Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Exponential ReLU Neural Network Approximation Rates for Point and Edge Singularities

C. Marcati and J. A. A. Opschoor and P. C. Petersen and Ch. Schwab

Research Report No. 2020-65

October 2020
Latest revision: October 2020

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Exponential ReLU Neural Network Approximation Rates for Point and Edge Singularities

Carlo Marcati* Joost A. A. Opschoor* Philipp C. Petersen† Christoph Schwab*

October 19, 2020

Abstract

We prove exponential expressivity with stable ReLU Neural Networks (ReLU NNs) in $H^1(\Omega)$ for weighted analytic function classes in certain polytopal domains Ω, in space dimension $d = 2, 3$. Functions in these classes are locally analytic on open subdomains $D \subset \Omega$, but may exhibit isolated point singularities in the interior of Ω or corner and edge singularities at the boundary $\partial\Omega$. The exponential expression rate bounds proved here imply uniform exponential expressivity by ReLU NNs of solution families for several elliptic boundary and eigenvalue problems with analytic data. The exponential approximation rates are shown to hold in space dimension $d = 2$ on Lipschitz polygons with straight sides, and in space dimension $d = 3$ on Fichera-type polyhedral domains with plane faces. The constructive proofs indicate in particular that NN depth and size increase poly-logarithmically with respect to the target NN approximation accuracy $\varepsilon > 0$ in $H^1(\Omega)$. The results cover in particular solution sets of linear, second order elliptic PDEs with analytic data and certain nonlinear elliptic eigenvalue problems with analytic nonlinearities and singular, weighted analytic potentials as arise in electron structure models. In the latter case, the functions correspond to electron densities that exhibit isolated point singularities at the positions of the nuclei. Our findings provide in particular mathematical foundation of recently reported, successful uses of deep neural networks in variational electron structure algorithms.

Keywords: Neural networks, finite element methods, exponential convergence, analytic regularity, singularities

Subject Classification: 35Q40, 41A25, 41A46, 65N30

Contents

*Seminar for Applied Mathematics, ETH Zürich, CH8092 Zürich, Switzerland,
email: {carlo.marcati, joost.opschoor, christoph.schwab}@sam.math.ethz.ch
†Faculty of Mathematics and Research Platform Data Science @ Uni Vienna, University of Vienna, 1090, Vienna, Austria,
email: Philipp.Petersen@univie.ac.at

1 Introduction

The application of neural networks (NNs) as approximation architecture in numerical solution methods of partial differential equations (PDEs), possibly on high-dimensional parameter- and state-spaces, has attracted significant and increasing attention in recent years. We mention only [47, 6, 38, 39, 45] and the references therein. In these works, the solution of elliptic and parabolic boundary value problems is approximated by NNs which are found by minimization of a residual of the NN in the PDE.

A necessary condition for the performance of the mentioned NN-based numerical approximation methods is a high rate of approximation which is to hold uniformly over the solution set associated with the PDE under consideration. For elliptic boundary and eigenvalue problems, the function classes that weak solutions of the problems belong to are well known. Moreover, in many cases, representation systems such as splines or polynomials that achieve optimal linear or nonlinear approximation rates for functions belonging to these function classes have been identified. For functions belonging to a Sobolev or Besov type smoothness space of finite differentiation order such as continuously differentiable or Sobolev-regular functions on a bounded domain, upper bounds for the approximation rate by NNs were established for example in [50, 11, 51, 25, 48]. Here, we only mentioned results that use the ReLU activation function. Besides, for PDEs, the solutions of which have a particular structure, approximation rates of the solution that go beyond classical smoothness-based results can be achieved, such as in [7, 44, 22, 2, 19]. Again, we confine the list to publications with approximation rates for NNs with the ReLU activation function (referred to as ReLU NNs below).

In the present paper, we analyze approximation rates provided by ReLU NNs for solution classes of linear and nonlinear elliptic source and eigenvalue problems on polygonal and polyhedral domains. Mathematical results on weighted analytic regularity [13, 14, 12, 1, 3, 15, 32, 5, 28, 31] imply that these classes consist of functions that are *analytic with possible corner, edge, and corner-edge singularities*.

Our analysis provides, for the aforementioned functions, approximation errors in Sobolev norms that decay exponentially in terms of the number of parameters M of the ReLU NNs.

1.1 Contribution

The principal contribution of this work is threefold:

1. We prove, in Theorem 4.3, a general result on the approximation by ReLU NNs of weighted analytic function classes on $Q := (0, 1)^d$, where $d = 2, 3$. The analytic regularity of solutions is quantified via countably normed, analytic classes, based on weighted Sobolev spaces of Kondrat'ev type in Q, which admit corner and, in space dimension $d = 3$, also edge singularities. Such classes were introduced, e.g., in [3, 1, 12, 13, 14, 5] and in the references there. We prove exponential

expression rates by ReLU NNs in the sense that for a number M of free parameters of the NNs, the approximation error is bounded, in the H^1-norm, by $C\exp(-bM^{1/(2d+1)})$ for constants $b, C > 0$.

2. Based on the ReLU NN approximation rate bound of Theorem 4.3, we establish, in Section 5, approximation results for solutions of different types of PDEs by NNs with ReLU activation. Concretely, in Section 5.1.1, we study the reapproximation of solutions of nonlinear Schrödinger equations with singular potentials in space dimension $d = 2, 3$. We prove that for solutions which are contained in weighted, analytic classes in Ω, ReLU NNs (whose realizations are continuous, piecewise affine) with at most M free parameters yield an approximation with accuracy of the order $\exp(-bM^{1/(2d+1)})$ for some $b > 0$. Importantly, this convergence is in the $H^1(\Omega)$-norm. In Section 5.1.2, we establish the same exponential approximation rates for the eigenstates of the Hartree-Fock model with singular potential in $\mathbb{R}^3$. This result constitutes the first, to our knowledge, mathematical underpinning of the recently reported, high efficiency of various NN-based approaches in variational electron structure computations, e.g., [37, 18, 16].

   In Section 5.2, we demonstrate the same approximation rates also for elliptic boundary value problems with analytic coefficients and analytic right-hand sides, in plane, polygonal domains Ω. The approximation error of the NNs is, again, bound in the $H^1(\Omega)$-norm. We also infer an exponential NN expression rate bound for corresponding traces, in $H^{1/2}(\partial\Omega)$ and for viscous, incompressible flow.

   Finally, in Section 5.3, we obtain the same asymptotic exponential rates for the approximation of solutions to elliptic boundary value problems, with analytic data, on so-called Fichera-type domains $\Omega_F \subset \mathbb{R}^3$ (being, roughly speaking, finite unions of tensorized hexahedra). These solutions exhibit corner, edge and corner-edge singularities.

3. The exponential approximation rates of the ReLU NNs established here are based on emulating corresponding variable grid and degree ("hp") piecewise polynomial approximations. In particular, our construction comprises tensor product hp-approximations on Cartesian products of geometric partitions of intervals. In Theorem A.25, we establish novel *tensor product hp-approximation results* for weighted analytic functions on Q of the form $\|u - u_{\mathsf{hp}}\|_{H^1(Q)} \leq C\exp(-b\sqrt[2d]{N})$ for $d = 1, 2, 3$, where N is the number of degrees of freedom in the representation of u_{hp} and $C, b > 0$ are independent of N (but depend on u). The geometric partitions employed in Q and the architectures of the constructed ReLU NNs are of tensor product structure, and generalize to space dimension $d > 3$.

   We note that hp-approximations based on non-tensor-product, geometric partitions of Q into hexahedra have been studied before e.g. in [40, 41] in space dimension $d = 3$. There, the rate of $\|u - u_{\mathsf{hp}}\|_{H^1(Q)} \lesssim \exp(-b\sqrt[5]{N})$ was found. Being based on tensorization, the present construction of exponentially convergent, tensorized hp-approximations in Appendix A does not invoke the rather involved polynomial trace liftings in [40, 41], and is interesting in its own right: the geometric and mathematical simplification comes at the expense of a slightly smaller (still exponential) rate of approximation. Moreover, we expect that this construction of u_{hp} will allow a rather direct derivation of rank bounds for tensor structured function approximation of u in Q, generalizing results in [20, 21] and extending [30] from point to edge and corner-edge singularities.

1.2 Neural network approximation of weighted analytic function classes

The proof strategy that yields the main result, Theorem 4.3, is as follows. We first establish exponential approximation rates in the Sobolev space H^1 for tensor-product, so-called hp-finite elements for weighted analytic functions. Then, we re-approximate the corresponding quasi-interpolants by ReLU NNs.

The emulation of hp-finite element approximation by ReLU NNs is fundamentally based on the approximate multiplication network formalized in [50]. Based on the approximate multiplication operation and an extension thereof to errors measured with respect to $W^{1,q}$-norms, for $q \in [1, \infty]$, we established already in [34] a reapproximation theorem of univariate splines of order $p \in \mathbb{N}$ on arbitrary meshes with $N \in \mathbb{N}$ cells. There, we observed that there exists a NN that reapproximates a variable-order, free-knot spline u in the H^1-norm up to an error of $\varepsilon > 0$ with a number of free parameters that scales logarithmically in ε and $|u|_{H^1}$, linearly in N and quadratically in p. We recall this result in Proposition 3.7 below.

From this, it is apparent by the triangle inequality that, in univariate approximation problems where hp-finite elements yield exponential approximation rates, also ReLU NNs achieve exponential approximation rates, (albeit with a possibly smaller exponent, because of the quadratic dependence on p, see [34, Theorem 5.12]).

The extension of this result to higher dimensions for high-order finite elements that are built from univariate finite elements by tensorization is based on the underlying compositionality of NNs. Because of that, it is possible to compose a NN implementing a multiplication of d inputs with d approximations of univariate finite elements. We introduce a formal framework describing these operations in Section 3.

We remark that for high-dimensional functions with a radial structure, of which the univariate radial profile allows an exponentially convergent hp-approximation, exponential convergence was obtained in [34, Section 6] by composing ReLU approximations of univariate splines with an exponentially convergent approximation of the Euclidean norm, obtaining exponential convergence without the curse of dimensionality.

1.3 Outline

The manuscript is structured as follows: in Section 2, in particular Section 2.2, we review the weighted function spaces which will be used to describe the weighted analytic function classes in polytopes Ω that underlie our approximation results. In Section 2.3, we present an approximation result by tensor-product hp-finite elements for functions from the weighted analytic class. A proof of this result is provided in Appendix A. In Section 3 we review definitions of NNs and a "ReLU calculus" from [7, 36] whose operations will be required in the ensuing NN approximation results.

In Section 4, we state and prove the key results of the present paper. In Section 5, we illustrate our results by deducing novel NN expression rate bounds for solution classes of several concrete examples of elliptic boundary-value and eigenvalue problems where solutions belong to the weighted analytic function classes introduced in Section 2. Some of the more technical proofs of Section 5 are deferred to Appendix B. In Section 6, we briefly recapitulate the principal mathematical results of this paper and indicate possible consequences and further directions.

2 Setting and functional spaces

We start by recalling some general notation that will be used throughout the paper. We also introduce some tools that are required to describe two and three dimensional domains as well as the associated weighted Sobolev spaces.

2.1 Notation

For $\alpha \in \mathbb{N}_0^d$, define $|\alpha| := \alpha_1 + \cdots + \alpha_d$ and $|\alpha|_\infty := \max\{\alpha_1, \dots, \alpha_d\}$. When we indicate a relation on $|\alpha|$ or $|\alpha|_\infty$ in the subscript of a sum, we mean the sum over all multi-indices that fulfill that relation: e.g., for a $k \in \mathbb{N}_0$

$$\sum_{|\alpha|\leq k} = \sum_{\alpha\in\mathbb{N}_0^d : |\alpha|\leq k}.$$

For a domain $\Omega \subset \mathbb{R}^d$, $k \in \mathbb{N}_0$ and for $1 \leq p \leq \infty$, we indicate by $W^{k,p}(\Omega)$ the classical $L^p(\Omega)$-based Sobolev space of order k. We write $H^k(\Omega) = W^{k,2}(\Omega)$. We introduce the norms $\|\cdot\|_{W^{1,p}_{\mathrm{mix}}(\Omega)}$ as

$$\|v\|^p_{W^{1,p}_{\mathrm{mix}}(\Omega)} := \sum_{|\alpha|_\infty\leq 1} \|\partial^\alpha v\|^p_{L^p(\Omega)},$$

with associated spaces

$$W^{1,p}_{\mathrm{mix}}(\Omega) := \left\{ v \in L^p(\Omega) : \|v\|_{W^{1,p}_{\mathrm{mix}}(\Omega)} < \infty \right\}.$$

We denote $H^1_{\mathrm{mix}}(\Omega) = W^{1,2}_{\mathrm{mix}}(\Omega)$. For $\Omega = I_1 \times \cdots \times I_d$, with bounded intervals $I_j \subset \mathbb{R}$, $j = 1, \dots, d$, $H^1_{\mathrm{mix}}(\Omega) = H^1(I_1) \otimes \cdots \otimes H^1(I_d)$ with Hilbertian tensor products. Throughout, C will denote a generic positive constant whose value may change at each appearance, even within an equation.

The ℓ^p norm, $1 \leq p \leq \infty$, on $\mathbb{R}^n$ is denoted by $\|x\|_p$. The number of nonzero entries of a vector or matrix x is denoted by $\|x\|_0$.

Three dimensional domain. Let $\Omega \subset \mathbb{R}^3$ be a bounded, polygonal/polyhedral domain. Let $\mathcal{C}$ denote a set of isolated points, situated either at the corners of Ω or in the interior of Ω (that we refer to as the singular corners in either case, for simplicity), and let $\mathcal{E}$ be a subset of the edges of Ω (the singular edges). Furthermore, denote by $\mathcal{E}_c \subset \mathcal{E}$ the set of singular edges abutting at a corner c. For each $c \in \mathcal{C}$ and each $e \in \mathcal{E}$, we introduce the following weights:

$$r_c(x) := |x - c| = \mathrm{dist}(x, c), \qquad r_e(x) := \mathrm{dist}(x, e), \qquad \rho_{ce}(x) := \frac{r_e(x)}{r_c(x)} \quad \text{for } x \in \Omega.$$

For $\varepsilon > 0$, we define edge-, corner-, and corner-edge neighborhoods:

$$\Omega_e^\varepsilon := \left\{x \in \Omega : r_e(x) < \varepsilon \text{ and } r_c(x) > \varepsilon, \forall c \in \mathcal{C}\right\}, \Omega_c^\varepsilon := \left\{x \in \Omega : r_c(x) < \varepsilon \text{ and } \rho_{ce}(x) > \varepsilon, \forall e \in \mathcal{E}\right\},$$
$$\Omega_{ce}^\varepsilon := \left\{x \in \Omega : r_c(x) < \varepsilon \text{ and } \rho_{ce}(x) < \varepsilon\right\}.$$

We fix a value $\widehat{\varepsilon} > 0$ small enough so that $\Omega_c^{\widehat{\varepsilon}} \cap \Omega_{c'}^{\widehat{\varepsilon}} = \varnothing$ for all $c \neq c' \in \mathcal{C}$ and $Q_{ce}^{\widehat{\varepsilon}} \cap \Omega_{ce'}^{\widehat{\varepsilon}} = \Omega_e^{\widehat{\varepsilon}} \cap \Omega_{e'}^{\widehat{\varepsilon}} = \varnothing$ for all singular edges $e \neq e'$. In the sequel, we omit the dependence of the subdomains on $\widehat{\varepsilon}$. Let also

$$\Omega_{\mathcal{C}} := \bigcup_{c\in\mathcal{C}} \Omega_c, \qquad \Omega_{\mathcal{E}} := \bigcup_{e\in\mathcal{E}} \Omega_e, \qquad \Omega_{\mathcal{C}\mathcal{E}} := \bigcup_{c\in\mathcal{C}} \bigcup_{e\in\mathcal{E}_c} \Omega_{ce},$$

and

$$\Omega_0 := \Omega \setminus \overline{(\Omega_{\mathcal{C}} \cup \Omega_{\mathcal{E}} \cup \Omega_{\mathcal{C}\mathcal{E}})}.$$

In each subdomain Ω_{ce} and Ω_e, for any multi-index $\alpha \in \mathbb{N}_0^3$, we denote by $\alpha_\parallel$ the multi-index whose component in the coordinate direction parallel to e is equal to the component of α in the same direction, and which is zero in every other component. Moreover, we set $\alpha_\perp := \alpha - \alpha_\parallel$.

Two dimensional domain. Let $\Omega \subset \mathbb{R}^2$ be a polygon. We adopt the convention that $\mathcal{E} := \varnothing$. For $c \in \mathcal{C}$, we define

$$Q_c^\varepsilon := \left\{x \in \Omega : r_c(x) < \varepsilon\right\}.$$

As in the three dimensional case, we fix a sufficiently small $\widehat{\varepsilon} > 0$ so that $\Omega_c^{\widehat{\varepsilon}} \cap \Omega_{c'}^{\widehat{\varepsilon}} = \varnothing$ for $c \neq c' \in \mathcal{C}$ and write $\Omega_c = \Omega_c^{\widehat{\varepsilon}}$. Furthermore, $\Omega_{\mathcal{C}}$ is defined as for $d = 3$, and $\Omega_0 := \Omega \setminus \overline{\Omega_{\mathcal{C}}}$.

2.2 Weighted spaces with nonhomogeneous norms

We introduce classes of weighted, analytic functions in space dimension $d = 3$, as arise in analytic regularity theory for linear, elliptic boundary value problems in polyhedra, in the particular form introduced in [5]. There, the structure of the weights is in terms of Cartesian coordinates which is particularly suited for the presently adopted, tensorized approximation architectures.

The definition of the corresponding classes when $d = 2$ is analogous. For a *weight exponent vector* $\underline{\gamma} = \{\gamma_c, \gamma_e,\ c \in \mathcal{C}, e \in \mathcal{E}\}$, we introduce the *nonhomogeneous, weighted Sobolev norms*

$$\begin{aligned}\|v\|_{\mathcal{J}^k_{\underline{\gamma}}(\Omega)} := \sum_{|\alpha|\le k} \|\partial^\alpha v\|_{L^2(\Omega_0)} &+ \sum_{c\in\mathcal{C}} \sum_{|\alpha|\le k} \|r_c^{(|\alpha|-\gamma_c)_+} \partial^\alpha v\|_{L^2(\Omega_c)} \\ &+ \sum_{e\in\mathcal{E}} \sum_{|\alpha|\le k} \|r_e^{(|\alpha_\perp|-\gamma_e)_+} \partial^\alpha v\|_{L^2(\Omega_e)} \\ &+ \sum_{c\in\mathcal{C}} \sum_{e\in\mathcal{E}_c} \sum_{|\alpha|\le k} \|r_c^{(|\alpha|-\gamma_c)_+} \rho_{ce}^{(|\alpha_\perp|-\gamma_e)_+} \partial^\alpha v\|_{L^2(\Omega_{ce})}\end{aligned}$$

where $(x)_+ = \max\{0, x\}$. Moreover, we define the associated function space by

$$\mathcal{J}^k_{\underline{\gamma}}(\Omega; \mathcal{C}, \mathcal{E}) := \left\{v \in L^2(\Omega) : \|v\|_{\mathcal{J}^k_{\underline{\gamma}}(\Omega)} < \infty\right\}.$$

Furthermore,

$$\mathcal{J}^\infty_{\underline{\gamma}}(\Omega; \mathcal{C}, \mathcal{E}) = \bigcap_{k\in\mathbb{N}_0} \mathcal{J}^k_{\underline{\gamma}}(\Omega; \mathcal{C}, \mathcal{E}).$$

For $A, C > 0$, we define the space of weighted analytic functions with nonhomogeneous norm as

$$\begin{aligned}\mathcal{J}^\varpi_{\underline{\gamma}}(\Omega; \mathcal{C}, \mathcal{E}; C, A) := \Big\{v \in \mathcal{J}^\infty_{\underline{\gamma}}(\Omega; \mathcal{C}, \mathcal{E}) :\ & \sum_{|\alpha|=k} \|\partial^\alpha v\|_{L^2(\Omega_0)} \le CA^k k!, \\ & \sum_{|\alpha|=k} \|r_c^{(|\alpha|-\gamma_c)_+} \partial^\alpha v\|_{L^2(\Omega_c)} \le CA^k k! \quad \forall c \in \mathcal{C}, \\ & \sum_{|\alpha|=k} \|r_e^{(|\alpha_\perp|-\gamma_e)_+} \partial^\alpha v\|_{L^2(\Omega_e)} \le CA^k k! \quad \forall e \in \mathcal{E}, \\ & \sum_{|\alpha|=k} \|r_c^{(|\alpha|-\gamma_c)_+} \rho_{ce}^{(|\alpha_\perp|-\gamma_e)_+} \partial^\alpha v\|_{L^2(\Omega_{ce})} \le CA^k k! \\ & \forall c \in \mathcal{C} \text{ and } \forall e \in \mathcal{E}_c, \forall k \in \mathbb{N}_0\Big\}.\end{aligned} \tag{2.1}$$

Finally, we denote

$$\mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega; \mathcal{C}, \mathcal{E}) \coloneqq \bigcup_{C,A>0} \mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega; \mathcal{C}, \mathcal{E}; C, A).$$

2.3 Approximation of weighted analytic functions on tensor product geometric meshes

The approximation result of weighted analytic functions via NNs that we present below is based on emulating an approximation strategy of tensor product hp-finite elements. In this section, we present this hp-finite element approximation. Let $I \subset \mathbb{R}$ be an interval. A *partition of I into $N \in \mathbb{N}$ intervals* is a set $\mathcal{G}$ such that $|\mathcal{G}| = N$, all elements of $\mathcal{G}$ are disjoint, connected, and open subsets of I, and

$$\bigcup_{U \in \mathcal{G}} \overline{U} = \overline{I}.$$

We denote, for all $p \in \mathbb{N}_0$, by $\mathbb{Q}_p(\mathcal{G})$ the piecewise polynomials of degree p on $\mathcal{G}$.

One specific partition of $I = [0, 1]$ is given by the *one dimensional geometrically graded grid*, which for $\sigma \in (0, 1/2]$ and $\ell \in \mathbb{N}$, is given by

$$\mathcal{G}^{\ell}_1 \coloneqq \left\{ J^{\ell}_k,\, k = 0, \dots, \ell \right\}, \quad \text{where } J^{\ell}_0 \coloneqq (0, \sigma^{\ell}) \quad \text{and} \quad J^{\ell}_k \coloneqq (\sigma^{\ell-k+1}, \sigma^{\ell-k}),\ k = 1, \dots, \ell. \tag{2.2}$$

Theorem 2.1. *Let $d \in \{2, 3\}$ and $Q \coloneqq (0, 1)^d$. Let $\mathcal{C} = \{c\}$ where c is one of the corners of Q and let $\mathcal{E} = \mathcal{E}_c$ contain the edges adjacent to c when $d = 3$, $\mathcal{E} = \varnothing$ when $d = 2$. Further assume given constants $C_f, A_f > 0$, and*

$$\begin{aligned} &\underline{\gamma} = \{\gamma_c : c \in \mathcal{C}\}, && \text{with } \gamma_c > 1, \text{ for all } c \in \mathcal{C} && \text{if } d = 2,\\ &\underline{\gamma} = \{\gamma_c, \gamma_e : c \in \mathcal{C}, e \in \mathcal{E}\}, && \text{with } \gamma_c > 3/2 \text{ and } \gamma_e > 1, \text{ for all } c \in \mathcal{C} \text{ and } e \in \mathcal{E} && \text{if } d = 3. \end{aligned}$$

Then, there exist $C_p > 0$, $C_L > 0$ such that, for every $0 < \varepsilon < 1$, there exist $p, L \in \mathbb{N}$ with

$$p \leq C_p(1 + |\log(\varepsilon)|), \quad L \leq C_L(1 + |\log(\varepsilon)|),$$

so that there exist piecewise polynomials

$$v_i \in \mathbb{Q}_p(\mathcal{G}^L_1) \cap H^1(I), \qquad i = 1, \dots, N_{1\mathrm{d}},$$

with $N_{1\mathrm{d}} = (L+1)p+1$, and, for all $f \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E}; C_f, A_f)$ there exists a d-dimensional array of coefficients

$$c = \left\{ c_{i_1 \dots i_d} : (i_1, \dots, i_d) \in \{1, \dots, N_{1\mathrm{d}}\}^d \right\}$$

such that

1. *For every $i = 1, \dots N_{1\mathrm{d}}$, $\mathrm{supp}(v_i)$ intersects either a single interval or two neighboring subintervals of $\mathcal{G}^L_1$. Furthermore, there exist constants C_v, b_v depending only on C_f, A_f, σ such that*

$$\|v_i\|_{\infty} \leq 1, \qquad \|v_i\|_{H^1(I)} \leq C_v \varepsilon^{-b_v}, \qquad \forall i = 1, \dots, N_{1\mathrm{d}}. \tag{2.3}$$

2. *There holds*

$$\|f - \sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} c_{i_1\dots i_d} \phi_{i_1 \dots i_d}\|_{H^1(Q)} \leq \varepsilon \qquad \text{with} \qquad \phi_{i_1\dots i_d} = \bigotimes_{j=1}^{d} v_{i_j},\ \forall i_1, \dots, i_d = 1, \dots, N_{1\mathrm{d}}. \tag{2.4}$$

3. *$\|c\|_{\infty} \leq C_2(1 + |\log(\varepsilon)|)^d$ and $\|c\|_1 \leq C_c(1 + |\log(\varepsilon)|)^{2d}$, for $C_2, C_c > 0$ independent of p, L, ε.*

We present the proof in Subsection A.9.3 after developing an appropriate framework of hp-approximation in Section A.

3 Basic ReLU neural network calculus

In the sequel, we distinguish between a neural network, as a collection of weights, and the associated *realization of the NN*. This is a function that is determined through the weights and an activation function. In this paper, we only consider the so-called ReLU activation:

$$\varrho : \mathbb{R} \to \mathbb{R} : x \mapsto \max\{0, x\}.$$

Definition 3.1 ([36, Definition 2.1]). *Let $d, L \in \mathbb{N}$. A* neural network Φ with input dimension d and L layers *is a sequence of matrix-vector tuples*

$$\Phi = \big((A_1, b_1), (A_2, b_2), \dots, (A_L, b_L)\big),$$

where $N_0 := d$ and $N_1, \dots, N_L \in \mathbb{N}$, and where $A_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b_\ell \in \mathbb{R}^{N_\ell}$ for $\ell = 1, ..., L$.

For a NN Φ, we define the associated realization of the NN Φ *as*

$$\mathrm{R}(\Phi) : \mathbb{R}^d \to \mathbb{R}^{N_L} : x \mapsto x_L =: \mathrm{R}(\Phi)(x),$$

where the output $x_L \in \mathbb{R}^{N_L}$ results from

$$\begin{aligned} x_0 &:= x, \\ x_\ell &:= \varrho(A_\ell\, x_{\ell-1} + b_\ell) \quad \text{for } \ell = 1, \dots, L-1, \\ x_L &:= A_L\, x_{L-1} + b_L. \end{aligned}$$

Here ϱ is understood to act component-wise on vector-valued inputs, i.e., for $y = (y^1, \dots, y^m) \in \mathbb{R}^m$, $\varrho(y) := (\varrho(y^1), \dots, \varrho(y^m))$. We call $N(\Phi) := d + \sum_{j=1}^{L} N_j$ the number of neurons of the NN Φ, $\mathrm{L}(\Phi) := L$ *the* number of layers *or* depth, $\mathrm{M}_j(\Phi) := \|A_j\|_0 + \|b_j\|_0$ *the* number of nonzero weights in the j-th layer, *and* $\mathrm{M}(\Phi) := \sum_{j=1}^{L} \mathrm{M}_j(\Phi)$ *the* number of nonzero weights of Φ, *also referred to as its* size. *We refer to N_L as the* dimension of the output layer of Φ.

3.1 Concatenation, parallelization, emulation of identity

An essential component in the ensuing proofs is to construct NNs out of simpler building blocks. For instance, given two NNs, we would like to identify another NN so that the realization of it equals the sum or the composition of the first two NNs. To describe these operations precisely, we introduce a formalism of operations on NNs below. The first of these operations is the concatenation.

Proposition 3.2 (NN concatenation, [36, Remark 2.6]). *Let $L_1, L_2 \in \mathbb{N}$, and let Φ^1, Φ^2 be two NNs of respective depths L_1 and L_2 such that $N_0^1 = N_{L_2}^2 =: d$, i.e., the input layer of Φ^1 has the same dimension as the output layer of Φ^2.*

Then, there exists a NN $\Phi^1 \odot \Phi^2$, called the sparse concatenation of Φ^1 and Φ^2, *such that $\Phi^1 \odot \Phi^2$ has $L_1 + L_2$ layers, $\mathrm{R}(\Phi^1 \odot \Phi^2) = \mathrm{R}(\Phi^1) \circ \mathrm{R}(\Phi^2)$ and $\mathrm{M}\left(\Phi^1 \odot \Phi^2\right) \leq 2\,\mathrm{M}\left(\Phi^1\right) + 2\,\mathrm{M}\left(\Phi^2\right)$.*

The second fundamental operation on NNs is parallelization, achieved with the following construction.

Proposition 3.3 (NN parallelization, [36, Definition 2.7]). *Let $L, d \in \mathbb{N}$ and let Φ^1, Φ^2 be two NNs with L layers and with d-dimensional input each. Then there exists a NN $\mathrm{P}(\Phi^1, \Phi^2)$ with d-dimensional input and L layers, which we call the* parallelization of Φ^1 and Φ^2, *such that*

$$\mathrm{R}\left(\mathrm{P}\left(\Phi^1, \Phi^2\right)\right)(x) = \left(\mathrm{R}\left(\Phi^1\right)(x), \mathrm{R}\left(\Phi^2\right)(x)\right), \text{ for all } x \in \mathbb{R}^d$$

and $\mathrm{M}(\mathrm{P}(\Phi^1, \Phi^2)) = \mathrm{M}(\Phi^1) + \mathrm{M}(\Phi^2)$.

Proposition 3.3 requires two NNs to have the same depth. If two NNs have different depth, then we can artificially enlarge one of them by concatenating with a NN that implements the identity. One possible construction of such a NN is presented next.

Proposition 3.4 (NN emulation of Id, [36, Remark 2.4]). *For every $d, L \in \mathbb{N}$ there exists a NN $\Phi^{\mathrm{Id}}_{d,L}$ with $\mathrm{L}(\Phi^{\mathrm{Id}}_{d,L}) = L$ and $\mathrm{M}(\Phi^{\mathrm{Id}}_{d,L}) \leq 2dL$, such that $\mathrm{R}(\Phi^{\mathrm{Id}}_{d,L}) = \mathrm{Id}_{\mathbb{R}^d}$.*

Finally, we sometimes require a parallelization of NNs that do not share inputs.

Proposition 3.5 (Full parallelization of NNs with distinct inputs, [7, Setting 5.2]). *Let $L \in \mathbb{N}$ and let*

$$\Phi^1 = \left(\left(A_1^1, b_1^1\right), \ldots, \left(A_L^1, b_L^1\right)\right), \quad \Phi^2 = \left(\left(A_1^2, b_1^2\right), \ldots, \left(A_L^2, b_L^2\right)\right)$$

be two NNs with L layers each and with input dimensions $N_0^1 = d_1$ and $N_0^2 = d_2$, respectively.

Then there exists a NN, denoted by $\mathrm{FP}(\Phi^1, \Phi^2)$, with d-dimensional input where $d = (d_1 + d_2)$ and L layers, which we call the full parallelization of Φ^1 and Φ^2, *such that for all $x = (x_1, x_2) \in \mathbb{R}^d$ with $x_i \in \mathbb{R}^{d_i}$, $i = 1, 2$*

$$\mathrm{R}\left(\mathrm{FP}\left(\Phi^1, \Phi^2\right)\right)(x_1, x_2) = \left(\mathrm{R}\left(\Phi^1\right)(x_1), \mathrm{R}\left(\Phi^2\right)(x_2)\right)$$

and $\mathrm{M}(\mathrm{FP}(\Phi^1, \Phi^2)) = \mathrm{M}(\Phi^1) + \mathrm{M}(\Phi^2)$.

Proof. Set $\mathrm{FP}\left(\Phi^1, \Phi^2\right) := \left(\left(A_1^3, b_1^3\right), \ldots, \left(A_L^3, b_L^3\right)\right)$ where, for $j = 1, \ldots, L$, we define

$$A_j^3 := \begin{pmatrix} A_j^1 & 0 \\ 0 & A_j^2 \end{pmatrix} \text{ and } b_j^3 := \begin{pmatrix} b_j^1 \\ b_j^2 \end{pmatrix}.$$

All properties of $\mathrm{FP}\left(\Phi^1, \Phi^2\right)$ claimed in the statement of the proposition follow immediately from the construction. □

3.2 Emulation of multiplication and piecewise polynomials

In addition to the basic operations above, we use two types of functions that we can approximate especially efficiently with NNs. These are high dimensional multiplication functions and univariate piecewise polynomials. We first give the result of an emulation of a multiplication in arbitrary dimension.

Proposition 3.6 ([11, Lemma C.5], [35, Proposition 2.6]). *There exists a constant $C > 0$ such that, for every $0 < \varepsilon < 1$, $d \in \mathbb{N}$ and $M \geq 1$ there is a NN $\Pi_{\varepsilon,M}^d$ with d-dimensional input- and one-dimensional output, so that*

$$\left| \prod_{\ell=1}^d x_\ell - \mathrm{R}(\Pi_{\varepsilon,M}^d)(x) \right| \leq \varepsilon, \text{ for all } x = (x_1, \ldots, x_d) \in [-M, M]^d,$$

$$\left| \frac{\partial}{\partial x_j} \prod_{\ell=1}^d x_\ell - \frac{\partial}{\partial x_j} \mathrm{R}(\Pi_{\varepsilon,M}^d)(x) \right| \leq \varepsilon, \text{ for almost every } x = (x_1, \ldots, x_d) \in [-M, M]^d \text{ and all } j = 1, \ldots, d,$$

and $\mathrm{R}(\Pi_{\varepsilon,M}^d)(x) = 0$ if $\prod_{\ell=1}^d x_\ell = 0$, for all $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Additionally, $\Pi_{\varepsilon,M}^d$ satisfies

$$\max\left\{ \mathrm{L}\left(\Pi_{\varepsilon,M}^d\right), \mathrm{M}\left(\Pi_{\varepsilon,M}^d\right)\right\} \leq C\left(1 + d\log(dM^d/\varepsilon)\right).$$

In addition to the high-dimensional multiplication, we can efficiently approximate univariate continuous, piecewise polynomial functions by realizations of NNs with the ReLU activation function.

Proposition 3.7 ([34, Proposition 5.1]). *There exists a constant $C > 0$ such that, for all $\boldsymbol{p} = (p_i)_{i \in \{1,\ldots,N_{\mathrm{int}}\}} \subset \mathbb{N}$, for all partitions $\mathcal{T}$ of $I = (0,1)$ into N_{int} open, disjoint, connected subintervals I_i, $i = 1, \ldots, N_{\mathrm{int}}$, for all $v \in S_{\boldsymbol{p}}(I, \mathcal{T}) := \{v \in H^1(I) : v|_{I_i} \in \mathbb{P}_{p_i}(I_i), i = 1, \ldots, N_{\mathrm{int}}\}$, and for every $0 < \varepsilon < 1$, there exist NNs $\{\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\}_{\varepsilon \in (0,1)}$ such that for all $1 \leq q' \leq \infty$ it holds that*

$$\begin{aligned}
\left\| v - \mathrm{R}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right)\right\|_{W^{1,q'}(I)} &\leq \varepsilon \left|v\right|_{W^{1,q'}(I)}, \\
\mathrm{L}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right) &\leq C(1 + \log(p_{\max}))\left(p_{\max} + \left|\log\varepsilon\right|\right), \\
\mathrm{M}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right) &\leq CN_{\mathrm{int}}(1 + \log(p_{\max}))\left(p_{\max} + \left|\log\varepsilon\right|\right) + C\sum_{i=1}^{N_{\mathrm{int}}} p_i\left(p_i + \left|\log\varepsilon\right|\right),
\end{aligned}$$

where $p_{\max} := \max\{p_i : i = 1, \ldots, N_{\mathrm{int}}\}$. In addition, $\mathrm{R}\left(\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\right)(x_j) = v(x_j)$ for all $j \in \{0, \ldots, N_{\mathrm{int}}\}$, where $\{x_j\}_{j=0}^{N_{\mathrm{int}}}$ are the nodes of $\mathcal{T}$.

Remark 3.8. *It is not hard to see that the result holds also for $I = (a,b)$, where $a, b \in \mathbb{R}$, with $C > 0$ depending on $(b-a)$. Indeed, for any $v \in H^1((a,b))$ the concatenation of v with the invertible, affine map $T : x \mapsto (x-a)/(b-a)$ is in $H^1((0,1))$. Applying Proposition 3.7 yields NNs $\{\Phi_\varepsilon^{v,\mathcal{T},\boldsymbol{p}}\}_{\varepsilon \in (0,1)}$ approximating $v \circ T$ to an appropriate accuracy. Concatenating these networks with the 1-layer NN (A_1, b_1), where $A_1 x + b_1 = T^{-1}x$ yields the result. The explicit dependence of $C > 0$ on $(b-a)$ can be deduced from the error bounds in $(0,1)$ by affine transformation.*

4 Exponential approximation rates by realizations of NNs

We now establish several technical results on the *exponentially consistent* approximation by realizations of NNs with ReLU activation of univariate and multivariate tensorized polynomials. These results will be used to establish Theorem 4.3, which yields exponential approximation rates of NNs for functions in the weighted, analytic classes introduced in Section 2.2. They are of independent interest, as they imply that spectral and pseudospectral methods can, in principle, be emulated by realizations of NNs with ReLU activation.

