IMA Journal of Numerical Analysis (2022) **42**, 1023–1049
https://doi.org/10.1093/imanum/draa103
Advance Access publication on 16 March 2021

Equivalence of local- and global-best approximations, a simple stable local commuting projector, and optimal *hp* approximation estimates in $\boldsymbol{H}(\mathrm{div})$

Alexandre Ern
CERMICS, Ecole des Ponts, 77455 Marne-la-Vallée, France, and Inria, 2 rue Simone Iff, 75589 Paris, France

Thirupathi Gudi
Department of Mathematics, Indian Institute of Science, Bangalore, India 560012

Iain Smears
Department of Mathematics, University College London, London WC1E 6BT, UK

and

Martin Vohralík*
Inria, 2 rue Simone Iff, 75589 Paris, France, and CERMICS, Ecole des Ponts, 77455 Marne-la-Vallée, France
*Corresponding author: martin.vohralik@inria.fr

[Received on 18 May 2020; revised on 24 November 2020]

Given an arbitrary function in $\boldsymbol{H}(\mathrm{div})$, we show that the error attained by the global-best approximation by $\boldsymbol{H}(\mathrm{div})$-conforming piecewise polynomial Raviart–Thomas–Nédélec elements under additional constraints on the divergence and normal flux on the boundary is, up to a generic constant, equivalent to the sum of independent local-best approximation errors over individual mesh elements, without constraints on the divergence or normal fluxes. The generic constant only depends on the shape-regularity of the underlying simplicial mesh, the space dimension, and the polynomial degree of the approximations. The analysis also gives rise to a stable, local, commuting projector in $\boldsymbol{H}(\mathrm{div})$, delivering an approximation error that is equivalent to the local-best approximation. We next present a variant of the equivalence result, where robustness of the constant with respect to the polynomial degree is attained for unbalanced approximations. These two results together further enable us to derive rates of convergence of global-best approximations that are fully optimal in both the mesh size h and the polynomial degree p, for vector fields that only feature elementwise the minimal necessary Sobolev regularity. We finally show how to apply our findings to derive optimal *a priori hp*-error estimates for mixed and least-squares finite element methods applied to a model diffusion problem.

Keywords: best approximation; piecewise polynomial; localization; $\boldsymbol{H}(\mathrm{div})$ Sobolev space; Raviart–Thomas–Nédélec space; minimal regularity; optimal error bound; commuting projector; mixed finite element method; least-squares method; *a priori* error estimate.

1. Introduction

Interpolation operators that approximate a given function with weak gradient, curl, or divergence by a piecewise polynomial of degree p are fundamental in numerical analysis. Typically, this is done over a computational domain Ω covered by a mesh $\mathscr{T}$ with characteristic size h. The canonical interpolation operators associated with the degrees of freedom commute with the appropriate differential operators and they are projectors, i.e., they leave the interpolated function invariant if it is already a piecewise polynomial; the term 'commuting projector' is commonly used in the literature in such a case. They are also local (defined independently on each element K of the mesh $\mathscr{T}$) and they lead to optimal approximation error bounds with respect to the mesh size h. However, canonical interpolation operators have two main deficiencies. First, they can act on a given function only if it possesses more regularity beyond the minimal H^1, $\boldsymbol{H}(\mathrm{div})$, or $\boldsymbol{H}(\mathrm{curl})$. And secondly, they are not well suited to deriving approximation error bounds that are optimal in the polynomial degree p.

1.1 *Interpolation operators and hp-approximation*

The projection-based interpolation operators, see Demkowicz & Buffa (2005), Demkowicz (2008), and the references therein, lead to *optimal approximation properties* in the *mesh size h* and *quasi-optimal approximation properties* in the *polynomial degree p* (up to logarithmic factors). They were derived under a conjecture of existence of commuting and polynomial-preserving extension operators from the boundary of the given element K to its interior, which was later established in Demkowicz *et al.* (2008, 2009, 2012); the approximation results are summarized in Demkowicz *et al.* (2012, Theorem 8.1). Thus, these operators essentially lift the second drawback of the canonical interpolation operators described above (up to logarithmic factors), while still being local commuting projectors. However, they again require more regularity beyond the minimal H^1, $\boldsymbol{H}(\mathrm{div})$, or $\boldsymbol{H}(\mathrm{curl})$.

In the particular case of $\boldsymbol{H}(\mathrm{div})$, which constitutes the focus of the present work, the normal component of the interpolant on each mesh face is fully dictated by the normal component of the interpolated function on that face, which requires $\boldsymbol{H}^s(\mathrm{div})$ regularity with $s > 0$, which is slightly more than $\boldsymbol{H}(\mathrm{div})$. Some further refinements can be found in Bespalov & Heuer (2011) and Ern & Guermond (2018). Recently, a local commuting projector that has *optimal p-approximation properties* (does not feature the logarithmic factors) has been devised by Melenk & Rojik (2020). To define the projector though, higher regularity is needed with, in particular, $\boldsymbol{H}^s(\mathrm{div})$, $s \geqslant 1$, in the case of interest here.

The issue of constructing (quasi-)interpolation projectors under the minimal regularities H^1, $\boldsymbol{H}(\mathrm{div})$, and $\boldsymbol{H}(\mathrm{curl})$ has been addressed before; cf., e.g., Clément (1975), Scott & Zhang (1990), and Bernardi & Girault (1998) in the H^1 case, Nochetto & Stamm (2019) in the $\boldsymbol{H}(\mathrm{div})$ case, and Bernardi & Hecht (2007) in the $\boldsymbol{H}(\mathrm{curl})$ case; see also the references therein. Stability and h-optimal approximation estimates in any L^p-norm, $1 \leqslant p \leqslant \infty$, have recently been achieved by Ern & Guermond (2017) in a unified setting for a wide range of finite elements encompassing the whole discrete de Rham sequence. The arguments used in Ern & Guermond (2017) are somewhat different from those in the previous references: a projection onto the fully discontinuous (broken) piecewise polynomial space is applied first, followed by an averaging operator to ensure the appropriate H^1, $\boldsymbol{H}(\mathrm{div})$, or $\boldsymbol{H}(\mathrm{curl})$ trace continuity. Unfortunately, all of the quasi-interpolation projectors mentioned in this paragraph do not commute with the appropriate differential operators and, moreover, they are only shown to be optimal in h but not in p.

1.2 *Stable local commuting projectors under minimal regularity*

Constructing projectors applicable under the *minimal regularities* H^1, $\boldsymbol{H}(\mathrm{div})$, and $\boldsymbol{H}(\mathrm{curl})$ that would in addition be *commuting*, *stable*, and *locally defined* represents a long-standing effort. Stability, commutativity, and the projection property were obtained by Christiansen & Winther (2008) by composing the canonical interpolation operators with mollification, following some earlier ideas in particular from Schöberl (2001, 2005); cf. also Ern & Guermond (2016) for a shrinking technique avoiding the need for extensions outside the domain and Licht (2019) for essential boundary conditions only prescribed on a part of the boundary of Ω. These operators are, however, not locally defined. This last remaining issue was finally remedied in Falk & Winther (2014), where a patch-based construction resembling that of the Clément operator is introduced. However, no approximation properties are discussed, and stability is achieved only in the graph space of the appropriate differential operator, e.g., $\boldsymbol{H}(\mathrm{div})$, but not $\boldsymbol{L}^2$ for the case of interest here.

1.3 *Equivalence of local-best and global-best approximations*

In a seemingly rather unconnected recent result, Veeser (2016) showed that the error in the *best approximation* of a given scalar-valued function in H^1 by *continuous* piecewise polynomials is *equivalent* up to a generic constant to that by *discontinuous* piecewise polynomials. This result is termed equivalence of global- and local-best approximations. A predecessor result in the lowest-order case $p = 1$ and up to data oscillation can be easily deduced from Carstensen *et al.* (2012, Theorem 2.1 and inequalities (3.2), (3.5), and (3.6)); see also the references therein. Equivalences between approximations by different numerical methods are studied in Carstensen *et al.* (2012). A similar result is also given in Aurada *et al.* (2013, Proposition 3.1), and an improvement of the dependence of the equivalence constant on the polynomial degree in two space dimensions is developed in Canuto *et al.* (2017, Theorem 4).

This equivalence result might be surprising at first glance, since the local-best error is clearly smaller than the global-best one. The twist comes from the fact that any H^1 function is continuous in the sense of traces, so one does not gain in approximating it by discontinuous piecewise polynomials. For finite element discretizations of coercive problems, this result in particular allows one to obtain estimates by a direct application of the Deny–Lions/Bramble–Hilbert lemma that does not require invoking an additional conforming and locally stable interpolation operator; see Gudi (2010) or Carstensen & Schedensack (2015) for examples of tools sharing the same spirit in *a priori* error analysis. Another important application is in approximation classes in *a posteriori*-based convergence and optimality; cf. Veeser (2016).

1.4 *Main results of the paper*

Our main results can be divided into three parts.

(1) A simple stable local commuting projector defined under the minimal $\boldsymbol{H}(\mathrm{div})$ regularity

We define an interpolation operator on the entire $\boldsymbol{H}(\mathrm{div})$ space that is a projector, enjoys a commuting property with the divergence operator, is locally defined over patches of elements, and is stable in $\boldsymbol{L}^2$ up to an *hp* data oscillation term for the divergence. It also takes into account essential (no-flux) boundary conditions on only a part of the computational domain and it achieves, on each element, an error equivalent to local-best errors over a patch of neighboring elements. All these results are summarized in Definition 3.1 and Theorem 3.2 below. Our main tool for defining the projector is the *equilibrated flux reconstruction*. This technique has been traditionally used in *a posteriori* error analysis of *primal* finite element methods derived from H^1-formulations; see Destuynder & Métivet (1999), Luce & Wohlmuth (2004), Braess & Schöberl (2008), Ern & Vohralík (2013, 2015), Becker *et al.* (2016), and the references therein. We now employ it here in the context of *a priori* error analysis of *dual* approximations in $\boldsymbol{H}(\mathrm{div})$.

(2) Equivalence of local- and global-best approximations in $\boldsymbol{H}(\mathrm{div})$ under minimal regularity

In Theorem 3.3, we show that the *global* best-approximation error defined in (3.10) is, up to a generic constant, *equivalent* to the *local*-best approximation errors defined by elementwise minimizations (3.11). This actually results from the properties of the above projector. This extends the results of Carstensen *et al.* (2012), Aurada *et al.* (2013), Veeser (2016), Canuto *et al.* (2017) to the $\boldsymbol{H}(\mathrm{div})$ case, where we are importantly able to remove *constraints* on *both* the *normal trace* interelement continuity and the *divergence*.

(3) Optimal *hp*-approximation estimates in $\boldsymbol{H}(\mathrm{div})$

Our third main result is Theorem 3.6 where we derive *hp-approximation estimates*. These estimates feature the following four properties: (i) they request *no global regularity* of the approximated function $\boldsymbol{v}$ beyond $\boldsymbol{H}(\mathrm{div})$; (ii) only the *minimal local (elementwise)* $\boldsymbol{H}^s$*-regularity*, $s \geqslant 0$, is needed; (iii) the convergence rates are *fully optimal* in both the *mesh size h* and the *polynomial degree p*, in particular featuring no logarithmic factor of the polynomial degree p; (iv) *no higher-order norms of the divergence* of $\boldsymbol{v}$ appear in the bound whenever $s \geqslant 1$. This improves on Demkowicz & Buffa (2005), Demkowicz (2008) in removing the suboptimality with respect to the polynomial degree, on Demkowicz & Buffa (2005), Demkowicz (2008), Melenk & Rojik (2020) in reducing the regularity requirements, and on approximations using Clément-type operators in removing the need for regularity assumptions over the (overlapping) elemental patches while reducing it instead to (nonoverlapping) elements.

1.5 *Applications to mixed finite element and least-squares mixed finite element methods*

The above results can be immediately turned into fully optimal *hp a priori error estimates* for two popular classes of numerical methods for second-order elliptic partial differential equations: mixed finite element methods and least-squares methods, as we show in Lemmas 6.1–6.3. Note also that an immediate application of the commuting projector of Definition 3.1 in the context of mixed finite elements is the construction of a *Fortin operator* under *minimal* $\boldsymbol{H}(\mathrm{div})$ *regularity*.

1.6 *Organization of the paper*

The rest of the paper is organized as follows. In Section 2 we introduce the setting and notation. In Section 3 we state our main results, whereas Sections 4 and 5 are concerned with their proofs and Section 6 with the application. A result on polynomial-degree-robust equivalence between constrained and unconstrained best approximations in $\boldsymbol{H}(\mathrm{div})$ on a simplex is presented in Appendix A; it is of independent interest. We present our results in two or three space dimensions only since this is the current limitation of Lemma 4.4, which hinges on Costabel & McIntosh (2010) and Demkowicz *et al.* (2012). Lemma 4.4, with the involved constant independent of the polynomial degree p, is only crucial for Theorem 3.6; the other results actually hold in arbitrary space dimension. Variable polynomial degrees can be taken into account by proceeding as in, e.g., Dolejší *et al.* (2016). We avoid this here for the sake of clarity of exposition.

2. Setting and notation

2.1 *Domain* Ω*, space* $\boldsymbol{H}_{0,\Gamma_N}(\mathrm{div},\Omega)$*, and simplicial mesh* $\mathscr{T}$

Let $\Omega \subset \mathbb{R}^d$ for $d \in \{2,3\}$ be an open, bounded, connected polygon or polyhedron with Lipschitz boundary Γ. Let $\mathscr{T}$ be a given conforming, simplicial, shape-regular, and possibly locally refined mesh of Ω, i.e., $\overline{\Omega} = \cup_{K\in\mathscr{T}} K$, where any K is a closed simplex and the intersection of two different simplices is either an empty set or their common vertex, edge, or face. Let Γ_{D} be a (possibly empty) closed subset of Γ, and let $\Gamma_{\mathrm{N}} := \Gamma \setminus \Gamma_{\mathrm{D}}$ be its (relatively open) complement in Γ, with the assumption that $\mathscr{T}$ matches Γ_{D} and Γ_{N} in the sense that every boundary face of the mesh $\mathscr{T}$ is fully contained either in Γ_{D} or in $\overline{\Gamma_{\mathrm{N}}}$. Let $\boldsymbol{L}^2(\Omega) := L^2(\Omega;\mathbb{R}^d)$ and $\boldsymbol{H}(\mathrm{div},\Omega) := \{\boldsymbol{v} \in \boldsymbol{L}^2(\Omega),\ \nabla{\cdot}\boldsymbol{v} \in L^2(\Omega)\}$. Furthermore, we define the space $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega) := \{\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div},\Omega),\ \boldsymbol{v}{\cdot}\boldsymbol{n} = 0 \text{ on } \Gamma_{\mathrm{N}}\}$, where $\boldsymbol{v}{\cdot}\boldsymbol{n} = 0$

on Γ_{N} means that $\langle \boldsymbol{v}\cdot\boldsymbol{n}, \varphi\rangle_{\Gamma} = 0$ for all functions $\varphi \in H^1(\Omega)$ that have vanishing trace on Γ_{D}; here $\langle \boldsymbol{v}\cdot\boldsymbol{n}, \varphi\rangle_{\Gamma} := \int_{\Omega} [\boldsymbol{v}\cdot\nabla\varphi + (\nabla\cdot\boldsymbol{v})\varphi]$. For an open subset $\omega \subset \Omega$, let $\boldsymbol{L}^2(\omega) := L^2(\omega;\mathbb{R}^d)$ and $\boldsymbol{H}(\mathrm{div},\omega) := \{\boldsymbol{v} \in \boldsymbol{L}^2(\omega),\ \nabla\cdot\boldsymbol{v} \in L^2(\omega)\}$. We also denote by $(\cdot,\cdot)_{\omega}$ and $\|\cdot\|_{\omega}$ the L^2-inner product and norm for scalar- or vector-valued functions on ω. In the special case where $\omega = \Omega$, we drop the subscript, i.e., $(\cdot,\cdot) := (\cdot,\cdot)_{\Omega}$ and $\|\cdot\| := \|\cdot\|_{\Omega}$. The diameter of ω is denoted by h_{ω}, and its outward unit normal on its boundary by $\boldsymbol{n}_{\omega}$.

