{\cdot}(\boldsymbol{x} - \boldsymbol{x}_D)) + \underline{f} + \underline{\nabla f}{\cdot}(\boldsymbol{x} - \boldsymbol{x}_D)$, $\underline{f}, \underline{\nabla f}$ denoting mean values over D and $\boldsymbol{x}_D$ the barycenter of D.)

Exercise 23.6 (Inverse inequality). Prove (23.19). (*Hint*: use (23.15b).)

Exercise 23.7 (Approximation with $\mathcal{J}_h^{\mathrm{c}}$). Let $r \in [0, k+1]$ and $p \in [1,\infty]$. Let $\boldsymbol{g} \in \boldsymbol{W}^{r,p}(D)$ be such that $\nabla{\times}\boldsymbol{g} \in \boldsymbol{W}^{r,p}(D)$. Prove that $\|\boldsymbol{g} - \mathcal{J}_h^{\mathrm{c}}(\boldsymbol{g})\|_{\boldsymbol{L}^p(D)} \le ch^r |\boldsymbol{g}|_{\boldsymbol{W}^{r,p}(D)}$ and $\|\nabla{\times}(\boldsymbol{g} - \mathcal{J}_h^{\mathrm{c}}(\boldsymbol{g}))\|_{\boldsymbol{L}^p(D)} \le ch^r |\nabla{\times}\boldsymbol{g}|_{\boldsymbol{W}^{r,p}(D)}$. (*Hint*: use Theorem 23.12.)

Exercise 23.8 (Best approximation in L^p). We propose an alternative proof of Corollary 22.9 on quasi-uniform meshes. Let $h \in \mathcal{H}$ be the mesh-size of $\mathcal{T}_h$ and set $\delta := \epsilon h$ in (23.4) with ϵ fixed small enough. Prove that $\inf_{f_h \in P_k(\mathcal{T}_h)} \|f - f_h\|_{L^p(D;\mathbb{R}^q)} \le ch^r \ell_D^{-r} \|f\|_{W^{r,p}(D;\mathbb{R}^q)}$ for all $r \in [0, k+1]$, all $p \in [1,\infty)$, and all $f \in W^{r,p}(D;\mathbb{R}^q)$. (*Hint*: admit as a fact that there is c, uniform, s.t. $\delta^s |\mathcal{K}_\delta(f)|_{W^{s,p}(D;\mathbb{R}^q)} \le c(\delta/\ell_D)^t \|f\|_{W^{t,p}(D;\mathbb{R}^q)}$ for all $s \ge t \ge 0$, then use $\mathcal{I}_h \circ \mathcal{K}_\delta$.)

Exercise 23.9 ($\boldsymbol{Z}_0^{\mathrm{c},p}(D) = \ker(\gamma^{\mathrm{c}})$). Let $p \in (1,\infty)$ and let $\boldsymbol{Z}_0^{\mathrm{c},p}(D) := \overline{\boldsymbol{\mathcal{C}}_0^\infty(D)}^{\boldsymbol{Z}^{\mathrm{c},p}(D)}$. We want to prove that $\boldsymbol{Z}_0^{\mathrm{c},p}(D) = \ker(\gamma^{\mathrm{c}})$ with the trace map $\gamma^{\mathrm{c}} : \boldsymbol{Z}^{\mathrm{c},p}(D) \to \boldsymbol{W}^{-\frac{1}{p},p}(\partial D)$ s.t. $\langle \gamma^{\mathrm{c}}(\boldsymbol{v}), \boldsymbol{l} \rangle := \int_D \boldsymbol{v}{\cdot}\nabla{\times}\boldsymbol{w}(\boldsymbol{l}) \, \mathrm{d}x - \int_D (\nabla{\times}\boldsymbol{v}){\cdot}\boldsymbol{w}(\boldsymbol{l}) \, \mathrm{d}x$ for all $\boldsymbol{v} \in \boldsymbol{Z}^{\mathrm{c},p}(D)$ and all $\boldsymbol{l} \in \boldsymbol{W}^{\frac{1}{p},p'}(\partial D)$, where $\boldsymbol{w}(\boldsymbol{l}) \in \boldsymbol{W}^{1,p}(D)$ is such that $\gamma^{\mathrm{d}}(\boldsymbol{w}(\boldsymbol{l})) = \boldsymbol{l}$ (see §4.3). (i) Show that $\boldsymbol{Z}_0^{\mathrm{c},p}(D) \subset \ker(\gamma^{\mathrm{c}})$. (*Hint*: $\mathcal{K}_\delta^{\mathrm{g}}(\boldsymbol{w}) \to \boldsymbol{w}$ in $\boldsymbol{W}^{1,p}(D)$ as $\delta \to 0$ for all $\boldsymbol{w} \in \boldsymbol{W}^{1,p}(D)$ and $\gamma^{\mathrm{g}} : W^{1,p}(D) \to W^{\frac{1}{p},p'}(\partial D)$ is surjective.) (ii) Let $\boldsymbol{v} \in \ker(\gamma^{\mathrm{c}})$. Show that $\nabla{\times}\widetilde{\boldsymbol{v}} = \widetilde{\nabla{\times}\boldsymbol{v}} \in \boldsymbol{L}^p(\mathbb{R}^d)$, where for every function $\boldsymbol{v}$ defined in D, $\widetilde{\boldsymbol{v}}$ denotes its zero-extension to $\mathbb{R}^d$. (iii) Show that $\ker(\gamma^{\mathrm{c}}) \subset \boldsymbol{Z}_0^{\mathrm{c},p}(D)$. (*Hint*: use the mollification operator $\mathcal{K}_{\delta,0}^{\mathrm{c}}$ defined in (23.23).)

Appendices

Appendix A

Banach and Hilbert spaces

The goal of this appendix is to recall basic results on Banach and Hilbert spaces. To stay general, we consider complex vector spaces, i.e., vector spaces over the field $\mathbb{C}$ of complex numbers. The case of real vector spaces is recovered by replacing the field $\mathbb{C}$ by $\mathbb{R}$, by removing the real part symbol $\Re(\cdot)$ and the complex conjugate symbol $\overline{\cdot}$, and by interpreting the symbol $|\cdot|$ as the absolute value instead of the modulus.

A.1 Banach spaces

Let V be a complex vector space.

Definition A.1 (Norm). *A* norm *on V is a map $\|\cdot\|_V : V \to \mathbb{R}_+ := [0, \infty)$ satisfying the following three properties:*

(i) *Definiteness:* $[\, \|v\|_V = 0 \,] \iff [\, v = 0 \,].$
(ii) *1-homogeneity:* $\|\lambda v\|_V = |\lambda| \, \|v\|_V$ *for all $\lambda \in \mathbb{C}$ and all $v \in V$.*
(iii) *Triangle inequality:* $\|v + w\|_V \le \|v\|_V + \|w\|_V$ *for all $v, w \in V$.*

For every norm $\|\cdot\|_V : V \to \mathbb{R}_+ := [0, \infty)$, the function $d(x, y) := \|x - y\|_V$, for all $x, y \in V$, is a metric (or distance).

Remark A.2 (Definiteness). Item (i) can be slightly relaxed by requiring only that $[\, \|v\|_V = 0 \,] \implies [\, v = 0 \,]$, since the 1-homogeneity implies that $[\, v = 0 \,] \implies [\, \|v\|_V = 0 \,]$. $\qquad\square$

Definition A.3 (Seminorm). *A* seminorm *on V is a map $|\cdot|_V : V \to \mathbb{R}_+$ satisfying only the statements* (ii) *and* (iii) *above, i.e., 1-homogeneity and the triangle inequality.*

Definition A.4 (Banach space). *A vector space V equipped with a norm $\|\cdot\|_V$ is called* Banach space *if every Cauchy sequence in V has a limit in V.*

Definition A.5 (Equivalent norms). *Two norms* $\|\cdot\|_{V,1}$ *and* $\|\cdot\|_{V,2}$ *are said to be* equivalent *on V if there exists a positive real number c such that*

$$c\,\|v\|_{V,2} \le \|v\|_{V,1} \le c^{-1}\,\|v\|_{V,2}, \qquad \forall v \in V. \tag{A.1}$$

Whenever (A.1) *holds true, V is a Banach space for the norm* $\|\cdot\|_{V,1}$ *if and only if it is a Banach space for the norm* $\|\cdot\|_{V,2}$.

Remark A.6 (Finite dimension). If V is finite-dimensional, all the norms in V are equivalent. This result is false in infinite-dimensional vector spaces. Actually, the unit ball in V is a compact set (for the norm topology) if and only if V is finite-dimensional; see Brezis [48, Thm. 6.5], Lax [131, §5.2]. □

A.2 Bounded linear maps and duality

Definition A.7 (Linear, antilinear map). *Let V, W be complex vector spaces. A map $A : V \to W$ is said to be* linear *if $A(v_1 + v_2) = A(v_1) + A(v_2)$ for all $v_1, v_2 \in V$ and $A(\lambda v) = \lambda A(v)$ for all $\lambda \in \mathbb{C}$ and all $v \in V$, and it is said to be* antilinear *if $A(v_1 + v_2) = A(v_1) + A(v_2)$ for all $v_1, v_2 \in V$ and $A(\lambda v) = \overline{\lambda} A(v)$ for all $\lambda \in \mathbb{C}$ and all $v \in V$.*

Definition A.8 (Bounded (anti)linear map). *Assume that V and W are equipped with norms $\|\cdot\|_V$ and $\|\cdot\|_W$, respectively. The (anti)linear map $A : V \to W$ is said to be* bounded *or* continuous *if*

$$\|A\|_{\mathcal{L}(V;W)} := \sup_{v \in V} \frac{\|A(v)\|_W}{\|v\|_V} < \infty. \tag{A.2}$$

In this book, we systematicaly abuse the notation by implicitly assuming that the argument in this type of supremum is nonzero. Bounded (anti)linear maps in Banach spaces are called operators.

