Revisiting Conditional Whitney Forms: From Structure Preservation to Physics Recovery

Pavlos Kallinikidis
Department of Electrical and Systems Engineering
University of Pennsylvania
Philadelphia, PA 19104
pgkallin@seas.upenn.edu

Paris Perdikaris
Department of Mechanical Engineering and Applied Mechanics
University of Pennsylvania
Philadelphia, PA 19104
pgp@seas.upenn.edu

George J. Pappas
Department of Electrical and Systems Engineering
University of Pennsylvania
Philadelphia, PA 19104
pappasg@seas.upenn.edu

Abstract

Conditional Whitney forms have recently emerged as a promising framework at the intersection of scientific machine learning and finite element analysis. They offer a solid theoretical foundation for enforcing structure preservation in complex learning settings. However, their use so far has been restricted to tasks where structural constraints can be satisfied with simple, yet inaccurate, physical representations. In this work, we analyze why existing formulations reduce to typical unconstrained reformulations, circumventing actual physics recovery, highlight the necessity of incorporating additive structure pertaining to the governing physics of the system and validate experimentally our theoretical insights.

1 Introduction

Partial Differential Equations (PDEs) form the foundation for describing and modeling complex physical systems across engineering and sciences. Traditional PDE solvers, such as the Finite Element Method (Brenner and Carstensen, 2004), are notoriously computationally intensive, lack generalizability across different parameters and do not apply in settings where the governing physics are only partially known. Motivated by the achievements of machine learning, operator learning (Kovachki et al., 2023; Lu et al., 2021) emerged as a very powerful learning paradigm for learning maps between infinite-dimensional function spaces. Neural operators address the limitations of traditional solvers and show great promise in both predictive performance and efficiency (Bodnar et al., 2025), but integrate physics mainly as inductive bias (Li et al., 2024; Karniadakis et al., 2021) and do not guarantee physical realizability (e.g. conservation laws).

At the intersection of finite element exterior calculus (FEEC) (Arnold, 2018; Trask et al., 2022) and operator learning, conditional Whitney forms (CWF) (Kinch et al., 2025) were proposed as a framework that enables the construction of learnable reduced finite element spaces, in which conservation laws can be strictly enforced. CWF indeed provide a solid theoretical foundation for ensuring conservation and stability. However, the equality-constrained optimization problem, as posed in the original formulation of Kinch et al. (2025) and applied in subsequent experiments, leads to trivial conservation without actual physics recovery. We provide an intuitive, yet rigorous,

Machine Learning and the Physical Sciences Workshop, NeurIPS 2025.

interpretation of this phenomenon and highlight the necessity of extra regularization to mitigate it. Furthermore, we reformulate the optimization problem to enable actual-physics recovery and provide experimental validation for the effectiveness of our approach.

2 Theoretical Formulation

2.1 Preliminaries

Inspired by Whitney forms (Lohi and Kettunen, 2021) in FEEC, Actor et al. (2024) showed that any partition of unity ($\{\psi_i(\cdot)\}$ s.t. $\psi_i(\cdot) \geq 0$ and $\sum_i \psi_i(\cdot) = 1$) can be used as the basis for the construction of a family of finite element spaces that constitute a mimetic discretization (Castillo and Miranda, 2013) of the de Rham complex, essentially providing exact discrete analogues of the standard differential operators (e.g. gradient, divergence) and a generalized Stokes theorem.

In the 2D setting, using the finite element spaces:

$$W^0(\Omega) = \text{span}\{\psi_i^0(\cdot)\}, \quad W^1(\Omega) = \text{span}\{\psi_{ij}^1(\cdot) = \psi_j^0 \nabla \psi_i^0 - \psi_i^0 \nabla \psi_j^0\}$$

to solve a PDE of the following conservative form:

$$\nabla \cdot (\nabla u + N[u; \phi]) = S, \qquad u = 0 \quad on\ \partial\Omega$$

mixed Galerkin form seeks $(u, F) \in W^0(\Omega) \times W^1(\Omega)$ such that for all $(q, v) \in W^0(\Omega) \times W^1(\Omega)$:

$$-(F, \nabla q)_\Omega = (S, q)_\Omega, \qquad (F, v)_\Omega = (\nabla u, v)_\Omega + (N[u; \phi], v)_\Omega.$$