2.2 *Elements, vertices, faces, and patches of elements*

For any mesh element $K \in \mathcal{T}$, its diameter is denoted by h_K, and we set $h := \max_{K\in\mathcal{T}} h_K$. Let $\mathcal{V}_{\Omega}$ denote the set of interior vertices of $\mathcal{T}$, i.e., the vertices contained in Ω. Let $\mathcal{V}_{\Gamma}$ denote the set of vertices of $\mathcal{T}$ on the boundary Γ, and set $\mathcal{V} := \mathcal{V}_{\Omega} \cup \mathcal{V}_{\Gamma}$. We divide $\mathcal{V}_{\Gamma}$ into two disjoint sets $\mathcal{V}_{\mathrm{D}}$ and $\mathcal{V}_{\mathrm{N}}$, where $\mathcal{V}_{\mathrm{D}}$ contains all vertices in Γ_{D} (recalling that Γ_{D} is assumed to be closed) and $\mathcal{V}_{\mathrm{N}}$ consists of all vertices in Γ_{N}. For each vertex $\boldsymbol{a} \in \mathcal{V}$, define the patch $\mathcal{T}_{\boldsymbol{a}} := \{K \in \mathcal{T}, \boldsymbol{a}$ is a vertex of $K\}$ and the corresponding open subdomain $\omega_{\boldsymbol{a}} := \{\cup_{K\in\mathcal{T}_{\boldsymbol{a}}} K\}^{\circ}$. The piecewise affine Lagrange finite element basis function associated with a vertex $\boldsymbol{a} \in \mathcal{V}$ is denoted by $\psi_{\boldsymbol{a}}$. Let $\mathcal{F}$ denote the set of all $(d-1)$-dimensional faces of $\mathcal{T}$. By convention, we consider faces to be closed sets. For an element $K \in \mathcal{T}$, we denote the set of all faces of K by $\mathcal{F}_K$ and the set of all vertices of K by $\mathcal{V}_K$. For each interior vertex $\boldsymbol{a} \in \mathcal{V}_{\Omega}$, we let $\mathcal{F}_{\boldsymbol{a}}^{\mathrm{in}}$ denote the set of all faces that contain the vertex $\boldsymbol{a}$ (and thus do not lie on the boundary of $\omega_{\boldsymbol{a}}$). For boundary vertices $\boldsymbol{a} \in \mathcal{V}_{\Gamma}$, let $\mathcal{F}_{\boldsymbol{a}}^{\mathrm{in}}$ collect the faces that contain the vertex $\boldsymbol{a}$ but do not lie on the Dirichlet boundary Γ_D. The mesh shape-regularity parameter is defined as $\kappa_{\mathcal{T}} := \max_{K\in\mathcal{T}} h_K/\varrho_K$, where ϱ_K is the diameter of the largest ball inscribed in K.

2.3 *Piecewise polynomial and Raviart–Thomas–Nédélec spaces*

Let $p \geqslant 0$ be a non-negative integer. For $S \in \{K, F\}$, where $K \in \mathcal{T}$ is an element and $F \in \mathcal{F}$ is a face, we define $\mathcal{P}_p(S)$ as the space of all polynomials of total degree at most p on S. If $\widetilde{\mathcal{T}}$ denotes a subset of elements of $\mathcal{T}$, $\mathcal{P}_p(\widetilde{\mathcal{T}}) := \{v_h \in L^2(\Omega),\ v_h|_K \in \mathcal{P}_p(K)\ \ \forall K \in \widetilde{\mathcal{T}}\}$ is the space of piecewise polynomials of degree at most p over $\widetilde{\mathcal{T}}$. Typically, $\widetilde{\mathcal{T}}$ will be either the whole mesh $\mathcal{T}$ or the vertex patch $\mathcal{T}_{\boldsymbol{a}}$ as defined above. We define the piecewise Raviart–Thomas–Nédélec space $\boldsymbol{RTN}_p(\mathcal{T}) := \{\boldsymbol{\varsigma}_{\mathcal{T}} \in \boldsymbol{L}^2(\Omega),\ \boldsymbol{\varsigma}_{\mathcal{T}}|_K \in \boldsymbol{RTN}_p(K)\ \forall K \in \mathcal{T}\}$, where $\boldsymbol{RTN}_p(K) := \mathcal{P}_p(K;\mathbb{R}^d) + \boldsymbol{x}\mathcal{P}_p(K)$ and $\mathcal{P}_p(K;\mathbb{R}^d)$ denotes the space of $\mathbb{R}^d$-valued functions defined on K with each component being a polynomial of degree at most p in $\mathcal{P}_p(K)$. Note that with this choice of notation, functions in the space $\boldsymbol{RTN}_p(\mathcal{T})$ do not necessarily belong to $\boldsymbol{H}(\mathrm{div},\Omega)$; thus, $\boldsymbol{RTN}_p(\mathcal{T}) \cap \boldsymbol{H}(\mathrm{div},\Omega)$ is a proper subspace of $\boldsymbol{RTN}_p(\mathcal{T})$, which is classically characterized as those functions in $\boldsymbol{RTN}_p(\mathcal{T})$ having a continuous normal component across interior mesh faces. Moreover, $\boldsymbol{RTN}_p(\mathcal{T})$ is a subspace of $\boldsymbol{C}^1(\mathcal{T}) := \{\boldsymbol{v} \in \boldsymbol{L}^2(\Omega),\ \ \boldsymbol{v}|_K \in \boldsymbol{C}^1(K)\ \forall K \in \mathcal{T}\}$, the space of piecewise (broken) first-order componentwise differentiable vector-valued fields over $\mathcal{T}$. To avoid confusion between piecewise smooth and globally smooth functions, we denote the elementwise gradient and the elementwise divergence by $\nabla_{\mathcal{T}}$ and by $\nabla_{\mathcal{T}}\cdot$, respectively.

2.4 *L^2-orthogonal projection and elementwise canonical interpolant*

For each polynomial degree $p \geqslant 0$, let $\Pi_{\mathcal{T}}^p : L^2(\Omega) \to \mathcal{P}_p(\mathcal{T})$ denote the L^2-orthogonal projection of order p. Similarly, let Π_F^p denote the L^2-orthogonal projection of order p on a face $F \in \mathcal{F}$, which maps $L^2(F)$ to $\mathcal{P}_p(F)$. Let $\boldsymbol{I}_{\mathcal{T}}^p : \boldsymbol{C}^1(\mathcal{T}) \to \boldsymbol{RTN}_p(\mathcal{T})$ be the *elementwise* canonical (Raviart–Thomas–

Nédélec) interpolant. The domain of $\boldsymbol{I}^p_{\mathscr{T}}$ can be taken to be a (much) larger space than $\boldsymbol{C}^1(\mathscr{T})$, but not as large as piecewise $\boldsymbol{H}(\mathrm{div})$ fields; the present choice is sufficient for our purposes. For any $\boldsymbol{v} \in \boldsymbol{C}^1(\mathscr{T})$, the interpolant $\boldsymbol{I}^p_{\mathscr{T}}\boldsymbol{v}$ is defined separately on each element $K \in \mathscr{T}$ by the conditions

$$\begin{aligned} ((\boldsymbol{I}^p_{\mathscr{T}}\boldsymbol{v})|_K \cdot \boldsymbol{n}_K, q_K)_F &= (\boldsymbol{v}|_K \cdot \boldsymbol{n}_K, q_K)_F \quad && \forall q_K \in \mathscr{P}_p(F),\ \forall F \in \mathscr{F}_K, \\ (\boldsymbol{I}^p_{\mathscr{T}}\boldsymbol{v}, \boldsymbol{r}_K)_K &= (\boldsymbol{v}, \boldsymbol{r}_K)_K && \forall \boldsymbol{r}_K \in \mathscr{P}_{p-1}(K;\mathbb{R}^d), \end{aligned} \tag{2.1}$$

where $\boldsymbol{v}|_K \cdot \boldsymbol{n}_K$ denotes the normal trace of $\boldsymbol{v}|_K$, the restriction of $\boldsymbol{v}$ to K. Note that (2.1) implies that $((\boldsymbol{I}^p_{\mathscr{T}}\boldsymbol{v})|_K \cdot \boldsymbol{n}_K)|_F = \Pi^p_F((\boldsymbol{v}|_K \cdot \boldsymbol{n}_K)|_F)$ for all faces $F \subset \mathscr{F}_K$. A useful property of the operator $\boldsymbol{I}^p_{\mathscr{T}}$ is the commuting identity:

$$\nabla_{\mathscr{T}} \cdot (\boldsymbol{I}^p_{\mathscr{T}}\boldsymbol{v}) = \Pi^p_{\mathscr{T}}(\nabla_{\mathscr{T}} \cdot \boldsymbol{v}) \quad \forall \boldsymbol{v} \in \boldsymbol{C}^1(\mathscr{T}). \tag{2.2}$$

2.5 *Spaces for patchwise equilibration*

In the spirit of Braess *et al.* (2009) and Ern *et al.* (2017), Ern & Vohralík (2013, 2015), we finally define the local mixed finite element spaces $\boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ by

$$\boldsymbol{V}_p(\omega_{\boldsymbol{a}}) := \begin{cases} \{\boldsymbol{\varsigma_a} \in \boldsymbol{RTN}_p(\mathscr{T}_{\boldsymbol{a}}) \cap \boldsymbol{H}(\mathrm{div};\omega_{\boldsymbol{a}}),\ \boldsymbol{\varsigma_a} \cdot \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = 0 \text{ on } \partial\omega_{\boldsymbol{a}}\} & \text{if } \boldsymbol{a} \in \mathscr{V}_\Omega \cup \mathscr{V}_N, \\ \{\boldsymbol{\varsigma_a} \in \boldsymbol{RTN}_p(\mathscr{T}_{\boldsymbol{a}}) \cap \boldsymbol{H}(\mathrm{div};\omega_{\boldsymbol{a}}),\ \boldsymbol{\varsigma_a} \cdot \boldsymbol{n}_{\omega_{\boldsymbol{a}}} = 0 \text{ on } \partial\omega_{\boldsymbol{a}} \setminus \Gamma^{\boldsymbol{a}}_{\mathrm{D}}\} & \text{if } \boldsymbol{a} \in \mathscr{V}_D, \end{cases} \tag{2.3}$$

where $\Gamma^{\boldsymbol{a}}_{\mathrm{D}}$ contains those boundary faces from Γ_{D} that share the vertex $\boldsymbol{a}$. In particular, we observe that when $\partial\omega_{\boldsymbol{a}} \cap \Gamma_{\mathrm{N}} \neq \emptyset$, then $\boldsymbol{\varsigma_a} \cdot \boldsymbol{n} = 0$ on Γ_{N} for any $\boldsymbol{\varsigma_a} \in \boldsymbol{V}_p(\omega_{\boldsymbol{a}})$. As a result of the above definitions, it follows that the zero extension to all of Ω of any $\boldsymbol{\varsigma_a} \in \boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ belongs to $\boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$.

3. Main results

This section collects our main results.

3.1 *A simple stable local commuting projector in* $\boldsymbol{H}_{0,\Gamma_N}(\mathrm{div},\Omega)$

We first construct a simple, local, and stable commuting projector defined over the entire space $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ that leads to an approximation error equivalent to the local-best approximation error.

Recall the definition of the broken Raviart–Thomas–Nédélec interpolant $\boldsymbol{I}^p_{\mathscr{T}}$ from (2.1) and that of the piecewise polynomial patchwise $\boldsymbol{H}(\mathrm{div};\omega_{\boldsymbol{a}})$-conforming spaces $\boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ from (2.3). Recall also that zero extensions of elements of $\boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ belong to $\boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ and that $\psi_{\boldsymbol{a}}$ is the piecewise affine Lagrange finite element basis function associated with the vertex $\boldsymbol{a}$.

Definition 3.1 (A simple locally defined mapping from $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ to $\boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$). Let $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ be arbitrary. Let $\boldsymbol{\tau}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T})$ be defined elementwise by

$$\boldsymbol{\tau}_{\mathscr{T}}|_K := \mathop{\arg\min}_{\substack{\boldsymbol{\varsigma}_K \in \boldsymbol{RTN}_p(K) \\ \nabla \cdot \boldsymbol{\varsigma}_K = \Pi^p_{\mathscr{T}}(\nabla \cdot \boldsymbol{v})}} \|\boldsymbol{v} - \boldsymbol{\varsigma}_K\|_K \qquad \forall K \in \mathscr{T}. \tag{3.1}$$

For each mesh vertex $\boldsymbol{a} \in \mathscr{V}$, let $\boldsymbol{\sigma}_{\boldsymbol{a}} \in \boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ be defined by

$$\boldsymbol{\sigma}_{\boldsymbol{a}} := \underset{\substack{\boldsymbol{\varsigma}_{\boldsymbol{a}} \in \boldsymbol{V}_p(\omega_{\boldsymbol{a}}) \\ \nabla\cdot\boldsymbol{\varsigma}_{\boldsymbol{a}} = \Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla\cdot\boldsymbol{v} + \nabla\psi_{\boldsymbol{a}}\cdot\boldsymbol{\tau}_{\mathscr{T}})}}{\arg\min} \|\boldsymbol{\varsigma}_{\boldsymbol{a}} - \boldsymbol{I}^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}})\|_{\omega_{\boldsymbol{a}}}. \tag{3.2}$$

Extending $\boldsymbol{\sigma}_{\boldsymbol{a}}$ from the $\omega_{\boldsymbol{a}}$ to the rest of Ω by zero, we define $P^p_{\mathscr{T}}(\boldsymbol{v}) \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$ by

$$P^p_{\mathscr{T}}(\boldsymbol{v}) := \boldsymbol{\sigma}_{\mathscr{T}} := \sum_{\boldsymbol{a}\in\mathscr{V}} \boldsymbol{\sigma}_{\boldsymbol{a}}. \tag{3.3}$$

The justification that the construction of $P^p_{\mathscr{T}}(\boldsymbol{v})$ is well defined is given in Section 4.1 below. The first step (3.1) considers the elementwise $\boldsymbol{L}^2$-norm local-best approximation that defines the *discontinuous piecewise RTN polynomial* $\boldsymbol{\tau}_{\mathscr{T}}$ closest to $\boldsymbol{v}$ under the divergence constraint. At this step, crucially, the minimal regularity $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$ is sufficient; note that we only work with $\boldsymbol{v}$ volumewise and that no normal component of $\boldsymbol{v}$ on a face is requested to exist, in contrast to the common interpolation operators discussed in Section 1. Essentially, this step brings us into a piecewise polynomial setting, where we will stay henceforth, with $\boldsymbol{\tau}_{\mathscr{T}}$ being the best approximation on each mesh element $K \in \mathscr{T}$.

As $\boldsymbol{\tau}_{\mathscr{T}}$ from (3.1) is discontinuous from one mesh element to the other, the second step in (3.2) can be seen as *smoothing* $\boldsymbol{\tau}_{\mathscr{T}}$ over the vertex patch subdomains $\omega_{\boldsymbol{a}}$ to obtain an $\boldsymbol{H}(\mathrm{div}; \omega_{\boldsymbol{a}})$-conforming approximation $\boldsymbol{\sigma}_{\boldsymbol{a}}$ with a suitably prescribed divergence. An important ingredient is the employment of the elementwise canonical Raviart–Thomas–Nédélec interpolant $\boldsymbol{I}^p_{\mathscr{T}}$ (this is well justified, since the argument $\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}$ is a discrete object). The crucial role of $\boldsymbol{I}^p_{\mathscr{T}}$ is to decrease the order of $\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}$, which lies in $\boldsymbol{RTN}_{p+1}(\mathscr{T})$ because of the multiplication by the hat function $\psi_{\boldsymbol{a}}$, back to $\boldsymbol{RTN}_p(\mathscr{T})$. A similar construction below, see (5.3), will avoid the use of $\boldsymbol{I}^p_{\mathscr{T}}$ and will be crucial for our *hp*-optimal approximation estimates. Finally, in the third step (3.3), the approximations $\boldsymbol{\sigma}_{\boldsymbol{a}}$ are summed into $P^p_{\mathscr{T}}(\boldsymbol{v})$, thereby producing an $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$-conforming piecewise polynomial from the RTN space of order p. The overall procedure is motivated by *equilibrated flux reconstructions* coming from *a posteriori* error estimation, as in Destuynder & Métivet (1999), Braess & Schöberl (2008), Ern & Vohralík (2013). Here we adapt those techniques to the purpose of *a priori* error analysis.

Our first main result, whose proof is postponed to Section 4, is the following.

