Thèse n° 9158

Collapse-invariant properties of spaces equipped with signals or directions

Présentée le 26 août 2022

Faculté des sciences de la vie
Laboratoire pour la topologie et les neurosciences
Programme doctoral en mathématiques

pour l'obtention du grade de Docteur ès Sciences

par

Stefania EBLI

Acceptée sur proposition du jury

Prof. M. Viazovska, présidente du jury
Prof. K. Hess Bellwald, directrice de thèse
Prof. C. Landi, rapporteuse
Prof. M. Schaub, rapporteur
Prof. P. Vandergheynst, rapporteur

École
polytechnique
fédérale
de Lausanne

2022

Acknowledgements

I would like to start by thanking my supervisor Kathryn Hess for supporting me during this incredible journey. I am extremely grateful for all the opportunities you gave me to explore many different topics from maths to neuroscience.

Thank you to the members of my jury, Claudia Landi, Michael Schaub, Pierre Vandergheynst and Maryna Viazovska.

I have been extremely lucky to have wonderful collaborators from all over the word. Thank you to the Sunbeth group: Robin Belton, Robyn Brooks, Lisbeth Fajstrup, Brittany Terese Fasy, Catherine Ray, Nicole Sanderson, and Elizabeth Vidaurre. Thank you to Carina's group: Carina Curto, Alice Patania, Daniela Egas, Kate Morrison, and Nicole Sanderson. To my collaborators at EPFL: Gard Spreemann and Michaël Defferrard. And a special thank you to the DMT team: Celia and Kelly.

To my PhD sisters Celia and Adélie. For all the moments we shared and to how we grew together in this journey. Thank you also to all the other PhDs in the group Aras, Hao, Sam, Olivia, Lyne, Kelly.

To Daniela, Nicolas and Gard. For all the time spent in Geneva struggling with neuroscience data but always having fun.

A special thank to Guillaume and Daniela. For your friendship and unconditional support throughout these years.

To Kelly, Darrick, Nicolas, Celia and Adélie. For not being only amazing colleagues but for all the time we spent chatting, partying and enjoying time together.

Last but not least, to my family. For all their unconditional love and to always support my dreams.

Abstract

Collapsing cell complexes was first introduced in the 1930's as a way to deform a space into a topological-equivalent subspace with a sequence of elementary moves. Recently, discrete Morse theory techniques provided an efficient way to construct deformation retracts collapsing one space into the other while preserving global topological properties. This type of collapse, called a Morse matching, has been widely used to speed up computations in (persistent) homology by reducing the size of complexes.

Unlike classical collapses, in this thesis we consider topological spaces equipped with signals or directions. The main goal is then to reduce the size of the spaces while preserving as much as possible of both the topological structure and the properties of the signals or the directions.

In the first part of the thesis we explore collapsing in topological signal processing. In this context, each signal on the cells of a complex is processed using the combinatorial Laplacian and the resultant Hodge decomposition.

In Article 2.1 we provide an approach to signal compression and reconstruction on chain complexes that leverages the tools of algebraic discrete Morse theory. We first prove that any deformation retract of real finite-dimensional based chain complex is equivalent to a Morse matching. We then study the interaction between the Hodge decomposition and signal compression and reconstruction. Specifically, we prove that parts of a signal's Hodge decomposition are preserved under compression and reconstruction for specific classes of discrete Morse deformation retracts of a given based chain complex. Finally, we provide an algorithm to compute Morse matchings with minimal reconstruction error.

Complementary to our theoretic results in topological signal processing, we provide two applications in this field. Article 2.2 extends graph convolutional neural networks to simplicial complexes, while Article 2.3 presents a novel algorithm, inspired by the well-known spectral clustering algorithm, to embed simplices in a Euclidean space.

The object of our studies in the second part of the thesis is topological spaces equipped with a sense of direction. In the directed setting, the topology of the space is characterized by directed paths between fixed initial and terminal points.

Motivated by applications in concurrent programs, we focus on directed Euclidean cubical complexes and their spaces of directed paths.

In Article 3.1 we define a notion of directed collapsibility for Euclidean cubical complexes using the notion of past links, a combinatorial local representations of cubical complexes. We show that this notion of collapsability preserves given properties of the directed path spaces. In particular, we give sufficient conditions for a directed Euclidean cubical complex to have a contractible or a connected space of directed paths from a fixed initial vertex.

In Article 3.2 we extend these results, providing further conditions for directed collapses to preserve the contractability or conneectdness of spaces of directed paths. Furthermore, we provide simple combinatorial conditions for preserving the topology of past links. These conditions are the first step towards developing an algorithm that checks at each iteration if a collapse preserves certain properties of the directed space.

Keywords: Collapsing topological spaces, topological signal processing, discrete Morse theory, directed cubical complexes, concurrent programs.

Résumé

Les collapses de complexes cellulaires ont été introduit pour la première fois dans les années 1930 comme un moyen de déformer un espace en un sous-espace topologiquement équivalent à l'aide d'une suite de mouvements élémentaires. Récemment, les techniques de la théorie de Morse discrète ont fourni un moyen efficace de construire des rétractions déformations collapsant un espace dans l'autre tout en préservant les propriétés topologiques globales. Ce type de collapse , appelé correspondance de Morse, a été largement utilisé pour accélérer les calculs en homologie (persistante) en réduisant la taille des complexes.

Contrairement aux collapses classiques, nous considérons dans cette thèse des espaces topologiques dotés de signaux ou de directions. L'objectif principal est alors de réduire la taille des espaces, tout en préservant autant que possible à la fois la structure topologique et les propriétés des signaux ou des directions.

Dans la première partie de la thèse, nous explorons les collapses dans le traitement topologique du signal. Dans ce contexte, chaque signal sur les cellules d'un complexe est traité en utilisant le Laplacien combinatoire et la décomposition de Hodge qui en résulte.

Dans l'Article 2.1, nous proposons une approche de la compression et de la reconstruction des signaux sur les complexes de chaînes qui exploite les outils de la théorie algébrique discrète de Morse. Nous prouvons d'abord que toute rétraction par déformation d'un complexe de chaînes réel de dimension finie est équivalente à une correspondance de Morse. Nous étudions ensuite l'interaction entre la décomposition de Hodge et la compression et la reconstruction du signal. Plus précisément, nous prouvons que certaines parties de la décomposition de Hodge d'un signal sont préservées par la compression et la reconstruction pour des classes spécifiques de rétractions de déformation de Morse discrètes d'un complexe en chaîne basé donné. Enfin, nous fournissons un algorithme permettant de calculer les correspondances de Morse avec une erreur de reconstruction minimale.

En complément de nos résultats théoriques sur le traitement du signal topologique, nous fournissons deux applications dans ce domaine. L'Article 2.2 étend les réseaux de neurones convolutionnels de graphes aux complexes simpliciaux,

tandis que l'Article 2.3 présente un nouvel algorithme, inspiré de l'algorithme bien connu de regroupement spectral, pour intégrer les simpliciaux dans un espace euclidien.

L'objet de nos études dans la deuxième partie de la thèse est constitué par les espaces topologiques dotés d'un sens de direction. Dans le cadre dirigé, la topologie de l'espace est caractérisée par des chemins dirigés entre des points initiaux et terminaux fixes. Motivés par des applications dans les programmes concurrents, nous nous concentrons sur les complexes cubiques euclidiens dirigés et leurs espaces de chemins dirigés.

Dans l'Article 3.1 nous définissons une notion de collapsibilité dirigée pour les complexes cubiques euclidiens en utilisant la notion de liens passés, une représentation locale combinatoire des complexes cubiques. Nous montrons que cette notion de collapsabilité préserve des propriétés données des espaces de chemins dirigés. En particulier, nous donnons des conditions suffisantes pour qu'un complexe cubique euclidien dirigé ait un espace de chemins dirigés contractile ou connecxe à partir d'un sommet initial fixe.

Dans l'Article 3.2, nous étendons ces résultats, en fournissant des conditions supplémentaires sur le moment où les collapses dirigés préservent la contractabilité ou la connexité des espaces de chemins dirigés. De plus, nous fournissons des conditions combinatoires simples pour préserver la topologie des liens passés. Ces conditions sont la première étape vers le développement d'un algorithme qui vérifie à chaque itération si un collapse préserve certaines propriétés de l'espace dirigé.

Mots clés: Collapses des espaces topologiques, traitement du signal topologique, théorie discrète de Morse, complexes cubiques dirigés, programmation concurrente.

Contents

Chapter 1

Introduction

This thesis consists of a collection of five papers co-written by the author [1, 2, 3, 4, 5] and connected by the leitmotif of collapsing topological spaces. Collapses are commonly used in topology to reduce the size of a space while preserving its global topological features. Unlike classical collapses, we consider here spaces with additional information. Specifically, we consider topological spaces endowed with signals or directions. Our goal is then to collapse these spaces while preserving as much as possible of both their topology and the properties of the signals or the direction.

In this introduction we offer an non-technical overview of this problem. The aim is to provide the reader with general intuition on how signals and direction create additional properties of topological spaces and how questions on collapsing these variables naturally arise. In Section 1.1 we recall classical notions of collapsing for topological spaces. We then contextualize the problem of collapsing spaces endowed with signals or directions, respectively in Section 1.2 and Section 1.3. Finally, in Section 1.2.1 and Section 1.3.1 we describe our contributions to these fields. The articles about collapsing spaces with signal are then included in Chapter 2, while Chapter 3 presents the articles on collapsing spaces with directions.

1.1 Collapsing Topological Spaces

In this section, we first recall some basic notions in algebraic topology. We refer the reader to [6] for a more detailed exposition. Then we present two methods to collapse topological spaces. We start by introducing elementary collapses, a sequence of elementary moves to reduce the size of a space while preserving its topological features. Finally, we present the main concepts of discrete Morse

theory, focusing on how to collapse spaces via acyclic partial matchings.

1.1.1 Background in Algebraic Topology

Definition 1.1 (Regular Cell Complexes)**.** A finite *regular cell complex* is a topological space X with a partition into subspaces $\{X_\alpha\}_{\alpha \in P_X}$, called *cells*, satisfying the following conditions.

1. For each $x \in X$, every sufficiently small neighborhood of x intersects finitely many X_α.
2. For all α, β, $\overline{X}_\alpha \cap X_\beta \neq \emptyset$ if and only if $X_\beta \subseteq \overline{X}_\alpha$.
3. (Regularity) For every α, there is a homeomorphism φ_α of a closed ball in $\mathbb{R}^{n_\alpha}$ to X_α that maps the interior of the ball homeomorphically onto X_α.

A cell X_α is called an *n-cell* (or an an *n-dimensional cell*) if it is homeomorphic to $\mathbb{R}^n$ for some n. The space given by the union of cells of dimension less or equal than n is called the *n-skeleton* of X and denoted by X_n. A map φ_α satisfying condition (3) such that the boundary of the closed ball is contained in X_{n-1} is called a *characteristic map* of X_α.

Note that condition (2) provides a poset structure on the indexing set P_X given by the relation $\beta \lhd \alpha \Leftrightarrow X_\beta \subseteq \overline{X}_\alpha$, while condition (3) guarantees that this poset structure encodes all the topological information about X. As a consequence, regular cell complexes are essentially combinatorial objects whose structure is determined by the cells and their relations in the poset.

Intuitively, regular cell complexes can be constructed through an iterative gluing procedure. One starts with a set of 0-cells (vertices), then the 1-cells (edges) are attached to these by gluing the end-points of closed line segments to them. Then one can glue the 2-cells (closed disks) by attaching their boundary to any cycle, and so on in higher dimensions. The class of regular cell complexes includes simplicial complexes and Euclidean cubical complexes, which will be at the center our studies in Chapter 2 and in Chapter 3 respectively.

Example 1.2. Simplicial complexes are cell complexes constructed by gluing together points, line segments, triangles, and their n-dimensional counterparts (see Figure 1.1).

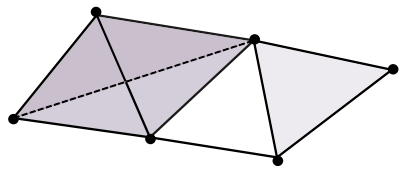
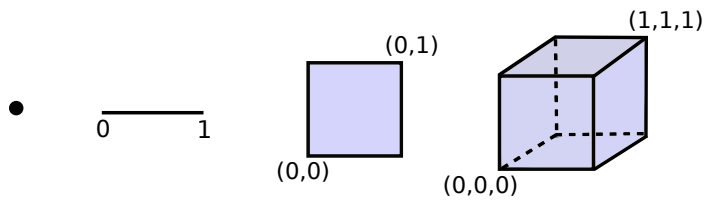


Figure 1.1: A simplicial complex.

Example 1.3. Euclidean cubical complexes are cell complexes whose n-cells are unitary cubes in $\mathbb{R}^n$ (see Figure 1.2-A and Figure 1.2-B).

A Elementary Cubes



B Euclidean Cubical Complex

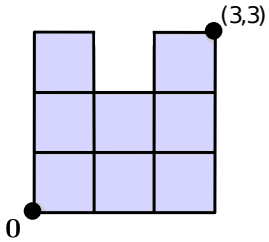


Figure 1.2: A Euclidean cubical complex and elementary cubes up to dimension 3.

The attaching map structure of the cells in a complex can be encoded in a purely algebraic framework, called a *chain complex*. To construct the chain complex of a regular cell complex, we need a boundary operator and a notion of orientation–a signed *incidence function* on P_X.

Definition 1.4 (Incidence Function)**.** An incidence function associated to a regular cell complex X is a map $[\bullet : \bullet] : P_X \times P_X \to \{-1, 0, 1\}$ satisfying the following conditions.

1. If $[\sigma, \tau] \neq 0$, then $\sigma \lhd \tau$, and there are no cells between σ and τ in the incidence poset.
2. For any $\sigma \lhd \tau$, $\sum_{\gamma \in P_X} [\sigma : \gamma][\gamma : \tau] = 0$.

Definition 1.5 (Chain complex)**.** The *n-th chain group* $\mathbf{C}_n(X)$ of a regular cell complex X is the free $\mathbb{R}$-module with basis the n-cells of X. Given an incidence function $[\bullet : \bullet] : P_X \times P_X \to \{-1, 0, 1\}$, we can define a *boundary operator* $\partial_n : \mathbf{C}_n(X) \to \mathbf{C}_{n-1}(X)$ by

$$\partial_n \tau = \sum_{\sigma \in P_X} [\sigma : \tau]\sigma.$$

This leads us to define the *chain complex* over $\mathbb{R}$ associated to X as:

$$\ldots \to \mathbf{C}_{n+1}(X) \xrightarrow{\partial_{n+1}} \mathbf{C}_n(X) \xrightarrow{\partial_n} \mathbf{C}_{n-1}(X) \to \ldots.$$

Remark 1.6 (Based Chain Complex)**.** Chain complexes are a purely algebraic objects defined on a basis given by the cells of the complex. This structure is generalized by *based chain complexes* [7], chain complexes with a graded structure over a given base–not necessarily provided by geometric objects. Based chain complexes will play a pivotal role in Article 2.1.

Note that condition (2) of Definition 1.4 implies that $\operatorname{Im} \partial_{n+1} \subseteq \operatorname{Ker} \partial_n$ for all n.

Definition 1.7 (Homology)**.** Let X be a regular cell complex. The *homology* of the real chain complex $\mathbf{C}(X)$ is the family of modules

$$\mathrm{H}_n(\mathbf{C}(X), \mathbb{R}) = \frac{\operatorname{Ker} \partial_n}{\operatorname{Im} \partial_{n+1}}, n \in \mathbb{N}.$$

The elements of $\operatorname{Ker} \partial_n$ are called *n-cycles*, and the elements of $\operatorname{Im} \partial_{n+1}$ are called *n-boundaries.*

Given two topological spaces X and Y, we now define an algebraic analogue of the notion of deforming one space into the other while preserving topological properties.

A *chain map* between two chain complexes $(\mathbf{C}, \partial)$ and $(\mathbf{C}', \partial')$ is given by $\mathbb{R}$-module homomorphisms $\Psi_n : \mathbf{C}_n \longrightarrow \mathbf{C}_n'$ for each n, such that

$$\Psi_{n-1} \circ \partial_n = \partial_n' \circ \Psi_n.$$

It is easy to see that a chain map $\Psi : \mathbf{C} \to \mathbf{C}'$ induces homomorphisms $\mathrm{H}_n(\Psi) : \mathrm{H}_n(\mathbf{C}) \to \mathrm{H}_n(\mathbf{C}')$ in homology. A *chain homotopy* between two chain maps $\Psi, \Phi : \mathbf{C} \longrightarrow \mathbf{C}'$ is given by maps $h_n : \mathbf{C}_n \longrightarrow \mathbf{C}'_{n+1}$, for all $n \geq 0$ such that

$$\Psi_n - \Phi_n = \partial'_{n+1} \circ h_n + h_{n-1} \circ \partial_n.$$

Two chain homotopic maps induce the same homomorphism on homology modules. For two chain complexes $(\mathbf{C}, \partial)$ and $(\mathbf{C}', \partial')$, a pair of chain maps

$$\Psi : \mathbf{C} \to \mathbf{C}' \text{ and } \Phi : \mathbf{C}' \to \mathbf{C}$$

are *chain equivalances* if $\Psi \circ \Phi$ and $\Phi \circ \Psi$ are chain homotopic to the identities on $\mathbf{C}'$ and $\mathbf{C}$, respectively. Note that this implies that the maps induced on the homology modules by Φ and Ψ are isomorphisms. The chain equivalences Ψ and Φ form a deformation retract of the chain complexes $\mathbf{C}$ and $\mathbf{C}'$ if $\Psi \circ \Phi$ is the identity map on $\mathbf{C}'$. Deformation retracts will be often depicted as the following diagram

$$\mathbf{C}' \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C} \circlearrowleft h$$

where h is a chain homotopy from $\Phi\Psi$ to $\mathrm{Id}_{\mathbf{C}}$.

1.1.2 Background on Elementary Collapses

Given two regular cell complexes and their associated chain complexes, it is usually very hard to decide whether there exists a deformation retract between them. In the 1930's J. H. C. Whitehead introduced the notion of collapsing as a simple way to deform a space into a subspace by a finite sequence of elementary moves called *elementary collapses* [8].

Definition 1.8 (Elementary Collapses)**.** Let X be a cell complex and Y a subcomplex of X. Then X collapses to Y by an *elementary collapse* if the following conditions hold.

1. $X = Y \cup \sigma \cup \tau$, where $\dim(\tau) = n$, $\dim(\sigma) = n - 1$ and $\sigma, \tau \notin Y$
2. There exist a pair of disks (D_n, D_{n-1}) and a map $\phi : D_n \to X$ such that

(a) ϕ is the characteristic map for τ.

(b) $\phi|_{D_{n-1}}$ is the characteristic map for σ.

(c) $\phi(P_{n-1}) \subseteq Y_{n-1}$ where $P_{n-1} = \overline{\partial_n D_n \setminus D_{n-1}}$.

Note that the second condition ensures that by removing σ and τ, from X we obtain a subcomplex of X with the correct attaching map structure.

We say that X *collapses* to Y if there exists a finite sequence of elementary collapses from X to Y.

Example 1.9. In the setting of simplicial complexes, collapsibility is a completely combinatorial notion. Definition 1.8 is equivalent to checking the following conditions. An elementary collapse in a simplicial complex X consists of removing a pair of simplices (σ, τ) such that

- $\dim(\sigma) = \dim(\tau) - 1$.
- τ is *maximal*–it is not contained in any other simplex than itself.
- σ is a *free face*–the only simplices containing it are τ and σ.

Figure 1.3 shows an elementary collapse of a pair (σ, τ) in a simplcial complex.

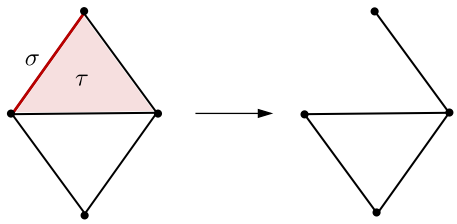


Figure 1.3: Elementary collapse in a simplicial complex removing (σ, τ).

The following theorem ensures that collapsing a cell complex X into a complex Y preserves its homology.

Theorem 1.10 ([8])**.** *Let X be a cell complex such that X collapses to Y. Then there exists a deformation retract*

$$\begin{tikzcd} \mathbf{C}(Y) \arrow[r, shift right, "\Phi"'] & \mathbf{C}(X) \arrow[l, shift right, "\Psi"'] \arrow[loop right, "h"] \end{tikzcd}$$

Remark 1.11. Although elementary collapses are simple moves to reduce a cell complex, it is not always possible to verify the homotopy-type of a space by collapsing it. For example, the R. H. Bing's house with two rooms [9] is contractible (homotopy equivalent to a point), but not collapsible.

1.1.3 Background in Discrete Morse Theory

Discrete Morse theory was originally developed by Forman in the 1990s as a discrete version of Morse theory [10]. The main goal of discrete Morse theory is to find critical cells on a cell complex, in order to collapse its structure while preserving its topological features. Unlike elementary collapses, discrete Morse theory reduces a complex to another complex that is not necessarily a subcomplex via a sequence of allowed collapses defined by *acyclic partial matchings.*

Definition 1.12 (Acyclic partial matching)**.** An *acyclic partial matching* $\mathcal{V}$ on a regular cell complex X consists of a partition of the cells of X into three disjoint sets $\mathcal{M}, \mathcal{D}$ and $\mathcal{U}$ along with a bijection $\mu : \mathcal{D} \to \mathcal{U}$ satisfying:

1. $\sigma \triangleleft \mu(\sigma)$ for every $\sigma \in \mathcal{D}$,
2. the transitive closure of the binary relation defined by $\sigma' \prec_\mu \sigma$ whenever $\sigma' \triangleleft \mu(\sigma)$ is a partial order on the cells of $\mathcal{D}$.

The cells of $\mathcal{M}$ are called the *critical cells* of $\mathcal{V}$. The acyclic partial matching $\mathcal{V}$ is called *Morse matching.*

Example 1.13. In Figure 1.4 we show a Morse matching $\mathcal{V}$ on a cell complex X. The pairs $(\sigma, \mu(\sigma))$ in the Morse matching $\mathcal{V}$ are visually depicted by blue arrows running from the cell σ to the cell $\mu(\sigma)$, see Figure 1.4-left. Alternatively, one can visually depict the Morse matching in the Hasse diagram associated to X. The Hasse diagram of X is the directed graph whose nodes are the cells of X and whose edges are defined as follows. For every cell τ and σ we draw an arrow from τ to σ if $\sigma \triangleleft \tau$. The edges corresponding to pairs $(\sigma, \mu(\sigma))$ in $\mathcal{V}$ are reversed, namely we draw an arrow from σ to $\mu(\sigma)$, see Figure 1.4-right. The acyclicity condition in the definition of Morse matching is satisfied if such a graph doesn't contain directed cycles [10]. Critical cells are represented by red nodes in the graphs. The collapsed cell complex $X_\mathcal{V}$ is showed in Figure 1.4-center.

A classical result of discrete Morse theory states that a cell complex equipped with an acyclic partial matching $\mathcal{V}$ can be deformed into a cell complex with number of cells equal to the number of critical cells in $\mathcal{V}$ [10]. In practice, for a cell complex X and a partial matching $\mathcal{V}$, one can construct the deformation retract as follows. We define the *collapsed complex* $X_\mathcal{V}$ as

$$X_\mathcal{V} = X \setminus \bigcup_{\sigma \in \mathcal{D}} (\sigma, \mu(\sigma)).$$

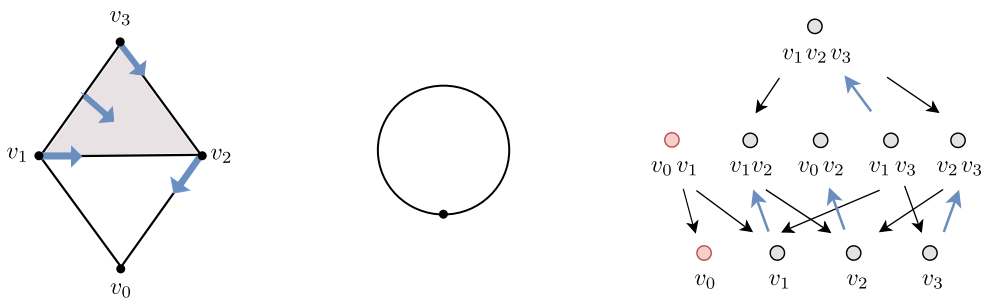


Figure 1.4: Morse matching $\mathcal{V}$ on a cell complex X. Left: The cell complex X with the Morse matching depicted by blue arrows. Right: The Morse matching $\mathcal{V}$ represented on the Hasse diagram of X. Center: The collapsed cell complex $X_{\mathcal{V}}$.

One can associate to $X_{\mathcal{V}}$ a chain complex by constructing a boundary operator $\partial_{\mathcal{V}}$ as described in [11]. Then, $X_{\mathcal{V}}$ is related to X by the following theorem.

Theorem 1.14 ([10, 7, 11])**.** *Let X be a regular cell complex and $\mathcal{V}$ an acyclic partial matching on X. Then there exist a deformation retract*

$$\begin{tikzcd} \mathbf{C}(X_{\mathcal{V}}) \arrow[r, shift right, "\Phi"'] & \mathbf{C}(X) \arrow[l, shift right, "\Psi"'] \arrow[loop right, "h"] \end{tikzcd}$$

The theorem above shows how discrete Morse theory allow us to reduce the size of complexes and to preserve their topological features by constructing deformation retracts via acyclic partial matching.

Remark 1.15. Finding a Morse matching with given properties is usually a hard problem. For example, the task of finding a matching that minimises the number of critical cells is already known to be NP-hard [12].

1.2 Collapsing in Topological Signal Processing

In this section we contextualize the articles presented in Chapter 2. We first present an overview on topological signal processing, a broad term used to refer to the field concerned with topological spaces endowed with signals. Then we discuss the idea of collapsing such spaces, and we provide intuition on the properties one would like to preserve. Finally, we summarize our contributions and discuss future perspectives.

Topological Signal Processing Signal processing research concerns the analysis, synthesis, and modification of signals, which can be broadly defined as functions conveying "information about the behavior or attributes of some phenomenon [13]". Examples of such signals include value of pixels of an image, temperatures of cities in a given area, or any biological measurement.

Historically, signal processing was defined for data lying on metric spaces. Techniques in this field have been extended only recently to a larger family of topological spaces [14], such as graphs, simplicial complexes, and cell complexes [15, 16, 17, 18]. Here, signals are functions on the cells, the building blocks of the spaces, and relationships between signals are encoded by how cells are attached. For instance, on a graph, functions on nodes are pairwise related by the edges that connect them.

From a formal point of view, a signal s on a complex X can be thought as an element in the real chain complex $\mathbf{C}(X)$. This mathematical object, when endowed with a degree-wise inner product, allow us to define the combinatorial Laplacian Δ_n [19], a natural generalization of the classical time-delay operator [15, 20]. The eigenvectors of the Laplacian play the role of a 'topological' Fourier basis transforming a signal into a topologically meaningful coordinate system. Additionally, the combinatorial Laplacian gives rise to the combinatorial Hodge decomposition [19]

$$\mathbf{C}_n(X) = \operatorname{Im} \partial_{n+1} \oplus \operatorname{Ker} \Delta_n \oplus \operatorname{Im} \partial_n^{\dagger},$$

the components of which each have their own topological interpretation [15]. For instance, Eckmann proved that $\operatorname{Ker} \Delta_n$ is isomorphic to the n-th homology group $\mathrm{H}_n(\mathbb{R})$ [19].

This Fourier representation has proven to be useful in multiple applications [21, 22]. In graph signal processing, it has been exploited for signal smoothing and denoising [23, 24], node embeddings via Laplacian eigenmaps [25, 26], graph neural networks [27, 28], and signal compression and reconstruction [29]. Some of these applications have recently been extended to signal processing on cell and simplicial complexes [30].

In this context, we present two articles providing novel applications of signal processing on simplicial complex. In particular, Article 2.2 extends graph convolutional neural networks to simplicial complexes, while Article 2.3 presents an algorithm to embed simplices inspired by Laplacian eigenmaps.

Collapsing in Toplogical Signal Processing In Section 1.1, we discussed how discrete Morse theory efficiently generates deformation retracts that reduce the size of a cell complex while preserving its global topological structure. For this reason, this technique has been widely used to compress 3D images [31] or to speed up computations of (persistent) homology [11].

The main question we address here is how such deformation retracts act on signals on a cell complex. Specifically, from a topological signal processing point of view, it is interesting to study which parts of a signal's projection on the Hodge decomposition are preserved.

Recall that a deformation retract of a chain complex $\mathbf{C}$ onto $\mathbf{D}$ consists of chain equivalences (Ψ, Φ, h) in the diagram

$$\mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C} \circlearrowleft h$$

and a chain homotopy $h : \mathbf{C} \to \mathbf{D}$ between $\Phi\Psi$ and $\mathrm{Id}_{\mathbf{C}}$ such that $\Psi\Phi = \mathrm{Id}_{\mathbf{D}}$.

It is well known that usually neither chain or cochain maps between two complexes with degree-wise inner product respect the grading of the Hodge decomposition. Therefore, we focus on a different notion of preservation by examining the effect of applying $\Phi\Psi$ to an element $s \in \mathbf{C}_n$. Here the composition $\Phi\Psi s$ encodes the compression and reconstruction of a signal $s \in \mathbf{C}$ when performing a retraction to a chain complex $\mathbf{D}$. In order to understand and evaluate how compression and reconstruction changes the signal, one can study the *topological reconstruction error* given by difference $s - \Phi\Psi s$. Our main contribution in Article 2.1 is a theorem proving that, for a class of deformation retracts (Ψ, Φ, h), called $(n, n-1)$-*free*, the topological reconstruction error has trivial cocycle reconstruction. In particular, we prove s and $\Phi\Psi s$ agree when projected on the cocycles $\operatorname{Ker} \partial_{n+1}^{\dagger}$.

This results provides a clear topological interpretation of the compression and reconstruction via discrete Morse theory. Additionally, it opens up to computational questions on how to minimize the reconstruction of the signal that is not preserved. Article 2.1 offers a detailed discussion on this problem leading to algorithms for computing optimal single collapses.

Such algorithms were motivated by defining pooling layers preserving topological features in simplicial neural networks, developed in Article 2.2. Despite the recent development of further types of neural networks on topological spaces [32, 33, 34, 35], such architectures do not include pooling layers that take into account topological properties of both the signals and the spaces. We envision to

define such layers using the theory developed in Article 2.1, which would allow to reduce the size of a complex and pool the signal preserving topological components.

Collapsing using discrete Morse theory has potential application also in the harmonic clustering algorithm developed in Article 2.3. Here we developed a clustering method for simplices of any dimensions via the eigenvectors of the Laplacian. The algorithm for optimal collapses presented in Article 2.1 could be use a pre-process method to reduce the size of a complex to speed up the computations of the eigenvectors necessary for clustering.

1.2.1 Contributions in Topological Signal Processing

In this framework, we present three articles providing both novel theory and applications of topological signal processing. Our contributions can be summarized as follows.

- Article 2.1, "Morse Theoretic Signal Compression and Reconstruction on Chain Complexes", presents a novel approach to signal compression and reconstruction on based chain complexes leveraging tools of discrete Morse theory. We first prove that any deformation retract of real degree-wise finite-dimensional chain complexes is equivalent to a Morse matching. We then study how reducing the size of complexes by particular types of Morse matchings changes the signal on its cells. In particular, we define specific classes of discrete Morse theory deformation retracts (Ψ, Φ, h) that perfectly preserve parts of the signal's Hodge decomposition. Specifically, they are characterised by the following (co)cycle reconstruction properties (Theorem 4.5 Article 2.1).

  1. (Cocycle Reconstruction) A signal $s \in \mathbf{C}_n$ and its reconstruction $\Phi\Psi s$ have the same cocycle information:
  $$\mathrm{Proj}_{\mathrm{Ker}\,\partial_{n+1}^{\dagger}}(\Phi\Psi s - s) = 0 \text{ for all } s \in \mathbf{C}_n.$$
  2. (Cycle Reconstruction) A signal $s \in \mathbf{C}_{n-1}$ and the adjoint of the reconstruction $\Psi^{\dagger}\Phi^{\dagger}s$ have the same cycle information:
  $$\mathrm{Proj}_{\mathrm{Ker}\,\partial_{n-1}}(\Psi^{\dagger}\Phi^{\dagger}s - s) = 0 \text{ for all } s \in \mathbf{C}_{n-1}.$$

  Moreover, we provide an algorithm for computing optimal single collapses that minimize the difference between the reconstructed and original sig-

nals. The performance of these algorithms is tested on synthetic data and compared to random collapses.

- Article 2.2, "Simplicial Neural Networks", proposes a generalization of graph convolutional neural networks (GCNNs) to data lying on simplicial complexes. We define a convolutional layer via the topological Fourier transform using the eigenvectors of the combinatorial Laplacian. Similarly to convolutional neural networks (CNNs) and GCNNs, this allows us to take into account the locality of data lying on a simplicial complex and achieve linear computational complexity. Finally, the performance of simplicial neural networks is evaluated on the task of imputing missing data on coauthorship complexes constructed from real-world data.

- Article 2.3, "A Notion of Harmonic Clustering in Simplicial Complexes" presents a novel clustering algorithm for simplicial complexes. Inspired by the classical spectral clustering algorithm, we provide an algorithm that utilizes the eigenvectors of the Laplacian to embed simplices of any dimension in a Euclidean space. Our method, unlike classical spectral clustering, involves the eigenvectors associated to the zero eigenvalues. Using several synthetic examples we show computationally that the algorithm produces clusters of simplices in a way that is sensitive to the homology of the complex.

While all the theory presented in the articles above can be defined for general based chain complexes–thus is immediately generalized to cell complexes and sheaves–in Article 2.2 and Article 2.3, we focus on simplicial complexes. This choice is dictated by applications. Our examples, coming from synthetic and real-world data, provide intuition on the theory using only simplicial complexes. On the other hand, the theory in Article 2.1 is developed for based chain complexes, a more general framework encompassing the chain complexes associated to cell and simplicial complexes.

Related Work Signal compression and reconstruction on graphs has been widely studied in different settings, ranging from the theory of sampling and reconstruction to autoencoders [23, 36, 37]. Recently, Barbarossa et al. [15, 22] presented a theory for sampling and reconstruction of bandlimited signals on simplicial complexes, as well as a method for topology inference based on the

signal supported on the simplices [38]. It is worth remarking that these theories restrict exclusively to signals on edges. Another article that is closely related to the signal reconstruction problem is that of Schaub et al. [30], where the authors use interpolation on edge-data lying on simplicial complexes to predict signals of unlabeled edges.

The theory developed in our Article 2.1 differs from previous work in two aspects. First, we deal with based chain complexes, a general algebraic framework that includes the chain complexes associated to cell and simplcial complexes. Second, our method for signal compression and reconstruction exploits Morse collapses. To the best of our knowledge, no other result merging these two topics has been published.

The theory presented in Article 2.2 is an extension of graph convolutional neural networks [27] to simplicial complexes. Other convolutional architectures based on the Laplacians [39, 40] were presented at the same conference as Article 2.2. In [39] they focus only on edge-data, whereas [40] presents the same framework for cell complexes. These articles generated a subsequent interest for neural networks for higher-structured data. Among these, Bodnar et al. [33, 32] defined a more general architecture for message passing on simplicial and cell complexes, allowing features on cells of different dimensions to interact. They proved that their architecture has more expressive power than simplicial neural networks, by showing that the number of linear regions of the functions represented by message passing is higher than in simplicial neural networks.

Interestingly, the work of Nanda et al. [41] encompasses aspects of both the harmonic clustering algorithm (Article 2.3) and simplcial neural networks [2]. Exploiting properties of a particular graph Laplacian, they define novel homologically-aware graph neural networks, whose input is built from simplicial complexes. Theoretical and computational results on the properties of the eigenvectors of the simplicial Laplacian to detect homological features are presented in [42], with promising results for the problem of the shortest homologous loop detection. Other types of embeddings based on Laplacian eigenvectors are presented in [30], where edge trajectories are projected and analyzed in the embedding space of the harmonics.

Future Perspectives We propose three applications of topological signal processing: to signal compression and reconstruction, to neural networks and to cell embedding. The theoretical results presented in Article 2.1 are framed in the gen-

eral algebraic setting of based chain complexes, which applies to a wide spectrum of contexts including such as simplicial complexes, cell complexes, and sheaves. We merge concepts from Hodge theory and discrete Morse theory, to gain insight into how signals on complexes are compressed and reconstructed when performing particular types of collapses. These results were motivated by the desire to define a meaningful, topologically-aware pooling layer for the simplicial neural networks presented in Article 2.2. For this reason, we envision that the next research steps in this direction will involve including such types of pooling in existing experiments and studying how collapses could be learned by the neural networks.

In the context of harmonic clustering, the results presented in [42], could be extended to the eigenvectors corresponding to non-zero eigenvalues, shedding light on the connection with spectral clustering.

Finally, from a computational perspective, a current limitation arises from the restriction to modelling data with simplicial or cell complexes. In fact, real-world data is not often naturally modelled as a cell complex with signals on its cells. Much effort should go in this direction, to find applications where higher-order relations can be efficiently represented by cells or simplicial complexes.

Author Contributions All papers were developed throughout regular discussion with the co-authors. The contributions of the author of this thesis can be found in all the aspects of the papers, from the theoretical results to the development of the code and experiments.

1.3 Collapsing in Directed Cubical Complexes

In this section we provide an overview of the articles presented in Chapter 3. We first introduce directed Euclidean cubical complexes and provide some motivating examples. We then provide intuition on how to reduce such complexes while preserving directional properties.

Directed Cubical Complexes Recently, topological spaces equipped with a notion of direction have received substantial attention in the field of algebraic topology [43, 44, 45]. In these spaces, direction usually comes from a notion of time incorporated in their structure. More precisely, the notion of direction

consists in specifying which paths in the space can be considered as "increasing" or "directed". The goal is then to study the topology of these spaces, and, in particular, to find suitable notions of algebraic invariants for the *space of directed paths*.

In applications directed spaces appear when modelling concurrent programs [44], dynamical systems [46] or motion planning [47], and multiple notions of direction have been proposed to deal with these different type of spaces. Motivated by verifying the execution of concurrent programs, in this thesis we focus on a model of directed spaces arising from this application. Concurrent programs are multiple processes running in parallel, the execution of which usually depends on the correct sequential schedule of its processes. For instance, it might happen that if two process run at the same time, the program will not be executed correctly.

Verifying combinatorially such executions is usually a hard problem, where the complexity can grow exponentially with the number of processes. A simplification of this problem can be obtained by modelling concurrent programs as directed Euclidean cubical complexes where each axis represents a sequence of actions a process completes in the program execution, and paths respecting the time directions represent executions of programs [44].

In this context, two executions of a program are equivalent if their corresponding directed paths are in the same equivalence class for a suitable notion of directed homotopy. Then, the verification of concurrent programs boils down to verifying only one execution from each connected component of the space of directed paths, see Figure 1.5.

Collapsing Directed Cubical Complexes In our work we address the question of how we can further simplify directed cubical complexes to easily compute equivalence classes of directed paths corresponding to equivalent executions of the concurrent program. The goal is to reduce the size of the directed cubical complex via suitable collapses that preserve desirable properties of the directed path spaces.

Although elementary collapses preserve the homotopy type of the underlying space, this type of collapsing in directed Euclidean cubical complexes may not preserve topological properties of spaces of directed paths. On the other hand, defining collapses directly on the spaces of directed paths is infeasible due to their continuous nature. This force us to utilize the notion of *past links*, a combinato-

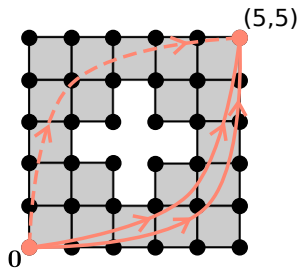


Figure 1.5: The Swiss-Flag Directed Cubical Complex. This complex models a concurrent program with two processes that have two shared resources with limited capacity. There are two distinct directed paths (solid and dashed) up to directed homotopy equivalence. The two distinct paths represent which process uses each shared resources first. See Article 3.1, Example 2.2 for a more detailed explanation.

rial local representations of cubical complexes at vertices, to define the so-called *link preserving directed collapses (LPDC)*. These types of collapses preserve the homotopy of past links, which is intrinsically related to the homotopy of the directed path spaces. Therefore, the natural question is *which exact properties of directed path spaces are preserved by LPDCs?* Answering such questions would allow us to simplify the verification of execution of concurrent programs in a reduced Euclidean cubical complex, see Figure 1.6.

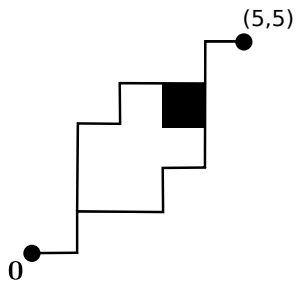


Figure 1.6: The collapses Swiss-Flag Directed Cubical Complex. This complex is the obtain by a sequence of LPDCs on the Swiss-Flag. It is easier to see in this figure that there are only two distinct paths from the initial to final vertex.

1.3.1 Contributions in Directed Cubical Complexes

Our contributions can be summarized as follows.

- Article 3.1, "Towards Directed Collapsability" shows how the topology of the past links associated to vertices in a directed Euclidean cubical complex is intrinsically related to spaces of directed paths. Specifically, we prove that the contractability and/or connectedness of past links of vertices in directed Euclidean cubical complexes with a minimum vertex implies that all spaces of directed paths with initial point are also contractible and/or connected. Additionally, we provide a partial converse in case of connectedness. These results motivate us to define link-preserving directed collapses (LPDC) as cubical collapses preserving the homotopy of past links. As an immediate consequence, we show that, if we start with a directed Euclidean cubical complex where all past links are connected and/or contractible, then any sequence of LPDCs results in a directed Euclidean cubical complex with connected and/or contractible directed path spaces.

- Article 3.2, "Combinatorial Conditions for Directed Collapsing" studies combinatorial conditions for collapsing unit cubes in a directed Euclidean cubical complex for which we can obtain an LPDC. Specifically, we show that there exists simple combinatorial conditions on the vertices of a collapse of unit cubes for being an LPDC. In practice, this condition allows us to develop an algorithm that can check at each step if a collapse is preserving properties of the directed space. Moreover, we further develop results of Article 3.1 that provide conditions on LPDCs to preserve the contractability and/or connectedness of directed path spaces. Finally, we also provide examples in which these properties are not preserved by LPDCs.

It is worth remarking that in both articles we restrict to directed topological spaces arising from Euclidean cubical complexes. These are however not the only type of directed spaces presented in the literature, see for example [43]. Our choice is dictated by applications to concurrent programs and allows us to borrow definitions and results from collapsing cubical complexes in the undirected sense. Furthermore, in article 3.1 we use the term directed collapses instead of link preserving directed collapsed, introduced in Article 3.2. Although they refer to the same mathematical object, we decided to introduce a more prescriptive terminology to underline that directed collapses preserve properties about past links and not necessarily about directed path spaces.

Related Work. Many authors studied how to reduce the size of cubical complexes via different types of collapses. For example, Vanessa Robins et al. [31], present an algorithm to reduce the size of images using discrete Morse collapses built from the grey-scale image values. In [48] they instead apply a coreduction algorithm to compute persistent homology of both cubical and simplicial datasets efficiently.

Our articles build upon the work on directed topological spaces presented in [49, 50, 44, 51, 43] and, differently from other work, our notion of collapsing takes into account the directionality of the space. In particular, under certain conditions, we can preserve properties of the spaces of directed paths. To the best of our knowledge, no other publications have explored this research direction.

Future Perspective. In Article 3.1 and 3.2, we prove that topological properties of past links of directed cubical complexes are inherently connected to the topology of their directed path spaces. We then define link preserving directed collapses (LPDCs), a novel notion of collapsing for directed cubical complexes that preserves topological properties of past links. Finally, we provide an easy combinatorial condition to determine when a collapse is an LPDC. In various settings LPDCs preserve spaces of directed paths, however, there exist examples of directed cubical complexes in which LPDCs do not preserve topological properties of the spaces of directed paths. A direction for future research could be to further investigate whether there are additional conditions on LPDCs to preserve spaces of directed paths between two given vertices. This would be another step towards developing algorithms that compress directed Euclidean cubical complexes and preserve directed topology. Another application of directed cubical complexes might include multiparameter persistent homology. The first question to address is how to find an appropriate representation of multi-paramenter persistence spaces via directed cubical complexes. Ideally, in this representation equivalence classes of directed paths would correspond to filtrations along which persistence modules are equivalent.

Author Contributions. The articles presented in Chapter 3 are the product of one of the working groups at the Women in Topology (WIT) workshop at MSRI in November 2017. Since then the group has been regularly meeting online, and all the results were developed throughout regular discussion with the co-authors. The author of this thesis has contributed in all the aspect of the papers from the

theoretical results to the writing of the manuscripts.

Chapter 2

Articles in Topological Signal Processing

2.1 Morse Theoretic Signal Compression and Reconstruction on Chain Complexes

(joint work with Celia Hacker and Kelly Maggs)
Submitted to the Journal of Applied and Computational Topology

Abstract

At the intersection of Topological Data Analysis (TDA) and machine learning, the field of cellular signal processing has advanced rapidly in recent years. In this context, each signal on the cells of a complex is processed using the combinatorial Laplacian and the resulting Hodge decomposition. Meanwhile, discrete Morse theory has been widely used to speed up computations by reducing the size of complexes while preserving their global topological properties.

In this paper, we provide an approach to signal compression and reconstruction on chain complexes that leverages the tools of algebraic discrete Morse theory. The main goal is to reduce and reconstruct a based chain complex together with a set of signals on its cells via deformation retracts, preserving as much as possible the global topological structure of both the complex and the signals.

We first prove that any deformation retract of real degree-wise finite-dimensional based chain complexes is equivalent to a Morse matching. We will then study how the signal changes under particular types of Morse matching, showing its reconstruction error is trivial on specific components of the Hodge decomposition. Furthermore, we provide an algorithm to compute Morse matchings with minimal reconstruction error.

1 Introduction

The analysis of signals supported on topological objects such as graphs or simplicial complexes is a fast-growing field combining techniques from topological data analysis, machine learning and signal processing [2, 32, 33]. The emerging field of simplicial and cellular signal processing falls within this paradigm [1, 34, 35], and here the combinatorial Laplacian Δ_n plays a pivotal role. In this context, a signal takes the form of a real-valued chain (or cochain) on a chain complex $(\mathbf{C}, \partial)$ endowed with a degree-wise inner product. In particular, the eigenvectors of Δ_n, called the Hodge basis, serve as a 'topological' Fourier basis to transform a signal into a topologically meaningful coordinate system [10, 35]. Additionally, the combinatorial Laplacian gives rise to the combinatorial Hodge decomposition [11]:

$$\mathbf{C}_n = \operatorname{Im} \partial_{n+1} \oplus \operatorname{Ker} \Delta_n \oplus \operatorname{Im} \partial_n^{\dagger},$$

the components of which each have their own topological interpretation [1] and respect the eigendecomposition of Δ_n.

The goal of the paper is to investigate signal compression and reconstruction over cell complexes by combining tools of Hodge theory and discrete Morse theory. We take an entirely algebraic approach to this problem, working at the level of degree-wise finite-dimensional *based* chain complexes endowed with inner products. The classical example is the chain complex of a cell complex equipped with its canonical cellular basis, but more general constructions such as cellular sheaves fit into this framework as well. This algebraic perspective not only gives us greater flexibility, but also helps to illuminate connections between Hodge theory and discrete Morse theory that occur only at the level of chain complexes.

