Quantitative aspects of Morse and Floer theory
Math 635, USC Spring 2022

Kyler Siegel

Contents

Introduction

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

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

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 $\mathrm{ind}_f(p)$.*

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 $\mathrm{ind}_f(p) = 0$ if and only if p is a local minimum, $\mathrm{ind}_f(p) = n$ 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 $a > \max f$
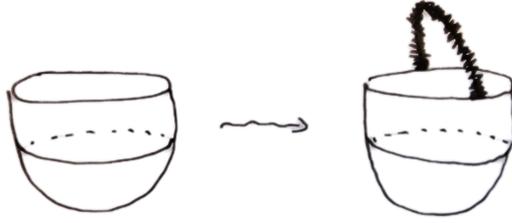
FIGURE 1.1. Attaching a 2-dimensional 1-handle.
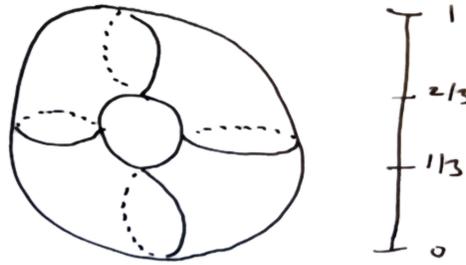
FIGURE 1.2. The height function $\mathbb{T}^2 \to \mathbb{R}$.

- $S_{\leq a} \subset S_{\leq a'}$ for $a \leq a'$.

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.*

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

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

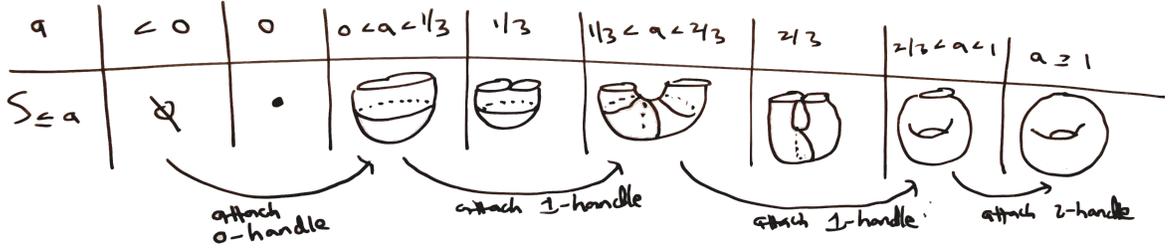$$C_i^{\mathrm{Morse}}(f; \mathbb{K}) := \mathbb{K}\langle \mathrm{crit}_i(f) \rangle$$

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:

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

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 $p_+ \in C_k^{\mathrm{Morse}}(f; \mathbb{K})$ and $p_- \in C_{k-1}^{\mathrm{Morse}}(f; \mathbb{K})$, we consider maps $u : \mathbb{R} \to M$ satisfying

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

Then we have $\partial_k(p_-) = \mathrm{sign}(u)p_+ + \cdots$, 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, $\varepsilon(u) \in \{1, -1\}$ (which we can ignore if say $\mathbb{K} = \mathbb{Z}/2$). We then put

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

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

As a first consequence, observe that we must have

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

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:

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

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 $\mathrm{crit}_0 = \{p_0\}$, $\mathrm{crit}_1 = \{p_1\}$, $\mathrm{crit}_2 = \{p_2, p_3\}$.*
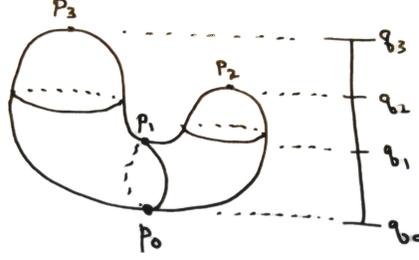
FIGURE 1.4. A Morse function $S^2 \to \mathbb{R}$ with four critical points.

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*

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

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

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

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

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

which computes the Morse cohomology 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 $H_{\mathrm{Morse}}^i(-f) \cong H_{n-i}^{\mathrm{Morse}}(f)$, 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* $C_*^{\mathrm{Morse}}(f)$
   (2) to this filtered chain complex $C_*^{\mathrm{Morse}}(f)$ 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* $d_{\mathrm{int}}(V, W)$
   (3) for two barcodes $\mathcal{B}, \mathcal{B}'$, we have the *bottleneck distance* $d_{\mathrm{bot}}(\mathcal{B}, \mathcal{B}')$.

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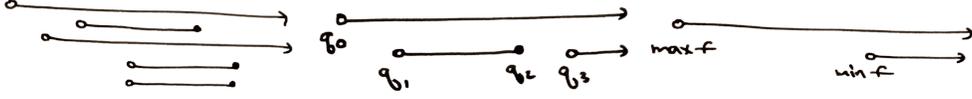
FIGURE 2.1. Left: an example of a barcode, i.e. a multiset of intervals. Note that some are finite and some are infinite, and some intervals may be repeated. Center: the barcode of the Morse function $h : S^2 \to$ depicted in Figure 1.4. Right: the barcode of a Morse function $f : S^2 \to \mathbb{R}$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

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

$\square$

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 $\bigcup\limits_{r \in \mathbb{R}} C^{<r} = C$.*

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*

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

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 $C_*^{\mathrm{Morse}}(f)$ 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 $\lim\limits_{s \to \pm\infty} u(s) = p_\pm$.

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 $T_x A + T_x B = T_x M$. 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!).*

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

$$C_*^{\mathrm{Morse}}(f) \xrightarrow[\psi]{\phi} C_*^{\mathrm{Morse}}(g) \ ,$$

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 $V(f)_t := H_*^{\mathrm{Morse}}(S_{<t})$, 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 $t - \varepsilon < s < t$.*

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*

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

and we put

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

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:

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

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).*

More generally, this is a nontrivial morphism from $\mathbb{F}(a,b]$ to $\mathbb{F}(c,d]$ if and only if $c \leq a < d \leq b$.

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** $\mathrm{Sh}_V^\delta : V \to V[\delta]$ is defined by $(\mathrm{Sh}_V^\delta)_t = \pi_{t,t+\delta}$.*

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:* $V \underset{\mathrm{Sh}_V^{2\delta}}{\xrightarrow{\hspace{0.3cm} F \hspace{0.3cm}} W[\delta] \xrightarrow{\hspace{0.3cm} G[\delta] \hspace{0.3cm}} V[2\delta]} \quad W \underset{\mathrm{Sh}_W^{2\delta}}{\xrightarrow{\hspace{0.3cm} G \hspace{0.3cm}} V[\delta] \xrightarrow{\hspace{0.3cm} F[\delta] \hspace{0.3cm}} W[2\delta]}$.

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*

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

EXERCISE 3.14. *The interleaving distance is a **pseudometric**, i.e. it is symmetric and satisfies the triangle inequality. However, a priori $d_{\mathrm{int}}(V, W)$ 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 $d_{\mathrm{int}}(V, W) < \infty$ 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.

We will also sometimes write e.g. $\{1, 1, \pi, \pi, \pi\}$ to denote the multiset with two copies of 1 and three copies of π.

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 $\leq 2\delta$ 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.

REMARK 3.19. *Recall that $\mathcal{B}$ and $\mathcal{B}'$ are defined to be multisets of intervals. Strictly speaking, the bijection involved in Definition 3.18 is a bijection between sub-mulitsets of $\mathcal{B}$ and $\mathcal{B}'$. This means that if an interval is repeated some number of times in $\mathcal{B}$ or $\mathcal{B}'$, then some number (possibly zero) of its copies are involved in the bijection, and they need not all pair with the same interval on the other side.*

EXAMPLE 3.20. *Put $\mathcal{B} = \{(0, 3], (0, 3], (1, 2]\}$ and $\mathcal{B}' = \{(-1/2, 7/2], (1/2, 5/2]\}$. Then $\mathcal{B}$ and $\mathcal{B}'$ are $1/2$-matched, where one possible bijection is*

$$(0, 3] \longleftrightarrow (-1/2, 7/2]$$
$$(0, 3] \longleftrightarrow (1/2, 5/2]$$
$$(1, 2] \longleftrightarrow \varnothing,$$

where the last line means that we "throw out" the interval $(1, 2]$, i.e. it is not involved in the bijection.

REMARK 3.21. *Note that in the definition of a δ-matching, we can throw away any number of intervals of length $\leq 2\delta$ in $\mathcal{B}$ and $\mathcal{B}'$, but we are also allowed to have some of them participate in the bijection. We will sometimes refer to the intervals of length $\leq 2\delta$ as "short intervals". For example, $\mathcal{B} = \{(-.9, .9]\}$ and $\mathcal{B}' = \{(-1.5, 1.5]\}$ are 1-matched, with the short interval $(-.9, .9]$ necessarily participating in the bijection.*

DEFINITION 3.22. *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 3.23. *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.*

REMARK 3.24. *One manifestation of the above exercise is that we can compose a δ-matching μ between $\mathcal{B}$ and $\mathcal{B}'$ and a δ'-matching μ' between $\mathcal{B}$ and $\mathcal{B}''$ to get a $(\delta+\delta')$-matching between $\mathcal{B}$ and $\mathcal{B}''$. Namely, we "compose whenever possible", i.e. if $I \in \mathcal{B}$ is paired under μ with $J \in \mathcal{B}'$ and J is paired under μ' with $K \in \mathcal{B}''$, then I is paired under $\mu' \circ \mu$ with K.*

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

THEOREM 3.25 ("Isometry Theorem"). *The association $V \mapsto \mathcal{B}(V)$ is an isometry from the set of persistence modules e'q'uipped 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*

$$\sigma \in \Sigma \quad and \quad \tau \subset \sigma \quad \implies \quad \tau \in \Sigma.$$

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 $\phi : S \to \{1, \ldots, |S|\}$ *is a chosen bijection, and* $e_1, \ldots, e_{|S|}$ *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*

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

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

DEFINITION 4.4. *Let* X *be a finite metric space, with distance function* $d : X \times X \to \mathbb{R}_{\geq 0}$*. Given* $\varepsilon \in \mathbb{R}_{>0}$*, the* **Vietoris–Rips** $\mathrm{VR}(X, \varepsilon)$ *is the simplicial complex with vertex set* X*, such that* $\sigma = \{x_0, \ldots, x_k\} \subset X$ *is a* k*-simplex if and only if we have* $d(x_i, x_j) < \varepsilon$ *for all* $1 \leq i, j \leq k$*.*
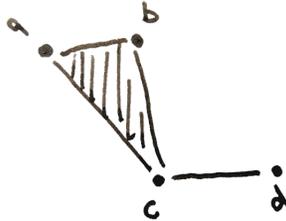

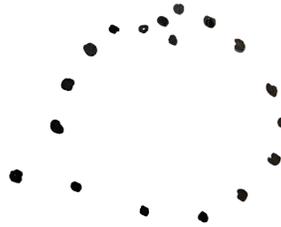FIGURE 4.1. An example of simplicial complex with 4 vertices

FIGURE 4.2. A point cloud which resembles a circle.

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

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

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 $\mathrm{VR}(X, \varepsilon)$ 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 $\mathrm{VR}(X, \varepsilon)$ is quite sensitive to the choice of ε. In the example of Figure 4.2, it is evident that $\mathrm{VR}(X, \varepsilon)$ 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 $\varepsilon > 0$ 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

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

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 $1 \leq i < j \leq \infty$.

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

$$(5.2) \qquad b_{r,s} = \sum_{\substack{c_i \leq r \\ c_j \geq s}} m_{c_i, c_j}.$$

Here the sum is over all $1 \le i, j \le M + 1$ such that $c_i < r$ and $c_j \ge s$, 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 $s \le b$, and it contributes 0 otherwise. For $1 \le i < j \le M + 1$ we have

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

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:

$$(5.3) \qquad b_{c_{i+1}, c_j} - b_{c_i, c_j} = \sum_{\substack{\alpha \le i \\ \beta \ge j}} m_{c_\alpha, c_\beta} - \sum_{\substack{\alpha \le i-1 \\ \beta \ge j}} m_{c_\alpha, c_\beta}$$

$$(5.4) \qquad = \sum_{\substack{\alpha = i \\ \beta \ge j}} m_{c_\alpha, c_\beta},$$

and similarly

$$(5.5) \qquad b_{c_{i+1}, c_{j+1}} - b_{c_i, c_{j+1}} = \sum_{\substack{\alpha = i \\ \beta \ge j+1}} m_{c_\alpha, c_\beta}.$$

Subtracting (5.5) from (5.4), we have

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

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 $t \le t_0$*
- $\pi_{r,s} : W_r \to W_s$ *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 $a < b \le \infty$.*

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 $\operatorname{spec}(V) = \{c_1, \ldots, c_M, c_{M+1} = \infty\}$. Note that we have $\operatorname{spec}(W) \subset \operatorname{spec}(V)$, and t_0 as in Definition 5.3 must lie in $\operatorname{spec}(V)$. 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 $k = \gamma, \ldots, \delta - 1$ 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 $k = \gamma, \gamma + 1, \ldots$. For $t \in (y_{c_k}, y_{c_{k+1}})$ for some k, we put $y_t := \pi_{t, c_{k+1}}^{-1}(y_{c_{k+1}})$. 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.

Computing persistent homology

Let us now try to get a better feel for persistent homology by computing it in a concrete example. We borrow the running example from [**ZC**], and we recommend that article for a much more detailed discussion and additional tricks to speed up the computation.

Let us start by considering the simplicial complex represented in Figure 6.1. Being a filled square, its (ordinary) homology over $\mathbb{Z}$ is trivial except for a copy of $\mathbb{Z}$ in degree 0. However, if we wish to compute its homology directly from the direction, one way is to represent each of the boundary maps ∂_i by a matrix M_i, and then compute the Smith normal forms of these matrices.

DEFINITION 6.1. *Given an $m \times n$ matrix M with coefficients in a principal ideal domain R. Its* **Smith normal form** *is an $m \times n$ matrix D of the form $D = AMB$, where*
- *A is an invertible $m \times m$ matrix*
- *B is an invertible $n \times n$ matrix*
- *the only nonzero entries of D are $D_{1,1}, \ldots, D_{k,k}$ for some $k \leq \min(m,n)$, where $D_{i,i}$ divides $D_{i+1,i+1}$ for $i = 1, \ldots, k-1$.*

The entries $D_{1,1}, \ldots, D_{k,k}$, called the **elementary divisors** of M, are uniquely determined up to multiplication by units.

In particular, $\mathbb{Z}$ is a principal ideal domain. It turns out that one way to compute the homology of a chain complex over $\mathbb{Z}$ is to compute the Smith normal (or more specifically the elementary divisors) of the matrix representatives of the differentials. Assume that each C_i is free abelian group, and let r_i denote its rank. Suppose that $\partial_i : C_i \to C_{i-1}$ has matrix representative M_i with respect to the chosen bases of C_i and C_{i-1}. Suppose that M_i has elementary divisors $\lambda_{i,1}, \ldots, \lambda_{i,k_i}$. Then it follows that
- $\ker \partial_i$ has rank $r_i - k_i$
- $\operatorname{im} \partial_i$ has rank k_i
- H_i is isomorphic to $\mathbb{Z}/\lambda_{i+1,1} \oplus \cdots \oplus \mathbb{Z}/\lambda_{i+1,k_{i+1}} \oplus \mathbb{Z}^{r_i - k_i - k_{i+1}}$.

Given a matrix M, we can apply elementary column operators (this corresponds to changing the basis of the domain) and elementary row operations (this corresponds to
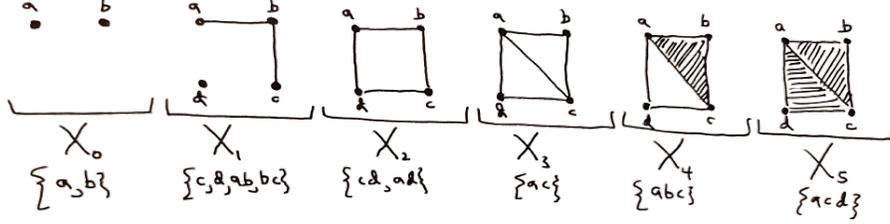

FIGURE 6.1. A simplicial complex.

