Proceedings of the International Congress of Mathematicians
Hyderabad, India, 2010

Arnold's diffusion: from the *a priori* unstable to the *a priori* stable case.

Patrick Bernard *

Abstract. We expose some selected topics concerning the instability of the action variables in *a priori* unstable Hamiltonian systems, and outline a new strategy that may allow to apply these methods to *a priori* stable systems.

Mathematics Subject Classification (2000). 37J40, 37J50, 37C29, 37C50, 37J50.

Keywords. Arnold's diffusion, normally hyperbolic cylinder, partially hyperbolic tori, homoclinic intersections, Weak KAM solutions, variational methods, action minimization.

1. Introduction

A very classical problem in dynamics consists in studying the Hamiltonian system on the symplectic manifolds $T^*\mathbb{T}^n = \mathbb{T}^n \times \mathbb{R}^n$ generated by the Hamiltonian

$$\begin{aligned} H_\epsilon : \mathbb{T} \times T^*\mathbb{T}^n = \mathbb{T} \times \mathbb{T}^n \times \mathbb{R}^n &\longrightarrow \mathbb{R} \\ (t,x,y) &\longmapsto \frac{1}{2}\|y\|^2 + \epsilon G(t,x,y) \end{aligned} \tag{1}$$

where ϵ is a small perturbation parameter. More general unperturbed systems $h(y)$ can be considered instead of $\|y\|^2/2$, but we restrict to that particular case in the present paper in order to simplify some notations. For $\epsilon = 0$, the system is integrable, and the momenta y are integrals of motion. For $\epsilon > 0$, these variables undergo small oscillations. KAM theory implies that these oscillations remain permanently bounded for many initial conditions. For other initial conditions, a large evolution might be possible. By Nekhoroshev theory, it must be extremely slow. The questions we discuss in the present text is whether this large evolution is actually possible, and to what geometric structures it is associated.

Let us consider a resonant momentum $y_0 = (I_0, 0) \in \mathbb{R}^m \times \mathbb{R}^r = \mathbb{R}^n$, and assume that I_0 is not resonant, which means that $k \cdot I_0$ never belongs to $\mathbb{Z}$ for $k \in \mathbb{Z}^m, k \neq 0$. In order to study the dynamics near the torus $\{y = y_0\}$, it is useful to introduce the notations $x = (\theta, q) \in \mathbb{T}^m \times \mathbb{T}^r$, and $y = (I, p) \in \mathbb{R}^m \times \mathbb{R}^r$, $m + r = n$. In

*Membre de l'IUF

the neighborhood of the torus $\{y = y_0\}$, the dynamics is approximated by the averaged system

$$\frac{1}{2}\|y\|^2 + \epsilon V(q),$$

where

$$V(q) = \int_{\mathbb{T}\times\mathbb{T}^m} G(t,\theta,q,y_0)d\theta dt.$$

Following a classical idea of Poincaré and Arnold, we can try to exploit this observation by considering the system

$$H(t,\theta,q,I,p) = \frac{1}{2}\|p\|^2 + \frac{1}{2}\|I\|^2 - \epsilon V(q) - \mu R(t,\theta,q,I,p) \tag{2}$$

with a second perturbation parameter μ independent from ϵ. We assume that V has a unique non-degenerate minimum, say at $q = 0$. Fixing $\epsilon > 0$, we can study this system for $\mu > 0$ small enough, which is a simpler problem which may give some hints about the dynamics of (1). The reason why instability is more easily proved in (2) than in (1) is the presence of the hyperbolic fixed point at $(0,0)$ of the (q,p) component of the averaged system. Studying (2) for $\mu > 0$ small enough is thus called the *a priori* unstable problem, or the *a priori* hyperbolic problem. In contrast, the Hamiltonian (1) is called *a priori* stable. The *a priori* unstable case is by now quite well understood for $m = 1$, see [46, 17, 18, 26, 5, 48] for example. The *a priori* unstable case for $m > 1$ and the *a priori* stable case can be considered as widely open, in spite of the important announcements of John Mather in [42]. The starting point in the study of (2) is the famous paper of Arnold, [1]. In this paper, Arnold introduced a particular *a priori* unstable system where some geometric structures associated to diffusion, partially hyperbolic tori (that he called whiskered), their stable and unstable manifolds, and heteroclinic connections, can be almost explicitly described. This geometric structure have been called a transition chain. Most of the subsequent works on the *a priori* unstable problem have consisted in trying to find transition chains in more general cases, but understanding the general *a priori* unstable Hamiltonian have required a change of paradigm: from partially hyperbolic tori to normally hyperbolic cylinders. The variational methods introduced by John Mather in [41] and Ugo Bessi in [9] have also been very influential.

Transforming the understanding gained on the dynamics of (2) to informations on the *a priori* stable case is not an easy task. Since we understand the system (2) when $m = 1$ the first attempt should be to study (1) in the neighborhood of an $(n-1)$-resonant line, for example the line consisting of momenta of the form $y = (I, 0), I \in \mathbb{R}$. We could hope to prove the existence of drift along such a line by using the *a priori* unstable approximations near each value of y. However, we face the problem that an approximation like (2) holds only in the neighborhood of the torus $\{y = (I, 0)\}$ when the frequency $I \in \mathbb{R}$ is irrational. Near the torus $\{y = (I, 0)\}$ with I rational, one should use an approximation of the form

$$H(t,x,y) = \frac{1}{2}\|y\|^2 - \epsilon W(x) - \mu R(t,x,y)$$

and different methods must be used. This is often called the problem of double resonances when $n = 2$. We will call it the problem of *maximal resonances*.

Our general goal in this paper is to study *a priori* unstable systems with a sufficient generality to be able to gain informations on the *a priori* stable case. We start with a relatively detailed description of the Arnold's example in Section 2, which is also an occasion to settle some notations and introduce some important objects, like the partially hyperbolic tori, their stable and unstable manifolds, and the associated generating functions. Working with these generating functions allows to highlight the connections between the various classical approaches, geometric methods, variational methods, and weak KAM theory. Then, from the end of Section 2 to Section 3, we progressively generalize the setting and indicate how the methods introduced on the example of Arnold can be improved to face the new occurring difficulties. We present the Large Gap Problem, which prevents Arnold's mechanism from being directly applied to general *a priori* unstable systems, and explain how the presence of a normally hyperbolic cylinder can be used to solve this Problem and prove instability in general *a priori* unstable systems. In section 4 we give a new result from [6], on the existence of normally hyperbolic cylinders in the *a priori* stable situation, which should allow to apply the tools exposed in the previous sections to *a priori* stable systems. This suggests a possible strategy to prove the following conjecture:

Conjecture 1. *For a typical perturbation G, there exists two positive numbers ϵ_0 and δ, such that, for each $\epsilon \in]0, \epsilon_0[$, The system (1) has an orbit*

$$(\theta(t), q(t), \dot{\theta}(t), \dot{q}(t)) : \mathbb{R} \longrightarrow \mathbb{T} \times \mathbb{T}^{n-1} \times \mathbb{R} \times \mathbb{R}^{n-1}$$

such that $\sup_t \dot{\theta} - \inf_t \dot{\theta} > \delta$.

We are currently working on this program in collaboration with Vadim Kaloshin and Ke Zhang. The same conjecture can be stated with a more general unperturbed system $h(y)$, and the same proof should work provided h is convex and smooth. Our strategy of proof does not consist in solving the maximal resonance problem, but rather in observing that the conjectured statement can be reached without solving that difficulty. In that respect, what we expose is much easier than the project of Mather as announced in [42]. The result is weaker since only limited diffusion is obtained. The maximal resonance problem has to be solved in order to prove the existence of global diffusion on a whole resonant line, or from one resonant line to another. Our strategy, on the other hand, has the advantage of working with all $n \geqslant 2$, while Mather is limited to $n = 2$ at the moment.