Theorem 3.4 from Kinch et al. (2025) yields an equivalent representation of the above formulation as:

$$\delta_0^\top M_1 \hat{F} = M_0 \hat{S}, \quad M_1 \hat{F} - M_1 \delta_0 \hat{u} - M_1 \hat{N}[\hat{u}; \phi] = 0, \tag{1}$$

where $\hat{(\cdot)}$ denotes a discretized version of a variable, M_0, M_1 denote the mass matrices associated with $W^0(\Omega)$, $W^1(\Omega)$ and δ_0 denotes a generalized incidence matrix between $W^0(\Omega)$ and $W^1(\Omega)$. Using the closeness of partitions of unity under convex combinations, we can map the degrees of freedom of a fixed partition of unity (fine-grained) to the degrees of a reduced one (coarse-grained) through a simple matrix multiplication with a learnable column stochastic matrix:

$$W(z; \theta) = \begin{bmatrix} W^{\text{int}}(z; \theta) & 0 \\ 0 & W^{\text{bnd}} \end{bmatrix}, \tag{2}$$

which is conditioned on a variable input z and maps boundary to boundary and interior to interior degrees of freedom. Taken together, (1) and (2) yield the following PDE-constrained learning problem in the reduced space, where operator N was replaced by a neural network $\mathcal{NN}$:

$$\min_{\hat{u}_i, \theta, \phi} \quad \sum_i \| \sum_j \hat{u}_{i,j} \cdot \psi_j^0(x; z_i, \theta) - u_{i_{\text{target}}}(x) \|_\Omega^2 \qquad i \in \{1, ...N_{\text{samples}}\}, \quad x \in \Omega$$
$$\text{subject to} \quad \mathcal{F}(\hat{u}_i; \phi, \theta, z_i) := \delta_0^\top M_1(z_i; \theta)(\delta_0 \hat{u}_i + \mathcal{NN}[\hat{u}_i; z_i, \phi]) - M_0(z_i; \theta) \hat{S}(z_i) = 0 \tag{3}$$

For more details on the derivations of CWF components, we refer the reader to Kinch et al. (2025); Actor et al. (2024).

2.2 Trivial Conservation

As posed in (3), the learning problem is related to the inverse problem of discovering a conservation equation applied to a reduced space of learnable finite elements. However, conservation is achieved in a trivial way, which, essentially, means that the constraint can be omitted, as it does not restrict the search space of candidate solutions $u_i(x)$. Thus, the learning problem can be solved as an unconstrained regression problem in $W^0(\Omega)$ and the conservation constraint can be satisfied with a standard post-processing step in $W^1(\Omega)$, as it only really imposes a source-flux outflow balance. Formally, the redundancy of the PDE constraint is summarized in the following proposition.

Proposition 1 *Let $u(x) := \sum_i c_i \lambda_i(x)$ s.t. $0 \leq u(x) \leq 1$ and $u(x) = 0$ on $\partial\Omega$, where $\{\lambda_i(x)\}_{i=1}^N$ denotes a fine-grained partition of unity. When function $f(\cdot)$ is not restricted to any specific structure and for any source $s(x)$, there is a trivial reduced partition $\{\psi_j(x)\}$, in which $u(x)$ can be exactly described and satisfies a conservation equation of the form:*

$$\delta_0^\top M_1 \hat{f} = M_0 \hat{s}, \qquad M_1 \hat{f} = M_1 f(\hat{u}), \tag{4}$$

for an infinite number of vectors $\hat{f}$.

Proof: We define $u(x) := \sum_i c_i \lambda_i(x)$, where $\{\lambda_i(x)\}_{i=1}^N$ denotes a fine-grained partition. We can construct a coarse-grained partition $\{\psi_j(x)\}$ as follows.

$$\psi_0(x) := \sum\nolimits_i c_i^{\text{int}} \lambda_i^{\text{int}}(x), \quad \psi_1(x) := \sum\nolimits_i (1 - c_i^{\text{int}}) \lambda_i^{\text{int}}(x), \quad \psi_2(x) := \sum\nolimits_i \lambda_i^{\text{bnd}}(x).$$