Our approach to compressing and reconstructing signals over complexes involves deformation retracts of based chain complexes, which have the advantage of reducing the size of complexes while preserving their homology. A deformation retract of a chain complex $\mathbf{C}$ onto $\mathbf{D}$ consists of a pair of chain maps (Ψ, Φ)

$$\begin{tikzcd} \mathbf{D} \arrow[r, shift right, "\Phi"'] & \mathbf{C} \arrow[l, shift right, "\Psi"'] \arrow[loop right, "h"] \end{tikzcd}$$

such that $\Psi\Phi = \mathrm{Id}_{\mathbf{D}}$ and a chain homotopy $h : \mathbf{C} \to \mathbf{C}$ between $\Phi\Psi$ and $\mathrm{Id}_{\mathbf{C}}$. In this context, the map Ψ is used to compress the signal s onto the reduced complex $\mathbf{D}$, and Φ serves to reconstruct it back in $\mathbf{C}$. Thus, for every $s \in \mathbf{C}$ one can compute the difference $\Phi\Psi s - s$, called the *topological reconstruction error*, to understand and evaluate how compression and reconstruction changes the signal.

Among the many topological methods to reduce the size of complexes [36, 40], discrete Morse theory [12, 13] provides the perfect tool to efficiently generate such deformation retracts of chain complexes. This technique has already been used with great success in the compression of 3D images [40], persistent homology [29] and cellular sheaves [8]. In this paper we utilise Sköldberg's *algebraic* version of discrete Morse theory [37, 38]. It takes as input a based chain complex $\mathbf{C}$ and, by reducing its based structure with respect to a Morse matching M, returns a smaller, chain-equivalent complex $\mathbf{C}^M$. The first result presented in this article connects the Hodge decomposition of a complex with discrete Morse theory by defining a natural pairing in the Hodge basis. In particular, we show that *any* deformation retract (Ψ, Φ, h) of degree-wise finite-dimensional, based chain complexes of real inner product spaces can be obtained from a Morse matching over the Hodge basis of a certain sub-complex. This process, called the *Morsification* of (Ψ, Φ, h), is described in Theorem 3.7. In the second part of the paper, we study how the topological reconstruction error associated to a deformation retract (Ψ, Φ, h) is distributed amongst the three components of the Hodge decomposition. We define a class of deformation retracts (Ψ, Φ, h), called $(n, n-1)$-*free*, for which the topological reconstruction error has trivial (co)cycle reconstruction. Specifically, they are characterised by the following properties (Theorem 4.5).

1. (Cocycle Reconstruction) A signal $s \in \mathbf{C}_n$ and its reconstruction $\Phi\Psi s$ encode the same cocycle information:

$$\mathrm{Proj}_{\mathrm{Ker}\,\partial_{n+1}^{\dagger}}(\Phi\Psi s - s) = 0 \text{ for all } s \in \mathbf{C}_n.$$

2. (Cycle Reconstruction) A signal $s \in \mathbf{C}_{n-1}$ and the adjoint of the reconstruction $\Psi^{\dagger}\Phi^{\dagger}s$ have the same cycle information:

$$\mathrm{Proj}_{\mathrm{Ker}\,\partial_{n-1}}(\Psi^{\dagger}\Phi^{\dagger}s - s) = 0 \text{ for all } s \in \mathbf{C}_{n-1}.$$

Moreover, the Morsification concept defined above simplifies many of the proofs and allows them to be extended into a more general framework (Corollary 4.6).

Finally, we study how the topological reconstruction error of $(n, n-1)$-free deformation retracts can be minimized while maintaining (co)cycle reconstruction. We develop an iterative algorithm to find the retract (Ψ, Φ) that minimizes the norm of the topological reconstruction error for a given signal $s \in \mathbf{C}$. Our algorithm is inspired by the reduction pair algorithms in [8, 25, 29] and, like these algorithms, computes a single Morse matching at each step with the additional requirement of minimizing the norm. We show that its computational

complexity is linear when the complex is sparse, and discuss bounds on how well the iterative process approximates the optimal deformation retract. Finally, we show computationally that iterating single optimal collapses leads to topological reconstruction loss that is significantly lower than that arising from performing sequences of random collapses.

The paper is structured as follows. In Section 2, we present the necessary background in algebraic topology, discrete Hodge theory, and algebraic discrete Morse theory, giving the definitions and main results that will be used throughout the paper. Section 3 introduces the notion of Hodge matching, which allows us to prove that every deformation retract of a degree-wise finite-dimensional based chain complex $\mathbf{C}$ of real inner product spaces is equivalent to a Morse retraction (see Morsification Theorem 3.7). In Section 4 we investigate the interaction between deformation retracts and Hodge theory. The main results, Theorem 4.5 and Corollary 4.6, utilise the Morsification theorem to prove that $(n, n-1)$-free (sequential) Morse matchings preserve (co)cycles. Section 4.3 presents an additional result that explains how the reconstruction $\Phi\Psi s$ can be understood as a sparsification of the signal s (see Lemma 4.10). Finally, Section 5 is dedicated to presenting algorithms to minimize the topological reconstruction error in case of iterative single pairings (see Algorithms 1 and 2).

Related Work. Many articles incorporate topology into the loss or reconstruction error function [5, 14, 26, 30], however, these deal almost exclusive with point cloud data. At the same time discrete Morse theory has been used in conjunction with machine learning in [22] for image processing, but not in the context of reconstruction error optimisation.

The notion of taking duals (over $\mathbb{Z}$) of discrete Morse theoretic constructions is featured in [13]. There, the dual flow is over $\mathbb{Z}$, whereas we work with adjoint flow over $\mathbb{R}$, for which the orthogonality considerations are somewhat different, as discussed in Appendix A.2.

On the computational side, the articles [8, 24, 25, 29] involve algorithms to reduce chain complexes over arbitrary PIDs, including those of cellular sheaves but do not investigate the connection with the combinatorial Laplacian (or sheaf Laplacian). Our algorithms are based on the coreduction algorithms of [24, 25], with the additional requirement of a topological loss minimization.

To the best of our knowledge, the only other contemporary work that examines the link between the combinatorial Hodge decomposition and discrete Morse theory is [7], linking the coefficients of the characteristic equation of Δ_n to the n-dimensional paths in an acyclic partial matching.

2 Background

In this section, for the sake of completeness, we first recall some basic notions in algebraic topology. We refer the reader to [18] for a more detailed exposition. Then we present the main concepts of algebraic discrete Morse theory and finally, we discuss the foundations of discrete Hodge theory.

Algebraic Discrete Morse Theory. For two chain complexes $(\mathbf{C}, \partial)$ and $(\mathbf{D}, \partial')$, a pair of chain maps $\Psi : \mathbf{C} \to \mathbf{D}$ and $\Phi : \mathbf{D} \to \mathbf{C}$ are *chain equivalances* if $\Phi \circ \Psi : \mathbf{C} \to \mathbf{C}$ and $\Psi \circ \Phi : \mathbf{D} \to \mathbf{D}$ are chain homotopic to the identities on $\mathbf{C}$ and $\mathbf{D}$, respectively. Note that this implies that the maps induced on the homology modules by Φ and Ψ are isomorphisms. The chain equivalences Ψ and Φ form a *deformation retract* of the chain complexes $\mathbf{C}$ and $\mathbf{D}$ if $\Psi \circ \Phi$ is the identity map on $\mathbf{D}$. Deformation retracts will be often depicted as the following diagram.

$$\mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C} \circlearrowleft h$$

With a slight abuse of notation, we denote such deformation retract by the pair (Ψ, Φ) instead of (Ψ, Φ, h). Throughout the paper we will be working with the following notion of *based* chain complexes, as defined in [37], which in this context are chain complexes with a graded structure.

Definition 2.1. Let R be a commutative ring. A *based chain complex* of R-modules is a pair $(\mathbf{C}, I)$, where $\mathbf{C}$ is a chain complex of R-modules and $I = \{I_n\}_{n \in \mathbb{N}}$ is a set of mutually disjoint sets such that for all n and all $\alpha \in I_n$ there exist $C_\alpha \subseteq \mathbf{C}_n$ such that $\mathbf{C}_n = \bigoplus_{\alpha \in I_n} C_\alpha$.

Similarly, a based cochain complex is a cochain complex with an indexing set and graded decomposition as above. The components of the boundary operator ∂_n are denoted $\partial_{\beta,\alpha} : C_\alpha \to C_\beta$ for all $\alpha \in I_n$ and $\beta \in I_{n-1}$. We will refer to the elements of I_n as the *n-cells* of $(\mathbf{C}, I)$, and if $\partial_{\beta,\alpha} \neq 0$, we say that β is a *face* of α. If $\mathbf{C}$ is endowed with a degree-wise inner product, we say that I is an orthogonal base if $C_\alpha \perp C_\beta$ for all $\alpha \neq \beta \in I$.

Remark 2.2. In this paper, working with combinatorial Hodge theory means that, if not specified otherwise, we restrict our study to degree-wise finite-dimensional chain complexes over $\mathbb{R}$ with an inner product on each of the chain module $\mathbf{C}_n$.[1] Moreover, we will refer to degree-wise finite-dimensional based chain complexes as finite-type based chain complexes.

[1] We leave the original definition here to emphasise that algebraic discrete Morse theory works in more generality.

The following examples motivate such a choice of terminology for based chain complexes.

Example 2.3. In the special case where $(\mathbf{C}, I)$ is a finite-type based chain complex over $\mathbb{R}$ and $C_\alpha \cong \mathbb{R}$ for all $\alpha \in I$, we can think of I as a choice of basis, and each $\partial_{\beta,\alpha} \in \mathsf{Hom}(\mathbb{R}, \mathbb{R}) = \mathbb{R}$ as the (β, α)-entry in the boundary matrix multiplying on the left with respect to such a basis.

Example 2.4 (CW complexes)**.** The chain complex associated to a finite CW complex with a basis given by its cells is an example of a based chain complex (see [18] for a precise definition of CW complex). For two cells σ, τ in a CW complex $\mathcal{X}$, denote the degree of the attaching map of σ to τ by $[\sigma : \tau]$ and write $\sigma \triangleright \tau$ whenever they are incident[2]. For two incident cells, $\partial_{\tau,\sigma}$ is multiplication by $[\sigma : \tau]$.

Example 2.5 (Cellular Sheaves)**.** Here we present the main definitions for cellular sheaves, following the more detailed exposition of sheaf Laplacians found in [17]. A *cellular sheaf* of finite dimensional Hilbert spaces over a regular[3] CW complex $\mathcal{X}$ consists of an assignment of a vector space $\mathcal{F}(\sigma)$ to each cell $\sigma \in \mathcal{X}$ and a linear map $\mathcal{F}_{\tau \triangleleft \sigma} : \mathcal{F}(\tau) \to \mathcal{F}(\sigma)$ to each pair of incident cells $\sigma \triangleright \tau$. This defines a cochain complex,with

$$\mathbf{C}_n = \bigoplus_{\tau \in \mathcal{X}_n} \mathcal{F}(\tau),$$

where $\mathcal{X}_n$ denotes the set of n-cells of $\mathcal{X}$, and coboundary maps $\delta_n : \mathbf{C}_n \to \mathbf{C}_{n+1}$ defined component-wise by $\delta_{\sigma,\tau} = [\sigma : \tau]\mathcal{F}_{\tau \triangleleft \sigma} : C_\tau \to C_\sigma$.

Using the inner product on $\mathbf{C}_n$ induced by the inner product on each Hilbert space $\mathcal{F}(\sigma)$, one can define a boundary map $\partial_n : \mathbf{C}_{n+1} \to \mathbf{C}_n$ as the adjoint of the coboundary map δ_n. This chain complex is an example of a based chain complex, where the n-cells of the base correspond the n-cells of the underlying indexing complex.

Discrete Morse theory was originally introduced by Forman in [12] as a combinatorial version of classical Morse theory. Here we present its fundamental ideas in a purely algebraic setting, following the exposition in [37].

Definition 2.6. Let $(\mathbf{C}, I)$ be a finite-type based chain complex with base I. We denote by $\mathcal{G}(\mathbf{C}, I)$ the *graph of the complex*, which is the directed graph consisting of vertices I and edges $\alpha \to \beta$ whenever $\partial_{\beta,\alpha}$ is non-zero. When clear from the context we will denote $\mathcal{G}(\mathbf{C}, I)$ by $\mathcal{G}(\mathbf{C})$. For a subset of edges E of $\mathcal{G}(\mathbf{C})$, denote by $\mathcal{G}(\mathbf{C})^E$ the graph $\mathcal{G}(\mathbf{C})$ with the edges of E reversed.

[2]Here, incident means that the closure $\overline{\sigma}$ of σ contains τ.

[3]Regular here indicates that the attaching maps are homeomorphisms.

Using these notions we can define a Morse matching as follows.

Definition 2.7. An *(algebraic) Morse matching* M on a based complex $(\mathbf{C}, I)$ is a selection of edges $\alpha \to \beta$ in $\mathcal{G}(\mathbf{C})$ such that

1. each vertex in $\mathcal{G}(\mathbf{C})$ is adjacent to at most one edge in M;
2. for each edge $\alpha \to \beta$ in M, the map $\partial_{\beta,\alpha}$ is an isomorphism;
3. the relation on each I_n given by $\alpha \succ \beta$ whenever there exists a directed path from α to β in $\mathcal{G}(\mathbf{C})^M$ is a partial order.

For context, the third condition corresponds to acyclicity in the classical Morse matching definition, where directed paths akin to gradient flow-lines – which are non-periodic – in the smooth Morse theory setting [28].

When there is an edge $\alpha \to \beta$ in M, we say that α and β are *paired* in M, and refer to them as a $(\dim \alpha, \dim \alpha - 1)$*-pairing.* We use M^0 to denote the elements of I that are not paired by M, and refer to them as *critical cells* of the pairing. For a directed path $\gamma = \alpha, \sigma_1, \ldots, \sigma_k, \beta$ in the graph $\mathcal{G}(\mathbf{C}, I)^M$, the *index* $\mathcal{I}(\gamma)$ of γ is then defined as

$$\mathcal{I}(\gamma) = \epsilon_n \partial^{\epsilon_n}_{\beta,\sigma_n} \circ \ldots \circ \epsilon_1 \partial^{\epsilon_1}_{\sigma_2,\sigma_1} \circ \epsilon_0 \partial^{\epsilon_0}_{\sigma_1,\alpha} : C_\alpha \to C_\beta$$

where $\epsilon_i = -1$ if $\sigma_i \to \sigma_{i+1}$ is an element of M, and 1 otherwise. For any $\alpha, \beta \in I$, we define the *summed index* $\Gamma_{\alpha,\beta}$ to be

$$\Gamma_{\beta,\alpha} = \sum_{\gamma:\alpha\to\beta} \mathcal{I}(\gamma) : C_\alpha \to C_\beta,$$

the sum over all possible paths from α to β. In the case that there are no paths from $\alpha \to \beta$ then $\Gamma_{\beta,\alpha} = 0$.

The theorem below is the main theorem of algebraic Morse theory. While this theorem was originally proved in [38], here we state it in the form presented in [37] where it is proved as a corollary of the Homological Perturbation Lemma ([37], Theorem 1, [4,15]). This proof provides an explicit description of the chain homotopy $h : \mathbf{C} \to \mathbf{C}$ that witnesses the fact that the algebraic Morse reduction is a homotopy equivalence.

Theorem 2.8 (Sköldberg, [37])**.** *Let $(\mathbf{C}, I)$ be a based chain complex indexed by I, and M a Morse matching. For every $n \geq 0$ let*

$$\mathbf{C}^M_n = \bigoplus_{\alpha \in I_n \cap M^0} C_\alpha.$$

The diagram

$$\mathbf{C}^M \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C} \circlearrowleft h$$

where for $\alpha \in M^0 \cap I_n$ *and* $x \in C_\alpha$

$$\partial_{\mathbf{C}^M}(x) = \sum_{\beta \in M^0 \cap I_{n-1}} \Gamma_{\beta,\alpha}(x) \qquad \Phi(x) = \sum_{\beta \in I_n} \Gamma_{\beta,\alpha}(x)$$

and for $\alpha \in I_n$ *and* $x \in C_\alpha$

$$\Psi(x) = \sum_{\beta \in M^0 \cap I_{n-1}} \Gamma_{\beta,\alpha}(x) \qquad h(x) = \sum_{\beta \in I_{n+1}} \Gamma_{\beta,\alpha}(x)$$

is a deformation retract[4] *of chain complexes.*

We refer to the finite-type based chain complex $(\mathbf{C}^M, \partial_{C^M}, I \cap M^0)$ as the *Morse chain complex*. Moreover, we call this deformation retract of $\mathbf{C}$ into $\mathbf{C}^M$ the *Morse retraction* induced by M.

Example 2.9. Given a based chain complex $(\mathbf{C}, I)$ and a single $(n+1, n)$-pairing $M = (\alpha \to \beta)$, Lemma 2.8 can be used to get a simple closed form of the updated complex $(\mathbf{C}^M, \partial_{\mathbf{C}^M})$ as well as the chain equivalences. We write them explicitly here, and will refer to them throughout the paper.

- For every $\tau, \sigma \in M^0$, the Morse boundary operator is
$$\partial^{\mathbf{C}^M}_{\tau,\sigma} = \partial_{\tau,\sigma} - \partial_{\tau,\alpha}\partial^{-1}_{\beta,\alpha}\partial_{\beta,\sigma}.$$
- The map Ψ is the identity except at components C_α and C_β, where it is
$$\Psi^M_n\big|_{C_\beta} = \sum_{\tau \in I_n \setminus \alpha} -\partial_{\tau,\alpha}\partial^{-1}_{\beta,\alpha} \qquad \Psi^M_{n+1}\big|_{C_\alpha} = 0.$$
- The map Φ is the identity except at components C_η for each $\eta \in M^0 \cap I_{n+1}$, where it is
$$\Phi^M_{n+1}\big|_{\mathbf{C}_\eta} = Id_{\mathbf{C}_\eta} - \partial^{-1}_{\beta,\alpha}\partial_{\beta,\eta}.$$

Note that these equations are identical to those appearing in [25, 29] in the case that each component C_α is of dimension 1.

When $(\mathbf{C}, I)$ is a finite-type based chain complex of real inner product spaces, the adjoints of the maps in Theorem 2.8 play an important role in later sections. Their discrete Morse theoretic interpretation in terms of flow, however, hinges on the orthogonality of the base of $\mathbf{C}$ (see Appendix A.2). We will require the following basic result of linear algebra regarding adjoints throughout the paper.

[4] In fact the result is stronger. Specifically the maps form a *strong* deformation retract.

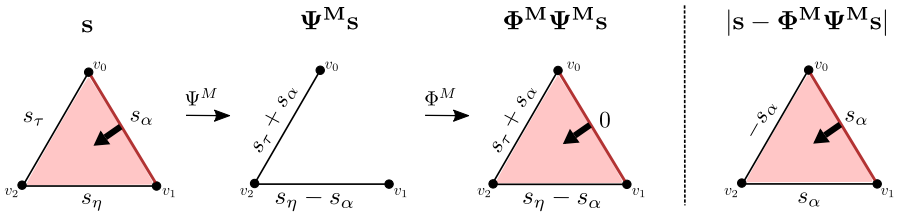


Figure 1: The chain maps Ψ and Φ operating on a signal $s \in \mathbf{C}_1$.

Lemma 2.10. *Let V be an finite dimensional inner product space and $W \subseteq V$ be a subspace. The adjoint of the inclusion map $i : W \to V$ is the orthogonal projection* $\mathrm{Proj}_W = i^\dagger$ *onto W.*

Example 2.11. Let $(\mathbf{C}, I)$ be the canonical based chain complex associated to the cell complex in Figure 1, (left). Following the standard convention of discrete Morse theory, we visually depict a pairing $\alpha \to \beta$ by an arrow running from the cell β to the cell α. We consider the single $(2,1)$-pairing $M = (\alpha, \beta)$, depicted by the black arrow. Figure 1 illustrates how the maps Ψ^M and Φ^M, made explicit Example 2.9, operate on $s \in \mathbf{C}_1$.

Remark 2.12. Motivated by the emerging field of cellular signal processing, we refer to elements $s \in \mathbf{C}_n$ as *signals* ([1, 35]).

In the next definition we introduce the concept of *sequential Morse matching*, an iterative sequence of Morse matchings. This type of matching, unlike a Morse matching, has a low computational cost to reduce the chain complex to a minimal number of critical cells. We discuss this in detail in Section 5.

Definition 2.13. A *sequential Morse matching* $\underline{M}$ on a based chain complex $(\mathbf{C}, I)$ is a finite sequence of Morse matchings, $M_{(1)}, \ldots, M_{(n)}$ and bases $I_1, \ldots, I_n$ such that the following conditions hold.

1. $M_{(1)}$ is a Morse matching on $(\mathbf{C}, I)$.
2. $M_{(j+1)}$ is a Morse matching in $(\mathbf{C}^{M_{(j)}}, I_j)$ for every $j \in \{1, \ldots, n-1\}$.
3. $\mathbf{C}^{M_{(j)}}$ is a based complex over $I_j \subseteq I_k$ for every $1 \leq j \leq k \leq n$.

We denote by $(\mathbf{C}^{\underline{M}}, \partial_{\mathbf{C}_{\underline{M}}})$ the based chain complex obtained from $\mathbf{C}$ by iteratively composing the Morse matchings in the sequential Morse matching $\underline{M}$, implying that $(\mathbf{C}^{\underline{M}}, \partial_{\mathbf{C}_{\underline{M}}}) = (\mathbf{C}^{M_{(n)}}, \partial_{\mathbf{C}_{M_{(n)}}})$. Note that in this case, the critical cells of each individual matching in $\underline{M}$ form a nested sequence $M^0_{(1)} \supseteq \cdots \supseteq M^0_{(n)}$. We denote by $\underline{M}^0$ the set of *critical cells* of the sequential Morse matching $\underline{M}$ and define it to be the set of critical cells in the last Morse matching in the sequence, namely $\underline{M}^0 = M^0_{(n)}$.

Combinatorial Laplacians. For a finite-type based chain complex $\mathbf{C}$ over $\mathbb{R}$ with boundary operator ∂ and inner products $\langle \cdot, \cdot \rangle_n$ on each $\mathbf{C}_n$, define $\partial_n^\dagger : \mathbf{C}_n \to \mathbf{C}_{n+1}$ as the adjoint of ∂_n, i.e., the map that satisfies $\langle \sigma, \partial_n^\dagger \tau \rangle_n = \langle \partial_n \sigma, \tau \rangle_{n-1}$ for all $\sigma \in \mathbf{C}_n$ and $\tau \in \mathbf{C}_{n-1}$. The adjoint maps form a cochain complex

$$\ldots \xleftarrow{\partial_{n+1}^\dagger} \mathbf{C}_n \xleftarrow{\partial_n^\dagger} \mathbf{C}_{n-1} \xleftarrow{\partial_n^\dagger} \ldots$$

where $(\partial^\dagger)^2 = 0$ follows from the adjoint relation.

Remark 2.14. If ∂_n is represented as a matrix in a given basis, and the inner products with respect to that basis are represented as $\langle \sigma, \tau \rangle_n = \sigma^T W_n \tau$ where each W_n is a positive-definite symmetric matrix, then the matrix form of the adjoint is given by $\partial_n^\dagger = (W_n^{-1}) \partial_n^T W_{n-1}$. Note that in our definition the inner product matrix W_n does not necessarily preserve the orthogonality of the standard cellular or simplicial basis in case we are working with cell complexes. In practice, other authors require W_n to be a diagonal matrix to keep the standard basis orthogonal [21]. In this way the coefficients of W_n can be thought as weights on the n-cells, see Appendix A.1

Definition 2.15. The *combinatorial Laplacian* is then defined as the sequence of operators

$$(\Delta_n = \partial_n^\dagger \partial_n + \partial_{n+1} \partial_{n+1}^\dagger : \mathbf{C}_n \longrightarrow \mathbf{C}_n)_{n \geq 0}.$$

For each n, the two summands can be further delineated into

1. the *n-th up-Laplacian* $\Delta_n^+ = \partial_{n+1} \partial_{n+1}^\dagger : \mathbf{C}_n \to \mathbf{C}_n$ and
2. the *n-th down-Laplacian* $\Delta_n^- = \partial_n^\dagger \partial_n : \mathbf{C}_n \to \mathbf{C}_n$.

The fundamental results concerning the combinatorial Laplacian were proved by Eckmann in the 1940s [11].

Theorem 2.16. *(Eckmann,* [11]*) If $\mathbf{C}$ is a finite-type based chain complex over $\mathbb{R}$ equipped with an inner product in each degree, then for all $n \geq 0$*

1. *$H_n(\mathbf{C}) \cong \operatorname{Ker} \Delta_n$, and*
2. *$\mathbf{C}_n$ admits an orthogonal decomposition*

$$\mathbf{C}_n \cong \operatorname{Im} \partial_{n+1} \oplus \operatorname{Ker} \Delta_n \oplus \operatorname{Im} \partial_n^\dagger. \tag{1}$$

The decomposition in the second point, called the *combinatorial Hodge decomposition*, is the finite-dimensional analogue of the Hodge decomposition for smooth differential forms. Two additional orthogonal decompositions associated with adjoints that we will use frequently are

$$\mathbf{C}_n = \operatorname{Ker} \partial_{n+1}^\dagger \oplus \operatorname{Im} \partial_{n+1} = \operatorname{Ker} \partial_n \oplus \operatorname{Im} \partial_n^\dagger. \tag{2}$$

Singular value decomposition. Let V, W be real finite-dimensional inner-product spaces. Let $f : V \to W$ be a linear map and $f^\dagger : W \to V$ its adjoint. The Spectral Theorem states that $f^\dagger f$ and $ff^\dagger$ have the same set of real eigenvalues Λ. Moreover, the singular value decomposition guarantees that there exist orthonormal bases $\mathcal{R}(f)$ and $\mathcal{L}(f)$ of V and W formed by eigenvectors of $f^\dagger f$ and $ff^\dagger$ such that for each non-zero $\lambda \in \Lambda$ there exists a unique $v \in \mathcal{R}(f)$ and a unique $w \in \mathcal{L}(f)$ such that

$$f(v) = \sqrt{\lambda} w.$$

We denote by $\mathcal{L}_+(f)$ and $\mathcal{R}_+(f)$ the subsets of $\mathcal{L}(f)$ and $\mathcal{R}(f)$ respectively corresponding to non-zero eigenvalues. Consider now $f = \partial_n : \mathbf{C}_n \to \mathbf{C}_{n-1}, n \geq 0$, the boundary operators associated to a based chain complex. Note that $\mathcal{L}_+(\partial_{n+1})$ and $\mathcal{R}_+(\partial_n)$, the sets of eigenvectors with positive eigenvalues of $\Delta_n^+ = \partial_{n+1}\partial_{n+1}^\dagger$ and $\Delta_n^- = \partial_n^\dagger \partial_n$, form orthonormal bases for $\operatorname{Im} \partial_{n+1}$ and $\operatorname{Im} \partial_n^\dagger$, respectively (by Equation (2)). In the next section we will see how these eigenvectors together with the Hodge decomposition will allow us to define a canonical Morse matching.

3 Morsification of Deformation Retracts

The aim of this section is to prove that every deformation retract of a finite-type based chain complex $\mathbf{C}$ over $\mathbb{R}$ equipped with degree-wise inner products is equivalent to a Morse retraction, with a canonical choice of basis. We first introduce the notion of the *Hodge matching* on $\mathbf{C}$, a Morse matching defined over the eigenbasis of the combinatorial up and down Laplacians Δ_n^+ and Δ_n^-. We can see the matching obtained by Hodge decomposition and the eigenvectors of Δ_n^+ and Δ_n^- as a *canonical* Morse matching.

3.1 Hodge Matchings

The following concept marries the discrete Morse theoretic notion of pairing to the pairing inherent to the eigendecomposition of Δ_n^+ and Δ_n^-, which is intrinsically connected to the Hodge decomposition of a finite real chain complex.

Definition 3.1 (Hodge basis)**.** Let $\mathbf{C}$ be a finite-type based chain complex over $\mathbb{R}$. A *Hodge basis* of $\mathbf{C}$ is the basis given by $I^\Delta = \{I_n^\Delta\}_{n\in\mathbb{N}}$, where

$$I_n^\Delta = \mathcal{L}_+(\partial_{n+1}) \bigcup \mathcal{R}_+(\partial_n) \bigcup \mathcal{B}(\operatorname{Ker} \Delta_n),$$

for some choice of bases $\mathcal{L}_+(\partial_{n+1}), \mathcal{R}_+(\partial_n)$ and $\mathcal{B}(\operatorname{Ker} \Delta_n)$.

Observe that in the definition above each set in I_n^Δ forms a basis for one of the components in the Hodge decomposition (see Equation 1). Our discussion on the singular value decomposition ensures that Hodge bases always exist.

Definition 3.2 (Hodge matching)**.** Let $\mathbf{C}$ be a finite-type based chain complex of real inner product spaces, and let I^Δ be a Hodge basis. The *Hodge matching* on $(\mathbf{C}, I^\Delta)$ is

$$M^\Delta := \bigcup_i \{v \in \mathcal{R}_+(\partial_i) \to w \in \mathcal{L}_+(\partial_i) \mid \partial_i v = \sigma w, \sigma \neq 0\}.$$

Lemma 3.3. *For a finite-type based chain complex $(\mathbf{C}, I^\Delta)$ of real inner product spaces and I^Δ be a Hodge basis. The Hodge matching M^Δ on $(\mathbf{C}, I^\Delta)$ is a Morse matching and satisfies*

1. $(M^\Delta)_n^0 = \operatorname{Ker} \Delta_n$*, where $\Delta : \mathbf{C} \to \mathbf{C}$ is the combinatorial Laplacian of $\mathbf{C}$ and*
2. $\partial^{M^\Delta} = 0$.

Proof. The description of orthonormal bases $\mathcal{L}(\partial_n)$ and $\mathcal{R}(\partial_n)$ described at the end Section 2 implies that each cell is adjacent to at most one other cell in $\mathcal{G}(\mathbf{C})^{M^\Delta}$. This means there are no nontrivial paths from any n-cell to any other n-cell for all n in $\mathcal{G}(\mathbf{C})^{M^\Delta}$. Thus, condition (3) in Definition 2.7 is trivially satisfied, and M^Δ indeed constitutes a Morse matching. By definition,

$$\operatorname{Im} \partial_{n+1} = \operatorname{span} \mathcal{L}_+(\partial_{n+1}) \text{ and } \operatorname{Im} \partial_n^\dagger = \operatorname{span} \mathcal{R}_+(\partial_n),$$

and all basis elements are paired. The remaining basis elements of $\mathbf{C}_n$ are critical, and constitute $(M^\Delta)_n^0 = \operatorname{Ker} \Delta_n$ for all n. Since there are no non-trivial paths, ∂^{M^Δ} agrees with the boundary operator ∂ of $\mathbf{C}$ on $\operatorname{Ker} \Delta$, which is indeed the zero map.

□

We call the data

$$\operatorname{Ker} \Delta \underset{\Phi^{M^\Delta}}{\overset{\Psi^{M^\Delta}}{\leftrightarrows}} \mathbf{C} \circlearrowleft h$$

the *Hodge retraction* of $(\mathbf{C}, I^\Delta)$. Noting that the maps Φ^{M^Δ}, Ψ^{M^Δ} are chain equivalences reproves Eckmann's result that $\operatorname{Ker} \Delta$ is isomorphic to the homology $\mathrm{H}(\mathbf{C})$ of the original complex.

The same proof also encompasses the case of cellular sheaves discussed in [17]. Note that here, a Hodge matching will be over a Hodge base I^Δ rather than the one specified by the cellular structure of the indexing complex. Nevertheless, since $\operatorname{Ker} \Delta$ does not depend on the choice of base, the result is the same.

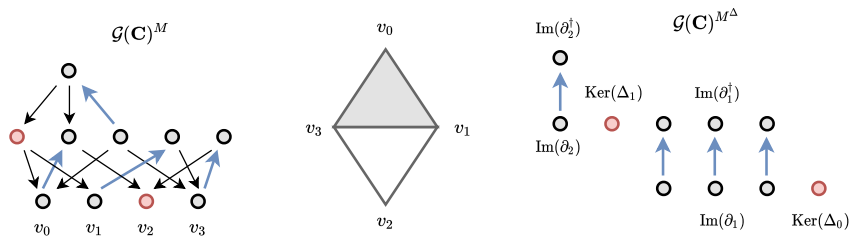


Figure 2: Two choices of bases and Morse matchings for the $\mathbb{R}$-valued chain complex of a simplicial complex. Edges in the Morse matchings are highlighted in blue and critical cells in red.

Example 3.4. In Figure 2 we depict two different choice of bases – the standard cellular basis and the Hodge basis – for the celllular chain complex of the pictured simplicial complex. Two matchings M and M^Δ are visualized through their corresponding Morse graphs $\mathcal{G}(\mathbf{C})^M$ and $\mathcal{G}(\mathbf{C})^{M^\Delta}$. The structure of the singular value decomposition of ∂ and ensuing Hodge matching 'straightens out' the connections in the matching graph, as pictured in Figure 2.

3.2 Morsification Theorem

In this section, we say that two deformation retracts

$$\mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C} \circlearrowleft h \qquad \text{and} \qquad \mathbf{D}' \underset{\Phi'}{\overset{\Psi'}{\leftrightarrows}} \mathbf{C}' \circlearrowleft h'$$

are *equivalent* if there exist isomorphisms of chain complexes, $f : \mathbf{D} \to \mathbf{D}'$ and $g : \mathbf{C} \to \mathbf{C}'$ such that the diagrams

$$\begin{tikzcd} \mathbf{D} \arrow[d, "f", "\cong"'] & \mathbf{C} \arrow[l, "\Psi"'] \arrow[d, "g", "\cong"'] \\ \mathbf{D}' & \mathbf{C}' \arrow[l, "\Psi'"] \end{tikzcd} \qquad \begin{tikzcd} \mathbf{D} \arrow[r, "\Phi"] \arrow[d, "f", "\cong"'] & \mathbf{C} \arrow[d, "g", "\cong"'] \\ \mathbf{D}' \arrow[r, "\Phi'"'] & \mathbf{C}' \end{tikzcd}$$

commute. Our goal is to show that any deformation retraction of finite-type chain complexes of real inner product spaces is isomorphic to a Morse retraction (Theorem 3.7).

In the special case that $\mathbf{C} = \mathbf{C}'$ and g is the identity, the commutativity of the diagrams above implies that

$$\Phi'\Psi' = \Phi f^{-1} f \Psi = \Phi\Psi. \tag{3}$$

Thus, to study the topological reconstruction error of a deformation retract, it is enough to study that of an equivalent deformation retract of the original

complex. Two equivalent deformation retracts over a shared domain $\mathbf{C}$ may have different homotopies, however, they are related by

$$\partial h + h\partial = 1 - \Phi\Psi = 1 - \Phi'\Psi' = \partial h' + h'\partial.$$

The main theorem of this section relies on the observation that deformation retracts share a number of characteristics with projection maps in linear algebra i.e. a linear endomorphism $P : V \to V$ of a vector space V satisfying $P^2 = P$. For any projection map, there exists a decomposition $V = \operatorname{Im} P \oplus \operatorname{Ker} P$ such that P can be decomposed as

$$P = 1_{\operatorname{Im} P} + 0 : \operatorname{Im} P \oplus \operatorname{Ker} P \to \operatorname{Im} P \oplus \operatorname{Ker} P.$$

The following lemma describes an analogous structure for real chain complexes, where a deformation retraction plays the role of a projection.

Lemma 3.5. *For any deformation retract*

$$\begin{tikzcd} \mathbf{D} \arrow[r, shift right, "\Phi"'] & \mathbf{C} \arrow[l, shift right, "\Psi"'] \arrow[loop right, "h"] \end{tikzcd}$$

of chain complexes over $\mathbb{R}$*,*

$$\mathbf{C} = \operatorname{Ker} \Psi \oplus \operatorname{Im} \Phi. \tag{4}$$

as chain complexes.

Proof. The deformation retraction condition $\Psi\Phi = \mathrm{Id}_{\mathbf{D}}$ implies that

$$(\Phi_n \Psi_n)^2 = \Phi_n \Psi_n \Phi_n \Psi_n = \Phi_n \Psi_n,$$

i.e., each component $\Phi_n \Psi_n$ of $\Phi\Psi$ is a projection operator. Thus there is a splitting of vector spaces

$$\mathbf{C}_n = \operatorname{Ker}(\Phi\Psi)_n \oplus \operatorname{Im}(\Phi\Psi)_n$$

for each n. Since $\Phi\Psi$ is a chain map, the decomposition above commutes with the boundary operator of $\mathbf{C}$, whence

$$\mathbf{C} = \operatorname{Ker} \Phi\Psi \oplus \operatorname{Im} \Phi\Psi$$

as chain complexes. Lastly, Ψ is surjective and Φ is injective since $\Psi\Phi = \mathrm{Id}_{\mathbf{D}}$, implying that $\operatorname{Im} \Phi\Psi = \operatorname{Im} \Phi$ and $\operatorname{Ker} \Phi\Psi = \operatorname{Ker} \Psi$. $\square$

The decomposition defined in Equation 4 has an interesting interpretation when passing to homology: all of the non-trivial homology of $\mathbf{C}$ arises from the $\operatorname{Im} \Phi$ component of the decomposition. One way to think of this decomposition is that $\operatorname{Ker} \Psi$ is the component of $\mathbf{C}$ that is discarded by the deformation retraction, whereas $\operatorname{Im} \Phi$ is preserved.

Lemma 3.6. *Under the hypotheses of Lemma 3.5*

1. $\mathrm{H}(\mathbf{C}) \cong \mathrm{H}(\operatorname{Im}\Phi)$, *and*

2. $\mathrm{H}(\operatorname{Ker}\Psi) = 0$.

Proof. Since Ψ is a weak equivalence, $\mathrm{H}(\mathbf{C}) \cong \mathrm{H}(\mathbf{D})$. Since $\Psi\Phi = \mathrm{Id}_{\mathbf{D}}$, Φ is injective, so $\mathbf{D} \xrightarrow{\Phi} \operatorname{Im}\Phi$ is an isomorphism of chain complexes, proving point (1). Since $\mathbf{C} = \operatorname{Ker}\Psi \oplus \operatorname{Im}\Phi$ by Equation 4, it follows that $\mathrm{H}(\operatorname{Ker}\Psi) = 0$. $\square$

Theorem 3.7 (Morsification)**.** *Any deformation retract*

$$\mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C} \circlearrowleft h$$

of finite-type chain complexes of real inner product spaces is equivalent to a Morse retraction $(\Psi^{\mathcal{M}}, \Phi^{\mathcal{M}})$ *over* $\mathbf{C}$ *.*

Notation 3.8. We refer to the pairing $\mathcal{M}$ in this theorem as the *Morsification* of a deformation retract.

Proof. Define a pairing $\mathcal{M} = \widetilde{M}^{\Delta} \sqcup \widehat{M}$ on $\mathbf{C}$ as the union of a Hodge pairing $\widetilde{M}^{\Delta}$ on $\operatorname{Ker}\Psi$ (which is given the subspace inner product) and the trivial pairing $\widehat{M}$ on $\operatorname{Im}\Phi$. We previously showed that $\mathbf{C} = \operatorname{Ker}\Psi \oplus \operatorname{Im}\Phi$ and $\mathrm{H}(\mathbf{C}) = \mathrm{H}(\operatorname{Im}\Phi)$, implying that $\mathrm{H}(\operatorname{Ker}\Psi) = 0$. Consequently, all the basis elements in $\operatorname{Ker}\Psi$ are paired by the Hodge pairing, and further, the Morse retraction maps

$$\mathrm{H}(\operatorname{Ker}\Psi) \cong 0 \underset{\Phi^{\widetilde{M}^{\Delta}}}{\overset{\Psi^{\widetilde{M}^{\Delta}}}{\leftrightarrows}} \operatorname{Ker}\Psi$$

defined by the matching $\widetilde{M}^{\Delta}$ are trivial.

On the other hand, since $\widehat{M}$ is the trivial pairing, the entirety of $\operatorname{Im}\Phi$ is critical in the pairing $\mathcal{M}$. Further, the Morse boundary operator $\partial^{\widehat{M}}$ is the same as the boundary operator on $\mathbf{C}$, implying $\mathbf{C}^{M} = \operatorname{Im}\Phi$ and that the maps

$$\mathbf{C}^{M} \cong \operatorname{Im}\Phi \underset{\Phi^{\widehat{M}}}{\overset{\Psi^{\widehat{M}}}{\leftrightarrows}} \operatorname{Im}\Phi$$

are identities. We conclude that $\Phi^{\mathcal{M}}\Psi^{\mathcal{M}} = i_{\operatorname{Im}\Phi} \circ \pi_{\operatorname{Im}\Phi}$, where $i_{\operatorname{Im}\Phi} : \operatorname{Im}_{\Phi} \hookrightarrow \mathbf{C}$ is the inclusion.

Now we show that this is equivalent to the original deformation retract. To do so, first note that $\Phi : D \to \operatorname{Im}\Phi$ is an isomorphism. We then need to show that the following diagram

$$\begin{tikzcd} & \mathbf{C} \arrow[dl, "\Psi"'] \arrow[dr, "\Psi^{\mathcal{M}}"] & \\ \mathbf{D} \arrow[rr, "\cong", "\Phi"'] & & \operatorname{Im}\Phi \end{tikzcd}$$

commutes. For any $(s, \Phi(t)) \in \mathbf{C} = \operatorname{Ker}\Psi \oplus \operatorname{Im}\Phi$, we have

$$\Phi\Psi(s, \Phi(t)) = (\Phi\Psi(s), \Phi\Psi\Phi(t)) = (0, \Phi(t)) = i \circ \pi_{\operatorname{Im}\Phi}(s, \Phi(t)) = \Phi^{\mathcal{M}}\Psi^{\mathcal{M}}(s, \Phi(t))$$

as required. Finally, to see that

$$\begin{tikzcd} & \mathbf{C} & \\ \mathbf{D} \arrow[ur, "\Phi"] \arrow[rr, "\cong", "\Phi"'] & & \operatorname{Im}\Phi \arrow[ul, "\Phi^{\mathcal{M}}"'] \end{tikzcd}$$

commutes simply note that $\Phi^{\mathcal{M}}$ is the inclusion map. □

Remark 3.9. When the original deformation retract comes from a Morse matching, the subspace $\operatorname{Im}\Phi = \operatorname{Im}\Phi\Psi = \operatorname{Ker}(1 - \Phi\Psi)$ is the space of *flow-invariant chains* used by Forman in his foundational articles [12, 13]. The difference here is that these chains are linear combinations of genuine critical cells, albeit for a Morse matching in a new base.

It is not difficult to see that the Morsification of a deformation retract is unique up to a choice of bases in the eigenspaces of Δ^+ and Δ^-, and that each such choice produces equivalent deformation retracts. Combining Theorem 3.7 with Equation 3, we get a simple expression for the reconstruction error of a deformation retract in terms of the paired cells in its Morsification.

Corollary 3.10. *Any deformation retract*

$$\begin{tikzcd} \mathbf{D} \arrow[r, shift right, "\Phi"'] & \mathbf{C} \arrow[l, shift right, "\Psi"'] \arrow[loop right, "h"] \end{tikzcd}$$

of finite-type chain complexes of real inner product spaces and Morsification $\mathcal{M}$

$$1 - \Psi\Phi = \sum_{\alpha \in I^{\mathcal{M}} \setminus \mathcal{M}^0} i_\alpha \circ \pi_\alpha$$

Proof. By Equation 3 and Theorem 3.7, we have

$$1 - \Phi\Psi = 1 - i_{\operatorname{Im}\Phi} \circ \pi_{\operatorname{Im}\Phi} = i_{\operatorname{Ker}\Phi\Psi} \circ \pi_{\operatorname{Ker}\Phi\Psi} = \sum \alpha \in I^{\mathcal{M}} \setminus \mathcal{M}^0 i_\alpha \circ \pi_\alpha$$

which proves the statement, noting that the paired cells in $\mathcal{M}$ span $\operatorname{Ker}\Psi$. □

In the case that the deformation retract arises from a Morse matching on a based complex, the Morsification construction will most likely alter the base. However, the number of pairings and critical cells in each dimension are related, as described in the following proposition.

Notation 3.11. For a sequential Morse matching $\underline{M}$ on a based chain complex $(\mathbf{C}, I)$, let $\underline{M}_n^-$ and $\underline{M}_n^+$ denote the elements of I_n that are the union of all start and endpoints respectively of edges in each of the matchings $M_{(i),n}$ for all i. This means that

$$I_n = \underline{M}_n^- \sqcup \underline{M}_n^0 \sqcup \underline{M}_n^+.$$

Further, let

$$|\underline{M}_n^*| = \sum_{\alpha \in \underline{M}_n^*} \dim C_\alpha$$

where $* \in \{+, -, 0\}$, and the subscript n refers to the dimension of the cells.

Proposition 3.12. *Let $\underline{M}$ be a sequential Morse matching on a finite-type based chain complex $(\mathbf{C}, I)$ of real inner product spaces and $\mathcal{M}$ be its Morsification. Then*

$$|\mathcal{M}_n^*| = |\underline{M}_n^*|$$

for $* \in \{+, -, 0\}$, in each dimension $n \geq 0$.

Proof. By Theorem 3.7 we know that $\mathbf{C}^{\underline{M}} \cong \mathbf{C}^{\mathcal{M}}$, implying that the dimensions spanned by critical cells

$$\left|\underline{M}_n^0\right| = \dim \mathbf{C}_n^{\underline{M}} = \dim \mathbf{C}_n^{\mathcal{M}} = \left|\mathcal{M}_n^0\right|$$

are equal for all n. This implies that

$$\left|\underline{M}_n^+\right| + \left|\underline{M}_n^-\right| = \dim \mathbf{C}_n - \dim \mathbf{C}_n^{\underline{M}} = \left|\mathcal{M}_n^+\right| + \left|\mathcal{M}_n^-\right| \tag{5}$$

where we have used the identity $\dim \mathbf{C}_n = \left|\underline{M}_n^+\right| + \left|\underline{M}_n^-\right| + \left|\underline{M}_n^0\right|$.

Since the chain complex is concentrated in non-negative degrees, cells in dimension 0 can be paired only with elements in dimension 1, implying that $\left|\underline{M}_0^-\right| = \left|\mathcal{M}_0^-\right| = 0$. Combining this with Equation 5 we conclude that $\left|\underline{M}_0^+\right| = \left|\mathcal{M}_0^+\right|$. The bijection between cells paired up in dimension i with those paired down in dimension $i+1$ then implies that

$$\left|\underline{M}_1^-\right| = \left|\underline{M}_0^+\right| = \left|\mathcal{M}_0^+\right| = \left|\mathcal{M}_1^-\right|,$$

and, again using Equation 5, that $\left|\underline{M}_1^+\right| = \left|\underline{M}_1^+\right|$. By inductively performing this procedure, we prove the result for all n as required. $\square$

It is not difficult to see that two equivalent Morse retractions of $\mathbf{C}$ must have the same Morsification. Thus the above proposition then implies that when two sequential Morse retractions $\underline{M}$ and $\underline{M}'$ of a complex $\mathbf{C}$ under two different bases I and I' are equivalent, there are equalities between the number of dimensions paired up $\left|\underline{M}_n^+\right| = \left|\underline{M}_n'^+\right|$ and down $\left|\underline{M}_n^-\right| = \left|\underline{M}_n^-\right|$ for all n. Notably, this occurs independently of the bases I and I'.

4 (Co)cycle Preservation and Sparsification

Discrete Morse theory aims to reduce the dimension of a chain complex while preserving its homology. Meanwhile, for combinatorial Hodge theory, understanding the effect of deformation on the components of the Hodge decomposition is of equal importance. However, because of the 'adjointness' inherent in the Hodge decomposition, neither chain or cochain maps between two complexes usually respect the grading of the Hodge decomposition.

Here, we define a different notion of preservation by examining the effect of applying either $\Phi\Psi$ or $\Psi^\dagger\Phi^\dagger$ to an element $s \in \mathbf{C}_n$. For a pair of chain maps

$$\mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C}$$

we define the *topological reconstruction error* at $s \in \mathbf{C}$ as $\Phi\Psi s - s \in \mathbf{C}$. The goal of this section is to examine the projection of $\Phi\Psi s - s$ on the different components of the Hodge decomposition. In particular, we describe which components of the signal are preserved and discarded by $\Phi\Psi$ when the deformation retract arises from a $(n, n-1)$-free Morse matching, a special type of (sequential) Morse matchings described in the next section. Further, we show that for such matchings the reconstruction $\Phi\Psi s$ (or $\Psi^\dagger\Phi^\dagger s$) is supported only on the critical cells, and serves to sparsify the data on the original complex while preserving the (co)cycle information.

