An averaging theorem for Hamiltonian dynamical systems in the thermodynamic limit

A. Carati*

January 20, 2006

ABSTRACT

It is shown how to perform some steps of perturbation theory if one assumes a measure–theoretic point of view, i.e. if one renounces to control the evolution of the single trajectories, and the attention is restricted to controlling the evolution of the measure of some meaningful subsets of phase–space. For a system of coupled rotators, estimates uniform in N for finite specific energy can be obtained in quite a direct way . This is achieved by making reference not to the sup norm, but rather, following Koopman and von Neumann, to the much weaker L^2 norm.

Running title: An averaging theorem in the thermodynamic limit

*Università di Milano, Dipartimento di Matematica Via Saldini 50, 20133 Milano (Italy)
E-mail: carati@mat.unimi.it

1 Introduction

A very much discussed problem is the question whether Nekhoroshev–type theorems can have some relevance for the foundations of Statistical Mechanics. In its heuristic formulation ("*actions remain frozen to their initial values up to exponentially long times*") this theorem seems to grasp the essential feature of the Fermi–Pasta–Ulam phenomenon: the energy remains confined to the low frequency modes, while the energies (i.e., up to a factor, the actions) of the high frequency modes remain frozen up to very large times. On the other hand, in Nekhoroshev theorem the estimate for the time of freezing is of the type $T \simeq T_0 \exp(\varepsilon^*/\varepsilon)^{1/N}$, where ε is the pertubative parameter (for example the specific energy) and N is the number of degrees of freedom of the system. Thus, for systems of interest to Statistical Mechanics, in which N is very large, the exponential estimate in the above formula disappears (see [1], [7]). These considerations seem to indicate that Classical Perturbation Theory may be useless for the aims of statistical mechanics.

The following remark is however in order. The aim of Perturbation Theory, as it was developed until now, is to give the most accurate description of every trajectory of a dynamical system, and this enforces, at a technical level, the use of the sup norm in the phase space for the estimate of the relevant quantities. On the other hand, for the aims of Statistical Mechanics a control over any single trajectory is completely irrelevant (the knowledge of the values of 10^{23} actions instant by instant is enormously redundant for estimating, for example, the specific heat). So one can limit oneself to control only the evolution of some significant quantity, for example the energy of some subsystem of the complete system. In this case the dependence on the total number N of degrees of freedom changes drastically, and in fact estimates uniform in N were obtained (see [2], [3]); however, such estimates turn out to be valid only for finite total energy E, namely for vanishing specific energy E/N in the thermodynamic limit $N \to +\infty$. Very recently, in the same spirit, estimates uniform in N for non–vanishing specific energy E/N where given (see [4]) in the Fermi–Pasta–Ulam problem, but only for a special class of initial data.

But, for the pourpose of the Statistical Mechanics, the author feels that also this weakened approach is unnecessarily strong, because one pretends to control some dynamical variable "initial data by initial data", without taking into account any statistical aspect. Instead, a measure–theoretic point of view ought to be taken, namely one should renounce to control the evolution of the single trajectories, and the attention should be restricted to controlling the evolution of the measure of some meaningful subset of phase–space: actually, in this case, estimates uniform in N for finite specific energy can be obtained in quite a direct way. This is achieved by making reference not to the sup norm, but rather to some much weaker integral norm, tipically, following Koopman and von Neumann (see [5], [6]), the L^2 norm, which will be the one used in the present paper.

To this end, we first show how, for a generic system, an estimate of the rate of mixing for any invariant measure μ can be given. This is shown in Section 2. Then, by considering a concrete example (a system of N rotators with nearest neighbour interaction), we show that at least three steps of the perturbative construction can be performed. This will be obtained by making use of the method of the direct construction of integrals of motion. The corresponding estimates show that the mixing rate is much smaller than the one estimated directly from the equations of motion. To this second task Sections 3, 4 and 5 will be devoted. In particular, in Section 3 a normal form theorem is given, from which the estimate of the mixing rate follows as a Corollary. Section 4 is devoted to an accurate discussion of the first step of perturbation theory, while two further steps are performed in Section 5, leading to the proof of the theorem. A technical Lemma is proven in an appendix.

2 The estimate of the mixing rate

Consider a Hamitonian system with Hamiltonian function H on a phase space $\mathcal{M}$, endowed with a finite invariant measure μ (so that we can suppose $\mu(\mathcal{M}) = 1$). It is well known that the existence of a smooth integral of motion $f(\mathbf{x})$ independent of the energy implies that the system is not ergodic (on a single surface of constant energy). Indeed, obviously, the two sets $A = \{x : |f(\mathbf{x}) - \bar{f}| < k\}$ and $B = \{x : |f(\mathbf{x}) - \bar{f}| > 2k\}$, where $\bar{f} \stackrel{\mathrm{def}}{=} \int f(\mathbf{x})\,\mathrm{d}\,\mu$ is the expectation of f and k a positive constant, are invariant disjoint nontrivial sets (considering for example the Gibbs measure, and a not too large value of k).

Suppose now f is only a quasi–constant of motion, in the sense that (we denote by $[\cdot,\cdot]$ the Poisson bracket of two functions) $[H, f]$ is small in L^2 norm; the problem of finding such a function will be one of the main themes discussed later. In such a case the sets A and B are no more invariant: denoting by A_t the set evolved from A according to the dynamics, one expects that $A_t \cap B \neq \emptyset$.

But if the evolution is slow (in the mean), one expects that it will take some time in order that, at a given point of A, the value of the function f_t, i.e. the evolution of f (to be defined in a moment), grows from the value (smaller than k) it has at time $t = 0$ to the values (larger than $2k$) that f

has in B, i.e., in terms of sets, the measure of $A_t \cap B$ is expected to remain small up to a certain time.

In order to give a rigorous form to such rather vague reasoning, we begin with introducing the notion of *mixing time*. We recall that one defines a system to be (strongly) mixing if $\mu(A_t \cap B) \to \mu(A)\mu(B)$ as $t \to +\infty$. But, as especially pointed out by von Neumann, it is also of interest to have an estimate of the actual relaxation time, i.e. the time at which the limit value is actually reached. This is particular relevant if $\mu(A_t \cap B)$ grows slowly. This justifies the following definition

Definition 1 (mixing time) *We define the mixing time t_{mix} for the two sets A and B, defined as above, by $t_{\mathrm{mix}} = \sup t^*$, where t^* is such that*

$$\mu(A_t \cap B) < \frac{1}{2}\mu(A)\mu(B) \tag{1}$$

for all $0 < t < t^*$.

The choice of the factor 1/2 to define the degree of mixing is a matter of convention, and it will be appear later that any other choice α with $0 < \alpha < 1$ would work as well.

So, the previous considerations can be restated in terms of the mixing time, by saying that the presence of a quasi–constant of motion f is expected to lead to a large value of t_{mix}. In order to prove this fact we have first of all to recall in which way the function f evolves with time. Denoting by Φ^t the flow generated by the equations of motion, we define the evolution of the function $f(\mathbf{x})$ by $f_t \stackrel{\mathrm{def}}{=} f \circ \Phi^{-t}$ (the definition $f_t = f \circ \Phi^t$ is often adopted, but the difference is irrelevant). So f_t is a time–dependent constant of motion, i.e., satisfies the Liouville equation

$$\partial_t f_t + [H, f_t] = 0 \ .$$

Thus, even if at time $t = 0$ the derivative of f_t is small, it may happen that $\| f_t - f \|_2$ (the L^2–norm of $f_t - f$) becomes large as time increases, so that the intersection $A_t \cap B$ too may become large. One actually has the following first, simple, perturbative result

