

Enforcing exact physics in scientific machine learning: A data-driven exterior calculus on graphs

Nathaniel Trask^{a,*}, Andy Huang^b, Xiaozhe Hu^c

^a Center for Computing Research, Sandia National Laboratories, United States of America

^b Radiation and Electrical Science, Sandia National Laboratories, United States of America

^c Department of Mathematics, Tufts University, United States of America

ARTICLE INFO

Article history:

Received 28 January 2021

Received in revised form 8 January 2022

Accepted 9 January 2022

Available online 20 January 2022

Keywords:

Scientific machine learning

Data driven modeling

Discrete exterior calculus

de Rham complex

PDE-constrained optimization

Graph neural networks

ABSTRACT

As traditional machine learning tools are increasingly applied to science and engineering applications, physics-informed methods have emerged as effective tools for endowing inferences with properties essential for physical realizability. While promising, these methods generally enforce physics weakly via penalization. To enforce physics strongly, we turn to the exterior calculus framework underpinning combinatorial Hodge theory and physics-compatible discretization of partial differential equations (PDEs). Historically, these two fields have remained largely distinct, as graphs are strictly topological objects lacking the metric information fundamental to PDE discretization. We present an approach where this missing metric information may be learned from data, using graphs as coarse-grained mesh surrogates that inherit desirable conservation and exact sequence structure from the combinatorial Hodge theory. The resulting data-driven exterior calculus (DDEC) may be used to extract structure-preserving surrogate models with mathematical guarantees of well-posedness. The approach admits a PDE-constrained optimization training strategy which guarantees machine-learned models enforce physics to machine precision, even for poorly trained models or small data regimes. We provide analysis of the method for a class of models designed to reproduce nonlinear perturbations of elliptic problems and provide examples of learning $H(\text{div})/H(\text{curl})$ systems representative of subsurface flows and electromagnetics.

Published by Elsevier Inc.

1. Introduction

Network models of physical systems are ubiquitous throughout the sciences and engineering. The electronic circuit models typically learned in first year undergraduate physics may similarly be used to describe fluid, mechanical or heat transfer systems in corresponding hydraulic circuit, mass-spring-damper, or thermal circuit analogies, respectively [68,69,29,42,13,61,22,52]. Such network models represent discrete representations of conservation laws corresponding to classical control volume analysis and their physical properties are intricately tied to the underlying graph topology [60,7], allowing modeling of a number of thermodynamic principles [19]. In contemporary machine learning (ML), graph neural networks (GNNs) attach a more “black-box” message passing model to a network, inferring “object-relation”/causal relationships from unstructured data to obtain data representations and model pair-wise interactions [33–35,10,21]. Generally

* Corresponding author.

E-mail address: natrask@sandia.gov (N. Trask).

however, applications of GNNs have not focused upon preserving physical structures in their network topology, although some recent works have considered how the interplay between microstructure and mechanics can be encoded via graphs [64].

In this work we present a framework to extract efficient data-driven network models which exactly preserve desirable mathematical structures of the underlying physics. The framework we introduce is general, but we focus on a particular application in which we assume access to high-fidelity data and aim to learn a low dimensional network as an efficient structure-preserving surrogate model. The resulting model bears similarities to reduced-order models (ROMs), however the control volume analysis pursued in the current approach may provide advantages in preserving structural properties that have proven challenging for the variational derivation underpinning ROMs [54,53].

Solutions of PDEs generally rely upon computational meshes partitioning space into disjoint curvilinear cells. The shared topological structure offered by both networks and meshes allow a unified analysis in terms of the exterior calculus. For PDEs, the discrete exterior calculus (DEC) and related spatially compatible discretizations encapsulate a range of so-called mimetic discretizations providing discrete solutions mimicking physical properties of the continuous problem [27,4,3,16]. These methods generally preserve conservation properties and spectral representations of operators, provide a coordinate-free means of prescribing physics on manifolds, and allow handling of the non-trivial null-spaces required in electromagnetics. In topological data analysis, combinatorial Hodge theory has emerged as a tool for analyzing flows on graphs along with their spectral and homological properties, e.g., [47,39,46,8,11,45,59,63]. These techniques are supported by a graph calculus providing generalizations of gradient, curl, and divergence operators admitting interpretation as discrete exterior derivatives. As graphs are purely topological however, this graph calculus lacks any metric information, and the operators do not converge in a meaningful way to the familiar vector calculus necessary to model physical systems.

The key observation of the current work is that one may exploit the shared structure of graphs and meshes and use data to endow the graph calculus with “missing” metric information, in the process learning a data-driven exterior calculus (DDEC). Conservation laws are encoded onto the graph via trainable div/curl operators satisfying a Stokes theorem, while “black-box” DNNs may parameterize fluxes. This guarantees that the mathematical structures related to conservation and exact sequence properties are independent of the errors incurred during training. Consequently, this allows us to prove a number of properties of the resulting model, independent of the quantity of available data or particular local minimizer found during training.

The incorporation of physical principles into scientific machine learning tasks has surged in recent years, as it has become apparent that “off-the-shelf” ML tools often fail to provide robust predictions for science and engineering applications [6]. Physics-informed approaches, and in particular physics-informed neural networks (PINNs), have achieved a range of successes by introducing regularizers that penalize deviations from physical properties [56,55,44]. While simple to implement and effective for a range of problems, physics in these cases are only enforced to within optimization error and may hold to a relatively coarse tolerance (for open problems regarding their training, see e.g. [65,67]). Some applications however require constraints to hold to machine precision; e.g. in electrodynamics it is critical that the solenoidal constraint hold to machine precision to handle the involution condition [9,26], and in forward modeling contexts compatible/mimetic discretizations provide approximations guaranteeing such properties hold by construction [70,49,17,14]. We therefore propose that the current work constitutes *structure-preserving* machine learning imposing physics strongly, in contrast to *physics-informed* machine learning which imposes physics weakly. Later in the manuscript, we will see that encoding physics via exterior calculus resolves many issues with PINNs [65,67,30]: conservation and exact sequence structure is preserved to machine precision; learning physics as a black-box perturbation of a Hodge-Laplacian allows a simple stability theory and discovery of physics without assuming a priori model form; well-posedness allows rigorous application of equality constrained optimizers, eliminating problematic calibration of penalty parameters; for problems with non-trivial null-spaces a discrete Hodge decomposition allows rigorous imposition of gauge conditions; non-oscillatory solutions are obtained in the presence of material discontinuities; hyperparameters are eliminated from training, apart from learning rate; and for linear elliptic problems, data is reproduced to machine precision. While the simplicity of PINNs makes them a powerful tool for inverse problems, particularly in combination with operator regression [20,66], problems such as electromagnetism and magnetohydrodynamics mandate a structure-preserving approach. The current work therefore targets a separate class of problems inaccessible to existing physics-informed approaches.

1.1. Paper organization

We first recall necessary exterior calculus fundamentals before introducing our data-driven exterior calculus. While the theory is abstract, we focus on applying it to learn nonlinear perturbations of $\text{div} - \text{grad}$ and $\text{div} - \text{curl}$ model problems as canonical examples of physics requiring structure preservation. We next provide numerical analysis, establishing conditions under which the learned model has a unique solution. We establish that the data-driven exterior calculus inherits the desirable properties of the graph calculus and use them to analyze the well-posedness of a class of nonlinear elliptic problems. A necessary implementation question is how to obtain a graph to define the model upon. We provide a specific example considering coarse-graining of a high-fidelity mesh, associated with either a finite element simulation or histograms binning experimental data, and show how the relevant commuting diagrams are preserved under coarsening. Next, we introduce a PDE-constrained optimizer to fit the model to data, allowing enforcement of physics exactly via equality constraint. The numerical analysis implies that the forward problem associated with the equality constraint is always well-posed given mild

conditions on the architecture. Finally, we provide several numerical results which validate the analysis for canonical elliptic problems involving scalar and magnetic diffusion which mandate compatibility, demonstrating how one may learn efficient physics-preserving surrogates from high-fidelity data.

2. The graph exterior calculus

We recall first the graph calculus, which serves as the foundation for DDEC. Let $\mathbf{N} = \{n_i\}_{i=1}^{N_N}$ denote a set of nodes. We embed $\mathbf{N}$ in $\mathbb{R}^d$ by associating with each node a unique position $\mathbf{p}_i \in \mathbb{R}^d$, $i \in 1, \dots, N_N$. We define a k -clique as an ordered tuple consisting of k nodes, i.e. $t_k = [n_1, \dots, n_k]$. A k -clique has positive orientation if $\pi = \{i_1, \dots, i_k\}$ is an even permutation of $\{1, \dots, k\}$ and negative otherwise. Via the embedding of the graph, we may associate with each k -clique the $(k-1)$ -simplex defined as the convex hull of the vertices $s_k = [\mathbf{p}_1, \dots, \mathbf{p}_k]$, for which the $k \leq d$ distinct points span a k -dimensional hyperplane. A k -chain c_k may then be defined as a linear combination of $(k+1)$ -cliques, and we denote the set of k -chains by C_k . One may introduce a *boundary operator* $\partial_k : C_{k+1} \rightarrow C_k$ defined via

$$\partial_k[n_1, \dots, n_k] = \sum_i^k (-1)^{i-1} [n_1, \dots, n_{i-1}, \hat{n}_i, n_{i+1}, n_k], \quad (1)$$

where $\hat{\cdot}$ denotes an omitted entry, and which satisfies the property $\partial_{k-1} \circ \partial_k = 0$. When the dependence upon k is clear, we will sometimes write the coboundary simply as ∂ - we will adopt this convention for similar operators throughout. With these definitions in hand we may finally introduce the chain complex as the following exact sequence pairing k -chains and boundary operators.

$$C_0 \xleftarrow{\partial_0} C_1 \xleftarrow{\partial_1} \dots \xleftarrow{\partial_{d-1}} C_d \quad (2)$$

with the standard convention that ∂_{-1} maps C_0 to the empty set.

Note that in this graph context, the specific realization of the chain complex may be qualitatively different from the DEC setting. Traditionally in compatible discretizations, the complex is realized by partitioning the domain of interest into a collection of disjoint simplices to obtain C_d , and then deriving lower dimensional mesh entities C_k , $k < d$ via the boundary operator. In contrast, for the graph setting one may obtain overlapping simplices, in the sense that given unique $t_k, t'_k \in C_1$, $t_k \cap t'_k \neq \emptyset$. We will later discuss details regarding specific choice of chain complex, but for now keep the presentation abstract.

We next associate real numbers with the graph entities constituting the chain complex. For each set of chains C_k , we introduce the dual set of cochains C^k consisting of linear functionals acting on C_k . Given $\phi \in C^k$, we denote the value associated with the k -chain $t_{i_1 i_2 \dots i_k}$ via the shorthand $\phi_{i_1 i_2 \dots i_k} := \phi(t_{i_1 i_2 \dots i_k})$. Note that cochains inherit the orientation of the underlying chains, e.g. $\phi_{ij} = -\phi_{ji}$ via the definition of π . Introducing the *coboundary operator* $\delta_k : C^k \rightarrow C^{k+1}$, we next arrive at the following cochain complex

$$C^0 \xrightarrow{\delta_0} C^1 \xrightarrow{\delta_1} \dots \xrightarrow{\delta_{d-1}} C^d \quad (3)$$

We will formally denote the pairing between boundary and coboundary operators via the inner product

$$\langle \phi, \partial_k t \rangle = \langle \delta_k \phi, t \rangle. \quad (4)$$

In the traditional DEC setting, one would arrive at a definition of the coboundary via the generalized Stokes theorem, defined by the dual pairing of δ_k and ∂_k via $\int_{\partial\omega} w = \int_{\omega} \delta w$. In the graph setting, we identify δ_k algebraically as the adjoint of the matrix representing ∂_k . For example, this gives rise to the following combinatorial gradient, assigning to the 2-clique t_{ij} the function

$$\delta_0 \phi_{ij} = \phi_j - \phi_i. \quad (5)$$

Similarly we may obtain the combinatorial curl by assigning to the 3-clique t_{ijk} the function

$$\delta_1 \phi_{ijk} = \phi_{ij} + \phi_{jk} + \phi_{ki}. \quad (6)$$

One may easily see that $\delta_1 \circ \delta_0 = 0$. For the purposes of this work, we will not require δ_k for $k > 1$, however the definition extends naturally to $1 < k \leq d$, and one may show that $\delta_k \circ \delta_{k-1} = 0$. In this manner, the coboundary operator inherits the exact sequence property of the boundary operator.