The subsequent transformation matrix W of (2) is :

$$W = \begin{bmatrix} \begin{bmatrix} \mathbf{c}_{int}^T \\ (\mathbf{1} - \mathbf{c}_{int})^T \end{bmatrix} & 0 \\ 0 & \mathbf{1}^T \end{bmatrix},$$

we trivially derive $u(x) = 1 \cdot \psi_0(x) + 0 \cdot \psi_1(x) + 0 \cdot \psi_2(x)$ and denote $\hat{u} := [1, 0, 0]^T$. Equipped with the constructed partition of unity, we can express a conservation equation as in (4). Since everything is known and $f(\cdot)$ is not limited to any specific form, we only have to solve the first system in (4). As we have used $\psi_2(x)$ only to apply the homogeneous boundary conditions, the system admits a reduced form as $(\delta_0^\top)_{1:2,:} M_1 \hat{f} = M_0 \hat{s}$, in which $M_0 \hat{s}$ denotes the projection of the source only on the interior nodes. Since δ_0 is an incidence matrix, $(\delta_0^\top)_{1:2,:}$ takes the following form:

$$(\delta_0^\top)_{1:2,:} = \begin{bmatrix} -1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix},$$

which is a full-row rank matrix and M_1 is a symmetric positive-definite matrix as the mass matrix of linearly independent finite elements; thus, the system is underdetermined, admitting an infinite number of solutions and concluding the proof.

Remark 1 *We can drop the assumptions pertaining to $u(x)$ in Proposition 1 by slight modifications to the proof. We can generalize for any u s.t. $u_{min} \leq u(x) \leq u_{max}$ by setting $\hat{u} := [u_1, u_2]^T$, where $u_1 \geq u_{max}$ and $u_2 \leq u_{min}$ and rescaling $\mathbf{c}_{int}$. Furthermore, we can drop the homogeneous boundary conditions with a decomposition of the fine-grained boundary degrees of freedom, identical to the decomposition used for the interior ones.*

2.3 Flux Regularization

Adopting a mixed-space approach to express a data-driven generalized conservation law, CWFs are endowed with tools to treat boundary conditions, source-flux outflow balance and other structure-preserving properties in a precise way. The block-diagonal structure of (2) plays a critical role in the preservation of the structure. It allows for the performance of the lift, which is necessary to impose (in-)homogeneous Dirichlet boundary conditions, and for the quantification of the flux outflow, which is necessary for the source-flux outflow balance claims.

However, the findings of Section 2.2 indicate that the lack of any structure in the reconstructed flux $\hat{f}$ leads to the triviality of (3). In practice, this translates into trivial reduced partitions and flux hallucinations, contradicting the initial selection of a mixed-space approach to describe the physics while preserving a conservation law. Furthermore, experimental evidence shows that, in settings where a machine learning model is used for the inference of $W(z; \theta)$, the predictive performance of CWFs is almost identical to that of the original model, hinting at the findings of Section 2.2.

As we emphasize the importance of flux regularization, we propose the reformulation of (3), adding a flux reconstruction term as follows.

$$\begin{aligned} \min_{\hat{u}_i, \theta, \phi} \quad & \sum_i \| \sum_j \hat{u}_{i,j} \psi_j^0(x; z_i, \theta) - u_{i_{\text{target}}}(x) \|_\Omega^2 + \\ & + \lambda \sum_i \| \sum_{jk} \hat{f}_{i,jk} \psi_{jk}^1(x; z_i, \theta) - f_{i_{\text{target}}}(x) \|_\Omega^2, \\ \text{subject to} \quad & \delta_0^\top M_1(z_i; \theta) \hat{f}_i - M_0(z_i; \theta) \hat{s}(z_i) = 0, \\ & M_1(z_i; \theta)(\hat{f}_i - \delta_0 \hat{u}_i - \mathcal{NN}(\hat{u}_i; \phi, z_i)) = 0. \end{aligned} \tag{5}$$

We adopt the above data-driven approach of flux reconstruction, as learning optimal geometries for the representation of the governing physics is one of the main desiderata of CWFs. To be more precise, the above form of $\hat{f}_i$ contains both a structural assumption and a learnable component, as $\delta_0 \hat{u}_i$ exactly represents the gradient term of flux (diffusion); see Actor et al. (2024), and $\mathcal{NN}(\hat{u}_i; \phi, z_i)$ denotes a learnable nonlinear flux term (convection).

