A note on micro-instability for Hamiltonian systems close to integrable

Abed Bounemoura * and Vadim Kaloshin †

January 12, 2015

Abstract

In this note, we consider the dynamics associated to a perturbation of an integrable Hamiltonian system in action-angle coordinates in any number of degrees of freedom and we prove the following result of "micro-diffusion": under generic assumptions on h and f, there exists an orbit of the system for which the drift of its action variables is at least of order $\sqrt{\varepsilon}$, after a time of order $\sqrt{\varepsilon}^{-1}$. The assumptions, which are essentially minimal, are that there exists a resonant point for h and that the corresponding averaged perturbation is non-constant. The conclusions, although very weak when compared to usual instability phenomena, are also essentially optimal within this setting.

1 Introduction and result

1.1 Introduction

Let $n \geq 2$ be an integer, $B = B_1 \subseteq \mathbb{R}^n$ be the unit open ball with respect to the supremum norm $|\,.\,|$ and $\mathbb{T}^n := \mathbb{R}^n/\mathbb{Z}^n$. Consider a smooth (at least C^2) Hamiltonian function H defined on the domain $\mathbb{T}^n \times B$ of the form

$$H(\theta, I) = h(I) + \varepsilon f(\theta, I), \quad \varepsilon \geq 0, \quad (\theta, I) \in \mathbb{T}^n \times B, \tag{H}$$

and its associated Hamiltonian system

$$\begin{cases} \dot{\theta}(t) = \partial_I H(\theta(t), I(t)) = \partial_I h(I(t)) + \varepsilon \partial_I f(\theta(t), I(t)), \\ \dot{I}(t) = -\partial_\theta H(\theta(t), I(t)) = -\varepsilon \partial_\theta f(\theta(t), I(t)). \end{cases}$$

For $\varepsilon = 0$, the system is stable in the sense that the action variables $I(t)$ of all solutions are constant, and these solutions are quasi-periodic. Now for $\varepsilon \neq 0$ but sufficiently small, the celebrated KAM theorem ([Kol54], [Arn63], [Mos62]) and Nekhoroshev theorem ([Nek77], [Nek79]) assert that the system, provided it is real-analytic, retains some stability properties: for a generic h and all f, "most" solutions are quasi-periodic and the action variables of all solutions are almost constant for a very long interval of time.

Yet in the same setting, Arnold conjectured in the sixties that for a generic h and for $n \geq 3$, the following phenomenon of instability should occur: "for any points I' and I'' on

*CNRS - CEREMADE, Université Paris Dauphine & IMCCE, Observatoire de Paris (abedbou@gmail.com)
†University of Maryland at College Park (vadim.kaloshin@gmail.com)

the connected level hypersurface of h in the action space there exist orbits connecting an arbitrary small neighborhood of the torus $I = I'$ with an arbitrary small neighborhood of the torus $I = I''$, provided that ε is sufficiently small and that f is generic" (see [Arn94]).

Since Arnold's original example of such a phenomenon ([Arn64]), this question has been investigated extensively, but only recently solutions to this conjecture have appeared for convex h (see [KZ12], [Che13] for $n = 3$ and [KZ14b] for $n = 4$ and a progress for any $n > 4$ in [BKZ11], [KZ14a]). For non-convex integrable Hamiltonians that posses a "super-conductivity channel" (that is, a rational subspace contained in an energy level), it is very simple to construct examples of perturbation having unstable solutions (see [Mos60], [Nek79]), and this is also true for a generic perturbation for $n = 2$ ([BK14]). Apart from these two classes of integrable Hamiltonians (the convex ones and the ones that posses a super-conductivity channel), nothing is known, even for a specific perturbation.

It is our purpose here to show, using the method of [BK14], that for a generic integrable Hamiltonian (this generic condition being the existence of a resonant point) and for a generic perturbation (the associated averaged perturbation is non-constant), one has a phenomenon of "micro-instability": existence of a solution whose action variables drift of order $\sqrt{\varepsilon}$ after a time of order $\sqrt{\varepsilon}^{-1}$.