2. The example of Arnold and some extensions

Following Arnold [1], we consider the Hamiltonian

$$H(t, \theta, q, I, p) = \frac{1}{2}\|p\|^2 + \frac{1}{2}\|I\|^2 + \epsilon(\cos(2\pi q) - 1)(1 + \mu f(t, \theta, q)) \qquad (3)$$

with $(t, \theta, q, I, p) \in \mathbb{T} \times \mathbb{T} \times \mathbb{T} \times \mathbb{R} \times \mathbb{R}$. We will often use the corresponding Lagrangian

$$L(t, \theta, q, \dot{\theta}, \dot{q}) = \frac{1}{2}\|\dot{q}\|^2 + \frac{1}{2}\|\dot{\theta}\|^2 + \epsilon(1 - \cos(2\pi q))(1 + \mu f(t, \theta, q)).$$

We will see $\epsilon > 0$ as a fixed parameter, and discuss mainly the small parameter μ. When $\mu = 0$, the variable I is an integral of motion. Our goal is to study its evolutions for $\mu > 0$. The form of the perturbation is chosen in such a way that the two-dimensional tori

$$\mathcal{T}(a) = \mathbb{T}^2 \times \{0\} \times \{a\} \times \{0\}, \quad a \in \mathbb{R}$$

are invariant in the extended phase space, and carry a linear motion of frequency $(1, a)$. By studying invariant manifolds attached to these invariant tori, Arnold discovered a remarkable diffusion mechanism, now called the Arnold Mechanism, that we are now going to describe. In the case $\mu = 0$, the tori $\mathcal{T}(a)$ appear as the products of the hyperbolic fixed point $\{0, 0\}$ of the pendulum in (q, p) by the invariant torus $\mathbb{T} \times \mathbb{T} \times \{a\}$ of the integrable system in the (t, θ, I) space. They are thus partially hyperbolic, and have stable and unstable manifolds, which coincide and can be given explicitly as

$$\mathcal{W}(a) = \big\{\big(t, \theta, q, a, \pm\partial_q S_0(q)\big) : (t, \theta, q) \in \mathbb{T}^3\big\}$$

with

$$S_0(q) = \frac{2\sqrt{\epsilon}}{\pi}(1 - \cos(\pi q)). \tag{4}$$

The coincidence and compactness of these stable manifolds is a very special feature of the unperturbed case $\mu = 0$. For $\mu \neq 0$, the tori $\mathcal{T}(a)$ still have stable and unstable manifolds which can be described as follows: There exists two functions

$$S^{\pm}_{a,\mu}(t, \theta, q) : \mathbb{T} \times \mathbb{T} \times [-3/4, 3/4] \longrightarrow \mathbb{R}, \tag{5}$$

which converge to $\pm S_0$ when $\mu \longrightarrow 0$, and such that the graphs

$$\mathcal{W}^{\pm}_{\mu}(a) = \big\{\big(t, \theta, q \mod 1, a + \partial_\theta S^{\pm}(t, \theta, q), \partial_q S^{\pm}(t, \theta, q)\big)\big\}$$

are pieces of the stable and unstable manifolds of the torus $\mathcal{T}(a)$. More precisely, the set $\mathcal{W}^+(a)$ is negatively invariant under the extended Hamiltonian flow, and

$$\mathcal{T}(a) = \bigcap_{t \leqslant 0} \varphi^t\big(\mathcal{W}^+(a)\big)$$

while the set $\mathcal{W}^-(a)$ is positively invariant under the extended Hamiltonian flow, and

$$\mathcal{T}(a) = \bigcap_{t \geqslant 0} \varphi^t\big(\mathcal{W}^-(a)\big).$$

The functions $S^{\pm}_a$ solve the Hamilton-Jacobi equation

$$\partial_t S + H(t, \theta, q, a + \partial_\theta S, \partial_q S) = a^2/2,$$

which merely says that the invariant manifolds are contained in the energy level of the torus. The functions $S_a^\pm$ have an expression in terms of the action:

$$S_a^+(t,\theta,q) = \int_{-\infty}^{\tau} L(s,\theta^+(s),q^+(s),\dot\theta^+(s),\dot q^+(s)) - a\dot\theta^+(s) + a^2/2ds \tag{6}$$
$$S_a^-(t,\theta,q) = \int_{\tau}^{+\infty} L(s,\theta^-(s),q^-(s),\dot\theta^-(s),\dot q^-(s)) - a\dot\theta^-(s) + a^2/2ds,$$

where τ is any real number such that $\tau \mod 1 = t$, and $(\theta^\pm(s),q^\pm(s))$ is the solution of the Euler-Lagrange equations such that

$$\theta^\pm(\tau) = \theta, q^\pm(\tau) = q \mod 1, \dot\theta^\pm(\tau) = a + \partial_\theta S^\pm(t,\theta,q), \dot q^\pm(\tau) = \partial_q S^\pm(t,\theta,q).$$

Note that the result does not depend on the choice of τ.

2.1. Homoclinic orbits. If $(T,\Theta,Q) \in \mathbb{T}\times\mathbb{T}\times]1/4,3/4[$ is a critical point of the function

$$\Delta_a(t,\theta,q) = S_a^+(t,\theta,q) - S_a^-(t,\theta,q-1),$$

then the point

$$\begin{aligned}&\big(T,\Theta,Q \mod 1, a+\partial_\theta S_a^+(T,\Theta,Q), \partial_q S_a^+(T,\Theta,Q)\big)\\ =&(T,\Theta,(Q-1) \mod 1, a+\partial_\theta S_a^-(T,\Theta,Q-1), \partial_q S_a^-(T,\Theta,Q-1)\big)\end{aligned}$$

obviously belongs both to $\mathcal{W}^+(a)$ and $\mathcal{W}^-(a)$, hence it is a homoclinic point. It is a transversal homoclinic point if in addition the Hessian of Δ_a has rank two (it can not have rank 3 because the intersection is necessarily one-dimensional). It is not obvious at this point that the function Δ_a necessarily has critical points on the domain $\mathbb{T}\times\mathbb{T}\times]1/4,3/4[$. When μ is small enough, this follows from:

Lemma 2. *If (T,Q) is a critical point of the function $\bar\Delta_a : (t,q) \longmapsto \Delta_a(t,q,1/2)$, then $(T,Q,1/2)$ is a critical point of Δ_a, hence the manifolds $\mathcal{W}^-(a)$ and $\mathcal{W}^+(a)$ intersect above $(T,\Theta,1/2) \in \mathbb{T}^3$. This homoclinic point is transversal if and only if the Hessian of $\bar\Delta_a$ at (T,Q) is a non-degenerate 2×2 matrix.*

Note that the function $\bar\Delta_a$ is defined on $\mathbb{T}^2$, and therefore it has critical points. Proof. We have $\partial_t S^+(T,Q,1/2) = \partial_t S^-(T,Q,-1/2)$, let us denote by e this value. We also have $\partial_\theta S^+(T,Q,1/2) = \partial_\theta S^-(T,Q,-1/2)$, we denote by I this value. It is enough to prove that $\partial_q S^+(T,Q,1/2) = \partial_q S^-(T,Q,-1/2)$. In order to do so, it is enough to observe that $\partial_q S^+$ is the only non-negative solution of the equation

$$e + H(T,\Theta,1/2,a+I,.) = a^2/2,$$

and that precisely the same characterization is true for $\partial_q S^-(T,Q,-1/2)$. Note that the equation above has two solutions, and that we can discriminate between them because we work in a perturbative setting which gives us rough informations on the signs. In more general situation, this is a source of difficulty. □

2.2. Heteroclinic orbits. We have proved the existence of homoclinic orbits. But what is interesting for Arnold diffusion are heteroclinic orbits between different tori. We can deduce the existence of a heteroclinic orbit between $\mathcal{T}(a)$ and $\mathcal{T}(a')$ provided we can find a critical point of the function

$$\mathbb{T} \times \mathbb{R} \times]1/4, 3/4[\ni (t, \theta, q) \longmapsto S_a^+(t, \theta, q) - S_{a'}^-(t, \theta, q-1) + (a - a')\theta,$$

where we have lifted the functions S without changing their names. As before, we can limit ourselves to finding critical points of the function

$$\Sigma_{a,a'} : \mathbb{T} \times \mathbb{R} \ni (t, \theta) \longmapsto S_a^+(t, \theta, 1/2) - S_{a'}^-(t, \theta, -1/2) + (a - a')\theta, \tag{7}$$

but the term $(a - a')\theta$ prevents us from finding them using a global variational method when $a' \neq a$. This reflects the fact that we are studying a non exact Lagrangian intersection problem. For $\mu = 0$, heteroclinics do not exist. However, recalling that $\bar{\Delta}_a(t, \theta) = \Delta_a(t, \theta, 1/2)$, we have:

Lemma 3. *If the function $\bar{\Delta}_a(t, q)$ has a non-degenerate critical point, then the functions $\Sigma_{a,a'}$ and $\Sigma_{a',a}$ both have a non-degenerate critical point provided a' is sufficiently close to a.*

