

# The KAM theorem through Dirichlet's box and Khintchine's transference principles

Abed Bounemoura \*

March 14, 2013

## Abstract

In this paper, we give a new proof of the classical KAM theorem which avoids small divisors and relies on two basic principles of Diophantine approximation: Dirichlet's box and Khintchine transference principles.

## 1 Introduction and statement of the result

### 1.1 Introduction

**1.** KAM theory is part of the theory of perturbation of quasi-periodic motions in dynamical systems, which was initiated by Kolmogorov ([Kol54]), Arnold ([Arn63a],[Arn63b]) and Moser ([Mos62]) in the context of Hamiltonian systems. It is often considered “as one of the most important achievements in the qualitative theory of ordinary differential equations in the whole second half of the twentieth century”, and has been the subject of extensive investigations (see [Sev03] and references therein).

**2.** A model problem, already considered by the founders of the theory ([Kol53], [Arn61] and [Mos66]), deals with vector fields on the torus and can be explained as follows. Let  $\alpha$  be a non-zero vector in  $\mathbb{R}^n$ , and  $X_\alpha = \alpha$  the constant vector field on the  $n$ -dimensional torus  $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$  equals to  $\alpha$ . The integral curves of  $X_\alpha$  describe, by definition, quasi-periodic motions with frequency  $\alpha$ , and the problem is to understand the fate of these solutions when a small but arbitrary perturbation  $P$  is added to  $X_\alpha$ . In general, the vector field  $X_\alpha$  is not structurally stable, as the perturbation will induce a shift of frequency so that the vector field  $X_\alpha + P$  cannot be conjugated to  $X_\alpha$ . However, if the frequency  $\alpha$  satisfies a Diophantine condition and  $P$  is sufficiently regular, and if moreover we are allowed to first shift the frequency according to the perturbation, then the resulting vector field can indeed be conjugated to  $X_\alpha$ : this is the content of the classical KAM theorem, first proved by Arnold in [Arn61]) (we refer to Theorem 1.1, §1.2, for a precise statement).

**3.** In the classical approach to the KAM theorem, and more generally in the classical theory of perturbation of quasi-periodic motions, a central role is played by the equation

$$[V, X_\alpha] = P - [P]_\alpha, \quad [P]_\alpha = \lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t (X_\alpha^s)^* P ds \quad (1)$$

---

\* abedbou@gmail.com, IHES, 35 route de Chartres, 91440 Bures-sur-Yvette (abed@ihes.fr)

where  $V$  is the unknown vector field and  $[P]_\alpha$  is the time-average of  $P$  along the flow  $X_\alpha^s$  of  $X_\alpha$  ( $(X_\alpha^s)^*P$  is the pull-back of  $P$  by  $X_\alpha^s$ ). Assuming that  $\alpha$  is non-resonant, that is if the components of  $\alpha$  are independent over the field of rational numbers  $\mathbb{Q}$ , it follows from Birkhoff's ergodic theorem that  $[P]_\alpha$  coincides with the space average, that is

$$[P]_\alpha = [P] = \int_{\mathbb{T}^n} P$$

where the integration on  $\mathbb{T}^n$  is with respect to its Haar measure. Equation (1), which is usually called the homological equation, can be seen as the linearized version of the conjugacy equation we need to solve, and the time-one map  $\Phi = V^1$  of the vector field  $V$  is then used as a building block to construct the sought conjugacy by an iterative scheme. It is precisely in trying to solve the equation (1) that small divisors arise. Geometrically, one needs to integrate along the integral curves of  $X_\alpha$ , and these curves are not closed (they densely fill the torus  $\mathbb{T}^n$ ). Analytically, one needs to invert the operator  $[\cdot, X_\alpha]$  acting on the space of smooth vector fields, and this operator is unbounded. Indeed, this operator can be diagonalized in a Fourier basis: letting  $e_k(\theta) = e^{ik \cdot \theta}$  for  $k \in \mathbb{Z}^n$  and  $V = (V_1, \dots, V_n)$ ,  $P = (P_1, \dots, P_n)$ , if  $V_j = \sum_{k \in \mathbb{Z}^n} V_{j,k} e_k$  and  $P_j = \sum_{k \in \mathbb{Z}^n} P_{j,k} e_k$  for  $1 \leq j \leq n$ , then the solution is given by

$$V_{j,0} = 0, \quad V_{j,k} = (ik \cdot \alpha)^{-1} P_{j,k}, \quad k \in \mathbb{Z}^n \setminus \{0\},$$

and so the absolute values  $|k \cdot \alpha|^{-1}$  of its eigenvalues can be arbitrarily large if the supremum norm  $|k|$  is arbitrarily large. It is then useful to quantify how the quantities  $|k \cdot \alpha|^{-1}$  grow with  $|k|$  by introducing the function

$$\Psi_\alpha(Q) = \max \{|k \cdot \alpha|^{-1} \mid k \in \mathbb{Z}^n, 0 < |k| \leq Q\}, \quad Q \geq 1. \quad (2)$$

A vector  $\alpha$  is said to satisfy a Diophantine condition if the function  $\Psi_\alpha(Q)$  grows at most as a power of  $Q$  (see §1.2 for a more precise definition). To overcome the effect of small divisors, for a given parameter  $Q \geq 1$  a classical approach is to approximate  $P$  by a trigonometric vector field  $P_Q$  and to solve the approximate equation

$$[V, X_\alpha] = P_Q - [P_Q] = P_Q - [P] \quad (3)$$

using Fourier analysis as above: the norm of the vector field  $V$ , and hence the transformation  $\Phi = V^1$ , can be essentially controlled in terms of  $\Psi(Q)$ , and the term  $P - P_Q$  is simply considered as an "error" that causes no trouble in the iteration scheme. We refer to the very nice survey [Pös01] for a detailed exposition of the classical approach in the Hamiltonian setting, and to [Pös11] for a variant of the classical approach (proposed by Rüssmann in [Rüs10], using a somehow "optimal" approximation  $P_Q$ ) in our setting.