The complex vector space composed of the bounded linear maps from V to W is denoted by $\mathcal{L}(V;W)$. One readily verifies that the map $\|\cdot\|_{\mathcal{L}(V;W)}$ defined in (A.2) is indeed a norm on $\mathcal{L}(V;W)$.

Proposition A.9 (Banach space). *Assume that W is a Banach space. Then $\mathcal{L}(V;W)$ equipped with the norm* (A.2) *is also a Banach space. The same statement holds true for the complex vector space composed of all the bounded antilinear maps from V to W.*

Proof. See Rudin [170, p. 87], Yosida [202, p. 111]. □

Example A.10 (Continuous embedding). Assume that $V \subset W$ and that there is a real number c such that $\|v\|_W \le c\|v\|_V$ for all $v \in V$. This means that the embedding of V into W is continuous. We say that V is continuously embedded into W, and we write $V \hookrightarrow W$. □

The dual of a real Banach space V is composed of the bounded linear maps from V to $\mathbb{R}$. The same definition can be adopted if V is a complex space, but to stay consistent with the formalism considered in the weak formulation of complex-valued PDEs, we define the dual space as being composed of bounded antilinear maps from V to $\mathbb{C}$.

Definition A.11 (Dual space). *Let V be a complex vector space. The* dual space *of V is denoted by V' and is composed of the bounded antilinear maps from V to $\mathbb{C}$. An element $A \in V'$ is called* bounded antilinear form, *and its action on an element $v \in V$ is denoted either by $A(v)$ or $\langle A, v \rangle_{V',V}$.*

Owing to Proposition A.9, V' is a Banach space with the norm

$$\|A\|_{V'} = \sup_{v \in V} \frac{|A(v)|}{\|v\|_V} = \sup_{v \in V} \frac{|\langle A, v \rangle_{V',V}|}{\|v\|_V}, \qquad \forall A \in V'. \tag{A.3}$$

Remark A.12 (Linear vs. antilinear form). If $A : V \to \mathbb{C}$ is an antilinear form, then $\overline{A}$ (defined by $\overline{A}(v) := \overline{A(v)} \in \mathbb{C}$ for all $v \in V$) is a linear form. $\square$

A.3 Hilbert spaces

Let V be a complex vector space.

Definition A.13 (Inner product). *An* inner product *on V is a map $(\cdot, \cdot)_V : V \times V \to \mathbb{C}$ satisfying the following three properties:* (i) *Sesquilinearity (the prefix sesqui means one and a half): $(\cdot, w)_V$ is a linear map for all fixed $w \in V$, whereas $(v, \cdot)_V$ is an antilinear map for all fixed $v \in V$. If V is a real vector space, the inner product is a bilinear map (i.e., it is linear in both of its arguments).* (ii) *Hermitian symmetry: $(v, w)_V = \overline{(w, v)_V}$ for all $v, w \in V$.* (iii) *Positive definiteness: $(v, v)_V \geq 0$ for all $v \in V$ and $[(v, v)_V = 0] \iff [v = 0]$. (Notice that $(v, v)_V$ is always real owing to the Hermitian symmetry and that $(0, \cdot)_V = (\cdot, 0)_V = 0$ owing to sesquilinearity.)*

Proposition A.14 (Cauchy–Schwarz). *Let $(\cdot, \cdot)_V$ be an inner product on V. By setting*

$$\|v\|_V := (v, v)_V^{\frac{1}{2}}, \qquad \forall v \in V, \tag{A.4}$$

one defines a norm on V. This norm is said to be induced by the inner product. Moreover, we have the Cauchy–Schwarz inequality

$$|(v, w)_V| \leq \|v\|_V \|w\|_V, \qquad \forall v, w \in V. \tag{A.5}$$

Definition A.15 (Hilbert space). *A* Hilbert space *V is an inner product space that is complete with respect to the induced norm (and is therefore a Banach space).*

Theorem A.16 (Riesz–Fréchet). *Let V be a complex Hilbert space. For all $A \in V'$, there exists a unique $v \in V$ s.t. $(v,w)_V = \langle A, w \rangle_{V',V}$ for all $w \in V$, and we have $\|v\|_V = \|A\|_{V'}$.*

Proof. See Brezis [48, Thm. 5.5], Lax [131, p. 56], Yosida [202, p. 90].   □

A.4 Compact operators

Definition A.17 (Compact operator). *Let V, W be two complex Banach spaces. The operator $T \in \mathcal{L}(V; W)$ is said to be* compact *if from every bounded sequence $(v_n)_{n \in \mathbb{N}}$ in V, one can extract a subsequence $(v_{n_k})_{k \in \mathbb{N}}$ such that the sequence $(T(v_{n_k}))_{k \in \mathbb{N}}$ converges in W. Equivalently T is said to be compact if T maps the unit ball in V into a relatively compact set in W (that is, a set whose closure in W is compact).*

Example A.18 (Compact embedding). Assume that $V \subset W$ and that the embedding of V into W is compact. Then from every bounded sequence $(v_n)_{n \in \mathbb{N}}$ in V, one can extract a subsequence that converges in W.   □

Proposition A.19 (Composition). *Let W, X, Y, Z be four Banach spaces and let $A \in \mathcal{L}(Z; Y)$, $K \in \mathcal{L}(Y; X)$, $B \in \mathcal{L}(X; W)$ be three operators. Assume that K is compact. Then the operator $B \circ K \circ A$ is compact.*

The following compactness result is used at several instances in this book. The reader is referred to Tartar [189, Lem. 11.1] and Girault and Raviart [107, Thm. 2.1, p. 18] for a slightly more general statement and references.

Lemma A.20 (Peetre–Tartar). *Let X, Y, Z be three Banach spaces. Let $A \in \mathcal{L}(X; Y)$ be an injective operator and let $T \in \mathcal{L}(X; Z)$ be a compact operator. Assume that there is $c > 0$ such that $c\|x\|_X \leq \|A(x)\|_Y + \|T(x)\|_Z$ for all $x \in X$. Then there is $\alpha > 0$ such that*

$$\alpha\|x\|_X \leq \|A(x)\|_Y, \qquad \forall x \in X. \tag{A.6}$$

Proof. We prove (A.6) by contradiction. Assume that there is a sequence $(x_n)_{n \in \mathbb{N}}$ of X s.t. $\|x_n\|_X = 1$ and $\|A(x_n)\|_Y$ converges to zero as $n \to \infty$. Since T is compact and the sequence $(x_n)_{n \in \mathbb{N}}$ is bounded, there is a subsequence $(x_{n_k})_{k \in \mathbb{N}}$ s.t. $(T(x_{n_k}))_{k \in \mathbb{N}}$ is a Cauchy sequence in Z. Owing to the inequality

$$\alpha\|x_{n_k} - x_{m_k}\|_X \leq \|A(x_{n_k}) - A(x_{m_k})\|_Y + \|T(x_{n_k}) - T(x_{m_k})\|_Z,$$

$(x_{n_k})_{k \in \mathbb{N}}$ is a Cauchy sequence in X. Let x be its limit, so that $\|x\|_X = 1$. The boundedness of A implies $A(x_{n_k}) \to A(x)$, and $A(x) = 0$ since $A(x_{n_k}) \to 0$. Since A is injective, $x = 0$, which contradicts $\|x\|_X = 1$.   □

We finish this section with a striking property of compact operators.

Theorem A.21 (Approximability and compactness). *Let V, W be Banach spaces. If there exists a sequence $(T_n)_{n\in\mathbb{N}}$ of operators in $\mathcal{L}(V;W)$ of finite rank (i.e., $\dim(\mathrm{im}(T_n)) < \infty$ for all $n \in \mathbb{N}$) such that $\lim_{n\to\infty}\|T - T_n\|_{\mathcal{L}(V;W)} = 0$, then T is compact. Conversely, if W is a Hilbert space and $T \in \mathcal{L}(V;W)$ is a compact operator, then there exists a sequence of operators in $\mathcal{L}(V;W)$ of finite rank, $(T_n)_{n\in\mathbb{N}}$, such that $\lim_{n\to\infty}\|T - T_n\|_{\mathcal{L}(V;W)} = 0$.*

Proof. See Brezis [48, pp. 157–158]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

A.5 Interpolation between Banach spaces

Interpolation between Banach spaces is often used to combine known results to derive new results that could be difficult to obtain directly. An important application is the derivation of functional inequalities in fractional-order Sobolev spaces (see §2.2.2). There are many interpolation methods; see, e.g., Bergh and Löfström [18], Tartar [189], and the references therein. For simplicity we focus on the real interpolation K-method; see [18, §3.1] and [189, Chap. 22].

