

# SIMULTANEOUS DIOPHANTINE APPROXIMATION IN CLASSICAL PERTURBATION THEORY: WHY AND WHAT FOR?

**P.Lochak**

Institut de Mathématiques  
175 rue du Chevaleret, F-75013 Paris, France

**Abstract.** The use of simultaneous diophantine approximation in perturbation theory is quite recent, having been introduced by the author around 1990. After recalling part of the motivation and some basic facts, we give here a brief survey of some results which have been obtained using this alternative to the classical approach via small divisors techniques. The main purpose is to underline some important ideas and provide a concise reading guide.

## 1. Introduction: from physics to arithmetic.

(1.1) Small divisors in perturbation theory have been around for more than two centuries (at first somewhat implicitly). One comes across these quantities when trying to inductively sharpen the computation of a trajectory of the perturbed system starting from a solution of the unperturbed one, or in more modern terms, stemming from Poincaré’s thesis, trying to build a normal form for the perturbed system, that is an adapted coordinate system. Letting  $\omega \in \mathbf{R}^n$  denote an  $n$ -vector one is confronted with integrating along the linear flow defined by  $\omega$  on the  $n$ -torus  $\mathbf{T}^n$ , in other words one has to deal with the first order, linear, constant coefficients partial differential equation:

$$\omega \cdot \nabla \chi(q) = g(q). \quad (1)$$

Here  $\chi$  and  $g$  are scalar functions of  $q \in \mathbf{T}^n$ ,  $\nabla \chi$  denotes the gradient of  $\chi$  and the dot denotes ordinary scalar product. Assume for a moment that there are no commensurability relations among the components of  $\omega$  and that  $g$  has zero spatial average. Then expand  $g$  as a Fourier series:

$$g(q) = \sum_{k \in \mathbf{Z}^n} g_k e^{ikq}, \quad (2)$$

where  $kq = k \cdot q$  and  $g_0 = 0$ . The solution of (1) with zero spatial average reads:

$$\chi(q) = \sum_{k \in \mathbf{Z}^n} \frac{-ig_k}{\omega \cdot k} e^{ikq}, \quad (3)$$

where the term with  $k = 0$  vanishes by definition. The denominators (or divisors)  $\omega \cdot k$  are then inevitably “small” for certain “large” values of  $k \in \mathbf{Z}^n$

and the quantitative analysis of this phenomenon has been the subject of an enormous amount of work.

**(1.2)** This approach using Fourier series and unavoidably generating small divisors has been so much part of the subject that it may be surprising that there actually exists an alternative approach to some of the most important problems in Hamiltonian perturbation theory. It was first proposed and developed in [L1] with, as an immediate example of application a very substantial improvement (both conceptually and quantitatively) and simplification of the by now “classical” results presented in [N]. In this text our aim is to review the motivation, the general context and the recent achievements which this viewpoint has made possible. All the material surveyed here has been written up in detail so that we may afford to be concise and privilege general suggestive arguments. The text is divided into sections which serve essentially to stress the various subjects we broach; their varying lengths certainly do not reflect the importance of these themes. We have appended a relatively modest bibliography, which is by no means complete; it should however be amply sufficient for a start, and further reading suggestions can be easily collected using modern means of bibliographical research, which in our view should help make reference lists shorter, not longer.

**(1.3)** Recall that the famous Fermi-Pasta-Ulam experiment, one of the first “numerical experiment” ever, was originally conceived in order to check a prediction by Einstein about heat conduction. Einstein’s law is certainly correct, but the experiment gave rise to many surprises, which are still far from being completely understood. In the same vein, physicists in Milano in the early eighties were trying to understand more rigorously the connection between classical dynamics and statistical thermodynamics. An important point is that in many of these “toy models” that one-dimensional lattices provide, equipartition of energy takes a “surprisingly” long time to settle: the system exhibits less “chaos” than would be expected. This is indeed the case in the FPU model but it should be noted that this model consists of a perturbation of a chain of harmonic oscillators, i.e. linear or *isochronous* objects, whereas the experiments in Milano also involved perturbations of chains of rotators, i.e. nonlinear (integrable) or *anisochronous* objects; this is a key difference from the point of view of classical dynamics. In any case, reviving reasonings due to Boltzmann and Jeans at the beginning of the twentieth century the Milano group conjectured that the time needed for energy equipartition to settle (if and when it settles at all) may be due to a phenomenon which can be roughly described as the freezing of the fast degrees of freedom. This in principle opens the possibility of understanding the foundations of statistical physics ([BGG2]) in a new way, hypothetically also to shed some light on the basic quantization principle, namely Planck’s law for the black body radiation.

Clearly this extremely terse summary does not do justice to the subject but we hope it may arouse the curiosity of the reader to explore these fascinating (and still mysterious!) areas.

(1.4) So the story started in fact from thermodynamics, in particular from infinite dimensional classical systems (to which we briefly return in §7 below). Note that this is not without precedent and for instance one of the primary goals of Kolmogorov in devising his famous theorem on the preservation of invariant tori ([K]; see also his 1954 Amsterdam ICM address) was to test Boltzmann’s chaotic hypothesis; indeed Kolmogorov’s result shows that near-integrable systems are not ergodic on an energy surface because of the existence of a set of invariant tori of nonzero measure. At this point, seeking for mathematical tools, the Milano group came across what is now known as “Nekhoroshev’s theorem”, which was completely unknown in the western world at that time, and they rewrote a version of the original proof in the convex case ([BGG1]). Roughly speaking Nekhoroshev’s results deal with the long-time stability of near-integrable systems. Following in part suggestions of V.Arnold he proves a general result showing that perturbations of “truly nonlinear” systems are stable over exponentially long (w.r.t. to the inverse of the size of the perturbation) intervals of time; see below for –some– more detail. This is a fundamental result in Hamiltonian perturbation theory and it may be slightly surprising that it came so late. Moreover there did not even exist at the time any result of the kind over *polynomial* times that use the nonlinearity of the unperturbed Hamiltonian. One should beware of the fact that these results are truly different and more elaborate than those which concern the perturbations of linear (or isochronous, harmonic) nonresonant systems, as developed in particular in the famous book by N.N.Bogoliubov and Y.A.Mitropolosky (see e.g. [LM] for detail).

The results contained in [N] were actually obtained around 1970 (see the 1971 announcement by the author) but it took a long time to write them up. Actually [N] does not contain the “technicalities” which appeared in a later (1979) paper by the same author. So these are all “post-KAM” results, which again, historically speaking may appear as a little surprising; we would like to stress however that the basic techniques are actually the same as in “KAM theory”, drawing from the usual toolbox of perturbation theory (we return to this point in §8). In any case it took 15 years and the interest of a group of physicists for the foundations of thermodynamics and quantum physics for these important results to reach the western world (and unfortunately be turned in part into a shallow key- and catch-word – like the magic “KAM”).

(1.5) Let us fix some basic notation. We consider classical near-integrable Hamiltonian systems with Hamilton’s function of the form:

$$H(p, q) = h(p) + \varepsilon f(p, q), \tag{4}$$

where  $(p, q) \in \mathbf{R}^n \times \mathbf{T}^n = T^*\mathbf{T}^n$  are action-angle variables for the unperturbed Hamiltonian  $h$ . We write  $\omega(p) = \nabla h(p) \in \mathbf{R}^n$  for the associated frequency map. If the function  $h$  is *convex*, the map  $p \mapsto \omega(p)$  is a local diffeomorphism between action and frequency space; if  $h$  is *quasiconvex* i.e. the restriction of  $h$  to a level set  $h(p) = E$  is convex, then  $\omega$  is a local diffeomorphism between the said energy level and the projective frequency vector (i.e.  $\omega$  as viewed in  $(n - 1)$ -projective space). Details can be found in [N], [L1], [Pö1] and many other places. A weaker condition to be placed on  $h$  is *steepness*, to which we briefly return in §6.

