A GEOMETRIC MECHANISM FOR DIFFUSION IN HAMILTONIAN SYSTEMS OVERCOMING THE LARGE GAP PROBLEM: ANNOUNCEMENT OF RESULTS

AMADEU DELSHAMS, RAFAEL DE LA LLAVE, AND TERE M. SEARA

Abstract. We present a geometric mechanism for diffusion in Hamiltonian systems. We also present tools that allow to verify it in a concrete model. In particular, we verify it in a system which presents the large gap problem.

1. Background and preliminaries

About 40 years ago, the paper [Arn64] introduced the remarkable Hamiltonian

\[ H_{\varepsilon, \mu}(A_1, A_2, \varphi_1, \varphi_2, t) = H_0 + \varepsilon H_{\varepsilon} + \mu H_{\mu} \]

\[ = \frac{1}{2} A_1^2 + \frac{1}{2} A_2^2 + \varepsilon (\cos \varphi_1 - 1) + \mu (\cos \varphi_1 - 1)(\sin \varphi_2 + \cos t) \]

and showed that for \( 0 < \mu \ll \varepsilon \ll 1 \), the flow of Hamiltonian (1) contains orbits such that, for some \( T > 0 \), \( |A_2(T) - A_2(0)| \geq 1 \).

This result is in marked contrast with the behavior for \( \varepsilon = \mu = 0 \) for which \( A_1, A_2 \) are conserved quantities and it is all the more surprising since the KAM theorem (proved slightly earlier than [Arn64]) showed that the phase space of the system—except for a set of measure \( O(\varepsilon + \mu)^{1/2} \)—is covered by quasi-periodic solutions for which \( |A(t) - A(0)| \leq K(\varepsilon + \mu)^{1/2} \) for all real \( t \), where \( A = (A_1, A_2) \) and \( K \) is a constant independent of \( \varepsilon, \mu \).

Motivated by applications, related phenomena of instability were studied numerically and a wealth of heuristic understanding was accumulated (see e.g. [Chi79]).

An essential ingredient in the geometric description of the argument in [Arn64] is that the flow of Hamiltonian (1) preserves for all \( \varepsilon, \mu \) a one-parameter family of 2-dimensional tori

\[ T_a = \{(A, \varphi, t) : A_1 = 0, A_2 = a, \varphi_1 = 0, (\varphi_2, t) \in \mathbb{T}^2 \}. \]

This is a consequence of the fact that all the terms in (1) which involve \( \varepsilon, \mu \) contain as a factor \( (\cos \varphi_1 - 1) \) which vanishes up to second order in \( \varphi_1 \).

When \( \varepsilon > 0 \), the tori \( T_a \) have 3-dimensional stable and unstable manifolds, so they are called whiskered tori. By performing first order perturbative
calculations in $\mu$, it is possible to show that the stable and unstable manifolds of nearby tori cross transversally.

As shown in a classic calculation in [Poi99, §403–409] (modern expositions emphasizing the Lagrangian character of the manifolds and its consequences are [DR97, DG00]) the distance between the modified manifolds associated to $T_a$ can be expressed as $\mu \nabla \phi_2(t) L(a, \phi_2, t) + O(\mu^2)$.

In the case of (1), the function $L$, usually called Melnikov potential, has the form

$$L(a, \phi_2, t) = 2\pi \left( \frac{a^2 \sin \phi_2}{\sinh(\frac{a}{2\sqrt{\epsilon}})} - \frac{\cos t}{\sinh(\frac{t}{2\sqrt{\epsilon}})} \right).$$

Note that the function $L$ is exponentially small in $\epsilon$, but if we choose

$$\mu \ll \max \left( \exp(-\frac{\pi}{2}a\epsilon^{-1/2}), \exp(-\frac{\pi}{2}\epsilon^{-1/2}) \right),$$

the first order in $\mu$ of the perturbative calculation of the intersection is much larger than the remainder and it allows us to conclude that the stable and unstable manifolds of a torus $T_a$ intersect transversally. Hence, the unstable manifold of $T_a$ intersects transversally the stable manifold of $T_{a+\delta}$ for $\delta \ll |\mu \nabla L|$.