Theorem 1 *Let μ be an invariant measure, and $f \in L^2(\mathrm{d}\,\mu)$ be such that*

$$\| \, [H, f] \, \|_2 \le \eta \| \, f \, \|_2 \ , \tag{2}$$

with a positive constant η, where $\| \cdot \|_2$ denotes the $L^2(\mathrm{d}\,\mu)$ norm. If f_t is the evolution of the function f, then one has

$$\| \, f_t - f \, \|_2 \le \eta t \, \| \, f \, \|_2 \ . \tag{3}$$

Proof. Introduce the difference $\delta \stackrel{\text{def}}{=} f_t - f$. As f_t satisfies the Liouville equation and f is time–independent, one has $\partial_t \delta = \partial_t f_t = -[H, f_t]$, which in terms of δ takes the form

$$\partial_t \delta = -[H, \delta] + g \ , \tag{4}$$

with $g \stackrel{\text{def}}{=} -[H, f]$. It is well known that, μ being invariant, the solutions of the Liouville equation are generated by a one–parameter group $\hat{U}(t)$ of unitary operators (see [5]) in the sense that $f_t = \hat{U}(t) f$. As $\delta(0) = 0$, the solution of equation (4) is given by

$$\delta = \int_0^t \hat{U}(t-s) g \, \mathrm{d}\, s \ , \tag{5}$$

so that, $\hat{U}$ being unitary (i.e., $\| \hat{U}(t-s) g \|_2 = \| g \|_2$), one gets the estimate

$$\| \delta \|_2 \leq \int_0^t \| \hat{U}(t-s) g \|_2 \, \mathrm{d}\, s = t \| g \|_2 \leq \eta t \| f \|_2$$

i.e., the thesis. Q.E.D

We are now in a position to give a simple estimate of the measure of the intersection $A_t \cap B$. Notice that, if $\mathbf{x} \in A_t \cap B$, then one has both $|f(\mathbf{x}) - \bar{f}| > 2k$ (since $\mathbf{x} \in B$), and $|f_t(\mathbf{x}) - \bar{f}| < k$ (since $f_t(x) = f(\Phi^{-t} x)$ and $\Phi^{-t} x \in A$). So one has

$$\begin{aligned} k\mu(A_t \cap B) &\leq \int_{A_t \cap B} \Big| |f(\mathbf{x}) - \bar{f}| - |f_t(\mathbf{x}) - \bar{f}| \Big| \, \mathrm{d}\, \mu \leq \\ &\leq \int_{A_t \cap B} |f(\mathbf{x}) - f_t(\mathbf{x})| \, \mathrm{d}\, \mu \leq \Big(\int_{A_t \cap B} \mathrm{d}\, \mu \Big)^{1/2} \Big(\int_{A_t \cap B} |f(\mathbf{x}) - f_t(\mathbf{x})|^2 \, \mathrm{d}\, \mu \Big)^{1/2} \\ &\leq \Big(\mu(A_t \cap B) \Big)^{1/2} \| f_t - f \|_2 \ . \end{aligned}$$

Thus, by (3) one gets

$$\mu(A_t \cap B) \leq \eta^2 t^2 \frac{\| f \|_2^2}{k^2} \ . \tag{6}$$

So, we have proved the following theorem (analogous to that of Chebyshev)

Theorem 2 *Let μ be an invariant finite measure, and $f \in L^2(\mathrm{d}\,\mu)$ have the property $\|\,[H,f]\,\|_2 \le \eta \|\,f\,\|_2$. Define the sets A and B by*

$$A = \{\mathbf{x} : |f(\mathbf{x}) - \bar{f}| \le k\}\ , \quad B = \{\mathbf{x} : |f(\mathbf{x}) - \bar{f}| \ge 2k\}\ ,$$

with $\bar{f} = \int f \,\mathrm{d}\,\mu$ and k a positive constant. Then the estimate (6) holds.

Relation (6) allows one to give a lower bound to the mixing time. In fact using (6) one has $\mu(A_t \cap B) < \frac{1}{2}\mu(A)\mu(B)$ for all t such that

$$t < \frac{k\sqrt{2}}{\eta\|\,f\,\|_2}\Big(\mu(A)\mu(B)\Big)^{1/2}\ , \tag{7}$$

so that one gets the estimate

$$t_{\mathrm{mix}} \ge \frac{k\sqrt{2}}{\eta\|\,f\,\|_2}\Big(\mu(A)\mu(B)\Big)^{1/2}\ , \tag{8}$$

One sees that $t_{\mathrm{mix}} \to +\infty$ as $\eta \to 0$, so that a sort of continuity is recovered. If f is a constant of motion, the two sets A and B remain separated for all times; if the time derivative of f is small, then A and B remain "quasi" separated (at least in measure) for very long times, which tend to infinity with the vanishing of the derivative $\dot{f}$, namely of η.

A comment on relation (8): up to now we have considered the constant k as a free parameter. But, as k is a measure of the deviation of f from its expectation $\bar{f}$, it is meaningful to take it of the same order of magnitude as the standard deviation of f,

$$\delta_f \stackrel{\mathrm{def}}{=} \left[\int (f(\mathbf{x}) - \bar{f})^2 \,\mathrm{d}\,\mu\right]^{1/2}\ .$$

Otherwise it could happen that the measure of A or that of B be essentially zero and the estimate (8) trivial. So, in the rest of the paper we fix $k = \delta_f$, and our estimate (8) becomes

$$t_{\mathrm{mix}} \ge \frac{\sqrt{2}\,\delta_f}{\eta\|\,f\,\|_2}\Big(\mu(A)\mu(B)\Big)^{1/2}\ . \tag{9}$$

3 The periodic chain of rotators

In the rest of the paper we tackle the problem of constructing, for a concrete system, a function f which has a slow evolution, i.e., satisfies (2) with a small η. The system we consider is a classical one, a chain of $2N$ rotators with nearest neighbour trigonometric coupling and periodic boundary conditions, i.e., the system with Hamiltonian

$$H = \sum_{j=-N+1}^{N} \frac{p_j^2}{2} - \sum_{j=-N}^{N} V_0 \cos(q_{j+1} - q_j) \ , \qquad q_N = q_{-N} \ , \tag{10}$$

where $q_j \in \mathbf{T}^1$, $p_j \in \mathbf{R}$ and V_0 is a positive constant. As an invariant measure we take the Gibbs one at inverse temperature β, defined by

$$\mathrm{d}\,\mu = \frac{1}{Z} \exp(-\beta H)\,\mathrm{d}\,\mathbf{x} \ , \quad Z = \int \exp(-\beta H)\,\mathrm{d}\,\mathbf{x} \tag{11}$$

with $\mathbf{x} = (q_{-N+1}, \ldots, p_N)$, and $\mathrm{d}\,\mathbf{x} = \mathrm{d}\,q_{-N+1} \ldots \mathrm{d}\,p_N$.

For notational simplicity we will perform the (noncanonical) change of coordinates $\tilde{q}_j = q_j$, $\tilde{p}_j = \beta^{1/2} p_j$, and a change of time $\tau = \beta^{-1/2} t$. This being understood, we drop tildes, and denote $\tilde{q}_j$ by q_j and $\tilde{p}_j$ by p_j. The resulting equations of motion can be deduced from the Hamiltonian function

$$H = \sum_{j=-N+1}^{N} \frac{p_j^2}{2} - \varepsilon \sum_{j=-N}^{N} \cos(q_{j+1} - q_j) \ , \qquad q_N = q_{-N} \ , \tag{12}$$

where we have denoted $\varepsilon = \beta V_0$. Correspondingly, the Gibbs measure becomes

$$\mathrm{d}\,\mu = \frac{1}{Z} \exp(-H)\,\mathrm{d}\,\mathbf{x} \ , \quad Z = \int \exp(-H)\,\mathrm{d}\,\mathbf{x} \ .$$

From the form of the Hamiltonian it is apparent that ε is our small parameter, because for $\varepsilon = 0$ our system is formally integrable, having, as constants of motions, all the functions p_j. For small ε, one has instead

$$[H, p_j] = \varepsilon \Big(\sin(q_{j+1} - q_j) - \sin(q_j - q_{j-1}) \Big) \ .$$

From this it follows that the momenta p_j themselves have a slow evolution, or are quasi–integrals, because they satisfy the relation (2) with $\eta = 2\varepsilon$, i.e.,

$$\|\, [H, p_j] \,\|_2 \le 2\varepsilon = 2\varepsilon \|\, p_j \,\|_2 \ .$$

This follows making use of the facts that $|\sin x| \le 1$ and that the p_j, being normally distributed with unit variance and zero mean, have the property $\|\, p_j \,\|_2 = 1$. So, applying the estimate (9) one finds that the mixing time is $\tau_{\mathrm{mix}} \sim \varepsilon^{-1}$, which in terms of the original, non–rescaled time, gives the estimate

$$t_{\mathrm{mix}} \sim \varepsilon^{-1/2} \ . \tag{13}$$

But actually the mixing time is much smaller, because perturbation theory up to third order leads to the following

Theorem 3 (normal form construction) *For any j, there exists a function f_j of the form $f_j = p_j + \varepsilon^{3/5} X_j(\mathbf{p}, \mathbf{q})$ having the properties*

$$\| \, [H, f_j] \, \|_2 \le C_1 \varepsilon^{1+\frac{3}{5}} \tag{14}$$

$$\| \, X_j \, \|_2 \le C_2 \ , \tag{15}$$

with two positive constants C_1 and C_2 independent of ε and N.