**4.** Recently, in a joint work with Fischler ([BF12]), we proposed a fundamentally different approach to the classical theory of perturbation of quasi-periodic motions, replacing the small divisors problem by a method of periodic approximations. To explain this method, let us assume without loss of generality that  $\alpha = (1, \alpha_1, \dots, \alpha_{n-1})$  with  $|\alpha_j| \leq 1$  for  $1 \leq j \leq n$  (this can always be achieved, dividing  $\alpha$  by its supremum norm  $|\alpha|$  and re-ordering its components if necessary); such a simplification will enable us to deal with rational approximations only. So let  $\omega \in \mathbb{Q}^n \setminus \{0\}$  be a rational vector with minimal denominator  $q$ , and consider the equation

$$[V, X_\omega] = P - [P]_\omega, \quad [P]_\omega = \lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t (X_\omega^s)^* P ds. \quad (4)$$

It is easy to see that the time-average  $[P]_\omega$  has the following simple expression

$$[P]_\omega = q^{-1} \int_0^q P \circ X_\omega^t dt = \int_0^1 P \circ X_{q\omega}^t dt.$$

Unlike equation (1), equation (4) is easily solved without Fourier expansions by the following simple integral formula

$$V = q \int_0^1 (P - [P]_\omega) \circ X_{q\omega}^t dt$$

and there is no small divisors: geometrically, the integral curves of  $X_\omega$  are closed, and analytically, the inverse operator of  $[\cdot, X_\omega]$  is bounded (by  $q$ , with respect to any translation-invariant norm on the space of vector fields on the torus). This was first used by Lochak in connection with the Nekhoroshev theorem ([Loc92]). In [BF12], we proved (as a particular case) the following result in Diophantine approximation: if  $\alpha$  is non-resonant, then there exists  $n$  rational vectors  $\omega_1, \dots, \omega_n$ , of denominators  $q_1, \dots, q_n$ , such that  $q_1\omega_1, \dots, q_n\omega_n$  form a basis of  $\mathbb{Z}^n$  and such that, up to constants depending only on  $n$ , the distance  $|\alpha - \omega_i|$  is bounded by  $(q_i Q)^{-1}$  and the denominators  $q_i$  are bounded by  $\Psi_\alpha(Q)$ , where  $\Psi_\alpha$  is the function defined in (2). This result allows us to reduce the study of the equation (1) (or the approximate equation (3)) to the simpler equation (4), avoiding small divisors and the use of Fourier expansions as a consequence. Indeed, let  $P_0 = P$  and defined inductively  $P_j = [P_{j-1}]_{\omega_j}$  for  $1 \leq j \leq n$ , then we solve the  $n$  equations

$$[V_j, X_{\omega_j}] = P_{j-1} - [P_{j-1}]_{\omega_j}, \quad 1 \leq j \leq n, \quad (5)$$

which are of the type (4), and whose solutions are “controlled” in terms of  $q_j$ , hence in terms of  $\Psi_\alpha(Q)$ . Then  $|\alpha - \omega_j|$  are simply considered as “errors”,  $P_n = [\dots [P]_{\omega_1} \dots]_{\omega_n} = [P]$  as one can easily check using the fact that  $q_1\omega_1, \dots, q_n\omega_n$  is a basis of the lattice  $\mathbb{Z}^n$ , and the composition of time-one maps  $V_1^1 \circ \dots \circ V_n^1$  will give a transformation  $\Phi$  which has a similar effect as the transformation constructed by solving the equation (3).

**5.** There is a duality between the classical approach and the approach given in [BF12]. The classical approach deals with equation (1) by replacing the perturbation by a simpler (trigonometric) perturbation keeping the integrable vector field fixed (this is equation (3)), whereas in [BF12], we deal with equation (1) by replacing the integrable vector field by simpler (periodic) vector fields, keeping the perturbation fixed (these are equations (5)). In the words of Lochak (see [Loc02]), the classical approach pertains to linear Diophantine approximation, whereas [BF12] should pertain to simultaneous Diophantine approximation.

**6.** The purpose of this article is to present yet another approach to the KAM theorem, which is in some sense intermediate between the usual approach (for instance [Pös11]) and the approach given in [BF12], as it mixes both linear and simultaneous Diophantine approximation. The idea is quite simple. Instead of using  $n$  rational approximations with suitable properties (as was done in [BF12]), the existence of which is non-trivial, we will just use one rational approximation, the existence of which is trivial by Dirichlet’s box principle: for any  $Q \geq 1$ , there exists a rational vector  $\omega$  of denominator  $q$  such that  $|\alpha - \omega| \leq (qQ)^{-1}$  and  $1 \leq q \leq Q^{n-1}$ . Then we approximate the integrable vector field  $X_\alpha$  by the periodic vector field  $X_\omega$ , we solve equation (4) and we consider  $|\alpha - \omega|$  as an “error”. But this is clearly not sufficient, as the equation involves the average  $[P]_\omega$  instead of  $[P]$ , so we need furthermore to be able to consider  $[P]_\omega - [P]$  as an “error”. It is easy to see that the Fourier expansion of

$[P]_\omega - [P]$  contains only harmonics associated to integers  $k$  such that  $k \cdot \omega = 0$ , hence using the decay of Fourier coefficients of regular vector fields,  $[P]_\omega - [P]$  can be considered small enough provided that  $|k|$  is large enough. This will be proved in Proposition 2.3 below, using Khintchine transference principle which makes a connection between linear and simultaneous Diophantine approximation. One can say that our approach here avoids small divisors, since we deal only with the equation (4), even though we make use of Fourier expansions to estimate the remainder.

## 1.2 Statement of the result

1. Before stating the result, let us describe precisely the setting.

Let  $n \in \mathbb{N}$ ,  $n \geq 2$ . Without loss of generality, we may assume that

$$\alpha = (1, \tilde{\alpha}) \in \mathbb{R} \times \mathbb{R}^{n-1}, \quad \tilde{\alpha} = (\alpha_1, \dots, \alpha_{n-1}) \in [-1, 1]^{n-1}.$$

First we need to impose a Diophantine condition on  $\alpha$ , and for simplicity we will assume that  $\alpha$  satisfies a classical Diophantine condition. For  $\tau \geq 0$ , we define

$$\Omega^{n-1}(\tau) = \left\{ x \in [-1, 1]^{n-1} \mid \exists \gamma > 0, \forall k \in \mathbb{Z}^{n-1} \setminus \{0\}, \|k \cdot x\|_{\mathbb{Z}} \geq \gamma |k|^{-(1+\tau)(n-1)} \right\},$$

where  $\|\cdot\|_{\mathbb{Z}}$  denotes the distance to the lattice  $\mathbb{Z}$ , that is  $\|y\|_{\mathbb{Z}} = \min_{k \in \mathbb{Z}} |y - k|$ . Hence  $\tilde{\alpha} \in \Omega^{n-1}(\tau)$  if and only if there exists a constant  $\gamma = \gamma(\tilde{\alpha})$  such that for all  $k \in \mathbb{Z}^{n-1} \setminus \{0\}$ ,  $\|k \cdot \tilde{\alpha}\|_{\mathbb{Z}} \geq \gamma |k|^{-(1+\tau)(n-1)}$ , and it is not a restriction to assume that  $\gamma \leq 1$ .

