6. L.Chierchia and G.Gallavotti 1994 (Annales IHP):

This paper is quite long (140 pages) but we shall be brief, keeping in store loads of observations, not all minor. As explicitly written by G.Gallavotti, sections 1-8 are preliminary. Much of section 1-7 is devoted to redoing from scratch, often without the proper citations, what has been done in other places. There is a rather long list of authors and works on variants of the KAM theory in a partially hyperbolic situation which would have deserved quite specific references. We postpone a comment on section 8, which is also related to article 9 below.

Section 9 contains definitions and the main assertion of section 10, namely the existence of "large homoclinic angles" is wrong. Specifically, Assertion d) of Theorem 3 turned out to be wrong; it was allegedly proved in Appendix 13, consisting of a long, essentially uncheckable computation. The fact that something there is not plausible should have been *a priori* clear from the partly heuristic connection between the size of the splitting and the "speed of diffusion", as discussed by B.V.Chirikov (1979) and formalized (but not rigorized) in [RMS]. The fact that it took so long to spot this implausibility is essentially due to the size of the paper and the way it is written, which make it extremely hard to detect such crucial conceptual gaps.

Sections 11 and 12 depend on section 10 and are thus invalidated as well. With more work, it seems that the existence of drift in the celestial mechanical model discussed there can and will be proved (see 10 below). Such a result would present an interest from the point of view of celestial mechanics but will *not* show the existence of diffusion (or drift) for the perturbation of an honest "*a priori* stable" and nondegenerate integrable Hamiltonian. Simply because the model studied in Sections 11-12 is not of that kind and presents some specific and essentially simpler features.

Note that with a method as simple as the one employed in section 8, the transition times that one finds are essentially forced to be at least exponential (see below equation (8.21)). However it is expected (see [RMS] and/or [Comp]) that the drift in such *a priori* unstable systems is generically *polynomially* slow in the small parameter. This is likely to hold true in this case (but has not been proved). At any rate, the exponentials appearing in section 8 are simply an artifact of the method used; note that at the beginning of the paragraph containing formula (8.22) it is acknowledged that they constitute at best a convenient arbitrary choice. They cannot possibly be connected to any version of "Nekhoroshev's theorem", which certainly does not apply in the *a priori* unstable case at hand.

9. G.Gallavotti 1997 "Fast Arnold diffusion in isochronous systems" (preprint, September 1997):

The comments below apply to the version of this paper deposited in the University of Texas electronic archives in September 1997. The first eight lines are very difficult to interpret. In particular, one reads that the geometric (nonvariational) method employed in section 8 of article 6 above "maintains its interest in spite of the better estimates coming from variational methods because it is the only one which, so far, is robust enough to apply to anisochronous systems". However, precisely the variational papers quoted by the author in this very paragraph deal with anisochronous systems, namely Arnold's original system.

In section 2 we learn that G.Gentile's 1995 paper is "the basic paper" on KAM for 1-hyperbolic tori. Note that in [Ge], it is not clear that the proof is complete and indeed we learn in the present paper that "a proof is essentially" in [Ge]. In any case it applies only to a rather particular case. One should recall that the old-fashioned "classical" proofs are complete, were published long ago and apply to much more general situations.

In the conclusion of section 3, the author stresses "the conceptual difference with respect to the variational approach, which accounts for the impressive difference" for the bounds on the transition times. Furthermore, in section 5 we read: "The result is an extremely large diffusion time T (namely exponential in $\mathcal{N}$). *Nevertheless the estimate that comes out of the above scheme seems essentially optimal.* And then the problem is: 'how is it possible that by other methods (e.g. variational methods of [Be],[Br]) one can get *far better* estimates?' A reason may be that the variational methods are less constructive: less so than above."

The use of the term "optimal" in this passage is misleading. It would be more accurate to say: with this simple direct approach one cannot do better. Furthermore, the reasoning suggested here and developed at length in the article to explain the difference in diffusion times is not correct. The variational techniques in [Be], [Br] do keep track of the location of the constructed orbits, contrarily to what the author asserts several times. Moreover, geometric methods have been pushed beyond the "[CG]-method", yielding techniques and results that are not drastically different from the variational ones.

Section 6 contains the original result of the paper (theorem 2). The two page justification is quite hard to follow (beware of the unusual definition of μ) and cannot be considered as providing a formal proof. But one can make the following possibly important observations which if correct simply invalidate the paper: It is not clear from the last line of the paragraph how one gets the upper bound on T_i, nor even how to read this expression. Apparently (taking the exponent i to be inside the log), T_i is supposed to increase linearly with i, producing the stated bound on the drift time (because $\sum_{i=1}^{N} i = O(N^2)$). This is hardly plausible and in sharp contrast with what intuition suggests and with what actually happens in the variational papers. It implies in particular that as μ (or ε) tends to 0, one constructs orbits whose starting points do not tend to the first torus; indeed the ergodization time E' and the time T_1 remain bounded independently of μ.

