De Rham compatible Deep Neural Network FEM

Neural Networks 165 (2023) 721–739

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

De Rham compatible Deep Neural Network FEM

Marcello Longo [a], Joost A.A. Opschoor [a,*], Nico Disch [b], Christoph Schwab [a], Jakob Zech [b]

[a] Seminar for Applied Mathematics, ETH Zürich, Rämistrasse 101, CH–8092 Zürich, Switzerland
[b] IWR, Universität Heidelberg, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 14 January 2022
Received in revised form 2 June 2023
Accepted 4 June 2023
Available online 9 June 2023

Keywords:
De Rham complex
Finite Elements
Lavrentiev gap
Neural networks
PINNs

ABSTRACT

On general regular simplicial partitions $\mathcal{T}$ of bounded polytopal domains $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, we construct *exact neural network (NN) emulations* of all lowest order finite element spaces in the discrete de Rham complex. These include the spaces of piecewise constant functions, continuous piecewise linear (CPwL) functions, the classical "Raviart–Thomas element", and the "Nédélec edge element". For all but the CPwL case, our network architectures employ both ReLU (rectified linear unit) and BiSU (binary step unit) activations to capture discontinuities. In the important case of CPwL functions, we prove that it suffices to work with pure ReLU nets. Our construction and DNN architecture generalizes previous results in that no geometric restrictions on the regular simplicial partitions $\mathcal{T}$ of Ω are required for DNN emulation. In addition, for CPwL functions our DNN construction is valid in any dimension $d \geq 2$. Our "FE-Nets" are required in the variationally correct, structure-preserving approximation of boundary value problems of electromagnetism in nonconvex polyhedra $\Omega \subset \mathbb{R}^3$. They are thus an essential ingredient in the application of e.g., the methodology of "physics-informed NNs" or "deep Ritz methods" to electromagnetic field simulation via deep learning techniques. We indicate generalizations of our constructions to higher-order compatible spaces and other, non-compatible classes of discretizations, in particular the "Crouzeix–Raviart" elements and Hybridized, Higher Order (HHO) methods.

1. Introduction

Recent years have seen the emergence of Deep Neural Network (DNN) based methods for the numerical approximation of solutions to partial differential equations (PDEs for short). In one class of proposed methods, DNNs serve as *approximation architectures* in a suitable, weak form of the PDE of interest. In E and Yu (2018), for elliptic, self-adjoint PDEs the variational principle associated to the PDE is computationally minimized over suitable DNNs, so that the energy functional of the physical system of interest gives rise to a consistent loss function for the training of the DNN. Numerical solutions obtained from training the approximating DNN in this way correspond to approximate variational solutions of the PDE under consideration.

The recently promoted "physics-informed NNs" (PiNNs), e.g. Raissi, Perdikaris, and Karniadakis (2019), Yang, Meng, and Karniadakis (2021) and references there, insert DNN approximations with suitably smooth activations (e.g. softmax or tanh) as approximation architecture into the strong form of the governing PDE. Approximate solutions are obtained by numerical minimization of loss functions obtained by discretely enforcing smallness of the residual at collocation points in the spatio-temporal domain. While empirically successful in a large number of test cases, also DNN based approximations are subject to the fundamental paradigm that "stability and consistency implies convergence". A key factor of recent successful DNN deployment in numerical PDE solution is their excellent approximation properties, in particular on high-dimensional state- and parameter-spaces, e.g. Mhaskar and Poggio (2016), Petersen and Voigtlaender (2018), Schwab and Zech (2019) and the references there. High smoothness of DNNs with smooth activations may, however, preclude convergence of so-called "deep Ritz" approaches where loss functions in DNN training are derived from energies in variational principles (E & Yu, 2018), even for linear, deterministic and well posed PDEs.

To leverage the methodology of PiNNs and e.g., the variational Ritz method for computational electromagnetics, computational magneto-hydrodynamics etc., *structure-preserving DNNs* must be adopted. We provide here, therefore, *de Rham complex compatible DNN emulations* of the standard, lowest order finite element spaces on regular, simplicial triangulations of polytopal domains $\Omega \subset \mathbb{R}^d$, $d = 2, 3$. These spaces satisfy exact (de Rham) sequence properties, and also spawn discrete boundary complexes on $\partial\Omega$ that satisfy exact sequence properties for the surface

* Corresponding author.
E-mail addresses: marcello.longo@sam.math.ethz.ch (M. Longo), joost.opschoor@sam.math.ethz.ch (J.A.A. Opschoor), n.disch@stud.uni-heidelberg.de (N. Disch), christoph.schwab@sam.math.ethz.ch (C. Schwab), jakob.zech@uni-heidelberg.de (J. Zech).

https://doi.org/10.1016/j.neunet.2023.06.008

divergence and curl operators $\mathrm{div}_{\partial\Omega}$ and $\mathrm{curl}_{\partial\Omega}$. These in turn enable "neural boundary elements" for computational electromagnetism, recently proposed in Aylwin, Henríquez, and Schwab (2023).

1.1. Previous work

The connection between DNNs with Rectified Linear Unit (ReLU for short) activation and continuous, piecewise linear (CPwL) spline approximation spaces has been known for some time: *nodal discretizations* based on CPwL finite element methods (FEM) can be emulated by ReLU NNs (e.g. as introduced in Arora, Basu, Mianjy, & Mukherjee, 2018 and He, Li, Xu, & Zheng, 2020).

When CPwL finite elements are applied to, for example, weak formulations of the time-harmonic Maxwell equations, they are known to converge to the correct solution, generally, only for convex polygons or polyhedra: if Ω has re-entrant corners or edges, then with[1]

$$X_N(\Omega) := H^0(\mathrm{div},\Omega)\cap H^0(\mathrm{curl},\Omega)\cap\{u : u\times n = 0 \text{ on } \partial\Omega\},$$

where n is a unit normal vector to the boundary $\partial\Omega$ of Ω, the vector fields $[H^1(\Omega)]^3\cap X_N(\Omega)$ are closed in $X_N(\Omega)$ *without being dense*, see, e.g., Costabel (1991), Costabel, Dauge, and Nicaise (1999). For such nonconvex polyhedra, the weak solution to the time-harmonic Maxwell's equations is generally not contained in $[H^1(\Omega)]^3$.

Since any discrete conforming space based on a standard nodal finite element method is contained in $[H^1(\Omega)]^3$, nodal FEM in this situation converges to a wrong solution (in $[H^1(\Omega)]^3$) as the meshwidth tends to zero (respectively as the width of the corresponding NN tends to infinity) (Costabel & Dauge, 1999). Similar issues will arise for PiNN numerical approximations of low-regularity solutions for $H^0(\mathrm{curl},\Omega)$-based PDEs such as the time-harmonic Maxwell equations. They will persist also for DNN surrogates with more regular activation functions such as $ReLU^k$ for $k\in\mathbb{N}$ and sigmoidal or softmax activations. On bounded sets, such NNs realize Lipschitz continuous functions, which are in $[H^1(\Omega)]^3$ and therefore may converge to an incorrect solution.

A second broad class of variational models, where continuous nodal FEM may cause problems, are "deep Ritz" type approaches such as in E and Yu (2018), which attempt to minimize energy functionals. For certain nonlinear problems the so-called "Lavrentiev gap" incurred by CPwL approximation architectures is known to be a fundamental obstruction to obtain convergent families of discrete minimizers, see e.g., Ball (2001), Marcellini (1989), Zhikov (1983). Again, relaxing continuity below H^1-conformity is known to remedy this issue; see, e.g. Balci, Ortner, and Storn (2022) and the discussion and references there. Accordingly, in Section 7.2 of the present paper we present CR-Net, a DNN emulation of the Crouzeix–Raviart element with BiSU and ReLU activations, on general regular, simplicial partitions of polytopal domains $\Omega\subset\mathbb{R}^d$, $d\geq 2$, which, when used in a deep Ritz method style approach for variational problems, affords convergent sequences of DNN approximations of minimizers. CR-Net will also afford advantages in variational image segmentation (e.g. Chambolle & Pock, 2020 and the references there).

Structure preservation in scientific machine learning is also the topic of Trask, Huang, and Hu (2022). For machine learning models on graphs, a data driven exterior calculus is introduced which strongly enforces physical laws, e.g. those in the de Rham complex, while allowing for additional information to be learned from data.

[1] Definitions of the (standard) spaces $H^1(\Omega)$, $H^0(\mathrm{div},\Omega)$ and $H^0(\mathrm{curl},\Omega)$ are recalled in Section 1.4.2.

1.2. Contributions

The purpose of the present paper is the design of DNNs which emulate *exactly*, on arbitrary regular, simplicial partitions $\mathcal{T}$ of polytopal domains $\Omega\subset\mathbb{R}^d$, $d=2,3$, the FE spaces $\mathrm{S}^1_1(\mathcal{T},\Omega)$ (continuous, piecewise linear functions), $\mathrm{N}_0(\mathcal{T},\Omega)$ (the Nédélec element), $\mathrm{RT}_0(\mathcal{T},\Omega)$ (the Raviart–Thomas element) and $\mathrm{S}^0_0(\mathcal{T},\Omega)$ (the piecewise constant functions). The precise definitions of these spaces will be given in Section 1.4.

We provide constructions of DNNs based on a combination of ReLU (2.1) and BiSU (Binary Step Unit) (2.2) activations, which emulate these classical, lowest-order FE spaces in the de Rham complex on a regular, simplicial partition $\mathcal{T}$ of Ω. We underline that our construction of NNs which emulate, in particular, the classical "Courant Finite Elements" $\mathrm{S}^1_1(\mathcal{T},\Omega)$, as well as $\mathrm{S}^0_0(\mathcal{T},\Omega)$ and $\mathrm{RT}_0(\mathcal{T},\Omega)$, applies to polytopal domains Ω of any dimension $d\geq 2$. For the practically relevant space $\mathrm{S}^1_1(\mathcal{T},\Omega)$, the so-called "continuous, piecewise linear (CPwL) functions", we provide DNN constructions based on ReLU activation only, which work in arbitrary, finite dimension $d\geq 2$ (the univariate case $d=1$ being trivial).

Our constructions accommodate general, regular simplicial partitions $\mathcal{T}$ of Ω. In particular, apart from regularity of the simplicial partition $\mathcal{T}$ of the polytopal domain Ω, no further constraints of geometric nature are imposed on $\mathcal{T}$, in arbitrary dimension $d\geq 2$. Our results on ReLU NN emulation of CPwL functions in Section 4 therefore unify and quantitatively improve earlier ones such as, e.g., He et al. (2020, Section 3), which covered only CPwL FE spaces on particular triangulations of Ω. Our main results, Propositions 5.1 and 5.7 and Theorem 5.5 in Section 5, provide mathematically exact DNN realizations of the lowest order FE spaces in the exact sequence (1.2) on general regular, simplicial partitions of the contractible, polytopal domain Ω. In our main results using ReLU and BiSU activations, the network size scales linearly with the cardinality $|\mathcal{T}|$ of $\mathcal{T}$. For the ReLU NN emulation of CPwL functions, the network size is in general of the order $|\mathcal{T}|\log(|\mathcal{T}|)$, which can be improved to order $|\mathcal{T}|$ for shape regular meshes.

1.3. Layout

The structure of this paper is as follows. In Section 1.4 we introduce the de Rham complex. In Section 2, we recapitulate notation and basic definitions for the NNs which we consider. We also review a basic NN calculus that shall be used subsequently in order to derive several properties of the proposed NN architectures.

Sections 3 and 4 contain the core material of the paper: in Section 3, using ReLU and BiSU activations, we provide explicit emulations for bases of all the FE spaces considered in this paper, without geometric conditions on the regular triangulations $\mathcal{T}$ of Ω. In Section 4 we show that for the special case of the emulation of CPwL functions, the same can be achieved employing solely ReLU activations. In both sections, we show that the network size depends only moderately on the space dimension d and the shape regularity of the partition.

In Section 5 we combine NN emulations of basis functions from Sections 3 and 4 to provide emulations of the FE spaces and discuss the implications of our results for function approximation by NNs in the respective Sobolev spaces. Section 6 provides a construction of NN emulations for compatible spaces on the boundary $\Gamma=\partial\Omega$ of the polytopal domains. These spaces are required in the deep neural network approximation of boundary integral equations in electromagnetics, among others, as discussed in Buffa et al. (2020), Buffa, Hiptmair, von Petersdorff, and Schwab (2003) and the references there. Finally, in Section 7 we present conclusions and explain how our analysis may be extended to higher order polynomial spaces and to certain Finite Element families which are non-compatible with (1.1).

$$\mathbb{R} \xrightarrow{i} H^1(\Omega) \xrightarrow{\text{grad}} H^0(\text{curl}, \Omega) \xrightarrow{\text{curl}} H^0(\text{div}, \Omega) \xrightarrow{\text{div}} L^2(\Omega) \xrightarrow{o} \{0\}. \tag{1.1}$$

Box I.

1.4. Notation and Finite Element spaces

We recall definitions of the de Rham complex and corresponding lowest order FE spaces.

1.4.1. Meshes

The term "mesh" shall denote certain simplicial partitions of polyhedral domains $\Omega \subseteq \mathbb{R}^d$ for some $d \in \mathbb{N} = \{1, 2, \ldots\}$. To specify these, for $k \in \{0, \ldots, d\}$ we define a k-simplex T by $T = \text{conv}(\{a_0, \ldots, a_k\}) \subset \mathbb{R}^d$, for some $a_0, \ldots, a_k \in \mathbb{R}^d$ which do not all lie in one affine subspace of dimension $k-1$, and where

$$\text{conv}(Y) := \left\{ x = \sum_{y \in Y} \lambda_y y \,:\, \lambda_y > 0 \text{ and } \sum_{y \in Y} \lambda_y = 1 \right\}$$

denotes the open convex hull. By $|T|$ we will denote the k-dimensional Lebesgue measure of a k-simplex. We consider a simplicial mesh $\mathcal{T}$ on Ω of d-simplices, i.e. $\mathcal{T}$ satisfies that $\overline{\Omega} = \bigcup_{T \in \mathcal{T}} \overline{T}$ and $T \cap T' = \emptyset$, for all $T \neq T'$. We assume that $\mathcal{T}$ is a *regular* partition, i.e. for all distinct $T, T' \in \mathcal{T}$ it holds that $\overline{T} \cap \overline{T'}$ is the closure of a k-subsimplex of T for some $k \in \{0, \ldots, d-1\}$, i.e. there exist $a_0, \ldots, a_d \in \overline{\Omega}$ such that $T = \text{conv}(\{a_0, \ldots, a_d\})$ and $\overline{T} \cap \overline{T'} = \text{conv}(\{a_0, \ldots, a_k\})$. The *shape-regularity constant* $C_{\text{sh}} := C_{\text{sh}}(\mathcal{T})$ of a simplicial partition $\mathcal{T}$ of Ω is $C_{\text{sh}} := \max_{T \in \mathcal{T}} \frac{h_T}{r_T} > 0$. Here $h_T := \text{diam}(T)$ and r_T is the radius of the largest ball contained in $\overline{T}$. Let $\mathcal{V}$ be the set of vertices of $\mathcal{T}$. We also let $\mathcal{F}, \mathcal{E}$ be the sets of $(d-1)$- and 1-subsimplices of $\mathcal{T}$, whose elements are called *faces* and *edges*, respectively, that is

$$\begin{aligned}
\mathcal{F} &:= \{f \subset \overline{\Omega} \,:\, \exists T = \text{conv}(\{a_0, \ldots, a_d\}) \in \mathcal{T}, \exists i \in \{0, \ldots, d\} \\
&\qquad \text{with } f = \text{conv}(\{a_0, \ldots, a_d\} \backslash \{a_i\})\}, \\
\mathcal{E} &:= \{e \subset \overline{\Omega} \,:\, \exists T = \text{conv}(\{a_0, \ldots, a_d\}) \in \mathcal{T}, \exists i, j \in \{0, \ldots, d\}, \\
&\qquad i \neq j, \text{ with } e = \text{conv}(\{a_i, a_j\})\}.
\end{aligned}$$

We denote the boundary of Ω by $\partial\Omega$ and the *skeleton* of $\mathcal{T}$ by $\partial\mathcal{T} := \bigcup_{T \in \mathcal{T}} \partial T$.

1.4.2. De Rham complex

We write $H^1(\Omega)$, $H^0(\text{div}, \Omega)$, $H^0(\text{curl}, \Omega)$ to indicate the following Sobolev spaces

$$\begin{aligned}
H^1(\Omega) &:= \{v \in L^2(\Omega) : \nabla v \in [L^2(\Omega)]^d\}, \\
H^0(\text{curl}, \Omega) &:= \{v \in [L^2(\Omega)]^d : \text{curl}\, v \in [L^2(\Omega)]^d\} \quad \text{for } d = 2, 3, \\
H^0(\text{div}, \Omega) &:= \{v \in [L^2(\Omega)]^d : \text{div}\, v \in L^2(\Omega)\},
\end{aligned}$$

where the two-dimensional curl is defined as $\text{curl}\, v = (-\partial_1 v_2, \partial_2 v_1)$. These are Hilbert spaces. Specifically, let us assume that $\Omega \subset \mathbb{R}^3$ is a contractible Lipschitz[2] domain with connected boundary $\partial\Omega$. Then, it is well-known that the following de Rham complex is an exact sequence (e.g. Ern and Guermond (2021, Proposition 16.14) and the references there), see the equation given in Box I: Here, i denotes an injection and o denotes the zero operator.

[2] That is with boundary parametrized locally by Lipschitz continuous maps (Ern & Guermond, 2021, Definition 3.2).

1.4.3. First order discrete de Rham complex

Finite dimensional subspaces preserving this structure are usually required to fit into a *discrete de Rham complex* (e.g. (Ern & Guermond, 2021, Proposition 16.15), see the equation given in Box II). We next define these spaces.

Throughout, denote by $\mathbb{P}_k$ the set of polynomials of degree at most $k \in \mathbb{N} \cup \{0\}$. For a given regular triangulation $\mathcal{T}$ of a domain Ω, $S^1_1(\mathcal{T}, \Omega)$ is the class of continuous, piecewise linear (CPwL) functions on $\mathcal{T}$, i.e.

$$S^1_1(\mathcal{T}, \Omega) := \{v \in H^1(\Omega) \,:\, v|_T \in \mathbb{P}_1, \ \forall T \in \mathcal{T}\} \subset H^1(\Omega). \tag{1.3}$$

Moreover, $S^0_0(\mathcal{T}, \Omega)$ is the class of piecewise constant functions on the partition $\mathcal{T}$

$$S^0_0(\mathcal{T}, \Omega) := \{v \in L^2(\Omega) \,:\, v|_T \equiv c_T \in \mathbb{R}, \ \forall T \in \mathcal{T}\} \subset L^2(\Omega). \tag{1.4}$$

Next, we recall $\text{RT}_0(\mathcal{T}, \Omega)$, the lowest order Raviart–Thomas space. Define the vector-valued polynomial space $\mathbb{RT}_0 = (\mathbb{P}_0)^d \oplus x\mathbb{P}_0$ and, for all $f \in \mathcal{F}$, let n_f denote a unit normal to the face f. Denote by $[v \cdot n_f]_f$ the jump of the normal component of a vector field across f, that is

$$[v \cdot n_f]_f(x_0) = \lim_{\epsilon \searrow 0} (v(x_0 + \epsilon n_f) - v(x_0 - \epsilon n_f)) \cdot n_f \qquad \text{for all} \quad x_0 \in f.$$

The Raviart–Thomas finite element space of lowest order is (e.g. Ern and Guermond (2021, Section 14.1))

$$\begin{aligned}
\text{RT}_0(\mathcal{T}, \Omega) := \{&v \in (L^1(\Omega))^d \,:\, v|_T \in \mathbb{RT}_0 \ \forall T \in \mathcal{T} \text{ and} \\
&[v \cdot n_f]_f = 0 \ \forall f \in \mathcal{F}, f \subset \Omega\}.
\end{aligned} \tag{1.5}$$

This space has one degree of freedom per face $f \in \mathcal{F}$ and it satisfies $\text{RT}_0(\mathcal{T}, \Omega) \subset H^0(\text{div}, \Omega)$.

To define the lowest order Nédélec space $\text{N}_0(\mathcal{T}, \Omega)$, for $d = 2$ define $\mathbb{NE}_0 = (\mathbb{P}_0)^d \oplus \mathbb{P}_0(-x_2, x_1)$, and for $d = 3$ define $\mathbb{NE}_0 = (\mathbb{P}_0)^d \oplus x \times (\mathbb{P}_0)^d$. Let $[v \times n_f]_f$ denote the jump of the tangential component of a vector field across f, that is

$$[v \times n_f]_f(x_0) = \lim_{\epsilon \searrow 0} (v(x_0 + \epsilon n_f) - v(x_0 - \epsilon n_f)) \times n_f \qquad \text{for all} \quad x_0 \in f.$$

Then the Nédélec finite element space of lowest order (Ern & Guermond, 2021, Section 15.1) reads

$$\begin{aligned}
\text{N}_0(\mathcal{T}, \Omega) := \{&v \in (L^1(\Omega))^d \,:\, v|_T \in \mathbb{NE}_0 \ \forall T \in \mathcal{T} \text{ and} \\
&[v \times n_f]_f = 0 \ \forall f \in \mathcal{F}, f \subset \Omega\}.
\end{aligned} \tag{1.6}$$

This space has one degree of freedom per edge $e \in \mathcal{E}$ and it satisfies $\text{N}_0(\mathcal{T}, \Omega) \subset H^0(\text{curl}, \Omega)$. For $d = 2$, $\text{N}_0(\mathcal{T}, \Omega)$ is closely related to $\text{RT}_0(\mathcal{T}, \Omega)$ (see (3.13)). For each of the spaces in (1.2), we state in Section 3 a basis of the space. The FE spaces from (1.2) have the advantage of being *conforming*, i.e., they are finite dimensional spaces, each strictly contained in the respective Sobolev space in (1.1). Furthermore, the ($\mathcal{T}$-dependent) projections $\Pi_{S^1_1}, \Pi_{\text{N}_0}, \Pi_{\text{RT}_0}, \Pi_{S^0_0}$ on these subspaces introduced in Ern and Guermond (2021, Sec. 19.3) commute with the differential operators as shown in the following diagram (Ern & Guermond, 2021, Lemma 19.6):

$$\begin{array}{ccccccc}
H^1(\Omega) & \xrightarrow{\text{grad}} & H^0(\text{curl}, \Omega) & \xrightarrow{\text{curl}} & H^0(\text{div}, \Omega) & \xrightarrow{\text{div}} & L^2(\Omega) \\
\downarrow{\scriptstyle \Pi_{S^1_1}} & & \downarrow{\scriptstyle \Pi_{\text{N}_0}} & & \downarrow{\scriptstyle \Pi_{\text{RT}_0}} & & \downarrow{\scriptstyle \Pi_{S^0_0}} \\
S^1_1(\mathcal{T}, \Omega) & \xrightarrow{\text{grad}} & \text{N}_0(\mathcal{T}, \Omega) & \xrightarrow{\text{curl}} & \text{RT}_0(\mathcal{T}, \Omega) & \xrightarrow{\text{div}} & S^0_0(\mathcal{T}, \Omega)
\end{array}$$

$$\mathbb{R} \xrightarrow{i} \mathrm{S}^1_1(\mathcal{T}, \Omega) \xrightarrow{\mathrm{grad}} \mathrm{N}_0(\mathcal{T}, \Omega) \xrightarrow{\mathrm{curl}} \mathrm{RT}_0(\mathcal{T}, \Omega) \xrightarrow{\mathrm{div}} \mathrm{S}^0_0(\mathcal{T}, \Omega) \xrightarrow{o} \{0\}. \tag{1.2}$$

Box II.