FIGURE 6.2. A filtered simplicial complex.

changing the basis of the codomain) until M in Smith normal form, $D = AMB$, where A and B correspond to the change of basis matrices.

In the example represented by Figure 6.1, ∂_1 has matrix representative

$$
\begin{array}{c c}
 & \begin{array}{ccccc} ab & bc & cd & ad & ac \end{array} \\
\begin{array}{c} a \\ b \\ c \\ d \end{array} &
\left(\begin{array}{ccccc}
-1 & 0 & 0 & -1 & -1 \\
1 & -1 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 1 \\
0 & 0 & 1 & 1 & 0
\end{array}\right)
\end{array}
$$

and after changing bases this becomes

$$
\begin{array}{c c}
 & \begin{array}{ccccc} cd & bc & ab & z_1 & z_2 \end{array} \\
\begin{array}{c} d-c \\ c-b \\ b-a \\ a \end{array} &
\left(\begin{array}{ccccc}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{array}\right)
\end{array}
$$

with $z_1 := ad - bc - cd - ab$ and $z_2 := ac - bc - ab$.

Now consider the (discretely) filtered simplicial complex $X_0 \subset \cdots \subset X_5$ represented by Figure 6.2. Computing the persistent homology over $\mathbb{Z}/2$ is equivalent to viewing it as a module over $\mathbb{Z}/2[t]$ and computing its ordinary homology. In turn, we can achieve this by computing the Smith normal form of the matrix representatives of the differentials, now viewed as matrices with coefficients in $\mathbb{Z}/2[t]$.

For example, ∂_1 can be represented as a matrix over $(\mathbb{Z}/2)[t]$ by:

$$
\partial_1 =
\begin{array}{c c}
 & \begin{array}{ccccc} ab & bc & cd & ad & ac \end{array} \\
\begin{array}{c} a \\ b \\ c \\ d \end{array} &
\left(\begin{array}{ccccc}
t & 0 & 0 & t^2 & t^3 \\
t & t & 0 & 0 & 0 \\
0 & 1 & t & 0 & t^2 \\
0 & 0 & t & t & 0
\end{array}\right)
\end{array}.
$$

For example the 1-simplex ac first appears in X_3, whereas the 0-simplex c first appears in X_1. Therefore the entry in the column ac and row c is t^2, represented a filtration difference of two. Note that only the *difference* in filtration is recorded by the t powers of the matrix entries. Recall that we are actually computing the homology of a *graded* $(\mathbb{Z}/2)[t]$-module, and the actual filtration level of e.g. ac is taken into account by this extra grading data.

Now, after elementary row and column operations, we get the Smith normal form:

$$
\partial_1 = \begin{array}{c} \\ c+tb \\ c+d \\ b+a \\ a \end{array}
\begin{array}{ccccc} bc & cd & ab & z_1 & z_2 \\ \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & t & 0 & 0 & 0 \\ 0 & 0 & t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right) \end{array}
$$

where $z_1 = ad + t \cdot bc + cd + t \cdot ab$ and $z_2 = ac + t^2 \cdot ab + t^2 \cdot bc$. Note that all of these basis elements are homogeneous.

We can now read off the H_0 barcode. In general, a bar with left endpoint $k \in \mathbb{Z}$ and right endpoint $l \in \mathbb{Z}_{\geq k}$ corresponds a cycle σ first appearing in X_k, and a chain τ with $\partial \tau = \sigma$ first appearing in X_l. Since $\partial_0 = 0$, each of the basis elements $c + tb, c + d, b + a, a$ is a cycle.

- Since a is not a boundary, it corresponds a bar with left endpoint at 0 and no right end point.
- Corresponding to $\partial_1(bc) = c + tb$ we get a bar with both left and right endpoints at 1.
- Corresponding to $\partial(cd) = t(c + d)$, we get a bar with left endpoint at 1 and right endpoint at 2.
- Corresponding to $\partial(ab) = t(b + a)$, we get a bar with left endpoint at 1 and right endpoint at 0.

Now consider $\partial_2 : C_2 \to C_1$. After performing elementary row operations which exactly mirror the elementary column operations we performed for ∂_1, we can write it in terms of the basis bc, cd, ab, z_1, z_2 for C_1. Moreover, it is easy to check that the elements $c + tb, c + d, b + a$ cannot appear in the image of ∂_2, as we must have $\partial_1 \circ \partial_2 = 0$. We have

$$
\partial_2 = \begin{array}{c} \\ z_2 \\ z_1 \end{array}
\begin{array}{cc} abc & acd \\ \left(\begin{array}{cc} t & t^2 \\ 0 & t^3 \end{array} \right), \end{array}
$$

and after changing bases we arrive at the Smith normal form:

$$
\partial_2 = \begin{array}{c} \\ z_2 \\ z_1 \end{array}
\begin{array}{cc} abc & acd+t\cdot abc \\ \left(\begin{array}{cc} t & 0 \\ 0 & t^3 \end{array} \right). \end{array}
$$

It follows that there are two H_1 bars, one with left endpoint 3 and right endpoint 4, and another one with left endpoint 2 and right endpoint 5. Note that there are no bars corresponding to H_2, since there are no 2-cycles. Also, the unique bar with no right endpoint corresponds to the fact that the homology of X is 1-dimensional in degree 0 and trivial otherwise.

EXERCISE 6.2. *Check some of the computations in this lecture.*

LECTURE 7

The isometry theorem

Our goal for this lecture is to discuss the proof of Theorem 3.25, which states that for any pair of persistence modules V, W we have $d_{\mathrm{int}}(V, W) = d_{\mathrm{bot}}(\mathcal{B}(V), \mathcal{B}(W))$. For concreteness, we will be assuming as before that V and W are persistence modules over a field $\mathbb{F}$ which are finite type and lower semicontinuous. We will mostly follow the exposition of [**PRSZ**, §3], which in turn is based on [**BL**]. For the sake of brevity we will give only a sketch, referring the reader to the aforementioned references for more details.

7.1. Bottleneck distance bounds interleaving distance

We first establish the bound $d_{\mathrm{int}}(V, W) \leq d_{\mathrm{bot}}(\mathcal{B}(V), \mathcal{B}(W))$. Assume that there is a δ-matching between $\mathcal{B}(V)$ and $\mathcal{B}(W)$. Our goal is to find a δ-interleaving between V and W.

By the definition of a δ-matching (see §3), there is a bijection between some of the intervals of $\mathcal{B}(V)$ and some of the intervals of $\mathcal{B}(W)$, such that:

- all of the intervals of length $> 2\delta$ are involved in this bijection
- if an interval I in $\mathcal{B}(V)$ is paired with an interval J in $\mathcal{B}(W)$, then the left endpoint of I lies at distance $\leq \delta$ from the left endpoint of J, and similarly for the right endpoints.

EXERCISE 7.1. *Show that for $V = \mathbb{F}(a, b]$, we have*

$$
\mathrm{Sh}_V^\delta = \begin{cases} \mathbb{1} & if \ \ t \in (a, b - \delta] \\ 0 & otherwise. \end{cases}
$$

In particular, Sh_V^δ is the zero morphism if and only if $b - a \leq \delta$.

Suppose that $I = (a, b] \in \mathcal{B}(V)$ is matched with $J = (c, d] \in \mathcal{B}(W)$. In particular, we must have $|a - c| \leq \delta$ and $|b - d| \leq \delta$. We have $\mathbb{F}(J)[\delta] = (c - \delta, d - \delta]$. Note that since $c - \delta \leq a$ and $d - \delta \leq b$, the endpoints of $\mathbb{F}(J)[\delta]$ lie to the left of the corresponding endpoints of $\mathbb{F}(I)$.

Let us define a (possibly trivial) morphism $F_I : \mathbb{F}(I) \to \mathbb{F}(J)[\delta]$ by

$$
(F_I)_t = \begin{cases} \mathbb{1} & if \ \ t \in (a, b] \cap (c - \delta, d - \delta] \\ 0 & otherwise. \end{cases}
$$

Similarly, define a morphism $G_J : \mathbb{F}(J) \to F(I)[\delta]$ by

$$
(G_J)_t = \begin{cases} \mathbb{1} & if \ \ t \in (c, d] \cap (a - \delta, b - \delta] \\ 0 & otherwise. \end{cases}
$$

Then the composition $G_J[\delta] \circ F_I$ is

$$(G_J \circ F_I)_t = \begin{cases} \mathbb{1} & \text{if} \quad t \in (a,b] \cap (c-\delta, d-\delta] \cap (a-2\delta, b-2\delta] \\ 0 & \text{otherwise.} \end{cases}$$

Note that we have $(c-\delta, d-\delta] \supset (a,b] \cap (a-2\delta, b-2\delta] = (a, b-2\delta]$, and hence $G_J[\delta] \circ F_I$ is precisely $\mathrm{Sh}^{2\delta}_{\mathbb{F}(I)} : \mathbb{F}(I) \to \mathbb{F}(I)[2\delta]$ (recall Definition 3.11). Similarly, we have $F_I[\delta] \circ G_J = \mathrm{Sh}^{2\delta}_{\mathbb{F}(J)}$.

We now define morphisms $F : V \to W[\delta]$ and $G : W \to V[\delta]$ as follows. Recall that by the normal form theorem, V and W are direct sums of interval modules, with one copy for each bar of the corresponding barcode. For an unmatched interval $I \in \mathcal{B}(V)$, we put $F|_I = 0$. Note that if I is an unmatched interval then it necessarily has length $\leq 2\delta$, and hence $\mathrm{Sh}^{2\delta}_{\mathbb{F}(I)}$ is the zero morphism. If $I \in \mathcal{B}(V)$ is matched with $J \in \mathcal{B}(W)$, we define $F|_I$ to have image in $F(J)[\delta]$, and to be given by $F_I : F(I) \to F(J)[\delta]$ as above. We define the morphism G similarly, by putting together G_J as J ranges over matched intervals in $\mathcal{B}(W)$, and taking G to be zero on the remaining summands. By the above discussion, we have $G[\delta] \circ F = \mathrm{Sh}^{2\delta}_V$ and $F[\delta] \circ G = \mathrm{Sh}^{2\delta}_W$, and hence V and W are δ-interleaved.

7.2. Interleaving distance bounds bottleneck distance

Now suppose that V and W are δ-interleaved. We seek to find a δ-matching of $\mathcal{B}(V)$ and $\mathcal{B}(W)$.

Let us begin with a simple special case as a warmup:

EXAMPLE 7.2. *Given interval modules $V = \mathbb{F}(I)$ and $W = \mathbb{F}(J)$ which are δ-interleaved, we seek to prove that $\mathcal{B}(V)$ and $\mathcal{B}(W)$ are δ-matched. Put $I = (a,b]$ and $J = (c,d]$. If both of the intervals I and J have length $\leq 2\delta$, then their barcodes are immediately δ-matched. Otherwise, assume without loss of generality that $\mathrm{len}(I) > 2\delta$. Then $\mathrm{Sh}^{2\delta}_V$ is not the zero morphism, and hence we have nontrivial morphisms $F : V \to W[\delta]$ and $G : W \to V[\delta]$. In particular, we must have $c - \delta \leq a < d - \delta \leq b$ and $a - \delta \leq c < b - \delta \leq d$ (c.f. Example 3.9). Since $|a - c| \leq \delta$ and $|b - d| \leq d$, there is an obvious δ-matching of $\mathcal{B}(V)$ and $\mathcal{B}(W)$.*

In general, if $V \to W$ is any morphism, we will define an "induced matching" of $\mathcal{B}(V)$ and $\mathcal{B}(W)$, which we denote by $\mu(F)$. This association is not generally functorial, i.e. for composable morphisms F and G, $\mu(G \circ F)$ is not equal to the composed matching $\mu(G) \circ \mu(G)$. Nevertheless, we will have:

LEMMA 7.3. $F : V \to W[\delta]$ *and* $G : W \to V[\delta]$ *define a δ-interleaving, then $\mu(F)$ will be a δ-matching between $\mathcal{B}(V)$ and $\mathcal{B}(W)$.*

We will satisfy ourselves with defining the induced matching $\mu(F)$, leaving the details of Lemma 7.3 to the references.

Firstly, suppose that $F : V \to W$ is injective, in the sense that each $F_t : V_t \to W_t$ is injective. For each $b \in \mathbb{R} \cup \{\infty\}$, let

$$(a_1, b] \supset \cdots \supset (a_k, b]$$

be the bars in $\mathcal{B}(V)$ having right endpoint b, in order of nonincreasing length. Similarly, let

$$(a'_1, b] \supset \cdots \supset (a'_k, b]$$

be the bars in $\mathcal{B}(W)$ with right endpoint b.

EXERCISE 7.4. *We have $k \leq l$.*

DEFINITION 7.5. *If $F : V \to W$ is an injective morphism, we define a matching $\mu_{\mathrm{inj}}(F)$ between $\mathcal{B}(V)$ and $\mathcal{B}(W)$ as follows. For each $b \in \mathbb{R} \cup \{\infty\}$, we match*

$$(a_i, b] \longleftrightarrow (a_i', b] \quad for \quad i = 1, \ldots, k$$
$$\varnothing \longleftrightarrow (a_i', b] \quad for \quad i = k+1, \ldots, l.$$

Now suppose that $F : V \to W$ is a surjective morphism. For $a \in \mathbb{R} \cup \{\infty\}$, let

$$(a, b_1] \supset \cdots \supset (a, b_k]$$

denote the bars of $\mathcal{B}(V)$ with left endpoint a, in order of nondecreasing length, and similarly let

$$(a, b_1'] \supset \cdots \supset (a, b_l']$$

denote the bars of $\mathcal{B}(W)$ with left endpoint a.

EXERCISE 7.6. *We have $k \geq l$.*

DEFINITION 7.7. *If $F : V \to W$ is a surjective morphism, we define a matching $\mu_{\mathrm{sur}}(F)$ between $\mathcal{B}(V)$ and $\mathcal{B}(W)$ as follows. For each $a \in \mathbb{R} \cup \{\infty\}$, we match*

$$(a, b_i] \longleftrightarrow (a, b_i'] \quad for \quad i = 1, \ldots, l$$
$$(a, b_i] \longleftrightarrow \varnothing \quad for \quad i = l+1, \ldots, k.$$

We now define $\mu(F)$ for any morphism F by a simple trick:

DEFINITION 7.8. *Given any morphism $F : V \to W$, we write it as the composition*

$$V \xrightarrow{F_{\mathrm{sur}}} \mathrm{im}\,(F) \xrightarrow{F_{\mathrm{inj}}} W,$$

where F_{sur} is surjective and is given by simply restricting the codomain to $\mathrm{im}\,(F)$ (or more precisely, doing this for each F_t), and F_{inj} is injective is given by the natural inclusion. We then put

$$\mu(F) := \mu_{\mathrm{inj}}(F_{\mathrm{inj}}) \circ \mu_{\mathrm{sur}}(F_{\mathrm{sur}}).$$

EXAMPLE 7.9. *Put $V := \mathbb{F}(1,3] \oplus \mathbb{F}(1,2]$ and $W := \mathbb{F}(3,4] \oplus \mathbb{F}(0,2]$. Let $F : V \to W$ be the morphism which is trivial on the factor $\mathbb{F}(1,3]$, and restricts to the other factor as the "identity whenever possible" morphism $\mathbb{F}(1,2] \to \mathbb{F}(0,2]$. Then $\mu_{\mathrm{sur}}(F_{\mathrm{sur}})$ is*

$$(1,3] \longleftrightarrow (1,2]$$
$$(1,2] \longleftrightarrow \varnothing,$$

and $\mu_{\mathrm{inj}}(F_{\mathrm{inj}})$ is

$$(1,2] \longleftrightarrow (0,2]$$
$$\varnothing \longleftrightarrow (3,4],$$

and hence $\mu(F)$ is

$$(1,3] \longleftrightarrow (0,2]$$
$$(1,2] \longleftrightarrow \varnothing$$
$$\varnothing \longleftrightarrow (3,4].$$

7.3. Boundary depth and its generalizations

The boundary depth played an important role in the proof of Proposition 2.2. Recall that we stated during that proof that it is 2-Lipshitz as a function on the space of barcodes. We are now ready to prove (a generalization of) this fact.