1.2 Result

Let H be as in (H), we assume it is of class C^3 and

$$|h|_{C^2(B)} \leq 1, \quad |f|_{C^3(\mathbb{T}^n \times B)} \leq 1 \tag{1}$$

where $|\,.\,|_{C^2(B)}$ (respectively $|\,.\,|_{C^3(\mathbb{T}^n\times B)}$) denotes the standard C^2-norm for functions defined on B (respectively the standard C^3-norm for functions defined on $\mathbb{T}^n \times B$). Our first general assumption is on the integrable Hamiltonian h.

$(A.1)$. There exists $I^* \in B$ such that $\omega := \partial_I h(I^*)$ is resonant but non-zero, that is $k \cdot \omega = 0$ for some (but not all) $k \in \mathbb{Z}^n \setminus \{0\}$, where $\cdot$ denotes the Euclidean scalar product.

We denote by Λ the real subspace of $\mathbb{R}^n$ spanned by the $\mathbb{Z}$-module $\{k \in \mathbb{Z}^n \mid k \cdot \omega = 0\}$. By assumption, the dimension of Λ is at least 1 and at most $n-1$. Without loss of generality, we will assume that $I^* = 0$. For our second assumption, we define

$$f_\omega(\theta, I) := \lim_{t\to+\infty} \frac{1}{t}\int_0^t f(\theta + s\omega, I)ds, \quad f_\omega^*(\theta) := f_\omega(\theta, I^*) = f_\omega(\theta, 0).$$

Expanding f in Fourier series, $f(\theta, I) = \sum_{k\in\mathbb{Z}^n} f_k(I)e^{i2\pi k\cdot\theta}$, we have a more explicit expression

$$f_\omega^*(\theta) = \sum_{k\in\mathbb{Z}^n\cap\Lambda} f_k(I^*)e^{i2\pi k\cdot\theta} = \sum_{k\in\mathbb{Z}^n\cap\Lambda} f_k(0)e^{i2\pi k\cdot\theta}.$$

Our second general assumption is as follows.

$(A.2)$ The function f_ω^* is non-constant, that is there exists $\theta^* \in \mathbb{T}^n$ such that $|\partial_\theta f_\omega^*(\theta^*)| = \lambda > 0$.

From (1) we necessarily have $\lambda \leq 1$. In order to state precisely our theorem, we need further definitions. Let $\Lambda^\perp$ be the orthogonal complement of Λ (observe that $\Lambda^\perp$ is nothing but the minimal rational subspace of $\mathbb{R}^n$ containing ω). Let us define $\Psi = \Psi_\omega$ by

$$\Psi(Q) = \max\left\{|k \cdot \omega|^{-1} \mid k \in \Lambda^\perp \cap \mathbb{Z}^n,\ 0 < |k| \leq Q\right\}. \tag{2}$$

This is well-defined for $Q \geq Q_\omega$, where $Q_\omega \geq 1$ is a constant depending on ω (see [BF13]). Then for $x \geq Q_\omega \Psi(Q_\omega)$, we define $\Delta = \Delta_\omega$ by

$$\Delta(x) = \sup\{Q \geq Q_\omega \mid Q\Psi(Q) \leq x\}. \tag{3}$$

We can finally state our theorem.

Theorem 1. *Let H be as in* (H) *satisfying* (1)*, and assume that* $(A.1)$ *and* $(A.2)$ *holds true. There exist positive constants $\kappa = \kappa(n, |\omega|, \Lambda)$, $\mu_0 = \mu_0(n, |\omega|, \Lambda, \lambda)$, $c = c(\lambda, \Lambda)$ and $\delta = \delta(\lambda, \Lambda)$ such that if*

$$0 < \mu(\sqrt{\varepsilon}) := \left(\Delta\left(\kappa\sqrt{\varepsilon}^{-1}\right)\right)^{-1} \leq \mu_0,$$

then there exists a solution $(\theta(t), I(t))$ of the system (H) *such that*

$$|I(\tau) - I(0)| \geq c\sqrt{\varepsilon}, \quad \tau := \delta/\sqrt{\varepsilon}.$$

Moreover, there exists a positive constant $C = C(n, |\omega|, \Lambda)$ such that this solution satisfies

$$d(I(0), I^*) \leq C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}), \quad d(I(t) - I(0), \Lambda) \leq C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}), \quad 0 \leq t \leq \tau$$

where d is the distance induced by the supremum norm.