4.1 NN-based approximation of univariate, piecewise polynomial functions

We start with the following corollary to Proposition 3.7. It quantifies stability and consistency of realizations of NNs with ReLU activation for the emulation of the univariate, piecewise polynomial basis functions in Theorem 2.1.

Corollary 4.1. *Let $I = (a, b) \subset \mathbb{R}$ be a bounded interval. Fix $C_p > 0$, $C_v > 0$, and $b_v > 0$. Let $0 < \varepsilon_{\mathsf{hp}} < 1$ and $p, N_{1\mathrm{d}}, N_{\mathrm{int}} \in \mathbb{N}$ be such that $p \le C_p(1 + |\log \varepsilon_{\mathsf{hp}}|)$ and let $\mathcal{G}_{1\mathrm{d}}$ be a partition of I into N_{int} open, disjoint, connected subintervals and, for $i \in \{1, \dots, N_{1\mathrm{d}}\}$, let $v_i \in \mathbb{Q}_p(\mathcal{G}_{1\mathrm{d}}) \cap H^1(I)$ be such that $\mathrm{supp}(v_i)$ intersects either a single interval or two adjacent intervals in $\mathcal{G}_{1\mathrm{d}}$ and $\|v_i\|_{H^1(I)} \le C_v \varepsilon_{\mathsf{hp}}^{-b_v}$, for all $i \in \{1, \dots, N_{1\mathrm{d}}\}$.*

Then, for every $0 < \varepsilon_1 \le \varepsilon_{\mathsf{hp}}$, and for every $i \in \{1, \dots, N_{1\mathrm{d}}\}$, there exists a NN $\Phi^{v_i}_{\varepsilon_1}$ such that

$$\|v_i - \mathrm{R}\,(\Phi^{v_i}_{\varepsilon_1})\|_{H^1(I)} \le \varepsilon_1 |v_i|_{H^1(I)}, \tag{4.1}$$

$$\mathrm{L}\,(\Phi^{v_i}_{\varepsilon_1}) \le C_4(1 + |\log(\varepsilon_1)|)(1 + \log(1 + |\log(\varepsilon_1)|)), \tag{4.2}$$

$$\mathrm{M}\,(\Phi^{v_i}_{\varepsilon_1}) \le C_5(1 + |\log(\varepsilon_1)|^2), \tag{4.3}$$

for constants $C_4, C_5 > 0$ depending on $C_p > 0$, $C_v > 0$, $b_v > 0$ and $(b - a)$ only. In addition, $\mathrm{R}\,(\Phi^{v_i}_{\varepsilon_1})\,(x_j) = v_i(x_j)$ for all $i \in \{1, \dots, N_{1\mathrm{d}}\}$ and $j \in \{0, \dots, N_{\mathrm{int}}\}$, where $\{x_j\}_{j=0}^{N_{\mathrm{int}}}$ are the nodes of $\mathcal{G}_{1\mathrm{d}}$.

Proof. Let $i = 1, \dots, N_{1\mathrm{d}}$. For v_i as in the assumption of the corollary, we have that either $\mathrm{supp}(v_i) = \overline{J}$ for a unique $J \in \mathcal{G}_{1\mathrm{d}}$ or $\mathrm{supp}(v_i) = \overline{J \cup J'}$ for two neighboring intervals $J, J' \in \mathcal{G}_{1\mathrm{d}}$. Hence, there exists a partition $\mathcal{T}_i$ of I of at most four subintervals so that $v_i \in S_{\boldsymbol{p}}(I, \mathcal{T}_i)$, where $\boldsymbol{p} = (p)_{i \in \{1,\dots,4\}}$.

Because of this, an application of Proposition 3.7 with $q' = 2$ and Remark 3.8 yields that for every $0 < \varepsilon_1 \le \varepsilon_{\mathsf{hp}} < 1$ there exists a NN $\Phi^{v_i}_{\varepsilon_1} := \Phi^{v_i, \mathcal{T}_i, \boldsymbol{p}}_{\varepsilon_1}$ such that (4.1) holds. In addition, by invoking $p \lesssim 1 + |\log(\varepsilon_{\mathsf{hp}})| \le 1 + |\log(\varepsilon_1)|$, we observe that

$$\mathrm{L}\,(\Phi^{v_i}_{\varepsilon_1}) \le\ C(1 + \log(p))\,(p + |\log(\varepsilon_1)|) \lesssim 1 + |\log(\varepsilon_1)|\,(1 + \log(1 + |\log(\varepsilon_1)|)).$$

Therefore, there exists $C_4 > 0$ such that (4.2) holds. Furthermore,

$$\begin{aligned}\mathrm{M}\,(\Phi^{v_i}_{\varepsilon_1}) &\le 4C(1 + \log(p))\,(p + |\log(\varepsilon_1)|) + C \sum_{i=1}^{4} p(p + |\log(\varepsilon_1)|) \\ &\lesssim p^2 + |\log(\varepsilon_1)|\, p + (1 + \log(p))\,(p + |\log(\varepsilon_1)|)\,.\end{aligned}$$

We use $p \lesssim 1 + |\log(\varepsilon_1)|$ and obtain that there exists $C_5 > 0$ such that (4.3) holds. □

4.2 Emulation of functions with singularities in cubic domains by NNs

Below we state a result describing the efficiency of re-approximating continuous, piecewise tensor product polynomial functions in a cubic domain, as introduced in Theorem 2.1, by realizations of NNs with the ReLU activation function.

Theorem 4.2. *Let $d \in \{2, 3\}$, let $I = (a, b) \subset \mathbb{R}$ be a bounded interval, and let $Q = I^d$. Suppose that there exist constants $C_p > 0$, $C_{N_{1\mathrm{d}}} > 0$, $C_v > 0$, $C_c > 0$, $b_v > 0$, and, for $0 < \varepsilon \le 1$, assume there exist $p, N_{1\mathrm{d}}, N_{\mathrm{int}} \in \mathbb{N}$, and $c \in \mathbb{R}^{N_{1\mathrm{d}} \times \dots N_{1\mathrm{d}}}$, such that*

$$N_{1\mathrm{d}} \le C_{N_{1\mathrm{d}}}(1 + |\log \varepsilon|^2), \quad \|c\|_1 \le C_c(1 + |\log \varepsilon|^{2d}), \quad p \le C_p(1 + |\log \varepsilon|).$$

Further, let $\mathcal{G}_{1\mathrm{d}}$ be a partition of I into N_{int} open, disjoint, connected subintervals and let, for all $i \in \{1, \dots, N_{1\mathrm{d}}\}$, $v_i \in \mathbb{Q}_p(\mathcal{G}_{1\mathrm{d}}) \cap H^1(I)$ be such that $\mathrm{supp}(v_i)$ intersects either a single interval or two neighboring subintervals of $\mathcal{G}_{1\mathrm{d}}$ and

$$\|v_i\|_{H^1(I)} \le C_v \varepsilon^{-b_v}, \qquad \|v_i\|_{L^\infty(I)} \le 1, \qquad \forall i \in \{1, \dots, N_{1\mathrm{d}}\}.$$

Then, there exists a NN $\Phi_{\varepsilon,c}$ such that

$$\left\| \sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} c_{i_1\dots i_d} \bigotimes_{j=1}^{d} v_{i_j} - \mathrm{R}\left(\Phi_{\varepsilon,c}\right) \right\|_{H^1(Q)} \leq \varepsilon. \tag{4.4}$$

Furthermore, there holds

$$\left\|\mathrm{R}\left(\Phi_{\varepsilon,c}\right)\right\|_{L^\infty(Q)} \leq (2^d+1)C_c(1+|\log\varepsilon|^{2d}),\ \mathrm{M}(\Phi_{\varepsilon,c}) \leq C(1+|\log\varepsilon|^{2d+1}), \mathrm{L}(\Phi_{\varepsilon,c}) \leq C(1+|\log\varepsilon|\log(|\log\varepsilon|)),$$

where $C > 0$ depends on C_p, $C_{N_{1\mathrm{d}}}$, C_v, C_c, b_v, d, and $(b-a)$ only.

Proof. Assume $I \neq \varnothing$ as otherwise there is nothing to show. Let $C_I \geq 1$ be such that $C_I^{-1} \leq (b-a) \leq C_I$. Let $c_{v,\max} := \max\{\|v_i\|_{H^1(I)} \colon i \in \{1,\dots,N_{1\mathrm{d}}\}\} \leq C_v\varepsilon^{-b_v}$, let $\varepsilon_1 := \min\{\varepsilon/(2\cdot d\cdot(c_{v,\max}+1)^d \cdot \|c\|_1), 1/2, C_I^{-1/2}C_v^{-1}\varepsilon^{b_v}\}$, and let $\varepsilon_2 := \min\{\varepsilon/(2\cdot(\sqrt{d}+1)\cdot(c_{v,\max}+1)\cdot\|c\|_1), 1/2\}$.

Construction of the neural network. Invoking Corollary 4.1 we choose, for $i = 1,\dots,N_{1\mathrm{d}}$, NNs $\Phi^{v_i}_{\varepsilon_1}$ so that

$$\|\mathrm{R}(\Phi^{v_i}_{\varepsilon_1}) - v_i\|_{H^1(I)} \leq C_v\varepsilon_1\varepsilon^{-b_v} \leq 1.$$

It follows that for all $i \in \{1,\dots,N_{1\mathrm{d}}\}$

$$\|\mathrm{R}\left(\Phi^{v_i}_{\varepsilon_1}\right)\|_{H^1(I)} \leq \|\mathrm{R}\left(\Phi^{v_i}_{\varepsilon_1}\right) - v_i\|_{H^1(I)} + \|v_i\|_{H^1(I)} \leq 1 + c_{v,\max} \tag{4.5}$$

and that, by Sobolev imbedding,

$$\begin{aligned}\|\mathrm{R}\left(\Phi^{v_i}_{\varepsilon_1}\right)\|_\infty &\leq \|\mathrm{R}\left(\Phi^{v_i}_{\varepsilon_1}\right) - v_i\|_\infty + \|v_i\|_\infty \leq C_I^{1/2}\|\mathrm{R}\left(\Phi^{v_i}_{\varepsilon_1}\right) - v_i\|_{H^1(I)} + 1 \\ &\leq C_I^{1/2}C_v\varepsilon_1\varepsilon^{-b_v} + 1 \leq 2.\end{aligned} \tag{4.6}$$

Then, let Φ_{basis} be the NN defined as

$$\Phi_{\mathrm{basis}} := \mathrm{FP}\left(\mathrm{P}(\Phi^{v_1}_{\varepsilon_1},\dots,\Phi^{v_{N_{1\mathrm{d}}}}_{\varepsilon_1}),\dots,\mathrm{P}(\Phi^{v_1}_{\varepsilon_1},\dots,\Phi^{v_{N_{1\mathrm{d}}}}_{\varepsilon_1})\right), \tag{4.7}$$

where the full parallelization is of d copies of $\mathrm{P}(\Phi^{v_1}_{\varepsilon_1},\dots,\Phi^{v_{N_{1\mathrm{d}}}}_{\varepsilon_1})$. Note that Φ_{basis} is a NN with d-dimensional input and $dN_{1\mathrm{d}}$-dimensional output. Subsequently, we introduce the $N_{1\mathrm{d}}^d$ matrices $E^{(i_1,\dots,i_d)} \in \{0,1\}^{d\times dN_{1\mathrm{d}}}$ such that, for all $(i_1,\dots,i_d) \in \{1,\dots,N_{1\mathrm{d}}\}^d$,

$$E^{(i_1,\dots,i_d)}a = \{a_{(j-1)N_{1\mathrm{d}}+i_j} : j = 1,\dots,d\} \qquad \text{for all } a = (a_1,\dots,a_{dN_{1\mathrm{d}}}) \in \mathbb{R}^{dN_{1\mathrm{d}}}.$$

Note that, for all $(i_1,\dots,i_d) \in \{1,\dots,N_{1\mathrm{d}}\}^d$,

$$\mathrm{R}((E^{(i_1,\dots,i_d)},0)\odot\Phi_{\mathrm{basis}}) : (x_1,\dots,x_d) \mapsto \left\{\mathrm{R}(\Phi^{v_{i_j}}_{\varepsilon_1})(x_j) : j = 1,\dots,d\right\}.$$

Then, we set

$$\Phi_\varepsilon := \mathrm{P}\left(\Pi^d_{\varepsilon_2,2}\odot(E^{(i_1,\dots,i_d)},0) : (i_1,\dots,i_d) \in \{1,\dots,N_{1\mathrm{d}}\}^d\right)\odot\Phi_{\mathrm{basis}}, \tag{4.8}$$

where $\Pi^d_{\varepsilon_2,2}$ is according to Proposition 3.6. Note that, by (4.6), the inputs of $\Pi^d_{\varepsilon_2,2}$ are bounded in absolute value by 2. Finally, we define

$$\Phi_{\varepsilon,c} := ((\mathrm{vec}(c)^\top, 0))\odot\Phi_\varepsilon,$$

where $\mathrm{vec}(c) \in \mathbb{R}^{N_{1\mathrm{d}}^d}$ is the reshaping such that, for all $(i_1,\dots,i_d) \in \{1,\dots,N_{1\mathrm{d}}\}^d$

$$(\mathrm{vec}(c))_i = c_{i_1\dots i_d}, \qquad \text{with } i = 1 + \sum_{j=1}^{d}(i_j - 1)N_{1\mathrm{d}}^{j-1}. \tag{4.9}$$

See Figure 1 for a schematic representation of the NN $\Phi_{\varepsilon,c}$.

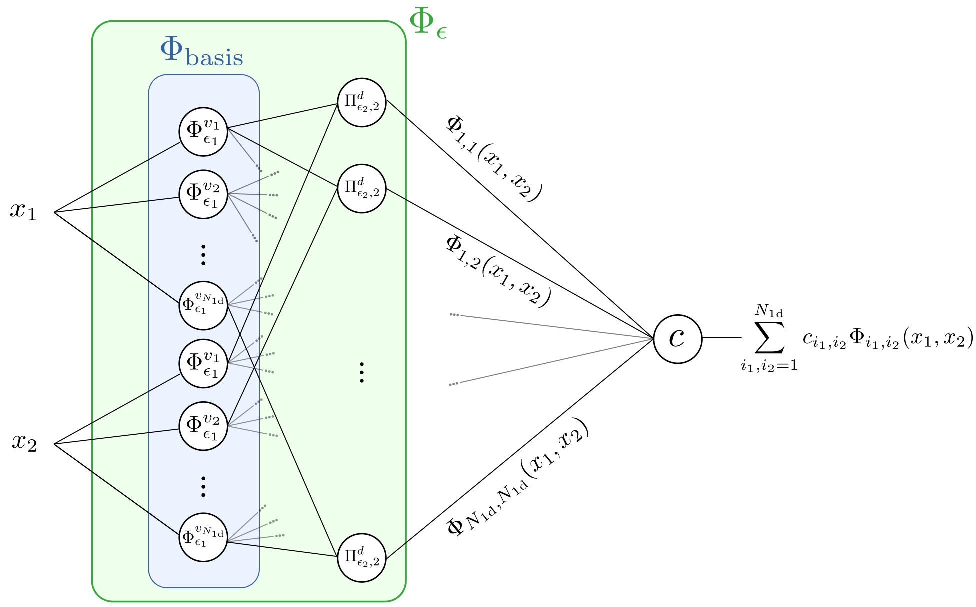

Figure 1: Schematic representation of the neural network $\Phi_{\varepsilon,c}$, for the case $d=2$ constructed in the proof of Theorem 4.2. The circles represent subnetworks (i.e., the neural networks $\Phi^{v_i}_{\varepsilon_1}$, $\Pi^d_{\varepsilon_2,2}$, and $((\text{vec}(c)^\top, 0))$). Along some branches, we indicate $\Phi_{i,k}(x_1,x_2) = \mathrm{R}\left(\Pi^2_{\varepsilon_2,2} \odot ((E^{(i,k)},0)) \odot \Phi_{\text{basis}}\right)(x_1,x_2)$.

Approximation accuracy. Let us now analyze if $\Phi_{\varepsilon,c}$ has the asserted approximation accuracy. Define, for all $(i_1,\dots,i_d) \in \{1,\dots,N_{1\text{d}}\}^d$

$$\phi_{i_1\dots i_d} = \bigotimes_{j=1}^d v_{i_j},$$

Furthermore, for each $(i_1,\dots,i_d) \in \{1,\dots,N_{1\text{d}}\}^d$, let $\Phi_{i_1\dots i_d}$ denote the NNs

$$\Phi_{i_1\dots i_d} = \Pi^d_{\varepsilon_2,2} \odot ((E^{(i_1,\dots,i_d)},0)) \odot \Phi_{\text{basis}}.$$

We estimate by the triangle inequality that

$$\begin{aligned}\left\| \sum_{i_1,\dots,i_d=1}^{N_{1\text{d}}} c_{i_1\dots i_d}\phi_{i_1\dots i_d} - \mathrm{R}(\Phi_{\varepsilon,c}) \right\|_{H^1(Q)} &= \left\| \sum_{i_1,\dots,i_d=1}^{N_{1\text{d}}} c_{i_1\dots i_d}\phi_{i_1\dots i_d} - \sum_{i_1,\dots,i_d=1}^{N_{1\text{d}}} c_{i_1\dots i_d}\,\mathrm{R}(\Phi_{i_1\dots i_d}) \right\|_{H^1(Q)} \\ &\le \sum_{i_1,\dots,i_d=1}^{N_{1\text{d}}} |c_{i_1\dots i_d}|\, \|\phi_{i_1\dots i_d} - \mathrm{R}(\Phi_{i_1\dots i_d})\|_{H^1(Q)}\,. \end{aligned} \tag{4.10}$$

We have that

$$\|\phi_{i_1\dots i_d} - \mathrm{R}(\Phi_{i_1\dots i_d})\|_{H^1(Q)} = \left\| \bigotimes_{j=1}^d v_{i_j} - \mathrm{R}\left(\Pi^d_{\varepsilon_2,2}\right) \circ \left[\mathrm{R}\left(\Phi^{v_{i_1}}_{\varepsilon_1}\right),\dots,\mathrm{R}\left(\Phi^{v_{i_d}}_{\varepsilon_1}\right)\right] \right\|_{H^1(Q)}$$

and, by another application of the triangle inequality, we have that

$$\begin{aligned}\|\phi_{i_1\dots i_d} - \mathrm{R}\left(\Phi_{i_1\dots i_d}\right)\|_{H^1(Q)} &\le \left\| \bigotimes_{j=1}^d v_{i_j} - \bigotimes_{j=1}^d \mathrm{R}\left(\Phi^{v_{i_j}}_{\varepsilon_1}\right) \right\|_{H^1(Q)} \\ &\quad + \left\| \bigotimes_{j=1}^d \mathrm{R}\left(\Phi^{v_{i_j}}_{\varepsilon_1}\right) - \mathrm{R}\left(\Pi^d_{\varepsilon_2,2}\right) \circ \left[\mathrm{R}\left(\Phi^{v_{i_1}}_{\varepsilon_1}\right),\dots,\mathrm{R}\left(\Phi^{v_{i_d}}_{\varepsilon_1}\right)\right] \right\|_{H^1(Q)} \\ &\le \left\| \bigotimes_{j=1}^d v_{i_j} - \bigotimes_{j=1}^d \mathrm{R}\left(\Phi^{v_{i_j}}_{\varepsilon_1}\right) \right\|_{H^1(Q)} + (\sqrt{d}+1)\varepsilon_2(c_{v,\max}+1), \end{aligned} \tag{4.11}$$

where the last estimate follows from Proposition 3.6 and the chain rule:

$$\left\| \bigotimes_{j=1}^{d} \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_j}}\right) - \mathrm{R}\left(\Pi_{\varepsilon_2,2}^{d}\right) \circ \left[\mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_1}}\right), \dots, \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_d}}\right)\right] \right\|_{L^2(Q)} \leq \varepsilon_2$$

and

$$\begin{aligned}
&\left| \bigotimes_{j=1}^{d} \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_j}}\right) - \mathrm{R}\left(\Pi_{\varepsilon_2,2}^{d}\right) \circ \left[\mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_1}}\right), \dots, \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_d}}\right)\right] \right|_{H^1(Q)}^2 \\
&\quad = \sum_{k=1}^{d} \left\| \frac{\partial}{\partial x_k} \bigotimes_{j=1}^{d} \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_j}}\right) - \frac{\partial}{\partial x_k} \mathrm{R}\left(\Pi_{\varepsilon_2,2}^{d}\right) \circ \left[\mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_1}}\right), \dots, \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_d}}\right)\right] \right\|_{L^2(Q)}^2 \\
&\quad = \sum_{k=1}^{d} \left\| \left(\bigotimes_{\substack{j=1 \\ j\neq k}}^{d} \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_j}}\right) - \left(\frac{\partial}{\partial x_k} \mathrm{R}\left(\Pi_{\varepsilon_2,2}^{d}\right)\right) \circ \left[\mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_1}}\right), \dots, \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_d}}\right)\right] \right) \left(\frac{\partial}{\partial x} \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_k}}\right)\right) \right\|_{L^2(Q)}^2 \\
&\quad \leq \sum_{k=1}^{d} \varepsilon_2^2 \left\| \frac{\partial}{\partial x} \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_k}}\right) \right\|_{L^2(I)}^2 \leq d\varepsilon_2^2 (c_{v,\max}+1)^2,
\end{aligned}$$

where we used (4.5). We now use (4.6) to bound the first term in (4.11): for $d = 3$, we have that, for all $(i_1, \dots, i_d) \in \{1, \dots, N_{1\mathrm{d}}\}^d$,

$$\begin{aligned}
\left\| \bigotimes_{j=1}^{d} v_{i_j} - \bigotimes_{j=1}^{d} \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_j}}\right) \right\|_{H^1(Q)} &\leq \left\| (v_{i_1} - \mathrm{R}(\Phi_{\varepsilon_1}^{v_{i_1}})) \otimes \bigotimes_{j=2}^{d} v_{i_j} \right\|_{H^1(Q)} \\
&\quad + \left\| \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_j}}\right) \otimes \left(v_{i_2} - \mathrm{R}\left(\Phi_{\varepsilon_1}^{v_{i_2}}\right)\right) \otimes v_{i_d} \right\|_{H^1(Q)} \\
&\quad + \left\| \bigotimes_{j=1}^{d-1} \mathrm{R}(\Phi_{\varepsilon_1}^{v_{i_j}}) \otimes (v_{i_d} - \mathrm{R}(\Phi_{\varepsilon_1}^{v_{i_d}})) \right\|_{H^1(Q)} =: (\mathrm{I}).
\end{aligned}$$

For $d = 2$, we end up with a similar estimate with only two terms. By the tensor product structure, it is clear that $(\mathrm{I}) \leq d\varepsilon_1(c_{v,\max}+1)^d$. We have from (4.10) and the considerations above that

$$\left\| \sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} c_{i_1\dots i_d} \phi_{i_1\dots i_d} - \mathrm{R}\left(\Phi_{\varepsilon,c}\right) \right\|_{H^1(Q)} \leq \|c\|_1 \cdot \left(d \cdot \varepsilon_1 \cdot (c_{v,\max}+1)^d + (\sqrt{d}+1) \cdot \varepsilon_2 \cdot (c_{v,\max}+1)\right) \leq \varepsilon.$$

This yields (4.12).

Bound on the L^∞ norm of the neural network. As we have already shown, $\|\mathrm{R}\left(\Phi_{\varepsilon_1}^{v_i}\right)\|_\infty \leq 2$. Therefore, by Proposition 3.6, $\|\mathrm{R}\left(\Phi_\varepsilon\right)\|_\infty \leq 2^d + \varepsilon_2$. It follows that

$$\|\mathrm{R}\left(\Phi_{\varepsilon,c}\right)\|_\infty \leq \|c\|_1 \left(2^d + \varepsilon_2\right) \leq (2^d+1)C_c(1+|\log\varepsilon|^{2d}).$$

Size of the neural network. Bounds on the size and depth of $\Phi_{\varepsilon,c}$ follow from Proposition 3.6 and Corollary 4.1. Specifically, we start by remarking that there exists a constant $C_1 > 0$ depending on C_v, b_v, C_I and d only, such that $|\log(\varepsilon_1)| \leq C_1(1+|\log\varepsilon|)$. Then, by Corollary 4.1, there exist constants C_4, $C_5 > 0$ depending on $C_p, C_v, b_v, (b-a)$, and d only such that for all $i = 1, \dots, N_{1\mathrm{d}}$,

$$\mathrm{L}\left(\Phi_{\varepsilon_1}^{v_i}\right) \leq C_4(1+|\log\varepsilon|)(1+\log(1+|\log\varepsilon|)) \text{ and } \mathrm{M}\left(\Phi_{\varepsilon_1}^{v_i}\right) \leq C_5(1+|\log\varepsilon|^2).$$

Hence, by Propositions 3.5 and 3.3, there exist C_6, $C_7 > 0$ depending on $C_p, C_v, b_v, (b-a)$, and d only such that

$$\mathrm{L}(\Phi_{\mathrm{basis}}) \leq C_6(1+|\log\varepsilon|)(1+\log(1+|\log\varepsilon|)) \text{ and } \mathrm{M}(\Phi_{\mathrm{basis}}) \leq C_7 d N_{1\mathrm{d}}(1+|\log\varepsilon|^2).$$

Then, remarking that for all $(i_1, \dots, i_d) \in \{1, \dots, N_{1\mathrm{d}}\}^d$ there holds $\|E^{(i_1,\dots,i_d)}\|_0 = d$ and, by Propositions 3.2, 3.6, and 3.3, we have

$$\mathrm{L}(\Phi_\varepsilon) \leq C_8(1+|\log\varepsilon|)(1+\log(1+|\log\varepsilon|)), \quad \mathrm{M}(\Phi_\varepsilon) \leq C_9\left(N_{1\mathrm{d}}^d(1+|\log\varepsilon|) + \mathrm{M}(\Phi_{\mathrm{basis}})\right).$$

For $C_8, C_9 > 0$ depending on $C_p, C_v, b_v, (b-a), d$ and C_c only. Finally, we conclude that there exists a constant $C_{10} > 0$ depending on $C_p, C_v, b_v, (b-a), d$ and C_c only such that

$$\mathrm{L}(\Phi_{\varepsilon,c}) \leq C_{10}(1+|\log\varepsilon|)(1+\log(1+|\log\varepsilon|)).$$

Using also the fact that $N_{1\mathrm{d}} \leq C(1+|\log\varepsilon|^2)$ for $C>0$ independent of ε and since $d \geq 2$,

$$\mathrm{M}(\Phi_{\varepsilon,c}) \leq C_{11}(1+|\log\varepsilon|)^{2d+1},$$

for a constant $C_{11} > 0$ depending on $C_p, C_v, b_v, (b-a), d$ and C_c only. □

Next, we state our main approximation result, which describes the approximation of singular functions in $(0,1)^d$ by realizations of NNs.

Theorem 4.3. *Let $d \in \{2,3\}$ and $Q := (0,1)^d$. Let $\mathcal{C} = \{c\}$ where c is one of the corners of Q and let $\mathcal{E} = \mathcal{E}_c$ contain the edges adjacent to c when $d=3$, $\mathcal{E} = \varnothing$ when $d=2$. Assume furthermore that $C_f, A_f > 0$, and*

$$\underline{\gamma} = \{\gamma_c : c \in \mathcal{C}\}, \qquad \text{with } \gamma_c > 1, \text{ for all } c \in \mathcal{C} \qquad \text{if } d=2,$$
$$\underline{\gamma} = \{\gamma_c, \gamma_e : c \in \mathcal{C}, e \in \mathcal{E}\}, \quad \text{with } \gamma_c > 3/2 \text{ and } \gamma_e > 1, \text{ for all } c \in \mathcal{C} \text{ and } e \in \mathcal{E} \qquad \text{if } d=3.$$

Then, for every $f \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E};C_f,A_f)$ and every $0<\varepsilon<1$, there exists a NN $\Phi_{\varepsilon,f}$ so that

$$\|f - \mathrm{R}(\Phi_{\varepsilon,f})\|_{H^1(Q)} \leq \varepsilon. \tag{4.12}$$

In addition, $\|\mathrm{R}(\Phi_{\varepsilon,f})\|_{L^\infty(Q)} = \mathcal{O}(|\log\varepsilon|^{2d})$ for $\varepsilon \to 0$. Also, $\mathrm{M}(\Phi_{\varepsilon,f}) = \mathcal{O}(|\log\varepsilon|^{2d+1})$ and $\mathrm{L}(\Phi_{\varepsilon,f}) = \mathcal{O}(|\log\varepsilon|\log(|\log\varepsilon|))$, for $\varepsilon \to 0$.

Proof. Denote $I := (0,1)$ and let $f \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E};C_f,A_f)$ and $0<\varepsilon<1$. Then, by Theorem 2.1 (applied with $\varepsilon/2$ instead of ε) there exists $N_{1\mathrm{d}} \in \mathbb{N}$ so that $N_{1\mathrm{d}} = \mathcal{O}((1+|\log\varepsilon|)^2)$, $c \in \mathbb{R}^{N_{1\mathrm{d}}\times\cdots\times N_{1\mathrm{d}}}$ with $\|c\|_1 \leq C_c(1+|\log\varepsilon|^{2d})$, and, for all $(i_1,\dots,i_d) \in \{1,\dots,N_{1\mathrm{d}}\}^d$,

$$\phi_{i_1\dots i_d} = \bigotimes_{j=1}^{d} v_{i_j},$$

such that the hypotheses of Theorem 4.2 are met, and

$$\left\| f - \sum_{i_1,\dots i_d=1}^{N_{1\mathrm{d}}} c_{i_1\dots i_d}\phi_{i_1\dots i_d}\right\|_{H^1(Q)} \leq \frac{\varepsilon}{2}.$$

We have, by Theorem 2.1 and the triangle inequality, that for $\Phi_{\varepsilon,f} := \Phi_{\varepsilon/2,c}$

$$\|f - \mathrm{R}(\Phi_{\varepsilon,f})\|_{H^1(Q)} \leq \frac{\varepsilon}{2} + \left\| \sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} c_{i_1\dots i_d}\phi_{i_1\dots i_d} - \mathrm{R}(\Phi_{\varepsilon/2,c})\right\|_{H^1(Q)}.$$

Then, the application of Theorem 4.2 (with $\varepsilon/2$ instead of ε) concludes the proof of (4.12). Finally, the bounds on $\mathrm{L}(\Phi_{\varepsilon,f}) = \mathrm{L}(\Phi_{\varepsilon/2,c})$, $\mathrm{M}(\Phi_{\varepsilon,f}) = \mathrm{M}(\Phi_{\varepsilon/2,c})$, and on $\|\mathrm{R}(\Phi_{\varepsilon,f})\|_{L^\infty(Q)} = \|\mathrm{R}(\Phi_{\varepsilon/2,c})\|_{L^\infty(Q)}$ follow from the corresponding estimates of Theorem 4.2. □

Theorem 4.3 admits a straightforward generalization to functions with multivariate output, so that each coordinate is a weighted analytic function with the same regularity. Here, we denote for a NN Φ with N-dimensional output, $N \in \mathbb{N}$, by $\mathrm{R}(\Phi)_n$ the n-th component of the output (where $n \in \{1,\dots,N\}$).

Corollary 4.4. *Let $d \in \{2,3\}$ and $Q := (0,1)^d$. Let $\mathcal{C} = \{c\}$ where c is one of the corners of Q and let $\mathcal{E} = \mathcal{E}_c$ contain the edges adjacent to c when $d=3$; $\mathcal{E} = \varnothing$ when $d=2$. Let $N_f \in \mathbb{N}$. Further assume that $C_f, A_f > 0$, and*

$$\underline{\gamma} = \{\gamma_c : c \in \mathcal{C}\}, \qquad \text{with } \gamma_c > 1, \text{ for all } c \in \mathcal{C} \qquad \text{if } d=2,$$
$$\underline{\gamma} = \{\gamma_c, \gamma_e : c \in \mathcal{C}, e \in \mathcal{E}\}, \quad \text{with } \gamma_c > 3/2 \text{ and } \gamma_e > 1, \text{ for all } c \in \mathcal{C} \text{ and } e \in \mathcal{E} \qquad \text{if } d=3.$$

Then, for all $\boldsymbol{f} = (f_1,\dots,f_{N_f}) \in \left[\mathcal{J}^{\varpi}_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E};C_f,A_f)\right]^{N_f}$ and every $0<\varepsilon<1$, there exists a NN $\Phi_{\varepsilon,\boldsymbol{f}}$ with d-dimensional input and N_f-dimensional output such that, for all $n = 1,\dots,N_f$,

$$\left\|f_n - \mathrm{R}(\Phi_{\varepsilon,\boldsymbol{f}})_n\right\|_{H^1(Q)} \leq \varepsilon. \tag{4.13}$$

In addition, $\|\mathrm{R}(\Phi_{\varepsilon,\boldsymbol{f}})_n\|_{L^\infty(Q)} = \mathcal{O}(|\log\varepsilon|^{2d})$ for every $n = \{1,\dots,N_f\}$, $\mathrm{M}(\Phi_{\varepsilon,f}) = \mathcal{O}(|\log\varepsilon|^{2d+1} + N_f|\log\varepsilon|^{2d})$ and $\mathrm{L}(\Phi_{\varepsilon,f}) = \mathcal{O}(|\log\varepsilon|\log(|\log\varepsilon|))$, for $\varepsilon \to 0$.

Proof. Let Φ_ε be as in (4.8) and let $c^{(n)} \in \mathbb{R}^{N_{1\mathrm{d}}\times\cdots\times N_{1\mathrm{d}}}$, $n = 1, \dots, N_f$ be the matrices of coefficients such that, in the notation of the proof of Theorems 4.2 and 4.3, for all $n \in \{1, \dots, N_f\}$,

$$\left\| f_n - \sum_{i_1,\dots i_d=1}^{N_{1\mathrm{d}}} c^{(n)}_{i_1\dots i_d}\phi_{i_1\dots i_d} \right\|_{H^1(Q)} \leq \frac{\varepsilon}{2}.$$

We define, for vec as defined in (4.9), the NN $\Phi_{\varepsilon,\boldsymbol{f}}$ as

$$\Phi_{\varepsilon,\boldsymbol{f}} \coloneqq \mathrm{P}\left(((\mathrm{vec}(c^{(1)})^\top, 0)), \dots, ((\mathrm{vec}(c^{(N_f)})^\top, 0)) \right) \odot \Phi_\varepsilon.$$

The estimate (4.13) and the L^∞-bound then follow from Theorem 4.2. The bound on $\mathrm{L}(\Phi_{\varepsilon,\boldsymbol{f}})$ follows directly from Theorem 4.2 and Proposition 3.2. Finally, the bound on $\mathrm{M}(\Phi_{\varepsilon,\boldsymbol{f}})$ follows by Theorem 4.2 and Proposition 3.2, as well as, from the observation that

$$\mathrm{M}\left(\mathrm{P}\left(((\mathrm{vec}(c^{(1)})^\top, 0)), \dots, ((\mathrm{vec}(c^{(N_f)})^\top, 0)) \right)\right) \leq N_f N_{1\mathrm{d}}^d \leq C N_f (1 + |\log\varepsilon|^{2d}),$$

for a constant $C > 0$ independent of N_f and ε. □

5 Exponential expression rates for solution classes of PDEs

In this section, we develop Theorem 4.3 into several exponentially decreasing upper bounds for the rates of approximation, by realizations of NNs with ReLU activation, for solution classes to elliptic PDEs with singular data (such as singular coefficients or domains with nonsmooth boundary). In particular, we consider elliptic PDEs in two-dimensional *general* polygonal domains, in three-dimensional domains that are a union of cubes, and elliptic eigenvalue problems with isolated point singularities in the potential which arise in models of electron structure in quantum mechanics.

In each class of examples, the solution sets belong to the class of weighted analytic functions introduced in Subsection 2.2. However, the approximation rates established in Section 4 only hold on tensor product domains with singularities on the boundary. Therefore, we will first extend the exponential NN approximation rates to functions which exhibit singularities on a set of isolated points internal to the domain, arising from singular potentials of nonlinear Schrödinger operators. In Section 5.2, we demonstrate, using an argument based on a partition of unity, that the approximation problem on general polygonal domains can be reduced to that on tensor product domains and Fichera-type domains, and establish exponential NN expression rates for linear elliptic source and eigenvalue problems. In Section 5.3, we show exponential NN expression rates for classes of weighted analytic functions on two- and three-dimensional Fichera-type domains.