4.1 $(n, n-1)$-free Matchings

Definition 4.1. A Morse matching M is said to be $(n, n-1)$*-free* if $|M_n^-| = 0$.

An equivalent condition is that $|M_{n-1}^+| = 0$. Put simply, a Morse matching is $(n, n-1)$-free if no n-cells are paired with $(n-1)$-cells. In what follows, the mantra is that preservation of (co)cycle information in dimension $n-1$ (or n) is equivalent to absence of such pairings. We define an $(n, n-1)$*-free sequential Morse matching* $\underline{M} = (M_{(1)}, \ldots, M_{(k)})$ to be a sequential Morse matching where all M_i are $(n, n-1)$-free Morse matchings.

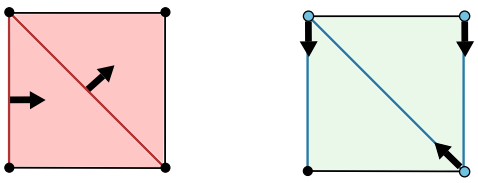


Figure 3: Two Morse matchings – the left is $(1,0)$-free and the right is $(2,1)$-free.

Example 4.2. Figure 3 shows a $(1,0)$-free and a $(2,1)$-free matching. The matchings are computed on the cellular chain complex of the depicted cell complex, based with the standard cellular basis. We visually depicted the pairings in the macthings by black arrows. Note that being $(n, n-1)$-free does not necessarily prohibit all n or $(n-1)$-cells from appearing in the matching, implying that $(n, n-1)$-free matchings can still lead to dimension reduction of both $\mathbf{C}_n$ and $\mathbf{C}_{n-1}$.

Example 4.3. If $\mathbf{C}$ is finite-type chain complex of real inner product spaces such that $\partial_n = 0$, then the Hodge matching M^Δ is $(n, n-1)$-free for some choice of Hodge basis I^Δ.

The corollary below, which follows immediately from Proposition 3.12, shows that the property of being $(n, n-1)$-free is not an artifact of our choice of basis. Namely, if two Morse matchings are equivalent, then either they are both $(n, n-1)$-free or neither is.

Corollary 4.4. *A sequential Morse matching $\underline{M}$ on a based chain complex $(\mathbf{C}, I)$ is $(n, n-1)$-free if and only if its Morsification $\mathcal{M}$ is $(n, n-1)$-free.*

4.2 (Co)cycle Preservation for $(n, n-1)$-free Matchings

The following reconstruction theorem shows that both the topological reconstruction error of the deformation retract and its adjoint are supported on non-kernel components of the Hodge decomposition.

Theorem 4.5 (Reconstruction)**.** *Suppose that M is a Morse matching on a finite-type based chain complex $(\mathbf{C}, I)$ of real inner product spaces. Let*

$$\begin{tikzcd} \mathbf{C}^M \arrow[r, shift right, "\Phi"'] & \mathbf{C} \arrow[l, shift right, "\Psi"'] \arrow[loop right, "h"] \end{tikzcd}$$

be the deformation retract given by Theorem 2.8. Then

1. *for all $s \in \mathbf{C}_n$,*

$$\mathrm{Proj}_{\mathrm{Ker}\,\partial^\dagger_{n+1}}(\Phi\Psi s - s) = 0, \text{ and}$$

2. for all $s \in \mathbf{C}_{n-1}$,

$$\mathrm{Proj}_{\mathrm{Ker}\,\partial_{n-1}}(\Psi^\dagger\Phi^\dagger s - s) = 0$$

if and only if M is a $(n, n-1)$-free matching.

Proof. We first prove that if M is a $(n, n-1)$-free matching, then conditions (1) and (2) hold. If $M_n^- = \emptyset$, then there are no paths in $\mathcal{G}(\mathbf{C})^M$ from an $(n-1)$-cell to an n-cell. Theorem 2.8 then implies that $h_{n-1}(x) = 0$ for all $\alpha \in I_{n-1}$ and $x \in C_\alpha$, whence

$$(\Phi\Psi - 1)_n = \partial_{n+1}h_n + h_{n-1}\partial_n = \partial_{n+1}h_n. \tag{6}$$

The first claim now follows from the orthogonal decomposition

$$\mathbf{C}_n = \mathrm{Ker}\,\partial_{n+1}^\dagger \oplus \mathrm{Im}\,\partial_{n+1}.$$

The argument above also shows that $h_{n-1}^\dagger = 0$, since the adjoint of the zero map is the zero map. By taking the adjoint of Equation 6 one dimension lower, it then follows that

$$(\Psi^\dagger\Phi^\dagger - 1)_{n-1} = (\Phi\Psi - 1)_{n-1}^\dagger = \partial_{n-1}^\dagger h_{n-2}^\dagger + h_{n-1}^\dagger\partial_n^\dagger = \partial_{n-1}^\dagger h_{n-2}^\dagger.$$

The second claim is then a consequence of the orthogonal decomposition $\mathbf{C}_{n-1} = \mathrm{Ker}\,\partial_{n-1} \oplus \mathrm{Im}\,\partial_{n-1}^\dagger$.

For the other direction we will prove the contrapositive statement. It is sufficient to show that if the Morse matching is not $(n, n-1)$-free, then there exists $s \in \mathbf{C}_n$ such that

$$\mathrm{Proj}_{\mathrm{Ker}\,\partial_{n+1}^\dagger}(\Phi\Psi s - s) \neq 0.$$

The Morse matching M is $(n, n-1)$-free if and only if its Morsification $\mathcal{M}$ is $(n, n-1)$-free (Corollary 4.4) and, further, $1 - \Phi^M\Psi^M = 1 - \Phi^{\mathcal{M}}\Psi^{\mathcal{M}}$ (Equation 3). Therefore, it is sufficient to prove the contrapositive statement for the Morsification.

Since the Morsification is not $(n, n-1)$-free, there exists an $(n, n-1)$-pair $\alpha \to \beta$ such that $\partial_{\beta,\alpha}$ is an isomorphism. Recall that by 3.10, we have that $(\Phi^{\mathcal{M}}\Psi^{\mathcal{M}} - 1)x = x$ for $x \in C_\alpha$. The orthogonal decomposition of $\mathbf{C}_n$ implies that

$$x = \mathrm{Proj}_{\mathrm{Ker}\,\partial_n}x + \mathrm{Proj}_{\mathrm{Im}\,\partial_n^\dagger}x.$$

Applying ∂_n and using the fact that $\partial_n(x) \neq 0$, we obtain

$$0 \neq \partial_n\mathrm{Proj}_{\mathrm{Ker}\,\partial_n}x + \partial_n\mathrm{Proj}_{\mathrm{Im}\,{}_n^\dagger}x = \partial_n\mathrm{Proj}_{\mathrm{Im}\,\partial_n^\dagger}x.$$

Since $\operatorname{Im}\partial_n^\dagger \subseteq \operatorname{Ker}\partial_{n+1}^\dagger$, this implies that

$$0 \neq \operatorname{Proj}_{\operatorname{Ker}\partial_{n+1}^\dagger} x = \operatorname{Proj}_{\operatorname{Ker}\partial_{n+1}^\dagger}(\Phi^{\mathcal{M}}\Psi^{\mathcal{M}} - 1)x = \operatorname{Proj}_{\operatorname{Ker}\partial_{n+1}^\dagger}(\Phi^{M}\Psi^{M} - 1)x,$$

which proves our statement.

□

The utility of the theorem above is that an $(n, n-1)$-free matching M reduces the dimension of $\mathbf{C}_n$, while perfectly preserving the n-cocycles of a signal $s \in \mathbf{C}_n$ under the reconstruction $\Phi_n\Psi_n$. The extent of this reduction depends on the $(n+1, n)$-pairs in M. Indeed, the direct sum of the components $\bigoplus_{\alpha \in M_n^+} C_\alpha$ of n-cells in such pairs is isomorphic to the subspace $\operatorname{Ker}\Psi_n$ discarded by the deformation retract. One way to see this is using the fact that the Morsification has the same pair structure as the sequential Morse matching, and the Morsification $\Phi^{\mathcal{M}}$ is zero on non-critical cells.

If, on the other hand, one is interested in preserving the cycle information of a signal $s \in \mathbf{C}_{n-1}$, then one can use the adjoint maps $\Phi^\dagger\Psi^\dagger$ to perform a similar procedure. Namely, an $(n, n-1)$-free matching M will perfectly preserve the $(n-1)$-cycle part of s under the reconstruction $\Psi_{n-1}^\dagger\Phi_{n-1}^\dagger$. Analogously to the dual case, the extent of reduction depends on the $(n-1, n-2)$-pairings, where the subspace $\bigoplus_{\alpha \in M_{n-1}^-} C_\alpha$ is isomorphic to the discarded subspace $\operatorname{Ker}\Phi_{n-1}^\dagger$.

Using Morsification, we can extend the (co)cycle reconstruction theorem to $(n, n-1)$-free sequential Morse matchings.

Corollary 4.6. *Let $\underline{M}$ be a sequential Morse matching on a based chain complex $(\mathbf{C}, I)$. Then the (co)cycle preservation conditions* (1) *and* (2) *of Theorem 4.5 hold if and only if $\underline{M}$ is $(n, n-1)$-free.*

Proof. By Corollary 4.4 we know that $\underline{M}$ is $(n, n-1)$-free if and only if its Morsification $\mathcal{M}$ is $(n, n-1)$-free. Further, we know that

$$1 - \Phi^{\underline{M}}\Psi^{\underline{M}} = 1 - \Phi^{\mathcal{M}}\Psi^{\mathcal{M}}$$

by Equation 3. Then the statement follows by applying Theorem 4.5 to $\mathbf{C}$ and $\mathcal{M}$. □

One may wonder whether there is a proof by induction that follows directly from Theorem 4.5. The problem with using induction is that each chain complex in the sequential Morse matching has a different Hodge decomposition, and that the maps between them do not necessarily respect the grading. So Theorem 4.5 implies the (co)cycle preservation conditions will be satisfied between the i-th and $(i+1)$-th chain complexes but not necessarily between $\mathbf{C}$ and $\mathbf{C}^{\mathcal{M}}$.

In the general case of deformation retracts that do not arise from a Morse matching, combining Theorem 4.5 and Corollary 4.4 yields the following.

Corollary 4.7. *Let* (Φ, Ψ) *be a deformation retract of based finite-type chain complexes* $(\mathbf{C}, I)$ *and* $(\mathbf{D}, I')$ *of real inner product spaces. Then the (co)cycle preservation conditions* (1) *and* (2) *of Theorem 4.5 hold if and only if the Morsification* $\mathcal{M}$ *associated to* (Φ, Ψ) *is* $(n, n-1)$*-free.*

4.3 Sparsification for $(n, n-1)$-free Matchings

In the previous section, we showed how a signal's projection onto each Hodge component is related to that of its reconstruction. In addition, one would like to know how the reconstructed signal sits in the complex with respect to the base on which the Morse matching is constructed.

In this section we will show that, for a $(n, n-1)$-free (sequential) Morse matching, the image of $\Phi_n \Psi_n$ is supported only on the critical cells M_n^0 of I_n. Intuitively, applying $\Phi_n \Psi_n$ can be thought of as a form of sparsification which preserves one of either cycles or cocycles (Theorem 4.5).

Lemma 4.8. *Let* M *be an* $(n, n-1)$*-free matching of an orthogonally based finite-type chain complex* $(\mathbf{C}, I)$ *of real inner product spaces. Then*

1. $\Phi_n : \mathbf{C}_n^M \to \mathbf{C}$ *and*
2. $\Psi_{n-1}^\dagger : \mathbf{C}_{n-1}^M \to \mathbf{C}$

are subspace inclusions and, thus, isometries.

Proof. By Theorem 2.8

$$\Phi_n = \sum_{\alpha \in M_n^0} \sum_{\beta \in I_n} \Gamma_{\beta,\alpha}.$$

A path in $\mathcal{G}(\mathbf{C})^M$ starting at an n-dimensional critical cell must first step down a dimension. Since M is $(n, n-1)$-free, it cannot return to dimension n. This shows that the only paths starting at critical cells in dimension n are trivial and hence

$$\Phi_n(x) = \sum_{\beta \in I_n} \Gamma_{\beta,\alpha}(x) = x$$

for all $x \in \mathbf{C}_\alpha$, $\alpha \in M_n^0$.

For point (2), recall that

$$\Psi_{n-1} = \sum_{\alpha \in M_{n-1}^0} \sum_{\beta \in I_{n-1}} \Gamma_{\alpha,\beta}.$$

When $\alpha \in \underline{M}^0_{n-1}$, all non-trivial paths in $\mathcal{G}(\mathbf{C})^M$ from $\beta \in I_{n-1}$ to α must pass through dimension n. However, this is impossible since M is $(n, n-1)$-free, implying all paths out of critical cells in dimension $(n-1)$ to cells in dimension $(n-1)$ are trivial and $\sum_{\beta \in I_{n-1}} \Gamma_{\alpha,\beta} = \pi_\alpha$. This yields

$$\Psi_{n-1} = \sum_{\alpha \in M^0_{n-1}} \pi_\alpha = \pi_{\mathbf{C}^M}.$$

According to Lemma 2.10, the inclusion $i : \mathbf{C}^M \to \mathbf{C}$ is the adjoint of the *orthogonal* projection $\mathrm{Proj}_{\mathbf{C}^M}$, and is not necessarily the same as the categorical projection $\pi_{\mathbf{C}^M}$. However, the condition that the base I is orthogonal, implies that $\mathbf{C}^M$ is is indeed orthogonal to $\mathbf{C}/\mathbf{C}^M$, and that $\Psi^\dagger_{n-1}$ is the inclusion map $i : \mathbf{C}^M \hookrightarrow \mathbf{C}$ as required. □

Remark 4.9. The condition that the base is orthogonal is also important for having a discrete Morse theoretic interpretation of the adjoint in terms of backwards flow within the Morse graph $\mathcal{G}(\mathbf{C})^M$. We explain this perspective in detail in Appendix A.2.

Given that the composition of a sequence of inclusions of sub-spaces is again an inclusion, Lemma 4.8 holds equally well for $(n, n-1)$-free *sequential* Morse matchings.

Corollary 4.10 (Sparsification)**.** *Let $\underline{M}$ be an $(n, n-1)$-free sequential Morse matching of an orthogonally based chain complex $(\mathbf{C}, I)$. Then*

1. $$\Phi^{\underline{M}}_n \Psi^{\underline{M}}_n(s) \in \bigoplus_{\alpha \in M^0 \cap I_n} C_\alpha \text{ for all } s \in \mathbf{C}_n$$

2. $$\Psi^{\underline{M}\dagger}_{n-1} \Phi^{\underline{M}\dagger}_{n-1}(s) = \bigoplus_{\beta \in M^0 \cap I_{n-1}} C_\beta \text{ for all } s \in \mathbf{C}_{n-1}.$$

Proof. By definition we know that

$$\Psi^{\underline{M}}_n(s) \in \bigoplus_{\alpha \in \underline{M}^0 \cap I_n} C_\alpha = \mathbf{C}^{\underline{M}}_n \quad \text{and} \quad \Phi^{\underline{M}\dagger}{}_{n-1}(s) \in \bigoplus_{\beta \in \underline{M}^0 \cap I_{n-1}} C_\beta = \mathbf{C}^{\underline{M}}_{n-1}.$$

The result then follows from Lemma 4.8, which implies that both $\Phi^{\underline{M}}_n$ and $\Psi^{\underline{M}\dagger}_{n-1}$ are compositions of subspace inclusions. □

Example 4.11. In this example we consider the based chain complex $\mathbf{C}$ associated to the cell complex $\mathcal{X}$ in Figure 4-A. We work with the standard basis

generated by the n-cells and the standard boundary operator ∂_*. The signal $s \in \mathbf{C}_1$ is obtained by randomly sampling from $[0, 1]$. We consider the $(1, 0)$-free matching M in Figure 4-C, where there are two 1-cells are paired with two 2-cells, denoted by the arrows. All the other cells are critical.

In Figure 4-A we show how the signal s is transformed by the maps Φ^M and Ψ^M induced by the $(1, 0)$-free matching M. The absolute value of the reconstruction error, $|s - \Phi^M \Psi^M|$ is shown in Figure 5-B. As proved in Theorem 4.5, we observe in Figure 5-D that the reconstructed signal $\Phi^M \Psi^M s$ is perfectly preserved on $\operatorname{Ker} \partial_1 = \operatorname{Ker} \Delta_1 \oplus \operatorname{Im} \partial_1^\dagger$, and all changes in the reconstructed signal are contained in $\operatorname{Im} \partial_2$. Note that $\Phi_1^M \Psi_1^M s$ is supported only on the critical 1-cells as proved in Lemmas 4.10 and 4.8.

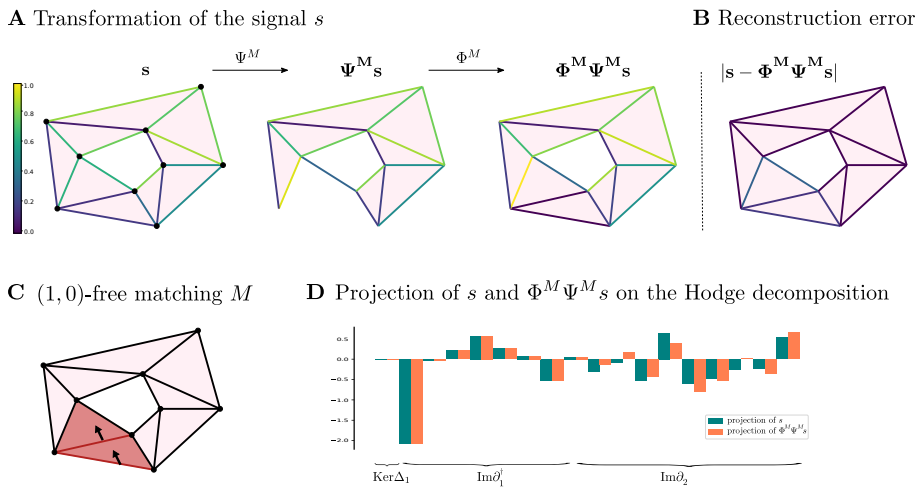


Figure 4: The life-cycle and reconstruction error of a signal $s \in \mathbf{C}$ in the standard basis of a simplicial complex under the maps associated to a Morse matching.

5 Algorithms and Experiments

The goal of this section is to reduce a based complex $(\mathbf{C}, I)$ together with a signal $s \in \mathbf{C}$ (or set of signals $\mathcal{S} \subset \mathbf{C}$) via a sequential Morse matching while trying to minimize the norm of the topological reconstruction error.

We propose the following procedure to iterativly reduce a based chain complex $(\mathbf{C}, I)$ with signal s via a sequential Morse matching. The method is inspired by the classical reduction pair algorithm described in [24, 25] but differs in the optimization step in (1).

1. If $\partial \neq 0$, select a single pairing $\alpha \to \beta$ in $(\mathbf{C}, \partial)$ minimizing $\|s - \Phi\Psi s\|$.
2. Reduce $\mathbf{C}$ to $\mathbf{C}^M$ and repeat with $\mathbf{C} = \mathbf{C}^M$ and $\partial = \partial^{C^M}$.

Note that this procedure differs as well from that of Nanda et al. which, in the context of both persistent homology [29] and cellular sheaves [8], requires an actual Morse matching. The details of the algorithm are provided in Section 5.1 (see Algorithm 1 and Algorithm 2), where we also show that their computational complexity is linear in the number of $(n+1)$-cells. In Section 5.2 we discuss the behaviour of the norm of the topological reconstruction error when performing this type of iterated reduction. In Section 5.3 we prove that such an algorithm converges to a based chain complex with the minimal number of critical cells. Finally, in Section 5.4 we provide experiments on synthetic data.

Remark 5.1. Since in most of the applications $\dim C_\alpha = 1$ for all $\alpha \in I$, we will work with this assumption throughout the following sections. Thus, without loss of generality, we will refer to the elements of I_n as a basis of $\mathbf{C}_n$ and denote $\partial_{\beta,\alpha} = [\alpha : \beta]$ (see Example 2.3 for more details).

5.1 Algorithms for Optimal (sequential) Morse Matchings

For a pair of chain maps

$$\mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C}$$

between based chain complex with inner product on each $\mathbf{C}_n$ and $\mathbf{D}_n$, and a signal $s \in \mathbf{C}_n$, define the *topological loss* of the maps (Φ, Ψ) over s to be the norm of the topological reconstruction error

$$\mathcal{L}_s(\Psi, \Phi) = \langle s - \Phi\Psi s, s - \Phi\Psi s\rangle_{\mathbf{C}_n}^{1/2} = \|s - \Phi\Psi s\|_{\mathbf{C}_n}. \tag{7}$$

For a finite subset $\mathcal{S} \subset \mathbf{C}_n$, the loss is defined to be the sum

$$\mathcal{L}_{\mathcal{S}}(\Psi, \Phi) = \sum_{s \in \mathcal{S}} \mathcal{L}_s(\Psi, \Phi)$$

of the individual losses. The loss of a single collapse can be given a closed form by using Theorem 2.8, in the case of a deformation retract associated to a Morse matching.

Specifically, suppose we have a single $(n+1, n)$-pairing $\alpha \to \beta$. Theorem 2.8 implies that the homotopy h maps β to $-\frac{1}{[\alpha:\beta]}\alpha$ and is zero elsewhere. For a signal $s \in \mathbf{C}_n$, using the equations developed in Example 2.9, we have

$$\mathcal{L}_s(\Psi, \Phi) = \|(1 - \Phi\Psi)s\|_{\mathbf{C}_n} = \left\|\partial_n h_n s\right\|_{\mathbf{C}_n} = \left\| \frac{s_\beta}{[\alpha : \beta]} \cdot \partial_{n+1}(\alpha) \right\|_{\mathbf{C}_n} \tag{8}$$

where s_β is the component of s on basis element β. Similarly, for a signal $s \in \mathbf{C}_{n+1}$ we have a *dual* topological loss

$$\mathcal{L}_s(\Phi^\dagger, \Psi^\dagger) = \|(1 - \Psi^\dagger\Phi^\dagger)s\|_{n+1} = \|\partial_{n+1}^\dagger h_n^\dagger s\|_n \tag{9}$$

If I is an orthogonal basis for $\mathbf{C}$, Theorem A.2 implies that we can write this loss as

$$\mathcal{L}_s(\Phi^\dagger, \Psi^\dagger) = \left\| s_\alpha \frac{\partial^\dagger_{n+1}(\beta)}{[\alpha : \beta]} \right\|_{\mathbf{C}_{n+1}}$$

Note that to write a compact form for Equation (7), in case M is not a single Morse matching, one needs to sum over all possible non-trivial paths in Theorem 2.8. Therefore finding the matching M minimizing this norm would be computationally expensive, if not infeasible. On the other hand, it is not hard to find the single $(n+1, n)$-pairing $\alpha \to \beta$ minimizing the topological loss in Equation (8). Therefore, as a first approach towards finding an approximate solution of the problem, we begin by studying optimal matchings by restricting to iterated single pairings.

Remark 5.2. Naturally, one can ask the same questions about finding the optimal pairing minimizing the topological loss for $\Psi^\dagger\Phi^\dagger s - s$. Given the duality of the problem, we will present algorithms and experiments only for $\Phi\Psi s - s$. The algorithms and computations for the dual topological loss can be found by dualizing the chain and boundary maps.

Given a finite-type based chain complex $(\mathbf{C}, I)$ of real inner product spaces and a signal s on the n-cells, our goal is now to find the the $(n+1, n)$-pairing $\alpha \to \beta$ minimizing the topological loss in Equation (8). Computing the minimum and its arguments for a single pair boils down to storing for each $(n+1)$-cell τ in the basis the face σ where the quantity

$$\frac{|s_\sigma|}{|[\tau : \sigma]|} \left\| \partial_{n+1}\tau \right\|_n$$

is minimal, and choosing among all the $(n+1)$-cells the one realizing the minimum of $\mathcal{L}_s$.

Example 5.3. Consider the based chain complex associated to a simplicial complex $\mathcal{X}$ with basis induced by its cells and ∂_* the standard boundary operator. Let s be a signal on the n-cells. The minimum of the reconstruction loss $\mathcal{L}_s$ in Equation (8) is then realized on the n-cell β, where $|s_\beta|$ is minimum, paired with any of its cofaces α. Note that the minimum and its argument might not be unique.

Following the idea above, Algorithm 1 returns a single $(n+1, n)$-pairing $\alpha \to \beta$ that minimizes the topological loss for a given based chain complex $(\mathbf{C}, I)$ and signal s.

Algorithm 1 Perform a single optimal pairing

Input A based chain complex $\mathbf{C}$ with basis I, a signal on $\mathbf{C}_n$, ∂_{n+1}, the non-zero $n+1$-boundary.

Output A a single $(n+1, n)$-pairing $\alpha \to \beta$ which minimize the topological loss.

function OptimalPairing($\mathbf{C}$, I, `signal`, ∂_{n+1})

  for each $n+1$-cell τ in I_{n+1} **do**

    `OptCol`$[\tau]$=0 ▷ `OptCol` keeps track of the face which realizes the optimal collapse on τ

  end for

  for each $n+1$-cell τ in I_{n+1} **do**

    `ValOptCol`$[\tau]$=∞ ▷ `ValOptCol` keeps track of the value of the optimal collapse on τ

  end for

  for each $n+1$-cell τ in I_{n+1} **do**

    for each face ξ of τ in $\mathcal{F}_\tau$ **do**

      $x \leftarrow \dfrac{|\texttt{signal}[\xi]|}{|[\tau : \xi]|} \|\partial_{n+1}(\tau)\|_{\mathbf{C}_n}$

      `ValOptCol`$[\tau]$←minimum(x, `ValOptCol`$[\tau]$)

    end for

    $\sigma \leftarrow \text{random}\left(\underset{\xi \in \mathcal{F}_\tau}{\arg\min}\left(\dfrac{|\texttt{signal}[\xi]|}{|[\tau : \xi]|} \|\partial_{n+1}(\tau)\|_{\mathbf{C}_n} = \texttt{ValOptCol}[\tau]\right)\right)$

    ▷ σ is randomly chosen among the faces of τ which have minimal reconstruction loss

    `OptCol`$[\tau] \leftarrow \sigma$

  end for

  `TotalMin`←minimum(`ValOptCol`) ▷ The value `TotalMin` is the minimum reconstruction loss.

  $\alpha \leftarrow \text{random}(\underset{\tau \in I_{n+1}}{\arg\min}(\texttt{ValOptCol}=\texttt{TotalMin}))$ ▷ The $n+1$ cell α to collapse is randomly chosen among the $n+1$ cells where the reconstruction loss is minimal.

  $D \leftarrow$ `OptCol`$[\alpha]$ ▷ The n cell β to collapse is the face of τ obtaining minimal reconstruction loss.

  return (α, β)

end function

The computational complexity of Algorithm 1 is $O(pc^2) + O(p)$, where $p = \dim \mathbf{C}_{n+1}$ and $c = \max_{\tau \in I_{n+1}} |\partial_{n+1}\tau|$. The first term follows from the fact that we need to iterate through all the $(n+1)$-cells and their faces, computing the minimum of lists of size at most c. The second summand follows from the fact that the final step of the algorithm requires computations of the minimum of a list of size at most c. Since the first summand dominates the second one, the computational complexity of Algorithm 1 is $O(pc^2)$. We assume that in most of the computations we are dealing with sparse based chain complexes, i. e. based chain complexes in which the number of n-cells in the boundary of an $(n+1)$-cell is at most a constant $c \ll p$. In this case the computational complexity of Algorithm 1 is $O(p)$.

In practice, one would like to further reduce the size of a based chain complex. In Algorithm 2 we provide a way to perform a sequence of single optimal collapses. For a based chain complex $\mathbf{C}$ and a signal s, the algorithm computes at each iteration a single optimal pairing (α, β) and it updates $(\mathbf{C}, \partial)$ to $(\mathbf{C}^M, \partial_{C_M})$ and the signal s to $\Psi^M s$.

Algorithm 2 Perform k single optimal pairings

Input A based chain complex $\mathbf{C}$ with basis I, a signal on $\mathbf{C}_n$, ∂_{n+1} the non-zero $(n+1)$-boundary and parameter k of the number of single optimal collapses to perform.

OutputA based chain complex $\mathbf{C}^M$ with basis $I^M \subseteq I$ and its boundary ∂_{C_M} obtained by iteratively computing k optimal pairings starting from $\mathbf{C}$.

```
function k-OptimalPairings(C, I, signal, ∂_{n+1}, k)
    i ← 1
    while i ≤ k do
        (α, β) ← OptimalUpCollapse(C, I, signal, ∂_{n+1})
        (C, ∂, I) ← (C^M, ∂_{C_M}, I_M)
        signal ← Ψ(signal)
        i ← i + 1
    end while
return C , ∂
end function
```

In fact, Algorithm 2 consists of the classical reduction pair algorithm proposed in [24, 25] with the additional step of the loss minimization. If applied only to a $(n, n-1)$-free sequential Morse matching, Algorithm 2 will converge to a based chain complex with given dimensions, as we prove in Proposition 5.9. Otherwise, if applied to cells of every size, it allows us to reduce a chain complex up to a minimal number of critical n-cells, as proved in [25]. We state again this result in Section 5.3. At the same time, the algorithm constructs a $(n, n-1)$-free sequential Morse matching, therefore the original signal is perfectly reconstructed on part of the Hodge decomposition, as proved in Theorem 4.5.

Finally, a further justification for the choice of this iterative algorithm, is that the loss on the original complex is bounded by the sum of the losses in the iterative step. We further discuss this in the next section.

5.2 Conditional Loss

The computational advantages outlined above are dictated by the fact that Algorithm 2 iteratively searches for optimal pairings. One important detail to understand is then how the loss function interacts with such iterated reductions. For a diagram of chain maps

$$\mathbf{E} \underset{\Phi'}{\overset{\Psi'}{\leftrightarrows}} \mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C}$$

and $s \in \mathbf{C}_n$, define the conditional loss to be

$$\mathcal{L}_s(\Psi', \Phi' \mid \Psi, \Phi) = \mathcal{L}_{\Psi(s)}(\Psi', \Phi') = \|\Psi s - \Phi'\Phi'\Psi s\|_{\mathbf{D}_n}.$$

In practice, we will generate a sequential Morse matching by taking a series of collapses and optimising the conditional loss at each step.

Lemma 5.4. *Let $C, D,$ and E be inner product spaces and suppose we have a diagram of linear maps*

$$E \underset{\phi'}{\overset{\psi'}{\leftrightarrows}} D \underset{\phi}{\overset{\psi}{\leftrightarrows}} C$$

where ϕ is an isometry. Then for all $s \in C$ we have

$$\|(1 - \phi\phi'\psi'\psi)s\|_C \leq \|(1 - \phi\psi)s\|_C + \|(1 - \phi'\psi')\psi(s)\|_D.$$

Proof. Using the triangle inequality and the fact that ϕ is an isometry, we have

$$\begin{aligned}
\|(1 - \phi\phi'\psi'\psi)s\|_C &= \|(1 - \phi\psi)s + \phi(1 - \phi'\psi')\psi(s)\|_C \\
&\leq \|(1 - \phi\psi)s\|_C + \|\phi(1 - \phi'\psi')\psi(s)\|_C \\
&= \|(1 - \phi\psi)s\|_C + \|(1 - \phi'\psi')\psi(s)\|_D
\end{aligned}$$

as required. □

The following corollary justifies the approach of minimizing the conditional loss at each step. It states that the loss on the original complex will be bounded by the sum of the conditional losses. Note that the same result and proof also work for the adjoint case where $s \in \mathbf{C}_{n-1}$, as long as the complex is orthogonally based.

Corollary 5.5. *Suppose we have a diagram of chain maps*

$$\mathbf{E} \underset{\Phi'}{\overset{\Psi'}{\leftrightarrows}} \mathbf{D} \underset{\Phi}{\overset{\Psi}{\leftrightarrows}} \mathbf{C}$$

where each step arises from an $(n, n-1)$-free Morse matching. Then for all $s \in \mathbf{C}_n$

$$\mathcal{L}_s(\Psi'\Psi, \Phi\Phi') \leq \mathcal{L}_s(\Psi, \Psi) + \mathcal{L}_s(\Psi', \Phi' \mid \Psi, \Phi)$$

Proof. In the Sparsification Lemma 4.10, we showed that taking $(n, n-1)$-matchings implied that Φ_n, Φ'_n are isometries. The result then follows from applying the lemma above. □

5.3 Reduction Pairings and Convergence

The following proposition ensures that the reduction pair algorithm proposed in [25], which is the foundation of Algorithm 2, converges in a finite (and predetermined) number of steps to the homology of $\mathbf{C}$. This advantage of being able to maximally reduce a based complex is in contrast with the well-studied NP-hard problem [23] of finding Morse matchings. In this section, we will prove an analogous result for $(n, n-1)$-free matchings.

Theorem 5.6 (Kaczynski et al. [25])**.** *Let $(\mathbf{C}, I)$ be a finite-type based chain complex over $\mathbb{R}$, where $\dim C_\alpha = 1$ for all $\alpha \in I$. The iteration of the following procedure*

1. *If $\partial \neq 0$, select a single pairing $\alpha \to \beta$ in $(\mathbf{C}, \partial)$.*
2. *Reduce $\mathbf{C}$ to $\mathbf{C}^M$ and repeat with $\mathbf{C} = \mathbf{C}^M$ and $\partial = \partial_{C_M}$.*

converges to the complex $\mathrm{H}(\mathbf{C})$ with $\partial = 0$ after

$$N = \frac{1}{2} \sum_n (\dim \mathbf{C}_n - \dim \mathrm{H}_n(\mathbf{C}))$$

steps.

To prove a similar result for $(n, n-1)$-free matchings, we first prove two lemmas describing how the dimensions of the summands in the Hodge decomposition of $\mathbf{C}^M$ relate to those of $\mathbf{C}$ when M is a single pairing.

Lemma 5.7. *Let $M = (\alpha \to \beta)$ be an $(n+1, n)$-pairing of a based complex $(\mathbf{C}, I)$. Then*

$$\operatorname{Im} \partial_n^M = \operatorname{Im} \partial_n$$

Proof. Since no $(n-1)$-cells are deleted by M, $\mathbf{C}_{n-1} = \mathbf{C}^M_{n-1}$. The formulas in the background section in Example 2.9 show that $\partial^M_n = \partial_n|_{\mathbf{C}^M_n}$, implying that $\operatorname{Im}\partial^M_n + \partial_n(C_\beta) = \operatorname{Im}\partial_n$. To prove the statement it then suffices to show that $\partial_n(C_\beta)$ is contained in $\operatorname{Im}\partial^M_n$. Using $\partial_n\partial_{n+1} = 0$ and the fact that $\partial_{\alpha,\beta}$ is an isomorphism, we then have that

$$0 = \partial_n(\partial_{n+1}(C_\alpha)) = \partial_n(\partial_{\beta,\alpha}(C_\alpha) + \sum_{\tau\in I_n\setminus\beta} \partial_{\tau,\alpha}(C_\alpha))$$
$$\Rightarrow \partial_n(C_\beta) = -\partial_n(\sum_{\tau\in I_n\setminus\beta} \partial_{\tau,\alpha}(C_\alpha)) \subseteq \operatorname{Im}\partial^M_n.$$

which proves the result. □

Note that while the images of both ∂^M_n and ∂_n agree, the eigendecomposition of their correspondent up- and down-Laplacians may not be related in a straightforward way. In other words, the combinatorial Laplacian eigenbases for $\mathbf{C}_{n-1}$ and $\mathbf{C}^M_{n-1}$ can be rather different, even though the corresponding summands of their Hodge decompositions have the same dimensions.

Lemma 5.8. *Let $M = (\alpha \to \beta)$ be an $(n+1, n)$-pairing of a finite-type based complex $(\mathbf{C}, I)$ of real inner product spaces. Then*

$$\dim \operatorname{Im}(\partial^M_i)^\dagger = \dim \operatorname{Im}\partial^M_i = \begin{cases} \dim \operatorname{Im}\partial_i - \dim C_\beta & i = n+1 \\ \dim \operatorname{Im}\partial_i & \textit{else} \end{cases} \tag{10}$$

Proof. The left equality is a basic property of adjoints. For the right equality, note that (1) $\mathbf{C} \simeq \mathbf{C}^M$ implies $\dim \operatorname{Ker}\Delta^M_i = \dim \operatorname{Ker}\Delta_i$ for all i and (2)Lemma 5.7 implies that $\dim \operatorname{Im}(\partial^M_n)^\dagger = \dim \operatorname{Im}\partial^\dagger_n$. Together these imply that

$$\dim \mathbf{C}_n - \dim \mathbf{C}^M_n = \dim \operatorname{Im}\partial_{n+1} - \dim \operatorname{Im}\partial^M_{n+1} = \dim C_\beta.$$

Equivalently, this says that $\dim \operatorname{Im}\partial^\dagger_{n+1} - \dim \operatorname{Im}(\partial^M_{n+1})^\dagger = \dim C_\alpha$, and now all of the change in dimension from $\mathbf{C}$ to $\mathbf{C}^M$ has been accounted for. □

We can now state the convergence theorem for the $(n, n-1)$-sequential Morse matchings over $\mathbb{R}$ in Algorithm 2. Along with homology, $\dim \operatorname{Im}\partial_n$ and $\dim \operatorname{Im}\partial^\dagger_n$ provide a (strict) upper bound on how many pairings we can make in an $(n, n-1)$-free sequential Morse matching.

Proposition 5.9 (Convergence)**.** *Let $(\mathbf{C}, I)$ be a finite-type based chain complex over $\mathbb{R}$ with inner products. Then Algorithm 2 for $(n, n-1)$-free Morse matchings*

converges to a based chain complex $\mathbf{D}$ *such that*

$$\mathbf{D}_i \cong \begin{cases} \mathrm{H}(\mathbf{C}_i) \oplus \operatorname{Im} \partial_i^\dagger & i = n \\ \mathrm{H}(\mathbf{C}_i) \oplus \operatorname{Im} \partial_{i+1} & i = n-1 \\ \mathrm{H}(\mathbf{C}_i) & \textit{else} \end{cases}$$

where $\partial_i^{\mathbf{D}} = 0$ *for all* $i \neq n$.

Proof. Given the conditions on the basis assumed at the beginning of the section, $\partial_{\alpha,\beta}$ is an isomorphism if and only if it is a multiplication by a non-zero element of $\mathbb{R}$. Hence, $\partial_i = 0$ if and only if we are not able to make any more $(i, i-1)$-pairings, implying the process must converge to some complex $\mathbf{D}$ with $\partial_i^{\mathbf{D}} = 0$ for all $i \neq n$. Since $\mathbf{D}$ is weakly equivalent to $\mathbf{C}$, this proves that $\mathbf{D}_i = \mathrm{H}_i(\mathbf{D}) = \mathrm{H}_i(\mathbf{C})$ for all $i \notin \{n, n-1\}$.

By Lemma 5.8, each $(n+1, n)$-pairing reduces the dimension of $\operatorname{Im} \partial_{n+1}$ by 1, and each $(n-1, n-2)$-pairing reduces the dimension of $\operatorname{Im} \partial_{n-1}^\dagger$ by 1. One can iterate the process of either $(n+1, n)$-pairing or $(n-1, n-2)$-pairing, until $\dim \operatorname{Im} \partial_{n+1} = 0$ or $\dim \operatorname{Im} \partial_{n-1}^\dagger = 0$ respectively. Thus, the isomorphism in the Lemma follows from this itarative process and from the Hodge decomposition of $\mathbf{D}_i$. □

5.4 Experiments

In this section we provide examples of how algorithms 1 and 2 can be applied to compress and reconstruct signals on synthetic complexes. Moreover, we show computationally that the topological reconstruction loss of a sequence of optimal pairings given by algorithm 2 is significantly lower than the loss when performing sequences of random collapses (see Figure 5 and Figure 8). Our main goal is to provide a proof of concept for the theoretical results and algorithms of this article rather than an exhaustive selection of experiments. The code for the experiments can be found in [39].

Example 5.10. In this example we consider the cell complex $\mathcal{X}$ in Figure 5.A-left, constructed as the alpha complex of points sampled uniformly at random in the cube $[0,1] \times [0,1]$. We work with the basis given the cells of $\mathcal{X}$ and the standard boundary operator ∂. The signal s on the 1-cells is given by the height function on the 1-cells. The example illustrates a $(1,0)$-free sequential Morse matching $\underline{M}$ obtained by iterating Algorithm 2 for $k = 120$. Note that the optimal matchings correspond to 1-cells where the signal is lower (see Figure 5.A-center). This can be explained by Remark 5.3 and the fact that Equation (8) favors collapsing cells with lower signal even when $\mathcal{X}$ is not a simplicial complex.

The absolute value of the reconstruction error after the sequential Morse matching $\underline{M}$ is shown in Figure 5.B. As expected from Equation (8), the error is mainly concentrated on the 1-cells that are in the boundaries of the collapsed 2-cells. Further, the map $\Phi^{\underline{M}}$ is an inclusion as showed in Lemma 4.10. In panel C of Figure 5 we show the projection of the signal s and the reconstructed signal $\Phi^{\underline{M}}\Psi^{\underline{M}}s$ on the Hodge decomposition. By Theorem 4.5 the signal is perfectly reconstructed on $\operatorname{Ker}\partial_1 = \operatorname{Ker}\Delta_1 \oplus \operatorname{Im}\partial_1^{\dagger}$, and only $\operatorname{Im}\partial_2$ contains non-trivial reconstruction error. Due to formatting constraints, we show the projection onto only 30 (randomly chosen) vectors of the Hodge basis in $\operatorname{Im}\partial_1^{\dagger}$ and $\operatorname{Im}\partial_2$.

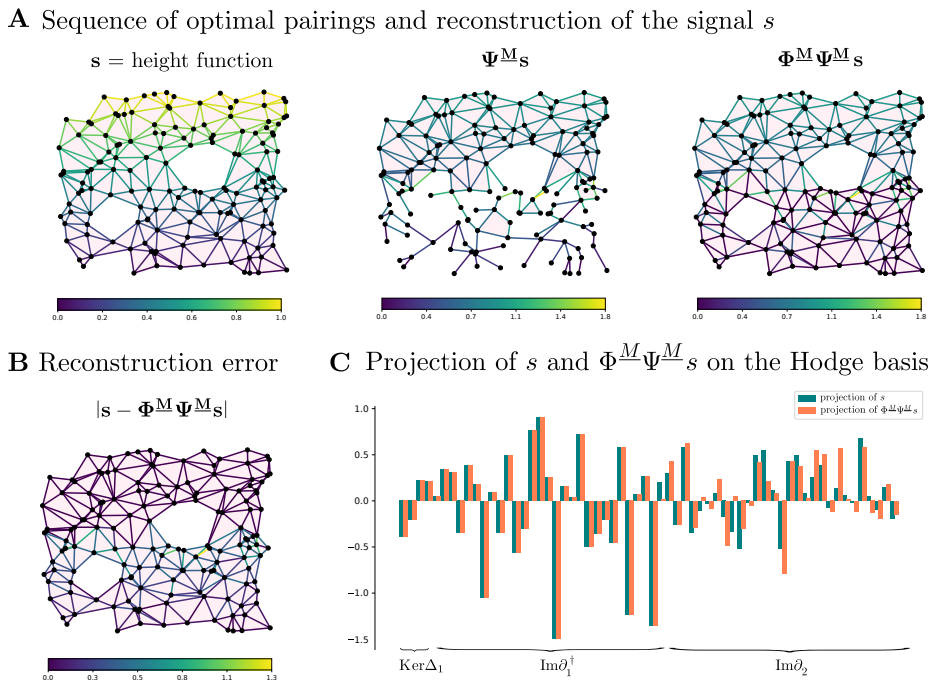


Figure 5: Optimal $(1,0)$-free sequential Morse matching (M) obtained by iterating Algorithm 2 for $k = 120$ on $(2,1)$-pairs. The signal s on the 1-cells is given by the height function.

In Figure 6 we propose the same example as above with a non-geometric function on the 1-cells. Specifically, the signal s on the 1-cells is given by sampling uniform at random in $[0,1]$ and the $(1,0)$-free sequential Morse matching $\underline{M}$ is obtained by iterating Algorithm 2 until all 2-cells were removed.

To quantify how low the topological reconstruction loss is after performing a sequential Morse matching with optimal pairings, we compare the reconstruction loss after a sequence of k optimal matchings with the reconstruction loss after a

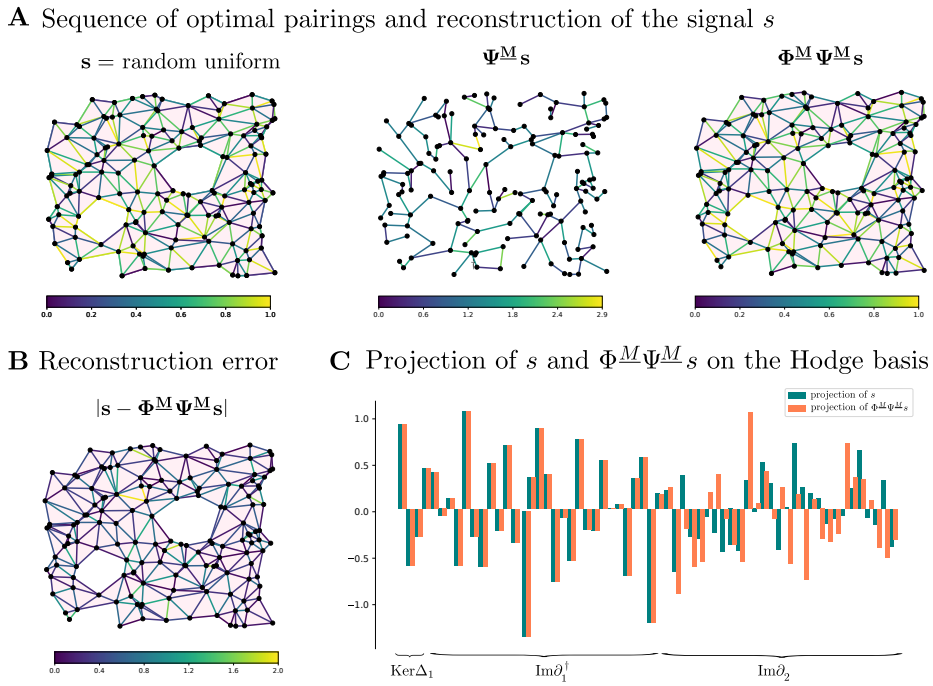


Figure 6: Optimal $(1,0)$-free sequential Morse matching $\underline{M}$ obtained by iterating Algorithm 2 until all 2-cells were removed. The signal s on the 1-cells is given by sampling uniform at random in $[0,1]$.

sequence of k random matchings.

Example 5.11. In this example we compare the sequence of optimal collapses presented in Example 5.10 in Figure 5 and in Figure 6 respectively with sequence of random collapses. In particular, we consider the complex $\mathcal{X}$ of Example 5.10 with signal on the 1-cells s given by the height function as in Figure 7 and signal s given by sampling uniformly at random in $[0,1]$ as in Figure 6. Instead of finding a sequence of $(2,1)$-pairings minimizing the reconstruction loss, at each step of algorithm 2 we will randomly remove a $(2,1)$-pair. We apply this procedure for $k = 120$ iterations in case s is the height function of the 1-cells and until all 2-cells are removed when the signal s is sampled uniform at random in $[0,1]$.