We next let $(\cdot, \cdot)_k$ denote an inner product mapping $C^k \times C^k \rightarrow \mathbb{R}$. This inner product induces a *codifferential operator* $\delta_k^* : C^{k+1} \rightarrow C^k$ via the pairing $(v, \delta_k^* u)_k = (\delta_k v, u)_{k+1}$. In this manner, the careful choice of $(\cdot, \cdot)_k$ will endow the codifferential with desirable approximation properties, however we note that independent of the choice of inner products the

codifferential again inherits the exact sequence properties of the coboundary operator so that $\delta_{k-1}^* \circ \delta_k^* = 0$. This follows trivially from the definition, so that for all $v \in C^{k-2}$ and all $u \in C^k$

$$(v, \delta_{k-1}^* \delta_k^* u)_{k-2} = (\delta_{k-1} v, \delta_k^* u)_{k-1} = (\delta_k \delta_{k-1} v, u)_k = 0. \quad (7)$$

Finally, we will contrast how the choice of inner product typically used in the graph exterior calculus precludes the use of graph boundary/coboundary operators in discretizing PDE. In the graph context, one selects as $(\cdot, \cdot)_k$ the ℓ_2 inner-product:

$$(x, y)_k = (x, y) := \sum_{i=1}^{\dim(C^k)} x_i y_i, \quad (8)$$

and its induced norm is denoted by $\|\cdot\|_k$. In this case, the codifferential δ_k^* may be identified as the transpose of the matrix associated with the coboundary δ_k . For the remainder of this work, we will assume ℓ_2 inner products in the definition of the codifferential unless otherwise specified, so that δ_0^* and δ_1^* correspond to the graph divergence and curl.

Besides the properties that $\delta_k \circ \delta_{k-1} = 0$ and $\delta_{k-1}^* \circ \delta_k^* = 0$, the Hodge-Laplacian is defined as follows

$$L_k = \delta_{k-1} \delta_{k-1}^* + \delta_k^* \delta_k. \quad (9)$$

From the combinatorial Hodge theorem, the combinatorial Laplacian decomposes the space C^k as follows

$$C^k = \text{im}(\delta_{k-1}) \oplus \ker(L_k) \oplus \text{im}(\delta_k^*), \quad (10)$$

which is the Hodge decomposition [28]. One direct consequence of the Hodge decomposition (10) is the following Poincaré inequalities,

$$\begin{aligned} \|z_k\|_k &\leq c \|\delta_k z\|_{k+1}, \quad z_k \in \text{im}(\delta_k^*), \\ \|z_k\|_k &\leq c^* \|\delta_{k-1}^* z_k\|_{k-1}, \quad z_k \in \text{im}(\delta_{k-1}). \end{aligned}$$

Here c and c^* are Poincaré constants which depend upon the topological structure of the graph. Another consequence of the Hodge decomposition (10) is that the Hodge Laplacian L_k is positive semidefinite and the dimension of the null space is equal to the dimension of the homology $\ker(\delta_k)/\text{im}(\delta_{k-1})$. This implies the invertibility of the Hodge Laplacian L_k on the orthogonal complement of the null space.

In the next section, we will introduce a data-driven exterior calculus and establish analogues of these properties.

3. The data-driven exterior calculus

We now parameterize these graph calculus operators in a manner which allows recovery of traditional DEC schemes for PDE discretizations as a specific instance. We consider the de Rham complex as a prototypical means of analyzing structure-preserving differential operators in $\mathbb{R}^d$, and seek to develop its discrete approximation on a graph. In general the construction presented here may be used to approximate any exact sequence: we restrict our presentation to the de Rham complex as a single example to support later discretization of PDEs in $\mathbb{R}^d$. In three dimensions, the de Rham complex for $\Omega \in \mathbb{R}^3$ is given by

$$C^\infty(\Omega) \xrightleftharpoons[\text{div}^*]{\text{grad}} [C^\infty(\Omega)]^3 \xrightleftharpoons[\text{curl}^*]{\text{curl}} [C^\infty(\Omega)]^3 \xrightleftharpoons[\text{grad}^*]{\text{div}} C^\infty(\Omega) \quad (11)$$

Here, coboundary operators grad , curl , and div satisfying $\text{curl} \circ \text{grad} = \text{div} \circ \text{curl} = 0$; codifferential operators div^* , curl^* , and grad^* satisfying $\text{curl}^* \circ \text{grad}^* = \text{div}^* \circ \text{curl}^* = 0$. To support later two-dimensional examples, we will also consider the restriction to two dimensions $\Omega \in \mathbb{R}^2$.

$$C^\infty(\Omega) \xrightleftharpoons[\text{curl}^*]{\text{curl}} [C^\infty(\Omega)]^2 \xrightleftharpoons[\text{grad}^*]{\text{div}} C^\infty(\Omega) \quad (12)$$

For completeness, we recall the alternate complex in $\mathbb{R}^2$

$$C^\infty(\Omega) \xrightleftharpoons[\text{div}^*]{\text{grad}} [C^\infty(\Omega)]^2 \xrightleftharpoons[\text{curl}^*]{\text{curl}} C^\infty(\Omega). \quad (13)$$

In this work however, we consider only (12) as the natural complex to obtain conservation properties.

3.1. Data-driven coboundaries and codifferentials

Considering the general case where $\Omega \subset \mathbb{R}^d$, we define new coboundary and codifferential operators by associating learnable metric information with the graph operators as follows,

$$\mathbf{d}_k := \mathbf{B}_{k+1} \delta_k \mathbf{B}_k^{-1}, \quad \text{and} \quad \mathbf{d}_k^* = \mathbf{D}_k^{-1} \delta_k^* \mathbf{D}_{k+1}^*. \quad (14)$$

Here, $\mathbf{B}_k$ and $\mathbf{D}_k$ denote a diagonal tensor with positive entries weighting corresponding elements of C^k . These tensors are understood to encode metric information; in traditional DEC/covolume discretizations of PDEs, the diagonal entries correspond to oriented measures of mesh entities (e.g. lengths, areas, volumes) [50,51]. In this work, we treat these terms as abstract measures which may be learned from data. Positivity of entries may be implemented in several ways; we choose to parameterize each diagonal entry as the square of a trainable parameter.

For example, in $\mathbb{R}^3$, we have

$$GRAD = \mathbf{B}_1 \delta_0 \mathbf{B}_0^{-1} \quad CURL = \mathbf{B}_2 \delta_1 \mathbf{B}_1^{-1} \quad DIV = \mathbf{B}_3 \delta_2 \mathbf{B}_2^{-1} \quad (15)$$

$$DIV^* = \mathbf{D}_0^{-1} \delta_0^* \mathbf{D}_1 \quad CURL^* = \mathbf{D}_1^{-1} \delta_1^* \mathbf{D}_2 \quad GRAD^* = \mathbf{D}_2^{-1} \delta_2^* \mathbf{D}_3 \quad (16)$$

In $\mathbb{R}^2$, we have

$$CURL = \mathbf{B}_1 \delta_0 \mathbf{B}_0^{-1}, \quad DIV = \mathbf{B}_2 \delta_1 \mathbf{B}_1^{-1}, \quad (17)$$

$$CURL^* = \mathbf{D}_0^{-1} \delta_0^* \mathbf{D}_1, \quad GRAD^* = \mathbf{D}_1^{-1} \delta_1^* \mathbf{D}_2. \quad (18)$$

Here, we denote discrete graph operations in capital letters, and use lower case for continuum counterparts. These metric tensors naturally induce the following chain maps,

$$\begin{array}{ccccccc} C^0 & \xrightarrow{\mathbf{d}_0} & C^1 & \xrightarrow{\mathbf{d}_1} & C^2 & \xrightarrow{\mathbf{d}_2} & C^3 & \xrightarrow{\mathbf{d}_3} & \dots & \xrightarrow{\mathbf{d}_{d-1}} & C^d \\ \uparrow \mathbf{B}_0 & & \uparrow \mathbf{B}_1 & & \uparrow \mathbf{B}_2 & & \uparrow \mathbf{B}_3 & & & & \uparrow \mathbf{B}_d \\ C^0 & \xleftarrow{\delta_1^*} & C^1 & \xleftarrow{\delta_2^*} & C^2 & \xleftarrow{\delta_3^*} & C^3 & \xleftarrow{\delta_4^*} & \dots & \xleftarrow{\delta_{d-1}^*} & C^d \\ \uparrow \mathbf{D}_0 & & \uparrow \mathbf{D}_1 & & \uparrow \mathbf{D}_2 & & \uparrow \mathbf{D}_3 & & & & \uparrow \mathbf{D}_d \\ C^0 & \xleftarrow{\mathbf{d}_0^*} & C^1 & \xleftarrow{\mathbf{d}_1^*} & C^2 & \xleftarrow{\mathbf{d}_2^*} & C^3 & \xleftarrow{\mathbf{d}_3^*} & \dots & \xleftarrow{\mathbf{d}_{d-1}^*} & C^d \end{array} \quad (19)$$

Based on the definitions (15) and (16), it is easy to verify that the diagram (19) is commutative, i.e.,

$$\mathbf{B}_{k+1} \delta_k = \mathbf{d}_k \mathbf{B}_k, \quad \text{and} \quad \mathbf{D}_k^{-1} \delta_k^* = \mathbf{d}_k^* \mathbf{D}_{k+1}^{-1}.$$

In $\mathbb{R}^3$, we have,

$$\mathbf{B}_1 \delta_0 = GRAD \mathbf{B}_0, \quad \mathbf{B}_2 \delta_1 = CURL \mathbf{B}_1, \quad \mathbf{B}_3 \delta_2 = DIV \mathbf{B}_2,$$

and

$$\delta_0^* \mathbf{D}_0^{-1} = DIV^* \mathbf{D}_1^{-1}, \quad \mathbf{D}_1^{-1} \delta_1^* = CURL^* \mathbf{D}_2^{-1}, \quad \mathbf{D}_2^{-1} \delta_2^* = GRAD^* \mathbf{D}_3^{-1}.$$

And in $\mathbb{R}^2$, similarly, we have

$$\mathbf{B}_1 \delta_0 = CURL \mathbf{B}_0, \quad \mathbf{B}_2 \delta_1 = DIV \mathbf{B}_1,$$

and

$$\delta_0^* \mathbf{D}_0^{-1} = CURL^* \mathbf{D}_1^{-1}, \quad \mathbf{D}_1^{-1} \delta_1^* = GRAD^* \mathbf{D}_2^{-1}.$$

Theorem 3.1. The discrete derivatives $\mathbf{d}_k$ in (14) form an exact sequence if the simplicial chain complex (C^k, δ_k) is exact, and in particular $\mathbf{d}_{k+1} \circ \mathbf{d}_k = 0$.

Proof. Because (19) is a commutative diagram, the chain maps take cycles to cycles and boundaries to boundaries. Therefore, if the simplicial complex is exact, then (15) forms an exact sequence. Moreover, $\mathbf{d}_{k+1} \circ \mathbf{d}_k = 0$ can be verified by the definitions (14). $\square$

Remark 3.1. In $\mathbb{R}^3$, we have $CURL \circ GRAD = DIV \circ CURL = 0$. And in $\mathbb{R}^2$, we have $DIV \circ CURL = 0$.

Theorem 3.2. The discrete derivatives d_k^* in (14) form an exact sequence if the simplicial cochain complex (C^k, d_k) is exact, and in particular $d_k^* \circ d_{k+1}^* = 0$.

Proof. The conclusion follows from the fact that (19) is commutative and the definitions (14). The proof is essentially the same as the proof of Theorem 3.1 $\square$

Remark 3.2. In $\mathbb{R}^3$, we have $DIV^* \circ CURL^* = CURL^* \circ GRAD^* = 0$. And in $\mathbb{R}^2$, we have $CURL^* \circ GRAD^* = 0$

We will refer to this collection of operators as a data-driven exterior calculus, with the understanding that the metric information encoded in $\mathbf{B}_k$ and $\mathbf{D}_k$ will be learned from data. Note that in the traditional low-order compatible/mimetic schemes, these metric tensors contain geometric information related to the oriented measures of mesh entities, such as cell volumes, face moments, etc. [1]. Following from the exact sequence property, this exterior calculus structure inherits the following other properties of the graph calculus.

