Stochastic Arnold Diffusion of deterministic systems

V. Kaloshin

December 5, 2016

- Two main examples: Asteroid belt and Arnold’s example.
- Ergodicity (almost all orbits over a long time behave the same), mixed behavior: regular and stochastic.
- (Nearly) integrable systems, regular & stochastic behavior.
- KAM and regular (quasiperiodic) behavior on a set of positive measure.
- Stochastic behavior: horseshoe, Anosov, geodesic flows on manifolds of negative curvature. Can it occur on a set of positive measure?

- Building instabilities:
  - Arnold Example, Arnold mechanism and Mather mechanism
  - Weak KAM theory.
  - Variational shadowing lemma.
  - Arnold Example, Invariant Laminations mechanism
  - Invariant laminations and random iteration of cylinder maps
  - Stochasticity: Diffusion process, CLT for random iterations of cylinder maps
1 Dynamics in the Asteroid belt

We would like to analyze dynamics of Asteroids in the Asteroid belt.

- **Location** between orbits of Mars and Jupiter.
- **Quantity** \( \sim 10^6 \) objects of diameter \( \geq 1 \text{km} \).
- **Total mass** \( \sim 4\% \) of mass of the Moon.

**Goal:** Understand the dynamics of many orbits in the Asteroid belt.

Consider the distribution of Asteroid according to orbital period, see e.g. https://en.wikipedia.org/wiki/Asteroid(belt)

The Puzzle: This distribution has gaps called *Kirkwood gaps*.

Heuristic: Outside of gaps motions are regular (or quasi-periodic), Inside of gaps — stochastic.

Notice that equations of motion are deterministic and have randomness, stochasticity comes from initial conditions!

1.1 Mathematical model

Let \( q_i \in \mathbb{R}^3, i = 0, 1, \ldots, N \) be a finite collection of point masses. According to Newton’s law equations of motion are

\[
m_i \ddot{q}_i = \sum_{j \neq i} \frac{m_i m_j (q_j - q_i)}{|q_j - q_i|^3}.
\]

Let the Sun be \( q_0 \), Jupiter be \( q_1 \).

**Assumptions**

- Neglect interactions among Asteroids.
Neglect effect of Asteroids on the Sun and Jupiter.

Neglect effect of Mars (and other planets).

\[ \Rightarrow \] The Sun and Jupiter form a two body problem.

\[ \Rightarrow \] (Kepler) Orbits of the Sun and Jupiter, denoted \( q_0(t) \) and \( q_1(t) \), are elliptic.

One can choose a coordinate frame so that \( m_0 q_0(t) + m_1 q_1(t) \equiv 0 \). The number \( \mu = \frac{m_1}{m_0 + m_1} \) is called mass ratio. In reality it is small and \( \sim 10^{-3} \).

\[ \text{Figure 1: The Kirkwood gaps} \]

After normalization equations of motion become

\[ \ddot{q} = \frac{(1 - \mu)(q_1 - q(t))}{|q_0 - q(t)|^3} + \frac{\mu(q_1 - q(t))}{|q_1 - q(t)|^3}. \] (1)

**Mathematical Goal:** Understand the dynamics of many orbits of this equation of motion.
**KAM theory** In a region away from collisions and small $\mu$ for most of initial conditions orbits are quasiperiodic.

**Stochasticity Conjecture** For a positive measure set of initial conditions orbits behavior of eccentricity is stochastic, where stochasticity is due to a random choice of initial condition.

We shall make this conjecture more concrete later.

## 2 Mechanical and Hamiltonian systems

Consider an evolution of a particle $x : \mathbb{R} \to \mathbb{R}^N$, $N \geq 1$. According to the Newton Second Law the acceleration vector, defined by

$$
\ddot{x}(t_0) = \frac{d^2x}{dt^2} |_{t=t_0},
$$

is proportional to the force applied to the particle. We define a force as a function

$$
F : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^N
$$

and equations of motion become

$$
\ddot{x} = F(x, \dot{x}, t)
$$

and are called *Newton’s equation*. Call systems satisfying this equation — *mechanical systems*. Denote $(\cdot \cdot \cdot)$ the the scalar (or dot) product in $\mathbb{R}^N$ and $\|x\| = \sqrt{(x \cdot x)}$ be the length of $x$.

**Example of mechanical systems**

Let $U : \mathbb{R}^N \to \mathbb{R}$ be a differentiable function, called *potential*. Denote by $\partial U/\partial x$ the gradient of $U$. Consider $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ and

$$
\dot{x} = -\frac{dU}{dx},
$$

where $\frac{dU}{dx} = (\partial_{x_1} U, \ldots, \partial_{x_N} U)$. 

4
1. A stone falling to the Earth. In the case $N = 1$ and $U = -gx$ for Galileo’s $g \approx 9.8 \text{m/s}^2$.

More generally, for some $g \in \mathbb{R}^N$, called acceleration vector, we have

$$U(x) = (g \cdot x).$$

2. Falling from great height

Let $U(x) = -\frac{k}{\|x\|}$, i.e. $U$ is inversely proportional to the distance $\|x\|$.

3. Oscillations of a spring

Let $N = 1$ and $U(x) = -\alpha x^2$.

More generally, let $U(x) \geq 0$, $U(x) = 0$.

4. The mathematical pendulum

Let $N = 1$ and $U(x) = -k \sin x$ for some $k > 0$.

See also Chapter 1, [1].

5. Product systems

One can consider $X = (x_1, x_2) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$ for some $N_1, N_2 > 0$ and each component $x_1$ and $x_2$ evolves according to one of the above equations.

More generally, for a manifold $M$ and a cotangent bundle $T^*M$ one can define so-called canonical coordinate $(q, p) \in T^*M$. Then for a function smooth $H(q, p)$ called Hamiltonian we have the following equations of motion

$$\begin{cases}
\dot{p} = \frac{\partial H}{\partial q} \\
\dot{q} = -\frac{\partial H}{\partial p}.
\end{cases}$$

A Hamiltonian function can also depend on time, i.e. $H(q, p, t)$. Often time dependence assumed to be periodic, i.e. $t \in \mathbb{T} = \mathbb{R}/\mathbb{Z}$.

**Definition 1.** We call the Hamiltonian equation (2) of $n$ degrees of freedom, if the dimension of $M$ equals $n$; For time-periodic Hamiltonian with $\dim M = n$, the degree of freedom is $n + \frac{1}{2}$.

- All of the above examples with $M = \mathbb{R}^N$ for a proper $N$, $x = q$, $p = \dot{x}$,

$$H = \frac{(p \cdot p)}{2} + U(x).$$
• Geodesic flows on Riemannian manifolds Let \((M, g)\) be a smooth Riemannian manifold with a metric \(g\). One can define a flow on the cotangent bundle \(T^*M\) locally minimizing \(g\)-distance between nearby points on the same orbit. One can represent this flow as a Hamiltonian system. See Chapter 2, [3]

• For many other examples, see e.g. https://en.wikipedia.org/wiki/Hamiltonian(underline)mechanics

Two main examples:

Example 1 Arnold’s example is given by the product of

the Mathematical Pendulum \(\times\) a nonlinear spring+ \(O(\varepsilon)\).

For example, the Hamiltonian can have the form

\[
H_\varepsilon(q, p) = \frac{p_1^2}{2} + k \sin q_1 + \frac{p_2^2}{2} + U(q_2) + \varepsilon H_1(q, p, t), \quad \text{where } q_i, p_i \in \mathbb{R}.
\] (3)

\(U\) is even, \(U(q_2) \geq 0\), \(U(0) = U'(0) = 0\), \(U'(q_2) \neq 0\), \(q_2 \neq 0\) and \(|U(q_2)| \to +\infty\) as \(|q_2| \to +\infty\) and \(\varepsilon\) is small. For example, \(U(q_2) = q_2^2 + q_2^4\), the function \(H_1\) is assumed to be periodic in \(q_1\) and \(t\). Notice that in action-angle variable the condition is a lot easier to state, see (6).

Example 2

Motion of Asteroids in the Asteroid belt.

The Hamiltonan has the form

\[
H(q, p, t) = \frac{(p \cdot p)}{2} + \frac{1 - \mu}{|q - q_0(t)|} + \frac{\mu}{|q - q_1(t)|}, \quad \text{where } q, p \in \mathbb{R}^3,
\]

and \(q_0(t), q_1(t)\) are elliptic orbits of the Sun and Jupiter, given by Kepler’s law.

The associated problem is called the Restricted Three Body Problem.
Theorem (Arnold) For a positive measure set of initial conditions orbits are quasiperiodic.

In this setting the Stochasticity conjecture takes the following form:

Stochasticity conjecture For a positive measure set of initial conditions in a proper time scale eccentricity behaves as a stochastic diffusion process, where stochasticity is due to a random choice of initial conditions.

3 Ergodicity

Let $X$ be a set and $\Sigma$ be a $\sigma$-algebra over $X$. A non-negative function $\mu$ from $\Sigma$ to $[0, \infty) \cup +\infty$ is called measure if

- $\mu(\emptyset) = 0$
- $\mu$ satisfies countable additivity, i.e. for all countable collections $\{E_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\Sigma$: $\mu(\bigcup_{k=1}^{\infty} \Sigma_k) = \sum_{k=1}^{\infty} \mu(\Sigma_k)$.

The pair $(X, \Sigma)$ is called a measurable space and elements of $\Sigma$ are called measurable sets.

A triple $(X, \Sigma, \mu)$ is called a measure space. A probability measure is a measure with total measure one, i.e. $\mu(X) = 1$. A probability space is a measure space with a probability measure.

Let $T : X \to X$ be a measure-preserving transformation on a probability space $(X, \Sigma, \mu)$, i.e. for every $E$ in $\Sigma$ with $\mu(T^{-1}(E)) = \mu(E)$. Call a set $E$ in $\Sigma$ invariant if $T^{-1}(E) = E$. A measure-preserving transformation $T$ ergodic if for every $E$ invariant set either $\mu(E) = 0$ or $\mu(E) = 1$.

Time and space averages!

Let $f$ is a $\mu$-integrable function, i.e. $f \in L^1(\mu)$, i.e. $\int |f(x)| \, d\mu(x) < +\infty$. Then we define the following averages:

Time average: defined as the average (if it exists) over iterations of $T$ starting from some initial point $x$:

$$\hat{f}(x) = \lim_{n \to +\infty} \frac{1}{n} \sum_{k=1}^{n-1} f(T^k x).$$
Space average: If $\mu(X)$ is finite and nonzero, we can consider the space or phase average of $f$:

$$\bar{f} = \frac{1}{\mu(X)} \int f(y) d\mu(y).$$

**Ergodic theorem** If $T$ is ergodic, then for $\mu$-almost all $x$ the time average equals the space average, i.e.

$$\hat{f}(x) = \bar{f}.$$ 

Let $X$ be a topological space.

**Corollary 2.** In the above setting for $\mu$-almost all $x$ the corresponding orbit $\{T^k x\}_{k \in \mathbb{Z}}$ is dense in $X$.

**Example 1.** Let $X = \mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the circle, $\Sigma$ be the $\sigma$-algebra of measurable sets, $\mu$ be the Lebesgue measure. Let $f_\alpha : x \to x + \alpha \pmod{1}$ be the rotation by an irrational angle $\alpha \notin \mathbb{Q}$. Then $f_\alpha$ is ergodic.

**Example 2.** For the space probability space $(X, \Sigma, \mu)$ as in the previous example. Let $h_2 : x \to 2x \pmod{1}$ be the doubling map. Then $h_2$ is ergodic.

We say that a measure preserving $T : X \to X$ transformation of a probability space $(X, \Sigma, \mu)$ has **mixed behavior** if there are at least two invariant sets $R$ and $S$ both of positive measure such that “behavior of orbits” in $R$ and in $S$ are different. For example, almost periodic in $R$ and stochastic (in a certain sense) in $S$.

## 4 Integrable systems

**Definition 3** (Hamiltonian). For a $n$–dimensional smooth and boundless manifold $M$, We call an ODE on $T^*M$ being **Hamiltonian**, if there exists a function $H : T^*M \to \mathbb{R}$ and

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

for any $(q, p) \in T^*M$.  

To make the notion of a conserved quantity more precise, we give a formal definition. Let $M$ be a manifold of dimension $n$ and $T^*M$ be the cotangent bundle. Let $(p, q) = (p_1, \ldots, p_n, q_1, \ldots, q_n)$ be the canonical coordinates. Define the canonical non-degenerate two form
\[
\omega = \sum_{i=1}^{n} dp_k \wedge dq_k.
\]
The map $\Phi$ is called canonical if it preserves the canonical form $\omega$, i.e. $\Phi^* \omega = \omega$. Define the Poisson bracket $\{ \cdot, \cdot \}$ of two $C^1$ functions $F$ and $H$ by $\{ F, H \} = \omega(\nabla F, \nabla H)$, where $\nabla F$ and $\nabla H$ are the gradients of $F$ and $G$ resp.

Definition 4. $F : M^{2n} \to \mathbb{R}$ is a first integral of the Hamiltonian $(H)$ if $\dot{F} = \{ F, H \} = 0$.

Definition 5. Two real-valued functions $F_1, F_2$ on $M^{2n}$ are in involution if $\{ F_1, F_2 \} \equiv 0$.

Definition 6. A Hamiltonian $(H)$ of $n$ degrees of freedom is completely integrable (or simply integrable) if it has $n$ first integrals in involution:
$$\{ F_i, F_j \} \equiv 0 \text{ for all } i \neq j \in \{1, \cdots, n\}$$

4.1 First Integrals of the Two Body Problem

For example, we claim that the two body problem is completely integrable. Without loss of generality, we may assume it is planar, so $T^*M^{2n} = \mathbb{R}^8$, so we need to find 4 first integrals to show it is completely integrable. We have $q_1, q_2 \in \mathbb{R}^2$, $m_1, m_2 > 0$. We claim that the energy $H$, the angular momentum $G$, and the linear momentum $L$ from the system are first integrals. Since the linear momentum is a vector, the 3 quantities $H$, $G$, and $L = (L_1, L_2)$ really give 4 conserved quantities. These quantities are
\[
H = \sum \frac{|p_i|^2}{2m_i} - \sum \frac{m_im_j}{|q_i - q_j|}, \quad G = \sum q_i \times p_i, \quad L = \sum p_i.
\]
To verify, we compute the appropriate Poisson brackets:
\[
\begin{align*}
\{ G, H \} &= \dot{G} = 0 \text{ by Kepler’s second law}, \\
\{ L_i, H \} &= \dot{L}_i = 0 \text{ since the center of mass is fixed at 0 (for } i = 1, 2), \\
\{ G, L_i \} &= L_j \text{ (for } i, j \in \{1, 2\}, i \neq j) \text{ by Jacobi’s identity}, \\
\{ L_1, L_2 \} &= 0 \text{ by Jacobi’s identity},
\end{align*}
\]
so the two body problem is completely integrable.