The construction of the function f_j is performed in the next two Sections, using the method of the direct construction of a first integral (see for example [8]), and implementing three steps of the perturbative construction.

It is clear that the estimate (14) leads to an estimate of the mixing time of order

$$t_{\mathrm{mix}} \simeq \varepsilon^{-1/2-3/5} \ , \tag{16}$$

which is much larger than (13). There remains open the question of how many steps of the perturbative construction can be performed. If one could prove that the construction can be performed to all orders, one would obtain a mixing time exponentially large, thus recovering the analog of Nekhoroshev theorem, with however a complete elimination of N.

For a proof of (16) one has to estimate the other quantities entering formula (9). This is provided by the following Lemma

Lemma 1 *For any j, consider the function $f_j = p_j + \varepsilon^{3/5} X_j(\mathbf{p}, \mathbf{q})$, with $X_j \in L^2(\mathrm{d}\,\mu)$. Then one has ($\delta$ denoting standard deviation)*

i) $\bar{f}_j = O(\varepsilon^{3/5})$

ii) $\delta^2_{f_j} = \delta^2_{p_j} + O(\varepsilon^{3/5}) = 1 + O(\varepsilon^{3/5})$

iii) *for the sets $A_\varepsilon = \{\mathbf{x} : |f_j - \bar{f}_j| < \delta_{f_j}\}$ and $B_\varepsilon = \{\mathbf{x} : |f_j - \bar{f}_j| > 2\delta_{f_j}\}$ one has*

$$\mu(A_\varepsilon) = \mu(A_0) + O(\varepsilon^{2/5}) \ , \quad \mu(B_\varepsilon) = \mu(B_0) + O(\varepsilon^{2/5}) \ .$$

Proof. The proof goes as follows.

i) This is immediate. One has $\bar{f}_j = \bar{p}_j + \varepsilon^{3/5} \bar{X}_j$. On the other hand $\bar{p}_j = 0$, and $|\bar{X}_j| \le \int |X_j| \, \mathrm{d}\,\mu \le \| \, X_j \, \|_2$.

ii) This is obtained by a simple computation. Indeed one has

$$\begin{aligned} \delta^2_{f_j} &= \int (f_j - \bar{f}_j)^2 \, \mathrm{d}\,\mu \\ &= \int p_j^2 \, \mathrm{d}\,\mu + 2\varepsilon^{3/5} \int p_j X_j \, \mathrm{d}\,\mu + \varepsilon^{6/5} \int X_j^2 \, \mathrm{d}\,\mu - (\bar{f}_j)^2 \ . \end{aligned}$$

Now, $\int p_j^2 \,\mathrm{d}\,\mu = \delta_{p_j}^2$ since $\bar{p}_j = 0$, while, by the Schwarz inequality, $|\int p_j X_j \,\mathrm{d}\,\mu| \leq \| p_j \|_2 \| X_j \|_2$. The result is then obtained by estimating $\bar{f}_j$ through i).

iii) We show only the first inequality, because the second one is proved in the same way. We start noticing the trivial relation $\delta_{X_j} = \int (X_j - \bar{X}_j)^2 \,\mathrm{d}\,\mu \leq \| X_j \|_2^2$, so that, introducing the set $C = \{\mathbf{x} : |X_j - \bar{X}_j| \geq \varepsilon^{-1/5}\}$, by Chebyshev theorem one gets

$$\mu(C) \leq \varepsilon^{2/5} \| X_j \|_2^2 = O(\varepsilon^{2/5}) \ .$$

Now, the complementary set A_ε / C is contained in the set $A' = \{\mathbf{x} : |p_j| \leq \sigma_{f_j} + \varepsilon^{2/5}\}$, because in A_ε / C one has $|X_j - \bar{X}_j| \leq \varepsilon^{-1/5}$. The measure of the set A' can be readily evaluated, recalling that p_j is normally distributed, and that in addition, by ii), one has $\delta_{f_j} = 1 + \varepsilon^{3/5}$. One thus finds $\mu(A') = \mu(A_0) + O(\varepsilon^{2/5})$, and so one gets the thesis using $\mu(A_\varepsilon) = \mu(C) + \mu(A_\varepsilon / C) \leq \mu(C) + \mu(A')$. **Q.E.D.**

4 The first Perturbative Step

We have now to show how the quasi–constants of motion f_j entering Theorem 3 are constructed. The first perturbative step is performed in the present section. From the Hamiltonian (12) we obtain the following equations of motion

$$\begin{cases} \dot{p}_j = \varepsilon\Big(\sin(q_{j+1} - q_j) - \sin(q_j - q_{j-1}) \Big) \\ \dot{q}_j = p_j \ . \end{cases} \tag{17}$$

It is well known (see [9]) how a normal form (which is however formal for $N \gg 1/\varepsilon$) can be constructed for this equation. However, in this simple case it is possible to find a first–order integral directly, avoiding the use of the normal form techniques. This is obtained by recalling that in virtue of the equation of motion one has the relation

$$\begin{aligned} \varepsilon \sin(q_{j+1} - q_j) = &- \frac{\mathrm{d}}{\mathrm{d}\,t} \left(\varepsilon \frac{\cos(q_{j+1} - q_j)}{p_{j+1} - p_j} \right) + \\ &+ \varepsilon^2 \frac{\cos(q_{j+1} - q_j)}{(p_{j+1} - p_j)^2} \Big(\sin(q_{j+2} - q_{j+1}) - 2 \sin(q_{j+1} - q_j) + \sin(q_j - q_{j-1}) \Big) \ . \end{aligned}$$

and the analogous one for $\varepsilon \sin(q_j - q_{j-1})$. So, in the region $p_j \neq p_{j\pm1}$, if we define

$$\tilde{X}_j^{(1)} \stackrel{\mathrm{def}}{=} \frac{\cos(q_{j+1} - q_j)}{p_{j+1} - p_j} - \frac{\cos(q_j - q_{j-1})}{p_j - p_{j-1}} \ ,$$

we find that the function $\tilde{f}_j^{(1)} = p_j + \varepsilon \tilde{X}_j^{(1)}$ evolves with velocity of order ε^2, i.e., is slower than p_j. It is obvious that, due to the presence of the denominators (the small divisors), the function $\tilde{X}_j^{(1)}$ is not in L^2, and so is useless for the estimates.

This example also shows in a very clear way the difficulty of applying the standard perturbation techniques for large N. Indeed, in order to have a slow evolution (of order ε^2), it is not enough to restrict the initial data to the region $|p_j - p_{j\pm 1}| > \sigma$ (with σ a positive parameter), but one has to secure that such an inequality also holds for times or order ε^{-2}. On the other hand, this cannot be secured, because $p_{j\pm 1}$ evolve in general on a time scale of order ε^{-1}. One way to secure that $p_{j\pm 1}$ do not evolve too much is to consider the functions $\tilde{X}_{j\pm 1}^{(1)}$, and choose initial data such that $|p_{j\pm 1} - p_{j\pm 2}| > \sigma$. But then we have the problem of the evolution of the variables $p_{j\pm 2}$. Thus, one is forced to iterate this procedure, so that our hypothesis can be secured only in a set of the form $C = \{\mathbf{x} : |p_j - p_{j+1}| > \sigma, \forall j\}$. On the other hand a simple computation shows that one has $\mu(C) \simeq (1-\sigma)^N$, i.e., that the set C has essentially a vanishing measure for N large.