Figure 7.A shows the projection on the Hodge basis of s and $\Phi^{\underline{M}}\Psi^{\underline{M}}s$ when s is the height function and Figure 7.B shows the same result for s sampled uniform at random. Due to formatting constraints, we show the projection onto only 30 (randomly chosen) vectors of the Hodge basis in $\mathrm{Im}\,\partial_1^\dagger$ and $\mathrm{Im}\,\partial_2$. Note that, for both types of signal, the projection of the reconstructed signal $\Phi^{\underline{M}}\Psi^{\underline{M}}s$ and s on $\mathrm{Im}\,\partial_2$ differ significantly more than the the projection on $\mathrm{Im}\,\partial_2$ of

the reconstructed error and the signal in the case of optimal sequential Morse matching presented in Example 5.10 (see Figure 5.D and Figure 6.D)

Figure 7: Projection of the signal and the reconstructed signal on the Hodge basis after a sequence of random parings.

The quantitative results shown in the previous examples can be strengthened by comparing the value of the topological reconstruction loss for random and optimal sequence of pairings. In the next example we show that, for different types of both geometric and random signals, the topological reconstruction loss is significantly lower in sequentially optimal matchings than in random matchings.

Example 5.12. We consider again the same complex $\mathcal{X}$ as in Example 5.10. Figure 8 shows the value of the topological reconstruction loss after a sequence optimal and random pairings. We took sequences of length $k = 1, 2, \ldots 244$, terminating when all 2-cells were reduced. In panel A we consider a signal on the 1-cells sampled from a uniform distribution in $[0, 1]$, in panel B the signal is the height function on the 1-cells, in panel C the signal is sampled from a normal distribution (mean 0.5 and standard deviation 0.1), and in panel D the signal is given by the distance of the middle point of the 1-cells from the center of the cube $[0, 1] \times [0, 1]$. The blue curve is the average over 10 instantiations of optimal pairings while the green curve is the average over 10 instantiations of random pairings. The filled opaque bars show the respective mean square errors. Note that for all type of functions, the loss for the optimal pairings is significantly lower than the loss of random pairings.

6 Discussion

Contributions. The contributions of this paper are threefold. First we demonstrated that any deformation retract (Φ, Ψ) of finite-type based chain complexes over $\mathbb{R}$ is equivalent to a deformation retract $(\Phi^{\mathcal{M}}, \Psi^{\mathcal{M}})$ associated to a Morse

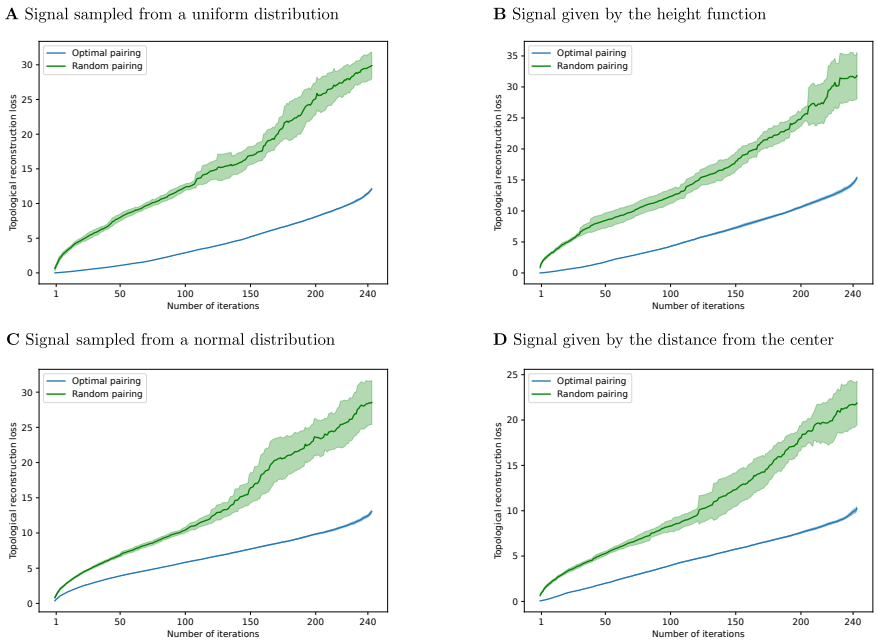


Figure 8: Topological reconstruction error for sequences of optimal and random up-collapses with different lengths.

matching $\mathcal{M}$ in a given basis. Second, we proved that the reconstruction error $s - \Phi\Psi s$, associated to any signal $s \in \mathbf{C}_n$ and deformation retract $(\Phi^{\mathcal{M}}, \Psi^{M})$, is contained in specific components of the Hodge decomposition if and only if $\mathcal{M}$ is a $(n, n-1)$-free (sequential) Morse matching. In the more general case, we showed that the reconstruction error associated to a deformation retract of a based chain complex is contained in specific parts of the Hodge decomposition if and only if its Morsification $\mathcal{M}$ is $(n, n-1)$-free. Moreover, we proved that the composition $\Phi^{M}\Psi^{M}s$ can be thought as a sparsification of the signal s in the $(n, n-1)$-free case. Finally, on the computational side, we designed and implemented algorithms that calculate (sequential) matchings that minimize the norm of topological reconstruction error. Further, we demonstrated computationally that finding a sequence of optimal matchings with our algorithm performs significantly better than randomly collapsing.

Limitations. The type of collapses that preserve cocycles involve chain maps, and those that preserve cycles involve the adjoints of these maps. This has two main limitations. The first one is that one can pick only one of the two features to be encoded at a time. The second limitation is the fact that chain maps do not necessarily send cocycles in $\mathbf{C}$ to cocycles in $\mathbf{D}$, and dually for cochain maps.

The proof of Theorem 4.5 hints at the difficulties of trying to define chain maps that preserve cocycles and dually cochain maps that preserve cycles. Namely,

to preserve cocycles with chain maps in dimension n, Morsification and Corollary 3.10 yield some insight, saying that this will occur only when the paired n-cells of Morsification lie in $\partial_n^\dagger$. A sufficient condition for this is that $\operatorname{Ker}\Psi \perp \operatorname{Im}\Phi$, in which case $\partial_n^\dagger\Big|_{\operatorname{Ker}\Psi} = (\partial_n|_{\operatorname{Ker}\Phi\Psi})^\dagger$ (See Appendix A.2). This rarely occurs in the standard CW or sheaf bases.

6.1 Applications and Future Work

Algorithms for optimal collapses. In this paper we minimize the reconstruction error by considering only single collapses. It would be desirable to find algorithms either for the optimal $(n, n-1)$-free Morse matchings, with no restriction on the length of the sequence, or for optimal $(n, n-1)$-free Morse matchings of given length k. We speculate that this task is likely to be NP-hard, given that the simpler task of finding a matching that minimises the number of critical cells is already known to be NP-hard [23, 27]. In this case, it would be useful to develop algorithms to approximate optimal matchings. These could be then used to compare how far away the reconstruction error of a sequence of k optimal pairings (Algorithm 2) is from the reconstruction error of a optimal collapse of size k.

Applications with inner products. In this paper, we have chosen examples that are helpful to visually illustrate the key results. However, the theory is built to accommodate a far larger class of applications. Examples where our theory may be useful for performing reductions that respect the inner product structure include the following

- **Markov-based heat diffusion.** The foundational work of [6] introduces a graph-theoretic model of heat diffusion on a point cloud, and can be framed in terms of combinatorial (graph) Laplacians. Here, distance kernel functions induce a weighting function on the nodes and edges of fully connected graph over the points. This weighting function is equivalent to specifying an inner product on $\mathbf{C}$ where the standard basis vectors are orthogonal [21].

- **Triangulated manifolds.** If M is a Riemannian manifold with smooth triangulation K, then $\mathbf{C}(K;\mathbb{R})$ has an inner product structure that converges to the canonical inner product on the de Rham complex $\Omega(M)$ under a certain type of subdivision [9]. This inner product on $\mathbf{C}(K;\mathbb{R})$ – and variations thereof – are useful in discrete Exterior calculus and its applications [19, 20].

The main theorems of this paper will hold in any of the circumstances described above, and provide a discrete Morse theoretic procedure for signal compression that is aware of the geometric information contained in the inner product structure.

Pooling in cell neural networks. Complementary to theoretical ideas, this research direction may have potential applications in pooling layers in neural networks for data structured on complexes or sheaves, such as in [3, 10, 16]. One could use Algorithm 2 to reduce the complex for a fixed sized k and then the map Φ to send the signal onto the reduced complex. We also envision that in pooling layers one could learn the $(n, n-1)$-free Morse matchings.

Acknowledgements K. M. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 859860, S. E. and C. H. were supported by the NCCR Synapsy grant of the Swiss National Science Foundation. S. E. was also supported by Swiss National Science Foundation under grant No. 200021-172636 (Ebli)

The authors would like to acknowledge Kathryn Hess for her detailed feedback and insightful discussions.

References

[1] Sergio Barbarossa and Stefania Sardellitti, *Topological signal processing over simplicial complexes*, IEEE Transactions on Signal Processing **68** (2020), 2992–3007.

[2] Sergio Barbarossa, Stefania Sardellitti, and Elena Ceci, *Learning from signals defined over simplicial complexes*, 2018 ieee data science workshop (dsw), 2018, pp. 51–55.

[3] Cristian Bodnar, Fabrizio Frasca, Yu Guang Wang, Nina Otter, Guido Montúfar, P. Liò, and M. Bronstein, *Weisfeiler and Lehman go topological: Message passing simplicial networks*, Proceedings of the 38th International Conference on Machine Learning PMLR **139** (2021), 1026–1037.

[4] Ronald Brown, *The twisted eilenberg-zilber theorem*, In 'simposio di topologia (messina, 1964)', edizioni oderisi, gubbio, 1965.

[5] Mathieu Carrière, Frédéric Chazal, Marc Glisse, Yuichi Ike, and Hariprasad Kannan, *A note on stochastic subgradient descent for persistence-based functionals: convergence and practical aspects*, CoRR (2020), available at `2010.08356`.

[6] Ronald R. Coifman and Stéphane Lafon, *Diffusion maps*, Applied and Computational Harmonic Analysis **21** (2006).

[7] Ivan Contreras and Andrew R. Tawfeek, *On discrete gradient vector fields and laplacians of simplicial complexes*, arXiv (2021), available at `2105.05388`.

[8] Justin Curry, Robert Ghrist, and Vidit Nanda, *Discrete Morse Theory for Computing Cellular Sheaf Cohomology*, Foundations of Computational Mathematics **16** (2016), no. 4.

[9] Jozef Dodziuk, *Finite-difference approach to the hodge theory of harmonic forms*, American Journal of Mathematics **98** (1976), no. 1, 79–104.

[10] Stefania Ebli, Michaël Defferrard, and Gard Spreemann, *Simplicial neural networks*, Topological Data Analysis and Beyond workshop at NeurIPS, 2020.

[11] Beno Eckmann, *Harmonische Funktionen und Randwertaufgaben in einem Komplex*, Commentarii Mathematici Helvetici **17** (1944), no. 1, 240–255.

[12] Robin Forman, *Morse theory for cell complexes*, Advances in Mathematics **134** (1998), no. 1, 90–145.

[13] ———, *Discrete Morse Theory and the Cohomology Ring*, Transactions of the American Mathematical Society **354** (2002oct), no. 12, 5063–5085.

[14] Rickard Brüel Gabrielsson, Bradley J. Nelson, Anjan Dwaraknath, and Primoz Skraba, *A topology layer for machine learning*, Proceedings of the twenty third international conference on artificial intelligence and statistics, 202026, pp. 1553–1563.

[15] V. K. A. M. Gugenheim, *On the chain-complex of a fibration*, Illinois Journal of Mathematics **16** (1972), 398–414.

[16] Jakob Hansen and Thomas Gebhart, *Sheaf neural networks* (2019), available at `2012.06333`.

[17] Jakob Hansen and Robert Ghrist, *Toward a spectral theory of cellular sheaves*, Journal of Applied and Computational Topology **3** (2019), no. 4.

[18] Allen Hatcher, *Algebraic topology*, Cambridge University Press, 2002.

[19] Ralf Hiptmair, *Finite elements in computational electromagnetism*, Acta Numerica **11** (2002), 237–339.

[20] Anil Nirmal Hirani, *Discrete exterior calculus*, California Institute of Technology, 2003.

[21] Danijela Horak and Jürgen Jost, *Spectra of combinatorial laplace operators on simplicial complexes*, Advances in Mathematics **244** (2013), 303–336.

[22] Xiaoling Hu, Yusu Wang, Li Fuxin, Dimitris Samaras, and Chao Chen, *Topology-aware segmentation using discrete Morse theory*, International conference on learning representations, 2021.

[23] Michael Joswig and Marc E. Pfetsch, *Computing optimal morse matchings*, SIAM J. Discret. Math. **20** (2006), 11–25.

[24] Tomasz Kaczynski, Konstantin Mischaikow, and Marian Mrozek, *Computational homology*, Vol. 157, Springer Science & Business Media, 2006.

[25] Tomasz Kaczyński, Marian Mrozek, and Maciej Ślusarek, *Homology computation by reduction of chain complexes*, Computers & Mathematics with Applications **35** (1998), no. 4, 59–70.

[26] Kwangho Kim, Jisu Kim, Manzil Zaheer, Joon Sik Kim, Frédéric Chazal, and Larry A. Wasserman, *Pllay: Efficient topological layer based on persistent landscapes*, Neurips, 2020.

[27] Francisco Martinez-Figueroa, *Optimal discrete morse theory simplification (expository survey)* (2021), available at `2111.05774`.

[28] John Milnor, *Morse theory*, Princeton University Press, 1969.

[29] Konstantin Mischaikow and Vidit Nanda, *Morse theory for filtrations and efficient computation of persistent homology*, Discrete & Computational Geometry **50** (2013), no. 2, 330–353.

[30] Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt, *Topological autoencoders*, Proceedings of the 37th international conference on machine learning, 2020, pp. 7045–7054.

[31] Facundo Mémoli, Zhengchao Wan, and Yusu Wang, *Persistent laplacians: properties, algorithms and implications* (2021), available at `2012.02808`.

[32] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst, *Graph signal processing: Overview, challenges, and applications*, Proceedings of the IEEE **106** (2018), no. 5, 808–828.

[33] Michael Robinson, *Topological signal processing*, Vol. 81, Springer, 2014.

[34] T. Mitchell Roddenberry, Michael T. Schaub, and Mustafa Hajij, *Signal processing on cell complexes* (2021), available at `2110.05614`.

[35] Michael T. Schaub, Yu Zhu, Jean-Baptiste Seby, T. Mitchell Roddenberry, and Santiago Segarra, *Signal processing on higher-order networks: Livin' on the edge... and beyond*, Signal Processing **187** (2021), 108149.

[36] Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, Eurographics symposium on point-based graphics, 2007.

[37] Emil Sköldberg, *Algebraic Morse theory and homological perturbation theory*, Algebra Discrete Math. **26** (2018), 124–129.

[38] Emil Sköldberg, *Morse theory from an algebraic viewpoint*, Transactions of the American Mathematical Society **358** (200601), 115–129.

[39] Ebli Stefania, Hacker Celia, and Maggs Kelly, *Morse theoretic signal compression and reconstruction on chain complexes*, GitHub, 2022. Availabe at `https://github.com/stefaniaebli/dmt-signal-processing`.

[40] P. Wood, A. P. Sheppard, and V. Robins, *Theory and algorithms for constructing discrete Morse complexes from grayscale digital images*, IEEE Transactions on Pattern Analysis and Machine Intelligence **33** (2011aug), no. 08, 1646–1658.

A Adjoints and Discrete Morse Theory

A.1 Matrix Representation of Adjoints and Weights

In this appendix we include a lengthier discussion about inner products and weight functions. To begin, we state a basic result about the matrix representation of the adjoint in finite-dimensional inner product spaces.

Proposition A.1. *Let V and W be finite-dimensional inner product spaces where*

$$\langle v_1, v_2 \rangle_V = v_1^T A v_2$$

and

$$\langle w_1, w_2 \rangle_W = w_1^T B w_2$$

for some fixed bases of V and W, where A, B are positive definite symmetric matrices. If $T : V \to W$, then the adjoint $T^\dagger : W \to V$ of T satisfies

$$T^\dagger = (A^{-1})^T T^T B^T.$$

The idea is that inner products are a vehicle to incorporate data with weights on the simplices into the linear algebraic world of combinatorial Laplacians. In particular, as mentioned in Remark 2.14, there is a one-to-one correspondence between inner products where elementary simplicial (co)chains form an orthogonal basis and weight matrices on the simplices. In the literature there are two approaches to associate weights to the simplices.

Firstly, the work of [31] begins by letting $\partial_n : \mathbf{C}_n(\mathcal{X}) \to \mathbf{C}_{n-1}(\mathcal{X})$ be the standard cellular boundary operator on a simplicial complex $\mathcal{X}$, and defines an inner product structure with respect to a basis given by the simplices via

$$\langle \sigma, \tau \rangle_n = \sigma^t W_n \tau,$$

where where each W_n is a diagonal matrix. The diagonal entries of W_n can be thought as weights on the n-cells. Then the coboundary operator $\partial_n^\dagger : \mathbf{C}_{n-1}(\mathcal{X}) \to \mathbf{C}_n(\mathcal{X})$, is given by

$$\partial_n^\dagger = W_n^{-1} \partial_n^T W_{n-1}$$

following the proposition above.

The second approach, exemplified by the work of [21], starts instead with the standard coboundary operator on a simplicial complex $\mathcal{X}$, $\delta_n = \partial_n^T : \mathbf{C}_{n-1}(\mathcal{X}) \to \mathbf{C}_n(\mathcal{X})$. Here the inner product structure on $\mathbf{C}_n(\mathcal{X})$ with respect to a basis given by the simplices is defined instead to be

$$\langle \sigma, \tau \rangle_n = \sigma^t W_n \tau,$$

where where each W_n is a diagonal matrix, the entries of which can be thought as weights on the n-cells. In this approach, the boundary operator is then written as

$$\delta_n^\dagger = W_{n-1}^{-1}\delta_n^T W_n. \tag{11}$$

Because we are working with discrete Morse theory, which conventionally is built for homology, we take the approach of always beginning with a boundary operator before constructing its adjoint operator. If one starts by defining a weighted boundary operator

$$\tilde{\partial}_n = W_{n-1}^{-1}\partial_n W_n,$$

then the adjoint operator induced by the weighted inner product yields

$$\tilde{\partial}_n^\dagger = W_n^{-1}W_n\partial_n^T W_{n-1}^{-1}W_{n-1} = \partial_n^T.$$

In other words, the adjoint of this weighted boundary operator is the standard coboundary operator, recovering the method of [21].

A.2 The Adjoint of a Morse Retraction

In this section, we explain why the orthogonality condition on the base I of a based chain complex $\mathbf{C}$ is important for establishing a discrete Morse theoretic interpretation when taking adjoints in Theorem 2.8. One can of course take the adjoint of the maps in this theorem to construct a deformation retract of the adjoint cochain complex, along with a coboundary operator, cochain weak-equivalences, and a cochain homotopy between them. However, only in the special case of an orthogonal base can these maps be decomposed in terms of adjoint flow backwards along paths in the original matching graph $\mathcal{G}(\mathbf{C})^M$.

Adjoint paths and flow. Suppose we have a Morse matching M on any based finite-type chain complex $\mathbf{C}$ over $\mathbb{R}$ with inner products. One can always define a notion of adjoint flow. First, observe that

$$\partial_{\beta,\alpha} = 0 \Leftrightarrow \partial_{\beta,\alpha}^\dagger = 0$$

and further

$$\partial_{\beta,\alpha}^\dagger \text{ isomorphism } \Leftrightarrow \partial_{\beta,\alpha} \text{ isomorphism.}$$

The opposite digraph $\mathcal{G}^{op}(\mathbf{C})^M$ (same vertices with edges reversed) of the directed graph $\mathcal{G}(\mathbf{C})^M$ then has an analogous relationship with the adjoint of the boundary operator. Namely, there is an edge $\beta \to \alpha$ whenever $\partial_{\beta,\alpha}^\dagger$ is non-zero, and a reversed edge $\beta \to \alpha$ in $\mathcal{G}^{op}(\mathbf{C})^M$ whenever $\alpha \to \beta$ is in M and $\partial_{\beta,\alpha}^\dagger$ is an

isomorphism. The same cells are unpaired in the adjoint world as the original one, and thus the critical cells of both are the same.

For a directed path $\gamma = \alpha, \sigma_1, \ldots, \sigma_k, \beta$ in the graph $\mathcal{G}(\mathbf{C})^M$, the *adjoint index* $I^\dagger(\gamma)$ of γ is written as

$$\mathcal{I}^\dagger(\gamma) = \epsilon_0 \partial^{\epsilon_0 \dagger}_{\alpha,\sigma_0} \circ \ldots \circ \epsilon_1 \partial^{\epsilon_{n-1}\dagger}_{\sigma_{n-2},\sigma_{n-1}} \circ \epsilon_n \partial^{\epsilon_n \dagger}_{\sigma_n,\beta} : C_\beta \to C_\alpha$$

where $k_i = -1$ if $\sigma_i \to \sigma_{i+1}$ is an element of M, and 1 otherwise. For any $\alpha, \beta \in I$, we can interpret this as following the path backwards and taking the adjoint of each map. The adjoint of the summed index also has a similar structure:

$$\Gamma^\dagger_{\beta,\alpha} = \sum_{\gamma:\alpha\to\beta} \mathcal{I}^\dagger(\gamma) : C_\beta \to C_\alpha.$$

where the sum runs over all paths γ from $\alpha \to \beta$ in $\mathcal{G}(\mathbf{C})^M$ or, equivalently, over all paths $\beta \to \alpha$ in $\mathcal{G}^{op}(\mathbf{C})^M$.

Main theorem for adjoint matching. To see what can go wrong, we need to be careful to distinguish categorical projections – those that simply delete components of a direct sum – from orthogonal projections that arise from the inner product structure.

Let $f : C = \oplus_\alpha C_\alpha \to D = \oplus_\beta D_\beta$ be a map of finite-type graded Hilbert spaces, based by I and J respectively. Each component $f_{\beta,\alpha}$ can be thought of as the composition of maps

$$f_{\beta,\alpha} : C_\alpha \xrightarrow{i_\alpha} C \xrightarrow{f} D \xrightarrow{\pi_\beta} D_\beta \tag{12}$$

such that we recover the total map f via sums

$$f = \sum_{\alpha,\beta} f_{\beta,\alpha}.$$

In a Hilbert space, the the inclusion i_α is adjoint to the *orthogonal* projection Proj_{C_α} onto C_α (Lemma 2.10), which not necessarily the categorical projection π_α. The categorical projection map π_α agrees with Proj_{C_α} if and only if

$$C_\alpha \perp C_{\alpha'} \tag{13}$$

for all $\alpha' \in I \setminus \alpha$. If this equation holds for both $\alpha \in I$ and $\beta \in J$, then the adjoint of the component map

$$(f_{\beta,\alpha})^\dagger : D_\beta \xrightarrow{\pi^\dagger_\beta} D \xrightarrow{f^\dagger} D \xrightarrow{i^\dagger_\alpha} D_\alpha.$$

agrees with the component maps of the adjoint

$$(f^\dagger)_{\alpha,\beta} : D_\beta \xrightarrow{i_\beta} D \xrightarrow{f^\dagger} D \xrightarrow{\pi_\alpha} D_\alpha.$$

If Equation 13 holds for all $\alpha \in I$ and $\beta \in J$, then

$$f^\dagger = \bigoplus_{\alpha,\beta} (f_{\beta,\alpha})^\dagger.$$

In other words, the adjoint commutes with the direct sum.

The reasoning above underpins why orthogonal components lead to a natural interpretation of the adjoint maps of 2.8 in terms of the adjoint flow. If this is the case, we can take the adjoint of 2.8 everywhere to prove the following important result.

Theorem A.2 (Sköldberg, [37])**.** *Let* $\mathbf{C}$ *be a finite-dimensional chain complex indexed by an orthogonal base* I*,* M *a Morse matching, and*

$$\mathbf{C}_n^M = \bigoplus_{\alpha \in I_n \cap M^0} C_\alpha.$$

The diagram

$$\mathbf{C}^M \underset{\Psi^\dagger}{\overset{\Phi^\dagger}{\leftrightarrows}} \mathbf{C} \circlearrowleft h^\dagger$$

is a deformation retract of cochain complexes, where for $x \in C_\beta$ *with* $\beta \in I_n$*,*

- $(\partial^\dagger_{\mathbf{C}^M})_n(x) = \sum_{\alpha \in M^0 \cap I_n} \Gamma^\dagger_{\beta,\alpha}(x)$
- $\Phi^\dagger_n(x) = \sum_{\alpha \in I_n} \Gamma^\dagger_{\beta,\alpha}(x)$
- $\Psi^\dagger_n(x) = \sum_{\alpha \in M^0 \cap I_n} \Gamma^\dagger_{\beta,\alpha}(x)$
- $h^\dagger_n(x) = \sum_{\alpha \in I_{n-1}} \Gamma^\dagger_{\beta,\alpha}(x)$

In most circumstances – weighted Laplacians, cellular sheaves, etc. – there is indeed an orthogonal basis. However, in the Morsification Lemma 3.7, we perform a reduction on the left component of

$$\operatorname{Ker} \Psi \oplus \operatorname{Im} \Phi$$

which, in general, is *not* orthogonal to $\operatorname{Im} \Phi$. One needs to be careful in such situations not to utilise the adjoint flow decompositions given in Theorem A.2.

2.2 Simplicial Neural Networks

(joint work with Michaël Defferrard and Gard Spreemann)

Accepted in Workshop on Topological Data Analysis and Beyond, NeurIps 2020.

Abstract

We present simplicial neural networks (SNNs), a generalization of graph neural networks to data that live on a class of topological spaces called simplicial complexes. These are natural multi-dimensional extensions of graphs that encode not only pairwise relationships but also higher-order interactions between vertices—allowing us to consider richer data, including vector fields and n-fold collaboration networks. We define an appropriate notion of convolution that we leverage to construct the desired convolutional neural networks. We test the SNNs on the task of imputing missing data on coauthorship complexes. Code and data are available at https://github.com/stefaniaebli/simplicial_neural_networks.

1 Introduction

The key to convolutional neural networks (CNNs) lies in the way they employ convolution as a local and shift-invariant operation on Euclidean spaces, e.g. $\mathbb{R}$ for audio or $\mathbb{R}^2$ for images. Recently, the concept of CNNs has been extended to more general spaces to exploit different structures that may underlie the data: This includes spherical convolutions for rotationally invariant data [1, 2, 3], more general convolutions on homogeneous spaces [4, 5, 6], or convolutions on graphs [7, 8].

Graph neural networks (GNNs) have proven to be an effective tool that can take into account irregular graphs to better learn interactions in the data [9, 10]. Although graphs are useful in describing complex systems of irregular relations in a variety of settings, they are intrinsically limited to modeling pairwise relationships. The advance of topological methods in machine learning [11, 12, 13], and the earlier establishment of *topological data analysis (TDA)* [14, 15, 16, 17] as a field in its own right, have confirmed the usefulness of viewing data as topological spaces in general, or in particular as simplicial complexes. The latter can be thought of as a higher-dimensional analog of graphs [18, 19]. We here take the view that structure is encoded in *simplicial complexes*, and that these represent n-fold interactions. In this setting, we present *simplicial neural networks (SNNs)*, a neural network framework that take into account locality of data living over a simplicial complex in the same way a GNN does for graphs or a conventional CNN does for grids.

Higher-order relational learning methods, of which hypergraph neural networks [20] and motif-based GNNs [21] are examples, have already proven useful in some applications, e.g. protein interactions [22]. However the mathematical theory underneath the notion of convolution in these approaches does not have clear connections with the global topological structure of the space in question. This leads us to believe that our method, motivated by Hodge–de Rham theory, is far better suited for situations where topological structure is relevant, such as perhaps in the processing of data that exists naturally as vector fields or data that is sensitive to the space's global structure.

2 Proposed method

Simplicial complexes. A simplicial complex is a collection of finite sets closed under taking subsets. We call a member of a simplicial complex K a *simplex* of *dimension* p if it has cardinality $p+1$, and denote the set of all such p-simplices K_p. A p-simplex has $p+1$ *faces* of dimension $p-1$, namely the subsets omitting one element. We denote these $[v_0, \ldots, \hat{v}_i, \ldots, v_p]$ when omitting the i'th element. If a simplex σ is a face of τ, we say that τ is a *coface* of σ. While this definition is entirely combinatorial, there is a geometric interpretation, and it will make sense to refer to and think of 0-simplices as *vertices*, 1-simplices as *edges*, 2-simplices as *triangles*, 3-simplices as *tetrahedra*, and so forth (see Figure 1, (b)).

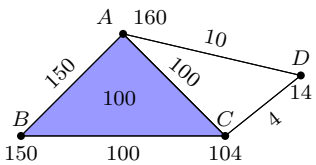
Let $C^p(K)$ be the set of functions $K_p \to \mathbb{R}$, with the obvious vector space structure. These *p-cochains* will encode our data. Define the linear *coboundary* maps $\delta^p : C^p(K) \to C^{p+1}(K)$ by

$$\delta^p(f)([v_0, \ldots, v_{p+1}]) = \sum_{i=0}^{p+1} (-1)^i f([v_0, \ldots, \hat{v}_i, \ldots, v_{p+1}]).$$

Observe that this definition can be thought of in geometric terms: The support of $\delta^p(f)$ is contained in the set of $(p+1)$-simplices that are cofaces of the p-simplices that make up the support of f.

Papers	Authors	Citations
Paper I	A, B, C	100
Paper II	A, B	50
Paper III	A, D	10
Paper IV	C, D	4

(a)



(b)

$$\begin{array}{c} \begin{array}{ccccc} \text{AB} & \text{AC} & \text{AD} & \text{BC} & \text{CD} \end{array} \\ \begin{pmatrix} 3 & 0 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 & -1 \\ 1 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 3 & -1 \\ 0 & -1 & 1 & -1 & 2 \end{pmatrix} \end{array} \begin{array}{c} \\ \text{AB} \\ \text{AC} \\ \text{AD} \\ \text{BC} \\ \text{CD} \end{array}$$

(c)

Figure 1: Constructing a simplicial complex from data. (a) Coauthorship data. (b) Coauthorship complex with corresponding cochains from the data. (c) Degree-1 Laplacian L_1 of the coauthorship complex.

Simplicial Laplacians. We are in this paper concerned with finite abstract simplicial complexes, although our method is applicable to a much broader setting, e.g. CW-complexes. In analogy with Hodge–de Rham theory [23], we define the *degree-i simplicial Laplacian* of a simplicial complex K as the linear map

$$\begin{aligned}
&\mathcal{L}_i : C^i(K) \to C^i(K) \\
&\mathcal{L}_i = \mathcal{L}_i^{\text{up}} + \mathcal{L}_i^{\text{down}} = \delta^{i*} \circ \delta^i + \delta^{i-1} \circ \delta^{i-1*},
\end{aligned}$$

where δ^{i*} is the adjoint of the coboundary with respect to the inner product (typically the one making the indicator function basis orthonormal). In most practical applications, the coboundary can be represented as a sparse matrix B_i and the Laplacians can be efficiently computed as $L_i = B_i^\intercal B_i + B_{i-1} B_{i-1}^\intercal$. The matrices L_0 and B_0 are the classic graph Laplacian and incidence matrix. Note that the Laplacians carry valuable topological information about the complex: The kernel of the k-Laplacian is isomorphic to the k-(co)homology of its associated simplicial complex [24, 25][1].

Simplicial convolution. A convolutional layer is of the form $\psi \circ (f * \varphi_W)$, where $*$ denotes convolution, φ_W is a function *with small support* parameterized by learnable weights W, and ψ is some nonlinearity and bias. This formulation of CNNs lends itself to a spectral interpretation that we exploit to extend CNNs to a much more general setting.

Following [8] and motivated by the fact that the discrete Fourier transform of a real-valued function on an n-dimensional cubical grid coincides with its decomposition into a linear combination of the eigenfunctions of the graph Laplacian for that grid, we define the Fourier transform of real p-cochains on a simplicial complex with Laplacians $\mathcal{L}_p$ as

$$\begin{aligned}
&\mathcal{F}_p : C^p(K) \to \mathbb{R}^{|K_p|} \\
&\mathcal{F}_p(c) = \left(\langle c, e_1 \rangle_p, \langle c, e_2 \rangle_p, \ldots, \langle c, e_{|K_p|} \rangle_p \right),
\end{aligned}$$

where the e_i's are the eigencochains of $\mathcal{L}_p$ ordered by eigenvalues $\lambda_1 \leq \cdots \leq \lambda_{|K_p|}$. The function $\mathcal{F}_p$ is invertible since $\mathcal{L}_p$ is diagonalizable; explicitly, if we write $U \operatorname{diag}(\Lambda) U^\intercal$ for a normalized eigendecomposition, the orthonormal

[1] In other words, the number of zero-eigenvalues of the k-Laplacian corresponds to the number of k-dimensional holes in the simplicial complex.

matrices U and $U^\intercal$ represent $\mathcal{F}_p^{-1}$ and $\mathcal{F}_p$, respectively. This is the foundation for Barbarossa's development of signal processing on simplicial complexes [26].

Recall that on the function classes for which it is defined, the classical Fourier transform satisfies $\mathcal{F}(f * g) = \mathcal{F}(f)\mathcal{F}(g)$, where the right-hand side denotes pointwise multiplication. This will be our definition of convolution in the simplicial setting. Indeed, for cochains $c, c' \in C^p(K)$ we *define* their convolution as the cochain

$$c *_p c' = \mathcal{F}_p^{-1}\left(\mathcal{F}_p(c)\mathcal{F}_p(c')\right).$$

Within this framework, we are led to define a *simplicial convolutional layer* with input p-cochain c and weights W as being of the form

$$\psi \circ \left(\mathcal{F}_p^{-1}(\varphi_W) *_p c\right)$$

for some as of yet unspecified $\varphi_W \in \mathbb{R}^{|K_p|}$. To ensure the central property that a convolutional layer be localizing, we demand that φ_W be a low-degree polynomial in $\Lambda = (\lambda_1, \ldots, \lambda_{|K_p|})$, namely

$$\varphi_W = \sum_{i=0}^{N} W_i \Lambda^i = \sum_{i=0}^{N} W_i(\lambda_1^i, \lambda_2^i, \ldots, \lambda_{|K_p|}^i),$$

for small N. In signal processing parlance, one would say that such a convolutional layer *learns filters that are low-degree polynomials in the frequency domain.*

The reason for restricting the filters to be these low-degree polynomials is best appreciated when writing out the convolutional layer in a basis. Let L_p^i denote the i'th power of the matrix for $\mathcal{L}_p$ in, say, the standard basis for $C^p(K)$, and similarly for c. Then (ignoring ψ),

$$\mathcal{F}_p^{-1}(\varphi_W) *_p c = \sum_{i=0}^{N} W_i U \operatorname{diag}(\Lambda^i) U^\intercal c = \sum_{i=0}^{N} W_i \left(U \operatorname{diag}(\Lambda) U^\intercal\right)^i c = \sum_{i=0}^{N} W_i L_p^i c.$$

This is important for three reasons, like for traditional CNNs. First, the convolution can be efficiently implemented by N sparse matrix-vector multiplications: This reduces the computational complexity from $\mathcal{O}(|K_p|^2)$ to $\mathcal{O}(\xi|K_p|)$ where ξ is the density factor. Second, the number of weights to be learned is reduced from $\mathcal{O}(|K_p|)$ to $\mathcal{O}(1)$. Third, the operation is N-localizing in the sense that if two simplices σ, τ are more than N hops apart, then a degree-N convolutional layer does not cause interaction between $c(\sigma)$ and $c(\tau)$ in its output (see the supplementary material). Those local interactions (in the spatial domain) can be interpreted as message-passing between simplices.

3 Experimental results

As many real-world datasets contain missing values, missing data imputation is an important problem in statistics and machine learning [27, 28]. Leveraging the structure underlying the data, GNNs have recently proved to be a powerful tool for this task [29]. Extending this view to higher-dimensional structure, we evaluate the performance of SNNs in imputing missing data over simplicial complexes.

Data. A *coauthorship complex (CC)* [19] is a simplicial complex where a paper with k authors is represented by a $(k-1)$-simplex. The added subsimplices of the $(k-1)$-simplex are interpreted as collaborations among subsets of authors—a natural hierarchical representation that would be missed by the hypergraph representation of papers as hyperedges between authors. In general, a simplicial complex representing n-fold interactions (e.g. between authors) can be constructed as the one-mode projection of a multipartite graph (e.g. a paper-author bipartite graph). The $(k-1)$-cochains are given by the number of citations attributed to the given collaborations of k authors. See Figure 1 and 4, and the supplementary material for details. We sampled (see the supplementary material) two coauthorship complexes—CC1 and CC2, see Table 2 for statistics—from the Semantic Scholar Open Research Corpus [30], a dataset of over 39 million research papers with authors and citations.

Method. We evaluated the performance of the SNNs on the task of imputing missing citations on the k-cochains (for $k = 0, 1, 2$) of the extracted coauthorship complexes. As in a typical pipeline for this task [28], missing values are artificially introduced by replacing a portion of the values with a constant. Specifically, given a fixed coauthorship complex, missing data is introduced at random on the k-cochains at 5 rates: $10\%, 20\%, 30\%, 40\%$, and 50%. The SNN is given as input the k-cochains on which missing citations are substituted by the median of known citations (as a reasonable first guess) and is trained to minimize the L_1 norm over known citations. We trained SNNs made of 3 layers with 30 convolutional filters of degree $N = 5$ with Leaky ReLu for 1000 iterations with the Adam optimizer and a learning rate of 10^{-3}.

Results. Figure 2 shows the mean accuracy and absolute error distribution (see the supplementary material for definitions) of the SNN in inputing missing citations on CC1. Observe that the distribution of the prediction error accumulates close to zero.

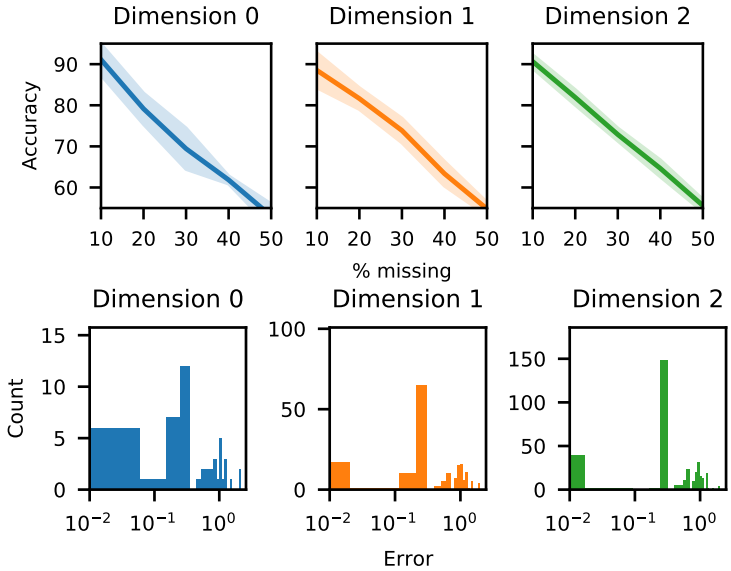


Figure 2: Performance of SNNs. Top: Mean accuracy ± standard deviation over 5 samples in imputing missing citations on CC1. Bottom: Absolute error distribution over 1 sample for 40% missing citations on CC1.

Table 1 shows the performance of two baselines: missing values inferred as (i) the mean or median of all known values, and (ii) the mean of the $(k-1)$ and $(k+1)$ neighboring simplices. SNNs well outperform these baselines. Comparison with stronger imputation algorithms is left for future work.

To demonstrate that our filters transfer across complexes, we evaluated how accurately an SNN trained on one coauthorship complex can impute missing citations on a different complex. We found that when imputing citations on CC1, a SNN trained on CC2 is almost as good as one trained on CC1 (compare Figures 2 and 3). We expect this result as coauthorship complexes share a similar structure, and the same process underlies the generation of citations across coauthorship complexes.

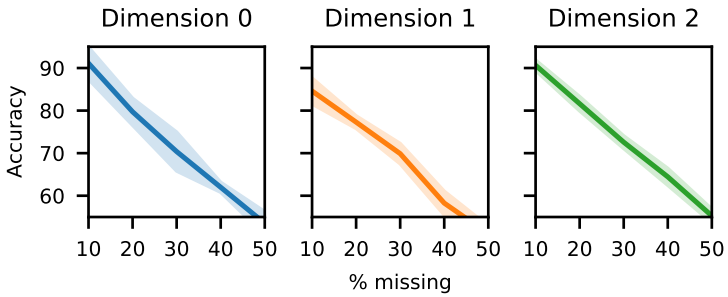


Figure 3: Performance on CC1 with an SNN trained on CC2.

Method	Dimension 0	Dimension 1	Dimension 2
Global Mean	3.30 ± 0.82	5.75 ± 1.28	2.96 ± 0.49
Global Median	7.78 ± 2.70	10.44 ± 1.00	12.50 ± 0.63
Neighbors Mean	11.88 ± 5.29	24.15 ± 1.85	27.38 ± 1.18

Table 1: Performance of baselines: mean accuracy $\pm$ standard deviation over 5 samples for 30% missing citations on CC1.

4 Conclusion and future work

We introduced a mathematical framework to design neural networks for data that live on simplicial complexes and provided preliminary results on their ability to impute missing data. Future work might include: (i) comparing SNNs with state-of-the-art imputation algorithms, (ii) using SNNs to solve vector field problems, (iii) generalizing coarsening and pooling to simplicial complexes, (iv) using boundaries and coboundaries to mix data structured by relationships of different dimensions, and (v) studying the expressive power of SNNs.

Unrelated to the simplicial nature of this work, we would like to emphasize how the spectral language was key to developing and even formulating our method. On homogeneous spaces, convolutions are defined as inner-products with filters shifted by the actions of a symmetry group of the space. They are the most general shift-invariant linear operators. On non-homogeneous spaces however, the spectral language yields generalized convolutions which are inner-products with *localized* filters [31, Sec. 2.4]. Those too are invariant to any symmetry the space might have. Convolutions exploit the space's structure to reduce learning complexity by sharing learnable weights through

shifts and localizations of filters.

S.E. and G.S. were supported by the Swiss National Science Foundation grant number 200021_172636, and would like to thank K. Hess for valuable discussions.

References

[1] Taco S Cohen et al. "Spherical CNNs". In: *International Conference on Learning Representations (ICLR)*. 2018. arXiv: 1801.10130.

[2] Carlos Esteves et al. "Learning SO(3) Equivariant Representations with Spherical CNNs". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018. arXiv: 1711.06721.

[3] Michaël Defferrard et al. "DeepSphere: a graph-based spherical CNN". In: *International Conference on Learning Representations (ICLR)*. 2020. URL: https://openreview.net/forum?id=B1e3OlStPB.

[4] Taco Cohen and Max Welling. "Group Equivariant Convolutional Networks". In: vol. 48. Proceedings of Machine Learning Research. 2016, pp. 2990–2999.

[5] Risi Kondor and Shubhendu Trivedi. "On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups". In: (2018), pp. 2747–2755. arXiv: 1802.03690.

[6] Daniel E Worrall et al. "Harmonic networks: Deep translation and rotation equivariance". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, pp. 5028–5037.

[7] Joan Bruna et al. "Spectral Networks and Locally Connected Networks on Graphs". In: *International Conference on Learning Representations (ICLR)*. 2014. arXiv: 1312.6203.

[8] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. "Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering". In: *Advances in Neural Information Processing Systems 29*. 2016, pp. 3844–3852.

[9] M. M. Bronstein et al. "Geometric Deep Learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42.

[10] Z. Wu et al. "A Comprehensive Survey on Graph Neural Networks". In: *IEEE Transactions on Neural Networks and Learning Systems* (2020), pp. 1–21.

[11] Rickard Brüel Gabrielsson et al. "A Topology Layer for Machine Learning". In: vol. 108. Proceedings of Machine Learning Research. 2020, pp. 1553–1563.

[12] C. Hofer, R. Kwitt, and M. Niethammer. "Learning Representations of Persistence Barcodes". In: *J. Mach. Learn. Res.* 20 (2019), 126:1–126:45.

[13] Bastian Rieck et al. "Neural Persistence: A Complexity Measure for Deep Neural Networks Using Algebraic Topology". In: International Conference on Learning Representations. 2019.

[14] Gunnar Carlsson. "Topology and Data". In: *Bull. Amer. Math. Soc.* 46.2 (2009), pp. 255–308.

[15] Frederic Chazal and Bertrand Michel. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists.* 2017. arXiv: 1710.04019.

[16] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction.* American Mathematical Soc., 2010.

[17] Robert Ghrist. "Barcodes: the persistent topology of data". In: *Bulletin of the American Mathematical Society* 45.1 (2008), pp. 61–75.

[18] Terrence J. Moore et al. "Analyzing collaboration networks using simplicial complexes: A case study". In: *Proceedings IEEE INFOCOM Workshops* (2012), pp. 238–243.

[19] Alice Patania, Giovanni Petri, and Francesco Vaccarino. "The shape of collaborations". In: *EPJ Data Science* 6.1 (2017).

[20] Yifan Feng et al. "Hypergraph Neural Networks". In: *Proceedings of the AAAI Conference on Artificial Intelligence.* Vol. 33. 01. 2019.

[21] F. Monti, K. Otness, and M. M. Bronstein. "MotifNet: a motif-based Graph Convolutional Network for directed graphs". In: *2018 IEEE Data Science Workshop (DSW).* 2018, pp. 225–228.

[22] Ze Xiao and Yue Deng. "Graph embedding-based novel protein interaction prediction via higher-order graph convolutional network". In: *PLoS One* 15 (2020).

[23] Ib Madsen and Jørgen Tornehave. *From calculus to cohomology: de Rham cohomology and characteristic classes.* Cambridge University Press, 1997.

[24] Beno Eckmann. "Harmonische Funktionen und Randwertaufgaben in einem Komplex". In: *Commentarii Mathematici Helvetici* 17.1 (1944), pp. 240–255.

[25] Danijela Horak and Jürgen Jost. "Spectra of combinatorial Laplace operators on simplicial complexes". In: *Advances in Mathematics* 244 (2013), pp. 303–336.

[26] Sergio Barbarossa, Stefania Sardellitti, and Elena Ceci. "Learning from signals defined over simplicial complexes". In: *2018 IEEE Data Science Workshop (DSW).* IEEE. 2018, pp. 51–55.

[27] Roderick J A Little and Donald B Rubin. *Statistical Analysis with Missing Data.* USA: John Wiley & Sons, Inc., 1986. ISBN: 0471802549.

[28] Fulufhelo Nelwamondo, Shakir Mohamed, and Tshilidzi Marwala. "Missing Data: A Comparison of Neural Network and Expectation Maximisation Techniques". In: *Current Science* 93 (May 2007).

[29] Indro Spinelli, Simone Scardapane, and Aurelio Uncini. "Missing data imputation with adversarially-trained graph convolutional networks". In: *Neural Networks* 129 (2020), pp. 249–260.

[30] Waleed Ammar et al. "Construction of the Literature Graph in Semantic Scholar". In: *NAACL.* 2018.

[31] Nathanaël Perraudin et al. "DeepSphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications". In: *Astronomy and Computing* 27 (2019), pp. 130–146.

A Supplementary material

Simplicial distance and the localizing property of the Laplacian. Suppose that σ and τ are p-simplices for which $(\nu_0, \nu_1, \ldots, \nu_d)$ is the shortest sequence of p-simplices with the property that $\nu_0 = \sigma$, $\nu_d = \tau$, and each ν_i shares a face or a coface with ν_{i-1}, and a face or a coface with ν_{i+1}. We say that d is the *simplicial distance* between σ and τ. Then for all $N < d$, the entry of L_p^N corresponding to σ and τ is 0, and so the filter does not cause

interaction between $c(\sigma)$ and $c(\tau)$. This is analogous to a size-d ordinary CNN layer not distributing information between pixels that are more than d pixels apart. We will refer to N as the *degree* of the convolutional layer, but one may well wish to keep in mind the notion of *size* from traditional CNNs.