Proof. The theory of partial hyperbolicity implies that the stable and unstable manifolds $\mathcal{W}_\mu^\pm(a)$ depend regularly on the parameter a. As a consequence, their generating functions $S_a^\pm$ also regularly depend on a, and the functions $\Sigma_{a,a'}$ depend regularly on a and a'. The result follows since $\Sigma_{a,a} = \bar{\Delta}_a$. □

We say that $a_0, a_1, \ldots, a_k$ is an elementary transition chain if the functions Σ_{a_{i-1},a_i} have non-degenerate critical points. We will sometimes use the same terminology for the different requirement that these functions have isolated local minima. From Lemma 3, we deduce:

Proposition 1. *Let μ be given and sufficiently small. Let $[a^-, a^+]$ be an interval such that each of the functions $\bar{\Delta}_{a,\mu}, a \in [a^-, a^+]$ have a non-degenerate critical point, which means that each of the tori $\mathcal{T}_\mu(a), a \in [a^-, a^+]$ has a transversal homoclinic orbit. Then there exists an elementary transition chain $a^- = a_0, a_1, \ldots, a_k = a^+$.*

Proof. Let us consider the set $A \subset [a^-, a^+]$ of points that can be reached from a^- by a transition chain. The set A is open : If $a' \in A$, then there exists a transition chain $a^- = a_0, a_1, \ldots a_k = a'$ and, by Lemma 3, the sequence $a^- = a_0, a_1, \ldots a_k, a_{k+1}$ is a transition chain when a_{k+1} is sufficiently close to a. The set A is closed : Let a be in the closure of A. By Lemma 3, the pair a, a' is a transition chain when a' is close to a. Since a is in the closure of A, the point a' can be taken in A. Then, there exists a transition chain $a^- = a_0, \ldots, a_k = a'$, and then the longer sequence $a^- = a_0, \ldots, a_k, a_{k+1} = a$ is a transition chain between a_0 and a, hence $a \in A$. Being open, closed and not empty (it contains a_0), the set A is equal to $[a^-, a^+]$. □

The existence of transition chains implies the existence of diffusion orbits. This is proved by Arnold invoking an “obstruction property”. This obstruction property is a characteristic of the local dynamics near the partially hyperbolic tori. It has been proved by Jean-Pierre Marco in [38], see also [21, 31]. The most appealing way to understand the geometric shadowing of transition chains is to use the following statement of Jacky Cresson [22], which can be seen as a strong obstruction property:

Lemma 4. *If there exists a transversal heteroclinic between $\mathcal{T}(a)$ and $\mathcal{T}(a')$ and a transversal heteroclinic between $\mathcal{T}(a')$ and $\mathcal{T}(a'')$, then there exists a transversal heteroclinic between $\mathcal{T}(a)$ and $\mathcal{T}(a'')$.*

This Lemma implies:

Corollary 5. *If $a_0, a_1, \dots, a_k$ is an elementary transition chain, then there exists a transversal heteroclinic orbit between $\mathcal{T}(a_0)$ and $\mathcal{T}(a_k)$.*

Putting everything together, we obtain:

Theorem 1. *Let μ be given and sufficiently small. Let $[a^-, a^+]$ be an interval such that each of the functions $\bar{\Delta}_{a,\mu}, a \in [a^-, a^+]$ have a non-degenerate critical point. Then there exists a heteroclinic orbit between $\mathcal{T}(a^-)$ and $\mathcal{T}(a^+)$.*

2.3. Poincaré-Melnikov approximation. We have constructed diffusion orbits under the assumption that transversal homoclinics exist. We have proved that homoclinic orbits necessarily exist, and one may argue that transversality should hold for typical systems, we will come back on this later. However, it is useful to be able to check whether transversality holds in a given system. A classical approach consists in proving the existence of non-degenerate critical points of the functions $\bar{\Delta}_{a,\mu}$ defined in Lemma 3 by expanding them in power series of μ. As a starting point the generating functions $S_a^{\pm}$ can be expanded as follows:

$$\begin{aligned} S_a^+(t,\theta,q) &= S_0(q) + \mu M_a^+(t,\theta,q) + O(\mu^2), \\ S_a^-(t,\theta,q) &= -S_0(q) - \mu M_a^-(t,\theta,q) + O(\mu^2), \end{aligned} \tag{8}$$

where $S_0(q) = \frac{2\sqrt{\epsilon}}{\pi}(1-\cos(\pi q))$ is the generating function of the unperturbed manifolds, and $M^{\pm}$ are the so-called Poincaré-Melnikov integrals,

$$M_a^+(t,\theta,q) = \epsilon \int_{-\infty}^{t} F\Big(s, \theta + a(s-t), \frac{2}{\pi}\arctan\big(e^{2\pi\sqrt{\epsilon}(s-t)}\tan(\pi q/2)\big)\Big) ds$$
$$M_a^-(t,\theta,q) = \epsilon \int_{t}^{+\infty} F\Big(s, \theta + a(s-t), \frac{2}{\pi}\arctan\big(e^{2\pi\sqrt{\epsilon}(t-s)}\tan(\pi q/2)\big)\Big) ds$$

where $F(t,\theta,q) = (1-\cos(2\pi q))f(t,\theta,q)$. To better understand these formula, it is worth recalling that

$$s \longmapsto \frac{2}{\pi}\arctan\big(e^{2\pi\sqrt{\epsilon}(s-t)}\tan(\pi q/2)\big)$$

is the homoclinic orbit of the system $\|p\|^2/2+\epsilon(\cos(2\pi q)-1)$ which takes the value q at time t. The formula above are similar to (6), but the integration is performed on unperturbed trajectories, which are explicitly known. For $q\in]1/4,3/4[$, we obtain:

$$\Delta_a(t,\theta,q)=S_a^+(t,\theta,q)-S_a^-(t,\theta,q-1)=\mu M_a(t,\theta,q)+O(\mu^2),$$

where M_a is the Poincaré-Melnikov integral

$$\begin{aligned} M_a(t,\theta,q) &= M_a^+(t,\theta,q)+M_a^-(t,\theta,q-1)\\ &=\epsilon\int_{\mathbb{R}} F\Big(s,\theta+a(s-t),\frac{2}{\pi}\arctan\left(e^{2\pi\sqrt{\epsilon}(s-t)}\tan(\pi q/2)\right)\Big)ds. \end{aligned}$$

In the specific case studied by Arnold, where $f(t,\theta,q)=\cos(2\pi\theta)+\cos(2\pi t)$, the Melnikov integral can be computed explicitly through residues, we obtain:

$$M_a(t,q,1/2)=\frac{a}{\mathrm{sh}(\pi a/2\sqrt{\epsilon})}\cos(2\pi\theta)+\frac{1}{\mathrm{sh}(\pi/2\sqrt{\epsilon})}\cos(2\pi t),$$

it has a non-degenerate minimum at $(t,q)=(0,0)$. We can conclude, following Arnold:

Theorem 2 (Arnold, [1]). *Let us consider the Hamiltonian (3) with $f(t,\theta,q)=\cos(2\pi\theta)+\cos(2\pi t)$ and $\mu>0$ small enough. Given two real numbers $a^-<a^+$, there exists an orbit $(\theta(t),q(t),I(t),p(t))$ and a time $T>0$ such that $I(0)\leqslant a^-$ and $I(T)\geqslant a^+$.*

2.4. Bessi's variational mechanism.

Ugo Bessi introduced in [9] a very interesting approach to study the system (3), see also [10, 11]. In order to describe this approach, let us define the function

$$\begin{aligned} &A_a:\mathbb{R}\times\mathbb{T}\times]1/4,3/4[\times\mathbb{R}\times\mathbb{T}\times]1/4,3/4[\longrightarrow\mathbb{R}\\ &\quad((t_1,\theta_1,q_1),(t_2,\theta_2,q_2))\longmapsto\min\int_{t_1}^{t_2}L(s,\theta(s),q(s),\dot\theta(s),\dot q(s))-a\dot\theta(s)+a^2/2\,ds, \end{aligned}$$

where the minimum is taken on the set of C^1 curves $(\theta(s),q(s)):[t_1,t_2]\longrightarrow\mathbb{T}\times\mathbb{R}$ such that

$$(\theta(t_1),q(t_1))=(\theta_1,q_1-1)\quad\text{and}\quad(\theta(t_2),q(t_2))=(\theta_2,q_2).$$

