Symplectic Integrators and Persistent Homology

Linden Disney-Hogg

Contents

1 Introduction

I am going to assume you are familiar with the Arnold-Liouville Theorem, so we will jump in to the two main topics and then hopefully link them together. We will state the results we want to know from Hamiltonian dynamics: we consider a dynamical system given by Hamilton's equations

$$\dot{q} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial q}$$

for some Hamiltonian $H = H(q, p)$. Note this can be written as

$$\dot{y} = J\nabla H = \{y, H\}$$

where $y = (q, p)^T$ and $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$. If we introduce the operator D_H by $D_H f = \{f, H\}$ we have that the formal solution to the ODE is

$$y(t_0 + t) = \exp(t D_H) y(t_0)$$

when H is autonomous.

> **Lemma 1.1.** *The Hamiltonian is conserved, i.e $\dot{H} = 0$.*

> **Lemma 1.2.** *The symplectic form $dq \wedge dp$ is preserved under time evolution.*

We will also be using a simple specific example to illustrate some points, and this will be the case of the simple harmonic oscillator with

$$H(q, p) = \frac{1}{2}(p^2 + q^2)$$

This has the known exact solution

$$\begin{pmatrix} q(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} q(0) \\ p(0) \end{pmatrix}$$

We may easily verify the above lemmas in this case.

2 Symplectic Integrators

> **Remark.** *The content of this section is covered in [12]*

For our purpose an **integrator** will be a numerical method to approximate the solution to an ODE.

Definition 2.1. *Given a 1st order ODE $\dot{y}(t) = f(t, y(t))$, the **Euler method** to approximate the solution with initial condition $y(t_0) = y_0$ is*

$$y_{n+1} = y_n + hf(t_n, y_n),\ t_n = t_0 + nh,\ y(t_n) \approx y_n$$

for some h taken to be small.

Remark. *For our current applications, we will only by considering autonomous Hamiltonians, so we can wlog take $t_0 = 0$*

Example 2.2. *Let us see how the Euler method applies to the SHO:*

$$\begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & h \\ -h & 1 \end{pmatrix} \begin{pmatrix} q_n \\ p_n \end{pmatrix}$$

We can see that

$$H(q_{n+1}, p_{n+1}) = (1 + h^2)H(q_n, p_n)$$
$$dq_{n+1} \wedge dp_{n+1} = (1 + h^2)(dq_n \wedge dp_n)$$

(the point is that the matrix is not symplectic) and so we do not get the nice properties we want from our Hamiltonian system. It is possible to rectify the problems in the Euler method (at least for the symplectic form) if we change the integrator to

$$\begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & h \\ -h & 1 - h^2 \end{pmatrix} \begin{pmatrix} q_n \\ p_n \end{pmatrix}$$

More generally if we have a Hamiltonian $H(q, p) = T(p) + V(q)$ we can make an integrator that preserves the symplectic form by

$$q_{n+1} = q_n + hT'(p_n),\ p_{n+1} = p_n - hV'(q_{n+1})$$

This work because it is the composition of two symplectic maps.

Remark. *You might wonder how you would come up with such a transform. One idea is to take a generating function for the transform $(q, p) \mapsto (Q, P)$ given by*

$$W(q, P) = qP + hH(q, P)$$

(think HJE). Then the equations for the generating function are

$$Q = \frac{\partial W}{\partial P} = q + h \left.\frac{\partial H}{\partial P}\right|_{(q,P)}$$

$$p = \frac{\partial W}{\partial q} = P + h \left.\frac{\partial H}{\partial q}\right|_{(q,P)}$$

Alternatively, note that because the Poisson bracket is linear we have $D_H = D_T + D_V$. $\exp(tD_T), \exp(tD_V)$ are both symplectic transforms, and so by extension so is $\prod_{i=1}^{k} \exp(c_i tD_t) \exp(d_i tD_V)$ for $c_i, d_i \in \mathbb{R}$. If we can choose c_i, d_i s.t

$$S(t) = \prod_{i=1}^{k} \exp(c_i tD_t) \exp(d_i tD_V) = \exp(t(D_T + D_V) + \mathcal{O}(t^{n+1}))$$

then this transform approximates time evolution. Furthermore, these constituent transforms can be calculated in a closed form as

$$q_{i+1} = q_i + c_{i+1} tT'(p_i),\ p_{i+1} = p_i - d_{i+1} tV'(q_{i+1})$$

We see that the simplest approximation is when $k = 1$ and we have $c_1 = 1 = d_1$. Another case that we may use is that of $k = 2$, $c_1 = \frac{1}{2} = c_2$, $d_1 = 1$, $d_2 = 0$. Calculating the c, d coefficients is hard without using the techniques of [11].