Therefore, it is possible to construct a sequence of tori $\{T_{a_i}\}_{i=0}^{N}$ such that the unstable manifold of one intersects transversally the stable manifold of the next and $a^1 = 0, a^N = 1$.

Then, an argument—which can be made essentially topological—shows that there is an orbit following all the intersections. More details on the paper [Arn64] can be found in [AA67, DLS03a, FM03].

In [Arn63, p. 176] one can find the formulation of the unsolved problem of establishing that the mechanism in [Arn64] happens for many Hamiltonian systems. The formulation of the problem already includes the hint that one would need to consider resonances more carefully.

If instead of the term containing $\mu$ in (1) we would have included a more general function which does not happen to vanish to second order at $\phi_1 = 0$, the family $T_a$ would not have been preserved since the tori $T_a$ for which the frequency $A_2 = a$ is rational are destroyed (See [Poi99, §81], [Tre89]).

At resonances (regions where the frequency $A_2$ is rational), the family of tori is interrupted by gaps whose size is at most of order $O(\mu^{1/2})$. This is much larger than the distance that the manifolds move in the range of parameters (2) that we required to ensure that the first order perturbation theory in $\mu$ gave conclusive results. The fact that the gaps on the family of tori are larger than the size of the intersections of the stable and unstable manifolds—for the ranges of perturbations considered—is what in the literature is called the large gap problem.

The heuristic discoveries of numerical exploration show that the phenomenon of diffusion is strongest precisely on resonances. Indeed this is the most important guide in the numerical exploration and in applications. See
Overcoming the large gap problem

[Chi79, Ten82, CLSV85]. In contrast, the diffusion exhibited in [Arn64] happens precisely—by design—in a zone which is devoid of resonances since the resonant tori are preserved.

Hence, there seems to have been a divergence between the mathematical literature—that has aimed to verifying the mechanism of [Arn64]—and the physical literature that has explored the role of resonances. See [Dou88, DLC83, FM01, Gal99, Moe96] for rigorous mathematical verifications of the existence of the mechanism of [Arn64] in some specially constructed models as well as [BB02, BBB03] and references there for a functional analysis approach. See [DLS00] and [BT99] for rigorous verification variations of the mechanism in perturbed geodesic flows, which has also been considered in [Mat95]. In all the cases above, the gaps between KAM tori are smaller than the splitting of separatrices. (A more extensive bibliography is available in [DLS03a].)

See also [CG94, CG98] for a rather complete set of tools. See [Ten82, CLSV85] as well as [Chi79] for heuristic discussions of the importance of resonances in diffusion. Among conferences which include papers with the mathematical and physical points of view on diffusion we mention [HRS83, Sim99].

The goal of this note is to describe a geometric mechanism that overcomes the large gap problem. We also describe a rather detailed verification of the existence of this mechanism in a concrete problem. A more detailed manuscript with full proofs is available in [DLS03a].

Similar problems have been considered in [Tre03, CY03], which we received after finishing this paper.

We also mention that there are detailed announcements of results using variational methods [Mat02] as well as shorter announcements [Xia98].

2. The two dynamics mechanism for diffusion (heuristic)

The first main idea of our proposed mechanism is that, at the same time that resonances destroy KAM whiskered tori, they generate some other objects that can also be used as part of transition chains that generate diffusion.

An example which is well known is that perturbations of integrable twist maps break resonant tori but leave hyperbolic periodic orbits—with their stable and unstable manifolds—and elliptic islands. Note that neither of those objects have analogues in the integrable map. (See [Poi99, §74, 79, 81] [LW89] for some higher dimensional extensions). All these objects generated by the resonances will take up the role of the invariant tori that are destroyed by the resonances.

One convenient way of organizing the many different objects that we encounter is to identify normally hyperbolic invariant manifolds $\tilde{\Lambda}$ that are present in the system. It happens frequently that the stable and unstable manifolds of $\tilde{\Lambda}$ intersect transversally. This is commonly described as saying that there are transverse homoclinic intersections to $\tilde{\Lambda}$. An easy dimension
count shows that the intersection of these stable and unstable manifolds will typically contain a manifold that has the same dimension as $\tilde{\Lambda}$.