Naturally, based on the Hodge Laplacians (9) and (19), we can define generalized data-driven Hodge-Laplacians as follows,

$$\Delta_k = d_{k-1} d_k^* + d_k^* d_k$$

For example, for practical applications, we consider the following Hodge-Laplacians in $\mathbb{R}^2$:

$$\Delta_1 := CURL \circ CURL^* + GRAD^* \circ DIV$$

$$\Delta_2 := DIV \circ GRAD^*.$$

The Hodge decomposition (10) also can be generalized naturally. Here, we choose $\langle \mathbf{v}, \mathbf{w} \rangle_k := \langle \mathbf{v}, \mathbf{w} \rangle_{\mathbf{D}_k \mathbf{B}_k^{-1}} = (\mathbf{D}_k \mathbf{B}_k^{-1} \mathbf{v}, \mathbf{w})$ and denote its induced norm as $\|\cdot\|_k$.

Theorem 3.3 (Hodge Decomposition). For C^k , the following decomposition holds

$$C^k = \text{im}(d_{k-1}) \bigoplus_k \ker(\Delta_k) \bigoplus_k \text{im}(d_k^*), \quad (20)$$

where $\bigoplus_k$ means the orthogonality with respect to the $\langle \cdot, \cdot \rangle_k$ -inner product.

Proof. Since (19) is a commutative diagram, following from the Hodge decomposition (10), C^k can be naturally decomposed into three parts, $\text{im}(d_{k-1})$, $\text{im}(d_k^*)$, and $\ker(\Delta_k)$. Next we show this decomposition is orthogonal with respect to the $\langle \cdot, \cdot \rangle_k$ -inner product. For $\mathbf{u}_{k-1} \in C^{k-1}$ and $\mathbf{u}_{k+1} \in C^{k+1}$, we have

$$\begin{aligned} & \langle d_{k-1} \mathbf{u}_{k-1}, d_k^* \mathbf{u}_{k+1} \rangle_k \\ &= (\mathbf{B}_k \delta_{k-1} \mathbf{B}_{k-1}^{-1} \mathbf{u}_{k-1}, \mathbf{D}_k^{-1} \delta_k^* \mathbf{D}_{k+1} \mathbf{u}_{k+1})_{\mathbf{D}_k \mathbf{B}_k^{-1}} \\ &= (\delta_{k-1} \mathbf{B}_{k-1}^{-1} \mathbf{u}_{k-1}, \delta_k^* \mathbf{D}_{k+1} \mathbf{u}_{k+1}) \\ &= (\delta_k \delta_{k-1} \mathbf{B}_{k-1}^{-1} \mathbf{u}_{k-1}, \mathbf{D}_{k+1} \mathbf{u}_{k+1}) \\ &= 0 \end{aligned}$$

For $\mathbf{h}_k \in \ker(\Delta_k)$, we have $d_{k-1}^* \mathbf{h}_k = 0$ and $d_k \mathbf{h}_k = 0$, which implies $\delta_{k-1}^* \mathbf{D}_k \mathbf{h}_k = 0$ and $\delta_k \mathbf{B}_k^{-1} \mathbf{h}_k = 0$. And then for $\mathbf{u}_{k-1} \in C^{k-1}$,

$$\begin{aligned} & \langle d_{k-1} \mathbf{u}_{k-1}, \mathbf{h}_k \rangle_k \\ &= (\mathbf{B}_k \delta_{k-1} \mathbf{B}_{k-1}^{-1} \mathbf{u}_{k-1}, \mathbf{h}_k)_{\mathbf{D}_k \mathbf{B}_k^{-1}} \\ &= (\mathbf{D}_k \delta_{k-1} \mathbf{B}_{k-1}^{-1} \mathbf{u}_{k-1}, \mathbf{h}_k) \\ &= (\mathbf{B}_{k-1}^{-1} \mathbf{u}_{k-1}, \delta_{k-1}^* \mathbf{D}_k \mathbf{h}_k) \\ &= 0. \end{aligned}$$

On the other hand, for $\mathbf{u}_{k+1} \in C^{k+1}$, we have

$$\begin{aligned} & \langle d_k^* \mathbf{u}_{k+1}, \mathbf{h}_k \rangle_k \\ &= (\mathbf{D}_k^{-1} \delta_k^* \mathbf{D}_{k+1} \mathbf{u}_{k+1}, \mathbf{h}_k)_{\mathbf{D}_k \mathbf{B}_k^{-1}} \\ &= (\mathbf{B}_k^{-1} \delta_k^* \mathbf{D}_{k+1} \mathbf{u}_{k+1}, \mathbf{h}_k) \end{aligned}$$

$$\begin{aligned}
&= (\mathbf{D}_{k+1} \mathbf{u}_{k+1}, \delta_k \mathbf{B}_k^{-1} \mathbf{h}_k) \\
&= 0.
\end{aligned}$$

Thus, the decomposition is orthogonal with respect to the $\langle \cdot, \cdot \rangle_k$ -inner product, which completes the proof. $\square$

For example, in $\mathbb{R}^3$, we have the following Hodge decomposition when $k = 1$,

$$C^1 = \text{im}(\text{GRAD}) \bigoplus_k \ker(\Delta_1) \bigoplus_k \text{im}(\text{CURL}^*),$$

and when $k = 2$

$$C^2 = \text{im}(\text{CURL}) \bigoplus_k \ker(\Delta_2) \bigoplus_k \text{im}(\text{GRAD}^*).$$

In $\mathbb{R}^2$, we have the following Hodge decomposition when $k = 1$

$$C^1 = \text{im}(\text{CURL}) \bigoplus_k \ker(\Delta_1) \bigoplus_k \text{im}(\text{GRAD}^*).$$

Based on the Hodge decomposition, we have the following Poincaré inequality.

Theorem 3.4 (Poincaré inequality). *For each k , there exists a constant $c_{P,k}$ such that*

$$\|\mathbf{z}_k\|_k \leq c_{P,k} \|\mathbf{d}_k \mathbf{z}_k\|_{k+1}, \quad \mathbf{z}_k \in \text{im}(\mathbf{d}_k^*),$$

and another constant $c_{P,k}^*$ such that

$$\|\mathbf{z}_k\|_k \leq c_{P,k}^* \|\mathbf{d}_{k-1}^* \mathbf{z}_k\|_{k-1}, \quad \mathbf{z}_k \in \text{im}(\mathbf{d}_{k-1}).$$

Thus, for $\mathbf{u}_k \in C^k$, we have

$$\inf_{\mathbf{h}_k \in \ker(\Delta_k)} \|\mathbf{u}_k - \mathbf{h}_k\|_k \leq C \left(\|\mathbf{d}_k \mathbf{u}_k\|_{k+1} + \|\mathbf{d}_{k-1}^* \mathbf{u}_k\|_{k-1} \right),$$

where constant $C > 0$ only depends on $c_{P,k}$ and $c_{P,k}^*$.

Proof. These inequalities are a direct consequence of the Hodge decomposition (20). $\square$

Since we are considering matrix representation, by direct computation, we can see that

$$c_{P,k} = \lambda_{\min}(\mathbf{d}_k^* \mathbf{d}_k)^{-1/2},$$

where $\lambda_{\min}$ denotes the smallest non-trivial eigenvalue. For example, in $\mathbb{R}^3$, when $k = 0$,

$$c_{P,0} = \lambda_{\min}(\mathbf{d}_0^* \mathbf{d}_0)^{-1/2} = \lambda_{\min}(\delta_0^* \mathbf{D}_1 \mathbf{B}_1 \delta_0)^{-1/2} = \lambda_{\min}(\text{DIV}^* \circ \text{GRAD})^{-1/2}.$$

Note that

$$\min_i \{(\mathbf{D}_1 \mathbf{B}_1)_{ii}\} \lambda(\delta_0^* \delta_0) \leq \lambda(\delta_0^* \mathbf{D}_1 \mathbf{B}_1 \delta_0) \leq \max_i \{(\mathbf{D}_1 \mathbf{B}_1)_{ii}\} \lambda(\delta_0^* \delta_0).$$

This implies

$$\min_i \{(\mathbf{D}_1 \mathbf{B}_1)_{ii}\} \lambda_{\min}(\delta_0^* \delta_0) \leq c_{P,0}^{-2} \leq \max_i \{(\mathbf{D}_1 \mathbf{B}_1)_{ii}\} \lambda_{\min}(\delta_0^* \delta_0),$$

which relates the Poincaré constant with $\lambda_{\min}(\delta_0^* \delta_0)$, also known as the Fiedler value of the graph Laplacian $L_0 = \delta_0^* \delta_0$. Classical works provide bounds upon the Fiedler eigenvalue in terms of the size and degree of a given graph, see for example [23].

Another consequence of the Hodge decomposition (20) is the invertibility of the Hodge Laplacian Δ_k once we account for its kernel.

Theorem 3.5 (Invertibility of Hodge Laplacian). *The k^{th} -order Hodge Laplacian Δ_k is positive-semidefinite, with the dimension of its null-space equal to the dimension of the corresponding homology $H^k = \ker(\mathbf{d}_k) / \text{im}(\mathbf{d}_{k-1})$.*

Proof. For $\mathbf{u}_k \in C^k$, we have

$$\begin{aligned} & \langle \Delta_k \mathbf{u}_k, \mathbf{u}_k \rangle_k \\ &= ((\mathbf{B}_k \delta_{k-1} \mathbf{B}_{k-1}^{-1} \mathbf{D}_{k-1}^{-1} \delta_{k-1}^* \mathbf{D}_k + \mathbf{D}_k^{-1} \delta_k^* \mathbf{D}_{k+1} \mathbf{B}_{k+1} \delta_k \mathbf{B}_k^{-1}) \mathbf{u}_k, \mathbf{u}_k)_{\mathbf{D}_k \mathbf{B}_k^{-1}} \\ &= ((\mathbf{D}_k \delta_{k-1} \mathbf{B}_{k-1}^{-1} \mathbf{D}_{k-1}^{-1} \delta_{k-1}^* \mathbf{D}_k + \mathbf{B}_k^{-1} \delta_k^* \mathbf{D}_{k+1} \mathbf{B}_{k+1} \delta_k \mathbf{B}_k^{-1}) \mathbf{u}_k, \mathbf{u}_k) \\ &= \langle \mathbf{d}_{k-1}^* \mathbf{u}_k, \mathbf{d}_{k-1}^* \mathbf{u}_k \rangle_{k-1} + \langle \mathbf{d}_k \mathbf{u}_k, \mathbf{d}_k \mathbf{u}_k \rangle_{k+1} \end{aligned}$$

which shows that Δ_k is positive-semidefinite. The second part follows directly from the Hodge decomposition (20). $\square$

Theorem 3.5 means that the following linear system of the Hodge Laplacian

$$\Delta_k \mathbf{u}_k = \mathbf{f}_k, \quad (21)$$

is solvable as long as $\mathbf{f}_k \in C^k$ is orthogonal to $\ker(\Delta_k)$ with respect to the $\langle \cdot, \cdot \rangle_k$ -inner product.

