# Directed persistence

J.F. Jardine*

Department of Mathematics
University of Western Ontario
London, Ontario, Canada

jardine@uwo.ca

March 29, 2020

**Abstract**

Oriented Vietoris-Rips complexes arise naturally from listings of the underlying data sets. The path category (fundamental category) invariant for a finite oriented simplicial complex is computable, via an algorithm that is recalled here. A stability theorem for oriented complexes is proved, which specializes to a stability result for path categories.

## Introduction

Traditionally, topological data analysis is a story of non-oriented complexes.

Starting with a finite data set $X$ in a metric space $Z$, one constructs a filtration

$$X = V_0(X) \subset \dots \subset V_s(X) \subset V_t(X) \subset \dots$$

consisting of subcomplexes of a big simplex. The simplices of $V_s(X)$ are the finite subsets $\sigma$ of $X$ such that $d(x,y) \leq s$ for all $x,y \in \sigma$, with respect to the metric $d$ on $Z$. The Vietoris-Rips filtration is the collection of the simplicial complexes $\{V_s(X)\}$ that is associated to the data set $X$. If $s$ is sufficiently large, the full set $X$ is one of the simplices of $X$, and $V_s(X)$ is the big simplex under discussion in that case.

In its simplest form, persistent homology is the study of the behaviour of the systems $\{H_kV_s(X))\}$ of homology groups, while clustering is based on the sets of path components $\pi_0V_s(X)$.

The Vietoris-Rips complexes $V_s(X)$ are classical unoriented simplicial complexes, which are completely determined by their posets $P_s(X)$ of simplices.

*Supported by NSERC.

But one way or another, the data set $X$ is a list that is stored in a computer, and therefore has a total ordering. The ordering itself, which must be specified, can be an important part of experimental design.

Another way of saying this is that $X$ is identified with an injective function $X : \mathbf{N} \to Z$, where $\mathbf{N} = \{0, 1, \dots, N\}$ is the ordinal number having $N+1$ elements. The corresponding big simplex is the oriented simplicial complex $\Delta^N$, and the imbeddings $V_s(X) \subset \Delta^N$ give the Vietoris-Rips complexes the structure of oriented simplicial complexes.

Standard topological data analysis is a study of classical homotopy theoretic invariants, such as path components or homology groups, which do not see changes of orientation. This short paper is a preliminary study of invariants associated to Vietoris-Rips complexes equipped with fixed orientations.

The path category $P(K)$ of a simplicial complex $K$ is perhaps the most elementary of these invariants. The term "path category" originated in concurrency theory [1], while $P(K)$ is called the "fundamental category" (and is usually denoted by $\tau_1(K)$) in higher category theory [5].

The category $P(K)$ has the vertices of $K$ as objects, and its morphisms $x \to y$ are equivalence classes of oriented paths, or strings of 1-simplices, from $x$ to $y$, where the equivalence relation is defined by directed homotopies across 2-simplices. If $K$ is the triangulation of a finite cubical complex $L$, then the path category $P(K)$ is the category of execution paths in the higher dimensional automaton that is defined by $L$. More recently, path categories have been used to model ambiguity of execution for programs in distributed computing theory.

The functor $K \mapsto P(K)$ is the left adjoint of the nerve functor $C \mapsto BC$ for small categories $C$, and as such is similar to the fundamental groupoid functor $K \mapsto \pi(K)$, which is also defined by a left adjunction. The difference is that one can write down an algorithm to completely specify $P(K)$ from the simplices of $K$, while this is famously not possible for the fundamental groupoid.

The path category $P(K)$ and the algorithm that computes it are reviewed in the first two sections of this paper. This algorithm uses a 2-category "resolution" $P_2(K)$ for $P(K)$.

The path category algorithm is exponential in the size of $K$, but can be considerably simplified for computing the set of morphisms $P(K)(x, y)$ (execution paths) from the vertex (state) $x$ to $y$. This simplification is reviewed in the second section of this paper.