Simplicial complexes as the projections of bipartite graphs. Given a bipartite graph X-Y, the simplicial projection on Y is the simplicial complex whose $(k-1)$-simplices are the sets of k vertices in Y that have at least one common neighbor in X. Cochains on the simplicial projection are naturally given by weights on X: Given any $(k-1)$-simplex $[y_1, \ldots, y_k]$ and its neighboring vertices $\{x_1, \ldots, x_j\} \subseteq X$, one can define a $(k-1)$-cochain as $\phi(\{x_1, \ldots, x_j\})$, for any function $\phi : \mathcal{P}(X) \to \mathbb{R}$. In our coauthorship application, ϕ is the sum and the weight of a paper is the number of times it is cited. See Figure 4.

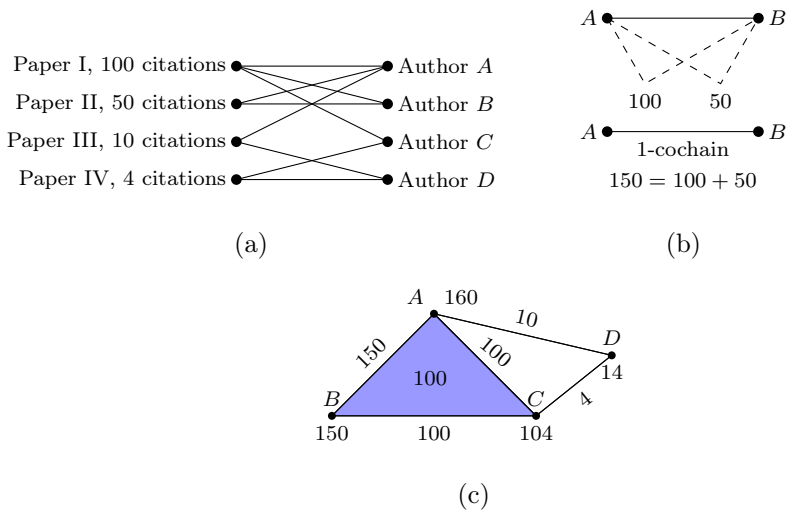


Figure 4: Constructing a simplicial complex and its cochain from a bipartite graph. (a) Paper-author bipartite graph (same data as in Figure 1). (b) The 1-simplex $[A, B]$ is included in the coauthorship complex since authors A and B collaborated on papers I and II. The 1-cochain on $[A, B]$ is given by the sum of their common papers' citations. (c) Resulting coauthorship complex with cochains.

Sampling papers. From the Semantic Scholar Open Research Corpus [30], we excluded papers with fewer than 5 citations or more than 10 authors. To sample a CC, we sampled 80 papers (corresponding to maximal simplices

Dimension	0	1	2	3	4	5	6	7	8	9	10
CC1	352	1474	3285	5019	5559	4547	2732	1175	343	61	5
CC2	1126	5059	11840	18822	21472	17896	10847	4673	1357	238	19

Table 2: Number of simplices of the two coauthorship complexes sampled from Semantic Scholar.

in the CC) by performing a random walk (of length 80, from a randomly chosen starting paper) on the graph whose vertices represent papers and edges connect papers sharing at least one author.

Mean accuracy and absolute error. A missing value is considered to be correctly imputed if the imputed value differs by at most 10% from the true value. The *accuracy* is the percentage of missing values that has been correctly imputed and the *absolute error* is the magnitude of the difference between the imputed and true value. For each rate of missing values, we compute the *mean accuracy* $\pm$ standard deviation over 5 samples with randomly damaged portions.

2.3 A Notion of Harmonic Clustering in Simplicial Complexes

(joint work with Gard Spreemann)

Accepted in 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019.

Abstract

We outline a novel clustering scheme for simplicial complexes that produces clusters of simplices in a way that is sensitive to the homology of the complex. The method is inspired by, and can be seen as a higher-dimensional version of, graph spectral clustering. The algorithm involves only sparse eigenproblems, and is therefore computationally efficient. We believe that it has broad application as a way to extract features from simplicial complexes that often arise in topological data analysis.

1 Introduction

An important objective in modern machine learning, and part of many scientific and data analysis pipelines, is *clustering* [4]. By clustering, we generally mean the separation of data into groups, in a way that is somehow meaningful for the domain-specific relationships that govern the underlying data and problem in question. However, the demands that the clustering scheme should satisfy are of course inherently vague.

For data that form a point cloud in Euclidean space, and where one expects k clusters to exist, one may employ elementary methods such as *k-means clustering* [37]. For data in a more abstract "similarity space", for which no obviously meaningful Euclidean embedding exists, researchers invented the schemes [35, 34] that we today refer to as *hierarchical clustering*. Alternatively, one can derive a graph structure from some notion of similarity between the data points. Treating the data points as vertices of a graph allows one to exploit the popular and highly successful spectral clustering techniques [41, 30] which developed from the field of spectral graph theory [9].

Although the graph structure provides us with additional information about the data, graphs are intrinsically limited to modeling pairwise interactions. The success of topological methods in studying data, and the parallel establishment of *topological data analysis (TDA)* as a field [13, 42] (see also [6, 7, 12, 15] for modern introductions and surveys), have confirmed the usefulness of viewing data through a higher-dimensional analog of graphs [28, 32]. Such a higher-dimensional analog is called a *simplicial complex*, a mathematical object whose structure can describe n-fold interactions between points. Their ability to capture hidden patterns in the data has led to various applications from biology [16, 33] to material science [21]. Recent work has also expanded classical graph-centric results — such as Cheeger inequalities [17, 5], isoperimetric inequalities [31] and spectral methods [22] — to simplicial complexes. This leads naturally towards a novel domain of "spectral TDA" methods.

In this paper we present the *harmonic clustering algorithm*, a novel clustering scheme inspired by the well-known spectral clustering algorithm for graphs. Our method, like spectral clustering, does not require any parameter optimization and involves only computing the smallest eigenvalue eigenvectors of a sparse matrix. The harmonic clustering algorithm is applied directly to a simplicial complex and it outputs a clustering of the *simplices* (of a

fixed degree) that is sensitive to the homological structure of the complex, something that is highly relevant in TDA. Moreover, since simplices can encode interactions of higher order than just the pairs captured by graphs, our algorithm allows us to cluster complex community structures rather than just the entities they comprise.

Our method can be seen as complimentary to the one presented in [5].

1.1 Spectral graph theory

The method we present in this paper does not require many formal results from spectral graph theory. The notions relevant for our purposes are described below for the sake of completeness.

In its simplest form, the *Laplacian* of an undirected and unweighted finite graph G is taken to be the positive definite matrix $L = D - A$, where A is the adjacency matrix of G and D its diagonal degree matrix (i.e. the row/column sums of A). The *normalized Laplacian* is then defined as $\bar{L} = D^{-1/2} L D^{-1/2}$. For reasons that will become clear later on (see 2.1), we will write $C_0(G)$ for the free real vector space generated by the vertices of G, and consider L as the matrix of a linear map $C_0(G) \to C_0(G)$ in this basis.

Already in the middle of the 19th century it was clear that the eigenvalue spectrum of L has a lot to say about G, as is evident from as early as a historic theorem of Kirchhoff relating the eigenvalues of the Laplacian with the number of spanning trees in the graph [25]. From the 1950s, graph theorists and quantum chemists were independently discovering more relationships between a graph and the eigenspectrum of its Laplacian. However, the publication of the book [10] may be said to mark the start of *spectral graph theory* as a field in its own right. A modern introduction to the field and references to the results listed below can be found in [9].

The spectrum of L encodes information about the connectivity of the graph. For instance, the number of connected components of the graph is equal to the dimension of the kernel of L. Moreover, the eigenvectors associated to the zero eigenvalues, also called *harmonic representatives*, take constant values on connected components. A perhaps more interesting result is given by the Cheeger constant [8], a measure of how far away a connected graph is from being disconnected by bounding the smallest non-zero eigenvalue of L.

Theorem 1 (Cheeger, 1969 [8]; see also e.g. [9])**.** *Let $G = (V; E)$ be a*

finite, connected, undirected, unweighted graph. Write $\mathrm{cut}(G)$ *for the triples* $(S, \bar{S}, \partial S)$ *with* $S, \bar{S} \subseteq V$ *and* $\partial S \subseteq E$ *such that* $S \sqcup \bar{S} = V$ *and*

$$\partial S = \{(u, w) \in E : u \in S, w \in \bar{S}\}.$$

Define the Cheeger constant *of* G *as*

$$h(G) = \min_{(S,\bar{S},\partial S)\in \mathrm{cut}(G)} \frac{|\partial S|}{\min(\sum_{u\in S} \deg(u), \sum_{w\in \bar{S}} \deg(w))}.$$

Then the first non-zero eigenvalue λ_1 *of the graph's normalized Laplacian satisfies*

$$2h(G) \geq \lambda_1 \geq \frac{(h(G))^2}{2}.$$

A partition of V as $S \sqcup \bar{S}$ that attains the Cheeger constant is called a *Cheeger cut*. It is known that finding an exact Cheeger cut is an NP-hard problem [38]. One of the best known approaches to approximating the Cheeger cut is *the spectral clustering method*, which takes the first non-zero eigenvalue eigenvector of the graph Laplacian as a relaxed real-valued solution of the original discrete optimization problem [41]. Namely, the smallest non-zero eigenvector of $\bar{L}$, also called *the Fiedler vector* or *the connectivity vector* [14], can be exploited to find the best partition of the graph into two "almost disconnected" components. The Cheeger cut can be easily generalized to find $k+1$ "almost disconnected" components using the k first non-zero eigenvectors of the graph Laplacian [41].

1.2 Graph spectral clustering

The Fiedler vector being a relaxed solution of the Cheeger cut has implications for clustering the vertices of a graph into "almost disconnected" components [41, 3]. For the remainder of this section we will assume that the graph under consideration is connected.

Graph spectral clustering of a graph $G = (V, E)$ with Laplacian L works in two steps. First, one uses the information encoded in the lowest-eigenvalue eigenvectors of L to map V into low-dimensional Euclidean space. One thereafter uses standard k-means or any applicable Euclidean clustering technique on the points in the image of this map, before pulling back to G. Specifically, we will write $e_1, e_2, \ldots, e_n$ for the eigenvectors associated with the n first non-zero eigenvalues of L. One defines a function, also called a

spectral embedding, $\phi : C_0(G) \to \mathbb{R}^n$ by

$$\phi(v) = (\langle v, e_1 \rangle, \langle v, e_2 \rangle, \ldots, \langle v, e_n \rangle), \tag{1}$$

where $\langle \bullet, \bullet \rangle$ is the inner product on $C_0(G)$ that makes V orthonormal. As a finite Euclidean point cloud, $\operatorname{im} \phi$ is then clustered in $\mathbb{R}^n$ by standard k-means or any suitable clustering algorithms. The clustering obtained is then pulled back to V. Figure 1 shows an example. Observe that in this case, mapping into the real line using only the Fiedler vector would suffice (i.e. $n = 1$).

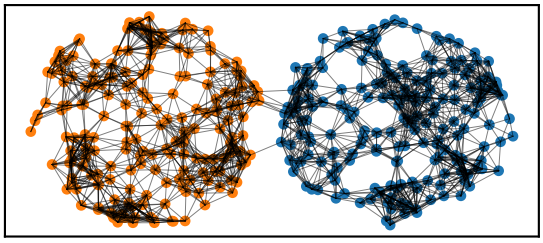


Figure 1: Graph spectral clustering of the nodes of a graph with two well-connected components weakly interconnected. Clustering using the Fiedler vector produces as clusters the well-connected components.

As pointed out in [41], spectral clustering is one of the standard approaches to identify groups of "similar behavior" in empirical data. It is therefore not surprising that it has been successfully employed in many fields ranging from computer science and statistics to biology and social science. Moreover, compared to other approaches, such as Gaussian Mixture Models clustering, spectral clustering does not require any parameter optimization and can be solved efficiently by standard linear algebra methods.

2 Harmonic clustering in simplicial complexes

Our method is inspired by spectral clustering in graphs, but applies instead to a higher-dimensional analog, namely *simplicial complexes*. Instead of clustering only vertices, which are the zero-dimensional building blocks of graphs and simplicial complexes, the method clusters independently building blocks of any dimension.

This section outlines the prerequisite basic constructions from algebraic topology before describing our method. A reader interested in more background on algebraic topology is directed to standard textbooks [18].

Those wishing a quick overview of method can view it in algorithmic form in figure 3.

2.1 Algebraic topology

A *simplicial complex* is a collection of finite sets closed under taking subsets. We refer to a set in a simplicial complex as a *simplex* of *dimension* p if it has cardinality $p+1$. Such a p-simplex has $p+1$ *faces* of dimension $p-1$, namely the sets omitting one element, which we will denote as $(v_0, \ldots, \hat{v}_i, \ldots, v_p)$ when omitting the i'th element. While this definition is entirely combinatorial, we will soon see that there is a geometric interpretation, and it will make sense to refer to and think of 0-simplices as *vertices*, 1-simplices as *edges*, 2-simplices as *triangles*, 3-simplices as *tetrahedra*, and so forth.

Let $C_p(K)$ be the free real vector space with basis K_p, the set of p-simplices in a simplicial complex K. The elements of $C_p(K)$ are called *p-chains*. These vector spaces come equipped with *boundary maps*, namely linear maps defined by

$$\partial_p : C_p \to C_{p-1}$$
$$\partial_p((v_0, \ldots, v_p)) = \sum_{i=0}^{p} (-1)^i (v_0, \ldots, \hat{v}_i, \ldots, v_p)$$

with the convention that $C_{-1}(K) = 0$ and $\partial_0 = 0$ for convenience. Figure 2 shows how the boundary maps give a geometric interpretation of simplicial complexes.

One readily verifies that $\partial_p \circ \partial_{p+1} = 0$, and so $C_\bullet(K)$ is a real *chain complex*. By the *p'th homology vector space* of K we will mean the p'th homology vector space of this chain complex, namely

$$H_p(K) = H_p(C_\bullet(K)) = \ker \partial_p / \operatorname{im} \partial_{p+1}.$$

The elements of $\ker \partial_p$ are called *p-cycles*, while those of $\operatorname{im} \partial_{p+1}$ are called *p-boundaries*, as can be seen geometrically in Figure 2. The *Betti numbers* are the dimensions of the homology vector spaces, and we write $\beta_p(K) = \dim H_p(K)$. Intuitively, the Betti numbers count connected components, non-bounding loops, non-bounding cavities, and so forth.

We emphasize again that this is homology with *real* coefficients, not integer or finite field, as is common in TDA.

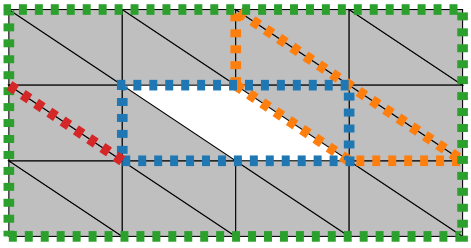


Figure 2: A simplicial complex K with 20 0-simplices, 38 1-simplices (the edges) and 22 2-simplices (the filled triangles), with some highlighted 1-chains. The highlighted simplices in these represent the edges with non-zero coefficient in each chain (the unfamiliar reader is invited to fill in possible values for these coefficients). The red 1-chain consists of a single 1-simplex, and is neither a cycle nor a boundary. The orange 1-chain has trivial boundary, and is therefore a cycle. It is not a representative of any non-trivial homology class, for it is the boundary of 2-chain consisting of the three 2-simplices it encloses. The green and the blue 1-chains are cycles that represent the same homology class (intuitively the 2-dimensional hole in the middle). $H_0(K)$ is 1-dimensional, K's single connected component, while $H_1(K)$ is 1-dimensional due to the central hole.

2.2 Simplicial Laplacians

We are in this paper concerned with finite simplicial complexes, and assume that they are built in a way that encodes useful information about the data being studied. We will briefly discuss the case where each simplex in K comes equipped with extra data — including, but not limited to the filtration/weighting information that is ubiquitous in TDA — or with a normalization factor derived from the complex's structure, in the form of a function $w : K \to \mathbb{R}_+$. The latter is analogous to the various normalization schemes that are often used in graph spectral theory. Our computational experiments, however, will only consider the case $w = 1$.

The weights are encoded into the chain complex by endowing each degree with the inner product that makes all simplices orthogonal, and a simplex have norm given by the weight, i.e.

$$\langle \bullet, \bullet \rangle_i : C_i(K) \times C_i(K) \to \mathbb{R}$$
$$\langle \sigma, \tau \rangle_i = \begin{cases} w(\sigma)^2 & \text{if } \sigma = \tau \\ 0 & \text{otherwise.} \end{cases}$$

Further discussions of weighting schemes can be found in [22].

We place no further assumptions on the simplicial complex that we take as input. In particular, it is not necessary for it to come equipped with some

embedding into Euclidean space, nor do we demand that it triangulates a Riemannian manifold. Therefore dualities like the Hodge star, which is used to construct the Hodge–de Rham Laplacian in the smooth setting [26] that motivates us, are unavailable for our method. The same is true for discrete versions of the Hodge star, such as that of Hirani [20]. Instead of dualizing with respect to a Hodge star, to define a discrete version of the Laplacian for simplicial complexes, we simply take the linear adjoint of the boundary operator with respect to the inner product, defining $\partial_i^* : C_{i-1} \to C_i$ by

$$\langle \partial_i^* \sigma, \tau \rangle_i = \langle \sigma, \partial_i \tau \rangle_{i-1} \quad \forall \sigma \in K_{i-1}, \tau \in K_i.$$

In analogy with Hodge–de Rham theory, we then define the *degree-i simplicial Laplacian* of a simplicial complex K as the linear operator $\mathcal{L}_i : C_i(K) \to C_i(K)$ such that

$$\begin{aligned}
\mathcal{L}_i &= \mathcal{L}_i^{\text{up}} + \mathcal{L}_i^{\text{down}} \\
\mathcal{L}_i^{\text{up}} &= \partial_{i+1} \circ \partial_{i+1}^* : C_i(K) \to C_i(K) \\
\mathcal{L}_i^{\text{down}} &= \partial_i^* \circ \partial_i : C_i(K) \to C_i(K).
\end{aligned}$$

The *harmonics* are defined as

$$\mathcal{H}_i(K) = \ker \mathcal{L}_i.$$

Observe that there are p Laplacians for a complex of dimension p. In most practical applications, the matrices for the Laplacians are very sparse and can easily be computed as a product of sparse boundary matrices and their transposes.

The following discrete version of the important *Hodge decomposition theorem* is a simple exercise in linear algebra in the current setting.

Theorem 2 (Eckmann, 1944 [11])**.** *The vector spaces of chains decompose orthogonally as*

$$C_i(K) \cong \mathcal{H}_i(K) \oplus \operatorname{im} \partial_{i+1} \oplus (\ker \partial_i)^{\perp}.$$

Moreover,

1. $\mathcal{H}_i(K) \cong H_i(K)$
2. *the harmonics are both cycles and cocycles (i.e. cycles with respect to* ∂_{i+1}^**)*

3. *the harmonics are the L^2-minimal representatives of their (co)homology classes, i.e. if $h \in \mathcal{H}_i(K)$ and $h \sim z \in \ker \partial_i$ are homologous, then $\langle h, h \rangle_i \leq \langle z, z \rangle_i$.*

The first detailed work on the spectral properties of this kind of simplicial Laplacian was carried out by Horak and Jost [22]. Recently Steenbergen et al. [36] provided a notion of a higher dimensional Cheeger constant for simplicial complexes. At the same time, Gundert and Szedlák [17] proved a lower bound for a modified version of the higher dimensional Cheeger constant for simplicial complex which was later generalized to weighted complexes by Braxton et al. Mukherjee and Steenbergen [29] developed an appropriate notion of random walks on simplicial complexes, and related the asymptotic properties of these walks to the simplicial Laplacians and harmonics. It is worth mentioning that, to the best of our knowledge, no connection between the eigenvectors of the simplicial Laplacian and an optimal cut for simplices in higher dimensions is known.

Our contribution is a notion of spectral clustering for simplicial complexes using the harmonics.

2.3 Harmonic clustering

Observe that the ordinary graph Laplacian, as described in section 1.1, is just the matrix of $\mathcal{L}_0 = \mathcal{L}_0^{\text{up}}$ in the standard basis for $C_0(G)$. The function ϕ in equation (1) can thus be seen as projecting the 0-simplices onto a subspace of low-but-nonzero-eigenvalue eigenvectors of $\mathcal{L}_0$. The zero part of the spectrum is not used. Theorem 2 makes the reason clear: harmonics in $\mathcal{H}_0(G)$ have the same coefficient for every vertex in a connected component of G. As connectivity information is easy to obtain anyway, there is little use in adding these eigenvectors to the subspace that ϕ projects onto. This is not so for the higher Laplacians. In fact, our method *primarily* uses the harmonics, and only optionally ventures into the non-zero part of the eigenspectrum.

In what follows, K is a fixed simplicial complex arising from data. The particulars of *how* K was built from data is outside the scope of this paper, and is a topic that is well-studied in the field of TDA in general. Our goal is to obtain a useful clustering of K_p for some chosen p. We assume that K is of low "homological complexity" in degree p, by which we mean that $\beta_p(K)$ is small (less than 10, say).

Analogously to ϕ above, we define the *harmonic embedding*

$$\psi : K_p \to \mathbb{R}^{\beta_p(K)}$$
$$\psi = \xi \circ \mathrm{proj}_{\mathcal{H}_p(K)} \circ i,$$

where $i : K_p \hookrightarrow C_p(K)$, $\mathrm{proj} : C_p(K) \to \mathcal{H}_p(K)$ is orthogonal projection, and $\xi : \mathcal{H}_p(K) \to \mathbb{R}^{\beta_p(K)}$ is any vector space isomorphism. In practice, we simply pick an orthonormal basis $h_1, \ldots, h_{\beta_p(K)}$ for $\mathcal{H}_p(K)$ and let

$$\psi(\sigma) = \left(\langle \sigma, h_1 \rangle_p, \langle \sigma, h_2 \rangle_p, \ldots, \langle \sigma, h_{\beta_p(K)} \rangle_p \right).$$

In many situations of practical use, it turns out that many points in $\operatorname{im} \psi$ lie along one-dimensional subspaces of $\mathbb{R}^{\beta_p(K)}$. The membership of a point $\psi(\sigma)$ in such a subspace is what is used to cluster the p-simplex σ (or to leave it unclustered in case it is not judged to be sufficiently close to lying in one of the subspaces). This amounts to clustering K_p by performing Euclidean subspace clustering of $\operatorname{im} \psi$. A variety of Euclidean subspace clustering methods are available, but are outside the scope of this paper. Examples include *independent component analysis* [23], *SUBCLU* [24], and density maximization on $\mathbb{S}^{\beta_p(K)-1}$ (or, more precisely, on $\mathbb{RP}^{\beta_p(K)-1}$), which itself has a multitude of approaches, including purely TDA-based ones by means of persistent homology of sublevel sets.

We point out that the choice of the isomorphism $\xi : \mathcal{H}_p(K) \to \mathbb{R}^{\beta_p(K)}$ does not matter on a theoretical level. It may, however, have practical implications for how easy it is to perform subspace clustering. In experiments we typically choose ξ to be the isomorphism that sends h_i to the standard basis vector e_i. Choosing a different orthonormal basis for $\mathcal{H}_p(K)$ then just amounts to an element of $\mathrm{SO}(\beta_p(K))$ acting on $\operatorname{im} \psi$.

Figure 3 summarizes our method in algorithmic form.

3 Experimental results

In this section we present experimental results for the harmonic clustering algorithm on synthetic data. Specifically, we focus on clustering the edges of various constructed simplicial complexes. The outcomes of our experiments suggest that the harmonic clustering algorithm provides clusters sensitive to the homology of the complex. Comparing our results with those of the traditional spectral clustering algorithm applied to the graph underlying the

Require: Integer $p \geq 0$; simplicial complex K with $\beta_p = \dim(H_p(K))$ small, $K_p = \{\sigma_1, \ldots, \sigma_N\}$, and inner products $\langle \bullet, \bullet \rangle_p$ on $C_p(K)$.

$L_p \leftarrow$ matrix for $\mathcal{L}_p$

$(h_1, \ldots, h_{\beta_p}) \leftarrow$ orthonormal basis for $\ker \mathcal{L}_p$ (computed using iterative methods [19] on L_p)

for $i = 1$ **to** $i = N$ **do**

  $x_i \leftarrow \left(\langle \sigma_i, h_1 \rangle_p, \ldots, \langle \sigma_i, h_{\beta_p} \rangle_p \right)$

end for

$(a_1, \ldots, a_k) \leftarrow \text{subcluster}(x_1, \ldots, x_N)$

for $i = 1$ **to** $i = k$ **do**

  $c_i \leftarrow \{\sigma_j \in K_p : j \in a_i\}$

end for

Ensure: Homologically sensitive clustering $c_1, \ldots, c_k$ of p-simplices in K.

Figure 3: Our method in algorithmic form. The subroutine subcluster refers to any Euclidean subspace clustering scheme, such as independent component analysis [23], SUBCLU [24], or density maximization on $\mathbb{S}^{\beta_p(K)-1}$ (or, more precisely, on $\mathbb{RP}^{\beta_p(K)-1}$). The latter can be done using methods from TDA, for example by means of persistent homology of certain sublevel sets. Note that there may be unclustered simplices, i.e. it may happen that $\cup_{i=1}^k c_i \neq K_p$.

simplicial complex reinforces the idea that our algorithm reveals substantially different patterns in data compared to the classical method.

Below, we consider four simplicial complexes. Three of them are complexes built from Euclidean point clouds by standard methods from TDA, while one is a triangulation of a torus. We reiterate that our method works with abstract simplicial complexes without utilizing any embedding of these into an ambient space. Euclidean point clouds just happen to be a good and common source of simplicial complexes in TDA, and allow for visualization of the obtained clustering in a way that easily relates to the original data.

An important step in preprocessing many kinds of input data in TDA is constructing a simplicial complex satisfying certain theoretical properties. In particular, if the input data come from points sampled from a topological space $X \subset \mathbb{R}^n$, one may wish for the homology of the complex to coincide with the homology of X.

Two constructions for which some such guarantees exist are the *alpha complex* [1] and the *Vietoris–Rips (VR) complex* [40]. Both can be seen as taking a point cloud and producing a *filtered* simplicial complex K, i.e. a

sequence $(K_t)_{t\in\mathbb{R}_+}$ with the property that $K_s \subset K_t$ whenever $s \leq t$. We wish to work with a single simplicial complex, not a filtration, so we use persistent homology (see e.g. [15]) to find the filtration scale t for which K_t has the appropriate homology. Of course, since in practice one probably has little or no knowledge of X itself, one cannot necessarily know the "correct" t to consider. However, it is often the case in TDA that long-lived homological features — that is to say, homology classes that remain non-trivial under the induced maps $H_p(K_s) \to H_p(K_t)$ for large $t - s$ — express interesting properties of the underlying space. We therefore choose a K_t to consider by looking for a scale t within the range of a manageable number of long-lived features and few short-lived ones in the degree under consideration.

In the following experiments, we simplify the setup in the algorithm in figure 3 by performing the subroutine subcluster in a somewhat ad hoc semi-manual way. Specifically, all the images of the ψ's lie in $\mathbb{R}^2$ or $\mathbb{R}^3$ in these experiments, so we manually pick out the subspaces $V_1, \ldots, V_k$ in question. Then, the points in $\operatorname{im}\psi$ are orthogonally projected onto each of the subspaces. A point $\psi(\sigma)$ is determined to lie on subspace V_i if $\operatorname{proj}_{V_i}(\psi(\sigma)/\|\psi(\sigma)\|)$ has norm at least 0.98, while the onto all other subspaces has norm less than 0.02. The simplex σ is then said to be in cluster number i. If the above is not true for any of the subspaces, σ is considered unclustered.

In many of the experiments that follow, many points in $\operatorname{im}\psi$ end up determined as "unclustered" because they project well onto neither of the 1-dimensional subspaces, or project too well onto multiple of them, as described in 2.3. This is not necessarily a problem, as the parts that are clustered contain a lot of useful information. Moreover, the problem can be reduced by choosing less *ad hoc* subspace clustering methods than we are currently employing.

To ease visualization, we focus on simplicial complexes that naturally live in $\mathbb{R}^2$ or $\mathbb{R}^3$ because they arise from point clouds.

3.1 Wedge of a sphere and two circles

In this experiment, we consider a noisy sampling of $X = \mathbb{S}^2 \vee \mathbb{S}^1 \vee \mathbb{S}^1$ realized as a central unit sphere with unit circles wedged onto antipodal points. We sampled 1000 points uniformly randomly from the central sphere, adding radial uniform noise with amplitude 0.01. The circles were sampled using 100 points each, again with a radial noise of 0.01. This yields a point cloud

with 1200 points, which is shown in figure 4. The VR complex is certainly a suboptimal choice of simplicial complex to build on this kind of data, but we chose it to demonstrate that our method works well also for such an overly dense complex. The complex, constructed at scale 1/2 and denoted K within this section, has 35722 1-simplices and 485189 2-simplices, and the Betti numbers are $\beta_0(K) = 1$, $\beta_1(K) = 2$, $\beta_2(K) = 1$, as for X itself.

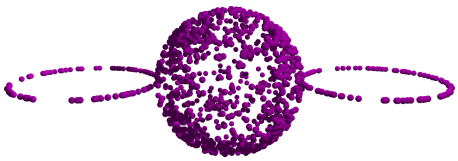


Figure 4: The point sample under consideration in 3.1.

We focus on clustering the 1-simplices of the complex. The image of ψ in $\mathbb{R}^2$ is shown in figure 5. The points are colored according to which of the two one-dimensional subspaces they are deemed to belong to. The determination was made by a simple criterion of projecting well enough onto one of the lines, but not the other. Points that project well onto both or neither are considered unclustered and shown as red. Figure 6 shows this clustering pulled back to the complex itself, excluding the unclustered edges. Observe how the method separates the 1-simplices of the VR complex in a manner that is sensitive to the two non-bounding cycles that generate homology in degree 1 (the two circles).

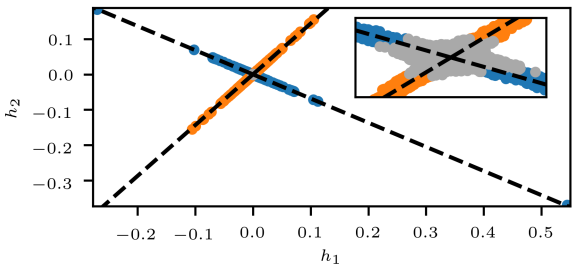


Figure 5: The image of ψ for the 1-simplices in the VR complex from the experiment in 3.1. The dashed lines indicate the subspaces used for clustering. The inset shows a detailed view near the origin, where one can see a large number of points in gray that are unclustered due to them projecting too well onto both subspaces.

We also repeated the experiment with one of the circles in X moved to be attached to the other circle instead of the sphere. This space is obviously homotopy equivalent to X, but is geometrically very different. figure 7 shows

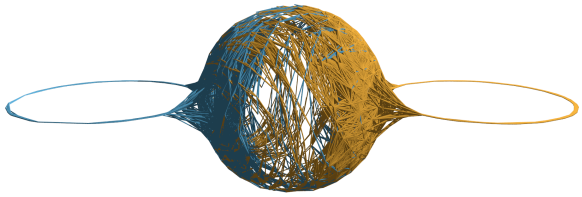


Figure 6: The clustering from figure 5 pulled back to the 1-simplices of the VR complex from 3.1, which is here drawn in $\mathbb{R}^3$ using the coordinates of the points for visualization purposes only. The unclustered 1-simplices, 19254 in number, are not drawn.

the result. Observe that the sphere is now captured by its adjacent circle, and that the unclustered edges tend to be those near where the two circles intersect.

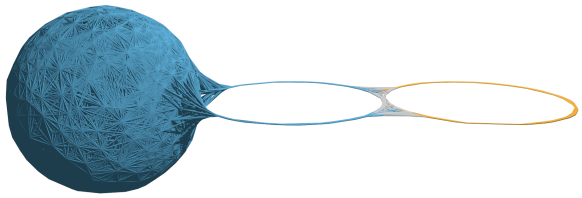


Figure 7: The result of clustering the rearranged point cloud from 3.1. Again the 1-simplices of the VR complex are clustered in a way respecting the generators of 1-homology. The unclustered simplices, 467 in number, are drawn in gray. (That the sphere appears solid is only a visualization artifact; the 2-simplices are not drawn.)

3.2 Punctured plane

In this experiment we uniformly randomly sample 1000 points from a unit square in $\mathbb{R}^2$ with three disks of radius 1/10 cut out. The points are seen as faint does in figure 8. We construct the alpha complex at parameter 0.1, and denote it by K in this section. It has Betti numbers $\beta_0(K) = 1$, $\beta_1(K) = 3$ and $\beta_{i>1}(K) = 0$. There are 2914 1-simplices and 1912 2-simplices. We again focus on the 1-simplices for clustering. The codomain of ψ is now $\mathbb{R}^3$, and the image is clustered according to three 1-dimensional linear subspaces. The result is shown in figure 8, and we again observe how the obtained clustering occurs with respect to the punctures of the square.

3.3 Torus

We next perform clustering of the edges of two different tori.

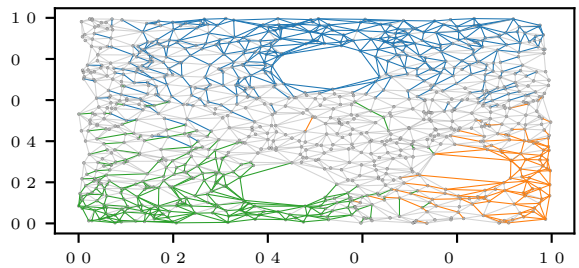


Figure 8: The point cloud of the experiment in 3.2 is shown as faint dots. The punctures can be seen in near $(0.5, 0.8)$, $(0.4, 0.2)$ and $(0.8, 0.3)$, and one observes that the clustered 1-simplices (blue, green, orange, respectively) follow the punctures. The gray 1-simplices are unclustered. The 2-simplices have not been drawn.

3.3.1 From a point cloud

We uniformly randomly sampled 1500 points from the unit square and map these under $(\varphi, \theta) \mapsto ((2 + \sin(2\pi\varphi)) \cos(2\pi\theta), (2 + \sin(2\pi\varphi)) \sin(2\pi\theta), \cos(2\pi\varphi))$ to produce a point sample of a torus in $\mathbb{R}^3$. The points were then given a uniformly random noise of amplitude 0.01 in both radii. Again a VR complex K was built, at scale 0.8. It has 35270 1-simplices and 377873 2-simplices, and has the homology of a torus, i.e. $\beta_0(K) = 1$, $\beta_1(K) = 2$, $\beta_2(K) = 1$. VR was chosen in order for the clustering task to be more complicated than in a more orderly alpha complex.

Figure 9 shows the image in $\mathbb{R}^2$ of K_1 under ψ. The subspaces for clustering are somewhat harder to make out than before, but they can still be found. The result of clustering by them can be seen in figure 10. Observe that the two clusters respect the two independent unfilled loops of the torus.

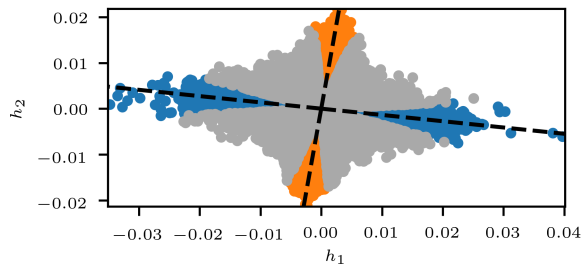


Figure 9: The clustering of the 1-simplices from the simplicial complex obtained from the sampled torus in the experiment in 3.3.1. The unclustered points are shown in gray. They are 23103 in number.

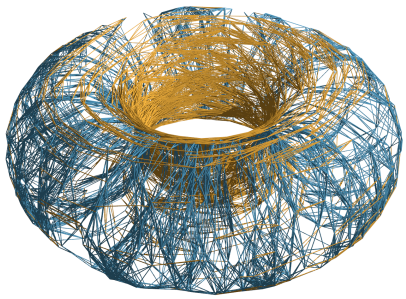


Figure 10: The clustering in figure 9 pulled back to 1-simplices of the torus from the experiment in 3.3.1. The unclustered ones are not shown, something which may make the torus appear broken.

3.3.2 A triangulation of the flat torus

As a smaller, more abstract and noise-free example, we consider a triangulation of a flat torus. The considered triangulation consists of a simplicial complex with 9 vertices, 27 1-simplices and 18 2-simplices. The image of its 1-simplices in $\mathbb{R}^2$ under ψ is shown in figure 11. The arrangement into a perfect hexagon means that there are in fact *three* subspaces that can be chosen for clustering. The clusters are shown in figure 12. The arrangement into a hexagon, and therefore the result of three instead of the expected two clusters, disappear if one breaks some of the symmetry in the triangulation, for example by having some of the diagonal edges go the opposite direction.

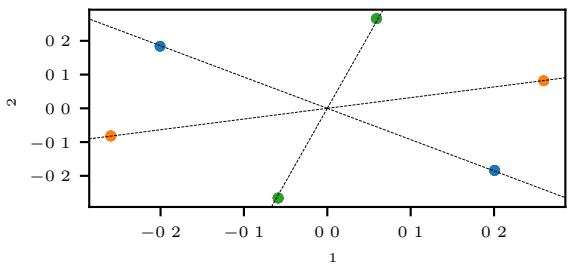


Figure 11: The clustering of the 1-simplices from the simplicial complex obtained as a triangulation of the flat torus from the experiment in 3.3.2. Note that many points overlap. Three clusters are given by points lying on three different linear subspaces.

3.4 Clustering 2-simplices

We have illustrated our method only on 1-simplices so far for ease of visualization. To point out that it also performs well in other dimensions, we

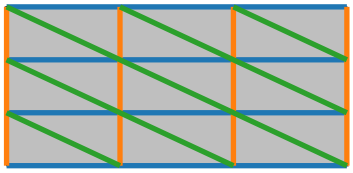


Figure 12: A triangulation of the flat torus represented as a rectangle with pairs of opposing edges identified. The 1-simpilces are clustered into three groups (orange, blue and green). Those in orange and blue are representatives of the two 1-homology classes of the complex, whereas the green ones are a linear combination of the others.

sampled 1000 points (each) from two spheres of radius 1 centered at $(-1, 0, 0)$ and $(1, 0, 0)$, each with a radial uniform random noise with amplitude 0.01. We computed the alpha complex K at parameter 0.3, so as to create a rather messy region between the spheres. There are 8851 1-simplices and 10478 2-simplices, and $\beta_0(K) = 1$, $\beta_1(K) = 0$ and $\beta_2(K) = 2$ as expected. Our clustering method performs as expected, producing clusters of K_2 that correspond to homological features, as is shown in figure 13.

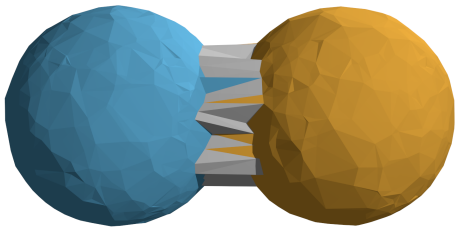


Figure 13: The output of our method when clustering the 2-simplices from the complex in the experiment in 3.4 are the blue and orange clusters. The 1485 gray simplices are unclustered.

3.5 Comparison with graph spectral clustering

It is worth comparing clustering obtained from our method with the ones obtained by clustering the nodes of the graph underlying each simplicial complex using the graph spectral clustering algorithm. Figure 14 shows the results of graph spectral clustering on the nodes of the graph underlying the complex in figure 7. The two first graph Laplacian eigenvectors were used to map the nodes into $\mathbb{R}^2$, and then k-means was used to find two clusters. Similarly, figure 15 displays three clusters on the nodes of the graph underlying the complex representing a punctured plane with three holes in figure 8. The nodes are mapped to $\mathbb{R}^3$ using the three first graph Laplacian

eigenvectors, after which k-means was used to find three clusters. In both cases we see that the clusters do not reflect any obviously meaningful property of the underlying data, unlike our method, which clusters in a way sensitive to homology.

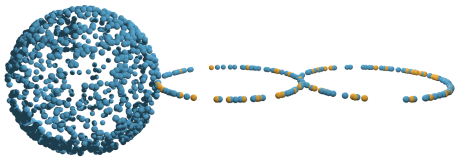


Figure 14: Graph spectral clustering of the vertices of the graph underlying the VR complex of figure 7.

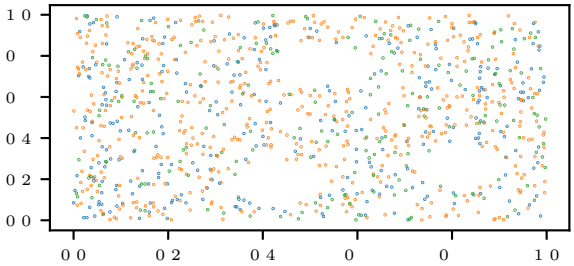


Figure 15: Graph spectral clustering of the vertices of the graph underlying the alpha complex from figure 8.

4 Conclusions and future work

In this paper we have presented a novel clustering method for simplicial complexes, one that is sensitive to the homology of the complex. We see the method as a contribution to an emerging field of spectral TDA [27, 2]. Our results suggest that the algorithm can be used to extract homological features for simplices of any degree. Experiments in various simplicial complexes demonstrate the ability of the method to accurately detect edges belonging to different non-bounding cycles. Similar results, not shown in this article for practical considerations of visualization, have been obtained by clustering simplices in higher dimensions. The sub-problem of the structure of the linear subspaces in the image of ψ, and how to accurately cluster based on demand, require further investigation of both a mathematical and a algorithmic nature.

While our method seems robust to noise in the underlying data, a more thorough investigation into the output's dependence on noise, and the output's

dependence on the scale at which a point-cloud-derived simplicial complex is built, is warranted.

Moreover, it has not eluded us that the method as outlined is not restricted to clustering just simplices. Other finitely generated chain complexes, such as discrete Morse complexes or cubical complexes, naturally lend themselves to the same analysis. One may also want to consider if there are theoretical implications even in the smooth case.

Further development will also include enlarging the target of the projection in ψ to include non-zero eigenvectors of $\mathcal{L}_p$, as in graph spectral clustering. Preliminary results indicate that this yields a further refinement of the homologically sensitive clusters into "fair" subdivisions.

Finally, further work needs to explore the effects of weighting. Both structural weighting, i.e. deriving weights from the local connectivity properties of the complex, as is often done with graph spectral clustering, and weighting originating from the underlying data itself, as is common in TDA.

A potential future application that we suspect fits our method well is collaboration networks [32], where n-fold collaborations clearly cannot accurately be encoded as $\binom{n}{2}$ pairwise ones.

Acknowledgments

Alpha complexes and persistent homology were computed using *GUDHI* [39]. Eigenvector computations were done with *SLEPc* [19].

We would like to thank K. Hess for valuable discussions.

Both authors were supported by the Swiss National Science Foundation grant number 200021_172636.

References

[1] Nataraj Akkiraju, Herbert Edelsbrunner, Michael Facello, Ping Fu, EP Mücke, and Carlos Varela. "Alpha shapes: definition and software". In: *Proceedings of the 1st International Computational Geometry Software Workshop*. Vol. 63. 1995, p. 66.

[2] Sergio Barbarossa, Stefania Sardellitti, and Elena Ceci. "Learning from signals defined over simplicial complexes". In: *2018 IEEE Data Science Workshop (DSW)*. IEEE. 2018, pp. 51–55.

[3] Mikhail Belkin and Partha Niyogi. "Laplacian eigenmaps and spectral techniques for embedding and clustering". In: *Advances in neural information processing systems*. 2002, pp. 585–591.

[4] Pavel Berkhin. "A Survey of Clustering Data Mining Techniques". In: *Grouping Multidimensional Data. Springer, Berlin, Heidelberg* (2006).

[5] Osting Braxton, Palande Sourabh, and Wang Bei. *Towards Spectral Sparsification of Simplicial Complexes Based on Generalized Effective Resistance*. 2017. arXiv: `1708.08436`.

[6] Gunnar Carlsson. "Topology and Data". In: *Bull. Amer. Math. Soc.* 46.2 (2009), pp. 255–308.

[7] Frederic Chazal and Bertrand Michel. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists*. 2017. arXiv: `1710.04019`.

[8] Jeff Cheeger. "A lower bound for the smallest eigenvalue of the Laplacian". In: *Proceedings of the Princeton conference in honor of Professor S. Bochner*. 1969, pp. 195–199.

[9] Fan R. K. Chung. *Spectral graph theory*. Vol. 92. Regional Conference Series in Mathematics. American Mathematical Society, 1997.

[10] Dragoš M. Cvetković, Michael Doob, and Horst Sachs. *Spectra of Graphs. Theory and Application*. Pure and Applied Mathematics. Academic Press, 1979.

[11] Beno Eckmann. "Harmonische Funktionen und Randwertaufgaben in einem Komplex". In: *Commentarii Mathematici Helvetici* 17.1 (1944), pp. 240–255.

[12] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[13] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. "Topological persistence and simplification". In: *Proceedings 41st Annual Symposium on Foundations of Computer Science*. IEEE. 2000, pp. 454–463.

[14] Miroslav Fiedler. "Algebraic connectivity of graphs". In: *Czechoslovak mathematical journal* 23.2 (1973), pp. 298–305.

[15] Robert Ghrist. "Barcodes: the persistent topology of data". In: *Bulletin of the American Mathematical Society* 45.1 (2008), pp. 61–75.

[16] Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. "Clique topology reveals intrinsic geometric structure in neural correlations". In: *Proceedings of the National Academy of Sciences* 112.44 (2015), pp. 13455–13460.

[17] Anna Gundert and May Szedlák. "Higher dimensional Cheeger inequalities". In: *Proceedings of the thirtieth annual symposium on Computational geometry*. 2014, p. 181.

[18] Allen Hatcher. *Algebraic Topology.* Cambridge University Press, 2002.

[19] Vicente Hernandez, Jose E. Roman, and Vicente Vidal. "SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems". In: *ACM Trans. Math. Software* 31.3 (2005), pp. 351–362.

[20] Anil N. Hirani. "Discrete Exterior Calculus". PhD thesis. California Insitute of Technology, 2003.

[21] Yasuaki Hiraoka, Takenobu Nakamura, Akihiko Hirata, Emerson G. Escolar, Kaname Matsue, and Yasumasa Nishiura. "Hierarchical structures of amorphous solids characterized by persistent homology". In: *Proceedings of the National Academy of Sciences* 113.26 (2016), pp. 7035–7040.

[22] Danijela Horak and Jürgen Jost. "Spectra of combinatorial Laplace operators on simplicial complexes". In: *Advances in Mathematics* 244 (2013), pp. 303–336.

[23] Aapo Hyvärinen and Erkki Oja. "Independent component analysis: algorithms and applications". In: *Neural networks* 13.4-5 (2000), pp. 411–430.

[24] Karin Kailing, Hans-Peter Kriegel, and Peer Kröger. "Density-connected subspace clustering for high-dimensional data". In: *Proceedings of the 2004 SIAM international conference on data mining.* SIAM. 2004, pp. 246–256.

[25] Gustav Kirchhoff. "Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird". In: *Annalen der Physik* 148.12 (1847), pp. 497–508.

[26] Ib Madsen and Jørgen Tornehave. *From calculus to cohomology: de Rham cohomology and characteristic classes.* Cambridge University Press, 1997.

[27] Joshua L Mike and Jose A Perea. *Geometric Data Analysis Across Scales via Laplacian Eigenvector Cascading.* 2018. arXiv: `1812.02139`.

[28] Terrence J. Moore, Robert J. Drost, Prithwish Basu, Ram Ramanathan, and Anantharam Swami. "Analyzing collaboration networks using simplicial complexes: A case study". In: *Proceedings IEEE INFOCOM Workshops* (2012), pp. 238–243.

[29] Sayan Mukherjee and John Steenbergen. "Random walks on simplicial complexes and harmonics". In: *Random structures & algorithms* 49.2 (2016), pp. 379–405.