DEFINITION 7.10. *Given a barcode $\mathcal{B}$, let $\beta_k(\mathcal{B})$ be the largest of the kth longest finite bar, or 0 if there are fewer than k finite bars.*

So given a barcode, we get real numbers $\beta_1(\mathcal{B}) \geq \beta_2(\mathcal{B}) \geq \dots$.

PROPOSITION 7.11. *For any pair of barcodes $\mathcal{B}, \mathcal{B}'$ and any $k \in \mathbb{Z}_{\geq 1}$ we have*

$$|\beta_k(\mathcal{B}) - \beta_k(\mathcal{B}')| \leq 2d_{\mathrm{bot}}(\mathcal{B}, \mathcal{B}').$$

The proof of Proposition 7.11 will involve the following purely combinatorial lemma (we refer the reader to [**PRSZ**, §4.1] for a proof).

LEMMA 7.12 ("Matching Lemma"). *Given $b_1 \leq \dots \leq b_N$ and $c_1 \leq \dots \leq c_N$ real numbers, we have*

$$\min_{\sigma \in S_N} \max_{i \in \{1,\dots,N\}} |b_i - c_{\sigma(i)}| = \max_{i \in \{1,\dots,N\}} |b_i - c_i|.$$

Here S_N is the permutation group on N elements. In order words, if we view $\sigma \in S_N$ as a "matching" between the b_i and c_i, and we define the "cost" of such a matching by $\max_{i \in \{1,\dots,N\}}$, then the optimal matching is given simply by the naive one corresponding to the identity permutation.

PROOF OF PROPOSITION 7.11. Suppose that we have a δ matching μ between $\mathcal{B}$ and $\mathcal{B}'$. Let $I_1, \dots, I_N$ be the matched bars of $\mathcal{B}$, ordered such that $\mathrm{len}(I_1) \geq \dots \geq \mathrm{len}(I_N)$. Similarly, let $J_1, \dots, J_N$ be the matched bars of $\mathcal{B}'$, ordered such that $\mathrm{len}(J_1) \geq \dots \geq \mathrm{len}(J_N)$. The matching μ corresponds to a permutation $\tau \in S_N$ such that we match $I_i \longleftrightarrow J_{\tau(i)}$ for $i = 1, \dots, N$. In particular, we have

$$|\mathrm{len}(I_i) - \mathrm{len}(J_{\tau(i)})| \leq 2\delta$$

for $i = 1, \dots, N$. We will prove

(7.1) $$\beta_k(\mathcal{B}) - \beta_k(\mathcal{B}') \leq 2\delta,$$

with the analogous inequality with $\mathcal{B}$ and $\mathcal{B}'$ swapped then following by symmetry.

We can assume $\beta_k(\mathcal{B}) > 2\delta$, since otherwise (7.1) is immediate. Then the k largest bars of $\mathcal{B}$ all have length $> 2\delta$, and hence are all matched under μ, meaning that $I_1, \dots, I_k$ are the k longest bars of $\mathcal{B}$, whence $\beta_k(\mathcal{B}) = \mathrm{len}(I_k)$. Applying Lemma 7.12, we have

$$\begin{aligned}
\mathrm{len}(I_k) - \mathrm{len}(J_k) &\leq \max_{i \in \{1,\dots,N\}} |\mathrm{len}(I_i) - \mathrm{len}(J_i)| \\
&= \min_{\sigma \in S_N} \max_{i \in \{1,\dots,N\}} |\mathrm{len}(I_i) - \mathrm{len}(J_{\sigma(i)})| \\
&\leq \max_{i \in \{1,\dots,N\}} |\mathrm{len}(I_i) - \mathrm{len}(J_{\tau(i)})| \\
&\leq 2\delta.
\end{aligned}$$

Note that there could be some bars of $\mathcal{B}$ which are longer than J_k but do not participate in the matching, if $\mathrm{len}(J_k) \leq 2\delta$. At any rate, J_k is the k'-longest bar of $\mathcal{B}'$ for some $k' \geq k$, and we have $\mathrm{len}(J_k) \leq \beta_k(\mathcal{B}')$, and hence

$$\beta_k(\mathcal{B}) = \mathrm{len}(I_k) \leq \mathrm{len}(J_k) + 2\delta \leq \beta_k(\mathcal{B}') + 2\delta.$$

$\square$

LECTURE 8

The Hamiltonian diffeomorphism group

Having discussed the abstract theory of persistence modules in the last few lectures, we now introduce now additional geometry. In this lecture we discuss the group of symplectomorphisms of a symplectic manifold, and its notable subgroup of Hamiltonian diffeomorphisms.

8.1. Symplectic manifold basics

Some familiarity with symplectic geometry is helpful but we will attempt to give a self-contained treatment, starting with some basic definitions.

DEFINITION 8.1. *A **symplectic manifold** is a smooth manifold M equipped with a differential two-form $\omega \in \Omega^2(M)$ which is closed and nondegenerate. The latter means that for each $p \in M$, the map $T_pM \to T_p^*M$ sending v to $\omega(v, -)$ is an isomorphism.*

In particular, M must be even-dimensional. As usual, we will often abuse notation and denote our symplectic manifold simply by M when the symplectic two-form ω is implicit.

EXAMPLE 8.2. *The standard symplectic form on $\mathbb{R}^{2n}$ with coordinates $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ is given by $\omega_{\mathrm{std}} := \sum_{i=1}^n dx_i \wedge dy_i$.*

DEFINITION 8.3. *Let (M, ω) be a symplectic manifold, and let $H : M \to \mathbb{R}$ be a smooth function. Since ω is nondegenerate, there is a unique vector field $\nabla_\omega(H)$ defined by*

$$\omega(\nabla_\omega(H), -) = dH.$$

*This is the **symplectic gradient** of H. It is also called the **Hamiltonian vector field associated to** H, denoted by X_H.*

REMARK 8.4. *Borrowing terminology from physics, we will also refer to H as a "Hamiltonian", but from the mathematical point of view it is simply a smooth function.*

As the name suggests, $\nabla_\omega(H)$ is defined just like the usual gradient of a function, except using a symplectic form instead of a Riemannian metric. However, its behavior is very different. Whereas the Riemannian gradient of H points in the direction of maximal increase, the symplectic gradient is tangent to the level sets of H. More precisely, we have

LEMMA 8.5. *For any $H : M \to \mathbb{R}$, $\nabla_\omega(H)$, we have $dH(X_H) = 0$.*

PROOF. By definition, $dH(X_H) = \omega(X_H, X_H)$, and this vanishes because ω is a differential two-form and hence skew-symmetric. □

8.2. Hamiltonian flows and symplectomorphisms

If M is closed, then by flowing along the vector field X_H, we get for each $t \in \mathbb{R}$ a diffeomorphism $Fl_H^t \in \mathrm{Diff}(M)$. Finding the flow is equivalent to solving the ODE

$$\dot\gamma(t) = X_H(\gamma(t)), \quad \gamma(0) = p$$

30

for each initial condition $p \in M$, which leads to $Fl_H^t(p) = \gamma(t)$. Note that finding the flow of a vector field on a manifold which is noncompact can be problematic, as integral curves could run off to infinity in finite time, but this is easily ruled out e.g. if we assume that the vector field has compact support (a similar discussion applies to the case that M has boundary).

EXAMPLE 8.6 (classical mechanics). *Consider a particle with one degree of freedom, e.g. a bead moving along a wire. Let $\mathbb{R}^2$ with coordinates (q, p) be its phase space, where q describes the position of the bead along the wire and p is the corresponding momentum coordinate. Imagine also that there is some force on the bead, e.g. gravity, and we suppose the force is conservative, meaning that it is given by the negative gradient of a potential energy function $V(q)$. The Hamiltonian $H : \mathbb{R}^2 \to \mathbb{R}$ is then the total energy of the system, i.e. potential energy plus kinetic energy:*

$$H(q, p) = V(q) + \tfrac{1}{2}p^2.$$

One can easily check that the corresponding Hamiltonian vector field is

$$X_H = p\partial_q - (\partial_q V)\partial_p.$$

Hamilton's formulation of classical mechanics states that the time evolution of the particle is given by flowing along the Hamiltonian vector field X_H. In the present example this amounts to the ODE

$$\dot{q}(t) = p(t)$$
$$\dot{p}(t) = -\partial_q V(q).$$

This system is easily seen to be equivalent to Newton's second law $\ddot{q} = -\partial_q V$.

Another basic result about Hamiltonian flows is that they preserve the symplectic form.

DEFINITION 8.7. *Given a symplectic manifold (M, ω), let*

$$\mathrm{Symp}(M, \omega) := \{\Phi \in \mathrm{Diff}(M) \mid \Phi^*\omega = \omega\}$$

denote the set of diffeomorphisms which preserve the symplectic two-form. This is easily seen to be a subgroup of $\mathrm{Diff}(M)$, and its elements are called **symplectomorphisms**.

LEMMA 8.8. *Given a Hamiltonian $H : M \to \mathbb{R}$, we have $Fl_H^t \in \mathrm{Symp}(M, \omega)$ for each t.*

PROOF. It suffices to show that the Lie derivative of ω in the direction of X_H is zero, and for this we can apply Cartan's formula:

$$\mathcal{L}_{X_H}(\omega) = d(\omega(X_H, -)) + (d\omega)(X_H, -, -) = d(dH) = 0.$$

$\square$

8.3. Hamiltonian diffeomorphism group

We will also want to consider flows of **time-dependent Hamiltonians**, i.e. smooth functions $H : M \times I \to \mathbb{R}$. Given such an H, we denote by $H_t : M \to \mathbb{R}$ the function $H(-, t)$. We will sometimes view H as a family of ordinary functions $\{H_t\}_{t \in [0,1]}$, $H_t \in C^\infty(M)$. Note that we get a corresponding time-dependent family of Hamiltonian vector fields $\{X_{H_t}\}_{t \in [0,1]}$. We will denote the corresponding flow (now given by solving a time-dependent ODE) by Fl_H^t or $Fl_{\{H_t\}}^t$. We will say that $H : M \times I \to \mathbb{R}$ is **autonomous** if it is independent of t. In the future, will usually assume by default that our Hamiltonians are time-dependent, unless stated otherwise. Note that for an autonomous Hamiltonian we have $Fl_H^s \circ Fl_H^t = Fl_H^{s+t}$, but this need not hold for a general time-dependent Hamiltonian.

We are now ready to define the Hamiltonian diffeomorphism group:

DEFINITION 8.9. *Let (M, ω) a closed symplectic manifold, we put*

$$\mathrm{Ham}(M, \omega) := \left\{ Fl_H^1 \mid H : M \times I \to \mathbb{R} \right\}.$$

It is not so important that we take the time 1 flow:

EXERCISE 8.10. *For any $\tau \in [0, 1]$ and $F : M \times I \to \mathbb{R}$, we have $Fl_F^\tau = Fl_G^1$, where $G : M \times I \to \mathbb{R}$ is given by $G_t = \tau F_{\tau t}$.*

Unlike $\mathrm{Diff}(M)$ or $\mathrm{Symp}(M, \omega)$, it is not immediately obvious that $\mathrm{Ham}(M, \omega)$ is a group. This follows from the following proposition, which in fact shows that $\mathrm{Ham}(M, \omega)$ is a *normal* subgroup of $\mathrm{Symp}(M, \omega)$.

PROPOSITION 8.11.
  (1) *For $\phi, \psi \in \mathrm{Ham}(M, \omega)$, we have $\phi \circ \psi \in \mathrm{Ham}(M, \omega)$. In fact, if $\phi_t = Fl_{\{F_t\}}^t$ and $\psi_t = Fl_{\{G_t\}}^t$, then we have $\phi_t \circ \psi_t = Fl_{\{H_t\}}^t$, where $H_t := F_t + G_t \circ \phi_t^{-1}$.*
  (2) *For $\phi \in \mathrm{Ham}(M, \omega)$, we have $\phi^{-1} \in \mathrm{Ham}(M, \omega)$. In fact, if $\phi_t = Fl_{\{F_t\}}^t$, then we have $\phi_t^{-1} = Fl_{\{H_t\}}^t$, where $H_t := -F_t \circ \phi_t$.*
  (3) *For $\phi \in \mathrm{Ham}(M\omega)$ and $\theta \in \mathrm{Symp}(M, \omega)$, we have $\theta \circ \phi \circ \theta^{-1} \in \mathrm{Ham}(M, \omega)$. In fact, if $\phi_t = Fl_{\{F_t\}}^t$, then we have $\theta \circ \phi_t \circ \theta^{-1} = Fl_{\{H_t\}}^t$, where $H_t := F_t \circ \theta^{-1}$.*

Before proving the above proposition, we prove a simple lemma which will be convenient to invoke later on:

LEMMA 8.12. *Let M be a symplectic manifold, let $K : M \to \mathbb{R}$ be a smooth function, and let $Q : M \to M$ be a symplectomorphism. Then we have $X_{K \circ Q^{-1}} = Q_*(X_K)$.*

PROOF. Recall that, by definition of the pushforward of vector fields, we have

$$(Q_* X_K)|_p = dQ|_{Q^{-1}(p)}(X_K|_{Q^{-1}(p)}) \in T_p M$$

for each $p \in M$. Here $X_K|_p \in T_p M$ denotes the value of the vector field X_K at p; we also denote this sometimes by $X_K(p)$ when no confusion should arise from notational overload. We also denote by $dQ|_{Q^{-1}(p)} : T_{Q^{-1}(p)} M \to T_p M$ the derivative of Q at the point $Q^{-1}(p)$, and so on. We will use write e.g. $dQ(v) = dQ|_p(v)$ for $v \in T_p M$.

By nondegeneracy of the symplectic form ω, it suffices to show that for any $v \in T_p M$ we have

$$\omega(X_{K \circ Q^{-1}}|_p, v) = \omega(Q_*(X_K)|_p, v).$$

Since Q is a symplectomorphism, we have $\omega(dQ|_p(w), dQ|_p(v)) = \omega(v, w)$ for any $v, w \in T_p M$. Therefore, we have

$$\begin{aligned}
\omega(Q_*(X_K)|_p, v) &= \omega(dQ|_{Q^{-1}(p)}(X_K|_{Q^{-1}(p)}), v) \\
&= \omega(X_K|_{Q^{-1}(p)}, dQ|_{Q^{-1}(p)}^{-1} v) \\
&= dK(dQ^{-1}(v)) \\
&= d(K \circ Q^{-1})(v) \\
&= \omega(X_{K \circ Q^{-1}}|_p, v),
\end{aligned}$$

as desired. $\square$

We can also summarize the above lemma by writing

(8.1) $$dQ(X_K(p)) = X_{K \circ Q^{-1}}(Q(p)) \quad \text{for all} \quad p \in M.$$

PROOF OF PROPOSITION 8.11. For (1), put $\theta_t := \phi_t \circ \psi_t$. For $p \in M$ we have

$$\partial_t(\theta_t(p)) = \partial_t(\phi_t \circ \psi_t(p))$$
$$= X_{F_t}(\phi_t \circ \psi_t(p)) + d\phi_t(X_{G_t}(\psi(p)))$$
$$= X_{F_t}(\phi_t \circ \psi_t(p)) + X_{G_t \circ \phi_t^{-1}}(\phi_t \circ \psi_t(p))$$
$$= X_{F_t + G_t \circ \phi_t^{-1}}(\phi_t \circ \psi_t(p))$$
$$= X_{H_t}(\theta_t(p))$$

for $H_t = F_t + G_t \circ \phi_t^{-1}$, and therefore we have $\theta_t = Fl^t_{\{H_t\}}$.

