CrossMark

Birkhoff coordinates for the Toda lattice in the limit of infinitely many particles with an application to FPU

D. Bambusi [*], A. Maspero

Dipartimento di Matematica, Università degli Studi di Milano, Via Saldini 50, I-20133 Milano, Italy

A R T I C L E   I N F O

A B S T R A C T

In this paper we study the Birkhoff coordinates (Cartesian action angle coordinates) of the Toda lattice with periodic boundary condition in the limit where the number N of the particles tends to infinity. We prove that the transformation introducing such coordinates maps analytically a complex ball of radius R/N^α (in discrete Sobolev-analytic norms) into a ball of radius R'/N^α (with $R, R' > 0$ independent of N) if and only if $\alpha \geq 2$. Then we consider the problem of equipartition of energy in the spirit of Fermi–Pasta–Ulam. We deduce that corresponding to initial data of size R/N^2, $0 < R \ll 1$, and with only the first Fourier mode excited, the energy remains forever in a packet of Fourier modes exponentially decreasing with the wave number. Finally we consider the original FPU model and prove that energy remains localized in a similar packet of Fourier modes for times one order of magnitude longer than those covered by previous results which is the time of formation of the packet. The proof of the theorem on Birkhoff coordinates is based on a new quantitative version of a Vey type theorem by Kuksin and Perelman which could be interesting in itself.

© 2015 Elsevier Inc. All rights reserved.

 * Corresponding author.
    E-mail addresses: dario.bambusi@unimi.it (D. Bambusi), alberto.maspero@unimi.it (A. Maspero).

1. Introduction and main result

It is well known that the Toda lattice, namely the system with Hamiltonian

$$H_{Toda}(p,q) = \frac{1}{2}\sum_{j=0}^{N-1} p_j^2 + \sum_{j=0}^{N-1} e^{q_j - q_{j+1}} \ , \tag{1.1}$$

and periodic boundary conditions $q_N = q_0$, $p_N = p_0$, is integrable [36,21]. Thus, by standard Arnold–Liouville theory the system admits action angle coordinates. However the actual introduction of such coordinates is quite complicated (see [18,15]) and the corresponding transformation has only recently been studied analytically in a series of papers by Henrici and Kappeler [23,24]. In particular these authors have proved the existence of global Birkhoff coordinates, namely canonical coordinates (x_k, y_k) analytic on the whole $\mathbb{R}^{2N}$, with the property that the kth action is given by $(x_k^2 + y_k^2)/2$. The construction of Henrici and Kappeler, however is not uniform in the size of the chain, in the sense that the map Φ_N introducing Birkhoff coordinates is globally analytic for any fixed N, but it could (and actually does) develop singularities as $N \to +\infty$. Here we prove some analyticity properties fulfilled by Φ_N uniformly in the limit $N \to +\infty$. Precisely we consider complex balls centered at the origin and prove that Φ_N maps analytically a ball of radius R/N^α in discrete Sobolev-analytic norms into a ball of radius R'/N^α, with $R, R' > 0$ independent of N if and only if $\alpha \geq 2$. Furthermore we prove that the supremum of Φ_N over a complex ball of radius R/N^α diverges as $N \to +\infty$ when $\alpha < 1$.

In order to prove upper estimates on Φ_N we apply to the Toda lattice a Vey type theorem [39] for infinite dimensional systems recently proved by Kuksin and Perelman [30]. Actually, we need to prove a new quantitative version of Kuksin–Perelman's theorem. We think that such a result could be interesting in itself.

The lower estimates on the size of Φ_N are proved by constructing explicitly the first term of the Taylor expansion of Φ_N through Birkhoff normal form techniques; in particular we prove that the second differential $d^2\Phi_N(0)$ at the origin diverges like N^2.

We finally apply the result to the problem of equipartition of energy in the spirit of Fermi–Pasta–Ulam. We prove that in the Toda lattice, corresponding to initial data with energy E/N^3 ($0 < E \ll 1$) and with only the first Fourier mode excited, the energy remains forever in a packet of Fourier modes exponentially decreasing with the wave number. Then we consider the original FPU model and prove that, corresponding to the same initial data, energy remains in an exponentially localized packet of Fourier modes for times of order N^4 (see Theorem 1.16 below), namely for times one order of magnitude longer than those covered by previous results (see [2], see also [35,20]). This is relevant in view of the fact that the time scale of formation of the packet is N^3 (see [2]), so the result of the present paper allows to conclude that the packet persists over a time much longer than the one needed for its formation.

1.1. Birkhoff coordinates for the Toda lattice

We come to a precise statement of the main results of the present paper. Consider the Toda lattice in the subspace characterized by

$$\sum_j q_j = 0 = \sum_j p_j \tag{1.2}$$

which is invariant under the dynamics. Introduce the discrete Fourier transform $\mathcal{F}(q) = \hat{q}$ defined by

$$\hat{q}_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} q_j e^{2i\pi jk/N}, \qquad k \in \mathbb{Z}, \tag{1.3}$$

and consider $\hat{p}_k$ defined analogously. Due to (1.2) one has $\hat{p}_0 = \hat{q}_0 = 0$ and furthermore $\hat{p}_k = \hat{p}_{k+N}$, $\hat{q}_k = \hat{q}_{k+N}$, $\forall k \in \mathbb{Z}$, so we restrict to $\{\hat{p}_k, \hat{q}_k\}_{k=1}^{N-1}$. Corresponding to real sequences (p_j, q_j) one has $\overline{\hat{q}_k} = \hat{q}_{N-k}$ and $\overline{\hat{p}_k} = \hat{p}_{N-k}$.

Introduce the linear Birkhoff variables

$$X_k = \frac{\hat{p}_k + \hat{p}_{N-k} - i\omega_k(\hat{q}_k - \hat{q}_{N-k})}{\sqrt{2\omega_k}},$$

$$Y_k = \frac{\hat{p}_k - \hat{p}_{N-k} + i\omega_k(\hat{q}_k + \hat{q}_{N-k})}{i\sqrt{2\omega_k}}, \quad k = 1, \ldots, N-1, \tag{1.4}$$

where $\omega_k \equiv \omega\left(\frac{k}{N}\right) := 2\sin(k\pi/N)$; using such coordinates, which are symplectic, the quadratic part

$$H_0 := \sum_{j=0}^{N-1} \frac{p_j^2 + (q_j - q_{j+1})^2}{2} \tag{1.5}$$

of the Hamiltonian takes the form

$$H_0 = \sum_{k=1}^{N-1} \omega\left(\frac{k}{N}\right) \frac{X_k^2 + Y_k^2}{2}. \tag{1.6}$$

With an abuse of notations, we re-denote by H_{Toda} the Hamiltonian (1.1) written in the coordinates (X, Y). The following theorem is due to Henrici and Kappeler:

Theorem 1.1. *(See [24].) For any integer $N \geq 2$ there exists a global real analytic symplectic diffeomorphism $\Phi_N : \mathbb{R}^{N-1} \times \mathbb{R}^{N-1} \to \mathbb{R}^{N-1} \times \mathbb{R}^{N-1}$, $(X, Y) = \Phi_N(x, y)$ with the following properties:*

(i) The Hamiltonian $H_{Toda} \circ \Phi_N$ is a function of the actions $I_k := \frac{x_k^2 + y_k^2}{2}$ only, i.e. (x_k, y_k) are Birkhoff variables for the Toda Lattice.

(ii) The differential of Φ_N at the origin is the identity: $d\Phi_N(0, 0) = \mathbb{1}$.

Our main results concern the analyticity properties of the map Φ_N as $N \to \infty$. To come to a precise statement we have to introduce a suitable topology in $\mathbb{C}^{N-1} \times \mathbb{C}^{N-1}$.

For any $s \geq 0$, $\sigma \geq 0$ introduce in $\mathbb{C}^{N-1} \times \mathbb{C}^{N-1}$ the discrete Sobolev-analytic norm

$$\|(X, Y)\|_{\mathcal{P}^{s,\sigma}}^2 := \frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s} \, e^{2\sigma [k]_N} \, \omega\left(\tfrac{k}{N}\right) \frac{|X_k|^2 + |Y_k|^2}{2} \tag{1.7}$$

where

$$[k]_N := \min(|k|, |N - k|) \ .$$

The space $\mathbb{C}^{N-1} \times \mathbb{C}^{N-1}$ endowed by such a norm will be denoted by $\mathcal{P}^{s,\sigma}$. We denote by $B^{s,\sigma}(R)$ the ball of radius R and center 0 in the topology defined by the norm $\|.\|_{\mathcal{P}^{s,\sigma}}$. We will also denote by $B_{\mathbb{R}}^{s,\sigma} := B^{s,\sigma}(R) \cap (\mathbb{R}^{N-1} \times \mathbb{R}^{N-1})$ the *real* ball of radius R.

Remark 1.2. When $\sigma = s = 0$ the norm (1.7) coincides with the energy norm rescaled by a factor $1/N$ (the rescaling factor will be discussed in Remark 1.11). We are particularly interested in the case $\sigma > 0$ since, in such a case, states belonging to $\mathcal{P}^{s,\sigma}$ are exponentially decreasing in Fourier space. The consideration of positive values of s will be needed in the proof of the main theorem.

Our main result is the following theorem.

Theorem 1.3. *For any $s \geq 0$, $\sigma \geq 0$ there exist strictly positive constants $R_{s,\sigma}$, $C_{s,\sigma}$, such that for any $N \geq 2$, the map Φ_N is analytic as a map from $B^{s,\sigma}(R_{s,\sigma}/N^2)$ to $\mathcal{P}^{s,\sigma}$ and fulfills*

$$\sup_{\|(x,y)\|_{\mathcal{P}^{s,\sigma}} \leq R/N^2} \|\Phi_N(x, y) - (x, y)\|_{\mathcal{P}^{s+1,\sigma}} \leq C_{s,\sigma} \frac{R^2}{N^2} \ , \quad \forall R < R_{s,\sigma}. \tag{1.8}$$

The same estimate is fulfilled by the inverse map Φ_N^{-1} possibly with a different $R_{s,\sigma}$.

Remark 1.4. The estimate (1.8) controls the size of the nonlinear corrections in a norm which is stronger than the norm of (x, y), showing that $\Phi_N - \mathbb{1}$ is 1-smoothing. The proof of this kind of smoothing effect was actually the main aim of the work by Kuksin and Perelman [30], which proved it for KdV. Subsequently Kappeler, Schaad and Topalov [28] proved that such a smoothing property holds also globally for the KdV Birkhoff map.

Remark 1.5. As a consequence of (1.8) one has

$$\Phi_N\left(B^{s,\sigma}\left(\frac{R}{N^2}\right)\right) \subset B^{s,\sigma}\left(\frac{R}{N^2}(1+C_{s,\sigma}R)\right), \quad \forall R < R_{s,\sigma}, \forall N \geq 2 \qquad (1.9)$$

and the same estimate is fulfilled by the inverse map Φ_N^{-1}, possibly with a different $R_{s,\sigma}$.

Corollary 1.6. *For any $s \geq 0$, $\sigma \geq 0$ there exist strictly positive constants $R_{s,\sigma}$, $C_{s,\sigma}$, with the following property. Consider the solution $v(t) \equiv (X(t), Y(t))$ of the Toda Lattice corresponding to initial data $v_0 \in B^{s,\sigma}\left(\frac{R}{N^2}\right)$ with $R \leq R_{s,\sigma}$; then one has*

$$v(t) \in B^{s,\sigma}\left(\frac{R}{N^2}(1+C_{s,\sigma}R)\right), \quad \forall t \in \mathbb{R}. \qquad (1.10)$$

In order to state a converse of Theorem 1.3 consider the second differential $Q^{\Phi_N} := d^2\Phi_N(0,0)$ of Φ_N at the origin; $Q^{\Phi_N} : \mathcal{P}^{s,\sigma} \to \mathcal{P}^{s,\sigma}$ is a quadratic polynomial in the phase space variables.[1]

Theorem 1.7. *For any $s \geq 0$, $\sigma \geq 0$ there exist strictly positive R, C, $N_{s,\sigma} \in \mathbb{N}$, such that, for any $N \geq N_{s,\sigma}$, $\alpha \in \mathbb{R}$, the quadratic form Q^{Φ_N} fulfills*

$$\sup_{v \in B^{s,\sigma}_{\mathbb{R}}\left(\frac{R}{N^\alpha}\right)} \left\| Q^{\Phi_N}(v,v) \right\|_{\mathcal{P}^{s,\sigma}} \geq CR^2 N^{2-2\alpha}. \qquad (1.11)$$

Remark 1.8. Roughly speaking, one can say that, as $N \to \infty$, the real diffeomorphism Φ_N develops a singularity at zero in the second derivative.

Using Cauchy estimate (see Subsection 3.2) one immediately gets the following corollary.

Corollary 1.9. *Assume that for some $s \geq 0$, $\sigma \geq 0$ there exist strictly positive R, R' and $\alpha \geq 0$, $\alpha' \in \mathbb{R}$, $N_{s,\sigma} \in \mathbb{N}$, s.t., for any $N \geq N_{s,\sigma}$, the map Φ_N is analytic in the complex ball $B^{s,\sigma}(R/N^\alpha)$ and fulfills*

$$\Phi_N\left(B^{s,\sigma}\left(\frac{R}{N^\alpha}\right)\right) \subset B^{s,\sigma}\left(\frac{R'}{N^{\alpha'}}\right), \qquad (1.12)$$

then one has $\alpha' \leq 2(\alpha - 1)$.

Remark 1.10. A particular case of Corollary 1.9 is $\alpha < 1$, in which one has that the image of a ball of radius $RN^{-\alpha}$ under Φ_N is unbounded as $N \to \infty$.

A further interesting case is that of $\alpha = \alpha'$, which implies $\alpha \geq 2$, thus showing that the scaling R/N^2 is the best possible one in which a property of the kind of (1.9) holds.

---

[1] Actually according to the estimate (1.8) it is smooth as a map $\mathcal{P}^{s,\sigma} \to \mathcal{P}^{s+1,\sigma}$.