3.2. Nonlinear perturbation of Hodge-Laplacian problems

In many cases, it is helpful to consider the mixed form of the Hodge-Laplacian problem (21) as it naturally provides connections to integral balance laws and conservation principles [5,2]. To this end, let us introduce a new variable $\mathbf{w}_{k-1} := \mathbf{d}_{k-1}^* \mathbf{u}_k$ corresponding to a generalized flux, and the mixed form of (21) as follows,

$$\mathbf{w}_{k-1} - \mathbf{d}_{k-1}^* \mathbf{u}_k = 0, \quad (22)$$

$$\mathbf{d}_{k-1} \mathbf{w}_{k-1} + \mathbf{d}_k^* \mathbf{u}_k = \mathbf{f}_k. \quad (23)$$

This class of problems describes several canonical second-order elliptic operators; for example, in $\mathbb{R}^2$, for $k=2$ we obtain the Darcy flow model

$$\begin{aligned} \mathbf{F} + \kappa \nabla \phi = 0 & \rightarrow \mathbf{w}_1 - \text{GRAD}^* \mathbf{u}_0 = 0 \\ \nabla \cdot \mathbf{F} = f & \quad \text{DIV} \mathbf{w}_1 = \mathbf{f}_0, \end{aligned}$$

and for $k=1$ we obtain the magnetostatics model, after applying a vector potential for the magnetic field and applying a suitable gauge condition [15,18].

$$\begin{aligned} \nabla \times \mathbf{J} = \mathbf{f} & \rightarrow \mathbf{w}_0 - \text{CURL}^* \mathbf{u}_1 = 0, \\ \nabla \cdot \mathbf{B} = 0 & \quad \text{CURL} \mathbf{w}_0 + \text{GRAD}^* \circ \text{DIV} \mathbf{u}_1 = \mathbf{f}_1. \\ \mathbf{J} = \mu \mathbf{B} \end{aligned}$$

While this model form is appropriate for learning, e.g. diffusion coefficients corresponding to second-order elliptic problems, realistic problems require accounting for nonlinearities. With this in mind, we introduce a nonlinear perturbation of the fluxes while leaving the relevant conservation statements untouched. This preserves the exterior calculus structure while incorporating data into fluxes only, which are traditionally more empirical. Any parameterization may be used for the nonlinearities, but we consider deep neural networks. As a result, we obtain a nonlinear perturbation of a Hodge-Laplacian problem in the mixed form as follows,

$$\mathbf{w}_{k-1} = \mathbf{d}_{k-1}^* \mathbf{u}_k + \epsilon \mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{u}_k; \xi), \quad (24)$$

$$\mathbf{d}_{k-1} \mathbf{w}_{k-1} + \mathbf{d}_k^* \mathbf{u}_k = \mathbf{f}_k, \quad (25)$$

where $\mathcal{NN} : \mathbb{R}^{\dim(C^{k-1})} \rightarrow \mathbb{R}^{\dim(C^{k-1})}$ denotes a vector of $\dim(C^{k-1})$ neural networks parameterizing a black-box generalized flux. The parameter ϵ is introduced for notational convenience and may take a value of either 1 or 0 if one wishes to include or not include nonlinearity, respectively. The corresponding primal form is

$$\Delta_k \mathbf{u}_k + \epsilon \mathbf{d}_{k-1} \circ \mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{u}_k; \xi) = \mathbf{f}_k \quad (26)$$

Later in this section, we will theoretically show that when the second term is sufficiently small in an appropriate norm, the nonlinear problem (24) and (25) remains well-posed. First, let us look at some examples. In $\mathbb{R}^2$, when $k=2$, we have

$$\begin{aligned} \text{DIV} \circ \text{GRAD}^* \mathbf{u}_0 + \epsilon \text{DIV} \circ \mathcal{NN}(\text{GRAD}^* \mathbf{u}_0; \xi) &= \mathbf{f}_0 \\ &= \Delta_0 \mathbf{u}_2 + \epsilon \text{DIV} \circ \mathcal{NN}(\text{GRAD}^* \mathbf{u}_0; \xi) = \mathbf{f}_0 \end{aligned}$$

and when $k = 1$, we have

$$\begin{aligned} \text{GRAD}^* \circ \text{DIV} \mathbf{u}_1 + \text{CURL} \circ \text{CURL}^* \mathbf{u}_1 + \epsilon \text{CURL} \circ \mathcal{NN}(\text{CURL}^* \mathbf{u}_1; \xi) &= \mathbf{f}_1, \\ &= \Delta_1 \mathbf{u}_1 + \epsilon \text{CURL} \circ \mathcal{NN}(\text{CURL}^* \mathbf{u}_1; \xi) = \mathbf{f}_1. \end{aligned}$$

3.3. Well-posedness

Next we investigate the well-posedness of the perturbed Hodge-Laplacian problem (24)-(25). Consider the space $\mathbb{V} = C^k / \ker(\Delta_k)$, we introduce the following weak formulation of (26),

$$a(\mathbf{u}_k, \mathbf{v}) + \epsilon \langle \mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{u}_k), \mathbf{d}_{k-1}^* \mathbf{v} \rangle_{k-1} = \langle \mathbf{f}, \mathbf{v} \rangle_k, \quad \mathbf{v} \in \mathbb{V}, \quad (27)$$

where

$$a(\mathbf{u}, \mathbf{v}) := \langle \mathbf{d}_{k-1}^* \mathbf{u}, \mathbf{d}_{k-1}^* \mathbf{v} \rangle_{k-1} + \langle \mathbf{d}_k \mathbf{u}, \mathbf{d}_k \mathbf{v} \rangle_{k+1}, \quad \forall \mathbf{u}, \mathbf{v} \in \mathbb{V},$$

and its induced norm is $\|\mathbf{u}\|_a := \sqrt{a(\mathbf{u}, \mathbf{u})} = \sqrt{\|\mathbf{d}_{k-1}^* \mathbf{u}\|_{k-1}^2 + \|\mathbf{d}_k \mathbf{u}\|_{k+1}^2}$. We assume Lipschitz continuity of the nonlinear perturbation, i.e. that there exists a constant $L_N > 0$, such that,

$$\|\mathcal{NN}(\mathbf{v}_{k-1}) - \mathcal{NN}(\mathbf{w}_{k-1})\|_{k-1} \leq L_N \|\mathbf{v}_{k-1} - \mathbf{w}_{k-1}\|_{k-1}, \quad \forall \mathbf{v}_{k-1}, \mathbf{w}_{k-1} \in C^{k-1}. \quad (28)$$

In addition, we also assume that

$$\mathcal{NN}(\mathbf{0}) = \mathbf{0}, \quad (29)$$

which means that the nonlinear perturbation reduces to a linear gradient closure in the limit as the gradient becomes small.

The main tool we use is the Leray-Schauder fixed point theorem [32]. We define $T : \mathbb{V} \mapsto \mathbb{V}$ such that for each $\mathbf{w} \in \mathbb{V}$, $\tilde{\mathbf{u}} := T(\mathbf{w}) \in \mathbb{V}$ is given as the solution of the following linear problem,

$$a(\tilde{\mathbf{u}}, \mathbf{v}) + \epsilon \langle \mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{w}), \mathbf{d}_{k-1}^* \mathbf{v} \rangle_{k-1} = \langle \mathbf{f}, \mathbf{v} \rangle_k, \quad \forall \mathbf{v} \in \mathbb{V}. \quad (30)$$

The map T is clearly continuous and, therefore, compact in the finite dimensional space $\mathbb{V}$. The solvability of (27) is thus equivalent to the solvability of the equation $\mathbf{u}_k = T(\mathbf{u}_k)$ in $\mathbb{V}$, which is a fixed point problem.

If $\lambda > 0$ and $\mathbf{w}$ satisfies $T(\mathbf{w}) = \lambda \mathbf{w}$, then

$$\lambda a(\mathbf{w}, \mathbf{v}) + \epsilon \langle \mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{w}), \mathbf{d}_{k-1}^* \mathbf{v} \rangle_{k-1} = \langle \mathbf{f}, \mathbf{v} \rangle_k, \quad \forall \mathbf{v} \in \mathbb{V}.$$

By choosing $\mathbf{v} = \mathbf{w}$, we obtain

$$\lambda \|\mathbf{w}\|_a^2 \leq \|\mathbf{f}\|_{-a} \|\mathbf{w}\|_a + \epsilon \|\mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{w})\|_{k-1} \|\mathbf{d}_{k-1}^* \mathbf{w}\|_{k-1},$$

after application of the inequality $\langle \mathbf{f}, \mathbf{w} \rangle_k \leq \|\mathbf{f}\|_{-a} \|\mathbf{w}\|_a$, which follows directly from the definition of the negative norm, i.e.,

$\|\mathbf{f}\|_{-a} = \sup_{\mathbf{v} \in \mathbb{V}} \frac{\langle \mathbf{f}, \mathbf{v} \rangle_k}{\|\mathbf{v}\|_a}$. By the Lipschitz continuity assumption (28) and assumption (29), we have

$$\|\mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{w})\|_{k-1} = \|\mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{w}) - \mathcal{NN}(\mathbf{0})\|_{k-1} \leq L_N \|\mathbf{d}_{k-1}^* \mathbf{w}\|_{k-1}.$$

Using the fact that $\|\mathbf{d}_{k-1}^* \mathbf{w}\|_{k-1} \leq \|\mathbf{w}\|_a$, we have

$$\lambda \|\mathbf{w}\|_a^2 \leq (\|\mathbf{f}\|_{-a} + \epsilon L_N \|\mathbf{w}\|_a) \|\mathbf{w}\|_a,$$

therefore,

$$\lambda \leq \frac{\|\mathbf{f}\|_{-a} + \epsilon L_N \|\mathbf{w}\|_a}{\|\mathbf{w}\|_a}.$$

Thus, $\lambda < 1$ holds true for any $\mathbf{w}$ being on the boundary of the ball in $\mathbb{V}$ centered at the origin with radius

$$\rho \geq \frac{\|\mathbf{f}\|_{-a}}{(1 - \epsilon L_N)}.$$

Consequently, the Leray-Schauder fixed point theorem implies that the nonlinear map T has a fixed point in any ball centered at the origin with radius $\rho \geq \frac{\|\mathbf{f}\|_{-a}}{(1 - \epsilon L_N)}$. This fixed point is a solution of the equation (27).

Theorem 3.6. Assume (28) and (29) hold. If $\epsilon L_N < 1$, the equation (27) has an unique solution $\mathbf{u}_k \in \mathbb{V}$ satisfies

$$\|\mathbf{u}_k\|_a \leq \frac{\|\mathbf{f}\|_{-a}}{(1 - \epsilon L_N)}. \quad (31)$$

Proof. The existence has been discussed before the theorem. Let $\mathbf{u}_k$ be the solution and $\mathbf{v} = \mathbf{u}_k$ in (27), we have

$$\|\mathbf{u}_k\|_a^2 \leq (\|\mathbf{f}\|_{-a} + \epsilon L_N \|\mathbf{u}_k\|_a) \|\mathbf{u}_k\|_a,$$

which implies (31).

Next we prove uniqueness of (27) under the same assumptions. Let $\mathbf{u}_k$ and $\bar{\mathbf{u}}_k$ be two solutions of (27). Denoting $\mathbf{e} = \mathbf{u}_k - \bar{\mathbf{u}}_k$, we have,

$$a(\mathbf{e}, \mathbf{v}) + \epsilon \langle \mathcal{NN}(\mathbf{d}_{k-1}^* \mathbf{u}_k), \mathbf{d}_{k-1}^* \mathbf{v} \rangle_{k-1} - \epsilon \langle \mathcal{NN}(\mathbf{d}_{k-1}^* \bar{\mathbf{u}}_k), \mathbf{d}_{k-1}^* \mathbf{v} \rangle_{k-1} = 0.$$

Letting $\mathbf{v} = \mathbf{e}$ and using the Lipschitz continuity (28), we arrive at,

$$\|\mathbf{e}\|_a^2 \leq \epsilon L_N \|\mathbf{d}_{k-1}^* \mathbf{e}\|_{k-1}^2 \leq \epsilon L_N \|\mathbf{e}\|_a^2$$

which implies the uniqueness of the solution since $\epsilon L_N < 1$. This completes the proof. $\square$

We finally consider design of an architecture which satisfies this condition to ensure the extracted model is solvable, considering multilayer perceptrons as canonical architectures [57].

Remark 3.3. In our implementation, we consider a component-wise defined neural net which has the following structure

$$\mathcal{NN}(\mathbf{x}) := \begin{pmatrix} \mathcal{NN}^1(x_1) \\ \vdots \\ \mathcal{NN}^n(x_n) \end{pmatrix}, \quad \mathcal{NN}^i(x_i) = \psi_{L_i}^i \circ T_{L_i}^i \circ \cdots \circ \psi_1^i \circ T_1^i(x_i)$$

where $T_\ell^i(\mathbf{x}_\ell^i) := M_\ell^i \mathbf{x}_\ell^i + \mathbf{b}_\ell^i$ is an affine function and ψ_ℓ^i is a nonlinear activation function. If we assume ψ_ℓ^i are c_ℓ^i -Lipschitz nonlinear functions (e.g., ReLU, Leaky ReLU, tanh, and sigmoid are 1-Lipschitz functions) and define $\mathcal{N}_\ell^i(x_i) := T_\ell^i \circ \psi_{\ell-1}^i \cdots \circ \psi_1^i \circ T_1^i(x_i)$, then the Lipschitz constant L_N (28) of $\mathcal{NN}$ can be estimated as follows,

$$\begin{aligned} L_N &\leq \sup_{\mathbf{x} \in \mathbb{R}^n} \|\nabla \mathcal{NN}(\mathbf{x})\|_{k-1} = \sup_{\mathbf{x} \in \mathbb{R}^n} \sup_{\mathbf{y} \in \mathbb{R}^n} \frac{\|[\nabla \mathcal{NN}(\mathbf{x})]\mathbf{y}\|_{k-1}}{\|\mathbf{y}\|_{k-1}} \\ &= \sup_{\mathbf{x} \in \mathbb{R}^n} \|\nabla \mathcal{NN}(\mathbf{x})\| = \sup_{\mathbf{x} \in \mathbb{R}^n} \left\| \text{diag} \left(\frac{d\mathcal{NN}^1(x_1)}{dx_1}, \dots, \frac{d\mathcal{NN}^n(x_n)}{dx_n} \right) \right\| \\ &\leq \max_i \left\{ \sup_{x_i \in \mathbb{R}} \|\text{diag}((\psi_L^i)'(\mathcal{N}_L^i(x_i))) M_L^i \cdots M_2^i \text{diag}((\psi_1^i)'(\mathcal{N}_1^i(x_i))) M_1^i\| \right\} \\ &\leq \max_i \left\{ \prod_{\ell=1}^{L_i} c_\ell^i \|M_\ell^i\| \right\} \end{aligned}$$

Therefore, if we choose $\epsilon < \left(\max_i \left\{ \prod_{\ell=1}^{L_i} c_\ell^i \|M_\ell^i\| \right\} \right)^{-1}$, then the assumption $\epsilon L_N < 1$ would be satisfied. We also note that a more accurate upper bound of the Lipschitz constant L_N can be computed numerically using the advanced algorithms developed in [58], providing a tighter bound on ϵ .