It is conjectured and widely believed that for \( n \geq 3 \), the \( n \) body problem is not integrable. All the known integrals remain energy, angular momentum, and linear momentum, and for every additional body added the dimension of the phase space increases by four.

**Conjecture 7.** There are no other analytic first integrals for the \( n \) body problem.

Here is the famous Arnold-Liouville theorem on integrable Hamiltonians.

**Theorem 8.** Suppose the Hamiltonian \( (H) \) is integrable with \( H = F_1, \ldots, F_n \) as smooth first integrals. Consider \( \Gamma_a = \{ x \in M^{2n} : F_j(x) = a_j, 1 \leq j \leq n \} \), \( a \in \mathbb{R}^n \). Suppose \( \nabla F_j \) are linearly independent on \( \Gamma_a \), i.e.

\[
\text{Rank}(\nabla F_1, \cdots, \nabla F_n) = n.
\]

Then

1. \( \Gamma_a \) is a smooth \( n \)-dimensional submanifold invariant with respect to \( (H) \).

2. If \( \Gamma_a \) is compact and connected, then it is diffeomorphic to the \( n \)-torus \( T^n = \{ \theta = (\theta_1, \ldots, \theta_n) : \theta_i \in S^1 \} \).

3. In a neighborhood \( U \) of \( \Gamma_a \) there is a canonical map \( \Phi : (p,q) \to (I,\theta) \) and an implicitly defined function \( H_0(I(p,q)) = H(p,q) \) with

\[
\begin{align*}
\dot{\theta} &= \partial_I H_0(I) = \omega(I), \\
\dot{I} &= 0.
\end{align*}
\]

4. The flow of \( (H) \) defines a periodic or almost periodic motion on \( \Gamma_a \), i.e., \( \frac{d\theta}{dt} = \omega \).

The coordinates \( (\theta, I) \) are called *action-angle*: \( \theta \in \mathbb{T}^n \) be an \( n \)-dimensional angle and \( I \in \mathbb{R}^n \) be an action. The function \( \omega(I) \) is called the frequency map.

**Example 3.** Let \( X = \mathbb{T}^n \times \mathbb{R}^n \), \( n \geq 1 \), \( \Sigma \) be the set of measurable sets, and \( \mu \) be the Lebesgue measure. The \( (X, \Sigma, \mu) \) be a probability space. If \( \omega(I) = I \), then the time one map in (5) has the form

\[
\begin{align*}
\theta &\mapsto \theta + I \pmod{1}, \\
I &\mapsto I.
\end{align*}
\]
In this case the frequency map $\omega(I) = I$ is the identity map and $H_0(I) = \sum_{j=1}^{n} I_j^2/2$.

It turns out that in action-angle variable it is easier to write an Arnold’s example (3). Let $q_1, \theta, t \in \mathbb{T}$ be angle, $p_1, I \in \mathbb{R}$ be actions. Then the Hamiltonian can have the form

$$H_\varepsilon(q, p) = \frac{p_1^2}{2} + k \sin q_1 + \frac{f(I)}{2} + \varepsilon H_1(q_1, p_1, \theta, I, t),$$

where $f'(I)$ is strictly monotone for $I > 0$, i.e. $f''(I) \neq 0$ for $I > 0$, $H_1$ is assumed to be periodic in $t$.

**Exercise 1.** Consider a nonlinear spring with a potential $U(x), x \in \mathbb{R}$ such that $U$ is even, $U(x) \geq 0$, $U(0) = U'(0) = 0$, $U'(x) \neq 0$ for all $x \neq 0$. The Hamiltonian $H(x, p) = p^2/2 + U(x)$ is completely integrable and can be written in action-angle variables $(\theta, I) \in \mathbb{T} \times \mathbb{R}_+$. One can compute period of this string at each level set $\{H = E\}, \ E > 0$. Compute period $T(E)$ and notice that strict monotonicity of $T(E)$ implies strict monotonicity of $f(I)$.

### 4.2 Regular vs chaotic behavior

We start with preliminary analysis of the above three examples:

1. For the map $f_\alpha : x \to x + \alpha \pmod{1}$ with either irrational or rational $\alpha$ for a pair of nearby initial conditions $x, x'$ and any $n \in \mathbb{Z}$ we have

   $$|f_\alpha^n x - f_\alpha^n x'| = |x - x'|.$$

   In this case, there is no sensitive dependence on initial conditions.

2. For the map $h_2 : x \to 2x \pmod{1}$ for a pair of nearby initial conditions $x, x'$ and any positive $n \in \mathbb{Z}$ we have

   $$|h_2^n x - h_2^n x'| \leq \min\{2^n|x - x'|, 1\}.$$

   In this case, there is sensitive dependence on initial conditions as nearby orbits go apart exponentially fast.

   The exponent $\lambda$ of growth of the distance, i.e.

   $$|h_2^n x - h_2^n x'| \sim e^{\lambda n}|x - x'|$$

   is called Lyapunov exponents.
3. For the map of the cylinder \((\theta, I) \in \mathbb{T}^n \times \mathbb{R}^n\) given by (5), denoted \(F\), for a pair of nearby initial conditions \((\theta, I), (\theta', I')\) and any positive \(m \in \mathbb{Z}\) we have

\[
m|I' - I| \leq |F^m(\theta', I') - F^m(\theta, I)| \leq m|((\theta', I') - (\theta, I)|.
\]

In this case, there is weak sensitive dependence on initial conditions as nearby orbits go apart linearly.

**Exercise 2.** Construct example so that nearby orbits go apart as \(O(m^2)\). How about \(m^k\) for any positive integer \(k\)?

## 5 KAM theorem and Regular Behavior

Notice that all orbits of the equation (5) are either periodic or almost periodic. Consider the Hamiltonian in action angle coordinates \(H_0(I)\). The whole phase space is completely foliated into invariant tori \(\mathbb{T}^n \times \{I_0\}\), and on each of these tori the flow is linear with “frequencies”

\[
\omega = (\omega_1, \ldots, \omega_n) = \omega = \partial_I H_0(I).
\]

The components of \(\omega\) provide integrals in involution for the motion of the system. Such an integrable system is called *nondegenerate*, if these integrals are also functionally independent, which is expressed by the condition

\[
\det \frac{\partial \omega}{\partial I} = \det \partial^2_{II} H_0 \neq 0.
\]

Notice that this condition implies that \(\omega\) is a local diffeomorphism and “frequencies” are varying in an open set. Therefore, there are always both periodic and almost periodic orbits.

Consider a Hamiltonian perturbation of \(H_0\) given by

\[
H_\varepsilon(\theta, I) = H_0(I) + \varepsilon H_1(\theta, I).
\]

We want to persist some of these tori \(\mathbb{T}^n \times \{I_0\}\) for small \(\varepsilon\). But it was already known to Poincare that tori with periodic orbits generally break up under small perturbations, and for a long time it was an open question whether any torus can be continued for \(\varepsilon \neq 0\). Finally, by the work of Kolmogorov, Arnold,
and Moser, it turned out that those tori will persist whose frequencies are not only rationally independent but satisfy a "small denominator condition"

\[ |(\omega \cdot k)| \geq \frac{\gamma}{\|k\|^\tau}, \quad 0 \neq k \in \mathbb{Z}^n, \quad (7) \]

with \( \gamma > 0 \). Call a vector satisfying this condition \((\gamma, \tau)\)-diophantine and denote \( \mathcal{D}_{\gamma, \tau} \). Call a vector simply diophantine if it is \((\gamma, \tau)\)-diophantine for some \( \gamma > 0 \) and \( \tau > n - 1 \).

**Exercise 3.**  
- Show that almost all points in \( \mathbb{R}^n \) satisfy such a condition for some \( \gamma \) while \( \tau > n - 1 \) is kept fixed (Lebesgue full measure).
- The set \( \mathcal{D}_{\gamma, \tau} \) is nowhere dense.
- Let \( \Omega \subset \mathbb{R}^n \) be an bounded open set. Then the complement to the set of \((\gamma, \tau)\)-diophatine numbers \( \Omega \setminus \mathcal{D}_{\gamma, \tau} \) has measure \( \sim c\gamma \), where \( c \) depends on measure of \( \Omega \).

Fix \( \tau > n - 1 \). For an open set \( \Omega \in \mathbb{R}^n \) denote by

\[ \Omega_\gamma = \Omega \cap \mathcal{D}_{\gamma, \tau} \]

the Cantor set of \( \gamma \)-diophatine numbers.

**Kolmogorov–Arnold–Moser Theorem** (Autonomous) Let \( \Omega \subset \mathbb{R}^n \) be an open set and \( H_0 : \Omega \to \mathbb{R} \) be a real analytic Hamiltonian, which is nondegenerate for \( I \in \Omega \). Let the perturbed Hamiltonian \( H_\varepsilon = H_0 + \varepsilon H_1 \) be real analytic. Then, for a sufficiently small \( \varepsilon \) portional to \( \gamma^2 \), the perturbed system possesses an analytic invariant \( n \)-dimensional tori \( T_\omega \) with a linear flow for all \( \omega \in \mathcal{D}_\gamma \). Moreover, if \( \Omega \) is bounded, then the union of these tori occupies all but \( O(\sqrt{\varepsilon}) \) part of \( \mathbb{T}^n \times \Omega \), i.e.

\[ \text{Mes}(\Omega \times \mathbb{T}^n) \setminus \text{Mes}(\bigcup_{w \in \mathcal{D}_\gamma} T_w) \leq c \cdot \sqrt{\varepsilon} \]

with \( c \) depending on \( n \) and \( \|H_0\|_{C^2} \).

Sometimes one needs to consider a time-periodic perturbation of the integral systems, which is formally expressed by

\[ H_\varepsilon(\theta, I, t) = H_0(I) + \varepsilon H_1(\theta, I, t), \quad (\theta, I, t) \in \mathbb{T}^n \times \Omega \times \mathbb{T}. \quad (8) \]
In order to reduce to the autonomous case one can complete this system into an autonomous one by adding one variable $E$ conjugate to time $t$ and have

$$\tilde{H}_\varepsilon(\theta, I, t, E) = E + H_\varepsilon(\theta, I, t),$$

with the unperturbed integral part $\tilde{H}_0(\tilde{I}) = E + H_0(I)$, $\tilde{I} = (I, E)$. Notice that

$$\det \partial_{ij}^2 \tilde{H}_0(\tilde{I}) = 0$$

so the previous Theorem is no longer available for this case, but we can verify that $\tilde{H}_0(\tilde{I})$ is so called isoenergetically non-degenerate by

$$\det \begin{pmatrix} \partial_{ij}^2 \tilde{H}_0(\tilde{I}) & \nabla \tilde{H}_0 \\ \nabla \tilde{H}_0 & 0 \end{pmatrix} \neq 0,$$

and we indeed have a non-autonomous version of KAM theorem w.r.t. this condition:

**Kolmogorov–Arnold–Moser Theorem (Non-autonomous)** Let $\Omega \subset \mathbb{R}^n$ be an open set and $H_0 : \Omega \to \mathbb{R}$ be a real analytic Hamiltonian, which is isoenergetically non-degenerate for $I \in \Omega$. Let the perturbed Hamiltonian $H_\varepsilon = H_0 + \varepsilon H_1$ be real analytic. Then, for a sufficiently small $\varepsilon$ proportional to $\gamma^2$, the perturbed system possesses an analytic invariant $n$-dimensional tori $T_w$ with a linear flow for all $\omega \in D_\gamma$. Moreover, if $\Omega$ is bounded, then the union of these tori occupies all but $O(\sqrt{\varepsilon})$ part of $\mathbb{T}^n \times \Omega$, i.e.

$$\operatorname{Mes}(\Omega \times \mathbb{T}^n) \setminus \operatorname{Mes}(\bigcup_{w \in D_\gamma} T_w) \leq c \cdot \sqrt{\varepsilon}$$

with $c$ depending on $n$ and $\|H_0\|_{C^2}$.

5.1 The Energy Reduction

We need to specify that, the autonomous case can be transferred into the time-periodic case: For a fixed energy level $H_\varepsilon(\theta, I) = E$, if

$$\partial_{I_n} H_\varepsilon(\theta, I) \neq 0,$$

then there should be a function $I_n(\theta, \tilde{I}, E)$ satisfying

$$H_\varepsilon(\theta, \tilde{I}, I_n(\theta, \tilde{I}, E)) = E, \quad (\tilde{\theta}, \tilde{I}) := (\theta_1, \cdots, \theta_{n-1}, I_1, \cdots, I_{n-1}).$$
due to the implicit function theorem. Benefit from the continuity of \( H_\epsilon \), we can find an open interval \( U_E \subset \mathbb{R} \) containing \( E \), such that

\[
\partial_{I_n} H_\epsilon(\theta, I) \bigg|_{U_E} \neq 0.
\]

Then \( I_n(\hat{\theta}, \hat{I}, \theta_n, E) \) can be interpreted by a non-autonomous Hamiltonian with \( \theta_n \) playing the role of time \(-t\), i.e.

\[
\frac{d\hat{\theta}}{d\theta_n} = \frac{\partial I_n}{\partial \hat{I}}, \quad \frac{d\hat{I}}{d\theta_n} = -\frac{\partial I_n}{\partial \hat{\theta}} \tag{9}
\]

with \((\hat{\theta}, \hat{I}) \in T^*\mathbb{T}^{n-1}, \theta_n \in \mathbb{T}\).

**Theorem 9.** Suppose \( M \) is a \( n + 1 \) dimensional manifold, on which we can define a Hamiltonian flow (2). Restricted on the energy surface \( S_E = \{ H = E \} \), the Hamiltonian equations can be transferred into

\[
\begin{align*}
\frac{dp_i}{dt} &= \partial_{q_i} P_{n+1} \\
\frac{dq_i}{dt} &= -\partial_{p_i} P_{n+1} \tag{10}
\end{align*}
\]

where the function \( P_{n+1}(P, Q, t; E) \) is implicitly defined by

\[
H(P, P_{n+1}, Q, -t) = E.
\]

**Remark.** Notice that the time-periodic system has dimension \( 2n + 1 \), which one dimension less of the original autonomous system. Since \( P_{n+1}(P, Q, t; E) \) can be considered as a new Hamiltonian defined on the phase space \((P, Q, t) \in T^*\mathbb{T}^n \times \mathbb{T}^1\), the corresponding degree of freedom is \( n + \frac{1}{2} \).