Nekhoroshev’s main result (in [N]) states that *action variables of analytic perturbations of (analytic) steep systems are stable over exponentially long times*. Namely assuming that  $h$  and  $f$  are analytic and that  $h$  is steep, for small enough  $\varepsilon$  one has the stability estimate:

$$\|p(t) - p(0)\| < c\varepsilon^b \quad \text{for} \quad |t| < c \exp(c\varepsilon^a). \quad (5)$$

Here the  $c$ ’s are (different) constants; so are the exponents  $a$  and  $b$  which satisfy  $0 < a, b < 1$  and have come to be known (since [L1,2]) as the *stability exponents*. Details of the original statement can of course be found in [N], which is filled with interesting remarks (but does not contain the details of the proof, as mentioned above).

By far the most important case for physical applications is that when  $h$  is convex (or possibly quasiconvex), if only because kinetic energy is usually convex so that for instance energetic particles in a fixed potential provide physically relevant examples. So are weakly interacting rotators as mentioned in §1.3 above ( $h(p) = \frac{1}{2}p^2$ ) but *not* weakly interacting harmonic oscillators ( $h(p) = \omega \cdot p$  for a *fixed*  $\omega \in \mathbf{R}^n$ ) as in the work of Bogoliubov-Mitropolski and the FPU experiment.

The timescale over which the system is stable is given by the exponent  $a$ . If  $h$  is convex,  $a = a(n)$  depends on the number of degrees of freedom only and decreases with  $n$ . In the original paper ([N]), N.N.Nekhoroshev found a dependence of the form  $a(n) = O(1/n^2)$  which however was just dictated by the proof method and had no particular meaning.

**(1.6)** Returning to a possible use of such results in order to understand the foundations of thermodynamics (if only on toy models!) we are faced with at least two problems:

- Find an intrinsic and “universal” meaning of the time of stability;
- Find a meaningful way of going to the thermodynamical limit.

In more detail, the first problem consists in drawing an intrinsic line between stability and instability: on what timescale does instability actually switch on? This concerns at first finite dimensional systems. Instability in that case takes the form of the so-called “Arnold diffusion”, after [A1] whose

title is of course more adequate and explicit. It so happens that a highly conjectural answer was contained in [C]. We return to this question in §§4, 9.

The second question is quite dramatic: if  $a$  decreases to 0 as  $n$  goes to infinity, there does not remain much to say at the thermodynamical limit. Now the lattices which one uses in order to model statistical mechanics have of course additional spatial structures which make it possible to say much more specific things, as is also the case in KAM theory. But one first needs so to speak to *localize* the results. In other words (5) above states a stability result for *all* initial conditions  $(p(0), q(0))$  in some open region of phase space, with a uniform timescale governed by the first stability exponent  $a$ . This is just not good enough for anything to survive at the thermodynamical limit. What slowly emerged from physical considerations and numerical experiments in the Milano group is that in fact resonances can be a cause of stability. This is in our opinion an extremely important phenomenon, both from a physical and a mathematical standpoint, to which we return (again too briefly!) in §5.

Simultaneous approximation to some extent answers both problems at one shot. More precisely it solves the first question in a transparent way (see §4 below) for perturbations of convex (or quasiconvex) integrable systems. As mentioned above, these feature the overwhelming majority of the systems of physical interest. Of course one has to except here the systems which are linear and those which are not steep; in the first case, there is a well-developed simpler theory (in large part due to Birkhoff if not Poincaré), and in the second case stability simply breaks. Simultaneous approximation also helps in the investigation of general steep systems, for which a general answer is not known at present. Note that steep non quasiconvex systems do occur in nature (see §6) although they are much rarer; incidentally, this does not preclude steep systems from being generic in a precise mathematical sense, whereas quasiconvex systems alone certainly are not.

It would be too much to say that the second question has been “answered”. Simultaneous approximation has provided a clear mathematical meaning to the “stability of resonances” (which before [L1] was in large part an astonishing physical and numerical phenomenon) with a way to prove it in an effortless way. This all occurs for a fixed system and in particular a fixed number  $n$  of degrees of freedom. In other words one gets a local and most likely optimal (see §§5, 9) version of the stability theorems, again for analytic perturbations of (quasi)convex systems. These results do make it possible to pass to the thermodynamical limit for certain lattice models, at least to some extent. Let us venture to say that for these infinite dimensional systems, discrete and also continuous, simultaneous approximation has provided some of the best existing results, with again transparent and rather effortless proofs (see §7), but that there clearly remains a lot to be done.

## 2. First steps in simultaneous approximation and some basic dualities.

We recall the definition and very first results in simultaneous approximation; a standard reference is [Sc] and [L1,2] contain further information on the connection between approximation and dynamics.

**(2.1)** Let  $\alpha \in \mathbf{R}^n$  be a nonzero vector ( $n > 1$ ). Linear approximation, alias small divisors in dynamical terms, is concerned with the approximation by the integral lattice  $\mathbf{Z}^n$  of the hyperplane  $\alpha^\perp$  orthogonal to  $\alpha$ . Simultaneous approximation examines the approximation of the *line* directed along  $\alpha$ . In order to make this more precise, let us introduce the following notation:

$$\langle \alpha \rangle = \min_{\zeta \in \mathbf{Z}^n} (\max_{i=1, \dots, n} |\alpha_i - \zeta_i|). \quad (6)$$

In other words  $\langle \alpha \rangle$  denotes the distance of  $\alpha$  to the integer lattice using the norm given by the largest coordinate. In these terms simultaneous approximation studies the sequence  $\langle q\alpha \rangle$ ,  $q \in \mathbf{Z}_+$ . Note that one can also consider the continuous function  $\langle t\alpha \rangle$ ,  $t \in \mathbf{R}_+$ ; the difference is the same as between homogeneous and inhomogeneous approximations: it is essentially equivalent to approximate  $\alpha$  using the discrete sequence or to approximate the  $(n+1)$ -vector  $(\alpha, 1)$  using the continuous scheme. Roughly speaking using real instead of integral multipliers makes the dimension drop by one. Here we will go on using the discrete sequence  $\langle q\alpha \rangle$ ,  $q \in \mathbf{Z}_+$ .

The term “simultaneous approximation” comes from the fact that one is trying to approximate all the components of  $\alpha$  by rational numbers with a common denominator. Note that in any dimension  $n$ , including  $n = \infty$  in the good cases (see below), we are dealing with a line (dimension 1) whereas  $\alpha^\perp$  has dimension  $n - 1$ . From that point of view simultaneous approximation is in some sense the most natural type of approximation and indeed it is probably the most commonly used in diophantine arithmetic, but one should remark that simultaneous and linear approximation are actually extreme cases, corresponding to vertical and horizontal vectors. Other intermediate types of approximation are possible, corresponding to rectangular matrices (see [Sc]).