When the time interval t_2-t_1 is very large, the minimizing trajectory in the definition of A_a roughly looks like the concatenation of an orbit positively asymptotic to $\mathcal{T}(a)$ followed by an orbit negatively asymptotic to $\mathcal{T}(a)$. Using this observation, and recalling the formula (6), it is possible to prove that

$$\begin{aligned} &A_a((t_1,\theta_1,q_1),(t_2+k,\theta_2,q_2))\longrightarrow\\ &S_a^+(t_2\mod 1,\theta_2,q_2)-S_a^-(t_1\mod 1,\theta_1,q_1-1) \end{aligned}$$

when $k \longrightarrow \infty$. Fixing the real numbers $a_0, a_1, \dots, a_k$ and the integers $\tau_1, \dots, \tau_k$, we consider the discrete action functional

$$\begin{aligned}
&S^+_{a_0}(t_1 \mod 1, \theta_1 \mod 1, 1/2) + (a_0 - a_1)\theta_1 \\
&+A_{a_1}((t_1, \theta_1 \mod 1, 1/2), (t_2 + \tau_2, \theta_2 \mod 1, 1/2)) + (a_1 - a_2)\theta_2 \\
&+A_{a_2}((t_2, \theta_2 \mod 1, 1/2), (t_3 + \tau_3, \theta_3 \mod 1, 1/2)) + (a_2 - a_3)\theta_3 \\
&+\cdots \\
&+A_{a_{k-1}}(t_{k-1} + \tau_{k-1}, \theta_{k-1} \mod 1, 1/2), (t_k, \theta_k \mod 1, 1/2)) + (a_{k-1} - a_k)\theta_k \\
&-S^-_{a_k}(t_k \mod 1, \theta_k \mod 1, 1/2)
\end{aligned}$$

defined on $(]-1,1[\times]-1,1[)^k$. It is not hard to check that local minima of this discrete action functional give heteroclinics between the Torus $\mathcal{T}(a_0)$ and the torus $\mathcal{T}(a_k)$. In order to prove that local minima exist, observe that this functional is approximated by

$$\Sigma_{a_0,a_1}(t_1 \mod 1, \theta_1) + \cdots + \Sigma_{a_{k-1},a_k}(t_k \mod 1, \theta_k)$$

when the integers τ_i are large enough, with the functions Σ as defined in (7). This limit functional has the remarkable structure that the variables (t_i, θ_i) are separated. This break-down of the action functional into a sum of independent functions is sometimes called an anti-integrable limit, it is related to the obstruction property of the invariant tori, to the λ-Lemma, and to the Shilnokov's Lemma, see [15]. The limit functional has an isolated local minimum provided each of the functions Σ_{a_{i-1},a_i} has one, which is equivalent to say that $a_0, a_1, \dots, a_k$ is an elementary transition chain. In this case, the integers τ_i can be chosen large enough so that the action functional above has a local minimum, which gives a heteroclinic orbit between $\mathcal{T}(a_0)$ and $\mathcal{T}(a_k)$. Technically, this method has several advantages. In our presentation we introduced the generating functions $S^{\pm}_{a_i}$ of the invariant manifolds of the involved tori in order to stress the relations between the geometric and the variational method, and also because this works in a more general setting, see [15] for example. In our context and when $\mu > 0$ is small enough, it is easier to directly approximate the functions A_a in terms of the Melnikov integrals, and to use the following approximation for the action functional with large τ_i and small μ without the intermediate step through $S^{\pm}$:

$$\begin{aligned}
&\mu M_{a_0}(t_0 \mod 1, \theta_0 \mod 1, 1/2) + (a_1 - a_0)\theta_0 + \cdots + \\
&\qquad \mu M_{a_k}(t_k \mod 1, \theta_k \mod 1, 1/2) + (a_k - a_{k-1})\theta_k.
\end{aligned}$$

The corresponding calculations, performed in [9], are much more elementary than those required to derive the expansions (8).

2.5. Remarks on estimates. We have up to that point carefully avoided to discuss the subtle and important aspect of explicit estimates. In order to complete rigorously the proof of Theorem 2, we should prove the existence of a threshold $\mu_0(\epsilon)$ such that the Melnikov approximation holds, simultaneously for all a, when

$0 < \mu < \mu_0(\epsilon)$. This can actually been done, with

$$\mu_0(\epsilon) = e^{-\frac{C}{\sqrt{\epsilon}}},$$

but it is not simple, since it requires to study carefully the expansions of the functions $S_a^{\pm}$ and how the coefficients depend on a and ϵ. This is related to the so-called splitting problem, see [37]. As we mentioned above the approach of Bessi allows to prove that Theorem 2 holds for $0 < \mu < \mu_0(\epsilon)$ without estimating the splitting.

It is also important to give time estimates, that is to estimate the time needed for the variable I to perform a large evolution. Once again, this is closely related to the splitting estimates, although these can be avoided by using the method of Bessi. One should distinguish two different problems. Either we fix ϵ, and try to estimate the time as a function of μ, or we take μ as a function of ϵ, say $\mu = \mu_0(\epsilon)/2$, and try to estimate the time as a function of ϵ.

The second problem is especially important, because it is relevant for the study of the *a priori* stable problem. Once again, Ugo Bessi obtained the first estimate,

$$T = e^{\frac{C}{\sqrt{\epsilon}}}.$$

Estimating the time on examples allows to test the optimality of Nekhoroshev exponents, see [39, 36, 49] for works in that direction, see also [16] concerning the question of time estimates.

It is worth mentioning also that in the first problem, estimating the time as a function of μ, the estimate is polynomial, and not exponentially small. This was first understood by Pierre Lochak, and proved by Bessi's method in [2], where the estimate $T = C/\mu^2$ is given, see also [23]. The optimal estimate is $T = C|\ln \mu|/\mu$, as was conjectured by Lochak in [35] and proved by Berti, Biasco and Bolle in [8], see also [7].

Returning to the question of the threshold of validity, let us discuss what happens when μ is increased above $\mu_0(\epsilon)$. The content of Section 2.3 on finding transversal homoclinics via the Poincaré-Melnikov approximation breaks down, but the geometric constructions of the earlier sections is still valid. Theorem 1 holds as long as the invariant tori $\mathcal{T}(a)$ remain partially hyperbolic, and that their stable and unstable manifold can be represented by generating functions like (5). Actually, the methods we are now going to expose allow even to relax this last assumption. Being able to treat larger values of μ is especially important in view of the possible applicability to the *a priori* stable problem.

2.6. Higher dimensions. Let us now discuss the following immediate generalization in higher dimensions of Arnold's example:

$$H(t, \theta, q, I, p) = \frac{1}{2}\|p\|^2 + \frac{1}{2}\|I\|^2 - \epsilon V(q)(1 + \mu f(t, \theta, q))$$

with $(t, \theta, q, I, p) \in \mathbb{T} \times \mathbb{T}^m \times \mathbb{T}^r \times \mathbb{R}^m \times \mathbb{R}^r$, where $V(q)$ is a non-negative function having a unique non-degenerate minimum at $q = 0$, with $V(0) = 0$. The main difference with the example of Arnold appears for $r > 1$. In this case, the system is not integrable even for $\mu = 0$. There still exists a family of partially hyperbolic tori of dimension m,

$$\mathcal{T}(a) := \{(t, \theta, 0, a, 0), (t, \theta) \in \mathbb{T} \times \mathbb{T}^m\}$$

parametrized by $a \in \mathbb{R}^m$, but the system $\|p\|^2/2 - \epsilon V(q)$ is not necessarily integrable any more. As a consequence we do not know explicitly the stable and unstable manifolds of the hyperbolic fixed point $(0, 0)$, and so we do not have a perturbative setting to describe the stable and unstable manifolds of the hyperbolic tori $\mathcal{T}(a)$. This is also what happens for $r = 1$ if μ is not small enough. There is no obvious generalization of the generating functions $S_a^{\pm}$ in that setting, because the stable and unstable manifolds are not necessarily graphs over a prescribed domain. The proof of the existence of homoclinic orbits as given in 2.1 thus breaks down. The existence of homoclinic orbits in that setting can still be proved by global variational methods, as is now quite well understood, see [14, 28, 20, 3, 27] for example.

