DISCRETE DE RHAM-HODGE THEORY

By

Rundong Zhao

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Computer Science – Doctor of Philosophy

2020

ABSTRACT

DISCRETE DE RHAM-HODGE THEORY

By

Rundong Zhao

We present a systematic treatment to 3D shape analysis based on the well-established de Rham-Hodge theory in differential geometry and topology. The computational tools we developed are widely applicable to research areas such as computer graphics, computer vision, and computational biology. We extensively tested it in the context of 3D structure analysis of biological macromolecules to demonstrate the efficacy and efficiency of our method in potential applications. Our contributions are summarized in the following aspects.

First, we present a compendium of discrete Hodge decompositions of vector fields, which provides the primary building block of the de Rham-Hodge theory for computations performed on the commonly used tetrahedral meshes embedded in the 3D Euclidean space. Built on the foundations of the Hodge decomposition in the continuous setting, our implementation of a five-component orthogonal decomposition generically splits, for a variety of boundary conditions, any given discrete vector field expressed as discrete differential forms into two potential fields, as well as three additional harmonic components that arise from the topology or boundary of the domain. The resulting decomposition is proper and mimetic, in the sense that the theoretical dualities on the kernel spaces of vector Laplacians valid in the continuous case (including correspondences to cohomology and homology groups) are exactly preserved in the discrete realm.

Second, we present a real-world application of the above computational tool to 3D shape analysis on biological macromolecules. Biological macromolecules have intricate structures that underpin their biological functions. Understanding their structure-function relationships remains a challenge due to their structural complexity and functional variability. We introduce de Rham-Hodge theory as a unified paradigm for analyzing the geometry, topology, flexibility, and Hodge modal analysis of biological macromolecules. Geometric characteristics and topological invariants are

obtained either from the Helmholtz-Hodge decomposition of the scalar, vector and/or tensor fields of a macromolecule or from the spectral analysis of various Laplace-de Rham operators defined on the molecular manifolds. We propose Laplace-de Rham spectrum based models for predicting macromolecular flexibility. We further construct a Laplace-de Rham-Helfrich operator for revealing cryo-EM natural frequencies. Extensive experiments are carried out to demonstrate that the proposed de Rham-Hodge paradigm is one of the most versatile tools for the multiscale modeling and analysis of biological macromolecules and subcellular organelles. The proposed de Rham-Hodge paradigm has potential applications to subcellular organelles and the structure construction from medium or low-resolution cryo-EM maps, and functional predictions from massive biomolecular datasets.

Finally, we extend the above method to an evolutionary de Rham-Hodge method to provide a unified paradigm for the multiscale geometric and topological analysis of evolving manifolds constructed from a filtration, which induces a family of evolutionary de Rham complexes. While the present method can be easily applied to closed manifolds, the emphasis is given to more challenging compact manifolds with 2-manifold boundaries, which require appropriate analysis and treatment of boundary conditions on differential forms to maintain proper topological properties. Three sets of unique evolutionary Hodge Laplacians are proposed to generate three sets of topology-preserving singular spectra, for which the multiplicities of zero eigenvalues correspond to exactly the persistent Betti numbers of dimensions 0, 1, and 2. Additionally, three sets of non-zero eigenvalues further reveal both topological persistence and geometric progression during the manifold evolution. Extensive numerical experiments are carried out to demonstrate the potential of the proposed paradigm for data representation and shape analysis of both point cloud data and density maps.

Our work on the decomposition of vector fields, spectral shape analysis on static shapes, and evolving shapes has already shown its effectiveness in biomolecular applications and will lead to a rich set of features for machine learning-based shape analysis currently under development.

To my dear family

ACKNOWLEDGMENTS

As a very rewarding experience of my life, my Ph.D. journey almost comes to an end. I couldn't imagine that I would have had such an exciting and enjoyable experience without help from many people.

I would like to express my greatest thanks to my Ph.D. advisor, Dr. Yiying Tong. Besides his broad knowledge in the ocean of mathematics and physics, where he guided me to experience the ultimate beauty created by human, Yiying is also very kind-hearted, who truly cares about my psychological construction to peacefully face the world full of problems. I still remember how many times I almost gave up during a deadline, but he just pulled me back with his words of encouragement. Instead of an academic mentor, Yiying is truly more like a life mentor of mine, whose perspectives have a profound influence to potentially help me grow up mentally.

I would like to thank my co-advisor, Dr. Guo-Wei Wei, who not only provided a lot of interdisciplinary advice on how to proceed with my research but also gave me insightful guidance on how to make a life decision after finishing my Ph.D. study.

I would like to thank my collaborator Dr. Mathieu Desbrun, who gave me the opportunity to visit Caltech for half a year, where I had my first-hand experience with some of the smartest people on Earth. Besides the encouragement and advice from him to work on a fantastic research topic, I could feel his strong sense of humor and enthusiasm, when I realized that doing research is such an enjoyable experience.

I also would like to thank Dr. Xiaoming Liu, who provided valuable ideas on combining computer graphics and computer vision researches, and Dr. Benjamin Schmidt, who helped me have the first taste of beautiful geometry and topology, to be on my committee.

It is with great pleasure to work with many excellent researchers as well. I would like to thank Dr. Jin Huang for bringing me into the fantastic world of computer graphics; Dr. Beibei Liu and Dr. Xiaojun Wang for the help in getting started with my graphics research; Dr. Zixuan Cang, Dr. Jiahui Chen, and Dr. Menglun Wang for the help on providing research resources and assistance

on computational biology; Hayam Abdelrahman, Ze Zhang and Emily Ribando-Gros for sharing interesting research ideas.

Last but not the least, I would like to thank my parents, who unconditionally supported me and my decision to accept the challenge of pursuing a Ph.D. degree, such that I didn't need to worry about anything other than working on interesting and challenging researches.

I'll never forget such a great experience in pursuing advances in knowledge with a lot of nice people accompanying me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Understanding 3D shape plays a crucial role in research fields such as computer graphics, computer vision, and computational biology. This dissertation tries to approach this problem through a well-established mathematical theory, called de Rham - Hodge theory, to establish a systematic method to extract features and therefore understand 3D shapes. Applications on understanding 3D biological macromolecule structures in either topology and geometry are illustrated in detail.

The structure of the dissertation is as follows. In Chapter 2, computational foundations and discretization strategy are established through a famous computational theory called Discrete Exterior Calculus (DEC) exerted on a discrete approximation of smooth manifolds, such as triangle/tetrahedral meshes. The key contribution is figuring out the Laplacian operators with assigned boundary conditions, which finally gives the full discrete 3D Hodge decompositions as the building block of discrete de Rham - Hodge theory. In Chapter 3, The Laplacian operators, which contain rich information of the 3D shapes due to the integration of boundary conditions, are analyzed to unveil features of the 3D shape in the form of the spectrum, where topological and geometric features can be extracted straightforwardly. In Chapter 4, inspired by the theory of persistent homology, we introduce a family of manifolds into our framework, extracting additional 3D features in an evolutionary manner. And finally, in Chapter 5, we give a brief summary of our discrete de Rham-Hodge Theory and its applications.

1.2 3D Hodge Decompositions of Edge- and Face-based Vector Fields

1.2.1 Background

The existence of orthogonal decompositions of a given vector field into gradient and curl terms (that can be integrated into potentials) along with non-integrable parts (that are due to the topology of the domain) is a fundamental property leveraged in a variety of static and dynamical problems — for instance, fluid simulation to enforce incompressibility. The mathematical foundations behind such decompositions were developed using the theory of differential forms for any finite-dimensional compact manifold without boundary early on [10], but were fully extended to manifolds with boundaries much more recently [11].

1.2.2 Challenges

The analysis and processing of vector fields over *surfaces* have received plenty of attention in recent years. Consequently, the resulting computational tools needed to achieve a Hodge decomposition have been well documented and tested on various applications; see, e.g., recent surveys on surface vector field analysis [12, 13]. For the case of vector fields over 3D bounded domains, discussions about the Hodge decomposition are significantly scarcer: while the usefulness of the Hodge decomposition is as prevalent as in 2D, the existing literature lacks a rigorous computational treatment of the full-blown decomposition over 3D domains of arbitrary topology. Our paper fills this void by offering both the theoretical foundations and a practical linear-algebra based implementation of a five-term Hodge decomposition of vector fields expressed as discrete forms for the most common boundary conditions used in computational science.

A variety of books present detailed expositions of the Hodge decomposition from a mathematical perspective (see [14, 15] for two examples using a formulation based on differential forms), but provide no hints on computational approaches to implementing a discrete decomposition in the case of finite-dimensional vector field representations. Even more applied treatments (such as [16] which discusses the case of complicated topology at length) are often based on a Biot-Savart construction
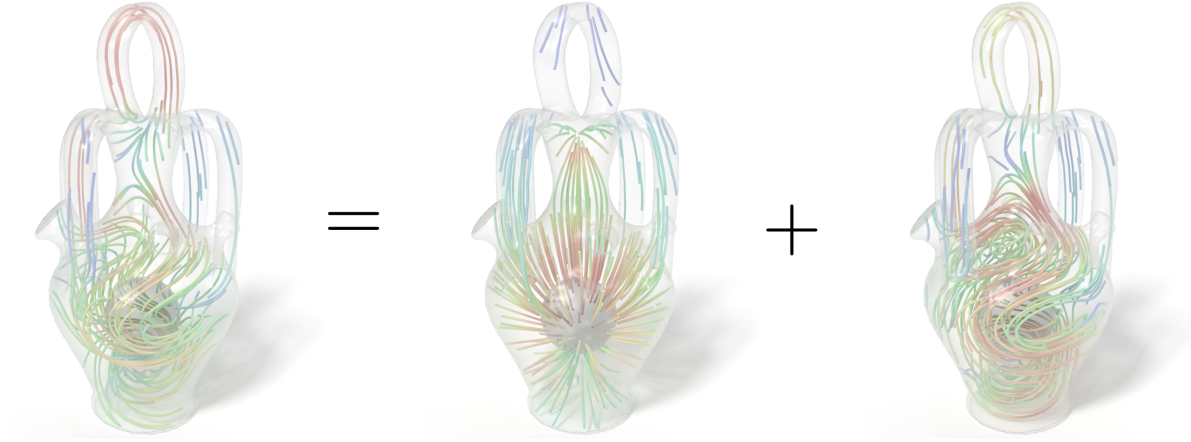
Figure 1.1: **Helmholtz decomposition.** Vector field in a vase with a spherical cavity decomposed into a gradient and a curl field, but with a nonzero L_2 inner product between these two resulting components.

relying on volume integrals to prove the existence and uniqueness of the decomposition, but leave the computational aspects to realize such a decomposition mostly unaddressed. For the simpler case of a two-component decomposition (known as the Helmholtz decomposition), a number of papers describe how to compute the scalar and potential vector potentials [17], but no mention is made of the validity of the implied discretization of the cohomology and whether its dimensionality matches the continuous case based on the discrete choices of divergence, curl and gradient operators. Yet, the numerical issues generated by a failure to capture the proper cohomologies are well documented by now – see, e.g., the spurious (i.e., aphysical) modes in computational electromagnetism [18], or the typical checkerboard patterns in Poisson solves.

The most common use of a 3D vector field decomposition is arguably in incompressible fluid simulation; however, not all components are needed in this context since a simple pressure projection is typically used to remove all divergence [19, 20]. Most approaches in fluid dynamics discuss the case of ball-like topology, with the exception of methods using vorticity to reconstruct the velocity field, e.g., [21, 22], which contain discussions on the treatment of domains with nonzero genus. Even in these cases, the decomposition is not comprehensive due to the specificities of typical boundary conditions in fluid animation. Finally, [23] provides a thorough survey of recent

progress on 2D and 3D vector field decompositions in graphics and visualization, but also laments the lack of computational methods providing a five-component decomposition with proper discrete cohomology: until recently, 3D decompositions were mostly achieved for piecewise-constant vector fields on tetrahedral meshes as in [24], extending the 2D variational approach of [25]); however, this decomposition overly inflates the size of the space of harmonic fields [13], leading to the wrong dimensionality of the cohomology. A cohomologically-correct five-component decomposition was very recently introduced for 2D surfaces in [26, 27] under the name of "boundary-aware" Hodge decomposition; a corresponding 3D five-component decomposition was proposed in [3], extending the 2D decomposition from [28] and 3D decomposition from [29]. However, it assumes piecewise-constant vector fields, making its extension to higher order basis functions unclear and its ability to handle mixed types of boundary conditions (common in applications like fluid simulation) limited. Moreover, gauge conditions were not discussed, thus preventing efficient implementations purely based on symmetric positive definite matrices in 3D. Finally, an alternative way for visualization and analysis of 2D or 3D vector fields in bounded domains is to create a natural boundary condition for the gradient and curl components as suggested in [30]. However, the lack of orthogonality between the resulting components limits its use in other applications.

1.2.3 Contributions

We describe both the mathematical formulations and practical computations of a five-component decomposition of vector fields in $\mathbb{R}^3$. We begin with a review of Hodge theory expressed using differential forms, then provide its discretization using Discrete Exterior Calculus (DEC [31]). We offer:

- a practical procedure for five-component decompositions based on discrete vector fields provided as discrete 1-forms (edge values) or 2-forms (face values) on a tetrahedral mesh;

- a thorough discussion on the enforcement of boundary conditions using DEC discretization to ensure the correct cohomology (with the proper dimensionality of the topology-induced non-

integrable parts of the vector field);

- and an effective method for solving the relevant Poisson equations with rank deficiency using only symmetric matrices.

Our exposition aims at serving both practitioners (as we spell out all the matrices involved and numerical treatments of their rank deficiency) and theoretically-minded researchers (as we carefully explain how the discrete setting mimics both boundary conditions and cohomologies).

1.3   de Rham-Hodge Analysis and Modeling of Biomolecules

1.3.1   Background

One of the most amazing aspects of biological science is the intrinsic structural complexity of biological macromolecules and its associated functions. The understanding of how changes in macromolecular structural complexity alter their function remains one of the most challenging issues in biophysics, biochemistry, structural biology, and molecular biology. This understanding depends crucially on our ability to model three-dimensional (3D) macromolecular shapes from original experimental data and to extract geometric and topological information from the architecture of molecular structures. Very often, macromolecular functions depend not only on native structures but also on nascent, denatured or unfolded states. As a result, understanding the structural instability, flexibility, and collective motion of macromolecules is of vital importance. Structural bioinformatics searches for patterns among diverse geometric, topological, instability and dynamic features to deduce macromolecular function. Therefore, the development of efficient and versatile computational tools for extracting macromolecular geometric characteristics, topological invariants, instability spots, flexibility traits, and mode analysis is a key to infer their functions, such as binding affinity, folding, folding stability change upon mutation, reactivity, catalyst efficiency, allosteric effects, etc.

Geometric modeling and characterization of macromolecular 3D shapes have been an active research topic for many decades. Surface models not only provide a visual basis for understanding

macromolecular 3D shapes and but also bridge the gap between experimental data and theoretical modeling, such as generalized Born and Poisson-Boltzmann models for biomolecular electrostatics [32, 33]. A space-filling model with van der Waals spheres was introduced by Corey, Pauling, and Koltun [34]. Solvent accessible surface (SAS) and solvent excluded surface were proposed [35, 36] to provide a more elaborate 3D description of biomolecular structures. However, these surface definitions admit geometric singularities, which lead to computational instability. Smooth surfaces, including Gaussian surfaces [37, 38, 39, 40, 41], skinning surfaces [42], minimal molecular surface [43] and flexibility-rigidity index (FRI) surfaces [44, 45], were constructed to mitigate the computational difficulty.

Another important property of macromolecules is their structural instability or flexibility. Such property measures macromolecular intrinsic ability to respond to external stimuli. Flexibility is known to be crucial for biomolecular binding, reactivity, allosteric signaling, and order-disorder transition [46]. It is typically studied by standard techniques, such as normal mode analysis (NMA) [47, 48, 49, 50], Gaussian network model (GNM) [51] and anisotropic network model (ANM) [52]. These methods have the computational complexity of $\mathcal{O}(N^3)$, with N being the number of unknowns. As a geometric graph-based method, FRI was introduced to reduce the computational complexity and improve the accuracy of GNM [44, 1]. NMA and ANM offer the collective motions, which, as manifested in normal modes, may facilitate the functionally important conformational variations of macromolecules.

The aforementioned Gaussian surface or FRI surface defines a manifold structure embedded in 3D, which makes the analysis of geometry and topology accessible by differential geometry and algebraic topology. Recently, differential geometry has been introduced to understand macromolecular structure and function [53, 54]. In general, protein surface has many atomic scale concave and convex regions which can be easily characterized by Gaussian curvature and/or mean curvature. In particular, the concave regions of a protein surface at the scale of a few residues are potential ligand binding pockets. Differential geometry-based algorithms in both Lagrangian and Cartesian formulations have been developed to generate multiscale representations of biomolecules. Recently, a

6

geometric flow based algorithm has been proposed to detect protein binding pockets [55]. Morse functions and Reeb graphs are employed to characterize the hierarchical pocket and sub-pocket structure [55, 56].

More recently, persistent homology [57, 58], a new branch of algebraic topology, has become a popular approach for the topological simplification of macromolecular structural complexity [59, 60, 61]. Topological invariants are macromolecular connected components, rings, and cavities. Topological analysis is able to unveil the topology-function relationship, such as ion channel open/ close, ligand binding/disassociation, and protein folding/unfolding. However, persistent homology neglects chemical and biological information during its geometric abstraction. Element-specific persistent homology has been introduced to retain crucial chemical and biological information during the topological simplification [62]. It has been integrated with deep learning to predict various biomolecular properties, including protein-ligand binding affinities and protein folding stability changes upon mutation [63].

1.3.2 Challenges

It is interesting to note that most current theoretical models for macromolecules are built from classical mechanics, namely, computational electromagnetics, fluid mechanics, elasticity theory, and molecular mechanics based on Newton's law. These approaches lead to multivalued scalar, vector and tensor fields, such as macromolecular electrostatic potential, ion channel flow, protein anisotropic motion, and molecular dynamics trajectories. Biomolecular cryogenic electron microscopy (cryo-EM) maps are also scalar fields. Mathematically, macromolecular multivalued scalar, vector, and tensor fields contain rich geometric, topological, stability, flexibility and Hodge mode information that can be analyzed to reveal molecular function. Unfortunately, unified geometric and topological analysis of macromolecular multivalued fields remains scarce. It is more challenging to establish a unified mathematical framework to further analyze macromolecular flexibility and Hodge modes. There is a pressing need to develop a unified theory for analyzing the geometry, topology, flexibility, and collective motion of macromolecules so that many existing

7

methods can be calibrated to better uncover macromolecular function, dynamics, and transport.

1.3.3 Contributions

The objective of the present work is to construct a unified theoretical paradigm for analyzing the geometry, topology, flexibility and Hodge mode of macromolecules in order to reveal their function, dynamics, and transport. To this end, we introduce de Rham-Hodge theory for the modeling and analysis of macromolecules. De Rham-Hodge theory is a cornerstone of contemporary differential geometry, algebraic topology, geometric algebra, and spectral geometry [64, 65, 66]. It provides not only the Helmholtz-Hodge decomposition to uncover the interplay between geometry and topology and the conservation of certain physical observables, but also the spectral representation of the underlying multivalued fields, which further unveils the geometry and topology. Specifically, as a ubiquitous computational tool, the Helmholtz-Hodge decomposition of various vector fields, such as electromagnetic fields [67], velocity fields [68], and deformation fields [52], can reveal their underlying geometric and topological features (see a survey [23]). Additionally, de Rham-Hodge theory interconnects classic differential geometry, algebraic topology and partial differential equation (PDE) and provides a high-level representation of vector calculus and the conservation law in physics. Finally, the spectra of Laplace-de Rham operators in various differential forms also contain the underlying geometric and topological information and provides a starting point for the theoretical modeling of macromolecular flexibility and Hodge modes. The corresponding computational tool is discrete exterior calculus (DEC) [69, 70, 71, 72]. Lim discussed discrete Hodge Laplacians on graphs, which might not recover all the properties of the Laplace-de Rham operator [73].

De Rham-Hodge theory has had great success in theoretical physics, such as electrodynamics, gauge theory, quantum field theory, quantum gravity, etc. However, this versatile mathematical tool has not been applied to biological macromolecules, to the best of our knowledge. The proposed de Rham-Hodge framework seamlessly unifies previously developed differential geometry, algebraic topology, spectral graph theory, and PDE based approaches for biological macromolecules [74]. Our specific contributions are summarized as follows

- We provide a spectral analysis tool based on de Rham-Hodge theory to extract geometric and topological features of macromolecules. In addition to the traditional spectra of scalar Hodge Laplacians, we enrich the spectra by using vector Hodge Laplacians with various boundary conditions.

- We construct a de Rham-Hodge theory-based analysis tool for the orthogonal decomposition of various vector fields, such as electric field, magnetic field, velocity field from molecular dynamics and displacement field, associated with macromolecular modeling, analysis, and computation.

- We propose a novel multiscale flexibility model based on the spectra of various Laplace-de Rham operators. This new method is applied to the Debye-Waller factor prediction of a set of 364 proteins [1]. By comparison with experimental data, we show that our new model outperforms GNM, the standard bearer in the field [51, 1].

- We introduce a multiscale Hodge mode model by constraining a vector Laplace-de Rham operator with a Helfrich curvature potential. The resulting Laplace-de Rham-Helfrich operator is applied to analyzing the Hodge modes of cryo-EM data. Unlike previous normal mode analysis which assumes harmonic potential around the equilibrium, our approach allows unharmonic motions far from the equilibrium. The multi-resolution nature of the present method makes it a desirable tool for the multiscale analysis of macromolecules, protein complexes, subcellular structures, and cellular motions.

- We demonstrate electrostatic field analysis based on Hodge decomposition and eigenfield analysis. The eigenfield analysis is applied on the reaction potential calculated by solving the Poisson-Boltzmann equation. We show that local dominant Hodge eigenfields exist for electrostatic analysis.

Our results are twofold: we first describe our contribution to computational tools for Laplace-de Rham operators based on the simplicial tessellation of volumes bounded by biomolecular sur-

9

faces, then we present the modeling and analysis of de Rham-Hodge theory for biological macro-molecules.

1.4 Evolutionary de Rham-Hodge Method

1.4.1 Background

The de Rham-Hodge theory reveals that the cohomology of an oriented closed Riemannian manifold can be represented by harmonic forms. It also holds for an oriented compact Riemannian manifold with boundary by forcing certain boundary conditions, such as absolute and relative cohomology [75]. This theory has been proved to be fundamentally important throughout algebraic geometry. It studies differential geometry and algebraic topology with partial differential equations (PDEs). The understanding of the de Rham-Hodge theory requires a variety of contemporary mathematical techniques including differential geometry, algebraic geometry, elliptic PDE, abstract algebra, topology, et al.

The de Rham-Hodge theory has a wide range of applications, including not only mathematics, but also graphics/visualization [76, 72], physics/fluids [77], vision/robotics [78, 79] and astrophysics/geophysics [80, 81]. Among all these applications, most of them rely upon the Hodge theory result, i.e., the Helmholtz-Hodge decomposition. It is one of the fundamental theorems in dynamical problems, describing a vector field into the gradient and curl components.

Due to the orthogonal decomposition, the analysis of vector fields becomes easier since certain properties such as incompressibility and vorticity of fluid dynamics can be studied on the orthogonal subspace. Such an orthogonal decomposition was first applied on a finite-dimensional compact manifold without boundary [64], and then was developed for manifolds with boundaries [11]. Pushed by the visualization community, the implementation of orthogonal decomposition integrates a variety of boundary conditions with discrete vector fields expressed as discrete differential forms into two potential fields and harmonic fields [72]. The boundary conditions of the decomposition preserve orthogonality. The duality revealed by tangential and normal boundary conditions provides compact spectral representations of the Laplace operators in the de Rham-Hodge theory. The

spectra of de Rham-Laplace operators provide a quantitative approach to understanding topological spaces and geometry characteristics of manifolds and have been applied to biomolecular modeling and analysis [82]. The development of discrete exterior calculus (DEC) is the driving force for de Rham-Hodge theory analysis and application [71, 83].

With the advancements in data development and computational software, persistent homology has been promoted as a new multiscale approach for data analysis [84, 85]. The traditional topological approaches describe the topology of a given object without invoking the metric or coordinate representations. Whereas, persistent homology bridges algebraic topology and multiscale analysis. The essential difference is that persistent homology analyzes the persistence of the topological space through a filtration process, which is a family of simplicial complexes under a series of inclusion maps. Therefore a series of complexes is constructed based on filtration, which captures topological features changing over a range of spatial scales and reveals the features' topological persistence. In some sense, persistent homology can embed geometric information to topological invariants such that "birth" and "death" of connected components, rings, or cavities can be monitored by topological measurements during geometric scale changes. The original idea of varying scales was introduced by Frosini and Landi [86] and by Robins in 1990s [87]. Edelsbrunner et al. formulated the persistent homology and developed the first efficient computational algorithm [88]. Zomorodian and Carlsson generalized the mathematical theory [84]. Persistent homology has stimulated much theoretical development[89, 85, 90, 91, 92, 93]. Among them, persistent spectral graph generates both topological persistence and spectral analysis [93]. Persistent homology has been applied to a variety of fields, including image analysis [94, 95, 96, 97], image retrieval [98], chaotic dynamics verification [99, 100], sensor network [101], complex network [102, 103], data analysis [104, 105], computer vision [96], shape recognition [106], and computational biology [59, 107, 108, 109, 110].

One of the first integrations of persistent homology and machine learning was developed for protein classification in 2015 [111]. Since then, persistent homology has been utilized as one of the most successful methods for the multiscale representation of complex biomolecular data [112, 113, 114]. Two other multiscale representations of complex biomolecular data have also been

proposed and found tremendous success in worldwide competitions in computer-aided drug design [115, 116]. One of them is based on multiscale graphs [117], or more precisely, multiscale weighted colored graphs [118]. Eigenvalues of the graph Laplacians of multiscale weighted colored graphs were shown to provide some of the most powerful representations of protein-ligand binding interactions [119]. The other representation utilizes the curvatures computed from multiscale interactive molecular manifolds [120]. The multiscale shape analysis offers an efficient means to discriminate similar geometries. A common feature that is crucial to the success of the aforementioned three mathematical data representations is that they either create a family of multiscale topological spaces, or generate a family of multiscale graphs, or construct a family of manifolds, indicating the importance of the multiscale analysis in the representation of complex data with intricate internal structures.

1.4.2 Challenges

In the last few decades, geometric analysis has made great progress in understanding shapes that evolve in time. Geometric flows [121] or geometric evolution equations have been extensively studied in mathematics [122, 123, 124], and many processes by which a curve or surface can evolve, such as the Gauss curvature flow and the mean curvature flow. Numerical techniques based on level sets were devised by Osher and Sethian [125] and have been extended and applied by many others in geometric flow analysis [126, 127, 128]. More recently, as the progress in contemporary life sciences, a large number of problems of unveiling the structure-function relationship of biomolecules and understanding of biomolecular systems, requires multiscale geometric modeling and analysis [129, 126, 130]. However, compared with the investigations on curves and surfaces, a small amount of geometric explorations focuses on the evolution of compact manifolds specific to $\mathbb{R}^3$ due to the difficulty of computations. Additionally, it is rare to resolve topology from a nonlinear geometric PDE. Using a minimal molecular surface model [129], Wang and Wei studied the topological persistence via the evolutionary profiles of the Laplace-Beltrami flow [131]. As a result, features of topological invariants are computed from the geometric PDE based filtration. In fact, there has

been much effort in pure mathematics to understand the convergence of Riemannian manifolds in terms of sequences of submanifolds in metric spaces. However, the involved Gromov-Hausdorff distance can be computationally very difficult.

1.4.3 Contributions

Inspired by the aforementioned ideas, we introduce an evolutionary de Rham-Hodge method for data representation. The present evolutionary de Rham-Hodge method is developed by integrating differential geometry, algebraic topology, and multiscale analysis. It is noted that the fusion of algebraic topology and multiscale analysis leads to persistent homology, the combination of differential geometry and multiscale analysis renders manifold convergence [132], while the union of differential geometry and algebraic topology results in the de Rham-Hodge theory. For a given dataset, using the evolutionary filtration developed in early work [131], we construct a sequence of evolving manifolds that lead to a geometry-embedded filtration under inclusion maps. The evolutionary de Rham-Hodge method is established on this sequence of manifolds. In general, the evolution of the manifolds can be either topological persistence which involves topological changes or geometric progression which does not involve topological changes. We are interested in both the data analysis by evolutionary Hodge decompositions associated with various differential forms and the data representations via the evolutionary spectra of de Rham Laplace operators defined on the sequence of manifolds. The evolutionary spectra reveal both the topological invariants and the geometric shapes of evolving manifolds. Such an evolutionary spectral analysis has great potential to "hear the shape of a drum".

In this work, we concern both close 2-manifolds and compact manifolds in $\mathbb{R}^3$ with boundaries, which require the enforcement of appropriate boundary conditions on differential forms to ensure topological properties. Much effort has been given to the understanding and implementation of appropriate boundary conditions for the evolutionary de Rham-Hodge method, which results in three sets of unique evolutionary Hodge Laplacians. The multiplicities of the zero eigenvalues of these evolutionary Hodge Laplacians provide the 0th, 1st, and 2nd persistent Betti numbers. Their

non-zero eigenvalues further portray the geometric shape and topological characteristics of data.

CHAPTER 2

3D HODGE DECOMPOSITIONS OF EDGE- AND FACE-BASED VECTOR FIELDS

2.1   Math Background

Before delving into the actual discrete notion of Hodge decomposition, we present some background on the continuous notions that we wish to numerically emulate.
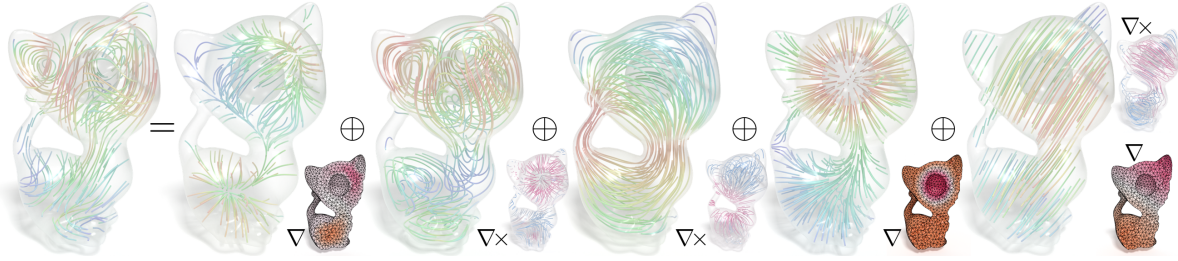
Figure 2.1: **Five-Component Vector Field Decomposition.** On a tetrahedral mesh of the kitten with a spherical cavity, a vector field is decomposed into a gradient field with zero potential on the boundary, a curl field with its vector potential orthogonal to the boundary, a pair of tangential and normal harmonic fields, and a harmonic field that is both a gradient and a curl field. Potential fields are shown in the corners of their corresponding components.

2.1.1   Helmholtz decomposition

In a bounded domain embedded in 3D Euclidean space, any vector field $\mathbf{v}$ can be expressed as the sum of the gradient of a scalar potential f and the curl of a vector potential $\mathbf{u}$, a two-component decomposition known as the *Helmholtz decomposition*, i.e.,

$$\mathbf{v} = \nabla f + \nabla \times \mathbf{u}.$$

The fields f and $\mathbf{u}$ can be constructed, for instance, using Green's functions of the Laplacian operator through volume and boundary surface integrals. However, this decomposition is, in general, not an *orthogonal* decomposition, i.e., the L_2-inner product between ∇f and $\nabla \times \mathbf{u}$ is not necessarily 0, and is not even unique without imposing proper boundary conditions (see Fig. 1.1). In

	order 0	order 1	order 2	order 3
form	f^0 f	$\mathbf{v}^1(\mathbf{a})$ $\mathbf{v}\cdot\mathbf{a}$	$\mathbf{v}^2(\mathbf{a},\mathbf{b})$ $\mathbf{v}\cdot(\mathbf{a}\times\mathbf{b})$	$f^3(\mathbf{a},\mathbf{b},\mathbf{c})$ $f[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]$
d $[dd=0]$	df^0 $(\nabla f)^1$	$d\mathbf{v}^1$ $(\nabla\times\mathbf{v})^2$	$d\mathbf{v}^2$ $(\nabla\cdot\mathbf{v})^3$	df^3 0
$\star$ $[\star\star=1]$	$\star f^0$ f^3	$\star\mathbf{v}^1$ $\mathbf{v}^2$	$\star\mathbf{v}^2$ $\mathbf{v}^1$	$\star f^3$ f^0
δ $[\delta\delta=0]$	δf^0 0	$\delta\mathbf{v}^1$ $(-\nabla\cdot\mathbf{v})^0$	$\delta\mathbf{v}^2$ $(\nabla\times\mathbf{v})^1$	δf^3 $(-\nabla f)^2$
$\wedge$ [(anti-)commute]	$f^0\wedge g^0$ $(fg)^0$	$f^0\wedge\mathbf{v}^1$ $(f\mathbf{v})^1$	$f^0\wedge\mathbf{v}^2, \mathbf{v}^1\wedge\mathbf{u}^1$ $(f\mathbf{v})^2, (\mathbf{v}\times\mathbf{u})^2$	$f^0\wedge g^3, \mathbf{v}^1\wedge\mathbf{u}^2$ $(fg)^3, (\mathbf{v}\cdot\mathbf{u})^3$

Figure 2.2: **Exterior vs. traditional calculus**: odd rows show exterior calculus notations, and even rows give their more conventional expressions in 3D.

practical problems, boundary conditions are often crucial, e.g., the slip wall (tangential) boundary conditions for fluid simulation, and the normal boundary condition for the electric field at an ideal conductor boundary. As the orthogonality between the gradient and curl parts are highly relevant for efficiency and accuracy in computational applications, a more general decomposition, called the *Helmholtz-Hodge decomposition* is called for; but it now involves components that are no longer integrable. Yet, these non-integrable parts are finite-dimensional and directly related to the topology of the domain through correspondences established by Poincaré, de Rham, and Hodge, as we briefly discuss next before spelling out the five-component decomposition.

2.1.2 Vector fields through differential forms

Hodge theory is more conveniently and concisely described by differential k-forms and the exterior calculus based on these forms. While this notational formalism is more involved than the traditional vector notation, both are strictly equivalent, and exterior calculus more clearly identifies topological *vs.* metric operators; the reader unfamiliar with this equivalence is referred to tutorials [31, 133]; we also provide a lookup table to peruse in Fig. 2.2 that summarize relevant equivalences (specific to 3D).

Forms as scalar or vector fields A k-form ω^k is a pointwise multilinear mapping from k vectors to a scalar such that if two input vectors are swapped, the sign of the output is switched. Thus a 0- or 3-form in our $\mathbb{R}^3$ setting has only one degree of freedom (DoF) per point, and can be simply identified with a single-component field f (since they represent, respectively, a scalar field and a density field), while a 1- or 2-form has three DoF per point, and can be identified with a vector field $\mathbf{v}$. We will use f^0, f^3, $\mathbf{v}^1$, and $\mathbf{v}^2$ to denote f seen as a 0- or 3-form and $\mathbf{v}$ as a 1- or 2-form respectively.[1] How the DoFs are used in the antisymmetric linear map is listed in Fig. 2.2.

Operators on forms Due to their antisymmetric tensorial nature, k-forms can be integrated on any k-submanifold. Additionally, the exterior derivative (or *differential*) d^k is an antisymmetriza-tion of the partial derivatives of a k-form to produce a $(k{+}1)$-form that satisfies the Stokes' theorem over any $(k{+}1)$-submanifold $\mathcal{R}$ in M:

$$\int_{\mathcal{R}} d^k \omega^k = \int_{\partial \mathcal{R}} \omega^k.$$

Consequently, one can readily verify that $d^k d^{k-1} = 0$. Depending on the form degree that it is applied to, it encompasses the classical gradient, curl and divergence operators in one consistent type. In the remainder of this paper, we will often omit the superscript of d since it can be directly inferred from the type of its operand. Note that we conventionally call a form *closed* if its differential is zero. Additionally, the wedge product $\wedge$ is defined as an antisymmetrization of the tensor product of two mappings (a p-form and a q-form) to produce a $(p{+}q)$-form: for $p{+}q > 3$, it is 0 since no degrees of freedom are left after antisymmetrization. Finally, the Hodge k-star $\star^k$ (or Hodge dual; we will also omit its superscript at times since the operand disambiguates its identity) is an isomorphism from a k-form ω^k to a $(3{-}k)$-form $(\star\omega)^{3-k}$ by treating them as the same DoF used in mapping $3{-}k$ vectors (instead of k vectors in the same Euclidean coordinate system) to a scalar. Combinations of the basic operators can be constructed. For instance, $\delta^k = (-1)^k \star^{k+1} d^k \star^k$

---

[1]This notation will allow us to keep the "musical" isomorphisms $\sharp$ and $\flat$ hidden to simplify expressions.
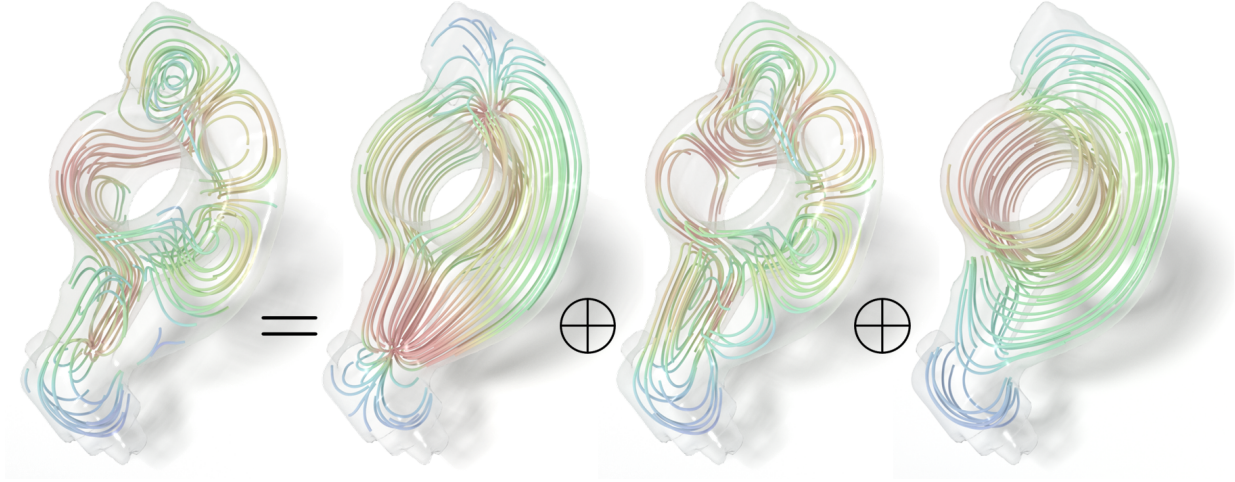
Figure 2.3: **Helmholtz-Hodge decomposition.** In this example, a tangential field is decomposed into the orthogonal sum of a tangential gradient field, a tangential curl field, and a tangential harmonic field.

is usually called the *codifferential* operator (acting on a k-form and returning a $(k-1)$-form).

Inner products of forms On a compact manifold M, the space of k-forms $\Omega^k(M)$ is a Hilbert space when equipped with the inner product between two k-forms α and β defined as:

$$\langle \alpha, \beta \rangle = \int_M \alpha \wedge \star \beta = \int_M \beta \wedge \star \alpha.$$

In our setting, it corresponds to the L_2-inner product between scalar fields for 0- or 3-forms, and to the L_2-inner product between vector fields for 1- or 2-forms.

2.1.3 Hodge decomposition for boundaryless manifolds

Based on the linear map d^k on a boundaryless manifold, there exists an orthogonal decomposition of the space Ω^k written as

$$\Omega^k = \ker d^k \oplus \operatorname{im} \delta^{k+1},$$

where $\ker$ denotes the kernel of an operator, $\oplus$ indicates an orthogonal sum of subspaces, and im denotes the image of an operator. This decomposition is simply a consequence of the fact that the

kernel of a linear operator is the orthogonal complement of the range of its adjoint operator. Note that we have

$$\langle d\alpha, \beta \rangle = \langle \alpha, \delta\beta \rangle + \int_{\partial M} \alpha \wedge \star\beta, \tag{2.1}$$

which implies that δ is formally the adjoint of d only for boundaryless manifolds, i.e., when $\partial M = \emptyset$.

The kernel component can be further decomposed by noticing that $\ker d^k = \operatorname{im} d^{k-1} + \mathcal{H}^k$, where $\mathcal{H}^k = \ker d^k / \operatorname{im} d^{k-1}$ is the quotient space between the kernel of d^k and the image of d^{k-1} (also known as the de Rham cohomology). This property is simply a consequence of the important property $d^k d^{k-1} = 0$. As observed by Hodge, we can turn the direct sum into an *orthogonal* sum instead by picking one particular representative for each equivalence class in $\mathcal{H}^k$: the unique one that is orthogonal to $\operatorname{im} d^{k-1}$. As $\Omega^k = \ker \delta^k \oplus \operatorname{im} d^{k-1}$, one realizes that, in fact, $\mathcal{H}^k$ is isomorphic to $\ker d^k \cap \ker \delta^k$, i.e., to the space of *harmonic* k-forms that are both closed and coclosed—we will also denote it as $\mathcal{H}^k$ due to the natural isomorphism. Given this newfound orthogonality, we reach the Hodge decomposition theorem, which states:

$$\Omega^k = \operatorname{im} d^{k-1} \oplus \operatorname{im} \delta^{k+1} \oplus \mathcal{H}^k. \tag{2.2}$$

In other words, a k-form $\omega \in \Omega^k$ can be decomposed into the orthogonal sum of an exact form $d\alpha$, a coexact form $\delta\beta$, and a harmonic form $h \in \mathcal{H}^k$ (a form is "exact" if it is the differential of another form; it is "coexact" if it is a codifferential instead). When $k = 1, 2$, this decomposition can be identified with its vector calculus equivalent, often referred to as the Helmholtz-Hodge decomposition in 3D:

$$\mathbf{v} = \nabla f + \nabla \times \mathbf{u} + \mathbf{h}. \tag{2.3}$$

According to de Rham's theorem, $\mathcal{H}^k$ is isomorphic to a space called the singular cohomology $H^k(M)$, which is in turn isomorphic to the homology $H_{n-k}(M)$ by Poincaré duality, where H_k can be understood as the space of non-contractible k-dimensional closed manifolds. The dimensionality of $H_k(M)$ is a finite topological invariant, often referred to as the k-th Betti number $\beta_k = \dim H_k(M)$. Based on Eq. (2.1), $\mathcal{H}^k$ can also be equivalently defined through $\mathcal{H}^k = \{\alpha \in \Omega^k | \Delta\alpha = 0\}$, where the (de Rham) *Laplacian* Δ is defined as $\Delta = d\delta + \delta d$, with thus a finite-

type	f^0	$\mathbf{v}^1$	$\mathbf{v}^2$	f^3
tangential	unrestricted	$\mathbf{v} \cdot \mathbf{n} = 0$	$\mathbf{v} \parallel \mathbf{n}$	$f\vert_{\partial M} = 0$
normal	$f\vert_{\partial M} = 0$	$\mathbf{v} \parallel \mathbf{n}$	$\mathbf{v} \cdot \mathbf{n} = 0$	unrestricted

Table 2.1: **Boundary conditions for scalar and vector fields.** This table shows the definition of tangential and normal boundary conditions.

dimensional kernel. In fact, Δ is self-adjoint due to Eq. (2.1), and we can decompose k-forms as

$$\Omega^k = \operatorname{im} \Delta^k \oplus \ker \Delta^k. \tag{2.4}$$

This fact suggests a simple computational approach to the Helmholtz-Hodge decomposition: if we can fix the finite rank deficiency of the Laplacian Δ, by projecting $\mathbf{v}$ to $\ker \Delta$ to get $\mathbf{h}$, and get $\mathbf{w}$ through $\Delta \mathbf{w} = \mathbf{v} - \mathbf{h}$, we can have the Hodge decomposition through $f = \nabla \cdot \mathbf{w}$ and $\mathbf{u} = \nabla \times \mathbf{w}$. However, in practice, our domain in 3D Euclidean space is always bounded and thus with a boundary—in which case, the boundary condition and orthogonality of the subspaces must be treated very carefully as we describe next.

2.1.4 Hodge decomposition for manifolds with boundary

To ensure adjointness of operators in the presence of boundary, there certainly are a variety of choices. A choice consistent with physical boundary conditions is to force the form α in the decomposition to be tangential to the boundary (we call a form α "tangential" or "parallel" if $\star\alpha$ is zero when applied to tangent vectors of the boundary), or normal to the boundary (we call a form α "normal" if α is zero when applied to tangent vectors of the boundary).

Consequently, we can construct a Hodge decomposition as proposed in [134] through

$$\Omega^k = d\Omega_n^{k-1} \oplus \delta\Omega_t^{k+1} \oplus \mathcal{H}^k, \tag{2.5}$$

where Ω_t^{k+1} is the space of tangential forms (also known as Neumann forms since their normal components are fixed), Ω_n^{k-1} is the space of normal forms (or Dirichlet forms), and $\mathcal{H}^k = \ker d^k \cap \ker \delta^k$. Note that being both closed and coclosed is *stronger* than satisfying $\Delta\omega = 0$ when $\partial M \neq \emptyset$. (This

point is important in our context, and we will come back to it in Sec. 2.3.) Nevertheless, one still has the orthogonality between the subspaces, using the adjointness of the operators with Dirichlet and Neumann boundary conditions for the potentials (sometimes called parallel and normal boundary conditions). However, $\mathcal{H}^k$ is infinite-dimensional in this case.

Complete decomposition Friedrichs [135] proposed two ways to decompose $\mathcal{H}^k$ orthogonally based on tangential or normal boundary conditions: $\mathcal{H}^k = \mathcal{H}_t^k \oplus (d\Omega^{k-1} \cap \mathcal{H}^k)$ as shown in Figure 2.4, or $\mathcal{H}^k = \mathcal{H}_n^k \oplus (\delta\Omega^{k+1} \cap \mathcal{H}^k)$, which can be combined into the following five-component (Hodge-Morrey-Friedrichs) decomposition:

$$\Omega^k = d\Omega_n^{k-1} \oplus \delta\Omega_t^{k+1} \oplus (\mathcal{H}_t^k + \mathcal{H}_n^k) \oplus (d\Omega^{k-1} \cap \delta\Omega^{k+1}), \tag{2.6}$$

where the sum of the latter three terms spans the harmonic space $\mathcal{H}^k$, while $\mathcal{H}_t^k$ is the tangential harmonic space, $\mathcal{H}_n^k$ is the normal harmonic space, and the last term is *both* exact and coexact. Friedrichs also noted that $\mathcal{H}_t^k \cong H^k(M)$ and $\mathcal{H}_n^k \cong H^k(M, \partial M)$, i.e., these two special harmonic spaces are isomorphic to, respectively, the aforementioned (absolute) cohomology $H^k(M)$ now for a manifold with boundary, and the relative cohomology $H^k(M, \partial M)$ for which two k-forms differing only by a k-form on the boundary are treated as equivalent. In general, $\mathcal{H}_t^k$ and $\mathcal{H}_n^k$ are not L_2-orthogonal with each other. However, they *are* orthogonal for domains in $\mathbb{R}^3$ according to [11], as both the absolute and relative homologies are due to the boundary (we can always patch up the boundary to turn the domain into a ball). This indicates that we have a 5-component orthogonal decomposition, which is consistent with the work of [16]:

$$\omega^k = d\alpha_n^{k-1} \oplus \delta\beta_t^{k+1} \oplus h_t^k \oplus h_n^k \oplus \eta^k, \tag{2.7}$$

where η^k is the part that is both exact and coexact. Note that three components can be expressed through potentials under boundary conditions that ensure orthogonality (η being exact and coexact, it can be written as the differential or the codifferential of a potential), while the other two components h_t^k and h_n^k belong to finite-dimensional spaces spanned by harmonic basis fields determined by topology (resp., absolute and relative homologies). As a reminder, the dimensionality of the
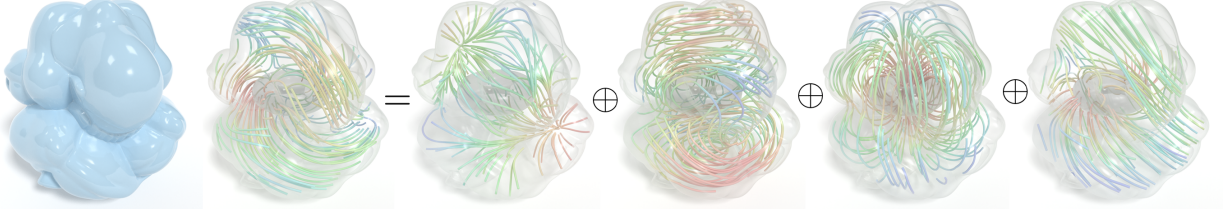
Figure 2.4: **Hodge-Morrey-Friedrichs decomposition.** For any 3D bounded domain, a vector field can always be decomposed into the orthogonal sum of the gradient of a scalar field vanishing on the boundary, the curl of a normal field, a tangential harmonic fields, and a harmonic gradient field.

spaces of both the h_t^k and h_n^{3-k} components in our 3D setup is the Betti number β_k, with a rather intuitive topological meaning: it is the number of components for $k=0$, tunnels for $k=1$, cavities for $k=2$, or simply 0 for $k=3$.

