



# Stratifying the space of barcodes using Coxeter complexes

Benjamin Brück<sup>1</sup> · Adélie Garin<sup>2</sup>

Received: 8 March 2022 / Revised: 23 September 2022 / Accepted: 24 September 2022 /  
Published online: 26 December 2022  
© The Author(s) 2022

## Abstract

Embeddings of the space of barcodes in Euclidean spaces are unstable due to the permutation of the bars of a barcode. We use tools from geometric group theory to produce a stratification of the space  $\mathcal{B}_n$  of barcodes with  $n$  bars that takes into account these permutations. This gives insights in the combinatorial structure of  $\mathcal{B}_n$ . The top-dimensional strata are indexed by permutations associated to barcodes as defined by Kanari, Garin and Hess. More generally, the strata correspond to marked double cosets of parabolic subgroups of the symmetric group  $\text{Sym}_n$ . This subdivides  $\mathcal{B}_n$  into regions that consist of barcodes with the same averages and standard deviations of birth and death times and the same permutation type. We obtain coordinates that form a new invariant of barcodes, extending the one of Kanari–Garin–Hess. This description also gives rise to metrics on  $\mathcal{B}_n$  that coincide with modified versions of the bottleneck and Wasserstein metrics.

**Keywords** Barcodes · Topological data analysis · Coxeter complex · Geometric group theory

**Mathematics Subject Classification** 51F15 · 55N31 · 20F55 · 57N80

## 1 Introduction

Barcodes Carlsson (2009); Edelsbrunner et al. (2008); Ghrist (2008) are topological summaries of the persistent homology of a filtered space. The barcode  $B$  associated to a *filtration*  $\{X_t\}_{t \in \mathbb{R}}$  is a multiset of points  $(b, d) \in \mathbb{R}^2$ . It summarises the creation and destruction of homology classes while varying the parameter  $t$ , which is often interpreted as “time”. A bar  $(b, d) \in B$  corresponds to a homology cycle appearing in

---

✉ Adélie Garin  
adelie.garin@gmail.com

<sup>1</sup> ETH Zurich, Zurich, Switzerland

<sup>2</sup> École Polytechnique Fédérale de Lausanne: Ecole Polytechnique Federale de, Lausanne, Switzerland

$X_b$  and becoming a boundary in  $X_d$ . The first element of the pair  $(b, d)$  is called the *birth* and the second one the *death*.

Persistent homology has applications in many fields, from biology Byrne et al. (2019); Gameiro et al. (2015); Kanari et al. (2018); Reimann et al. (2017) to material science Delgado-Friedrichs et al. (2014); Lee et al. (2018); Robins et al. (2016), astronomy Heydenreich et al. (2021) and climate science Muszynski et al. (2019). In many of these applications, it is necessary to study statistics on barcodes. Unfortunately, the space of barcodes is not a Hilbert space, which means that it can be difficult to apply statistical methods to it. Several ways to overcome the issue exist, such as the creation of kernels to map barcodes into a Hilbert space Bubenik (2015); Carrière et al. (2015); Adams et al. (2017); Di Fabio et al. (2015).

In this paper, we tackle this issue from a different perspective. We use combinatorial tools from geometric group theory to define new coordinates for describing barcodes. These coordinates divide the space of barcodes into regions indexed by the averages and the standard deviations of births and deaths and by the permutation type of a barcode as defined in Kanari et al. (2020); Curry et al. (2021). By associating to a barcode the coordinates of its region, we define a new invariant of barcodes. This opens the door to doing statistics on barcodes using methods from the field of permutation statistics.

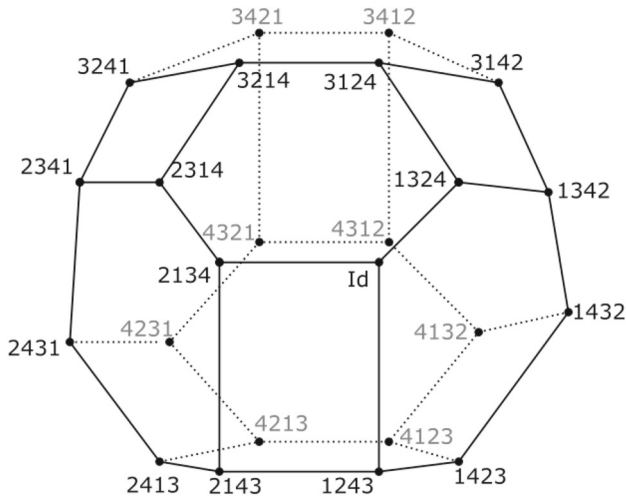
## 1.1 Motivation

The motivation for this work is to understand the space of barcodes from a combinatorial and geometric point of view and to show that it almost has a locally Euclidean structure.

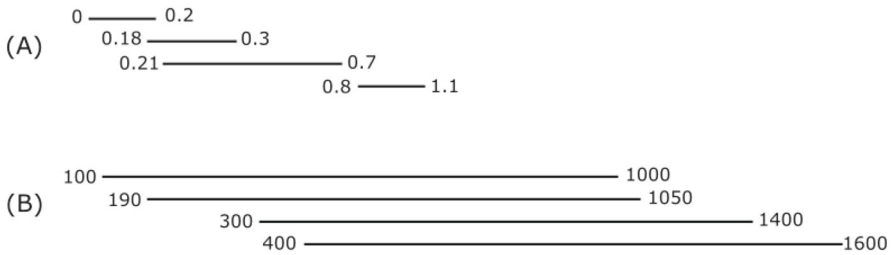
We call a barcode *strict* if there are no two pairs in it that have the same birth or death. It was observed in Kanari et al. (2020) that to a strict barcode  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}}$  with  $n$  bars, one can associate a permutation  $\sigma_B \in \text{Sym}_n$ . It is the permutation such that the bar with the  $i$ -th smallest death has the  $\sigma_B(i)$ -th smallest birth. This divides the set of strict barcodes with  $n$  bars into  $n!$  equivalence classes, one for each element of the symmetric group  $\text{Sym}_n$ . Based on this observation, one can study the combinatorial properties of strict barcodes by describing these equivalence classes—or equivalently, the elements of  $\text{Sym}_n$ —and the relations between them.

A first approach to this, taken in Kanari et al. (2020); Curry et al. (2021), is to consider the Cayley graph of the symmetric group with respect to the generating set given by adjacent transpositions  $(i, i + 1)$ . This yields a combinatorial representation of the elements of  $\text{Sym}_n$ . It tells us how a pair of permutations can be transformed into one another using transpositions one step at a time. However, it yields no information about “higher order relations” that exist among larger sets of permutations.

A way to resolve this is to add higher dimensional cells to the Cayley graph and to consider it more geometrically as a cell complex instead of as a (combinatorial) graph. A first approach would be to use that the Cayley graph of  $\text{Sym}_n$  is the 1-skeleton of the permutohedron Postnikov (2009) of order  $n$ , see Fig. 1. This observation embeds the Cayley graph into a polyhedral decomposition of the  $(n - 2)$ -sphere. As this is a more geometric object, it allows to continuously “walk” from one permutation to another.



**Fig. 1** The permutohedron Postnikov (2009) of order 4 is a polyhedral decomposition of the sphere where each vertex corresponds to an element of the symmetric group  $Sym_4$ . Its 1-skeleton is the Cayley graph of  $Sym_4$  (see also Fig. 5)



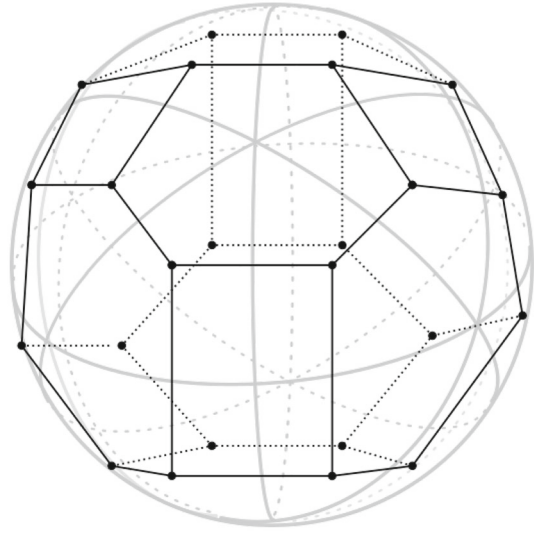
**Fig. 2** Two barcodes with the same associated permutation (the identity [1234]) but with large differences in their birth and death values

The problem is that only the vertices (and not the higher dimensional cells) of the permutohedron have an interpretation in terms of elements of the symmetric group. Furthermore, this representation lacks a notion of “size” for barcodes. For instance, the two barcodes depicted in Fig. 2 lie in the same equivalence class, i.e. have the same associated permutation.

The alternative that we suggest to overcome these problems is to work with *Coxeter complexes* instead of permutohedra. The Coxeter complex associated to  $Sym_n$  is the dual of the permutohedron of order  $n$  (see Fig. 3). It forms a simplicial decomposition of the  $(n - 2)$ -sphere and is well-studied in the context of reflection groups and Tits buildings.

For us, it has the advantage that its top-dimensional simplices correspond in a natural way to permutations and only passing through a face of lower dimension changes such a permutation. This allows for a better description of continuous changes between different permutations. It also has the advantage that it comes with an embedding in  $\mathbb{R}^n$ , where the additional two real parameters that are needed to describe positions

**Fig. 3** The permutohedron of order 4 (black) is the dual of the Coxeter complex  $\Sigma(\text{Sym}_4)$  (grey)



relative to this  $(n - 2)$ -dimensional space have a natural interpretation in terms of the “size” of barcodes. Moreover, using the Coxeter complex description for barcodes allows to define the permutation type of *any* barcode. For non-strict barcodes, it is defined only up to *parabolic subgroups* of  $\text{Sym}_n$ , i.e. subgroups that are generated by sets of adjacent transpositions.

## 1.2 Contributions

In this paper, we use Coxeter complexes to develop a description of the set  $\mathcal{B}_n$  of barcodes with  $n$  bars with coordinates that have natural interpretations when doing statistics with barcodes. These coordinates define a stratification of  $\mathcal{B}_n$  where the top-dimensional strata are indexed by the symmetric group  $\text{Sym}_n$ . Our main contributions can be summarised as follows.

**Theorem 1.1** *Let  $\mathcal{B}_n$  denote the set of barcodes with  $n$  bars.*

1.  $\mathcal{B}_n$  can in a natural way be seen as a subset of a quotient  $\text{Sym}_n \backslash \mathbb{R}^{2n}$ .
2.  $\mathcal{B}_n$  is stratified over the poset of marked double cosets of parabolic subgroups of  $\text{Sym}_n$ .
3. Using this description, one obtains a decomposition of  $\mathcal{B}_n$  into different regions. Each region is characterised as the set of all barcodes having the same average birth and death, the same standard deviation of births and deaths and the same permutation type  $\sigma_B \in \text{Sym}_n$ .
4. This description gives rise to metrics on  $\mathcal{B}_n$  that coincide with modified versions of the bottleneck and Wasserstein metrics.

Intuitively, this means that there are two types of small perturbations of a barcode. One is to perturb it such that one obtains another barcode with the same permutation

type. Such a perturbation takes place in a Euclidean subspace (a single stratum) of  $\mathcal{B}_n$ . The other is to change the permutation type and hence going from one Euclidean area (i.e. stratum) to another. For more detailed and formal statements of these results, see Proposition 4.2, Theorem 4.9, Corollary 4.10 and Proposition 5.2.

