Lecture notes for Math 635: quantitative aspects of Morse and Floer theory
USC Spring 2022

Kyler Siegel

Contents

LECTURE 1

Introduction

Let M^n be a closed smooth manifold of dimension n. There are many equivalent models for its (co)homology:

- singular homology $H_*^{\text{sing}}(M;\mathbb{Z})$
- simplicial homology $H_*^{\text{simp}}(M;\mathbb{Z})$
- de Rham cohomology $H^*_{\text{dR}}(M;\mathbb{R})$
- Morse homology $H_*^{\text{Morse}}(M;\mathbb{Z})$,

etc. A basic fact from algebraic topology is that singular and simplicial homology are isomorphic, although the latter requires a choice of triangulation. Similarly, a basic fact from smooth manifold theory is that de Rham cohomology (i.e. closed differential forms modulo exact ones) is isomorphic to singular or simplicial cohomology over the real numbers.

The last model, Morse homology, is the one most relevant for this course. We begin by recalling the basics of Morse theory.

1.1. First look at Morse theory

DEFINITION 1.1. *A smooth function $f : M \to \mathbb{R}$ is* **Morse** *if all of its critical points are nondegenerate. Here $p \in M$ is a* **critical point** *if $df|_p = 0$ (i.e. $\partial_{x_1} f(p) = \cdots = \partial_{x_n} f(p) = 0$ in local coordinates $x_1, \ldots, x_n$ near p) and such a critical point is* **nonsingular** *if the Hessian $d^2 f|_p$ is nonsingular (i.e. the determinant of the symmetric $n \times n$ matrix $(\partial_{x_i}\partial_{x_j} f(p))$ is nonzero).*

LEMMA 1.2 (Morse lemma). *If $f : M \to \mathbb{R}$ is a Morse function and $p \in M$ is critical point, we can find local coordiantes $x_1, \ldots, x_n$ near p such that*

$$f(x_1, \ldots, x_n) = f(p) - x_1^2 - \cdots - x_s^2 + x_{s+1}^2 + \cdots + x_n^2.$$

Here s is called the **(Morse) index** *of f at p, denoted by* $\text{ind}_f(p)$.

Remark 1.3. If $p \in M$ is a **regular point** (i.e. not a critical point), then we can find local coordinates near p such that $f(x_1, \ldots, x_n) = x_1$. This is a consequence of the inverse function theorem. ◊

Example 1.4. We have $\text{ind}_f(p) = 0$ if and only if p is a local minimum, $\text{ind}_f(p) = n$ if and only if p is a local maximum, and otherwise p is a saddle point. ◊

The basic philosophy of Morse theory is to study the relationship between the topology of M and properties of a Morse function $f : M \to \mathbb{R}$. Consider the sublevel sets

$$S_{\leq a} := f^{-1}((-\infty, a]).$$

We have:

- $S_{\leq a} = \varnothing$ for $a < \min f$
- $S_{\leq a} = M$ for $a > \max f$
- $S_{\leq a} \subset S_{\leq a'}$ for $a \leq a'$.

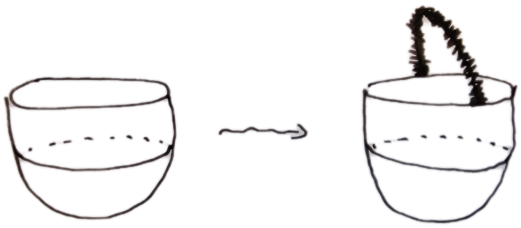

FIGURE 1.1. Attaching a 2-dimensional 1-handle.

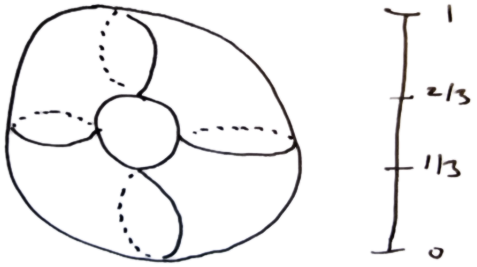

FIGURE 1.2. The height function $\mathbb{T}^2 \to \mathbb{R}$.

NOTATION 1.5. *Let* $\mathrm{crit}_i(f)$ *denote the set of index* i *critical points of* f*, and put* $\mathrm{crit}(f) := \cup_{i=0}^n \mathrm{crit}_i(f)$*. Note that* $f(\mathrm{crit}(f)) \subset \mathbb{R}$ *is the set of critical values of* f*.*

The following could be called the "fundamental theorem of Morse theory":

THEOREM 1.6. *For* $a \leq a'$*,* $S_{\leq a'}$ *deformation retracts onto* $S_{\leq a}$ *if* $[a, a'] \cap f(\mathrm{crit}(f)) = \varnothing$*. On the other hand, if (for simplicity) there is a single critical point* p *in* $f^{-1}([a, a'))$ *with* $a < f(p) < a'$*, then* $S_{\leq a}$ *is homotopy equivalent to* $S_a \cup H$*, where* H *is an* n*-dimensional handle of index* $\mathrm{ind}_f(p)$*.*

Recall that a n-dimensional handle of index k is of the form $\mathbb{D}^k \times \mathbb{D}^{n-k}$, glued to a manifold with boundary along $\partial\mathbb{D}^k \times \mathbb{D}^{n-k}$. This can be viewed as a "thickened" version of attaching a k-cell to a CW complex. See Figure 1.1 for a cartoon.

Example 1.7. Figure 1.2 depicts a Morse function $\mathbb{T}^2 \to \mathbb{R}$ given by the height. Figure 1.3 shows how the sublevel sets vary with a and the corresponding handle attachments as promised by Theorem 1.6 ◊

Remark 1.8. A function $f : \mathbb{T}^2 \to \mathbb{R}$ must have a global minimum and a global maximum since $\mathbb{T}^2$ is compact. From the above theorem, it's easy to see that it must also have at least one saddle point, i.e. there's no way to get $\mathbb{T}^2$ by a sequence of 0-handle and 2-handle attachments. ◊

1.2. Morse homology basics and consequences

Let $f : M^n \to \mathbb{R}$ be a Morse function. Let μ be a Riemannian metric on M which satisfies the Morse–Smale condition (roughly this means that μ is "generic"). Put

$$C_i^{\mathrm{Morse}}(f; \mathbb{K}) := \mathbb{K}\langle \mathrm{crit}_i(f) \rangle$$

