Arnold's diffusion in isochronous systems

G. Gallavotti

Università di Roma 1, Fisica
3 settembre 1998 (revised)

*Abstract: I discuss an illustration of a mechanism for Arnold's diffusion following a non-variational approach, and an improvement of the related estimates for the diffusion time.**

Keywords: Arnold's diffusion, homoclinic splitting, KAM

§1: Introduction.

Arnold's diffusion was established in simple paradigmatic examples by Arnold, [A]. Since that paper several methods aiming at extending its validity to more general systems have been developed: this was done either by following methods sometimes called "geometric methods" close to the original approach, [CG],[C],[M], or by "variational methods", [Be], [Br]. In the approach [CG] one finds estimates, for the time necessary for a diffusion of $O(1)$ in the space of the action variables, that are terribly big as functions of the size ε of the perturbation when it approaches 0 (their order is $\gg \exp O(\varepsilon^{-1})$); the variational method instead gives better estimates, "*fast*",([Be], their order is $\exp O(\varepsilon^{-1/2})$), and even very good, "*polynomial*", ones ([Br], their order is $O(\varepsilon^{-2})$).

Recently remarkable progress has been achieved in the geometric approach via the papers [M] and the impressive "summa" [C], who have been able to recover not only the best variational results but to extend them to the cases discussed in [CG], greatly improving the bounds obtained there, and to many substantially new cases of applicative interest. The work [C] gives an extensive bibliography to which I refer. However the subject is still presented at a very technical level, and the relation of the new methods with those in [CG] is not transparent.

Here I first illustrate (§5) the method of [CG] by developing it with the aim of showing existence of diffusion. This may lead to a clarification of a method not appropriately quoted in the literature and which maintains its interest, because of its relative simplicity, in spite of the better estimates coming from the quoted alternative methods. If explicit estimates are avoided one gains enormously in simplicity: this kind of approach was probably the one meant in [A] where the problem was first posed and solved without bothering to give the (fairly obvious, see §5) details. What follows in §5 applies also to the Arnold's case, but I prefer to illustrate it in a case that is even simpler.

Furthermore I show (§6) that if a new idea is added to the method of [CG] then one can get a "fast" (still exponential) estimate for the drift time. This bound is derived in detail and is conceptually independent of the other works.

Consider Hamiltonians $\mathcal{H}$ with three degrees of freedom described by coordinates $I \in R$, $\underline{A}_0 = (A_{01}, A_{02}) \in R^2$ and angles $\varphi \in T^1$, $\underline{\alpha} = (\alpha_1, \alpha_2) \in T^2$:

$$\mathcal{H} = \underline{\omega} \cdot \underline{A}_0 + \frac{I^2}{2} + g^2(\cos\varphi - 1) + \varepsilon f(\varphi, \underline{\alpha}) \stackrel{def}{=} \mathcal{H}_0 + \varepsilon f(\varphi, \underline{\alpha}) \tag{1.1}$$

where $\underline{\omega} = (\omega_1, \omega_2) \in R^2$ is a vector with Diophantine constants C, τ, *i.e.* such that for all integer components vectors $\underline{\nu} = (\nu_1, \nu_2)$ it is $|\underline{\omega} \cdot \underline{\nu}|^{-1} \le C|\underline{\nu}|^\tau$ if $\underline{\nu} \ne \underline{0}$; the *perturbation* f is supposed to be a (fixed) trigonometric polynomial of degree N: $f(\varphi, \underline{\alpha}) = \sum_{0<|\underline{\nu}|<N, |n|<N} f_{n,\underline{\nu}} \cos(n\varphi + \underline{\nu} \cdot \underline{\alpha})$. The subject being fairly well under-

---

* The first version of this paper is archived in: chao-dyn@xyz.lanl.gov # 9709011.

stood we do not need to be really very careful about units so that some coefficients in (1.1) have been set equal to 1.

One can also use Jacobi's hyperbolic coordinates, which we denote p_0, q_0, to describe the pendulum $\frac{1}{2}I^2 + g^2(\cos\varphi - 1)$ near the unstable point $I = \varphi = 0$, see for instance Appendix A1 of[G4]. In the new coordinates the pendulum Hamiltonian becomes $J(p_0 q_0)$ with $J'(x) \stackrel{def}{=} \frac{dJ(x)}{dx} = g + \sum_{n=1}^{\infty} g_n x^n \stackrel{def}{=} g(x)$ and the total Hamiltonian becomes:

$$\mathcal{H} = \underline{\omega} \cdot \underline{A}_0 + J(p_0 q_0) + \varepsilon f_0(\underline{\alpha}, p_0, q_0) \tag{1.2}$$

where $f_0(\underline{\alpha}, p_0, q_0) = f(\varphi, \underline{\alpha})$.

The function f_0, still a trigonometric polynomial in $\underline{\alpha}$, has the property: $f_0(\underline{\alpha}, p, q) = f_0(-\underline{\alpha}, q, p) = f_0(-\underline{\alpha}, -p, -q)$ and we shall call *parity* the 4–elements group of transformations generated by $P_1 : (\underline{\alpha}, p, q) \longleftrightarrow (-\underline{\alpha}, q, p)$ and $P_2 : (\underline{\alpha}, p, q) \longleftrightarrow (-\underline{\alpha}, -p, -q)$. We call P_0, P_1, P_2, P the group elements; and we say that f_0 has "even parity". If $F(\underline{\alpha}, p, q) = -F(P_j(\underline{\alpha}, q, p)) \stackrel{def}{=} P_j F(\underline{\alpha}, q, p)$, $j = 1, 2$, we say that F has odd parity: for instance $\partial_{\underline{\alpha}} f_0$ has "odd parity". The p derivative F or the q derivative G of an even function have the property $F = P_1 G = -P_2 F, G = P_1 F = -P_2 G$. The Jacobi's map in general transforms functions of φ, $\underline{\alpha}$ with given parity in $\underline{\alpha}, \varphi$ in the ordinary sense into functions with the same parity in the $(\underline{\alpha}, p, q)$ variables.

§2. Invariant tori and nearby flow.

We look for a change of coordinates $(\underline{A}_0, \underline{\alpha}, p_0, q_0) \longleftrightarrow (\underline{A}, \underline{\psi}, p, q)$ which integrates locally (1.2) near the unstable equilibrium of the pendulum. More precisely so that in the new coordinates the motion is:

$$\underline{A} = const, \;\; \underline{\psi} \to \underline{\psi} + \underline{\omega}\, t, \;\; p \to p e^{-(1+\gamma)gt}, \;\; q \to q e^{(1+\gamma)gt} \tag{2.1}$$

where $g = g(pq), \gamma = \gamma(pq)$ and $x \to \gamma(x)$ is a suitable function analytic in x near $x = 0$ while $g(x) = J'(x)$ (see (1.2)). We shall attempt to write the change of coordinates:

$$\begin{aligned} \underline{A}_0 &= \underline{A} + \underline{H}(\underline{\psi}, p, q), \qquad & p_0 &= p + L(\underline{\psi}, p, q) \\ \underline{\alpha} &= \underline{\psi}, & q_0 &= q + \tilde{L}(\underline{\psi}, p, q) \stackrel{def}{=} q + L(-\underline{\psi}, q, p) \end{aligned} \tag{2.2}$$

where $\underline{H}(\underline{\psi}, p, q)$ has even parity, and zero $\underline{\psi}$ average at $p = q = 0$. Setting $\not\partial \stackrel{def}{=} q\partial_q - p\partial_p$ and imposing that (2.2) and (2.1) verify the equations of motion one gets:

$$\begin{aligned} \big(g(x)\not\partial + \underline{\omega} \cdot \partial_{\underline{\psi}}\big)\, \underline{H} &= -\,\varepsilon \partial_{\underline{\alpha}} f_0(\underline{\psi}, p_0, q_0) - g(x)\gamma(x)\not\partial\, \underline{H} \\ \big(g(x) + g(x)\not\partial + \underline{\omega} \cdot \partial_{\underline{\psi}}\big)\, L &= -\,\varepsilon \partial_{q_0} f_0(\underline{\psi}, p_0, q_0) - (g(x_0) - g(x))p_0 + \\ &\quad + \gamma(x) g(x) p - g(x)\gamma(x)\not\partial\, L \end{aligned} \tag{2.3}$$

where $x = pq$, $x_0 = p_0 q_0$, $p_0 = p + L$, $q_0 = q + \tilde{L}$. In fact the second equation is independent on the first. Both can be shown to admit, for ε small enough as we shall always suppose below, a solution analytic in ε and divisible by ε. Note that the unknowns are $\underline{H}, L, \gamma$. This is a very simple theorem. A "classical" (*i.e.* by quadratic iterations) proof can be derived from §5 in [CG] where the harder anisochronous case is detailed; or see [G4] (p. 2–3 and p. 9) for a self contained elementary analysis (see also [DGJS] where the same theorem is also explicitly proved in p. 52–62). A better proof is essentially in the basic paper [Ge]: basic in the sense that it discusses the general theory of the manifolds from what I think is a "basic" point of view: *i.e.* it follows the Eliasson's method using bounds on power series expansions, [E] (as developed in [G2], [G3] [GG]).