**(2.2)** The first result in simultaneous approximation is the following elementary theorem due to Dirichlet:

**Theorem (Dirichlet):** *For any  $\alpha \in \mathbf{R}^n$  and any real  $Q > 1$  there exists an integer  $q \in \mathbf{Z}_+$ , with  $1 \leq q < Q$ , such that  $\langle q\alpha \rangle \leq Q^{-1/n}$ .*

It turns out that one needs only apply this result in order to prove the strongest possible results in Hamiltonian perturbation concerning stability over exponentially long times. Note that the exponent  $1/n$  on the r.h.s. is optimal.

Another important and rather mysterious sequence of integers attached to a vector  $\alpha$  is provided by its best periods, which for  $n = 2$  are given by the denominators of the convergent of the associated continuous fraction. In general define the sequence  $(q_i)$  by induction:  $q_0 = 1$  and  $\langle q\alpha \rangle \geq \langle q_i\alpha \rangle$  for  $q < q_{i+1}$ . One has  $\langle q_i\alpha \rangle = |q_i\alpha - p_i|_\infty$  for some integral vectors  $p_i \in \mathbf{Z}^n$ . The vectors  $\alpha_i = q_i^{-1}p_i$  with rational components (and denominators  $q_i$ ) are called the best (Dirichlet) approximations of  $\alpha$ .

**(2.3)** It may be useful to list some of the ‘dualities’ which underly this arithmetic, geometric and dynamical situation:

- At the level of arithmetic the correspondence between simultaneous and linear approximation is given by the classical *transference theorems* (essentially due to Khintchin). We refer to [Sc] and to [Ban] for more recent results in this active area, which is related to the ubiquitous problem of finding integral points in convex bodies. In [GL] a transference result is proved in a setting tailored to the needs of Hamiltonian dynamics, namely for the so-called Bryuno vectors.
- At the level of pure geometry, transference results rest on the *projective duality* between lines and hyperplanes, much as above between  $\alpha$  and  $\alpha^\perp$ .
- In terms of ergodic theory, at least for the elementary case of linear flows on the torus (which is what Hamiltonian perturbation theory actually needs to consider), the duality is nothing but the classical one between time and space averages, with ergodic theorems to bridge the gap. To put it more precisely (see [L1,2] for more), let  $\omega \in \mathbf{R}^n$  be a nonzero vector and  $\mathbf{T}^n = \mathbf{R}^n/\mathbf{Z}^n$  be again the standard torus. Then the simultaneous approximation of  $\omega$ , and in particular the sequence of periods  $(q_i)$  as above, contains information about the approximate *recurrence times* of the linear flow along  $\omega$ , whereas small divisors measure the spatial spreading of distributions.
- Another way to express the above is simply to say that the two forms of approximations correspond to the two sides of the *Fourier transform*: simultaneous approximation is on the side of “particles” (i.e. configuration space, before applying Fourier) whereas linear approximation is on the side of “waves” (after applying Fourier). To illustrate this point, we stress again that in [L1,2] is shown a (probably optimal) theorem in Hamiltonian perturbation theory without resorting at all to Fourier transform and small divisors, which were hitherto the two basic ingredients of these kinds of results and proofs.

### 3. One-phase averaging: on the side of analysis.

The use of simultaneous approximation had the effect of drastically reducing the analytic input in the results connected with stability over exponentially long times. In fact it is enough to analyze the case of one-phase averaging, where no small divisors appear at all. This is explained in detail in [L1,2] and has lots of consequences under various circumstances, several of which

will be mentioned below. Let us very briefly review a few salient points with references.

**(3.1)** First the situation is not connected in an essential way with the symplectic character of the problem. One can in particular trace it to a note by A.Neishtadt [Ne] in which the following question is examined: Consider the differential equation

$$\frac{dx}{dt} = \varepsilon f(x, t), \quad (7)$$

where  $x \in \mathbf{R}^n$ ,  $f$  is analytic w.r.t.  $x$ , and Lipschitz and periodic with period 1 w.r.t.  $t$ ;  $\varepsilon$  is a small parameter. One seeks to eliminate the time dependence, that is to find a near-to-identity change of variable  $x = y + \varepsilon u(y, t, \varepsilon)$  such that the transformed equation would be autonomous, that is would read:

$$\frac{dy}{dt} = \varepsilon g(y, \varepsilon), \quad (8)$$

with  $g$  the transformed vector field. For a discussion of this problem and its importance as a paradigm, see in particular [L1] (Appendix 2) and [L2] (§2) (see also [LM, Chapter 8]).

Here is an informal targeted summary of some important properties. Usually, i.e. generically in the adequate topology (taking in particular domains of definition into account, which we did not mention), there do *not* exist analytic functions  $u$  and  $g$  such that (8) holds, that is such that the system becomes autonomous. On the positive side there does exist a unique solution  $(u, g)$  in the category of formal series (w.r.t.  $\varepsilon$ ) with analytic coefficients.

Write  $u = \sum_{n \geq 0} u_n(y, t)$ ,  $g = \sum_{n \geq 0} g_n(y)$  for the formal solution. The coefficients  $u_n$  and  $g_n$  are analytic functions, as mentioned above. The next fact is that, using suitable analytic norms, the norms of  $u_n$  and  $g_n$  increase roughly like  $n!$ . This was conjectured in [L1] and proved in [Sa] and then in [RS] in a more general setting. An immediate consequence is that one can truncate the formal solution to a large order  $N$  ( $N = N(\varepsilon)$  on the order of  $\varepsilon^{-1}$ ) so that the truncated *analytic* solution  $(u_N, g_N)$  satisfies (8), up to an exponentially small remainder. Namely, if we define  $y_N$  such that  $x = y_N + \varepsilon u_N(y_N, t, \varepsilon)$ , we find that:

$$\frac{dy_N}{dt} = \varepsilon g_N(y_N, \varepsilon) + O(\exp(-w/\varepsilon)) \quad (9)$$

for some  $w > 0$ ; here  $w$  stands for “width”, as it is connected in a subtle way with the width of analyticity of the data, essentially the vector field  $f$ . One thus recovers the main result of [Ne], although from a slightly different viewpoint. We refer to [RS] for a thorough discussion of the various implications, using the Laplace and Borel transforms.

**(3.2)** Another way to express the above is to say that the formal solution  $(u, g)$  lies in some Gevrey class, as investigated in [Sa]. Given a formal series:

$$a(\varepsilon) = \sum_{n \geq 0} a_n \varepsilon^n, \quad (10)$$

with coefficients  $a_n$  in a Banach space equipped with a norm  $\|\cdot\|$ , we say that it lies in the Gevrey class with index  $\alpha$  ( $\alpha \geq 1$ ) or simply that it is a Gevrey- $\alpha$  series if the coefficients satisfy:

$$\|a_n\| < CM^n (n!)^{\alpha-1} \quad (11)$$

for all  $n \geq 0$  and some constants  $C > 0$ ,  $M > 0$ . There is no universal convention for the labeling of Gevrey indices. We adopt here the convention of [MS] to which we refer for basic definitions. These definitions also appear in [L1] (Appendix 2) with a different convention, coming from a seminal paper by J.-P. Ramis; the reader will also find much more detail and further references in [RS].

A germ of function near the origin is said to be Gevrey- $\alpha$  if it has an asymptotic expansion in a germ of sector at the origin which is given by a Gevrey- $\alpha$  series. We refer to any of the papers quoted above for detail. Here we simply remark that with the present convention analytic functions are Gevrey-1 and we can rephrase the discussion in §3.1 above by saying that analytic (Gevrey-1) data lead to Gevrey-2 series (or functions, somewhat loosely speaking). This phrasing makes it plausible that starting from Gevrey  $\alpha$  data one should get Gevrey- $(\alpha + 1)$  series after averaging. This is indeed proved in [MS]. We may extract from this the following motto: *In the absence of small divisors, averaging (alias normalization) lets the Gevrey index increase by 1*. In other words the  $n$ -th coefficient of the relevant series gets multiplied by a factor  $n!$ . Note that this complex of notions is (of course!) not foreign to Poincaré: cf. *Méthodes Nouvelles*, volume 2, where he speaks about “grands multiplicateurs” in contrast with “petits diviseurs”.