Remark 1.11. A state (X, Y) is in the ball $B^{s,\sigma}(R/N^2)$ if and only if there exist interpolating periodic functions (β, α), namely functions s.t.

$$p_j = \beta\left(\frac{j}{N}\right), \quad q_j - q_{j+1} = \alpha\left(\frac{j}{N}\right), \qquad (1.13)$$

which are analytic in a strip of width σ and have a Sobolev-analytic norm of size R/N^2. More precisely, given a state (p, q) one considers its Fourier coefficients $(\hat{p}, \hat{q})$ and the corresponding X, Y variables; define

$$\alpha(x) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{q}_k \left(1 - e^{-2\pi i k/N}\right) e^{-2\pi i x k}, \qquad \beta(x) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{p}_k e^{-2\pi i x k}$$

which fulfill (1.13). Then the Sobolev-analytic norms of α and β are controlled by $\|(X, Y)\|_{\mathcal{P}^{s,\sigma}}$. For example one has

$$\|(\alpha, \beta)\|_{H^s}^2 := \|\alpha\|_{L^2}^2 + \|\beta\|_{L^2}^2 + \frac{1}{(2\pi)^{2s}} \|\partial_x^s \alpha\|_{L^2}^2 + \frac{1}{(2\pi)^{2s}} \|\partial_x^s \beta\|_{L^2}^2 = \|(X, Y)\|_{\mathcal{P}^{s,0}}^2,$$

where $\|\alpha\|_{L^2}^2 := \int_0^1 |\alpha(x)|^2 \, dx$. In particular we consider here states with Sobolev-analytic norm of order R/N^2 with $R \ll 1$. The factor $1/N$ in the definition of the norm was introduced to get correspondence between the norm of a state and the norm of the interpolating functions.

Remark 1.12. As a consequence of Remark 1.11, the order in N of the solutions we are describing with Theorem 1.3 is the same of the solutions studied in the papers [2] and [3,5,4].

Remark 1.13. The results of Theorem 1.3 and Theorem 1.7 extend to states with discrete Sobolev–Gevrey norm defined by

$$\|(X, Y)\|_{\mathcal{P}^{s,\sigma,\nu}}^2 := \frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s} \, e^{2\sigma [k]_N^\nu} \, \omega\left(\frac{k}{N}\right) \frac{|X_k|^2 + |Y_k|^2}{2} \qquad (1.14)$$

where $0 \leq \nu \leq 1$. As a consequence of Remark 1.11, these states are interpolated by periodic functions with regularity Gevrey ν.

This paper is part of a project aiming at studying the dynamics of periodic Toda lattices with a large number of particles, in particular its asymptotics. First results in this project were obtained in the papers [3,5,4]. They are based on the Lax pair representation of the Toda lattice in terms of periodic Jacobi matrices. The spectrum of these matrices leads to a complete set of conserved quantities and hence determines the Toda Hamiltonian and the dynamics of Toda lattices, such as their frequencies. In order

to study the asymptotics of Toda lattices for a large number N of particles one therefore needs to work in two directions: on the one hand one has to study the asymptotics of the spectrum of Jacobi matrices as $N \to \infty$ and on the other hand, one needs to use tools of the theory of integrable systems in order to effectively extract information on the dynamics of Toda lattices from the periodic spectrum of periodic Jacobi matrices.

The limit of a class of sequences of $N \times N$ Jacobi matrices as $N \to \infty$ has been formally studied already at the beginning of the theory of the Toda lattices (see e.g. [36]). However, as pointed out in [5], these studies only allowed to (formally) compute the asymptotics of the spectrum in special cases. In particular, Toda lattices, which incorporated right and left moving waves could not be analyzed at all in this way. In [5], based on an approach pioneered in [11], the asymptotics of the spectra of sequences of Jacobi matrices corresponding to states of the form (1.13) were rigorously derived by the means of semiclassical analysis. It turns out that in such a limit the spectrum splits into three parts: one group of eigenvalues at each of the two edges of the spectrum within an interval of size $O(N^{-2})$, whose asymptotics are described by certain Hill operators, and a third group of eigenvalues, consisting of the bulk of the spectrum, whose asymptotics coincides with the one of Toda lattices at the equilibrium – see [5] for details.

In [4] the asymptotics of the eigenvalues obtained in [5] were used in order to compute the one of the actions and of the frequencies of Toda lattices. In particular it was shown that the asymptotics of the frequencies at the two edges involve the frequencies of two KdV solutions. The tools used in [4] are those of the theory of infinite dimensional integrable systems as developed in [27] and adapted to the Toda lattice in [23].

The present paper takes up another important topic in the large number of particle limit of periodic Toda lattices: we study the Birkhoff coordinates near the equilibrium in the limit of large N to provide precise estimates on the size of complex balls around the equilibrium in Fourier coordinates and the corresponding size in Birkhoff coordinates. Our analysis allows to describe the evolution of Toda lattices with large number of particles in the original coordinates and to obtain an application to the study of FPU lattices (on which we will comment in the next section).

We remark that the obtained estimates on the size of the complex balls are optimal. In our view this is a strong indication that beyond such a regime the standard tools of integrable systems become inadequate for studying the asymptotic features of the dynamics of the periodic Toda lattices as $N \to \infty$.

The proofs of our results are based on a novel technique developed in [30] to show a Vey type theorem for the KdV equation on the circle which we adapt here to the study of Toda lattices, developing in this way another tool for the study of periodic Toda lattices with a large number of particles. We remark that for our arguments to go through, we need to assume an additional smallness condition on the set of states admitted as initial data: the states are required to be interpolated by functions α and β with Sobolev-analytic norm of size R/N^2, with $R \ll 1$ sufficiently small. (In the papers [3,5,4], the size R can be arbitrarily large.)

1.2. On the FPU metastable packet

In this subsection we recall the phenomenon of the formation of a packet of modes in the FPU chain and state our related results. First of all we recall that the FPU (α, β)-model is the Hamiltonian lattice with Hamiltonian function which, in suitable rescaled variables, takes the form

$$H_{FPU}(p, q) = \sum_{j=0}^{N-1} \frac{p_j^2}{2} + U(q_j - q_{j+1}) \ , \tag{1.15}$$

$$U(x) = \frac{x^2}{2} + \frac{x^3}{6} + \beta \frac{x^4}{24} \ . \tag{1.16}$$

We will consider the case of periodic boundary conditions: $q_0 = q_N$, $p_0 = p_N$.

Remark 1.14. One has

$$H_{FPU}(p, q) = H_{Toda}(p, q) + (\beta - 1)H_2(q) + H^{(3)}(q),$$

where

$$H_l(q) := \sum_{j=0}^{N-1} \frac{(q_j - q_{j+1})^{l+2}}{(l+2)!} \ , \quad \forall l \geq 2 \ ,$$

$$H^{(3)} := -\sum_{l \geq 3} H_l \ .$$

Introduce the energies of the normal modes by

$$E_k := \frac{|\hat{p}_k|^2 + \omega \left(\frac{k}{N}\right)^2 |\hat{q}_k|^2}{2} \ , \quad 1 \leq k \leq N - 1 \ , \tag{1.17}$$

correspondingly denote by

$$\mathcal{E}_k := \frac{E_k}{N} \tag{1.18}$$

the specific energy in the kth mode. Note that since p, q are real variables, one has $\mathcal{E}_k = \mathcal{E}_{N-k}$.

In their celebrated numerical experiment Fermi, Pasta and Ulam [16], being interested in the problem of foundation of statistical mechanics, studied both the behavior of $\mathcal{E}_k(t)$ and of its time average

$$\langle \mathcal{E}_k \rangle(t) := \frac{1}{t} \int_0^t \mathcal{E}_k(s) ds \ .$$

They observed that, corresponding to initial data with $\mathcal{E}_1(0) \neq 0$ and $\mathcal{E}_k(0) = 0$ $\forall k \neq 1, N-1$, the quantities $\mathcal{E}_k(t)$ present a recurrent behavior, while their averages $\langle \mathcal{E}_k \rangle(t)$ quickly relax to a sequence $\bar{\mathcal{E}}_k$ exponentially decreasing with k. This is what is known under the name of FPU packet of modes.

Subsequent numerical observations have investigated the persistence of the phenomenon for large N and have also shown that after some quite long time scale (whose precise length is not yet understood) the averages $\langle \mathcal{E}_k \rangle(t)$ relax to equipartition (see e.g. [9,10,7,8]). This is the phenomenon known as metastability of the FPU packet.

The idea of exploiting the vicinity of FPU with Toda in order to study the dynamics of FPU goes back to [15], in which the authors performed some numerical investigations studying the evolution of the Toda invariants in the dynamics of FPU. A systematic numerical study of the evolution of the Toda invariants in FPU, paying particular attention to the dependence on N of the phenomena, was performed by Benettin and Ponno [7] (see also [8]). In particular such authors put into evidence the fact that the FPU packet seems to have an infinite lifespan in the Toda lattice. Furthermore they showed that the relevant parameter controlling the lifespan of the packet in the FPU model is the distance of FPU from the corresponding Toda lattice.

Our Theorem 1.3 yields as a corollary the effective existence and infinite persistence of the packet in the Toda lattice and also an estimate of its lifespan in the FPU system, estimate in which the effective parameter is the distance between Toda and FPU.

It is convenient to state the results for Toda and FPU using the small parameter

$$\mu := \frac{1}{N}$$

as in [2].

The following corollary is an immediate consequence of Corollary 1.6.

Corollary 1.15. *Consider the Toda lattice* (1.1). *Fix* $\sigma > 0$, *then there exist constants* R_0, C_1, *such that the following holds true. Consider an initial datum with*

$$\mathcal{E}_1(0) = \mathcal{E}_{N-1}(0) = R^2 e^{-2\sigma} \mu^4 \;\; , \quad \mathcal{E}_k(0) \equiv \mathcal{E}_k(t)\big|_{t=0} = 0 \;\; , \quad \forall k \neq 1, N-1 \quad (1.19)$$

with $R < R_0$. *Then, along the corresponding solution, one has*

$$\mathcal{E}_k(t) \leq R^2 (1 + C_1 R) \mu^4 e^{-2\sigma k} \;\; , \quad \forall 1 \leq k \leq \lfloor N/2 \rfloor \;\; , \quad \forall t \in \mathbb{R} \; . \quad (1.20)$$

For the FPU model we have the following corollary

Theorem 1.16. *Consider the FPU system* (1.15). *Fix* $s \geq 1$ *and* $\sigma \geq 0$; *then there exist constants* R'_0, C_2, T, *such that the following holds true. Consider a real initial datum fulfilling* (1.19) *with* $R < R'_0$, *then, along the corresponding solution, one has*

$$\mathcal{E}_k(t) \leq \frac{16R^2\mu^4 e^{-2\sigma k}}{k^{2s}} \ , \quad \forall 1 \leq k \leq \lfloor N/2 \rfloor \ , \quad |t| \leq \frac{T}{R^2\mu^4} \cdot \frac{1}{|\beta - 1| + C_2 R\mu^2} \cdot$$

(1.21)

Furthermore, for $1 \leq k \leq N-1$, consider the action $I_k := \frac{x_k^2 + y_k^2}{2}$ of the Toda lattice and let $I_k(t)$ be its evolution according to the FPU flow. Then one has

$$\frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2(s-1)} e^{2\sigma[k]_N} \omega\left(\frac{k}{N}\right) |I_k(t) - I_k(0)| \leq C_3 R^2 \mu^5 \qquad \text{for } t \text{ fulfilling (1.21)}$$

(1.22)

Remark 1.17. The estimates (1.21) are stronger than the corresponding estimates given in [2], which are

$$\mathcal{E}_k(t) \leq C_1\mu^4 e^{-\sigma k} + C_2\mu^5 \ , \quad \forall 1 \leq k \leq \lfloor N/2 \rfloor \ , \quad |t| \leq \frac{T}{\mu^3}.$$

First, the time scale of validity of (1.21) is one order longer than that of [2]. Second we show that as β approaches the value corresponding to the Toda lattice (1 in our units) the time of stability improves. Third the exponential estimate of $\mathcal{E}_k$ as a function of k is shown to hold also for large values of k (the μ^5 correction is missing). Finally in [2] it was shown that T/μ^3 is the time of formation of the metastable packet. So we can now conclude that the time of persistence of the packet is at least one order of magnitude larger (namely μ^{-4}) with respect to the time needed for its formation.

Remark 1.18. We recall also the result of [20] in which the authors obtained a control of the dynamics for longer time scales, but for initial data with much smaller energies.

Remark 1.19. Recently some results on energy sharing in FPU in the thermodynamic limit [31] (see also [12,13,19]) have also been obtained, however such results are not able to explain the formation and the stability of the FPU packet of modes.

Remark 1.20. In this paper we did not address the observation of near-quasiperiodicity of solutions in the numerical experiments on FPU. One can think to use KAM theory in order to prove existence of quasiperiodic motions in FPU. This is possible in view of the fact that Toda lattice has good action angle coordinates and that the action to frequency map in the Toda lattice is nondegenerate. Results more or less in this line have been proved in [34,25], however the results of these papers do not persist in the limit $N \to \infty$, and it is very hard to identify the dependence on N of the threshold for the applicability of KAM theory to FPU.

2. A quantitative Kuksin–Perelman Theorem

2.1. Statement of the theorem

In this section we state and prove a quantitative version of Kuksin–Perelman Theorem which will be used to prove Theorem 1.3. It is convenient to formulate it in the framework of weighted ℓ^2 spaces, that we are going now to recall.

For any $N \leq \infty$, given a sequence $w = \{w_k\}_{k=1}^N$, $w_k > 0 \ \forall k \geq 1$, consider the space ℓ^2_w of complex sequences $\xi = \{\xi_k\}_{k=1}^N$ with norm

$$\|\xi\|_w^2 := \sum_{k=1}^N w_k^2 |\xi_k|^2 < \infty. \tag{2.1}$$

Denote by $\mathcal{P}^w$ the complex Banach space $\mathcal{P}^w := \ell^2_w \oplus \ell^2_w \ni (\xi, \eta)$ endowed with the norm $\|(\xi, \eta)\|_w^2 := \|\xi\|_w^2 + \|\eta\|_w^2$. We denote by $\mathcal{P}^w_{\mathbb{R}}$ the real subspace of $\mathcal{P}^w$ defined by

$$\mathcal{P}^w_{\mathbb{R}} := \left\{ (\xi, \eta) \in \mathcal{P}^w \ : \ \eta_k = \bar{\xi}_k \ \forall 1 \leq k \leq N \right\}. \tag{2.2}$$

We will denote by $B^w(\rho)$ (respectively $B^w_{\mathbb{R}}(\rho)$) the ball in the topology of $\mathcal{P}^w$ (respectively $\mathcal{P}^w_{\mathbb{R}}$) with center 0 and radius $\rho > 0$.

Remark 2.1. In the case of the Toda lattice the variables (ξ, η) are defined by

$$\xi_k = \frac{\hat{p}_k + i\omega\left(\frac{k}{N}\right)\hat{q}_k}{\sqrt{2\omega\left(\frac{k}{N}\right)}}, \quad \eta_k = \frac{\hat{p}_{N-k} - i\omega\left(\frac{k}{N}\right)\hat{q}_{N-k}}{\sqrt{2\omega\left(\frac{k}{N}\right)}}, \quad 1 \leq k \leq N-1, \tag{2.3}$$

and their connection with the real Birkhoff variables is given by

$$X_k = \frac{\xi_k + \eta_k}{\sqrt{2}}, \quad Y_k = \frac{\xi_k - \eta_k}{i\sqrt{2}}, \quad 1 \leq k \leq N-1. \tag{2.4}$$

We denote by $\mathcal{P}^1$ the Banach space of sequences in which all the weights w_k are equal to 1. For $\mathcal{X}$, $\mathcal{Y}$ Banach spaces, we shall write $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ to denote the set of linear and bounded operators from $\mathcal{X}$ to $\mathcal{Y}$. For $\mathcal{X} = \mathcal{Y}$ we will write just $\mathcal{L}(\mathcal{X})$.