In what follows we *fix* ε so small that the above functions $\underline{H}, L, \tilde{L}$ are well defined by their power series in ε. The variables $\underline{A}$ will be called *average actions* because they are the average values with respect to time of the actions of the motions that take place on the corresponding invariant tori.

§3. Stable and unstable manifolds.

From (2.2) we can read the following facts (of course if ε is small enough):

(1) Phase space contains a family of invariant tori $\mathcal{T}(\underline{A})$ parameterized by $\underline{A} \in R^2$ and obtained by setting $p = q = 0$. The average position of the tori is precisely $\underline{A}$, because $\underline{H}$ can (and will) be chosen to have zero average: average with respect to $\underline{\psi}$ or to time.[1]

(2) Given $\underline{A}$ and setting $q = 0$, $p \neq 0$ one obtains a surface whose points are parameterized by $\underline{\psi}, p$ and which is a local piece of the stable manifold $W^s(\underline{A})$ of $\mathcal{T}(\underline{A})$. The quantity $-(1+\gamma(x))g(x)$ is the "Lyapunov exponent" of $W^s(\underline{A})$. Likewise, setting $p = 0$, $q \neq 0$ one defines a local piece of the unstable manifold $W^u(\underline{A})$ of $\underline{A}$ and $(1+\gamma(x))g(x)$ is the corresponding exponent.

(3) Given $\underline{A}$ and setting $q \neq 0, p \neq 0$ one parameterizes the rest of phase space *near* $\mathcal{T}(\underline{A})$. In this part of phase space the motion is in some sense very regular.

(4) However the motions, just described locally, are globally more interesting and chaotic. In fact generically $W^u(\underline{A}), W^s(\underline{A}')$ do intersect transversally if $\underline{A}$, $\underline{A}'$ are close enough (depending on ε) and $\mathcal{T}(\underline{A}), \mathcal{T}(\underline{A}')$ have the same energy. In such cases $W^u(\underline{A}) \cap W^s(\underline{A}')$ consists of trajectories, or *heteroclinic intersections*, running asymptotically around $\mathcal{T}(\underline{A})$ as $t \to -\infty$ and around $\mathcal{T}(\underline{A}')$ as $t \to +\infty$.
The symmetry of the problem implies, (as it is well known, see for instance §9 of [CG]), that if $\underline{A} = \underline{A}'$ and $\varphi = \pi$ then the point of $W^s(\underline{A})$ with $\varphi = \pi$, $\underline{\alpha} = \underline{0}$ is *homoclinic*, *i.e.* it is on a trajectory on $W^u(\underline{A}) \cap W^s(\underline{A})$ (provided ε is small so that the results of §2 apply). Generically the intersection $W^u(\underline{A}) \cap W^s(\underline{A}')$ exists for all $\underline{A}$, $\underline{A}'$ close enough and is *transversal* in the sense that if we fix $\varphi = \pi$ (or φ to any other value $\neq 0, 2\pi$) then any pair of tangents to $W^u(\underline{A})$ and $W^s(\underline{A}')$ at common points form an angle $\geq \mu > 0$; the bound μ depends on ε, of course, and generically can be taken proportional to ε.

(5) Analytically we can write $\underline{A}' + \underline{A}^s(\underline{\alpha}, \varphi)$ and $\underline{A} + \underline{A}^u(\underline{\alpha}, \varphi)$ the parametric equation of the manifolds $W^u(\underline{A}')$ and $W^s(\underline{A})$, so representable at least for φ away from 0 or 2π, see (2.2) (*e.g.* see [CG] or [G3]). The functions $\underline{A}^u$, $\underline{A}^s$ do not depend on $\underline{A}$ because of isochrony:[2] to see this note that the evolution equations for $I, \varphi, \underline{\alpha}$ do not involve the $\underline{A}$'s; explicit expressions for $\underline{A}^s$, $\underline{A}^u$ can be found in in [G3] or [GGM].
The equation for the $\underline{\alpha}$ value of an intersection point in $W^u(\underline{A}) \cap W^s(\underline{A}')$ with $\varphi = \pi$ (say) is just $\underline{Q}(\underline{\alpha}) \stackrel{def}{=} \underline{A}^s(\underline{\alpha}, \pi) - \underline{A}^u(\underline{\alpha}, \pi) = \underline{A}' - \underline{A}$, where usually $\underline{Q}(\underline{\alpha})$ is called the *splitting vector* at $\underline{\alpha}$ (and $\varphi = \pi$). The angles between the tangent planes to $W^u(\underline{A})$ and $W^s(\underline{A})$ at the homoclinic intersection at $\varphi = \pi$, $\underline{\alpha} = \underline{0}$ are related to the eigenvalues of the *intersection matrix* $D_{ij} \stackrel{def}{=} \partial_i Q_j(\underline{\alpha})|_{\underline{\alpha}=\underline{0}}$ which is *also the Jacobian* of the implicit equation $\underline{Q}(\underline{\alpha}) = \underline{A}' - \underline{A}$ near $\underline{\alpha} = \underline{0}$..
It follows from the classical Melnikov's theory of splitting (see for instance [G3]) that the eigenvalues of D have *generically* values of order $O(\varepsilon)$ so that the angles between tangents to $W^u(\underline{A})$ and $W^s(\underline{A})$ at $\underline{\alpha} = \underline{0}, \varphi = \pi$ will have size $O(\varepsilon)$ (and $\det D = O(\varepsilon^2)$). The genericity condition is a very simple algeraic condition on the coefficients of the

[1] This is a simple special case of a property which becomes rather nontrivial in more interesting "anisochronous" Thirring models. Such models (see [T] and [G2], [G3]) differ from (1.1) because of a possible addition to $\mathcal{H}_0$ of an extra term $\frac{1}{2K}\underline{A}^2$ with K constant; then the average action $\underline{A}$ of the motion on a torus is directly related to the gradient of the *unperturbed* Hamiltonian via $\underline{\omega} = \partial_{\underline{A}} \mathcal{H}_0(\underline{A})$, *i.e.* the frequencies are not "twisted" by the perturbation (a fact apparently "discovered" in [G3]).

[2] Note that the parametric equation for the I variable needs not to be specified as it follows from the ones for the $\underline{A}$'s via the energy conservation.

polynomial f and is easily verified in many examples: the very simplest being perhaps $f(\underline{\alpha},\varphi)=\cos(\alpha_1+\varphi)+\cos(\alpha_2+\varphi)$.

The non vanishing of the intersection matrix determinant, and its interpretation as Jacobian of the implicit equation for the heteroclinic intersections, implies that the latter exist always, as soon as the average actions of the tori are close enough (and the tori have the same energy, of course).

(6) One can define also the splitting in the φ variables: call $X^u_-(t), X^s_-(t)$ the values at time t of the φ coordinate of the point on the unstable or stable manifold which at time $t=0$ has coordinates $(\underline{A}^u(\underline{\alpha},\pi), \underline{\alpha}, I^u(\underline{\alpha},\pi),\pi)$ or $(\underline{A}^s(\underline{\alpha},\pi), \underline{\alpha}, I^s(\underline{\alpha},\pi),\pi)$ and one sets $\Delta(t)=X^s_-(t)-X^u_-(t)$, which also depends on $\underline{\alpha}$.

(7) Finally a definition: Let $\underline{A}_0, \underline{A}_1,\ldots,\underline{A}_{\mathcal{N}}$ be a sequence such that $|\underline{A}_j-\underline{A}_{j+1}|$ is so small that $W^u(\underline{A}_j)\cap W^s(\underline{A}_{j+1})$ have a transversal heteroclinic intersection, in the above sense, with intersection angles $\ge\mu$ at $\varphi=\pi$. We call such a chain a *heteroclinic chain* or *ladder*. As remarked in (5) one finds generically and in most simple examples $\mu=O(\varepsilon)=O(\mathcal{N}^{-1})$.

