Posets, metric spaces, and topological data analysis

Rick Jardine

University of Western Ontario

September 16, 2020

Basic setup

X data set: Z metric space, $X \in D(Z) =$ finite subsets of Z.

• $P_s(X) =$ poset of subsets $\sigma \subset X$ such that $d(x,y) \leq s$ for all $x, y \in \sigma$. $s \in [0, \infty)$.

$P_s(X)$ is the poset of simplices of **Vietoris-Rips complex** $V_s(X)$. The **nerve** $BP_s(X)$ is **barycentric subdivision** of $V_s(X)$ (same homotopy type).

We have poset inclusions

$$\sigma : P_s(X) \subset P_t(X), \;\; s \leq t,$$

$P_0(X) = X$, and $P_t(X) = \mathcal{P}(X)$ (all subsets of X) for t suff large.

• $k \geq 0$: $P_{s,k}(X) \subset P_s(X)$ subposet of simplices σ such that each element $x \in \sigma$ has at least k distinct neighbours y such that $d(x,y) \leq s$.

$P_{s,k}(X)$ is the poset of simplices of the **degree Rips complex** $L_{s,k}(X)$. $BP_{s,k}(X)$ is the **barycentric subdivision** of $L_{s,k}(X)$

The nerve construction

The **nerve** BC of a category C is a simplicial set with n-simplices BC_n given by the set of strings of arrows

$$a_0 \to a_1 \to \cdots \to a_n$$

of length n in C, equivalently functors $\mathbf{n} \to C$, where

$$\mathbf{n} = \{0, 1, \dots, n\},$$

with the obvious poset structure.

Composition with the functors $\theta : \mathbf{m} \to \mathbf{n}$ defines the simplicial structure of BC.

Examples: 1) $B\mathbf{n} = \Delta^n$, the standard n-simplex in simplicial sets.

2) $BG = K(G, 1)$ for a group G, classifies principal G-bundles.

BC is also called the **classifying space** of C.

Stability

Theorem 1 (Rips stability).

Suppose $X \subset Y$ in $D(Z)$ such that $d_H(X,Y) < r$. There is a homotopy commutative diagram (homotopy interleaving)

$$\begin{array}{ccc} P_s(X) & \xrightarrow{\sigma} & P_{s+2r}(X) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ P_s(Y) & \xrightarrow[\sigma]{} & P_{s+2r}(Y) \end{array}$$

Corollary 2 (Stability for persistence invariants).

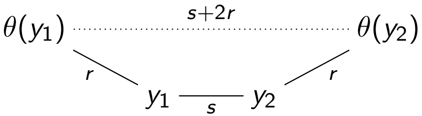
Same assumptions as Theorem 1. There are commutative diagrams

$$\begin{array}{ccc} H_k(V_s(X)) & \xrightarrow{\sigma} & H_k(V_{s+2r}(X)) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ H_k(V_s(Y)) & \xrightarrow[\sigma]{} & H_k(V_{s+2r}(Y)) \end{array}$$

There is a corresponding statement for π_0 (clusters).

Sketch proof

$y \in Y$: there is $\theta(y) \in X$ st. $d(y, \theta(y)) < r$ (from $d_H(X, Y) < r$).
$x \in X$: $\theta(x) = x$.



$\sigma = \{y_1, \ldots, y_k\}$ in $P_s(Y)$, then

$$\sigma \cup \theta(\sigma) = \{y_1, \ldots, y_k, \theta(y_1), \ldots, \theta(y_k)\} \in P_{s+2r}(Y)$$

and there are homotopies (natural transformations)

$$\sigma \ \subseteq \ \sigma \cup \theta(\sigma) \ \supseteq \ \theta(\sigma).$$

between poset morphisms $P_s(Y) \to P_{s+2r}(Y)$.