3 Experiments

Experimental Setup We probe the validity of theoretical insights in Section 2.2 and the effectiveness of the proposed reformulation of Section 2.3 in a 2D stationary advection diffusion problem with an inhomogeneous advection field. The PDE for this problem is described below.

$$\nabla \cdot (\nabla u(x,y) + a(x,y)u(x,y)) = S(x,y), \qquad (x,y) \in \Omega := [0,1]^2$$

where $S(x,y) := 20sin(\pi x)cos(\pi y)$. Adopting the idea of the Darcy flow dataset from (Takamoto et al., 2022), we create different realizations of this PDE by dividing Ω into 2 subdomains Ω_1 and Ω_2, as shown in Figure 1, and applying different velocities within these domains. The velocity field $a(x,y)$ is defined as follows.

$$a(x,y) = \begin{cases} -5 \cdot \mathbf{1} & \text{if } (x,y) \in \Omega_1 \\ -0.1 \cdot \mathbf{1} & \text{if } (x,y) \in \Omega_2 \end{cases},$$

We divide a data set of $10,000$ samples into $9,500$ training and 500 testing samples.

Model Setup We use a CViT (Wang et al., 2024) model for the inference of learnable Whitney forms. CViT combines a vision transformer encoder, a grid-based coordinate embedding and a query-wise cross-attention mechanism with state-of-the-art results. We select a $Q1$ basis on a 128×128 quadrilateral mesh as the fine-grained partition of unity and select 8 learnable basis functions for the coarse-grained partition.

Results The results are summarized in Table 1 . Surprisingly, we observe that the regularized method does not harm the reconstruction of the distribution u, while it drastically improves the reconstruction of the flux F (see Figure 2). We also see that the unregularized method learns uniform fluxes, achieving flux balance through adjustments near the boundaries (see Figure 2b), showing that the equality constraint of (3) imposes only an artificial outflow-source balance.

Table 1: Reconstruction MSE for both distribution u and flux F in the unregularized and regularized settings. We use $\lambda = 0.01$ for the regularized method.

Method	Dist. Loss	Flux Loss
Unregularized	$1.5 \cdot 10^{-6}$	$3.0 \cdot 10^{0}$
Regularized	$1.5 \cdot 10^{-6}$	$1.6 \cdot 10^{-3}$

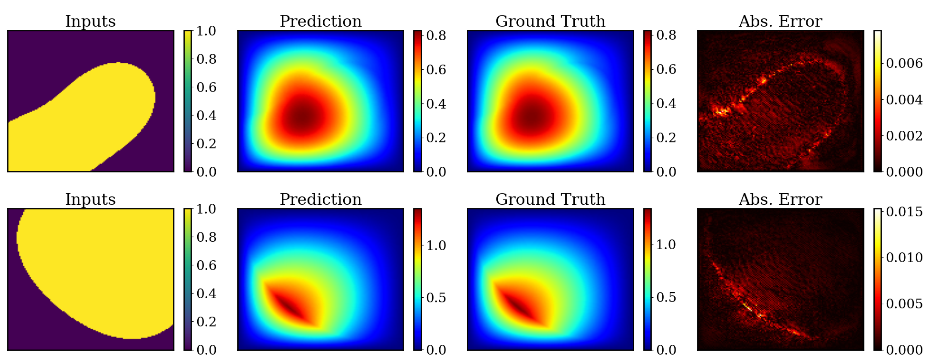
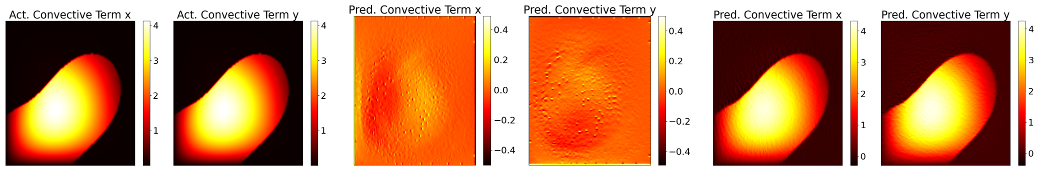


Figure 1: *2D Advection Diffusion.* Representative subdomains, predictions and point-wise errors with the regularized method.