**Exercise 4.** Prove that if \( H_\epsilon \) is nearly integrable, then induced \( I_n \) is also nearly integrable.

**Example 4.** (Needs to be modified to include nearly integrable case)

Consider a smooth Hamiltonian

\[
H_\epsilon(q_1, \ldots, q_{n+1}, p_1, \ldots, p_{n+1}) = H_0(p_1, \ldots, p_{n+1}) + \varepsilon H_1(q, p).
\]

Fix \( \delta > 0 \) and \( E > \delta \) and \(|p_{n+1}| > \delta\). Then

\[
p_{n+1} = \pm \left( 2E - \sum_{j=1}^{n} p_j^2 - 2\varepsilon H_1(q_1, \ldots, q_{n+1}, p_1, \ldots, p_{n+1}) \right)^{1/2}
\]
which has a solution

\[ P_{n+1}(P, Q, t; E) = \pm \left( 2E - \sum_{j=1}^{n} p_j^2 \right)^{1/2} + \varepsilon \Delta P(P, Q, t : E), \]

where \( P = (p_1, \ldots, p_n) \), \( Q = (q_1, \ldots, q_n) \) and \( \Delta P(\cdot) \) is a smooth function.

Now let’s turn back to the KAM theorem, instead of giving a detailed proof of that, we will make the demonstration via two low dimensional examples, which are related with the aforementioned settings.

**Example 5.** For \( n = 1 \), any Hamiltonian system on \( T^*\mathbb{T}^1 \) is integral, cause \( H(\theta, I) \) is a naturally first integral. If for a fixed energy level \( H(\theta, I) = E \), we have \( \partial H \bigg|_E \neq 0 \), by the Implicit Function Theorem we can solve the implicit equation \( H(\theta, \mathcal{I}(\theta, E)) = E \). The solution

\[ \mathcal{T} = \{ (\theta, I(\theta, E)) : \theta \in \mathbb{T} \} \]

describes an invariant torus.

Notice that for any sufficiently small perturbation \( \varepsilon H_1(\theta, I) \), the condition \( \partial H \bigg|_E \neq 0 \) implies that the condition \( \partial H + \varepsilon H_1 \bigg|_E \neq 0 \). Thus, there exists an torus \( \mathcal{I}_\varepsilon(\theta, E) \) satisfying

\[ H(\theta, \mathcal{I}_\varepsilon(\theta, E)) + \varepsilon H_1(\theta, \mathcal{I}_\varepsilon(\theta, E)) = E, \quad \forall \theta \in \mathbb{T}. \]

Moreover, this torus is a small perturbation of the torus \( \{ I = I_* \} \), where \( H_0(I_*) = E \).

**Remark.** We need to specify that the deformed torus with the same frequency \( \omega \) may lie a different energy surface compare to the unperturbed torus. Here is a trivial example: \( H(I) = I^2/2 \) and \( H_\varepsilon(I) = H(I) + \varepsilon I \), for \( w_0 = 1 \), the unperturbed torus \( \mathcal{T}_0 = \{ (\theta, 1) : \theta \in \mathbb{T} \} \) lies on \( \{ H(\theta, I) = 1/2 \} \), whereas \( \mathcal{T}_\varepsilon = \{ (\theta, 1 - \varepsilon) : \theta \in \mathbb{T} \} \) lies on \( \{ H_\varepsilon(\theta, I) = (1 - \varepsilon^2)/2 \} \).

**Example 6.** For \( n = 1.5 \), which means the Hamiltonian is time-periodic \( H_\varepsilon(\theta, I, t) = H_0(I) + \varepsilon H_1(\theta, I, t) \) with \( (\theta, I, t) \in T^*\mathbb{T} \times \mathbb{T} \), we can process the KAM iteration for the following Poincaré map:
5.2 Poincare maps and further dimension reduction

In the previous section we showed that orbits of the autonomous system (2) satisfy a time-periodic equation (10), which has one dimension less. Now we describe the procedure of reducing \((2n+1)\)-dimensional time-periodic system to a \(2n\)-dimensional map.

Let \(\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{T}^n, I = (I_1, \ldots, I_n) \in \mathbb{R}^n\).

\[
\tilde{H}_\varepsilon(\theta, I, t) = \tilde{H}_0(I) + \varepsilon \tilde{H}_1(\theta, I, t) \tag{11}
\]

be a time-periodic Hamiltonian. Consider the time one map

\[
f_\varepsilon : (\theta, I) \to (\theta', I') \tag{12}
\]

sending \((\theta, I, 0)\) to \((\theta', I', 1)\). This is a \(2n\)-dimensional symplectic map, i.e. it preserves the canonical form \(\omega = dI \wedge d\theta\). Corresponding to aforementioned Example, we study the case \(n = 1\).

Call a \(2n\)-dimensional symplectic map \(f : (\theta, I) \to (\theta', I')\) integrable if it has the form \((\theta, I) \to (\theta + \omega(I) \mod 1, I)\) for some map \(\omega(I) \in \mathbb{R}^n\). Call a \(2n\)-dimensional symplectic map \(f\) nearly integrable if it can be written in the form \(f = f_0 + \varepsilon f_1\), where \(f_1\) is sufficiently smooth and \(\varepsilon\) can be chosen sufficiently small.

Exercise 5. Show that if \(\tilde{H}_\varepsilon\) is nearly integrable, then so is its Poincare map \(f_\varepsilon\).

5.3 Exact Area Preserving Twist maps

Let \((\theta, I) \in \mathbb{T} \times \mathbb{R} = \mathbb{A}\) be coordinates on the cylinder \(\mathbb{A}\). Call a \(C^1\) smooth map \(f : (\theta, I) \to (\theta', I')\). Denote by \(\hat{f}(x, I)\) the list of this map to a map \(\mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}\), i.e. \(\hat{f}(x + 1, I) = (x' + 1, I')\) for all \((x, I) \in \mathbb{R}^2\).

We call \(f\)

- **area-preserving** if it preserves the area form \(d\theta \wedge dI\).
- **exact** if for any noncontractible curve \(\gamma\) the flux of \(\gamma\) is zero, i.e. the area above \(\gamma\) below \(f(\gamma)\) equals the area below \(\gamma\) above \(f(\gamma)\).
- **twist** if \(\partial I' / \partial x > 0\), i.e. image \(f(L_{x_0})\) of every vertical line \(L_{x_0} = \{x = x_0\}\) is monotonically twisted in the \(x\)-component.
Call $f$ satisfying the first two items a twist map and satisfying all three items an EAPT map.

**Theorem 10.** (Moser) Any EAPT map can be given as the time one map of a time-periodic Hamiltonian $\tilde{H}(x,I,t)$ satisfying the Legendre condition

$$\partial_{II}^2 \tilde{H}(x,I,t) > 0 \quad \text{for all } (x,I,t) \in \mathbb{R} \times \mathbb{R} \times \mathbb{T}.$$

**Example 7** (Standard Map). Let $a$ be a real number. Consider the map of a cylinder

$$f_a: (\theta,I) \mapsto (\theta + I + a \sin 2\pi \theta \pmod{1}, I + \varepsilon \sin 2\pi \theta).$$

Its lift has the form

$$\hat{f}_a: (x,I) \mapsto (x + I + a \sin 2\pi x \pmod{1}, I + \varepsilon \sin 2\pi x).$$

**Exercise 6.** Prove that $f_a$ is an EAPT map for any $a \in \mathbb{R}$.

**Example 8** (The Billiard Map). Let $\Omega \subset \mathbb{R}^2$ be a strictly convex domain. Let the boundary $\partial \Omega$ be $C^r$ smooth with $r \geq 3$. Fix a point $P_0 \in \partial \Omega$ and let $s$ be the length parametrization, i.e. $P_s$ is at distance $s$ from $P_0$ in the clockwise direction. Define a map

$$f_\Omega: (s,I) \mapsto (s',I'),$$

where $\varphi \in [0,\pi]$ be angle of a ray with the boundary and $I = \cos \varphi \in [-1,1]$. This is called a billiard map, see Fig.2.

**Exercise 7.** Prove that $f_\Omega: \mathbb{T} \times [-1,1] \rightarrow \mathbb{T} \times [-1,1]$ is an EAPT map.

**HINT:** prove that the generating function is $-\|P_0(s)-P_1(s')\|_e$, where $\|\cdot\|_e$ is the Euclid norm on $\mathbb{R}^2$.

### 5.4 Generating Functions

We say that $\hat{f}: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ has a generating function $h: \mathbb{R}^2 \rightarrow \mathbb{R}$, $h(x,x')$ if the map $\hat{f}(x,I) = (x',I')$ for all $(x,I) \in \mathbb{R}^2$ satisfies

$$\begin{cases} \partial_1 h(x,x') = -y \\ \partial_2 h(x,x') = y'. \end{cases}$$

(13)
Figure 2: the incident angle always equals the reflective angle

Example 9 (Standard Map). The generating function for the standard map is
\[ h(x, x') = \frac{1}{2} |x' - x|^2 - \frac{\varepsilon}{2\pi} \cos 2\pi x, \quad x, x' \in \mathbb{R}. \]

Example 10 (Billiard Map). The generating function for the billiard map is
\[ h(s, s') = -\|P_0(s) - P_1(s')\|_e, \]
where \(\|\cdot\|_e\) is the Euclid norm on \(\mathbb{R}^2\).

Theorem 11. Any \(C^1\) smooth twist map \(f\) possesses a generating function \(h\), i.e. (13) holds for all \((x, I) \in \mathbb{R}^2\). Moreover, \(f\) is exact if and only if \(h\) is periodic in the sense that \(h(x + 1, x' + 1) = h(x, x')\) for all \((x, x') \in \mathbb{R}^2\).

Exercise 8. Prove that for a nearly integrable Hamiltonian \(H_\varepsilon(\theta, I, t) = H_0(I) + \varepsilon H_1(\theta, I, t)\) with \(H_0(I)\) be positively definite, the Poincaré map of \(\Sigma_0\) is an EAPT map with the generating function by
\[ h(x, X) = \inf_{\substack{\gamma(0) = x \\ \gamma(1) = X}} \int_0^1 L(\gamma, \dot{\gamma}, t)dt \]
where the Lagrangian function \(L(x, v, t)\) defined on \(TT \times T\) can be derived via the Legendre transformation:
\[ L(x, v, t) = \max_{p \in \mathbb{R}} \{ p \cdot v - H(x, p, t) \}. \]
HINT: use Moser's theorem.

In a view of this exercise the equation (13) can be viewed as the Euler-Largange equation. Indeed, if \( \hat{f}^k(x,I) = (x_k,I_k) \), \( k \in \mathbb{Z} \) is an orbit, then for the formal sum

\[
\sum_{k \in \mathbb{Z}} h(x_k,x_{k+1})
\]

the extrema for each \( k \in \mathbb{Z} \) satisfy the equation

\[
\partial_1 h(x_k,x_{k+1}) + \partial_2 h(x_{k-1},x_k) = 0.
\]

In general, the sequences \((x_k)_{k \in \mathbb{Z}}\) satisfying this equation called stationary.

**Exercise 9.** Each orbit \( \hat{f}^k(x,I) = (x_k,I_k) \), \( k \in \mathbb{Z} \) defines a stationary configuration \((x_k)_{k \in \mathbb{Z}}\) and each stationary configuration \((x_k)_{k \in \mathbb{Z}}\) defines an orbit \((x_k,I_k)\), \( k \in \mathbb{Z} \) by setting \( I_k = \partial_1 h(x_kx_{k+1}) \).

Call a generating function \( h(x,x') \) nearly integrable if it can be written in the form \( h(x,x') = h_0(x' - x) + \varepsilon h_1(x,x') \), where \( h_1 \) has bounded \( C^3 \) norms and \( \varepsilon \) can be chosen sufficiently small.

**Exercise 10.** Show that if \( f_\varepsilon \) is nearly integrable, then so is its generating function \( h_\varepsilon \).

### 5.5 Equation for an invariant curve for EAPT

Suppose \( \hat{\Gamma}^\omega \) is an invariant curve of a lifted EAPT \( \hat{f} : (x,I) \to (x',I') \), where \( (x,I) \in \mathbb{R}^2 \). Moreover, suppose that in the \((x,I)\)-coordinates there is a parametrization

\[
\hat{\Gamma}^\omega = \{ w(\theta) = (u(\theta),v(\theta)) \in \mathbb{R}^2 | \theta \in \mathbb{R} \},
\]

such that for some real \( \omega \in \mathbb{R} \) we have

\[
(IE) \quad \hat{f}(w(\theta)) = w(\theta + \omega), \quad \forall \theta \in \mathbb{R}.
\]

This, in particular, means that \( \hat{f} \) restricted on \( \Gamma^\omega \) is conjugated to the rigid rotation \( x \to x + \omega \).

Notice that \( u(\theta) - \theta, v(\theta) \) are both 1-periodic in \( \theta \). Denote by \( \Gamma^\omega \) the natural projection of \( \hat{\Gamma}^\omega \) onto the cylinder \( A \) and by the projection of \( \hat{f} \) to an EAPT map \( f \).

Denote by \( u^\pm(\theta) := u(\theta \pm \omega) \), then we have the following criteria
Theorem 12. \( \tilde{\Gamma}^{\omega} \) satisfies (IE) iff \( u(\theta) \) satisfies the following 2\textsuperscript{nd} order difference equation holds

\[
(EL) : \quad E(u(\theta)) = \partial_1 h(u, u^+)(\theta) + \partial_2 h(u^-, u)(\theta) \equiv 0, \quad \forall \theta \in \mathbb{R}. \quad (15)
\]

Remark. For the standard map, the 2\textsuperscript{nd} order equation becomes:

\[
u(\theta) + (\theta) - 2u(\theta) + u(\theta) = \varepsilon \sin 2\pi u(\theta).
\]

For a nearly integrable map with a generating function \( h_\varepsilon(x, x') = h_0(x' - x) + \varepsilon h_1(x, x') \) the 2\textsuperscript{nd} order equation becomes:

\[
\partial h_0(u^+(\theta) - u(\theta)) + \partial h_0(u(\theta) - u^-(\theta)) = -\varepsilon (\partial_1 h_1(u, u^+)(\theta) + \partial_2 h_1(u^-, u)(\theta)).
\]

So we successfully transfer the existence of KAM torus to the existence of function \( u(\theta) \) of (15).