Then we need to impose a regularity condition on the perturbation, and we will assume that it is real analytic. Recall that  $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$ , and we define  $\mathbb{T}_{\mathbb{C}}^n = \mathbb{C}^n / \mathbb{Z}^n$ . For  $z = (z_1, \dots, z_n) \in \mathbb{C}^n$ , we define  $|z| = \max_{1 \leq j \leq n} |z_j|$ , and given  $s > 0$ , we define a complex neighbourhood of  $\mathbb{T}^n$  in  $\mathbb{T}_{\mathbb{C}}^n$  by

$$\mathbb{T}_s^n = \{\theta \in \mathbb{T}_{\mathbb{C}}^n \mid |(\operatorname{Im}(\theta_1), \dots, \operatorname{Im}(\theta_n))| < s\}.$$

Without loss of generality, we may also assume that  $s \leq 1$ .

For a small parameter  $\varepsilon \geq 0$ , we will consider bounded real-analytic vector fields on  $\mathbb{T}_s^n$ , of the form

$$X = X_\alpha + P, \quad X_\alpha = \alpha = (1, \tilde{\alpha}), \quad \tilde{\alpha} \in \Omega^{n-1}(\tau), \quad |P|_s = \sup_{z \in \mathbb{T}_s^n} |P(z)| \leq \varepsilon. \quad (*)$$

By real-analytic, we mean that the vector field is analytic and is real valued for real arguments. For any such vector field  $Y$ , we shall denote by  $Y^t$  its time- $t$  map for values of  $t \in \mathbb{C}$  which makes sense, and given another vector field  $Z$ , we denote by  $[Y, Z]$  their Lie bracket. Moreover, for a real-analytic embedding  $\Phi : \mathbb{T}_r^n \rightarrow \mathbb{T}_s^n$ ,  $r \leq s$ , and for a real-analytic vector field  $Y$  which is well-defined on the image of  $\Phi$ , we let  $\Phi^*Y$  be the pull-back  $Y$ , which is well-defined on  $\mathbb{T}_r^n$ .

2. We can finally state the result, whose content is the classical KAM theorem for constant vector fields on the torus first proved by Arnold.

**Theorem 1.1.** *Let  $X$  be as in (\*). Then there exist constants  $\varepsilon_* \leq 1$  and  $C_1, C_2 \geq 1$ , which depend only on  $n, s$  and  $\tilde{\alpha}$ , such that if  $\varepsilon \leq \varepsilon_*$ , there exist a unique constant  $\beta \in \mathbb{C}^n$  and a real-analytic embedding  $\Phi : \mathbb{T}_{s/2}^n \rightarrow \mathbb{T}_s^n$  such that*

$$\Phi^*(X + X_\beta) = X_\alpha$$

with the estimates

$$|\Phi - \text{Id}|_{s/2} \leq C_1 \varepsilon, \quad |\beta| \leq C_2 \varepsilon.$$

Let us now add some remarks on the above theorem.

Firstly, we only considered vectors  $\alpha$  satisfying a classical Diophantine condition, but as in [BF12], the approach can be easily extended to a more general class of frequency vectors, even though we don't know yet what is the weakest condition on  $\alpha$  we can reach through our method.

Secondly, also as in [BF12], we have decided to formulate our result in the setting of perturbation of constant vector fields on the torus, but using further elementary techniques, for instance as described in [Pös01], we could have formulated our result in the context of perturbation of integrable Hamiltonian systems without any difficulties, provided the Hamiltonian is real-analytic and the integrable Hamiltonian is non-degenerate in the sense of Kolmogorov.

Finally, the constants  $\varepsilon_*$ ,  $C_1$ ,  $C_2$  in the statement depend on  $n$ ,  $s$  and on  $\tilde{\alpha}$  only through  $\tau$ ,  $\gamma$  and the constant  $\bar{\gamma}$  that appears in Theorem 2.2. We could have easily provided explicit values for these constants, but for clarity we decided not to do so, and in the sequel, when convenient, we will sometimes use a  $\cdot$  in replacement of any constant  $C \geq 1$  depending only on  $n, s, \tau, \gamma$  and  $\bar{\gamma}$ : that is, an expression of the form  $u \ll v$  means that there exists a constant  $C \geq 1$ , that depends only on the above set of parameters, such that  $u \leq Cv$ . Similarly, we will use the notations  $u \cdot \ll v$ ,  $u = \cdot v$  and  $u \cdot = v$ .

## 2 Proof of Theorem 1.1

The proof of Theorem 1.1 will be given in §2.3, based on a quasi-periodic averaging result we will prove in §2.2. Such a deduction is classical, but the novelty lies in the proof of the quasi-periodic averaging result, which will simply follow from a consequence of Dirichlet's box and Khintchine transference principles, as explained in §2.1, and from simple analytical estimates that are exposed in the Appendix A.

### 2.1 Dirichlet's approximation and Khintchine transference theorems

**1.** We start by recalling the classical theorem of Dirichlet on the approximation of an arbitrary non-zero vector by vector with rational components.

**Theorem 2.1** (Dirichlet). *Let  $x \in \mathbb{R}^{n-1} \setminus \{0\}$ , and  $Q \geq 1$ . Then there exists  $(q, p) \in \mathbb{N} \times \mathbb{Z}^{n-1}$  such that*

$$|qx - p| \leq Q^{-1}, \quad 1 \leq q \leq Q^{n-1}.$$

For a proof we refer to [Sch80] or [Cas57].