[30] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. "On Spectral Clustering: Analysis and an algorithm". In: *Advances In Neural Information Processing Systems.* MIT Press, 2001, pp. 849–856.

[31] Ori Parzanchevski, Ron Rosenthal, and Ran J. Tessler. "Isoperimetric inequalities in simplicial complexes". In: *Combinatorica* 36.2 (2016), pp. 195–227.

[32] Alice Patania, Giovanni Petri, and Francesco Vaccarino. "The shape of collaborations". In: *EPJ Data Science* 6.1 (2017).

[33] Michael W. Reimann et al. "Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function". In: *Frontiers in Computational Neuroscience* 11 (2017), p. 48.

[34] Peter H A Sneath. "The application of computers to taxonomy". In: *Journal of General Microbiology* 17 (1957), pp. 201–226.

[35] Thorvald Sørensen. "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons". In: *Biologiske Skrifter* 5.4 (1948), pp. 1–34.

[36] John Steenbergen, Caroline Klivans, and Sayan Mukherjee. "A Cheeger-type inequality on simplicial complexes". In: *Advances in Applied Mathematics* 56 (2014), pp. 56–77.

[37] Hugo Steinhaus. "Sur la division des corp materiels en parties". In: *Bull. Acad. Polon. Sci* 3.4 (1957), pp. 801–804.

[38] Arthus Szlam and Xavier Bresson. "Total variation and Cheeger cuts". In: *Proceedings of the 27th International Conference on Machine Learning*. 2010, pp. 1039–1046.

[39] The GUDHI Project. *GUDHI User and Reference Manual*. GUDHI Editorial Board, 2015. URL: `http://gudhi.gforge.inria.fr/doc/latest/`.

[40] Leopold Vietoris. "Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen". In: *Mathematische Annalen* 97.1 (1927), pp. 454–472.

[41] Ulrike von Luxburg. "A tutorial on spectral clustering". In: *Statistics and computing* 17.4 (2007), pp. 395–416.

[42] Afra Zomorodian and Gunnar Carlsson. "Computing persistent homology". In: *Discrete & Computational Geometry* 33.2 (2005), pp. 249–274.

Chapter 3

Articles in Directed Topology

3.1 Towards Directed Collapsability

(joint work with Robin Belton, Robyn Brooks, Lisbeth Fajstrup, Brittany Terese Fasy, Catherine Ray, Nicole Sanderson, and Elizabeth Vidaurre)

Published in Advances in Mathematical Sciences, AWM Series, Springer, 2020.

Abstract

In the directed setting, the spaces of directed paths between fixed initial and terminal points are the defining feature for distinguishing different directed spaces. The simplest case is when the space of directed paths is homotopy equivalent to that of a single path; we call this the *trivial space of directed paths*. Directed spaces that are topologically trivial may have non-trivial spaces of directed paths, which means that information is lost when the direction of these topological spaces is ignored. We define a notion of directed collapsibility in the setting of a directed Euclidean cubical complex using the spaces of directed paths of the underlying directed topological space, relative to an initial or a final vertex. In addition, we give sufficient conditions for a directed Euclidean cubical complex to have a contractible or a connected space of directed paths from a fixed initial vertex. We also give sufficient conditions for the path space between two vertices in a Euclidean cubical complex to be disconnected. Our results have applications to speeding up the verification process of concurrent programming and to understanding partial executions in concurrent programs.

1 Introduction

Spaces that are equipped with a direction have only recently been given more attention from a topological point of view. The spaces of directed paths are the defining feature for distinguishing different directed spaces. One reason for studying directed spaces is their application to the modeling of concurrent programs, where standard algebraic topology does not provide the tools needed [4]. Concurrent programming is used when multiple processes need to access shared resources. Directed spaces are models for concurrent program, where paths respecting the time directions represent executions of programs. In such models, executions are equivalent if their execution paths are homotopic through a family of directed paths. This observation has already led to new insights and algorithms. For instance, verification of concurrent programs is simplified by verifying one execution from each connected component of the space of directed paths; see [4, 5].

While equivalence of executions is clearly stated in concurrent programming, equivalence of the directed topological spaces themselves is not well understood. Directed versions of homotopy groups and homology groups are not agreed upon. Directed homeomorphism is too strong; whereas, directed homotopy equivalence is often too weak, to preserve the properties of the concurrent programs. In classical (undirected) topology, the concept of simplifying a space by a sequence of collapses goes back to J.H.C. Whitehead [11], and has been studied in [1, 6], among others. However, a definition for a directed collapse of a Euclidean cubical complex that preserves spaces of directed paths is notably missing from the literature.

In this article, we consider spaces of directed paths in Euclidean cubical complexes. Our objects of study are spaces of directed paths relative to a fixed pair of endpoints. We show how local information of the past links of vertices in a Euclidean cubical complex can provide global information on the spaces of directed paths. As an example, our results are applied to study the spaces of directed paths in the well-known dining philosophers problem. Furthermore, we define directed collapse so that a directed collapse of a Euclidean cubical complex preserves the relevant spaces of directed paths in the original complex. Our theoretical work has applications to simplifying verification of concurrent programs without loops, and better understanding partial executions in those concurrent programs.

We begin, in Section 2, with two motivating examples of how the execu-

tion of concurrent programs can be modeled by Euclidean cubical complexes and directed path spaces. In Section 3, we introduce the notions of spaces of directed paths and Euclidean cubical complexes. Given the directed structure of these Euclidean cubical complexes, we do not consider the link of a vertex but the *past* link of it. In Section 4, we give results on the topology of the spaces of directed paths from an initial vertex to other vertices in terms of past links. Theorem 4.1 gives sufficient conditions on the past links of every vertex of a complex so that spaces of directed paths are contractible. Theorem 4.2 gives conditions that are sufficient for the spaces of directed paths to be connected. In Theorem 4.8, we give sufficient conditions on the past link of a vertex so that the space of directed paths from the initial vertex to that vertex is disconnected. In Section 5, we describe a method of collapsing one complex into a simpler complex, while preserving the directed path spaces.

2 Concurrent Programs and Directed Path Spaces

We illustrate how to organize possible executions of concurrent programs using Euclidean cubical complexes and directed spaces. An execution is a scheduling of the events that occur in a program in order to compute a specific task. In Example 2.1, we describe the dining philosophers problem. In Example 2.2, we illustrate how to model executions of concurrent programs in the context of the dining philosophers problem in the case of two philosophers.

Example 2.1 (Dining Philosophers)**.** The dining philosophers problem originally formulated by E. Dijkstra [2] and reformulated by T. Hoare [7] illustrates issues that arise in concurrent programs. Consider n philosophers sitting at a round table ready to eat a meal. Between each pair of neighboring philosophers is a chopstick for a total of n chopsticks. Each philosopher must eat with the two chopsticks lying directly to the left and right of her. Once the philosopher is finished eating, she must put down both chopsticks. Since there are only n chopsticks, the philosophers must share the chopsticks in order for all of them to eat. The dining philosopher problem is to design a concurrent program where all n philosophers are able to eat once for some finite amount of time.

A design of a program is a choice of actions for each philosopher. One

example of a design of a program is where each of the n philosophers does the following:

1. Wait until the right chopstick is available, then pick it up.
2. Wait until the left chopstick is available, then pick it up.
3. Eat for some finite amount of time.
4. Put down the left chopstick.
5. Put down the right chopstick.

While correct executions of this program are possible (e.g., where the philosophers take turns eating alone), this design has states in which every philosopher has picked up the chopstick to her right and is waiting for the other chopstick. Such a situation exemplifies a *deadlock* in concurrent programming, an execution that gets "stuck" and never finishes.

The design described above also has states that cannot occur. For example, consider the dining philosophers problem when $n = 2$. The state in which both philosophers are finished eating and one is still holding onto chopstick a while the other is holding chopstick b would imply that a philosopher was able to eat with only one chopstick—an example of an *unreachable* state in concurrent programming.

The dining philosophers problem illustrates the difficulties in designing concurrent programs. Difficulties arise since each philosopher must use chopsticks that must be shared with the neighboring philosophers. Analogously, in concurrent programming, multiple processes must access shared resources that have a finite capacity.

The next example illustrates how to model executions of the dining philosophers problem with a Euclidean cubical complex. When the problem consists of two philosophers, the Euclidean cubical complex used to model the dining philosophers problem is often referred to as the *Swiss Flag*.

Example 2.2 (Swiss Flag)**.** In the language of concurrent programming, the two philosophers represent two processes denoted by T_1 and T_2. The two chopsticks represent shared resources denoted by a and b. One process is executing the program $P_aP_bV_bV_a$ and the other process is executing the program $P_bP_aV_aV_b$. Here, P means that a process has a *lock* on that resource

while V means that a process *releases* a resource. To model this concurrent program with a Euclidean cubical complex, we construct a 5×5 grid where the x-axis is labeled by $P_a P_b V_b V_a$, each a unit apart, and the y-axis is labeled by $P_b P_a V_a V_b$, each also a unit apart (see Figure 1). The region $[1,4] \times [2,3]$ represents when both T_1 and T_2 have a lock on a. In the dining philosophers problem, a single chopstick can only be held by one philosopher at a given time. The mutual exclusion of the chopsticks translates to the shared resources, a and b, each having *capacity* one, where the capacity of a resource is the number of processes that can have access to the resource simultaneously. We call the region $[1,4] \times [2,3]$ *forbidden* since T_1 and T_2 cannot have a lock on a at the same time. The region $[2,3] \times [1,4]$ represents when both T_1 and T_2 have a lock on b. This region is also forbidden. The set complement of the interior of $[1,4] \times [2,3] \cup [2,3] \times [1,4]$ in $[0,5] \times [0,5]$ is called the Swiss flag and is the Euclidean cubical complex modeling this program design for the dining philosophers problem.

In general, the Euclidean cubical complex modeling a concurrent program is the complement of the interior of the forbidden region. An execution is a directed path from the initial point to the terminal point. Executions are equivalent if they give the same output given the same input, which can be interpreted geometrically as the corresponding paths are dihomotopic in the path space. The Swiss flag has two distinct directed paths up to homotopy equivalence: one corresponding to T_1 using the shared resources first, and the other corresponding to T_2 using the shared resources first. See Figure 1.

3 Past Links as Obstructions

In this section, we introduce the notions of spaces of directed paths and Euclidean cubical complexes. The (relative) past link of a vertex of a Euclidean cubical complex is defined as a simplicial complex. Studying the contractibility and connectedness of past links gives us insight on the contractibility and connectedness of certain spaces of directed paths.

Definition 3.1 (d-space)**.** A *d-space* is a pair $(X, \overrightarrow{P}(X))$, where X is a topological space and $\overrightarrow{P}(X) \subseteq P(X) := X^{[0,1]}$ is a family of paths on X (called *dipaths*) that is closed under non-decreasing reparametrizations and concatenations, and contains all constant paths.

For every x, y in X, let $\overrightarrow{P}_x^y(X)$ be the family of *dipaths from x to y*:

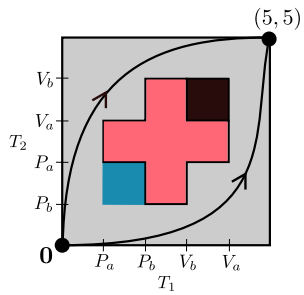


Figure 1: The Swiss Flag. . The pink region is the forbidden region. Any bi-monotone path outside of F is a possible execution. The set of all executions of two processes, T_1 and T_2, is called the *state-space* Two regions in the state space are of particular interest. The black region is the set of all unreachable states, and the blue region is the set of all states that are doomed to never complete. A state is doomed if any path starting at that state leads to a deadlock. The black curves in the figure are two possible paths in this directed space.

$$\overrightarrow{P}_x^y(X) := \{\alpha \in \overrightarrow{P}(X) : \alpha(0) = x \text{ and } \alpha(1) = y\}.$$

In particular, consider the following directed space: the *directed real line* $\overrightarrow{\mathbb{R}}$ is the directed space constructed from the real line whose family of dipaths $\overrightarrow{P}(\mathbb{R})$ consists of all non-decreasing paths. The *Euclidean space* $\overrightarrow{\mathbb{R}^n}$ is the n-fold product $\overrightarrow{\mathbb{R}} \times \cdots \times \overrightarrow{\mathbb{R}}$ with family of dipaths the n-fold product $\overrightarrow{P}(\mathbb{R}^n) = \overrightarrow{P}(\mathbb{R}) \times \cdots \times \overrightarrow{P}(\mathbb{R})$.

Furthermore, we can solely focus on the family of dipaths in a d-space and endow it with the compact open topology.

Definition 3.2 (Space of Directed Paths)**.** In a d-space $(X, \overrightarrow{P}(X))$, the *space of directed paths* from x to y is the family $\overrightarrow{P}_x^y(X)$ with the compact open topology.

By topologizing the space of directed paths, we may now use topological reasoning and comparison. Since $\overrightarrow{P}_x^y(X)$ does not have directionality, contractibility and other topological features are defined as in the classical case. Moreover, observe that the set $\overrightarrow{P}_x^y(X)$ might have cardinality of the continuum, but is considered trivial if it is homotopy equivalent to a point.

The d-spaces that we consider in this article are constructed from Euclidean cubical complexes. Let $\mathbf{p} = (p_1, \ldots, p_n), \mathbf{q} = (q_1, \ldots, q_n) \in \mathbb{R}^n$. We

write $\mathbf{p} \preceq \mathbf{q}$ if and only if $p_i \leq q_i$ for all $i = 1, \ldots, n$. Furthermore, we denote by $\mathbf{q} - \mathbf{p} := (q_1 - p_1, \ldots, q_n - p_n)$ the component-wise difference between $\mathbf{q}$ and $\mathbf{p}$, $|\mathbf{p}| := \sum_{i=1}^{n} p_i$ is the element-wise sum, or one-norm, of $\mathbf{p}$. Similarly to the one-dimensional case, the interval $[\mathbf{p}, \mathbf{q}]$ is defined as $\{\mathbf{x} \in \mathbb{R}^n : \mathbf{p} \preceq \mathbf{x} \preceq \mathbf{q}\}$.

Definition 3.3 (Euclidean Cubical Complex)**.** Let $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n$. If $\mathbf{q}, \mathbf{p} \in \mathbb{Z}^n$ and $\mathbf{q} - \mathbf{p} \in \{0, 1\}^n$, then the interval $[\mathbf{p}, \mathbf{q}]$ is an *elementary cube in* $\mathbb{R}^n$ of dimension $|\mathbf{q} - \mathbf{p}|$. A *Euclidean cubical complex* $K \subseteq \mathbb{R}^n$ is the union of elementary cubes.

Remark 3.4. A Euclidean cubical complex K is a subset of $\mathbb{R}^n$ and it has an associated abstract cubical complex. By a slight abuse of notation, we do not distinguish these.

Every cubical complex K inherits the directed structure from the Euclidean space $\overrightarrow{\mathbb{R}^n}$, described after Definition 3.1. An elementary cube of dimension d is called a d-cube. The m-skeleton of K, denoted by K_m, is the union of all elementary cubes contained in K that have dimension less than or equal to m. The elements of the zero-skeleton are called the vertices of K. A vertex $\mathbf{w} \in K_0$ is said to be *minimal* (resp., *maximal*) if $\mathbf{w} \preceq \mathbf{v}$ (resp., $\mathbf{w} \succeq \mathbf{v}$) for every vertex $\mathbf{v} \in K_0$.

Following [12], we define the (relative) past link of a vertex of a Euclidean cubical complex as a simplicial complex. Let Δ^{n-1} denote the complete simplicial complex with vertices $\{1, \ldots, n\}$. Simplices of Δ^{n-1} is be identified with elements $\mathbf{j} \in \{0, 1\}^n$. That is, every subset $S \subseteq \{1, \ldots, n\}$ is mapped to the n-tuple with entry 1 in the k-th position if k belongs to S and 0 otherwise. The topological space associated to the simplicial complex Δ^{n-1} is the one given by its geometric realization.

Definition 3.5 (Past Link)**.** In a Euclidean cubical complex K in $\mathbb{R}^n$, the *past link*, $lk^-_{K,\mathbf{w}}(\mathbf{v})$, of a vertex $\mathbf{v}$, with respect to another vertex $\mathbf{w}$ is the simplicial subcomplex of Δ^{n-1} defined as follows: $\mathbf{j} \in lk^-_{K,\mathbf{w}}(\mathbf{v})$ if and only if $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K \cap [\mathbf{w}, \mathbf{v}]$.

Remark 3.6. While K is a *cubical* complex, the past link of a vertex in K is always a *simplicial* complex.

Remark 3.7. Often the vertex $\mathbf{w}$ and the complex K are understood. In this case, we denote the past link of $\mathbf{v}$ by $lk^-(\mathbf{v})$.

Remark 3.8. Other definitions of the (past) link are found in the literature. Unlike Definition 3.5, (past) links are usually subcomplexes of K. However, the (past) links found in other literature are homeomorphic to the (past) link of Definition 3.5.

In the following example, we show that a vertex $\mathbf{v}$ can have past links with different homotopy type depending on what the initial vertex $\mathbf{w}$ is. We consider as a Euclidean cubical complex the open top box (Figure 2) and the past links of the vertex $\mathbf{v} = (1, 1, 1)$, with respect to the vertices $\mathbf{w} = \mathbf{0}$ and $\mathbf{w}' = (0, 0, 1)$.

Example 3.9 (Open Top Box)**.** Let $L \subset \mathbb{R}^3$ be the Euclidean cubical complex consisting of all of the edges and vertices in the elementary cube $[\mathbf{0}, \mathbf{v}]$ and five of the six two-cubes, omitting the elementary two-cube $[(0, 0, 1), \mathbf{v}]$, i.e., the top of the box. Because the elementary one-cube $[\mathbf{v} - (0, 0, 1), \mathbf{v}] \subseteq L \cap [\mathbf{0}, \mathbf{v}] = L$, $lk^-_{L,\mathbf{0}}(\mathbf{v})$ contains the vertex in Δ^2 corresponding to $\mathbf{j} = (0, 0, 1)$. Similarly, because the elementary two-cube $[\mathbf{v} - (0, 1, 1), \mathbf{v}] \subseteq L$, the past link $lk^-_{L,\mathbf{0}}(\mathbf{v})$ contains the edge in Δ^2 corresponding to $\mathbf{j} = (0, 1, 1)$. However, because the elementary two-cube $[\mathbf{v} - (1, 1, 0), \mathbf{v}]$ is not contained in L, $lk^-_{L,\mathbf{0}}(\mathbf{v})$ does not include the edge corresponding to $\mathbf{j} = (1, 1, 0)$. Instead taking the initial vertex to be $\mathbf{w} = (0, 0, 1)$, we get that $lk^-_{L,\mathbf{w}}(\mathbf{v})$ consists of the two vertices corresponding to $\mathbf{j} = (0, 1, 0)$ and $\mathbf{j}' = (1, 0, 0)$. See Figure 2.

4 The Relationship Between Past Links and Path Spaces

In this section, we illustrate how to use past links to study spaces of directed paths with an initial vertex of $\mathbf{0}$. In particular, the contractibility and connectedness of all past links guarantees the contractibility and connectedness of spaces of directed paths. We also provide a partial converse to the result concerning connectedness.

Theorem 4.1 (Contractibility)**.** *Let $K \subset \mathbb{R}^n$ be a Euclidean cubical complex with minimal vertex $\mathbf{0}$. Suppose for all $\mathbf{k} \in K_0$, $\mathbf{k} \neq \mathbf{0}$, the past link $lk^-_{\mathbf{0}}(\mathbf{k})$ is contractible. Then, all spaces of directed paths $\overrightarrow{P}^{\mathbf{k}}_{\mathbf{0}}(K)$ are contractible.*

Proof. By [12, Prop. 5.3], if $\overrightarrow{P}^{\mathbf{k}-\mathbf{j}}_{\mathbf{0}}(K)$ is contractible for all $\mathbf{j} \in \{0, 1\}^n$, $\mathbf{j} \neq \mathbf{0}$, and $\mathbf{j} \in lk^-(\mathbf{k})$, then $\overrightarrow{P}^{\mathbf{k}}_{\mathbf{0}}(K)$ is homotopy equivalent to $lk^-(\mathbf{k})$. Hence, it

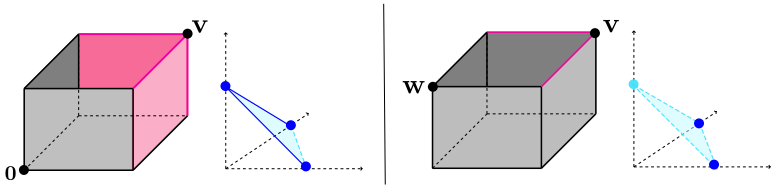


Figure 2: The Open Top Box. Left: the open top box and the geometric realization of the past link of the red vertex $\mathbf{v} = (1, 1, 1)$, with respect to the black vertex $\mathbf{0}$. The geometric realization of $lk^-_{L,\mathbf{0}}(\mathbf{v})$ contains two edges of a triangle, since the two red faces are included in $[\mathbf{0}, \mathbf{v}]$ and three vertices, since the three red edges are included in $[\mathbf{0}, \mathbf{v}]$. Right: the open top box and the geometric realization of the past link of the red vertex $\mathbf{v} = (1, 1, 1)$, with respect to the black vertex $\mathbf{w} = (0, 0, 1)$. The geometric realization of $lk^-_{L,\mathbf{w}}(\mathbf{v})$ consists only of two vertices of a triangle, since the two red edges are included in $[\mathbf{w}, \mathbf{v}]$.

suffices to see that all the spaces $\overrightarrow{P}^{\mathbf{k}-\mathbf{j}}_{\mathbf{0}}(K)$ are contractible. This follows by structural induction on the partial order on vertices in K.

The start is at $\overrightarrow{P}^{\mathbf{0}+\mathbf{e}_i}_{\mathbf{0}}(K)$, where $\mathbf{e}_i$ is the i-th unit vector, and $\mathbf{0} + \mathbf{e}_i \in K_0$. If the edge $[\mathbf{0}, \mathbf{0} + \mathbf{e}_i]$ is in K, then $\overrightarrow{P}^{\mathbf{0}+\mathbf{e}_i}_{\mathbf{0}}(K)$ is contractible. Otherwise, $lk^-_{\mathbf{0}}(\mathbf{0} + \mathbf{e}_i)$ is empty, which contradicts the hypothesis that all of the past links are contractible. By structural induction, using also that $\overrightarrow{P}^{\mathbf{0}}_{\mathbf{0}}$ is contractible, the theorem now holds. □

Now, we give an analogous sufficient condition for when spaces of directed paths are connected. We provide two different proofs of Theorem 4.2. The first proof shows how we can use [9, Prop. 2.20] to get our desired result. The second proof uses notions from category theory and is based on the fact that the colimit of connected spaces over a connected category is connected.

Theorem 4.2 (Connectedness)**.** *With K as above, suppose all past links $lk^-_{\mathbf{0}}(\mathbf{k})$ of all vertices $\mathbf{k} \neq \mathbf{0}$ are connected. Then, for all $\mathbf{k} \in K_0$, all spaces of directed paths $\overrightarrow{P}^{\mathbf{k}}_{\mathbf{0}}(K)$ are connected.*

In this first proof we show that [9, Prop. 2.20] is an equivalent condition to all past links being connected.

Proof. In [9, Prop. 2.20], a local condition is given that ensures that all spaces of directed paths to a certain final point are connected. Here, we explain how the local condition is equivalent to all past links being connected.

Their condition is in terms of the local future; however, we reinterpret this in terms of local past instead of local future. Since we consider all spaces of directed paths *from* a point (as opposed to *to* a point), then reinterpreting the result in terms of local past is the right setting we should look at. The local condition is the following: for each vertex, $\mathbf{v}$, and all pairs of edges $[\mathbf{v}-\mathbf{e}_r, \mathbf{v}]$, $[\mathbf{v}-\mathbf{e}_s, \mathbf{v}]$ in K, there is a sequence of two-cells $\{[\mathbf{v}-\mathbf{e}_{k_i}-\mathbf{e}_{l_i}, \mathbf{v}]\}_{i=1}^m$, each of which is in K such that $l_i = k_{i+1}$ for $i = 1, \ldots, m-1$, $k_1 = r$ and $l_m = s$. Now, we show that this local condition is equivalent to ours. In the past link considered as a simplicial complex, such a sequence of two-cells corresponds to a sequence of edges from the vertex r to the vertex s. For $x, y \in lk^-(\mathbf{v})$, they are both connected to a vertex via a line. And those vertices are connected. Hence, the past link is connected.

Vice versa: Suppose $lk^-(\mathbf{v})$ is connected. Let p, q be vertices in $lk^-(\mathbf{v})$ and let $\gamma : I \to lk^-(\mathbf{v}) \in \Delta^{n-1}$ be a path from p to q. The sequence of simplices traversed by γ, $S_1, S_2, \ldots, S_k$, satisfies $S_i \cap S_{i+1} \neq \emptyset$. Moreover, the intersection is a simplex. Let $p_i \in S_i \cap S_{i+1}$. A sequence of pairwise connected edges connecting p to q is constructed by such sequences from p_i to p_{i+1} in S_{i+1} thus providing a sequence of two-cells similar to the requirement in [9]. Hence, by [9], if all past links of all vertices are connected, then all $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{k}}$ are connected □

This second proof of Theorem 4.2 has a more categorical flavor.

Proof. We give a more categorical argument which is closer to the proof of Theorem 4.1. In [10, Prop. 2.3 and Equation 2.2], the space of directed paths $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{k}}$ is given as a colimit over $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{k}-\mathbf{j}}$. The indexing category is $\mathcal{J}_{\mathcal{K}}$ with objects $\{\mathbf{j} \in \{0,1\}^n : [\mathbf{k}-\mathbf{j}] \subseteq K\}$ and morphisms $\mathbf{j} \to \mathbf{j}'$ for $\mathbf{j} \geq \mathbf{j}'$ given by inclusion of the simplex $\Delta^{\mathbf{j}} \subset \Delta^{\mathbf{j}'}$. The geometric realization of the index category is the past link which with our requirements is connected. The colimit of connected spaces over a connected category is connected. Hence, by induction as above, beginning with edges from $\mathbf{0}$, the directed paths $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{k}-\mathbf{j}}$ are all connected and the conclusion follows. □

Remark 4.3. Our conjecture is that similar results for k-connected past links should follow from the k-connected Nerve Lemma.

Remark 4.4. The statements of both Theorem 4.1 and Theorem 4.2 concern past links and path spaces defined with respect to a fixed initial vertex.

To see why past links depend on their initial vertex, consider the open top box of Example 3.9. All past links in L with respect to the initial vertex $\mathbf{0}$ are contractible, but $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{w}'}(L)$, where $\mathbf{w}' = (0,0,1)$ and $\mathbf{v} = (1,1,1)$, is not contractible. It is in fact two points. Note, this does not contradict Theorem 4.1, which only asserts that $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{0}}(L)$ is contractible; see Figure 2.

We now show how Theorem 4.1 and Theorem 4.2 can be used to study the spaces of the directed paths in slight modifications of the dining philosophers problem.

Example 4.5 (Three Concurrent Processes Executing the Same Program)**.** We consider a modification of Example 2.1 where we have three processes and two resources each with capacity two. All processes are executing the program $P_aP_bV_bV_a$. The Euclidean cubical complex modeling this situation has three dimensions, each representing the program of a process. Since each resource has capacity two, it is not possible to have a three way lock on any of the resources. The three processes have a lock on a in the region $[P_a, V_a]^{\times 3}$, which is the cube $[(1,1,1),(4,4,4)]$. Similarly, the three processes have a lock on b in the region $[P_b, V_b]^{\times 3}$ which is the cube $[(2,2,2),(3,3,3)]$. The forbidden region is the union of these two sets which is $[(1,1,1),(4,4,4)]$. We can model this concurrent program as a three-dimensional Euclidean cubical complex and the forbidden region is the inner $3 \times 3 \times 3$ cube.

In order to analyze the connectedness and contractibility of the spaces of directed paths with initial vertex $\mathbf{0}$, we study the past links of the vertices of K. First, we show that not all past links are contractible. Let $\mathbf{v} = (4,4,4)$. Then, $lk^-_{K,\mathbf{0}}(\mathbf{v})$ consists of all $\mathbf{j} \in \{0,1\}^3$ except $(1,1,1)$. The past link does not contain $(1,1,1)$ because the cube $[(3,3,3),(4,4,4)]$ is not contained in K, but $[\mathbf{v}-\mathbf{j}, \mathbf{v}] \subset K$ for all other $\mathbf{j}$. Therefore, $lk^-_{K,\mathbf{0}}(\mathbf{v})$ is the boundary of the two simplex (see Figure 3). Because the boundary of the two simplex is not contractible, the hypothesis of Theorem 4.1 is not satisfied. Hence, we cannot use Theorem 4.1 to study the contractibility of the spaces of directed paths.

Next, we show that all past links are connected. If we directly compute the past link $lk^-_{K,\mathbf{0}}(\mathbf{k})$ for all $\mathbf{k} \in K_0$, we find that the past link consists of either a zero simplex, one simplex, the boundary of the two simplex, or a two simplex. All these past links are connected. Theorem 4.2 implies that for all $\mathbf{k} \in K_0$, the space of directed paths, $\overrightarrow{P}^{\mathbf{k}}_{\mathbf{0}}(K)$ is connected.

We can generalize this example to n processes and two resources with

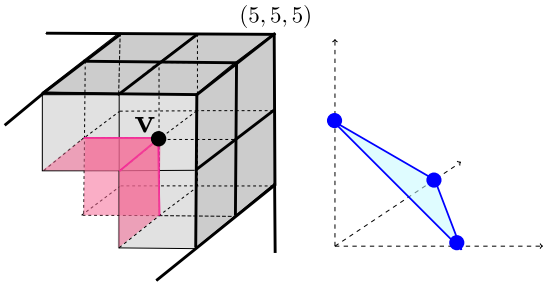


Figure 3: Three processes, same program. Illustrating $lk^-_{K,\mathbf{0}}(\mathbf{v})$ where K is the cube $[\mathbf{0},(5,5,5)]$ minus the inner cube, $[(1,1,1),(4,4,4)]$, and $\mathbf{v} = (4,4,4)$. The geometric realization of the simplicial complex $lk^-_{K,\mathbf{0}}(\mathbf{v})$ is the boundary of the two simplex since the three pink faces and edges are included in $[\mathbf{0},\mathbf{v}]$.

capacity $n-1$ where all processes are executing the program $P_aP_bV_bV_a$. For all n, Theorem 4.2 shows that all spaces of directed paths are connected.

The converse of Theorem 4.2 is not true. To see this, and give the conditions under which the converse does hold, we need to introduce the following definition:

Definition 4.6 (Reachable)**.** The point $x \in K$ is *reachable* from $\mathbf{w} \in K_0$ if there is a path from $\mathbf{w}$ to x. A subcomplex of K is induced by the set of points that are reachable from a vertex $\mathbf{w}$.

Example 4.7 (Boundary of the $3\times3\times3$ Cube with Top Right Cube)**.** Let K be the Euclidean cubical complex that is the boundary of the $3\times3\times3$ cube along with the cube $[(2,2,2),(3,3,3)]$. Observe that all spaces of directed paths with initial vertex $\mathbf{0}$ are connected. However, K has a disconnected past link at $\mathbf{v} = (3,2,2)$. If we consider the subcomplex $\hat{K}$ that is reachable from $\mathbf{0}$, then $\hat{K}$ is the boundary of the $3\times3\times3$ cube. The past links of all vertices in $\hat{K}$ are connected. This motivates the conditions given in Theorem 4.8 of removing the unreachable points of a Euclidean cubical complex. The connected components of a disconnected past link in the remaining complex can then be represented by directed paths from the initial point and not only locally.

Theorem 4.8 (Realizing Obstructions)**.** *Let K be a Euclidean cubical complex with initial vertex $\mathbf{0}$. Let $\hat{K} \subset K$ be the subcomplex reachable from $\mathbf{0}$. If*

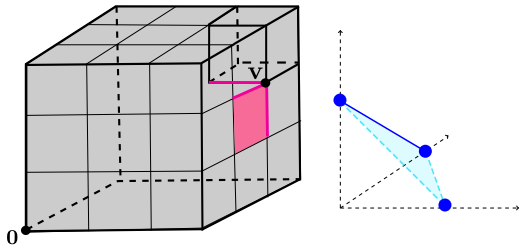


Figure 4: Motivating reachability condition. Let K be the boundary of the $3 \times 3 \times 3$ cube union with $[(2,2,2),(3,3,3)]$. Then, the geometric realization of the simplicial complex $lk^-_{K,\mathbf{0}}(\mathbf{v})$ is an edge and a point since the three pink edges and one face are included in $[(0,0,0),\mathbf{v}]$.

for $\mathbf{v} \in \hat{K}_0$, the past link in $\hat{K}$ is disconnected, then the path space $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{0}}(K)$ is disconnected.

Proof. Let $\mathbf{v}$ be a vertex such that $lk^-_{K,\mathbf{0}}(\mathbf{v})$ is disconnected and let $\mathbf{j}_1, \mathbf{j}_2$ be vertices in $lk^-_{\hat{K}}(v)$ in different components. The edges $[\mathbf{v}-\mathbf{j}_i, \mathbf{v}]$ are then in $\hat{K}$ and, in particular, $\mathbf{v}-\mathbf{j}_i \in \hat{K}_0$. Hence, there are paths $\mu_i : \overrightarrow{I} \to \hat{K}$ such that $\mu_i(0) = \mathbf{0}$ and $\mu_i(1) = \mathbf{v}-\mathbf{j}_i$.

By [3], there are $\hat{\mu}_i$ which are dihomotopic to μ_i and such that $\hat{\mu}_i$ is combinatorial, i.e., a sequence of edges in $\hat{K}$. Let γ_i be the concatenation of $\hat{\mu}_i$ with the edge $[\mathbf{v}-\mathbf{j}_i, \mathbf{v}]$.

Suppose for contradiction that γ_1 and γ_2 are connected by a path in $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{0}}(K)$. Let $H : \overrightarrow{I} \times I \to K$ be such a path with $H(t,0) = \gamma_1(t)$ and $H(t,1) = \gamma_2(t)$. Since $H(t,s)$ is reachable from $\mathbf{0}$, H maps to $\hat{K}$.

By [3], there is a combinatorial approximation $\hat{H} : \overrightarrow{I} \times I \to \hat{K}_2$ to the 2-skeleton of $\hat{K} \subset K$. Let B be the open ball centered around $\mathbf{v}$ with radius $1/2$. Since $\hat{H}$ is continuous, the inverse image of B under $\hat{H}$ is a neighborhood of $\{1\} \times I \subset \overrightarrow{I} \times I$. For $0 < \epsilon < 1/2$, this neighborhood contains a strip $(1-\epsilon, 1] \times I$ (by compactness of I). Then $\hat{H}(1-\epsilon/2 \times I)$ gives a path connecting the two edges $[\mathbf{v}-\mathbf{j}_i, \mathbf{v}]$. This path traverses a sequence of 2-cubes (the carriers). These correspond to a sequence of edges in the past link that connect $\mathbf{j}_1$ and $\mathbf{j}_2$, which contradicts the assumption that they are in different components. Therefore, γ_1 and γ_2 correspond to two points in $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{0}}(K)$ that are not connected by a path. $\square$

In general, the reachability condition in Theorem 4.8 eliminates the spurious disconnected past links that could appear in the unreachable parts of

a Euclidean cubical complex.

Example 4.9. To see how Theorem 4.8 can be applied, consider Example 2.2, the Swiss flag. The Swiss flag has two vertices with disconnected past links with respect to $\mathbf{0}$ namely $(4,3)$ and $(3,4)$. These disconnected past links imply that Theorem 4.2 is inconclusive. If the unreachable section of the Swiss flag is removed, we obtain a new Euclidean cubical complex in which the vertex $\mathbf{v} = (4,4)$ has a disconnected past link, consisting of two points. By Theorem 4.8, the path space $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{v}}(K)$ is also disconnected. In fact, $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{v}}(K)$ has two points, representing the dihomotopy classes of paths which pass above the forbidden region, and those paths which pass below.

The disconnected path space, $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{v}}(K)$, found in the previous example helps illustrate the following: given two vertices $\mathbf{w}$ and $\mathbf{v}$ in a Euclidean cubical complex K, if the path space $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K)$ is disconnected, then there exists a vertex in $[\mathbf{w}, \mathbf{v}]$ that has a disconnected past link with respect to $\mathbf{w}$ (the vertices $(4,3)$ and $(3,4)$ in the Swiss flag). If $\mathbf{w} = \mathbf{0}$, then we get the contrapositive of Theorem 4.2. If K is reachable from $\mathbf{0}$, Theorem 4.8 allows us to draw conclusions about the space of directed paths.

5 Directed Collapsibility

To simplify the underlying topological space of a d-space while preserving topological properties of the associated space of directed paths, we introduce the process of directed collapse. The criteria we require to perform directed collapse on Euclidean cubical complexes involves the topology of the past links of the vertices of the complex. We defined the past links as simplicial complexes that are not themselves directed, so our topological criteria are in the usual sense.

Definition 5.1 (Directed Collapse)**.** Let K be a Euclidean cubical complex with initial vertex $\mathbf{0}$. Consider $\sigma, \tau \in K$ such that $\tau \subsetneq \sigma$, σ is maximal, and no other maximal cube contains τ. Let $K' = K \setminus \{\gamma \in K | \tau \subseteq \gamma \subseteq \sigma\}$. K' is a *directed (cubical) collapse* of K if, for all $\mathbf{v} \in K'_0$, $lk^-_K(\mathbf{v})$ is homotopy equivalent to $lk^-_{K'}(\mathbf{v})$. The pair τ, σ is then called a *collapsing pair.*

K' is a *directed 0-collapse* of K if for all $\mathbf{v} \in K'_0$, $lk^-_K(\mathbf{v})$ is connected if and only if $lk^-_{K'}(\mathbf{v})$ is connected.

Remark 5.2. As in the simplicial case, when we remove σ from the abstract cubical complex, the effect on the geometric realization is to remove the interior of the cube corresponding to σ.

Remark 5.3. Note for finding collapsing pairs, (τ, σ), using Definition 5.1, with the geometric realization of σ given by the elementary cube, $[\mathbf{w}-\mathbf{j}, \mathbf{w}]$, it is sufficient to only check $\mathbf{v} \in K_0'$ such that $\mathbf{v} = \mathbf{w} - \mathbf{j}'$ where $\mathbf{j} - \mathbf{j}' > 0$. Otherwise the past links, $lk_K^-(\mathbf{v})$ and $lk_{K'}^-(\mathbf{v})$, are equal.

Definition 5.4 (Past Link Obstruction)**.** Let $\mathbf{w} \in K_0$. A *past link obstruction (type-∞)* in K with respect to $\mathbf{w}$ is a vertex $\mathbf{v} \in K_0$ such that $lk_{K,\mathbf{w}}^-(\mathbf{v})$ is not contractible. A *past link obstruction (type-0)* in K with respect to $\mathbf{w}$ is a vertex $\mathbf{v} \in K_0$ such that $lk_{K,\mathbf{w}}^-(\mathbf{v})$ is not connected.

Directed collapses preserve some topological properties of the space of directed paths. In particular:

Corollary 5.5. *If there are no type-∞ past link obstructions, then all spaces of directed paths from the initial point are contractible. If there are no type-0 past link obstructions, all spaces of directed paths from the initial point are connected.*

Proof. Contractibility is a direct consequence of Theorem 4.1. Likewise, connectedness follows from Theorem 4.2. □

Corollary 5.6 (Invariants of Directed Collapse)**.** *If we have a sequence of directed collapses from K to K', then there are no obstructions in K iff there are no obstructions in K'.*

Remark 5.7 (Past Link Obstructions are Inherently Local)**.** The past link of a vertex is constructed using local (rather than global) information from the cubical complex. Therefore, a past link obstruction is also a local property, which is not dependent on the global construction of the cubical complex.

Below, we provide a few motivating examples for our definition of directed collapse. In general, we want our directed collapses to preserve all spaces of directed paths between the initial vertex and any other vertex in our cubical complex. Notice, τ from Definition 5.1 is a *free face* of K. Performing a directed collapse with an arbitrary free face of a directed space K

with minimal element $\mathbf{0} \in K_0$ and maximal element $\mathbf{1} \in K_0$ can modify the individual spaces of directed paths $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{0}}(K)$ and $\overrightarrow{P}^{\mathbf{1}}_{\mathbf{v}}(K)$ for $\mathbf{v} \in K_0$.

When $\overrightarrow{P}^{\mathbf{1}}_{\mathbf{v}}(K) = \emptyset$, we call $\mathbf{v}$ a *deadlock*. When $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{0}}(K) = \emptyset$, we call $\mathbf{v}$ *unreachable*. Deadlocks and unreachable vertices are in a sense each others opposites. Notice if we take the same directed space K yet reverse the direction of all dipaths, then deadlocks become unreachable vertices and vice versa. However, as Example 5.8 and Example 5.9 illustrate, the creation of an unreachable vertex in the process of a directed collapse might result in a past link obstruction at a neighboring vertex while the creation of a deadlock does not.

Example 5.8 (3 x 3 Grid, Deadlocks & Unreachability)**.** Let K be the Euclidean cubical complex in $\mathbb{R}^2$ that is the 3×3 grid. Consider the Euclidean cubical complexes K' and K'' obtained by removing (τ, σ) with $\tau = [(1,3),(2,3)], \sigma = [(1,2),(2,3)]$ and (τ', σ') with $\tau' = [(1,0),(2,0)]$, $\sigma' = [(1,0),(2,1)]$, respectively. While K' is a directed collapse of K, K'' is not a directed collapse of K because K'' introduces a past link obstruction at $(2,1)$. So, (τ, σ) is a collapsing pair while (τ', σ') is not. Collapsing K to K' creates a deadlock at $(1,3)$ but this does not change the space of directed paths from the designated start vertex $\mathbf{0}$ to any of the vertices between $\mathbf{0}$ and the designated end vertex $(3,3)$ (see K' in Figure 5). However, collapsing K to K'' creates an unreachable vertex $(2,0)$ from the start vertex $\mathbf{0}$ (see K'' in Figure 5) which does change the space of directed paths from $\mathbf{0}$ to $(2,0)$ to be empty. Hence not all spaces of directed paths starting at $\mathbf{0}$ are preserved. This motivates our definition of directed collapse.

Our next example shows how directed collapses can be performed with collapsing pairs (τ, σ) when τ is of codimension one and greater.

Example 5.9 (3 x 3 grid, Edge & Vertex Collapses)**.** Consider again the Euclidean cubical complex K from Example 5.8. If we allow a collapsing pair (τ, σ) with τ of dimension greater than 0, we may introduce deadlocks or unreachable vertices. In particular, collapsing the free edge $\tau = [(1,3),(2,3)]$ of the top blue square $\sigma = [(1,2),(2,3)]$ in Figure 6 changes the space of directed paths $\overrightarrow{P}^{(3,3)}_{(1,3)}(K)$ from being trivial to empty in $K \backslash \{\gamma | \tau \subseteq \gamma \subseteq \sigma\}$. Yet we care about preserving the space of directed paths from our designated start vertex $\mathbf{0}$ to any of the vertices (i,j) with $0 \leq i, j \leq 3$ since we ultimately are interested in preserving the path space $\overrightarrow{P}^{(3,3)}_{\mathbf{0}}(K)$. Because of this, such

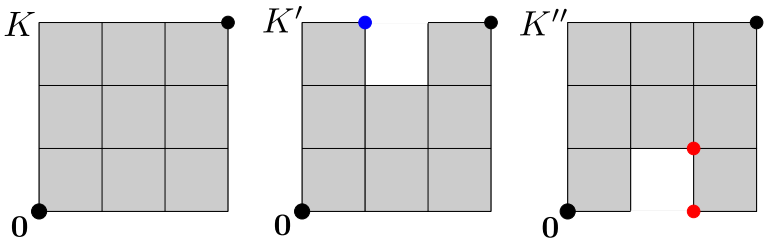
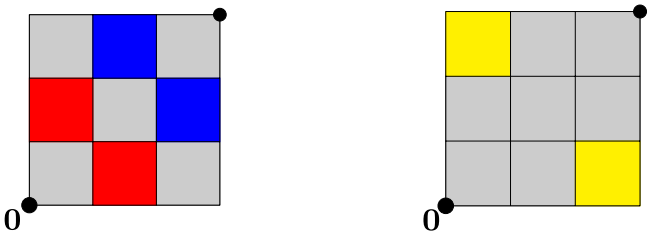


Figure 5: Illustrating Example 5.8. On the left: the cubical complex K with initial vertex $\mathbf{0}$ and final vertex $(3,3)$. In the center: The cubical complex K' which is a directed collapse of K. The deadlock in blue does not change the space of directed paths from $\mathbf{0}$ to any of the vertices between $\mathbf{0}$ and $(3,3)$. On the right: the cubical complex K'' which is not a directed collapse of K. The space of directed paths into the unreachable red vertex, $(2,0)$, becomes empty. The empty path space is reflected in the topology of the past link of the red vertex $(2,1)$ (see Example 5.9).

collapses should be allowed in our directed setting. Note that, in these cases, the past link of all vertices remains contractible. However, collapsing the free edge $\tau' = [(1,0),(2,0)]$ of the bottom red square $\sigma' = [(1,0),(2,1)]$ in Figure 6 changes the path space $\overrightarrow{P}_{\mathbf{0}}^{(2,0)}(K)$ from being trivial to empty. This change is reflected in the non-contractible past link of $(2,1)$ in $K \backslash \{\gamma | \tau' \subseteq \gamma \subseteq \sigma'\}$ that consists of the two vertices $\mathbf{j} = (1,0)$ and $\mathbf{j}' = (0,1)$ but not the edge $\mathbf{j}'' = (1,1)$ connecting them. Restricting our collapsing pairs to only include τ of dimension 0 allows for only two potential collapses, the corner vertices $(0,3)$ and $(3,0)$ into the yellow squares $[(0,2),(1,3)]$ and $[(2,0),(3,1)]$, respectively. Neither of these collapses create deadlocks or unreachable vertices and the contractibility of the past link at all vertices is preserved. Performing these corner vertex collapses exposes new free vertices that can be a part of subsequent collapses.

Lastly, we explain how the Swiss flag can be collapsed using a sequence of zero-collapses. The Swiss flag contains uncountably many paths between the initial and final vertex. After performing the sequence of zero-collapses as described in Example 5.10, the Swiss flag has only two paths up to reparametrization between the initial and final vertex. These two paths represent the two dihomotopy classes of paths that exists for the Swiss flag. Referring back to concurrent programming, we interpret the two paths as two inequivalent executions: either the first process holds a lock on the two



(a) Edge Collapse (b) Vertex Collapse

Figure 6: Illustrating Example 5.9. On the left: the collapsing of the free edge in the blue squares is an admitted directed collapse. The collapsing of the free edge in the red squares is not an admitted directed collapse. On the right: the collapsing of the free vertex in the yellow squares is an admitted directed collapse.

resources then releases them so the other process can place a lock on the resources or vice versa.

Example 5.10 (0-collapsing the Swiss Flag)**.** The Swiss flag considered as a Euclidean cubical complex in the 5×5 grid has vertices with connected past links, except at $(4, 3)$ and $(3, 4)$. The vertex $(2, 2)$ and the cube $[1, 2] \times [1, 2]$ are a 0-collapsing pair. The vertex $(3, 3)$ and the cube $[3, 4] \times [3, 4]$ are not, since that collapse would produce a disconnected past link at $(4, 4)$. A sequence of 0-collapses preserving the initial and final point will give a one-dimensional Euclidean cubical complex and one 2-cube. Specifically, we get the edges $[0, 1] \times \{0\}$, $\{1\} \times [0, 1]$, $\{1\} \times [1, 3]$, $[1, 3] \times \{1\}$, $[1, 2] \times \{3\}$, $\{3\} \times [1, 2]$, $\{2\} \times [3, 4]$, $[3, 4] \times \{2\}$, $[2, 3] \times \{4\}$, $\{4\} \times [2, 3]$, the square $[3, 4] \times [3, 4]$, and lastly the edges $\{4\} \times [4, 5]$ and $[4, 5] \times \{5\}$.