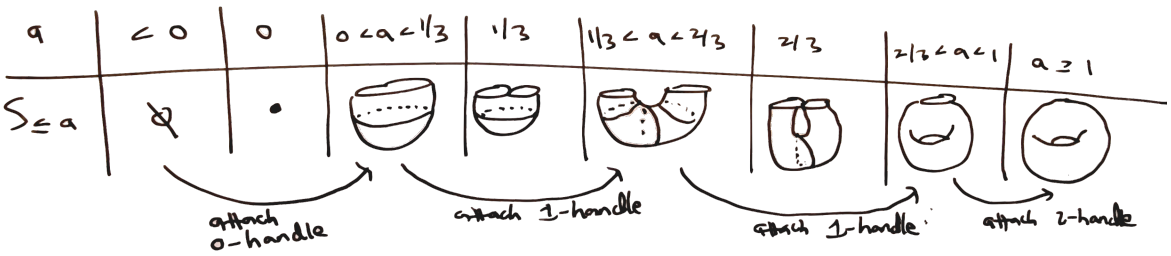

FIGURE 1.3. The sublevel sets $S_{\leq a}$ of the height function $\mathbb{T}^2 \to \mathbb{R}$ as a increases.

denote the module over some chosen coefficient ring $\mathbb{K}$ which is freely generated by the index i critical points of f. We have boundary operators:

$$C_n^{\text{Morse}}(f;\mathbb{K}) \xrightarrow{\partial_n} C_{n-1}^{\text{Morse}}(f;\mathbb{K}) \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_1} C_0^{\text{Morse}}(f;\mathbb{K}).$$

The precise definition of ∂_k is somewhat technical, but roughly it counts "gradient flow lines" between critical points with Morse index differing by 1. Namely, given $p_+ \in C_k^{\text{Morse}}(f;\mathbb{K})$ and $p_- \in C_{k-1}^{\text{Morse}}(f;\mathbb{K})$, we consider maps $u : \mathbb{R} \to M$ satisfying

$$\begin{cases} \lim_{s\to\pm\infty} u(s) = p_\pm \\ (\partial_s u)(s) = -(\nabla_\mu f)(u(s)) \text{ for all } s \in \mathbb{R}. \end{cases}$$

Then we have $\partial_k(p_-) = \text{sign}(u)p_+ + \cdots$, with contributions from all other gradient flow lines with negative asymptotic p_-. Here the gradient flow lines are counted modulo translation in the s variable, and with a certain sign, $\epsilon(u) \in \{1,-1\}$ (which we can ignore if say $\mathbb{K} = \mathbb{Z}/2$). We then put

$$H_k^{\text{Morse}}(f;\mathbb{K}) := \ker(\partial_k)/\text{im}(\partial_{k+1}).$$

THEOREM 1.9. *We have* $H_k^{\text{Morse}}(f;\mathbb{K}) \cong H_k^{\text{sing}}(M;\mathbb{K})$.

As a first consequence, observe that we must have

$$|\text{crit}_i(f)| \geq \text{rank}\, H_i(M;\mathbb{Z}).$$

Example 1.10. Any Morse function $f : \mathbb{T}^2 \to \mathbb{R}$ must have a least *two* saddle points, since $H_2(\mathbb{T}^2;\mathbb{Z}) \cong \mathbb{Z}^2$ has rank two. ◊

A second consequence is:

$$\sum_{i=0}^{n} (-1)^i |\text{crit}_i(f)| = \chi(M),$$

where $\chi(M)$ denotes the Euler characteristic of M. Indeed, a basic fact about chain complexes is that the alternating sum of ranks is unchanged after passing to homology.

Example 1.11. Figure 1.4 depicts a Morse function $S^2 \to \mathbb{R}$ with four critical points. Namely, we have $\text{crit}_0 = \{p_0\}$, $\text{crit}_1 = \{p_1\}$, $\text{crit}_2 = \{p_2, p_3\}$. ◊

Example 1.12. There exists a Morse function $f : \mathbb{CP}^n \to \mathbb{R}$ all of whose critical points have even Morse index. Hence the Morse complex is

$$C_{2n}^{\text{Morse}}(f;\mathbb{K}) \to 0 \to C_{2n-2}^{\text{Morse}}(f;\mathbb{K}) \to \cdots \to 0 \to C_0^{\text{Morse}}(f;\mathbb{K}),$$

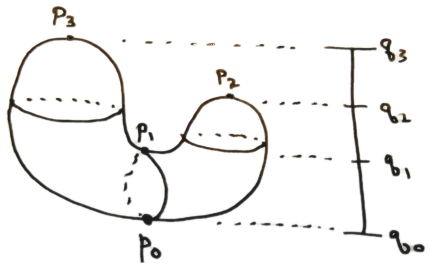

FIGURE 1.4. A Morse function $S^2 \to \mathbb{R}$ with four critical points.

and therefore we have $H_i^{\text{Morse}}(f;\mathbb{K}) = C_i^{\text{Morse}}(f;\mathbb{K})$ for $i = 0,\dots,n$. In particular,

$$H_i(\mathbb{CP}^n;\mathbb{Z}) \cong \mathbb{Z}^{|\text{crit}_i(f)|}.$$

◊

Example 1.13. We have $p \in \text{crit}_i(f)$ if and only if $p \in \text{crit}_{n-i}(-f)$. By dualizing the Morse complex for f, we get a complex

$$C_n^\vee(f) \xleftarrow{\partial_n^\vee} C_{n-1}^\vee(f) \xleftarrow{\partial_{n-1}^\vee} \leftarrow \cdots \xleftarrow{\partial_1^\vee} C_0^\vee(f)$$

which computes the Morse *co*homology of f, and this is precisely identified with the Morse complex of $-f$:

$$C_0(-f) \leftarrow C_1(-f) \leftarrow \cdots \leftarrow C_n(-f).$$

We then have $H^i_{\text{Morse}}(-f) \cong H^{\text{Morse}}_{n-i}(f)$, which is Poincaré duality. ◊

In Example 1.11, we have $\partial_1(p_1) = p_0 - p_0 = 0$, $\partial_2(p_2) = p_1$, and $\partial_2(p_3) = p_1$. Note that there are e.g gradient flow trajectories from p_2 to p_0, but these do not appear in the appear complex since these critical points index differing by two rather than one. We see that p_1 is a boundary, and hence does not contribute to the homology of S^2. A natural question is whether the critical point p_1, and particularly the gradient flow trajectory from p_2 to p_1, is "visible" in some sense. It turns out that this flow line does not contribute to the Morse homology, but it does contribute to the *filtered* Morse homology. In fact, we will see the filtered Morse homology imposes restrictions on the "geometry" of the function $f : S^2 \to \mathbb{R}$.

LECTURE 2

Quantitative Morse homology

2.1. Toy problem

Let M^n be a closed manifold. For a function $f : M \to \mathbb{R}$, let $||f|| := \max_{x \in M} |f(x)|$ denote its "uniform norm". In the following, let $h : S^2 \to \mathbb{R}$ denote the Morse function with four critical points depicted in Figure 1.4. Let $\mathcal{F}$ denote the set of Morse functions on S^2 with exactly two critical points (note that these necessarily have index 0 and 2 respectively).

PROBLEM 2.1. *What is* $\inf_{f \in \mathcal{F}} ||h - f||$*? In other words, how well can the Morse function* h *with four critical points be approximated by a Morse function with only two critical points?*

We give a (partial) answer to the above toy problem with the following:

PROPOSITION 2.2. *We have* $\inf_{f \in \mathcal{F}} ||h - f|| \geq \frac{1}{2}(q_2 - q_1)$.