Remark 2.2. In the application to the Toda lattice with N particles we will use a finite, but not fixed N and weights of the form $w_k^2 = w_{N-k}^2 = N^3 k^{2s} e^{2\sigma k}$, $1 \leq k \leq \lfloor N/2 \rfloor$.

Given two weights w^1 and w^2, we will say that $w^1 \leq w^2$ iff $w_k^1 \leq w_k^2$, $\forall k$. Sometimes, when there is no risk of confusion, we will omit the index w from the different quantities.

In $\mathcal{P}^1$ we will use the scalar product

$$\left\langle (\xi^1, \eta^1), (\xi^2, \eta^2) \right\rangle_c := \sum_{k=1}^{N} \xi_k^1 \bar{\xi}_k^2 + \eta_k^1 \bar{\eta}_k^2 \ . \tag{2.5}$$

Correspondingly, the scalar product and symplectic form on the real subspace $\mathcal{P}_{\mathbb{R}}^w$ are given for $\xi^1 \equiv (\xi^1, \bar{\xi}^1)$ and $\xi^2 \equiv (\xi^2, \bar{\xi}^2)$ by

$$\left\langle \xi^1, \xi^2 \right\rangle := 2Re \sum_{k=1}^{N} \xi_k^1 \overline{\xi_k^2} \ , \qquad \omega_0(\xi^1, \xi^2) := \left\langle E\, \xi^1, \xi^2 \right\rangle \ , \tag{2.6}$$

where $E := -i$.

Given a smooth $F : \mathcal{P}_{\mathbb{R}}^w \to \mathbb{C}$, we denote by X_F the Hamiltonian vector field of F, given by $X_F = J\nabla F$, where $J = E^{-1}$. For $F, G : \mathcal{P}_{\mathbb{R}}^w \to \mathbb{C}$ we denote by $\{F, G\}$ the Poisson bracket (with respect to ω_0): $\{F, G\} := \left\langle \nabla F, J\nabla G \right\rangle$ (provided it exists). We say that the functions F, G commute if $\{F, G\} = 0$.

In order to state the main abstract theorem we start by recalling the notion of normally analytic map, exploited also in [33] and [1].

First we recall that a map $\tilde{P}^r : (\mathcal{P}^w)^r \to \mathcal{B}$, with $\mathcal{B}$ a Banach space, is said to be *r-multilinear* if $\tilde{P}^r(v^{(1)}, \dots, v^{(r)})$ is linear in each variable $v^{(j)} \equiv (\xi^{(j)}, \eta^{(j)})$; a *r-multi-linear* map is said to be *bounded* if there exists a constant $C > 0$ such that

$$\left\| \tilde{P}^r(v^{(1)}, \dots, v^{(r)}) \right\|_{\mathcal{B}} \leq C \left\| v^{(1)} \right\|_w \cdots \left\| v^{(r)} \right\|_w \qquad \forall v^{(1)}, \dots, v^{(r)} \in \mathcal{P}^w.$$

Correspondingly its norm is defined by

$$\left\| \tilde{P}^r \right\| := \sup_{\| v^{(1)} \|_w, \dots, \| v^{(r)} \|_w \leq 1} \left\| \tilde{P}^r(v^{(1)}, \dots, v^{(r)}) \right\|_{\mathcal{B}}.$$

A map $P^r : \mathcal{P}^w \to \mathcal{B}$ is a *homogeneous polynomial* of order r if there exists a r-multilinear map $\tilde{P}^r : (\mathcal{P}^w)^r \to \mathcal{B}$ such that

$$P^r(v) = \tilde{P}^r(v, \dots, v) \quad \forall v \in \mathcal{P}^w \ . \tag{2.7}$$

A r-homogeneous polynomial is bounded if it has finite norm

$$\| P^r \| := \sup_{\| v \|_w \leq 1} \| P^r(v) \|_{\mathcal{B}} \ .$$

Remark 2.3. Clearly $\| P^r \| \leq \| \tilde{P}^r \|$. Furthermore one has $\left\| \tilde{P}^r \right\| \leq e^r \| P^r \|$ – cf. [32].

It is easy to see that a multilinear map and the corresponding polynomial are continuous (and analytic) if and only if they are bounded.

Let $P^r : \mathcal{P}^w \to \mathcal{B}$ be a homogeneous polynomial of order r; assume $\mathcal{B}$ separable and let $\{\mathbf{b}_n\}_{n \geq 1} \subset \mathcal{B}$ be a basis for the space $\mathcal{B}$. Expand P^r as follows

$$P^r(v) \equiv P^r(\xi, \eta) = \sum_{\substack{|K|+|L|=r \\ n \geq 1}} P_{K,L}^{r,n} \xi^K \eta^L \mathbf{b}_n, \tag{2.8}$$

where $K, L \in \mathbb{N}_0^N$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $|K| := K_1 + \cdots + K_N$, $\xi \equiv \{\xi_j\}_{j \geq 1}$ and $\xi^K \equiv \xi_1^{K_1} \cdots \xi_N^{K_N}$, $\eta^L \equiv \eta_1^{L_1} \cdots \eta_N^{L_N}$.

Definition 2.4. The modulus of a polynomial P^r is the polynomial $\underline{P}^r$ defined by

$$\underline{P}^r(\xi, \eta) := \sum_{\substack{|K|+|L|=r \\ n \geq 1}} \left| P_{K,L}^{r,n} \right| \xi^K \eta^L \mathbf{b}_n. \tag{2.9}$$

A polynomial P^r is said to have *bounded modulus* if $\underline{P}^r$ is a bounded polynomial.

A map $F : \mathcal{P}^w \to \mathcal{B}$ is said to be an *analytic germ* if there exists $\rho > 0$ such that $F : B^w(\rho) \to \mathcal{B}$ is analytic. Then F can be written as a power series absolutely and uniformly convergent in $B^w(\rho)$: $F(v) = \sum_{r \geq 0} F^r(v)$. Here $F^r(v)$ is a homogeneous polynomial of degree r in the variables $v = (\xi, \eta)$. We will write $F = O(v^n)$ if in the previous expansion $F^r(v) = 0$ for every $r < n$.

Definition 2.5. An analytic germ $F : \mathcal{P}^w \to \mathcal{B}$ is said to be *normally analytic* if there exists $\rho > 0$ such that

$$\underline{F}(v) := \sum_{r \geq 0} \underline{F}^r(v) \tag{2.10}$$

is absolutely and uniformly convergent in $B^w(\rho)$. In such a case we will write $F \in \mathcal{N}_\rho(\mathcal{P}^w, \mathcal{B})$. $\mathcal{N}_\rho(\mathcal{P}^w, \mathcal{B})$ is a Banach space when endowed by the norm

$$|\underline{F}|_\rho := \sup_{v \in B^w(\rho)} \|\underline{F}(v)\|_{\mathcal{B}}. \tag{2.11}$$

Let $U \subset \mathcal{P}_{\mathbb{R}}^w$ be open. A map $F : U \to \mathcal{B}$ is said to be a *real analytic germ* (respectively *real normally analytic*) on U if for each point $u \in U$ there exist a neighborhood V of u in $\mathcal{P}^w$ and an analytic germ (respectively normally analytic germ) which coincides with F on $U \cap V$.

Remark 2.6. It follows from Cauchy inequality that the Taylor polynomials F^r of F satisfy

$$\|\underline{F}^r(v)\|_{\mathcal{B}} \leq |\underline{F}|_\rho \frac{\|v\|_w^r}{\rho^r} \qquad \forall v \in B^w(\rho) . \tag{2.12}$$

Remark 2.7. Since $\forall r \geq 1$ one has $\|F^r\| \leq \|\underline{F}^r\|$, if $F \in \mathcal{N}_\rho(\mathcal{P}^w, \mathcal{B})$ then the Taylor series of F is uniformly convergent in $B^w(\rho)$.

The case $\mathcal{B} = \mathcal{P}^w$ will be of particular importance; in this case the basis $\{\mathbf{b}_j\}_{j\geq 1}$ will coincide with the natural basis $\{\mathbf{e}_j\}_{j\geq 1}$ of such a space (namely the vectors with all components equal to zero except the jth one which is equal to 1). We will consider also the case $\mathcal{B} = \mathcal{L}(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$ (bounded linear operators from $\mathcal{P}^{w^1}$ to $\mathcal{P}^{w^2}$), where w^1 and w^2 are weights. Here the chosen basis is $\mathbf{b}_{jk} = \mathbf{e}_j \otimes \mathbf{e}_k$ (labeled by 2 indexes).

Remark 2.8. For $v \equiv (\xi, \eta) \in \mathcal{P}^1$, we denote by $|v|$ the vector of the modulus of the components of v: $|v| = (|v_1|, \ldots, |v_N|)$, $|v_j| := (|\xi_j|, |\eta_j|)$. If $F \in \mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$ then $\underline{dF}(|v|)|u| \leq d\underline{F}(|v|)|u|$ (see [30]) and therefore, for any $0 < d < 1$, Cauchy estimates imply that $dF \in \mathcal{N}_{(1-d)\rho}(\mathcal{P}^{w^1}, \mathcal{L}(\mathcal{P}^{w^1}, \mathcal{P}^{w^2}))$ with

$$|\underline{dF}|_{\rho(1-d)} \leq \frac{1}{d\rho} |\underline{F}|_\rho \ , \tag{2.13}$$

where $\underline{dF}$ is computed with respect to the basis $\mathbf{e}_j \otimes \mathbf{e}_k$.

Following Kuksin and Perelman [30] we will need also a further property.

Definition 2.9. A normally analytic germ $F \in \mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$ will be said to be of class $\mathcal{A}_{w^1, \rho}^{w^2}$ if $F = O(v^2)$ and the map $v \mapsto dF(v)^* \in \mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{L}(\mathcal{P}^{w^1}, \mathcal{P}^{w^2}))$. Here $dF(v)^*$ is the adjoint operator of $dF(v)$ with respect to the standard scalar product (2.5). On $\mathcal{A}_{w^1, \rho}^{w^2}$ we will use the norm

$$\|F\|_{\mathcal{A}_{w^1, \rho}^{w^2}} := |\underline{F}|_\rho + \rho\,|\underline{dF}|_\rho + \rho\,|\underline{dF^*}|_\rho \,. \tag{2.14}$$

Remark 2.10. Assume that for some $\rho > 0$ the map $F \in \mathcal{A}_{w^1, \rho}^{w^2}$, then for every $0 < d \leq \frac{1}{2}$ one has $|\underline{F}|_{d\rho} \leq 2d^2 |\underline{F}|_\rho$ and $\|F\|_{\mathcal{A}_{w^1, d\rho}^{w^2}} \leq 6d^2 \|F\|_{\mathcal{A}_{w^1, \rho}^{w^2}}$.

A real normally analytic germ $F : B_{\mathbb{R}}^{w^1}(\rho) \to \mathcal{P}_{\mathbb{R}}^{w^2}$ will be said to be of class $\mathcal{N}_\rho(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{P}_{\mathbb{R}}^{w^2})$ (respectively $\mathcal{A}_{w^1, \rho}^{w^2}$) if there exists a map of class $\mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$ (respectively $\mathcal{A}_{w^1, \rho}^{w^2}$), which coincides with F on $B_{\mathbb{R}}^{w^1}(\rho)$. In this case we will also denote by $|\underline{F}|_\rho$ (respectively $\|F\|_{\mathcal{A}_{w^1, \rho}^{w^2}}$) the norm defined by (2.11) (respectively (2.14)) of the complex extension of F.

Let now $F : U \subset \mathcal{P}^{w^1} \to \mathcal{P}^{w^2}$ be an analytic map. We will say that F *is real for real sequences* if $F(U \cap \mathcal{P}_{\mathbb{R}}^{w^1}) \subseteq \mathcal{P}_{\mathbb{R}}^{w^2}$, namely $F(\xi, \eta) = (F_1(\xi, \eta), F_2(\xi, \eta))$ satisfies $\overline{F_1(\xi, \bar{\xi})} = F_2(\xi, \bar{\xi})$. Clearly, the restriction $F|_{U \cap \mathcal{P}_{\mathbb{R}}^{w^1}}$ is a real analytic map.

We come now to the statement of the Vey Theorem.

Fix $\rho > 0$ and let $\Psi : B_{\mathbb{R}}^{w^1}(\rho) \to \mathcal{P}_{\mathbb{R}}^{w^1}$, $\Psi = \mathbb{1} + \Psi^0$ with $\mathbb{1}$ the identity map and $\Psi^0 \in \mathcal{A}_{w^1, \rho}^{w^2}$. Write Ψ component-wise, $\Psi = \left\{ (\Psi_j, \overline{\Psi}_j) \right\}_{j \geq 1}$, and consider the foliation

defined by the functions $\left\{ |\Psi_j(v)|^2 /2 \right\}_{j \geq 1}$. Given $v \in \mathcal{P}^w_{\mathbb{R}}$ we define the leaf through v by

$$\mathcal{F}_v := \left\{ u \in \mathcal{P}^w_{\mathbb{R}} : \; \frac{|\Psi_j(u)|^2}{2} = \frac{|\Psi_j(v)|^2}{2} \; , \; \forall j \geq 1 \right\} . \qquad (2.15)$$

Let $\mathcal{F} = \bigcup_{v \in \mathcal{P}^w_{\mathbb{R}}} \mathcal{F}_v$ be the collection of all the leaves of the foliation. We will denote by $T_v \mathcal{F}$ the tangent space to $\mathcal{F}_v$ at the point $v \in \mathcal{P}^w_{\mathbb{R}}$. A relevant role will also be played by the function $I = \{I_j\}_{j \geq 1}$ whose components are defined by

$$I_j(v) \equiv I_j(\xi, \bar{\xi}) := \frac{|\xi_j|^2}{2} \quad \forall j \geq 1 . \qquad (2.16)$$

The foliation they define will be denoted by $\mathcal{F}^{(0)}$.

Remark 2.11. Ψ maps the foliation $\mathcal{F}$ into the foliation $\mathcal{F}^{(0)}$, namely $\mathcal{F}^{(0)} = \Psi(\mathcal{F})$.