5.1 Nonlinear eigenvalue problems with isolated point singularities

Point singularities emerge in the solutions of elliptic eigenvalue problems, as arise, for example, for electrostatic interactions between charged particles that are modelled mathematically as point sources in $\mathbb{R}^3$. Other problems that exhibit point singularities appear in general relativity, and for electron structure models in quantum mechanics. We concentrate here on the expression rate of "ab initio" NN approximation of the electron density near isolated singularities of the nuclear potential. Via a ReLU-based partition of unity argument, an exponential approximation rate bound for a single, isolated point singularity in Theorem 5.1 is extended in Corollary 5.4 to electron densities corresponding to potentials with multiple point singularities at a priori known locations, modeling (static) molecules.

The numerical approximation in ab initio electron structure computations with NNs has been recently reported to be competitive with other established, methodologies (e.g. [37, 18] and the references there). The exponential ReLU expression rate bounds obtained here can, in part, underpin competitive performances of NNs in (static) electron structure computations.

5.1.1 Nonlinear Schrödinger equations

Let $\Omega = \mathbb{R}^d/(2\mathbb{Z})^d$, where $d \in \{2, 3\}$, be a flat torus and let $V : \Omega \to \mathbb{R}$ be a potential such that $V(x) \geq V_0 > 0$ for all $x \in \Omega$ and there exists $\delta > 0$ and $A_V > 0$ such that

$$\|r^{2+|\alpha|-\delta}\partial^\alpha V\|_{L^\infty(\Omega)} \leq A_V^{|\alpha|+1}|\alpha|! \qquad \forall \alpha \in \mathbb{N}_0^d, \tag{5.1}$$

where $r(x) = \mathrm{dist}(x, (0, \ldots, 0))$. For $k \in \{0, 1, 2\}$, we introduce the Schrödinger eigenproblem that consists in finding the smallest eigenvalue $\lambda \in \mathbb{R}$ and an associated eigenfunction $u \in H^1(\Omega)$ such that

$$(-\Delta + V + |u|^k)u = \lambda u \quad \text{in } \Omega, \quad \|u\|_{L^2(\Omega)} = 1. \tag{5.2}$$

There holds the following approximation result.

Theorem 5.1. *Let $k \in \{0, 1, 2\}$ and $(\lambda, u) \in \mathbb{R} \times H^1(\Omega)\backslash\{0\}$ be a solution of the eigenvalue problem* (5.2) *with minimal λ, where V satisfies* (5.1)*.*

Then, for every $0 < \varepsilon \le 1$ there exists a NN $\Phi_{\varepsilon,u}$ such that

$$\|u - \mathrm{R}(\Phi_{\varepsilon,u})\|_{H^1(Q)} \le \varepsilon. \tag{5.3}$$

In addition, as $\varepsilon \to 0$,

$$\mathrm{M}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|^{2d+1}),\ \mathrm{L}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))\ .$$

Proof. Let $\mathcal{C} = \{(0, \ldots, 0)\}$ and $\mathcal{E} = \varnothing$. The regularity of u is a consequence of [27, Theorem 2] (see also [28, Corollary 3.2] for the linear case $k = 0$): there exists $\gamma_c > d/2$ and $C_u, A_u > 0$ such that $u \in \mathcal{J}^{\varpi}_{\gamma_c}(\Omega; \mathcal{C}, \mathcal{E}; C_u, A_u)$. Here, γ_c and the constants C_u and A_u depend only on, V_0, A_V and δ in (5.1), and on k in (5.2).

Then, for all $0 < \varepsilon \le 1$, by Theorem 4.2 and Proposition A.25, there exists a NN $\Phi_{\varepsilon,u}$ such that (5.3) holds. Furthermore, there exist constants C_1, $C_2 > 0$ dependent only on V_0, A_V, δ, and k, such that

$$\mathrm{M}(\Phi_{\varepsilon,u}) \le C_1(1 + |\log(\varepsilon)|^{2d+1}) \text{ and } \mathrm{L}(\Phi_{\varepsilon,u}) \le C_2\big(1 + |\log(\varepsilon)|\big)\big(1 + \log(1 + |\log(\varepsilon)|)\big).$$

□

5.1.2 Hartree-Fock model

The Hartree-Fock model is an approximation of the full many-body representation of a quantum system under the Born-Oppenheimer approximation, where the many-body wave function is replaced by a sum of Slater determinants. Under this hypothesis, for $M, N \in \mathbb{N}$, the Hartree-Fock energy of a system with N electrons and M nuclei with positive charges Z_i at isolated locations $R_i \in \mathbb{R}^3$, reads

$$E_{\mathrm{HF}} = \inf\Bigg\{ \sum_{i=1}^N \int_{\mathbb{R}^3} \big(|\nabla\varphi_i|^2 + V|\varphi_i|^2\big) + \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho(x)\rho(y)}{|x-y|}dxdy - \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\tau(x,y)^2}{|x-y|}dxdy : (\varphi_1, \ldots, \varphi_N) \in H^1(\mathbb{R}^3)^N \text{ such that } \int_{\mathbb{R}^3} \varphi_i\varphi_j = \delta_{ij} \Bigg\}, \tag{5.4}$$

where δ_{ij} is the Kronecker delta, $V(x) = -\sum_{i=1}^M Z_i/|x - R_i|$, $\tau(x, y) = \sum_{i=1}^N \varphi_i(x)\varphi_i(y)$, and $\rho(x) = \tau(x, x)$, see, e.g., [23, 24]. The Euler-Lagrange equations of (5.4) read

$$(-\Delta + V(x))\varphi_i(x) + \int_{\mathbb{R}^3} \frac{\rho(y)}{|x-y|}dy\varphi_i(x) - \int_{\mathbb{R}^3} \frac{\tau(x,y)}{|x-y|}\varphi_i(y)dy = \lambda_i\varphi_i(x), \qquad i = 1, \ldots, N, \text{ and } x \in \mathbb{R}^3 \tag{5.5}$$

with $\int_{\mathbb{R}^3} \varphi_i\varphi_j = \delta_{ij}$.

Remark 5.2. *It has been shown in* [23] *that, if $\sum_{k=1}^M Z_k > N - 1$, there exists a ground state $\varphi_1, \ldots, \varphi_N$ of* (5.4)*, solution to* (5.5)*.*

The following statement gives exponential expression rate bounds of the NN-based approximation of electronic wave functions in the vicinity of one singularity (corresponding to the location of a nucleus) of the potential.

Theorem 5.3. *Assume that* (5.5) *has N real eigenvalues $\lambda_1, \ldots, \lambda_N$ with associated eigenfunctions $\varphi_1, \ldots, \varphi_N$, such that $\int_{\mathbb{R}^3} \varphi_i\varphi_j = \delta_{ij}$. Fix $k \in \{1, \ldots, M\}$, let R_k be one of the singularities of V and let $a > 0$ such that $|R_j - R_k| > 2a$ for all $j \in \{1, \ldots, M\} \setminus \{k\}$. Let Ω_k be the cube $\Omega_k = \big\{x \in \mathbb{R}^3 : \|x - R_k\|_\infty \le a\big\}$.*

Then there exists a NN $\Phi_{\varepsilon,\varphi}$ such that $\mathrm{R}(\Phi_{\varepsilon,\varphi}) : \mathbb{R}^3 \to \mathbb{R}^N$, satisfies

$$\|\varphi_i - \mathrm{R}(\Phi_{\varepsilon,\varphi})_i\|_{H^1(\Omega_k)} \le \varepsilon, \qquad \forall i \in \{1, \ldots, N\}. \tag{5.6}$$

In addition, as $\varepsilon \to 0$, $\|\mathrm{R}(\Phi_{\varepsilon,\varphi})_i\|_{L^\infty(\Omega_k)} = \mathcal{O}(|\log\varepsilon|^6)$ for every $i = \{1, \ldots, N\}$,

$$\mathrm{M}(\Phi_{\varepsilon,\varphi}) = \mathcal{O}(|\log(\varepsilon)|^7 + N\,|\log(\varepsilon)|^6),\ \ \mathrm{L}(\Phi_{\varepsilon,\varphi}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|)).$$

Proof. Let $\mathcal{C} = \{(0,0,0)\}$ and $\mathcal{E} = \varnothing$ and fix $k \in \{1,\dots,M\}$. From the regularity result in [29, Corollary 1], see also [8, 9], there exist C_φ, A_φ, and $\gamma_c > 3/2$ such that $(\varphi_1,\dots,\varphi_N) \in \left[\mathcal{J}^\varpi_{\gamma_c}(\Omega_k;\mathcal{C},\mathcal{E};C_\varphi,A_\varphi)\right]^N$. Then, (5.6), the L^∞ bound and the depth and size bounds on the NN $\Phi_{\varepsilon,\varphi}$ follow from the hp approximation result in Proposition A.25 (centered in R_k by translation), from Theorem 4.2, as in Corollary 4.4. □

The arguments in the preceding subsections applied to wave functions for a single, isolated nucleus modelled by the singular potential V as in (5.1) can then be extended to give upper bounds on the approximation rates achieved by realizations of NNs of the wave functions in a bounded, sufficiently large domain containing all singularities of the nuclear potential in (5.4).

Corollary 5.4. *Assume that* (5.5) *has N real eigenvalues $\lambda_1,\dots,\lambda_N$ with associated eigenfunctions $\varphi_1,\dots,\varphi_N$, such that $\int_{\mathbb{R}^3}\varphi_i\varphi_j = \delta_{ij}$. Let $a_i, b_i \in \mathbb{R}$, $i = 1,2,3$, and $\Omega = \bigtimes_{i=1}^d (a_i,b_i)$ such that $\{R_j\}_{j=1}^M \subset \Omega$. Then, for every $0<\varepsilon<1$, there exists a NN $\Phi_{\varepsilon,\varphi}$ such that $\mathrm{R}(\Phi_{\varepsilon,\varphi}) : \mathbb{R}^3 \to \mathbb{R}^N$ and*

$$\|\varphi_i - \mathrm{R}(\Phi_{\varepsilon,\varphi})_i\|_{H^1(\Omega)} \le \varepsilon, \qquad \forall i = 1,\dots,N. \tag{5.7}$$

Furthermore, as $\varepsilon \to 0$ $\mathrm{M}(\Phi_{\varepsilon,\varphi}) = \mathcal{O}(|\log(\varepsilon)|^7 + N|\log(\varepsilon)|^6)$ and $\mathrm{L}(\Phi_{\varepsilon,\varphi}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))$.

Proof. The proof is based on a partition of unity argument. We only sketch it at this point, but will develop it in detail in the proof of Theorem 5.6. Let $\mathcal{T}$ be a tetrahedral, regular triangulation of Ω, and let $\{\kappa_k\}_{k=1}^{N_\kappa}$ be the hat-basis functions associated to it. We suppose that the triangulation is sufficiently refined to ensure that, for all $k \in \{1,\dots,N_\kappa\}$, exists a cube $\widetilde{\Omega}_k \subset \Omega$ such that $\mathrm{supp}(\kappa_k) \subset \widetilde{\Omega}_k$ and that there exists at most one $j \in \{1,\dots,M\}$ such that $\overline{\widetilde{\Omega}}_k \cap R_j \neq \varnothing$.

For all $k \in \{1,\dots,N_\kappa\}$, by [17, Theorem 5.2], which is based on [49], there exists a NN Φ^{κ_k} such that

$$\mathrm{R}(\Phi^{\kappa_k})(x) = \kappa_k(x), \qquad \forall x \in \Omega.$$

For all $0<\varepsilon<1$, let

$$\varepsilon_1 \coloneqq \frac{\varepsilon}{2N_\kappa\left(\max_{k\in\{1,\dots,N_\kappa\}}\|\kappa_k\|_{W^{1,\infty}(\Omega)}\right)}.$$

For all $k \in \{1,\dots,N_\kappa\}$ and $i \in \{1,\dots,N\}$, there holds $\varphi_i|_{\widetilde{\Omega}_k} \in \mathcal{J}^\varpi_\gamma(\widetilde{\Omega}_k;\{R_1,\dots,R_M\}\cap\overline{\widetilde{\Omega}}_k,\varnothing)$. Then there exists a NN $\Phi^k_{\varepsilon_1,\varphi}$, as defined in Theorem 5.3, such that

$$\|\varphi_i - \mathrm{R}(\Phi^k_{\varepsilon_1,\varphi})_i\|_{H^1(\widetilde{\Omega}_k)} \le \varepsilon_1, \qquad \forall i \in \{1,\dots,N\}. \tag{5.8}$$

Let

$$C_\infty \coloneqq \max_{k\in\{1,\dots,N_\kappa\}} \sup_{\hat{\varepsilon}\in(0,1)} \frac{\|\mathrm{R}(\Phi^k_{\hat{\varepsilon},\varphi})\|_{L^\infty(\widetilde{\Omega}_k)}}{1+|\log\hat{\varepsilon}|^6} < \infty$$

where the finiteness is due to Theorem 5.3. Then, we denote

$$\varepsilon_\times \coloneqq \frac{\varepsilon}{2N_\kappa(|\Omega|^{1/2}+1+\max_{i=1,\dots,N}|\varphi_i|_{H^1(\Omega)}+\max_{k=1,\dots,N_\kappa}\|\kappa_k\|_{W^{1,\infty}(\Omega)}|\Omega|^{1/2})}$$

and $M_\times(\varepsilon_1) \coloneqq C_\infty(1+|\log\varepsilon_1|^6)$. As detailed in the proof of Theorem 5.6 below, after concatenating with identity NNs and possibly after increasing the constants, we assume that $\mathrm{L}(\Phi^k_{\varepsilon_1,\varphi})$ is independent of k and that the bound on $\mathrm{M}(\Phi^k_{\varepsilon_1,\varphi})$ is independent of k, and that the same holds for Φ^{κ_k}, $k = 1,\dots,N_\kappa$.

Let now, for $i \in \{1,\dots,N\}$, $E_i : \mathbb{R}^{N+1} \to \mathbb{R}^2$ be the matrices such that, for all $x = (x_1,\dots,x_{N+1})$, $E_i x = (x_i, x_{N+1})$. Let also $A \in \mathbb{R}^{N\times N_\kappa}$ be a matrix of ones. Then, we introduce the NN

$$\Phi_{\varepsilon,\varphi} = (A,0)\odot \mathrm{P}\left(\left\{\mathrm{P}\left(\left\{\Pi^2_{\varepsilon_\times,M_\times(\varepsilon_1)}\odot(E_i,0)\right\}_{i=1}^N\right)\odot \mathrm{P}(\Phi^k_{\varepsilon_1,\varphi},\Phi^{\mathrm{Id}}_{1,L}\odot\Phi^{\kappa_k})\right\}_{k=1}^{N_\kappa}\right), \tag{5.9}$$

where $L \in \mathbb{N}$ is such that $\mathrm{L}(\Phi^{\mathrm{Id}}_{1,L}\odot\Phi^{\kappa_k}) = \mathrm{L}(\Phi^k_{\varepsilon_1,\varphi})$, from which it follows that $\mathrm{M}(\Phi^{\mathrm{Id}}_{1,L}) \le C\,\mathrm{L}(\Phi^k_{\varepsilon_1,\varphi})$. There holds, for all $i \in \{1,\dots,N\}$,

$$\mathrm{R}(\Phi_{\varepsilon,\varphi})(x)_i = \sum_{k=1}^{N_\kappa}\mathrm{R}(\Pi^2_{\varepsilon_\times,M_\times(\varepsilon_1)})(\mathrm{R}(\Phi^k_{\varepsilon_1,\varphi})(x)_i,\kappa_k(x)), \qquad \forall x \in \Omega.$$

By the triangle inequality, [33, Theorem 2.1], (5.8), and Proposition 3.6, for all $i \in \{1, \dots, N\}$,

$$
\begin{aligned}
&\|\varphi_i - \mathrm{R}(\Phi_{\varepsilon,\varphi})_i\|_{H^1(\Omega)} \\
&\quad \le \|\varphi_i - \sum_{i=1}^{N_\kappa} \kappa_k \,\mathrm{R}(\Phi^k_{\varepsilon_1,\varphi})_i\|_{H^1(\Omega)} + \sum_{k=1}^{N_\kappa} \|\,\mathrm{R}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)})\left(\mathrm{R}(\Phi^k_{\varepsilon_1,\varphi})_i, \kappa_k\right) - \kappa_k\, \mathrm{R}(\Phi^k_{\varepsilon_1,\varphi})_i\|_{H^1(\Omega_k)} \\
&\quad \le N_\kappa \left(\max_{k\in\{1,\dots,N_\kappa\}} \|\kappa_k\|_{W^{1,\infty}(\Omega)} \right) \varepsilon_1 + N_\kappa(|\Omega|^{1/2} + 1 + \max_{i=1,\dots,N} |\varphi_i|_{H^1(\Omega)} + \max_{k=1,\dots,N_\kappa} \|\kappa_k\|_{W^{1,\infty}(\Omega)} |\Omega|^{1/2})\varepsilon_\times \\
&\quad \le \varepsilon.
\end{aligned}
$$

The asymptotic bounds on the size and depth of $\Phi_{\varepsilon,\varphi}$ can then be derived from (5.9), using Theorem 5.3, as developed in more detail in the proof of Theorem 5.6 below. □

5.2 Elliptic PDEs in polygonal domains

We establish exponential expressivity for realizations of NNs with ReLU activation of solution classes to elliptic PDEs in polygonal domains Ω, the boundaries $\partial\Omega$ of which are Lipschitz and consist of a finite number of straight line segments. Notably, $\Omega \subset \mathbb{R}^2$ need not be a finite union of axiparallel rectangles. In the following lemma, we construct a partition of unity in Ω subordinate to an open covering, of which each element is the affine image of one out of three *canonical patches*. Remark that we admit corners with associate angle of aperture π; this will be instrumental, in Corollaries 5.11 and 5.12, for the imposition of different boundary conditions on $\partial\Omega$. The three canonical patches that we consider are listed in Lemma 5.5, item [P2]. Affine images of $(0,1)^2$ are used away from corners of $\partial\Omega$ and when the internal angle of a corner is smaller than π. Affine images of $(-1,1)\times(0,1)$ are used near corners with internal angle π. PDE solutions may exhibit point singularities near such corners e.g. if the two neighboring edges have different types of boundary conditions. Affine images of $(-1,1)^2 \setminus (-1,0]^2$ are used near corners with internal angle larger than π. In the proof of Theorem 5.6, we use on each patch Theorem 4.3 or a result from Subsection 5.3 below.

A triangulation $\mathcal{T}$ of Ω is defined as a finite partition of Ω into open triangles K such that $\bigcup_{K\in\mathcal{T}} \overline{K} = \overline{\Omega}$. A *regular triangulation* of Ω is, additionally, a triangulation $\mathcal{T}$ of Ω such that, for any two neighboring elements $K_1, K_2 \in \mathcal{T}$, $\overline{K}_1 \cap \overline{K}_2$ is either a corner of both K_1 and K_2 or an entire edge of both K_1 and K_2. For a regular triangulation $\mathcal{T}$ of Ω, we denote by $S_1(\Omega,\mathcal{T})$ the space of functions $v \in C(\Omega)$ such that for every $K \in \mathcal{T}$, $v|_K \in \mathbb{P}_1$.

We postpone the proof of Lemma 5.5 to Appendix B.1.

Lemma 5.5. *Let $\Omega \subset \mathbb{R}^2$ be a polygon with Lipschitz boundary, consisting of straight sides, and with a finite set $\mathcal{C}$ of corners. Then, there exists $N_p \in \mathbb{N}$, a regular triangulation $\mathcal{T}$ of $\mathbb{R}^2$, such that for all $K \in \mathcal{T}$ either $K \subset \Omega$ or $K \subset \Omega^c$. Moreover, there exists a partition of unity $\{\phi_i\}_{i=1}^{N_p} \subset [S_1(\Omega,\mathcal{T})]^{N_p}$ such that*

[P1] $\operatorname{supp}(\phi_i) \cap \Omega \subset \Omega_i$ *for all* $i = 1, \dots, N_p$,

[P2] *for each $i \in \{1,\dots,N_p\}$, there exists an affine map $\psi_i\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $\psi_i^{-1}(\Omega_i) = \widehat{\Omega}_i$ for*

$$\widehat{\Omega}_i \in \{(0,1)^2, \Omega_{DN}, \Omega_F\}, \quad \textit{with} \quad \Omega_{DN} := (-1,1)\times(0,1), \quad \Omega_F := (-1,1)^2 \setminus (-1,0]^2;$$

[P3] $\mathcal{C} \cap \overline{\Omega}_i \subset \psi_i(\{(0,0)\})$ *for all* $i \in \{1,\dots,N_p\}$.

The following statement, then, provides expression rates for the NN approximation of functions in weighted analytic classes in polygonal domains.

Theorem 5.6. *Let $\Omega \subset \mathbb{R}^2$ be a polygon with Lipschitz boundary consisting of straight sides and with a finite set $\mathcal{C}$ of corners. Let $\underline{\gamma} = \{\gamma_c : c \in \mathcal{C}\}$ such that $\min \underline{\gamma} > 1$. Then, for all $u \in \mathcal{J}^\varpi_{\underline{\gamma}}(\Omega;\mathcal{C},\varnothing)$ and for every $0 < \varepsilon < 1$, there exists a NN $\Phi_{\varepsilon,u}$ such that*

$$\|u - \mathrm{R}(\Phi_{\varepsilon,u})\|_{H^1(\Omega)} \le \varepsilon. \tag{5.10}$$

In addition, as $\varepsilon \to 0$,

$$\mathrm{M}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|^5), \quad \mathrm{L}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)| \log(|\log(\varepsilon)|)).$$

Proof. We introduce, using Lemma 5.5, a regular triangulation $\mathcal{T}$ of $\mathbb{R}^2$, an open cover $\{\Omega_i\}_{i=1}^{N_p}$ of Ω, and a partition of unity $\{\phi_i\}_{i=1}^{N_p} \in [S_1(\Omega,\mathcal{T})]^{N_p}$ such that the properties [P1] – [P3] of Lemma 5.5 hold.

We define $\hat{u}_i := u_{|\Omega_i} \circ \psi_i : \widehat{\Omega}_i \to \mathbb{R}$. Since $u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega; \mathcal{C}, \varnothing)$ with $\min \underline{\gamma} > 1$ and since the maps ψ_i are affine, we observe that for every $i \in \{1, \dots, N_p\}$, there exists $\underline{\gamma}$ such that $\min \underline{\gamma} > 1$ and $\hat{u}_i \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(\widehat{\Omega}_i, \{(0,0)\}, \varnothing)$, because of [P2] and [P3]. Let

$$\varepsilon_1 := \frac{\varepsilon}{2N_p \max_{i\in\{1,\dots,N_p\}} \|\phi_i\|_{W^{1,\infty}(\Omega)} \left(\| \det J_{\psi_i} \|_{L^\infty((0,1)^2)} \left(1 + \| \| J_{\psi_i^{-1}} \|_2 \|^2_{L^\infty(\Omega_i)} \right) \right)^{1/2}}.$$

By Theorem 4.3 and by Lemma 5.21 and Theorem 5.16 in the forthcoming Subsection 5.3, there exist N_p NNs $\Phi^{\hat{u}_i}_{\varepsilon_1}$, $i \in \{1, \dots, N_p\}$, such that

$$\|\hat{u}_i - \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1})\|_{H^1(\widehat{\Omega}_i)} \le \varepsilon_1, \qquad \forall i \in \{1, \dots, N_p\}, \tag{5.11}$$

and there exists $C_\infty > 0$ independent of ε_1 such that, for all $i \in \{1, \dots, N_p\}$ and all $\hat{\varepsilon} \in (0,1)$

$$\| \mathrm{R}(\Phi^{\hat{u}_i}_{\hat{\varepsilon}}) \|_{L^\infty(\widehat{\Omega}_i)} \le C_\infty (1 + |\log \hat{\varepsilon}|^4).$$

The NNs given by Theorem 4.3, Lemma 5.21 and Theorem 5.16, which we here denote by $\widetilde{\Phi}^{\hat{u}_i}_{\varepsilon_1}$ for $i = 1, \dots, N_p$, may not have equal depth. Therefore, for all $i = 1, \dots, N_p$ and suitable $L_i \in \mathbb{N}$ we define $\Phi^{\hat{u}_i}_{\varepsilon_1} := \Phi^{\mathrm{Id}}_{1,L_i^1} \odot \widetilde{\Phi}^{\hat{u}_i}_{\varepsilon_1}$, so that the depth is the same for all $i = 1, \dots, N_p$. To estimate the size of the enlarged NNs, we use the fact that the size of a NN is not smaller than the depth unless the associated realization is constant. In the latter case, we could replace the NN by a NN with one non-zero weight without changing the realization. By this argument, we obtain for all $i = 1, \dots, N_p$ that $\mathrm{M}(\Phi^{\hat{u}_i}_{\varepsilon_1}) \le 2\,\mathrm{M}(\Phi^{\mathrm{Id}}_{1,L_i}) + 2\,\mathrm{M}(\widetilde{\Phi}^{\hat{u}_i}_{\varepsilon_1}) \le C \max_{j=1,\dots,N_p} \mathrm{L}(\widetilde{\Phi}^{\hat{u}_j}_{\varepsilon_1}) + C\,\mathrm{M}(\widetilde{\Phi}^{\hat{u}_i}_{\varepsilon_1}) \le C \max_{j=1,\dots,N_p} \mathrm{M}(\widetilde{\Phi}^{\hat{u}_j}_{\varepsilon_1})$. Furthermore, as shown in [17], there exist NNs Φ^{ϕ_i}, $i \in \{1, \dots, N_p\}$, such that

$$\mathrm{R}(\Phi^{\phi_i})(x) = \phi_i(x), \qquad \forall x \in \Omega,\ \forall i \in \{1, \dots, N_p\}.$$

Here we use that $\mathcal{T}$ is a partition $\mathbb{R}^2$, so that ϕ_i is defined on all of $\mathbb{R}^2$ and [17, Theorem 5.2] applies, which itself is based on [49]. Similarly to the previously handled case of $\Phi^{\hat{u}_i}_{\varepsilon_1}$, we can assume that Φ^{ϕ_i} for $i = 1, \dots, N_p$ all have equal depth and that the size of Φ^{ϕ_i} is bounded independent of i.

Since by [P2] the mappings ψ_i are affine and invertible, it follows that ψ_i^{-1} is affine for every $i \in \{1, \dots, N_p\}$. Thus, there exist NNs $\Phi^{\psi_i^{-1}}$, $i \in \{1, \dots, N_p\}$, of depth 1, such that

$$\mathrm{R}(\Phi^{\psi_i^{-1}})(x) = \psi_i^{-1}(x), \qquad \forall x \in \Omega_i,\ \forall i \in \{1, \dots, N_p\}. \tag{5.12}$$

Next, we define

$$\varepsilon_\times := \frac{\varepsilon}{2N_p(|\Omega|^{1/2} + 1 + |u|_{H^1(\Omega)} + \max_{i=1,\dots,N_p} \|\phi_i\|_{W^{1,\infty}(\Omega)} |\Omega|^{1/2})}$$

and $M_\times(\varepsilon_1) := C_\infty(1 + |\log \varepsilon_1|^4)$. Finally, we set

$$\Phi_{\varepsilon,u} := ((\underbrace{1, \dots, 1}_{N_p \text{ times}}), 0) \odot \mathrm{P}\left(\left\{ \Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)} \odot \mathrm{P}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}}, \Phi^{\mathrm{Id}}_{1,L} \odot \Phi^{\phi_i}) \right\}_{i=1}^{N_p} \right), \tag{5.13}$$

where $L \in \mathbb{N}$ is such that $\mathrm{L}(\Phi^{\hat{u}_1}_{\varepsilon_1} \odot \Phi^{\psi_1^{-1}}) = \mathrm{L}(\Phi^{\mathrm{Id}}_{1,L} \odot \Phi^{\phi_1})$, which yields that $\mathrm{M}(\Phi^{\mathrm{Id}}_{1,L}) \le C\,\mathrm{L}(\Phi^{\hat{u}_1}_{\varepsilon_1} \odot \Phi^{\psi_1^{-1}})$.

Approximation accuracy. By (5.13), we have for all $x \in \Omega$,

$$\mathrm{R}(\Phi_{\varepsilon,u})(x) = \sum_{i=1}^{N_p} \mathrm{R}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)}) \left(\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})(x), \mathrm{R}(\Phi^{\phi_i})(x) \right).$$

Therefore,

$$\begin{aligned} \|u - \mathrm{R}(\Phi_{\varepsilon,u})\|_{H^1(\Omega)} &\le \|u - \sum_{i=1}^{N_p} \phi_i \, \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})\|_{H^1(\Omega)} \\ &\quad + \sum_{i=1}^{N_p} \| \mathrm{R}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)}) \left(\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}}), \phi_i \right) - \phi_i \, \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})\|_{H^1(\Omega)} \\ &= (I) + (II). \end{aligned} \tag{5.14}$$

We start by considering term (I). For each $i \in \{1,\dots,N_p\}$, thanks to (5.11), there holds, with $\|J_{\psi_i^{-1}}\|_2^2$ denoting the square of the matrix 2-norm of the Jacobian of ψ_i^{-1},

$$
\begin{aligned}
\|u - \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})\|_{H^1(\Omega_i)} &= \|\hat{u}_i \circ \psi_i^{-1} - \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1}) \circ \psi_i^{-1}\|_{H^1(\Omega_i)} \\
&= \left(\int_{\widehat{\Omega}_i} \left(|\hat{u}_i|^2 + \left\| J_{\psi_i^{-1}} \nabla \left(\hat{u}_i - \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1}) \right) \right\|_2^2 \right) \det J_{\psi_i} dx \right)^{1/2} \\
&\leq \varepsilon_1 \left(\|\det J_{\psi_i}\|_{L^\infty(\widehat{\Omega}_i)} + \|\det J_{\psi_i}\|_{L^\infty(\widehat{\Omega}_i)} \| \|J_{\psi_i^{-1}}\|_2^2 \|_{L^\infty(\Omega_i)} \right)^{1/2} \\
&\leq \varepsilon_2 := \varepsilon_1 \max_i \left(\|\det J_{\psi_i}\|_{L^\infty(\widehat{\Omega}_i)} + \|\det J_{\psi_i}\|_{L^\infty(\widehat{\Omega}_i)} \| \|J_{\psi_i^{-1}}\|_2^2 \|_{L^\infty(\Omega_i)} \right)^{1/2}.
\end{aligned}
\tag{5.15}
$$

By [33, Theorem 2.1],

$$
(I) \leq N_p \varepsilon_2 \max_{i \in \{1,\dots,N_p\}} \|\phi_i\|_{W^{1,\infty}(\Omega)} \leq \frac{\varepsilon}{2}. \tag{5.16}
$$

We now consider term (II) in (5.14). There holds, by Theorem 4.3 and (5.12),

$$
\|\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})\|_{L^\infty(\Omega_i)} = \|\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1})\|_{L^\infty(\widehat{\Omega}_i)} \leq C_\infty (1 + |\log \varepsilon_1|^4)
$$

for all $i \in \{1,\dots,N_p\}$. Furthermore, by [P1], $\phi_i(x) = 0$ for all $x \in \Omega \setminus \Omega_i$ and, by Proposition 3.6,

$$
\mathrm{R}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)}) \left(\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})(x), \phi_i(x) \right) = 0, \qquad \forall x \in \Omega \setminus \Omega_i.
$$

From (5.15), we also have

$$
|\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})|_{H^1(\Omega_i)} \leq |u|_{H^1(\Omega_i)} + \|u - \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})\|_{H^1(\Omega_i)} \leq 1 + |u|_{H^1(\Omega_i)}.
$$

Hence,

$$
\begin{aligned}
(II) &= \sum_{i=1}^{N_p} \|\mathrm{R}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)}) \left(\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}}), \phi_i \right) - \phi_i \, \mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})\|_{H^1(\Omega_i)} \\
&\leq \sum_{i=1}^{N_p} \left(\|\mathrm{R}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)})(a,b) - ab\|_{W^{1,\infty}([-M_\times(\varepsilon_1), M_\times(\varepsilon_1)]^2)} \left(|\Omega|^{1/2} + |\mathrm{R}(\Phi^{\hat{u}_i}_{\varepsilon_1} \odot \Phi^{\psi_i^{-1}})|_{H^1(\Omega_i)} + |\phi_i|_{H^1(\Omega_i)} \right) \right) \\
&\leq N_p \varepsilon_\times \left(|\Omega|^{1/2} + 1 + |u|_{H^1(\Omega_i)} + |\Omega|^{1/2} \max_{i=1,\dots,N_p} \|\phi_i\|_{W^{1,\infty}(\Omega)} \right) \\
&\leq \frac{\varepsilon}{2}.
\end{aligned}
\tag{5.17}
$$

The asserted approximation accuracy follows by combining (5.14), (5.16), and (5.17).

Size of the neural network. To bound the size of the NN, we remark that N_p and the sizes of $\Phi^{\psi_i^{-1}}$ and of Φ^{ϕ_i} only depend on the domain Ω. Furthermore, there exist constants $C_{\Omega,i}$, $i = 1,2,3$, that depend only on Ω and u such that

$$
\begin{gathered}
|\log \varepsilon_1| \leq C_{\Omega,1}(1 + |\log \varepsilon|), \qquad |\log \varepsilon_\times| \leq C_{\Omega,2}(1 + |\log \varepsilon|), \\
|\log M_\times(\varepsilon_1)| \leq C_{\Omega,3}(1 + \log(1 + |\log \varepsilon|)).
\end{gathered}
\tag{5.18}
$$

From Theorem 4.3 and Proposition 3.6, in addition, there exist constants $C^L_{\hat{u}}, C^M_{\hat{u}}, C_\times > 0$ such that, for all $0 < \varepsilon_1, \varepsilon_\times \leq 1$,

$$
\begin{gathered}
\mathrm{L}(\Phi^{\hat{u}_i}_{\varepsilon_1}) \leq C^L_{\hat{u}}(1 + |\log \varepsilon_1|)(1 + \log(1 + |\log \varepsilon_1|)), \qquad \mathrm{M}(\Phi^{\hat{u}_i}_{\varepsilon_1}) \leq C^M_{\hat{u}}(1 + |\log \varepsilon_1|^5), \\
\max(\mathrm{M}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)}), \mathrm{L}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)})) \leq C_\times(1 + \log(M_\times(\varepsilon_1)^2/\varepsilon_\times)).
\end{gathered}
\tag{5.19}
$$

Then, by (5.13), we have

$$
\begin{aligned}
\mathrm{L}(\Phi_{\varepsilon,u}) &= 1 + \mathrm{L}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)}) + \max_{i=1,\dots,N_p} \left(\mathrm{L}(\Phi^{\hat{u}_i}_{\varepsilon_1}) + \mathrm{L}(\Phi^{\psi_i^{-1}}) \right), \\
\mathrm{M}(\Phi_{\varepsilon,u}) &\leq C \left(N_p + \mathrm{M}(\Pi^2_{\varepsilon_\times, M_\times(\varepsilon_1)}) + \sum_{i=1}^{N_p} \left(\mathrm{M}(\Phi^{\hat{u}_i}_{\varepsilon_1}) + \mathrm{M}(\Phi^{\psi_i^{-1}}) + \mathrm{M}(\Phi^{\mathrm{Id}}_{1,L}) + \mathrm{M}(\Phi^{\phi_i}) \right) \right).
\end{aligned}
\tag{5.20}
$$

The desired depth and size bounds follow from (5.18), (5.19), and (5.20). This concludes the proof. □

The exponential expression rate for the class of weighted, analytic functions in Ω by realizations of NNs with ReLU activation in the $H^1(\Omega)$-norm established in Theorem 5.6 implies an exponential expression rate bound on $\partial\Omega$, via the trace map and the fact that $\partial\Omega$ can be *exactly parametrized by the realization of a shallow NN with ReLU activation*. This is relevant for NN-based solution of boundary integral equations.

Corollary 5.7. (*NN expression of Dirichlet traces*) *Let $\Omega \subset \mathbb{R}^2$ be a polygon with Lipschitz boundary and a finite set $\mathcal{C}$ of corners. Let $\underline{\gamma} = \{\gamma_c : c \in \mathcal{C}\}$ such that $\min \underline{\gamma} > 1$. For any connected component Γ of $\partial\Omega$, let $\ell_\Gamma > 0$ be the length of Γ, such that there exists a continuous, piecewise affine parametrization $\theta : [0, \ell_\Gamma] \to \mathbb{R}^2 : t \mapsto \theta(t)$ of Γ with finitely many affine linear pieces and $\left\| \frac{d}{dt}\theta \right\|_2 = 1$ for almost all $t \in [0, \ell_\Gamma]$.*

Then, for all $u \in \mathcal{J}^\varpi_{\underline{\gamma}}(\Omega; \mathcal{C}, \varnothing)$ and for all $0 < \varepsilon < 1$, there exists a NN $\Phi_{\varepsilon,u,\theta}$ approximating the trace $Tu := u_{|\Gamma}$ such that

$$\|Tu - \mathrm{R}(\Phi_{\varepsilon,u,\theta}) \circ \theta^{-1}\|_{H^{1/2}(\Gamma)} \leq \varepsilon. \tag{5.21}$$

In addition, as $\varepsilon \to 0$,

$$\mathrm{M}(\Phi_{\varepsilon,u,\theta}) = \mathcal{O}(|\log(\varepsilon)|^5), \quad \mathrm{L}(\Phi_{\varepsilon,u,\theta}) = \mathcal{O}(|\log(\varepsilon)| \log(|\log(\varepsilon)|)).$$

Proof. We note that both components of θ are continuous, piecewise affine functions on $[0, \ell_\Gamma]$, thus they can be represented exactly as realization of a NN of depth two, with the ReLU activation function. Moreover, the number of weights of these NNs is of the order of the number of affine linear pieces of θ. We denote the parallelization of the NNs emulating exactly the two components of θ by Φ^θ.

