Saistītas struktūras topoloģiskā datu analīzē

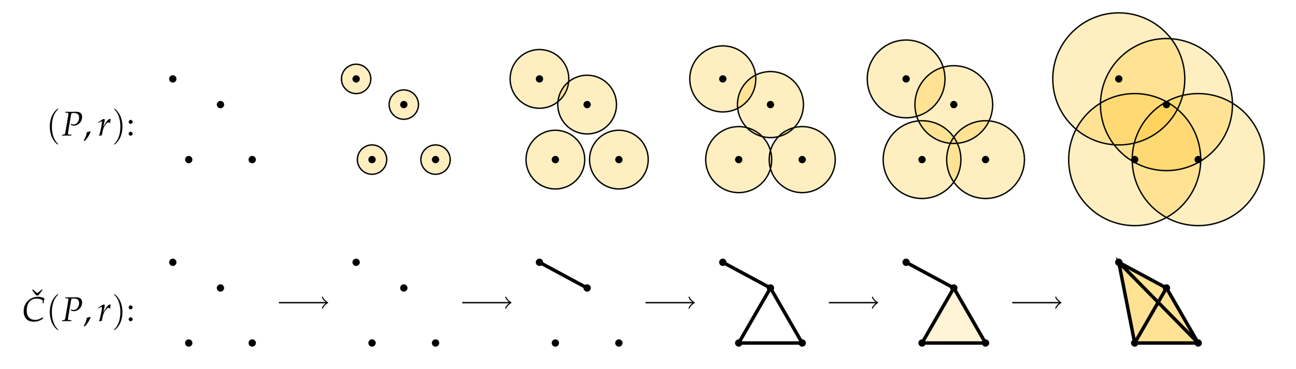
Secondary structures in topological data analysis

Jānis Lazovskis
University of Illinois at Chicago

Ievads — Introduction

Kā salīdzināt divas dažādas datu kopas? Kādas izskatās datu kopas "starp" šīm divām kopām? Ģeometrija, topoloģija un algebra dod precīzas atbildes.

Šis projekts radās no nesen izveidojušās topoloģiskās datu analīzes apakšjomas, kas pēta datu kopu attiecības ar homoloģijas grupām un homotopiju teoriju. Mērķis ir pārtulkot homoloģijas stabilitāti datu kopu stabilitātē.



$$H_0:\quad \mathbf{Z}^4 \longrightarrow \mathbf{Z}^4 \longrightarrow \mathbf{Z}^3 \longrightarrow \mathbf{Z} \longrightarrow \mathbf{Z} \longrightarrow \mathbf{Z}$$

$$H_1:\quad 0 \longrightarrow 0 \longrightarrow 0 \longrightarrow \mathbf{Z} \longrightarrow 0 \longrightarrow 0$$

How do we compare two samples from the same data set? What do other samples "in between" these two look like? Geometry, topology, and algebra give us answers.

This project is motivated by **persistent homology**, a subfield of topological data analysis which gives descriptions of sets using homology groups and homotopy theory. Homology is stable under perturbations, which we translate into the stability of data sets.

Topoloģijas pamati — Topological basics

topoloģiska telpa. Punktu kopa X ar vaļējas apakškopas $U \subseteq X$ jēdzienu.

nepārtraukta funkcija. Topoloģisku telpu atbilstība $f\colon X \to Y$ ar īpašību, ka katrai vaļējai apakškopai $U \subseteq Y$, kopa $f^{-1}(U) \subseteq X$ ir arī vaļēja.

daļēji sakārtota kopā. Punktu kopa X ar refleksīvu, anti-simetrisku un tranzitīvu attiecību $\leqslant$. Daļēji sakārtotas kopas ir topoloģiskas telpas, kurām vaļējas kopas rada kopas $U_x := \{y \in X : x \leqslant y\}$.

stratifikācija. Nepārtraukta funkcija no topoloģiskas telpas uz daļēji sakārtotu kopu. Stratifikācija ir *konusveidīga*, ja katram punktam ir apkārtne, kas ir kona stratifikācija.

šķipsna. Atbilstība $\mathcal{F}$ starp topoloģiskas telpas X vaļējām kopām U un objektiem $\mathcal{F}(U)$, ar saderīgām funkcijām $\rho_{UV}\colon \mathcal{F}(U) \to \mathcal{F}(V)$ kad $V \subseteq U$. Tās *stiebrs* punktā $x \in X$ ir $\varinjlim_{U \ni x} \mathcal{F}(U)$.