The main theorem of this section is the following

Theorem 2.12 (*Quantitative version of Kuksin–Perelman Theorem*). *Let w^1 and w^2 be weights with $w^1 \leq w^2$. Consider the space $\mathcal{P}^{w^1}_{\mathbb{R}}$ endowed with the symplectic form ω_0 defined in (2.6). Let $\rho > 0$ and assume $\Psi : B^{w^1}_{\mathbb{R}}(\rho) \to \mathcal{P}^{w^1}_{\mathbb{R}}$, $\Psi = \mathbb{1} + \Psi^0$ and $\Psi^0 \in \mathcal{A}^{w^2}_{w^1,\rho}$. Define*

$$\epsilon_1 := \left\| \Psi^0 \right\|_{\mathcal{A}^{w^2}_{w^1,\rho}} . \qquad (2.17)$$

Assume that the functionals $\{\frac{1}{2} |\Psi_j(v)|^2\}_{j \geq 1}$ pairwise commute with respect to the symplectic form ω_0, and that ρ is so small that

$$\epsilon_1 < 2^{-34} \rho. \qquad (2.18)$$

Then there exists a real normally analytic map $\widetilde{\Psi} : B^{w^1}_{\mathbb{R}}(a\rho) \to \mathcal{P}^{w^1}_{\mathbb{R}}$, $a = 2^{-48}$, with the following properties:

i) *$\widetilde{\Psi}^* \omega_0 = \omega_0$, so that the coordinates $z := \widetilde{\Psi}(v)$ are canonical;*
ii) *the functionals $\left\{ \frac{1}{2} \left| \widetilde{\Psi}_j(v) \right|^2 \right\}_{j \geq 1}$ pairwise commute with respect to the symplectic form ω_0;*
iii) *$\mathcal{F}^{(0)} = \widetilde{\Psi}(\mathcal{F})$, namely the foliation defined by Ψ coincides with the foliation defined by $\widetilde{\Psi}$;*
iv) *$\widetilde{\Psi} = \mathbb{1} + \widetilde{\Psi}^0$ with $\widetilde{\Psi}^0 \in \mathcal{A}^{w^2}_{w^1,a\rho}$ and $\left\| \widetilde{\Psi}^0 \right\|_{\mathcal{A}^{w^2}_{w^1,a\rho}} \leq 2^{17} \epsilon_1.$*

The following corollary holds:

Corollary 2.13. *Let $H : \mathcal{P}_{\mathbb{R}}^{w^1} \to \mathbb{R}$ be a real analytic Hamiltonian function. Let Ψ be as in Theorem 2.12 and assume that for every $j \geq 1$, $|\Psi_j(v)|^2$ is an integral of motion for H, i.e.*

$$\{H, |\Psi_j|^2\} = 0 \quad \forall j \geq 1. \tag{2.19}$$

Then the coordinates (x_j, y_j) defined by $x_j + iy_j = \widetilde{\Psi}_j(v)$ are real Birkhoff coordinates for H, namely canonical conjugated coordinates in which the Hamiltonian depends only on $(x_j^2 + y_j^2)/2$.

Proof. Since $\Psi = \mathbb{1} + \Psi^0$, the functions $\Psi_j(v)$ can be used as coordinates in a suitable neighborhood of 0 in $\mathcal{P}_{\mathbb{R}}^w$. Let $\widetilde{\Psi}$ be the map in the statement of Theorem 2.12. Denote $F_l(v) := \frac{1}{2}\left|\widetilde{\Psi}_l(v)\right|^2$. Since the foliation defined by the functions $\{F_l\}_{l \geq 1}$ and the foliation defined by $\{|\Psi_j|^2\}_{j \geq 1}$ coincide (Theorem 2.12*iii*)), each F_l is constant on the level sets of $\{|\Psi_j|^2\}_{j \geq 1}$. It follows that each F_l is a function of $\{|\Psi_j|^2\}_{j \geq 1}$ only. Since $\forall j \geq 1$, $|\Psi_j|^2$ is an integral of motion for H, the same is true for F_l, $\forall l \geq 1$. Define now, in a suitable neighborhood of the origin, the coordinates $(z, \bar{z})$ by $z_j \equiv \widetilde{\Psi}_j$, $\bar{z}_j \equiv \overline{\widetilde{\Psi}_j}$. Of course $F_l = \frac{|z_l|^2}{2}$. By (2.19) it follows then that

$$0 = \{H, z_l\bar{z}_l\} = \frac{1}{i}\left(\frac{\partial H}{\partial z_l}z_l - \frac{\partial H}{\partial \bar{z}_l}\bar{z}_l\right). \tag{2.20}$$

Since $d\widetilde{\Psi}(0) = \mathbb{1}$ (Theorem 2.12*iv*)), $\widetilde{\Psi}$ is invertible and its inverse $\widetilde{\Phi}$ satisfies $\widetilde{\Phi} = \mathbb{1} + \widetilde{\Phi}^0$ with $\widetilde{\Phi}^0 \in \mathcal{A}_{w^1,a\mu\rho}^{w^2}$ and $\left\|\widetilde{\Phi}^0\right\|_{\mathcal{A}_{w^1,a\mu\rho}^{w^2}} \leq 2\left\|\widetilde{\Psi}^0\right\|_{\mathcal{A}_{w^1,a\rho}^{w^2}} \leq 2^{18}\epsilon_1$ (Lemma A.3*ii*) in Appendix A).

Expand now $H \circ \widetilde{\Phi}$ in Taylor series in the variables $(z, \bar{z})$:

$$H \circ \widetilde{\Phi}(z, \bar{z}) = \sum_{\substack{r \geq 2, \\ |\alpha| + |\beta| = r}} H_{\alpha,\beta}^r z^\alpha \bar{z}^\beta.$$

Then equation (2.20) implies that in each term of the summation $\alpha = \beta$, therefore $H \circ \widetilde{\Phi}$ is a function of $|z_1|^2, \ldots, |z_N|^2$. Define now the real variables (x, y) as in the statement, then the claim follows immediately. $\quad\square$

2.2. Proof of the quantitative Kuksin–Perelman Theorem

In this section we recall and adapt Eliasson's proof [14] of the Vey Theorem following [30]. As we anticipated in the introduction, the novelty of our approach is to add quantitative estimates on the Birkhoff map $\widetilde{\Psi}$ of Theorem 2.12. In Appendix A we show that the class of normally analytic maps is closed under several operations like composition,

inversion and flow-generation, and provide new quantitative estimates which will be used during the proof below.

The idea of the proof of Theorem 2.12 is to consider the functions $\{\Psi_j(v)\}_{j\geq 1}$ as non-canonical coordinates, and to look for a coordinate transformation introducing canonical variables and preserving the foliation $\mathcal{F}^{(0)}$ (which is the image of $\mathcal{F}$ in the noncanonical variables).

This will be done in two steps both based on the standard procedure of Darboux Theorem that we now recall. In order to construct a coordinate transformation φ transforming the closed nondegenerate form Ω_1 into a closed nondegenerate form Ω_0, then it is convenient to look for φ as the time 1 flow φ^t of a time-dependent vector field Y^t. To construct Y^t one defines $\Omega_t := \Omega_0 + t(\Omega_1 - \Omega_0)$ and imposes that

$$0 = \tfrac{d}{dt}\big|_{t=0} \varphi^{t*}\Omega_t = \varphi^{t*}\left(\mathcal{L}_{Y^t}\Omega_t + \Omega_1 - \Omega_0\right) = \varphi^{t*}\left(d(Y^t \lrcorner \Omega_t) + d(\alpha_1 - \alpha_0)\right)$$

where α_1, α_0 are potential forms for Ω_1 and Ω_0 (namely $d\alpha_i = \Omega_i$, $i = 0, 1$) and $\mathcal{L}_{Y^t}$ is the Lie derivative of Y^t. Then one gets

$$Y^t \lrcorner \Omega_t + \alpha_1 - \alpha_0 = df \tag{2.21}$$

for each f smooth; then, if Ω_t is nondegenerate, this defines Y^t. If Y^t generates a flow φ^t defined up to time 1, the map $\varphi := \varphi^t\big|_{t=1}$ satisfies $\varphi^*\Omega_1 = \Omega_0$. Thus, given Ω_0 and Ω_1, the whole game reduces to study the analytic properties of Y^t and to prove that it generates a flow.

A non-constant symplectic form Ω will always be represented through a linear skew-symmetric invertible operator E as follows:

$$\Omega(v)(u^{(1)}; u^{(2)}) = \langle E(v)u^{(1)}; u^{(2)}\rangle , \quad \forall u^{(1)}, u^{(2)} \in T_v \mathcal{P}_\mathbb{R}^w \simeq \mathcal{P}_\mathbb{R}^w. \tag{2.22}$$

We denote by $\{F, G\}_\Omega$ the Poisson bracket with respect to Ω: $\{F, G\}_\Omega := \langle \nabla F, J\nabla G\rangle$, $J := E^{-1}$.

Similarly we will represent 1-forms through the vector field A such that

$$\alpha(v)(u) = \langle A(v), u\rangle, \quad \forall u \in T_v \mathcal{P}_\mathbb{R}^w. \tag{2.23}$$

Define $\omega_1 := (\Psi^{-1})^*\omega_0$, and let E_{ω_1} be the operator representing the symplectic form ω_1. The first step consists in transforming ω_1 to a symplectic form whose "average over $\mathcal{F}^{(0)}$" coincides with ω_0.

So we start by defining precisely what "average of k-forms" means. To this end consider the Hamiltonian vector fields $X_{I_l}^0$ of the functions $I_l \equiv \frac{|v_l|^2}{2}$ through the symplectic form ω_0; they are given by

$$X_{I_l}^0(v) = \mathrm{i}\nabla I_l(v) = \mathrm{i}v_l \mathbf{e}_l, \quad \forall l \geq 1. \tag{2.24}$$

For every $l \geq 1$ the corresponding flow $\phi_l^t \equiv \phi_{X_{I_l}^0}^t$ is given by

$$\phi_l^t(v) = \left(v_1, \cdots, v_{l-1}, e^{it}v_l, v_{l+1}, \cdots\right) .$$

Remark that the map ϕ_l^t is linear in v, 2π periodic in t and its adjoint satisfies $(\phi_l^t)^* = \phi_l^{-t}$.

Given a k-form α on $\mathcal{P}_{\mathbb{R}}^w$ ($k \geq 0$), we define its average by

$$M_j\alpha(v) = \frac{1}{2\pi}\int_0^{2\pi}((\phi_j^t)^*\alpha)(v)dt, \quad j \geq 1 , \qquad \text{and} \qquad M\alpha(v) = \int_{\mathcal{T}}[(\phi^\theta)^*\alpha]\,d\theta$$

(2.25)

where $\mathcal{T}$ is the (possibly infinite dimensional) torus, the map $\phi^\theta = (\phi_1^{\theta_1} \circ \phi_2^{\theta_2} \cdots)$ and $d\theta$ is the Haar measure on $\mathcal{T}$.

Remark 2.14. In the particular cases of 1 and 2-forms it is useful to compute the average in terms of the representations (2.22) and (2.23). Thus, for $v, u^{(1)}, u^{(2)} \in \mathcal{P}_{\mathbb{R}}^w$, if

$$\alpha(v)u^{(1)} = \langle A(v); u^{(1)}\rangle , \quad \omega(v)(u^{(1)}, u^{(2)}) = \langle E(v)\,u^{(1)}; u^{(2)}\rangle ,$$

one has

$$(M\alpha)(v)u^{(1)} = \langle (MA)(v); u^{(1)}\rangle , \quad \text{with} \quad MA(v) = \int_{\mathcal{T}}\phi^{-\theta}A(\phi^\theta(v))\,d\theta \quad (2.26)$$

and

$$(M\omega)(v)(u^{(1)}, u^{(2)}) = \langle (ME)(v)u^{(1)}; u^{(2)}\rangle , \quad \text{with} \quad ME(v) = \int_{\mathcal{T}}\phi^{-\theta}E(\phi^\theta(v))\phi^\theta\,d\theta.$$

(2.27)

Remark 2.15. The operator M commutes with the differential operator d and the rotations ϕ^θ. In particular $MA(v)$ and $ME(v)$ as in (2.26), (2.27) satisfy

$$\phi^\theta MA(v) = MA(\phi^\theta v), \quad \phi^\theta ME(v)u = ME(\phi^\theta v)\phi^\theta u, \qquad \forall \theta \in \mathcal{T} .$$

We study now the analytic properties of ω_1 and of its potential form α_{ω_1}. In the rest of the section denote by $S := \sum_{n=1}^{\infty} 1/n^2$ and by

$$\mu := 1/e(32S)^{1/2} \approx 0.0507 . \tag{2.28}$$

Lemma 2.16. *Let $\Phi := \Psi^{-1}$ and ω_1 be as above. Assume that $\epsilon_1 \leq \rho/e$. Then the following holds:*

(i) $E_{\omega_1} = -\mathrm{i} + \Upsilon_{\omega_1}$, *with* $\Upsilon_{\omega_1} \in \mathcal{N}_{\mu\rho}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{L}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{P}_{\mathbb{R}}^{w^2}))$ *and*

$$|\underline{\Upsilon_{\omega_1}}|_{\mu\rho} \leq \frac{8\epsilon_1}{\mu\rho} . \tag{2.29}$$

(ii) Define

$$W_{\omega_1}(v) := \int_0^1 \Upsilon_{\omega_1}(tv)tv \; dt , \tag{2.30}$$

then $W_{\omega_1} \in \mathcal{A}_{w^1, \mu^3\rho}^{w^2}$ and $\|W_{\omega_1}\|_{\mathcal{A}_{w^1, \mu^3\rho}^{w^2}} \leq 8\epsilon_1$. Moreover the 1-form $\alpha_{W_{\omega_1}} := \langle W_{\omega_1}; . \rangle$ satisfies $d\alpha_{W_{\omega_1}} = \omega_1 - \omega_0$.