By continuity of the trace operator $T : H^1(\Omega) \to H^{1/2}(\partial\Omega)$ (e.g. [10, 4]), there exists a constant $C_\Gamma > 0$ such that for all $v \in H^1(\Omega)$ it holds $\|Tv\|_{H^{1/2}(\Gamma)} \leq C_\Gamma \|v\|_{H^1(\Omega)}$, and without loss of generality we may assume $C_\Gamma \geq 1$.

Next, for any $\varepsilon \in (0,1)$, let $\Phi_{\varepsilon/C_\Gamma,u}$ be as given by Theorem 5.6. Define $\Phi_{\varepsilon,u,\theta} := \Phi_{\varepsilon/C_\Gamma,u} \odot \Phi^\theta$. It follows that

$$\left\|Tu - \mathrm{R}(\Phi_{\varepsilon,u,\theta}) \circ \theta^{-1}\right\|_{H^{1/2}(\Gamma)} = \left\|T\left(u - \mathrm{R}(\Phi_{\varepsilon/C_\Gamma,u})\right)\right\|_{H^{1/2}(\Gamma)} \leq C_\Gamma \left\|u - \mathrm{R}(\Phi_{\varepsilon/C_\Gamma,u})\right\|_{H^1(\Omega)} \leq \varepsilon.$$

The bounds on its depth and size follow directly from Proposition 3.2, Theorem 5.6, and the fact that the depth and size of Φ^θ are independent of ε. This finishes the proof. ☐

Remark 5.8. *The exponent 5 in the bound on the NN size $\mathrm{M}(\Phi_{\varepsilon,u,\theta})$ in Corollary 5.7 is likely not optimal, due to it being transferred from the NN rate in Ω.*

The proof of Theorem 5.6 established exponential expressivity of realizations of NNs with ReLU activation for the analytic class $\mathcal{J}^\varpi_{\underline{\gamma}}(\Omega; \mathcal{C}, \varnothing)$ in Ω. This implies that realizations of NNs can approximate, with exponential expressivity, solution classes of elliptic PDEs in polygonal domains Ω. We illustrate this by formulating concrete results for three problem classes: second order, linear, elliptic source and eigenvalue problems in Ω, and viscous, incompressible flow. To formulate the results, we specify the assumptions on Ω.

Definition 5.9 (Linear, second order, elliptic divergence-form differential operator with analytic coefficients)**.** *Let $d \in \{2, 3\}$ and let $\Omega \subset \mathbb{R}^d$ be a bounded domain. Let the coefficient functions $a_{ij}, b_i, c : \overline{\Omega} \to \mathbb{R}$ be real analytic in $\overline{\Omega}$, and such that the matrix function $A = (a_{ij})_{1 \leq i,j \leq d} : \Omega \to \mathbb{R}^{d \times d}$ is symmetric and uniformly positive definite in Ω. With these functions, we define the linear, second order, elliptic divergence-form differential operator $\mathcal{L}$ acting on $w \in C_0^\infty(\Omega)$ via (summation over repeated indices $i, j \in \{1, \ldots, d\}$)*

$$(\mathcal{L}w)(x) := -\partial_i(a_{ij}(x)\partial_j w(x)) + b_j(x)\partial_j w(x) + c(x)w(x) , \quad x \in \Omega .$$

Setting 5.10. *We assume that $\Omega \subset \mathbb{R}^2$ is an open, bounded polygon with boundary $\partial\Omega$ that is Lipschitz and connected. In addition, $\partial\Omega$ is the closure of a finite number $J \geq 3$ of straight, open sides Γ_j, i.e., $\Gamma_i \cap \Gamma_j = \varnothing$ for $i \neq j$ and $\partial\Omega = \bigcup_{1 \leq j \leq J} \overline{\Gamma_j}$. We assume the sides are enumerated cyclically, according to arc length, i.e. $\Gamma_{J+1} = \Gamma_1$. By n_j, we denote the exterior unit normal vector to Ω on Γ_j and by $\mathbf{c}_j := \overline{\Gamma_{j-1}} \cap \overline{\Gamma_j}$ the corner j of Ω.*

With $\mathcal{L}$ as in Definition 5.9, we associate on boundary segment Γ_j a boundary operator $\mathcal{B}_j \in \{\gamma_0^j, \gamma_1^j\}$, i.e. either the Dirichlet trace γ_0 or the distributional (co-)normal derivative operator γ_1, acting on $w \in C^1(\overline{\Omega})$ via

$$\gamma_0^j w := w|_{\Gamma_j} , \qquad \gamma_1^j w := (A\nabla w) \cdot n_j|_{\Gamma_j}, \quad j = 1, ..., J . \tag{5.22}$$

We collect the boundary operators $\mathcal{B}_j$ in $\mathcal{B} := \{\mathcal{B}_j\}_{j=1}^J$.

The first corollary addresses exponential ReLU expressibility of solutions of the source problem corresponding to $(\mathcal{L}, \mathcal{B})$.

Corollary 5.11. *Let Ω, $\mathcal{L}$, and $\mathcal{B}$ be as in Setting 5.10 with $d = 2$. For f analytic in $\overline{\Omega}$, let u denote a solution to the boundary value problem*

$$\mathcal{L}u = f \ \text{ in } \ \Omega, \qquad \mathcal{B}u = 0 \ \text{ on } \ \partial\Omega \,. \tag{5.23}$$

Then, for every $0 < \varepsilon < 1$, there exists a NN $\Phi_{\varepsilon,u}$ such that

$$\|u - \mathrm{R}(\Phi_{\varepsilon,u})\|_{H^1(\Omega)} \leq \varepsilon. \tag{5.24}$$

In addition, $\mathrm{M}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|^5)$ and $\mathrm{L}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))$, as $\varepsilon \to 0$.

Proof. The proof is obtained by verifying weighted, analytic regularity of solutions. By [5, Theorem 7.2], there exists $\underline{\gamma}$ such that $\min \underline{\gamma} > 1$ and $u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega; \mathcal{C}, \varnothing)$. Then, the application of Theorem 5.6 concludes the proof. □

Next, we address NN expression rates for eigenfunctions of $(\mathcal{L}, \mathcal{B})$.

Corollary 5.12. *Let Ω, $\mathcal{L}$, $\mathcal{B}$ be as in Setting 5.10 with $d = 2$ and let $0 \neq w \in H^1(\Omega)$ be an eigenfunction of the elliptic eigenvalue problem*

$$\mathcal{L}w = \lambda w \text{ in } \Omega, \qquad \mathcal{B}w = 0 \text{ on } \partial\Omega. \tag{5.25}$$

Then, for every $0 < \varepsilon < 1$, there exists a NN $\Phi_{\varepsilon,w}$ such that

$$\|w - \mathrm{R}(\Phi_{\varepsilon,w})\|_{H^1(\Omega)} \leq \varepsilon. \tag{5.26}$$

In addition, $\mathrm{M}(\Phi_{\varepsilon,w}) = \mathcal{O}(|\log(\varepsilon)|^5)$ and $\mathrm{L}(\Phi_{\varepsilon,w}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))$, as $\varepsilon \to 0$.

Proof. The statement follows from [5] and Theorem 5.6 as in Corollary 5.11. □

The analytic regularity of solutions u in the proof of Theorem 5.6 also holds for certain nonlinear, elliptic PDEs. We illustrate it for the velocity field of viscous, incompressible flow in Ω.

Corollary 5.13. *Let $\Omega \subset \mathbb{R}^2$ be as in Setting 5.10. Let $\nu > 0$ and let $\boldsymbol{u} \in H^1_0(\Omega)^2$ be the velocity field of the Leray solutions of the viscous, incompressible Navier-Stokes equations in Ω, with homogeneous Dirichlet ("no slip") boundary conditions*

$$-\nu\Delta\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} + \nabla p = \boldsymbol{f} \text{ in } \Omega, \qquad \nabla\cdot\boldsymbol{u} = 0 \text{ in } \Omega, \qquad \boldsymbol{u} = \boldsymbol{0} \text{ on } \partial\Omega, \tag{5.27}$$

where the components of $\boldsymbol{f}$ are analytic in $\overline{\Omega}$ and such that $\|\boldsymbol{f}\|_{H^{-1}(\Omega)}/\nu^2$ is small enough so that $\boldsymbol{u}$ is unique.

Then, for every $0 < \varepsilon < 1$, there exists a NN $\Phi_{\varepsilon,\boldsymbol{u}}$ with two-dimensional output such that

$$\|\boldsymbol{u} - \mathrm{R}(\Phi_{\varepsilon,\boldsymbol{u}})\|_{H^1(\Omega)} \leq \varepsilon. \tag{5.28}$$

In addition, $\mathrm{M}(\Phi_{\varepsilon,\boldsymbol{u}}) = \mathcal{O}(|\log(\varepsilon)|^5)$ and $\mathrm{L}(\Phi_{\varepsilon,\boldsymbol{u}}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))$, as $\varepsilon \to 0$.

Proof. The velocity fields of Leray solutions of the Navier-Stokes equations in Ω satisfy the weighted, analytic regularity $\boldsymbol{u} \in \left[\mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega; \mathcal{C}, \varnothing)\right]^2$, with $\min\underline{\gamma} > 1$, see [31]. Then, the application of Theorem 5.6 concludes the proof. □

5.3 Elliptic PDEs in Fichera-type polyhedral domains

Fichera-type polyhedral domains $\Omega \subset \mathbb{R}^3$ are, loosely speaking, closures of finite, disjoint unions of (possibly affinely mapped) axiparallel hexahedra with $\partial\Omega$ Lipschitz. In Fichera-type domains, analytic regularity of solutions of linear, elliptic boundary value problems from acoustics and linear elasticity in displacement formulation has been established in [5]. As an example of a boundary value problem covered by [5] and our theory, consider $\Omega_F := (-1,1)^d \setminus (-1,0]^d$ for $d = 2, 3$, displayed for $d = 3$ in Figure 2. We introduce the *setting for elliptic problems with analytic coefficients in Ω_F*. Note that the boundary of Ω_F is composed of 6 edges when $d = 2$ and of 9 faces when $d = 3$.

Setting 5.14. *We assume that $\mathcal{L}$ is an elliptic operator as in Definition 5.9. On each edge (if $d = 2$) or face (if $d = 3$) Γ_j, $j \in \{1, \ldots, 3d\}$ of $\partial\Omega_F$, we introduce the boundary operator $\mathcal{B}_j \in \{\gamma_0, \gamma_1\}$, where γ_0 and γ_1 are defined as in* (5.22). *We collect the boundary operators $\mathcal{B}_j$ in $\mathcal{B} := \{\mathcal{B}_j\}_{j=1}^{3d}$.*

For a right hand side f, the elliptic boundary value problem we consider in this section is then

$$\mathcal{L}u = f \text{ in } \Omega_F, \quad \mathcal{B}u = 0 \text{ on } \partial\Omega_F. \tag{5.29}$$

The following extension lemma will be useful for the approximation of the solution to (5.29) by NNs. We postpone its proof to Appendix B.2.

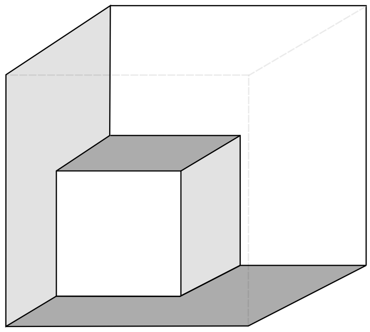

Figure 2: Example of Fichera-type corner domain.

Lemma 5.15. *Let $d \in \{2,3\}$ and $u \in W^{1,1}_{\mathrm{mix}}(\Omega_F)$. Then, there exists a function $v \in W^{1,1}_{\mathrm{mix}}((-1,1)^d)$ such that $v|_{\Omega_F} = u$. The extension is stable with respect to the $W^{1,1}_{\mathrm{mix}}$ norm.*

We denote the set containing all corners (including the re-entrant one) of Ω_F as

$$\mathcal{C}_F = \{-1,0,1\}^d \setminus (-1,\dots,-1).$$

When $d = 3$, for all $c \in \mathcal{C}_F$, then we denote by $\mathcal{E}_c$ the set of edges abutting at c and we denote $\mathcal{E}_F := \bigcup_{c\in\mathcal{C}_F} \mathcal{E}_c$.

Theorem 5.16. *Let $u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega_F;\mathcal{C}_F,\mathcal{E}_F)$ with*

$$\begin{aligned} &\underline{\gamma} = \{\gamma_c : c \in \mathcal{C}_F\}, && \text{with } \gamma_c > 1, \text{ for all } c \in \mathcal{C}_F && \text{if } d=2,\\ &\underline{\gamma} = \{\gamma_c,\gamma_e : c \in \mathcal{C}_F, e\in\mathcal{E}_F\}, && \text{with } \gamma_c > 3/2 \text{ and } \gamma_e > 1, \text{ for all } c \in \mathcal{C}_F \text{ and } e \in \mathcal{E}_F && \text{if } d=3. \end{aligned}$$

Then, for any $0<\varepsilon<1$ there exists a NN $\Phi_{\varepsilon,u}$ so that

$$\|u - \mathrm{R}(\Phi_{\varepsilon,u})\|_{H^1(\Omega_F)} \le \varepsilon. \tag{5.30}$$

In addition, $\|\mathrm{R}(\Phi_{\varepsilon,u})\|_{L^\infty(\Omega_F)} = \mathcal{O}(1+|\log\varepsilon|^{2d})$, for $\varepsilon\to 0$. Also, $\mathrm{M}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|^{2d+1})$ and $\mathrm{L}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))$, for $\varepsilon\to 0$.

Proof. By Lemma 5.15, we extend the function u to a function $\tilde u$ such that

$$\tilde u \in W^{1,1}_{\mathrm{mix}}((-1,1)^d) \quad \text{and} \quad \tilde u|_{\Omega_F} = u.$$

Note that, by the stability of the extension, there exists a constant $C_{\mathrm{ext}}>0$ independent of u such that

$$\|\tilde u\|_{W^{1,1}_{\mathrm{mix}}((-1,1)^d)} \le C_{\mathrm{ext}}\|u\|_{W^{1,1}_{\mathrm{mix}}(\Omega_F)}. \tag{5.31}$$

Since $u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega_F;\mathcal{C}_F,\mathcal{E}_F)$, there holds $u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(S;\mathcal{C}_S,\mathcal{E}_S)$ for all

$$S \in \left\{ \bigtimes_{j=1}^{d} (a_j, a_j+1/2) : (a_1,\dots,a_d) \in \{-1,-1/2,0,1/2\}^d \right\} \text{ such that } S\cap\Omega_F \neq \varnothing \tag{5.32}$$

with $\mathcal{C}_S = \overline{S}\cap\mathcal{C}_F$ and $\mathcal{E}_S = \{e\in\mathcal{E}_F : e \subset \overline{S}\}$. Since $S\subset\Omega_F$ and $\tilde u|_{\Omega_F} = u|_{\Omega_F}$, we also have

$$\tilde u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(S;\mathcal{C}_S,\mathcal{E}_S) \text{ for all } S \text{ satisfying (5.32)}.$$

By Theorem A.25 exist $C_p>0$, $C_{\widetilde N_{\mathrm{1d}}}>0$, $C_{\widetilde N_{\mathrm{int}}}>0$, $C_{\tilde v}>0$, $C_{\tilde c}>0$, and $b_{\tilde v}>0$ such that, for all $0<\varepsilon\le 1$, there exists $p\in\mathbb{N}$, a partition $\mathcal{G}_{\mathrm{1d}}$ of $(-1,1)$ into $\widetilde N_{\mathrm{int}}$ open, disjoint, connected subintervals, a d-dimensional array $c\in\mathbb{R}^{\widetilde N_{\mathrm{1d}}\times\cdots\times\widetilde N_{\mathrm{1d}}}$, and piecewise polynomials $\tilde v_i \in \mathbb{Q}_p(\mathcal{G}_{\mathrm{1d}})\cap H^1((-1,1))$, $i=1,\dots,\widetilde N_{\mathrm{1d}}$, such that

$$\widetilde N_{\mathrm{1d}} \le C_{\widetilde N_{\mathrm{1d}}}(1+|\log\varepsilon|^2), \quad \widetilde N_{\mathrm{int}} \le C_{\widetilde N_{\mathrm{int}}}(1+|\log\varepsilon|), \quad \|c\|_1 \le C_{\tilde c}(1+|\log\varepsilon|^{2d}), \quad p\le C_p(1+|\log\varepsilon|)$$

and

$$\|\tilde v_i\|_{H^1(I)} \le C_{\tilde v}\varepsilon^{-b_{\tilde v}}, \qquad \|\tilde v_i\|_{L^\infty(I)} \le 1, \qquad \forall i\in\{1,\dots,\widetilde N_{\mathrm{1d}}\}.$$

Furthermore,

$$\|u - v_{\mathsf{hp}}\|_{H^1(\Omega_F)} = \|\tilde{u} - v_{\mathsf{hp}}\|_{H^1(\Omega_F)} \leq \frac{\varepsilon}{2}, \qquad v_{\mathsf{hp}} = \sum_{i_1,\dots,i_d=1}^{\widetilde{N}_{1\mathrm{d}}} \widetilde{c}_{i_1\dots i_d} \bigotimes_{j=1}^{d} \tilde{v}_{i_j}.$$

Due to the stability (5.31) and to Lemmas A.21 and A.22, there holds

$$\|\widetilde{c}\|_1 \leq C N_{\mathrm{int}}^{2d} \|u\|_{\mathcal{J}^d_{\underline{\gamma}}(\Omega_F)},$$

i.e., the bound on the coefficients $\widetilde{c}$ is independent of the extension $\tilde{u}$ of u. By Theorem 4.2, there exists a NN $\Phi_{\varepsilon,u}$ with the stated approximation properties and asymptotic size bounds. The bound on the $L^\infty(\Omega_F)$ norm of the realization of $\Phi_{\varepsilon,u}$ follows as in the proof of Theorem 4.3. □

Remark 5.17. *Arguing as in Corollary 5.7, a NN with ReLU activation and two-dimensional input can be constructed so that its realization approximates the Dirichlet trace of solutions to* (5.29) *in* $H^{1/2}(\partial\Omega_F)$ *at an exponential rate in terms of the NN size* M*.*

The following statement now gives expression rate bounds for the approximation of solutions to the Fichera problem (5.29) by realizations of NNs with the ReLU activation function.

Corollary 5.18. *Let f be an analytic function on $\overline{\Omega}_F$ and let u be a solution to* (5.29) *with operators $\mathcal{L}$ and $\mathcal{B}$ as in Setting 5.14 and with source term f. Then, for any $0 < \varepsilon < 1$ there exists a NN $\Phi_{\varepsilon,u}$ so that*

$$\|u - \mathrm{R}\left(\Phi_{\varepsilon,u}\right)\|_{H^1(\Omega_F)} \leq \varepsilon. \tag{5.33}$$

In addition, $\mathrm{M}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|^{2d+1})$ *and* $\mathrm{L}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))$*, for* $\varepsilon \to 0$*.*

Proof. By [5, Corollary 7.1 and Theorem 7.4], there exists $\underline{\gamma}$ such that $\gamma_c - d/2 > 0$ for all $c \in \mathcal{C}_F$ and $\gamma_e > 1$ for all $e \in \mathcal{E}_F$ such that $u \in \mathcal{J}^\varpi_{\underline{\gamma}}(\Omega_F; \mathcal{C}_F, \mathcal{E}_F)$. An application of Theorem 5.16 concludes the proof. □

Remark 5.19. *By* [5, Corollary 7.1 and Theorem 7.4]*, Corollary 5.18 holds verbatim also under the hypothesis that the right-hand side f is weighted analytic, with singularities at the corners/edges of the domain; specifically,* (5.33) *and the size bounds on the NN $\Phi_{\varepsilon,u}$ hold under the assumption that there exists $\underline{\gamma}$ such that $\gamma_c - d/2 > 0$ for all $c \in \mathcal{C}_F$ and $\gamma_e > 1$ for all $e \in \mathcal{E}_F$ such that*

$$f \in \mathcal{J}^\varpi_{\underline{\gamma}-2}(\Omega_F; \mathcal{C}_F, \mathcal{E}_F).$$

Remark 5.20. *The numerical approximation of solutions for* (5.29) *with a NN in two dimensions has been investigated e.g. in* [26] *using the so-called 'PINNs' methodology. There, the loss function was based on minimization of the residual of the NN approximation in the strong form of the PDE. Evidently, a different (smoother) activation than the ReLU activations considered here had to be used. Starting from the approximation of products by NNs with smoother activation functions introduced in* [44, Sec.3.3] *and following the same line of reasoning as in the present paper, the results we obtain for ReLU-based realizations of NNs can be extended to large classes of NNs with smoother activations and similar architecture.*

Furthermore, in [6, Section 3.1]*, a slightly different elliptic boundary value problem is numerically approximated by realizations of NNs. Its solutions exhibit the same weighted, analytic regularity as considered in this paper. The presently obtained approximation rates by NN realizations extend also to the approximation of solutions for the problem considered in* [6]*.*

In the proof of Theorem 5.6, we require in particular the approximation of weighted analytic functions on $(-1,1)\times(0,1)$ with a corner singularity at the origin. For convenient reference, we detail the argument in this case.

Lemma 5.21. *Let $d = 2$ and $\Omega_{DN} := (-1,1) \times (0,1)$. Denote $\mathcal{C}_{DN} = \{-1,0,1\} \times \{0,1\}$. Let $u \in \mathcal{J}^\varpi_{\underline{\gamma}}(\Omega_{DN}; \mathcal{C}_{DN}, \varnothing)$ with $\underline{\gamma} = \{\gamma_c : c \in \mathcal{C}_{DN}\}$, with $\gamma_c > 1$ for all $c \in \mathcal{C}_{DN}$.*

Then, for any $0 < \varepsilon < 1$ there exists a NN $\Phi_{\varepsilon,u}$ so that

$$\|u - \mathrm{R}\left(\Phi_{\varepsilon,u}\right)\|_{H^1(\Omega_{DN})} \leq \varepsilon. \tag{5.34}$$

In addition, $\|\,\mathrm{R}\left(\Phi_{\varepsilon,u}\right)\|_{L^\infty(\Omega_{DN})} = \mathcal{O}(1 + |\log\varepsilon|^4)$ *, for* $\varepsilon \to 0$*. Also,* $\mathrm{M}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|^5)$ *and* $\mathrm{L}(\Phi_{\varepsilon,u}) = \mathcal{O}(|\log(\varepsilon)|\log(|\log(\varepsilon)|))$*, for* $\varepsilon \to 0$*.*

Proof. Let $\tilde{u} \in W^{1,1}_{\text{mix}}((-1,1)^2)$ be defined by

$$\begin{cases} \tilde{u}(x_1, x_2) = u(x_1, x_2) & \text{for all } (x_1, x_2) \in (-1,1) \times [0,1), \\ \tilde{u}(x_1, x_2) = u(x_1, 0) & \text{for all } (x_1, x_2) \in (-1,1) \times (-1,0), \end{cases}$$

such that $\tilde{u}|_{\Omega_{DN}} = u$. Here we used that there exist continuous imbeddings $\mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega_{DN}; \mathcal{C}_{DN}, \varnothing) \hookrightarrow W^{1,1}_{\text{mix}}(\Omega_{DN}) \hookrightarrow C^0(\overline{\Omega_{DN}})$ (see Lemma A.22 for the first imbedding), i.e. u can be extended to a continuous function on $\overline{\Omega_{DN}}$.

As in the proof of Lemma 5.15, this extension is stable, i.e. there exists a constant $C_{\text{ext}} > 0$ independent of u such that

$$\|\tilde{u}\|_{W^{1,1}_{\text{mix}}((-1,1)^d)} \le C_{\text{ext}} \|u\|_{W^{1,1}_{\text{mix}}(\Omega_{DN})}. \tag{5.35}$$

Because $u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(\Omega_{DN}; \mathcal{C}_{DN}, \varnothing)$, it holds with $\mathcal{C}_S = \overline{S} \cap \mathcal{C}_{DN}$ that $u \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(S; \mathcal{C}_S, \varnothing)$ for all

$$S \in \left\{ \bigtimes_{j=1,2} (a_j, a_j + 1/2) : (a_1, a_2) \in \{-1, -1/2, 0, 1/2\} \times \{0, 1/2\} \right\}.$$

The remaining steps are the same as those in the proof of Theorem 5.16. □

6 Conclusions and extensions

We review the main findings of the present paper and outline extensions of the present results, and perspectives for further research.

6.1 Principal mathematical results

We established exponential expressivity of realizations of NNs with the ReLU activation function in the Sobolev norm H^1 for functions which belong to certain countably normed, weighted analytic function spaces in cubes $Q = (0,1)^d$ of dimension $d = 2, 3$. The admissible function classes comprise functions which are real analytic at points $x \in Q$, and which admit analytic extensions to the open sides $F \subset \partial Q$, but may have singularities at corners and (in space dimension $d = 3$) edges of Q. We have also extended this result to cover exponential expressivity of realizations of NNs with ReLU activation for solution classes of linear, second order elliptic PDEs in divergence form in plane, polygonal domains and of elliptic, nonlinear eigenvalue problems with singular potentials in three space dimensions. Being essentially an approximation result, the DNN expression rate bound in Theorem 5.6 will apply to any elliptic boundary value problem in polygonal domains where weighted, analytic regularity is available. Apart from the source and eigenvalue problems, such regularity is in space dimension $d = 2$ also available for linearized elastostatics, Stokes flow and general elliptic systems [12, 15, 5].

The established approximation rates of realizations of NNs with ReLU activation are fundamentally based on a novel exponential upper bound on approximation of weighted analytic functions via tensorized *hp* approximations on multi-patch configurations in finite unions of axiparallel rectangles/hexahedra. The *hp* approximation result is presented in Theorem A.25 and of independent interest in the numerical analysis of spectral elements.

The proofs of exponential expressivity of NN realizations are, in principle, constructive. They are based on explicit bounds on the coefficients of *hp* projections and on corresponding emulation rate bounds for the (re)approximation of modal *hp* bases.

6.2 Extensions and future work

The tensor structure of the *hp* approximation considered here limited geometries of domains that are admissible for our results. *Curvilinear, mapped domains* with analytic domain maps will allow corresponding approximation rates, with the NN approximations obtained by composing the present constructions with NN emulations of the domain maps and the fact that compositions of NNs are again NNs.

The only activation function considered in this work is the ReLU. Following the same strategy, similar expression rate bounds can be obtained for functions with smoother, nonlinear activation functions. We refer to Remark 5.20 and to the discussion in [44, Sec. 3.3].

The principal results in Section 5.1 yield exponential expressivity of realizations of NNs with ReLU activation for singular eigenvalue problems with multiple, isolated point singularities as arise in electron-structure computations *for static molecules with known loci of the nuclei*. Inspection of our proofs reveals that the expression rate bounds are robust with respect to perturbations of the nuclei sites; only interatomic distances enter the constants in the expression rate bounds of Section 5.1.2. Given the *closedness of NNs under composition*, obtaining similar expression rates also for solutions of the *vibrational Schrödinger equation* appears in principle possible.

The presently proved deep ReLU NN expression rate bounds can, in connection with recently proposed residual-based DNN training methodologies (e.g., [46]), imply exponential convergence rates of numerical PDE solutions based on machine learning approaches.

A Tensor product hp approximation

In this section, we construct the *hp* tensor product approximation which will then be emulated to obtain the NN expression rate estimates. We denote $Q = (0,1)^d$, $d \in \{2,3\}$ and introduce the set of corners $\mathcal{C}$,

$$\mathcal{C} = \begin{cases} \{(0,0)\} & \text{if } d = 2, \\ \{(0,0,0)\} & \text{if } d = 3, \end{cases} \tag{A.1}$$

and the set of edges $\mathcal{E}$,

$$\mathcal{E} = \begin{cases} \varnothing & \text{if } d = 2, \\ \{\{0\}\times\{0\}\times(0,1), \{0\}\times(0,1)\times\{0\}, (0,1)\times\{0\}\times\{0\}\} & \text{if } d = 3. \end{cases} \tag{A.2}$$

The results in this section extend, by rotation or reflection, to the case where $\mathcal{C}$ contains any of the corners of Q and $\mathcal{E}$ is the set of the adjacent edges when $d = 3$. Most of the section addresses the construction of exponentially consistent *hp*-quasiinterpolants in the reference cube $(0,1)^d$; in Section A.10 the analysis will be extended to domains which are specific finite unions of such patches.

A.1 Product geometric mesh and tensor product hp space

We fix a geometric mesh grading factor $\sigma \in (0, 1/2]$. Furthermore, let

$$J_0^\ell = (0, \sigma^\ell) \qquad \text{and} \qquad J_k^\ell = (\sigma^{\ell-k+1}, \sigma^{\ell-k}),\ k = 1,\dots,\ell.$$

In $(0,1)$, the geometric mesh with ℓ layers is $\mathcal{G}_1^\ell = \{J_k^\ell : k = 0,\dots,\ell\}$. Moreover, we denote the nodes of $\mathcal{G}_1^\ell$ by $x_0^\ell = 0$ and $x_k^\ell = \sigma^{\ell-k+1}$ for $k = 1,\dots,\ell+1$. In $(0,1)^d$, the d-dimensional tensor product geometric mesh is[1]

$$\mathcal{G}_d^\ell = \left\{ \bigtimes_{i=1}^d K_i, \text{ for all } K_1,\dots,K_d \in \mathcal{G}_1^\ell \right\}.$$

For an element $K = \bigtimes_{i=1}^d J_{k_i}^\ell$, $k_i \in \{0,\dots,\ell\}$, we denote by d_c^K the distance from the singular corner, and d_e^K the distance from the closest singular edge. We observe that

$$d_c^K = \left(\sum_{i=1}^d \sigma^{2(\ell-k_i+1)} \right)^{1/2} \tag{A.3}$$

and

$$d_e^K = \min_{(i_1,i_2)\in\{1,2,3\}^2} \left(\sum_{i\in\{i_1,i_2\}} \sigma^{2(\ell-k_i+1)} \right)^{1/2}. \tag{A.4}$$

The *hp tensor product space* is defined as

$$X_{\mathsf{hp},d}^{\ell,p} \coloneqq \{v \in H^1(Q) : v_{|K} \in \mathbb{Q}_p(K),\ \text{for all } K \in \mathcal{G}_d^\ell\},$$

where $\mathbb{Q}_p(K) \coloneqq \operatorname{span}\left\{ \prod_{i=1}^d (x_i)^{k_i} : k_i \le p, i = 1,\dots,d \right\}$. Note that, by construction, $X_{\mathsf{hp},d}^{\ell,p} = \bigotimes_{i=1}^d X_{\mathsf{hp},1}^{\ell,p}$.

For positive integers p and s such that $1 \le s \le p$, we will write

$$\Psi_{p,s} \coloneqq \frac{(p-s)!}{(p+s)!}. \tag{A.5}$$

Additionally, we will denote, for all $\sigma \in (0,1/2]$,

$$\tau_\sigma \coloneqq \frac{1-\sigma}{\sigma} \in [1,\infty). \tag{A.6}$$

[1]We assume *isotropic tensorization*, i.e. the same σ and the same number of geometric mesh layers in each coordinate direction; all approximation results remain valid (with possibly better numerical values for the constants in the error bounds) for anisotropic, co-ordinate dependent choices of ℓ and of σ.

A.2 Local projector

We denote the reference interval by $I = (-1,1)$ and the reference cube by $\widehat{K} = (-1,1)^d$. We also write $H^1_{\mathrm{mix}}(\widehat{K}) = \bigotimes_{i=1}^d H^1(I) \supset H^d(\widehat{K})$. Let $p \geq 1$: we introduce the univariate projectors $\widehat{\pi}_p : H^1(I) \to \mathbb{P}_p(I)$ as

$$\begin{aligned}(\widehat{\pi}_p \hat{v})(x) &= \hat{v}(-1) + \sum_{n=0}^{p-1} \left(\hat{v}', \frac{2n+1}{2} L_n\right) \int_{-1}^{x} L_n(\xi) d\xi \\ &= \hat{v}(-1)\left(\frac{1-x}{2}\right) + \hat{v}(1)\left(\frac{1+x}{2}\right) + \sum_{n=1}^{p-1} \left(\hat{v}', \frac{2n+1}{2} L_n\right) \int_{-1}^{x} L_n(\xi) d\xi,\end{aligned} \tag{A.7}$$

where L_n is the nth Legendre polynomial, L^∞ normalized, and $(\cdot,\cdot)$ is the scalar product of $L^2((-1,1))$. Note that

$$(\widehat{\pi}_p \hat{v})(\pm 1) = \hat{v}(\pm 1), \qquad \forall \hat{v} \in H^1(I). \tag{A.8}$$

For $p \in \mathbb{N}$, we introduce the projection on the reference element $\widehat{K}$ as $\widehat{\Pi}_p = \bigotimes_{i=1}^d \widehat{\pi}_p$. For all $K \in \mathcal{G}^\ell_d$, we introduce an affine transformation from K to the reference element

$$\Phi_K : K \to \widehat{K} \quad \text{such that} \quad \Phi_K(K) = \widehat{K}. \tag{A.9}$$

Remark that since the elements are axiparallel, the affine transformation can be written as a d-fold product of one dimensional affine transformations $\phi_k : J^\ell_k \to I$, i.e., supposing that $K = \bigtimes_{i=1}^d J^\ell_{k_i}$, there holds

$$\Phi_K = \bigotimes_{i=1}^d \phi_{k_i}.$$

Let $K \in \mathcal{G}^\ell_d$ and let k_i, $i = 1,\dots,d$ be the indices such that $K = \bigtimes_{i=1}^d J^\ell_{k_i}$. Define, for $w \in H^1(J^\ell_{k_i})$,

$$\pi^{k_i}_p w = \left(\widehat{\pi}_p (w \circ \phi^{-1}_{k_i})\right) \circ \phi_{k_i}.$$

For v defined on K such that $v \circ \Phi^{-1}_K \in H^1_{\mathrm{mix}}(\widehat{K})$ and for $(p_1,\dots,p_d) \in \mathbb{N}^d$, we introduce the local projection operator

$$\Pi^K_{p_1 \dots p_d} = \bigotimes_{i=1}^d \pi^{k_i}_{p_i}. \tag{A.10}$$

We also write

$$\Pi^K_p v = \Pi^K_{p \dots p} v = \left(\widehat{\Pi}_p (v \circ \Phi^{-1}_K)\right) \circ \Phi_K. \tag{A.11}$$

For later reference, we note the following property of $\Pi^K_p v$:

Lemma A.1. *Let K_1, K_2 be two axiparallel cubes that share one regular face F (i.e., F is an entire face of both K_1 and K_2). Then, for $v \in H^1_{\mathrm{mix}}(\mathrm{int}\,(\overline{K}_1 \cup \overline{K}_2))$, the piecewise polynomial*

$$\Pi^{K_1 \cup K_2}_p v = \begin{cases} \Pi^{K_1}_p v & \text{in } K_1, \\ \Pi^{K_2}_p v & \text{in } K_2 \end{cases}$$

is continuous across F.

Proof. This follows directly from (A.8). □

A.3 Global projectors

We introduce, for $\ell, p \in \mathbb{N}$, the univariate projector $\pi^{\ell,p}_{\mathsf{hp}} : H^1((0,1)) \to X^{\ell,p}_{\mathsf{hp},1}$ as

$$\left(\pi^{\ell,p}_{\mathsf{hp}} u\right)(x) = \begin{cases} \left(\pi^0_1 u\right)(x) & \text{if } x \in J^\ell_0, \\ \left(\pi^k_p u\right)(x) & \text{if } x \in J^\ell_k,\ k \in \{1,\dots,\ell\}. \end{cases} \tag{A.12}$$

Note that for all $\ell \in \mathbb{N}$, for $x \in J^\ell_0$

$$\left(\pi^0_1 u\right)(x) = u(0) + \sigma^{-\ell}\left(u(\sigma^\ell) - u(0)\right) x.$$

The d-variate *hp* quasi-interpolant is then obtained by tensorization, i.e.

$$\Pi^{\ell,p}_{\mathsf{hp},d} := \bigotimes_{i=1}^d \pi^{\ell,p}_{\mathsf{hp}}. \tag{A.13}$$

Remark A.2. *By the nodal exactness of the projectors, the operator $\Pi^{\ell,p}_{\mathsf{hp},d}$ is continuous across interelement interfaces (see Lemma A.1), hence its image is contained in $H^1((0,1)^d)$. The continuity can also be observed from the expansion in terms of continuous, globally defined basis functions given in Proposition A.24.*

Remark A.3. *The projector $\Pi^{\ell,p}_{\mathsf{hp},d}$ is defined on a larger space than $H^1_{\mathrm{mix}}(Q)$ as specified below (e.g. Remark A.20).*

A.4 Preliminary estimates

The projector on $\widehat{K}$ given by

$$\widehat{\Pi}_{p_1\dots p_d} := \bigotimes_{i=1}^{d} \widehat{\pi}_{p_i} \tag{A.14}$$

has the following property.