More precisely, given a normally hyperbolic invariant manifold $\tilde{\Lambda}$ whose stable and unstable manifolds $W^s_{\tilde{\Lambda}}$ intersect transversally, we can associate to it two dynamical systems:

**The inner map:** This is simply, the time one map restricted to $\tilde{\Lambda}$ and we will denote it by $F$.

**The scattering map or outer map:** We also associate another dynamical system to the homoclinic excursions along the prescribed connection.

More precisely, given a family of orbits $\gamma \subset W^s_{\tilde{\Lambda}} \cap W^u_{\tilde{\Lambda}}$ such that the intersection of the stable and unstable manifolds is transversal along $\gamma$—hence $\gamma$ should be a manifold of the same dimension of $\tilde{\Lambda}$—we define the scattering map $S$ associated to $\gamma$ by setting $x_+ = S(x_-)$ when we can find $z \in \gamma$ such that $d(\Phi_t(z), \Phi_t(x_{\pm})) \leq C e^{-\lambda |t|}$ as $t \to \pm \infty$. Here $\Phi_t$ denotes the dynamics of the system and $\lambda > 0$ is the exponent of the normal rate of contraction of the normally hyperbolic manifold.

That is, when performing a homoclinic excursion, the system goes from resembling the orbit of $x_-$ to resembling the orbit of $x_+ = S(x_-)$.

Since the map $S$ is formulated in terms of normally hyperbolic invariant manifolds it is possible to compute it using perturbation theory of normally hyperbolic manifolds. It admits a rather explicit expansion $S = S_0 + \varepsilon S_1 + \varepsilon^2 S_2 + \cdots$.

We note that the scattering map can be considered as an alternative to the standard methods in Melnikov theory. Conceptually, it relies less on coordinates, so that it can be computed quite effectively and can work for all points in $\tilde{\Lambda}$ independently of the motion they experience.

By intermingling the two dynamics, we construct sequences of points $(x_N)_{N \in \mathbb{N}}$, where $x_N = F^{n_N} \circ S \circ \cdots \circ S \circ F^{n_1}(x)$, which diffuse.

Such a sequence can be described heuristically as a pseudo-orbit which stays “parked” near the invariant manifold $\tilde{\Lambda}$ along some convenient times $n_i$ and, at certain times suitably chosen, performs homoclinic excursions.

Under many circumstances, there are variants of the obstruction property which ensure that there are orbits of the original system that track $x_N$ closely. Notably if all the $x_N$ lie in non-resonant tori invariant under $F$, we can use the obstruction argument in [AA67] (a very careful implementation can be found in [FM01]).

As a consequence, if $x_N$ experiences diffusion—and is such that the obstruction property can be applied—then there is an orbit $\{\Phi_t(y)\}$ in the original system which also experiences diffusion.

Note that in Example (1) the manifold $A_1 = 0$, $\varphi_1 = 0$ is normally hyperbolic for $|\varepsilon| > 0$. Hence it persists for $|\mu| \ll 1$.

We emphasize that the persistence of the normally hyperbolic invariant manifold is true for any perturbation $H_\mu$, not just the one used in (1).
If we take the perturbation $H_{\mu}$ to be a generic one we can expect that the dynamics in the manifold $\Lambda$ includes not only the tori that survive from $H_0 + \varepsilon H_\varepsilon$ but also the dynamics associated to the resonances.

3. Rigorous verification in a model

In [DLS03a] we have undertaken a very detailed rigorous verification of the existence of the mechanism described above for models of the form

$$H_\varepsilon = \pm \left( \frac{1}{2} A_1^2 + V(\varphi_1) \right) + \frac{1}{2} A_2^2 + \varepsilon h(A_1, A_2, \varphi_1, \varphi_2, t; \varepsilon)$$

The system (3) presents the large gap problem and has been a very popular model for the study of diffusion since [HM82].