Proof. By Lemma A.3 one has that $\Phi = (\mathbb{1} + \Psi^0)^{-1} = \mathbb{1} + \Phi^0$ with $\Phi^0 \in \mathcal{A}_{w^1, \mu\rho}^{w^2}$ and $\|\Phi^0\|_{\mathcal{A}_{w^1, \mu\rho}^{w^2}} \leq 2 \|\Psi^0\|_{\mathcal{A}_{w^1, \rho}^{w^2}} \leq 2\epsilon_1$. To prove (i), just remark that

$$E_{\omega_1}(v) = d\Phi^*(v)(-\mathrm{i})d\Phi(v) = -\mathrm{i} + d\Phi^0(v)^*(-\mathrm{i})d\Phi(v) - \mathrm{i}d\Phi^0(v) =: -\mathrm{i} + \Upsilon_{\omega_1}(v)$$

and use the results of Lemma A.3. To prove (ii), use Poincaré construction of the potential of ω_1 which gives

$$\alpha_{\omega_1}(v)u := \langle \int_0^1 E_{\omega_1}(tv)tv, u \rangle dt = \alpha_0(v)u + \langle W_{\omega_1}(v), u \rangle, \quad W_{\omega_1}(v) = \int_0^1 \Upsilon_{\omega_1}(tv)tv \; dt ,$$

where α_0 is the potential for ω_0. In order to prove the analytic properties of W_{ω_1}, note that $W_{\omega_1}(v) = \int_0^1 (H_1(tv) + H_2(tv))dt$ where $H_1(v) := -\mathrm{i} \, d\Phi^0(v)v$ and $H_2(v) := d\Phi^0(v)^*(-\mathrm{i})d\Phi(v)v \equiv d\Phi^0(v)^*(-\mathrm{i}v + H_1(v))$. Thus, by Lemma A.3, one gets that $\|H_1\|_{\mathcal{A}_{w^1, \mu^2\rho}^{w^2}} \leq 2 \|\Phi^0\|_{\mathcal{A}_{w^1, \mu\rho}^{w^2}} \leq 4\epsilon_1$ and $\|H_2\|_{\mathcal{A}_{w^1, \mu^3\rho}^{w^2}} \leq 2 \|\Phi^0\|_{\mathcal{A}_{w^1, \mu^2\rho}^{w^2}} \leq 4\epsilon_1$. Thus the estimate on W_{ω_1} follows. $\square$

Remark 2.17. One has $M\alpha_{\omega_1} - \alpha_0 = M\alpha_{W_{\omega_1}} = \langle MW_{\omega_1}, \cdot \rangle$ and $\|MW_{\omega_1}\|_{\mathcal{A}_{w^1, \mu^3\rho}^{w^2}} \leq \|W_{\omega_1}\|_{\mathcal{A}_{w^1, \mu^3\rho}^{w^2}}$.

We are ready now for the first step.

Lemma 2.18. *There exists a map $\hat{\varphi} : B_{\mathbb{R}}^{w^1}(\mu^5\rho) \to \mathcal{P}_{\mathbb{R}}^{w^1}$ such that $(\mathbb{1} - \hat{\varphi}) \in \mathcal{A}_{w^1, \mu^5\rho}^{w^2}$ and*

$$\|\mathbb{1} - \hat{\varphi}\|_{\mathcal{A}_{w^1, \mu^5\rho}^{w^2}} \leq 2^5\epsilon_1 . \tag{2.31}$$

Moreover $\hat\varphi$ satisfies the following properties:

(i) *$\hat\varphi$ commutes with the rotations ϕ^θ, namely $\phi^\theta \hat\varphi(v) = \hat\varphi(\phi^\theta v)$ for every $\theta \in \mathcal{T}$.*
(ii) *Denote $\hat\omega_1 := \hat\varphi^* \omega_1$, then $M\hat\omega_1 = \omega_0$.*

Proof. We apply the Darboux procedure described at the beginning of this section with $\Omega_0 = \omega_0$ and $\Omega_1 = M\omega_1$. Then Ω_t is represented by the operator $\hat{E}^t_{\omega_1} := (-\mathrm{i} + t(ME_{\omega_1} + \mathrm{i}))$. Write equation (2.21), with $f \equiv 0$, in terms of the operators defining the symplectic forms, getting the equation $\hat{E}^t_{\omega_1} \hat{Y}^t = -MW_{\omega_1}$ (see also Remark 2.17). This equation can be solved by inverting the operator $\hat{E}^t_{\omega_1}$ by Neumann series:

$$\hat{Y}^t := -(-\mathrm{i} + tM\Upsilon_{\omega_1})^{-1} MW_{\omega_1} \ . \tag{2.32}$$

By the results of Lemma 2.16 and Remark 2.17, $\hat{Y}^t$ is of class $\mathcal{A}^{w^2}_{w^1, \mu^4 \rho}$ and fulfills

$$\sup_{t \in [0,1]} \left\| \hat{Y}^t \right\|_{\mathcal{A}^{w^2}_{w^1, \mu^4 \rho}} \leq 2 \left\| MW_{\omega_1} \right\|_{\mathcal{A}^{w^2}_{w^1, \mu^3 \rho}} \leq 2^4 \epsilon_1 \ . \tag{2.33}$$

By Lemma A.4 the vector field $\hat{Y}^t$ generates a flow $\hat\varphi^t : B^{w^1}_\mathbb{R}(\mu^5 \rho) \to \mathcal{P}^{w^1}$ such that $\hat\varphi^t - \mathbb{1}$ is of class $\mathcal{A}^{w^2}_{w^1, \mu^5 \rho}$ and satisfies

$$\left\| \hat\varphi^t - \mathbb{1} \right\|_{\mathcal{A}^{w^2}_{w^1, \mu^5 \rho}} \leq 2 \sup_{t \in [0,1]} \left\| \hat{Y}^t \right\|_{\mathcal{A}^{w^2}_{w^1, \mu^4 \rho}} \leq 2^5 \epsilon_1.$$

Therefore the map $\hat\varphi \equiv \hat\varphi^t|_{t=1}$ exists, satisfies the claimed estimate (2.31) and furthermore $\hat\varphi^* M\omega_1 = \omega_0$.

We prove now item (i). The claim follows if we show that the vector field $\hat{Y}^t$ commutes with rotations. To this aim consider equation (2.32), and define $\hat{J}^t_{\omega_1}(v) = (\hat{E}^t_{\omega_1}(v))^{-1}$. By construction the operator $\hat{E}^t_{\omega_1}$ commutes with rotations (cf. Remark 2.15), namely $\forall \theta_0 \in \mathcal{T}$ one has $\phi^{\theta_0} \hat{E}^t_{\omega_1}(v)u = \hat{E}^t_{\omega_1}(\phi^{\theta_0}(v))\phi^{\theta_0} u$. Then it follows that

$$\phi^{\theta_0} \hat{Y}^t(v) = -\phi^{\theta_0} \hat{J}^t_{\omega_1}(v) MW_{\omega_1}(v) = -\hat{J}^t_{\omega_1}(\phi^{\theta_0}(v))\phi^{\theta_0} MW_{\omega_1}(v)$$

$$= -\hat{J}^t_{\omega_1}(\phi^{\theta_0}(v)) MW_{\omega_1}(\phi^{\theta_0}(v)) = \hat{Y}^t(\phi^{\theta_0}(v)).$$

This proves item (i). Item (ii) then follows from item (i) since, defining $\hat\omega_1 = \hat\varphi^* \omega_1$, one has the chain of identities $M\hat\omega_1 = M\hat\varphi^* \omega_1 = \hat\varphi^* M\omega_1 = \omega_0$. $\quad\square$

The analytic properties of the symplectic form $\hat\omega_1$ can be studied in the same way as in Lemma 2.16; we get therefore the following corollary:

Corollary 2.19. *Denote by $E_{\hat{\omega}_1}$ the symplectic operator describing $\hat{\omega}_1 = \hat{\varphi}^* \omega_1$. Then*

(i) $E_{\hat{\omega}_1} = -\mathrm{i} + \Upsilon_{\hat{\omega}_1}$, with $\Upsilon_{\hat{\omega}_1} \in \mathcal{N}_{\mu^5 \rho}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{L}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{P}_{\mathbb{R}}^{w^2}))$ and $\left|\underline{\Upsilon_{\hat{\omega}_1}}\right|_{\mu^5 \rho} \leq 2^7 \frac{\epsilon_1}{\mu \rho}$.

(ii) Define $W(v) := \int_0^1 \Upsilon_{\hat{\omega}_1}(tv) tv \, dt$, then $W \in \mathcal{A}_{w^1, \mu^7 \rho}^{w^2}$ and $\|W\|_{\mathcal{A}_{w^1, \mu^7 \rho}^{w^2}} \leq 2^7 \epsilon_1$.

Furthermore the 1-form $\alpha_W := \langle W, . \rangle$ satisfies $d\alpha_W = \hat{\omega}_1 - \omega_0$.

Finally we will need also some analytic and geometric properties of the map

$$\check{\Psi} := \hat{\varphi}^{-1} \circ \Psi. \tag{2.34}$$

The functions $\{\check{\Psi}(v)\}_{j \geq 1}$ forms a new set of coordinates in a suitable neighborhood of the origin whose properties are given by the following corollary:

Corollary 2.20. *The map $\check{\Psi} : B_{\mathbb{R}}^{w^1}(\mu^8 \rho) \to \mathcal{P}_{\mathbb{R}}^{w^1}$, defined in (2.34), satisfies the following properties:*

(i) $d\check{\Psi}(0) = \mathbb{1}$ and $\check{\Psi}^0 := \check{\Psi} - \mathbb{1} \in \mathcal{A}_{w^1, \mu^8 \rho}^{w^2}$ with $\left\|\check{\Psi}^0\right\|_{\mathcal{A}_{w^1, \mu^8 \rho}^{w^2}} \leq 2^8 \epsilon_1$.

(ii) $\mathcal{F}^{(0)} = \check{\Psi}(\mathcal{F})$, namely the foliation defined by $\check{\Psi}$ coincides with the foliation defined by Ψ.

(iii) The functionals $\{\frac{1}{2}\left|\check{\Psi}_j\right|^2\}_{j \geq 1}$ pairwise commute with respect to the symplectic form ω_0.

Proof. By Lemma A.3 the map $\hat{\varphi}$ is invertible in $B_{\mathbb{R}}^{w^1}(\mu^6 \rho)$ and $\hat{\varphi}^{-1} = \mathbb{1} + g$, with $g \in \mathcal{A}_{w^1, \mu^6 \rho}^{w^2}$ and $\|g\|_{\mathcal{A}_{w^1, \mu^6 \rho}^{w^2}} \leq 2^6 \epsilon_1$. Then $\check{\Psi} = \mathbb{1} + \check{\Psi}^0$ where $\check{\Psi}^0 = \Psi^0 + g \circ (\mathbb{1} + \Psi^0)$. By Remark 2.10, $\left\|\Psi^0\right\|_{\mathcal{A}_{w^1, \mu^7 \rho}^{w^2}} \leq 6\mu^{14}\epsilon_1$, thus Lemma A.3i) implies that $\check{\Psi}^0 \in \mathcal{A}_{w^1, \mu^8 \rho}^{w^2}$ and moreover $\left\|\check{\Psi}^0\right\|_{\mathcal{A}_{w^1, \mu^8 \rho}^{w^2}} \leq 6\mu^{14}\epsilon_1 + 2^7\epsilon_1 \leq 2^8\epsilon_1$. Items (ii) and (iii) follow from the fact that, by Lemma 2.18(i), $\hat{\varphi}$ commutes with the rotations (see also the proof of Corollary 2.13). $\quad\square$

The second step consists in transforming $\hat{\omega}_1$ into the symplectic form ω_0 while preserving the functions I_l. In order to perform this transformation, we apply once more the Darboux procedure with $\Omega_1 = \hat{\omega}_1$ and $\Omega_0 = \omega_0$. However, we require each leaf of the foliation to be invariant under the transformation. In practice, we look for a change of coordinates φ satisfying

$$\varphi^* \Omega_1 = \Omega_0 , \tag{2.35}$$

$$I_l(\varphi(v)) = I_l(v), \quad \forall l \geq 1 . \tag{2.36}$$

In order to fulfill the second equation, we take advantage of the arbitrariness of f in equation (2.21). It turns out that if f satisfies the set of differential equations given by

$$df(X_{I_l}^0) - (\alpha_1 - \alpha_0)(X_{I_l}^0) = 0, \quad \forall l \geq 1 \tag{2.37}$$

then equation (2.36) is satisfied (as it will be proved below). Here α_1 is the potential form of $\hat{\omega}_1$ and is given by $\alpha_1 := \alpha_0 + \alpha_W$, where α_W is defined in Corollary 2.19. However, (2.37) is essentially a system of equations for the potential of a 1-form on a torus, so there is a solvability condition. In Lemma 2.23 below we will prove that the system (2.37) has a solution if the following conditions are satisfied:

$$d(\alpha_1 - \alpha_0)|_{T\mathcal{F}^{(0)}} = 0 \, , \tag{2.38}$$

$$M(\alpha_1 - \alpha_0)|_{T\mathcal{F}^{(0)}} = 0 \, . \tag{2.39}$$

In order to show that these two conditions are fulfilled, we need a preliminary result. First, for $v \in \mathcal{P}_{\mathbb{R}}^w$ fixed, define the symplectic orthogonal of $T_v\mathcal{F}^{(0)}$ with respect to the form $\omega^t := \omega_0 + t(\hat{\omega}_1 - \omega_0)$ by

$$(T_v\mathcal{F}^{(0)})^{\angle_t} := \left\{ h \in \mathcal{P}_{\mathbb{R}}^w : \omega^t(v)(u, h) = 0 \ \forall u \in T_v\mathcal{F}^{(0)} \right\}. \tag{2.40}$$

Lemma 2.21. *For $v \in B_{\mathbb{R}}^{w^1}(\mu^5\rho)$, one has $T_v\mathcal{F}^{(0)} = (T_v\mathcal{F}^{(0)})^{\angle_t}$.*

Proof. First of all we have that, since for any couple of functions F, G and any change of coordinates Φ, one has

$$\{F \circ \Phi, G \circ \Phi\}_{\Phi^*\omega_0} = \{F, G\}_{\omega_0} \circ \Phi \, ,$$

it follows that

$$\{I_l, I_m\}_{\omega_1} = \left\{ |\Psi_l|^2, |\Psi_m|^2 \right\}_{\omega_0} = 0 \, , \quad \forall l, m \geq 1$$

and

$$\{I_l, I_m\}_{\hat{\omega}_1} \circ \hat{\varphi}^{-1} = \left\{ I_l \circ \hat{\varphi}^{-1}, I_m \circ \hat{\varphi}^{-1} \right\}_{\omega_1}$$

but, by the property of invariance with respect to rotations of $\hat{\varphi}$ (and therefore of $\hat{\varphi}^{-1}$), $I_j \circ \hat{\varphi}^{-1}$ is a function of $\{I_l\}_{l \geq 1}$ only, and therefore the above quantity vanishes and one has $\forall l, m$

$$0 = \{I_l(v), I_m(v)\}_{\hat{\omega}_1} = \langle \nabla I_l(v), J_{\hat{\omega}_1}(v)\nabla I_m(v) \rangle = \langle v_l\mathbf{e}_l, J_{\hat{\omega}_1}(v)v_m\mathbf{e}_m \rangle \quad \forall l, m \geq 1.$$

$$\tag{2.41}$$

Define $\Sigma_v := \mathrm{span}\,\{v_l \mathbf{e}_l,\ l \geq 1\}$. The identities (2.41) imply that $J_{\hat{\omega}_1}(v)(\Sigma_v) \subseteq \Sigma_v^{\perp} \equiv \mathrm{i}\Sigma_v$. By Corollary 2.19$(i)$, $E_{\hat{\omega}_1}(v)$ is an isomorphism for $v \in B_{\mathbb{R}}^{w^1}(\mu^5 \rho)$, so the same is true for its inverse $J_{\hat{\omega}_1}(v)$. Hence $J_{\hat{\omega}_1}(v)(\Sigma_v) = \mathrm{i}\Sigma_v$ and $\Sigma_v = E_{\hat{\omega}_1}(v)(\mathrm{i}\Sigma_v)$ and

$$\hat{\omega}_1(X_{I_l}^0, X_{I_m}^0) = \langle E_{\hat{\omega}_1}(v)(\mathrm{i}v_l \mathbf{e}_l), \mathrm{i}v_m \mathbf{e}_m \rangle = 0, \quad \forall\, l, m \geq 1. \tag{2.42}$$