IDEA OF PROOF. We proceed as follows (with the various ingredients to be introduced shortly):

(1) to f we associate a *filtered chain complex* $C_*^{\text{Morse}}(f)$
(2) to this filtered chain complex $C_*^{\text{Morse}}(f)$ we associate a *persistence module* $V(f)$
(3) to the persistence module $V(f)$ we associate a *barcode* $\mathcal{B}(V(f))$
(4) barcode $\mathcal{B}(V(f))$ has a *boundary depth* $\beta_1(\mathcal{B}(V(f))) \in \mathbb{R}_{\geq 0}$.

Roughly, a persistence module is a collection of $\mathbb{K}$ modules V_t indexed by the real numbers, along with maps $V_s \to V_t$ for all $s < t$ which are coherent under compositions. A barcode is roughly a multiset of interals - see Figure 2.1 for a pictorial representation. The boundary depth of a barcode is by definition the length of the longest finite bar, or zero in the event that there are no finite bars.

In turns out that all of these objects admit natural metrics:

(1) for functions $f, g : M \to \mathbb{R}$ we use the uniform distance $||f - g||$
(2) for two persistence modules V, W, we have the *interleaving distance* $d_{\text{int}}(V, W)$
(3) for two barcodes $\mathcal{B}, \mathcal{B}'$, we have the *bottleneck distance* $d_{\text{bot}}(\mathcal{B}, \mathcal{B}')$.

Moreover, we have:

(1) the association $f \mapsto V(f)$ is a 1-Lipschitz map from the set of Morse functions equipped with the uniform distance to the set of persistence modules equipped with the interleaving distance
(2) the association $V \mapsto \mathcal{B}(V)$ is an isometry from the set of persistence modules equipped with the interleaving distance to the set of barcodes equipped with the bottleneck distance
(3) the boundary depth is a 2-Lipshitz map from the set of barcodes equipped with the bottleneck distace to $\mathbb{R}$ (with the Euclidean metric).

FIGURE 2.1. Left: an example of a barcode, i.e. a multiset of intervals. Note that some are finite and some are infinite, and some intervals may be repeated. Center: the barcode of the Morse function $h : S^2 \to$ depicted in Figure 1.4. Right: the barcode of a Morse function $f : S^2 \to \mathbb{R}$ with exactly two critical points.

Recall that a map $F : X \to Y$ between metric space (X, μ_X) and (Y, μ_Y) is *K-Lipschitz* if $\mu_Y(F(x_1), F(x_2)) \leq K\mu_X(x_1, x_2)$ for all $x_1, x_2 \in X$. Moreover, we claim that the barcodes of h and $f \in \mathcal{F}$ are as depicted in Figure 2.1. As we will see, the infinite bars correspond to homology classes, while the finite bars record features of a more quantitative nature.

Taking this all for granted for the moment, we can now finish off the proof. Namely, for $f \in \mathcal{F}$ we have:

$$\begin{aligned} ||f - h|| \geq d_{\text{int}}(V(f), V(h)) = d_{\text{bot}}(\mathcal{B}(V(f)), \mathcal{B}(V(g))) &\geq \tfrac{1}{2}|\beta_1(\mathcal{B}(V(f))) - \beta_1(\mathcal{B}(V(h)))| \\ &= \tfrac{1}{2}|0 - (q_2 - q_1)| \\ &= \tfrac{1}{2}|q_2 - q_1|. \end{aligned}$$

□

2.2. Filtered chain complexes

Let us now introduce some of these objects in the above proof more carefully. We begin with:

DEFINITION 2.3. *A* **filtered chain complex** *is a chain complex (C, ∂) and a collection of subcomplexes $C^{<r} \subset C$, $r \in \mathbb{R}$, such that $C^{<r} \subset C^{<r'}$ for $r < r'$ and $\bigcup_{r \in \mathbb{R}} C^{<r} = C$.*

Here $C^{<r}$ being a subcomplex of C means that from the differential ∂ preserves $C^{<r}$, given an induced differential $C^{<r} \to C^{<r}$ (which we often still denote by ∂). Note that the chain complex in C need not be graded, although it will often be the case that our chain complex has a natural $\mathbb{Z}$ grading (as in singular chains) or at least a $\mathbb{Z}/2$ grading. We typically write C_* instead of C if we wish to emphasize the grading.

Remark 2.4. In the future we will often want to impose some additional technical assumptions on our filtered chain complexes. For example, it is often useful to assume that $C^{<r} = \{0\}$ for r sufficiently small. ◊

Example 2.5. Given a continuous function $f : M \to \mathbb{R}$, put

$$C_k^{\text{sing},<r} := C_k^{\text{sing}}(S_{<a}),$$

with $S_{<a} := f^{-1}((-\infty, a))$. In other words, C_k^{sing} consists of linear combinations of continuous maps from the k-simplex Δ^k to M whose images lie "below" the level set $\{f = a\}$. This makes C_*^{sing} into the structure of a filtered chain complex. ◊

2.3. Filtered Morse theory

If $f : M \to \mathbb{R}$ is a Morse function, then the Morse complex $C_*^{\text{Morse}}(f)$ is also naturally a filtered chain complex. In order to explain this, let us expose a few more details about

the Morse complex. Recall that we put $C_k^{\mathrm{Morse}}(f) = \mathbb{K}\langle \mathrm{crit}_k(f)\rangle$. The differential $\partial_k : C_k^{\mathrm{Morse}}(f) \to C_{k-1}^{\mathrm{Morse}}(f)$ is a $\mathbb{K}$-linear map whose value on a basis element $p_- \in \mathrm{crit}_k(f)$ takes the form

$$\partial_k(p_-) = \sum_{p_+ \in \mathrm{crit}_{k-1}(f)} \#\left(\mathcal{M}(p_-;p_+)/\mathbb{R}\right) p_+.$$

Here $\mathcal{M}(p_-;p_+)$ denotes the moduli space of gradient descent flow lines from p_- to p_+, i.e. the space of maps $u : \mathbb{R} \to M$ such that $\partial_s u = -\nabla_\mu f \circ u$ and $\lim_{s\to\pm\infty} u(s) = p_\pm$.

Recall that μ is a chosen Riemannian metric on M satisfies the Morse–Smale condition, defined as follows. Since $-\nabla_\mu f$ is a vector field on a closed manifold M, it has a time-t flow for all $t \in \mathbb{R}$, which we denote by $\phi_t \in \mathrm{Diff}(M)$.