For (2), putting $\phi_t = Fl^t_{\{F_t\}}$ and $\psi_t = Fl^t_{\{-F_t \circ \phi_t\}}$, observe that by (1) we have

$$\phi_t \circ \psi_t = Fl^t_{\{F_t - F_t \circ \phi_t \circ \phi_t^{-1}\}} = \mathbb{1}.$$

For (3), observe that using (8.1) we have

$$\partial_t(\theta \circ \phi_t \circ \theta^{-1}(p)) = d\theta(X_{\{F_t\}}(\phi_t(\theta^{-1}(p))))$$
$$= X_{\{F_t \circ \theta^{-1}\}}(\theta \circ \phi_t \circ \theta^{-1}(p)),$$

and hence $\theta \circ \phi_t \circ \theta^{-1} = Fl^t_{\{F_t \circ \theta^{-1}\}}$. $\qquad\qquad\square$

The following theorem of Banyaga (which we will simply state without proof) resolves two natural questions about $\mathrm{Ham}(M, \omega)$. Firstly, given a path $\gamma : I \to \mathrm{Ham}(M, \omega)$, by definition each $\gamma(t)$ is the time-1 flow of some time-dependent Hamiltonian on M, but it is not immediately clear that the vector field $\partial_t \gamma$ is itself the Hamiltonian vector field of some function $M \to \mathbb{R}$. Secondly, it is natural to ask wonder whether $\mathrm{Ham}(M, \omega)$ is in a sense the "smallest" geometrically meaningful subgroup of $\mathrm{Symp}(M, \omega)$, or whether $\mathrm{Ham}(M, \omega)$ itself has nontrivial normal subgroups.

THEOREM 8.13. *Let (M, ω) be a closed symplectic manifold.*

(1) *If $\{\gamma_t\}_{t \in [0,1]}$ is any smooth path in $\mathrm{Ham}(M, \omega)$ with $\gamma_0 = \mathbb{1}$, then we have $\gamma_t = Fl^t_H$ for some $H : M \times I \to \mathbb{R}$.*

(2) *$\mathrm{Ham}(M, \omega)$ is a simple group, i.e. it has no nontrivial proper normal subgroups.*

8.4. The C^∞ topology

As we have seen, $\mathrm{Ham}(M, \omega)$ is a group. It is also naturally a topological space. Indeed, given closed smooth manifolds M and N, there is a natural topology on the set $C^\infty(M, N)$ of smooth maps $F : M \to N$. In particular, this puts a topology on $C^\infty(M, M)$, along with its subspaces $\mathrm{Diff}(M), \mathrm{Symp}(M, \omega), \mathrm{Ham}(M, \omega)$, and so on.

For the reader's convenience, let us briefly recall the definition of the topology on $C^\infty(M, N)$. Roughly speaking, two functions are considered "close" in this topology if all their partial derivatives are "close".

DEFINITION 8.14. *Let M and N be smooth manifolds, and assume that M is compact. A* **subbasic neighborhood** *is a subset $Q \subset C^\infty(M, N)$ of the following form. Suppose that we have open subsets $U \subset M$ and $V \subset N$, and local coordinate charts $\phi : U \to \mathbb{R}^{\dim M}$ and $\psi : V \to \mathbb{R}^{\dim N}$. Fix*

- *$r \in \mathbb{Z}_{\geq 0}$*
- *$\varepsilon \in \mathbb{R}_{>0}$*
- *a compact subset $K \subset U$*

- a smooth map $F_0 : M \to N$ such that $F_0(K) \subset V$.

Let Q be the set of all smooth maps $F : M \to N$ such that

- $F(K) \subset V$
- for each $p \in \phi(K)$, all of the partial derivatives of $\psi \circ F \circ \phi^{-1}$ at p of order at most r differ from those of $\psi \circ F_0 \circ \phi^{-1}$ by at most ε.

The topology on $C^\infty(M, N)$ is the smallest one having subbasic neighborhoods as open sets, i.e. open sets in $C^\infty(M, N)$ are unions of finite intersections of subbasic neighborhoods.

REMARK 8.15. *If M is noncompact we can still define the above topology on $C^\infty(M, N)$, but since it only controls convergence on compact subsets of M it is called the "weak topology", and there is also a finer "strong topology" (see [**Hir**, §2.1] for more details).*

In particular, using the above topology we can define:

DEFINITION 8.16. *We denote by $\mathrm{Symp}_0(M, \omega) \subset \mathrm{Symp}(M, \omega)$ the connected component of the identity map. Note that this is also a subgroup.*

Similarly, we denote by $\mathrm{Diff}_0(M) \subset \mathrm{Diff}(M)$ the connected component of $\mathbb{1}$. Note that $\mathrm{Ham}(M, \omega)$ is already connected, thanks to Exercise 8.10.

8.5. Normalized Hamiltonians and symplectic isotopies

By Theorem 8.13(1), the derivative of a path of Hamiltonian diffeomorphisms is a Hamiltonian vector field. Note that the natural linear map $H \mapsto X_H$ from smooth functions on M to Hamiltonian vector fields on M is by definition surjective, and its kernel consists of locally constant functions on M. In particular, adding a constant to a function does not change its Hamiltonian vector field. It will often be natural to add in an extra constraint to remove this ambiguity.

DEFINITION 8.17. *Given a closed symplectic manifold (M^{2n}, ω), the* **mean value** *of a Hamiltonian $H : M \to \mathbb{R}$ is*

$$\langle H \rangle := \frac{\int_M H \omega^{\wedge n}}{\int_M \omega^{\wedge n}}.$$

DEFINITION 8.18. *If (M, ω) is a closed symplectic manifold, an autonomous Hamiltonian $H : M \to I$ is* **normalized** *if $\langle H \rangle = 0$. Similarly, a time-dependent Hamiltonian $H : M \times I \to \mathbb{R}$ is normalized if $\langle H_t \rangle = 0$ for each $t \in I$.*

PROPOSITION 8.19. *If (M, ω) is a closed symplectic manifold with $H^1(M; \mathbb{R}) = 0$, then we have $\mathrm{Ham}(M, \omega) = \mathrm{Symp}_0(M, \omega)$.*

PROOF. Given $\phi \in \mathrm{Symp}_0(M, \omega)$, we can find a family of symplectomorphisms $\{\phi_t\}_{t \in [0,1]}$ with $\phi_0 = \mathbb{1}$ and $\phi_1 = \phi$. Let $V_t = \partial_t \phi_t$ be the corresponding family of vector fields, i.e. for each $p \in M$ we have $V_t|_{\phi_t(p)} = \partial_t(\phi_t(p))$. Since $\phi_t^* \omega = \omega$ for all t, we have $\mathcal{L}_{V_t}(\omega) = 0$, where $\mathcal{L}_{V_t}(\omega)$ denotes the Lie derivative of the two-form ω_t in the direction of the vector field V_t. Using Cartan's formula for the Lie derivative, we have

$$0 = \mathcal{L}_{V_t}(\omega) = d\iota_{V_t}\omega + \iota_{V_t}d\omega = d\iota_{V_t}\omega.$$

This shows that $\iota_{V_t}\omega$ is a closed one-form on M, and therefore it is exact since $H^1(M; \mathbb{R}) = 0$. There is then, for each $t \in I$, a unique normalized Hamiltonian $H_t : M \to \mathbb{R}$ such that $V_t = X_{H_t}$. We therefore have $\phi_t = Fl^t_{\{H_t\}}$, and in particular $\phi = \phi_1 \in \mathrm{Ham}(M, \omega)$.   □

EXERCISE 8.20. *Verify in the above proof that $H : M \times I \to \mathbb{R}$ is smooth.*

A vector field V on M such that $\iota_V\omega$ is closed is called a **symplectic vector field**.

8.6. Universal covers and flux

For some purposes, we will need to consider not just $\mathrm{Ham}(M,\omega)$ or $\mathrm{Symp}(M,\omega)$ but also their universal covers. As usual, one can realize the universal cover $\widetilde{S}$ of a topological space S as the space of all paths which start at a fixed basepoint, modulo homotopies rel endpoints. The projection map $\widetilde{S} \to S$ then corresponds to the endpoint of the path.

DEFINITION 8.21. *Let* $\widetilde{\mathrm{Symp}}_0(M,\omega)$ *denote the space of all paths* $\{\phi_t\}_{t\in[0,1]}$ *of symplectomorphisms such that* $\phi_0 = \mathbb{1}$, *modulo homotopies between paths fixed at* $t = 0$ *and* $t = 1$.

Similarly, $\widetilde{\mathrm{Ham}}(M,\omega)$ *is the space of all paths* $\{\phi_t\}_{t\in[0,1]}$ *of Hamiltonian diffeomorphisms such that* $\phi_0 = \mathbb{1}$, *modulo homotopies rel endpoints.*

We will often write e.g. $\{\phi_t\} \in \widetilde{\mathrm{Symp}}_0(M,\omega)$ or $\{\phi_t\} \in \widetilde{\mathrm{Ham}}(M,\omega)$ with the implicit understanding that $\{\phi_t\}$ is defined up to homotopy rel endpoints. We will sometimes write $[\{\phi_t\}]$ if we wish to emphasize that we are considering an equivalence class of paths. Sometimes for brevity we will also denote an element of $\widetilde{\mathrm{Ham}}(M,\omega)$ simply by ϕ, when no confusion should arise.

Now consider some $[\{\phi_t\}_{t\in[0,1]}] \in \widetilde{\mathrm{Symp}}_0(M,\omega)$. If $H^1(M;\mathbb{R}) \neq 0$, then Proposition 8.19 does not apply, but we can instead introduce a measure of how far the path is from being Hamiltonian:

DEFINITION 8.22. *We define the* **flux** *of a path* $\{\phi_t\}_{t\in[0,1]}$ *of symplectomorphisms by*

$$\mathrm{Flux}(\{\phi_t\}) := \int_0^1 [\iota_{V_t}\omega]dt \in H^1(M;\mathbb{R}),$$

where $V_t = \partial_t\phi_t$ *is the associated symplectic vector field and* $[\iota_{V_t}\omega] \in H^1(M;\mathbb{R})$ *is the corresponding de Rham cohomology class.*

It turns out that the flux in unaffected by homotopies rel endpoints, and therefore it actually descends to a homomorphism

$$\mathrm{Flux} : \widetilde{\mathrm{Symp}}_0(M,\omega) \to H^1(M;\mathbb{R}).$$

Note that $\mathrm{Flux}(\{\phi_t\}) = 0$ if $\{\phi_t\} \in \widetilde{\mathrm{Ham}}(M,\omega)$, since in that case the corresponding vector field $V_t = \partial_t\phi_t$ is exact for $t \in [0,1]$. In fact, for $\{\phi_t\} \in \widetilde{\mathrm{Symp}}_0(M,\omega)$, one can show that $\mathrm{Flux}(\{\phi_t\}) = 0$ if and only if $\{\phi_t\}$ is homotopic rel endpoints to a path in $\mathrm{Ham}(M,\omega)$.

It is easy to see that $\mathrm{Flux} : \widetilde{\mathrm{Symp}}_0(M,\omega) \to H^1(M;\mathbb{R})$ is surjective, since any class in $H^1(M;\mathbb{R})$ corresponds to a symplectic vector field on M and its flow generates an element of $\widetilde{\mathrm{Symp}}_0(M,\omega)$. By the above discussion, the kernel of Flux is identified with $\widetilde{\mathrm{Ham}}(M,\omega)$. It is natural to ask whether Flux descends to a map defined on $\mathrm{Symp}_0(M,\omega)$ itself. The obstruction comes from the flux of loops in $\mathrm{Symp}_0(M,\omega)$ based at $\mathbb{1}$.

DEFINITION 8.23. *Let* $\Gamma_\omega \subset H^1(M;\mathbb{R})$ *denote the set of all fluxes of paths* $\{\phi_t\} \in \mathrm{Symp}_0(M,\omega)$ *with* $\phi_0 = \phi_1 = \mathbb{1}$.

Then we get a well-defined homomorphism $\mathrm{Flux} : \mathrm{Symp}_0(M,\omega) \to H^1(M;\mathbb{R})/\Gamma_\omega$. The following theorem of Ono resolved an old question known as the "flux conjecture":

THEOREM 8.24 ([**Ono**]). *For a closed symplectic manifold* (M,ω), $\Gamma_\omega \subset H^1(M;\mathbb{R})$ *is a discrete subgroup, and we have*

$$\mathrm{Symp}_0(M,\omega)/\mathrm{Ham}(M,\omega) \cong H^1(M,\omega)/\Gamma_\omega.$$

LECTURE 9

The Hofer metric and displacement energy

Having discussed $\text{Ham}(M,\omega)$ as both a group and a topological space, we now explain its natural bi-invariant metric, the existence of which is a "wonder" of symplectic geometry.

9.1. The Hofer metric and its basic properties

For a Hamiltonian $H : M \to \mathbb{R}$, put

$$||H|| := \max_M H - \min_M H.$$

Note that $||H||$ is unaffected if we add a constant to H. For a path $\{\phi_t\}$ in $\text{Ham}(M,\omega)$, we define its length by

$$\text{len}(\{\phi_t\}) := \int_0^1 ||H_t||dt,$$

where $\partial_t \phi_t = X_{H_t}$.

DEFINITION 9.1 (The Hofer metric). *We $\phi, \psi \in \text{Ham}(M,\omega)$,*

$$d_{\text{Hof}}(\phi,\psi) := \inf\{\text{len}(\{\gamma_t\}_{t\in[0,1]} \mid \gamma_t \in \text{Ham}(M,\omega),\ \gamma_0 = \phi,\ \gamma_1 = \psi\},$$

i.e. the infimal length of a path in $\text{Ham}(M,\omega)$ joining ϕ and ψ.

Since $\mathbb{1} \in \text{Ham}(M,\omega)$ is the canonical choice of basepoint, we can think of the Hofer metric as giving a measure of the "size" of any Hamiltonian diffeomorphism:

DEFINITION 9.2 (The Hofer norm). *For $\phi \in \text{Ham}(M,\omega)$, put $||\phi||_{\text{Hof}} := d_{\text{Hof}}(\phi, \mathbb{1})$.*

LEMMA 9.3. *The Hofer metric satisfies the following basic properties for any ϕ, ψ, θ in $\text{Ham}(M,\omega)$:*

(1) *(nonnegativity) $d_{\text{Hof}}(\phi,\psi) \geq 0$ and $d_{\text{Hof}}(\phi,\phi) = 0$*
(2) *(symmetry) $d_{\text{Hof}}(\phi,\psi) = d_{\text{Hof}}(\psi,\phi)$*
(3) *(triangle inequality) $d_{\text{Hof}}(\phi,\psi) \leq d_{\text{Hof}}(\phi,\theta) + d_{\text{Hof}}(\theta,\psi)$*
(4) *(bi-invariance) $d_{\text{Hof}}(\phi,\psi) = d_{\text{Hof}}(\theta \circ \phi, \theta \circ \psi) = d_{\text{Hof}}(\phi \circ \theta, \psi \circ \theta)$.*

EXERCISE 9.4. *Check the above properties.*

The first three properties are equivalent to saying that d_{Hof} is a **pseudometric**. In fact, d_{Hof} is a genuine metric, but compared to the other properties this is rather nontrivial and requires some machinery to prove:

THEOREM 9.5. [**Hof, Pol, LM**] *For any closed symplectic manifold (M,ω), the Hofer metric is nondegenerate, i.e. we have $d(\phi,\psi) > 0$ if $\phi \neq \psi$.*

REMARK 9.6. *By bi-invariance, nondegeneracy of $d_{\text{Hof}}(-,-)$ follows from nondegeneracy of $|| - ||_{\text{Hof}}$, i.e. it suffices to prove that $||\phi||_{\text{Hof}} = 0$ only if $\phi = \mathbb{1}$.*

We will discuss an approach to the proof of Theorem 9.5 in the coming lectures. Following [**PR**], we will also use this as an excuse to introduce several other tools and abstractions.

Theorem 9.5 is quite fortunate, since it turns out that if d_{Hof} were not nondegenerate then it would be completely trivial:

EXERCISE 9.7. *Let G be any simple group, and let d be a pseudometric on G which is bi-invariant, i.e. such that $d(gg_1, gg_2) = d(g_1g, g_2g) = d(g_1, g_2)$ for any $g, g_1, g_2 \in G$. Then either d is nondegenerate, or else d is identically zero.*

Hint: show that $\{h \in G \mid d(\mathbb{1}, h) = 0\}$ is a normal subgroup of G, and then invoke simplicity.