Equations defining the potentials   Applying the codifferential δ to both sides of Eq. (A.1), one finds that the form potential α_n^{k-1} satisfying

$$\delta\omega^k = \delta d\alpha_n^{k-1}. \tag{2.8}$$

This equation can be highly underdetermined as $d\alpha_n^{k-1} = d\big(\alpha_n^{k-1} + d\gamma\big)$ for any $\gamma \in \Omega^{k-2}$. As $d\gamma$ does not influence the exact component, it is referred to as a **gauge** field and can be arbitrarily fixed through various gauge conditions. For example, we can enforce $\delta\alpha^{k-1}=0$, turning the above equation into $\delta\omega = (\delta d + d\delta)\alpha^{k-1} = \Delta\alpha^{k-1}$. Since the rank deficiency of Δ restricted to the space of normal forms Ω_n^{k-1} is $\dim\mathcal{H}_n^{k-1}$, it is finite, so we can leverage this property to solve the corresponding linear system as we will see in Sec. 2.2.7 when we discuss discretizations and computations. Similarly, applying the differential d to both sides of Eq. (A.1) shows that

$$d\omega^k = \Delta\beta_t^{k+1}, \tag{2.9}$$

which has a rank deficiency of $\dim\mathcal{H}_t^{k+1}$ on tangential forms. We will also show in Sec. 2.2.8 that seeking a potential in $d\Omega \cap \delta\Omega$ for $k=1,2$ (i.e., for vector fields) can be achieved by solving a Laplace equation with Neumann boundary condition.

2.2 Discrete Decomposition of Vector Fields

We now assume that our 3D domain M is discretized in the form of a tetrahedral mesh. We can then use the discrete exterior calculus (DEC [136, 31, 137]) as our primary tool to represent discrete differential forms, as DEC preserves the key identity $d \circ d = 0$ for simplicial meshes. We will show that our *discrete* five-component decomposition exhibits the desired orthogonality defined by the L_2-inner product between discrete differential forms, as well as the proper cohomology dimensionality for tangential and normal forms—thus exactly mimicking the continuous case and preventing the presence of spurious terms.

Given a tetrahedral mesh M, we denote the set of vertices, edges, faces and tets as $\mathcal{V}/\mathcal{E}/\mathcal{F}/\mathcal{T}$ respectively. We refer to the boundary (triangle) mesh as B, with the boundary vertex/edge/face sets as $\mathcal{V}_b/\mathcal{E}_b/\mathcal{F}_b$. Finally, we assume w.l.o.g that the domain is connected; otherwise one can treat each connected component separately.

2.2.1 Discrete forms as values on mesh elements

A continuous k-form can be discretized very naturally on a mesh: one can integrate it against every oriented k-simplex of the tet mesh. The resulting set of scalar values (one per oriented k-simplex) can then be seen as a discrete k-form; see [31, 138] for details on how to reconstruct a continuous vector or scalar field from such discrete forms. In 3D, as we discussed in Sec. 2.1.2, vector fields can be interpreted as 1- or 2- forms, while scalar fields are 0- or 3- forms. So we will consider as an input vector field either a set of edge values (where each edge is given a fixed, but arbitrary orientation), or a set of face values (where, again, each face is given an arbitrary fixed orientation) encoding the associated discrete form. The whole discrete decomposition will then split the input discrete form into values on vertices, edges, faces, and/or tets, plus a few non-integrable components depending on the topology of the domain and the boundary condition.
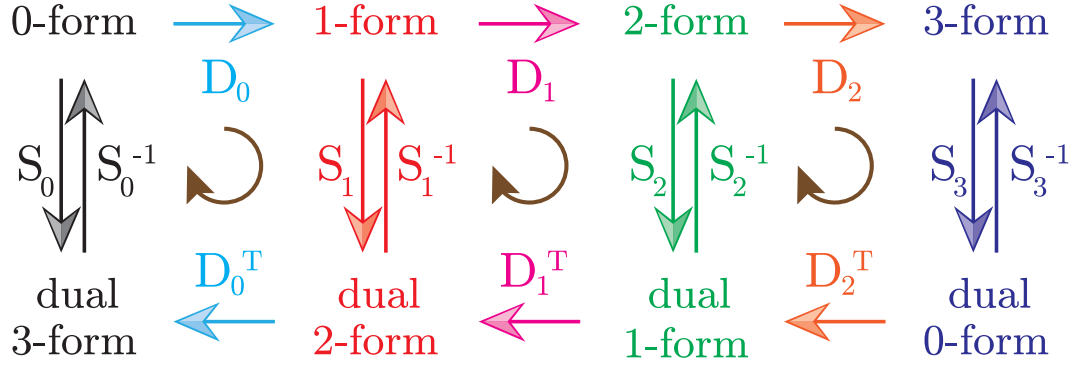
Figure 2.5: **Discrete de Rham cohomology.** The DEC linear operators provide a cohomology associated with the combinatorial operators D_k such that $D_{k+1}D_k = 0$ and the Hodge duality through the discrete Hodge stars S_k.

2.2.2 de Rham complex

Based on Stokes' theorem, the integral of $d\omega$ for a k-form ω on a $(k+1)$-simplex is simply the signed sum of the integral of ω on the boundary faces of the simplex, where the sign is determined by the relative orientation between the simplex and a particular face. Thus, the exterior derivative d^k is simply encoded in the discrete setting as a matrix D_k which stores the signed incidence between $(k+1)$-simplices and k-simplices [31]; it is thus very sparse and completely combinatorial. The identity $d \circ d = 0$ can be easily verified to hold ($D_{k+1}D_k = 0$) with this discrete definition since the boundary of a boundary of an element is always empty. The Hodge star $\star^k$ is treated as a mapping from a discrete form ω^k (one value per k-simplex) to one value per corresponding $(n-k)$-cells on a dual mesh—typically, the Voronoi dual structure of the tet mesh. The values on dual Voronoi $(n-k)$-cells are treated as the integral of an $(n-k)$-form stored on the dual mesh, and referred to as a *dual* discrete form. Thus, we will have two types of discrete forms (called primal forms $\in \Omega^k$ and dual forms $\in \widetilde{\Omega}^k$). Their isomorphism is through the Hodge duality $\star^k$, which, in the discrete setting, can be implemented as a *diagonal* matrix S_k, with diagonal entries representing the ratio between the $(n-k)$-volume of the Voronoi cell and the k-volume of the corresponding primal k-simplex. Other more accurate Hodge star matrices can be used (such as the Galerkin Hodge star [139]), but they must remain symmetric positive definite (SPD) to guarantee the correct dimensionality of the discrete cohomologies. We discuss how to construct sparse linear systems

for non-diagonal Hodge stars after the exposition based on diagonal Hodge stars. We refer to the collection of discrete form spaces connected by the discrete counterparts of the d and $\star$ operators as the *discrete de Rham complex*, mimicking its continuous counterpart, see Figure 2.5.

2.2.3   On the subtleties of boundary treatment

In order to provide a correct computational procedure to find the desired five-component decomposition, boundary values must be treated with caution: a naive derivation of operators without careful boundary treatment can lose the key adjoint properties that we seek to preserve. For instance, the general Laplacian operator for 1-forms is expressed in the continuous setting as $\Delta_1 = d\delta + \delta d$, the analog of the component-wise scalar Laplacian $\Delta = -\nabla\nabla\cdot + \nabla\times\nabla\times \equiv -\nabla^2$ of vector fields. Since we have discrete operators for d and δ, one could be tempted to directly define a discrete Laplacian L_1 as the symmetric matrix (corresponding to $\star\Delta_1$) through:

$$L_1 = D_1^T S_2 D_1 + S_1 D_0 S_0^{-1} D_0^T S_1.$$

This term-by-term conversion then corresponds to a discrete Dirichlet energy for discrete 1-forms $V \in \Omega_P^k$ defined as

$$\tfrac{1}{2} V^T L_1 V = \tfrac{1}{2}(D_1 V)^T S_2 (D_1 V) + \tfrac{1}{2}(S_0^{-1} D_0^T S_1 V)^T S_0 (S_0^{-1} D_0^T S_1 V).$$

However, one realizes that this energy mimics only the *non-boundary part* of the continuous identity (where $\mathbf{n}$ is the boundary normal):

$$\int_M \mathbf{v} \cdot \Delta\mathbf{v} = \int_M (\nabla \cdot \mathbf{v})^2 + \int_M (\nabla \times \mathbf{v})^2 - \int_{\partial M} \left[\mathbf{v}\nabla \cdot \mathbf{v} + \mathbf{v} \times \nabla \times \mathbf{v}\right] \cdot \mathbf{n}. \qquad (2.10)$$

In other words, the continuous de Rham Laplacian (Fig. 2.6) implicitly contains boundary terms that are not zero except for very specific types of vector fields, thus adding spurious terms. In the following two sections, we describe how to construct discrete operators that properly treat the typical boundary conditions required in practical computations. We begin with the *tangential Laplace operator*, i.e., the de Rham Laplace operator for vector fields that are *tangent* to the domain boundary, before turning our attention to the *normal Laplace operator*.

25

$$\Delta_0 = \star_0^{-1} d_0^T \star_1 d_0$$

$$\Delta_1 = \star_1^{-1} d_1^T \star_2 d_1 + d_0 \star_0^{-1} d_0^T \star_1$$

$$\Delta_2 = \star_2^{-1} d_2^T \star_3 d_2 + d_1 \star_1^{-1} d_1^T \star_2$$

$$\Delta_3 = d_2 \star_2^{-1} d_2^T \star_3$$

Figure 2.6: **Continuous de Rham Laplacian.** This figure shows the definition of continuous de Rham Laplacian.

2.2.4 Tangential vector Laplacian operator

Note that we have the choice between a primal 1-form or a primal 2-forms to represent a 3D vector field $\mathbf{u}$, which is equivalent to choosing a dual 2-form or a dual 1-form respectively. In order to assemble a proper discrete *tangential* vector Laplacian, we first discuss the case when our input is a primal 2-form $V \in \Omega_n^2 \subset \Omega^2$ (recall that $\mathbf{v}$ is tangential means that its corresponding 2-form $\mathbf{v}^2$ is normal), followed by the case of a primal 1-form.
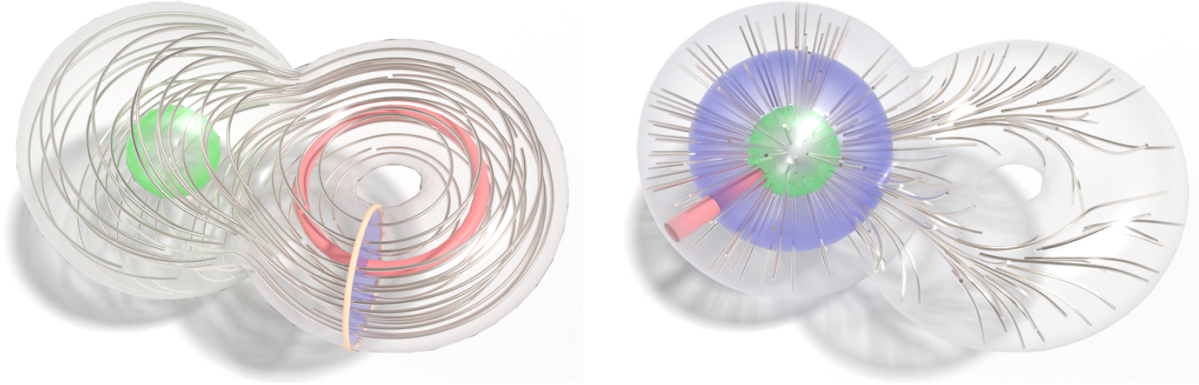
Figure 2.7: **Absolute and relative homologies.** Homology generators and corresponding harmonic fields on a topological torus with a spherical cavity inside. The red loop (left) around the tunnel represents the first homology, and the blue membrane is its dual in the second relative homology. The red curve (first relative homology generator, right) is a loop when the boundary is considered as a point, and the blue membrane is its dual in the second homology. Each harmonic field has the same circulation (resp., flux) on all loops (resp., membranes) that can deform into each other in the domain.

2-form version   For an input primal 2-form $V \in \Omega_n^2$ to be normal (and thus corresponding to a tangential vector field $\mathbf{v}$) simply means that the boundary face values are zero, i.e., the flux through boundary triangle faces is null: $\forall f \in \mathcal{F}_b$, $V_f = 0$.

Forcing the divergence calculation to consider the fluxes through boundary faces as zero is equivalent to simply removing the columns of D_2 that correspond to the boundary faces in order to keep only the interior face values. We denote the resulting matrix as $D_{2,\text{int}}$.

This idea is trivial to generalize: for $k = 0, 1, 2, 3$, we can create a version of D_k *restricted to the interior elements* using

$$D_{k,\text{int}} = P_{k+1} D_k P_k^T,$$

where P_k is the projection (or selection) matrix turning a discrete k-form to a *restricted* k-form containing only the values assigned to interior k-simplices. (Note while we use $D_{k,\text{int}}$ for conciseness, one can also implement this idea through a matrix $\overline{D}_{k,\text{int}} = P_{k+1}^T P_{k+1} D_k P_k^T P_k$ (this time, without altering the size of the original matrix D_k), as it directly zeroes out the elements in the rows and columns corresponding to boundary elements.)

Similar to $D_{2,\text{int}}$ for divergence, one can show that $D_{1,\text{int}}$ provides the correct curl calculation for tangential vector fields. In addition to removing the columns of D_1^T corresponding to boundary faces, it is necessary to remove the rows in D_1^T that correspond to the boundary edges: otherwise, the term corresponding to $\nabla \times \mathbf{v}$ would include a fictitious term assuming the tangential components along the boundary to be zero. More precisely, as shown in Fig. 2.8, D_1^T sums up the integrals along the yellow dual polyline around the red boundary edge. Defining this term as the curl would amount to setting to zero the line integral along the dotted boundary path that forms a closed loop with the yellow polyline.

We also denote the discrete Hodge star for *interior* k-forms as

$$S_{k,\text{int}} = P_k S_k P_k^T.$$
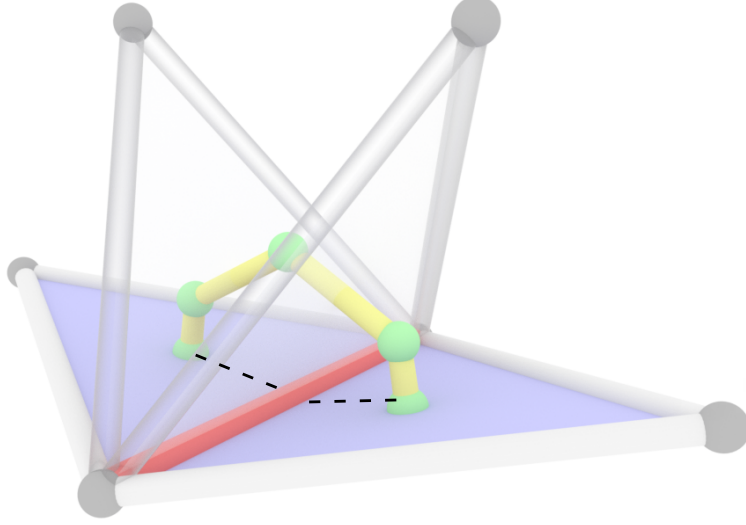
Figure 2.8: **Curl calculation on boundary edge.** This figure shows how to compute curl on boundary edge for tangential boundary condition.

Notice that $S_{3,\text{int}} = S_3$, since all tets are interior tets. Now the normal 2-form Laplacian can be expressed as

$$L_{2,n} = D_{2,\text{int}}^T S_3 D_{2,\text{int}} + S_{2,\text{int}} D_{1,\text{int}} S_{1,\text{int}}^{-1} D_{1,\text{int}}^T S_{2,\text{int}},$$

With this expression, we can verify that the harmonic forms defined by $h \in \ker L_{2,n}$ indeed correspond to the relative cohomology

$$H^2(M, \partial M) = \ker D_{2,\text{int}} / \operatorname{im} D_{1,\text{int}}.$$

Indeed, we first note that both terms in $L_{2,n}$ are semi-positive definite. Thus $h^T L_{2,n} h = 0$ indicates $(D_{2,\text{int}} h)^T S_3 (D_{2,\text{int}} h) = 0$, which means $h \in \ker D_{2,\text{int}}$ as S_3 is positive definite. Similarly, $D_{1,\text{int}}^T S_{2,\text{int}} h = 0$, which implies $\forall V \in \operatorname{im} D_{1,\text{int}}, V^T S_{2,\text{int}} h = W^T D_{1,\text{int}}^T S_{2,\text{int}} h = 0$ (where $V = D_{1,\text{int}} W$), thus h is orthogonal to $\operatorname{im} D_{1,\text{int}}$. Consequently, h is the unique representative for its equivalence class in the quotient space, and we have the following theorem.

> *Discrete de Rham's Theorem for Normal 2-Forms.* The space of discrete harmonic 2-forms normal to the boundary (i.e., our discrete counterpart to the de Rham cohomology $\mathcal{H}_{\text{dR},n}^2$) is isomorphic to the second (singular) relative cohomology group $\ker L_{2,n} \cong H^2(M, \partial M)$.

By Lefschetz duality, $H^2(M, \partial M) \cong H_1(M)$, the first homology group, which represents the independent "tunnels" of the shape M (see Figure 3.4). So the dimension of $\ker L_{2,n}$ is $\beta_1 \equiv$
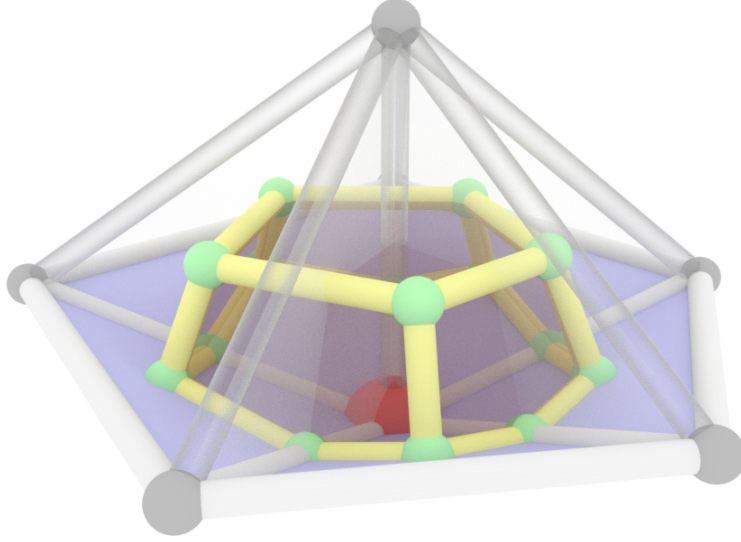
Figure 2.9: **Flux calculation on boundary vertex.** This figure shows how to compute flux on boundary vertex for normal boundary condition.

$\dim H_1(M)$, exactly the sum of the genus for each connected component of the boundary ∂M. We can thus safely use a typical eigensolver to find the β_1 unit eigenvectors associated with the smallest eigenvalues of $L_{2,n}h = \lambda S_2 h$ (these eigenvalues will be 0 up to numerical accuracy): these are the basis of all harmonic 2-forms normal to the boundary. We assemble them into a (tall) matrix $\mathbb{H}_{2,n}$ of size $|\mathcal{F} \backslash \mathcal{F}_b| \times \beta_1$ as: $\mathbb{H}_{2,n} = [h_1 \dots h_{\beta_1}]$.

1-form version   If a 1-form discretization is used for input tangential vector fields $V \in \Omega_t^1$, a direct term-by-term discretization actually holds, i.e., the discrete tangential 1-form Laplacian is simply:

$$L_{1,t} = D_1^T S_2 D_1 + S_1 D_0 S_0^{-1} D_0^T S_1.$$

This is, in light of the previous case, not surprising: a full discretization of the vector field as a discrete one-form should also include one value per boundary Voronoi region (the intersection of the dark dual

polyhedra dual to the red boundary vertex on the boundary surface as shown in Fig. 2.9), stored as U_b: otherwise, the dual 2-form $S_1 U$ cannot be integrated over the boundary by lack of information. Consequently, the matrices D_1 and D_0^T should be augmented accordingly; but forcing the 1-form to be tangential means that the extra rows/columns must be suppressed anyway just like

in the previous case: the use of the original discrete exterior derivatives is thus justified. Since $S_0^{-1} D_0^T S_1$ is now the divergence operator, and at the boundary, the fluxes through interior dual faces are summed while getting no contribution from boundary (tangential condition), the resulting operator properly captures its continuous counterpart. We thus have a similar isomorphism theorem.

> *Discrete de Rham's Theorem for Tangential 1-Forms.* The space of discrete harmonic 1-forms tangential to the boundary is isomorphic to the first cohomology group $\ker L_{1,t} \cong H^1(M)$.

Since the Hodge duality holds for singular cohomology ($H^1(M) \cong H^2(M, \partial M)$), $\ker L_{2,n}$ and $\ker L_{1,t}$ are isomorphic (see Figure 3.4). The dimensionality of $\ker L_{1,t}$ is again $\dim H_2(M, \partial M) = \dim H_1(M) = \beta_1$. Solving for the first β_1 eigenvectors associated with the smallest eigenvalues of $L_{1,t} h = \lambda S_1 h$ (which will be 0 up to numerical accuracy) is thus also a viable approach to computing a basis. We assemble them into a (tall) matrix $\mathbb{H}_{1,t}$ of size $|\mathcal{E}| \times \beta_1$ as: $\mathbb{H}_{1,t} = [h_1 \dots h_{\beta_1}]$.

2.2.5 Normal vector Laplacian operator

The discrete expressions of the two normal Laplacian operators can be obtained by basically mirroring the arguments used earlier, as we now review for completeness.

1-form version For an input primal 1-form $V \in \Omega_n^1$ to represent a normal vector field (i.e., a 1-form normal to the boundary), one must clearly have: $\forall e \in \mathcal{E}_{\mathcal{B}}, V_e = 0$. Thus, modifying D_1 to become $D_{1,\text{int}}$ by removing the columns of the boundary edges as earlier is required. Moreover, the discrete dual divergence operator D_0^T must also be altered to become $D_{0,\text{int}}^T$ by removing the rows corresponding to boundary vertices: otherwise, a fictitious term in divergence $\nabla \cdot \mathbf{v}$ would (erroneously) assume the boundary fluxes to be zero. The discrete 1-form normal Laplacian is then

$$L_{1,n} = D_{1,\text{int}}^T S_{2,\text{int}} D_{1,\text{int}} + S_{1,\text{int}} D_{0,\text{int}} S_{0,\text{int}}^{-1} D_{0,\text{int}}^T S_{1,\text{int}}.$$

2-form version Similar to the tangential 1-form $V \in \Omega_n^2$ case, we can augment the discrete tangential 2-form (corresponding to a normal vector field) with additional variables—this time, values of

the line integral along each boundary dual edge to encode the tangential components of the 1-form. Setting these extra terms to 0 turns out to be equivalent to using the original D_k and S_k matrices to assemble the Laplacian, hence:

$$L_{2,t} = D_2^T S_3 D_2 + S_2 D_1 S_1^{-1} D_1^T S_2.$$

Like in the tangential case, the two related theorems follow.

Discrete de Rham's Theorem for Normal 1-Forms. The space of discrete harmonic 1-forms normal to the boundary is isomorphic to the first relative cohomology group $\ker L_{1,n} \cong H^1(M, \partial M)$.

Discrete de Rham's Theorem for Tangential 2-Forms. The space of discrete harmonic 2-forms tangential to the boundary is isomorphic to the second cohomology group $\ker L_{2,t} \cong H^2(M)$.

The dimension of the harmonic space is $\dim H^2(M) = \dim H_2(M) \equiv \beta_2$, which is the number of connected components of the boundary minus 1 (see Figure 3.4). Solving for the first β_2 eigenvectors associated with the smallest eigenvalues of $L_{1,n} h = \lambda S_1 h$ and $L_{2,t} h = \lambda S_2 h$ (which will be 0 up to numerical accuracy) provides us with the basis of these tangential harmonic spaces. As earlier, we assemble them into two (tall) matrices $\mathbb{H}_{1,n}$ and $\mathbb{H}_{2,t}$ of size $|\mathcal{E} \backslash \mathcal{E}_b| \times \beta_2$ and $|\mathcal{F}| \times \beta_2$.

2.2.6 Normal and tangential scalar Laplacian operators

For the case of the Laplacian operator of scalar functions (represented as 0- or 3-forms), the exact same construction applies—but the expressions are simpler as only one part of the $d\delta + \delta d$ general expression is nonzero in these cases. We find:

$$L_{0,t} = D_0^T S_1 D_0 \qquad\qquad L_{3,n} = S_{3,\text{int}} D_{2,\text{int}} S_{2,\text{int}}^{-1} D_{2,\text{int}}^T S_{3,\text{int}}$$

$$L_{0,n} = D_{0,\text{int}}^T S_{1,\text{int}} D_{0,\text{int}} \qquad\qquad L_{3,t} = S_3 D_2 S_2^{-1} D_2^T S_3.$$

As in the continuous case, the rank deficiency of $L_{0,t}$ and $L_{3,n}$ is $\dim H^0(M) = \dim H^3(M, \partial M) = \dim H_0(M) \equiv \beta_0$, i.e., the number of connected components of the domain. The rank deficiency

of $L_{0,t}$ and $L_{3,n}$ is, instead, $\dim H^3(M) = \dim H^0(M, \partial M) = \dim H_3(M) \equiv \beta_3 = 0$ as we cannot have a non-empty boundary of the 3D domain.

The reader may notice that just like for vector Laplacians, the normal Laplacians (where "normal" is meant in the differential form sense, i.e., with an n subscript) involve interior elements only, while the tangential Laplacians are assembled from full-blown differential and star operators. Thus the following formula can be used for any k (where terms containing an index < 0 or > 3 are considered null):

$$L_{k,t} = D_k^T S_{k+1} D_k + S_k D_{k-1} S_{k-1}^{-1} D_{k-1}^T S_k$$
$$L_{k,n} = D_{k,\text{int}}^T S_{k+1,\text{int}} D_{k,\text{int}} + S_{k,\text{int}} D_{k-1,\text{int}} S_{k-1,\text{int}}^{-1} D_{k-1,\text{int}}^T S_{k,\text{int}}.$$

2.2.7 Five-component decomposition

We are now ready to introduce our computational approach to evaluate the five-component decompositions, which depending on whether we start from a 1-form V^1 or 2-form V^2 input, reads

$$\boxed{V^k = D_{k-1}\alpha^{k-1} + S_k^{-1} D_k^T S_{k+1}\beta^{k+1} + h_t^k + h_n^k + \eta^k} \text{ for } k = 1, 2.$$

They both correspond to the same vector field decomposition in vector calculus, i.e., $\mathbf{v} = \nabla f + \nabla \times \mathbf{u} + \mathbf{h}_t + \mathbf{h}_n + \nabla e$, where f is a scalar function that vanishes on ∂M (therefore, ∇f is orthogonal to the boundary), $\mathbf{u}$ is a vector potential that is orthogonal to ∂M (and thus $\nabla \times \mathbf{u}$ is a tangential vector field at the boundary), $\mathbf{h}_t$ is a tangential harmonic field, $\mathbf{h}_n$ is a normal harmonic field, and e is a harmonic scalar function (because of the exact and coexact nature of this last term, one can equivalently write it in vector calculus also as the curl $\nabla \times \mathbf{e}$ of a harmonic vector field $\mathbf{e}$).

Equations to solve for potentials For the 1-form decomposition, one uses our preassembled Laplacian matrices to solve the two discrete form potentials α^0 (on vertices) and β^2 (on faces):

$$L_{0,n}\,\alpha^0 = D_{0,\text{int}}^T P_1 S_1 V^1, \qquad L_{2,t}\beta^2 = S_2 D_1 V^1.$$

For the 2-form decomposition, we solve for α^1 (on edges) and β^3 (on tets) instead through:

$$L_{3,t}\beta^3 = S_3 D_2 V^2, \qquad L_{1,n}\,\alpha^1 = D_{1,\text{int}}^T P_2 S_2 V^2.$$

Topological parts   The next two parts are evaluated by just projecting the input form onto the eigenvectors of the vector Laplacian:

$$h_t^k = \mathbb{H}_{k,t}\mathbb{H}_{k,t}^T S_k V^k, \quad h_n^k = P_k^T \mathbb{H}_{k,n}\mathbb{H}_{k,n}^T P_k S_k V^k,$$

since $\mathbb{H}\mathbb{H}^T$ is a projection over the rows of $\mathbb{H}$.

Last term   The fifth element, i.e., the 1-form η^1 or 2-form η^2, can finally be deduced from the other four components by subtracting them from the input. Note that η is completely determined by either its normal component at the boundary or its tangential components at the boundary. We will also introduce two alternative ways to directly compute it through either of its potentials in Sec. 2.2.8.

Resolving rank deficiency   The only technical issue in implementing the above linear solves is that some of the Laplacian matrices involved do not have full rank. Fortunately, we know *exactly* their rank deficiency, as well as a basis of their kernel (the associated harmonic forms). For instance,
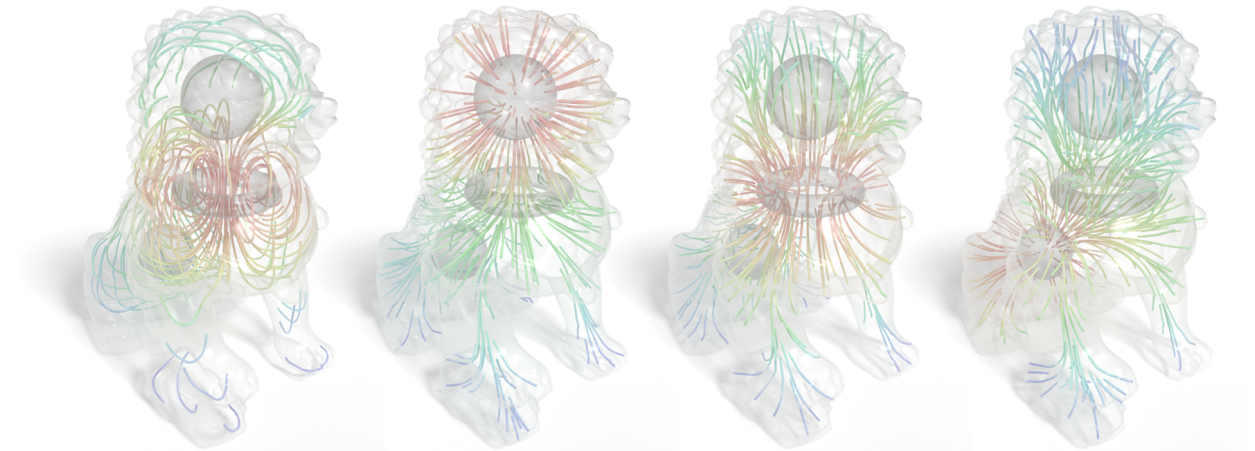
Figure 2.10: **Harmonic field basis.** Shown are $(\beta_1\!=\!1)$ tangential and $(\beta_2\!=\!3)$ normal harmonic basis fields spanning the corresponding harmonic spaces.
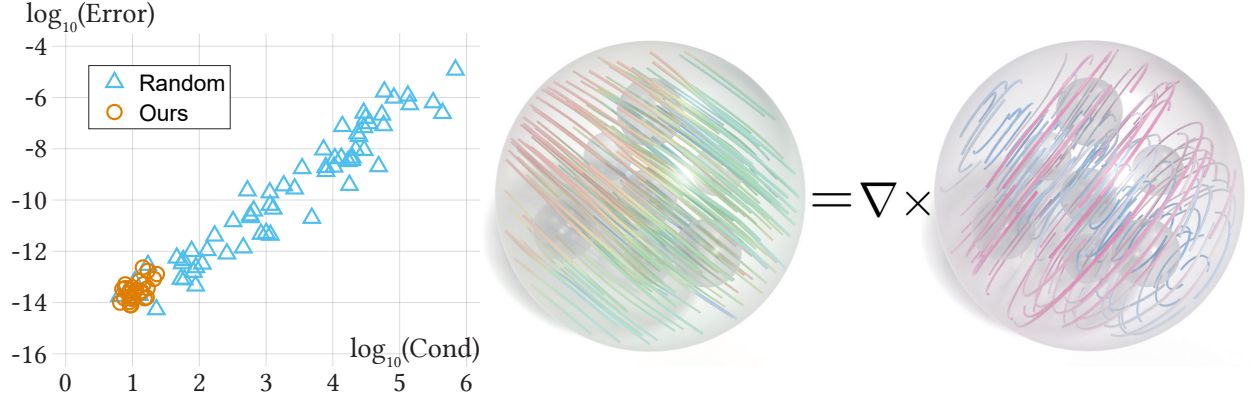
Figure 2.11: **Resolving rank deficiency.** Randomly selected index sets to remove degeneracy of linear systems may result in very large inaccuracies in the solution of the linear system, unless our simple heuristic is used.

for an equation of the form $L_{2,t}x = y$, we know that the linear system is indefinite since $L_{2,t}$ has a rank deficiency of β_2. One way to get a definite linear system is to add the constraint $\mathbb{H}_{2,t}^T S_2 x = 0$; but the system is now rectangular. One could instead solve $\left(L_{2,t} + S_2 \mathbb{H}_{2,t} \mathbb{H}_{2,t}^T S_2\right) x = y$ efficiently with an iterative solver since $S_2 \mathbb{H}_{2,t} \mathbb{H}_{2,t}^T S_2$ is dense but can be multiplied with a vector in $\mathcal{O}(n)$ time, where n is the number of faces.

Inspired by numerical strategies to pick a subset of columns in order to obtain an optimal condition number with high probability [140], we propose instead a simpler alternative since we already know the kernel and its topological origin. We randomly pick β_2 face indices of x. We assemble the small square sub-block of $\mathbb{H}_{2,t}$ corresponding to these indices, and check its condition number. After having tried $10\beta_2$ such randomly selected index sets, we pick the one with the lowest condition number among those with a determinant higher than the lowest 10% determinants — or we stop early if a condition number happens to be below a given safe threshold (we pick 5.0). This simple procedure has always performed reliably in all of our tests (see Figure 2.11). Once we find a good set of indices, we remove these indices from x, along with the corresponding rows and columns of $L_{2,t}$, project the right hand side to $b - S_2 \mathbb{H}_{2,t} \mathbb{H}_{2,t}^T b$, and remove the β_2 indices of this resulting vector as well. This smaller (yet still symmetric) linear system will then have full rank (since we fixed the null space), and a solution of the original equation is the solution of this non-degenerate system where the few missing indices are set to 0. Note that we can finally project this solution
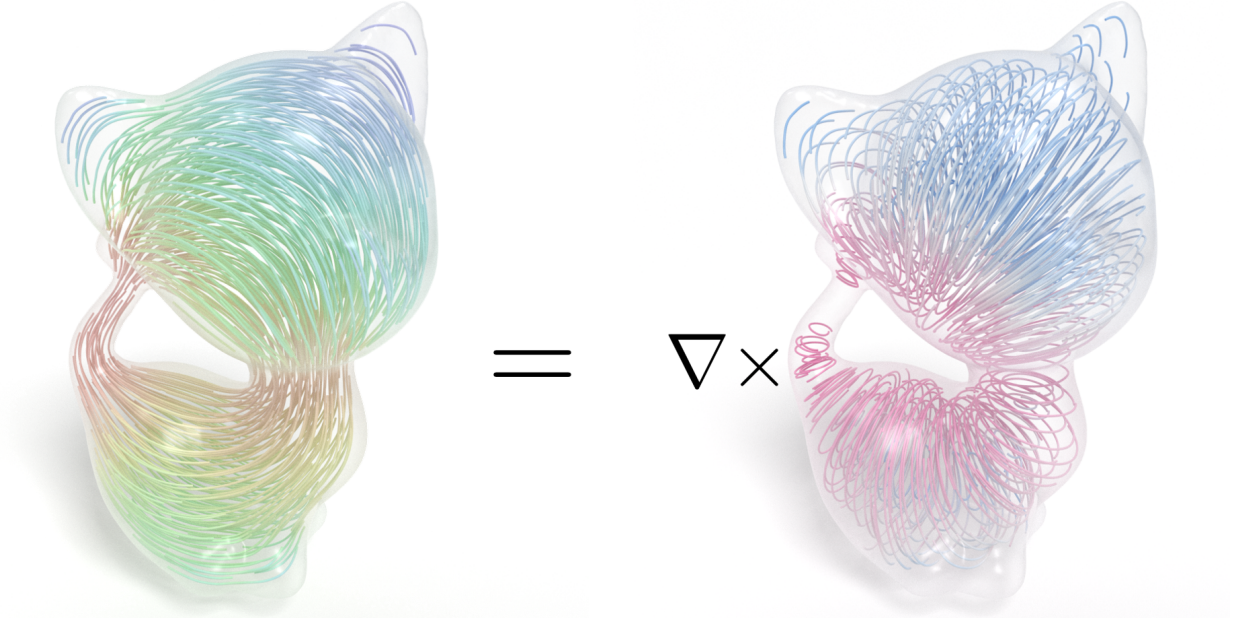
Figure 2.12: **Vector potential for tangential harmonic field.** For a tangential harmonic vector field (left) inside his kitten model forming a torus, we can compute its vector potential (right) whose curl is the original field.

to the space containing no harmonic fields using $\mathbb{H}_{2,t}$, if needed. Other rank deficiencies are fixed similarly.

2.2.8 Potentials for the harmonic components

While we proposed a simple eigen-based procedure to compute the tangential and normal harmonic spaces, we can exploit the fact that our domain is embedded in $\mathbb{R}^3$ to directly compute potentials that define the two topological and harmonic terms $\mathbf{h}_t$ and $\mathbf{h}_n$[2]. Depending on how the decomposition is used in practical applications, this alternative approach may be more efficient. For completeness, we also describe how to extract the potential (either as a gradient or a curl) of the fifth term η.

---

[2]These potentials are not of the same nature as α and β: from Eq. (2.5), one can see that harmonic parts can *not* be written as the d of a normal form or the δ of a tangential form. But they are, however, in the range of d and the range of δ, so we can find potentials for them—just not with the same boundary conditions, hence the commonly-used term of "non-integrable" to describe these topological terms.

Tangential harmonic space   If one has already computed the generators for $H_2(M, \partial M)$, i.e., a set of independent surfaces in M that have their boundary loops in ∂M, we can construct one gradient field per generator that will be a *tangential* harmonic field. The gradient field is constructed by simulating a cut through the generator, allowing the potential to have two different (edge or face) values on the generator that differ exactly by 1 as done in global parameterization methods for quad meshing purposes [141]; we can then solve a Laplace equation with a single element fixed to remove the null space. Once these gradients are found, we run a Gram-Schmidt procedure to obtain an orthonormal basis for these tangential harmonic fields. Another simple strategy is to first restrict the computation to the nonzero genus boundary components. For each handle loop (that is, a non-contractible loop of the boundary which can be contracted inside the domain volume), one may build a tangential harmonic vector field $\mathbf{h}_t$ on the surface such that the circulation around the handle loop is 1; one can then extend $\mathbf{h}_t$ to the volume by solving a vector Laplace equation $\Delta \mathbf{h}_t = 0$ (i.e., either $L_{1,t} h_t^1 = 0$ or $L_{2,n} h_n^2 = 0$ depending on the decomposition being targeted) with Dirichlet boundary condition on all components of $\mathbf{h}_t$ at the boundary (for all other connected components of the boundary, it is set to 0). We will have β_1 such vector potentials, and they will span the entire space of tangential harmonic field, providing an alternative to the construction of $\mathbb{H}_{1,t}$ and $\mathbb{H}_{2,n}$. Moreover, the vector potential $\boldsymbol{\psi}_t$ of each basis element can be solved through $\Delta \boldsymbol{\psi}_t = \nabla \times \mathbf{h}_t$ with boundary conditions $(\nabla \times \boldsymbol{\psi}_t) \times \mathbf{n} = \mathbf{h}_t \times \mathbf{n}$ and by forcing the normal component of $\boldsymbol{\psi}_t$ to be 0 to impose the gauge condition,; e.g., for h_n^2, we can solve for the potential ψ_t^1 through $L_{1,t} \psi_t^1 = D_1^T S_2 h_n^2$, where the righthand side computes the tangential component of h_n^2 (i.e., it generates the tangential component of h_n^2 across the dual of boundary edges and produces 0 for the interior); we proceed for h_t^1 in a similar fashion, with $L_{2,n} \psi_n^2 = S_{2,\text{int}} D_{1,\text{int}} P_1 h_t^1$ this time (where $D_{1,\text{int}} P_1 h_t^1$ contains only the negative tangential component along boundary edges).

Normal harmonic space   We can similarly construct the elements of the kernels of the *normal* Laplacian matrices directly. Through the duality to $H^1(M, \partial M)$, we can represent these harmonic functions as combinations of simple gradient fields $\mathbf{h}_n = \nabla \phi$ (where ϕ be a discrete 3-form (resp.,
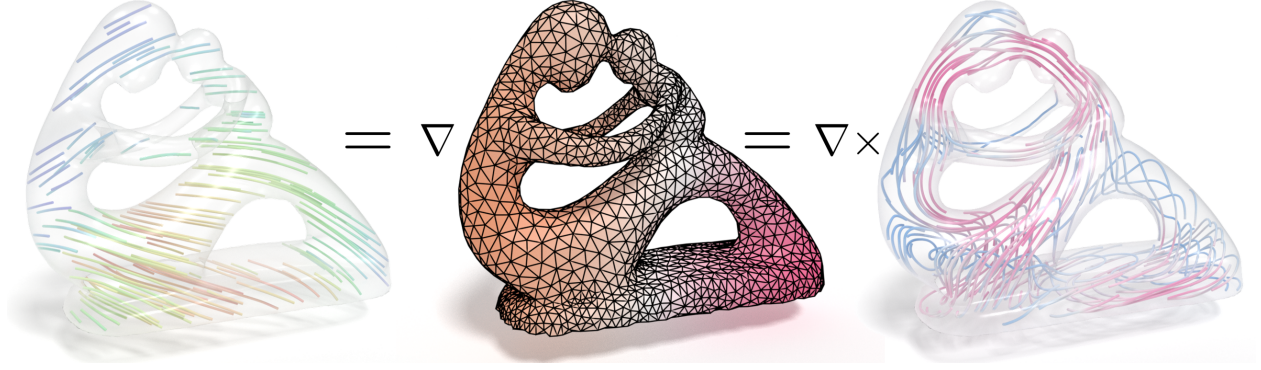
Figure 2.13: **Potentials for the exact and coexact field.** In any 3D volume, the fifth vector component η in our decomposition (left) can be expressed both as a gradient field (middle) and a curl field (right).

0-form) when the input is a 2-form (resp., 1-form)), which are the solution of $\Delta\phi=0$ with Dirichlet boundary conditions $\phi=1$ on one of the connected component of the boundary mesh, and $\phi=0$ on the rest of the boundary. This is a 3D extension of the procedure proposed in the Appendix of [141], and essentially corresponds to the problem of finding a static electric field with ideal conductor boundary for given potentials on the boundary. For 1-form inputs, it is solved on the graph of primal vertices and edges, while for 2-form inputs, it is solved on the dual graph for tets and faces. An additional Gram-Schmidt procedure is also necessary if one requires an orthonormal basis. If the potentials of these normal harmonic basis elements are needed, we can solve them in a mirrored way to the tangential case through: $L_{0,t}\phi_t^0 = D_0^T S_1 h_n^1$ and $L_{3,n}\phi_n^3 = S_3 D_{2,\text{int}} P_2 h_t^2$.

Potential(s) for fifth term   From η^1/η^2, we can solve for their scalar potential e ($\eta^1 = D_2 e_t^0 \equiv de_t^0$ or $\eta^2 = S_{2,\text{int}}^{-1} D_{2,\text{int}}^T S_3 e_n^3 \equiv \delta e_n^3$) through:

$$L_{0,t}e_t^0 = D_0^T S_1 \eta^1, \quad \text{or} \quad L_{3,n}e_n^3 = S_3 D_{2,\text{int}} P_2 \eta^2,$$

where the right hand side only contains nonzero terms at the boundary (enforcing $\nabla e \cdot \mathbf{n} = \eta \cdot \mathbf{n}$). This is essentially the same setup as solving for potentials of normal harmonic fields. For such Neumann boundary conditions, we also need to fix $\beta_0 = \dim H_0(M)$ variables, since we can add one constant function to each connected component of the domain without changing the actual fifth component. Likewise for the vector potential $\mathbf{e}$ ($\eta^1 = S_{1,\text{int}}^{-1} D_{1,\text{int}}^T P_2 S_2 e_n^2 \equiv \delta e_n^2$ or $\eta^2 = D_1 e_t^1 \equiv de_t^1$)

by solving the same type of Laplace equation with boundary conditions $\nabla \times \mathbf{e} \times \mathbf{n} = \eta \times \mathbf{n}$ as for potentials of tangential harmonic fields, i.e.,

$$L_{2,n} e_n^2 = S_{2,\text{int}} D_{1,\text{int}} P_1 \eta^1 \quad \text{or} \quad L_{1,t} e_t^1 = D_1^T S_2 \eta^2.$$

Observe that directly applying δ to e_n^3 (resp., e_n^2) only provides correct values for η^2 (resp., η^1) on interior elements . Still, these potentials offer enough information for extrapolation to boundary elements through harmonicity of η^2: if each tet contains at most one boundary face, η^2 on that face is the negated sum of the other three fluxes; likewise for η^1 if each boundary edge is incident to at least one triangle with only one boundary edge. If the input mesh does not satisfy these conditions, local splits of offending tets and triangles can be applied. Alternatively, e_n^3 (resp., e_n^2) can be supplemented with one value per boundary face (resp., edge) for δ to generate the correct gradient (resp., curl) on each boundary element.

2.2.9   Counting argument

Both to further enhance the understanding of our discrete vector decomposition and to offer yet another approach to convince oneself that the counting is correct, we now review the number of degrees of freedom (DoFs) within each component in both representations. For the 1-form representation, $d\alpha^0$ contains $|\mathcal{V}| - |\mathcal{V}_B|$ DoFs, i.e. one value per interior vertex; $\delta\beta^2$ contains $|\mathcal{F}| - |\mathcal{T}| - \beta_2$ since we start with $|\mathcal{F}|$ values but need to get rid of $\dim \ker \delta_2 = \dim \operatorname{im} \delta_3 + \dim H_2$. The non-integrable components h_t^1 and h_n^1 provide β_1 and β_2 DoFs respectively. Finally, η^1 provides $|\mathcal{V}_B| - \beta_2 - \beta_0$ DoFs, because $\beta_2 + \beta_0$ is the number of connected components of the boundary, and on each of them the total flux is 0. From the Poincaré-Euler formula

$$|\mathcal{V}| - |\mathcal{E}| + |\mathcal{F}| - |\mathcal{T}| = \beta_0 - \beta_1 + \beta_2 - \beta_3,$$

we can then verify that the number of values of the input 1-form ($|\mathcal{E}|$) is indeed the sum of the above DoFs (with $\beta_3 = 0$ in 3D) . For the 2-form representation, following similar arguments, the DoFs for the five components are in the same order: $|\mathcal{E}| - |\mathcal{E}_B| - |\mathcal{V}| - |\mathcal{V}_B| - \beta_2$, $|\mathcal{T}|$, β_2, β_1, and

$|\mathcal{F}_B| - \beta_2 - \beta_0$. Noting that $|\mathcal{V}_B| - |\mathcal{E}_B| + |\mathcal{F}_B| = 2(\beta_0 - \beta_1 + \beta_2)$ as it is the sum of the Euler characteristic $2 - 2g$ of each boundary component (one should not use the Euler characteristic of the volumetric domain!), we can again verify that they sum up to $|\mathcal{F}|$, as expected. We recap the various numbers of DoFs in Tab. A.1.

	1-form DoFs	2-form DoFs
ω	$\|\mathcal{E}\|$	$\|\mathcal{F}\|$
$d\alpha$	$\|\mathcal{V}\|-\|\mathcal{V}_B\|$	$\|\mathcal{E}\|-\|\mathcal{E}_B\|+\|\mathcal{V}\|-\|\mathcal{V}_B\|-\beta_2$
$\delta\beta$	$\|\mathcal{F}\|-\|\mathcal{T}\| - \beta_2$	$\|\mathcal{T}\|$
h_t	β_1	β_2
h_n	β_2	β_1
η	$\|\mathcal{V}_B\|-\beta_2-\beta_0$	$\|\mathcal{F}_B\|-\beta_2-\beta_0$

Table 2.2: **List of DoFs for 1-form and 2-form decompositions.**

2.3 Variational nature of our decomposition

Because we made sure our discrete treatment is closely mimicking the continuous five-component Hodge decomposition, it is directly related to variational approaches to vector decomposition based on L_2 energies. In particular, we point out that our discrete treatment can be understood as a particular enforcement of harmonicity with zero divergence and curl boundary conditions to enforce proper orthogonal projections.

Continuous notion of harmonicity Because we are in $\mathbb{R}^3$, recall from Eq. (2.10) that the Laplace quadratic form satisfies:

$$\int_M \mathbf{v} \cdot \Delta\mathbf{v} = \int_M (\nabla \cdot \mathbf{v})^2 + (\nabla \times \mathbf{v})^2 \tag{2.11}$$

only if the boundary integral vanishes, i.e.:

$$\int_{\partial M} (\mathbf{v}\nabla \cdot \mathbf{v} + \mathbf{v} \times \nabla \times \mathbf{v}) \cdot \mathbf{n} = 0$$

Our choice of gauge in the decomposition proposed in Sec. 2.2.7, in fact, enforces the latter equality since it implies that we discretely enforce $\nabla \times \mathbf{v} = 0$ (with tangential $\mathbf{v}$) or $\nabla \cdot \mathbf{v} = 0$ (with normal $\mathbf{v}$) to make this boundary integrand identically zero. This is precisely why our harmonic forms are not only harmonic ($\Delta\mathbf{v} = 0$) in the interior of the domain, but have these boundary conditions enforced

as well — explaining why we stated in Sec. 2.1.4 that forcing the forms in $\mathcal{H}^k$ to be closed and coclosed (hence, curl- and divergence-free) is *stronger* than just the notion of interior harmonicity. In the continuous setting, the consequence of these boundary conditions is that our construction can then be understood as forcing the tangential vector fields to satisfy one Dirichlet condition $\mathbf{v}\cdot\mathbf{n}=0$ (tangentiality) and two Neumann conditions $\mathbf{n}\cdot\nabla_{\mathbf{t}_1}\mathbf{v}=0$, $\mathbf{n}\cdot\nabla_{\mathbf{t}_2}\mathbf{v}=0$ (where $\mathbf{t}_1$ and $\mathbf{t}_2$ are two local tangent direction forming a coordinate frame with the surface normal $\mathbf{n}$) to enforce a zero curl. On the other hand, the normal vector fields are constrained to satisfy two Dirichlet conditions $\mathbf{v}\cdot\mathbf{t}_1=0$ and $\mathbf{v}\cdot\mathbf{t}_2=0$, and one Neumann condition $\mathbf{n}\cdot\nabla\mathbf{v_n}=0$ to enforce a zero divergence. These conditions are consistent with the formulation on generic manifold cases from [142]. With these three conditions on the boundary added to the condition of harmonicity, the space of harmonic forms is finite-dimensional.