Homotopies

A natural transformation h between functors $f, g : C \to D$ is a diagram of functors

$$\begin{array}{ccc} C & & \\ {\scriptstyle i_0}\downarrow & \searrow^{f} & \\ C \times 1 & \xrightarrow{h} & D \\ {\scriptstyle i_1}\uparrow & \nearrow_{g} & \\ C & & \end{array} \qquad \begin{array}{ccc} f(a) & \xrightarrow{h} & g(a) \\ {\scriptstyle f(\alpha)}\downarrow & & \downarrow{\scriptstyle g(\alpha)} \\ f(b) & \xrightarrow[h]{} & g(b) \end{array}$$

where $1 = \{0 \leq 1\}$, $i_\epsilon(a) = (a, \epsilon)$.

$$B(C \times 1) \cong BC \times B1 = BC \times \Delta^1$$

$$\begin{array}{ccc} BC & & \\ {\scriptstyle i_0}\downarrow & \searrow^{f} & \\ BC \times \Delta^1 & \xrightarrow{h} & BD \\ {\scriptstyle i_1}\uparrow & \nearrow_{g} & \\ BC & & \end{array}$$

Stability: degree Rips

$D(Z)$ = finite subsets of a metric space Z.

Theorem 3.

Suppose $X \subset Y$ in $D(Z)$ such that $d_H(X^{k+1}_{dis}, Y^{k+1}_{dis}) < r$. There is a homotopy commutative diagram

$$\begin{array}{ccc} P_{s,k}(X) & \xrightarrow{\sigma} & P_{s+2r,k}(X) \\ i\downarrow & \overset{\theta}{\nearrow} & \downarrow i \\ P_{s,k}(Y) & \xrightarrow[\sigma]{} & P_{s+2r,k}(Y) \end{array}$$

X^{k+1}_{dis} is the set of subsets $\sigma \subset X$ having $k+1$ elements.

$X^{k+1}_{dis} \subset Z^{k+1}$.

Equivalence up to shift

NB: $V_*(X) := BP_*(X)$ for a while.

Suppose that $X \subset Y$ in $D(Z)$ and we have a homotopy interleaving

$$\begin{array}{ccc} V_s(X) & \xrightarrow{\sigma} & V_{s+2r}(X) \\ {\scriptstyle i}\downarrow & \overset{\theta}{\nearrow} & \downarrow{\scriptstyle i} \\ V_s(Y) & \xrightarrow[\sigma]{} & V_{s+2r}(Y) \end{array}$$

σ are **shift** maps.

1) $i : \pi_0 V_*(X) \to \pi_0 V_*(Y)$ is a $2r$-**monomorphism**: if $i([x]) = i([y])$ in $\pi_0 V_s(Y)$ then $\sigma([x]) = \sigma([y])$ in $\pi_0 V_{s+2r}(X)$

2) $i : \pi_0 V_*(X) \to \pi_0 V_*(Y)$ is a $2r$-**epimorphism**: given $[y] \in \pi_0 V_s(Y)$, $\sigma([y]) = i([x])$ for some $[x] \in \pi_0 V_{s+2r}(X)$.

3) All $i : \pi_n(V_*(X), x) \to \pi_n(V_*(Y), i(x))$ are $2r$-**isomorphisms**.

The map $i : V_*(X) \to V_*(Y)$ is a $2r$-**equivalence** of systems.

Systems

A **system** of spaces is a functor $X : [0, \infty) \to s\mathrm{Set}$, aka. a **diagram** of simplicial sets with index category $[0, \infty)$.

A **map** of systems $X \to Y$ is a natural transformation of functors defined on $[0, \infty)$.

$s\mathrm{Set}^{[0,\infty)}$ is the category of systems and natural transformations.

Examples

1) The functor $V_*(X)$, $s \mapsto V_s(X) = BP_s(X)$ is a system of spaces, for a data set $X \subset Z$.

2) If $X \subset Y$ in $D(Z)$, the induced maps $P_s(X) \to P_s(Y)$, $V_s(X) \to V_s(Y)$ define maps of systems

$P_*(X) \to P_*(Y)$ (posets) and $V_*(X) \to V_*(Y)$ (spaces).