For these reasons we say that these spaces are *de Rham compatible*.

These spaces also appear in the Helmholtz decomposition of vector fields in bounded, contractible polyhedral domains $\Omega \subset \mathbb{R}^3$. For every vector field $v \in [L^2(\Omega)]^3$ there exist $\varphi \in H^1(\Omega)$ and $\psi \in H^0(\mathrm{curl}, \Omega) \cap H^0(\mathrm{div}, \Omega)$ such that $v = \mathrm{grad}\,\varphi + \mathrm{curl}\,\psi$, see e.g. Amrouche, Bernardi, Dauge, and Girault (1998, Section 3.5).

2. Neural networks

To accommodate for both continuous components and discontinuous components in the functions we want to emulate, we consider neural networks where several different activation functions are used throughout the network. We define neural networks as a collection of parameters and for each position in the network (also called neuron or unit) we specify the activation function used. Associated to such a neural network is a function, called realization, which is the iterated composition of affine transformations defined in terms of the parameters and non-linear activation functions.

2.1. Feedforward NNs

For $d, L \in \mathbb{N}$, a *neural network* Φ with input dimension $d \geq 1$ and number of layers $L \geq 1$, comprises a finite collection of activation functions $\boldsymbol{\varrho} = \{\varrho_\ell\}_{\ell=1}^L$ and a finite sequence of matrix–vector tuples, i.e.

$$\Phi = ((A_1, b_1, \varrho_1), (A_2, b_2, \varrho_2), \ldots, (A_L, b_L, \varrho_L)).$$

For $N_0 := d$ and *numbers of neurons* $N_1, \ldots, N_L \in \mathbb{N}$ *per layer*, for all $\ell = 1, \ldots, L$ it holds that $A_\ell \in \mathbb{R}^{N_\ell \times N_{\ell-1}}$ and $b_\ell \in \mathbb{R}^{N_\ell}$, and that ϱ_ℓ is a list of length N_ℓ of *activation functions* $(\varrho_\ell)_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, N_\ell$, acting on node i in layer ℓ.

The *realization* of $\Phi : \mathbb{R}^{N_0} \to \mathbb{R}^{N_L}$ as a map is the function

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

where

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

Here, for $\ell = 1, \ldots, L-1$, the list of activation functions ϱ_ℓ of length N_ℓ is effected componentwise: for $y = (y_1, \ldots, y_{N_\ell}) \in \mathbb{R}^{N_\ell}$ we denote $\varrho_\ell(y) = ((\varrho_\ell)_1(y_1), \ldots, (\varrho_\ell)_{N_\ell}(y_{N_\ell}))$. I.e., $(\varrho_\ell)_i$ is the activation function applied in position i of layer ℓ.

We call the layers indexed by $\ell = 1, \ldots, L-1$ *hidden layers*, in those layers activation functions are applied. No activation is applied in the last layer of the NN. For consistency of notation, we define $\varrho_L := \mathrm{Id}_{\mathbb{R}^{N_L}}$.

We refer to $L(\Phi) := L$ as the *depth* of Φ. For $\ell = 1, \ldots, L$ we denote by $M_\ell(\Phi) := \|A_\ell\|_0 + \|b_\ell\|_0$ the *size of layer* ℓ, which is the number of nonzero components in the weight matrix A_ℓ and the bias vector b_ℓ, and call $M(\Phi) := \sum_{\ell=1}^L M_\ell(\Phi)$ the *size* of Φ. Furthermore, we call d and N_L the *input dimension* and the *output dimension*, and denote by $M_{\mathrm{in}}(\Phi) := M_1(\Phi)$ and $M_{\mathrm{out}}(\Phi) := M_L(\Phi)$ the size of the first and the last layer, respectively.

Our networks will use two different activation functions. Firstly, we use the *Rectified Linear Unit* (*ReLU*) activation

$$\rho(x) = \max\{0, x\}. \tag{2.1}$$

We will often use the elementary identities $\rho(x) + \rho(-x) = |x|$, $\rho(x) - \rho(-x) = x$ to construct composite functions. Networks which only contain ReLU activations realize continuous, piecewise linear functions. By *ReLU NNs* we refer to NNs which only have ReLU activations, including networks of depth 1, which do not have hidden layers and realize affine transformations. Secondly, for the emulation of discontinuous functions, we additionally use the *Binary Step Unit* (*BiSU*) activation

$$\sigma(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0, \end{cases} \tag{2.2}$$

which is also called *Heaviside* function. Alternatively, the BiSU can be defined to equal $\frac{1}{2}$ in $x = 0$. That function, which we denote by $\tilde{\sigma}$, can be expressed in terms of σ via $2\tilde{\sigma}(x) = \sigma(x) + 1 - \sigma(-x)$ for all $x \in \mathbb{R}$. Hence, for every NN with $\tilde{\sigma}$ as activation function, there exists a network with σ-activations instead, with proportional depth and size.

2.2. Operations on NNs

In the following sections, we will construct NNs from smaller networks using a ReLU-based *calculus of NNs*, which we now recall from Petersen and Voigtlaender (2018).

Proposition 2.1 (*Parallelization of NNs Petersen and Voigtlaender (2018, Definition 2.7)*). *For $d, L \in \mathbb{N}$ let $\Phi^1 = \left((A_1^{(1)}, b_1^{(1)}, \varrho_1^{(1)}), \ldots, (A_L^{(1)}, b_L^{(1)}, \varrho_L^{(1)})\right)$ and $\Phi^2 = \left((A_1^{(2)}, b_1^{(2)}, \varrho_1^{(2)}), \ldots, (A_L^{(2)}, b_L^{(2)}, \varrho_L^{(2)})\right)$ be two NNs with input dimension d and depth L. Let the* parallelization $\mathrm{P}(\Phi^1, \Phi^2)$ *of Φ^1 and Φ^2 be defined by*

$$\begin{aligned} \mathrm{P}(\Phi^1, \Phi^2) &:= ((A_1, b_1, \varrho_1), \ldots, (A_L, b_L, \varrho_L)), \\ A_1 &= \begin{pmatrix} A_1^{(1)} \\ A_1^{(2)} \end{pmatrix}, \quad A_\ell = \begin{pmatrix} A_\ell^{(1)} & 0 \\ 0 & A_\ell^{(2)} \end{pmatrix}, \quad \text{for } \ell = 2, \ldots, L, \\ b_\ell &= \begin{pmatrix} b_\ell^{(1)} \\ b_\ell^{(2)} \end{pmatrix}, \quad \varrho_\ell = \begin{pmatrix} \varrho_\ell^{(1)} \\ \varrho_\ell^{(2)} \end{pmatrix}, \quad \text{for } \ell = 1, \ldots, L. \end{aligned}$$

Then,

$$\begin{aligned} &\mathrm{R}(\mathrm{P}(\Phi^1, \Phi^2))(x) = (\mathrm{R}(\Phi^1)(x), \mathrm{R}(\Phi^2)(x)), \quad \text{for all } x \in \mathbb{R}^d, \\ &L(\mathrm{P}(\Phi^1, \Phi^2)) = L, \qquad M(\mathrm{P}(\Phi^1, \Phi^2)) = M(\Phi^1) + M(\Phi^2). \end{aligned}$$

The parallelization of more than two NNs is done by repeated application of Proposition 2.1.

Proposition 2.2 (*Sum of NNs*). *For $d, N, L \in \mathbb{N}$ let $\Phi^1 = \left((A_1^{(1)}, b_1^{(1)}, \varrho_1^{(1)}), \ldots, (A_L^{(1)}, b_L^{(1)}, \varrho_L^{(1)})\right)$ and $\Phi^2 = \left((A_1^{(2)}, b_1^{(2)}, \varrho_1^{(2)}), \ldots, (A_L^{(2)}, b_L^{(2)}, \varrho_L^{(2)})\right)$ be two NNs with input dimension d, output dimension N and depth L. Let the* sum $\Phi^1 + \Phi^2$ *of Φ^1 and Φ^2 be defined by*

$$\begin{aligned} \Phi^1 + \Phi^2 &:= ((A_1, b_1, \varrho_1), \ldots, (A_L, b_L, \varrho_L)), \\ A_1 &= \begin{pmatrix} A_1^{(1)} \\ A_1^{(2)} \end{pmatrix}, \quad b_1 = \begin{pmatrix} b_1^{(1)} \\ b_1^{(2)} \end{pmatrix}, \quad \varrho_1 = \begin{pmatrix} \varrho_1^{(1)} \\ \varrho_1^{(2)} \end{pmatrix}, \\ A_\ell &= \begin{pmatrix} A_\ell^{(1)} & 0 \\ 0 & A_\ell^{(2)} \end{pmatrix}, \quad b_\ell = \begin{pmatrix} b_\ell^{(1)} \\ b_\ell^{(2)} \end{pmatrix}, \quad \varrho_\ell = \begin{pmatrix} \varrho_\ell^{(1)} \\ \varrho_\ell^{(2)} \end{pmatrix}, \\ &\text{for } \ell = 2, \ldots, L-1. \end{aligned}$$

$$A_L = \begin{pmatrix} A_L^{(1)} & A_L^{(2)} \end{pmatrix}, \quad b_L = b_L^{(1)} + b_L^{(2)}, \quad \varrho_L = \mathrm{Id}_{\mathbb{R}^N}.$$

Then,

$$\mathrm{R}(\Phi^1 + \Phi^2)(x) = \mathrm{R}(\Phi^1)(x) + \mathrm{R}(\Phi^2)(x), \quad \text{for all } x \in \mathbb{R}^d,$$
$$L(\Phi^1 + \Phi^2) = L, \qquad M(\Phi^1 + \Phi^2) \le M(\Phi^1) + M(\Phi^2).$$

Next, we define the sparse concatenation of two NNs, which realizes exactly the composition of the realizations of the two networks using the fact that we allow the ReLU activation. See Fig. 2.1 for a sketch of the NN structure. The sparse concatenation is a construction which allows to bound the size of the concatenation as 2 times the sum of the sizes of the individual networks. This bound does not hold if we combine the affine transformation of the output layer of Φ^2 with the affine transformation of the input layer of Φ^1.

Proposition 2.3 (*Sparse Concatenation of NNs Petersen and Voigtlaender (2018, Remark 2.6)*)**.** *For $L^{(1)}, L^{(2)} \in \mathbb{N}$, let $\Phi^1 = ((A_1^{(1)}, b_1^{(1)}, \varrho_1^{(1)}), \dots, (A_{L^{(1)}}^{(1)}, b_{L^{(1)}}^{(1)}, \varrho_{L^{(1)}}^{(1)}))$ and $\Phi^2 = \big((A_1^{(2)}, b_1^{(2)}, \varrho_1^{(2)}), \dots, (A_{L^{(2)}}^{(2)}, b_{L^{(2)}}^{(2)}, \varrho_{L^{(2)}}^{(2)})\big)$ be two NNs with depths $L^{(1)}$ and $L^{(2)}$, respectively, such that $N_{L^{(2)}}^{(2)} = N_0^{(1)}$, i.e. the output dimension of Φ^2 equals the input dimension of Φ^1. Let the* sparse concatenation $\Phi^1 \odot \Phi^2$ *of Φ^1 and Φ^2 be a NN of depth $L := L^{(1)} + L^{(2)}$ defined by*

$$\Phi^1 \odot \Phi^2 := ((A_1, b_1, \varrho_1), \dots, (A_L, b_L, \varrho_L)),$$
$$(A_\ell, b_\ell, \varrho_\ell) = (A_\ell^{(2)}, b_\ell^{(2)}, \varrho_\ell^{(2)}), \quad \text{for } \ell = 1, \dots, L^{(2)} - 1,$$
$$A_{L^{(2)}} = \begin{pmatrix} A_{L^{(2)}}^{(2)} \\ -A_{L^{(2)}}^{(2)} \end{pmatrix}, \quad b_{L^{(2)}} = \begin{pmatrix} b_{L^{(2)}}^{(2)} \\ -b_{L^{(2)}}^{(2)} \end{pmatrix}, \quad \varrho_{L^{(2)}} = \begin{pmatrix} \rho \\ \vdots \\ \rho \end{pmatrix},$$
$$A_{L^{(2)}+1} = \begin{pmatrix} A_1^{(1)} & -A_1^{(1)} \end{pmatrix}, \quad b_{L^{(2)}+1} = b_1^{(1)}, \quad \varrho_{L^{(2)}+1} = \varrho_1^{(1)},$$
$$(A_\ell, b_\ell, \varrho_\ell) = (A_{\ell - L^{(2)}}^{(1)}, b_{\ell - L^{(2)}}^{(1)}, \varrho_{\ell - L^{(2)}}^{(1)}), \quad \text{for } \ell = L^{(2)} + 2, \dots, L^{(1)} + L^{(2)}.$$

Then, it holds that

$$\mathrm{R}(\Phi^1 \odot \Phi^2) = \mathrm{R}(\Phi^1) \circ \mathrm{R}(\Phi^2), \qquad L(\Phi^1 \odot \Phi^2) = L^{(1)} + L^{(2)},$$
$$M(\Phi^1 \odot \Phi^2) \le M(\Phi^1) + M_{\mathrm{in}}(\Phi^1) + M_{\mathrm{out}}(\Phi^2) + M(\Phi^2) \le 2M(\Phi^1) + 2M(\Phi^2).$$

Proposition 2.1 only applies to networks of equal depth. To parallelize two networks of unequal depth, the shallowest can be concatenated with a network that emulates the identity using Proposition 2.3. One example of ReLU NNs that emulate the identity is provided by the following proposition.

Proposition 2.4 (*ReLU NN emulation of $\mathrm{Id}_{\mathbb{R}^d}$ Petersen and Voigtlaender (2018, Remark 2.4)*)**.** *For all $d, L \in \mathbb{N}$, there exists a ReLU NN $\Phi_{d,L}^{\mathrm{Id}}$ with input dimension d, output dimension d and depth L which satisfies $\mathrm{R}(\Phi_{d,L}^{\mathrm{Id}}) = \mathrm{Id}_{\mathbb{R}^d}$, $L(\Phi_{d,L}^{\mathrm{Id}}) = L$ and $M(\Phi_{d,L}^{\mathrm{Id}}) \le 2dL$.*

2.3. Expression of specific functions

In the following, we will need ReLU NNs which emulate the minimum or maximum of $d \in \mathbb{N}$ inputs. These are provided in Lemma 2.5. We also need to multiply values from a bounded interval $[-\kappa, \kappa]$ for $\kappa > 0$ by values from the discrete set $\{0, 1\}$, which is the range of the BiSU defined in (2.2). A ReLU NN which emulates such multiplications exactly is constructed in the proof of Proposition 2.8 below.

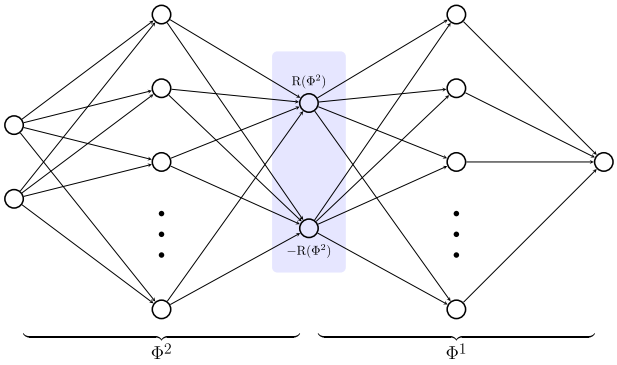


Fig. 2.1. Illustration of Proposition 2.3. In blue, the additional ReLU layer that evaluates the positive and negative parts of $\mathrm{R}(\Phi^2)$.

Lemma 2.5 (*ReLU NN emulation of* min *and* max*, He et al. (2020, Proof Theorem 3.1)*)**.** *For all $d \in \mathbb{N}$, there exist ReLU NNs $\Phi_d^{\max}$ and $\Phi_d^{\min}$ which satisfy*

$$\mathrm{R}(\Phi_d^{\max})(x) = \max\{x_1, \dots, x_d\}, \quad \text{for all } x \in \mathbb{R}^d,$$
$$\mathrm{R}(\Phi_d^{\min})(x) = \min\{x_1, \dots, x_d\}, \quad \text{for all } x \in \mathbb{R}^d,$$
$$L(\Phi_d^{\max}) = L(\Phi_d^{\min}) \le 2 + \log_2(d),$$
$$M(\Phi_d^{\max}) = M(\Phi_d^{\min}) \le Cd.$$

Here, the constant $C > 0$ is independent of d and of the NN sizes and depths.

In space dimension $d = 1$, we may take $\Phi_1^{\min} := \Phi_1^{\max} := \Phi_{1,2}^{\mathrm{Id}}$.

Remark 2.6. The network $\Phi_d^{\max}$ is obtained by repeated applications of $\Phi_2^{\max}$, which itself can for instance be constructed as

$$\Phi_2^{\max} := \left(\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \rho \\ \rho \\ \rho \end{pmatrix} \right), \left(\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, 0, \mathrm{Id}_{\mathbb{R}} \right) \right).$$

We point out that this construction of $\Phi_2^{\max}$ leads to a slightly more efficient representation of $\Phi_d^{\max}$ than the one given in He et al. (2020, Theorem 3.1), as it requires less neurons, weights and biases. However, this will merely improve the constant C in Lemma 2.5, but not the stated asymptotic d-dependence of $L(\Phi_2^{\max})$ and $M(\Phi_2^{\max})$. The remark applies verbatim to min networks.

Remark 2.7 (*Min/max with Recurrent Nets*)**.** The d-dependence can be completely avoided by admitting recurrent neural nets (RNNs), i.e., RNNs can express the maximum of d inputs with a network of size, depth and width $O(1)$. We briefly sketch the idea: An RNN allows for information to flow backwards, i.e., we can take the output of $\Phi_2^{\max}$ in time step t as one of its inputs at time step $t + 1$. With the initialization $\tilde{x}_0 := x_1$, this leads to the iteration

$$\tilde{x}_t = \Phi_2^{\max}(\tilde{x}_{t-1}, x_t),$$

where the network receives in step t the input x_t. Then the network's output $\tilde{x}_n$ in step n equals $\max\{x_1, \dots, x_n\}$. The remark applies verbatim to min networks.

The following proposition provides the exact ReLU NN emulation of products of elements from a bounded interval $[-\kappa, \kappa]$ for $\kappa > 0$ by elements from the discrete set $\{0, 1\}$. The network depth and size are independent of κ.

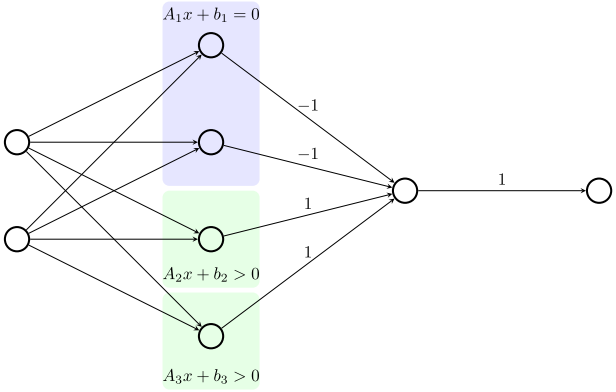


Fig. 2.2. Illustration of Lemma 2.9 for $d = 2$, $n = 1$, $N = 3$. In blue, the neurons corresponding to one hyperplane ($i = 1$), in green the half-spaces ($i = 2, 3$).

Proposition 2.8. *For all $d \in \mathbb{N}$ and $\kappa > 0$ there exists a ReLU NN $\Phi^{\times}_{d,\kappa}$*

$$\mathrm{R}(\Phi^{\times}_{d,\kappa})(x_1, \dots, x_d, y) = xy = (x_1 y, \dots, x_d y)^{\top},$$
$$\text{for all } x \in [-\kappa, \kappa]^d \text{ and } y \in \{0, 1\},$$
$$L(\Phi^{\times}_{d,\kappa}) \leq 2, \qquad M(\Phi^{\times}_{d,\kappa}) \leq 12d.$$

A proof of Proposition 2.8 is given in the Appendix.

The exact BiSU emulation of indicator functions is the topic of the following lemma. An illustration of the network defined in the lemma is given in Fig. 2.2.

Lemma 2.9 (*Emulation of Indicator Functions*)**.** *For $d, N \in \mathbb{N}$ and $n \in \{0, \dots, N\}$, let $A_1, \dots, A_N \in \mathbb{R}^{1\times d}$ and $b_1, \dots, b_N \in \mathbb{R}^1$ be such that*

$$\Omega := \bigcap_{i=1,\dots,n} \{x \in \mathbb{R}^d : A_i x + b_i = 0\} \cap \bigcap_{i=n+1,\dots,N} \{x \in \mathbb{R}^d : A_i x + b_i > 0\} \neq \emptyset.$$

Let the NN $\Phi^{\mathbb{1}}_{\Omega}$ with layer sizes $N_0 = d$, $N_1 = N+n$ and $N_2 = 1 = N_3$ be defined as

$$\Phi^{\mathbb{1}}_{\Omega} := \left(\left(\begin{pmatrix} A_1 \\ -A_1 \\ \vdots \\ A_n \\ -A_n \\ A_{n+1} \\ \vdots \\ A_N \end{pmatrix}, \begin{pmatrix} b_1 \\ -b_1 \\ \vdots \\ b_n \\ -b_n \\ b_{n+1} \\ \vdots \\ b_N \end{pmatrix}, \begin{pmatrix} \sigma \\ \sigma \\ \vdots \\ \sigma \\ \sigma \\ \sigma \\ \vdots \\ \sigma \end{pmatrix} \right), (A, b, \varrho), (1, 0, \mathrm{Id}_{\mathbb{R}}) \right),$$
$$A := \begin{pmatrix} -1 & \cdots & -1 & 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{1\times(N+n)},$$
$$b := -(N - n - \tfrac{1}{4}) \in \mathbb{R}^1, \quad \varrho := \sigma,$$

where the first $2n$ elements of A equal -1 and the last $N - n$ equal 1.