The proof is quite easy in our context, let us give a rapid sketch. We first define a function A_a similar to the one appearing in Section 2.4, but slightly different:

$$\begin{aligned}
&A_a : \mathbb{R} \times \mathbb{T}^m \times \mathbb{T}^r \times \mathbb{R} \times \mathbb{T}^m \times \mathbb{T}^r \longrightarrow \mathbb{R} \\
&\quad((t_1, \theta_1, q_1), (t_2, \theta_2, q_2)) \longmapsto \min \int_{t_1}^{t_2} L(s, \theta(s), q(s), \dot\theta(s), \dot q(s)) - a\dot\theta(s) + a^2/2\, ds,
\end{aligned}$$

where the minimum is taken on the set of curves $(\theta(s), q(s)) : [t_1, t_2] \longrightarrow \mathbb{T}^m \times \mathbb{T}^r$ such that $(\theta(t_i), q(t_i)) = (\theta_i, q_i)$ for $i = 1$ or 2. Let us set

$$\xi(a) := \liminf_{\mathbb{N} \ni k \longrightarrow \infty} A_a((0, 0, q_0), (k, 0, q_1)),$$

and consider a sequence of minimizing extremals

$$(\theta_i(t), q_i(t)) : [0, k_i] \longrightarrow \mathbb{T}^m \times \mathbb{T}^r$$

such that $(\theta_i(0), q_i(0)) = (0, q_0)$, $(\theta_i(k_i), q_i(k_i)) = (0, q_1)$, $k_i \longrightarrow \infty$, and

$$\int_0^{k_i} L(s, \theta_i(s), q_i(s), \dot\theta_i(s), \dot q_i(s)) - a\dot\theta_i(s) + a^2/2\, ds \longrightarrow \xi(a).$$

Let M be a submanifold of $\mathbb{T} \times \mathbb{T}^m \times \mathbb{T}^r$ which separates $\mathbb{T} \times \mathbb{T}^m \times \{q_0\}$ from $\mathbb{T} \times \mathbb{T}^m \times \{q_1\}$, and let $T_i \in [0, k_i]$ be a time such that $(T_i \mod 1, \theta_i(T_i), q_i(T_i)) \in M$, and let τ_i be the integer part of T_i. It is not hard to check that the curves $(\theta_i(t - \tau_i), q_i(t - \tau_i))$ converge (up to a subsequence) uniformly on compact sets to a limit $(\theta_\infty(t), q_\infty(t)) : \mathbb{R} \longrightarrow \mathbb{T}^m \times \mathbb{T}^r$. This limit curve satisfies

$$\int_{-\infty}^{\infty} L(s, \theta_\infty(s), q_\infty(s), \dot\theta_\infty(s), \dot q_\infty(s)) - a\dot\theta_\infty(s) + a^2/2\, ds = \xi(a), \tag{9}$$

and the corresponding orbit is a heteroclinic from $\mathcal{T}_0(a)$ to $\mathcal{T}_1(a)$. We call minimizing heteroclinics (for the lifted system) those which have minimal action, or in other words those which satisfy (9). In the original system (before taking the covering), we call minimizing homoclinic orbit a homoclinic which lifts to a minimizing heteroclinic.

Let us now try to establish some connections between the present discussion and the proof of the existence of homoclinic orbits given in 2.1. We define two functions on $\mathbb{T} \times \mathbb{T}^m \times \mathbb{T}^r$:

$$S_a^-(t,\theta,q) = -\liminf_{\mathbb{N}\ni k\longrightarrow\infty} A_a\big((t,\theta,q),(k,0,q_1)\big) \tag{10}$$

$$S_a^+(t,\theta,q) = \liminf_{\mathbb{N}\ni k\longrightarrow\infty} A_a\big((0,0,q_0),(t+k,\theta,q)\big). \tag{11}$$

Note that

$$S_a^+(0,0,q_1) = -S_a^-(0,0,q_0) = \xi(a).$$

The functions $S_a^\pm$, whose definition is basic both in Mather's ([41]) and in Fathi's ([29]) theory, share many features with those introduced in (5), that's why we use the same name. Let us state some of their properties:

The function S_a^- is non-positive and it vanishes on $\mathbb{T} \times \mathbb{T}^m \times \{q_1\}$ (and only there). Moreover, it is smooth around this manifold, which is a transversally non-degenerate critical manifold. Let us chose a small $\delta > 0$. The set

$$\mathcal{W}^-_{loc}(a) := \Big\{\big(t,\theta,q,a+\partial_\theta S_a^-,\partial_q S_a^-(t,\theta,q)\big), \quad S_a^-(t,\theta,q) > -\delta\Big\}$$

is a positively invariant local stable manifold of $\mathcal{T}_1(a)$.

Similarly, S_a^+ is non-negative, it is null on $\mathbb{T} \times \mathbb{T}^m \times \{q_0\}$, and smooth around it, and this critical manifold is transversally non-degenerate. The set

$$\mathcal{W}^+_{loc}(a) := \Big\{\big(t,\theta,q,a+\partial_\theta S_a^+,\partial_q S_a^+(t,\theta,q)\big), \quad S_a^+(t,\theta,q) < \delta\Big\}$$

is a negatively invariant local unstable manifold of $\mathcal{T}_0(a)$.

The functions $S_a^\pm$ also have a global meaning. Let us give the details for S^+. For each point (T,Θ,Q), there exists a real number $\tau \in \mathbb{R}$ and at least one solution $(\theta(s),q(s)) : (-\infty,\tau] \longrightarrow \mathbb{T}^m \times \mathbb{T}^r$ of the Euler-Lagrange equations such that $(\tau \mod 1, \theta(\tau), q(\tau)) = (T,\Theta,Q)$, and which is calibrated by S_a^+ in the following sense: The relation

$$\begin{aligned} &S_a^+(t \mod 1,\theta(t),q(t)) - S_a^+(s \mod 1,\theta(s),q(s)) \\ &= \int_s^t L(\sigma,\theta(\sigma),q(\sigma),\dot\theta(\sigma),\dot q(\sigma)) - a\dot\theta(\sigma) + a^2/2\, d\sigma \end{aligned}$$

holds for all $s < t \leqslant \tau$. The corresponding orbit is asymptotic either to $\mathcal{T}_0(a)$ or to $\mathcal{T}_1(a)$ when $s \longrightarrow -\infty$. It is not easy in general to determine whether the asymptotic torus is $\mathcal{T}_0(a)$ or $\mathcal{T}_1(a)$ but the following Lemma is not hard to prove:

Lemma 6. *If $S_a^+(T,\Theta,Q) < \xi(a)$, then each calibrated curve*

$$(\theta(s),q(s)) : (-\infty,\tau] \longrightarrow \mathbb{T}^m \times \mathbb{T}^r$$

satisfying $(\tau \bmod 1, \theta(\tau), q(\tau)) = (T, \Theta, Q)$*, is* α*-asymptotic to* $\mathcal{T}_0(a)$*, and satisfies*

$$\int_{-\infty}^{\tau} L(\sigma, \theta(\sigma), q(\sigma), \dot\theta(\sigma), \dot q(\sigma)) - a\dot\theta(\sigma) + a^2/2\, d\sigma = S_a^+(T, \Theta, Q).$$

If the function S_a^+ is differentiable at (T, Θ, Q) then there is one and only one calibrated curve as above, it is characterized by the equations

$$\dot\theta(\tau) = a + \partial_\theta S_a^+(\tau, \Theta, Q), \quad \dot q(\tau) = \partial_q S_a^+(\tau, \Theta, Q).$$

Formally, the critical points of the difference $S_a^+ - S_a^-$ correspond to heteroclinic orbits (in the lifted system). By studying a bit more carefully the relations between the calibrated curves and the differentiability properties of the functions $S_a^\pm$ (which is one of the central aspects of Fathi's Weak KAM theory, see [29]), this idea can be made rigorous as follows:

Lemma 7. *If* (T, Θ, Q) *is a local minimum of the function* $S_a^+ - S_a^-$*, then both* S_a^+ *and* S_a^- *are differentiable at the point* (T, Θ, Q)*, we have*

$$\big(T, \Theta, Q, a + \partial_\theta S^-, \partial_q S^-\big) = \big(T, \Theta, Q, a + \partial_\theta S^+, \partial_q S^+\big),$$

and the orbit of this point is either a heteroclinic between $\mathcal{T}_0(a)$ *and* $\mathcal{T}_1(a)$ *or a homoclinic to* $\mathcal{T}_0(a)$ *or to* $\mathcal{T}_1(a)$ *in the system lifted to the covering, and thus it projects to an orbit homoclinic to* $\mathcal{T}(a)$ *in the original system.*

Although it is not obvious *a priori* that a local minimum of the function $S_a^+ - S_a^-$ exists away from $q = q_0$ and $q = q_1$, this follows from the existence of minimizing heteroclinics, that we already proved. More precisely, we have:

- The minimal value of $S_a^+ - S_a^-$ is $\xi(a)$.
- The point (T, Θ, Q) is a global minimum of $S_a^+ - S_a^-$ if and only if either $Q \in \{q_0, q_1\}$ or the orbit of the point $\big(T, \Theta, Q, a + \partial_\theta S^-, \partial_q S^-\big) = \big(T, \Theta, Q, a + \partial_\theta S^+, \partial_q S^+\big)$, is a minimizing heteroclinic between $\mathcal{T}_0(a)$ and $\mathcal{T}_1(a)$.
- The set of minima of the function $S_a^+ - S_a^-$ properly contains $\mathbb{T} \times \mathbb{T}^m \times \{q_0\} \cup \mathbb{T} \times \mathbb{T}^m \times \{q_1\}$.