The general class of intended applications arises as follows. Suppose that the injection $X : \mathbf{N} \to Z$ defines an oriented data set in a metric space $Z$. Write $K_s(X)$ for the subcomplex of $\Delta^N$ whose non-degeneralte simplices are the (ordered) subsets $\sigma = \{x_0, \dots, x_k\}$ such that $d(x_i, x_j) \leq s$ for all $i, j$. The complex $K_s(X)$ is the oriented Vietoris-Rips complex — it is the Vietoris-Rips complex $V_s(X)$, with the orientation inherited from the ordering on $X$.

This general setup should be of particular use for time series $X : \mathbf{N} \to Z \times \mathbb{R}$, or "space-time" data sets, in which the real coordinate is a time variable. Intuitively, the total ordering on $X$ would be lexicographic with respect to the time variable, but the ordering of the members $(x,t)$ of the data set having a fixed time $t$ would need to be specified.

For an oriented data set $X$, the associated system of path categories $s \mapsto PK_s(X)$ has some familiar properties: $PK_0(X)$ is the discrete category on the set $X$, while $PK_s(X)$ is the totally ordered (hence contractible) poset $\mathbf{N} = \{0,1,\dots,N\}$ for $s$ sufficently large.

One is interested in the way the "communication channels" $PK_sX(x,y)$ change with varying $s$ for $x \leq y$ in $X$. There are no channels from $x$ to $y$ for $s=0$, while there is only one channel $x \to y$ in $PK_s(X)(x,y)$ for $s$ sufficiently large, and all sets of channels $PK_s(X)(x,y)$ can be computed with sufficiently robust computing equipment.

The path category is an orientation-dependent invariant that can be computed in a persistence context, and it is non-traditional in that it does not respect ordinary weak equivalences. In particular, there is a weak homotopy equivalence $BP_s(X) \xrightarrow{\simeq} K_s(X)$ since $BP_s(X)$ is the barycentric subdivison of $K_s(X)$, but the path category $P(BP_s(X)) = P_s(X)$ is a poset which does not reflect the structure of the path category $PK_s(X)$ of $K_s(X)$.

The path category functor is an invariant of higher category theory, but the general higher category approach is not quite right for persistence, essentially because the path category construction takes simplicial homotopies to natural transformations instead of natural isomorphisms.

Stability results (Theorem 2 and Corollary 3) are proved for ordered data sets in the third section, in which this distinction appears. These results have the form that one would expect from the results of [4], but their proofs are a little more rigid in the oriented case.

# Contents



# 1 The path category

The *path category* functor $P : s\mathbf{Set} \to \mathbf{cat}$ assigns a small category $PX$ to a simplicial set $X$.

The functor $P$ is left adjoint to the nerve functor $B : \mathbf{cat} \to s\mathbf{Set}$. As such, because the nerve $BC$ of a small category is a 2-coskeleton, $PX$ is defined by taking the free category on the graph $\operatorname{sk}_1 X$ defined by the vertices and

1-simplices of $X$, subject to relations $d_1\sigma = d_0\sigma \cdot d_2\sigma$

$$\begin{array}{ccc} x_0 & \xrightarrow{d_2\sigma} & x_1 \\ & {\scriptstyle d_1\sigma} \searrow & \downarrow {\scriptstyle d_0\sigma} \\ & & x_2 \end{array}$$

that are defined by the 2-simplices $\sigma : \Delta^2 \to X$ of $X$, where $x_0, x_1, x_2$ are the vertices

$$\Delta^0 \to \Delta^2 \xrightarrow{\sigma} X$$

of $\sigma$.

Here are some elementary observations:

1) The inclusion $\mathrm{sk}_2(X) \subset X$ of the 2-skeleton of $X$ induces a natural isomorphism
$$P(\mathrm{sk}_2(X)) \cong P(X).$$

2) The counit of the adjunction $\epsilon : P(BC) \to C$ is an isomorphism of categories for all small categories $C$.