Let V_0 and V_1 be two normed vector spaces that are continuously embedded into a common topological vector space $\mathcal{V}$. Then $V_0 + V_1$ is a normed vector space with the (canonical) norm $\|v\|_{V_0+V_1} := \inf_{v=v_0+v_1}(\|v_0\|_{V_0} + \|v_1\|_{V_1})$. Moreover, if V_0 and V_1 are Banach spaces, then $V_0 + V_1$ is also a Banach space; see [18, Lem. 2.3.1]. For all $v \in V_0 + V_1$ and all $t > 0$, we define

$$K(t, v) := \inf_{v=v_0+v_1}(\|v_0\|_{V_0} + t\|v_1\|_{V_1}). \tag{A.7}$$

For all $t > 0$, $v \mapsto K(t, v)$ defines a norm on $V_0 + V_1$ that is equivalent to the canonical norm. One can verify that the function $t \mapsto K(t, v)$ is nondecreasing and concave (and therefore continuous) and that the function $t \mapsto \frac{1}{t}K(t, v)$ is increasing.

Definition A.22 (Interpolated space). *Let $\theta \in (0, 1)$ and let $p \in [1, \infty]$. The interpolated space $[V_0, V_1]_{\theta,p}$ is defined to be the vector space*

$$[V_0, V_1]_{\theta,p} := \{v \in V_0 + V_1 \mid \|t^{-\theta}K(t, v)\|_{L^p(\mathbb{R}_+; \frac{dt}{t})} < \infty\}, \tag{A.8}$$

where $\|\varphi\|_{L^p(\mathbb{R}_+; \frac{dt}{t})} := \left(\int_0^\infty |\varphi(t)|^p \frac{dt}{t}\right)^{\frac{1}{p}}$ for all $p \in [1, \infty)$ and $\|\varphi\|_{L^\infty(\mathbb{R}_+; \frac{dt}{t})} := \sup_{0<t<\infty}|\varphi(t)|$. This space is equipped with the norm

$$\|v\|_{[V_0,V_1]_{\theta,p}} := \|t^{-\theta}K(t, v)\|_{L^p(\mathbb{R}_+; \frac{dt}{t})}. \tag{A.9}$$

If V_0 and V_1 are Banach spaces, so is $[V_0, V_1]_{\theta,p}$.

Remark A.23 (Value for θ). Since $K(t, v) \geq \min(1, t)\|v\|_{V_0+V_1}$, the space $[V_0, V_1]_{\theta,p}$ reduces to $\{0\}$ if $t^{-\theta}\min(1, t) \notin L^p(\mathbb{R}_+; \frac{dt}{t})$. In particular, $[V_0, V_1]_{\theta,p}$ is trivial if $\theta \in \{0, 1\}$ and $p < \infty$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark A.24 (Gagliardo set). The function $t \mapsto K(t, v)$ has a simple geometric interpretation. Introducing the Gagliardo set $G(v) := \{(x_0, x_1) \in \mathbb{R}^2 \mid v = v_0 + v_1 \text{ with } \|v_0\|_{V_0} \leq x_0 \text{ and } \|v_1\|_{V_1} \leq x_1\}$, one can verify that $G(v)$ is convex and that $K(t, v) = \inf_{v \in \partial G(v)}(x_0 + t x_1)$, so that the map $t \mapsto K(t, v)$ is one way to explore the boundary of $G(v)$; see [18, p. 39]. $\square$

Remark A.25 (Intersection). The vector space $V_0 \cap V_1$ can be equipped with the (canonical) norm $\|v\|_{V_0 \cap V_1} := \max(\|v\|_{V_0}, \|v\|_{V_1})$. One can verify that $K(t, v) \leq \min(1, t)\|v\|_{V_0 \cap V_1}$ for all $v \in V_0 \cap V_1$, which implies the boundedness of the embedding $V_0 \cap V_1 \hookrightarrow [V_0, V_1]_{\theta, p}$ for all $\theta \in (0, 1)$ and all $p \in [1, \infty]$. Hence, if $V_0 \subset V_1$, then $V_0 \hookrightarrow [V_0, V_1]_{\theta, p}$. $\square$

Lemma A.26 (Continuous embedding). *Let $\theta \in (0, 1)$ and $p, q \in [1, \infty]$ with $p \leq q$. Then we have $[V_0, V_1]_{\theta, p} \hookrightarrow [V_0, V_1]_{\theta, q}$.*

Theorem A.27 (Riesz–Thorin, interpolation of operators). *Let $A : V_0 + V_1 \rightarrow W_0 + W_1$ be a linear operator that maps V_0 and V_1 boundedly to W_0 and W_1, respectively. Then for all $\theta \in (0, 1)$ and all $p \in [1, \infty]$, A maps $[V_0, V_1]_{\theta, p}$ boundedly to $[W_0, W_1]_{\theta, p}$. Moreover, we have*

$$\|A\|_{\mathcal{L}([V_0, V_1]_{\theta, p}; [W_0, W_1]_{\theta, p})} \leq \|A\|_{\mathcal{L}(V_0; W_0)}^{1-\theta} \|A\|_{\mathcal{L}(V_1; W_1)}^{\theta}. \tag{A.10}$$

Proof. See [189, Lem. 22.3]. $\square$

Theorem A.28 (Lions–Peetre, reiteration). *Let $\theta_0, \theta_1 \in (0, 1)$ with $\theta_0 \neq \theta_1$. Assume that $[V_0, V_1]_{\theta_0, 1} \hookrightarrow W_0 \hookrightarrow [V_0, V_1]_{\theta_0, \infty}$ and $[V_0, V_1]_{\theta_1, 1} \hookrightarrow W_1 \hookrightarrow [V_0, V_1]_{\theta_1, \infty}$. Then for all $\theta \in (0, 1)$ and all $p \in [1, \infty]$, $[W_0, W_1]_{\theta, p} = [V_0, V_1]_{\eta, p}$ with equivalent norms, where $\eta := (1 - \theta)\theta_0 + \theta\theta_1$.*

Proof. See Tartar [189, Thm. 26.2]. $\square$

Theorem A.29 (Lions–Peetre, extension). *Let V_0, V_1, F be three Banach spaces. Let $A \in \mathcal{L}(V_0 \cap V_1; F)$. Then A extends into a linear continuous map from $[V_0, V_1]_{\theta, 1; J}$ to F iff*

$$\exists c < \infty, \qquad \|A(v)\|_F \leq c\|v\|_{V_0}^{1-\theta} \|v\|_{V_1}^{\theta}, \qquad \forall v \in V_0 \cap V_1. \tag{A.11}$$

Proof. See [189, Lem. 25.3]. $\square$

Theorem A.30 (Interpolation of dual spaces). *Let $\theta \in (0, 1)$ and $p \in [1, \infty)$. Then $[V_0, V_1]_{\theta, p}' = [V_1', V_0']_{1-\theta, p'}$ where $p' := \frac{p}{p-1}$ (with the convention that $p' := \infty$ if $p = 1$).*

Proof. See [189, Lem. 41.3] or Bergh and Löfström [18, Thm. 3.7.1]. $\square$

Appendix B

Differential calculus

This appendix briefly overviews some basic facts of differential calculus concerning Fréchet derivatives and their link to the notions of gradient, Jacobian matrix, and Hessian matrix.

B.1 Fréchet derivative

Let V, W be Banach spaces and let U be an open set in V. The space $C^0(U; W)$ consists of those functions $f : U \to W$ that are continuous in U.

Definition B.1 (Fréchet derivative). *Let $f \in C^0(U; W)$. We say that f is Fréchet differentiable (or differentiable) at $x \in U$ if there is a bounded linear operator $Df(x) \in \mathcal{L}(V; W)$ such that*

$$\lim_{h \to 0} \frac{\|f(x + h) - f(x) - Df(x)(h)\|_W}{\|h\|_V} = 0. \tag{B.1}$$

The operator $Df(x)$ is called Fréchet derivative of f at x. If the map $Df : U \to \mathcal{L}(V; W)$ is continuous, we say that f is of class C^1 in U, and we write $f \in C^1(U; W)$.

The above process can be repeated to define $D(Df)(x)$. For an integer $n \geq 2$, let us denote by $\mathcal{M}_n(V, \ldots, V; W)$ the space spanned by the multilinear maps from $V \times \ldots \times V$ (n times) to W. Upon identifying $\mathcal{L}(V; \mathcal{L}(V; W))$ with $\mathcal{M}_2(V, V; W)$ and setting $D^2 f(x) := D(Df)(x)$, we have $D^2 f(x) \in \mathcal{M}_2(V, V; W)$. The n-th Fréchet derivative of f at x is defined recursively as being the Fréchet derivative of $D^{n-1} f$ at x for all $n \geq 2$, that is,

$$D^n f(x) \in \mathcal{M}_n(\underbrace{V, \ldots, V}_{n \text{ times}}; W).$$

If $D^n f : U \to \mathcal{M}_n(V, \ldots, V; W)$ is continuous, we write $f \in C^n(U; W)$.

Let us restate some elementary properties of the Fréchet derivative (for the chain rule, the reader is referred, e.g., to Cartan [64, pp. 28-96], Ciarlet and Raviart [78, p. 227]). For an integer $n \geq 1$, $\mathcal{S}_n$ denotes the set of permutations of the integer set $\{1{:}n\} := \{1, \ldots, n\}$.