Instead, if we want to control the evolution of the measure of the sets, and not the single trajectories, this kind of problems is not met. In fact, one can limit oneself to perform the normalization only in the non–resonant zone $|p_j - p_{j\pm 1}| > \sigma$, and keep the action p_j unaltered in the resonant one. The idea is thus to define

$$f_j^{(1)} \stackrel{\mathrm{def}}{=} \begin{cases} p_j & \text{for} \quad |p_j - p_{j\pm 1}| < \sigma \\ p_j + \varepsilon \tilde{X}_j^{(1)} & \text{for} \quad |p_j - p_{j\pm 1}| > \sigma \ , \end{cases}$$

in such a way that the region where the derivative of $f_j^{(1)}$ is large, has a small measure (of order σ). Now, choosing in an appropriate way σ as a function of ε (we will take $\sigma = \varepsilon^{2/5}$), one can obtain that the L^2–norm of $\dot{f}_j^{(1)}$ becomes less than the L^2–norm of $\dot{p}_j$, notwithstanding the fact that these two functions have the same sup–norm.

In order to give this idea a clear mathematical content, we need to introduce some objetcs. First of all we need a truncation function $\zeta(x)$ of $\mathcal{C}^\infty$ class, i.e. a function having the properties stated in

Lemma 2 *For every sufficiently small (positive) constant σ, there exists a $\mathcal{C}^\infty(\mathbf{R})$ function $\zeta(x)$ such that:*

i) *one has $\zeta(x) = 1$ for $|x| < \sigma$, and $\zeta(x) = 0$ for $|x| > 3\sigma$;*

ii) *for all $n \in \mathbf{N}$, one has $|\partial_x^n \zeta(x)| < c_n \sigma^{-n}$, where c_n are numerical constants independent of x and σ; moreover $\partial_x^n \zeta(x) = 0$ for $|x| < \sigma$ and $|x| > 3\sigma$;*

iii) *having defined* $\mathcal{Z}(x) \stackrel{\text{def}}{=} \int \zeta(x)\,\mathrm{d}\,x$ *and* $\mathcal{Z}^{(2)}(x) \stackrel{\text{def}}{=} \int \mathcal{Z}(x)\,\mathrm{d}\,x$, *then for* $|x| > 3\sigma$ *one has* $\mathcal{Z}(x) = 0$ *and* $\mathcal{Z}^{(2)}(x) = 0$. *Moreover the estimates* $|\mathcal{Z}(x)| < 4|x|$ *and* $|\mathcal{Z}^{(2)}(x)| < |x|^2/2$ *hold.*

These are standard properties of truncation functions, the only unusual one being iii), the meaning of which will become clear in the next Section, when we will go beyond the first order. The proof of this Lemma is deferred to the Appendix.

Furthermore, we define the integer vectors $\mathbf{e}_j \in \mathbf{Z}^{2N}$ as the standard basis vectors, i.e., those having all components vanishing but the j–th one, which is equal to one. Analogously we define the vectors $\delta_j \stackrel{\text{def}}{=} \mathbf{e}_{j+1} - \mathbf{e}_j$. We finally define the set M_j^1 made up of the four vectors $\pm\delta_j, \pm\delta_{j-1}$.

In order to have formulas with the minimal number of indices, from now on we concentrate on the case $j = 0$, but it will be obvious that all formulas are valid for a generic value of j. The equation of motion for p_0 can be rewritten as

$$\dot{p}_0 = \sum_{\mathbf{k}\in M_0^1} \frac{\varepsilon}{2i} c_{\mathbf{k}} \exp(i\mathbf{k}\cdot\mathbf{q}) \ ,$$

where the constants $c_{\mathbf{k}} = \pm 1$ come from the expression of the sine in terms of complex exponentials. Now, using the function $\zeta(x)$ one can separate the resonant part from the non resonant one as follows

$$\dot{p}_0 = \varepsilon \sum_{\mathbf{k}\in M_0^1} c_{\mathbf{k}} \zeta(\mathbf{k}\cdot\mathbf{p}) \frac{\exp(i\mathbf{k}\cdot\mathbf{q})}{2i} + \varepsilon \sum_{\mathbf{k}\in M_0^1} c_{\mathbf{k}} \big(1 - \zeta(\mathbf{k}\cdot\mathbf{p})\big) \frac{\exp(i\mathbf{k}\cdot\mathbf{q})}{2i} \ .$$

Thus, by integrating by parts the second term at the r.h.s., one gets the identity

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\,t} \left(p_0 + \varepsilon \sum_{\mathbf{k}\in M_0^1} c_{\mathbf{k}} \frac{1 - \zeta(\mathbf{k}\cdot\mathbf{p})}{\mathbf{k}\cdot\mathbf{p}} \, \frac{\exp(i\mathbf{k}\cdot\mathbf{q})}{2} \right) &= \varepsilon \sum_{\mathbf{k}\in M_0^1} c_{\mathbf{k}} \zeta(\mathbf{k}\cdot\mathbf{p}) \frac{\exp(i\mathbf{k}\cdot\mathbf{q})}{2i} \\ + \varepsilon \sum_{\mathbf{k}\in M_0^1} \frac{c_{\mathbf{k}}}{2} \partial_x \left. \frac{1 - \zeta(x)}{x} \right|_{\mathbf{k}\cdot\mathbf{p}} \exp(i\mathbf{k}\cdot\mathbf{q})\, \mathbf{k}\cdot\dot{\mathbf{p}} \ . & \end{aligned} \tag{18}$$

Now the term $\mathbf{k}\cdot\dot{\mathbf{p}}$ is of order ε, because

$$\mathbf{k}\cdot\dot{\mathbf{p}} = \varepsilon \sum_j \sum_{\mathbf{k}'\in M_j^1} \mathbf{k}'\cdot\mathbf{e}_j \frac{c_{\mathbf{k}'}}{2i} \exp(i\mathbf{k}'\cdot\mathbf{q}) \ ,$$

and moreover, since $\mathbf{k} \in M_0^1$, only the terms with $j = -2, \ldots, 1$ do not vanish (for the other values of j one has $\mathbf{k}' \cdot \mathbf{e}_j = 0$). So, one gets

$$\exp(i\mathbf{k} \cdot \mathbf{q})\mathbf{k} \cdot \dot{\mathbf{p}} = \varepsilon \sum_{\mathbf{k}' \in \{M_j^1\}} \mathbf{k}' \cdot \mathbf{e}_j \frac{c_{\mathbf{k}'}}{2i} \exp\big(i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{q}\big)$$

where the summation is performed over all sets M_j^1 with $j = -2, \ldots, 1$. Now, introducing the function

$$X_0^{(1)} \stackrel{\text{def}}{=} \sigma \sum_{\mathbf{k} \in M_0^1} c_{\mathbf{k}} \frac{1 - \zeta(\mathbf{k} \cdot \mathbf{p})}{\mathbf{k} \cdot \mathbf{p}} \frac{\exp(i\mathbf{k} \cdot \mathbf{q})}{2} \ , \tag{19}$$

we can rewrite (18) in the form

$$\frac{\mathrm{d}}{\mathrm{d}\, t}\left(p_0 + \varepsilon \sigma^{-1} X_0^{(1)}\right) = \varepsilon \sum_{\mathbf{k} \in M_0^1} c_{\mathbf{k}} \zeta(\mathbf{k} \cdot \mathbf{p}) \frac{\exp(i\mathbf{k} \cdot \mathbf{q})}{2i} + \mathcal{R}_1 \ , \tag{20}$$

where the remainder $\mathcal{R}_1$ has the form

$$\mathcal{R}_1 \stackrel{\text{def}}{=} \varepsilon^2 \sum_{\mathbf{k}_1 \in M_0^1} \sum_{\mathbf{k}_2 \in \{M_j^1\}} \mathbf{k}_2 \cdot \mathbf{e}_j \frac{c_{k_1} c_{k_2}}{4i} \, \partial_x \frac{1 - \zeta(x)}{x} \bigg|_{\mathbf{k}_1 \cdot \mathbf{p}} \exp\big(i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{q}\big) \ . \tag{21}$$