Homotopy types

Homotopy theories of systems: oldest is the **projective structure** (Bousfield-Kan, 1972):

A map $f : X \to Y$ is a **weak equivalence** (resp. **fibration**) if each map $X_s \to Y_s$ is a weak equiv. (resp. Kan fibration) of simplicial sets.

A map $A \to B$ is a **projective cofibration** if it has the left lifting property with respect all maps which are trivial fibrations.

Lemma 4.

Suppose that $X \subset Y$ in $D(Z)$. Then $V_*(X) \to V_*(Y)$ is a projective cofibration.

The map $V_*(X) \to V_*(Y)$ is also a **sectionwise** cofibration, i.e. all maps $V_s(X) \to V_s(Y)$ are monomorphisms.

Fibrations and weak equivalences for the projective structure are defined sectionwise.

Controlled equivalences

Suppose that $f : X \to Y$ is a map of systems. Say that f is an r-**equivalence** if

1) the map $f : \pi_0(X) \to \pi_0(Y)$ is an r-isomorphism of systems of sets
2) the maps $f : \pi_k(X_s, x) \to \pi_k(Y_s, f(x))$ are r-isomorphisms of systems of groups, for all $s \geq 0$, $x \in X_s$, $k \geq 1$.

Observation: Suppose given a diagram of systems

$$\begin{array}{ccc} X_1 & \xrightarrow{f_1} & Y_1 \\ \text{sect } \simeq\downarrow & & \downarrow\simeq \text{ sect} \\ X_2 & \xrightarrow[f_2]{} & Y_2 \end{array}$$

Then f_1 is an r-equivalence iff f_2 is an r-equivalence.

Examples: Stability results. A sectionwise equivalence is a 0-equivalence.

A **controlled equivalence** is a map which is an r-equivalence for some $r \geq 0$.

Triangle axiom

Lemma 5.

Suppose given a commutative triangle

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & {\scriptstyle h}\searrow & \downarrow{\scriptstyle g} \\ & & Z \end{array}$$

If one of the maps is an r-equivalence, a second is an s-equivalence, then the third map is a $(r+s)$-equivalence.

Proof.

Set theory. □

Theorem 6.

Suppose that $i : A \to B$ is a sectionwise cofibration and an r-equivalence, and suppose given a pushout

$$\begin{array}{ccc} A & \longrightarrow & C \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle i_*} \\ B & \longrightarrow & D \end{array}$$

Then i_* is a sectionwise cofibration and a $2r$-equivalence.

Sketch (Whitehead theorem): There is a $2r$-interleaving

$$\begin{array}{ccc} A_s & \longrightarrow & FA_{s+2r} \\ \downarrow & \nearrow & \downarrow \\ B_s & \longrightarrow & FB_{s+2r} \end{array}$$

for a sectionwise fibrant model of i. The class of cofibrations admitting $2r$-interleavings is closed under pushout.

Persistent homotopy theory

(A'): Suppose given a commutative diagram

$$\begin{array}{ccc} A & \longrightarrow & C \\ & \searrow \quad \nearrow & \\ & B & \end{array}$$

If one of the maps is an r-equivalence, another is an s-equivalence, then the third is an $(r+s)$-equivalence.

(B): The composite of two cofibrations is a cofibration. Any isomorphism is a cofibration.

(C'): Cofibrations are closed under pushout. Given a pushout

$$\begin{array}{ccc} A & \rightarrow & C \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle i_*} \\ B & \rightarrow & D \end{array}$$

with i a cofibration and r-equivalence, then i_* is a cofibration and a $2r$-equivalence.

(D): For any object A there is at least one cylinder object $A \otimes \Delta^1$.

(E): All objects are cofibrant.

This is an adjusted list of axioms for a category of cofibations structure — works for projective or sectionwise cofibrations.

Lemma 7 (left properness).

Suppose given a pushout

$$\begin{array}{ccc} A & \xrightarrow{u} & C \\ {\scriptstyle i}\downarrow & & \downarrow \\ B & \xrightarrow[u_*]{} & D \end{array}$$

where i is a cofibration and u is an r-equivalence. Then u_* is a $2r$-equivalence.