3) The path category functor preserves products: there is a natural isomorphism
$$P(X \times Y) \cong P(X) \times P(Y).$$

4) There is a natural isomorphism of groupoids
$$G(PX) \xrightarrow{\cong} \pi(X)$$
where $G(PX)$ is the free groupoid on the category $PX$ and $\pi(X)$ is the fundamental groupoid of $X$.

The definition of $PX$ uses only the 2-skeleton of $X$, and statement 1) follows. Statement 2) is an exercise, and Statement 4) is an adjointness argument.

One proves statement 3) by observing that every product $X \times Y$ is a colimit of products of simplices $\Delta^n \times \Delta^m = B(\mathbf{n} \times \mathbf{m})$.

Statements 2) and 3) together imply that there is an isomorphism

$$P(X \times \Delta^1) \cong P(X) \times \mathbf{1},$$

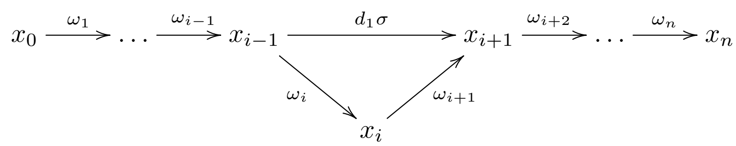
where $\mathbf{1}$ is the ordinal number poset $\{0, 1\}$, so that the path category functor takes homotopies to natural transformations.

Suppose that $x$ and $y$ are vertices of $X$. The morphisms from $x$ to $y$ in the free category on the graph $\mathrm{sk}_1 X$ consist of strings $P$ of 1-simplices (or paths)

$$P : x = x_0 \xrightarrow{\omega_1} x_1 \xrightarrow{\omega_2} \dots \xrightarrow{\omega_n} x_n = y,$$

where the length $n$ of the path can vary. Composition in this category is defined by concatenation of paths. These paths represent morphisms from $x$ to $y$ in the path category $PX$, after the following identifications are made:

1) If there is a picture

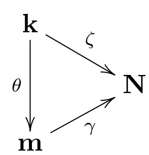



where $\sigma$ is a 2-simplex with boundary $(d_0\sigma, d_1\sigma, d_2\sigma) = (\omega_{i+1}, d_1\sigma, \omega_i)$, then the path $P$ and the (shorter) top string in the diagram represent the same morphism of $PX$.

2) The degeneracies $s_0(x)$ define identities in $PX$, since the 2-simplices $s_0\omega$ and $s_1\omega$ for a 1-simplex $\omega : x \to y$ have boundaries $(\omega, \omega, s_0(x))$ and $(s_0(y), \omega, \omega)$, respectively.

**Example**: There is an isomorphism $P\Delta^N \cong \mathbf{N}$, where $\mathbf{N}$ is the ordinal number poset $\{0, 1, \dots, N\}$, since $\Delta^N = B\mathbf{N}$.

The poset $\mathbf{N}$ has a resolution $\mathbf{N}_s$, which is a 2-category with 0-cells given by the objects $\{0, 1, \dots, N\}$, 1-cells $\gamma : i \to j$ given by ordinal number maps (paths) $\gamma : \mathbf{m} \to \mathbf{N}$ with $\gamma(0) = i$ and $\gamma(m) = j$. A 2-cell $\theta : \zeta \to \gamma$ is a diagram of ordinal number maps



(1)

where $\theta$ is endpoint preserving in the sense that $\theta(0) = 0$ and $\theta(k) = m$.

We can restrict the structure to a sub-2-category $N(\mathbf{N}_s)$ by assuming that all 1-cells and 2-cells are non-degenerate simplices in the sense that they are monomorphisms of ordinal numbers. Then the path component category $\pi_0 N(\mathbf{N}_s)$ is a copy of $\mathbf{N}$. Here, $\pi_0 N(\mathbf{N}_s)$ is a category with

$$\pi_0 N(\mathbf{N}_s)(i, j) = \pi_0(N(\mathbf{N}_s)(i, j)).$$

In other words, one computes path components for each category of morphisms $N(\mathbf{N}_s)(i, j)$.