As a consequence, the trajectory $(\theta(t), q(t), \dot\theta(t), \dot q(t))$ is a minimizing heteroclinic if and only if $(S_a^+ - S_a^-)(t \bmod 1, \theta(t), q(t)) = \xi(a)$ for each $t \in \mathbb{R}$ (and if $q(t)$ is not identically q_0 or q_1). This minimizing heteroclinic is called isolated if, for some $t \in \mathbb{R}$, the point $(\theta(t), q(t))$ is an isolated minimum of the function

$$(\theta, q) \longmapsto (S_a^+ - S_a^-)(t \mod 1, \theta, q).$$

Now we have proved that the stable and unstable manifolds of the torus $\mathcal{T}(a)$ necessarily intersect, let us suppose that there exists a compact and connected set $A \subset \mathbb{R}^m$ such that the intersection is transversal for $a \in A$. By a continuity argument as in Proposition 1, we conclude that any two points a^- and a^+ in A can be connected by a transition chain, that is a sequence $a_0 = a^-, a_1, \dots, a_n = a^+$ such that the unstable manifold of $\mathcal{T}(a_{i-1})$ transversally intersects the stable manifold of $\mathcal{T}(a_i)$. We would like to deduce the existence of a transversal heteroclinic orbit between $\mathcal{T}(a^-)$ and $\mathcal{T}(a^+)$, but I do not know whether the higher codimensional analog of Cresson's transitivity Lemma 4 holds. However, the weaker obstruction property proved in [21, 31] is enough to imply the existence of orbits connecting any neighborhood of $\mathcal{T}(a^-)$ to any neighborhood of $\mathcal{T}(a^+)$. It is also possible to build shadowing orbits using a variational approach. We need the slightly different assumption that $A \subset \mathbb{R}^m$ is a compact connected set such that, for all $a \in A$, all the minimizing homoclinics of $\mathcal{T}(a)$ are isolated. For each a^- and a^+ in A, it is then possible to construct by a variational method similar to Section 2.4 a heteroclinic orbit between $\mathcal{T}(a^-)$ and $\mathcal{T}(a^+)$.

3. The general *a priori* unstable case

A very specific feature of all the examples studied so far is that the perturbation preserves the partially hyperbolic invariant tori $\mathcal{T}(a), a \in \mathbb{R}^m$. We now discuss the general *a priori* unstable system (2).

3.1. The Large Gap Problem. Let us assume that $r = 1$ and try to apply the method of Section 2. There is no explicit invariant torus any more, but KAM methods can be applied to prove the existence of many partially hyperbolic tori. More precisely, there exists a diffeomorphism

$$\omega_\mu(a) : \mathbb{R}^m \longrightarrow \mathbb{R}^m,$$

close to the identity, such that an invariant quasiperiodic Torus $\mathcal{T}_\mu(a)$ of frequency $\omega_\mu(a)$ exists, and is close to $\mathcal{T}(a)$, provided the frequency $\omega_\mu(a)$ satisfies some Diophantine condition. Moreover, for such values of a, the local stable and unstable manifolds $\mathcal{W}^\pm_\mu(a)$ can be generated by functions

$$S^\pm_{a,\mu}(t, \theta, q) : \mathbb{T} \times \mathbb{T}^m \times [-3/4, 3/4] \longrightarrow \mathbb{R},$$

as earlier. So we have exactly the same picture as in Section 2, except that the objects are defined only on a subset $A_\mu \subset \mathbb{R}^m$ of parameters. In order to reproduce the mechanism of Section 2, we must find elementary transitions chains $a_0, \dots, a_k$ in $\mathbb{R}^m$, with the additional requirement that $a_i \in A_\mu$. It is necessary at this point to describe a bit more the set A_μ. Roughly, the KAM methods allow to prove the existence of the Torus $\mathcal{T}_\mu(a)$ provided a belongs to

$$A_\mu = \big\{ a : \quad k \cdot (1, \omega_\mu(a)) \geqslant \frac{\sqrt{\mu}}{\|k\|^\tau} \quad \forall k \in \mathbb{Z}^{m+1} - \{0\} \big\}$$

for some constant $\tau \geqslant m + 1$. This set A_μ is totally disconnected, hence it is not possible to apply a continuity method like in Proposition 1 in order to prove the existence of a transition chain. We must be more quantitative, which is possible when μ is so small that the Poincaré-Melnikov approximation is valid. In that regime, we have $\bar{\Delta}_{a,\mu} \approx \mu M_a$, where M_a has a non-degenerate critical point. The conclusion of Lemma 3 can then be proved to hold under the more explicit condition that $\|a' - a\| \leqslant C\mu$. In other words, the sequence $a_0, a_1, \ldots, a_k$ is an elementary transition chain if $\|a_i - a_{i-1}\| \leqslant C\mu$. However, the gaps in A_μ have a width of size $\sqrt{\mu} > C\mu$. As a consequence, for small μ, it seems impossible to build long transition chains, and the method fails. This is the Large Gap Problem, see [35]. Even if there are classes of examples where the method can be applied because more tori exist in some regions of phase space, see [13, 19, 8] for example, the generic case seems out of range.

3.2. Normally hyperbolic invariant cylinder.

The Large Gap problem has now been solved, at least in the case where $m = 1$, see [46, 17, 18, 26, 5, 44]. We will not discuss and compare all these solutions here, but just expose some general ideas which arise from them.

An important new point of view is to focus on the whole cylinder $\mathcal{C} = \cup_a \mathcal{T}(a)$ rather than on each of the tori $\mathcal{T}(a)$ individually. This cylinder is Normally hyperbolic in the sense of [34, 30], and thus it is preserved in the perturbed system. This new point of view is very natural, it appears in [43, 24], and then in many other papers. The deformed cylinder $\mathcal{C}_\mu$ contains all the preserved tori $\mathcal{T}_\mu(a)$ obtained by KAM theory. The restricted dynamics is described by an *a priori* stable system on $\mathbb{T} \times \mathbb{T}^m \times \mathbb{R}^m$. If $m > 1$, we are confronted to our lack of understanding of the *a priori* stable situation. If $m = 1$, however, the restricted system is the suspension of an area preserving twist map, and we can exploit the good understanding of these systems given by Birkhoff theory which has also been interpreted (and extended) variationally in the works of Mather [40, 41]. We consider this case ($m = 1$) from now on. The invariant 2-tori which are graphs are of particular importance (they correspond to rotational invariant circles of the time-one map). To each of these invariant graphs, we can associate two real numbers, the rotation number ω (defined from Poincaré theory of circle homeomorphisms), and the area a, which is the symplectic area of the domain of the cylinder $\mathcal{C}_\mu \cap \{t = 0\}$ delimited by the zero section and by the invariant graph under consideration. If a given invariant graph $\mathcal{T}$ of the restricted dynamics has irrational rotation number (or is completely periodic), then there is no other invariant graph with the same area a. We can take a two-covering and associate to this graph two functions $S_a^{\pm}$ by formula similar to (10). They generate the local stable and unstable manifold of the Torus $\mathcal{T}$, the correspond to the global minima of the difference of the so-called barrier function $S_a^+ - S_a^-$. Minimal homoclinics and isolated minimal homoclinics to $\mathcal{T}$ can be defined as in Section 2.6. The existence of minimal homoclinics can be proved basically in the same way as it was there.