Definition 2.6. *Given a critical point $p \in \mathrm{crit}(f)$, the* **stable manifold** *is*

$$W^s(p) := \{x \in M \mid \lim_{t\to\infty} \phi_t(x) = p\}.$$

Similarly, the **unstable manifold** *is*

$$W^u(p) := \{x \in M \mid \lim_{t\to-\infty} \phi_t(x) = p\}.$$

Roughly, $W^s(p)$ is the set of all points which "descend" to p, and $W^u(p)$ is the set of all points which 'ascend" to p. It turns out that $W^u(p)$ is diffeomorphic to an open disk of dimension $\mathrm{ind}_f(p)$, and $W^s(p)$ is diffeomorphic to an open disk of codimension $\mathrm{ind}_f(p)$, i.e. dimension $n - \mathrm{ind}_f(p)$.

Definition 2.7. *The pair (f,μ) is* **Morse–Smale** *if $W^s(p)$ and $W^u(q)$ intersect transversely for all $p, q \in \mathrm{crit}(f)$.*

Recall that two submanifolds $A, B \subset M$ are said to intersect transversely if for every $x \in A \cap B$ we have $T_xA + T_xB = T_xM$. In particular, this holds vacuously if $A \cap B = \varnothing$.

A consequence of the above definition is that for any $p_-, p_+ \in \mathrm{crit}(f)$ with $\mathrm{ind}(p_-) > \mathrm{ind}(p_+)$, $\mathcal{M}(p_-;p_+)$ is a smooth manifold of dimension $\mathrm{ind}(p_-) - \mathrm{ind}(p_+)$. Moreover, there is a free $\mathbb{R}$-action on $\mathcal{M}(p_-;p_+)$ given simply by translating. Namely, for $r \in \mathbb{R}$ we put $r \cdot u := u_r$, where $u_r(s) = u(s + r)$. Then the quotient manifold $\mathcal{M}(p_-;p_+)/\mathbb{R}$ is also a smooth manifold of dimension $\mathrm{ind}(p_-) - \mathrm{ind}(p_+) - 1$ (not necessarily closed or compact!).

It turns out that we can orient each of the moduli spaces $\mathcal{M}(p_-;p_+)/\mathbb{R}$, at least if M is oriented. In brief, these orientations are naturally induced by picking (arbitrarily) orientations on all of the stable manifolds of critical points of f, which in turn induce orientations on all of the unstable manifolds. In the case $\mathrm{ind}(p_-) = \mathrm{ind}(p_+) + 1$, $\mathcal{M}(p_-;p_+)/\mathbb{R}$ becomes an oriented 0-manifold, which means a collection of points, each with an associated sign $\pm$ attached to it. We then denote by $\#\mathcal{M}(p_-;p_+)$ the signed count of such points. Note that this count is only legitimate if the number of such points is finite. This turns out to be the case by a compactness theorem.

Now observe that $\#\mathcal{M}(p_-;p_+)/\mathbb{R}$ can only be nonzero if $f(p_-) \geq f(p_+)$. Indeed, f decreases along gradient descent trajectories. Therefore if we put

$$C_k^{\mathrm{Morse},<r}(f) := \mathbb{K}\langle p \in \mathrm{crit}_k(f) \mid f(p) < r\rangle,$$

then the Morse differential ∂_k maps $C_k^{\mathrm{Morse},<r}(f)$ to $C_{k-1}^{\mathrm{Morse},<r}(f)$. This makes $C_*^{\mathrm{Morse}}(f)$ into a filtered chain complex.

Remark 2.8. Various nontrivial analytic facts go into the proof that $\partial_{\mathrm{Morse}}$ squares to zero (and similarly for various other structural properties of Morse homology). In brief, the proof

proceeds by analyzing the moduli spaces $\mathcal{M}(p_-; p_+)/\mathbb{R}$ whenever $\mathrm{ind}(p_-) = \mathrm{ind}(p_+) + 2$. These are 1-dimensional manifolds which are not typically compact, but they admit natural compactifications in terms of "broken flow lines" going from p_- to an intermediate point q with $\mathrm{ind}(q) = \mathrm{ind}(p_-) - 1$, and then proceeding from q to p_+. Proving this requires a compactness theorem, which states that we really do get something compact after adding in these broken flow lines, and also gluing theorem, which states that this compactification really does have the structure of a manifold with boundary. Combining this with the fact that the (signed) count of boundary points of an (oriented) 1-dimensional manifold with boundary is zero translates into the algebraic relation $(\partial_{\mathrm{Morse}})^2 = 0$ (check this!). $\Diamond$

For two different Morse functions $f, g : M \to \mathbb{R}$, we know that the corresponding Morse homologies $H_*^{\mathrm{Morse}}(f)$ and $H_*^{\mathrm{Morse}}(g)$ are isomorphic, since these are both isomorphic to the singular homology of M. However, the filtered chain complexes $C_*^{\mathrm{Morse}}(f)$ and $C_*^{\mathrm{Morse}}(g)$ are *not* isomorphic. There are natural chain homotopy equivalences

$$C_*^{\mathrm{Morse}}(f) \underset{\psi}{\overset{\phi}{\rightleftarrows}} C_*^{\mathrm{Morse}}(g) \ ,$$

meaning that ϕ and ψ are chain maps and the compositions $\phi \circ \psi$ and $\psi \circ \phi$ are homotopic to the identity, but these do not preserve filtrations. More precisely, we have

$$\phi(C_*^{\mathrm{Morse},<r}(f)) \subset C_*^{\mathrm{Morse},<r+\delta}(g) \quad \text{and} \quad \psi(C_*^{\mathrm{Morse},<r}(g)) \subset C_*^{\mathrm{Morse},<r+\delta}(f),$$

where $\delta = ||f - g||$.

LECTURE 3

Persistence modules and barcodes

We begin with:

DEFINITION 3.1. *A* **persistence module** *(over a field $\mathbb{F}$) is a family of finite dimensional vector spaces $\{V_t\}_{t\in\mathbb{R}}$, along with linear maps $\pi_{s,t} : V_s \to V_t$ for all $s < t$ such that we have $\pi_{t,r} \circ \pi_{s,t} = \pi_{s,r}$ for all $s < t < r$.*

Example 3.2. Given a filtered chain complex $\{C_*^{<r}\}_{r\in\mathbb{R}}$, we get a persistence module by putting $V_t := H(C_*^{<r})$, and letting $\pi_{s,t} : H(C_*^{<s}) \to H(C_*^{<t})$ be the map induced by the inclusion $C_*^{<s} \subset C_*^{<t}$. Note that $\pi_{s,t}$ is not necessarily injective. ◊

Example 3.3. Given a Morse function $f : M \to \mathbb{R}$, we get a persistence module $V(f)$ with $V(f)_t := H_*^{\text{Morse}}(S_{<t})$, where $S_{<t} := \{x \in M \mid f(x) < t\}$. We also get persistence submodules for each $k = 0, \dots, \dim M$ by looking only at degree k homology classes. Note that we could also replace Morse homology with e.g. singular homology. ◊

DEFINITION 3.4. *Let us call a persistence module V* **finite type** *if*

- *for all but finitely many $t \in \mathbb{R}$, there is an open neighborhood U of t such that $\pi_{r,s}$ is an isomorphism for all $r, s \in U$ with $r < s$*
- *for some t_0, we have $V_t = \{0\}$ for all $t \le t_0$.*