Observe that $\mu(\sqrt{\varepsilon})$ always converge to zero as ε goes to zero, more slowly than $\sqrt{\varepsilon}$: for instance, if ω is periodic (a multiple of a rational vector), then $\mu(\sqrt{\varepsilon})$ is exactly of order $\sqrt{\varepsilon}$, and if ω is resonant-Diophantine (meaning that it is not rational but the function Ψ defined above grows at most as a power), then $\mu(\sqrt{\varepsilon})$ is of order a power of $\sqrt{\varepsilon}$. In general, the speed of convergence to zero can be arbitrarily slow. Yet the quantity $\sqrt{\varepsilon}\mu(\sqrt{\varepsilon})$ is always smaller than $\sqrt{\varepsilon}$, and so the statement implies that the $\sqrt{\varepsilon}$-drift occurs along the resonant direction Λ, as in the transverse direction the variation of the action is of order $\sqrt{\varepsilon}\mu(\sqrt{\varepsilon})$ during the interval of time considered.

1.3 Some comments

Let us now briefly discuss the assumptions and conclusions of Theorem 1.

First, if the assumption $(A.1)$ is not satisfied, that is if the image of the gradient map $\partial_I h$ does not contain a resonant point (which means that this image is contained in a non-resonant line), it is not hard to see that the conclusions of the theorem do not hold true: for all solutions and for all $0 \leq t \leq \tau$, the variation of the action variables cannot be of order $\sqrt{\varepsilon}$ (or put it differently, in order to have a drift of order $\sqrt{\varepsilon}$ one needs a time strictly larger than τ). Indeed, one can prove in this case (using normal form techniques) that the system can be (globally) conjugated to another system which consists of an integrable part plus a perturbation whose size is of order $\varepsilon\mu(\sqrt{\varepsilon})$: this implies that for times $0 \leq t \leq \tau$, the variation of the action variables of all solutions is of order at most $\sqrt{\varepsilon}\mu(\sqrt{\varepsilon})$. Then, if the assumption

$(A.2)$ is not satisfied, it is also easy to see that the conclusions of the theorem do not hold true for any solution starting close to $I^* = 0$: indeed, looking at Lemma 2 below, one would get a (local, defined around $I^* = 0$) conjugacy to a perturbation of an integrable system, with a perturbation whose size is again of order $\varepsilon\mu(\sqrt{\varepsilon})$.

Concerning the conclusions, it is also plain to remark that at the time τ the variation of the action variables cannot be larger than $\sqrt{\varepsilon}$, up to a constant. But more is true in the special case where h is convex (or quasi-convex) and Λ is a hyperplane (which is equivalent to ω being a periodic vector): for a time T which is very large (any fixed power of $\sqrt{\varepsilon}^{-1}$ if H is smooth of even $\exp\left(\sqrt{\varepsilon}^{-1}\right)$ if H is real-analytic), the variation of the action variables of the solution given by Theorem 1 is of order $\sqrt{\varepsilon}$ for times $0 \leq t \leq T$ (see [Loc92] for the analytic case and [Bou10] for the smooth case). In this situation, one has the curious fact that the variation of the action variables is exactly of order $\sqrt{\varepsilon}$ (in the sense that it is bigger than some small constant times $\sqrt{\varepsilon}$ and smaller than some large constant times $\sqrt{\varepsilon}$) during the very long interval of time $\tau \leq t \leq T$.

2 Proof of the result

The proof of Theorem 1 follows the strategy of [BK14]. On a $\sqrt{\varepsilon}$-neighborhood of the point I^*, we will conjugate our Hamiltonian to a simpler Hamiltonian (a resonant normal form plus a small remainder) for which the result will be obtained by simply looking at the equations of motion. Using the fact that the conjugacy is given by a symplectic transformation which is close to identity, the result for our original Hamiltonian will follow. The normal form will be stated in §2.1, and the proof of Theorem 1 will be given in §2.2.