3.4. Construction of chain complex

The model introduced in the previous section assumes access to an underlying graph to apply the DDEC to. Motivated by our surrogate modeling application, we assume access to a very fine polygonal mesh and access to a high-fidelity PDE solution defined as oriented moments of mesh entities (i.e. cell average scalar potentials, face average fluxes, edge average circulations). We next show how a coarsening of the underlying fine mesh preserves some structure allowing a particularly simple implementation; essentially, given graph grad/curl/div matrices on the fine mesh, we derive coarsening matrices that encode proper orientations for the coarsened complex and the desired $\mathbf{d}_k$ and $\mathbf{d}_k^*$ operators. We consider here graph-cut coarsening available in packages such as METIS [40], which partitions the domain Ω into N_c disjoint volumetric subdomains

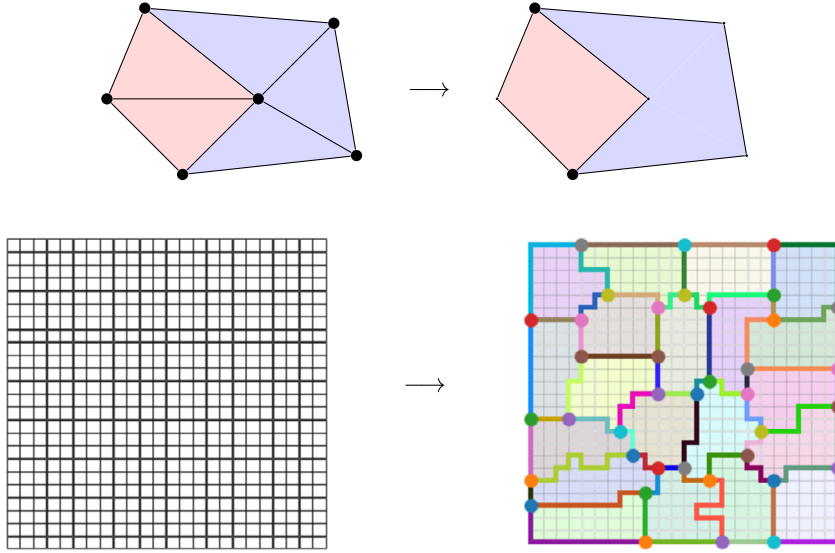


Fig. 1. The coarsening process takes a fine mesh discretizing Ω in d -cells, partitions them into disjoint subdomains and uses the fine scale mesh to derive coarsened k -cells and coboundary operators ∂_k with consistent orientations. Values on the coarsened k -cells correspond to sums of the constituent fine-grained moments with proper accounting of orientation. *Top:* five 2-cells coarsened into a pair of 2-cells. *Bottom:* a representative coarsening of a Cartesian mesh. Coarsened 0-, 1-, and 2-cells are each colored differently and are a formal linear sum of elements of the fine mesh.

(or d -cells), and derive appropriately oriented lower degree k -cells. For an unweighted set of partitions, specification of the parameter N_c in METIS fully specifies the process for generating N_c subdomains. In Fig. 1 we provide a cartoon of the process. While the following presentation provides a mathematical description of the process via a commutative diagram, it practically will provide a simple implementation of the coarse coboundary operators δ_k in terms of a few simple matrices, greatly simplifying implementation.

Remark 3.4. This is one particular construction appropriate for synthetic data which takes advantage of available adjacency matrices of an underlying fine mesh. We stress however that one may apply the calculus to *any* appropriately defined graph. For example, in experimental contexts the bins associated with histograms may be used instead, or the calculus may be applied to graphs with no associated partition of space at all (see e.g. [39]). We further comment that the graph-cut coarsening assumed here provides a quasi-uniform partition of space that does not take advantage of the data; in another work we consider spectral graph partitioning strategies to obtain data-driven partitions tuned to give optimal representations of data [37].

We assume the high-fidelity solution is associated with a d -dimensional chain complex, with k -cells denoted by F_k for $k < d$ and $\partial_k^{fine} : F_{k+1} \rightarrow F_k$ as defined in Section 2, and coboundary $\delta_k^{fine} : F^k \rightarrow F^{k+1}$ encoded via the adjacency matrices generally available in mesh data structures. Here, we use the symbol F and superscript *fine* to denote fine scale objects. Our goal is to construct a coarse complex, denoted by the symbol C , that can be used in the DDEC framework as we discussed in Section 3.1-3.3. The construction contains two steps. First, we will coarsen the fine level set of the k -chain, F_k , to obtain the coarse level set of the k -chain, C_k . This can be done inductively by starting with coarsening the d -cells. Then we will define the coarse boundary and coboundary operators by relating the fine and coarse level complex properly.

Constructing C_k and C^k To inductively define C_k from F_k , we start with coarsening the d -cells. Given a partitioning of the N_f d -cells into N_c disjoint, connected subsets, denoted by $F_d = \sqcup_{i=1}^{N_c} P_i$. We define the space of coarse d -chain C_d as a subspace of F_d based on the partition $\{P_i\}_{i=1}^{N_c}$ with the following natural inclusion

$$\iota_d \in \mathbb{R}^{N_f \times N_c}, \quad (\iota_d)_{ij} = \begin{cases} 1, & f_i \in P_j, \\ 0, & f_i \notin P_j. \end{cases}$$

This has a dual C^d identified as a subspace of F^d .

Now, consider the image of C_d under ∂_{d-1}^{fine} , consisting of the fine d -chains lying at partition interfaces. We define the element $c_{ij} := P_i \cap P_j$ (oriented consistently) for $i, j = 1, 2, \dots, N_c$ and their span as the coarse $(d-1)$ chains C_{d-1} . Then we can proceed inductively. With the chain and cochain spaces indexed by $d-k$, consider the image of C_{d-k} under the map ∂_{d-k-1}^{fine} . This consists of the $(d-k-1)$ -interfaces between the coarse $d-k$ -chains. Define the element

$$c_{i_1 i_2 \dots i_{d-k-1}} = P_{i_1} \cap P_{i_2} \cap \dots \cap P_{i_{d-k-1}}$$

The span of these elements is the space of coarse $(d - k - 1)$ -chains C_{d-k-1} . As before, by identifying C_{d-k-1} as a subset of F_{d-k-1} , we can define the dual cochains C^{d-k-1} .

Remark 3.5. It may be the case that $c_{i_1 i_2 \dots i_{d-k-1}}$ is disjoint. As a choice of implementation detail, each of the connected components can be taken as a separate chain, or their disjoint union can be taken as a single chain. What results from this choice is the treatment of parallel fluxes between partitions as either: a sum of individual distinct fluxes between the partitions, or the sum of those fluxes as a single effective quantity between the partitions. We choose the former in our computational examples.

By construction, we have $C_k \subset F_k$, for $k = 0, \dots, d$. Therefore, there is a natural inclusion $\iota_k : C_k \hookrightarrow F_k$ whose entries are 1, -1 , or 0 depending on the partition and orientation. Furthermore, we also have $C^k \subset F^k$ and the corresponding natural inclusions are defined as $\iota^k := \iota_{d-k}$, for $k = 0, \dots, d$.

Building boundary and coboundary operators In order to define the coarse boundary and coboundary operators that connects C_k and C^k , we first define the linear projections $\pi_k : F_k \mapsto C_k$ as

$$\pi_k = \left(\iota_k^T \iota_k \right)^{-1} \iota_k^T,$$

and $\pi^k : F^k \mapsto C^k$ as

$$\pi^k = \left(\left(\iota^k \right)^T \iota^k \right)^{-1} \left(\iota^k \right)^T.$$

Note that, $\pi^k = \pi_{d-k}$ since $\iota^k = \iota_{d-k}$.

With these in hand, we define the coarse boundary operators as follows,

$$\partial_k := \pi_k \cdot \partial_k^{fine} \cdot \iota_{k+1},$$

and the coarse coboundary operators

$$\delta_k := \pi^{k+1} \cdot \delta_k^{fine} \cdot \iota^k.$$

Remark 3.6. The linear projections π_k and π^k are least-squares projections. Since the natural inclusions ι_k and ι^k are constructed based on partitions, $\iota_k^T \iota_k$ and $(\iota^k)^T \iota^k$ are diagonal matrices with diagonal entries equal to the number of k -chains in the corresponding partition. Thus, in a practical implementation, we can simply use $\pi_k = \iota_k^T$ and $\pi^k = (\iota^k)^T$ and, based on the definition (14), $(\iota_k^T \iota_k)^{-1}$ and $((\iota^k)^T \iota^k)^{-1}$ are implicitly absorbed in $\mathbf{B}_k$ and $\mathbf{D}_k$.

To summarize, in the case of $d = 2$, all of the spaces discussed are related through the following commutative diagram:

$$\begin{array}{ccccccc}
 0 & \longrightarrow & F^0 & \xrightarrow{\delta_0^{fine}} & F^1 & \xrightarrow{\delta_1^{fine}} & F^2 \longrightarrow 0 \\
 & & \downarrow \pi^0 & \swarrow \iota^0 & \downarrow \pi^1 & \swarrow \iota^1 & \downarrow \pi^2 \swarrow \iota^2 \\
 0 & \longrightarrow & C^0 & \xrightarrow{\delta_0} & C^1 & \xrightarrow{\delta_1} & C^2 \longrightarrow 0 \\
 & & \downarrow & & \downarrow & & \downarrow \\
 0 & \longleftarrow & F_0 & \xleftarrow{\partial_0^{fine}} & F_1 & \xleftarrow{\partial_1^{fine}} & F_2 \longleftarrow 0 \\
 & & \downarrow \pi_0 & \swarrow \iota_0 & \downarrow \pi_1 & \swarrow \iota_1 & \downarrow \pi_2 \swarrow \iota_2 \\
 0 & \longleftarrow & C_0 & \xleftarrow{\partial_0} & C_1 & \xleftarrow{\partial_1} & C_2 \longleftarrow 0
 \end{array} \tag{32}$$

The top-front row of this diagram corresponds to the middle row of the chain complex (19). The DDEC complex results from perturbing these operators as in (14).

4. PDE-constrained optimization

We finally turn toward the question of how to fit a model of the form (26) to data. We assume access to data via the coarsening process of the previous section, and denote by $\mathbf{X}_{data}$ a vector concatenating the coarsened $\mathbf{w}_{k+1}$ (flux) and $\mathbf{u}_k$ (potential) degrees of freedom. We may have only partial data: observations of possibly a single field $\mathbf{w}_{k+1}$ or $\mathbf{u}_k$, or observations on a subset of the chain complex. We then concisely express the boundary value problem in (26) via

the nonlinear operator $\mathcal{L}_\xi[\mathbf{X}] = \mathbf{f}$, where we have lumped all model parameters into ξ (i.e. $\mathbf{B}_k$, $\mathbf{D}_k$, and neural network weights and biases), and refer to this as the *forward problem*. We postpone a problem-specific discussion of how boundary conditions will be imposed for the following section, and assume the forward problem is prescribed such that BCs are imposed naturally. We seek a solution to the following quadratic program with nonlinear equality constraints

$$\min_{\mathbf{X}, \xi} \|\mathbf{X} - \mathbf{X}_{data}\|^2 \quad (33)$$

$$\text{such that } \mathcal{L}_\xi[\mathbf{X}] = \mathbf{f}, \quad (34)$$

where $\|\cdot\|$ denotes the ℓ_2 norm.