We shall prove the following theorem ("Arnold's diffusion" or "drift"):

Theorem 1: *Let $\underline{A}_0, \underline{A}_1,\ldots,\underline{A}_{\mathcal{N}}$ be a heteroclinic chain: for any $\delta>0$ there are trajectories starting within δ of $\mathcal{T}(\underline{A}_0)$ and arriving after a finite time T_{drift} within δ of $\mathcal{T}(\underline{A}_{\mathcal{N}})$.*

I shall give a complete proof of it (§5), again, along the lines of [CG] for the sake of illustrating the *simplicity* of the method (due to Arnold). The purpose being of showing the conceptual difference with respect to the variational approaches, which accounts for the impressive difference in the time scale of T_{drift} compared with [Be], [Br] or with the estimate in theorem 2 below (see (6.9)). In §6 I give a more refined, yet very simple and detailed, proof getting explicit and much better bounds ("fast") although still far from the best in the literature.

§4. Geometric concepts.

Let $2\kappa>0$ be smaller than the radius of the disk in the (p,q)–plane where the functions in (2.2) are defined. We call κ a "target parameter".

To visualize the geometry of the problem involving 2–dimensional tori and their 3–dimensional stable and unstable manifolds, in the 5–dimensional energy surface, we shall need the following geometric objects:

•(a) a point X_i, heteroclinic between $\mathcal{T}(\underline{A}_i)$ and $\mathcal{T}(\underline{A}_{i+1})$, which has local coordinates, see (2.2), $X_i=(\underline{A}_i, \underline{\psi}_i, 0,\kappa)$.

•(b) the equations, at fixed $q=\kappa$, of the connected part of $W^s(\underline{A}_{i+1})$ containing X_i, in the local coordinates near $\mathcal{T}(\underline{A}_i)$; they will be written as:

$$Y_i(\underline{\psi})=(\underline{A}^s_{i+1}(\underline{\psi}), \underline{\psi}, p^s_{i+1}(\underline{\psi}),\kappa) \tag{4.1}$$

with $|\underline{\psi}-\underline{\psi}_i|<\zeta$ for some $\zeta>0$ (i–independent): it is $\underline{A}^s_{i+1}(\underline{\psi}_i)=\underline{A}_i$, $p^s_{i+1}(\underline{\psi}_i)=0$ because we require $Y_i(\underline{\psi}_i)=X_i$. There are constants F',F such that $|\underline{A}^s_{i+1}(\underline{\psi})-\underline{A}^s_{i+1}(\underline{\psi}_i)|$ and $\max_{|\underline{\psi}-\underline{\psi}_i|=fixed}|p^s_{i+1}(\underline{\psi})|$ are bounded, for ζ small enough, below by $F'|\underline{\psi}-\underline{\psi}_i|$ and above by $F|\underline{\psi}-\underline{\psi}_i|$; the constants F',F have size bounded below by $O(\mu)$ (by the transversality assumption in the definition of heteroclinic chain, (7) of §3).

Note that $W^s(\underline{A}_{i+1})$ also contains a part with local equations $(\underline{A}_{i+1}, \underline{\psi}, p, 0)$ which *should not* to be confused with the previous one described by the function $Y_i(\underline{\psi})$. This is more easily understood by looking at the meaning of the above objects in the original

$(\underline{A}, \underline{\alpha}, I, \varphi)$ coordinates: in a way the first part of $W^s(\underline{A}_{i+1})$ is close to $\varphi = 0$ and the second to $\varphi = 2\pi$. They can be close because of the periodicity, but they are conceptually quite different.

•(c) a point $P_i = Y_i(\tilde{\underline{\psi}}_i)$ with $|\tilde{\underline{\psi}}_i - \underline{\psi}_i| = r_i$, where $\tilde{\underline{\psi}}_i, r_i$ will be determined recursively, and a neighborhood B_i:

$$B_i = \{|\underline{A} - \underline{A}^s_{i+1}(\underline{\psi})| < \kappa^2 r_i',\ |\underline{\psi} - \tilde{\underline{\psi}}_i| < r_i',\ |p - p^s_{i+1}(\underline{\psi})| < \kappa r_i',\ q = \kappa\} \tag{4.2}$$

where $r_i' < r_i$ are scales < 1 and to be determined recursively. If $\overline{g}, 2\overline{g}$ are lower/upper bounds to $(1+\gamma(x))g(x)$ for $|x| < 4\kappa^2$, the point P_i evolves in a time $T_i \simeq \overline{g}^{-1}\log\kappa^{-1}$ into a point X_i' near $\mathcal{T}(\underline{A}'_{i+1})$ which has local coordinates $X_i' = (\underline{A}_{i+1}, \underline{\psi}'_i, \kappa, 0)$. Note that any point $(\underline{A}, \underline{\psi}, p, q)$ will evolve, provided it does not exit the neighborhood where the local coordinates are defined, into a point of the form $(\underline{A}, \underline{\psi} + \underline{\omega}\, T_{in}, q, p)$ after a time $T_{in} = -g(x)^{-1}(1+\gamma(x))^{-1}\log qp^{-1}$ if $x = pq$, because of the special hyperbolic form of the time evolution, see (2.1): we shall call this time the *interchange time* of the "last two coordinates" and we shall repeatedly use it.

The choice of B_0 is rather arbitrary and we take $r_0 = \zeta$ (ζ is defined after (4.1)) and $r_0' = \frac{1}{2}r_0$, choosing $\tilde{\underline{\psi}}_0$ arbitrarily (at distance r_0 from $\underline{\psi}_0$).

•(d) The points ξ of the set B_i are mapped by the time evolution to points that, at the beginning at least, come close to $\mathcal{T}(\underline{A}_{i+1})$ and in a time $\tau(\xi)$ acquire local coordinates near $\mathcal{T}(\underline{A}_{i+1})$ with $p = \kappa$ exactly: the time $\tau(\xi)$ is of the order of $\overline{g}^{-1}\log\kappa^{-1}$.

If S_t is the time evolution flow for the system (1.1) we write $S\xi = S_{\tau(\xi)}\xi$ (note that S depends also on i). Then S maps the set B_i into a set SB_i containing:

$$B_i' = \{|\underline{A} - \underline{A}_{i+1}| < \frac{1}{E}\kappa^2 r_i',\ |\underline{\psi} - \underline{\psi}'_i| < \frac{1}{E}r_i',\ p = \kappa,\ |q| < \frac{\kappa}{E}r_i'\} \tag{4.3}$$

because all the points in B_i with $\underline{A} = \underline{A}^s_{i+1}(\underline{\psi})$, $p = p^s_{i+1}(\underline{\psi})$, $q = \kappa$ evolve (each taking its own time) to points with $\underline{A} = \underline{A}_{i+1}$, $p = \kappa$, $q = 0$ and $\underline{\psi}$ close to $\underline{\psi}'_i$, by the definitions. Here E is a bound on the Jacobian matrix of S. The latter, being essentially a flow over a time $O(\overline{g}^{-1}\log\kappa^{-1})$, has derivatives bounded i–independently: since we suppose that ε is "small enough" we could take $E = 1 + b\varepsilon$ for some $b > 0$ if, as often the case, $|\underline{A}_i - \underline{A}_{i+1}| < O(\varepsilon)$.

§5. The [CG]-method of proof of the theorem.

Consider the points $Y_{i+1}(\underline{\psi}) \in W^s(\underline{A}_{i+2})$ with coordinates $(\underline{A}^s_{i+2}(\underline{\psi}), \underline{\psi}, p^s_{i+2}(\underline{\psi}), \kappa)$. They will evolve backwards in time so that $\underline{A}$ stays constant, $\underline{\psi}$ evolves quasi-periodically hence "rigidly", and $p^s_{i+2}(\underline{\psi})$ evolves to κ while the q–coordinate evolves from κ to $q = p^s_{i+2}(\underline{\psi})$ (because pq stays constant, see (c) above). The time for this evolution is $T_{\underline{\psi}} \simeq \overline{g}^{-1}\log\kappa|p^s_{i+2}(\underline{\psi})|^{-1} \xrightarrow[\underline{\psi}\to\underline{\psi}_{i+1}]{} +\infty$.

Therefore there is a sequence $\underline{\psi}^n \xrightarrow[n\to+\infty]{} \underline{\psi}_{i+1}$ such that $|p^s_{i+2}(\underline{\psi}^n)| > 0$, $p^s_{i+2}(\underline{\psi}^n) \to 0$ $\underline{A}^s_{i+2}(\underline{\psi}^n) \to \underline{A}_{i+1}$ *and* $\underline{\psi}^n - \underline{\omega}\, T_{\underline{\psi}^n} \xrightarrow[n\to\infty]{} \underline{\psi}'_i$, as a consequence of the Diophantine properties of $\underline{\omega}$. So that there is $\tilde{\underline{\psi}}_{i+1} \stackrel{def}{=} \underline{\psi}^n$ with n suitable and a point $P_{i+1} = (\underline{A}^s_{i+2}(\tilde{\underline{\psi}}_{i+1}), \tilde{\underline{\psi}}_{i+1}, p^s_{i+2}(\tilde{\underline{\psi}}_{i+1}), \kappa) \in W^s(\underline{A}_{i+2})$ (actually infinitely many) which evolves, backwards in time, from P_{i+1} to a point of B_i'.