Definition 8. *An invariant graph is called a transition torus if it has irrational rotation number (or if it is foliated by periodic orbits), and if all its minimal homoclinic orbits are isolated.*

Transition tori can be used to build transition chains in the same way as partially hyperbolic quasiperiodic tori with transversal homoclinics. Let $A \subset \mathbb{R}$ be the set of areas of transition tori. To each $a \in A$ is attached a unique transition torus $\mathcal{T}_\mu(a)$ (note that this torus may be only Lipschitz, and is not necessarily quasiperiodic). If A contains an interval $[a^-, a^+]$, then the existence of a heteroclinic orbit between $\mathcal{T}_\mu(a^-)$ and $\mathcal{T}_\mu(a^+)$ can be proved by already exposed methods (considering the way we have chosen our definitions, a variational method should be used, but a parallel geometric theory could certainly be given).

In general, the set A is totally disconnected, and transition chains can't be obtained by a simple continuity method. If we make the additional hypothesis that all invariant graphs of the restricted dynamics are transition tori, then the set A is closed and a connected component $]a^-, a^+[$ of its complement corresponds to a "region of instability" of the restricted system in the terminology of Birkhoff. More precisely, the tori $\mathcal{T}_\mu(a^-)$ and $\mathcal{T}_\mu(a^+)$ enclose a cylinder which does not contain any invariant graph. The theory of Birkhoff then implies that there exist orbits of the restricted dynamics connecting an arbitrarily small neighborhood of $\mathcal{T}_\mu(a^-)$ to an arbitrarily small neighborhood of $\mathcal{T}_\mu(a^+)$. This gives an indication about how to solve the large gap problem: use the Birkhoff orbits to cross regions of instability, and the Arnold homoclinic mechanism to cross transition circles. It is by no means obvious to prove the existence of actual orbits shadowing that kind of structure. In order to do so, one should first put these mechanisms into a common framework. The variational framework seems appropriate, although a geometric approach is also possible. The Birkhoff theory was described and extended using variational methods by Mather in [40], and he proposed a new variational formalism adapted to higher dimensional situations in [41]. On the other hand, Bessi's method indicates how to put Arnold's mechanism into a variational framework. These heuristics lead to:

Theorem 3. *Let $[a^-, a^+]$ be a given interval. If all the invariant graphs of area $a \in [a^-, a^+]$ of the restricted dynamics are transition tori, then there exists an orbit $(\theta(t), q(t), \dot\theta(t), \dot q(t))$ and a time $T > 0$ such that $\dot\theta(0) \leqslant a^-$ and $\dot\theta(T) \geqslant a^+$.*

This theorem is proved using variational methods and weak KAM theory in [5], Section 11, where it is deduced from more general abstract results. It also almost follows from [18], Theorem 5.1, which is another general abstract result proved by elaborations on Mather's variational methods [41], see also [4]. Applying that result of Cheng and Yan, however, would require a minor additional generic hypothesis on the restricted dynamics. In the case where $r = 1$, a slightly weaker version of Theorem 3 could also be deduced from the earlier paper of Chen and Yan [17]. Under different sets of hypotheses, results in the same spirit have been obtained by geometric methods in [32, 33]. At the moment, these methods do not reach statements as general as Theorem 3, but they apply in contexts where the variational methods can't be used.

The following variant of Theorem 3 may deserve attention in connection to the Arnold Mechanism: Assume that a^- and a^+ belong to A, or in other words that there exist transition tori $\mathcal{T}_\mu(a^\pm)$. These tori enclose a compact invariant piece $\mathcal{C}_\mu[a^-, a^+]$ of the invariant cylinder. If all the invariant graphs contained in $\mathcal{C}_\mu[a^-, a^+]$ are transition tori, then we say that $\mathcal{C}_\mu[a^-, a^+]$ is a transition channel.

The proof of Theorem 3 also implies that, if $\mathcal{C}_\mu[a^-, a^+]$ is a transition channel, then there exists a heteroclinic orbit connecting $\mathcal{T}_\mu(a^-)$ to $\mathcal{T}_\mu(a^+)$.

Theorem 3 proves the existence of diffusion under "explicit" conditions. These conditions are hard to check on a given system, but they seem to hold for typical systems. It is much harder than one may expect to prove a precise statement in that direction, but it was achieved by Cheng and Yan in [17, 18]. The main difficulty comes from the condition on the isolated minimal homoclinics. Actually, it is not hard to prove that the homoclinics to a given torus are isolated for a typical perturbation, but we need the condition to hold for all the tori simultaneously. Since there are uncountably many tori, it is necessary to understand the regularity of the map $a \longmapsto S_a^\pm$. Recall that the functions $S_a^\pm$ are well-defined provided there exists an invariant graph of area a which has irrational rotation number or is foliated by periodic orbits. We call $\tilde{A}$ this set of areas, it contains A. Cheng and Yan prove that the map $a \longmapsto S_a^\pm$ is Hölder continuous on $\tilde{A}$, and deduce the genericity result using an unpublished idea of John Mather.

4. Back to the *a priori* stable case

The main objects in Arnold's mechanism are partially hyperbolic tori, that he called whiskered tori. It was proved by Treshchev [46], that whiskered tori exist in the *a priori* stable situation, see also [27, 45]. However, because of the Large Gap Problem, it seems difficult to prove directly the existence of transition chains made of whiskered tori. Actually, small transition chains do exist, because the density of KAM tori increases near a given one, but the length of these chains gets small when ϵ gets small, hence these chains do not produce instability of the action variables in general.

The modern paradigm on the *a priori* unstable case that we exposed in Section 3.2 elects 3-dimensional normally hyperbolic invariant cylinders as the important structure. It is well-known that normally hyperbolic invariant cylinders exist in the *a priori* stable case. For example, each 2-dimensional whiskered torus has a center manifold, which is a 3-dimensional normally hyperbolic invariant cylinder, see *e. g.*[12]. Actually, it is simpler to prove directly the existence of normally hyperbolic invariant cylinders, this involves no small divisors. However, the most direct proofs seem to produce "small" normally hyperbolic cylinders, which means that their size is getting small with ϵ, so that we face the same problem as above when we had small transition chains. The main statement of [6] is that "large" normally hyperbolic cylinders exist, meaning that their size is bounded from below independently of ϵ.

In order to be more specific, let us select a resonant momentum of the form $y_0 = (I_0, 0) \in \mathbb{R} \times \mathbb{R}^{n-1}$, with I_0 Diophantine. Assuming that the corresponding averaged potential V has a unique minimum and that this minimum is non-degenerate, we have:

Theorem 4 ([6]). *There exists two intervals $[a^-, a^+] \subset J$, J open, both* independent from ϵ*, and $\epsilon_0 > 0$ such that, for $\epsilon \in]0, \epsilon_0[$ the following holds:*

There exists a C^1 map

$$(Q,P) : \mathbb{T} \times \mathbb{T} \times J \ni (t,\theta,I) \longmapsto \big(Q(t,\theta,I), P(t,\theta,I)\big) \in \mathbb{T}^{n-1} \times \mathbb{R}^{n-1}$$

such that the flow is tangent to the graph Γ of (Q,P). Moreover, there exist two real numbers $a_0 < a^-$ and $a_1 > a^+$ in J (which depend on ϵ) such that the Treshchev tori $\mathcal{T}(a_0)$ and $\mathcal{T}(a_1)$ exist and are contained in Γ. The part Γ_0^1 of Γ delimited by these two tori is then a compact invariant manifold with boundary of the flow, it is normally hyperbolic. It is equivalent to say that it is partially hyperbolic with a central distribution equal to the tangent space of Γ. The inner dynamics is the suspension of an area-preserving twist map (where the area is the one induced from the ambient symplectic form).

It is then reasonable to expect that, under generic additional hypotheses, Γ_0^1 is a transition channel as defined in Section 3.2, and thus that $\mathcal{T}(a_0)$ and $\mathcal{T}(a_1)$ are connected by a heteroclinic orbit. We are currently exploring that program in collaboration with Vadim Kaloshin and Ke Zhang. It is important to observe that the map (Q,P) is not C^1-close to $(0,0)$, and that the inner dynamics is not close to integrable. Fortunately, Theorem 3 allows such a generality.

References

[1] V. I. Arnold, *Instability of dynamical systems with several degrees of freedom*, *Sov. Math. Doklady* **5**, 581–585 (1964).

[2] P. Bernard. *Perturbation d'un hamiltonien partiellement hyperbolique*, Comptes rendus de l'Académie des sciences. Série 1, Mathématique **323** no. 2, 189–194, (1996).