The estimate of the r.h.s of (20) is then very simple. Indeed one has

$$\begin{aligned} \int \Big|\zeta(\mathbf{k} \cdot \mathbf{p}) \exp(i\mathbf{k} \cdot \mathbf{q})\Big|^2 \mathrm{d}\,\mu &= \int \Big|\zeta(\mathbf{k} \cdot \mathbf{p})\Big|^2 \mathrm{d}\,\mu \\ &\le \int_{|p_0 - p_{\pm 1}| < 3\sigma} \exp\left(-\frac{p_0^2 + p_{\pm 1}^2}{2}\right) \mathrm{d}\, p_0 \, \mathrm{d}\, p_{\pm 1} = 6\sigma \ , \end{aligned}$$

so that

$$\| \sum_{\mathbf{k} \in M_0^1} c_{\mathbf{k}} \zeta(\mathbf{k} \cdot \mathbf{p}) \frac{\exp(i\mathbf{k} \cdot \mathbf{q})}{2i} \|_2 \le 4\sigma^{1/2} \ . \tag{22}$$

Instead, in order to estimate the term $\mathcal{R}_1$ given by (21), one has to estimate a finite sum of integrals of the type

$$\begin{aligned} \int \bigg|\partial_x \frac{1 - \zeta(x)}{x}\bigg|_{\mathbf{k} \cdot \mathbf{p}} \exp(i\mathbf{k} \cdot \mathbf{q})\bigg|^2 \mathrm{d}\,\mu &= \int \bigg|\partial_x \frac{1 - \zeta(x)}{x}\bigg|_{k \cdot \mathbf{p}}\bigg|^2 \mathrm{d}\,\mu \\ &\le \int_{|p_0 - p_{\pm 1}| > 3\sigma} \bigg|\partial_x \, \frac{1 - \zeta(x)}{x}\bigg|_{p_0 - p_{\pm 1}}\bigg|^2 \exp\left(-\frac{p_0^2 + p_{\pm 1}^2}{2}\right) \mathrm{d}\, p_0 \, \mathrm{d}\, p_{\pm 1} \ . \end{aligned}$$

We can estimate the derivative appearing in the last term, using ii) of Lemma 2 for the derivative of $\zeta(x)$, and the fact that the denominator is bounded away from zero (since $|x| > \sigma$). One has thus

$$\left| \partial_x \left. \frac{1-\zeta(x)}{x} \right|_{p_0 - p_{\pm 1}} \right|^2 \leq \mathsf{C}\sigma^{-4} \ ,$$

with a given constant C, and using this bound in the above formula one finds

$$\int \left| \partial_x \left. \frac{1-\zeta(x)}{x} \right|_{k\cdot\mathbf{p}} \exp(i\mathbf{k}\cdot\mathbf{q}) \right|^2 \mathrm{d}\,\mu \leq \mathsf{C}\sigma^{-4} \ .$$

Thus, there exists a numerical constant $\mathcal{C}_1$ such that

$$\| \mathcal{R}_1 \|_2 \leq \mathcal{C}_1 \sigma^{-2} \varepsilon^2 \ . \tag{23}$$

In exactly the same way one can show, from the expression (19) for the function $X_0^{(1)}$, that there exists a numerical constant $\mathcal{C}_2$, such that

$$\| X_0^{(1)} \|_2 \leq \mathcal{C}_2 \ . \tag{24}$$

Now, taking $\sigma = \varepsilon^{2/5}$, from estimates (23) and (22) one gets that the function $f_0^{(1)} = p_0 + \varepsilon^{3/5} X_0^{(1)}$ has a time derivative of order $\varepsilon^{1+1/5}$, i.e., it evolves more slowly than the action p_0. To obtain a time–evolution as slow as the one implied by (14) of Theorem 3, one needs to perform two more perturbative steps. This will accomplished in the next Section.

5 The next steps of Perturbative Theory and Proof of Theorem 3

By symmetry, letting $(\mathbf{k}_1, \mathbf{k}_2) \to (-\mathbf{k}_1, -\mathbf{k}_2)$, one can easily check that in the expression (21) for $\mathcal{R}_1$ the terms with $\mathbf{k}_1 + \mathbf{k}_2 = 0$ are lacking. So, in complete analogy with what was done for the first step, we can integrate by parts all the terms, obtaining, at least in the region where $|(\mathbf{k}_1 + \mathbf{k}_2)\cdot\mathbf{p}| > \sigma$, a remainder of order ε^3/σ^4, while the resonant zone $|(\mathbf{k}_1 + \mathbf{k}_2)\cdot\mathbf{p}| < \sigma$ will give a contribution to the remainder of order $\varepsilon^2/\sigma^{3/2}$. For $\sigma = \varepsilon^{2/5}$ the two contributions are of the same order $\varepsilon^{1+2/5}$. The question remains of understanding how to treat the first–order resonant term of the remainder, because in this case the phase of the terms $\exp(i\mathbf{k}\cdot\mathbf{q})$ is slow, and so we cannot take the average.

But now one has to consider that slow angles appear only in the time derivatives of slow actions (the quantities $\mathbf{k}\cdot\mathbf{p}$). So one could argue that the terms with slow angles may be replaced by the time derivatives of $\mathbf{k}\cdot\mathbf{p}$ plus some terms containing fast angles. In this case, the terms with the time derivative of the slow action would give a total derivative, while the terms with fast angles could be averaged away. Indeed, things go exactly in this way.

In fact, in our case in which $\mathbf{k}=\pm\delta_j$, $j=0,-1$, one has that $\mathbf{k}'=\mathbf{e}_{j+1}+\mathbf{e}_j$ is orthogonal to $\mathbf{k}$, so that $\mathbf{k}'\cdot\mathbf{p}$ is "fast". In addition, from the equations of motion one gets directly the relation

$$\begin{aligned}\varepsilon\sin(\delta_j\cdot\mathbf{q}) &= \frac{1}{2}(\mathbf{e}_{j+1}+\mathbf{e}_j)\cdot\dot{\mathbf{p}}-\frac{1}{2}(\mathbf{e}_{j+1}-\mathbf{e}_j)\cdot\dot{\mathbf{p}}+\varepsilon\sin(\delta_{j-1}\cdot\mathbf{q})\\ &= \mathbf{k}'\cdot\dot{\mathbf{p}}-\frac{1}{2}\mathbf{k}\cdot\dot{\mathbf{p}}+\varepsilon\sin(\mathbf{k}''\cdot\mathbf{q})\ .\end{aligned}$$

where we have set $\mathbf{k}''=\delta_{j-1}$ (remember that $\mathbf{k}=\delta_j$). From this it follows that the first term at the r.h.s of (20) can be rewritten as

$$\begin{aligned}\varepsilon\sum_{\mathbf{k}\in M_0^1}\frac{c_{\mathbf{k}}\zeta(\mathbf{k}\cdot\mathbf{p})}{2i}\exp(i\mathbf{k}\cdot\mathbf{q}) &= \frac{1}{2}\sum_{\mathbf{k}\in M_0^1}-\zeta(\mathbf{k}\cdot\mathbf{p})\mathbf{k}\cdot\dot{\mathbf{p}}\\ &\quad+\frac{1}{2}\sum_{\mathbf{k}\in M_0^1}\zeta(\mathbf{k}\cdot\mathbf{p})\mathbf{k}'\cdot\dot{\mathbf{p}}+\varepsilon\sum_{\mathbf{k}\in M_0^1}\zeta(\mathbf{k}\cdot\mathbf{p})\sin(\mathbf{k}''\cdot\mathbf{q})\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(-\frac{1}{2}\sum_{\mathbf{k}\in M_0^1}\mathcal{Z}(\mathbf{k}\cdot\mathbf{p})\right)+\frac{\varepsilon}{2}\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in M_{\mathbf{k}_1}^2}c_{\mathbf{k}_2}^1\zeta(\mathbf{k}_1\cdot\mathbf{p})\exp(i\mathbf{k}_2\cdot\mathbf{q})\ ,\end{aligned}$$

where $c_{\mathbf{k}_2}^1$ are numerical constants (less than 2 in absolute value), and the sets $M_{\mathbf{k}_1}^2$ are made up of the vectors $\pm\delta_{j-1}$, $\pm\delta_{j-1}$ (remember that $\mathbf{k}_1=\pm\delta_j$), as one can check using the relation $\mathbf{k}'\cdot\dot{\mathbf{p}}=\varepsilon\Big(\sin(\delta_{j+1}\cdot\mathbf{q})-\sin(\delta_{j-1}\cdot\mathbf{q})\Big)$.