Lemma A.4 ([43, Propositions 5.2 and 5.3])**.** *Let $d=3$, $(p_1,p_2,p_3)\in\mathbb{N}^3$, and $(s_1,s_2,s_3)\in\mathbb{N}^3$ with $1\le s_i\le p_i$. Then the projector $\widehat{\Pi}_{p_1p_2p_3}: H^1_{\mathrm{mix}}(\widehat{K})\to\mathbb{Q}_{p_1,p_2,p_3}(\widehat{K})$ satisfies that*

$$\begin{aligned}\|v-\widehat{\Pi}_{p_1p_2p_3}v\|^2_{H^1(\widehat{K})} \le C_{\mathrm{appx1}}\Bigg(&\Psi_{p_1,s_1}\sum_{\alpha_1,\alpha_2\le 1}\|\partial^{(s_1+1,\alpha_1,\alpha_2)}v\|^2_{L^2(\widehat{K})}\\ &+\Psi_{p_2,s_2}\sum_{\alpha_1,\alpha_2\le 1}\|\partial^{(\alpha_1,s_2+1,\alpha_2)}v\|^2_{L^2(\widehat{K})}\\ &+\Psi_{p_3,s_3}\sum_{\alpha_1,\alpha_2\le 1}\|\partial^{(\alpha_1,\alpha_2,s_3+1)}v\|^2_{L^2(\widehat{K})}\Bigg),\end{aligned} \tag{A.15}$$

for all $v\in H^{s_1+1}(I)\otimes H^{s_2+1}(I)\otimes H^{s_3+1}(I)$. Here, C_{appx1} is independent of (p_1,p_2,p_3), (s_1,s_2,s_3) and v.

Remark A.5. *In space dimension $d=2$, a result analogous to Lemma A.4 holds, see [43].*

Lemma A.6. *Let $d=3$, $(p_1,p_2,p_3)\in\mathbb{N}^3$, and $(s_1,s_2,s_3)\in\mathbb{N}^3$ with $1\le s_i\le p_i$. Further, let $\{i,j,k\}$ be a permutation of $\{1,2,3\}$. Then, the projector $\widehat{\Pi}_{p_1p_2p_3}: H^1_{\mathrm{mix}}(\widehat{K})\to\mathbb{Q}_{p_1,p_2,p_3}(\widehat{K})$ satisfies*

$$\begin{aligned}\|\partial_{x_i}\left(v-\widehat{\Pi}_{p_1p_2p_3}v\right)\|^2_{L^2(\widehat{K})} \le C_{\mathrm{appx2}}\Bigg(&\Psi_{p_i,s_i}\sum_{\alpha_1,\alpha_2\le 1}\|\partial^{s_i+1}_{x_i}\partial^{\alpha_1}_{x_j}\partial^{\alpha_2}_{x_k}v\|^2_{L^2(\widehat{K})}\\ &+\Psi_{p_j,s_j}\sum_{\alpha_1\le 1}\|\partial_{x_i}\partial^{s_j+1}_{x_j}\partial^{\alpha_1}_{x_k}v\|^2_{L^2(\widehat{K})}\\ &+\Psi_{p_k,s_k}\sum_{\alpha_1\le 1}\|\partial_{x_i}\partial^{\alpha_1}_{x_j}\partial^{s_k+1}_{x_k}v\|^2_{L^2(\widehat{K})}\Bigg),\end{aligned} \tag{A.16}$$

for all $v\in H^{s_1+1}(I)\otimes H^{s_2+1}(I)\otimes H^{s_3+1}(I)$. Here, $C_{\mathrm{appx2}}>0$ is independent of (p_1,p_2,p_3), (s_1,s_2,s_3), and v.

Proof. Let $(p_1,p_2,p_3)\in\mathbb{N}^3$, and $(s_1,s_2,s_3)\in\mathbb{N}^3$, be as in the statement of the lemma. Also, let $i\in\{1,2,3\}$ and $\{j,k\}=\{1,2,3\}\setminus\{i\}$. By Lemma A.4, there holds

$$\begin{aligned}\|\partial_{x_i}(v-\widehat{\Pi}_p v)\|^2_{L^2(\widehat{K})} \le C_{\mathrm{appx1}}\Bigg(&\Psi_{p_1,s_1}\sum_{\alpha_1,\alpha_2\le 1}\|\partial^{(s_1+1,\alpha_1,\alpha_2)}v\|^2_{L^2(\widehat{K})}\\ &+\Psi_{p_2,s_2}\sum_{\alpha_1,\alpha_2\le 1}\|\partial^{(\alpha_1,s_2+1,\alpha_2)}v\|^2_{L^2(\widehat{K})}\\ &+\Psi_{p_3,s_3}\sum_{\alpha_1,\alpha_2\le 1}\|\partial^{(\alpha_1,\alpha_2,s_3+1)}v\|^2_{L^2(\widehat{K})}\Bigg).\end{aligned} \tag{A.17}$$

With a $C_{\mathrm{appx1}}>0$ independent of (p_1,p_2,p_3), (s_1,s_2,s_3), and v. Let now, $\overline{v}_i: I^2\to\mathbb{R}$ such that

$$\overline{v}_i(x_j,x_k) = \int_I v(x_1,x_2,x_3)dx_i.$$

We denote $\tilde{v} := v - \overline{v}_i$ and, remarking that $\partial_{x_i}\overline{v}_i = \partial_{x_i}\widehat{\Pi}_p\overline{v}_i = 0$, we apply (A.17) to $\tilde{v}$, so that

$$\|\partial_{x_i}(v - \widehat{\Pi}_p v)\|^2_{L^2(\widehat{K})} \leq C\Bigg(\Psi_{p_1,s_1}\sum_{\alpha_1,\alpha_2\leq 1}\|\partial^{(s_1+1,\alpha_1,\alpha_2)}\tilde{v}\|^2_{L^2(\widehat{K})} + \Psi_{p_2,s_2}\sum_{\alpha_1,\alpha_2\leq 1}\|\partial^{(\alpha_1,s_2+1,\alpha_2)}\tilde{v}\|^2_{L^2(\widehat{K})} + \Psi_{p_3,s_3}\sum_{\alpha_1,\alpha_2\leq 1}\|\partial^{(\alpha_1,\alpha_2,s_3+1)}\tilde{v}\|^2_{L^2(\widehat{K})}\Bigg). \tag{A.18}$$

By the Poincaré inequality, it holds for all $\alpha_1 \in \{0,1\}$ that

$$\|\partial^{s_j+1}_{x_j}\partial^{\alpha_1}_{x_k}\tilde{v}\|^2_{L^2(\widehat{K})} \leq C\|\partial_{x_i}\partial^{s_j+1}_{x_j}\partial^{\alpha_1}_{x_k}v\|^2_{L^2(\widehat{K})} \quad\text{and}\quad \|\partial^{\alpha_1}_{x_j}\partial^{s_k+1}_{x_k}\tilde{v}\|^2_{L^2(\widehat{K})} \leq C\|\partial_{x_i}\partial^{\alpha_1}_{x_j}\partial^{s_k+1}_{x_k}v\|^2_{L^2(\widehat{K})}.$$

Using the fact that $\partial_{x_i}\tilde{v} = \partial_{x_i}v$ in the remaining terms of (A.18) concludes the proof. □

A.4.1 One dimensional estimate

The following result is a consequence of, e.g., [41, Lemma 8.1] and scaling.

Lemma A.7. *There exists $C > 0$ such that for all $\ell \in \mathbb{N}$, all integer $0 < k \leq \ell$, all integers $1 \leq s \leq p$, all $\gamma > 0$, and all $v \in H^{s+1}(J^\ell_k)$*

$$h^{-2}\|v - \pi^k_p v\|^2_{L^2(J^\ell_k)} + \|\nabla(v - \pi^k_p v)\|^2_{L^2(J^\ell_k)} \leq C\tau_\sigma^{2(s+1)}\Psi_{p,s}h^{2(\min\{\gamma-1,s\})}\||x|^{(s+1-\gamma)_+}v^{(s+1)}\|^2_{L^2(J^\ell_k)} \tag{A.19}$$

where $h = |J^\ell_k| \simeq \sigma^{\ell-k}$.

Proof. From [41, Lemma 8.1], there exists $C > 0$ independent of p, k, s, and v such that

$$h^{-2}\|v - \pi^k_p v\|^2_{L^2(J^\ell_k)} + \|\nabla(v - \pi^k_p v)\|^2_{L^2(J^\ell_k)} \leq C\Psi_{p,s}h^{2s}\|v^{(s+1)}\|^2_{L^2(J^\ell_k)}.$$

In addition, for all $k = 1, \dots, \ell$, there holds $x|_{J^\ell_k} \geq \frac{\sigma}{1-\sigma}h$. Hence, for all $\gamma < s+1$,

$$h^{2s}\|v^{(s+1)}\|^2_{L^2(J^\ell_k)} \leq \tau_\sigma^{2(s+1-\gamma)}h^{2\gamma-2}\|x^{s+1-\gamma}v^{(s+1)}\|^2_{L^2(J^\ell_k)}.$$

This concludes the proof. □

A.4.2 Estimate at a corner in dimension $d = 2$

We consider now a setting with a two dimensional corner singularity. Let $\beta \in \mathbb{R}$, $\mathfrak{K} = J^\ell_0 \times J^\ell_0$, $r(x) = |x - x_0|$ with $x_0 = (0,0)$, and define the corner-weighted norm $\|v\|_{\mathcal{J}^2_\beta(\mathfrak{K})}$ by

$$\|v\|^2_{\mathcal{J}^2_\beta(\mathfrak{K})} := \sum_{|\alpha|\leq 2}\|r^{(|\alpha|-\beta)_+}\partial^\alpha v\|^2_{L^2(\mathfrak{K})}.$$

Lemma A.8. *Let $d = 2$, $\beta \in (1,2)$. There exists $C_1, C_2 > 0$ such that for all $v \in \mathcal{J}^2_\beta(\mathfrak{K})$*

$$\sum_{\alpha\in\mathbb{N}^2_0:|\alpha|\leq 1}\|\partial^\alpha(\pi^0_1 \otimes \pi^0_1)v\|_{L^2(\mathfrak{K})} \leq C_1\left(\|v\|_{H^1(\mathfrak{K})} + \sum_{\alpha\in\mathbb{N}^2_0:|\alpha|=2}\sigma^{(\beta-1)\ell}\|r^{2-\beta}\partial^\alpha v\|_{L^2(\mathfrak{K})}\right). \tag{A.20}$$

and

$$\sum_{\alpha\in\mathbb{N}^2_0:|\alpha|\leq 1}\sigma^{-\ell(1-|\alpha|)}\|\partial^\alpha(v - (\pi^0_1 \otimes \pi^0_1)v)\|_{L^2(\mathfrak{K})} \leq C_2\sigma^{\ell(\beta-1)}\sum_{\alpha\in\mathbb{N}^2_0:|\alpha|=2}\|r^{2-\beta}\partial^\alpha v\|_{L^2(\mathfrak{K})}. \tag{A.21}$$

Proof. Denote by c_i, $i = 1, \dots, 4$ the corners of $\mathfrak{K}$ and by ψ_i, $i = 1, \dots, 4$ the bilinear functions such that $\psi_i(c_j) = \delta_{ij}$. Then,

$$(\pi^0_1 \otimes \pi^0_1)v = \sum_{i=1}^4 v(c_i)\psi_i.$$

Therefore, writing $h = \sigma^\ell$, we have

$$\|(\pi^0_1 \otimes \pi^0_1)v\|_{L^2(\mathfrak{K})} \leq \sum_{i=1,\dots,4}|v(c_i)|\|\psi_i\|_{L^2(\mathfrak{K})} \leq 4\|v\|_{L^\infty(\mathfrak{K})}|\mathfrak{K}|^{1/2} \leq 4h\|v\|_{L^\infty(\mathfrak{K})}. \tag{A.22}$$

With the imbedding $\mathcal{J}^2_\beta((0,1)^2) \hookrightarrow L^\infty((0,1)^2)$ which is valid for $\beta > 1$ (which follows e.g. from Lemma A.22 and $W^{1,1}_{\text{mix}}((0,1)^2) \hookrightarrow L^\infty((0,1)^2)$), a scaling argument gives

$$h^2\|v\|^2_{L^\infty(\mathfrak{K})} \leq Ch^2 \left(h^{-2}\|v\|^2_{L^2(\mathfrak{K})} + |v|^2_{H^1(\mathfrak{K})} + \sum_{|\alpha|=2} h^{2\beta-2}\|r^{2-\beta}\partial^\alpha v\|^2_{L^2(\mathfrak{K})} \right),$$

so that we obtain

$$\|(\pi_1^0 \otimes \pi_1^0)v\|^2_{L^2(\mathfrak{K})} \leq C \left(\|v\|^2_{L^2(\mathfrak{K})} + h^2|v|^2_{H^1(\mathfrak{K})} + \sum_{|\alpha|=2} h^{2\beta}\|r^{2-\beta}\partial^\alpha v\|^2_{L^2(\mathfrak{K})} \right). \tag{A.23}$$

For any $|\alpha| = 1$, denoting $v_0 = v(0,0)$ and using the fact that $(\pi_1^0 \otimes \pi_1^0)v_0 = v_0$ hence $\partial^\alpha(\pi_1^0 \otimes \pi_1^0)v_0 = 0$,

$$\|\partial^\alpha(\pi_1^0 \otimes \pi_1^0)v\|_{L^2(\mathfrak{K})} = \|\partial^\alpha(\pi_1^0 \otimes \pi_1^0)(v-v_0)\|_{L^2(\mathfrak{K})} \leq \sum_{i=1,\dots,4} |(v-v_0)(c_i)|\|\partial^\alpha\psi_i\|_{L^2(\mathfrak{K})} \leq C\|v-v_0\|_{L^\infty(\mathfrak{K})}. \tag{A.24}$$

With the imbedding $\mathcal{J}^2_\beta((0,1)^2) \hookrightarrow L^\infty((0,1)^2)$, Poincaré's inequality, and rescaling we obtain

$$\|\partial^\alpha(\pi_1^0 \otimes \pi_1^0)v\|^2_{L^2(\mathfrak{K})} \leq C \left(|v|^2_{H^1(\mathfrak{K})} + \sum_{|\alpha|=2} h^{2\beta-2}\|r^{2-\beta}\partial^\alpha v\|^2_{L^2(\mathfrak{K})} \right),$$

which finishes the proof of (A.20). To prove (A.21), note that by the Sobolev imbedding of $W^{2,1}(\mathfrak{K})$ into $H^1(\mathfrak{K})$ and by scaling, we have

$$\sum_{|\alpha|\leq 1} h^{|\alpha|-1}\|\partial^\alpha(v - (\pi_1^0 \otimes \pi_1^0)v)\|_{L^2(\mathfrak{K})} \leq C \sum_{|\alpha|\leq 2} h^{|\alpha|-2}\|\partial^\alpha(v - (\pi_1^0 \otimes \pi_1^0)v)\|_{L^1(\mathfrak{K})}.$$

By classical interpolation estimates [4, Theorem 4.4.4], we additionally conclude that

$$\sum_{|\alpha|\leq 1} h^{|\alpha|-2}\|\partial^\alpha(v - (\pi_1^0 \otimes \pi_1^0)v)\|_{L^1(\mathfrak{K})} \leq C|v|_{W^{2,1}(\mathfrak{K})}.$$

Using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\sum_{|\alpha|\leq 1} h^{|\alpha|-1}\|\partial^\alpha(v - (\pi_1^0 \otimes \pi_1^0)v)\|_{L^2(\mathfrak{K})} &\leq C \sum_{|\alpha|=2} \|\partial^\alpha v\|_{L^1(\mathfrak{K})} \\
&\leq C \sum_{|\alpha|=2} \|r^{-2+\beta}\|_{L^2(\mathfrak{K})}\|r^{2-\beta}\partial^\alpha v\|_{L^2(\mathfrak{K})} \\
&\leq C \sum_{|\alpha|=2} h^{\beta-1}\|r^{2-\beta}\partial^\alpha v\|_{L^2(\mathfrak{K})}
\end{aligned}$$

where we also have used, in the last step, the facts that $r(x) \leq \sqrt{2}h$ for all $x \in \mathfrak{K}$ and that $\beta > 1$. □

A.5 Interior estimates

The following lemmas give the estimate of the approximation error on the elements not belonging to edge or corner layers. For $d = 3$, all $\ell \in \mathbb{N}$, all $k_1, k_2, k_3 \in \{0, \dots, \ell\}$ and all $K = J^\ell_{k_1} \times J^\ell_{k_2} \times J^\ell_{k_3}$, we denote, by $h_\parallel$ the length of K in the direction parallel to the closest singular edge, and by $h_{\perp,1}$ and $h_{\perp,2}$ the lengths of K in the other two directions. If an element has multiple closest singular edges, we choose one of those and consider it as "closest edge" for all points in that element. When considering functions from $\mathcal{J}^d_\gamma(Q)$, γ_e will refer to the weight of this closest edge. Similarly, we denote by $\partial_\parallel$ (resp. $\partial_{\perp,1}$ and $\partial_{\perp,2}$) the derivatives in the direction parallel (resp. perpendicular) to the closest singular edge.

Lemma A.9. *Let $d = 3$, $\ell \in \mathbb{N}$ and $K = J^\ell_{k_1} \times J^\ell_{k_2} \times J^\ell_{k_3}$ for $0 < k_1, k_2, k_3 \leq \ell$. Let also $v \in \mathcal{J}^\varpi_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E}; C_v, A_v)$ with $\gamma_c \in (3/2, 5/2)$, $\gamma_e \in (1, 2)$. Then, there exists $C > 0$ dependent only on σ, C_{appx2}, C_v and $A > 0$ dependent only on σ, A_v such that for all $1 \leq s \leq p$*

$$\|\partial_\parallel(v - \Pi^K_p v)\|^2_{L^2(K)} \leq C\Psi_{p,s}A^{2s+6}\left((d^K_c)^2 + (d^K_c)^{2(\gamma_c-1)}\right)((s+3)!)^2, \tag{A.25}$$

where $\partial_\parallel$ is the derivative in the direction parallel to the closest singular edge.

Proof. We write $d_a = d_a^K$, $a \in \{c, e\}$. There holds

$$d_c^2 = \left(\frac{\sigma}{1-\sigma}\right)^2 (h_\parallel^2 + h_{\perp,1}^2 + h_{\perp,2}^2), \qquad d_e^2 = \left(\frac{\sigma}{1-\sigma}\right)^2 (h_{\perp,1}^2 + h_{\perp,2}^2).$$

Denoting $\hat{v} = v \circ \Phi_K^{-1}$ and $\widehat{\Pi}_p \hat{v} = \Pi_p^K v \circ \Phi_K^{-1} = \widehat{\Pi}_p (v \circ \Phi_K)$, using the result of Lemma A.6 and rescaling, we have

$$\begin{aligned}
\|\widehat{\partial}_\parallel (\hat{v} - \widehat{\Pi}_p \hat{v})\|_{L^2(\widehat{K})}^2 &\leq C_{\mathrm{appx2}} \Psi_{p,s} \frac{h_\parallel}{h_{\perp,1} h_{\perp,2}} \Bigg(\sum_{\alpha_1,\alpha_2 \leq 1} h_\parallel^{2s} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K)}^2 \\
&\quad + \sum_{\alpha_1 \leq 1} h_{\perp,1}^{2s+2} h_{\perp,2}^{2\alpha_1} \|\partial_\parallel \partial_{\perp,1}^{s+1} \partial_{\perp,2}^{\alpha_1} v\|_{L^2(K)}^2 \\
&\quad + \sum_{\alpha_1 \leq 1} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2s+2} \|\partial_\parallel \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{s+1} v\|_{L^2(K)}^2 \Bigg) \\
&= C_{\mathrm{appx2}} \Psi_{p,s} \frac{h_\parallel}{h_{\perp,1} h_{\perp,2}} \Big((I) + (II) + (III) \Big).
\end{aligned} \tag{A.26}$$

Denote $K_c = K \cap Q_c$, $K_e = K \cap Q_e$, $K_{ce} = K \cap Q_{ce}$, and $K_0 = K \cap Q_0$. Furthermore, we indicate

$$(I)_c = \sum_{\alpha_1,\alpha_2 \leq 1} h_\parallel^{2s} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_c)}^2,$$

and do similarly for the other terms of the sum (II) and (III) and the other subscripts e, ce, 0. Remark also that $r_{i|_K} \geq d_i$, $i \in \{c, e\}$, and that for $a, b \in \mathbb{R}$ holds $r_c^a r_e^b = r_c^{a+b} \rho_{ce}^b$.

We will also write $\widetilde{\gamma} = \gamma_c - \gamma_e$. We start by considering the term $(I)_{ce}$. Let $\alpha_1 = \alpha_2 = 1$; then,

$$\begin{aligned}
h_\parallel^{2s} h_{\perp,1}^2 h_{\perp,2}^2 \|\partial_\parallel^{s+1} \partial_{\perp,1} \partial_{\perp,2} v\|_{L^2(K_{ce})}^2 &\leq \tau_\sigma^{2s+4} d_c^{2s} d_e^4 \|\partial_\parallel^{s+1} \partial_{\perp,1} \partial_{\perp,2} v\|_{L^2(K_{ce})}^2 \\
&\leq \tau_\sigma^{2s+4} d_c^{2\widetilde{\gamma}-2} d_e^{2\gamma_e} \|r_c^{s+3-\gamma_c} \rho_{ce}^{2-\gamma_e} \partial_\parallel^{s+1} \partial_{\perp,1} \partial_{\perp,2} v\|_{L^2(K_{ce})}^2 ,
\end{aligned}$$

where τ_σ is as in (A.6). Furthermore, if $\alpha_1 + \alpha_2 \leq 1$ and $s + 1 + \alpha_1 + \alpha_2 - \gamma_c \geq 0$,

$$\begin{aligned}
h_\parallel^{2s} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_{ce})}^2 &\leq \tau_\sigma^{2s+2(\alpha_1+\alpha_2)} d_c^{2s} d_e^{2(\alpha_1+\alpha_2)} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_{ce})}^2 \\
&\leq \tau_\sigma^{2s+2(\alpha_1+\alpha_2)} d_c^{2\gamma_c-2} \|r_c^{s+1+\alpha_1+\alpha_2-\gamma_c} \partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_{ce})}^2,
\end{aligned}$$

where we have also used $d_e \leq d_c$. Therefore,

$$(I)_{ce} \leq \tau_\sigma^{2s+4} d_c^{2\gamma_c-2} \sum_{\alpha_1,\alpha_2 \leq 1} \|r_c^{s+1+\alpha_1+\alpha_2-\gamma_c} \rho_{ce}^{(\alpha_1+\alpha_2-\gamma_e)_+} \partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_{ce})}^2.$$

If $s + 1 + \alpha_1 + \alpha_2 - \gamma_c < 0$, then $s = 1$ and $\alpha_1 = \alpha_2 = 0$, thus

$$(I)_{ce} \leq \tau_\sigma^{2s+4} d_c^2 \|r_c^{(s+1+\alpha_1+\alpha_2-\gamma_c)_+} \rho_{ce}^{(\alpha_1+\alpha_2-\gamma_e)_+} \partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_{ce})}^2.$$

Then, if $s + 1 + \alpha_1 + \alpha_2 - \gamma_c \geq 0$

$$\begin{aligned}
(I)_c &= \sum_{\alpha_1,\alpha_2 \leq 1} h_\parallel^{2s} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_c)}^2 \\
&\leq \tau_\sigma^{2s+4} \sum_{\alpha_1,\alpha_2 \leq 1} d_c^{2s} d_e^{2(\alpha_1+\alpha_2)} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_c)}^2 \\
&\leq \tau_\sigma^{2s+4} d_c^{2\gamma_c-2} \sum_{\alpha_1,\alpha_2 \leq 1} \|r_c^{(s+1+\alpha_1+\alpha_2-\gamma_c)_+} \partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_c)}^2
\end{aligned}$$

where the last inequality follows also from $d_e \leq d_c$. If $s + 1 + \alpha_1 + \alpha_2 - \gamma_c < 0$, then the same bound holds with $d_c^{2\gamma_c-2}$ replaced by d_c^2. Similarly,

$$\begin{aligned}
(I)_e &= \sum_{\alpha_1,\alpha_2 \leq 1} h_\parallel^{2s} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_e)}^2 \\
&\leq \tau_\sigma^{2s+4} \sum_{\alpha_1,\alpha_2 \leq 1} d_c^{2s} d_e^{2\alpha_1+2\alpha_2-2(\alpha_1+\alpha_2-\gamma_e)_+} \|r_e^{(\alpha_1+\alpha_2-\gamma_e)_+} \partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_e)}^2 \\
&\leq \tau_\sigma^{2s+4} d_c^{2s} \sum_{\alpha_1,\alpha_2 \leq 1} \|r_e^{(\alpha_1+\alpha_2-\gamma_e)_+} \partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|_{L^2(K_e)}^2,
\end{aligned}$$

where we used that $d_e \leq 1$. The bound on $(I)_0$ follows directly from the definition:

$$(I)_0 = \sum_{\alpha_1,\alpha_2\leq 1} h_\parallel^{2s} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1}\partial_{\perp,1}^{\alpha_1}\partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K_0)} \leq \tau_\sigma^{2s+4} d_c^{2s} \sum_{\alpha_1,\alpha_2\leq 1} \|\partial_\parallel^{s+1}\partial_{\perp,1}^{\alpha_1}\partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K_0)}.$$

Using (2.1), there exists $C > 0$ dependent only on C_v and σ and $A > 0$ dependent only on A_v and σ such that

$$(I) \leq CA^{2s+6}((s+3)!)^2 \left(d_c^2 + d_c^{2\gamma_c - 2}\right). \tag{A.27}$$

We then apply the same argument to the terms (II) and (III). Indeed,

$$\begin{aligned}
(II)_{ce} &= \sum_{\alpha_1\leq 1} h_{\perp,1}^{2s+2} h_{\perp,2}^{2\alpha_1} \|\partial_\parallel \partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K_{ce})} \\
&\leq \tau_\sigma^{2s+4} \sum_{\alpha_1\leq 1} d_e^{2s+2+2\alpha_1} \|\partial_\parallel \partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K_{ce})} \\
&\leq \tau_\sigma^{2s+4} \sum_{\alpha_1\leq 1} d_c^{2\widetilde{\gamma}-2} d_e^{2\gamma_e} \|r_c^{s+2+\alpha_1-\gamma_c} \rho_{ce}^{s+1+\alpha_1-\gamma_e} \partial_\parallel \partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K_{ce})}
\end{aligned}$$

and the estimate for $(III)_{ce}$ follows by exchanging $h_{\perp,1}$ and $\partial_{\perp,1}$ with $h_{\perp,2}$ and $\partial_{\perp,2}$ in the inequality above. The estimates for $(II)_{c,e,0}$ and $(III)_{c,e,0}$ can be obtained as for $(I)_{c,e,0}$:

$$\begin{aligned}
(II)_c &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1\leq 1} d_c^{2\gamma_c-2} \|r_c^{s+2+\alpha_1-\gamma_c} \partial_\parallel \partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K_c)}, \\
(II)_e &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1\leq 1} d_e^{2\gamma_e} \|r_e^{s+1+\alpha_1-\gamma_e} \partial_\parallel \partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K_e)}, \\
(II)_0 &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1\leq 1} d_e^{2s+2} \|\partial_\parallel \partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K_0)}.
\end{aligned}$$

Therefore, we have

$$(II), (III) \leq CA^{2s+6}(d_c^2 + d_c^{2\gamma_c-2})((s+3)!)^2. \tag{A.28}$$

We obtain, from (A.26), (A.27), and (A.28) that there exists $C > 0$ (dependent only on σ, C_{appx2}, C_v and $A > 0$ (dependent only on σ, A_v) such that

$$\|\widehat{\partial}_\parallel(\hat{v} - \widehat{\Pi}_p \hat{v})\|^2_{L^2(\widehat{K})} \leq C \frac{h_\parallel}{h_{\perp,1}h_{\perp,2}} \Psi_{p,s} A^{2s+6}(d_c^2 + d_c^{2\gamma_c-2})((s+3)!)^2.$$

Considering that

$$\|\partial_\parallel(v - \Pi_p v)\|^2_{L^2(K)} \leq \frac{h_{\perp,1}h_{\perp,2}}{h_\parallel} \|\widehat{\partial}_\parallel(\hat{v} - \widehat{\Pi}_p \hat{v})\|^2_{L^2(\widehat{K})}$$

completes the proof. □

Lemma A.10. *Let $d = 3$, $\ell \in \mathbb{N}$ and $K = J_{k_1}^\ell \times J_{k_2}^\ell \times J_{k_3}^\ell$ for $0 < k_1, k_2, k_3 \leq \ell$. Let also $v \in \mathcal{J}_{\underline{\gamma}}^{\varpi}(Q; \mathcal{C}, \mathcal{E}; C_v, A_v)$ with $\gamma_c \in (3/2, 5/2)$, $\gamma_e \in (1, 2)$. Then, there exists $C > 0$ dependent only on σ, C_{appx2}, C_v and $A > 0$ dependent only on σ, A_v such that for all $p \in \mathbb{N}$ and all $1 \leq s \leq p$*

$$\begin{aligned}
&\|\partial_{\perp,1}(v - \Pi_p^K v)\|^2_{L^2(K)} + \|\partial_{\perp,2}(v - \Pi_p^K v)\|^2_{L^2(K)} \\
&\qquad\qquad \leq C\Psi_{p,s} A^{2s+6} \left((d_c^K)^{2(\gamma_c-1)} + (d_c^K)^{2(\gamma_e-1)}\right) ((s+3)!)^2,
\end{aligned} \tag{A.29}$$

where $\partial_{\perp,1}$, $\partial_{\perp,2}$ are the derivatives in the directions perpendicular to the closest singular edge.

Proof. The proof follows closely that of Lemma A.9 and we use the same notation. From Lemma A.6 and rescaling, we have

$$\begin{aligned}
\|\widehat{\partial}_{\perp,1}(\hat{v} - \widehat{\Pi}_p \hat{v})\|^2_{L^2(\widehat{K})} &\leq C_{\mathrm{appx2}} \Psi_{p,s} \frac{h_{\perp,1}}{h_\parallel h_{\perp,2}} \Bigg(\sum_{\alpha_1\leq 1} h_\parallel^{2s+2} h_{\perp,2}^{2\alpha_1} \|\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K)} \\
&\qquad + \sum_{\alpha_1,\alpha_2\leq 1} h_\parallel^{2\alpha_1} h_{\perp,1}^{2s} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K)} \\
&\qquad + \sum_{\alpha_1\leq 1} h_\parallel^{2\alpha_1} h_{\perp,2}^{2s+2} \|\partial_\parallel^{\alpha_1}\partial_{\perp,1}\partial_{\perp,2}^{s+1} v\|^2_{L^2(K)} \Bigg) \\
&= C_{\mathrm{appx2}} \Psi_{p,s} \frac{h_{\perp,1}}{h_\parallel h_{\perp,2}} \Big((I) + (II) + (III) \Big).
\end{aligned} \tag{A.30}$$

As before, we will write $\widetilde{\gamma} = \gamma_c - \gamma_e$. We start by considering the term $(I)_{ce}$. When $\alpha_1 = 1$,

$$\begin{aligned} h_\parallel^{2s+2} h_{\perp,2}^2 \|\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}v\|^2_{L^2(K_{ce})} &\leq \tau_\sigma^{2s+4} d_c^{2s+2} d_e^2 \|\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+4} d_c^{2\widetilde{\gamma}} d_e^{2\gamma_e-2} \|r_c^{s+3-\gamma_c}\rho_{ce}^{2-\gamma_e}\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}v\|^2_{L^2(K_{ce})}, \end{aligned}$$

where $d_c^{2\widetilde{\gamma}} d_e^{2\gamma_e-2} \leq d_c^{2\gamma_c-2}$. Furthermore, if $\alpha_1 = 0$,

$$\begin{aligned} h_\parallel^{2s+2}\|\partial_\parallel^{s+1}\partial_{\perp,1}v\|^2_{L^2(K_{ce})} &\leq \tau_\sigma^{2s+2} d_c^{2s+2}\|\partial_\parallel^{s+1}\partial_{\perp,1}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+2} d_c^{2\gamma_c-2}\|r_c^{s+2-\gamma_c}\partial_\parallel^{s+1}\partial_{\perp,1}v\|^2_{L^2(K_{ce})}. \end{aligned}$$

Therefore,

$$(I)_{ce} \leq \left(\frac{1-\sigma}{\sigma}\right)^{2s+4} d_c^{2\gamma_c-2} \sum_{\alpha_1\leq 1} \|r_c^{s+2+\alpha_1-\gamma_c}\rho_{ce}^{(1+\alpha_1-\gamma_e)_+}\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}^{\alpha_1}v\|^2_{L^2(K_{ce})}.$$

The estimates for $(I)_{c,e,0}$ follow from the same technique:

$$\begin{aligned} (I)_e &\leq \sum_{\alpha_1\leq 1} \tau_\sigma^{2s+4} d_c^{2s+2}\|r_e^{(1+\alpha_1-\gamma_e)_+}\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}^{\alpha_1}v\|^2_{L^2(K_e)}, \\ (I)_c &\leq \sum_{\alpha_1\leq 1} \tau_\sigma^{2s+4} d_c^{2\gamma_c-2}\|r_c^{s+2+\alpha_1-\gamma_c}\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}^{\alpha_1}v\|^2_{L^2(K_c)}, \\ (I)_0 &\leq \sum_{\alpha_1\leq 1} \tau_\sigma^{2s+4} d_c^{2s+2}\|\partial_\parallel^{s+1}\partial_{\perp,1}\partial_{\perp,2}^{\alpha_1}v\|^2_{L^2(K_0)}. \end{aligned}$$

Hence, from (2.1), there exists $C > 0$ dependent only on C_v and σ and $A > 0$ dependent only on A_v and σ such that

$$(I) \leq CA^{2s+6}((s+3)!)^2 d_c^{2\gamma_c-2}. \tag{A.31}$$

We then apply the same argument to the terms (II) and (III). Indeed, if $s+1+\alpha_1+\alpha_2-\gamma_c \geq 0$

$$\begin{aligned} (II)_{ce} &= \sum_{\alpha_1,\alpha_2\leq 1} h_\parallel^{2\alpha_1} h_{\perp,1}^{2s} h_{\perp,2}^{2\alpha_2}\|\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+4}\sum_{\alpha_1,\alpha_2\leq 1} d_c^{2\alpha_1} d_e^{2s+2\alpha_2}\|\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+4}\sum_{\alpha_1\leq 1} d_c^{2\widetilde{\gamma}} d_e^{2\gamma_e-2}\|r_c^{s+1+\alpha_1+\alpha_2-\gamma_c}\rho_{ce}^{s+1+\alpha_2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+4}\sum_{\alpha_1\leq 1} d_c^{2\gamma_c-2}\|r_c^{s+1+\alpha_1+\alpha_2-\gamma_c}\rho_{ce}^{s+1+\alpha_2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})}, \end{aligned}$$

where at the last step we have used that $\gamma_e > 1$ and $d_e \leq d_c$. If $s+1+\alpha_1+\alpha_2-\gamma_c < 0$, then

$$\begin{aligned} (II)_{ce} &= \sum_{\alpha_1,\alpha_2\leq 1} h_\parallel^{2\alpha_1} h_{\perp,1}^{2s} h_{\perp,2}^{2\alpha_2}\|\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+4}\sum_{\alpha_1,\alpha_2\leq 1} d_c^{2\alpha_1} d_e^{2s+2\alpha_2}\|\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+4}\sum_{\alpha_1\leq 1} d_c^{2\alpha_1} d_e^{2s+2\alpha_2}(d_e/d_c)^{-2s-2-2\alpha_2+2\gamma_e}\|\rho_{ce}^{s+1+\alpha_2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})} \\ &\leq \tau_\sigma^{2s+4}\sum_{\alpha_1\leq 1} d_c^{2s+2-2\gamma_e} d_e^{2\gamma_e-2}\|\rho_{ce}^{s+1+\alpha_2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})}. \end{aligned}$$

Thus, using $d_e \leq d_c$,

$$(II)_{ce} \leq \tau_\sigma^{2s+4}\sum_{\alpha_1\leq 1}(d_c^{2s}+d_c^{2\gamma_c-2})\|r_c^{(s+1+\alpha_1+\alpha_2-\gamma_c)_+}\rho_{ce}^{s+1+\alpha_2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_{ce})}.$$