THEOREM 3.2 (Commutativity, projection, approximation, and stability of $P^p_{\mathscr{T}}$). Let a mesh $\mathscr{T}$ of Ω and a polynomial degree $p \geqslant 0$ be fixed. Then the operator $P^p_{\mathscr{T}}$ from Definition 3.1 maps $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$ to $\boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$ and

$$\nabla\cdot P^p_{\mathscr{T}}(\boldsymbol{v}) = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v}) \qquad \forall\, \boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega), \tag{3.4}$$

$$P^p_{\mathscr{T}}(\boldsymbol{v}) = \boldsymbol{v} \qquad \forall\, \boldsymbol{v} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega). \tag{3.5}$$

Thus, $P^p_{\mathscr{T}}$ is a projection from $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ onto $\boldsymbol{RTN}_p(\mathscr{T})\cap\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ such that the commuting property (3.4) is satisfied; this can be cast into the commuting diagram

$$\begin{array}{ccc}
\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega) & \xrightarrow{\nabla\cdot} & L^2(\Omega) \\
\Big\downarrow P^p_{\mathscr{T}} & & \Big\downarrow \Pi^p_{\mathscr{T}} \\
\boldsymbol{RTN}_p(\mathscr{T})\cap\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega) & \xrightarrow{\nabla\cdot} & \mathscr{P}_p(\mathscr{T}).
\end{array}$$

Furthermore, for any $\boldsymbol{v}\in\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ and any $K\in\mathscr{T}$, we have the local approximation and stability bounds

$$\begin{aligned}
&\|\boldsymbol{v}-P^p_{\mathscr{T}}(\boldsymbol{v})\|_K^2+\left[\frac{h_K}{p+1}\|\nabla\cdot(\boldsymbol{v}-P^p_{\mathscr{T}}(\boldsymbol{v}))\|_K\right]^2 \\
&\qquad\leqslant C\sum_{K'\in\mathscr{T}_K}\left\{\min_{\boldsymbol{\varsigma}_{K'}\in\boldsymbol{RTN}_p(K')}\|\boldsymbol{v}-\boldsymbol{\varsigma}_{K'}\|_{K'}^2+\left[\frac{h_{K'}}{p+1}\|\nabla\cdot\boldsymbol{v}-\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_{K'}\right]^2\right\},
\end{aligned}\tag{3.6}$$

$$\|P^p_{\mathscr{T}}(\boldsymbol{v})\|_K^2\leqslant C\sum_{K'\in\mathscr{T}_K}\left\{\|\boldsymbol{v}\|_{K'}^2+\left[\frac{h_{K'}}{p+1}\|\nabla\cdot\boldsymbol{v}-\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_{K'}\right]^2\right\},\tag{3.7}$$

$$\|P^p_{\mathscr{T}}(\boldsymbol{v})\|_K^2+h_\Omega^2\|\nabla\cdot P^p_{\mathscr{T}}(\boldsymbol{v})\|_K^2\leqslant C\sum_{K'\in\mathscr{T}_K}\left\{\|\boldsymbol{v}\|_{K'}^2+h_\Omega^2\|\nabla\cdot\boldsymbol{v}\|_{K'}^2\right\},\tag{3.8}$$

where $\mathscr{T}_K:=\cup_{\boldsymbol{a}\in\mathscr{V}_K}\mathscr{T}_{\boldsymbol{a}}$ are the neighboring elements of K, and recalling that h_Ω denotes the diameter of Ω. The constant C above only depends on the space dimension d, the shape-regularity parameter $\kappa_{\mathscr{T}}$ of $\mathscr{T}$, and the polynomial degree p.

Property (3.7) readily implies that $P^p_{\mathscr{T}}$ is globally $\boldsymbol{L}^2$-stable up to *hp* data oscillation of the divergence, since summing (3.7) over the mesh elements leads to

$$\|P^p_{\mathscr{T}}(\boldsymbol{v})\|^2\leqslant C\left\{\|\boldsymbol{v}\|^2+\sum_{K\in\mathscr{T}}\left[\frac{h_K}{p+1}\|\nabla\cdot\boldsymbol{v}-\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K\right]^2\right\}\qquad\forall\,\boldsymbol{v}\in\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega).\tag{3.9a}$$

Similarly, from (3.8) we infer that $P^p_{\mathscr{T}}$ is $\boldsymbol{H}(\mathrm{div})$-stable, since

$$\|P^p_{\mathscr{T}}(\boldsymbol{v})\|^2+h_\Omega^2\|\nabla\cdot P^p_{\mathscr{T}}(\boldsymbol{v})\|^2\leqslant C\big[\|\boldsymbol{v}\|^2+h_\Omega^2\|\nabla\cdot\boldsymbol{v}\|^2\big]\qquad\forall\,\boldsymbol{v}\in\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega).\tag{3.9b}$$

The projector $P^p_{\mathscr{T}}$ in Definition 3.1 and Theorem 3.2 improves on Christiansen & Winther (2008) in that the construction is *local*, and on Falk & Winther (2014) in that it is, up to data oscillation, *stable in* $\boldsymbol{L}^2$ (i.e., with respect to $\|\boldsymbol{v}\|$), see (3.9a), rather than only in $\boldsymbol{H}(\mathrm{div})$ (i.e., with respect to $\big[\|\boldsymbol{v}\|^2+h_\Omega^2\|\nabla\cdot\boldsymbol{v}\|^2\big]^{1/2}$), see (3.9b). We note more precisely that, for the divergence term, (3.7) improves the bound of Falk & Winther (2014, Theorem 5.2, (5.2)) since, in particular, we have $\|\nabla\cdot\boldsymbol{v}-\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K$ in place of $\|\nabla\cdot\boldsymbol{v}\|_K$, whereas (3.8) is similar to the combination of the bounds of Falk & Winther (2014, Theorem 5.2, (5.2) and (5.3)). Additionally, the terms with $\|\nabla\cdot\boldsymbol{v}-\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K$ are of '*hp* data oscillation' type, containing the weight factors $h_K/(p+1)$, so that they tend optimally to zero for both $h\to 0$ and $p\to\infty$. The projection operator $P^p_{\mathscr{T}}$ defined here also satisfies the *commuting property* with the divergence operator (3.4), in contrast to Ern & Guermond (2017).

3.2 *Equivalence of local- and global-best approximations in $\boldsymbol{H}_{0,\Gamma_N}(\mathrm{div},\Omega)$*

For any function $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$, we consider the *global-best approximation error* $E_{\mathscr{T},p}(\boldsymbol{v})$ defined as the best approximation, in a weighted norm, from $\boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$, subject to a constraint on the divergence:

$$[E_{\mathscr{T},p}(\boldsymbol{v})]^2 := \min_{\substack{\boldsymbol{\varsigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega) \\ \nabla\cdot\boldsymbol{\varsigma}_{\mathscr{T}} = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})}} \|\boldsymbol{v} - \boldsymbol{\varsigma}_{\mathscr{T}}\|^2_{\Omega} + \sum_{K\in\mathscr{T}} \left[\frac{h_K}{p+1}\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K\right]^2. \tag{3.10}$$

Note that the global minimization in (3.10) is subject to three constraints: the global minimizer $\boldsymbol{\varsigma}_{\mathscr{T}}$ has to have the normal trace continuous across the mesh faces, the normal trace on Γ_{N} equal to zero (whenever relevant), and the divergence equal to $\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})$. We further consider the *local-best approximation errors* defined on each element $K \in \mathscr{T}$ by

$$[e_{K,p}(\boldsymbol{v})]^2 := \min_{\boldsymbol{\varsigma}_K \in \boldsymbol{RTN}_p(K)} \|\boldsymbol{v} - \boldsymbol{\varsigma}_K\|^2_K + \left[\frac{h_K}{p+1}\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K\right]^2. \tag{3.11}$$

Note that the local minimization in (3.11) is completely *constraint-free:* the local minimizer $\boldsymbol{\varsigma}_K$ is completely free on the faces of the mesh element K, including those in Γ_{N} (whenever relevant), and neither is a subject to any divergence constraint. Furthermore, since $\Pi^p_{\mathscr{T}}$ is the L^2-orthogonal projection onto the broken polynomial space $\mathscr{P}_p(\mathscr{T})$, we have $\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K = \min_{q\in\mathscr{P}_p(K)}\|\nabla\cdot\boldsymbol{v} - q\|_K$. Thus, the local approximation errors $e_{K,p}(\boldsymbol{v})$ involve the local-best approximation errors in $\boldsymbol{L}^2$ plus a weighted L^2 best approximation error of the divergence.

In a direct consequence of Theorem 3.2, we now show that the global-best error $E_{\mathscr{T},p}(\boldsymbol{v})$ is in fact *equivalent* to the root-mean square sum of the local-best errors $e_{K,p}(\boldsymbol{v})$ over all elements of the mesh. This may seem surprising at a first sight, since the latter might initially seem to be (much) smaller.

THEOREM 3.3 (Equivalence of local- and global-best approximations). There exists a constant C depending only on the space dimension d, the shape-regularity parameter $\kappa_{\mathscr{T}}$ of $\mathscr{T}$, and the polynomial degree $p \geqslant 0$, such that, for any $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$,

$$\left[E_{\mathscr{T},p}(\boldsymbol{v})\right]^2 \leqslant C \sum_{K\in\mathscr{T}} \left[e_{K,p}(\boldsymbol{v})\right]^2 \leqslant C\left[E_{\mathscr{T},p}(\boldsymbol{v})\right]^2. \tag{3.12}$$

Proof. The second inequality in (3.12) follows immediately from the definitions in (3.10) and (3.11): indeed, the global minimization set is (much) smaller than the local ones. To prove the first inequality, consider an arbitrary function $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$. Then Theorem 3.2 shows that the projection $P^p_{\mathscr{T}}(\boldsymbol{v}) \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ satisfies the constraints of the global minimization set in (3.10) due to its commuting property (3.4). Therefore, it suffices to pick the function $P^p_{\mathscr{T}}(\boldsymbol{v})$ from the minimization set, sum the bound in the local approximation property (3.6) over all mesh elements, and invoke the shape-regularity of the mesh, which implies that the number of neighbors a mesh cell can have is uniformly bounded from above. □

REMARK 3.4 (Necessity of the divergence error terms). Although the scaled divergence terms $\frac{h_K}{p+1}\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K$ take an identical form in both $E_{\mathscr{T},p}(\boldsymbol{v})$ and $e_{K,p}(\boldsymbol{v})$, they cannot be removed from the local contributions $e_{K,p}(\boldsymbol{v})$. Otherwise, it would be possible to choose a sequence of functions $\boldsymbol{v}$ in $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ approaching a function $\boldsymbol{\tau}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T})$ but $\boldsymbol{\tau}_{\mathscr{T}} \notin \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ such that the middle term in (3.12) would tend to zero, but $E_{\mathscr{T},p}(\boldsymbol{v})$ would remain uniformly bounded away from zero.

REMARK 3.5 (Equivalence with constraint on the right-hand side). Theorem 3.3 also straightforwardly implies the slightly weaker property

$$\begin{aligned}\left[E_{\mathscr{T},p}(\boldsymbol{v})\right]^2 &\leqslant C\sum_{K\in\mathscr{T}}\left\{\min_{\substack{\boldsymbol{\varsigma}_K\in\boldsymbol{RTN}_p(K)\\ \nabla\cdot\boldsymbol{\varsigma}_K=\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})|_K}}\|\boldsymbol{v}-\boldsymbol{\varsigma}_K\|_K^2+\left[\frac{h_K}{p+1}\|\nabla\cdot\boldsymbol{v}-\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K\right]^2\right\}\\ &\leqslant C\left[E_{\mathscr{T},p}(\boldsymbol{v})\right]^2\end{aligned}$$

with the same constant C, where the minimization problems in the middle term include a constraint on the divergence to mirror the divergence constraint in $E_{\mathscr{T},p}(\boldsymbol{v})$.

3.3 *Optimal-order hp-approximation estimates in* $\boldsymbol{H}_{0,\Gamma_N}(\mathrm{div},\Omega)$

We finally focus on functions with some additional elementwise regularity. For any $s\geqslant 0$ and any mesh element $K\in\mathscr{T}$, let $\boldsymbol{H}^s(K)$ denote the space of vector-valued fields in $\boldsymbol{L}^2(K)$ with each component in $H^s(K)$. Recall the definition (3.10) of $E_{\mathscr{T},p}(\boldsymbol{v})$. Our third and last main result, whose proof is postponed to Section 5, delivers *hp-optimal* convergence rates for vector fields in $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ with the minimally necessary additional elementwise $\boldsymbol{H}^s(K)$-regularity.

THEOREM 3.6 (*hp*-optimal approximation estimates under minimal $\boldsymbol{H}^s(K)$-regularity). Let $s\geqslant 0$ and let $\boldsymbol{v}\in\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ be such that

$$\boldsymbol{v}|_K\in\boldsymbol{H}^s(K)\quad\forall K\in\mathscr{T}.$$

Let the polynomial degree $p\geqslant 0$. Then there exists a constant C, depending only on the regularity exponent s, the space dimension d, and the shape-regularity parameter $\kappa_{\mathscr{T}}$ of $\mathscr{T}$, such that

$$[E_{\mathscr{T},p}(\boldsymbol{v})]^2\leqslant C\left\{\sum_{K\in\mathscr{T}}\left[\frac{h_K^{\min(s,p+1)}}{(p+1)^s}\|\boldsymbol{v}\|_{\boldsymbol{H}^s(K)}\right]^2+\delta_{s<1}\left[\frac{h_K}{p+1}\|\nabla\cdot\boldsymbol{v}\|_K\right]^2\right\},\tag{3.13}$$

where $\delta_{s<1}:=1$ if $s<1$ and $\delta_{s<1}:=0$ if $s\geqslant 1$.

Notice that the above generic constant C is independent of p and that the bound (3.13) is optimal with respect to both the mesh size h and the polynomial degree p for arbitrary regularity index $s\geqslant 0$.

4. Proof of Theorem 3.2 (commutativity, projection, approximation, and stability of $P^p_{\mathscr{T}}$)

The proof of Theorem 3.2 is split into several parts. First, in Section 4.1 we analyse essential properties of the construction of the mapping $P^p_{\mathscr{T}}$ from Definition 3.1. We next establish the commuting property (3.4) in Section 4.2. Then, in Section 4.3, we prove the statement (3.6) on the approximation properties of $P^p_{\mathscr{T}}$. This is the most technical part of the proof. Finally, in Section 4.4 we conclude by proving the remaining three statements (3.5), (3.7), and (3.8) (the projection property, $\boldsymbol{L}^2$ stability, and $\boldsymbol{H}(\mathrm{div})$ stability).

4.1 *Justification of the construction of* $P_{\mathscr{T}}^p$

We start by showing that the operator $P_{\mathscr{T}}^p$ of Definition 3.1 is well defined on $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$. Recall the notation from Section 2.2.

LEMMA 4.1 (Discrete weak divergence of $\boldsymbol{L}^2$-projection). For any function $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$, let $\boldsymbol{\tau}_{\mathscr{T}}$ be defined elementwise in (3.1). Then

$$(\nabla\cdot\boldsymbol{v},\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\tau}_{\mathscr{T}},\nabla\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} = 0 \quad \forall \boldsymbol{a} \in \mathscr{V}_\Omega \cup \mathscr{V}_N. \tag{4.1}$$

Proof. First, observe that for any vertex $\boldsymbol{a} \in \mathscr{V}_\Omega \cup \mathscr{V}_{\mathrm{N}}$, the hat function $\psi_{\boldsymbol{a}}$ belongs to $H^1_{\Gamma_{\mathrm{D}}}(\Omega)$ owing to the conformity of $\mathscr{T}$ with respect to the Dirichlet and Neumann boundary sets. Therefore, $(\nabla\cdot\boldsymbol{v},\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} + (\boldsymbol{v},\nabla\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} = 0$, where we use the fact that $\omega_{\boldsymbol{a}}$ is the support of $\psi_{\boldsymbol{a}}$. Since $\nabla\psi_{\boldsymbol{a}}$ is a constant vector on each element K, the Euler–Lagrange equations for (3.1) imply that

$$(\boldsymbol{\tau}_{\mathscr{T}},\nabla\psi_{\boldsymbol{a}})_K = (\boldsymbol{v},\nabla\psi_{\boldsymbol{a}})_K \qquad \forall K \in \mathscr{T}_{\boldsymbol{a}}. \tag{4.2}$$

Consequently, $(\boldsymbol{\tau}_{\mathscr{T}},\nabla\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} = (\boldsymbol{v},\nabla\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}}$, and (4.1) follows. □

We now show that the local minimization problems (3.2) give well-defined local contributions $\boldsymbol{\sigma}_{\boldsymbol{a}}$.