To obtain this description of  $\mathcal{B}_n$  we proceed as follows. A barcode is an (unordered) multiset of  $n$  pairs of real numbers (births and deaths). It can hence be seen as a point in the quotient space  $\text{Sym}_n \setminus (\mathbb{R}^n \times \mathbb{R}^n)$ , where the action of  $\text{Sym}_n$  permutes the coordinate pairs. Since the birth is smaller than the death for every barcode,  $\mathcal{B}_n$  is a proper subset of this quotient of  $\mathbb{R}^{2n}$ .

The Coxeter complex  $\Sigma(\text{Sym}_n)$  associated to  $\text{Sym}_n$  is a simplicial complex whose geometric realisation is homeomorphic to an  $(n - 2)$ -sphere. Hence, we can decompose  $\mathbb{R}^n$  as

$$\mathbb{R}^n \cong \text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R},$$

where  $\text{cone}(\Sigma(\text{Sym}_n)) = (\Sigma(\text{Sym}_n) \times [0, \infty)) / (x, 0) \sim (y, 0) \cong \mathbb{R}^{n-1}$ . This decomposition allows one to describe each point  $x \in \mathbb{R}^n$  via coordinates  $x_\theta, \bar{x}, \|v_x\|$ , where  $x_\theta$  specifies a point on the Coxeter complex,  $\|v_x\|$  is the ‘‘cone parameter’’ and  $\bar{x}$  parametrises the remaining  $\mathbb{R}$  (for details, see Proposition 3.2, where the naming becomes clear as well). In summary, this describes  $\mathcal{B}_n$  as a subset of

$$\mathcal{B}_n \subset \text{Sym}_n \setminus ((\text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R} \times \text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R})).$$

We call the coordinates that we obtain from this description *Coxeter coordinates*. It turns out that for each barcode, these coordinates are  $b_\theta, \bar{b}, \|v_b\|$  and  $d_\theta, \bar{d}, \|v_d\|$ , where  $\bar{b}$  and  $\bar{d}$  are the averages of the births and deaths,  $\|v_b\|$  and  $\|v_d\|$  are their standard deviations and the coordinates  $b_\theta$  and  $d_\theta$  describe the permutation equivalence class of the barcode of Kanari et al. (2020); Curry et al. (2021). The stratification one obtains is induced by the simplicial structure of  $\Sigma(\text{Sym}_n)$ . Hence, each stratum is Euclidean.

The advantages of these new coordinates are two-fold: Firstly, using points in Coxeter complexes, one obtains coordinates that uniquely specify barcodes and are yet compatible with the combinatorial structure of  $\mathcal{B}_n$  given by permutation equivalence classes. Secondly, one resolves the earlier-mentioned problem that permutation equivalence classes themselves carry no notion of ‘‘size’’: The decomposition of  $\mathcal{B}_n$  into regions subdivides these equivalence classes by also taking into account the averages and standard deviations of births and deaths. This makes these regions a finer invariant than the permutation type. Therefore, they offer a new way to study statistics of barcodes by using both the average and standard deviation of births and deaths, which are commonly used summaries in Topological Data Analysis (TDA), and permutation statistics tools. The latter include the number of descents for instance, or the inversion numbers, which have proven useful for the study of the inverse problem for trees and barcodes Kanari et al. (2020); Curry et al. (2021).

### 1.3 Related work

This paper is a follow-up of the work started in Kanari et al. (2020); Curry et al. (2021) to study the space of barcodes from a combinatorial point of view. It extends the approach of considering permutations to classify barcodes to a finer classification that also takes into account the average and standard deviation of births and deaths. In Xu (2020), the author also observes a connection between barcodes and the symmetric group in a different setting, by studying the space of barcode bases using Schubert cells. Similarly, Jacquard et al. (2022) also studies the space of barcode bases.

The idea of giving coordinates to the space of barcodes is not new Di Fabio et al. (2015); Kališnik (Feb 2019). For example, the space of barcodes was given tropical coordinates in Kališnik (Feb 2019). In Adcock et al. (2013), it is mentioned that the space of barcodes can be identified with the  $n$ -fold symmetric product of  $\mathbb{R}^2$ , and the authors study the corresponding algebra of polynomials associated to the variety.

Finally, defining a polyhedral structure on a space to study statistics has been done for spaces of (phylogenetic) trees Billera et al. (2001); Grindstaff and Owen (2018). The connection between phylogenetic trees, merge trees and barcodes is studied in Curry et al. (2021). The polyhedral structure defined in this paper and in Billera et al. (2001) seem to be related, but we leave this as future work.

### 1.4 Overview

In Sect. 2 we review the necessary background on barcodes and on Coxeter complexes. We use a standard way of realising  $\text{Sym}_n$  as a reflection group to explain what we mean with “Coxeter coordinates” on  $\mathbb{R}^n$  in Sect. 3. We then describe the space  $\mathcal{B}_n$  of barcodes with  $n$  bars in terms of  $\text{Sym}_n \setminus \mathbb{R}^n \times \mathbb{R}^n$  in Sect. 4.1, before adapting the coordinates of  $\mathbb{R}^n$  to  $\mathcal{B}_n$  in Sect. 4.2. In Sect. 4.3, we describe the stratification of  $\mathcal{B}_n$  induced by these coordinates. Corollary 4.10 decomposes the space of barcodes into regions indexed by the average and standard deviation of the births and deaths and the permutation associated to a barcode. Finally, in Sect. 5, we show that  $\mathcal{B}_n$  can be given metrics inspired by the bottleneck and Wasserstein distances and that it defines an isometry between a subset of  $\text{Sym}_n \setminus \mathbb{R}^n \times \mathbb{R}^n$  and  $\mathcal{B}_n$ .

## 2 Background

### 2.1 Background on TDA

We start by reviewing the necessary background on TDA. For the reader who is completely new to this, we refer to the reviews Carlsson (2009); Edelsbrunner et al. (2008); Ghrist (2008). Even though this work focuses on the space of barcodes and could be apprehended from a purely combinatorial point of view, we shortly mention where barcodes arise in the field of TDA. This section is not necessary for the understanding of this paper, and we will give the combinatorial definition of barcodes that we use in the next section.

Barcodes are topological summaries of a *filtered topological space*, i.e. a sequence of spaces ordered by inclusion. To obtain a barcode from a filtered space, one computes homology at each step and considers the maps induced by the inclusions. The output is called a *persistence module*, and it summarises the evolution of the homology at each step of the filtration.

More precisely, let  $\{X_t\}_{t \in \mathbb{R}}$  be a filtered topological space, that is, each  $X_t$  is a topological space and  $X_t \subseteq X_{t'}$  if  $t \leq t'$ . The  $k$ -th persistence module associated to  $\{X_t\}_{t \in \mathbb{R}}$  is given by  $\mathbb{H}_k(\{X_t\}_{t \in \mathbb{R}})$ , where  $\mathbb{H}_k$  denotes the  $k$ -th homology functor (over a field  $\mathbb{k}$ ). The Crawley–Bovey Theorem Crawley-Boevey (2015) states that under mild tameness conditions on  $\{X_t\}_{t \in \mathbb{R}}$ , the associated persistence module can be decomposed as a direct sum of interval modules  $\bigoplus_{j \in \mathcal{J}} \mathbb{k}_{I_j}^{\oplus n_j}$ , where the *interval module*  $\mathbb{k}_{I_j}$  is the free  $\mathbb{k}$ -module of rank 1 on the interval  $I_j \subseteq \mathbb{R}$ , with identity maps internal to  $I_j$ , and is 0 elsewhere. This decomposition is unique up to reordering. Each interval represents the lifetime of a cycle in the filtered space. For instance, if a 1-cycle (a loop) appears in the topological space  $X_{b_j}$  for the first time and becomes a boundary (gets “filled in”) in  $X_{d_j}$ , then this 1-cycle will be represented by the interval  $I_j = [b_j, d_j)$ . The *barcode* associated to the persistence module is the multiset

$$B = \{I_j\}_{j \in \mathcal{J}},$$

where each interval  $I_j$  appears  $n_j$  times. Usually, each  $I_j$  is a half open interval  $I_j = [b_j, d_j)$ , where  $b_j$  is called the *birth* of the homological feature corresponding to  $I_j$  and  $d_j$  is called its *death*. If the interval  $I_j$  is a half infinite interval, i.e. it is of the form  $[b_i, \infty)$ , it is called an *essential class*.

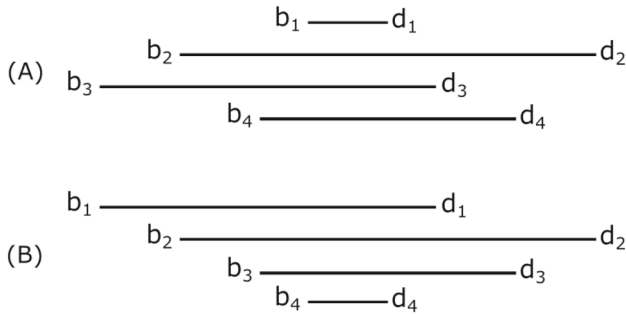
In this paper, we will identify such an interval with the pair  $(b_j, d_j)$ , since we are mostly interested in the combinatorics of the pairs and not the corresponding persistence module. Moreover,  $b_j$  and  $d_j$  will always take finite values in  $\mathbb{R}$ .

### 2.1.1 The space of barcodes

We introduce here the main definitions used in this paper. We start by a more combinatorial definition of barcodes that we will use in this article.

**Definition 2.1** A *barcode*  $\{(b_i, d_i)\}_{i \in J}$  is a multiset of pairs  $(b_i, d_i) \in \mathbb{R}^2$  such that  $b_i < d_i$  for each  $i \in J$  and  $|J| < \infty$ . Each such pair is called a *bar*; its first coordinate  $b_i$  is called the *birth* (time) and the second one  $d_i$  is called its *death* (time). A barcode is called *strict* if  $b_i \neq b_j$  and  $d_i \neq d_j$  for  $i \neq j$ . We let  $\mathcal{B}_n$  denote the set of barcodes with  $n$  bars and  $\mathcal{B}_n^{st}$  the set of strict barcodes with  $n$  bars.

**Remark 2.2** The reader familiar with persistent homology will notice that we suppose that the bars corresponding to essential classes have finite values instead of being half-open intervals. This is usually the case in practical applications, where such essential classes are given finite values for representing them on a computer. We also assume that every barcode consists of only finitely many bars.



**Fig. 4** **A** A barcode with 4 bars. **B** The same barcode with a different indexing where the bars are ordered by increasing birth times

**Remark 2.3** The definition of strict barcodes was first introduced in Kanari et al. (2020) to define the bijection between the symmetric group on  $n$  elements and some equivalence classes of barcodes that we introduce in the next section. The setting in this paper is slightly different from Kanari et al. (2020) and Curry et al. (2021), because all the barcodes considered there are specific to merge trees and arise from their 0-th persistent homology. This is why the definition of a strict barcode in Kanari et al. (2020) and Curry et al. (2021) assumes the existence of an essential bar  $(b_0, d_0)$  that contains all the others. In this paper however, barcodes can come from arbitrary filtrations in arbitrary dimension, and such a bar  $(b_0, d_0)$  need not exist. Therefore we slightly adapt the definition of a strict barcode and the relation to the symmetric group in the next sections.

In practice, for finite barcodes, the indexing set  $J$  is commonly the set  $\{1, \dots, n\}$ , giving the bars in the barcode an arbitrary but fixed ordering. We will also adopt this convention from now on. Note however that reordering the bars might change the indexing, but not the underlying barcode (see Example 2.4). It can sometimes be convenient to assume that the indexing is such that the births are ordered increasingly  $b_1 < b_2 < \dots < b_n$ , but we do not make this assumption in this paper unless specified.