EXERCISE 9.8. *Convince yourself that the topology on $\mathrm{Ham}(M, \omega)$ induced by the Hofer metric is strictly weaker than the C^∞ topology.*

9.2. Displacement energy

The Hofer metric gives a natural way of measuring quantifying various other geometric aspects of a symplectic manifold. One particularly relevant notion is that of displacement energy:

DEFINITION 9.9. *Let (M, ω) be a symplectic manifold and let $A \subset M$ be any subset. The* **displacement energy** *of A is*

$$e_{\mathrm{Hof}}(A) := \inf\{||\phi||_{\mathrm{Hof}} \mid \phi \in \mathrm{Ham}(M, \omega),\ \phi(A) \cap A = \varnothing\}.$$

We say that A is **displaceable** *if $e_{\mathrm{Hof}}(A) < \infty$, i.e. there exists $\phi \in \mathrm{Ham}(M, \omega)$ such that $\phi(A) \cap A = \varnothing$.*

PROPOSITION 9.10. *For any nonempty open subset $A \subset M$, we have $e_{\mathrm{Hof}}(A) > 0$.*

Taking Theorem 9.5 for granted for the moment, we show explain how to deduce the above proposition. We start with an elementary lemma. For $\phi, \psi \in \mathrm{Ham}(M, \omega)$ we denote their **commutator** by $[\phi, \psi] := \phi\psi\phi^{-1}\psi^{-1}$.

LEMMA 9.11. *For any $\phi, \psi \in \mathrm{Ham}(M, \omega)$, we have $||[\phi, \psi]||_{\mathrm{Hof}} \leq 2||\psi||_{\mathrm{Hof}}$.*

PROOF. We have

$$\begin{aligned}
||\phi\psi\phi^{-1}\psi^{-1}||_{\mathrm{Hof}} &= d_{\mathrm{Hof}}(\phi\psi\phi^{-1}\psi^{-1}, \mathbb{1}) \\
&= d_{\mathrm{Hof}}(\phi\psi\phi^{-1}, \psi) \\
&\leq d_{\mathrm{Hof}}(\phi\psi\phi^{-1}, \mathbb{1}) + d_{\mathrm{Hof}}(\mathbb{1}, \psi) \\
&= 2||\psi||_{\mathrm{Hof}}.
\end{aligned}$$

$\square$

PROOF OF PROPOSITION 9.10, ASSUMING THEOREM 9.5. We might as well assume that the nonempty open subset A is displaceable, since otherwise there is nothing to prove. Suppose that θ displaces A. Evidently we have $||\theta||_{\mathrm{Hof}} > 0$, but this does not prove $e_{\mathrm{Hof}}(A) > 0$ since the definition of the latter involves an infimum. We seek to find a lower bound for $||\theta||_{\mathrm{Hof}}$ which is independent of θ, whence the infimum defining $e_{\mathrm{Hof}}(A)$ will be positive.

Let $\phi, \psi \in \mathrm{Ham}(M, \omega)$ be Hamiltonian diffeomorphisms supported in A, and such that $\phi \circ \psi \neq \psi \circ \phi$ (see Exercise 9.12). Then by nondegeneracy of the Hofer norm we have

$||[\phi, \psi]||_{\mathrm{Hof}} > 0$. Note that $\theta\psi^{-1}\theta^{-1}$ is supported in $\theta(A)$, which is disjoint from A, and hence $\theta\psi^{-1}\theta^{-1}$ commutes with ϕ. We then have

$$[\phi, [\psi, \theta]] = \phi(\psi\theta\psi^{-1}\theta^{-1})\phi^{-1}(\theta\psi\theta^{-1}\psi^{-1})$$
$$= \phi\psi\phi^{-1}\psi^{-1}$$
$$= [\phi, \psi],$$

and hence by Lemma 9.11 we have

$$||[\phi, \psi]||_{\mathrm{Hof}} = ||[\phi, [\psi, \theta]]||_{\mathrm{Hof}}$$
$$\leq 2||[\psi, \theta]||_{\mathrm{Hof}}$$
$$\leq 4||\theta||_{\mathrm{Hof}},$$

and therefore $e_{\mathrm{Hof}}(A) \geq \frac{1}{4}||[\phi, \psi]||_{\mathrm{Hof}} > 0$. $\square$

EXERCISE 9.12. *Let A be a nonempty open subset of a symplectic manifold (M, ω). Prove that there exist $\phi, \psi \in \mathrm{Ham}(M, \omega)$ such that*

- *ϕ and ψ are supported in A, i.e. $\phi(p) = \psi(p) = p$ for any $p \in M \setminus A$*
- *ϕ and ψ do not commute, i.e. $\phi \circ \psi \neq \psi \circ \phi$ as elements of $\mathrm{Ham}(M, \omega)$.*

Some possible hints: Recall that by Darboux's theorem there is an open subset of A which is symplectomorphic to a ball in Euclidean space with its standard symplectic structure. Note that we can cut off the support of a Hamiltonian diffeomorphism by cutting off the generating Hamiltonians. Recall that in dimension two symplectomorphisms are just area-preserving diffeomorphisms.

LECTURE 10

Ham as an infinite-dimensional Lie group

In this lecture we would like to further develop the idea that $\text{Ham}(M,\omega)$ is an infinite-dimensional Lie group, where the group multiplication is given by composition and the topology is the C^∞ topology (see Definition 8.14). Our motivation is to put the Hofer metric in greater context and to compare it with related constructions for finite-dimensional Lie groups. Since this lecture is mostly intended to build intuition, we will not give a formal definition of "infinite-dimensional Lie group", although we will assume familiarity will finite-dimensional Lie groups.

10.1. Finsler metrics

DEFINITION 10.1. *A* **Finsler metric** *on a finite-dimensional manifold M is a choice of asymmetric norm on each tangent space. More precisely, it is a continuous function $\mu : TM \to \mathbb{R}_{\geq 0}$, which is smooth away from the zero section, such that for all $p \in M$ and $v, w \in T_pM$ we have:*

- *(subadditivity) $\mu(v + w) \leq \mu(v) + \mu(w)$*
- *(positive homogeneity) $\mu(\lambda v) = \lambda\mu(v)$ for all $\lambda \in \mathbb{R}_{\geq 0}$*
- *(positive definiteness) $\mu(v) > 0$ if v is nonzero.*

Given a Finsler metric μ on M, we define the length of a curve $\gamma : [a, b] \to M$ by

$$\text{len}_\mu(\gamma) := \int_a^b \mu(\dot\gamma(t))dt,$$

and we define the distance between two points $p, q \in M$ by

$$d_\mu(p, q) := \inf\{\text{len}(\gamma) \mid \gamma : I \to M, \ \gamma(0) = p, \ \gamma(1) = q\}.$$

One can check that d_μ satisfies the axioms of a metric except for possibly symmetry, and it is also symmetric if μ is fully homogeneous, i.e. $\mu(\lambda v) = \lambda\mu(v)$ for all (possibly negative) $\lambda \in \mathbb{R}$.

If G is a finite-dimensional Lie group, recall that its Lie algebra $\mathfrak{g}$ is the space of vector fields on G which are invariant under left multiplication by elements of G. Since a left-invariant vector field is determined by its value at the identity element $e \in G$, we have an identification $\mathfrak{g} \approx T_eG$. The Lie bracket on $\mathfrak{g}$ is defined using the Lie bracket (i.e. $[V, W]f = V(W(f)) - W(V(f))$) of the corresponding left-invariant vector fields. Recall that G acts on instead by conjugation, and by linearizing at the identity element we get the adjoint action

$$\text{Ad} : G \to \text{End}(\mathfrak{g}).$$

Note that any norm $||-||$ on T_eG induces a left-invariant Finsler metric on G by putting $\mu(v) = \mu(dL_{g^{-1}}(v))$ for any $g \in G$ and $v \in T_gG$ (here $L_g : G \to G$ denotes left multiplication by g). If $||-||$ is in addition Ad-invariant, i.e. $||v|| = ||Ad_g(v)||$ for any $v \in \mathfrak{g}$ and $g \in G$,

then this prescription gives a Finsler metric which is also right invariant, whence the induced metric d_μ on G is bi-invariant.

It turns out that having a bi-invariant Finsler metric on a finite dimensional Lie group is quite restrictive.

PROPOSITION 10.2. *If G is a simple finite-dimensional Lie group with a bi-invariant Finsler metric, then G is compact.*

In particular, the existence of the Hofer metric on $\mathrm{Ham}(M, \omega)$ should be viewed as a purely infinite-dimensional phenomenon.

10.2. The infinite-dimensional setting

Let M be a closed smooth manifold. We view $\mathrm{Diff}(M)$ as an infinite-dimensional Lie group, with multiplication given by composition and topology given by the C^∞ topology. The Lie algebra is $T_{\mathbb{1}}\mathrm{Diff}(M)$, which is identified with the vector space of smooth vector fields on M, and with Lie bracket given by the usual Lie bracket of vector fields.

EXERCISE 10.3. *For any differential k-form α and vector fields A, B, we have*

$$[\mathcal{L}_A, \iota_B]\alpha := \mathcal{L}_A \iota_B \alpha - \iota_B \mathcal{L}_A \alpha = \iota_{[A,B]}\alpha.$$

Hint: the Lie derivative satisfies the Leibiz rule

$$\mathcal{L}_A(\alpha(B_1, \ldots, B_k)) = (\mathcal{L}_A\alpha)(B_1, \ldots, B_k) + \alpha(\mathcal{L}_A B_1, B_2, \ldots, B_k) + \cdots + \alpha(B_1, \ldots, B_{k-1}, \mathcal{L}_A B_k)$$

for vector fields $A, B_1, \ldots, B_k$. Recall that $\mathcal{L}_A B_i = [A, B_i]$.

Suppose that A and B are symplectic vector fields, i.e. $d(\iota_A \omega) = d(\iota_B \omega) = 0$. Then using Exercise 10.3 we have

$$\begin{aligned}
\iota_{[A,B]}\omega &= [\mathcal{L}_A, \iota_B]\omega \\
&= d\iota_A \iota_B \omega + \iota_A d\iota_B \omega - \iota_B d\iota_A \omega - \iota_B \iota_A d\omega \\
&= d(\omega(B, A)).
\end{aligned}$$

This shows that $[A, B] = -X_{\omega(A,B)}$[1] is a Hamiltonian vector field.

In particular, if $A = X_F$ and $B = X_G$ are Hamiltonian vector fields, then we have

$$(10.1) \qquad\qquad\qquad\qquad [X_F, X_G] = -\{F, G\},$$

where $\{F, G\} := \omega(X_F, X_G)$ is the **Poisson bracket** of F and G.

EXERCISE 10.4. *For any two Hamiltonians $F, G : M \to \mathbb{R}$, their Poisson bracket $\{F, G\}$ is normalized.*

Hint: show that we have

$$dF \wedge dG \wedge \omega^{\wedge(n-1)} = -\tfrac{1}{n}\{F, G\}\omega^{\wedge n}.$$

In other words, given a closed symplectic manifold (M, ω), the Lie algebra of $\mathrm{Ham}(M, \omega)$ is the space $\mathcal{H}$ of normalized Hamiltonian functions $H : M \to \mathbb{R}$, and the Lie bracket is given by the Poisson bracket.

Given $\phi \in \mathrm{Ham}(M, \omega)$ and $H \in \mathcal{H}$, one can check that the adjoint action is given by

$$\mathrm{Ad}_\phi(H) = H \circ \phi^{-1} \in \mathcal{H}.$$

---

[1]Note the sign flip here! This is rather unpleasant but unavoidable based on our conventions.

By the above discussion, we can pick any norm $|| - ||'$ on $\mathcal{H}$ such that $||H||' = ||H \circ \phi||'$ for all $\phi \in \text{Ham}(M, \omega)$, and this will induce bi-invariant pseudometric on $\text{Ham}(M, \omega)$. For example, we could take the norm

$$||H||_{L^p} := \left(\int_M |H|^p \omega^{\wedge n} \right)^{1/p}$$

for $p \in [1, \infty)$. in particular, in the case of Ham explain that there are many viable possibilities, e.g. any L^p metric would work. As it turns out, any such choice will either will result in something which is either degenerate or equivalent to the Hofer metric:

THEOREM 10.5. [**BO**] *Let* $|| - ||'$ *be a norm on* $\mathcal{H}$ *which is invariant under the adjoint action of* $\text{Ham}(M, \omega)$ *and which is continuous with respect to the* C^∞ *topology on* $\text{Ham}(M, \omega)$. *Then either the induced bi-invariant pseudometric on* $\text{Ham}(M, \omega)$ *is trivial, or else* $|| - ||'$ *is equivalent to the uniform norm.*

Recall that two norms $|| - ||_1$ and $|| - ||_2$ on a vector space V are *equivalent* if there exist positive constants C_1, C_2 such that we have $C_1||v||_1 \leq ||v||_2 \leq C_2||v||_1$ for all $v \in V$.

The action spectrum

11.1. The action of a contractible periodic Hamiltonian orbit

Let (M, ω) be a closed symplectic manifold. We begin by defining the action spectrum of a time-dependent Hamiltonian $H : M \to \mathbb{R}$, which turns out to depend only on the flow as an element of $\widetilde{\mathrm{Ham}}(M, \omega)$. We then define spectral invariants, which input elements of $\widetilde{\mathrm{Ham}}(M, \omega)$ and take values in the associated action spectrum. The existence of certain subadditive spectral invariants, which we will establish later using filtered Floer homology, will play a key role in our proof that the Hofer metric is nondegenerate.

Recall that elements of $\widetilde{\mathrm{Ham}}(M, \omega)$ are paths in $\mathrm{Ham}(M, \omega)$ which begin at $\mathbb{1}$, modulo homotopies rel endpoints. This inherits a group structure via $\{\phi_t\} \cdot \{\psi_t\} := \{\phi_t \circ \psi_t\}$ for $\{\phi_t\}, \{\psi_t\} \in \widetilde{\mathrm{Ham}}(M, \omega)$.

EXERCISE 11.1. *Show that the group multiplication is equivalently given by concatenation. More precisely, $\{\phi_t\} \cdot \{\phi_t\}$ is homotopic rel endpoints to the path $\{\phi_{2t}\}$ for $t \in [0, 1/2]$, followed by $\{\phi_1 \circ \phi_{2t-1}\}$ for $t \in [1/2, 1]$.*

Given $\{\phi_t\} \in \widetilde{\mathrm{Ham}}(M, \omega)$, denote the fixed points of $\phi_1 \in \mathrm{Ham}(M, \omega)$ by

$$\mathrm{Fix}(\phi_1) := \{p \in M \mid \phi_1(p) = p\}.$$

Note that for $p \in \mathrm{Fix}(\phi_1)$, we get an associated closed loop $\gamma_p : [0, 1] \to M$ based at p via $\gamma_p(t) = \phi_t(p)$.

DEFINITION 11.2. *Given any contractible loop $\gamma : S^1 \to M$, a **capping disk** is a smooth map $D : \mathbb{D}^2 \to M$ such that $D|_{\partial \mathbb{D}^2} = \gamma$.*

DEFINITION 11.3. *Let $H : M \times I \to \mathbb{R}$ be a time-dependent Hamiltonian, with Hamiltonian flow $\{\phi_t\} \in \widetilde{\mathrm{Ham}}(M, \omega)$. Suppose that $p \in \mathrm{Fix}(\phi_1)$, and assume that the loop $\gamma_p : [0, 1] \to M$ is contractible. Let D be a capping disk for γ_p. The **action** of the pair (p, D) is*

$$(11.1) \qquad \mathcal{A}_H(p, D) := -\int_{\mathbb{D}^2} D^* \omega + \int_0^1 H(\phi_t(p), t) dt.$$

If D and D' are two different capping disks for γ_p, observe that we have

$$\mathcal{A}_F(p, D) - \mathcal{A}_F(p, D') = \int_{S^2} u^* \omega,$$

where $u : S^2 \to M$ is given by gluing together D' to D along their common boundary. More preicsely, defining the integral $\int u^* \omega$ requires a choice of orientation on S^2, and we use the orientation induced by that of D' and the opposite orientation on D.