Dirichlet energy   In flat $\mathbb{R}^3$, the oft-used Dirichlet energy of vector fields can be converted to a volume integration using the Laplacian and a boundary term through integration by part:

$$\int_M |\nabla\mathbf{v}|^2 = \int_M \mathbf{v}\cdot\Delta\mathbf{v} + \int_{\partial M} \mathbf{v}\cdot\nabla_{\mathbf{n}}\mathbf{v}. \tag{2.12}$$

Notice now that with the boundary conditions we enforced, by Eq. (2.11), we have three energies that match: the harmonic energy, the Laplace quadratic form, and the Dirichlet energy, i.e.,

$$\int_M (\nabla\cdot\mathbf{v})^2 + (\nabla\times\mathbf{v})^2 = \int_M \mathbf{v}\cdot\Delta\mathbf{v} = \int_M |\nabla\mathbf{v}|^2.$$

Variational nature   Due to its L_2-orthogonality, one can conceive our decomposition as orthogonal projections onto subspaces — and thus as a variational problem. For instance the projection of V^1 onto $D_0\alpha^0$ can be seen as minimization of

$$\langle V^1 - D_0\alpha^0, V^1 - D_0\alpha^0 \rangle = \langle V^1, V^1 \rangle - 2\langle V^1, D_0\alpha^0 \rangle + \langle D_0\alpha^0, D_0\alpha^0 \rangle.$$

The scalar function α^0 is then entirely determined by adding a gauge to enforce that α is zero on the boundary. This type of variational arguments were already leveraged for a three-component 3D decomposition in [24] (extending the 2D work of [25]); however, their choice of space of discrete

vector fields (piecewise linear vector potential) did not lead to a cohomology-preserving discretization. Using DEC, instead, allows a discretization that captures the topological aspects correctly.

Our discrete Laplacian can really be seen as the counterpart of the continuous Laplacian, with particular boundary conditions (compatible with gauge conditions) added at the boundary to offer proper L_2-orthogonality of the various terms of the Hodge decomposition.

2.4 Extensions and specializations

While we described a discrete decomposition of vector fields given as 1- or 2-forms with a particularly canonical choice of gauges, we can extend our approach to different gauges in order to get different potentials, derive smaller, more specialized decompositions, or use non-diagonal Hodge star matrices without hindering efficiency.

2.4.1 Helmholtz decomposition

Our five-component decomposition can be trivially condensed into the two-component Helmholtz decomposition we described in Sec. 2.1.1: ∇f, $\mathbf{h}_n$, and η can all be expressed as a gradient field, and $\nabla \times \mathbf{u}$, $\mathbf{h}_t$, and η can be expressed as a curl field; no matter how we split the η term, we will get the expected decomposition of the type

$$\mathbf{v} = \nabla \phi + \nabla \times \boldsymbol{\psi}.$$

However, if one wishes to ensure the orthogonality between the two components, we must put η entirely in the gradient part (resulting in a tangential curl field), or entirely in the curl part (resulting in a normal gradient field).

2.4.2 Specialized inputs

In some contexts, we can assume the input to be a tangential or normal vector field. In these cases, it is possible to specialize our decomposition and make it a three-component or even two-component

41

decomposition instead. We provide a few examples to illustrate how this can be useful (and more efficient) in practice.

Tangential inputs   If we know that input vector field $\mathbf{v}$ is tangential, we can directly solve for a tangential vector field (called *Newtonian potential*) $\mathbf{w}$, such that the continuous decomposition becomes: $\mathbf{v} = \Delta \mathbf{w} + \mathbf{h}_t$. The discrete version is straightforward: one can solve for a 1-form w^1 or a 2-form w^2 based on the degree of the input form through:

$$L_{1,t} w_t^1 = S_1(V^1 - h_t^1) \quad \text{or} \quad L_{2,n} w_n^2 = P_2 S_2(V^2 - h_n^2),$$

since $\mathbf{v} - \mathbf{h}_t$ is in the image of Δ for tangential fields; note that the (tangential) harmonic part is computed directly by projection with the basis $\mathbb{H}$. This decomposition can be turned into a three-component decomposition as well through

$$\mathbf{v} = -\nabla(\nabla \cdot \mathbf{w}) + \nabla \times (\nabla \times \mathbf{w}) + \mathbf{h}_t.$$

We can further shift part of the curl field to the gradient field to make every component tangential: we can solve for the normal vector potential, and shift the rest to the gradient part.

Normal inputs   For normal vector fields, a similar approach leads to a two- or three-component decomposition:

$$\mathbf{v} = \Delta \mathbf{w} + \mathbf{h}_n \quad \text{or} \quad \mathbf{v} = -\nabla(\nabla \cdot \mathbf{w}) + \nabla \times (\nabla \times \mathbf{w}) + \mathbf{h}_n.$$

The discrete treatment to find the normal vector field $\mathbf{w}$ as a 1- or 2-form becomes then:

$$L_{1,n} w_n^1 = P_1 S_1(V^1 - h_n^1) \quad \text{or} \quad L_{2,t} w_t^2 = S_2(V^2 - h_n^2),$$

Again, one can shift part of the gradient field in the three-term decomposition to make the gradient field part normal to the boundary, which will make the curl field part normal to the boundary.
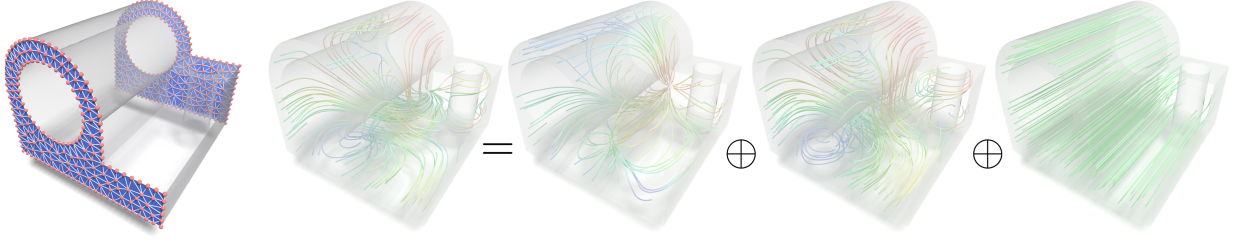
Figure 2.14: **Mixed boundary conditions.** Our orthogonal decomposition extends naturally to mixed boundary conditions as well; in this example, no constraints are set on the blue regions, but tangential conditions are set on the rest of the boundary.

2.4.3 Mixed boundary conditions

Mixed and/or partial boundary conditions are sometimes required. Orthogonal decomposition into gradient and curl fields with boundary conditions and topology-determined finite-dimensional harmonic space can be established in the same fashion through relative homologies. In general, the boundary is the disjoint union of tangential, normal, and unconstrained regions: $\partial M = \partial M_t \cup \partial M_n \cup \partial M_u$. Sticking to the original full decomposition will lead to some components not satisfying the boundary conditions. One can make each component follow the given boundary conditions by replacing the original boundary conditions to enforce, instead:

$$\star d\alpha|_{\partial M_t} = \star h_n|_{\partial M_t} = \star\eta|_{\partial M_t} = 0, \quad \delta\beta|_{\partial M_n} = h_t|_{\partial M_n} = \eta|_{\partial M_n} = 0,$$

$$\alpha|_{\partial M_n \cup \partial M_u} = h_n|_{\partial M_n \cup \partial M_u} = 0, \ \star\beta|_{\partial M_t \cup \partial M_u} = \star h_t|_{\partial M_t \cup \partial M_u} = 0.$$

From a practical standpoint, we simply have to impose the typical conditions on α, β, h_n and h_t on the unconstrained regions to provide orthogonality and the uniqueness of the decomposition; one can add η to any combination of the other components to create the "natural" unconstrained behavior for chosen components as we described for the three-component decomposition. In order to solve for the various potentials, we can define an altered Laplacian $L_{k,A}$ for k-forms that are normal on a boundary region $A \subset \partial M$ and tangential on $\partial M \backslash A$ through the following matrix expression:

$$L_{k,A} = D_{k,A}^T S_{k+1,A} D_{k,A} + S_{k,A} D_{k-1,A} S_{k-1,A}^{-1} D_{k-1,A}^T S_{k,A},$$

where $S_{k,A} = P_{k,A} S_k P_{k,A}^T$, $D_{k,A} = P_{k+1,A} D_k P_{k,A}^T$, and $P_{k,A}$ is the projection of a full k-form on M to a k-form restricted to $M \backslash A$. With this definition, we can solve for α using $L_{k-1,\partial M_n \cup \partial M_u}$, for β using $L_{k+1,\partial M_n}$, for h_n using $L_{k,\partial M_n \cup \partial M_u}$, and for h_t using $L_{k,\partial M_n}$.

Note that h_t and h_n are no longer necessarily L_2-orthogonal for general mixed boundary conditions as in the generic Hodge-Morrey-Friedrichs decomposition (Equation (2.6)). Nevertheless, we can either combine the two components and create an orthonormal basis for the span of both low dimensional subspaces, and/or combine one of them to the other components. Let's use the case $\partial M = \partial M_t \cup \partial M_u$ (i.e., where we only want to impose tangential forms on parts of the boundary) as an illustrative example. We propose the following decomposition by combining some of the components,

$$\omega^1 = d\alpha^0 + \delta\beta^2 + h_n^1$$

All three components are tangential on ∂M_t, α^0 is 0 on ∂M_u, β^2 is tangential on ∂M_t (i.e., the corresponding vector field is normal), and h_n^1 is normal on ∂M_u. Following the derivation of typical boundary conditions, we find that h_n^1 correspond to the relative homology $H_2(M, \partial M_t)$, or equivalently through its Hodge dual, to $H_1(M, \partial M_u)$. Thus, the space for h_n^1 is finite dimensional; Figure 2.14 shows such an example where it is two dimensional.

Finally, the orthogonality of the various terms of the resulting five-component decomposition is properly enforced. E.g.,

$$\langle d\alpha^0, h_n^1 \rangle = \langle \alpha^0, \delta h_n^1 \rangle + \int_{\partial M} \alpha^0 \wedge \star h_n^1,$$

which is zero since $\delta h_n^1 = 0$, the boundary integral on ∂M_u is 0 as $\alpha^0 = 0$ there, and the leftover integral on ∂M_t vanishes due to the boundary condition on $\star h_n^1$. The exact same argument holds for the $\delta\beta^2$ term, while the orthogonality between $\delta\beta^2$ and h_n^1 is established through similar arguments. Note that when ∂M_u is replaced by ∂M_n, the case $\partial M = \partial M_t \cup \partial M_n$ is recovered.

2.4.4 Friedrichs decompositions

Finally, we note that η in the four-component decomposition can be merged with $\mathbf{h}_t$ to form a subspace that is both harmonic and a curl field, or with $\mathbf{h}_n$ to form a subspace that is both harmonic

and a gradient field. Both four-component decompositions are sometimes called the Friedrichs decomposition.

2.4.5 Non-diagonal Hodge star

While the low-order "diagonal Hodge star" is often preferred in graphics due to its optimal sparsity [31], a variety of other discrete Hodge stars have been proposed [143]. Of particular interest are the Galerkin Hodge stars [144, 145] which offer higher-order accuracy of approximation, at the price of requiring still sparse, but non-diagonal matrix representations. As they are symmetric positive definite, our decompositions apply without modification. However, S_k^{-1} can become a dense matrix, making the evaluation of the Laplace matrices much less efficient. We outline a procedure that only uses sparse matrices for the decomposition to be still strictly L_2-orthogonal according to a non-diagonal Hodge star matrix S_k.

We first note that among the necessary discrete Laplacians for the decomposition of V^k ($k = 1, 2$), only $L_{k+1,t}$ involves S_k^{-1}. In other words, we can compute $D_{k-1}\alpha^{k-1}$, h_t^k, h_n^k, and $\eta = D_{k-1}\phi^{k-1}$ with the non-diagonal S_k. While it may be necessary to replace S_{k-1}^{-1} and S_{k-2}^{-1} by sparse substitute matrices $\tilde{S}_{k-1}$ and $\tilde{S}_{k-2}$ (e.g., identity matrices) to keep those systems sparse, it does not influence the actual accuracy of the decomposition: first, the L_2-orthogonality in Ω_k for the components depends on S_k, which is not altered; second, the harmonic spaces remain the same since the kernels remain in $\ker d \cap \ker \delta$ under normal/tangential boundary conditions; third, the potential α may deviate from satisfying $\delta\alpha = 0$ exactly, but the error lies within the gauge field, so $D_{k-1}\alpha^{k-1}$ is still accurate.

For the final component, note that $\gamma \equiv \omega - d\alpha - h_t - h_n - \eta$ is in $\operatorname{im} L_{k+1,t}$, so $S_{k+1}\beta = \gamma$ has a solution in $\Omega_{k+1,t}$. This means that

$$\left\langle D_k^T S_{k+1}\beta - S_k\gamma, D_k^T S_{k+1}\beta - S_k\gamma \right\rangle_{\tilde{S}_k}$$

can be minimized to exactly 0 in any weighted L_2-inner product $\langle \cdot \rangle_{\tilde{S}}$, where $\tilde{S}_k$ is an arbitrary

45

sparse SPD matrix. We can thus solve for the exact β without the inverse matrix S_k^{-1} through

$$(D_{k+1}^T S_{k+2} D_{k+1} + S_{k+1} D_k \tilde{S}_k D_k^T S_{k+1})\beta = S_{k+1} D_k \tilde{S}_k S_k \gamma.$$

If we take $\tilde{S}_k = S_k^{-1}$, the matrix is the same tangential Laplacian used for solving for β in our DEC decomposition; but we can now accommodate non-diagonal Hodge matrices as $\tilde{S}$ can be chosen arbitrarily: we will still find the exact potential satisfying $S_k^{-1} D_k^T S_{k+1}\beta = \gamma$. This construction extends to arbitrary Hodge stars the approach described in [146], where the authors realized that when the Galerkin Hodge star S_k (computed using Whitney forms, and thus non-diagonal) is multiplied by D_{k-1} on the right and D_{k-1}^T on the left, the result is no different than if the Galerkin Hodge star was replaced by a diagonal "lumped" matrix.

2.5 Experiments

Decomposition zoo. In Figure 2.1, we perform the full five-component vector field decomposition using a discrete 1-form representation. The connected domain contains one outside and one inside boundary components, with genus 1 and 0 respectively, thus $\beta_0 = 1, \beta_1 = 1, \beta_2 = 1, \beta_3 = 0$. We further evaluate the vector potential of the tangential harmonic component, the scalar potential of the normal harmonic component, and both potentials of the fifth (exact, coexact) component. We also numerically verified the L_2-orthogonality of the five terms. In Fig. 2.10, we provide a depiction of all the harmonic field basis vectors for a model with a more complex topology (two spherical and one toroidal cavities).

To demonstrate the non-orthogonality when no boundary condition is imposed, we show in Figure 1.1 a decomposition into the sum of a gradient field and a curl field, resulting from the five-component decomposition and after merging $d\alpha + h_n + \frac{1}{2}\eta$ and $\delta\beta + h_t + \frac{1}{2}\eta$ (or just summing up the potentials). Note that the L_2-inner product between the two is then $\frac{1}{4}\langle \eta, \eta \rangle$.

When the input is a tangential field as in Figure 2.3, its Helmholtz-Hodge decomposition contains three tangential fields, the gradient field $d\alpha + h_n + \eta$, the curl field $\delta\beta$, and the tangential harmonic field (non-integrable in the sense that it cannot be seen as the curl of a *normal* vector potential). It is also possible to obtain either one of the two four-component Hodge-Morrey-Friedrichs
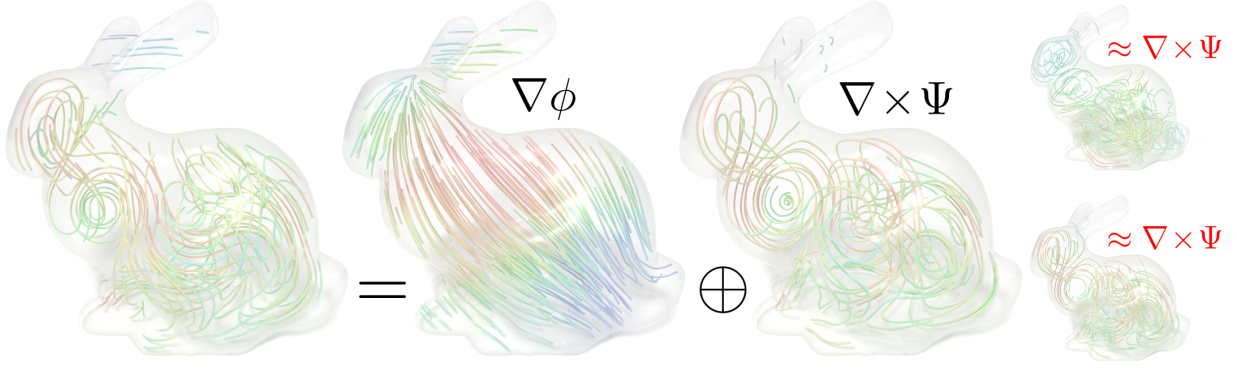
Figure 2.15: **Non-diagonal Hodge star.** Even for higher-order accurate Hodge stars, our decomposition still only requires sparse linear systems. Using a diagonal S_1 in the Laplacian produces inaccurate potentials (right), whether we use a curl operator with a diagonal (top) or non-diagonal (bottom) S_1.

decompositions; e.g., in Figure 2.4, we decompose the input into $d\alpha$, $\delta\beta$, h_t, and a harmonic gradient field $h_n+\eta$, which is harmonic with a scalar potential.

In Figure 2.14 for mixed boundary conditions, we create a case mimicking fluids passing through a domain with two openings. As described in Section 2.4.3, the normal harmonic space (vector field normal to the unrestricted boundary and tangential to the tangential boundary condition region) is two dimensional, which can also be constructed through eigensolvers or through the corresponding relative homology. The gradient component can be constructed by solving a Poisson equation with the divergence of the input on the right hand side, and the same tangential boundary conditions. The rest can be expressed as the curl of a vector potential that is orthogonal to the boundary outside the openings.

Using the Galerkin Hodge star associated with Whitney basis functions [139], the potential β for the $\delta\beta$ term in Figure 2.15 is accurate with our approach. If the diagonal Hodge star S_D is used instead in the Laplacian to compute a different potential $\tilde{\beta}$, then $S_1^{-1}D_1^T\tilde{\beta}$ has a deviation from the γ term defined in Sec. 2.4.5 of around 48% (Fig. 2.15(top right)), but it still is orthogonal to the other components; if one tries $S_D^{-1}D_1^T\tilde{\beta}$ for consistency, then there is still a 1.5% deviation from γ (Fig. 2.15(bottom right)) and a 1.5% error in L_2-inner product with the other components is now present.

Our decomposition is also demonstrated on a simulated channel flow. The velocity field was generated with the OpenFOAM software, with a forced velocity on the round inlet and outlet, with free-slip and no-transfer boundary conditions on the interior walls. Our decomposition thus sets all regions away from the inlet and outlet with tangential conditions.

Performance and accuracy. For completeness, we also tested the assembly of the matrices on a laptop. For models with around 25K tets, we can perform the necessary solves using Conjugate Gradient in 2s even on a regular laptop with our unoptimized code. Note that if we prefactorize (through Cholesky decomposition) the Laplacians, we can much more efficiently perform the decomposition of arbitrary fields on the same domain through forward and backward substitutions, in less than a second. As shown in Figure 2.11, our condition number based strategy to choose rows and columns to eliminate the null space of the Laplacian matrices is very effective in maintaining the accuracy of the linear system. Note that when working with non-diagonal Hodge stars, we can also either use Conjugate Gradient or precompute a Cholesky decomposition for the evaluation of the curl in (Figure 2.15): for a 10K tet mesh, the iterative CG solve takes less than 1s, whereas the Cholesky factorization of the non-diagonal Hodge star takes 5s—but allows fast repeated evaluations.
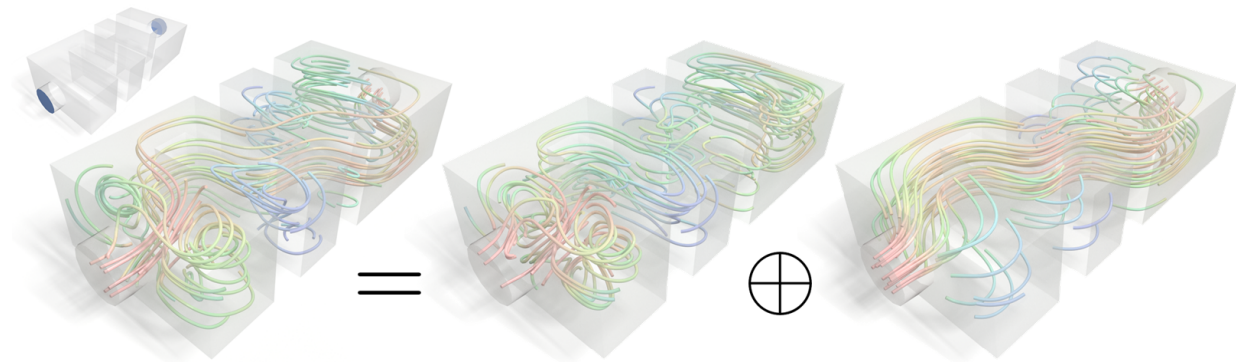
Figure 2.16: **Decomposition of a channel flow simulation.** For a simulated channel flow (inlet and outlet in blue), the resulting vector field is decomposed into a curl field and a harmonic field, with the blue regions are set as unconstrained and all other boundary regions as tangential.å

Comparison to [3]. The only other existing 5-component 3D decomposition and our approach are based on very different discretization methods: Poelke represents a discrete vector field as piecewise constant per tetrahedron, while we use discrete 1- or 2-forms. In this sense, our work is complementary. Yet, and while cohomologies are preserved in both approaches, our representation also requires fewer DoFs as input, as the number of edges or faces is always smaller than three times the number of tets. Our approach also tackles the full 5-component decomposition using *only symmetric semi-positive definite matrices with smaller sizes*, resulting in higher computational efficiency: numerical experiments confirm that differential form based discretization leads to better accuracy, partially due to their exact line integral and flux sampling (i.e., linear precision vs piecewise-constant precision of the representation). Moreover, it is straightforward for us to formulate the relation of mixed boundary conditions to relative cohomologies, or to extend our construction using a higher-order L_2-inner product. Finally, our eigensolver also produces L_2-orthonormal basis for the cohomology more efficiently than the non-L_2-orthonormal basis obtained in [3] through singular value decomposition of rectangular matrices.

DE RHAM-HODGE ANALYSIS AND MODELING OF BIOMOLECULES

3.1 Theoretical modeling and analysis

This section introduces de Rham-Hodge theory for the analysis of biomolecules. To establish notation, we provide a brief review of de Rham-Hodge theory. Then, we introduce topological structure-preserving analysis tools, such as discrete exterior calculus (DEC) [70], discretized differential forms, and Hodge-Laplacians, on the compact manifolds enclosing biomolecular boundaries. We use simple finite-dimensional linear algebra to computationally realize our structure-preserving analysis on various differential forms. We construct appropriate physically-relevant boundary conditions on biomolecular manifolds to facilitate various scalar and vector Laplace-de Rham operators such that the resulting spectral bases are consistent with three basic singular value decompositions of the gradient, curl and divergence operators through dualities.

3.1.1 De Rham-Hodge theory for macromolecules

While the spectral analysis can be carried out using scalar, vector and tensor calculus, differential forms and exterior calculus are required in de Rham-Hodge theory to reveal the intrinsic relations between differential geometry and algebraic topology on biomolecular manifolds. Since biomolecular shapes can be described as 3-manifolds with a 2-manifold boundary in the 3D Euclidean space, we represent scalar and vector fields on molecular manifolds as well as their interconversion through differential forms. As a generalization of line integral and flux calculation of vector fields, a differential k-form $\omega^k \in \Omega^k(M)$ is a field that can be integrated on a k-dimensional submanifold of M, which can be mathematically defined through a rank-k antisymmetric tensor defined on a manifold M. By treating it as a multi-linear map from k vectors spanning the tangent space to a scalar, it turns an infinitesimal k-dimensional cell into a scalar, whose sum over all cells in a tessellation of a k-dimensional submanifold produces the integral in the limit of infinitesimal cell size. In $\mathbb{R}^3$,

0-forms and 3-forms have one degree of freedom at each point and can be regarded as scalar fields, while 1-forms and 2-forms have three degrees of freedom, and can be interpreted as vector fields.

The differential operator (also called exterior derivative) d can be seen as a unified operator that corresponds to gradient (∇), curl ($\nabla \times$) , and divergence ($\nabla \cdot$) when applied to 0-, 1-, and 2-forms, mapping them to 1-, 2-, and 3-forms, respectively. On a boundaryless manifold, a codifferential operator δ is the adjoint operator under L_2-inner product of the fields (integral of pointwise inner product over the whole manifold), which corresponds to $-\nabla \cdot$, $\nabla \times$, and $-\nabla$, for 1-, 2-, and 3-forms, mapping them to 0-, 1-, and 2-forms, respectively.

One key property of $d : \Omega^k(M) \rightarrow \Omega^{k+1}(M)$ is that $dd = 0$, which allows the space of differential forms Ω^k to form a chain complex, which is called the de Rham complex

$$0 \rightarrow \Omega^0(M) \xrightarrow[(\nabla)]{d} \Omega^1(M) \xrightarrow[(\nabla \times)]{d} \Omega^2(M) \xrightarrow[(\nabla \cdot)]{d} \Omega^3(M) \xrightarrow{d} 0. \tag{3.1}$$

It also matches the identities of second derivatives for vector calculus in $\mathbb{R}^3$, i.e., $(\nabla \times)\nabla = 0$ and $(\nabla \cdot)\nabla \times = 0$. The topological property associated with differential forms is given by the de Rham cohomology,

$$H_{dR}^k(M) = \frac{\ker d^k}{\mathrm{im} d^{k-1}}. \tag{3.2}$$

The de Rham theorem states that the de Rham cohomology is isomorphic to the singular cohomology, which is derived purely from the topology of the biomolecular manifold.

The Hodge k-star $\star^k$ (also called Hodge dual) is a linear map from a k-from to its dual form, $\star^k : \Omega^k(M) \rightarrow \Omega^{n-k}(M)$. Given two k-forms $\alpha, \beta \in \Omega^k(M)$, the ($L_2$-)inner product between them can be defined along with star operator as

$$\langle \alpha, \beta \rangle = \int_M \alpha \wedge \star\beta = \int_M \beta \wedge \star\alpha. \tag{3.3}$$

Under the inner products, the adjoint operators of d are the codifferential operators $\delta^k : \Omega^k(M) \rightarrow \Omega^{k-1}(M)$, $\delta^k = (-1)^k \star d\star$ satisfies $\delta\delta = 0$. Hodge further established the isomorphism

$$H_{dR}^k(M) \cong H_\Delta^k(M), \tag{3.4}$$

where $H_\Delta^k(M) = \{\omega | \Delta\omega = 0\}$ is the kernel of the Laplace-de Rham operator $\Delta \equiv d\delta + \delta d = (d + \delta)^2$, also known as the space of harmonic forms. A corollary of the result is the Hodge decomposition,

$$\omega = d\alpha + \delta\beta + h, \tag{3.5}$$

which is an L_2-orthogonal decomposition of any form ω into d and δ of two potential fields $\alpha \in \Omega^{k-1}(M)$ and $\beta \in \Omega^{k+1}(M)$ respectively, and a harmonic form $h \in H_\Delta^k(M)$. This means that harmonic forms are the non-integrable parts of differential forms, which form a finite dimensional space determined by the topology of the biomolecular domain due to de Rham's and Hodge's theorems.

3.1.2 Macromolecular spectral analysis

The Laplace-de Rham operator $\Delta = d\delta + \delta d$, when restricted to a 3D object embedded in the 3D Euclidean space, is simply $-\nabla^2$. As it is a self-adjoint operator with a finite dimensional kernel, it can be used to build spectral bases for differential forms. For irregularly shaped objects, these bases can be very complicated. However, for a simple geometry, these bases are well-known functions. For example, 0-forms on a unit circle can be expressed as the linear combination of sine and cosine functions, which are eigenfucntions of the Laplacian for 0-forms Δ_0. Similarly, spherical harmonics are eigenfunctions of Δ_0 on a sphere and it has also been extended to manifold harmonics on Riemannian 2-manifolds.

We further extend the analysis to any rank k and to 3D shapes such as macromolecular shapes where analysis can be carried out in two types of cases. In the first type, one may treat the surface of the molecular shape as a boundaryless compact manifold and analyzes any field defined on such a 2D surface. In fact, this approach is relevant to protein surface electrostatic potentials or the behavior of cell membrane or mitochondrial ultrastructure. In this work, we shall restrain from any further exploration in this direction. In the second type, we consider the volumetric data enclosed by a macromolecular surface. As a result, the molecular shape has a boundary. In this setting, the harmonic space becomes infinite-dimensional unless certain boundary conditions are enforced. In

particular, tangential or normal boundary conditions (also called Dirichlet or Neumann boundary conditions, respectively) are enforced to turn the harmonic space into a finite-dimensional space corresponding to algebraic topology constructions that lead to absolute and relative homologies.

We first discuss the natural separation of the eigenbasis functions into curl-free and div-free fields in the continuous theory, assuming that the boundary condition is implicitly enforced, before providing details on the discrete exterior calculus with the boundary taken into consideration.

Given any eigenfield ω of the Laplacian,

$$\Delta\omega = \lambda\omega, \tag{3.6}$$

we can decompose it into $\omega = d\alpha + \delta\beta + h$. For $\lambda \neq 0$, $h = 0$, based on $dd = 0$ and $\delta\delta = 0$, it is easy to see that both $d\alpha$ and $\delta\beta$ are eigenfunctions of Δ with eigenvalue λ due to the uniqueness of the decomposition, unless one of them is 0. It is typically the case that ω is either a curl field or a gradient field, otherwise, λ has a multiplicity of at least 2, in which case both eigenfields associated with λ are the linear combinations of the same pair of the gradient field and the curl field.

3.1.3 Discrete spectral analysis of differential forms

In a simplicial tessellation of a manifold mesh, d^k is implemented as a matrix D_k, which is a signed incidence matrix between $(k+1)$-simplices and k-simplices. We provide the details in Sec. 3.3, but the defining property in de Rham-Hodge theory is preserved through such a discretization: $D_{k+1}D_k = 0$. The adjoint operator δ^k is implemented as $S_{k-1}^{-1}D_{k-1}^T S_k$, where S_k is a mapping from a discrete k-form to a discrete $(n-k)$-form on the dual mesh, which can be treated as a discretization of the L_2-inner product of k-forms. As S_k is always a symmetric positive matrix, the L_2-inner product between two discrete k-forms can be expressed as $(\omega_1^k)^T S_k \omega_2^k$. The discrete Hodge Laplacian maps a discrete k-form to a discrete $n-k$-form which is defined as

$$L_k = D_k^T S_{k+1} D_k + S_k D_{k-1} S_{k-1}^{-1} D_{k-1}^T S_k, \tag{3.7}$$

which is a symmetric matrix and $S_k^{-1}L_k$ corresponds to Δ_k. The eigenbasis functions are found through a generalized eigenvalue problem,

$$L_k\omega^k = \lambda^k S_k\omega^k. \tag{3.8}$$

Depending on whether the tangential or normal boundary condition is enforced, D_k includes or excludes the boundary elements respectively. Thus, the boundary condition is built into discrete linear operators. When we need to distinguish these two cases, we use $L_{k,t}$ and $L_{k,n}$ to denote the tangential and normal boundary conditions respectively.

In general, it is not necessarily efficient to take the square root of the discrete Hodge star operator, $S_k^{\frac{1}{2}}$ or to compute its inverse, S_k^{-1}. However, for analysis, we can always convert a generalized eigenvalue problem in Eq. (3.8) into a regular eigenvalue problem,

$$\bar{L}_k\bar{\omega}^k \equiv S_k^{-\frac{1}{2}}L_kS_k^{-\frac{1}{2}}\bar{\omega}^k = \lambda^k\bar{\omega}^k, \tag{3.9}$$

where $\bar{\omega} \equiv S_k^{\frac{1}{2}}\omega$. We can further rewrite the symmetrically modified Hodge Laplacian as

$$\bar{L}_k = \bar{D}_k^T\bar{D}_k + \bar{D}_{k-1}\bar{D}_{k-1}^T, \tag{3.10}$$

where $\bar{D}_k \equiv S_{k+1}^{\frac{1}{2}}D_kS_k^{-\frac{1}{2}}$ must satisfy $\bar{D}_{k+1}\bar{D}_k = 0$. Now the L_2-inner product between two discrete differential forms in the modified space is simply $(\bar{\omega}_1^k)^T\bar{\omega}_2^k$, and the adjoint operator of $\bar{D}_k$ is simply $\bar{D}_k^T$.

Now the partitioning of the eigenbasis functions into harmonic fields, gradient fields, and curl fields for 1-forms and 2-forms and their relationship can be understood from the singular value decomposition of the differential operator

$$\bar{D}_k = U_{k+1}\Sigma_kV_k^T, \tag{3.11}$$

where U_{k+1} and V_k are orthogonal matrices, and Σ_k is a rectangular matrix that only has nonzero entries on the diagonal, which can be sorted in ascending order as $\sqrt{\lambda_i^k}$ with trailing zeros. As the Hodge decomposition is an orthogonal decomposition, each column of V_k that corresponds to a nonzero singular value $\sqrt{\lambda_i^k}$ is orthogonal to any column of U_k that corresponds to a nonzero
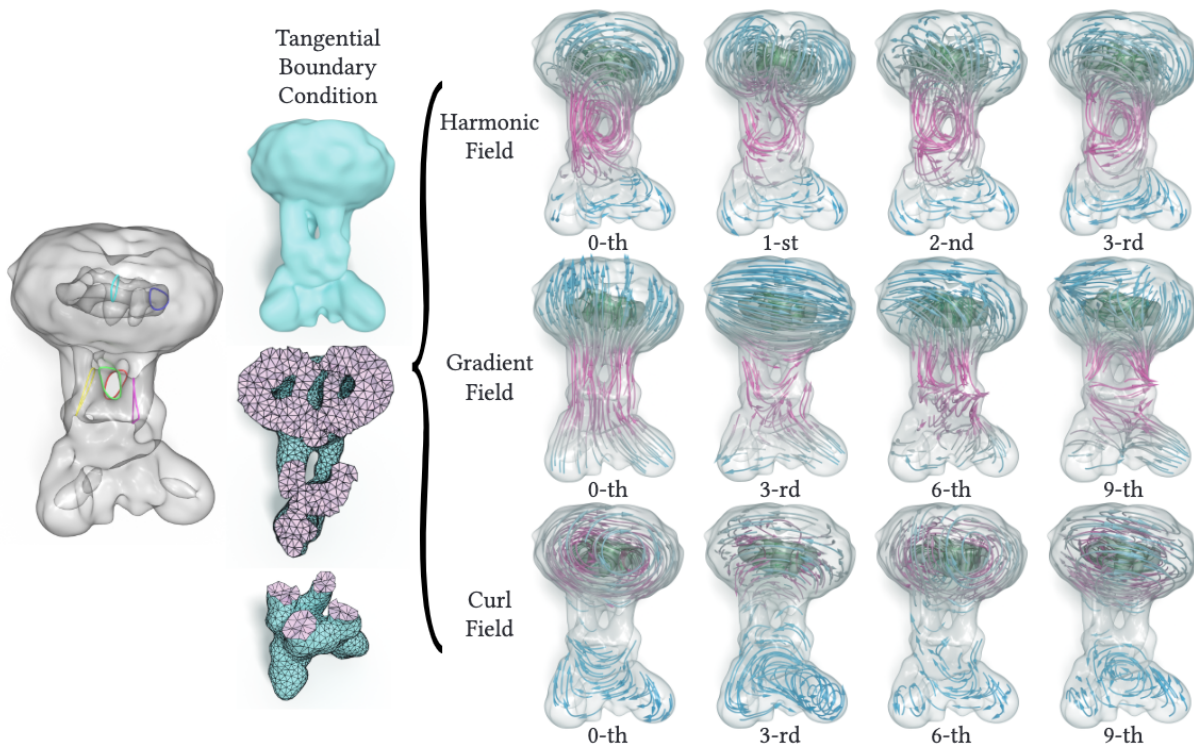
Figure 3.1: **Illustration of tangential spectra of a cryo-EM map EMD 7972** Topologically, EMD 7972 [4] has 6 handles and 2 cavities. The left column is the original shape and its anatomy showing the topological complexity. On the right-hand side of the parenthesis, the first row shows tangential harmonic eigenfields, the second row shows tangential gradient eigenfields, and the third row shows tangential curl eigenfields. The credit for the leftmost picture belongs to Hayam Mohamed.

$\sqrt{\lambda_j^{k-1}}$. Here V_k and U_k, together with the finite dimensional set of harmonic forms h_k (which satisfy both $D_k h_k = 0$ and $D_{k-1}^T h_k = 0$), span the entire space of k-forms. Moreover, the spectrum (i.e., set of eigenvalues) of the symmetric modified Hodge Laplacian in Eq. (3.10) consists of 0s, the set of λ_i^k's, and the set of λ_j^{k-1}'s. Note that, in the spectral basis, taking derivatives $\bar{D}$ (or $\bar{D}^T$) is simply performed through multiplying the corresponding singular values, and integration is done through division by the corresponding singular values, mimicking the situation in the traditional Fourier analysis for scalar fields.

3.1.4 Boundary conditions and dualities in 3D molecular manifolds

Overall, appropriate boundary conditions are prescribed to preserve the orthogonal property of the Hodge decomposition. In 3D molecular manifolds, 0- and 3-forms can be seen as scalar fields and 1-

and 2-forms as vector fields. For the spectral analysis of scalar fields (0-forms or 3-forms), two types of typical boundary conditions are used: Dirichlet boundary condition $f|_{\partial M} = f_0$ and Neumann boundary condition $\mathbf{n} \cdot \nabla f|_{\partial M} = g_0$, where f_0 and g_0 are functions on the boundary ∂M and $\mathbf{n}$ is the unit normal on the boundary. For spectral analysis, harmonic fields satisfying the arbitrary boundary conditions can be dealt with through spectral analysis of f_0 or g_0 on the boundary, and the following boundary conditions are used for the volumetric function f. The normal 0-forms (tangential 3-forms) satisfy

$$f|_{\partial M} = 0, \tag{3.12}$$

and the tangential 0-forms (normal 3-forms) satisfy

$$\mathbf{n} \cdot \nabla f|_{\partial M} = 0. \tag{3.13}$$

For the spectral analysis of vector fields, boundary conditions are for the three components of the field. Based on the de Rham-Hodge theory, it is more convenient to also use two types of boundary conditions. For tangential vector field (representing tangential 1-forms or normal 2-forms) $\mathbf{v}$, we use the Dirichlet boundary condition for the normal component and the Neumann condition for the tangential components:

$$\mathbf{v} \cdot \mathbf{n} = 0, \quad \mathbf{n} \cdot \nabla(\mathbf{v} \cdot \mathbf{t}_1) = 0, \quad \mathbf{n} \cdot \nabla(\mathbf{v} \cdot \mathbf{t}_2) = 0, \tag{3.14}$$

where $\mathbf{t}_1$ and $\mathbf{t}_2$ are two local tangent directions forming a coordinate frame with the unit normal $\mathbf{n}$. The corresponding spectral fields are shown in Fig 3.1. For normal vector field (representing normal 1-forms or tangential 2-forms) $\mathbf{v}$, we use the Neumann boundary condition on the normal component, and the Dirichlet boundary condition on the tangential components:

$$\mathbf{v} \cdot \mathbf{t}_1 = 0, \quad \mathbf{v} \cdot \mathbf{t}_2 = 0, \quad \mathbf{n} \cdot \nabla(\mathbf{v} \cdot \mathbf{n}) = 0. \tag{3.15}$$

The corresponding spectral fields are shown in Fig. 3.2. Aside from the harmonic spectral fields, there are two types of fields involved for the spectral fields of both boundary conditions—the set of divergence-free fields (also called curl fields) and the set of curl-free fields (also called gradient

fields). In summary, the above four boundary conditions account for both types of boundary conditions of all four differential forms, since the tangential boundary conditions of k-forms are equivalent to the normal boundary conditions of $n-k$-forms.

3.1.5 Reduction and analysis

For the four types of k-forms ($k \in \{0, 1, 2, 3\}$ in $\mathbb{R}^3$) in combinations with the two types of boundary conditions (tangential and normal), there are 8 different Laplace-de Rham operators ($L_{k,t}$ and $L_{k,n}$) in total. However, based on Eq. (3.10), the nonzero-parts of the spectrum L_k can be assembled from the singular values of $\bar{D}_k$ and $\bar{D}_{k-1}$. Thus, for each type of boundary conditions, there are only three spectra associated with $\bar{D}_0$, $\bar{D}_1$, and $\bar{D}_2$, since $\bar{D}_3 = 0$ for 3D space (one still has eight Laplace-de Rham operators). Moreover, according to the Hodge duality discussed in above paragraph, there is a one-to-one mapping between tangential k-forms and normal $(3-k)$-forms, which further identifies $\bar{D}_{0,t}$ with $\bar{D}_{2,n}^T$, $\bar{D}_{0,n}$ with $\bar{D}_{2,t}^T$, and $\bar{D}_{1,t}$ with $\bar{D}_{1,n}^T$. As a result, one has four independent Laplace-de Rham operators. Finally, due to the self-adjointness, there are only 3 intrinsically different spectra: 1) The first contains singular values of the gradient operator $D_{0,t}$ on tangential scalar potential fields (or equivalently, the singular values of the divergence operator $D_{2,n}$ on tangential gradient fields) as shown in Fig. 3.3 **b**; 2) The second contains singular values of the gradient operator $D_{0,n}$ on normal scalar potential fields (or equivalently, the singular values of the divergence operator $D_{2,t}$ on normal gradient fields) as shown in Fig. 3.3 **c**; 3) The third contains singular values of the curl operator $D_{1,t}$ applied to tangential curl fields (or equivalently, the singular values of the curl operator $D_{1,n}$ applied to normal curl fields) as shown in Fig. 3.3 **d**.

As the discussed above, each of the 8 Hodge Laplacians defined for smooth fields on a smooth shape has a spectrum that is simply the combination of one or two of the 3 sets of singular values along with possibly a 0. However, the numerical evaluation of the singular values of the differential operators for tangential k-forms $\bar{D}_{k,t}$ can differ from those of the discrete operators for normal $3-k$-forms $\bar{D}_{2-k,n}^T$, as shown in Fig. 3.3 **d**. One immediate reason is that the degrees of freedom (DoFs) associated for tangential/normal scalar/vector fields represented as tangential forms are not

Normal
Boundary
Condition

Gradient
Fields

0-th 1-st

2-nd 3-rd

Curl
Fields

0-th 1-st

2-nd 3-rd
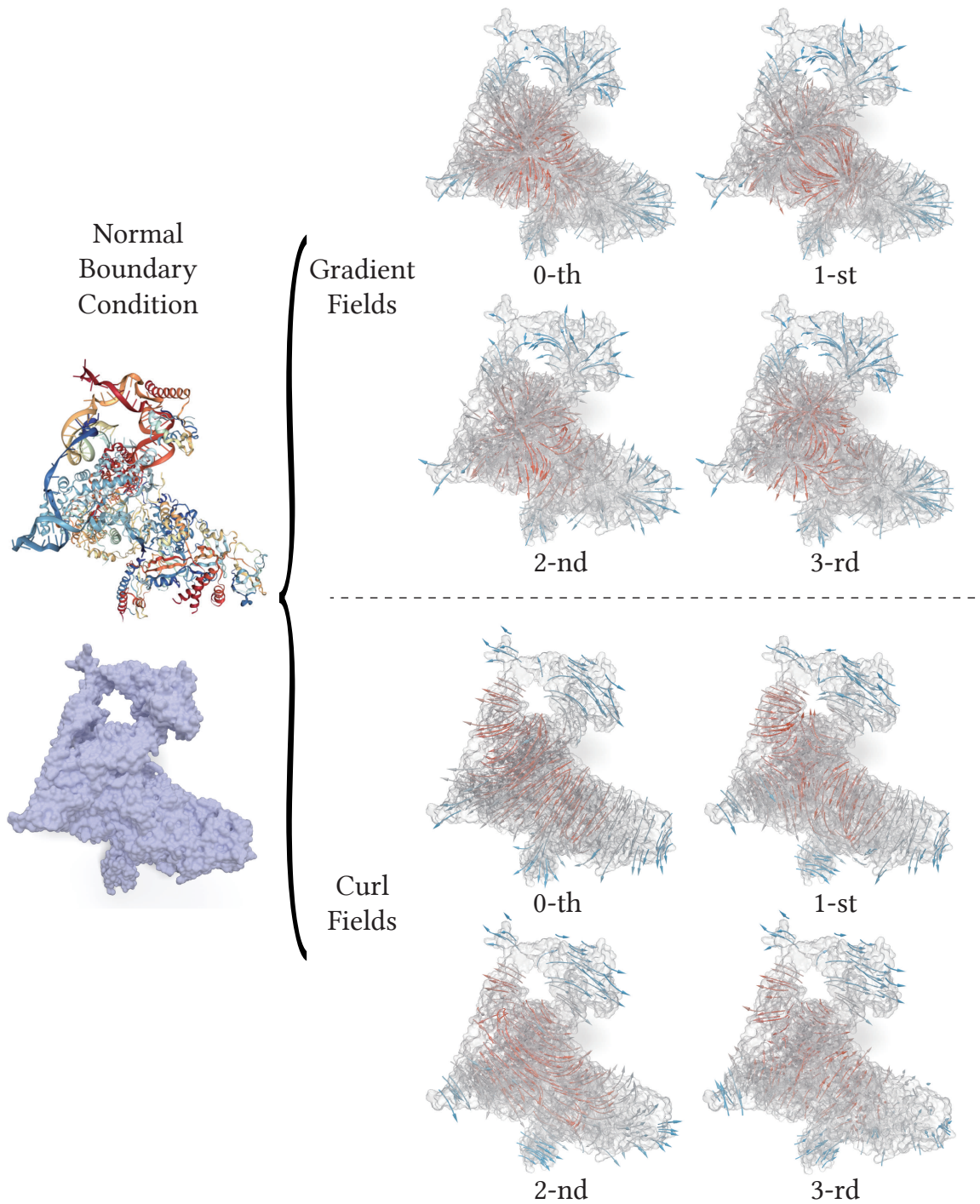
Figure 3.2: **Illustration of the normal spectra of protein and DNA complex 6D6V** Topologically, the crystal structure of 6D6V [5] has 1 handle. The left column shows the secondary structure and the solvent excluded surface (SES). On the right-hand side, the first two rows show normal gradient eigenfields, and the last two rows show normal curl eigenfields.
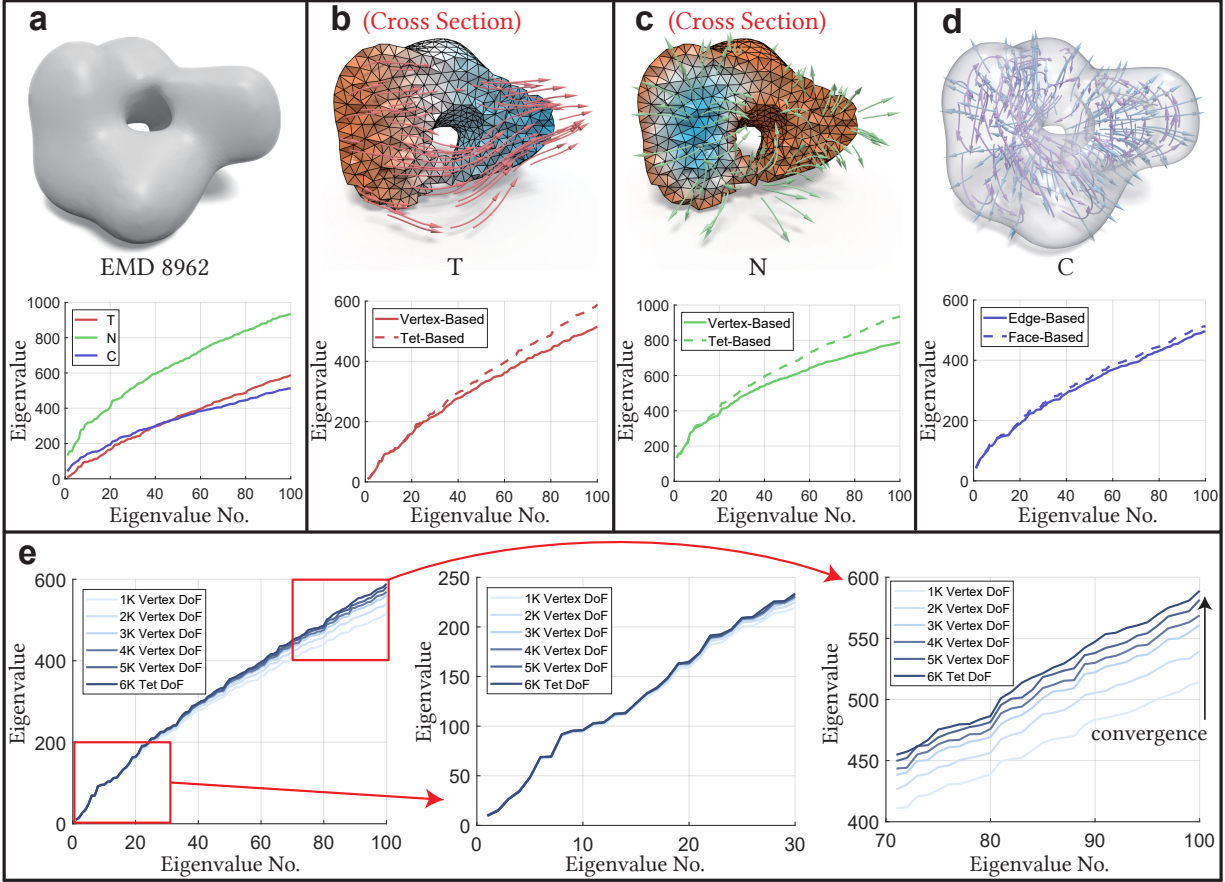
Figure 3.3: **Illustration of Hodge Laplacian spectra** This figure shows the properties of 3 spectral groups, namely, tangential gradient eigenfields (T), normal gradient eigenfields (N), and curl eigenfields (C), for EMD 8962 [6]. **a** shows the original input surface and 3 distinct spectral groups. **b** shows the cross section of a typical tangential gradient eigenfield and the distribution of eigenvalues for group T. **c** shows the cross section of a typical normal gradient eigenfield and the distribution of eigenvalues for group N. **d** shows a typical curl eigenfield and the distribution of eigenvalues for group C. **e** The left chart shows the convergence of spectra in the same spectral group due to the increase in the mesh size, i.e., the DoFs from 1,000 (1K) to 6,000 (6K). Obviously, low order eigenvalues converge fast (middle chart) and high order eigenvalues converge slowly (right chart).

the same as those represented by normal forms on a given tessellation, leading to different sampling accuracies. For example, the tessellation of the shape in Fig. 3.3 consists of approximately $1,000$ vertices, $7,000$ edges, $10,000$ triangles and $5,000$ tetrahedra. Thus, each tangential 0-form only has $1,000$ DoFs, and each normal 3-form has $5,000$. Hence, $L_{3,n}$ is capable of handling higher frequency signals in any given smooth scalar field than $L_{0,t}$ when we approach the Nyquist frequencies of the sampling. The convergence of both discretizations for the same continuous operator can be observed with increasing DoFs for both differential forms under refinement of the tet meshes (Fig. 3.3 **e** left). For low frequencies (smallest eigenvalues), there is a good agreement to begin with (Fig. 3.3 **e** mid), while for any given high frequency, the convergence with increased resolutions can be clearly observed (Fig. 3.3 **e** right).