There is also a **patching lemma**.

ep-metric spaces (following Spivak)

An extended pseudo-metric space (**ep-metric space**) (X, D) is a set X and a function $D : X \times X \to [0, \infty]$ such that

1) $D(x, x) = 0$,
2) $D(x, y) = D(y, x)$,
3) $D(x, z) \leq D(x, y) + D(y, z)$.

- Can have distinct x, y such that $D(x, y) = 0$ ("pseudo").
- Can have u, v such that $D(u, v) = \infty$ ("extended").

Every metric space (Y, d) is an ep-metric space via composition

$$Y \times Y \xrightarrow{d} [0, \infty) \subset [0, \infty].$$

A **morphism** $f : (X, d_X) \to (Y, d_Y)$ of ep-metric spaces is a function $f : X \to Y$ such that

$$d_Y(f(x), f(y)) \leq d_X(x, y) \text{ (non-expanding).}$$

$ep - \text{Met}$ is the category of ep-metric spaces and their morphisms.

Quotient construction

(X, d) an ep-metric space and $p : X \to Y$ a surjective function.

For $x, y \in Y$ set

$$D(x, y) = \inf_P \sum_{i=0}^{k} d(x_i, y_i),$$

"Polygonal path" P : pairs (x_i, y_i) in X with $x = p(x_0)$, $y = p(x_k)$, $p(y_i) = p(x_{i+1})$.

For $x, y \in X$, $P : x, y$ is path from $p(x)$ to $p(y)$, so $D(p(x), p(y)) \leq d(x, y)$.

Polygonal paths concatenate, so $D(x, z) \leq D(x, y) + D(y, z)$.

$D(x, x) = 0$ and $D(x, y) = D(y, x)$.

Quotient map $p : (X, d) \to (Y, D)$ satisfies universal property.

1) Suppose (X_i, d_i), $i \in I$ is a set of ep-metric spaces. There is an ep-metric D on $\bigsqcup_i X_i$, with

$$D(x,y) = \begin{cases} d_i(x,y) & \text{if } x, y \in X_i, \\ \infty & \text{if } x, y \text{ are in different summands.} \end{cases}$$

$\bigsqcup_i (X_i, d_i)$ is a **coproduct** in $ep - \mathrm{Met}$.

2) Suppose given morphisms $f, g : (X, d_X) \to (Y, d_Y)$ in $ep - \mathrm{Met}$. Form the set theoretic coequalizer

$$X \underset{g}{\overset{f}{\rightrightarrows}} Y \xrightarrow{p} C,$$

Then p is a surjective function, and we give C the quotient ep-metric D.

$$(X, d_X) \underset{g}{\overset{f}{\rightrightarrows}} (Y, d_Y) \xrightarrow{p} (C, D)$$

is a **coequalizer** in $ep - \mathrm{Met}$.

Excision

Suppose given subspaces $X, Y \subset Z$ of a finite metric space (Z, d) such that $Z = X \cup Y$.

$X \cup_{em} Y$ is **pushout** in $ep - \mathrm{Met}$.

The map $X \cup_{ep} Y \to Z$ is the identity on the underlying sets.

The metric D on $X \cup_{ep} Y$ is defined by

$$D(x, y) = \inf_{P} \{\sum_{i} d(x_i, x_{i+1})\},$$

indexed over sequences $P : x = x_0, x_1, \ldots, x_k = y$ such that each pair (x_i, x_{i+1}) is either in X or in Y.

Fact: The induced map

$$\pi_0(V_s(X, d) \cup V_s(Y, d)) \to \pi_0 V_s(X \cup_{ep} Y, D)$$

is a bijection.

Question: What about homotopy/homology groups?

Simplex categories

$s\mathrm{Set}^{[0,\infty]}$ is the diagram category of functors $[0,\infty] \to s\mathrm{Set}$.