Lastly a word on appendix 2. The simple reasoning in [G1] proves a simple estimate for the filling time, with exponent $\tau + d + 1$. Here too, it is likely that this estimate is "optimal", given the method. But in fact, using a more refined approach, S.Dumas improved the exponent to $\tau + d/2$ (J. Dyn. and Diff. Eq. 1991), and improving further on that, J.Bourgain, F.Golse and B.Wennberg produced the probably optimal answer as far as the exponent is concerned; it is equal to just τ (see Commun. Math. Phys. 190, 1998). All this is in sharp constrast with the passing mention "see [BGL] (sic) for an alternative proof".

Nota bene: Subsequently and because a first version of the present review was circulated, another version of this preprint has been posted *bearing the same date* (September 9th 1997) as the old version. Most remarks made above still apply and the texts are almost undistinguishable, bearing in particular the same title and same date. Yet the second version does not undertake to prove a *polynomial* but only an *exponential* estimate for the instability time. In particular, the first lines of section 6 and the paragraph containing (6.5) have been slightly modified; especially in the line preceding formula (6.5), the quantity N^2 has been replaced with 2^N: a substantial mathematical difference. This paper was written in order to illustrate and improve on the "[CG]-method", i.e. the direct geometric estimates of section 8 in article 6 above. Section 6 of the present paper is supposed to "add a new idea". In the present state, it does not present a mathematical proof of what is *now* (i.e. in the second version) asserted; moreover what is asserted is weaker than what has already been rigorously proved via other methods.

10. G.Gallavotti, G.Gentile, V.Mastropietro, " Separatrix splitting for systems with three timescales" (submitted to Commun. in Math. Phys. in October 1997; hereafter [GGM1]), and "Hamilton-Jacobi equation, heteroclinic chains and Arnold diffusion" (first deposited in the Texas archives in December 1997; hereafter [GGM2]):

These two manuscripts deal with a special and relatively simple case of the multidimensional splitting problem, with a view of using the results to study a (degenerate form of) Arnold diffusion, in particular to prove the existence of diffusion in the example of Sections 11-12 of article 6 above (they do not achieve this goal though). This latter aspect is the subject matter of [GGM2] and it is a little odd that the material has been spread out over two articles (apparently [GGM2] was not meant for publication when first written up; but it was deposited in the archives).

First the fact that three timescales are present precludes in effect the presence of small divisors (as mentioned in passing on l.15 of p. 23 of [GGM1], although this is one of the crucial points here). This fact is of course well-known and has been used by many authors over the years: frequencies which are not of the same order of magnitude cannot possibly "resonate". This says that we are not facing a truly multidimensional (or multifrequency) problem from the point of view of perturbation theory, inasmuch as arithmetical difficulties are absent. From the point of view of "diffusion" it can be considered as a local problem modelling the vicinity of a double resonance in a 3 d.o.f. near integrable Hamiltonian where two simple resonances meet, with one being "weak" and the other "strong".

The application of KAM theory in this setting is well-known. Especially, theorem 1 presented in [GGM1] is not new, including integrability on the perturbed invariant manifolds (cf. especially works by S.Graff and

D.Treshev). The existence of "many" invariant tori together with the attending invariant manifolds is addressed again in [GGM2], leading to theorems 2 and 3 there. These results are not new either: the idea of combining higher order averaging (normal forms) with KAM theory is actually quite old. Its first appearance may perhaps be traced to a paper by A.Neishtadt (J. Pure and Applied Mech., 1981) in which one can find in essence the results presented here, the proof using iterative methods being quite short and simple. One may also look at a 1984 paper by the same author published in the same journal. More recently, independent works by several authors, notably A.Morbidelli and A.Giorgilli, L.Niederman as well as A.Delshams and P.Gutierrez have pushed related ideas further on (of course the works of these authors do not simply repeat each other although there is a certain overlap in the results and techniques). We note in passing that the reference list of [GGM2] comprises 12 items, of which 11 are purely "internal" (note also that even the title of [GGM1] is wrongly quoted there, in a significant way...) and the last one refers to the work of L.H.Eliasson which originally inspired the development of these "direct methods".