The estimates for $(II)_{c,e,0}$ and $(III)_{ce,c,e,0}$ can be obtained as above:

$$(II)_e \leq \tau_\sigma^{2s+4}\sum_{\alpha_1\leq 1} d_e^{2\gamma_e-2}\|r_e^{s+1+\alpha_2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2}v\|^2_{L^2(K_e)},$$

if $s+1+\alpha_1+\alpha_2-\gamma_c \geq 0$, then

$$(II)_c \leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} d_c^{2\gamma_c-2} \| r_c^{s+1+\alpha_1+\alpha_2-\gamma_c} \partial_\parallel^{\alpha_1} \partial_{\perp,1}^{s+1} \partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K_c)},$$

if $s+1+\alpha_1+\alpha_2-\gamma_c < 0$, then

$$(II)_c \leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} d_c^{2s} \| \partial_\parallel^{\alpha_1} \partial_{\perp,1}^{s+1} \partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K_c)},$$

so that

$$\begin{aligned}
(II)_c &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} (d_c^{2s} + d_c^{2\gamma_c-2}) \| \partial_\parallel^{\alpha_1} \partial_{\perp,1}^{s+1} \partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K_c)},\\
(II)_0 &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} d_c^{2s} \| \partial_\parallel^{\alpha_1} \partial_{\perp,1}^{s+1} \partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K_0)},\\
(III)_{ce} &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} d_c^{2\gamma_c-2} \| r_c^{s+2+\alpha_1-\gamma_c} \rho_{ce}^{s+2-\gamma_e} \partial_\parallel^{\alpha_1} \partial_{\perp,1} \partial_{\perp,2}^{s+1} v\|^2_{L^2(K_{ce})},\\
(III)_e &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} d_e^{2\gamma_e-2} \| r_e^{s+2-\gamma_e} \partial_\parallel^{\alpha_1} \partial_{\perp,1} \partial_{\perp,2}^{s+1} v\|^2_{L^2(K_e)},\\
(III)_c &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} d_c^{2\gamma_c-2} \| r_c^{s+2+\alpha_1-\gamma_c} \partial_\parallel^{\alpha_1} \partial_{\perp,1} \partial_{\perp,2}^{s+1} v\|^2_{L^2(K_c)},\\
(III)_0 &\leq \tau_\sigma^{2s+4} \sum_{\alpha_1 \leq 1} d_e^{2s+2} \| \partial_\parallel^{\alpha_1} \partial_{\perp,1} \partial_{\perp,2}^{s+1} v\|^2_{L^2(K_0)}.
\end{aligned}$$

Therefore, we have

$$(II) + (III) \leq CA^{2s+6} (d_c^{2\gamma_c-2} + d_c^{2\gamma_e-2}) ((s+3)!)^2. \tag{A.32}$$

We obtain, from (A.30), (A.31), and (A.32) that there exists $C > 0$ dependent only on σ, C_{appx2}, C_v and $A > 0$ dependent only on σ, A_v such that

$$\|\widehat{\partial}_{\perp,1}(\hat{v} - \widehat{\Pi}_p \hat{v})\|^2_{L^2(\widehat{K})} \leq C \frac{h_{\perp,1}}{h_\parallel h_{\perp,2}} \Psi_{p,s} A^{2s+6} \left(d_c^{2(\gamma_c-1)} + d_c^{2(\gamma_e-1)} \right) ((s+3)!)^2.$$

Considering that

$$\|\partial_{\perp,1}(v - \Pi_p v)\|^2_{L^2(K)} \leq \frac{h_\parallel h_{\perp,2}}{h_{\perp,1}} \|\widehat{\partial}_{\perp,1}(\hat{v} - \widehat{\Pi}_p \hat{v})\|^2_{L^2(\widehat{K})}$$

and considering that the estimate for the other term at the left-hand side of (A.29) is obtained by exchanging $\{h, \partial\}_{\perp,1}$ with $\{h, \partial\}_{\perp,2}$ completes the proof. □

Lemma A.11. *Let $d = 3$, $\ell \in \mathbb{N}$ and $K = J^\ell_{k_1} \times J^\ell_{k_2} \times J^\ell_{k_3}$ for $0 < k_1, k_2, k_3 \leq \ell$. Let also $v \in \mathcal{J}^\varpi_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E}; C_v, A_v)$ with $\gamma_c \in (3/2, 5/2)$, $\gamma_e \in (1, 2)$. Then, there exists $C > 0$ dependent only on σ, C_{appx1}, C_v and $A > 0$ dependent only on σ, A_v such that for all $p \in \mathbb{N}$ and all $1 \leq s \leq p$*

$$\|v - \Pi_p^K v\|^2_{L^2(K)} \leq C \Psi_{p,s} A^{2s+6} \left(d_c^{2(\gamma_c-1)} + d_c^{2(\gamma_e-1)} \right) ((s+3)!)^2. \tag{A.33}$$

Proof. The proof follows closely that of Lemmas A.9 and A.10; we use the same notation. From Lemma A.4 and rescaling, we have

$$\begin{aligned}
\|\hat{v} - \widehat{\Pi}_p \hat{v}\|^2_{L^2(\widehat{K})} \leq C_{\mathrm{appx1}} \Psi_{p,s} \frac{1}{h_\parallel h_{\perp,1} h_{\perp,2}} \Bigg(&\sum_{\alpha_1,\alpha_2 \leq 1} h_\parallel^{2s+2} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1} \partial_{\perp,1}^{\alpha_1} \partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K)} \\
&+ \sum_{\alpha_1,\alpha_2 \leq 1} h_\parallel^{2\alpha_1} h_{\perp,1}^{2s+2} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{\alpha_1} \partial_{\perp,1}^{s+1} \partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K)} \\
&+ \sum_{\alpha_1,\alpha_2 \leq 1} h_\parallel^{2\alpha_1} h_{\perp,1}^{2\alpha_2} h_{\perp,2}^{2s+2} \|\partial_\parallel^{\alpha_1} \partial_{\perp,1}^{\alpha_2} \partial_{\perp,2}^{s+1} v\|^2_{L^2(K)} \Bigg).
\end{aligned} \tag{A.34}$$

Most terms at the right-hand side above have already been considered in the proofs of Lemmas A.9 and A.10, and the terms with $\alpha_1 = \alpha_2 = 0$ can be estimated similarly; the observation that

$$\|v - \Pi_p v\|^2_{L^2(K)} \leq h_\parallel h_{\perp,1} h_{\perp,2} \|\hat{v} - \widehat{\Pi}_p \hat{v}\|^2_{L^2(\widehat{K})}$$

concludes the proof. □

We summarize Lemmas A.9 to A.11 in the following result.

Lemma A.12. *Let $d = 3$, $\ell \in \mathbb{N}$ and $K = J^{\ell}_{k_1} \times J^{\ell}_{k_2} \times J^{\ell}_{k_3}$ such that $0 < k_1, k_2, k_3 \leq \ell$. Let also $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E}; C_v, A_v)$ with $\gamma_c \in (3/2, 5/2)$, $\gamma_e \in (1, 2)$. Then, there exists $C > 0$ dependent only on σ, C_{appx1}, C_{appx2}, C_v and $A > 0$ dependent only on σ, A_v such that for all $p \in \mathbb{N}$ and all $1 \leq s \leq p$*

$$\|v - \Pi^K_p v\|^2_{H^1(K)} \leq C\Psi_{p,s} A^{2s+6}\left(d_c^{2(\gamma_c-1)} + d_c^{2(\gamma_e-1)}\right)((s+3)!)^2. \tag{A.35}$$

We then consider elements on the faces (but not abutting edges) of Q.

Lemma A.13. *Let $d = 3$, $\ell \in \mathbb{N}$ and $K = J^{\ell}_{k_1} \times J^{\ell}_{k_2} \times J^{\ell}_{k_3}$ such that $k_j = 0$ for one $j \in \{1, 2, 3\}$ and $0 < k_i \leq \ell$ for $i \neq j$. For all $p \in \mathbb{N}$ and all $1 \leq s \leq p$, let $p_j = 1$ and $p_i = p \in \mathbb{N}$ for $i \neq j$. Let also $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E}; C_v, A_v)$ with $\gamma_c \in (3/2, 5/2)$, $\gamma_e \in (1, 2)$. Then, there exists $C > 0$ dependent only on σ, C_{appx1}, C_{appx2}, C_v and $A > 0$ dependent only on σ, A_v such that*

$$\|v - \Pi^K_{p_1p_2p_3} v\|^2_{H^1(K)} C\left(\Psi_{p,s} A^{2s+6}(d^K_c)^{2(\min(\gamma_c,\gamma_e)-1)}((s+3)!)^2 + (d^K_e)^{2(\min(\gamma_c,\gamma_e)-2)}\sigma^{2\ell}A^8\right). \tag{A.36}$$

Proof. We write $d_a = d^K_a$, $a \in \{c, e\}$. Suppose, for ease of notation, that $j = 3$, i.e. $k_3 = 0$. The projector is then given by $\Pi^K_{pp1} = \pi^{k_1}_p \otimes \pi^{k_2}_p \otimes \pi^0_1$. Also, we denote $h_{\perp,2} = \sigma^\ell$ and $\partial_{\perp,2} = \partial_{x_3}$. By (A.16),

$$\begin{aligned}
\|\partial_\parallel(v - \Pi^K_{pp1}v)\|^2_{L^2(K)} \leq C_{\mathrm{appx2}} \Bigg(&\Psi_{p,s}\Bigg(\sum_{\alpha_1,\alpha_2\leq 1} h_\parallel^{2s} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{2\alpha_2} \|\partial_\parallel^{s+1}\partial_{\perp,1}^{\alpha_1}\partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K)} \\
&+ \sum_{\alpha_1\leq 1} h_{\perp,1}^{2s+2} h_{\perp,2}^{2\alpha_1} \|\partial_\parallel \partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K)}\Bigg) \\
&+ \sum_{\alpha_1\leq 1} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{4} \|\partial_\parallel \partial_{\perp,1}^{\alpha_1}\partial_{\perp,2}^{2} v\|^2_{L^2(K)}\Bigg) \\
= C_{\mathrm{appx2}}\Big((I) + (II) + (III)\Big)&.
\end{aligned}$$

The bounds on the terms (I) and (II) can be derived as in Lemma A.9, and give

$$(I) + (II) \leq C\Psi_{p,s}A^{2s+6}\left((d^K_c)^2 + (d^K_c)^{2(\gamma_c-1)}\right)((s+3)!)^2.$$

We consider then term (III): with the usual notation, writing $\widetilde{\gamma} = \gamma_c - \gamma_e$,

$$\begin{aligned}
(III)_{ce} &= \sum_{\alpha_1\leq 1} h_{\perp,1}^{2\alpha_1} h_{\perp,2}^{4} \|\partial_\parallel \partial_{\perp,1}^{\alpha_1}\partial_{\perp,2}^{2} v\|^2_{L^2(K_{ce})} \\
&\leq \sum_{\alpha_1\leq 1} \tau_\sigma^{4+2\alpha_1} d_c^{2\widetilde{\gamma}-2} d_e^{2\gamma_e-4}\sigma^{4\ell} \|r_c^{3+\alpha_1-\gamma_c}\rho_{ce}^{2+\alpha_1-\gamma_e}\partial_\parallel \partial_{\perp,1}^{\alpha_1}\partial_{\perp,2}^{2} v\|^2_{L^2(K_{ce})} \\
&\leq C\tau_\sigma^6 d_c^{2\widetilde{\gamma}-2} d_e^{2\gamma_e-4}\sigma^{4\ell}A^8.
\end{aligned} \tag{A.37}$$

Note that $d_c \geq d_e$ and

$$d_c^{\widetilde{\gamma}} d_e^{\gamma_e} \leq \begin{cases} 1^{\widetilde{\gamma}} d_e^{\gamma_e} & \text{if } \widetilde{\gamma} \geq 0 \\ d_e^{\widetilde{\gamma}} d_e^{\gamma_e} & \text{if } \widetilde{\gamma} \geq 0 \end{cases} \leq d_e^{\min(\gamma_c,\gamma_e)}, \tag{A.38}$$

where we have also used that $d_c \leq 1$. Hence,

$$(III)_{ce} \leq C\tau_\sigma^6 d_e^{2\min(\gamma_e,\gamma_c)-6}\sigma^{4\ell}A^8 \leq C\tau_\sigma^6 d_e^{2\min(\gamma_e,\gamma_c)-4}\sigma^{2\ell}A^8. \tag{A.39}$$

The bounds on the terms $(III)_{c,e,0}$ follow by the same argument:

$$\begin{aligned}
(III)_e &\leq C\tau_\sigma^6 d_e^{2\gamma_e-4}\sigma^{4\ell}A^8, \\
(III)_c &\leq C\tau_\sigma^6 d_c^{2\gamma_c-6}\sigma^{4\ell}A^8 \leq C\tau_\sigma^6 d_e^{2\gamma_c-4}\sigma^{2\ell}A^8, \\
(III)_0 &\leq C\tau_\sigma^6\sigma^{4\ell}A^8.
\end{aligned}$$

Then,

$$\begin{aligned}
\|\partial_{\perp,1}(v-\Pi^K_{pp1}v)\|^2_{L^2(K)} \leq C_{\text{appx2}} &\Bigg(\frac{(p-s)!}{(p+s)!}\Bigg(\sum_{\alpha_1\leq 1} h_{\|}^{2s+2} h_{\perp,2}^{2\alpha_1} \|\partial_{\|}^{s+1}\partial_{\perp,1}\partial_{\perp,2}^{\alpha_1} v\|^2_{L^2(K)} \\
&\quad + \sum_{\alpha_1,\alpha_2\leq 1} h_{\|}^{2\alpha_1} h_{\perp,1}^{2s} h_{\perp,2}^{2\alpha_2} \|\partial_{\|}^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2}^{\alpha_2} v\|^2_{L^2(K)}\Bigg) \\
&\quad + \sum_{\alpha_1\leq 1} h_{\|}^{2\alpha_1} h_{\perp,2}^{4} \|\partial_{\|}^{\alpha_1}\partial_{\perp,1}\partial_{\perp,2}^{2} v\|^2_{L^2(K)}\Bigg) \\
&\leq C_{\text{appx2}}\Big((I)+(II)+(III)\Big).
\end{aligned}$$

The bounds on the first two terms at the right-hand side above can be obtained as in Lemma A.10:

$$(I)+(II) \leq C\Psi_{p,s}A^{2s+6}\left((d_c^K)^{2(\gamma_c-1)} + (d_c^K)^{2(\gamma_e-1)}\right)((s+3)!)^2,$$

while the last term can be bounded as in (A.39),

$$\begin{aligned}
(III)_{ce} &\leq \tau_\sigma^6 d_c^{2\widetilde{\gamma}} d_e^{2\gamma_e-6}\sigma^{4\ell}A^8 \leq C\tau_\sigma^6 d_e^{2\min(\gamma_c,\gamma_e)-4}\sigma^{2\ell}A^8,\\
(III)_{e} &\leq \tau_\sigma^6 d_e^{2\gamma_e-6}\sigma^{4\ell}A^8 \leq C\tau_\sigma^6 d_e^{2\gamma_e-4}\sigma^{2\ell}A^8,\\
(III)_{c} &\leq \tau_\sigma^6 d_c^{2\gamma_c-6}\sigma^{4\ell}A^8 \leq C\tau_\sigma^6 d_e^{2\gamma_c-4}\sigma^{2\ell}A^8,\\
(III)_{0} &\leq \tau_\sigma^6 \sigma^{4\ell}A^8,
\end{aligned}$$

so that

$$\sum_{\alpha_1\leq 1} h_{\|}^{2\alpha_1} h_{\perp,2}^{4} \|\partial_{\|}^{\alpha_1}\partial_{\perp,1}\partial_{\perp,2}^{2} v\|^2_{L^2(K)} \leq C d_e^{2\min(\gamma_c,\gamma_e)-4}\sigma^{2\ell}A^8.$$

The same holds true for the last term of the gradient of the approximation error, given by

$$\begin{aligned}
\|\partial_{\perp,2}(v-\Pi^K_{pp1}v)\|^2_{L^2(K)} \leq C_{\text{appx2}} &\Bigg(\Psi_{p,s}\Bigg(\sum_{\alpha_1\leq 1} h_{\|}^{2s+2} h_{\perp,1}^{2\alpha_1} \|\partial_{\|}^{s+1}\partial_{\perp,1}^{\alpha_1}\partial_{\perp,2} v\|^2_{L^2(K)} \\
&\quad + \sum_{\alpha_1\leq 1} h_{\|}^{2\alpha_1} h_{\perp,1}^{2s+2} \|\partial_{\|}^{\alpha_1}\partial_{\perp,1}^{s+1}\partial_{\perp,2} v\|^2_{L^2(K)}\Bigg) \\
&\quad + \sum_{\alpha_1,\alpha_2\leq 1} h_{\|}^{2\alpha_1} h_{\perp,1}^{2\alpha_2} h_{\perp,2}^{2} \|\partial_{\|}^{\alpha_1}\partial_{\perp,1}^{\alpha_2}\partial_{\perp,2}^{2} v\|^2_{L^2(K)}\Bigg) \\
&\leq C_{\text{appx2}}\Big((I)+(II)+(III)\Big).
\end{aligned}$$

From Lemma A.10, we obtain

$$(I)+(II) \leq C\Psi_{p,s}A^{2s+6}\left((d_c^K)^{2(\gamma_c-1)} + (d_c^K)^{2(\gamma_e-1)}\right)((s+3)!)^2,$$

whereas for the third term, it holds that if $\alpha_1+\alpha_2+2-\gamma_c\geq 0$

$$(III)_{ce} \leq \tau_\sigma^6 d_c^{2\widetilde{\gamma}} d_e^{2\gamma_e-4}\sigma^{2\ell}A^8 \leq C\tau_\sigma^6 d_e^{2\min(\gamma_c,\gamma_e)-4}\sigma^{2\ell}A^8, \quad (III)_c \leq \tau_\sigma^6 d_c^{2\gamma_c-4}\sigma^{2\ell}A^8,$$

and if $\alpha_1+\alpha_2+2-\gamma_c<0$, then

$$(III)_{ce} \leq \tau_\sigma^6 d_e^{2\gamma_e-4}\sigma^{2\ell}A^8, \quad (III)_c \leq \tau_\sigma^6\sigma^{2\ell}A^8,$$

and for all $\alpha_1+\alpha_2+2-\gamma_c\in\mathbb{R}$, $(III)_e$ and $(III)_0$ satisfy the bounds that $(III)_{ce}$ and $(III)_c$ satisfy in case $\alpha_1+\alpha_2+2-\gamma_c<0$, so that

$$\|\partial_{\perp,2}(v-\Pi^K_{pp1}v)\|^2_{L^2(K)} \leq C\left(\Psi_{p,s}A^{2s+6}((s+3)!)^2 d_c^{2(\min(\gamma_c,\gamma_e)-1)} + A^8 d_e^{2(\min(\gamma_c,\gamma_e)-2)}\sigma^{2\ell}\right).$$

Finally, the bound on the $L^2(K)$ norm of the approximation error can be obtained by a combination of the estimates above. ☐

The exponential convergence of the approximation in internal elements (i.e., elements not abutting a singular edge or corner) follows, from Lemmas A.9 to A.13.

Lemma A.14. *Let $d=3$ and $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E})$ with $\gamma_c > 3/2$, $\gamma_e > 1$. There exists a constant $C_0 > 0$ such that if $p \geq C_0\ell$, there exist constants $C, b > 0$ such that for every $\ell \in \mathbb{N}$ holds*

$$\sum_{K: d_e^K > 0} \|v - \Pi^{\ell,p}_{\mathsf{hp},d} v\|^2_{H^1(K)} \leq C e^{-b\ell}.$$

Proof. We suppose, without loss of generality, that $\gamma_c \in (3/2, 5/2)$, and $\gamma_e \in (1,2)$. The general case follows from the inclusion $\mathcal{J}^{\varpi}_{\underline{\gamma}_1}(Q;\mathcal{C},\mathcal{E}) \subset \mathcal{J}^{\varpi}_{\underline{\gamma}_2}(Q;\mathcal{C},\mathcal{E})$, valid for $\gamma_1 \geq \gamma_2$. Fix any $C_0 > 0$ and choose $p \geq C_0\ell$. For all $A > 0$ there exist $C_1, b_1 > 0$ such that (see, e.g., [43, Lemma 5.9])

$$\forall p \in \mathbb{N}: \quad \min_{1 \leq s \leq p} \Psi_{p,s} A^{2s} (s!)^2 \leq C_1 e^{-b_1 p}.$$

From (A.35) and (A.36), there holds

$$\begin{aligned}\sum_{K: d_e^K > 0} \|v - \Pi^{\ell,p}_{\mathsf{hp},d} v\|^2_{H^1(K)} &\leq C_2 \left(\sum_{K: d_e^K > 0} e^{-b_1\ell} (d_c^K)^{2(\min(\gamma_c,\gamma_e)-1)} + \sum_{K: d_e^K > 0, d_f^K = 0} (d_e^K)^{2(\min(\gamma_e,\gamma_c)-2)} \sigma^{2\ell} \right) \\ &= C_2\big((I) + (II)\big),\end{aligned}$$

where d_f^K indicates the distance of an element K to one of the faces of Q. There holds directly $(I) \leq C\ell^2 e^{-b_1\ell}$. Furthermore, because $(\min(\gamma_c,\gamma_e) - 2) < 0$,

$$\begin{aligned}(II) &\leq 6\sigma^{2\ell} \sum_{k_1=1}^{\ell} \sum_{k_2=1}^{k_1} \sigma^{2(\ell-k_2)(\min(\gamma_e,\gamma_c)-2)} \\ &\leq C\sigma^{2\ell} \sum_{k_1=1}^{\ell} \sigma^{2\ell(\min(\gamma_c,\gamma_e)-2)} \\ &\leq C\ell\sigma^{2(\min(\gamma_c,\gamma_e)-1)\ell}.\end{aligned}$$

Adjusting the constants at the exponent to absorb the terms in ℓ and ℓ^2, we obtain the desired estimate. ☐

A similar statement holds when $d = 2$, and the proof follows along the same lines.

Lemma A.15. *Let $d = 2$ and $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E})$ with $\gamma_c > 1$. There exists a constant $C_0 > 0$ such that if $p \geq C_0\ell$, there exist constants $C, b > 0$ such that*

$$\sum_{K: d_c^K > 0} \|v - \Pi^{\ell,p}_{\mathsf{hp},d} v\|^2_{H^1(K)} \leq C e^{-b\ell}, \qquad \forall \ell \in \mathbb{N}.$$

A.6 Estimates on elements along an edge in three dimensions

In the following lemma, we consider the elements K along one edge, but separated from the singular corner.

Lemma A.16. *Let $d = 3$, $e \in \mathcal{E}$ and let $K \in \mathcal{G}_3^\ell$ be such that $d_c^K > 0$ for all $c \in \mathcal{C}$ and $d_e^K = 0$. Let $C_v, A_v > 0$. Then, if $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E};C_v,A_v)$ with $\gamma_c \in (3/2, 5/2)$, $\gamma_e \in (1,2)$, there exist $C, A > 0$ such that for all $p \in \mathbb{N}$ and all $1 \leq s \leq p$, with $(p_1,p_2,p_3) \in \mathbb{N}^3$ such that $p_\parallel = p$, $p_{\perp,1} = 1 = p_{\perp,2}$,*

$$\|v - \Pi^K_{p_1p_2p_3} v\|^2_{H^1(K)} \leq C \left(\sigma^{2\{\min\{\gamma_c - 1, s\}(\ell-k)} \Psi_{p,s} A^{2s} ((s+3)!)^2 + \sigma^{2(\min(\gamma_e,\gamma_c)-1)\ell} \right), \tag{A.40}$$

where $k \in \{1,\dots,\ell\}$ is such that $d_c^K = \sigma^{\ell-k+1}$.

Proof. We suppose that $K = J_k^\ell \times J_0^\ell \times J_0^\ell$ for some $k \in \{1,\dots,\ell\}$, the elements along other edges follow by symmetry. This implies that the singular edge is parallel to the first coordinate direction. Furthermore, we denote

$$\Pi^K_{p11} = \pi_p^k \otimes (\pi_1^0 \otimes \pi_1^0) = \pi_\parallel \otimes \pi_\perp.$$

For $\alpha = (\alpha_1,\alpha_2,\alpha_3) \in \mathbb{N}_0^3$, we write $\alpha_\parallel = (\alpha_1, 0, 0)$ and $\alpha_\perp = (0, \alpha_2, \alpha_3)$. Also,

$$h_\parallel = |J_k^\ell| = \sigma^{\ell-k}(1-\sigma) \qquad h_\perp = \sigma^\ell.$$

We have

$$v - \Pi^K_{p11} v = v - \pi_\perp v + \pi_\perp\left(v - \pi_\parallel v\right). \tag{A.41}$$

We start by considering the first terms at the right-hand side of the above equation. We also compute the norms over $K_{ce} = K \cap Q_{ce}$; the estimate on the norms over $K_c = K \cap Q_c$ and $K_e = K \cap Q_e$ follow by similar or simpler arguments. By (A.21) from Lemma A.8, we have that if $\gamma_c < 2$

$$
\begin{aligned}
\sum_{|\alpha_\perp|\leq 1} h_\perp^{-2(1-|\alpha_\perp|)} \|\partial^{\alpha_\perp}(v - \pi_\perp v)\|^2_{L^2(K_{ce})} &\lesssim h_\perp^{2(\gamma_e-1)} \sum_{|\alpha_\perp|=2} \|r_e^{2-\gamma_e} \partial^{\alpha_\perp} v\|^2_{L^2(K_{ce})} \\
&\lesssim h_\|^{2(\gamma_c-\gamma_e)} h_\perp^{2(\gamma_e-1)} \sum_{|\alpha_\perp|=2} \|r_c^{(2-\gamma_c)_+} \rho_{ce}^{2-\gamma_e} \partial^{\alpha_\perp} v\|^2_{L^2(K_{ce})} \\
&\lesssim \sigma^{2k(\gamma_e-1)} \sigma^{2(\ell-k)(\gamma_c-1)} A^4 \lesssim \sigma^{2\ell(\min\{\gamma_c,\gamma_e\}-1)} A^4,
\end{aligned}
\tag{A.42a}
$$

whereas for $\gamma_c \geq 2$

$$
\begin{aligned}
\sum_{|\alpha_\perp|\leq 1} h_\perp^{-2(1-|\alpha_\perp|)} \|\partial^{\alpha_\perp}(v - \pi_\perp v)\|^2_{L^2(K_{ce})} &\lesssim h_\perp^{2(\gamma_e-1)} \sum_{|\alpha_\perp|=2} \|r_e^{2-\gamma_e} \partial^{\alpha_\perp} v\|^2_{L^2(K_{ce})} \\
&\lesssim \sigma^{2\ell(\gamma_e-1)} A^4.
\end{aligned}
\tag{A.42b}
$$

On K_e, the same bound holds as on K_{ce} for $\gamma_c \geq 2$, and on K_c the same bounds hold as on K_{ce} for $\gamma_c < 2$. By the same argument, for $|\alpha_\|| = 1$,

$$
\begin{aligned}
\|\partial^{\alpha_\|}(v - \pi_\perp v)\|^2_{L^2(K_{ce})} &= \|(\partial^{\alpha_\|} v) - \pi_\perp(\partial^{\alpha_\|} v)\|^2_{L^2(K_{ce})} \\
&\lesssim h_\perp^{2\gamma_e} \sum_{|\alpha_\perp|=2} \|r_e^{2-\gamma_e} \partial^{\alpha_\perp} \partial^{\alpha_\|} v\|^2_{L^2(K_{ce})} \\
&\lesssim h_\|^{2\widetilde{\gamma}-2} h_\perp^{2\gamma_e} \sum_{|\alpha_\perp|=2} \|r_c^{3-\gamma_c} \rho_{ce}^{2-\gamma_e} \partial^{\alpha} v\|^2_{L^2(K_{ce})} \\
&\lesssim \sigma^{2(\ell-k)(\gamma_c-1)} \sigma^{2k(\gamma_e-1)} A^6 \lesssim \sigma^{2\ell(\min\{\gamma_c,\gamma_e\}-1)} A^6,
\end{aligned}
\tag{A.43a}
$$

and

$$
\|(\partial^{\alpha_\|} v) - \pi_\perp(\partial^{\alpha_\|} v)\|^2_{L^2(K_e)} \lesssim \sigma^{2\ell\gamma_e} A^6, \tag{A.43b}
$$

$$
\|(\partial^{\alpha_\|} v) - \pi_\perp(\partial^{\alpha_\|} v)\|^2_{L^2(K_c)} \lesssim \sigma^{2(\ell-k)(\gamma_c-1)} \sigma^{2k(\gamma_e-1)} A^6 \lesssim \sigma^{2\ell(\min\{\gamma_c,\gamma_e\}-1)} A^6. \tag{A.43c}
$$

We now turn to the second part of the right-hand side of (A.41). We use (A.20) from Lemma A.8 so that

$$
\begin{aligned}
&\sum_{|\alpha_\perp|\leq 1} \|\partial^{\alpha_\perp} \pi_\perp(v - \pi_\| v)\|^2_{L^2(K)} \\
&\quad \lesssim \sum_{|\alpha_\perp|\leq 1} \|\partial^{\alpha_\perp}(v - \pi_\| v)\|^2_{L^2(K)} + \sum_{|\alpha_\perp|=2} h_\perp^{2(\gamma_e-1)} \|r_e^{2-\gamma_e} \partial^{\alpha_\perp}(v - \pi_\| v)\|^2_{L^2(K)}.
\end{aligned}
\tag{A.44}
$$

By Lemma A.7 we have, recalling that $\alpha_\| = s+1$ and $1 \leq s \leq p$, for all $|\alpha_\perp| \leq 1$,

$$
\begin{aligned}
\|\partial^{\alpha_\perp}(v - \pi_\| v)\|^2_{L^2(K)} &= \|(\partial^{\alpha_\perp} v) - \pi_\|(\partial^{\alpha_\perp} v)\|^2_{L^2(K)} \\
&\lesssim \tau_\sigma^{2s+2} h_\|^{2\min\{\gamma_c,s+1\}} \Psi_{p,s} \||x_1|^{(s+1-\gamma_c)_+} \partial^{\alpha_\|} \partial^{\alpha_\perp} v)\|^2_{L^2(K)},
\end{aligned}
$$

and, for all $|\alpha_\perp| = 2$, using that $\pi_\|$ and multiplication by r_e commute, because r_e does not depend on x_1,

$$
\begin{aligned}
\|r_e^{2-\gamma_e} \partial^{\alpha_\perp}(v - \pi_\| v)\|^2_{L^2(K)} &= \|(r_e^{2-\gamma_e} \partial^{\alpha_\perp} v) - \pi_\|(r_e^{2-\gamma_e} \partial^{\alpha_\perp} v)\|^2_{L^2(K)} \\
&\lesssim \tau_\sigma^{2s+2} h_\|^{2\min\{\gamma_c,s+1\}} \Psi_{p,s} \||x_1|^{(s+1-\gamma_c)_+} r_e^{2-\gamma_e} \partial^{\alpha_\|} \partial^{\alpha_\perp} v)\|^2_{L^2(K)}.
\end{aligned}
$$

Then, remarking that $|x_1| \lesssim r_c \lesssim |x_1|$, combining (A.44) with the two inequalities above we obtain

$$
\begin{aligned}
&\sum_{|\alpha_\perp|\leq 1} \|\partial^{\alpha_\perp} \pi_\perp(v - \pi_\| v)\|^2_{L^2(K)} \\
&\quad \lesssim \tau_\sigma^{2s+2} \Psi_{p,s} h_\|^{2\min\{\gamma_c-1,s\}} h_\|^2 \left(\sum_{|\alpha_\perp|\leq 1} \|r_c^{(s+1-\gamma_c)_+} \partial^{\alpha} v\|^2_{L^2(K)} \right. \\
&\qquad\qquad \left. + \sum_{|\alpha_\perp|=2} h_\perp^{2(\gamma_e-1)} \|r_c^{(s+1-\gamma_c)_+} r_e^{2-\gamma_e} \partial^{\alpha} v\|^2_{L^2(K)} \right).
\end{aligned}
$$

Adjusting the exponent of the weights, replacing $h_\parallel$ and $h_\perp$ with their definition, we find that there exists $A > 0$ depending only on σ and A_v such that

$$\begin{aligned}
&\sum_{|\alpha_\perp|\leq 1} \|\partial^{\alpha_\perp}\pi_\perp(v-\pi_\parallel v)\|^2_{L^2(K_{ce})} \\
&\quad\lesssim \tau_\sigma^{2s+2}\Psi_{p,s} h_\parallel^{2\min\{\gamma_c-1,s\}} h_\parallel^2 \Bigg(\sum_{|\alpha_\perp|\leq 1} h_\parallel^{-2|\alpha_\perp|} \|r_c^{(s+1+|\alpha_\perp|-\gamma_c)_+}\partial^\alpha v\|^2_{L^2(K_{ce})} \\
&\qquad\qquad + \sum_{|\alpha_\perp|=2} h_\perp^{2(\gamma_e-1)} h_\parallel^{-2\gamma_e} \|r_c^{s+3-\gamma_c}\rho_{ce}^{2-\gamma_e}\partial^\alpha v\|^2_{L^2(K_{ce})} \Bigg) \\
&\quad\lesssim \sigma^{2(\ell-k)\min\{\gamma_c-1,s\}}\Psi_{p,s} A^{2s+4}((s+3)!)^2,
\end{aligned} \tag{A.45a}$$

and similarly

$$\sum_{|\alpha_\perp|\leq 1} \|\partial^{\alpha_\perp}\pi_\perp(v-\pi_\parallel v)\|^2_{L^2(K_e)} \lesssim \sigma^{2(\ell-k)\min\{\gamma_c,s+1\}}\Psi_{p,s} A^{2s+4}((s+3)!)^2, \tag{A.45b}$$

and the estimate on K_c is the same as that on K_{ce}. Similarly to (A.44), using first (A.23) from the proof of Lemma A.8, and then Lemma A.7

$$\begin{aligned}
&\sum_{|\alpha_\parallel|\leq 1} \|\partial^{\alpha_\parallel}\pi_\perp(v-\pi_\parallel v)\|^2_{L^2(K)} \\
&\quad\lesssim \sum_{|\alpha_\parallel|\leq 1} \Bigg(\sum_{|\alpha_\perp|\leq 1} h_\perp^{2|\alpha_\perp|} \|\partial^{\alpha_\perp}\partial^{\alpha_\parallel}(v-\pi_\parallel v)\|^2_{L^2(K)} + \sum_{|\alpha_\perp|=2} h_\perp^{2\gamma_e} \|r_e^{2-\gamma_e}\partial^{\alpha_\perp}\partial^{\alpha_\parallel}(v-\pi_\parallel v)\|^2_{L^2(K)} \Bigg) \\
&\quad\lesssim \tau_\sigma^{2s+2}\Psi_{p,s} h_\parallel^{2\min\{\gamma_c-1,s\}} \Bigg(\sum_{|\alpha_\parallel|=s+1}\sum_{|\alpha_\perp|\leq 1} h_\perp^{2|\alpha_\perp|} \|r_c^{(s+1-\gamma_c)_+}\partial^{\alpha_\parallel}\partial^{\alpha_\perp} v\|^2_{L^2(K)} \\
&\qquad\qquad + \sum_{|\alpha_\parallel|=s+1}\sum_{|\alpha_\perp|=2} h_\perp^{2\gamma_e} \|r_e^{2-\gamma_e} r_c^{(s+1-\gamma_c)_+}\partial^{\alpha_\parallel}\partial^{\alpha_\perp} v\|^2_{L^2(K)} \Bigg).
\end{aligned}$$

As before, there exists $A > 0$ depending only on σ and A_v such that

$$\begin{aligned}
&\sum_{|\alpha_\parallel|\leq 1} \|\partial^{\alpha_\parallel}\pi_\perp(v-\pi_\parallel v)\|^2_{L^2(K_{ce})} \\
&\quad\lesssim \tau_\sigma^{2s+2}\Psi_{p,s} h_\parallel^{2\min\{\gamma_c-1,s\}} \Bigg(\sum_{|\alpha_\parallel|=s+1}\sum_{|\alpha_\perp|\leq 1} h_\perp^{2|\alpha_\perp|} h_\parallel^{-2|\alpha_\perp|} \|r_c^{(s+1+|\alpha_\perp|-\gamma_c)_+}\partial^\alpha v\|^2_{L^2(K_{ce})} \\
&\qquad\qquad + \sum_{|\alpha_\parallel|=s+1}\sum_{|\alpha_\perp|=2} h_\perp^{2\gamma_e} h_\parallel^{-2\gamma_e} \|r_c^{s+3-\gamma_c}\rho_{ce}^{2-\gamma_e}\partial^\alpha v\|^2_{L^2(K_{ce})} \Bigg) \\
&\quad\lesssim \sigma^{2(\ell-k)\min\{\gamma_c-1,s\}}\Psi_{p,s} A^{2s+4}((s+3)!)^2,
\end{aligned} \tag{A.46a}$$

and

$$\begin{aligned}
&\sum_{|\alpha_\parallel|\leq 1} \|\partial^{\alpha_\parallel}\pi_\perp(v-\pi_\parallel v)\|^2_{L^2(K_e)} \\
&\quad\lesssim \tau_\sigma^{2s+2}\Psi_{p,s} h_\parallel^{2\min\{\gamma_c-1,s\}} \Bigg(\sum_{|\alpha_\parallel|=s+1}\sum_{|\alpha_\perp|\leq 1} h_\perp^{2|\alpha_\perp|} \|r_c^{(s+1-\gamma_c)_+}\partial^\alpha v\|^2_{L^2(K_e)} \\
&\qquad\qquad + \sum_{|\alpha_\parallel|=s+1}\sum_{|\alpha_\perp|=2} h_\perp^{2\gamma_e} \|r_c^{(s+1-\gamma_c)_+} r_e^{2-\gamma_e}\partial^\alpha v\|^2_{L^2(K_e)} \Bigg) \\
&\quad\lesssim \sigma^{2(\ell-k)\min\{\gamma_c-1,s\}}\Psi_{p,s} A^{2s+4}((s+3)!)^2,
\end{aligned} \tag{A.46b}$$

and the estimate on K_c is the same as that on K_{ce}. The assertion now follows from (A.42), (A.43), (A.45), and (A.46), upon possibly adjusting the value of the constant A. □

Lemma A.17. *Let $d = 3$ and $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E})$ with $\gamma_c > 3/2$, $\gamma_e > 1$. There exists a constant $C_0 > 0$ such that if $p \geq C_0 \ell$, there exist constants $C, b > 0$ such that*

$$\sum_{K: d_c^K > 0, d_e^K = 0} \|v - \Pi^{\ell,p}_{\mathsf{hp},d} v\|_{H^1(K)} \leq C e^{-b\ell}, \qquad \forall \ell \in \mathbb{N}.$$

Proof. As in the proof of Lemma A.14, we may assume that $\gamma_c \in (3/2, 5/2)$ and $\gamma_e \in (1, 2)$. The proof of this statements follows by summing over the right-hand side of (A.40), i.e.,

$$\begin{aligned}\sum_{K: d_c^K > 0, d_e^K = 0} \|v - \Pi^{\ell,p}_{\mathsf{hp},d} v\|^2_{H^1(K)} &\leq C \left(\sum_{k=1}^{\ell} \sigma^{2\min\{\gamma_c - 1, s\}(\ell - k)} \Psi_{p,s} A^{2s} ((s+3)!)^2 + \sigma^{2(\min(\gamma_c, \gamma_e) - 1)\ell} \right) \\ &= C((I) + (II)).\end{aligned}$$

We have $(II) \lesssim \ell \sigma^{2(\min(\gamma_c, \gamma_e) - 1)\ell}$; the observation that for all $A > 0$ there exist $C_1, b_1 > 0$ such that

$$\min_{1 \leq s \leq p} \Psi_{p,s} ((s+3)!)^2 A^{2s} \leq C_1 e^{-b_1 p},$$

(see, e.g., [43, Lemma 5.9]). Combining with $p \geq C_0 \ell$ concludes the proof. □

A.7 Estimates at the corner

The lemma below follows from classic low-order finite element approximation results and from the embedding $\mathcal{J}^2_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E}) \subset H^{1+\theta}(Q)$, valid for a $\theta > 0$ if $\gamma_c - d/2 > 0$, for all $c \in \mathcal{C}$, and, when $d = 3$, $\gamma_e > 1$ for all $e \in \mathcal{E}$ (see, e.g., [41, Remark 2.3]).