We have the related theorem:

Theorem 2.3 (Yoshida, 1990)**.** *The above integrator describes the exact time evolution under the modified Hamiltonian*

$$\tilde{H}(h) = T + V + \frac{h}{2}\{V, T\} + \frac{h^2}{12}\left(\{\{T, V\}, V\} + \{\{V, T\}, T\}\right) + \dots$$

This actually comes from Baker-Campbell-Hausdorff.

Proof. Apply BCH which says that $\exp X \exp Y = \exp Z$ for

$$Z = X + Y + \frac{1}{2}[X, Y] + \dots$$

and then using $[D_f, D_g] = D_{\{f,g\}}$ $\qquad\qquad\qquad\square$

The important thing about this theorem is that, because $\tilde{H}$ is conserved, the error terms must be non-secular, which is non-generic and tells us that the error in H is indeed $\mathcal{O}(h)$.

Remark. *We might wonder if it is possible to construct an integrator that preserves the energy as well as the symplectic form. It is a result from [14] and [2] that if H has no other conserved quantities then this is not possible.*

Example 2.4. *With the SHO, we find*

$$\tilde{H}(h) = \frac{1}{2}(p^2 + q^2) + \frac{h}{2}pq + \frac{h^2}{12}(p^2 + q^2) + \frac{h^3}{12}pq + \dots$$

In this case this actually continues and we have that

$$\frac{1}{2}(p^2 + q^2) + \frac{h}{2}pq$$

is a constant of motion.

Remark. *If we do not have Hamiltonian which separates into $T + V$ we are not immediately ruled out, as we may use the modified*

$$H(q, \xi) + H(\eta, p) + \frac{1}{2}\omega\left(|q - \eta|^2 + |p - \xi|^2\right)$$

introduce by Tao [10]. The results introduced on how to make sure that this new system well approximates the last requires H be integrable, which will be the paradigm we consider, and the closeness can be improved by making ω larger. More specifically, an lth order integrator has global error $\mathcal{O}(Th^l\omega)$ up to time $T = \mathcal{O}(\min(\delta^{-l}\omega^{-1}, \omega^{0.5}))$
This is then helped by another result:

Lemma 2.5. *If the system is integrable, then the global error of the symplectic grows linearly for exponentially long times. See [9].*

There are alternatively implicit methods that do not require the Hamiltonian to separate, but I will not consider these. To see more on this see [13].

We have another useful property of symplectic integrators that relates to KAM theory:

Theorem 2.6 (KAM informally). *If we perturb the Hamiltonian of an integrable system in a sufficiently nice way, the invariant tori remain.*

Theorem 2.7 (SIA and KAM informally). *Symplectic integrators for small enough time step keep invariant tori*

3 Persistent Homology

We now introduce the story of what is the persistent homology and how to calculate it. We will largely follow [6]

The first ingredient we want will be some data X which forms a metric space with distance d. We will take some finite sample $S \subset X$).

Remark. *For our purposes we will be considering $X \subset \mathbb{R}^N$ and take d to be the Euclidean distance.*

We want to try to detect some underlying structure of the data

Example 3.1. *If we took some $h \ll 1$ and data*

$$S = \left\{ x_k = \begin{pmatrix} \cos kh \\ -\sin kh \end{pmatrix} \mid k = 1, \dots, \left\lceil \frac{2\pi}{h} \right\rceil \right\} \subset X = S^1 \subset \mathbb{R}^2$$

then we may want to somehow identify the circle these data points are laying on.