DEFINITION 3.5. *A persistence module V is* **lower semi-continuous** *if for any $t \in \mathbb{R}$ there exists ϵ such that $\pi_{s,t}$ is an isomorphism for any $t - \epsilon < s < t$.*

The following example plays the role of a basic building block in the theory of persistence modules:

Example 3.6. For $a < b \le \infty$, we have a persistence module $\mathbb{F}(a, b]$, called an "interval module", where

$$\mathbb{F}(a,b]_t = \begin{cases} \mathbb{F} & t \in (a,b] \\ 0 & \text{otherwise,} \end{cases}$$

and we put

$$\pi_{s,t} = \begin{cases} \mathbb{1} & s,t \in (a,b],\ s < t \\ 0 & \text{otherwise.} \end{cases}$$

◊

Remark 3.7. Observe that the interval module $\mathbb{F}(a, b]$ is finite type and lower semi-continuous. However, if we were define e.g. $\mathbb{F}[a, b)$ analogously, the lower semocontinuity property would no longer hold. ◊

DEFINITION 3.8. *A* **morphism** *$F : V \to W$ between persistence modules V, W consists of a family of linear maps $F_t : V_t \to W_t$ for $t \in \mathbb{R}$ such that for each $s, t \in \mathbb{R}$ the following*

diagram commutes:

$$\begin{array}{ccc} V_s & \longrightarrow & V_t \\ {\scriptstyle F_s}\downarrow & & \downarrow{\scriptstyle F_t} \\ W_s & \longrightarrow & W_t. \end{array},$$

where the horizontal arrows are the structural maps $\pi_{s,t}$ for the persistence modules V and W.

Example 3.9. There is a natural morphism F of persistence modules $\mathbb{F}(1,2] \to \mathbb{F}(0,2]$ where $F_t = \mathbb{1}$ for $t \in (1,2]$, and $F_t = 0$ otherwise. On the other hand, any morphism from $\mathbb{F}(0,1] \to \mathbb{F}(0,2]$ is necessarily trivial (check this). ♢

In order to define the interleaving distaince d_{int}, we introduce a little bit more formalism.

Definition 3.10. *Given a persistence module V and $\delta \in \mathbb{R}$, let $V[\delta]$ denote its* δ**-shift**, *the persistence module with $V[\delta]_t = V_{t+\delta}$ and structural maps $\pi[\delta]_{s,t} = \pi_{s+\delta,t+\delta}$.*

Similarly, given a morphism $F : V \to W$, we have the δ**-shifted morphism** *$F[\delta] : V[\delta] \to W[\delta]$ defined by $F[\delta]_t = F_{t+\delta}$.*

Note that the structural maps $\pi_{t,t+\delta}$ always define a morphism from V to its shift $V[\delta]$:

Definition 3.11. *Given a persistence module V and $\delta \in \mathbb{R}$, the* **shift morphism** *$\text{Sh}_V^\delta : V \to V[\delta]$ is defined by $(\text{Sh}_V^\delta)_t = \pi_{t,t+\delta}$.*

Definition 3.12. *Given $\delta > 0$, two persistence modules V, W are* δ**-interleaved** *if there exist persistence morphisms $F : V \to W[\delta]$ and $G : W \to V[\delta]$ such that the following diagrams commute:* $V \xrightarrow{F} W[\delta] \xrightarrow{G[\delta]} V[2\delta]$ ($\text{Sh}_V^{2\delta}$) $\quad W \xrightarrow{G} V[\delta] \xrightarrow{F[\delta]} W[2\delta]$ ($\text{Sh}_W^{2\delta}$) .

There is a natural way of composing two morphisms $F : V \to W$ and $G : W \to Q$, and there is also a notion of identity morphism $\mathbb{1}_V$ from a persistence module V to itself. We will say that two persistence modules V and W are *isomorphic* if there exist morphisms $F : V \to W$ and $G : W \to V$ such that $G \circ F = \mathbb{1}_V$ and $F \circ G = \mathbb{1}_W$. One can view a δ-interleaving as an "isomorphism up to an error of δ".

Definition 3.13. *Given two persistence modules V, W, their* **interleaving distance** *is defined by*

$$d_{\text{int}}(V,W) := \inf\{\delta > 0 \mid V, W \text{ are } \delta\text{-interleaved}\}.$$

Exercise 3.14. *The interleaving distance is a* **pseudometric**, *i.e. it is symmetric and satisfies the triangle inequality. However, a priori $d_{\text{int}}(V,W)$ could be infinity.*

Exercise 3.15. *Given a finite type persistence module V, there exists $t_1 \in \mathbb{R}$ such that $\pi_{s,t}$ is an isomorphism for all $s,t \geq t_1$. Let V_∞ denote V_t for $t \geq t_1$ (or, more precisely, V_∞ is the direct limit of our direct system).*

For finite type persistence modules V, W, show that $d_{\text{int}}(V,W) < \infty$ if and only if $\dim V_\infty = \dim W_\infty$.

In the future, it will sometimes be convenient to use the language of multisets:

Definition 3.16. *A* **multiset** *is a set S together with a function $m : S \to \mathbb{Z}_{\geq 1}$. We view a multiset as a set except that each element can be repeated a finite number of times, and we view $m(x)$ as the "multiplicity" of the element $x \in S$.*

Equivalently, we can view a multiset as a set of pairs $(x, m_x) \in S \times \mathbb{Z}_{\geq 1}$, where each x appears only once.

THEOREM 3.17 (Normal form theorem for persistence modules). *Let V be a persistence module over a field $\mathbb{F}$ which is finite type and lower semi-continuous. Then there is a unique finite multiset of left open right closed intervals $\{(I_i, m_i)\}_{i=1}^N$ such that*

$$V \cong \mathbb{F}(I_1)^{\oplus m_1} \oplus \cdots \oplus \mathbb{F}(I_N)^{\oplus m_N}.$$

Here each I_i is of the form $(a_i, b_i]$ with $a_i < b_i \leq \infty$ and $m_i \in \mathbb{Z}_{\geq 1}$.

Let $\mathcal{B}(V) := \{(I_i, m_i)\}_{i=1}^N$ denote the multiset as in Theorem 3.17. We call this the *barcode* associated to the persistence module V.

We next define the bottleneck distance between barcodes. Given a barcode $\mathcal{B}$ and $\delta > 0$, let $\mathcal{B}_\delta$ denote the result after throwing away all bars of length $\leq \delta$. For an interval $I = (a, b]$, put $I^{-\delta} := (a - \delta, b + \delta]$.