Lemma B.2 (Leibniz product rule). *Let $f \in C^n(U; W_1)$, $g \in C^n(U; W_2)$, $n \geq 1$, and let $b : W_1 \times W_2 \to W_3$ be a bilinear map, where U is an open set in V and V, W_1, W_2 are Banach spaces. The following holds true for all $x \in U$:*

$$D^n b(f(x), g(x)) = \sum_{l \in \{0:n\}} \binom{n}{l} b(D^{n-l} f(x), D^l g(x)), \qquad \forall x \in U. \quad \text{(B.2)}$$

Theorem B.3 (Symmetry). *Let V, W be Banach spaces. Let $n \geq 2$ and let $\mathcal{S}_n$ be the set of the permutations of $\{1{:}n\}$. Let $f \in C^n(U; W)$ where U is an open set in V. Then $D^n f$ is symmetric, i.e.,*

$$D^n f(x)(h_1, \ldots, h_n) = D^n f(x)(h_{\sigma(1)}, \ldots, h_{\sigma(n)}), \qquad \forall x \in U, \quad \text{(B.3)}$$

for all $\sigma \in \mathcal{S}_n$ and all $h_1, \ldots, h_n \in V$.

Theorem B.3 with $n := 2$ is often called *Clairaut or Schwarz theorem* in the literature.

Lemma B.4 (Chain rule). *Let $f \in C^n(U; W_1)$ and $g \in C^n(W_1; W_2)$, $n \geq 1$, where V, W_1, W_2 are Banach spaces and let U be an open set in V. Then we have*

$$D^n(g \circ f)(x)(h_1, \ldots, h_n) = \sum_{\sigma \in \mathcal{S}_n} \sum_{l \in \{1:n\}} \sum_{1 \leq r_1 + \ldots + r_l = n} \frac{1}{l! r_1! \ldots r_l!} \times \quad \text{(B.4)}$$

$$D^l g(f(x))(D^{r_1} f(x)(h_{\sigma(1)}, \ldots, h_{\sigma(s_1)}), \ldots, D^{r_l} f(x)(h_{\sigma(s_{l-1}+1)}, \ldots, h_{\sigma(n)})).$$

with $s_0 := 0$, $s_1 := r_1$, $s_2 := r_1 + r_2$, $\ldots$, $s_{l-1} := r_1 + \ldots + r_{l-1}$.

The identity (B.4) is often called *Faà di Bruno's formula* in the literature.

Example B.5 For $n = 1$, (B.4) yields

$$D(f \circ g)(x)(h) = Dg(f(x))(Df(x)(h)),$$

i.e., $D(f \circ g)(x) = Dg(f(x)) \circ Df(x)$. $\qquad \qquad \square$

B.2 Vector and matrix representation

Assume that $V = \mathbb{R}^d$ and let $\{e_1, \ldots, e_d\}$ be the canonical Cartesian basis of $\mathbb{R}^d$. (We use boldface notation for elements in V). Let U be an open set

of $\mathbb{R}^d$. We say that f is differentiable in the direction $\boldsymbol{e}_i$ at $\boldsymbol{x} \in U$ if there is an element in W, say $\partial_i f(\boldsymbol{x}) \in W$, such that $\lim_{t \to 0} |t|^{-1}(f(\boldsymbol{x} + t\boldsymbol{e}_i) - f(\boldsymbol{x}) - t\partial_i f(\boldsymbol{x})) = 0$. If f is Fréchet differentiable at $\boldsymbol{x}$, it is differentiable along any direction $\boldsymbol{e}_i$ for $i \in \{1{:}d\}$ (the converse is not necessarily true), and we have

$$\partial_i f(\boldsymbol{x}) = Df(\boldsymbol{x})(\boldsymbol{e}_i). \tag{B.5}$$

More generally, let $\alpha := (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ be a multi-index. The number $|\alpha| := \alpha_1 + \dots + \alpha_d$ is called the *length* of α. For all $f \in C^n(U; W)$ and every multi-index α s.t. $|\alpha| = n$, we write

$$\partial^\alpha f(\boldsymbol{x}) := \underbrace{\partial_1 \dots \partial_1}_{\alpha_1 \text{ times}} \dots \underbrace{\partial_d \dots \partial_d}_{\alpha_d \text{ times}} f(\boldsymbol{x}) = D^n f(\boldsymbol{x})(\underbrace{\boldsymbol{e}_1, \dots, \boldsymbol{e}_1}_{\alpha_1 \text{ times}}, \dots, \underbrace{\boldsymbol{e}_d, \dots, \boldsymbol{e}_d}_{\alpha_d \text{ times}}),$$
$$\tag{B.6}$$

and the order of the partial derivatives is irrelevant owing to Theorem B.3.

Let us finally assume that W is also finite-dimensional, e.g., $W := \mathbb{R}^m$ or $W := \mathbb{C}^m$. For $m = 1$, we adopt the convention that the *gradient* of f at $\boldsymbol{x}$, say $\nabla f(\boldsymbol{x})$, is the column vector with components

$$(\nabla f(\boldsymbol{x}))_i := \partial_i f(\boldsymbol{x}), \qquad \forall i \in \{1{:}d\}. \tag{B.7}$$

Identifying $\boldsymbol{h}$ with a column vector in $\mathbb{R}^d$, the action of $Df(\boldsymbol{x})$ is such that the following identities hold true for all $\boldsymbol{h} = \sum_{i \in \{1:d\}} h_i \boldsymbol{e}_i \in \mathbb{R}^d$:

$$Df(\boldsymbol{x})(\boldsymbol{h}) = \sum_{i \in \{1:d\}} \partial_i f(\boldsymbol{x}) h_i = (\nabla f(\boldsymbol{x}), \boldsymbol{h})_{\ell^2(\mathbb{R}^d)}, \tag{B.8}$$

where $(\cdot, \cdot)_{\ell^2(\mathbb{R}^d)}$ denotes the Euclidean product in $\mathbb{R}^d$. Assuming that $m \geq 2$, consider a basis of $\mathbb{R}^m$ and define the $m{\times}d$ *Jacobian matrix* of f at $\boldsymbol{x}$, say $\mathbb{J}_f(\boldsymbol{x})$, by its entries

$$(\mathbb{J}_f(\boldsymbol{x}))_{ij} := \partial_j f_i(\boldsymbol{x}), \qquad \forall i, j \in \{1{:}d\}, \tag{B.9}$$

where f_i is the i-th component of f in the chosen basis. Then we have

$$Df(\boldsymbol{x})(\boldsymbol{h}) = \mathbb{J}_f(\boldsymbol{x})\boldsymbol{h}, \qquad \forall \boldsymbol{h} \in \mathbb{R}^d. \tag{B.10}$$

Note that when $m = 1$, $\mathbb{J}_f(\boldsymbol{x})$ is the transpose of the gradient of f at $\boldsymbol{x}$, i.e., $\mathbb{J}_f(\boldsymbol{x}) = (\nabla f(\boldsymbol{x}))^\mathsf{T}$. For a scalar-valued function f, one can introduce the (symmetric) $d{\times}d$ *Hessian matrix* at $\boldsymbol{x}$, say $H_f(\boldsymbol{x})$, with entries

$$(H_f)_{ij} := \partial_{ij} f(\boldsymbol{x}), \qquad \forall i, j \in \{1{:}d\}, \tag{B.11}$$

leading to the following representation:

$$D^2 f(\boldsymbol{x})(\boldsymbol{h}_1, \boldsymbol{h}_2) = \boldsymbol{h}_1^\mathsf{T} H_f(\boldsymbol{x}) \boldsymbol{h}_2 = \boldsymbol{h}_2^\mathsf{T} H_f(\boldsymbol{x}) \boldsymbol{h}_1, \qquad \forall \boldsymbol{h}_1, \boldsymbol{h}_2 \in \mathbb{R}^d. \tag{B.12}$$

References

[1] M. Abramowitz, I. Stegun, *Handbook of Mathematical Functions*, 9th edn. (Dover Publications Inc., New York, NY, 1972). Page 61

[2] G. Acosta, R.G. Durán, An optimal Poincaré inequality in L^1 for convex domains. Proc. Amer. Math. Soc. **132**(1), 195–202 (2004). Page 34

[3] R.A. Adams, J.J.F. Fournier, *Sobolev Spaces*, vol. 140, 2nd edn. Pure and Applied Mathematics (Elsevier/Academic Press, Amsterdam, The Netherlands, 2003). Pages 3, 9, 10, 11, 19, 22, 25, 42

[4] R. Agelek, M. Anderson, W. Bangerth, W. Barth, On orienting edges of unstructured two- and three-dimensional meshes. ACM Trans. Math. Software **44**, 5/1–22 (2017). Page 118

[5] M. Ainsworth, A posteriori error estimation for discontinuous Galerkin finite element approximation. SIAM J. Numer. Anal. **45**(4), 1777–1798 (2007). Page 144

[6] M. Ainsworth, J. Coyle, Hierarchic finite element bases on unstructured tetrahedral meshes. Internat. J. Numer. Methods Engrg. **58**(14), 2103–2130 (2003). Pages 81, 118, 170, 184

[7] M. Ainsworth, G. Andriamaro, O. Davydov, Bernstein-Bézier finite elements of arbitrary order and optimal assembly procedures. SIAM J. Sci. Comput. **33**(6), 3087–3109 (2011). Page 81

[8] A. Alonso, A. Valli, An optimal domain decomposition preconditioner for low-frequency time-harmonic Maxwell equations. Math. Comp. **68**(226), 607–631 (1999). Page 211

[9] C. Amrouche, C. Bernardi, M. Dauge, V. Girault, Vector potentials in three-dimensional non-smooth domains. Math. Methods Appl. Sci. **21**(9), 823–864 (1998). Page 211

[10] D.N. Arnold, G. Awanou, The serendipity family of finite elements. Found. Comput. Math. **11**(3), 337–344 (2011). Page 71

[11] D.N. Arnold, R.S. Falk, R. Winther, Finite element exterior calculus, homological techniques, and applications. Acta Numer. **15**, 1–155 (2006). Pages 197, 198, 292