On the other hand, $\bar{D}_k \bar{D}_k^T$ and $\bar{D}_k^T \bar{D}_k$ will have strictly the same set of nonzero eigenvalues. For instance, the spectrum of $L_{0,t}$ and the partial spectrum of $L_{1,t}$ that corresponds to gradient fields are identical, since $\bar{D}_{0,t} \bar{D}_{0,t}^T$ and $\bar{D}_{0,t}^T \bar{D}_{0,t}$ have the same nonzero eigenvalues.

For eigenfields vector Laplacians represented as 1-forms or 2-forms, i.e. the eigenfields of L_1 or L_2, we can observe some typical traits in the distributions of eigenvalues under normal and tangential boundary conditions. The normal boundary condition tends to allow more gradient eigenfields associated with eigenvalues below a given threshold than those under the tangential boundary condition for eigenvalues below the same threshold. We conjecture that it is due to the more stringent Dirichlet boundary condition on the potential scalar fields than the Neumann boundary condition on the potential scalar fields. The relation between the tangential boundary condition gradient-type eigenfields and curl-type eigenfields for low-frequency range seems to be highly dependent on the shape (see Fig. 3.3 **b** and **d**). Fig. 3.1 shows different vector eigenfields for tangential boundary condition with EMD 7972 surface. The first row shows different harmonic fields corresponding to the number of handles of the shape, the second row shows different gradient fields and the third row shows different curl fields. Fig. 3.2 shows different vector eigenfields for normal boundary condition with the protein and DNA complex crystal structure 6D6V. Since there are no cavities for this shape, there are no harmonic fields. The first row shows different gradient fields and the

second row shows different curl fields. Note that the scalar potentials for gradient fields and the vector potentials for curl fields are also themselves eigenfields associated with the same eigenvalues, although for different Laplacians.

Summarizing the above discussion on the properties of Laplacian spectra for 3D shape, we propose the following suggestions for practical spectral analysis.

- Only 3 independent spectra (e.g., singular values of $D_{0,t}$, $D_{1,t}$, and $D_{2,t}$) are necessary to avoid redundancy.

- Laplace-de Rham operators with higher DoFs can be used for more accurate calculation (at a higher computational cost) given the same tessellation.

- When computing eigenvalues given the same high-frequency truncation threshold, the differences in the numbers of eigenvalues in the 3 spectra vary with the shape.

3.2   Macromolecular modeling and analysis

Biological macromolecules and their complexes offer a rich variety of geometric and topological features, which often exhibit close relations with their functionalities. For instance, protein pockets can often be identified as a geometrically concave region on the protein surface, or as a topological cavity of an offset surface. Ion channels that regulate important biological functions can be usually associated with a topological tunnel. Mitochondrial ultrastructures admit various geometric and topological complexity which is related to their functions [147]. Hence, a unified approach for quantitatively analyzing such geometric and topological features is in great need. Our de Rham-Hodge analysis and Laplace-de Rham operator modeling provide such a unified approach for capturing both geometric and topological features simultaneously.

Our de Rham-Hodge analysis offers a powerful new tool for characterizing macromolecular geometry, identifying macromolecular topology, and modeling macromolecular structural flexibility and collective motion. We have carried out extensive computational experiments using protein

structural datasets and cryo-EM maps to demonstrate the utility and usefulness of the proposed de Rham-Hodge tools and models.

3.2.1 Molecular shape generation

The geometric modeling of macromolecular 3D shapes bridges the gap between experimental data and theoretical models for macromolecular function, dynamics, and transport. To carry out our de Rham-Hodge analysis on a macromolecule or a protein complex, we need a given domain containing the 3D macromolecular shape. Theoretically, such a domain for a macromolecule can be generated by taking an isosurface of a cryo-EM map or constructed from the atomic coordinates of the macromolecule. For a given set of atomic coordinates $\mathbf{r}_i, i = 1, 2, \cdots, N$, van der Waals surface, solvent accessible surface, and the solvent excluded surface can be constructed. However, these surfaces are typically singular, leading to computational instability for de Rham-Hodge analysis. Alternatively, minimal molecular surface (MMS) generated by differential geometry, Gaussian surface [41], and flexibility rigidity index (FRI) surface [44, 1] are computationally preferred and used widely in many studies. In fact, FRI surface is simpler than MMS and more stable than Gaussian surface [45]. To generate an FRI surface, we use a discrete-to-continuum mapping to define an unnormalized molecular density [44, 45]

$$\rho(\mathbf{r}, \eta) = \sum_{j=1}^{N} \Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta) \tag{3.16}$$

where η is a scale parameter and in this paper, it is set to twice of the atomic van der Waals radius r_j. Φ is density estimator that satisfies the following admissibility conditions

$$\Phi\left(\|\mathbf{r} - \mathbf{r}_j\|; \eta\|\right) = 1, \quad \text{as} \quad \|\mathbf{r} - \mathbf{r}_j\| \to 0, \tag{3.17}$$

$$\Phi\left(\|\mathbf{r} - \mathbf{r}_j\|; \eta\|\right) = 0, \quad \text{as} \quad \|\mathbf{r} - \mathbf{r}_j\| \to \infty. \tag{3.18}$$

Monotonically decaying radial basis functions are all admissible. Commonly used correlation kernels include generalized exponential functions

$$\Phi\left(\|\mathbf{r} - \mathbf{r}_j\|; \eta\|\right) = e^{-\left(\|\mathbf{r} - \mathbf{r}_j\|/\eta\right)^{\kappa}}, \quad \kappa > 0; \tag{3.19}$$

and generalized Lorentz functions

$$\Phi\left(\|\mathbf{r} - \mathbf{r}_j\|; \eta\right) = \frac{1}{1 + \left(\|\mathbf{r} - \mathbf{r}_j\|/\eta\right)^\nu}, \quad \nu > 0. \tag{3.20}$$

The Gaussian kernel ($\kappa = 2$) is employed in this work.

A family of biomolecular domains can be defined by varying level set parameter $c > 0$

$$M = \{\mathbf{r}|\rho(\mathbf{r}, \eta) \geq c\}. \tag{3.21}$$

3.2.2 Topological analysis

In this work, we discuss topology in the mathematical sense. Therefore, topological features are those stable structural characteristics that do not change with deformation, such as the number of connected components, the number of holes on each connected components, and the number of cavities. They are captured in the null spaces of the corresponding Laplace-de Rham operators. In other words, the invariant spaces associated with the eigenvalue of 0, i.e., the lowest ends of the spectra. Specifically, the dimension of the null space of $L_{1,t}$ and $L_{2,n}$ is the same as the number of tunnels as shown in Fig. 3.4 **a**. The dimension of the null space of $L_{1,n}$ and $L_{2,t}$ provides the number of cavities as shown in Fig. 3.4 **b**. The dimension of the $L_{0,t}$ is equal to the number of connected components. In persistent homology, the geometric measurement for characterizing the persistence of a topological feature has been proven crucial to the practical use of these otherwise overly stable features. The eigenfields associated with the eigenvalue 0 in our spectral analysis can also provide such information. For instance, the strength of the eigenfield associated with the eigenvalue 0 for $L_{1,t}$ can indicate how narrow the handle/tunnel is in the region. In the tangential harmonic fields of Fig. 3.1, the colors show the strength of eigenfields such that red colors stand for high strengths and indigo colors stand for low strengths. One can see that strengths are higher in middle narrow tunnels than top and bottom parts.
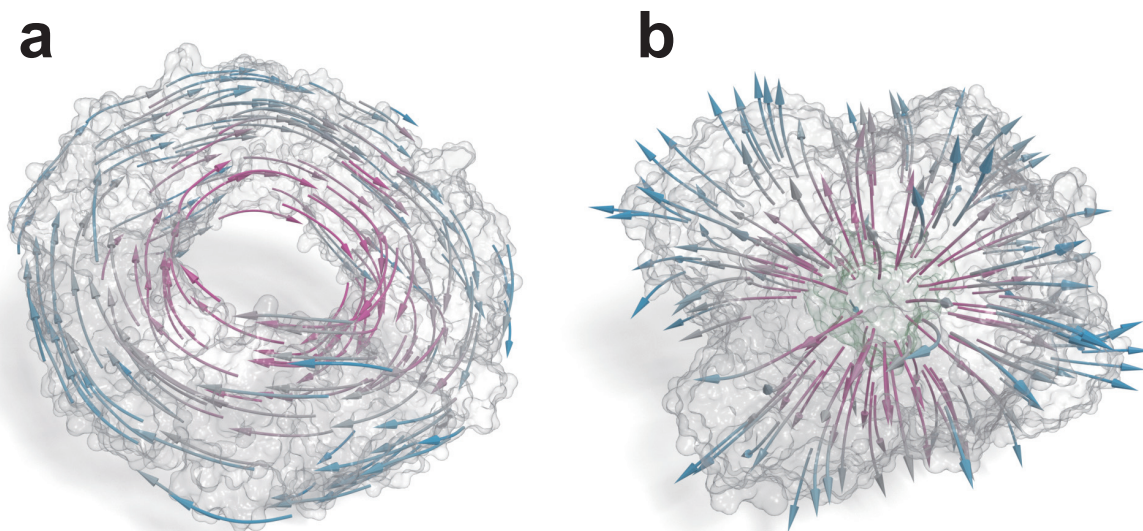
Figure 3.4: **Illustration of topological analysis. a.** Eigenfields by null space of tangential Laplace-de Rham operators correspond to handles. **b.** Eigenfields by null space of normal Laplace-de Rham operators correspond to cavities.

3.2.3 Geometric analysis

Although the spectra of the Laplace-de Rham operators do not uniquely determine the geometry (sometimes referred to as "you cannot hear the shape of the drum"), they do provide key information when comparing shapes, which, sometimes, is referred to as shape "DNA". Thus, the traits of the non-zero parts of the spectra can be regarded as geometrical features. These geometrical features are rigid transformation invariant. The scalar Hodge Laplacian spectrum has already been used in computer graphics and computer vision to distinguish various structures in shape analysis and shape retrieval. It has also been extended to 1-form Hodge Laplacian on surfaces for the purpose of shape analysis. However, on surfaces, L_1 spectrum is identical to L_0 spectrum, except that the multiplicity is doubled for nonzero eigenvalues. Note that the multiplicity for the zero eigenvalue is determined by the number of genus instead of the number of connected components for scalar Hodge Laplacian. In our 3D extension, we have three unique spectra for each molecule. Fig. 3.5 shows non-zero spectrum traits for 3 simple proteins (PDB IDs: 2Z5H [148], 6HU5 [149], and 5HY9 [150]), where the clear distinction among the spectra can be observed. We have tested on
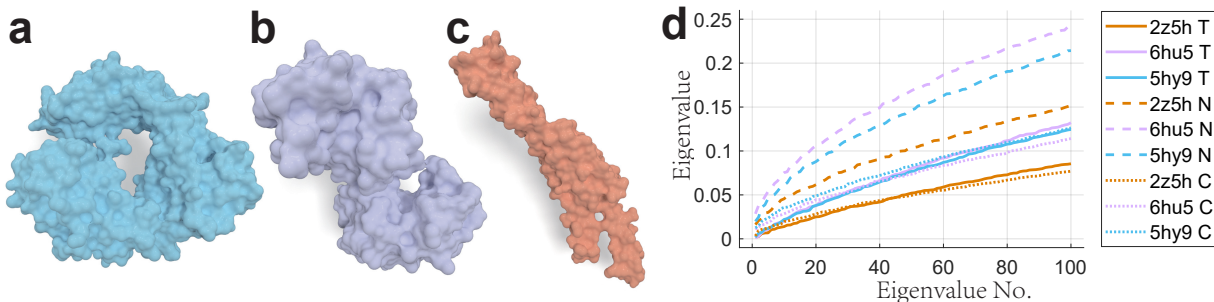
Figure 3.5: **Illustration of geometric analysis.** The geometry of different molecules (PDB IDs: 2Z5H (**a**), 6HU5 (**b**), and 5HY9 (**c**)) can be captured by three groups of different Hodge Laplacian spectra with clear separations shown in **d**. Note that the color of the line plot corresponds to the color of the molecules. The solid lines show tangential gradient (T) spectrum, the dashed lines show the normal gradient (N) spectrum, and the dot lines show the curl spectrum (C). While there is a possibility that certain spectral sets may be close to each other (see groups T of proteins 6HU5 and 5HY9), the other 2 groups of spectra (see groups N and C of proteins 6HU5 and 5HY9) will show a clear difference. In addition, our topological features will also provide a definite difference. For example, protein 6HU5 has trivial topology (ball), but protein 5HY9 has a handle.

various biomolecules and observed the same discriminating ability of the spectra on these shapes.

Geometric analysis and topological analysis based on the de Rham-Hodge theory can be readily applied to characterizing biomolecules in machine learning and to biomolecular modeling. To further demonstrate the capability of de Rham-Hodge spectral analysis for macromolecular analysis, we propose a set of de Rham-Hodge models for protein flexibility analysis and a vector de Rham model for biomolecular Hodge mode analysis.

3.2.4 Flexibility analysis

Biomolecular flexibility analysis and B factor prediction have been commonly performed by normal mode analysis [47, 48, 49, 50, 46] and Gaussian network model (GNM) [51]. The flexibility is strongly correlated to protein functions, such as structural support, catalyzing chemical reactions, and allosteric regulation [151]. Recently, graph theory-based FRI has been shown to outperform other methods [1]. However, all of the aforementioned methods are based on the discrete coordinate representation of biomolecules. As such, it is not very convenient to use these methods for flexibility analysis at different scales. For example, for some large macromolecules, such as an HIV

viral capsid which involves millions of atoms, one may wish to analyze their flexibility at atomic, residue, protein domain, protein, and protein complex scales by using a unified approach so that the results from cross-scales can be compared on an equal footing. However, current approaches cannot provide such a unified cross-scale flexibility analysis. In this work, we introduce a de Rham-Hodge theory-based model to quantitatively analyze macromolecular flexibility cross many scales.

We assume that the de Rham-Hodge B factor at the ith atom estimated by L_k is given by

$$B_{k,i}^{\text{dRH}} = a \sum_j \frac{1}{\lambda_j^k} \left[\omega_j^k(\mathbf{r})(\omega_j^k(\mathbf{r}'))^T \right]_{\mathbf{r}=\mathbf{r}_i, \mathbf{r}'=\mathbf{r}_i}, \forall \lambda_j^k > 0, \tag{3.22}$$

where a is a parameter to be determined by the least squares regression. Its value depends on structural resolution, diffraction intensity, experimental method (i.e, x-ray scattering, electron microscopy, etc), number of diffraction angles, experimental temperature, sample quality, and structure reconstruction method. In the computation, the value of $\omega_j^k(\mathbf{r})$ is given on a set of mesh points. The linear regression over a cutoff radius d is used to obtain the required values in atomic centers $\mathbf{r}_i$ where the B-factor values are reported. $L_{0,t}$ is applied in test cases.

We perform numerical experiments to confirm that our flexibility analysis on C-alpha atoms is robust and reliable. In fact, our method can analyze the flexibility of all atoms or a subset of atoms. The cutoff radius is set to 7 Å. Our method involves several parameters including level set value c and grid spacing r and cutoff radius d (see Fig. 3.6). In Fig. 3.7, it shows statistics of the average Pearson correlation coefficient with various parameters on te test set of 364 proteins.

Level set The level set parameter c in Eq. (3.21) controls the general distance from the surface to C-alpha atoms (see Fig. 3.6 **a**). A larger level set value will result in a smaller domain with richer topology structures, including many tunnels and cavities. A smaller level set value will make the surface fatter so that it will lead to a ball-like shape.

Grid spacing The grid spacing r controls the density of tetrahedrons of the mesh. A finer mesh will lead to a better prediction but is computationally more expensive (see Fig. 3.6 **b**).

Cutoff radius The parameter cutoff radius d controls the linear regression region around the specific C-alpha atom (tets within the radius d to the specific C-alpha atom which is colored purple in Fig. 3.6 **d**). Our approach will potentially introduce a denser mesh, which will lead to small
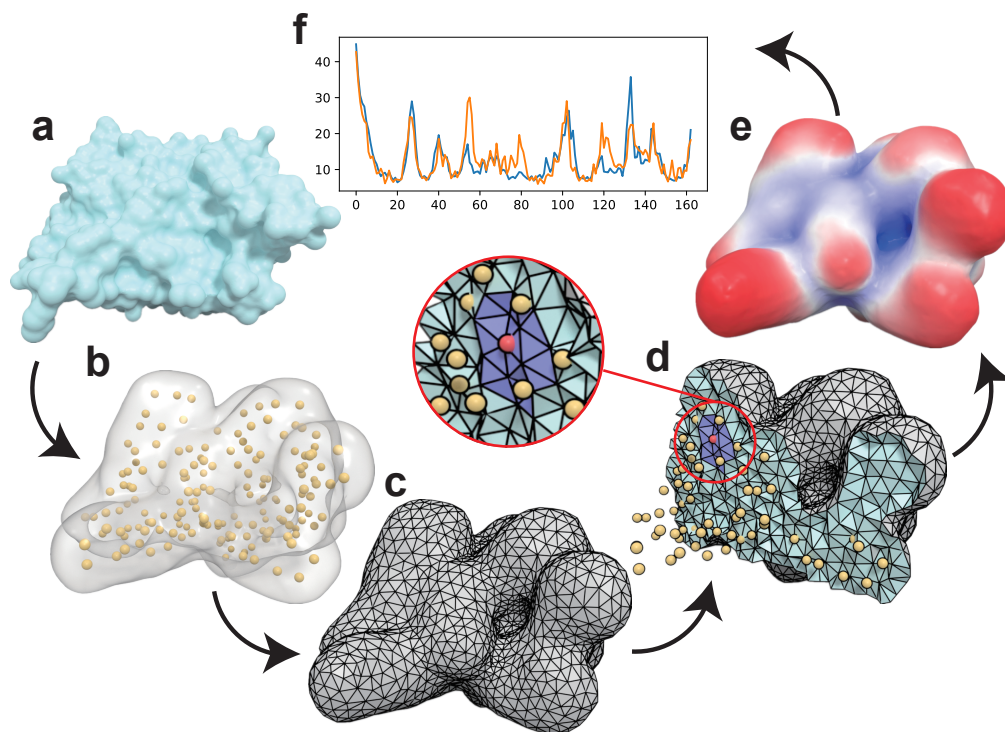
Figure 3.6: **Illustration of the procedure for flexibility analysis**. We use protein 3VZ9 [7] as an example to demonstrate our procedure from **a** to **f**. **a** shows the input protein crystal structure. **b** shows that only C-alpha atoms (yellow spheres) are considered in this case. We assign a Gaussian kernel to each C-alpha atoms and extract the level set surface (transparent surface) as our computation domain. **c** shows that standard tetrahedral mesh is generated with the domain (boundary faces are gray, inner faces are indigo). We use a standard matrix diagonalization procedure to obtain eigenvalues and eigenvectors. B factor at each mesh vertex is computed as shown in Eq. (3.22). **d** B factor at the position of a C-alpha atom is obtained by the linear regression using within the nearby region (for the red C-alpha atom, the linear regression region is colored as purple, which is within the cutoff radius.) **e** shows the predicted B factors on the surface. **f** shows the predicted B factors at C-alpha atoms (orange), compared with the experimental B factors in the PDB file (blue). Our prediction for 3VA9 has the Pearson correlation coefficient of 0.8081.
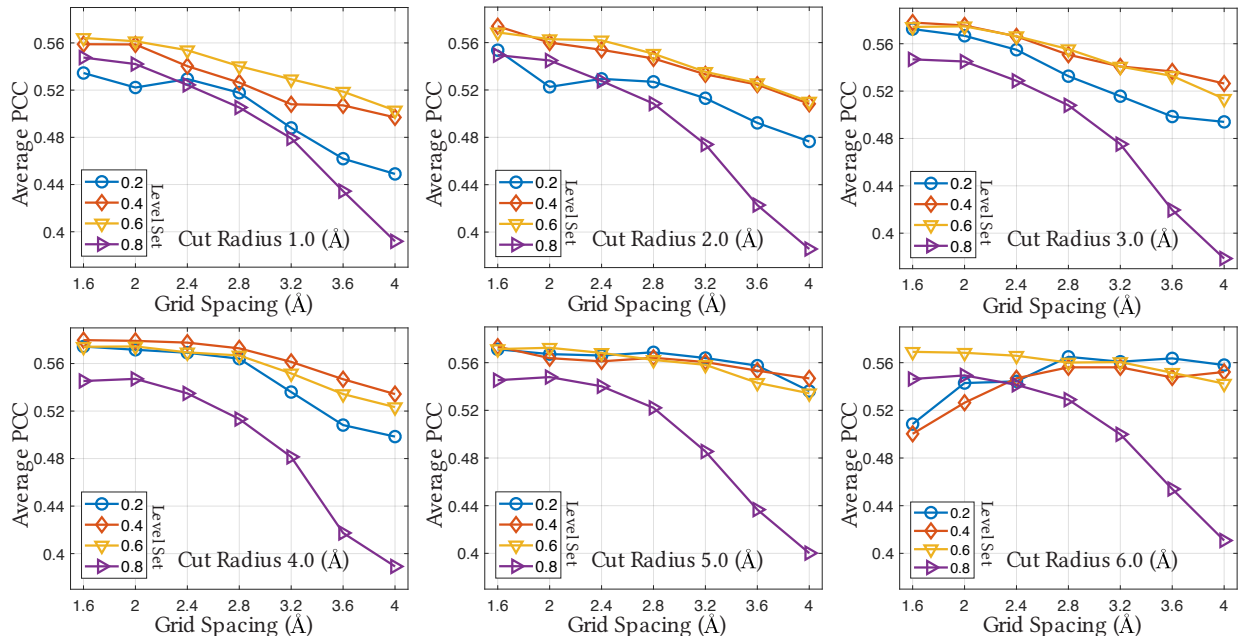
Figure 3.7: **Flexibility prediction results.** Statistics of the average Pearson correlation coefficient (PCC) with various parameters on the test set of 364 proteins. Each plot has the same cutoff radius varying from 1.0 Å to 6.0 Å with interval 1.0 Å. In each plot, the level set value varies from 0.2 to 0.8 with interval 0.2 shown by different lines; the grid spacing varies from 1.6 Å to 4.0 Å with interval 0.4 Å shown in horizontal axis.
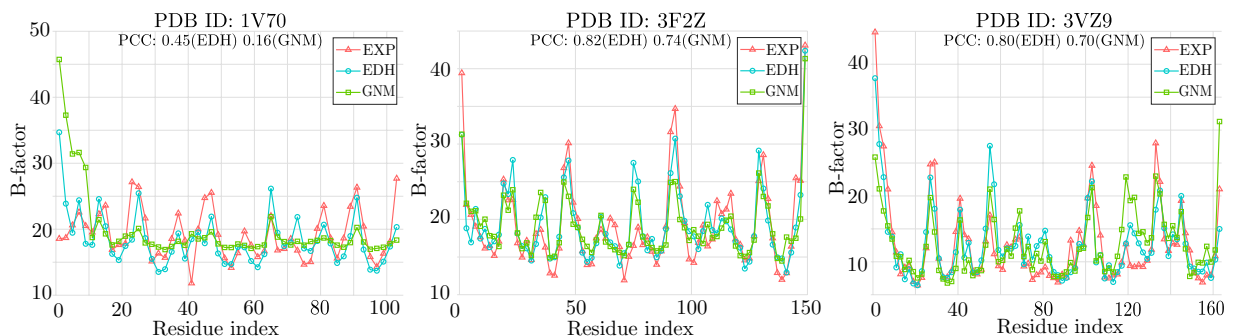
Figure 3.8: **Illustration of B-factor prediction.** We use proteins 1V70 [8], 3F2Z and 3VZ9 as examples to show our predictions compared with the experiments. The red lines with triangles are the ground truth from experimental data. The blue lines with circles are predictions from our method (EDH). The green lines with cubes are predictions from Gaussian network method (GNM).

local vibrations (high frequencies introduced due to the increasing number of matrix elements) that should be filtered out. This treatment is the same as throwing away higher frequencies.

We consider a benchmark test set of 364 proteins studied in earlier work [1] to systematically validate our method. Our test indicates that the best parameters are $c = 0.4$, $r = 1.6$ Å, $d = 4.0$ Å.

68

		Grid Spacing (Å)						
		1.6	2.0	2.4	2.8	3.2	3.6	4.0
Level Set	0.2	0.574	0.572	0.569	0.564	0.536	0.508	0.498
	0.4	**0.580**	0.579	0.578	0.573	0.561	0.547	0.534
	0.6	0.574	0.574	0.569	0.567	0.552	0.534	0.523
	0.8	0.545	0.547	0.535	0.513	0.481	0.417	0.389

Table 3.1: **Results for flexibility prediction.** The average Pearson correlation coefficient for predicting 364 proteins at cutoff radius 4.0 Å. The overall best average Pearson correlation coefficient is 0.580, compared to that of 0.565 for GNM on the same dataset [1].

Fig. 3.8 shows several examples with the best parameters and comparisons with GNM. Table 3.1 shows the average Pearson correlation coefficient of predicting the benchmark set of 364 proteins [1] at a cutoff radius 4.0 Å, which includes the overall best average Pearson correlation coefficient at grid spacing 1.6 Å and level set value 0.4 . The contour level value should not be too large such that only those C-alpha atoms that are close enough to each other will have interactions, as well as not be too small such that enough geometric and topological features are preserved. The cutoff radius should be a proper value such that higher frequencies are mitigated while lower frequencies are well kept. There is not much of influence of resolution if the previous 2 parameters are well set (see statistics at cutoff radius 5 Å). This provides the foundation for analyzing large protein complex with coarse resolution.

The proposed flexibility analysis can be easily extended to analyze the flexibility of a cryo-EM data at given level set. The computed (relative) B factors are located at vertices but can be interpolated to any desirable location if necessary. Due to the multi-resolution nature of our approach, computational cost is determined by the number of unknowns, i.e., the mesh size. For a given computational domain, the mesh size depends on the grid spacing. Therefore, for large macromolecules with millions of atoms, which is intractable for coordinate-based methods, the proposed de Rham-Hodge approach can still be very efficient.

The commonly used method that produces the B-factors that wind up in the PDB files is the least squares fit. This method connects diffraction intensity profiles and structural model predicted densities in the PDB with B factors. In our model, we connect experimental structures (the co-

ordinates of structural model predicted densities) and B factors in the PDB files with our Hodge eigenvalue and eigenvector-based model.

3.2.5 Hodge mode analysis

Normal mode analysis is an important approach for understanding biomolecular collective behavior, residue coupling, protein domain motion, and protein-protein interaction, reaction pathway, allosteric signaling, and enzyme catalysis [47, 48, 49, 50, 46]. However, normal model analysis becomes very expensive for large biomolecules. In particular, it is difficult to carry out the anisotropic network model (ANM) analysis [52] for cryo-EM maps which do not have atomic coordinates. Virtual particle-based ANM methods were proposed to tackle this problem [152, 153]. Being based on the harmonic potential assumption, these methods are restricted to relatively small elastic motions. In this work, we propose an entirely different strategy for biological macromolecular anisotropic motion analysis based on de Rham-Hodge theory.

Laplace-de Rham operator It is noted that a mass-spring system is underlying many earlier successful elastic network models. This system describes the interconversion between the kinetic energy and potential energy during the dynamic motion. In our construction, we take advantage of de Rham-Hodge theory. In fact, de Rham-Hodge theory provides a general framework to model the dynamic behavior of macromolecules. In the present work, we just illustrate this approach with special construction.

In order for de Rham-Hodge theory to be able to describe anisotropic motions, we utilize the 1-form Laplace-de Rham operator

$$\Delta_1 = d_0 \star_0^{-1} d_0 \star_1 + \star_1^{-1} d_1 \star_2 d_1, \tag{3.23}$$

where d_k denote exterior derivatives on $\Omega^k(M)$ and $\star_k$ denote Hodge star operators. Note that 2-form Laplace-de Rham operator works similarly well but we will limit our discussion with 1-form. The first term on the right hand side of Eq. (3.23) is the quadratic energy form measuring the total

70

divergence energy, while the second term measures the total curl energy. Both terms are kinetic energy physically or Dirichlet energy mathematically.

Laplace-de Rham-Helfrich operator   Physically, a potential energy term is required to constrain the elastic motion of biological macromolecules. There are many options, such as Willmore energy, which minimize the difference between two principle curvatures. Additionally, Helfrich introduced a curvature energy for modeling cell membrane or closed lipid vesicles [154, 128]. In our case, we assume the curvature energy of the form

$$V = \mu \int_{\partial M} (H - H_0)^2 dA, \tag{3.24}$$

where μ is the molecular bending rigidity, H is the mean curvature on the molecular surface and H_0 is the spontaneous curvature of the molecule. The potential energy in Eq. (3.24) is defined on the compact manifold enclosing a smooth molecular surface.

Conceptually, our curvature model deals with a dynamical system with a thin shell having thickness much smaller than other dimensions. Computationally, the 2D curvature model serves as a boundary condition to complete the Laplace-de Rham operator on a macromolecule. The curvature energy increases as the mean curvature H deforms away from its rest state. Therefore, H is a function of the surface displacement. The quadratic energy generated from surface deformation is given by (see [155] for discretization details)

$$Q = \partial^2 V / \partial X^2 \tag{3.25}$$

where X is a displacement vector field on the surface. Due to the isomorphism between vector fields and 1-forms, we can evaluate the volumetric 1-form ω as a displacement vector field and restrict it to the boundary surface. We denote the restriction as a linear operator G,

$$X = G\omega. \tag{3.26}$$

Then the quadratic form for the curvature energy in terms of the 1-form is $G^T Q G$. Finally, the total 1-form quadratic energy is given by the following one-parameter Laplace-de Rham-Helfrich
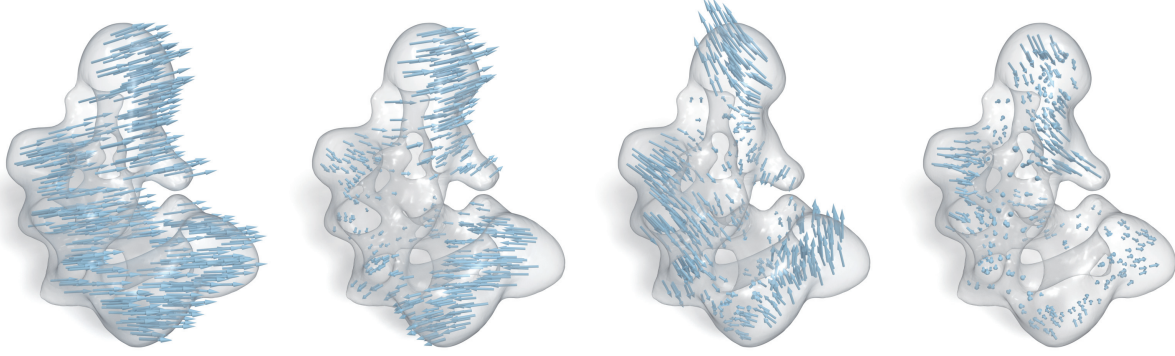
Figure 3.9: **Hodge modes of EMD 1258.** The 0-th, 4-th, 8-th and 12-th Hodge modes are shown.

operator

$$E_\mu = d_0 \star_0^{-1} d_0^T \star_1 + \star_1^{-1} d_1^T \star_2 d_1 + G^T Q G \qquad (3.27)$$

We can solve the eigenvalue problem for the Laplace-de Rham-Helfrich operator E_μ to extract the natural vibration modes of biomolecules. It is a standard procedure to assemble required matrix G and Q together with our Laplace-de Rham matrix.

In fact, an advantage of the proposed anisotropic motion theory is that it allows to treat the divergence energy and curl energy differently. For example, we can introduce a bulk modulus-type of parameter λ to the divergence energy term, which leads to a weighted Laplace-de Rham operator. As a result, we have a two-parameter Laplace-de Rham-Helfrich operator

$$E_{\lambda\mu} = \lambda \cdot d_0 \star_0^{-1} d_0^T \star_1 + \star_1^{-1} d_1^T \star_2 d_1 + G^T Q G. \qquad (3.28)$$

We need to choose appropriate weight parameters λ and μ. Generally, the two-parameter Laplace-de Rham-Helfrich operator and boundary condition matrix can be tuned separately. What we would like to achieve is letting the curvature energy drive the motion and let our system penalize the compressibility (i.e., the divergence energy). Therefore, we select an appropriate λ at a different scale and choose $\mu > \lambda > 1$.

Modal analysis, compared to fluctuation analysis, provides more information. In addition to the description of flexibility, modal analysis also provides the collective motion of a molecule and its potential function. The dynamics of a macromolecule can be described by the linear combination

of its natural modes. Fig. 3.9 shows several Hodge modes for core spliceosomal components, EMD 1258 [156], which indicates the success of our Laplace-de Rham-Helfrich operator.

It is noted that the original Laplace-de Rham operator with appropriate boundary conditions admits the orthogonal Hodge decomposition in terms of divergence-free, curl-free, and harmonic eigenmodes. In contrast, the Laplace-de Rham-Helfrich operator does not preserve these properties. Nonetheless, the eigenmodes generated by the Laplace-de Rham-Helfrich operator are mutually orthogonal and subject to different physical interpretations. For example, the first three eigenmodes are associated with 3D translational motions. Therefore, the operator is translational invariant. The modes in Fig. 3.9 have little to do with the topological singularity of EMD 1258.

Additionally, the eigenmodes in Fig. 3.1 have a fixed boundary. In contrast, boundaries of eigen-modes generated with the Laplace-de Rham-Helfrich operator as shown in Fig. 3.9 are allowed to change. The Laplace-de Rham-Helfrich operator can predict significant macromolecular deforma-tions, which are controllable with two weight parameters, λ and μ. In contrast, existing normal mode analysis methods can only admit small deformations due to the use of the harmonic potential.

Moreover, due to its continuous nature, the proposed Laplace-de Rham-Helfrich operator can be easily employed for the Hodge mode analysis at any given scale. It can be directly applied to the analysis of cryo-EM maps and other volumetric data at an arbitrary scale. One specific example of potential applications is the analysis of subcellular organelles, such as mitochondrial ultrastructure and endoplasmic reticulum.

Finally, the proposed Laplace-de Rham-Helfrich model is phenomenological in nature but can describe physical observations. Like the Navier-Stokes equation for fluid mechanics and the Ginzburg-Landau equation for superconductivity, Laplace-de Rham-Helfrich model is not rigorously derived from the fundamental laws of physics or first principles.

3.2.6   Field decomposition and analysis

Our Laplace-de Rham operators constructed from different boundary conditions can also perform vector field decomposition tasks. Follow the discussion of boundary conditions in Sec. 3.1, a Hodge
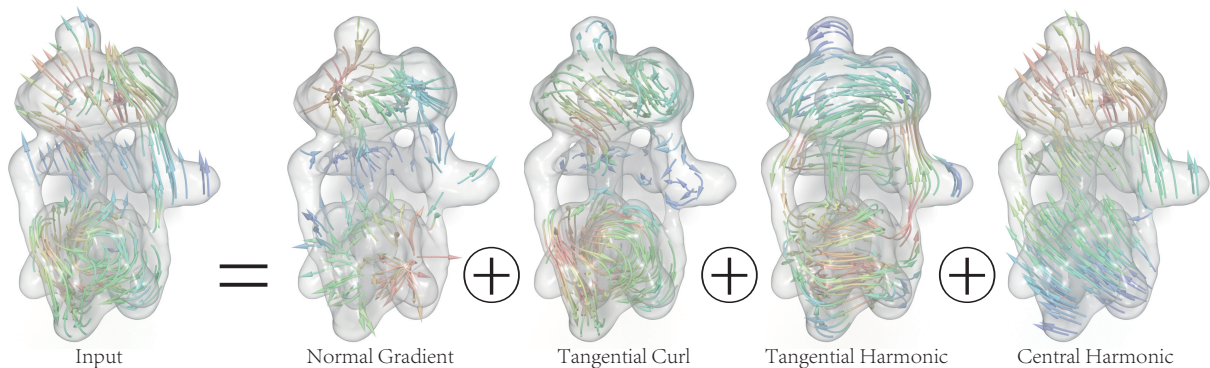
Figure 3.10: **Biological flow decomposition** Illustration of a synthetic vector field in EMD 1590 that is decomposed into several mutually orthogonal components based on different boundary conditions.

decomposition for a k-form bounded manifolds in 3D is constructed as $\omega^k = d\alpha_n^{k-1} + \delta\beta_t^{k+1} + h^k$, where α_n^{k-1} is in the space of normal $(k-1)$-forms Ω_n^{k-1}, β_t^{k+1} is in the space of tangential $(k+1)$-forms and h^k is in H_Δ^k. Moreover, H_Δ^k is further decomposed based on boundary conditions and a five-component orthogonal decomposition [16] is given as

$$\omega^k = d\alpha_n^{k-1} + \delta\beta_t^{k+1} + h_t^k + h_n^k + \eta^k \tag{3.29}$$

where h_t^k is a tangential harmonic form, h_n^k is a normal harmonic form, and η^k is central harmonic form which is both exact and coexact. There are naturally various vector fields existing in biomolecules, such as electric fields, magnetic fields, and elastic displacement fields. De Rham-Hodge theory can help provide a mutually orthogonal decomposition to investigate source, sink and vortex features presented in those fields. An example of this analysis is given in Fig. 3.10 for a synthetic vector field on a vacuolar ATPase motor, EMD 1590 [157]. We expect this decomposition becomes more interesting for biomolecular electric fields, dipolar fields, and magnetic fields. Various components from the decomposition can be naturally used as the components of machine learning feature vectors. Moreover, each orthogonal component can be represented in the basis fromed by eigenfields of Laplace-de Rham operators, and the low-frequency coefficients can be used as machine learning features as well. The following session illustrates an example of an eigenfield representation, for the gradient of the reaction potential for molecular electrostatics.

74
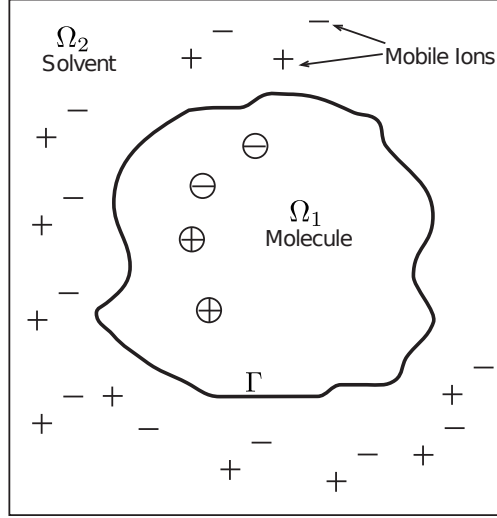
Figure 3.11: **The PB implicit solvent model.** Γ is the molecular surface separating space into the solute region Ω_1 and the solvent region Ω_2.

Electrostatics analysis   Electrostatic interactions are of paramount importance in biomolecular simulations due to their ubiquitous existence and vital contribution to force fields. Two major types of electrostatic analyses are the qualitative analysis for general electrostatic characteristics and the quantitative analysis for statistical, thermodynamic and kinetic observables. An important two-scale implicit solvent model for electrostatic analysis is the Poisson-Boltzmann (PB) model [158, 159], in which the explicit water molecules are treated as a dielectric continuum and the dissolved electrolytes are modeled with the Boltzmann distribution. The PB model has been widely applied in biomolecular simulations such as protein structures [160], protein-protein interactions [161], pKa [162, 163, 164], membranes [165], binding energies [166], solvation free energies [167], etc.

The Poisson-Boltzmann model for a solvated molecule The PB model is illustrated in Fig. 3.11, in which the molecular surface Γ separates the solute domain Ω_1 and the solvent domain Ω_2. The molecule domain Ω_1 consists of a set of atomic charges q_k located at atomic centers $\mathbf{x}_k$ for $k = 1, ..., N_c$. In domain Ω_2, a Boltzmann distribution describes the free ions. For computational purposes, the Boltzmann term is often linearized.

Thus the electrostatic potential $\phi(\mathbf{x})$ here satisfies the linearized PB equation,

$$-\nabla \cdot \epsilon(\mathbf{x})\nabla\phi(\mathbf{x}) + \bar{\kappa}^2(\mathbf{x})\phi(\mathbf{x}) = \sum_{k=1}^{N_c} q_k \delta(\mathbf{x} - \mathbf{x}_k), \tag{3.30}$$

Eigenvector		e_0	e_1	e_2	e_3	e_4	$\cdots$	e_{10}	$\cdots$	e_{100}
p-p	$\langle \omega, e_i \rangle^2$	0.538	0.006	0.025	0.000	0.000	$\cdots$	0.001	$\cdots$	0.000
	$\sum_{j=1}^{i} \langle \omega, e_j \rangle^2$	0.538	0.544	0.569	0.569	0.569	$\cdots$	0.576	$\cdots$	0.928
n-p	$\langle \omega, e_i \rangle$	0.002	0.479	0.000	0.000	0.01	$\cdots$	0.000	$\cdots$	0.000
	$\sum_{j=1}^{i} \langle \omega, e_j \rangle^2$	0.002	0.481	0.481	0.481	0.482	$\cdots$	0.556	$\cdots$	0.906

Table 3.2: **Results for two point charges.** Example 1 considers two cases: p-p for two positive charges and n-p for a negative charge on the left and a positive charge on the right. Here, $\langle \omega, e_i \rangle$ is the inner product of the normalized electrostatic reaction field ω with i-th eigenvector, which is normalized too. The second row of each case is the squared sum of inner products. The sum recovers the normalized electrostatic reaction field if summation is carried out over the inner products with all the eigenfields according to Parseval's theorem.

where $\epsilon(\mathbf{x})$ is the piecewise-constant dielectric function

$$\epsilon(\mathbf{x}) = \begin{cases} \epsilon_1, & \mathbf{x} \in \Omega_1, \\ \epsilon_2, & \mathbf{x} \in \Omega_2, \end{cases} \tag{3.31}$$

and $\bar{\kappa}$ is the screening parameter with the relation $\bar{\kappa}^2 = \epsilon_2 \kappa^2$, where κ is the inverse Debye length measuring the ionic length. The interface conditions on the molecular surface are

$$\phi_1(\mathbf{x}) = \phi_2(\mathbf{x}), \quad \epsilon_1 \frac{\partial \phi_1(\mathbf{x})}{\partial n} = \epsilon_2 \frac{\partial \phi_2(\mathbf{x})}{\partial n}, \quad \mathbf{x} \in \Gamma, \tag{3.32}$$

where ϕ_1 and ϕ_2 are the limit values when approaching the interface from the inside and the outside domains, n is the outward unit normal vector on Γ, and the normal derivatives are $\frac{\partial \phi_i}{\partial n} = n \cdot \nabla \phi_i$. The PB model assumes the far-field boundary condition of $\lim_{|\mathbf{x}| \to \infty} \phi(\mathbf{x}) = 0$. Taking interface Γ as the solvent excluded surface, the PB model is usually solved numerically. Two types of methods have been developed: grid-based finite-difference and finite-element methods discretize the entire domain [168, 169, 170], such as MIBPB [171, 172]; and boundary element methods discretize only the molecular surface [173, 174, 175, 176, 177]. We use boundary element methods according to the same surface mesh used as the molecular surface and the boundary for our volumetric manifold, for the simplicity of calculating the reaction potential.

Solving PB model and reaction potential   A well-conditioned boundary integral form of PB implicit solvent model is derived by applying Green's second identity and properties of fundamental
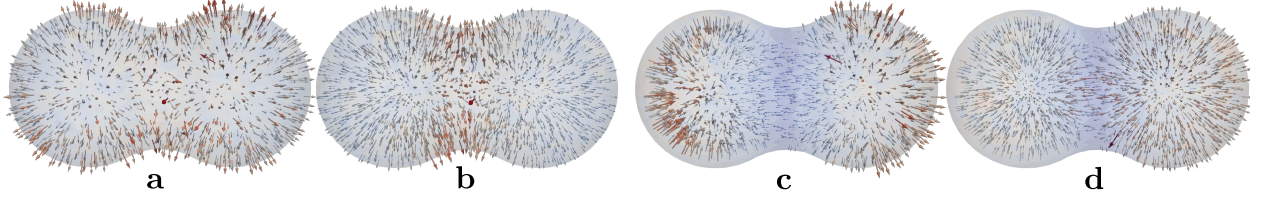
Figure 3.12: **Two point charges. a** the force field of two positive charges; **b** the first eigenvector; **c** the force field of one negative and one positive charges; **c** the second eigenvector.

solutions to Eq. (3.30), which yields the electrostatic potential,

$$\phi(\mathbf{x}) = \int_{\Gamma} \left[G_0(\mathbf{x}, \mathbf{y}) \frac{\partial \phi(\mathbf{y})}{\partial n} - \frac{\partial G_0(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{y}}} \phi(\mathbf{y}) \right] dS_{\mathbf{y}} + \sum_{k=1}^{N_c} q_k G_0(\mathbf{x}, \mathbf{y}_k), \quad \mathbf{x} \in \Omega_1, \quad (3.33a)$$

$$\phi(\mathbf{x}) = \int_{\Gamma} \left[-G_\kappa(\mathbf{x}, \mathbf{y}) \frac{\partial \phi(\mathbf{y})}{\partial n} + \frac{\partial G_\kappa(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{y}}} \phi(\mathbf{y}) \right] dS_{\mathbf{y}}, \quad \mathbf{x} \in \Omega_2, \quad (3.33b)$$

where the Green's function for Coulomb interaction is $G_0(\mathbf{x}, \mathbf{y}) = \frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}$ and the Green's function for the screened Coulomb interaction $G_k(\mathbf{x}, \mathbf{y}) = \frac{e^{-\kappa|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}$. Then applying the interface condition in Eq. (3.32) with the differentiation of electrostatic potential in each domain yield a set of boundary integral equations relating the surface potential ϕ_1 and its normal derivative $\partial\phi_1/\partial n$ on Γ,

$$\frac{1}{2}(1+\epsilon)\phi_1(\mathbf{x}) = \int_{\Gamma} \left[K_1(\mathbf{x}, \mathbf{y}) \frac{\partial \phi_1(\mathbf{y})}{\partial n} + K_2(\mathbf{x}, \mathbf{y})\phi_1(\mathbf{y}) \right] dS_{\mathbf{y}} + S_1(\mathbf{x}), \quad \mathbf{x} \in \Gamma, \quad (3.34a)$$

$$\frac{1}{2}\left(1+\frac{1}{\epsilon}\right)\frac{\partial \phi_1(\mathbf{x})}{\partial n} = \int_{\Gamma} \left[K_3(\mathbf{x}, \mathbf{y}) \frac{\partial \phi_1(\mathbf{y})}{\partial n} + K_4(\mathbf{x}, \mathbf{y})\phi_1(\mathbf{y}) \right] dS_{\mathbf{y}} + S_2(\mathbf{x}), \quad \mathbf{x} \in \Gamma, \quad (3.34b)$$

where $\epsilon = \epsilon_2/\epsilon_1$. As given in Eqs. (3.35a-3.35b) and (3.36), the kernels $K_{1,2,3,4}$ and source terms $S_{1,2}$ are linear combinations of the Coulomb and screened Coulomb interactions, and their first and second order normal derivatives,

$$K_1(\mathbf{x}, \mathbf{y}) = G_0(\mathbf{x}, \mathbf{y}) - G_\kappa(\mathbf{x}, \mathbf{y}), \quad K_2(\mathbf{x}, \mathbf{y}) = \epsilon \frac{\partial G_\kappa(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{y}}} - \frac{\partial G_0(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{y}}}, \quad (3.35a)$$

$$K_3(\mathbf{x}, \mathbf{y}) = \frac{\partial G_0(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{x}}} - \frac{1}{\epsilon} \frac{\partial G_\kappa(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{x}}}, \quad K_4(\mathbf{x}, \mathbf{y}) = \frac{\partial^2 G_\kappa(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{x}}\partial n_{\mathbf{y}}} - \frac{\partial^2 G_0(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{x}}\partial n_{\mathbf{y}}}, \quad (3.35b)$$

and the source terms $S_{1,2}$ are

$$S_1(\mathbf{x}) = \frac{1}{\epsilon_1} \sum_{k=1}^{N_c} q_k G_0(\mathbf{x}, \mathbf{y}_k), \quad S_2(\mathbf{x}) = \frac{1}{\epsilon_1} \sum_{k=1}^{N_c} q_k \frac{\partial G_0(\mathbf{x}, \mathbf{y}_k)}{\partial n_{\mathbf{x}}}. \quad (3.36)$$

Once the potential and normal derivative of the potential on boundary of Eqs. (3.33a) and (3.33b) are solved, the reaction potential $\phi_{\text{reac}}(\mathbf{x}) = \phi(\mathbf{x}) - S_1(\mathbf{x})$ and for $\mathbf{x} \in \Omega_1$ it is given

Eigenvector		e_0	e_1	e_2	e_3	e_4	$\cdots$	e_{10}	$\cdots$	e_{100}
p-p-p-p	$\langle\omega, e_i\rangle^2$	0.547	0.017	0.000	0.001	0.000	$\cdots$	0.000	$\cdots$	0.001
	$\sum_{j=1}^{i}\langle\omega, e_j\rangle^2$	0.547	0.564	0.564	0.565	0.565	$\cdots$	0.566	$\cdots$	0.853
p-p-n-n	$\langle\omega, e_i\rangle^2$	0.008	0.268	0.211	0.001	0.000	$\cdots$	0.001	$\cdots$	0.000
	$\sum_{j=1}^{i}\langle\omega, e_j\rangle^2$	0.008	0.276	0.487	0.488	0.488	$\cdots$	0.546	$\cdots$	0.839
p-n-p-n	$\langle\omega, e_i\rangle^2$	0.005	0.198	0.272	0.005	0.000	$\cdots$	0.000	$\cdots$	0.000
	$\sum_{j=1}^{i}\langle\omega, e_j\rangle^2$	0.005	0.203	0.475	0.480	0.480	$\cdots$	0.533	$\cdots$	0.840
p-n-n-p	$\langle\omega, e_i\rangle^2$	0.002	0.002	0.002	0.434	0.000	$\cdots$	0.000	$\cdots$	0.000
	$\sum_{j=1}^{i}\langle\omega, e_j\rangle^2$	0.002	0.004	0.006	0.440	0.440	$\cdots$	0.459	$\cdots$	0.839
p-p-p-n	$\langle\omega, e_i\rangle^2$	0.434	0.055	0.000	0.047	0.000	$\cdots$	0.000	$\cdots$	0.001
	$\sum_{j=1}^{i}\langle\omega, e_j\rangle^2$	0.434	0.489	0.489	0.536	0.536	$\cdots$	0.557	$\cdots$	0.848

Table 3.3: **Results for four point charges.** Example 2 considers four charges arranged in five cases, namely p-p-p-p, p-p-n-n, p-n-p-n, p-n-n-p, and p-p-p-n, where "p" stands for positive and "n" stands for negative, specified in the order of top left, top right, bottom left, and bottom right. Here, $\langle\omega, e_i\rangle$ is the inner product of the normalized electrostatic reaction field ω with i-th eigenvector, which is normalized too. The second row of each case is the squared sum of inner products. The sum recovers the normalized electrostatic reaction field if summation is carried out over the inner products with all the eigenfields according to Parseval's theorem.

as

$$\phi_{\text{reac}}(\mathbf{x}) = \int_{\Gamma}\left[G_0(\mathbf{x}, \mathbf{y})\frac{\partial\phi(\mathbf{y})}{\partial n} - \frac{\partial G_0(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{y}}}\phi(\mathbf{y})\right]dS_{\mathbf{y}}. \tag{3.37}$$

Numerically solving boundary integral forms of PB model requires speedup techniques, for which we directly apply the software package presented in [178]. The reaction potential describes the potential caused by the solvent and solute near their interface. It is important to calculate the electrostatic solvation energy, given as $\Delta G_{\text{sol}} = \frac{1}{2}\sum_{k=1}^{N_c}q_k\phi_{\text{reac}}(\mathbf{x}_k)$, where N_c is the number of charges and q_k are charges.