We often represent a barcode by the set of intervals  $[b_i, d_i] \subset \mathbb{R}$  (as in Fig. 4). Another common way to represent barcodes is what is called a *persistence diagram*, where the pairs  $(b_i, d_i)$  are represented as points in  $\mathbb{R}^2$  (as in Fig. 8). These points lie above the diagonal since  $b_i < d_i$  for all  $i$ .

**Example 2.4** Figure 4 shows an example of a strict barcode with two different indexing conventions.

To turn the set of barcodes into a topological space, one needs to specify a topology. One option to do this is by introducing the bottleneck or Wasserstein distances, two commonly used metrics for barcodes. Intuitively, the bottleneck distance between two barcodes  $B$  and  $B'$  tries all possible matchings between the bars of  $B$  and the bars of  $B'$  and chooses the one that minimises the “energy” required to move the matched pair of bars with maximal separation. However, it does not only consider matching of bars between  $B$  and  $B'$  but also with points on the diagonal  $\Delta = \{(x, x) \mid x \in \mathbb{R}\}$ .



**Definition 2.5** Let  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}}$  and  $B' = \{(b'_i, d'_i)\}_{i \in \{1, \dots, m\}}$  be two barcodes. The *bottleneck distance* between  $B$  and  $B'$  is

$$d_B(B, B') = \min_{\gamma} \max_{x \in B} \|x - \gamma(x)\|_{\infty},$$

where  $\gamma$  runs over all possible matchings, i.e. maps that assign to each bar  $(b_i, d_i) \in B$  either a bar in  $B'$  or a point in the diagonal  $\Delta$ , such that no point of  $B'$  is in the image more than once. Here,  $\|\cdot\|_{\infty}$  is the  $l^{\infty}$ -norm on  $\mathbb{R}^2$ .

**Remark 2.6** The permutation  $\gamma$  acts as a “reindexing” of the indices of  $B$  and  $B'$ , and in particular ensures that  $d_B(B, B')$  does not depend on any indexing of the bars.

The *Wasserstein distance* is defined in a similar way by taking the sum over all  $l_2$ -distances between  $x$  and  $\gamma(x)$  instead:

$$d_W(B, B') = \min_{\gamma} \sqrt{\sum_{x \in B} \|x - \gamma(x)\|_2^2}.$$

**Remark 2.7** Note that in general, the barcodes  $B$  and  $B'$  need not have the same number of bars. The diagonal allows matchings between barcodes with different number of bars, since “unmatched” bars can be sent to the diagonal. In this paper however, we study the set of barcodes  $\mathcal{B}_n$  with exactly  $n$  bars (for arbitrary, but fixed  $n$ ) and restrict ourselves to this case.

We are mainly interested in  $\mathcal{B}_n$  as a set and the main results we prove do not depend on the metric that is chosen on  $\mathcal{B}_n$ . We will still with a slight abuse of notation mostly talk of  $\mathcal{B}_n$  as a *space*, without specifying a specific metric on it. An exception to that is Sect. 5, where we explain how a metric  $\tilde{d}_B$  on  $\mathcal{B}_n$ , which is closely related to the bottleneck distance, occurs in an alternative description of the set  $\mathcal{B}_n$  that we work with later on.

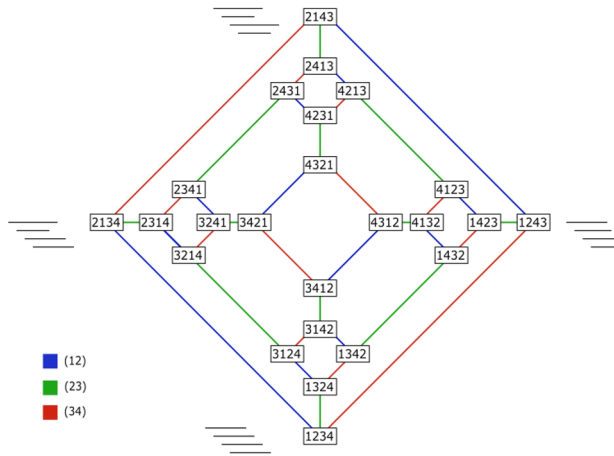
### 2.1.2 Relation to the symmetric group

We write  $\text{Sym}_n$  for the symmetric group on  $n$  letters, i.e. the group of all permutations of  $\{1, \dots, n\}$ . We usually use the one-line notation for permutations. That is, we specify  $\sigma \in \text{Sym}_n$  by the its image of the ordered set  $\{1, \dots, n\}$ , e.g. we write  $\sigma = [132] \in \text{Sym}_3$  if  $\sigma(1) = 1, \sigma(2) = 3$  and  $\sigma(3) = 2$ . We make an exception for transpositions to simplify the notation: the transposition that switches  $i$  and  $j$  is denoted by  $(i, j)$ .

**Definition 2.8** Kanari et al. (2020) Let  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}} \in \mathcal{B}_n^{st}$  be a strict barcode. If we order the births increasingly such that  $b_{i_1} < \dots < b_{i_n}$ , the indexing in  $\{1, \dots, n\}$  gives a permutation  $\tau_b$  by  $\tau_b(k) = i_k$ , i.e.  $\tau_b$  is the (unique) permutation such that

$$b_{\tau_b(1)} < \dots < b_{\tau_b(n)}. \tag{1}$$

Similarly, ordering the deaths  $d_{j_1} < \dots < d_{j_n}$  gives rise to a permutation  $\tau_d$  with  $\tau_d(k) = j_k$ . The *permutation  $\sigma_B$  associated to  $B$*  is defined as  $\sigma_B = \tau_b^{-1} \tau_d$ ; it tracks the ordering of the death values with respect to the birth values.



**Fig. 5** (Figure from Kanari et al. (2020)) The Cayley graph of  $Sym_4$  generated by the three transpositions (12), (23), (34). Four barcodes are drawn next to the extremities of the graphs (permutations [1234], [2134], [2143], [1243]) to illustrate a typical barcode corresponding to each permutation

**Remark 2.9** The permutations  $\tau_b$  and  $\tau_d$  both depend on the indexing choice of the  $b_i$  and  $d_i$ . However, the permutation  $\sigma$  does not depend on any indexing of the births and deaths, it is intrinsic to the multiset  $B$ . Indeed,  $\sigma_B$  can be defined directly as the permutation that sends the  $i$ -th death (in increasing order) to the  $\sigma(i)$ -th birth (idem). If we assume that the births are ordered increasingly, then  $\tau_b = \text{id}$  and  $\sigma_B$  can be defined directly by  $\sigma_B = [j_1 j_2 \dots j_n]$ , the indices of the deaths when they are ordered increasingly.

**Example 2.10** Figure 4A shows an example of a strict barcode. Its birth permutation is  $\tau_b = [3241]$ , since

$$b_3 < b_2 < b_4 < b_1.$$

Similarly, its death permutation is  $\tau_d = [1342]$ , since  $d_1 < d_3 < d_4 < d_2$ . The permutation  $\sigma_B$  associated to the barcode of Fig. 4A is  $\sigma_B = [4132] = \tau_b^{-1} \tau_d$ . Figure 4B shows the same barcode with the bars ordered by birth times. The corresponding permutations  $\tau_b = [1234]$  and  $\tau_d = [4132]$  are different, but the product  $\sigma_B = \tau_b^{-1} \tau_d = [4132]$  is the same, as it does not depend on the indexing of the bars. Further examples are depicted in Fig. 5.

We extend Definition 2.8 to non-strict barcodes in Sect. 4.3.

## 2.2 Background on Coxeter groups and complexes

### 2.2.1 Coxeter groups

*Coxeter groups* form a family of groups that was defined by Tits in its modern form. They are abstract versions of reflection groups; in fact, the family of finite Coxeter

groups coincides with the family of finite reflection groups. Besides their close connections to geometry and topology Davis et al. (2015), Coxeter groups have a rich combinatorial theory Björner et al. (2005). They appear in many areas of mathematics, e.g. as Weyl groups in Lie theory. We will view  $\text{Sym}_n$  as one of the most basic examples of a Coxeter group.

Usually, one does not consider a Coxeter group  $W$  by itself but instead a *Coxeter system*  $(W, S)$ , where  $S$  is a generating set of  $W$  that consists of involutions called the *simple reflections*. In what follows, we will tacitly assume that such a set of simple reflections is always fixed when we talk about a Coxeter group  $W$ . In the case where  $W = \text{Sym}_n$ , we will take  $S$  to be the set of adjacent transpositions  $S = \{(i, i + 1) \mid 1 \leq i \leq n - 1\}$ . A rank- $(|S| - 1 - k)$  (standard) *parabolic subgroup* of  $W$  is a subgroup of the form  $P_T = \langle T \rangle$ , where  $T \subset S$  is a subset of size  $(|S| - 1 - k)$ .

### 2.2.2 Coxeter complexes

Each Coxeter group  $W$  can be assigned a simplicial complex  $\Sigma(W)$ , the *Coxeter complex*, that is equipped with an action of  $W$ . If  $W$  is a finite group with set of simple reflections  $S$ , the complex  $\Sigma(W)$  is a triangulation of a sphere of dimension  $|S| - 1$ . Coxeter complexes have nice combinatorial properties and are in particular colourable flag complexes Abramenko et al. (2008) [Sect. 1.6] that are shellable Björner (1984).

The top-dimensional simplices of  $\Sigma(W)$  are in one-to-one correspondence with the elements of the group  $W$ . Furthermore, one recovers the Cayley graph of  $(W, S)$  as the *chamber graph* of  $\Sigma(W)$ , i.e. the graph that has a vertex for each top-dimensional simplex of  $\Sigma(W)$  and an edge connecting two vertices if the corresponding simplices share a codimension-1 face Abramenko et al. (2008) [Corollary 1.75].

More generally, the set of  $k$ -simplices in  $\Sigma(W)$  is in one-to-one correspondence with the cosets of rank- $(|S| - 1 - k)$  parabolic subgroups of  $W$ :

**Definition 2.11** The *Coxeter complex*  $\Sigma(W)$  of the Coxeter system  $(W, S)$  is defined as the simplicial complex

$$\Sigma(W) = \bigcup_{T \subseteq S} W/P_T = \{\tau P_T \mid \tau \in W, T \subseteq S\},$$

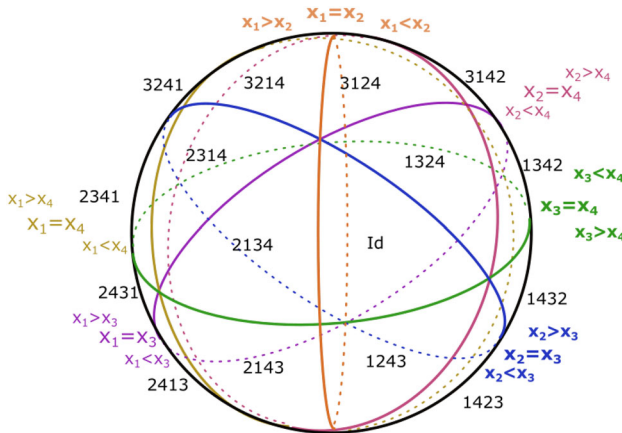
where each simplex  $\tau P_T$  has dimension<sup>1</sup>  $\dim(\tau P_T) = |S \setminus T| - 1$  and the face relation is defined by the partial order

$$\tau P_T \leq \tau' P_{T'} \Leftrightarrow \tau P_T \supseteq \tau' P_{T'}. \tag{2}$$

The group  $W$  acts simplicially on  $\Sigma(W)$  by left multiplication on the cosets,  $\gamma \cdot (\tau P) := \gamma\tau P$ .