[12] D.N. Arnold, R.S. Falk, R. Winther, Finite element exterior calculus: from Hodge theory to numerical stability. Bull. Amer. Math. Soc. (N.S.) **47**(2), 281–354 (2010). Page 197

[13] A. Axelsson, A. McIntosh, Hodge decompositions on weakly Lipschitz domains, in *Advances in Analysis and Geometry. New Developments Using Clifford Algebras*, ed. by T. Qian, T. Hempfling, A. McIntosh, F. Sommen, Trends in Mathematics (Birkhäuser, Basel, Switzerland, 2004), pp. 3–29. Page 29

[14] I. Babuška, Error-bounds for finite element method. Numer. Math. **16**, 322–333 (1970/1971). Page v

[15] R.E. Bank, H. Yserentant, On the H^1-stability of the L_2-projection onto finite element spaces. Numer. Math. **126**, 361–381 (2014). Page 285

[16] R.G. Bartle, *A Modern Theory of Integration*, vol. 32. Graduate Studies in Mathematics (American Mathematical Society, Providence, RI, 2001). Pages 3, 9, 10, 11

[17] M. Bebendorf, A note on the Poincaré inequality for convex domains. Z. Anal. Anwendungen **22**(4), 751–756 (2003). Page 34

[18] J. Bergh, J. Löfström. *Interpolation Spaces. An Introduction*, Grundlehren der mathematischen Wissenschaften (Springer, Berlin-New York, 1976). No. 223. Pages 12, 307, 308

[19] A. Bermúdez, R. Rodríguez, P. Salgado, Numerical treatment of realistic boundary conditions for the eddy current problem in an electrode via Lagrange multipliers. Math. Comp. **74**(249), 123–151 (2005). Page 211

[20] C. Bernardi, Optimal finite element interpolation on curved domains. SIAM J. Numer. Anal. **26**, 1212–1240 (1989). Page 154

[21] C. Bernardi, V. Girault, A local regularization operator for triangular and quadrilateral finite elements. SIAM J. Numer. Anal. **35**(5), 1893–1916 (1998). Page 279

[22] C. Bernardi, Y. Maday, *Spectral methods, in Handbook of Numerical Analysis*, vol. V (North-Holland, Amsterdam, The Netherlands, 1997), pp. 209–485. Page 141

[23] S. Bertoluzza, The discrete commutator property of approximation spaces. C. R. Acad. Sci. Paris, Sér. I **329**(12), 1097–1102 (1999). Page 286

[24] W. Blaschke, *Kreis und Kugel* (Verlag von Veit & Comp, Leipzig, Germany, 1916). Page 36

[25] A. Blouza, H. Le Dret, An up-to-the-boundary version of Friedrichs's lemma and applications to the linear Koiter shell model. SIAM J. Math. Anal. **33**(4), 877–895 (2001). Page 290

[26] M.G. Blyth, H. Luo, C. Pozrikidis, A comparison of interpolation grids over the triangle or the tetrahedron. J. Engrg. Math. **56**(3), 263–272 (2006). Page 81

[27] P. Bochev, J.M. Hyman, Principles of mimetic discretizations of differential operators, in *Compatible Spatial Discretization, vol. 142, The IMA Volumes in Mathematics and its Applications*, ed. by D. Arnold, P. Bochev, R. Lehoucq, R.A. Nicolaides, M. Shashkov (Springer, Berlin, Germany, 2005), pp. 89–120. Page 112

[28] D. Boffi, L. Gastaldi, Interpolation estimates for edge finite elements and application to band gap computation. Appl. Numer. Math. **56**(10–11), 1283–1292 (2006). Page 211

[29] D. Boffi, F. Brezzi, M. Fortin, *Mixed Finite Element Methods and Applications*, vol. 44. Springer Series in Computational Mathematics (Springer, Heidelberg, Germany, 2013). Pages 167, 170

[30] M. Boman, Estimates for the L_2-projection onto continuous finite element spaces in a weighted L_p-norm. BIT **46**(2), 249–260 (2006). Page 285

[31] M. Bonazzoli, F. Rapetti, High-order finite elements in numerical electromagnetism: degrees of freedom and generators in duality. Numer. Algorithms **74**(1), 111–136 (2017). Pages 167, 181

[32] J. Bonelle, A. Ern, Analysis of compatible discrete operator schemes for elliptic problems on polyhedral meshes. ESAIM: Math. Model. Numer. Anal. **48**(2), 553–581 (2014). Page 112

[33] A. Bonito, J.-L. Guermond, F. Luddens, An interior penalty method with C^0 finite elements for the approximation of the Maxwell equations in heterogeneous media: convergence analysis with minimal regularity. ESAIM Math. Model. Numer. Anal. **50**(5), 1457–1489 (2016). Pages 287, 298

[34] J.P. Borthagaray, P. Ciarlet Jr., On the convergence in H^1-norm for the fractional Laplacian. SIAM J. Numer. Anal. **57**(4), 1723–1743 (2019). Page 21

[35] L. Bos, M.A. Taylor, B.A. Wingate, Tensor product Gauss-Lobatto points are Fekete points for the cube. Math. Comp. **70**(236), 1543–1547 (2001). Pages 70, 71

[36] A. Bossavit, *Electromagnétisme en vue de la modélisation*, vol. 14, SMAI Series on Mathematics and Applications (Springer, Paris, France, 1993). See also *Computational Electromagnetism, Variational Formulations, Complementary, Edge Elements* (Academic Press, San Diego, CA, 1998). Page 181

[37] A. Bossavit, Computational electromagnetism and geometry. J. Japan Soc. Appl. Electromagn. Mech. 7-8, 150–9 (no 1), 294–301 (no 2), 401–8 (no 3), 102–9 (no 4), 203–9 (no 5), 372–7 (no 6), (1999–2000). Page 112

[38] J. Bourgain, H. Brezis, P. Mironescu, Another look at Sobolev spaces, in *Optimal Control and Partial Differential Equations* (2001), pp. 439–455. Page 21

[39] D. Braess, *Finite Elements*, 3rd edn. (Cambridge University Press, Cambridge, UK, 2007). Translated from the German by Larry L, Schumake. Page 129

[40] J.H. Bramble, S.R. Hilbert, Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation. SIAM J. Numer. Anal. **7**, 112–124 (1970). Page 129

[41] J.H. Bramble, S.R. Hilbert, Bounds for a class of linear functionals with applications to Hermite interpolation. Numer. Math. **16**, 362–369 (1970/1971). Page 153

[42] J.H. Bramble, J.E. Pasciak, A.H. Schatz, The construction of preconditioners for elliptic problems by substructuring. I. Math. Comp. **47**(175), 103–134 (1986). Page 221

[43] J.H. Bramble, J.E. Pasciak, O. Steinbach, On the stability of the L^2 projection in $H^1(\Omega)$. Math. Comp. **71**(237), 147–156 (2002). Page 285

[44] S.C. Brenner, Two-level additive Schwarz preconditioners for nonconforming finite elements, in *Domain Decomposition Methods in Scientific and Engineering Computing*, ed. by D.E. Keyes, J. Xue (AMS, 1993). Proceedings of the 7th International Conference on Domain Decomposition Methods. http://www.ddm.org/DD07/index-neu.htm. Page 277

[45] S.C. Brenner, Two-level additive Schwarz preconditioners for nonconforming finite element methods. Math. Comp. **65**(215), 897–921 (1996). Page 277

[46] S.C. Brenner, Poincaré-Friedrichs inequalities for piecewise H^1 functions. SIAM J. Numer. Anal. **41**(1), 306–324 (2003). Page 277

[47] S.C. Brenner, L.R. Scott, *The Mathematical Theory of Finite Element Methods*, vol. 15, 3rd edn., Texts in Applied Mathematics (Springer, New York, NY, 2008). Pages 129, 154

[48] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext (Springer, New York, NY, 2011). Pages 3, 9, 10, 11, 12, 22, 23, 25, 31, 34, 42, 289, 304, 306, 307

[49] F. Brezzi, J. Douglas Jr., L.D. Marini, Recent results on mixed finite element methods for second order elliptic problems, *Vistas in Applied Mathematics*, Transl. Ser. Math. Engrg. (Optimization Software, New York, NY, 1986), pp. 25–43. Pages 169, 171

[50] F. Brezzi, J. Douglas Jr., R.G. Durán, M. Fortin, Mixed finite elements for second order elliptic problems in three variables. Numer. Math. **51**(2), 237–250 (1987). Pages 169, 171

[51] F. Brezzi, J. Douglas Jr., M. Fortin, L.D. Marini, Efficient rectangular mixed finite elements in two and three space variables. RAIRO Modél. Math. Anal. Numér. **21**(4), 581–604 (1987). Page 171

[52] E. Burman, A. Ern, Continuous interior penalty hp-finite element methods for advection and advection-diffusion equations. Math. Comp. **76**(259), 1119–1140 (2007). Page 277

[53] E. Burman, A. Ern, A continuous finite element method with face penalty to approximate Friedrichs' systems. M2AN Math. Model. Numer. Anal. **41**(1), 55–76 (2007). Page 277

[54] A.P. Calderón, Lebesgue spaces of differentiable functions and distributions. Proc. Sympos. Pure Math. **4**, 33–49 (1961). Page 24

[55] M. Campos Pinto, *Développement et analyse de méthodes adaptatives pour les équations de transport*. PhD thesis, University Pierre et Marie Curie - Paris VI, France (2006). https://tel.archives-ouvertes.fr/tel-00129013. Page 24

