

EXETER COLLEGE, UNIVERSITY OF OXFORD

Topics in Applied Topology: Actions, Knots and Proteins



Panayiota Naya Maria Yerolemou

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2022

Contents

I	Morse Theory for Complexes of Groups	11
1	Loop-Free Categories	15
2	Group Actions	16
3	Complexes of Groups	20
3.1	The Complex of Groups Associated to an Action	20
3.2	Development and the Basic Construction	25
4	The Discrete Flow Category	29
4.1	Entrance Paths	29
4.2	Discrete Morse Theory	30
4.3	Localisation	32
5	The Morse Complex of Groups	34
6	Equivariant Homotopy Equivalence	41
7	An Example	48
II	A Generalised Clock Theorem	53
8	Knot Diagrams and Kauffman States	57
9	From Knots to Discrete Morse Matchings	61
10	Counting Discrete Morse Matchings	67
11	A Generalised Clock Theorem	71
12	Complexes	83
III	Knotoids and Proteins	89
13	Knotted Proteins	92
13.1	Existing Summaries of Knottiness	93
13.2	Deep and Shallow Knotting	95
14	Knotoids	95
15	Folding Hypotheses	97
16	KnotoEMD for Trefoil Proteins	99
16.1	Methods	100
16.2	Results	101
16.3	Conclusions	107

17	Supplementary Information	108
17.1	Trefoil Proteins	108
17.2	Double- and Single-Loop Open Trefoils	109
17.3	Computations	111
18	Towards the Exact Knotoid Distribution	116
	Bibliography	121

To Zoe, for putting everything into perspective.

Acknowledgements

First and foremost, my thanks go to Vidit Nanda and Ulrike Tillmann for giving me the opportunity to complete a DPhil at Oxford. Thanks to Vidit for his supervision and support, and for giving me the academic freedom to explore different areas of mathematics. Thanks also to Ulrike Tillmann and André Henriques for their additional guidance, and to the Alan Turing Institute for funding my DPhil.

Thanks to my examiners, Marc Lackenby and Kathryn Hess, whose feedback improved the clarity of this thesis.

Special thanks go to Daniele, Agnese and Luci who have greatly enriched my professional and personal life over the past few years. Thanks for all of the mentorship, personal support, and for teaching me so much fun maths.

To Karen, Sarah and Ollie - thanks for putting up with me!

Finally, thanks to my family for being a constant source of comfort, support and food.

Introduction

THIS THESIS consists of three parts in which we explore different topics in pure and applied topology; in Part I, we develop an equivariant version of discrete Morse theory compatible with complexes of groups. In Part II, we link discrete Morse theory to knot projections, and in Part III we apply tools from knotoid theory to the study of knotted proteins. The two overall themes running through the thesis are discrete Morse theory and knot theory.

In Part I we construct a version of discrete Morse theory for simplicial complexes endowed with group actions. We characterise the compatibility of an acyclic partial matching on the quotient space with an overlaid complex of groups. We use the discrete flow category of any such compatible matching to build the corresponding Morse complex of groups, and prove that the development of the Morse complex of groups recovers the original simplicial complex up to equivariant homotopy equivalence.

In Part II we obtain a simple and complete characterisation of which matchings on the Tait graph of a knot diagram induce a discrete Morse function (dMf) on S^2 . We show these dMfs are in bijection with certain rooted spanning forests in the Tait graph. We use this to count the number of such dMfs with a closed formula involving the graph Laplacian. We then simultaneously generalise Kauffman's Clock Theorem and Kenyon-Propp-Wilson's correspondence in two different directions; we first prove that the image of the correspondence induces a bijection on perfect dMfs, then we show that all perfect matchings, subject to an admissibility condition, are related by a finite sequence of click and clock moves. Finally, we study and compare the matching and discrete Morse complexes associated to the Tait graph, in terms of partial Kauffman states, and provide some computations.

Understanding how knotted proteins fold is a challenging problem in biology. Researchers have proposed several models for their folding pathways, based on theory, simulations and experiments. The geometry of proteins with the same knot type can vary substan-

tially and recent simulations reveal different folding behaviour for deeply and shallow knotted proteins. In Part III, we analyse proteins forming open-ended trefoil knots by introducing a topologically inspired statistical metric that measures their entanglement. By looking directly at the geometry and topology of their native states, we are able to probe different folding pathways for such proteins. In particular, the folding pathway of shallow knotted carbonic anhydrases involves the creation of a double-looped structure, contrary to what has been observed for other knotted trefoil proteins. We validate this with Molecular Dynamics simulations. By leveraging the geometry and local symmetries of knotted proteins' native states, we provide the first numerical evidence of a double-loop folding mechanism in trefoil proteins.

Part I

Morse Theory for Complexes of Groups

Introduction

The work in this section is joint with Vidit Nanda, and appears in a pre-print (Yerolemou and Nanda, 2022).

In this work, we are interested in two ideas: actions of finite groups on simplicial complexes, and discrete Morse theory. In general, when an action on a topological space is not free, the quotient is not guaranteed to have the same structure as the original space. Under mild conditions on a simplicial action (called *regularity*), we can guarantee a simplicial quotient and simplicial orbit map. The extra data recording the stabilisers decorate the quotient, which together form a *complex of groups*. The complex of groups associated to the trivial action is called the *constant complex of groups*.

Complexes of groups have been studied by Corson (Corson, 1992) and Haefliger (Haefliger, 1991, 1992) in geometric group theory. Accompanying any complex of groups associated to an action is a natural embedding into the constant complex of groups. This pair, consisting of a complex of groups $\mathbf{F}$ and an embedding Φ , can be ‘unfolded’ into a category $\mathcal{D}(\mathbf{F}, \Phi)$ that is endowed with a natural G -action and is G -equivariantly isomorphic to the face poset of the original simplicial complex. The category $\mathcal{D}(\mathbf{F}, \Phi)$ is called the *development* of the pair $(\mathbf{F}, \Phi)$, and the isomorphism result is called the *Basic Construction* (Bridson and Häfliger, Ch. III.C, Thm 2.13).

Discrete Morse theory, developed by Forman (Forman, 1998), is a tool that has been used extensively in the field of applied topology, for example, to increase the efficiency of persistent homology computations (Mischaikow and Nanda, 2013), to develop scalable methods to analyse 3D greyscale images (Robins et al., 2011) and to reconstruct graph-like structures hidden in noisy data (Dey et al., 2018).

Discrete Morse theory provides a mechanism by which one can reduce the size of a simplicial complex by systematically removing ‘well-behaved’ pairs of simplices, whilst preserving the homotopy-type of the space. The collection of these pairs is called an *acyclic partial matching*. An explicit description of this reduced space is given by Nanda, who first introduced the *discrete flow category* (Nanda, 2019). This category is necessarily poset-enriched in order to guarantee that its classifying space has the right homotopy type.

We combine the theories described above by defining a condition on when acyclic partial matchings are compatible with the data of a complex of groups on the quotient. This enables us to track how the complex of groups over the quotient morphs as the collapses occur. We call the modified complex of groups over the discrete flow category the *Morse complex of groups*.

In order to apply the theory developed by Bridson and Haefliger, we first generalise the relevant work to the poset-enriched setting; we prove a generalised version of the Basic Construction for poset-enriched categories. This is then applicable to a complex of groups defined over the discrete flow category. We then prove that the development of this modified complex of groups has the same homotopy type as the original space. Moreover, this procedure retains the G -equivariance property. Our main result can be stated as follows:

Theorem 0.1. *Let G act regularly on a simplicial complex X , let Y denote the quotient X/G , and let $(\mathbf{F}, \Phi)$ denote the complex of groups and embedding associated to this action, for some choice of lifts and transfers. If Σ is an acyclic partial matching on Y compatible with $\mathbf{F}$, then we can construct the Morse complex of groups over the discrete flow category $\mathbf{Flo}_\Sigma[Y]$, and associated embedding Ψ , and the classifying space of the development $\mathcal{D}(\mathbf{M}, \Psi)$ is G -equivariantly homotopy equivalent to X .*

Outline: The work is organised as follows: in Section 1 we introduce loop-free poset-enriched categories, which will be used throughout. We then define group actions on these categories in Section 2, which leads to a well-defined quotient, and introduce *regular* actions on simplicial complexes. In Section 3 we expand the usual definition of complexes of groups to poset-enriched categories, and prove a generalised version of the *Basic Construction* (Proposition 3.11). We then define the discrete flow category in Section 4 and introduce our construction of the *Morse complex of groups* in Section 5, showing it is well-defined. In Section 6 we prove our main theorem (Theorem 6.1), relating to the G -equivariant homotopy-equivalence between the development of the Morse complex of groups over the quotient, and the original space. We present a detailed example in Section 7.

1 Loop-Free Categories

It is well-known that we can summarise a simplicial complex K with its *face poset*, $\mathbf{Fac}[K]$ ¹. This category has objects in bijective correspondence with the simplices of K , and a morphism $x \rightarrow y$ whenever y is a proper face of x . The geometric realisation $|K|$ is homeomorphic to the geometric realisation of this category, so $\mathbf{Fac}[K]$ really is just another representation of the same combinatorial data. Intuitively, we can see that the face poset is ‘loopless’ because the morphisms flow in only one direction, i.e. *down* in dimension, which prevents the formation of any directed cycles. See Figure 1 for an example.

Loop-free categories (Definition 1.1) are a generalisation of poset-like categories, but without having to arise from a simplicial complex.

Definition 1.1. A small category $\mathcal{C}$ is **loop-free** if two conditions hold:

1. for each object x , the poset $\mathcal{C}(x, x)$ contains only the identity morphism, and
2. for objects $x \neq y$, if $\mathcal{C}(x, y)$ is nonempty then $\mathcal{C}(y, x)$ must be empty.

Condition (2) captures the directedness of the morphisms. Note that all face posets arising from simplicial complexes are loop-free, but there are loop-free categories that cannot arise in this way. For example, consider the loop-free category associated to any polygon with more than three sides (see Figure 2).

A simple example of a category failing to be loop-free is the category associated to any non-trivial group G , which has a single object, and a morphism for every element in G (see Figure 3).

We need to work with categories with some additional structure on the hom-sets, and it will become apparent why in Section 4.

Definition 1.2. A small category $\mathcal{C}$ is said to be **poset-enriched** if for any pair of objects x and y in $\mathcal{C}$, the set of morphisms $\mathcal{C}(x, y)$ from x to y is endowed with a partial order $\Rightarrow$ so that across any triple x, y and z of such objects, the map

$$\mathcal{C}(x, y) \times \mathcal{C}(y, z) \xrightarrow{\circ} \mathcal{C}(x, z)$$

sending (f, f') to the composite $f' \circ f$ is monotone.

That is, whenever we have $f_0 \Rightarrow f_1$ in the poset $\mathcal{C}(x, y)$ and $f'_0 \Rightarrow f'_1$ in the poset $\mathcal{C}(y, z)$, then we also have the order relation $f'_0 \circ f_0 \Rightarrow f'_1 \circ f_1$ between composites in the poset $\mathcal{C}(x, z)$.

We will henceforth abbreviate loop-free poset-enriched categories as **LP-categories**. Figure 4 gives an example of such a category.

¹ Wachs 2006 WACHS, Michelle L.: Poset topology: tools and applications. In: *arXiv preprint math/0602226* (2006)

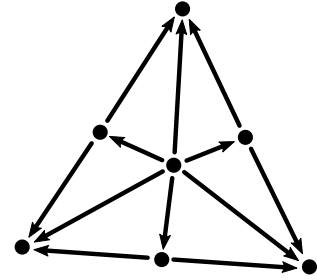


Figure 1: The face poset of the 2-simplex.

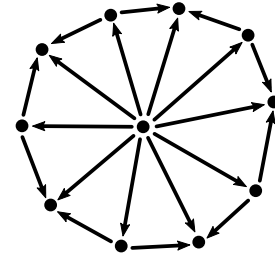


Figure 2: The loop-free category associated to a pentagon. This cannot have arisen as the face poset of any simplicial complex.

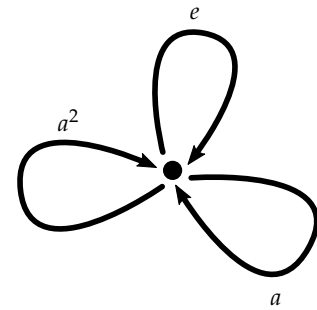


Figure 3: The category associated to a non-trivial group, e.g. $C_3 = \langle a \rangle$, is not loop-free because it fails to satisfy condition (1); it has $|G|$ morphisms from a single object to itself.

Definition 1.3. Given two LP categories $\mathcal{C}$ and $\mathcal{D}$, an **LP functor** $F : \mathcal{C} \rightarrow \mathcal{D}$ is an ordinary functor with the additional requirement that for any pair of objects x, y in $\mathcal{C}$ the induced map

$$F_{xy} : \mathcal{C}(x, y) \rightarrow \mathcal{D}(Fx, Fy)$$

is order-preserving.

To pass from a category to a topological space, we use the construction of a category's classifying space.

Definition 1.4 (Geometric nerve). The **geometric nerve** $\Delta \mathcal{C}$ of an LP-category $\mathcal{C}$ is the simplicial set whose vertices are the objects $\mathcal{C}_{\text{ob}}$, and whose n -simplices spanning objects $x_0, \dots, x_n$ consist of composable morphisms $f_{ij} : x_i \rightarrow x_j$ satisfying $f_{ik} \Rightarrow f_{jk} \circ f_{ij}$, for all $0 \leq i \leq j \leq k \leq n$.

Definition 1.5 (Classifying space). The **classifying space** of an LP-category is the geometric realisation of its nerve.

For an example, See Figure 5.

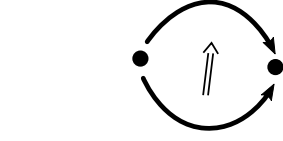
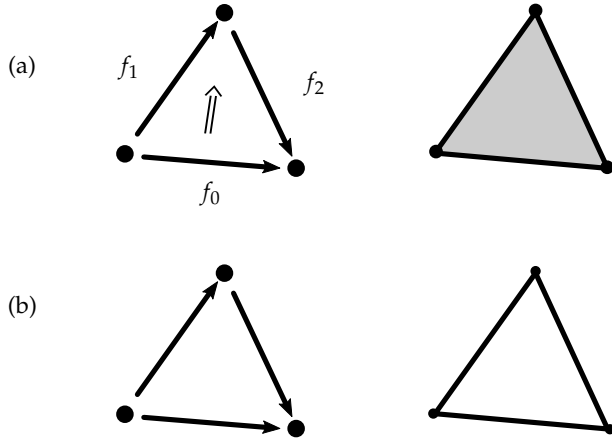


Figure 4: A small, poset-enriched category.

Figure 5: (a) An LP category and its classifying space; we add a 2-simplex spanned by all three objects because we have the relation $f_0 \Rightarrow f_2 \circ f_1$. (b) The effect of poset-enrichment: without the order relation we get a different classifying space.

2 Group Actions

In Section 3 we will be interested in pseudofunctors with codomain a 2-categorical extension of the category of groups, where the higher morphisms arise naturally from conjugation. Recall that a 2-category has morphisms between morphisms, i.e. a 2-category is a category enriched in small categories. See (nLab authors, 2022) for a full definition.

The 2-category **Grp** is defined as follows:

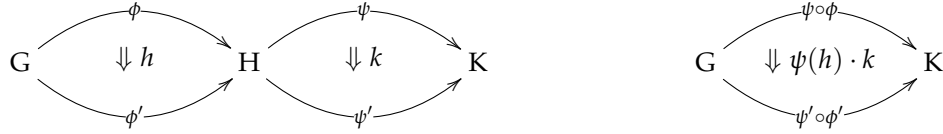
1. its objects are all groups;

2. the 1-morphisms $\mathbf{Grp}(G, H)$ are group homomorphisms $G \rightarrow H$; and,
3. the 2-morphisms $\phi \Rightarrow \phi'$ in $\mathbf{Grp}(G, H)$ consist of all group elements h in H that satisfy $\phi(g) = h \cdot \phi'(g) \cdot h^{-1}$ for all g in G .

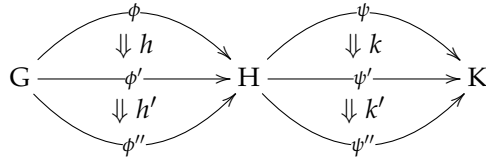
The 1-morphisms in $\mathbf{Grp}$ are composed in the usual manner for group homomorphisms. On the other hand, the 2-morphisms in $\mathbf{Grp}$ are often called *twisting elements*, and can be composed in two different ways. The *vertical* composition sends each pair (h, h') below to their product $h \cdot h'$ in H :



We indicate this type of composition (contravariantly) by $h' * h = h \cdot h'$ and note that h^{-1} serves as an inverse for h . Thus, for each pair of objects G, H the set $\mathbf{Grp}(G, H)$ forms a 1-category whose morphisms are all invertible (i.e., a groupoid). Conversely, the *horizontal* composite sends the pair (h, k) below to the product $\psi(h) \cdot k$ in K :



This composition is denoted by $k \circ h = \psi(k) \cdot h$. For any configuration of objects, 1 and 2-morphisms in $\mathbf{Grp}$ of the form the *interchange*



law holds between the horizontal and vertical compositions; in other words, we have

$$(k' * k) \circ (h' * h) = (k' \circ h') * (k \circ h),$$

which confirms that $\mathbf{Grp}$ is a 2-category.

The *automorphism group* $\text{Aut}(\mathcal{C})$ of an LP-category $\mathcal{C}$ consists of all invertible LP-functors $F : \mathcal{C} \rightarrow \mathcal{C}$. The group structure on $\text{Aut}(\mathcal{C})$ is obtained by the usual composition of such functors.

Definition 2.1. An **action** of a group G on the LP-category $\mathcal{C}$ is a group homomorphism $\alpha : G \rightarrow \text{Aut}(\mathcal{C})$; in particular, each group element g is sent by α to an invertible functor $\alpha_g : \mathcal{C} \rightarrow \mathcal{C}$. Moreover, for every $g \in G$ and non-identity morphism $f : x \rightarrow y$ in $\mathcal{C}$, we require:

1. $\alpha_g(x) \neq y \ \forall g \in G$, and
2. if $\alpha_g(x) = x$ then $\alpha_g(f) = f$.

Fix a group G , an LP-category $\mathcal{C}$, and an action $\alpha : G \rightarrow \text{Aut}(\mathcal{C})$ as described above. Denote images of the form $\alpha_g(x)$ and $\alpha_g(f)$ by gx and gf respectively. For each object x and morphism $f : x \rightarrow y$ in $\mathcal{C}$ we have the associated *object-orbits* and *morphism-orbits* given by

$$Gx := \{gx \mid g \in G\} \quad \text{and} \quad Gf := \{gf \mid g \in G\};$$

similarly, we have the corresponding *stabilisers*

$$G_x := \{g \in G \mid gx = x\} \quad \text{and} \quad G_f := \{g \in G \mid gf = f\},$$

both of which are subgroups of G .

Example 2.2. Fix a simplicial complex X and a group G which acts on X by invertible simplicial self-maps. Such an action is called *regular* (see (Bredon, 1972, Chapter III.1)) if two conditions hold for every simplex x — writing $(v_0, \dots, v_d)$ for the vertices of x ,

1. if g is in the stabiliser G_x then g fixes each v_i ; and
2. if $(g_0v_0, \dots, g_dv_d)$ is a simplex in X then there exists a group element g such that $g \cdot x = (g_0v_0, \dots, g_dv_d)$.

Any G -action on X can be made regular by passing to the second barycentric subdivision of X (Bredon, 1972, Chapter III.1 Prop. 1.1). In particular, if we have a regular simplicial action on X , then the induced action on $\mathbf{Fac}[X]$ satisfies both requirements of Definition 2.1.

Henceforth, we will talk about actions on a simplicial complex X and the induced action on $\mathbf{Fac}[X]$ interchangeably (see Figure 6). The motivation for regularity is to guarantee a simplicial quotient, which fails in very simple examples if the conditions are not satisfied. For example, consider C_3 acting on the 2-simplex by rotation, which fails to meet condition (1) of regularity. The orbit space is not simplicial, because it consists of one cell in each dimension, one for each orbit.

The requirements in Definition (2.1) are necessary to obtain a quotient category that is also loopless. We will formally define the quotient shortly, but for now we can assume that objects in the quotient are orbit classes of objects in $\mathcal{C}$, and similarly for morphisms. If

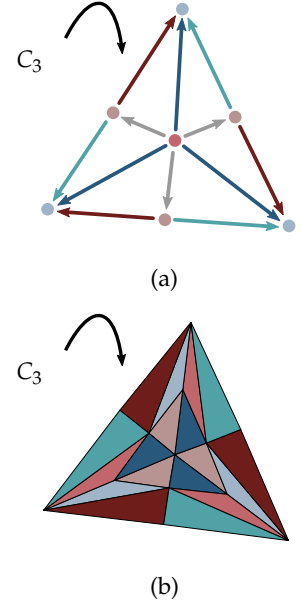


Figure 6: Actions that do and do not satisfy the conditions in Definition 2.1: the colours correspond to orbits. (a) The action of C_3 on the face poset of the 2-simplex is invalid because condition (1) is violated. (b) If we pass to the second barycentric subdivision, the action is valid.

we remove requirement (1), then we allow inversions of morphisms, which generates loops in the quotient. Consider a category with exactly two objects and two morphism $f : x \rightarrow y$ and $g : y \rightarrow x$ (see Figure 7). If C_2 acts by permuting the two objects, then the quotient has one orbit class $x \sim y$ and one non-trivial morphism class from this object class to itself.

When we introduce complexes of groups (in particular, the one summarising a group action), we will require that for any $f : x \rightarrow y$, $G_x \leq G_y$. Condition (2) of Definition 2.1 guarantees this, because if $g \cdot x = x$, then by (2), $g \cdot f = f$, and so $g \cdot y = y$ (see Figure 8). This condition also allows for a well-defined partial order on the quotient, shown explicitly in the proof of Proposition 2.3.

Given g in G and $f : x \rightarrow y$ in $\mathcal{C}$, the source and target of gf lie in Gx and Gy respectively. For any pair of $\mathcal{C}$ -morphisms $f : x \rightarrow y$ and $f' : y \rightarrow z$, the composite of any $f_1 \in Gf$ with any $f_2 \in Gf'$ must lie in the orbit $G(f' \circ f)$, by condition 2 of a group action. Thus, we have a well-defined (loopfree) 1-category $\mathcal{Q}$ whose objects are $\{Gx \mid x \text{ is an object of } \mathcal{C}\}$, and similarly for morphisms. In order to define a suitable quotient for the given G -action on $\mathcal{C}$, it remains to define a partial order on the morphism-orbits.

Proposition 2.3. *Consider the binary relation on the set of all morphisms $Gx \rightarrow Gy$ in $\mathcal{Q}$ prescribed by the following rule: $Gf \Rightarrow Gf'$ whenever there exists some $g \in G$ which satisfies $gf \Rightarrow f'$ in $\mathcal{C}$. This binary relation forms a partial order on the set $\mathcal{Q}(Gx, Gy)$.*

Proof. We seek to establish reflexivity, anti-symmetry and transitivity. To this end, note first that $Gf \Rightarrow Gf$ always holds in $\mathcal{Q}(Gx, Gy)$ because $f = \text{id}_G f \Rightarrow f$. Next, if we have $Gf \Rightarrow Gf' \Rightarrow Gf$, then there exist some g_0 and g_1 in G satisfying $g_0 f \Rightarrow f'$ and $g_1 f' \Rightarrow f$, and so

$$g_1 g_0 f \Rightarrow g_1 f' \Rightarrow f.$$

Now, $g_1 g_0 f \Rightarrow f$, so $g_1 g_0$ fixes the source, hence $g_1 g_0$ fixes f by requirement (2) of Definition 2.1. So in fact $f' = g_1^{-1} f$ lies in the orbit Gf and we have $Gf = Gf'$. Finally, assume that $Gf \Rightarrow Gf'$ and $Gf' \Rightarrow Gf''$ both hold in $\mathcal{Q}(Gx, Gy)$, so $gf \Rightarrow f'$ and $g'f' \Rightarrow f''$ hold for some g and g' in G . Then, we immediately have $g'gf \Rightarrow f''$, and hence $Gf \Rightarrow Gf''$ as desired. $\square$

We have now arrived at the desired quotient category.

Definition 2.4. The **quotient** LP-category $\mathcal{Q} = \mathcal{C} / G$ corresponding to the action of G on $\mathcal{C}$ is defined by the following data:

1. its objects are G -orbits of $\mathcal{C}$'s objects,

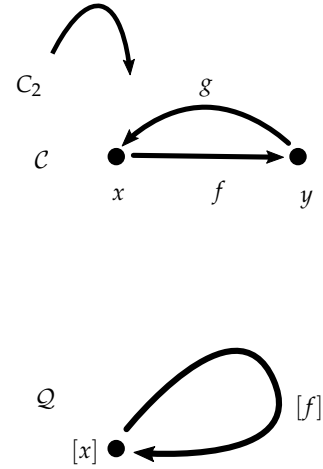


Figure 7: Removing condition (1) from Definition 2.1 allows for loops in the quotient.

2. its morphisms $Gx \rightarrow Gx'$ are G -orbits of $\mathcal{C}$'s morphisms $x \rightarrow x'$, partially ordered by $Gf \Rightarrow Gf'$ if and only if $gf \Rightarrow f'$ holds in $\mathcal{C}$ for some g in G .

The composite of $Gf_1 : Gx_0 \rightarrow Gx_1$ and $Gf_2 : Gx_1 \rightarrow Gx_2$ equals $G(f_2 \circ f_1)$.

There is a surjective **orbit functor** $\mathbf{P}_G : \mathcal{C} \twoheadrightarrow \mathcal{Q}$ (of LP-categories) that sends each x to Gx and each $f : x \rightarrow y$ to $Gf : Gx \rightarrow Gy$.

3 Complexes of Groups

COMPLEXES OF GROUPS are a generalisation of well-studied graphs of groups ², and can be viewed as the discrete analogue to orbifolds. They are our object of study because they can be used to summarise the data of a simplicial group action.

² Bass 1993 Bass, Hyman: Covering theory for graphs of groups. In: *Journal of Pure and Applied Algebra* 89 (1993), Nr. 1, S. 3–47. – ISSN 0022–4049

3.1 The Complex of Groups Associated to an Action

We generalise this existing construction, due to Bridson and Haefliger ³, to LP categories.

Definition 3.1 (Complex of groups). A **complex of groups** $\mathbf{F}$ over an LP-category $\mathcal{C}$ assigns:

1. to each object x of $\mathcal{C}$ a group $\mathbf{F}x$,
2. to each pair of objects x, y in $\mathcal{C}$ a functor $\mathbf{F} = \mathbf{F}_{xy}$ from the poset $\mathcal{C}(x, y)$ to the groupoid $\mathbf{Grp}(\mathbf{F}x, \mathbf{F}y)$ with $\mathbf{F}f : \mathbf{F}x \rightarrow \mathbf{F}y$ injective for all $f : x \rightarrow y$ in $\mathcal{C}$; and
3. to every pair $(f, f') \in \mathcal{C}(x, y) \times \mathcal{C}(y, z)$ of composable morphisms in $\mathcal{C}$, a twisting element $\gamma(f, f') = \gamma_{\mathbf{F}}(f, f')$ of the form $\mathbf{F}f' \circ \mathbf{F}f \Rightarrow \mathbf{F}(f' \circ f)$. More precisely, $\gamma(f, f')$ is a group element in $\mathbf{F}z$ such that

$$\mathbf{F}f' \circ \mathbf{F}f(k) = \gamma(f, f') \cdot \mathbf{F}(f' \circ f)(k) \cdot \gamma(f, f')^{-1},$$

for all $k \in \mathbf{F}x$.

These assignments are subject to the following axioms: first, given any morphism $f : x \rightarrow y$ in $\mathcal{C}$, both the twisting elements $\gamma(\text{id}_y, f)$ and $\gamma(f, \text{id}_x)$ equal the identity natural transformation on the functor $\mathbf{F}f$. And second, across any triple of composable morphisms in $\mathcal{C}$,

$$x_0 \xrightarrow{f_1} x_1 \xrightarrow{f_2} x_2 \xrightarrow{f_3} x_3,$$

³ Bridson and Häfliger BRIDSON, M.R. ; HÄFLIGER, A.: *Metric Spaces of Non-Positive Curvature*. Springer Berlin Heidelberg (Grundlehren der mathematischen Wissenschaften). – ISBN 9783540643241

the associated twisting elements satisfy the following *cocycle condition*:

$$\gamma(f_2, f_3) \cdot \gamma(f_1, f_3 \circ f_2) = \mathbf{F}f_3(\gamma(f_1, f_2)) \cdot \gamma(f_2 \circ f_1, f_3)$$

in the group $\mathbf{F}x_3$.

Equivalently, $\mathbf{F} : \mathcal{C} \rightarrow \mathbf{Grp}$ is a pseudofunctor that sends every morphism $f : x \rightarrow y$ in $\mathcal{C}$ to an injective group homomorphism $\mathbf{F}f : \mathbf{F}x \hookrightarrow \mathbf{F}y$ in $\mathbf{Grp}$. The cocycle condition described above is designed to ensure that the following associativity diagram commutes in the groupoid $\mathbf{Grp}(\mathbf{F}x_0, \mathbf{F}x_3)$:

$$\begin{array}{ccc} \mathbf{F}f_3 \circ \mathbf{F}f_2 \circ \mathbf{F}f_1 & \xRightarrow{\text{id}_{\mathbf{F}f_3} \circ \gamma(f_1, f_2)} & \mathbf{F}f_3 \circ \mathbf{F}(f_2 \circ f_1) \\ \Downarrow \gamma(f_2, f_3) \circ \text{id}_{\mathbf{F}f_1} & & \Downarrow \gamma(f_2 \circ f_1, f_3) \\ \mathbf{F}(f_3 \circ f_2) \circ \mathbf{F}f_1 & \xRightarrow{\gamma(f_1, f_3 \circ f_2)} & \mathbf{F}(f_3 \circ f_2 \circ f_1) \end{array}$$

Example 3.2. The simplest example of a complex of groups over an LP category $\mathcal{C}$ is the constant one, denoted $\underline{G}_{\mathcal{C}}$. It assigns the same group G to every object in $\mathcal{C}$, the identity homomorphism to every morphism in $\mathcal{C}$ and all twisting elements are trivial.

Definition 3.3 (Morphism of complexes of groups). A **morphism** $\Phi : \mathbf{F} \rightarrow \mathbf{F}'$ of complexes of groups over the LP-category $\mathcal{C}$ assigns

1. to each object x of $\mathcal{C}$ a 1-morphism $\Phi_x : \mathbf{F}x \rightarrow \mathbf{F}'x$ in $\mathbf{Grp}$, and
2. to each 1-morphism $f : x \rightarrow y$ of $\mathcal{C}$ a twisting element $\gamma(f) = \gamma_{\Phi}(f)$ in $\mathbf{Grp}$ of the form $\Phi_y \circ \mathbf{F}f \Rightarrow \mathbf{F}'f \circ \Phi_x$:

$$\begin{array}{ccc} \mathbf{F}x & \xrightarrow{\Phi_x} & \mathbf{F}'x \\ \mathbf{F}f \downarrow & \gamma(f) \nearrow & \downarrow \mathbf{F}'f \\ \mathbf{F}y & \xrightarrow{\Phi_y} & \mathbf{F}'y \end{array}$$

so that the following two axioms hold:

1. the *identity* axiom requires $\gamma(\text{id}_x)$ to be the identity $1_{\mathbf{F}'x} : \Phi_x \Rightarrow \Phi_x$ for each object x of $\mathcal{C}$, and
2. for each pair $x \xrightarrow{f_0} y \xrightarrow{f_1} z$ of composable 1-morphisms of $\mathcal{C}$, the *coherence* axiom imposes a relation

$$\Phi_z(\gamma_{\mathbf{F}}(f_0, f_1)) \cdot \gamma_{\Phi}(f_1 \circ f_0) = \gamma_{\Phi}(f_1) \cdot \mathbf{F}'f_1(\gamma_{\Phi}(f_0)) \cdot \gamma_{\mathbf{F}'}(f_0, f_1) \quad (1)$$

in the group $\mathbf{F}'z$ among the associated twisting elements.

The coherence axiom described above enforces the commutativity of the following pentagon in the groupoid $\mathbf{Grp}(\mathbf{F}x, \mathbf{F}'z)$:

$$\begin{array}{ccc}
 & \Phi_z \circ \mathbf{F}f_1 \circ \mathbf{F}f_0 & \xrightarrow{\gamma_\Phi(f_1) \circ \text{id}_{\mathbf{F}f_0}} \mathbf{F}'f_1 \circ \Phi_y \circ \mathbf{F}f_0 \\
 \text{id}_{\Phi_z} \circ \gamma_{\mathbf{F}}(f_0, f_1) \swarrow & & \searrow \text{id}_{\mathbf{F}'f_1} \circ \gamma_\Phi(f_0) \\
 \Phi_z \circ \mathbf{F}(f_1 \circ f_0) & & \mathbf{F}'f_1 \circ \mathbf{F}'f_0 \circ \Phi_x \\
 \searrow \gamma_\Phi(f_1 \circ f_0) & & \swarrow \gamma_{\mathbf{F}'}(f_0, f_1) \circ \text{id}_{\Phi_x} \\
 & \mathbf{F}'(f_1 \circ f_0) \circ \Phi_x &
 \end{array}$$

We call $\Phi : \mathbf{F} \rightarrow \mathbf{F}'$ an epi/mono/iso-morphism of complexes of groups if each Φ_x is an epi/mono/iso-morphism of groups.

We can construct a complex of groups associated to an action. The remarkable thing here is that this single object, plus an embedding of it into the constant complex of groups, provides all of the data necessary to build back a category that is isomorphic to the one with which we started ⁴. We will see this process in detail in the next section.

⁴ This result is called the *Basic Construction*.

In order to construct the aforementioned complex of groups, we must first introduce *lifts* and *transfers*.

Definition 3.4. A **lift function** for the G -action on $\mathcal{C}$ is an assignment $\lambda : \mathcal{Q}_{\text{ob}} \rightarrow \mathcal{C}_{\text{ob}}$ of an object $\lambda(y)$ in $\mathcal{C}$ to each object y in the quotient $\mathcal{Q}$ for which $\mathbf{P}_G \circ \lambda(y) = y$ holds.

Note that, in contrast to the orbit functor $\mathbf{P}_G$, a lift λ need not extend to a functor in general. The difficulty here is that morphisms $y \rightarrow y'$ in $\mathcal{Q}$ may not lift to morphisms $\lambda(y) \rightarrow \lambda(y')$ in $\mathcal{C}$. In general, for each $h : y \rightarrow y'$ in $\mathcal{Q}$, we are only guaranteed the existence of a (not necessarily unique) group element $g \in G$ and a unique morphism $f : \lambda(y) \rightarrow g\lambda(y')$ in $\mathcal{C}$ so that $\mathbf{P}_G(f) = h$.

Definition 3.5. A choice of **transfer elements** for a given lift function $\lambda : \mathcal{Q}_{\text{ob}} \rightarrow \mathcal{C}_{\text{ob}}$ assigns to each morphism $h : y \rightarrow y'$ in $\mathcal{Q}$ an element $\sigma := \sigma(h)$ in G so that there is a unique morphism $f : \lambda(y) \rightarrow \sigma^{-1}\lambda(y')$ in $\mathcal{C}$ satisfying $\mathbf{P}_G(f) = h$. We implicitly require the transfer element $\sigma(\text{id}_y)$ assigned to each identity morphism in $\mathcal{Q}$ to equal the identity element 1_G in G .

Given a choice of lift function and transfer elements for the action of a group G on an LP-category $\mathcal{C}$, there is a standard way to construct a pair $(\mathbf{F}, \Phi)$ where $\mathbf{F}$ is a complex of groups on the quotient $\mathcal{Q} = \mathcal{C} / G$ and $\Phi : \mathbf{F} \hookrightarrow \underline{G}_{\mathcal{Q}}$ is an injective morphism on the constant G -valued complex of groups on $\mathcal{Q}$.

Definition 3.6. Let G be a group acting on an LP-category $\mathcal{C}$ with quotient $\mathcal{Q}$; let $\lambda : \mathcal{Q}_{\text{ob}} \rightarrow \mathcal{C}_{\text{ob}}$ be an associated lift function and $\{\sigma(h) \mid h : y \rightarrow y' \text{ in } \mathcal{Q}\}$ a choice of transfer elements for λ . The **complex of groups $\mathbf{F}$ associated to these choices** is obtained by assigning

1. to each object y the stabiliser $\mathbf{F}y := G_{\lambda(y)}$ of its lift;
2. to each morphism $h : y \rightarrow y'$ the group homomorphism $\mathbf{F}h : \mathbf{F}y \rightarrow \mathbf{F}y'$ given by conjugation with $\sigma(h)$, i.e. $\mathbf{F}h(k) = \sigma(h)k\sigma(h)^{-1}$ for all $k \in \mathbf{F}y$;
3. to each order relation $h \Rightarrow h'$ the group element $\sigma(h \Rightarrow h') := \sigma(h) \cdot \sigma(h')^{-1}$; and finally,
4. to each pair $h_1 : y_1 \rightarrow y_2$ and $h_2 : y_2 \rightarrow y_3$ of composable morphisms, the twisting element

$$\gamma_{\mathbf{F}}(h_1, h_2) := \sigma(h_2) \cdot \sigma(h_1) \cdot \sigma(h_2 \circ h_1)^{-1}.$$

Similarly, the **associated monomorphism $\Phi : \mathbf{F} \hookrightarrow \underline{G}_{\mathcal{Q}}$** is given by assigning

1. to each y the inclusion $G_{\lambda(y)} \hookrightarrow G$, and
2. to each $h : y \rightarrow y'$ the twisting element $\gamma_{\Phi}(h) := \sigma(h)$.

We will sometimes refer to the associated monomorphism as the *associated embedding* into the constant complex of groups.

It has been shown in (Bridson and Häfliger, Chapter III.C.2.9) that $\mathbf{F}$ and Φ as described above satisfy the requirements of Definitions 3.1 and 3.3 respectively when defined over small loopfree categories; moreover, making a different choice of lift function and transfer elements produces an isomorphic pair $(\mathbf{F}, \Phi)$. Moreover, we have the additional data of a partial order on each hom-set, so we check that $\mathbf{F}$ is well-defined.

Proposition 3.7. *The construction above yields a well-defined complex of groups and associated morphism over an LP category.*

The proof follows from Lemmas 3.8, 3.9 and 3.10.

Lemma 3.8. *The assignment $\mathbf{F}_{yy'}$ is a functor $\mathcal{Q}(y, y') \rightarrow \mathbf{Grp}(\mathbf{F}y, \mathbf{F}(y'))$ for every pair of objects (y, y') .*

Proof. For every $f : y \rightarrow y'$, the map $\mathbf{F}f$ is an injective group homomorphism because it is conjugation by the chosen transfer element. The homomorphism $\mathbf{F}f$ maps from $\mathbf{F}y$ to $\mathbf{F}y'$: take any k in $\mathbf{F}y$ and recall that the transfer $\sigma(f)$ satisfies that there is a unique morphism

$$\lambda(f) : \lambda(y) \rightarrow \sigma(f)^{-1}\lambda(y')$$

such that $\mathbf{P}(\lambda(f)) = f$. By Definition 2.1(2), the morphism

$$k\lambda(f) : k\lambda(y) \rightarrow k\sigma(f)^{-1}\lambda(y')$$

is equal to $\lambda(f)$, since $k\lambda(y) = \lambda(y)$. Hence,

$$k\sigma(f)^{-1}\lambda(y') = \sigma(f)^{-1}\lambda(y'),$$

which implies that $\sigma(f) \cdot k \cdot \sigma(f)^{-1} \in \mathbf{F}y'$, i.e., $\mathbf{F}f(k) \in \mathbf{F}y'$.

To show that the elements $\sigma := \sigma(f \Rightarrow f')$ are well-defined, we have to check that for any $k \in \mathbf{F}y$, we have $\mathbf{F}f(k) = \sigma \cdot \mathbf{F}f'(k) \cdot \sigma^{-1}$. We have,

$$\begin{aligned} \sigma \cdot \mathbf{F}f'(k) \cdot \sigma^{-1} &= \sigma(f)\sigma(f')^{-1} \cdot \left(\sigma(f') \cdot k \cdot \sigma(f')^{-1} \right) \cdot \sigma(f')\sigma(f)^{-1} \\ &= \sigma(f) \cdot k \cdot \sigma(f)^{-1}. \end{aligned}$$

Lastly, we check that $\mathbf{F}_{yy'}$ satisfies the identity and associativity laws. The identity law follows immediately from the fact that the transfer element $\sigma(\text{id}_y)$ is the identity element in G , so $\mathbf{F}_{yy'}(\text{id}_y) = \text{id}_{\mathbf{F}y'}$.

For the associativity law to hold, we need to check that for any three morphisms $f, f', f'' \in \mathcal{Q}(y, y')$ satisfying $f \Rightarrow f' \Rightarrow f''$, the identity $\sigma(f \Rightarrow f') \cdot \sigma(f' \Rightarrow f'') = \sigma(f \Rightarrow f'')$ holds:

$$\begin{aligned} \sigma(f \Rightarrow f') \cdot \sigma(f' \Rightarrow f'') &= \left(\sigma(f)\sigma(f')^{-1} \right) \left(\sigma(f')\sigma(f'')^{-1} \right) \\ &= \sigma(f)\sigma(f'')^{-1} \\ &= \sigma(f \Rightarrow f''), \end{aligned}$$

as desired. □

Lemma 3.9. *For any pair (f_0, f_1) of composable morphisms, the twisting element $\gamma_{\mathbf{F}}(f_0, f_1)$ is well-defined and satisfies the cocycle condition.*

Proof. For each pair of composable morphisms $y \xrightarrow{f_0} y' \xrightarrow{f_1} y''$, we need to check that the twisting element $\gamma(f_0, f_1)$ satisfies

$$\mathbf{F}f_1 \circ \mathbf{F}f_0(k) = \gamma(f_0, f_1) \cdot \mathbf{F}(f_1 \circ f_0)(k) \cdot \gamma(f_0, f_1)^{-1}$$

for any $k \in \mathbf{F}y$. The left-hand side yields

$$\sigma(f_1)\sigma(f_0) \cdot k \cdot \sigma(f_0)^{-1}\sigma(f_1)^{-1},$$

and the right-hand side yields

$$(\sigma(f_1)\sigma(f_0)) \cdot k \cdot (\sigma(f_1)\sigma(f_0))^{-1},$$

so the equation above holds.

The twisting elements also satisfy the cocycle condition: for any sequence of composable morphisms

$$y \xrightarrow{f_0} y' \xrightarrow{f_1} y'' \xrightarrow{f_2} y'''$$

both $\gamma(f_1, f_2)\gamma(f_0, f_2 \circ f_1)$ and $\mathbf{F}f_2(\gamma(f_0, f_1))\gamma(f_1 \circ f_0, f_2)$ are equal to

$$\sigma(f_2)\sigma(f_1)\sigma(f_0)\sigma(f_2 \circ f_1 \circ f_0)^{-1},$$

as required. $\square$

Lemma 3.10. *The morphism $\Phi : \mathbf{F} \rightarrow \underline{\mathbf{G}}_{\mathcal{Q}}$ is well-defined and satisfies the coherence axiom.*

Proof. For any $f : y \rightarrow y'$, it follows immediately from the fact that Φ_y and $\Phi_{y'}$ are inclusions into $\mathbf{G}$ and $\mathbf{F}f$ is conjugation by $\gamma_{\Phi} = \sigma(f)$, that $\gamma_{\Phi} : \Phi_{y'} \circ \mathbf{F}f \Rightarrow \Phi_y$.