DEFINITION 11.4. *The **group of periods** of (M, ω) is the subgroup of $\mathbb{R}$ given by*

$$L_\omega := \left\{ \int_{S^2} u^* \omega \mid u : S^2 \to M \right\}.$$

Therefore we can also say that, if γ_y is contractible, the action $\mathcal{A}_H(y)$ is well-defined in $\mathbb{R}/L_\omega$.

EXAMPLE 11.5. *Let ω_{std} be a symplectic form on S^2 such that $\int_{S^2} \omega_{\mathrm{std}} = 1$, and let $\pi_i : S^2 \times S^2 \to S^2$ be projection onto the ith factor for $i = 1, 2$. If we endow $S^2 \times S^2$ with the symplectic form $a\pi_1^*\omega_{\mathrm{std}} + b\pi_2^*\omega_{\mathrm{std}}$ for some $a, b \in \mathbb{R}_{>0}$, then L_ω is the subgroup of $\mathbb{R}$ generated by a, b.*

PROPOSITION 11.6. *Let (M, ω) be a closed symplectic manifold, and suppose that $F, G : M \times I \to \mathbb{R}$ are two time-dependent Hamiltonians which both generate the same element of $\widetilde{\mathrm{Ham}}(M, \omega)$. Suppose that $y \in \mathrm{Fix}(Fl_F^1) = \mathrm{Fix}(Fl_G^1)$, and the associated loops in M are contractible. Then we have $\mathcal{A}_F(y) = \mathcal{A}_G(y) \mod L_\omega$.*

Note that F, G generating the same element of $\widetilde{\mathrm{Ham}}(M, \omega)$ means that $Fl_F^1 = Fl_G^1$, and then $\{Fl_F^t\}$ is homotopic to $\{Fl_G^t\}$ rel endpoints. In particular, the loops $\{Fl_F^t(y)\}$ and $\{Fl_G^t\}$ are homotopic, so one is contractible if and only if the other is.

11.2. Two-parameter families of Hamiltonian diffeomorphisms

Before proving Proposition 11.6, we will need a preliminary lemma.

LEMMA 11.7. [**Ban**] *Let M be a smooth manifold, and let $\{\phi_{s,t} \in \mathrm{Diff}(M)\}$ be a 2-parameter family of diffeomorphisms. Let $X_{s,t}$ and $Y_{s,t}$ be the associated 2-parameter families of vector fields on M satisfying*

$$X_{s,t}(\phi_{s,t})(p) = \partial_t \phi_{s,t}(p)$$
$$Y_{s,t}(\phi_{s,t})(p) = \partial_s \phi_{s,t}(p)$$

for all $p \in M$. Then we have

$$\partial_s X_{s,t} - \partial_t Y_{s,t} = [X_{s,t}, Y_{s,t}].$$

PROOF. Endow $\mathbb{R}^2$ with coordinates (s, t), and let $\pi : M \to \mathbb{R}^2 \to \mathbb{R}^2$ denote the projection to $\mathbb{R}^2$. We can view any (s, t)-dependent vector field on M as a single vector field on $M \times \mathbb{R}^2$ which is tangent to each fiber $M \times \{(s_0, t_0)\}$ of π. Let $\widetilde{X}$ denote the vector field on $M \times \mathbb{R}^2$ given by $\widetilde{X} = X_{s,t} + \partial_t$. Similarly, put $\widetilde{Y} := Y_{s,t} + \partial_s$ as a vector field on $M \times \mathbb{R}^2$ which is tangent to the fibers of π. Observe that there is a foliation of $M \times \mathbb{R}^2$ by leaves of the form $\{\phi_{s,t}(p) \mid (s, t) \in \mathbb{R}^2\}$ for $p \in M$. Moreover, $\widetilde{X}$ and $\widetilde{Y}$ are tangent to these leaves, and so by the Frobenius integrability theorem we have that $[\widetilde{X}, \widetilde{Y}]$ is also tangent to these leaves. Not that we have $d\pi([\widetilde{X}, \widetilde{Y}]) = [d\pi(\widetilde{X}), d\pi(\widetilde{Y})] = [\partial_s, \partial_t] = 0$, and hence $[\widetilde{X}, \widetilde{Y}]$ is also tangent to the fibers of π. Since the leaves of the foliation intersect the fibers of π transversely, we conclude that $[\widetilde{X}, \widetilde{Y}] = 0$, and therefore we have

$$0 = [\widetilde{X}, \widetilde{Y}] = [X_{s,t}, Y_{s,t}] + [\partial_t, Y_{s,t}] + [X_{s,t}, \partial_s] + [\partial_s, \partial_t]$$
$$= [X_{s,t}, Y_{s,t}] + \partial_t Y_{s,t} - \partial_s X_{s,t},$$

which immediately gives the desired formula. $\qquad\square$

COROLLARY 11.8. *Let (M, ω) be a closed symplectic manifold, and let $\phi_{s,t} \in \mathrm{Ham}(M, \omega)$ be a 2-parameter family of Hamiltonian diffeomorphisms. Let $A_{s,t}, B_{s,t} : M \to \mathbb{R}$ be the corresponding 2-parameter families of normalized Hamiltonians such that we have*

$$\partial_t \phi_{s,t}(p) = X_{A_{s,t}}(\phi_{s,t}(p))$$
$$\partial_s \phi_{s,t}(p) = X_{B_{s,t}}(\psi_{s,t}(p))$$

for all $p \in M$. Then we have

$$\partial_s A_{s,t} - \partial_t B_{s,t} = -\{A_{s,t}, B_{s,t}\}.$$

PROOF. Using Lemma 11.7 and (10.1), we have

$$\partial_s X_{A_{s,t}} - \partial_t X_{B_{s,t}} = [X_{A_{s,t}}, X_{B_{s,t}}] = -X_{\{A_{s,t}, B_{s,t}\}},$$

and hence

$$-d\{A_{s,t}, B_{s,t}\} = \omega(\partial_s X_{A_{s,t}}, -) - \omega(\partial_t X_{B_{s,t}}, -)$$
$$= \partial_s dA_{s,t} - \partial_t dB_{s,t},$$

and hence $\partial_s A_{s,t} - \partial_t B_{s,t}$ agrees with $-\{A_{s,t}, B_{s,t}\}$ up to a constant $C_{s,t}$, which can be readily seen to be zero using Exercise 10.4. $\qquad\square$

11.3. Invariance of action under homotopies

PROOF OF PROPOSITION 11.6. Put $\psi := Fl_F^1 = Fl_G^1 \in \mathrm{Ham}(M, \omega)$. Since F and G generate the same element of $\widetilde{\mathrm{Ham}}(M, \omega)$, we can find a 2-parameter family $\{\phi_{s,t}\}_{s,t \in [0,1]}$ in $\mathrm{Ham}(M, \omega)$ such that

- $\phi_{0,t} = Fl_F^t$ and $\phi_{1,t} = Fl_G^t$
- $\phi_{s,0} = \mathbb{1}$ for all $s \in [0,1]$
- $\phi_{s,1} = \psi$ for all $s \in [0,1]$.

Let y be a fixed point of ψ, and assume that the loops γ_0, γ_1 defined by $\gamma_i(t) = \phi_{i,t}$ are contractible.

Suppose that D is a capping disk for γ_0. Let $u : [0,1] \times S^1 \to M$ be the cylinder in M defined by $u(s,t) = \phi_{s,t}(y)$. Then we can concatenate D with u to get a capping disk D' for γ_1. It suffices to establish

$$\mathcal{A}_F(y, D) = \mathcal{A}_G(y, D').$$

Using (11.1), we have

$$(11.2) \quad \mathcal{A}_G(y, D') - \mathcal{A}_F(y, D) = -\int_{[0,1] \times S^1} u^* \omega + \int_0^1 G(\gamma_1(t), t)dt - \int_0^1 F(\gamma_0(t), t)dt.$$

As in Corollary 11.8, let $A_{s,t}, B_{s,t} : M \to \mathbb{R}$ be 2-parameter families of normalized Hamiltonian vector fields such that

$$\partial_t \phi_{s,t}(p) = X_{A_{s,t}}(\phi_{s,t}(p))$$
$$\partial_s \phi_{s,t}(p) = X_{B_{s,t}}(\phi_{s,t}(p))$$

for all $p \in M$, so that we have

$$\partial_s A_{s,t} - \partial_t B_{s,t} = \{A_{s,t}, B_{s,t}\}.$$

Note that we have

- $A_{0,t} = F_t$ and $A_{1,t} = G_t$
- $B_{s,0} = B_{s,1} = 0$.

We then have

$$-\int_{[0,1]\times S^1} u^*\omega = \int_0^1\int_0^1 \omega(\partial_s u,\partial_t u)dsdt$$

$$= -\int_0^1\int_0^1 \omega(X_{A_{s,t}},X_{B_{s,t}})\circ u\,dsdt$$

$$= -\int_0^1\int_0^1 \{A_{s,t},B_{s,t}\}\circ u\,dsdt$$

$$= \int_0^1\int_0^1 (\partial_s A_{s,t} - \partial_t B_{s,t})\circ u\,dsdt$$

$$= \int_0^1\int_0^1 (\partial_s(A_{s,t}\circ u) - dA_{s,t}\circ\partial_s u - \partial_t(B_{s,t}\circ u) + dB_{s,t}\circ\partial_t u)\,dsdt$$

$$= \int_0^1\int_0^1 (\partial_s(A_{s,t}\circ u) - \partial_t(B_{s,t}\circ u))\,dsdt - 2\int_0^1\int_0^1 \{A_{s,t},B_{s,t}\}\circ u\,dsdt,$$

and therefore

$$\int_{[0,1]\times S^1} u^*\omega = \int_0^1\int_0^1 (\partial_s(A_{s,t}\circ u) - \partial_t(B_{s,t}\circ u))\,dsdt$$

$$= \int_0^1 (A_{1,t} - A_{0,t})\circ u\,dt$$

$$= \int_0^1 G(\gamma_1(t),t) - \int_0^1 F(\gamma_0(t),t),$$

which combined with (11.2) readily gives the desired result. $\qquad\square$

11.4. The action spectrum

Given $\phi = [\{\phi_t\}] \in \widetilde{\mathrm{Ham}}(M,\omega)$ and $y \in \mathrm{Fix}(\phi_1)$ such that the corresponding loop γ_y in M is contractible, we put $\mathcal{A}_\phi(y,D) := \mathcal{A}_F(y,D)$, where $F : M \times I \to \mathbb{R}$ is any normalized time-dependent Hamiltonian such that $[\{Fl_F^t\}] = [\{\phi_t\}] \in \widetilde{\mathrm{Ham}}(M,\omega)$, and D is a choice of capping disk of γ_y.

DEFINITION 11.9. *Given* $\phi = [\{\phi_t\}] \in \widetilde{\mathrm{Ham}}(M,\omega)$, *its* **action spectrum** $\mathrm{spec}(\phi) \subset \mathbb{R}$ *is the set of all actions* $\mathcal{A}_\phi(y,D)$, *where* y *varies fixed points of* ϕ_1 *such that* γ_y *is contractible, and* D *is a capping disk of* γ_y.

A particularly nice situation is when $\phi_1 \in \mathrm{Ham}(M,\omega)$ is **nondegenerate**, meaning that for each $y \in \mathrm{Fix}(\phi_1)$, the linear map $d\phi_1|_y : T_yM \to T_yM$ does not have 1 as an eigenvalue. In this case, the fixed points of ϕ_1 are isolated, and hence there are only finitely many (assuming M is compact). We will say that $[\{\phi_t\}] \in \widetilde{\mathrm{Ham}}(M,\omega)$ is nondegenerate if ϕ_1 is. In this case, $\mathrm{spec}([\{\phi_t\}])$ is a finitely generated module over $\mathbb{Z}[L_\omega]$.

EXERCISE 11.10. *Show that a Hamiltonian diffeomorphism* $\phi \in \mathrm{Ham}(M,\omega)$ *is nondegenerate if and only if its graph* $\mathrm{Gr}(\phi) := \{(x,\phi(x) \mid x \in M\} \subset M \times M$ *is transverse to the diagonal* $\Delta := \{(x,x) \mid x \in M\} \subset M \times M$. *Conclude that the fixed points of* ϕ *are isolated.*

REMARK 11.11. *Suppose that* M *is* **symplectically aspherical**, *i.e.* $\int_{S^2} u^*\omega = 0$ *for all* $u : S^2 \to M$ *(e.g. this holds whenever* $\pi_2(M) = 0$*). Then* $\mathcal{A}_F(y,D)$ *is in fact independent of the choice of capping disk* D.

LECTURE 12

Subadditive spectral invariants

12.1. The key ingredient

We begin with a definition:

DEFINITION 12.1. *Let (M, ω) be a closed symplectic manifold. A* **subadditive spectral invariant** *is a function $c : \widetilde{\text{Ham}}(M, \omega) \to \mathbb{R}$ satisfying the following axioms for any $\phi, \psi \in \widetilde{\text{Ham}}(M, \omega)$:*

- *(1) (conjugation invariance) $c(\phi\psi\phi^{-1}) = c(\psi)$*
- *(2) (subadditivity) $c(\phi\psi) \leq c(\phi) + c(\psi)$*
- *(3) (stability) $\int_0^1 \min(F_t - G_t)dt \leq c(\phi) - c(\psi) \leq \int_0^1 \max(F_t - G_t)dt$, where $F, G : M \times I \to \mathbb{R}$ are normalized Hamiltonians generating ϕ, ψ respectively*
- *(4) (spectrality) $c(\phi) \in \text{spec}(\phi)$ if ϕ is nondegenerate.*

The two conditions can be viewed as compatibility with the group structure on $\widetilde{\text{Ham}}(M, \omega)$. The third condition is akin to a continuity property for c. The last condition, which connects with the previous lecture, ties c with dynamical properties of ϕ and makes the exists of c rather nontrivial to establish.

We following theorem will be our key nontrivial input from Floer theory:

THEOREM 12.2. *Every closed symplectic manifold (M, ω) admits a subadditive spectral invariant $c : \widetilde{\text{Ham}}(M, \omega) \to \mathbb{R}$ which moreover satisfies $c(\mathbb{1}) = 0$.*

For the next several lectures, we will take Theorem 12.2 as a black box and use it to deduce various important applications. Along the way, we will develop some formalism which is independently interesting and useful.

12.2. Nondegeneracy of the Hofer metric

Given any subadditive spectral invariant $c : \widetilde{\text{Ham}}(M, \omega)$, we make the following further definitions:

DEFINITION 12.3. *Let $q_c : \widetilde{\text{Ham}}(M, \omega) \to \mathbb{R}$ be defined by $q_c(\phi) := c(\phi) + c(\phi^{-1})$.*

Note $q_c(\phi)$ is always nonnegative, since using subadditivity we have

$$c(\mathbb{1}) = c(\mathbb{1} \cdot \mathbb{1}) \leq c(\mathbb{1}) + c(\mathbb{1}),$$

and hence

$$0 \leq c(\mathbb{1}) = c(\phi \cdot \phi^{-1}) \leq c(\phi) + c(\phi^{-1}) = q(\phi).$$

We will view $q_c(\phi)$ as a measure of the "size" of ϕ which is a kind of algebraic analogue of the Hofer norm $||\phi||_{\text{Hof}}$.

DEFINITION 12.4. *For an open subset $U \subset M$, we define the* **spectral displacement energy** *by*

$$e_c(U) := \inf\{q_c([\{\phi_t\}]) \mid [\{\phi_t\}] \in \widetilde{\operatorname{Ham}}(M\omega), \ \phi_1(U) \cap U = \varnothing\}.$$

Note that $e_c(U)$ is defined just like the usual displacement energy $e_{\operatorname{Hof}}(U)$, but using q_c in place of the Hofer norm.