[56] M. Campos Pinto, E. Sonnendrücker, Gauss-compatible Galerkin schemes for time-dependent Maxwell equations. Math. Comp. **302**, 2651–2685 (2016). Page 277

[57] C. Canuto, A. Quarteroni, Approximation results for orthogonal polynomials in Sobolev spaces. Math. Comp. **38**(157), 67–86 (1982). Page 141

[58] C. Canuto, M.Y. Hussaini, A. Quarteroni, T.A. Zang, *Spectral Methods*. Scientific Computation (Springer, Berlin, Germany, 2006). Page 81

[59] C. Carstensen. Quasi-interpolation and a posteriori error analysis in finite element methods. M2AN Math. Model. Numer. Anal. **33**(6), 1187–1202 (1999). Page 279

[60] C. Carstensen, Merging the Bramble-Pasciak-Steinbach and the Crouzeix-Thomée criterion for H^1-stability of the L^2-projection onto finite element spaces. Math. Comp. **71**(237), 157–163 (2002). Page 285

[61] C. Carstensen, An adaptive mesh-refining algorithm allowing for an H^1-stable L^2-projection onto Courant finite element spaces. Constr. Approx. **20**(4), 549–564 (2004). Page 285

[62] C. Carstensen, S.A. Funken, Constants in Clément-interpolation error and residual based a posteriori error estimates in finite element methods. East-West J. Numer. Math. **8**(3), 153–175 (2000). Page 144

[63] C. Carstensen, R. Verfürth, Edge residuals dominate a posteriori error estimates for low order finite element methods. SIAM J. Numer. Anal. **36**(5), 1571–1587 (1999). Page 279

[64] H. Cartan, *Cours de calcul différentiel.* Collection Méthodes (Hermann, Paris, France, 1997). Page 310

[65] S.N. Chandler-Wilde, D.P. Hewett, A. Moiola, Interpolation of Hilbert and Sobolev spaces: quantitative estimates and counterexamples. Mathematika **61**(2), 414–443 (2015). Pages 33, 283

[66] Q. Chen, I. Babuška, Approximate optimal points for polynomial interpolation of real functions in an interval and in a triangle. Comput. Methods Appl. Mech. Engrg. **128**(3–4), 405–417 (1995). Page 81

[67] A. Chernov, Optimal convergence estimates for the trace of the polynomial L^2-projection operator on a simplex. Math. Comp. **81**(278), 765–787 (2012). Page 144

[68] S.H. Christiansen, Stability of Hodge decompositions in finite element spaces of differential forms in arbitrary dimension. Numer. Math. **107**(1), 87–106 (2007). See also Preprint 2005:19, The Univeristy of Oslo, Department of Mathematics, https://www.duo.uio.no/handle/10852/10573. Page 292

[69] S.H. Christiansen, On eigenmode approximation for Dirac equations: differential forms and fractional Sobolev spaces. Math. Comp. **87**(310), 547–580 (2018). Page 297

[70] S.H. Christiansen, F. Rapetti, On high order finite element spaces of differential forms. Math. Comp. **85**(298), 517–548 (2016). Pages 167, 181

[71] S.H. Christiansen, R. Winther, Smoothed projections in finite element exterior calculus. Math. Comp. **77**(262), 813–829 (2008). Page 292

[72] S.-K. Chua, R.L. Wheeden, Estimates of best constants for weighted Poincaré inequalities on convex domains. Proc. London Math. Soc. (3) **93**(1), 197–226 (2006). Page 34

[73] P. Ciarlet Jr., Analysis of the Scott-Zhang interpolation in the fractional order Sobolev spaces. J. Numer. Math. **21**(3), 173–180 (2013). Pages 145, 282

[74] P. Ciarlet Jr., J. Zou, Fully discrete finite element approaches for time-dependent Maxwell's equations. Numer. Math. **82**(2), 193–219 (1999). Page 211

[75] P.G. Ciarlet, *Mathematical Elasticity. Vol. I*, vol. 20, Studies in Mathematics and its Applications (North-Holland Publishing Co., Amsterdam, The Netherlands, 1988). Page 106

[76] P.G. Ciarlet. *Basic Error Estimates for Elliptic Problems*, ed. by P.G. Ciarlet, J.-L. Lions, volume II: Finite Element Methods of Handbook of Numerical Analysis, chapter 2 (North Holland, Amsterdam, The Netherlands, 1991). Pages 51, 154, 156

[77] P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*, vol. 40, Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, Reprint of the 1978 original (North-Holland, Amsterdam, The Netherlands, 2002). Pages 129, 156

[78] P.G. Ciarlet, P.-A. Raviart, Interpolation theory over curved elements, with applications to finite element methods. Comput. Methods Appl. Mech. Engrg. **1**, 217–249 (1972). Pages 129, 148, 150, 156, 310

[79] P.G. Ciarlet, P.-A. Raviart, General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods. Arch. Rational Mech. Anal. **46**, 177–199 (1972). Pages 129, 158

[80] P. Clément, Approximation by finite element functions using local regularization. RAIRO, Anal. Num. **9**, 77–84 (1975). Page 279

[81] B. Cockburn, G. Kanschat, D. Schötzau, A note on discontinuous Galerkin divergence-free solutions of the Navier-Stokes equations. J. Sci. Comput. **31**(1–2), 61–73 (2007). Page 277

[82] M. Costabel, M. Dauge, Crack singularities for general elliptic systems. Math. Nachr. **235**, 29–49 (2002). Page 29

[83] M. Costabel, A. McIntosh, On Bogovskiǐ and regularized Poincaré integral operators for de Rham complexes on Lipschitz domains. Math. Z. **265**(2), 297–320 (2010). Page 197

[84] R. Courant, Variational methods for the solution of problems of equilibrium and vibrations. Bull. Amer. Math. Soc. **49**, 1–23 (1943). Pages v, 233

[85] R. Courant, K. Friedrichs, H. Lewy, Über die partiellen Differenzengleichungen der mathematischen Physik. Math. Ann. **100**(1), 32–74 (1928). Page 36

[86] M. Crouzeix, 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), 3–75 (1973). Page 82

[87] M. Crouzeix, V. Thomée, The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces. Math. Comp. **48**(178), 521–532 (1987). Page 285

[88] F. Demengel, G. Demengel, Functional Spaces, for the Theory of Elliptic Partial Differential Equations, Universitext (Springer, London, UK; EDP Sciences, Les Ulis, France, 2012). Translated from the 2007 French original by Reinie Erné. Pages 3, 24

[89] A. Denjoy, L. Felix, P. Montel, Henri Lebesgue, le savant, le professeur, l'homme. Enseignement Math. (2), **3**, 1–18 (1957). Page 8

[90] J. Deny, J.-L. Lions, Les espaces du type de Beppo Levi. Ann. Inst. Fourier (Grenoble) **5**, 305–370 (1954). Page 129

[91] M. Dubiner, Spectral methods on triangles and other domains. J. Sci. Comput. **6**(4), 345–390 (1991). Page 81

[92] T. Dupont, R. Scott, Polynomial approximation of functions in Sobolev spaces. Math. Comp. **34**(150), 441–463 (1980). Page 34

[93] N. Dyn, D. Levine, J.A. Gregory, A butterfly subdivision scheme for surface interpolation with tension control. ACM Trans. Graph. **9**(2), 160–169 (1990). Page 160

[94] P. Erdős, Problems and results on the theory of interpolation. II. Acta Math. Acad. Sci. Hungar. **12**, 235–244 (1961). Page 68

[95] K. Eriksson, C. Johnson, Adaptive finite element methods for parabolic problems. II. Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. SIAM J. Numer. Anal. **32**(3), 706–740 (1995). Page 285

[96] A. Ern, J.-L. Guermond, Mollification in strongly Lipschitz domains with application to continuous and discrete de Rham complexes. Comput. Methods Appl. Math. **16**(1), 51–75 (2016). Pages 44, 287, 288, 298

[97] A. Ern, J.-L. Guermond, Finite element quasi-interpolation and best approximation. ESAIM Math. Model. Numer. Anal. **51**(4), 1367–1385 (2017). Pages 34, 145, 279, 281, 282

[98] A. Ern, M. Vohralík, A posteriori error estimation based on potential and flux reconstruction for the heat equation. SIAM J. Numer. Anal. **48**(1), 198–223 (2010). Page 277

[99] L.C. Evans, *Partial Differential Equations*, vol. 19. Graduate Studies in Mathematics (American Mathematical Society, Providence, RI, 1998). Pages 3, 18, 19, 22, 23, 25, 34

[100] R.S. Falk, R. Winther, Local bounded cochain projections. Math. Comp. **83**(290), 2631–2656 (2014). Page 297

[101] L. Fejér, Lagrangesche Interpolation und die zugehörigen konjugierten Punkte. Math. Ann. **106**(1), 1–55 (1932). Page 68

[102] K.O. Friedrichs, The identity of weak and strong extensions of differential operators. Trans. Amer. Math. Soc. **55**, 132–151 (1944). Page 287

[103] F. Fuentes, B. Keith, L. Demkowicz, S. Nagaraj, Orientation embedded high order shape functions for the exact sequence elements of all shapes. Comput. Math. Appl. **70**(4), 353–458 (2015). Pages 84, 167, 181

[104] E. Gagliardo, Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in n variabili. Rend. Sem. Mat. Univ. Padova **27**, 284–305 (1957). Page 31