Returning to [GGM1], the bulk of it is devoted to proving theorem 2, that is to vindicate the "naive" Melnikov computation. Before analysing it, it should be recalled that the hitherto most successful track in terms of getting lower bounds for the splitting distance or angle was initiated in a remarkable 1984 paper by V.Lazutkin. This paper deals with the two dimensional case, more precisely the case of the standard map, which is significantly more difficult than the now standard example of the rapidly forced pendulum (or variants of it). The results for the standard map are not complete there but part of the method was then successfully applied to Hamiltonian flows (as opposed to maps) with small parameters (as opposed to *without* a small parameter). It turned out that this part does extend to higher dimensions. The first important result is precisely to bound the difference between the splitting angle and the result of the linear (Poincaré-Melnikov) computation, and the key point is a direct analog of the extension lemma which was first used by V.Lazutkin. In turn, the underlying idea about the usefulness of this extension lemma is contained in a "trivial" (yet crucial!) lemma in elementary harmonic analysis (or one-dimensional complex function theory; it amounts to lifting a contour of integration) which says that if one can get a uniform bound for an analytic function over a "wide" horizontal strip centered on the real axis, one gets a much better bound on the restriction of the function to the real axis. A recent paper by A.Delshams, V.Gelfreich, A.Jorba and T.Seara ([DGJS]; now published in CMP, 1997) uses this method in the multidimensional isochronous case. Note that the "trivial" lemma remains one-dimensional, as it concerns the plane of the complex time. A paper by M.Rudnev and S.Wiggins ([RW]) dealing with the anisochronous case and using the same technique has appeared in Physica D (1998) but as it stands contains a mistake in its theorem 2.1 (which also invalidates theorems 2.2 and 2.3 as stated). An erratum will be shortly issued by the authors and a large part of the statements can hopefully be recovered. It is quite important to insist that both papers ([DGJS] and [RW]) do confront (to various extents) the arithmetic questions inherent to the problem, as was first explained in [RMS]. Specifically, if the perturbation is polynomial, as in the original example given by Arnold and as in [GGM1,2] no small divisors enter in the denominators of the Melnikov integrals, nor in the higher order terms of the perturbation series. As a result, the analysis is greatly simplified and this case (corresponding to a resonance of maximal order) can be viewed essentially as a parametrized one-frequency case. In particular, evaluating the Melnikov function does not pose any difficult problem and the exponent in the exponentially small splitting is the same as in the one-frequency case. It is independent of the dimension (which in [GGM1,2] is the lowest possible). These phenomena underly the analysis in [RMS] and this is precisely why the original Arnold model was generalized in that paper. Indeed it was recognized there that in the generic case of a perturbation with infinitely many harmonics (and no gaps in the Fourier series), small divisors (corresponding to secondary resonances) appear in the denominator of the Melnikov function, making its evaluation much harder. But this is also a much more interesting case as it was found that the splitting exponent should then coincide, via a key heuristic computation appearing in [RMS], with the optimal stability exponent derived there. At present the rigourous evaluation of the Melnikov function under these generic conditions has been performed in [DGJS] in a very special 3D-case and in §6 of [RW] (which is self-contained and independent of the error in theorem 2.1) for the general 3D-case. Higher dimensional situations have not been touch and seem difficult to attack, for lack of a higher dimensional analogue of continuous fractions.

Returning to theorem 2 in [GGM1] it seems that the isochronous case (Hamiltonian 2.1) can be treated as in [DGJS]. The only case which is not formally covered in the existing literature (taking into account the error in [RW]) would thus be the anisochronous variant of theorem 2. But actually the main point again is that the cases investigated in the paper do not really address the difficulties which are specific to the multifrequency case, as explained above. It should also be added that it is unclear whether the arguments presented in [GGM1] (§6 and §7) can be viewed as a complete "proof" of theorem 2. It seems that it would constitute a formidable task to try and give a complete self-contained version of the arguments presented there, especially in the anisochronous case discussed in §7 (see many examples of appeal to outside material in these two sections or to the fact that one could "follow the same path" as somewhere else).

Applications of the results to Arnold diffusion, more properly to the construction of heteroclinic chains (the authors do not take up the subject of constructing shadowing orbits, which is a separate issue) appear in §8 of [GGM1] and §6 of [GGM2]. The "easy case" of isochronous (hence gapless) systems which is dealt with in §8 of [GGM1] is really a toy problem and the application is essentially obvious, as the authors themselves seem to imply. The application to anisochronous systems in [GGM2] leading to theorem 4 there amounts to the observation that the splitting is independent of the diophantine condition under the given circumstances (widely separated timescales, polynomial perturbation); see (ii) on p.10 of [GGM2]. This is a nice elementary remark, which under the given circumstances provides a way out of the the pervasive problem of "bridging the gaps", but one should recall that this is indeed a simplified setting, whose investigation does not really help attacking the core of the matter. In some sense, the situation is quite reminiscent of the original example of Arnold, in which the possibility of applying an implicit function theorem relies on the same fact. More justification would actually be needed in the example given by Arnold; it is partly provided in a recent paper by P.Perfetti.