Then, for all $x \in \mathbb{R}^d$

$$\mathrm{R}(\Phi^{\mathbb{1}}_{\Omega})(x) = \begin{cases} 1 & \text{if } x \in \Omega, \\ 0 & \text{otherwise}, \end{cases} \qquad L(\Phi^{\mathbb{1}}_{\Omega}) = 3,$$
$$M(\Phi^{\mathbb{1}}_{\Omega}) \leq (d + 2)(N + n) + 2.$$

A proof of Lemma 2.9 is provided in the Appendix.

3. NN emulation of lowest order conforming finite element shape functions

Consider a bounded polytopal domain $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}\backslash\{1\}$, and a regular simplicial partition $\mathcal{T}$ of Ω.

In Sections 3.1–3.4, we will present neural network emulations of the lowest order conforming FEM spaces for $H^1(\Omega)$, $H^0(\mathrm{curl}, \Omega)$, $H^0(\mathrm{div}, \Omega)$ and $L^2(\Omega)$.[3] These finite-dimensional spaces appear naturally in structure-preserving discretizations of the de Rham complex. They are a key ingredient for *variationally consistent DNN emulations* of differential operators appearing in the de Rham complex. For each type of shape function, we explicitly define a network which emulates that shape function exactly. Global approximations can be obtained by taking a linear combination of these shape functions using Proposition 2.2 (scalar multiples of shape functions are obtained by scaling all weights and biases of the output layer). We will detail this in Proposition 5.1.

For shape functions which are discontinuous after extending them to Ω by the value zero outside their domain of definition, we use Lemma 2.9 based on BiSU activation to emulate indicator functions of (parts of) their domain of definition. We then use Proposition 2.8 based on ReLU activation to multiply a continuous, piecewise linear function, which is equal to the shape function on part of Ω, by the indicator function of that part of the domain.

The following lemma provides NN emulations of possibly discontinuous, piecewise linear functions, and will be used repeatedly in Sections 3.1–3.4. A sketch of the NN structure is given in Fig. 3.1.

Lemma 3.1 (*Emulation of Piecewise Linear Functions*)**.** *For $d, s, \mu \in \mathbb{N}$ let $\Omega \subset \mathbb{R}^d$ be a bounded polytope and $\mathcal{T}$ be a regular, simplicial partition of Ω with $s = |\mathcal{T}|$ elements, $\mathcal{T} = \{T_i\}_{i=1,\dots,s}$. Let $u : \Omega \to \mathbb{R}^{\mu}$ be a function which for all $i = 1, \dots, s$ satisfies $u|_{T_i} \in [\mathbb{P}_1]^{\mu}$ and $u|_{T_i}(x) = A^{(i)}x + b^{(i)}$, $x \in T_i$.*

Then, for any

$$\kappa \geq \max_{i=1,\dots,s} \sup_{x\in T_i} \|A^{(i)}x + b^{(i)}\|_{\infty}, \tag{3.1}$$

$$\Phi^{PwL}_u := \sum_{i=1}^{s} \Phi^{\times}_{\mu,\kappa} \odot \mathrm{P}\left(\Phi^{\mathrm{Id}}_{\mu,2} \odot \left(\left(A^{(i)}, b^{(i)}, \mathrm{Id}_{\mathbb{R}^{\mu}}\right)\right), \Phi^{\mathbb{1}}_{T_i}\right) \tag{3.2}$$

satisfies $u(x) = \mathrm{R}\left(\Phi^{PwL}_u\right)(x)$ for all $x \in \cup_{i=1}^{s} T_i$ and $\mathrm{R}\left(\Phi^{PwL}_u\right)(x) = 0$ for all $x \in \mathbb{R}^d \setminus \cup_{i=1}^{s} T_i$. Furthermore, if $\|A^{(i)}\|_0 + \|b^{(i)}\|_0 \leq m$ for all $i = 1, \dots, s$, then there exists $C > 0$ independent of d and $\mathcal{T}$ such that

$$L(\Phi^{PwL}_u) = 5\,, \qquad M(\Phi^{PwL}_u) \leq Cs(\mu + m + d^2).$$

Proof. Firstly, we observe that indeed $u(x) = \mathrm{R}\left(\Phi^{PwL}_u\right)(x)$ for all $x \in \cup_{i=1}^{s} T_i$ and $\mathrm{R}\left(\Phi^{PwL}_u\right)(x) = 0$ for all $x \in \mathbb{R}^d \setminus \cup_{i=1}^{s} T_i$.

Secondly, we estimate

$$\begin{aligned} L(\Phi^{PwL}_u) &= L(\Phi^{\times}_{\mu,\kappa}) + L(\Phi^{\mathbb{1}}_{T_i}) = 5, \\ M(\Phi^{PwL}_u) &\leq s\left(2M\left(\Phi^{\times}_{\mu,\kappa}\right)\right. \\ &\qquad \left. + \; 2M\left(\mathrm{P}\left(\Phi^{\mathrm{Id}}_{\mu,2} \odot \left(\left(A^{(i)}, b^{(i)}, \mathrm{Id}_{\mathbb{R}^{\mu}}\right)\right), \Phi^{\mathbb{1}}_{T_i}\right)\right)\right) \\ &\leq s\left(2M\left(\Phi^{\times}_{\mu,\kappa}\right) + 4M(\Phi^{\mathrm{Id}}_{\mu,2}) + 4M\left(\left(\left(A^{(i)}, b^{(i)}, \mathrm{Id}_{\mathbb{R}^{\mu}}\right)\right)\right)\right. \\ &\qquad \left. + \; 2M\left(\Phi^{\mathbb{1}}_{T_i}\right)\right) \\ &\leq s(C\mu + C\mu + Cm + Cd^2) \leq Cs(\mu + m + d^2), \end{aligned}$$

[3] Throughout this section, we will regularly refer to Ern and Guermond (2021), where only Lipschitz domains are considered. We stress that the finite element spaces $\mathrm{S}^0_0(\mathcal{T}, \Omega)$, $\mathrm{N}_0(\mathcal{T}, \Omega)$, $\mathrm{RT}_0(\mathcal{T}, \Omega)$ and $\mathrm{S}^1_1(\mathcal{T}, \Omega)$ can be defined on regular, simplicial triangulations $\mathcal{T}$ of all bounded, polytopal domains Ω and that our NN emulation results from this section apply to all such Ω.

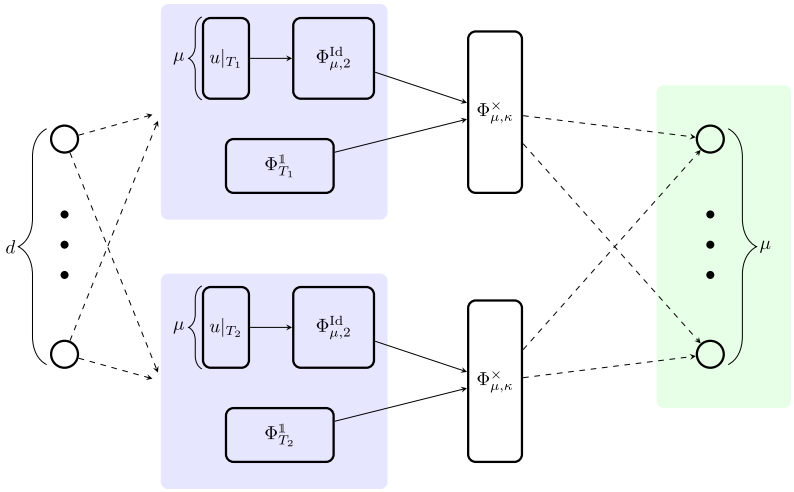


Fig. 3.1. Illustration of Lemma 3.1 for $s = 2$. Solid lines indicate sparse concatenation, dashed lines visualize layers integrated in the blocks. Each of the blue groups represents a parallelization from Proposition 2.1. The elementwise assembly is done in the output layer (green).

where we applied in the last line Lemma 2.9 with $N = d + 1$, $n = 0$.

3.1. Piecewise constants S^0_0

The lowest order approximation space for $L^2(\Omega)$ is the finite dimensional subspace $\mathrm{S}^0_0(\mathcal{T}, \Omega)$ from (1.4). A basis is given by $\{\theta^{\mathrm{S}^0_0}_T\}_{T\in\mathcal{T}}$, whose elements are indicator functions $\theta^{\mathrm{S}^0_0}_T := \mathbb{1}_T$. They can be expressed by applying Lemma 2.9 with $N = d + 1$ and $n = 0$: for all $T = \mathrm{conv}(\{a_0, \dots, a_d\}) \in \mathcal{T}$, we define $(A_i, b_i) \in \mathbb{R}^{1\times(d+1)}$, $i = 1, \dots, d + 1$ by the relations

$$(A_i, b_i)\begin{pmatrix} (a_0)_1 & & (a_d)_1 \\ & \ddots & \\ (a_0)_d & & (a_d)_d \\ 1 & \cdots & 1 \end{pmatrix} = \boldsymbol{e}_i^\top, \qquad \text{for all } i = 1, \dots, d + 1, \tag{3.3}$$

where $(\boldsymbol{e}_i)_j = \delta_{ij}$, so that $T = \bigcap_{i=1,\dots,d+1}\{x \in \mathbb{R}^d : A_i x + b_i > 0\}$. Then there exists $C > 0$ independent of d and $\mathcal{T}$ such that for all $T \in \mathcal{T}$ the NN $\Phi^{\mathrm{S}^0_0}_T := \Phi^{\mathbb{1}}_T$ satisfies

$$\theta^{\mathrm{S}^0_0}_T = \mathrm{R}(\Phi^{\mathrm{S}^0_0}_T), \quad L(\Phi^{\mathrm{S}^0_0}_T) = 3, \quad M(\Phi^{\mathrm{S}^0_0}_T) \le (d+2)(d+1)+2 \le Cd^2. \tag{3.4}$$

3.2. Raviart–Thomas elements RT_0

We introduce a basis of $\mathrm{RT}_0(\mathcal{T}, \Omega)$ from (1.5), and provide a NN emulation of those basis functions. For $f \subset \partial\Omega$, we define $s(f) = 1$ and $\theta^{\mathrm{RT}_0}_f(x) := \frac{|f|}{d|T_1|}(x - a_1)\mathbb{1}_{T_1}$, where $f \subset \overline{T_1}$, $T_1 \in \mathcal{T}$ and a_1 is the only vertex of T_1 that does not belong to $\overline{f}$.[4] For interior faces $f \subset \Omega$ we define $s(f) = 2$ and construct $\theta^{\mathrm{RT}_0}_f$ by assembling local shape functions of the neighboring simplices T_1, T_2 with $\overline{f} = \overline{T_1} \cap \overline{T_2}$, (Ern & Guermond, 2021, Equation (14.3))

$$\theta^{\mathrm{RT}_0}_f(x) := \begin{cases} \frac{|f|}{d|T_1|}(x - a_1) & \text{if } x \in T_1, \\ -\frac{|f|}{d|T_2|}(x - a_2) & \text{if } x \in T_2, \\ 0 & \text{if } x \notin \overline{T_1 \cup T_2}, \end{cases} \tag{3.5}$$

where a_1, a_2 are the only vertices of T_1, T_2, respectively, not belonging to $\overline{f}$. The functions $\{\theta^{\mathrm{RT}_0}_f\}_{f\in\mathcal{F}}$ form a basis of $\mathrm{RT}_0(\mathcal{T}, \Omega)$ (see, e.g., Ern and Guermond (2021, Proposition 14.1)).

Proposition 3.2. *Given $f \in \mathcal{F}$, let $\{T_i\}_{i=1}^{s(f)}$ be the simplices adjacent to f and let $a_i := (\mathcal{V} \cap \overline{T_i}) \setminus \overline{f} \in \mathbb{R}^d$, $i = 1, \dots, s(f)$. Then*

$$\Phi^{\mathrm{RT}_0}_f := \sum_{i=1}^{s(f)}(-1)^{i-1}\Phi^{\times}_{d,\kappa} \odot \mathrm{P}\left(\Phi^{\mathrm{Id}}_{d,2} \odot \left(\left(\tfrac{|f|}{d|T_i|}\mathrm{Id}_{d\times d}, -\tfrac{|f|}{d|T_i|}a_i, \mathrm{Id}_{\mathbb{R}^d}\right)\right), \Phi^{\mathbb{1}}_{T_i}\right) \tag{3.6}$$

satisfies $\theta^{\mathrm{RT}_0}_f(x) = \mathrm{R}(\Phi^{\mathrm{RT}_0}_f)(x)$ for a.e. $x \in \Omega$, for any

$$\kappa \ge \max_{i=1,\dots,s(f)} \sup_{x\in T_i} \tfrac{|f|}{d|T_i|}\|x - a_i\|_\infty. \tag{3.7}$$

In addition, there exists an absolute constant $C > 0$ independent of d and $\mathcal{T}$ such that for all $f \in \mathcal{F}$

$$L(\Phi^{\mathrm{RT}_0}_f) = 5, \qquad M(\Phi^{\mathrm{RT}_0}_f) \le Cd^2 s(f) \le 2Cd^2.$$

Remark 3.3. We note that the right-hand side of (3.7) is bounded from above by a constant which only depends on d and the shape regularity constant C_{sh}.

Proof of Proposition 3.2. Firstly, we observe that indeed $\theta^{\mathrm{RT}_0}_f(x) = \mathrm{R}\left(\Phi^{\mathrm{RT}_0}_f\right)(x)$ for all $x \in \Omega \setminus \partial\mathcal{T}$, where $\partial\mathcal{T} := \bigcup_{T\in\mathcal{T}} \partial T$.

Secondly, we apply Lemma 3.1 with $\mu = d$, $m = 2d$ and $s = s(f)$.

Alternatively, we can build the same shape functions by enforcing strongly, via ReLU activation, continuity of the component normal to f, as imposed in (1.5). We select the unit normal vector n_f to f pointing towards T_2 and an orthonormal system $\{t_1, \dots, t_{d-1}\}$ spanning the hyperplane tangent to f. Then, we decompose

$$\theta^{\mathrm{RT}_0}_f(x) = (\theta^{\mathrm{RT}_0}_f(x) \cdot n_f)n_f + \sum_{j=1}^{d-1}(\theta^{\mathrm{RT}_0}_f(x) \cdot t_j)t_j. \tag{3.8}$$

Thus, it suffices to compute separately $\theta^{\mathrm{RT}_0}_f(x) \cdot n_f$ and $\theta^{\mathrm{RT}_0}_f(x) \cdot t_j$ and to take the linear combination (3.8) in the last layer. The proof of the next proposition will be given in the Appendix.

[4] We use a different normalization of the shape functions than in Ern and Guermond (2021, Section 14.1). This is inconsequential for the ensuing analysis.

Proposition 3.4. *Given $f \in \mathcal{F}$, let $\{T_i\}_{i=1}^{s(f)}$ be the simplices adjacent to f. Then there exist $A_{n_f}^{(i)} \in \mathbb{R}^{1\times d}$, $b_{n_f}^{(i)} \in \mathbb{R}$, $i = 1, \dots, s(f)$ such that*

$$\Phi_f^{\mathrm{RT}_0,\perp} := \Phi_{1,1}^{\times} \odot \mathrm{P}\left(\Phi_{s(f)}^{\min} \odot \left(\left(\begin{pmatrix} A_{n_f}^{(1)} \\ \vdots \\ A_{n_f}^{(s(f))} \end{pmatrix}, \begin{pmatrix} b_{n_f}^{(1)} \\ \vdots \\ b_{n_f}^{(s(f))} \end{pmatrix}, \mathrm{Id}_{\mathbb{R}^{s(f)}}\right)\right), \sum_{i=1}^{s(f)} \Phi_{T_i}^{\mathbb{1}} + \Phi_f^{\mathbb{1}}\right) \tag{3.9}$$

satisfies $\theta_f^{\mathrm{RT}_0}(x) \cdot n_f = \mathrm{R}\left(\Phi_f^{\mathrm{RT}_0,\perp}\right)(x)$ for a.e. $x \in \Omega$ and every $x \in f$. There also exist $A_{t_j}^{(i)} \in \mathbb{R}^{1\times d}$, $b_{t_j}^{(i)} \in \mathbb{R}$, $i = 1, \dots, s(f)$, $j = 1, \dots, d-1$ such that

$$\Phi_f^{\mathrm{RT}_0,t_j} := \sum_{i=1}^{s(f)} \Phi_{1,\kappa}^{\times} \odot \mathrm{P}\left(\Phi_{1,2}^{\mathrm{Id}} \odot \left(A_{t_j}^{(i)}, b_{t_j}^{(i)}, \mathrm{Id}_{\mathbb{R}}\right), \Phi_{T_i}^{\mathbb{1}}\right) \tag{3.10}$$

satisfies $\theta_f^{\mathrm{RT}_0}(x) \cdot t_j = \mathrm{R}\left(\Phi_f^{\mathrm{RT}_0,t_j}\right)(x)$ for a.e. $x \in \Omega$, where

$$\kappa \geq \max_{\substack{i=1,\dots,s(f)\\ j=1,\dots,d-1}} \sup_{x\in T_i} \left|A_{t_j}^{(i)} x + b_{t_j}^{(i)}\right|. \tag{3.11}$$

In addition, there exists a constant $C > 0$ that is independent of d and $\mathcal{T}$ such that for all $f \in \mathcal{F}$

$$L(\Phi_f^{\mathrm{RT}_0,\perp}) = 5, \qquad L(\Phi_f^{\mathrm{RT}_0,t_j}) = 5,$$
$$M(\Phi_f^{\mathrm{RT}_0,\perp}) \leq Cd^2 s(f) \leq 2Cd^2, \qquad M(\Phi_f^{\mathrm{RT}_0,t_j}) \leq Cd^2 s(f) \leq 2Cd^2.$$

Remark 3.5. The right-hand side of Eq. (3.11) is bounded from above by a constant which only depends on d and the shape regularity constant C_{sh} of the mesh $\mathcal{T}$.

Corollary 3.6. *For all $f \in \mathcal{F}$, the NN*

$$\Phi_f^{\mathrm{RT}_0,*} := \left(\left(n_f, 0, \mathrm{Id}_{\mathbb{R}^d}\right)\right) \odot \Phi_f^{\mathrm{RT}_0,\perp} + \sum_{j=1}^{d-1} \left(\left(t_j, 0, \mathrm{Id}_{\mathbb{R}^d}\right)\right) \odot \Phi_f^{\mathrm{RT}_0,t_j} \tag{3.12}$$

satisfies $\theta_f^{\mathrm{RT}_0}(x) = \mathrm{R}\left(\Phi_f^{\mathrm{RT}_0,*}\right)(x)$ for a.e. $x \in \Omega$ and $(\theta_f^{\mathrm{RT}_0}(x) \cdot n_f) n_f = \mathrm{R}\left(\Phi_f^{\mathrm{RT}_0,*}\right)(x)$ for all $x \in f$. In addition, there exists $C > 0$ independent of d and $\mathcal{T}$ such that for all $f \in \mathcal{F}$

$$L(\Phi_f^{\mathrm{RT}_0,*}) = 6, \qquad M(\Phi_f^{\mathrm{RT}_0,*}) \leq Cd^3.$$

Proof. We estimate the network size and depth as follows:

$$L(\Phi_f^{\mathrm{RT}_0,*}) = 6,$$
$$M(\Phi_f^{\mathrm{RT}_0,*}) \leq 2M\left(\left(\left(n_f, 0, \mathrm{Id}_{\mathbb{R}^d}\right)\right)\right) + 2M(\Phi_f^{\mathrm{RT}_0,\perp}) + \sum_{j=1}^{d-1}\left(2M\left(\left(\left(t_j, 0, \mathrm{Id}_{\mathbb{R}^d}\right)\right)\right) + 2M(\Phi_f^{\mathrm{RT}_0,t_j})\right) \leq Cd^3.$$

3.3. Nédélec elements N_0

In this section we restrict ourselves to the space dimension $d \in \{2, 3\}$. For $d = 2$, we can relate the Nédélec basis functions to the Raviart–Thomas basis. In fact, one can verify that, for an edge $f \in \mathcal{E} = \mathcal{F}$ (which is also a face), the finite element basis $\{\theta_f^{\mathrm{N}_0}\}_{f\in\mathcal{F}}$ for N_0 satisfies

$$\theta_f^{\mathrm{N}_0} \cdot t_f = \theta_f^{\mathrm{RT}_0} \cdot n_f, \quad \text{and} \quad \theta_f^{\mathrm{N}_0} \cdot n_f = -\theta_f^{\mathrm{RT}_0} \cdot t_f, \tag{3.13}$$

where $n_f := ((n_f)_1, (n_f)_2)$ is a unit normal vector to f as in (3.8) and $t_f = (-(n_f)_2, (n_f)_1)$ is a unit vector tangent to f. Hence, a NN emulation for $\theta_f^{\mathrm{N}_0}$ can be derived from Proposition 3.2 or Corollary 3.6.

We now focus on the case $d = 3$. A basis $\{\theta_e^{\mathrm{N}_0}\}_{e\in\mathcal{E}}$ for $\mathrm{N}_0(\mathcal{T}, \Omega)$ can be constructed by assembling local shape functions of all simplices $T_1 \dots, T_{s(e)}$, $s(e) \in \mathbb{N}$ sharing an edge e. We fix $e \in \mathcal{E}$ and a unit vector t_e tangent to e, and denote the midpoint of e by m_e. We denote by $\tilde{e}(i)$ the only edge of T_i that does not share a vertex with e, and let $t_{\tilde{e}(i)}$ be a unit vector tangent to $\tilde{e}(i)$, directed in such a way that $t_e \cdot [(m_e - m_{\tilde{e}(i)}) \times t_{\tilde{e}(i)}] > 0$.

Then,

$$\theta_e^{\mathrm{N}_0}(x) := \begin{cases} \dfrac{(x - m_{\tilde{e}(i)}) \times t_{\tilde{e}(i)}}{t_e \cdot [(m_e - m_{\tilde{e}(i)}) \times t_{\tilde{e}(i)}]} & \text{if } x \in T_i,\ i = 1, \dots, s(e), \\ 0 & \text{if } x \notin \overline{\bigcup_{i=1\dots,s(e)} T_i}. \end{cases} \tag{3.14}$$

Note that

$$\mathfrak{s}(\mathcal{E}) := \max_{e\in\mathcal{E}} s(e) \tag{3.15}$$

is bounded from above by a constant only dependent on the shape regularity constant C_{sh} of $\mathcal{T}$. See Ern and Guermond (2021, Remark 11.5 and Proposition 11.6).