To avoid handling the equality constraint, one could introduce a single scalar penalty parameter λ and use a gradient descent optimizer to solve

$$\min_{\xi} \|\mathbf{X} - \mathbf{X}_{data}\|^2 + \lambda \|\mathcal{L}_\xi[\mathbf{X}] - \mathbf{f}\|^2. \quad (35)$$

This approach resembles currently popular approaches such as physics-informed neural networks [56] and is simple to implement in machine learning libraries using automatic differentiation to implement first-order optimization schemes. However, λ becomes a hyperparameter introducing well-known issues with training [67,65], and ultimately results in $\mathcal{L}_\xi[\mathbf{X}] = \mathbf{f}$ holding only to within optimization error and a subsequent loss of structure preservation. We instead enforce the equality constraint exactly, introducing Lagrange multipliers λ as follows.

$$\min_{\xi, \mathbf{X}, \lambda} \mathbf{L}_\xi(\mathbf{X}, \lambda) \quad (36)$$

$$\text{such that } \mathcal{L}_\xi[\mathbf{X}] = \mathbf{f}, \quad (37)$$

$$\text{where } \mathbf{L}_\xi(\mathbf{X}, \lambda) = \|\mathbf{X} - \mathbf{X}_{data}\|^2 + \lambda^\top (\mathcal{L}_\xi[\mathbf{X}] - \mathbf{f}) \quad (38)$$

Stationarity of the Karush-Kuhn Tucker conditions requires that the gradient of $\mathbf{L}_\xi(\mathbf{X}, \lambda)$ with respect to $\mathbf{X}$, λ and ξ be zero. This yields the following set of three necessary conditions for a minimizer of (36), which we will iteratively solve. If this fixed point iteration converges and all three are satisfied, then one has obtained a minimizer.

- **Forward problem:** The condition $\partial_{\lambda} \mathbf{L}_\xi(\mathbf{X}, \lambda) = 0$ requires that the forward problem is solved: $\mathcal{L}_\xi[\mathbf{X}] = \mathbf{f}$. Assuming ξ fixed, one may solve with a Newton-Raphson method, requiring calculation of the Jacobian of $\mathcal{L}_\xi$, which we denote $J_{\xi, \mathbf{X}}$. Following the analysis in Section 3, this problem is guaranteed to be solvable provided $\epsilon L_n < 1$.
- **Adjoint problem:** Enforcing $\partial_{\mathbf{X}} \mathbf{L}_\xi(\mathbf{X}, \lambda) = 0$ provides the linear adjoint problem $J_{\xi, \mathbf{X}}^\top \lambda = -2(\mathbf{X} - \mathbf{X}_{data})$ for the Lagrange multipliers. Having solved the forward problem in the previous step, the Jacobian is already available.
- **Model update:** The remaining condition $\partial_{\xi} \mathbf{L}_\xi(\mathbf{X}, \lambda) = 0$ does not readily admit solution with second-order optimizers due to the neural networks embedded in the nonlinear perturbations of (26). It is well-known that neural networks admit a complex optimization landscape due to linear dependence with many suboptimal local minima. With this in mind, we apply a single step of a first-order gradient optimizer instead, providing a small perturbation of the model at each iteration.

This process ensures that the physics imposed by the carefully designed model formed in the previous sections hold to machine precision at each iteration, even in scenarios with limited training data. Asymptotically, the added complexity compared to typical gradient-descent approaches is that of solving the forward problem at each iteration, which is generally inexpensive for the low-dimensional models under consideration and converges rapidly. The remaining complexity lies in calculating the relevant derivatives for Newton, which may be simply calculated with the same automatic differentiation used to perform the gradient descent step.

Considering that the assumed model form is nonlinear, it is necessary to train simultaneously on a variety of boundary conditions to learn the nonlinear response across a range of conditions - we must therefore assimilate multiple solutions to the nonlinear problem simultaneously. With that in mind, we present in Algorithm 1 a batch-training strategy for handling training data $\mathcal{D} = \{\mathbf{X}_{data, i}\}_{i=1}^{N_{batches}}$ consisting of $N_{batches}$ solutions.

When fitting a model to a system governed by a linear PDE, it is desirable to obtain a parsimonious representation which learns the metric information in $a(u, v)$ while setting the contribution from $\mathcal{NN}$ to zero; by linearity, such a model will generalize well to different boundary conditions, whereas a nonlinear one will handle out-of-distribution inference poorly. While one can impose this directly by setting $\epsilon = 0$, we propose a regularizer to include nonlinear terms only if necessary. In light of Remark 3.3, assuming 1-Lipschitz activation functions are used, one may indirectly control the magnitude of the Lipschitz constant relative to the Poincare inequality, by introducing the regularization of Equation (36)

$$\mathbf{L}_\xi(\mathbf{X}, \lambda) + \epsilon_L \sum_i \left(\prod_{\ell=1}^{L_i} \|M_\ell^i\| \right)^2, \quad (39)$$

Data: Training solutions $\mathcal{D}$, parameter initialization ξ_0 , tolerance ϵ_{TOL}
Result: ξ
for $e \in \{1, \dots, \text{epochs}\}$ **do**
 $\mathbf{X} \leftarrow \mathbf{0}$;
 for $\mathbf{X}_{data,i} \in \mathcal{D}$ **do**
 Solve forward problem;
 while $\|\mathcal{L}_\xi[\mathbf{X}] - f\| > \epsilon_{TOL}$ **do**
 $J_{\xi,\mathbf{X}} \leftarrow \nabla_{\mathbf{X}} \mathcal{L}_\xi(\mathbf{X})$;
 Solve $\Delta = -(J_{\xi,\mathbf{X}})^{-1} \mathcal{L}_\xi(\mathbf{X})$;
 $\mathbf{X} \leftarrow \mathbf{X} + \Delta$;
 end
 Solve adjoint problem;
 $\lambda \leftarrow 2(J_{\xi,\mathbf{X}})^{-\top} (\mathbf{X}_{data,i} - \mathbf{X})$;
 Apply gradient descent update;
 $\xi \leftarrow GD(\mathcal{L}_\xi(\mathbf{X}, \lambda), \xi)$;
 end
end

Algorithm 1: Application of equality constrained optimizer batched over multiple solutions $\mathbf{X}_{i,data}$. By $GD(l, p)$ we denote a gradient descent update of the parameters p to minimize the loss l . For this work we apply the Adam optimizer [41].

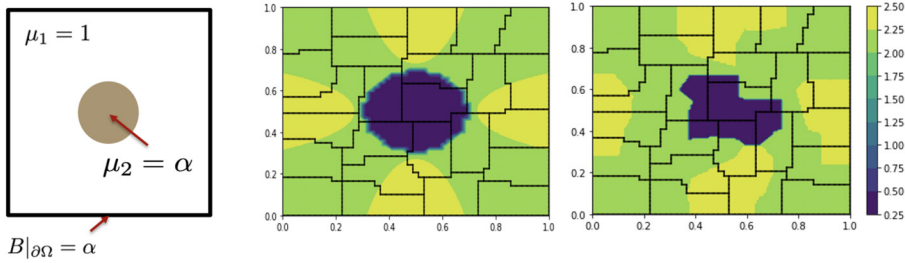


Fig. 2. We consider a cylindrical inclusion in a 2D domain for steady diffusion problems in subsurface flow and magnetostatics (left), considering diffusion coefficients which vary with the boundary conditions magnitude to obtain a nonlinear response. (See Fig. 2.) The high-fidelity solution is coarsened, obtaining a target solution on the corresponding k -chains (center). The coarsened graph model is trained to reproduce moments on each subdomain (right). Pictured here is the resulting magnetic field defined on coarsened 0-chains, with nearest neighbor interpolant of surrogate solution. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

where $\epsilon_L > 0$ is a penalty term to be calibrated, and $\|M_\ell^i\|$ denotes the 2-norm of the weights in the ℓ^{th} layer of the i^{th} neural network. Introduction of the regularizer does not alter the forward or adjoint solves during training.

5. Numerical results

In the remainder we present results for two canonical $H(\text{div})$ and $H(\text{curl})$ problems from subsurface flow and magnetostatics. In both examples we consider a cylindrical inclusion of radius a embedded within the unit square centered at the origin, whose material properties are prescribed by

$$\mu_\alpha(\mathbf{x}) = \begin{cases} \alpha, & \text{if } \|\mathbf{x}\| < a \\ 1, & \text{otherwise} \end{cases} \quad (40)$$

Treatment of the material interface at $\|\mathbf{x}\| = a$ without introducing spurious oscillations is a hallmark of mimetic discretizations and stems from the exact treatment of interface conditions via Stokes theorem, namely, that the jump in flux across an interface be zero.

For both problems we impose boundary conditions by identifying appropriate cochains on the boundary, replacing their corresponding row of the Jacobian matrix with zeros and a one on the diagonal, and setting the desired value on the right hand side. For the introductory nature of this paper this is sufficient, but we note that the imposition of boundary conditions is a rich topic in the discrete exterior calculus literature, with more complex applications requiring a deeper consideration of the interaction between boundary conditions and the discrete exterior calculus [24,43,12].

The parameterization of nonlinear fluxes requires specification of activation function and network width/depth. We use small multilayer perceptron networks of $O(10)$ parameters and varying activation functions to highlight robustness to architectural hyperparameters. Network hidden layers are initialized with the He initializer [36], the final linear layer is initialized to zero, and minimization is performed with Adam [41]. The only hyperparameter requiring calibration is the learning rate in Adam, which we calibrate by starting with a value of 0.01 and reducing until we obtain a good result. Training data is

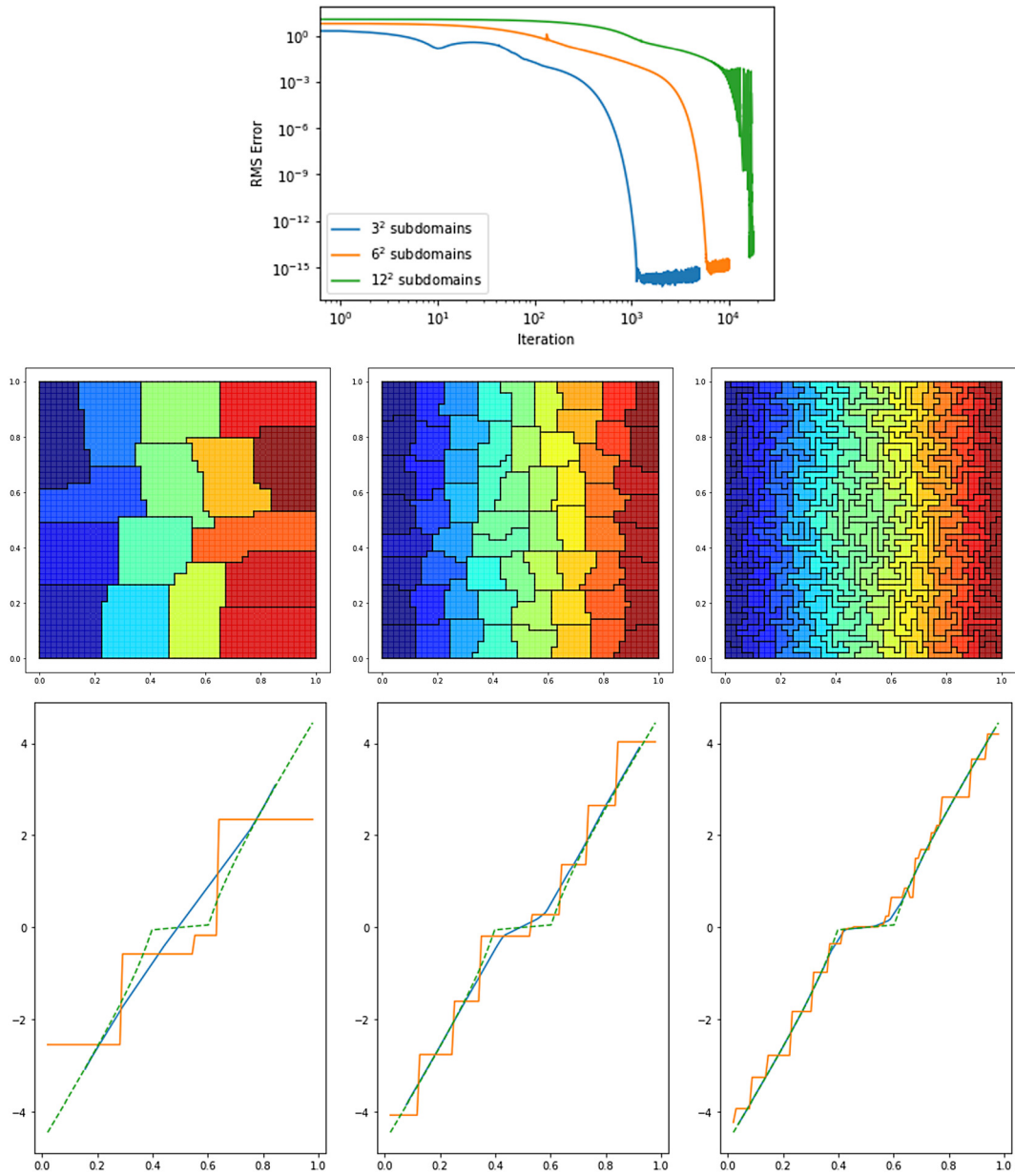


Fig. 3. Solutions for problem D1. *Top:* RMS error vs number of subdomains in coarsened mesh, using learning rate of 0.05, 0.005, 0.0005 for 3^2 , 6^2 , 12^2 subdomains, respectively, demonstrating ability to capture training data to machine precision. *Center:* piecewise constant plots of pressure for obtained model. *Bottom:* Comparison of pressure along $y = 0.5$ for model solution (orange) and Delaunay interpolant (blue) to training solution (dashed green) for increasing resolution, using coarsened subdomain centroids to compute interpolant.

generated via a high-fidelity DEC discretization on a grid of 50^2 cells. Accuracy is presented in terms of the root-mean-square (RMS) error of the cochains, $RMS(y) = \sqrt{\frac{1}{N} \sum_{i=1}^N y_i^2}$.