As mentioned above, these papers still do not fix the mistakes in article 6 above. More precisely, theorem 2 in [GGM1] proves that essentially the opposite of what is stated in section 10 of 6 is true. Namely the splitting is exponentially small and the linearized part gives the correct answer. Without the large angle theorem at hand, proving the existence of drift in the example of sections 11-12 of article 6 is not so easy, but perhaps can be achieved (cf. the rather hazy discussion in §8 of [GGM2]). Again though it would still not provide a "real" example of Arnold diffusion i.e. one which occurs when perturbing a genuine "*a priori* stable" integrable Hamiltonian. But it would be interesting as an example drawn from the "real world".

11. Papers on "twistless tori"; in particular "Twistless KAM tori etc." by G.Gallavotti, Reviews in Mathematical Physics 1994 (hereafter [G3]), "Twistless KAM tori" by G.Gallavotti, CMP 1994 (hereafter [CMP]) and "Non recursive proof of the KAM theorem" by G.Gallavotti and G.Gentile, Erg. Th. and Dyn. Syst. 1995 (hereafter [ETDS]):

The first remark to be made is quite simple: none of the papers above contains a result nearly as general as the one derived in the Appendix below. More precisely, in the notation of proposition 1 of the Appendix, the above papers address the cases where $g = 0$, f is independent of I_1 and is indeed an even trigonometric polynomial in the angles ϕ. The name "Thirring model" was coined by the authors to refer to this special case. We add that the appendix gives a proof which is complete, granted only the original paper by Kolmogorov and the detailed elementary estimates given in a later paper (see the appendix for reference). Moreover the pace in the appendix is leasurely and the core of the proof is hardly two-page long.

Taking this remark into account and the words of the authors of the papers at their face value, we shall succinctly address the following questions: How do "direct mehods" relate to "classical" (i.e. iterative) methods? What is the role of "physics" in these matters? It goes without saying that we do not pretend to actually answer these questions; we shall be content with providing some elements and leave it to the reader to draw conclusions if she or he feels the need to do so.

As to the first question, we first recall that it all started with the work of H.Eliasson who proved in a "direct" way that the Linstedt series describing the quasi-periodic solutions (invariant tori) in a KAM-type situation do indeed converge. His proof was a tour de force and several people (especially the authors of the papers reviewed here) undertook to understand and clarify it, giving several expositions of related results, including the ones reviewed here. What is unclear is to what extent the twistless property (not mentioned in H.Eliasson's work) was supposed to be a *new* result, not accessible to iterative methods (the appendix makes it clear that it *is* quite accessible and even easy).

In [G3] we read (p.346, remark 2): "If some of the inertia moments J_i [corresponding to μ^{-1} in the notation of the appendix] are $+\infty$ the above theorem is an easy consequence of the classical KAM theorem (or better of its proof)". This statement corresponds to the case $\mu = 0$ in the appendix and correctly states that the iterative methods do recover this case. The quotation then continues with: "However the bound b [radius of convergence] obtained via the classical proof depends on the twist rate, i.e. on the maximum among the J_i *which are not* $+\infty$ [underlined by the authors], and diverges as the rate approaches 0. This non uniformity is quite surprising but it is an artifact of the classical proof (as a direct careful analysis of the latter also shows)." This seems to imply that the twistless property can be seen only via direct mehods, an incorrect statement in view of the appendix. We apologize to the reader for dwelling too long on this discussion but it seems to us that this is quite an important point: Have new results been proved using direct methods? Have classical results been recovered in their full generality using direct methods? The answer to both questions is simply "no". Returning to the twistless property we look as the author suggests at page 363 in [G3]. There again one should distinguish between the "no-twist" case ($\mu = 0$) and the "twistless" case (i.e. $\mu \to 0$; it would be more appropriate to call it the "vanishing twist" case). The former case is trivial when the perturbation is action independent, in particular a trigonometric polynomial (the "Thirring model"); this takes care of the remarks around (4.16). Concerning the twistless case we read (p.363): "The convergence of the formal series for the tori equations studied in Sec. 7 will yield a radius of convergence b^{-1}, independent of the J_i (a result stronger than the usual KAM theorem relying on the twist property). Hence we show, by direct bounds, that the just posed twistless KAM problem has a solution (a fact that could be checked also by a careful examination of some of the classical proofs of KAM theorem, as mentioned in the introduction)." Again it is quite hard to interpret this quotation, as the two parenthetical remarks seem to contradict each other (at least in spirit) and as in the introduction we learned (as quoted above) that the "non uniformity is quite surprising but is an artifact of the classical proof". Apparently the second parenthetical remark says that the "artifact" *could* be repaired; the appendix to the present paper simply shows that it does not exist. In any case the conclusions of this rather intricate discussion are very briefly summarized in subsequent papers. In [CMP] we simply read (first page of the paper): "The results will be uniform in T [$= \mu$ of the appendix] (hence the name "twistless": this is not a contradiction with the necessity of a twist rate in the general problems)." We leave it to the reader to appreciate how much novelty is claimed here and we apologize again for this perhaps unnecessarily detailed discussion, but the material is indeed quite abundant, perhaps strikingly so compared to the mere three pages of the appendix below.