4 Discussion

Summary This work sheds light on the recently proposed framework of conditional Whitney forms (Kinch et al., 2025). Building upon foundational ideas of finite element exterior calculus, CWF



(a) Ground Truth (b) Unregularized Convective Term (c) Regularized Convective Term

Figure 2: *2D Advection Diffusion.* Representative flux reconstruction for both methods. We choose to only show the learnable part of the flux. The regularized method achieves to capture the discontinuity of the convective term.

introduce a learning paradigm that operates in mixed spaces of finite elements. Thus, it is equipped with strong theoretical tools to embed structure-preserving properties, such as conservation laws and boundary conditions into learnable reduced-order models. However, the solution of the learning problem, as originally proposed, leads to trivial representations of both geometry and physics. We essentially show that, when no specific structure of the physics is assumed, the constrained-optimization problem is equivalent to an unconstrained regression task, followed by a standard post-processing step. We also stress the importance of physics regularization and propose a data-driven approach to recover the actual physics. Finally, our theoretical insights are supported by empirical results on a 2D advection diffusion problem.

Further discussion As mixed FEM-ML approaches have received increasing attention, (Rezaei et al., 2024; Bouziani and Boullé, 2024; Ouyang et al., 2025; Farsi et al., 2025) we acknowledge conditional Whitney forms as a very promising direction for the design of reliable machine learning models with guaranteed physical realizability. Yet, we reckon that they have only been employed in learning problems, where the structure-preserving properties can be trivially achieved through the selection of a mixed-space approach and do not mingle with the task of actual physics recovery, as a mixed-space approach would suggest. We hope that our findings will motivate further work on this framework. Future work could provide better insights into both experimental and theoretical aspects, such as performance in highly nonlinear settings, the trade-off between physics recovery and stability and efficient architectural designs.

Acknowledgements

We would like to acknowledge support from the US Department of Energy under the Advanced Scientific Computing Research (ASCR) program (grant DE-SC0024563).

References

Jonas A Actor, Xiaozhe Hu, Andy Huang, Scott A Roberts, and Nathaniel Trask. Data-driven whitney forms for structure-preserving control volume analysis. *Journal of Computational Physics*, 496: 112520, 2024.

Douglas N Arnold. *Finite element exterior calculus*. SIAM, 2018.

Cristian Bodnar, Wessel P Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A Weyn, Haiyu Dong, et al. A foundation model for the earth system. *Nature*, pages 1–8, 2025.

Nacime Bouziani and Nicolas Boullé. Structure-preserving operator learning. *arXiv preprint arXiv:2410.01065*, 2024.

Susanne C Brenner and Carsten Carstensen. Finite element methods. *Encyclopedia of computational mechanics*, 1:73–114, 2004.

José E Castillo and Guillermo F Miranda. *Mimetic discretization methods*. CRC Press, 2013.

Ado Farsi, Nacime Bouziani, and David A Ham. Missing physics discovery through fully differentiable finite element-based machine learning. *arXiv preprint arXiv:2507.15787*, 2025.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Brooks Kinch, Benjamin Shaffer, Elizabeth Armstrong, Michael Meehan, John Hewson, and Nathaniel Trask. Structure-preserving digital twins via conditional neural whitney forms. *arXiv preprint arXiv:2508.06981*, 2025.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024.

Jonni Lohi and Lauri Kettunen. Whitney forms and their extensions. *Journal of Computational and Applied Mathematics*, 393:113520, 2021.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Weihang Ouyang, Yeonjong Shin, Si-Wei Liu, and Lu Lu. Neural-operator element method: Efficient and scalable finite element method enabled by reusable neural operators. *arXiv preprint arXiv:2506.18427*, 2025.

Shahed Rezaei, Reza Najian Asl, Kianoosh Taghikhani, Ahmad Moeineddin, Michael Kaliske, and Markus Apel. Finite operator learning: Bridging neural operators and numerical methods for efficient parametric solution and optimization of pdes. *arXiv preprint arXiv:2407.04157*, 2024.

Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. *Advances in Neural Information Processing Systems*, 35:1596–1611, 2022.

Nathaniel Trask, Andy Huang, and Xiaozhe Hu. Enforcing exact physics in scientific machine learning: a data-driven exterior calculus on graphs. *Journal of Computational Physics*, 456:110969, 2022.

Sifan Wang, Jacob H Seidman, Shyam Sankaran, Hanwen Wang, George J Pappas, and Paris Perdikaris. Cvit: Continuous vision transformer for operator learning. *arXiv preprint arXiv:2405.13998*, 2024.