Since ω^t is a linear combination of ω_0 and $\hat{\omega}_1$, the previous formula implies that $\omega^t(v)(X_{I_l}^0, X_{I_m}^0) = 0$ for every $t \in [0, 1]$ and $v \in B_{\mathbb{R}}^{w^1}(\mu^5 \rho)$, hence $T_v \mathcal{F}^{(0)} \subseteq (T_v \mathcal{F}^{(0)})^{\angle_t}$. Now assume by contradiction that the inclusion is strict: then there exists $u \in (T_v \mathcal{F}^{(0)})^{\angle_t}$, $\|u\| = 1$, such that $u \notin T_v \mathcal{F}^{(0)}$. Decompose $u = u_\top + u_\perp$ with $u_\top \in T_v \mathcal{F}^{(0)}$ and $u_\perp \in (T_v \mathcal{F}^{(0)})^{\perp}$. Due to the bilinearity of $\omega(v)^t$, we can always assume that $u \equiv u_\perp$. Then for every $l \geq 1$

$$dI_l(v)(-\mathrm{i}u) = \langle \nabla I_l(v), -\mathrm{i}u \rangle = \langle -\mathrm{i}X_{I_l}^0(v), -\mathrm{i}u \rangle = \langle X_{I_l}^0(v), u \rangle = 0 \qquad \forall\, l \geq 1$$

since $X_{I_l}^0(v) \in T_v \mathcal{F}^{(0)}$. Hence $\mathrm{i}u \in T_v \mathcal{F}^{(0)}$ and therefore $\omega^t(v)(-\mathrm{i}u, u) = 0$. Furthermore it holds that

$$\omega^t(0)(\mathrm{i}u, u) = \omega_0(-\mathrm{i}u, u) = \langle \mathrm{i}^2 u, u \rangle = -1.$$

It follows that for $v \in B_{\mathbb{R}}^{w^1}(\mu^5 \rho)$ one has $\|tM\Upsilon_{\hat{\omega}_1}(v)\|_{\mathcal{L}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{P}_{\mathbb{R}}^{w^1})} \leq 1/2$, thus $\omega^t(v)(\mathrm{i}u, u) = -1 + \langle tM\Upsilon_{\hat{\omega}_1}(v)\mathrm{i}u, u \rangle < 0$, leading to a contradiction. $\quad\square$

We can now prove the following lemma:

Lemma 2.22. *The solvability conditions* (2.38), (2.39) *are fulfilled.*

Proof. Condition (2.38) follows by equation (2.42), since

$$d(\alpha_1 - \alpha_0)(X_{I_l}^0, X_{I_m}^0) = \hat{\omega}_1(X_{I_l}^0, X_{I_m}^0) - \omega_0(X_{I_l}^0, X_{I_m}^0) = 0, \qquad \forall\, l, m \geq 1.$$

We analyze now (2.39). We claim that in order to fulfill this condition, one must have that $\hat{\omega}_1$ satisfies $M\hat{\omega}_1 = \omega_0$, which holds by Lemma 2.18(ii). Indeed, since

$$0 = M\hat{\omega}_1 - \omega_0 = M(\hat{\omega}_1 - \omega_0) = Md(\alpha_1 - \alpha_0) = dM(\alpha_1 - \alpha_0),$$

there exists a function g such that $M(\alpha_1 - \alpha_0) = dg$. But $Mdg = M(M(\alpha_1 - \alpha_0)) = M(\alpha_1 - \alpha_0) = dg$, therefore $g = Mg$, so g is invariant by rotations. Hence $0 = \frac{d}{dt}\big|_{t=0}\, g(\phi_l^t) = dg(X_{I_l}^0) = M(\alpha_1 - \alpha_0)(X_{I_l}^0)$, $\forall\, l \geq 1$, thus also (2.39) is satisfied. $\quad\square$

We show now that the system (2.37) can be solved and its solution has good analytic properties:

Lemma 2.23 *(Moser). If conditions* (2.38) *and* (2.39) *are fulfilled, then equation* (2.37) *has a solution f. Moreover, denoting $h_j := (\alpha_1 - \alpha_0)(X^0_{I_j})$, the solution f is given by the explicit formula*

$$f(v) = \sum_{j=1}^{\infty} f_j(v), \qquad f_j(v) = M_1 \cdots M_{j-1} L_j h_j \tag{2.43}$$

where

$$L_j g = \frac{1}{2\pi} \int_0^{2\pi} t g(\phi_j^t) dt .$$

Finally $f \in \mathcal{N}_{\mu^7 \rho}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathbb{C})$, $\nabla f \in \mathcal{N}_{\mu^7 \rho}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{P}_{\mathbb{R}}^{w^2})$ and

$$\left| \underline{f} \right|_{\mu^7 \rho} \leq 2^{10} \epsilon_1 \mu^7 \rho, \qquad \left| \underline{\nabla f} \right|_{\mu^7 \rho} \leq 2^{11} \epsilon_1 . \tag{2.44}$$

Proof. Denote by θ_j the time along the flow generated by $X^0_{I_j}$, then one has $dg(X^0_{I_j}) = \frac{\partial g}{\partial \theta_j}$, so that the equations to be solved take the form

$$\frac{\partial f}{\partial \theta_j} = h_j, \qquad \forall j \geq 1. \tag{2.45}$$

Clearly $\frac{\partial}{\partial \theta_j} M_j h_j = 0$, and by (2.38) it follows that

$$\frac{\partial}{\partial \theta_l} M_j h_j = M_j \frac{\partial h_j}{\partial \theta_l} = M_j \frac{\partial h_l}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} M_j h_l = 0, \qquad \forall l, j \geq 1,$$

which shows that $M_j h_j$ is independent of all the θ's, thus $M_j h_j = M h_j$. Furthermore, by (2.39) one has $M h_j = 0$, $\forall j \geq 1$. Now, using that $\frac{\partial}{\partial \theta_j} L_j g = g - M_j g$, one verifies that f_j defined in (2.43) satisfies

$$\frac{\partial f_j}{\partial \theta_l} = \begin{cases} 0 & \text{if } l < j \\ M_1 \cdots M_{j-1} h_j & \text{if } l = j \\ M_1 \cdots M_{j-1} h_l - M_1 \cdots M_j h_l & \text{if } l > j \end{cases}$$

where, for $j = 1$, we defined $M_1 \cdots M_{j-1} h_l = h_l$. Thus the series $f(v) := \sum_{j \geq 1} f_j(v)$, if convergent, satisfies (2.45).

We prove now the convergence of the series for f and ∇f. First we define, for $\theta \in \mathcal{T}$,

$$\Theta_j^\theta := \phi_1^{\theta_1} \cdots \phi_j^{\theta_j} \qquad \forall j \geq 1 ,$$

then by (2.43) one has

$$f_j(v) = \int_{\mathcal{T}^j} \theta_j h_j(\Theta_j^\theta v)\, d\theta^j \ , \tag{2.46}$$

$$\nabla f_j(v) = \int_{\mathcal{T}^j} \Theta_j^{-\theta}\theta_j \nabla h_j(\Theta_j^\theta v)\, d\theta^j \ , \tag{2.47}$$

where $\mathcal{T}^j$ is the j-dimensional torus and $d\theta^j = \frac{d\theta_1}{2\pi}\cdots\frac{d\theta_j}{2\pi}$. Now, using that

$$h_j(v) = \langle W(v), X_{I_j}^0(v)\rangle = Re(\mathrm{i}W_j(v)\bar{v}_j) \qquad \forall j \geq 1$$

one gets that $\underline{f_j}(|v|) \leq 2\pi\, \underline{h_j}(|v|) \leq 2\pi\, \underline{W_j}(|v|)|v_j|$, therefore $\underline{f}(|v|) \leq \sum_{j=1}^{\infty} \underline{f_j}(|v|) \leq 2\pi\, \|\underline{W}(|v|)\|_{w^1}\, \|v\|_{w^1}$ and it follows that $\left|\underline{f}\right|_{\mu^7\rho} \leq 2\pi\, |\underline{W}|_{\mu^7\rho}\,\mu^7\rho$. This proves the convergence of the series defining f.

Consider now the gradient of h_j, whose kth component is given by

$$[\nabla h_j(v)]_k = Re\left(\mathrm{i}\frac{\partial W_j(v)}{\partial v_k}\bar{v}_j\right) + \delta_{j,k}\, Re\,(\mathrm{i}W_j(v))\ .$$

Inserting the formula displayed above in (2.47) we get that ∇f_j is the sum of two terms. We begin by estimating the second one, which we denote by $(\nabla f_j)^{(2)}$. The kth component of $(\nabla f)^{(2)} := \sum_j (\nabla f_j)^{(2)}$ is given by

$$\left[(\nabla f(v))^{(2)}\right]_k = \left[\sum_j (\nabla f_j(v))^{(2)}\right]_k = \int_{\mathcal{T}^k} \Theta_k^{-\theta}\theta_k\, Re\,(\mathrm{i}W_k(\Theta_k^\theta v))\, d\theta^k \ , \tag{2.48}$$

thus, for any $v \in B_{\mathbb{R}}^{w^1}(\mu^7\rho)$ one has $\left[(\underline{\nabla f}(|v|))^{(2)}\right]_k \leq 2\pi\, \underline{W_k}(|v|)$, and therefore

$$\left|(\underline{\nabla f})^{(2)}\right|_{\mu^7\rho} \leq 2\pi\, |\underline{W}|_{\mu^7\rho} \leq \pi 2^8 \epsilon_1.$$

We come to the other term, which we denote by $(\nabla f_j)^{(1)}$. Its kth component is given by

$$\left[(\nabla f_j(v))^{(1)}\right]_k = \int_{\mathcal{T}^j} \Theta_j^{-\theta}\theta_j Re\left(\mathrm{i}\frac{\partial W_j}{\partial v_k}(\Theta_j^\theta v)\overline{\phi_j^{\theta_j}v_j}\right) d\theta \ . \tag{2.49}$$

Then $\underline{\nabla f_j}(|v|) \leq 2\pi\frac{\overline{\partial W_j}}{\partial v_k}(|v|)|v_j| = 2\pi[\underline{dW}(|v|)]_k^j |v_j|$.

It follows that the kth component of the function $(\nabla f)^{(1)} := \sum_j (\nabla f_j)^{(1)}$ satisfies

$$\left[(\underline{\nabla f}(|v|))^{(1)}\right]_k \leq \left[\sum_j (\underline{\nabla f_j}(|v|))^{(1)}\right]_k \leq 2\pi \sum_j [\underline{dW}(|v|)]_k^j |v_j| \ .$$

Therefore $\left|(\nabla f)^{(1)}\right|_{\mu^7 \rho} \leq 2\pi \left\|W\right\|_{\mathcal{A}^{w^2}_{w^1,\mu^7 \rho}} \leq \pi 2^8 \epsilon_1$. This is the step at which the control of the norm of the modulus $\underline{dW^*}$ of dW^* is needed. Thus the claimed estimate for ∇f follows. $\square$

We can finally apply the Darboux procedure in order to construct an analytic change of coordinates φ which satisfies (2.35) and (2.36).

Lemma 2.24. *There exists a map* $\varphi : B^{w^1}_{\mathbb{R}}(\mu^9 \rho) \to \mathcal{P}^{w^1}_{\mathbb{R}}$ *which satisfies* (2.35). *Moreover* $\varphi - \mathbb{1} \in \mathcal{N}_{\mu^9 \rho}(\mathcal{P}^{w^1}_{\mathbb{R}}, \mathcal{P}^{w^2}_{\mathbb{R}})$, $\varphi - \mathbb{1} = O(v^2)$ *and*

$$\left|\varphi - \mathbb{1}\right|_{\mu^9 \rho} \leq 2^{14} \epsilon_1 . \tag{2.50}$$

Proof. As anticipated just after Corollary 2.20, we apply the Darboux procedure with $\Omega_0 = \omega_0$, $\Omega_1 = \hat{\omega}_1$ and f solution of (2.37). Then equation (2.21) takes the form

$$Y^t = (-\mathrm{i} + t \Upsilon_{\hat{\omega}_1})^{-1} (\nabla f - W), \tag{2.51}$$

where $\Upsilon_{\hat{\omega}_1}$ and W are defined in Corollary 2.19. By Lemma 2.23 and Corollary 2.19, the vector field Y^t is of class $\mathcal{N}_{\mu^8 \rho}(\mathcal{P}^{w^1}_{\mathbb{R}}, \mathcal{P}^{w^2}_{\mathbb{R}})$ and

$$\sup_{t \in [0,1]} \left|Y^t\right|_{\mu^8 \rho} < 2(2^{11} \epsilon_1 + 2^7 \epsilon_1) < 2^{13} \epsilon_1.$$

Thus Y^t generates a flow $\varphi^t : B^{w^1}_{\mathbb{R}}(\mu^9 \rho) \to \mathcal{P}^{w^1}_{\mathbb{R}}$, defined for every $t \in [0,1]$, which satisfies (cf. Lemma A.4)

$$\left|\varphi^t - \mathbb{1}\right|_{\mu^9 \rho} \leq 2^{14} \epsilon_1, \quad \forall t \in [0,1] .$$

Thus the map $\varphi := \varphi^t|_{t=1}$ exists and satisfies the claimed properties. $\square$

We prove now that the map φ of Lemma 2.24 satisfies also equation (2.36).