From the expression of $M_{\mathbf{k}_1}^2$ one checks that $\mathbf{k}_2\neq\mathbf{k}_1$, so we have rewritten the resonant term as a total derivative plus some non–resonant terms. Now the non resonant terms can be integrated by parts to give

$$
\begin{aligned}
\varepsilon \sum_{\mathbf{k}\in M_0^1} \frac{c_{\mathbf{k}}\zeta(\mathbf{k}\cdot\mathbf{p})}{2i}\exp(i\mathbf{k}\cdot\mathbf{q}) = \frac{\mathrm{d}}{\mathrm{d}\,t}\Bigg[-\frac{1}{2}\sum_{\mathbf{k}\in M_0^1} \mathcal{Z}(\mathbf{k}\cdot\mathbf{p}) + \\
-\frac{\varepsilon}{2}\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in M^2_{\mathbf{k}_1}} c^1_{\mathbf{k}_2}\frac{\zeta(\mathbf{k}_1\cdot\mathbf{p})\Big(1-\zeta(\mathbf{k}_2\cdot\mathbf{p})\Big)}{\mathbf{k}_2\cdot\mathbf{p}}\exp(i\mathbf{k}_2\cdot\mathbf{q})\Bigg] + \\
+\frac{\varepsilon}{2}\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in M^2_{\mathbf{k}_1}} c^1_{\mathbf{k}_2}\zeta(\mathbf{k}_1\cdot\mathbf{p})\zeta(\mathbf{k}_2\cdot\mathbf{p})\exp(i\mathbf{k}_2\cdot\mathbf{q}) + \mathcal{R}_2\ ,
\end{aligned}
\tag{25}
$$

with

$$
\begin{aligned}
\mathcal{R}_2 = \frac{\varepsilon}{2}\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in M^2_{\mathbf{k}_1}} \frac{c^1_{\mathbf{k}_2}}{4}\Bigg[\frac{\partial}{\partial x}\zeta(x)|_{\mathbf{k}_1\cdot\mathbf{p}}\frac{\Big(1-\zeta(\mathbf{k}_2\cdot\mathbf{p})\Big)}{\mathbf{k}_2\cdot\mathbf{p}}\mathbf{k}_1\cdot\dot{\mathbf{p}} \\
+\zeta(\mathbf{k}_1\cdot\mathbf{p})\frac{\partial}{\partial x}\frac{\Big(1-\zeta(x)\Big)}{x}\Big|_{\mathbf{k}_2\cdot\mathbf{p}}\mathbf{k}_2\cdot\dot{\mathbf{p}}\Bigg]\exp(i\mathbf{k}_2\cdot\mathbf{q})\ .
\end{aligned}
\tag{26}
$$

The remainder $\mathcal{R}_2$ can be simply estimated using the estimate for the derivatives of ζ (the estimate ii) of Lemma 2) in the same way in which $\mathcal{R}_1$ was estimated in Section 4. The only difference is the presence of the terms $\zeta(\mathbf{k}_1\cdot\mathbf{p})$ which restrict the computation of the integrals to a region of measure σ, giving a smaller value. One has indeed

$$
\|\,\mathcal{R}_2\,\|_2 \le C_3\varepsilon^2\sigma^{-3/2}\ , \tag{27}
$$

where C_3 a positive numerical constant.

Now, we turn to the remainder $\mathcal{R}_1$ at the r.h.s of expression (20). As already said, every term can be integrated by parts (outside of the resonant region), and so, using the function $\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)$ to split the phase space in the resonant region and the non resonant one, we get

$$
\begin{aligned}
\mathcal{R}_1 = \varepsilon^2\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in\{M_j^1\}} \mathbf{k}_1\cdot\mathbf{e}_j\frac{c^1_{k_1}}{4i}\frac{\partial}{\partial x}\frac{1-\zeta(x)}{x}\Big|_{\mathbf{k}_1\cdot\mathbf{p}}\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big) \\
\exp\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big) + \frac{\mathrm{d}}{\mathrm{d}\,t}\Bigg[-\varepsilon^2\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in\{M_j^1\}}\mathbf{k}_1\cdot\mathbf{e}_j\frac{c^1_{k_1}}{4}\frac{\partial}{\partial x}\frac{1-\zeta(x)}{x}\Big|_{\mathbf{k}_1\cdot\mathbf{p}} \\
\frac{1-\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)}{(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}}\exp\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big)\Bigg] + \mathcal{R}_3\ ,
\end{aligned}
$$

with

$$\mathcal{R}_3 \stackrel{\text{def}}{=} \varepsilon^2 \sum_{\mathbf{k}_1\in M_0^1} \sum_{\mathbf{k}_2\in\{M_j^1\}} \mathbf{k}_1\cdot\mathbf{e}_j \frac{c_{k_1}^1}{4} \frac{\partial}{\partial x}\frac{1-\zeta(x)}{x}\bigg|_{\mathbf{k}_1\cdot\mathbf{p}} \frac{\partial}{\partial x}\frac{1-\zeta(x)}{x}\bigg|_{(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}}$$
$$(\mathbf{k}_1+\mathbf{k}_2)\cdot\dot{\mathbf{p}}\,\exp\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big)\ .$$

Now, defining

$$X_0^{(2)} = X_0^{(1)} + \frac{\sigma}{2\varepsilon}\sum_{\mathbf{k}\in M_0^1}\mathcal{Z}(\mathbf{k}\cdot\mathbf{p}) +$$
$$+\sigma\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in M_{\mathbf{k}_1}^2} c_{\mathbf{k}_2}^1 \frac{\zeta(\mathbf{k}_1\cdot\mathbf{p})\Big(1-\zeta(\mathbf{k}_2\cdot\mathbf{p})\Big)}{\mathbf{k}_2\cdot\mathbf{p}}\exp(i\mathbf{k}_2\cdot\mathbf{q}) +$$
$$+\varepsilon\sigma\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in\{M_j^1\}} \mathbf{k}_1\cdot\mathbf{e}_j\frac{c_{k_1}^1}{4}\frac{\partial}{\partial x}\frac{1-\zeta(x)}{x}\bigg|_{\mathbf{k}_1\cdot\mathbf{p}}$$
$$\frac{1-\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)}{(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}}\exp\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big)\ ,$$

we find

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\Big(p_0+X_0^{(2)}\Big) = \frac{\varepsilon}{2}\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in M_{\mathbf{k}_1}^2} c_{\mathbf{k}_2}^1\zeta(\mathbf{k}_1\cdot\mathbf{p})\zeta(\mathbf{k}_2\cdot\mathbf{p})\exp(i\mathbf{k}_2\cdot\mathbf{q}) +$$
$$+\varepsilon^2\sum_{\mathbf{k}_1\in M_0^1}\sum_{\mathbf{k}_2\in\{M_j^1\}} \mathbf{k}_1\cdot\mathbf{e}_j\frac{c_{k_1}^1}{4i}\frac{\partial}{\partial x}\frac{1-\zeta(x)}{x}\bigg|_{\mathbf{k}_1\cdot\mathbf{p}}\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big) \tag{28}$$
$$\exp\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big) + \mathcal{R}_2 + \mathcal{R}_3\ .$$

The second step is then accomplished. The estimate can be performed in a very simple way, by estimating the L^2–norm as the sup of the function times the measure (to the power 1/2) of its support (i.e. of the region in which the function does not vanish). One finds in this way

$$\begin{aligned}
&\|\,\mathcal{Z}(\mathbf{k}\cdot\mathbf{p})\,\|_2 \le const\,\sigma^{3/2}\\
&\|\,\varepsilon\zeta(\mathbf{k}_1\cdot\mathbf{p})\zeta(\mathbf{k}_2\cdot\mathbf{p})\exp(i\mathbf{k}_1\cdot\mathbf{p})\,\|_2 \le const\,\sigma\varepsilon\\
&\|\,\mathcal{R}_2\,\|_2 \le const\,\sigma^{-3/2}\varepsilon^2\\
&\|\,\mathcal{R}_3\,\|_2 \le const\,\sigma^{-4}\varepsilon^3\ ,
\end{aligned}$$

i.e. that, for $\sigma=\varepsilon^{2/5}$, all terms are of order $\varepsilon^{1+2/5}$. One has then

$$\|\,p_0+\frac{\varepsilon}{\sigma}X_0^{(2)}\,\|_2 \le C_3\varepsilon^{1+\frac{2}{5}}$$
$$\|\,X_0^{(2)}\,\|_2 \le C_4\ ,$$

with certain numerical constants C_3 and C_4.