Suppose that K is a simplicial set and $s \in [0,\infty]$.

$$L_sK_t = \begin{cases} K & \text{if } t \geq s, \\ \emptyset & \text{if } t < s. \end{cases}$$

Fact: $\hom(L_sK, X) = \hom(K, X_s)$ for $X : [0,\infty] \to s\mathrm{Set}$

A **simplex** of X is a morphism $\sigma : L_s\Delta^n \to X$ (an n-simplex of X_s). A **morphism of simplices** is a commutative diagram

$$\begin{array}{ccc} L_t\Delta^m & \xrightarrow{\theta} & L_s\Delta^n \\ & {\scriptstyle\tau}\searrow \quad \swarrow{\scriptstyle\sigma} & \\ & X & \end{array} \qquad \begin{array}{ccc} \Delta^m & \xrightarrow{\theta} & \Delta^n \\ {\scriptstyle\tau}\downarrow & & \downarrow{\scriptstyle\sigma} \\ X_t & \longleftarrow & X_s \end{array}$$

The **simplex category** is denoted by Δ/X.

Write U_s^n for the metric space structure on $\mathbf{n} = \{0, 1, \dots, n\}$ with $d(i,j) = s$.

If $\theta : \mathbf{m} \to \mathbf{n}$ is a poset map and $s \leq t$, then θ induces an ep-metric space map $\theta : U_t^m \to U_s^n$, since

$$d(\theta(i), \theta(j)) \leq s \leq t = d(i,j), \text{ if } i \neq j.$$

The assignment which takes a simplex $\sigma : L_s\Delta^n \to X$ to the space U_s^n defines a functor $\Delta/X \to ep - \text{Met}$. Then

$$\text{Re}(X) := \varinjlim_{L_s\Delta^n \to X} U_s^n$$

is the **realization** of X in $ep - \text{Met}$.

If $Y \in ep - \text{Met}$ then $S(Y)$ is the diagram with

$$S(Y)_{s,n} = \hom(U_s^n, Y).$$

Singular functor $S : ep - \text{Met} \to s\text{Set}^{[0,\infty]}$ is **right adjoint** to Re.

Realization, II

Suppose that $X : [0, \infty] \to s\mathrm{Set}$ is a diagram. Then

$$\varinjlim_{s} X_s = X_\infty$$

because ∞ is terminal in $[0, \infty]$.

Lemma 8.

Given $X : [0, \infty] \to s\mathrm{Set}$, *then* $\mathrm{Re}(X) \cong (X_\infty, D)$, *where*

$$D(x, y) = \inf_{P} \{\sum_{i=0}^{k} s_i\}$$

where P *is a polygonal path consisting of* 1*-simplices* $x_i \to y_i$ *in* X_{s_i}, *with* $x_0 \mapsto x$, $x_k \mapsto y$, *and* y_i, x_{i+1} *have the same image in* X_∞.

Examples

(X, d) a totally ordered finite ep-metric space.

The Vietoris-Rips complex $V_s(X)$ is a simplicial set with n-simplices

$$x_0 \leq x_1 \leq \cdots \leq x_n$$

with $d(x_i, x_j) \leq s$ for all i, j. $BP_s(X)$ is the barycentric subdivision of $V_s(X)$. Here, $s \leq \infty$.

Lemma 9.

X a totally ordered finite ep-metric space. Then $\mathrm{Re}(V_*(X)) \cong X$.

Proof.

$D(x, y) = \inf_P \sum d(x_i, x_{i+1})$, indexed over sequences $x = x_0, x_1, \ \ldots, x_k = y$. Then $d(x, y) \leq D(x, y)$ by the triangle identity.

$D(x, y) \leq d(x, y)$ because (x, y) is a path in some X_s. □

Canonical map

(X, d) a totally ordered finite ep-metric space.

The canonical map $\eta : V(X) \to S(\text{Re}(V(X)))$ has the form

$$\eta : V(X) \to S(X),$$

where η takes the simplex $x_0 \leq x_1 \leq \cdots \leq x_n$ to the sequence $(x_0, x_1, \ldots, x_k)$ (forgets the ordering — the sequence is a "bag of words").

Theorem 10.