Proposition 3.7. *Given $e \in \mathcal{E}$, let $A_e^{(i)} \in \mathbb{R}^{3\times 3}$, $b_e^{(i)} \in \mathbb{R}^3$ be such that for $i = 1, \dots, s(e)$*

$$A_e^{(i)} x := \frac{x \times t_{\tilde{e}(i)}}{t_e \cdot [(m_e - m_{\tilde{e}(i)}) \times t_{\tilde{e}(i)}]} \quad \forall x \in \mathbb{R}^3,$$
$$b_e^{(i)} := -\frac{m_{\tilde{e}(i)} \times t_{\tilde{e}(i)}}{t_e \cdot [(m_e - m_{\tilde{e}(i)}) \times t_{\tilde{e}(i)}]}.$$

Then

$$\Phi_e^{\mathrm{N}_0} := \sum_{i=1}^{s(e)} \Phi_{3,\kappa}^{\times} \odot \mathrm{P}\left(\Phi_{3,2}^{\mathrm{Id}} \odot \left(\left(A_e^{(i)}, b_e^{(i)}, \mathrm{Id}_{\mathbb{R}^3}\right)\right), \Phi_{T_i}^{\mathbb{1}}\right) \tag{3.16}$$

satisfies $\theta_e^{\mathrm{N}_0}(x) = \mathrm{R}\left(\Phi_e^{\mathrm{N}_0}\right)(x)$ for a.e. $x \in \Omega$, for any κ such that

$$\kappa \geq \max_{i=1,\dots,s(e)} \sup_{x\in T_i} \frac{\|(x - m_{\tilde{e}(i)}) \times t_{\tilde{e}(i)}\|_\infty}{t_e \cdot [(m_e - m_{\tilde{e}(i)}) \times t_{\tilde{e}(i)}]}. \tag{3.17}$$

Furthermore, there exists $C > 0$ independent of $\mathcal{T}$ such that for all $e \in \mathcal{E}$

$$L(\Phi_e^{\mathrm{N}_0}) = 5, \quad M(\Phi_e^{\mathrm{N}_0}) \leq Cs(e) \leq C\mathfrak{s}(\mathcal{E}).$$

Proof. Firstly, we observe that indeed $\theta_e^{\mathrm{N}_0}(x) = \mathrm{R}\left(\Phi_e^{\mathrm{N}_0}\right)(x)$ for all $x \in \Omega \setminus \partial\mathcal{T}$. Secondly, we use Lemma 3.1 with $\mu = d = 3$, $m = d^2 + d = 12$ and $s = s(e)$.

3.4. CPwL elements S_1^1

In this section, we provide a construction based on element-by-element assembly of the shape functions, similar to that in the previous sections, using both ReLU and BiSU activations. A basis $\{\theta_p^{\mathrm{S}_1^1}\}_{p\in\mathcal{V}}$ of $\mathrm{S}_1^1(\mathcal{T}, \Omega)$ is uniquely defined by the relations $\theta_{p_i}^{\mathrm{S}_1^1}(p_j) = \delta_{ij}$, for $p_i, p_j \in \mathcal{V}$. Define $s(p) := |\{T \in \mathcal{T} : p \in \overline{T}\}| \in \mathbb{N}$. Note that

$$\mathfrak{s}(\mathcal{V}) := \max_{p\in\mathcal{V}} s(p) \tag{3.18}$$

is bounded from above by a constant only dependent on d and the shape regularity constant C_{sh} of $\mathcal{T}$. See Ern and Guermond (2021, Remark 11.5 and Proposition 11.6), which generalize to

space dimension $d > 3$. The following proposition is analogous to Propositions 3.2 and 3.7.

Proposition 3.8. *Given $p \in \mathcal{V}$, let $T_1, \dots, T_{s(p)} \in \mathcal{T}$, $s(p) \in \mathbb{N}$ denote the simplices adjacent to p. Let $A_p^{(i)} \in \mathbb{R}^{1\times d}$, $b_p^{(i)} \in \mathbb{R}^1$ for $i = 1, \dots, s(p)$ be such that*

$$(A_p^{(i)}, b_p^{(i)}) \begin{pmatrix} p_1 & (a_{i,1})_1 & & (a_{i,d})_1 \\ \vdots & & \ddots & \\ p_d & (a_{i,1})_d & & (a_{i,d})_d \\ 1 & 1 & \cdots & 1 \end{pmatrix} = (1, 0, \dots, 0),$$

where the points $a_{i,j} \in \mathbb{R}^d$ are such that $T_i = \mathrm{conv}(p, a_{i,1}, \dots, a_{i,d})$. Then

$$\Phi_p^{\mathsf{S}_1^1} := \sum_{i=1}^{s(p)} \Phi_{1,1}^{\times} \odot \mathrm{P}\left(\Phi_{1,2}^{\mathrm{Id}} \odot \left(A_p^{(i)}, b_p^{(i)}, \mathrm{Id}_{\mathbb{R}}\right), \Phi_{T_i}^{\mathbb{1}}\right) \tag{3.19}$$

satisfies $\theta_p^{\mathsf{S}_1^1}(x) = \mathrm{R}(\Phi_p^{\mathsf{S}_1^1})(x)$ for a.e. $x \in \Omega$. Furthermore, there exists $C > 0$ independent of $\mathcal{T}$ such that for all $p \in \mathcal{V}$

$$L(\Phi_p^{\mathsf{S}_1^1}) = 5, \qquad M(\Phi_p^{\mathsf{S}_1^1}) \le Cs(p)d^2 \le C\mathfrak{s}(\mathcal{V})d^2.$$

Proof. Observing that $\theta_p^{\mathsf{S}_1^1}(x) = \mathrm{R}(\Phi_p^{\mathsf{S}_1^1})(x)$ for all $x \in \Omega \setminus \partial\mathcal{T}$, we use Lemma 3.1 with $\mu = 1$, $m = d + 1$ and $s = s(p)$ to estimate the NN size.

4. ReLU NN emulation of CPwL shape functions

For continuous shape functions which vanish on the boundary of their support, one can construct NN emulations using the ReLU activation function alone, as shown in He et al. (2020, Section 3) for regular, simplicial meshes *with convex patches*. The purpose of this section is to extend these results to arbitrary regular, simplicial partitions $\mathcal{T}$ of polytopal domains $\Omega \subset \mathbb{R}^d$, in any space dimension $d \ge 2$, using only ReLU activations, significantly improving the network size bounds from He et al. (2020, Theorem 5.2). In the sequel, for a vertex $p \in \mathcal{V}$ we write

$$\omega(p) := \bigcup_{i=1\dots,s(p)} \overline{T}_i, \tag{4.1}$$

where $T_1, \dots, T_{s(p)} \in \mathcal{T}$ denote the simplices adjacent to p. We call $\omega(p)$ a *patch*. One key assumption in He et al. (2020, Section 3) was that $\omega(p)$ is convex for all vertices $p \in \mathcal{V}$.

Removing this assumption is the main topic of Section 4.2. We remark that the construction given in Section 3.4 also does not require convexity of the patches and, since no minimum is computed, the depth of the network is independent of the input dimension d and the maximum number of elements meeting in one point $\mathfrak{s}(\mathcal{V})$. In this section we avoid the use of BiSU activations, which could be considered not natural for the emulation of continuous functions in $\mathsf{S}_1^1(\mathcal{T}, \Omega)$.

4.1. Regular, simplicial partitions $\mathcal{T}$ with convex patches

Under the assumption of convexity of patches, the hat basis functions $\{\theta_p^{\mathsf{S}_1^1}\}_{p\in\mathcal{V}} \subset \mathsf{S}_1^1(\mathcal{T}, \Omega)$ satisfy He et al. (2020, Lemma 3.1)

$$\theta_p^{\mathsf{S}_1^1}(x) = \max\left\{0, \min_{i=1,\dots,s(p)} A_p^{(i)}x + b_p^{(i)}\right\}, \tag{4.2}$$

with $A_p^{(i)} \in \mathbb{R}^{1\times d}$, $b_p^{(i)} \in \mathbb{R}^1$ such that $A_p^{(i)}x + b_p^{(i)} = \theta_p^{\mathsf{S}_1^1}|_{T_i}(x)$ for all $T_i \subset \omega(p)$, $i = 1, \dots, s(p)$.

We now recall the emulation of the shape functions from He et al. (2020), and show the dependence of the constants on d. We remark that the explicit d-dependence was not studied in He et al. (2020).

Proposition 4.1 (He et al. (2020, Theorem 3.1))**.** *Consider $p \in \mathcal{V}$ for which $\omega(p)$ is convex and let $T_1, \dots, T_{s(p)} \in \mathcal{T}$, $s(p) \in \mathbb{N}$ be the simplices adjacent to p. Let $(A_p^{(i)}, b_p^{(i)}) \in \mathbb{R}^{(d+1)\times 1}$, $i = 1, \dots, s(p)$ be as in (4.2).*

Then

$$\Phi_p^{CPwL} := \left((1, 0, \rho), (1, 0, \mathrm{Id}_{\mathbb{R}})\right) \odot \Phi_{s(p)}^{\min} \odot \left(\left(\begin{pmatrix} A_p^{(1)} \\ \vdots \\ A_p^{(s(p))} \end{pmatrix}, \begin{pmatrix} b_p^{(1)} \\ \vdots \\ b_p^{(s(p))} \end{pmatrix}, \mathrm{Id}_{\mathbb{R}^{s(p)}}\right)\right)$$

satisfies $\mathrm{R}(\Phi_p^{CPwL})(x) = \theta_p^{\mathsf{S}_1^1}(x)$ for all $x \in \Omega$ and there exists $C > 0$ independent of d and $\mathcal{T}$ such that for all $p \in \mathcal{V}$

$$L(\Phi_p^{CPwL}) \le 5 + \log_2(s(p)), \qquad M(\Phi_p^{CPwL}) \le Cs(p)d.$$

The depth only depends on $\mathcal{T}$ through $s(p)$.

Proof. The network depth and size can be bounded as

$$\begin{aligned} L(\Phi_p^{CPwL}) &= L\left(\left((1, 0, \rho), (1, 0, \mathrm{Id}_{\mathbb{R}})\right)\right) + L(\Phi_{s(p)}^{\min}) \\ &\quad + L\left(\left(\left(\begin{pmatrix} A_p^{(1)} \\ \vdots \\ A_p^{(s(p))} \end{pmatrix}, \begin{pmatrix} b_p^{(1)} \\ \vdots \\ b_p^{(s(p))} \end{pmatrix}, \mathrm{Id}_{\mathbb{R}^{s(p)}}\right)\right)\right) \\ &= 2 + L(\Phi_{s(p)}^{\min}) + 1 \le 3 + (2 + \log_2(s(p))) \\ &= 5 + \log_2(s(p)), \\ M(\Phi_p^{CPwL}) &\le CM\left(\left((1, 0, \rho), (1, 0, \mathrm{Id}_{\mathbb{R}})\right)\right) + CM(\Phi_{s(p)}^{\min}) \\ &\quad + CM\left(\left(\left(\begin{pmatrix} A_p^{(1)} \\ \vdots \\ A_p^{(s(p))} \end{pmatrix}, \begin{pmatrix} b_p^{(1)} \\ \vdots \\ b_p^{(s(p))} \end{pmatrix}, \mathrm{Id}_{\mathbb{R}^{s(p)}}\right)\right)\right) \\ &\le C(2 + s(p) + s(p)(d + 1)) \le Cs(p)d. \end{aligned}$$

The preceding result can be used to construct emulations of shape functions on non-convex patches which only use the ReLU activation.

4.2. Regular, simplicial partitions $\mathcal{T}$ including non-convex patches

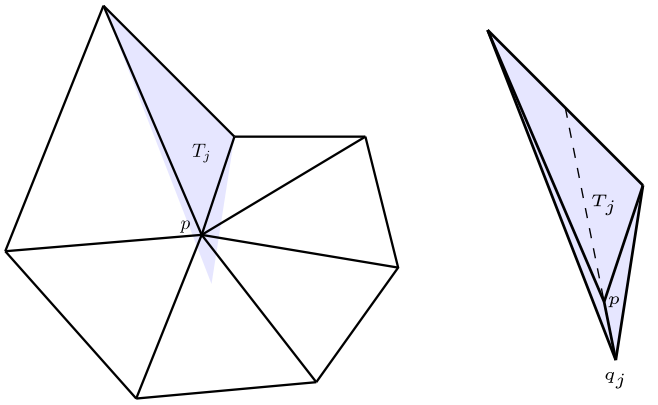
We now extend Section 4.1 to non-convex patches, i.e. we show that ReLU NNs can emulate CPwL functions on arbitrary regular, simplicial meshes in $d \in \mathbb{N}$ dimensions. To present this result in Theorem 4.3 below, we introduce some notation (see Fig. 4.1).

Given $p \in \mathcal{V}$, let $T_1, \dots, T_{s(p)} \in \mathcal{T}$ denote the simplices adjacent to p. For all $j = 1, \dots, s(p)$, let $a_0 := p$ and $a_1, \dots, a_d \in \mathbb{R}^d$ be such that $T_j = \mathrm{conv}(\{a_0, \dots, a_d\})$ and let $q_j := p + \delta_j \sum_{i=1}^d (p - a_i)$ for some sufficiently small $\delta_j > 0$. Then we define

$$\tilde{T}_{ij} := \begin{cases} \mathrm{conv}(\{q_j, a_0, \dots, a_d\} \setminus \{a_i\}) & \text{if } i \in \{1, \dots, d\}, \\ T_j & \text{if } i = 0. \end{cases} \tag{4.3}$$

Furthermore, set

$$\tilde{\omega}_j(p) := \bigcup_{i=0}^{d} \overline{\tilde{T}_{ij}}. \tag{4.4}$$



(a) $\omega(p)$ (b) $\tilde{\omega}_j(p)$

Fig. 4.1. The patches $\omega(p)$ and (shaded) $\tilde{\omega}_j(p) \subset \omega(p)$.

We build basis functions for $\mathrm{S}^1_1(\mathcal{T}, \Omega)$ starting from the hat functions $\tilde{\theta}^{\mathrm{S}^1_1}_{p,j} \in C^0(\Omega)$ for $j = 1, \dots, s(p)$ defined by

$$\begin{aligned} &\tilde{\theta}^{\mathrm{S}^1_1}_{p,j}(p) = 1 \text{ and } \tilde{\theta}^{\mathrm{S}^1_1}_{p,j}(q) = 0 \text{ for all other vertices } q \text{ of } \tilde{\omega}_j(p), \\ &\tilde{\theta}^{\mathrm{S}^1_1}_{p,j}|_{\tilde{T}_{ij}} \in \mathbb{P}_1 \text{ for all } i = 0, \dots, d, \\ &\tilde{\theta}^{\mathrm{S}^1_1}_{p,j}|_{\Omega \setminus \tilde{\omega}_j(p)} = 0. \end{aligned} \tag{4.5}$$

In Theorem 4.2 we show that CPwL basis functions with non-convex support $\omega(p)$ can be computed as the maximum of $s(p)$ many CPwL basis functions with convex support, whose ReLU NN emulation was given in Section 4.1. This maximum can be emulated exactly by a ReLU NN using the constructions in Section 2, as shown in Theorem 4.3. We obtain the same bound on the ReLU NN size as the bound on the NN size in Proposition 3.8. The proofs of these results are postponed to the Appendix.

Theorem 4.2. *For all $p \in \mathcal{V}$, let $T_1, \dots, T_{s(p)} \in \mathcal{T}$, $s(p) \in \mathbb{N}$ be the simplices adjacent to p. Then, for all $p \in \mathcal{V}$ and all $x \in \omega(p)$*

$$\theta^{\mathrm{S}^1_1}_p(x) = \max_{j=1,\dots,s(p)} \tilde{\theta}^{\mathrm{S}^1_1}_{p,j}(x) = \max_{j=1,\dots,s(p)} \max\left\{0, \min_{i\in\{0,\dots,d\}} \tilde{A}^{(i,j)}_p x + \tilde{b}^{(i,j)}_p\right\}, \tag{4.6}$$

where each $x \mapsto \tilde{A}^{(i,j)}_p x + \tilde{b}^{(i,j)}_p$ is a globally linear function fulfilling $(\tilde{A}^{(i,j)}_p x + \tilde{b}^{(i,j)}_p)|_{\tilde{T}_{ij}} = \tilde{\theta}^{\mathrm{S}^1_1}_{p,j}|_{\tilde{T}_{ij}}$.

Theorem 4.3. *For all $p \in \mathcal{V}$ let $T_1, \dots, T_{s(p)} \in \mathcal{T}$, $s(p) \in \mathbb{N}$ be the simplices adjacent to p. For $\tilde{\theta}^{\mathrm{S}^1_1}_{p,j}$, $j = 1, \dots, s(p)$ defined in (4.5), let $\tilde{\Phi}^{CPwL}_{p,j}$, $j = 1, \dots, s(p)$ be the NNs from Proposition 4.1 satisfying $\mathrm{R}(\tilde{\Phi}^{CPwL}_{p,j}) = \tilde{\theta}^{\mathrm{S}^1_1}_{p,j}$ on Ω.*

Then

$$\Phi^{CPwL}_p := \Phi^{\max}_{s(p)} \odot \mathrm{P}(\tilde{\Phi}^{CPwL}_{p,1}, \dots, \tilde{\Phi}^{CPwL}_{p,s(p)}) \tag{4.7}$$

satisfies $\mathrm{R}(\Phi^{CPwL}_p)(x) = \theta^{\mathrm{S}^1_1}_p(x)$ for all $x \in \Omega$ and

$$L(\Phi^{CPwL}_p) \le 7 + \log_2(s(p)) + \log_2(d+1), \quad M(\Phi^{CPwL}_p) \le Cd^2 s(p).$$

5. NN emulation of lowest order conforming FE spaces. Approximation rates.

Having defined explicit constructions of NN emulations of shape functions for all finite elements in the discrete de Rham complex of the lowest polynomial order (1.2), we are now in position to formulate and prove our main results: *exact NN emulations* of each of the lowest order FE spaces in the de Rham complex, on regular, simplicial partitions $\mathcal{T}$ of polytopal domains $\Omega \subset \mathbb{R}^d$. For $\blacklozenge \in \{\mathrm{S}^1_1, \mathrm{N}_0, \mathrm{RT}_0, \mathrm{S}^0_0\}$, we obtain a vector space of NNs $\mathcal{NN}(\blacklozenge; \mathcal{T}, \Omega) = \{\Phi^{\blacklozenge, v} : v \in \blacklozenge(\mathcal{T}, \Omega)\}$ such that the realization of each NN $\Phi^{\blacklozenge, v}$ equals v a.e. in Ω.

With the networks $\mathcal{NN}(\blacklozenge; \mathcal{T}, \Omega)$ at hand, we may lift known approximation results for finite elements to obtain constructive NN approximations of arbitrary functions in the Sobolev spaces belonging to the de Rham complex (1.1).

Accordingly, we first construct NN emulations of the FE spaces in Proposition 5.1, from which the approximation results are derived in Theorem 5.5. To present the next statement, we define $\mathfrak{s}(\mathcal{F}) := \max_{f\in\mathcal{F}} s(f) \le 2$, $\mathfrak{s}(\mathcal{T}) := 1$ and $s(T) := 1$ for all $T \in \mathcal{T}$.

Proposition 5.1. *Let $\Omega \subset \mathbb{R}^d$, $d \ge 2$, be a bounded, polytopal domain. For every regular, simplicial triangulation $\mathcal{T}$ of Ω and every $\blacklozenge \in \{\mathrm{S}^1_1, \mathrm{N}_0, \mathrm{RT}_0, \mathrm{S}^0_0\}$ (with the Nédélec space $\blacklozenge = \mathrm{N}_0$ excluded if $d > 3$), there exists a NN $\Phi^{\blacklozenge} := \Phi^{\blacklozenge(\mathcal{T},\Omega)}$ with ReLU and BiSU activations, which in parallel emulates the shape functions $\{\theta^{\blacklozenge}_i\}_{i\in\mathcal{I}}$ for $\mathcal{I} \in \{\mathcal{V}, \mathcal{E}, \mathcal{F}, \mathcal{T}\}$, respectively, that is $\mathrm{R}(\Phi^{\blacklozenge}) : \Omega \to \mathbb{R}^{|\mathcal{I}|}$ satisfies*

$$\mathrm{R}(\Phi^{\blacklozenge})(x)_i = \theta^{\blacklozenge}_i(x) \quad \text{for a.e. } x \in \Omega \text{ and all } i \in \mathcal{I}.$$

There exists $C > 0$ independent of d and $\mathcal{T}$ such that

$$L(\Phi^{\blacklozenge}) = \begin{cases} 5 & \text{if } \blacklozenge \in \{\mathrm{S}^1_1, \mathrm{N}_0, \mathrm{RT}_0\}, \\ 3 & \text{if } \blacklozenge = \mathrm{S}^0_0, \end{cases}$$

$$M(\Phi^{\blacklozenge}) \le Cd^2 \sum_{i\in\mathcal{I}} s(i) \le Cd^2 \mathfrak{s}(\mathcal{I}) \dim(\blacklozenge(\mathcal{T}, \Omega)).$$

For $\blacklozenge \in \{\mathrm{S}^1_1, \mathrm{N}_0, \mathrm{RT}_0, \mathrm{S}^0_0\}$ and for every FE function $v = \sum_{i\in\mathcal{I}} v_i \theta^{\blacklozenge}_i \in \blacklozenge(\mathcal{T}, \Omega)$, there exists a NN $\Phi^{\blacklozenge, v} := \Phi^{\blacklozenge(\mathcal{T},\Omega), v}$ with ReLU and BiSU activations, such that for a constant $C > 0$ independent of d and $\mathcal{T}$

$$\mathrm{R}(\Phi^{\blacklozenge, v})(x) = v(x) \quad \text{for a.e. } x \in \Omega,$$

$$L(\Phi^{\blacklozenge, v}) = \begin{cases} 5 & \text{if } \blacklozenge \in \{\mathrm{S}^1_1, \mathrm{N}_0, \mathrm{RT}_0\}, \\ 3 & \text{if } \blacklozenge = \mathrm{S}^0_0, \end{cases}$$

$$M(\Phi^{\blacklozenge, v}) \le Cd^2 \sum_{i\in\mathcal{I}} s(i) \le Cd^2 \mathfrak{s}(\mathcal{I}) \dim(\blacklozenge(\mathcal{T}, \Omega)).$$

The layer dimensions and the lists of activation functions of $\Phi^{\blacklozenge}$ and $\Phi^{\blacklozenge, v}$ are independent of v and only depend on $\mathcal{T}$ through $\{s(i)\}_{i\in\mathcal{I}}$ and $|\mathcal{I}| = \dim(\blacklozenge(\mathcal{T}, \Omega))$.

For each $\blacklozenge \in \{\mathrm{S}^1_1, \mathrm{N}_0, \mathrm{RT}_0, \mathrm{S}^0_0\}$, the set

$$\mathcal{NN}(\blacklozenge; \mathcal{T}, \Omega) := \{\Phi^{\blacklozenge, v} : v \in \blacklozenge(\mathcal{T}, \Omega)\}, \tag{5.1}$$

together with the linear operation

$$\Phi^{\blacklozenge, v} \widehat{+} \lambda \Phi^{\blacklozenge, w} := \Phi^{\blacklozenge, v+\lambda w}, \qquad \text{for all } v, w \in \blacklozenge(\mathcal{T}, \Omega) \text{ and } \lambda \in \mathbb{R} \tag{5.2}$$

is a vector space, and the map $\mathrm{R}(\cdot) : \mathcal{NN}(\blacklozenge; \mathcal{T}, \Omega) \to \blacklozenge(\mathcal{T}, \Omega)$ is a linear isomorphism.