In the next section we look for \( \tilde{\Gamma}^{\omega} \) so that it corresponds to just the KAM torus of \( \tilde{H}_\varepsilon \) with the frequency \((\omega, 1)\).

5.6 Newton Iteration Method

Goal: suppose \( E(u_0) \leq \varepsilon \) is small and \( E'(u_0) \sim O(1) \), then in a neighborhood of \( u_0 \) there exists a unique solution making \( E(u^*) = 0 \).

The following is a scheme towards this goal, which is actually the essence of the Newton Iteration Method. For \(|v| \leq \varepsilon \ll 1\),

\[
E(u_0 + v) = E(u_0) + E'(u_0)v + Q(v, v),
\]

where \(|Q(v, v)| \leq \varepsilon^2\) is a quadratic term.

**Step 1:** solve \( E(u_0) + E'(u_0)v = 0 \), and denote \( u_1 = u_0 + v \), so we finish the first step iteration by

\[
E(u_1) = Q(v, v) \lesssim E^2(u_0);
\]

**Step n:** make \( E(u_n) + E'(u_n)v_n = 0 \), and denote \( u_{n+1} = u_n + v_n \). Similarly we get

\[
E(u_{n+1}) = Q(v_n, v_n) \lesssim E^2(u_n) \lesssim \cdots \lesssim E^n(u_0).
\]

Remark. We have to confess that, in the previous scheme, \( v_n \) has less regularity than \( u_n \), and that would cause some difficulty !!!
5.7 A further KAM exploration: from (IE) to a circle diffeomorphism

Now we turn back to (15), and further explore the relationship between it and the KAM iteration. Recall that \( f_\varepsilon \) is nearly integrable due to Exercise 4, that implies \(|u(\theta) - \theta| \leq \varepsilon\). So \( f_\varepsilon|\Gamma_w \) becomes a circle diffeomorphism:

\[
\text{(Model Problem):} \quad f_0 : x \to x + w \pmod{1}, \quad f_\varepsilon : x \to x + w + \varepsilon v(x) \pmod{1},
\]

find \( v(x) \) such that these two circle diffeomorphisms are conjugated via the following commutative graph:

\[
\begin{array}{ccc}
\mathbb{T} & \xrightarrow{f_\varepsilon} & \mathbb{T} \\
\uparrow \varphi_\varepsilon & & \uparrow \varphi_\varepsilon \\
\mathbb{T} & \xrightarrow{f_0} & \mathbb{T} 
\end{array}
\]

Formally we write by \( \varphi_\varepsilon(x) = x + \varepsilon \psi(x) \), then

\[
f_\varepsilon = \varphi_\varepsilon \circ f_0 \circ \varphi_\varepsilon^{-1} = x + w + \varepsilon \psi(x + w) - \varepsilon \psi(x) + O(\varepsilon^2). \tag{16}
\]

On the other side, \( f_\varepsilon(x) = x + w + \varepsilon v(x) + O(\varepsilon^2) \), so we deduce the following cohomology equation:

\[
\psi(x + w) - \psi(x) = v(x).
\]

**Definition 13.** Denote by \( \mathbb{W}_r := \{ f \in C^w(\mathbb{T}, \mathbb{R}) | \|f\|_r = \sup_{|\theta| \leq r} |f(\theta)| \} \), where \( r \) is called the analytic radius.

**Exercise 11.** The Fourier expansion \( f(\theta) = \sum_{k \in \mathbb{Z}} f_k e^{ik\theta} \) satisfies

\[
|f_k| \leq \|f\|_r \exp(-|k|r), \quad \forall k \in \mathbb{Z}
\]

for \( f \in \mathbb{W}_r \).

**Lemma 14 (Iteration).** Suppose \( w \) is \( \gamma \)-Diophantine, \( v \in \mathbb{W}_r \) is of zero mean, i.e. \( v_0 = 0 \). Then the cohomology equation has a unique solution \( \psi \in \mathbb{W}_{r'} \) with zero mean. Moreover, for any \( 0 < r' < r \),

\[
\|\psi\|_{r'} \leq \frac{c(\gamma)\|v\|_r}{|r - r'|^3}.
\]
Proof. Formally calculating \( \psi(x+w)-\psi(x) = \sum_{k \in \mathbb{Z}} \psi_k(e^{i2\pi kw}-1)e^{ikx} \), which implies

\[
\psi_k = \frac{v_k}{e^{i2\pi kw} - 1}, \quad (k \neq 0, \psi_0 = 0).
\]

Recall that \(|qw-p| \geq \gamma \cdot |q|^{-2}\) for all \(q,p \in \mathbb{Z}\), which is due to the Diophantine condition, then

\[
|e^{i2\pi kw} - e^{i2\pi p}| \geq \frac{c(\gamma)}{|k|^2}.
\]

(17)

On the other side,

\[
|v_k| \leq \|v\|_r \exp(-|k|r), \quad \forall k \in \mathbb{Z}
\]

Then

\[
|\psi_k| \leq \|v\|_r e^{-|k|r} \exp(1(|k|) |k|^2)
\]

\[
\leq \|v\|_r e^{-|k|s} e^{-|k|(r-s)c^{-1}(\gamma)|k|^2}.
\]

Recall that \(xe^{-x} \leq e^{-1}\) for \(x \geq 1\), then for \(x = \frac{a|k|}{b}\)

\[
e^{-a|k||k|^b} \leq e^{-\left(\frac{b}{a}\right)^b}, \quad \forall k \in \mathbb{N}.
\]

Substitute by \(a = r-s, b = 2\) then

\[
|\psi_k| < c_1(\gamma)\|v\|_r \exp(-s|k|) \frac{\exp(-s|k|)}{(r-s)^2}.
\]

Finally we get

\[
\|\psi\|_r' \leq \sum_k |\psi_k| e^{-r'|k|} \leq \frac{c_1(\gamma)\|v\|_r}{(r-s)^2} (1 - \exp(-(s-r')))
\]

\[
\leq \frac{c_1(\gamma)\|v\|_r}{(r-s)^2} \frac{1}{(s-r')^2}
\]

\[
\leq \frac{c_1(\gamma)\|v\|_r}{(r-r')^2}
\]

(18)

for \(s = \frac{r+r'}{2}\). So we finish the proof. \(\square\)
5.8 the KAM iteration of generating functions

The proof we present here is due to Moser and can be found in Moser-Levi [4]. A multidimensional version of the same approach can be found in [5]. Let’s rewrite the discrete (EL) equation here:

$$E(u(\theta)) = \partial_1 h(u, u^+) + \partial_2 h(u^-, u) = 0, \quad \forall \theta \in \mathbb{R}.$$  

For nearly integrable generating function $h_\varepsilon(x, x') = h_0(x' - x) + \varepsilon h_1(x, x')$, we can transfer (EL) into

$$h_0'(u^+ - u) - h_0'(u - u^-) = -\varepsilon[\partial_1 h_1(u, u^+) + \partial_2 h_1(u^-, u)]$$

w.r.t $h_0''(\cdot) \neq 0$.

Remark. • Actually, (EL) is the 1st order variational equation of the following Percival Variational Principle:

$$\delta \int_0^1 h(u, u^+) d\theta = 0. \quad (P.V.P).$$

• the mean value of $u_0 \mathbb{E}(u)$ equals zero, i.e.

$$\int_0^1 u_0 \mathbb{E}(u) d\theta = 0.$$

Theorem 15 (the KAM theorem of twist maps). • Let $\mathcal{D} \subset \mathbb{C}^2$ be an open set of complex plane such that a generating function $h(x, x')$ is real analytic and analytic in $\mathcal{D}$. Moreover, the complex expansion still satisfies

$$h(x, x') = h(x + 1, x' + 1), \quad \forall (x, x') \in \mathcal{D}.$$

• Let $\min_{\mathcal{D}} |\partial_{12} h| > \kappa > 0$ for some $\kappa > 0$. Suppose there exists $M > 0$, such that $\|h\|_{C^3(\mathcal{D})} < M$. Denote by

$$\mathcal{D}_R := \{(x, x') \in \mathcal{D} | \text{whose } R-\text{neighborhood belongs to } \mathcal{D}\}.$$

• $u_0(\theta) - \theta \in \mathbb{W}_r$ with $0 < r < 1$, and $(u_0, u^+) \in \mathcal{D}_R$. There exists $N_0 > 0$, such that

$$|(u_0)_\theta| < N_0, \quad |(u_0)^{-1}_\theta| < N_0 \forall |\Im \theta| < r.$$
For any diophantine \( \omega \), i.e. \( \exists \gamma > 0 \), such that
\[
|p - q\omega| > \gamma \cdot |q|^{-2}
\]  
for all \( p, q \in \mathbb{Z} \), if some \( h, u_0 \) satisfy all the previous conditions, then \( \delta = \delta(r, h, M, N_0, \gamma, R) \), if \( |E(u_0)|_r < \delta \), then there exists a unique \( u \) close to \( u_0 \), such that
\[
E(u) = 0
\]
and \( u(\theta) - \theta \in \mathbb{W}_{r/2} \) with mean value of \( u(\theta) - \theta \) being zero.

In the next section we will give a proof for this Theorem.

5.9 The homological equation and Newton Iteration method

Let’s recall the (EL) equation: if \( u = u_0 + v \), then
\[
E(u) = E(u_0) + E'(u_0)v + Q(v, v)
\]
with
\[
E'(u)v = [\partial_{11}h(u, u^+) + \partial_{22}h(u^+, u)]v + [\partial_{12}h(u, u^+)v^+ + \partial_{12}h(u^+, u)v^-].
\]
We can apply the Newton Iteration method to this equation, i.e.

- **Step 1:** solve \( E'(u_0)v_0 = -E(u_0) \). Let \( u_1 = u_0 + v_0 \), then \( E(u_1) = Q(v_0, v_0) \sim E^2(u_0) \).
- **Step n:** solve \( E'(u_n)v_n = -E(u_n) \). Let \( u_{n+1} = u_n + v_n \), then \( E(u_{n+1}) = Q(v_n, v_n) \sim E^2(u_n) \sim \cdots \sim E^{2^n}(u_0) \).

But we still have to face the decay of regularity!!!

We call
\[
E'(u)v = -E(u) \quad (HE)
\]
the homological equation.

Recall that we will face the quadratic derivatives in (HE), so we should improve it into a manageable case. In other words, we need to make a transformation from solving ‘quadratic equation’ to solving a couple of ‘linear equations’ !!!
Let’s multiply a $u_\theta$ on both sides,

$$u_\theta \mathbb{E}'(u)v = -u_\theta \mathbb{E}(u).$$

Recall that $|\mathbb{E}(u)| < \delta$ and $|v| < \delta$, then $v \cdot \frac{d}{d\theta} \mathbb{E}(u) = v \mathbb{E}'(u)u_\theta$ is a quadratic small term. So we can get

$$u_\theta \mathbb{E}'(u)v - v \mathbb{E}'(u)u_\theta = -u_\theta \mathbb{E}(u), \quad (20)$$

which can be further transferred into

$$\partial_{12}h(u, u^+)(u_\theta v^+ - u_\theta^+ v) - \partial_{12}h(u^-, u)(u_\theta v - u_\theta^- v) = -u_\theta \mathbb{E}(u).$$

Introduce $w = \frac{u}{u_\theta} \approx v$, then it becomes

$$\partial_{12}h(u, u^+)u_\theta u_\theta^+(w^+ - w) - \partial_{12}h(u^-, u)u_\theta u_\theta^-(w - w^-) = -u_\theta \mathbb{E}(u).$$

Let’s define the discrete gradient by

$$\nabla f = f^+ - f, \quad \nabla^* f = f - f^-,$$

then formally we get a manageable homological equation by:

$$\nabla^* (\partial_{12}h u_\theta u_\theta^+ \nabla w) = g, \quad g = -u_\theta \mathbb{E}(u). \quad (MHE)$$

**Lemma 16.** Suppose $(u, u^+) \in \mathcal{D}$ for $|Im\theta| < r$ and $|u_\theta| < N$, $|u_\theta^{-1}| < N$, then (MHE) has a unique solution $w \in \mathcal{W}_\rho$ of zero average for all $0 < \rho < r$ and $v = u_\theta w$ satisfying

$$|v|_\rho \leq \frac{c}{(r - \rho)^6} \|\mathbb{E}(u)\|_r, \quad |v_\theta|_\rho \leq \frac{c}{(r - \rho)^7} \|\mathbb{E}(u)\|_r,$$

where $c = c(M, N, R)$.

**Proof.** Denote by $p = (\partial_{12}h u_\theta u_\theta^+)^{-1}$ for short, then

$$\nabla^* \psi = g, \quad p^{-1}\nabla w = \psi + \mu$$

where $\mu$ is a constant equal to $-\frac{\int p^* d\theta}{\int p^*}$, which implies

$$|\mu| \leq R MN^4 c \frac{|g|_r}{(r - \rho)^3}.$$
Due to Lemma 17, $|\psi|_{r'} \leq \frac{c(\gamma)}{(r - r')^2} |g|_r$, for all $0 < r' < r$, then

$$|w|_{\rho} \leq c_1 \frac{|p(\psi + \mu)|_{r'}}{(r' - \rho)^3} \leq \frac{c_2 |g|_r}{(r' - \rho)^3(r - r')^3}.$$

Let $r' = \frac{r + \rho}{2}$, then $|w|_{\rho} \leq \frac{c_3 \|E(u)\|_r}{(r - \rho)^3}$. Recall that $|v| \leq |w| \cdot |u_\theta|$, then

$$|v_\theta|_{\rho} \leq \frac{c_4 \|E(u)\|_r}{(r - \rho)^7}.$$

\[ \square \]

**Lemma 17** (Deease of domain of analyticity). $w$ is a Diophantine frequency, $g \in \mathbb{W}_r$ of zero mean, then $\nabla^\ast \psi(\theta) = g(\theta)$ has a unique solution $\psi \in \mathbb{W}_{r'}$ of zero mean, $\forall 0 < r' < r$. Moreover,

$$|\psi|_{r'} \leq c_2(\gamma) \frac{|g|_r}{(r - r')^3}. \quad (21)$$

**Proof.** The Fourier coefficients of $g$ and $\psi$ are related via the following:

$$\psi_n = \frac{g_n}{e^{i2\pi nw} - 1} \neq 0, \ \psi_0 = 0.$$

The Diophantine condition (19) gives the lower bound on the denominators:

$$|e^{i2\pi nw} - 1| > c(\gamma) \frac{1}{|n|^2},$$

while $g \in \mathbb{W}_r$ implies

$$|g_n| \leq |g|_r e^{-2\pi |n|r}.$$

Using previous inequalities we obtain, for $0 < r' < s < r$,

$$|\psi_n| \leq |g|_r c^{-1}(\gamma) e^{-2\pi r|n|} |n|^2 = |g|_r c^{-1}(\gamma) e^{-2\pi s|n|} e^{2\pi(r-s)|n|} |n|^2$$

We estimate the right hand side by using the estimate $xe^{-x} \leq e^{-1}$ for positive $x$, which for $x = a|n|/b$ gives the inequality

$$e^{-a|n|} |n|^b \leq e^{-b(b/a)^b}$$

for all $n$; with $a = 2\pi(r - s)$ and $b = 2$ we obtain the result

$$|\psi_n| \leq c_1(\gamma) |g|_r \frac{1}{(r - s)^2} e^{-|n|2\pi s}.$$
From this estimate we obtain
\[
|\psi|_{r'} \leq \sum |\psi_n|e^{ln|2\pi'|} \leq \frac{2c_1|g|^r}{(r-s)^2} (1 - e^{-2\pi(s-r'))}^{-1} \leq \frac{2c_1|g|^r}{(r-s)^2(s-r')}
\]
which for \( s = (r+r')/2 \) gives the desired estimate (21) for \( \psi \). This completes the proof.