DEFINITION 3.18. *Two barcodes $\mathcal{B}, \mathcal{B}'$ are* **δ-matched** *if, after throwing some bars of length $\leq \delta$ in $\mathcal{B}$ and $\mathcal{B}'$, there is a bijection between the remaining bars of $\mathcal{B}$ and $\mathcal{B}'$ such that such that the endpoints of corresponding intervals lie at distance $\leq \delta$ from each other.*

The last condition is equivalent to having $I \subset J^{-\delta}$ and $J \subset I^{-\delta}$ whenever I and J are paired under the bijection.

DEFINITION 3.19. *The* **bottleneck distance** *between two barcodes $\mathcal{B}, \mathcal{B}'$ is*

$$d_{\text{bot}}(\mathcal{B}, \mathcal{B}') := \inf\{\delta \mid \mathcal{B}, \mathcal{B}' \text{ are } \delta\text{-matched}\}.$$

EXERCISE 3.20. *The bottleneck distance d_{bot} satisfies the axioms of a metric (at least for barcodes at a finite distance from a given one), i.e. it is symmetric, satisfies the triangle inequality, and is nondegenerate.*

By Theorem 3.17, there is a bijective correspondence between persistence modules and barcodes. In fact, we have:

THEOREM 3.21. *The association $V \mapsto \mathcal{B}(V)$ is an isometry from the set of persistence modules equipped with the interleaving distance to the set of barcodes equipped with the bottleneck distance.*

In particular, d_{int} is nondegenerate.

LECTURE 4

The normal form theorem, I

4.1. Topological data analysis

We begin by introducing an important source of persistence modules, namely finite metric spaces. This plays a central role in so-called "topological data analysis" (see e.g. [Car]), a framework which has become very popular in recent years.

DEFINITION 4.1. *A (finite)* **abstract simplicial complex** *is a pair (S, Σ), where S is a finite set (the "vertices") and Σ is a set of subsets of S (the "simplices") which is closed under passing to further subsets, i.e. such that*

$$\sigma \in \Sigma \quad \textit{and} \quad \tau \subset \sigma \quad \Longrightarrow \quad \tau \in \Sigma.$$

The **geometrical realization** *of (S, Σ) is a topological space defined by*

$$|(S, \Sigma)| := \bigcup_{\sigma \in \Sigma} \mathrm{conv}\{e_{\phi(s)}\}_{s \in \sigma},$$

where $\phi : S \to \{1, \dots, |S|\}$ is a chosen bijection, and $e_1, \dots, e_{|S|}$ are the standard basis vectors for $\mathbb{R}^{|S|}$.

Remark 4.2. The geometric realization of (S, Σ) can be understood as taking for each $\sigma \in \Sigma$ a standard simplex dimension $|\sigma| - 1$, and then gluing these together along facets, with gluings determined by the subset relations in Σ. Recall that the standard k-simplex is

$$\Delta^k := \{(x_0, \dots, x_k) \in \mathbb{R}^{k+1} \mid x_0 + \cdots + x_k = 1,\ x_0, \dots, x_k \geq 0\}.$$

$\Diamond$

Example 4.3. In the example depicted in Figure 4.1, we have vertex set $S = \{a, b, c, d\}$ and simplices $\Sigma = \{\varnothing, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{a, c\}, \{b, c\}, \{c, d\}, \{a, b, c\}\}$. $\Diamond$

DEFINITION 4.4. *Let X be a finite metric space, with distance function $d : X \times X \to \mathbb{R}_{\geq 0}$. Given $\epsilon \in \mathbb{R}_{>0}$, the* **Vietoris–Rips** $\mathrm{VR}(X, \epsilon)$ *is the simplicial complex with vertex set X,*

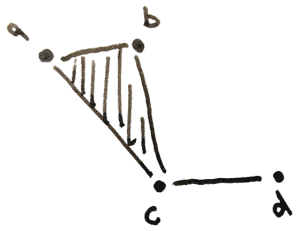

FIGURE 4.1. An example of simplicial complex with 4 vertices

FIGURE 4.2. A point cloud which resembles a circle.

such that $\sigma = \{x_0, \dots, x_k\} \subset X$ *is a* k*-simplex if and only if we have* $d(x_i, x_j) < \epsilon$ *for all* $1 \leq i, j \leq k$.

It is helpful to think of the family of Vietoris–Rips complexes $\mathrm{VR}(X, \epsilon)$ as ϵ ranges over all positive real numbers:

- for $\epsilon > 0$ sufficiently small, $\mathrm{VR}(X, \epsilon)$ is a discrete space, i.e. there are no k-simplices for $k \geq 1$
- for $\epsilon \gg 0$, $\mathrm{VR}(X, \epsilon)$ is equivalent to a $(|X| - 1)$-dimensional simplex $\Delta^{|X|-1}$, i.e. every subset of X corresponds to a simplex in $\mathrm{VR}(X, \epsilon)$
- $\mathrm{VR}(X, \epsilon) = \mathrm{VR}(X, \epsilon')$ unless there is a pair $x, x' \in X$ with $d(x, x') \in [\epsilon, \epsilon']$.

The basic premise of topological data analysis is as follows. Given some collection of data points, along with a natural way of measuring pairwise distances, we get a finite metric space X. For example, starting with a "point cloud", i.e. a finite subset of $\mathbb{R}^N$, we typically measure distances using the standard Euclidean metric. Now we imagine that there is some "true" topological space which underlies our dataset. For example, for the point cloud depicted in Figure 4.2, visual inspection suggests that our data comes from a space which is homeomorphic (or at least homotopy equivalent) to a circle.

Granted, since our dataset is finite and most likely involves some noise (i.e. small random errors), we cannot simply take e.g. the subspace topology on X, as that would just give the discrete topology. Rather, the simplicial complex $\mathrm{VR}(X, \epsilon)$ is one way of trying to cook up a topological space which hopefully approximates the "true" topology (note that there are also several common variations on the Vietoris–Rips complex which arise in practice applications - see e.g. **[Car]**). However, as the above discussion illustrates, the topology of $\mathrm{VR}(X, \epsilon)$ is quite sensitive to the choice of ϵ. In the example of Figure 4.2, it is evident that $\mathrm{VR}(X, \epsilon)$ will not be homotopy equivalent to a circle if we take ϵ too small or too large. In this example the best choice is to take ϵ to be roughly the distance between any two "consecutive" points. However, for a general dataset with many more points and of much higher dimension it is often fair from clear what a good choice of ϵ is.

Fortunately, we can avoid making any choice of $\epsilon > 0$ by considering all possibilities simultaneously and studying the resulting persistence module $\{H_*(\mathrm{VR}(X, t))\}_{t \in \mathbb{R}_{>0}}$ and its corresponding barcode $\mathcal{B}$. Long bars in $\mathcal{B}$ should correspond to homology classes which "persist" for a large range of ϵ, and hence are topologically significant for our dataset. By constrast, we will see various short bars which correspond to random noise and hence are essentially negligible. For the example in Figure 4.2, we expect to see two long bars, one in homological degree zero and one in homological degree one (corresponding to the homology groups of the circle), plus a large number of short bars in various degrees which we can safely ignore.