[105] F.D. Gaspoz, C.-J. Heine, K.G. Siebert, Optimal grading of the newest vertex bisection and H^1-stability of the L_2-projection. IMA J. Numer. Anal. **36**(3), 1217–1241 (2016). Page 285

[106] M. Gerritsma, An introduction to a compatible spectral discretization method. Mech. Adv. Mater. Struc. **19**(1–3), 48–67 (2012). Page 112

[107] V. Girault, P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations. Theory and Algorithms*. Springer Series in Computational Mathematics (Springer, Berlin, Germany, 1986). Pages 158, 306

[108] V. Girault, L.R. Scott, Analysis of a two-dimensional grade-two fluid model with a tangential boundary condition. J. Math. Pures Appl. (9) **78**(10), 981–1011 (1999). Page 290

[109] J. Gopalakrishnan, L.E. García-Castillo, L.F. Demkowicz, Nédélec spaces in affine coordinates. Comput. Math. Appl. **49**(7–8), 1285–1294 (2005). Page 181

[110] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, vol. 24. Monographs and Studies in Mathematics (Pitman (Advanced Publishing Program), Boston, MA, 1985). Pages 3, 24, 25, 28, 29, 30, 31, 32, 33, 37, 42

[111] T. Gudi, A new error analysis for discontinuous finite element methods for linear elliptic problems. Math. Comp. **79**(272), 2169–2189 (2010). Page 277

[112] I. Harari, T.J.R. Hughes, What are C and h?: Inequalities for the analysis and design of finite element methods. Comput. Methods Appl. Mech. Engrg. **97**(2), 157–192 (1992). Page 140

[113] J. Heinonen, *Lectures on Lipschitz Analysis*, vol. 100, Report (Department of Mathematics and Statistics. University of Jyväskylä, Finland, 2005). Page 20

[114] J.S. Hesthaven, From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex. SIAM J. Numer. Anal. **35**(2), 655–676 (1998). Page 68

[115] J.S. Hesthaven, S. Gottlieb, D. Gottlieb, *Spectral Methods for Time-Dependent Problems*, vol. 21, Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, Cambridge, UK, 2007). Page 68

[116] N. Heuer, On the equivalence of fractional-order Sobolev semi-norms. J. Math. Anal. Appl. **417**(2), 505–518 (2014). Page 34

[117] R. Hiptmair, Canonical construction of finite elements. Math. Comp. **68**(228), 1325–1346 (1999). Pages 167, 181

[118] S. Hofmann, M. Mitrea, M. Taylor, Geometric and transformational properties of Lipschitz domains, Semmes-Kenig-Toro domains, and other classes of finite perimeter domains. J. Geom. Anal. **17**(4), 593–647 (2007). Pages 287, 288

[119] R.H.W. Hoppe, B. Wohlmuth, Element-oriented and edge-oriented local error estimators for nonconforming finite element methods. RAIRO Modél. Math. Anal. Numér. **30**(2), 237–263 (1996). Page 277

[120] P. Houston, D. Schötzau, T.P. Wihler, Energy norm a posteriori error estimation of *hp*-adaptive discontinuous Galerkin methods for elliptic problems. Math. Models Methods Appl. Sci. **17**(1), 33–62 (2007). Page 277

[121] C. Johnson, A. Szepessy, On the convergence of a finite element method for a nonlinear hyperbolic conservation law. Math. Comp. **49**(180), 427–444 (1987). Page 286

[122] O.A. Karakashian, 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). Page 277

[123] G.E. Karniadakis, S.J. Sherwin, *Spectral/hp Element Methods for Computational Fluid Dynamics*, 2nd edn. Numerical Mathematics and Scientific Computation (Oxford University Press, New York, NY, 2005). Pages 64, 81

[124] T. Kato, Estimation of iterated matrices, with application to the von Neumann condition. Numer. Math. **2**, 22–29 (1960). Page 56

[125] R.C. Kirby, Fast simplicial finite element algorithms using Bernstein polynomials. Numer. Math. **117**(4), 631–652 (2011). Page 81

[126] R.C. Kirby, Low-complexity finite element algorithms for the de Rham complex on simplices. SIAM J. Sci. Comput. **36**(2), A846–A868 (2014). Page 81

[127] R. Kornhuber, D. Peterseim, H. Yserentant, An analysis of a class of variational multiscale methods based on subspace decomposition. Math. Comp. **87**(314), 2765–2774 (2018). Page 277

[128] A. Kroó, On the exact constant in the L_2 Markov inequality. J. Approx. Theory **151**(2), 208–211 (2008). Page 140

[129] A. Kroó, S. Révész, On Bernstein and Markov-type inequalities for multivariate polynomials on convex bodies. J. Approx. Theory **99**(1), 134–152 (1999). Page 140

[130] N. Kuznetsov, A. Nazarov, Sharp constants in the Poincaré, Steklov and related inequalities (a survey). Mathematika **61**(2), 328–344 (2015). Page 36

[131] P.D. Lax. *Functional Analysis*. Pure and Applied Mathematics (Wiley-Interscience, New York, NY, 2002). Pages 304, 306

[132] M. Lenoir, Optimal isoparametric finite elements and error estimates for domains involving curved boundaries. SIAM J. Numer. Anal. **23**(3), 562–580 (1986). Pages 154, 156

[133] J. Leray, Sur le mouvement d'un liquide visqueux emplissant l'espace. Acta Math. **63**(1), 193–248 (1934). Page 287

[134] M.W. Licht, Smoothed projections and mixed boundary conditions. Math. Comp. **88**(316), 607–635 (2019). Page 297

[135] J.-L. Lions, E. Magenes, *Non-Homogeneous Boundary Value Problems and Applications*. Vols. I, II (Springer, New York-Heidelberg, 1972). Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181–182. Pages 33, 283

[136] F.W. Luttmann, T.J. Rivlin, Some numerical experiments in the theory of polynomial interpolation. IBM J. Res. Develop. **9**, 187–191 (1965). Page 68

[137] Y. Maday, O. Mula, A.T. Patera, M. Yano, The generalized empirical interpolation method: stability theory on Hilbert spaces with an application to the Stokes equation. Comput. Methods Appl. Mech. Engrg. **287**, 310–334 (2015). Page 56

[138] J. Malý, W.P. Ziemer, *Fine Regularity of Solutions of Elliptic Partial Differential Equations*, vol. 51. Mathematical Surveys and Monograph (American Mathematical Society, Providence, RI, 1997). Pages 3, 10, 18, 19, 25

[139] J.E. Marsden, T.J.R. Hughes, *Mathematical Foundations of Elasticity* (Dover Publications Inc., New York, NY, 1994). Corrected reprint of the 1983 original. Page 106

[140] V. Maz'ya, T. Shaposhnikova, On the Bourgain, Brezis, and Mironescu theorem concerning limiting embeddings of fractional Sobolev spaces. J. Funct. Anal. **195**(2), 230–238 (2002). Page 21

[141] W. McLean, *Strongly Elliptic Systems and Boundary Integral Equations* (Cambridge University Press, Cambridge, UK, 2000). Pages 24, 31

[142] C. Meray, Observations sur la légitimité de l'interpolation. Ann. Sci. Ecole Normale Sup. **3**(1), 165–176 (1884). Page 68

[143] N.G. Meyers, J. Serrin, $H = W$. Proc. Nat. Acad. Sci. U.S.A. **51**, 1055–1056 (1964). Page 22

[144] P. Monk, Analysis of a finite element method for Maxwell's equations. SIAM J. Numer. Anal. **29**(3), 714–729 (1992). Page 211

[145] P. Monk, *Finite Element Methods for Maxwell's Equations*, Numerical Mathematics and Scientific Computation (Oxford University Press, New York, NY, 2003). Pages 106, 167, 181

[146] P. Monk, E. Süli, The adaptive computation of far-field patterns by a posteriori error estimation of linear functionals. SIAM J. Numer. Anal. **36**(1), 251–274 (1999). Page 144

[147] P. Morin, R.H. Nochetto, K.G. Siebert, Data oscillation and convergence of adaptive FEM. SIAM J. Numer. Anal. **38**(2), 466–488 (2000). Page 285

[148] C.B. Morrey, Jr., *Multiple Integrals in the Calculus of Variations*. Die Grundlehren der mathematischen Wissenschaften, Band 130 (Springer New York, Inc., New York, NY, 1966). Page 135

[149] E.J. Nanson, Note on hydrodynamics. Messenger Math. **VII**, 182–185 (1877–1878). Page 108

[150] J. Nečas, Sur une méthode pour résoudre les équations aux dérivées partielles de type elliptique, voisine de la variationnelle. Ann. Scuola Norm. Sup. Pisa **16**, 305–326 (1962). Page v

[151] J.-C. Nédélec, Mixed finite elements in $\mathbb{R}^3$. Numer. Math. **35**(3), 315–341 (1980). Pages 167, 181, 184

[152] J.-C. Nédélec, A new family of mixed finite elements in $\mathbb{R}^3$. Numer. Math. **50**, 57–81 (1986). Page 184

[153] P. Oswald, On a BPX-preconditioner for P_1 elements. Computing **51**, 125–133 (1993). Page 277

[154] R.G. Owens, Spectral approximations on the triangle. R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **454**(1971), 857–872 (1998). Page 81

[155] S. Özişik, B. Rivière, T. Warburton, On the constants in inverse inequalities in L_2. Technical Report TR10-19, Rice University (2010). Page 140