Lemma 18. For \( u(\theta) \) satisfying all the conditions of Lemma 16, and \( \tilde{u} = u + v \) satisfying \( (\tilde{u}, \tilde{u}^+) \in \mathcal{D} \) for all \( |Im\theta| < \rho \) for \( 0 < \rho < r \), then
\[
|E(\tilde{u})|_{\rho} \leq c^* \|E(u)\|_{r}^2 \frac{r}{(r-\rho)^8},
\]
where \( c^* = c^*(M,N,\gamma) \).

Proof. With \( u \) as specified, we estimate the error
\[
|E(\tilde{u})|_{\rho} \equiv |E(u + v)|_{\rho} = |E(u) + E'(u)v + Q|_{\rho}
\]
with \( Q \) being a quadratic remainder. From (20) we obtain
\[
u \theta E(u) + u \theta E'(u)v = vE'(u)u \theta,
\]
or \( E(u) + E'(u)v = wE'(u)u \theta = w \frac{d}{d\theta} E(u) \). Using Lemma 16 and the Cauchy estimate \( \left| \frac{d}{d\theta} E(u) \right|_{\rho} \leq c \frac{E(u)}{r-\rho} \), we obtain
\[
|E(u) + E'(u)v|_{\rho} \leq c_4 \frac{E(u)}{r-\rho} < c_4 \frac{E(u)}{(r-\rho)^8},
\]
where \( c_4 = c_4(M,N,\gamma) \). Moreover, the error \( Q \) is quadratically small in terms of \( |E(u)|_{r} \). Indeed, by Taylor’s formula we have for some \( 0 \leq t \leq 1 \):
\[
Q = \frac{1}{2} \frac{d^2}{dt^2} E(u + tv).
\]
Then we have
\[
|Q|_{\rho} \leq c_5 |v|_{\rho}^2,
\]
where \( c_5 \) depends only on \( |h|_{C^3} \). From this, we obtain the desired estimate.
Proof of Theorem 15: Benefit from Lemma 16, Lemma 17 and Lemma 18, the strategy at the beginning of this section is available now. So we can repeat this iteration for infinite times, with the corresponding parameters $r_n = \frac{r}{2} + 2^{-n} \frac{r}{2}$, and 

$$\varepsilon_{n+1} = |\mathbb{E}(u_{n+1})|_{r_{n+1}} \lesssim \varepsilon_n^2 c(M, N, \gamma, r) 2^{16n}.$$ 

By taking $\eta_n = 2^{16(n+1)} c \varepsilon_n$ we get $\eta_{n+1} \leq \eta_n^2$. Once $\eta_0 = 2^{16} \varepsilon_0 < 1$, then 

$$\eta_n \leq \eta_0^{2^n} \to 0$$ 

as $\varepsilon_n \to 0$.

6 Arnold’s Example and Normally hyperbolic invariant cylinder (NHIC)

Let’s start from the following example:

Example 11 (Arnold). For $\varepsilon \ll k \ll 1$, 

$$H_{\varepsilon}(x, t) = \frac{p^2}{2} + k(\cos 2\pi q - 1) + \frac{r^2}{2} + \varepsilon H_1(x, t)$$

where $x = (p, q, r, \theta)$ with $(q, \theta) \in \mathbb{T}^2$, $(p, r) \in \mathbb{R}^2$ and $H_1(x, t)$ is 1-periodic of time $t$.

For $\varepsilon = 0$, $H_{\varepsilon}$ is totally integrable, i.e. the phase space $x \in T^*\mathbb{T}^2$ is a product of individual ‘rotator’ and ‘pendulum’ systems. We can easy find that $(p, q) = (0, Z)$ are a list of copies of saddle points, with the heteroclinic orbits $\mathcal{H}_0^+(t) = (p^+(t), q^+(t))$ satisfying 

$$\frac{(p^\pm)^2(t)}{2} + k(\cos 2\pi q\pm(t) - 1) = 0.$$ 

Moreover, $\mathcal{H}_0^+ \to (0, 1)$ as $t \to +\infty$, and $\mathcal{H}_0^+ \to (0, 0)$ as $t \to -\infty$; $\mathcal{H}_0^- \to (0, 1)$ as $t \to -\infty$, and $\mathcal{H}_0^- \to (0, 0)$ as $t \to +\infty$.

******************************************************************************

**Goal:** For any $A < B$ and a generic perturbation $\varepsilon H_1$, we can always find an orbit $x(t)$, such that $r(0) < A$, $r(T) > B$ for some $T > 0$. (AD)
Hypothesis (1). \( H_1(0, 0, r, \theta, t) \equiv 0, \) for all \((r, \theta, t) \in \mathbb{R} \times \mathbb{T}^2, \) then

\[
\Lambda_0 = \{(p, q, t) = (0, 0, 0)\}
\]
is a 2-dimensional invariant cylinder.

Consider the Poincaré map \( F_\varepsilon : x \rightarrow x', \) restricted on the time section \( \Sigma_0 = \{t = 0\}. \) It is given by the time one map of \( H_\varepsilon. \)

Notice that when \( F_\varepsilon \) is restricted to the cylinder \( \Lambda_0, \) it has a family of invariant curves with either periodic or quasiperiodic motion in each, i.e.

\[
F_\varepsilon|_{\Lambda_0} : (r, \theta) \rightarrow (r, \theta + r).
\]
Let \( X \) be an invariant set, i.e. \( F_\varepsilon(X) = X \) given by either the cylinder \( \Lambda_0 \) or an invariant circle \( T_\omega = \Lambda_0 \cap \{r = \omega\}. \) Define the stable (unstable) manifold by

\[
W^s_\varepsilon(X) = \{x| F_\varepsilon^n x \rightarrow X \text{ as } n \rightarrow +\infty\},
\]

\[
W^u_\varepsilon(X) = \{x| F_\varepsilon^n x \rightarrow X \text{ as } n \rightarrow -\infty\}.
\]

Remark. For \( \varepsilon = 0, \) the invariant manifolds of the cylinder coincide and are given by 3-dimensional manifolds

\[
W^s(\Lambda_0) = W^u(\Lambda_0) = \{x| \frac{p^2}{2} + k(cos 2\pi q - 1) = 0\}.
\]
The invariant manifolds of the invariant circle \( T_\omega \) also coincide and are given by 2-dimensional manifolds

\[
W^s(T_\omega) = W^u(T_\omega) = \{x| \frac{p^2}{2} + k(cos 2\pi q - 1) = 0, \ r = \omega\}.
\]

For \( \varepsilon > 0 \) the invariant manifolds of \( \Lambda_0 \) (resp. \( T_\omega \)) no longer need to coincide and are immersed 3-dimensional (resp. 2-dimensional) manifolds.

Arnold’s Mechanism:

Definition 19. Call a collection of invariant curves \((T_{\omega_0}, \cdots, T_{\omega_n})\) transition chain if

\[
W^u(T_{\omega_i}) \cap W^s(T_{\omega_{i+1}}) \quad \forall i = 0, \cdots, n - 1,
\]
i.e. the intersection is non-empty and transverse.
Lemma 20. If there exists a transition chain such that $\omega_0 < A$ and $\omega_n > B$, then there exist orbits satisfying (AD).

Remark. In the original paper of Arnold, he choose

$$H_1 = (1 - \cos 2\pi q)(\cos 2\pi \theta + \cos 2\pi t)$$

of which the transition chain indeed exists. Besides, for $w_i$ away from 0, $|\omega_i - \omega_{i+1}| \lesssim e^{-\sqrt{\varepsilon}}$.

Now we observe that to prove the Goal, we should find the transition chain first. For general $H_1$ with $\varepsilon \neq 0$, Hypothesis (1) no longer holds, which urges us to find a deformative invariant cylinder $\Lambda_\varepsilon$ close to $\Lambda_0$, then try to find the transition chain of the invariant curves $T_{w_i}$ on it.

6.1 Existence of NHIC and Conley isolating block

Definition 21 (Fenichel, HPS). Let $F : M \rightarrow M$ be a $C^1$ diffeomorphism. We call $\Lambda \subset M$ a normally hyperbolic invariant manifold (NHIM), if $\Lambda = F(\Lambda)$ and there exists a decomposition of the bundle space, i.e.

$$TM = T\Lambda \bigoplus E^u \bigoplus E^s,$$

such that $\exists C > 0$, $0 < \lambda < \mu < 1$ satisfying for all $x \in \Lambda$,

$$v \in E^s_x \iff |dF^n(x)v| \leq C\lambda^n|v|, \forall n \geq 0,$$

$$v \in E^u_x \iff |dF^n(x)v| \leq C\lambda^n|v|, \forall n \leq 0,$$

$$v \in T_x\Lambda \iff |dF^n(x)v| \leq C\mu^{-|n|}|v|, \forall n \in \mathbb{Z}.$$

Example 12. For $\varepsilon = 0$, $M = T^*\mathbb{T}^2$, $\Lambda = \Lambda_0 = \{x|p = q = 0\}$, then we can see that $E^s$ is tangent to the stable direction of the saddle and $E^u$ is tangent to the unstable direction of the saddle. Moreover,

$$\lambda = \exp(\text{negative eigenvalue of the saddle}),$$

$$\mu = 1 + C\varepsilon, \quad C = C(H_1) \text{ is a constant}.$$

Theorem 22. For all $\varepsilon H_1$ be $C^3$ small perturbation, system $H_\varepsilon$ preserves a $2 -$dimensional NHIC $\Lambda_\varepsilon$ close to $\Lambda_0$.  

31
Remark. This Theorem is a special application of HPS’s invariant manifold theorem in our setting. From the definition of NHIM, we can see that at least we need $\Lambda$ be $C^1$ sub manifold (if it exists). This implies that we need to prove the existence of topologically invariant set first (isolating block), then raise the smoothness in the tangent bundle space (cone condition).

6.1.1 Conley Isolating Block

For a $C^1$ differential equation
\[ \dot{x} = F(x), \quad x \in \mathbb{R}^n, \tag{22} \]
we can decompose $\mathbb{R}^n$ by $\mathbb{R}^{n_u} \times \mathbb{R}^{n_s} \times \mathbb{R}^{n_c}$, of which $F = (F^s, F^u, F^c)$. We choose $\Omega = B^u \times B^s \times B^c$, where $B^* (s = u, s, c)$ are all Euclidean balls. Assume $|L(x)|$ is bounded with
\[
L(x) = dF(x) = \begin{pmatrix}
L_{uu} & L_{us} & L_{uc} \\
L_{su} & L_{ss} & L_{sc} \\
L_{cu} & L_{cs} & L_{cc}
\end{pmatrix}.
\]
We call such a $\Omega$ isolating block, if the following Hypothesis holds:

Hypothesis (2). Isolating Block
\[
F^c = 0 \text{ on } B^u \times B^s \times \partial B^c,
\]
\[
|F_u(u, s, c)u| > |u| \text{ on } \partial B^u \times B^s \times B^c,
\]
\[
|F_s(u, s, c)s| < |s| \text{ on } B^u \times \partial B^s \times B^c.
\]

Theorem 23 (Existence). If $\Omega$ is an isolating block, there exist a topological invariant set in $\Omega$.

Proof. We just need to take $\cap_{i \in \mathbb{Z}} F^i(B^u \times B^s \times B^c)$ and Hypothesis (2) ensures the intersection is not empty. \qed

Hypothesis (3). Cone Condition For all $x \in \Omega$,
\[
L_{uu}(x) \geq \alpha I, \quad L_{ss}(x) \leq -\alpha I,
\]
\[
\sum_{ij} |L_{ij}(x)| \leq m, \quad ij \neq ss, uu.
\]
Exercise 12. Let $B^c = 0$, Prove that $F$ satisfying Hypothesis (2) and (3) with $\alpha > 2m$ will preserve a saddle point in $B^u \times B^s$.

Theorem 24. Assume Hypothesis (2) and (3) hold, then for $K = \frac{m}{\alpha - 2m} \leq \frac{1}{\sqrt{2}}$, $\mathbb{W}^{sc} = \{x|x(t) \in \Omega, \forall t \geq 0\}$ is a graph of $C^1$ function $W^{sc}$; $\mathbb{W}^{uc} = \{x|x(t) \in \Omega, \forall t \leq 0\}$ is a graph of $C^1$ function $W^{uc}$; $\mathbb{W}^{c} = \{x|x(t) \in \Omega, \forall t \in \mathbb{R}\}$ is a graph of $C^1$ function $W^{c}$; Here $W^{sc} : B^s \times B^c \rightarrow B^u$, $W^{uc} : B^u \times B^c \rightarrow B^s$ and $W^{c} : B^c \rightarrow B^u \times B^s$. Moreover, $|dW^{sc}| \leq K$, $|dW^{uc}| \leq K$ and $|dW^{c}| \leq 2m$.

Corollary 25. Theorem 24 leads to Theorem 22 for small $\varepsilon$, and we can estimate how small $\varepsilon$ is !!!