Eigenfield decomposition   The 1-form electrostatic reaction field ω is generated from the gradient of the reaction potential $\nabla\phi_{\text{reac}}$ by taking line integral on each edge. Our goal is to project ω onto the eigenvectors of Hodge Laplacian by L_2-inner products of Eq. (3.3). The molecular surface Γ created by the solute and the solvent is considered as the boundary of the volumetric manifold M. The space of k-forms $\Omega_k(M)$ is a Hilbert space equipped with the aforementioned L_2-inner products. Therefore, the corresponding 1-form of the electrostatic reaction field inside the molecule
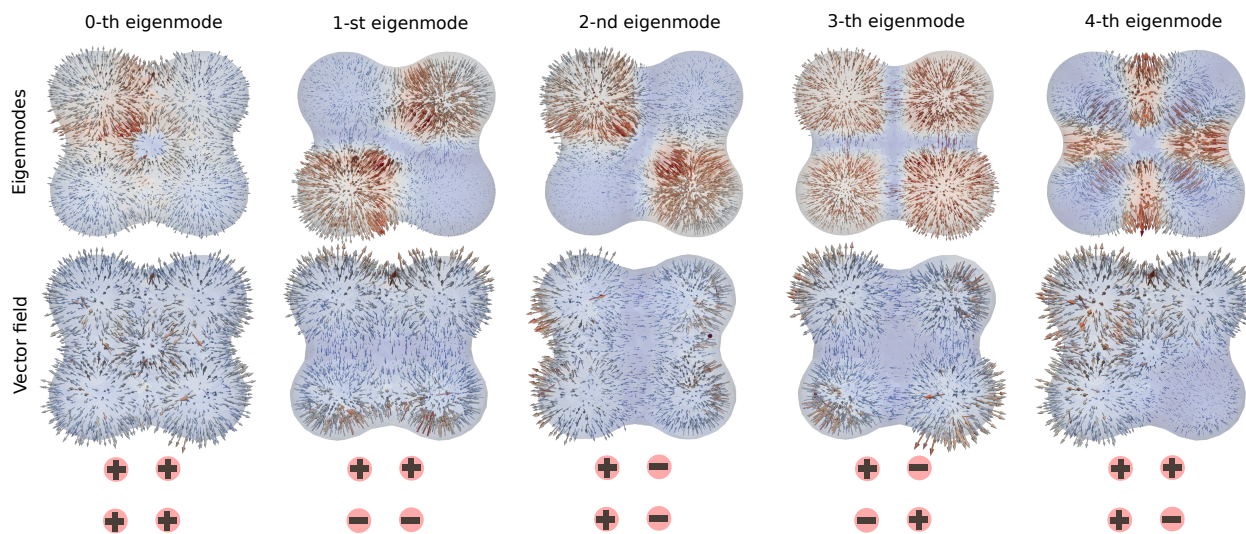
Figure 3.13: **Four point charges.** The first row shows the first five eigenmodes. The second row shows vector fields under corresponding charge combinations.

surface is in the space $\Omega_1(M)$. Moreover, as shown in Eq. (3.29), aside from a harmonic component, the gradient of the reaction potential is in the spaced of normal gradient fields, which is spanned by the eigenvectors corresponding to the normal gradient fields. Represented in the basis formed by these eigenvectors, the electrostatic reaction field (without the harmonic component) is a linear combination of these eigenvectors. However, the coefficients are with only large absolute values for certain modes, since dominant eigenmodes often exist due to the geometry characteristics of molecular domain. We illustrate the Hodge mode decomposition for two examples. Table 3.2 shows the square of coefficients of i-th eigenvector projected on the electrostatic reaction field ω as $\langle \omega, e_i \rangle^2$, and the their sums. The dominant eigenvectors for p-p and n-p are the first and second eigenvectors respectively as shown in Fig. 3.12, in which the eigenvectors are sorted in ascending order of their corresponding eigenvalues. As the number of eigenvectors increases, the difference between the electrostatic reaction field and the approximated electrostatic reaction field decreases. Table 3.3 shows another example with four changes arranged in five ways as shown in Fig. 3.13.

The first case has four positive charges. The first Hodge eigenvector is the dominant mode among all the eigenvectors as shown in Fig. 3.13. In the second and third cases, where two same type charges either located either in the top-bottom or right-left manner, the second and third Hodge

eigenvectors dominate their electrostatic reaction fields. The dominant Hodge eigenvector for the third case is the forth Hodge mode. The last case illustrates a molecule that has three positive charges and one negative charge, for which, the first Hodge eigenvector is the dominant mode. In all cases, the accumulated contributions of the first 11 Hodge modes have a similar magnitude. This method is readily applicable to the electrostatic reaction field analysis of complex biomolecular systems and to the general Hodge mode analysis of any biomolecular vector fields.

3.3 Method preliminaries

We provide the details for our design of computational tools, data structures, and parameters in our implementation of the present de Rham-Hodge spectral analysis. Through efficient implementation, our method is highly scalable and capable of handling molecular data ranging from protein crystal structures to cryo-EM maps.

3.3.1 Simplicial complex generation

The domain of our Laplace-de Rham operators is first tessellated into a simplicial complex, which is a tetrahedral mesh in our 3D case. There are quite a few well-developed software packages for tetrahedral mesh generation given a boundary with a surface triangle mesh as input. We chose CGAL (Computational geometry algorithm library) over others for its superior control on element quality.

In theory, we can generate tetrahedral meshes with any highly accurate closed surface. However, macromolecule complexes with atom-level resolution often make the output mesh intractable with typical computing platforms. Moreover, a dense mesh is unnecessary for the calculation of the low-frequency range of the spectrum. Thus, we produce a coarse resolution with a spatial sampling density higher than twice the spatial frequencies (wavenumbers, i.e., square root of eigenvalues of the Laplacians) of the geometrical and topological features to be computed in the given biomolecule complexes.

For protein crystal structures, we tested the construction of the surface using only the C_α po-

sitions. First, a Gaussian kernel is assigned to each atomic position to approximate the electron density. Then, a level set surface is generated to construct the contour of the protein closely enclosing the high electron density regions.

For cryo-EM data, to produce a smooth contour surface, Gaussian kernels are associated with data points. Other approaches, such as mean curvature flow [43, 55] can be used as well. When dealing with noisy and densely sampled data, we can carefully choose the level set that corresponds to a fairly smooth contour surface that encloses the original cryo-EM data.

Given a volumetric data, we can either directly use CGAL to produce a tetrahedral mesh, or first convert it to a triangular surface mesh through the marching cubes algorithm, and use that to generate a tetrahedral mesh. Different sampling densities are tested to meet typical quality requirements while balancing computational cost and mesh quality.

3.3.2 Discrete exterior calculus

As a topological structure-preserving discretization of the exterior calculus on differential forms, discrete exterior calculus (DEC) has been widely applied in the recent years for various successful applications on geometrical problems and finite element analysis, including meshing and computational electromagnetics [67]. It is an appropriate tool for our de Rham-Hodge analysis of biomolecules, as all the related operations, including exterior derivatives and the Hodge stars, are represented as matrices that preserve the defining properties in the continuous setting. More precisely, the discrete exterior derivative operators strictly satisfies $D_{k+1}D_k = 0$, mimicking $d_{k+1}d_k = 0$, and the discrete Hodge star operators are realized by symmetric positive definite matrices. Hence, the discrete Laplace-de Rham operators can be assembled using finite dimensional linear algebra with the aforementioned three distinct spectra.

To allow replication of our results, we recap our implementation of DEC [72]. We start by a tetrahedral tessellation of the volumetric domain, i.e., a tetrahedral mesh, which is the collection of a vertex set $\mathcal{V}$, an edge set $\mathcal{E}$, a triangle set $\mathcal{F}$, and a tetrahedron set $\mathcal{T}$. The vertices are points in 3D Euclidean space, the edges/triangles/tetrahedra are represented as 1-/2-/3-simplices,

81
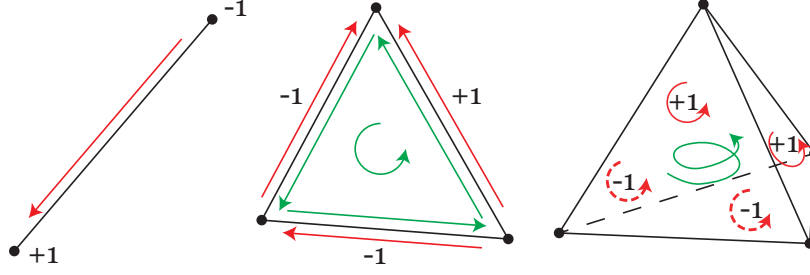
Figure 3.14: **Illustration of orientation.** Pre-assigned orientation is colored in red. Induced orientation by ∂ is colored in green. The vertices are assumed to have a positive pre-assigned orientation. Therefore, the induced orientation from edge orientation is $+1$ at the head and -1 at the tail. For a triangle facet, $+1$ is assigned whenever the pre-assigned orientation conforms with the induced orientation, and -1 vice versa. A similar rule applies to tets which obey a right-hand orientation with the normal pointing outward. Non-adjacent vertices give 0.

i.e., pairs/triples/quadruples of vertex indices respectively, and regarded as the convex hull of these vertices. We further choose an arbitrary orientation for each k-simplex, which is an order set of $k+1$ vertices, up to an even permutation. We denote an oriented k-simplex as

$$\sigma = [v_0, v_1, ..., v_k]. \tag{3.38}$$

The boundary operator is defined as

$$\partial\sigma = \sum_{i=0}^{k}(-1)^i[v_0, v_1, ..., \widehat{v_i}, ..., v_k], \tag{3.39}$$

where $\widehat{v_i}$ means that the i-the vertex is omitted. Thus the boundary operator will take all the 1-degree lower faces of σ with an induced orientation. We will take the following strategy to handle orientation in the implementation. We usually assign each tet an orientation such that, when applying the boundary operator, each facet has an outward pointing orientation. The total boundary of the tet mesh conforms naturally with the surface with outward pointing orientation. But for each edge and facet, we pre-assign an orientation by increasing indices of incident vertices. In this case, we need to take care of the boundary operator when there is a conflict between the pre-assigned orientation and the induced orientation. The algorithm for calculating the cohomology basis of boundary operators is similar to the algorithm in simplicial homology [179]. However, DEC needs further constructions.
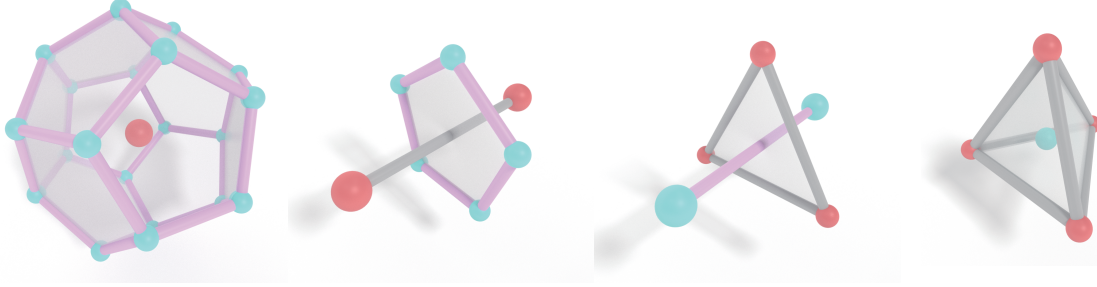
Figure 3.15: **Illustration of the primal and dual elements of the tetrahedral mesh.** All the red vertices are mesh primal vertices. All the indigo vertices are dual vertices at circumcenter of each tet. All the gray edges are primal edges. All the pink edges are dual edges connecting adjacent dual vertices. The first chart shows the dual cell of a primal vertex. The second chart shows the dual facet of the primal edge. The third chart shows the dual edge of the primal facet. The last chart shows the dual vertex of the primal cell (tet).

Scalar fields are naturally encoded as 0-forms and 3-forms. A 0-form is the same with the finite element method such that the coefficients are sampled on vertices equipped with basis functions. A 3-form is, different from a 0-form, stored per tet as volume integration of the scalar field. Vector fields are naturally encoded as 1-form and 2-form. A 1-form is sampled by the line integral on each oriented edge. A 2-form is sampled by surface flux on each oriented facet. Whitney forms [180] can help convert forms back to piecewise linear vector fields on each tet, which can be used in, e.g., the construction of the operator G.

We will store discrete k-forms as column vectors. Then as mentioned before, all the discrete operators can be formed as matrices applying on the column vectors. Then we start to construct discrete exterior derivative and discrete Hodge star matrices. Suppose we are dealing with discrete differential form $d\omega$ on simplices σ, according to Stokes' theorem

$$\int_{\partial\sigma} \omega = \int_{\sigma} d\omega, \tag{3.40}$$

$d\omega$ is just an oriented summation of ω on facets of σ. So the discrete exterior derivative operator D_k is just a matrix filled with $-1, 0, 1$ (see Fig. 3.14), depending on whether the pre-assigned orientation is conforming with the induced orientation. The preservation of Stokes' theorem is what guarantees the preservation of the de Rham cohomology, as the discrete de Rham k-cohomology is isomorphic to the simplicial $n - k$-homology due to the boundary operator, which is in turn
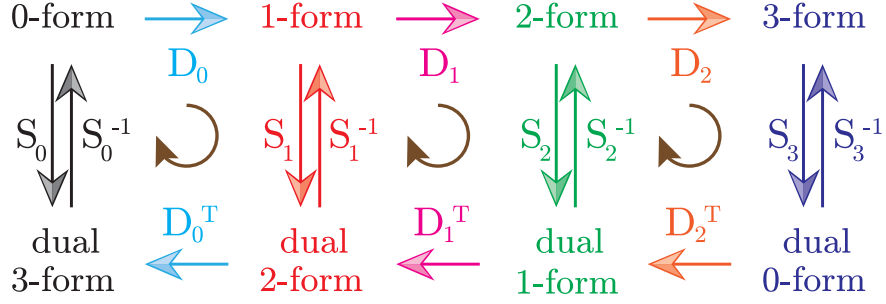
Figure 3.16: **Illustration of cohomology.** This figure illustrates the relation by exterior derivative and Hodge star operators. The assembly of Laplacian operator L_k is just starting from primal k-forms, multiplying matrices along the circular direction.

isomorphic to singular k-cohomology and thus to the continuous de Rham k-cohomology.

One can easily observe that the discrete exterior derivative operators for dual forms are merely D_k^T. The discrete Hodge star operator S_k is just converting primal form and dual form back a forth by following equation

$$\frac{1}{|\sigma_k|} \int_{\sigma_k} \omega = \frac{1}{|*\sigma_k|} \int_{*\sigma_k} \star\omega. \tag{3.41}$$

Each primal element in the tet mesh has one corresponding dual element (see Fig. 3.15). So the discrete Hodge star operator is merely a diagonal matrix. Note that here we use a diagonal matrix to approximate the Hodge star operator, where non-diagonal Hodge star with higher accuracy can be applied as well. But a diagonal Hodge star is enough for our current application. The diagonal Hodge star matrix just has diagonal entries as dual element volume over primal element volume. For example, given a 1-form on each edge, applying the Hodge star is turning the primal 1-form into dual 2-form stored on each dual facets. This can be interpreted as we sample the vector field at the center of the edge. One way is to compute the 1-form as the sampled vector integrated the primal edge as the line integral, the other way is to compute the 2-form as the sampled vector integrated on the dual facet as vector flux. So the transition can be encoded as a number dual element volume over primal element volume. See Fig. 3.16 for relations between differential forms and operators.

Once we have these related matrices for discrete operators, we are ready to construct the Lapla-

cian matrix L_k for $k = 0, ..., 3$ as

$$L_0 = D_0^T S_1 D_0, \qquad\qquad L_1 = D_1^T S_2 D_1 + S_1 D_0 S_0^{-1} D_0^T S_1,$$
$$L_2 = D_2^T S_3 D_2 + S_2 D_1 S_1^{-1} D_1^T S_2, \ L_3 = S_3 D_2 S_2^{-1} D_2^T S_3, \tag{3.42}$$

where D_k are pre-assembled discrete exterior derivatives, S_k are discrete Hodge star matrices and L_k correspond to $\star\Delta_k$. The assembly of Laplace-de Rham operators L_k are just starting from primal k-forms, multiplying matrices along the circular direction as shown in Fig. 3.16. Note that the usual Hodge Laplacian matrix is not symmetric generally. In practice, we usually left multiply by Hodge star to turn it into a symmetric one. After this, we need to take care of the boundary conditions. Boundary condition treatment can be incorporated when assembling d matrices. Recall that the d matrices are merely for creating an oriented summation of discrete differential forms stored on simplices. We can just delete corresponding columns and rows for boundary elements. We use $L_{k,t}$ to denote Laplace-de Rham operator with boundary elements, and $L_{k,n}$ to denote those without boundary elements [142].

Finally, the spectral analysis can be done with a generalized eigenvalue problem in Eq. (3.8).

The smallest eigenvalues and their corresponding eigenvectors are associated with useful low frequencies. In principle, large eigenvalues also contain useful information but are often impaired by large computational errors. We use an eigensolver with parameter starting from small magnitude eigenvalues.

EVOLUTIONARY DE RHAM-HODGE METHOD

4.1 A primer on de Rham-Hodge theory

To introduce the evolutionary de Rham-Hodge method, we briefly review the de Rham-Hodge theory to establish notation. We first discuss differential geometry and de Rham complex on smooth manifolds before reviewing the Hodge decomposition. Then, we illustrate the DEC discretization of the de Rham-Laplace operators and analyze their spectra.

4.1.1 Differential geometry and de Rham complex

Differential geometry is the study of shapes that can be represented by smooth manifolds of an arbitrary dimension. A differential k-form $\omega^k \in \Omega^k(M)$ is an antisymmetric covariant tensor of rank k on manifold M. Roughly speaking, at each point of M, it is a linear map from an array of k vectors into a number, which switches sign if any two of the vectors are swapped. In general, it gives a uniform approach to define the integrals over curves, surfaces, volumes or higher-dimensional oriented submanifolds of M. More precisely, the antisymmetric rank-k covariant tensor linearly maps k edges from the first vertex of each k-simplex in a tessellation of the k-submanifold into a number, creating a Riemann sum that converges to an integral independent of the tessellation.

In $\mathbb{R}^3$, 0-forms and 3-forms can be recognized as scalar fields, as the antisymmetry permits one degree of freedom (DoF) per point, whereas 1-forms and 2-forms are considered vector fields as they require three DoFs per point. Our following discussion is specific to 3-dimensional (3D) volumes bounded by 2-manifolds in $\mathbb{R}^3$.

The *differential* operator (i.e., exterior derivative) d^k maps from the space of k-form on manifold, $\Omega^k(M)$ to $\Omega^{k+1}(M)$. It can be regarded as an antisymmetrization of the partial derivatives of a k-form. As such, it is a linear map $d^k : \Omega^k(M) \to \Omega^{k+1}(M)$ that satisfies the Stokes' theorem

over any $(k+1)$-submanifold $\mathcal{S}$ in M:

$$\int_{\mathcal{S}} d^k \omega^k = \int_{\partial \mathcal{S}} \omega^k, \tag{4.1}$$

where $\partial \mathcal{S}$ is the boundary of $\mathcal{S}$ and $\omega^k \in \Omega^k(M)$ is an arbitrary k-form. Consequently, a key property of differential operator, $d^k d^{k-1} = 0$, follows from that boundaries are boundaryless ($\partial \partial \mathcal{S} = 0$). This implies that an *exact* form (image of a $(k-1)$-form under differential) is *closed* (i.e., is in the kernel of differential). The differential operator indeed provides a unification of a number of commonly used operators in 3D vector field analysis. Depending on the degree k of differential forms, d^k can be regarded as gradient (∇), curl ($\nabla \times$) and divergence ($\nabla \cdot$) operators for 0-, 1- and 2-forms, respectively, e.g., d^0 takes the gradient of a scalar field (representing a 0-form) to a vector field (representing a 1-form).

With the linear spaces of k-forms treated as abelian groups under addition and the linear maps d treated as group homomorphisms, they form a sequence that fits the definition of a *cochain complex* as $d^k d^{k-1} = 0$. This cochain complex of differential forms on a smooth manifold M is known as the *de Rham complex*:

$$0 \longrightarrow \Omega^0(M) \xrightarrow{d^0} \Omega^1(M) \xrightarrow{d^1} \Omega^2(M) \xrightarrow{d^2} \Omega^3(M) \xrightarrow{d^3} 0.$$

Note that d^3 maps 3-forms to 4-forms, but k-forms for $k > 3$ are always zero in $\mathbb{R}^3$ due to antisymmetry.

The *Hodge k-star* $\star^k$ (also called *Hodge dual*) is linear map (and hence also a group isomorphism) from a k-from to its dual form, $\star^k : \Omega^k(M) \to \Omega^{n-k}(M)$. Due to the antisymmetry, both k-forms and their dual $(n-k)$-forms have the same DoF $\binom{n}{k} = \binom{n}{n-k}$. More specifically, for an orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$, $\star^k(\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \cdots \wedge \mathbf{e}_{i_k}) = \mathbf{e}_{j_1} \wedge \mathbf{e}_{j_2} \wedge \cdots \wedge \mathbf{e}_{j_{n-k}}$, where $\wedge$ denotes the antisymmetrized tensor product, and $(i_1, ..., i_k, j_1, ..., j_{n-k})$ is an even permutation of $\{1, 2, ..., n\}$. The associated $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ is a basis for 1-forms, and $\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_k}$ form a basis for k-forms.

As $\star^k$ and d^k can only operate on k-forms, we can omit the superscript of the forms or the operators when the dimension is clear from the context. The (L_2-)inner product of differential

	order 0	order 1	order 2	order 3
form	f^0	$\mathbf{v}^1(\mathbf{a})$	$\mathbf{v}^2(\mathbf{a},\mathbf{b})$	$f^3(\mathbf{a},\mathbf{b},\mathbf{c})$
	f	$\mathbf{v}\cdot\mathbf{a}$	$\mathbf{v}\cdot(\mathbf{a}\times\mathbf{b})$	$f[(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}]$
d	df^0	$d\mathbf{v}^1$	$d\mathbf{v}^2$	df^3
	$(\nabla f)^1$	$(\nabla\times\mathbf{v})^2$	$(\nabla\cdot\mathbf{v})^3$	0
$\star$	$\star f^0$	$\star\mathbf{v}^1$	$\star\mathbf{v}^2$	$\star f^3$
	f^3	$\mathbf{v}^2$	$\mathbf{v}^1$	f^0
δ	δf^0	$\delta\mathbf{v}^1$	$\delta\mathbf{v}^2$	δf^3
	0	$(-\nabla\cdot\mathbf{v})^0$	$(\nabla\times\mathbf{v})^1$	$(-\nabla f)^2$
$\wedge$	$f^0\wedge g^0$	$f^0\wedge\mathbf{v}^1$	$f^0\wedge\mathbf{v}^2, \mathbf{v}^1\wedge\mathbf{u}^1$	$f^0\wedge g^3, \mathbf{v}^1\wedge\mathbf{u}^2$
	$(fg)^0$	$(f\mathbf{v})^1$	$(f\mathbf{v})^2, (\mathbf{v}\times\mathbf{u})^2$	$(fg)^3, (\mathbf{v}\cdot\mathbf{u})^3$

Table 4.1: **Exterior vs. traditional calculus.** Exterior (odd rows) vs. traditional (even rows) calculus in $\mathbb{R}^3$. f^0, $\mathbf{v}^1$, $\mathbf{v}^2$ and f^3 stand for 0-, 1-, 2- and 3-forms with their components stored in either a scalar field f or vector field $\mathbf{v}$.

forms for two k-forms $\alpha, \beta \in \Omega^k(M)$ can be defined as

$$\langle \alpha, \beta \rangle = \int_M \alpha \wedge \star\beta = \int_M \beta \wedge \star\alpha. \tag{4.2}$$

Under these inner products, the adjoint operators of d are the *codifferential* operators $\delta^k \colon \Omega^k(M) \to \Omega^{k-1}(M)$, $\delta^k = (-1)^k \star^{4-k} d^{3-k}\star^k$ for $k = 1, 2, 3$. In 3D, they can be identified with $-\nabla\cdot$, $\nabla\times$ and $-\nabla$ for δ^k, $k = 1, 2, 3$ respectively in vector field analysis. Equipped with codifferential operators δ^k, the spaces of differential forms now constitute a bi-directional chain complex,

$$\Omega^0(M) \underset{\delta^1}{\overset{d^0}{\rightleftarrows}} \Omega^1(M) \underset{\delta^2}{\overset{d^1}{\rightleftarrows}} \Omega^2(M) \underset{\delta^3}{\overset{d^2}{\rightleftarrows}} \Omega^3(M).$$

Finally, the exterior calculus notations and their counterparts in traditional calculus are summarized in Table 4.1. The exterior calculus operations are strictly equivalent to the vector calculus operation in flat 3-dimensional space. A 0- or 3-form can be identified as a scalar function $f : M \subset \mathbb{R}^3 \to \mathbb{R}$, while a 1- or 2-form is identified with a vector field $\mathbf{v} : M \to \mathbb{R}^3$. Thus, we can use $f^0, \mathbf{v}^1, \mathbf{v}^2$ or f^3 to denote a scalar field f or vector field $\mathbf{v}$ regarded as a 0-, 1-, 2- or 3-form, respectively.

4.1.2 Hodge decomposition for manifolds

Hodge theory can be seen as the study of nonintegral parts (cohomology) of (scalar/vector) fields through the analysis of differential operators. Thus, it is often conveniently and concisely described by differential k-forms and the exterior calculus of these forms, as discussed in the previous section.

We first establish the aforementioned adjointness between the differential and codifferential operators. Through integration by part and the Stokes' theorem Eq. (4.1),

$$\langle d\alpha, \beta \rangle = \langle \alpha, \delta\beta \rangle + \int_{\partial M} \alpha \wedge \star\beta. \tag{4.3}$$

Thus, either for a boundaryless manifold ($\partial M = \emptyset$) or for forms that vanish on boundary ($\alpha|_{\partial M} = 0$ or $\star\beta|_{\partial M} = 0$), the boundary integral vanishes, i.e., $\int_{\partial M} \alpha \wedge \star\beta = 0$. In such cases, the adjointness, $\langle d\alpha, \beta \rangle = \langle \alpha, \delta\beta \rangle$, implies that d and δ satisfy the important property of adjoint operators—the kernel of a linear operator is the orthogonal complement of the range of its adjoint operator.

If we denote the space of *normal forms* as $\Omega_n^k = \{\omega \in \Omega^k | \omega|_{\partial M} = 0\}$, and the space of *tangential forms* as $\Omega_t^k = \{\omega \in \Omega^k | \star\omega|_{\partial M} = 0\}$, the orthogonal complementarity can be expressed as $\Omega^k = \ker\delta^k \oplus d\Omega_n^{k-1}$ and $\Omega^k = \ker d^k \oplus \delta\Omega_t^{k+1}$. With $\mathrm{im}\, d^{k-1} \subset \ker d^k$ (based on the property of the cochain complex $d^k d^{k-1} = 0$), the complementarity restricted to $\ker d^k$ implies

$$\ker d^k = \mathcal{H}^k \oplus d\Omega_n^{k-1}, \tag{4.4}$$

where $\mathcal{H}^k = \ker d^k \cap \ker \delta^k$ is the space of *harmonic forms*, which are defined to be both closed and coclosed. Substituting the above equation into $\Omega^k = \ker d^k \oplus \delta\Omega_t^{k+1}$, we obtain the three-component Hodge decomposition,

$$\Omega^k = d\Omega_n^{k-1} \oplus \delta\Omega_t^{k+1} \oplus \mathcal{H}^k. \tag{4.5}$$

Thus, any $\omega \in \Omega^k$ can be uniquely expressed as a sum of three k-forms from the three orthogonal subspaces,

$$\omega = d\alpha_n + \delta\beta_t + h, \tag{4.6}$$

where $\alpha_n \in \Omega_n^{k-1}$, $\beta_t \in \Omega_t^{k+1}$, and $h \in \mathcal{H}^k$. Note that the potentials α and β do not have to be unique, and a variety of gauge conditions can be specified to make them unique.

4.1.2.1 Boundaryless manifolds

When $\partial M = \emptyset$, $\Omega^k = \Omega^k_t = \Omega^k_n$, we can establish an isomorphism between the cohomology (of the de Rham complex described in the previous section) and the harmonic space, as was developed by Hodge.

In this case, Eq. (4.4) can be written as

$$\ker d^k = \mathcal{H}^k \oplus \operatorname{im} d^{k-1}. \tag{4.7}$$

Thus, we can find a unique element in $\mathcal{H}^k$ that corresponds to each equivalence class in the *de Rham cohomology* $H^k_{dR} = \ker d^k / \operatorname{im} d^{k-1}$ (quotient spaces induced by the de Rham cochain complex). This bijection implies $\mathcal{H}^k \cong H^k_{dR}$, which indicates $\mathcal{H}^k$ is a finite-dimensional space with its dimension determined by the topology of the manifold.

Moreover, we can identify $\mathcal{H}^k$ as the kernel of a particular second-order differential operator, the *de Rham-Laplace operator*, or *Hodge Laplacian*, defined as $\Delta^k \equiv d^{k-1}\delta^k + \delta^{k+1}d^k$. Through the adjointness between d and δ, we have

$$\langle \Delta\alpha, \alpha \rangle = \langle (d\delta + \delta d)\alpha, \alpha \rangle = \langle d\alpha, d\alpha \rangle + \langle \delta\alpha, \delta\alpha \rangle. \tag{4.8}$$

Denoting $\mathcal{H}^k_\Delta \equiv \ker \Delta^k$, the above equation implies that $\mathcal{H}^k_\Delta = \ker \Delta^k = \ker d^k \cap \ker \delta^k = \mathcal{H}^k$ for boundary-less manifolds.

As a direct consequence, we rewrite Eq. (4.5) as

$$\Omega^k = \operatorname{im} d^{k-1} \oplus \operatorname{im} \delta^{k+1} \oplus \mathcal{H}^k_\Delta. \tag{4.9}$$

The importance of the decomposition lies in that the first two components can be expressed as the derivatives of some potential functions, and the last non-integral part is spanned by the finite-dimensional harmonic space, whose dimension is determined by the topology of the domain due to the above-mentioned isomorphism. For example, for Ω^k with $k = 1, 2$, this decomposition is often recognized as the Helmholtz-Hodge decomposition of vector calculus in 3D, $\mathbf{v}^1 = \nabla f^0 + \nabla \times \mathbf{u}^2 + \mathbf{h}^1$, and $\mathbf{v}^2 = -\nabla f^3 + \nabla \times \mathbf{u}^1 + \mathbf{h}^2$.

type	f^0	$\mathbf{v}^1$	$\mathbf{v}^2$	f^3
tangential	unrestricted	$\mathbf{v} \cdot \mathbf{n} = 0$	$\mathbf{v} \parallel \mathbf{n}$	$f\vert_{\partial M} = 0$
normal	$f\vert_{\partial M} = 0$	$\mathbf{v} \parallel \mathbf{n}$	$\mathbf{v} \cdot \mathbf{n} = 0$	unrestricted

Table 4.2: **Boundary conditions of tangential and normal form.**

4.1.2.2 Manifolds with boundary

For 3-manifolds with 2-manifold boundary, we need additional boundary conditions to have a finite dimensional kernel for the Laplacians, as in this case, $\mathcal{H} = \ker d \cap \ker \delta \subsetneq \mathcal{H}_\Delta$. Through integration by part with the boundary, we have

$$\langle \Delta \alpha, \alpha \rangle = \langle (d\delta + \delta d)\alpha, \alpha \rangle = \langle d\alpha, d\alpha \rangle + \langle \delta \alpha, \delta \alpha \rangle + \int_{\partial M} (\delta \alpha \wedge \star \alpha - \alpha \wedge \star d\alpha). \qquad (4.10)$$

Thus, if we can eliminate the boundary integral by restricting the space of forms, the kernel of Δ will be the intersection of the kernel of d and δ. Indeed, there are a variety of choices to satisfy boundary conditions, e.g., forcing the support of the differential form to be in the interior of manifolds. However, an option that is consistent with common physical boundary conditions is to restrict the differential form α in the decomposition to be tangential to the boundary $\star \alpha\vert_{\partial M} = 0$ or normal to the boundary $\alpha\vert_{\partial M} = 0$ as we have required for the potentials.

Then, one natural choice to eliminate both terms in the boundary integral is to force $d\alpha$ to be tangential when α is tangential and force $\delta \alpha$ to be normal when α is normal. In other words, we modify the definition Ω_t to be the space of tangential forms with tangential differential, i.e., $\alpha_t \in \Omega_t$ if and only if

$$\star \alpha_t\vert_{\partial M} = 0, \quad \star d\alpha_t\vert_{\partial M} = 0. \qquad (4.11)$$

Similarly, we modify the definition of Ω_n to be the space of normal forms with normal codifferential, i.e., $\alpha_n \in \Omega_n$ if and only if

$$\alpha_n\vert_{\partial M} = 0, \quad \delta \alpha_n\vert_{\partial M} = 0. \qquad (4.12)$$

To illustrate the boundary conditions explicitly, we consider a moving frame, which is formed at each boundary point by two tangent vectors of the boundary surface $\mathbf{t}_1$ and $\mathbf{t}_2$ and the normal vector

to the surface $\mathbf{n}$, with the typical convention that they form a right-hand orthonormal frame with the normal pointing outward. As a 1-form $\mathbf{v}^1$ is tangential if $\star \mathbf{v}^1(\mathbf{t}_1, \mathbf{t}_2) = \mathbf{v}^2(\mathbf{t}_1, \mathbf{t}_2) = \mathbf{v} \cdot (\mathbf{t}_1 \times \mathbf{t}_2) = \mathbf{v} \cdot \mathbf{n} = 0$, it matches the condition that the corresponding vector field is tangential to the boundary. Similarly, a 1-form $\mathbf{v}^1$ is normal to the boundary, if $\mathbf{v}^1(\mathbf{t}_i) = \mathbf{v} \cdot \mathbf{t}_i = 0$ for $i = 1, 2$, thus it is the equivalent to $\mathbf{v}$ is normal to the boundary. For a 2-form $\mathbf{v}^2$, its normal (tangential) boundary condition is the same as the tangential (normal) boundary condition of $\mathbf{v}^1$. Therefore, normal (tangential) 2-forms should have their corresponding vector fields tangential (normal, resp.) to the boundary. Additionally, tangential 3-forms (normal 0-forms) are zero on the boundary whereas normal 3-forms (tangential 0-forms) automatically satisfy the boundary condition. In Table 4.2, we summarized these choices of the boundary conditions for tangential and normal k-forms in 3D.

In vector field representation, the boundary conditions Eqs. (4.11) and (4.12) are equivalent to the following. The choice of a 1-form in Ω_t^1 (a 2-form in Ω_n^2) is equivalent to enforcing a tangential vector field $\mathbf{v}$ to have its curl to be normal to the boundary, i.e., adding two homogeneous Neumann boundary conditions to the (Dirichlet-type) tangentiality,

$$\mathbf{v} \cdot \mathbf{n} = 0, \quad \nabla_{\mathbf{n}}(\mathbf{v} \cdot \mathbf{t}_1) = 0, \quad \nabla_{\mathbf{n}}(\mathbf{v} \cdot \mathbf{t}_2) = 0. \tag{4.13}$$

For a normal vector field $\mathbf{v}$ (1-forms in Ω_n^1 or 2-forms in Ω_t^2), it amounts to adding one homogeneous Neumann boundary condition derived from the zero divergence on the boundary to the (Dirichlet-type) orthogonality constraints,

$$\mathbf{v} \cdot \mathbf{t}_1 = 0, \quad \mathbf{v} \cdot \mathbf{t}_2 = 0, \quad \nabla_{\mathbf{n}}(\mathbf{v} \cdot \mathbf{n}) = 0. \tag{4.14}$$

For an unrestricted function f (tangential 0-forms or normal 3-forms), it amounts to forcing its gradient to be tangential at the boundary (Neumann-type),

$$\nabla_{\mathbf{n}} f|_{\partial M} = 0, \tag{4.15}$$

and a function f for tangential 3-forms (normal 0-forms) satisfies the homogeneous Dirichlet boundary condition

$$f|_{\partial M} = 0. \tag{4.16}$$

92

With these modified boundary conditions, we still have the same Hodge decomposition,

$$\Omega^k = d\Omega_n^{k-1} \oplus \delta\Omega_t^{k+1} \oplus \mathcal{H}^k. \tag{4.17}$$

This is due to the fact that $d\Omega_n$ (or $\delta\Omega_t$) remains the same regardless of whether Ω_n (or Ω_t) contains the additional boundary conditions, as they can be seen as part of the gauge condition that restricts the potentials but not their differential (codifferential).

As mentioned above, with the boundary, $\mathcal{H}^k$ is no longer finite dimensional or the kernel of of Laplacians $\mathcal{H}_\Delta^k$. However, if we restrict Δ to Ω_t or Ω_n and denote the corresponding operator as Δ_t and Δ_n respectively, we can still find finite dimensional kernels $\mathcal{H}_{\Delta_t}^k$ and $\mathcal{H}_{\Delta_n}^k$ that correspond to $\mathcal{H}^k \cap \Omega_t$ or $\mathcal{H}^k \cap \Omega_n$ orthogonal to $\mathrm{im}\, d$ and $\mathrm{im}\, \delta$.

In fact, the harmonic space $\mathcal{H}^k$ can be further decomposed into tangential, normal harmonic forms and exact-coexact harmonic forms $\mathcal{H}^k = (\mathcal{H}_{\Delta_t}^k + \mathcal{H}_{\Delta_n}^k) \oplus (d\Omega^{k-1} \cap \delta\Omega^{k+1})$ as proposed by Friedrichs [135]. Moreover, in flat 3D space, all three subspaces are orthogonal to each other. The third space can be seen as the infinite-dimensional space of solutions to Laplace equations in dimension $k \pm 1$ with either normal or orthogonal boundary conditions. Thus, we can focus on the Laplacian operators that are either tangential or normal for analysis.

In total, there are 8 different Hodge Laplacians (Δ_t^k and Δ_n^k for $k = 0, 1, 2, 3$) and 8 associated finite dimensional harmonic spaces. Friedrichs also noted that for manifolds with boundary, the tangential harmonic spaces are isomorphic to the absolute de Rham cohomology $\mathcal{H}_{\Delta_t}^k \cong H^k(M)$, and the normal harmonic spaces are isomorphic to the relative de Rham cohomology $\mathcal{H}_{\Delta_n}^k \cong H^k(M, \partial M)$. From the dimensionality of the corresponding homology (Betti numbers) of the manifold M, together with the Hodge duality between $\mathcal{H}_{\Delta_t}^k$ and $\mathcal{H}_{\Delta_n}^{3-k}$, we can obtain the dimensions of all these harmonic spaces: $\beta_k = \dim \mathcal{H}_{\Delta_t}^k = \dim \mathcal{H}_{\Delta_n}^{3-k}$. Roughly, speaking, β_0 is the number of connected components, β_1 is the number of rings, β_2 is the number of cavities, and β_3 is 0 as M in flat 3D cannot contain any noncontractible topological 3-sphere.

4.1.3 Discrete forms and spectral analysis

In practical applications, the de Rham-Hodge theory is often computed for decompositions and spectral analysis. In both cases, the discretization of exterior derivatives is required. We follow one typical discretization of the exterior calculus on differential forms, the discrete exterior calculus (DEC) [83]. A major technical aspect is the handling of arbitrarily complex geometric shapes in 3D. In spectral analysis, the Hodge Laplacian operators and their boundary conditions are to be implemented such that the key topological property of $d \circ d = 0$, which defines the de Rham cohomology, is preserved in the discrete version by DEC in complex computational domains. First, the domain of differential forms, in this case, a 3-manifold embedded in 3D Euclidean space is tessellated into a 3D simplicial complex, i.e., a tetrahedral mesh. Any k-form ω is represented by its integral on oriented k-D elements (k-simplex) of the mesh, listed as a vector W with the length equaling the number of k-simplices. More specifically, a discrete 0-form is the assignment of one real number per vertex, a discrete 1-form is the assignment of one value per oriented edge, a discrete 2-form is the assignment of one value per oriented triangle, and a discrete 3-form is the assignment of one value per tetrahedron (tet). The choice of orientation per k-simplex is arbitrary since the antisymmetry of a k-form guarantees that the integral on that k-simplex only changes its sign.

Now the linear operator d^k is represented by a sparse matrix D_k, which is implemented as the transpose of the signed incidence matrix between k-simplices and $(k+1)$-simplices, with the sign determined by mutual orientation. Furthermore, an arbitrary orientation for each k-simplex is chosen up to an even permutation, which is an order set of $k+1$ vertices. An oriented k-simplex is defined as

$$\sigma = [v_0, v_1, ..., v_k]. \tag{4.18}$$

The boundary operator ∂ is defined as

$$\partial \sigma = \sum_{i=0}^{k} (-1)^i [v_0, v_1, ..., \hat{v}_i, ..., v_k], \tag{4.19}$$

where $\hat{v}_i$ means that the ith vertex is removed. The discrete boundary operator will take all the 1-degree discrete lower faces of σ with an induced orientation. Thus the discrete exterior deriva-

tive operator D_k is just a matrix filled with $-1, 0, 1$. The discrete Hodge star matrices S_k is just converting primal forms and dual forms by the following equation

$$\frac{1}{|\sigma_k|} \int_{\sigma_k} \omega = \frac{1}{|*\sigma_k|} \int_{*\sigma_k} \star\omega. \tag{4.20}$$

Thus, the discrete Hodge star operator is a diagonal matrix. This can be seen as the consequence of the aforementioned Stokes' theorem, because the integral of $d\omega$ on each $(k+1)$-simplex is exactly the sum of the integral of ω on the boundary of the $(k+1)$-simplex, which is the union of its consistently oriented k-simplex faces.

Thus, the defining property in de Rham-Hodge theory $D_{k+1}D_k = 0$ is preserved through as the boundary of the boundary is empty. As shown in Fig. 4.1, the adjoint operator δ^k is implemented as $S_{k-1}^{-1} D_{k-1}^T S_k$, where S_k is discretization of the L_2-inner product between two discrete k-forms such that $(W_1^k)^T S_k W_2^k$ is an approximation of $\langle \omega_1^k, \omega_2^k \rangle$. In this work, we use the lowest order diagonal matrices for S_k for simplicity, but higher-order Galerkin matrices for k-form basis can be developed with proper treatment on matrix inversion for better accuracy. Such a discrete Hodge star operator can also be seen as a mapping from a discrete k-form to a discrete dual $(3-k)$-form defined on the basis associated with dual elements of a dual mesh to the tet mesh. Obviously, this field needs more effort from the computational mathematics community

With both the differential operators and the Hodge stars discretized, the discrete counterpart of a Hodge Laplacian Δ^k is defined as $S_k^{-1} L_k$ through products and summations of these matrices following the continuous version, here

$$L_k = D_k^T S_{k+1} D_k + S_k D_{k-1} S_{k-1}^{-1} D_{k-1}^T S_k. \tag{4.21}$$

The reason that L_k is used frequently as the discrete Hodge Laplacian instead of $S_k^{-1} L_k$ is its symmetry. Alternatively, we can also see L_k as the quadratic form on the space of discrete k-forms, such that $W^T L_k W$ is an approximation of $\langle \omega, \Delta\omega \rangle$.

In our analysis of volumetric shapes, we conjecture that the evolution of topological and geometric structures is related not only to the null spaces of Hodge Laplacians, but also to the general spectra of these operators, in particular, those eigenvalues that are close to zero. The associated

$$
\begin{array}{ccccccc}
\text{0-form} & \xrightarrow{\;D_0\;} & \text{1-form} & \xrightarrow{\;D_1\;} & \text{2-form} & \xrightarrow{\;D_2\;} & \text{3-form} \\
S_0^{-1}\!\uparrow\!\downarrow S_0 & & S_1^{-1}\!\uparrow\!\downarrow S_1 & & S_2^{-1}\!\uparrow\!\downarrow S_2 & & S_3^{-1}\!\uparrow\!\downarrow S_3 \\
\text{dual 3-form} & \xleftarrow[D_0^T]{} & \text{dual 2-form} & \xleftarrow[D_1^T]{} & \text{dual 1-form} & \xleftarrow[D_2^T]{} & \text{dual 0-form}
\end{array}
$$

Figure 4.1: **Discrete de Rham cohomology.** D_k is the combinatorial operators such that $D_{k+1}D_k = 0$; S_k is the discrete Hodge stars.

eigen differential forms can be found through a generalized eigenvalue problem for the discrete Hodge Laplacian and Hodge star operators.

$$
L_k W^k = \lambda^k S_k W^k. \tag{4.22}
$$

For illustration purpose, we can reformulate Eq. (4.22) as a regular eigenvalue problem,

$$
\bar{L}_k \bar{W}^k = \lambda^k \bar{W}^k, \tag{4.23}
$$

where $\bar{L}_k = S_k^{-1/2} L_k S_k^{-1/2}$ and $\bar{W}^k = S_k^{1/2} W^k$. Then, to partition the spectrum of the modified discrete Hodge Laplacian, we express it as the sum of two semi-positive-definite matrices,

$$
\bar{L}_k = \bar{D}_k^T \bar{D}_k + \bar{D}_{k-1}\bar{D}_{k-1}^T, \tag{4.24}
$$

where $\bar{D}_k = S_{k+1}^{1/2} D_k S_k^{-1/2}$. We can observe that the cohomology structure is maintained as $\bar{D}_{k+1}\bar{D}_k = 0$. Moreover, now the adjoint operator of $\bar{D}_k$, in the L_2 inner products defined by the Hodge stars, is simply its transpose $\bar{D}_k^T$. Thus, the entire spectrum of $\bar{L}_k$ can be studied through the singular value decomposition of the discrete differential operator

$$
\bar{D}_k = U_{k+1} \Sigma_k V_k^T, \tag{4.25}
$$

where U_{k+1} and V_k are orthogonal matrices, and Σ_k is a rectangular diagonal matrix with non-negative real elements. We can recognize the nonzero spectra of the modified Hodge Laplacian as the union of the squares of the nonzero entries from Σ_k and Σ_{k-1}, since

$$
\bar{L}_k = V_k \Sigma_k^2 V_k^T + U_k \Sigma_{k-1}^2 U_k^T. \tag{4.26}
$$

Note that for 0- or 3-forms, one of the Σ's contains only zeros.