We also remark that saying we are in a no small divisors situation is a kind of oversimplification, at least from the physicist’s viewpoint. What happens is that there are in fact *two* frequencies, namely 1 and  $\varepsilon$ . Remark now that a linear combination  $k + \varepsilon \ell$ ,  $k, \ell \in \mathbf{Z}$  can be small only for  $(|k| + |\ell|)$  at least on the order of  $\varepsilon^{-1}$ . For analytic data, Fourier coefficients decrease exponentially, so one will thus generate remainders roughly of size  $\exp(-w/\varepsilon)$ . This gives another rationale for exponentially small remainders in the situation of (formally) one-phase normalization (see [LM], Chapter 8 for more along this line).

Before leaving the subject we stress once more the importance of the problem above, especially now that in many situations analysis and arithmetic have

been uncoupled by using simultaneous approximation. It is akin to the problem of linearizing germs of holomorphic plane vector fields which are tangent to the identity, although more difficult to analyze. Following J.Écalle, this is because equational resurgence is more accessible than parametric resurgence. In fact there surely remain global properties to be better understood here. To be specific, consider  $f$  as nice as possible, e.g.  $f$  polynomial in  $x$  (and possibly a trigonometric polynomial in  $t$ ). Then  $u$  and  $g$  are Gevrey-2 (with the present labeling), so taking their Borel transforms w.r.t.  $\varepsilon$  one gets germs of analytic functions on the Borel transform of the  $\varepsilon$ -plane. Question: are these germs “resurgent”? In other words, can they be continued “without end”, without encountering natural boundaries?

**(3.3)** What one actually needs to know in order to tackle Hamiltonian higher dimensional problems is in some sense simpler, because one is looking for coarser information, although more parameters come into play and one needs to take into account the symplectic nature of the problem. A flexible statement is given in [L2] (Theorem 1). It certainly admits a large number of variants, some of which are discussed there. We refer again to [MS] for the case of systems with Gevrey regularity (as opposed to *analytic* data).

Here we will only mention the way one can solve the “homological” or linearizing equation (1) in that setting, bypassing Fourier transform completely (cf. [L1], Lemma 2). Returning to (1), the point is that one needs only consider *periodic* vectors. That is, far from solving (1) for a “nonresonant” vector  $\omega$ , one picks  $\omega$  such that exists  $T > 0$  with  $T\omega \in \mathbf{Z}^n$ . One can take the smallest such  $T$ , which is no other but the period of the linear flow along  $\omega$  on the torus  $\mathbf{T}^n$ . Note that at variance with §2 we do consider a real and not necessarily integral  $T$ . One can solve (1) if (and only if) the function  $g$  has zero average, which here can be taken as the *time* average. In other words assume that:

$$\bar{g} = \frac{1}{T} \int_0^T g(q + \omega t) dt = 0. \quad (12)$$

Then the solution of (1) explicitly reads:

$$\chi(q) = \frac{1}{T} \int_0^T g(q + \omega t) t dt. \quad (13)$$

The proof is two lines (integrate by parts) and this shows that  $\|\chi\| \leq \frac{T}{2} \|g\|$  for any translation invariant norm  $\|\cdot\|$  defined on the space of measurable functions on the torus (e.g. any  $L^p$  norm).

#### **4. Long-time stability of near integrable nonlinear Hamiltonian systems: global results.**

**(4.1)** The main result contained in [N] has been somewhat informally reviewed in §1.5 and is encapsulated there in the italicized sentence. Note that for

Hamiltonian systems the general averaging theorems (reviewed in [LM]) take the form of the stability of the action variables. This is because the motion of these variables is governed by the gradient  $\partial H/\partial q$  which has zero average on the torus, so that the averaged vector field vanishes identically (to all orders). Of course the Hamiltonian character of the motion also has the effect that one can devise results over much longer times. Again these results use essentially the same techniques that are put to use in ordinary KAM theory – or rather vice versa. This is to point out that they are in fact neither “more difficult” nor “easier” in essence: the major initial insight of Kolomogorov was much more about understanding what relatively slight modifications were necessary in order to show the persistence of a torus.

We will not take up here a detailed discussion of the results. It can be found in particular in [L2] and in [LMS] (§2.6) from a slightly different viewpoint. Here however is roughly the state of affairs, concerning the timescale over which the investigated systems are stable. That is we will confine ourselves to a very brief discussion of the value of the exponent  $a$  in (5). The value of  $b$ , i.e. of the *radius of stability*, is discussed in [L1] and slightly more thoroughly in [L2] (§4).

The value of this exponent  $a$  is clearly not just a technical matter: indeed it draws the limit between stability and instability, between “perturbation theory” and “Arnold diffusion”. Here we consider the global case only (which before [L1] was the only existing one), that is independently of the initial conditions  $(p(0), q(0))$  of the trajectory – see below §5 for the local case.

In [N] there appear explicit lower bounds for  $a$  in the general steep case, which however are clearly very pessimistic. As we mentioned in §1.5, when specialized to the quasiconvex case they produce a value of  $a$  which decreases like the inverse of the square of the number of degrees of freedom. Now starting from the other side, that is from the viewpoint of *instability* B.V.Chirikov had reached heuristic conclusions (in the late seventies) which pointed to the fact that (in our language)  $a$  should decrease like the inverse of the number of degrees of freedom. His remarkable discussion can be found in [C], §7 (see also [CV]). *Part of* it is taken up in [L1], §5.2 and formalized, again from the point of view of instability; since then it is often quoted and/or taken for granted, but obviously –almost– never read, which is a pity, as it still contains subtle information to be extracted (*part of* this information is discussed in [LMS], §2.6). In any case, the point here, from the point of view of stability, is that B.V.Chirikov predicted the correct value for the stability exponent of convex system, namely  $a(n) = 1/(2n)$ . That this is most likely the “correct” value means that it is most likely optimal in the sense that over longer timescales (larger values of  $a$ ) instability generically should prevail. We will hardly go into this here (see however §9) but would like to explain again how simultaneous approximation made it possible to predict

the exact same value – and prove its validity. We naturally refer to [L1,2] and references therein. J.Pöschel immediately – but a posteriori – showed (in [Pö1]) how one can retrieve this value by refining the “classical” construction of [N]. Looking at the construction it seems however difficult or even hopeless to “understand” from there why that particular value is natural, i.e. optimal. Note that it is absolutely natural that there should exist such a “classical” version, since again linear and simultaneous approximation carry essentially the same information, as shown by the various transference results. The point is that they should precisely be thought of as two facets of the same reality and that using simultaneous approximation may shed light on some important problems, to the point of making them crystal clear, whereas they may appear much more intricate using linear approximation, which sometimes (as is the case here) simply fails to provide the necessary intuition.