The identity axiom is immediately satisfied because the transfer element associated to any identity morphism is the identity element in $\mathbf{G}$.

Lastly, we check the coherence axiom, for any sequence of composable morphisms $y \xrightarrow{f_0} y' \xrightarrow{f_1} y''$. The left-hand side of the coherence axiom (Equation (1)) yields:

$$\begin{aligned} \Phi_{y''}(\gamma_{\mathbf{F}}(f_0, f_1)) \cdot \gamma_{\Phi}(f_1 \circ f_0) &= \left(\sigma(f_1)\sigma(f_0)\sigma(f_1 \circ f_0)^{-1} \right) \cdot \sigma(f_1 \circ f_0) \\ &= \sigma(f_1)\sigma(f_0), \end{aligned}$$

and the right-hand side yields:

$$\begin{aligned} \gamma_{\Phi}(f_1) \cdot \underline{\mathbf{G}}f_1(\gamma_{\Phi}(f_0)) \cdot \gamma_{\underline{\mathbf{G}}}(f_1, f_2) &= \sigma(f_1) \cdot \text{id}(\sigma(f_0)) \cdot 1_{\mathbf{G}} \\ &= \sigma(f_1)\sigma(f_0), \end{aligned}$$

as required. $\square$

3.2 Development and the Basic Construction

In this section we prove a mild generalisation of Theorem 2.13 from (Bridson and Häfliger, Chapter III.C) to the poset-enriched setting. Throughout, we fix an LP-category $\mathcal{Q}$ and a group $\mathbf{G}$.

Proposition 3.11. *Let $\mathbf{F} : \mathcal{Q} \rightarrow \mathbf{Grp}$ be a complex of groups and $\Phi : \mathbf{F} \rightarrow \underline{\mathbf{G}}_{\mathcal{Q}}$ a monomorphism to the constant $\mathbf{G}$ -valued complex of groups over $\mathcal{Q}$. Associated to this pair $(\mathbf{F}, \Phi)$ is a new LP-category $\mathcal{D} = \mathcal{D}(\mathbf{F}, \Phi)$ equipped with a natural $\mathbf{G}$ -action so that*

1. *the quotient $\mathcal{D} / \mathbf{G}$ is isomorphic to $\mathcal{Q}$; moreover,*
2. *there exist choices of lift function and transfer elements for the $\mathbf{G}$ -action on $\mathcal{D}$ which produce the pair $(\mathbf{F}, \Phi)$; and finally,*

3. if $\mathbf{F}$ and Φ are the complex of groups and monomorphism associated to a G -action on an LP-category $\mathcal{C}$ (subject to choices of lifts and transfers), then there exists a G -equivariant isomorphism of categories between $\mathcal{C}$ and $\mathcal{D}$.

The LP-category $\mathcal{D}$ whose existence and uniqueness (up to equivariant isomorphism) are guaranteed by the above result is called the **development** of the given pair $(\mathbf{F}, \Phi)$.

For each object x of $\mathcal{Q}$ and each group element g in G , let

$$\mathbf{c}_x : G \rightarrow G / \Phi_x(\mathbf{F}x)$$

be the map (of sets) which assigns to each group element g the corresponding left coset $g \cdot \Phi_x(\mathbf{F}x)$. The development $\mathcal{D}$ is defined as follows:

1. its objects are pairs $(x, \mathbf{c}_x(g))$ where x is an object of $\mathcal{Q}$ and g is an element of G ;
2. the morphisms $(x, \mathbf{c}_x(g)) \rightarrow (y, \mathbf{c}_y(h))$ are all pairs $(f, \mathbf{c}_x(g))$ where $f : x \rightarrow y$ is a morphism in $\mathcal{Q}$ and the following cosets coincide:

$$\mathbf{c}_y(h) = \mathbf{c}_y(g \cdot \gamma_\Phi(f)^{-1}).$$

Two such morphisms satisfy the partial order relation $(f, \mathbf{c}_x(g)) \Rightarrow (f', \mathbf{c}_x(g))$ within $\mathcal{D}$ whenever $f \Rightarrow f'$ holds in $\mathcal{Q}$, and the element $\gamma_\Phi(f) \cdot \gamma_\Phi(f')^{-1}$ is in $\Phi_y(\mathbf{F}y)$. This condition ensures that the targets of the morphisms are the same, in order for them to be comparable in $\mathcal{D}$. The composite

$$(x, \mathbf{c}_x(g_1)) \xrightarrow{(f_1, \mathbf{c}_x(g_1))} (y, \mathbf{c}_y(g_2)) \xrightarrow{(f_2, \mathbf{c}_y(g_2))} (z, \mathbf{c}_z(g_3)) \quad (2)$$

is given by $(f_2 \circ f_1, \mathbf{c}_x(g_1))$, and the group G acts on $\mathcal{D}$ as follows: the element $h \in G$ sends each object $(x, \mathbf{c}_x(g))$ to $(x, \mathbf{c}_x(h \cdot g))$, and similarly for morphisms.

We first check that the morphisms $(f, \mathbf{c}_x(g))$ of $\mathcal{D}$ are well-defined, i.e., independent of the choice of representative element g of the coset.

Proposition 3.12. *Let $f : x \rightarrow y$ be a morphism in $\mathcal{Q}$. If $\mathbf{c}_x(g) = \mathbf{c}_x(h)$ holds for g, h in G , then we also have*

$$\mathbf{c}_y(g \cdot \gamma_\Phi(f)^{-1}) = \mathbf{c}_y(h \cdot \gamma_\Phi(f)^{-1}).$$

Proof. Let k be any element of the group $\mathbf{F}x$; since $\gamma_\Phi(f)$ is a twisting element of the form $\Phi_y \circ \mathbf{F}f \Rightarrow \Phi_x$ by Definition 3.3, we have

$$\gamma_\Phi(f) \cdot \Phi_x(k) \cdot \gamma_\Phi(f)^{-1} = \Phi_y \circ \mathbf{F}f(k).$$

Since $\mathbf{F}f$ takes values in $\mathbf{F}y$, the cosets $\mathbf{c}_y(\Phi_x(k) \cdot \gamma_\Phi(f)^{-1})$ and $\mathbf{c}_y(\gamma_\Phi(f)^{-1})$ are equal regardless of k , so the desired conclusion follows. $\square$

Next, we show that the composition of $(f_1, \mathbf{c}_x(g_1))$ and $(f_2, \mathbf{c}_y(g_2))$, as described in (2), produces a morphism $(x, \mathbf{c}_x(g_1)) \rightarrow (z, \mathbf{c}_z(g_3))$ in $\mathcal{D}$.

Proposition 3.13. *If $\mathbf{c}_y(g_2) = \mathbf{c}_y(g_1 \cdot \gamma_\Phi(f_1)^{-1})$ and $\mathbf{c}_z(g_3) = \mathbf{c}_z(g_2 \cdot \gamma_\Phi(f_2)^{-1})$ hold, then we also have $\mathbf{c}_z(g_3) = \mathbf{c}_z(g_1 \cdot \gamma_\Phi(f_2 \circ f_1)^{-1})$.*

Proof. Since the cosets $\mathbf{c}_y(g_2)$ and $\mathbf{c}_y(g_1 \cdot \gamma_\Phi(f_1)^{-1})$ are equal, there exists some $h \in \mathbf{F}y$ satisfying $\gamma_\Phi(f_1) \cdot g_1^{-1} \cdot g_2 = \Phi_y(h)$, and so

$$g_2 = g_1 \cdot \gamma_\Phi(f_1)^{-1} \cdot \Phi_y(h).$$

Using this expression for g_2 along with our assumption that $\mathbf{c}_z(g_3) = \mathbf{c}_z(g_2 \cdot \gamma_\Phi(f_2)^{-1})$, we obtain

$$\begin{aligned} \mathbf{c}_z(g_3) &= \mathbf{c}_z\left(g_1 \cdot \gamma_\Phi(f_1)^{-1} \cdot \Phi_y(h) \cdot \gamma_\Phi(f_2)^{-1}\right) \\ &= \mathbf{c}_z\left(g_1 \cdot \gamma_\Phi(f_1)^{-1} \cdot \gamma_\Phi(f_2)^{-1}\right) && \text{by Prop 3.12} \\ &= \mathbf{c}_z\left(g_1 \cdot \gamma_\Phi(f_2 \circ f_1)^{-1}\right), \end{aligned}$$

as desired. $\square$

It is not difficult to confirm that $\mathcal{Q}$ is the quotient $\mathcal{D} / G$. To obtain $(\mathbf{F}, \Phi)$ for the G -action on $\mathcal{D}$ as per Definition 3.6, one uses the lift function $x \mapsto (x, \mathbf{c}_y(1_G))$ and the transfer elements $f \mapsto \gamma_\Phi(f)$ for each object x and morphism f of $\mathcal{Q}$. Thus, it only remains to establish the universal property from assertion (3) of Proposition 3.11.

Proposition 3.14. *Let $\mathcal{C}$ be any LP-category upon which G acts with quotient $\mathcal{C} / G$ equal to $\mathcal{Q}$. If there exists a choice of lifts and transfers for this action which produces $(\mathbf{F}, \Phi)$ as the associated complex of groups and monomorphism to $\underline{G}_\mathcal{Q}$, then there exists a G -equivariant isomorphism of LP-categories $\mathcal{D}(\mathbf{F}, \Phi) \cong_G \mathcal{C}$.*

Proof. Let $\lambda : \mathcal{Q}_{\text{ob}} \rightarrow \mathcal{C}_{\text{ob}}$ and $\{\sigma(q) \mid q : x \rightarrow y \text{ in } \mathcal{Q}\}$ be any choice of lift function and associated transfer elements for the G -action on $\mathcal{C}$ that produce $\mathbf{F}$ and Φ . We define a functor $\mathbf{H} : \mathcal{D}(\mathbf{F}, \Phi) \rightarrow \mathcal{C}$ as follows: $\mathbf{H}$ sends each object $(x, \mathbf{c}_x(g))$ in $\mathcal{D}$ to $g\lambda(x)$ in $\mathcal{C}$. The surjectivity of $\mathbf{H}$ on objects is clear; injectivity follows easily from the fact that $g\lambda(x) = h\lambda(x)$ if and only if $h^{-1}g \in \Phi_x(\mathbf{F}x)$, in which case $\mathbf{c}_x(g) = \mathbf{c}_x(h)$.

Given any morphism $f : x \rightarrow y$ in $\mathcal{Q}$, the transfer element $\gamma_\Phi(f)$ satisfies that

$$\tilde{f} : \lambda(x) \rightarrow \gamma_\Phi(f)^{-1}\lambda(y)$$

is the unique morphism in $\mathcal{C}$ with source $\lambda(x)$ such that $\mathbf{P}(\tilde{f}) = f$. We set the $\mathbf{H}$ -image of the morphism $(f, \mathbf{c}_x(g))$ to be $g\tilde{f}$.

We first check that $\mathbf{H}$ is order-preserving, i.e. that

$$(f, \mathbf{c}_x(g)) \Rightarrow (f', \mathbf{c}_x(g)) \text{ if and only if } g\tilde{f} \Rightarrow g\tilde{f}'.$$

If $(f, \mathbf{c}_x(g)) \Rightarrow (f', \mathbf{c}_x(g))$ in $\mathcal{D}$, then $f \Rightarrow f'$ in $\mathcal{Q}$ and $\gamma_\Phi(f) \cdot \gamma_\Phi(f')^{-1} \in \Phi_y(\mathbf{F}y)$. By definition of the quotient, $f \Rightarrow f'$ in $\mathcal{Q}$ if and only if there exists $k \in G$ such that $k\tilde{f} \Rightarrow \tilde{f}'$. But $\tilde{f}$ and $\tilde{f}'$ have the same source and target, and there is a unique morphism in the orbit of $\tilde{f}$ with this source, hence, $k = \text{id}_G$ and $g\tilde{f} \Rightarrow g\tilde{f}'$. The other direction follows from the fact that, if $g\tilde{f} \Rightarrow g\tilde{f}'$ in $\mathcal{C}$, then they must have the same target and so $\gamma_\Phi(f) \cdot \gamma_\Phi(f')^{-1} \in \Phi_y(\mathbf{F}y)$.

The fact that $\mathbf{H}$ is bijective on hom-sets follows from an argument similar to the one above, and some simple computations. The fact that $\mathbf{H}$ is G -equivariant follows almost directly from the definition. $\square$

We introduce a technical lemma relating the development of complexes of groups over isomorphic LP categories, that will be used in Section 6.

Lemma 3.15. *Suppose that $\mathbf{E} : X \rightarrow X'$ is an isomorphism of LP categories, and that $\mathbf{F} : X \rightarrow \mathbf{Grp}$ and $\mathbf{F}' : X' \rightarrow \mathbf{Grp}$ are complexes of group such that the following diagram commutes:*

$$\begin{array}{ccc} X & \xrightarrow{\mathbf{E}} & X' \\ \mathbf{F} \downarrow & \searrow \mathbf{F}' & \\ \mathbf{Grp} & & \end{array}$$

If $\Phi : \mathbf{F} \rightarrow \underline{G}_X$ and $\Phi' : \mathbf{F}' \rightarrow \underline{G}_{X'}$ are morphisms of complexes of groups such that $\Phi_f = \Phi'_{\mathbf{E}f}$ for every morphism f in X , then there is a G -equivariant isomorphism of LP categories

$$\mathcal{D}(\mathbf{F}, \Phi) \cong_G \mathcal{D}(\mathbf{F}', \Phi').$$

Proof. First note that by the commutativity of the diagram, the groups $\mathbf{F}x$ and $\mathbf{F}'\mathbf{E}x$ are equal for every $x \in X$; hence, the cosets $g \cdot \Phi(\mathbf{F}x)$ and $g \cdot \Phi'_{\mathbf{E}x}(\mathbf{F}'\mathbf{E}x)$ are equal for any group element g , and so the indices $[G : \Phi_x(\mathbf{F}x)]$ and $[G : \Phi'_{\mathbf{E}x}(\mathbf{F}'\mathbf{E}x)]$ are equal.

Each object $x \in X$ generates the set $\{(x, g \cdot \Phi_x(\mathbf{F}x)) \mid \forall g \in G\}$ of objects in $\mathcal{D}(\mathbf{F}, \Phi)$. This is clearly in bijection with the set of objects $\{(\mathbf{E}x, g \cdot \Phi'_{\mathbf{E}x}(\mathbf{F}'\mathbf{E}x)) \mid \forall g \in G\}$ in $\mathcal{D}(\mathbf{F}', \Phi')$, since $\mathbf{E}$ is an isomorphism.

The poset of morphisms

$$(x, g \cdot \Phi_x(\mathbf{F}x)) \rightarrow (y, g \cdot \Phi_f^{-1} \Phi_y(\mathbf{F}y))$$

in $\mathcal{D}(\mathbf{F}, \Phi)$ is isomorphic to the poset of morphisms

$$(\mathbf{E}x, g \cdot \Phi'_{\mathbf{E}x}(\mathbf{F}'\mathbf{E}x)) \rightarrow (\mathbf{E}y, g \cdot (\Phi'_{\mathbf{E}f})^{-1} \Phi'_{\mathbf{E}y}(\mathbf{F}'\mathbf{E}y))$$

in $\mathcal{D}(\mathbf{F}', \Phi')$ by the arguments above, in addition to the fact that we assume that $\Phi_f = \Phi'_{\mathbf{E}f}$.

The correspondence between the partial order on morphisms in $\mathcal{D}(\mathbf{F}, \Phi)$ and $\mathcal{D}(\mathbf{F}', \Phi')$ follows from the fact that $f \Rightarrow f'$ in X if and only if $\mathbf{E}f \Rightarrow \mathbf{E}f'$ in X' , and $\Phi_f = \Phi'_{\mathbf{E}f}$ for any morphism $f \in X$. $\square$

4 The Discrete Flow Category

The aim in this section is to define the *discrete flow category*; this category encodes all of the information of a discrete gradient vector field on a simplicial complex - the main tool extracted from discrete Morse theory. We use these tools ultimately to reduce the size of a simplicial complex, with a view to applying them on the quotient of an action.

Throughout this section, X will denote a simplicial complex. We write $x > x'$ for simplices x and x' if x' is a face of x .

4.1 Entrance Paths

We have already seen that for a simplicial complex (or more generally, a regular CW complex), the face poset $\mathbf{Fac}[X]$ summarises all of the combinatorial information about X , and its classifying space has the same homotopy type as X . In particular, recall that for any pair of simplices there is at most one morphism between them.

The *entrance path category* can be considered as an expansion of the face poset, with the same objects, recording all possible adjacency paths between simplex pairs⁵. More precisely, an *entrance path* from x to x' is a strictly descending sequence of faces from x to x' . Each f in the set of entrance paths $\mathbf{Ent}[X](x, x')$ has the form

$$f = (x > x_1 > \cdots > x_k > x')$$

for some length $k \geq 0$ where every x_i is a simplex of X . These entrance paths are partially ordered by refinement, so we have $f' \Rightarrow f$ whenever f' is obtained by removing some of the intermediate x_i from f . Note that the only entrance path from each simplex x to itself is the trivial path (x) of length zero.

Definition 4.1 (Entrance path category). The **entrance path category** of X is the LP-category $\mathbf{Ent}[X]$ prescribed by the following data:

⁵ It will become more apparent why we want to keep track of all such paths later on in the section; readers already familiar with discrete Morse theory might already notice the potential utility, given that a discrete gradient vector field on X does not guarantee a unique gradient flow between two critical cells in the reduced CW complex.

1. its objects are the simplices of X , and
2. the poset of morphisms $x \rightarrow x'$ is $\mathbf{Ent}[X](x, x')$.

The composite of $f := (x > x_0 > \dots > x_k > x')$ and $f' := (x' > x'_0 > \dots > x_\ell > x'')$ is the concatenated path

$$f' \circ f := (x > x_0 > \dots > x_k > x' > x'_0 > \dots > x_\ell > x''),$$

which lies in the poset $\mathbf{Ent}[X](x, x'')$.

Proposition 4.2. (*Nanda, 2019, Proposition 3.3*) *The classifying space of $\mathbf{Ent}[X]$ is homotopy equivalent to X .*

Figure 9 provides an example of the entrance path category of a simplicial complex, and illustrates why we require poset enrichment to obtain the right homotopy-type.

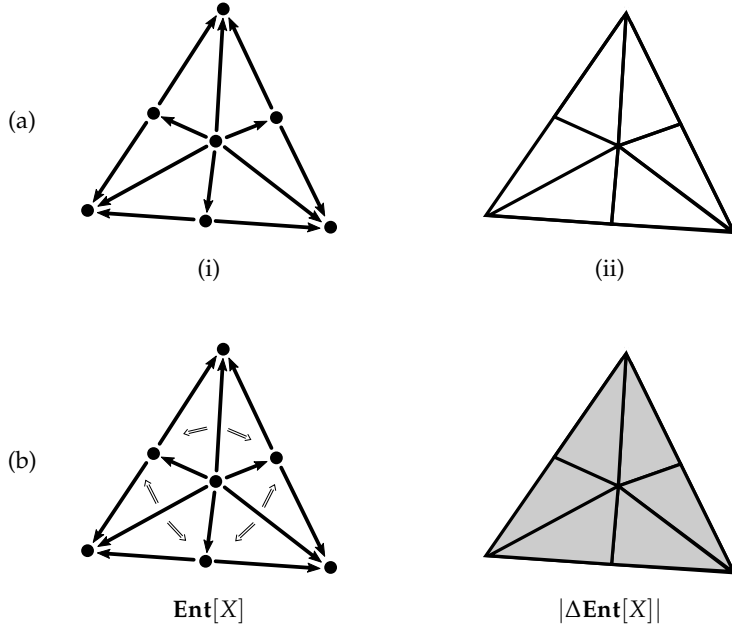


Figure 9: The 2-simplex and its entrance path category. **(a)** (i) illustrates the entrance path category of the 2-simplex, with no partial order. For any triple $x > y > z$ of simplices, there is not equality of morphisms between the composition $(y > z) \circ (x > y)$ and $(x > z)$, as there was in the face poset. Hence, the required 2-simplices do not exist in the geometric nerve (ii). **(b)** In the entrance path category, we have the relation $x > z \Rightarrow x > y > z$ in the hom-set $\mathbf{Ent}[X](x, z)$ yielding the desired 2-simplex spanned by the objects x, y, z .

4.2 Discrete Morse Theory

Motivated by the desire to simplify simplicial complexes, we can apply *elementary collapses* to remove simplices in a homology-preserving way. Recall that an elementary collapse is the removal of a pair of simplices $x > y$ where x is the only co-face of y , they differ in dimension by 1, and x is not the face of any other simplex. Figure 10 illustrates an example of such a collapse.

Of course, this ‘nice’ situation does not always occur, and discrete Morse theory provides a formal way to find ‘good’ pairs of simplices to remove.

The original theory was developed by Forman in 1998⁶, where the author presents a combinatorial analogue of smooth Morse theory, starting from the definition of a discrete Morse function (a real-valued function on the set of simplices with some conditions). For a comprehensive exposition of discrete Morse theory, see for example, (Forman, 1998; Knudson, 2015). For our purposes, we will jump directly to discrete gradient vector fields. In the literature, these are commonly called *acyclic partial matchings*.

Definition 4.3. A set $\Sigma = \{(x_\bullet > y_\bullet)\}$ of simplex-pairs in X is called an **acyclic partial matching** if it satisfies three axioms:

1. **dimension:** if $(x > y) \in \Sigma$, then $\dim x - \dim y = 1$;
2. **partition:** if $(x > y) \in \Sigma$, then neither x nor y lies in any other pair of Σ ; and,
3. **acyclicity:** the transitive closure of the binary relation

$$(x_0 > y_0) \blacktriangleright (x_1 > y_1) \text{ whenever } x_0 > y_1 \text{ in } X$$

generates a partial order on Σ .⁷

We can take a graph-theoretic approach to the definition of acyclic partial matchings by passing to the *poset graph* $\mathcal{P}(X)$ of a simplicial complex; this is a directed graph with a vertex for each simplex, and an edge from x to y if and only if y is a codimension one face of x . Recall that a matching in a graph is a subset of edges sharing no vertices. An acyclic partial matching on X is a matching in the poset graph such that, if the orientation of all edges in the matching is reversed, there are no directed cycles. We denote the modified poset graph by $\mathcal{P}'(X)$. This approach was first introduced by Chari⁸.

We can also think of Σ as a collection of (well-behaved) morphisms in $\mathbf{Ent}[X]$, rather than pairs of simplices; the utility of this perspective becomes clearer in Section 4.3.

The simplices of X which do not appear in any Σ -pair are called **critical**; the gradient paths from a critical simplex w to a critical simplex z are decorated by zigzags of $\blacktriangleright$ -descending Σ elements, e.g.,

$$\zeta := (w > y_0 < x_0 > y_1 < x_1 > \cdots > y_k < x_k > z).$$

Here each backward-pointing $y_i < x_i$ corresponds to an element $(x_i > y_i)$ in Σ , and the fact that y_{i+1} is a face of x_i ensures that $(x_i > y_i) \blacktriangleright (x_{i+1} > y_{i+1})$. An inductive argument establishes the following result:

Theorem 4.4 (Corollary 3.5 (Forman, 1998)). *If Σ is an acyclic partial matching on X , then X is homotopy equivalent to a CW complex whose cells are in bijective correspondence with the critical simplices of Σ .*

⁶ Forman 1998 FORMAN, Robin: Morse Theory for Cell Complexes. In: *Advances in Mathematics* 134 (1998), S. 90–145

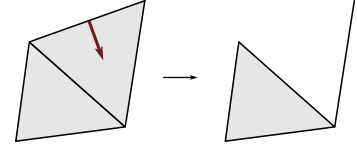


Figure 10: Applying an elementary collapse doesn't change the homotopy type. Trying to remove simplex pairs in the interior of the complex is more complicated, because we do not have details of how the remaining cells are attached to each other.

⁷ Some readers might wonder how these matchings relate to functions; see (Knudson, 2015, Theorem 6.19) for a bridge between the two viewpoints.

⁸ Chari 2000 CHARI, Manoj K.: On discrete Morse functions and combinatorial decompositions. In: *Discrete Mathematics* 217 (2000), Nr. 1-3, S. 101–113

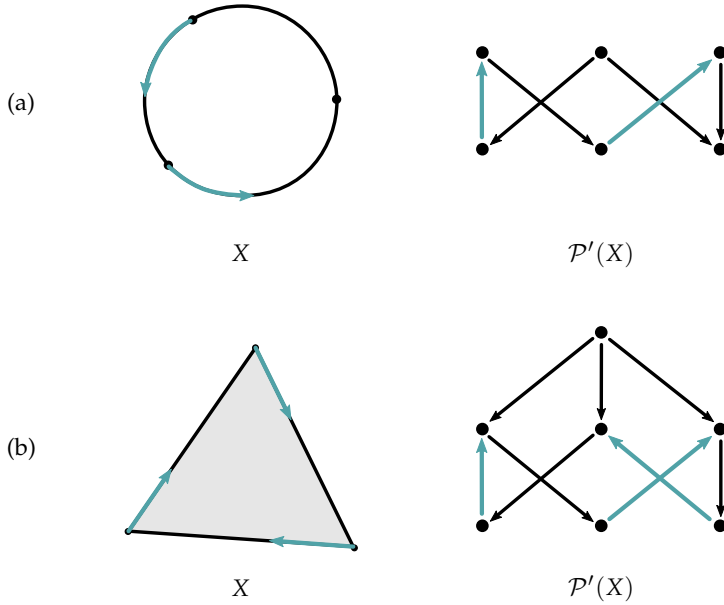


Figure 11: (a) A simplicial complex with an acyclic partial matching Σ , consisting of two pairs $\{x_\bullet > y_\bullet\}$, indicated with arrows on the complex $y \rightarrow x$. Σ can also be viewed in the modified poset graph. (b) A partial matching that is not acyclic, which can be seen most readily in the modified poset graph where we can identify an oriented cycle on the bottom two rows.

4.3 Localisation

Theorem 4.4 provides an inductive rather than constructive proof; in particular, even if we start with a simplicial complex, we are not guaranteed to be homotopy-equivalent to a regular CW complex. For example, any maximal matching on a simplicial structure X on S^2 will have exactly one critical simplex in dimension 0 and one in dimension 2, which means X is homotopy equivalent to an irregular CW complex. The crucial information we are missing is knowledge of the new attachments after the removal of all non-critical pairs of simplices.

To obtain a more explicit description of this homotopy equivalence in terms of Σ -zigzags (such as ζ), we adopt the perspective of (Nanda, 2019). The goal is to construct an LP-category whose objects are critical simplices, whose morphisms $w \rightarrow z$ consist of Σ -zigzags, and whose classifying space is homotopy equivalent to X .

Let Σ be an acyclic partial matching in X , and let Σ^+ denote the union of Σ with all identity morphisms in $\mathbf{Ent}[X]$.

Definition 4.5. The localised entrance path category $\mathbf{Loc}_\Sigma \mathbf{Ent}[X]$ is the category given by the following data:

1. The set of objects is the same as for $\mathbf{Ent}[X]$.
2. For each pair of objects w, z , every morphism is an equivalence

class of finite zigzags of the form

$$w \xrightarrow{g_0} y_0 \xleftarrow{f_0} x_0 \xrightarrow{g_1} y_1 \xleftarrow{f_1} \dots \xleftarrow{f_k} x_k \xrightarrow{g_{k+1}} z$$

where every g_i is an arbitrary morphism in $\mathbf{Ent}[X]$, every f_i is in Σ^+ and the equivalence relation is defined by the following:

- Two zigzags are *horizontally equivalent* if they differ by intermediate identity morphisms, and
- Two zigzags are *vertically equivalent* if they form the rows of a commutative diagram in $\mathbf{Ent}[X]$

$$\begin{array}{ccccccccccccccc} w & \xrightarrow{g_0} & y_0 & \xleftarrow{f_0} & x_0 & \xrightarrow{g_1} & y_1 & \xleftarrow{f_1} & \dots & \xleftarrow{f_k} & x_k & \xrightarrow{g_{k+1}} & z \\ \parallel & \nearrow & \downarrow u_0 & \nwarrow & \downarrow v_0 & \nearrow & \downarrow u_1 & \nwarrow & \dots & \nwarrow & \downarrow v_k & \nearrow & \parallel \\ w & \xrightarrow{g'_0} & y'_0 & \xleftarrow{f'_0} & x'_0 & \xrightarrow{g'_1} & y'_1 & \xleftarrow{f'_1} & \dots & \xleftarrow{f'_k} & x'_k & \xrightarrow{g'_{k+1}} & z \end{array}$$

where $u_i, v_i \in \Sigma^+$ for all i .

3. For $\zeta, \zeta' \in \mathbf{Loc}_\Sigma \mathbf{Ent}[X](w, z)$, $\zeta' \Rightarrow \zeta$ if and only if there are zigzags representing ζ and ζ' that fit into a commutative diagram of the form

$$\begin{array}{ccccccccccccccc} w & \xrightarrow{g_0} & y_0 & \xleftarrow{f_0} & x_0 & \xrightarrow{g_1} & y_1 & \xleftarrow{f_1} & \dots & \xleftarrow{f_k} & x_k & \xrightarrow{g_{k+1}} & z \\ \parallel & \nearrow & \downarrow u_0 & \nwarrow & \downarrow v_0 & \nearrow & \downarrow u_1 & \nwarrow & \dots & \nwarrow & \downarrow v_k & \nearrow & \parallel \\ w & \xrightarrow{g'_0} & y'_0 & \xleftarrow{f'_0} & x'_0 & \xrightarrow{g'_1} & y'_1 & \xleftarrow{f'_1} & \dots & \xleftarrow{f'_k} & x'_k & \xrightarrow{g'_{k+1}} & z \end{array}$$

where $u_i, v_i \in \Sigma^+$ for all i and the representatives of ζ and ζ' form the top and bottom rows, respectively.

4. Composition is given by concatenating representative zigzags, where the two adjacent forward-pointing morphisms are composed in $\mathbf{Ent}[X]$.

This category comes equipped with a localisation functor:

$$\mathbf{L}_\Sigma : \mathbf{Ent}[X] \rightarrow \mathbf{Loc}_\Sigma \mathbf{Ent}[X].$$

$\mathbf{L}_\Sigma$ is the identity on objects and it sends each morphism to its equivalence class. The localisation functor induces a homotopy equivalence of classifying spaces:

Theorem 4.6. (*Nanda, 2019, Theorem 4.1*) *The localisation LP functor $\mathbf{L}_\Sigma : \mathbf{Ent}[X] \rightarrow \mathbf{Loc}_\Sigma \mathbf{Ent}[X]$ induces a homotopy-equivalence of classifying spaces.⁹*

⁹ Definition 4.5 and Theorem 4.6 can actually be stated in greater generality for *cellular* categories, which are (informally) LP categories where each poset of morphisms is either empty or has a ‘minimal’ element.

The **flow category** $\mathbf{Flo}_\Sigma[X]$ is the full subcategory of $\mathbf{Loc}_\Sigma\mathbf{Ent}[X]$ spanned by all critical objects. The flow category is naturally embedded into $\mathbf{Loc}_\Sigma\mathbf{Ent}[X]$ via an inclusion functor

$$J : \mathbf{Flo}_\Sigma[X] \rightarrow \mathbf{Loc}_\Sigma\mathbf{Ent}[X],$$

which also induces a homotopy equivalence of classifying spaces (Nanda, 2019, Theorem 5.7). Combining these homotopy equivalences leads to the following result, (see (Nanda, 2019, Theorem 1.1)).

Theorem 4.7. *If Σ is any acyclic partial matching on the simplicial complex X , then the classifying space $|\Delta\mathbf{Flo}_\Sigma[X]|$ of Σ 's flow category is homotopy equivalent to X .*

In Figure 12 we can see examples of the discrete flow category for cell structures on spaces with different homotopy types.

For example, the first row (a) shows, on the left, a simplicial structure on S^1 with two 0-simplices and two 1-simplices. The acyclic partial matching Σ consists of exactly one pair of simplices, with the matching indicated by the arrow. In the middle column, we construct the flow category: in this example, the flow category has exactly two objects (one unmatched 0-cell and one unmatched 1-cell). There are exactly two ways (up to the equivalence relation) of traversing from the critical 1-cell to the critical 0-cell: either directly via the existing entrance path, or by zigzagging via the unique morphism in Σ . These contribute the two morphisms in the discrete flow category, and they are not related by the partial order in $\mathbf{Flo}$, yielding the classifying space seen on the right.

5 The Morse Complex of Groups

In this section, we construct what we call the *Morse complex of groups*: a complex of groups over the discrete flow category. Whilst the complex of groups in Definition 3.6 summarises a group action, the Morse version tracks how this layer of data changes after applying discrete Morse theory on the quotient.

In order to proceed with the construction, we must first define what it means for the data of an acyclic partial matching to be compatible with the data of a complex of groups.

Definition 5.1. An acyclic partial matching Σ on a simplicial complex X is **compatible** with a complex of groups $\mathbf{F}:\mathbf{Fac}[X] \rightarrow \mathbf{Grp}$ if it satisfies the following conditions, for every morphism $f : x \rightarrow x'$ in Σ :

1. the $\mathbf{F}$ -image $\mathbf{F}f$ is an isomorphism, and

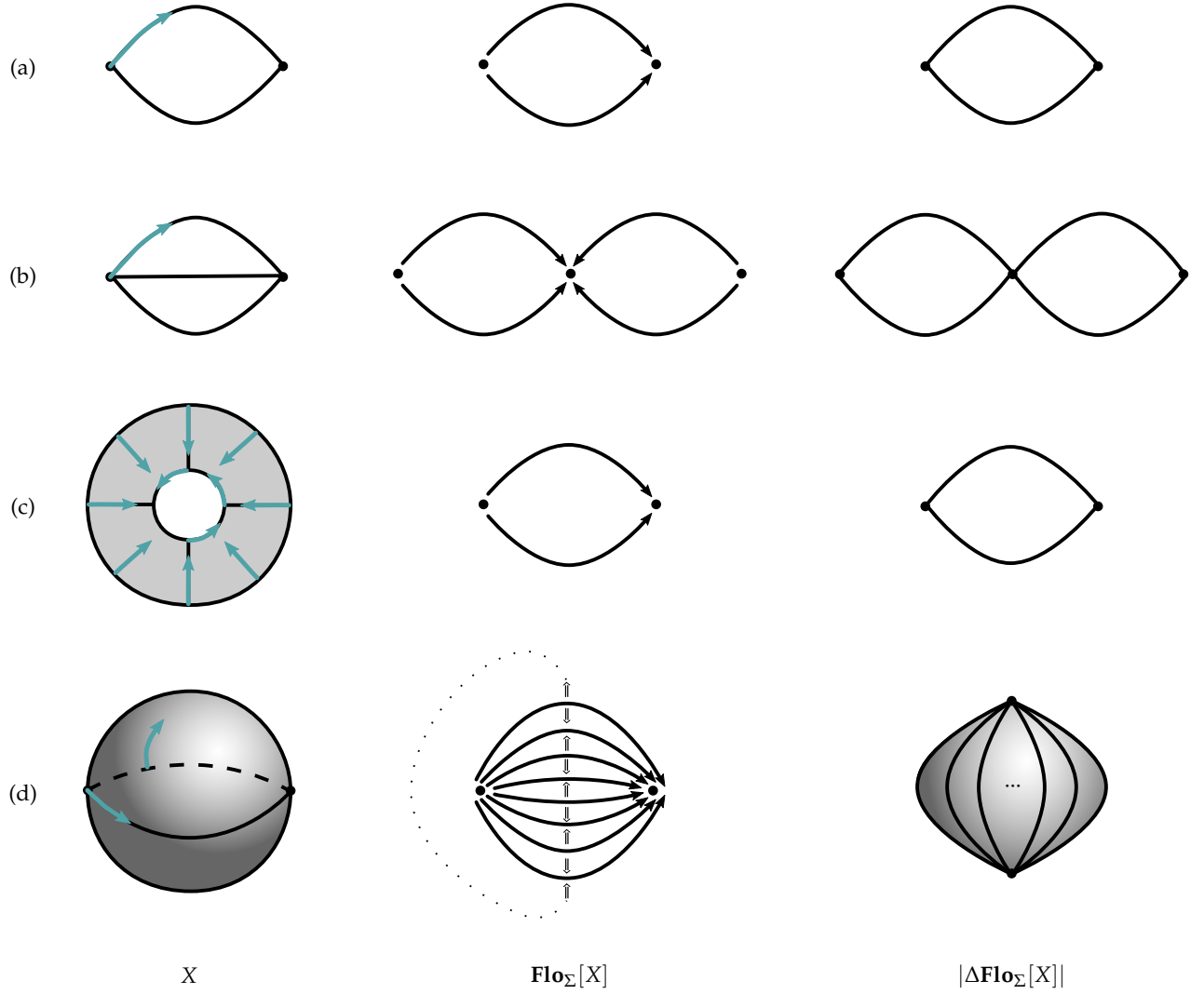


Figure 12: A variety of examples of the discrete flow category of simplicial complexes with acyclic partial matchings, and their classifying spaces. (a) S^1 , (b) $S^1 \vee S^1$, (c) an annulus, (d) S^2 : the cell structure of the classifying space resembles a ‘beach ball’ built from 8 segments. This example illustrates the necessity for poset enrichment because without it, the classifying space would have the homotopy type of the wedge of eight circles.

2. whenever we have

$$\begin{array}{ccc} w & \xrightarrow{g} & x \\ & \searrow g' & \downarrow f \\ & & x' \end{array} \quad \text{or} \quad \begin{array}{ccc} x & & \\ \downarrow f & \searrow h' & \\ x' & \xrightarrow{h} & z \end{array}$$

the $\mathbf{F}$ -images are, respectively,

$$\begin{array}{ccc} \mathbf{F}w & \xrightarrow{\mathbf{F}g} & \mathbf{F}x \\ & \searrow \mathbf{F}g' & \downarrow \mathbf{F}f \\ & & \mathbf{F}x' \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathbf{F}x & & \\ \downarrow \mathbf{F}f & \searrow \mathbf{F}h' & \\ \mathbf{F}x' & \xrightarrow{\mathbf{F}h} & \mathbf{F}z \end{array}$$

Let G act regularly on a simplicial complex X , and let $\mathbf{F} : Y \rightarrow \mathbf{Grp}$ be the complex of groups associated to this action, where Y denotes the quotient space¹⁰. Let $\Phi : \mathbf{F} \rightarrow \underline{G}$ be the associated morphism of complexes of groups and let Σ be an $\mathbf{F}$ -compatible acyclic partial matching on Y . We construct the Morse complex of groups

$$\mathbf{M:Flo}_\Sigma[Y] \rightarrow \mathbf{Grp}$$

as follows.

1. For all objects $y \in \mathbf{Flo}_\Sigma[Y]$, $\mathbf{M}y := \mathbf{F}y$.
2. For each equivalence class of 1-morphisms represented by some zigzag ζ

$$w \xrightarrow{h_0} y_0 \xleftarrow{f_0} x_0 \xrightarrow{h_1} y_1 \xleftarrow{f_1} \cdots \xleftarrow{f_k} x_k \xrightarrow{h_{k+1}} z$$

the homomorphism $\mathbf{M}\zeta : \mathbf{M}w \rightarrow \mathbf{M}z$ is the composite

$$\mathbf{F}h_{k+1} \circ \mathbf{F}f_k^{-1} \circ \cdots \circ \mathbf{F}h_1 \circ \mathbf{F}f_0^{-1} \circ \mathbf{F}h_0,$$

i.e., $\mathbf{M}\zeta$ is conjugation by the product of transfer elements

$$\sigma(\zeta) := \sigma(h_{k+1}) \cdot \sigma(f_k)^{-1} \cdots \sigma(h_1) \cdot \sigma(f_0)^{-1} \cdot \sigma(h_0).$$

3. For two morphisms $\zeta, \zeta' : w \rightarrow z$ in $\mathbf{Flo}_\Sigma[Y]$ satisfying $\zeta' \Rightarrow \zeta$, fitting into a diagram

$$\begin{array}{ccccccccccc} w & \xrightarrow{h_0} & y_0 & \xleftarrow{f_0} & x_0 & \xrightarrow{h_1} & y_1 & \xleftarrow{f_1} & \cdots & \xleftarrow{f_k} & x_k & \xrightarrow{h_{k+1}} & z \\ \downarrow & \nearrow & \downarrow & \nwarrow & \downarrow & \nearrow & \downarrow & \nwarrow & & \nwarrow & \downarrow & \nearrow & \downarrow \\ w & \xrightarrow{h'_0} & y'_0 & \xleftarrow{f'_0} & x'_0 & \xrightarrow{h'_1} & y'_1 & \xleftarrow{f'_1} & \cdots & \xleftarrow{f'_k} & x'_k & \xrightarrow{h'_{k+1}} & z \end{array} \quad (3)$$

¹⁰ Recall that we are using X and $\mathbf{Fac}[X]$ interchangeably, and similarly for Y .

where ζ, ζ' form the top and bottom rows, respectively, the 2-morphism

$$\mathbf{M}(\zeta' \Rightarrow \zeta) : \mathbf{M}\zeta' \Rightarrow \mathbf{M}\zeta$$

is defined to be the horizontal composition of group elements $t_0 \circ t_1 \circ \dots \circ t_{k+1}$ along the following diagram

$$\begin{array}{ccccccccc}
 & \text{Fu}_0 \circ \text{Fh}_0 & & \text{Fu}_0 \circ \text{Ff}_0 & & \text{Fu}_1 \circ \text{Fh}_1 & & \text{Fu}_k \circ \text{Ff}_k & & \text{Fh}_{k+1} \\
 & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow \\
 \text{Fw} & \xrightarrow{t_0} & \text{Fy}'_0 & \xrightarrow{1_G} & \text{Fx}_0 & \xrightarrow{t_1} & \dots & \xrightarrow{1_G} & \text{Fx}_k & \xrightarrow{t_{k+1}} & \text{Fz} \\
 & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
 & \text{Fh}'_0 & & \text{Ff}'_0 \circ \text{Fv}_0 & & \text{Fh}'_1 \circ \text{Fv}_0 & & \text{Ff}'_k \circ \text{Fv}_k & & \text{Fh}'_{k+1} \circ \text{Fv}_k
 \end{array} \quad (4)$$

where each t_i is the 2-morphism assigned by $\mathbf{F}$.¹¹

4. For any composable morphisms $y \xrightarrow{\zeta} y' \xrightarrow{\zeta'} y''$, the twisting element is

$$\gamma_{\mathbf{M}}(\zeta, \zeta') = \sigma(\zeta') \cdot \sigma(\zeta) \cdot \sigma(\zeta' \circ \zeta)^{-1}.$$

¹¹ Note that we have identity twisting elements in all backwards portions of the diagram because the $\mathbf{F}$ -image of identity 2-morphisms are identities.