Lemma A.18. *Let $d \in \{2, 3\}$, $K = \bigtimes_{i=1}^d J_0^\ell$. Then, if $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E})$ with*

$$\begin{aligned} &\gamma_c > 1, \text{ for all } c \in \mathcal{C}, && \text{if } d = 2, \\ &\gamma_c > 3/2 \text{ and } \gamma_e > 1, \text{ for all } c \in \mathcal{C} \text{ and } e \in \mathcal{E}, && \text{if } d = 3, \end{aligned}$$

there exists a constant $C_0 > 0$ independent of ℓ such that if $p \geq C_0 \ell$, there exist constants $C, b > 0$ such that

$$\|v - \Pi^{\ell,p}_{\mathsf{hp},d} v\|_{H^1(K)} \leq C e^{-b\ell}.$$

A.8 Exponential convergence

The exponential convergence of the approximation in the full domain Q follows then from Lemmas A.14, A.15, A.17, and A.18.

Proposition A.19. *Let $d \in \{2, 3\}$, $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E})$ with*

$$\begin{aligned} &\gamma_c > 1, \text{ for all } c \in \mathcal{C}, && \text{if } d = 2, \\ &\gamma_c > 3/2 \text{ and } \gamma_e > 1, \text{ for all } c \in \mathcal{C} \text{ and } e \in \mathcal{E}, && \text{if } d = 3. \end{aligned}$$

Then, there exist constants $c_p > 0$ and $C, b > 0$ such that, for all $\ell \in \mathbb{N}$,

$$\|v - \Pi^{\ell, c_p \ell}_{\mathsf{hp},d} v\|_{H^1(Q)} \leq C e^{-b\ell}.$$

With respect to the dimension of the discrete space $N_{\mathrm{dof}} = \dim(X^{\ell, c_p \ell}_{\mathsf{hp},d})$, the above bound reads

$$\|v - \Pi^{\ell, c_p \ell}_{\mathsf{hp},d} v\|_{H^1(Q)} \leq C \exp(-b N_{\mathrm{dof}}^{1/(2d)}).$$

A.9 Explicit representation of the approximant in terms of continuous basis functions

Let $p \in \mathbb{N}$. Let $\hat{\zeta}_1(x) = (1+x)/2$ and $\hat{\zeta}_2 = (1-x)/2$. Let also $\hat{\zeta}_n(x) = \frac{1}{2} \int_{-1}^x L_{n-2}(\xi) d\xi$, for $n = 3, \dots, p+1$, where L_{n-2} denotes the $L^\infty((-1,1))$-normalized Legendre polynomial of degree $n-2$ introduced in Section A.2. Then, fix $\ell \in \mathbb{N}$ and write $\zeta_n^k = \hat{\zeta}_n \circ \phi_k$, $n = 1, \dots, p+1$ and $k = 0, \dots, \ell$, with the affine map $\phi_k : J_k^\ell \to (-1, 1)$ introduced in Section A.2. We construct those functions explicitly: denoting $J_k^\ell = (x_k, x_{k+1})$ and $h_k = |x_{k+1} - x_k|$, there holds, for $x \in J_k^\ell$,

$$\zeta_1^k(x) = \frac{1}{h_k}(x - x_k), \qquad \zeta_2^k(x) = \frac{1}{h_k}(x_{k+1} - x), \tag{A.47}$$

and

$$\zeta_n^k(x) = \frac{1}{h_k}\int_{x_k}^{x} L_{n-2}(\phi_k(\eta))d\eta \qquad n = 3,\dots,p+1. \tag{A.48}$$

Then, for any element $K \in \mathcal{G}_3^\ell$, with $K = J_{k_1}^\ell \times J_{k_2}^\ell \times J_{k_3}^\ell$, there exist coefficients $c_{i_1\dots i_d}^K$ such that

$$\Pi_{\mathsf{hp},d}^{\ell,p} u_{|K}(x_1,x_2,x_3) = \sum_{i_1,i_2,i_3=1}^{p+1} c_{i_1\dots i_d}^K \zeta_{i_1}^{k_1}(x_1)\zeta_{i_2}^{k_2}(x_2)\zeta_{i_3}^{k_3}(x_3), \quad \forall (x_1,x_2,x_3)\in K \tag{A.49}$$

by construction. We remark that, whenever $i_j > 2$ for all $j = 1,2,3$, the basis functions vanish on the boundary of the element:

$$\left(\zeta_{i_1}^{k_1}\zeta_{i_2}^{k_2}\zeta_{i_3}^{k_3}\right)_{|\partial K} = 0 \qquad \text{if } i_j \geq 3,\ j=1,2,3.$$

Furthermore, write

$$\psi_{i_1\dots i_d}^K(x_1,x_2,x_3) = \zeta_{i_1}^{k_1}(x_1)\zeta_{i_2}^{k_2}(x_2)\zeta_{i_3}^{k_3}(x_3)$$

and consider $t_{i_1\dots i_d} = \#\{i_j \leq 2,\ j=1,2,3\}$. We have

- if $t_{i_1\dots i_d} = 1$, then $\psi_{i_1\dots i_d}^K$ is not zero only on one face of the boundary of K,
- if $t_{i_1\dots i_d} = 2$, then $\psi_{i_1\dots i_d}^K$ is not zero only on one edge and neighboring faces of the boundary of K,
- if $t_{i_1\dots i_d} = 3$, then $\psi_{i_1\dots i_d}^K$ is not zero only on one corner and neighboring edges and faces of the boundary of K.

Similar arguments hold when $d = 2$.

A.9.1 Explicit bounds on the coefficients

We derive here a bound on the coefficients of the local projectors with respect to the norms of the projected function. We will use that

$$\|L_i\circ\phi_k\|_{L^2(J_k^\ell)} = \left(\frac{h_k}{2}\right)^{1/2}\|L_i\|_{L^2((-1,1))} = \left(\frac{h_k}{2i+1}\right)^{1/2} \qquad \forall i\in\mathbb{N}_0,\ \forall k\in\{0,\dots,\ell\}. \tag{A.50}$$

Remark A.20. *As mentioned in Remark A.3, the hp-projector $\Pi_{\mathsf{hp},d}^{\ell,p}$ can be defined for more general functions than $u \in H^1_{\mathrm{mix}}(Q)$. As follows from Equations* (A.53), (A.57), (A.61) *and* (A.64) *below, the projector is also defined for $u\in W^{1,1}_{\mathrm{mix}}(Q)$.*

Lemma A.21. *There exist constants C_1, C_2 such that, for all $u \in W^{1,1}_{\mathrm{mix}}(Q)$, all $\ell\in\mathbb{N}$, all $p\in\mathbb{N}$*

$$|c_{i_1\dots i_d}^K| \leq C\left(\prod_{j=1}^d i_j\right)\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)} \qquad \forall K\in\mathcal{G}_d^\ell,\ \forall (i_1,\dots,i_d)\in\{1,\dots,p+1\}^d \tag{A.51}$$

and for all $(i_1,\dots,i_d)\in\{1,\dots,p+1\}^d$

$$\sum_{K\in\mathcal{G}_3^\ell}|c_{i_1\dots i_d}^K| \leq C\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)}\begin{cases}\left(\prod_{j=1}^d i_j\right) & \text{if } t_{i_1\dots i_d}=0,\\ (\ell+1)\left(\sum_{j_1=1}^d\sum_{j_2=j_1+1}^d i_{j_1}i_{j_2}\right) & \text{if } t_{i_1\dots i_d}=1,\\ (\ell+1)^2\left(\sum_{j=1}^d i_j\right) & \text{if } t_{i_1\dots i_d}=2,\\ (\ell+1)^d & \text{if } t_{i_1\dots i_d}=3.\end{cases} \tag{A.52}$$

Proof. Let $d=3$ and $K = J_{k_1}^\ell\times J_{k_2}^\ell\times J_{k_3}^\ell \in \mathcal{G}_3^\ell$.
Internal modes. We start by considering the case of the coefficients of internal modes, i.e., c_{i_1,i_2,i_3}^K as defined in (A.49) for $i_n\geq 3$, $n=1,2,3$. Let then $i_1,i_2,i_3\in\{3,\dots,p+1\}$ and write $L_n^k = L_n\circ\phi_k$: there holds

$$c_{i_1,i_2,i_3}^K = (2i_1-3)(2i_2-3)(2i_3-3)\int_K(\partial_{x_1}\partial_{x_2}\partial_{x_3}u(x_1,x_2,x_3))\,L_{i_1-2}^{k_1}(x_1)L_{i_2-2}^{k_2}(x_2)L_{i_3-2}^{k_3}(x_3)dx_1dx_2dx_3. \tag{A.53}$$

If $u\in W^{1,1}_{\mathrm{mix}}(K)$, since $\|L_n\|_{L^\infty(-1,1)}=1$ for all n, we have

$$|c_{i_1\dots i_d}^K| \leq (2i_1-3)(2i_2-3)(2i_3-3)\|\partial_{x_1}\partial_{x_2}\partial_{x_3}u\|_{L^1(K)} \qquad i_n\geq 3,\ n=1,2,3, \tag{A.54}$$

hence,

$$\sum_{K\in\mathcal{G}_3^\ell} |c^K_{i_1\dots i_d}| \le (2i_1-3)(2i_2-3)(2i_3-3)\|\partial_{x_1}\partial_{x_2}\partial_{x_3}u\|_{L^1(Q)} \qquad i_n\ge 3,\ n=1,2,3. \tag{A.55}$$

Face modes. We continue with face modes and fix, for ease of notation, $i_1 = 1$. We also denote $F = J^\ell_{k_2}\times J^\ell_{k_3}$. The estimates will then also hold for $i_1 = 2$ and for any permutation of the indices by symmetry. We introduce the trace inequality constant $C^{T,1}$, independent of K, such that, for all $v\in W^{1,1}(Q)$ and $\hat{x}\in(0,1)$,

$$\|v(\hat{x},\cdot,\cdot)\|_{L^1(F)} \le \|v(\hat{x},\cdot,\cdot)\|_{L^1((0,1)^2)} \le C^{T,1}\left(\|v\|_{L^1(Q)} + \|\partial_{x_1}v\|_{L^1(Q)}\right). \tag{A.56}$$

This follows from the trace estimate in [42, Lemma 4.2] and from the fact that

$$\|v(\hat{x},\cdot,\cdot)\|_{L^1((0,1)^2)} \le C\min\left\{\frac{1}{|1-\hat{x}|}\|v\|_{L^1((\hat{x},1)\times(0,1)^2)} + \|\partial_{x_1}v\|_{L^1((\hat{x},1)\times(0,1)^2)},\right.$$
$$\left.\frac{1}{|\hat{x}|}\|v\|_{L^1((0,\hat{x})\times(0,1)^2)} + \|\partial_{x_1}v\|_{L^1((0,\hat{x})\times(0,1)^2)}\right\}.$$

There holds, for $i_2, i_3\in\{3,\dots,p+1\}$,

$$c^K_{1,i_2,i_3} = (2i_2-3)(2i_3-3)\int_F\left(\partial_{x_2}\partial_{x_3}u(x^\ell_{k_1},x_2,x_3)\right)L^{k_2}_{i_2-2}(x_2)L^{k_3}_{i_3-2}(x_3)dx_2dx_3. \tag{A.57}$$

Since the Legendre polynomials are L^∞ normalized and using the trace inequality (A.56),

$$|c^K_{1,i_2,i_3}| \le (2i_2-3)(2i_3-3)\|(\partial_{x_2}\partial_{x_3}u)(x^\ell_{k_1},\cdot,\cdot)\|_{L^1(F)} \le C^{T,1}(2i_2-3)(2i_3-3)\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)}. \tag{A.58}$$

Summing over all internal faces, furthermore,

$$\begin{aligned}\sum_{K\in\mathcal{G}_3^\ell}|c^K_{1,i_2,i_3}| &\le (2i_2-3)(2i_3-3)\sum_{k_1=0}^{\ell}\|(\partial_{x_2}\partial_{x_3}u)(x^\ell_{k_1},\cdot,\cdot)\|_{L^1((0,1)^2)}\\ &\le C^{T,1}(\ell+1)(2i_2-3)(2i_3-3)\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)}.\end{aligned} \tag{A.59}$$

Edge modes. We now consider edge modes. Fix for ease of notation $i_1 = i_2 = 1$; as before, the estimates will hold for $(i_1,i_2)\in\{1,2\}^2$ and for any permutation of the indices. By the same arguments as for (A.56), there exists a trace constant $C^{T,2}$ such that, denoting $e = J^\ell_{k_3}$, for all $v\in W^{1,1}((0,1)^2)$ and for all $\hat{x}\in(0,1)$,

$$\|v(\hat{x},\cdot)\|_{L^1(e)} \le \|v(\hat{x},\cdot)\|_{L^1((0,1))} \le C^{T,2}\left(\|u\|_{L^1((0,1)^2)} + \|\partial_{x_2}u\|_{L^1((0,1)^2)}\right). \tag{A.60}$$

By definition,

$$c^K_{1,1,i_3} = (2i_3-3)\int_e\left(\partial_{x_3}u(x^\ell_{k_1},x^\ell_{k_2},x_3)\right)L^{k_3}_{i_3-2}(x_3)dx_3. \tag{A.61}$$

Using (A.56) and (A.60)

$$|c^K_{1,1,i_3}| \le (2i_3-3)\|(\partial_{x_3}u)(x^\ell_{k_1},x^\ell_{k_2},\cdot)\|_{L^1(e)} \le C^{T,1}C^{T,2}(2i_3-3)\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)}. \tag{A.62}$$

Summing over edges, in addition,

$$\begin{aligned}\sum_{K\in\mathcal{G}_3^\ell}|c^K_{1,1,i_3}| &\le (2i_3-3)\sum_{k_1=0}^{\ell}\sum_{k_2=0}^{\ell}\|(\partial_{x_3}u)(x^\ell_{k_1},x^\ell_{k_2},\cdot)\|_{L^1((0,1))}\\ &\le C^{T,1}C^{T,2}(\ell+1)^2(2i_3-3)\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)}.\end{aligned} \tag{A.63}$$

Node modes. Finally, we consider the coefficients of nodal modes, i.e., $c^K_{i_1,i_2,i_3}$ for $i_1,i_2,i_3\in\{1,2\}$, which by construction equal function values of u, e.g.

$$c_{111} = u(x^\ell_{k_1},x^\ell_{k_2},x^\ell_{k_3}). \tag{A.64}$$

The Sobolev imbedding $W^{1,1}_{\mathrm{mix}}(Q)\hookrightarrow L^\infty(Q)$ and scaling implies the existence of a uniform constant C_{imb} such that, for any $v\in W^{1,1}_{\mathrm{mix}}(Q)$

$$\|v\|_{L^\infty(K)} \le \|v\|_{L^\infty(Q)} \le C_{\mathrm{imb}}\|v\|_{W^{1,1}_{\mathrm{mix}}(Q)}.$$

Then, by construction,

$$|c^K_{i_1,i_2,i_3}| \leq \|u\|_{L^\infty(K)} \leq C_{\mathrm{imb}}\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)} \qquad \forall i_1, i_2, i_3 \in \{1,2\}. \tag{A.65}$$

Summing over nodes, it follows directly that

$$\sum_{K\in\mathcal{G}^\ell_3} |c^K_{i_1,i_2,i_3}| \leq \sum_{K\in\mathcal{G}^\ell_3} \|u\|_{L^\infty(K)} \leq C_{\mathrm{imb}}(\ell+1)^3\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)} \qquad \forall i_1, i_2, i_3 \in \{1,2\}. \tag{A.66}$$

We obtain (A.51) from (A.54), (A.58), (A.62), and (A.65). Furthermore, (A.52) follows from (A.55), (A.59), (A.63), and (A.66). The estimates for the case $d = 2$ follow from the same argument. ☐

The following lemma shows the continuous imbedding of $\mathcal{J}^d_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E})$ into $W^{1,1}_{\mathrm{mix}}(Q)$, given sufficiently large weights $\underline{\gamma}$.

Lemma A.22. *Let $d \in \{2,3\}$. Let $\underline{\gamma}$ be such that $\gamma_c > d/2$, for all $c \in \mathcal{C}$ and (if $d = 3$) $\gamma_e > 1$ for all $e \in \mathcal{E}$. There exists a constant $C > 0$ such that, for all $u \in \mathcal{J}^d_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E})$,*

$$\|u\|_{W^{1,1}_{\mathrm{mix}}(Q)} \leq C\|u\|_{\mathcal{J}^d_{\underline{\gamma}}(Q)}.$$

Proof. We recall the decomposition of Q as

$$\overline{Q} = \overline{Q_0} \cup \overline{Q_\mathcal{C}} \cup \overline{Q_\mathcal{E}} \cup \overline{Q_{\mathcal{C}\mathcal{E}}},$$

where $Q_\mathcal{E} = Q_{\mathcal{C}\mathcal{E}} = \varnothing$ if $d = 2$. There holds

$$\|u\|_{W^{1,1}_{\mathrm{mix}}(Q_0)} \leq C|Q_0|^{1/2}\|u\|_{H^d(Q_0)} \leq C|Q_0|^{1/2}\|u\|_{\mathcal{J}^d_{\underline{\gamma}}(Q)}. \tag{A.67}$$

We now consider the subdomain Q_c, for any $c \in \mathcal{C}$. There holds, with constant C that depends only on γ_c and on $|Q_c|$,

$$\begin{aligned}\|u\|_{W^{1,1}_{\mathrm{mix}}(Q_c)} &= \|u\|_{W^{1,1}(Q_c)} + \sum_{\substack{2\leq|\alpha|\leq d\\|\alpha|_\infty\leq 1}} \|\partial^\alpha u\|_{L^1(Q_c)}\\ &\leq C|Q_c|^{1/2}\|u\|_{H^1(Q_c)} + C\sum_{\substack{2\leq|\alpha|\leq d\\|\alpha|_\infty\leq 1}} \|r_c^{-(|\alpha|-\gamma_c)_+}\|_{L^2(Q_c)}\|r_c^{(|\alpha|-\gamma_c)_+}\partial^\alpha u\|_{L^2(Q_c)}\\ &\leq C\|u\|_{\mathcal{J}^d_{\underline{\gamma}}(Q)},\end{aligned} \tag{A.68}$$

where the last inequality follows from the fact that $\gamma_c > d/2$, hence the norm $\|r_c^{-(|\alpha|-\gamma_c)_+}\|_{L^2(Q_c)}$ is bounded for all $|\alpha| \leq d$. Consider then $d = 3$ and any $e \in \mathcal{E}$. Suppose also, without loss of generality, that $\gamma_c - \gamma_e > 1/2$ and $\gamma_e < 2$ (otherwise, it is sufficient to replace γ_e by a smaller $\widetilde{\gamma}_e$ such that $1 < \widetilde{\gamma}_e < \gamma_c - 1/2$ and $\gamma_e < 2$ and remark that $\mathcal{J}^d_{\underline{\gamma}}(Q;\mathcal{C},\mathcal{E}) \subset \mathcal{J}^d_{\underline{\widetilde{\gamma}}}(Q;\mathcal{C},\mathcal{E})$ if $\widetilde{\gamma}_e < \gamma_e$). Since $\gamma_e > 1$, then $\|r_e^{-|\alpha_\perp|+\gamma_e}\|_{L^2(Q_e)}$ is bounded by a constant depending only on γ_e and $|Q_e|$ as long as α is such that $|\alpha_\perp| \leq 2$. Hence, denoting by $\partial_\parallel$ the derivative in the direction parallel to e,

$$\begin{aligned}\|u\|_{W^{1,1}_{\mathrm{mix}}(Q_e)} &= \|u\|_{W^{1,1}(Q_e)} + \sum_{|\alpha_\perp|=1}\|\partial_\parallel\partial^{\alpha_\perp}u\|_{L^1(Q_e)} + \sum_{\alpha_1=0,1}\|\partial_\parallel^{\alpha_1}\partial_{\perp,1}\partial_{\perp,2}u\|_{L^1(Q_e)}\\ &\leq C|Q_e|^{1/2}\left(\|u\|_{H^1(Q_e)} + \sum_{|\alpha_\perp|=1}\|\partial_\parallel\partial^{\alpha_\perp}v\|_{L^2(Q_e)}\right)\\ &\qquad + C\sum_{\alpha_1=0,1}\|r_e^{-2+\gamma_e}\|_{L^2(Q_e)}\|r_e^{2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}\partial_{\perp,2}u\|_{L^2(Q_e)}\\ &\leq C\|u\|_{\mathcal{J}^3_{\underline{\gamma}}(Q)}.\end{aligned} \tag{A.69}$$

Since $x_\parallel \leq r_c(x) \leq \widehat{\varepsilon}$ for all $x \in Q_{ce}$, and due to the fact that $Q_{ce} \subset \left\{x_\parallel \in (0,\widehat{\varepsilon}), (x_{\perp,1}, x_{\perp,2}) \in (0,\widehat{\varepsilon}^2)^2\right\}$, there holds

$$\|r_c^{-(\gamma_e+1-\gamma_c)_+}r_e^{-2+\gamma_e}\|_{L^2(Q_{ce})} \leq \|x_\parallel^{-(\gamma_e+1-\gamma_c)_+}\|_{L^2((0,\widehat{\varepsilon}))}\|r_e^{-2+\gamma_e}\|_{L^2((0,\widehat{\varepsilon}^2)^2)} \leq C,$$

for a constant C that depends only on $\widehat{\varepsilon}$, γ_c, and γ_e. Hence,

$$
\begin{aligned}
\|u\|_{W^{1,1}_{\mathrm{mix}}(Q_{ce})} &= \|u\|_{W^{1,1}(Q_{ce})} + \sum_{|\alpha_\perp|=1} \|\partial_\parallel \partial^{\alpha_\perp} u\|_{L^1(Q_{ce})} + \sum_{\alpha_1=0,1} \|\partial_\parallel^{\alpha_1} \partial_{\perp,1}\partial_{\perp,2} u\|_{L^1(Q_{ce})} \\
&\leq C|Q_{ce}|^{1/2}\|u\|_{H^1(Q_e)} + C\sum_{|\alpha_\perp|=1} \|r_c^{-(2-\gamma_c)_+}\|_{L^2(Q_{ce})} \|r_c^{(2-\gamma_c)_+}\partial_\parallel \partial^{\alpha_\perp} u\|_{L^2(Q_{ce})} \\
&\quad + C\sum_{\alpha_1=0,1} \|r_c^{-(\alpha_1+\gamma_e-\gamma_c)_+} r_e^{-2+\gamma_e}\|_{L^2(Q_{ce})} \|r_c^{(\alpha_1+2-\gamma_c)_+}\rho_{ce}^{2-\gamma_e}\partial_\parallel^{\alpha_1}\partial_{\perp,1}\partial_{\perp,2} u\|_{L^2(Q_{ce})} \\
&\leq C\|u\|_{\mathcal{J}^3_{\underline{\gamma}}(Q)},
\end{aligned}
\tag{A.70}
$$

with C independent of u. Combining inequalities (A.67) to (A.70) concludes the proof. □

The following statement is a direct consequence of the two lemmas above and the fact that $\left\|\psi^K_{i_1\dots i_d}\right\|_{L^\infty(K)} \leq 1$ for all $K \in \mathcal{G}^\ell_3$ and all $i_1, \dots, i_d \in \{1, \dots, p+1\}$.

Corollary A.23. *Let $\underline{\gamma}$ be such that $\gamma_c - d/2 > 0$, for all $c \in \mathcal{C}$ and, if $d = 3$, $\gamma_e > 1$ for all $e \in \mathcal{E}$. There exists a constant $C > 0$ such that for all $\ell, p \in \mathbb{N}$ and for all $u \in \mathcal{J}^d_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E})$,*

$$
\|\Pi^{\ell,p}_{\mathsf{hp},d} u\|_{L^\infty(Q)} \leq C p^{2d} \|u\|_{\mathcal{J}^d_{\underline{\gamma}}(Q)}.
$$

A.9.2 Basis of continuous functions with compact support

It is possible to construct a basis for $\Pi^{\ell,p}_{\mathsf{hp},d}$ in Q such that all basis functions are continuous and have compact support. For all $\ell \in \mathbb{N}$ and all $p \in \mathbb{N}$, extend to zero outside of their domain of definition the functions ζ^k_n defined in (A.47) and (A.48), for $k = 0, \dots, \ell$ and $n = 1, \dots, p+1$. We introduce the univariate functions with compact support $v_j : (0,1) \to \mathbb{R}$, for $j = 1, \dots, (\ell+1)p + 1$ so that $v_1 = \zeta^0_2$, $v_{\ell+2} = \zeta^\ell_1$,

$$
v_k = \zeta^{k-2}_1 + \zeta^{k-1}_2, \qquad \text{for all } k = 2, \dots, \ell+1 \tag{A.71}
$$

and

$$
v_{\ell+2+k(p-1)+n} = \zeta^k_{n+2}, \qquad \text{for all } k = 0, \dots, \ell \text{ and } n = 1, \dots, p-1.
$$

Proposition A.24. *Let $\ell \in \mathbb{N}$ and $p \in \mathbb{N}$. Furthermore, let $u \in \mathcal{J}^d_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E})$ with $\underline{\gamma}$ such that $\gamma_c - d/2 > 0$ and, if $d = 3$, $\gamma_e > 1$. Let $N_{1\mathrm{d}} = (\ell+1)p + 1$. There exists an array of coefficients*

$$
c = \left\{ c_{i_1\dots i_d} : (i_1, \dots, i_d) \in \{1, \dots, N_{1\mathrm{d}}\}^d \right\}
$$

such that

$$
\left(\Pi^{\ell,p}_{\mathsf{hp},d} u\right)(x_1, \dots, x_d) = \sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} c_{i_1\dots i_d} \prod_{j=1}^{d} v_{i_j}(x_j) \qquad \forall (x_1, \dots, x_d) \in Q. \tag{A.72}
$$

Furthermore, there exist constants $C_1, C_2 > 0$ independent of ℓ, p, and u, such that

$$
|c_{i_1\dots i_d}| \leq C_1 (p+1)^d \|u\|_{\mathcal{J}^d_{\underline{\gamma}}(Q)} \qquad \forall i_1, \dots, i_d \in \{1, \dots, N_{1\mathrm{d}}\}^d
$$

and

$$
\sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} |c_{i_1\dots i_d}| \leq C_2 \left(\sum_{t=0}^{d} (\ell+1)^t (p+1)^{2(d-t)} \right) \|u\|_{\mathcal{J}^d_{\underline{\gamma}}(Q)}.
$$

Proof. The statement follows directly from the construction of the projector, see (A.49), and from the bounds in Lemmas A.21 and A.22. In particular, (A.72) holds because the element-wise coefficients related to ζ^{k-1}_2 and to ζ^{k-2}_1 are equal: it follows from Equations (A.57), (A.61) and (A.64) that $c^K_{1i_2\dots i_d} = c^{K'}_{2i_2\dots i_d}$ for all $i_2, \dots, i_d \in \{1, \dots, p+1\}$, all $K = J^\ell_{k_1} \times J^\ell_{k_2} \times J^\ell_{k_3} \in \mathcal{G}^\ell_3$ satisfying $k_1 < \ell$ and $K' = J^\ell_{k_1+1} \times J^\ell_{k_2} \times J^\ell_{k_3} \in \mathcal{G}^\ell_3$. The same holds for permutations of $i_1, \dots, i_d$. Because $(v_k)^{(\ell+1)p+1}_{k=1}$ are continuous, this again shows continuity of $\Pi^{\ell,p}_{\mathsf{hp},d} u$ (Remark A.2). The last estimate is obtained with (A.52):

$$
\sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} |c_{i_1\dots i_d}| \leq \sum_{t=0}^{d} \sum_{\substack{i_1,\dots,i_d=1 \\ t_{i_1\dots i_d}=t}}^{p+1} \left(\sum_{K \in \mathcal{G}^\ell_d} |c_{i_1\dots i_d}| \right) \leq C_2 \left(\sum_{t=0}^{d} (\ell+1)^t (p+1)^{2(d-t)} \right) \|u\|_{\mathcal{J}^d_{\underline{\gamma}}(Q)}.
$$

□

A.9.3 Proof of Theorem 2.1

Proof of Theorem 2.1. Fix A_f, C_f, and $\underline{\gamma}$ as in the hypotheses. Then, by Proposition A.19, there exists c_p, C_{hp}, $b_{\mathsf{hp}} > 0$ such that for every $\ell \in \mathbb{N}$ and for all $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}}(Q; \mathcal{C}, \mathcal{E}; C_f, A_f)$, there exists $v^{\ell}_{\mathsf{hp}} \in X^{\ell, c_p\ell}_{\mathsf{hp},d}$ such that (see Section A.1 for the definition of the space $X^{\ell, c_p\ell}_{\mathsf{hp},d}$)

$$\|v - v^{\ell}_{\mathsf{hp}}\|_{H^1(Q)} \leq C_{\mathsf{hp}} e^{-b_{\mathsf{hp}}\ell}.$$

For $\varepsilon > 0$, we choose

$$L := \left\lceil \frac{1}{b_{\mathsf{hp}}} \left|\log(\varepsilon / C_{\mathsf{hp}})\right| \right\rceil, \tag{A.73}$$

so that

$$\|v - v^{L}_{\mathsf{hp}}\|_{H^1(Q)} \leq \varepsilon.$$

Furthermore, $v^L_{\mathsf{hp}} = \sum_{i_1,\dots,i_d}^{N_{1\mathrm{d}}} c_{i_1\dots i_d}\phi_{i_1\dots i_d}$ and, for all $(i_1, \dots, i_d) \in \{1, \dots, N_{1\mathrm{d}}\}^d$, there exists v_{i_j}, $j = 1, \dots, d$ such that $\phi_{i_1\dots i_d} = \bigotimes_{j=1}^d v_{i_j}$, see Section A.9.2 and Proposition A.24. By construction of v_i in (A.71), and by using (A.47) and (A.48), we observe that $\|v_i\|_{L^\infty(I)} \leq 1$ for all $i = 1, \dots, N_{1\mathrm{d}}$. In addition, (A.50), demonstrates that

$$\|v_i\|_{H^1(I)} \leq \frac{2}{|\mathrm{supp}(v_i)|^{1/2}\deg(v_i)^{1/2}} \leq 2\sigma^{-L/2} \qquad \forall i \in \{1, \dots, N_{1\mathrm{d}}\}.$$

Then, by (A.73),

$$\sigma^{-L} \leq \sigma^{-\frac{1}{b_{\mathsf{hp}}}\log(C_{\mathsf{hp}})} \varepsilon^{-\frac{1}{b_{\mathsf{hp}}}\log(1/\sigma)}.$$

This concludes the proof of Items 1 and 2. Finally, Item 3 follows from Proposition A.24 and the fact that $p \leq C_p\left(1 + |\log(\varepsilon)|\right)$ for a constant $C_p > 0$ independent of ε. □

A.10 Combination of multiple patches

The approximation results in the domain $Q = (0,1)^d$ can be generalized to include the combination of multiple patches. We give here an example, relevant for the PDEs considered in Section 5. For the sake of conciseness, we show a single construction that takes into account all singularities of the problems in Section 5. We will then use this construction to prove expression rate bounds for realizations of NNs.

Let $a > 0$ and $\Omega = (-a, a)^d$. Denote the set of corners

$$\mathcal{C}_\Omega = \bigtimes_{j=1}^{d} \{-a, 0, a\}, \tag{A.74}$$

and the set of edges

$$\mathcal{E}_\Omega = \begin{cases} \varnothing & \text{if } d = 2, \\ \bigcup_{j=1}^d \bigtimes_{k=1}^{j-1}\{-a,0,a\} \times \{(-a,-a/2), (-a/2, 0), (0, a/2), (a/2, a)\} \times \bigtimes_{k=j+1}^{d}\{-a,0,a\} & \text{if } d = 3. \end{cases} \tag{A.75}$$

We introduce the affine transformations $\psi_{1,+} : (0,1) \to (0, a/2)$, $\psi_{2,+} : (0,1) \to (a/2, a)$, $\psi_{1,-} : (0,1) \to (-a/2, 0)$, $\psi_{2,-} : (0,1) \to (-a, -a/2)$ such that

$$\psi_{1,\pm}(x) = \pm\frac{a}{2}x, \qquad \psi_{2,\pm}(x) = \pm\left(a - \frac{a}{2}x\right).$$

For all $\ell \in \mathbb{N}$, define then

$$\widetilde{\mathcal{G}}^{\ell}_1 = \bigcup_{i\in\{1,2\}, \star\in\{+.-\}} \psi_{i,\star}(\mathcal{G}^{\ell}_1).$$

Consequently, for $d = 2, 3$, denote $\widetilde{\mathcal{G}}^{\ell}_d = \{\bigtimes_{i=1}^d K_i : K_1, \dots, K_d \in \widetilde{\mathcal{G}}^{\ell}_1\}$, see Figure 3. The *hp* space in $\Omega = (-a, a)^d$ is then given by

$$\widetilde{X}^{\ell,p}_{\mathsf{hp},d} = \{v \in H^1(\Omega) : v_{|K} \in \mathbb{Q}_p(K), \text{ for all } K \in \widetilde{\mathcal{G}}^{\ell}_d\}.$$

Finally, recall the definition of $\pi^{\ell,p}_{\mathsf{hp}}$ from (A.12) and construct

$$\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} : W^{1,1}((-a,a)) \to \widetilde{X}^{\ell,p}_{\mathsf{hp},1}$$

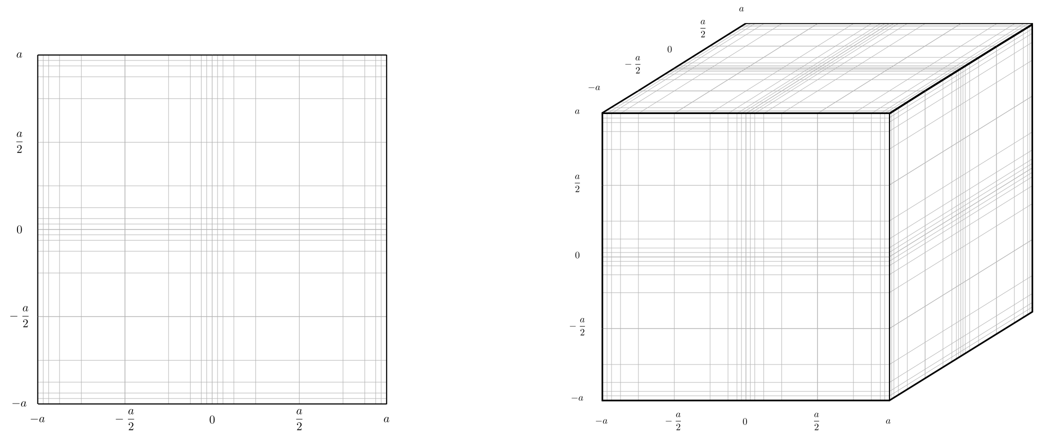

Figure 3: Multipatch geometric tensor product meshes $\widetilde{\mathcal{G}}^{\ell}_{d}$, for $d = 2$ (left) and $d = 3$ (right).

such that, for all $v \in W^{1,1}((-a,a))$,

$$\left(\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} v\right)|_{(0,\frac{a}{2})} = \left(\pi^{\ell,p}_{\mathsf{hp}}(v|_{(0,\frac{a}{2})} \circ \psi_{1,+})\right) \circ \psi^{-1}_{1,+}, \qquad \left(\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} v\right)|_{(\frac{a}{2},a)} = \left(\pi^{\ell,p}_{\mathsf{hp}}(v|_{(\frac{a}{2},a)} \circ \psi_{2,+})\right) \circ \psi^{-1}_{2,+},$$
$$\left(\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} v\right)|_{(-\frac{a}{2},0)} = \left(\pi^{\ell,p}_{\mathsf{hp}}(v|_{(-\frac{a}{2},0)} \circ \psi_{1,-})\right) \circ \psi^{-1}_{1,-}, \qquad \left(\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} v\right)|_{(-a,-\frac{a}{2})} = \left(\pi^{\ell,p}_{\mathsf{hp}}(v|_{(-a,-\frac{a}{2})} \circ \psi_{2,-})\right) \circ \psi^{-1}_{2,-}. \tag{A.76}$$

Then, the global hp projection operator $\widetilde{\Pi}^{\ell,p}_{\mathsf{hp},d} : W^{1,1}_{\mathrm{mix}}(\Omega) \to \widetilde{X}^{\ell,p}_{\mathsf{hp},d}$ is defined as

$$\widetilde{\Pi}^{\ell,p}_{\mathsf{hp},d} = \bigotimes_{i=1}^{d} \widetilde{\pi}^{\ell,p}_{\mathsf{hp}}.$$

Theorem A.25. *For $a > 0$, let $\Omega = (-a,a)^d$, $d = 2,3$. Denote by Ω^k, $k = 1,\dots,4^d$ the patches composing Ω, i.e., the sets $\Omega^k = \bigtimes_{j=1}^{d}(a^k_j, a^k_j + a/2)$ with $a^k_j \in \{-a, -a/2, 0, a/2\}$. Denote also $\mathcal{C}^k = \mathcal{C}_\Omega \cap \overline{\Omega}^k$ and $\mathcal{E}^k = \{e \in \mathcal{E}_\Omega : e \subset \overline{\Omega}^k\}$, which contain one singular corner, and three singular edges abutting that corner, as in* (A.1) *and* (A.2).