Lemma 2.25. *Let f be as in* (2.43) *and φ^t be the flow map of the vector field Y^t defined in* (2.51). *Then $\forall l \geq 1$ one has $I_l(\varphi^t(v)) = I_l(v)$, for each $t \in [0,1]$.*

Proof. The following chain of equivalences follows from Lemma 2.21 and the Darboux equation (2.21):

$$I_l(\varphi^t(v)) = I_l(v) \iff 0 = \frac{d}{dt} I_l(\varphi^t(v)) = dI_l(Y^t(v)) \iff Y^t(v) \in T_v \mathcal{F}^{(0)}$$

$$\iff Y^t(v) \in (T_v \mathcal{F}^{(0)})^{\angle_t} \iff \left(\omega^t_v(Y^t(v), X^0_{I_l}(v)) = 0 , \ \forall l \geq 1\right)$$

$$\iff \alpha_1(X^0_{I_l}) - \alpha_0(X^0_{I_l}) = df(X^0_{I_l}) \quad \forall l \geq 1 .$$

In turn the last property follows since f is a solution of (2.37). $\square$

We can finally prove the quantitative version of the Kuksin–Perelman Theorem.

Proof of Theorem 2.12. Consider the map φ of Lemma 2.24. Since $d\varphi(0) = \mathbb{1}$, φ is invertible in $B_{\mathbb{R}}^{w^1}(\mu^{10}\rho)$ and $\varphi^{-1} = \mathbb{1} + g_1$ with $g_1 \in \mathcal{N}_{\mu^{10}\rho}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{P}_{\mathbb{R}}^{w^2})$ and $\left|\underline{g_1}\right|_{\mu^{10}\rho} \leq 2\left|\underline{\varphi - \mathbb{1}}\right|_{\mu^9\rho} \leq 2^{15}\epsilon_1$ (cf. Lemma A.2). Define now

$$\widetilde{\Psi} := \varphi^{-1} \circ \check{\Psi}.$$

It's easy to check that $\widetilde{\Psi}^*\omega_0 = \omega_0$, thus proving that $\widetilde{\Psi}$ is symplectic. By equation (2.36) one has $I_l(\widetilde{\Psi}(v)) = I_l(\check{\Psi}(v))$ for every $l \geq 1$, therefore $\widetilde{\Psi}$ and $\check{\Psi}$ define the same foliation, which coincides also with the foliation defined by Ψ, cf. Corollary 2.20. Similarly one proves that the functionals $\left\{\frac{1}{2}\left|\widetilde{\Psi}_j(v)\right|\right\}_{j\geq 1}$ pairwise commute with respect to the symplectic form ω_0. We have thus proved items i)–iii) of Theorem 2.12.

We prove now item iv). Clearly $d\widetilde{\Psi}(0) = \mathbb{1}$, and $\widetilde{\Psi}^0 := \widetilde{\Psi} - \mathbb{1} = \check{\Psi}^0 + g_1 \circ (\mathbb{1} + \check{\Psi}^0)$ is of class $\mathcal{N}_{\mu^{11}\rho}(\mathcal{P}_{\mathbb{R}}^{w^1}, \mathcal{P}_{\mathbb{R}}^{w^2})$. Moreover, by Remark 2.10 and Corollary 2.20(i), one has $\left|\underline{\check{\Psi}^0}\right|_{\mu^{11}\rho} \leq 2\mu^6\left|\underline{\check{\Psi}^0}\right|_{\mu^8\rho} \leq \mu^6 2^9 \epsilon_1 \leq \mu^{11}\rho$ by condition (2.18). Thus $\left|\underline{\mathbb{1} + \check{\Psi}^0}\right|_{\mu^{11}\rho} \leq \mu^{10}\rho$ and by Lemma A.1

$$\left|\underline{\widetilde{\Psi}_0}\right|_{\mu^{11}\rho} \leq \left|\underline{\check{\Psi}^0}\right|_{\mu^{11}\rho} + \left|\underline{g_1 \circ (\mathbb{1} + \check{\Psi}^0)}\right|_{\mu^{11}\rho} \leq \left|\underline{\check{\Psi}^0}\right|_{\mu^{11}\rho} + \left|\underline{g_1}\right|_{\mu^{10}\rho} \leq 2^8\epsilon_1 + 2^{15}\epsilon_1 \leq 2^{16}\epsilon_1.$$

We are left to prove that $\widetilde{\Psi}^0 \in \mathcal{A}_{w^1, \mu^{12}\rho}^{w^2}$. Since $\widetilde{\Psi}^*\omega_0 = \omega_0$, one has $d\widetilde{\Psi}(v)^*(-\mathrm{i})\,\widetilde{\Psi}(v) = -\mathrm{i}$, from which it follows that $\widetilde{\Psi}^0$ satisfies

$$d\widetilde{\Psi}^0(v)^* = \mathrm{i}\,d\widetilde{\Psi}^0(v)\left(\mathbb{1} + d\widetilde{\Psi}^0(v)\right)^{-1}\mathrm{i}$$

and therefore $\widetilde{\Psi}^0 \in \mathcal{A}_{w^1, \mu^{12}\rho}^{w^2}$ with $\left\|\widetilde{\Psi}^0\right\|_{\mathcal{A}_{w^1, \mu^{12}\rho}^{w^2}} < 2^{17}\epsilon_1$. $\square$

3. Toda lattice

3.1. Proof of Theorem 1.3 and Corollary 1.6

We consider the Toda lattice with N particles and periodic boundary conditions on the positions q and momenta p: $q_{j+N} = q_j$, $p_{j+N} = p_j$, $\forall j \in \mathbb{Z}$. As anticipated in Section 1, we restrict to the invariant subspace characterized by (1.2). The phase space of the system is $\mathcal{P}^{s,\sigma}$, where $s \geq 0$, $\sigma \geq 0$ and it is defined in terms of the linear, complex,

Birkhoff variables (ξ, η) (defined in (2.3)). We endow the phase space with the symplectic form[2] $\Omega_0 = -\mathrm{i} \sum_{k=1}^{N-1} d\xi_k \wedge d\eta_k$.

We will denote by $\mathcal{P}_{\mathbb{R}}^{s,\sigma}$ the real subspace of $\mathcal{P}^{s,\sigma}$ in which $\eta_k = \bar{\xi}_k \ \forall 1 \leq k \leq N-1$, endowed with the norm (1.7), and by $B_{\mathbb{R}}^{s,\sigma}(\rho)$ the ball in $\mathcal{P}_{\mathbb{R}}^{s,\sigma}$ with center 0 and radius $\rho > 0$. The main step of the proof of Theorem 1.3 is the construction of the functions $\{\Psi_j\}_{1 \leq j \leq N-1}$. This is based on a detailed analysis of the spectrum of the Jacobi matrix appearing in the Lax pair representation of the Toda lattice. So we start by recalling the elements of the theory needed for our development. Introduce the translated Flaschka coordinates [17] by

$$(b, a) = \Theta(p, q), \qquad (b_j, \ a_j) := (-p_j, \ e^{\frac{1}{2}(q_j - q_{j+1})} - 1). \qquad (3.2)$$

The translation of the a variables by 1 is useful in order to keep the equilibrium point at $(b, a) = (0, 0)$. Recall that the variables b, a are constrained by the conditions

$$\sum_{j=0}^{N-1} b_j = 0, \ \prod_{j=0}^{N-1} (1 + a_j) = 1 \ .$$

Introduce Fourier variables $(\hat{b}, \hat{a})$ for the Flaschka coordinates by (1.3). In these variables

$$E_k = \frac{|\hat{b}_k|^2 + 4|\hat{a}_k|^2}{2} + O(\hat{a}^3), \qquad 1 \leq k \leq N-1 \ . \qquad (3.3)$$

The Jacobi matrix whose spectrum forms a complete set of integrals of motions for the Toda lattice is given by [38]

$$L(b, a) := \begin{pmatrix} b_0 & 1 + a_0 & 0 & \ldots & 1 + a_{N-1} \\ 1 + a_0 & b_1 & 1 + a_1 & \ddots & \vdots \\ 0 & 1 + a_1 & b_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 + a_{N-2} \\ 1 + a_{N-1} & \ldots & 0 & 1 + a_{N-2} & b_{N-1} \end{pmatrix}. \qquad (3.4)$$

It is useful to double the size of $L(b, a)$, redefining

---

[2] So that the Hamilton equations become

$$\dot{\xi}_k = \mathrm{i} \frac{\partial H}{\partial \eta_k}, \qquad \dot{\eta}_k = -\mathrm{i} \frac{\partial H}{\partial \xi_k} \ . \qquad (3.1)$$

$$L_{b,a} := \begin{pmatrix} b_0 & 1+a_0 & \cdots & & 0 & 0 & \cdots & 0 & 1+a_{N-1} \\ 1+a_0 & b_1 & \ddots & & \vdots & 0 & \cdots & & 0 \\ \vdots & \ddots & \ddots & 1+a_{N-2} & & \vdots & & & \vdots \\ 0 & \ddots & 1+a_{N-2} & b_{N-1} & 1+a_{N-1} & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 1+a_{N-1} & b_0 & 1+a_0 & \cdots & & 0 \\ 0 & \cdots & & 0 & 1+a_0 & b_1 & \ddots & & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & & 1+a_{N-2} \\ 1+a_{N-1} & \cdots & 0 & 0 & 0 & \ddots & 1+a_{N-2} & b_{N-1} \end{pmatrix}. \tag{3.5}$$

Consider the eigenvalues of $L_{b,a}$ and order them in the non-decreasing sequence

$$\lambda_0(b,a) < \lambda_1(b,a) \le \lambda_2(b,a) < \ldots < \lambda_{2N-3}(b,a) \le \lambda_{2N-2}(b,a) < \lambda_{2N-1}(b,a)$$

where one has that where the sign $\le$ appears equality is possible, while it is impossible in the correspondence of a sign $<$. Define the quantities

$$\gamma_j(b,a) := \lambda_{2j}(b,a) - \lambda_{2j-1}(b,a), \qquad 1 \le j \le N-1; \tag{3.6}$$

$\gamma_j(b,a)$ is called *jth spectral gap*. The quantities $\{\gamma_j^2\}_{1 \le j \le N-1}$ form a complete set of commuting integrals of motions, which are regular also at $(b,a) = (0,0)$. Furthermore one has $H(b,a) = H(\gamma_1^2(b,a), \ldots, \gamma_{N-1}^2(b,a))$ [6]. A spectral gap is said to be *closed* if $\gamma_j(b,a) = 0$.

The following Theorem 3.1 ensures that the assumptions of Theorem 2.12 are fulfilled by the Toda lattice.

Theorem 3.1. *There exists $\epsilon_* > 0$, independent of N, and an analytic map*

$$\Psi : \left(B^{s,\sigma}\left(\frac{\epsilon_*}{N^2}\right), \Omega_0 \right) \to \mathcal{P}^{s,\sigma}, \quad (\xi, \eta) \mapsto (\phi(\xi,\eta), \psi(\xi,\eta)) \tag{3.7}$$

such that:

(Ψ1) Ψ *is real for real sequences, namely* $\overline{\phi_k(\xi,\bar\xi)} = \psi_k(\xi,\bar\xi)\ \forall k$.
(Ψ2) *For every $1 \le j \le N-1$, and for $(\phi,\psi) \in B^{s,\sigma}\left(\frac{\epsilon_*}{N^2}\right) \cap \mathcal{P}^{s,\sigma}_{\mathbb{R}}$, one has*

$$\gamma_j^2 = \tfrac{2}{N}\omega\left(\tfrac{j}{N}\right)|\psi_j|^2 = \tfrac{2}{N}\omega\left(\tfrac{k}{N}\right)|\varphi_j|^2 \ .$$

(Ψ3) $\Psi(0,0) = (0,0)$ *and* $d\Psi(0,0) = \mathbb{1}$.
(Ψ4) *There exist constants $C_1, C_2 > 0$, independent of N, such that for every $0 < \epsilon \le \epsilon_*$, the map $\Psi^0 := \Psi - \mathbb{1} \in \mathcal{N}_{\epsilon/N^2}\left(\mathcal{P}^{s,\sigma}, \mathcal{P}^{s+1,\sigma}\right)$ and $[d\Psi^0]^* \in \mathcal{N}_{\epsilon/N^2}(\mathcal{P}^{s,\sigma}, \mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{P}^{s+1,\sigma}))$. Furthermore one has*

$$\left|\underline{\Psi^0}\right|_{\epsilon/N^2} \le C_1 \frac{\epsilon^2}{N^2}; \qquad \left|\underline{[d\Psi^0]^*}\right|_{\epsilon/N^2} \le C_2 \epsilon \ . \tag{3.8}$$

The main point is (Ψ4), in which the estimates of the domain of definition of the map Ψ hold uniformly in the limit $N \to \infty$.

We show now how Theorem 1.3 follows from Kuksin–Perelman Theorem 2.12.

Proof of Theorem 1.3. Introduce the weights $w^1 := \{N^{3/2}[k]_N^s e^{\sigma[k]_N} \omega \left(\frac{k}{N}\right)^{1/2}\}_{k=1}^{N-1}$ and $w^2 := \{N^{3/2}[k]_N^{s+1} e^{\sigma[k]_N} \omega \left(\frac{k}{N}\right)^{1/2}\}_{k=1}^{N-1}$ and consider the map Ψ of Theorem 3.1 as a map from $\mathcal{P}^{w^1}$ in itself. Since for any $(\xi, \eta) \in \mathcal{P}^{w^1}$ one has that

$$\|(\xi,\eta)\|_{\mathcal{P}^{w^1}} \equiv N^2 \|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \ , \tag{3.9}$$

it follows by scaling that there exists a constant $C_3 > 0$, independent of N, such that

$$\left\|\Psi^0\right\|_{\mathcal{A}_{w^1,\rho}^{w^2}} \leq C_3 \rho^2 \ .$$

Thus, for any $\rho \leq \rho_* \equiv \min\left(\frac{2^{-34}}{C_3}, \epsilon_*\right)$, Ψ satisfies condition (2.18). Thus we can apply Theorem 2.12 to the map Ψ, getting the existence of a symplectic real analytic map $\widetilde{\Psi}$ defined on $B^{w^1}(a\rho_*)$ which satisfies $i)$–$iv)$ of Theorem 2.12.

By Lemma A.3 the map $\widetilde{\Psi}$ is invertible in $B^{w^1}(\mu a\rho_*)$ and its inverse Φ satisfies $\Phi = \mathbb{1} + \Phi^0$ with $\Phi^0 \in \mathcal{A}_{w^1,\mu a\rho_*}^{w^2}$. To get the statement of the theorem simply reexpress the map Φ in terms of real variables (x,y), (X,Y) and denote such a map by Φ_N. $\quad\square$

Remark 3.2. By the proof of Theorem 1.3 above one deduces the estimate

$$\sup_{\|(\phi,\psi)\|_{\mathcal{P}^{s,\sigma}} \leq R_{s,\sigma}/N^2} \left\|d\Phi^0(\phi,\psi)^*\right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma},\mathcal{P}^{s+1,\sigma})} \leq C_{s,\sigma} R_{s,\sigma} \ , \tag{3.10}$$

for some $C_{s,\sigma} > 0$, independent of N.

The rest of this subsection is devoted to the proof of Theorem 3.1.