LEMMA 4.2 (Existence and uniqueness of local problems (3.2)). For each vertex $\boldsymbol{a} \in \mathscr{V}$, there exists a unique $\boldsymbol{\sigma}_{\boldsymbol{a}} \in \boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ satisfying (3.2).

Proof. The minimization problem (3.2) is equivalent to a mixed finite element problem in the patch subdomain $\omega_{\boldsymbol{a}}$. For Dirichlet boundary vertices $\boldsymbol{a} \in \mathscr{V}_{\mathrm{D}}$, this problem is well posed with a unique minimizer since the space $\boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ of (2.3) does not impose the normal constraint everywhere on $\partial\omega_{\boldsymbol{a}}$. For interior and Neumann vertices $\boldsymbol{a} \in \mathscr{V}_\Omega \cup \mathscr{V}_{\mathrm{N}}$, the source term in the divergence constraint satisfies the compatibility condition

$$(\Pi_{\mathscr{T}}^p(\psi_{\boldsymbol{a}}\nabla\cdot\boldsymbol{v} + \nabla\psi_{\boldsymbol{a}}\cdot\boldsymbol{\tau}_{\mathscr{T}}),1)_{\omega_{\boldsymbol{a}}} = (\nabla\cdot\boldsymbol{v},\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\tau}_{\mathscr{T}},\nabla\psi_{\boldsymbol{a}})_{\omega_{\boldsymbol{a}}} = 0,$$

where the second equality follows from Lemma 4.1. Therefore, $\boldsymbol{\sigma}_{\boldsymbol{a}}$ is also well defined for interior and Neumann vertices $\boldsymbol{a} \in \mathscr{V}_\Omega \cup \mathscr{V}_{\mathrm{N}}$. □

It follows from Lemma 4.2 that $P_{\mathscr{T}}^p(\boldsymbol{v}) \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ is well defined for every $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$.

4.2 *Proof of the commuting property* (3.4)

We are now ready to establish the following result.

LEMMA 4.3 (Commuting property (3.4)). $P_{\mathscr{T}}^p$ satisfies (3.4).

Proof. Since the functions $\{\psi_{\boldsymbol{a}}\}_{\boldsymbol{a}\in\mathscr{V}}$ form a partition of unity over Ω, i.e., $\sum_{\boldsymbol{a}\in\mathscr{V}}\psi_{\boldsymbol{a}} = 1$, and consequently $\sum_{\boldsymbol{a}\in\mathscr{V}}\nabla\psi_{\boldsymbol{a}} = \mathbf{0}$, we find from (3.3) and (3.2) that

$$\nabla{\cdot}P^p_{\mathscr{T}}(\boldsymbol{v}) = \sum_{\boldsymbol{a}\in\mathscr{V}}\nabla{\cdot}\boldsymbol{\sigma}_{\boldsymbol{a}} = \sum_{\boldsymbol{a}\in\mathscr{V}}\left\{\Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla{\cdot}\boldsymbol{v} + \nabla\psi_{\boldsymbol{a}}{\cdot}\boldsymbol{\tau}_{\mathscr{T}})\right\} = \Pi^p_{\mathscr{T}}(\nabla{\cdot}\boldsymbol{v}). \tag{4.3}$$

□

4.3 *Proof of the approximation property* (3.6)

Let us start with two useful technical results. For a given vertex $\boldsymbol{a} \in \mathscr{V}$, let the space $H^1_*(\omega_{\boldsymbol{a}})$ be defined by

$$H^1_*(\omega_{\boldsymbol{a}}) := \begin{cases}\{\varphi \in H^1(\omega_{\boldsymbol{a}}), (\varphi,1)_{\omega_{\boldsymbol{a}}} = 0\} & \text{if } \boldsymbol{a} \in \mathscr{V}_\Omega \cup \mathscr{V}_N,\\ \{\varphi \in H^1(\omega_{\boldsymbol{a}}), \varphi|_{\partial\omega_{\boldsymbol{a}}\cap\Gamma^{\boldsymbol{a}}_{\mathrm{D}}} = 0\} & \text{if } \boldsymbol{a} \in \mathscr{V}_D,\end{cases} \tag{4.4}$$

where we recall that $\Gamma^{\boldsymbol{a}}_{\mathrm{D}}$ contains those boundary faces from Γ_{D} that share the vertex $\boldsymbol{a}$. Recall also the discrete spaces $\boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ defined in (2.3). The following result has been shown in Braess *et al.* (2009, Theorem 7) in two space dimensions and Ern & Vohralík (2020, Corollaries 3.3, 3.6, and 3.8) in three space dimensions, crucially building on Costabel & McIntosh (2010) and Demkowicz *et al.* (2012) to achieve the independence of the involved constant on the polynomial degree p.

Lemma 4.4 (Stability of patchwise flux equilibration). Let a vertex $\boldsymbol{a} \in \mathscr{V}$ be fixed, and let $g_{\boldsymbol{a}} \in \mathscr{P}_p(\mathscr{T}_{\boldsymbol{a}})$ and $\boldsymbol{\tau}_{\boldsymbol{a}} \in \boldsymbol{RTN}_p(\mathscr{T}_{\boldsymbol{a}})$ be given discontinuous piecewise polynomials with the condition $(g_{\boldsymbol{a}},1)_{\omega_{\boldsymbol{a}}} = 0$ if $\boldsymbol{a} \in \mathscr{V}_\Omega \cup \mathscr{V}_{\mathrm{N}}$. Then there exists a constant C, depending only on the space dimension d and the mesh shape-regularity parameter $\kappa_{\mathscr{T}}$, such that

$$\min_{\substack{\boldsymbol{v}_{\boldsymbol{a}}\in\boldsymbol{V}_p(\omega_{\boldsymbol{a}})\\ \nabla\cdot\boldsymbol{v}_{\boldsymbol{a}}=g_{\boldsymbol{a}}}} \|\boldsymbol{v}_{\boldsymbol{a}} - \boldsymbol{\tau}_{\boldsymbol{a}}\|_{\omega_{\boldsymbol{a}}} \leqslant C \sup_{\substack{\varphi\in H^1_*(\omega_{\boldsymbol{a}})\\ \|\nabla\varphi\|_{\omega_{\boldsymbol{a}}}=1}} \left\{(g_{\boldsymbol{a}},\varphi)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\tau}_{\boldsymbol{a}},\nabla\varphi)_{\omega_{\boldsymbol{a}}}\right\}.$$

We shall also use the following auxiliary bound for face terms based on the bubble-function technique of Verfürth, cf. Verfürth (2013), from *a posteriori* error analysis.

Lemma 4.5 (Bound on face terms). Let a mesh face $F \in \mathscr{F}$ be fixed, and let $\mathscr{T}_F$ be the set of one or two mesh elements $K \in \mathscr{T}$ to which F belongs, with ω_F the corresponding open subdomain. Let h_F denote the diameter of F. Then, there exists a constant C, depending only on the space dimension d, the mesh shape-regularity parameter $\kappa_{\mathscr{T}}$, and the polynomial degree p, such that

$$h_F^{1/2}\|q_h\|_F \leqslant C \sup_{\substack{\varphi\in H^1(\omega_F)\\ \varphi=0 \text{ on } \partial\omega_F\setminus F\\ \|\nabla\varphi\|_{\omega_F}=1}} (q_h,\varphi)_F \qquad \forall q_h \in \mathscr{P}_p(F).$$

We are now ready to prove statement (3.6) from Theorem 3.2, where we now employ the shorthand notation $e_{K,p}(\boldsymbol{v})$ from (3.11). Let $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ be arbitrary. Since it follows from

$\nabla{\cdot}P^p_{\mathcal{T}}(\boldsymbol{v}) = \Pi^p_{\mathcal{T}}(\nabla{\cdot}\boldsymbol{v})$ that

$$\frac{h_K}{p+1}\|\nabla{\cdot}\boldsymbol{v} - \nabla{\cdot}P^p_{\mathcal{T}}(\boldsymbol{v})\|_K \leqslant e_{K,p}(\boldsymbol{v}),$$

it only remains to prove that

$$\|\boldsymbol{v} - P^p_{\mathcal{T}}(\boldsymbol{v})\|_K \leqslant C\left\{\sum_{K'\in\mathcal{T}_K} e_{K',p}(\boldsymbol{v})^2\right\}^{\frac{1}{2}} \qquad \forall K\in\mathcal{T}. \tag{4.5}$$

For this purpose we proceed in two steps.

Step 1: Bound on $\boldsymbol{\sigma}_{\boldsymbol{a}}$. Recall that $\boldsymbol{\sigma}_{\boldsymbol{a}}$ is defined in (3.2) with $\boldsymbol{\tau}_{\mathcal{T}}$ defined elementwise in (3.1).

LEMMA 4.6 (Bound on $\boldsymbol{\sigma}_{\boldsymbol{a}}$). There exists a constant C, depending only on d, $\kappa_{\mathcal{T}}$ and p, such that

$$\|\boldsymbol{\sigma}_{\boldsymbol{a}} - \boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}})\|_{\omega_{\boldsymbol{a}}} \leqslant C\left\{\sum_{K\in\mathcal{T}_{\boldsymbol{a}}} [e_{K,p}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}} \qquad \forall \boldsymbol{a}\in\mathcal{V}. \tag{4.6}$$

Proof. First, since $\boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}}) \in \boldsymbol{RTN}_p(\mathcal{T}_{\boldsymbol{a}})$, we can apply Lemma 4.4 to $\boldsymbol{\sigma}_{\boldsymbol{a}}$, with the choices $\boldsymbol{\tau}_{\boldsymbol{a}} := \boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}})$ and $g_{\boldsymbol{a}} := \Pi^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\nabla{\cdot}\boldsymbol{v} + \nabla\psi_{\boldsymbol{a}}{\cdot}\boldsymbol{\tau}_{\mathcal{T}}) \in \mathcal{P}_p(\mathcal{T}_{\boldsymbol{a}})$ to obtain

$$\|\boldsymbol{\sigma}_{\boldsymbol{a}} - \boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}})\|_{\omega_{\boldsymbol{a}}} \leqslant C \sup_{\substack{\varphi\in H^1_*(\omega_{\boldsymbol{a}})\\ \|\nabla\varphi\|_{\omega_{\boldsymbol{a}}}=1}} \left\{(g_{\boldsymbol{a}},\varphi)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}}),\nabla\varphi)_{\omega_{\boldsymbol{a}}}\right\}, \tag{4.7}$$

where the space $H^1_*(\omega_{\boldsymbol{a}})$ is defined in (4.4). Let $h_{\omega_{\boldsymbol{a}}}$ denote the diameter of $\omega_{\boldsymbol{a}}$ and recall the Poincaré inequality $\|v\|_{\omega_{\boldsymbol{a}}} \leqslant Ch_{\omega_{\boldsymbol{a}}}\|\nabla v\|_{\omega_{\boldsymbol{a}}}$ on $H^1_*(\omega_{\boldsymbol{a}})$, with a constant C depending only on the dimension d and on $\kappa_{\mathcal{T}}$. Moreover, note that the shape-regularity of the mesh implies that $h_{\omega_{\boldsymbol{a}}} \approx h_K \approx h_F$ for all $K\in\mathcal{T}_{\boldsymbol{a}}$ and all $F\in\mathcal{F}^{\mathrm{in}}_{\boldsymbol{a}}$.

Define, for any $\boldsymbol{\varsigma}_{\mathcal{T}} \in \boldsymbol{C}^1(\mathcal{T})$, the jump $[\![\boldsymbol{\varsigma}_{\mathcal{T}}]\!]$ on an interior face F shared by two mesh elements K_+ and K_- by $[\![\boldsymbol{\varsigma}_{\mathcal{T}}]\!] := (\boldsymbol{\varsigma}_{\mathcal{T}}|_{K_+})|_F - (\boldsymbol{\varsigma}_{\mathcal{T}}|_{K_-})|_F$; here $\boldsymbol{n}_F := \boldsymbol{n}_{K_-} = -\boldsymbol{n}_{K_+}$ is the unit normal to F that points outward K_- and inward K_+. Similarly, if F is a boundary face, then we define $[\![\boldsymbol{\varsigma}_{\mathcal{T}}]\!] := \boldsymbol{\varsigma}_{\mathcal{T}}|_F$. To bound the right-hand side of (4.7) consider an arbitrary $\varphi\in H^1_*(\omega_{\boldsymbol{a}})$ such that $\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} = 1$. Then using integration by parts elementwise, we find that

$$\begin{aligned}\left(\boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}}),\nabla\varphi\right)_{\omega_{\boldsymbol{a}}} &= \sum_{F\in\mathcal{F}^{\mathrm{in}}_{\boldsymbol{a}}} \left([\![\boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}})]\!]{\cdot}\boldsymbol{n}_F,\varphi\right)_F - \sum_{K\in\mathcal{T}_{\boldsymbol{a}}} \left(\nabla{\cdot}\boldsymbol{I}^p_{\mathcal{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}}),\varphi\right)_K\\ &= \sum_{F\in\mathcal{F}^{\mathrm{in}}_{\boldsymbol{a}}} \left(\Pi^p_F(\psi_{\boldsymbol{a}}[\![\boldsymbol{\tau}_{\mathcal{T}}]\!]{\cdot}\boldsymbol{n}_F),\varphi\right)_F - \left(\Pi^p_{\mathcal{T}}(\nabla_{\mathcal{T}}{\cdot}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathcal{T}})),\varphi\right)_{\omega_{\boldsymbol{a}}}.\end{aligned}$$

Here, in the first identity, the set of faces can be restricted to $\mathcal{F}^{\mathrm{in}}_{\boldsymbol{a}}$; indeed, for interior vertices, this follows from the fact that $\psi_{\boldsymbol{a}}$ vanishes on $\partial\omega_{\boldsymbol{a}}$ whereas, for boundary vertices, $\varphi\in H^1_*(\omega_{\boldsymbol{a}})$ vanishes on $\Gamma^{\boldsymbol{a}}_{\mathrm{D}}$.

The second identity is then obtained from the definition of the elementwise canonical interpolant $\boldsymbol{I}^p_{\mathscr{T}}$ in (2.1) and the commutation identity (2.2). Expanding $\nabla_{\mathscr{T}}\cdot(\psi_a \boldsymbol{\tau}_{\mathscr{T}}) = \nabla\psi_{\boldsymbol{a}}\cdot\boldsymbol{\tau}_{\mathscr{T}} + \psi_{\boldsymbol{a}}\nabla_{\mathscr{T}}\cdot\boldsymbol{\tau}_{\mathscr{T}}$ and simplifying give

$$
\begin{aligned}
(g_{\boldsymbol{a}}, \varphi)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{I}^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}), \nabla\varphi)_{\omega_{\boldsymbol{a}}} &= (\Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla_{\mathscr{T}}\cdot(\boldsymbol{v}-\boldsymbol{\tau}_{\mathscr{T}})), \varphi)_{\omega_{\boldsymbol{a}}} \\
&\quad + \sum_{F\in\mathscr{F}^{\mathrm{in}}_{\boldsymbol{a}}} \left(\Pi^p_F(\psi_{\boldsymbol{a}}[\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F), \varphi\right)_F.
\end{aligned} \tag{4.8}
$$

We now bound the two terms on the right-hand side of (4.8) separately.