To make this question specific, we are going to focus on **homology inference**, that is can we work out the homology groups of X from a sample S. To do this we will introduce some constructions:

Definition 3.2. *Let $\mathcal{U} = \{U_i\}_{i \in I}$ be an open cover of X. Then the **nerve** of $\mathcal{U}$ is*

$$N(\mathcal{U}) = \left\{ J \subseteq I \mid \bigcap_{j \in J} U_j \neq \emptyset \right\}$$

This is a simplicial complex with vertices I and simplices $\{i_0, \dots, i_k\}$ when $\cap_{m=1}^k U_{i_m} \neq \emptyset$.

Aside. *If you haven't come across simplices before, here I will give the Clark Barwick version:*

> **Definition 3.3.** *We define the category Δ with objects given by totally ordered finite sets*
>
> $$[n] = \{0 < 1 < \cdots < n\}$$
>
> *and morphisms given by non-decreasing maps.*

> **Definition 3.4.** *A **simplicial set** if a functor $X : \Delta^{op} \to \mathbf{Set}$.*

Theorem 3.5 (Nerve Theorem [3]). *Let $\mathcal{U}$ be an open cover of X s.t. each $U \in \mathcal{U}$ is contractible and every non-empty intersection is contractible, then $N(\mathcal{U})$ is homotopic to $\cup_{U \in \mathcal{U}} U$*

This motivates the following definition:

Definition 3.6. *The **Čech complex** on S at scale ϵ is is*

$$\check{C}_\epsilon(S) = N\left(\{B_\epsilon(x) \mid x \in S\}\right)$$

Lemma 3.7. $\epsilon < \epsilon' \Rightarrow \check{C}_\epsilon \subset \check{C}_{\epsilon'}$.

This is a complex that should, intuitively, give us the right homology for X if we choose a good enough sample S. We can make precise what we mean about "good enough" with the following definition and result:

Definition 3.8. *The **injectivity radius** of a manifold X at $y \in X$ is*

$$inj(y) = \sup \{r \mid \exp : B_r \to X \text{ is injective}\}$$

The injectivity radius of the manifold is the inf of the injectivity radius at each point.

The injectivity radius has a natural interpretation given by the following lemma:

Lemma 3.9. *The norm of the second fundamental form is bounded in all directions by $\frac{1}{\tau}$.*

Theorem 3.10. *Let $X \subset \mathbb{R}^N$ be a smooth compact submanifold with injectivity radius τ and S a sample with sufficient density s.t.*

$$\max_{x \in S, y \in X} d(x, y) < \frac{\epsilon}{2}$$

for $\epsilon < \sqrt{\frac{3}{5}}\tau$ then $\check{C}_{2\epsilon}(S)$ deformation retracts onto X.

Proof. See [4]. $\square$

This is good because while the homology of a generic topological space is hard to compute, it can be computed quite easily for simplicial complexes. Let us see how this is done in the case of integral homology (cf later when we do homology over a field):

Let Y be the simplicial complex associated to S, then $C_n(Y)$ be the free abelian group with the basis given by the m_n singular n-simplices. We construct te **standard matrix representation** of $\partial_n : C_n(Y) \to C_{n-1}(Y)$, M_n, as the $m_{n-1} \times m_n$ matrix with entries in $\{0, \pm 1\}$ showing given the action of ∂_n on the basis elements. We use smith normal form algorithms (which we can do as $\mathbb{Z}$ is a Euclidean domain) to find

$$\tilde{M}_n = \begin{pmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_{l_n} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

where $d_i | d_{i+1}$, $l_n = \text{rank}(M_n)$. This smith normal form reduction also gives us the corresponding basis wrt which M_n is diagonal. The collection of matrices $\left\{\tilde{M}_n\right\}$ gives us all the information we need to work out $H_\bullet(Y)$ as

- The torsion coefficients of $H_{p-1}(Y)$ are given by the d_i of $\tilde{M}_p$

- The rank of $H_p(Y)$ is $\beta_p = (m_p - l_p) - l_{p+1}$.

The problem we have now is that, without knowing properties of X beforehand, we don't know what ϵ to choose. The idea then this - to increase ϵ to get a **filtered simplicial complex**, compute the homology at each step, and see how it changes.