We note that, from the explicit form (31) of $\mathcal{Z}(x)$ given in appendix, for $|\mathbf{k}\cdot\mathbf{p}|<\sigma$ one has $\mathcal{Z}(x)=x$, so that in the resonant region one has

$$p_0+\frac{1}{2}\mathcal{Z}(\mathbf{k}\cdot\mathbf{p})=\frac{1}{2}\mathbf{k}'\cdot\mathbf{p}\ ,$$

i.e. in the resonant region our function coincides with the fast action.

At this point one can ask whether it is possible to perform more steps of the perturbative construction, or even an infinite number of them. It is well known that, in the process of the direct construction of an integral of motion, insurmountable difficulties are found in the resonant case. An example of these difficulties was met at the second step, when we had to deal with terms of the type $\zeta(\mathbf{k}\cdot\mathbf{p})\sin(\mathbf{k}\cdot\mathbf{q})$. To perform the third step one analogously has to deal with terms of the type

$$\mathcal{N}_1\stackrel{\text{def}}{=}\zeta(\mathbf{k}_1\cdot\mathbf{p})\zeta(\mathbf{k}_2\cdot\mathbf{p})\sin(\mathbf{k}_2\cdot\mathbf{q})\ ,$$

and

$$\mathcal{N}_2\stackrel{\text{def}}{=}\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)\frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\sin\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big)\ ,$$

in the remainder at the r.h.s of relation (28). At the succesive steps we will find other resonant terms having a form always different from those of the previous steps, and at present we were unable to find a recurrent scheme to perfom an arbitrary number of steps. We limit ourselvws to show briefly how the resonant term at the r.h.s of (28) can be dealt with, and so how the third step of the construction can be performed. In fact the terms $\mathcal{R}_2$ and $\mathcal{R}_3$ are non resonant and thus can be integrated by parts (giving rise, at fourth order, to other resonant terms).

We begin considering the terms of the type $\mathcal{N}_1$. Using the explicit form of the vectors $\mathbf{k}_1$ and $\mathbf{k}_2$, one can check that

$$\sin(\mathbf{k}_2\cdot\mathbf{q})=\alpha_1\mathbf{k}_1\cdot\dot{\mathbf{p}}+\alpha_2\mathbf{k}_2\cdot\dot{\mathbf{p}}+\sum_{\mathbf{k}_3\in M^2_{k_1,K_2}}\beta_{k_3}\sin(\mathbf{k}_3\cdot\mathbf{q})\ ,\quad \mathbf{k}_3\neq\mathbf{k}_1,\mathbf{k}_2\ ,$$

α_i and β_{k_3} being numerical constants, and $M^2_{k_1,K_2}$ a given (finite) set of integer vectors. One thus gets

$$\zeta(\mathbf{k}_1\cdot\mathbf{p})\zeta(\mathbf{k}_2\cdot\mathbf{p})\sin(\mathbf{k}_2\cdot\mathbf{q})=\alpha_1\zeta(\mathbf{k}_2\cdot\mathbf{p})\frac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{Z}(\mathbf{k}_1\cdot\mathbf{p})+$$
$$+\alpha_2\zeta(\mathbf{k}_1\cdot\mathbf{p})\frac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{Z}(\mathbf{k}_2\cdot\mathbf{p})+\mathcal{R}_4\ ,$$

where $\mathcal{R}_4$ is non–resonant. Finally we have the relation

$$\zeta(\mathbf{k}_1\cdot\mathbf{p})\zeta(\mathbf{k}_2\cdot\mathbf{p})\sin(\mathbf{k}_2\cdot\mathbf{q}) = \frac{\mathrm{d}}{\mathrm{d}\,t}\Big(\alpha_1\zeta(\mathbf{k}_2\cdot\mathbf{p})\mathcal{Z}(\mathbf{k}_1\cdot\mathbf{p})+ \\ + \alpha_2\zeta(\mathbf{k}_1\cdot\mathbf{p})\mathcal{Z}(\mathbf{k}_2\cdot\mathbf{p})\Big) - \alpha_1\zeta'(\mathbf{k}_2\cdot\mathbf{p})\mathcal{Z}(\mathbf{k}_1\cdot\mathbf{p})\mathbf{k}_2\cdot\dot{\mathbf{p}}+ \\ - \alpha_2\zeta'(\mathbf{k}_1\cdot\mathbf{p})\mathcal{Z}(\mathbf{k}_2\cdot\mathbf{p})\mathbf{k}_1\cdot\dot{\mathbf{p}} + \mathcal{R}_4 \ .$$

At this point one can check that the terms at the r.h.s. outside the time–derivative are non–resonant, and thus can be integrated by parts. The resulting terms are of order $\varepsilon\sigma^{3/2}$ (the terms which are triply resonating) and of order $\varepsilon^2\sigma^{-1}$ (the ones which are integrated by parts in a doubly resonating region of measure σ^2).

The other resonant term $\mathcal{N}_2$ can be treated in a similar way. One can check (using the explicit expressions of $\mathbf{k}_1$ and $\mathbf{k}_2$) that

$$\varepsilon\exp\big(i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big) = i(\mathbf{k}_1+\mathbf{k}_2)\cdot\dot{\mathbf{p}}\exp(i\mathbf{k}_1\cdot\mathbf{q}) + \varepsilon\mathcal{P}_1 \ ,$$

where $\varepsilon\mathcal{P}_1$ is a non–resonant trigonometric polinomial. One has then

$$\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)\frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\exp\big(i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big) = \\ \frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\exp(i\mathbf{k}_1\cdot\mathbf{q})\frac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{Z}\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big) + \mathcal{R}_5 \ ,$$

where $\mathcal{R}_5$ is non–resonant and can be integrated by parts. The first term at the r.h.s gives instead, as usual,

$$\frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\exp(i\mathbf{k}_1\cdot\mathbf{q})\frac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{Z}\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big) = \\ \frac{\mathrm{d}}{\mathrm{d}\,t}\left(\frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\exp(i\mathbf{k}_1\cdot\mathbf{q})\mathcal{Z}\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)\right) + \\ + \varepsilon\left(x\frac{\partial}{\partial x}\frac{1-\zeta(x)}{x}\right)\Bigg|_{\mathbf{k}_1\cdot\mathbf{p}}\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)\exp(i\mathbf{k}_1\cdot\mathbf{p}) + \mathcal{R}_6 \ ,$$

where $\mathcal{R}_6$ are non resonant terms. Instead, the second term at the r.h.s is again a resonant one, but it can be transformed into a total time–derivative (plus some non–resonant terms) using again a relation of the kind

$$\varepsilon\exp(i\mathbf{k}_1\cdot\mathbf{p}) = (\mathbf{k}_1+\mathbf{k}_2)\cdot\dot{\mathbf{p}} + \varepsilon\mathcal{P}_2 \ ;$$

where again $\mathcal{P}_2$ is a non-resonant trigonometrical polinomial. With some simple algebra one finally gets

$$\zeta\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)\frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\sin\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big)=$$
$$\frac{\mathrm{d}}{\mathrm{d}\,t}\Bigg[\mathcal{Z}\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)\frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\exp\big(i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{q}\big)$$
$$+\,\mathbf{k}_1\cdot\mathbf{p}\frac{\partial}{\partial\mathbf{k}_1\cdot\mathbf{p}}\frac{1-\zeta(\mathbf{k}_1\cdot\mathbf{p})}{\mathbf{k}_1\cdot\mathbf{p}}\mathcal{Z}^{(2)}\big((\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{p}\big)\Bigg]+\mathcal{R}_8\ ,$$

where $\mathcal{R}_8$ is a non–resonant term. In this way it is clear how is it possible to perform three steps of the construction, and at the same time how complicated becomes the procedure of performing further steps. In any case, performing the estimate and putting $\sigma=\varepsilon^{2/5}$, the estimate of Theorem 3 is obtained

6 Conclusions

In this paper a perturbative scheme is introduced which may be called "*uniform*", in the sense that it can be applied at the same time both in the nonresonant region and in the resonant ones.