6 Discussion

Directed topological spaces have a rich underlying structure and many interesting applications. The analysis of this structure requires tools that are not fully developed, and a further investigation into these methods will lead to a better understanding of directed spaces. In particular, the development of these notions, such as directed collapse, may lead to a better understanding of equivalence of directed spaces and their spaces of directed paths.

Interestingly, when comparing directed collapse with the notion of cu-

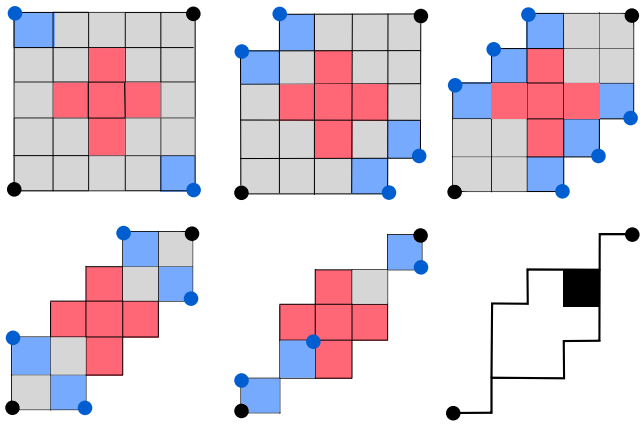


Figure 7: Zero-collapsing the Swiss Flag. A sequence of zero-collapses is presented from the top left to bottom right. At each stage, the faces and vertices shaded in blue represent the zero-collapsing pairs. The result of the sequence is shown in the bottom right which is a one-dimensional Euclidean cubical complex and one two-cube.

bical collapse in the undirected case, two main contrasts arise. First, the notion of directed collapse is stronger than that of cubical collapse; any directed collapse is a cubical collapse, but not all cubical collapses satisfy the past link requirement of directed collapse. However, directed collapse is not related to existing notions of dihomotopy equivalence which involve continuous maps between topological spaces that preserve directed paths. Hence, directed collapse contrasts cubical collapse in the undirected case since any two spaces related by cubical collapses are homotopic. This contrast suggests the need for dihomotopy equivalence with respect to an initial point.

Directed collapse may not preserve dihomotopy equivalence, so we can collapse more than, e.g., Kahl. By Theorem 5.6, if K' is a directed collapse of K with respect to $\mathbf{v}$ and K' has trivial spaces of directed paths from $\mathbf{v}$, then so does K. Similarly, if all spaces of directed paths are connected in K', then all spaces of directed paths are connected in K. Hence, our definition of directed collapsibility preserves spaces of directed paths with an initial vertex of $\mathbf{0}$. Preserving spaces of directed paths allows us to study more types of concurrent programs and preserve notions of partial executions.

We plan to pursure many future avenues of research in the directed

topological setting. First, we hope to find necessary and sufficient conditions for a pair of cubical cells (τ, σ) to be a collapsing pair. The key will be to have a better understanding of what removing a cubical cell does to the past link of a complex. Additionally, we would like to find directed conterparts to the various types of simplicial collapses. For example, is there a notion of strong directed collapse? As strong collapse also considers the link of a vertex, a consideration of how strong collapse extends to a directed setting seems natural.

Next, we would like to learn more about past link obstructions. We know that performing a directed collapse will not alter the space of directed paths of a Euclidean cubical complex; however, if we are unable to perform a directed collapse due to a past link obstruction, what happens to the space of directed paths? Theorem 4.8 is a start in understanding what happens to spaces of directed paths for 0 collapses. Another question may be, in what way are obstructions of type ∞ realized as non-contractible spaces of directed paths?

Another direction of research we hope to pursue is defining a way to compute a directed homology that is collapsing invariant. Even the two-dimensional setting (where the cubes are at most dimension two) has proved to be difficult, as adding one two-cell can have various effects, depending on the past links of the vertices involved. We would like to classify the spaces where such a dynamic programming approach would work.

Lastly, many computational questions arise on how to implement the collapse of a directed cubical complex. In [8], an example of collapsing a three-dimensional cubical complex is implemented in C++. This algorithm could be used as a model when handling the directed complex.

Many interesting theoretical and computational questions continue to emerge in the field of directed topology. We hope that our research excites others in studying cubical complexes in the directed setting.

Acknowledgments This research is a product of one of the working groups at the Women in Topology (WIT) workshop at MSRI in November 2017. This workshop was organized in partnership with MSRI and the Clay Mathematics Institute, and was partially supported by an AWM ADVANCE grant (NSF-HRD 1500481). In addition, LF and BTF further collaborated at the Hausdorff Research Institute for Mathematics during

the Special Hausdorff Program on Applied and Computational Algebraic Topology (2017).

The authors also thank the generous support of NSF. RB is partially supported by the NSF GRFP (grant no. DGE 1649608). BTF is partially supported by NSF CCF 1618605. CR is partially supported by the NSF GRFP (grant no. DGE 1842165). SE is supported by the Swiss National Science Foundation (grant no. 200021-172636)

References

[1] Jonathan Ariel Barmak and Elias Gabriel Minian. Strong homotopy types, nerves and collapses. *Discrete Comput. Geom.*, 47(2):301–328, 2012.

[2] Edsger W. Dijkstra. Two starvation-free solutions of a general exclusion problem. 1977. Manuscript EWD625, from the archives of UT Austin, https://www.cs.utexas.edu/users/EWD/.

[3] Lisbeth Fajstrup. Dipaths and dihomotopies in a cubical complex. *Advances in Applied Mathematics*, 35(2):188 – 206, 2005.

[4] Lisbeth Fajstrup, Eric Goubault, Emmanuel Haucourt, Samuel Mimram, and Martin Raussen. *Directed Algebraic Topology and Concurrency.* Springer Publishing Company, Inc., 1st edition, 2016.

[5] Lisbeth Fajstrup, Eric Goubault, and Martin Raussen. Algebraic topology and concurrency. *Theoretical Comuter Science*, pages 241–271, 2006.

[6] Robin Forman. A user's guide to discrete Morse theory. *Sém. Lothar. Combin.*, 48, 12 2001.

[7] Charles A.R. Hoare. Communicating sequential processes. *Commun. ACM*, 21(8):666–677, August 1978.

[8] Jacques-Olivier Lachaud. Cubical complex collapse, 2017.

[9] Martin Raussen. On the classification of dipaths in geometric models for concurrency. *Mathematical Structures in Computer Science*, 10(4):427–457, 2000.

[10] Martin Raussen and Krzysztof Ziemiański. Homology of spaces of directed paths on euclidean cubical complexes. *Journal of Homotopy and Related Structures*, 9(1):67–84, Apr 2014.

[11] John H.C. Whitehead. Simplicial spaces, nuclei and m-groups. *Proceedings of the London Mathematical Society*, s2-45(1):243–327.

[12] Krzysztof Ziemiański. On execution spaces of PV-programs. *Theoretical Computer Science*, 619:87–98, 2016.

3.2 Combinatorial Conditions for Directed Collapsing

(joint work with Robin Belton, Robyn Brooks, Lisbeth Fajstrup, Brittany Terese Fasy, Nicole Sanderson, and Elizabeth Vidaurre)

Accepted in Advances in Mathematical Sciences, AWM Series, Springer, 2021.

Abstract

While collapsibility of CW complexes dates back to the 1930s, collapsibility of *directed* Euclidean cubical complexes has not been well studied to date. The classical definition of collapsibility involves certain conditions on pairs of cells of the complex. The direction of the space can be taken into account by requiring that the past links of vertices remain homotopy equivalent after collapsing. We call this type of collapse a *link-preserving directed collapse*. In the undirected setting, pairs of cells are removed that create a deformation retract. In the directed setting, topological properties—in particular, properties of spaces of directed paths—are not always preserved. In this paper, we give computationally simple conditions for preserving the topology of past links. Furthermore, we give conditions for when link-preserving directed collapses preserve the contractability and connectedness of spaces of directed paths. Throughout, we provide illustrative examples.

1 Introduction

A directed Euclidean cubical complex is a subset of $\mathbb{R}^n$ comprising a finite union of directed unit cubes. Directed paths (i.e., paths that are nondecreasing in all coordinates) and spaces of directed paths are the objects of study in this paper. In particular, we address the question of how to simplify directed Euclidean complexes without significantly changing the spaces of directed paths.

This model is motivated by several applications, where each axis of the model corresponds to a parameter of the application (e.g., time). In particular, Euclidean cubical complexes are used to model concurreny in computer programming [4–6,21], hybrid dynamical systems [20], and motion planning [7]. Consider the application to concurrency. In this example, each axis represents a sequence of actions a process completes in the program execution. The complex itself corresponds to "compatible" parameters (i.e., when the processes can execute simultaneously). Cubes missing from the complex correspond to parameters for which the processes cannot execute simultaneously for some reason, such as when they require the same resources with limited capacity; see Fig. 1. A directed path (dipath) in the complex represents a, possibly partial, program execution. Such executions are equivalent if the corresponding dipaths are *directed homotopic*. Simplifying the complexes allows for a more compact representation of the execution space, which, in turn, reduces the complexity of validating correctness of concurrent programs.

A non-trivial Euclidean cubical complex contains uncountably many dipaths and more information than we need for understanding the topology of the spaces of dipaths. The main question we ask is, *How can we simplify a directed Euclidean cubical complex while still preserving spaces of dipaths*?

Past links are local representations of a Euclidean cubical complex at vertices. They were introduced in [21] as a means to show that any finite homotopy type can be realized as a connected component of the space of execution paths for some PV-model. In [1], we found conditions for when the local information of past links preserve the global information on the homotopy type of spaces of dipaths. Because of these relationships between past links and dipath spaces, we define collapsing in terms of past links. We call this type of collapsing *link-preserving directed collapse* (LPDC). We aim to compress a Euclidean cubical complex by LPDCs before attempting to answer questions about dipath spaces.

The main result of this paper is Theorem 3.9, which provides a simple criterion for such a collapsing to be allowed: *A pair of cubes (τ, σ) is an LPDC pair if and only if it is a collapsing pair in the non-directed sense and τ does not contain the minimum vertex of σ.* This condition greatly simplifies the defini-

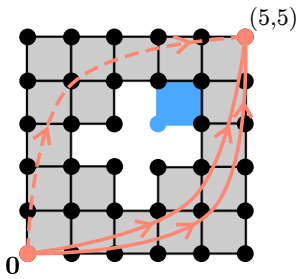


Figure 1: The Swiss Flag and Three Directed Paths. The gray and blue squares are the two-cubes of a Euclidean cubical complex. The bi-monotone increasing paths are directed paths starting at $(0, 0)$ and ending at $(5, 5)$. This complex has a cross-shaped hole in the middle. As a consequence, the solid directed paths are directed homotopic while the dashed directed path is not directed homotopic to either of the other directed paths. Each point highlighted in blue is *unreachable*, meaning that we cannot reach any point highlighted in blue without breaking bi-monotonicity in a path starting at $(0, 0)$. This complex models the dining philsophers problem, a well-known example in concurrency, where two processes require two shared resources with limited capacity [4,12]. The two distinct paths (solid and dashed) represent which process uses both shared resources first.

tion of LPDC and is easy to add to a collapsing algorithm for Euclidean cubical complexes in the undirected setting. Algorithms and implementations in this setting already exist such as in [15]. Furthermore, we provide conditions for when LPDCs preserve the contractability and connectedness of dipath spaces (Section 4) along with a discussion of some of the limitations (Section 5). This work provides a start at the mathematical foundations for developing polynomial time algorithms that collapse Euclidean cubical complexes and preserve dipath spaces.

2 Background

This paper builds on our prior work [1], as well as work by others [6, 9, 10, 16, 21]. In this section, we recall the definitions of directed Euclidean cubical complexes, which are the objects that we study in this paper. Then, we discuss the relationship between spaces of directed paths and past links in directed Euclidean cubical complexes. For additional background on directed topology (including generalizations of the definitions below), we refer the reader to [5]. We also assume the reader is familiar with the notion of homotopy equivalence of topological spaces (denoted using $\simeq$ in this paper) and homotopy between

paths as presented in [11].

2.1 Directed Spaces and Euclidean Cubical Complexes

Let n be a positive integer. A *(closed) elementary cube* in $\mathbb{R}^n$ is a product of closed intervals of the following form:

$$[v_1 - j_1, v_1] \times [v_2 - j_2, v_2] \times \ldots \times [v_n - j_n, v_n] \subsetneq \mathbb{R}^n, \tag{1}$$

where $\mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{Z}^n$ and $\mathbf{j} = (j_1, j_2, \ldots, j_n) \in \{0, 1\}^n$. We often refer to elementary cubes simply as *cubes*. The dimension of the cube is the number of unit entries in the vector $\mathbf{j}$; specifically, the dimension of the cube in Eq. (1) is the sum: $\sum_{i=1}^{n} j_i$. In particular, when $\mathbf{j} = \mathbf{0} := (0, 0, \ldots, 0)$, the elementary cube is a single point and often denoted using just $\mathbf{v}$. If τ and σ are elementary cubes such that $\tau \subseteq \sigma$, we say that τ is a face of σ and that σ is a coface of τ. Cubical sets were first introduced in the 1950s by Serre [17] in a more general setting; see also [2, 8, 13].

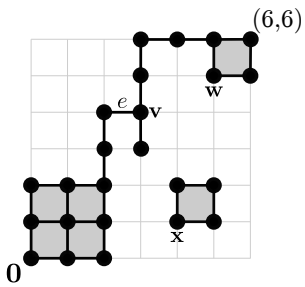


Figure 2: Euclidean cubical complex in $\mathbb{R}^2$ with 24 zero-cubes (vertices), 28 one-cubes (edges), and six two-cubes (squares). By construction, all elementary cubes in a directed Euclidean cubical complex are axis aligned. Consider the vertex $\mathbf{v} = (3, 4)$. The edge $e = [(2, 4), (3, 4)]$ (written $e = [2, 3] \times [4, 4]$ in the notation of Eq. (1)) is one of the two lower cofaces of $\mathbf{v}$. Since e is not a face of any two-cube, e is a maximal cube (since it is not a face of a higher-dimensional cube).

Elementary cubes stratify $\mathbb{R}^n$, where two points $x, y \in \mathbb{R}^n$ are in the same stratum if and only if they are members of the same set of elementary cubes; we call this the *cubical stratification of* $\mathbb{R}^n$. Each stratum in the stratification is either an open cube or a single point. A *Euclidean cubical complex* $(K, \mathcal{K})$ is a subspace $K \subsetneq \mathbb{R}^n$ that is equal to the union of a finite set of elementary cubes, together with the stratification $\mathcal{K}$ induced by the cubical stratification of $\mathbb{R}^n$; see Fig. 2. We topologize K using the subspace topology with the standard topology

on $\mathbb{R}^n$. By construction, if $\sigma \in \mathcal{K}$, then all of its faces are necessarily in $\mathcal{K}$ as well. If $\sigma \in \mathcal{K}$ with no proper cofaces, then we say that σ is a *maximal cube* in K. We denote the set of closed cubes in $(K, \mathcal{K})$ by $\overline{\mathcal{K}}$; the set of closed cubes in $\overline{\mathcal{K}}$ is in one-to-one correspondence with the open cubes in $\mathcal{K}$. Specifically, vertices in $\overline{\mathcal{K}}$ correspond to vertices in $\mathcal{K}$ and all other elementary cubes in $\overline{\mathcal{K}}$ correspond to their interiors in $\mathcal{K}$. Throughout this paper, we denote the set of zero-cubes in $\mathcal{K}$ by $\mathrm{verts}(K)$ and note that $\mathrm{verts}(K) \subsetneq \mathbb{Z}^n$, since all cubes in $(K, \mathcal{K})$ are elementary cubes.

The *product order on* $\mathbb{R}^n$, denoted $\preceq$, is the partial order such that for two points $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ and $\mathbf{q} = (q_1, q_2, \ldots, q_n)$ in $\mathbb{R}^n$, we have $\mathbf{p} \preceq \mathbf{q}$ if and only if $p_i \leq q_i$ for each coordinate i. Using this partial order, we define the interval of points in $\mathbb{R}^n$ between $\mathbf{p}$ and $\mathbf{q}$ as

$$[\mathbf{p}, \mathbf{q}] := \{\mathbf{x} \mid \mathbf{p} \preceq \mathbf{x} \preceq \mathbf{q}\}.$$

The point $\mathbf{p}$ is the minimum vertex of the interval and $\mathbf{q}$ is the maximum vertex of the interval, with respect to $\preceq$. Notationally, we write this as $\min([\mathbf{p}, \mathbf{q}]) := \mathbf{p}$ and $\max([\mathbf{p}, \mathbf{q}]) := \mathbf{q}$. When $\mathbf{q} \in \mathbb{Z}^n$ and $\mathbf{p} = \mathbf{q} + \mathbf{j}$, for some $\mathbf{j} \in \{0, 1\}^n$, the interval $[\mathbf{p}, \mathbf{q}]$ is an elementary cube as defined in Eq. (1). If, in addition, $\mathbf{j}$ is not the zero vector, then we say that $[\mathbf{v} - \mathbf{j}, \mathbf{v}]$ is a *lower coface* of $\mathbf{v}$.

Using the fact that the partial order $(\mathbb{R}^n, \preceq)$ induces a partial order on the points in K, we define directed paths in K as the set of nondecreasing paths in K: A *path* in K is a continuous map from the unit interval $I = [0, 1]$ to K. We say that a path $\gamma \colon I \to K$ goes from $\gamma(0)$ to $\gamma(1)$. Letting K^I denote the set of all paths in K, the set of *directed paths* (or *dipaths* for short) is

$$\overrightarrow{P}(K) := \{\gamma \in K^I \mid \forall i, j \text{ s.t. } 0 \leq i \leq j \leq 1, \gamma(i) \preceq \gamma(j)\}.$$

We topologize $\overrightarrow{P}(K)$ using the compact-open topology. For $\mathbf{p}, \mathbf{q} \in K$, we denote the subspace of dipaths from $\mathbf{p}$ to $\mathbf{q}$ by $\overrightarrow{P}^{\mathbf{q}}_{\mathbf{p}}(K)$. We refer to $(K, \overrightarrow{P}(K))$ as a *directed Euclidean cubical complex*.[1] The connected components of $\overrightarrow{P}^{\mathbf{q}}_{\mathbf{p}}(K)$ are exactly the equivalence classes of dipaths, up to dihomotopy. If two dipaths, f and g are homotopic through a continuous family of dipaths, then f and g are called *dihomotopic*.

Given a directed complex, certain subcomplexes are of interest:

[1] Directed Euclidean cubical complexes are an example of a more general concept known as *directed space* (d-spaces). To define a d-space, we have a topological space X and we define a set of dipaths $P'(X) \subseteq X^I$ that contains all constant paths, and is closed under taking nondecreasing reparameterizations, concatenations, and subpaths. Indeed, $\overrightarrow{P}(K)$ satisfies these properties.

Definition 2.1 (Special Complexes)**.** Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\mathbf{p} \in \text{verts}(K)$ and let σ be an elementary cube (that need not be in $\mathcal{K}$).

1. The complex above $\mathbf{p}$ is $K_{\mathbf{p}\preceq} := \{\mathbf{q} \in K \mid \mathbf{p} \preceq \mathbf{q}\}$.
2. The complex below $\mathbf{p}$ is $K_{\preceq\mathbf{p}} := \{\mathbf{q} \in K \mid \mathbf{q} \preceq \mathbf{p}\}$.
3. The reachable complex from $\mathbf{p}$ is $\text{reach}(K, \mathbf{p}) := \{\mathbf{q} \in K \mid \overrightarrow{P}^{\mathbf{q}}_{\mathbf{p}}(K) \neq \emptyset\}$.
4. The complex restricted to σ is
$$K|_\sigma := \bigcup\{\tau \in \mathcal{K} \mid \min\sigma \preceq \min\tau \preceq \max\tau \preceq \max\sigma\}.$$
5. If $K = I^n$, then we call $(K, \mathcal{K})$ the *standard unit cubical complex* and often denote it by $(I^n, \mathcal{I})$. If $K = I^n + \mathbf{x}$ for some $\mathbf{x} \in \mathbb{Z}^n$, then K is a full-dimensional unit cubical complex.

2.2 Past Links of Directed Cubical Compelxes

An *abstract simplicial complex* is a finite collection $\mathcal{S}$ of sets that is closed under the subset relation, i.e., if $A \in \mathcal{S}$ and B is a set such that $\emptyset \neq B \subseteq A$, then $B \in \mathcal{S}$. The sets in $\mathcal{S}$ are called *simplices*. If the simplex A has $k+1$ elements, then we say that the dimension of A is $\dim(A) := k$, and we say A is a k-simplex. For example, the zero-simplices are the singleton sets and are often referred to as vertices. Since every element of a set $A \in \mathcal{S}$ gives rise to a singleton set in the finite set $\mathcal{S}$, A must be finite.

In a topological space embedded in $\mathbb{R}^n$, the link of a point $\mathbf{v}$ is constructed by intersecting an arbitrarily small $(n-1)$-sphere around $\mathbf{v}$ with the space itself. In $\mathbb{R}^n$, the link of a point is an $(n-1)$-sphere. Moreover, if $\mathbf{v} \in \mathbb{Z}^n$, the link inherits the stratification as a subcomplex of $\mathbb{R}^n$, and can be represented as a simplicial complex whose i-simplices are in one-to-one correspondence with the $(i+1)$-dimensional cofaces of $\mathbf{v}$. The past link of $\mathbf{v}$ is the restriction of the link using the set of lower cofaces of $\mathbf{v}$ instead of all cofaces. Thus, we can represent each simplex in the past link as a vector in $\{0,1\}^n \setminus \{\mathbf{0}\}$, where the vector $\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\}$ represents the cube $[\mathbf{v}-\mathbf{j}, \mathbf{v}]$ in the simplex-cube correspondence. As a simplicial complex, the past link of $\mathbf{v}$ in $\mathbb{R}^n$ has n vertices $\{x_i\}_{1\leq i\leq n}$, and $\mathbf{j}$ represents the simplex $\{x_i | 1 \leq i \leq n, j_i = 1\}$ of dimension $||\mathbf{j}||_1 - 1$; for example, $(1,0,0)$ represents a vertex and $(1,0,1)$ represents an edge. We are now ready to define the past link of a vertex in a Euclidean cubical complex:

(a) Past Link of $\mathbf{v}$ (b) Past Link of $\mathbf{x}$

Figure 3: Past link in the Open Top Box. (a) The maximum vertex of this complex is $\mathbf{v} = (v_1, v_2, v_3)$. The past link $\mathrm{lk}_K^-(\mathbf{v})$ is the simplicial complex comprising three vertices and two edges (shown in blue/cyan). These simplices are in one-to-one correspondence with the set of lower cofaces of $\mathbf{v}$ (highlighted in pink). For example, the edges of $\mathrm{lk}_K^-(\mathbf{v})$, which are labeled e_1 and e_2, are in one-to-one correspondence with the elementary two-cubes that are lower cofaces of $\mathbf{v}$ ($\sigma_1 = [(v_1-1, v_2, v_3-1), \mathbf{v}]$ and $\sigma_2 = [(v_1, v_2-1, v_3-1), \mathbf{v}]$, respectively). In the vector notation for simplices of $\mathrm{lk}_K^-(\mathbf{v})$, we write $e_1 = (1,0,1)$ and $e_2 = (0,1,1)$. (b) The past link of a vertex $\mathbf{x}$ that is neither the minimum nor the maximum vertex in the complex.

Definition 2.2 (Past Link)**.** Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\mathbf{v} \in \mathbb{Z}^n$. The *past link* of $\mathbf{v}$ is the following simplicial complex:

$$\mathrm{lk}_K^-(\mathbf{v}) := \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid [\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K\}.$$

As a set, the past link represents all elementary cubes in K for which $\mathbf{v}$ is the maximum vertex. As a simplicial complex, it describes (locally) the different types of dipaths to or through $\mathbf{v}$ in K; see Fig. 3.

We conclude this section with a lemma summarizing properties of the past link, most of which follow directly from definitions:

Lemma 2.3 (Properties of Past Links)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Then, the following statements hold for all $\mathbf{v} \in \mathbb{Z}^n$:*

1. $\mathrm{lk}_K^-(\mathbf{v}) = \bigcup_{\mathbf{p} \in \mathbb{R}^n} \mathrm{lk}_{K_{\mathbf{p} \preceq}}^-(\mathbf{v})$.
2. *If $(K', \mathcal{K}')$ is a subcomplex of $(K, \mathcal{K})$, then $\mathrm{lk}_{K'}^-(\mathbf{v}) \subseteq \mathrm{lk}_K^-(\mathbf{v})$.*
3. $\mathrm{lk}_K^-(\mathbf{v}) = \mathrm{lk}_{K_{\preceq \mathbf{v}}}^-(\mathbf{v})$.
4. *If there exists $\mathbf{w} \in \mathbb{Z}^n$ such that $K = [\mathbf{w} - \mathbf{1}, \mathbf{w}]$, then $\mathrm{lk}_K^-(\mathbf{w})$ is the complete simplicial complex on n vertices.*
5. $\mathrm{lk}_K^-(\mathbf{v})$ *is a subcomplex of the complete simplicial complex on n vertices.*

Proof. Statement *1*: If $K = \emptyset$, then all past links are empty and the equality trivially holds. If $K \neq \emptyset$, then $\text{verts}(K)$ is a finite non empty set. Thus, there exists $\mathbf{q} \in \mathbb{R}^n$ such that for all $\mathbf{w} \in \text{verts}(K)$, $\mathbf{q} \preceq \mathbf{w}$. Let $\mathbf{j} \in \text{lk}^-_K(\mathbf{v})$. Then, $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K$ and so $\mathbf{v} - \mathbf{j} \in \text{verts}(K)$. Hence, $\mathbf{q} \preceq \mathbf{v} - \mathbf{j}$, which means that $\mathbf{j} \in \text{lk}^-_{K_{\mathbf{q}\preceq}}(\mathbf{v}) \subseteq \bigcup_{\mathbf{p}\in\mathbb{R}^n} \text{lk}^-_{K_{\mathbf{p}\preceq}}(\mathbf{v})$. The reverse inclusion follows from the fact that each of these statements holds if and only if.

Statement *2*: Observe that if $\mathbf{j} \in \text{lk}^-_{K'}(\mathbf{v})$, then, by definition of the past link, $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K'$. Since $K' \subseteq K$, we have $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K' \subseteq K$. Therefore, we can conclude that $\mathbf{j} \in \text{lk}^-_K(\mathbf{v})$.

Statement *3*: By Statement *2* (which we just proved), we have the following inclusion $\text{lk}^-_{K_{\preceq\mathbf{v}}}(\mathbf{v}) \subseteq \text{lk}^-_K(\mathbf{v})$. To prove the inclusion in the other direction, let $\mathbf{j} \in \text{lk}^-_K(\mathbf{v})$. Since $\mathbf{v} - \mathbf{j} \preceq \mathbf{v}$, then $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K_{\preceq\mathbf{v}}$. Therefore, we conclude that $\text{lk}^-_K(\mathbf{v}) \subseteq \text{lk}^-_{K_{\preceq\mathbf{v}}}(\mathbf{v})$.

Statement *4*: Since $K = [\mathbf{w} - \mathbf{1}, \mathbf{w}]$, we know that K is full-dimensional, and so for all $\mathbf{j} \in \{0,1\}^n$, $[\mathbf{w} - \mathbf{j}, \mathbf{w}] \subseteq K$. Thus, by definition of past link, we have that the past link of $\mathbf{w}$ is: $\text{lk}^-_K(\mathbf{w}) := \{0,1\}^n \setminus \{\mathbf{0}\}$, which is the complete simplicial complex on n vertices.

Statement *5*: Let $L = K \cap [\mathbf{v} - \mathbf{1}, \mathbf{v}]$. By definition of past link, we know $\text{lk}^-_L(\mathbf{v}) = \text{lk}^-_K(\mathbf{v})$. By Statement *2*, since L is a subcomplex of $[\mathbf{v} - \mathbf{1}, \mathbf{v}]$, we know $\text{lk}^-_L(\mathbf{v}) \subseteq \text{lk}^-_{[\mathbf{v}-\mathbf{1},\mathbf{v}]}(\mathbf{v})$. By Statement *4*, $\text{lk}^-_{[\mathbf{v}-\mathbf{1},\mathbf{v}]}(\mathbf{v})$ is the complete simplicial complex on n vertices. Therefore, $\text{lk}^-_K(\mathbf{w})$ is the complete simplicial complex on n vertices. □

2.3 Relationship Between Past Links and Path Spaces

The topology of the past links is intrinsically related to spaces of dipaths. Specifically, in [1] we prove that the contractability and/or connectedness of past links of vertices in directed Euclidean cubical complexes with a minimum vertex[2] implies that all spaces of dipaths with $\mathbf{w}$ as initial point are also contractible and/or connected.

Theorem 2.4 (Contractability [1, Theorem 1])**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$ that has a minimum vertex $\mathbf{w}$. If, for all vertices $\mathbf{v} \in \text{verts}(K)$, the past link $\text{lk}^-_K(\mathbf{v})$ is contractible, then the space $\overrightarrow{P}^{\mathbf{k}}_{\mathbf{w}}(K)$ is contractible for all $\mathbf{k} \in \text{verts}(K)$.*

An analogous theorem for connectedness also holds.

[2] In [1], the minimum (initial) vertex was often assumed to be $\mathbf{0}$ for ease of exposition. We restate the lemmas and theorems here using more general notation, where K has a minimum vertex $\mathbf{w}$.

Theorem 2.5 (Connectedness [1, Theorem 2])**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$ that has a minimum vertex $\mathbf{w}$. Suppose that, for all $\mathbf{v} \in \mathrm{verts}(K)$, the past link $\mathrm{lk}_K^-(\mathbf{v})$ is connected. Then, for all $\mathbf{k} \in \mathrm{verts}(K)$, the space $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{k}}(K)$ is connected.*

Furthermore, we proved a partial converse to Theorem 2.5. Specifically, the converse holds only if K is a reachable directed Euclidean cubical complex as defined in Statement *3* of Definition 2.1. This is expected: Properties of parts of the directed Euclidean complex which are not reachable from $\mathbf{w}$, do not influence the dipath spaces from $\mathbf{w}$.

Theorem 2.6 (Realizing Obstructions [1, Theorem 3])**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\mathbf{w} \in \mathrm{verts}(K)$, and let $L = \mathrm{reach}(K, \mathbf{w})$. Let $\mathbf{v} \in \mathrm{verts}(L)$. If the past link $\mathrm{lk}_L^-(\mathbf{v})$ is disconnected, then the space $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K)$ is disconnected.*

3 Directed Collapsing Pairs

Although simplicial collapses preserve the homotopy type of the underlying space [14, Proposition 6.14] and hence of all path spaces, this type of collapsing in directed Euclidean cubical complexes may not preserve topological properties of spaces of dipaths. In this section, we study a specific type of collapsing called a link-preserving directed collapse. We define link-preserving directed collapses in Section 3.1 and give properties of link-preserving directed collapses in Section 3.2.

3.1 Link-Preserving Directed Collapses

Since we are interested in preserving the dipath spaces through collapses, the results from Section 2.3 motivate us to study a type of directed collapse (DC) via past links, introduced in [1]. However, we call it a *link-preserving directed collapse* (LPDC) (as opposed to a *directed collapse*) since we show in the last sections of this paper that when the spaces of dipaths starting from the minimum vertex are not connected, the following definition of collapse does not preserve the number of components.

Definition 3.1 (Link Preserving Directed Collapse)**.** Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\sigma \in \mathcal{K}$ be a maximal cube, and let τ be a proper face of σ such that no other maximal cube contains τ (in this case, we say that τ is a *free face* of σ). Then, we define the (τ, σ)-collapse of K as the subcomplex obtained by removing everything in between τ and σ:

$$K' = K \setminus \{\gamma \in \mathcal{K} \mid \overline{\tau} \subseteq \overline{\gamma} \subseteq \overline{\sigma}\}, \tag{2}$$

and let $\mathcal{K}'$ denote the stratification of the set K' induced by the cubical stratification of $\mathbb{R}^n$ (thus, $\mathcal{K}' \subsetneq \mathcal{K}$).

We call the directed Euclidean cubical complex $(K', \mathcal{K}')$ a *link-preserving directed collapse (LPDC)* of $(K, \mathcal{K})$ if, for all $\mathbf{v} \in \text{verts}(K')$, the past link $\text{lk}^-_K(\mathbf{v})$ is homotopy equivalent to $\text{lk}^-_{K'}(\mathbf{v})$ (denoted $\text{lk}^-_K(\mathbf{v}) \simeq \text{lk}^-_{K'}(\mathbf{v})$). The pair (τ, σ) is then called an *LPDC pair*.

Remark 3.2 (Simplicial Collapses)**.** The study of simplicial collapses is known as *simple homotopy theory* [3, 19], and traces back to the work of Whitehead in the 1930s [18]. The idea is very similar: If C is an abstract simplicial complex and $\alpha \in C$ such that α is a proper face of exactly one maximal simplex β, then the following complex is the *α-collapse of C in the simplicial setting*:

$$C' = C \setminus \{\gamma \in C \mid \alpha \subseteq \gamma \subseteq \beta\}.$$

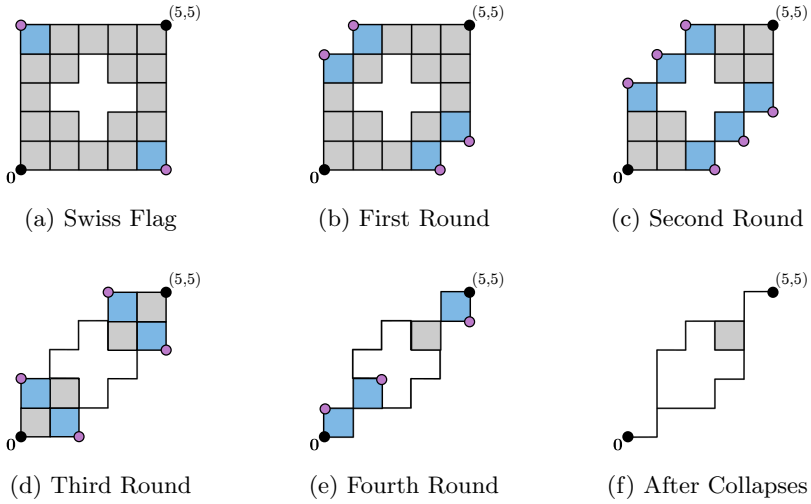
Note that we use only the free face (α) when defining a simplicial collapse, as doing so helps to distinguish between discussing a simplicial collapse and a directed Euclidean cubical collapse. In addition, we always explicitly state "in the simplicial setting" when talking about a simplicial collapse.

Applying a sequence of LPDCs to a directed Euclidean cubical complex can reduce the number of cubes, and hence can more clearly illustrate the number of dihomotopy classes of dipaths within the directed Euclidean cubical complex. For an example, see Fig. 4. However, it is not necessarily true that LPDCs preserve dipath spaces. We discuss the relationship between dipath spaces and LPDCs in Section 4.

3.2 Properties of LPDCs

We give a combinatorial condition for a collapsing pair (τ, σ) to be an LPDC pair; namely, the condition is that τ does not contain the vertex $\min(\sigma)$. From the definition of an LPDC, we see that finding an LPDC pair requires computing the past link of *all* vertices in $\text{verts}(K')$. In [1], we discussed how we can reduce the check down to only the vertices in σ since no other vertices have their past links affected. In this paper, we prove we need to only check *one* condition to determine if we have an LPDC pair. The one simple condition dramatically reduces the number of computations we need to perform in order to verify we have an LPDC. This result given in Theorem 3.9 depends on the following lemmas about the properties of past links on vertices.

Lemma 3.3 (Properties of Past Links in a Vertex Collapse)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\sigma \in \mathcal{K}$ and $\tau, \mathbf{v} \in \text{verts}(\sigma)$ such*



(a) Swiss Flag (b) First Round (c) Second Round

(d) Third Round (e) Fourth Round (f) After Collapses

Figure 4: Collapsing the Swiss Flag. A sequence of vertex collapses is presented from the top left to bottom right. At each stage, the faces and vertices shaded in blue and purple represent the vertex collapsing pairs with the blue Euclidean cube being σ and the purple vertex being τ. The result of the sequence of LPDCs is shown in (f) is a one-dimensional directed Euclidean cubical complex and one two-cube. Observe that this directed Euclidean cubical complex clearly illustrates the two dihomotopy classes of $\overrightarrow{P}_{\mathbf{0}}^{(5,5)}(K)$.

that $\tau \preceq \mathbf{v}$. If τ is a free face of σ and K' is the (τ, σ)-collapse, then the following two statements hold:

1. $\mathrm{lk}^-_{K|_\sigma}(\mathbf{v}) = \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \min(\sigma) \preceq \mathbf{v} - \mathbf{j}\}$.
2. $\mathrm{lk}^-_{K'|_\sigma}(\mathbf{v}) = \mathrm{lk}^-_{K|_\sigma}(\mathbf{v}) \setminus \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \mathbf{v} - \mathbf{j} \preceq \tau\}$.

Proof. To ease notation, we define the following two sets:

$$J := \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \min(\sigma) \preceq \mathbf{v} - \mathbf{j}\}$$
$$I := \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \mathbf{v} - \mathbf{j} \preceq \tau\}.$$

First, we prove Statement *1* (that $\mathrm{lk}^-_{K|_\sigma}(\mathbf{v}) = J$). We start with the forward inclusion. Let $\mathbf{j} \in \mathrm{lk}^-_{K|_\sigma}(\mathbf{v})$. By the definition of past links (see Definition 2.2), we know that $[\mathbf{v}-\mathbf{j}, \mathbf{v}] \subseteq K|_\sigma$. By the definition of $K|_\sigma$ (see Definition 2.1), we know that $\min(\sigma) \preceq \min([\mathbf{v}-\mathbf{j}, \mathbf{v}]) = \mathbf{v}-\mathbf{j}$. This implies $\mathbf{j} \in J$. Therefore, $\mathrm{lk}^-_{K|_\sigma}(\mathbf{v}) \subseteq J$. For the backward inclusion, let $\mathbf{j} \in J$. Then, since $\mathbf{v} \in \mathrm{verts}(\sigma)$ and σ is an elementary cube by assumption, and $\min(\sigma) \preceq \mathbf{v} - \mathbf{j}$ by definition of J, we have $\mathbf{v} - \mathbf{j} \in \mathrm{verts}(\sigma)$. Since $\sigma \in \mathcal{K}$, all faces must be in $\mathcal{K}$; hence, $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K|_\sigma$.

Therefore, $\mathbf{j} \in \mathrm{lk}^-_{K|_\sigma}(\mathbf{v})$, and so $\mathrm{lk}^-_{K|_\sigma}(\mathbf{v}) \supseteq J$. Since we have both inclusions, then Statement *1* holds.

Now, we prove Statement *2* (that $\mathrm{lk}^-_{K'|_\sigma}(\mathbf{v}) = J \setminus I$). Again, we prove the inclusions in both directions. For the forward inclusion, let $\mathbf{j} \in \mathrm{lk}^-_{K'|_\sigma}(\mathbf{v})$. By Statement *2* of Lemma 2.3, we have $\mathrm{lk}^-_{K'|_\sigma}(\mathbf{v}) \subseteq \mathrm{lk}^-_{K|_\sigma}(\mathbf{v})$, and so, we obtain $\mathbf{j} \in \mathrm{lk}^-_{K|_\sigma}(\mathbf{v}) = J$. Next, we must show that $\mathbf{j} \notin I$. Assume, for a contradiction, that $\mathbf{j} \in I$. Then, by definition of I, $\mathbf{v} - \mathbf{j} \preceq \tau$. Since $\tau \preceq \mathbf{v}$, we obtain the partial order $\mathbf{v} - \mathbf{j} \preceq \tau \preceq \mathbf{v}$. This implies that $[\tau, \mathbf{v}] \subseteq [\mathbf{v} - \mathbf{j}, \mathbf{v}]$. Since $[\mathbf{v} - \mathbf{j}, \mathbf{v}]$ is an elementary cube in $K'|_\sigma$, then its face $[\tau, \mathbf{v}]$ must also be an elementary cube in $K'|_\sigma$. Setting $\overline{\gamma} = [\tau, \mathbf{v}]$ and observing $\tau = \overline{\tau} \subseteq \overline{\gamma} \subseteq \overline{\sigma}$, we observe that γ is not an elementary cube in K' by Eq. (2). This gives us a contradiction and so $\mathbf{j} \notin I$. Therefore, $\mathrm{lk}^-_{K'|_\sigma}(\mathbf{v}) \subseteq J \setminus I$.

Finally, we prove the backward inclusion of Statement *2*. Let $\mathbf{j} \in J \setminus I$. Then, by Statement *1*, $\mathbf{j} \in \mathrm{lk}^-_{K|_\sigma}(\mathbf{v})$ and either $\tau \prec \mathbf{v} - \mathbf{j}$ or τ is not comparable to $\mathbf{v} - \mathbf{j}$ under $\preceq$. Thus, by Eq. (2), $[\mathbf{v} - \mathbf{j}, \mathbf{v}]$ is an elementary cube of $K'|_\sigma$. Thus, by Definition 2.2, we have that $\mathbf{j} \in \mathrm{lk}^-_{K'|_\sigma}(\mathbf{v})$. Hence, $J \setminus I \subseteq \mathrm{lk}^-_{K'|_\sigma}(\mathbf{v})$, and so Statement *2* holds. □

Using Lemma 3.3, we see why τ cannot be the vertex $\min(\sigma)$ when performing an LPDC. If $\tau = \min(\sigma)$, then

$$\begin{aligned}
\mathrm{lk}^-_{K'|_\sigma}(\mathbf{v}) &= \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \min(\sigma) \preceq \mathbf{v} - \mathbf{j}\} \\
&\quad \setminus \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \mathbf{v} - \mathbf{j} \preceq \min(\sigma)\} \\
&= \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \min(\sigma) \preceq \mathbf{v} - \mathbf{j} \text{ and } \mathbf{v} - \mathbf{j} \succ \min(\sigma)\} \\
&= \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \min(\sigma) \prec \mathbf{v} - \mathbf{j}\} \\
&= \{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \mathbf{j} \prec \mathbf{v} - \min(\sigma)\}.
\end{aligned}$$

If $\mathbf{v}$ is the maximum vertex of σ, then we obtain $\mathrm{lk}^-_{K'|_\sigma}(\mathbf{v}) = \{0,1\}^n \setminus \{\mathbf{0}, \mathbf{v} - \min(\sigma)\}$. This computation gives us the following corollary, which we illustrate in Fig. 5 when K is a single closed three-cube.

Corollary 3.4 (Caution for a $(\min(\sigma), \sigma)$-Collapse)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\sigma \in \mathcal{K}$, $\tau = \min(\sigma)$, and $\mathbf{v} \in \mathrm{verts}(\sigma)$. If τ is a free face and K' is the (τ, σ)-collapse, then the past link of $\mathbf{v}$ in $K'|_\sigma$ is:*

$$\{\mathbf{j} \in \{0,1\}^n \setminus \{\mathbf{0}\} \mid \mathbf{j} \prec \mathbf{v} - \min(\sigma)\}$$

In particular, if $\mathbf{v} = \max(\sigma)$ and $k = \dim(\sigma)$, then the past link is the complete complex on k elements before the collapse, and, after the collapse, it is homeomorphic to $\mathbb{S}^{k-2}$. Thus, (τ, σ) is not an LPDC pair.

(a) Initial Complex (b) After Collapse

Figure 5: Removing the minimum vertex of a cube. Consider the directed Euclidean cubical complex in (a), which as a subset of $\mathbb{R}^3$ is a single closed three-cube; call this three-cube σ. Letting $\tau = \min(\sigma)$, we observe that the past link of $\mathbf{v} = \max(\sigma)$ is contractible before the (τ, σ)-collapse and is homeomorphic to $\mathbb{S}^1$ after the collapse. Thus, the past links before and after the collapse are not homotopy equivalent, and so this collapse is not an LPDC.

The following lemma shows under which condition a directed Euclidean cubical collapse induces a simplicial collapse in the past link.

Lemma 3.5 (Vertex Collapses that Induce Simplicial Collapse of Past Links)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\sigma \in \mathcal{K}$ and $\tau, \mathbf{v} \in$ $\mathrm{verts}(\sigma)$ such that $\tau \preceq \mathbf{v}$ and $\tau \neq \min(\sigma)$. If τ is a free face of σ and K' is the (τ, σ)-collapse, then $\mathrm{lk}^-_{K'}(\mathbf{v})$ is the $(\mathbf{v} - \tau)$-collapse of $\mathrm{lk}^-_K(\mathbf{v})$ in the simplicial setting.*

Proof. Consider $K_{\preceq \mathbf{v}}$. Since $\tau, \mathbf{v} \in \mathrm{verts}(\sigma)$ and σ is maximal in K, we know $[\min(\sigma), \mathbf{v}]$ and $[\tau, \mathbf{v}]$ are elementary cubes in $K_{\preceq \mathbf{v}}$. Since τ is a free face of σ, we further know that $[\min(\sigma), \mathbf{v}]$ is the only maximal proper coface of $[\tau, \mathbf{v}]$ in $K_{\preceq \mathbf{v}}$. By definition of past link (Definition 2.2), we then have that $\mathbf{v} - \min(\sigma)$ and $\mathbf{v} - \tau$ are simplices in $\mathrm{lk}^-_{K_{\preceq \mathbf{v}}}(\mathbf{v})$, and $\mathbf{v} - \min(\sigma)$ is the only maximal proper coface of $\mathbf{v} - \tau$ in $\mathrm{lk}^-_{K_{\preceq \mathbf{v}}}(\mathbf{v})$. Hence, $\mathbf{v} - \tau$ is free in $\mathrm{lk}^-_{K_{\preceq \mathbf{v}}}(\mathbf{v})$. Moreover, $\mathrm{lk}^-_{K'_{\preceq \mathbf{v}}}(\mathbf{v})$ is the $(\mathbf{v} - \tau)$-collapse of $\mathrm{lk}^-_{K_{\preceq \mathbf{v}}}(\mathbf{v})$. One can see this by using Statement *2* of Lemma 3.3 by which $\mathrm{lk}^-_{K'_{\preceq \mathbf{v}}}(\mathbf{v})$ can be characterized as the $(\mathbf{v} - \tau)$-collapse of $\mathrm{lk}^-_{K_{\preceq \mathbf{v}}}(\mathbf{v})$.

By Statement *3* of Lemma 2.3, we know that $\mathrm{lk}^-_K(\mathbf{v}) = \mathrm{lk}^-_{K_{\preceq \mathbf{v}}}(\mathbf{v})$ and that $\mathrm{lk}^-_{K'}(\mathbf{v}) = \mathrm{lk}^-_{K'_{\preceq \mathbf{v}}}(\mathbf{v})$, which concludes this proof. □

Next, we prove two lemmas concerning relationships of the past link of a vertex in the original directed Euclidean cubical complex and in the collapsed directed Euclidean cubical complex. These relationships depend on where $\mathbf{v}$ is located with respect to τ. In the first lemma, we consider the case where $\min(\tau) \not\preceq$

(a) Initial Complex (b) After Collapse

Figure 6: Past link of an "uncomparable" vertex before and after a collapse. Consider the directed Euclidean cubical complex shown, comprising a single three-cube σ and all of its faces. Let $\tau = [\mathbf{x}, \mathbf{y}]$. Since $\mathbf{v}$ and $\max(\tau) = \mathbf{y}$ are not comparable, by Lemma 3.6, the past link of $\mathbf{v}$ is the same before and after the collapse. Indeed, we see that this is the case for this example. The past link of $\mathbf{v}$ is the complete complex on two vertices, both before and after.

$\mathbf{v}$, and we present a sufficient condition for past links in K and the (τ, σ)-collapse to be equal. See Fig. 6 for an example that illustrates the result of this lemma.