**Remark 2.12** With a slight abuse of notation, we will in what follows often use the cosets  $\tau P$  to also denote simplices in the geometric realisation of the Coxeter complex.

<sup>1</sup> Note that we take the (combinatorial) convention that this simplicial complex has a unique face of dimension  $-1$ . This face does not appear in the geometric realisation.



**Fig. 6** The geometric realisation of the Coxeter complex  $\Sigma(\text{Sym}_4)$ . The permutation corresponding to each triangle of the front of the sphere is indicated in black. The hyperplanes  $x_i = x_j$  depicted in colours correspond to the transpositions  $(i, j)$ . The hyperplanes corresponding to adjacent transpositions  $(i, i + 1)$  are in boldface. A detailed description of how to obtain such a geometric realisation of the Coxeter complex can be found in Sect. 3

To be coherent with the definition of a stratification (Definition 2.13), we will always consider these simplices to be closed.

### 2.2.3 The Coxeter complex

$\Sigma(\text{Sym}_n)$  For the case  $W = \text{Sym}_n$  that we are interested in, the Coxeter complex  $\Sigma(\text{Sym}_n)$  is of dimension  $n - 2$  and is isomorphic to the barycentric subdivision of the boundary of an  $(n - 1)$ -simplex. It can be realised geometrically as a triangulation of the  $(n - 2)$ -sphere. This complex is the dual to the permutohedron of order  $n$  (see Fig. 3). Figure 6 depicts the Coxeter complex  $\Sigma(\text{Sym}_4)$ . The top-dimensional simplices of  $\Sigma(\text{Sym}_n)$  are in one-to-one correspondence with the elements of  $\text{Sym}_n$ . Two such simplices share a codimension-1 face if and only if the corresponding permutations differ by precomposing with an adjacent transposition  $(i, i + 1)$ , i.e. by exchanging two neighbouring entries of the permutation. As a consequence, if  $x$  lies in the interior of a maximal simplex of the geometric realisation of  $\Sigma(\text{Sym}_n)$ , it can be assigned a permutation  $\tau \in \text{Sym}_n$ . If  $x$  lies on a face of dimension  $k$ , then  $\tau$  is well-defined only up to applying an element of a parabolic subgroup  $P \leq \text{Sym}_n$  that is generated by  $|S| - 1 - k = n - 2 - k$  adjacent transpositions. A concrete embedding of  $\Sigma(\text{Sym}_n)$  in  $\mathbb{R}^n$  will be described in more detail in Sect. 3.

For later reference, we note that the identification  $S^{n-2} \cong \Sigma(\text{Sym}_n)$  gives a stratification of the sphere by its simplicial decomposition. The strata are the (closed) simplices of the geometric realisation and the stratification is over the partially ordered set (poset) specified by Eq. (2).

**Definition 2.13** Bridson et al. (1999) A set  $X$  is stratified over a poset  $\mathcal{P}$  if there exists a collection of subsets  $\{X_i\}_{i \in \mathcal{P}}$  of  $X$  such that:

1.  $X = \bigcup_i X_i$ ;
2.  $i \leq j$  if and only if  $X_i \subseteq X_j$ ;
3. If  $X_i \cap X_j \neq \emptyset$ , then it is a union of strata;
4. For every  $x \in X$ , there exists a unique  $i_x \in \mathcal{P}$  such that  $\bigcap_{X_i \ni x} X_i = X_{i_x}$ .

Each  $X_i$  is called a *stratum*.

### 3 Coxeter complex coordinates on $\mathbb{R}^n$

In this section, we describe  $\mathbb{R}^n$  as the product of a cone over the Coxeter complex  $\Sigma(\text{Sym}_n)$  with a 1-dimensional space orthogonal to it. This description is obtained by describing a standard way for realising  $\text{Sym}_n$  as a reflection group Abramenko et al. (2008) [Example 1.11]. In terms of Coxeter groups, this is often called the “dual representation”, see e.g. Abramenko et al. (2008) [Sect. 2.5.2]. Example 3.4 below goes through the following steps in detail for the case  $n = 3$ .

In what follows, we will consider  $\mathbb{R}^n$  with the  $l^2$ -norm  $\|\cdot\|$  that is induced by the standard scalar product  $\langle \cdot, \cdot \rangle$ . We let  $e_1, \dots, e_n$  denote the standard basis. The symmetric group  $\text{Sym}_n$  acts on  $\mathbb{R}^n$  by permuting this standard basis. This action can be expressed in coordinates as

$$\gamma \cdot (x_1, \dots, x_n) = (x_{\gamma^{-1}(1)}, \dots, x_{\gamma^{-1}(n)}). \tag{3}$$

It is norm-preserving and fixes the 1-dimensional subspace  $L = \langle e \rangle$  spanned by  $e := e_1 + \dots + e_n = (1, \dots, 1)$ . Hence, there is an induced action on the orthogonal complement  $V = e^\perp$ , which can be described as

$$V = \{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 0 \}.$$

Note that  $L$  is the subspace consisting of all  $(x_1, \dots, x_n) \in \mathbb{R}^n$  where  $x_i = x_j$  for all  $i, j$ . So in particular, every  $(x_1, \dots, x_n) \in \mathbb{R}^n \setminus L$  has at least two coordinates that are different from one another.

The subspace  $V$  has a natural structure of a cone over the Coxeter complex  $\Sigma(\text{Sym}_n)$  associated to  $\text{Sym}_n$ , see Remark 3.3. The transposition  $(i, j) \in \text{Sym}_n$  acts on  $V$  by orthogonal reflection along the hyperplane

$$\{ (x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i = x_j \},$$

permuting the  $i$ -th and  $j$ -th coordinates. Let  $\mathcal{H}$  be the collection of all these hyperplanes, and let  $S_r$  denote the  $(n - 2)$ -sphere of radius  $r > 0$  around the origin in  $V$  (with respect to the norm induced by the restriction of the standard scalar product on  $\mathbb{R}^n$ ), i.e.  $S_r = \{ v \in V \mid \|v\| = r \}$ .

**Lemma 3.1** (Abramenko et al. (2008) [Examples 1.10, 1.4.7 & 1.81]) *The hyperplanes  $\mathcal{H}$  induce a triangulation of  $S_r$ . The resulting simplicial complex  $\Sigma$  is isomorphic to the Coxeter complex  $\Sigma(\text{Sym}_n)$  as  $\text{Sym}_n$ -spaces.*

The set of points  $x \in \mathbb{R}^n$  such that all coordinates are different is the *configuration space*

$$\text{Conf}_n(\mathbb{R}) = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid i \neq j \implies x_i \neq x_j\}.$$

The previous lemma describes how a permutation in  $\text{Sym}_n$  can be associated to each point  $x \in \text{Conf}_n(\mathbb{R})$ . To understand why this is true, observe that if  $C$  is a connected component of  $S_r \setminus \bigcup \mathcal{H}$ , then for all  $(x_1, \dots, x_n) \in C$ :

- If  $i \neq j$ , then  $x_i \neq x_j$ , i.e.  $(x_1, \dots, x_n) \in \text{Conf}_n(\mathbb{R})$ ;
- If  $(y_1, \dots, y_n) \in C$ , then  $y_i < y_j$  if and only if  $x_i < x_j$ .

In particular, there is a unique  $\tau \in \text{Sym}_n$  such that

$$(x_1, \dots, x_n) \in C \iff x_{\tau(1)} < x_{\tau(2)} < \dots < x_{\tau(n)}. \tag{4}$$

In other words, the order of the elements  $x_1, \dots, x_n$  is given by  $\tau((1, \dots, n))$ , see Fig. 6 above for the case  $n = 4$ . The connected components of  $S_r \setminus \bigcup \mathcal{H}$  are exactly the (interiors of) the maximal simplices of  $\Sigma$ . Sending each such component  $C$  to the facet of  $\Sigma(\text{Sym}_n)$  that corresponds to the permutation  $\tau$  defined by Eq. 4 gives the desired isomorphism  $\Sigma \cong \Sigma(\text{Sym}_n)$ .

Using spherical coordinates, we can express every point  $v \in V$  in terms of a radial component  $r > 0$  and an angular component, which is equivalent to specifying a point  $v_\theta \in S_r$  (i.e. a point in the geometric realisation of  $\Sigma(\text{Sym}_n)$ ). The upshot of this is that we obtain a new set of coordinates for points in  $\mathbb{R}^n \setminus L$ .

**Proposition 3.2** *Let  $n \geq 2$ . There exist two projection maps*

$$p : \mathbb{R}^n \longrightarrow \mathbb{R} \times \mathbb{R}_{\geq 0} : x \mapsto, (\bar{x}, \|v_x\|),$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  and  $\|v_x\| = (\sum_{i=1}^n |x_i - \bar{x}|^2)^{1/2}$ , and

$$q : \mathbb{R}^n \setminus L \longrightarrow \Sigma(\text{Sym}_n)$$

that define a bijection

$$(p|_{\mathbb{R}^n \setminus L}, q) : \mathbb{R}^n \setminus L \longrightarrow \mathbb{R} \times \mathbb{R}_{> 0} \times \Sigma(\text{Sym}_n).$$

Let  $\text{Sym}_n$  act on  $\mathbb{R}^n$  by permuting the coordinates (Eq. 3) and on the product  $\mathbb{R} \times \mathbb{R}_{> 0} \times \Sigma(\text{Sym}_n)$  by extending the action on  $\Sigma(\text{Sym}_n)$  trivially on the first two factors. Then the map  $(p|_{\mathbb{R}^n \setminus L}, q)$  is  $\text{Sym}_n$ -equivariant.

**Proof** For every  $x \in \mathbb{R}^n$ , the orthogonal decomposition  $\mathbb{R}^n = \langle e \rangle \oplus V$  gives a unique way to write  $x = \bar{x} \cdot e + v_x$  with  $\bar{x} \in \mathbb{R}$  and  $v_x \in V$ , where

$$\bar{x} = \frac{\langle e, x \rangle}{\langle e, e \rangle} = \sum_{i=1}^n x_i/n = \frac{1}{n} \sum_{i=1}^n x_i.$$

We can describe the projection  $v_x = x - \bar{x} \cdot e \in V$  in spherical coordinates. Its norm (the radius of the sphere) is

$$\|v_x\| = \|x - \bar{x} \cdot e\| = \left( \sum_{i=1}^n |x_i - \bar{x}|^2 \right)^{1/2},$$

so  $v_x$  is determined by this value together with a point  $x_\theta$  on the  $(n - 2)$ -sphere  $S_{\|v_x\|}$ , or equivalently on the geometric realisation of  $\Sigma(\text{Sym}_n)$ . Notice that  $x \in L$  if and only if  $v_x = 0$ , as the line  $L$  intersects  $V$  at its origin.

We define the map  $p : \mathbb{R}^n \rightarrow \mathbb{R} \times \mathbb{R}_{\geq 0} : x \mapsto (\bar{x}, \|v_x\|)$  and the map  $q : \mathbb{R}^n \setminus L \rightarrow S^{n-2} : x \mapsto x_\theta$ . The point  $x_\theta$  is well-defined since  $x \notin L$  and therefore there exist  $i, j$  such that  $x_i \neq x_j$ . It is easy to see that  $(p|_{\mathbb{R}^n \setminus L}, q)$  is a bijection, i.e. that given  $c_1 \in \mathbb{R}, c_2 \in \mathbb{R}_{>0}$  and  $c_3 \in \Sigma(\text{Sym}_n)$ , there is a unique  $x \in \mathbb{R}^n \setminus L$  such that  $c_1 = \bar{x}, c_2 = \|v_x\|$  and  $c_3 = x_\theta$ .