In brief, given a finite metric space X, viewed as a dataset, we have a barcode $\mathcal{B}$, which we expect to be a good proxy for topological (or at least homological) "features" of our dataset. In practical applications, we could e.g. feed the barcode $\mathcal{B}$ (or some function applied to it) into an auxiliary machine learning algorithm as part of our pipeline.

4.2. Several versions of the normal form theorem

Recall that a **partially ordered set** is a set S equipped with a binary relation "$\leq$" which is reflexive, anti-symmetric, and transitive. It is **totally ordered** if for any two elements x, y we have either $x \leq y$ or $y \leq x$.

DEFINITION 4.5. *Let $\mathcal{P}$ be a partially ordered set, viewed as a category with one object for each element x, with a unique morphism $x \to y$ whenever we have $x \leq y$. Given another category $\mathcal{C}$, a* **$\mathcal{P}$-persistence object** *in $\mathcal{C}$ is a functor from $\mathcal{P}$ to $\mathcal{C}$. Similarly, a* **morphism** *between two $\mathcal{P}$-persistence objects in $\mathcal{C}$ is a natural transformation between the corresponding functors.*

In particular, a $\mathbb{Z}_{\geq 0}$-persistence $\mathbb{F}$-vector space (over a field $\mathbb{F}$) consists of a collection of vector spaces V_i, $i \in \mathbb{Z}_{\geq 0}$, along with linear maps $V_i \to V_{i+1}$ for each $i \in \mathbb{Z}_{\geq 0}$. A $\mathbb{Z}_{\geq 0}$-persistence abelian group is the same thing as a nonnegatively graded module over the polynomial ring $\mathbb{Z}[t]$. Unfortunately there is apparently no nice classification theorem for graded modules over $\mathbb{Z}[t]$, but the situation becomes much more favorable if we work over a field and make a tameness assumption (akin to our finite type condition in the case of $\mathbb{R}$-persistence).

DEFINITION 4.6. *A $\mathbb{Z}_{\geq 0}$-persistent $\mathbb{F}$-vector space V is* **tame** *if each V_i is finite dimensional and the persistence map $V_i \to V_{i+1}$ is an isomorphism for i sufficiently large.*

Recall that finitely generated modules over a principle ideal domain (e.g. $\mathbb{Z}$ or $\mathbb{F}[t]$) uniquely decompose into a direct sum of free and cyclic modules. This is a generalization of the classification theorem for finitely generated abelian groups (see e.g. [**DF**]). In particular, a graded extension of this result states that for any nonnegatively graded finitely generated $\mathbb{F}[t]$-module V we have

$$V \cong \bigoplus_{i=1}^{M} \mathbb{F}[t][-a_i] \oplus \bigoplus_{j=1}^{N} \mathbb{F}[t]/(t^{l_j})[-b_j],$$

where for a module M we denote by $M[-a]$ the same module but with the grading shifted upward by a.

PROPOSITION 4.7. *A $\mathbb{Z}_{\geq 0}$-persistence $\mathbb{F}$-vector space is tame if and only if the corresponding $\mathbb{F}[t]$-module is finitely generated.*

Using Proposition 4.7, we get a corresponding normal form theorem for tame $\mathbb{Z}_{\geq 0}$-persistent $\mathbb{F}$-vector spaces. Under this translation, each free factor $\mathbb{F}[t][-a]$ corresponds to an interval module $\mathbb{F}[a, \infty)$, while each cyclic factor $\mathbb{F}[t]/(t^l)[-b]$ corresponds to an interval module $\mathbb{F}[b, b + l]$.

Let us also briefly mention a more general version of Theorem 3.17:

THEOREM 4.8 ([**CB**]). *Let R be a totally ordered set, and let V be an R-persistence $\mathbb{F}$-vector space such that V_r is finite-dimensional for each $r \in R$. Then V isomorphic to a direct sum of (possibly infinitely many) interval modules, and moreover this decomposition is unique up to reordering the factors.*

LECTURE 5

The normal form theorem II

In this lecture we prove Theorem 3.17. We first prove the uniqueness part, and then existence. We mostly follow the exposition in [**PRSZ**, §2.1].

5.1. Proof of the normal form theorem: uniqueness

We begin by proving the uniqueness part of Theorem 3.17. That is, assuming we have an isomorphism

$$V \cong \bigoplus_{i=1}^{N} \mathbb{F}(a_i, b_i]^{\oplus m_i}, \tag{5.1}$$

we need to show that the intervals $(a_i, b_i]$ and their multiplicities m_i are uniquely determined (up to reordered) by the isomorphism type of V.

Example 5.1. Consider the persistence module $V := \mathbb{F}(0,3] \oplus \mathbb{F}(1,2]$ versus $V' := \mathbb{F}(0,2] \oplus \mathbb{F}(1,3]$. Uniqueness implies that these cannot be isomorphic, since they have different barcodes. However, note that we have $\dim V_t = \dim V'_t$ for all $t \in \mathbb{R}$. To distinguish them, we need to look at the corresponding persistence maps $\pi_{s,t}$ and $\pi'_{s,t}$. For example, we can observe that $\pi_{1/2,5/2} : V_{1/2} \to V_{5/2}$ is nonzero, whereas $\pi_{1/2,5/2} : V'_{1/2} \to V'_{5/2}$ is necessarily trivial. It is easy to check that the ranks of these maps would have to coincide if V and V' were isomorphic. ◊

It will be convenient to introduce the following language:

Definition 5.2. *Given a persistence module V, a point $t \in \mathbb{R}$ is* **spectral** *if for any neighborhood U of t there exist $r < s$ in U such that $\pi_{r,s}$ is not an isomorphism. Let* $\mathrm{spec}(V)$ *denote the set of all spectral points of V, together with ∞.*

Let $c_1 < \cdots < c_M$ denote the endpoints (both left and right) of all finite intervals involved in our decomposition (5.1), and put $c_{M+1} := \infty$. Note that we have $\mathrm{spec}(V) = \{c_1, \ldots, c_{M+1}\}$. Moreover, if V and V' are isomorphic persistence modules then we have $\mathrm{spec}(V) = \mathrm{spec}(V')$.

For $r < s$, let $m_{r,s}$ denote the multiplicity of the interval $(r, s]$ in our decompostion (in particular $m_{r,s} = 0$ the interval $(r, s]$ does not appear). It suffices to show that $m_{r,s}$ is determined from the isomorphism type of V. Since $m_{r,s} = 0$ unless $r, s \in \mathrm{spec}(V)$, it suffices to extract the numbers m_{c_i,c_j} for all $1 \le i < j \le \infty$.