---

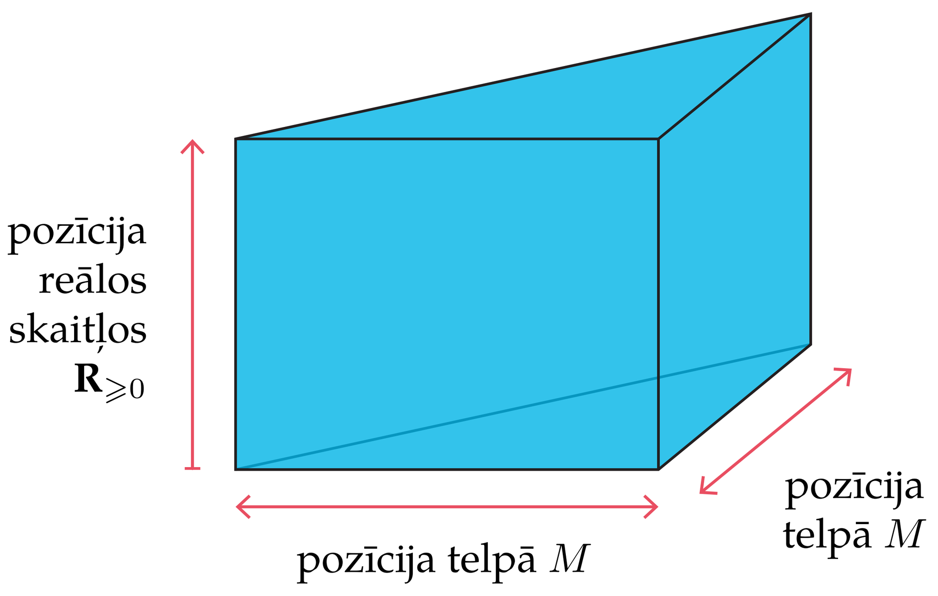
topological space. A set X with a notion of an open set $U \subseteq X$.

continuous function. An assignment $f\colon X \to Y$ of topological spaces such that $f^{-1}(U) \subseteq X$ is open fo every $U \subseteq Y$ open.

partially ordered set. A set X with a relation $\leqslant$ that satisfies reflexivity, anti-symmetry, and transitivity. Partially ordered sets are topological spaces, whose open sets are generated by the sets $U_x := \{y \in X : x \leqslant y\}$.

stratification. A continuous function from a topological space to a partially ordered set. A stratification is *conical* if every point has a neighborhood that is a stratified cone.

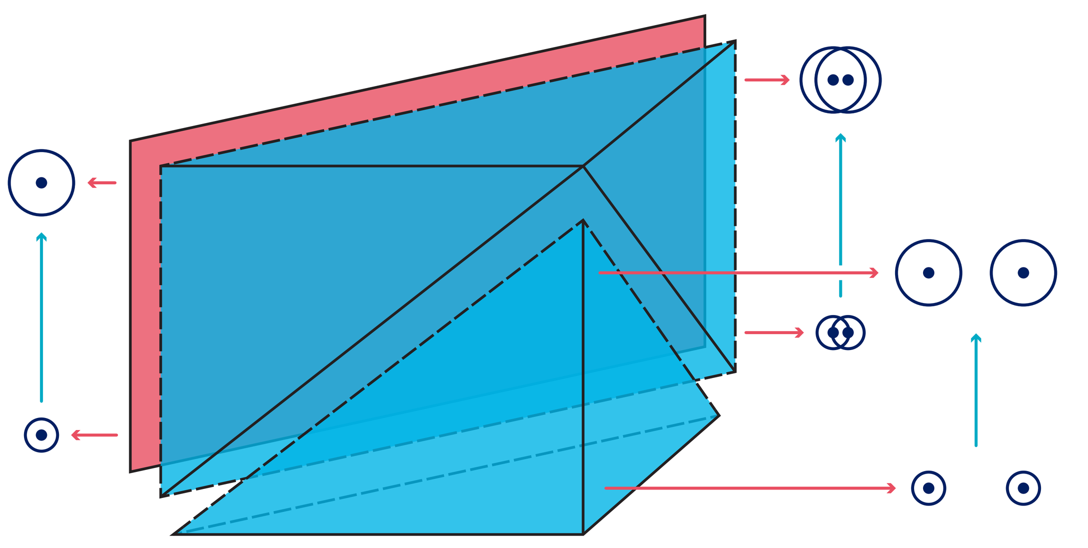
sheaf. An assignment $\mathcal{F}$ of objects $\mathcal{F}(U)$ to every open set U of a topological space X, with compatible functions $\rho_{UV}\colon \mathcal{F}(U) \to \mathcal{F}(V)$ whenever $V \subseteq U$. Its *stalk* at $x \in X$ is the categorical construction $\varinjlim_{U \ni x} \mathcal{F}(U)$.