Definition 3.11. *The pth **persistent homology** of the filtered simplicial complex $K = K_1 \subset \cdots \subset K_l$ is the pair $(\{H_p(K_i) \mid 1 \le i \le l\}, \{f_{ij} \mid 1 \le i \le j \le l\})$ where $f_{ij} : K_i \hookrightarrow K_j$ is the inclusion map.*

Remark. *We assigning a filtered simplicial complex to our data, we usually include a $K_0 = S$*

Aside. *Here we will actually want to be considering homology with coefficients in some field k. This is quite technical so I will not cover it in full, but I will lay out some of the important points.*

> **Lemma 3.12.** *Any abelian group A can be viewed as a $\mathbb{Z}$-module by*
> $$n \cdot a = \underbrace{a + \cdots + a}_{n \text{ times}}$$

Now we think of integral homology, or homology with $\mathbb{Z}$ coefficients as a functor

$$H_p(\cdot) : \boldsymbol{Top} \overset{C(\cdot)}{\to} \boldsymbol{Ch}_{\geq 0}(\mathbb{Z}) \overset{H_p}{\to} \boldsymbol{Ab}$$

where $\boldsymbol{Top}$ is the category of topological spaces, $\boldsymbol{Ch}_{\geq}(\mathbb{Z})$ is the category of chain complexes of abelian group, where the $\mathbb{Z}$ demonstates that we are thinking of the abelian groups as $\mathbb{Z}$-modules, and $\boldsymbol{Ab}$ is the category of abelian groups.

To extend to homology with coefficients we define $C(X; A)$ to be the chain complex with $C_k(X; A) = C_k(X) \otimes A$ and chain maps $\partial \otimes \text{id}_A$. This gives us

$$H_p(\cdot; A) : \boldsymbol{Top} \overset{C(\cdot)}{\to} \boldsymbol{Ch}_{\geq 0}(\mathbb{Z}) \overset{\otimes A}{\to} \boldsymbol{Ch}_{\geq 0}(\mathbb{Z}) \overset{H_p}{\to} \boldsymbol{Ab}$$

We now have the following lemma:

> **Lemma 3.13.** *Let A be an abelian group, R a commutative ring, and M an R-module, then the map*
> $$R \times A \times M \to A \times M$$
> $$(r, a, m) \mapsto (a, rm)$$
> *induces an R-module structure on $A \otimes M$, defining a functor $\otimes : \boldsymbol{Ab} \times R\boldsymbol{Mod} \to R\boldsymbol{Mod}$.*

This lets us define homology with module-valued coefficients as

$$H_p(\cdot; M) : \boldsymbol{Top} \overset{C(\cdot)}{\to} \boldsymbol{Ch}_{\geq 0}(\mathbb{Z}) \overset{\otimes M}{\to} \boldsymbol{Ch}_{\geq 0}(R) \overset{H_p}{\to} \boldsymbol{Ab}$$

An important example is the following:

> **Example 3.14.** *When $R = k$ is a field and we take the module $M = k$ to get that H_p is a functor into vector spaces over k.*

We have an important theorem that relates integral homology to homology with coefficients:

> **Theorem 3.15** (Universal Coefficient Theorem). *Let X be a topological space and A an abelian group. Then there is a SES*
> $$0 \to H_p(X; \mathbb{Z}) \otimes A \to H_p(X; A) \to Tor(H_{p-1}(X; \mathbb{Z}), A) \to 0$$

This Tor functor is far too complicated, but all we need to know about it is the following:

> **Lemma 3.16.** *Let B be an abelian group and $p \geq 0$. Then*
> $$Tor(B, \mathbb{Z}_p) \cong \ker(p) = \{b \in B \mid p \cdot b = 0\}$$
> *where we are viewing B as a $\mathbb{Z}$-module.*

> **Remark.** *We can see that the choice of coefficients does affect the homology, and so this choice is important in calculating persistent homology. For more information see [5]*

The persistent homology is an example of a more general structure:

Definition 3.17. *A **persistence module** $\mathcal{M}$ is a faimly of R-modules M_i with homomorphisms $\phi_i :$ $M_i \to M_{i+1}$.*

Definition 3.18. *A persistence module is of **finite type** if each M_i is finitely generated and $\exists m$ s.t $\forall i \geq m$ the ϕ_i are isomorphisms.*

Aside. *We now know that we are going to want to be dealing with homology which is going to be some R-module, so we want some structure theorems. This aside comes mainly from [15].*