Remark 5.2. Note that $\sum_{i\in\mathcal{I}} s(i) \le c(\mathcal{I}, d)|\mathcal{T}|$, where $c(\mathcal{V}, d) = d+1$ is the number of vertices of a d-simplex, $c(\mathcal{E}, d)$ the number of edges of a d-simplex, $c(\mathcal{F}, d)$ the number of faces of a d-simplex and $C(\mathcal{T}, d) = 1$. We obtain this inequality by observing that each element $T \in \mathcal{T}$ contributes $+1$ to $c(\mathcal{I}, d)$ terms $s(i)$. Therefore, we also have the bound $M(\Phi^{\blacklozenge}) \le Cd^2 c(\mathcal{I}, d)|\mathcal{T}|$, independent of the shape regularity constant C_{sh} of $\mathcal{T}$. The same bound holds for $M(\Phi^{\blacklozenge, v})$.

Definition 5.3. For a given polytopal domain $\Omega \subset \mathbb{R}^d$, $d \geq 2$ and a regular, simplicial triangulation $\mathcal{T}$ on Ω, we call the network $\Phi^{\blacklozenge}$ defined in Proposition 5.1 a $\blacklozenge$*-basis net*.

Proof of Proposition 5.1. We define $\Phi^{\blacklozenge(\mathcal{T},\Omega)}$ as the parallelization of networks from Propositions 3.8, 3.7, 3.2 or Eq. (3.4), namely $\Phi^{\blacklozenge(\mathcal{T},\Omega)} := \mathrm{P}(\{\Phi_i^{\blacklozenge}\}_{i\in\mathcal{I}})$, from which the formula for the realization, the formula for the NN depth and the bound on the NN size of $\Phi^{\blacklozenge(\mathcal{T},\Omega)}$ directly follow with Proposition 2.1.

The NN $\Phi^{\blacklozenge(\mathcal{T},\Omega),v}$ is defined as the sum $\Phi^{\blacklozenge(\mathcal{T},\Omega),v} := \sum_{i\in\mathcal{I}} v_i \Phi_i^{\blacklozenge}$, where the sum of NNs is as defined in Proposition 2.2, and where the NNs $v_i\Phi_i^{\blacklozenge}$ are obtained from those in Propositions 3.8, 3.7, 3.2 and Eq. (3.4) by scaling all weights and biases in the last layer by v_i. The formula for the realization, the formula for the depth and the bound on the NN size follow with Proposition 2.2.

By comparing the definition of the parallelization in Proposition 2.1 and the sum in Proposition 2.2, we observe that their hidden layers are equal. Therefore, the hidden layers of $\Phi^{\blacklozenge(\mathcal{T},\Omega),v}$ and of $\Phi^{\blacklozenge(\mathcal{T},\Omega)}$ coincide.

By definition of $\Phi^{\blacklozenge(\mathcal{T},\Omega),v}$ as linear combination of the basis NNs $\{\Phi_i^{\blacklozenge}\}_{i\in\mathcal{I}}$, which are the same for all v, the NN $\Phi^{\blacklozenge(\mathcal{T},\Omega),v}$ is determined uniquely by the coefficients $\{v_i\}_{i\in\mathcal{I}}$. Therefore, $\mathrm{R}(\cdot) : \mathcal{NN}(\blacklozenge;\mathcal{T},\Omega) \to \blacklozenge(\mathcal{T},\Omega)$ is a bijection. With the linear operations defined in (5.2), this map is linear by definition, thus a linear isomorphism.

Remark 5.4. For all $v = \sum_{i\in\mathcal{I}} v_i\theta_i^{\blacklozenge} \in \blacklozenge(\mathcal{T},\Omega)$, for $\boldsymbol{v} = (v_i)_{i\in\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|}$, the network $\Phi^{\blacklozenge,v}$ can be obtained from $\Phi^{\blacklozenge}$ as follows. Denoting the last layer weight matrix and bias vector of $\Phi_i^{\blacklozenge}$ by $A^{(i)}$ and $b^{(i)}$, those of $\Phi^{\blacklozenge}$ are given by $A = \mathrm{diag}(A^{(i_1)},\dots,A^{(i_{|\mathcal{I}|})})$ and $b = ((b^{(i_1)})^\top,\dots,(b^{(i_{|\mathcal{I}|})})^\top)^\top$, and those of $\Phi^{\blacklozenge,v}$ are given by $(v_{i_1}A^{(i_1)},\dots,v_{i_{|\mathcal{I}|}}A^{(i_{|\mathcal{I}|})})$ and $\sum_{i\in\mathcal{I}} v_i b^{(i)}$ for an enumeration $i_1,\dots,i_{|\mathcal{I}|}$ of $\mathcal{I}$.

Note that the sum defined in (5.2) differs from the sum of neural networks from Proposition 2.2. In (5.2), the hidden layers of $\Phi^{\blacklozenge,v+\lambda w}$ are independent of v, w and λ and depend only on $\blacklozenge(\mathcal{T},\Omega)$. These hidden layers coincide with those of $\Phi^{\blacklozenge}$, which emulates a basis of $\blacklozenge(\mathcal{T},\Omega)$.

For all $v \in \blacklozenge(\mathcal{T},\Omega)$ there exists a unique NN $\Phi^{\blacklozenge,v} \in \mathcal{NN}(\blacklozenge;\mathcal{T},\Omega)$ which realizes v. However, there exist many other NNs, not in $\mathcal{NN}(\blacklozenge;\mathcal{T},\Omega)$, with the same realization.

We apply the previous results to quasi-uniform, shape-regular families of meshes $\{\mathcal{T}_h\}_{h>0}$ in dimension $d = 2, 3$. For $V = H^1(\Omega)$, $H^0(\mathrm{curl},\Omega)$ for $d = 3$, $H^0(\mathrm{div},\Omega)$ or $L^2(\Omega)$, define the template for the respective smoothness space $V^\bullet \subset V$ as follows

$$\begin{aligned}
&V = H^1(\Omega) \longleftrightarrow V^\bullet = H^2(\Omega),\\
\text{for } d=3:\quad &V = H^0(\mathrm{curl},\Omega) \longleftrightarrow V^\bullet = H^1(\mathrm{curl},\Omega)\\
&\quad := \{v \in [H^1(\Omega)]^d \,:\, \mathrm{curl}\, v \in [H^1(\Omega)]^d\},\\
&V = H^0(\mathrm{div},\Omega) \longleftrightarrow V^\bullet = H^1(\mathrm{div},\Omega)\\
&\quad := \{v \in [H^1(\Omega)]^d \,:\, \mathrm{div}\, v \in H^1(\Omega)\},\\
&V = L^2(\Omega) \longleftrightarrow V^\bullet = H^1(\Omega).
\end{aligned} \tag{5.3}$$

We arrive at the following result.

Theorem 5.5. *Given a bounded, contractible polytopal Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, assume that $(V,\blacklozenge) \in \{(H^1(\Omega),\mathrm{S}_1^1),(H^0(\mathrm{curl},\Omega),\mathrm{N}_0),(H^0(\mathrm{div},\Omega),\mathrm{RT}_0),(L^2(\Omega),\mathrm{S}_0^0)\}$, that the regularity space $V^\bullet \subset V$ is as in (5.3) and that $d = 3$ if $V = H^0(\mathrm{curl},\Omega)$.*

Assume given a family $\{\mathcal{T}_h\}_{h>0}$ of regular, simplicial partitions of the polytopal domain Ω which are uniformly shape-regular and quasi-uniform with respect to the mesh-size parameter h.

Then there exists a constant $C > 0$ (depending only on the shape regularity parameter C_{sh} of the family $\{\mathcal{T}_h\}_h$ and on d) such that for all $h > 0$ and for every $v \in V^\bullet$, there exists $\Phi_h \in \mathcal{NN}(\blacklozenge;\mathcal{T}_h,\Omega)$ such that

$$\|v - \mathrm{R}(\Phi_h)\|_V \leq Ch\|v\|_{V^\bullet}$$

and

$$L(\Phi_h) = \begin{cases} 5 & \text{if } \blacklozenge \in \{\mathrm{S}_1^1,\mathrm{N}_0,\mathrm{RT}_0\},\\ 3 & \text{if } \blacklozenge = \mathrm{S}_0^0,\end{cases} \qquad M(\Phi_h) \leq Ch^{-d}.$$

Proof. Let $V_h = \blacklozenge(\mathcal{T}_h,\Omega)$ denote the lowest order FE space corresponding to V. By Proposition 5.1, for all $v_h \in V_h$ there exists a NN $\Phi_h := \Phi^{\blacklozenge,v_h} \in \mathcal{NN}(\blacklozenge;\mathcal{T}_h,\Omega)$ such that $\mathrm{R}(\Phi_h)(x) = v_h(x)$ for a.e. $x \in \Omega$. In particular, for all $v_h \in V_h$ and all $v \in V^\bullet$

$$\|v - \mathrm{R}(\Phi_h)\|_V = \|v - v_h\|_V.$$

We can then apply the approximation results e.g. Ern and Guermond (2021, Theorem 11.13) for $V = H^1(\Omega)$, (Alonso & Valli, 1999, Equations (5.7) and (5.8)) for $V = H^0(\mathrm{curl},\Omega)$ in case $d = 3$, Ern and Guermond (2021, Theorem 16.4) for $V = H^0(\mathrm{div},\Omega)$ and Poincaré's inequality for $V = L^2(\Omega)$. More precisely, for a constant C only dependent on C_{sh} and on d, for all $v \in V^\bullet$ there exists $v_h \in V_h$ for which

$$\|v - v_h\|_V \leq Ch\|v\|_{V^\bullet}. \tag{5.4}$$

The formula for $L(\Phi_h)$ follows from Proposition 5.1. In addition, the bound on the NN size follows from Proposition 5.1, together with $\dim(V_h) \sim h^{-d}$ as $h \downarrow 0$ and the fact that $\mathfrak{s}(\mathcal{I})$, $\mathcal{I} \in \{\mathcal{V},\mathcal{E},\mathcal{F},\mathcal{T}\}$ is bounded from above by a constant depending only on C_{sh}.

Remark 5.6. In (5.4) in the proof of the theorem, the choice of v_h (depending on v, given in the cited references) is made to have the approximation property (5.4). However, other choices of $v_h \in V_h$ based on interpolation or quasi-interpolation can equally be emulated with NNs. In Ern and Guermond (2017, Corollary 5.3), the authors give a particular definition of quasi-interpolants in $V_h = \blacklozenge(\mathcal{T}_h,\Omega)$ for $\blacklozenge \in \{\mathrm{S}_1^1,\mathrm{N}_0,\mathrm{RT}_0\}$, requiring minimal regularity of the function v. This gives existence of a constant $C > 0$ that is independent of v, h such that for all $v \in [W^{r,p}(\Omega)]^{d_L}$ there exists a $\Phi_h \in \mathcal{NN}(\blacklozenge;\mathcal{T}_h,\Omega)$ satisfying, for any $p \in [1,\infty]$, $r \in \{0,1\}$ or any $p \in [1,\infty)$, $r \in (0,1)$

$$\|v - \mathrm{R}(\Phi_h)\|_{[L^p(\Omega)]^{d_L}} \leq Ch^r|v|_{[W^{r,p}(\Omega)]^{d_L}}. \tag{5.5}$$

Here $d_L = d$ if $V_h = \mathrm{RT}_0(\mathcal{T}_h,\Omega)$ or $V_h = \mathrm{N}_0(\mathcal{T}_h,\Omega)$ and $d_L = 1$ otherwise. See e.g. Ern and Guermond (2021, Section 2.2) for a definition of the Sobolev space $W^{r,p}(\Omega)$ in which this result is stated.

The following analogue of Proposition 5.1 for ReLU emulation of S_1^1 also holds.

Proposition 5.7. *Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$, be a bounded, polytopal domain. For every regular, simplicial triangulation $\mathcal{T}$ of Ω, there exists a NN $\Phi^{CPwL} := \Phi^{CPwL(\mathcal{T},\Omega)}$ with only ReLU activations, which in parallel emulates the shape functions $\{\theta_i^{\mathrm{S}_1^1}\}_{i\in\mathcal{I}}$ for $\mathcal{I} = \mathcal{V}$. That is, $\mathrm{R}(\Phi^{CPwL}) : \Omega \to \mathbb{R}^{|\mathcal{I}|}$ satisfies*

$$\mathrm{R}(\Phi^{CPwL})(x)_i = \theta_i^{\mathrm{S}_1^1}(x) \quad \text{for all } x \in \Omega \text{ and all } i \in \mathcal{I}.$$

There exists $C > 0$ independent of d and $\mathcal{T}$ such that

$$\begin{aligned}
L(\Phi^{CPwL}) &\leq 8 + \log_2(\mathfrak{s}(\mathcal{I})) + \log_2(d+1),\\
M(\Phi^{CPwL}) &\leq C|\mathcal{I}|\log_2(\mathfrak{s}(\mathcal{I})) + Cd^2\sum_{i\in\mathcal{I}} s(i)\\
&\leq Cd^2\mathfrak{s}(\mathcal{I})\dim(\mathrm{S}_1^1(\mathcal{T},\Omega)).
\end{aligned}$$

For all $v = \sum_{i\in\mathcal{I}} v_i \theta_i^{S_1^1} \in S_1^1(\mathcal{T}, \Omega)$*, there exists a NN* $\Phi^{CPwL,v} := \Phi^{CPwL(\mathcal{T},\Omega),v}$ *with only ReLU activations, such that for a constant* $C > 0$ *independent of* d *and* $\mathcal{T}$

$$\begin{aligned}
&\mathrm{R}(\Phi^{CPwL,v})(x) = v(x) \quad \text{for all } x \in \Omega,\\
&L(\Phi^{CPwL,v}) \le 8 + \log_2(\mathfrak{s}(\mathcal{I})) + \log_2(d+1),\\
&M(\Phi^{CPwL,v}) \le C|\mathcal{I}|\log_2(\mathfrak{s}(\mathcal{I})) + Cd^2\sum_{i\in\mathcal{I}} s(i)\\
&\qquad\le Cd^2\mathfrak{s}(\mathcal{I})\dim(S_1^1(\mathcal{T},\Omega)).
\end{aligned}$$

The layer dimensions and the lists of activation functions of Φ^{CPwL} *and* $\Phi^{CPwL,v}$ *are independent of* v *and only depend on* $\mathcal{T}$ *through* $\{s(i)\}_{i\in\mathcal{I}}$ *and* $|\mathcal{I}| = \dim(S_1^1(\mathcal{T},\Omega))$.

The set $\mathcal{NN}(CPwL;\mathcal{T},\Omega) := \{\Phi^{CPwL,v} : v \in S_1^1(\mathcal{T},\Omega)\}$ *together with the linear operation* $\Phi^{CPwL,v} \widehat{+} \lambda\Phi^{CPwL,w} := \Phi^{CPwL,v+\lambda w}$ *for all* $v, w \in S_1^1(\mathcal{T},\Omega)$ *and all* $\lambda \in \mathbb{R}$ *is a vector space.*

The realization map $\mathrm{R}(\cdot) : \mathcal{NN}(CPwL;\mathcal{T},\Omega) \to S_1^1(\mathcal{T},\Omega)$ *is a linear isomorphism.*

Proof. We define $\Phi^{CPwL(\mathcal{T},\Omega)}$ as the parallelization of networks from Theorem 4.3, namely $\Phi^{CPwL(\mathcal{T},\Omega)} := \mathrm{P}(\{\Phi^{\mathrm{Id}}_{1,L_i} \odot \Phi_i^{CPwL}\}_{i\in\mathcal{I}})$ for $L_i = 1 + \max_{j\in\mathcal{I}} L(\Phi_j^{CPwL}) - L(\Phi_i^{CPwL})$, such that all components of the parallelization have equal depth. For the depth and size of the components, we obtain with Theorem 4.3

$$\begin{aligned}
&L(\Phi^{\mathrm{Id}}_{1,L_i} \odot \Phi_i^{CPwL}) \le 1 + (7 + \log_2(\mathfrak{s}(\mathcal{I})) + \log_2(d+1)),\\
&M(\Phi^{\mathrm{Id}}_{1,L_i} \odot \Phi_i^{CPwL}) \le CM(\Phi^{\mathrm{Id}}_{1,L_i}) + CM(\Phi_i^{CPwL})\\
&\qquad\le C(8 + \log_2(\mathfrak{s}(\mathcal{I})) + \log_2(d+1)) + Cd^2 s(i),
\end{aligned}$$

from which the stated results follow with Proposition 2.1 by the same arguments as in the proof of Proposition 5.1.

The NN $\Phi^{CPwL(\mathcal{T},\Omega),v}$ is defined as the sum $\Phi^{CPwL(\mathcal{T},\Omega),v} := \sum_{i\in\mathcal{I}} v_i \Phi^{\mathrm{Id}}_{1,L_i} \odot \Phi_i^{CPwL}$. The results now follow from Proposition 2.2 as in the proof of Proposition 5.1.

Definition 5.3 and Remark 5.4 apply, with $CPwL$ instead of $\blacklozenge$, $S_1^1(\mathcal{T},\Omega)$ instead of $\blacklozenge(\mathcal{T},\Omega)$ and $\Phi^{\mathrm{Id}}_{1,L_i} \odot \Phi_i^{CPwL}$ instead of $\Phi_i^{\blacklozenge}$. In addition, a result analogous to Theorem 5.5 follows from Proposition 5.7, with the formula for the depth replaced by $L(\Phi_h) \le C$ for a constant $C > 0$ only dependent on C_{sh} and d.

6. Neural emulation of trace spaces

In the previous sections, we have developed ReLU NN emulations of the lowest order, de Rham compatible Finite Elements on cellular complexes in the bounded Lipschitz polyhedral domains $\Omega \subset \mathbb{R}^3$. In certain applications, however, corresponding boundary complexes are required; we mention only variational boundary integral equations which arise in computational electromagnetism (e.g. Buffa et al., 2020, 2003 and the references there). We approximate traces on the boundary $\Gamma = \partial\Omega$, which is a finite union of plane sides, with the network constructions developed in Section 5 for $d = 2$. As has been emphasized e.g. in Buffa et al. (2020), trace spaces of the spaces occurring in the de Rham complex satisfy exact sequence properties derived from the compatibility of the corresponding sequences in Ω. We refer to Buffa et al. (2020, 2003) and the references there for a definition and basic properties of these spaces. We recall the trace operators (e.g. from Buffa et al. (2020, Definition 2.1)):

$$\gamma_0 : H^1(\Omega) \to H^{1/2}(\Gamma): \qquad \gamma_0(u)(x_0) = \lim_{x\to x_0} u(x), \tag{6.1a}$$

$$\check{\gamma}_0 : H^0(\mathrm{curl},\Omega) \to H^{-1/2}(\mathrm{curl}_\Gamma,\Gamma): \qquad \check{\gamma}_0(u)(x_0) = \lim_{x\to x_0} u(x) - (u(x)\cdot n_{x_0})n_{x_0}, \tag{6.1b}$$

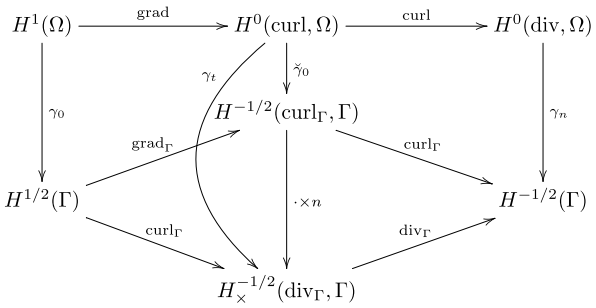


Fig. 6.1. Boundary complex.

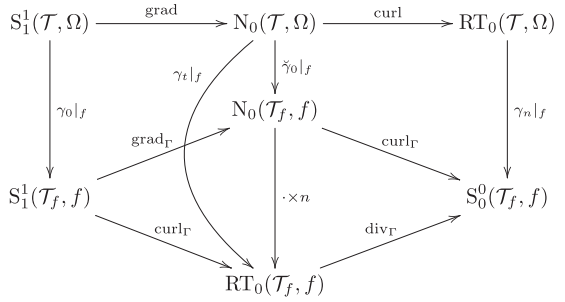


Fig. 6.2. Discrete boundary complex.

$$\gamma_t : H^0(\mathrm{curl},\Omega) \to H_\times^{-1/2}(\mathrm{div}_\Gamma,\Gamma): \qquad \gamma_t(u)(x_0) = \lim_{x\to x_0} u(x)\times n_{x_0}, \tag{6.1c}$$

$$\gamma_n : H^0(\mathrm{div},\Omega) \to H^{-1/2}(\Gamma): \qquad \gamma_n(u)(x_0) = \lim_{x\to x_0} u(x)\cdot n_{x_0}, \tag{6.1d}$$

for almost all $x_0 \in \Gamma$, where we use x to denote points in Ω, and where n_{x_0} denotes the outward unit normal to Γ in x_0. These trace operators render the diagram in Fig. 6.1 commutative (e.g. Buffa et al. (2020, Figure 2)). The trace operators in (6.1) are surjective (e.g. Buffa et al. (2020, Theorem 1)), thus the fractional Sobolev spaces on Γ in (6.1) comprise precisely all traces of elements of the respective function spaces on Ω. In addition, the trace operators in (6.1) are continuous with respect to the norms defined in Buffa et al. (2020, Section 2), see Buffa et al. (2020, Theorem 1).

Given a regular simplicial partition $\mathcal{T}$ of Ω, for each face f of Ω, the set $\mathcal{T}_f = \{\mathrm{int}(f \cap \overline{T}) : T \in \mathcal{T}\}$ is a regular, simplicial triangulation of f (where the interior $\mathrm{int}(\ldots)$ is defined with respect to the subspace topology on the face f). Discretizations of the trace spaces can be defined as the traces in the sense of (6.1) of the finite element spaces on Ω (see Fuentes, Keith, Demkowicz, and Nagaraj (2015, Section 1.6)). The corresponding diagram for the lowest order conforming FEM spaces also commutes (Fig. 6.2).