Lemma 26. If Hypothesis (3) holds, then

$$\pi^{sc}\mathbb{W}^{sc} = B^s \times B^c; \quad <1>$$
$$\pi^{uc}\mathbb{W}^{uc} = B^u \times B^c; \quad <2>$$

Moreover, the closure $\overline{\mathbb{W}}^{sc}$ and $\overline{\mathbb{W}}^{uc}$ satisfies

$$\overline{\mathbb{W}}^{sc} \subset B^u \times B^s \times B^c, \quad \overline{\mathbb{W}}^{uc} \subset B^s \times B^u \times B^c. \quad <3>$$

Here $\pi^* (\ast = s, u, c)$ is the standard projection to each subspace.

Proof. $T^+(x) = \min\{t > 0, x(t) \in \partial\Omega\}$. If $T^+(x) < \infty$, $x \notin W^{sc}$. By Hypothesis (3), $x(T^+(x)) \in \partial B^u \times B^s \times B^c$, which implies $T^+(x)$ is locally continuous and is $C^1$ dependent of the initial position $x$. Then we can define the retraction:

Definition 27. We call the map $f : X \rightarrow Y \subset X$ a retraction, if $f|_Y = id$.

We can see that there exists no retraction from $B^u$ to $\partial B^u$. To prove $<1>$, we assume that $(s, c) \in B^s \times B^c$ such that $\mathbb{W}^{sc} \cap B^u \times \{s\} \times \{c\} = \emptyset$, which implies

$$\pi^{sc}\mathbb{W}^{sc} \cap (s, c) = \emptyset.$$ 

Then we actually find a retraction $u \in B^u \rightarrow \pi^u x(T^+(u, s, c)) \in \partial B^u$, which leads to a contradiction. $<2>$ can be proved in a similar way.

For $<3>$, notice that in a small neighborhood of $\partial B^u$, $T^+(x) < \infty$. Then $\overline{\mathbb{W}}^{sc} \cap (N(\partial B^u) \times \bar{B}^s \times \bar{B}^c) = \emptyset$. This leads to $\overline{\mathbb{W}}^{sc} \subset B^u \times \bar{B}^s \times \bar{B}^c$. The other cases follows the same way. 

$\square$
6.1.2 the cone condition and graph transform

Recall that previous Theorem 23 ensures the existence of a topological invariant set in $\Omega$, but we need to prove the smoothness (at leaf $C^1$) then get Theorem 24.

Besides, we can see that the topological invariant set $W_c$ is actually the intersection of the following forward invariant and backward invariant sets:

$$W_{sc} = \cap_{i \in \mathbb{Z}^+} F^i(\Omega);$$
$$W_{uc} = \cap_{i \in \mathbb{Z}^-} F^i(\Omega);$$
$$W_c = W_{sc} \cap W_{uc}.$$

Benefit from Hypothesis (3), we can prove that $W_{uc}$ (resp. $W_{sc}$) are $C^1$ smooth and then $W_c$ is also $C^1$: For a point $z \in \Omega$ define a map from $T_z \Omega$ to its neighborhood $U \subset \Omega$ by considering the exponential map $\exp_z(v) \to \Omega$. By definition $\exp_z(0) = z$ and for a unit vector $v$ we have $\exp_z(tv)$ to be the position of geodesic starting at $z$ and for a unit vector $v$ after time $t$. For the Euclidean components we assume that the metric is flat and the corresponding exponential map is linear.

Now let’s denote by the following unstable (resp. stable) cone in $T_z \Omega$ by

$$K^u(z) = \{(u, s, c) | |u| \geq |s| + |c|\};$$
$$K^s(z) = \{(u, s, c) | |s| \geq |u| + |c|\};$$

with $z \in \Omega$ and $(u, s, c) \in T_z \Omega$ is the coordinate corresponding to the components of $\Omega = B^u \times B^s \times B^c$.

**Lemma 28.** For $z \in W_{uc}$, $K^s(z)$ is backward invariant of the flow (22). Respectively, for $z \in W_{sc}$, $K^u(z)$ is forward invariant.

**Remark.** Due to the exponential map, we can get that $W_{uc}$ and $W_{sc}$ are both Lipschitz graphs. Actually, we can automatically raise the smoothness to $C^1$, this is because the topological invariant set $W_c$ is unique and every Lipschitz function is almost everywhere differentiable.

At last, let’s point out that due to the invariant theorem we can continue to raise the smoothness of $W_c$, see [2] for more details.
Large Gap problem and weak KAM theory

In last section, Theorem 24 implies NHIC could preserve for sufficiently small $\varepsilon$, even if Hypothesis (1) is not available. Then due to Arnold’s mechanism, we need to find a transition chain first, then have chance to construct a diffusion orbit.

However, the foliation structure of NHIC will be destructed by generic $H_1$, and the KAM theorem only ensures that a sparse set of KAM tori preserves, although they may occupy a large measure part.

Example 13. Verify that the greatest $p$-deviation between any two points in the zero energy level of $H = \frac{p^2}{2} + \varepsilon(\cos q - 1)$ with $(p, q) \in T^*T$ is $2\sqrt{\varepsilon}$.

Previous example implies that under a generic perturbation of the NHIC, there would be a Birkhoff unstable domain of the width $O(\sqrt{\varepsilon})$, i.e. the two adjacent KAM tori on the NHIC will be $O(\sqrt{\varepsilon})$ faraway. $|T_w - T_{w'}| \gtrsim \sqrt{\varepsilon}$ in turns implies that $|w - w'| \gtrsim \sqrt{\varepsilon}$.

On the other side, the Melnikov function implies that $W^u(T_w) \cap W^s(T_{w'}) \neq \emptyset$ holds only for $|w - w'| \lesssim \varepsilon$ (resp. $W^s(T_w) \cap W^u(T_{w'})$). Here $\varepsilon \ll \sqrt{\varepsilon}$ cause a so called ‘Large Gap Problem’, i.e. we couldn’t use the Arnold’s mechanism to overcome the Birkhoff unstable domain.

Goal: For any $w \in [w', w'']$ with $T_{w'}$, and $T_{w''}$ are two adjacent non contractible invariant circles, we can fill the gap with invariant set $A_w$ (Aubry Mather set) of the rotational number $w$, such that $W^u(A_{w_k}) \cap W^s(A_{w_{k+1}}) \neq \emptyset$ (resp. $W^s(A_{w_k}) \cap W^u(A_{w_{k+1}}) \neq \emptyset$) for at least one sequence $\{w_i\}_{i=1}^N \subset [w', w'']$. This $\{w_i\}$ will form a transition chain and we can now across the Birkhoff unstable domain.

7.1 weak KAM theory

Definition 29. Suppose $M$ is a smooth, compact manifold without boundary, thence call a Hamiltonian $H : T^*M \to \mathbb{R}$ be Tonelli, if
• $H(x, p)$ is at least $C^2$ smooth;

• $H(x, p)$ is fibre convex, i.e. $\partial^{2}_{pp} H$ is positively definite for all $(x, p) \in T^* M$;

• $H(x, p)$ is superlinear, i.e. $\lim_{|p| \to \infty} \frac{H(x, p)}{|p|} = \infty$.

Example 14. • (Geodesic Flow) $H(x, p) = \frac{1}{2} \|p\|^2$;

• (Mechanical) $H(x, p) = \frac{1}{2} \|p\|^2 + U(x)$;

• (Mané’s Hamiltonian) for any vector field $X(x)$ on $M$, $H(x, p) = \frac{1}{2} |p|^2 + \langle p, X(x) \rangle$.

Definition 30. Suppose $M$ is a smooth, compact manifold without boundary, thence call a Lagrangian $L : TM \to \mathbb{R}$ be Tonelli, if

• $L(x, v)$ is at least $C^2$ smooth;

• $L(x, v)$ is fibre convex, i.e. $\partial^2_{vv} L$ is positively definite for all $(x, v) \in TM$;

• $L(x, v)$ is superlinear, i.e. $\lim_{|v| \to \infty} \frac{L(x, v)}{|v|} = \infty$.

This Lagrangian is highly related with the following variation problem: denote by

$$A_L(x, x', a, b) = \int_{a}^{b} L(\gamma(s), \dot{\gamma}(s)) ds$$

the Action function of the $C^1$-piecewise path $\gamma : [a, b] \to M$ with fixed endpoints $\gamma(a) = x, \gamma(b) = x'$, the following Lemma holds:

Lemma 31. The critical value of $A_L(x, x', a, b)$ corresponds to curves $\gamma$ satisfying the Euler Lagrange equation:

$$\frac{d}{dt} L_v(\gamma, \dot{\gamma}) = L_x(\gamma, \dot{\gamma}), \quad (EL)$$

for all $t \in [a, b]$.

Remark. For autonomous Lagrangian $L(x, v)$, $(EL)$ is complete, i.e. any solution of $(EL)$ can be expanded for all $t \in \mathbb{R}$. If $L(x, v, t)$ is time periodic, then we have to add the following item in the definition of Tonelli Lagrangian:
• (EL) flow is complete for all \( t \in \mathbb{R} \).

**Theorem 32** (Legendre Transformation). *There is an one to one correspondence between Tonelli Hamiltonian and Tonelli Lagrangian:*

\[
L(x,v) = \max_{p \in T^*M} \{ \langle p, v \rangle - H(x,p) \}. \quad (*)
\]

Respectively, \( H(x,p) = \max_{v \in TM} \{ \langle p, v \rangle - L(x,v) \} \).

Recall that \( \mathcal{L} : (x,p) \to (x,v) \) is a diffeomorphism due to the convexity of \( H(x,p) \) and actually \( v = \partial_p H(x,p) \). We call \( \mathcal{L} \) the Fenichel Legendre transformation. Due to this Legendre transformation, we get:

**Example 15.**

• *(Geodesic Flow)* \( L(x,v) = \frac{1}{2} \|v\|^2 \);

• *(Mechanical)* \( L(x,v) = \frac{1}{2} \|v\|^2 - U(x) \);

• *(Mané’s Lagrangian)* \( L(x,v) = \frac{1}{2} \|v - X(x)\|^2 \).

**Lemma 33.** \( \mathcal{L} \) sends an orbit of \( (H) \) into an orbit of \( (EL) \).

**Definition 34.** We call \( \mathcal{T}_w \) a (maximal) KAM torus with \( w \in \mathbb{R}^n \), if

• \( \mathcal{T}_w = \{(x,c + du(x))\} \) is a Lagrangian graph;

• \( \mathcal{T}_w \) is invariant of the Hamiltonian flow;

• the flow on \( \mathcal{T}_w \) is conjugated to the rigid flow \( \dot{x} = w \) (or \( x \to x + wt \) (mod 1)).

**Remark.** If \( \langle k,w \rangle \neq 0 \) for all \( k \in \mathbb{Z}^n \setminus \{0\} \), then any orbit on \( \mathcal{T}_w \) is dense. Moreover, due to the Birkhoff ergodic Theorem, there exists a unique invariant probability measure \( \mu \) which supports on \( \mathcal{T}_w \) and \( \mu = \varphi^* \text{Leb} \) with \( \text{Leb} \) is the Lebesgue measure on \( \mathbb{T}^n \) and \( \varphi \) conforms to the commutative diagram:

\[
\begin{array}{ccc}
\mathbb{T} & \xrightarrow{\varphi^* \mu} & \mathbb{T} \\
\uparrow \varphi & & \uparrow \varphi \\
\mathbb{T} & \xrightarrow{\rho^*_w} & \mathbb{T}.
\end{array}
\]

**Theorem 35** (Herman). *If \( \langle k,w \rangle \neq 0 \) for all \( k \in \mathbb{Z}^n \setminus \{0\} \), any KAM torus has to be a maximal KAM torus.*
Definition 36. Denote by $\mathcal{M}$ the space of all probability invariant measures on $TM$, then we define by

$$w = [\mu] = \int_{TM} v d\mu$$

the rotational vector of $\mu$.

Remark. If $\mu$ is ergodic, then due to the Birkhoff ergodic Theorem, we can find a (EL) curve $\gamma$, such that

$$\lim_{T \to \infty} \frac{1}{T} \int_{0}^{T} \dot{\gamma}(t) dt = \int_{TM} v d\mu = [\mu].$$

So essentially we can see that $w$ is the averaging velocity of the ergodic curve $\gamma$.

Exercise 13. For integrable system $H(x, p) = \frac{1}{2}p^2$, the probability measure $\mu$ supported on $T_w$ has the rational vector $w$.

Theorem 37. The unique measure (invariant, probability) supported on $T_w$ satisfies

$$\int L d\mu = \inf_{[\mu] = w} \int L d\mu.$$

Now we expand previous idea: for any $\rho \in \mathbb{R}^n$, can we find an invariant probability measure $\mu_\rho$, such that $[\mu_\rho] = \rho$?

In the universal cover of $T^n$, for any fixed point $x_0$, we introduce the following variational principle:

$$A(\rho, x_0, T) := \min_{\gamma(0) = x_0, \gamma(T) = x_0 + \rho T} \int_{0}^{T} L(\gamma(s), \dot{\gamma}(s), s) ds,$$

where ‘min’ is for all the absolute continuous curves.

Remark. If $L(x, v, t)$ is time-periodic, we need to impose the ‘completeness’ of (EL). Due to the Tonelli Theorem, the minimizer of $A(\rho, x_0, T)$ is always achievable and we can improve the smoothness of the minimizer $\gamma$ to $C^2([0, T], T^n)$. 
Suppose $\mu_T$ is the singular probability measure supported on $\{(\gamma(s), \dot{\gamma}(s)) | s \in [0, T]\}$, i.e., for all function $f(x, v)$ with compact support in $T\mathbb{T}^n$,
\[
\int_{T\mathbb{T}^n} f \, d\mu_T = \frac{1}{T} \int_0^T f(\gamma(s), \dot{\gamma}(s)) \, ds.
\]
Recall that $\mu_T$ is not invariant, but indeed
\[
\rho = \frac{\gamma(T) - \gamma(0)}{T} = \frac{1}{T} \int_0^T \dot{\gamma}(s) \, ds = \int v \, d\mu_T.
\]
Due to Krylov-Bogolybov Theorem, the space of all probability measures on $T\mathbb{T}^n$ is weak* compact, which implies $\lim_{T \to \infty} \mu_T$ exists, and each limit measure $\mu_\infty$ is invariant with
\[
\int v \, d\mu_\infty = \rho.
\]
Recall that if $\mu_\infty$ is ergodic, we can use the Birkhoff ergodic theorem and find at least one $(\gamma, \dot{\gamma}) \in \text{supp} (\mu_\infty)$, such that
\[
\lim_{T \to \infty} \frac{1}{T} \int_0^T \dot{\gamma}(t) \, dt = \int_{TM} v \, d\mu_\infty = \rho. \tag{23}
\]
If $\mu_\infty$ is not ergodic, the following example shows that we are not able to find such a supported orbit $\gamma$ such that (23) holds:

**Example 16 (Hedlund).** We choose a Riemannian Hamiltonian $H(q, p) = \frac{1}{2} \|p\|_h^2$, where $(q, p) \in T^*\mathbb{T}^3$ and $x \in \mathbb{R}^3$ is the lift of $q \in \mathbb{T}^3$. Along $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$, we construct 3 'high way' $T_j$, $j = 1, 2, 3$, i.e., the Riemannian metric $\langle \cdot, \cdot \rangle_h$ is sufficiently small; outside $T_j$ the metric is Euclidean, see Fig. 3.