**2.** For  $\tau \geq 0$ , we already defined

$$\Omega^{n-1}(\tau) = \left\{ x \in [-1, 1]^{n-1} \mid \exists \gamma > 0, \forall k \in \mathbb{Z}^{n-1} \setminus \{0\}, \|k \cdot x\|_{\mathbb{Z}} \geq \gamma |k|^{-(1+\tau)(n-1)} \right\},$$

and now we define

$$\Omega_{n-1}(\tau) = \left\{ x \in [-1, 1]^{n-1} \mid \exists \bar{\gamma} > 0, \forall q \in \mathbb{N}^*, \|qx\|_{\mathbb{Z}^{n-1}} \geq \bar{\gamma} q^{-(1+\tau)(n-1)} \right\}$$

where  $\|\cdot\|_{\mathbb{Z}^{n-1}}$  denotes the distance to the lattice  $\mathbb{Z}^{n-1}$ , that is  $\|y\|_{\mathbb{Z}^{n-1}} = \min_{k \in \mathbb{Z}^{n-1}} |y - k|$ . We shall use the following statement, which is a particular case of a transference theorem of Khintchine.

**Theorem 2.2** (Khinchine). *For  $\tau \geq 0$ ,  $\Omega^{n-1}(\tau) \subseteq \Omega_{n-1}((n-1)\tau)$ . That is, if  $\tilde{\alpha} \in \Omega^{n-1}(\tau)$ , then there exists  $\bar{\gamma} = \bar{\gamma}(\tilde{\alpha})$  such that for any  $q \in \mathbb{N}^*$ ,*

$$\|q\tilde{\alpha}\|_{\mathbb{Z}^{n-1}} \geq \bar{\gamma}q^{-(1+(n-1)\tau)(n-1)^{-1}}.$$

More generally, we have  $\Omega_{n-1}(\tau((n-2)\tau + n - 1)^{-1}) \subseteq \Omega^{n-1}(\tau) \subseteq \Omega_{n-1}((n-1)\tau)$  for any  $\tau \geq 0$ , which implies in particular that  $\Omega^{n-1}(0) = \Omega_{n-1}(0)$ . For a proof, we refer to [Sch80] or [Cas57] (see also [Ber03] for another related and interesting transference result).

**3.** The proposition below, which is a consequence of Dirichlet's and Khinchine's Theorems, will be our main tool to derive our quasi-periodic averaging result.

**Proposition 2.3.** *Let  $\alpha = (1, \tilde{\alpha}) \in \mathbb{R}^n \setminus \{0\}$  with  $\tilde{\alpha} \in \Omega^{n-1}(\tau)$ , and  $Q \geq 1$ . Then there exists a vector  $\omega = (1, q^{-1}p) \in \mathbb{Q}^n$ , with  $(q, p) \in \mathbb{N} \times \mathbb{Z}^{n-1}$ , such that:*

$$(i) \quad |\alpha - \omega| = |\tilde{\alpha} - q^{-1}p| \leq (qQ)^{-1}, \quad 1 \leq q < Q^{n-1};$$

(ii) *for any  $k \in \mathbb{Z}^n \setminus \{0\}$  satisfying  $k \cdot \omega = 0$ , we have  $|k| \geq \gamma^* Q^{a-1}$  with*

$$\gamma^* = \left( n^{-1} \bar{\gamma} \bar{\gamma}^{(n-1)(1+(n-1)\tau)^{-1}} \right)^{(n+(n-1)\tau)^{-1}}, \quad a = 1 + (n-1)\tau.$$

*Proof.* The first part of the statement follows easily from Theorem 2.1 applied to  $x = \tilde{\alpha}$ , so it remains to prove the second part. Since  $\tilde{\alpha} \in \Omega^{n-1}(\tau)$ , we can use Theorem 2.2 and together with (i), we obtain

$$Q^{-1} \geq |q\tilde{\alpha} - p| \geq \|q\tilde{\alpha}\|_{\mathbb{Z}^{n-1}} \geq \bar{\gamma}q^{-(1+(n-1)\tau)(n-1)^{-1}}$$

which implies

$$q \geq (\bar{\gamma}Q)^{(n-1)(1+(n-1)\tau)^{-1}}. \quad (6)$$

Now let  $k \in \mathbb{Z}^n \setminus \{0\}$  such that  $k \cdot \omega = 0$ , and let us write  $k = (k_0, \tilde{k}) \in \mathbb{Z} \times \mathbb{Z}^{n-1}$ . Necessarily  $\tilde{k}$  is non-zero. Now  $k \cdot \omega = 0$  is equivalent to  $qk \cdot \omega = qk_0 + \tilde{k} \cdot p = 0$ , hence

$$q(k_0 + \tilde{k} \cdot \tilde{\alpha}) = q(k_0 + \tilde{k} \cdot \tilde{\alpha}) - (qk_0 + \tilde{k} \cdot p) = (q\tilde{\alpha} - p) \cdot \tilde{k}. \quad (7)$$

On the one hand, using Cauchy-Schwarz inequality and (i), we have

$$|(q\tilde{\alpha} - p) \cdot \tilde{k}| \leq n|q\tilde{\alpha} - p| |\tilde{k}| \leq nQ^{-1} |\tilde{k}|,$$

and on the other hand, using the fact that  $\tilde{\alpha} \in \Omega^{n-1}(\tau)$ , we have

$$|q(k_0 + \tilde{k} \cdot \tilde{\alpha})| \geq q \|\tilde{k} \cdot \tilde{\alpha}\| \geq q\gamma |\tilde{k}|^{-(1+\tau)(n-1)}.$$

These last two inequalities, together with the equality (7), implies

$$|\tilde{k}|^{(1+\tau)(n-1)+1} = |\tilde{k}|^{n+(n-1)\tau} \geq n^{-1} \gamma q Q,$$

and using (6), this gives

$$\begin{aligned} |\tilde{k}|^{n+(n-1)\tau} &\geq n^{-1} \bar{\gamma} \bar{\gamma}^{(n-1)(1+(n-1)\tau)^{-1}} Q^{(n-1)(1+(n-1)\tau)^{-1}+1} \\ &= n^{-1} \bar{\gamma} \bar{\gamma}^{(n-1)(1+(n-1)\tau)^{-1}} Q^{(n+(n-1)\tau)(1+(n-1)\tau)^{-1}}. \end{aligned}$$

This proves that  $|\tilde{k}| \geq \gamma^* Q^{a-1}$ , with  $\gamma^*$  and  $a$  as in the statement, and this proves (ii) since  $|k| \geq |\tilde{k}|$ .  $\square$

## 2.2 Quasi-periodic averaging

Recall that  $a = 1 + (n - 1)\tau \geq 1$  have been defined in the statement of Proposition 2.3, now we define three additional constants

$$b = 4^{na} > 1, \quad c = b^{-1}(1 - b^{-1})^{-1} < 1, \quad d = c + 1 = (1 - b^{-1})^{-1} \geq 1. \quad (8)$$

Our aim here is to prove the following quasi-periodic averaging result.