Hence we can define $r_{i+1} = |\tilde{\underline{\psi}}_{i+1} - \underline{\psi}_{i+1}|$ and r'_{i+1} small enough so that the backward motion of the points in B_{i+1} enters in due time into B_i'. It follows that the set B_i evolves in time so that all the points of B_{i+1} are on trajectories of points of B_i, *for all* $i = 1, \ldots, \mathcal{N}$. Hence all points of $B_{\mathcal{N}}$ will be reached by points starting in B_0.

This completes the proof. All constants can be estimated explicitly, even though this is somewhat long and cumbersome, see [CG]. The result is an extremely large drift time

T_{drift} (namely the value at $\mathcal{N}$ of a composition of $\mathcal{N}$ exponentials! at least this is the estimate I get after correcting an error in §8 of [CG]: the error is minor but leads to substantially worse bounds).

Nevertheless the estimate that comes out of the above scheme seems essentially optimal. And then the problem is: "how is it possible that by other methods (*e.g.* variational methods of [Be],[Br]) one can get *far better* estimates?

The above argument is quite close to the proof of the "obstruction property" in [C], p. 34: hence the latter work shows that the above analysis misses some key idea that is exploited in the papers [M], [C]; perhaps the possibility of setting up a symbolic dynamics around the tori and exploiting it in the bounds.

The difference with respect to the variational methods may be due to the fact that they are "less constructive": less so than the above. The "fast drifting" trajectory exists but there seems to be no algorithm to determine it, not even the sequence of its "close encounters" with the invariant tori that generates drift: which is in fact *preassigned* in the above method. This certainly might account for a difference in the estimates. In fact the above construction is far too rigid: we pretend not only that drift takes place but also that it takes place via a path that visits closely a *prescribed sequence* of tori in an essentially *predetermined* way. In §6 a less constructive method is proposed and used to obtain bounds: which, however, turn out to be still far from polynomial. Of course a better understanding of why the results are so different with the different methods is highly desirable: but my efforts to understand satisfactorily this point only led to the improvement in §6 below, which has nevertheless some interest as it introduces the notion of elastic heteroclinic chain which I think might be useful also for the analysis of the anisochronous cases.

§6. Fast diffusion: elastic heteroclinic chains.

The following adds a new idea to the method exposed in §5, allowing us to improve the super-exponential estimate of [CG] mentioned there. Below ε will be fixed small enough, and $\overline{g}, 2\overline{g}$ will be lower and upper bounds, respectively, to $g(x)(1+\gamma(x))$, see (2.1).

Let y be the curvilinear abscissa of a curve $y \to \underline{A}(y)$, $y \in [0, y_{\max}]$, in the "average action space" such that the tori $\mathcal{T}(\underline{A}(y))$ have fixed energy. Then (evaluating the energy at the homoclinic point $\underline{\alpha} = \underline{0}, \varphi = \pi$ and using that the I coordinate of the points on $W^u(\underline{A}), W^s(\underline{A})$ do not depend on $\underline{A}$, because of isochrony, (2.2) one sees that $\underline{\omega} \cdot \underline{A}(y)$ is constant so that the line $y \to \underline{A}(y)$ is parallel to $\underline{\omega}^{\perp} = (\omega_2, -\omega_1)$.

Hence $\underline{A}(y) = \underline{A}_0 + \underline{w}^{\perp} y$ with $\underline{w}^{\perp} \stackrel{def}{=} \frac{\underline{\omega}^{\perp}}{|\underline{\omega}|}$ and $\underline{A}'(y) \stackrel{def}{=} \frac{\partial \underline{A}(y)}{dy} \equiv \underline{w}^{\perp}$.

Define $y \to \underline{A}(y)$, $y \in [0, \overline{y}_{\max}]$, to be an *elastic heteroclinic chain* with flexibility parameters $\beta, \vartheta > 0$ and splitting μ if:

(i) for all $|y-y'| < \vartheta\mu$ there is a heteroclinic intersection between the stable and unstable manifolds of $\mathcal{T}(\underline{A}(y))$ and $\mathcal{T}(\underline{A}(y'))$ with splitting angles $\geq \mu$ at $\varphi = \pi$.

(ii) the intersection matrix $D \stackrel{def}{=} \mu D_o$ at $\varphi = \pi$, $\underline{\alpha} = \underline{0}$ verifies:

$$(\underline{w}^{\perp} \cdot D_o^{-1} \underline{w}^{\perp}) \stackrel{def}{=} \beta \neq 0, \qquad \underline{w}^{\perp} = \frac{\underline{\omega}^{\perp}}{|\underline{\omega}|} \tag{6.1}$$

Remarks:

(a) The above definition is a special case of a natural more general definition relevant for higher dimensions and for anisochronous systems. For instance in the case of anisochronous systems, in which a term $\underline{A}^2/2K$, with $K > 0$ constant, is added to (1.1) one has to require that $y \to \underline{A}(y)$ is a simple rectifiable curve and that, uniformly in $y \in [0, y_{\max}]$, (6.1) holds with D replaced by the intersection matrix D_y, and $\underline{\omega}$ replaced by $\underline{\omega}(\underline{A}(y)) = \underline{\omega} + \underline{A}(y)K^{-1}$.

But in the anisochronous cases the condition that for all y there is the torus $\mathcal{T}(\underline{A}(y))$, called "no gap property", is strongly restrictive and quite artificial (although it is verified in the example in [A], see also [P]).

Below we consider, without further mention, only the isochronous models in (1.1) and in this case D, D_o are y–independent because of isochrony, see (5) in §3.

Condition (6.1) is a transversality property: in the case of (1.1) it holds generically (in the perturbation f and for ε small) and in this case it is a consequence of (i). Thus examples exist and are elementary, and generically $\mu = O(\varepsilon)$. A simple concrete example is provided by the already mentioned perturbation $f(\underline{\alpha}, \varphi) = \cos(\alpha_1 + \varphi) + \cos(\alpha_2 + \varphi)$.

Below we shall also suppose, without further mention, that $\mu = O(\varepsilon)$, i.e. that we consider a generic case. The greater generality of the above definition is meant to clarify a notion that might seem special for the isochronous case, and for future reference.

(b) thus every sequence $y_0, y_1, \ldots, y_{\mathcal{N}}$ with $|y_i - y_{i+1}| < \vartheta\mu$ is a heteroclinic chain in the sense of §3, and the theorem proved in §5 applies to it. A elastic heteroclinic chain with parameter ϑ is also elastic with parameter $\vartheta' < \vartheta$. Hence it will not be restrictive to suppose that ϑ is as small as needed.

(c) if ϑ is small enough so that the first order Taylor's expansions of the splitting vector $\underline{Q}(\underline{\alpha})$, see §3, (5), are "good" approximations we deduce (by applying the implicit functions theorem) that a heteroclinic intersection at $\varphi = \pi$ between $W^s(\underline{A}(y))$ and $W^u(\underline{A}(y+\delta))$ takes place at:

$$\begin{aligned}
&\underline{\alpha}_y(\delta) = D_o^{-1}\,\underline{w}^{\perp}\vartheta' + O(\vartheta'^2) && \text{for} \quad \delta = \mu\vartheta',\ |\vartheta'| \le \vartheta \\
&\frac{1}{2}\beta\,|\tilde\vartheta| < |(\underline{\alpha}_y(\delta') - \underline{\alpha}_y(\delta''))\cdot \underline{w}^{\perp}| < 2\beta\,|\tilde\vartheta|, && \forall\ \delta' = \mu\vartheta',\ \delta'' = \mu\vartheta''
\end{aligned} \tag{6.2}$$

for ϑ small enough, for $|\vartheta'|, |\vartheta''| < \vartheta$, and having set $\tilde\vartheta = \vartheta' - \vartheta''$.

(d) A geometrical consequence of (6.1), (6.2) is that when y varies by δ (so that $\underline{A}(y)$ varies in R^2 *orthogonally* to $\underline{\omega}$ by $O(\delta)$), then the heteroclinic intersection $\underline{\alpha}_y(\delta)$ between $W^s(\underline{A}(y))$ and $W^u(\underline{A}(y+\delta))$ is displaced away from $\underline{0}$ with a component *in the direction orthogonal to* $\underline{\omega}$ of size $O(\delta\mu^{-1})$, provided $\delta\mu^{-1} = \vartheta'$ is small enough.