**(4.2)** Why should one get for convex systems a stability time which goes like  $\exp(w/\varepsilon^a)$  with an exponent  $a = 1/(2n)$  (and  $w$  some positive, system dependent constant)? Why should the “speed of Arnold diffusion” go like  $\exp(-w/\varepsilon^a)$ ? Note that it should really not be taken for granted that the answers coincide, as far as the value of  $a$  is concerned. As mentioned above, the second question is discussed in [C] and [L1], §V.2 (and also in [LMS]). A rigorous answer to the first question is of course one of the important points in [L1,2]. The meaning of the factor  $1/(2n)$  can be decomposed as follows: rewrite  $2n$  as  $2 \times 1 \times n$ ; then the factor 2 comes from the fact that  $\sqrt{\varepsilon}$  is the natural parameter; in other words  $f = ma$  is a *second* order equation. The factor 1 is for Gevrey-1, as explained in §3 (see also below); and lastly the factor  $n$  or rather  $1/n$  comes from Dirichlet theorem (§2.2). This hopefully “explains” why this value is natural, probably indeed optimal. It may also give the feeling that the connection with *instability* is not so obvious. And indeed, as explained in [L1], §V.2, one gets this *same* value from a rather different viewpoint, stemming from a discussion of the splitting of invariant manifolds and a reasoning in *linear* approximation, but in a different environment.

Quite recently, and mainly for the purpose of studying *instability*, J.-P.Marco and D.Sauzin extended (in [MS]) the stability results to the case of Gevrey systems. In the notation of §§1.5 and 3.2, if  $H$  is in the Gevrey class of index  $\alpha \geq 1$  and if  $h$  is convex, their result reads as in (5) where now  $a = a(n, \alpha)$  is given by:  $a = 1/(2\alpha n)$ . This is clearly in accordance with the above: the analytic case corresponds to  $\alpha = 1$ . Of course technical work is needed to give flesh to that prediction (see [MS]) and that work goes essentially into the analytic part, as the geometry and arithmetic do not depend on the regularity of the data.

## 5. Local results: the stability of resonances.

We come to a phenomenon which we believe is of great *physical* significance. We very briefly outlined in §§1.3, 1.4, how it is actually at the origin of the present story and we urge the reader to explore the literature quoted there and its ramifications. Below are just a few words on part of the mathematical aspects, for which the introduction of simultaneous approximation was absolutely essential (see also §7 below).

**(5.1)** The main point, when reduced to approximation, looks incredibly simple. Let again  $\omega \in \mathbf{R}^n$  be a nonzero vector and assume that in fact  $\omega = (0, \omega')$ , where  $0 \in \mathbf{R}^m$  is the null vector and  $\omega' \in \mathbf{R}^d$  is nonzero; here of course  $n = m + d$ . Then clearly the simultaneous approximation of  $\omega$  is *identical* to that of  $\omega'$ . First conclusion: one may replace  $n$  with  $d$  in the statement of Dirichlet theorem (see §2.2). Application to Hamiltonian perturbation theory: In the situation of §1.5, consider a resonance of multiplicity  $m$  ( $0 \leq m < n$ ), corresponding to a resonance surface of dimension  $d = n - m$ . Then after a symplectic linear change of variables the corresponding frequency vectors will look like  $\omega$  above. Consider an initial condition  $(p(0), q(0))$  which is close enough to the said resonance surface: then the trajectory will satisfy (5) where  $a = 1/(2d)$  i.e.  $n$  has been replaced by  $d$ . So the time of stability gets much longer, since the stability exponent increases. Details and amplifications can be found in [L1,2]; they are quite easy mathematically speaking, but before the introduction of simultaneous approximation such results had not even been conjectured (see however [C], §7 and the discussion in [LMS]). These local stability results are proved in the same way for systems of Gevrey class (see [MS]). We remark that rigorously speaking this enhancement of stability *via* resonance will be proved only after precise *instability* results are available (cf. §9 below). That is one should prove that the local exponent  $a = 1/(2d)$  is indeed generically optimal in a suitable category of analytic data. This is plausible but far from being demonstrated to-date. However the optimality of the analogous local exponents in the Gevrey class has been essentially proved in [MS].

**(5.2)** The physical importance of the above should definitely be emphasized, including for finite dimensional systems, that is in particular notwithstanding its relevance for thermodynamics. In physics (and mathematics!) resonance is usually thought of as a cause of *instability*, indeed in these essentially elliptic systems as *the* major cause of instability. This is certainly true when referring to *linear* systems, which give rise to *parametric resonance* (see [LM], Chapter 8 for an account from the point of view of Hamiltonian adiabatic invariance and stability over exponentially long times). Usually again, including in terms of engineering, the instability caused by resonance can be damped by *friction*,

i.e. passing into the dissipative, irreversible world. Here on the contrary, it is *nonlinearity*, or more specifically *convexity* which makes resonance into a cause of *stability* over exponentially long times, retaining the Hamiltonian, reversible character of the dynamics. Convexity is quite natural from a physical and mathematical standpoint, corresponding to kinetic energies and Riemannian metrics, so that the above applies to numerous classes of systems of physical interest, including in celestial mechanics (see §6.1). It also raises the question as to whether this universal feature is actually specific of convexity, that is whether or not it is a dynamical feature which is intrinsically associated to perturbations of convex integrable Hamiltonian systems (see [L1], §IV.4 for a short discussion).

## 6. Finite dimensional variations on the theme.

Here are several important variations and applications concerning the stability of finite dimensional systems. We provide hardly more than a reading guide, hoping to arouse the curiosity of the reader.

**(6.1)** Applications to *celestial mechanics*, more specifically to the problem of stability in the planetary problem, are discussed in [N]. Again simultaneous approximation has more than a word to say here, as explained in [L1], §IV.1. This was thoroughly investigated in [Ni1], which contains by far the best rigorous stability estimates in that context to-date. Local stability is also explored, including in realistic situations. It should be noted for non specialists (among which the author of these lines) that in celestial mechanics, preparing so to speak the Hamiltonian, starting from just writing it down in the proper coordinates, is a very nontrivial affair. Fortunately many generations of hard-working astronomers have laid the ground for that. Then one has to determine domains of analyticity, and this is also far from easy. Much of [Ni1] is devoted to these tasks, in particular to determining suitable domains of analyticity for the planetary problem. This is where a large part of the hard technical work goes, without direct connection to the stability problems and in fact the part of [Ni1] dealing with these questions turned out to be useful for other purposes.

We remark that the way simultaneous approximation enables one to break the difficulties into pieces, in particular to solve only 1-phase averaging problems, makes it also possible to reach much better thresholds of validity, that is much larger values of the perturbation parameter  $\varepsilon$  for which the results are valid. This is strikingly true in celestial mechanics where the systems are very intricate and is amply demonstrated in [Ni1], where numerical examples are given. Quite often they are not exactly realistic in the sense that they are not directly applicable to the actual solar system, but that could be remedied by using a computer assisted proof. In other words, one could prepare the sys-

tem by performing a certain number of steps of the perturbative scheme using formal manipulations, and only then apply the results. Since the theoretical thresholds in the results are not so low, as compared for instance with the kind of thresholds one gets in the same situation for KAM theory, the number of steps to be performed on the computer is not a priori very large – in fact it can often be on the order of just a few. We simply conclude that it is to be hoped that such computer assisted experiments – and proofs – will actually be attempted.