**Proposition 2.4.** *Let  $Y = X + S$ , with  $X$  as in (\*) and  $|S|_s \leq d\varepsilon$ . For any  $0 < \sigma < s$  and  $Q \geq 1$ , assume that*

$$Q^n \varepsilon \leq 1, \quad Q^{-1} \sigma^{-1} \ll 1, \quad Q \sigma^{1-n} e^{-\gamma^* Q^{a-1} \sigma/2} \leq 1. \quad (9)$$

Then there exists a real analytic embedding  $\Phi_1 : \mathbb{T}_{s-\sigma}^n \rightarrow \mathbb{T}_s^n$  such that

$$\Phi_1^* Y = X_\alpha + S + [P] + P^+, \quad [P] = \int_{\mathbb{T}^n} P,$$

with the estimates

$$|\Phi_1 - \text{Id}|_{s-\sigma} \leq Q^{n-1} \varepsilon, \quad |P^+|_{s-\sigma} \leq b^{-1} \varepsilon.$$

The above statement does not concern  $X$  but the modified vector field  $X + S$ , where  $S$  is a sufficiently small arbitrary vector field.  $S$  does not play any role here, but it will become useful for the proof of Theorem 1.1 to deal with the shift of frequency (caused by the average  $[P]$ ) at each step of the iterative scheme. The choice of the constant  $b > 1$  (and subsequently the choice of  $c$  and  $d$ ) is rather arbitrary, but was made in order to simplify the proof of Theorem 1.1.

*Proof of Proposition 2.4.* Since  $\alpha = (1, \tilde{\alpha}) \in \mathbb{R}^n \setminus \{0\}$  with  $\tilde{\alpha} \in \Omega^{n-1}(\tau)$ , for  $Q \geq 1$  we can apply Proposition 2.3: there exists a vector  $\omega = (1, q^{-1}p) \in \mathbb{Q}^n$ , with  $(q, p) \in \mathbb{N} \times \mathbb{Z}^{n-1}$ , and by (i),

$$|\alpha - \omega| = |\tilde{\alpha} - q^{-1}p| \leq (qQ)^{-1}, \quad 1 \leq q \leq Q^{n-1}.$$

We set

$$[P]_\omega = \int_0^1 P \circ X_{q\omega}^t dt, \quad V = q \int_0^1 (P - [P]_\omega) \circ X_{q\omega}^t dt.$$

Since  $|P|_s \leq \varepsilon$ , then obviously  $|[P]_\omega|_s \leq \varepsilon$  and  $|V|_s \leq q\varepsilon$ . Since  $q \leq Q^{n-1}$ , from the first and second part of (9), we have in particular

$$q\varepsilon \ll Q^{n-1} \varepsilon \ll Q^{-1} \ll \sigma \quad (10)$$

hence by Lemma A.1 (applied with  $\varsigma = \sigma$ ), the map  $V^1 : \mathbb{T}_{s-\sigma}^n \rightarrow \mathbb{T}_s^n$  is a well-defined real-analytic embedding and

$$|V^1 - \text{Id}|_{s-\sigma} \leq |V|_s \leq q\varepsilon \leq Q^{n-1} \varepsilon.$$

Let us define  $\varpi = \alpha - \omega$  so that  $Y = X + S = X_\alpha + S + P = X_\omega + X_\varpi + S + P$ . Now we can write

$$(V^1)^* Y = (V^1)^* X_\omega + (V^1)^* (X_\varpi + S + P), \quad (11)$$

and using the general equality

$$\frac{d}{dt}(V^t)^*F = (V^t)^*[F, V]$$

for an arbitrary vector field  $F$ , we can apply Taylor's formula with integral remainder to the right-hand side of (11), at order two for the first term and at order one for the second term, and we get

$$(V^1)^*Y = X_\omega + [X_\omega, V] + \int_0^1 (1-t)(V^t)^*[[X_\omega, V], V]dt + X_\varpi + S + P + \int_0^1 (V^t)^*[X_\varpi + S + P, V]dt.$$

Now let us check that the equality  $[V, X_\omega] = P - [P]_\omega$  holds true: let us denote  $G = P - [P]_\omega$  and  $DV$  the differential of  $V$ , then since  $X_\omega$  is a constant vector field, we have

$$[V, X_\omega] = DV.\omega = q \int_0^1 D(G \circ X_{q\omega}^t).\omega dt = \int_0^1 D(G \circ X_{q\omega}^t).q\omega dt$$

so using the chain rule

$$[V, X_\omega] = \int_0^1 \frac{d}{dt}(G \circ X_{q\omega}^t)t dt$$

and an integration by parts

$$[V, X_\omega] = (G \circ X_{q\omega}^t)t \Big|_0^1 - \int_0^1 G \circ X_{q\omega}^t dt = G,$$

where in the last equality,  $G \circ X_{q\omega}^1 = G$  since  $q\omega \in \mathbb{Z}^n$  and the integral vanishes since  $[G] = 0$ . So using the equality  $[V, X_\omega] = P - [P]_\omega$ , that can be written as  $[X_\omega, V] + P = [P]_\omega$ , we have

$$(V^1)^*Y = X_\omega + X_\varpi + S + [P]_\omega + \int_0^1 (1-t)(V^t)^*[[X_\omega, V], V]dt + \int_0^1 (V^t)^*[X_\varpi + S + P, V]dt,$$

and if we set

$$P_t = tP + (1-t)[P]_\omega, \quad \tilde{P} = \int_0^1 (V^t)^*[X_\varpi + S + P_t, V]dt, \quad P^+ = \tilde{P} + [P]_\omega - [P]$$

and use again the equality  $[X_\omega, V] = [P]_\omega - P$  we eventually obtain

$$(V^1)^*Y = X_\omega + X_\varpi + S + [P]_\omega + \tilde{P} = X_\alpha + S + [P] + P^+.$$

Letting  $\Phi_1 = V^1$ , it remains only to estimate  $P^+$ .