(X, d) a totally ordered finite ep-metric space. Then each map

$$\eta : V_s(X) \to S(X)_s$$

is a weak equivalence of simplicial sets.

Proof uses simplicial approximation techniques. Show that $BNV_s(X) \to BNS(X)_s$ and $\pi : \text{sd}\, S(X)_s \to BNS(X)_s$ are weak equivalences.

UMAP (Healy-McInnes)

(X, d) is a finite ep-metric space.

Choose a **neighbourhood set** N_x for each $x \in X$.

Set $r_x = \max_{y \in N_x} \ d(x, y)$.

e.g. N_x is set of k nearest neighbours if X is totally ordered, for some k.

Set $(U_x, d_x) = \vee_{y \in N_x} \ (\{x, y\}, d)$ in $ep - \text{Met}$.
Then $d_x(y_1, y_2) = d(x, y_1) + d(x, y_2)$ for $y_1, y_2 \in N_x$.

Extend to an ep-metric d_x on X by setting $d_x(y, z) = \infty$ if either y or z is outside of U_x.

We have inclusions $X \subset V(X, d_x)$, $x \in X$. Form iterated pushout

$$V(X, N) = \vee_X \ V(X, d_x) \simeq \vee_X \ S(X, d_x).$$

The metrics d_x can be rescaled, but the diagram $V(X, N)$ is "the" **UMAP complex**.

UMAP algorithm: apply TDA machinery (e.g. π_0) to $V(X, N)$.

Comparisons

(X,d) is a finite totally ordered ep-metric space, with neighbourhoods $N = \{N_x,\ x \in X\}$.

$$\phi : V(X,N) = \vee_X\ V(X,d_x) \to V(X) \text{ canonical map.}$$

(x,y) in X is a **neighbourhood pair** if $y \in N_x$ or $x \in N_y$.

Graph $\Gamma(X,N) \overset{i}{\subset} V(X)$ with vertices X and edges all nbhd pairs.

Lemma 11.

$$\begin{array}{ccc} \pi_0 V(X,N)_s & \xrightarrow{\cong} & \pi_0\Gamma(X,N)_s \\ & \searrow^{\phi_*} & \downarrow i_* \\ & & \pi_0 V_s(X) \end{array}$$

If all 1*-simplices of* $V_s(X)$ *are nbhd pairs, then* i_* *is an iso.*

Example: $N_x = k$-nearest neighbours, $r_x = \max_{y \in N_x} d(x,y)$, $s < r_x$ for all x.

Comparisons II

Fact: $V(X,N)_\infty$ is a big wedge of circles.

$V(X,d_X) = \Delta^X = \Delta^N$ for $N+1 = |X|$, so $V(X,N)_\infty = \vee_N \Delta^n$ ($N+1$ summands).

Define $\mathrm{N} \to \Delta^N = X_i$, $0 \leq i \leq k$, $Y = \vee_{\mathrm{N}} X_i$ (iterated pushout).

Each X_i is contractible, so $Y/X_0 \simeq Y$, and

$$Y/X_0 = (X_1/\mathrm{N}) \vee \cdots \vee (X_k/\mathrm{N}) = (\Delta^N/\mathrm{N}) \vee \cdots \vee (\Delta^N/\mathrm{N})$$

and each

$$\begin{aligned}\Delta^N/\mathrm{N} &\simeq \Sigma\mathrm{N} \simeq \Sigma(S^0 \vee \cdots \vee S^0) \ \ (N \text{ summands, N pointed by } 0)\\ &\simeq S^1 \vee \cdots \vee S^1.\end{aligned}$$

Consequence: $V(X,N)_\infty \simeq \vee_{i=1}^{N^2} S^1$ $(N = |X| - 1)$.

References

J.F. Jardine.
Metric spaces and homotopy types.
Preprint, http://uwo.ca/math/faculty/jardine/, 2020.

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

Leland McInnes and John Healy.
Accelerated hierarchical density based clustering.
In *2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, November 18-21, 2017*, pages 33–42, 2017.

D.I. Spivak.
Metric realization of fuzzy simplicial sets.
Preprint, 2009.