Upon parametrizing each face of Ω by a polygon in $\mathbb{R}^2$, we can construct NN approximations of the traces on f. We parametrize each face f by an affine bijection $F_f : D_f \to f$ for some polygon $D_f \subset \mathbb{R}^2$, which can be partitioned by $\mathcal{T}_{D_f} := \{F_f^{-1}(T) : T \in \mathcal{T}_f\}$. Functions in $S_1^1(\mathcal{T}_f,f)$, $\mathrm{RT}_0(\mathcal{T}_f,f)$ and $S_0^0(\mathcal{T}_f,f)$ can be pulled back to D_f. In particular,

$$\begin{aligned}
\{u\circ F_f : u \in S_1^1(\mathcal{T}_f,f)\} &= S_1^1(\mathcal{T}_{D_f},D_f),\\
\{u\circ F_f : u \in S_0^0(\mathcal{T}_f,f)\} &= S_0^0(\mathcal{T}_{D_f},D_f).
\end{aligned}$$

NN emulations of these spaces have already been provided in Propositions 5.1 and 5.7. The spaces $\mathrm{RT}_0(\mathcal{T}_f,f)$ and $\mathrm{RT}_0(\mathcal{T}_{D_f},D_f)$ are related by the Piola transform. For J denoting the Jacobian

of F_f,

$$\{\det(J)J^{-1}(u\circ F_f)\,:\,u\in \mathrm{RT}_0(\mathcal{T}_f,f)\}=\mathrm{RT}_0(\mathcal{T}_{D_f},D_f).$$

Thus, for $u\in \mathrm{RT}_0(\mathcal{T}_f,f)$, a network that emulates $u\circ F_f: D_f\to\mathbb{R}^3$ is given by $\det(J^{-1})J\Phi$, for a NN $\Phi\in\mathcal{NN}(\mathrm{RT}_0;\mathcal{T}_{D_f},D_f)$ from Proposition 5.1 emulating $\det(J)J^{-1}(u\circ F_f)\in\mathrm{RT}_0(\mathcal{T}_{D_f},D_f)$. Here, ReLU activations imply that the affine transformation $\det(J^{-1})J$ can be emulated exactly either by applying this transformation to the weights and biases of the output layer of Φ, or by concatenating Φ with a ReLU NN of depth one. In both cases, the network size is increased by at most Cd^2 (with $C>0$ independent of d and $\mathcal{T}$), and the network depth is increased by 0 respectively 1.

The shape functions of $\mathrm{N}_0(\mathcal{T}_f,f)$ equal those of $\mathrm{RT}_0(\mathcal{T}_f,f)$ up to a rotation. As explained in Section 3.3, we can use results from Section 3.2 for the NN emulation of the $\mathrm{N}_0(\mathcal{T}_{D_f},D_f)$ shape functions. Therefore, for $u\in\mathrm{N}_0(\mathcal{T}_f,f)$, a network that emulates $u\circ F_f: D_f\to\mathbb{R}^3$ is given by $\det(J^{-1})J\Phi$, for a NN $\Phi\in\mathcal{NN}(\mathrm{N}_0;\mathcal{T}_{D_f},D_f)$ from Proposition 5.1 emulating $\det(J)J^{-1}(u\circ F_f)\in\mathrm{N}_0(\mathcal{T}_{D_f},D_f)$.

The preceding discussion in this section can be summarized as follows:

Proposition 6.1. *Assume given a bounded polytopal domain $\Omega\subset\mathbb{R}^3$ with boundary $\Gamma=\partial\Omega$ consisting of a finite union of plane, polygonal faces f. For a regular, simplicial partition $\mathcal{T}$ of Ω, and for a face $f\subset\Gamma$ of Ω, consider the regular, simplicial partition $\mathcal{T}_f=\{\mathrm{int}(f\cap\overline{T})\,:\,T\in\mathcal{T}\}$ of f with edges $\mathcal{E}_f=\{e\,:\,e\subset\overline{f}\}$ and vertices $\mathcal{V}_f=\{v\,:\,v\in\overline{f}\}$ (i.e., obtained as "trace" of $\mathcal{T}$ on $f\subset\Gamma$). Let $F_f: D_f\to f$ be a bijective affine parametrization of f for some polygonal parameter domain $D_f\subset\mathbb{R}^2$ partitioned by $\mathcal{T}_{D_f}:=\{F_f^{-1}(T)\,:\,T\in\mathcal{T}_f\}$. In the following, C only depends on the shape regularity constant of the simplicial partition $\mathcal{T}_f$. Then we have the following.*

(i) For all $\blacklozenge\in\{\mathrm{S}_1^1,\mathrm{N}_0,\mathrm{RT}_0,\mathrm{S}_0^0\}$ there exists a NN $\Phi^{\blacklozenge,F_f}:=\Phi^{\blacklozenge(\mathcal{T}_f,f),F_f}$ with ReLU and BiSU activations, which in parallel emulates $\{\theta_i^{\blacklozenge}\circ F_f\}_{i\in\mathcal{I}}$ for $\mathcal{I}\in\{\mathcal{V}_f,\mathcal{E}_f,\mathcal{F}_f,\mathcal{T}_f\}$, i.e. $\mathrm{R}(\Phi^{\blacklozenge,F_f}): D_f\to\mathbb{R}^{|\mathcal{I}|}$ satisfies

$$\mathrm{R}(\Phi^{\blacklozenge,F_f})(x)_i=\theta_i^{\blacklozenge}\circ F_f(x)\quad\text{for a.e. } x\in D_f \text{ and all } i\in\mathcal{I}.$$

(ii) There exists $C>0$ independent of $\mathcal{T}$ such that

$$L(\Phi^{\blacklozenge,F_f})=\begin{cases}5 & \text{if } \blacklozenge\in\{\mathrm{S}_1^1,\mathrm{N}_0,\mathrm{RT}_0\},\\ 3 & \text{if } \blacklozenge=\mathrm{S}_0^0,\end{cases}$$

$$M(\Phi^{\blacklozenge,F_f})\le C\dim(\blacklozenge(\mathcal{T}_f,f)).$$

(iii) For all $v\in\blacklozenge(\mathcal{T}_f,f)$, there exists a DNN $\Phi^{\blacklozenge,v,F_f}:=\Phi^{\blacklozenge(\mathcal{T}_f,f),v,F_f}$ with BiSU and ReLU activations, satisfying the same depth and size bounds as $\Phi^{\blacklozenge,F_f}$, such that $\mathrm{R}(\Phi^{\blacklozenge,v,F_f})=v\circ F_f$ a.e. in D_f. The set $\mathcal{NN}(\blacklozenge;\mathcal{T}_f,f;F_f):=\{\Phi^{\blacklozenge,v,F_f}\,:\,v\in\blacklozenge(\mathcal{T}_f,f)\}$ together with the linear operation $\Phi^{\blacklozenge,v,F_f}\widehat{+}\lambda\Phi^{\blacklozenge,w,F_f}:=\Phi^{\blacklozenge,v+\lambda w,F_f}$ for all $v,w\in\blacklozenge(\mathcal{T}_f,f)$ and all $\lambda\in\mathbb{R}$ is a vector space.

(iv) There also exists a DNN $\Phi^{CPwL,F_f}:=\Phi^{CPwL(\mathcal{T}_f,f),F_f}$ of depth C and size at most $C\dim(\mathrm{S}_1^1(\mathcal{T}_f,f))$, with only ReLU activations, such that

$$\mathrm{R}(\Phi^{CPwL,F_f})(x)_i=\theta_i^{\mathrm{S}_1^1}\circ F_f(x)\quad\text{for all } x\in\overline{D_f} \text{ and all } i\in\mathcal{I}.$$

(v) For every $v\in\mathrm{S}_1^1(\mathcal{T}_f,f)$ there exists a DNN $\Phi^{CPwL,v,F_f}:=\Phi^{CPwL(\mathcal{T}_f,f),v,F_f}$ with only ReLU activations, which satisfies the same depth and size bounds as Φ^{CPwL,F_f}, and $\mathrm{R}(\Phi^{CPwL,v,F_f})=v\circ F_f$ everywhere in $\overline{D_f}$. The set $\mathcal{NN}(CPwL;\mathcal{T}_f,f;F_f):=\{\Phi^{CPwL,v,F_f}\,:\,v\in\mathrm{S}_1^1(\mathcal{T}_f,f)\}$ together with the linear operation

$$\Phi^{CPwL,v,F_f}\widehat{+}\lambda\Phi^{CPwL,w,F_f}:=\Phi^{CPwL,v+\lambda w,F_f}$$

for all $v,w\in\mathrm{S}_1^1(\mathcal{T}_f,f)$ and all $\lambda\in\mathbb{R}$

is a vector space.

7. Extensions and conclusions

We conclude this paper by indicating some extensions of the main results, as well as further possible directions of research.

7.1. Higher order polynomial spaces

For polynomial degree $k\in\mathbb{N}$ and space dimension $d\ge 2$ denote in the following by $\mathbb{P}_k:=\mathrm{span}\{\prod_{j=1}^d x_j^{\nu_j}\,:\,\sum_{j=1}^d\nu_j\le k\}$ the space of d - variate polynomials of total degree at most k. As observed in Li, Tang, and Yu (2020), networks employing the "ReLU^r"[5] activation

$$\rho_r(x):=\rho(x)^r=\max\{0,x\}^r$$

for some fixed integer $r\ge 2$, can be used to express multivariate polynomials in $\mathbb{P}_k$ exactly. We use here a formulation of this result from Opschoor, Schwab, and Zech (2022),[6] extended to vector-valued polynomials by parallelization:

Proposition 7.1 (*Opschoor et al. (2022, Proposition 2.14)*)**.** *Fix d, $\mu\in\mathbb{N}$, $r\in\mathbb{N}$, $r\ge 2$ and a polynomial degree $k\in\mathbb{N}$.*

Then there exists a constant $C>0$ independent of d, μ and k but depending on r such that for any multivariate polynomial $w\in[\mathbb{P}_k]^\mu$ there is a NN Φ_w, employing ReLU^r activation, such that $\mathrm{R}(\Phi_w)(x)=w(x)$, for all $x\in\mathbb{R}^d$ and such that $M(\Phi_w)\le C\mu(k+1)^d$ and $L(\Phi_w)\le Cd\log_2(k+1)$.

Combining Proposition 7.1 with Proposition 2.8 and Lemma 2.9, by a similar argument as in Lemma 3.1 we obtain a generalization of this result to piecewise polynomial functions on regular, simplicial partitions for all interelement-conformities which arise from compatibility with the complex (1.1).

Lemma 7.2 (*Emulation of Piecewise Higher Order Polynomial Elements*)**.** *Let $r\in\mathbb{N}$, $r\ge 2$. For $d,s,\mu,k\in\mathbb{N}$ let $\Omega\subset\mathbb{R}^d$ be a bounded polytope with boundary $\partial\Omega$ being a finite union of plane, polytopal faces and let $\mathcal{T}$ be a regular, simplicial partition of Ω with $s=|\mathcal{T}|$ elements, $\mathcal{T}=\{T_i\}_{i=1,\dots,s}$. Let $u:\Omega\to\mathbb{R}^\mu$ be a function that for all $i=1,\dots,s$ satisfies $u|_{T_i}\in[\mathbb{P}_k]^\mu$.*

Then there exists a NN Φ_u^{PwP} employing ReLU, ReLU^r and BiSU activations and satisfies $u(x)=\mathrm{R}\left(\Phi_u^{PwP}\right)(x)$ for all $x\in\cup_{i=1}^s T_i$ and $\mathrm{R}\left(\Phi_u^{PwP}\right)(x)=0$ for all $x\in\mathbb{R}^d\setminus\cup_{i=1}^s T_i$. Furthermore,

$$L(\Phi_u^{PwP})\le Cd\log_2(k+1)\,,\qquad M(\Phi_u^{PwP})\le Cs\mu(k+1)^d.$$

Here the constant C is independent of $\mathcal{T}$, d, s, μ and of k but depends on r.

Our results thus straightforwardly extend to piecewise polynomial spaces of arbitrarily high order, covering all de Rham compatible element families on simplicial partitions on polytopes as described in Fuentes et al. (2015). Importantly, as in the case of low-order finite elements, the network size only scales linearly in the number $s=|\mathcal{T}|$ of simplices of the triangulation $\mathcal{T}$. Similarly, also the results of Section 6 extend to higher order polynomials.

We now state the corresponding generalization of Proposition 5.1 for three types of higher order finite elements. For arbitrary polynomial degree $k\in\mathbb{N}$ we recall the Lagrange FE space

$$\mathrm{S}_k^1(\mathcal{T},\Omega):=\{v\in H^1(\Omega)\,:\,v|_T\in\mathbb{P}_k,\ \forall T\in\mathcal{T}\}\subset H^1(\Omega)\qquad(7.1)$$

[5] Also referred to as "rectified power unit" (RePU).

[6] We apply this result here with the multiindex set $\Lambda:=\{(\nu_1,\dots,\nu_d)\in\mathbb{N}_0^d\,:\,\sum_{j=1}^d\nu_j\le k\}$, which has cardinality bounded by $(k+1)^d$. Here, we denoted $\mathbb{N}_0=\{0,1,\dots\}$.

from Ern and Guermond (2021, Section 7.4). For all $k \in \mathbb{N}_0$ we recall $\mathbb{RT}_k = (\mathbb{P}_k)^d \oplus x\,\mathrm{span}\{x^\alpha : \alpha \in \mathbb{N}_0^d, |\alpha| = k\}$ and the Raviart–Thomas FE space

$$\mathrm{RT}_k(\mathcal{T},\Omega) := \{v \in (L^1(\Omega))^d : v|_T \in \mathbb{RT}_k\ \forall T \in \mathcal{T} \text{ and } [v\cdot n_f]_f = 0\ \forall f \subset \Omega\} \subset H^0(\mathrm{div},\Omega) \tag{7.2}$$

from Ern and Guermond (2021, Sections 14.2 and 14.3), and let for all $k \in \mathbb{N}_0$

$$\mathrm{S}_k^0(\mathcal{T},\Omega) := \{v \in L^1(\Omega) : v|_T \in \mathbb{P}_k\ \forall T \in \mathcal{T}\} \subset L^2(\Omega). \tag{7.3}$$

For all $k \in \mathbb{N}$ and $\blacklozenge \in \{\mathrm{S}_k^1, \mathrm{RT}_{k-1}, \mathrm{S}_k^0\}$ and for a suitable index set $\mathcal{I}$ we will denote by $\{\theta_i^\blacklozenge\}_{i\in\mathcal{I}}$ any collection of shape functions of $\blacklozenge(\mathcal{T},\Omega)$, each of which is supported on $s(i) \le \mathfrak{s}(\mathcal{I})$ elements of $\mathcal{T}$.

Proposition 7.3. *Let $\Omega \subset \mathbb{R}^d$, $d \ge 2$, be a bounded, polytopal domain and let $r \in \mathbb{N}$, $r \ge 2$ be the power in the ReLU^r activation, and let $k \ge 1$ denote the element degree.*

Then we have the following.

(i) For every regular, simplicial triangulation $\mathcal{T}$ of Ω, every $k \in \mathbb{N}$ and every $\blacklozenge \in \{\mathrm{S}_k^1, \mathrm{RT}_{k-1}, \mathrm{S}_k^0\}$ there exists a NN $\Phi^\blacklozenge := \Phi^{\blacklozenge(\mathcal{T},\Omega)}$ with ReLU, ReLU^r and BiSU activations, which in parallel emulates the basis functions $\{\theta_i^\blacklozenge\}_{i\in\mathcal{I}}$, that is $\mathrm{R}(\Phi^\blacklozenge): \Omega \to \mathbb{R}^{|\mathcal{I}|}$ satisfies

$$\mathrm{R}(\Phi^\blacklozenge)(x)_i = \theta_i^\blacklozenge(x) \quad \text{for a.e. } x \in \Omega \text{ and all } i \in \mathcal{I}.$$

(ii) There exists $C > 0$ independent of d, k and $\mathcal{T}$, but depending on r, such that with $\mu = 1$ if $\blacklozenge \in \{\mathrm{S}_k^1, \mathrm{S}_k^0\}$ and $\mu = d$ if $\blacklozenge = \mathrm{RT}_{k-1}$,

$$L(\Phi^\blacklozenge) \le Cd\log(k+1), \qquad M(\Phi^\blacklozenge) \le C\mu(k+1)^d \sum_{i\in\mathcal{I}} s(i).$$

(iii) For every FE function $v = \sum_{i\in\mathcal{I}} v_i\theta_i^\blacklozenge \in \blacklozenge(\mathcal{T},\Omega)$, exists a NN $\Phi^{\blacklozenge,v} := \Phi^{\blacklozenge(\mathcal{T},\Omega),v}$ with ReLU, ReLU^r and BiSU activations, such that for a constant $C > 0$ independent of d, k and $\mathcal{T}$, but depending on r,

$$\mathrm{R}(\Phi^{\blacklozenge,v})(x) = v(x) \quad \text{for a.e. } x \in \Omega,$$
$$L(\Phi^{\blacklozenge,v}) \le Cd\log(k+1), \qquad M(\Phi^{\blacklozenge,v}) \le C\mu(k+1)^d \sum_{i\in\mathcal{I}} s(i).$$

The layer dimensions and the lists of activation functions of $\Phi^\blacklozenge$ and $\Phi^{\blacklozenge,v}$ are independent of v and only depend on $\mathcal{T}$ through $\{s(i)\}_{i\in\mathcal{I}}$ and $|\mathcal{I}| = \dim(\blacklozenge(\mathcal{T},\Omega))$.

(iv) For each $\blacklozenge \in \{\mathrm{S}_k^1, \mathrm{RT}_{k-1}, \mathrm{S}_k^0\}$,

$$\mathcal{NN}(\blacklozenge;\mathcal{T},\Omega) := \{\Phi^{\blacklozenge,v} : v \in \blacklozenge(\mathcal{T},\Omega)\}, \tag{7.4}$$

together with the linear operation $\Phi^{\blacklozenge,v} \widehat{+} \lambda\Phi^{\blacklozenge,w} := \Phi^{\blacklozenge,v+\lambda w}$ for all $v, w \in \blacklozenge(\mathcal{T},\Omega)$ and $\lambda \in \mathbb{R}$ is a vector space, and the map $\mathrm{R}(\cdot) : \mathcal{NN}(\blacklozenge;\mathcal{T},\Omega) \to \blacklozenge(\mathcal{T},\Omega)$ is a linear isomorphism.

Proof. For all $i \in \mathcal{I}$ let $\Phi_i^\blacklozenge$ be the NN approximation of $\theta_i^\blacklozenge$ from Lemma 7.2. Possibly after concatenating each $\Phi_i^\blacklozenge$ with $\Phi_{\mu,2}^{\mathrm{Id}}$, which only affects the constant C in the bounds on depth and size from Lemma 7.2, we may assume that $L(\Phi_i^\blacklozenge) \ge 3 = L(\Phi_{T_i}^{\mathbb{1}})$. From the proof of Proposition 7.1, where Φ_w is the μ-fold parallelization of a ReLU^r network from Opschoor et al. (2022, Proposition 2.14) applied with $\Lambda = \{\boldsymbol{\nu} \in \mathbb{N}_0^d : \sum_{j=1}^d \nu_j \le k\}$, we see that the depth and the layer dimensions depend on d, μ and r, but not on w. The same holds for the network in Lemma 7.2, by an argument similar to that in the proof of Lemma 3.1. We can therefore define $\Phi^{\blacklozenge(\mathcal{T},\Omega)} := \mathrm{P}\left(\{\Phi_i^\blacklozenge\}_{i\in\mathcal{I}}\right)$ and the sum $\Phi^{\blacklozenge(\mathcal{T},\Omega),v} := \sum_{i\in\mathcal{I}} v_i\Phi_i^\blacklozenge$ and obtain the linear structure of $\mathcal{NN}(\blacklozenge;\mathcal{T},\Omega)$ by the same arguments as in the proof of Proposition 5.1.

It remains to prove the formula for the realization and to estimate the NN depth and size. Firstly, we observe that indeed $\theta_i^\blacklozenge(x) = \mathrm{R}\left(\Phi^\blacklozenge\right)(x)_i$ for all $x \in \Omega \setminus \partial\mathcal{T}$, where $\partial\mathcal{T} := \bigcup_{T\in\mathcal{T}} \partial T$, and $\mathrm{R}\left(\Phi^\blacklozenge\right)(x)_i = 0$ else.

Secondly, we apply Lemma 7.2 with $s = s(i)$ and $\mu = 1$ if $\blacklozenge \in \{\mathrm{S}_k^1, \mathrm{S}_k^0\}$ and $\mu = d$ if $\blacklozenge = \mathrm{RT}_{k-1}$. We obtain that

$$L(\Phi^\blacklozenge) \le Cd\log_2(k+1),$$
$$M(\Phi^\blacklozenge) \le \sum_{i=1}^{s(i)} M(\Phi_i^\blacklozenge) \le \sum_{i\in\mathcal{I}} Cs(i)\mu(k+1)^d,$$

and the same bounds hold for the depth and size of $\Phi^{\blacklozenge,v}$.

Definition 5.3 applies, and also Remark 5.4.

Remark 7.4. ReLU NNs (and thus also ReLU+BiSU NNs) are known to be efficient at *approximating* multivariate polynomials, see e.g. Liang and Srikant (2017), Opschoor et al. (2022), Yarotsky (2017). Thus, also ReLU+BiSU (rather than ReLU+ReLU^r+BiSU) networks could be employed to extend our results to higher order polynomial spaces, however only in an approximate sense.

The resulting PwL NN realizations may violate the discrete exact sequence property however.

7.2. Crouzeix–Raviart elements CR_0

While this work focused on conformal discretization of functions in the compatible spaces in (1.1), the result of Lemma 3.1 is more general and includes the non-conformal Crouzeix–Raviart elements (e.g. Ern and Guermond (2021, Section 7.5)) of lowest order for $d \ge 2$. Due to the importance and widespread use of the Crouzeix–Raviart elements (e.g. Balci et al., 2022; Chambolle & Pock, 2020; Crouzeix & Falk, 1989 and the references there), we state a NN emulation result of these elements. For $d \ge 2$ and a polytopal domain $\Omega \subset \mathbb{R}^d$, let $\mathcal{T}$ be a regular, simplicial triangulation of Ω as in Section 3. The lowest order Crouzeix–Raviart FE space is defined as

$$\mathrm{CR}_0(\mathcal{T},\Omega) := \{v \in L^1(\Omega) : v|_T \in \mathbb{P}_1\ \forall T \in \mathcal{T} \text{ and } \int_f [v]_f = 0\ \forall f \in \mathcal{F}, f \subset \Omega\}, \tag{7.5}$$

where $[v]_f$ denotes the jump of a function across f, that is, given a unit normal vector n_f to f, $[v]_f(x_0) = \lim_{\epsilon\searrow 0}(v(x_0+\epsilon n_f) - v(x_0-\epsilon n_f))$ for all $x_0 \in f$. Analogously to the case of Raviart–Thomas FE, the space $\mathrm{CR}_0(\mathcal{T},\Omega)$ has one degree of freedom per face $f \in \mathcal{F}$. The corresponding shape functions are, for $f \subset \partial\Omega$ and thus $s(f) = 1$, $\theta_f^{\mathrm{CR}_0}(x) := d(\frac{1}{d} - (1 - \frac{|f|(x-a_1)\cdot n_f}{d|T_1|}))\mathbb{1}_{T_1}$, where $f \subset \overline{T_1}$, $T_1 \in \mathcal{T}$ and a_1 is the only vertex of T_1 that does not belong to $\overline{f}$. For interior faces $f \subset \Omega$ and thus $s(f) = 2$, we construct $\theta_f^{\mathrm{CR}_0}$ by assembling local shape functions of the neighboring simplices T_1, T_2 with $\overline{f} = \overline{T_1} \cap \overline{T_2}$,

$$\theta_f^{\mathrm{CR}_0}(x) := \begin{cases} d(\frac{1}{d} - (1 - \frac{|f|(x-a_1)\cdot n_f}{d|T_1|})) & \text{if } x \in T_1, \\ d(\frac{1}{d} - (1 + \frac{|f|(x-a_2)\cdot n_f}{d|T_2|})) & \text{if } x \in T_2, \\ 0 & \text{if } x \notin \overline{T_1 \cup T_2}, \end{cases} \tag{7.6}$$

where a_1, a_2 are the only vertices of T_1, T_2, respectively, not belonging to $\overline{f}$. The following proposition allows us to apply Proposition 5.1, Definition 5.3 and Remark 5.4 with $\blacklozenge = \mathrm{CR}_0$.