(e) The value $\varphi = \pi$ is not special in many respects and the same remains true if one looks at the displacement of the heteroclinic intersection at any other section located away from the tori by a fixed distance $\kappa > 0$, if ε is small enough. In fact consider the intersection matrix $D(t)$ evaluated along the heteroclinic trajectory at a time t after the passage through $\varphi = \pi$. From the equations of motion its evolution is:

$$D(t) = D - \varepsilon \int_0^t \partial_{\underline{\alpha}\varphi} f(\varphi(\tau), \underline{\omega}\,\tau)\, \partial_{\underline{\alpha}} \Delta(\tau)\, d\tau \tag{6.3}$$

where $\Delta(t)$ denotes the splitting in the φ–coordinates (and $|\Delta(t)| < O(\varepsilon)$), defined in remark (6) of §3, and $\varphi(t)$ is the heteroclinic evolution of φ: hence $D(t) = D + O(\varepsilon^2)$ (while $D = O(\varepsilon)$) for t bounded, by Melnikov's theory (see also (5.5) in [GGM]).

In particular if we look at the $\underline{\psi}$–coordinate $\underline{\psi}_y(\delta)$ of the heteroclinic intersection point at $q = \kappa$, on the same heteroclinic trajectory, and compare it with the position of the homoclinic point $\underline{\psi}_y(0)$ of $\mathcal{T}(\underline{A}(y))$ at $q = \kappa$ then we can say that, for some constants $2b_1, 2b_0$ (the factor 2 is for later convenience) it is:

$$|\underline{w}^{\perp}\cdot(\underline{\psi}_y(\delta') - \underline{\psi}_y(\delta''))| \in [2b_1\tilde\vartheta, 2b_0\tilde\vartheta] \tag{6.4}$$

with $\tilde\vartheta = (\delta' - \delta'')\mu^{-1}$; the constants b_0, b_1 depend on the constant κ prefixed at the beginning of §6, and on β.

Theorem 2: *Suppose that $y \to \underline{A}(y)$ is elastic in the above sense, then fixed a, b there exist heteroclinic chains $\underline{A}_0 = \underline{A}(y_0), \underline{A}_1 = \underline{A}(y_1), \ldots, \underline{A}_{\mathcal{N}} = \underline{A}(y_{\mathcal{N}})$ with $y_0 = 0, y_{\mathcal{N}} = y_{\max}$ along which the drift time is $\overline{g}^{-1} e^{O(\mu^{-1})}$.*

The estimates proceed by performing the construction of §5 *without* fixing *a priori* the heteroclinic chain: we construct it inductively, by trying to optimize (as well as we can) various choices. *The proof below is divided, to break it into several trivial statements, into a few propositions and lemmata each of which is marked by a •.*

Using the notations of §4 assume that y_j have been constructed for $j \le i+1$ together with $\tilde{\underline{\psi}}_j, r_j, r'_j, B_j, \underline{\psi}'_j, B'_j$ for $j \le i$, verifying $r_i < \vartheta$. We must define $y_{i+2}, \tilde{\underline{\psi}}_{i+1}, r_{i+1}, r'_{i+1}$.

The set B_0 is fixed as in the paragraph following (4.2) above, and $y_1 - y_0 = \mu\vartheta, r_0, r'_0 = \frac{1}{2} r_0$ are arbitrarily chosen (positive) and we also require $r_0 < \vartheta$ and ϑ small.

•$_1$ Let E be as in §5 and let E' be so large that if $T^0 = \overline{g}^{-1} \log E' E^{-1}$ the points $\underline{\omega}\, t$, $t \in [0, T^0]$, fill the torus within $\frac{1}{2} b_1 \vartheta$ (see (6.4) for the definition of b_1). This means that E' is very large.[3]

•$_2$ Let $X_{i+1}(y)$ be heteroclinic between $\mathcal{T}(\underline{A}_{i+1})$ and $\mathcal{T}(\underline{A}(y))$ for $y \in [y_{i+1} + \frac{1}{2}\mu\vartheta, y_{i+1} + \mu\vartheta]$. We choose to look for y_{i+2} among such y's to be sure that every time i increases by one unit then y_i increases by $\frac{1}{2}\mu\vartheta$ at least, so that after $\mathcal{N}$ steps, with $\mathcal{N} = O(\mu^{-1})$ we shall have reached the "upper extreme $y_{\max}$ of the chain".

Let the local coordinates of $X_{i+1}(y)$ be $(\underline{A}_{i+1}, \underline{\psi}_{i+1}(y), 0, \kappa)$ (see §4 for the notations). Let, see (4.1):

$$Y_{i+1,y}(\underline{\psi}) = (\underline{A}^s_{i+2,y}(\underline{\psi}), \underline{\psi}, p^s_{i+2,y}(\underline{\psi}), \kappa) \tag{6.5}$$

be the equation of $W^s(\underline{A}(y))$ in the local coordinates around the torus $\mathcal{T}(\underline{A}_{i+1})$ near $X_{i+1}(y)$. We may suppose that:

$$|\underline{A}^s_{i+2,y}(\underline{\psi}) - \underline{A}_{i+1}|, |p^s_{i+2,y}(\underline{\psi})| < b_3 \mu |\underline{\psi} - \underline{\psi}_{i+1}(y)| \tag{6.6}$$

for some b_3 of $O(1)$ and we may suppose $b_3 > 1$, for simplicity. This expresses simply the analyticity in $\underline{\psi}$ and ε of the stable manifold (note that $\underline{A}^s_{i+1,y}(\underline{\psi}) - \underline{A}_{i+1}$ and $p^s_{i+1,y}(\underline{\psi})$ vanish at $\underline{\psi}_{i+1}(y)$, *i.e.* at the heteroclinic point.

•$_3$ Suppose r small: a first approximation to $\tilde{\underline{\psi}}_{i+1}$ will be obtained by *fixing y at the left extreme $\overline{y}$ of its interval of variation* (which is $[y_{i+1} + \frac{1}{2}\mu\vartheta, y_{i+1} + \mu\vartheta]$) and by choosing a point $\underline{\psi}_{i+1,\overline{y},r}$ at distance r from $\underline{\psi}_{i+1}(\overline{y})$ along a line Λ on which we can be sure that $p^s_{i+2,\overline{y}}(\underline{\psi})$ does not vanish. For instance we can take the straight line Λ_{60^o} at 60^o from the gradient $\underline{a}_{i+2}(\overline{y})$ of $p^s_{i+2,\overline{y}}(\underline{\psi})$ at $\underline{\psi} = \underline{\psi}_{i+1}(\overline{y})$. In this way ($\cos 60^o = \frac{1}{2}$):

$$\lambda \stackrel{def}{=} |p^s_{i+2,\overline{y}}(\underline{\psi}_{i+1,\overline{y},r})| \simeq \frac{1}{2} \max_{|\underline{\psi} - \underline{\psi}_{i+1}(\overline{y})| = r} |p^s_{i+2,\overline{y}}(\underline{\psi})| \tag{6.7}$$

and $\lambda \in [b_2 \mu r, b_3 \mu r]$, for some $b_3, b_2 = O(1) > 0$, by the assumption on the splitting angles (which implies that the modulus of the gradient $\underline{a}_{i+2}(y) \stackrel{def}{=} \partial_{\underline{\psi}} p^s_{i+2,y}(\underline{\psi}_{i+1}(y))$ is in $[b_2\mu, b_3\mu]$ for some constants $b_2, b_3 > 0$ of $O(1)$). The constants b_2, b_3 depend on the "target" parameter κ, fixed once and for all, see beginning of §4, and b_3 can be taken to be the same constant in (6.6). Let $d = b_2/2b_3$.

---

[3] One can take $E' = E \exp O(C\vartheta^{-\tau})$ estimating by $O(C\delta^{-\tau})$ the time needed to a quasi periodic rotation of the torus with vector $\underline{\omega}$, Diophantine with constants C, τ, to fill with lines parallel to $\underline{\omega}$ and within δ the whole torus T^2. I discuss this estimate in Appendix A1 as an aside, since here only finiteness of E' is required, (a trivial fact): this gives me the chance of discussing a simple conjecture.

$\bullet_4$ To improve the approximation for $\tilde{\underline{\psi}}_{i+1}$ note that as r varies in the range $d\frac{r'_i}{4b_3E'} < r < \frac{r'_i}{4b_3E}$ the point $\underline{\psi}_{i+1,\overline{y},r}$ varies and λ varies by a factor not smaller than $2E'/E$ by our definition of d. Hence the time $T(r)$ necessary in order that the backward evolution of the point $Y_{i+1,\overline{y}}(\underline{\psi}_{i+1,\overline{y},r}) = (\underline{A}^s_{i+2,\overline{y}}(\underline{\psi}_{i+1,\overline{y},r}), \underline{\psi}_{i+1,\overline{y},r}, p^s_{i+2,\overline{y}}(\underline{\psi}_{i+1,\overline{y},r}),\kappa)$ *interchanges* the last two coordinates, will vary by an amount $\ge T^0 = \overline{g}^{-1}\log E'/E = O(C\vartheta^{-\tau})$, see footnote [3] and comment (c) in §4.