[156] A.T. Patera, Spectral methods for spatially evolving hydrodynamic flows, in *Spectral Methods for Partial Differential Equations (Hampton, VA, 1982)* (SIAM, Philadelphia, PA, 1984), pp. 239–256. Page 71

[157] L.E. Payne, H.F. Weinberger, An optimal Poincaré inequality for convex domains. Arch. Rational Mech. Anal. **5**, 286–292 (1960). Page 34

[158] H. Poincaré, Sur les equations aux dérivées partielles de la physique mathématique. Am. J. Math. **12**(3), 211–294 (1890). Pages 35, 36

[159] H. Poincaré, Sur les équations de la physique mathématique. Rendiconti del Circolo Matematico di Palermo **8**(1), 57–155 (1894). Page 36

[160] A.C. Ponce, J. Van Schaftingen, The continuity of functions with N-th derivative measure. Houston J. Math. **33**(3), 927–939 (2007). Page 24

[161] M.J.D. Powell, M.A. Sabin, Piecewise quadratic approximations on triangles. ACM Trans. Math. Software **3**(4), 316–325 (1977). Page 58

[162] J. Proriol, Sur une famille de polynômes à deux variables orthogonaux dans un triangle. C. R. Acad. Sci. Paris **245**, 2459–2461 (1957). Page 81

[163] F. Rapetti, A. Bossavit, Whitney forms of higher degree. SIAM J. Numer. Anal. **47**(3), 2369–2386 (2009). Pages 167, 181

[164] P.-A. Raviart, J.-M. Thomas, A mixed finite element method for second order elliptic problems, in *Mathematical aspects of the finite element method*, vol. 606, Lecture Notes in Mathematics, ed. by E.M.I. Galligani (Springer, New York, NY, 1977). Page 167

[165] P.-A. Raviart, J.M. Thomas, Primal hybrid finite element methods for 2nd order elliptic equations. Math. Comp. **31**(138), 391–413 (1977). Page 167

[166] S. Rebay, Efficient unstructured mesh generation by means of Delaunay triangulation and Bowyer-Watson algorithm. J. Comput. Phys. **106**, 125–138 (1993). Page 98

[167] T.J. Rivlin, *An Introduction to the Approximation of Functions* (Dover Publications Inc., New York, NY, 1981). Corrected reprint of the 1969 original, Dover books on Advanced Mathematics. Page 68

[168] M.E. Rognes, R.C. Kirby, A. Logg, Efficient assembly of $H(\mathrm{div})$ and $H(\mathrm{curl})$ conforming finite elements. SIAM J. Sci. Comput. **31**(6), 4130–4151 (2009/10). Page 106

[169] W. Rudin. *Principles of Mathematical Analysis*, 3rd edn. International Series in Pure and Applied Mathematics (McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976). Pages 3, 4, 5, 6

[170] W. Rudin, *Real and Complex Analysis*, 3rd edn. (McGraw-Hill Book Co., New York, NY, 1987). Pages 3, 7, 8, 9, 10, 11, 15, 304

[171] C. Runge, Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. Zeit. Math. Physik **46**, 224–243 (1901). Page 68

[172] J. Schöberl, Commuting quasi-interpolation operators for mixed finite elements. Technical Report ISC-01-10-MATH, Texas A&M University (2001). URL www.isc.tamu.edu/publications-reports/tr/0110.pdf. Page 288

[173] J. Schöberl, A multilevel decomposition result in $H(\mathrm{curl})$, in *Multigrid, Multilevel and Multiscale Methods, EMG, 2005*, ed. by P. Wesseling, C.W. Oosterlee, P. Hemker (2005). Page 292

[174] J. Schöberl, A posteriori error estimates for Maxwell equations. Math. Comp. **77**(262), 633–649 (2008). Page 288

[175] J. Schöberl, C. Lehrenfeld, Domain decomposition preconditioning for high order hybrid discontinuous Galerkin methods on tetrahedral meshes, in *Advanced Finite Element Methods and Applications*, vol. 66. Lect. Notes Appl. Comput. Mech. (Springer, Heidelberg, Germany, 2013), pp. 27–56. Page 277

[176] J. Schöberl, S. Zaglmayr, High order Nédélec elements with local complete sequence properties. COMPEL **24**(2), 374–384 (2005). Pages 170, 184

[177] C. Schwab, *p- and hp-Finite Element Methods*, Numerical Mathematics and Scientific Computation (The Clarendon Press, Oxford University Press, New York, NY, 1998). Pages 139, 140

[178] R.L. Scott, S. Zhang, Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math. Comp. **54**(190), 483–493 (1990). Page 279

[179] S.L. Sobolev, Sur un théorème d'analyse fonctionnelle. Rec. Math. [Mat. Sbornik] N.S. **4**(46), 471–497 (1938). Page 287

[180] S.L. Sobolev, *Applications of Functional Analysis in Mathematical Physics*, 2nd edn., vol. VII. Translations of Mathematical Monographs (American Mathematical Society, Providence, RI, 1963). Pages 3, 23

[181] E.M. Stein, *Singular Integrals and Differentiability Properties of Functions* (Princeton University Press, Princeton, NJ, 1970). Page 24

[182] W. Stekloff, (aka V.A. Steklov). Problème de refroidissement d'une barre hétérogène. Ann. Fac. Sci. Toulouse Sci. Math. Sci. Phys. (2) **3**(3), 281–313 (1901). Page 36

[183] V.A. Steklov, On the expansion of a given function into a series of harmonic functions (in Russian). Communications de la Société mathématique de Kharkow, 2ème Série **5**, 60–73 (1897). URL http://mi.mathnet.ru/khmo200. Page 36

[184] V.A. Steklov, The problem of cooling of an heterogeneous rigid rod (in Russian). Communications de la Société mathématique de Kharkow, 2ème Série **5**, 136–181 (1897). URL http://mi.mathnet.ru/khmo222. Page 36

[185] V.A. Steklov, On the expansion of a given function into a series of harmonic functions (in Russian). Communications de la Société mathématique de Kharkow, 2ème Série **6**, 57–124 (1899). URL http://mi.mathnet.ru/khmo183. Page 36

[186] R. Stevenson, Optimality of a standard adaptive finite element method. Found. Comput. Math. **7**(2), 245–269 (2007). Page 285

[187] R. Stevenson, The completion of locally refined simplicial partitions created by bisection. Math. Comp. **77**(261), 227–241 (2008). Page 285

[188] D.B. Szyld, The many proofs of an identity on the norm of oblique projections. Numer. Algorithms **42**(3–4), 309–323 (2006). Page 56

[189] L. Tartar, *An Introduction to Sobolev Spaces and Interpolation Spaces*, vol. 3. UMI, Bologna, Italy, Lecture Notes of the Unione Matematica Italiana (Springer, Berlin, Germany, 2007). Pages 3, 12, 19, 20, 21, 23, 24, 33, 44, 282, 283, 306, 307, 308

[190] M.A. Taylor, B.A. Wingate, R.E. Vincent, An algorithm for computing Fekete points in the triangle. SIAM J. Numer. Anal. **38**(5), 1707–1720 (2000). Page 81

[191] L.N. Trefethen, J.A.C. Weideman, Two results on polynomial interpolation in equally spaced points. J. Approx. Theory **65**(3), 247–260 (1991). Page 68

[192] C. Truesdell, R. Toupin, The classical field theories, in *Handbuch der Physik, Band III/1*, pages 226–793; appendix (Springer, Berlin, Germany, 1960), pp. 794–858. With an appendix on tensor fields by J.L. Ericksen. Page 108

[193] A. Veeser, R. Verfürth, Explicit upper bounds for dual norms of residuals. SIAM J. Numer. Anal. **47**(3), 2387–2405 (2009). Page 144

[194] A. Veeser, R. Verfürth, Poincaré constants for finite element stars. IMA J. Numer. Anal. **32**(1), 30–47 (2012). Pages 34, 279

[195] R. Verfürth, On the constants in some inverse inequalities for finite element functions. Technical report, Ruhr-Universität Bochum (2004). Page 146

[196] T. Warburton, J.S. Hesthaven, On the constants in hp-finite element trace inverse inequalities. Comput. Methods Appl. Mech. Engrg. **192**(25), 2765–2773 (2003). Page 142

[197] J.P. Webb, Hierarchal vector basis functions of arbitrary order for triangular and tetrahedral finite elements. IEEE Trans. Antennas Propag. **47**(8), 1244–1253 (1999). Page 181

[198] A. Weil, Sur les théorèmes de de Rham. Commentarii Math. Helvetici **26**, 119–145 (1952). Pages 167, 181

[199] H. Whitney, *Geometric Integration Theory* (Princeton University Press, Princeton, NJ, 1957). Pages 44. 167, 181

[200] D.R. Wilhelmsen, A Markov inequality in several dimensions. J. Approx. Theory **11**, 216–220 (1974). Page 139

[201] J. Xu, L. Zikatanov, Some observations on Babuška and Brezzi theories. Numer. Math. **94**(1), 195–202 (2003). Page 56

[202] K. Yosida, *Functional Analysis*. Classics in Mathematics (Springer, Berlin, Germany, 1995). Reprint of the sixth (1980) edition. Pages 3, 304, 306

[203] M. Zlámal, Curved elements in the finite element method. I. SIAM J. Numer. Anal. **10**, 229–240 (1973). Page 154

[204] M. Zlámal, Curved elements in the finite element method. II. SIAM J. Numer. Anal. **11**, 347–362 (1974). Page 154

Index