Let $b_{r,s}$ denote the rank of the map $\pi_{r,s}$. In the spirit of Example 5.1, this depends only on the isomorphism type of V. At the same time, in terms of the decomposition (5.1), we have

$$b_{r,s} = \sum_{\substack{c_i \le r \\ c_j \ge s}} m_{c_i,c_j}. \tag{5.2}$$

Here the sum is over all $1 \leq i, j \leq M+1$ such that $c_i < r$ and $c_j \geq s$, and (5.2) follows from the observation that each interval module $\mathbb{F}(a,b]$ contributes 1 to $m_{r,s}$ if $r > a$ and $s \leq b$, and it contributes 0 otherwise. For $1 \leq i < j \leq M+1$ we have

$$b_{c_i,c_j} = \sum_{\substack{\alpha<i \\ \beta\geq j}} m_{c_\alpha,c_\beta} = \sum_{\substack{\alpha\leq i-1 \\ \beta\geq j}} m_{c_\alpha,c_\beta}.$$

We should be able to determine the multiplicity of the interval $(c_i, c_j]$ by measuring how b_{c_i,c_j} changes as we increment or decrement i and j. More formally, we have:

$$b_{c_{i+1},c_j} - b_{c_i,c_j} = \sum_{\substack{\alpha\leq i \\ \beta\geq j}} m_{c_\alpha,c_\beta} - \sum_{\substack{\alpha\leq i-1 \\ \beta\geq j}} m_{c_\alpha,c_\beta} \tag{5.3}$$

$$= \sum_{\substack{\alpha=i \\ \beta\geq j}} m_{c_\alpha,c_\beta}, \tag{5.4}$$

and similarly

$$b_{c_{i+1},c_{j+1}} - b_{c_i,c_{j+1}} = \sum_{\substack{\alpha=i \\ \beta\geq j+1}} m_{c_\alpha,c_\beta}. \tag{5.5}$$

Subtracting (5.5) from (5.4), we have

$$\left(b_{c_{i+1},c_j} - b_{c_i,c_j}\right) - \left(b_{c_{i+1},c_{j+1}} - b_{c_i,c_{j+1}}\right) = \sum_{\substack{\alpha=i \\ \beta=j}} m_{c_\alpha,c_\beta} = m_{c_i,c_j}. \tag{5.6}$$

5.2. Proof of the normal form theorem: existence

The following definition will be convenient for our proof:

Definition 5.3. *Given a persistence module V, a persistence submodule $W \subset V$ is* **semi-surjective** *if there exists $t_0 \in \mathbb{R}$ such that:*

- *$W_t = V_t$ for $t \leq t_0$*
- *$\pi_{r,s} : W_r \to W_s$ is surjective for any $t_0 < r < s$.*

The key lemma is:

Lemma 5.4. *Let V be a persistence module which is finite type and lower semi-continuous, and let $W \subsetneq V$ be a proper persistence submodule which is semi-surjective. Then there exists another semi-surjective persistence submodule $W_\# \subset V$ which is isomorphic to the direct sum $W \oplus \mathbb{F}(a,b]$ for some $a < b \leq \infty$.*

The existence part of Theorem 3.17 now follows immediately from Lemma 5.4 by induction, starting with the trivial submodule $\{0\} \subset V$. It therefore remains to prove the lemma.

Let $W \subset V$ be as in Lemma 5.4. As before, put $\mathrm{spec}(V) = \{c_1, \ldots, c_M, c_{M+1} = \infty\}$. Note that we have $\mathrm{spec}(W) \subset \mathrm{spec}(V)$, and t_0 as in Definition 5.3 must lie in $\mathrm{spec}(V)$. For some $\gamma \in \{1, \ldots, M\}$ we have $W_{c_i} = V_{c_i}$ for $i = 1, \ldots, \gamma - 1$ and $W_{c_\gamma} \subsetneq V_{c_\gamma}$. Pick any $z_{c_\gamma} \in V_{c_\gamma} \setminus W_{c_\gamma}$. For $k > \gamma$, put

$$z_{c_k} := \pi_{c_\gamma,c_k}(z_{c_\gamma}) \in V_{c_k}.$$

There are two possibilities to consider:

(1) We have $z_{c_k} \notin W_{c_k}$ for $k = \gamma, \gamma+1, \ldots, M+1$.

(2) For some $k > \gamma$ we have $z_{c_k} \in W_{c_k}$.

Case (1) will correspond to the interval $(a, b]$ in Lemma 5.4 being infinite (i.e. $b = \infty$), while (2) will correspond to a finite interval. Let us focus on (2), case (1) being similar.

For $k > \gamma$, put $z_{c_k} := \pi_{c_\gamma, c_k}(z_{c_\gamma}) \in V_{c_k}$. Let δ be the smallest $k > \gamma$ for which $z_{c_k} \in W_{c_k}$. By semisurjectivity, we can find $x_{c_\gamma} \in W_{c_\gamma}$ such that $\pi_{c_\gamma, c_\delta}(x_{c_\gamma}) = z_{c_\delta}$. Put $y_{c_\gamma} := z_{c_\gamma} - x_{c_\gamma}$ and $y_{c_k} := \pi_{c_\gamma, c_k}(y_{c_\gamma})$ for $k > \gamma$. Note that we have $y_{c_k} \in V_{c_k} \setminus W_{c_k}$ for $k = \gamma, \dots, \delta - 1$ and $y_{c_\delta} = 0$.

Now let us extend the definition of $y_t \in V_t$ to all $t \in \mathbb{R}$ as follows. Firstly, for $t \leq c_{\gamma-1}$ or $t > c_{\delta-1}$ we put $y_t := 0$. Recall that y_{c_k} is already defined for $k = \gamma, \gamma + 1, \dots$. For $t \in (y_{c_k}, y_{c_{k+1}})$ for some k, we put $y_t := \pi^{-1}_{t, c_{k+1}}(y_{c_{k+1}})$. The fact that $\pi_{t, c_{k+1}}$ is invertible comes from the fact that there are no spectral points in $[t, c_{k+1})$, together with the lower semicontinuity property.

EXERCISE 5.5. *Putting $P_t := \mathbb{F}\langle y_t \rangle$ for $t \in \mathbb{R}$, this defines a persistence submodule $P \subset V$ such that:*

- *P is isomorphic to $\mathbb{F}(c_{\gamma-1}, c_{\delta-1}]$*
- *$W_t \cap P_t = \{0\}$ for all t, and together W and P span a persistence submodule $W_\# := W + P$ which isomorphic to the direct sum $W \oplus \mathbb{F}(c_{\gamma-1}, c_{\delta-1}]$*
- *$W_\# \subset V$ is semi-surjective.*

This completes the proof of Lemma 5.4.

Bibliography

[Car] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society* **46**(2009), 255–308. 4.1, 4.1

[CB] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *Journal of Algebra and its Applications* **14**(2015), 1550066. 4.8

[DF] David S Dummit and Richard M Foote. *Abstract algebra*. Prentice Hall Englewood Cliffs, NJ, 1991. 4.2

[PRSZ] Leonid Polterovich, Daniel Rosen, Karina Samvelyan, and Jun Zhang. Topological persistence in geometry and analysis. *arXiv:1904.04044* (2019). 5