Definition 3.19. *A **graded ring** is a ring R with decomposition $R = \bigoplus_{i \in \mathbb{Z}} R_i$ s.t multiplication is a bilinear $R_n \otimes R_m \to R_{n+m}$.*

Example 3.20. *The polynomial ring $R[t]$ is naturally graded by the degree, i.e. $R_i = R \cdot t^i$*

Definition 3.21. *A **graded module** over a graded ring R is an R-module M s.t $M = \bigoplus_{i \in \mathbb{Z}} M_i$ and the action of R is via $R_n \otimes M_m \to M_{n+m}$.*

Theorem 3.22. *Let R be a PID. Then every finitely-generated R-module has a unique decomposition*

$$M \cong R^\beta \oplus \left(\bigoplus_{i=1}^m {}^R\!\big/_{(d_i)} \right)$$

where $d_i \in R, \beta \in \mathbb{Z}$ s.t. $d_i | d_{i+1}$.
If R is now a graded ring, and M a graded module, then M has a unique decomposition

$$M \cong \left(\bigoplus_{i=1}^n \Sigma^{\alpha_i} R \right) \oplus \left(\bigoplus_{j=1}^m \Sigma^{\gamma_j} {}^R\!\big/_{(d_i)} \right)$$

where $\alpha_i, \gamma_j \in \mathbb{Z}$ and Σ^α represents an α-shift upwards in degree.
*In both the d_i are called the **torsion coefficients** as the quotient part is the torsion submodule.*

Think back to how this tells us the structure of our homology. It will now tell us about or persistence homology more generally.

Theorem 3.23. *Given a persistence module $\mathcal{M}$, define the graded $R[t]$-module*

$$\alpha(\mathcal{M}) = \bigoplus_{i \geq 0} M_i$$

with action of t as $t \cdot (m_0, m_1, m_2, \dots) = (0, \phi_0(m_0), \phi_1(m_1), \dots)$. Then the map α defines an equivalence of categories between persistence modules of finite type over R and finitely generated, non-negatively graded $R[t]$-modules.

We now know that understanding persistence modules must correspond to understanding finitely-generated graded $R[t]$-modules. We know how to do this only when $R[t]$ is a PID, so we must restrict to the case $R = k$ a field to be able to make any more theoretical progress, by the following result:

Proposition 3.24. *$R[t]$ a PID $\Leftrightarrow$ R a field.*

Remark. *When we have set our ring as a field k there is an alternative way of thinking of a persistence module as a functor $(\mathbb{R}, \geq) \to \mathrm{Vect}_k$ from the real numbers as a posetal category. Morphisms between*

We can now introduce the objects we need.

Definition 3.25. *A $\mathcal{P}$-interval is an ordered pair (i, j) with $0 \leq i < j \in \mathbb{Z}^\infty$.*

To a $\mathcal{P}$-interval we can associate a graded $k[t]$-module with a map Q by

$$Q(i, j) = \Sigma^i \cdot {k[t]}/{(t^{j-i})}$$

$$Q(i, \infty) = \Sigma^i k[t]$$

and so to a set $\mathcal{S} = \{(i_1, j_1), \ldots, (i_k, j_k)\}$ we have

$$Q(\mathcal{S}) = \bigoplus_{l=1}^{k} Q(i_l, j_l)$$

The map Q is in fact a bijection, so we have the corollary:

Corollary 3.26. *Isomorphism classes of persistence modules biject with finite sets of $\mathcal{P}$-intervals.*

We now think about the implication of this bijection when our persistence module is the pth persistent homology of a filtered simplicial complex K. This bijection is s.t each $\mathcal{P}$ interval (i, j) corresponds to a basis element of the homology that appears in K_i, remains until K_{j-1}, and then disappears in K_j.