In fact we also have to briefly address an even more down-to-earth question: do papers using direct methods always *prove* the statements they contain? Again we shall not answer the question but simply give elements drawn from [G3]. This paper is 70 page long and contains two results, labeled as theorems 1 and 2. We are here interested in the first one which is actually called "twistless KAM". As mentioned already, the setting is that of the "Thirring model". After 40 pages of preparation the reader comes (in §7) to the "proof" of theorem 1. Strangely enough the first thing she or he is told is that "this section has heuristic nature". Then one learns the bad news that one has to impose an extra arithmetical condition (condition (7.1)) which was not mentioned in the statement of the theorem. This "strong" diophantine condition is discussed again in [CMP] and [ETDS] and it finally emerges that it is superflous. Yet the conclusion of all this in [ETDS] is quite curious. The main goal of the paper is to show that "in fact such an hypothesis can be relaxed" (i.e. the strong diophantine condition is unnecessary; note that of course it does not appear in the appendix to the present paper). Yet the authors conclude (first page, second paragraph): "In our opinion this shows that a hypothesis like the strong diophantine condition, or something similar to it, is very natural, as it simplifies the structure of the proof [...]." So the assumption is both superfluous and natural; why not conclude that the proof is not "natural"? Returning to [G3], he "heuristic" section 7 closes with the promise of a "complete analysis" in Appendices A3 and A4. It does not leap to the eyes that Appendix A3 is less heuristic than section 7: it contains one clear statement, namely Bruno's lemma, which is reproved in Appendix A4 and indeed constitutes the entire content of the latter appendix. The rest of Appendix 3 consists in a discussion which is quite hard to follow and leads to a semiformal statement italicized at the bottom of p.407. It seems that armed with this statement, one should be able to conclude the proof of the statement of theorem 1, supplemented by the unexpected arithmetical condition (7.1). This analysis could be pursued in much more detail and would apply to many other papers. This is particularly the case for the "proof" of the anisochronous case in [GGM1] section 7 (see item 10 above), which closes with an appeal to the techniques of [G3].

Now a few words about the role of "physics". Note that the remarks which follow are in part subjective and personal and should not be put on the same footing as the many mathematical *facts* we have tried to gather in these reviews. In the papers we are discussing, as well as in many others, the *vocabulary* of physics frequently enters: Thirring model, renormalizable quantum field theory, ultraviolet divergence, Feynmann integrals etc. It is certainly not easy to understand how far this is *technically* helpful, i.e. how much of the *knowhow* of these fields actually enters in a meaningful way. For instance, as we saw already, the term "Thirring model" is simply shorthand for "trigonometric polynomial". In the classical iterative proof, looking at this particular case does not help; it simply obscures the concepts, and this may be why proposition 1 of the appendix was not stated and proved before. In the attempts at using renormalization techniques in order to prove *more* than what the iterative methods yield, it is natural (but that could be misleading too) to start with such a simple model. Indeed D.Escande for example used an even more particular model (the "two-wave Hamiltonian" which is a particular case of the "Thirring model") and derived very interesting properties, most of which have remainded beyond the possibility of a rigorous proof (by *any* method). Other researchers like B.V.Chirikov have of course taken a decidedly "physical" path and gained some precious insights which, on the other hand, are very difficult to prove or disprove. Now in the articles under scrutiny here, no such "physical statements" are present and the goal is to prove mathematical statements which are in fact quite within reach of ordinary mathematical techniques. Rigorous versions of renormalization techniques still do not exist in this setting but certainly then it would be interesting to draw from "physics", proving statements which "mathematicians" just did not think of or had no tools to vindicate. This seems hardly to be the case here.

Concerning field theory, we simply add the following. Hard graphical enumerative techniques have been developed in mathematical field theory. In particular, the techniques developed by C.Itzykson and coworkers enabled them to derive some aymptotic estimates which were instrumental in the proof of the Witten's conjectures by M.Kontsevitch (see his paper in CMP, 1992, section 3). There some key features are present: namely one considers graphs which are *not* of genus 0, *a fortiori* have non trivial homotopy (are not trees) and the automorphisms of the object are one of the main sources of difficulty. In the papers under review here, none of these features is present, and it is unclear what is gained by the translation of elementary graphical enumerative arguments into "field theory". We close with a quotation from [GGM1] (see item 10), first page of section 6: "In the present case the graphs will be, topologically, trees: very unusual graphs from the point of view of field theory, where loops are often the main source of interest and non triviality. On the other hand the graphs have nodes with arbitrary large coordination number: also unusual in quantum field theories (with polynomial interactions)."