$\bullet_5$ This implies, by continuity and by the size of T_0, that there will be a value $r(\overline{y})$ such that the "backward motion" $\underline{\psi}_{i+1}(\overline{y}) \to \underline{\psi}_{i+1}(\overline{y}) - \underline{\omega}t$ of duration $T(r(\overline{y}))$ has $\underline{\psi}$–coordinate close to $\underline{\psi}'_i$ within $\frac{1}{2}b_1\vartheta$ *and on the line* ℓ *orthogonal to* $\underline{\omega}$ *through* $\underline{\psi}'_i$.
We can also *prefix on which side* of it it will be. Remark that as y increases past $\overline{y}$ the point $\underline{\psi}_{i+1}(y)$ moves with a displacement having a non zero component in the direction parallel to ℓ (by the second of (6.1) and the first of (6.2)). Therefore we shall choose the side so that the component of the displacement along ℓ is *towards* $\underline{\psi}'_i$: this is convenient for reasons that will become clear below.

$\bullet_6$ I now imagine varying y in its interval of variation $[y_{i+1}+\frac{1}{2}\mu\vartheta, y_{i+1}+\mu\vartheta]$ and select $r(y)$, hence $\underline{\psi}_{i+1,y,r(y)}$, so that the time of intechange of the last two coordinates of the point $Y_{i+1,y}(\underline{\psi}_{i+1,y,r(y)})$ *does not change.* The latter time is, by (2.1), $g(p\kappa)^{-1}(1+\gamma(p\kappa))^{-1}\log\kappa|p|^{-1}$ if $p = p^s_{i+2,y}(\underline{\psi}_{i+1,y,r(y)})$, because the motion is "exponential" and preserves the product of the last two coordinates (see (2.1))
Hence fixing the interchange time means determining $\underline{\psi} = \underline{\psi}_{i+1,y,r(y)}$ so that p is constant. Although it might be clear that this can be done I describe in some detail the way that I follow in Appendix A2 which also gives the quantitative information that:

$$\frac{d}{8b_3E'}r'_i < r(y) < \frac{1}{2b_3E'}r'_i \tag{6.8}$$

and, therefore, the point $Y_{i+1,y}(\underline{\psi})$ remains in the neighborhood where the local coordinates are defined.

$\bullet_7$ For each y the point $\underline{\psi}_{i+1}(y) - \underline{\omega}T(r(\overline{y}))$ will either "fall short" or "long" of the line ℓ orthogonal to $\underline{\omega}$ through $\underline{\psi}'_i$: but "only by a length bounded by $\le 2|D_o^{-1}\underline{w}^{\perp}|\vartheta$". In fact by construction the vector $\underline{\psi}_{i+1}(\overline{y}) - \underline{\omega}T(r(\overline{y}))$ is exactly on the line ℓ and the maximum variation that $|\underline{\psi}_{i+1}(y) - \underline{\psi}_{i+1}(\overline{y})|$ can undergo as y varies by at most $\frac{1}{2}\mu\vartheta$ is bounded by the first of (6.2).

$\bullet_8$ Hence *by a suitable rotation of the direction of the line* Λ_{60° *along which we choose the point* $\underline{\psi}_{i+1,y,r(y)}$ we can change the size of $p^s_{i+2,y}$ by a factor $1+O(\vartheta)$ and arrange that at the time T_{in} when the last two coordinates are interchanged $\underline{\psi}_{i+1}(y) - \underline{\omega}T_{in}$ is *exactly* on the line ℓ orthogonal to $\underline{\omega}$ through $\underline{\psi}'_i$.
In fact our choice of the line Λ_{60° on which $\underline{\psi}_{i+1,y,r(y)}$ is selected, neither orthogonal nor parallel to the gradient of $p_{i+2,y}(\underline{\psi})$, shows that we can change in this way $|p^s_{i+2,y}|$ by *up to* a factor about 2, *i.e.* by a factor $\gg 1+O(\vartheta)$ if ϑ is small, and *down to* a factor 0.[4]
So that, by continuity, we can find a line Λ' slightly off Λ_{60° by an angle of $O(\vartheta)$ and a point $\underline{\psi}$ on it such that in its interchange time T_{in} the *(other)* point $\underline{\psi}_{i+1}(y)$ ends on the target line ℓ.

[4] Note that $p^s_{i+2,y}(\underline{\psi})$ vanishes in correspondence of the heteroclinic point, *i.e.* at $\underline{\psi}_{i+1}(y)$, *as well as* on a curve through the heteroclinic point value $\underline{\psi}_{i+1}(y)$ by the transversality of the splitting and by the implicit functions theorem.

We still call $\underline{\psi}_{i+1,y,r(y)}$ the new (and final) choice of $\underline{\psi}$ on Λ'. *Note, however, that the point* $\underline{\psi}_{i+1,y,r(y)}$ *will not end on the line* ℓ *but it will miss it by at most a distance* $r(y) < r_i'/(2b_3E)$, because $\underline{\psi}_{i+1,y,r(y)}$ is $r(y)$ apart from $\underline{\psi}_{i+1}(y)$, which by the construction ends on ℓ at the interchange time T_{in}: the point reached on ℓ will be away by at most $b_1\vartheta$ from $\underline{\psi}_i'$.

The angle of the needed rotation will be of $O(\vartheta)$ off the line Λ_{60^o} at 60^o degrees to the gradient $\partial_{\underline{\psi}}\, p_{i+2,y}(\underline{\psi}_{i+1}(y)$, because the velocity of the quasi periodic motion has size of order $O(1)$ (*i.e.* it is $|\underline{\omega}|$) so that a time variation of up to $O(\vartheta)$ suffices for a displacement of $r(y) < \vartheta$ (recall that $r_i' < r_i < \vartheta$, as stipulated at the beginning).

$\bullet_9$ As we vary y we find, by continuity, a point y^* such that $\underline{\psi}_{i+1}(y^*) - \underline{\omega}\, T_{in} = \underline{\psi}_i'$ because $\underline{\psi}_{i+1}(y)$ has a component along $\underline{w}^{\perp}$ which varies by $b_1\vartheta$ at least, see (6.4), *and in the right direction* towards $\underline{\psi}_i'$ by the above proposition $\bullet_5$.

$\bullet_{10}$ Setting $r^* = r(y^*)$ and $\tilde{\underline{\psi}}_{i+1} = \underline{\psi}_{i+1,r^*,y^*}$ we see that the evolution of $Y_{i+1,y^*}(\tilde{\underline{\psi}}_{i+1})$ leads to a point which has $\underline{\psi}$–coordinate close within $\frac{r_i'}{2b_3E}$ to the coordinate $\underline{\psi}_i'$ of the point $X_i' = (\underline{A}_{i+1}, \underline{\psi}_i', \kappa, 0)$ (around which the already inductively known set B_i' is constructed, see (4.3)), because $\tilde{\underline{\psi}}_{i+1}$ is within $r_i'/2b_3E$ of $\underline{\psi}_{i+1}(y)$ by construction. *Note that this is just a continuity statement: hence it is non constructive*, as much as the other continuity arguments used above.

We set $r_{i+1} = r^* > d\frac{r_i'}{8b_3E'}$, $r_{i+1}' = \gamma r_{i+1}^2$ with γ small enough see that the points of B_{i+1} defined by such parameters via (4.2) evolve backward in time to fall inside B_i' at their last interchange time.

However the interchange time T_{in} varies when $\underline{\psi}$ varies in the disk of radius r_{i+1}' centered at a point at distance r_{i+1} from the heteroclinic point. And it is proportional to the logarithm of the inverse of $|p_{i+2}(\underline{\psi})|$; the latter is a function essentially linear in $\underline{\psi}$.[5] Hence it varies bounded by a factor proportional to $\log\frac{r_{i+1}+r_{i+1}'}{r_{i+1}-r_{i+1}'}$.

The latter variation has size $O(\frac{r_{i+1}'}{r_{i+1}})$ which *must be* $< O(r_i')$ *because* we want that the backward evolution of the points in B_{i+1} is, at their interchange time, inside B_i' and the velocity $\underline{\omega}$ of the quasi periodic motion is of $O(1)$. Hence if the time varies by $O(r_i')$ the resulting displacement of the final value of $\underline{\psi}$ will be of $O(r_i')$: recalling that $r_{i+1} \in [\frac{d}{8b_3E'}r_i', \frac{1}{2b_3E}r_i']$ we get a quadratic recursion for the definition of the r_i: unavoidable in the above scheme.