To bound the first term, we consider first the case $p \geqslant 1$: using the divergence constraint on $\boldsymbol{\tau}_{\mathscr{T}}$ in (3.1), the orthogonality of the L^2-projections, the approximation bound $\|\varphi - \Pi^{p-1}_{\mathscr{T}}\varphi\|_K \leqslant C\frac{h_K}{p+1}\|\nabla\varphi\|_K$ (note that $\frac{1}{p} \leq \frac{2}{p+1}$ for all $p \geqslant 1$), along with $\|\psi_{\boldsymbol{a}}\|_{\infty,\omega_{\boldsymbol{a}}} = 1$ and $\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} = 1$, we find that there is a constant C, depending only on d and $\kappa_{\mathscr{T}}$, such that

$$
\begin{aligned}
(\Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla_{\mathscr{T}}\cdot(\boldsymbol{v}-\boldsymbol{\tau}_{\mathscr{T}})), \varphi)_{\omega_{\boldsymbol{a}}} &= \left(\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v}), \psi_{\boldsymbol{a}}\Pi^p_{\mathscr{T}}(\varphi - \Pi^{p-1}_{\mathscr{T}}\varphi)\right)_{\omega_{\boldsymbol{a}}} \\
&\leqslant C\left\{\sum_{K\in\mathscr{T}_{\boldsymbol{a}}} \frac{h_K^2}{(p+1)^2}\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K^2\right\}^{\frac{1}{2}} \leqslant C\left\{\sum_{K\in\mathscr{T}_{\boldsymbol{a}}}[e_{K,p}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}}.
\end{aligned}
$$

For $p = 0$, we instead apply the Cauchy–Schwarz inequality, the stability of the L^2-projection, the Poincaré inequality on $H^1_*(\omega_{\boldsymbol{a}})$ and $\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} = 1$ to get

$$
\begin{aligned}
|(\Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla_{\mathscr{T}}\cdot(\boldsymbol{v}-\boldsymbol{\tau}_{\mathscr{T}})), \varphi)_{\omega_{\boldsymbol{a}}}| &\leqslant C\|\nabla_{\mathscr{T}}\cdot(\boldsymbol{v}-\boldsymbol{\tau}_{\mathscr{T}})\|_{\omega_{\boldsymbol{a}}} h_{\omega_{\boldsymbol{a}}}\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} \\
&\leqslant C\left\{\sum_{K\in\mathscr{T}_{\boldsymbol{a}}}[e_{K,p}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}},
\end{aligned}
$$

where C depends only on d and $\kappa_{\mathscr{T}}$.

To bound the second term on the right-hand side of (4.8), we recall the trace inequality

$$
\|\varphi\|_F^2 \leqslant C\left(\|\nabla\varphi\|_K\|\varphi\|_K + h_K^{-1}\|\varphi\|_K^2\right),
$$

for any $\varphi \in H^1(K)$ and $F \in \mathscr{F}_K$, where C depends only on d and $\kappa_{\mathscr{T}}$. Combined with the Poincaré inequality on $H^1_*(\omega_{\boldsymbol{a}})$ and $\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} = 1$, this gives

$$
\sum_{F\in\mathscr{F}^{\mathrm{in}}_{\boldsymbol{a}}} |\left(\Pi^p_F(\psi_{\boldsymbol{a}}[\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F), \varphi\right)_F| \leqslant C\left\{\sum_{F\in\mathscr{F}^{\mathrm{in}}_{\boldsymbol{a}}} h_F\|[\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F\|_F^2\right\}^{\frac{1}{2}},
$$

with C depending only on d and $\kappa_{\mathscr{T}}$. Finally, we invoke Lemma 4.5, yielding, for each $F \in \mathscr{F}_{\boldsymbol{a}}^{\text{in}}$,

$$h_F^{1/2}\|[\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F\|_F \leqslant C \sup_{\substack{w\in H^1(\omega_F)\\ w=0 \text{ on } \partial\omega_F\setminus F\\ \|\nabla w\|_{\omega_F}=1}} ([\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F, w)_F, \tag{4.9}$$

where now the constant C depends on the polynomial degree p in addition to d and $\kappa_{\mathscr{T}}$. Fix $w \in H^1(\omega_F)$ such that $w = 0$ on $\partial\omega_F \setminus F$ and $\|\nabla w\|_{\omega_F} = 1$. By definition, $F \in \mathscr{F}_{\boldsymbol{a}}^{\text{in}}$ means that F is either an internal face shared by two simplices or a Neumann boundary face. Then the zero extension of w to Ω belongs to $H^1_{\Gamma_{\mathrm{D}}}(\Omega)$. Since $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$, we infer from the definition of the weak divergence that

$$(\nabla\cdot\boldsymbol{v}, w)_{\omega_F} + (\boldsymbol{v}, \nabla w)_{\omega_F} = 0.$$

Consequently, developing $([\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F, w)_F$ shows that

$$\begin{aligned}
|([\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F, w)_F| &= |(\nabla_{\mathscr{T}}\cdot\boldsymbol{\tau}_{\mathscr{T}}, w)_{\omega_F} + (\boldsymbol{\tau}_{\mathscr{T}}, \nabla w)_{\omega_F}| \\
&\leqslant |(\nabla_{\mathscr{T}}\cdot(\boldsymbol{\tau}_{\mathscr{T}} - \boldsymbol{v}), w - \Pi^p_{\mathscr{T}} w)_{\omega_F}| + |(\boldsymbol{\tau}_{\mathscr{T}} - \boldsymbol{v}, \nabla w)_{\omega_F}| \\
&\leqslant \|\nabla_{\mathscr{T}}\cdot(\boldsymbol{\tau}_{\mathscr{T}} - \boldsymbol{v})\|_{\omega_F}\|w - \Pi^p_{\mathscr{T}} w\|_{\omega_F} + \|\boldsymbol{\tau}_{\mathscr{T}} - \boldsymbol{v}\|_{\omega_F}\|\nabla w\|_{\omega_F} \\
&\leqslant C \sum_{K\in\mathscr{T}_F}\left\{\|\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}}\|_K^2 + \frac{h_K^2}{(p+1)^2}\|\nabla\cdot(\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}})\|_K^2\right\}^{1/2},
\end{aligned}$$

owing to the Cauchy–Schwarz inequality, the orthogonality of the L^2-projection, and the approximation bound $\|w - \Pi^p_{\mathscr{T}} w\|_K \leqslant C\frac{h_K}{p+1}\|\nabla w\|_K$. Hence, Lemma A1 below implies that

$$\sum_{F\in\mathscr{F}_{\boldsymbol{a}}^{\text{in}}} \left(\Pi_F^p(\psi_{\boldsymbol{a}}[\![\boldsymbol{\tau}_{\mathscr{T}}]\!]\cdot\boldsymbol{n}_F), \varphi\right)_F \leqslant C\left\{\sum_{K\in\mathscr{T}_{\boldsymbol{a}}}[e_{K,p}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}},$$

where the constant C depends only on d, $\kappa_{\mathscr{T}}$, and the polynomial degree p via (4.9). Combining these bounds implies (4.6). □

Step 2: Bound on $\|\boldsymbol{v} - P^p_{\mathscr{T}}(\boldsymbol{v})\|_K$.
Let $K \in \mathscr{T}$. In this second and last step, we first show that

$$\|P^p_{\mathscr{T}}(\boldsymbol{v}) - \boldsymbol{\tau}_{\mathscr{T}}\|_K \leqslant C\left\{\sum_{K'\in\mathscr{T}_K}[e_{K',p}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}}. \tag{4.10}$$

Recalling that $\mathscr{V}_K$ denotes the set of vertices of the element K, using the partition of unity $\sum_{\boldsymbol{a}\in\mathscr{V}_K}\psi_{\boldsymbol{a}}|_K = 1$ and the linearity of the elementwise canonical interpolant $\boldsymbol{I}^p_{\mathscr{T}}$ (2.1), as well as the definition (3.3) of $P^p_{\mathscr{T}}(\boldsymbol{v})$ and the fact that $\boldsymbol{\tau}_{\mathscr{T}} = \boldsymbol{I}^p_{\mathscr{T}}(\boldsymbol{\tau}_{\mathscr{T}})$, we find that

$$\left(P^p_{\mathscr{T}}(\boldsymbol{v}) - \boldsymbol{\tau}_{\mathscr{T}}\right)|_K = \left(P^p_{\mathscr{T}}(\boldsymbol{v}) - \boldsymbol{I}^p_{\mathscr{T}}(\boldsymbol{\tau}_{\mathscr{T}})\right)|_K = \sum_{\boldsymbol{a}\in\mathscr{V}_K} \left(\boldsymbol{\sigma}_{\boldsymbol{a}} - \boldsymbol{I}^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}})\right)|_K.$$

Thus

$$\|P^p_{\mathscr{T}}(\boldsymbol{v}) - \boldsymbol{\tau}_{\mathscr{T}}\|^2_K = \left\| \sum_{\boldsymbol{a}\in\mathscr{V}_K} \left(\boldsymbol{\sigma}_{\boldsymbol{a}} - \boldsymbol{I}^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}})\right)\right\|^2_K \leqslant (d+1) \sum_{\boldsymbol{a}\in\mathscr{V}_K} \|\boldsymbol{\sigma}_{\boldsymbol{a}} - \boldsymbol{I}^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}})\|^2_{\omega_{\boldsymbol{a}}},$$

and Lemma 4.6 then yields (4.10).

Finally, having obtained (4.10), the main bound (4.5) then follows from the triangle inequality and Lemma A1, since

$$\|\boldsymbol{v} - P^p_{\mathscr{T}}(\boldsymbol{v})\|_K \leqslant \|\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}}\|_K + \|\boldsymbol{\tau}_{\mathscr{T}} - P^p_{\mathscr{T}}(\boldsymbol{v})\|_K \leqslant C \left\{ \sum_{K'\in\mathscr{T}_K} [e_{K',p}(\boldsymbol{v})]^2 \right\}^{\frac{1}{2}}.$$

This completes the proof of the approximation property (3.6) from Theorem 3.2.

4.4 *Proof of the projection property* (3.5)*,* $\boldsymbol{L}^2$ *stability* (3.7)*, and* $\boldsymbol{H}(\mathrm{div})$ *stability* (3.8)

To prove (3.5), we observe that if $\boldsymbol{v} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$, then it follows from the definition (3.11) that $e_{K,p}(\boldsymbol{v}) = 0$ for all $K \in \mathscr{T}$, and thus (3.5) follows immediately from (3.6).

To prove (3.7), we observe that, for any $K \in \mathscr{T}$, the triangle inequality yields

$$\|P^p_{\mathscr{T}}(\boldsymbol{v})\|_K \leqslant \|\boldsymbol{v}\|_K + \|\boldsymbol{v} - P^p_{\mathscr{T}}(\boldsymbol{v})\|_K.$$

The first term is trivially contained in the right-hand side of (3.7). Bounding the second one by (3.6) the definition (3.11) of $e_{K,p}(\boldsymbol{v})$ implies that

$$e_{K,p}(\boldsymbol{v}) \leqslant \|\boldsymbol{v}\|_K + \frac{h_K}{(p+1)} \|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K.$$

This shows that (3.7) holds true.

Finally, from (3.7), the bound in (3.8) follows immediately since $\frac{h_K}{p+1} \leqslant h_\Omega$ and since both terms $\|\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K$ and $\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K$ are bounded by $\|\nabla\cdot\boldsymbol{v}\|_K$.

5. Proof of Theorem 3.6 (*hp*-optimal approximation estimates under minimal $\boldsymbol{H}^s(K)$-regularity)

We present here a proof of Theorem 3.6. First, in Section 5.1 we derive an unbalanced but polynomial-degree-robust bound in Proposition 5.1. Then, in Section 5.2 we combine it with Theorem 3.3.

5.1 *Polynomial-degree-robust one-sided bound*

We present here an auxiliary result that gives a bound where the global-best approximation error (3.10) is bounded in terms of the sums of local-best approximation errors (3.11) with a constant that is *robust with respect to the polynomial degree*, but where the polynomial degree in the local approximation errors is $(p-1)$ instead of p. As a result, in contrast to Theorem 3.3, this is a one-sided inequality and not an equivalence, and it is valid only for $p \geqslant 1$ and not for $p \geqslant 0$.

PROPOSITION 5.1 (Polynomial-degree-robust bound). There exists a constant C, depending only on the space dimension d and the shape-regularity parameter $\kappa_{\mathscr{T}}$ of $\mathscr{T}$, such that, for any $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ and any $p \geqslant 1$,

$$[E_{\mathscr{T},p}(\boldsymbol{v})]^2 \leqslant C \sum_{K\in\mathscr{T}} [e_{K,p-1}(\boldsymbol{v})]^2. \tag{5.1}$$

Let us stress that (5.1) is similar to the first inequality in (3.12), but the constant C above is independent of the polynomial degree p.

The rest of this section is devoted to the proof of Proposition 5.1, performed in the spirit of that of Theorem 3.3. We start by adapting Definition 3.1 as follows.

DEFINITION 5.2 (Alternative locally defined mapping from $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ to $\boldsymbol{RTN}_p(\mathscr{T})\cap\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$). Let $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ be arbitrary. Let $\boldsymbol{\tau}_{\mathscr{T}}$ be defined elementwise by

$$\boldsymbol{\tau}_{\mathscr{T}}|_K := \mathop{\arg\min}_{\substack{\boldsymbol{\varsigma}_K\in\boldsymbol{RTN}_{p-1}(K)\\ \nabla\cdot\boldsymbol{\varsigma}_K=\Pi^{p-1}_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})}} \|\boldsymbol{v}-\boldsymbol{\varsigma}_K\|_K \qquad \forall\, K\in\mathscr{T}. \tag{5.2}$$

For each mesh vertex $\boldsymbol{a} \in \mathscr{V}$, the patchwise contributions $\boldsymbol{\sigma}_{\boldsymbol{a}}$ are now defined as

$$\boldsymbol{\sigma}_{\boldsymbol{a}} := \mathop{\arg\min}_{\substack{\boldsymbol{\varsigma}_{\boldsymbol{a}}\in\boldsymbol{V}_p(\omega_{\boldsymbol{a}})\\ \nabla\cdot\boldsymbol{\varsigma}_{\boldsymbol{a}}=\Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla\cdot\boldsymbol{v})+\nabla\psi_{\boldsymbol{a}}\cdot\boldsymbol{\tau}_{\mathscr{T}}}} \|\boldsymbol{\varsigma}_{\boldsymbol{a}}-\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}\|_{\omega_{\boldsymbol{a}}}, \tag{5.3}$$

with the spaces $\boldsymbol{V}_p(\omega_{\boldsymbol{a}})$ still defined in (2.3). Finally, after extending each $\boldsymbol{\sigma}_{\boldsymbol{a}}$ from $\omega_{\boldsymbol{a}}$ to the rest of Ω by zero, the equilibrated flux reconstruction $\boldsymbol{\sigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ is defined as

$$\boldsymbol{\sigma}_{\mathscr{T}} := \sum_{\boldsymbol{a}\in\mathscr{V}} \boldsymbol{\sigma}_{\boldsymbol{a}}. \tag{5.4}$$

The prescription (5.2) is similar to (3.1) up to the employment of the polynomial degree $p-1$ in place of p. Likewise, (5.3) is similar to (3.2), up to two subtle differences: the canonical interpolant $\boldsymbol{I}^p_{\mathscr{T}}$ is no longer used (since from (5.2), $\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}$ already lies in $\boldsymbol{RTN}_p(\mathscr{T})$), and the projection $\Pi^p_{\mathscr{T}}$ in the divergence constraint only concerns the term $\psi_{\boldsymbol{a}}\nabla\cdot\boldsymbol{v}$. Definition 5.2 is also well posed. In particular, existence and uniqueness for the local minimization problems (5.3) follow as in Lemma 4.2, since the orthogonality property (4.2) also holds here, implying that (4.1) is satisfied with the above definitions.

Also, just as in (4.3), we deduce that

$$\nabla\cdot\boldsymbol{\sigma}_{\mathscr{T}} = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v}).$$

We continue with the following lemma.