5.1. Problem 1: Darcy

We consider as training data solutions to the system of equations

$$\begin{aligned} \mathbf{F} + \mu_\alpha \nabla \phi &= 0 & \mathbf{x} \in \Omega \\ \nabla \cdot \mathbf{F} &= f & \mathbf{x} \in \Omega \end{aligned}$$

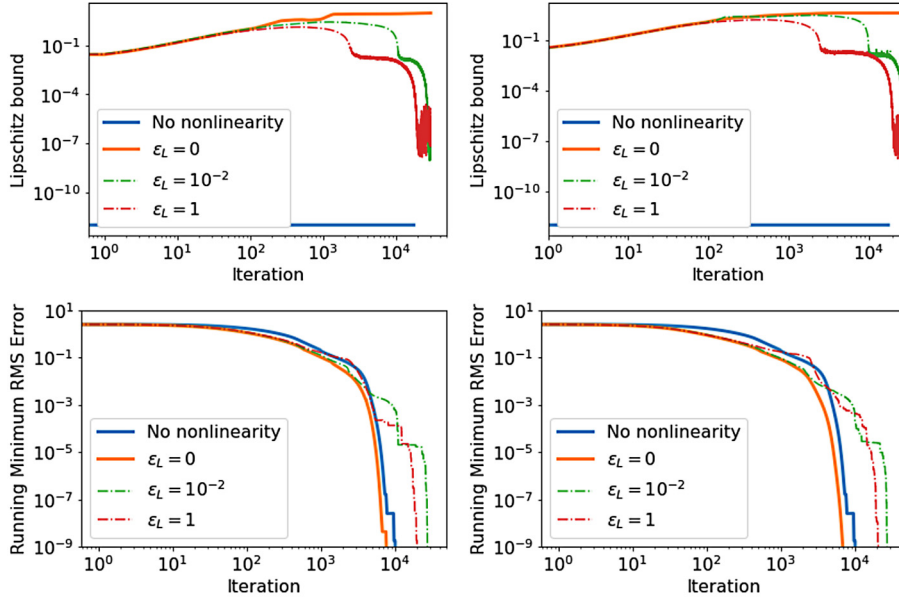


Fig. 4. Study for problem D1. Exploring choice of bias to learn parsimonious linear operators. In light of Remark 3.3, upper bound on Lipschitz constant L_N (top) quantifies magnitude of nonlinearity given extant of bias, with Theorem 3.6 guaranteeing well-posedness for $L_N < 1$. Plots of corresponding error (bottom) demonstrate ability to achieve machine precision either by explicitly turning off nonlinearity (setting $\epsilon = 0$) or penalizing operator norm of neural networks. Both ReLU (left) and tanh (right) activations are considered, with accuracy appearing robust to choice of ϵ_L . All runs are conducted with 9 partitions, a learning rate of 10^{-4} , and shallow networks of width 10.

$$\mathbf{F} \cdot \hat{\mathbf{n}} = g \quad \mathbf{x} \in \partial\Omega$$

obtained via the scheme in [51]. Boundary conditions are imposed upon the 1-cochains encoding the fluxes of $\mathbf{F}$ through subdomain boundaries, and we stress that the resulting DEC method will guarantee that $\int_{\Omega} \nabla \cdot \mathbf{F} dx = \int_{\partial\Omega} g dA$. For this problem we will consider two scenarios.

Darcy problem 1 (D1): We take $f = 0$ and $g = \langle 1, 0 \rangle \cdot \langle 1, 0 \rangle$ consistent with applying a unit horizontal flux and consider training data associated with a fixed α . In this setting, the PDE is linear and will not require a contribution from the neural network, and we use this to gauge the method's ability to recover a PDE discretization consistent with a single solution (i.e. $N_{data} = 1$). In Fig. 3 we demonstrate how the scheme is able to recover the high-fidelity solution moments to machine precision while varying the number of partitions. In Fig. 4 we demonstrate how a parsimonious linear operator can be obtained either by explicitly setting $\epsilon = 0$ or by calibrating the Lipschitz regularizer ϵ_L .

Darcy problem 2 (D2): We take $f = 0$ and $g = \langle \alpha, 0 \rangle \cdot \langle 1, 0 \rangle$ consistent with applying a horizontal flux which scales with the diffusion parameter. This corresponds to a material which becomes more conductive as the current/flux increases, and train over $N_{data} = 3$ solutions corresponding to $\mu \in \{1, 2, 4\}$. Capturing the nonlinear behavior of this problem requires learning of nonlinear fluxes which depend upon the magnitude of the potential. Fig. 5 provides a summary and discussion of results.

5.2. Problem 2: magnetostatics

We consider as training data solutions to the system of equations

$$\begin{aligned} \nabla \times \mathbf{J} &= \mathbf{f} & \mathbf{x} \in \Omega \\ \nabla \cdot \mathbf{B} &= 0 & \mathbf{x} \in \Omega \\ \mathbf{B} &= \mu_{\alpha} \mathbf{J} & \mathbf{x} \in \Omega \\ \mathbf{B} \cdot \hat{\mathbf{n}} &= g & \mathbf{x} \in \partial\Omega \end{aligned}$$

obtained via the scheme in [50]. For this 2D problem, the magnetic field may be identified with a scalar ($\mathbf{B} = \langle 0, 0, B(x, y) \rangle$). Boundary conditions are imposed upon the 0-cochains encoding the fluxes of $\mathbf{B}$ through subdomain boundaries, and the 1-cochains encoding the magnetic potential on the boundary are fixed to a value of zero. For this problem we will consider only the nonlinear generalization of D2, taking $f = 0$ and $g = \alpha$ to obtain a nonlinear material whose permittivity increases with the magnitude of applied magnetic field (Fig. 6). We take $N_{data} = 4$ and sample μ uniformly from $\{1, 10\}$. In the same manner as the Darcy problem, the resulting model is able to recover a range of α 's and the resulting discontinuity in the magnetic field.

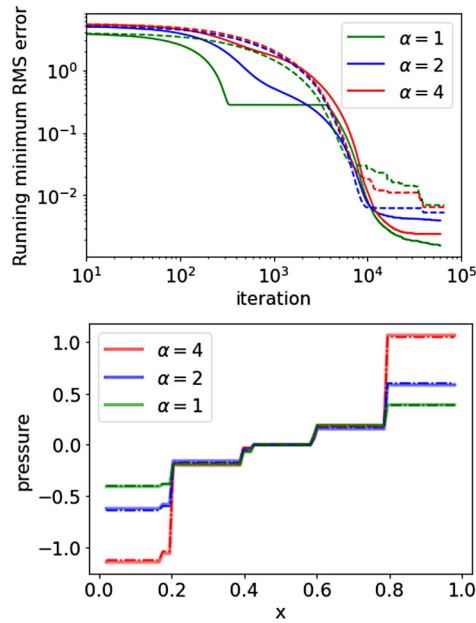


Fig. 5. Study for problem D2. A single nonlinear model is extracted which may be solved for increasing α . *Top*: Running minimum of RMS error over each batch during training for: two layer width 5 elu network (solid lines) and single layer width 10 ReLU network (dashed lines). *Bottom*: Comparison of pressure profile along $y = 0.5$ for increasing α . True solution given by solid line, while learned surrogate is given by dashed line. For both architectures, learning is performed on a chain complex of 49 subdomains using a learning rate of 5×10^{-5} .

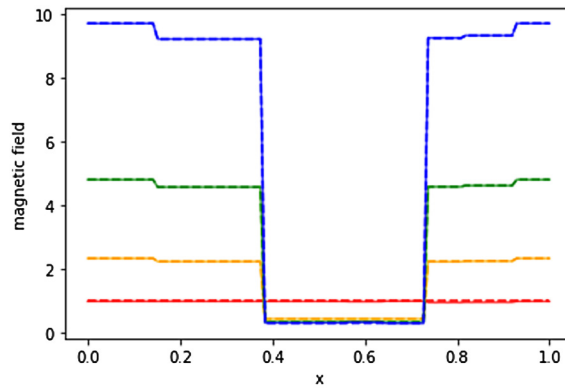


Fig. 6. Profile of magnetic field along $y = 0.5$ line for magnetostatics problem. The trained model reproduces the coarsened moments of the problem and avoids introduction of any spurious oscillations at the material interface, with learned coarse solution (dashed) reproducing moments of exact coarse solution (solid). For this problem 25 subdomains were used with a learning rate of 10^{-4} and a single layer width 10 network with parametric ReLU activation.

6. Conclusions

We have presented a new data-driven exterior calculus which parameterizes the classical graph calculus. The resulting framework allows training of discrete exterior calculus operators which incorporates metric information from data while preserving exact sequence structure and invertibility of Hodge Laplacians - both of which are important for handling a range of physical systems. Analysis of the resulting system shows that when nonlinear perturbations are applied to a class of second-order elliptic problems the system remains well-posed under general conditions, allowing the machine learning of nonlinear elliptic systems. Numerical results demonstrate the practical aspects of the approach.

For the sake of introductory exposition, we have restricted ourselves to elementary elliptic problems in the current work. The DDEC framework however may be applied to a broad range of more sophisticated problems; for example, we are currently using it to discover surrogates for semiconductor physics governed by nonlinear drift-diffusion equations (using [48]), and where the resulting network model can be embedded within circuit simulators such as Xyce [31,38].

Abstractly, the DDEC framework provides a structure-preserving means of parameterizing Dirichlet-to-Neumann maps [62] governing multiscale physics, generalizing previous works restricted to resistor networks [25] while supporting machine

learning tasks. In future work we will provide details regarding how this may be incorporated into a workflow to develop provably stable multiscale models that preserve structure at both fine and coarse scales.

CRediT authorship contribution statement

Nathaniel Trask: Conceptualization, Methodology, Software, Supervision, Writing – original draft. **Andy Huang:** Methodology, Software, Writing – original draft. **Xiaoze Hu:** Formal analysis, Writing – original draft.

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.

Acknowledgement

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

N. Trask has been supported by the U.S. Department of Energy, Office of Advanced Scientific Computing Research under the Collaboratory on Mathematics and Physics-Informed Learning Machines for Multiscale and Multiphysics Problems (PhILMs) project and the U.S. Department of Energy, Office of Advanced Scientific Computing Research under the Early Career Research Program. A. Huang has been supported under the Sandia National Laboratories Laboratory Directed Research and Development (LDRD) program. The work of X. Hu is partially supported by the National Science Foundation under grant DMS-1812503 and CCF-1934553.

SAND Number: SAND2020-14261 O.