Proposition 7.5. *Given $f \in \mathcal{F}$, let $\{T_i\}_{i=1}^{s(f)}$ be the simplices adjacent to f and let $a_i := (\mathcal{V} \cap \overline{T_i}) \setminus \overline{f} \in \mathbb{R}^d$, $i = 1, \dots, s(f)$. Then, there exist*

$A_{f,T_i} \in \mathbb{R}^{1\times d}$, $b_{f,T_i} \in \mathbb{R}$, $i = 1, \dots, s(f)$ *such that*

$$\Phi_f^{\mathrm{CR}_0} := \sum_{i=1}^{s(f)} \Phi_{1,\kappa}^{\times} \odot \mathrm{P}\left(\Phi_{1,2}^{\mathrm{Id}} \odot \left(\left(A_{f,T_i}, b_{f,T_i}, \mathrm{Id}_{\mathbb{R}}\right)\right), \Phi_{T_i}^{\mathbb{1}}\right) \tag{7.7}$$

satisfies $\theta_f^{\mathrm{CR}_0}(x) = \mathrm{R}(\Phi_f^{\mathrm{CR}_0})(x)$ *for a.e.* $x \in \Omega$, *for any*

$$\kappa \geq d - 1. \tag{7.8}$$

In addition, there exists a constant $C > 0$ *that is independent of* d *and* $\mathcal{T}$ *such that for all* $f \in \mathcal{F}$

$$L(\Phi_f^{\mathrm{CR}_0}) = 5, \qquad M(\Phi_f^{\mathrm{CR}_0}) \leq Cd^2 s(f) \leq 2Cd^2.$$

Proof. The values of A_{f,T_i}, b_{f,T_i} can be read from (7.6). Similar to Proposition 3.2, $\theta_f^{\mathrm{CR}_0}(x) = \mathrm{R}(\Phi_f^{\mathrm{CR}_0})(x)$ for all $x \in \Omega \setminus \partial\mathcal{T}$, where $\partial\mathcal{T} := \bigcup_{T\in\mathcal{T}} \partial T$.

We conclude applying Lemma 3.1 with $\mu = 1$, $m = d + 1$ and $s = s(f)$.

The same idea carries over to higher order Crouzeix–Raviart elements and canonical hybrid elements (Ern & Guermond, 2021, Section 7.6), along the lines of Section 7.1.

7.3. Domains of general topology

In our discussion of the de Rham complex (see Section 1.4.2) we assumed throughout that the physical domain Ω is contractible. This renders the topology of Ω trivial: its Betti-numbers are $b_0 = 1$, $b_1 = b_2 = b_3 = 0$. As is well-known, for polytopal domains Ω with a non-trivial topology (e.g. domains $\Omega \subset \mathbb{R}^3$ with voids) in (1.1) the cohomology spaces

$$\mathcal{H}_0 := \text{Ker grad/Im } i \qquad \mathcal{H}_1 := \text{Ker curl/Im grad}$$
$$\mathcal{H}_2 := \text{Ker div /Im curl} \qquad \mathcal{H}_3 := L^2(\Omega)/\text{Im div}$$

are nontrivial. Our neural network emulation results are given without topological restrictions on the bounded polytopal domain Ω. Therefore, the presently proposed DNN emulations of de Rham compatible FE spaces on simplicial partitions preserve these cohomology spaces provided that the corresponding discrete homology spaces $\mathcal{H}_i(\mathcal{T})$ for the FE spaces in Ω are isomorphic to $\mathcal{H}_i$. This property has been verified for several large classes of FE spaces (see, e.g., Di Pietro, Droniou, & Pitassi, 2022 and the references there).

7.4. Conclusions

The present construction of deep NN emulations of de Rham compatible Finite Element spaces was given for the lowest order Finite Element families on regular, simplicial partitions $\mathcal{T}$ of Ω. Generalizing recent work (He et al., 2020), we provided exact emulation of continuous piecewise linear functions ("Courant" Finite Elements) on arbitrary, regular simplicial partitions in any space dimension by ReLU networks. As shown, for uniformly shape regular partitions the network size in this construction merely scales linearly with the number of elements.

As is well known (e.g. Fuentes et al. (2015) and the reference there) the presently emulated, lowest order element families are embedded in hierarchies of higher-order Finite Element families for arbitrary polynomial order. We argued that admitting higher order, so-called ReLU^r activations with $r \in \mathbb{N}$, $r \geq 2$ allows to exactly emulate the higher order element families from Fuentes et al. (2015) along the lines of the present constructions.

Compatible constructions similar to the ones developed here are also possible on *affine partitions* $\mathcal{T}$ (comprising elements that are affine images of reference elements) which contain other element shapes, in particular quadrilaterals ($d = 2$) and hexahedral elements ($d = 3$). We refer to Fuentes et al. (2015, Sec. 4 and 6) for details on the shape functions.

The present results, in particular Proposition 6.1, can be the basis to extend the recently proposed frameworks of "PiNN" (Raissi et al., 2019) and "deep Ritz" (E & Yu, 2018) for DNN discretization of PDEs to larger classes of PDEs, and to corresponding boundary integral formulations (see, e.g., Sauter & Schwab, 2011 for such methods, and Aylwin et al. (2023) for a realization of this approach for a model problem). While in this paper we mainly concentrated on the de Rham formalism, our ideas and proofs naturally extend also to compatible discretizations of more general structures, as occur in the so-called Finite Element Exterior Calculus (FEEC) (e.g. Arnold, Falk, & Winther, 2006 and the references there).

Similarly, with Lemma 7.2 other nonconforming FEM such as Hybridized, High Order ("HHO") FEM can be emulated with appropriate functionals which account for element interface unknowns and reduced interelement conformity, see, e.g. Cicuttin, Ern, and Pignet (2021, Prop. 1.8).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

ChS acknowledges stimulating discussions at the workshop "Deep learning and partial differential equations MDLW03 15 November 2021 to 19 November 2021" at the Isaac Newton Institute, Cambridge, UK, during the program "Mathematics of deep learning MDL 1 July 2021 to 17 December 2021". Excellent online conferencing and discussion facilitation by the INI and sabbatical leave from ETH Zürich during the autumn term 2021 are warmly acknowledged.

Appendix. Proofs

A.1. Proofs from Section 2

Proof of Proposition 2.8. This proof is in two steps. In Step 1, we define a function of x, y that computes the desired output for $d = 1$. In Step 2, we construct a NN which exactly emulates that function d times and estimate its depth and size.

Step 1. For $x, y \in \mathbb{R}$ let

$$f(x, y) := \tfrac{1}{2}\left(\rho(x + y) + \rho(-x - y) - \rho(x - y) - \rho(-x + y)\right).$$

Note that for all $x \in [-1, 1]$ and $y \in [0, 1]$ such that $|x| \leq y \leq 1$ it holds that $f(x, y) = \frac{1}{2}\rho(x + y) + 0 - 0 - \frac{1}{2}\rho(-x + y) = x$ and that for all $x \in [-1, 1]$ it holds that $f(x, 0) = \frac{1}{2}\rho(x) + \frac{1}{2}\rho(-x) - \frac{1}{2}\rho(-x) - \frac{1}{2}\rho(x) = 0$. Hence, f satisfies

$$f(x, y) = xy, \text{ for all } x \in [-1, 1] \text{ and } y \in \{0, 1\},$$

and thus for all $x \in [-\kappa, \kappa]$ and $y \in \{0, 1\}$ it follows that

$$\kappa f(\tfrac{x}{\kappa}, y) = xy.$$

Step 2. For $d = 1$, let

$$\Phi^{\times}_{1,\kappa} := \left(\left(\begin{pmatrix} \frac{1}{\kappa} & 1 \\ -\frac{1}{\kappa} & -1 \\ \frac{1}{\kappa} & -1 \\ -\frac{1}{\kappa} & 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \rho \\ \rho \\ \rho \\ \rho \end{pmatrix}\right), \left(\begin{pmatrix} \frac{\kappa}{2} & \frac{\kappa}{2} & -\frac{\kappa}{2} & -\frac{\kappa}{2} \end{pmatrix}, 0, \mathrm{Id}_{\mathbb{R}}\right)\right),$$

which satisfies

$$\mathrm{R}(\Phi^{\times}_{1,\kappa})(x,y) = \kappa f(\tfrac{x}{\kappa}, y) \quad \text{for all } (x,y) \in \mathbb{R}^2,$$
$$L(\Phi^{\times}_{1,\kappa}) = 2, \qquad M(\Phi^{\times}_{1,\kappa}) = 8 + 4 = 12.$$

Similarly, with $u_1 := (\frac{1}{\kappa}, -\frac{1}{\kappa}, \frac{1}{\kappa}, -\frac{1}{\kappa})^\top$, $u_2 := (1,-1,-1,1)^\top$ and $u_3 := (\frac{\kappa}{2}, \frac{\kappa}{2}, -\frac{\kappa}{2}, -\frac{\kappa}{2})$, we define for $d > 1$ the following network with layer sizes $N_0 = d+1$, $N_1 = 4d$ and $N_2 = d$:

$$\Phi^{\times}_{d,\kappa} := \left(\left(\begin{pmatrix} u_1 & & u_2 \\ & \ddots & \vdots \\ & u_1 & u_2 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \rho \\ \vdots \\ \rho \end{pmatrix}\right), \left(\begin{pmatrix} u_3 & & \\ & \ddots & \\ & & u_3 \end{pmatrix}, \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} \mathrm{Id}_{\mathbb{R}} \\ \vdots \\ \mathrm{Id}_{\mathbb{R}} \end{pmatrix}\right)\right),$$

which satisfies

$$\mathrm{R}(\Phi^{\times}_{d,\kappa})(x_1,\dots,x_d,y) = (\kappa f(\tfrac{x_1}{\kappa}, y), \dots, \kappa f(\tfrac{x_d}{\kappa}, y))^\top \in \mathbb{R}^d, \quad \text{for all } (x,y) \in \mathbb{R}^d \times \mathbb{R},$$
$$L(\Phi^{\times}_{d,\kappa}) = 2, \qquad M(\Phi^{\times}_{d,\kappa}) = 12d.$$

Proof of Lemma 2.9. From

$$1 - \sigma(y) - \sigma(-y) = \begin{cases} 1 & \text{if } y = 0, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for all } y \in \mathbb{R},$$

it follows that for all $x \in \mathbb{R}^d$

$$\mathrm{R}(\Phi^{\mathbb{1}}_{\Omega})(x) = \sigma\left(\sum_{i=1}^{n}(1 - \sigma(A_i x + b_i) - \sigma(-A_i x - b_i)) + \sum_{i=n+1}^{N} \sigma(A_i x + b_i) - (N - \tfrac{1}{4})\right) = \begin{cases} 1 & \text{if } x \in \Omega, \\ 0 & \text{otherwise,} \end{cases}$$
$$L(\Phi^{\mathbb{1}}_{\Omega}) = 3, \qquad M(\Phi^{\mathbb{1}}_{\Omega}) \le ((N+n)d + (N+n)) + ((N+n)+1) + 1 = (d+2)(N+n) + 2.$$

A.2. Proofs from Section 3

Proof of Proposition 3.4. Below, we prove the result for $f \subset \Omega$, i.e. $s(f) = 2$. The case $f \subset \partial\Omega$, i.e. $s(f) = 1$, follows analogously.

Observe that we can write $T_i = \mathrm{conv}(\{a_i, p_1, \dots, p_d\})$ where $\{p_1, \dots, p_d\} := \mathcal{V} \cap \bar{f}$. The point values $\theta_f^{\mathrm{RT}_0}(p_j) \cdot n_f = 1$, $\forall j = 1, \dots, d$ are well-defined by continuity of $\theta_f^{\mathrm{RT}_0} \cdot n_f$ across f. Therefore, we can take $A^{(i)}_{n_f} \in \mathbb{R}^{1\times d}$, $b^{(i)}_{n_f} \in \mathbb{R}$, $i = 1, 2$ to be the matrices and vectors solving

$$(A^{(i)}_{n_f}, b^{(i)}_{n_f}) \begin{pmatrix} (a_i)_1 & (p_1)_1 & & (p_d)_1 \\ \vdots & & \ddots & \\ (a_i)_d & (p_1)_d & & (p_d)_d \\ 1 & 1 & \cdots & 1 \end{pmatrix} = (0, 1, \dots, 1), \tag{A.1}$$

where $(0,1,\dots,1) = (-1)^{i-1}\frac{|f|}{d|T_i|}(0, (p_1 - a_i)\cdot n_f, \dots, (p_d - a_i)\cdot n_f)$. With this choice, since $(\theta_f^{\mathrm{RT}_0}(x)\cdot n_f) \in [0,1]$ for $x \in T_1 \cup T_2 \cup f$, it holds that $\theta_f^{\mathrm{RT}_0}(x)\cdot n_f = \mathrm{R}\left(\Phi_f^{\mathrm{RT}_0,\perp}\right)(x)$ for a.e. $x \in \Omega$ and every $x \in f$. On the other hand, the discontinuous tangential component can be assembled element by element, as in Proposition 3.2: matrices and vectors $(A^{(i)}_{t_j}, b^{(i)}_{t_j}) \in \mathbb{R}^{1\times(d+1)}$, $i = 1, 2$ and $j = 1, \dots, d-1$ which only depend on T_i and t_j can be computed as in (A.1), but with different right-hand sides, namely $(-1)^{i-1}\frac{|f|}{d|T_i|}(0, (p_1 - a_i)\cdot t_j, \dots, (p_d - a_i)\cdot t_j) \in \mathbb{R}^{1\times(d+1)}$.

Finally, we estimate the network depth and size. For $d = 2$, as in Remark 2.6 let

$$\Phi^{\min}_2 := \left(\left(\begin{pmatrix} -1 & 1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \rho \\ \rho \\ \rho \end{pmatrix}\right), \left(\begin{pmatrix} -1 & 1 & -1 \end{pmatrix}, 0, \mathrm{Id}_{\mathbb{R}}\right)\right),$$
$$L(\Phi^{\min}_2) = 2, \qquad M(\Phi^{\min}_2) = 7.$$

In particular, we use that $L(\Phi^{\min}_2) + 1 = L(\Phi^{\mathbb{1}}_{T_1} + \Phi^{\mathbb{1}}_f + \Phi^{\mathbb{1}}_{T_2}) = 3$, i.e. in (3.9) both components in the parallelization have equal depth. Also, because the networks for $s(f) = 1$ have smaller sizes than those for $s(f) = 2$, we will only estimate the sizes of the latter. In the bound on the size of $\Phi^{\mathbb{1}}_{T_1} + \Phi^{\mathbb{1}}_f + \Phi^{\mathbb{1}}_{T_2}$ in (3.9), we use for T_1, T_2 Lemma 2.9 with $N = d+1$ and $n = 0$, whereas for f we use Lemma 2.9 with $N = d+1$ and $n = 1$. The size of the network in (3.10) is estimated using Lemma 3.1 with $\mu = 1$, $m = d+1$ and $s = 2$.

$$\begin{aligned} L(\Phi_f^{\mathrm{RT}_0,\perp}) &= L(\Phi^{\times}_{1,1}) + L(\Phi^{\mathbb{1}}_{T_1} + \Phi^{\mathbb{1}}_f + \Phi^{\mathbb{1}}_{T_2}) = 5, \\ M(\Phi_f^{\mathrm{RT}_0,\perp}) &\le 2M(\Phi^{\times}_{1,1}) \\ &\quad + 2M\left(\mathrm{P}\left(\Phi^{\min}_2 \odot \left(\left(\begin{pmatrix} A^{(1)}_{n_f} \\ A^{(2)}_{n_f} \end{pmatrix}, \begin{pmatrix} b^{(1)}_{n_f} \\ b^{(2)}_{n_f} \end{pmatrix}, \mathrm{Id}_{\mathbb{R}^2}\right)\right), \Phi^{\mathbb{1}}_{T_1} + \Phi^{\mathbb{1}}_f + \Phi^{\mathbb{1}}_{T_2}\right)\right) \\ &\le 2M(\Phi^{\times}_{1,1}) + 4M(\Phi^{\min}_2) \\ &\quad + 4M\left(\left(\left(\begin{pmatrix} A^{(1)}_{n_f} \\ A^{(2)}_{n_f} \end{pmatrix}, \begin{pmatrix} b^{(1)}_{n_f} \\ b^{(2)}_{n_f} \end{pmatrix}, \mathrm{Id}_{\mathbb{R}^2}\right)\right)\right) \\ &\quad + 2M(\Phi^{\mathbb{1}}_{T_1}) + 2M(\Phi^{\mathbb{1}}_f) + 2M(\Phi^{\mathbb{1}}_{T_2}) \\ &\le C(C + C + Cd + Cd^2 + Cd^2 + Cd^2) \le Cd^2. \end{aligned}$$

A.3. Proofs from Section 4

In this section we give proofs of Theorems 4.2 and 4.3. Our proof strategy is to write a non-convex patch as a suitable union of convex ones, and thereby reduce the problem to the convex case.

Lemma A.1. *For $d \in \mathbb{N}$, let $T = \mathrm{conv}(\{a_0, \dots, a_d\})$ be a simplex and $\delta > 0$. Define $q := a_0 + \delta\sum_{i=1}^{d}(a_0 - a_i)$.*

Then $T_\delta := \mathrm{conv}(\{q, a_1, \dots, a_d\})$ is a simplex and $a_0 \in T_\delta$.

Proof. Without loss of generality, $a_0 = 0$. To show that T_δ is a simplex, it suffices to verify $a_0 = 0 \in T_\delta$, as it then follows that $T \subset T_\delta$, i.e. T_δ has nonempty interior and is thus a simplex. By definition,

$$T_\delta = \left\{\alpha_0\left(\delta\sum_{i=1}^{d} -a_i\right) + \sum_{i=1}^{d}\alpha_i a_i : \sum_{i=0}^{d}\alpha_i = 1 \text{ and } \alpha_i > 0\right\}.$$

Therefore, $a_0 \in T_\delta$ is equivalent to

$$\alpha_0\delta\sum_{i=1}^{d} a_i = \sum_{i=1}^{d}\alpha_i a_i,$$

which holds if and only if $\alpha_0\delta = \alpha_i$ for all $i = 1, \dots, d$. A viable choice satisfying $\sum_{i=0}^{d} \alpha_i = 1$ and $\alpha_i > 0$ is $\alpha_0 = (1 + d\delta)^{-1}$ and $\alpha_i = \delta(1 + d\delta)^{-1}$ for all $i = 1, \dots, d$, and thus $a_0 \in T_\delta$.

Proposition A.2. *Given a simplex $T = \operatorname{conv}(\{a_0, \dots, a_d\})$ and a point $p \in T$, let*

$$T_i := \operatorname{conv}(\{p, a_0, \dots, a_d\} \setminus \{a_i\}) \quad \text{for all } i \in \{0, \dots, d\}.$$

Then

$$\overline{\bigcup_{i\in\{0,\dots,d\}} T_i} = \overline{T} \tag{A.2}$$

and this is a patch, and $\{T_i : i \in \{0, \dots, d\}\}$ is a regular simplicial partition of T.

Proof. Let $p \in T$, i.e.

$$p = \sum_{i=0}^{d} \alpha_i a_i, \quad \alpha_i > 0 \text{ and } \sum_{i=0}^{d} \alpha_i = 1.$$

First we show that T_0 is a simplex, which by symmetry implies that T_i is a simplex for all $i \in \{0, \dots, d\}$. It suffices to check that $p - a_1 \notin \operatorname{span}\{a_2 - a_1, \dots, a_d - a_1\}$, since then $\{p - a_1, a_2 - a_1, \dots, a_d - a_1\}$ is a set of linearly independent vectors. This is true since $\{a_0 - a_1, a_2 - a_1, \dots, a_d - a_1\}$ are linearly independent vectors, $\alpha_1 = 1 - \sum_{i\neq 1} \alpha_i$ and thus

$$p - a_1 = \sum_{i\neq 1} \alpha_i (a_i - a_1),$$

with $\alpha_0 > 0$ does not belong to $\operatorname{span}\{a_2 - a_1, \dots, a_d - a_1\}$. Furthermore, $\bigcup_{i\in\{0,\dots,d\}} \overline{T_i} \subset \overline{T}$ follows by the fact that $p \in T$ implies $T_i \subset T$ for all $i \in \{0, \dots, d\}$.

Next we show $\bigcup_{i\in\{0,\dots,d\}} \overline{T_i} \supset \overline{T}$. Fix $\bar{p} := \sum_{j=0}^{d} \gamma_j a_j$ with $\gamma_j \geq 0$ satisfying $\sum_{j=0}^{d} \gamma_j = 1$, i.e. $\bar{p}$ is an arbitrary point in $\overline{T}$. We wish to show that $\bar{p} \in \overline{T_i}$ for some $i \in \{0, \dots, d\}$, i.e.

$$\sum_{j=0}^{d} \gamma_j a_j = \sum_{j\neq i}^{d} \beta_j a_j + \beta_i \sum_{j=0}^{d} \alpha_j a_j \quad \text{for some} \quad \beta_j \geq 0, \quad \sum_{j=0}^{d} \beta_j = 1.$$

This is equivalent to

$$\sum_{j=0}^{d} (\gamma_j - \beta_i \alpha_j) a_j = \sum_{j\neq i}^{d} \beta_j a_j. \tag{A.3}$$

We now show that there exist $(\beta_j)_{j=0}^{d}$ for which this holds. Let

$$i \in \operatorname*{argmin}_{j\in\{0,\dots,d\}} \frac{\gamma_j}{\alpha_j}, \tag{A.4}$$

which is well-defined because $\gamma_j \geq 0$ and $\alpha_j > 0$ for all $j \in \{0, \dots, d\}$. Since $\sum_{j=0}^{d} \alpha_j = 1 = \sum_{j=0}^{d} \gamma_j$, for i in (A.4) it must hold $\frac{\gamma_i}{\alpha_i} \leq 1$. Eq. (A.3) holds if $\gamma_i - \beta_i\alpha_i = 0$ and $\gamma_j - \beta_i\alpha_j = \beta_j$ for all $j \neq i$. The former is satisfied for $\beta_i = \frac{\gamma_i}{\alpha_i} \in [0, 1]$ and the latter is satisfied if $\beta_j = \gamma_j - \beta_i\alpha_j$, which implies $\beta_j = \alpha_j(\frac{\gamma_j}{\alpha_j} - \frac{\gamma_i}{\alpha_i}) \geq 0$ and $\beta_j = \alpha_j(\frac{\gamma_j}{\alpha_j} - \frac{\gamma_i}{\alpha_i}) \leq \gamma_j \leq 1$. It is left to show that $\sum_{j=0}^{d} \beta_j = 1$. We have

$$\begin{aligned}\sum_{j\neq i}^{d} \beta_j + \beta_i &= \sum_{j\neq i}^{d} \alpha_j(\tfrac{\gamma_j}{\alpha_j} - \tfrac{\gamma_i}{\alpha_i}) + \tfrac{\gamma_i}{\alpha_i} \\ &= \sum_{j\neq i}^{d} \gamma_j + \tfrac{\gamma_i}{\alpha_i}\left(1 - \sum_{j\neq i}^{d} \alpha_j\right) = \sum_{j\neq i}^{d} \gamma_j + \tfrac{\gamma_i}{\alpha_i}\alpha_i = 1.\end{aligned}$$

We found $i \in \{0, \dots, d\}$ and $(\beta_j)_{j=0}^{d}$ for which (A.3) holds. Thus $\bar{p} \in \overline{T_i}$, and $\bigcup_{i\in\{0,\dots,d\}} \overline{T_i} \supset \overline{T}$.