**Exercise 14.** Use the Legendre transformation we can get $L(x, v) = \frac{1}{2} \|v\|_h^2$. Prove that for $\rho = (1, 1, 1)$, we can find an invariant measure $\mu$ satisfying $[\mu] = \rho$ can be decomposed into
\[
\mu = \frac{\mu_1 + \mu_2 + \mu_3}{3}
\]
with $\mu_j$ is a measure supported in $T_j$ with $[\mu_j] = e_j$, $j = 1, 2, 3$. 

Lemma 38. $f : \mathbb{T}^n \rightarrow \mathbb{R}$ is $C^1$, then $\int_{\mathbb{T}^n} \langle df, v \rangle \, d\mu = 0$ for all invariant measure $\mu$.

Proof. Let $(x_0, v_0) \in \text{supp}(\mu)$, and $(x_t, v_t)$ be the (EL) flow.

$$
\int \langle df, v \rangle \, d\mu = \frac{1}{T} \int_0^T dt \int_{\mathbb{T}^n} \langle df, v \rangle \, d(\varphi^t_0)_* \mu
$$

$$
= \lim_{T \to \infty} \frac{1}{T} \int_0^T dt \int_{\mathbb{T}^n} \langle df, v \rangle \bigg|_{(x_t, v_t)} \, d\mu
$$

$$
= \lim_{T \to \infty} \frac{1}{T} \int_{\mathbb{T}^n} d\mu \int_0^T df(x_t)v_t dt
$$

$$
= \lim_{T \to \infty} \int_{\mathbb{T}^n} d\mu \frac{f(x_T) - f(x_0)}{T}
$$

$$
= 0.
$$

So we finish the proof. \qed

Now we turn back to the (maximal) KAM torus: Suppose the frequency is $w$ and $\mu_w$ is the corresponding invariant measure on $\mathcal{T}_w$, then

$$
\int_{\mathbb{T}^n} L(x, v) d\mu_w = \int_{\mathbb{T}^n} L(x, v) - \langle c + du, v \rangle d\mu_w + \int_{\mathbb{T}^n} \langle c, v \rangle d\mu_w
$$
because of Lemma 38. By the Legendre transformation and $H(x, c + du(x)) = \alpha(c)$,
\[
\int_{T^\mathbb{T}} L(x, v) d\mu_w = \langle c, w \rangle - \alpha(c).
\]

**Proposition 39.** For any other invariant measure $\tilde{\mu}$, we always have
\[
\int_{T^\mathbb{T}} L - c \cdot v d\tilde{\mu} \geq \int_{T^\mathbb{T}} L - c \cdot v d\mu_w.
\]

**Proof.** Due to the Fenichel Legendre Inequality we get
\[
H(x, c + du(x)) + L(x, v) \geq \langle v, c + du \rangle.
\]
Define by $L_c(x, v) := L(x, v) - c \cdot v$, then
\[
\int_{T^\mathbb{T}} L_c \, d\tilde{\mu} \geq -\alpha(c)
\]
and we finish the proof. \qed

**Definition 40.** We define the $\alpha-$function by
\[
\alpha(c) = -\min_{\mu \in \mathcal{M}_{inv}} \int_{T^\mathbb{T}} L_c \, d\mu
\]
where $\mathcal{M}_{inv}$ is the set of all invariant measures. We call the minimizers of (24) $c-$minimal measures. We call $\mathcal{M}_L(c) := \cup \text{supp}(\mu_c)$ the $c-$Mather set, where `$\cup$' is taken for all $c-$minimal measures.

**Corollary 41.** For any invariant measure $\tilde{\mu}$ with $\rho$ being the rotation vector,
\[
\int L d\tilde{\mu} \geq \int L d\mu_w
\]
with $\mu_w$ supported on $T_w$.

**Proposition 42.** Suppose $\gamma$ is an orbit on $T_w$, then for all $T > 0$, we have
\[
\inf_{\begin{subarray}{c}
\xi(0) = \gamma(a) \\
\xi(T) = \gamma(b)
\end{subarray}} \int_0^T L(d\xi(t)) - c\dot{\xi}(t) + \alpha(c)dt \geq \int_a^b L(d\gamma(t)) - c\dot{\gamma}(t) + \alpha(c)dt,
\]
i.e. for any interval $[a, b]$, $\gamma$ restricted on it is a free time minimizer.
Proof.

\[
\int_0^T L(d\xi(t)) - c\dot{\xi}(t) + \alpha(c)dt = \int_0^T L(d\xi(t)) - (c + du(\xi))\dot{\xi}(t) + \alpha(c)dt \\
+ u(\xi(T)) - u(\xi(0)) \\
\geq u(\gamma(b)) - u(\gamma(a))
\]

because \(L(d\xi(t)) - (c + du(\xi))\dot{\xi}(t) + \alpha(c) \geq 0\) due to the Fenichel Legendre inequality. On the other side, we know that

\[
\int_a^b L(d\gamma(t)) - c\dot{\gamma}(t) + \alpha(c)dt = u(\gamma(b)) - u(\gamma(a))
\]

because \(H(x, c + du(x)) = \alpha(c)\).

\[
\begin{align*}
\text{Definition 43.} & \quad \text{For fixed } x, y \in \mathbb{T}^n, k \in \mathbb{R} \text{ and } c \in \mathbb{R}^n, \\
& \text{the Mañé Potential function is defined by} \\
& \quad \varphi_{c,k}(x, y) = \inf_{T>0} \int_0^T L_c(d\gamma)dt + kT,
\end{align*}
\]

where \(\inf\) is for all the absolute conti. curve with \(\gamma(0) = x\) and \(\gamma(T) = y\).

Proposition 44.  
- For all \(k \geq \alpha(c)\), \(x, y \in \mathbb{T}^n\), \(\varphi_{c,k}(x, y) > -\infty\); for all \(k < \alpha(c)\), \(x, y \in \mathbb{T}^n\), \(\varphi_{c,k}(x, y) = -\infty\).
- For all \(x, y, z \in \mathbb{T}^n\), \(\varphi_{c,\alpha(c)}(x, y) \leq \varphi_{c,\alpha(c)}(x, z) + \varphi_{c,\alpha(c)}(z, y)\).

Definition 45. We call an (EL) orbit \(\gamma\) **global \(c\)-minimizer** if \(\forall [a, b] \subset \mathbb{R}, \gamma|[a,b]\) is the minimizer of \(A_L(\gamma(a), \gamma(b), a, b)\), and **regular \(c\)-minimizer** if \(\forall [a, b] \subset \mathbb{R}, \gamma|[a,b]\) is the minimizer \(\varphi_{c,\alpha(c)}(\gamma(a), \gamma(b))\). If \(\gamma\) is a regular \(c\)-minimizer satifying

\[
\varphi_{c,\alpha(c)}(\gamma(a), \gamma(b)) + \varphi_{c,\alpha(c)}(\gamma(b), \gamma(a)) = 0,
\]

then we call it **static \(c\)-minimizer**. Usually we use \(\tilde{G}(c), \tilde{N}(c)\) and \(\tilde{A}(c)\) to represent the union of all the \(d\gamma|_{t\in\mathbb{R}}\) where \(\gamma\) is a global, regular, static \(c\)-minimizer relatively.

Theorem 46.  
- \(\tilde{M}(c) \subset \tilde{A}(c) \subset \tilde{N}(c) \subset \tilde{G}(c) \subset \{L_{\tilde{H}}^{-1}(x, p) \subset T\mathbb{T}^n|H(x, p) = \alpha(c)\}\).
\begin{itemize}
  \item Suppose $\pi : T\mathbb{T}^n \to \mathbb{T}^n$ be a standard projection, then $\pi|_{\tilde{A}(c)}$ and $\pi|_{\tilde{M}(c)}$ is injective and $\pi^{-1}|_{\tilde{A}(c)}$ and $\pi^{-1}|_{\tilde{M}(c)}$ is Lipschitz.
\end{itemize}

**Example 17** (Pendulum). For $L(x, v) = \frac{v^2}{2} + (1 - \cos 2\pi x)$ and the conjugated $H(x, p) = \frac{p^2}{2} + (\cos 2\pi x - 1)$, we denote by the energy level 
$$\gamma_{E}^{\pm} := \{(x, v) | v = \pm \sqrt{2(E + (1 - \cos 2\pi x))}\}, \quad E \geq 0.$$ 
Since the pendulum is integrable, so everything is clear and we can use these to illustrate the previous definitions:

\begin{itemize}
  \item $\alpha(0) = 0$ and $\alpha(c) = \alpha(-c)$ for all $c \in \mathbb{R}$;
  \item $\tilde{M}(0) = \tilde{A}(0) = \tilde{N}(0) = \tilde{G}(0) = \{(0, 0)\}$;
  \item $\alpha(c)$ is convex continuous and superlinear;
  \item for all $c \in (-\frac{4}{\pi}, \frac{4}{\pi})$, $\alpha(c) = 0$ and $\tilde{M}(c) = \tilde{A}(c) = \tilde{N}(c) = \tilde{G}(c) = \{(0, 0)\}$;
  \item for $c = \frac{4}{\pi}$, $\tilde{M}(\frac{4}{\pi}) = \{(0, 0)\}$ but $\tilde{A}(\frac{4}{\pi}) = \tilde{N}(\frac{4}{\pi}) = \tilde{G}(\frac{4}{\pi}) = \gamma_{E}^{+}$ and for $c = -\frac{4}{\pi}$, $\tilde{M}(-\frac{4}{\pi}) = \{(0, 0)\}$ but $\tilde{A}(-\frac{4}{\pi}) = \tilde{N}(-\frac{4}{\pi}) = \tilde{G}(-\frac{4}{\pi}) = \gamma_{E}^{-}$.
  \item for $c > \frac{4}{\pi}$, there exists a unique $E_{c} > 0$ such that 
$$c = \int_{0}^{1} \sqrt{2E_{c} + 2(1 - \cos 2\pi q)}dq $$
and $\tilde{M}(c) = \tilde{A}(c) = \tilde{N}(c) = \tilde{G}(c) = \gamma_{E_{c}}^{+}$.
  \item for $c < -\frac{4}{\pi}$, there exists a unique $E_{c} > 0$ such that 
$$c = -\int_{0}^{1} \sqrt{2E_{c} + 2(1 - \cos 2\pi q)}dq $$
and $\tilde{M}(c) = \tilde{A}(c) = \tilde{N}(c) = \tilde{G}(c) = \gamma_{E_{c}}^{-}$.
\end{itemize}

**Definition 47.** The $\beta-$function defined by $\beta : H_{1}(\mathbb{T}^{n}, \mathbb{R}) \to \mathbb{R}$ satisfies 
$$\beta(h) = \inf_{\mu \in \mathcal{M}_{inv}} \int Ld\mu.$$
Lemma 48 (Mather). \( \alpha(c) \) and \( \beta(h) \) are both convex, super linear continuous functions; they are conjugated via

\[
\beta(h) = \max_c \{ \langle c, h \rangle - \alpha(c) \}
\]

and

\[
\alpha(c) = \max_h \{ \langle c, h \rangle - \beta(h) \}.
\]

Example 18. Suppose \( H(p) \) is a Tonelli integrable Hamiltonian and \( L(v) \) is the Legendre transformation, then \( \alpha(c) = H(c) \) and \( \beta(h) = L(h) \).

Since \( \alpha(c) \) and \( \beta(h) \) are convex, so we can define the super differential set by the following:

Definition 49.

\[
\partial \beta(h) := \{ c \in H^1(T^n, \mathbb{R}) | \beta(h') - \beta(h) \geq \langle h' - h, c \rangle, \forall h' \in H^1(T^n, \mathbb{R}) \},
\]

\[
\partial \alpha(c) := \{ h \in H^1(T^n, \mathbb{R}) | \alpha(c') - \alpha(c) \geq \langle c' - c, h \rangle, \forall c' \in H^1(T^n, \mathbb{R}) \}.
\]

Proposition 50.

\[
 c \in \partial \beta(h) \iff \langle c, h \rangle = \alpha(c) + \beta(h) \iff h \in \partial \alpha(c).
\]

Example 19. For previous integrable system \( H(p) \), the conjugated \( c \) and \( h \) satisfy \( c = \nabla L(h) \) and \( h = \nabla H(c) \).

7.2 Arnold’s Example and the Aubry Mather interpretation

Now let’s turn back to the Arnold’s example: \( x = (q, \theta) \in \mathbb{T}^2 \) and \( y = (p, r) \in \mathbb{R}^2 \)

\[
H_\varepsilon(y, x, t) = \frac{p^2}{2} + (\cos 2\pi q - 1) + \frac{r^2}{2} + \varepsilon H_1(y, x, t),
\]

\[
L_\varepsilon(v, x, t) = \frac{v_q^2}{2} - (\cos 2\pi q - 1) + \frac{v_\theta^2}{2} + \varepsilon L_1(v, x, t).
\]

Recall that the Legendre transformation \( \mathcal{L}_H : (p, q, r, \theta, t) \to (p + \varepsilon \partial_p H_1, q, r + \varepsilon \partial_r H_1, \theta, t) \) is a diffeomorphism.

Exercise 15. Verify the existence of \( L_1 \) via \( \mathcal{L}_H \).
Exercise 16. Suppose Hypothesis 1 holds, then

- $L(v_q, v_\theta, q = 0, \theta, t) \equiv 0$.
- $1 - \cos 2\pi q + \varepsilon L_1 \geq 0$ for all $(v, x, t) \in T\mathbb{T}^2 \times \mathbb{T}$ with $|v|$ bounded, and $'=\,'$ holds only for $q = 0$.