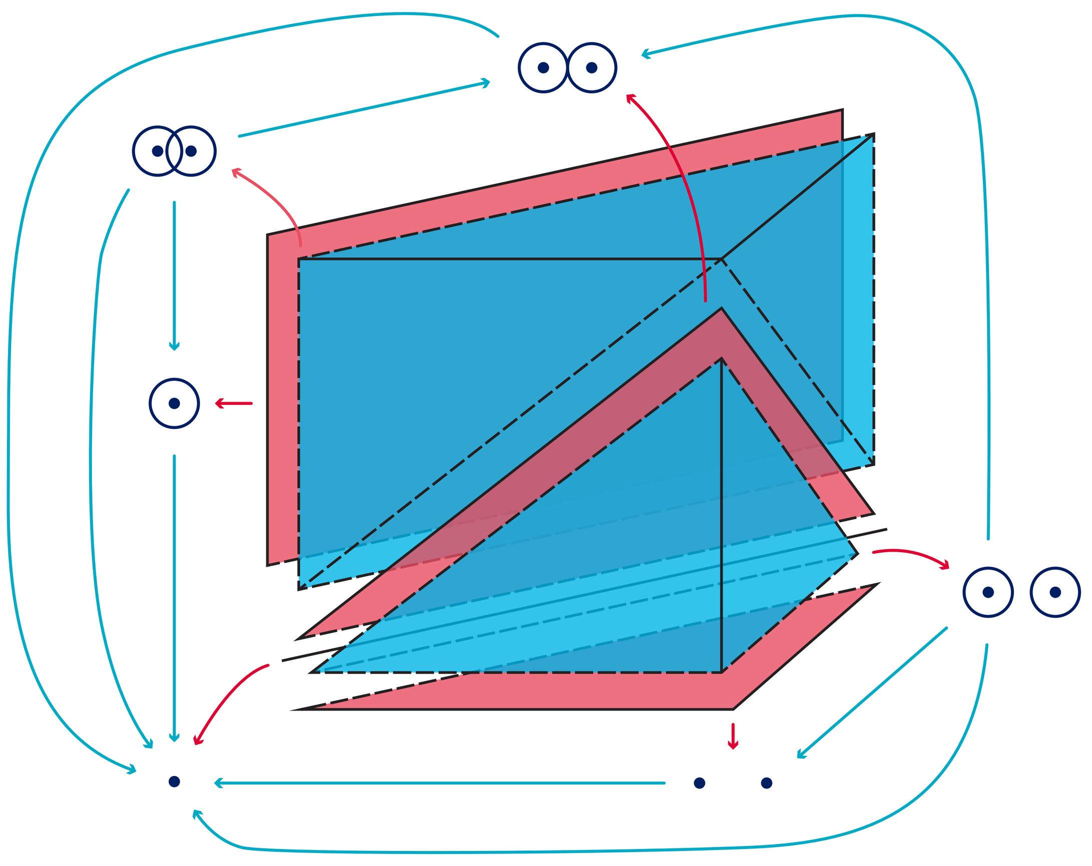


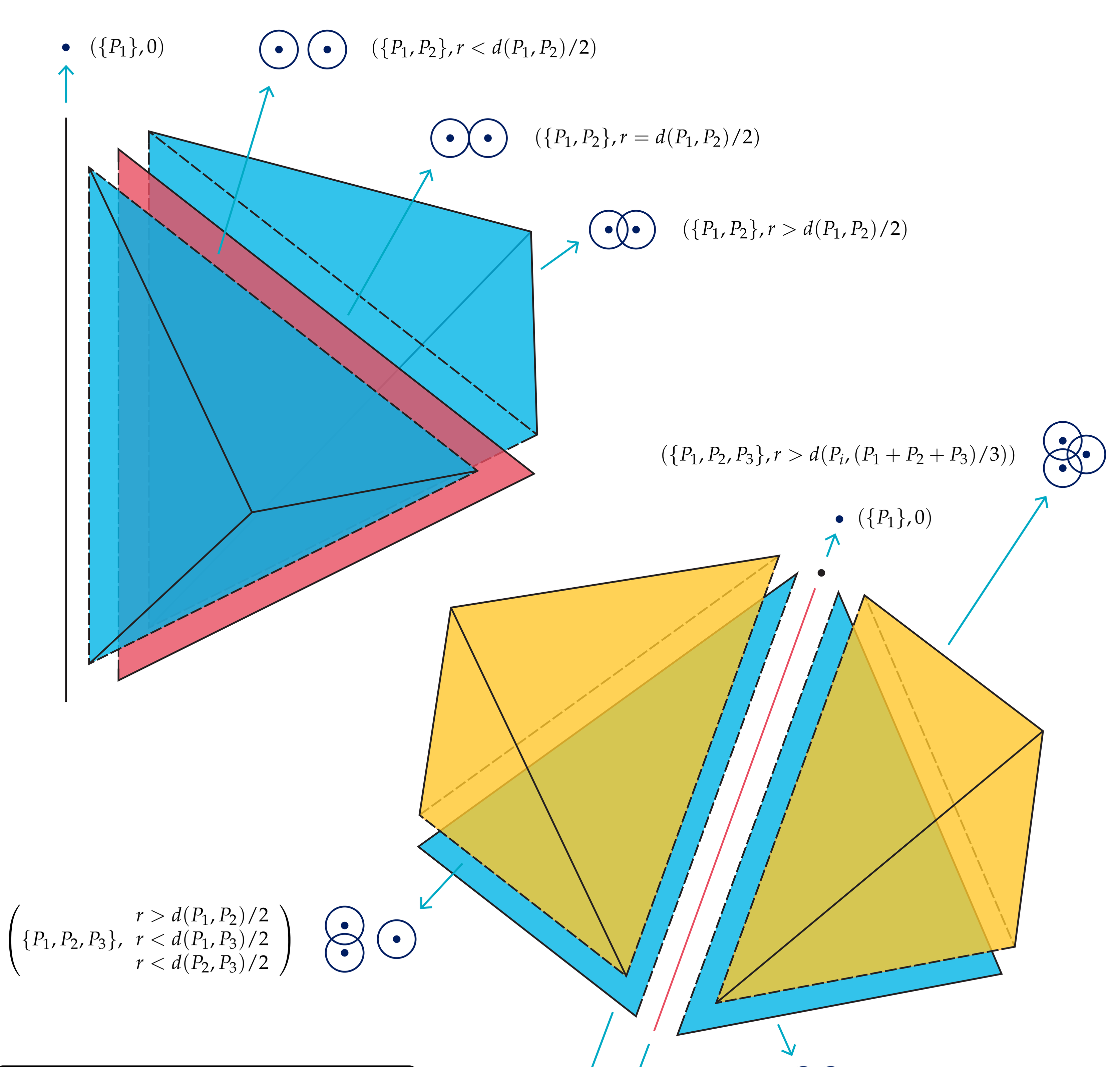
E.G. 2

Telpa $M = [0, 1]$ un tās stratifikācija caur funkcijām Č un $\check{C}_f$.

The space $M = [0, 1]$ and its stratification via Č and $\check{C}_f$.