The fact that  $(p|_{\mathbb{R}^n \setminus L}, q)$  is  $\text{Sym}_n$ -equivariant follows from Lemma 3.1 and because permuting the coordinates of  $x \in \mathbb{R}^n$  changes neither the average  $\frac{1}{n} \sum_i x_i$  nor the standard deviation  $(\sum_i |x_i - \bar{x}|^2)^{1/2}$ . □

To summarise, every point  $x = (x_1, \dots, x_n) \in \mathbb{R}^n \setminus L$  determines the following three things:

1. Its projection to  $L$ , given by  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \in \mathbb{R}$ ;
2. The norm of its projection to  $V$ , given by  $\|v_x\| = (\sum_{i=1}^n |x_i - \bar{x}|^2)^{1/2} \in \mathbb{R}_{>0}$ ;
3. A point  $x_\theta$  in the geometric realisation of the Coxeter complex  $\Sigma(\text{Sym}_n)$  associated to  $\text{Sym}_n$ .

Furthermore,  $x$  is uniquely determined by these three coordinates.

**Remark 3.3** There is an isomorphism  $\mathbb{R}_{>0} \times \Sigma(\text{Sym}_n) \cong \text{cone}(\Sigma(\text{Sym}_n)) \setminus \{*\}$ , where

$$\text{cone}(\Sigma(\text{Sym}_n)) = (\Sigma(\text{Sym}_n) \times [0, \infty)) / (x, 0) \sim (y, 0)$$

and  $*$  is the cone point, i.e. the equivalence class of  $(x, 0)$ . Since  $\mathbb{R}^n = \mathbb{R}^n \setminus L \sqcup L$ , the above map  $(p|_{\mathbb{R}^n \setminus L}, q)$  gives rise to a decomposition  $\mathbb{R}^n \cong \text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R}$ . Indeed, the line  $L \subset \mathbb{R}^n$  corresponds to points  $x \in \mathbb{R}^n$  with  $v_x = 0$ , which could be seen as “spheres of radius 0” in the projection  $q$ .

**Example 3.4** We go through the previous construction in detail for the case of  $\mathbb{R}^3$  equipped with the natural action of the symmetric group  $\text{Sym}_3$ , illustrating the example in Fig. 7. Consider  $\mathbb{R}^3 = \langle e_1, e_2, e_3 \rangle$ . The symmetric group  $\text{Sym}_3$  acts on it by permuting the coordinates of each vector  $(x_1, x_2, x_3)$ :

$$\gamma \cdot (x_1, x_2, x_3) = (x_{\gamma^{-1}(1)}, x_{\gamma^{-1}(2)}, x_{\gamma^{-1}(3)}).$$

Each  $\gamma \in \text{Sym}_3$  can be written as a product of transpositions  $(i, j)$  and its action on  $\mathbb{R}^3$  is given by the performing the corresponding sequence of reflections along the

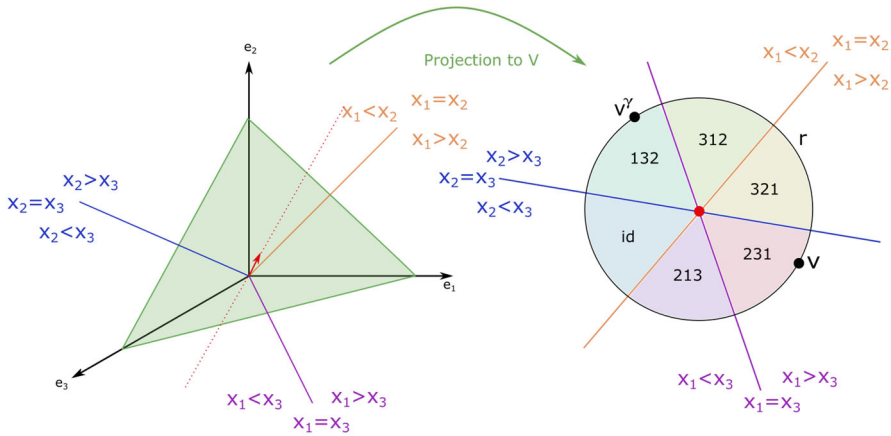


Fig. 7 Example of the decomposition of  $\mathbb{R}^3$  in Coxeter coordinates

hyperplanes  $x_i = x_j$ . The three (2-dimensional) planes corresponding to the equations  $x_1 = x_2$ ,  $x_2 = x_3$  and  $x_1 = x_3$  are indicated as lines on the left hand side of Fig. 7 to make the picture clearer. The subspace  $L$  that is invariant under this action is spanned by the vector  $(1, 1, 1) = e$ , shown in red in Fig. 7.

We can define new coordinates on  $\mathbb{R}^3$ , lying in  $\langle e \rangle = L$  and  $e^\perp = V$ , a 2-dimensional subspace whose affine shift is depicted in green in Fig. 7, reflecting the decomposition of  $\mathbb{R}^3$  into a product of  $\langle e \rangle$  and  $V$ . A point  $x \in \mathbb{R}^3$  can now be written as  $\bar{x} \cdot e + v_x$ , where  $\bar{x} \in \mathbb{R}$  and  $v_x \in V$ .

We show on the right hand side of Fig. 7 how  $V$ , represented as  $\mathbb{R}^2$ , has the structure of a cone over a Coxeter complex. The figure shows the projections of the planes  $x_1 = x_2$ ,  $x_2 = x_3$  and  $x_1 = x_3$  and the intersection of  $V$  with the subspace  $\langle e \rangle$  (red dot). To obtain the cone structure on  $V$ , we give it spherical coordinates (i.e. polar coordinates in this case). The first coordinate is the radius  $r$ , which determines a 1-sphere centred at the origin (the black circle). On the circle, a point  $v_x$  is determined by an angle  $x_\theta$ . Intersecting the circle with the hyperplanes, we decompose it into  $|\text{Sym}_3| = 6$  (coloured) strata indexed by the symmetric group and forget about the angle  $x_\theta$ . For instance, if  $v = (v_1, v_2, v_3)$  with  $v_2 < v_3 < v_1$ , the point  $v$  lies in the stratum indexed by [231]; this is the unique region that lies on those sides of the hyperplanes that satisfy  $x_1 > x_2$ ,  $x_2 < x_3$  and  $x_1 > x_3$ .

Let  $\gamma = (12)$ . It acts on  $v$  via  $\gamma \cdot v = (v_{\gamma^{-1}(1)}, v_{\gamma^{-1}(2)}, v_{\gamma^{-1}(3)}) = (v_2, v_1, v_3)$ . We denote its image by  $v^\gamma := \gamma \cdot v$ . The order of the coordinates of  $v^\gamma$  satisfies  $v_1^\gamma \leq v_3^\gamma \leq v_2^\gamma$ , so  $v^\gamma$  lies in the stratum indexed by the permutation [132]. The image  $v^\gamma$  of  $v$  through the action of  $\gamma$  corresponds to the reflection through the hyperplane  $x_1 = x_2$ .

**Remark 3.5** There are two special cases in Proposition 3.2, when  $x_i = x_j$  for all  $i, j$ , i.e.  $(x_1, \dots, x_n) \in L$  and when  $x_i \neq x_j$  for all  $i \neq j$ , i.e.  $(x_1, \dots, x_n) \in \text{Conf}_n(\mathbb{R})$ . For the former, we have  $p(x) = (\bar{x}, \|v_x\|) = (x_i, 0)$  and  $x_\theta$  is not defined. For the latter,  $q(x) = x_\theta$  lies in the interior of a top-dimensional simplex of  $\Sigma(\text{Sym}_n)$ . Hence, it determines a *unique* element  $\tau_x \in \text{Sym}_n$ . In fact, these are just the two extremes of a family of situations that can occur:



If  $x_i = x_j$  for some  $i \neq j$ , then  $x_\theta$  lies on the corresponding hyperplane in  $\mathcal{H}$  and hence on a lower-dimensional face of  $\Sigma(\text{Sym}_n)$ . There exists a permutation  $\tau \in \text{Sym}_n$  such that

$$x_{\tau(1)} \leq x_{\tau(2)} \leq \dots \leq x_{\tau(n)},$$

but  $\tau$  is not unique. It is defined only up to multiplication by the subgroup

$$P = \{ \gamma \in \text{Sym}_n \mid x_{\tau(i)} = x_{\tau\gamma(i)} \}.$$

Note that  $P$  is generated by adjacent transpositions  $(i, i + 1)$ , i.e. it is of the form  $\langle T \rangle$ , where  $T \subset S$  is a subset of the set  $S$  of simple reflections of  $\text{Sym}_n$ . Hence, it is a parabolic subgroup of  $\text{Sym}_n$  (see Sect. 2.2). The number of adjacent transpositions in  $P$  depends on how many coordinates of  $(x_1, \dots, x_n)$  agree, or, equivalently, the number of hyperplanes in  $\mathcal{H}$  it lies on. Intuitively speaking, one could phrase this as “the more of the  $x_i$ ’s take the same value, the less ‘permutation information’ is left”. The coset

$$\tau P = \{ \rho \in \text{Sym}_n \mid x_{\rho(1)} \leq \dots \leq x_{\rho(n)} \},$$

corresponds to the lowest dimensional face of  $\Sigma(\text{Sym}_n)$  that  $x$  lies on. It depends only on the values of the  $x_i$ , not on the choice of  $\tau$ . If  $x \in L$ , we have  $\tau P = \text{Sym}_n$ . This could be interpreted as the degenerate case where  $x_\theta$  lies on the unique  $(-1)$ -dimensional face of  $\Sigma(\text{Sym}_n)$  (see Definition 2.11).

### 4 Coxeter coordinates for the space of barcodes

We are finally ready to turn to our main goal, namely to describe a stratification of  $\mathcal{B}_n$ . Recall that this will decompose  $\mathcal{B}_n$  into different regions, where each region is characterised as the set of all barcodes having the same average birth and death, the same standard deviation of births and deaths and the same permutation type.

#### 4.1 Describing $\mathcal{B}_n$ as a quotient

In this section, we describe  $\mathcal{B}_n$  as a subset of a quotient of  $\mathbb{R}^{2n}$ . This will be used in the next section to equip this space with Coxeter complex coordinates.

Let  $X := \text{Sym}_n \backslash \mathbb{R}^n \times \mathbb{R}^n$ , where  $\text{Sym}_n$  acts diagonally by permuting the coordinates, i.e. for  $\gamma \in \text{Sym}_n$ , we set

$$\gamma \cdot (x_1, \dots, x_n, y_1, \dots, y_n) = (x_{\gamma^{-1}(1)}, \dots, x_{\gamma^{-1}(n)}, y_{\gamma^{-1}(1)}, \dots, y_{\gamma^{-1}(n)}).$$

The elements of  $X$  are equivalence classes of tuples  $(x_1, \dots, x_n, y_1, \dots, y_n) \in \mathbb{R}^n \times \mathbb{R}^n$ , which are denoted by  $[x_1, \dots, x_n, y_1, \dots, y_n]$ .

**Remark 4.1** We write  $X := \text{Sym}_n \setminus \mathbb{R}^n \times \mathbb{R}^n$  to emphasise that  $\text{Sym}_n$  acts from the left on this space. The reason we stress this is that later on, we will combine the statements here with descriptions of the Coxeter complex. There, the simplices are given by cosets  $\tau P$  and the symmetric group acts on them by *left* multiplication.