Suppose that $K$ is a finite oriented simplicial complex, so that it admits an imbedding $K \subset \Delta^N$. We restrict the 2-category structure $N(\mathbf{N}_s)$ to a 2-category $NP_2(K)$. The 0-cells of $NP_2(K)$ are the vertices of $K$, a 1-cell $\gamma : x \to y$ is a non-degenerate simplex $\gamma : \mathbf{m} \to \mathbf{N}$ with $\gamma(0) = x$ and $\gamma(m) = y$, such that the string

$$x = \gamma(0) \xrightarrow{\omega_1} \gamma(1) \xrightarrow{\gamma_2} \dots \xrightarrow{\gamma_m} \gamma(m)$$

consists of 1-simplices of $K$. A 2-cell of $NP_2(K)$ consists of a diagram as in (1) such that the non-degenerate simplex $\theta$ is defined as a composite $\theta = d^{i_r} \cdot \dots \cdot d^{i_1}$, where each morphism $d^{i_j}$ is defined by a 2-simplex of $K$. Then Theorem 4.4 of [1] asserts that the canonical functor

$$\pi_0 NP_2(K) \to P(K) \tag{2}$$

is an isomorphism. Again, path components are computed for each category of morphisms $NP_2(K)(x,y)$, and we are asserting that there is a bijection

$$\pi_0(NP_2(K)(x,y)) \cong P(K)(x,y).$$

## 2 The path category algorithm

The isomorphism of (2) amounts to a 2-category resolution of the path category $P(K)$ for a finite oriented simplicial complex $K$, and we use this resolution to give an algorithm for computing the path category $P(K)$.

Broadly speaking this algorithm proceeds as follows:

1) Restrict to the 2-skeleton $\mathrm{sk}_2(K)$

2) Find all strings of non-degenerate 1-simplices

$$P : v_0 \xrightarrow{\omega_1} v_1 \xrightarrow{\omega_1} \dots \xrightarrow{\omega_k} v_k$$

in $K$.

3) Find all morphisms in the category $NP_2(K)(v,w)$ (composites of 2-simplices), for each pair of vertices $v \leq w$ of $K$.

4) Compute $\pi_0 NP_2(K)(v,w)$ for all $v \leq w$.

This is the path category algorithm. It was first coded by G. Denham in Macaulay (2008), and was later coded (and refined) in C by M. Misamore. Misamore's code is posted on github.com and hackage.haskell.org. See also [6].

There is a very big caveat: the path category algorithm is exponential in the size of the underlying complex, and toy examples can be found that will choke any given computer.

The practical game, however, is to attack smaller parts of the 2-category $NP_2(K)$ by computing path components of individual categories $NP_2(K)(v,w)$. There is a sequence of complexity reduction tricks for the analysis of a single set of morphisms $P(K)(v,w)$ that are developed in [3], and are described below. The general idea is to cut down the complex $K$ using full imbeddings $L \subset K$ where $v, w$ are vertices of $L$.

A subcomplex $L \subset K$ is *full* if

1) the subcomplex $L$ contains every path

$$v = x_0 \to x_1 \to \cdots \to x_n = w$$

of $K$, for all $v \leq w$ in $L$, and

2) if $\sigma$ is a simplex of $K$ such that all of its vertices are in $L$, then $\sigma$ is a simplex of $L$

Then one has the following:

**Lemma 1.** *Suppose that $L$ is a full subcomplex of $K$.*

*Then the induced functor $P(L) \to P(K)$ is a fully faithful imbedding. In particular, the functions $P(L)(v,w) \to P(K)(v,w)$ are bijections for all vertices $v, w$ of $L$.*

*Proof.* The proof of this result is an exercise. □

**Example**: Suppose that $v \leq w$ are vertices of $K$, and let $K[v,w] \subset K$ be the subcomplex of simplices $\sigma$ whose vertices $x$ satisfy $v \leq x \leq w$. Then $K[v,w] \subset K$ is a full imbedding.