It is left to show that for all $m \neq n \in \{0, \dots, d\}$ the intersection of $\overline{T_m}$ and $\overline{T_n}$ is the closure of a sub-simplex of both. Consider

$$\overline{T_m} = \left\{ \sum_{j\neq m}^{d} \beta_j a_j + \beta_m p \ : \ \sum_{j=0}^{d} \beta_j = 1 \text{ and } \beta_j \geq 0 \right\},$$

$$\overline{T_n} = \left\{ \sum_{j\neq n}^{d} \beta_j a_j + \beta_n p \ : \ \sum_{j=0}^{d} \beta_j = 1 \text{ and } \beta_j \geq 0 \right\}.$$

Then

$$\overline{T_m} \cap \overline{T_n} = \left\{ \sum_{j\neq m,n}^{d} \beta_j a_j + \beta p \ : \ \beta + \sum_{j\neq m,n}^{d} \beta_j = 1 \text{ and } \beta, \beta_j \geq 0 \right\},$$

which, by definition, is the closure of a sub-simplex of both $\overline{T_m}$ and $\overline{T_n}$.

For a set $S \subset \mathbb{R}^d$, in the following we call $x \in S$ a *star point* of S iff for all $y \in S \setminus \{x\}$ holds $\operatorname{conv}(\{x, y\}) \subset S$.

Lemma A.3. *Let $p \in \mathcal{V} \cap \operatorname{int}\Omega$ and let $T_1, \dots, T_{s(p)} \in \mathcal{T}$ be the simplices adjacent to p. For each $j = 1, \dots, s(p)$, we denote by P_{T_j} the hyperplane passing through all vertices of T_j except p. Then, any $x \in \operatorname{int}\omega(p)$ that is on the same side of the hyperplane P_{T_j} as p for all $j = 1, \dots, s(p)$ is a star point for the patch $\omega(p)$.*

Proof. We divide the proof of the claim in three steps:

Step 1. We claim that $\partial\omega(p) \subset \bigcup_{j=1}^{s(p)} P_{T_j}$ for all $p \in \mathcal{V} \cap \operatorname{int}\Omega$. To prove this, we define a regular partition $\mathcal{T}'$ of $\mathbb{R}^d$ that extends $\mathcal{T}$, i.e. such that $\mathcal{T} \subset \mathcal{T}'$. For every point z on $\partial\omega(p)$, z is on the boundary of an element $T \subset \omega(p)$ and of an element $T' \subset \mathbb{R}^d \setminus \omega(p)$, $T, T' \in \mathcal{T}'$. By regularity of $\mathcal{T}'$, $\overline{T} \cap \overline{T'}$ is the closure of a subsimplex f of both $\overline{T}$ and $\overline{T'}$. Because T is a simplex with p as one of its vertices, if z is not in P_T, then f touches p, which implies that $T' \subset \omega(p)$ and gives a contradiction.

Step 2. We show that any star point x of $\operatorname{int}\omega(p)$ is a star point of $\omega(p)$. Given $q \in \omega(p)$, define a sequence $\{q_n\} \subseteq \operatorname{int}\omega(p)$ such that $q_n \to q$, then for all $t \in [0, 1]$ we obtain $\operatorname{int}\omega(p) \ni xt + q_n(1 - t) \to xt + q(1 - t) \in \omega(p)$, as $\omega(p)$ is closed.

Step 3. Assume that $x \in \operatorname{int}\omega(p)$ is not a star point of $\omega(p)$. By Step 2, there exists $q \in \operatorname{int}\omega(p)$ and $t \in [0, 1]$ such that $tx + (1 - t)q \notin \operatorname{int}\omega(p)$. Therefore, there exist $\underline{t}, \bar{t} \in (0, 1)$ satisfying $\underline{t} < \bar{t}$, and $T \in \mathcal{T}$, $T \subset \omega(p)$, such that, using Step 1, $\bar{t}x + (1 - \bar{t})q \in P_T$ and $sx + (1 - s)q \in T$, $\forall s \in (\underline{t}, \bar{t})$. Since p and T lie on the same side of P_T, $sx + (1 - s)q$ lies on the other side of P_T than p for all $s \in (\bar{t}, 1]$. For $s = 1$, this implies that x is on the other side of the hyperplane P_T than p. Thus if $x \in \operatorname{int}\omega(p)$ is not a star point, it lies on the other side of at least one hyperplane P_{T_j} than p. Therefore, if a point $x \in \operatorname{int}\omega(p)$ is on the same side of P_{T_j} as p for all $j = 1, \dots, s(p)$, then x is a star point for $\omega(p)$, by contradiction.

Remark A.4. The converse implication of Lemma A.3 holds as well: the set of all star points of the patch $\omega(p)$ coincides with the intersection $\bigcap_{j=1}^{s(p)} \overline{H_j}$ of all closed half-spaces $\overline{H_j}$, where $H_j \subset \mathbb{R}^d$ is defined as the set of all points that lie on the same side of P_{T_j} as p.

In the following, denote by $B_\epsilon(p) \subset \mathbb{R}^d$ the ball of radius $\epsilon > 0$ centered at $p \in \mathbb{R}^d$, with respect to the Euclidean norm.

Lemma A.5. *For all $p \in \mathcal{V} \cap \operatorname{int}\Omega$, there exists $\epsilon > 0$ such that $B_\epsilon(p) \subset \omega(p)$ and such that every $x \in B_\epsilon(p)$ is a star point of $\omega(p)$.*

Proof. For $j = 1, \dots, s(p)$ denote by $H_j \subset \mathbb{R}^d$ the open half-space containing all points that lie on the same side of P_{T_j} as p. Then $\bigcap_{j=1}^{s(p)} H_j$ is open, contains p and is a subset of the set of all star-points of $\omega(p)$ by Lemma A.3.

For interior vertices we obtain the following result. Recall the definitions of $\tilde{T}_{ij}$ and $\tilde{\omega}_j(p)$ given in (4.3) and (4.4), respectively.

Lemma A.6. *Given $p \in \mathcal{V} \cap \operatorname{int}\Omega$, let $T_1, \dots, T_{s(p)} \in \mathcal{T}$ be the simplices adjacent to p. For all $j = 1, \dots, s(p)$, let $a_0 := p$ and $a_1, \dots, a_d \in \mathbb{R}^d$ be such that $T_j = \operatorname{conv}(\{a_0, \dots, a_d\})$ and let $q_j := p + \delta_j \sum_{i=1}^d (p - a_i)$ for some sufficiently small $\delta_j > 0$. Then $\tilde{\omega}_j(p)$ is convex and $\overline{T_j} \subset \tilde{\omega}_j(p) \subset \omega(p)$. The sets $\{\tilde{T}_{ij}\}_{i=0,\dots,d}$ form a regular partition of $\tilde{\omega}_j(p)$.*

Proof. For $\epsilon > 0$ as in Lemma A.5, let $\delta_j > 0$ in the definition of q_j be such that $\|q_j - p\|_2 = \frac{\epsilon}{2}$. Then $q_j \in B_\epsilon(p) \subset \omega(p)$ is a star point of $\omega(p)$ by Lemma A.5. Therefore, $\tilde{T}_{ij} \subset \omega(p)$ for all $i = 1, \dots, d$, and thus $\tilde{\omega}_j(p) \subset \omega(p)$. It also holds that $\overline{T_j} = \overline{\tilde{T}_{0j}} \subset \tilde{\omega}_j(p)$. After observing that $\operatorname{int}\tilde{\omega}_j(p) = (T_j)_{\delta_j}$ in the notation of Lemma A.1, Lemma A.1 shows that $\tilde{\omega}_j(p)$ is a simplex and thus convex. The sets $\{\tilde{T}_{ij}\}_{i=0,\dots,d}$ form a regular partition of $\tilde{\omega}_j(p)$ by Proposition A.2 ($q_j, a_1, \dots, a_d, p$ in the notation of this proof correspond to $a_0, \dots, a_d, p$ in the notation of the lemma).

Corollary A.7. *Let $\tilde{\omega}_j(p)$ be as in Lemma A.6, then*

$$\omega(p) = \bigcup_{j=1}^{s(p)} \tilde{\omega}_j(p).$$

Proof. Using $\overline{T_j} \subset \tilde{\omega}_j(p) \subset \omega(p)$ for all $j = 1, \dots, s(p)$ we have

$$\omega(p) = \bigcup_{j=1}^{s(p)} \overline{T_j} \subset \bigcup_{j=1}^{s(p)} \tilde{\omega}_j(p) \subset \omega(p).$$

Corollary A.8. *In the notation of Lemma A.6, for $p \in \mathcal{V} \cap \operatorname{int}\Omega$ and $j = 1, \dots, s(p)$, let $\tilde{\theta}_{p,j}^{S_1^1} \in C^0(\Omega)$ be the hat function on $\tilde{\omega}_j(p)$ defined in (4.5). For all $p \in \mathcal{V} \cap \operatorname{int}\Omega$ and $j = 1, \dots, s(p)$, these functions can be written as*

$$\tilde{\theta}_{p,j}^{S_1^1}(x) = \max\left\{0, \min_{i=0,\dots,d} \tilde{A}_p^{(i,j)} x + \tilde{b}_p^{(i,j)}\right\}, \quad x \in \Omega,$$

where each $x \mapsto \tilde{A}_p^{(i,j)} x + \tilde{b}_p^{(i,j)}$ is a globally linear function fulfilling $(\tilde{A}_p^{(i,j)} x + \tilde{b}_p^{(i,j)})|_{\tilde{T}_{ij}} = \tilde{\theta}_{p,j}^{S_1^1}|_{\tilde{T}_{ij}}$.

Proof. Since $\tilde{\omega}_j(p)$ is a convex patch, the statement follows by Proposition 4.1 (which corresponds to He et al. (2020, Theorem 3.1)). The function $x \mapsto \tilde{A}_p^{(i,j)} x + \tilde{b}_p^{(i,j)}$ in our notation corresponds to g_k in the notation of He et al. (2020), and $\tilde{\theta}_{p,j}^{S_1^1}$ corresponds to ϕ_i.

Lemma A.9. *For all $p \in \mathcal{V}$, $j, k = 1, \dots, s(p)$ and $i = 0, \dots, d$, let $x \mapsto \tilde{A}_p^{(i,j)} x + \tilde{b}_p^{(i,j)}$ be as defined in Corollary A.8 and let $x \mapsto A_p^{(k)} x + b_p^{(k)}$ be the function defined by $(A_p^{(k)} x + b_p^{(k)})|_{T_k} = \theta_p^{S_1^1}|_{T_k}$. Then,*

$$0 \le \tilde{A}_p^{(i,j)} x + \tilde{b}_p^{(i,j)} \le A_p^{(k)} x + b_p^{(k)}, \quad \forall x \in T_k \cap \tilde{T}_{ij},$$

for all $j, k = 1, \dots, s(p)$ and $i = 0, \dots, d$.

Proof. First consider $p \in \mathcal{V} \cap \operatorname{int}\Omega$, such that $p \in \operatorname{int}\omega(p)$. For all $j, k \in \{1, \dots, s(p)\}$ and $i \in \{0, \dots, d\}$, we first note that $\tilde{A}_p^{(i,j)} p + \tilde{b}_p^{(i,j)} = A_p^{(k)} p + b_p^{(k)} = 1$ as well as $\tilde{A}_p^{(i,j)} y + \tilde{b}_p^{(i,j)} = 0$ for all y on the face of $\tilde{T}_{ij}$ opposite to p. Similarly, $A_p^{(k)} y + b_p^{(k)} = 0$ for all y on the face of T_k opposite to p. Next, let $x \in T_k \cap \tilde{T}_{ij}$ in case this set is nonempty. Let L be the halfline starting in p through x. Note that $p \in \operatorname{int}\tilde{\omega}_j(p) \subset \operatorname{int}\omega(p)$ is a star point of both $\tilde{\omega}_j(p)$ and $\omega(p)$. It follows from $\tilde{\omega}_j(p) \subset \omega(p)$ that the intersection point $y_1 \in L \cap \partial\tilde{\omega}_j(p)$ is closer to p than (or equal to) the intersection point $y_2 \in L \cap \partial\omega(p)$. Because $y \mapsto \tilde{A}_p^{(i,j)} y + \tilde{b}_p^{(i,j)}$ linearly interpolates between the value 1 in p and 0 in y_1, and $y \mapsto A_p^{(k)} y + b_p^{(k)}$ linearly interpolates between 1 in p and 0 in y_2, it follows that $\tilde{A}_p^{(i,j)} y + \tilde{b}_p^{(i,j)} \le A_p^{(k)} y + b_p^{(k)}$ for all points y between p and y_1, which includes x. Finally, the first inequality in the lemma follows from Corollary A.8.

For $p \in \mathcal{V} \cap \partial\Omega$, we can apply the argument above after extending $\mathcal{T}$ to a regular, simplicial partition of all of $\mathbb{R}^d$, of which only the elements touching p are relevant.

Proof of Theorem 4.2. For all $j = 1, \dots, s(p)$, applying Lemma A.9 for all $i = 0, \dots, d$ and all $k = 1, \dots, s(p)$ shows that $0 \le \tilde{\theta}_{p,j}^{S_1^1}(x) \le \theta_p^{S_1^1}(x)$ for all $x \in \tilde{\omega}_j(p)$. Together with $\tilde{\theta}_{p,j}^{S_1^1}(x) = 0$ for all $x \in \Omega \setminus \tilde{\omega}_j(p)$, this shows that $0 \le \tilde{\theta}_{p,j}^{S_1^1}(x) \le \theta_p^{S_1^1}(x)$ for all $x \in \Omega$. To finish the proof, recall that for all $j = 1, \dots, s(p)$ and $x \in T_j$

$$\theta_p^{S_1^1}(x) = A_p^{(j)} x + b_p^{(j)} = \tilde{A}_p^{(0,j)} x + \tilde{b}_p^{(0,j)} = \tilde{\theta}_{p,j}^{S_1^1}(x).$$

The first and the last equality hold by definition, and the second holds because both functions are linear and equal the value 1 in p and 0 in the other vertices of T_j.

Proof of Theorem 4.3. Because $\tilde{\omega}_j(p) = \cup_{i=0}^d \overline{\tilde{T}_{ij}}$ is a regular partition of the convex set $\tilde{\omega}_j(p)$ by Lemma A.6, we can apply Proposition 4.1 and it follows that for all $j = 1, \dots, s(p)$

$$L(\tilde{\Phi}_{p,j}^{CPwL}) \le 5 + \log_2(d+1), \quad M(\tilde{\Phi}_{p,j}^{CPwL}) \le Cd(d+1) \le Cd^2$$

and that all NNs $\{\tilde{\Phi}_{p,j}^{CPwL}\}_{j=1}^{s(p)}$ have equal depth, see Proposition 4.1.

The fact that $\mathrm{R}(\Phi_p^{CPwL})(x) = \theta_p^{S_1^1}(x)$ for all $x \in \Omega$ follows from Theorem 4.2, and the network depth and size are bounded as follows:

$$\begin{aligned} L(\Phi_p^{CPwL}) &= L(\Phi_{s(p)}^{\max}) + L(\tilde{\Phi}_{p,1}^{CPwL}) \\ &\le 2 + \log_2(s(p)) + 5 + \log_2(d+1), \\ M(\Phi_p^{CPwL}) &\le 2M(\Phi_{s(p)}^{\max}) + 2\sum_{j=1}^{s(p)} M(\tilde{\Phi}_{p,j}^{CPwL}) \\ &\le Cs(p) + s(p)Cd^2 \le Cd^2 s(p). \end{aligned}$$

References

Alonso, A., & Valli, A. (1999). An optimal domain decomposition preconditioner for low-frequency time-harmonic Maxwell equations. *Mathematics of Computation*, *68*(226), 607–631.

Amrouche, C., Bernardi, C., Dauge, M., & Girault, V. (1998). Vector potentials in three-dimensional non-smooth domains. *Mathematical Methods in the Applied Sciences*, *21*(9), 823–864.

Arnold, D. N., Falk, R. S., & Winther, R. (2006). Differential complexes and stability of finite element methods. II. The elasticity complex. In *IMA vol. math. appl.: vol. 142, Compatible spatial discretizations* (pp. 47–67). New York: Springer.

Arora, R., Basu, A., Mianjy, P., & Mukherjee, A. (2018). Understanding deep neural networks with rectified linear units. In *International conference on learning representations*. arXiv:1611.01491.

Aylwin, R., Henríquez, F., & Schwab, C. (2023). ReLU neural network Galerkin BEM. *Journal of Scientific Computing*, *95*(2), 41.

Balci, A. K., Ortner, C., & Storn, J. (2022). Crouzeix-Raviart finite element method for non-autonomous variational problems with Lavrentiev gap. *Numerische Mathematik*, *151*(4), 779–805.

Ball, J. M. (2001). Singularities and computation of minimizers for variational problems. In *London math. soc. lecture note ser.: vol. 284, Foundations of computational mathematics (Oxford, 1999)* (pp. 1–20). Cambridge: Cambridge Univ. Press.

Buffa, A., Dölz, J., Kurz, S., Schöps, S., Vázquez, R., & Wolf, F. (2020). Multipatch approximation of the de Rham sequence and its traces in isogeometric analysis. *Numerische Mathematik*, *144*(1), 201–236.

Buffa, A., Hiptmair, R., von Petersdorff, T., & Schwab, C. (2003). Boundary element methods for Maxwell transmission problems in Lipschitz domains. *Numerische Mathematik*, *95*(3), 459–485.

Chambolle, A., & Pock, T. (2020). Crouzeix-Raviart approximation of the total variation on simplicial meshes. *Journal of Mathematical Imaging and Vision*, *62*(6–7), 872–899.

Cicuttin, M., Ern, A., & Pignet, N. (2021). *Springerbriefs in mathematics ser.*, *Hybrid high-order methods : A primer with applications to solid mechanics*. Cham: Springer International Publishing AG.

Costabel, M. (1991). A coercive bilinear form for Maxwell's equations. *Journal of Mathematical Analysis and Applications*, *157*(2), 527–541.

Costabel, M., & Dauge, M. (1999). Maxwell and Lamé eigenvalues on polyhedra. *Mathematical Methods in the Applied Sciences*, *22*(3), 243–258.

Costabel, M., Dauge, M., & Nicaise, S. (1999). Singularities of Maxwell interface problems. *M2AN Mathematical Modelling and Numerical Analysis*, *33*(3), 627–649.

Crouzeix, M., & Falk, R. S. (1989). Nonconforming finite elements for the Stokes problem. *Mathematics of Computation*, *52*(186), 437–456.

Di Pietro, D. A., Droniou, J., & Pitassi, S. (2022). Cohomology of the discrete de Rham complex on domains of general topology. arXiv:2209.00957.

E, W., & Yu, B. (2018). The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, *6*(1), 1–12.

Ern, A., & Guermond, J.-L. (2017). Finite element quasi-interpolation and best approximation. *ESAIM Mathematical Modelling and Numerical Analysis*, *51*(4), 1367–1385.

Ern, A., & Guermond, J.-L. (2021). Finite elements I—Approximation and interpolation. *Texts in applied mathematics: vol. 72*, (p. xii+325). Cham: Springer, [2021] ©.

Fuentes, F., Keith, B., Demkowicz, L., & Nagaraj, S. (2015). Orientation embedded high order shape functions for the exact sequence elements of all shapes. *Computers & Mathematics with Applications*, *70*(4), 353–458.

He, J., Li, L., Xu, J., & Zheng, C. (2020). ReLU deep neural networks and linear finite elements. *Journal of Computational Mathematics*, *38*.

Li, B., Tang, S., & Yu, H. (2020). PowerNet: Efficient representations of polynomials and smooth functions by deep neural networks with rectified power units. *Journal of Mathematical Study*, *53*(2), 159–191.

Liang, S., & Srikant, R. (2017). Why deep neural networks for function approximation? In *Proc. of ICLR 2017* (pp. 1–17).

Marcellini, P. (1989). Regularity of minimizers of integrals of the calculus of variations with nonstandard growth conditions. *Archive for Rational Mechanics and Analysis*, *105*(3), 267–284.

Mhaskar, H. N., & Poggio, T. (2016). Deep vs. shallow networks: An approximation theory perspective. *Analysis and Applications*, *14*(6), 829–848.

Opschoor, J. A. A., Schwab, C., & Zech, J. (2022). Exponential ReLU DNN expression of holomorphic maps in high dimension. *Constructive Approximation*, *55*(1), 537–582.

Petersen, P., & Voigtlaender, F. (2018). Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Networks*, *108*, 296–330.

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707.

Sauter, S. A., & Schwab, C. (2011). *Springer series in computational mathematics: vol. 39*, *Boundary element methods* (p. xviii+561). Berlin: Springer-Verlag, Translated and expanded from the 2004 German original.

Schwab, C., & Zech, J. (2019). Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ. *Analysis and Applications, Singapore*, *17*(1), 19–55.

Trask, N., Huang, A., & Hu, X. (2022). Enforcing exact physics in scientific machine learning: A data-driven exterior calculus on graphs. *Journal of Computational Physics*, *456*, Article 110969.

Yang, L., Meng, X., & Karniadakis, G. E. (2021). B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics*, *425*, Article 109913.

Yarotsky, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Networks*, *94*, 103–114.

Zhikov, V. V. (1983). Questions of convergence, duality and averaging for functionals of the calculus of variations. *Izv. Akad. Nauk SSSR Ser. Mat.*, *47*(5), 961–998.