2.1 A normal form lemma

Before starting the proof, it will be more convenient to assume that the subspace Λ is generated by the first d vectors of the canonical basis of $\mathbb{R}^n$, for $1 \leq d \leq n-1$. This is no restriction, as by a linear symplectic change of coordinates one can always write $\omega = (0, \tilde{\omega}) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ for some non-resonant vector $\tilde{\omega} \in \mathbb{R}^{n-d}$. This enables us to get rid of the dependence on Λ in the constants involved. Since $\Lambda^\perp \cap \mathbb{Z}^n = \{0\} \times \mathbb{Z}^{n-d}$, the function Ψ defined in (2) takes the simpler form

$$\Psi(Q) = \max\left\{ |k \cdot \tilde{\omega}|^{-1} \mid k \in \mathbb{Z}^{n-d},\ 0 < |k| \leq Q \right\}$$

and is well-defined for $Q \geq 1$ (that is one can take $Q_\omega = 1$). The function Δ introduced in (2) is then defined for $x \geq \Psi(1) = |\tilde{\omega}|^{-1}$ and we have

$$\Delta(x) = \sup\{Q \geq 1 \mid Q\Psi(Q) \leq x\}.$$

Observe also that in this situation, we simply have

$$f_\omega(\theta, I) = f_\omega(\theta_1, \ldots, \theta_d, I) = \int_{\mathbb{T}^{n-d}} f(\theta_1, \ldots, \theta_d, \theta_{d+1}, \ldots, \theta_n, I) d\theta_{d+1} \ldots d\theta_n$$

and hence

$$f_\omega^*(\theta_1, \ldots, \theta_d) = f_\omega(\theta_1, \ldots, \theta_d, I^*) = f_\omega(\theta_1, \ldots, \theta_d, 0).$$

The assumption $(A.2)$ thus reduces to the existence of a point $\theta^* = (\theta_1^*, \ldots, \theta_d^*) \in \mathbb{T}^d$ and a constant $0 < \lambda \leq 1$ such that $|\partial_\theta f_\omega^*(\theta^*)| = \lambda$. Here's the statement of our normal form lemma.

Lemma 2. *Let H be as in* (H) *satisfying* (1). *There exist positive constants $\kappa = \kappa(n, |\omega|)$, $\mu_0 = \mu_0(n, |\omega|)$ and $C = C(n, |\omega|)$ such that if*

$$\mu(\sqrt{\varepsilon}) := \Delta\left(\kappa\sqrt{\varepsilon}^{-1}\right)^{-1} \leq \mu_0, \tag{4}$$

then there exists a symplectic map $\Phi : \mathbb{T}^n \times B_{2\sqrt{\varepsilon}} \to \mathbb{T}^n \times B_{3\sqrt{\varepsilon}}$ of class C^2 such that

$$H \circ \Phi(\theta, I) = h(I) + \varepsilon f_\omega(\theta_1, \dots, \theta_d, I) + \tilde{f}_\varepsilon(\theta, I)$$

with the estimates

$$|\Pi_I \Phi - \mathrm{Id}|_{C^0(\mathbb{T}^n \times B_{2\sqrt{\varepsilon}})} \leq C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}), \tag{5}$$

$$|\partial_\theta \tilde{f}_\varepsilon|_{C^0(\mathbb{T}^n \times B_{2\sqrt{\varepsilon}})} \leq C\varepsilon\mu(\sqrt{\varepsilon}), \tag{6}$$

$$|\partial_I \tilde{f}_\varepsilon|_{C^0(\mathbb{T}^n \times B_{2\sqrt{\varepsilon}})} \leq C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}). \tag{7}$$

This is a very special case of Theorem 1.3 of [Bou13], to which we refer for a proof (strictly speaking, Theorem 1.3 of [Bou13] would require in our situation the integrable Hamiltonian h to be of class C^5, but one can see from the proof that C^2 is in fact sufficient).

2.2 Proof of Theorem 1

Theorem 1 will be easily deduced from Lemma 2.

Proof of Theorem 1. We start by choosing $\varepsilon > 0$ sufficiently small so that (4) holds true. Then we apply Lemma 2: there exists a symplectic map $\Phi : \mathbb{T}^n \times B_{2\sqrt{\varepsilon}} \to \mathbb{T}^n \times B_{3\sqrt{\varepsilon}}$ of class C^2 such that

$$H \circ \Phi(\theta, I) = h(I) + \varepsilon f_\omega(\theta_1, \dots, \theta_d, I) + \tilde{f}_\varepsilon(\theta, I)$$

with the estimates (5), (6) and (7). Obviously f_ω has unit C^3-norm since this is the case for f. Consider the solution $(\theta(t), I(t))$ of the Hamiltonian $H \circ \Phi$, starting at $I(0) = I^* = 0$, $(\theta_1(0), \dots, \theta_d(0)) = \theta^* \in \mathbb{T}^d$ and with $(\theta_{d+1}(0), \dots, \theta_n(0)) \in \mathbb{T}^{n-d}$ arbitrary. It satisfies the following equations:

$$\begin{cases} \dot{I}(t) = -\varepsilon\partial_\theta f_\omega(\theta_1(t), \dots, \theta_d(t), I(t)) - \partial_\theta \tilde{f}_\varepsilon(\theta(t), I(t)), \\ \dot{\theta}(t) = \partial_I h(I(t)) + \varepsilon\partial_I f_\omega(\theta_1(t), \dots, \theta_d(t), I(t)) + \partial_I \tilde{f}_\varepsilon(\theta(t), I(t)). \end{cases} \tag{8}$$

Let us fix $\delta := \sqrt{\lambda/6}$ and let $\tau = \delta/\sqrt{\varepsilon}$. From the first equation of (8) and the estimate (6), one has

$$|I(t) - I(0)| \leq \delta\sqrt{\varepsilon} + \delta C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}) \leq \delta\sqrt{\varepsilon} + \delta\sqrt{\varepsilon} = 2\delta\sqrt{\varepsilon}, \quad 0 \leq t \leq \tau, \tag{9}$$

up to taking μ_0 smaller than C^{-1}. Using the fact that $\partial_I h(I(0)) = \partial_I h(0) = (0, \tilde{\omega}) \in \mathbb{R}^d \times \mathbb{R}^{n-d}$ which follows from our first assumption and the choice of $I(0)$, this last estimate, together with the fact that h has unit C^2-norm, imply that

$$\max_{1 \leq i \leq d} |\partial_{I_i} h(I(t))| = \max_{1 \leq i \leq d} |\partial_{I_i} h(I(t)) - \partial_{I_i} h(I(0))| \leq 2\delta\sqrt{\varepsilon}, \quad 0 \leq t \leq \tau.$$

From the second equation of (8) and the estimate (7), we obtain from the last estimate

$$\max_{1 \leq i \leq d} |\dot{\theta}_i(t)| \leq 2\delta\sqrt{\varepsilon} + \varepsilon + C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}), \quad 0 \leq t \leq \tau.$$

Taking μ_0 small enough with respect to C and λ (and hence δ), the sum of the last two terms of the right-hand side of the inequality above can me made smaller than $\delta\sqrt{\varepsilon}$, thus we can ensure that

$$\max_{1\leq i\leq d} |\dot{\theta}_i(t)| \leq 3\delta\sqrt{\varepsilon}, \quad 0 \leq t \leq \tau$$

and hence

$$\max_{1\leq i\leq d} |\theta_i(t) - \theta_i(0)| \leq 3\delta^2, \quad 0 \leq t \leq \tau.$$

Now from our second assumption and the choice of the initial condition, we have

$$|\varepsilon\partial_\theta f_\omega(\theta_1(0), \dots, \theta_d(0), I(0))| = |\varepsilon\partial_\theta f_\omega^*(\theta_1^*, \dots, \theta_d^*)| = \varepsilon\lambda > 0$$

so, using the last estimate and our choice of δ, we get

$$|\varepsilon\partial_\theta f_\omega(\theta_1(t), \dots, \theta_d(t), I(0))| \geq \varepsilon\lambda - 3\varepsilon\delta^2 = \varepsilon\lambda - \varepsilon\lambda/2 = \varepsilon\lambda/2, \quad 0 \leq t \leq \tau.$$

From (9), taking μ_0 smaller with respect to λ, we can make sure that

$$|\varepsilon\partial_\theta f_\omega(\theta_1(t), \dots, \theta_d(t), I(t))| \geq \varepsilon\lambda/3, \quad 0 \leq t \leq \tau$$

but also, from the estimate (6),

$$|\varepsilon\partial_\theta f_\omega(\theta_1(t), \dots, \theta_d(t), I(t)) + \partial_\theta \tilde{f}_\varepsilon(\theta(t), I(t))| \geq \varepsilon\lambda/4, \quad 0 \leq t \leq \tau.$$