Appendix: twistless invariant tori in nearly integrable Hamiltonian systems

In this note we prove a result about the preservation of tori in nearly integrable Hamiltonian systems, which are "twistless" in the sense of [G1], [G2]. The result follows from a variant of the scheme originally proposed by A.N.Kolmogorov in [K], which has been nicely detailed in [BGGS] "with a mostly pedagogical intent". We shall follow the latter paper, which itself closely follows [K], including in the notation; we use Lie transforms as in [BGGS] as they appear to be somewhat more convenient than the canonical transform formalism of [K]. Because by now (1998) this or essentially equivalent algorithms have been repeated hundreds of times we shall be somewhat sketchy in the exposition but it should be clear that the proofs below are complete, granted only [K] with the estimates spelled out as in [BGGS].

Let us briefly and incompletely recall the well-known setting; for details we refer to [BGGS]. The only minor differences with this paper (and [K]) are notational. Here we shall use (I,ϕ) rather than (p,q) for the action-angle variables and in fact we need to divide these into two groups. So let $n = n_1 + n_2$ be the number of degrees of freedom, and let $(I_k, \phi_k) \in \mathbf{R}^{n_k} \times \mathbf{T}^{n_k}$ $(k = 1, 2)$; we write $I = (I_1, I_2) \in \mathbf{R}^n$, $\phi = (\phi_1, \phi_2) \in \mathbf{T}^n$, and we adopt a similar notation for all the intervening quantities, which formally live in the tensor algebra of the phase space (I,ϕ). We use a rather concise notation for these tensorial (scalars and matrices in fact) quantities but that should cause no ambiguities. For example, if ω is a frequency vector and I an action vector, we write indifferently ωI or $\omega \cdot I$ for the ordinary scalar product of ω and I; in the same vein, CI^2 should be read as $CI \cdot I$, with C a square matrix. The result reads as follows:

Proposition 1: Consider the nearly integrable Hamiltonian:

$$H(I,\phi) = \omega I + \frac{1}{2} C_1 I_1^2 + \frac{1}{2} \mu C_2 I_2^2 + \varepsilon f(I_1, \phi) + \varepsilon \mu g(I, \phi),$$

where ε and μ are real parameters. Assume that the symmetric matrices C_1 and C_2 are invertible, that the functions f and g are defined and analytic near $I = 0$, and that the vector ω is diophantine ($|\omega \cdot k| \geq \gamma |k|^{-\tau}$ for $k \in \mathbf{Z}^n \setminus \{0\}$ and some $\gamma > 0$, $\tau \geq n - 1$).

Then for $|\mu| \leq 1$ and $|\varepsilon| \leq \varepsilon_0$ with $\varepsilon_0 > 0$ independent of *μ, there exists an invariant tori of frequency ω for H; it is ε-close to $I = 0$ (again independently of μ) and the flow on it is conjugate to the rotation with frequency ω.*

Remarks: There are several variants of the statement which could be proved in the same way. Here f might depend on ε and g on ε and μ, in which case one requires analyticity in ε (but not necessarily in μ). For rather trivial reasons we need to work on a bounded interval in μ so we assume w.l.o.g. that it is smaller than 1.

For a given $\mu \neq 0$ this is essentially Kolmogorov's statement. Not quite because we did not want the statement to be too cumbersome but that would be easy to fix. Namely we could consider a Hamiltonian $H(I, \phi, \varepsilon, \mu)$, analytic in (I, ϕ, ε) and C^2 in μ say, and give the appropriate conditions on its partial linearization at $\mu = 0$ which would lead to a seemingly more general statement than the one above and would recover the usual statement for fixed nonzero μ. Note however that proposition 2 below does reduce for fixed $\mu \neq 0$ to the parallel statement in [K] (see theorem 2 in [BGGS]). For $\mu = 0$, one is dealing with the

ω_2-quasiperiodic perturbation of a fully nonlinear integrable Hamiltonian. The statement says that the limit $\mu \to 0$ is actually *regular*, although the partial twist matrix μC_2 vanishes (hence the name "twistless tori").

Another remark is that we could consider μ as an n_2-vector i.e. deal with "multiple actionscale" Hamiltonians, rather than the two actionscale case of proposition 1. This amplification would come essentially from a proper reading of the formulas: μ would be an n_2-vector, μI_2 would mean componentwise multiplication (*not* scalar product), $\mu C_2 I_2^2$ would mean $(\mu I_2 \cdot C_2 I_2)$ etc.

Finally, we note that most of the modifications and observations which have been made during the past 40 years apply to this case. We shall very briefly discuss a few well-known issues at the end.