While this approach does not allow to control individual trajectories, it is well suited to study the ergodic properties of the dynamics (at least for the case of rotators), through the estimate of an integral norm of suitable funtions. A distintive feature of this approach is the fact that it can be applied to systems with an arbitrarily large number of degrees of freedom (at least in the case of rotators), so that it can be of use for systems of interest in statistical mechanics.

The idea of using in perturbation theory an L^p integral norm instead of the familiar sup norm, which is a fundamental ingredient of the present paper, was apparently introduced for the first time by A. Neishtadt in the paper [10], where a non–Hamiltonian system with a small number of degrees of freedom was considered. I thank A. Neishtadt for kindly pointing this out to me, after the reading of a preliminary version of the present paper.

Appendix: proof of Lemma 2

The proof of Lemma 2 is quite standard (apart from property iii)) and can be found in any text–book in partial differential equations. Consider a $C^\infty([-1,1])$ function $y(x)$ having all derivatives vanishing for $x\to\pm1$

(take for example $y(x) = \exp(1/(x^2-1))$). Such a function can obviously be extended smoothly to the whole real line by setting it equal to zero outside that interval. Introduce now the auxiliary function

$$\tilde{\zeta}(x) = \frac{1}{C} \int\limits_{-2}^{2} y(t-x)\, \mathrm{d}\, t \ , \qquad C \stackrel{\mathrm{def}}{=} \int\limits_{-1}^{1} y(t)\, \mathrm{d}\, t \ .$$

In terms of $z = t - x$ one equivalently can write

$$\tilde{\zeta}(x) = \frac{1}{C} \int\limits_{[-2-x,\, 2-x] \bigcap [-1,1]} y(z)\, \mathrm{d}\, z \ ,$$

from which it is apparent that for $|x| < 1$ one has $\tilde{\zeta}(x) = 1$ (because in such a case one has $[-2-x, 2-x] \bigcap [-1,1] = [-1,1]$), while for $|x| > 3$ one has $\tilde{\zeta}(x) = 0$ (in such a case one has instead $[-2-x, 2-x] \bigcap [-1,1] = \emptyset$). If in addition one has $y(x) \geq 0$ one also gets $0 \leq \tilde{\zeta}(x) \leq 1$. It is obvious that $\tilde{\zeta}(x)$ is a C^∞ function, and one can define the constants

$$C'_n \stackrel{\mathrm{def}}{=} \sup_{|x| \leq 3} \left| \frac{\mathrm{d}^n}{\mathrm{d}x^n} \tilde{\zeta}(x) \right| \ . \tag{29}$$

In the case of the exponential function $y(x) = \exp(1/(x^2-1))$, simple (numerical) estimates for the first three constants are

$$C'_0 = 1 \ , \quad C'_1 < 2 \quad \text{and} \quad C'_2 < 21 \ ; \tag{30}$$

the other ones growing quite rapidly. It also obvious that all the derivatives vanish for $|x| < 1$ and $|x| > 3$.

We take now

$$\mathcal{Z}^{(2)}(x) \stackrel{\mathrm{def}}{=} \frac{x^2}{2} \tilde{\zeta}\left(\frac{x}{\sigma}\right) \ ,$$

and consequently, $\mathcal{Z}(x)$ being the derivative of $\mathcal{Z}^{(2)}(x)$ and $\zeta(x)$ the derivative of $\mathcal{Z}(x)$, one gets

$$\begin{aligned} \mathcal{Z}(x) &= x\, \tilde{\zeta}\left(\frac{x}{\sigma}\right) + \frac{x^2}{2\sigma} \tilde{\zeta}'\left(\frac{x}{\sigma}\right) \\ \zeta(x) &= \tilde{\zeta}\left(\frac{x}{\sigma}\right) + \frac{2x}{\sigma} \tilde{\zeta}'\left(\frac{x}{\sigma}\right) + \frac{x^2}{2\sigma^2} \tilde{\zeta}''\left(\frac{x}{\sigma}\right) \ . \end{aligned} \tag{31}$$

The functions $\zeta(x)$, $\mathcal{Z}(x)$ and $\mathcal{Z}^{(2)}(x)$ vanish for $|x| > 3\sigma$, while for $|x| < \sigma$ they reduce to $\zeta(x) = 1$, $\mathcal{Z}(x) = x$ and $\mathcal{Z}^{(2)}(x) = x^2/2$ (recall that the derivatives of $\tilde{\zeta}$ vanish for $|x|<1$). So i) of Lemma 2 is proved. To prove ii), one remarks that

$$
\begin{aligned}
\frac{\mathrm{d}^n\zeta(x)}{\mathrm{d}x^n} &= \frac{\mathrm{d}^{n+2}}{\mathrm{d}x^{n+2}}\left(\frac{x^2}{2}\tilde{\zeta}\left(\frac{x}{\sigma}\right)\right) \\
&= \frac{1}{\sigma^n}\left((n^2+3n+2)\frac{\mathrm{d}^n\tilde{\zeta}}{\mathrm{d}x^n} + \frac{(n+2)x}{\sigma}\frac{\mathrm{d}^{n+1}\tilde{\zeta}}{\mathrm{d}x^{n+1}} + \frac{x^2}{2\sigma^2}\frac{\mathrm{d}^{n+2}\tilde{\zeta}}{\mathrm{d}x^{n+2}}\right) ,
\end{aligned}
$$

so that, recalling the bound (29), the constant c_n can be taken equal to

$$
c_n = 5C'_{n+2} + 6(n+2)C'_{n+1} + (n+3n+2)C'_n \ .
$$

Part iii) of Lemma 2 follows directly from the definition and from the explict bound (30) for the constants C'_0, C'_1 and C'_2.

References

[1] G. Benettin, L. Galgani, A. Giorgilli, Comm. Math. Phys. **121**, 557 (1989).

[2] G. Benettin, L. Galgani, A. Giorgilli, Phys. Lett. A **120**, 23 (1987).

[3] D. Bambusi, A. Giorgilli, J. Stat. Phys. **71**, 569 (1993).

[4] D. Bambusi, A. Ponno, Chaos **15**, 015107 (2005);
D. Bambusi, A. Ponno, *On metastability in FPU*, Comm. Math. Phys., in print.

[5] B.O. Koopman, Proc. Nat. Ac. Sc. **17**, 315 (1931).

[6] J. von Neumann, Proc. Nat. Ac. Sc. **18**, 70 (1932);
B.O. Koopman, J. von Neumann, Proc. Nat. Ac. Sc. **18**, 255 (1932);
J. von Neumann, Proc. Nat. Ac. Sc. **18**, 263 (1932).

[7] L. Galgani, A. Giorgilli, A. Martinoli, S. Vanzini, Physica D **59**, 334 (1992).

[8] A. Giorgilli, Ann. Ist. H. Poincaré Phys. Théor. **48**, 423 (1988).

[9] G. Benettin, L. Galgani, A. Giorgilli, N. Cim. **89** B, 89 (1985).

[10] A. Neishtadt, Sov. Phys. Dokl. **21**, 80 (1976).