Based on the Hodge decomposition Eq. (4.17), we can also notice that the columns of V_k that correspond to nonzero singular values in Eq. (4.26) are orthogonal to those of U_k, which means the entire k-form space is spanned by harmonic forms (eigen form with eigenvalue 0), and those column vectors of V_k and U_k.

For domains with boundaries, the tangential or normal forms are restricted by Dirichlet and/or Neumann boundary conditions, which can be implemented by whether to include the boundary elements or not for D_k. We denote the discrete differential operator for tangential (normal) k-forms as $D_{k,t}$ (respectively $D_{k,n}$). For the detail on the construction of these matrices, readers are referred to our previous work [72]. In summary, for the four types of k-form ($k = 0, 1, 2, 3$) with two boundary conditions, there are 8 different discrete Hodge Laplacians ($L_{k,t}$ and $L_{k,n}$) in total, such that

$$
\begin{aligned}
L_{k,t} &= D_{k,t}^T S_{k+1} D_{k,t} + S_k D_{t,k-1} S_{k-1}^{-1} D_{t,k-1}^T S_k, \\
L_{k,n} &= D_{k,n}^T S_{k+1} D_{k,n} + S_k D_{n,k-1} S_{k-1}^{-1} D_{n,k-1}^T S_k.
\end{aligned}
\tag{4.27}
$$

Based on the above singular value analysis, the non-zero spectrum of $\bar{L}_k$ is the union of squared singular values of $\bar{D}_k$ and those of $\bar{D}_{k-1}$. Therefore, for each type of boundary conditions, the spectra of the four discrete Hodge Laplacians only depend on the singular spectra of $\bar{D}_0$, $\bar{D}_1$ and $\bar{D}_2$. Furthermore, in Table 4.2, the same set of boundary conditions is shared between tangential 1-forms and the normal 2-forms, between tangential 2-forms and normal 1-forms, between normal 3-forms and tangential 0-forms, and between tangential 3-forms and normal 0-forms. This duality between tangential k-forms and normal $(3-k)$-forms is also present in the corresponding operators between these forms, more specifically, the equivalence exists between $\bar{D}_{0,t}$ and $\bar{D}_{2,n}^T$, $\bar{D}_{1,t}$ and $\bar{D}_{1,n}^T$, and $\bar{D}_{2,t}$ and $\bar{D}_{0,n}^T$. We thus reduce the 8 different spectra of Hodge Laplacians to 3 distinct sets of different singular spectra. We denote the set of singular values of $\bar{D}_{0,t}$ for the tangential gradient eigen field by T, the set of the singular values of $\bar{D}_{1,t}$ for the curl eigen field by C, and the set of the singular value set of $\bar{D}_{2,t}$ for tangential divergent eigen field by N.

Although each of the 8 spectra for Hodge Laplacians defined on smooth manifolds can be rep-

resented by the combination of one or two sets of the T, C and N, the numerical calculations of the singular values of the equivalent differential operators can deviate from these due to the different DoFs in the representations for different discrete forms, as well as the inaccuracy introduced by the approximation of Hodge star and differential operators. While the numerically computed singular values of tangential k-forms $\bar{D}_{k,t}$ can deviate from those of normal $(3-k)$-forms $\bar{D}_{2-k,n}^{T}$, as the observation in previous work [82], with increased resolution, the low frequencies converge reasonably well.

4.2 Evolutionary de Rham-Hodge method

In this section, we introduce the evolutionary de Rham-Hodge method to analyze the topological and geometric properties throughout the evolution of manifolds. We first discuss the existing data that motivates the present theoretic formulation. Then, we provide the mathematical description of manifold evolution, followed by the definitions of the associated persistence and progression. We extend the usual study of cohomology (associated to zero eigenvalues of Hodge Laplacians) to employing the leading small non-zero eigenvalues to facilitate the concepts of persistence and progression so that the variations of topological spaces (β_0, β_1 and β_2) can be traced to the changes in the eigenvalues away from or towards zero as the geometry evolves.

4.2.1 Data and their de Rham-Hodge analysis

Most commonly occurred data are closed manifolds, such as star surfaces, earth surfaces, brain surfaces, and molecular surfaces. The de Rham Laplace operator can be applied to compute eigenfunctions and eigenvalues for the geometric shape analysis. Another interesting type of data includes scalar or vector functions defined on closed manifolds, such as temperature or ocean currents on the earth's surface and in compact manifolds with boundaries, such as the electron densities or electrostatic potentials in proteins or the magnetic fields around the earth. The Hodge decomposition can be directly applied to these functions. For smooth scalar functions, surface contours can be specified to generate compact manifolds with boundaries. The geometric shape analysis via the de

Rham Laplace operator can be carried out. A special class of data is the density distributions, either obtained from cryogenic electron microscopy (cryo-EM), magnetic resonance imaging (MRI) or created from quantum mechanical calculations. In this situation, one can render a family of inclusion surfaces by systematically varying the density isovalues. The de Rham-Hodge analysis and modeling of this family of inclusion surfaces are the objects of the present theoretical development.

The evolutionary de Rham-Hodge method developed in this work can also be applied to point cloud data, such as stars in the universe, atoms in biomolecules, and the output of 3D scanning processes. In this situation, one can carry out a discrete to continuum map to create volumetric density functions from point clouds [2, 181]. Then, a family of inclusion surfaces can be obtained for the evolutionary de Rham-Hodge analysis.

Flexibility rigidity index (FRI) density is a useful tool to construct a continuous density distribution from a set of discrete point cloud data inputs. By selecting an isovalue from the FRI density, one can further generate a boundary surface, which composes the 3-manifold with a 2-manifold boundary. Moreover, one can also use the Gaussian dielectric function to generate density distributions [182, 183]. FRI density has been shown to be particularly straightforward to implement and computationally stable on any point cloud [181] and is defined by the following position-dependent rigidity (or density) function [2]

$$\rho(\mathbf{r}, \eta) = \sum_{j=1}^{N} \Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta) \tag{4.28}$$

where $\mathbf{r}$ is a point in space, N is the number of particles, $\mathbf{r}_j$ is the location of a data point j, η is a scaling parameter, and $\Phi(\cdot; \eta)$ is a correlation function, i.e., a real-valued monotonically decreasing function with the following admissibility conditions

$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta) = 1, \text{ as } \|\mathbf{r} - \mathbf{r}_j\| \to 0,$$
$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta) = 0, \text{ as } \|\mathbf{r} - \mathbf{r}_j\| \to \infty, \tag{4.29}$$

One used families of correlation functions is the generalized exponential functions

$$\Phi(\|\mathbf{r} - \mathbf{r}_j\|; \eta) = \exp(-(\|\mathbf{r} - \mathbf{r}_j\|/\eta)^{\kappa}), \quad \kappa > 0. \tag{4.30}$$

Here, the weight η is application-dependent, e.g., the multiplication of a scaling parameter and the van der Waals radius $r_{textrmvdw_j}$ of the atom at $\mathbf{r}_j$ for molecular data. In fact, η can be chosen as anisotropic function to induce a multidimensional persistent homology filtration [184]. In our numerical tests, we use the generalized exponential function with $\kappa = 2$, which is known as the Gaussian function. A family of 3-manifolds can be defined by a varying level set parameter (isovalue) $c \in (0, c_{\max})$, where $c_{\max} = \max \rho(\cdot, \eta)$,

$$M_c = \{\mathbf{r}|\rho(\mathbf{r}, \eta) \leq c_{\max} - c\}, \tag{4.31}$$

which has the level-set of ρ as its boundary $\partial M_c = \{\mathbf{r}|\rho(\mathbf{r}, \eta) = c_{\max} - c\}$.

4.2.2 Manifold evolution

Hodge theory studies the de Rham cohomology groups of a smooth manifold M, and established the bijection from equivalence classes in a cohomology group to a harmonic differential form in the null space of the corresponding Hodge Laplacian. While these harmonic forms associated with the zero eigenvalues in the spectra of Hodge Laplacians carry some geometric information in addition to the topology, the non-zero spectra provide richer geometric information than the multiplicity of zero. However, the geometry is not uniquely determined by the spectra of the Hodge Laplacians (even for planar shapes), as one cannot hear the shape of a drum [185]. Thus, we propose to extend the study of de Rham-Hodge theory to a family of smooth manifolds instead of one specific manifold and track the spectral changes in a sequence of manifolds. Such a family of manifolds controlled by a continuous filtration parameter is sometimes called the evolution of manifolds embedded in an ambient manifold, which in our case is the 3D Euclidean space.

The evolution of manifolds is often defined through a smooth map from a basic manifold B to a family of submanifold $\{M_c\}$ of an ambient manifold M at a given instant (the value of parameter c treated as time). More precisely, it is the smooth map $F : B \times [0, c_{\max}] \to M$ such that $F^c = F(\cdot, c)$ is an immersion for every c. The one-parameter family of subsets of M, $\{F^c(B)\}_{c \geq 0}$ is then called the *evolving manifold*. However, such a Hodge Lagrangian description makes it hard to handle

topological changes, especially if each mapping is restricted to be an embedding. Therefore, in this work, we directly use the Eulerian representation described by M_c in Eq. (4.31). This level-set bounded volume evolution handles both the geometric progression and topological changes in a consistent fashion. As Morse functions are dense in continuous functions, we can assume $\rho(\mathbf{r}, \eta)$ to be a Morse function without loss of generality, since otherwise, we can use symbolic perturbation to make it a Morse function. We can regularly sample the interval $(0, c_{\max})$ at n sample locations, forming an index set $I = \{c_0, c_1, ..., c_n\}$, such that none of the parameters are one of the isolated critical values through symbolic perturbation if necessary. Noting that M_c are only non-manifold when c is a critical point of the Morse function, the snapshots of the evolving manifold, $\{F^c\}_{c \in I}$, are all manifolds. Thus, they form a filtration of manifold M, with the inclusion map $\mathfrak{I}_{l,l+1} : M_l \hookrightarrow M_{l+1}$ linking each pair of consecutive manifolds and

$$M_0 \xrightarrow{\mathfrak{I}_{0,1}} M_1 \xrightarrow{\mathfrak{I}_{1,2}} M_2 \xrightarrow{\mathfrak{I}_{2,3}} \cdots \xrightarrow{\mathfrak{I}_{n-1,n}} M_n \xrightarrow{\mathfrak{I}_{n,n+1}} M = M_{c_{\max}}.$$

If (c_l, c_{l+p}) does not contain any critical points of $\rho(\mathbf{r}, \eta)$ and the largest critical value smaller than c_l is c_c, the inclusion map $\mathfrak{I}_{l,l+p} : M_l \hookrightarrow M_{l+p}$ is also homotopic to a homeomorphism from M_l to M_{l+p}, which can be constructed by moving every point $\mathbf{r}$ with $\rho(\mathbf{r}, \eta) > c_{\max} - c_c$ along the gradient integral line of $\rho(\cdot, \eta)$ to a point $\hat{\mathbf{r}}$ such that $\rho(\mathbf{r}, \eta) - \rho(\hat{\mathbf{r}}, \eta) = (c_{l+p} - c_l)e^{1 - \frac{c_l - c_c}{\rho(\mathbf{r}, \eta) - c_{\max} - c_c}}$. When the two parameter values are similar, one can also see that the above map is nearly isometric since the deformation is close to an identity map.

When (c_l, c_{l+p}) contains critical points of the Morse function, there is no smooth homeomorphism between M_l and M_{l+p} as the level set underwent topological changes. Without loss of generality, we can assume that there is only one critical point, which can be classified as (local) minimum, 1-saddle, 2-saddle, or (local) maximum, based on the signature of the Hessian of ρ. As all minima of ρ is at the value of 0, the interval may only contain the latter three types: if it is a maximum, one 2nd homology generator in M_l will be mapped to 0 in M_{l+p} for the mapping induced by the inclusion; if it is a 2-saddle, either M_l has a 1st homology generator mapped to 0 or M_{l+p} contains a 2nd homology generator not in the image of the induced mapping from $H(M_l)$

to $H(M_{l+p})$; similarly, if it is a 1-saddle, either M_l has a 0th homology generator mapped to 0 or M_{l+p} contains a 1st homology generator, not in the image of the induced mapping. Through the isomorphisms among the de Rham cohomology, singularly homology, simplicial homology, and simplicial cohomology, we can use the persistent homology to study the mapping between the de Rham cohomologies indirectly. However, we found that direct construction can reveal some additional insight on the relation and persistence of the harmonic forms across different manifolds, as we discuss next.

4.2.3 Persistence of harmonic forms

4.2.3.1 Normal harmonic forms

Drawing an analogy from persistent homology, we first attempt to construct a homomorphism from closed forms on M_l to closed forms on M_{l+p}, i.e., from $\ker d_l$ to $\ker d_{l+p}$, if we use the subscript l to denote the operator defined on M_l. For manifolds with boundary, one realizes that this is not possible for tangential forms through the isomorphism relations to cochain and chain spaces on simplicial complexes, but rather straightforward for normal forms in the discrete case. More specifically, we can map k-forms in M_l by setting values for simplices in $M_{l,p}^c = M_{l+p} \backslash M_l$ to 0, i.e., a 0-padded k-cochain on M_{l+p} as the image of a k-cochain on M_l assuming that M_l has a tessellation that is a subcomplex of the tessellation of M_{l+p}. The reason that the image of $\omega_l \in \ker d_l$ remains in $\ker d_{l+p}$ is that the value of $d\omega_{l+p}$ on any $(k{+}1)$-simplex with one or more faces in ∂M_l is still 0, as $\omega_l|_{\partial M_l} = 0$.

However, in the continuous case, setting ω to 0 in $M_{l,p}^c$ creates either discontinuity or at least large $\delta\omega$ near the boundary. A smoother extension of the ω from M_l to M_{l+p} can be defined by minimizing the Dirichlet energy $\langle d\omega, d\omega \rangle + \langle \delta\omega, \delta\omega \rangle$ in $M_{l,p}^c$, which leads to simply a Laplace equation $\Delta\omega = 0$. The boundary of $M_{l,p}^c$ is the union of ∂M_l and ∂M_{l+p} with the orientation of the former flipped. Recall that when ω is normal to the boundary i.e., $\omega_l|_{\partial M_l} = 0$, we also impose the condition that $\delta\omega$ is normal to the boundary ($\delta\omega_l|_{\partial M_l} = 0$). For the extension, we keep this

condition on ∂M_{l+p}, while on ∂M_l we impose the continuity instead, $\omega_{l+p}|_{\partial M_l} = \omega_l|_{\partial M_l}$. Note that the resulting Laplace equation has a finite kernel identical to that of Δ_n on $M_{l,p}^c$, so we can find a unique solution by forcing the solution to have 0 projection to this kernel [72].

For instance, if we have a normal 1-form ω_l to extend, we can impose the homogeneous boundary condition for the proxy vector field $\mathbf{v}$ on ∂M_{l+p} as in Eq. (4.14),

$$\mathbf{v}_{l+p} \cdot \mathbf{t}_1 = 0, \quad \mathbf{v}_{l+p} \cdot \mathbf{t}_2 = 0, \quad \nabla_{\mathbf{n}}(\mathbf{v}_{l+p} \cdot \mathbf{n}) = 0; \tag{4.32}$$

whereas on ∂M_l, we use a Dirichlet boundary condition for continuity $\mathbf{v}_{l+p} = \mathbf{v}_l$, i.e.,

$$\mathbf{v}_{l+p} \cdot \mathbf{n} = \mathbf{v}_l \cdot \mathbf{n}, \quad \mathbf{v}_{l+p} \cdot \mathbf{t}_1 = 0, \quad \mathbf{v}_{l+p} \cdot \mathbf{t}_2 = 0. \tag{4.33}$$

We denote the map through this harmonic extension as $\mathfrak{E}_{l,p}$, i.e., $\omega_{l+p} = \mathfrak{E}_{l,p}(\omega_l)$. However, the minimization of Dirichlet energy does not imply $\delta\omega_{l+p} = 0$ even when $\delta\omega_l = 0$. Nevertheless, $d\omega_{l+p} = 0$ is always possible, since otherwise, one would be able to perform a Hodge decomposition to find a tangential $(k+1)$-form β_t in $M_{l,p}^c$ and remove $d\omega_{l+p}$ by subtracting $\delta\beta_t$ from ω_{l+p}. An alternative is to restrict the extension to minimize $\langle \delta\omega, \delta\omega \rangle$ under the constraint $d\omega_{l+p} = 0$ in $M_{l,p}^c$, which results in a fourth-order bi-Laplace equation. Since this discussion is mainly for theoretical purposes, we assume the simple harmonic extension followed by a decomposition to enforce $d\omega_{l+p} = 0$ instead of a biharmonic extension. In Fig. 4.2 (a), we illustrate the implementation of boundary conditions for the extension of normal harmonic forms to the interior cavity. In this evolving process, the outside surface is fixed and the inner cavity shrinks to null in order that the manifold with a cavity extends into a solid ball. Under the boundary condition Eq. (4.33) on the interior surface, the input normal harmonic forms (thin lines) are extended into the cavity, which also preserve curl-free properties shown as thick lines in Fig. 4.2 (a).

Note that $d\mathfrak{E}(\omega)$ is a solution to the equation for solving the extension of $d\omega$, by the uniqueness we impose, it must be $\mathfrak{E}(d\omega)$. Thus, we can construct the following commutative diagram on the de Rham complexes for normal forms on the filtration of M:

(a) Normal harmonic forms   (b) Tangential harmonic forms
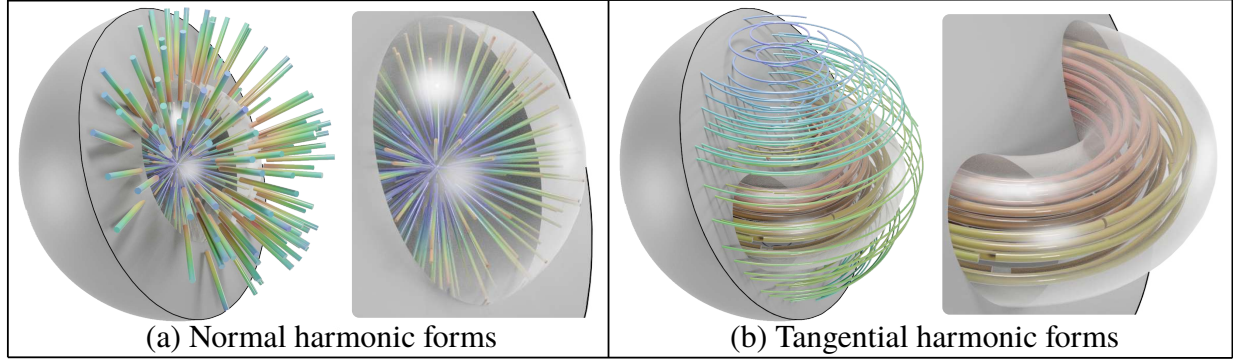
Figure 4.2: **Illustration of normal and tangential harmonic field extensions.** Thick lines are the inputs and thin lines are the extended outputs. Left charts in both (a) and (b) show harmonic fields and their extensions while right charts give meticulous detail of interior parts. (a) Normal harmonic forms. A solid ball with a cavity extends inward to a solid ball without cavity. The outside surface is fixed. (b) Tangential harmonic forms. A torus extends to a solid ball.

$$
\begin{array}{ccccccc}
\Omega_n^0(M_0) & \xrightarrow{d^0} & \Omega_n^1(M_0) & \xrightarrow{d^1} & \Omega_n^2(M_0) & \xrightarrow{d^2} & \Omega_n^3(M_0) \\
\downarrow{\scriptstyle \mathfrak{E}_{0,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{0,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{0,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{0,1}} \\
\Omega_n^0(M_1) & \xrightarrow{d^0} & \Omega_n^1(M_1) & \xrightarrow{d^1} & \Omega_n^2(M_1) & \xrightarrow{d^2} & \Omega_n^3(M_1) \\
\downarrow{\scriptstyle \mathfrak{E}_{1,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{1,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{1,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{1,1}} \\
\Omega_n^0(M_2) & \xrightarrow{d^0} & \Omega_n^1(M_2) & \xrightarrow{d^1} & \Omega_n^2(M_2) & \xrightarrow{d^2} & \Omega_n^3(M_2) \\
\downarrow{\scriptstyle \mathfrak{E}_{2,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{2,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{2,1}} & & \downarrow{\scriptstyle \mathfrak{E}_{2,1}} \\
\cdots & & \cdots & & \cdots & & \cdots
\end{array}
$$

which places the de Rham complex in the horizontal direction and the filtration-induced extensions in the vertical direction.

Now, we can discuss the direct relation of bases of normal harmonic forms induced by $\mathfrak{E}$. First, $\omega_n \in \ker d_l$ implies $\mathfrak{E}_{l,p}(\omega_n) \in \ker d_{l+p}$. Thus, there is an injective homomorphism from $\ker d_l$ to $\ker d_{l+p}$. This induces a homomorphism from the cohomology group $\ker d_l^k / \mathrm{im}\ d_l^{k-1}$ to $\ker d_{l+p}^k / \mathrm{im}\ d_{l+p}^{k-1}$, which, through de-Rham isomorphism between cohomology and harmonic spaces in M_l and M_{lp}, is equivalent to a homomorphism from the harmonic space $\mathcal{H}_{\Delta n,l}^k$ to $\mathcal{H}_{\Delta n,l+p}^k$. Instead of using the mapping between the equivalence classes, we can actually directly pick the unique harmonic representative $h_n \in \ker d^k \cup \ker \delta^{k+1} = \mathcal{H}_{\Delta_n}^k$ for each equivalence class in the cohomology, as we can pick the closed form that is orthogonal to $\mathrm{im}\ d^{k-1}$ which is

$\ker \delta^k$ due to the adjointness between d and δ. However, for $h_n \in \mathcal{H}^k_{\Delta n,l}$, its extension $\mathfrak{E}_{l,p}(h_n)$ is not necessarily an element of $\mathcal{H}^k_{\Delta n,l+p}$. Nevertheless, composed with the simple L_2 projection onto the finite dimensional normal harmonic space $P_{\mathcal{H}^k_{\Delta n,l+p}}$, we have the linear map (also a homomorphism) $\Psi_{n,l,p} = P_{\mathcal{H}^k_{\Delta n,l+p}} \circ \mathfrak{E}_{l,p} : \mathcal{H}^k_{\Delta n,l} \to \mathcal{H}^k_{\Delta n,l+p}$.

The map between these two normal harmonic spaces is neither necessarily injective nor necessarily surjective. In fact, if $h_n \in \mathcal{H}^k_{\Delta n,l}$ is not in $\operatorname{im} \Psi_{n,l-1,1}$, it is said to be born at index l; if p is the smallest integer such that $\Psi_{n,l,p}(h_n) = 0$, it is said to die at index $l+p$, with a persistence of p. This is consistent with the persistence of the relative cohomology $H^k(M, \partial M)$ and the (absolute) homology $H_{3-k}(M)$.

4.2.3.2 Tangential harmonic forms

As there is a one-to-one correspondence between tangential k-forms and normal $(3-k)$-forms, it is indeed sufficient to study the tangential forms only. For completeness and flexibility in numerical implementation, we provide a brief discussion on this dual case.

We first note that there is a homomorphism from coclosed forms on M_l to coclosed forms on M_{l+p}, i.e., from $\ker \delta_l$ to $\ker \delta_{l+p}$ when restricted to tangential forms $\Omega_t(M_l)$. The same harmonic extension $\mathfrak{E}_{l,p}$ can be obtained through the minimization of the Dirichlet energy $\langle d\omega, d\omega \rangle + \langle \delta\omega, \delta\omega \rangle$ in $M^c_{l,p}$. For tangential forms, $\star\omega_l|_{\partial M_l} = 0$, we also impose the condition that $d\omega$ is tangential to the boundary ($\star d\omega_l|_{\partial M_l} = 0$). We keep this condition on ∂M_{l+p}, on ∂M_l we impose continuity $\omega_{l+p}|_{\partial M_l} = \omega_l|_{\partial M_l}$ and $d\omega_{l+p}|_{\partial M_l} = d\omega_l|_{\partial M_l}$. A unique solution is again found by forcing it to have 0 projection to the kernel of a mixed-type boundary condition Laplace equation [72].

To illustrate it with a tangential 1-form ω_l, we can impose the homogeneous boundary condition for the proxy vector field $\mathbf{v}$ on ∂M_{l+p} as in Eq. (4.13),

$$\mathbf{v}_{l+p} \cdot \mathbf{n} = 0, \quad \nabla_{\mathbf{n}}(\mathbf{v}_{l+p} \cdot \mathbf{t}_1) = 0, \quad \nabla_{\mathbf{n}}(\mathbf{v}_{l+p} \cdot \mathbf{t}_2) = 0; \tag{4.34}$$

whereas on ∂M_l, the Dirichlet boundary condition $\mathbf{v}_{l+p} = \mathbf{v}_l$ is equivalent to

$$\mathbf{v}_{l+p} \cdot \mathbf{t}_1 = \mathbf{v}_l \cdot \mathbf{t}_1, \quad \mathbf{v}_{l+p} \cdot \mathbf{t}_2 = \mathbf{v}_l \cdot \mathbf{t}_2, \quad \mathbf{v}_{l+p} \cdot \mathbf{n} = 0. \tag{4.35}$$

In this case, we can enforce $\mathfrak{E}_{l,p}(\ker \delta_l) \subset \ker \delta_{l+p}$. For example, Fig. 4.2 (b) shows the extension of tangential harmonic forms from a torus to a solid sphere where both boundary conditions Eqs. (4.34) and (4.35) are applied. The inputs (thick lines) are only circulations shown in the right chart of Fig. 4.2 (b), while the extended outputs (thin lines) are tangential harmonic forms as well. Therefore, we can construct the following commutative diagram on the de Rham complexes for tangential forms on the filtration of M:

$$
\begin{array}{ccccccc}
\Omega_t^0(M_0) & \xleftarrow{\ \delta^1\ } & \Omega_t^1(M_0) & \xleftarrow{\ \delta^2\ } & \Omega_t^2(M_0) & \xleftarrow{\ \delta^3\ } & \Omega_t^3(M_0) \\
\downarrow{\scriptstyle\mathfrak{E}_{0,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{0,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{0,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{0,1}} \\
\Omega_t^0(M_1) & \xleftarrow{\ \delta^1\ } & \Omega_t^1(M_1) & \xleftarrow{\ \delta^2\ } & \Omega_t^2(M_1) & \xleftarrow{\ \delta^3\ } & \Omega_t^3(M_1) \\
\downarrow{\scriptstyle\mathfrak{E}_{1,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{1,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{1,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{1,1}} \\
\Omega_t^0(M_2) & \xleftarrow{\ \delta^1\ } & \Omega_t^1(M_2) & \xleftarrow{\ \delta^2\ } & \Omega_t^2(M_2) & \xleftarrow{\ \delta^3\ } & \Omega_t^3(M_2) \\
\downarrow{\scriptstyle\mathfrak{E}_{2,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{2,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{2,1}} & & \downarrow{\scriptstyle\mathfrak{E}_{2,1}} \\
\cdots & & \cdots & & \cdots & & \cdots
\end{array}
$$

Similar to the normal form case, through the composition with the simple L_2 projection onto the finite dimensional tangential harmonic space $P_{\mathcal{H}_{\Delta_t,l+p}^k}$, we have a linear map (also a homomorphism) between the tangential harmonic spaces of different manifolds in the filtration, $\Psi_{t,l,p} = P_{\mathcal{H}_{\Delta_t,l+p}^k} \circ \mathfrak{E}_{l,p} : \mathcal{H}_{\Delta_t,l}^k \to \mathcal{H}_{\Delta_t,l+p}^k$. If $h_t \in \mathcal{H}_{\Delta_t,l}^k$ is not in $\mathrm{im}\ \Psi_{t,l-1,1}$, it is said to be born at index l. If p is the smallest integer such that $\Psi_{t,l,p}(h_t) = 0$, it is said to die at index $l + p$, with a persistence of p. This is consistent with the persistence of the (absolute) cohomology $H^k(M)$ and the relative homology $H_{3-k}(M, \partial M)$.

4.2.3.3 Relation among persistent cohomologies under different boundary conditions

As discussed in section 4.1.3, with the duality through Hodge star, there are only three independent singular spectra T, N and C for the three differential/codifferential operators (two for gradient operators under tangential or normal conditions, and one curl operator with either tangential or normal boundary condition). The unions of these spectra produce all the eigenvalues of the eight possible Hodge Laplacians on an arbitrary compact manifold M embedded in a flat 3D space. Moreover, the intersections of spaces spanned by left or right singular vectors of singular value

0 for these operators form the tangential and normal harmonic spaces. Thus, we can restrict our discussion to either normal or tangential fields without loss of generality.

We now discuss the persistence from the perspective of evolving Hodge Laplacian operators. Note that the following discussion is to provide theoretical backgrounds for our proposed use of the evolution of eigenvalues, but not for implementations, since some of the operators discussed may not be sparse matrices when discretized. Recall that for any two manifolds M_l and M_{l+p} in any type of filtration, there is an inclusion map $\mathfrak{I}_{l,p} : M_l \hookrightarrow M_{l+p}$. We call M_{l+p} the p-*evolution manifold* of M_l. We can directly investigate whether a harmonic form in M_l survived in its p-evolution manifold, by defining a restricted subset $\tilde{\Omega}_p^k(M_l)$ of $\Omega^k(M_{l+p})$ and using it to define modified differential and codifferential operators on M_l. This restricted subset is given by

$$\tilde{\Omega}_p^k(M_l) = \{\omega \in \Omega^k(M_{l+p}) | d_{l+p}^k \omega \in \mathfrak{E}_{l,p}(\ker d_l^{k+1})\}. \tag{4.36}$$

This space can be equipped with a modified operator $\tilde{d}_{l+p}^k$ that maps it to $\Omega^{k+1}(M_l)$, which is defined as the compound of d_{l+p}^k followed by the pullback through the inclusion, i.e., $\tilde{d}_{l+p}^k = \mathfrak{I}_{l,p}^* \circ d_{l+p}^k$. Assuming that we use normal differential forms, we have $d_l^{k+1} \tilde{d}_{l+p}^k = 0$ on $\tilde{\Omega}_p^k(M_l)$ as a result of the definition of the restricted space. For $\omega \in \Omega^{k-1}(M_l)$, we have $d_{l+p}^{k-1}\mathfrak{E}_{l,p}(\omega) = \mathfrak{E}_{l,p}(d_{l+p}^{k-1}\omega) \in \mathfrak{E}_{l,p}(\ker d_l^k)$, thus $\mathfrak{E}_{l,p}(\Omega^{k-1}(M_l)) \subseteq \tilde{\Omega}_p^{k-1}(M_l)$ for $p \geq 0$. Therefore, we can construct the following the p-evolution differential form diagram

where $\tilde{\delta}_{l+p}^k$ denotes the adjoint operator of $\tilde{d}_{l+p}^k$. Based on this diagram, the p-evolution Hodge Laplacian $\Delta_{l,p}^k : \Omega^k(M_l) \to \Omega^k(M_l)$ can be defined on M_l as

$$\Delta_{l,p}^k = \delta_l^{k+1} d_l^k + \tilde{d}_{l+p}^{k-1} \tilde{\delta}_{l+p}^k, \tag{4.37}$$

which leads to the definition of the p-evolution harmonic space as $\mathcal{H}_{l,p}^k = \ker \Delta_{l,p}^k = \ker d_l^k \cap \ker \tilde{\delta}_{l+p}^k$. The p-evolution (tangential) k-form spectra are the sets of $\Delta_{l,p}^k$'s eigenvalues for $k =$
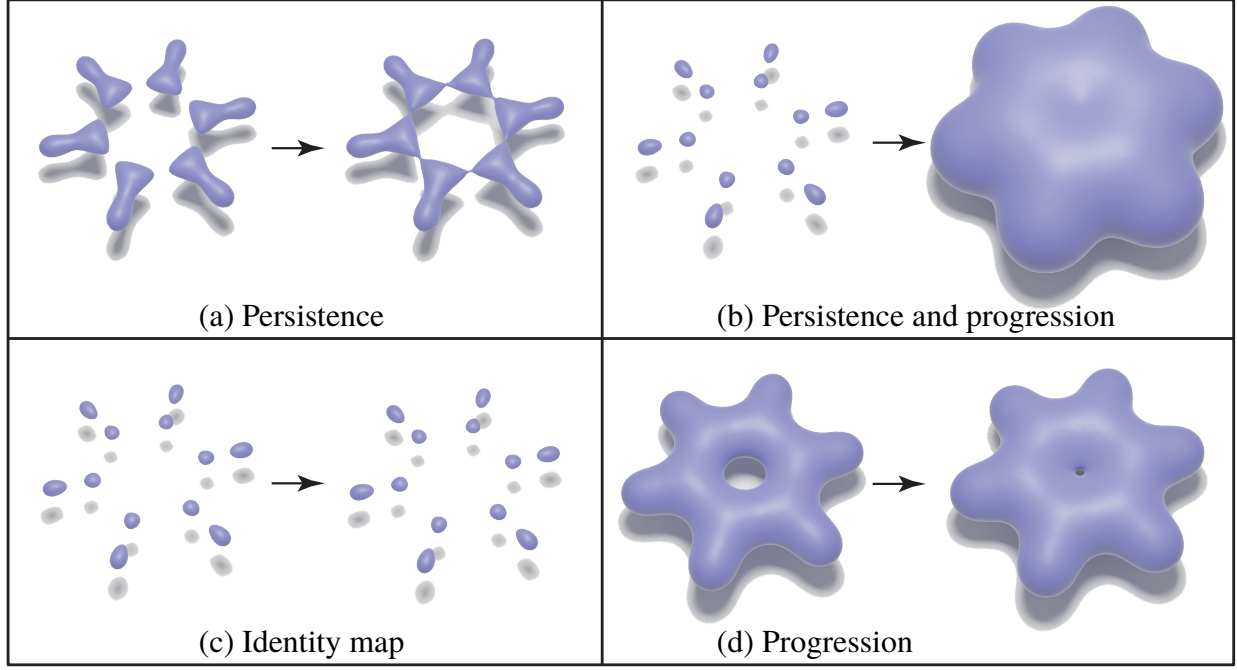
Figure 4.3: **Persistence and progression on benzene.**

$0, 1, 2, 3$. By comparing the p-evolution Laplace operator $\Delta_{l,p}^k$ and the Laplace operator $\Delta_{l,0}^k$, the eigenvalues of the unmodified part, $\delta_l^{k+1} d_l^k$, are preserved, and the eigenvalues involving the pull-back of the restricted operators are varying with p. Next, we examine the part involving $\tilde{d}_{l+p}^{k-1} \tilde{\delta}_{l+p}^k$. For any $\alpha \in \ker \tilde{\delta}_{l+p}^k$, and any $\tilde{\beta} \in \tilde{\Omega}_p^{k-1}(M_l)$, we have $0 = \langle \tilde{\delta}_{l+p}^k \alpha, \tilde{\beta} \rangle = \langle \alpha, \tilde{d}_{l+p}^{k-1} \tilde{\beta} \rangle$. For any $\beta \in \Omega^{k-1}(M_l)$, we have $\langle \delta_l^k \alpha, \beta \rangle = \langle \alpha, d_l^{k-1} \beta \rangle = \langle \alpha, \tilde{d}_{l+p}^{k-1} \mathfrak{E}_{l,p}(\beta) \rangle = 0$. Therefore, $\ker \tilde{\delta}_{l+p}^k \subset \ker \delta_l^k \subset \Omega^k(M_l)$.

Thus, in terms of persistent cohomology, we may examine the kernel of p-evolution Laplace operator for the persistence of topological features of M_l in M_{l+p}. In the perspective of spectral analysis, this change is reflected in the multiplicity of the eigenvalue 0, which changes if $\dim(\ker \tilde{\delta}_l^k) < \dim(\ker \delta_l^k)$, or remains unchanged when $\dim(\ker \tilde{\delta}_l^k) = \dim(\ker \delta_l^k)$. In the former case, as shown in Fig. 4.3 (a), multiplicity of 0 (the number of connected components) is reduced for $\Delta_{l,p}^0$, whereas $\Delta_{l,p}^1$ has a new 0 (a tunnel) that is not present in $\Delta_{l,p}^1$. For the latter case, the inclusion map is homotopic to a geometrical deformation of the manifold, which implies the same topology. Fig. 4.3 (d) illustrate an example where the size of tunnel shrinks, and the cohomology

groups are isomorphic.

The spectra are continuous when corresponding manifolds are continuously deforming, since, as discussed above, when the level set values are close, the deformation is close to an isometric, and the eigenvalues of Hodge Laplacian is determined by the metric tensor. In particular, the smallest non-zero eigenvalues are continuous if the dimension of null space is stable, but are typically non-differentiable when the multiplicity of eigenvalue 0 is changed. The birth of non-zero eigenvalues is the death of topological features, which signals the death of harmonic basis fields; whereas the birth of zero eigenvalues indicates the birth of topological features. Moreover, the changes in leading smallest non-zero eigenvalues can thus indicate possible pending topological changes as well as the geometric properties when the manifold evolves without topological changes.

For instance, for the l-th manifold of the filtration of M, $\{\lambda_{l,i}^T\}$, $\{\lambda_{l,i}^C\}$ and $\{\lambda_{l,i}^N\}$ give the eigenvalues of the T, C and N sets respectively. In particular, the multiplicities of the zero eigenvalues in $\lambda_{l,0}^T$, $\lambda_{l,0}^C$, and $\lambda_{l,0}^N$ are associated with Betti numbers β_0, β_1 and β_2, respectively. Additionally, $\lambda_{l,1}^T$, $\lambda_{l,1}^C$, and $\lambda_{l,1}^N$ are the first non-zero eigenvalues, which are known as the Fiedler values in graph theory, an indicator of how well the graph is connected.

In summary, the correspondence established by the spectral analysis provides us with tools to investigate both types of manifold evolution, with persistence for topological features and spectral progression for the geometric properties.

4.3   Evolutionary de Rham-Hodge analysis of geometric shapes

In this section, we present the application of the proposed evolutionary de Rham-Hodge method. We demonstrate the spectral analysis with evolutionary de Rham Laplace operators and illustrate their topological persistence and geometric progression associated with submanifolds in $\mathbb{R}^3$. The evolving manifolds in our studies are generated by applying Eq. (4.31) to point cloud datasets with a varying level set c, with a fixed scaling parameter η.

For clarity, the first three examples are simple point sets consisting of few points. The two-body set has the location coordinates in $\{(-1.5, 0, 0), (1.5, 0, 0)\}$, and for the four-body and eight-body
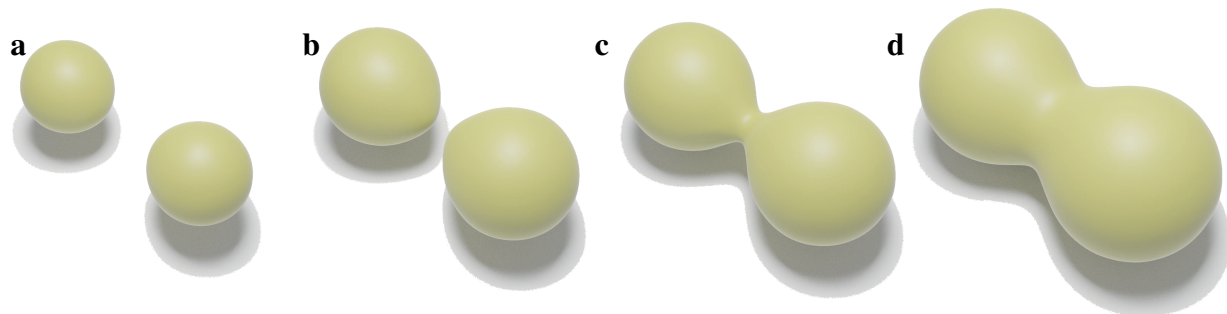
Figure 4.4: **Snapshots of evolving manifold with the two-body system. a**, **b**, **c** and **d** are snapshots from the beginning to the end. **b** and **c** show the transition of the Betti-0 number from 2 to 1.

sets. We duplicate the two-body set by translating ± 1.5 along the y-axis, and duplicate the four-body set by translating ± 1.5 along the z-axis respectively. Next, we present two concrete molecular examples with interesting topological and geometric features, benzene (C_6H_6) and fullerene (C_{60}). Lastly, we illustrate a cry-EM data (EMD-1776) which has interesting properties. We show in these proof-of-concept examples that the evolution of leading smallest eigenvalues provides additional information to that of the persistent Betti numbers, which are the same as those of persistent homology analysis. That is, we propose to extend the evaluation of the manifold evolution from persistent Betti numbers (i.e., the multiplicity of the zero eigenvalues of evolutionary de Rham Laplace operators) to a larger subset of the spectra.

4.3.1 Two-body system

Our first example illustrates the evolving manifold with a two-body system, in which the initial two connected components merge into one. In this evolution, only the number of components persistent β_0 changes from 2 to 1, with the other Bettie numbers remain at 0 throughout. As shown in Fig. 4.4, the two connected components gradually approach each other as the isovalue grows and eventually touch each other as more volume is enclosed.

The change in topology can be observed directly from the blue circle plots in Fig. 4.4, where persistent β_0 is dropped from 2 to 1 when c increased to around 0.6, and the curves for persistent β_1 and β_2 remained flat due to the lack of tunnels or cavities in the system. However, the persistent
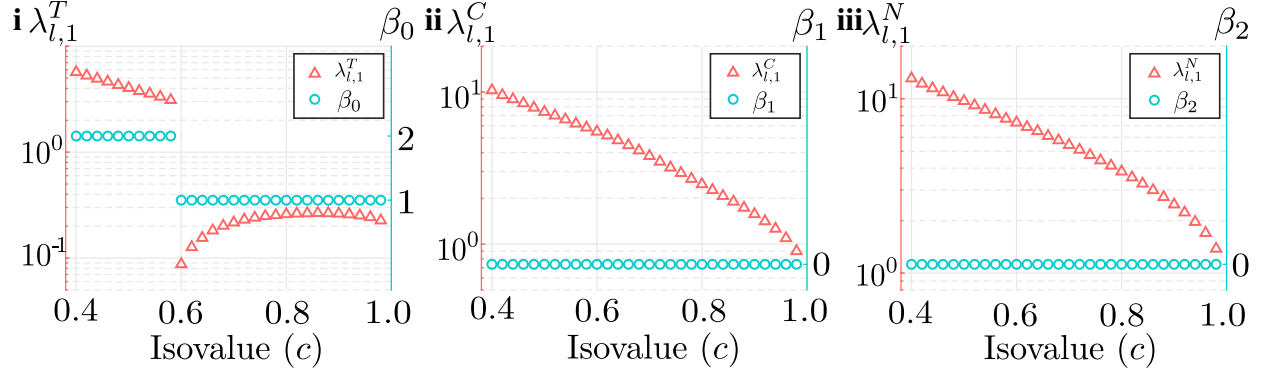
Figure 4.5: **Statistics for two-body system.** Eigenvalues and Betti numbers vs isovalue (c) of the two-body system with $\eta = 1.19$ and $\max(\rho) \approx 1.0$. **i** shows the smallest eigenvalues of the T set. The drops at $c = 0.6$ correspond to snapshots in Figs. 4.4 **b** and **c**. **ii** and **iii** show the smallest eigenvalues of the C and N sets respectively.

Betti numbers do not provide any information about the volume increase of the manifold during the evolution, or the increase in the size of the tube-like structure between the two blobs around the body centers after they touch. In contrast, the orange triangles in Fig. 4.5 show how the first nonzero eigenvalues (Fiedler values) in the three singular spectra (T, C and N) demonstrated both the topological transition and geometric progression in the evolving manifold.

First, one may observe that the discontinuity for the Fiedler values of the tangential gradient fields T coincides with the jump of persistent β_0 in Fig. 4.5 **i**, whereas the Fiedler values of the tangential/normal curl fields C and that of the normal gradient fields N are both smooth as shown in Figs. 4.5 **ii** and **iii**. These behaviors are consistent with the evolution process only having changes in the number of connected components. More precisely, the multiplicity of the eigenvalue zero in T is $\beta_0 = 2$ at the beginning, so the Fiedler values can be seen as the third eigenvalue, whereas after the merging, it is switched to be the second eigenvalue, which contributes to the discontinuity in its value. As we will see in later examples, this behavior for the persistence to be directly observable in the discontinuity of Fiedler values happening at the same isovalue when the Betti numbers jump to different integers is generic, which indicates that the birth of non-zero eigenvalue and the death of the harmonic basis are both linked to the death of topological features (homology generators). Moreover, as the tube between the two blobs is created, the extreme values of the first oscillation mode can be placed further apart along the line connecting the two atoms. Thus, $\lambda_{l,1}^T$ jumps to a
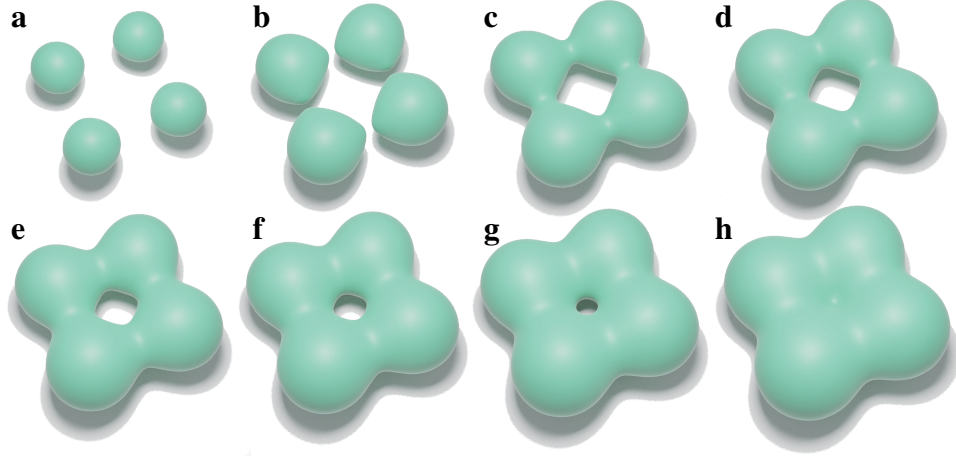
Figure 4.6: **Snapshots of evolving manifolds with the four-body system. a** is the initial point of four components; **b** and **c** show the transition of a ring formed and the persistent Betti-0 number changes from 4 to 1. **g** and **h** show the vanishing of the ring and the persistent Betti-1 number changes from 1 to 0.
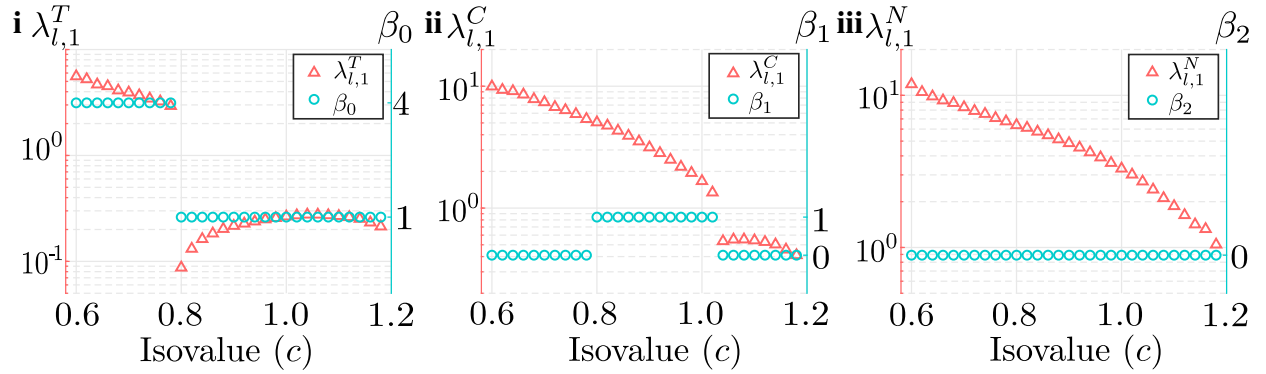
Figure 4.7: **Statistics for four-body system.** Eigenvalues and Betti numbers vs isovalue (c) of the four-body system with $\eta = 1.19$ and $\max(\rho) \approx 1.2$. **i** shows the smallest eigenvalues of the T set. At near $c = 0.80$, the persistent Betti-0 number changes from 4 to 1. **ii** shows the smallest eigenvalues of the C set. At around $c = 1.02$, the persistent Betti-1 number changes from 1 to 0. **iii** shows the smallest eigenvalues of the N set.

small value. It grows as the structure becomes stiffer when the narrow tube turns thicker before it eventually decays again as the entire shape turns softer as a ball with a growing radius. Figs. 4.5 **ii** and **iii** show the smoothness of $\lambda_{l,1}^{C}$ and $\lambda_{l,1}^{N}$ which is consistent with the invariant 1st and 2nd Betti numbers.

4.3.2 Four-body system

As another example, we explore an evolution that involves changes in both the number of components persistent β_0 and the number of tunnels β_1. With two points added to the two-body set to form a planar square, the evolving manifold can contain a tunnel for a range of isovalues, when each of the four components touches two neighbors to form a ring, which will eventually disappear as the level set value increases to the point that the tunnel in the middle is filled. During the same process, persistent β_0 drops from four to one when persistent β_1 increased to one with the formation of the tunnel, but persistent β_0 stays at 1 when persistent β_1 changes back to zero with the disappearance of the tunnel. The persistent Betti number β_2 remains unchanged as there is no cavity in the system.

In terms of the geometric measurements, the total volume continuously increases, and once the tunnel appears, the size of the handle dual to the tunnel also increases. Finally, at the time of disappearing of the tunnel, two concave surfaces are formed on each side of the blocked tunnel with the concavity decreases with an increasing level set parameter.