Our next order of business is to establish the following three results:

LEMMA A. *For any displaceable open subset $U \subset M$, we have $e_c(U) \leq 2c(\mathbb{1}) + e_{\operatorname{Hof}}(U)$.*

LEMMA B. *For any open and displacement subset $U \subset M$ and for any $\phi \in \widetilde{\operatorname{Ham}}_U(M,\omega)$ we have $q_c(\phi) \leq 2e_c(U)$.*

Here $\widetilde{\operatorname{Ham}}_U(M,\omega)$ denotes the subgroup of $\widetilde{\operatorname{Ham}}(M,\omega)$ given by Hamiltonian isotopies which are generated by a (not necessarily normalized) Hamiltonian $H : M \times I \to \mathbb{R}$ such that H_t is supported in U for all $t \in [0,1]$. Of course for any Hamiltonian supported in U we can always add a constant in order to make it normalized (i.e. have its integral over M equal to zero), but the result will generally no longer be supported in U.

LEMMA C. *For any $\phi = [\{\phi_t\}] \in \widetilde{\operatorname{Ham}}(M,\omega)$ such that $\phi_1 \neq \mathbb{1} \in \operatorname{Ham}(M,\omega)$, we have $q_c(\phi) > 0$.*

Assuming these lemmas, we can easily prove nondegeneracy of the Hofer norm, as well as:

THEOREM 12.5. *For any nonempty open subset $U \subset M$, we have $e_{\operatorname{Hof}}(U) > 0$.*

REMARK 12.6. *Note that this is exactly the same as Proposition 9.10, which we previously proved assuming nondegeneracy of the Hofer norm. We now slightly shuffle the logic and instead prove Theorem 12.5 assuming Theorem 12.2 as well as Lemmas A,B,C, and we then use Theorem 12.5 to easily deduce nondegeneracy of the Hofer norm below.*

PROOF OF THEOREM 12.5. We can assume without loss of generality that U is displaceable. Take any $\phi = [\{\phi_t\}] \in \widetilde{\operatorname{Ham}}_U(M,\omega)$ such that $\phi_1 \neq \mathbb{1}$. Let $c : \widetilde{\operatorname{Ham}}(M,\omega)$ be a subadditive spectral invariant with $c(\mathbb{1}) = 0$. Combining Lemmas A,B,C, we have

$$0 < q_c(\phi) \leq 2e_c(U) \leq 4c(\mathbb{1}) + 2e_{\operatorname{Hof}}(U) = 2e_{\operatorname{Hof}}(U).$$

$\square$

PROOF OF THEOREM 9.5. Observe that any $\phi \in \operatorname{Ham}(M,\omega)$ displaces some open subset $U \subset M$, whence by Theorem 12.5 and the definition of displacement energy we have

$$0 < e_{\operatorname{Hof}}(U) \leq ||\phi||_{\operatorname{Hof}}.$$

$\square$

12.3. Upper bound on the spectral displacement energy

Here we prove Lemma A. Firstly, given $\phi = [\{\phi_t\}] \in \widetilde{\operatorname{Ham}}(M,\omega)$, let us define

$$\widetilde{d}_{\operatorname{Hof}}(\mathbb{1}, \phi) := \inf\{\operatorname{len}(\{\psi_t\}) \mid [\{\psi_t\}] = [\{\phi_t\}]\}.$$

Note that this is just defined like $d_{\mathrm{Hof}}(\mathbb{1}, \phi_1)$, except that we only infimize only paths in $\mathrm{Ham}(M, \omega)$ from $\mathbb{1}$ to ϕ_1 which are homotopy rel endpoints to the path $\{\phi_t\}$. More generally, for any $\phi, \psi \in \widetilde{\mathrm{Ham}}(M, \omega)$ we put

$$\widetilde{d}_{\mathrm{Hof}}(\phi, \psi) := \widetilde{d}_{\mathrm{Hof}}(\mathbb{1}, \phi^{-1}\psi).$$

This defines a bi-invariant pseudonorm on $\widetilde{\mathrm{Ham}}(M, \omega)$.

REMARK 12.7. *As far as we are aware, it is unknown whether $\widetilde{d}_{\mathrm{Hof}}$ is nondegenerate for all closed symplectic manifolds (M, ω). Note that this does not follow directly from nondegeneracy of d_{Hof}, as in particular we would need to establish $\widetilde{d}(1, \phi) > 0$ when ϕ is a nontrivial loop in $\mathrm{Ham}(M, \omega)$ based at $\mathbb{1}$.*

LEMMA 12.8. *For any $\phi \in \widetilde{\mathrm{Ham}}(M, \omega)$ we have*

$$q_c(\phi) \leq 2c(\mathbb{1}) + \widetilde{d}_{\mathrm{Hof}}(\mathbb{1}, \phi).$$

Let $F : M \times I \to \mathbb{R}$ be a normalized time dependent Hamiltonian such that $[\{Fl_F^t\}] = [\{\phi_t\}] \in \widetilde{\mathrm{Ham}}(M, \omega)$. By the stability axiom, with $\psi = \mathbb{1}$ and $G \equiv 0$, we have

$$c(\phi) - c(\mathbb{1}) \leq \int_0^1 \max(F_t) dt.$$

Recall that ϕ^{-1} is generated by $\{-F_t \circ \phi_t\}$, and hence we similarly have

$$c(\phi^{-1}) - c(\mathbb{1}) \leq \int_0^1 \max(-F_t \circ \phi_t) dt = -\int_0^1 \min(F_t) dt.$$

Adding these gives

$$q(\phi) = c(\phi) + c(\phi^{-1}) \leq 2c(\mathbb{1}) + \int_0^1 \left(\max(F_t) - \min(F_t)\right) dt = 2c(\mathbb{1}) + \int_0^1 ||F_t|| dt,$$

and hence $q(\phi) \leq 2c(\mathbb{1}) + \widetilde{d}_{\mathrm{Hof}}(\mathbb{1}, [\{\phi_t\}])$.

PROOF OF LEMMA A. For any $\phi = [\{\phi_t\}] \in \widetilde{\mathrm{Ham}}(M, \omega)$ such that $\phi_1(U) \cap U = \varnothing$, we have by the previous lemma and the definition of $e_c(U)$

$$e_c(U) \leq q_c(\phi) \leq 2c(\mathbb{1}) + \widetilde{d}_{\mathrm{Hof}}(\mathbb{1}, \phi) \leq 2c(\mathbb{1}) + d_{\mathrm{Hof}}(\mathbb{1}, \phi_1).$$

Since $\phi_1 \in \mathrm{Ham}(M, \omega)$ is an arbitrary Hamiltonian diffeomorphism which displaces U, we have by the definition of $e_{\mathrm{Hof}}(U)$

$$e_c(U) \leq 2c(\mathbb{1}) + e_{\mathrm{Hof}}(U).$$

$\square$

The Calabi homomorphism

13.1. Definition and invariance

We next seek to prove Lemma B. Before doing so, let us introduce the Calabi homomorphism.

DEFINITION 13.1. *Let (M, ω) be a noncompact symplectic manifold, let $H : M \times I \to \mathbb{R}$ be a time-dependent Hamiltonian, and let $\phi \in \widetilde{\mathrm{Ham}}(M, \omega)$ denote its flow. We define the* **Calabi invariant** *by*

$$\mathrm{Cal}(H) := \int_{t=0}^{t=1} \left(\int_M H_t \omega^{\wedge n} \right) dt \in \mathbb{R}.$$

PROPOSITION 13.2. $\mathrm{Cal}(H)$ *depends only on $\phi \in \widetilde{\mathrm{Ham}}(M, \omega)$ both not on H. Namely, if $F, G : M \times I \to \mathbb{R}$ both generate the same element of $\widetilde{\mathrm{Ham}}(M, \omega)$, then we have $\mathrm{Cal}(F) = \mathrm{Cal}(G)$. Moreover, the induced map $\mathrm{Cal} : \widetilde{\mathrm{Ham}}(M, \omega) \to \mathbb{R}$ is a group homomorphism.*

REMARK 13.3. *Note that there is no analogue of* Cal *when (M, ω) is a symplectic manifold, since in this case after normalizing we would just get $\mathrm{Cal}(H) = 0$.*

PROOF OF PROPOSITION 13.2. Let $F, G : M \times I \to \mathbb{R}$ be time-dependent Hamiltonians with flows $\{\phi_t\}, \{\psi_t\}$ respectively, such that $[\{\phi_t\}] = [\{\psi_t\}] \in \widetilde{\mathrm{Ham}}(M, \omega)$. As in Corollary 11.8, let $\gamma_{s,t}$ be a 2-parameter family in $\mathrm{Ham}(M, \omega)$ such that $\gamma_{0,t} = \phi_t$ and $\gamma_{1,t} = \psi_t$, and $\gamma_{s,i}$ is s-independent for $i = 0, 1$.

and let $A_{s,t}, B_{s,t}$ be 2-parameter families of compactly supported functions $M \to \mathbb{R}$ such that we have

$$\partial_t \gamma_{s,t}(p) = X_{A_{s,t}}(\gamma_{s,t}(p))$$
$$\partial_s \gamma_{s,t}(p) = X_{B_{s,t}}(\gamma_{s,t}(p))$$

for all $p \in M$ (c.f. Banyaga's theorem). According to Corollary 11.8, we have

$$\partial_s A_{s,t} - \partial_t B_{s,t} = -\{A_{s,t}, B_{s,t}\}.$$

Note that we have $A_{0,t} = F$, $A_{1,t} = G$, and $B_{s,0} = B_{s,1} = 0$.

It suffices show that the quantity

$$\int_0^1 \int_M A_{s,t} \omega^{\wedge n} dt \in \mathbb{R}$$

is independent of s. Using Exercise 10.4, we compute

$$\partial_s \int_0^1 \int_M A_{s,t} \omega^{\wedge n} dt = \int_0^1 \int_M (\partial_t B_{s,t} - \{A_{s,t}, B_{s,t}\}) dt$$

$$= n \int_0^1 \int_M dA_{s,t} \wedge dG_{s,t} \wedge \omega^{\wedge (n-1)} dt,$$

and the latter vanishes since $dA_{s,t} \wedge dG_{s,t} \wedge \omega^{\wedge(n-1)}$ is an exact $2n$-form on M.

$\square$

Now suppose that (M, ω) is closed, and let $U \subset M$ be an open subset. If $H : M \times I \to \mathbb{R}$ is supported in U, we can view it as a time-dependent Hamiltonian on the noncompact symplectic manifold $(U, \omega|_U)$, and as such we denote its Calabi invariant by $\mathrm{Cal}_U(H)$. We thus get a well-defined homomorphism

$$\mathrm{Cal}_U : \widetilde{\mathrm{Ham}}_U(M, \omega) \to \mathbb{R}.$$

13.2. A lemma and the proof of Lemma B

As before, let (M, ω) be a closed symplectic manifold and let $c : \widetilde{\mathrm{Ham}}(M, \omega) \to \mathbb{R}$ be a subadditive spectral invariant.

LEMMA 13.4. *Let $U \subset M$ be an open subset which is displaceable, and let $F, G : M \times I \to \mathbb{R}$ be time-dependent Hamiltonians such that*

- F *is supported in U, and thus generates a flow $\{\phi_t\} \in \widetilde{\mathrm{Ham}}_U(M, \omega)$*
- G *generates a flow $\{\psi_t\} \in \widetilde{\mathrm{Ham}}(M, \omega)$ such that $\psi_1(U) \cap U = \varnothing$.*

Then we have

$$c(\psi\phi) = c(\psi) - \int_0^1 \langle F_t \rangle dt.$$

Here for $Q : M \to \mathbb{R}$ we put $\langle Q \rangle = \frac{\int_M Q\omega^{\wedge n}}{\int_M \omega^{\wedge n}}$.

PROOF OF LEMMA B. Consider any $\phi \in \widetilde{\mathrm{Ham}}_U(M, \omega)$, and suppose that $\psi \in \widetilde{\mathrm{Ham}}(M, \omega)$ displaces U. Then by Lemma 13.4 we have

$$c(\phi) = c(\psi^{-1}\psi\phi) \le c(\psi^{-1}) + c(\psi\phi)$$

$$= c(\psi^{-1}) + c(\psi) - \frac{\mathrm{Cal}_U(\phi)}{\int_M \omega^{\wedge n}}$$

$$= q(\psi) - \frac{\mathrm{Cal}_U(\phi)}{\int_M \omega^{\wedge n}}.$$

Combining this with the analogous bound for $c(\phi^{-1})$, and using the fact that $\mathrm{Cal}_U(\phi) = -\mathrm{Cal}_U(\phi^{-1})$, we have

$$q(\phi) = c(\phi) + c(\phi^{-1}) \le 2q(\psi),$$

and hence $q(\phi) \le 2e_c(U)$.

$\square$

PROOF OF LEMMA 13.4. We will explain the prove assuming that ψ_1 is nondegenerate. The general case then follows by approximating a degenerate Hamiltonian flow by a nondegenerate one and appealing to the stability axiom – see [**PR**, Proof of Prop. 4.3.5].

For $\tau \in [0, 1]$, let $\gamma^\tau \in \widetilde{\mathrm{Ham}}(M, \omega)$ be given by concatenating $\{\psi_t\}_{t \in [0,1]}$ with the path $\{\psi_1\phi_t\}_{t \in [0,\tau]}$, after reparametrizing time so that result takes place over $[0, 1]$ (rather than $[0, 1 + \tau]$). We seek to compare the action spectra of ψ_1 and $\gamma_1^\tau = \psi_1\phi_\tau$. Firstly, observe that we have $\mathrm{Fix}(\psi_1) = \mathrm{Fix}(\psi_1\phi_\tau)$. Indeed, any $p \in U$ cannot be fixed by either, since ψ_1 displaces U while ϕ_τ is supported in U. On the other hand, for $p \notin U$ we have $\psi_1\phi_\tau(p) = \psi_1(p)$.

Now consider some $p_0 \in \mathrm{Fix}(\psi_1) = \mathrm{Fix}(\psi_1\phi_\tau)$ such that the loop $\{\psi_t(p_0)\}_{t \in [0,1]}$ in M is contractible, and let D be a capping disk. Note that the loop $\{\gamma_t^\tau(p_0)\}_{t \in [0,1]}$ is, up to

reparametrizing time, given by the loop $\{\psi_t(p_0)\}_{t\in[0,1]}$ followed by a constant loop at p_0. Therefore we can also view D as a capping disk for $\{\gamma^\tau_t(p_0)\}_{t\in[0,1]}$, and for each $\tau \in [0,1]$ we have

$$\mathcal{A}_{\{\gamma^\tau_t\}_{t\in[0,1]}}(p_0, D) = \mathcal{A}_{\{\psi_t\}_{t\in[0,1]}}(p_0, D) + \int_{t=0}^{t=\tau} H_t(\psi\phi_t(p_0))dt,$$

where H is the *normalized* time-dependent Hamiltonian on M which generates the isotopy $\{\psi\phi_t\}_{t\in[0,\tau]}$. That is, we have

$$\partial_t\psi_1\phi_t(p) = X_{H_t}(\psi_1\phi_t(p))$$

for all $p \in M$ and $t \in [0, \tau]$, and hence $H_t = F_t \circ \psi_1^{-1}$ (c.f. (8.1)). Since $H_t = F_t - \langle F_t \rangle$ and F_t vanishes at $\phi_t(p_0) = p_0$, we have

$$\mathcal{A}_{\{\gamma^\tau_t\}_{t\in[0,1]}}(p_0, D) - \mathcal{A}_{\{\psi_t\}_{t\in[0,1]}}(p_0, D) = -\int_0^\tau \langle F_t \rangle dt,$$

whence

$$\mathrm{spec}(\gamma^\tau) = \mathrm{spec}(\psi) - \int_0^\tau \langle F_t \rangle dt.$$

Since ψ_1 is nondegenerate, it is easy to check that $\psi_1\phi_\tau$ is as well, so by the spectrality axiom we have

$$c(\gamma^\tau) \in \mathrm{spec}(\gamma^\tau),$$

or equivalently

$$c(\gamma^\tau) + \int_0^\tau \langle F_t \rangle dt \in \mathrm{spec}(\psi).$$

Since ψ is nondegenerate, it is easy to check that $\mathrm{spec}(\psi) \subset \mathbb{R}$ has no interior. As $c(\gamma^\tau)$ and $\int_0^\tau \langle F_t \rangle dt$ are both continuous functions of τ (convince yourself!), it follows that $c(\gamma^\tau) + \int_0^\tau \langle F_t \rangle$ is independent of τ. In particular, we have

$$c(\psi) = c(\gamma^0) = c(\gamma^1) + \int_0^1 \langle F_t \rangle dt = c(\psi\phi) + \int_0^1 \langle F_t \rangle dt$$

$\square$

REMARK 13.5. *In fact, for any (not necessarily nondegenerate)* $\psi \in \widetilde{\mathrm{Ham}}(M, \omega)$, $\mathrm{spec}(\psi) \subset \mathbb{R}$ *has measure zero – see* [**Oh**, *Lem. 2.2*].