References

- [1] James H. Adler, Casey Cavanaugh, Xiaoze Hu, Ludmil T. Zikatanov, A finite-element framework for a mimetic finite-difference discretization of Maxwell's equations, arXiv:2012.03148 [math.NA], December 2020.
- [2] Douglas Arnold, Richard Falk, Ragnar Winther, Finite element exterior calculus: from Hodge theory to numerical stability, *Bull. Am. Math. Soc.* 47 (2) (2010) 281–354.
- [3] Douglas N. Arnold, *Finite Element Exterior Calculus*, SIAM, 2018.
- [4] Douglas N. Arnold, Pavel B. Bochev, Richard B. Lehoucq, Roy A. Nicolaides, Mikhail Shashkov, *Compatible Spatial Discretizations*, vol. 142, Springer Science & Business Media, 2007.
- [5] Douglas N. Arnold, Richard S. Falk, Ragnar Winther, Finite element exterior calculus, homological techniques, and applications, *Acta Numer.* 15 (May 2006) 1–155.
- [6] Nathan Baker, Frank Alexander, Timo Bremer, Aric Hagberg, Yannis Kevrekidis, Habib Najm, Manish Parashar, Abani Patra, James Sethian, Stefan Wild, et al., Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence, Technical report, USDOE Office of Science (SC), Washington, DC (United States), 2019.
- [7] Paul Bamberg, Shlomo Sternberg, *A Course in Mathematics for Students of Physics*, vol. 2, 1991.
- [8] Sergio Barbarossa, Stefania Sardellitti, Topological signal processing over simplicial complexes, arXiv:1907.11577 [eess.SP], March 2020.
- [9] Timothy Barth, On the role of involutions in the discontinuous Galerkin discretization of Maxwell and magnetohydrodynamic systems, in: *Compatible Spatial Discretizations*, Springer, 2006, pp. 69–88.
- [10] Peter Battaglia, Razvan Pascanu, Matthew Lai, Danilo Jimenez Rezende, et al., Interaction networks for learning about objects, relations and physics, *Adv. Neural Inf. Process. Syst.* 29 (2016) 4502–4510.
- [11] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, Giovanni Petri, Networks beyond pairwise interactions: structure and dynamics, *Phys. Rep.* 874 (June 2020) 1–92.
- [12] René Beltman, M.J.H. Anthonissen, Barry Koren, Conservative polytopal mimetic discretization of the incompressible Navier–Stokes equations, *J. Comput. Appl. Math.* 340 (2018) 443–473.
- [13] A. Bloch, Electromechanical analogies and their use for the analysis of mechanical and electromechanical systems, *J. Inst. Electr. Eng., Part I, Gen.* 92 (52) (1945) 157–169.
- [14] P. Bochev, A. Robinson, Matching algorithms with physics: exact sequences of finite element spaces, in: D. Estep, S. Tavener (Eds.), *Collected Lectures on the Preservation of Stability Under Discretization*, SIAM, Philadelphia, 2001.
- [15] Pavel B. Bochev, Jonathan J. Hu, Christopher M. Siefert, Raymond S. Tuminaro, An algebraic multigrid approach based on a compatible gauge reformulation of Maxwell's equations, *SIAM J. Sci. Comput.* 31 (1) (2008) 557–583.
- [16] Pavel B. Bochev, James M. Hyman, Principles of mimetic discretizations of differential operators, in: *Compatible Spatial Discretizations*, Springer, 2006, pp. 89–119.
- [17] Alain Bossavit, *Computational Electromagnetism: Variational Formulations, Complementarity, Edge Elements*, Academic Press, 1998.
- [18] Alain Bossavit, “Stiff” problems in eddy-current theory and the regularization of Maxwell's equations, *IEEE Trans. Magn.* 37 (5) (2001) 3542–3545.
- [19] Peter C. Breedveld, Multibond graph elements in physical systems theory, *J. Franklin Inst.* 319 (1–2) (1985) 1–36.
- [20] Shengze Cai, Zhicheng Wang, Lu Lu, Tamer A. Zaki, George Em Karniadakis, DeepM&Mnet: inferring the electroconvection multiphysics fields based on operator approximation by neural networks, *J. Comput. Phys.* 436 (2021) 110296.
- [21] Michael B. Chang, Tomer Ullman, Antonio Torralba, Joshua B. Tenenbaum, A compositional object-based approach to learning physical dynamics, arXiv preprint, arXiv:1612.00341, 2016.

- [22] Qun Chen, Rong-Huan Fu, Yun-Chao Xu, Electrical circuit analogy for heat transfer analysis and optimization in heat exchanger networks, *Appl. Energy* 139 (2015) 81–92.
- [23] Fan R.K. Chung, Fan Chung Graham, *Spectral Graph Theory*, vol. 92, American Mathematical Soc., 1997.
- [24] Lorenzo Codecasa, Refoundation of the cell method using augmented dual grids, *IEEE Trans. Magn.* 50 (2) (2014) 497–500.
- [25] Edward B. Curtis, James A. Morrow, The Dirichlet to Neumann map for a resistor network, *SIAM J. Appl. Math.* 51 (4) (1991) 1011–1029.
- [26] Constantine M. Dafermos, Quasilinear hyperbolic systems with involutions, *Arch. Ration. Mech. Anal.* 94 (4) (1986) 373–389.
- [27] Mathieu Desbrun, Anil N. Hirani, Melvin Leok, Jerrold E. Marsden, *Discrete Exterior Calculus*, 2005.
- [28] Beno Eckmann, Harmonische funktionen und randwertaufgaben in einem komplex, *Comment. Math. Helv.* 17 (1) (1944) 240–255.
- [29] Floyd A. Firestone, A new analogy between mechanical and electrical systems, *J. Acoust. Soc. Am.* 4 (3) (1933) 249–267.
- [30] Olga Fuks, Hamdi A. Tchelepi, Limitations of physics informed machine learning for nonlinear two-phase transport in porous media, *J. Mach. Learn. Model. Comput.* 1 (1) (2020).
- [31] Xujiao Gao, Andy Huang, Nathaniel Trask, Shahed Reza, Physics-informed graph neural network for circuit compact model development, in: 2020 International Conference on Simulation of Semiconductor Processes and Devices, SISPAD, IEEE, 2020, pp. 359–362.
- [32] David Gilbarg, Neil S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, 2015.
- [33] Marco Gori, Gabriele Monfardini, Franco Scarselli, A new model for learning in graph domains, in: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, vol. 2, 2005, IEEE, 2005, pp. 729–734.
- [34] Will Hamilton, Zitao Ying, Jure Leskovec, Inductive representation learning on large graphs, in: *Advances in Neural Information Processing Systems*, 2017, pp. 1024–1034.
- [35] William L. Hamilton, Rex Ying, Jure Leskovec, Representation learning on graphs: methods and applications, *arXiv preprint*, arXiv:1709.05584, 2017.
- [36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.
- [37] Andy Huang, Nathaniel Trask, Christopher Brissette, Xiaozhe Hu, Greedy Fiedler spectral partitioning for data-driven discrete exterior calculus, in: *Proceedings of the AAAI 2021 Spring Symposium on Combining Artificial Intelligence and Machine Learning with Physical Sciences*, AAAI, 2021.
- [38] S. Hutchinson, E. Keiter, R. Hoekstra, H. Watts, A. Waters, T. Russo, R. Schells, S. Wix, C. Bogdan, The Xyce™ parallel electronic simulator—an overview, in: *Parallel Computing: Advances and Current Issues*, World Scientific, 2002, pp. 165–172.
- [39] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, Yinyu Ye, Statistical ranking and combinatorial Hodge theory, *Math. Program.* 127 (1) (2011) 203–244.
- [40] George Karypis, Vipin Kumar, Metis: a Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices, 1997.
- [41] Diederik P. Kingma, Jimmy Ba, Adam: a method for stochastic optimization, *arXiv preprint*, arXiv:1412.6980, 2014.
- [42] H.E. Koenig, W.A. Blackwell, Linear graph theory—a fundamental engineering discipline, *IRE Trans. Educ.* 3 (2) (1960) 42–49.
- [43] Jasper Kreeft, Artur Palha, Marc Gerritsma, Mimetic framework on curvilinear quadrilaterals of arbitrary order, *arXiv preprint*, arXiv:1111.4304, 2011.
- [44] Isaac E. Lagaris, Aristidis Likas, Dimitrios I. Fotiadis, Artificial neural networks for solving ordinary and partial differential equations, *IEEE Trans. Neural Netw.* 9 (5) (1998) 987–1000.
- [45] Lek-Heng Lim, Hodge Laplacians on graphs, *SIAM Rev.* 62 (3) (January 2020) 685–715.
- [46] S. Maletić, M. Rajković, Combinatorial Laplacian and entropy of simplicial complexes associated with complex networks, *Eur. Phys. J. Spec. Top.* 212 (1) (September 2012) 77–97.
- [47] Abubakr Muhammad, Magnus Egerstedt, Control Using Higher Order Laplacians in Network Topologies, 2006, p. 15.
- [48] Lawrence Musson, Xujiao Gao, Mihai Negoita, Andy Huang, Gary L. Hennigan, Charon: a radiation aware massively parallel tcad modeling code, Technical report, Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), 2018.
- [49] Jean-Claude Nédélec, Mixed finite elements in $\mathbb{R}^3$, *Numer. Math.* 35 (3) (1980) 315–341.
- [50] R.A. Nicolaides, Direct discretization of planar Div-Curl problems, *SIAM J. Numer. Anal.* 29 (1) (February 1992) 32–56.
- [51] Roy A. Nicolaides, Kathryn A. Trapp, Covolume discretization of differential forms, in: *Compatible Spatial Discretizations*, Springer, 2006, pp. 161–171.
- [52] Kwang W. Oh, Kangsun Lee, Byungwook Ahn, Edward P. Furlani, Design of pressure-driven microfluidic networks using electric circuit analogy, *Lab Chip* 12 (3) (2012) 515–545.
- [53] Mario Ohlberger, Stephan Rave, Reduced basis methods: success, limitations and future challenges, *arXiv preprint*, arXiv:1511.02021, 2015.
- [54] Alfio Quarteroni, Andrea Manzoni, Federico Negri, *Reduced Basis Methods for Partial Differential Equations: an Introduction*, vol. 92, Springer, 2015.
- [55] Maziar Raissi, Paris Perdikaris, George E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.* 378 (2019) 686–707.
- [56] Maziar Raissi, Paris Perdikaris, George Em Karniadakis, Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations, *arXiv preprint*, arXiv:1711.10561, 2017.
- [57] Frank Rosenblatt, Principles of neurodynamics. Perceptrons and the theory of brain mechanisms, Technical report, Cornell Aeronautical Lab Inc, Buffalo NY, 1961.
- [58] Kevin Scaman, Aladin Virmaux, Lipschitz regularity of deep neural networks: analysis and efficient estimation, *arXiv:1805.10965v2 [stat.ML]*, October 2019.
- [59] Michael T. Schaub, Austin R. Benson, Paul Horn, Gabor Lippner, Ali Jadbabaie, Random walks on simplicial complexes and the normalized Hodge 1-Laplacian, *SIAM Rev.* 62 (2) (January 2020) 353–391.
- [60] S. Smale, On the mathematical foundations of electrical circuit theory, *J. Differ. Geom.* 7 (1–2) (1972) 193–210.
- [61] Malcolm C. Smith, Synthesis of mechanical networks: the inerter, *IEEE Trans. Autom. Control* 47 (10) (2002) 1648–1662.
- [62] John Sylvester, Gunther Uhlmann, The Dirichlet to Neumann Map and Applications, *Inverse Problems in Partial Differential Equations*, vol. 42, SIAM Publications, Philadelphia, 1990, p. 101.
- [63] Joaquín J. Torres, Ginestra Bianconi, Simplicial complexes: higher-order spectral dimension and dynamics, *J. Phys.: Complex.* 1 (1) (May 2020) 015002.
- [64] Nikolaos Vlassis, Ran Ma, WaiChing Sun, Geometric deep learning for computational mechanics part I: anisotropic hyperelasticity, *arXiv preprint*, arXiv:2001.04292, 2020.
- [65] Sifan Wang, Yujun Teng, Paris Perdikaris, Understanding and mitigating gradient pathologies in physics-informed neural networks, *arXiv preprint*, arXiv:2001.04536, 2020.
- [66] Sifan Wang, Hanwen Wang, Paris Perdikaris, Learning the solution operator of parametric partial differential equations with physics-informed DeepOnets, *arXiv preprint*, arXiv:2103.10974, 2021.
- [67] Sifan Wang, Xinling Yu, Paris Perdikaris, When and why pinns fail to train: a neural tangent kernel perspective, *arXiv preprint*, arXiv:2007.14527, 2020.
- [68] Hermann Weyl, Repartición de corriente en una red conductora, *Rev. Mat. Hisp.-Am.* 5 (1923) 153–164.
- [69] Herbert S. Wilf, Frank Hilary, *Mathematical Aspects of Electrical Network Analysis*, SIAM-AMS Proceedings, vol. 3, American Mathematical Society, 1971.
- [70] Kane Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas Propag.* 14 (3) (1966) 302–307.