LEMMA 5.3 (Bound on $\boldsymbol{\sigma}_{\boldsymbol{a}}$). Let $\boldsymbol{\tau}_{\mathscr{T}}$ be given by (5.2) and let $\boldsymbol{\sigma}_{\boldsymbol{a}}$ be given by (5.3). Then there exists a constant C, depending only on d and $\kappa_{\mathscr{T}}$, such that

$$\|\boldsymbol{\sigma}_{\boldsymbol{a}} - \psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}\|_{\omega_{\boldsymbol{a}}} \leqslant C\left\{\sum_{K\in\mathscr{T}_{\boldsymbol{a}}}[e_{K,p-1}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}} \qquad \forall\boldsymbol{a}\in\mathscr{V}. \tag{5.5}$$

Proof. Fix a vertex $\boldsymbol{a}\in\mathscr{V}$. We rely on Lemma 4.4, where we take $\boldsymbol{\tau}_{\boldsymbol{a}} := \psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}$ and $g_{\boldsymbol{a}} := \Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla\cdot\boldsymbol{v}) + \nabla\psi_{\boldsymbol{a}}\cdot\boldsymbol{\tau}_{\mathscr{T}}$ in order to apply it to our construction (5.3) from Definition 5.2. This yields

$$\|\boldsymbol{\sigma}_{\boldsymbol{a}} - \psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}\|_{\omega_{\boldsymbol{a}}} \leqslant C \sup_{\substack{v\in H^1_*(\omega_{\boldsymbol{a}})\\ \|\nabla\varphi\|_{\omega_{\boldsymbol{a}}}=1}} \left\{(g_{\boldsymbol{a}},\varphi)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\tau}_{\boldsymbol{a}},\nabla\varphi)_{\omega_{\boldsymbol{a}}}\right\},$$

where the involved constant C is crucially independent of the polynomial degree p. Let $\varphi\in H^1_*(\omega_{\boldsymbol{a}})$ with $\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} = 1$ be fixed, where we recall that the space $H^1_*(\omega_{\boldsymbol{a}})$ is defined in (4.4). Then, the product $\psi_{\boldsymbol{a}}\varphi\in H^1_{\Gamma_{\rm D}}(\Omega)$ for any $\boldsymbol{a}\in\mathscr{V}$, and thus the definition of the weak divergence, implies that

$$\big(\boldsymbol{v},\nabla(\psi_{\boldsymbol{a}}\varphi)\big)_{\omega_{\boldsymbol{a}}} + \big(\nabla\cdot\boldsymbol{v},\psi_{\boldsymbol{a}}\varphi\big)_{\omega_{\boldsymbol{a}}} = 0.$$

The product rule and the orthogonality of the L^2-projection give

$$\begin{aligned}(g_{\boldsymbol{a}},\varphi)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\tau}_{\boldsymbol{a}},\nabla\varphi)_{\omega_{\boldsymbol{a}}} &= \big(\Pi^p_{\mathscr{T}}(\psi_{\boldsymbol{a}}\nabla\cdot\boldsymbol{v}),\varphi\big)_{\omega_{\boldsymbol{a}}} + (\nabla\psi_{\boldsymbol{a}}\cdot\boldsymbol{\tau}_{\mathscr{T}},\varphi)_{\omega_{\boldsymbol{a}}} + \big(\psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}},\nabla\varphi\big)_{\omega_{\boldsymbol{a}}}\\ &= \big(\nabla\cdot\boldsymbol{v},\psi_{\boldsymbol{a}}\Pi^p_{\mathscr{T}}(\varphi)\big)_{\omega_{\boldsymbol{a}}} + \big(\boldsymbol{\tau}_{\mathscr{T}},\nabla(\psi_{\boldsymbol{a}}\varphi)\big)_{\omega_{\boldsymbol{a}}}\\ &= \big(\nabla\cdot\boldsymbol{v},\psi_{\boldsymbol{a}}(\Pi^p_{\mathscr{T}}(\varphi)-\varphi)\big)_{\omega_{\boldsymbol{a}}} + \big(\boldsymbol{\tau}_{\mathscr{T}}-\boldsymbol{v},\nabla(\psi_{\boldsymbol{a}}\varphi)\big)_{\omega_{\boldsymbol{a}}}\\ &= \big(\psi_{\boldsymbol{a}}(\nabla\cdot\boldsymbol{v}-\Pi^{p-1}_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})),\Pi^p_{\mathscr{T}}(\varphi)-\varphi\big)_{\omega_{\boldsymbol{a}}} + \big(\boldsymbol{\tau}_{\mathscr{T}}-\boldsymbol{v},\nabla(\psi_{\boldsymbol{a}}\varphi)\big)_{\omega_{\boldsymbol{a}}},\end{aligned}$$

since $\psi_{\boldsymbol{a}}\Pi^{p-1}_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})$ is a piecewise polynomial of degree at most p. Therefore, we have

$$\begin{aligned}|(g_{\boldsymbol{a}},\varphi)_{\omega_{\boldsymbol{a}}} + (\boldsymbol{\tau}_{\boldsymbol{a}},\nabla\varphi)_{\omega_{\boldsymbol{a}}}| &\leqslant C\sum_{K\in\mathscr{T}_{\boldsymbol{a}}}\left[\frac{h_K}{p}\|\nabla\cdot\boldsymbol{v}-\Pi^{p-1}_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K\right]\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}}\\ &\quad + \|\boldsymbol{v}-\boldsymbol{\tau}_{\mathscr{T}}\|_{\omega_{\boldsymbol{a}}}\|\nabla(\psi_{\boldsymbol{a}}\varphi)\|_{\omega_{\boldsymbol{a}}}\\ &\leqslant C\left(1+\|\nabla(\psi_{\boldsymbol{a}}\varphi)\|_{\omega_{\boldsymbol{a}}}\right)\left\{\sum_{K\in\mathscr{T}_{\boldsymbol{a}}}[e_{K,p-1}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}},\end{aligned}$$

where we have used $\|\psi_{\boldsymbol{a}}\|_{\infty,\omega_{\boldsymbol{a}}} = 1$, the *hp* approximation bound $\|\varphi - \Pi^p_{\mathscr{T}}(\varphi)\|_K \leqslant C\frac{h_K}{p+1}\|\nabla\varphi\|_K \leqslant C\frac{h_K}{p}\|\nabla\varphi\|_K$, the Cauchy–Schwarz inequality, the scaling $\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} = 1$, and Lemma A1 below applied with $(p-1)$ in place of p. Finally, the bound (5.5) follows from the inequality $\|\nabla(\psi_{\boldsymbol{a}}\varphi)\|_{\omega_{\boldsymbol{a}}} \leqslant C\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} \leqslant C$ for all $\varphi \in H^1_*(\omega_{\boldsymbol{a}})$, owing to the Poincaré inequality on $H^1_*(\omega_{\boldsymbol{a}})$ and $\|\nabla\varphi\|_{\omega_{\boldsymbol{a}}} = 1$. □

We are now ready to complete the proof of Proposition 5.1. Let $\boldsymbol{v} \in \boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div},\Omega)$ and let $\boldsymbol{\sigma}_{\mathscr{T}}$ be given by Definition 5.2. The triangle inequality gives

$$\|\boldsymbol{v} - \boldsymbol{\sigma}_{\mathscr{T}}\| \leqslant \|\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}}\| + \|\boldsymbol{\tau}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathscr{T}}\|,$$

and, Lemma A1 below applied with $(p-1)$ in place of p allows us to bound the divergence-constrained minimization in (5.2) as

$$\|\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}}\| \leqslant C\left\{\sum_{K\in\mathscr{T}}[e_{K,p-1}(\boldsymbol{v})]^2\right\}^{\frac{1}{2}},$$

where C only depends on d and $\kappa_{\mathscr{T}}$. Finally, we obtain (5.1) from the estimate

$$\|\boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\tau}_{\mathscr{T}}\|^2 = \sum_{K\in\mathscr{T}}\left\|\sum_{\boldsymbol{a}\in\mathscr{V}_K}\left(\boldsymbol{\sigma}_{\boldsymbol{a}} - \psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}\right)\right\|_K^2 \leqslant (d+1)\sum_{\boldsymbol{a}\in\mathscr{V}}\|\boldsymbol{\sigma}_{\boldsymbol{a}} - \psi_{\boldsymbol{a}}\boldsymbol{\tau}_{\mathscr{T}}\|^2_{\omega_{\boldsymbol{a}}},$$

Lemma 5.3, and picking $\boldsymbol{\sigma}_{\mathscr{T}}$ in (3.10).

5.2 *Proof of Theorem 3.6*

The proof of Theorem 3.6 hinges on the bounds from Theorem 3.3 and Proposition 5.1. Recall the definitions (3.10) of $E_{\mathscr{T},p}(\boldsymbol{v})$ and (3.11) of $e_{K,p}(\boldsymbol{v})$. Recall also the notation $\delta_{s<1} := 1$ if $s < 1$ and $\delta_{s<1} := 0$ if $s \geqslant 1$. We proceed in two steps.

Step 1: Case $p \leqslant s$. We first suppose that $p \leqslant s$ and let $t := \min(s, p+1)$. Here we will employ Theorem 3.3. Since $\mathscr{P}_p(K;\mathbb{R}^d) \subset \boldsymbol{RTN}_p(K)$, well-known *hp*-approximation bounds (see e.g., Babuška & Suri, 1987, Lemma 4.1) imply that

$$[e_{K,p}(\boldsymbol{v})]^2 \leqslant C\left\{\left[\frac{h_K^t}{(p+1)^s}\|\boldsymbol{v}\|_{\boldsymbol{H}^s(K)}\right]^2 + \delta_{s<1}\left[\frac{h_K}{p+1}\|\nabla{\cdot}\boldsymbol{v}\|_K\right]^2\right\}, \tag{5.6}$$

for each $K \in \mathscr{T}$, with C depending only on s, d, $\kappa_{\mathscr{T}}$. Note that for $s < 1$ we applied here the trivial bound $\|\nabla{\cdot}\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla{\cdot}\boldsymbol{v})\|_K \leqslant \|\nabla{\cdot}\boldsymbol{v}\|_K$ as $\boldsymbol{v}|_K \in \boldsymbol{H}^s(K)$ is insufficient to improve the bound on the error of the divergence. Combining (5.6) with the first bound in (3.12) of Theorem 3.3 then implies that there exists a constant $C_{s,d,\kappa_{\mathscr{T}},p}$ depending only on s, d, $\kappa_{\mathscr{T}}$, and p, such that

$$[E_{\mathscr{T},p}(\boldsymbol{v})]^2 \leqslant C_{s,d,\kappa_{\mathscr{T}},p}\sum_{K\in\mathscr{T}}\left\{\left[\frac{h_K^t}{(p+1)^s}\|\boldsymbol{v}\|_{\boldsymbol{H}^s(K)}\right]^2 + \delta_{s<1}\left[\frac{h_K}{p+1}\|\nabla{\cdot}\boldsymbol{v}\|_K\right]^2\right\}.$$

Define then the constant $C^{\star}_{s,d,\kappa_{\mathcal{T}}} := \max_{0\leqslant p\leqslant s} C_{s,d,\kappa_{\mathcal{T}},p}$, so that, for all $p \leqslant s$,

$$[E_{\mathcal{T},p}(\boldsymbol{v})]^2 \leqslant C^{\star}_{s,d,\kappa_{\mathcal{T}}} \sum_{K\in\mathcal{T}} \left\{ \left[\frac{h_K^t}{(p+1)^s} \|\boldsymbol{v}\|_{\boldsymbol{H}^s(K)} \right]^2 + \delta_{s<1} \left[\frac{h_K}{p+1} \|\nabla\cdot\boldsymbol{v}\|_K \right]^2 \right\}.$$

This implies (3.13) for any $p \leqslant s$ with constant $C = C^{\star}_{s,d,\kappa_{\mathcal{T}}}$.

Step 2: Case $p > s$. Now consider the case $p > s$; since p is an integer, this implies that $p \geqslant 1$. Here we rely on Proposition 5.1. The approximation bounds, similarly to (5.6), imply that there exists a constant C, depending only on s, d, and $\kappa_{\mathcal{T}}$, such that

$$[e_{K,p-1}(\boldsymbol{v})]^2 \leqslant C \left\{ \left[\frac{h_K^s}{p^s} \|\boldsymbol{v}\|_{\boldsymbol{H}^s(K)} \right]^2 + \delta_{s<1} \left[\frac{h_K}{p} \|\nabla\cdot\boldsymbol{v}\|_K \right]^2 \right\},$$

for all $K \in \mathcal{T}$. Note that $p+1 \leqslant 2p$ for all $p \geqslant 1$, so that the terms p^s in the denominators above can be replaced by $(p+1)^s$ at the cost of an extra s-dependent constant, and similarly for $\frac{1}{p} \leq \frac{2}{(p+1)}$. Hence, the inequality (5.1) of Proposition 5.1 and summation over the elements of $\mathcal{T}$ show that there exists a constant $C^{\sharp}_{s,d,\kappa_{\mathcal{T}}}$, depending only on s, d, and $\kappa_{\mathcal{T}}$, such that (3.13) holds with constant $C = C^{\sharp}_{s,d,\kappa_{\mathcal{T}}}$ for all $p > s$.

Conclusion. Combining Steps 1 and 2 shows that (3.13) holds for general $s \geqslant 0$ and $p \geqslant 0$ with a constant C that can be taken as $\max\{C^{\star}_{s,d,\kappa_{\mathcal{T}}}, C^{\sharp}_{s,d,\kappa_{\mathcal{T}}}\}$, which then depends only on s, d, and $\kappa_{\mathcal{T}}$.

REMARK 5.4 (Full *hp*-optimality). Theorem 3.6 shows that optimal-order convergence rates with respect to both the mesh sizes h_K and the polynomial degree p can be obtained despite the unfavorable dependence of the constant C on the polynomial degree p in Theorem 3.3 and unbalanced polynomial degrees in Proposition 5.1.

6. Application to *a priori* error estimates

In this section we show how to apply the results of Section 3 to the *a priori* error analysis of mixed finite element methods and least-squares mixed finite element methods for a model diffusion problem.

6.1 *Mixed finite element methods*

Let us consider the dual mixed finite element method for the Poisson model problem, following Raviart & Thomas (1977), Nédélec (1980), Roberts & Thomas (1991), or Boffi *et al.* (2013). Let $f \in L^2(\Omega)$ and $\Gamma_{\mathrm{N}} = \emptyset$ for simplicity, so that $\boldsymbol{H}_{0,\Gamma_{\mathrm{N}}}(\mathrm{div}, \Omega)$ becomes $\boldsymbol{H}(\mathrm{div}, \Omega)$. Consider the Laplace problem of finding $u : \Omega \to \mathbb{R}$ such that

$$-\Delta u = f \qquad \text{in } \Omega, \tag{6.1a}$$

$$u = 0 \qquad \text{on } \partial\Omega. \tag{6.1b}$$

The primal weak formulation of (6.1) reads as follows: find $u \in H_0^1(\Omega)$ such that

$$(\nabla u, \nabla v) = (f, v) \qquad \forall\, v \in H_0^1(\Omega). \tag{6.2}$$

The dual weak formulation of (6.1) then reads as follows: find $\boldsymbol{\sigma} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ and $u \in L^2(\Omega)$ such that

$$(\boldsymbol{\sigma}, \boldsymbol{v}) - (u, \nabla\cdot\boldsymbol{v}) = 0 \qquad \forall\, \boldsymbol{v} \in \boldsymbol{H}(\mathrm{div}, \Omega), \tag{6.3a}$$

$$(\nabla\cdot\boldsymbol{\sigma}, q) = (f, q) \qquad \forall\, q \in L^2(\Omega). \tag{6.3b}$$

Classically, the solutions u from (6.2) and (6.3) coincide and $\boldsymbol{\sigma} = -\nabla u$. The dual mixed finite element method of order $p \geqslant 0$ for problem (6.3) then looks for the pair $\boldsymbol{\sigma}_{\mathrm{M}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ and $u_{\mathrm{M}} \in \mathscr{P}_p(\mathscr{T})$ such that

$$(\boldsymbol{\sigma}_{\mathrm{M}}, \boldsymbol{\varsigma}_{\mathscr{T}}) - (u_{\mathrm{M}}, \nabla\cdot\boldsymbol{\varsigma}_{\mathscr{T}}) = 0 \qquad \forall\, \boldsymbol{\varsigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega), \tag{6.4a}$$

$$(\nabla\cdot\boldsymbol{\sigma}_{\mathrm{M}}, q_{\mathscr{T}}) = (f, q_{\mathscr{T}}) \qquad \forall\, q_{\mathscr{T}} \in \mathscr{P}_p(\mathscr{T}). \tag{6.4b}$$

It is immediate to check from (6.3b) and (6.4b) that $\nabla\cdot\boldsymbol{\sigma}_{\mathrm{M}} = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{\sigma})$. Furthermore, the following *a priori* error characterization is classical; cf. Boffi *et al.* (2013). We include its proof to highlight the precise arguments.