There is a map  $\phi$  from the space of barcodes with  $n$  bars to  $X$  given by

$$\phi : \mathcal{B}_n \rightarrow X = \text{Sym}_n \setminus \mathbb{R}^n \times \mathbb{R}^n$$

$$\{(b_i, d_i)\}_{i \in \{1, \dots, n\}} \mapsto [b_1, \dots, b_n, d_1, \dots, d_n].$$

The image of  $\phi$  is independent of the choice of indices for the bars of the barcode because the action of  $\text{Sym}_n$  is factored out. The map  $\phi$  is clearly injective, but it is not surjective as the birth time of a homology class is always smaller than its death time. The image of  $\phi$  is the subspace  $Y$  of  $X$  given by

$$Y := \text{Sym}_n \setminus \{(x_1, \dots, x_n, y_1, \dots, y_n) \in \mathbb{R}^n \times \mathbb{R}^n \mid x_i < y_i \forall i\}.$$

For later reference, we note this observation in the following.

**Proposition 4.2** *The map  $\phi$  defines a bijection  $\mathcal{B}_n \rightarrow Y \subset \text{Sym}_n \setminus \mathbb{R}^n \times \mathbb{R}^n$ .*

In Sect. 5, we equip  $\mathcal{B}_n$  with metrics inspired by the bottleneck and Wasserstein distances. The map  $\phi$  is an isometry with respect to these metrics.

### 4.2 Coxeter complexes for birth and death

We now introduce the Coxeter complex coordinates for  $\mathcal{B}_n$ . These coordinates are obtained by applying the map  $(p|_{\mathbb{R}^n \setminus L}, q)$  of Proposition 3.2 to the two copies of  $\mathbb{R}^n$  in  $Y$ .

**Theorem 4.3** *Every barcode  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}} \in \mathcal{B}_n$  such that at least two of the  $b_i$  and two of the  $d_i$  are different from each other determines the following five data:*

1. Its average birth time  $\bar{b} = \sum_{i=1}^n b_i/n \in \mathbb{R}$ ;
2. Its average death time  $\bar{d} = \sum_{i=1}^n d_i/n \in \mathbb{R}$ ;
3. Its birth standard deviation  $\|v_b\| = (\sum_{i=1}^n |b_i - \bar{b}|^2)^{1/2} \in \mathbb{R}_{>0}$ ;
4. Its death standard deviation  $\|v_d\| = (\sum_{i=1}^n |d_i - \bar{d}|^2)^{1/2} \in \mathbb{R}_{>0}$ ;
5. An orbit  $\text{Sym}_n \cdot (b_\theta, d_\theta) \in \text{Sym}_n \setminus \Sigma(\text{Sym}_n) \times \Sigma(\text{Sym}_n)$ .

Furthermore, these five data uniquely determine  $B$ .

**Proof** Let  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}}$  be such that at least two  $b_i$  and two  $d_i$  are different. By assumption, both  $(b_1, \dots, b_n)$  and  $(d_1, \dots, d_n)$  are points in  $\mathbb{R}^n \setminus L$ . The image of  $B$  under  $\phi$  (Proposition 4.2) is

$$\phi(B) = [b_1, \dots, b_n, d_1, \dots, d_n] \in \text{Sym}_n \setminus (\mathbb{R}^n \setminus L \times \mathbb{R}^n \setminus L).$$

Since the map  $(p|_{\mathbb{R}^n \setminus L}, q)$  is  $\text{Sym}_n$ -equivariant (Proposition 3.2), it induces a bijection

$$\begin{aligned} & \text{Sym}_n \setminus (\mathbb{R}^n \setminus L \times \mathbb{R}^n \setminus L) \\ & \cong \text{Sym}_n \setminus ((\mathbb{R} \times \mathbb{R}_{>0} \times \Sigma(\text{Sym}_n)) \times (\mathbb{R} \times \mathbb{R}_{>0} \times \Sigma(\text{Sym}_n))). \end{aligned}$$

The image of  $[b_1, \dots, b_n, d_1, \dots, d_n]$  under this bijection is the  $\text{Sym}_n$ -orbit of

$$(p|_{\mathbb{R}^n \setminus L}, q)^2(b_1, \dots, b_n, d_1, \dots, d_n) = (\bar{b}, \|v_b\|, b_\theta, \bar{d}, \|v_d\|, d_\theta).$$

The claim now follows since the action of  $\text{Sym}_n$  on  $(\bar{b}, \|v_b\|, b_\theta, \bar{d}, \|v_d\|, d_\theta)$  is trivial on  $\bar{b}, \|v_b\|, \bar{d}, \|v_d\|$  and is given by the action of  $\text{Sym}_n$  on the Coxeter complex  $\Sigma(\text{Sym}_n)$  for  $b_\theta, d_\theta$ .  $\square$

### 4.3 A stratification of $\mathcal{B}_n$

In this section, we describe the stratification that we obtain from the description of  $\mathcal{B}_n$  in terms of Coxeter complexes.

We start by extending Definition 2.8, the permutation assigned to a strict barcode, to the general case of  $\mathcal{B}_n$ . For non-strict barcodes, we cannot uniquely assign a permutation. However, there is a nice description of the set of all possible such permutations in terms of double cosets of parabolic subgroups:

**Definition 4.4** For a barcode  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}} \in \mathcal{B}_n$ , let  $\tau_b$  and  $\tau_d$  be elements of  $\text{Sym}_n$  such that  $b_{\tau_b(1)} \leq \dots \leq b_{\tau_b(n)}$  and  $d_{\tau_d(1)} \leq \dots \leq d_{\tau_d(n)}$ . Let

$$P_b^B = \{\gamma \in \text{Sym}_n \mid b_{\tau_b(i)} = b_{\tau_b\gamma(i)}\}, \quad P_d^B = \{\gamma \in \text{Sym}_n \mid d_{\tau_d(i)} = d_{\tau_d\gamma(i)}\}.$$

The double coset  $D_B$  associated to  $B$  is defined as  $D_B := P_b^B \tau_b^{-1} \tau_d P_d^B$ .

**Remark 4.5** Note that while  $\tau_b$  and  $\tau_d$  depend on the ordering of the barcode,  $P_b^B$  and  $P_d^B$  do not. The groups  $P_b^B$  and  $P_d^B$  are parabolic subgroups of  $\text{Sym}_n$ , as was observed in Remark 3.5. The cosets

$$\tau_b P_b^B = \{\rho \in \text{Sym}_n \mid b_{\rho(1)} \leq \dots \leq b_{\rho(n)}\}$$

and

$$\tau_d P_d^B = \{\rho \in \text{Sym}_n \mid d_{\rho(1)} \leq \dots \leq d_{\rho(n)}\},$$

which are the sets of permutations that preserve the order of the  $b_i$  and  $d_i$  respectively, do not depend on the indexing of  $B$  either. Hence, the double coset  $D_B = (\tau_b P_b^B)^{-1} \cdot \tau_d P_d^B$  is indeed an invariant of the barcode  $B$ . Furthermore, if  $B$  is a strict barcode, then  $P_b^B = \{\text{id}\} = P_d^B$ , so  $D_B = \{\tau_b^{-1} \tau_d\} = \{\sigma_B\}$  and we recover the definition of Kanari et al. (2020) as given in Definition 2.8.

**Example 4.6** Let

$$B = \{(b_1, d_1) = (1, 10), (b_2, d_2) = (2, 5), (b_3, d_3) = (4, 5), (b_4, d_4) = (4, 7)\} \in \mathcal{B}_4.$$

One has  $b_1 < b_2 < b_3 = b_4$  and  $d_2 = d_3 < d_4 < d_1$ . Let  $\tau_b = [1234]$  and  $\tau_d = [2341]$ . They satisfy  $b_{\tau_b(1)} \leq \dots \leq b_{\tau_b(4)}$  and  $d_{\tau_d(1)} \leq \dots \leq d_{\tau_d(4)}$  respectively, but so do  $\tau'_b = [1243]$  and  $\tau'_d = [3241]$ . In this case, one has  $P_b^B = \{\text{id}, (34)\}$ ,  $P_d^B = \{\text{id}, (12)\}$  and  $\tau_b P_b^B = \{[1234], [1243]\}$ ,  $\tau_d P_d^B = \{[2341], [3241]\}$ . The double coset

$$\begin{aligned} D_B &= \{\gamma_b \tau_b^{-1} \tau_d \gamma_d \mid \gamma_b \in P_b^B, \gamma_d \in P_d^B\} \\ &= \{\tau_b^{-1} \tau_d, \tau_b'^{-1} \tau_d, \tau_b^{-1} \tau'_d, \tau_b'^{-1} \tau'_d\} \\ &= \{[2341], [2431], [3241], [4231]\} \end{aligned}$$

is the set of all the permutations  $\sigma$  that satisfy that the  $j$ -th death (in increasing order) is paired with the  $\sigma(j)$ -th birth.

Recall that the Coxeter complex  $\Sigma(\text{Sym}_n)$  is a simplicial complex with simplices given by cosets of parabolic subgroups  $\tau P$ . This simplicial decomposition gives it the structure of a stratified space over the poset of cosets of parabolic subgroups equipped with reverse inclusion (see Sect. 2.2). Taking the cone and products of these simplices yields a decomposition of

$$\mathbb{R}^{2n} \cong \text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R} \times \text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R} \tag{5}$$

into strata that are compatible with the action of  $\text{Sym}_n$ , i.e. each stratum is sent to another stratum of same dimension by the action of  $\text{Sym}_n$ . This follows from Remark 3.3 and the fact that  $\Sigma(\text{Sym}_n)$  is stratified and the map  $(p|_{\mathbb{R}^n \setminus L}, q)$  of Proposition 3.2 is  $\text{Sym}_n$ -equivariant. The strata in Eq. (5) are indexed by pairs of cosets  $(\tau_1 P_1, \tau_2 P_2)$ , where  $\tau_1, \tau_2 \in \text{Sym}_n$  and  $P_1, P_2 \leq \text{Sym}_n$  are parabolic subgroups<sup>2</sup>. The partial ordering on these pairs is given component-wise by reverse inclusion [cf. Eq. (20)].

It follows that the quotient  $X = \text{Sym}_n \setminus \mathbb{R}^{2n}$  is stratified over the quotient  $\mathcal{P}$  of this poset by the action of  $\text{Sym}_n$ . More concretely,  $\mathcal{P}$  can be described as follows: The elements of  $\mathcal{P}$  are orbits of the form  $\text{Sym}_n \cdot (\tau_1 P_1, \tau_2 P_2)$ , where  $\tau_1, \tau_2 \in \text{Sym}_n$  and  $P_1, P_2 \leq \text{Sym}_n$  are parabolic subgroups. The partial ordering is given by

$$\text{Sym}_n \cdot (\tau_1 P_1, \tau_2 P_2) \leq \text{Sym}_n \cdot (\tau'_1 P'_1, \tau'_2 P'_2)$$

if there is  $\gamma \in \text{Sym}_n$  such that

$$\tau_1 P_1 \supseteq \gamma \tau'_1 P'_1 \text{ and } \tau_2 P_2 \supseteq \gamma \tau'_2 P'_2.$$

<sup>2</sup> Note that, following Remark 3.5, the points in  $\text{Conf}_n(\mathbb{R}) \times \text{Conf}_n(\mathbb{R}) \subset \mathbb{R}^n \times \mathbb{R}^n$  are exactly the ones that belong to the top-dimensional strata. Similarly, the points of  $L \times L \subset \mathbb{R}^n \times \mathbb{R}^n$  belong to the lowest dimensional strata, corresponding to the cone points in Eq. (5).