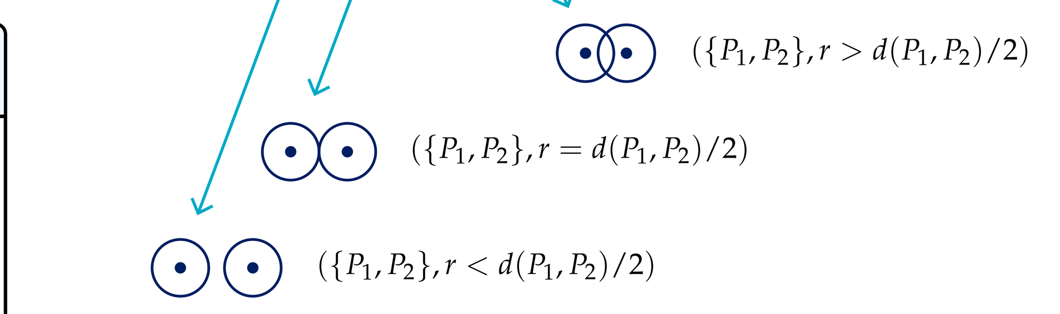




E.G. 1

Stratifikācijas zemu dimensiju daļas. Abām diagrammām ir kopīga divu dimensiju plakne.

Low dimensional pieces of a stratification. The two diagrams intersect in a 2-dimensional subspace.