Lemma 51. Let $c = (0, c_\theta) \in \mathbb{R}^2$ and Hypothesis 1 holds, then $\tilde{\mathcal{M}}(c) = \tilde{\mathcal{A}}(c) = \tilde{\mathcal{N}}(c) = \tilde{\mathcal{G}}(c) = \{p = q = 0, r = c_\theta, \theta \in \mathbb{T}\}$.

Proof. Because $1 - \cos 2\pi q + \varepsilon L_1 \geq 0$, then

$$L_\varepsilon - cv \geq \frac{v_q^2}{2} + \frac{(v_\theta - c_\theta)^2}{2} - \frac{c_\theta^2}{2},$$

which leads to $-\alpha(c) \geq -\frac{c_\theta^2}{2}$; On the other side, we just need to show that $-\frac{c_\theta^2}{2}$ is achievable, this is due to the action estimate of $\{p = q = 0, r = c_\theta\}$. \hfill \square

**********************************************************************************

Now let’s get rid of Hypothesis 1. Due to the invariant manifold theory, we know that $H_\varepsilon$ still preserves a NHIC $\Lambda_\varepsilon$ which is $\varepsilon-$close to $\Lambda_0$, as long as $\varepsilon$ sufficiently small; So the goal in this part is to prove that for $c = (0, c_\theta) \in \mathbb{R}^2$, $\tilde{\mathcal{A}}(c) \subset \Lambda_\varepsilon$. Before we prove this goal, let’s specify that all the Aubry Mather theory we introduced above can be applied directly to the time-periodic Hamiltonians.

Another observation is that since Hypothesis 1 is not available now, $\Lambda_\varepsilon \cap \{t = 0\}$ becomes a 2-dimensional cylinder and $\varphi^1_{H_\varepsilon}$ becomes a twist map; So generically the foliation structure of invariant tori no longer exists and we need to explore the possible states of the invariant sets for the twist maps.

Benefit from the graphic property of Aubry set and the dimensional restriction in the Arnold’s example, we can easily fix that

Exercise 17. Every orbit in $\tilde{\mathcal{A}}(c)$ will have a unified rotation direction $(\rho, 1)$, where ‘1’ is the free velocity of the time variable.

Proof. Hint: Each orbit in $\tilde{\mathcal{A}}(c)$ will converge to $\tilde{\mathcal{M}}(c)$ in distribution, this implies the averaging velocity, i.e. the rotational vector can be defined; On the other side, if two different orbits in $\tilde{\mathcal{A}}(c)$ have different rotation vector,
then the projected image in the configuration space must intersect with each other, this is because the dimensional restriction. But this would contradict with the graphic property of \( \tilde{A}(c) \).

**Theorem 52** (Poincaré Classification Theorem). Suppose \( f : T \to T \) is a orientation preserving home, and the rotation number is irrational, then there exists a continuous map \( h : T \to T \), such that \( h \circ f = R_\rho \circ h \) with \( R_\rho : x \to x + \rho \mod 1 \).

- If \( f \) is transitive, then \( h \) is also a homeo;
- If \( f \) is not transitive, \( h \) is not invertible (ghost stairway);

In item 1, the recurrent set \( \text{Rec}(f) \) is an invariant curve; In item 2, it’s a no where dense close set, which is the so called Denjoy set.

This theorem can help us to classify the Aubry set in the twist maps:

**Theorem 53.** If \( h \in \mathbb{R} \setminus \mathbb{Q} \), \( \tilde{A}(c) \cap \{t = 0\} \) will be either an invariant curve or a Denjoy set; If \( h = p/q \in \mathbb{Q} \), then \( \tilde{M}(c) \) consists of periodic orbits. Generically there exist \( c_{p/q}^- < c_{p/q}^+ \) such that

- for \( c \in (c_{p/q}^-, c_{p/q}^+) \), \( \tilde{A}(c) = \tilde{M}(c) \);
- for \( c = c_{p/q}^+ \), \( \tilde{A}(c) \setminus \tilde{M}(c) \) consists of \( p/q \)+ homoclinic (heteroclinic) orbits;
- for \( c = c_{p/q}^- \), \( \tilde{A}(c) \setminus \tilde{M}(c) \) consists of \( p/q \)− homoclinic (heteroclinic) orbits;

**Theorem 54.** For \( c = (0, c_\theta) \in H^1(T^2, \mathbb{R}) \) with \( c_\theta \in [c_-, c_+] \), the Arnold example \( H_\varepsilon \) preserves \( \Lambda_\varepsilon \) as long as \( \varepsilon \) sufficiently small; Moreover, \( \tilde{A}_\varepsilon(c) \subset \Lambda_\varepsilon \).

### 7.3 weak KAM interpretation of the transition chain

**Elementary \( \lambda \)-Lemma:** \( f : \mathbb{A} \to \mathbb{A} \) is an area preserving map, with \( p_i = f^{n_i}(p_i) \), \( i = 1, \cdots, N + 1 \) being a collection of saddle periodic points and \( W^u(p_i) \cap W^s(p_{i+1}) = \emptyset \) for all \( i = 1, \cdots, N \). Then there exists an orbit \( \theta = (f_x^k)_{k \in \mathbb{Z}} \) s.t.

\[
f_x^k \to \theta(p_1) = (f_{p_1}^k)_{k \in \mathbb{Z}}
\]
as \( n \to -\infty \) and
\[
f^k_x \to \theta(p_{N+1}) = (f^k_{p_{N+1}})_{k \in \mathbb{Z}}
\]
as \( n \to +\infty \).

Goal: Now we replace the periodic orbits by Aubry sets, we can get a generalized \( \lambda \)-Lemma, which reveals the dynamic essence of the transition chain in the space of pseudo graphs.

Definition 55. Suppose \( G \) is a Lipschitz graph from \( \pi G \subset \mathbb{T}^n \) to \( G \subset T\mathbb{T}^n \), where \( \pi \) is the standard projection. If \( G \) is Lagrangian, there exists a unique \( c \in H^1(\mathbb{T}^n, \mathbb{R}) \) with \( \eta \) be a \( C^1 \) closed one form satisfying \([\eta] = c\), such that
\[
G = G_{\eta,u} = \{(x, \eta_x + du)|x \in \mathbb{T}^n \text{ where } du(x) \text{ exists}\}
\]
We call \( G_{\eta,u} \) a pseudo graph.

Example 20. We list some well known pseudo graphs here:
- **KAM torus** \( \mathcal{T}_w = \{(x, c + du(x))|x \in \mathbb{T}^n\} \).
- **Aubry set** \( \mathcal{A}(c) \).

Definition 56. \( f : \mathbb{T}^n \to \mathbb{R} \) is called \( K \)-semi concave if for all \( x \in \mathbb{T}^n \), there exists \( p_x \in T^*\mathbb{T}^n \) such that
\[
f(x) - f(y) \leq p_x(x - y) + K|x - y|^2, \quad \forall y \in \mathbb{T}^n.
\]
Here \( p_x \) is called a super differential of \( f \); Generally we call \( f \) be semi concave if there exists \( K > 0 \) such that \( f \) is \( K \)-semi concave.

Remark. semi concave function is Lipschitz, which implies it’s differentiable almost every point.

Example 21. **(Pendulum)** \( H_0 = \frac{v^2}{2} + (\cos 2\pi q - 1) \), so there exists a natural closed 1-form \( \eta^E_q = \sqrt{2E + 2(1 - \cos 2\pi q)} \) such that
\[
G_{\eta^E,0} = \{(a, \eta_a)|a \in \mathbb{T}\}, \quad E \geq 0
\]
becomes a pseudo graph with \( G_{\eta^E,0} = \{H_0 = E\} \cap \{p \geq 0\} \).
• (Perturbed Pendulum) \( H_\varepsilon = H_0 + \varepsilon H_1(q, p, t) \) with \( H_1(0, 0, t) = 0 \), then we choose a pseudo graph

\[
G = \begin{cases} 
W_\varepsilon^u, & \text{for } q \in [0, 1/2), \\
W_\varepsilon^s, & \text{for } q \in (1/2, 1]. 
\end{cases}
\]

Then there exists \( \eta^0 \) and Lipschitz \( u \) such that \( G = G_{\eta^0, u} \).

Let \( S \) be a space of semi concave functions and \( P := H^1(\mathbb{T}^n, \mathbb{R}) \times (S/\mathbb{R}) \) be the space of pseudo graphs. Recall that adding a constant to a semi concave function would’t change the semi concavity, so we can endow a norm on \( S/\mathbb{R} \) by

\[
|u| = \frac{\max u - \min u}{2}
\]

and

\[
\|G_{c,u}\| = |c| + |u|
\]

**Definition 57.** For any \( u \in C^0(\mathbb{T}^n, \mathbb{R}) \), the following Lax-Oleinik mapping

\[
T_\eta u(x) := \min_{\gamma(0)=y,\gamma(1)=x} \{ u(y) + \int_0^1 L_\eta(d\gamma(t), t) + \alpha_L(\eta) dt \}
\]

sends \( u(x) \) into a semi concave function.

**Theorem 58.** • \( \forall \eta \in H^1(\mathbb{T}^n, \mathbb{R}), u, v \) semi concave functions,

\[
\|T_\eta u - T_\eta v\|_\infty \leq \|u - v\|_\infty.
\]

• There exists a unique operator \( \Phi : P \to P \) such that \( \Phi G_{\eta,u} = G_{\eta,T_\eta u} \). Moreover, \( \Phi \) is continuous and for any pseudo graph \( G \in P \), \( \Phi G \subset \varphi^1_H(G) \), where \( \varphi^1_H \) is the time-1-mapping of the Hamiltonian flow.

• For all \( c \in H^1(\mathbb{T}^n, \mathbb{R}), \Phi P_c \subset P_c \); Actually, \( \Phi \) has a fixed point in \( P_c \).

**Example 22** (Katzrelson-Ornstein Trimming). For previous perturbed pendulum \( H_\varepsilon \), we starts from the pseudo graph \( G \) which equals \((W^u(0) \cap \{ q \leq \frac{1}{2} \}) \cup (W^s(0) \cap \{ q > \frac{1}{2} \})\), then we can define \( \Phi \) via the following trimming principle: \( \Phi G = (\Phi^1_H W^u(0) \cap \{ q \leq q^* \}) \cup (\Phi^1_H W^s(0) \cap \{ q > q^* \}) \) where \( q^* \) is uniquely fixed by making the area between \( \Phi^1_H W^u(0), \Phi^1_H W^s(0) \) are equally separated, see Fig. 4.
Figure 4: The red parts correspond to $G$, blue parts is the expansion of $\varphi_H^{1}W^{\mu}(0)$ (resp. retreat of $\varphi_H^{1}W^{\nu}(0)$).

Example 23. For integrable Tonelli Lagrangian $L(x,v,t) = \frac{1}{2}v^2$, any $u \in C^0(\mathbb{T}^n, \mathbb{R})$ and $c \in H^1(\mathbb{T}^n, \mathbb{R})$, $T_n^c u \to \text{const}$ as $n \to \infty$.

Proof. Due to the definition, $T_n^c u(x) = \min_{y \in \mathbb{T}^n} u(y) + \int_0^n L_c(d\gamma(t)) + \alpha(c) dt = \min_{y \in \mathbb{T}^n} u(y) + \int_0^n |\dot{\gamma}(t)-c|^2 dt$, so we can choose $\gamma : [0,n] \in \mathbb{R}^n$ in the universal covering space with $\gamma(0) = \arg \min_{y \in \mathbb{T}^n} u(y), |\dot{\gamma}(t) - c| \leq \frac{\sqrt{2}}{n}$ for all $t \in [0,n]$ and $\gamma(n) \equiv x (\text{mod } \mathbb{Z}^n)$. Then we can see

$$\min_{y \in \mathbb{T}^n} u(y) \leq T_n^c u(x) \leq \min_{y \in \mathbb{T}^n} u(y) + \frac{1}{n}.$$  

Taking the limit we prove the conclusion.  

Definition 59. Denote by $\mathbb{V}_c$ the set of fixed points in $\mathcal{P}_c$, and $\mathbb{V} := \cup_c \mathbb{V}_c$. Then for any $G \in \mathbb{V}_c$, we can find the invariant sets

$$\mathcal{I}(G) := \cap_{n \geq 0} \varphi_{H}^{-n}(G),$$

then

$$\tilde{\mathcal{A}}(c) = \cap_{G \in \mathbb{V}, \mathcal{I}(G)}$$

49
\[ \tilde{N}(c) = \bigcup_{G \in \mathcal{V}_c} \mathcal{I}(G). \]

For two pseudo graphs \( G \) and \( G' \), we define \( G \vdash_N G' \) as follows:

\[ G \vdash_N G' \iff \bigcap_{n=1}^N \varphi^n G' \subseteq \bigcup_{n=1}^N \varphi^n G. \]

We call \( G \) forces \( G' \) and denote by \( G \vdash G' \) if there exists \( N > 0 \), such that \( G \vdash_N G' \). We call \( G \) forces \( c' \) if \( G \) forces some \( G' \in \mathcal{P}_{c'} \); Accordingly, we call \( c \) forces \( c' \) and denote by \( c \vdash c' \) if all \( G \in \mathcal{P}_c \) forces \( c' \).

**Definition 60.** We denote by \( c \dashv \vdash c' \) if \( c \vdash c' \) and \( c' \vdash c \).

**Proposition 61 (Bernard).**

\- \( c \dashv \vdash c' \) is an equivalent relation in \( H^1(\mathbb{T}^n, \mathbb{R}) \).

\- If \( c \dashv c' \), then any two pseudo graphs \( G \in \mathcal{P}_c \) and \( G' \in \mathcal{P}_{c'} \), there exists \( n \in \mathbb{Z} \) such that \( \varphi^n G \cap G' \neq \emptyset \).

\- If \( c \dashv c' \), there exist two heteroclinic orbits connecting \( \tilde{A}(c) \) to \( \tilde{A}(c') \) (resp. \( \tilde{A}(c') \) to \( \tilde{A}(c) \)).

\- Let \( \{ c_i \} \) be a list of cohomology classes, where \( c_i \vdash c_{i+1} \); For any neighborhood \( U_i \) containing \( \tilde{A}(c_i) \), there exists an orbit visiting \( U_i \) in turns.

**References**