In order to prove proposition 1, one first recasts (as in [K] and [BGSS]) the Hamiltonian into Kolmogorov normal form. Indeed proposition 1 is an immediate consequence of

Proposition 2: Consider the Hamiltonian:

$$H(I,\phi) = \omega I + \frac{1}{2}C_1(\phi)I_1^2 + \frac{1}{2}\mu C_2(\phi)I_2^2 + \varepsilon A(\phi) + \varepsilon B_1(\phi)I_1 + \mu\varepsilon B_2(\phi)I_2 + Q(I_1,\phi) + \mu R(I,\phi),$$

where ω is diophantine, the averages C_1 and C_2 of the symmetric matrices $C_1(\phi)$ and $C_2(\phi)$ over $\phi \in \mathbf{T}^n$ are invertible, $A(\phi)$ has zero average and Q (resp. R) is of order I_1^3 (resp. I^3). Here all the intervening functions are supposed to be defined and analytic in their respective arguments near $I = 0$. The conclusion is the same as in proposition 1.

Remark: In both the above propositions, we chose (in contrast with [K] and [BGGS]) to make the dependence on the paramaters ε and μ explicit, simply because there are two of them, and we hope this is typographically helpful. Yet we have dropped this dependence from the functions and we hope that the reader will mentally restore it. In fact, since the proof is iterative, one can and needs to assume that the functions may depend on the parameters.

As noted above, the proof of proposition 2 is almost identical to its analog in [K] (and [BGGS]). So we shall be sketchy and mention only the few points of difference.

First write $H = H^0 + \varepsilon H^1$ where $H^1 = A + B_1 I_1 + \mu B_2 I_2$ is the perturbation which we want to eliminate recursively. Let χ be the auxiliary Hamiltonian (the analog of the generating function in terms of Lie series) by means of which we shall perform one step of the perturbation scheme. Here we choose (again essentially as in [K]):

$$\chi(I,\phi) = \xi \cdot \phi + X(\phi) + Y_1(\phi)I_1 + \mu Y_2(\phi)I_2,$$

where $\xi \in \mathbf{R}^n$, the scalar function X and the vector functions Y_1 and Y_2 (of sizes n_1 and n_2 respectively) have to be determined. Note that for $\mu \neq 0$ the I_2-component ξ_2 of the mean translation of the torus is *a priori* of order 1 (not μ).

We perform one step of the perturbation scheme; the convergence will follow exactly as in the usual case (cf. [BGGS] §5). Denote by L_a the Liouville operator associated to a function a; that is $L_a(b) = \{a, b\}$ is the Poisson bracket of a with a function b. We modify H by taking the time ε of the flow of χ; so we get the new Hamiltonian $H' = \exp(\varepsilon L_\chi)H$ and as usual we have to pick χ so as to make the perturbative term smaller. We keep the same names for the variables when refering to H' (this is one of the convenient features of Lie series; see e.g. [BGGS] §3.1) but the new functions will get "primed" names (A becomes A' etc.). Again here we closely follow §4 of [BGGS] which completely explicits a few lines in [K].

The function χ is determined by solving the "homological equation", namely here by requiring that:

$$H^1 + \{\chi, H^0\} = cst + O(I_1^2) + \mu O(I^2).$$

So we compute the left-hand-side:

$$\begin{aligned}H^1+\{\chi,H^0\} &= -\xi\cdot\omega-\omega\frac{\partial X}{\partial\phi}+A(\phi)\\ &+[B_1(\phi)-C_1(\phi)(\xi_1+\frac{\partial X}{\partial\phi_1})-\omega\frac{\partial Y_1}{\partial\phi}]I_1\\ &+\mu[B_2(\phi)-C_2(\phi)(\xi_2+\frac{\partial X}{\partial\phi_2})-\omega\frac{\partial Y_2}{\partial\phi}]I_2+O(I_1^2)+\mu O(I^2).\end{aligned}$$

Here one should be a little careful when unravelling the notation: the keypoint is that in the last two lines, which essentially duplicate the usual result, one gets the *partial* ϕ-gradients of the scalar function $X(\phi)$ but the *total* gradients of the vector functions $Y_1(\phi)$ and $Y_2(\phi)$.

The system to be solved thus reads:

$$\begin{cases}\omega\frac{\partial X}{\partial\phi}=A(\phi)\\ \omega\frac{\partial Y_1}{\partial\phi}=B_1(\phi)-C_1(\phi)(\xi_1+\frac{\partial X}{\partial\phi_1})\\ \omega\frac{\partial Y_2}{\partial\phi}=B_2(\phi)-C_2(\phi)(\xi_2+\frac{\partial X}{\partial\phi_2}).\end{cases}$$

The first equation is the same as usual and can be solved because $A(\phi)$ has zero average and ω is diophantine. The last two equations are identical and indeed exactly duplicate the usual one (see [K] or eq. (4.14) in [BGGS]). So they can be solved for the appropriate (and unique) choice of $\xi=(\xi_1,\xi_2)$, because the averages of $C_1(\phi)$ and $C_2(\phi)$ are invertible, and using again the fact that ω is diophantine. Moreover, the estimates on ξ, X, Y_1 and Y_2 are clearly the same as usual, and for good reasons.