Fig. 4.7 shows all the Fiedler values varying over time, along with the relevant Betti numbers. As both β_0 and β_1 change during the evolution, $\lambda_{l,1}^T$ and $\lambda_{l,1}^C$ are non-differentiable for this example. On the other hand, β_2 is invariant and thus $\lambda_{l,1}^N$ is smooth. Fig. 4.7 **i** exhibits a similar pattern as the two-body case of $\lambda_{l,1}^T$. As the volume of the manifold increases, $\lambda_{l,1}^T$ decays until the four components are connected, at which point $\lambda_{l,1}^T$ drops to a much smaller value. After the discontinuity, the increasing handle size leads to an initial growth of $\lambda_{l,1}^T$ due to the increased stiffness of the system, before returning to the decreasing trend as the system becomes more flexible with the increase in the overall volume. In Fig. 4.7 **ii**, one may observe the difference compared with the first case as we introduce the changes in persistent β_1. When β_1 changes from zero to one through the connection of the four components, $\lambda_{l,1}^C$ does not actually change much, because the tangential/normal curl field is not largely influenced when the handle size is nearly zero. In stark contrast, $\lambda_{l,1}^C$ is discontinuous when β_1 changes back down to zero as the hole disappears. The behavior of $\lambda_{l,1}^C$ after the discontinuity is similar to that of $\lambda_{l,1}^T$, an initial increase in stiffness and then a decrease again. Moreover, by comparing Figs. 4.7 **i** and **ii**, we observe that the value of $\lambda_{l,1}^T$ starts to decrease just when $\lambda_{l,1}^C$
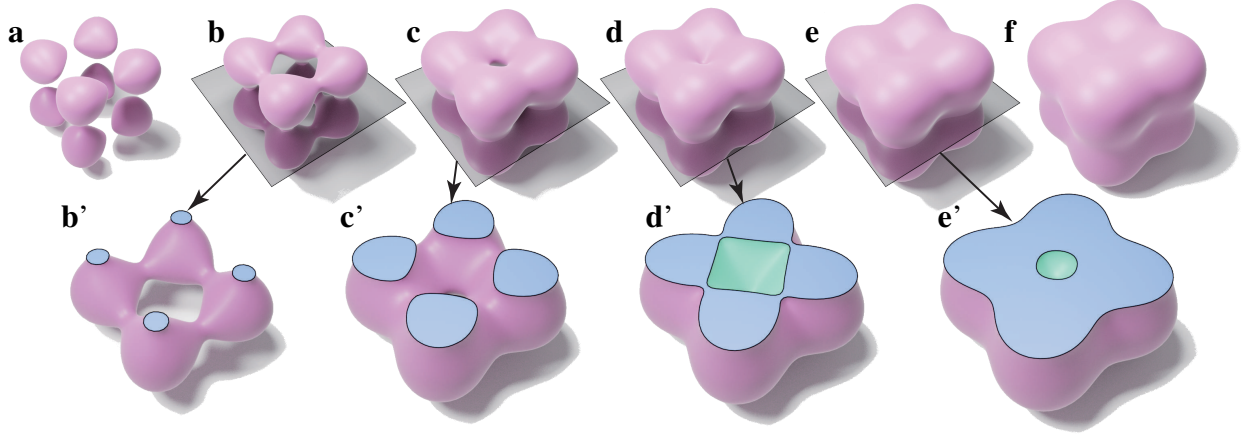
Figure 4.8: **Snapshots of evolving manifold with the eight-body system. a** presents the initial state with eight components. **b** and **c** show the formation of 6 tunnels when the persistent Betti-0 number changes from 8 to 1, and the persistent Betti-1 number changes from 0 to 5. **d** and **e** illustrate that a cavity appears, so the persistent Betti-1 number drops to 0 and the persistent Betti-2 number increases to 1. **f** shows a solid volume without cavity. The gray planes cut manifolds to create cross-section views to illustrate the process of the formation of cavity as shown in **b'**, **c'**, **d'** and **e'**.

is discontinuous, as the structural change in the tunnel also contributed to the "stiffness" of the tangential gradients. Finally, Fig. 4.7 **iii** shows the smooth Fiedler values $\lambda_{l,1}^N$ with an unchanged persistent β_2.

In summary, from the second example, one can notice that $\lambda_{l,1}^C$ can reveal the information of persistent β_1 and some geometric properties after the disappearance of the hole. In addition, the coincidental topological changes, the birth of hole that coincides with the death of a few connected components, can be distinguished by the spectral functions $\lambda_{l,1}^T$ and $\lambda_{l,1}^C$.

4.3.3 Eight-body system

[t]

We constructed the simple eight-body system to analyze the behavior of Hodge Laplacian spectra with an evolving cavity in the filtration. In this system, not only multiple connected components and multiple tunnels are involved, but a cavity also appears after the isovalue reaches a certain level before disappearing eventually. Thus, the dimension-2 Betti number β_2, which measures the

114
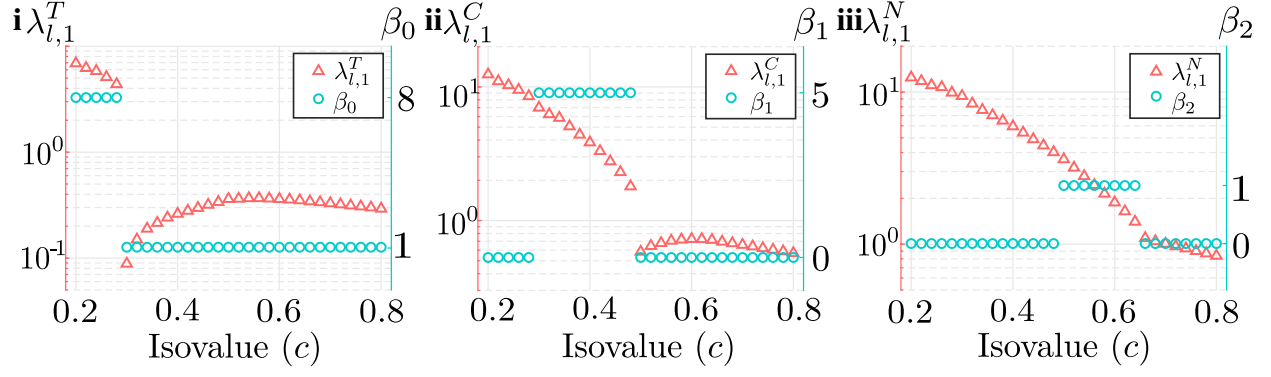
Figure 4.9: **Statistics for eight-body system.** Eigenvalues and Betti numbers vs isovalue (c) of the eight-body system with $\eta = 1.53$ and $\max(\rho) \approx 1.1$. **i** shows the Fiedler values of the T set and persistent Betti-0 numbers. **ii** shows the Fiedler values of the C set and persistent Betti-1 numbers. **iii** illustrates the comparison of $\lambda_{l,1}^{C}$ and persistent β_2.

number of cavities, changes during this process.

As shown in Fig. 4.8, the eight symmetric components start as blobs around eight vertices of a cube. Then they expand as the isovalue increases until they touch each other and form 6 rings, one for each face of the cube. At this point, persistent β_0 drops from 8 to 1, when persistent β_1 increases from 0 to 5 (as five of the six tunnels are independent homology generators). As the level set value increases to the point that the tunnels are filled, persistent β_1 drops back to 0, but persistent β_2 increases to 1 as a cavity formed inside the manifold. The cavity is filled up eventually, and persistent β_2 drops back to 0.

In Fig. 4.9, the Fiedler values as functions of isovalue are shown in Figs. 4.9 **i** and **ii**, which exhibit similar behaviors as in the first two examples. As in the previous example, the comparison between Figs. 4.9 **i** and **ii** shows that at $c = 0.3$ the spectral function $\lambda_{l,1}^{T}$ starts to decay when $\lambda_{l,1}^{C}$ is discontinuous. Different from the previous examples, the smallest eigenvalues in **iii** is no longer differentiable as persistent β_2 changes from one to zero near isovalue 0.5. Fig. 4.9 **iii** also indicates that at the isovalue where $\lambda_{l,1}^{N}$ is non-differentiable, $\lambda_{l,1}^{C}$ starts to decrease. Moreover, the simultaneous topological changes, the disappearance of tunnels and the appearance of the cavity, can be observed in $\lambda_{l,1}^{C}$. The disappearance of the cavity can be observed from $\lambda_{l,1}^{N}$. From these preliminary results of the evolutionary de Rham-Hodge method, one may observe that the singular values in different spectra taken as functions of the isovalue c not only illustrate the changes of
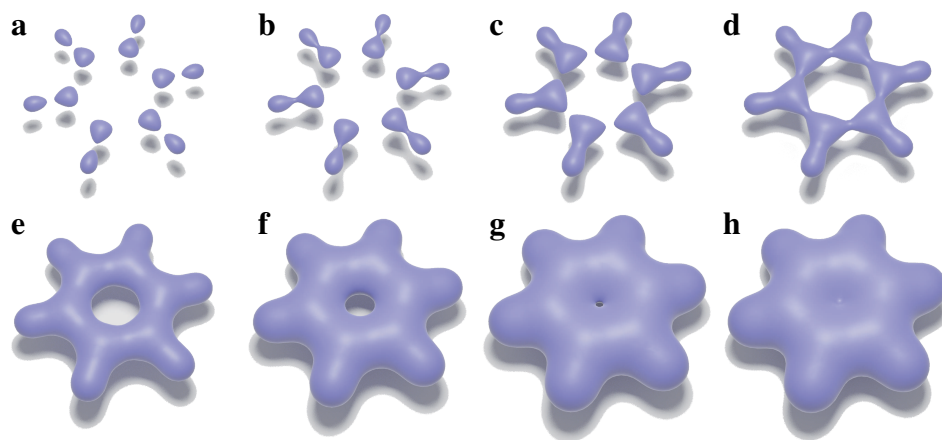
Figure 4.10: **Manifold evolution of benzene.** Manifold evolution of benzene with $\eta = 0.45 \times r_{\text{vdw}}$. **a** through **h** are snapshots from the start to the end. **a** and **b** show the transition of the persistent Betti-0 number from 12 to 6. **c** and **d** show the formation of a ring; The Betti-0 number changes from 6 to 1 and remains at one to the end, whereas the Betti-1 number changes from zero to one. **d**, **e**, **f** and **g** illustrate the deformation of the hexagonal tunnel to a round tunnel. From **g** to **h**, the ring disappears and the Betti-1 number changes from 1 back to 0.

topological features of different dimensions throughout the evolution of the manifold but also reveal the geometric features in different dimensions. Therefore, empirically, the importance of low frequencies rather than the multiplicity of the zeroth frequency can already be observed in these simplistic constructions for features of different dimensionality. In the following, we demonstrate similar characteristics of spectral functions in two molecular systems.

4.3.4 Benzene molecule

Benzene (C_6H_6) is a small organic chemical compound which consists of six carbon atoms in a planar hexagon ring and six hydrogen atoms each connected with one carbon atom. In this system, atoms have different van der Waals radii, one for carbon and another for hydrogen. The carbon atoms are closer to each other than the hydrogen atoms and form the benzene ring. Thus, benzene is a perfectly simple yet realistic example to illustrate the evolutionary de Rahm-Hodge method. With the benzene data, we use $\eta = 0.45$ to generate evolving manifolds.

The first evolving manifold of benzene is generated at $\eta = 0.45$. In the beginning, there are 12 components, with each smooth component center around one atom location as shown in Fig. 4.10 **a**.
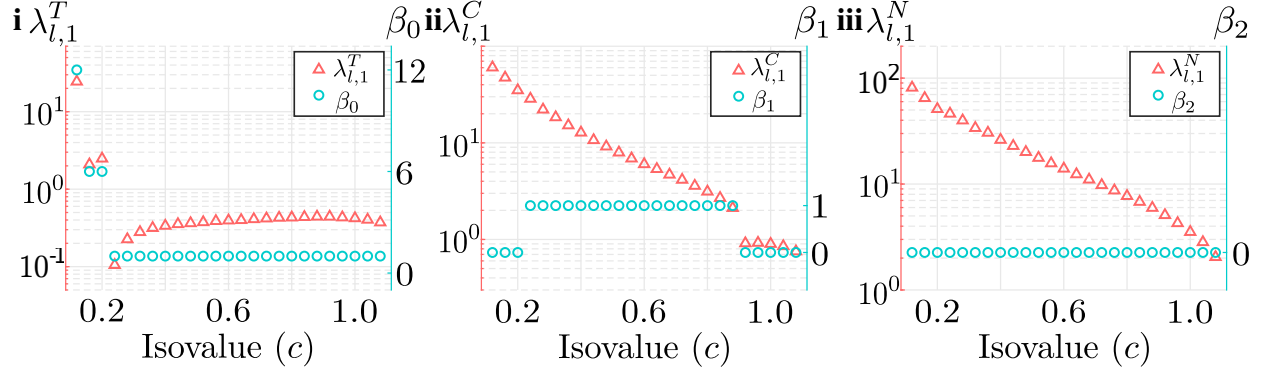
Figure 4.11: **Statistics for benzene.** Eigenvalues and Betti numbers vs isovalue (c) of the benzene system with $\eta = 0.45$ and $\max(\rho) \approx 1.1$. **i** shows the smallest eigenvalues of the T set. The drops at $c = 0.12$ correspond to snapshots in Figs. 4.10 **a** and **b**. The drops at $c = 0.22$ correspond to snapshots in Figs. 4.10 **c** and **d**. **ii** shows the smallest eigenvalue of the N set. The drops at $c = 0.9$ correspond to snapshots in Figs. 4.10 **g** and **h**. **iii** shows the smallest eigenvalues of the C set.

The van der Waals radius of carbon atoms is larger than that of hydrogen atoms, so the components associated with the carbon atoms are larger. From Fig. 4.10 **b** to Fig. 4.10 **c**, the originally separated components of the atoms start to connect pairwise, with a narrow tube formed between each hydrogen to its bonded carbon and thus, the persistent Betti-0 number is reduced to 6. The behavior of the manifold is similar to essentially six copies of our first example, the two-body system, until the six components of Fig. 4.10 **c** start to form a hexagonal ring, as shown in Fig. 4.10 **d**. At this point, there are six narrow tubes, one for each bond between two adjacent carbon atom pairs. As the density function continues to expand, the hexagonal ring evolves into a round cycle around a tunnel with a shrinking diameter. As the diameter of the tunnel reduces to zero at some parameter value between those of Fig. 4.10 **g** and Fig. 4.10 **h**, the noncontractible cycle disappears. During this topological change, the tiny cycle in the middle of the manifold in Fig. 4.10 **g** is filled up to form two concave surface patches in the middle of the manifold in Fig. 4.10 **h**. The final topology of this system remains as a single component with a volume larger than that of Fig. 4.10 **h**.

Fig. 4.11 shows the Fiedler values of the T, N and C sets and their relations with the persistent Betti numbers when seen as a function of varying isovalues. First, for the T set, $\lambda_{l,1}^T$ has two jumps at $c = 0.12$ and $c = 0.22$, which divide the $\lambda_{l,1}^T$ to three curve segments. Both discontinuities correspond to the decreases of the persistent Betti 0, from twelve to six, and then to one. As shown

in Figs. 4.11 **i**, $\lambda_{l,1}^T$ cannot only tell the topological changes but also give some additional information of a continuous portion of the evolution. After $c = 0.22$, $\lambda_{l,1}^T$ increases first and reaches its maximum at $c = 0.9$ when the ring just disappears, at which point the structure (for tangential gradients) starts to grow softer as an expanding blob instead of a thicker ring. Fig. 4.11 **ii** presents the jump of $\lambda_{l,1}^C$, which is correlated to the disappearance of the hole as indicated by the change of Betti-1 number from one to zero. After the jump, $\lambda_{l,1}^C$ also increases slightly first and decays in the end. There is no cavity involved, so the spectral function shows a steady progression for the C set as in our four-body example. One difference from that example is the finer grid used in the calculation, in order to handle the initial small components for the hydrogen atoms.

4.3.5   Buckminsterfullerene

The buckyball (C_{60}) has a beautiful structure composed of sixty carbon atoms. It has twenty hexagons and twelve pentagons that resemble the pattern on a soccer ball, which has a rich structure with both geometric symmetries and topology features. With our continuous density function, at certain values of η, the manifold evolution covers all the possible values of the persistent Betti-1 number allowed by the symmetry. However, it is difficult to cover all the topological space for a density function associated with a single kernel size η. Thus we propose to use a multiscale (with a few different kernel sizes) analysis of the manifold evolution. By using different η's to capture different sets of snapshots for the evolving manifolds, we can compare the spectra across different kernel sizes η as well as different control parameters c. We use the buckyball as an example for the multiscale analysis of manifold evolution, and demonstrate how the spectra provide information on the evolution of their topological spaces and geometric features.

For kernel scaling parameter $\eta = 0.5 \times r_{\text{vdw}}$, the manifold evolution starts with 60 components as shown in Fig. 4.12 **a**. The components start the expansion, each around the position of one carbon atom, and merge into larger connected components if they share a common pentagon in the skeleton structure as shown in Fig. 4.12 **b**. This leads to the changes in persistent β_0 (from 60 to 12) and persistent β_1 (from 0 to 12). Fig. 4.12 **c** shows the snapshot right after the appearance of

Figure 4.12: **Manifold evolution of fullerene.** Illustration of fullerene (C_{60}) manifold evolution with $\eta = 0.5 \times r_{vdw}$. **a** presents sixty components around carbon atom positions. **a** and **b** show that the components connect if they share a pentagonal hole, and persistent β_0 changes from 60 to 12 and persistent β_1 changes from 0 to 12. **c** shows the hexagonal holes are formed, resulting in the change of persistent β_0 to 1 and persistent β_1 to 31. (There are 32 rings, but only 31 are independent in terms of homology.) **c** and **d** show that the 12 pentagonal rings disappear and the persistent Betti-1 number drops from 31 to 19. **d** and **e** show that the 20 hexagonal rings disappear and a cavity forms inside, so that persistent β_1 drops to 0 and persistent β_2 increases to 1. The vertical plan cuts the manifolds that gives an illustration of cavity in **d'** and **e'**.

twenty hexagonal holes. Next, each hole starts to shrink. As each pentagonal hole has a smaller size than that of a hexagonal hole, we observe in Fig. 4.12 **c** to Fig. 4.12 **d**, the pentagonal holes disappear before the hexagonal holes also disappear. Simultaneous to the disappearance of hexagons, a cavity is created. In Fig. 4.12 **e** after the formation of the cavity, both the outer surface and the inner surface contain numerous regions of concavity and gradually, the shape evolves to resemble a slightly dented thick spherical shell.

For analysis of this evolution, Fig. 4.13 illustrates the eigenvalues and Betti numbers versus the isolvaue c. Fig. 4.13 **i** gives the Fiedler values (smallest eigenvalue) of the T set and β_0. This Betti number has two drops, from 60 to 12, and then to 1. Within each interval of isovalues with the same persistent Betti number, $\lambda_{l,1}^{T}$ is changing smoothly as expected from our discussion on homeomorphic shapes with a slowly evolving metric. Fig. 4.13 **ii** presents the information that the Fiedler values of the C set can offer. For the interval, $c \in [0.16, 0.5]$, persistent β_1 remains

119
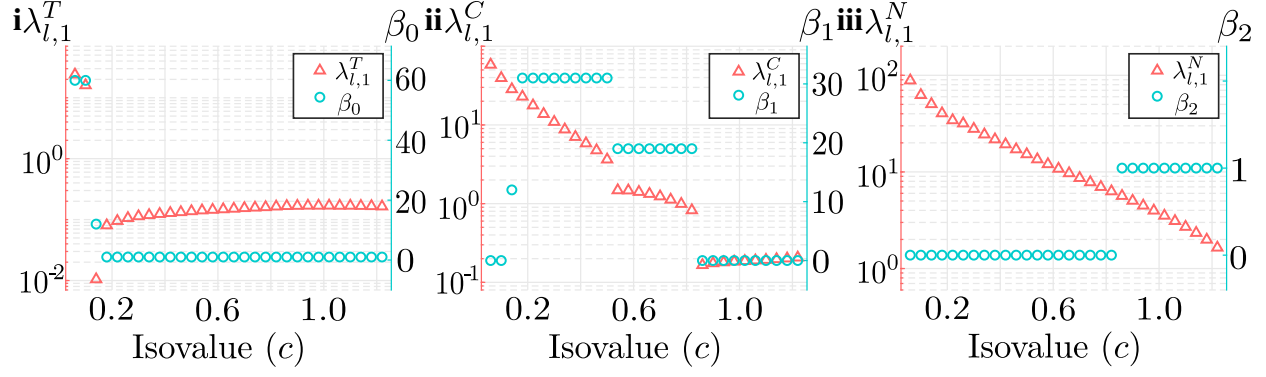
Figure 4.13: **Statistics for fullerene.** Eigenvalues and Betti numbers vs isovalue (c) of the fullerene (C_{60}) system with $\eta = 0.5 \times r_{\text{vdw}}$ and $\max(\rho) \approx 1.3$. **i** gives the Fiedler values of the T set and persistent β_0. **ii** presents the comparison of $\lambda^C_{l,1}$ and persistent β_1. **iii** shows the Fiedler values of the N set and persistent β_2.

at 31, and the continuous decrease in $\lambda^C_{l,1}$ shows that the geometric structure is "softer" for the curl fields as the handles grow thicker. Similarly, for intervals within which persistent β_1 equals to 19 or 1, $\lambda^C_{l,1}$ is a smooth function within each interval but is discontinuous at the boundary of these intervals where the topology transitions. The Fiedler values of the N set are given in Fig. 4.13 **iii**, which, although mostly smooth, also has changed in slope at isovalues associated with changes in connected components and tunnels. As the examples become more complex, the spectral functions also exhibit richer structure, with the advantage of indicating both topological persistence and geometric progression.

For large and dense point sets as in this fullerene, the shape of the manifold evolution is heavily influenced by the kernel size η. To show the importance of multiscale analysis, we create a second evolution with $\eta = 0.8 \times r_{\text{vdw}}$ and generate the snapshots in Fig. 4.14. For the initial isovalue, as seen in Fig. 4.14 **a**, the manifold consists of twelve pentagonal components. Unlike the evolution with $\eta = 0.5 \times r_{\text{vdw}}$, which contains pentagonal holes alongside hexagonal holes, here the pentagonal components are already with the holes filled before the hexagonal holes are even formed. Thus, the two evolutions cannot find a homeomorphism between their stages even if any isovalues are allowed, which implies that they can reveal different information regarding the system. As the components connect, twenty rings show up as in Figs. 4.14 **b** and 4.14 **c**, with decreasing diameters for increasing isovalues. Once the cavity is formed, the large inner surface shown in Fig. 4.14 **d**
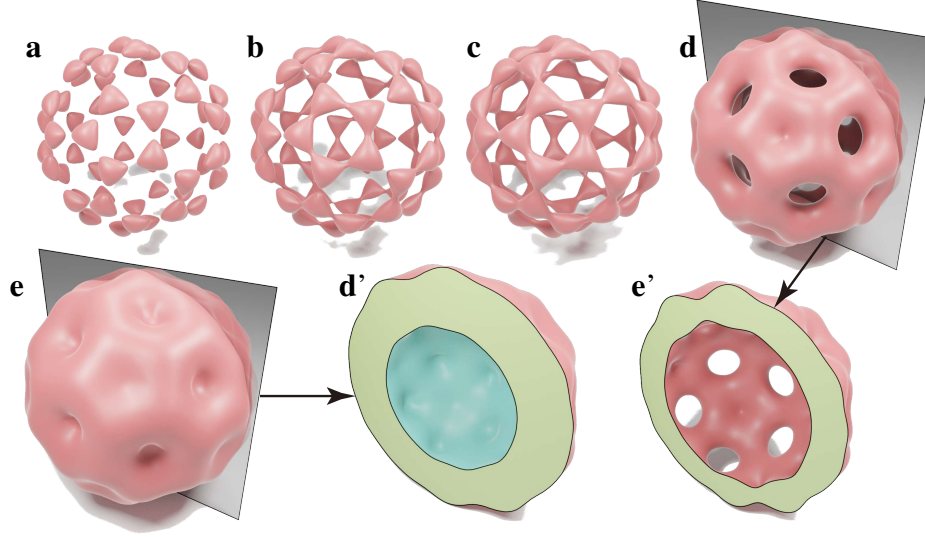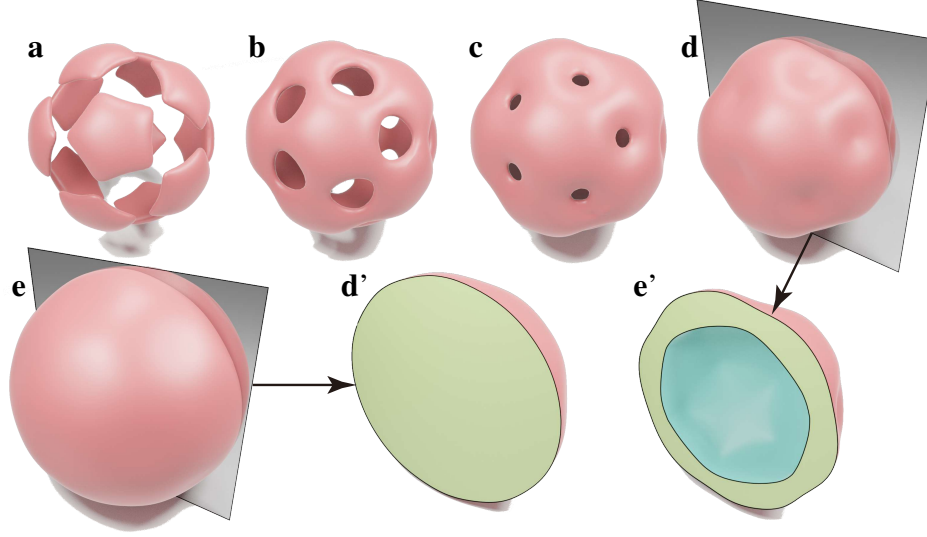
Figure 4.14: **Manifold evolution of fullerene.** Illustration of fullerene (C_{60}) manifold evolution with $\eta = 0.8 \times r_{\mathrm{vdw}}$. **a** shows 12 initial solid pentagonal components. **b** and **c** show the formation and contraction process of the 20 rings. **d** is the snapshot right after the formation of the cavity. **e** shows the final stage as a solid ball of this example.
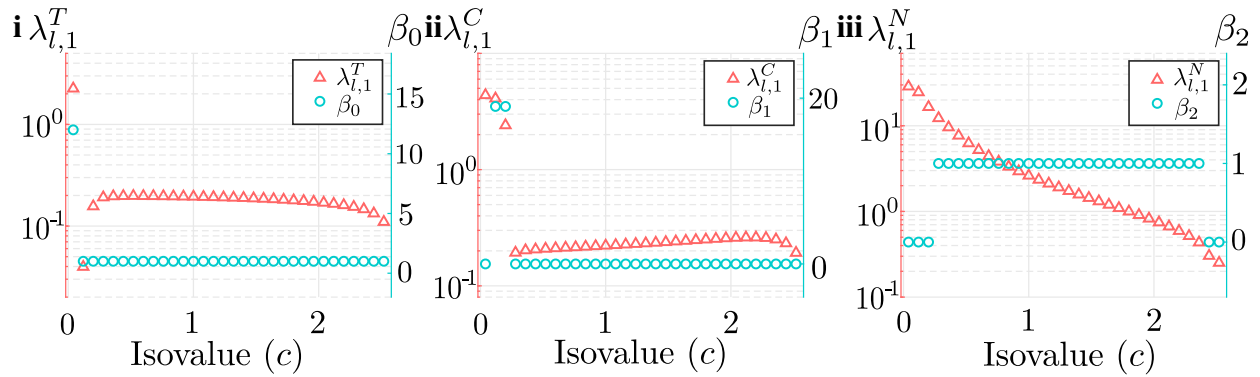
Figure 4.15: **Statistics for fullerene.** Eigenvalues and Betti numbers vs isovalue (c) of the fullerene (C_{60}) system with $\eta = 0.8 \times r_{\mathrm{vdw}}$; $\max \rho \approx 2.5$. **i** gives the Fiedler values of the T set and persistent β_0. **ii** presents the comparison of $\lambda_{l,1}^C$ and persistent β_1. **iii** shows the Fiedler values of the N set and persistent β_2.

starts to contract, and the manifold ends up as a solid ball in Fig. 4.14 **e**. As for the spectral functions, Fig. 4.15 shows three plots of the Fiedler values of the T, C and N sets and the persistent Betti numbers against the isovalues, respectively. Since the components connect right after first two snapshots, Fig. 4.15 **i** shows the drop of $\lambda_{l,1}^T$ in the third snapshot as persistent β_1 changes from 12 to 1. The Fiedler values $\lambda_{l,1}^T$ then increases before starting to decrease when persistent β_1 drops to 0 when the system can be seen as a shell growing softer with thicker membrane instead of a structure

growing stiffer with thicker supporting handles. Similarly, there are only a few snapshots for the evolving manifold to have rings as they are quickly filled up. In Fig. 4.15 **ii**, the Fiedler values $\lambda_{l,1}^C$ already decreases quickly before plunging to a small number at the point when holes disappear. During the period of the inner surface contracting and outer surface expanding, $\lambda_{l,1}^C$ increases first as the structure grows stiffer for curl fields, and then grows softer eventually near the very end of the manifold evolution. In the last plot of Fig. 4.15, $\lambda_{l,1}^N$ slightly increases at beginning and then decreases smoothly. The disappearance of the cavity is captured at the end of snapshots, thus there is a non-differentiable point at end of this spectral function. We see in this evolution again, that the progression of the manifold evolution can be observed in the spectral functions as well as the topological transitions.

4.4   Application

In this section, we present two examples to demonstrate the usefulness of the proposed evolutionary de Rham-Hodge method in biological applications. The first example shows the protein flexibility analysis by applying evolutionary de Rham-Hodge method and the second analyze the cryo-EM density map by using persistent spectra and topology.

4.4.1   Protein flexibility analysis

We apply the proposed evolutionary de Rham-Hodge method to biomolecular flexibility analysis. Protein flexibility is strongly correlated protein functions, such as structural support, catalyzing chemical reactions, and allosteric regulation. It can be measured by many experimental approaches, such as X-ray crystallograph and nuclear magnetic resonance (NMR) in terms of B-factors or Debye-Waller factors. Qualitative prediction of protein B-factors is important for understanding protein structure-function relationship. Many biophysical models, such as Gaussian network model (GNM) [186], anisotropic network model (ANM) [52], and FRI [2] have been developed in the past for such a prediction. Most of these methods are based on the graph network composed by selecting C^α carbon atoms as nodes and connections between nodes as edges. However, existing approaches
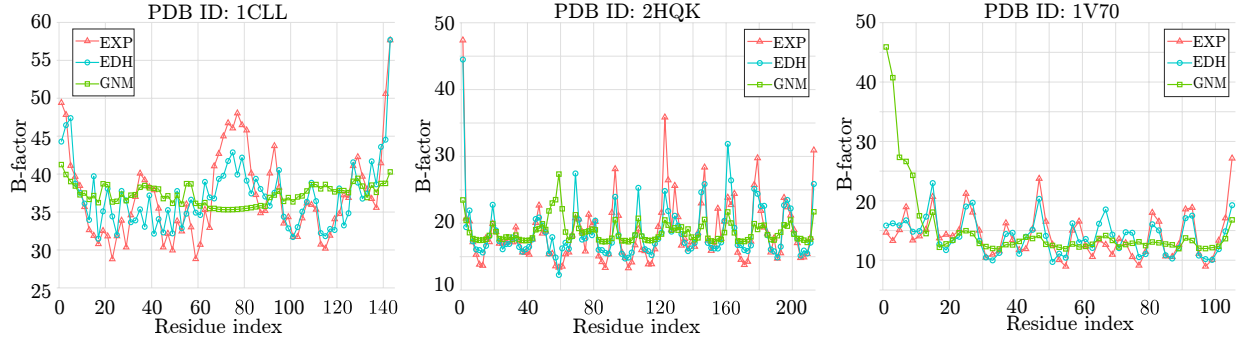
Figure 4.16: **Experimental and predicted B-factor values plotted per residue (PDB IDs: 1CLL, 2HQK and 1V70).** EXP: experimental values; EDH: evolutionary de Rham-Hodge (10 isovalues) method predicted values; GNM: Gaussian network method predicted values.
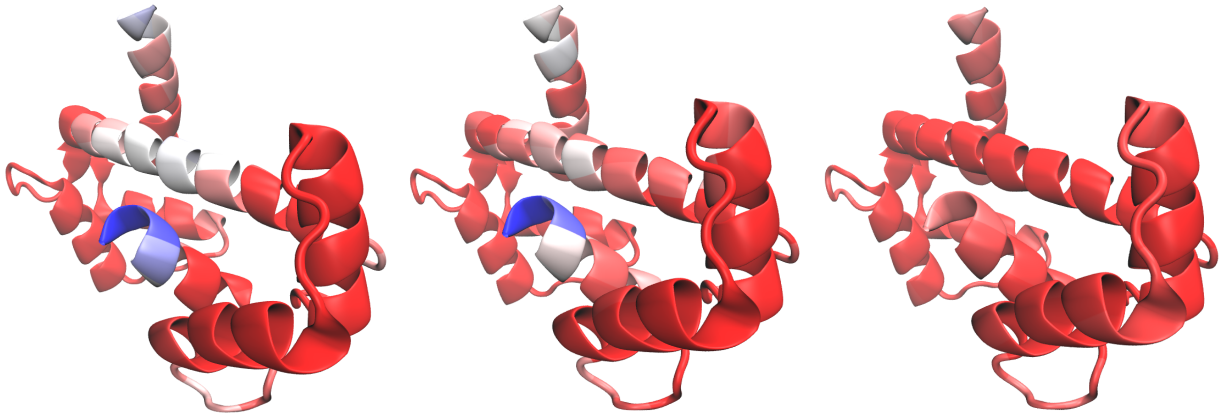
Figure 4.17: **The structure of calmodulin (PDB ID: 1CLL).** The structure of calmodulin (PDB ID: 1CLL) visualized in Visual Molecular Dynamics (VMD) [9] and colored by experimental B-factors (left), EDH (10 isovalues) predict B-factors (middle), and GNM predicted B-factors (right) with red representing the most flexible regions.

encounter much challenging for many macromolecules involving multiscale interactions. In the present study, we consider a few challenging test cases to demonstrate the utility and performance of the proposed evolutionary de Rham-Hodge method.

The evolutionary de Rham-Hodge method evaluates a manifold generated by Eq. (4.30) based on C^α carbon atoms and the B-factor at the i-th atom estimated by $\bar{L}_k$ in Eq (4.26) is given by

$$\mathrm{B}_{k,i}^{\mathrm{EDH}} = \sum_l a_l \sum_j \frac{1}{\lambda_{l,j}^k} \omega_{l,j}^k(\mathbf{r}_i)(\omega_{l,j}^k(\mathbf{r}_i))^T, \quad \forall \lambda_{l,j}^k > 0, \tag{4.38}$$

where a_l are parameters determined by a primitive machine learning algorithm (i.e., linear regression) for filtration parameter l. In our computation, discrete eigen fields $\omega_{l,j}^k$ are vectors of mesh

PDB ID	$N_{C\alpha}$	GNM[2]	mGNM[2]	EDH (10)	EDH (20)	EDH (40)
1CLL	292	0.261	0.763	0.789	0.797	0.850
1V70	105	0.162	0.750	0.754	0.772	0.858
2HQK	216	0.365	0.833	0.854	0.880	0.886
1WHI	122	0.270	0.484	0.640	0.711	0.794

Table 4.3: **Pearson correlation coefficients in B-factor predictions using GNM, mGNM, and EDH for four proteins.** Here, mGNM stands for multiscale GNM with two different kernels [2]. $N_{C\alpha}$ is the number of residues. In cases of EDH, three different isovalue sets are applied with 10, 20 and 40 points of equal spaces on the interval of $[0.1, 1.0]$.

points. Here, $\omega_{l,j}^{k}(\mathbf{r}_i)$ is computed by the interpolation of a neighborhood around i-th atom with a cutoff radius d. In our test, we use the grid spacing of mesh tetrahedron 1.6 Å, the cutoff radius $d = 4.0$ Å, and $\eta = 2.72$ Å. For a comparison, we consider the standard method, GNM, with its cutoff distance of 7 Å. In Fig. 4.16, predicted B-factors of three proteins (PDB IDs: 1CLL, 2HQK and 1V70) are presented together with their experimental results. In our method, 10 isovalues of equal spaces from 0.1 to 1 are calculated. The B-factors of C^α atoms predicted from the evolutionary de Rham-Hodge (EDH) method are more close to the experimental ones than those from GNM. Especially, Fig. 4.17 shows the flexibility of calmodulin of 1CLL obtained by experiment and theoretical predictions. Clearly, by a comparison with experimental results, EDH predictions are significantly better than those of GNM. Moreover, an advantage of evolutionary de Rham-Hodge method is that one can simply increases the number of isovalues to provide more geometry deformation information and attain better results. As shown in Table 4.3, the increase of the number of snapshots on the same interval delivers better predictions. The proposed EDH method outperforms other existing methods.

4.4.2 Evolutionary de Rham-Hodge analysis of cryo-EM density map

Cryo-electron microscopy (cryo-EM) is a power method for analyzing the structures of biological systems. Cryo-EM density maps are generated by bombarding samples by electron beams at cryogenic temperatures to improve the signal-to-noise ratio (SNR) and constructed from a large number of 2D images using computational methods. The projection (thin film) specimen scans
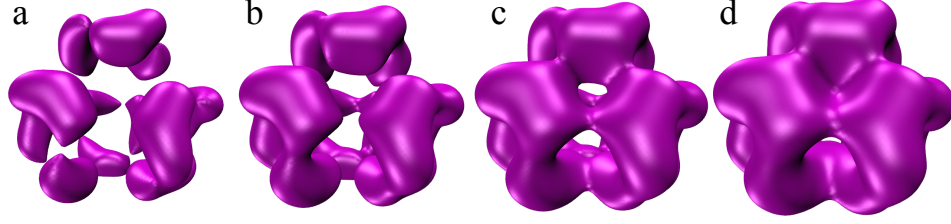
Figure 4.18: **Illustration of surfaces extracted with different isovalues for EMD-1776.** The isovalues for **a**, **b**, **c**, and **d** are 0.14, 0.10, 0.07, and 0.04, respectively. In **a**, β_0 is 12, and β_1 and β_2 are 0; In **b**, $\beta_0 = 4$, $\beta_1 = 4$, and $\beta_2 = 0$; In **c**, $\beta_0 = 1$, $\beta_1 = 13$, and $\beta_2 = 0$; In **d**, $\beta_0 = 1$, $\beta_1 = 9$, and $\beta_2 = 0$.
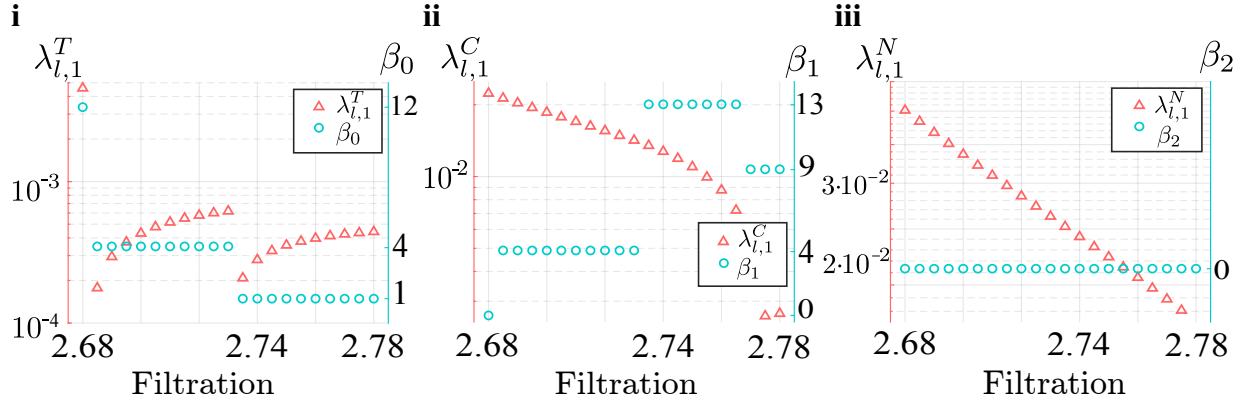
Figure 4.19: **Eigenvalues and Betti numbers vs filtration of the EMD-1776 density map.** The filtration goes from 2.68 (the largest isovalue (0.28) subtract by 0.14) to 2.78 (the largest isovalue (0.28) subtract by 0.04). **i** gives the Fiedler values of the T set and persistent β_0. **ii** presents the comparison of $\lambda_{l,1}^C$ and persistent β_1. **iii** shows the Fiedler values of the N set and persistent β_2.

collected from many different directions comprise the basis of cryo-EM images. A major advantage of cryo-EM is that it provides the image of specimens in a native environment without the need to grow crystals and another advantage is its capability of providing 3D mapping of entire cellular proteomes together with their detailed interactions at nanometer or subnanometer resolution [187, 188, 189]. After illustrating the evolutionary de Rham-Hodge analysis for the FRI density functions of known structures, we further consider a realistic cryo-EM data, EMD-1776, which is for eye lens chaperone alphaB-crystallin forms [190].

Here, we reveal the evolutionary spectra and persistent topology associated with the manifold evolution of EMD-1776 density map. Figure 4.18 depicts the surfaces extracted with different isovalues of EMD-1776. The isovalues for Figures 4.18 **a-b** are 0.14, 0.10, 0.07, and 0.04, respectively. Betti numbers in these Figures are given as $\beta_0 = 12$, $\beta_1 = 0$, and $\beta_2 = 0$ in Figure 4.18 **a**; $\beta_0 = 4$,

$\beta_1 = 4$, and $\beta_2 = 0$ in Figure 4.18 **b**; $\beta_0 = 1$, $\beta_1 = 13$, and $\beta_2 = 0$ in Figure 4.18 **c**; $\beta_0 = 1$, $\beta_1 = 9$, and $\beta_2 = 0$ in Figure 4.18 **d**. In Figure 4.19, the eigenvalues and Betti numbers of each filtration of the EMD-1776 system are presented. Note the filtration is generated by controlling the isovalue of cryo-EM data. The index shown for x-axis is calculated by subtracting the isovalue from the largest isovalue, in which the filtration has an inclusion relation. Similar to aforementioned results, eigenvalues illustrates the persistence of Betti number, but also depicts the geometry shape changing. In Figure 4.19 **i**, it shows that the eigenvalue $\lambda_{l,1}^{T}$ encounters discontinuity when the Betti-0 decreases from 12 to 4 and from 4 to 1. In Figure 4.19 **ii**, the eigenvalue $\lambda_{l,1}^{C}$ is discontinuous when the Betti-1 decreases from 13 to 9. This behavior is consistent with those of our earlier observations.

CHAPTER 5

CONCLUSION

We have presented a systematic method to analyze 3D shapes based on a famous mathematical theory called de Rham - Hodge theory, which could potentially benefit research areas such as computer graphics, computer vision and computational biology.

First, we have detailed the construction of a five-component decomposition of vector fields on triangulated 3D domains for a large variety of typical boundary conditions. Our approach was shown to be consistent with the continuous theory on vector field analysis, and to capture the proper kernel spaces due to nontrivial homology and cohomologies, either from the topology or from the boundary of 3D domains. We showed that the numerical procedure based on discrete exterior calculus leads to straightforward constructions of the desired boundary conditions. We discussed how to assemble all the matrices involved, as well as how to handle their known rank deficiency (based on the topology of the domain) to ensure fast computations. We expect this straightforward numerical tool to benefit computational applications involving volumes in 3D Euclidean space, such as in geometric modeling, electromagnetism, fluid dynamics, elasticity and biomolecular science. However, our decomposition is restricted to domains in $\mathbb{R}^3$ with Euclidean metric. It can however be extended to any 3-manifold that can be embedded in $\mathbb{R}^3$: the orthogonality between $\mathcal{H}_t$ and $\mathcal{H}_n$ only depends on the topology. Moreover, it is possible to extend it to k-forms on any simplicial tessellation of compact n-manifolds with boundaries if we lift the restriction on the orthogonality between those two components, and compute the harmonic vector fields also through eigensolvers instead of our efficient alternatives through potentials designed for 3D domains. As a special case, 0-forms and n-forms on n-manifolds can always be orthogonally decomposed into the divergence of a tangential gradient field plus β_0 constant fields. Exploring spectral analysis of our tangential and normal Laplacian operators is also an interesting direction of research.

Second, this work introduces the de Rham-Hodge theory as a unified paradigm to analyze biomolecular geometry, topology, flexibility and Hodge modes based on three-dimensional (3D)

127

coordinates or cryo-EM maps. Specifically, de Rham-Hodge spectral analysis has been carried out to reveal macromolecular geometric characteristics and topological invariants with normal and tangential boundary conditions. The Helmholtz-Hodge decomposition is employed to analyze the divergence-free, curl-free, and harmonic components of macromolecular vector fields. Based on the 0-form scalar Hodge-Laplacian, an accurate multiscale model is constructed to predict protein fluctuations. By equipping a vector Laplace-de Rham operator with a boundary constraint based on Helfrich-type curvature energy, a 1-form Laplace-de Rham-Helfrich operator is proposed to predict the Hodge modes of biomolecules, particularly cryo-EM maps. In addition to its versatile nature for a wide variety of modeling and analysis, the proposed de Rham-Hodge paradigm also provides a unified approach to handle biomolecular problems at various spatial scales and with different data formats. A state-of-the-art 3D discrete exterior calculus algorithm is developed to facilitate accurate, reliable and topological structure-preserving spectral analysis and modeling of biomolecules. Extensive numerical experiments indicate that the proposed de Rham-Hodge paradigm offers one of the most powerful tools for the modeling and analysis of biological macromolecules. The proposed de Rham-Hodge paradigm provides a solid foundation for a wide variety of other biological and biophysical applications. For example, the present de Rham-Hodge flexibility and Hodge mode analysis can be directly applied to subcellular organelles, such as vesicle, endoplasmic reticulum, Golgi apparatus, cytoskeleton, mitochondrion, vacuole, cytosol, lysosome, and centrosome, for which the existing atomistic biophysical approaches have very limited accessibility. Additionally, features extracted from de Rham-Hodge flexibility and Hodge mode analysis can be incorporated into deep neural networks for the structure reconstruction from medium and low-resolution cryoEM maps [191]. Finally, due to its ability to characterize geometric traits and describe topological invariants, the proposed de Rham-Hodge paradigm opens an entirely new direction for the quantitative structure-function analysis of molecular and macromolecular datasets. The integration of de Rham-Hodge features and machine learning algorithms for the predictions of protein-ligand binding affinity, protein-protein binding affinity, protein folding stability change upon mutation, drug toxicity, solubility, partition coefficient, permeability, and plasma protein binding are under our

128

consideration.

Third, We introduce an evolutionary de Rham-Hodge method to offer a unified multiscale geometric and topological representation of data. The evolutionary de Rham-Hodge method is applied to analyze the topological and geometric characteristics through the evolution of manifolds which are a family of 3D multiscale shapes constructed from an evolutionary filtration process. In addition to exactly the topological persistence that would be obtained from persistent homology, the analysis of the evolutionary spectra of Hodge Laplacian operators portrays geometric progression. Specifically, appropriate treatments of the Hodge Laplacian boundary conditions give rise to three unique sets of singular spectra associated with the tangential gradient eigen field (T), the curl eigen field (C), and the tangential divergent eigen field (N). The multiplicities of the zero eigenvalues corresponding to the T, C, and N sets of spectra are exactly the persistent Betti-0 (β_0), Betti-1 (β_1), and Betti-2 (β_2) numbers one would obtain from persistent homology. Using discrete exterior calculus in close manifolds or compact manifolds with boundary, we show that investigating the first non-zero eigenvalues, i.e., Fiedler values, of the T, C, and N sets of evolutionary spectra unveil both the persistence for topological features and the geometric progression for the shape analysis. For a proof-of-concept analysis, the evolutionary de Rham-Hodge method is applied to a few benchmark examples, including the two-body system, four-body system, eight-body system, benzene (C_6H_6), and buckminsterfullerene (C_{60}). Extensive numerical experiments demonstrate that the present evolutionary de Rham-Hodge method captures the multiscale geometric progression and topological persistence of data. The proposed evolutionary de Rham-Hodge method provides a solid foundation for a wide variety of applications, including shape analysis, image processing, computer vision, pattern recognition, computer aided design, network analysis, computational biology, and drug design. As a proof-of-concept, we demonstrate the proposed de Rham-Hodge modeling and analysis by the B-factor prediction of a few challenging cases for which the conventional methods encountered difficulties. By using both eigenfunctions and eigenvalues at various scales, we show that the present evolutionary de Rham-Hodge method outperforms existing methods in computational biophysics for protein flexibility analysis. Since the evolutionary de Rham-Hodge

129

method can reveal both topological persistence and geometric progression, it will offer a powerful multiscale representation of data for machine learning, including deep learning. Finally, the present evolutionary de Rham-Hodge method opens new opportunities in further theoretical developments in differential geometry, such as the introduction of multiscale analysis to Riemannian connection, tensor bundle, characteristic class, index theory, and K-theory.

Our work on the decomposition of vector fields, spectral shape analysis on static shapes, and evolving shapes has already shown its effectiveness in biomolecular applications. We envision that it will lead to a rich set of features for machine learning-based shape analysis.

APPENDICES

EFFICIENCY AND ACCURACY COMPARISONS FOR 3D HODGE DECOMPOSITION

In this companion note, we provide a comprehensive comparison with an existing volumetric Hodge decomposition for piecewise-constant vector fields, and offer pseudocode for key components of our vector field decomposition.

A.1   Comparison with [3]

We first note that the vector fields considered in [3] have very different degrees of freedom (DoFs) compared to ours, and the differential operators are also discretized as different matrices.

A.1.1   Differences in formulation

In our paper, we discuss two vector field representations, the 1-form and 2-form representations, which are two different formulations for vector fields linked through the Hodge duality; both differ from the piecewise-constant vector field (PCVF) formulation in [3]. To make our comparison more concise, we will illustrate the main differences using our 2-form decomposition (the same comments will hold for the 1-form version):

$$\omega^2 = d\alpha_n^1 \oplus \delta\beta_t^3 \oplus h_t^2 \oplus h_n^2 \oplus \eta^2, \tag{A.1}$$

where $\delta\beta$ is a normal gradient (NG) field, $d\alpha$ is the tangential curl (TC) field, h_t is the normal harmonic (NH) field, h_n is the tangential harmonic (TH) field, and η is the component that is both a gradient and a curl field, also known as central harmonic (CH) field.