We may now aim to calculate the persistent homology through the calculation of the $\mathcal{P}$-intervals, and the algorithm to compute the latter is more simple:

Let $\{e_j\}, \{\hat{e}_i\}$ represent the standard, homogeneous in the grading sense, bases of $C_n(Y), C_{n-1}(Y)$ now. $\deg e$ is the index of the K_i in which it first appears, and we can work out that in order to have M_n respect the grading it must be the case that

$$\deg e_j = \deg\left(\sum_i (M_n)_{ij}\hat{e}_i\right)$$

$$\Rightarrow \forall i,\ \deg e_j = \deg(M_n)_{ij}\hat{e}_j = \deg(M_n)_{ij} + \deg \hat{e}_i$$

$$\Rightarrow (M_n)_{ij} = (-1)^{(\ldots)} t^{\deg e_j - \deg \hat{e}_j} \quad \text{(exponent of -1 not specified)}$$

We want to reduce this matrix to identify the homology, but to make calculations of the $\mathcal{P}$-intervals we want to find a homogeneous basis that reduces the matrix (i.e. a homogeneous basis for $\ker \partial_n$)

Remark. *This point is quite key, as it is the reason that we cannot do **persistent** homology in $\mathbb{Z}[t]$. It is the identification of the same basis of cycles that persists through the skeleton, giving us birth and death times.*

We will find this basis by induction, assuming we have a homogeneous basis for $\ker \partial_{n-1}$. Certainly as $\partial_0 = 0$ we have our base case.

If $\{\hat{e}_i\}$ is the basis of $C_{n-1}(Y)$ that extends the homogeneous basis of $\ker \partial_{n-1}$, order the $\hat{e}_i$ in descending-degree order, represent ∂_n as M_n wrt the ordered basis $\hat{e}$ and the standard homogeneous basis of $C_n(Y)$, e. Apply Gaussian column elimination to get $\tilde{M}_n$, and the zero columns correspond to homogeneous elements of $\ker \partial_n$. It can be seen that this new basis is homogeneous as the column operations utilised all preserve homogeneity. We may then again do row operations that preserve homogeneity to make $\tilde{M}_n$ diagonal. This totals to give us the following result:

Proposition 3.27. *If in $\tilde{M}_n$, row i has pivot $(\tilde{M}_n)_{ij} = t^k$, then we get the $\mathcal{P}$-interval $(\deg \hat{e}_i, \deg \hat{e}_i + k)$ for H_{n-1}.*

The thing to notice here is that we didn't actually need to do the row operations, we only needed to know the pivot values.

Furthermore, as we know $M_n M_{n+1} = 0$, and that this is unchanged by elementary operations, we immediately know that all rows in M_{n+1} that correspond to pivot (i.e. non-zero) columns in M_n must be 0.

These two observations combined mean that we can heavily simplify the algorithm for computing persistence homology, by now computing all homology groups concurrently, incrementally. Let's describe this algorithm below in pseudo-code.

Let T be an array containing each simplex present in K, say $m \times 1$ dimensional, ordering the simplices by dimension, and within that by degree.

```
# input is K, a list of lists, where K[i] is the list of simplices contained
# in the ith complex in K. A simplex is a list of the vertices in the simplex
# If we label the 0-simplices, we get an orientated on simplces through the
# order on the elements of the simplex.
# We can extract the simplices and their degree from K
simplices = K[-1] # This is due to the inclusion we have
s = simplices
m = len(simplices)
l = len(K)
degrees = m*[0]
Kdim = max([len(s[i]) for i in range(len(s))])

for i in range(l-1):
        diff = list(Set(K[i+1]).difference(Set(K[i])))
        for j in range(len(diff)):
                ind = s.index(diff[j])
                degrees[ind] = i
```

```
# input is K, m, and s.
L = (K.dim()+1)*[[]] # We will store the P-intervals for H_k in L[k]
T = m*[[]]
marked = m*[0]
dims = [len(s[i]) for i in range(m)]

# s[j] will be the jth simplex.
for j in range(m):
        d = s[j].RemovePivotRows()
        # The method RemovePivotRows returns
        if (d==[[]]):
                Mark s[j] # marked[j]=1
        else:
                i = d.MaxIndex()
                # MaxIndex refers to the maximum index of the simplices
                # contained in d
                k = s[i].dim()
                # If we think of a simplex as a tuple, s.dim()=len(s)
                T[i] = [j,d]
                L[k] += [(s[i].deg(),s[j].deg())]

for j in range(m):
        if (s[j].is_marked() and T[j]==[]):
        #s[j].is_marked() == (marked[j]==1)
                k = s[j].dim()
                L[k] += [(s[j].deg(),infinity)]
```