A particular case of the results of [DLS03a] is:

**Theorem 1.** Assume that a Hamiltonian of the form (3) satisfies:

**H1.** The terms $V$ and $h$ in (3) are uniformly $C^r$ for $r \geq r_0$, which is sufficiently large.

**H2.** The potential $V : \mathbb{T} \to \mathbb{R}$ has a global maximum at 0 which is non-degenerate (i.e., the second derivative is not zero). We denote by $(A_1^0(t), \varphi_1^0(t))$ an orbit of the pendulum $\pm \left( \frac{1}{2} A_1^2 + V(\varphi_1) \right)$ homoclinic to $(0, 0)$.

**H3.** $h$ is a trigonometric polynomial in $\varphi$ and $t$:

$$h(A_1, A_2, \varphi_1, \varphi_2, t; \varepsilon) = \sum_{k,l \in \mathbb{N}} \hat{h}_{k,l}(A_1, A_2, \varphi_1; \varepsilon)e^{i(k\varphi_2 + lt)},$$

where $\mathbb{N} \subset \mathbb{Z}^2$ is a finite set.

**H4.** Consider the Poincaré function, also called Melnikov potential, associated to $h$ (and to the homoclinic orbit $(A_1^0, \varphi_1^0)$) mentioned in **H2**:

$$\mathcal{L}(A_2, \varphi_2, t) = -\int_{-\infty}^{+\infty} \left( h(A_1^0(\sigma), A_2, \varphi_1^0(\sigma), \varphi_2 + A_2\sigma, t + \sigma; 0) \right. - h(0, A_2, 0, \varphi_2 + A_2\sigma, t + \sigma; 0) \left. \right) d\sigma.$$

Assume that, for any value of $A_2 \in (a^-, a^+)$ there exists an open set $\mathcal{J}_{A_2} \subset \mathbb{T}^2$, with the property that when $(A_2, \varphi_2, t) \in H_-$, where

$$H_- = \bigcup_{A_2 \in (a^-, a^+)} \{A_2\} \times \mathcal{J}_{A_2} \subset (a^-, a^+) \times \mathbb{T}^2,$$

the map

$$\tau \in \mathbb{R} \mapsto \mathcal{L}(A_2, \varphi_2 - A_2\tau, t - \tau)$$

has a non-degenerate critical point $\tau$ which is locally given, by the implicit function theorem, in the form $\tau = \tau^*(A_2, \varphi_2, t)$ with $\tau^*$ a smooth function.
Assume moreover that for every \((A_2, \varphi_2, t) \in H_-,\) the function

\[
\frac{\partial L}{\partial \varphi_2}(A_2, \varphi_2 - A_2 \tau^*, t - \tau^*)
\]

is non-constant and negative (respectively positive).

**H5.** The perturbation terms \(h(A_1, A_2, \varphi_1, \varphi_2, t; 0)\), \(\frac{\partial h}{\partial \epsilon}(A_1, A_2, \varphi_1, \varphi_2, t; \epsilon)\), satisfy some non-degeneracy conditions that can be stated quite explicitly.

Then, for \(0 < |\epsilon| \leq \epsilon^*\), the system (3) has orbits such that \(A_2(0) \leq a^- + O(\epsilon)\), \(A_2(T) \geq a^+ + O(\epsilon)\) (respectively \(A_2(0) \geq a^- + O(\epsilon)\), \(A_2(T) \leq a^- + O(\epsilon)\)).

**Remark 2.** Since [CG94] is sometimes customary to distinguish between *a-priori stable* and *a-priori* unstable systems. Following this terminology, our model (3) is an a priori-unstable system because the unperturbed Hamiltonian has no global action angle variables (in fact, it has partially hyperbolic tori with homoclinic trajectories).

This distinction makes sense only for one-parameter families. When one considers two-parameter families as in [Arn64] or generic results, the results for a-priori unstable maps automatically imply results for perturbations of an a-priori stable Hamiltonian in a cusp residual set.