The meaning of the long-time stability of resonances is quite interesting to explore in terms of celestial mechanics. In some sense it is fair to say that speculations around such phenomena can be traced to Plato and closer to us to Kepler. Here are two remarks: Firstly it is a little hard in this domain not to think in an irreversible way. For instance the mechanism used in the stability proof of [N] is of irreversible nature although since we are dealing with reversible Hamiltonian systems, that mechanism should be taken only as a fictitious mathematical device. On the other hand, simultaneous approximation as used in [L1] (and many of the papers reviewed here) is of arithmetic or say geometric, rather than dynamical nature. In particular it does not really uncover any *instability* mechanism (see however [L1], §V.2 for a nonstandard picture). So although the stability of resonances is rigorously established, it is actually far from being dynamically understood, especially as far as a kind of competition with KAM stability is concerned. Here “competition” refers to the fact that KAM tori a priori live in very *nonresonant* region of phase space, and do also generate stability (see in particular [GM]). We believe these issues are really worth pursuing, including in celestial mechanics.

In [L1] (§IV.1) we mentioned interesting papers by A.M.Molchanov, dating from the late sixties and related to the above problem in the following rough form: are actual planets locked into resonance or not? These papers gave rise to an interesting controversy (see references in [L1]) but are on the whole forgotten or dismissed by specialists. It seems to us that Molchanov at least raises a very natural and in principle elementary question, which is not *per se* connected with celestial mechanics. Consider indeed the unit sphere in  $\mathbf{R}^n$  (the choice of the norm is not crucial here) and pick a point  $\omega$  on it at random, i.e. a vector of unit length. Is it likely that  $\omega$  is nearly resonant? Put this way, the question is of course meaningless and part of the problem is to find a meaningful formulation. We are actually asking for integer vectors  $k_i \in \mathbf{Z}^n$ ,  $i = 1, \dots, m$  ( $0 \leq m < n$ ) and (small) positive numbers  $\delta_i$ ,  $i = 1, \dots, m$ , such that for all  $i$ :

$$|\omega \cdot k_i| < \delta_i, \tag{14}$$

i.e.  $\omega$  is  $k_i$ -resonant to within  $\delta_i$ . This is still not a good enough formulation yet, and the point is to find plausible rules for the choice of  $(k_i, \delta_i)$  under given

circumstances, e.g. as in the case investigated by Molchanov, if  $\omega$  is actually the frequency vector associated to a planetary problem. The question is certainly relevant and natural: diophantine approximation, whether simultaneous or linear is usually preoccupied with asymptotic estimates ( $|k_i| \rightarrow \infty$ ). But what about the low-lying resonances and their relative measure? We refer to [L1] and of course especially to the papers of A.M.Molchanov for more, noting that these questions suggest numerical experiments which were not within the possibilities of computers thirty years ago but have now become perfectly feasible.

**(6.2)** Among the various possible transpositions we especially mention the stability of *elliptic fixed points* over exponentially long times. It was conjectured in [N] and was investigated also in [L1] (§IV.2), but without reaching a complete solution. The problem was then solved using simultaneous approximation in [Ni2] and also (simultaneously – no pun intended – and independently) using the classical method of [N] and [Pö1] by M. Guzzo, F.Fassò and G.Benettin (see the references e.g. in [Pö2]). The proof using simultaneous approximation was then rewritten and improved in [Pö2]. It is now transparent to the point that, as the author notes, it hardly differs anymore from the case of a near-integrable Hamiltonian. In particular it is probably optimal and lends itself easily to all kinds of variants (e.g. it could be easily transposed to the the Gevrey classes etc.).

**(6.3)** The rather technical case of perturbations of *steep not quasiconvex systems* had not been taken up since the original paper of N.N.Nekhoroshev ([N]) until recently, although in the mean time Y.S.Ilyashenko gave a nice characterization of steep *analytic* functions ([I]). This case of steep systems does occur in nature however, as demonstrated in [BFG] where the stability of a system well-known in celestial mechanics is studied. In [Ni3], L.Niederman combines linear and simultaneous approximation (i.e. the techniques of [N] – or in fact the improvement in [Pö1] – and [L1]) in order to produce a significant improvement and simplification of the results of [N] for general steep (not quasiconvex) systems.

The matter is not definitively settled though. In fact there is no conjecture at present predicting the potentially optimal stability exponent from the steepness exponents, that is connecting geometry and dynamics. It is not clear either whether simultaneous approximation can be used alone, although it does help, as demonstrated by [Ni3]. Nor is it clear whether there exist local results in the general steep case, that is whether or not long-time stability is enhanced by resonances for perturbations of steep not quasiconvex systems. This is of course just the other side of the question we asked in §5.2 above: is stability of resonances an intrinsic property of quasiconvex systems? À suivre.

**(6.4)** We finally mention a recent application of simultaneous approximation to Hamiltonian perturbation theory. In [Bl], A. Blaom studies perturbations of systems with *action-group coordinates*. Very roughly speaking the coordinates  $(p, q) \in T^*\mathbf{T}^n$  are replaced by coordinates on the cotangent bundle  $T^*G$  of a compact Lie group  $G$ . In other words the commutative group  $\mathbf{T}^n$  is replaced by an arbitrary compact (hence reductive) connected real Lie group  $G$ .

These systems have a long geometric history, which is retraced in [Bl] with references. Nowadays they are especially connected with the general theory of symplectic reduction, which also indicates that they indeed arise under very natural circumstances, depending on the symmetry group attached to the situation at hand. It seems however that [Bl] is one of the very first papers dealing with perturbation theory for such systems in a general geometric setting. Using simultaneous approximation the author obtains a stability theorem, much as in (5), but where now the role of the dimension  $n$  is played by the *rank* of the reductive group  $G$  (i.e. the dimension of a maximal torus), including in the expression of the stability exponent  $a(n)$  (see §4). Whereas simultaneous approximation is relatively easy to handle after the suitable, very interesting *geometric* preparatory work, it is not clear whether and how linear approximation could be used under the given circumstances.

## 7. Stability for infinite dimensional systems.

This is where the story started from and simultaneous approximation has turned out to be indeed quite useful in this setting. Here are some indications, again essentially of bibliographical nature.

**(7.1)** Two results about *discrete systems* can be found in [BG]. They concern chains of oscillators and rotators: the first one is about a variant of the FPU model, in connection with the foundations of thermodynamics (and quantum physics) and the recurring problem of energy equipartition. The second one deals with a chain of rotators and nonlinear localization (see also [L1], §IV.3), much in the same spirit. The authors first develop the formal apparatus necessary to deal with the infinite dimensional setting, introducing well-adapted functional analytic tools. They rely on periodic orbits, solving the homological equation as in §3.3, so that the almost trivial estimates can easily be transposed to the infinite dimensional setting. They then use Dirichlet theorem as in [L1,2] (see their §8) for the harder problem about rotators, in order to conclude. The paper contains indications and references on the physical relevance of the problems considered there (see also e.g. [BGG2] and the reference lists of its authors).

**(7.2)** The previous article ([BG]) has been the seed of further work, in particular by the first author who went on developing its framework and techniques, adapting it also to the case of *continuous systems*, i.e. to nonlinear

PDE's (see [Bam] and other papers by the same author). During the last decade of the century these equations, particularly the one-dimensional nonlinear Schrödinger equations, have been the object of intense study, trying to adapt and apply dynamical systems techniques in order to gain a better understanding of the phenomena. This includes in particular various extensions of KAM theory to the infinite dimensional case: see the works of W.Craig, S.Kuksin, J.Pöschel, E.Wayne and others.