We also describe the method `RemovePivotRows` used within:

```
# input is s, a given simplex
k = s.dim()
d = s.boundary() #i.e. [s[:i]+s[i+1:] for i in range(len(s))] over Z_2
d = [d[i] for i in range(len(d)) if marked[s.index(d[i])]==1]
while not (d==[[]]):
        i = d.MaxIndex() # max([s.index(d[j]) for j in range(len(d))])
        if (T[i]==[]):
                break
        unn = Set(d).union(Set(T[i][1]))
```

```
        inn = Set(d).intersection(Set(T[i][1]))
        elim = unn.difference(inn)
        d = list(elim)
```

RemovePivotRows first removes any rows that correspond to previously zero columns (i.e. columns where the boundary could be reduced to 0), and then with the remaining rows it does the Gaussian column elimination, using the knowledge of what the pivot element is (here is looks very simple using set differences as we are working in $\mathbb{Z}_2$).

4   Connecting the Dots

All that remains is to see how to knit the two together. The idea is a follows: upon being given a Hamiltonian system, we integrate it for a large number of steps with a symplectic integrator. This data should be a sample S of an invariant tori if the system is integrable (by our KAM theory) so we apple persistent homology to detect for this signature. Topological theory gives the following

Lemma 4.1. *For any abelian group A, $H_k(T^n, A) \cong A^{\binom{n}{k}}$*

This result means that we can expect to see the homology of a torus even when doing homology with coefficients.

Remark. *When doing this homology over field, because we will not detect any p-torsion (if we are working over $\mathbb{Z}_p$) we increase our false-positive ratio. There do now exists algorithms to calculate $\mathbb{Z}[t]$ persistent homology, and that is an approach we can investigate in the future. See [8].*

We can then take an array of different initial conditions, to see how the persistent homology varies as a function on phase space. It is then not impossible doing a hypothesis to see if the system is integrable. There is existing literature on hypothesis testing in the realm of persistent homology, see [7, 1].

One important feature we do need to tackle is the computational efficiency of such an approach. [6] considers the computational load of different complexes which can be formed in contrast to the Čech complex which are theoretical less intensive. One example which is implemented by the R package `TDAstats` is the **Vietoris-Rips complex**

Definition 4.2. *The **Vietoris-Rips Complex at scale ϵ on S** is*

$$VR_\epsilon(S) = \{\sigma \subseteq S \,|\, \forall x, y \in \sigma, \, d(x,y) \leq 2\epsilon\}$$

Such a complex ends up being easier to compute than the Čech complex, and approximates the Čech complex by the following lemma

Lemma 4.3. *If $S \subset \mathbb{R}^N$, then $\check{C}_\epsilon(S) \subseteq VR_\epsilon(S) \subseteq \check{C}_{\sqrt{2}\epsilon}(S)$.*

Aside. *It will be necessary later to understand exactly how the Vietoris-Rips complex is calculated, so we cover it here:*

Definition 4.4. *The ϵ-**neighbourhood graph** of S is the graph with vertices $V = S$ and edges*

$$E = \{(x,y) \in S \times S \,|\, x \neq y, \, d(x,y) \leq 2\epsilon\}$$

Definition 4.5. *The **clique complex** of a graph (V, E) is*

$$\{finite \; \sigma \subseteq V \,|\, \forall i, j \in \sigma, \, (i,j) \in E\}$$

We still realise a possible issue with the size of the information we need to store:

> **Lemma 4.6.** *The Vietoris-Rips complex of a sample S can contain at most $2^{|S|-1}$ simplices.*

This problem is slightly mitigated in our case as we know that $S \subset X \subset \mathbb{R}^N$, and lies on a codim-1 submanifold as we certainly know $\tilde{H}$ is conserved, so there cannot be any interesting homology H_p if $p \geq N$, so we need not compute simplices of order $p \geq N$. Hence the maximum number of simplices in our Vietoris-Rips complex is

$$\sum_{k=1}^{N-1} \binom{|S|}{k} = \mathcal{O}\left(|S|^{N-1}\right)$$

ss in all our applications, $|S| \gg N$. The calculation of the persistent homology is then $\mathcal{O}(m^3)$ if m is the number of simplices, so in the end our algorithm is $\mathcal{O}(|S|^{3(N-1)})$. If we are willing to increase our false positive rate, we can only take simplices of size up to n when looking for $T^n \subset \mathbb{R}^{2n} = \mathbb{R}^N$.