This quotient poset  $\mathcal{P}$  has a more explicit description in terms of another poset  $\mathcal{Q}$ , which consists of “marked” double cosets of parabolic subgroups:

**Definition 4.7** Let  $\mathcal{Q}$  be the poset consisting of all triples  $(P_1, P_1\sigma P_2, P_2)$ , where  $\sigma \in \text{Sym}_n$  and  $P_1, P_2 \leq \text{Sym}_n$  are parabolic subgroups and where

$$(P_1, P_1\sigma P_2, P_2) \leq (P'_1, P'_1\sigma P'_2, P'_2)$$

if and only if there is component-wise containment in the reverse direction,

$$P_1 \supseteq P'_1, P_2 \supseteq P'_2 \text{ and } P_1\sigma P_2 \supseteq P'_1\sigma P'_2.$$

A very similar poset is also studied as a two-sided version of the Coxeter complex by Hultman Hultman (2007) and Petersen Petersen (2018). We remark that  $\mathcal{Q}$  is different from the poset of all double cosets of the form  $P_1\sigma P_2$ : There can be  $P_1 \neq P'_1, P_2 \neq P'_2$  such that  $P_1\sigma P_2 = P'_1\sigma P'_2$  (see Petersen (2018)[Remark 4]).

**Lemma 4.8** *The map*

$$\begin{aligned} \phi : \mathcal{P} &\rightarrow \mathcal{Q} \\ \text{Sym}_n \cdot (\tau_1 P_1, \tau_2 P_2) &\mapsto (P_1, P_1\tau_1^{-1}\tau_2 P_2, P_2) \end{aligned}$$

*is an isomorphism of posets.*

**Proof** To see that  $\phi$  is a bijection of the underlying sets, consider the following map:

$$\begin{aligned} \psi : \mathcal{Q} &\rightarrow \mathcal{P} \\ (P_1, P_1\sigma P_2, P_2) &\mapsto \text{Sym}_n \cdot (P_1, \sigma P_2). \end{aligned}$$

It is easy to verify that  $\phi$  and  $\psi$  are independent of the choices of representatives and are inverse to one another. That  $\phi$  is indeed a map of posets, i.e. that it preserves the partial ordering, follows from elementary manipulations of cosets. □

**Theorem 4.9** *The set  $\mathcal{B}_n$  of barcodes with  $n$  bars is stratified over the poset  $\mathcal{Q}$ . The lowest dimensional stratum containing the barcode  $B$  is the stratum corresponding to  $(P_b^B, D_B, P_d^B) \in \mathcal{Q}$ . It is of the form*

$$\mathcal{B}_n^{(P_b^B, D_B, P_d^B)} = \left( \text{Sym}_n \cdot (\text{cone}(\tau_b P_b^B) \times \mathbb{R} \times \text{cone}(\tau_d P_d^B) \times \mathbb{R}) \right) \cap Y.$$

**Proof** Recall that  $\mathcal{B}_n \cong Y$  is a subset of  $X = \text{Sym}_n \setminus \mathbb{R}^{2n}$  (Proposition 4.2). As observed above,  $X$  is stratified over the poset  $\mathcal{P}$  and, by Lemma 4.8, this poset is isomorphic to  $\mathcal{Q}$ . It follows that  $\mathcal{B}_n$  is also stratified over  $\mathcal{Q}$ . The strata are obtained by taking the intersection with  $Y$ .

This stratification is induced by the simplicial structure of the Coxeter complexes in

$$X \cong \text{Sym}_n \setminus \left( \text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R} \times \text{cone}(\Sigma(\text{Sym}_n)) \times \mathbb{R} \right).$$

Hence, the strata that contain a barcode  $B \in \mathcal{B}_n$  only depend on the coordinate  $\text{Sym}_n \cdot (b_\theta, d_\theta) \in \text{Sym}_n \setminus \Sigma(\text{Sym}_n) \times \Sigma(\text{Sym}_n)$  that  $B$  determines by Theorem 4.3. As explained in Remark 3.5, the associated points  $b_\theta, d_\theta \in \Sigma(\text{Sym}_n)$  lie in the interior of the simplices  $\tau_b P_b^B, \tau_d P_d^B$ . Hence, the lowest dimensional stratum that contains  $B$  corresponds to the  $\text{Sym}_n$ -orbit of  $(\tau_b P_b^B, \tau_d P_d^B)$ .  $\square$

Let  $B$  be a strict barcode, that is,  $b_i \neq b_j$  and  $d_i \neq d_j$  for  $i \neq j$ . Then  $B$  is contained in the top-dimensional stratum

$$\mathcal{B}_n^{(\{\text{id}\}, \{\text{id}\} \tau_b^{-1} \tau_d \{\text{id}\}, \{\text{id}\})} = (\text{Sym}_n \cdot (\text{cone}(\tau_b \{\text{id}\}) \times \mathbb{R} \times \text{cone}(\tau_d \{\text{id}\}) \times \mathbb{R})) \cap Y.$$

Changing the representative of the  $\text{Sym}_n$ -orbit, this can be rewritten as

$$\mathcal{B}_n^{(\{\text{id}\}, \{\sigma_B\}, \{\text{id}\})} = (\text{Sym}_n \cdot (\text{cone}(\{\text{id}\}) \times \mathbb{R} \times \text{cone}(\sigma_B \{\text{id}\}) \times \mathbb{R})) \cap Y,$$

where  $\sigma_B = \tau_b^{-1} \tau_d$  is the permutation associated to  $B$  as in Definition 2.8. In particular, the strata containing strict barcodes are in one-to-one correspondence with the elements of  $\text{Sym}_n$ .

When one considers the cone and real line parameters in the stratification of Theorem 4.9, one obtains regions that are determined by the averages and standard deviations of Theorem 4.3 and by parabolic subgroups.

**Corollary 4.10** *The Coxeter coordinates of Theorem 4.3 decompose the space  $\mathcal{B}_n$  of barcodes with  $n$  bars into disjoint regions. The region containing the barcode  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}} \in \mathcal{B}_n$  is defined as the set of all barcodes  $B'$  such that:*

1. *Its average birth time is the same as that of  $B$ , i.e.  $\bar{b}' = \bar{b}$ ;*
2. *Its average death time is the same as that of  $B$ , i.e.  $\bar{d}' = \bar{d}$ ;*
3. *Its birth standard deviation is the same as that of  $B$ , i.e.  $\|v_{b'}\| = \|v_b\|$ ;*
4. *Its death standard deviation is the same as that of  $B$ , i.e.  $\|v_{d'}\| = \|v_d\|$ ;*
5.  *$P_{b'}^{B'} = P_b^B, P_{d'}^{B'} = P_d^B$  and  $D_B = D_{B'}$ .*

*For strict barcodes, the information of the last Item 5 is equivalent to specifying  $\sigma_B$ , the permutation associated to barcodes in Definition 2.8.*

### 5 A metric on $\mathcal{B}_n$

In this section, we explain how the description of  $\mathcal{B}_n$  given in Sect. 4.1 with  $\mathbb{R}^n$  equipped with the  $l^\infty$ -norm gives rise to a naturally defined metric  $\tilde{d}_B$  on  $\mathcal{B}_n$  that is closely related to the bottleneck distance. Similarly, the  $l^2$ -norm on  $\mathbb{R}^n$  leads to a modified Wasserstein distance  $\tilde{d}_W$  on  $\mathcal{B}_n$ .

To describe  $\tilde{d}_B$ , we equip  $\mathbb{R}^{2n}$  with the metric  $d_\infty$  induced by the  $l^\infty$ -norm. This metric induces a map  $X \times X \rightarrow \mathbb{R}$  on the quotient by taking the minimum value over all representatives of the corresponding equivalence classes:

$$d : X \times X \rightarrow \mathbb{R}$$

$$([x, y], [x', y']) \mapsto \min_{\substack{(\tilde{x}, \tilde{y}) \in [x, y], \\ (\tilde{x}', \tilde{y}') \in [x', y']}} d_\infty((\tilde{x}, \tilde{y}), (\tilde{x}', \tilde{y}')). \tag{6}$$

We will show that this map restricted to  $Y$  agrees with a modified version of the bottleneck distance.

**Definition 5.1** Let  $B = \{(b_i, d_i)\}_{i \in \{1, \dots, n\}}$  and  $B' = \{(b'_i, d'_i)\}_{i \in \{1, \dots, n\}}$  be two barcodes in  $\mathcal{B}_n$ . The *modified bottleneck distance* between  $B$  and  $B'$  is

$$\tilde{d}_B(B, B') := \min_{\gamma \in \text{Sym}_n} \max_{i \in \{1, \dots, n\}} \|(b_i, d_i) - (b'_{\gamma(i)}, d'_{\gamma(i)})\|_\infty.$$

where  $\|\cdot\|_\infty$  is the  $l^\infty$ -norm on  $\mathbb{R}^2$ .

Note that the difference between the modified bottleneck distance and the original bottleneck distance as defined in Definition 2.5 is that for the modified version, one does not allow to match points of the barcodes to the diagonal  $\Delta$  (see Fig. 8). Furthermore,  $\tilde{d}_B(B, B')$  is well-defined only if both  $B$  and  $B'$  contain the same number of bars, i.e. if they are both elements of the same  $\mathcal{B}_n$ . This is not necessary for the definition of the regular bottleneck distance, cf. Remark 5.3.

**Proposition 5.2** *The map  $d$  defines a metric on  $Y$  with respect to which  $\phi : (\mathcal{B}_n, \tilde{d}_B) \rightarrow (Y, d)$  is an isometry.*

**Proof** As observed before in Proposition 4.2,  $\phi$  maps  $\mathcal{B}_n$  bijectively onto  $Y$ . Hence, it is sufficient to show that for arbitrary barcodes  $B$  and  $B'$ ,

$$\tilde{d}_B(B, B') = d(\phi(B), \phi(B')).$$

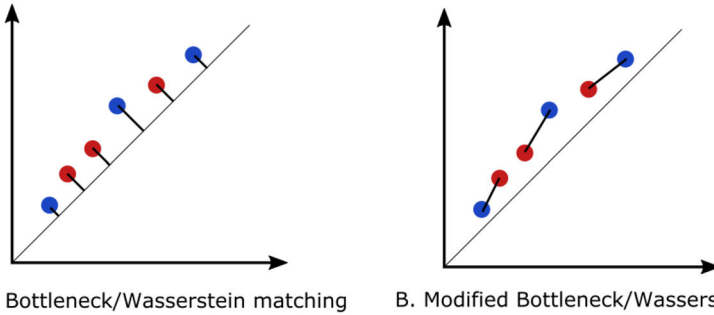
This follows from simply spelling out the definitions. For points  $(x, y)$  and  $(x', y')$  in  $\mathbb{R}^n \times \mathbb{R}^n$ ,

$$\begin{aligned} d_\infty((x, y), (x', y')) &= \max \{|x_1 - x'_1|, \dots, |x_n - x'_n|, |y_1 - y'_1|, \dots, |y_n - y'_n|\} \\ &= \max_{i=1, \dots, n} \max \{|x_i - x'_i|, |y_i - y'_i|\} \\ &= \max_{i=1, \dots, n} \|(x_i, y_i) - (x'_i, y'_i)\|_\infty, \end{aligned}$$

where  $\|\cdot\|_\infty$  is the  $l^\infty$ -norm on  $\mathbb{R}^2$ . Combining this with the definition of  $d$  on  $X$  [see Eq. (6)], we obtain

$$\begin{aligned} d(\phi(B), \phi(B')) &= \min_{\gamma \in \text{Sym}_n} d_\infty(\phi(B), \gamma \cdot \phi(B')) \\ &= \min_{\gamma \in \text{Sym}_n} \max_{i=1, \dots, n} \|(b_i, d_i) - (b'_{\gamma^{-1}(i)}, d'_{\gamma^{-1}(i)})\|_\infty. \end{aligned}$$

This is the same as the modified bottleneck distance of Definition 5.1. □



**Fig. 8** Two barcodes (red and blue) represented as persistence diagrams in  $\mathbb{R}^2$ . A. The matching that minimises the bottleneck or Wasserstein distance matches all the bars to the diagonal, as they are all very close to it. B. If bars are not allowed to be matched with the diagonal, the matching that minimises  $\|(b_i, d_i) - (b'_{\gamma(i)}, y'_{\gamma(i)})\|_\infty$  for the bottleneck distance or  $\sum_i \|(b_i, d_i) - (b'_{\gamma(i)}, y'_{\gamma(i)})\|_2$  respectively for the Wasserstein distance is different (color figure online)

Similarly, starting with  $\mathbb{R}^{2n}$  equipped with the  $l^2$ -norm, one can establish an isometry between  $Y$  and  $\mathcal{B}_n$  equipped with a modified Wasserstein distance instead.