Let $\mathcal{I} \subset \{1,\dots,4^d\}$ and let $v \in W^{1,1}_{\mathrm{mix}}(\Omega)$ be such that, for all $k \in \mathcal{I}$, there holds $v_{|\Omega^k} \in \mathcal{J}^{\varpi}_{\underline{\gamma}^k}(\Omega^k; \mathcal{C}^k, \mathcal{E}^k)$ with

$$\begin{aligned} &\gamma^k_c > 1, \text{ for all } c \in \mathcal{C}^k, && \text{if } d = 2,\\ &\gamma^k_c > 3/2 \text{ and } \gamma^k_e > 1, \text{ for all } c \in \mathcal{C}^k \text{ and } e \in \mathcal{E}^k, && \text{if } d = 3. \end{aligned}$$

Then, there exist constants $c_p > 0$ and $C, b > 0$ such that, for all $\ell \in \mathbb{N}$, with $p = c_p \ell$,

$$\|v - \widetilde{\Pi}^{\ell,p}_{\mathsf{hp},d} v\|_{H^1(\Omega^k)} \le C e^{-b\ell} \le C \exp(-b \sqrt[2d]{N_{\mathrm{dof}}}). \tag{A.77}$$

Here, $N_{\mathrm{dof}} = \mathcal{O}(\ell^{2d})$ denotes the overall number of degrees of freedom in the piecewise polynomial approximation. Furthermore, writing $\widetilde{N}_{1\mathrm{d}} = 4(\ell+1)p + 1$, there exists an array of coefficients

$$\widetilde{c} = \left\{ \widetilde{c}_{i_1\dots i_d} : (i_1,\dots,i_d) \in \{1,\dots,\widetilde{N}_{1\mathrm{d}}\}^d \right\}$$

such that

$$\left(\widetilde{\Pi}^{\ell,p}_{\mathsf{hp},d} v\right)(x_1,\dots,x_d) = \sum_{i_1,\dots,i_d=1}^{N_{1\mathrm{d}}} \widetilde{c}_{i_1\dots i_d} \prod_{j=1}^{d} \tilde{v}_{i_j}(x_j) \qquad \forall (x_1,\dots,x_d) \in \Omega,$$

where for all $j = 1,\dots,d$ and $i_j = 1,\dots,\widetilde{N}_{1\mathrm{d}}$, $\tilde{v}_{i_j} \in \widetilde{X}^{\ell,p}_{\mathsf{hp},1}$ with support in at most two, neighboring elements of $\widetilde{\mathcal{G}}^{\ell}_1$. Finally, there exist constants $C_1, C_2 > 0$ independent of ℓ such that

$$\|v_i\|_{H^1((-a,a))} \le C_1 \sigma^{-\ell/2} \qquad \forall i = 1,\dots,\widetilde{N}_{1\mathrm{d}}, \tag{A.78}$$

and

$$\sum_{i_1,\dots,i_d=1}^{\widetilde{N}_{1\mathrm{d}}} |\widetilde{c}_{i_1\dots i_d}| \le C_2 \sum_{j=0}^{d} (\ell+1)^j (p+1)^{2(d-j)} \|v\|_{W^{1,1}_{\mathrm{mix}}(\Omega)}. \tag{A.79}$$

Proof. The statement is a direct consequence of Propositions A.19 and A.24. We start the proof by showing that for any function $v \in W^{1,1}_{\text{mix}}(\Omega)$, the approximation $\widetilde{\Pi}^{\ell,p}_{\mathsf{hp},d} v$ is continuous; the rest of the theorem will then follow from the results in each sub-patch. Let now $w \in W^{1,1}((-a,a))$. Then, it holds that $\left(\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} w\right)|_I \in C(I)$, for all $I \in \{(0,a/2),(a/2,a),(-a/2,0),(-a,-a/2)\}$, by definition (A.76). Furthermore, by the nodal exactness of the local projectors, for $\tilde{x} \in \{-a/2, 0, a/2\}$, there holds

$$\lim_{x\to\tilde{x}^-} (\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} w)(x) = w(\tilde{x}) = \lim_{x\to\tilde{x}^+} (\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} w)(x),$$

implying then that $\widetilde{\pi}^{\ell,p}_{\mathsf{hp}} w$ is continuous. Since $\widetilde{\Pi}^{\ell,p}_{\mathsf{hp},d} = \bigotimes_{j=1}^d \widetilde{\pi}^{\ell,p}_{\mathsf{hp}}$, this implies that $\widetilde{\Pi}^{\ell,p}_{\mathsf{hp},d} v$ is continuous for all $v \in W^{1,1}_{\text{mix}}(\Omega)$. Fix $k \in \{1,\dots,4^d\}$ such that $v \in \mathcal{J}^{\varpi}_{\underline{\gamma}^k}(\Omega^k; \mathcal{C}^k, \mathcal{E}^k)$. There exist then, by Proposition A.19, constants $C, b, c_p > 0$ such that for all $\ell \in \mathbb{N}$

$$\|v - \widetilde{\Pi}^{\ell,c_p\ell}_{\mathsf{hp},d}\|_{H^1(\Omega^k)} \le C e^{-b\ell}.$$

Equation (A.77) follows. The bounds (A.78) and (A.79) follow from the construction of the basis functions (A.47)–(A.48) and from the application of Lemma A.21 in each patch, respectively. □

B Proofs of Section 5

B.1 Proof of Lemma 5.5

Proof of Lemma 5.5. For any two sets $X, Y \subset \Omega$, we denote by $\operatorname{dist}_\Omega(X,Y)$ the infimum of Euclidean lengths of paths in Ω connecting an element of X with one of Y. We introduce several domain-dependent quantities to be used in the construction of the triangulation $\mathcal{T}$ with the properties stated in the lemma.

Let $\mathcal{E}$ denote the set of edges of the polygon Ω. For each corner $c \in \mathcal{C}$ at which the interior angle of Ω is smaller than π (below called *convex corner*), we fix a parallelogram $G_c \subset \Omega$ and a bijective, affine transformation $F_c : (0,1)^2 \to G_c$ such that

- $F_c((0,0)) = c$,
- two edges of G_c coincide partially with the edges of Ω abutting at the corner c.

If at c the interior angle of Ω is greater than or equal to π (both are referred to by slight abuse of terminology as *nonconvex corner*), the same properties hold, with $F_c : (-1,1)\times(0,1) \to G_c$ if the interior angle equals π, and $F_c : (-1,1)^2 \setminus (-1,0]^2 \to G_c$ else, and with G_c having the corresponding shape. Let now

$$d_{c,1} \coloneqq \sup\{r > 0 : B_r(c) \cap \Omega \subset G_c\}, \qquad d_{\mathcal{C},1} \coloneqq \min_{c\in\mathcal{C}} d_{c,1}.$$

Then, for each $c \in \mathcal{C}$, let e_1 and e_2 be the edges abutting c, and define

$$d_{c,2} \coloneqq \operatorname{dist}_\Omega\left(e_1 \cap \left(B_{\frac{\sqrt{2}}{\sqrt{2}+1} d_{\mathcal{C},1}}(c)\right)^c, e_2 \cap \left(B_{\frac{\sqrt{2}}{\sqrt{2}+1} d_{\mathcal{C},1}}(c)\right)^c\right), \qquad d_{\mathcal{C},2} \coloneqq \min_{c\in\mathcal{C}} d_{c,2}.$$

Furthermore, for each $e \in \mathcal{E}$, denote $d_e := \infty$ if Ω is a triangle, otherwise

$$d_e \coloneqq \min\{\operatorname{dist}_\Omega(e, e_1) : e_1 \in \mathcal{E} \text{ and } \overline{e} \cap \overline{e_1} = \varnothing\}, \qquad d_{\mathcal{E}} = \min_{e\in\mathcal{E}} d_e.$$

Finally, for all $x \in \Omega$, let

$$n_e(x) \coloneqq \#\{e_1, e_2, \ldots \in \mathcal{E} : \operatorname{dist}_\Omega(x, \partial\Omega) = \operatorname{dist}_\Omega(x, e_1) = \operatorname{dist}_\Omega(x, e_2) = \ldots\}.$$

Then, in case Ω is a triangle, let d_0 be half of the radius of the inscribed circle, else let $d_0 := \frac{1}{3} d_{\mathcal{E}} < \frac{1}{2} d_{\mathcal{E}}$. It holds that

$$\operatorname{dist}_\Omega(\{x \in \Omega : n_e(x) \ge 3\}, \partial\Omega) \ge d_0 > 0.$$

For any shape regular triangulation $\mathcal{T}$ of $\mathbb{R}^2$, such that for all $K \in \mathcal{T}$, $K \cap \partial\Omega = \varnothing$, denote $\mathcal{T}_\Omega = \{K \in \mathcal{T} : K \subset \Omega\}$ and $h(\mathcal{T}_\Omega) = \max_{K\in\mathcal{T}_\Omega} h(K)$, where $h(K)$ denotes the diameter of K. Denote by $\mathcal{N}_\Omega$ the set of nodes of $\mathcal{T}$ that are in $\overline{\Omega}$. For any $n \in \mathcal{N}_\Omega$, define

$$\operatorname{patch}(n) \coloneqq \operatorname{int}\left(\bigcup_{K\in\mathcal{T}: n\in\overline{K}} \overline{K}\right).$$

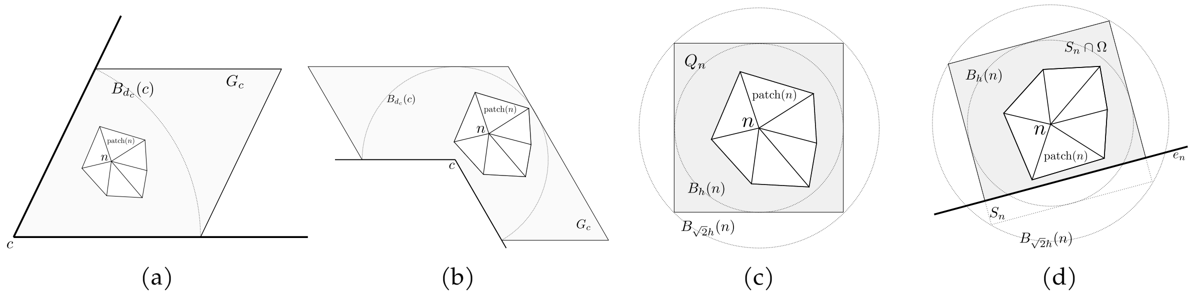

Figure 4: Patches Ω_n for nodes near a convex (a), nonconvex corner (b), for nodes in the interior of Ω (c), and near an edge (d).

Let $\mathcal{T}$ be a triangulation of $\mathbb{R}^2$ such that

$$h(\mathcal{T}_\Omega) \leq \min\left(\frac{d_0}{\sqrt{2}}, \frac{d_{\mathcal{C},1}}{\sqrt{2}+1}, \frac{d_{\mathcal{C},2}}{2\sqrt{2}}, \frac{d_{\mathcal{E}}}{2\sqrt{2}}\right), \tag{B.1}$$

and such that for all $K \in \mathcal{T}$ it holds $K \cap \partial\Omega = \varnothing$.

The hat-function basis $\{\phi_n\}_{n\in\mathcal{N}_\Omega}$ is a basis for $\mathbb{P}_1(\mathcal{T}_\Omega)$ such that $\operatorname{supp}(\phi_n) \subset \overline{\operatorname{patch}(n)}$ for all $n \in \mathcal{N}_\Omega$, and it is a partition of unity.

We will show that, for each $n \in \mathcal{N}_\Omega$, there exists a subdomain Ω_n with the desired properties, such that $\operatorname{patch}(n)\cap\Omega \subset \Omega_n$. We point to Figure 4 for an illustration of the patches Ω_n that will be introduced in the proof, for different sets of nodes.

For each $c \in \mathcal{C}$, let $\widehat{\mathcal{N}}_c = \{n \in \mathcal{N}_\Omega : \operatorname{patch}(n) \cap \Omega \subset G_c\}$. There holds

$$\mathcal{N}_c := \{n \in \mathcal{N}_\Omega : \operatorname{dist}_\Omega(n, c) \leq d_{\mathcal{C},1} - h(\mathcal{T}_\Omega)\} \subset \widehat{\mathcal{N}}_c.$$

Therefore, all the nodes $n \in \mathcal{N}_c$ are such that $\operatorname{patch}(n) \cap \Omega \subset G_c =: \Omega_n$. Denote then

$$\mathcal{N}_{\mathcal{C}} = \bigcup_{c\in\mathcal{C}} \mathcal{N}_c.$$

Note that, due to (B.1), there holds $\sqrt{2}h(\mathcal{T}_\Omega) \leq \frac{\sqrt{2}}{\sqrt{2}+1} d_{\mathcal{C},1} \leq d_{\mathcal{C},1} - h(\mathcal{T}_\Omega)$.

We consider the nodes in $\mathcal{N} \setminus \mathcal{N}_{\mathcal{C}}$. First, consider the nodes in

$$\mathcal{N}_0 := \{n \in \mathcal{N} \setminus \mathcal{N}_{\mathcal{C}} : \operatorname{dist}_\Omega(n, \partial\Omega) \geq \sqrt{2}h(\mathcal{T}_\Omega)\}.$$

For all $n \in \mathcal{N}_0$, there exists a square Q_n such that

$$\operatorname{patch}(n) \subset B_{h(\mathcal{T}_\Omega)}(n) \subset Q_n \subset B_{\sqrt{2}h(\mathcal{T}_\Omega)}(n) \subset \Omega,$$

see Figure 4c. Hence, for all $n \in \mathcal{N}_0$, we take $\Omega_n := Q_n$. Define

$$\mathcal{N}_{\mathcal{E}} := \mathcal{N} \setminus (\mathcal{N}_0 \cup \mathcal{N}_{\mathcal{C}}) = \left\{n \in \mathcal{N} : \operatorname{dist}_\Omega(n, c) > d_{\mathcal{C},1} - h(\mathcal{T}_\Omega), \forall c \in \mathcal{C}, \text{ and } \operatorname{dist}_\Omega(n, \partial\Omega) < \sqrt{2}h(\mathcal{T}_\Omega)\right\}.$$

For all $n \in \mathcal{N}_{\mathcal{E}}$, from (B.1) it follows that $\operatorname{dist}_\Omega(n, \partial\Omega) < \sqrt{2}h(\mathcal{T}_\Omega) \leq d_0$, hence $n_e(n) \leq 2$. Furthermore, suppose there exists $n \in \mathcal{N}_{\mathcal{E}}$ such that $n_e(n) = 2$. Let the two closest edges to n be denoted by e_1 and e_2, so that $\operatorname{dist}_\Omega(n, e_1) = \operatorname{dist}_\Omega(n, e_2) = \operatorname{dist}_\Omega(n, \partial\Omega) < \sqrt{2}h(\mathcal{T}_\Omega)$. If $\overline{e_1} \cap \overline{e_2} = \varnothing$, there must hold $\operatorname{dist}_\Omega(n, e_1) + \operatorname{dist}_\Omega(n, e_2) \geq d_{\mathcal{E}}$, which is a contradiction with $\operatorname{dist}_\Omega(n, \partial\Omega) < \sqrt{2}h(\mathcal{T}_\Omega) \leq d_{\mathcal{E}}/2$. If instead there exists $c \in \mathcal{C}$ such that $\overline{e_1} \cap \overline{e_2} = \{c\}$, then n is on the bisector of the angle between e_1 and e_2. Using that $2\sqrt{2}h(\mathcal{T}_\Omega) \leq d_{\mathcal{C},2}$, we now show that all such nodes belong either to $\mathcal{N}_{\mathcal{C}}$ or to $\mathcal{N}_0$, which is a contradiction to $n \in \mathcal{N}_{\mathcal{E}}$. Let $x_0 \in \Omega$ be the intersection of $B_{\frac{\sqrt{2}}{\sqrt{2}+1}d_{\mathcal{C},1}}(c)$ and the bisector. To show that $n \in \mathcal{N}_{\mathcal{C}} \cup \mathcal{N}_0$, it suffices to show that $\operatorname{dist}(x_0, e_i) \geq \sqrt{2}h(\mathcal{T}_\Omega)$ for $i = 1, 2$. Because $\frac{\sqrt{2}}{\sqrt{2}+1} d_{\mathcal{C},1} \leq d_{\mathcal{C},1} - h(\mathcal{T}_\Omega)$, it a fortiori holds for all points y in Ω on the bisector intersected with $\left(B_{d_{\mathcal{C},1}-h(\mathcal{T}_\Omega)}(c)\right)^c$, that $\operatorname{dist}(y, e_i) \geq \sqrt{2}h(\mathcal{T}_\Omega)$, which shows that if $\operatorname{dist}_\Omega(n, c) \geq d_{\mathcal{C},1} - h(\mathcal{T}_\Omega)$, then $n \in \mathcal{N}_0$. If c is a nonconvex corner, then $\operatorname{dist}(x_0, e_i) \geq \sqrt{2}h(\mathcal{T}_\Omega)$ for $i = 1, 2$ follows immediately from $\operatorname{dist}(x_0, e_i) = \operatorname{dist}(x_0, c) = \frac{\sqrt{2}}{\sqrt{2}+1} d_{\mathcal{C},1}$ and (B.1). To show that $\operatorname{dist}(x_0, e_i) \geq \sqrt{2}h(\mathcal{T}_\Omega)$, $i = 1, 2$ in case c is a convex corner, we make the following definitions (see Figure 5):

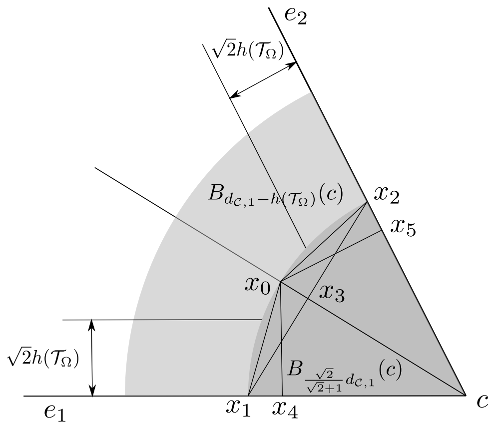

Figure 5: Situation near a convex corner c.

- For $i = 1, 2$, let x_i be the intersection of e_i and $B_{\frac{\sqrt{2}}{\sqrt{2}+1}d_{\mathcal{C},1}}(c)$,
- let x_3 be the intersection of $\overline{x_1x_2}$ with the bisector,
- and for $i = 1, 2$, let x_{i+3} be the orthogonal projection of x_0 onto e_i, which is an element of e_i because c is a convex corner.

Then $d_{c,2} = |\overline{x_1x_2}| = |\overline{x_1x_3}| + |\overline{x_3x_2}| = 2|\overline{x_ix_3}|$. Because the triangle cx_0x_{i+3} is congruent to cx_1x_3, it follows that $\operatorname{dist}(x_0, e_i) = |\overline{x_0x_{i+3}}| = |\overline{x_ix_3}| = \frac{1}{2}d_{c,2} \geq \sqrt{2}h(\mathcal{T}_\Omega)$. We can conclude with (B.1) that $n_e(n) = 1$ for all $n \in \mathcal{N}_\mathcal{E}$ and denote the edge closest to n by e_n. Let then S_n be the square with two edges parallel to e_n such that

$$\operatorname{patch}(n) \subset B_{h(\mathcal{T}_\Omega)}(n) \subset S_n \subset B_{\sqrt{2}h(\mathcal{T}_\Omega)}(n),$$

see Figure 4d, i.e. S_n has center n and sides of length $2h(\mathcal{T}_\Omega)$. For each $n \in \mathcal{N}_\mathcal{E}$, the connected component of $S_n \cap \Omega$ containing n is a rectangle:

(i) Note that for all edges e such that $\overline{e} \cap \overline{e_n} = \varnothing$, it holds that $S_n \cap e \subset B_{\sqrt{2}h(\mathcal{T}_\Omega)}(n) \cap e = \varnothing$. The latter holds because $\sqrt{2}h(\mathcal{T}_\Omega) \leq \frac{1}{2}d_\mathcal{E}$ and $\operatorname{dist}_\Omega(n, e_n) < \sqrt{2}h(\mathcal{T}_\Omega)$ imply $\operatorname{dist}_\Omega(n, e) \geq \sqrt{2}h(\mathcal{T}_\Omega)$.

(ii) From the previously given geometric argument considering a convex corner c and the two neighboring edges e_1 and e_2, showing that $\operatorname{dist}(x_0, e_i) \geq \sqrt{2}h(\mathcal{T}_\Omega)$ for $i = 1, 2$, we can additionally conclude that there is no $x \in \Omega \setminus B_{\frac{\sqrt{2}}{\sqrt{2}+1}d_{\mathcal{C},1}}(c)$ for which $\operatorname{dist}(x, e_n) < \sqrt{2}h(\mathcal{T}_\Omega)$ and such that there exists another edge e so that $\overline{e_n} \cap \overline{e} \neq \varnothing$ and $\operatorname{dist}(x, e) < \sqrt{2}h(\mathcal{T}_\Omega)$. This implies that $S_n \cap \partial\Omega \subset e_n$ or $S_n \cap \partial\Omega = \varnothing$.

Thus, the connected component of $S_n \cap \Omega$ containing n is a rectangle, which we define to be Ω_n.

Setting $N_p := \#\mathcal{N}_\Omega$ and $\{\Omega_i\}_{i=1,\dots,N_p} = \{\Omega_n\}_{n\in\mathcal{N}_\Omega}$ concludes the proof. □

B.2 Proof of Lemma 5.15

Proof of Lemma 5.15. Let $d = 3$ and denote $R = (-1, 0)^3$. Denote by O the origin, and let $E = \{e_1, e_2, e_3\}$ denote the set of edges of R abutting the origin. Let also $F = \{f_1, f_2, f_3\}$ denote the set of faces of R abutting the origin, i.e., the faces of R such that $f_i \subset \overline{R} \cap \overline{\Omega}_F$, $i = 1, 2, 3$. Let, finally, for each $f \in F$, $E_f = \{e \in E : e \subset \overline{f}$ denote the subset of E containing the edges neighboring f.

For each $e \in E$, define u_e to be the lifting of $u|_e$ into R, i.e., the function such that $u_e|_e = u_e$ and u_e is constant in the two coordinate directions perpendicular to e. Similarly, let, for each $f \in F$, u_f be such that $u_f|_f = u|_f$ and u_f is constant in the direction perpendicular to f.

We define $w : R \to \mathbb{R}$ as

$$w = u_0 + \sum_{e\in E}\big(u_e - u_0\big) + \sum_{f\in F}\Big(u_f - u_0 - \sum_{e\in E_f}(u_e - u_0)\Big) = u_0 - \sum_{e\in E} u_e + \sum_{f\in F} u_f, \tag{B.2}$$

where $u_0 = u(O)$. Since $u|_e \in W^{1,1}(e)$, $u|_f \in W^{1,1}_{\mathrm{mix}}(f)$ for all $e \in E$ and $f \in F$, there holds $u_e \in W^{1,1}_{\mathrm{mix}}(R)$ and $u_f \in W^{1,1}_{\mathrm{mix}}(R)$ for all $e \in E$ and $f \in F$ (cf. Equations (A.56) and (A.60)), hence $w \in W^{1,1}_{\mathrm{mix}}(R)$. Furthermore, note that

$$\big(u_e - u_0\big)|_{\widetilde{e}} = 0, \quad \text{for all } E \ni \widetilde{e} \neq e$$

and that

$$\big(u_f - u_0 - \sum_{e\in E_f} (u_e - u_0)\big)|_{\widetilde{f}} = 0, \quad \text{for all } F \ni \widetilde{f} \neq f.$$

From the first equality in (B.2), then, it follows that, for all $f \in F$,

$$w|_f = u_0 + \sum_{e\in E_f} \big(u_e|_f - u_0\big) + u_f|_f - u_0 - \sum_{e\in E_f} \big(u_e|_f - u_0\big) = u|_f.$$

Let the function v be defined as

$$v|_R = w, \qquad v|_{\Omega_F} = u. \tag{B.3}$$

Then, v is continuous in $(-1,1)^3$ and $v \in W^{1,1}_{\mathrm{mix}}((-1,1)^3)$. Now, for all $\alpha \in \mathbb{N}^3_0$ such that $|\alpha|_\infty \leq 1$,

$$\|\partial^\alpha u_e\|_{L^1(R)} = \|\partial^{\alpha^e_\parallel} u_e\|_{L^1(R)} = \|\partial^{\alpha^e_\parallel} u\|_{L^1(e)}, \qquad \forall e \in E,$$

where $\alpha^e_\parallel$ denotes the index in the coordinate direction parallel to e, and

$$\|\partial^\alpha u_f\|_{L^1(R)} = \|\partial^{\alpha^f_{\parallel,1}} \partial^{\alpha^f_{\parallel,2}} u_f\|_{L^1(R)} = \|\partial^{\alpha^f_{\parallel,1}} \partial^{\alpha^f_{\parallel,2}} u\|_{L^1(f)}, \qquad \forall f \in F,$$

where $\alpha^f_{\parallel,j}$, $j = 1, 2$ denote the indices in the coordinate directions parallel to f. Then, by a trace inequality (see [42, Lemma 4.2]), there exists a constant $C > 0$ independent of u such that

$$\|u_e\|_{W^{1,1}_{\mathrm{mix}}(R)} \leq C\|u\|_{W^{1,1}_{\mathrm{mix}}(\Omega_F)}, \qquad \|u_f\|_{W^{1,1}_{\mathrm{mix}}(R)} \leq C\|u\|_{W^{1,1}_{\mathrm{mix}}(\Omega_F)},$$

for all $e \in E$, $f \in F$. Then, by (B.2) and (B.3),

$$\|v\|_{W^{1,1}_{\mathrm{mix}}((-1,1)^d)} \leq C\|u\|_{W^{1,1}_{\mathrm{mix}}(\Omega_F)},$$

for an updated constant C independent of u. This concludes the proof when $d = 3$. The case $d = 2$ can be treated by the same argument. ☐

References

[1] I. Babuška and B. Guo. The *h-p* version of the finite element method for domains with curved boundaries. *SIAM J. Numer. Anal.*, 25(4):837–861, 1988.

[2] J. Berner, P. Grohs, and A. Jentzen. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. *SIAM J. Math. Data Sci.*, 2(3):631–657, 2020.

[3] P. Bolley, M. Dauge, and J. Camus. Régularité Gevrey pour le problème de Dirichlet dans des domaines à singularités coniques. *Comm. Partial Differential Equations*, 10(4):391–431, 1985.

[4] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[5] M. Costabel, M. Dauge, and S. Nicaise. Analytic regularity for linear elliptic systems in polygons and polyhedra. *Math. Models Methods Appl. Sci.*, 22(8):1250015, 63, 2012.

[6] W. E and B. Yu. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat.*, 6(1):1–12, 2018.

[7] D. Elbrächter, P. Grohs, A. Jentzen, and C. Schwab. DNN expression rate analysis of high-dimensional PDEs: Application to option pricing. *arXiv preprint 1809.07669*, 2018.

[8] H.-J. Flad, R. Schneider, and B.-W. Schulze. Asymptotic regularity of solutions to Hartree-Fock equations with Coulomb potential. *Math. Methods Appl. Sci.*, 31(18):2172–2201, 2008.

[9] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Østergaard Sørensen. Analytic structure of solutions to multiconfiguration equations. *J. Phys. A*, 42(31):315208, 11, 2009.

[10] E. Gagliardo. Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in n variabili. *Rend. Sem. Mat. Univ. Padova*, 27:284–305, 1957.

[11] I. Gühring, G. Kutyniok, and P. Petersen. Error bounds for approximations with deep ReLU neural networks in $W^{s,p}$ norms. *Anal. Appl. (Singap.)*, 18(05):803–859, 2020.

[12] B. Guo and I. Babuška. On the regularity of elasticity problems with piecewise analytic data. *Adv. in Appl. Math.*, 14(3):307–347, 1993.

[13] B. Guo and I. Babuška. Regularity of the solutions for elliptic problems on nonsmooth domains in $\mathbb{R}^3$. I. Countably normed spaces on polyhedral domains. *Proc. Roy. Soc. Edinburgh Sect. A*, 127(1):77–126, 1997.

[14] B. Guo and I. Babuška. Regularity of the solutions for elliptic problems on nonsmooth domains in $\mathbb{R}^3$. II. Regularity in neighbourhoods of edges. *Proc. Roy. Soc. Edinburgh Sect. A*, 127(3):517–545, 1997.

[15] B. Guo and C. Schwab. Analytic regularity of Stokes flow on polygonal domains in countably weighted Sobolev spaces. *J. Comput. Appl. Math.*, 190(1-2):487–519, 2006.

[16] J. Han, L. Zhang, and W. E. Solving many-electron Schrödinger equation using deep neural networks. *J. Comput. Phys.*, 399:108929, 8, 2019.

[17] J. He, L. Li, J. Xu, and C. Zheng. ReLU deep neural networks and linear finite elements. *J. Comp. Math.*, 38, 2020.

[18] J. Hermann, Z. Schätzle, and F. Noé. Deep neural network solution of the electronic Schrödinger equation. *arXiv preprint arXiv:1909.08423*, 2019.

[19] A. Jentzen, D. Salimova, and T. Welti. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *arXiv preprint arXiv:1809.07321*, 2018.

[20] V. Kazeev, I. Oseledets, M. Rakhuba, and C. Schwab. QTT-Finite-Element approximation for multiscale problems I: model problems in one dimension. *Adv. Comput. Math.*, 43(2):411–442, 2017.

[21] V. Kazeev and C. Schwab. Quantized tensor-structured finite elements for second-order elliptic PDEs in two dimensions. *Numer. Math.*, 138(1):133–190, 2018.

[22] F. Laakmann and P. Petersen. Efficient approximation of solutions of parametric linear transport equations by ReLU DNNs. *arXiv preprint arXiv:2001.11441*, 2020.

[23] E. H. Lieb and B. Simon. The Hartree-Fock theory for Coulomb systems. *Comm. Math. Phys.*, 53(3):185–194, 1977.

[24] P.-L. Lions. Solutions of Hartree-Fock equations for Coulomb systems. *Comm. Math. Phys.*, 109(1):33–97, 1987.

[25] J. Lu, Z. Shen, H. Yang, and S. Zhang. Deep network approximation for smooth functions. *arXiv preprint arXiv:2001.03040*, 2020.

[26] L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis. DeepXDE: A deep learning library for solving differential equations. *arXiv e-prints*, page arXiv:1907.04502, July 2019.

[27] Y. Maday and C. Marcati. Analyticity and *hp* discontinuous Galerkin approximation of nonlinear Schrödinger eigenproblems. *arXiv e-prints*, page arXiv:1912.07483, Dec. 2019.

[28] Y. Maday and C. Marcati. Regularity and *hp* discontinuous Galerkin finite element approximation of linear elliptic eigenvalue problems with singular potentials. *Math. Models Methods Appl. Sci.*, 29(8):1585–1617, 2019.

[29] Y. Maday and C. Marcati. Weighted analyticity of Hartree-Fock eigenfunctions. Technical Report 2020-59, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2020.

[30] C. Marcati, M. Rakhuba, and C. Schwab. Tensor Rank bounds for Point Singularities in $\mathbb{R}^3$. Technical Report 2019-68, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2019.

[31] C. Marcati and C. Schwab. Analytic regularity for the incompressible Navier-Stokes equations in polygons. *SIAM J. Math. Anal.*, 52(3):2945–2968, 2020.

[32] V. Maz'ya and J. Rossmann. *Elliptic Equations in Polyhedral Domains*, volume 162 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, Rhode Island, April 2010.

[33] J. M. Melenk and I. Babuška. The partition of unity finite element method: basic theory and applications. *Comput. Methods Appl. Mech. Engrg.*, 139(1-4):289–314, 1996.

[34] J. A. A. Opschoor, P. C. Petersen, and C. Schwab. Deep ReLU networks and high-order finite element methods. *Analysis and Applications*, 18(05):715–770, 2020.

[35] J. A. A. Opschoor, C. Schwab, and J. Zech. Exponential ReLU DNN expression of holomorphic maps in high dimension. Technical Report 2019-35_rev1, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2020. Accepted for publication in *Constructive Approximation*.

[36] P. Petersen and F. Voigtlaender. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Netw.*, 108:296 – 330, 2018.

[37] D. Pfau, J. S. Spencer, A. G. de G. Matthews, and W. M. C. Foulkes. Ab-initio solution of the many-electron Schrödinger equation with deep neural networks, 2019. arXiv:1909.02487.

[38] M. Raissi and G. E. Karniadakis. Hidden physics models: machine learning of nonlinear partial differential equations. *J. Comput. Phys.*, 357:125–141, 2018.

[39] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.

[40] D. Schötzau and C. Schwab. Exponential convergence for hp-version and spectral finite element methods for elliptic problems in polyhedra. *Math. Models Methods Appl. Sci.*, 25(9):1617–1661, 2015.

[41] D. Schötzau and C. Schwab. Exponential convergence of hp-FEM for elliptic problems in polyhedra: mixed boundary conditions and anisotropic polynomial degrees. *Found. Comput. Math.*, 18(3):595–660, 2018.

[42] D. Schötzau, C. Schwab, and T. P. Wihler. hp-dGFEM for second-order elliptic problems in polyhedra I: Stability on geometric meshes. *SIAM J. Numer. Anal.*, 51(3):1610–1633, 2013.

[43] D. Schötzau, C. Schwab, and T. P. Wihler. hp-DGFEM for second order elliptic problems in polyhedra II: Exponential convergence. *SIAM J. Numer. Anal.*, 51(4):2005–2035, 2013.

[44] C. Schwab and J. Zech. Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ. *Anal. Appl. (Singap.)*, 17(1):19–55, 2019.

[45] H. Sheng and C. Yang. PFNN: A Penalty-Free Neural Network Method for Solving a Class of Second-Order Boundary-Value Problems on Complex Geometries. *arXiv e-prints*, page arXiv:2004.06490, Apr. 2020.

[46] Y. Shin, Z. Zhang, and G. E. Karniadakis. Error estimates of residual minimization using neural networks for linear PDEs. *arXiv e-prints*, page arXiv:2010.08019, Oct. 2020.

[47] J. Sirignano and K. Spiliopoulos. DGM: a deep learning algorithm for solving partial differential equations. *J. Comput. Phys.*, 375:1339–1364, 2018.

[48] T. Suzuki. Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: optimal rate and curse of dimensionality. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019.

[49] J. M. Tarela and M. V. Martínez. Region configurations for realizability of lattice piecewise-linear models. *Math. Comput. Modelling*, 30(11-12):17–27, 1999.

[50] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Netw.*, 94:103 – 114, 2017.

[51] D. Yarotsky. Optimal approximation of continuous functions by very deep ReLU networks. volume 75 of *Proceedings of Machine Learning Research*, pages 639–649. PMLR, 06–09 Jul 2018.