For example, if we consider the a-priori stable system with two parameters

\[
H_{\delta, \epsilon} = \pm \left(\frac{1}{2} A_1^2 + \delta V(\varphi_1)\right) + \frac{1}{2} A_2^2 + \epsilon h(A_1, A_2, \varphi_1, \varphi_2, t; \epsilon)
\]

satisfying non-degeneracy conditions, we obtain similar results as in Theorem 1 if we assume, as in [Arn64], that \(\epsilon\) is (exponentially) small with respect to \(\delta\). Note, that in (8) the gaps between KAM tori at a non-degenerate primary resonance are \(B\epsilon^{1/2} + O(\epsilon)\), the splitting of separatrices is \(O(\exp(-C\delta^{-1/2})) + O(\epsilon^2)\) and the distance between secondary and primary tori is, again, smaller than \(O(\epsilon^{3/2})\) \(\square\)

The verification undertaken in [DLS03a] is very explicit so that, given \(h, V\), a finite number of calculations can establish that (3) satisfies the mechanism outlined heuristically above, mainly H4 and H5.

The verification undertaken in [DLS03a] consists of a sequence of perturbation theories. Most of them are somewhat standard in geometric theory of diffusion. They need to be developed with very quantitative statements so that they form a coherent toolkit. We highlight here the main steps.

(1) Using the theory or normally hyperbolic invariant manifolds, we show:

(a) The manifold \(\tilde{\Lambda} = \{A_1 = 0, \varphi_1 = 0\}\), which is invariant and normally hyperbolic for \(\epsilon = 0\), persists for \(|\epsilon| \ll 1\) giving rise to \(\tilde{\Lambda}_\epsilon\). We compute expansions for \(\tilde{\Lambda}_\epsilon\) and for the Hamiltonian restricted to it.
(b) Under the non-degeneracy conditions in $H4$, the stable and unstable manifolds $W^{s,u}_{\tilde{\Lambda}_{\varepsilon}}$ of the normally hyperbolic invariant manifold $\tilde{\Lambda}_{\varepsilon}$—which agree for $\varepsilon = 0$—have transverse intersections.

(c) Given the transverse intersections of the previous point, the scattering map $S$ can be computed in first order perturbation theory.

One of the conclusions of the calculations is that the scattering map $S$ is close to the identity in a $C^{r'}$ sense, where $r'$ can be taken arbitrarily large if $r$ is sufficiently large.

(2) The motion restricted to the normally hyperbolic invariant manifold $\tilde{\Lambda}_{\varepsilon}$ is analyzed as follows:

(a) Using that the system restricted to $\tilde{\Lambda}_{\varepsilon}$ is a periodic perturbation of a one degree of freedom Hamiltonian system, averaging theory shows that the system can be transformed up to small errors into a time independent system. Near the resonances, the system can be accurately described by systems similar to a pendulum.

(b) Far away from the resonances, using that a time independent one degree of freedom Hamiltonian is integrable, we can use the standard KAM theorem to show that there are closely spaced KAM tori.

(c) Near the resonances we switch to singular action-angle variables and show that, under appropriate non-degeneracy conditions, we can find

i) KAM tori close to the separatrices of the pendulum.

ii) Secondary KAM tori—tori which are contractible to a periodic orbit—close to the separatrices of the pendulum.

iii) Stable and unstable manifolds of periodic orbits close to the separatrices of the pendulum.

Note that the objects ii), iii) above are not present in the unperturbed system but are generated by the resonances. We refer to these objects as secondary objects.

We emphasize that the objects in i), ii), iii) can be made to be very close to each other. In particular, the secondary tori of ii) are very close to the KAM tori of i) and the stable and unstable manifolds of iii) are in between.

These secondary objects dovetail in the gap produced by the resonances among the KAM tori. See Figure 1.

(3) Under appropriate non-degeneracy conditions—which are reflected in the non-vanishing of certain terms of the perturbation theory of the map $S$—it will happen that the image under the scattering map $S$ of some of the objects considered in (b) and (c) of step 2 intersect transversally in $\tilde{\Lambda}_{\varepsilon}$ other such objects.
A moment’s reflection shows that if $S(\mathcal{V}_1), \mathcal{V}_2$ intersect transversally as submanifolds of $\hat{\Lambda}_\varepsilon$ then $W^u_{\mathcal{V}_1}$ intersects $W^s_{\mathcal{V}_2}$ transversally as submanifolds of the phase space.