$\bullet_{11}$ We find that, from the above proposition, that $r_i', r_i = O((\Gamma r_0)^{2^i})$ for some Γ (one can take $\Gamma = d\gamma/(8b_3E')$); so that the time $T_i = O(\overline{g}^{-1}\log r_i^{-1})$ necessary to hop one step along the chain is $O(\overline{g}^{-1}2^i)$ and the time T_{drift} for drifting along the chain is bounded above by $O(\overline{g}^{-1}2^{\mathcal{N}})$:

$$T_{drift} \le \overline{g}^{-1}2^{O(\mu^{-1})} \tag{6.9}$$

Recalling that ϑ is fixed, if $\mu = O(\varepsilon)$ (generic) this is $const\, e^{const\ \varepsilon^{-1}}$.

Remark: Therefore the exponential bound (6.9) is due to the rapid convergence to 0 of r_i, *i.e.* with a logarithm exponentially diverging, which arises from the quadratic recursion $r_{i+1}' = O(r_i'^2)$.

---

[5] Because the point $\tilde{\underline{\psi}}_{i+1}$ is still on a line Λ' very close, by the last remark in the preceding proposition, to Λ_{60^o}, at 60^o degrees to the gradient $\partial_{\underline{\psi}}\, p_{i+2,y}(\underline{\psi}_{i+1}(y^*))$.

§7. Concluding remarks. Very fast diffusion?

For a review on diffusion see [L]: in this paper the possibility of estimates of size of an inverse power of ε is proposed and discussed.

(1) The above non-variational proof gives results not directly comparable to the best known, [Be], [Br], based on a variational method and giving (in [Br]) a polynomial drift time of $O(\mu^{-2})$.

The papers [Be],[Br], deal with Arnold's example, [A], *i.e.* with a different case. However they make use in an essential way of the very similar structure of the model, *i.e.* of the fact that it admits a "gap-less" foliation into stable and unstable manifolds of invariant tori, see also [P]. It is hard to see how to improve the bounds of §6 in Arnold's example, if it is studied along the same lines.

The recent works [M], [C], also lead very close to the estimates in [Be],[Br] and, if I understand them correctly, they should apply also to the cases treated here and give polynomial estimates: hence the difference between the size of the bounds obtained by our approach and the ones obtained via variational methods or via geometric methods alternative to the ones exposed here remains (for me) a puzzle that I hope to understand in the future.

(2) It is worth stressing again that the methods of §5 apply every time there are "no gaps" around resonant tori and the homoclinic angles admit a nonzero lower bound: therefore they apply to the case in [A] with, in the notations of [A], $\mu = \varepsilon^c$ and c large enough.

In the isochronous models they apply, immediately, to a variety of cases: a non trivial one is the Hamiltonian (1.1) with $\underline{\omega} = (\eta^a, \eta^{-1/2})$, $a \ge 0$, $\varepsilon = \mu\eta^c$ with c large enough and, *possibly, even a further " monochromatic, strong and rapid"* perturbation $\beta f_0(\varphi, \lambda)$ like $\beta\cos(\lambda+\varphi)$ with $\beta = O(1)$.

Consider only values of η such that $|\underline{\omega}\cdot\underline{\nu}| > C\eta^d|\underline{\nu}|^{-\tau}$ for all $\underline{0} \ne \underline{\nu} \in Z^2$, and for some $C, d > 0$, see §2 in [GGM]. Then *by using the results of* [GGM] (§8) we see that if η is fixed small enough the homoclinic splitting is analytic in β for $|\beta| < O(\eta^{-1/2})$, while it *does not vanish* for β small (*i.e.* $\beta = O(\eta^c)$), generically in f (see [GGM], §6).

Hence it is not 0 for all $\beta < 2$ (say) *except, possibly, for finitely many values of* β. This means that in such *strongly perturbed systems* ($\beta = O(1)$) one still has elastic heteroclinic chains of arbitrary length, see §8 of [GGM], and therefore there is diffusion provable by the methods of §5,§6, except possibly in correspondence of finitely many values of β.

Furthermore the $\underline{A}$–independent (because of isochrony, see [GGM]) homoclinic angles can become large when β, μ approach their convergence radii and this gives us the possibility of "very fast" drift on time scales of $\sim O(1)$. In fact I think that the homoclinic splitting might be a monotonic function of ε, β for interesting classes of perturbations.

(3) An advantage of the technique of §5 is its flexibility which makes it immediately applicable, essentially without change, to anisochronous systems, see [CG].

(4) Constructivity, even partial (see comments in §5), seems the key to understanding the huge difference between the results of §5 and the variational results, or those of §6 above: diffusion time bounds in an inverse power of ε (in [Br] and §6) versus a super-exponential in the more constructive proposal in §5. A hint in this direction is provided by the bound in §6: by adding a new idea to the method of §5, *i.e.* of [CG], one can get a drift time estimate of $2^{-O(\varepsilon^{-1})}$ instead of the super-exponential of [CG], and §5. *But the theory becomes now less constructive*: not even the sequence of close encounters with invariant tori is determined constructively as continuity arguments are used.

(5) The method of §5 and of §6 seems related to the "windowing" analysis in the early work [Ea] and in [M], [C] as pointed out to me by P. Lochak and J. Cresson.

(6) Finally only drift in phase space is discussed here: but it is clear that heteroclinic chains do not need to "advance" at each step (*e.g.* a A–coordinate needs not to increase systematically): we can use heteroclinic chains that advance and back up at our prefixed wish (*e.g.* randomly). In this sense, there is no difference between drift and diffusion.

(7) The bound in theorem 2 has been independently derived by Valdinoci, [V], following the scheme of [CG]. The error in [CG] which led to an estimate better than the one the method could really give, see §5, can be avoided in the isochronous case and leads to a bound equivalent to (6.9).

Acknowledgments: I am indebted to P. Lochak stimulating comments and, in particular, to G. Gentile and V. Mastropietro for many discussions and help in revising the manuscript. This work is part of the research program of the European Network on: "Stability and Universality in Classical Mechanics", # ERBCHRXCT940460.

Appendix A1. Filling times of quasi periodic motions: a conjecture.

Let $(\omega_1,\ldots,\omega_d)=\underline{\omega}\in R^d$ be such that $|\underline{\omega}\cdot\underline{\nu}|^{-1}\le C|\underline{\nu}|^\tau$. Let $\chi(x),\chi_\perp(x)$ be C^∞–functions even and strictly positive for $|x|<\frac12\pi$, vanishing elsewhere and with integral 1. Let $\underline{\psi},\underline{\psi}_0\in T^d$ and $x(\underline{\psi})\,{\buildrel def\over=}\,\varepsilon^{-(d-1)}\chi(\underline{\omega}\cdot(\underline{\psi}-\underline{\psi}_0)/|\underline{\omega}|)\cdot\chi(\varepsilon^{-1}|P^\perp(\underline{\psi}-\underline{\psi}_0)|)$, $P^\perp=$ orthogonal projection on the plane orthogonal to $\underline{\omega}$.

The function x can be naturally regarded as defined and periodic on T^d: if $\hat\chi(\sigma)$ is the Fourier transform of χ as a function on R then the Fourier transform of x is $\hat\chi(\nu^\parallel)\hat\chi(\varepsilon|\underline{\nu}^\perp|)$, $\underline{\nu}$ integer components vector, $\nu^\parallel=\underline{\omega}\cdot\underline{\nu}/|\underline{\omega}|$, $\underline{\nu}^\perp=P^\perp\underline{\nu}$. The average $X=\overline{T}^{-1}\int_0^{\overline{T}}x(\underline{\omega}\,t)\,dt$ is:

$$X=1+\sum_{\underline{\nu}\ne\underline{0}}\hat x(\underline{\nu})e^{-i\underline{\psi}_0\cdot\underline{\nu}}\frac{1}{\overline{T}}\frac{e^{i\underline{\omega}\cdot\underline{\nu}\overline{T}}-1}{i\underline{\omega}\cdot\underline{\nu}}\ge 1-\frac{2C}{\overline{T}}\sum_{\underline{\nu}\ne\underline{0}}\Big|\hat\chi(\nu^\parallel)\hat\chi(\varepsilon|\underline{\nu}^\perp|)\Big|\,|\underline{\nu}|^\tau \qquad (A1.1)$$

Since the last sum is bounded above by $b\varepsilon^{-(\tau+d-1)}$ the average X is positive, *e.g.* $>\frac12$, for all $\underline{\psi}_0$ if $\overline{T}>4bC\varepsilon^{-(\tau+d-1)}$. This means that for $T>4bC\varepsilon^{-(\tau+d-1)}+\pi/|\underline{\omega}|$, hence for $T>BC\varepsilon^{-(\tau+d-1)}$ with B a suitable constant depending only on d, the torus will have been filled by the trajectory of any point within a distance ε. This proof is taken from (5), p. 111, of [G1], see [BGW] for an alternative proof and a much stronger result (*i.e.* with τ replacing $\tau+d-1$).