Proposition 5.2. *The 1-morphisms are well-defined i.e. for two zigzags $\zeta \sim \zeta'$, we have the equivalence $\mathbf{M}\zeta \equiv \mathbf{M}\zeta'$, and every $\mathbf{M}\zeta$ is injective.*

Proof. If ζ and ζ' are horizontally equivalent then clearly $\mathbf{M}\zeta \equiv \mathbf{M}\zeta'$ because $\text{Fid}_y = \text{id}_{\text{F}y}$, for all objects $y \in \text{Fac}[\mathcal{Y}]$. If ζ and ζ' are vertically equivalent, then they fit into a diagram as follows.

$$\begin{array}{ccccccccccccccc}
 w & \xrightarrow{h_0} & y_0 & \xleftarrow{f_0} & x_0 & \xrightarrow{h_1} & y_1 & \xleftarrow{f_1} & \dots & \xleftarrow{f_k} & x_k & \xrightarrow{h_{k+1}} & z \\
 \downarrow & \parallel & \downarrow & \parallel & \downarrow & \parallel & \downarrow & \parallel & \parallel & \parallel & \downarrow & \parallel & \downarrow \\
 w & \xrightarrow{h'_0} & y'_0 & \xleftarrow{f'_0} & x'_0 & \xrightarrow{h'_1} & y'_1 & \xleftarrow{f'_1} & \dots & \xleftarrow{f'_k} & x'_k & \xrightarrow{h'_{k+1}} & z
 \end{array}$$

By the second compatibility criterion for Σ , for each backward-pointing column we have

$$\text{Fu}_i \circ \text{Ff}_i = \text{Ff}'_i \circ v_i = \text{Ff}'_i \circ \text{Fv}_i,$$

and similarly, for each forward-pointing column we have

$$\text{Fu}_i \circ \text{Fh}_i = \text{Fh}'_i \circ \text{Fv}_{i-1},$$

with the convention that $v_{-1} = \text{id}_z$ and $u_{k+1} = \text{id}_w$. These equations yield the relations

$$\text{Ff}_i^{-1} = \text{Fv}_i^{-1} \circ \text{Ff}'_i^{-1} \circ \text{Fu}_i$$

$$\text{Fh}_i = \text{Fu}_i^{-1} \circ \text{Fh}'_i \circ \text{Fv}_{i-1},$$

for all suitable i , and are well-defined since $f_i, u_i, v_i \in \Sigma^+$ so their $\mathbf{F}$ -images are isomorphisms by the first compatibility criterion. Substituting into $\mathbf{M}\zeta$, we have that

$$\begin{aligned}
\mathbf{M}\zeta &= \mathbf{F}h_{k+1} \circ \mathbf{F}f_k^{-1} \circ \cdots \circ \mathbf{F}h_1 \circ \mathbf{F}f_0^{-1} \circ \mathbf{F}h_0 \\
&= (\mathbf{F}h'_{k+1} \circ \mathbf{F}v_k) \circ \cdots \circ (\mathbf{F}u_1^{-1} \circ \mathbf{F}h'_1 \circ \mathbf{F}v_0) \circ (\mathbf{F}v_0^{-1} \circ \mathbf{F}f'_0{}^{-1} \circ \mathbf{F}u_0) \circ (\mathbf{F}u_0^{-1} \circ \mathbf{F}h'_0) \\
&= \mathbf{F}h'_{k+1} \circ \mathbf{F}f'_k{}^{-1} \circ \cdots \circ \mathbf{F}h'_1 \circ \mathbf{F}f'_0{}^{-1} \circ \mathbf{F}h'_0 \\
&= \mathbf{M}\zeta',
\end{aligned}$$

as required. By construction, each $\mathbf{M}\zeta$ is the composite of finitely many injective homomorphisms, and is therefore injective. $\square$

Proposition 5.3. *If ζ, ζ' are two morphisms satisfying $\zeta' \Rightarrow \zeta$, fitting into a diagram of the form as in Diagram (3), then for every $n \leq k+1$, the horizontal composition of 2-morphisms $T_n := t_0 \circ \cdots \circ t_n$ satisfies*

$$T_n : \mathbf{M}\zeta'_n \Rightarrow \mathbf{F}u_n \circ \mathbf{M}\zeta_n, \quad (5)$$

where $\mathbf{M}\zeta_n$ is the truncated composition

$$\mathbf{F}h_n \circ \mathbf{F}f_{n-1}^{-1} \circ \cdots \circ \mathbf{F}h_1 \circ \mathbf{F}f_0^{-1} \circ \mathbf{F}h_0,$$

$u_{k+1} = id_z$ and the t_i are defined in the same way as in diagram 4.

In particular, the 2-morphisms of $\mathbf{M}$ are well-defined i.e.

$$T := t_0 \circ t_1 \circ \cdots \circ t_{k+1}$$

satisfies

$$\mathbf{M}\zeta'(s) = T \cdot \mathbf{M}\zeta(s) \cdot T^{-1}$$

for all $s \in \mathbf{M}w$.

Proof. We prove the first part of the proposition by induction on n .

By construction, $t_0 : \mathbf{F}h'_0 \Rightarrow \mathbf{F}u_0 \circ \mathbf{F}h_0$ satisfies

$$\mathbf{F}h'_0(s) = t_0 \cdot \mathbf{F}u_0 \circ \mathbf{F}h_0(s) \cdot t_0^{-1},$$

for all $s \in \mathbf{F}w$, so the base case holds. For the inductive step, suppose that Equation (5) holds for some $n-1 < k+1$. By definition, $T_n = \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(T_{n-1}) \right) \cdot t_n$, and we have two relations arising directly from the diagrams:

$$\mathbf{F}h'_n(\mathbf{F}v_{n-1}(s)) \cdot t_n = t_n \cdot \mathbf{F}u_n(\mathbf{F}h_n(s)) \quad (6)$$

$$\mathbf{F}f'_{n-1}{}^{-1}(r) = \mathbf{F}v_{n-1}{}^{-1} \left(\mathbf{F}f'_{n-1}{}^{-1}(\mathbf{F}u_{n-1}(r)) \right). \quad (7)$$

For all $s \in \mathbf{M}w$,

$$\begin{aligned}
\mathbf{M}\zeta'_n(s) \cdot T_n &= \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(\mathbf{M}\zeta'_{n-1}(s)) \right) \cdot \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(T_{n-1}) \right) \cdot t_n \\
&= \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(\mathbf{M}\zeta'_{n-1}(s) \cdot T_{n-1}) \right) \cdot t_n,
\end{aligned}$$

and similarly

$$\begin{aligned}
T_n \cdot \mathbf{F}u_n \circ \mathbf{M}\zeta_n(s) &= \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(T_{n-1}) \right) \cdot t_n \cdot \mathbf{F}u_n \left(\mathbf{F}h_n \left(\mathbf{F}f_{n-1}{}^{-1}(\mathbf{M}\zeta_{n-1}(s)) \right) \right) \\
&= \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(T_{n-1}) \right) \cdot \mathbf{F}h'_n \left(\mathbf{F}v_{n-1} \left(\mathbf{F}f_{n-1}{}^{-1}(\mathbf{M}\zeta_{n-1}(s)) \right) \right) \cdot t_n, && \text{by Equation (6)} \\
&= \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(T_{n-1}) \right) \cdot \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(\mathbf{F}u_{n-1}(\mathbf{M}\zeta_{n-1}(s))) \right) \cdot t_n, && \text{by Equation (7)} \\
&= \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(T_{n-1} \cdot \mathbf{F}u_{n-1}(\mathbf{M}\zeta_{n-1}(s))) \right) \cdot t_n \\
&= \mathbf{F}h'_n \left(\mathbf{F}f'_{n-1}{}^{-1}(\mathbf{M}\zeta'_{n-1}(s) \cdot T_{n-1}) \right) \cdot t_n,
\end{aligned}$$

where the final step follows from the inductive assumption. Hence, $\mathbf{M}\zeta'_n(s) \cdot T_n = T_n \cdot \mathbf{M}\zeta_n(s)$ for all $s \in \mathbf{M}w$. When $n = k + 1$, Equation (5) reduces to

$$t_0 \circ \cdots \circ t_{k+1} : \mathbf{M}\zeta' \Rightarrow \mathbf{M}\zeta,$$

since $\mathbf{F}u_{n+1} = \text{id}_{\mathbf{F}z}$, so the 2-morphisms $\mathbf{M}\zeta' \Rightarrow \zeta$ satisfy the conjugation criterion in **Grp**. $\square$

Whilst the most natural perspective of the definition of the 2-morphisms is the construction given above, the next result shows that it yields the same answer as $\sigma(\zeta') \cdot \sigma(\zeta)^{-1}$, where $\mathbf{M}\zeta = \text{Ad}(\sigma(\zeta))$.

Proposition 5.4. *For two morphisms $\zeta' \Rightarrow \zeta$, the twisting element $\mathbf{M}(\zeta' \Rightarrow \zeta) := t_0 \circ \cdots \circ t_{k+1}$ is equal to $\sigma(\zeta') \cdot \sigma(\zeta)^{-1}$.*

The proof follows immediately from Lemma 5.5 below, but first we set some notation. For any ζ, ζ' defined as per Diagram (3), let $\zeta_n := h_n \circ \cdots \circ f_0^{-1} \circ h_0$ be the truncated composition and similarly for ζ' . Let

$$T_n := t_0 \circ \cdots \circ t_n,$$

as above, which is a twisting element $\mathbf{M}\zeta'_n \Rightarrow \mathbf{F}u_n \circ \mathbf{M}\zeta_n$ by Proposition 5.3.

Lemma 5.5. *For any $n \leq k + 1$,*

$$\sigma(\zeta'_n)\sigma(\zeta_n)^{-1}\sigma(u_n)^{-1} = T_n, \tag{8}$$

with the convention that $u_{k+1} = \text{id}_z$.

Proof. We proceed by induction on n ; the base case follows immediately, since

$$t_0 = \sigma(h'_0)\sigma(h_0)^{-1}\sigma(u_0)^{-1} = \sigma(\zeta'_0)\sigma(\zeta_0)^{-1}\sigma(u_0)^{-1}.$$

For the inductive step, suppose the statement holds for some $n - 1$. The left-hand side of Equation (8) becomes:

$$\sigma(\zeta'_n)\sigma(\zeta_n)^{-1}\sigma(u_n)^{-1} = \left(\sigma(h'_n)\sigma(f'_{n-1})^{-1} \cdots \sigma(h'_0) \right) \cdot \left(\sigma(h_0)^{-1}\sigma(f_0) \cdots \sigma(f_{n-1})\sigma(h_n)^{-1} \right) \cdot \sigma(u_n)^{-1},$$

and the right-hand side becomes:

$$\begin{aligned}
T_n &= T_{n-1} \circ t_n \\
&= \mathbf{F}h'_n \left(\mathbf{F}v_{n-1} \left(\mathbf{F}v_{n-1}^{-1} \left(\mathbf{F}f'_{n-1}{}^{-1}(T_{n-1}) \right) \right) \right) \cdot t_n \\
&= \left(\sigma(h'_n) \sigma(f'_{n-1})^{-1} \cdot T_{n-1} \cdot \sigma(f'_{n-1}) \sigma(h'_n)^{-1} \right) \cdot \left(\sigma(h'_n) \sigma(v_{n-1}) \sigma(h_n)^{-1} \sigma(u_n)^{-1} \right) \\
&= \sigma(h'_n) \sigma(f'_{n-1})^{-1} \cdot \left(\sigma(\zeta'_{n-1}) \sigma(\zeta_{n-1})^{-1} \sigma(u_{n-1})^{-1} \right) \cdot \sigma(f'_{n-1}) \sigma(h'_n)^{-1} \sigma(h'_n) \sigma(v_{n-1}) \sigma(h_n)^{-1} \sigma(u_n)^{-1},
\end{aligned}$$

where the last step follows from the inductive assumption. This simplifies to:

$$\begin{aligned}
&\sigma(\zeta'_n) \cdot \sigma(\zeta_{n-1})^{-1} \left(\sigma(u_{n-1})^{-1} \sigma(f'_{n-1}) \sigma(v_{n-1}) \right) \sigma(h_n)^{-1} \cdot \sigma(u_n)^{-1} \\
&= \sigma(\zeta'_n) \cdot \left(\sigma(\zeta_{n-1})^{-1} \sigma(f_{n-1})^{-1} \sigma(h_n)^{-1} \right) \cdot \sigma(u_n)^{-1} \\
&= \sigma(\zeta'_n) \sigma(\zeta_n)^{-1} \sigma(u_n)^{-1}.
\end{aligned}$$

□

Proof of Proposition 5.4. The desired equality follows immediately from Lemma 5.5, setting $n = k + 1$. □

Proposition 5.6. $\mathbf{M}$ is a well-defined complex of groups.

Proof. Propositions 5.2, 5.3 and 5.4 leave us with just the cocycle condition to check: if ζ_0, ζ_1 and ζ_2 are three composable Σ -zigzags

$$\bullet \xrightarrow{\zeta_0} \bullet \xrightarrow{\zeta_1} \bullet \xrightarrow{\zeta_2} \bullet,$$

$$\begin{aligned}
\gamma(\zeta_1, \zeta_2) \cdot \gamma(\zeta_0, \zeta_2 \circ \zeta_1) &= \left(\sigma(\zeta_2) \sigma(\zeta_1) \sigma(\zeta_2 \circ \zeta_1)^{-1} \right) \cdot \left(\sigma(\zeta_2 \circ \zeta_1) \sigma(\zeta_0) \sigma(\zeta_2 \circ \zeta_1 \circ \zeta_0)^{-1} \right) \\
&= \sigma(\zeta_2) \sigma(\zeta_1) \sigma(\zeta_0) \sigma(\zeta_2 \circ \zeta_1 \circ \zeta_0)^{-1},
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{M}\zeta_2(\gamma(\zeta_0, \zeta_1)) \cdot \gamma(\zeta_1 \circ \zeta_0, \zeta_2) &= \left(\sigma(\zeta_2) \cdot \gamma(\zeta_0, \zeta_1) \cdot \sigma(\zeta_2)^{-1} \right) \cdot \left(\sigma(\zeta_2) \sigma(\zeta_1 \circ \zeta_0) \sigma(\zeta_2 \circ \zeta_1 \circ \zeta_0)^{-1} \right) \\
&= \gamma(\zeta_1, \zeta_2) \cdot \gamma(\zeta_0, \zeta_2 \circ \zeta_1),
\end{aligned}$$

as required. □

We define the embedding $\Psi : \mathbf{M} \rightarrow \underline{\mathbf{G}}_{\mathbf{Flo}_\Sigma[Y]}$ as follows:

1. for each object $y \in \mathbf{Flo}_\Sigma[Y]$, $\Psi_y : \mathbf{M}y \rightarrow \underline{\mathbf{G}}$ is just the inclusion, as before, and
2. for each morphism $\zeta : w \rightarrow z$, the twisting element Ψ_ζ is $\sigma(\zeta)$.

6 Equivariant Homotopy Equivalence

We are now ready to state and prove our main result, which combines the Basic Construction with discrete Morse theory:

Theorem 6.1. *Given a regular simplicial action of G on X , let $\mathbf{F}$ and Φ denote the complex of groups and monomorphism associated to this action, for some choice of lifts and transfers. If Σ is an $\mathbf{F}$ -compatible acyclic partial matching on the quotient Y , then the Morse complex of groups $\mathbf{M} : \mathbf{Flo}_\Sigma[Y] \rightarrow \mathbf{Grp}$ and associated monomorphism $\Psi : \mathbf{M} \rightarrow \underline{\mathbf{G}}_{\mathbf{Flo}_\Sigma[Y]}$ satisfy the following property: the classifying space of the development $\mathcal{D}(\mathbf{M}, \Psi)$ is G -equivariantly homotopy-equivalent to X .*

Throughout this section, we assume that the hypotheses of Theorem 6.1 hold. We will omit $\mathbf{Ent}$ from the notation for the localised entrance path category and discrete flow category, for simplicity.

The proof follows from a sequence of results, and the approach is summarised in Figure 13. The high-level strategy is to lift the acyclic partial matching Σ on Y to a matching $\tilde{\Sigma}$ in X , to push the G -action on X to $\mathbf{Flo}_{\tilde{\Sigma}}[X]$, and to show that $\mathbf{M}$ arises as the complex of groups associated to this action, for some choices dependent on the original choices made to construct $\mathbf{F}$. Then, by Proposition 3.11, $\mathcal{D}(\mathbf{M}, \Psi)$ is isomorphic to $\mathbf{Flo}_{\tilde{\Sigma}}[X]$, whose classifying space is homotopy-equivalent to X (Theorem 4.7).

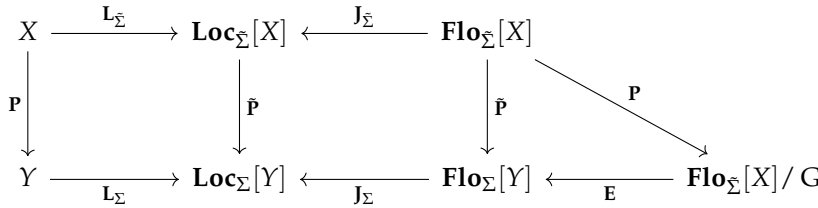


Figure 13: We prove the theorem by lifting the acyclic partial matching to a G -invariant one on X , and ultimately by showing that the development of $(\mathbf{M}, \Psi)$ is isomorphic to $\mathbf{Flo}_{\tilde{\Sigma}}[X]$.

Lemma 6.2. *If $\Sigma = \{f_i : x_i \rightarrow y_i\}$ is an acyclic partial matching on the quotient Y , then $\tilde{\Sigma} := \bigcup_i Gf_i$ is an acyclic partial matching on X .*

Proof. We have to prove that $\tilde{\Sigma}$ satisfies the dimension, partition and acyclicity conditions from Definition 4.3. The dimension property follows immediately since $\dim(\tilde{z}) = \dim(z)$ for any $\tilde{z}$ satisfying $\mathbf{P}(\tilde{z}) = z$.

Since Σ satisfies the partition property, it follows that, for $i \neq j$, $Gf_i \cap Gf_j = \emptyset$. So it remains to show that for a given $f \in \Sigma$, the orbit $Gf \subset \tilde{\Sigma}$ contains no repeated objects. This could occur in precisely two ways, as illustrated in Figure 14.

Case 1 cannot occur because there is a unique morphism with a given source in its orbit, so $g\tilde{f} = h\tilde{f}$. Suppose we are in Case 2; then

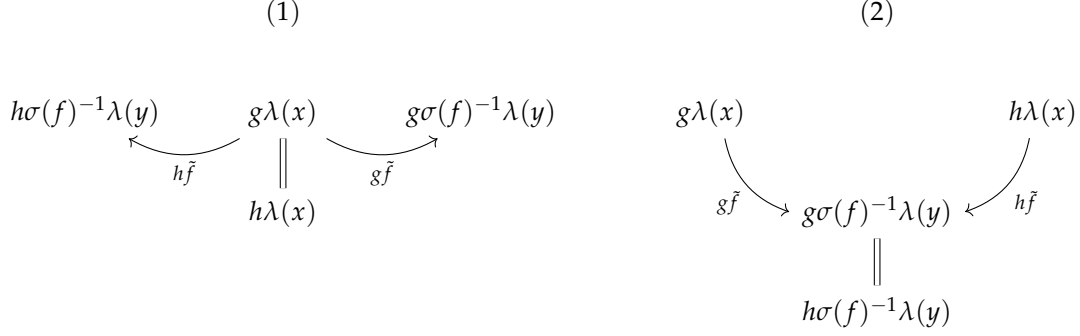


Figure 14: The two possible ways in which $\tilde{\Sigma}$ could violate the partition property of an acyclic partial matching.

$g\sigma(f)^{-1}\lambda(y) = h\sigma(f)^{-1}\lambda(y)$ and so $\sigma(f)h^{-1}g\sigma(f)^{-1} \in \mathbf{F}y$. Recall that since Σ is $\mathbf{F}$ -compatible, the morphism $\mathbf{F}f$ is an isomorphism for every $f \in \Sigma$. Hence, there is a $k \in \mathbf{F}x$ such that

$$\mathbf{F}f(k) = \sigma(f) \cdot k \cdot \sigma(f)^{-1} = \sigma(f) \cdot h^{-1}g \cdot \sigma(f)^{-1}.$$

So $k = h^{-1}g$, and $g\lambda(x) = h\lambda(x)$ and so we satisfy the partition property.

Suppose that $\tilde{\Sigma}$ is not acyclic, i.e. that there exists a sequence (writing $\tilde{x}_i$ for $\lambda(x_i)$):

$$g_1\sigma(f_1)^{-1}\tilde{y}_1 < g_1\tilde{x}_1 > g_2\sigma(f_2)^{-1}\tilde{y}_2 < \cdots > g_n\sigma(f_n)^{-1}\tilde{y}_n < g_n\tilde{x}_n,$$

with $g_n\tilde{x}_n > g_1\sigma(f_1)^{-1}\tilde{y}_1$. The projection of this sequence

$$y_1 < x_1 > y_2 < x_2 > \cdots > y_n < x_n$$

has the property that $x_n > y_1$.

This projected sequence could contain repeats if any of the $\tilde{x}_i$ are in the same orbit, for example. If, say, the pair $y_i < x_i$ is equal to $y_j < x_j$ for some $i < j$, remove from the projected sequence the entire portion between these two pairs, inclusive of $y_j < x_j$ and exclusive of $y_i < x_i$.

If this sequence has length strictly bigger than one, it contradicts the acyclicity of Σ , and we are done. This reduced sequence will have length equal to one if and only if all pairs involving $\tilde{y}_i$ and $\tilde{x}_i$ lie in the same orbit. This would require the existence of relations of the form $g_1\tilde{y} < \tilde{x} > g_2\tilde{y}$ for some group elements g_1 and g_2 , i.e., it would require two morphisms in the same orbit with the same source. By condition (2) of Definition 2.1, we must have that $g_1\tilde{y} = g_2\tilde{y}$, so the sequence cannot have length equal to one. $\square$

Using the action on X , we can define an action on $\mathbf{Loc}_{\tilde{\Sigma}}[X]$: the objects of both categories are the same, and the action on objects

remains the same. On morphisms, each $g \in G$ acts individually on each arrow in the zigzag. More precisely, if ζ is the $\tilde{\Sigma}$ zigzag

$$w \xrightarrow{h_0} y_0 \xleftarrow{f_0} x_0 \xrightarrow{h_1} y_1 \xleftarrow{f_1} \cdots \xleftarrow{f_k} x_k \xrightarrow{h_{k+1}} z,$$

then the morphism $g \cdot \zeta$ is defined to be

$$gw \xrightarrow{g \cdot h_0} gy_0 \xleftarrow{g \cdot f_0} gx_0 \xrightarrow{g \cdot h_1} gy_1 \xleftarrow{g \cdot f_1} \cdots \xleftarrow{g \cdot f_k} gx_k \xrightarrow{g \cdot h_{k+1}} gz.$$

The following result confirms that this action is well-defined, and also that it restricts to $\mathbf{Flo}_{\tilde{\Sigma}}[X]$.

Lemma 6.3. *The induced G -action on $\mathbf{Flo}_{\tilde{\Sigma}}[X]$ is well-defined.*

Proof. First note that the objects of $\mathbf{Loc}_{\tilde{\Sigma}}[X]$ are the same as for X , and so is the action on the objects; for the action to be well-defined on $\mathbf{Flo}_{\tilde{\Sigma}}[X]$, we need that the orbits of critical objects contain only critical objects, and similarly for non-critical objects. But this follows immediately from the fact that $\tilde{\Sigma}$ is G -invariant.

The G -invariance of $\tilde{\Sigma}$ also guarantees that $\tilde{\Sigma}$ -zigzags are sent to $\tilde{\Sigma}$ -zigzags under the action. That is, if ζ is the $\tilde{\Sigma}$ zigzag

$$w \xrightarrow{h_0} y_0 \xleftarrow{f_0} x_0 \xrightarrow{h_1} y_1 \xleftarrow{f_1} \cdots \xleftarrow{f_k} x_k \xrightarrow{h_{k+1}} z,$$

then the morphism $g \cdot \zeta$

$$gw \xrightarrow{g \cdot h_0} gy_0 \xleftarrow{g \cdot f_0} gx_0 \xrightarrow{g \cdot h_1} gy_1 \xleftarrow{g \cdot f_1} \cdots \xleftarrow{g \cdot f_k} gx_k \xrightarrow{g \cdot h_{k+1}} gz,$$

is also a $\tilde{\Sigma}$ -zigzag because each $g \cdot f_i$ is in $\tilde{\Sigma}$ by construction of $\tilde{\Sigma}$.

Next we check that if $\zeta \sim \tilde{\zeta}$ then $g\zeta \sim g\tilde{\zeta}$, for any $g \in G$. If ζ and $\tilde{\zeta}$ are horizontally equivalent, then their g -images are also horizontally equivalent because $g \cdot \text{id}_z = \text{id}_{g \cdot z}$ for any object z . If ζ and $\tilde{\zeta}$ are vertically equivalent, then they fit into a diagram as follows

$$\begin{array}{ccccccc} w & \longrightarrow & y_0 & \longleftarrow & x_0 & \longrightarrow & \cdots \longrightarrow z \\ \parallel & \nearrow & \downarrow & \nwarrow & \downarrow & \nearrow & \nwarrow \parallel \\ w & \longrightarrow & y'_0 & \longleftarrow & x'_0 & \longrightarrow & \cdots \longrightarrow z \end{array}$$

which maps to the following under g :

$$\begin{array}{ccccccc} g \cdot w & \longrightarrow & g \cdot y_0 & \longleftarrow & g \cdot x_0 & \longrightarrow & \cdots \longrightarrow g \cdot z \\ \parallel & \nearrow & \downarrow & \nwarrow & \downarrow & \nearrow & \nwarrow \parallel \\ g \cdot w & \longrightarrow & g \cdot y'_0 & \longleftarrow & g \cdot x'_0 & \longrightarrow & \cdots \longrightarrow g \cdot z \end{array}$$

Hence, $g \cdot \zeta$ is vertically equivalent to $g \cdot \tilde{\zeta}$ and so the induced action is well-defined on equivalence classes of morphisms.

Lastly, we need to check that the induced action still satisfies the two properties in Definition 2.1. Suppose that ζ is the morphism

$$w \xrightarrow{h_0} y_0 \xleftarrow{f_0} x_0 \xrightarrow{h_1} y_1 \xleftarrow{f_1} \cdots \xleftarrow{f_k} x_k \xrightarrow{h_{k+1}} z,$$

and suppose also that $g \cdot w = z$. Then the morphism $g \cdot \zeta$

$$z \xrightarrow{g \cdot h_0} g y_0 \xleftarrow{g \cdot f_0} g x_0 \xrightarrow{g \cdot h_1} g y_1 \xleftarrow{g \cdot f_1} \cdots \xleftarrow{g \cdot f_k} g x_k \xrightarrow{g \cdot h_{k+1}} g z$$

projects under the orbit functor to a cycle in $\mathbf{Flo}_\Sigma[Y]$, contradicting the acyclicity of Σ .

If instead that there is some $g \in G$ fixing w , then $gh_0 : w \rightarrow gy_0$. But, since the group action on X satisfies property 2 of Definition 2.1, we must have that $gh_0 = h_0$, and in particular that $gy_0 = y_0$. Since $\tilde{\Sigma}$ satisfies the partition property, $gx_0 = x_0$ and $gf_0 = f_0$, because y_0 cannot appear as the target of any other morphism in $\tilde{\Sigma}$. Applying this argument to the rest of the chain of morphisms forming ζ implies that, if g fixes w , then g fixes ζ . $\square$

We will construct a functor $\mathbf{E}$ between the quotient $\mathbf{Flo}_\Sigma[X]/G$ and the flow category on the original quotient $\mathbf{Flo}_\Sigma[Y]$, and show that it is an isomorphism of categories. This will enable us to relate our Morse complex of groups and the associated complex of groups over the orbit space of the action on $\mathbf{Flo}_\Sigma[X]$.

Let $\mathbf{P}$ denote the usual orbit functor from $\mathbf{Flo}_\Sigma[X]$ to $\mathbf{Flo}_\Sigma[X]/G$. Using $[\bullet]_X$ to denote orbit classes under the G -action on X , define $\tilde{\mathbf{P}} : \mathbf{Flo}_\Sigma[X] \rightarrow \mathbf{Flo}_\Sigma[Y]$ to be the functor mapping each $\tilde{\Sigma}$ -zigzag

$$\tilde{h}_{k+1} \circ \tilde{f}_k^{-1} \circ \cdots \circ \tilde{f}_0^{-1} \circ \tilde{h}_0$$

to the Σ -zigzag

$$[\tilde{h}_{k+1}]_X \circ [\tilde{f}_k^{-1}]_X \circ \cdots \circ [\tilde{f}_0^{-1}]_X \circ [\tilde{h}_0]_X.$$

$\tilde{\mathbf{P}}$ send each object to its orbit class under the G action on $\mathbf{Flo}_\Sigma[X]$.

Lemma 6.4. *$\tilde{\mathbf{P}}$ is a surjective LP functor.*

Proof. First we show that $\tilde{\mathbf{P}}$ is well-defined on equivalence classes; if two morphisms are horizontally equivalent then their $\tilde{\mathbf{P}}$ images differ by the $\mathbf{P}$ -image of an identity morphism, which is just the identity. Similarly, if two morphisms are vertically equivalent, then the $\tilde{\mathbf{P}}$ -image of the ladder-diagram also has equality in each square, because each square is just its $\mathbf{P}$ -image, and $\mathbf{P}$ is an LP functor. For the same reasons, $\tilde{\mathbf{P}}$ is order-preserving.

Every constituent morphism $h : x \rightarrow y$ in a Σ -zigzag can be lifted to a morphism in its orbit class with prescribed source $g\lambda(x)$, for any

$g \in G$. Any morphism in Σ is lifted to a morphism in $\tilde{\Sigma}$. Hence, we can find a $\tilde{\Sigma}$ morphism $\tilde{\zeta}$ in $\mathbf{Flo}_{\tilde{\Sigma}}[X]$ such that $\tilde{\mathbf{P}}(\tilde{\zeta}) = \zeta$. $\square$

$$\begin{array}{ccc}
 & \mathbf{Flo}_{\tilde{\Sigma}}[X] & \\
 \mathbf{P} \swarrow & & \searrow \tilde{\mathbf{P}} \\
 \mathbf{Flo}_{\tilde{\Sigma}}[X]/G & \xrightarrow{\mathbf{E}} & \mathbf{Flo}_{\Sigma}[Y]
 \end{array}$$

Observe that the sets of objects in $\mathbf{Flo}_{\tilde{\Sigma}}[X]/G$ and $\mathbf{Flo}_{\Sigma}[Y]$ are the same; first note that the G -action on X and $\mathbf{Loc}_{\tilde{\Sigma}}[X]$ is the same on objects¹², so the sets of orbit classes in $\mathbf{Loc}_{\tilde{\Sigma}}[X]/G$ and $\mathbf{Loc}_{\Sigma}[Y]$ are equal. Secondly, it is clear from the construction of $\tilde{\Sigma}$ that an object $[x] \in \mathbf{Loc}_{\Sigma}[Y]$ is Σ -critical if and only if gx is $\tilde{\Sigma}$ -critical for every $g \in G$. Hence, when we restrict to the flow categories, the sets of classes of objects are equal.

¹² Recall that these two categories have the same objects, i.e., elements of the poset $\mathbf{Fac}[X]$.

Thus, we define $\mathbf{E} : \mathbf{Flo}_{\tilde{\Sigma}}[X]/G \rightarrow \mathbf{Flo}_{\Sigma}[Y]$ to be the identity on objects. Since $\mathbf{P}$ is surjective, for each orbit class of morphisms $[\zeta] \in \mathbf{Flo}_{\tilde{\Sigma}}[X]/G$ there exists a morphism $\tilde{\zeta} \in \mathbf{Flo}_{\tilde{\Sigma}}[X]$ projecting to ζ . Define

$$\mathbf{E}([\zeta]) := \tilde{\mathbf{P}}(\tilde{\zeta}).$$

Clearly $\mathbf{E}(\text{id}_x) = \text{id}_{\mathbf{E}x} = \text{id}_x$ because the only morphisms in $\mathbf{Flo}_{\tilde{\Sigma}}[X]$ projecting to the identity are identities on $g\lambda(x)$, for all g in G . Any such map projects to id_x under $\tilde{\mathbf{P}}$. Also, $\mathbf{E}([\zeta] \circ [\zeta']) = \mathbf{E}([\zeta]) \circ \mathbf{E}([\zeta'])$, and any lifted morphism projecting to $\zeta \circ \zeta'$ can be separated into composable choices of $\tilde{\zeta}$ and $\tilde{\zeta}'$. Whilst choices made separately for $\mathbf{E}([\zeta]) \circ \mathbf{E}([\zeta'])$ may not be composable in $\mathbf{Flo}_{\tilde{\Sigma}}[X]$, they will lie in the same respective orbits, and so will project to the same morphisms under $\tilde{\mathbf{P}}$.

Proposition 6.5. *The functor $\mathbf{E} : \mathbf{Flo}_{\tilde{\Sigma}}[X]/G \rightarrow \mathbf{Flo}_{\Sigma}[Y]$ is a well-defined isomorphism of categories.*

Proof. This follows from the following claims:

1. *Claim:* The functor $\mathbf{E}$ is independent of choice of $\tilde{\zeta}$ in $\mathbf{Flo}_{\tilde{\Sigma}}[X]$.
Proof: If $\tilde{\zeta}$ and $\tilde{\zeta}'$ are two $\tilde{\Sigma}$ -zigzags such that $\mathbf{P}(\tilde{\zeta}) = \mathbf{P}(\tilde{\zeta}')$, then there exists a group element $g \in G$ such that $g\tilde{\zeta} = \tilde{\zeta}'$. But $\tilde{\mathbf{P}}(g\tilde{\zeta}) = \tilde{\mathbf{P}}(\tilde{\zeta})$ because each forward or backwards pointing arrow comprising $g\tilde{\zeta}$ and $\tilde{\zeta}$ are in the same orbit class.
2. *Claim:* If ζ and ζ' are representatives of the same orbit class, then $\mathbf{E}([\zeta]) = \mathbf{E}([\zeta'])$.
Proof: We can either make the same choice of lifted morphism in $\mathbf{Flo}_{\tilde{\Sigma}}[X]$, or note that by the argument in the proof of Claim (1),

two different choices in the same orbit map to the same morphism under $\tilde{\mathbf{P}}$.

3. *Claim:* The functor $\mathbf{E}$ is independent of choice of $\tilde{\Sigma}$ -equivalence class of $\tilde{\zeta}$.

Proof: Suppose first that $\tilde{\zeta}$ and $\tilde{\zeta}'$ are horizontally equivalent. By definition, they differ by an identity morphism, which maps to an identity morphism under $\tilde{\mathbf{P}}$. Hence, $\tilde{\mathbf{P}}(\tilde{\zeta})$ is horizontally equivalent to $\tilde{\mathbf{P}}(\tilde{\zeta}')$ in $\mathbf{Flo}_{\Sigma}[Y]$.

Suppose instead that $\tilde{\zeta}$ and $\tilde{\zeta}'$ are vertically equivalent, i.e. they fit into a ladder diagram, which is mapped by $\tilde{\mathbf{P}}$ to a diagram of the same form because $\tilde{\mathbf{P}}$ is an LP-functor (Lemma 6.4). So $\tilde{\mathbf{P}}(\tilde{\zeta})$ and $\tilde{\mathbf{P}}(\tilde{\zeta}')$ are vertically equivalent in $\mathbf{Flo}_{\Sigma}[Y]$.

4. *Claim:* The functor $\mathbf{E}$ is an isomorphism.

Proof: We prove this by showing that $\mathbf{E}$ is a bijection on the sets of objects, and $\mathbf{E}$ is an order-preserving bijection on each hom-set. The bijection on objects follows immediately since $\mathbf{E}$ is the identity on objects. $\mathbf{E}$ is injective on morphisms because:

$$\begin{aligned} \mathbf{E}([\zeta]) = \mathbf{E}([\zeta']) &\implies \tilde{\mathbf{P}}(\zeta) = \tilde{\mathbf{P}}(\zeta') \\ &\implies \tilde{\zeta} \text{ and } \tilde{\zeta}' \text{ are in the same orbit class} \\ &\implies [\zeta] = [\zeta']. \end{aligned}$$

Surjectivity follows from the fact that both $\mathbf{P}$ and $\tilde{\mathbf{P}}$ are surjective. Similarly, both functors are order-preserving, so $\mathbf{E}$ is also order-preserving.

Hence, the categories $\mathbf{Flo}_{\Sigma}[Y]$ and $\mathbf{Flo}_{\tilde{\Sigma}}[X]/G$ are isomorphic. $\square$

Proposition 6.5 tells us that finding an acyclic partial matching on the quotient and passing to the discrete flow category is the same (up to isomorphism) as finding an equivariant acyclic partial matching in the original space and taking the quotient of this larger flow category.

This gives us the tools to relate the Morse complex of groups over $\mathbf{Flo}_{\Sigma}[Y]$ to a complex of groups over the quotient $\mathbf{Flo}_{\tilde{\Sigma}}[X]/G$. In particular, we will use Lemma 3.15 to find a complex of groups over $\mathbf{Flo}_{\tilde{\Sigma}}[X]/G$, associated to the action, so that its development is isomorphic to that of $\mathbf{M}$.

Let λ denote the choice of lift function made to construct $\mathbf{F}$.

Proposition 6.6. *Associated to the action on $\mathbf{Flo}_{\tilde{\Sigma}}[X]$ there is a choice of lifts and transfers with associated complex of groups $\mathbf{M}'$ and monomorphism*

Ψ' such that the following diagram commutes,

$$\begin{array}{ccc}
 \mathbf{Flo}_{\Sigma}[X]/G & \xrightarrow{\mathbf{E}} & \mathbf{Flo}_{\Sigma}[Y] \\
 & \searrow \mathbf{M}' & \swarrow \mathbf{M} \\
 & \mathbf{Grp} &
 \end{array}$$

and $\Psi'_{\zeta} = \Psi_{\mathbf{E}\zeta}$ for every morphism $\zeta \in \mathbf{Flo}_{\Sigma}[X]/G$.

Proof. We define a lift function μ for $\mathbf{M}'$: for any object $[x]$, set $\mu([x]) := \lambda(\mathbf{E}([x]))$. For any morphism ζ with representative

$$w \xrightarrow{h_0} y_0 \xleftarrow{f_0} x_0 \xrightarrow{h_1} y_1 \xleftarrow{f_1} \dots \xleftarrow{f_k} x_k \xrightarrow{h_{k+1}} z,$$

choose the transfer element $\sigma := \sigma(\mathbf{E}\zeta)$, which is (by definition of the Morse complex of groups),

$$\sigma(\mathbf{E}\zeta) = \sigma(h_{k+1})\sigma(f_k)^{-1} \dots \sigma(h_1)\sigma(f_0)^{-1}\sigma(h_0).$$

We need to check that σ is a valid choice of transfer element, i.e., that the morphism $\lambda(\zeta) : \lambda(w) \rightarrow \sigma^{-1}\lambda(z)$ is the unique $\tilde{\Sigma}$ -zigzag with source $\lambda(w)$ such that $\mathbf{P}(\lambda(\zeta)) = \zeta$.

To simplify notation, write

$$\begin{aligned}
 \tilde{u} &:= \lambda(u), \text{ for any object } u \\
 \alpha_i &:= \sigma(h_0)^{-1}\sigma(f_0)\sigma(h_1)^{-1} \dots \sigma(h_i)^{-1}, \text{ and} \\
 \beta_i &:= \sigma(h_0)^{-1}\sigma(f_0)\sigma(h_1)^{-1} \dots \sigma(h_i)^{-1}\sigma(f_i).
 \end{aligned}$$

Note that $\beta_{k+1} = \sigma^{-1}$. Each β_i is such that $\beta_i \tilde{h}_i : \beta_i \tilde{x}_{i-1} \rightarrow \alpha_{i+1} \tilde{y}_i$ is the unique morphism with source $\beta_i \tilde{x}_{i-1}$ projecting to h_i under $\mathbf{P}$. Similarly, each α_i is such that $\alpha_i \tilde{f}_i^{-1} : \alpha_i \tilde{y}_i \rightarrow \beta_i \tilde{x}_i$ is the unique morphism in the orbit of $\tilde{f}_i^{-1}$ with source $\alpha_i \tilde{y}_i$. This follows from the definition of each transfer element $\sigma(h_i)$ and $\sigma(f_i)$.

Hence, $\tilde{\zeta}$ can be written explicitly as the composition:

$$\tilde{w} \xrightarrow{\tilde{h}_0} \alpha_0 \tilde{y}_0 \xrightarrow{\alpha_0 \tilde{f}_0^{-1}} \beta_0 \tilde{x}_0 \xrightarrow{\beta_0 \tilde{h}_1} \alpha_1 \tilde{y}_1 \xrightarrow{\alpha_1 \tilde{f}_1^{-1}} \dots \xrightarrow{\alpha_k \tilde{f}_k^{-1}} \beta_k \tilde{x}_k \xrightarrow{\beta_k \tilde{h}_{k+1}} \alpha_{k+1} \tilde{z},$$

which is a morphism such that $\mathbf{P}(\tilde{\zeta}) = \zeta$, with source $\lambda(w)$ and target $\sigma^{-1}\lambda(z)$, as required. Each arrow comprising ζ is unique (in X), and so $\tilde{\zeta}$ is the unique morphism in $\mathbf{Flo}_{\Sigma}[X]$ projecting to ζ . The commutativity of the diagram then follows immediately from this construction, because the 1-morphisms, 2-morphisms and twisting elements for $\mathbf{M}'$ and $\mathbf{M}$ are constructed in the same way from the transfer elements, and the maps on objects are the same.

To see that $\Phi_{\zeta} = \Phi_{\mathbf{E}\zeta}$ for any morphism ζ , recall that it is defined as conjugation by the transfer element corresponding to ζ . But this was chosen precisely so that $\sigma(\zeta) = \sigma(\mathbf{E}\zeta)$. $\square$

Remark 6.7. Note that by definition of the induced action on $\mathbf{Flo}_{\tilde{\Sigma}}[X]$, the localisation and inclusion functors $\mathbf{L}_{\tilde{\Sigma}}$ and $\mathbf{J}_{\tilde{\Sigma}}$ are G -equivariant.

Now we have all the tools necessary to prove Theorem 6.1.