The method is now to remove sources and sinks from a complex. A *source* is a vertex that has no inbound edges, and a *sink* is a vertex that has no outbound edges.

Here is a process for cutting down the complex $K[v,w]$:

1) Find the set $S$ of sources and sinks not equal to $v$ or $w$ in $K[v,w]$, and let $K_1 \subset K[v,w]$ be the subcomplex of simplices which do not have a vertex in $S$.

Then $K_1 \subset K[v,w]$ is a full subcomplex, since no path from $v'$ to $w'$ in $K[v,w]$ passes through a source or a sink. It follows that $P(K_1) \to P(K[v,w])$ is a fully faithful imbedding.

2) The subcomplex $K_1$ may have sources and sinks other than $v, w$. Remove them as above to produce a fully faithful imbedding $K_2 \subset K_1$.

3) Repeat these steps to produce a subcomplex $K_n$ which has no sources or sinks other than $v, w$.

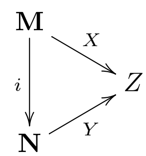
By construction, $P(K_n)(v,w) = P(K)(v,w)$. If $K_n = \{v,w\}$, then $K$ has no paths from $v$ to $w$ and $P(K)(v,w) = \emptyset$. Otherwise, there are paths from $v$ to $w$ in $K$, and $K_n$ is the full subcomplex of $K[v,w]$ of simplices $\sigma$ that have all vertices on a path from $v$ to $w$.

One then applies the path category algorithm to $K_n$ to find the morphism set $P(K)(v,w)$.

# 3 Stability

Suppose that $X \subset Y \subset Z$ are data sets in a metric space $Z$, where $Y$ has an orientation $Y : \mathbf{N} \to Z$.

Restriction of the orientation $Y : \mathbf{N} \to Z$ defines an orientation on $X$, and there is a commutative diagram of functions




where the function $i$ is an order preserving inclusion.

We write $K_s(X)$ for the oriented subcomplex of $\Delta^M$ whose simplices are oriented strings $[x_0, \dots, x_n]$ such that $d(x_i, x_j) \leq s$. The complex $K_s(X)$ is the oriented version of the Vietoris-Rips complex $V_s(X)$, and represents the same homotopy type.

We also have the usual filtration

$$X = K_0(X) \subset \cdots \subset K_s(X) \overset{\sigma}{\subset} K_t(X) \subset \cdots \subset \Delta^M, \ \ s < t,$$

of the simplex $\Delta^M$.

Suppose that the Hausdorff distance $d_H(X, Y)$ satisfies $d_H(X, Y) < r$ for some $r > 0$. This means that for every $y \in Y$ there is an $x \in X$ such that $d(x, y) < r$ for the metric $d$ of $Z$.

The ordering on $Y$ (and $X$) determines a retraction map $\theta : Y \to X$. In effect, given $y \in Y$, let $\theta(y)$ be the smallest element of $X$ such that $y \leq \theta(y)$ (if such exists). If $y > x$ for all $x \in X$, set $\theta(y) = \max_{x \in X}\{x\}$.

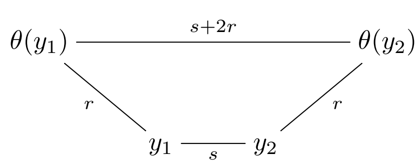
Suppose that $y_1 \leq y_2$. Then,

1) if $y_2 \leq \theta(y_2)$ then $y_1 \leq y_2 \leq \theta(y_2)$ so that $\theta(y_1) \leq \theta(y_2)$, and

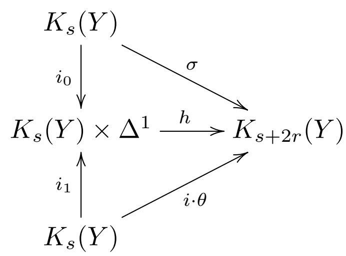
2) if $x \leq y_2$ for all $x \in X$, then $\theta(y_2) = \max_{x \in X}\ \{x\}$, so that $\theta(y_1) \leq \theta(y_2)$.