Let us first estimate  $\tilde{P}$ , and for that let us write  $U = [X_\varpi + S + P_t, V]$ . This is just a sum of three Lie brackets, and since

$$|X_\varpi|_s = |\varpi| = |\alpha - \omega| \leq (qQ^{-1}), \quad |S|_s \leq d\varepsilon, \quad |P_t|_s \leq \varepsilon, \quad |V|_s \leq q\varepsilon,$$

each of them can be estimated by Lemma A.2 (applied with  $\varsigma = \sigma/2$ ) and we obtain

$$|U|_{s-\sigma/2} \leq \sigma^{-1}q\varepsilon((qQ)^{-1} + d\varepsilon + \varepsilon) \leq \sigma^{-1}Q^{-1}\varepsilon + \sigma^{-1}q\varepsilon^2 \leq \sigma^{-1}Q^{-1}\varepsilon$$



where the last inequality follows from (10), as  $q\varepsilon \leq Q^{n-1}\varepsilon \leq Q^{-1}$ . Then using (10) again, we can apply Lemma A.3 (with  $\varsigma = \sigma/2$  and  $s - \sigma/2$  instead of  $s$ ) to obtain

$$|\tilde{P}|_{s-\sigma} \leq \sup_{|t| \leq 1} |(V^t)^* U|_{s-\sigma} \leq |U|_{s-\sigma/2} \leq \sigma^{-1} Q^{-1} \varepsilon.$$

Now let us estimate  $[P]_\omega - [P]$ , but first observe that if  $P = (P_1, \dots, P_n)$  and  $P_j = \sum_{k \in \mathbb{Z}^n} P_{j,k} e_k$  is the Fourier expansion of  $P_j$  for  $1 \leq j \leq n$  (recall that  $e_k(\theta) = e^{ik \cdot \theta}$ , for  $k \in \mathbb{Z}^n$ ), then

$$[P] = (P_{1,0}, \dots, P_{n,0})$$

and if  $[P]_\omega = (P_1^\omega, \dots, P_n^\omega)$  and  $P_j^\omega = \sum_{k \in \mathbb{Z}^n} P_{j,k}^\omega e_k$ , then

$$P_{j,k}^\omega = \begin{cases} P_{j,k}, & k \cdot \omega = 0, \\ 0, & k \cdot \omega \neq 0. \end{cases} \quad (12)$$

The first assertion is obvious, and the second follows from the computation below, where we denotes  $\bar{e}_k(\theta) = e^{-ik \cdot \theta}$ :

$$\begin{aligned} P_{j,k}^\omega &= \int_{\mathbb{T}^n} \bar{e}_k P_j^\omega = \int_{\mathbb{T}^n} \bar{e}_k \int_0^1 P_j \circ X_{q\omega}^t dt = \int_{\mathbb{T}^n} \bar{e}_k \int_0^1 \sum_{l \in \mathbb{Z}^n} P_{j,l} e_l e^{itl \cdot q\omega} dt \\ &= \sum_{l \in \mathbb{Z}^n} P_{j,l} \int_{\mathbb{T}^n} \bar{e}_k e_l \int_0^1 e^{itl \cdot q\omega} dt = P_{j,k} \int_0^1 e^{itk \cdot q\omega} dt. \end{aligned}$$

Now by Proposition 2.3, (ii), if  $k \cdot \omega = 0$  and  $k \neq 0$ , then  $|k| \geq \gamma^* Q^{a-1}$ , so we can apply Lemma A.4 (with  $K = \gamma^* Q^{a-1}$  and  $\varsigma = \sigma$ ) to obtain the estimate

$$|[P]_\omega - [P]| \leq \sigma^{-n} e^{-\gamma^* Q^{a-1} \sigma/2} |P|_s \leq \sigma^{-n} e^{-\gamma^* Q^{a-1} \sigma/2} \varepsilon$$

and therefore, using the last part of (9),

$$|[P]_\omega - [P]|_{s-\sigma} \leq Q^{-1} \sigma^{-1} \varepsilon.$$

Eventually, choosing the implicit constant in the second part of (9) sufficiently large, we obtain

$$|P^+|_{s-\sigma} \leq |\tilde{P}|_{s-\sigma} + |[P]_\omega - [P]|_{s-\sigma} \leq Q^{-1} \sigma^{-1} \varepsilon \leq b^{-1} \varepsilon$$

which ends the proof.  $\square$

### 2.3 Proof of Theorem 1.1

For a given  $r \geq 0$ , let

$$B_r(\alpha) = \{x \in \mathbb{C}^n \mid |\alpha - x| \leq r\}.$$

The proposition below is just a more convenient reformulation of Proposition 2.4, where we recall that the constants  $c$  and  $d$  have been defined in (8).

**Proposition 2.5.** *Let  $X$  be as in (\*). For any  $0 < \sigma < s$  and  $Q \geq 1$ , assume that*

$$Q^n \varepsilon \leq 1, \quad Q^{-1} \sigma^{-1} \ll 1, \quad Q \sigma^{1-n} e^{-\gamma^* Q^{a-1} \sigma/2} \leq 1. \quad (13)$$

*Then there exist an embedding  $\varphi_1 : B_{c\varepsilon}(\alpha) \rightarrow B_{d\varepsilon}(\alpha)$  and a real analytic embedding  $\Phi_1 : \mathbb{T}_{s-\sigma}^n \rightarrow \mathbb{T}_s^n$  such that for all  $x \in B_{c\varepsilon}(\alpha)$ ,*

$$\Phi_1^*(X_{\varphi_1(x)} + P) = X_x + P^+$$

*with the estimates*

$$|\Phi_1 - \text{Id}|_{s-\sigma} \leq Q^{n-1} \varepsilon, \quad |P^+|_{s-\sigma} \leq b^{-1} \varepsilon.$$