We can also make the following observation: Recall if we integrated out Hamiltonian system with time step h until time T we got a sample $S_h = \{x_i\}_{i=1}^M$ where $M = \frac{T}{h}$ and the approximation is s.t $x(nh) \approx x_n$. If instead we took timestep $\tilde{h} = \frac{h}{F}$ for some integer $F > 1$ then the sample $S_{\tilde{h}} = \{\tilde{x}_i\}_{i=1}^{MF}$ also approximates $x(nh) \approx \tilde{x}_{nF}$. If the integration scheme is of rth-order then the global error is $\mathcal{O}(h^{r-1})$, and so we see by taking the subset

$$\tilde{S} = \{\tilde{x}_{nF} \,|\, 1 \leq n \leq M\}$$

we keep $|S| = \left|\tilde{S}\right|$ but reduce the global error at $x(nh)$ by $\mathcal{O}(F^{1-r})$. Hence we can have our sample be as accurate as desirable, while keeping the density the same.

References

[1] Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, Aarti Singh. Confidence sets for persistence diagrams. *Ann. Statist.*, 42(6):pp. 2301–2339, 2014. doi: 10.1214/14-AOS1252.

[2] Zhong Ge. Equivariant symplectic difference schemes and generating functions. *Physica D: Nonlinear Phenomena*, 49(3):pp. 376 – 386, 1991. ISSN 0167-2789. doi:https://doi.org/10.1016/0167-2789(91)90154-2.

[3] R. Ghrist. *Elementary Applied Topology*. CreateSpace Independent Publishing Platform, 2014. ISBN 9781502880857.

[4] Partha Niyogi, Stephen Smale, Shmuel Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1):pp. 419–441, 2008. ISSN 1432-0444. doi:10.1007/s00454-008-9053-2.

[5] Ippei Obayashi, Michio Yoshiwaki. Field choice problem in persistent homology, 2019.

[6] Nina Otter, Mason A. Porter, Ulrike Tillmann, Peter Grindrod, Heather A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):p. 17, 2017. ISSN 2193-1127. doi: 10.1140/epjds/s13688-017-0109-5.

[7] Andrew Robinson, Katharine Turner. Hypothesis testing for topological data analysis. *Journal of Applied and Computational Topology*, 1(2):pp. 241–261, 2017. ISSN 2367-1734. doi:10.1007/s41468-017-0008-7.

[8] Ana Romero, Jónathan Heras, Julio Rubio, Francis Sergeraert. Defining and computing persistent Z-homology in the general case, 2014.

[9] Daniel Stoffer. On the qualitative behaviour of symplectic integrators. part ii. integrable systems. *Journal of Mathematical Analysis and Applications*, 217(2):pp. 501 – 520, 1998. ISSN 0022-247X. doi:https://doi.org/10.1006/jmaa.1997.5728.

[10] Molei Tao. Explicit symplectic approximation of nonseparable hamiltonians: Algorithm and long time performance. *Phys. Rev. E*, 94:p. 043303, 2016. doi:10.1103/PhysRevE.94.043303.

[11] Haruo Yoshida. Construction of higher order symplectic integrators. *Physics Letters A*, 150(5):pp. 262 – 268, 1990. ISSN 0375-9601. doi:https://doi.org/10.1016/0375-9601(90)90092-3.

[12] Haruo Yoshida. Symplectic integrators for hamiltonian systems: Basic theory. *Symposium - International Astronomical Union*, 152:p. 407–411, 1992. doi:10.1017/S0074180900091440.

[13] Haruo Yoshida. Recent progress in the theory and application of symplectic integrators. *Celestial Mechanics and Dynamical Astronomy*, 56(1):pp. 27–43, 1993. ISSN 1572-9478. doi:10.1007/BF00699717.

[14] Ge Zhong, Jerrold E. Marsden. Lie-poisson hamilton-jacobi theory and lie-poisson integrators. *Physics Letters A*, 133(3):pp. 134 – 139, 1988. ISSN 0375-9601. doi:https://doi.org/10.1016/0375-9601(88)90773-6.

[15] Afra Zomorodian, Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):pp. 249–274, 2005. ISSN 1432-0444. doi:10.1007/s00454-004-1146-y.