[3] P. Bernard, *Homoclinic orbits to invariant sets of quasi-integrable exact maps*, Ergodic Theory and Dynamical Systems **20** no. 6, 1583–1601 (2000).

[4] P. Bernard. *Connecting orbits of time dependent lagrangian systems,* Ann. Institut Fourier **52** no.5, 1533–1568 (2002).

[5] P. Bernard, *The dynamics of pseudographs in convex Hamiltonian systems*, J. A. M. S. **21** no. 3, 615–665 (2008).

[6] P. Bernard, *Large normally hyperbolic cylinders in a priori stable Hamiltonian systems*, preprint (2009).

[7] M. Berti, P. Bolle, *A functional analysis approach to Arnold diffusion*, Ann. I. H. P. Analyse Non Linaire **19** no. 4, 395–450 (2002).

[8] M. Berti, L. Biasco, P. Bolle, *Drift in phase space: a new variational mechanism with optimal diffusion time.* J. Math. Pures Appl. **82** no. 6, 613–664 (2003).

[9] U. Bessi, *An approach to Arnold's diffusion through the calculus of variations* Nonlinear Analysis, T. M. A., **26** no. 6, 1115-1135 (1996).

[10] U. Bessi, *Arnold's example with three rotators*, Nonlinearity, **10**, 763-781 (1997).

[11] U. Bessi, *Arnold's Diffusion with Two Resonances*, Journal of Differential Equations, **137** no. 2, 211-239 (1997).

[12] S.V. Bolotin, D.V. Treschev, *Remarks on the definition of hyperbolic tori of Hamiltonian systems* Regular and Chaotic dynamics, **5** no. 4, 401–412 (2000).

[13] S.V. Bolotin, D.V. Treschev, *Unbounded growth of energy in nonautonomous Hamiltonian systems* Nonlinearity **12**, 365–388 (1999).

[14] S.V. Bolotin, *Homoclinic orbits in invariant tori of Hamiltonian systems*, Dynamical systems in classical mechanics, 21–90, Amer. Math. Soc. Transl. Ser. 2, **168**, Amer. Math. Soc., Providence, RI, (1995).

[15] S.V. Bolotin, *Infinite number of homoclinic orbits to hyperbolic invariant tori of Hamiltonian systems*, Regular and Chaotic Dynamics **5** no. 2, 139–156 (2000).

[16] J. Bourgain, V. Kaloshin, *On diffusion in high-dimensional Hamiltonian systems,* J. Funct. Anal. **229** no. 1, 1–61 (2005).

[17] C.-Q. Cheng, J. Yan, *Existence of diffusion orbits in a priori unstable Hamiltonian systems*, J. Differential Geom. **67** no. 3, 457–517 (2004).

[18] C.-Q. Cheng, J. Yan, *Arnold Diffusion in Hamiltonian systems: the a priori unstable case*, J. Differential Geom. **82** no. 2, 229-277 (2009).

[19] L. Chierchia and G. Gallavotti, *Drift and diffusion in phase space* Ann. Inst. H. Poincaré Phys. Théor. **60** no. 1, 1-144 (1994).

[20] G. Contreras, G. Paternain, *Connecting orbits between static classes for generic Lagrangian systems*, Topology **41** no. 4, 645-666 (2002).

[21] J. Cresson, *A λ-lemma for partially hyperbolic tori and the obstruction property*, Lett. Math. Phys. **42** no. 4, 363–377 (1997).

[22] J. Cresson, *Un λ-lemme pour des tores partiellement hyperboliques*, C. R. A. S. série 1, **331** no. 1, 65–70 (2000).

[23] J. Cresson, *Temps d'instabilité des systèmes hamiltoniens initialement hyperboliques,* C. R. A. S. série 1, **332** no. 9, 831–834 (2001).

[24] A. Delshams, R. de la Llave, T. M. Seara, *A Geometric Approach to the Existence of Orbits with Unbounded Energy in Generic Periodic Perturbations by a Potential of Generic Geodesic Flows of* $\mathbb{T}^2$, Communications in Mathematical Physics, **209** no. 2, 353–392 (2000).

[25] A. Delshams, R. de la Llave, T. M. Seara, *A Geometric Mechanism for diffusion in Hamiltonian Systems Overcoming the Large Gap Problem: Heuristics and Rigorous Verification on a Model,* Mem. A.M.S. **179** no. 844 (2006).

[26] A. Delshams, R. de la Llave, T. M. Seara, Orbits of unbounded energy in quasi-periodic perturbations of geodesic flows, *Adv. in Math.* **202**,64-188 (2006).

[27] L. H. Eliasson, *Biasymptotic solutions of perturbed integrable Hamiltonian systems*, Bull. Braz. Math. Soc. **25** no. 1, 57-76 (1994).

[28] A. Fathi, Orbites heteroclines et ensemble de Peierls, *Comptes rendus de l'Académie des sciences. Série 1, Mathématique* **326** no. 10, 1213–1216 (1998).

[29] A. Fathi, *Weak KAM theorem in Lagrangian dynamics.* ghost book.

[30] N. Fenichel, *Persistence and smoothness of invariant manifolds for flows*, Indiana Univ. Math. J. **21**, 193–226 (1971).

[31] E. Fontich, P. Martin, *Differentiable invariant manifolds for partially hyperbolic tori*, Nonlinearity **13**, 1561–1593 (2000).

[32] M. Gidea, C. Robinson, *Shadowing orbits for transition chains of invariant tori*, Nonlinearity **20**, 1115-1143, (2007).

[33] M. Gidea, C. Robinson, *Obstruction argument for transition chains of Tori interspersed with gaps* Discrete Contin. Dyn. Syst. Ser. S **2** no. 2, 393–416 (2009).

[34] M.W. Hirsch, C.C. Pugh, M. Shub, *Invariant manifolds*, Lecture notes in Math. Springer Berlin, New York, (1977).

[35] P. Lochak, *Arnold diffusion; a compendium of remarks and questions*, in *Hamiltonian systems with three or more degrees of freedom*, 168–213, Kluwer Academic Publishers, (1999).

[36] P. Lochak, J-P. Marco, *Diffusion times and stability exponents for nearly integrable analytic systems,* Cent. Eur. J. Math. **3** no. 3, 342–397 (2005).

[37] P. Lochack, J. P. Marco, D. Sauzin, *On the Splitting of Invariant Manifolds in Multidimensional Near-Integrable Hamiltonian Systems*, Mem. A.M.S. **163** no. 775 (2003).

[38] J. P. Marco, *Transition le long des chaines de tores invariants pour les systèmes hamiltoniens analytiques*, Annales de l'I. H. P. Physique théorique **64** no. 2, 205–252 (1996).

[39] J. P. Marco, D. Sauzin, *Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems*, Publ. Math. Inst. Hautes tudes Sci. **96**, 199–275 (2003).

[40] J. N. Mather, *Variational construction of orbits of twist diffeomorphisms*, J.A.M.S. **4** no. 2, 207-263 (1991).

[41] J. N. Mather, *Variational construction of connecting orbits*, Ann. Inst. Fourier **43**, 1349-1368 (1993).

[42] J. N. Mather, *Arnold diffusion: announcement of results*, J. Math. Sci. (N. Y.) **124** no. 5, 5275–5289 (2004).

[43] R. Moeckel, *Transition Tori in the Five-Body Problem*, J.D.E. **129**, 290–314 (1996).

[44] R. Moeckel, *Generic drift on Cantor sets of annuli*, in *Celestial Mechanics*, Contemp. Math. 292, A.M.S., 163–171 (2002).

[45] L. Niederman, *Dynamics around simple resonant tori in nearly integrable Hamiltonian systems,* J. Differential Equations **161** no. 1, 1–41 (2000).

[46] D. Treshchev, *The Mechanism of destruction of resonance tori of Hamiltonian systems*, Math. USSR Sb. **68** no. 1, 181–203 (1991).

[47] D. Treshchev, *Evolution of slow variables in a priori unstable Hamiltonian systems*, *Nonlinearity* **17** no. 5, 1803–1841 (2004).

[48] Z. Xia, *Arnold diffusion and instabilities in Hamiltonian dynamics*, preprint (2002).

[49] K. Zhang, *Speed of Arnold diffusion for analytic Hamiltonian systems*, preprint (2009).

CEREMADE, UMR CNRS 7534, Place du Marechal de Lattre de Tassigny, 75775 Paris cedex 16, France
E-mail: patrick.bernard@ceremade.dauphine.fr