*Proof.* Recall that  $[P]$  is a constant,  $[P] \in \mathbb{C}^n$ , and as  $|P|_s \leq \varepsilon$ , then  $|[P]| \leq \varepsilon$ . So let us define  $\varphi_1 : B_{c\varepsilon}(\alpha) \rightarrow B_{d\varepsilon}(\alpha)$  to be the translation

$$\varphi_1(x) = x - [P], \quad x \in B_{c\varepsilon}(\alpha).$$

Take any  $x \in B_{c\varepsilon}(\alpha)$ , then  $|\varphi_1(x) - \alpha| \leq |x - \alpha| + |[P]| \leq (c+1)\varepsilon = d\varepsilon$  and we can write

$$X_{\varphi_1(x)} + P = X_\alpha + X_{\varphi_1(x)-\alpha} + P$$

and by condition (9), we can apply Proposition 2.4 with  $S = X_{\varphi_1(x)-\alpha}$ ,  $|S|_s \leq d\varepsilon$ , to find a real analytic embedding  $\Phi_1 : \mathbb{T}_{s-\sigma}^n \rightarrow \mathbb{T}_s^n$  such that

$$\Phi_1^*(X_{\varphi_1(x)} + P) = X_{\varphi_1(x)} + [P] + P^+ = X_{\varphi_1(x)+[P]} + P^+ = X_x + P^+$$

with the estimates

$$|\Phi_1 - \text{Id}|_{s-\sigma} \leq Q^{n-1} \varepsilon, \quad |P^+|_{s-\sigma} \leq b^{-1} \varepsilon.$$

This was the statement to prove. □

We will now use Proposition 2.5 as the building block of an iterative scheme leading to the proof of Theorem 1.1.

*Proof.* For a constant  $Q \geq 1$  to be chosen below, we define, for any  $m \in \mathbb{N}$ ,

$$\varepsilon_m = b^{-m} \varepsilon, \quad Q_m = b^{n-1} Q = 4^{am} Q, \quad \sigma_m = 2^{-m-2} s.$$

We also define  $s_0 = s$  and  $s_{m+1} = s_m - \sigma_m$  for  $m \in \mathbb{N}$ , and for convenience, we let  $\varepsilon_{-1} = d\varepsilon$ . We claim that for  $Q_0 = Q$  sufficiently large, the conditions

$$Q_m^n \varepsilon_m \leq 1, \quad Q_m^{-1} \sigma_m^{-1} \ll 1, \quad Q_m \sigma_m^{1-n} e^{-\gamma^* Q_m^{a-1} \sigma_m/2} \leq 1 \quad (14)$$

are satisfied for any  $m \in \mathbb{N}$ .

Indeed, we can choose  $Q = \cdot 1$  with a sufficiently large implicit constant so that the last two conditions in (14) are satisfied for  $m = 0$ , then if we define  $\varepsilon_* = Q^{-n}$ , the threshold  $\varepsilon \leq \varepsilon_*$  implies the first condition in (14) for  $m = 0$ . Next it is easy to see that  $Q_m^n \varepsilon_m = Q_0^n \varepsilon_0$  and  $Q_m \sigma_m = 4^{am} 2^{-m} Q_0 \sigma_0 \geq Q_0 \sigma_0$  since  $a \geq 1$ , therefore the first two conditions in (14) are satisfied for any  $m \in \mathbb{N}$ . For the last condition, let

$$M_m = Q_m \sigma_m^{1-n} e^{-\gamma^* Q_m^{a-1} \sigma_m/2},$$

observing that  $Q_m^{a^{-1}}\sigma_m = 2^m Q_0^{a^{-1}}\sigma_0$ , we have for any  $m \geq 1$ ,

$$M_{m+1}M_m^{-1} = 4^a 2^{n-1} e^{-\gamma^* Q_m^{a^{-1}}\sigma_m/2} \leq 4^a 2^{n-1} e^{-\gamma^* Q_0^{a^{-1}}\sigma_0}$$

and hence  $M_{m+1}M_m^{-1} \leq 1$ , up to taking a larger implicit constant if necessary in the definition  $Q = \cdot 1$ . So this proves the claim.

Next we claim that for any  $m \in \mathbb{N}$ , there exist an embedding  $\varphi^m : B_{c\varepsilon_{m-1}}(\alpha) \rightarrow B_{d\varepsilon}(\alpha)$  and a real analytic embedding  $\Phi^m : \mathbb{T}_{s_m}^n \rightarrow \mathbb{T}_s^n$  such that for all  $x_m \in B_{c\varepsilon_{m-1}}(\alpha)$ ,

$$(\Phi^m)^*(X_{\varphi^m(x_m)} + P) = X_{x_m} + P_m$$

with the estimates

$$|\Phi^m - \text{Id}|_{s_m} \leq \sum_{j=0}^{m-1} Q_j^{n-1} \varepsilon_j = Q^{n-1} \varepsilon \sum_{j=0}^{m-1} b^{-jn^{-1}}, \quad |P_m|_{s_m} \leq \varepsilon_m. \quad (15)$$

Indeed, for  $m = 0$ , choosing  $\varphi^0$  and  $\Phi^0$  to be the identity, and  $P_0 = P$ , there is nothing to prove. If we assume that the statement holds true for some  $m \in \mathbb{N}$ , then by (14) we can apply Proposition 2.5 (with  $\varepsilon_m, Q_m$  and  $\sigma_m$  instead of  $\varepsilon, Q$  and  $\sigma$ ) to the resulting vector field and an embedding  $\varphi_{m+1} : B_{c\varepsilon_m}(\alpha) \rightarrow B_{d\varepsilon_m}(\alpha)$  and a real analytic embedding  $\Phi_{m+1} : \mathbb{T}_{s_{m+1}}^n \rightarrow \mathbb{T}_{s_m}^n$  are constructed. It is then sufficient to let  $\varphi^{m+1} = \varphi^m \circ \varphi_{m+1}$ , which is well defined since  $d\varepsilon_{m+1} = c\varepsilon_m$ ,  $\Phi^{m+1} = \Phi^m \circ \Phi_{m+1}$ ,  $P_{m+1} = P_m^+$  and the estimates (15) are obvious.