This essentially finishes the proof. We mention for completeness the last insignificant departure from the usual scheme that needs be brought in the evaluation of the remainder $H'-H-\{\chi,H^0\}$. To this end, one has simply to estimate $\{\chi,H^1\}$ and $\{\chi,\{\chi,H\}\}$ (second order Taylor expansion; see [BGGS] §3.1). The minor point we want to make is that the μ-independent and the μ-dependent terms should not be lumped together. More precisely, one starts with estimates for A, B_1 and B_2 and wishes to derive similar estimates for A', B_1' and B_2'. As mentioned above, one first gets estimates for ξ, X, Y_1 and Y_2, the exact same as usual. Now decompose H^1 and χ as $H^1=H_0^1+\mu H_1^1$, $\chi=\chi_0+\mu\chi_1$, with $H_1^1=B_2I_2$ and $\chi_1=Y_2I_2$; each of H_0^1, H_1^1, χ_0 and χ_1 has thus already been estimated. It is now immediate to separately estimate the order 0 (w.r.t. μ) and order 1 parts of $\{\chi,H^1\}$ and $\{\chi,\{\chi,H\}\}$. Here H can be left as is and one does not even have to expand; just count the number of terms. This is a minor point, and the rest literally follows the standard case.

We will finish with some remarks on well-known issues, namely nondegeneracy conditions, arithmetical conditions and smoothness conditions.

Concerning the first of these topics, we required here that C_1 and C_2 be invertible, which is a condition on the second order jet of the unperturbed Hamiltonian. This can be considerably weakened, as in particular the work of H.Rüssman demonstrates. In order to apply his results in the present setting, consider again a Hamiltonian $H(I,\phi,\varepsilon,\mu)$ which is nearly integrable (and integrated), i.e. $H(I,\phi,0,\mu)$ is actually ϕ-independent. The nondegeneracy conditions will then be imposed on the ϕ-independent functions $H(I,\phi,0,0)$ and $\partial H/\partial\mu(I,\phi,0,0)$, by considering their jets (collections of higher order derivatives) up to a certain order. Above we simply restricted attention to the Hessian matrices (order 2 jet) of these two functions.

Coming to the question of the arithmetical conditions, we would like to call attention on the recent papers by A.Giorgilli and U.Locatelli (*ZAMP* **48**, 1997 and *MPEJ* **3**, 1997) in which it is shown how by suitably modifying (yet again...) Kolmogorov's original scheme, one can prove the preservation of invariant tori for frequencies satisfying Bruno's condition (only the case of a trigonometrical polynomial perturbation in the angles is explicitly treated there). We also recommend the detailed introductions of these papers. It is plausible that this work can be adapted to the cases investigated in the present note.

Finally concerning smoothness conditions, the usual remarks are in order and again adaptation from the standard case to the one at hand should be fairly straightforward (although possibly cumbersome). Since we have been working essentially straight from Kolmogorov's original paper, it may be instructive or amusing to insert in closing a short historical remark. Contrary to what is sometimes asserted, Kolmogorov was well aware of the fact that the phenomenon he was discovering is not specific to the analytic category. Yet this point seems to have given rise to the only "error" or say inaccuracy in [K]; he writes: "In essence, the considerations below are related to real functions, but impose rather significant conditions on the smoothness of the function $H(q,p,\theta)$ [θ is the small parameter], stronger than infinite differentiability. For simplicity, in what follows we assume that the function $H(q,p,\theta)$ is jointly analytic in the variables (q,p,θ)."

Smoothness is clearly a minor issue in [K], but it should be recalled that Kolmogorov was actually an expert on the subject. Indeed he started his "career" in 1922 (the result was published in 1923) at the age of 20 when, taking advantage of the NEP, he exhibited a Lebesgue integrable periodic function with almost everywhere divergent Fourier series. This gained him immediate recognition, especially in France. Yet in the above quotation, he seems to imply that one should impose Gevrey-type conditions on the perturbation. As is well-known, this inaccuracy was corrected by J.Moser who showed that via an appropriate use of J.Nash's smoothing operators, one can actually deal with perturbations which possess only a finite number of derivatives.

References

[BGGS] 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.

[G1] G.Gallavotti, Twistless KAM tori, Commun. in Math. Phys. **164**, 1994, 145-156.

[G2] G.Gallavotti, Twistless KAM tori etc., Reviews in Math. Phys. **6**, 1994, 343-411.

[K] A.N.Kolmogorov, On the preservation of conditionally periodic motions, Doklady AN **98**, 1954, 527-530.