Proof of Theorem 6.1. Given $\mathbf{F}$ and Φ , we can construct the Morse complex of groups $\mathbf{M}$ and associated morphism Ψ , as described in Section 5. Since Σ is $\mathbf{F}$ -compatible, we can build an equivariant acyclic partial matching $\tilde{\Sigma}$ on X (Lemma 6.2). Furthermore, the G -action on X pushes to an action on $\mathbf{Flo}_{\tilde{\Sigma}}[X]$ (Lemma 6.3), and the quotient $\mathbf{Flo}_{\tilde{\Sigma}}[X]/G$ is isomorphic to $\mathbf{Flo}_{\Sigma}[Y]$ (Proposition 6.5). There is a choice of lifts and transfers to build the complex of groups $\mathbf{M}'$ and morphism Ψ' associated to the action on $\mathbf{Flo}_{\tilde{\Sigma}}[X]$ making the diagram

$$\begin{array}{ccc} \mathbf{Flo}_{\tilde{\Sigma}}[X]/G & \xrightarrow{E} & \mathbf{Flo}_{\Sigma}[Y] \\ & \searrow \tilde{\mathbf{M}} \quad \swarrow \mathbf{M} & \\ & \mathbf{Grp} & \end{array}$$

commute (Proposition 6.6). Hence, $\mathcal{D}(\mathbf{M}, \Psi) \cong \mathcal{D}(\mathbf{M}', \Psi')$ as categories (Lemma 3.15). By the Basic Construction (Proposition 3.11), $\mathcal{D}(\mathbf{M}', \Psi')$ is G -equivariantly isomorphic to $\mathbf{Flo}_{\tilde{\Sigma}}[X]$, and the latter has classifying space homotopy-equivalent to X (Theorem 4.7). Since the localisation and inclusion functors, $\mathbf{L}_{\tilde{\Sigma}}$ and $\mathbf{J}_{\tilde{\Sigma}}$, are G -equivariant with respect to the induced G -action on $\mathbf{Loc}_{\tilde{\Sigma}}[X]$ and $\mathbf{Flo}_{\tilde{\Sigma}}[X]$ (Remark 6.7), the classifying space $|\Delta \mathcal{D}(\mathbf{M}, \Psi)|$ of the development of the Morse complex of groups and its embedding is G -equivariantly homotopy-equivalent to X , as required. $\square$

7 An Example

We present a detailed example to demonstrate the entire process, from a regular action on a simplicial complex to the higher development of the associated discrete Morse complex about an acyclic partial matching on the orbit space.

Let X be the triangulation of an annulus, as show in Figure 15, equipped with the (regular) left action of

$$D_8 = \langle r, s \mid r^4 = s^2 = e, r^3s = sr \rangle,$$

where r is the anticlockwise rotation by $\pi/2$ and s is a reflection in the y -axis. The orbit space Y in this case is a fundamental domain, *i.e.* a subcomplex of X ; it is a simplicial complex with four 0-simplices $\{v_0, \dots, v_3\}$, five 1-simplices $\{u_0, \dots, u_4\}$ and two 2-simplices $\{w_0, w_1\}$.

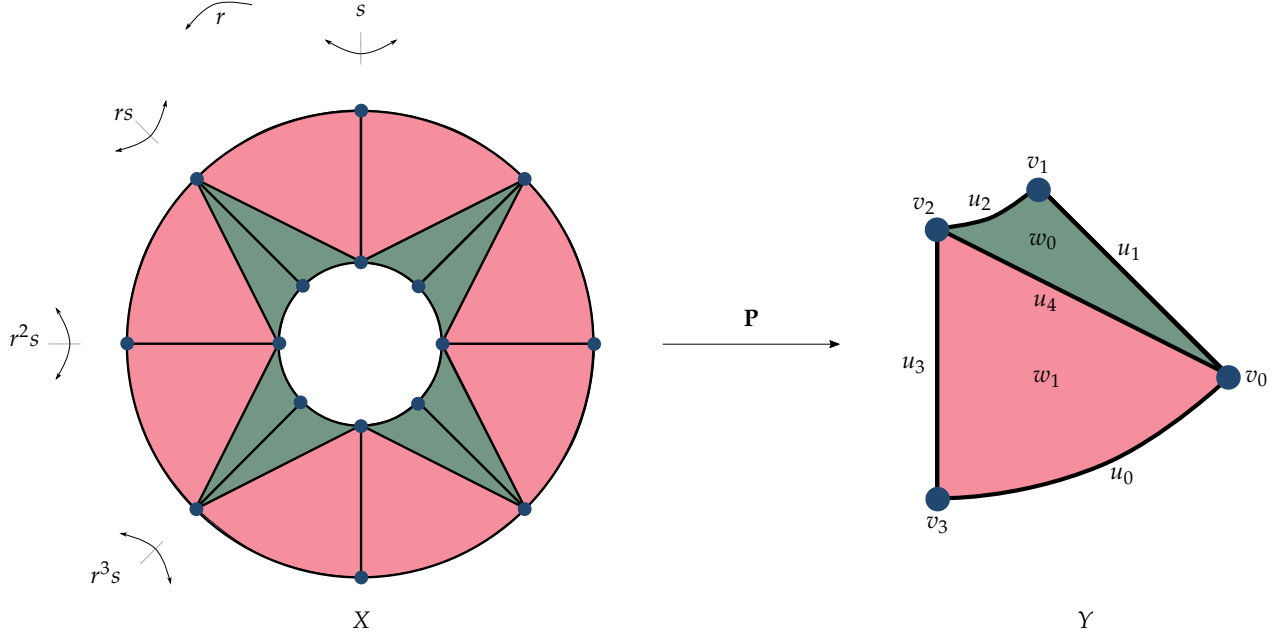
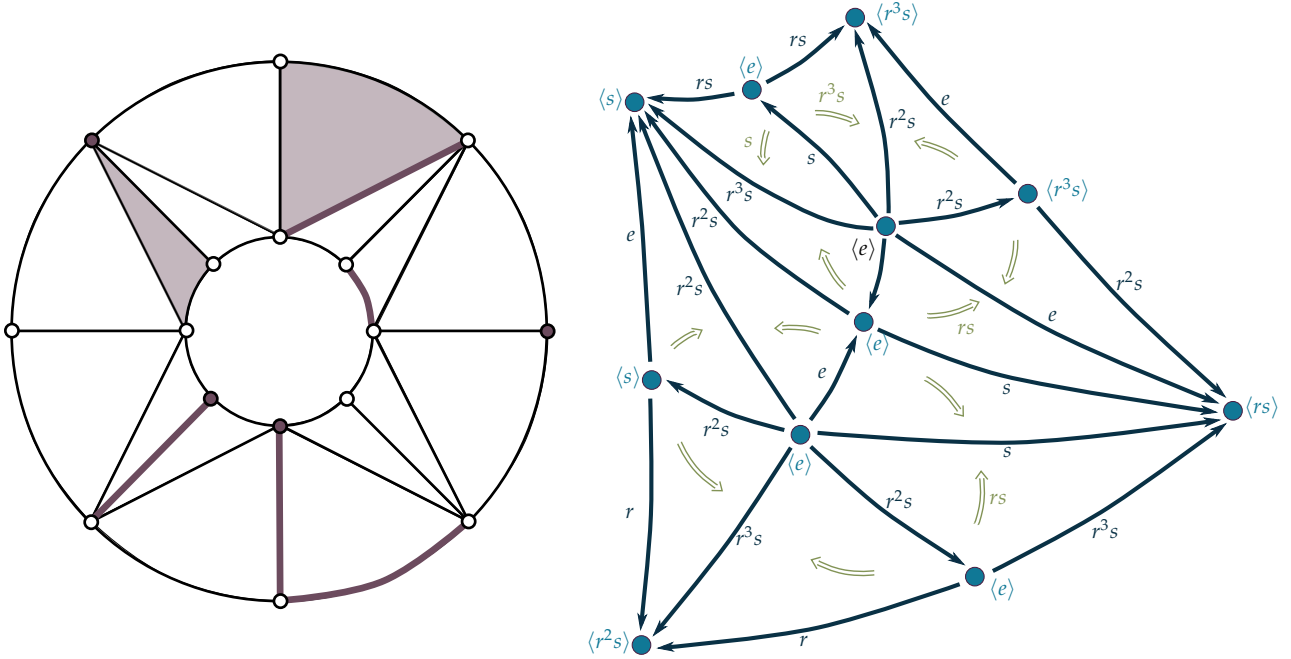


Figure 15: On the left, a triangulation X of the annulus equipped with a regular simplicial (left) group action of $D_8 = \langle r, s \mid r^4 = s^2 = e, r^3s = sr \rangle$, where r is the anticlockwise rotation by $\pi/2$ and s is the reflection in the y -axis. The two orbits of the 2-simplices are coloured in pink and green, respectively. On the right, the orbit space Y , a simplicial complex with four 0-simplices $\{v_0, \dots, v_3\}$, five 1-simplices $\{u_0, \dots, u_4\}$ and two 2-simplices $\{w_0, w_1\}$.

To construct the complex of groups associated to this action, as well as the morphism of complexes of groups, we make a choice of lifts and transfers; for each $y \in Y$ we choose a simplex $\bar{y} \in X$ such that $P(\bar{y}) = y$, and these choices are highlighted on the left of Figure 16. For each pair $x \rightarrow y$ of simplices related by a face relation in Y (i.e. for each pair of simplices related by a morphism in $\mathbf{Fac}[Y]$) we choose a group element $g \in G$ such that $g^{-1} \cdot \bar{y}$ is a face of $\bar{x}$ in X . For example, $\bar{w}_1$ and $\bar{u}_4$ are adjacent in X so their transfer is trivial; however, $r^2s \cdot \bar{u}_1$ is a face of $\bar{w}_0$ (whereas $\bar{u}_1$ is not), so in the complex of groups, the morphism between the stabilisers $\mathbf{F}w_0 \rightarrow \mathbf{F}u_1$ is conjugation by $(r^2s)^{-1} = r^2s$. After making all of these choices, we obtain a complex of groups over Y , as illustrated in Figure 16. The morphism $\Phi : \mathbf{F} \rightarrow \underline{G}$ follows immediately from the choice of $\bar{y}$ for each $y \in Y$ and transfers.

Equipped with a complex of groups and a morphism into $\underline{G}$, we have enough data to use Proposition 3.11 to build a simplicial complex that is G -equivariantly isomorphic to X . However, we are looking to reduce the size of Y by finding an acyclic partial matching on Y before building a complex of groups to reconstruct. We choose the smallest non-trivial matching consisting of a single pairing $\Sigma = \{f : u_4 \rightarrow w_1\}$. This leads to the discrete flow category displayed on the right hand side of Figure 17. The lifts of u_4 and w_1 are adjacent in X , yielding a compatible matching. On the right, the discrete flow category $\mathbf{Flo}_\Sigma[Y]$; there are no objects corresponding to



cells u_4 or w_1 as these are not critical. Each morphism is an equivalence class of zigzags under the relations given in Definition 4.5, with representative zigzags decorating the morphisms above. There is a (non-trivial) poset of morphisms for each of the pairs of objects (w_0, v_2) and (w_0, v_0) .

The discrete flow category is the domain of our Morse complex of groups $\mathbf{M}$, which is constructed in full at the top of Figure 18. At the bottom, we see part of its higher development, whose classifying space is G -equivariantly homotopy equivalent to X .

Figure 16: On the left, a choice of lifts for each simplex in the orbit space Y . On the right, the complex of groups associated to this group action and choice of lifts and transfers. The transfers determine the 1-morphisms, which are decorated on the right by the group element by which the morphism conjugate. Only non-trivial 2-morphisms are labelled.

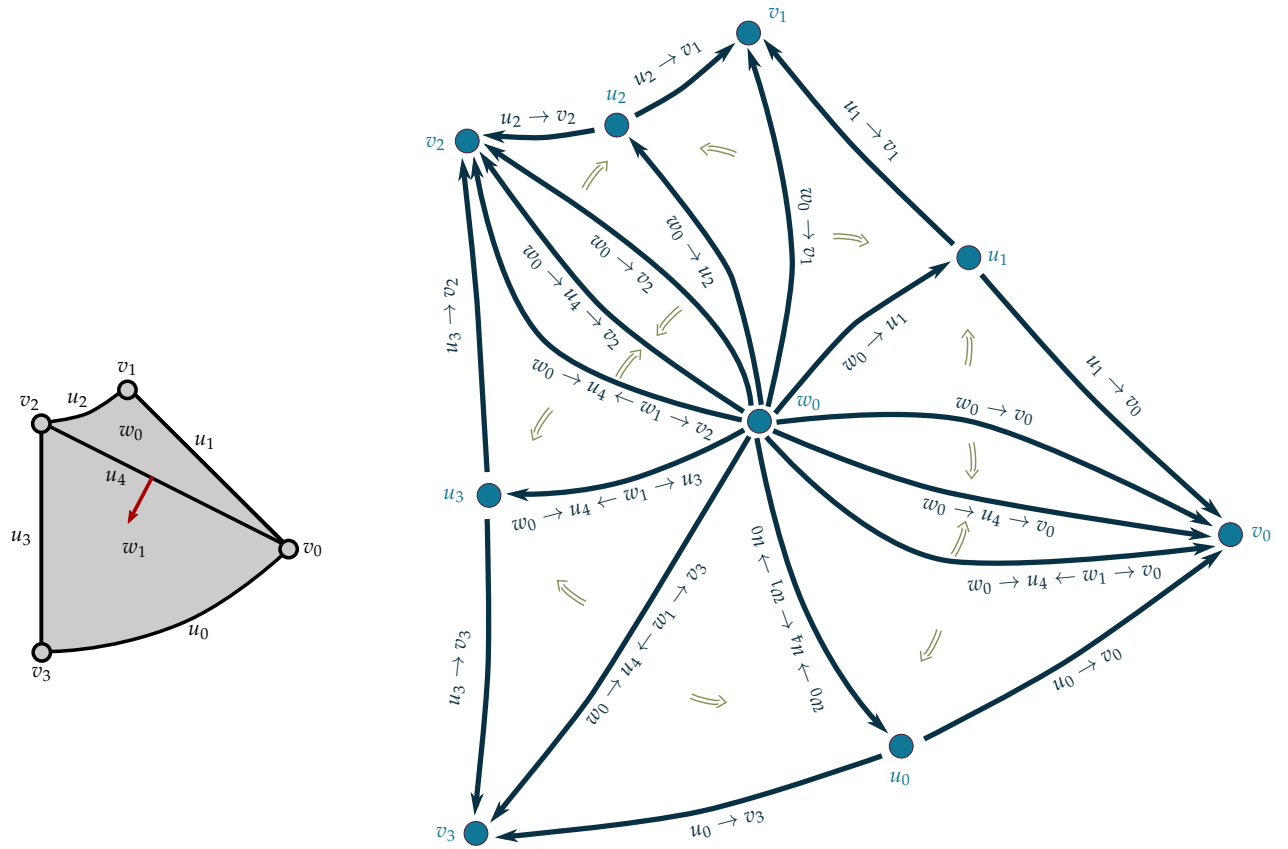


Figure 17: The discrete flow category.

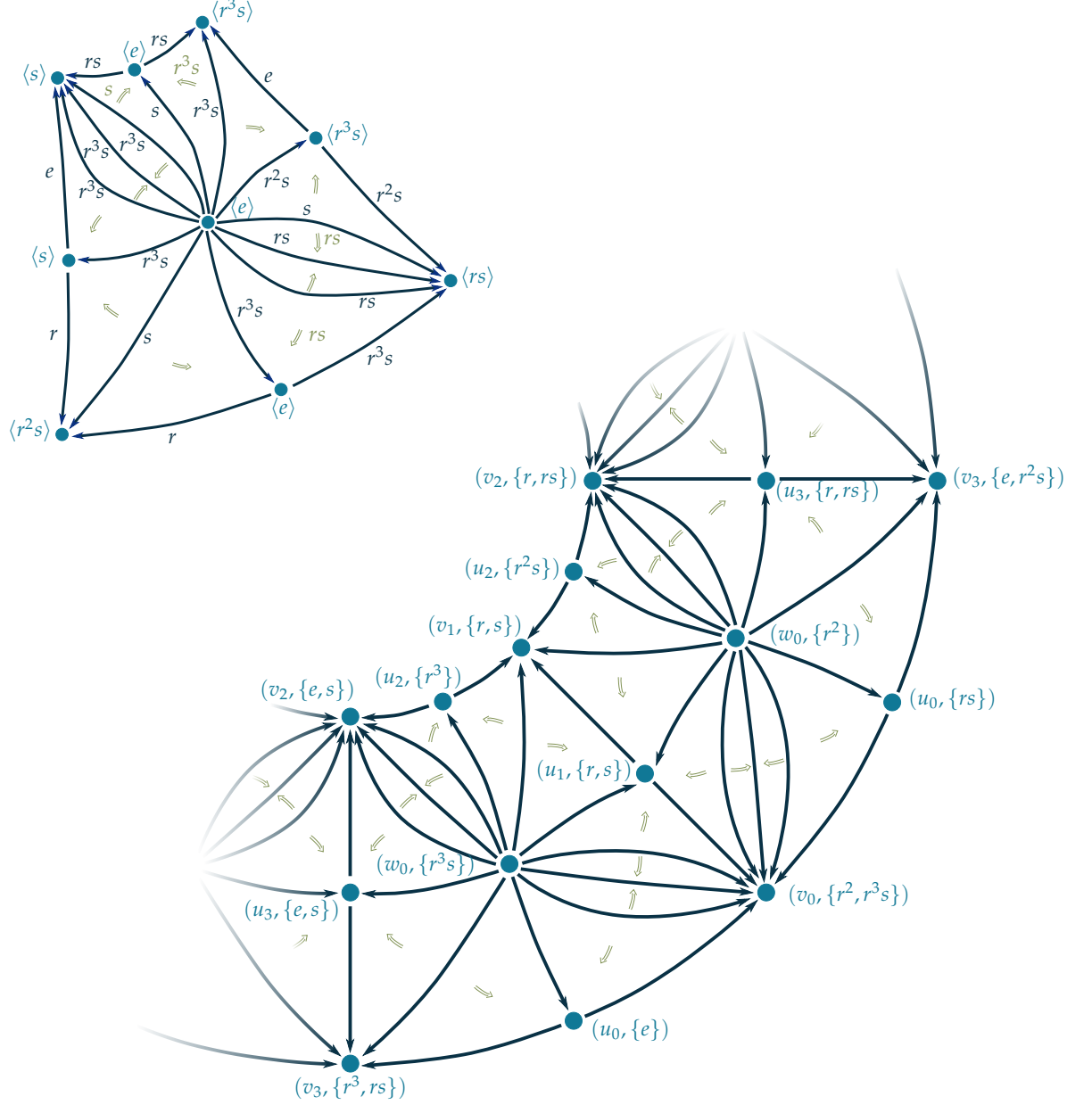


Figure 18: The discrete Morse complex (top) is a complex of groups over the discrete flow category $\mathbf{Flo}_\Sigma[Y]$. All 1-morphisms are conjugation by the group element decorating the arrow. All non-trivial 2-morphisms are labelled. Part of its higher development (bottom). The classifying space of the higher development is G -equivariantly homotopy equivalent to X , i.e. it has the homotopy-type of S^1 .

Part II

A Generalised Clock Theorem

The work in this section is joint with Daniele Celoria. This work has been published ¹³.

Given a graph G embedded in the 2-sphere, denote by G^* its plane dual, and by $\Gamma(G)$ the plane graph obtained by overlaying the two graphs. The vertices of $\Gamma(G)$ are divided into those coming from the vertices of the two original graphs and those arising as the intersection of dual edges. Fix a pair of vertices v and f , one in G and one in G^* . A celebrated result by Kenyon, Propp and Wilson (Kenyon et al., 2000, Theorem 1), known as the KPW correspondence, provides a map between spanning trees in G and perfect matchings on the graph obtained from $\Gamma(G)$ by removing v and f . This map is shown to be a bijection if v and f are “adjacent”, meaning that they are opposite vertices in a square face of $\Gamma(G)$. Otherwise, the map is only injective.

With a somewhat different point of view, in his *Formal Knot Theory* ¹⁴ Kauffman introduced what are now known as *Kauffman states*; these are a special kind of bijections between the set of crossings of a knot diagram and its regions, after two adjacent “forbidden” regions are excluded. In his famous Clock Theorem, he then proved that all Kauffman states are related by a finite sequence of simple swaps, known as *clock moves*. Every knot diagram D can be checkerboard coloured, thus producing two dual graphs, having as vertices the white (respectively black) regions, and crossings as edges.

Loosely speaking, Kauffman states are in bijection with spanning trees in the black graph, as well as with perfect matchings in the (Tait) graph $\Gamma(D)$ given by overlaying the two coloured graphs (see e.g. ¹⁵). In particular, the KPW correspondence reduces to Kauffman’s bijection for the black graphs with the forbidden regions as adjacent ones.

More recently, Cohen ¹⁶ showed how to associate a discrete Morse function –as defined by Forman ¹⁷– to a Kauffman state on a given knot diagram. More precisely, rather than an actual discrete Morse function, his result yields an equivalence class of such objects, which can be thought of as being perfect matchings on the balanced Tait graph (see Section 8 for the definition) of the diagram. One of the aims of this paper is to extend Cohen’s work bridging these graph theoretic properties and ideas stemming from knot theory. We tried to keep the exposition as self contained as possible, as well as accessible to people with graph or knot theoretic backgrounds.

Our starting point is to completely characterise the set of possible dMfs arising from a generalised version of Cohen’s construction

¹³ Celoria and Yerolemou 2021 CELORIA, Daniele ; YEROLEMOU, Naya: A discrete Morse perspective on knot projections and a generalised clock theorem. In: *The Electronic Journal of Combinatorics* (2021)

¹⁴ Kauffman 2006 KAUFFMAN, Louis H.: *Formal knot theory*. Courier Corporation, 2006

¹⁵ Cohen et al. 2014 COHEN, Moshe ; DASBACH, Oliver T. ; RUSSELL, Heather M.: A twisted dimer model for knots. In: *Fundamenta Mathematicae* 225 (2014), S. 57–74

¹⁶ Cohen 2012 COHEN, Moshe: A correspondence between complexes and knots. In: *arXiv preprint arXiv:1211.2553* (2012)

¹⁷ Forman 1998 FORMAN, Robin: Morse Theory for Cell Complexes. In: *Advances in Mathematics* 134 (1998), S. 90–145

(Theorem 9.4). This will allow us (Theorem 9.8) to prove the existence of a bijection between dMfs arising this way and rooted spanning orthogonal forests in the black and white graphs of the diagram, induced by a generalised version of Kauffman states, called *partial Kauffman states*.

We use this bijection to first count perfect dMfs in Proposition 10.1, then –using a symbolic version of the graph Laplacian for the two coloured graphs– we give a formula (Proposition 10.2) to count all dMfs, and compute them for a simple infinite family of diagrams in terms of Fibonacci numbers.

We then turn to perfect admissible matchings on $\Gamma(D)$, where admissible just means that exactly one vertex of each colour is unmatched.

We introduce a set of two moves, called *click path* and *click loop* moves; we first prove (Theorem 11.3) that click path moves induce bijections between the sets of perfect dMfs on S^2 with different critical points. This implies immediately that the image of the KPW correspondence only consists of perfect dMfs, regardless of the adjacency condition; furthermore (Corollary 11.4) every perfect dMf arises uniquely as the image of precisely one choice of spanning tree and one vertex of each colour.

We then prove a topological characterisation (Theorem 11.7) of the subgraphs induced by perfect admissible matchings on the black and white graph, as well as a combinatorial one (Theorem 11.8) in terms of Jordan resolutions.

With these results at hand we can state our main result, which is a simultaneous further generalisation of the Clock theorem and KPW’s correspondence:

Theorem 7.1 (Click-Clock). *If the knot diagram D is reduced, any two perfect admissible matchings on $\Gamma(D)$ are related by a finite sequence of click path, click loop and clock moves.*

One easy consequence of this result is that two perfect dMfs induced by Cohen’s construction can be transformed into one another by clock and click path moves – or only clock moves if they share the same critical points (this last part is the original Clock Theorem).

The proof of the Click-Clock theorem is surprisingly not straightforward, and relies on several seemingly unrelated graph-theoretic constructions. It is also noteworthy to point out that this generalisation of the Clock Theorem is independent from others that have appeared in the literature, such as Hine and Kálmán’s¹⁸ version for triangulated surfaces, and Zibrowius’ extension to tangles¹⁹.

Finally we introduce two simplicial complexes associated to the set

¹⁸ Hine and Kálmán 2018 HINE, Camden ; KÁLMÁN, Tamás: Clock theorems for triangulated surfaces. In: *arXiv preprint arXiv:1808.06091* (2018)

¹⁹ Zibrowius 2016 ZIBROWIUS, Claudius: On a Heegaard Floer theory for tangles. In: *arXiv preprint arXiv:1610.07494* (2016)

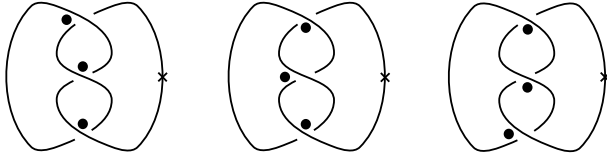
of partial Kauffman states, the *matching* and *discrete Morse* complexes. We study some of the properties of these mysterious complexes, and provide some sample computer aided computations of their homologies.

It remains an open question how to extract more information from these complexes associated to $\Gamma(D)$ and whether this information can be related to interesting features of the knot.

8 Knot Diagrams and Kauffman States

Consider a knot $K \subset S^3$ and a diagram D for K ; as this will not create confusion we will also denote by D the projection of the diagram, *i.e.* the 4-valent graph²⁰ on S^2 obtained by disregarding the over/under information at all of the crossings. Choose an arc a on D —that is, an edge in the projection— and call both a and the two regions of $S^2 \setminus D$ that have a as an edge *forbidden*. The pair (D, a) is usually called a *marked diagram* for K .

Definition 8.1. (Kauffman, 2006) A *Kauffman state* on (D, a) is a bijection between the crossings of D and the non-forbidden regions of $S^2 \setminus D$, such that each crossing c is assigned to one of the (at most 4) non-forbidden regions that are incident to c .



A common way of representing a Kauffman state on (D, a) is to mark the edge a , and specify the assigned region to a given crossing by placing a dot near it.

It was proved by Kauffman in (Kauffman, 2006) that every pair of Kauffman states on (D, a) is related by a finite sequence of *clock moves*, shown in Figure 20.

So, given a marked diagram (D, a) , one can construct the *clock graph* whose vertices are the Kauffman states $X(D, a)$, and edges are given by clock moves (see Figure 22). Clock graphs have been studied previously in (Abe, 2011), (Cohen and Teicher, 2014) and (Cohen et al., 2014), for example. The next objects we introduce have been previously studied in (Huggett et al., 2012) and (Cohen et al., 2014), for example.

Definition 8.2. Let D be a knot projection with a black and white checkerboard colouring²¹. The black graph $G_b(D)$ has black regions

²⁰ To avoid degenerate cases, we will always assume that the diagrams have at least one crossing.

Figure 19: The three Kauffman states on a minimal marked diagram (D, a) of the left trefoil. The forbidden edge a is marked with an asterisk.

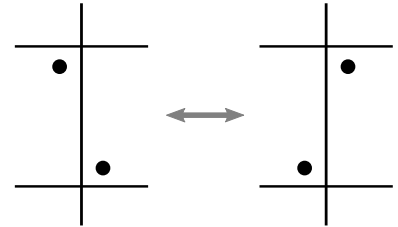


Figure 20: A clock move seen on a local portion of a diagram. The right configuration is obtained from the left one by a clockwise clock move. The two Kauffman states are assumed to coincide everywhere except on the two crossings involved in the clock move.

²¹ Note that this colouring always exists because D is 4-valent.

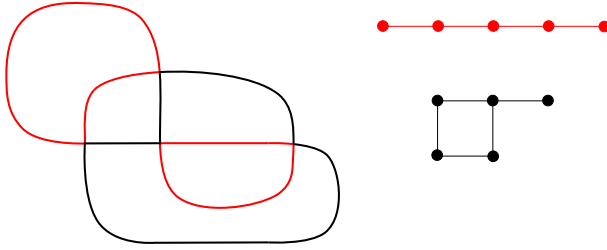


Figure 21: The projection of a minimal diagram of the figure 8 knot. Choosing any of the the red arcs as forbidden results in the red clock graph on the right and similarly for any choice of black arc.

as vertices and crossings as edges (the white graph $G_w(D)$ is defined analogously)²². The *overlaid Tait graph* $\Gamma(D)$ is the superimposition of G_b and G_w , using their natural embeddings in S^2 .

Note that the graphs G_b and G_w are plane duals, $\Gamma(D)$ divides the sphere into square regions, and

$$V(\Gamma(D)) = [V(G_w) \cup V(G_b)] \cup [E(G_w) \cap E(G_b)],$$

and edges in $\Gamma(D)$ are the half edges of the white and black graphs, connecting one of their vertices to a crossing (see e.g. (Kauffman, 2006) for a more detailed exposition).

If one discards the forbidden vertices (one from G_w and one from G_b), we obtain the *balanced overlaid Tait graph* $\Gamma(D, a)$. The vertices of this tri-partite graph can be split into those coming from black or white regions, and those corresponding to crossings.

In the figures, we will draw black/white regions as black/white dots respectively, and vertices coming from crossings as little crosses, as in Figure 22.

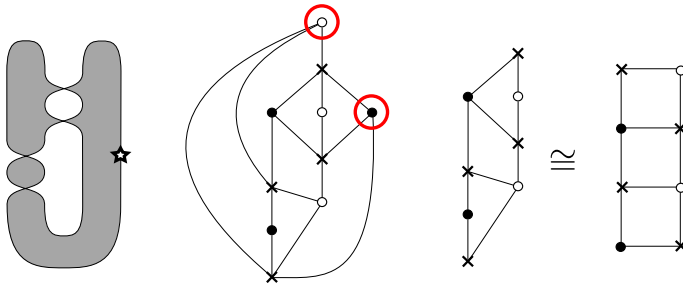


Figure 22: From left to right: a checkerboard coloured projection for a diagram D of the figure eight knot, the associated overlaid (unbalanced) Tait graph $\Gamma(D)$, and the balanced one $\Gamma(D, a)$ given by the choice of the arc a marked with an asterisk. Forbidden regions are circled in red.

Definition 8.3. (Kauffman, 2006) The *Jordan trail* associated to a Kauffman state $x \in X(D, a)$ is the curve formed from x by resolving all crossings as indicated in Figure 23.

We will see later on that Jordan trails associated to Kauffman states are always connected (see also Figure 25). We will expand this definition in Section 9 to include the cases where not all crossings are paired with a region (as in Figure 27).

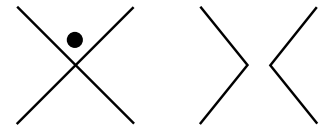


Figure 23: How to resolve a crossing involved in a Kauffman state.

²² We will omit D from the notation when it is clear to which diagram G_b is associated.

We recall here some graph-theoretic terminology that will be widely used throughout the rest of the chapter.

Definition 8.4. A *matching* on a graph G is a subset of edges in G sharing no common vertices. A matching is *perfect* if every vertex in G is incident to an edge in the matching and it is *maximal* if it is not a proper subset of another matching. A *maximum matching* is a matching containing the largest number of edges.

We write $\text{Tree}(G)$ to denote the set of all spanning trees in G . Let H be a subgraph of G , with $E(G) = \{e_1, \dots, e_m\}$ and $E(H) = \{e_i\}_{i \in I}$ for some $I \subseteq \{1, \dots, m\}$. A graph $H^\perp$ in the dual graph G^* *orthogonal to H* is any subgraph of G^* with edges a subset of $\{e_i^*\}_{i \notin I}$, where e_i^* is the unique edge in G^* intersecting e_i in $G \cup G^*$. It is a well-known result that the dual $H^\perp$ of a spanning tree H is itself a spanning tree in the dual graph, and that $H^\perp$ is completely determined by H . A forest is just a disjoint union of trees, and we allow isolated vertices to be connected components of a forest.

By definition, the set of perfect matchings of $\Gamma(D, a)$ is in bijective correspondence with $X(D, a)$ (see e.g. Figure 24), hence we will use these notions interchangeably throughout.

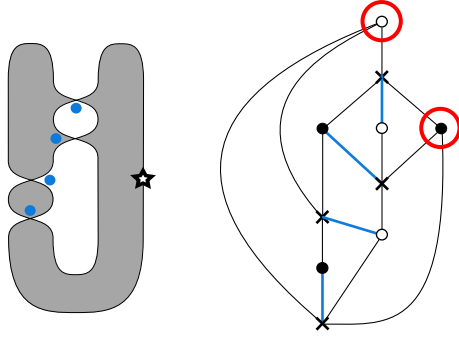
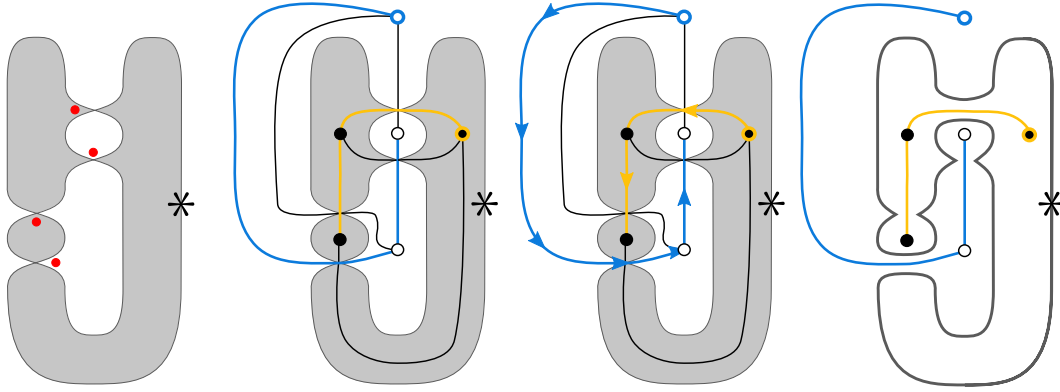


Figure 24: On the left is a Kauffman state on a diagram for the figure eight knot and on the right is the corresponding perfect matching on the balanced Tait graph. The forbidden regions are circled in red.

There is also a bijection between Kauffman states on (D, a) and rooted spanning trees in G_b (or G_w), after choosing a common root for the trees and their duals (see (Kauffman, 2006, Sec. 2)). The choice of these roots is dictated by the arc a : given a spanning tree $T \subset G_b$, take the unique vertex corresponding to a forbidden region as the root of the tree and orient all edges away from the root (see Figure 25). Let $T^* \subset G_w$ be the tree dual to T , and, again, orient the edges to flow away from the uniquely determined root. Each crossing lies at the centre of an oriented edge $u \rightarrow v$, so associate to each crossing the target v of the edge to which it belongs. This will always be a non-forbidden region because the two forbidden regions are, by construction, only ever the source of an edge. There is a bijection between crossings and non-forbidden regions because T and T^* are

dual and orthogonal, and they span all regions. In a similar way, one can construct a spanning tree from a Kauffman state.



We can now state a version of Kauffman's *Clock Theorem* linking Kauffman states, appropriately rooted trees in the black or white graphs and Jordan trails, suited for our purposes.

Theorem 8.5. (*Clock Theorem*) (Kauffman, 2006, Thm. 2.4) *There is a bijection between $X(D, a)$, $\text{Tree}(G_b)$ and Jordan trails of D . Moreover, all elements of $X(D, a)$ can be transformed in one another with a finite sequence of clock moves.*

Until now we only discussed the case where forbidden regions are determined by the choice of an arc on D . However (as also noted by Kauffman in (Kauffman, 2006, Sec. 2)) this hypothesis can be relaxed. Denote by (D, v_b, v_w) a diagram with exactly two forbidden regions, one black and one white, specified by the vertices $v_b \in G_b$ and $v_w \in G_w$, respectively and by $\Gamma(D, v_b, v_w)$ the subgraph of $\Gamma(D)$ with vertices v_b and v_w removed.

In (Kenyon et al., 2000, Thm. 1), Kenyon, Propp and Wilson independently generalise the work by Kauffman to the case where the two unmatched vertices v_b and v_w are not necessarily adjacent. Here by adjacent we mean that the two corresponding regions share an edge in the projection; this is also equivalent to requiring that v_b and v_w are opposite vertices in a square face of the Tait graph $\Gamma(D)$.

Their main result –recalled below– known as the *KPW correspondence* (or construction) is equivalent to the first part of Kauffman's Clock theorem, as stated in Section 8, in the case of adjacent unmatched vertices. If instead the unmatched vertices are not at the opposite sides²³ of a square in $\Gamma(D)$ we only get an injection between the rooted spanning trees of the black graph (with the extra choice of a vertex of the white graph) to perfect matchings in the corresponding balanced overlaid Tait graph.

Figure 25: Spanning trees determine Kauffman states: the Kauffman state on the left is completely determined by the oriented dual spanning trees in G_b and G_w , rooted at the coloured nodes, and displayed in the two central parts of the figure. On the right, the Jordan resolution determined by the Kauffman state and the trees. Note that the Jordan resolution is a neighbourhood of both trees.

²³ In their setting, the matchings obtained are all automatically admissible.

Theorem 8.6. (*Kenyon et al., 2000, Thm. 1*) Fix two vertices $v_b \in V(G_b)$ and $v_w \in V(G_w)$; then if these are adjacent, there is a bijective map from $\text{Tree}(G_b)$ to perfect matchings on $\Gamma(D, v_b, v_w)$. If instead the two vertices are not adjacent, the map is an injection.

The constructions and correspondences outlined here will be generalised in several different directions in the following Sections.

We will first extend the correspondence between perfect dMfs and rooted dual trees to non-maximal dMfs and *orthogonal rooted forests*, then we will see in detail what happens when considering non-adjacent roots in the perfect setting. We will then prove that the image of the KPW correspondence consists only of dMfs, and moreover, the elements in its image are in bijection with the set of perfect dMfs on a cell structure on S^2 induced by the given knot diagram.

What we have covered so far does not seem to be connected to clock moves, and hence to the second part of Kauffman's Clock theorem. We will remedy this in Section 11. There we prove a direct generalisation of his theorem, extended to perfect admissible matchings on $\Gamma(D)$, after introducing two further moves in addition to the usual clock ones.

9 From Knots to Discrete Morse Matchings

In this section we outline how to obtain matchings from a knot diagram and present the necessary conditions for such a matching to be acyclic. This is a generalisation of the construction presented by Cohen in (Cohen, 2012). We provide a proof for Cohen's assertion that a matching obtained from a Kauffman state via this construction is indeed a discrete Morse function in Proposition 9.5.

Note that henceforth when we say *discrete Morse function*, we really mean the equivalence class of all acyclic partial matchings induced by the same discrete Morse function. We write $\mathcal{H}$ for the poset graph. A discrete Morse function is *perfect* if the set of critical cells is minimal.

Remark 9.1. If X is a cell complex and $\Sigma = \{(x, \mu(x))\}_{x \in R}$ an acyclic partial matching on X , then for any pair $Y = (y, \mu(y)) \in \Sigma$, the matching $\Sigma \setminus Y$ is acyclic. This follows from the fact that since Σ is acyclic, there is a partial order $\triangleleft$ on R , defined by $x \triangleleft x'$ if and only if $x < \mu(x')$ and this is still a partial order on $R \setminus Y$.

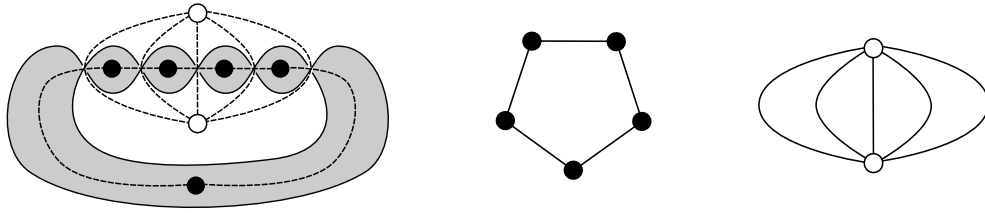
Cohen outlines a correspondence in (Cohen, 2012) and in (Cohen and Teicher, 2014, Sec. 8.2) between pairs $((D, a), x)$ with $x \in X(D, a)$ and discrete Morse functions on the 2-sphere with a cell structure determined by $\Gamma(D, a)$. More explicitly, let us define the map

$$\Xi : \{\text{knot projections}\} \longrightarrow \{\text{cellular decompositions of } S^2\}$$

such that $\Xi(D)$ is the cellular structure on S^2 whose 2-cells are given by the vertices of G_w , 1-cells by crossings in D , and 0-cells by the vertices of G_b .

The unbalanced Tait graph associated to (D, a) is a plane realisation of the abstract poset graph for this cellular structure; more precisely, we also need to orient the edges of $\Gamma(D)$ from white vertices to crossings, and from crossings to black vertices.

As an example (for one choice of the colouring), the minimal diagrams of the alternating torus knots $T_{2n+1,2}$ (see Figure 26) give the splitting of S^2 in two $(2n+1)$ -gons, attached along their boundary. It follows from the definition that $G_b(D)$ provides the 1-skeleton of $\Xi(D)$.



Cohen's main result from (Cohen, 2012) is that if we “extend” Ξ to pairs $((D, a), x)$ with $x \in X(D, a)$, then we get a discrete Morse function μ on S^2 with the cellular decomposition $\Xi(D)$. That is, the matchings on D induced by Kauffman states are mapped to acyclic matchings on the cellular structure $\Xi(D)$ on S^2 . A Kauffman state $x \in X(D, a)$ induces a matching in $\mathcal{H}(D)$ as it is a bijection between crossings and non-forbidden regions (corresponding to crossings and black or white vertices in $\Gamma(D)$). It is not immediately obvious that this induced matching is acyclic.

Cohen's construction induces maximum matchings; note that any perfect discrete Morse function on S^2 has exactly two critical (unmatched) cells: one in dimension 0 and one in dimension 2. The forbidden regions in (D, a) correspond precisely with one 2-cell (the white forbidden region) and one 0-cell (the black forbidden region), with any Kauffman state pairing up all other cells in $\Xi(D)$.

Using discrete Morse functions induced by Kauffman states gives rise to a restricted case of perfect discrete Morse functions, as the only two critical cells are incident to each other.

We define²⁴ *partial Kauffman states* to encompass non-maximum matchings, as well as maximum matchings with non-adjacent critical cells.

Definition 9.2. Consider a knot diagram D with n crossings $\{c_1, \dots, c_n\}$. The set of *partial Kauffman states* $\tilde{X}(D)$ is the set of injections from a subset $\{c_i\}_{i \in I}$ to $V(G_w) \sqcup V(G_b)$, such that each crossing is paired

Figure 26: The minimal projection D_{2n+1} for the alternating torus knots $T_{2,2n+1}$ (for $n = 2$), together with its black and white graphs. The black graph, a cycle of length $2n + 1$, provides the 1-skeleton of the cell structure on S^2 (which can be thought of as the equator). The two white vertices correspond to the two hemispheres, attached along their boundaries.

²⁴ We acknowledge that the term has also been used to describe the generators of certain knot Floer homology complexes (see e.g. Ozsváth and Szabó 2018), but there is no chance of confusion with the ones defined here.

with exactly one of its four adjacent regions, for all subsets $I \subset \{1, \dots, n\}$.

A partial Kauffman state (pKs) on the crossings indexed by I will be often represented by placing one dot –called a *component* of the pKs– near each crossing in $\{c_i\}_{i \in I}$, in one of the four regions incident to it. In the same way, if $v_b \in V(G_b)$ and $v_w \in V(G_w)$, we can define $\tilde{X}(D, v_b, v_w)$ as the set of pKs that do not have any component in the regions v_b and v_w . If, moreover, these two regions are adjacent, we will shorten the notation to $\tilde{X}(D, a)$, where a is the unique arc separating v_b and v_w .

A partial Kauffman state x is *admissible* if there is at least one unpaired region of each colour. x is *maximal* if it injects from the entire set of crossings of the diagram D .

Definition 9.3. The *Jordan resolution* $J(x)$ of a partial Kauffman state $x \in \tilde{X}(D)$ is the curve formed by resolving all dotted crossings as shown in Figure 23. It is the analogue of the Jordan trail of a Kauffman state; one example is displayed in Figure 27. Write $|J(x)|$ to denote the number of connected components of the Jordan resolution, i.e. the number of immersed curves, which we will call *Jordan cycles*.