To conclude, note that

$$\lim_{m \rightarrow +\infty} \varepsilon_m = 0, \quad \lim_{m \rightarrow +\infty} s_m = s/2,$$

and together with the estimates (15), when  $m$  goes to infinity,  $\varphi^m$  converges to a trivial map  $\varphi : \{\alpha\} \rightarrow B_{d\varepsilon}(\alpha)$ ,  $P_m$  converges to zero uniformly on every compact subsets of  $\mathbb{T}_{s/2}^n$ , while  $\Phi^m$  converges to an embedding  $\Phi : \mathbb{T}_{s/2}^n \rightarrow \mathbb{T}_s^n$ , uniformly on every compact subsets of  $\mathbb{T}_{s/2}^n$ . Since the space of real-analytic functions is closed for the topology of uniform convergence on compact subsets,  $\Phi$  is real-analytic, with the estimate

$$|\Phi - \text{Id}|_{s/2} \leq \left(1 - b^{-n^{-1}}\right)^{-1} Q^{n-1} \varepsilon.$$

Finally, at the limit we have  $\Phi^*(X_{\varphi(\alpha)} + P) = X_\alpha$ , hence if  $\beta \in \mathbb{C}^n$  is the unique vector such that  $\varphi(\alpha) = \alpha + \beta$ , then  $|\beta| \leq d\varepsilon$  and

$$\Phi^*(X + X_\beta) = X_\alpha.$$

This was the statement to prove. □

## A Technical estimates

In this appendix, we recall some technical estimates that we used for the proof of Proposition 2.4. These estimates are classical: proofs of Lemma A.1, Lemma A.2 and Lemma A.3 can be found in [BF12] and a proof of Lemma A.1 can be found in [BGG85].

**Lemma A.1.** *Let  $V$  be a bounded real-analytic vector field on  $\mathbb{T}_s^n$ ,  $0 < \varsigma < s$  and  $\tau = \varsigma|V|_s^{-1}$ . For  $t \in \mathbb{C}$  such that  $|t| < \tau$ , the map  $V^t : \mathbb{T}_{s-\varsigma}^n \rightarrow \mathbb{T}_s^n$  is a well-defined real-analytic embedding, and we have*

$$|V^t - \text{Id}|_{s-\varsigma} \leq |V|_s, \quad |t| < \tau.$$

Moreover,  $V^t$  depends analytically on  $t$ , for  $|t| < \tau$ .

**Lemma A.2.** *Let  $X$  and  $V$  be two bounded real-analytic vector fields on  $\mathbb{T}_s^n$ , and  $0 < \varsigma < s$ . Then*

$$|[X, V]|_{s-\varsigma} \leq 2\varsigma^{-1}|X|_s|V|_s.$$

**Lemma A.3.** *Let  $X$  and  $V$  be two bounded real-analytic vector fields on  $\mathbb{T}_s^n$ , and  $0 < \varsigma < s$ . Assume that  $|V|_s \leq (4e)^{-1}\varsigma$ . Then for all  $|t| \leq 1$ ,*

$$|(V^t)^* X|_{s-\varsigma} \leq 2|X|_s.$$

**Lemma A.4.** *Let  $X$  be a bounded real-analytic vector field on  $\mathbb{T}_s^n$ ,  $0 < \varsigma < s$  and  $K \geq 1$ . If  $X = (X_1, \dots, X_n)$  and  $X_j = \sum_{k \in \mathbb{Z}^n} X_{j,k} e_k$ , then the vector field  $X^K = (X_1^K, \dots, X_n^K)$  defined by*

$$X_j^K = \sum_{k \in \mathbb{Z}^n, |k| \geq K} X_{j,k} e_k, \quad 1 \leq j \leq n$$

satisfies

$$|X^K|_{s-\varsigma} \leq C_n \varsigma^{-n} e^{-K\varsigma/2} |X|_s$$

for a constant  $C_n \geq 1$  which depends only on  $n$ .

## References

- [Arn61] V.I. Arnol'd, *Small denominators. I. Mapping the circle onto itself*, Izv. Akad. Nauk SSSR Ser. Mat. **25** (1961), 21–86.
- [Arn63a] ———, *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.
- [Arn63b] ———, *Small denominators and problems of stability of motion in classical and celestial mechanics*, Russ. Math. Surv. **18** (1963), no. 6, 85–191.
- [Ber03] P. Bernard, *Une propriété de transfert en approximation diophantienne*, Ann. Fac. Sci. Toulouse Math. (6) **12** (2003), no. 4, 453–463.
- [BF12] A. Bounemoura and S. Fischler, *A diophantine duality applied to the KAM and Nekhoroshev theorems*, Math. Z. (2012), to appear.
- [BGG85] G. Benettin, L. Galgani, and A. Giorgilli, *A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems*, Celestial Mech. **37** (1985), 1–25.
- [Cas57] J.W.S. Cassels, *An introduction to Diophantine approximation*, Cambridge Tracts in Math. and Math. Phys., no. 45, Cambridge University Press, 1957.

- [Kol53] A. N. Kolmogorov, *On dynamical systems with an integral invariant on the torus*, Doklady Akad. Nauk SSSR (N.S.) **93** (1953), 763–766.
- [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.
- [Loc92] P. Lochak, *Canonical perturbation theory via simultaneous approximation*, Russ. Math. Surv. **47** (1992), no. 6, 57–133.
- [Loc02] ———, *Simultaneous Diophantine approximation in classical perturbation theory: why and what for?*, Progress in nonlinear science, Vol. 1 (Nizhny Novgorod, 2001), RAS, Inst. Appl. Phys., Nizhniï Novgorod, 2002, pp. 116–138.
- [Mos62] J. Moser, *On Invariant curves of Area-Preserving Mappings of an Annulus*, Nachr. Akad. Wiss. Göttingen **II** (1962), 1–20.
- [Mos66] ———, *A rapidly convergent iteration method and non-linear differential equations. II*, Ann. Scuola Norm. Sup. Pisa (3) **20** (1966), 499–535.
- [Pös01] J. Pöschel, *A lecture on the classical KAM theory*, Katok, Anatole (ed.) et al., Smooth ergodic theory and its applications (Seattle, WA, 1999). Providence, RI: Amer. Math. Soc. (AMS). Proc. Symp. Pure Math. 69, 707–732, 2001.
- [Pös11] ———, *KAM à la R*, Regul. Chaotic Dyn. **16** (2011), no. 1-2, 17–23.
- [Rüs10] H. Rüssmann, *KAM-iteration with nearly infinitely small steps in dynamical systems of polynomial character*, Discrete Contin. Dyn. Syst. Ser. S **3** (2010), no. 4, 683–718.
- [Sch80] W. Schmidt, *Diophantine approximation*, Lecture Notes in Math., no. 785, Springer, 1980.
- [Sev03] M. B. Sevryuk, *The classical KAM theory at the dawn of the twenty-first century*, Mosc. Math. J. **3** (2003), no. 3, 1113–1144, 1201–1202, {Dedicated to Vladimir Igorevich Arnold on the occasion of his 65th birthday}.