**Remark 5.3** Forgetting about the diagonal as done above opens the door to defining new metrics on barcodes by considering distances on  $\mathbb{R}^n \times \mathbb{R}^n$  and then taking the quotient as was done in this section. It could potentially be extended to barcodes with different number of bars. One could for instance imagine a map that forces matchings between as many bars as possible and then adds a positive weight equal to their distance to the diagonal to the unmatched bars if there are any. This is different from the bottleneck distance (or Wasserstein distance), which allows as many matchings as needed with the diagonal, see Fig. 8. When using barcodes to study data, bars close to the diagonal are usually considered as related to noise. However, there are cases where all the bars matter, for instance when the barcode is the one of a merge tree Kanari et al. (2020); Curry et al. (2021). In such a case, a new metric that does not take the diagonal into account could turn out useful. We leave this for future work.

### 6 Future directions

In this paper, we showed that the space  $\mathcal{B}_n$  of barcodes with  $n$  bars is stratified over the poset of marked double cosets of parabolic subgroups of  $\text{Sym}_n$ . A question that arises is how this could be extended to the whole space of barcodes, i.e. to the union  $\bigcup_{n \in \mathbb{N}} \mathcal{B}_n$ . An approach here would be to use appropriate inclusions  $\mathcal{B}_m \hookrightarrow \mathcal{B}_n$  for  $m \leq n$ . Note that on the group level, there are natural injections  $\text{Sym}_m \hookrightarrow \text{Sym}_n$ . On the level of simplicial complexes,  $\Sigma(\text{Sym}_n)$  also contains copies of  $\Sigma(\text{Sym}_m)$  for  $m \leq n$ .

It was shown in Kanari et al. (2020); Curry et al. (2021) that the permutation  $\sigma_B$  associated to a strict barcode  $B$  gives nice combinatorial insight on the number of merge trees that have the same barcode. This number, called the tree-realisation number (TRN), is derived directly from the permutation. It can also be used to do statistics



on barcodes. Our coordinates (Corollary 4.10) firstly extend this work to any (possibly non-strict) barcode and secondly return a finer invariant than just the permutation. A future direction would be to study this finer invariant defined by  $(\bar{b}, \bar{d}, \|v_b\|, \|v_d\|, \sigma_B)$ . It might be well-suited for studying statistical questions: The first four elements already have descriptions as averages and standard deviations. The behaviour of the permutation  $\sigma_B$  could be studied using tools from permutation statistics, such as the number of inversions or descents.

In a different direction, the description of  $\mathcal{B}_n$  in terms of Coxeter complexes allows to rephrase these combinatorial questions in more geometric terms. Using this geometric perspective might give new ways for studying invariants and statistics on barcodes.

It would be interesting to see if the geometric and combinatorial tools developed here can help to understand inverse problems in TDA as the ones in Kanari et al. (2020); Curry et al. (2021); Curry (2018); Leygonie et al. (2022). Since the merge tree to barcode problem is related to the symmetric group Kanari et al. (2020); Curry et al. (2021), it is also natural to ask whether the stratification that we obtain in Theorem 4.9 can be extended to the space of merge trees with  $n$  leaves.

Lastly, the modified bottleneck and Wasserstein distances seem to have a different behaviour than the usual ones. A deeper study of their properties and their potential extension to the space of barcodes (see Remark 5.3) is a natural next step to consider.

**Acknowledgements** This project started from discussions on the space of barcodes with Justin Curry and Brendan Mallery. The authors would also like to thank Jérôme Scherer, Kathryn Hess and Darrick Lee for the fruitful discussions and their useful comments on the manuscript.

**Funding** Open access funding provided by EPFL Lausanne

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abramenko, P., Brown, K.S.: Buildings. Graduate Texts in Mathematics, vol. 248. Springer, New York (2008)
- Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L.: Persistence images: a stable vector representation of persistent homology. *J. Mach. Learn. Res.* **18**(8), 1–35 (2017)
- Adcock, A., Carlsson, E., Carlsson, G.: The ring of algebraic functions on persistence bar codes. *Homol. Homot. Appl.* **18**, 04 (2013)

- Billera, L.J., Holmes, S.P., Vogtmann, K.: Geometry of the space of phylogenetic trees. *Adv. Appl. Math.* **27**(4), 733–767 (2001)
- Björner, A.: Some combinatorial and algebraic properties of Coxeter complexes and Tits buildings. *Adv. Math.* **52**(3), 173–212 (1984)
- Björner, A., Brenti, F.: *Combinatorics of Coxeter groups*. Graduate Texts in Mathematics, vol. 231. Springer, New York (2005)
- Bridson, M.R., Haefliger, A.: *Metric spaces of non-positive curvature*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319. Springer-Verlag, Berlin (1999)
- Bubenik, P.: Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16**(01), 77–102 (2015)
- Byrne, H.M., Harrington, H.A., Muschel, R., Reinert, G., Stolz, B.J., Tillmann, U.: Topological methods for characterising spatial networks: a case study in tumour vasculature (2019)
- Carlsson, G.: Topology and data. *Bull. Am. Math. Soc. (N.S.)* **46**(2), 255–308 (2009)
- Carrière, M., Oudot, S., Ovsjanikov, M.: Stable topological signatures for points on 3d shapes. *Comput. Graph. Forum* **34** (2015). <https://doi.org/10.1111/cgf.12692>
- Crawley-Boevey, W.: Decomposition of pointwise finite-dimensional persistence modules. *J. Algebra Appl.* **14**(05), 1550066 (2015)
- Curry, J.: The fiber of the persistence map for functions on the interval. *J. Appl. Comput. Topol.* **2**(3), 301–321 (2018)
- Curry, J., DeSha, J., Garin, A., Hess, K., Kanari, L., Mallery, B.: From Combinatorial and Probabilistic Aspects of a Topological Inverse Problem (2021). <https://doi.org/10.48550/ARXIV.2107.11212>
- Davis, M.W.: The geometry and topology of Coxeter groups. In *Introduction to modern mathematics*, volume 33 of *Adv. Lect. Math. (ALM)*, pp. 129–142. Int. Press, Somerville (2015)
- Delgado-Friedrichs, O., Robins, V., Sheppard, A.: Morse theory and persistent homology for topological analysis of 3D images of complex materials. In *2014 IEEE International Conference on Image Processing (ICIP)*, vol. 10, pp. 4872–4876 (2014)
- Di Fabio, B., Ferri, M.: Comparing persistence diagrams through complex vectors. In Murino, V., Puppo, E. (eds) *Image Analysis and Processing—ICIAP 2015*, pp. 294–305, Springer International Publishing, Cham (2015)
- Edelsbrunner, H., Harer, J.: Persistent homology—a survey. In *Surveys on Discrete and Computational Geometry*, volume 453 of *Contemp. Math.*, pp. 257–282. American Mathematical Society, Providence (2008)
- Gameiro, M., Hiraoka, Y., Izumi, S., Kramár, M., Mischaikow, K., Nanda, V.: A topological measurement of protein compressibility. *Jpn. J. Ind. Appl. Math.* **32**, 1–17 (2015)
- Ghrist, R.: Barcodes: the persistent topology of data. *Bull. Am. Math. Soc. (N.S.)* **45**(1), 61–75 (2008)
- Grindstaff, G., Owen, M.: Geometric comparison of phylogenetic trees with different leaf sets (2018). <https://doi.org/10.48550/ARXIV.1807.04235>
- Heydenreich, S., Brück, B., Harnois-Déraps, J.: Persistent homology in cosmic shear: constraining parameters with topological data analysis. *Astron. Astrophys.* **648**, A74 (2021)
- Hultman, A.: The combinatorics of twisted involutions in Coxeter groups. *Trans. Am. Math. Soc.* **359**(6), 2787–2798 (2007)
- Jacquard, E., Nanda, V., Tillmann, U.: The space of barcode bases for persistence modules. *J. Appl. Comput. Topol.* (2022). <https://doi.org/10.1007/s41468-022-00094-6>
- Kališnik, S.: Tropical coordinates on the space of persistence barcodes. *Found. Comput. Math.* **19**(1), 101–129 (2019)
- Kanari, L., Dlotko, P., Scolamiero, M., Levi, R., Shillcock, J., Hess, K., Markram, H.: A topological representation of branching neuronal morphologies. *Neuroinformatics* **16**(1), 3–13 (2018)
- Kanari, L., Garin, A., Hess, K.: From trees to barcodes and back again: theoretical and statistical perspectives. *Algorithms* **13** (2020). <https://doi.org/10.3390/a13120335>
- Lee, Y., Barthel, S., Dlotko, P., Moosavi, S.M., Hess, K., Smit, B.: High-throughput screening approach for nanoporous materials genome using topological data analysis: Application to zeolites. *J. Chem. Theory Comput.* **14**, 4427–4437 (2018)
- Leygonie, J., Tillmann, U.: The fiber of persistent homology for simplicial complexes. *J Pure Appl Algebra* **226** (2022). <https://doi.org/10.1016/j.jpaa.2022.107099>

- Muszynski, G., Kashinath, K., Kurlin, V., Wehner, M.F., Prabhat, M.: Topological data analysis and machine learning for recognizing atmospheric river patterns in large climate datasets. *Geosci. Model Dev.* **12**, 613–628 (2019)
- Petersen, T.K.: A two-sided analogue of the Coxeter complex. *Electr. J. Comb.* **25**(4), 28 (2018)
- Postnikov, A.: Permutohedra, associahedra, and beyond. *Int. Math. Res. Not. IMRN* **6**, 1026–1106 (2009)
- Reimann, M.W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., Dlotko, P., Levi, R., Hess, K., Markram, H.: Cliques of neurons bound into cavities provide a missing link between structure and function. *Front. Comput. Neurosci.* **11** (2017). <https://doi.org/10.3389/fncom.2017.00048>
- Robins, V., Saadatfar, M., Delgado-Friedrichs, O., Sheppard, A.P.: Percolating length scales from topological persistence analysis of micro-CT images of porous materials. *Water Resour. Res.* **52**(1), 315–329 (2016)
- Xu, C.: A correspondence between schubert cells and persistence diagrams. *Master thesis, Kyoto university, Supervisor: Yasuaki Hiraoka* (2020)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.