Since x is not necessarily maximal, $J(x)$ contains double points at all crossings not paired in x . Any connected, complete (i.e. all crossings are resolved) Jordan resolution is just a Jordan trail.

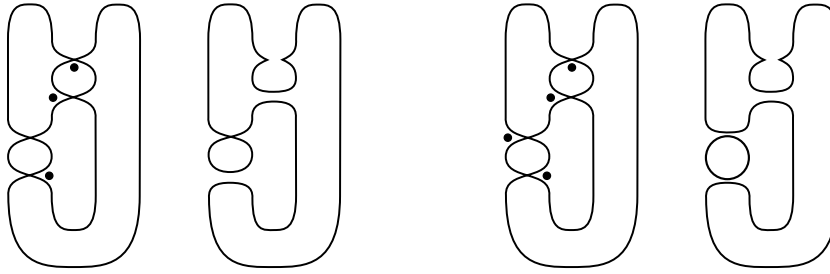


Figure 27: Two partial Kauffman states and their Jordan resolutions; on the left, a connected partial resolution with $|J(x)| = 1$, and on the right, a (disconnected) Jordan resolution with $|J(x)| = 2$. The pKs on the right is perfect but has non-adjacent unmatched regions; further it is not admissible.

We point out that, for a projection with n crossings, there are a priori 2^n possible complete resolutions. However only a small number of these are Jordan resolutions; Figure 28 shows a complete resolution which is not induced by any perfect matching on $\Gamma(D)$.

We say a simple loop γ in $\Gamma(D)$ is *monochromatic* if its vertices consist only of crossings and black regions or only of crossings and white regions. We say that a matching $x \in \tilde{X}(D)$ supports the monochromatic loop γ , if the edges of γ are alternatively composed by edges of x . One example is shown in Figure 29.

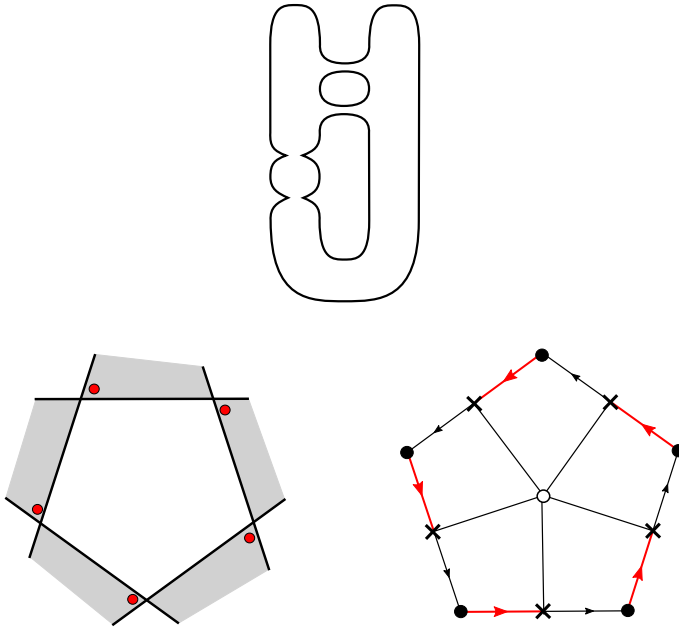


Figure 28: A complete resolution not induced by a perfect matching. In particular, there is no pKs inducing this resolution. Note that the circles are not concentric on S^2 (cf. Proposition 11.8).

Figure 29: An isolated critical point can give rise to partial matchings not inducing discrete Morse functions. The central white vertex is isolated because all adjacent crossings are matched with other vertices. The red dots/edges represent a partial Kauffman state supporting the black loop surrounding the white 2-cell.

The next result is the main tool for checking whether a matching induced by a partial Kauffman state is a discrete Morse function or not.

Theorem 9.4. *A partial Kauffman state x on $\Gamma(D)$ induces a discrete Morse function on $\Xi(D)$ if and only if x does not support a monochromatic loop.*

Proof. It is easy to see that if x supports a monochromatic loop, then its components on the loop give rise to an oriented cycle in the poset/Tait graph $\Gamma(D)$, as in Figure 29. Conversely, if x does not support any monochromatic loop, then all possible oriented loops in $\Gamma(D)$ must contain vertices of both colours. Hence, in each directed cycle in $\Gamma(D)$ we can find a portion of $\Gamma(D)$ that looks like one of the two configurations in Figure 30.

The top configuration can occur if and only if both vertices have been matched with the central crossing, as both arrows have been reversed, which contradicts the definition of a partial Kauffman state. The bottom configuration can occur in a directed cycle if and only if there is (eventually) a section of the cycle containing the top configuration (as there would have to be an arrow from a black vertex to a crossing and another from the crossing back to the white vertex on the left). Hence, there are no such directed cycles. $\square$

The next result is an extended version of the the construction given by Cohen in ²⁵.

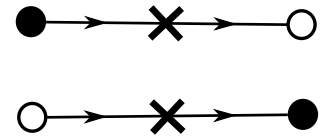


Figure 30: Any directed cycle (induced by a partial Kauffman state) in the Tait diagram containing vertices of both colours must contain these configurations.

²⁵ Cohen 2012 COHEN, Moshe: A correspondence between complexes and knots. In: *arXiv preprint arXiv:1211.2553* (2012)

Proposition 9.5. *Assume x is a maximal pKs on D . If the unpaired white and black vertices are two vertices of a square in $\Gamma(D)$, i.e. the two corresponding regions are adjacent in the diagram, then x induces a perfect discrete Morse function on $\Xi(D)$.*

Proof. By Theorem 9.4, it is sufficient to show that a maximal partial Kauffman state x with adjacent unpaired vertices does not admit a monochromatic loop on $\Gamma(D, a)$, where a is the unique arc common to the two unpaired regions. Assume, for a contradiction, that x contains a monochromatic loop γ . The loop divides S^2 , and consequently the vertices of $\Gamma(D, a)$, into three parts; those belonging to γ , and two other connected components R and R' . We can assume without loss of generality that the two forbidden regions are contained in R , as they are adjacent and so cannot be separated by γ . The thesis will then follow from the Lemma below.

Lemma 9.6. *Let x be a perfect and admissible pKs on $\Gamma(D)$, supporting a monochromatic loop γ . Then, both regions in the complement of γ contain exactly one unpaired vertex.*

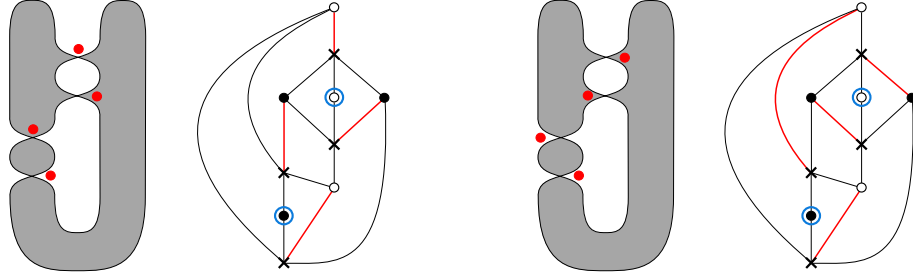
Proof. Again, call the two regions R and R' , and assume wlog that γ is black and has length $2n$, meaning that there are n black regions and n crossings as vertices. Consider the cellular structure on R induced by the restriction G'_b of the black graph to R . Since R is a disc, we have $1 = V(G'_b) - E(G'_b) + F(G'_b)$, where $F(G'_b) = V(G'_w)$ since the graphs are dual.

We can write $V(G'_b) = n + b^\circ$, $E(G'_b) = n + c^\circ$, $F(G'_b) = w^\circ$, where b°, c° and w° are the number of black vertices, crossings and white vertices, respectively, in the interior of R . Then, $1 + c^\circ = b^\circ + w^\circ$, and there is at least one unpaired white/black vertex. Since we are assuming that x is maximal, there can be at most one unpaired vertex in each region. $\square$

We can now conclude the Proposition 9.5 by noting that, by assumption, R contains two unpaired vertices. $\square$

Remark 9.7. If a matching admits a monochromatic loop γ , then it follows from the proof of Proposition 9.5 that γ necessarily contains at least one unpaired vertex on both regions it divides S^2 in. It is however easy to exhibit examples of knot diagrams admitting perfect matchings with an arbitrary number of monochromatic loops.

We conclude this section with a result relating discrete Morse functions to spanning forests, which should be thought of as an analogue of (Chari and Joswig, 2005, Prop. 3.1) and (Zibrowius, 2016, Thm. 1.17), and as a generalisation of the already mentioned (see also (Kenyon et al., 2000) and (Kauffman, 2006)) bijection between



Kauffman states and orthogonal pairs of rooted spanning trees in G_b and G_w .

Theorem 9.8. *There is a bijection between discrete Morse functions on $\Xi(D)$ induced by admissible partial Kauffman states and rooted orthogonal spanning forests in G_b and G_w . In particular, when the admissible partial Kauffman states are maximal, this reduces to a bijection between perfect discrete Morse functions and rooted orthogonal spanning trees.*

The latter part of the statement is just a part of Kauffman's Clock Theorem for non-necessarily adjacent forbidden regions. Before we can prove the theorem, we need to introduce a version of the trees induced by a Kauffman state, extended to admissible partial states.

Definition 9.9. Given a partial Kauffman state x inducing a discrete Morse function on $\Xi(D)$ we can associate to it a pair of subgraphs H_x^b, H_x^w of G_b and G_w respectively, in the same way as is done with perfect discrete Morse function: if x matches a 0/1-cell pair, we include in H_x^b the edge of G_b corresponding to the crossing giving the matched 1-cell in $\Xi(D)$. Likewise with 1/2-cell pairs and edges in H_x^w .

Proof of Theorem 9.8. First, the subgraphs H_x^b and H_x^w associated to x are in fact forests as a consequence of Theorem 9.4. Suppose that H_x^b and H_x^w are not orthogonal; then there exists an edge e in H_x^b intersecting an edge e^* in H_x^w . But e and e^* intersect at a unique crossing of D , and for both edges to be included in their respective graphs, the crossing must be matched with one black and one white region, contradicting the injectivity condition of a partial Kauffman state. Hence, H_x^b and H_x^w are orthogonal.

In this setting, critical points of x are given by unmatched vertices and crossings; each connected component (which is necessarily a tree –possibly composed by a single vertex), contains exactly one unmatched region. The root of each connected component in the forests is the unique unmatched vertex in each component.

Conversely, any such pair of spanning orthogonal forests produces a unique equivalence class of dMfs by applying the process in re-

Figure 31: On the left, a maximal matching inducing a dMf for a projection of the projection from Figure 22, with non-adjacent unpaired black and white regions (circled in blue). The red pKs shown on the right is maximal too, but contains two monochromatic loops, one white and one black. Hence, it does not induce any discrete Morse function.

verse; there are no induced directed cycles in $\Gamma(D)$ as the matching is induced by forests, hence it gives an equivalence class of dMfs. The roots are uniquely determined in a similar way to the description in Figure 25.

When we restrict to maximum pKs, there is exactly one unmatched vertex of each colour, corresponding to exactly one root for each forest, hence the forests contain exactly one connected component each. $\square$

We conclude this section with the following result, relating Jordan resolutions to dMfs.

Corollary 9.10. *An admissible partial Kauffman state $x \in \tilde{X}(D)$ induces a discrete Morse function on S^2 (through Cohen's construction) if and only if $J(x)$ is connected.*

Proof. The result follows easily from the characterisation provided of matchings induced by dMfs just provided, together with the fact that monochromatic loops always disconnect $J(x)$. $\square$

10 Counting Discrete Morse Matchings

Theorem 9.8 provides us with enough structure to count the number of discrete Morse functions on $\Xi(D)$, for a given knot diagram D . Figuring out the number of (perfect) dMfs for a given class of simplicial complexes is in general a challenging problem, and the precise number is known, for example, for the complete graph K_n , but unknown for the n -simplex for $n > 3$ (see respectively Sections 3 and 5 of ²⁶).

The Laplacian $L(G)$ of a graph G on n vertices is the $n \times n$ matrix

$$L(G) = D(G) - A(G),$$

where $D(G)$ is the diagonal matrix with vertex degrees on the diagonal and $A(G)$ is the adjacency matrix of G . It is well-known²⁷ that for a connected graph, the Laplacian has exactly one 0 eigenvalue, and that the non-zero eigenvalues $\lambda_2 \leq \dots \leq \lambda_n$ are strictly positive. Denote the characteristic polynomial of $L(G)$ by

$$p_G(t) = t^n + c_1 t^{n-1} + \dots + c_{n-1} t + c_n.$$

Since we are dealing only with connected graphs, we can assume that $c_n = 0$. To stress the dependence of the coefficients on the graph, we will sometimes write $c_i(G)$ instead of just c_i .

Given an integer $0 \leq e \leq |E(G)|$, let $\phi(G, e)$ denote the number of spanning forests in G with exactly e edges and let F_i^e denote a

²⁶ Chari 2000 CHARI, Manoj K.: On discrete Morse functions and combinatorial decompositions. In: *Discrete Mathematics* 217 (2000), Nr. 1-3, S. 101-113

²⁷ See Biggs 1993 for a complete source of related definitions and results.

subforest of G with e edges, for $i = 1, \dots, \phi(G, e)$. For each F_i^e , define

$$\rho(F_i^e) = m_1 \dots m_k,$$

where m_j is the number of vertices in the j -th connected component of F_i^e , so $\rho(F_i^e)$ counts the number of distinct ways of rooting the forest F_i^e . It is a well-known result (Biggs, 1993, Thm. 7.5) that

$$(-1)^e c_e(G) = \sum_{i=1}^{\phi(G,e)} \rho(F_i^e). \quad (9)$$

In other words, the e -th coefficient of p_G is (up to alternating sign) the number of rooted forests in G with e edges. In what follows, $|\det(K)|$ denotes the *knot determinant*, which is defined as the evaluation at -1 of the Alexander polynomial of K ²⁸.

Proposition 10.1. *Let D be a diagram of a knot K and L_b the Laplacian matrix of $G_b(D)$. Let $\lambda_2, \dots, \lambda_n$ be the non-zero eigenvalues of L_b , where $n = |V(G_b)|$. There are exactly*

$$\left(\prod_{i=2}^n \lambda_i \right) \cdot |V(G_w)|$$

perfect discrete Morse functions on $\Xi(D)$. Moreover, if D is a minimal diagram of an alternating knot K , then the expression can be written as

$$|\det(K)| \cdot |V(G_b)| \cdot |V(G_w)|.$$

Proof. By Kirchhoff's Matrix-Tree Theorem, there are

$$\frac{1}{n} \prod_{i=2}^n \lambda_i$$

unrooted spanning trees in G_b . For each such tree, we have $|V(G_b)|$ choices for the root in G_b and $|V(G_w)|$ for the choice of root in G_w , yielding the desired expression.

If D is a diagram of an alternating knot, then $|\det(K)|$ coincides with the number of (unrooted) spanning trees in either G_b or G_w ²⁹. □

In addition to enumerating the possible spanning trees, we can explicitly list them using the *symbolic Laplacian matrix* $L^{symb}(G)$. $L_{i,i}^{symb}$ is the formal sum of the edges incident to the vertex v_i and $L_{i,j}^{symb}$ is the negative formal sum of all edges connecting v_i to v_j , for $i \neq j$.

Let $p_G^{symb}(t, E(G))$ denote the characteristic polynomial

$$\det(L^{symb}(G) - t \cdot \text{Id}),$$

which is in $\mathbb{Z}[t, e_1, \dots, e_m]$, where $E(G) = \{e_1, \dots, e_m\}$. The coefficient of t^k is (up to an alternating sign) the formal sum of all possible

²⁸ Rolfsen 2003a ROLFSEN, D.: *Knots and Links*. AMS Chelsea Pub., 2003 (AMS Chelsea Publishing Series). – ISBN 9780821834367

²⁹ Burde and Zieschang 1985 BURDE, Gerhard ; ZIESCHANG, Heiner: *Knots*. Bd. 5 of de Gruyter Studies in Mathematics. Walter de Gruyter & Co., 1985

spanning forests in G with $m - k$ edges, counted with the possible ways of rooting them. In other words, each monomial in this coefficient consists of a formal product of the edges contained in a single forest, multiplied by the number of possible roots.

Using the symbolic Laplacian for both the black and white graphs, we can count the total number of discrete Morse functions on $\Xi(D)$:

Proposition 10.2. *Let $\#\mathcal{M}(D)$ denote the number of dMfs on $\Xi(D)$. Then we can compute $\#\mathcal{M}(D)$ by considering the product*

$$p_{G_b^*}^{symb}(-1, E(G_b)) \cdot p_{G_b}^{symb}(-1, E(G_b^*)) \in \mathbb{Z}[E(G_b), E(G_b^*)] / \langle e_i \cdot e_i^* \rangle_{i=1, \dots, |E(G_b)|} \quad (10)$$

and evaluating all variables e_i and e_i^* in 1 for $i = 1, \dots, m$ for a minimal representative.

Note that the evaluation of the polynomial in Equation (10) is just the characteristic polynomial of the symbolic Laplacian (up to a multiplication by t , which only affects the result up to a factor of -1) for the disjoint union of the graph G_b and its dual; here by disjoint union we mean that the two are to be thought of as being disjointly embedded in the plane. An alternative to evaluating in this quotient ring is to consider the product $p_{G_b^*}^{symb}(-1, E(G_b)) \cdot p_{G_b}^{symb}(-1, E(G_b))$ in the same m variables, discard all non square-free monomials, and then set all variables $E(G_b)$ equal to 1.

Proof of Proposition 10.2. Let us start by noting that the coefficients of the characteristic polynomial of the Laplacian of a graph alternate in sign, by Equation 9, so evaluating in $t = -1$ just gives the sum of the absolute values of the coefficients; moreover, by Theorem 9.8 the number of dMfs on $\Xi(D)$ coincides with number of rooted orthogonal spanning forests in G_b and its dual. Using the symbolic Laplacian we see that for a plane graph G , the evaluation $p_G^{symb}(-1, E(G)) \in \mathbb{Z}[E(G)]$ is a sum of monomials, each of which gives a rooted spanning forests in G , and whose coefficient is the number of possible rootings. In particular $p_G(-1, 1, \dots, 1)$, is the number of rooted spanning forests in G .

Now, the product in Equation (10) gives a sum of monomials in the variables $E(G_b)$ and $E(G_b^*)$. The condition $e_i \cdot e_i^* = 0$ is equivalent to the orthogonality of the forests. $\square$

Example 10.3. Using Propositions 10.1 and 10.2, we can count the total number of dMfs for a simple infinite family of knot projections, obtained from the minimal diagrams D_{2n+1} for the alternating torus knots $T_{2,2n+1}$, as in Figure 26.

Since the knot determinant of $T_{2,2n+1}$ is $2n + 1$ ³⁰, by Proposi-

³⁰ Rolfsen 2003a ROLFSEN, D.: *Knots and Links*. AMS Chelsea Pub., 2003 (AMS Chelsea Publishing Series). – ISBN 9780821834367

tion 9.5 there are $2(2n+1)^2$ perfect discrete Morse functions on $\Xi(D_{2n+1})$.

We can use the symbolic Laplacian as in Proposition 10.2 to generalise from perfect to general discrete Morse functions. The symbolic Laplacian matrices for the white and black graphs are respectively

$$L_w^{symb} = \begin{pmatrix} E & -E \\ -E & E \end{pmatrix}$$

where $E = \sum_{i=1}^{2n+1} e_i$, and

$$L_b^{symb} = \begin{pmatrix} e_1^* + e_{2n+1}^* & -e_1^* & 0 & \dots & 0 & -e_{2n+1}^* \\ -e_1^* & e_1^* + e_2^* & -e_2^* & \dots & 0 & 0 \\ 0 & -e_2^* & e_2^* + e_3^* & \dots & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & 0 & -e_{2n}^* \\ -e_{2n+1}^* & 0 & \dots & 0 & -e_{2n}^* & e_{2n}^* + e_{2n+1}^* \end{pmatrix}$$

The first symbolic characteristic polynomial is easily determined to be $p_{G_w}^{symb} = t(t - 2E)$. So,

$$\begin{aligned} \#\mathcal{M}(D_{2n+1}) &= \left((1 + 2E) \cdot p_{G_b}^{symb}(-1, e_1^*, \dots, e_{2n+1}^*) \right) \Big|_{e_i=1, e_i^*=1} \\ &= p_{G_b}^{symb}(-1, 1, \dots, 1) + 2 \left(\sum_{i=1}^{2n+1} e_i \cdot p_{G_b}^{symb}(-1, e_1^*, \dots, e_{2n+1}^*) \right) \Big|_{e_i=1, e_i^*=1}. \end{aligned} \quad (11)$$

Recall that all the products in Equation (11) are computed in the quotient polynomial ring

$$\mathbb{Z}[E(G_b), E(G_b^*)] / \langle e_i \cdot e_i^* \rangle_{i=1, \dots, |E(G_b)|}.$$

Using Theorems 1 and 2 from ³¹ (which provide expressions for the number of rooted spanning forests in path and cycle graphs in terms of Fibonacci numbers) we can see that

$$p_{G_b}^{symb}(-1, 1, \dots, 1) = \Phi_{4n+1} + \Phi_{4n+3} - 2$$

and

$$p_{G_b}^{symb}(-1, 1, \dots, 0, 1, \dots, 1) = \Phi_{4n+2},$$

where Φ_i is the i -th Fibonacci number. The second expression gives us the first term in Equation (11). To evaluate the second term in Equation (11), first recall that we are counting orthogonal forests in G_b and G_w . Hence, for each e_i in the sum $\sum_{i=1}^{2n+1} e_i$, we can consider the path graph obtained from G_b by deleting the edge e_i^* . Thus, we can use the second expression above, which counts rooted forests in path graphs. That is, the latter evaluation coincides with the evaluation in $(1, \dots, 1)$ of $p_{G_b}^{symb}(-1, e_1^*, \dots, e_{2n+1}^*) \cdot e_i$ in the quotient ring of

³¹ Chebotarev 2008 CHEBOTAREV, Pavel: Spanning forests and the golden ratio. In: *Discrete Applied Mathematics* 156 (2008), Nr. 5, S. 813–821

Equation (10) for some edge $e_i \subset E(G_w)$ (as the polynomial is symmetric, we get the same result for any edge e_i). Putting all together we obtain this aesthetically pleasing result:

$$\begin{aligned} \#\mathcal{M}(D_{2n+1}) &= \Phi_{4n+1} + \Phi_{4n+3} + (4n+2)\Phi_{4n+2} - 2 \\ &= 2\Phi_{4n+1} + (4n+3)\Phi_{4n+2} - 2. \end{aligned}$$

11 A Generalised Clock Theorem

In this section we generalise Kauffman's Clock Theorem to perfect admissible matchings, by introducing two new moves between partial Kauffman states. We also characterise the image of the KPW correspondence in terms of dMfs in Proposition 11.3.

From this point onwards we need an extra assumption on the projection of the knot diagrams considered, namely that they are *reduced*. This just means that the situation in Figure 32 is not allowed; in knot theory, crossings like those shown in Figure 32 are referred to as *nutgatory crossings*. This implies at once that both the black and white graphs are 2-connected, meaning that they can not be disconnected by removing a single vertex.

Definition 11.1. Consider two partial Kauffman states $x, x' \in \tilde{X}(D)$. A *click loop move* from x to x' consists of altering x on a single monochromatic loop on which x and x' disagree such that x' and the modified x agree on the loop. We say that x' and x are *click loop equivalent* if they both support the same monochromatic loops, and they coincide everywhere except on at least one of these loops (where they induce opposite orientations).

So, two click loop equivalent pKs are related by a finite number of click loop moves.

Definition 11.2. Consider two maximal partial Kauffman states $x \in \tilde{X}(D, v_w, v_b)$ and $x' \in \tilde{X}(D, v'_w, v'_b)$; we say that x and x' differ by a *click path move* $\rho \subset E(G_b)$ if x and x' induce the same unrooted tree on G_b , ρ is the unique black path in the branch of the tree determined by connecting the root v'_b to v_b , and the two partial states coincide everywhere except on ρ (see Figure 33). The same can be done if the two critical points differ only in the white graph.

The result below characterises the image of the KPW map as the space of perfect acyclic matchings. In other words, the injection guaranteed by KPW's correspondence becomes a bijection if we restrict the codomain to perfect dMfs.

Proposition 11.3. *Click path moves induce bijections between perfect dMfs on $\Gamma(D, v_w, v_b)$ and perfect dMfs on $\Gamma(D, v'_w, v'_b)$ for any pair of black/white vertices. Moreover, at most two such moves are always sufficient.*

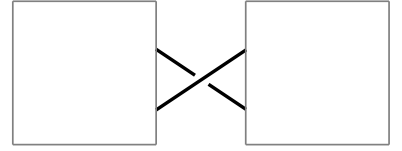


Figure 32: A knot diagram is reduced if it does not contain any such configuration; the two boxes represent portions of the diagram.

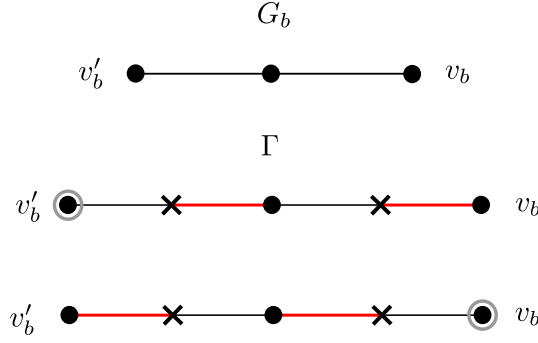


Figure 33: At the top, a path ρ of length 2 connecting the two roots in G_b , and on the bottom, the click path move associated with ρ acting on the two pKs. The two respective roots are circled in grey.

Proof. First of all, let us note that a click path move does not change the maximality of the matching. It then suffices to prove the existence of a bijection for $X(D, v_w, v_b)$ and $X(D, a)$, where a is an arc in D separating the two regions v_w and v'_b . Fix $x \in X(D, v_w, v_b)$, and denote by T_x the black spanning tree it supports. Call ρ the unique path in T_x connecting the root v_b to v'_b . A click path move along ρ transforms x into a matching $x' \in X(D, a)$ (again supporting T_x as a spanning tree). We can then conclude by observing that a click path move does not introduce or remove any monochromatic loops supported by either x or x' (see Figure 34); the only way this would be possible is if a supported loop could be adjacent to a path along which we can apply a click path move. However, we cannot obtain such a configuration without introducing an edge in the matching incident to two crossings. Hence, these moves preserve the acyclicity of maximal pKs by Theorem 9.4.

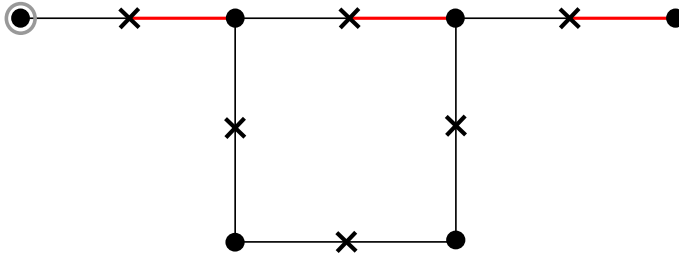


Figure 34: A click path move cannot introduce or remove any monochromatic loops supported by a perfect dMf; an example of the contradiction in the proof of Proposition 11.3.

The general statement then follows easily by noting that a single click path move along a black/white path suffices to change the two unmatched vertices of any perfect dMf on $\Xi(D)$. $\square$

Corollary 11.4. *There is a bijection between the union of all elements in the image of the KPW correspondence over all pairs $(v_b, v_w) \in G_b \times G_w$ and perfect dMfs on $\Xi(D)$.*

Proof. Each perfect matching on $\Gamma(D)$ which is in the image of the KPW correspondence is uniquely determined by the triple (T, v_b, v_w) ,

with $T \in \text{Tree}(G_b)$. By Theorem 9.8, each such triple is in bijective correspondence with a perfect discrete Morse function on $\Xi(D)$. $\square$

Neither click loop moves nor click path moves change the Jordan trails; clock moves and click loop moves do not change the position of the critical points. In fact, when perfect discrete Morse functions differ only by click path moves, we have the following result (cf. also Lemma 11.10).

Lemma 11.5. *Two perfect dMfs on $\Xi(D)$ share the same Jordan trail if and only if they support the same unrooted spanning trees.*

Proof. Let x and x' be two perfect dMfs such that $J(x) = J(x')$. Then at each crossing we must have one of the two possibilities shown in Figure 35, where the dots are, without loss of generality, in the black regions; in both cases we immediately see that the black edge between the two regions in Figure 35 must belong to the subgraph of the black graph supported by the dMfs; by Theorem 9.4 this subgraph can not contain cycles, thus it is a tree.

Conversely, if an edge of the black graph belongs to the common spanning tree, then on the corresponding crossing we are in the situation of Figure 35, and hence the two dMfs have the same Jordan trail. $\square$

Lemma 11.6. *Given a perfect admissible matching x on $\Gamma(D)$, a clock move can either leave $|J(x)|$ unchanged or change it by ± 2 .*

Proof. This follows easily by trying out all possible cases, shown in Figure 36, which shows (up to symmetries) how clock moves change the Jordan resolution. The number in the middle indicates how $|J(x)|$ changes under the clock move, while the dashed line indicates how the local picture eventually joins up. We divide the three possible types depending on their action on the resolution. Type I is the only one needed in the Clock theorem, Type II can only occur for matchings that are not dMfs (as necessarily $|J(x)| > 1$), and Type III moves are the only ones that can merge/split Jordan cycles. $\square$

Proposition 11.7. *If $x \in \tilde{X}(D)$ is a perfect and admissible matching, then each connected component of the black and white orthogonal subgraphs $H_x^b \subset G_b$ and $H_x^w \subset G_w$ it induces is either a tree or has the homotopy type of a circle. Furthermore, there is exactly one tree of each colour.*

Proof. Call H such a connected component, and assume without loss of generality that H is a subgraph of the black graph. Then $H \sim \bigvee^m S^1$ for some $m \geq 0$. Now, the Euler characteristic tells us that the difference between the number of black vertices and crossings is $1 - m$.

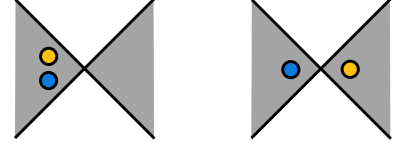


Figure 35: The two possibilities at each crossing: the two matchings either coincide or are on opposite sides of the crossing.

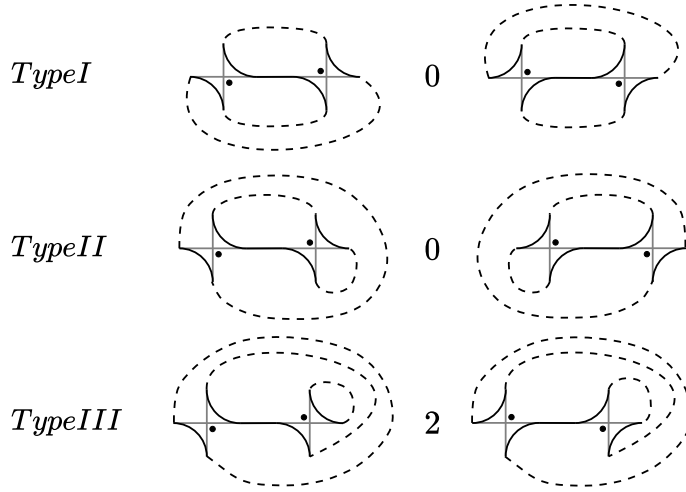


Figure 36: All cases (up to symmetries) of how clock moves change the Jordan resolution.

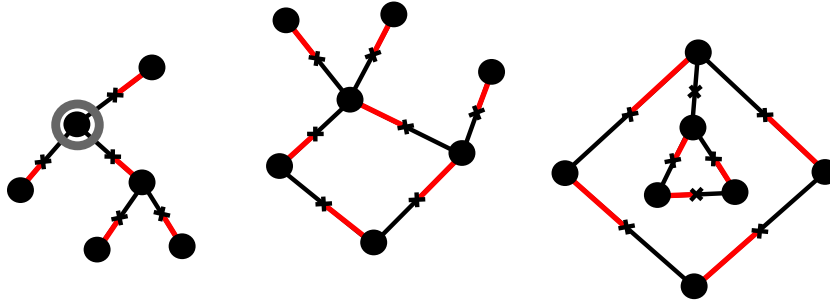


Figure 37: On the left, a tree with its circled root—which is the unpaired vertex; half edges are marked with a cross, and matching between crossings and vertices are highlighted in red. In the middle a component with the homotopy-type of a circle. Here there are no unmatched vertices, which must instead be in the complement in S^2 of this subgraph (cf. Prop. 9.5). The configuration on the right (with rank ≥ 2) cannot have all of its vertices matched.

Hence we can distinguish 3 cases, shown in Figure 37; if $m = 0$, then H is a tree, and there is always a perfect matching on (the first barycentric subdivision of) H leaving out exactly one black vertex. If $m = 1$, then H is homotopically a circle, so it is a simple cycle with some trees “attached” to it. In this case the number of black vertices coincides with the number of crossings, and there is (using the assumption that the matching is perfect) a perfect matching on H (in fact it is possible to prove that there are exactly two) which necessarily contains a monochromatic loop. On the other hand, if $m > 1$ there are more crossings than black dots, so there is no hope for a portion of a perfect admissible³² matching to be supported by H . As trees are the only components that can support an unmatched vertex, the admissibility of the matching implies that there is exactly one of each colour. $\square$

The following result characterises admissibility in terms of the parity of the Jordan resolution and is used in the proof of Theorem 11.9; as a consequence of the proof, we will also see that the subgraphs induced by admissible and perfect matchings are concentric.

³² As the only unmatched vertices can only be black or white, not crossings.

Proposition 11.8. *A perfect matching $x \in \tilde{X}(D)$ is admissible if and only if $|J(x)|$ is odd.*

Proof. It follows from the proof of Lemma 9.6 that each monochromatic loop supported by x has exactly one unpaired vertex in both discs into which it divides S^2 . Furthermore, by Proposition 11.7 each connected component of the black and white subgraphs H_x^b and H_x^w can contain at most one monochromatic loop.

Then, by Lemma 9.6, each connected component of $J(x)$ must have an unmatched vertex on both sides; here we are using the fact that monochromatic loops create connected components in the Jordan resolution. Since x is perfect, there are only two unmatched vertices in $\Gamma(D)$; this forces all Jordan cycles in $J(x)$ to be concentric on the 2-sphere (see Figure 38).

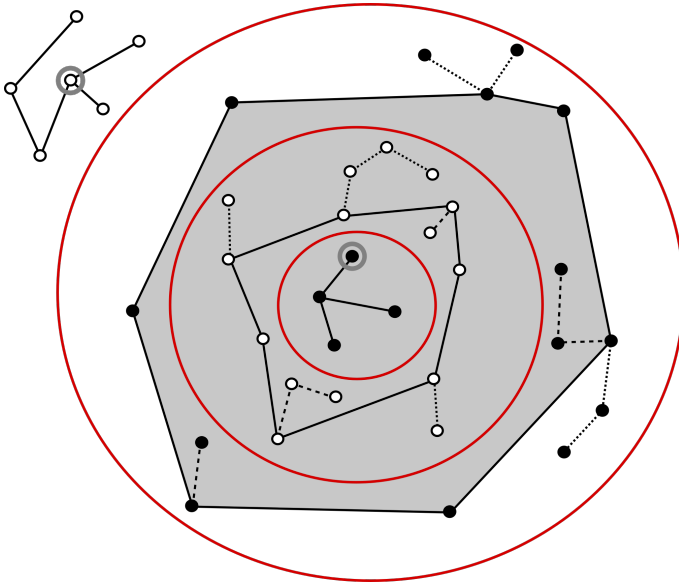


Figure 38: A schematic view of a generic admissible matching up to isotopy on S^2 ; the white and black graphs here are the subgraphs of G_w and G_b , respectively, induced by a maximal pKs. In the central disc, there is the isolated black rooted tree (roots are circled), while the white one is on the outside. The other components are the pseudotrees, and external trees are dotted while internal ones dashed. The only unmatched vertices are the roots of the two trees. The red circles represent the 3 Jordan cycles. The grey shaded region contains the innermost black cycle, its internal trees and the unique isolated black tree.

Connect the two unpaired vertices with a path in S^2 that avoids the crossings of D . Each time the path crosses an arc in the projection it changes colour. Then, note that the parity of the number of the intersections coincides with the parity of the number of circles in any possible Jordan resolution having those two vertices as the unmatched ones. $\square$

As a consequence of this last result, we see that for a given perfect admissible matching x , $|J(x)| - 1$ coincides with the number of monochromatic loops in x .

It follows from the last two results that the subgraph H_x^b of the black graph induced by a perfect admissible matching is a special case of a spanning *pseudoforest*. By this we mean that each connected

component is either a tree or has the homotopy type of a circle (it is a *pseudotree*). Up to isotopy of S^2 , the unique black tree can be thought of as being in the centre of all the concentric pseudotrees. We call this tree *isolated*; all other components of the black pseudoforest can be decomposed into three parts: a *cycle* (i.e. a single simple closed loop in G_b induced by a monochromatic loop), the *internal* trees and the *external* trees, each of which is attached to the cycle at exactly one vertex. We can distinguish between internal and external trees, according to whether they are respectively in the disc bounded by the cycle that contains the isolated tree of the same colour or not. The schematic picture is summarised in Figure 38.

We recall here the statement of the simultaneous generalisation of the Clock Theorem by Kauffman (Kauffman, 2006, Thm. 2.5) to perfect matchings and of the main result in (Kenyon et al., 2000, Thm. 1). In particular, Theorem 11.9 gives a way to organise all possible discrete Morse functions on a given cellular structure on S^2 in the image of Ξ , where each pair of discrete Morse functions is related by a finite sequence of moves.

Theorem 11.9 (Click-Clock). *If the diagram D is reduced, any two perfect admissible matchings on $\Gamma(D)$ are related by a finite sequence of click path, click loop and clock moves.*

As an easy consequence of the proof, two perfect dMfs can be transformed into one another by clock and click path moves – or only clock moves of Type I if they share the same critical points (this last part is Kauffman’s original Clock Theorem); furthermore, if these two dMfs share the same Jordan trail, then they differ by at most two click path moves.

Theorem 11.9 will follow from the next two lemmas; the proof of the first one is a straightforward generalisation of Lemma 11.5 to perfect admissible matchings. The proof of the second one is instead rather involved, and will occupy most of the remaining section.

Lemma 11.10. *Two perfect admissible matchings on $\Xi(D)$ have the same Jordan resolution if and only if they are related by a finite sequence of click path and loop moves.*

Proof. The “if” direction is trivial, as both kinds of moves preserve the Jordan resolution.

For the other implication, let x and x' denote two perfect admissible matchings satisfying $J(x) = J(x')$. Observe that $J(x) = J(x')$ if and only if all of the crossings in the corresponding diagrams resolve in the same way. This gives two possibilities, illustrated in Figure 35; the two matchings can either coincide at a crossing or be on opposite sides (on the two incident regions with the same colour).

If $x = x'$ then the statement follows immediately. If not, then there must be at least one crossing c on which the second possibility occurs. Without loss of generality, suppose we start at such a crossing, where the components of the two states are in the black regions. Draw an edge in G_b if the crossing through which the edge passes is as in the right of Figure 35. Starting from c , draw as many such edges as possible in both directions. If this path forms a loop then the two states are related by a click loop move. If not, then there is a (unique) path between the two unmatched black regions, one from x and one from x' , in which case the two states are related by a click path move. To see that this is actually a path, observe that by construction we never have a vertex of degree three or more, because this would imply the existence of at least two components in a single region from the same matching, which violates the definition of a partial Kauffman state. Since the matchings are admissible, they each have exactly one unmatched black region, hence this path must be between the two unmatched regions in the different matchings with the same colour. $\square$

Lemma 11.11. *Up to clock moves and click loop and path moves, the Jordan trail of a perfect admissible matching can be made connected.*

Since the proof of this last lemma is going to take a bit of work, we first show that, together with Lemma 11.10, it does in fact imply the Click-Clock theorem:

Proof of Theorem 11.9. It suffices to show that for a given perfect and admissible matching x on $\Gamma(D)$, we can convert it to a perfect dMf of $\Gamma(D)$ with prescribed adjacent critical points; the thesis would then follow by applying the “classical” Clock Theorem.

Using Lemma 11.11 we can apply a finite sequence of clock and click loop moves on x that make its Jordan trail connected. Call this new matching $\tilde{x}$; by Corollary 9.10 we know that $\tilde{x}$ is in fact a dMf on $\Gamma(D)$, and up to click path moves we can transform it into yet another dMf with prescribed and adjacent critical points, by Proposition 11.3. $\square$

Before we start the proof of Lemma 11.11, we need to introduce several new objects and study some of their properties. In particular we need the following *leaf spin* operation, first considered in ³³. A *leaf* is an edge with at least one incident vertex of degree one.

Definition 11.12. Let G be a plane graph, and $H \subset G$ a subgraph. If ℓ is a leaf in H which is not a leaf in G , define the clockwise (resp. counterclockwise) *spin* of H along ℓ as the subgraph H' of G obtained by

³³ Harary et al. 1983 HARARY, F ; MOKKEN, R ; PLANTHOLT, M: Interpolation theorem for diameters of spanning trees. In: *IEEE transactions on circuits and systems* 30 (1983), Nr. 7, S. 429–432

removing ℓ from H , and adding the first edge encountered by spinning ℓ around its isolated endpoint in a clockwise (resp. counterclockwise) fashion (see Figure 39).

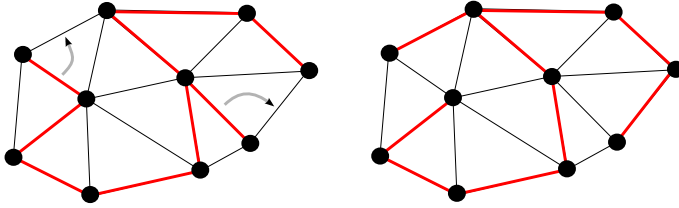


Figure 39: On the right a plane graph with a spanning tree (highlighted in red); on the right the effect of two leaf spins on the tree, one clockwise and one counterclockwise.

If H is a spanning forest (or pseudoforest) then leaf spins preserve this structure as pivoting on a leaf does not change the fact that it is a leaf. Hence, the graph H' obtained from H by a leaf spin is also a spanning forest (or pseudoforest). Furthermore, a leaf spin always leaves the cycles of a pseudoforest unchanged. We need to take a bit more care when dealing with rooted pseudoforests, as this will be the case we need. If a component of the pseudoforest contains the root as the terminal vertex of the leaf we want to spin, we first need to move the root along the unique edge to which it is incident. In other words we never spin a leaf around the root; this way the root does not switch between connected components. As a special case, if the rooted connected component consists of only one edge, spinning it will leave the root as an isolated vertex.

We are about to see how leaf spinning will yield the connection between concentric pseudoforests induced by admissible perfect matchings and certain clock moves on them, ultimately allowing us to conclude the proof of Lemma 11.11 and thus Theorem 11.9. More precisely, consider a pair of spanning pseudoforests induced by admissible perfect matchings, coinciding everywhere outside the first black cycle. We will prove that these are related by leaf spins (and hence by clock and click path moves).