It follows that $\theta : Y \to X$ is order preserving. Observe that $\theta(x) = x$ for all $x \in X$.

We have the picture of distance bounds



These pictures define natural transformations of functors on the simplex level, and hence induce simplicial set maps $\theta : K_s(Y) \to K_{s+2r}(X)$ and a homotopy

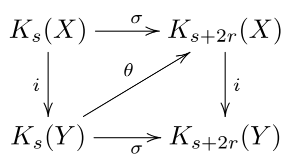


where $i$ denotes the inclusion of systems $i : K_s(X) \to K_s(Y)$.

We have proved the following:

**Theorem 2.** *Suppose that $X \subset Y$ are finite subsets of a metric space $Z$ such that $Y$ is totally ordered, and $d_H(X,Y) < r$. Define the order-preserving retraction map $\theta : Y \to X$ as above.*

*Then there are homotopy commutative diagrams*



(3)

*which are natural in $s$. The upper triangle commutes on the nose, and the lower triangle commutes up to a homotopy which fixes the image of $K_s(X)$.*

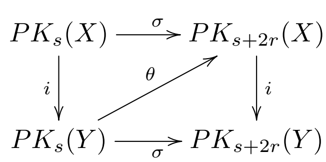
In the context for Theorem 2, write $PK_s(Y)$ for the path category of the oriented simplicial complex $K_s(Y)$.

Then $PK_0(Y) = Y$ is the discrete category with objects $Y$, and $P\Delta^Y$ is the poset $Y$, with its total ordering.

Applying the path category functor $P$ to the homotopy commutative diagram (3) gives the following:

**Corollary 3.** *Suppose given the setup $X \subset Y \subset Z$ of Theorem 2.*

*Then there is a diagram of functors*



(4)

*such that the upper triangle commutes, and the lower triangle commutes up a to natural transformation that is the identity on $PK_s(X)$.*

**Remarks**: 1) The diagram (4) is natural in $s$. The collection of these diagrams (for all $s$) is a homotopy interleaving of categories.

2) The homotopy (natural transformation) $i \cdot \theta \sim \sigma$ is not necessarily a natural isomorphism, so that the diagram (3) **is not** a homotopy commutative diagram of higher categories.

The variant of categorical homotopy theory that is relevant for the present application is older. It was pioneered by Quillen [7] and further developed by Thomason [8] — see also [2].

# References


[1] J. F. Jardine. Path categories and resolutions. *Homology Homotopy Appl.*, 12(2):231–244, 2010.

[2] J. F. Jardine. Homotopy theories of diagrams. *Theory Appl. Categ.*, 28:No. 11, 269–303, 2013.

[3] J.F. Jardine. Complexity reduction for path categories. Preprint, arXiv: 1909.08433 [math.AT], 2016.

[4] J.F. Jardine. Persistent homotopy theory. Preprint, arxiv: 2002:10013 [math.AT], 2020.

[5] A. Joyal. Notes on quasi-categories. Preprint, `http://ncatlab.org/nlab/show/Andre+Joyal`, 2008.

[6] Michael D. Misamore. Computing path categories of finite directed cubical complexes. *Applicable Algebra in Engineering, Communication and Computing*, pages 1–14, 2014.

[7] Daniel Quillen. Higher algebraic $K$-theory. I. In *Algebraic $K$-theory, I: Higher $K$-theories (Proc. Conf., Battelle Memorial Inst., Seattle, Wash., 1972)*, pages 85–147. Lecture Notes in Math., Vol. 341. Springer, Berlin, 1973.

[8] R. W. Thomason. Cat as a closed model category. *Cahiers Topologie Géom. Différentielle*, 21(3):305–324, 1980.