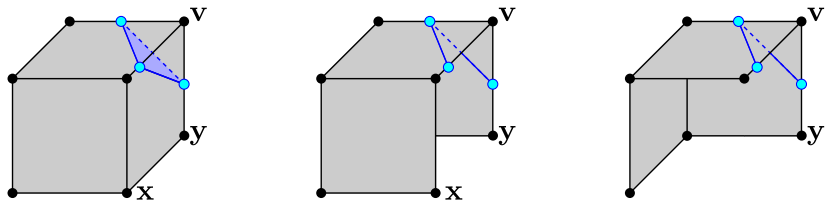
Lemma 3.6 (Condition for Past Links in K and K' to be Equal)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\tau, \sigma \in \mathcal{K}$ such that τ is a face of σ. If τ is a free face of σ and K' is the (τ, σ)-collapse, then, for all $\mathbf{v} \in \operatorname{verts}(K)$ such that $\max(\tau) \not\preceq \mathbf{v}$, we have $\operatorname{lk}^-_K(\mathbf{v}) = \operatorname{lk}^-_{K'}(\mathbf{v})$.*

Proof. By Statement *2* of Lemma 2.3, we have $\operatorname{lk}^-_{K'}(\mathbf{v}) \subseteq \operatorname{lk}^-_K(\mathbf{v})$. Thus, we only need to show $\operatorname{lk}^-_K(\mathbf{v}) \subseteq \operatorname{lk}^-_{K'}(\mathbf{v})$. Suppose $\mathbf{j} \in \operatorname{lk}^-_K(\mathbf{v})$. By the definition of the past link (see Definition 2.2), we know that $[\mathbf{v} - \mathbf{j}, \mathbf{v}]$ is an elementary cube in K. By assumption, $\max(\tau) \not\preceq \mathbf{v}$. Thus, by Eq. (2), $[\mathbf{v} - \mathbf{j}, \mathbf{v}]$ is not removed from K and thus is an elementary cube in K'. Thus, $\mathbf{j} \in \operatorname{lk}^-_{K'}(\mathbf{v})$. □

In the following lemma, we consider the case where $\max(\tau) \preceq \mathbf{v}$, and we present a sufficient condition for past links in the (τ, σ)-collapse and the $(\min(\tau), \sigma)$-collapse to be equal. See Fig. 7 for an example that illustrates this result.

Lemma 3.7 (Comparing Past Links in a General Collapse with Past Links in a Vertex Collapse)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$ such that there exists cubes $\tau, \sigma \in \mathcal{K}$ with $\min(\tau)$ a free face of σ. Let K' be the (τ, σ)-collapse and let $\widehat{K}$ be the $(\min(\tau), \sigma)$-collapse. If $\mathbf{v} \in \operatorname{verts}(K')$ and $\max(\tau) \preceq \mathbf{v}$, then $\mathbf{v} \in \operatorname{verts}(\widehat{K})$ and $\operatorname{lk}^-_{K'}(\mathbf{v}) = \operatorname{lk}^-_{\widehat{K}}(\mathbf{v})$.*

Proof. We first show $\mathbf{v} \in \operatorname{verts}(\widehat{K})$. If τ is a zero-cube (and hence in $\operatorname{verts}(K)$), then $K' = \widehat{K}$, which means that $\mathbf{v} \in \operatorname{verts}(\widehat{K})$. On the other hand, if τ is not a zero-cube, then we have $\min(\tau) \prec \max(\tau) \preceq \mathbf{v}$. In particular, $\min(\tau) \neq \mathbf{v}$. And



(a) Original Complex (b) The (τ, σ)-Collapse (c) The $(\mathbf{x}, \sigma)$-Collapse

Figure 7: Two collapses with same past links. For example, in the directed Euclidean cubical complex K shown in (a), let σ be the three-cube, and let $\tau = [\mathbf{x}, \mathbf{y}]$. We look at the past link of the vertex $\mathbf{v}$. In the original directed Euclidean cubical complex, the past link of $\mathbf{v}$ is the complete complex on three vertices. By Lemma 3.7, the past link of $\mathbf{v}$ is the same in both the (τ, σ)-collapse and the $(\mathbf{x}, \sigma)$-collapse since $\max(\tau) = \mathbf{y} \preceq \mathbf{v}$. By Lemma 3.8, we also know that the past links of $\mathbf{v}$ in K and the $(\mathbf{x}, \sigma)$-collapse are homotopy equivalent. Indeed, we see that this is the case.

so, by definition of $\widehat{K}$ as a $(\min(\tau), \sigma)$-collapse and since $\mathbf{v} \in \mathcal{K}$, we conclude that $\mathbf{v} \in \widehat{K}$.

Next, we show $\mathrm{lk}^-_{K'}(\mathbf{v}) = \mathrm{lk}^-_{\widehat{K}}(\mathbf{v})$. By Statement *2* of Lemma 2.3, we have $\mathrm{lk}^-_{\widehat{K}}(\mathbf{v}) \subseteq \mathrm{lk}^-_{K'}(\mathbf{v})$. Thus, what remains to be proven is $\mathrm{lk}^-_{K'}(\mathbf{v}) \subseteq \mathrm{lk}^-_{\widehat{K}}(\mathbf{v})$. Let $\mathbf{j} \in \mathrm{lk}^-_{K'}(\mathbf{v})$. By definition of the past link (Definition 2.2), we know that $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq K'$. Consider two cases: $\mathbf{v} - \mathbf{j} \preceq \min(\tau)$ and $\mathbf{v} - \mathbf{j} \not\preceq \min(\tau)$.

Case 1 ($\mathbf{v} - \mathbf{j} \preceq \min(\tau)$): Since $\mathbf{v} - \mathbf{j} \preceq \min(\tau) \preceq \max(\tau) \preceq \mathbf{v}$, we know that $\overline{\tau} \subseteq [\mathbf{v} - \mathbf{j}, \mathbf{v}]$. Thus, by Eq. (2), we have $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \nsubseteq K'$, which is a contradiction. So, Case 1 cannot happen.

Case 2 ($\mathbf{v} - \mathbf{j} \not\preceq \min(\tau)$): If $\mathbf{v} - \mathbf{j} \not\preceq \min(\tau)$, then, by the definition of a $(\min(\tau), \sigma)$-collapse in Definition 3.1, we know that $[\mathbf{v} - \mathbf{j}, \mathbf{v}] \subseteq \widehat{K}$ and thus $\mathbf{j} \in \mathrm{lk}^-_{\widehat{K}}(\mathbf{v})$.

Hence, $\mathrm{lk}^-_{K'}(\mathbf{v}) \subseteq \mathrm{lk}^-_{\widehat{K}}(\mathbf{v})$. Since we have both subset inclusions, we conclude $\mathrm{lk}^-_{K'}(\mathbf{v}) = \mathrm{lk}^-_{\widehat{K}}(\mathbf{v})$. □

In general, the minimal vertex of τ is not free in K and hence, there is no vertex collapse. In the main theorem, the previous lemma is applied to a subcomplex of K; specifically, it is applied to the restriction to the unit cube corresponding to σ, where all vertices, inculding $\min \tau$ are then free. The results carry over to K.

The next result states that vertex collapses result in homotopy equivalent past links as long as we are not collapsing the minimum vertex of the directed

Euclidean cubical complex.

Lemma 3.8 (Past Links in a Vertex Collapse)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\sigma \in \mathcal{K}$ and let $\tau \in \operatorname{verts}(\sigma)$ such that $\tau \neq \min(\sigma)$. Let $\mathbf{v} \in \operatorname{verts}(K)$ with $\mathbf{v} \neq \tau$. If τ is a free face of σ and K' is the (τ, σ)-collapse, then $\operatorname{lk}_K^-(\mathbf{v}) \simeq \operatorname{lk}_{K'}^-(\mathbf{v})$.*

Proof. We consider three cases:

Case 1 ($\mathbf{v} \notin \operatorname{verts}(\sigma)$): By definition of past link (Definition 2.2), if $\mathbf{v} \notin \operatorname{verts}(\sigma)$, then the past links $\operatorname{lk}_K^-(\mathbf{v})$ and $\operatorname{lk}_{K'}^-(\mathbf{v})$ are equal.

Case 2 ($\tau \not\preceq \mathbf{v}$): By Lemma 3.6, if $\tau = \max(\tau) \not\preceq \mathbf{v}$, again we have equality of the past links $\operatorname{lk}_K^-(\mathbf{v})$ and $\operatorname{lk}_{K'}^-(\mathbf{v})$.

Case 3 ($\mathbf{v} \in \operatorname{verts}(\sigma)$ and $\tau \preceq \mathbf{v}$): By Lemma 3.5, we know that $\operatorname{lk}_{K'}^-(\mathbf{v})$ is the $\mathbf{v} - \tau$-collapse of $\operatorname{lk}_K^-(\mathbf{v})$ in the simplicial setting. Since simplicial collapses preserve the homotopy type (see e.g., [14, Proposition 6.14]), we conclude $\operatorname{lk}_K^-(\mathbf{v}) \simeq \operatorname{lk}_{K'}^-(\mathbf{v})$.

□

We give an example of Lemma 3.8 in Fig. 7 by showing how the LPDC induces a simplicial collapse on past links.

Lastly, we are ready to prove the main result.

Theorem 3.9 (Main Theorem)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$ such that there exist cubes $\tau, \sigma \in \mathcal{K}$ with τ a free face of σ. Then, (τ, σ) is an LPDC pair if and only if $\min(\sigma) \notin \operatorname{verts}(\tau)$.*

Proof. Let $\mathbf{v} = \max(\sigma)$ and $k = \dim(\sigma)$. Let $(K', \mathcal{K}')$ be the (τ, σ)-collapse of K. Let $(L, \mathcal{L})$ be the cubical complex such that $L = K|_\sigma$. Since $\sigma \in K$, we know $L = \overline{\sigma}$ (i.e., L is a unit cube). Since L is a single unit cube and σ is a maximal elementary cube, all proper faces of σ, including τ and $\min(\tau)$, are free faces in L. Thus, let $(L', \mathcal{L}')$ be the (τ, σ)-collapse of L, and let $(\widehat{L}, \widehat{\mathcal{L}})$ be the $(\min(\tau), \sigma)$-collapse of L.

We first prove the forward direction by contrapositive (if $\min(\sigma) \in \operatorname{verts}(\tau)$, then (τ, σ) is not an LPDC pair). Assume $\min(\sigma) \in \operatorname{verts}(\tau)$. By Corollary 3.4, we obtain $\operatorname{lk}_L^-(\mathbf{v})$ is homeomorphic to $\mathbb{B}^{d-1}$ and $\operatorname{lk}_{\widehat{L}}^-(\mathbf{v})$ is homeomorphic to $\mathbb{S}^{k-2}$. Since $\min(\sigma) \in \operatorname{verts}(\tau)$, we know that $\min(\sigma) = \min(\tau)$. Since τ is a face of σ, we know $\max(\tau) \preceq \max(\sigma) = \mathbf{v}$. Since $\min(\sigma) = \min(\tau) \in \operatorname{verts}(\tau)$ and since τ is a proper face of σ, we know that $\mathbf{v} \neq \max(\tau)$. Thus, $\mathbf{v} \in \operatorname{verts}(L')$. Applying Lemma 3.7, we obtain $\operatorname{lk}_{L'}^-(\mathbf{v}) = \operatorname{lk}_{\widehat{L}}^-(\mathbf{v})$. Putting this all together, we have:

$$\operatorname{lk}_L^-(\mathbf{v}) \simeq \mathbb{B}^{d-1} \not\simeq \mathbb{S}^{k-2} \simeq \operatorname{lk}_{\widehat{L}}^-(\mathbf{v}) = \operatorname{lk}_{L'}^-(\mathbf{v}),$$

and so $\mathrm{lk}_L^-(\mathbf{v}) \not\simeq \mathrm{lk}_{L'}^-(\mathbf{v})$.

Since no faces of σ are in $\mathcal{K} \setminus \mathcal{L}$, the past link of $\mathbf{v}$ remains the same outside of L in both K and K'. Thus, $\mathrm{lk}_K^-(\mathbf{v}) \not\simeq \mathrm{lk}_{K'}^-(\mathbf{v})$ and so we conclude that (τ, σ) is not an LPDC pair, as was to be shown.

Next, we show the backwards direction. Suppose $\min(\sigma) \notin \mathrm{verts}(\tau)$. Let $\mathbf{v} \in \mathrm{verts}(K')$, and consider two cases: $\max(\tau) \not\preceq \mathbf{v}$ and $\max(\tau) \preceq \mathbf{v}$.

Case 1 ($\max(\tau) \not\preceq \mathbf{v}$): By Lemma 3.6, we have $\mathrm{lk}_K^-(\mathbf{v}) = \mathrm{lk}_{K'}^-(\mathbf{v})$. Hence, $\mathrm{lk}_K^-(\mathbf{v}) \simeq \mathrm{lk}_{K'}^-(\mathbf{v})$. Since $\mathbf{v}$ was arbitrarily chosen, we conclude that (τ, σ) is an LPDC pair.

Case 2 ($\max(\tau) \preceq \mathbf{v}$): By Lemma 3.7, we have that $\mathrm{lk}_{L'}^-(\mathbf{v}) = \mathrm{lk}_{\widehat{L}}^-(\mathbf{v})$. Since $\min(\sigma) \notin \mathrm{verts}(\tau)$, we know that $\min(\tau) \neq \min(\sigma)$. Applying Lemma 3.8, we obtain $\mathrm{lk}_L^-(\mathbf{v}) \simeq \mathrm{lk}_{\widehat{L}}^-(\mathbf{v})$. Again, since no faces of σ are removed from $\mathcal{K}$ and $\mathcal{K}'$ to obtain $\mathcal{L}$ and $\mathcal{L}'$, the past link of $\mathbf{v}$ remains the same outside of L in both K and K'. Thus, $\mathrm{lk}_K^-(\mathbf{v}) \simeq \mathrm{lk}_{K'}^-(\mathbf{v})$. Since $\mathbf{v}$ was arbitrarily chosen, we conclude that (τ, σ) is an LPDC pair. $\square$

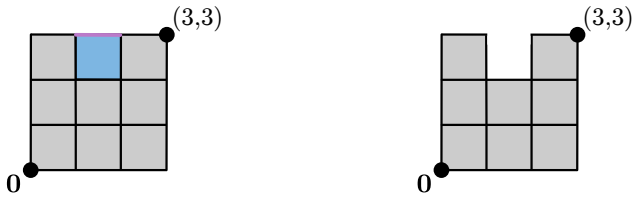
4 Preservation of Spaces of Dipaths

In [1], we proved several results on the relationships between past links and spaces of dipaths. One result, Theorem 2.4, states that for a directed Euclidean cubical complex with a minimum vertex, if all past links are contractible, then all spaces of dipaths starting at that minimum vertex are also contractible. If we start with a directed Euclidean cubical complex with a minimum vertex that has all contractible past links, then all spaces of dipaths from the minimum vertex are contractible by this theorem. We explain how those relationships extend to the LPDC setting in this section.

Applying an LPDC preserves the homotopy type of past links by definition. Hence, applying the theorem again, we see that any LPDC also has contractible dipath spaces from the minimum vertex. Notice that the minimum vertex is not removed in an LDPC, since it is a vertex and minimal in all cubes containing it (including the maximal cube). We give an example of this in Example 4.1.

Example 4.1 (3×3 filled grid)**.** Let K be the 3×3 filled grid. For all $\mathbf{v} \in \mathrm{verts}(K)$, $\mathrm{lk}_K^-(\mathbf{v})$ is contractible. By Theorem 2.4, this implies that all spaces of dipaths starting at $\mathbf{0}$ are contractible. Applying an LPDC such as the edge $[(1,3),(2,3)]$ results in contractible past links in K' and so all spaces of dipaths in K' are also contractible. See Fig. 8. We can generalize this example to any k^d filled grid where $k, d \in \mathbb{N}$.

An analogous result holds for connectedness (Theorem 2.5). If we start with a directed Euclidean cubical complex such that all past links are connected,



(a) Initial Complex (b) After Collapse

Figure 8: (a) The 3×3 filled grid has contractible past links and dipath spaces. The pair comprising of the purple edge $[(1,3),(2,3)]$ and the blue square $[(1,2),(2,3)]$ is an LPDC pair. (b) The result of performing the LPDC. All past links are contractible and so all dipath spaces are also contractible.

then all dipath spaces are connected. Any LPDC results in a directed Euclidean cubical complex that also has connected dipath spaces. See Example 4.2.

Example 4.2 (Outer Cubes of the $5 \times 5 \times 5$ Grid)**.** Let $K = [0,5]^3 \setminus [1,4]^3$, which, as an undirected complex, is homeomorphic to a thickened two-sphere. For all $\mathbf{v} \in \text{verts}(K)$, $\text{lk}_K^-(\mathbf{v})$ is connected. By Theorem 2.5, this implies that for all $\mathbf{v} \in \text{verts}(K)$, the space of dipaths $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{v}}(K)$ is connected. Applying an LPDC such as with the vertex $(5,0,0)$ in the cube $[(4,0,0),(5,1,1)]$ results in connected past links in K' and so all spaces of dipaths $\overrightarrow{P}_{\mathbf{0}}^{\mathbf{v}}(K')$ are connected. We can generalize this example to any k^d grid where $d \geq 3$ and the inner cubes of dimension d are removed.

Both Theorem 2.4 and Theorem 2.5 have assumptions on the topology of past links and results on the topology of spaces of dipaths from the minimum vertex. We may ask if the converse statements are true. Does knowing the topology of spaces of dipaths from the minimum vertex tell us anything about the topology of past links? The converse to Theorem 2.4 holds. To prove this, we first need a lemma whose proof appears in [21].

Lemma 4.3 (Homotopy Equivalence [21, Prop. 5.3])**.** *Let $(K,\mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$. Let $\mathbf{p},\mathbf{q} \in \mathbb{Z}^n$. If $\overrightarrow{P}_{\mathbf{p}}^{\mathbf{q}-\mathbf{j}}(K)$ is contractible for all $\mathbf{j} \in \text{lk}_K^-(\mathbf{q})$, then $\overrightarrow{P}_{\mathbf{p}}^{\mathbf{q}}(K) \simeq \text{lk}_{K_{\mathbf{p}\preceq}}^-(\mathbf{q})$.*

Thus, we obtain:

Theorem 4.4 (Contractability)**.** *Let $(K,\mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$ that has a minimum vertex $\mathbf{w}$. The following two statements are equivalent:*

1. *For all $\mathbf{v} \in \text{verts}(K)$, the space of dipaths $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K)$ is contractible.*

2. For all $\mathbf{v} \in \mathrm{verts}(K)$*, the past link* $\mathrm{lk}_K^-(\mathbf{v})$ *is contractible.*

Proof. By Theorem 2.4, we obtain Statement *2* implies Statement *1*.

Next, we show that Statement *1* implies Statement *2*. Let $\mathbf{v} \in \mathrm{verts}(K)$. For all $\mathbf{j} \in \mathrm{lk}_K^-(\mathbf{v})$, the cube $[\mathbf{v}-\mathbf{j}, \mathbf{v}]$ is a subset of K, which means that $\mathbf{v}-\mathbf{j} \in \mathrm{verts}(K)$. Thus, by assumption, all dipath spaces $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}-\mathbf{j}}(K)$ are contractible. By Lemma 4.3, we know that $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K) \simeq \mathrm{lk}_{K_{\mathbf{w}\preceq}}^-(\mathbf{v}) = \mathrm{lk}_K^-(\mathbf{v})$. Again, since $\mathbf{v} \in \mathrm{verts}(K)$, the dipath space $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K)$ is contractible. Therefore, $\mathrm{lk}_K^-(\mathbf{v})$ is contractible. □

As a consequence of this theorem, we know that if we start with a directed Euclidean cubical complex with contractible dipath spaces starting at the minimum vertex, then any LPDC also result in a directed Euclidean cubical complex with all contractible dipath spaces starting at the minimum vertex, and vice versa.

Corollary 4.5 (Preserving Directed Path Space Contractability)**.** *Let* $(K, \mathcal{K})$ *be a directed Euclidean cubical complex in* $\mathbb{R}^n$ *that has a minimum vertex* $\mathbf{w}$*. Let* $\tau, \sigma \in \mathcal{K}$ *such that* τ *is a face of* σ*. If* τ *is a free face of* σ*, let* $(K', \mathcal{K}')$ *be the* (τ, σ)*-collapse. If* K' *is an LPDC of* K*, then the spaces of dipaths* $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K)$ *are contractible for all* $\mathbf{v} \in \mathrm{verts}(K)$ *if and only if the spaces of dipaths* $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{k}}(K')$ *are contractible for all* $\mathbf{k} \in \mathrm{verts}(K')$*.*

Proof. We start with the forwards direction by assuming that the spaces of dipaths $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K)$ are contractible for all $\mathbf{v} \in \mathrm{verts}(K)$. Theorem 4.4 tells us that all past links $\mathrm{lk}_K^-(\mathbf{v})$ are contractible for all $\mathbf{v} \in \mathrm{verts}(K)$. This implies that $\mathrm{lk}_{K'}^-(\mathbf{k})$ is contractible for all $\mathbf{k} \in \mathrm{verts}(K')$ because K' is an LPDC of K. Applying Theorem 4.4 again, we see that all spaces of dipaths $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{k}}(K')$ are contractible for all $\mathbf{k} \in \mathrm{verts}(K')$.

Next we prove the backwards direction by assuming that the spaces of dipaths $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{k}}(K')$ are contractible for all $\mathbf{k} \in \mathrm{verts}(K')$. Let $\mathbf{v} \in \mathrm{verts}(K)$. Either $\mathbf{v} \in \mathrm{verts}(K')$ or $\mathbf{v} \notin \mathrm{verts}(K')$.

Case 1 ($\mathbf{v} \in \mathrm{verts}(K')$): By Theorem 4.4, we know that $\mathrm{lk}_{K'}^-(\mathbf{v})$ is contractible. Since K' is an LPDC of K, then $\mathrm{lk}_K^-(\mathbf{v})$ is also contractible.

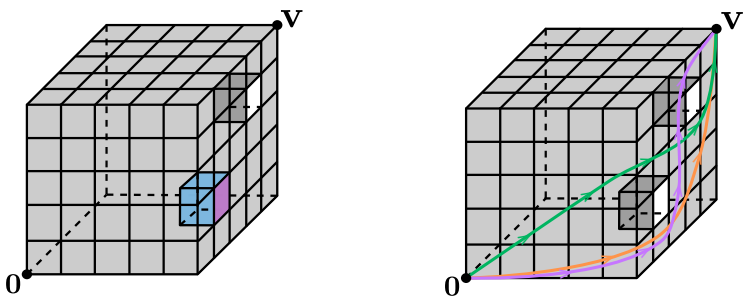
Case 2 ($\mathbf{v} \notin \mathrm{verts}(K')$): If $\mathbf{v} \notin \mathrm{verts}(K)$, then τ is a vertex and $\mathbf{v} = \tau$. Observe that $\mathrm{lk}_{\overline{\sigma}}^-(\tau)$ is contractible since $\overline{\sigma}$ is an elementary cube and τ does not contain $\min(\sigma)$. Furthermore, notice that $\mathrm{lk}_K^-(\tau) = \mathrm{lk}_\sigma^-(\tau)$ because τ is a free face of σ. Hence, $\mathrm{lk}_K^-(\tau)$ is contractible.

Therefore $\mathrm{lk}_K^-(\mathbf{v})$ is contractible for all $\mathbf{v} \in \mathrm{verts}(K)$. Applying Theorem 4.4, we get that $\overrightarrow{P}_{\mathbf{w}}^{\mathbf{v}}(K)$ is contractible for all $\mathbf{v} \in \mathrm{verts}(K)$. □

Using Theorem 2.5 and the partial converse to the connectedness theorem [1, Theorem 3], we get that any LPDC of a directed Euclidean cubical complex with connected dipath spaces and reachable vertices results in a directed Euclidean cubical complex with connected dipath spaces.

Corollary 4.6 (Condition for LPDCs to Preserve Connectedness of All Directed Path Spaces)**.** *Let $(K, \mathcal{K})$ be a directed Euclidean cubical complex in $\mathbb{R}^n$ that has a minimum vertex $\mathbf{w}$. Let $(L, \mathcal{L}) = \mathrm{reach}(K, \mathbf{w})$. Let (τ, σ) be an LPDC pair in L, and let L' be the (τ, σ)-collapse. The spaces of dipaths in $\overrightarrow{P}^{\mathbf{k}}_{\mathbf{w}}(L)$ are connected for all $\mathbf{v} \in \mathrm{verts}(L)$ if and only if the spaces of dipaths $\overrightarrow{P}^{\mathbf{v}}_{\mathbf{w}}(L')$ are connected for all $\mathbf{v} \in \mathrm{verts}(L')$.*

We note that reachability is a necessary condition. Below we give an example of a directed Euclidean cubical complex K that has all connected dipath spaces but an LPDC yields a directed Euclidean cubical complex with a disconnected path space.



(a) Original Complex (b) After Collapse

Figure 9: The bowling ball before and after the collapse described in Example 4.7. Observe $\overrightarrow{P}^{(5,5,5)}_{\mathbf{0}}(K)$ has one connected component. Additionally, $\sigma = [(4,1,1),(5,2,2)]$ (highlighted in blue) and $\tau = [(5,1,1),(5,2,2)]$ (highlighted in purple) is an LPDC pair. After collapsing (τ, σ), $\overrightarrow{P}^{(5,5,5)}_{\mathbf{0}}(K)$ changes from having one connected component to three connected components. The three connected components are represented by the three dipaths.

Example 4.7 (Bowling Ball)**.** Let K be the boundary of the $5 \times 5 \times 5$ grid union $[(4,1,1),(5,2,2)]$ and $[(4,3,3),(5,4,4)] \setminus [(5,3,3),(5,4,4)]$. See Fig. 9(a). Notice that some vertices of K are unreachable, for example, vertex $(4,1,1)$. Furthermore, all past links of vertices in K are connected and so all dipath spaces starting at $\mathbf{0}$ are also connected. After performing an LPDC with $\tau = [(5,1,1),(5,2,2)]$ and $\sigma = [(4,1,1),(5,2,2)]$, the dipath space between $\mathbf{0}$ and

$(5,5,5)$ changes from having one connected component to three connected components, as shown in the figure. This example shows that the reachability condition in Corollary 4.6 is necessary for preserving connnectedness in LPDCs.

LPDCs can also preserve dihomotopy classes of dipaths starting at the minimum vertex of many directed Euclidean cubical complexes that have disconnected past links. Recall the Swiss flag as discussed in Fig. 4. The Swiss flag has disconnected past links at $(3,4)$ and $(4,3)$, yet there exists a sequence of LPDCs that results in a directed Euclidean cubical complex that highlights the two dihomotopy classes of dipaths between $\mathbf{0}$ and (5.5). Example 4.8 gives another similar situation.

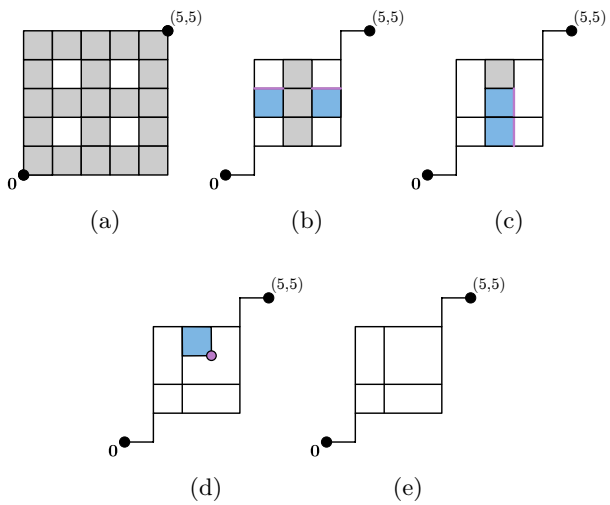


Figure 10: Link-preserving DCs of the window. A sequence of LPDCs is presented from (a)-(e). The directed Euclidean cubical complex in (b) comes from performing several vertex LPDCs to remove the two-cubes along the border of K. In (b)-(d), the LPDC pairs (τ, σ) are highlighted in purple and blue respectively. The result of the sequence of LPDCs is a graph of vertices and edges that more clearly illustrates the dihomotopy classes of dipaths in the dipath space.

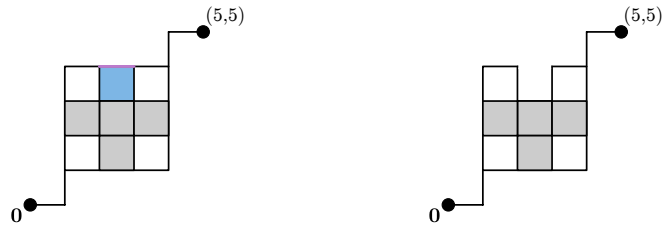
Example 4.8 (Window)**.** Let K be the 5×5 grid with the following two-cube interiors removed: $[(1,1),(2,2)]$, $[(3,1),(4,2)]$, $[(1,3),(2,4)]$, $[(3,3),(4,4)]$. See Fig. 10(a). K has disconnected past links at the vertices $(2,2),(4,2),(2,4),(4,4)$ so K does not satisfy Corollary 4.5 or Corollary 4.6. Observe that $\overrightarrow{P}(K)_{\mathbf{0}}^{(5,5)}$ has six connected components. We can perform a sequence of LPDCs that preserves the dihomotopy classes of dipaths between $\mathbf{0}$ and $(5,5)$ at each step. First, we apply vertex LPDCs to remove the two-cubes along the border. Then we can

apply four edge LPDCs and one vertex LPDC to get a graph of vertices and edges. This graph more clearly illustrates the six dihomotopy classes of dipaths in $\overrightarrow{P}(K)_{\mathbf{0}}^{(5,5)}$.

5 Discussion

LPDCs preserve spaces of dipaths in many examples (see Section 4), in particular, if they are all trivial in the sense of either all connected or all contractible and the directed Euclidean cubical complex is reachable for the minimum. However, LPDCs do not always preserve spaces of dipaths. We discuss some of those instances here. One limitation of LPDCs is that the number of components may increase after an LPDC as we saw in Example 4.7 or, as we see in Example 5.1, they may decrease.

Example 5.1 (A Sequence of LPDCs of the Window That Decreases the Number of Connected Components of the Dipath Space)**.** Consider K as given in Example 4.8. After applying vertex LPDCs that remove the two-cubes on the border of K, we can apply an LPDC to the edge $[(2,4),(3,4)]$. Now $\overrightarrow{P}(K')_{\mathbf{0}}^{(5,5)}$ has five connected components; whereas, the dipath space $\overrightarrow{P}(K)_{\mathbf{0}}^{(5,5)}$ has six connected components. See Fig. 11. This example shows that there are both "good"



(a) Initial Complex (b) After Collapse

Figure 11: Link-preserving DC of the window that changes dipath space. The LPDC of the edge $[(2,4),(3,4)]$ changes the dipath space between $\mathbf{0}$ to $(5,5)$ from having six connected components to five connected components.

and "bad" ways to apply a sequence of LPDCs to a directed Euclidean cubical complex. As illustrated in Example 4.8, there exists a sequence of LPDCs that preserves the six connected components in $\overrightarrow{P}(K)_{\mathbf{0}}^{(5,5)}$. However, if we perform a sequence of LPDCs that removes the edge $[(2,4),(3,4)]$ as in this example, then we get a directed Euclidean cubical complex that does not preserve the dihomotopy classes of dipaths in $\overrightarrow{P}(K)_{\mathbf{0}}^{(5,5)}$.

Example 5.1 illustrates the need to investigate other properties if we want to preserve dipath spaces when performing an LPDC.

In Example 4.7, the problem was the existence of unreachable vertices. In Example 5.1, the vertex $(2, 4)$ is a *deadlock* after the LPDC: only trivial dipaths initiate from there; whereas, before collapse, that was not the case. This seems to suggest that the introduction of new deadlocks should not be allowed; in practice, this would require an extra—but computationally easy—check on vertices of σ.

In the non-directed setting, if K' is obtained from K by collapsing a collapsing pair (τ, σ), then not only is the inclusion of K' in K a homotopy equivalence. K' is a deformation retract of K. The following example removes any hope of such a result in the directed setting:

Example 5.2 (LDPC of the Four-Cube With No Directed Retraction to the Collapsed Complex)**.** Let $(I^4, \mathcal{I}^4)$ be the standard unit four-cube. Let τ be the vertex $(1, 1, 0, 0)$, and σ be the cube $[\mathbf{0}, \mathbf{1}]$. Since τ is free and not the minimum vertex of σ, the pair (τ, σ) is an LPDC pair. Thus, let $(K', \mathcal{K}')$ be the collapsed complex. Next, we show that there is no directed retration, i.e., no directed map from I^4 to K' that is the identity on K'.

Suppose, for a contradiction, that $f : I^4 \to K'$ is such a directed retraction. Let $\mathbf{p}_1 = (0, 1, 0, 0)$, $\mathbf{p}_2 = (1, 0, 0, 0)$, $\mathbf{q}_1 = (1, 1, 1, 0)$, and $\mathbf{q}_2 = (1, 1, 0, 1)$. By the product order on $\mathbb{R}^4$, we have $\mathbf{p}_1, \mathbf{p}_2 \preceq \tau$ and $\tau \preceq \mathbf{q}_1, \mathbf{q}_2$. Since the points $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{q}_1$, and $\mathbf{q}_2$ are vertices of I^4 and are not equal to τ, we also know that $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{q}_1$, and $\mathbf{q}_2$ are points in K'. Since f is a directed retraction, we have that $\mathbf{p}_1 = f(\mathbf{p}_1) \preceq f(\tau)$ and that $\mathbf{p}_2 = f(\mathbf{p}_2) \preceq f(\tau)$. Similarly, we obtain that $f(\tau) \preceq f(\mathbf{q}_1) = \mathbf{q}_1$ and that $f(\tau) \preceq f(\mathbf{q}_2) = \mathbf{q}_2$.

Let $x_1, x_2, x_3, x_4 \in I$ such that $f(\tau) = (x_1, x_2, x_3, x_4)$. Then,

$$\begin{aligned}
\mathbf{p}_1 \preceq f(\tau) &\Rightarrow x_2 \geq 1 \text{ and hence } x_2 = 1,\\
\mathbf{p}_2 \preceq f(\tau) &\Rightarrow x_1 \geq 1 \text{ and hence } x_1 = 1,\\
f(\tau) \preceq \mathbf{q}_1 &\Rightarrow x_4 \leq 0 \text{ and hence } x_4 = 0,\\
f(\tau) \preceq \mathbf{q}_2 &\Rightarrow x_3 \leq 0 \text{ and hence } x_3 = 0.
\end{aligned}$$

Thus, $f(\tau) = (1, 1, 0, 0) = \tau$, which is not in K' and hence a contradiction. In fact, this argument extends to $(I^k, \mathcal{I}^k)$ for $k \geq 4$.

As further evidence that such a (τ, σ)-collapse does not preserve the directed topology, consider the spaces of dipaths in $(I^4, \mathcal{I}^4)$ and $(K', \mathcal{K}')$. We would need dipaths in the original space to map to dipaths in the collapsed space. However, notice that the dipath from $\mathbf{p}_1$ to $\mathbf{q}_1$ through τ cannot be mapped to a dipath in $(K', \mathcal{K}')$.

We observe that vertex LPDCs appear to not introduce the problems of unreachability and deadlocks. These observations lead us to suspect that studying unreachability, deadlocks, and vertex LPDCs can help us better understand when LPDCs preserve and do not preserve dipath spaces between the minimum and a given vertex. We leave this as future work.

In summary, we provide an easy criterion for determining when we have an LPDC pair, as well as discuss various settings for when LPDCs preserve spaces of dipaths. Fully understanding when LPDCs preserve spaces of dipaths between two given vertices is a step towards developing algorithms that compress directed Euclidean cubical complexes and preserve directed topology.

Acknowledgements This research is a product of one of the working groups at the Women in Topology (WIT) workshop at MSRI in November 2017. This workshop was organized in partnership with MSRI and the Clay Mathematics Institute, and was partially supported by an AWM ADVANCE grant (NSF-HRD 1500481).

This material is based upon work supported by the US National Science Foundation under grant No. DGE 1649608 (Belton) and DMS 1664858 (Fasy), as well as the Swiss National Science Foundation under grant No. 200021-172636 (Ebli).

We thank the Computational Topology and Geometry (CompTaG) group at Montana State University for giving helpful feedback on drafts of this work.

References

[1] Robin Belton, Robyn Brooks, Stefania Ebli, Lisbeth Fajstrup, Brittany Terese Fasy, Catherine Ray, Nicole Sanderson, and Elizabeth Vidaurre. Towards directed collapsibility. In Bahar Acu, Donatella Danielli, Marta Lewicka, A.N. Pati, S. Ramanathapuram Vancheeswara, and M.I. Teboh-Ewungkem, editors, *Advances in Mathematical Sciences*, Association for Women in Mathematics Series. Springer, United States, 2020.

[2] Ronald Brown and Philip J. Higgins. On the algebra of cubes. *Journal of Pure and Applied Algebra*, 21(3):233–60, 1981.

[3] Marshall M. Cohen. *A Course in Simple-Homotopy Theory*, volume 10. Springer Science & Business Media, 2012.

[4] Edsger W. Dijkstra. Two starvation-free solutions of a general exclusion problem. 1977. Manuscript EWD625, from the archives of UT Austin,

[5] Lisbeth Fajstrup, Eric Goubault, Emmanuel Haucourt, Samuel Mimram, and Martin Raussen. *Directed Algebraic Topology and Concurrency.* Springer, 1st edition, 2016.

[6] Lisbeth Fajstrup, Eric Goubault, and Martin Raussen. Algebraic topology and concurrency. *Theoretical Computer Science*, pages 241–271, 2006.

[7] Robert Ghrist. Configuration spaces, braids, and robotics. In *Braids: Introductory Lectures on Braids, Configurations and Their Applications*, pages 263–304. World Scientific, 2010.

[8] Eric Goubault. *The Geometry of Concurrency.* PhD thesis, École Normale Supérieure, Paris, France, 1995.

[9] Marco Grandis. Directed homotopy theory, I: The fundamental category. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, 44(4):281–316, 2003.

[10] Marco Grandis. *Directed Algebraic Topology: Models of Non-Reversible Worlds*, volume 13 of *New Mathematical Monographs.* Cambridge University Press, 2009.

[11] A. Hatcher. *Algebraic Topology.* Algebraic Topology. Cambridge University Press, 2002.

[12] Charles A.R. Hoare. Communicating sequential processes. *Commun. ACM*, 21(8):666–677, August 1978.

[13] John F. Jardine. Cubical homotopy theory: A beginning, 2002. Technical report.

[14] Dimitry Kozlov. *Combinatorial Algebraic Topology*, volume 21 of *Algorithms and Computation in Mathematics.* Springer-Verlag Berlin Heidelberg, 2008.

[15] Jacques-Olivier Lachaud. Cubical complex. software https://projet.liris.cnrs.fr/dgtal/doc/nightly/moduleCubicalComplex.html, 2018. part of the Toplogy package of the DGtal library.

[16] Martin Raussen and Krzysztof Ziemiański. Homology of spaces of directed paths on Euclidean cubical complexes. *Journal of Homotopy and Related Structures*, 9(1):67–84, Apr 2014.

[17] Jean-Pierre Serre. Homologie singulière des espaces fibrés. *Annals of Mathematics*, 54(3):425–505, 1951.

[18] John H. C. Whitehead. Simplicial spaces, nuclei and m-groups. *Proceedings of the London Mathematical Society*, s2-45(1):243–327, 1938.

[19] John H. C. Whitehead. Simple homotopy types. *American Journal of Mathematics*, 72(1):1–57, 1950.

[20] Rafal Wisniewski. Towards modelling of hybrid systems. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 911–916. IEEE, 2006.

[21] Krzysztof Ziemiański. On execution spaces of PV-programs. *Theoretical Computer Science*, 619:87–98, March 2016.

Bibliography

[1] S. Ebli, C. Hacker, and K. Maggs, "Morse theoretic signal compression and reconstruction on chain complexes," *ArXiv*, vol. abs/2203.08571, 2022.

[2] S. Ebli, M. Defferrard, and G. Spreemann, "Simplicial neural networks," *Workshop on Toplogical Data Analysis and Beyond, NeurIPS*, 2020.

[3] S. Ebli and G. Spreemann, "A notion of harmonic clustering in simplicial complexes," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 1083–1090.

[4] R. L. Belton, R. Brooks, S. Ebli, L. Fajstrup, B. T. Fasy, C. Ray, N. F. Sanderson, and E. Vidaurre, "Towards directed collapsibility," *Advances in Mathematical Sciences*, 2019.

[5] R. L. Belton, R. Brooks, S. Ebli, L. Fajstrup, B. T. Fasy, N. F. Sanderson, and E. Vidaurre, "Combinatorial conditions for directed collapsing," *ArXiv*, vol. abs/2106.01524, 2021.

[6] A. Hatcher, *Algebraic Topology*, ser. Algebraic Topology. Cambridge University Press, 2002. [Online]. Available: https://books.google.com/books?id=BjKs86kosqgC

[7] E. Sköldberg, "Algebraic Morse theory and homological perturbation theory," *Algebra Discrete Math.*, vol. 26, pp. 124–129, 2018. [Online]. Available: http://mi.mathnet.ru/adm675

[8] J. H. Whitehead, "Simplicial spaces, nuclei and m-groups," *Proceedings of the London Mathematical Society*, vol. s2-45, no. 1, pp. 243–327. [Online]. Available: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-45.1.243

[9] R. H. Bing, "Some aspects of the topology of 3-manifolds related to the poincaré conjecture," *Lectures on Modern Mathematics*, vol. 2, 1964.

[10] R. Forman, "Morse theory for cell complexes," *Advances in Mathematics*, vol. 134, no. 1, pp. 90–145, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001870897916509

[11] K. Mischaikow and V. Nanda, "Morse theory for filtrations and efficient computation of persistent homology," *Discrete & Computational Geometry*, vol. 50, no. 2, pp. 330–353, 2013. [Online]. Available: https://doi.org/10.1007/s00454-013-9529-6

[12] M. Joswig and M. E. Pfetsch, "Computing optimal morse matchings," *SIAM J. Discret. Math.*, vol. 20, pp. 11–25, 2006.

[13] R. Priemer, *Introductory signal processing.* World Scientific, 1991, vol. 6.

[14] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE signal processing magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[15] S. Barbarossa and S. Sardellitti, "Topological signal processing over simplicial complexes," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2992–3007, 2020.

[16] T. M. Roddenberry, M. T. Schaub, and M. Hajij, "Signal processing on cell complexes," 2021.

[17] S. Sardellitti, S. Barbarossa, and L. Testa, "Topological signal processing over cell complexes," *arXiv preprint arXiv:2112.06709*, 2021.

[18] L. J. Grady and J. Polimeni, "Discrete calculus - applied analysis on graphs for computational science," 2010.

[19] B. Eckmann, "Harmonische Funktionen und Randwertaufgaben in einem Komplex," *Commentarii Mathematici Helvetici*, vol. 17, no. 1, pp. 240–255, 1944. [Online]. Available: https://doi.org/10.1007/BF02566245

[20] M. Yang, E. Isufi, M. T. Schaub, and G. Leus, "Simplicial convolutional filters," *arXiv preprint arXiv:2201.11720*, 2022.

[21] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[22] S. Barbarossa, S. Sardellitti, and E. Ceci, "Larning from signal defined over simplicial complexes," in *2018 IEEE Data Science Workshop (DSW)*, 2018, pp. 51–55.

[23] S. Chen, A. Sandryhaila, J. M. Moura, and J. Kovacevic, "Signal denoising on graphs via graph filtering," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2014, pp. 872–876.

[24] D. Zhou and B. Schölkopf, "A regularization framework for learning from graph data," in *ICML 2004 Workshop on Statistical Relational Learning and Its Connections to Other Fields (SRL 2004)*, 2004, pp. 132–137.

[25] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[26] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[27] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in neural information processing systems*, vol. 29, pp. 3844–3852, 2016.

[28] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: going beyond euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[29] A. Grover, A. Zweig, and S. Ermon, "Graphite: Iterative generative modeling of graphs," in *International conference on machine learning*. PMLR, 2019, pp. 2434–2444.

[30] M. T. Schaub, Y. Zhu, J.-B. Seby, T. M. Roddenberry, and S. Segarra, "Signal processing on higher-order networks: Livin' on the edge... and beyond," *Signal Processing*, vol. 187, p. 108149, 2021.

[31] V. Robins, P. J. Wood, and A. P. Sheppard, "Theory and algorithms for constructing discrete morse complexes from grayscale digital images," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1646–1658, 2011.

[32] C. Bodnar, F. Frasca, Y. G. Wang, N. Otter, G. Montúfar, P. Lio, and M. Bronstein, "Weisfeiler and lehman go topological: Message passing simplicial networks," *arXiv preprint arXiv:2103.03212*, 2021.

[33] C. Bodnar, F. Frasca, N. Otter, Y. G. Wang, P. Liò, G. F. Montufar, and M. Bronstein, "Weisfeiler and lehman go cellular: Cw networks," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[34] C. Bodnar, F. D. Giovanni, B. P. Chamberlain, P. Liò, and M. Bronstein, "Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns," *ArXiv*, vol. abs/2202.04579, 2022.

[35] J. Hansen and T. Gebhart, "Sheaf neural networks," 2019.

[36] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

[37] C. Wang, S. Pan, G. Long, X. Zhu, and J. Jiang, "Mgae: Marginalized graph autoencoder for graph clustering," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 889–898.

[38] S. Barbarossa and S. Sardellitti, "Topological signal processing: Making sense of data building on multiway relations," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 174–183, 2020.

[39] E. Bunch, Q. You, G. Fung, and V. Singh, "Simplicial 2-complex convolutional neural nets," *arXiv preprint arXiv:2012.06010*, 2020.

[40] M. Hajij, K. Istvan, and G. Zamzmi, "Cell complex neural networks," *arXiv preprint arXiv:2010.00743*, 2020.

[41] A. D. Keros, V. Nanda, and K. Subr, "Dist2cycle: A simplicial neural network for homology localization," *arXiv preprint arXiv:2110.15182*, 2021.

[42] Y.-C. Chen and M. Meila, "The decomposition of the higher-order homology embedding constructed from the k-laplacian," *Advances in Neural Information Processing Systems*, vol. 34, 2021.

[43] M. Grandis, *Directed Algebraic Topology: Models of Non-Reversible Worlds*, ser. New Mathematical Monographs. Cambridge University Press, 2009.

[44] L. Fajstrup, E. Goubault, E. Haucourt, S. Mimram, and M. Raussen, *Directed Algebraic Topology and Concurrency*, 1st ed. Springer Publishing Company, Inc., 2016.

[45] M. Raussen, "On the classification of dipaths in geometric models for concurrency," *Mathematical Structures in Computer Science*, vol. 10, no. 4, pp. 427–457, 2000.

[46] R. Wisniewski, "Towards modelling of hybrid systems," in *Proceedings of the 45th IEEE Conference on Decision and Control*. IEEE, 2006, pp. 911–916.

[47] R. Ghrist, "Configuration spaces, braids, and robotics," in *Braids: Introductory Lectures on Braids, Configurations and Their Applications*. World Scientific, 2010, pp. 263–304.

[48] P. Dłotko and H. Wagner, "Simplification of complexes for persistent homology computations," *Homology, Homotopy and Applications*, vol. 16, no. 1, pp. 49–63, 2014.

[49] K. Ziemiański, "On execution spaces of PV-programs," *Theoretical Computer Science*, vol. 619, pp. 87–98, 2016.

[50] M. Raussen and K. Ziemiański, "Homology of spaces of directed paths on euclidean cubical complexes," *Journal of Homotopy and Related Structures*, vol. 9, no. 1, pp. 67–84, Apr 2014. [Online]. Available: https://doi.org/10.1007/s40062-013-0045-4

[51] M. Grandis, "Directed homotopy theory, I: The fundamental category," *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, vol. 44, no. 4, pp. 281–316, 2003.