Lemma 11.13. *If the spanning concentric pseudoforest H_x^b induced by the perfect admissible matching $x \in \tilde{X}(D)$ contains a leaf, then all of the pseudoforests obtained by spinning this leaf are induced by matchings that are clock and click path equivalent to x .*

Proof. Call $R \in V(G_b)$ the region corresponding to the terminal black vertex. Let us start by noting that a leaf in H_x^b is necessarily “surrounded” by white edges, as shown in Figure 40. Consider the two cases shown in Figure 40: in the left one we can spin the terminal black leaf in a clockwise fashion, and this corresponds to a sequence of clock moves. After one spin, this local configuration is just the one

we started with, only tilted by the angle of the spin. Hence, we can reach all other possible edges dual to the white ones through clock moves. In the second case we can start clocking/spinning the leaf in both directions, and also reach all other edges in G_b sharing the terminal vertex with the leaf.

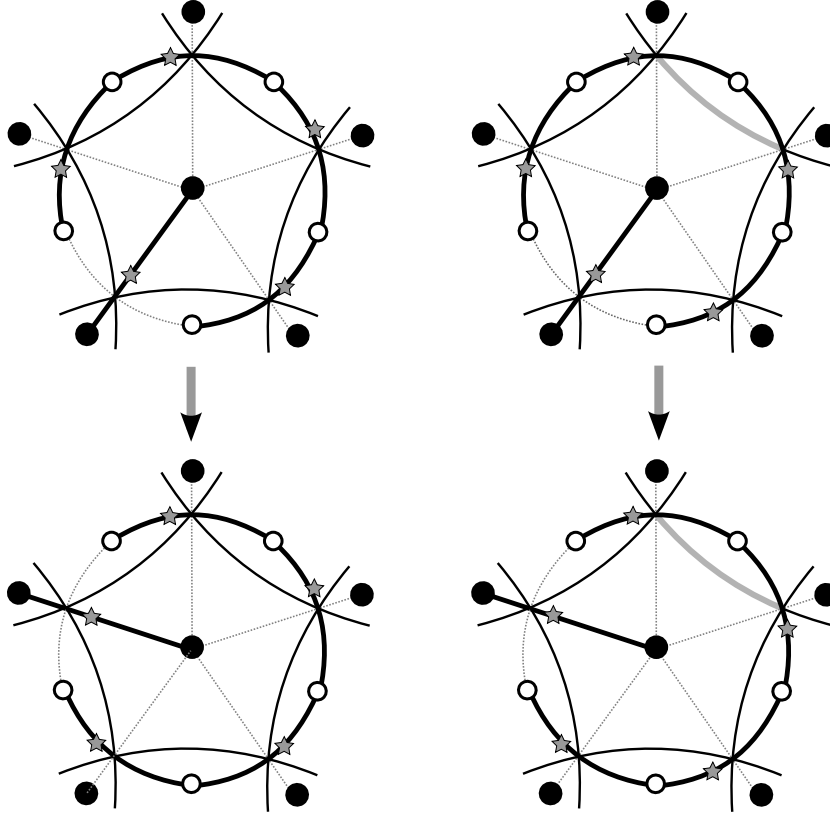


Figure 40: Bold black edges are those included in the subgraphs induced by the matching, and stars mark the dots of the associated pKs. In the top-left configuration, we can only spin the black leaf in a clockwise fashion, and this can be achieved through a sequence of clock moves, the result of which is shown in the bottom-left of the figure. In the top-right configuration we can spin and apply clock moves in both directions; note that we cannot clock clockwise (resp. counterclockwise) after we reach the two edges at the sides of the unique white region (the bold grey arc in the top-right) not matched with a crossing adjacent to the region R . In the bottom-right configuration we can see the result of applying a clockwise leaf spin to the top-right configuration.

It remains to show that these are all of the possible configurations containing leaves; *a priori* there are two possible things that could go wrong. Firstly, there could be no component of the pKs in nearby white regions that allows for a clock move in either direction (and hence the spin could not correspond to a clock move); this can be excluded by looking at the left part of Figure 41.

Lastly, applying multiple leaf spins might result in a configuration where more than one edge of R does not contain any component of the pKs on the crossings it has in common with R . This configuration contradicts the maximality of the pKs and the fact that there is a leaf, as illustrated in the right part of Figure 41.

Note that if any of these leaf spin moves involve the root, then we need to apply a suitable click path move to shift it away.

□

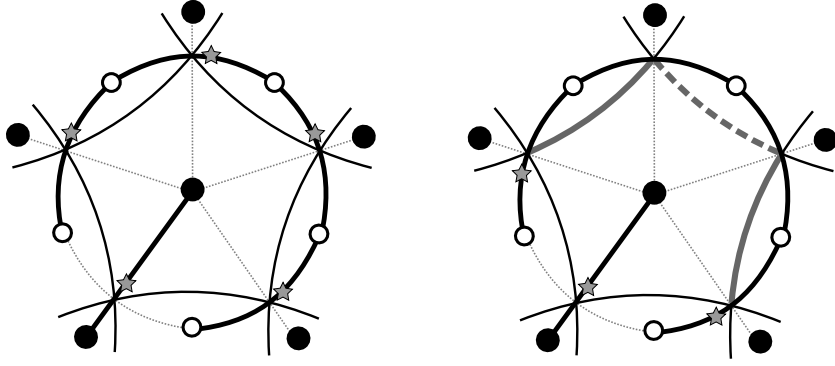


Figure 41: The two impossible configurations from the proof of Lemma 11.13. In the first case the stars near the only matched black edge force an impossible configuration: the two crossings in the top right can not be matched with any region, compatibly with the existence of the black leaf. In the second case, if we assume that the two bold grey arcs have no stars, there is no way of completing the partial Kauffman state in the grey bold and dashed edge (again, if we want to preserve the black leaf).

Remark 11.14. It follows immediately that spinning a leaf always correspond to clock moves of Types I and II (see Figure 36); in particular, a leaf spin can never merge different components of $J(x)$. Furthermore, a leaf spin cannot alter the cycle in a pseudoforest, but might change the cycle of the *dual* pseudoforest.

To complete the proof of Lemma 11.11 by proving the existence of a sequence of moves decreasing the number of Jordan cycles, we must take a detour to introduce some technical tools.

Definition 11.15. An *almost-tree* is the spanning forest obtained by removing exactly one edge from a spanning tree (see Figure 42). If the edge removed is a leaf, then its isolated vertex will be considered as a connected component.

Proposition 11.16. If G is a 2-connected plane graph, denote by $\widehat{G}_{aT}$ the graph whose vertices are almost-trees in G and edges are leaf spins. Then $\widehat{G}_{aT}$ is connected.

Proof. This is an easy consequence of (Harary et al., 1983, Thm. 1). The authors of ³⁴ prove that the related graph G_T , whose vertices are spanning trees and edges are leaf spins, is connected. We can conclude by noting that in $\widehat{G}_{aT}$ there are more possible moves that can be performed, as any almost-tree can arise after the removal of an edge in several trees in G . $\square$

Recall that Theorem 9.8 and Proposition 11.7 imply that for an admissible perfect matching $x \in \widetilde{X}(D)$ we get two induced pseudoforest subgraphs $H_x^w \subset G_w$ and $H_x^b \subset G_b$ composed by concentric pseudotrees alternating in colour.

Consider the subgraph $\overline{G}_b$ of G_b composed of the edges in the innermost black cycle—that is, the unique black pseudotree that surrounds the unique black tree—and all of the vertices and edges within.

³⁴ Harary et al. 1983 HARARY, F ; MOKKEN, R ; PLANTHOLT, M: Interpolation theorem for diameters of spanning trees. In: *IEEE transactions on circuits and systems* 30 (1983), Nr. 7, S. 429–432

In other words, $\overline{G}_b$ is the subgraph of G_b bounded by the unique black cycle, in the disc containing the internal black trees. Let $\overline{H}_x^b$ denote the intersection of $\overline{G}_b$ with the pseudoforest H_x^b . Note that $\overline{H}_x^b$ is spanning in $\overline{G}_b$, and has exactly two connected components: one is a pseudotree (containing only internal trees), and the other is the isolated black tree.

A finite 2-connected graph G embedded in the plane, divides $\mathbb{R}^2$ in regions; we call the edges of G in the boundary of the external infinite region its *boundary cycle*. With the assumptions above, this is in fact a simple cycle in G whose removal from the plane splits it into two discs.

Definition 11.17. For a given finite, leafless, 2-connected and planar graph G embedded in $\mathbb{R}^2$, call $C(G) \subset S^2$ its *boundary contraction*, the graph obtained by collapsing the boundary cycle of G to a point.

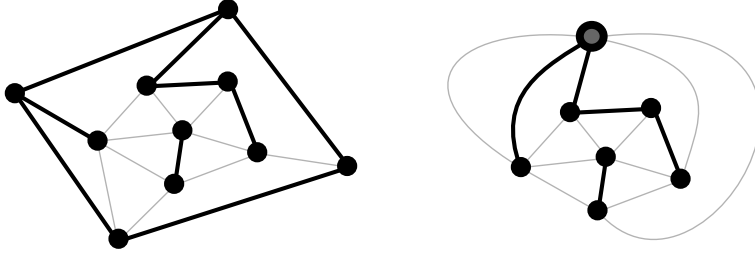


Figure 42: On the left one example of $\overline{G}_b$ graph, with $\overline{H}_x^b$ in bold. On the right the collapsed boundary cycle is the top grey vertex; we can also see how the isolated tree, together with the first black cycle and its internal trees, correspond to an almost-tree in the boundary contraction $C(\overline{G}_b)$.

By taking the boundary contraction we can associate to $\overline{H}_x^b \subset \overline{G}_b$ a spanning almost-tree $C(\overline{H}_x^b)$ in $C(\overline{G}_b)$, with one connected component consisting of the image of all the internal trees, and the other one of the image of the isolated tree.

Proof of Lemma 11.11. Let us denote by x a perfect admissible matching with $|J(x)| = 2k + 1$, since the number of components in the Jordan resolution is necessarily odd by Proposition 11.8. If $k = 0$ there is only one connected component, and we are done.

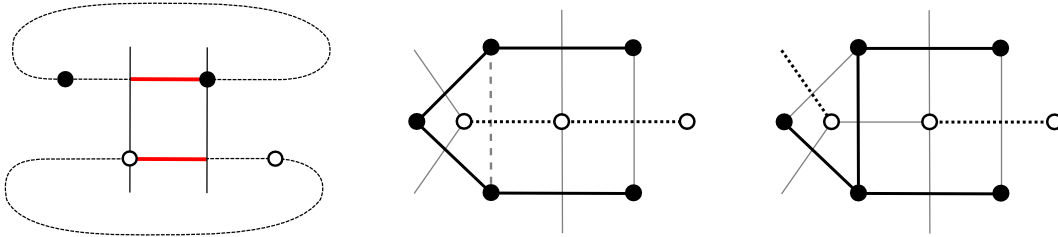
We want to show the existence of a sequence of clock moves and click path/loop moves that take x to a new perfect admissible matching x' with $|J(x')| = 1$. We do this by induction on k . The strategy is to reduce to a configuration akin to the left side of the third case in Figure 36, so that we can apply one final clock move in order to reduce the number of Jordan cycles by 2. The result then follows by the inductive step.

This can be achieved by suitably changing the subgraph $\overline{G}_b$ of the black graph, defined above.

The idea is the following: if we can make the two innermost cycles (i.e. the black cycle and the white cycle contained in H_x^w which

separates the black isolated tree from the black cycle) as close as possible (see the left of Figure 43), then –up to possibly a click loop move on either cycle– we can apply a clock move that merges the innermost Jordan cycles. Applying a click loop move ensures that there is at least one square in $\Gamma(D)$ with opposite edges of a square being matched, which makes applying a clock move possible.

We start by modifying the isolated black tree, by adding to it all the edges belonging the internal trees of the first pseudoforest $\bar{H}_x^b$. We can do this using clock moves and possibly click path moves thanks to Proposition 11.16. That is, we apply a sequence of leaf spins to move the internal trees of $\bar{H}_x^b$ to the isolated black tree. Leaf spins correspond to sequences of clock moves and changing the roots corresponds to click path moves.



Before we can conclude, we need to eliminate some local configurations that would prevent our strategy from succeeding; in particular, after spinning all internal black trees onto the unique isolated black tree as described in the previous paragraph, there remain configurations that prevent the application of a single clock move to reduce the number of Jordan cycles by 2, as illustrated in Figure 36. After applying the procedure described in the previous paragraph, all external white trees are in fact paths. This is because we “spun away” all black internal trees in $\bar{H}_x^b$. A configuration that would prevent applying the desired final Type III clock move is the presence of an external white tree surrounded by the unique black cycle, as illustrated in the centre of Figure 43. The problem here is that these external linear white trees separate the innermost white cycle from the innermost black cycle, preventing reaching the desired configuration (see the left part of Figure 43) where we can apply the final clock move. In particular, this occurs if there exists an edge in $\bar{G}_b \setminus \bar{H}_x^b$ with both vertices on the innermost black cycle (see the grey dashed edge in the central part of Figure 43).

We can overcome this problem by applying a sequence of leaf spins on these external linear white trees, which can be seen in the central and right parts of Figure 43. Note that these leaf spins have the effect of “shrinking” the innermost black cycle.

Figure 43: On the left: two monochromatic loops (dotted) sharing two opposite edges of a square –up to a click loop move on either one– lead to a component-reducing clock move of Type III. This is the desired end-state after applying all of the required leaf spins. In the centre: the local configuration that prevents reaching the desired configuration on the left. In particular, the presence of the dashed grey edge causes the problem. We can solve this problem by applying leaf spins in the white graph; the leaves that have to be spun away in the white graph are dotted. The right part shows the effect on the black cycle of spinning away a white leaf.

Finally, the first black and white cycles are adjacent (as in the left part of Figure 43), and hence we can perform a single clock move to reduce k by one, thus completing the inductive step. $\square$

Remark 11.18. It follows from the proof of Lemma 11.11 that if the perfect admissible matching x has $|J(x)| = 2k + 1$, then we can transform it into a dMf by using at most k clock moves of Type III, k click loop moves, and an unspecified number of clock moves of Type I/II. It is unclear if we can always avoid the use of click path moves.

12 Complexes

In this section we introduce the matching and Morse complexes – the objects encapsulating *all* matchings and acyclic matchings, respectively, associated to a given diagram (including both admissible and non-admissible matchings).

Definition 12.1. The *matching complex* of $\Gamma(D)$ is the simplicial complex $M(D)$ whose vertices are given by edges in $\Gamma(D)$, and n -simplices are matchings with $n + 1$ edges.

Remark 12.2. Alternatively, we can equip the set of partial Kauffman states $\tilde{X}(D)$ with a simplicial structure as follows: $n + 1$ distinct single component partial Kauffman states form an n -simplex in $\tilde{X}(D)$ if and only if their union is still a partial Kauffman state. This is clearly isomorphic to $M(D)$.

Another simplicial complex that can be associated to a diagram in this way is the Morse complex of $\Gamma(D)$, first introduced by Chari and Joswig in ³⁵ for general simplicial complexes.

Definition 12.3. The (*discrete*) *Morse complex* $\mathcal{M}(D)$ of $\Gamma(D)$ (or $\Gamma(D, a)$) is the simplicial complex whose vertices are the edges of $\Gamma(D)$ and whose n -simplices are spanned by $n + 1$ vertices such that the corresponding edges form an acyclic matching in $\Gamma(D)$.

Remark 12.4. By Remark 9.1, this is a well-defined simplicial complex. In fact, it is clear to see that $\mathcal{M}(D)$ is a subcomplex of $M(D)$ (similarly for marked diagrams). Note that the matching complex contains both admissible and non-admissible matchings whereas the discrete Morse complex contains only admissible matchings.

Let $f(D)$ denote the maximal number of arcs in the boundary of a face, maximising over all faces of $S^2 \setminus D$. By (Scoville and Zaremsky, 2020, Thm. 3.7), $\mathcal{M}(D)$ and $M(D)$ are connected – respectively, simply connected – if $4cr(D) \geq 2f(D) + 1$ and $4f(D) + 1$, where $cr(D)$ is the crossing number of D .

³⁵ Chari and Joswig 2005 CHARI, Manoj K. ; JOSWIG, Michael: Complexes of discrete Morse functions. In: *Discrete mathematics* 302 (2005), Nr. 1-3, S. 39–51

It is also possible to give more refined lower bounds on the connectivity of $M(D)$, using (Scoville and Zaremsky, 2020, Prop. 3.1): $M(D)$ is at least $\left(\lfloor \frac{4cr(D)-1}{2f(D)} \rfloor - 1\right)$ -connected. As an example, if D is the diagram in Figure 22, then $M(D)$ is 1-connected. The connectivity of the matching complex is also related to the existence of non-extendable pKs, that is, pKs that are not perfect, yet are not faces of simplices of maximal dimension. It is not hard to exhibit examples of non-extendable pKs of arbitrarily high codimension, as illustrated by Figure 44.

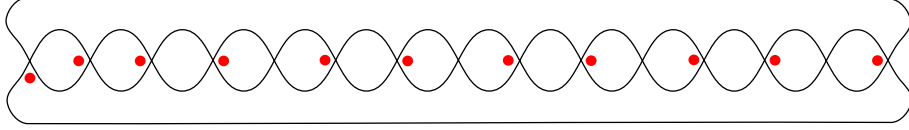


Figure 44: An example of an admissible and non-extendable pKs on the diagram of the $T_{15,2}$ torus knot.

Let $\tilde{X}_{pure}(D)$ denote the set of maximal pKs on the projection D , and use the same subscript for the sub-simplicial complexes of $M(D)$ and $\mathcal{M}(D)$ generated by the top-dimensional simplices.

The pure matching complex of a graph is the union of the pure parts of the matching complexes of all its spanning trees (Ayala et al., 2008, Thm. 8). We can prove an analogous result, generalising from graphs to the cellular structures on S^2 obtained from Cohen's construction. Denote by (T, v_b) a spanning tree $T \subseteq G_b$ rooted at v_b . Recall that a rooted spanning tree in G_b along with a root in G_w uniquely determines a Kauffman state, and hence a perfect acyclic matching in $\Gamma(D)$.

Let $\Delta(T, v_b, v_w)$ denote the simplex in $\mathcal{M}(D)$ spanned by the edges in $\Gamma(D)$ determined by T , v_b and v_w . Of course, $\Delta(T^*, v_w, v_b) = \Delta(T, v_b, v_w)$ since T uniquely determines T^* and vice-versa.

Lemma 12.5. *Let D be a diagram with $n + 1$ crossings. The simplicial complex $\mathcal{M}_{pure}(D)$ of perfect discrete Morse functions on $\Xi(D)$ is generated by the following set of maximal-dimensional simplices*

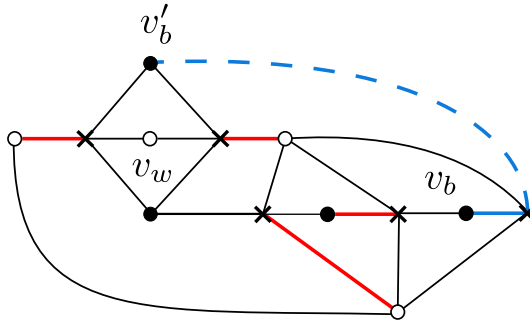
$$\bigcup_{\substack{T \in \text{Tree}(G_b) \\ v_w \in V(G_w) \\ v_b \in V(G_b)}} \Delta(T, v_b, v_w),$$

where each simplex is of dimension n .

Proof. Since each pair (T, v_b, v_w) uniquely identifies a perfect dMf on $\Xi(D)$ (see also e.g. Cor. 11.4), we get a bijection between the pure part of the Morse complex and the set of simplices above. $\square$

Theorem 11.9 gives us insight into the Morse and matching complexes; for example, two perfect dMfs with the same critical cells can

differ only by clock moves. At best, they can differ by exactly one clock move, in which case their corresponding simplices in $\mathcal{M}(D)$ are attached along an $n - 2$ dimensional face, where n is the dimension of the simplices. If two perfect dMfs differ by one click path move along a click path of length 2 (see Figure 45), then the corresponding simplices in $\mathcal{M}(D)$ are attached along an $n - 1$ dimensional face. This is unlike the case where we are studying the Morse and matching complexes of graphs (*i.e.* 1-dimensional simplicial complexes), where maximal-dimensional simplices are connected along unions of $(n - 2)$ -dimensional faces ³⁶.



Whilst it remains a hard problem to fully understand these complexes (even in just the pure case) the three moves between matchings do reveal some structure. We do point out however that complete knowledge of these moves in $\mathcal{M}(D)$ or $M(D)$ is not sufficient to determine their homotopy-type(s), indicating that the complexes themselves carry more information than just the set of moves. To see this more concretely, consider the clock graph, which captures all top-dimensional simplices and moves between them for a fixed pair of adjacent critical cells. This graph is not even enough to determine the homotopy-type of the pure Morse complex for adjacent regions, despite being one of the nicest cases dealt with so far, as can be appreciated with the example in Figure 46.

Before providing some sample computations of the homology of $M(D)$ and $\mathcal{M}(D)$ for small diagrams, we make a few remarks; first we want to highlight a striking difference between dMfs obtained through Cohen's construction, general dMfs and "regular" Morse functions; for concreteness let us restrict to the case of a triangulated smooth closed manifold X .

In this case, the matching complex of (the poset graph of) X can be thought of as a discrete analogue of the space of equivalence classes of gradient vector fields of smooth functions on X , and $\mathcal{M}(\mathcal{H}(X))$ as the subset of these gradient fields coming from functions that are Morse ³⁷. It is well-known that smooth Morse functions are dense

³⁶ Ayala et al. 2008 AYALA, R ; FERNÁNDEZ, LM ; QUINTERO, A ; VILCHES, JA: A note on the pure Morse complex of a graph. In: *Topology and its Applications* 155 (2008), Nr. 17-18, S. 2084–2089

Figure 45: The red edges together with the dashed blue one give a perfect dMf with (v_w, v_b) as forbidden regions, while the red edges with the solid blue one give a perfect dMf with forbidden regions (v_w, v'_b) . The union of the two blue edges gives the click path of length 1 between the two. Since the dMfs differ on only one edge, their simplices in $\mathcal{M}_{\text{pure}}(D)$ are attached along the $(n - 1)$ -dimensional face spanned by the red edges, where $n = 4$ is the dimension of a maximal simplex in $\mathcal{M}(D)$ since any matching consists of at most 5 edges.

³⁷ Benedetti 2016 BENEDETTI, Bruno: Smoothing discrete Morse theory. In: *Annali della Scuola Normale Superiore di Pisa-Classe di scienze* 16 (2016), Nr. 2, S. 335–368

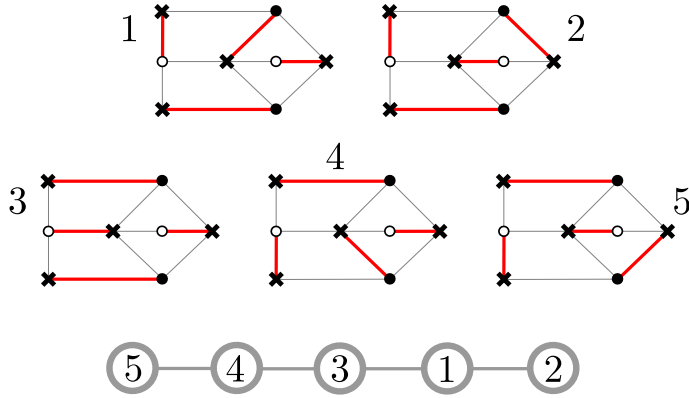


Figure 46: These are the 5 Kauffman states on the projection of the Figure eight knot in Figure 22, with the top-left (red) arc a chosen as the forbidden one. The Clock graph shown below in grey is clearly linear, but it is easy to compute that $\mathcal{M}_{\text{pure}}(D, a)$ is a 3-dimensional simplicial complex homotopic to a circle.

in the set of smooth functions on X , and that “small” isotopies are enough to transform any smooth function on X into a Morse one. In the discrete setting, this is no longer true; given a matching, it is generally very hard (see e.g. ³⁸, the discussion ³⁹, and the end of (Chari and Joswig, 2005, Sec. 2)) to find its minimal “distance” to a perfect dMf.

Of course, given any matching one can always obtain a dMf from it by simply removing one edge from each of its monochromatic loops. However, if we restrict to a perfect and admissible matching x , it is possible to use the Click-Clock theorem to show that the minimal distance between x and a *perfect* dMf is bounded below by the minimal number of clock and click path/loop moves needed to transform x in a dMf, and the proof of the theorem provides us with an almost explicit algorithm that allows us to perform such a transformation.

Finally, it would be interesting to tie some of the homological properties of $M(D)$ and $\mathcal{M}(D)$ to other known graph and knot theoretic invariants and quantities. There is much structure in these simplicial complexes that can be exploited in several ways.

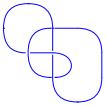
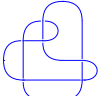
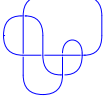
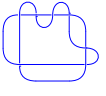
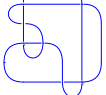
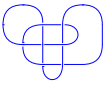
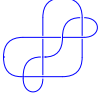
We list here the (reduced) homology of the matching and discrete Morse complexes for all minimal knot diagrams with up to 7 crossings. The computations are performed using a Sage ⁴⁰ program. The table displays the homology of the various simplicial complexes associated to minimal knot projections, up to 7 crossings. The notation $\mathbb{Z}_{(b)}^a$ denotes a generators of the homology in degree b . The diagrams are generated using SnapPy ⁴¹.

³⁸ Nicolaescu 2012 NICOLAESCU, Liviu I.: Combinatorial Morse flows are hard to find. In: *arXiv preprint arXiv:1202.2724* (2012)

³⁹ (<https://mathoverflow.net/users/20302/liviu-nicolaescu>) NICOLAESCU, Liviu N.: Combinatorial Morse functions and random permutations. MathOverflow. <https://mathoverflow.net/q/86193>. – URL:<https://mathoverflow.net/q/86193> (version: 2012-11-19)

⁴⁰ The Sage Developers 2020 THE SAGE DEVELOPERS: *SageMath, the Sage Mathematics Software System (Version 9.0)*, 2020. – <https://www.sagemath.org>

⁴¹ Culler et al. CULLER, Marc ; DUNFIELD, Nathan M. ; GOERNER, Matthias ; WEEKS, Jeffrey R.: *SnapPy, a computer program for studying the geometry and topology of 3-manifolds*. Available at <http://snappy.computop.org>,

Name	Minimal Projection	$\mathcal{M}(\mathbf{D})$	$\mathbf{M}(\mathbf{D})$	$\mathcal{M}_{\text{pure}}(\mathbf{D})$	$\mathbf{M}_{\text{pure}}(\mathbf{D})$
3 ₁		$\mathbb{Z}_{(1)}^4$	$\mathbb{Z}_{(2)}^4$	$\mathbb{Z}_{(1)}^4$	$\mathbb{Z}_{(2)}^4$
4 ₁		$\mathbb{Z}_{(2)}^{12}$	$\mathbb{Z}_{(2)}^5 \oplus \mathbb{Z}_{(3)}$	$\mathbb{Z}_{(2)}^{12}$	$\mathbb{Z}_{(2)}^5 \oplus \mathbb{Z}_{(3)}$
5 ₁		$\mathbb{Z}_{(2)} \oplus \mathbb{Z}_{(3)}^2$	$\mathbb{Z}_{(3)}^{19}$	$\mathbb{Z}_{(2)} \oplus \mathbb{Z}_{(3)}^2$	$\mathbb{Z}_{(3)}^9$
5 ₂		$\mathbb{Z}_{(3)}^6$	$\mathbb{Z}_{(3)}^{20}$	$\mathbb{Z}_{(3)}^6$	$\mathbb{Z}_{(3)}^{13}$
6 ₁		$\mathbb{Z}_{(3)}^6 \oplus \mathbb{Z}_{(4)}^4$	$\mathbb{Z}_{(4)}^{28}$	$\mathbb{Z}_{(3)}^3 \oplus \mathbb{Z}_{(4)}^4$	$\mathbb{Z}_{(3)}^2 \oplus \mathbb{Z}_{(4)}^3$
6 ₂		$\mathbb{Z}_{(3)}^{14} \oplus \mathbb{Z}_{(4)}^2$	$\mathbb{Z}_{(4)}^{30}$	$\mathbb{Z}_{(3)}^5 \oplus \mathbb{Z}_{(4)}^2$	$\mathbb{Z}_{(3)}^2 \oplus \mathbb{Z}_{(4)}$
6 ₃		$\mathbb{Z}_{(3)}^{26}$	$\mathbb{Z}_{(4)}^{34}$	$\mathbb{Z}_{(3)}^{32}$	$\mathbb{Z}_{(3)}^2 \oplus \mathbb{Z}_{(4)}^6$
7 ₁		$\mathbb{Z}_{(2)} \oplus \mathbb{Z}_{(5)}^2$	$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}$	$\mathbb{Z}_{(2)} \oplus \mathbb{Z}_{(5)}^2$	$\mathbb{Z}_{(4)}$
7 ₂		$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}^4$	$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}^8$	$\mathbb{Z}_{(3)}^3 \oplus \mathbb{Z}_{(5)}^4$	$\mathbb{Z}_{(4)}^{12}$
7 ₃		$\mathbb{Z}_{(4)}^9$	$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}^3$	$\mathbb{Z}_{(4)}^6$	$\mathbb{Z}_{(4)}^{12}$
7 ₄		$\mathbb{Z}_{(4)}^{10} \oplus \mathbb{Z}_{(5)}^2$	$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}^6$	$\mathbb{Z}_{(3)}^2 \oplus \mathbb{Z}_{(5)}^2$	$\mathbb{Z}_{(4)}^{16}$
7 ₅		$\mathbb{Z}_{(4)}^{23}$	$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}^9$	$\mathbb{Z}_{(4)}^8$	$\mathbb{Z}_{(4)}^{22}$
7 ₆		$\mathbb{Z}_{(4)}^{43}$	$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}^{14}$	$\mathbb{Z}_{(3)}^4 \oplus \mathbb{Z}_{(4)}^{12}$	$\mathbb{Z}_{(4)}^{26}$
7 ₇		$\mathbb{Z}_{(4)}^{50}$	$\mathbb{Z}_{(4)}^2 \oplus \mathbb{Z}_{(5)}^{14}$	$\mathbb{Z}_{(3)}^9 \oplus \mathbb{Z}_{(4)}^8$	$\mathbb{Z}_{(4)}^{30}$

Part III

Knotoids and Proteins

The work in Sections 14–17 is joint with Agnese Barbensi, Oliver Vipond, Barbara I. Mahler, Pawel Dabowski-Tumanski and Dimos Goundaroulis, and it has been published ⁴². The work in Section 18 is ongoing, and is joint with Agnese Barbensi and Oliver Vipond.

A small percentage of proteins deposited in the Protein Data Bank (Berman et al., 2007) (PDB) are known to fold into conformations that are non-trivially entangled. All the currently known knotted protein structures are catalogued in the online database KnotProt (Dabrowski-Tumanski et al., 2019). The vast majority of knotted proteins form simple knot-like structures. However, there are a few examples of proteins having more complex entanglement types (Dabrowski-Tumanski et al., 2019). Interestingly, even though they are not ubiquitous, knots in proteins have been conserved within species that are separated by millions of years of evolution (Sułkowska et al., 2009). This might imply that knottiness provides some advantages to proteins (Dabrowski-Tumanski et al., 2016), but this theory is still being contested (Jackson, 2020). Besides the role of knots in proteins, there are many open questions about the pathway that the backbone of a knotted protein follows to reach its native folded state. To this date, researchers have suggested several mechanisms for protein self-tying that are based on wet lab experiments, numerical simulations, mathematical theory or a combination of all the above (Jackson et al., 2017; Mallam, 2009; Sułkowska et al., 2012a; Li et al., 2012; Chwastyk and Cieplak, 2015a; Sułkowska et al., 2009; Covino et al., 2014; Lim and Jackson, 2015; Taylor, 2007; Najafi and Potestio, 2015; Noel et al., 2010; Wang et al., 2016; He et al., 2019; Mallam and Jackson, 2012; Dabrowski-Tumanski et al., 2018; Sułkowska, 2020; Flapan et al., 2019).

Recently, Flapan et al. (Flapan et al., 2019) proposed a novel topological model that contains as special cases two of the most popular folding theories (Taylor, 2007; Bölinger et al., 2010). This theory is agnostic in terms of protein families and it is based on the assumption that there exist two loops of different sizes, each containing up to two twists. It is shown that there are several possible pathways emerging from the different choices of parameters of the model and it is suggested that these pathways are enough to recover any knot-type found empirically thus far in proteins. Since this model is theoretical, physical interactions and dynamics that may affect the folding pathway are not taken into consideration.

In this work, we systematically analyse the native states of all deposited positive trefoil knotted proteins, aiming to determine potential folding pathways. We focus on trefoil proteins as these make up the majority of knotted proteins. Considering that the global underly-

⁴² Barbensi et al. 2021 BARBENSI, Agnese ; YEROLEMOU, Naya ; VIPOND, Oliver ; MAHLER, Barbara I. ; DABROWSKI-TUMANSKI, Pawel ; GOUNDAROULIS, Dimos: A Topological Selection of Folding Pathways from Native States of Knotted Proteins. In: *Symmetry* 13 (2021), Nr. 9. – ISSN 2073–8994

ing topology of all proteins in our data set of proteins is a trefoil, the geometry of the conformation plays a key role in detecting the most probable folding motif.

Our analysis is based on KnotoEMD, a statistical distance that we define on distributions of protein backbone projections considered as knotoids. Knotoids are knot-like open-ended topological structures that extend knot theory to the case of open curves (Turaev et al., 2012). The knotoid approach is a probabilistic method that does not require an artificial closure of the protein backbone. Multiple projections of the backbone are considered to determine the predominant knotoid type (Goundaroulis et al., 2020). It has been shown that knotoids provide a more refined overview of a protein's topology (Dabrowski-Tumanski et al., 2019). KnotoEMD is a specialisation of the earth mover's distance, i.e. the first Wasserstein metric (Pele and Werman, 2009), and it quantifies the distance between different knotoid distributions in the space of trefoil knotted proteins.

We observe the existence of a double-loop in the conformation of shallow knotted carbonic anhydrases, which suggest that their folding follows a pathway similar to the theoretical model of Flapan et al.. We subsequently perform computer simulations for the carbonic anhydrase with PDB entry 4QEF. The results confirm that this protein follows a complex folding pathway involving a double loop formation at the beginning of the process, although the specific mechanisms proposed by Flapan et al. are not observed. In contrast, a much simpler geometry is revealed for other families of shallow knotted trefoil proteins and for all the deeply knotted ones, providing strong evidence against a double-loop hypothesis in these cases. This is in agreement with the current consensus that different folding pathways should be accessible by shallow knotted proteins (Piejko et al., 2020; Chwastyk and Cieplak, 2015b), and with previous simulations of deeply knotted trefoil proteins (Sulkowska, 2020). We find that KnotoEMD extracts subtle geometric information from the knotoid distribution and groups together trefoil proteins by sequence similarity. This highlights the fact that KnotoEMD is a versatile and robust mathematical tool that captures subtle topological and geometric information of open-ended conformations and has potential beyond the scope of this research, as discussed in what follows.

13 *Knotted Proteins*

Mathematically, a knot is formally defined as a closed curve in 3-dimensional (3D) space, and two knots are deemed equivalent if one can be deformed into the other without cutting and pasting the curve (Rolfsen, 2003b). In order to identify the topological type of a linear

knotted protein, a classical approach consists in considering its backbone as a polygonal curve and then joining the two termini with a new segment. The resulting closed curve is then analysed as a knot, and its underlying knot type can be formally determined (Taylor, 2000). The obtained knot type is clearly dependent on the choice of segment connecting the end-points of the curve. There are several ways to create a closed curve out of an open one in an unbiased way, in order to approximate the topology of an open curve (Dorier et al., 2020). Perhaps the most popular approaches are those where one considers several different closures at once and then takes the probability distribution of knot types as a topological descriptor of the open curve (Millett et al., 2013; Sułkowska et al., 2012b). In such a probabilistic approach, the most likely outcome in the knot distribution defines the protein’s dominant topology and characterises its global entanglement type, i.e. the topology of the entire curve. The position and the relative size of the knotted portion of the protein can be identified by the so-called knot fingerprints (King et al., 2007).

We include a more detailed description of the methods mentioned above in the next section.

13.1 Existing Summaries of Knottiness

In this section, we outline some existing methods in more detail. These fall into two main categories: *closure methods* seek to create a closed curve out of an open one, so that ordinary knot classification tools are available. The second way to summarise open curves is by analysing its subchains.

Closure Methods

Given an open chain, there are several *single-closure* methods which create a closed curve by introducing a single segment joining the two termini. One example of this is the *direct closure* method, which introduces the shortest line segment between the termini. The main problem with this method is that it can often introduce more complicated entanglement because this line segment often passes close to the protein’s centre of mass.

Another example consists in taking a sphere enclosing the protein backbone, selecting a random edge direction and drawing a parallel edge from each terminus to the sphere. One can then join the ends of the edges, for example, with a great circle arc.

The main drawback of deterministic closure methods is the fundamental problem that there is almost never a unique way to create a closed curve from an open one, and so the choice of method impacts the resulting knot type. In particular, these methods are generally not

robust to the choices made.

A more popular approach is the *stochastic closure* method (Millett and Sheldon, 2005; Lua and Grosberg, 2006), which starts with a sphere enclosing the open curve. For each draw, one selects two points on the sphere uniformly at random, each connected to a single endpoint of the sphere, and then the points on the sphere are joined with a great circle arc. This process is repeated multiple times and the dominant knot type of the closed curves is taken to be the global knot type of the curve.

For a more thorough review of these methods, see (Millett et al., 2013).

The Knotting Fingerprint

A *slipknot* is chain such that the entire chain is not considered knotted (as per a closure method) but a portion of the curve is knotted, i.e. after deleting some residues at the endpoint(s) one obtains a knotted chain. Over 1000 slip-knotted proteins have been detected of the 144,000 analysed in the PDB (Faísca, 2015).

The *knotting fingerprint* is a lower-triangular matrix representation of a protein's knottiness, summarising the knot type of each possible subchain. The horizontal axis is indexed by the starting position of the subchain, and the vertical axis the end position. More precisely, the entry (i, j) is the knot type of the subchain starting at position j and ending at position i , with the axes ordered such that the lower left corner is indexed by $(n, 0)$ where n is the length of the chain. Hence, the entry in the bottom left corner of the fingerprint is the knot type of the entire chain. See Figure 47 for an example.

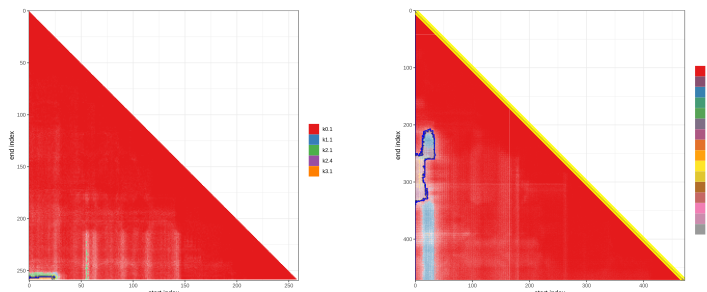


Figure 47: Knotting fingerprints, downloaded from KnotProt (Dabrowski-Tumanski et al., 2019). On the left is the knot fingerprint of protein with PDB code 6H36, which is knotted with type 3_1 . On the right is the fingerprint of protein with PDB code 6NPH, which is slipknotted.

This is a rich, informative representation of a protein's entanglement because it captures both knots and slipknots; a protein is knotted if the lower left corner of the fingerprint contains a block with non-trivial knot type, and it contains a slipknot if there is a non-trivial block shifted away from the bottom-left corner.

From the fingerprint, the *knot core* of a knotted or slip-knotted chain can be located, which is the portion of the chain that is knotted. The limitation with this summary is the difficulty in quantitatively comparing proteins.

13.2 Deep and Shallow Knotting

Recall that a protein is an open chain of amino acids, and we model the backbone of each protein as a piecewise linear curve p , where the vertices represent the amino acids. Each vertex in p is called a *residue*.

One method of locating the *knotted core* is to progressively trim residues from both ends of the chain until the knot type is trivial; the knot core is the shortest subchain for which a knot is detected, i.e. if one more residue were trimmed from either side, the knot type would be trivial. The portions of the chain between the termini and the knot core are called the *tails*, which are denoted N and C (because of the chemical composition of the amino acids at each end of the chain).

This leads to the notion of the *depth* a knotted protein: a protein is commonly considered to be *shallow* if the shortest tail consists of fewer than 20 residues, and is *deep* otherwise ⁴³.

Remark 13.1. We note here also that Barbensi and Goundaroulis ⁴⁴ present a slightly different definition of the depth of a knotted protein, that takes into account the length of the knotted core relative to the length of the tails:

$$D(p) = \frac{l_N(p)l_C(p)}{l_T(p)^2},$$

where $D(p)$ is the depth, $l_N(p)$, $l_C(p)$ are the lengths of the N and C tails, respectively, and $l_T(p)$ is the total length of the protein backbone.

14 Knotoids

Recently, knotoids have been used to study the entanglement of biopolymers (Goundaroulis et al., 2020; Dorier et al., 2018; Dabrowski-Tumanski et al., 2019; Goundaroulis et al., 2017a,b). Knotoids were introduced by Turaev (Turaev et al., 2012) as a generalisation of knots to open-ended curves. We introduce just the basic definitions that we will use to analyse proteins - see (Turaev et al., 2012) for a more thorough introduction.

Definition 14.1 (Knotoid). A *knotoid diagram* in S^2 is a generic immersion of the interval $[0, 1] \rightarrow S^2$ with finitely many transverse double points, and with over/under crossing information. The images v_0

⁴³ Taylor and Lin 2003 TAYLOR, William R. ; LIN, Kuang: Protein knots: a tangled problem. In: *Nature* 421 (2003), Nr. 6918, S. 25–25

⁴⁴ Barbensi and Goundaroulis 2021 BARBENSI, Agnese ; GOUNDAROULIS, Dimos: f-distance of knotoids and protein structure. In: *Proceedings of the Royal Society A* 477 (2021), Nr. 2246, S. 20200898

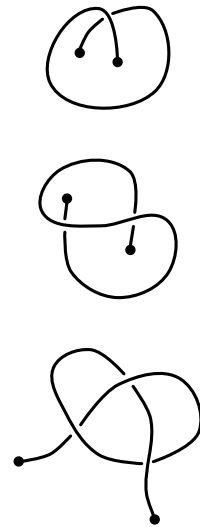


Figure 48: Examples of knotoid diagrams.

and v_1 of 0 and 1, respectively, are not double points. Knotoid diagrams are oriented from v_0 to v_1 . A *knotoid* is an equivalence class of knotoid diagrams, up to isotopies of S^2 and Reidemeister moves performed away from the endpoints. See Figure 48 for some examples.

Of course, unlike with knot diagrams, every knotoid diagram can be made trivial with a certain set of moves. *Forbidden moves* are moves on knotoid diagrams involving the endpoints: see Figure 49.