Many of these investigations are concerned with one-dimensional equations over finite intervals (and e.g. Dirichlet boundary conditions) so that expanding the solution in spatial Fourier series, one is confronted again with a *discrete* problem, namely with the “dynamical system” defined by the collection of the Fourier modes of the solution of the original PDE. This explains why several of the techniques developed in that field work for both lattices and PDE's, although lattices have in a sense an additional structure, namely the spatial one, allowing for specific restrictions, like nearest neighbor or at least short-range interactions. Of course there also exist typical PDE techniques as for instance developed by J.Bourgain in his study of the nonlinear Schrödinger equations.

Stability results over long but finite times are complementary to KAM-type results, just as in finite dimension. The former apply to open sets (for the appropriate topology) of initial conditions, but are valid only over finite (but hopefully long!) times, whereas the latter, which are valid for all times, apply typically to Cantor sets of initial condition. Again simultaneous approximation provides the best existing results for continuous systems in terms of finite time stability. In particular, generalizing and simplifying the results in [Bam], J.Pöschel showed (in [Pö3]) that one can use simultaneous approximation quite directly, the same way it is used in the finite dimensional case (especially as in [Ni2] and [Pö2]), provided the system is first prepared in the right way, to which end he uses a technique he developed with S.Kuksin in their joint work on KAM theory.

## 8. What about KAM theory?

**(8.1)** The situation here is very far from clear which is hardly surprising, since to-date it has been only casually investigated by a small handful of people (including the author) over a short period of time. Using simultaneous approximation in a meaningful way, one can prove a theorem about the conservation of invariant tori in the style of the original result by Kolmogorov in [K] (P.Bernard and P.L., unpublished) but after transference one is actually imposing stronger diophantine conditions than the usual polynomial condition as in [K], and a fortiori the Bryuno condition as in [Rü]. So one gets a result which is weaker than the one in [K], and it is not clear why and how to improve

on that state of affairs.

Simultaneous and linear approximations in some sense carry the same information, but it is encoded in a very different way and translation (i.e. transference results) is not necessarily easy. We have already mentioned the one-way transference result in [GL], from the usual linear Bryuno condition to a nice looking simultaneous condition which in *any* dimension looks like the one-dimensional usual Bryuno condition (of course simultaneous and linear approximations coincide in that one-dimensional case).

It is interesting to note why this result is one way only. We have already noted in §5.1 that if  $\omega \in \mathbf{R}^n$  is of the form  $\omega = (0, \omega')$  (0 denoting the null vector in  $\mathbf{R}^m$  for some  $m < n$ ) the simultaneous approximation of  $\omega$  is identical to that of  $\omega'$ , in particular in the sense that the sequence of best periods (see §2.2), which generalize the denominators of the convergent of the continued fraction, is indeed the same for  $\omega$  and  $\omega'$ . The situation is not quite the same with linear approximation, which does record the resonance module (here simply  $\mathbf{Z}^m \subset \mathbf{Z}^n$ , embedded in the standard way). This very simple phenomenon nonetheless lies at the root of the stability of resonances, a phenomenon which is far less visible using linear approximation; and indeed it was detected only using simultaneous approximation, building on previous physical considerations. But that very same property becomes a problem when it comes to – say – detecting invariant tori. For instance, closing one’s eyes and looking purely at the information provided by simultaneous approximation it is not even clear a priori what the dimension of the torus should be (not restricting oneself to Lagrangian tori a priori). In other words this insensitivity of simultaneous approximation to dimension, which is a very precious property in certain important situations as we have seen, becomes apparently a hindrance in others.

**(8.2)** Let us add some more words on the theme, enlarging the setting. Hamiltonian perturbation theory could conveniently be divided into ‘finite time’ theory, ‘geometric’ theory, and ‘beyond perturbation theory’. Finite time theory, which is often especially relevant for physics found its natural framework and scope surprisingly late as we have seen (starting with [N]). Geometric (or infinite time) perturbation theory has of course been studied a lot. It is essentially about finding invariant sets for a flow defined by a Hamiltonian as in (4); this comprises in particular finding fixed points (0-tori), periodic orbits (1-tori), invariant tori of all dimensions, Aubry-Mather sets (‘cantori’), invariant manifolds attached to the above objects etc. ‘Beyond perturbation theory’ refers to exploring the limits of the theory, i.e. *instability*. As mentioned above, especially in §§4,5 this is now relatively well understood in terms of timescales.

A point that we would like to make is that there is a strong unity in the

techniques that are used in finite time (or ‘classical’) and geometric perturbation theories. In particular KAM theory, which was started in [K] (see [BGG] for a very nice pedagogical elucidation) uses basically the same techniques as finite time perturbation theory. Or rather vice versa. Kolmogorov’s original paper ([K]) is actually in line with these classical techniques, adding essentially two ideas (localizing on a torus and using a fast – Newton – iteration technique). These are of course extremely important but it is a little odd that ‘KAM theory’ is often considered nowadays as a field by itself, although the techniques that are put to use are essentially classical analytical techniques (see [A2] for an account of the prehistory of ‘KAM’).

As we have sketched in this note, simultaneous approximation has proved extremely useful and efficient in terms of finite perturbation theory, in all sorts of settings. On the other hand combining finite time and geometric perturbation theory is surely desirable and this is why we emphasize the essential community of techniques. Some aspects of this task are well-known. For instance, in the vicinity of an invariant object one can in general build a finite time theory, taking the distance to the object as the small parameter. If the invariant object is a point, one arrives at Birkhoff’s theory, which is of linear or isochronous nature. But there may be more subtle aspects, as first exemplified in [GM]. And these may well require a mixture of linear and simultaneous approximations to be properly understood. In terms of stability, one is facing the following somewhat puzzling situation. Thinking in frequency space, the very resonant zones are very stable over very long times; on the other hand the very nonresonant zones are home for the invariant tori (which by definition are forever stable) whose neighboring zones are also very stable over very long times. This determines a subtle balance and it is really not clear how these two pieces of information combine together and with instability mechanisms, possibly beyond the classical one pioneered in [A1].

## 9. Stability, splitting of invariant manifolds and instability.

Having discussed or evoked the ‘finite time’ and ‘geometric’ sides of the theory, we add but essentially one remark about the connection between instability and simultaneous approximation in the framework of the original Arnold mechanism and referring in particular to [L3] for background material.

The seed of the connection between the concepts mentioned in the title lies of course in [A1]. In [C] (§7; see also [CV] and other papers by the same authors) a heuristic argument is presented about the *speed* of drift. This argument was formalised in [L1] (§V.2) and it still represents the main new phenomenon in the study of the splitting of *higher dimensional* invariant manifolds. Let us recall it in essence – and in words. The point is that in the case of higher dimensional manifolds, secondary resonances occur; mathematically

speaking they manifest themselves via *small divisors in the denominator of the Poincaré-Melnikov integral* (see [L1], §V.2). Note that a generalisation of V.Arnold's original model was also introduced there, precisely in order to detect these small divisors which do *not* occur in the original three-dimensional model; this model has now become widely used.

Summarizing again from our viewpoint here, that is paying attention to the form of approximation at play, one can put it roughly as follows, referring to [LMS] for full results and a discussion (in particular §2.6). The time of instability is connected with the size of the splitting of certain invariant manifolds, including from a local viewpoint, that is one can fully understand the counterpart of the stability of resonances discussed briefly in §5 above in terms of splitting of invariant manifolds (see [LMS] §2 for results and proofs). One then has to connect the size of the splitting with its linear approximation: in other words one has to determine whether the Poincaré-Melnikov integral actually approximates the size of the full nonlinear splitting. To this end one can estimate from above the difference between the two quantities, namely the full and the linearized splitting. These steps have now been understood in a fairly satisfactory way; note that in this respect [LMS] (§3) introduces a really new and more powerful method, using the Hamilton-Jacobi theory. Now, somewhat paradoxically, what remains to be understood is how to estimate the *linearized splitting* (i.e. the Poincaré-Melnikov integral). This difficulty was already noted in [L1], §V.2; indeed it emerged there that B.Chirikov's heuristic argument in [C], §7, could be precisely translated into a heuristic variational argument for estimating this kind of integrals.