Group pseudonorms, spectral width and the Poisson bracket inequality

14.1. Some geometric group theory

We begin by introducing some basic formalism some geometric group theory, for an abstract group G. In the next section we will directly apply these results in the case that $c : \widetilde{\mathrm{Ham}} \to \mathbb{R}$ is a subadditive spectral invariant. We mostly follow [**PR**, §3.5,§4.6].

DEFINITION 14.1. *A* **pseudo-norm** *on a group G is a function $\nu : G \to [0, \infty)$ such that, for any $g, h \in G$, we have:*

(i) (normalization) $\nu(e) = 0$ *(here $e \in G$ is the unit)*
(ii) (symmetry) $\nu(g) = \nu(g^{-1})$
(iii) (triangle inequality) $\nu(gh) \leq \nu(g) + \nu(h)$.

It is a **norm** *if moreover $\nu(g) > 0$ for whenever $g \neq e$.*

EXERCISE 14.2. *Let G be a group and let $c : G \to \mathbb{R}$ be a function such that for any $g, h \in G$ we have:*

(a) (conjugation invariance) $c(ghg^{-1}) = c(g)$
(b) (subadditivity) $c(gh) \leq c(g) + c(h)$.

Put $q(g) := c(g) + c(g^{-1})$ and $\nu(g) := \begin{cases} q(g) & g \neq e \\ 0 & g = e. \end{cases}$ Then ν is a pseudo-norm on G.

REMARK 14.3. *Note that for c as in Exercise 14.2, we have $c(e) = c(gg^{-1}) \leq c(g) + c(g^{-1}) = q(g)$ and in particular $c(e) \leq 2c(e)$, whence $0 \leq c(e)$. We therefore have $q(g) \geq 0$ for any $g \in G$.*

LEMMA 14.4. *For c as in Exercise 14.2, the limit $\sigma(g) := \lim_{k \to \infty} \frac{c(g^k)}{k}$ is defined for any $g \in G$.*

PROOF. Put $a_k := kc(g^{-1}) + c(g^k)$. We have

$$0 \leq q(g^k) = c(g^k) + c(g^{-k}) \leq a_k.$$

Also, note that we have

$$a_{k+l} \leq (k+l)c(g^{-1}) + c(g^k) + c(g^l) = a_k + a_l.$$

Since the seqence $\{a_k\}$ is subadditive, the limit $\lim_{k \to \infty} \frac{a_k}{k}$ exists by Fekete's lemma (see below). Since $\frac{a_k}{k} = c(g^{-1}) + \frac{c(g^k)}{k}$, it follows that the limit $\lim_{k \to \infty} \frac{c(g^k)}{k}$ also exists. $\qquad\square$

LEMMA 14.5 (Fekete's lemma). *Let $\{a_n\}_{n \in \mathbb{Z}_{\geq 1}}$ be a sequence of nonnegative real numbers which is subadditive, i.e. such that $a_{i+j} \leq a_i + a_j$ for all $i, j \in \mathbb{Z}_{\geq 1}$. Then the limit $\lim\limits_{k \to \infty} \frac{a_k}{k}$ exists and is equal to $\inf\limits_{k \geq 1} \frac{a_k}{k}$.*

PROPOSITION 14.6. *Let $c : G \to \mathbb{R}$ be as in Exercise 14.2, and let $\sigma : G \to \mathbb{R}$ be given by $\sigma(g) := \lim\limits_{k \to \infty} \frac{c(g^k)}{k}$ (this is well-defined by Lemma 14.4). Then for any $g, h \in G$ we have*

$$|\sigma(gh) - \sigma(g) - \sigma(h)| \leq \min(q(g), q(h)).$$

PROOF. For $g, h \in G$, observe that we have

$$(gh)^k = (gh) \cdots (gh) = (ghg^{-1})(g^2 hg^{-2}) \cdots (g^k hg^{-k})g^k = \theta g^k$$

where we put $\theta := (ghg^{-1})(g^2 hg^{-2}) \cdots (g^k hg^{-k})$. Using conjugation invariance and subadditivity, we have

$$c(\theta) \leq c(ghg^{-1}) + \cdots + c(g^k hg^{-k}) \leq kc(h).$$

Similarly, we have $\theta^{-1} = (g^k h^{-1} g^{-k}) \cdots (gh^{-1}g)$ and hence $c(\theta^{-1}) \leq kc(h^{-1})$. Therefore we have

$$c((gh)^k) \leq c(\theta) + c(g^k) \leq kc(h) + c(g^k).$$

Note that we have $0 \leq c(h^k) + c(h^{-k})$ and hence $-c(h^{-k}) \leq c(h^k)$. We then have

$$kc(h) = kq(h) - kc(h^{-1}) \leq kq(h) - c(h^{-k})$$
$$\leq kq(h) + c(h^k).$$

We therefore have

$$c((gh)^k) \leq kq(h) + c(h^k) + c(g^k),$$

i.e.

$$\frac{c((gh)^k)}{k} \leq q(h) + \frac{c(h^k)}{k} + \frac{c(g^k)}{k},$$

and hence

$$\sigma(gh) - \sigma(g) - \sigma(h) \leq q(h).$$

Swapping g and h and noting that $c(hg) = c(g^{-1}ghg) = c(gh)$, we also have

$$\sigma(gh) - \sigma(g) - \sigma(h) \leq q(g),$$

and hence

$$\sigma(gh) - \sigma(g) - \sigma(h) \leq \min(q(g), q(h)).$$

The proof that

$$\sigma(gh) - \sigma(g) - \sigma(h) \geq -\min(q(g), q(h))$$

is similar. $\square$

14.2. Poisson bracket inequality

Let (M, ω) be a closed symplectic manifold. Let $c : \widetilde{\mathrm{Ham}}(M, \omega) \to \mathbb{R}$ be a subadditive spectral invariant. Put $q_c(\phi) := c(\phi) + c(\phi^{-1})$.

DEFINITION 14.7. *The* **spectral width** *(with respect c) of an open subset $U \subset M$ is defined by*

$$w_c(U) := \sup\{q_c(\phi) \mid \phi \in \widetilde{\mathrm{Ham}}(M, \omega)\}.$$

For (not necessarily normalized) autonomous Hamiltonians $F, G : M \to \mathbb{R}$, let $\phi_F = \{\phi_F^t\}, \phi_G = \{\Phi_G^t\} \in \widetilde{\mathrm{Ham}}(M, \omega)$ denote their flows, and put

$$\Pi(F, G) := |\sigma(\phi_{F+G}) - \sigma(\phi_F) - \sigma(\phi_G)|.$$

Also, put

$$S(F, G) := \min(w_c(\mathrm{supp}\, F), w_c(\mathrm{supp}\, G)).$$

Here $\mathrm{supp}\, F, \mathrm{supp}\, G$ denote the *supports* of F, G respectively.

The following is the "Poisson bracket inequality", c.f. [**PR**, §4.6]:

PROPOSITION 14.8. *We have $\Pi(F, G) \leq \sqrt{2S(F, G)||\{F, G\}||}$.*

PROOF. By the Proposition 14.6, we have

$$|\sigma(\phi_F \phi_G)) - \sigma(\phi_F) - \sigma(\phi_G)| \leq \min(q(\phi_F), q(\phi_G))$$
$$\leq S(F, G).$$

Recall that by the stability axiom for c we have

$$c(\phi_F) - c(\phi_G) \leq \int_0^1 \max(F - G) dt.$$

EXERCISE 14.9. *We have also $\sigma(\phi_F) - \sigma(\phi_G) \leq \int_0^1 \max(F - G) dt$.*

Assume without loss of generality that we have $\langle F \rangle = \langle G \rangle = 0$. Then we have

$$\sigma(\phi_{F+G}) - \sigma(\phi_F \phi_G) \leq \int_0^1 \max(F + G - (F + G \circ (\phi_F^t)^{-1})) dt$$
$$\leq \int_0^1 ||G - G \circ (\phi_F^t)^{-1}|| dt$$
$$= \int_0^1 ||G \circ \phi_F^t - G|| dt.$$

Note that have

$$G \circ \phi_F^t - G = \int_{s=0}^{s=t} \partial_s (G \circ \phi_F^s) ds$$
$$= \int_0^t (dG \circ X_F) \circ (\phi_F^s) ds$$
$$= \int_0^t \{G, F\} \circ \phi_F^s ds$$
$$\leq t||\{G, F\}||,$$

and hence
$$|\sigma(\phi_{F+G}) - \sigma(\phi_F\phi_G)| \le \tfrac{1}{2}||\{G,F\}||.$$

We then have
$$\Pi(F,G) = |\sigma(\phi_{F+G}) - \sigma(\phi_F) - \sigma(\phi_G)| \le |\sigma(\phi_{F+G}) - \sigma(\phi_F\phi_G)| + |\sigma(\phi_F\phi_G) - \sigma(\phi_F) - \sigma(\phi_G)|$$
$$\le \tfrac{1}{2}||\{F,G\}|| + |\sigma(\phi_F\phi_G) - \sigma(\phi_F) - \sigma(\phi_G)|$$
$$\le \tfrac{1}{2}||\{F,G\}|| + S(F,G).$$

It follows that for any $s \in \mathbb{R}$ we have
$$\Pi(sF,sG) \le \tfrac{1}{2}||\{sF,sG\}|| + S(sF,sG).$$
Noting that each of these terms scale differently, this gives
$$\Pi(F,G) \le \tfrac{s}{2}||\{F,G\}|| + \frac{S(F,G)}{s}.$$

EXERCISE 14.10. *For $A,B \in \mathbb{R}$, th function $f(s) = sA + \frac{1}{s}B$ has minimum value $2\sqrt{A,B}$.*

By above the exercise, we have
$$\Pi(F,G) \le 2\sqrt{\tfrac{1}{2}||\{F,G]\}||S(F,G)} = \sqrt{2||\{F,G\}||S(F,G)},$$
which is the desired inequality. $\square$

We will not be making use of the Poisson bracket inequality except for the following corollary:

COROLLARY 14.11. *Suppose that we have $w_c(\operatorname{supp} F) = 0$ or $w_c(\operatorname{supp} G) = 0$. Then we have $\sigma(\phi_{F+G}) = \sigma(\phi_F) + \sigma(\phi_G)$.*

14.3. Proof of Lemma C

Our goal is to prove that for $\phi = [\{\phi_t\}] \in \widetilde{\operatorname{Ham}}(M,\omega)$ such that $\phi^1 \ne \mathbb{1}$, we have $q_c(\phi) > 0$.

LEMMA 14.12. *For $U \subset M$ open and displaceable and $\phi \in \widetilde{\operatorname{Ham}}_U(M,\omega)$, we have $\sigma(\phi) = -\frac{\operatorname{Cal}_U(\phi)}{\int_M \omega^{\wedge n}}$*

PROOF. Recall that by Lemma 13.4, if $\psi \in \widetilde{\operatorname{Ham}}(M,\omega)$ displaces U, then we have
$$c(\psi\phi) = c(\psi) - \frac{\operatorname{Cal}_U(\phi)}{\int_M \omega^{\wedge n}},$$
in which case we have
$$c(\phi) = c(\psi^{-1}\psi\phi) \le c(\psi\phi) + c(\psi^{-1})$$
$$= c(\psi) - \frac{\operatorname{Cal}_U(\phi)}{\int_M \omega^{\wedge n}} + c(\psi^{-1})$$
$$= q_c(\psi) - \frac{\operatorname{Cal}_U(\phi)}{\int_M \omega^{\wedge n}},$$
and therefore
$$c(\phi) \le e_c(U) - \frac{\operatorname{Cal}_U(\phi)}{\int_M \omega^{\wedge n}}.$$

Similarly, we have

$$c(\phi) \geq c(\psi\phi) - c(\psi) = -\frac{\mathrm{Cal}_U(\phi)}{\int_M \omega^{\wedge n}}.$$

Recalling that Cal_U is a homomorphism, we have then

$$-\frac{\mathrm{Cal}_U(\phi)}{\int_M \omega^{\wedge n}} \leq \frac{c(\phi^k)}{k} \leq \frac{e_c(U)}{k} - \frac{\mathrm{Cal}_U(\phi)}{\int_M \omega^{\wedge n}}$$

which immediately gives the desired result after taking the limit $k \to \infty$. $\qquad\square$

PROOF OF LEMMA C. Suppose by contradiction that we have $q_c(\phi) = 0$. Let $U \subset M$ be an open subset which is displaced by ϕ. Note that we have $e_c(\phi) = 0$.

EXERCISE 14.13. *We can find $\theta_1, \ldots, \theta_N \in \mathrm{Ham}(M, \omega)$ such that $\{\theta_i(U)\}$ is a cover of M.*

Let $\{\beta_i\}$ be a partition of unity subordinate to the cover $\{\theta_i(U)\}$, i.e. $\mathrm{supp}\,\beta_i \subset \theta_i(U)$ and $\sum_{i=1}^N \beta_i = 1$. We also have $e_c(\theta_i(U)) = e_c(U)$ for $i = 1, \ldots, N$ (convince yourself). Then we have

$$\omega(\mathrm{supp}\,\beta_i) \leq \omega(\theta_i(U)) \leq 2e_c(\theta_i(U)) = 0,$$

where the second inequality follows from Lemma B. We then have

$$
\begin{aligned}
0 = \sigma(\phi_{\beta_1 + \cdots + \beta_N}) &= \sigma(\phi_{\beta_1}) + \sigma(\phi_{\beta_2 + \cdots + \beta_N}) \\
&= \ldots \\
&= \sigma(\phi_{\beta_1}) + \cdots + \sigma(\phi_{\beta_N}) \\
&= -\frac{\sum_{i=1}^N \mathrm{Cal}_{\theta_i(U)}(\phi_{\beta_i})}{\int_M \omega^{\wedge n}} \\
&= \frac{-\sum_{i=1}^N \int_M \beta_i \omega^{\wedge n}}{\int_M \omega^{\wedge n}} \\
&= -1,
\end{aligned}
$$

which is a contradiction.

$\qquad\square$

Bibliography

[Ban]   Augustin Banyaga. Sur la structure du groupe des difféomorphismes qui préservent une forme symplectique. *Commentarii Mathematici Helvetici* **53**(1978), 174–227.

[BL]   Ulrich Bauer and Michael Lesnick. Induced matchings of barcodes and the algebraic stability of persistence. pages 355–364, 2014.

[BO]   Lev Buhovsky and Yaron Ostrover. On the uniqueness of Hofer's geometry. *Geometric and Functional Analysis* **21**(2010), 1296–1330.

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

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

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

[Hir]   Morris W Hirsch. *Differential topology*. Springer Science & Business Media, 2012.

[Hof]   Helmut Hofer. On the topological properties of symplectic maps. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics* **115**(1990), 25–38.

[LM]   François Lalonde and Dusa McDuff. The geometry of symplectic energy. *Annals of Mathematics* **141**(1995), 349–371.

[Oh]   Yong-Geun Oh. Chain level Floer theory and Hofer's geometry of the Hamiltonian diffeomorphism group. *Asian J. Math.* **6**(2002), 579–624.

[Ono]   Kaoru Ono. Floer–Novikov cohomology and the flux conjecture. *Geometric & Functional Analysis GAFA* **16**(2006), 981–1020.

[Pol]   Leonid Polterovich. Symplectic displacement energy for Lagrangian submanifolds. *Ergodic Theory and Dynamical Systems* **13**(1993), 357–367.

[PR]   Leonid Polterovich and Daniel Rosen. *Function theory on symplectic manifolds*. American Mathematical Soc., 2014.

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

[ZC]   Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry* **33**(2005), 249–274.