Given two knotoids k_1 and k_2 , the *forbidden move distance* or f -distance $d_f(k_1, k_2)$ between them is the minimal number of forbidden moves needed to transform k_1 to k_2 , across all representatives of the knotoids. All knotoid diagrams can be made trivial with a finite set of forbidden moves (Barbensi et al., 2018; Turaev et al., 2012).

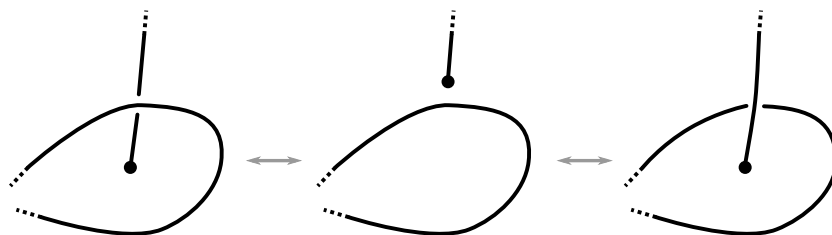


Figure 49: Forbidden moves performed locally on knotoid diagrams.

The Knotoid Distribution

Given an open curve in $\mathbb{R}^3$ (e.g. a protein backbone), we can choose a direction in which to take a planar projection on the curve. Recording the over/under information of the projection means that we can treat the projection as a knotoid. Since different directions of projections could result in different knotoid types, one can take all directions, yielding a distribution of knotoids on S^2 . This is known as the *knotoid distribution*. See Figure 51 for an illustration of how the knotoid distribution is obtained.

In practice (Dorier et al., 2018), a few thousand projections are sampled uniformly at random, and it is known that a dense enough sample provides a good approximation to the continuous knotoid distribution (Barbensi and Goundaroulis, 2021).

As explained in Section 13.1, there is not a unique way to assign a knot type to a protein; we take as our definition of a *positive trefoil protein* to be any protein whose dominant knotoid type in its knotoid distribution is a $+3_1$ (see Figure 50). This definition combines the stochastic and knotoid approaches. Note that this definition captures proteins containing trefoil slipknots.

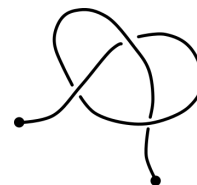
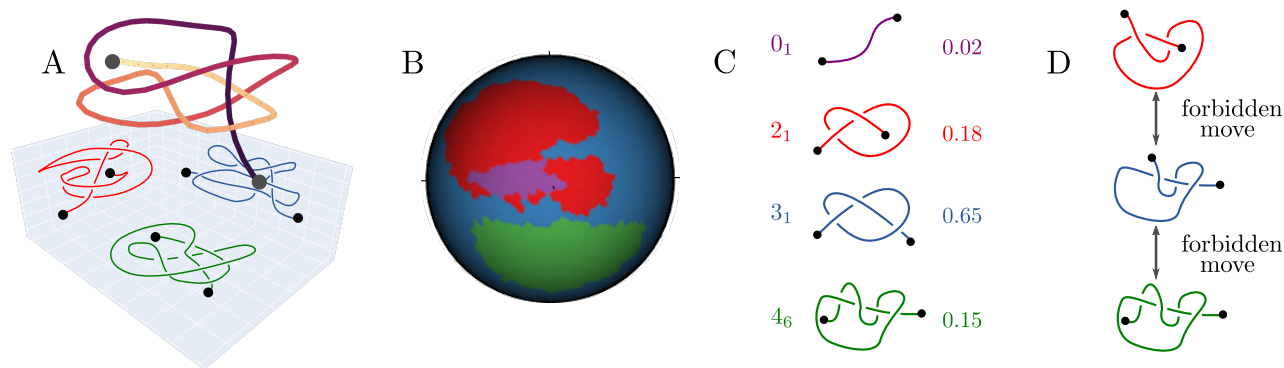


Figure 50: A diagram of a $+3_1$ knotoid.



15 Folding Hypotheses

Results over the past decade arising from theory, simulations and experiments (Jackson et al., 2017; Mallam, 2009; Sułkowska et al., 2012a; Li et al., 2012; Chwastyk and Cieplak, 2015a; Sułkowska et al., 2009; Covino et al., 2014; Lim and Jackson, 2015) reveal at least four potential folding mechanisms for trefoil proteins (Sulkowska, 2020). All of these pathways rely on the formation of a single twisted loop at one stage of the folding process, as shown in Figure 52A. They differ by the occurrence of N/C-terminus threading (Taylor, 2007; Najafi and Potestio, 2015; Noel et al., 2010), loop-flipping (aka mouse-trapping) (Covino et al., 2014; Wang et al., 2016; He et al., 2019), and by the presence of ribosomes and chaperones (Lim and Jackson, 2015; Mallam and Jackson, 2012; Dabrowski-Tumanski et al., 2018). A detailed description of various folding mechanisms is beyond the scope of this thesis. We point the interested reader to a review by Sułkowska and the references contained therein (Sulkowska, 2020). The single loop formation described by these folding mechanisms predicts a folded geometry resembling a somehow minimal open trefoil having only one pierced loop, as shown in Figure 52A and E, referred to henceforth as single loop (1-L) trefoil configurations. In reality, the geometry of all proteins is far more complex than these idealised representations; however, the smoothed, coarse geometric structure of a trefoil protein should resemble the minimal, single-looped geometry proposed by the hypotheses, if the protein folds according to one of these pathways.

Flapan et al. (Flapan et al., 2019) introduced a new theoretical model for the folding of knotted proteins based on loop flipping and numerical simulations of folding of complex knots (Bölinger et al., 2010). A characteristic feature of this model is the formation of two initial loops, as opposed to one (see Figure 52B). Similarly to above,

Figure 51: Knotoid Distribution and Forbidden Moves. (A) An open curve in 3D space with three different planar projections. Each projection defines a knotoid diagram, and different projections might yield non-equivalent knotoid diagrams. By considering all of the possible planar projections, we can associate a distribution of knotoids to the initial open curve. (B) The distribution of knotoids of the curve in (A) is visualised by colouring each point of a 2-sphere surrounding the curve according to the topological type of the corresponding projection. This picture is obtained using the software Knoto-ID. (C) The distribution of knotoids of the curve in (A). A minimal diagram is shown for each knotoid type. The knotoid distribution of each curve analysed in this work was computed using Knoto-ID. In practice, 5000 projections are computed for each curve, and this is known to be sufficient to properly approximate the continuous distribution. (D) Knotoid diagrams related by forbidden moves. Any knotoid diagram can be untangled by a finite sequence of forbidden moves.

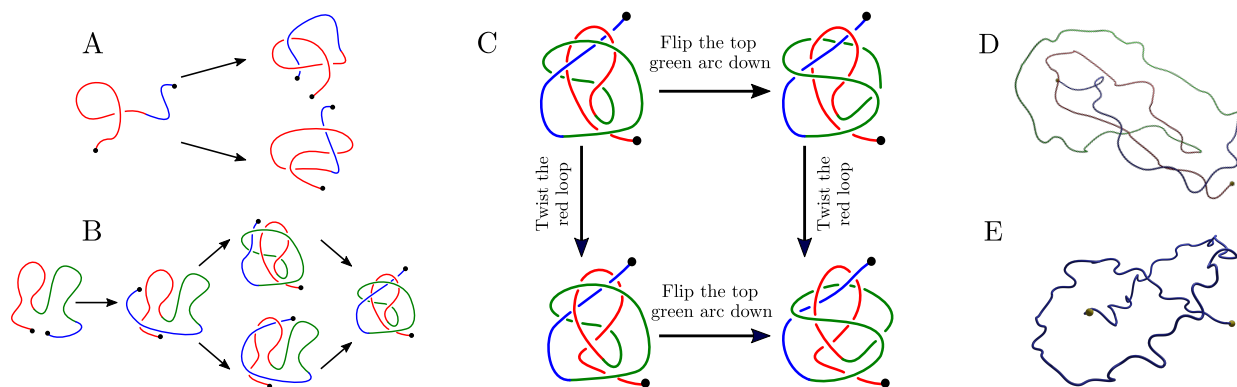


Figure 52: Folding Pathways for Trefoil Proteins. (A) A schematic representation of the single-loop folding theories. According to these theories, the formation of a twisted loop (native or not) is followed by either threading one terminus through the loop (top) or flipping the loop over the terminus (bottom). (B) A schematic representation of the double-loop folding theory. Two loops are formed and then approached by one terminus. This is followed by a combination of threading and flipping, in either order. (C) Simple 2-L configurations and moves between them. Two diagrams in the same row are related by a flipping of the green arc, while diagrams in a column are related by a twist of the red loop. Note that performing any of these moves is possible by fixing the majority of the curve and changing only one small portion. The eight other 2-L configurations are divided similarly into two groups of four. More details can be found in Section 17.2. (D) An example of a generated 2-L trajectory, rendered using Blender (Community, 2018). (E) An example of a generated 1-L trajectory, rendered using Blender (Community, 2018).

the formation of two loops predicted by this model should result in creating more complex, double-looped trefoil geometries. Knotting pathways in this model depend on the choice of several parameters, which regulate the number of twists in the loops and their relative positions. Twelve possible choices of parameters were identified, each yielding a different double-loop open positive trefoil (2-L) configuration. The twelve 2-L configurations are labelled by a two letter word in $\{L, R\}$ followed by a $\pm$ and two (signed) integers, describing the relative positions of the loops and the number of twists (Flapan et al., 2019) – a list of these names can be seen in Figure 54. Out of these twelve 2-L configurations, the least topologically complex are proposed as the most likely to occur as native states of trefoil proteins, while the more complex ones are discarded (Flapan et al., 2019) based on an approximation to the knot fingerprint. We observe that the twelve proposed 2-L configurations can be divided into three groups, where configurations within each group are related by local transformations, as illustrated in Figure 52C. More details on these groups can be found in Section 17; in particular, see Figure 56.

Henceforth, when we refer to 1-L or 2-L configurations, we mean open curves with geometries resembling the endpoints of the folding hypothesis illustrated in Figure 52A and B, respectively. Figure 52E is an example of a render of a 1-L configuration generated for our data-set, and Figure 52D is a render of a 2-L configuration. As mentioned in the paragraphs above, there is a total of twelve different 2-L configurations (see Figure 56), and one 1-L configuration.

Details on how we generate our data-set of configurations to compare with protein backbones can be found in Section 17.2.

16 KnotoEMD for Trefoil Proteins

In this section we introduce our distance between proteins, and the results of its application to trefoil proteins found in the PDB.

We state a definition of the earth mover's distance (EMD) from a linear programming perspective. Note that the EMD is also simply the first Wasserstein metric between two probability distributions.

Given a discrete distribution, we can represent it via a *signature* $P = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$, where p_i is the cluster representative and $w_{p_i} > 0$ is the weight of each cluster. The distance between the cluster types is known as the *ground distance*.

For example, we represent the knotoid distribution as the signature where each p_i is a knotoid class, and w_{p_i} is the proportion of the 5000 projections of that knotoid type. In practice, we include all knotoid classes up to and including 6 crossings, and have one cluster to count all of the knotoids with more than 6 crossings. This is done because the pairwise f -distances have been computed and tabulated for these knotoids only, and only a small proportion of directions arising from our data include more complicated knotoids.

Definition 16.1 (Earth mover's distance). Let $P = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$ and $Q = \{(q_1, w_{q_1}), \dots, (q_m, w_{q_m})\}$ be two signatures, and $D = (d_{ij})$ be the ground distance between clusters p_i and q_j . We seek to find the flow $F = (f_{ij})$, where f_{ij} is the flow from p_i to q_j , that minimises the total cost:

$$\min_F \sum_i \sum_j f_{ij} d_{ij},$$

subject to the constraints:

$$\begin{aligned} f_{ij} &\geq 0, 1 \leq i, j \leq m \\ \sum_j f_{ij} &\leq w_{p_i}, 1 \leq i \leq m \\ \sum_i f_{ij} &\leq w_{q_j}, 1 \leq j \leq m \\ \sum_i \sum_j f_{ij} &= \sum_i w_{p_i} = \sum_j w_{q_j}. \end{aligned}$$

Then, the **earth mover's distance** between P and Q is defined as:

$$EMD(P, Q) = \frac{\sum_i \sum_j f_{ij} d_{ij}}{\sum_i \sum_j f_{ij}}.$$

Given the backbones of two proteins, KnotoEMD is defined as the earth mover's distance between their knotoid distributions, with cost function (or ground distance) the f -distance.

Informally, KnotoEMD captures the cheapest way to transform one histogram of knotoid types to another, where the cost of moving

one unit of mass of type k_i to type k_j is the forbidden move distance between knotoids k_i and k_j .

Note that the total weight of each knotoid distribution associated to any protein in our data-set is the same, since we take a total of 5000 projections for each one.

16.1 Methods

The folding pathways described in Section 15 predict two different overarching trefoil geometries as an end result. Thus, we seek to differentiate trefoil proteins via their geometric entanglement in order to infer information about their respective folding pathways. Since we are working with proteins in their native state deposited in the PDB, we work under the premise that different folding pathways leave different geometric artefacts in the knot cores, such as a single or double loop in the proteins' native conformations. With this aim, we generate suitable curves of different lengths for the 1-L and twelve 2-L configurations predicted by the two folding models and then we apply numerous rounds of numerical perturbations to produce several hundred piecewise-linear open curves resembling the end-states of each of the pathways described by the models to compare with positive trefoil proteins. Our final data set consists of all the positive trefoil ($K+3_1$) proteins taken from KnotProt (Dabrowski-Tumanski et al., 2019) and a collection of generated open curves for the 1-L and 2-L configurations (see Figure 52D-E). More details on our data set can be found in Sections 17.1 and 17.2.

We then use the knotoid approach for each curve in the data set and compute its knotoid distribution. The knotoid distribution of each curve analysed in this work was computed using Knoto-ID (Dorier et al., 2018). In practice, 5000 projections are computed for each curve, and this is known to be sufficient to properly approximate the continuous distribution (Barbensi and Goundaroulis, 2021). Subsequently, we compute all pairwise KnotoEMD distances between distributions in the data set. All of these computed pairwise distances are contained in distance matrices that can be found in Section 17.3; for data visualisation purposes, we display our results using the dimensionality reduction technique, UMAP (McInnes et al., 2018), projecting the space of proteins and/or curves with distance given by KnotoEMD to a plane, whilst preserving the KnotoEMD distances as much as possible.

Our results suggest that the carbonic anhydrase 4QEF is entangled in a way similar to a 2-L configuration, so we selected it as a candidate protein for simulations to further probe its folding pathway. The coarse-grained simulations were performed in Gromacs 5 (Abraham

et al., 2015) using a $C\alpha$ structure-based model (Clementi et al., 2000) with parameters proposed by the SMOG server (Noel et al., 2016) and shadow algorithm (Noel et al., 2012). The native structure taken from PDB was first unfolded at a high temperature to obtain the starting frames for folding. Next, 200 folding trajectories with 300-500 millions of frames each were performed close to the folding temperature, where there is an equilibrium between folded and unfolded structures.

16.2 Results

We find that KnotoEMD is capable of detecting subtle geometric information from the knotoid distribution. This is reflected by the fact that it is able to group together trefoil proteins by sequence similarity. We differentiate between double-loop and single-loop open trefoils via this geometric and topological information, and detect fine structure in the space of trefoil proteins. The local geometric features detected suggest different folding pathways for trefoil proteins. In particular, we find that the proposed double-loop mechanism is the most likely for carbonic anhydrases forming a shallow trefoil.

Sequence similarity from the geometry of proteins' native states

KnotoEMD is sensitive to the geometry and topology of the entire open curve, and is therefore influenced by the depth of the knot core, locations of the termini, and the structure and entanglement of its core. These features are often shared by proteins with similar sequences, and these proteins are therefore seen as similar by KnotoEMD, as shown in Figure 53B. To show this, we cluster all trefoil proteins by first computing the pairwise sequence similarity scores, using PDB, and then performing single-linkage hierarchical clustering and pruning clusters at a threshold of 30% similarity. More details on our subdivision can be found in Section 17.1, and a dendrogram showing the clustering by sequence similarity is shown in Figure 53A. The structural features of the dendrogram are reflected almost perfectly by KnotoEMD; for example, the proteins represented by 4QEF appear to form two separate groups in the UMAP (McInnes et al., 2018) projection, coherently with the fact that two light blue clusters are merged quite late in the dendrogram to form the final one. Similarly, tight clusters in the dendrogram, such as the one represented by 6QQW, are compact in the UMAP projection. Moreover, we are able to detect a range of geometries, and to efficiently separate trefoil proteins by their geometric features. Indeed, by design,

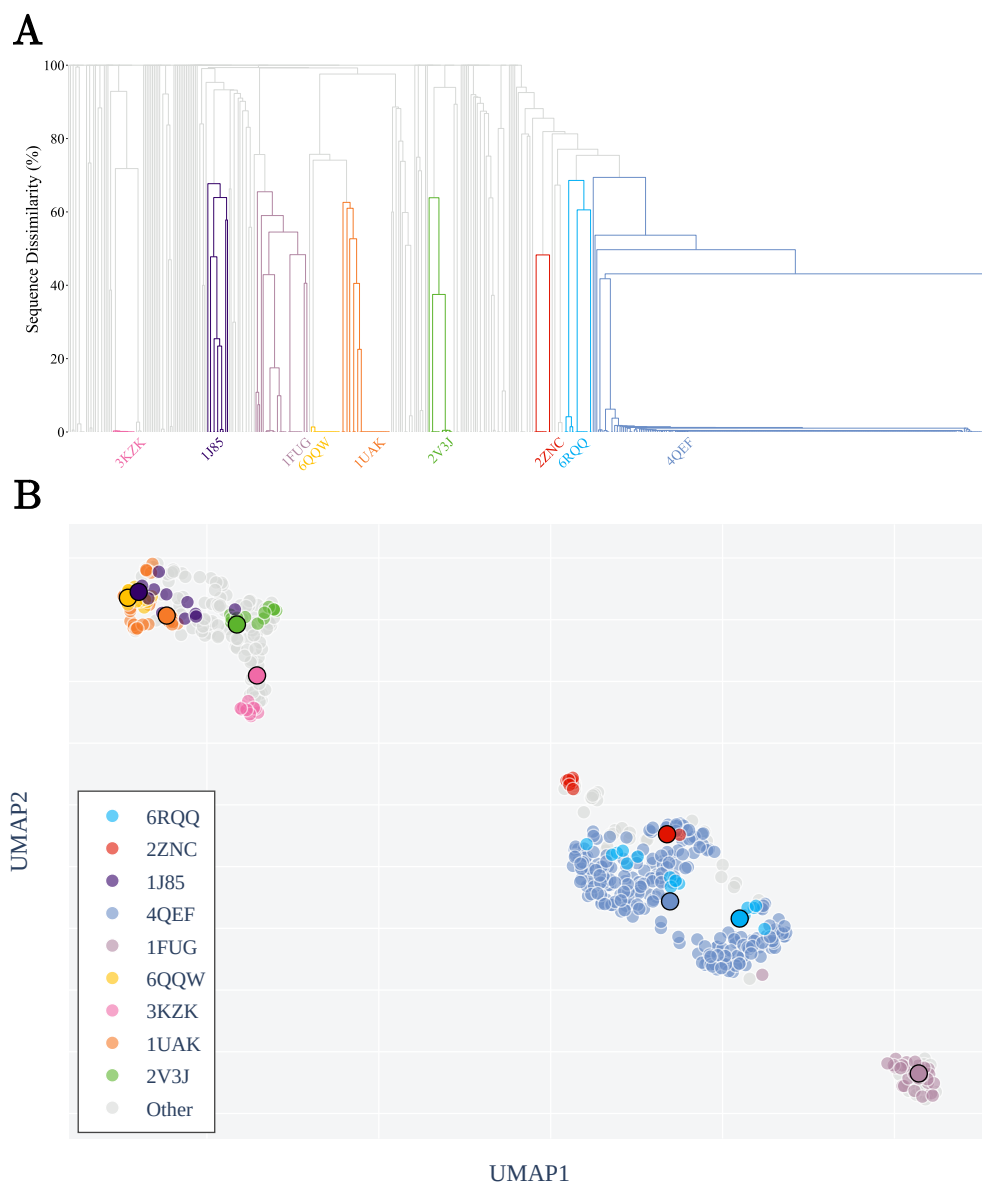


Figure 53: **KnotoEMD clusters trefoil proteins by sequence similarity.** (A) A dendrogram obtained from single-linkage hierarchical clustering of all trefoil proteins based on sequence similarity scores from PDB. (B) A UMAP projection of the space of trefoil proteins with KnotoEMD distances, when considering the entire protein backbone. Non-redundant proteins are indicated with thicker boundaries. We highlight the groups corresponding to proteins with sequences similar to 6RQQ, 2ZNC, 1J85, 4QEF, 1FUG, 6QQW, 3KZK, 1UAK, 2V3J, as these are the nine largest in our subdivision. Note that 2ZNC and 4QEF are all carbonic anhydrases with very similar structure. Similarly, 1UAK and 6QQW are both tRNA-methyltransferases. All of the deeply-knotted proteins are contained in the cluster on the left, and this cluster contains no shallow-knotted trefoil proteins. Depth is a dominant factor in determining the knotoid distribution, explaining why KnotoEMD separates proteins with a very deep knot from everything else. Finer structure in the space of proteins as detected by KnotoEMD can be seen in 3D-plots available in our GitHub repository (Yerolemou et al., 2021).

KnotoEMD takes into account both the dominant topological type of a protein and its overall structure; hence, it is able to detect features of open curves not captured by previously defined methods (King et al., 2007; Sułkowska et al., 2012b; Kufareva and Abagyan, 2011; Dabrowski-Tumanski et al., 2016) used to compare them. In particular, it is able to differentiate between proteins with the same dominant topology and at the same time distinguish pairs of homologous proteins with almost superimposable structures differing only by a strand passage, as in the case of the pair knotted AOT-Cases/unknotted OTCases (Dabrowski-Tumanski et al., 2016) (see Section 17.3).

Within this study, we investigate the geometry and topology of protein structures and compare their entanglement to knotting geometries predicted by the hypothesised knotting pathways. Since there is no restriction on the length of the two ends in the knotting hypotheses we consider, the pathways really only describe the entanglement of the core. In order to make the conformations proposed by the hypotheses comparable with the entangled proteins, we clip the ends of protein backbones up to their entangled core using a top-down approach (Dorier et al., 2018). When reducing the chain to the knotted core, the sequence similarity clustering is maintained, although it is slightly noisier in this case. The complete results on pairwise distances between protein cores can be found in Section 17.3.

KnotoEMD captures subtle geometric differences between double-loop and single-loop open trefoils

In a recent study, Piejko et al. highlight different folding behaviours for deeply knotted and shallow proteins (Piejko et al., 2020). It is then natural to ask whether there are geometric features of trefoil proteins, other than depth, that could be used to infer further information about their knotting mechanisms.

The knotoid distribution can vary substantially for different geometric realisations of an open-ended trefoil. We quantify such differences in the knotoid distributions by computing the pairwise KnotoEMD between the cores of our 1-L and 2-L trajectories, and the cores of trefoil proteins. As shown in Figure 54, the three groups of 2-L trajectories and the group of 1-L trajectories appear clearly separated, while 2-L trajectories within each of the three groups are close to each other. This demonstrates that KnotoEMD successfully captures subtle geometric structure, since configurations within the same group only differ by simple local moves (see Figure 52C), which only slightly impact the knotoid distribution and do not substantially change the overall geometry of the core.

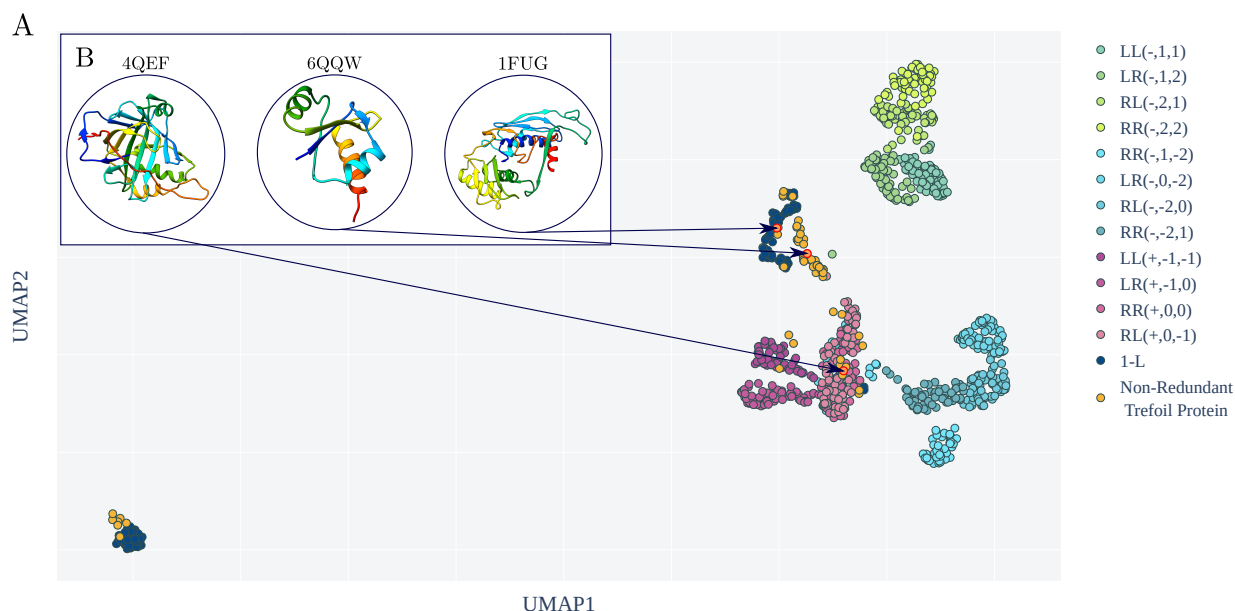


Figure 54: **Local geometric features suggest different folding pathways for trefoil proteins.** (A) 2D UMAP projection of the pairwise KnotoEMD between 1-L trajectories, 2-L trajectories and non-redundant trefoil proteins. Recall that 2-L configurations can be divided into three groups, in which configurations differ by small local transformations. 2-L configurations $LL(-,1,1)$, $LR(-,1,2)$, $RL(-,2,1)$, $RR(-,2,2)$ and $RR(-,1,-2)$, $LR(-,0,-2)$, $RL(-,-2,0)$, $RR(-,-2,1)$ form the two complex groups and $LL(+,-1,-1)$, $LR(+,-1,0)$, $RR(+,0,0)$, $RL(+,0,-1)$ the simple one. The clustering by KnotoEMD respects this subdivision almost perfectly. Similarly, 1-L configurations are clustered together, with the exception of a bottom-left group, which is separated due to oversimplification. The proteins appear distributed among the simple 2-L and the 1-L clusters. The small isolated clusters contains 1-L configurations and deeply knotted proteins whose core is oversimplified due to perturbation or noise in knot core reduction. (B) The knot cores of a shallow protein in the 2-L cluster (a human carbonic anhydrase, PDB entry 4QEF), a deeply knotted protein in the 1-L cluster (a tRNA-n1g37 methyltransferase, PDB entry 6QQW) and a shallow protein in the 1-L cluster (a S-adenosylmethionine synthetase, PDB entry 1FUG).

Local geometric features suggest different folding pathways for trefoil proteins

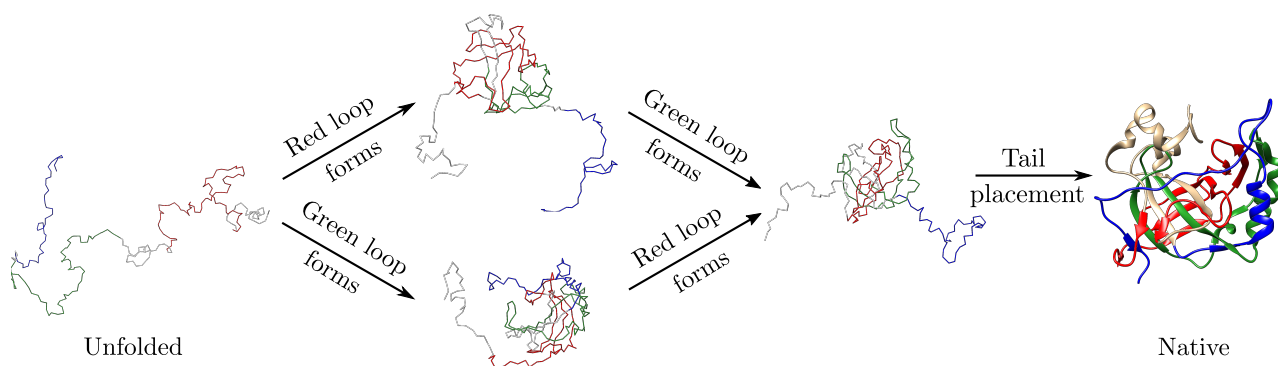
When analysing pairwise distances between protein and configuration cores, each trefoil protein core appears to have a knotoid distribution highly similar to either those of 1-L trajectories or of simple 2-L trajectories. This is particularly evident when looking at the distances between our generated trajectories and the list of non-redundant trefoil proteins taken from KnotProt (Dabrowski-Tumanski et al., 2019), as visualised using a UMAP (McInnes et al., 2018) projection in Figure 54. Our results are coherent with the fact that complex 2-L configurations exhibit knot fingerprints that are not compatible with those of trefoil proteins (Flapan et al., 2019). Indeed, trefoil proteins are distributed unevenly between the two simple groups (1-L and simple 2-L trajectories), with the vast majority having a knotted core with topological complexity indistinguishable from that of simple open-ended trefoil cores, i.e. from the 1-L trajectories. In particular, all of the non-redundant deeply knotted proteins we considered share this feature. Again, this is coherent with previously simulated folding pathways for deep trefoil proteins (Sulkowska, 2020; Piejko et al., 2020; Jackson et al., 2017), all involving the creation of a single twisted loop. This single loop can form either in a native state, as for the threading mechanisms, or unfolded, as in the loop flipping hypothesis. We emphasise that our computations demonstrate that the presence of two pierced loops in an open trefoil inevitably results in increased complexity of its knotoid distribution, making it substantially different from that of a simple trefoil; indeed, all three double-loop clusters appear clearly isolated in Figure 54. Hence, our results provide strong evidence that the double loop folding mechanisms (Flapan et al., 2019) can be excluded for deeply knotted trefoil proteins.

The same is true for the shallow protein 1FUG (see Figure 54), whose knot core appears very simple despite being considerably longer than the ones of deep proteins (ca. 256 amino acids), and even longer than many shallow ones. For these proteins, the cores' low topological complexity is shown in Figure 54, where the single-loop structure is clearly visible.

Note that there are two separate clusters of 1-L trajectories, each containing some proteins. For the 1-L trajectories, this stems from the fact that perturbing them might have resulted in an over-simplification of the curve, i.e. perturbing them in such a way that their knotoid distributions become trivial. Similarly, small errors in the knot core reduction of proteins with particularly simple core might give rise to an almost trivial knotoid distribution. In both cases, this happens

to curves whose actual core is very simple (which can be seen by inspection), so this does not affect the interpretation of the results. Proteins with the following PDB codes appear in the isolated cluster: 2YFK, 4KPP, 6AC6, 3L8U, 2V3J, 4L7I, 5AP8, all of which have very simple cores, excluding 6AC6.

There is a consensus that multiple folding pathways are possible for shallow proteins (Piejko et al., 2020; Chwastyk and Cieplak, 2015b; Rabbani et al., 2015) and this is reflected by our results. We find that, contrary to the aforementioned 1FUG which clusters with the 1-L trajectories, all carbonic anhydrase representatives, a group of proteins with very shallow trefoil knots, are clustered with simple 2-L trajectories. Hence, their knotoid distributions have a complexity coherent with the double-looped trefoil geometries. This suggests that a single-loop formation followed by either threading a terminus or flipping a loop over a terminus are unlikely folding mechanisms for carbonic anhydrases. A pathway requiring double loop formation, such as those described by Flapan et al. (Flapan et al., 2019) is the most likely among the proposed ones, as KnotoEMD shows that the knot core is tangled in a way most similar to a 2-L configuration.



To test this hypothesis we performed coarse-grained simulations of folding for the carbonic anhydrase with PDB code 4QEF. This protein features a very interesting "spine" of β -sheets, which joins together different parts of the protein. In particular, one can identify a small fragment of the chain extruding from the β -sheet core, which can be interpreted as one loop, shown in green in Figure 55, as well as a long fragment surrounding the entire β -core, which could be interpreted as the second loop, shown in red in Figure 55. The C-terminal tail protrudes from the inner loop and pierces the outer loop, eventually forming the knot. The folding of this protein starts from the formation of the β -sheet core and hence, the two loops are formed at the initial stage of the folding process. The β -sheet can

Figure 55: *The scheme of the folding pathway of a carbonic anhydrase (PDB code 4QEF). Folding starts from the unfolded conformation (left panel). Next one of the loops is formed (middle-left panel) followed by the formation of the second loop (middle right panel). Finally, both of the tails are structured, and the order of the tail attachment differs between simulations. The placement of the tails leads to the native structure (right panel).*

be formed starting either from the formation of the outer red loop, or the internal green loop, as shown in Figure 55. Once both loops are formed, the tails are then placed and the simulations suggest the existence of two different pathways, depending on which tail is placed first. Similar behaviour was observed in the case of the 5_2 -knotted UCH-L (Zhao et al., 2018), which was used as an example of a protein following a double-loop folding mechanism by Flapan et al. (Flapan et al., 2019). This mechanism, although not evidently confirming the mixed threading/mouse-trap movement event, is more complex than a simple threading mechanism observed for deeply knotted proteins.

The difficulties in unequivocal confirmation of the double-loop folding mechanism may be due to the fact that the mechanism detected is the result of some evolutionary optimisation. From a thermodynamic perspective, the folding speed is limited by the free energy barrier stemming from the knot formation. In particular, threading a free, dangling tail through an already formed loop is associated with a significant decrease of entropy, which is overcome by the concomitant formation of native contacts. This entropy-enthalpy interplay slows down the folding process, and would occur twice in a 'pure' 2-L folding mechanism (i.e. exactly like the one in Figure 52B), as there are two loops which have to be threaded. Therefore, from a kinetic point of view, such a mechanism could be optimised to reduce the threading burden. One obvious optimisation would be to make the folding similar to a single-threading mechanism, as noted also in (Flapan et al., 2019). In such a case, the observed double loop mechanism would be less evident, leaving only traces in the folding process or in the native structure. Our results indicate that these traces can be detected by KnotoEMD, giving material for follow-up studies, e.g. with mutation-based analysis.

16.3 Conclusions

Understanding how and why there are knots in proteins has been a challenging question in biology over the last few decades. One of the necessary steps to solving this question is to be able to efficiently distinguish and characterise different types of entanglement. The knotoid approach makes significant progress towards characterising the entanglement of knotted proteins since the knotoid distribution gives a topological and geometric description of an open-ended curve, richer than any closure method can give (Dabrowski-Tumanski et al., 2019). However, until now, there has been no systematic way to extract information from this distribution about specific geometric features of a curve, with the exception of very recent work by a

subgroup of the authors (Barbensi and Goundaroulis, 2021), in which the topological complexity given by the f -distance is used to separate deeply knotted trefoil proteins from shallow ones.

KnotoEMD provides a sophisticated tool able to precisely and efficiently compare the entanglement of open knotted curves. Remarkably, with KnotoEMD we are able to easily detect local geometric information within curves sharing the same global topology, which other methods have struggled to do. Importantly, we show that this geometric information can be used to reveal different folding pathways of trefoil proteins directly from their native states. Our methods are, of course, not restricted to studying folding hypotheses of proteins, since the topological and geometric information can be applied to other problems in protein science. In fact, they can be applied more generally to the study of entangled open polymers, such as DNA or RNA.

17 *Supplementary Information*

This section contains more detailed information about the data-sets used in the previous sections, and the results.

17.1 *Trefoil Proteins*

In our work we consider all the 517 proteins whose backbones form open ended positive trefoil knots catalogued in the KnotProt DataBase (Dabrowski-Tumanski et al., 2019) (as of August 2020). KnotProt classifies a protein as a positive trefoil if the dominant knotoid type in its knotoid distribution is a $+3_1$.

Restriction to Knot Core

The knot core of an open knot is defined as the shortest portion of the curve involved in the knot (Dabrowski-Tumanski et al., 2019; Dorier et al., 2018). For some of our computations we only needed to analyse the geometry and topology of knotted cores, rather than of the entire curves. In these cases, we computed the position of the knot core and of the two ends in each backbone with the software Knoto-ID (Dorier et al., 2018), that uses a top-down approach. We then removed both ends from the backbone, leaving a tolerance of one amino-acid on each side. The xyz files of each protein backbone and of their cores are available on our GitHub repository (Yerolemou et al., 2021)

Clustering by sequence-similarity

A sequence similarity search was performed for each of our 517 proteins using RCSB PDB Search API (Rose et al., 2020). Using the similarity scores returned by these searches we constructed a similarity matrix for the 517 proteins, assuming zero sequence similarity between proteins for which no score was reported. We converted the similarity matrix to a dissimilarity matrix by converting $\alpha\% \rightarrow (100 - \alpha)\%$ and performed single-linkage hierarchical clustering with the function `scipy.cluster.hierarchy.linkage` (Virtanen et al., 2020). We used a distance threshold of 70% dissimilarity to produce a clustering of the 517 proteins. We extracted the 9 largest clusters from these clusters to study.

List of non-redundant proteins

Among all the 517 trefoil proteins considered, we sometimes used as representatives the list of non-redundant ones given by KnotProt (Dabrowski-Tumanski et al., 2016). Note that in this list 7 of the proteins are obsolete PDB entries (Berman et al., 2007). In each of these cases, we replaced the entry with the corresponding updated PDB version.

The list of the 517 proteins we considered, the list of non-redundant ones and the sequence similarity clustering are all available on our GitHub repository (Yerolemou et al., 2021).

17.2 Double- and Single-Loop Open Trefoils

The twelve 2-L configurations and local moves between them

Flapan et al. identify twelve different 2-L configurations forming open ended trefoils, depending on the choice of several parameters, which regulate the number of twists in the loops and their relative positions. More precisely, the parameters prescribe the following characteristics:

- The mutual positions of the blue end and the two loops (indicated by a string of length two in R and L).
- The sign (+ or $-$) of the bottom crossing in each diagram.
- The (signed) number of (positive or negative) twists in the two loops (indicated by two integers a and b);

The twelve 2-L configurations are shown in Figure 56. We observe that the twelve proposed 2-L configurations can be divided into 3 groups, where configurations within each group are related by local transformations, as illustrated in Figure 56. Among these three

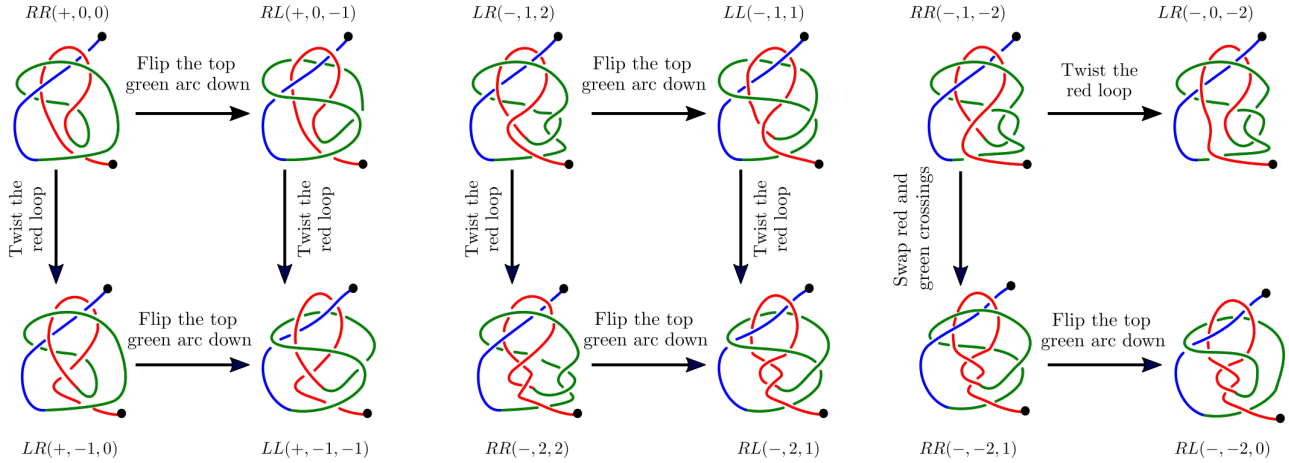


Figure 56: **2-L configurations and local moves between them.** The 12 2-L configurations differ by the number of twists in the loops and their relative positions. These configurations can be divided into groups related by local transformations that do not change the overall geometry of the curves. The 4 configurations on the left-hand side are the ones with lowest average crossing number. We will refer to them as *simple 2-L configurations*.

groups, the one whose configurations have the lower average crossing number (those forming the left-hand square in Figure 56) will be called *simple 2-L configurations*.

Generating our data-set of trajectories

Our data-set of trajectories was generated as follows.

- **Step 1: create the representative trajectories.** For each of the 12 2-L configurations we created two different representative piecewise-linear (PL) curves using the software KnotPlot (Scharein and Booth, 2002). In the same way, we created four different 1-L PL curves representing minimal (thus, admitting a projection with only 3 crossings) geometrical embeddings of open trefoils. All the curves were drawn to be quite shallow (i.e. with most of the curve involved in the knot).
- **Step 2: take different lengths of each curve.** We then subdivide each trajectory in three different ways, to obtain curves of length approximately (here the length is measured as the number of segments in the PL curve) 80, 160 and 240 (this is to match the different lengths of trefoil proteins' knotted cores). In this way we obtain a total of 6 different PL curves for each 2-L configuration, and 12 curves representing a 1-L configuration, for a total of 84 curves.
- **Step 3: perturb each curve.** We then generate 10 different trajectories for each of the 84 curves by performing numerical perturbations. The minimal distance d_m between vertices of each trajectory is determined. Each vertex is perturbed uniformly within a sphere

of radius d_m centred at the vertex. This step adds some randomness to a curve without breaking the geometry of the loops. The perturbation script is available in our GitHub repository (Yerolemou et al., 2021).

In this way, we end up with 60 different trajectories for each 2-L configuration, and 120 1-L trajectories, for a total of 840 different PL curves forming open ended trefoils with different geometries. The xyz coordinates of all the trajectories in our data-set are available on our GitHub repository (Yerolemou et al., 2021).

17.3 Computations

Knotoid Distributions

From the xyz file of each PL curve (either representing the entire chain of a protein or trajectory, or their reductions to the knot core) the knotoid distribution is computed using the software KnotoID (Dorier et al., 2018). More precisely, a total of 5000 projections are sampled and then subsequently analysed with polynomial invariants to detect their knotoid type. By results in (Barbensi and Goundaroulis, 2021) a sample size of 5000 is known to be optimal to well approximate the continuous knotoid distributions.

KnotoEMD

KnotoEMD is defined as the earth mover's distance between their knotoid distributions, with cost matrix given by the f -distance. Pairwise distances are computed using Python EMD wrapper *PyEMD* (Pele and Werman, 2008, 2009). The cost matrix containing the pairwise f -distances between knotoids with minimal crossing number at most 6 is obtained from results in (Barbensi and Goundaroulis, 2021). The cost matrix and the KnotoEMD computation script are available on our GitHub repository (Yerolemou et al., 2021). A small number of trefoil proteins and some of the most complicated 2-L trajectories have knotoid distribution containing a few occurrences of knotoids with minimal crossing number larger than 6. Since the current classification (Goundaroulis et al., 2019) of knotoids is restricted to minimal crossing number at most 6, these entries are collectively catalogued as "unknown". The f -distance between an unknown and any other knotoid is set to 2. Since the probability of occurrence of unknowns for the curves considered is very low, the overall aspect of our computations does not change for different choices of the distance between unknowns and other knotoids, such as for example 4 or 6.