From the first equation of (8) we finally have

$$|I(\tau) - I(0)| \geq \max_{1\leq i\leq d} |I_i(\tau) - I_i(0)| \geq \sqrt{\varepsilon}\lambda\delta/4 = \sqrt{\varepsilon}3\delta^3/2$$

but also, using the estimate (6),

$$\max_{d+1\leq j\leq n} |I_j(t) - I_j(0)| \leq C\delta\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}) \leq C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}), \quad 0 \leq t \leq \tau.$$

To conclude, using the estimate (5), this solution for $H \circ \Phi$ gives rise to a solution for H that, abusing notations, we still denote $(\theta(t), I(t))$ and such that, taking once again μ_0 smaller with respect to λ, satisfies

$$|I(\tau) - I(0)| \geq \sqrt{\varepsilon}\delta^3 := c\sqrt{\varepsilon}$$

and also, up to enlarging the constant C,

$$d(I(0), I^*) \leq C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}), \quad d(I(t) - I(0), \mathbb{R}^d \times \{0\}) \leq C\sqrt{\varepsilon}\mu(\sqrt{\varepsilon}).$$

This concludes the proof. □

Acknowledgements. The first author would like to thank IMPA for its hospitality. The second author acknowledges partial support of the NSF grant DMS-1402164.

References

[Arn63] V.I. Arnol'd, *Proof of a theorem of A.N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations*, Russ. Math. Surv. **18** (1963), no. 5, 9–36.

[Arn64] ______, *Instability of dynamical systems with several degrees of freedom*, Sov. Math. Doklady **5** (1964), 581–585.

[Arn94] ______, *Mathematical problems in classical physics*, Trends and perspectives in applied mathematics, Appl. Math. Sci., vol. 100, Springer, New York, 1994, pp. 1–20.

[BF13] A. Bounemoura and S. Fischler, *A diophantine duality applied to the KAM and Nekhoroshev theorems*, Math. Z. **275** (2013), no. 3, 1135–1167.

[BK14] A. Bounemoura and V. Kaloshin, *Generic fast diffusion for a class of non-convex Hamiltonians with two degrees of freedom*, Moscow Math. Journal **14** (2014), no. 2, 181–203.

[BKZ11] P. Bernard, V. Kaloshin, and K. Zhang, *Arnold diffusion in arbitrary degrees of freedom and crumpled 3-dimensional normally hyperbolic invariant cylinders*, Acta Mathematica (2011), conditionally accepted, arXiv:1112.2773.

[Bou10] A. Bounemoura, *Nekhoroshev theory for finitely differentiable quasi-convex Hamiltonians*, Journal of Differential Equations **249** (2010), no. 11, 2905–2920.

[Bou13] ______, *Normal forms, stability and splitting of invariant manifolds II. Finitely differentiable Hamiltonians*, Regul. Chaotic Dyn. **18** (2013), no. 3, 261–276.

[Che13] C.-Q. Cheng, *Arnold diffusion in nearly integrable hamiltonian systems*, Preprint (2013), arXiv:1207.4016v2.

[Kol54] A.N. Kolmogorov, *On the preservation of conditionally periodic motions for a small change in Hamilton's function*, Dokl. Akad. Nauk. SSSR **98** (1954), 527–530.

[KZ12] V. Kaloshin and K. Zhang, *A strong form of Arnold diffusion for two and a half degrees of freedom*, preprint (2012), arXiv:1212.1150.

[KZ14a] ______, *Partial averaging and dynamics of the dominant hamiltonian, with applications to arnold diffusion*, preprint (2014), arXiv:1410.1844.

[KZ14b] ______, *A strong form of arnold diffusion for three and a half degrees of freedom*, preprint (2014), http://www2.math.umd.edu/ vkaloshi/papers/announce-three-and-half.pdf.

[Loc92] P. Lochak, *Canonical perturbation theory via simultaneous approximation*, Russ. Math. Surv. **47** (1992), no. 6, 57–133.

[Mos60] J. Moser, *On the elimination of the irrationality condition and Birkhoff's concept of complete stability*, Bol. Soc. Mat. Mexicana (2) **5** (1960), 167–175.

[Mos62] ______, *On Invariant curves of Area-Preserving Mappings of an Annulus*, Nachr. Akad. Wiss. Göttingen **II** (1962), 1–20.

[Nek77] N.N. Nekhoroshev, *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems*, Russian Math. Surveys **32** (1977), no. 6, 1–65.

[Nek79] ______, *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems II*, Trudy Sem. Petrovs **5** (1979), 5–50.