LEMMA 6.1 (*A priori* bound for mixed finite element methods). Let $\boldsymbol{\sigma}_{\mathrm{M}}$ be the first component of the dual mixed finite element solution solving (6.4), approximating $\boldsymbol{\sigma}$ from (6.3). Then,

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}\| = \min_{\substack{\boldsymbol{\varsigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega) \\ \nabla\cdot\boldsymbol{\varsigma}_{\mathscr{T}} = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{\sigma})}} \|\boldsymbol{\sigma} - \boldsymbol{\varsigma}_{\mathscr{T}}\|.$$

Proof. Subtracting (6.4a) from (6.3a), we have

$$(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}, \boldsymbol{\varsigma}_{\mathscr{T}}) - (u - u_{\mathrm{M}}, \nabla\cdot\boldsymbol{\varsigma}_{\mathscr{T}}) = 0 \quad \forall\, \boldsymbol{\varsigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega). \tag{6.5}$$

Let $\boldsymbol{\sigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ be such that $\nabla\cdot\boldsymbol{\sigma}_{\mathscr{T}} = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{\sigma})$. Taking $\boldsymbol{\varsigma}_{\mathscr{T}} = \boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{M}}$ in (6.5), we obtain, since $\nabla\cdot\boldsymbol{\varsigma}_{\mathscr{T}} = 0$,

$$(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}, \boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{M}}) = 0.$$

Now, clearly

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}\|^2 = (\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}, \boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}) = (\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}, \boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}}) \leqslant \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}\|\, \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}}\|,$$

and hence $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}\| \leqslant \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}}\|$. Since $\boldsymbol{\sigma}_{\mathscr{T}}$ is arbitrary subject to the divergence constraint and can be taken as $\boldsymbol{\sigma}_{\mathrm{M}}$, we obtain the assertion. □

Thus, $\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{M}}\|$ can be readily estimated by using Theorems 3.3 and 3.6.

6.2 *Least-squares mixed finite element methods*

In this subsection we showcase the application of our results to the least-squares mixed finite element method discussed in Pehlivanov *et al.* (1994), Cai & Ku (2010), and Ku (2013); see also the references therein.

Again let $\Gamma_{\mathrm{N}} = \emptyset$ for simplicity and $f \in L^2(\Omega)$. Let $\boldsymbol{\sigma} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ and $u \in H_0^1(\Omega)$ be such that

$$(\boldsymbol{\sigma}, u) := \arg \min_{(\boldsymbol{p},v)\in \boldsymbol{H}(\mathrm{div},\Omega)\times H_0^1(\Omega)} \left\{ h_\Omega^2 \|\nabla{\cdot}\boldsymbol{p} - f\|^2 + \|\boldsymbol{p} + \nabla v\|^2 \right\},$$

where we recall that h_Ω is a length scale equal to the diameter of Ω. Then $\boldsymbol{\sigma} \in \boldsymbol{H}(\mathrm{div}, \Omega)$ and $u \in H_0^1(\Omega)$ solve the following system of equations:

$$(\boldsymbol{\sigma} + \nabla u, \nabla v) = 0 \qquad \forall\, v \in H_0^1(\Omega), \tag{6.6a}$$

$$h_\Omega^2(\nabla{\cdot}\boldsymbol{\sigma}, \nabla{\cdot}\boldsymbol{p}) + (\boldsymbol{\sigma} + \nabla u, \boldsymbol{p}) = h_\Omega^2(f, \nabla{\cdot}\boldsymbol{p}) \qquad \forall \boldsymbol{p} \in \boldsymbol{H}(\mathrm{div}, \Omega). \tag{6.6b}$$

Again, $\boldsymbol{\sigma}$ and u coincide with the solutions of (6.2) and (6.3). Let $p \geqslant 0$ and $q \geqslant 1$ denote two fixed polynomial degrees. The least-squares mixed finite element method for problem (6.6) consists of finding $\boldsymbol{\sigma}_{\mathrm{LS}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ and $u_{\mathrm{LS}} \in \mathscr{P}^q(\mathscr{T}) \cap H_0^1(\Omega)$ such that

$$(\boldsymbol{\sigma}_{\mathrm{LS}} + \nabla u_{\mathrm{LS}}, \nabla v_{\mathscr{T}}) = 0 \qquad \forall\, v_{\mathscr{T}} \in \mathscr{P}^q(\mathscr{T}) \cap H_0^1(\Omega), \tag{6.7a}$$

$$h_\Omega^2(\nabla{\cdot}\boldsymbol{\sigma}_{\mathrm{LS}}, \nabla{\cdot}\boldsymbol{p}_{\mathscr{T}}) + (\boldsymbol{\sigma}_{\mathrm{LS}} + \nabla u_{\mathrm{LS}}, \boldsymbol{p}_{\mathscr{T}}) = h_\Omega^2(f, \nabla{\cdot}\boldsymbol{p}_{\mathscr{T}}) \qquad \forall \boldsymbol{p}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega). \tag{6.7b}$$

Similarly to Lemma 6.1, we can obtain the following *a priori* error characterization.

LEMMA 6.2 (*A priori* bound for least-squares mixed finite element methods). Let $(\boldsymbol{\sigma}_{\mathrm{LS}}, u_{\mathrm{LS}})$ be the least-squares mixed finite element solution pair solving (6.7), approximating $(\boldsymbol{\sigma}, u)$ from (6.6). Then there exists a generic constant C, at most equal to 17, such that

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}\| + \|\nabla(u - u_{\mathrm{LS}})\| \leqslant C \left(\min_{\substack{\boldsymbol{\varsigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div},\Omega) \\ \nabla{\cdot}\boldsymbol{\varsigma}_{\mathscr{T}} = \Pi_{\mathscr{T}}^p(\nabla{\cdot}\boldsymbol{\sigma})}} \|\boldsymbol{\sigma} - \boldsymbol{\varsigma}_{\mathscr{T}}\| + \min_{v_{\mathscr{T}} \in \mathscr{P}^q(\mathscr{T}) \cap H_0^1(\Omega)} \|\nabla(u - v_{\mathscr{T}})\| \right).$$

Proof. Define the bilinear form $\mathscr{A}$ on $(\boldsymbol{H}(\mathrm{div}, \Omega) \times H_0^1(\Omega)) \times (\boldsymbol{H}(\mathrm{div}, \Omega) \times H_0^1(\Omega))$ by

$$\mathscr{A}(\boldsymbol{\sigma}, u; \boldsymbol{p}, v) := (\boldsymbol{\sigma} + \nabla u, \nabla v) + h_\Omega^2(\nabla{\cdot}\boldsymbol{\sigma}, \nabla{\cdot}\boldsymbol{p}) + (\boldsymbol{\sigma} + \nabla u, \boldsymbol{p}).$$

We have the following orthogonality from (6.6) and (6.7):

$$\mathscr{A}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}, u - u_{\mathrm{LS}}; \boldsymbol{p}_{\mathscr{T}}, v_{\mathscr{T}}) = 0, \tag{6.8}$$

for all $\boldsymbol{p}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div},\Omega)$ and for all $v_{\mathscr{T}} \in \mathscr{P}^q(\mathscr{T}) \cap H^1_0(\Omega)$. Moreover, the following coercivity is known from Pehlivanov *et al.* (1994): there exists a constant C such that

$$\mathscr{A}(\boldsymbol{p},v;\boldsymbol{p},v) \geqslant \frac{1}{C}\left(\|\boldsymbol{p}\|^2 + h_\Omega^2\|\nabla{\cdot}\boldsymbol{p}\|^2 + \|\nabla v\|^2\right) \qquad \forall\,(\boldsymbol{p},v) \in \boldsymbol{H}(\mathrm{div},\Omega)\times H^1_0(\Omega)). \tag{6.9}$$

Indeed, owing to the Cauchy–Schwarz and Young inequalities, we have, for any $0<\varepsilon<2$,

$$\begin{aligned}
\mathscr{A}(\boldsymbol{p},v;\boldsymbol{p},v) &= \|\nabla v\|^2 + (2-\varepsilon)(\boldsymbol{p},\nabla v) + h_\Omega^2\|\nabla{\cdot}\boldsymbol{p}\|^2 + \|\boldsymbol{p}\|^2 + \varepsilon(\nabla v,\boldsymbol{p})\\
&\geqslant \|\boldsymbol{p}\|^2 + \|\nabla v\|^2 - \frac{2-\varepsilon}{2}(\|\boldsymbol{p}\|^2 + \|\nabla v\|^2) + h_\Omega^2\|\nabla{\cdot}\boldsymbol{p}\|^2 - \varepsilon\|\nabla{\cdot}\boldsymbol{p}\|\|v\|\\
&\geqslant \frac{\varepsilon}{2}(\|\boldsymbol{p}\|^2 + \|\nabla v\|^2) + h_\Omega^2\|\nabla{\cdot}\boldsymbol{p}\|^2 - \varepsilon C_{\mathrm{PF}}h_\Omega\left(C_{\mathrm{PF}}h_\Omega\|\nabla{\cdot}\boldsymbol{p}\|^2 + \frac{1}{4C_{\mathrm{PF}}h_\Omega}\|\nabla v\|^2\right)\\
&= \frac{\varepsilon}{2}\|\boldsymbol{p}\|^2 + \frac{\varepsilon}{4}\|\nabla v\|^2 + \|\nabla{\cdot}\boldsymbol{p}\|^2\left(h_\Omega^2 - \varepsilon C_{\mathrm{PF}}^2h_\Omega^2\right),
\end{aligned}$$

where we have also employed the Green theorem $(\nabla v,\boldsymbol{p}) = -(\nabla{\cdot}\boldsymbol{p},v)$ and the Poincaré–Friedrichs inequality $\|v\| \leqslant C_{\mathrm{PF}}h_\Omega\|\nabla v\|$ (here h_Ω is the diameter of Ω and $C_{\mathrm{PF}} \leqslant 1$ a generic constant). The assertion (6.9) follows by choosing, e.g., $\varepsilon = h_\Omega^2/(2C_{\mathrm{PF}}^2h_\Omega^2)$. Note that, employing $C_{\mathrm{PF}} = 1$, the constant C in (6.9) can be taken as 8.

Now let $\boldsymbol{\varsigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T})\cap\boldsymbol{H}(\mathrm{div},\Omega)$ be such that $\nabla{\cdot}\boldsymbol{\varsigma}_{\mathscr{T}} = \Pi^p_{\mathscr{T}}(\nabla{\cdot}\boldsymbol{\sigma})$ and $v_{\mathscr{T}} \in \mathscr{P}^q(\mathscr{T})\cap H^1_0(\Omega)$ be an arbitrary function. Set $q_{\mathscr{T}} := v_{\mathscr{T}} - u_{\mathrm{LS}}$ and $\boldsymbol{p}_{\mathscr{T}} := \boldsymbol{\varsigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{LS}}$. Then using (6.8) and (6.9), we find

$$\begin{aligned}
\frac{1}{C}\left(\|\boldsymbol{p}_{\mathscr{T}}\|^2 + \|\nabla q_{\mathscr{T}}\|^2\right) &\leqslant \mathscr{A}(\boldsymbol{\varsigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{LS}}, v_{\mathscr{T}} - u_{\mathrm{LS}};\boldsymbol{p}_{\mathscr{T}},q_{\mathscr{T}})\\
&:= \mathscr{A}(\boldsymbol{\varsigma}_{\mathscr{T}} - \boldsymbol{\sigma}, v_{\mathscr{T}} - u;\boldsymbol{p}_{\mathscr{T}},q_{\mathscr{T}})\\
&:= (\boldsymbol{\varsigma}_{\mathscr{T}} - \boldsymbol{\sigma} + \nabla(v_{\mathscr{T}} - u),\nabla q_{\mathscr{T}}) + h_\Omega^2(\nabla{\cdot}(\boldsymbol{\varsigma}_{\mathscr{T}} - \boldsymbol{\sigma}),\nabla{\cdot}\boldsymbol{p}_{\mathscr{T}})\\
&\qquad + (\boldsymbol{\varsigma}_{\mathscr{T}} - \boldsymbol{\sigma} + \nabla(v_{\mathscr{T}} - u),\boldsymbol{p}_{\mathscr{T}}).
\end{aligned}$$

Since $\nabla{\cdot}\boldsymbol{\varsigma}_{\mathscr{T}} = \Pi^p_{\mathscr{T}}(\nabla{\cdot}\boldsymbol{\sigma})$ and $\nabla{\cdot}\boldsymbol{p}_{\mathscr{T}} \in \mathscr{P}^p(\mathscr{T})$, we have $(\nabla{\cdot}(\boldsymbol{\varsigma}_{\mathscr{T}} - \boldsymbol{\sigma}),\nabla{\cdot}\boldsymbol{p}_{\mathscr{T}}) = 0$. Using the Cauchy–Schwarz and the Young inequalities, we then obtain, with the constant C from (6.9),

$$\|\boldsymbol{p}_{\mathscr{T}}\| + \|\nabla q_{\mathscr{T}}\| \leqslant 2C\left(\|\boldsymbol{\sigma} - \boldsymbol{\varsigma}_{\mathscr{T}}\| + \|\nabla(u - v_{\mathscr{T}})\|\right),$$

which proves the claim owing to the triangle inequality and since $\boldsymbol{\varsigma}_{\mathscr{T}}$ and $v_{\mathscr{T}}$ are arbitrary. □

The two terms in the error bound from Lemma 6.2 are uncoupled. For the first one, we can again straightforwardly use Theorems 3.3 and 3.6. For the second one, the result of Veeser (2016) yields

$$\min_{v_{\mathscr{T}} \in \mathscr{P}^q(\mathscr{T})\cap H^1_0(\Omega)} \|\nabla(u - v_{\mathscr{T}})\|^2 \leqslant C\sum_{K\in\mathscr{T}}\min_{q_K\in\mathscr{P}^q(K)}\|\nabla(u - q_K)\|_K^2,$$

where the constant C depends only on the space dimension d, the shape-regularity parameter $\kappa_{\mathscr{T}}$ of $\mathscr{T}$, and the polynomial degree q, which is again optimal.

Finally, a localized estimate for the error $\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}})$ follows by the combination of the above results with the following lemma.

LEMMA 6.3 (*A priori* bound on the divergence for least-squares mixed finite element methods). Let $(\boldsymbol{\sigma}_{\mathrm{LS}}, u_{\mathrm{LS}})$ be the least-squares mixed finite element solution solving (6.7), approximating $(\boldsymbol{\sigma}, u)$ from (6.6). Then

$$h_\Omega^2 \|\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}})\|^2 \leqslant h_\Omega^2 \|\nabla\cdot\boldsymbol{\sigma} - \Pi_{\mathscr{T}}^p(\nabla\cdot\boldsymbol{\sigma})\|^2 + \|\nabla(u - u_{\mathrm{LS}})\|^2 + \min_{\substack{\boldsymbol{\varsigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div},\Omega) \\ \nabla\cdot\boldsymbol{\varsigma}_{\mathscr{T}} = \Pi_{\mathscr{T}}^p(\nabla\cdot\boldsymbol{\sigma})}} \|\boldsymbol{\sigma} - \boldsymbol{\varsigma}_{\mathscr{T}}\|^2.$$

Proof. Again, let $\boldsymbol{\sigma}_{\mathscr{T}} \in \boldsymbol{RTN}_p(\mathscr{T}) \cap \boldsymbol{H}(\mathrm{div}, \Omega)$ be such that $\nabla\cdot\boldsymbol{\sigma}_{\mathscr{T}} = \Pi_{\mathscr{T}}^p(\nabla\cdot\boldsymbol{\sigma})$. Using (6.6b) and (6.7b), we have

$$\begin{aligned}
h_\Omega^2 \|\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}})\|^2 &= h_\Omega^2 (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}), \nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}})) \\
&= h_\Omega^2 (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}), \nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}})) + h_\Omega^2 (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}), \nabla\cdot(\boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{LS}})) \\
&= h_\Omega^2 (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}), \nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}})) + h_\Omega^2 (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}), \nabla\cdot(\boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{LS}})) \\
&\quad - \mathscr{A}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}, u - u_{\mathrm{LS}}; \boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{LS}}, 0) \\
&= h_\Omega^2 (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}), \nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}})) - ((\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}) + \nabla(u - u_{\mathrm{LS}}), \boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{LS}}) \\
&= h_\Omega^2 (\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}}), \nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}})) - ((\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}) + \nabla(u - u_{\mathrm{LS}}), \boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}_{\mathrm{LS}}),
\end{aligned}$$

where we used that $(\nabla\cdot\boldsymbol{\sigma}_{\mathrm{LS}}, \nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}})) = (\nabla\cdot\boldsymbol{\sigma}_{\mathscr{T}}, \nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}})) = 0$ since both $\nabla\cdot\boldsymbol{\sigma}_{\mathrm{LS}}$ and $\nabla\cdot\boldsymbol{\sigma}_{\mathscr{T}}$ belong to $\mathscr{P}_p(\mathscr{T})$ in the last equality. Adding and subtracting $\boldsymbol{\sigma}$ in the second term on the right-hand side above and applying the Cauchy–Schwarz and Young inequalities imply that

$$\begin{aligned}
&h_\Omega^2 \|\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}})\|^2 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}\|^2 \\
&\quad = h_\Omega^2 \|\nabla\cdot\boldsymbol{\sigma} - \Pi_{\mathscr{T}}^p(\nabla\cdot\boldsymbol{\sigma})\|^2 - (\nabla(u - u_{\mathrm{LS}}), \boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}) \\
&\quad\quad - (\nabla(u - u_{\mathrm{LS}}), \boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}) - (\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}, \boldsymbol{\sigma}_{\mathscr{T}} - \boldsymbol{\sigma}) \\
&\quad \leqslant h_\Omega^2 \|\nabla\cdot\boldsymbol{\sigma} - \Pi_{\mathscr{T}}^p(\nabla\cdot\boldsymbol{\sigma})\|^2 + \|\nabla(u - u_{\mathrm{LS}})\|^2 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}}\|^2 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}}\|^2.
\end{aligned}$$

We infer that

$$h_\Omega^2 \|\nabla\cdot(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{LS}})\|^2 \leqslant h_\Omega^2 \|\nabla\cdot\boldsymbol{\sigma} - \Pi_{\mathscr{T}}^p(\nabla\cdot\boldsymbol{\sigma})\|^2 + \|\nabla(u - u_{\mathrm{LS}})\|^2 + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathscr{T}}\|^2.$$

This finishes the proof since $\boldsymbol{\sigma}_{\mathscr{T}}$ is arbitrary. □

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 647134 GATIPOR).