Since the objects we constructed in step 2 are closer than $\varepsilon^{3/2}$, and the scattering map moves objects by an amount $O(\varepsilon)$, some non-degeneracy assumptions will allow us to establish that they intersect transversally under the action of the scattering map. Therefore, all these objects form what is called a transition chain.

Note that the method allows us to establish without problem the existence of transition chains containing objects that have different topological type and which do not have analogues in the unperturbed problem.

(4) The obstruction methods can be adapted to show that there are orbits that track the escaping pseudo-orbits.

Even if some of the earlier proofs of the obstruction properties use coordinate systems, we point out that there are more recent proofs which make it clear that the obstruction property is true independently of a common system of coordinates in the tori.

The motion on the normally hyperbolic invariant manifold is depicted on Figure 1, where we indicated the objects and their relative positions and distances.

![Figure 1. Surface of section of $\hat{\Lambda}_\varepsilon$ illustrating the main invariant objects. The primary KAM tori and the secondary tori on the left. The primary tori and the stable and unstable manifolds of periodic orbits on the right.](image-url)
In Figure 2, we indicate the effect of the scattering map on the objects previously found.

![Figure 2](image-url)

**Figure 2.** Schematic illustration of the main invariant objects in $\tilde{\Lambda}_\epsilon$ as well as their images (dashed) under the scattering map.

### 4. Final remarks

The proof of Theorem 1 shows that there exists a transition chain with a torus $O(\epsilon)$ close to $T_{a^-}$ and another torus $O(\epsilon)$ close to $T_{a^+}$ and so that the tori composing this transition chain are closely spaced. Note that under hypothesis $H2$ there are indeed two homoclinic orbits of the pendulum $\pm(\frac{1}{2}A_1^2 + V(\varphi_1))$. Moreover, for each of these orbits, the function given in (6) will have several critical points, which are typically non-degenerate. If two of these choices give us intervals $[a_1^-, a_1^+]$, $[a_2^-, a_2^+]$ and $a_1^+ > a_2^-$, then, the result produces a transition chain that starts near $a_1^-$ and ends near $a_2^+$. Besides orbits that diffuse along the action $A_2$, the geometric mechanism used in the proof of Theorem 1 provides orbits that visit the tori in a somewhat arbitrary order.

Note that the transition chains produced in [Arn64] increase the action $O(\epsilon)$ per step in the transition chain. The transition chains produced here take steps $O(\epsilon^{1/2})$ at resonances. This agrees with the numerical and heuristic intuition which suggests that the diffusion is faster precisely at resonances.

We note that the verification in [DLS03a] uses mainly tools which are rather standard in the geometric approach to diffusion. Of course, they
require adaptations so that they can work together and form an efficient toolkit. We hope that this toolkit can be used for other problems. We also hope that new tools can be incorporated to the toolkit. Notably, variational methods and more topological methods. Indeed, a very similar mechanism to the one described in Section 2 was implemented in [DLS00] to give a geometric proof of the results in [Mat95]. In turn, ingredients of the geometric proof were used to simplify some of the variational arguments.

A similar toolkit has been used in [DLS03b] to verify the existence of orbits of unbounded energy of quasi-periodic perturbations of geodesic flows—for Riemannian, Lorentz or Finsler metrics—in many manifolds. We think it would be interesting to find variational analogues of the later result.

It seems to us that, if one does not insist on verifying the results for concrete systems, but rather show that the mechanism happens for quasi-integrable generic systems, the verification could be somewhat simplified. We hope that the modularity of the method could encourage to use topological or variational methods for some of the steps. We also note that the method presented here admits several variants with different quantitative properties. Diffusion seems to include all the variety of possible mechanisms.

5. Acknowledgments

The work described in this announcement has been supported by the Comisión Conjunta Hispano Norteamericana de Cooperación Científica y Tecnológica. A first version was prepared while R.L. was enjoying a Cátedra de la Fundación FBBV, and A.D. was visiting the Centre de Recerca Matemàtica. A.D. and T.S. have also been partially supported by the Catalan grant 2001SGR-70 and the INTAS grant 00-221, and R.L. by NSF grants.

REFERENCES


Overcoming the large gap problem