Degrees of freedom   The PCVF representation in [3] has $3|\mathcal{T}|$ DoFs for the input ω. The Nédélec edge element used for vector potential is actually the same as our Whitney 1-form basis, so α has $|\mathcal{E}| - |\mathcal{E}_B|$ DoFs. Additionally, the DoFs for 3 harmonic components are the same as well. However, the Crouzeix-Raviart element (nonconforming face-based piecewise linear scalar field) used for the

scalar potential contains $|\mathcal{F}| - |\mathcal{F}_B|$ DoFs. While both lead to valid cohomologies, the DoFs are different for the same mesh. For instance, our input 2-form contains $|\mathcal{F}|$ DoFs with $|\mathcal{T}|$ DoFs for scalar potential. The following equality reveals that the rest DoFs are the same by subtracting DoFs of scalar potential from total DoFs for both representation

$$3|\mathcal{T}| - (|\mathcal{F}| - |\mathcal{F}_B|) = |\mathcal{F}| - |\mathcal{T}|,$$

since the number of tets are linked to the number of faces through

$$4|\mathcal{T}| = 2|\mathcal{F}| - |\mathcal{F}_B|.$$

Our 1-form representation differs in DoFs for all the components of the decomposition except for the dimensionality of cohomologies.

Rank-deficient linear system L_2-projections are used in [3] to produce the different components. However, there is no explicit discussion on the gauge condition used for the curl component. While it is discussed that the curl operator contains a large rank deficiency, the stacked matrix of curl and gradient operators was incorrectly described as "almost-square" matrices on Page 99, which is not the case for 3D. Discussion on how to resolve rank deficient linear system is missing. In our approach, the gauge condition is explicitly specified and enforced through divergence of 1-forms with the dimension $|\mathcal{V}| - |\mathcal{V}_B|$, i.e., the number of interior vertices.

We show next numerical advantages of our approach, including a comparison to the statistics reported in Table 5.4 of [3].

A.1.2 Numerical tests

In the following, we use our 2-form representation to generate results for comparison with [3]. The results with a 1-form representation are similar.

A.1.2.1 Setting

We provide numerical comparisons evaluated on a PC with an *Intel(R) Xeon(R) CPU E5-1630 v3 @ 3.70GHz* processor and 16.0 GB RAM memory. All solvers are state-of-the-art implementations provided in *MATLAB 2017b*, including Cholesky decomposition (*chol*), QR decomposition (*qr*), partial eigenvalue problem solver (*eigs*) and partial singular value decomposition solver (*svds*).

Following [3], we test fields with analytic expressions on three models, a unit ball (radius 1), a unit ball with a cavity (outer sphere radius 1 and inner sphere radius 0.25), and a solid torus (major radius 1 and minor radius 0.25). NG and TC fields are tested on the unit ball, NH fields are tested on the unit ball with a cavity, TH fields are tested on the torus. The corresponding 3D triangulations are generated with *CGAL* 3D Delaunay Triangulation algorithms. They are listed below for self-containedness:

$$\mathcal{X}_{NG} := (x, y, z)$$
$$\mathcal{X}_{TC} := (y, -x, 0)$$
$$\mathcal{X}_{NH} := (x, y, z)/(x^2 + y^2 + z^2)^{3/2}$$
$$\mathcal{X}_{TH} := (y, -x)/(x^2 + y^2)$$
$$\mathcal{X}_{CH} := (x, y, z)/2$$

A.1.2.2 Comparisons of NG and TC

We explore the advantage, in terms of accuracy and efficiency, of our decomposition routines compared to [3]. The comparisons of NG and TC example fields are conducted on the unit ball. Because our approach has DoFs on primal cells (tetrahedrons) for the scalar potential while [3] has DoFs on primal facets, we will carefully adjust the mesh to have the same DoFs when comparing accuracy and efficiency. As seen in Fig A.1, our approach generally have much lower error rates. One of the reasons is that our sampling procedure through discrete differential forms is accurate in capturing interior line integrals and fluxes, i.e., a 1-form precisely samples line integral along each edge, and likewise for a 2-form for the flux across each triangle. The error in our case is mainly due to

inaccuracies of boundary approximation by mesh, i.e., the boundary facet normal of the tet mesh for the sphere are not exactly orthogonal to $\mathcal{X}_{TC}$. In terms of efficiency, our approach is faster when calculating the vector potential of TC component, even more so if we assume that the rank deficiency of [3] is fixed by the QR decomposition as mentioned in the original document (as we mentioned earlier, there was no explicit mention of gauge conditions). On the other hand, the NG component computations exhibits little difference, because both approaches involve solving an SPD linear system.
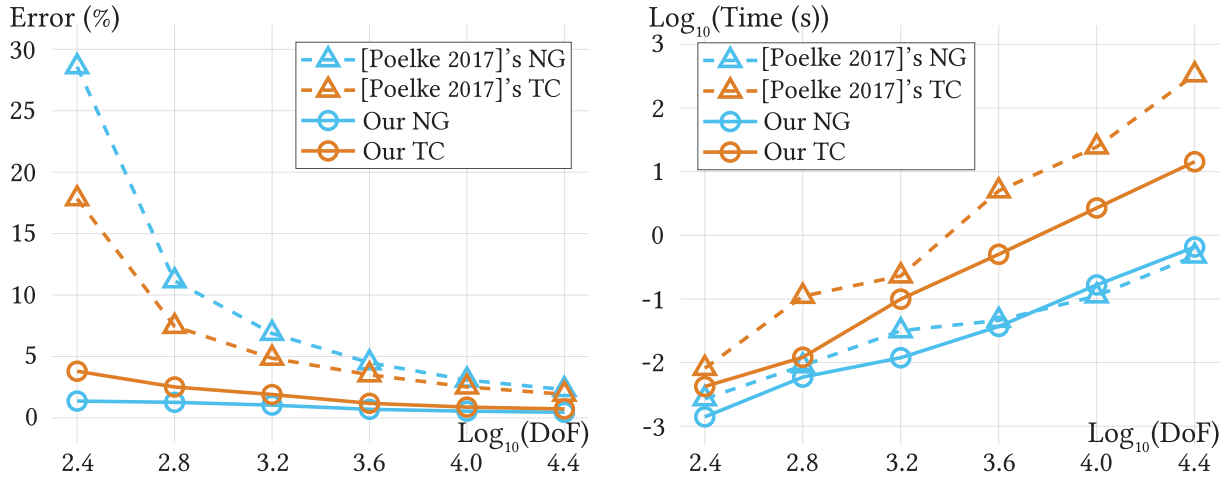
Figure A.1: **Comparisons on NG and TC fields.** *Left:* Our approach has lower error rates due to its linear-precise representation. *Right:* Our approach is faster for the TC component; the NG component computations show little difference because they both involve an SPD linear system.

A.1.2.3 Comparisons of NH and TH

We now explore the advantage, still in terms of accuracy and efficiency, of our harmonic fields computed from the kernel of our Laplacian matrix compared to cohomology representatives from [3]. The comparisons for the NH example field are conducted on the unit ball with a cavity, while the comparisons of the TH example field are conducted on the torus. We compare performance on the same meshes for each resolution. The resolutions are roughly controlled by a user-specified edge length parameter. Our approach also exhibits lower error rates on harmonic fields (see Fig. A.2, left) with shorter computational times (see Fig. A.2, right).
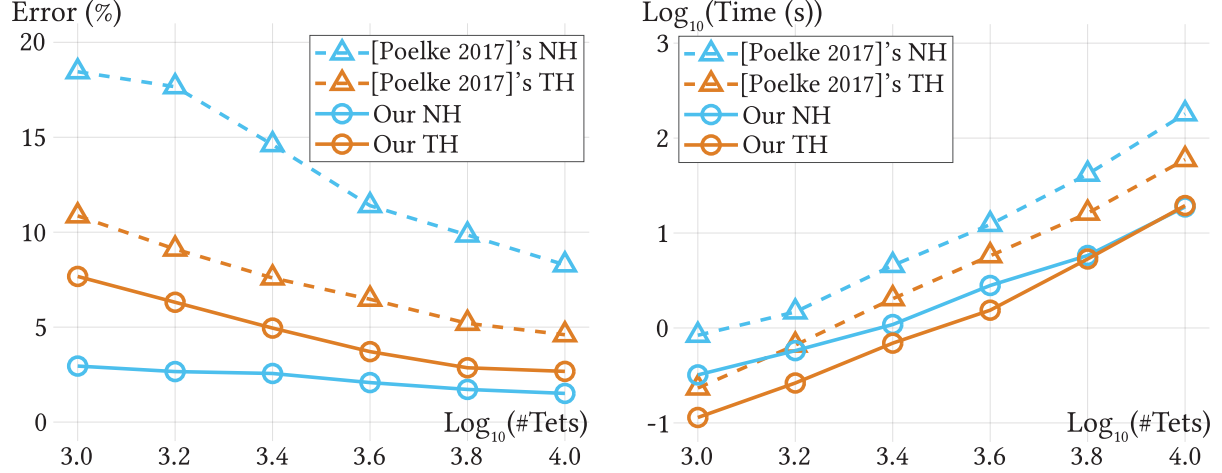
Figure A.2: **Comparisons on NH and TH fields.** *Left:* Our approach has lower error rates for harmonic fields. *Right:* Our approach has generally lower computation times due to our use of symmetric matrices of smaller sizes.

A.1.2.4 Comparison of L^2-norms of components

We also test whether an analytical vector field that is curl- or divergence-free as input can result in the correct components for the full decomposition: ideally, other components should have 0 L^2-norm. However, errors can be introduced due to discretization and residuals in the solves.

We first copy in Fig. A.1 the table of results as reported by [3] for reference. Since [3] did not provide details about the precise geometry of the unit ball with a cavity (inner radius) and the solid torus (major and minor radii), we create models with similar L^2-norms for the test fields. Fig. A.2 shows our reproduction of the tests in [3]. Finally, Fig. A.3 shows the L^2-norms of our proposed method with non-diagonal Galerkin Hodge star. Our method shares similar L^2-norms of each component for TC, NG and CH fields, but has lower noise for NH and TH fields.

	$\mathcal{X}_{TC}$	$\mathcal{X}_{NG}$	$\mathcal{X}_{CH}$	$\mathcal{X}_{NH}$	$\mathcal{X}_{TH}$
$\mathcal{X}$	1.27	1.55	1.75	6.14	3.17
TC Component	1.26	10e-13	10e-12	0.03	10e-3
NG Component	0.03	1.55	10e-13	1.00	0.17
CH Component	0.02	0.01	1.75	0.08	0.05
NH Component	10e-13	10e-12	10e-10	6.05	10e-11
TH Component	10e-12	10e-13	10e-12	10e-12	3.16

Table A.1: L^2**-norm of each component as reported in [3].**

136

	$\mathcal{X}_{TC}$	$\mathcal{X}_{NG}$	$\mathcal{X}_{CH}$	$\mathcal{X}_{NH}$	$\mathcal{X}_{TH}$
$\mathcal{X}$	1.2686	1.5504	1.7562	6.1383	3.1602
TC Component	1.2686	10e-16	10e-16	0.1752	0.0447
NG Component	10e-16	1.5540	10e-16	1.4430	0.1429
CH Component	0.0059	0.0055	1.7562	0.1786	0.1043
NH Component	10e-17	10e-15	10e-14	5.9611	10e-16
TH Component	10e-14	10e-15	10e-14	10e-14	3.1549

Table A.2: L^2**-norm of each component when we try to closely reproduce the tests in [3].**

	$\mathcal{X}_{TC}$	$\mathcal{X}_{NG}$	$\mathcal{X}_{CH}$	$\mathcal{X}_{NH}$	$\mathcal{X}_{TH}$
$\mathcal{X}$	1.2712	1.5611	1.7562	6.1435	3.1678
TC Component	1.2712	10e-16	10e-15	0.1499	0.0299
NG Component	10e-16	1.5611	10e-15	0.0162	0.0006
CH Component	0.0117	0.0088	1.7562	0.1186	0.1068
NH Component	0	0	0	6.1405	0
TH Component	0	0	0	0	3.1657

Table A.3: L^2**-norm of each component by our proposed method with the Galerkin Hodge star for Whitney basis functions.**

A.1.3   Summary

In summary, [3] and our approach are based on different discretization methods, with different basis elements. While cohomologies are preserved in both approaches, our approach tackles the full-blown five-component decomposition solely using symmetric semi-positive definite matrices with smaller size than what is proposed in Poelke's, which results in higher efficiency since we can leverage the efficiency of symmetric solvers. Experiments confirmed that differential form based discretization lead generally to better accuracy, partially due to our linear-precise line integral and flux sampling.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1]   K. Opron, K. L. Xia, and G. W. Wei, "Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis," *Journal of Chemical Physics*, vol. 140, no. 23410, p. 5, 2014.

[2]   K. Xia, X. Feng, Z. Chen, Y. Tong, and G.-W. Wei, "Multiscale geometric modeling of macromolecules i: Cartesian representation," *Journal of Computational Physics*, vol. 257, pp. 912–936, 2014.

[3]   K. Poelke, *Hodge-type decompositions for piecewise constant vector fields on simplicial surfaces and solids with boundary*. PhD thesis, FU Berlin, 2017.

[4]   R. Baradaran, C. Wang, A. F. Siliciano, and S. B. Long, "Cryo-em structures of fungal and metazoan mitochondrial calcium uniporters," *Nature*, vol. 559, no. 7715, pp. 580–584, 2018.

[5]   J. Jiang, Y. Wang, L. Sušac, H. Chan, R. Basu, Z. H. Zhou, and J. Feigon, "Structure of telomerase with telomeric dna," *Cell*, vol. 173, no. 5, pp. 1179–1190, 2018.

[6]   A. K. Singh, L. L. McGoldrick, E. C. Twomey, and A. I. Sobolevsky, "Mechanism of calmodulin inactivation of the calcium-selective trp channel trpv6," *Science advances*, vol. 4, p. 8, 2018.

[7]   T. Nishino, F. Rago, T. Hori, K. Tomii, I. M. Cheeseman, and T. Fukagawa, "Cenp-t provides a structural platform for outer kinetochore assembly," *The EMBO journal*, vol. 32, no. 3, pp. 424–436, 2013.

[8]   K. Hanawa-Suetsugu, S.-i. Sekine, H. Sakai, C. Hori-Takemoto, T. Terada, S. Unzai, J. R. Tame, S. Kuramitsu, M. Shirouzu, and S. Yokoyama, "Crystal structure of elongation factor p from thermus thermophilus hb8," *Proceedings of the National Academy of Sciences*, vol. 101, no. 26, pp. 9595–9600, 2004.

[9]   W. Humphrey, A. Dalke, and K. Schulten, "Vmd: visual molecular dynamics," *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[10]   W. V. D. Hodge, *The theory and applications of harmonic integrals*. Cambridge U. Press, 1941.

[11]   C. Shonkwiler, "Poincaré duality angles for riemannian manifolds with boundary," *arXiv preprint arXiv:0909.1967*, 2009.

[12]   A. Vaxman, M. Campen, O. Diamanti, D. Panozzo, D. Bommes, K. Hildebrandt, and M. Ben-Chen, "Directional Field Synthesis, Design, and Processing (star)," *Computer Graphics Forum*, 2016.

[13]   F. de Goes, M. Desbrun, and Y. Tong, "Vector field processing on triangle meshes," in *SIGGRAPH Course #27*, 2016.

[14] R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, Tensor Analysis, and Applications*, vol. 75 of *Applied Mathematical Sciences*. Springer-Verlag, 1988.

[15] G. Schwarz, *Hodge decomposition: a method for solving boundary value problems*. Springer-Verlag, 1995.

[16] J. Cantarella, D. DeTurck, and H. Gluck, "Vector calculus and the topology of domains in 3-space," *American math. monthly*, vol. 109, no. 5, pp. 409–442, 2002.

[17] C. Amrouche, C. Bernardi, M. Dauge, and V. Girault, "Vector potentials in three-dimensional non-smooth domains," *Mathematical Methods in the Applied Sciences*, vol. 21, no. 9, pp. 823–864, 1998.

[18] S. Caorsi, P. Fernandes, and M. Raffetto, "On the convergence of galerkin finite element approximations of electromagnetic eigenproblems," *SIAM J. Numerical Analysis*, vol. 38, no. 2, pp. 580–607, 2001.

[19] J. Stam, "Stable fluids," in *ACM SIGGRAPH proceedings*, pp. 121–128, 1999.

[20] F. Colin, R. Egli, and F. Y. Lin, "Computing a null divergence velocity field using smoothed particle hydrodynamics," *J. Comp. Phys.*, vol. 217, no. 2, pp. 680–692, 2006.

[21] S. Elcott, Y. Tong, E. Kanso, P. Schröder, and M. Desbrun, "Stable, circulation-preserving, simplicial fluids," *ACM Trans. Graph.*, vol. 26, no. 1, 2007.

[22] A. Chern, *Fluid dynamics with incompressible Schrödinger flow*. PhD thesis, California Institute of Technology, 2017.

[23] H. Bhatia, G. Norgard, V. Pascucci, and P.-T. Bremer, "The helmholtz-hodge decomposition– a survey," *IEEE Trans. Vis. Comp. Graph.*, vol. 19, no. 8, pp. 1386–1404, 2013.

[24] Y. Tong, S. Lombeyda, A. N. Hirani, and M. Desbrun, "Discrete multiscale vector field decomposition," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 445–452, 2003.

[25] K. Polthier and E. Preuß, "Variational approach to vector field decomposition," in *Data Visualization*, pp. 147–155, Springer, 2000.

[26] K. Poelke and K. Polthier, "Boundary-aware hodge decompositions for piecewise constant vector fields," *Computer-Aided Design*, vol. 78, pp. 126–136, 2016.

[27] F. Razafindrazaka, P. Yevtushenko, K. Poelke, K. Polthier, and L. Goubergrits, "Hodge decomposition of wall shear stress vector fields characterizing biological flows," 2018.

[28] D. N. Arnold and R. S. Falk, "A uniformly accurate finite element method for the reissner–mindlin plate," *SIAM Journal on Numerical Analysis*, vol. 26, no. 6, pp. 1276–1290, 1989.

[29] P. B. Monk, "A mixed method for approximating maxwell's equations," *SIAM Journal on Numerical Analysis*, vol. 28, no. 6, pp. 1610–1634, 1991.

[30] H. Bhatia, V. Pascucci, and P.-T. Bremer, "The natural helmholtz-hodge decomposition for open-boundary flow analysis," *IEEE Trans. Vis. Comp. Graph.*, vol. 20, no. 11, pp. 1566–1578, 2014.

[31] M. Desbrun, E. Kanso, and Y. Tong, "Discrete differential forms for computational modeling," in *Discrete Differential Geometry* (A. I. B. *et al.*, ed.), pp. 287–324, Birkhäuser Basel, 2008.

[32] V. Natarajan, P. Koehl, Y. Wang, and B. Hamann, "Visual analysis of biomolecular surfaces," in *Mathematical Methods for Visualization in Medicine and Life Science* (L. Linsen, H. Hagen, and B. Hamann, eds.), pp. 237–256, Springer Verlag: s, 2008.

[33] Z. Y. Yu, M. Holst, Y. Cheng, and J. A. McCammon, "Feature-preserving adaptive mesh generation for molecular shape modeling and simulation," *Journal of Molecular Graphics and Modeling*, vol. 26, pp. 1370–1380, 2008.

[34] R. B. Corey and L. Pauling, "Molecular models of amino acids, peptides, and proteins," *Review of Scientific Instruments*, vol. 24, no. 8, pp. 621–627, 1953.

[35] B. Lee and F. M. Richards, "The interpretation of protein structures: estimation of static accessibility," *Journal of molecular biology*, vol. 55, p. 3, 1971.

[36] F. M. Richards, "Areas, volumes, packing, and protein structure," *Annual review of biophysics and bioengineering*, vol. 6, no. 1, pp. 151–176, 1977.

[37] J. F. Blinn, "A generalization of algebraic surface drawing," *ACM Trans. Graph*, vol. 1, pp. 235–256, 1982.

[38] B. S. Duncan and A. D. Olson, "Shape analysis of molecular surfaces," *Biopolymers*, vol. 33, pp. 2–231, 1993.

[39] Q. Zheng, S. Yang, and G.-W. Wei, "Biomolecular surface construction by pde transform," *International journal for numerical methods in biomedical engineering*, vol. 28, no. 3, pp. 291–316, 2012.

[40] M. Chen, B. Tu, and B. Lu., "Triangulated manifold meshing method preserving molecular surface topology," *J. Mole. Graph. Model*, vol. 38, pp. 411–418, 2012.

[41] L. Li, C. Li, Z. Zhang, and E. Alexov, "On the dielectric "constant" of proteins: smooth dielectric function for macromolecular modeling and its implementation in delphi," *Journal of chemical theory and computation*, vol. 9, no. 4, pp. 2126–2136, 2013.

[42] H.-L. Cheng and X. Shi, "Quality mesh generation for molecular skin surfaces using restricted union of balls," *Computational Geometry*, vol. 42, no. 3, pp. 196–206, 2009.

[43] P. W. Bates, G. W. Wei, and S. Zhao, "Minimal molecular surfaces and their applications," *Journal of Computational Chemistry*, vol. 29, no. 3, pp. 380–91, 2008.

[44] K. L. Xia, K. Opron, and G. W. Wei, "Multiscale multiphysics and multidomain models — Flexibility and rigidity," *Journal of Chemical Physics*, vol. 139, no. 19410, p. 9, 2013.

[45] D. D. Nguyen, K. L. Xia, and G. W. Wei, "Generalized flexibility-rigidity index," *Journal of Chemical Physics*, vol. 144, no. 23410, p. 6, 2016.

[46] J. P. M. U. and, "and limitations of normal mode analysis in modeling dynamics of biomolecular complexes," *Structure*, vol. 13, pp. 373–180, 2005.

[47] N. Go, T. Noguti, and T. Nishikawa, "Dynamics of a small globular protein in terms of low-frequency vibrational modes," in *Proc. Natl. Acad. Sci*, (80), pp. 3696–3700, 1983.

[48] M. Tasumi, H. Takenchi, S. Ataka, A. M. Dwidedi, and S. Krimm, "Normal vibrations of proteins: Glucagon," *Biopolymers*, vol. 21, pp. 711–714, 1982.

[49] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. States, S. Swaminathan, and M. Karplus, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem*, vol. 4, pp. 187–217, 1983.

[50] M. Levitt, C. Sander, and P. S. Stern, "Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease and lysozyme," *J. Mol. Biol*, vol. 181, no. 3, pp. 423–447, 1985.

[51] I. Bahar, A. R. Atilgan, and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential," *Folding and Design*, vol. 2, pp. 173–181, 1997.

[52] A. R. Atilgan, S. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," *Biophysical journal*, vol. 80, no. 1, pp. 505–515, 2001.

[53] X. Feng, K. Xia, Y. Tong, and G.-W. Wei, "Geometric modeling of subcellular structures, organelles and large multiprotein complexes," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 28, pp. 1198–1223, 2012.

[54] K. L. Xia, X. Feng, Y. Y. Tong, and G. W. Wei, "Multiscale geometric modeling of macromolecules i: Cartesian representation," *Journal of Computational Physics*, vol. 275, pp. 912–936, 2014.

[55] R. Zhao, Z. Cang, Y. Tong, and G.-W. Wei, "Protein pocket detection via convex hull surface evolution and associated Reeb graph," *Bioinformatics*, vol. 34, 2018.

[56] T. K. Dey, F. Fan, and Y. Wang, "An efficient computation of handle and tunnel loops via reeb graphs," *ACM Trans. Graph*, vol. 32, p. 4, 2013.

[57] G. Carlsson, A. Zomorodian, A. Collins, and L. J. Guibas, "Persistence barcodes for shapes," *International Journal of Shape Modeling*, vol. 11, no. 2, pp. 149–187, 2005.

[58] H. Edelsbrunner and J. Harer, *Computational topology: An introduction*. American Mathematical Soc, 2010.

[59] Y. Yao, J. Sun, X. Huang, G. R. Bowman, G. Singh, M. Lesnick, L. J. Guibas, V. S. Pande, and G. Carlsson, "Topological methods for exploring low-density states in biomolecular folding pathways," *The Journal of chemical physics*, vol. 130, no. 14, p. 04B614, 2009.

[60] K. L. Xia and G. W. Wei, "Persistent homology analysis of protein structure, flexibility and folding," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 30, pp. 814–844, 2014.

[61] K. L. Xia, X. Feng, Y. Y. Tong, and G. W. Wei, "Persistent homology for the quantitative prediction of fullerene stability," *Journal of Computational Chemistry*, vol. 36, pp. 408–422, 2015.

[62] Z. X. Cang and G. W. Wei, "Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction ," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 34, p. 2, 2018.

[63] Z. X. Cang and G. W. Wei, "TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions," *PLOS Computational Biology*, vol. 13, p. 7, 2017.

[64] W. V. D. Hodge, *The theory and applications of harmonic integrals*. CUP Archive, 1989.

[65] R. Bott and L. W. Tu., *Differential forms in algebraic topology*, vol. 82. Science & Business Media: Springer, 2013.

[66] J. C. Mitchell, *Hodge decomposition and expanding maps on the flat tori*. PhD thesis, University of California, Berkeley, 1998.

[67] D. R. Hekstra, K. I. White, M. A. Socolich, R. W. Henning, V. Šrajer, and R. Ranganathan, "Electric-field-stimulated protein mechanics," *Nature*, vol. 540, no. 7633, pp. 400–405, 2016.

[68] J. G. D. L. Torre and V. A. Bloomfield, "Hydrodynamic properties of macromolecular complexes. i. translation.," *Biopolymers: Original Research on Biomolecules*, vol. 16, no. 8, pp. 1747–1763, 1977.

[69] A. N. Hirani, *Discrete exterior calculus*. PhD thesis, California Institute of Technology, 2003.

[70] M. Desbrun, A. N. Hirani, M. Leok, and J. E. Marsden, "Discrete exterior calculus," preprint, 2005.

[71] D. N. Arnold, R. S. Falk, and R. Winther, "Finite element exterior calculus, homological techniques, and applications," *Acta numerica*, vol. 15, pp. 1–155, 2006.

[72] R. Zhao, M. Desbrun, G.-W. Wei, and Y. Tong, "3d hodge decompositions of edge-and face-based vector fields," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, pp. 1–13, 2019.

[73] L.-H. Lim *arXiv preprint*, title = Hodge laplacians on graphs, year = 2015, archivePrefix = arXiv, eprint = 1507.05379.

[74] K. Xia and G.-W. Wei *arXiv preprint*, title = A review of geometric, topological and graph theory apparatuses for the modeling and analysis of biomolecular data, year = 2016, archivePrefix = arXiv, eprint = 1612.01735.

[75] D. B. Ray and I. M. Singer, "R-torsion and the laplacian on riemannian manifolds," *Advances in Mathematics*, vol. 7, no. 2, pp. 145–210, 1971.

[76] Y. Tong, S. Lombeyda, A. N. Hirani, and M. Desbrun, "Discrete multiscale vector field decomposition," in *ACM transactions on graphics (TOG)*, vol. 22, pp. 445–452, ACM, 2003.

[77] N. Foster and D. Metaxas, "Realistic animation of liquids," *Graphical models and image processing*, vol. 58, no. 5, pp. 471–483, 1996.

[78] H. Gao, M. K. Mandal, G. Guo, and J. Wan, "Singular point detection using discrete hodge helmholtz decomposition in fingerprint images," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1094–1097, IEEE, 2010.

[79] Y. Mochizuki and A. Imiya, "Spatial reasoning for robot navigation using the helmholtz-hodge decomposition of omnidirectional optical flow," in *2009 24th International Conference Image and Vision Computing New Zealand*, pp. 1–6, IEEE, 2009.

[80] N. N. Mansour, A. Kosovichev, D. Georgobiani, A. Wray, and M. Miesch, "Turbulence convection and oscillations in the sun," in *SOHO 14 Helio-and Asteroseismology: Towards a Golden Future*, vol. 559, p. 164, 2004.

[81] M. Akram and V. Michel, "Regularisation of the helmholtz decomposition and its application to geomagnetic field modelling," *GEM-International Journal on Geomathematics*, vol. 1, no. 1, pp. 101–120, 2010.

[82] R. Zhao, M. Wang, Y. Tong, and G.-W. Wei, "The de rham-hodge analysis and modeling of biomolecules," *arXiv preprint arXiv:1908.00572*, 2019.

[83] M. Desbrun, E. Kanso, and Y. Tong, "Discrete differential forms for computational modeling," in *Discrete differential geometry*, pp. 287–324, Springer, 2008.

[84] A. Zomorodian and G. Carlsson, "Computing persistent homology," *Discrete & Computational Geometry*, vol. 33, no. 2, pp. 249–274, 2005.

[85] H. Edelsbrunner and J. Harer, *Computational topology: an introduction*. American Mathematical Soc., 2010.

[86] P. Frosini and C. Landi, "Size theory as a topological tool for computer vision," *Pattern Recognition and Image Analysis*, vol. 9, no. 4, pp. 596–603, 1999.

[87] V. Robins, "Towards computing homology from finite approximations," in *Topology proceedings*, vol. 24, pp. 503–532, 1999.

[88] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," in *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pp. 454–463, IEEE, 2000.

[89] G. Carlsson, V. De Silva, and D. Morozov, "Zigzag persistent homology and real-valued functions," in *Proceedings of the twenty-fifth annual symposium on Computational geometry*, pp. 247–256, ACM, 2009.

[90] S. Chowdhury and F. Mémoli, "Persistent path homology of directed networks," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1152–1169, SIAM, 2018.

[91] Z. Cang, E. Munch, and G.-W. Wei, "Evolutionary homology on coupled dynamical systems," *arXiv preprint arXiv:1802.04677*, 2018.

[92] Z. Meng, D. V. Anand, Y. Lu, J. Wu, and K. Xia, "Weighted persistent homology for biomolecular data analysis," *arXiv preprint arXiv:1903.02890*, 2019.

[93] R. Wang, D. D. Nguyen, and G.-W. Wei, "Persistent spectral graph," *arXiv preprint arXiv:1912.04135*, 2019.

[94] G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian, "On the local behavior of spaces of natural images," *International journal of computer vision*, vol. 76, no. 1, pp. 1–12, 2008.

[95] D. Pachauri, C. Hinrichs, M. K. Chung, S. C. Johnson, and V. Singh, "Topology-based kernels with application to inference problems in alzheimer's disease," *IEEE transactions on medical imaging*, vol. 30, no. 10, pp. 1760–1770, 2011.

[96] G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson, and D. L. Ringach, "Topological analysis of population activity in visual cortex," *Journal of vision*, vol. 8, no. 8, pp. 11–11, 2008.

[97] P. Bendich, H. Edelsbrunner, and M. Kerber, "Computing robustness and persistence for images," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 6, pp. 1251–1260, 2010.

[98] P. Frosini and C. Landi, "Persistent betti numbers for a noise tolerant shape-based approach to image retrieval," *Pattern Recognition Letters*, vol. 34, no. 8, pp. 863–872, 2013.

[99] K. Mischaikow, M. Mrozek, J. Reiss, and A. Szymczak, "Construction of symbolic dynamics from experimental time series," *Physical Review Letters*, vol. 82, no. 6, p. 1144, 1999.

[100] T. Kaczynski, K. Mischaikow, and M. Mrozek, *Computational homology*, vol. 157. Springer Science & Business Media, 2006.

[101] V. De Silva, R. Ghrist, and A. Muhammad, "Blind swarms for coverage in 2-d.," in *Robotics: Science and Systems*, pp. 335–342, 2005.

[102] H. Lee, H. Kang, M. K. Chung, B.-N. Kim, and D. S. Lee, "Persistent brain network homology from the perspective of dendrogram," *IEEE transactions on medical imaging*, vol. 31, no. 12, pp. 2267–2277, 2012.

[103] D. Horak, S. Maletić, and M. Rajković, "Persistent homology of complex networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 03, p. P03034, 2009.

[104] P. Niyogi, S. Smale, and S. Weinberger, "A topological view of unsupervised learning from noisy data," *SIAM Journal on Computing*, vol. 40, no. 3, pp. 646–663, 2011.

[105] B. Wang, B. Summa, V. Pascucci, and M. Vejdemo-Johansson, "Branching and circular features in high dimensional data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 1902–1911, 2011.

[106] B. Di Fabio and C. Landi, "A mayer–vietoris formula for persistent homology with an application to shape recognition in the presence of occlusions," *Foundations of Computational Mathematics*, vol. 11, no. 5, p. 499, 2011.

[107] K. Xia and G.-W. Wei, "Persistent homology analysis of protein structure, flexibility, and folding," *International journal for numerical methods in biomedical engineering*, vol. 30, no. 8, pp. 814–844, 2014.

[108] K. Xia, X. Feng, Y. Tong, and G. W. Wei, "Persistent homology for the quantitative prediction of fullerene stability," *Journal of computational chemistry*, vol. 36, no. 6, pp. 408–422, 2015.

[109] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow, and V. Nanda, "A topological measurement of protein compressibility," *Japan Journal of Industrial and Applied Mathematics*, vol. 32, no. 1, pp. 1–17, 2015.

[110] V. Kovacev-Nikolic, P. Bubenik, D. Nikolić, and G. Heo, "Using persistent homology and dynamical distances to analyze protein binding," *Statistical applications in genetics and molecular biology*, vol. 15, no. 1, pp. 19–38, 2016.

[111] Z. Cang, L. Mu, K. Wu, K. Opron, K. Xia, and G.-W. Wei, "A topological approach for protein classification," *Computational and Mathematical Biophysics*, vol. 3, no. 1, 2015.

[112] Z. Cang and G.-W. Wei, "Topologynet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions," *PLoS computational biology*, vol. 13, no. 7, p. e1005690, 2017.

[113] Z. Cang and G.-W. Wei, "Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction," *International journal for numerical methods in biomedical engineering*, vol. 34, no. 2, p. e2914, 2018.

[114] Z. Cang, L. Mu, and G.-W. Wei, "Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening," *PLoS computational biology*, vol. 14, no. 1, p. e1005929, 2018.

[115] D. D. Nguyen, Z. Cang, K. Wu, M. Wang, Y. Cao, and G.-W. Wei, "Mathematical deep learning for pose and binding affinity prediction and ranking in d3r grand challenges," *Journal of computer-aided molecular design*, vol. 33, no. 1, pp. 71–82, 2019.

[116] D. D. Nguyen, K. Gao, M. Wang, and G.-W. Wei, "Mathdl: Mathematical deep learning for d3r grand challenge 4," *Journal of computer-aided molecular design*, pp. 1–17, 2019.

[117] K. Opron, K. Xia, and G.-W. Wei, "Communication: Capturing protein multiscale thermal fluctuations," 2015.

[118] D. Bramer and G.-W. Wei, "Multiscale weighted colored graphs for protein flexibility and rigidity analysis," *The Journal of chemical physics*, vol. 148, no. 5, p. 054103, 2018.

[119] D. Nguyen and G.-W. Wei, "Agl-score: Algebraic graph learning score for protein-ligand binding scoring, ranking, docking, and screening," *Journal of Chemical Information and Modeling*, 2019.

[120] D. D. Nguyen and G.-W. Wei, "Dg-gl: Differential geometry-based geometric learning of molecular datasets," *International journal for numerical methods in biomedical engineering*, vol. 35, no. 3, p. e3179, 2019.

[121] T. J. Willmore, *An introduction to differential geometry*. Courier Corporation, 2013.

[122] L. E.-J. Spruck, "Motion of level sets by mean curvature i," *J. Diff. Geom*, vol. 33, pp. 635–681, 1991.

[123] J. Gomes and O. Faugeras, "Using the vector distance functions to evolve manifolds of arbitrary codimension," in *International Conference on Scale-Space Theories in Computer Vision*, pp. 1–13, Springer, 2001.

[124] K. Mikula and D. Sevcovic, "A direct method for solving an anisotropic mean curvature flow of plane curves with an external force," *Mathematical Methods in the Applied Sciences*, vol. 27, no. 13, pp. 1545–1565, 2004.

[125] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations," *Journal of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.

[126] G.-W. Wei, "Differential geometry based multiscale models," *Bulletin of mathematical biology*, vol. 72, no. 6, pp. 1562–1622, 2010.

[127] T. Cecil, "A numerical method for computing minimal surfaces in arbitrary dimension," *Journal of Computational Physics*, vol. 206, no. 2, pp. 650–660, 2005.

[128] Q. Du, C. Liu, and X. Wang, "A phase field approach in the numerical study of the elastic bending energy for vesicle membranes," *Journal of Computational Physics*, vol. 198, no. 2, pp. 450–468, 2004.

[129] P. W. Bates, G.-W. Wei, and S. Zhao, "Minimal molecular surfaces and their applications," *Journal of Computational Chemistry*, vol. 29, no. 3, pp. 380–391, 2008.

[130] Z. Chen, N. A. Baker, and G.-W. Wei, "Differential geometry based solvation model ii: Lagrangian formulation," *Journal of mathematical biology*, vol. 63, no. 6, pp. 1139–1200, 2011.

[131] B. Wang and G.-W. Wei, "Object-oriented persistent homology," *Journal of computational physics*, vol. 305, pp. 276–299, 2016.

[132] S. Peters, "Convergence of riemannian manifolds," *Compositio Mathematica*, vol. 62, no. 1, pp. 3–16, 1987.

[133] C. Lessig, "A primer on differential forms," 2012. arXiv:1206.3323.

[134] C. B. Morrey, "A variational method in the theory of harmonic integrals, ii," *American Journal of Mathematics*, vol. 78, no. 1, pp. 137–170, 1956.

[135] K. O. Friedrichs, "Differential forms on riemannian manifolds," *Communications on Pure and Applied Mathematics*, vol. 8, no. 4, pp. 551–590, 1955.

[136] M. Desbrun, A. N. Hirani, and J. E. Marsden, "Discrete Exterior Calculus for variational problems in computer vision and graphics," in *Inter. Conf. on Decision and Control*, vol. 5, pp. 4902–4907, 2003.

[137] K. Crane, F. de Goes, M. Desbrun, and P. Schröder, "Digital geometry processing with discrete exterior calculus," in *SIGGRAPH Course #7*, 2013.

[138] K. Wang, Weiwei, Y. Tong, M. Desbrun, and P. Schröder, "Edge subdivision schemes and the construction of smooth vector fields," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1041–1048, 2006.

[139] A. Bossavit, "Computational electromagnetism and geometry. (5): The 'Galerkin Hodge'," *J. Japan Soc. Appl. Electromagn. & Mech.*, vol. 8, no. 2, pp. 203–209, 2000.

[140] J. A. Tropp, "Column subset selection, matrix factorization, and eigenvalue optimization," in *Symp. Disc. Algo.*, pp. 978–986, 2009.

[141] Y. Tong, P. Alliez, D. Cohen-Steiner, and M. Desbrun, "Designing quadrangulations with discrete harmonic forms," in *Symp. Geo. Proc.*, pp. 201–210, 2006.

[142] A. Demlow and A. N. Hirani, "A posteriori error estimates for finite element exterior calculus: the de rham complex," *Found. Comput. Math.*, vol. 14, no. 6, pp. 1337–1371, 2014.

[143] M. S. Mohamed, A. N. Hirani, and R. Samtaney, "Comparison of discrete hodge star operators for surfaces," *Comput. Aided Des.*, vol. 78, pp. 118–125, 2016.

[144] F. de Goes, M. Desbrun, M. Meyer, and T. DeRose, "Subdivision exterior calculus for geometry processing," *ACM Trans. Graph.*, vol. 35, no. 4, 2016.

[145] D. Arnold, *Finite Element Exterior Calculus*. SIAM, 2018.

[146] A. Bossavit and L. Kettunen, "Yee-like schemes on a tetrahedral mesh, with diagonal lumping," *Int. J. Num. Model. Elec. Net. Dev. Fields*, vol. 12, no. 1–2, pp. 129–142, 1999.

[147] L. C. Wollenman, M. R. V. Ploeg, M. L. Miller, Y. Zhang, and J. N. Bazil, "The effect of respiration buffer composition on mitochondrial metabolism and function," *PLoS One*, vol. 12, p. 11, 2017.

[148] K. Murakami, M. Stewart, K. Nozawa, K. Tomii, N. Kudou, N. Igarashi, Y. Shirakihara, S. Wakatsuki, T. Yasunaga, and T. Wakabayashi, "Structural basis for tropomyosin overlap in thin (actin) filaments and the generation of a molecular swivel by troponin-t," *Proceedings of the National Academy of Sciences*, vol. 105, no. 20, pp. 7200–7205, 2008.

[149] A. Lanza, E. Margheritis, E. Mugnaioli, V. Cappello, G. Garau, and M. Gemmi, "Nanobeam precession-assisted 3d electron diffraction reveals a new polymorph of hen egg-white lysozyme," *IUCrJ*, vol. 6, no. 2, pp. 178–188, 2019.

[150] A. Kuglstatter, M. Stihle, C. Neumann, C. Müller, W. Schaefer, C. Klein, J. Benz, R. P. Research, and E. Development, "Structural differences between glycosylated, disulfide-linked heterodimeric knob-into-hole fc fragment and its homodimeric knob–knob and hole–hole side products," *Protein Engineering, Design and Selection*, vol. 30, no. 9, pp. 649–656, 2017.

[151] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, "The energy landscapes and motions of proteins," *Science*, vol. 254, no. 5038, pp. 1598–1603, 1991.

[152] F. Tama, W. Wriggers, and C. L. B. III., "Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory," *Journal of molecular biology*, vol. 321, no. 2, pp. 297–305, 2002.

[153] D. Ming, Y. Kong, M. A. Lambert, Z. Huang, and J. Ma., "How to describe protein motion without amino acid sequence and atomic coordinates," *Proceedings of the National Academy of Sciences*, vol. 99, no. 13, pp. 8620–8625, 2002.

[154] W. Helfrich, "Elastic properties of lipid bilayers: Theory and possible experiments," *Zeitschrift für Naturforschung Teil C*, vol. 28, pp. 693–703, 1973.

[155] R. Tamstorf and E. Grinspun, "Discrete bending forces and their jacobians," *Graphical models*, vol. 75, no. 6, pp. 362–370, 2013.

[156] B. Sander, M. M. Golas, E. M. Makarov, H. Brahms, B. Kastner, R. Lührmann, and H. Stark, "Organization of core spliceosomal components u5 snrna loop i and u4/u6 di-snrnp within u4/u6. u5 tri-snrnp as revealed by electron cryomicroscopy," *Molecular cell*, vol. 24, no. 2, pp. 267–278, 2006.

[157] S. P. Muench, M. Huss, C. F. Song, C. Phillips, H. Wieczorek, J. Trinick, and M. A. Harrison, "Cryo-electron microscopy of the vacuolar atpase motor reveals its mechanical and regulatory complexity," *Journal of molecular biology*, vol. 386, no. 4, pp. 989–999, 2009.

[158] K. A. Sharp and B. Honig, "Electrostatic interactions in macromolecules - theory and applications," *Annual Review of Biophysics and Biophysical Chemistry*, vol. 19, pp. 301–332, 1990.

[159] F. Fogolari, A. Brigo, and H. Molinari, "The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology," *Journal of Molecular Recognition*, vol. 15, no. 6, pp. 377–92, 2002.

[160] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, *et al.*, "High-resolution crystal structure of an engineered human β2-adrenergic g protein–coupled receptor," *science*, vol. 318, no. 5854, pp. 1258–1265, 2007.

[161] F. Dong, M. Vijaykumar, and H. X. Zhou, "Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar," *Biophysical Journal*, vol. 85, no. 1, pp. 49–60, 2003.

[162] E. Alexov, E. L. Mehler, N. Baker, A. M. Baptista, Y. Huang, F. Milletti, J. Erik Nielsen, D. Farrell, T. Carstensen, M. H. Olsson, *et al.*, "Progress in the prediction of pka values in proteins," *Proteins: structure, function, and bioinformatics*, vol. 79, no. 12, pp. 3260–3275, 2011.

[163] J. Antosiewicz, J. A. McCammon, and M. K. Gilson, "The determinants of pK_as in proteins," *Biochemistry*, vol. 35, no. 24, pp. 7819–7833, 1996.

[164] J. E. Nielsen and J. A. McCammon, "Calculating pka values in enzyme active sites," *Protein Science*, vol. 12, no. 9, pp. 1894–1901, 2003.

[165] Y. Zhou, B. Lu, and A. A. Gorfe, "Continuum electromechanical modeling of protein-membrane interactions," *Physical Review E*, vol. 82, no. 4, p. 041923, 2010.

[166] D. D. Nguyen, B. Wang, and G. W. Wei, "Accurate, robust and reliable calculations of Poisson-Boltzmann binding energies," *Journal of Computational Chemistry*, vol. 38, pp. 941–948, 2017.

[167] J. A. Wagoner and N. A. Baker, "Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 22, pp. 8331–6, 2006.

[168] W. Im, D. Beglov, and B. Roux, "Continuum solvation model: electrostatic forces from numerical solutions to the Poisson-Boltzmann equation," *Computer Physics Communications*, vol. 111, no. 1-3, pp. 59–75, 1998.

[169] B. Honig and A. Nicholls, "Classical electrostatics in biology and chemistry," *Science*, vol. 268, no. 5214, pp. 1144–9, 1995.

[170] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 18, pp. 10037–10041, 2001.

[171] S. N. Yu, W. H. Geng, and G. W. Wei, "Treatment of geometric singularities in implicit solvent models," *Journal of Chemical Physics*, vol. 126, no. 24410, p. 8, 2007.

[172] D. Chen, Z. Chen, C. Chen, W. H. Geng, and G. W. Wei, "MIBPB: A software package for electrostatic analysis," *J. Comput. Chem*, vol. 32, pp. 657–670, 2011.

[173] A. Juffer, B. E., B. van Keulen, A. van der Ploeg, and H. Berendsen, "The electric potential of a macromolecule in a solvent: a fundamental approach," *J. Comput. Phys*, vol. 97, pp. 144–171, 1991.

[174] J. Liang and S. Subranmaniam, "Computation of molecular electrostatics with boundary element methods," *Biophys. J*, vol. 73, pp. 1830–1841, 1997.

[175] Y. N. Vorobjev and H. A. Scheraga, "A fast adaptive multigrid boundary element method for macromolecular electrostatic computations in a solvent," *Journal of Computational Chemistry*, vol. 18, no. 4, pp. 569–583, 1997.

[176] B. Lu, X. Cheng, and J. A. McCammon, "new-version-fast-multipole-method" accelerated electrostatic calculations in biomolecular systems.," *Journal of Computational Physics*, vol. 226, no. 2, pp. 1348–1366, 2007.

[177] W. Geng and R. Krasny, "A treecode-accelerated boundary integral poisson–boltzmann solver for electrostatics of solvated biomolecules," *Journal of Computational Physics*, vol. 247, pp. 62–78, 2013.

[178] J. Chen and W. Geng, "On preconditioning the treecode-accelerated boundary integral (tabi) poisson–boltzmann solver," *Journal of Computational Physics*, vol. 373, pp. 750–762, 2018.

[179] H. Edelsbrunner, D. Letscher, and A. Zomorodian, "Topological persistence and simplification," *Foundations of Computer Science*, vol. 2000, pp. 454–463, 2000.

[180] A. Bossavit, "Whitney forms: A class of finite elements for three-dimensional computations in electromagnetism," *IEE Proceedings A (Physical Science, Measurement and Instrumentation, Management and Education, Reviews)*, vol. 135, no. 8, pp. 493–500, 1988.

[181] D. D. Nguyen, K. Xia, and G.-W. Wei, "Generalized flexibility-rigidity index," *The Journal of chemical physics*, vol. 144, no. 23, p. 234106, 2016.

[182] T. Hazra, S. A. Ullah, S. Wang, E. Alexov, and S. Zhao, "A super-gaussian poisson–boltzmann model for electrostatic free energy calculation: smooth dielectric distribution for protein cavities and in both water and vacuum states," *Journal of mathematical biology*, vol. 79, no. 2, pp. 631–672, 2019.

[183] L. Wang, L. Li, and E. Alexov, "pka predictions for proteins, rna s, and dna s with the gaussian dielectric function using delphi pka," *Proteins: Structure, Function, and Bioinformatics*, vol. 83, no. 12, pp. 2186–2197, 2015.

[184] K. Xia and G.-W. Wei, "Multidimensional persistence in biomolecular data," *Journal of computational chemistry*, vol. 36, no. 20, pp. 1502–1520, 2015.

[185] M. Kac, "Can one hear the shape of a drum?," *The american mathematical monthly*, vol. 73, no. 4P2, pp. 1–23, 1966.

[186] I. Bahar, A. R. Atilgan, and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential," *Folding and Design*, vol. 2, no. 3, pp. 173–181, 1997.

[187] S. Nickell, C. Kofler, A. P. Leis, and W. Baumeister, "A visual approach to proteomics," *Nature reviews Molecular cell biology*, vol. 7, no. 3, pp. 225–230, 2006.

[188] C. V. Robinson, A. Sali, and W. Baumeister, "The molecular sociology of the cell," *Nature*, vol. 450, no. 7172, pp. 973–982, 2007.

[189] A. Leis, B. Rockel, L. Andrees, and W. Baumeister, "Visualizing cells at the nanoscale," *Trends in biochemical sciences*, vol. 34, no. 2, pp. 60–70, 2009.

[190] J. Peschek, N. Braun, T. M. Franzmann, Y. Georgalis, M. Haslbeck, S. Weinkauf, and J. Buchner, "The eye lens chaperone α-crystallin forms defined globular assemblies," *Proceedings of the National Academy of Sciences*, vol. 106, no. 32, pp. 13272–13277, 2009.

[191] D. Haslam, T. Zeng, R. Li, and J. He., "Exploratory studies detecting secondary structures in medium resolution 3d cryo-em images using deep convolutional neural networks," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 628–632, ACM, 2018.