Of course the above estimate $\overline{T}>O(\varepsilon^{-\tau-(d-1)})$ really deals with a quantity different from the minimum time of visit. It is an estimate of the minimum time beyond which all cylinders with height 1 (say) and basis of radius ε have not only been visited but they have been visited with a frequency that is, for all of them, larger than $\frac12$ of the asymptotic value (equal to ε^{d-1}): we can call the latter time the *first large frequency of visit time.* The difference between the two concepts explains the difference between the two estimates which are equally good, *i.e.* alternative, for the purposes of our analysis (and both too detailed since we only need that the minimum time of visit is finite). And I *conjecture* that both are optimal: the first is optimal as an estimate of the first visit time and the second as an estimate of the first large frequency of visit time.

Appendix A2. Fixing the time of interchange.

The differential condition on $\underline{\psi}$ is $\frac{d}{dy}(p^s_{i+2,y}(\underline{\psi})-p^s_{i+2,y}(\underline{\psi}_{i+1}(y))=0$ (having inserted $p^s_{i+2,y}(\underline{\psi}_{i+1}(y))=0$ for convenience) or, if prime denotes y–differentiation:

$$\begin{aligned}0=&\partial_{\underline{\psi}}p^s_{i+2,y}(\underline{\psi})\cdot(\underline{\psi}'-\underline{\psi}_{i+1}(y)')+(\partial_{\underline{\psi}}p^s_{i+2,y}(\underline{\psi})-\partial_{\underline{\psi}}p^s_{i+2,y}(\underline{\psi}_{i+1}(y)))\cdot\\&\cdot\underline{\psi}_{i+1}(y)'+(\partial_y p^s_{i+2,y})(\underline{\psi})-(\partial_y p^s_{i+2,y})(\underline{\psi}_{i+1}(y))\end{aligned}\qquad (A2.1)$$

which means that $r(y)\stackrel{def}{=}|\underline{\psi}'-\underline{\psi}_{i+1}(y)|$ verifies $|r(y)'|<C_1r(y)$ for some $C_1>0$ independent of μ because:

(1) all derivatives of $p^s_{i+2,y}$ with respect to the arguments $\underline{\psi},y$ are of order μ.
(2) the vector $\underline{\psi}-\underline{\psi}_{i+1}(y)$ has the form $r(y)\,\underline{w}_{60^\circ}(y)$ where $\underline{w}_{60^\circ}(y)$ is the unit vector parallel to the axis forming 60^o degrees with $\partial_{\underline{\psi}}p^s_{i+2,y}(\underline{\psi}_{i+1}(y))$, so that $(\underline{\psi}-\underline{\psi}_{i+1}(y))'=r'(y)\,\underline{w}_{60^\circ}(y)+r(y)\,\underline{w}'_{60^\circ}(y)$
(3) $\underline{\psi}_{i+1}(y)'$ has size $O(1)$; the second derivatives of $p^s_{i+1,y}(\underline{\psi})$ with respect to $\underline{\psi}$ have size $O(\mu)$ and the derivative of $\underline{w}_{60^\circ}(y)$ that can be computed by differentiating its expression (namely $\partial_{\underline{\psi}}p^s_{i+2}(\underline{\psi}_{i+1}(y))/|\partial_{\underline{\psi}}p^s_{i+2}(\underline{\psi}_{i+1}(y))|$) is of $O(1)$.

This fixes the y–derivatives of $r(y)\stackrel{def}{=}|\underline{\psi}-\underline{\psi}_{i+1}(y)|$ to have size $O(r(y))$ so that the variation of $r(y)$, as y varies in its interval of size $\frac12\mu\vartheta$ and starting at $\overline{y}=y_{i+1}+\frac12\mu\vartheta$, is bounded by $(e^{C_1\frac12\mu\vartheta}-1)\le C_1\mu\vartheta r(\overline{y})$ if $\vartheta C_1<\frac12$. Hence $\frac{d}{8b_3E'}r'_i<r(y)<\frac{1}{2b_3E'}r'_i$ and the point $Y_{i+1,y}(\underline{\psi})$ remains in the neighborhood where the local coordinates are defined.

References

[A] Arnold, V.: *Instability of dynamical systems with several degrees of freedom*, Sov. Mathematical Dokl., 5, 581-585, 1966.

[Be] Bessi, U.: it An approach to Arnold's diffusion through the Calculus of Variations, Nonlinear Analysis, 1995.

[Br] Bernard, P.: *Perturbation d'un hamiltonien partiellement hyperbolique*, C.R. Academie des Sciences de Paris, **323**, I, 189–194, 1996.

[BGW] Bourgain, J. Golse, F., Wennberg, S.: *The ergodisation time for linear flows on tori: application to kinetic theory*, to appear in Communications in Mathematical Physics, preprint 1995.

[C] Cresson, J.: *Symbolic dynamics for homoclinic partially hyperbolic tori and "Arnold diffusion"*, Preprint of Institut de mathematiques de Jussieux, 1997. And, mainly: *Propriétés d'instabilité des systèmes Hamiltoniens proches de systèmes intégrables*, Doctoral dissertation, L' Observatoire de Paris, Paris, 1997.

[CG] Chierchia, L., Gallavotti, G.: *Drift and diffusion in phase space*, Annales de l' Institut Poincarè, B, **60**, 1–144, 1994.

[DGJS] Delshams, S., Gelfreich, V.G., Jorba, A., Seara, T.M.: *Exponentially small splitting of separatrices under fast quasiperiodic forcing*, Communications in Mathematical Physics, **189**, 35–72, 1997.

[Ea] Easton, R.W.: *Orbit structure near trajectories biasymptotic to invariant tori*, in *Classical mechanics and dynamical systems*, ed. R. Devaney, Z. Nitecki, p. 55–67, Dekker, 1981.

[E] Eliasson, L.H.: *Absolutely convergent series expansions for quasi-periodic motions*, Mathematical Physics Electronic Journal, MPEJ, **2**, 1996.

[G1] Gallavotti, G.: The elements of mechanics, Springer, 1983.

[G2] Gallavotti, G.: *Twistless KAM tori*, Communications in Mathematical Physics **164**, 145–156, (1994).

[G3] Gallavotti, G.: *Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review*, Reviews on Mathematical Physics **6**, 343– 411, 1994.

[G4] Gallavotti, G.: *Hamilton–Jacobi's equation and Arnold's diffusion near invariant tori in a priori unstable isochronous systems*, Rendiconti del seminario matematico di Torino, in print; also in chao-dyn@lanl.gov #9710019.

[Ge] Gentile, G.: *A proof of existence of whiskered tori with quasi flat homoclinic intersections in a class of almost integrable systems*, Forum Mathematicum, **7**, 709–753, 1995. See also: *Whiskered tori with prefixed frequencies and Lyapunov spectrum*, Dynamics and Stability of Systems, **10**, 269–308, 1995.

[GG] G. Gallavotti, G. Gentile: *Majorant series convergence for twistless KAM tori*, Ergodic theory and dynamical systems, **15**, 857–869, 1995.

[GGM] G. Gallavotti, G. Gentile, V. Mastropietro: *Pendulum: separatrix splitting*, preprint, chao-dyn # 9709004: this paper will appear with a different, more informative, title "Separatrix splitting for systems with three time scales". And G. Gallavotti, G. Gentile, V. Mastropietro: *Melnikov's approximation dominance. Some examples.*, chao-dyn # 9804043, in print in Reviews in Mathematical Physics.

[L] Lochak, P.: *Arnold's diffusion: a compendium of remarks and questions*, Proceedings of 3DHAM, s'Agaro, 1995, in print.

[Ea] Easton, R.W.: *Orbit structure near trajectories biasymptotic to invariant tori*, in *Classical mechanics and dynamical systems*, ed. R. Devaney, Z. Nitecki, p. 55–67, Dekker, 1981.

[M] Marco, J.P.: *Transitions le long des chaines de tores invariants pour les systèmes hamiltoniens analytiques*, Annales de l' Institut Poincaré, **64**, 205–252, 1995.

[P] Perfetti, P.: *Fixed point theorems in the Arnol'd model about instability of the action–variables in phase space*, mp_arc@math.utexas.edu, #97-478, 1997, in print in Discrete and continuous dynamical systems.

[T] Thirring, W.: *Course in Mathematical Physics*, vol. 1, p. 133, Springer, Wien, 1983.

[V] Valdinoci, E.: private communication.

Internet: (Author's preprints): `http://ipparco.roma1.infn.it`
e-mail: `giovanni@ipparco.roma1.infn.it`