Objekti un metodes — Objects and methods

Katrai telpai M ir saistīta **Rana telpa** $\mathrm{Ran}(M)$, kurai dota topoloģija no kopu Hausdorfa attāluma. Katram objektam P Rana telpā ir saistīts Čeha rādijs $\check{c}r(P, r)$, kas atbilst slēgto ložu $B(p \in P, r)$ šķēluma izmēram, vai negatīvam attālumam līdz ložu netukšam šķēlumam. Rana telpa ir

Every space M has an associated **Ran space** $\mathrm{Ran}(M)$, with topology induced by the Hausdorff distance of subsets of M. Every element P in the Ran space has an associated Čech radius $\check{c}r(P, r)$, which gives the size of the intersection of the balls $B(p \in P, r)$, or the negative distance their non-empty intersection. Set

$$\mathrm{Ran}^{\leqslant n}(M) := \{P \subseteq M : 0 < |P| \leqslant n\}.$$

Definition: Let SC be the category of **simplicial complexes** and simplicial maps. A simplicial complex C is a pair $(V(C), S(C))$, where $V(C)$ is a set and $S(C) \subseteq P(V(C))$ is closed under taking subsets. Let SCF be the category of **frontier simplicial complexes** and simplicial maps. A frontier simplicial complex C is a triple $(V(C), S(C), F(C))$, where $(V(C), S(C))$ is a simplicial complex and $F(C) \subseteq S(C)$ is closed under taking supersets in $S(C)$. Define the Čech functions as

$$\begin{aligned} \check{C}\colon \mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant 0} &\to \mathsf{SC}, \\ (P, r) &\mapsto (V, S), \\ \check{C}_f\colon \mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant 0} &\to \mathsf{SCF}, \\ (P, r) &\mapsto (V, S, F), \end{aligned} \qquad \begin{aligned} V &= P, \\ P' \in S &\iff \check{c}r(P', r) \geqslant 0, \\ P' \in F &\iff \check{c}r(P', r) = 0. \end{aligned}$$

Order SC and SCF by setting $C \leqslant D$ whenever there is a simplicial map $D \to C$ that is surjective on vertices V (and injective on frontier simplices F). Define a functor from the **homotopy category** of SC-**exit paths** to SC by

$$\begin{aligned} F\colon \mathrm{Ho}(\mathrm{Sing}^{\mathsf{SC}}(\mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant 0})) &\to \mathsf{SC}, \\ (P, r) &\mapsto \check{C}(P, r), \\ ([\gamma]\colon (P, r) \to (Q, s)) &\mapsto \left(\check{\gamma}\colon \check{C}(Q, s) \to \check{C}(P, r)\right), \end{aligned}$$

where $\check{\gamma}$ is the unique homotopy-invariant simplicial map induced by the equivalence class of 1-simplices $[\gamma]$ in the homotopy category. This is a cofinal functor. Every not necessarily open $U \subseteq \mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant n}(M) \times \mathbf{R}_{\geqslant 0}$ induces another cofinal functor $F_U\colon \mathrm{Ho}(\mathrm{Sing}^{\mathsf{SC}}(U)) \to \mathsf{SC}$.

Rezultāti — Results

Teorēma. Dotajā kontekstā:

1. Čeha funkcijas Č un $\check{C}_f$ ir nepārtrauktas.
2. Ja M ir gabaliem lineāra, pastāv konusveidīga $\mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant 0}$ stratifikācija, kas sader ar $\check{C}_f$.
3. Funktors $\mathcal{F}(U) = \varinjlim F_U$ definē ko-stratificētu ko-šķipsnu uz $\mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant 0}$ ar vērtību lineārās topoloģiskās telpās. Tās ko-stiebrs ir $\mathcal{F}_{(P,r)} = \varprojlim_{U \ni (P,r)} \mathcal{F}(U) = \check{C}(P, r)$.

Theorem. In the context above:

1. The Čech maps Č and $\check{C}_f$ are continuous.
2. If M is piecewise-linear, there exists a conical stratification of $\mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant 0}$ compatible with $\check{C}_f$.
3. The functor $\mathcal{F}(U) = \varinjlim F_U$ defines an SC-valued coconstructbile cosheaf on $\mathrm{Ran}^{\leqslant n}(M) \times \mathbf{R}_{\geqslant 0}$, with costalk $\mathcal{F}_{(P,r)} = \varprojlim_{U \ni (P,r)} \mathcal{F}(U) = \check{C}(P, r)$.

Corollary. This cosheaf recovers the persistent homology of the point sample P on the subspace $\{P\} \times \mathbf{R}_{\geqslant 0}$.

Atsauksmes — References

[1] Jacob Lurie (2017), *Higher Algebra* (Section 5.5.1, Appendix A).

[2] Masahiro Shiota (1997), *Geometry of subanalytic and semialgebraic sets.*