We stress that this is in essence a pure problem in multidimensional Fourier series, which is restated explicitly and discussed in [LMS], §5.1. And a major point here is that it is an analytic problem in *linear approximation*. More than twenty years have elapsed since [C] was written and more than ten since the writing of [L1] and in spite of progresses which have been very briefly summarized above, the evaluation of the splitting of higher dimensional invariant hyperbolic manifolds still stumbles (with the exception of a few cases reviewed in [LMS], §5) over an analytic problem in linear approximation, stemming from the occurrence of secondary resonances. It may well be that simultaneous approximation does have its word to say here. Namely one could perhaps, just as in the stability problem, look at the multifrequency problem as a limiting case of one-frequency problems, although here the limiting process is far from clear, compared to the stability problem where at least for immediate purposes it reduces to the elementary Dirichlet theorem. This suggestion is discussed in somewhat more details in [LMS] (see in particular §3.6.3). This gives another example of a class of problems where simultaneous and linear approximations should perhaps cooperate...

## 10. References

- [A1] V.I.Arnold, Instability of dynamical systems with several degrees of freedom, *Soviet Math. Doklady* **5** (1964), 581-585.
- [A2] V.I.Arnold, From Hilbert's superposition problem to dynamical systems, in *The Arnoldfest. Proceedings of a conference in Honour of V.I.Arnold for his Sixtieth Birthday*, E.Bierstone et al. eds., Fields Institute Communications Series **24**, AMS Publ., 1999.
- [Bam] D.Bambusi, Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equations, *Math. Z.* **130** (1999), 345-387.
- [Ban] W.Banaszczyk, New bounds in some transference theorems in the geometry of numbers, *Math. Ann.* **296** (1993), 625-635.
- [Bl] A.D.Blaom, A geometric setting for Hamiltonian perturbation theory, *Memoirs of the AMS* **153**, no. 727 (2001).
- [BG] D.Bambusi, A.Giorgilli, Exponential stability of states close to resonance in infinite dimensional Hamiltonian systems, *J. Stat. Phys.* **71** (1993), 569-606.
- [BFG] G.Benettin, F.Fassò, M.Guzzo, Nekhoroshev stability of L4 and L5 in the spatial restricted three-body problem, *Regul. Chaotic Dyn.* **3** (1998), 56-72.
- [BGG1] G.Benettin, L.Galgani, A.Giorgilli, A proof of Nekhoroshev's theorem for the stability times in nearly integrable Hamiltonian systems, *Celestial Mechanics* **37** (1985), 1-25.
- [BGG2] G.Benettin, L.Galgani, A.Giorgilli, The dynamical foundations of classical statistical mechanics and the Boltzmann-Jeans conjecture, in *Seminar on Dynamical Systems*, Progr. Nonlinear Differential Equations Appl. **12**, Birkhäuser Publ., 1994, 3-14.
- [BGS] G.Benettin, L.Galgani, A.Giorgilli, J.-M.Strelcyn, A proof of Kolmogorov's theorem on invariant tori etc., *Il Nuovo Cimento* **79 B** (1984), 201-223.
- [C] B.V.Chirikov, A universal instability of many-dimensional oscillator systems, *Phys. Reports* **52** (1979), 263-379.
- [CV] B.V.Chirikov, V.V.Vecheslavov, KAM Integrability, in *Analysis, et cetera*, Academic Press, 1990, 219-236.
- [GL] F.Golse, P.Lochak, Sur l'approximation rationnelle simultanée des vecteurs de Bryuno, *CRAS* **324** (1997), 1047-1052.
- [GM] A.Giorgilli, A.Morbidelli, Invariant KAM tori and global stability for Hamiltonian systems, *Z. angew. Math. Phys.* **48** (1997), 102-134.
- [I] Y.S.Ilyashenko, A steepness test for analytic functions, *Russian Math. Surveys* **41** (1986), 229-230.
- [K] A.N.Kolmogorov, On the preservation of conditionally periodic motions, *Doklady AN* **98** (1954), 527-530.

- [L1] P.Lochak, Canonical perturbation theory via simultaneous approximation, *Russian Math. Surveys* **47** (1992), 57-133.
- [L2] P.Lochak, Hamiltonian perturbation theory: periodic orbits, resonances and intermittency, *Nonlinearity* **6** (1993), 885-904.
- [L3] P.Lochak, Arnold diffusion: a compendium of remarks and questions, in *Hamiltonian systems with three or more degrees of freedom*, C.Simó ed., Kluwer Acad. Publ., 1999, 168-183.
- [LM] P.Lochak, C.Meunier, *Multiphase averaging for classical systems*, Appl. Math. Sciences Series **72**, Springer Verlag, 1988.
- [LMS] P.Lochak, J.-P.Marco, D.Sauzin, On the splitting of invariant manifolds in multidimensional near-integrable Hamiltonian systems, Preprint, 1999, 120 pp.
- [MS] J.-P.Marco, D.Sauzin, Stability and instability for Gevrey quasi-convex near-integrable Hamiltonian systems, Preprint, 2001.
- [N] N.N.Nekhoroshev, An exponential estimate for the time of stability of nearly integrable Hamiltonian systems, *Russian Math. Surveys* **32** (1977), 1-65.
- [Ne] A.I.Neishtadt, On the separation of motions with a rapidly rotating phase, *J. Appl. Math. Mech.* **48** (1984), 133-139.
- [Ni1] L.Niedermaier, Stability over exponentially long times in the planetary problem, *Nonlinearity* **9** (1996), 1703-1751.
- [Ni2] L.Niedermaier, Nonlinear stability around an elliptic equilibrium point in a Hamiltonian system, *Nonlinearity* **11** (1998), 1465-1479.
- [Ni3] L.Niedermaier, Exponential stability for small perturbations of steep integrable Hamiltonian systems, to appear in *Ergodic Theory and Dynamical Systems*.
- [Pö1] J.Pöschel, Nekhoroshev estimates for quasi-convex Hamiltonian systems, *Math. Z.* **213** (1993), 187-216.
- [Pö2] J.Pöschel, On Nekhoroshev estimates at an elliptic equilibrium, *Internat. Math. Res. Notices* **4** (1999), 203-215.
- [Pö3] J.Pöschel, Nekhoroshev estimates for a nonlinear Schrödinger equation and a theorem by Bambusi, *Nonlinearity* **12** (1999), 1587-1600.
- [Rü] H.Rüssmann, On the frequencies of quasiperiodic solutions of analytic nearly integrable Hamiltonian systems, in *Seminar on Dynamical Systems*, Progr. Nonlinear Differential Equations Appl. **12**, Birkhäuser Publ., 1994, 160-183.
- [RS] J.-P.Ramis, R.Schäfke, Gevrey separation of fast and slow variables, *Nonlinearity* **9** (1996), 353-384.
- [Sa] D.Sauzin, Caractère Gevrey des solutions formelles d'un problème de moyennisation, *CRAS* **315** (1992), 991-995.
- [Sc] W.M.Schmidt, *Diophantine Approximation*, Lecture Notes in Math. **785**, Springer Verlag, 1980.