Distance matrices

This section contains a description of the various KnotoEMD matrices we computed. Figure 57 shows heat-maps for pairwise distances of entire chains and knot cores of trefoil proteins. Similarly, Figure 58 contains pairwise distances of our generated trajectories (restricted to the core), and distances between these and knot cores of trefoil proteins.

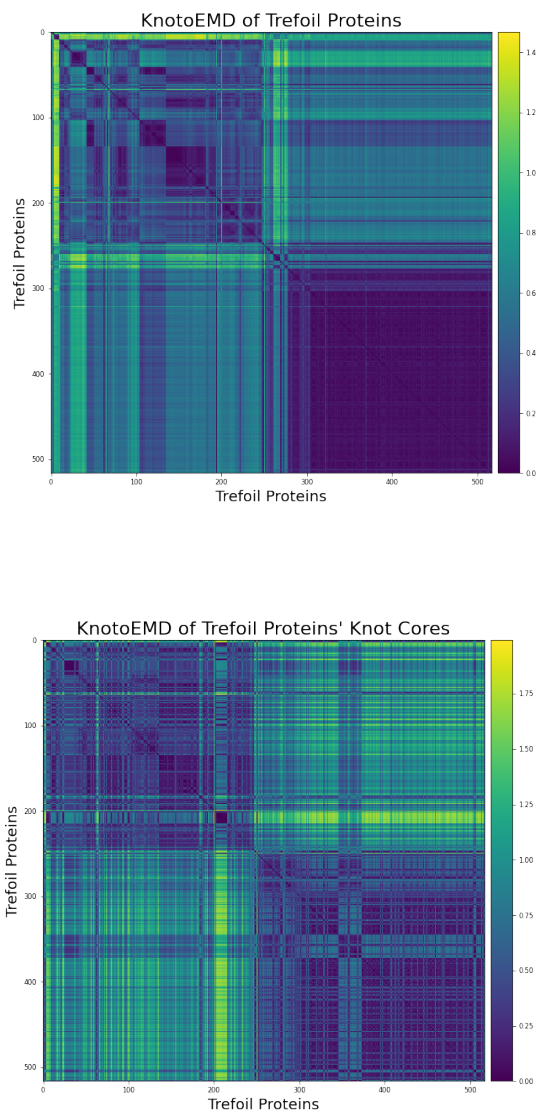


Figure 57: **Pairwise KnotoEMD on trefoil proteins.** At the top, knotoid distributions are computed from the entire chain, while at the bottom, they are computed from the knot core of each protein. Proteins are ordered according to sequence similarity: proteins in the same sequence similarity cluster are placed close to each other. The top noisy square represents pairwise distances between singletons (i.e. proteins that are not similar to any other protein considered). Note that pairwise distances between proteins in the same sequence similarity group are very low, showing that our distance recovers clustering by sequence similarity. This remains true when we only analyse the core, although the results are noisier in this case.

In all the relevant plots, proteins with similar sequences are placed closed to each other. The complete ordering and sequence similarity

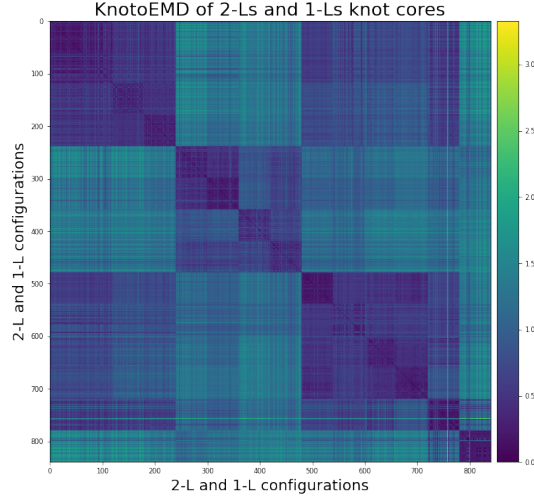
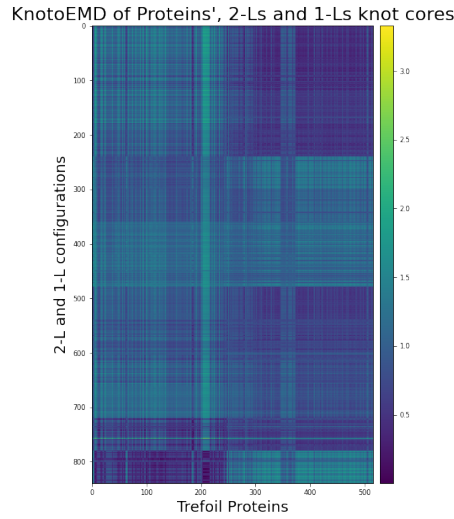


Figure 58: **KnotoEMD of 2-L configurations, 1-L configurations and trefoil proteins cores.** All the pairwise distances shown here are computed from the knotoid distributions of the knot core of each curve. Pairwise distances of our generated trajectories are shown on the left, and distance between trajectories and trefoil proteins are on the right. Note that KnotoEMD clusters each group of 2-L configurations, and the group of 1-L configurations. Similarly, distances from trajectories to proteins respect both the sequence similarity subdivision of proteins and the subdivision of trajectories.



clustering are available on our GitHub repository (Yerolemou et al., 2021). The generated trajectories are ordered as follows.

1. The simple 2-L trajectories, in the following order:

- the 60 trajectories for $RR(+, 0, 0)$;
- the 60 trajectories for $RL(+, 0, -1)$;
- the 60 trajectories for $LR(+, -1, 0)$;

- the 60 trajectories for $LL(+, -1, -1)$.
2. The first group of complex 2-L trajectories, in the following order:
 - the 60 trajectories for $LR(-, 1, 2)$;
 - the 60 trajectories for $LL(-, 1, 1)$;
 - the 60 trajectories for $RR(-, 2, 2)$;
 - the 60 trajectories for $RL(-, 2, 1)$.
 3. The second group of complex 2-L trajectories, in the following order:
 - the 60 trajectories for $LR(-, 0, -2)$;
 - the 60 trajectories for $RR(-, 1, -2)$;
 - the 60 trajectories for $RR(-, -2, 1)$;
 - the 60 trajectories for $RL(-, -2, 0)$.
 4. The 120 1-L trajectories.

UMAP projections

To visualise the spatial structure of the metric space of proteins equipped with KnotoEMD we used the non-linear dimension reduction technique UMAP (McInnes et al., 2018) on the KnotoEMD distance matrices. This dimension reduction technique requires choices of four parameters:

- `n_neighbors`: a constraint on the size of local neighbourhood considered in the dimension reduction.
- `min_dist`: the minimum distance separating points in the reduced dimension space.
- `metric`: the metric used to compare the points of the input space (in our case rows of a large distance matrix).
- `n_components`: the target dimension of the low dimensional space to which we project.

The parameter `n_neighbors` controls the emphasis on local vs global structure preservation in dimension reduction, and was set to 60 representing $\sim 10\%$ the number of points. The parameter `min_dist` affects the ability for points to tightly pack in the low dimensional embedding, and was set to 0.3 to balance fine and broad topological structure. Our plots were not highly sensitive to the choice of either the `n_neighbors` or `min_dist` parameter; the same clusters are observed when `n_neighbors` ranges across values tested in the range

10 – 200, and when `min_dist` ranges across values tested in the range 0 – 0.9. We set `metric = 'precomputed'` to indicate that our input is a distance matrix rather than a sample-feature matrix. We produced both 2 and 3 dimensional plots, and interactive versions are available in our GitHub repository (Yerolemou et al., 2021).

The knotted/unknotted homologous pair

It is important to notice that KnotoEMD captures at the same time information on the topology and geometry of open ended knots. Indeed, while it efficiently detects different structural and geometrical features among proteins forming open trefoils, it is at the same time able to distinguish proteins with almost superimposable structures differing only by a strand passage, as in the case of the pair knotted AOTCases/unknotted OTCases (Dabrowski-Tumanski et al., 2016), shown in Figure 59. The complete distance matrix and the order of the proteins considered are available on our GitHub repository (Yerolemou et al., 2021).

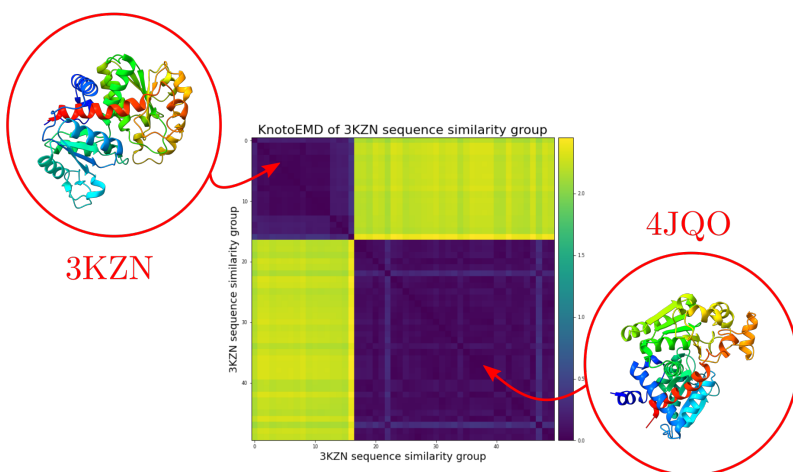


Figure 59: **KnotoEMD of knotted and unknotted proteins with sequence similar to 3KZN.** The deeply knotted AOTCases with PDB entry 3KZN has sequence highly similar to several unknotted proteins. The heat-map shows pairwise distances for proteins in this group. The first 17 entries represent the knotted ones, while the following ones are unknotted. The pair knotted AOTCases/unknotted OTCases, whose structures are almost superimposable (Dabrowski-Tumanski et al., 2016), is highlighted.

Abbreviations

The following abbreviations are used in Sections 13–17:

PDB	Protein Data Bank
2D	2-dimensional
3D	3-dimensional
1-L	Single loop open trefoil configuration
2-L	Double loop open trefoil configuration
AOTCases	N-acetylornithine transcarbamylase
OTCases	Ornithine Carbamoyltransferase
$C\alpha$	alpha-carbon
UMAP	Uniform Manifold Approximation and Projection for Dimension Reduction
SMOG	Structure-based Models for Biomolecules
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
tRNA	Transfer RNA
PL	Piecewise-linear

18 Towards the Exact Knotoid Distribution

The work in this section is ongoing, but we present the main ideas.

The aim of this work is to develop software that can compute the knotoid distribution of an open curve in 3D. In particular, the computation of the knotoid distribution reduces to two tasks:

1. Finding the boundaries of the regions in S^2 containing directions yielding the same knotoid type, and
2. Classifying the knotoid type of each region.

The current implementation ⁴⁵ produces an approximation to the knotoid distribution by sampling points on S^2 uniformly at random, classifying the knotoid type of each direction of projection, and computing the voronoi tessellation of S^2 . As mentioned earlier, in practice, a few thousand projections are computed in order to obtain a good approximation to the continuous distribution on the sphere. The main computational cost for the current implementation is the classification of the knotoid type of each projection; hence, we are motivated by two potential improvements with a new implementation:

1. Finding the exact boundaries, and
2. Computing knotoid invariants to classify only one point per region.

In particular, we seek to remove the need to compute knotoid invariants for multiple points per region, because this information is redundant. For simple curves, this could reduce the number of

⁴⁵ Dorier et al. 2018 DORIER, Julien ; GOUNDAROULIS, Dimos ; BENEDETTI, Fabrizio ; STASIAK, Andrzej: Knoto-ID: a tool to study the entanglement of open protein chains using the concept of knotoids. In: *Bioinformatics* 34 (2018), Nr. 19, S. 3402–3404

projections that need classifying from thousands to tens. For more complicated curves, for example, long protein backbones, this could reduce to hundreds.

Finding Region Boundaries

First, consider a smooth curve C in $\mathbb{R}^3$; the knotoid type of two nearby projections can only change if the knotoid diagrams differ by at least one forbidden move. Recall that forbidden moves are those moves involving the endpoints of C . Hence, all of the points on S^2 that could possibly be contained in the boundaries of regions, either side of which have a different knotoid classification, are directions where the endpoints align with another point on the curve.

More precisely, let e_0 and e_1 denote the two endpoints of C . For any point p in C , the directions

$$\pm \frac{(p - e_0)}{\|p - e_0\|}$$

and

$$\pm \frac{(p - e_1)}{\|p - e_1\|}$$

could all participate in the boundaries separating regions with different knotoid types. Note that we say ‘could’ because there can be moves involving endpoints that do not change the knotoid type.

Figure 60 illustrates how we trace out curves in S^2 by tracking

$$\pm \frac{(p - e_i)}{\|p - e_i\|}$$

for all $p \in C$, for each endpoint.

For a piecewise-linear curve, p ranges over the endpoints of the linear segments, and great circle arc segments interpolate between adjacent points on the sphere.

The Algorithm

The code currently implements the following pseudo-code for a curve C :

1. Find four curves $c_1, \dots, c_4$ of S^2 generated by $\pm \frac{(p - e_i)}{\|p - e_i\|}$ for each endpoint $i = 0, 1$.
2. Find all intersection points of c_i with each other.
3. Project to the plane to obtain a planar graph.
4. Find region boundaries via a planar graph face-finding algorithm.

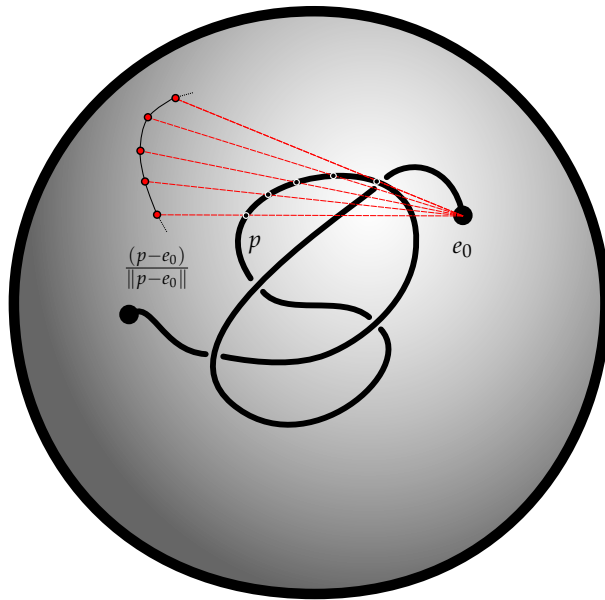


Figure 60: The only directions of projection of C that lie on region boundaries between different knotoid types are those in which either end-point line up with another point on the curve.

5. Find a single point contained inside each region (to use to classify the knotoid type of the projection of C in the direction of any point inside that region).

The classification of the points is not currently implemented. The in-progress code can be found in our GitHub repository ⁴⁶.

First Examples

We include some figures to demonstrate the current code functionality. For a simple open trefoil, we obtain the following 2D and 3D regions. Note that the boundaries of the regions in the plane are drawn as straight lines, but should be thought of as projections of great circle arcs.

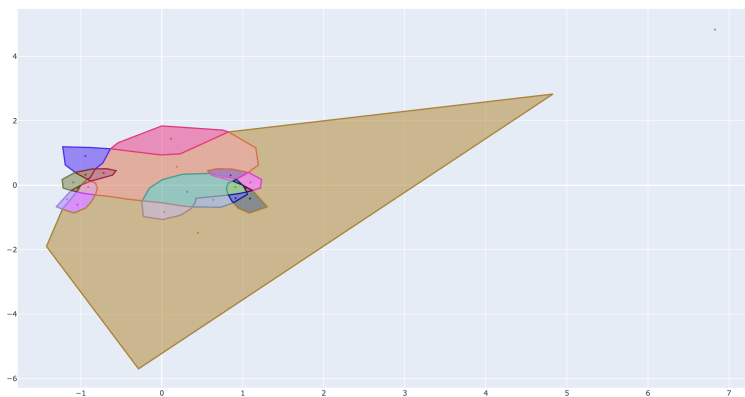


Figure 61: An illustration of the stereographic projection of regions found by our algorithm. The knotoid type of the projection of the curve in the direction given by any point in a single region is the same. Neighbouring regions can share the same knotoid type. The colours do not correspond to knotoid type.

⁴⁶ Yerolemov and Barbensi 2022
YEROLEMOU, Naya ; BARBENSI, Agnese: *Knotoid Distribution*.
<https://github.com/nyerolemov/knotoid-distribution>, 2022

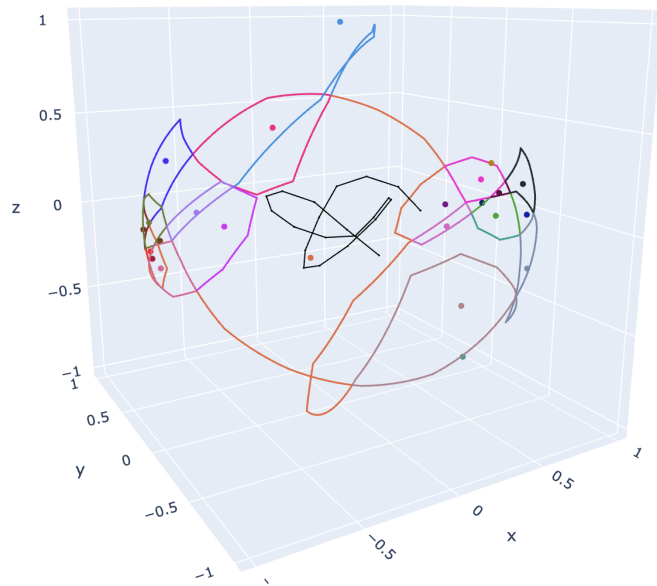


Figure 62: Visualising the regions on the sphere.

For the future, a complete implementation of the algorithm must be completed, including performance and complexity analysis.

Bibliography

- [Abe 2011] ABE, Yukiko: The clock number of a knot. In: *arXiv preprint arXiv:1103.0072* (2011)
- [Abraham et al. 2015] ABRAHAM, Mark J. ; MURTOLA, Teemu ; SCHULZ, Roland ; PÁLL, Szilárd ; SMITH, Jeremy C. ; HESS, Berk ; LINDAHL, Erik: GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. In: *SoftwareX* 1 (2015), S. 19–25
- [Ayala et al. 2008] AYALA, R ; FERNÁNDEZ, LM ; QUINTERO, A ; VILCHES, JA: A note on the pure Morse complex of a graph. In: *Topology and its Applications* 155 (2008), Nr. 17-18, S. 2084–2089
- [Barbensi et al. 2018] BARBENSI, Agnese ; BUCK, Dorothy ; HARRINGTON, Heather A. ; LACKENBY, Marc: Double branched covers of knotoids. In: *arXiv preprint arXiv:1811.09121* (2018)
- [Barbensi and Goundaroulis 2021] BARBENSI, Agnese ; GOUNDAROULIS, Dimos: f-distance of knotoids and protein structure. In: *Proceedings of the Royal Society A* 477 (2021), Nr. 2246, S. 20200898
- [Barbensi et al. 2021] BARBENSI, Agnese ; YEROLEMOU, Naya ; VIPOND, Oliver ; MAHLER, Barbara I. ; DABROWSKI-TUMANSKI, Pawel ; GOUNDAROULIS, Dimos: A Topological Selection of Folding Pathways from Native States of Knotted Proteins. In: *Symmetry* 13 (2021), Nr. 9. – ISSN 2073–8994
- [Bass 1993] BASS, Hyman: Covering theory for graphs of groups. In: *Journal of Pure and Applied Algebra* 89 (1993), Nr. 1, S. 3–47. – ISSN 0022–4049
- [Benedetti 2016] BENEDETTI, Bruno: Smoothing discrete Morse theory. In: *Annali della Scuola Normale Superiore di Pisa-Classe di scienze* 16 (2016), Nr. 2, S. 335–368
- [Berman et al. 2007] BERMAN, Helen ; HENRICK, Kim ; NAKAMURA, Haruki ; MARKLEY, John L.: The worldwide Protein Data Bank

- (wwPDB): ensuring a single, uniform archive of PDB data. In: *Nucleic acids research* 35 (2007), Nr. suppl_1, S. D301–D303
- [Biggs 1993] BIGGS, Norman: *Algebraic graph theory*. Bd. 67. Cambridge University Press, 1993
- [Bölinger et al. 2010] BÖLINGER, Daniel ; SUŁKOWSKA, Joanna I. ; HSU, Hsiao-Ping ; MIRNY, Leonid A. ; KARDAR, Mehran ; ONUCHIC, José N ; VIRNAU, Peter: A Stevedore's protein knot. In: *PLoS computational biology* 6 (2010), Nr. 4, S. e1000731
- [Bredon 1972] BREDON, Glen E.: *Introduction to Compact Transformation Groups*. Academic Press, 1972 (Pure and Applied Mathematics 46)
- [Bridson and Häfliger] BRIDSON, M.R. ; HÄFLIGER, A.: *Metric Spaces of Non-Positive Curvature*. Springer Berlin Heidelberg (Grundlehren der mathematischen Wissenschaften). – ISBN 9783540643241
- [Burde and Zieschang 1985] BURDE, Gerhard ; ZIESCHANG, Heiner: *Knots*. Bd. 5 of de Gruyter Studies in Mathematics. Walter de Gruyter & Co., 1985
- [Celoria and Yerolemou 2021] CELORIA, Daniele ; YEROLEMOU, Naya: A discrete Morse perspective on knot projections and a generalised clock theorem. In: *The Electronic Journal of Combinatorics* (2021)
- [Chari 2000] CHARI, Manoj K.: On discrete Morse functions and combinatorial decompositions. In: *Discrete Mathematics* 217 (2000), Nr. 1-3, S. 101–113
- [Chari and Joswig 2005] CHARI, Manoj K. ; JOSWIG, Michael: Complexes of discrete Morse functions. In: *Discrete mathematics* 302 (2005), Nr. 1-3, S. 39–51
- [Chebotarev 2008] CHEBOTAREV, Pavel: Spanning forests and the golden ratio. In: *Discrete Applied Mathematics* 156 (2008), Nr. 5, S. 813–821
- [Chwastyk and Cieplak 2015a] CHWASTYK, Mateusz ; CIEPLAK, Marek: Cotranslational folding of deeply knotted proteins. In: *Journal of Physics: Condensed Matter* 27 (2015), Nr. 35, S. 354105
- [Chwastyk and Cieplak 2015b] CHWASTYK, Mateusz ; CIEPLAK, Marek: Multiple folding pathways of proteins with shallow knots and co-translational folding. In: *The Journal of chemical physics* 143 (2015), Nr. 4, S. 07B611_1

- [Clementi et al. 2000] CLEMENTI, Cecilia ; NYMEYER, Hugh ;
 ONUCHIC, José N.: Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. In: *Journal of molecular biology* 298 (2000), Nr. 5, S. 937–953
- [Cohen 2012] COHEN, Moshe: A correspondence between complexes and knots. In: *arXiv preprint arXiv:1211.2553* (2012)
- [Cohen et al. 2014] COHEN, Moshe ; DASBACH, Oliver T. ; RUSSELL, Heather M.: A twisted dimer model for knots. In: *Fundamenta Mathematicae* 225 (2014), S. 57–74
- [Cohen and Teicher 2014] COHEN, Moshe ; TEICHER, Mina: Kauffman’s clock lattice as a graph of perfect matchings: a formula for its height. In: *The Electronic Journal of Combinatorics* 21 (2014), Nr. 4, S. P4–31
- [Community 2018] COMMUNITY, Blender O.: Blender - a 3D modelling and rendering package. (2018). <http://www.blender.org>
- [Corson 1992] CORSON, Jon M.: Complexes of groups. In: *Proceedings of the London Mathematical Society* 3 (1992), Nr. 1, S. 199–224
- [Covino et al. 2014] COVINO, Roberto ; ŠKRBIĆ, Tatjana ; BECCARA, Silvio A. ; FACCIOLO, Pietro ; MICHELETTI, Cristian: The role of non-native interactions in the folding of knotted proteins: insights from molecular dynamics simulations. In: *Biomolecules* 4 (2014), Nr. 1, S. 1–19
- [Culler et al.] CULLER, Marc ; DUNFIELD, Nathan M. ; GOERNER, Matthias ; WEEKS, Jeffrey R.: *SnapPy, a computer program for studying the geometry and topology of 3-manifolds*. Available at <http://snappy.computop.org>,
- [Dabrowski-Tumanski et al. 2018] DABROWSKI-TUMANSKI, Pawel ; PIEJKO, Maciej ; NIEWIECZERZAL, Szymon ; STASIAK, Andrzej ; SUŁKOWSKA, Joanna I.: Protein knotting by active threading of nascent polypeptide chain exiting from the ribosome exit channel. In: *The Journal of Physical Chemistry B* 122 (2018), Nr. 49, S. 11616–11625
- [Dabrowski-Tumanski et al. 2019] DABROWSKI-TUMANSKI, Pawel ; RUBACH, Pawel ; GOUNDAROULIS, Dimos ; DORIER, Julien ; SUŁKOWSKI, Piotr ; MILLETT, Kenneth C. ; RAWDON, Eric J. ; STASIAK, Andrzej ; SUŁKOWSKA, Joanna I.: KnotProt 2.0: a database of proteins with knots and other entangled structures. In: *Nucleic acids research* 47 (2019), Nr. D1, S. D367–D375

- [Dabrowski-Tumanski et al. 2016] DABROWSKI-TUMANSKI, Pawel ; STASIAK, Andrzej ; SULKOWSKA, Joanna I.: In search of functional advantages of knots in proteins. In: *PloS one* 11 (2016), Nr. 11, S. e0165986
- [Dey et al. 2018] DEY, Tamal K. ; WANG, Jiayuan ; WANG, Yusu: Graph Reconstruction by Discrete Morse Theory. (2018)
- [Dorier et al. 2018] DORIER, Julien ; GOUNDAROULIS, Dimos ; BENEDETTI, Fabrizio ; STASIAK, Andrzej: Knoto-ID: a tool to study the entanglement of open protein chains using the concept of knotoids. In: *Bioinformatics* 34 (2018), Nr. 19, S. 3402–3404
- [Dorier et al. 2020] DORIER, Julien ; GOUNDAROULIS, Dimos ; RAWDON, Eric J. ; STASIAK, Andrzej: Open Knots. In: *Encyclopedia of Knot Theory*. Chapman and Hall/CRC, 2020, Kapitel 84
- [Faísca 2015] FAÍSCA, Patrícia FN: Knotted proteins: A tangled tale of structural biology. In: *Computational and structural biotechnology journal* 13 (2015), S. 459–468
- [Flapan et al. 2019] FLAPAN, Erica ; HE, Adam ; WONG, Helen: Topological descriptions of protein folding. In: *Proceedings of the National Academy of Sciences* 116 (2019), Nr. 19, S. 9360–9369
- [Forman 1998] FORMAN, Robin: Morse Theory for Cell Complexes. In: *Advances in Mathematics* 134 (1998), S. 90–145
- [Goundaroulis et al. 2017a] GOUNDAROULIS, Dimos ; DORIER, Julien ; BENEDETTI, Fabrizio ; STASIAK, Andrzej: Studies of global and local entanglements of individual protein chains using the concept of knotoids. In: *Scientific reports* 7 (2017), Nr. 1, S. 1–9
- [Goundaroulis et al. 2019] GOUNDAROULIS, Dimos ; DORIER, Julien ; STASIAK, Andrzej: A systematic classification of knotoids on the plane and on the sphere. In: *arXiv preprint arXiv:1902.07277* (2019)
- [Goundaroulis et al. 2020] GOUNDAROULIS, Dimos ; DORIER, Julien ; STASIAK, Andrzej: Knotoids and protein structure. In: *Topology and Geometry of Biopolymers* 746 (2020), S. 185
- [Goundaroulis et al. 2017b] GOUNDAROULIS, Dimos ; GÜGÜMCÜ, Neslihan ; LAMBROPOULOU, Sofia ; DORIER, Julien ; STASIAK, Andrzej ; KAUFFMAN, Louis: Topological models for open-knotted protein chains using the concepts of knotoids and bonded knotoids. In: *Polymers* 9 (2017), Nr. 9, S. 444
- [Haefliger 1991] HAEFLIGER, André: Complexes of groups and orbihedra. In: *Group theory from a geometrical viewpoint*. 1991

- [Haeffliger 1992] HAEFLIGER, André: Extension of complexes of groups. In: *Annales de l'institut Fourier* Bd. 42, 1992, S. 275–311
- [Harary et al. 1983] HARARY, F ; MOKKEN, R ; PLANTHOLT, M: Interpolation theorem for diameters of spanning trees. In: *IEEE transactions on circuits and systems* 30 (1983), Nr. 7, S. 429–432
- [He et al. 2019] HE, Chengzhi ; LI, Shuai ; GAO, Xiaoqing ; XIAO, Adam ; HU, Chunguang ; HU, Xiaodong ; HU, Xiaotang ; LI, Hongbin: Direct observation of the fast and robust folding of a slipknotted protein by optical tweezers. In: *Nanoscale* 11 (2019), Nr. 9, S. 3945–3951
- [Hine and Kálmán 2018] HINE, Camden ; KÁLMÁN, Tamás: Clock theorems for triangulated surfaces. In: *arXiv preprint arXiv:1808.06091* (2018)
- [Huggett et al. 2012] HUGGETT, Stephen ; MOFFATT, Iain ; VIRDEE, Natalia: On the Seifert graphs of a link diagram and its parallels. In: *Mathematical Proceedings of the Cambridge Philosophical Society* Bd. 153 Cambridge University Press, 2012, S. 123–145
- [Jackson 2020] JACKSON, Sophie E.: Why are there knots in proteins? In: *Topology and Geometry of Biopolymers* 746 (2020), S. 129
- [Jackson et al. 2017] JACKSON, Sophie E. ; SUMA, Antonio ; MICHELETTI, Cristian: How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. In: *Current opinion in structural biology* 42 (2017), S. 6–14
- [Kauffman 2006] KAUFFMAN, Louis H.: *Formal knot theory*. Courier Corporation, 2006
- [Kenyon et al. 2000] KENYON, Richard W. ; PROPP, James G. ; WILSON, David B.: Trees and Matchings. In: *The Electronic Journal of Combinatorics* 7 (2000), Nr. R25, S. 2
- [King et al. 2007] KING, Neil P. ; YEATES, Eric O. ; YEATES, Todd O.: Identification of rare slipknots in proteins and their implications for stability and folding. In: *Journal of molecular biology* 373 (2007), Nr. 1, S. 153–166
- [Knudson 2015] KNUDSON, Kevin P.: *Morse theory: Smooth and discrete*. World Scientific Publishing Company, 2015
- [Kufareva and Abagyan 2011] KUFAREVA, Irina ; ABAGYAN, Ruben: Methods of protein structure comparison. In: *Homology Modeling*. Springer, 2011, S. 231–257

- [Li et al. 2012] LI, Wenfei ; TERAKAWA, Tsuyoshi ; WANG, Wei ; TAKADA, Shoji: Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. In: *Proceedings of the National Academy of Sciences* 109 (2012), Nr. 44, S. 17789–17794
- [Lim and Jackson 2015] LIM, Nicole C. ; JACKSON, Sophie E.: Mechanistic insights into the folding of knotted proteins in vitro and in vivo. In: *Journal of molecular biology* 427 (2015), Nr. 2, S. 248–258
- [Lua and Grosberg 2006] LUA, Rhonald C. ; GROSBERG, Alexander Y.: Statistics of knots, geometry of conformations, and evolution of proteins. In: *PLoS computational biology* 2 (2006), Nr. 5, S. e45
- [Mallam 2009] MALLAM, Anna L.: How does a knotted protein fold? In: *The FEBS journal* 276 (2009), Nr. 2, S. 365–375
- [Mallam and Jackson 2012] MALLAM, Anna L. ; JACKSON, Sophie E.: Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. In: *Nature chemical biology* 8 (2012), Nr. 2, S. 147–153
- [McInnes et al. 2018] MCINNES, Leland ; HEALY, John ; SAUL, Nathaniel ; GROSSBERGER, Lukas: UMAP: Uniform Manifold Approximation and Projection. In: *Journal of Open Source Software* 3 (2018), September, Nr. 29, S. 861. <http://dx.doi.org/10.21105/joss.00861>. – DOI 10.21105/joss.00861
- [Millett et al. 2013] MILLETT, Kenneth C. ; RAWDON, Eric J. ; STASIAK, Andrzej ; SUŁKOWSKA, Joanna I.: Identifying knots in proteins. In: *Biochemical Society Transactions* 41 (2013), Nr. 2, S. 533–537
- [Millett and Sheldon 2005] MILLETT, Kenneth C. ; SHELDON, Benjamin M.: Tying down open knots: a statistical method for identifying open knots with applications to proteins. In: *Physical And Numerical Models In Knot Theory: Including Applications to the Life Sciences*. World Scientific, 2005, S. 203–217
- [Mischaikow and Nanda 2013] MISCHAIKOW, Konstantin ; NANDA, Vidit: Morse theory for filtrations and efficient computation of persistent homology. In: *Discrete & Computational Geometry* 50 (2013), Nr. 2, S. 330–353
- [Najafi and Potestio 2015] NAJAFI, Saeed ; POTESTIO, Raffaello: Folding of small knotted proteins: Insights from a mean field coarse-grained model. In: *The Journal of chemical physics* 143 (2015), Nr. 24, S. 12B606_1

- [Nanda 2019] NANDA, Vidit: Discrete Morse theory and localization. In: *Journal of Pure and Applied Algebra* 223 (2019), Nr. 2, S. 459–488
- [Nicolaescu 2012] NICOLAESCU, Liviu I.: Combinatorial Morse flows are hard to find. In: *arXiv preprint arXiv:1202.2724* (2012)
- [(<https://mathoverflow.net/users/20302/liviu-nicolaescu>)] NICOLAESCU, Liviu N.: *Combinatorial Morse functions and random permutations*. MathOverflow. <https://mathoverflow.net/q/86193>. – URL:<https://mathoverflow.net/q/86193> (version: 2012-11-19)
- [nLab authors 2022] NLAB AUTHORS: *2-category*. <http://ncatlab.org/nlab/show/2-category>, August 2022. – Revision 48
- [Noel et al. 2016] NOEL, Jeffrey K. ; LEVI, Mariana ; RAGHUNATHAN, Mohit ; LAMMERT, Heiko ; HAYES, Ryan L. ; ONUCHIC, José N ; WHITFORD, Paul C.: SMOG 2: a versatile software package for generating structure-based models. In: *PLoS computational biology* 12 (2016), Nr. 3, S. e1004794
- [Noel et al. 2010] NOEL, Jeffrey K. ; SUŁKOWSKA, Joanna I. ; ONUCHIC, José N: Slipknotting upon native-like loop formation in a trefoil knot protein. In: *Proceedings of the National Academy of Sciences* 107 (2010), Nr. 35, S. 15403–15408
- [Noel et al. 2012] NOEL, Jeffrey K. ; WHITFORD, Paul C. ; ONUCHIC, José N: The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. In: *The Journal of Physical Chemistry B* 116 (2012), Nr. 29, S. 8692–8702
- [Ozsváth and Szabó 2018] OZSVÁTH, Peter ; SZABÓ, Zoltán: Kauffman states, bordered algebras, and a bigraded knot invariant. In: *Advances in Mathematics* 328 (2018), S. 1088–1198
- [Pele and Werman 2008] PELE, Ofir ; WERMAN, Michael: A linear time histogram metric for improved sift matching. In: *Computer Vision–ECCV 2008*, Springer, October 2008, S. 495–508
- [Pele and Werman 2009] PELE, Ofir ; WERMAN, Michael: Fast and robust earth mover’s distances. In: *2009 IEEE 12th International Conference on Computer Vision IEEE*, 2009, S. 460–467
- [Piejko et al. 2020] PIEJKO, Maciej ; NIEWIECZERZAL, Szymon ; SUŁKOWSKA, Joanna I.: The Folding of Knotted Proteins: Distinguishing the Distinct behavior of Shallow and Deep Knots. In: *Israel Journal of Chemistry* 60 (2020), Nr. 7, S. 713–724
- [Rabbani et al. 2015] RABBANI, Gulam ; AHMAD, Ejaz ; KHAN, Mohsin V. ; ASHRAF, Mohd T. ; BHAT, Rajiv ; KHAN, Rizwan H.:

- Impact of structural stability of cold adapted *Candida antarctica* lipase B (CaLB): in relation to pH, chemical and thermal denaturation. In: *Rsc Advances* 5 (2015), Nr. 26, S. 20115–20131
- [Robins et al. 2011] ROBINS, Vanessa ; WOOD, Peter J. ; SHEPPARD, Adrian P.: Theory and algorithms for constructing discrete Morse complexes from grayscale digital images. In: *IEEE Transactions on pattern analysis and machine intelligence* 33 (2011), Nr. 8, S. 1646–1658
- [Rolfsen 2003a] ROLFSEN, D.: *Knots and Links*. AMS Chelsea Pub., 2003 (AMS Chelsea Publishing Series). – ISBN 9780821834367
- [Rolfsen 2003b] ROLFSEN, Dale: *Knots and links*. Bd. 346. American Mathematical Soc., 2003
- [Rose et al. 2020] ROSE, Yana ; DUARTE, Jose M. ; LOWE, Robert ; SEGURA, Joan ; BI, Chunxiao ; BHIKADIYA, Charmi ; CHEN, Li ; ROSE, Alexander S. ; BITTRICH, Sebastian ; BURLEY, Stephen K. ; WESTBROOK, John D.: RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive. In: *Journal of Molecular Biology* (2020), November, 166704. <http://dx.doi.org/10.1016/j.jmb.2020.11.003>. – DOI 10.1016/j.jmb.2020.11.003
- [Scharein and Booth 2002] SCHAREIN, Robert G. ; BOOTH, Kellogg S.: Interactive knot theory with KnotPlot. In: *Multimedia Tools for Communicating Mathematics*. Springer, 2002, S. 277–290
- [Scoville and Zaremsky 2020] SCOVILLE, Nicholas A. ; ZAREMSKY, Matthew C. B.: Higher connectivity of the Morse complex. In: *arXiv preprint arXiv:2004.10481* (2020)
- [Sułkowska et al. 2012a] SUŁKOWSKA, Joanna I. ; NOEL, Jeffrey K. ; ONUCHIC, Jose N.: Energy landscape of knotted protein folding. In: *Proceedings of the National Academy of Sciences* 109 (2012), Nr. 44, S. 17783–17788
- [Sułkowska et al. 2012b] SUŁKOWSKA, Joanna I. ; RAWDON, Eric J. ; MILLETT, Kenneth C. ; ONUCHIC, Jose N. ; STASIAK, Andrzej: Conservation of complex knotting and slipknotting patterns in proteins. In: *Proceedings of the National Academy of Sciences* 109 (2012), Nr. 26, S. E1715–E1723
- [Sułkowska et al. 2009] SUŁKOWSKA, Joanna I. ; SUŁKOWSKI, Piotr ; ONUCHIC, Jose: Dodging the crisis of folding proteins with knots. In: *Proceedings of the National Academy of Sciences* 106 (2009), Nr. 9, S. 3119–3124

- [Sulkowska 2020] SULKOWSKA, Joanna I.: On folding of entangled proteins: knots, lassos, links and θ -curves. In: *Current opinion in structural biology* 60 (2020), S. 131–141
- [Taylor 2000] TAYLOR, William R.: A deeply knotted protein structure and how it might fold. In: *Nature* 406 (2000), Nr. 6798, S. 916–919
- [Taylor 2007] TAYLOR, William R.: Protein knots and fold complexity: some new twists. In: *Computational biology and chemistry* 31 (2007), Nr. 3, S. 151–162
- [Taylor and Lin 2003] TAYLOR, William R. ; LIN, Kuang: Protein knots: a tangled problem. In: *Nature* 421 (2003), Nr. 6918, S. 25–25
- [The Sage Developers 2020] THE SAGE DEVELOPERS: *Sage-Math, the Sage Mathematics Software System (Version 9.0)*, 2020. – <https://www.sagemath.org>
- [Turaev et al. 2012] TURAEV, Vladimir et al.: Knotoids. In: *Osaka Journal of Mathematics* 49 (2012), Nr. 1, S. 195–223
- [Virtanen et al. 2020] VIRTANEN, Pauli ; GOMMERS, Ralf ; OLIPHANT, Travis E. ; HABERLAND, Matt ; REDDY, Tyler ; COURNAPEAU, David ; BUROVSKI, Evgeni ; PETERSON, Pearu ; WECKESSER, Warren ; BRIGHT, Jonathan ; VAN DER WALT, Stéfan J. ; BRETT, Matthew ; WILSON, Joshua ; MILLMAN, K. J. ; MAYOROV, Nikolay ; NELSON, Andrew R. J. ; JONES, Eric ; KERN, Robert ; LARSON, Eric ; CAREY, C J. ; POLAT, İlhan ; FENG, Yu ; MOORE, Eric W. ; VANDERPLAS, Jake ; LAXALDE, Denis ; PERKTOLD, Josef ; CORMAN, Robert ; HENRIKSEN, Ian ; QUINTERO, E. A. ; HARRIS, Charles R. ; ARCHIBALD, Anne M. ; RIBEIRO, Antônio H. ; PEDREGOSA, Fabian ; VAN MULBREGT, Paul ; SCIPY 1.0 CONTRIBUTORS: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. In: *Nature Methods* 17 (2020), S. 261–272. <http://dx.doi.org/10.1038/s41592-019-0686-2>. – DOI 10.1038/s41592-019-0686-2
- [Wachs 2006] WACHS, Michelle L.: Poset topology: tools and applications. In: *arXiv preprint math/0602226* (2006)
- [Wang et al. 2016] WANG, Iren ; CHEN, Szu-Yu ; HSU, Shang-Te D.: Folding analysis of the most complex Stevedore’s protein knot. In: *Scientific reports* 6 (2016), Nr. 1, S. 1–11
- [Yerole mou and Barbensi 2022] YEROLEMOU, Naya ; BARBENSI, Agnese: *Knotoid Distribution*. <https://github.com/nyerole mou/knotoid-distribution>, 2022

- [Yerolemou and Nanda 2022] YEROLEMOU, Naya ; NANDA, Vidit: Morse Theory for Complexes of Groups. In: *arXiv preprint arXiv:2203.00539* (2022)
- [Yerolemou et al. 2021] YEROLEMOU, Naya ; VIPOND, Oliver ; GOUNDAROULIS, Dimos: *KnotoEMD for Proteins*. <https://github.com/nyerolemou/proteins-knotoEMD>, 2021
- [Zhao et al. 2018] ZHAO, Yani ; DABROWSKI-TUMANSKI, Pawel ; NIEWIECZERZAL, Szymon ; SULKOWSKA, Joanna I.: The exclusive effects of chaperonin on the behavior of proteins with 5_2 knot. In: *PLoS computational biology* 14 (2018), Nr. 3, S. e1005970
- [Zibrowius 2016] ZIBROWIUS, Claudius: On a Heegaard Floer theory for tangles. In: *arXiv preprint arXiv:1610.07494* (2016)