References

Aurada, M., Feischl, M., Kemetmüller, J., Page, M. & Praetorius, D. (2013) Each $H^{1/2}$-stable projection yields convergence and quasi-optimality of adaptive FEM with inhomogeneous Dirichlet data in $\mathbb{R}^d$. *ESAIM Math. Model. Numer. Anal*, **47**, 1207–1235.

Babuška, I. & Suri, M. (1987) The *h-p* version of the finite element method with quasi-uniform meshes. *RAIRO Modél. Math. Anal. Numér.*, **21**, 199–238.

Becker, R., Capatina, D. & Luce, R. (2016) Local flux reconstructions for standard finite element methods on triangular meshes. *SIAM J. Numer. Anal.*, **54**, 2684–2706.

Bernardi, C. & Girault, V. (1998) A local regularization operator for triangular and quadrilateral finite elements. *SIAM J. Numer. Anal.*, **35**, 1893–1916.

Bernardi, C. & Hecht, F. (2007) Quelques propriétés d'approximation des éléments finis de Nédélec, application à l'analyse a posteriori. *C. R. Math. Acad. Sci. Paris*, **344**, 461–466.

Bespalov, A. & Heuer, N. (2011) A new $\mathbf{H}$(div)-conforming p-interpolation operator in two dimensions. *ESAIM Math. Model. Numer. Anal.*, **45**, 255–275.

Boffi, D., Brezzi, F. & Fortin, M. (2013) Mixed finite element methods and applications. *Springer Series in Computational Mathematics*, vol. 44. Heidelberg: Springer.

Braess, D., Pillwein, V. & Schöberl, J. (2009) Equilibrated residual error estimates are p-robust. *Comput. Methods Appl. Mech. Engrg.*, **198**, 1189–1197.

Braess, D. & Schöberl, J. (2008) Equilibrated residual error estimator for edge elements. *Math. Comp.*, **77**, 651–672.

Cai, Z. & Ku, J. (2010) Optimal error estimate for the div least-squares method with data $f \in L^2$ and application to nonlinear problems. *SIAM J. Numer. Anal*, **47**, 4098–4111.

Canuto, C., Nochetto, R. H., Stevenson, R. & Verani, M. (2017) Convergence and optimality of *hp*-AFEM. *Numer. Math.*, **135**, 1073–1119.

Carstensen, C., Peterseim, D. & Schedensack, M. (2012) Comparison results of finite element methods for the Poisson model problem. *SIAM J. Numer. Anal.*, **50**, 2803–2823.

Carstensen, C. & Schedensack, M. (2015) Medius analysis and comparison results for first-order finite element methods in linear elasticity. *IMA J. Numer. Anal.*, **35**, 1591–1621.

Christiansen, S. H. & Winther, R. (2008) Smoothed projections in finite element exterior calculus. *Math. Comp.*, **77**, 813–829.

Clément, P. (1975) Approximation by finite element functions using local regularization. *RAIRO Anal. Numer.*, **9**, 77–84.

Costabel, M. & McIntosh, A. (2010) On Bogovskiĭ and regularized Poincaré integral operators for de Rham complexes on Lipschitz domains. *Math. Z.*, **265**, 297–320.

Demkowicz, L. (2008) Polynomial exact sequences and projection-based interpolation with application to Maxwell equations. *Mixed Finite Elements, Compatibility Conditions, and Applications* (D. Boffi & L. Gastaldi eds), Lecture Notes in Mathematics, vol. 1939. Berlin: Springer, pp. 101–158.

Demkowicz, L. & Buffa, A. (2005) H^1, H(curl) and H(div)-conforming projection-based interpolation in three dimensions. Quasi-optimal p-interpolation estimates. *Comput. Methods Appl. Mech. Engrg*, **194**, 267–296.

Demkowicz, L., Gopalakrishnan, J. & Schöberl, J. (2008) Polynomial extension operators. Part I. *SIAM J. Numer. Anal.*, **46**, 3006–3031.

Demkowicz, L., Gopalakrishnan, J. & Schöberl, J. (2009) Polynomial extension operators. Part II. *SIAM J. Numer. Anal.*, **47**, 3293–3324.

Demkowicz, L., Gopalakrishnan, J. & Schöberl, J. (2012) Polynomial extension operators. Part III. *Math. Comp.*, **81**, 1289–1326.

Destuynder, P. & Métivet, B. (1999) Explicit error bounds in a conforming finite element method. *Math. Comp.*, **68**, 1379–1396.

Dolejší, V., Ern, A. & Vohralík, M. (2016) *hp*-adaptation driven by polynomial-degree-robust a posteriori error estimates for elliptic problems. *SIAM J. Sci. Comput.*, **38**, A3220–A3246.

ERN, A. & GUERMOND, J.-L. (2016) Mollification in strongly Lipschitz domains with application to continuous and discrete de Rham complexes. *Comput. Methods Appl. Math.*, **16**, 51–75.

ERN, A. & GUERMOND, J.-L. (2017) Finite element quasi-interpolation and best approximation. *ESAIM Math. Model. Numer. Anal.*, **51**, 1367–1385.

ERN, A. & GUERMOND, J.-L. (2018). Quasi-optimal nonconforming approximation of elliptic PDEs with contrasted coefficients and minimal regularity. HAL Preprint 01964299, submitted for publication.

ERN, A., SMEARS, I. & VOHRALÍK, M. (2017) Discrete p-robust $\mathbf{H}$ (div)-liftings and a posteriori estimates for elliptic problems with H^{-1} source terms. *Calcolo*, **54**, 1009–1025.

ERN, A. & VOHRALÍK, M. (2013) Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM J. Sci. Comput.*, **35**, A1761–A1791.

ERN, A. & VOHRALÍK, M. (2015) Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations. *SIAM J. Numer. Anal.*, **53**, 1058–1081.

ERN, A. & VOHRALÍK, M. (2020) Stable broken H^1 and $\mathbf{H}$ (div) polynomial extensions for polynomial-degree-robust potential and flux reconstruction in three space dimensions. *Math. Comp*, **89**, 551–594.

FALK, R. S. & WINTHER, R. (2014) Local bounded cochain projections. *Math. Comp.*, **83**, 2631–2656.

GUDI, T. (2010) A new error analysis for discontinuous finite element methods for linear elliptic problems. *Math. Comp.*, **79**, 2169–2189.

KU, J. (2013) A comment on least-squares finite element methods with minimum regularity assumptions. *Int. J. Numer. Anal. Model.*, **10**, 899–903.

LICHT, M. W. (2019) Smoothed projections and mixed boundary conditions. *Math. Comp.*, **88**, 607–635.

LUCE, R. & WOHLMUTH, B. I. (2004) A local a posteriori error estimator based on equilibrated fluxes. *SIAM J. Numer. Anal.*, **42**, 1394–1414.

MELENK, J. M. & ROJIK, C. (2020) On commuting p-version projection-based interpolation on tetrahedra. *Math. Comp.*, **89**, 45–87.

NÉDÉLEC, J.-C. (1980) Mixed finite elements in $\mathbb{R}^3$. *Numer. Math.*, **35**, 315–341.

NOCHETTO, R. H. & STAMM, B. (2019) A posteriori error estimates for the electric field integral equation on polyhedra. *Contributions to Partial Differential Equations and Applications*, Computational Methods in Applied Sciences, vol. 47. Cham: Springer, pp. 371–394.

PEHLIVANOV, A. I., CAREY, G. F. & LAZAROV, R. D. (1994) Least-squares mixed finite elements for second-order elliptic problems. *SIAM J. Numer. Anal.*, **31**, 1368–1377.

RAVIART, P.-A. & THOMAS, J.-M. (1977) A mixed finite element method for 2nd order elliptic problems. *Mathematical Aspects of Finite Element Methods (Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975),* Lecture Notes in Mathematics, vol. 606. Berlin: Springer, pp. 292–315.

ROBERTS, J. E. & THOMAS, J.-M. (1991) Mixed and hybrid methods. *Handbook of Numerical Analysis, vol. II*. Amsterdam: North-Holland, pp. 523–639.

SCHÖBERL, J. (2001) Commuting quasi-interpolation operators for mixed finite elements. *Technical Report ISC-01-10-MATH*. Texas A&M University.

SCHÖBERL, J. (2005) A multilevel decomposition result in H(curl). *Multigrid, Multilevel and Multiscale Methods (EMG 2005)* (P. Wesseling, C. Oosterlee & P. Hemker eds). Scheveningen, The Hague, The Netherlands: TU Delft.

SCOTT, L. R. & ZHANG, S. (1990) Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, **54**, 483–493.

VEESER, A. (2016) Approximating gradients with continuous piecewise polynomial functions. *Found. Comput. Math.*, **16**, 723–750.

VERFÜRTH, R. (2013) A posteriori error estimation techniques for finite element methods. *Numerical Mathematics and Scientific Computation*. Oxford: Oxford University Press.

A. p-robust constrained–unconstrained equivalence on a simplex

In this appendix we present a way to remove the divergence constraint on a single simplex, and we do this in a polynomial-degree-robust way. This equivalence of constrained and unconstrained local-best approximations is an important consequence of the result of Costabel & McIntosh (2010, Corollary 3.4).

Recall the notation $e_{K,p}(\boldsymbol{v})$ from (3.11), where $\boldsymbol{RTN}_p(K) = \mathscr{P}_p(K;\mathbb{R}^d) + \boldsymbol{x}\mathscr{P}_p(K)$ is the Raviart–Thomas–Nédélec space of degree p on the simplex K, as well as that h_K denotes the diameter of K and ϱ_K the diameter of the largest ball inscribed in K.

Lemma A1 (Local p-robust constrained–unconstrained equivalence). Let a simplex $K \subset \mathbb{R}^d$, $d \geqslant 1$, and $\boldsymbol{v} \in \boldsymbol{H}(\mathrm{div};K)$ be fixed. Let $\boldsymbol{\tau}_{\mathscr{T}}$ be defined as in (3.1). Then there exists a constant C, depending only on the space dimension d and the shape-regularity parameter $\kappa_K := h_K/\varrho_K$ of K, such that

$$e_{K,p}(\boldsymbol{v}) \leqslant \|\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}}\|_K + \frac{h_K}{p+1}\|\nabla\cdot(\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}})\|_K \leqslant C e_{K,p}(\boldsymbol{v}). \tag{A.1}$$

Proof. Since $\nabla\cdot\boldsymbol{\tau}_{\mathscr{T}} = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})$ from (3.1), the first inequality in (A.1) is obvious, so we show the second one. Therein, $\frac{h_K}{p+1}\|\nabla\cdot(\boldsymbol{v}-\boldsymbol{\tau}_{\mathscr{T}})\|_K \leqslant e_{K,p}(\boldsymbol{v})$ trivially holds true for the same reason, so it remains only to bound $\|\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}}\|_K$.

Let $\widetilde{\boldsymbol{\tau}}_{\mathscr{T}}$ be the elementwise $\boldsymbol{L}^2$-projection of $\boldsymbol{v}$ into $\boldsymbol{RTN}_p(\mathscr{T})$, so that

$$[e_{K,p}(\boldsymbol{v})]^2 = \|\boldsymbol{v} - \widetilde{\boldsymbol{\tau}}_{\mathscr{T}}\|_K^2 + \frac{h_K^2}{(p+1)^2}\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K^2.$$

It follows from Costabel & McIntosh (2010, Corollary 3.4) that there exists $\boldsymbol{\varsigma}_K \in \boldsymbol{RTN}_p(K)$ such that $\nabla\cdot\boldsymbol{\varsigma}_K = \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})$ and

$$\|\boldsymbol{\varsigma}_K - \widetilde{\boldsymbol{\tau}}_{\mathscr{T}}\|_K \leqslant C \sup_{\substack{\varphi\in H_0^1(K)\\ \|\nabla\varphi\|_K=1}} \left\{(\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v}) - \nabla\cdot\widetilde{\boldsymbol{\tau}}_{\mathscr{T}}, \varphi)_K\right\}, \tag{A.2}$$

where C only depends on d and κ_K. Since $(\nabla\cdot\boldsymbol{v},\varphi)_K + (\boldsymbol{v},\nabla\varphi)_K = 0$, and since also $(\nabla\cdot\widetilde{\boldsymbol{\tau}}_{\mathscr{T}},\varphi)_K + (\widetilde{\boldsymbol{\tau}}_{\mathscr{T}},\nabla\varphi)_K = 0$ for all $\varphi \in H_0^1(K)$, we see that

$$(\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v}),\varphi)_K - (\nabla\cdot\widetilde{\boldsymbol{\tau}}_{\mathscr{T}},\varphi)_K = (\Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v}) - \nabla\cdot\boldsymbol{v}, \varphi - \Pi^p_{\mathscr{T}}(\varphi))_K - (\boldsymbol{v} - \widetilde{\boldsymbol{\tau}}_{\mathscr{T}}, \nabla\varphi)_K,$$

where we have also freely subtracted $\Pi^p_{\mathscr{T}}(\varphi)$. Therefore, the inequality (A.2) combined with the approximation bound $\|\varphi - \Pi^p_{\mathscr{T}}(\varphi)\|_K \leqslant C\frac{h_K}{p+1}\|\nabla\varphi\|_K$, with a constant C depending only on d and κ_K, implies that

$$\|\boldsymbol{\varsigma}_K - \widetilde{\boldsymbol{\tau}}_{\mathscr{T}}\|_K \leqslant C\left\{\|\boldsymbol{v} - \widetilde{\boldsymbol{\tau}}_{\mathscr{T}}\|_K^2 + \left[\frac{h_K}{p+1}\|\nabla\cdot\boldsymbol{v} - \Pi^p_{\mathscr{T}}(\nabla\cdot\boldsymbol{v})\|_K\right]^2\right\}^{\frac{1}{2}} = C e_{K,p}(\boldsymbol{v}).$$

Finally, owing to the triangle inequality $\|\boldsymbol{v} - \boldsymbol{\varsigma}_K\|_K \leqslant \|\boldsymbol{v} - \widetilde{\boldsymbol{\tau}}_{\mathscr{T}}\|_K + \|\widetilde{\boldsymbol{\tau}}_{\mathscr{T}} - \boldsymbol{\varsigma}_K\|_K$, we infer that $\|\boldsymbol{v} - \boldsymbol{\varsigma}_K\|_K \leqslant C e_{K,p}(\boldsymbol{v})$. Consequently, the definition of $\boldsymbol{\tau}_{\mathscr{T}}$ as the minimizer in (3.1) implies that $\|\boldsymbol{v} - \boldsymbol{\tau}_{\mathscr{T}}\|_K \leqslant \|\boldsymbol{v} - \boldsymbol{\varsigma}_K\|_K$, and this yields the second bound in (A.1). $\square$