In the following it will be convenient to consider the variables (b,a) defined in (3.2) dropping the conditions $\sum_{j=0}^{N-1} b_j = 0$ and $\prod_{j=0}^{N-1}(1+a_j) = 1$. Equation (3.3) suggests to introduce on the variables b, a the norm

$$\|(b,a)\|_{\mathcal{C}^{s,\sigma}}^2 := \frac{1}{2N} \sum_{k=0}^{N-1} \max(1,[k]_N^{2s}) e^{2\sigma[k]_N} \left(|\hat{b}_k|^2 + 4|\hat{a}_k|^2\right) \tag{3.11}$$

and to define the space

$$\mathcal{C}_{\mathbb{R}}^{s,\sigma} := \left\{(b,a) \in \mathbb{R}^N \times \mathbb{R}^N \ : \ \|(b,a)\|_{\mathcal{C}^{s,\sigma}} < \infty\right\} \ . \tag{3.12}$$

We will write $\mathcal{C}^{s,\sigma}$ for the complexification of $\mathcal{C}_{\mathbb{R}}^{s,\sigma}$.

In the following we will consider normally analytic map between the spaces $\mathcal{P}^{s,\sigma}$ and $\mathcal{C}^{s,\sigma}$. We need to specify the basis of $\mathcal{C}^{s,\sigma}$ that we will use to verify the property of being

normally analytic. While it is quite hard to verify this property when the basis is general, it turns out that it is quite easy to verify it using the basis of complex exponentials defined in (1.3). Indeed the norm (3.11) is given in terms of the Fourier variables. For the same reason, it will be convenient to express a map from $\mathcal{C}^{s,\sigma}$ to $\mathcal{P}^{s,\sigma}$ as a function of the Fourier variables $\hat{b}$, $\hat{a}$.

We prove now some analytic properties of the map Θ defined in (3.2). In the following we will denote by Θ_Ξ the map Θ expressed in the (ξ, η) variables.

Proposition 3.3. *The map Θ_Ξ satisfies the following properties:*

(Θ1) $\Theta_\Xi(0,0) = (0,0)$. *Furthermore let $d\Theta_\Xi(0,0)$ be the linearization of Θ_Ξ at $(\xi, \eta) = (0,0)$. Then $(B, A) = d\Theta_\Xi(0,0)[(\xi, \eta)]$ iff*

$$\widehat{B}_0 = 0, \qquad \widehat{B}_k = -\left(\tfrac{1}{2}\omega\left(\tfrac{k}{N}\right)\right)^{1/2}(\xi_k + \eta_{N-k}), \qquad 1 \le k \le N-1 \ ,$$

$$\widehat{A}_0 = 0, \qquad \widehat{A}_k = -\mathrm{i}\varpi_k\left(2\omega\left(\tfrac{k}{N}\right)\right)^{-1/2}(\xi_k - \eta_{N-k}), \qquad 1 \le k \le N-1, \quad (3.13)$$

where $\varpi_k := (1 - e^{-2\mathrm{i}\pi k/N})/2, \ \forall\, 1 \le k \le N-1$.
Moreover for any $s \ge 0$, $\sigma \ge 0$ there exist constants $C_{\Theta_1}, C_{\Theta_2} > 0$, independent of N, such that

$$\left\| d\Theta_\Xi(0,0) \right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma},\, \mathcal{C}^{s,\sigma})} \le C_{\Theta_1},$$

$$\left\| d\Theta_\Xi(0,0)^* \right\|_{\mathcal{L}(\mathcal{C}^{s+2,\sigma},\, \mathcal{P}^{s+1,\sigma})} \le \frac{C_{\Theta_2}}{N} \ . \tag{3.14}$$

(Θ2) *Let $\Theta_\Xi^0 := \Theta_\Xi - d\Theta_\Xi(0,0)$. For any $s \ge 0$, $\sigma \ge 0$, there exist constants $C_{\Theta_3}, C_{\Theta_4}, \epsilon_* > 0$, independent of N, such that the map $\Theta_\Xi^0 \in \mathcal{N}_{\epsilon_*/N^2}(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s+1,\sigma})$ and the map $[d\Theta_\Xi^0]^* \in \mathcal{N}_{\epsilon_*/N^2}(\mathcal{P}^{s,\sigma}, \mathcal{L}(\mathcal{C}^{s+2,\sigma}, \mathcal{P}^{s+1,\sigma}))$, and*

$$\left| \Theta_\Xi^0 \right|_{\epsilon/N^2} \equiv \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \le \epsilon/N^2} \left\| \Theta_\Xi^0(\xi,\eta) \right\|_{\mathcal{C}^{s+1,\sigma}} \le \frac{C_{\Theta_3}\epsilon^2}{N^2};$$

$$\left| [d\Theta_\Xi^0]^* \right|_{\epsilon/N^2} \equiv \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \le \epsilon/N^2} \left\| d\Theta_\Xi^0(\xi,\eta)^* \right\|_{\mathcal{L}(\mathcal{C}^{s+2,\sigma},\, \mathcal{P}^{s+1,\sigma})} \le \frac{C_{\Theta_4}\epsilon}{N^2}. \tag{3.15}$$

The proof of the proposition is postponed in Appendix C. Note that the estimates (3.14) and (3.15) imply that there exists a constant $C_{\Theta_5} > 0$, independent of N, such that for any $\rho \le \frac{\epsilon_*}{N^2}$ one has $\Theta_\Xi \in \mathcal{N}_\rho(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s,\sigma})$ and

$$\left| \Theta_\Xi \right|_\rho \le C_{\Theta_5}\, \rho \ . \tag{3.16}$$

We start now the perturbative construction of the Birkhoff coordinates for the Toda lattice, which is based on the construction of the spectrum and of the eigenfunctions

of $L_{b,a}$ (defined in (3.5)) as a perturbation of the free operator $L_0 := L_{b,a}|_{(b,a)=(0,0)}$. More precisely we decompose $L_{b,a} = L_0 + L_p$, where

$$
L_0 = \begin{pmatrix} 0 & 1 & 0 & \dots & 1 \\ 1 & 0 & 1 & \ddots & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 1 & \dots & \dots & 1 & 0 \end{pmatrix}, \quad L_p = \begin{pmatrix} b_0 & a_0 & 0 & \dots & a_{N-1} \\ a_0 & b_1 & a_1 & \ddots & \vdots \\ 0 & a_1 & b_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & a_{N-2} \\ a_{N-1} & \dots & \dots & a_{N-2} & b_{N-1} \end{pmatrix} \quad (3.17)
$$

and following the approach in [30,6,26] we apply Kato perturbation theory [29]. The next lemma characterizes completely the spectrum of L_0 as an operator on $\mathbb{C}^{2N}$:

Lemma 3.4. *Consider L_0 as an operator on $\mathbb{C}^{2N}$, then its eigenvalues and normalized eigenvectors are:*

eigenvalues   eigenvectors
$\lambda_0^0 = -2$, $f_{00}(k) = \frac{1}{\sqrt{2N}}(-1)^k$
$\lambda_{2j-1}^0 = \lambda_{2j}^0 = -2\cos\left(\frac{j\pi}{N}\right)$, $f_{2j-1,0}(k) = \frac{1}{\sqrt{2N}}e^{-i\rho_j k}$, $f_{2j,0}(k) = \frac{1}{\sqrt{2N}}e^{i\rho_j k}$, $1 \le j \le N-1$
$\lambda_{2N-1}^0 = 2$, $f_{2N-1,0}(k) = \frac{1}{\sqrt{2N}}$

where $0 \le k \le 2N-1$ and $\rho_j := \left(1 + \frac{j}{N}\right)\pi$. In particular the gaps of L_0 are all closed.

The proof is an easy computation and can be found in [23].

Remark 3.5. For $0 \le j, k \le \lfloor N/2 \rfloor$ one has $\left|\lambda_{2j}^0 - \lambda_{2k}^0\right|$, $\left|\lambda_{2N-j}^0 - \lambda_{2N-k}^0\right| \ge \frac{4|j^2-k^2|}{N^2}$.
In particular if $j \ne k$ then $\left|\lambda_{2j}^0 - \lambda_{2k}^0\right| \ge 1/N^2$.

We use now Kato perturbation theory of operators in order to introduce the main objects needed in the following and to give some preliminary estimates.

For $1 \le j \le N-1$ let $E_j(b,a)$ be the two-dimensional subspace spanned by the eigenvectors corresponding to the eigenvalues $\lambda_{2j-1}(b,a)$ and $\lambda_{2j}(b,a)$ of $L_{b,a}$. Analogously, let $E_0(b,a)$ (respectively $E_N(b,a)$) be the one-dimensional subspace spanned by the eigenvector of $\lambda_0(b,a)$ (respectively $\lambda_{2N-1}(b,a)$). Introduce the spectral projector on $E_j(b,a)$ defined by

$$
P_j(b,a) = -\frac{1}{2\pi i} \oint_{\Gamma_j} (L_{b,a} - \lambda)^{-1} \, d\lambda, \qquad 0 \le j \le N \quad (3.18)
$$

where, for $1 \le j \le N-1$, Γ_j is a closed path counter-clockwise oriented in $\mathbb{C}$ which encloses the eigenvalues $\lambda_{2j-1}(b,a)$ and $\lambda_{2j}(b,a)$ and does not contain any other eigenvalue of $L_{b,a}$. Analogously, Γ_0 (respectively Γ_N) encloses the eigenvalue $\lambda_0(b,a)$ (respectively $\lambda_{2N-1}(b,a)$) and no other eigenvalue of $L_{b,a}$. $P_j(b,a)$ maps $\mathbb{C}^{2N}$ onto $E_j(b,a)$ and, as we

will prove, is well defined for (b, a) small enough. $P_j(0, 0)$ will be denoted by P_{j0} and its range $E_j(0, 0)$, which will be denoted by E_{j0}, is given by

$$\operatorname{Im} P_{j0} = E_{j0}, \quad E_{j0} = \operatorname{span} \langle f_{2j,0}, f_{2j-1,0} \rangle .$$

Define also the transformation operators

$$U_j(b, a) = \left(\mathbb{1} - (P_j(b, a) - P_{j0})^2 \right)^{-1/2} P_j(b, a), \quad 1 \leq j \leq N - 1. \qquad (3.19)$$

U_j has the property of mapping isometrically E_{j0} into the subspace $E_j(b, a)$ spanned by the perturbed eigenvectors [29]. Remark, however, that in general the image of an unperturbed eigenvector is *not* an eigenvector itself. We prove now some properties of the just defined objects.

Lemma 3.6. *There exists a constant $C_{s,\sigma} > 0$, independent of N, such that the map $(b, a) \mapsto L_p(b, a)$ is analytic as a map from $\mathcal{C}^{s,\sigma}$ to $\mathcal{L}\left(\mathbb{C}^{2N} \right)$. Moreover*

$$\| L_p(b, a) \|_{\mathcal{L}(\mathbb{C}^{2N})} \leq C_{s,\sigma} \| (b, a) \|_{\mathcal{C}^{s,\sigma}} . \qquad (3.20)$$

Then by Kato theory one has the corollary

Corollary 3.7. *There exist constants $C_{s,\sigma}$, $\epsilon_* > 0$, independent of N, such that the following holds true:*

(i) *The spectrum of $L_{b,a}$ is close to the spectrum of L_0; in particular for any $(b, a) \in B^{\mathcal{C}^{s,\sigma}} \left(\frac{\epsilon_*}{N^2} \right)$*

$$\left| \lambda_{2j}(b, a) - \lambda_{2j}^0 \right|, \ \left| \lambda_{2j-1}(b, a) - \lambda_{2j-1}^0 \right| \leq C_{s,\sigma} \| (b, a) \|_{\mathcal{C}^{s,\sigma}} . \qquad (3.21)$$

(ii) *One has that $(b, a) \mapsto P_j(b, a)$ is analytic as a map from $B^{\mathcal{C}^{s,\sigma}} \left(\frac{\epsilon_*}{N^2} \right)$ to $\mathcal{L}(\mathbb{C}^{2N})$. Moreover for $(b, a) \in B^{\mathcal{C}^{s,\sigma}} \left(\frac{\epsilon_*}{N^2} \right)$ one has*

$$\| P_j(b, a) - P_{j0} \|_{\mathcal{L}(\mathbb{C}^{2N})} \leq C_{s,\sigma} \| (b, a) \|_{\mathcal{C}^{s,\sigma}} . \qquad (3.22)$$

(iii) *For each $1 \leq j \leq N - 1$, the maps U_j, defined in (3.19), are well defined from $B^{\mathcal{C}^{s,\sigma}} \left(\frac{\epsilon_*}{N^2} \right)$ to $\mathcal{L}(\mathbb{C}^{2N})$ and satisfy the following algebraic properties:*
(U1) *$\operatorname{Im} U_j(b, a) = E_j(b, a)$;*
(U2) *for (b, a) real, one has $\overline{U_j(b, a)f} = U_j(b, a)\bar{f}$;*
(U3) *for (b, a) real and $f \in E_{j0}$, one has $\| U_j(b, a)f \|_{\mathbb{C}^{2N}} = \| f \|_{\mathbb{C}^{2N}}$.*

Finally the following analytic property holds:

(U4) One has that $(b,a) \mapsto U_j(b,a)$ is analytic as a map from $B^{\mathcal{C}^{s,\sigma}}\left(\frac{\epsilon_*}{N^2}\right)$ to $\mathcal{L}(\mathbb{C}^{2N})$. Moreover for $(b,a) \in B^{\mathcal{C}^{s,\sigma}}\left(\frac{\epsilon_*}{N^2}\right)$ one has

$$\|U_j(b,a) - P_j(b,a)\|_{\mathcal{L}(\mathbb{C}^{2N})} \leq C_{s,\sigma} \|(b,a)\|^2_{\mathcal{C}^{s,\sigma}}. \tag{3.23}$$

The proofs of Lemma 3.6 and Corollary 3.7 can be found in Appendix D.

For $1 \leq j \leq N-1$ and $(b,a) \in B^{\mathcal{C}^{s,\sigma}}\left(\frac{\epsilon_*}{N^2}\right)$ define now the vectors

$$f_{2j-1}(b,a) := U_j(b,a)f_{2j-1,0}, \qquad \text{and} \qquad f_{2j}(b,a) := U_j(b,a)f_{2j,0} \tag{3.24}$$

which by property $(U1)$ belong to $E_j(b,a)$. Define also the maps

$$z_j(b,a) := \left(\tfrac{2}{N}\omega\left(\tfrac{j}{N}\right)\right)^{-1/2}\left\langle \left(L_{b,a} - \lambda^0_{2j}\right)f_{2j}(b,a), \overline{f_{2j}(b,a)}\right\rangle,$$

$$w_j(b,a) := \left(\tfrac{2}{N}\omega\left(\tfrac{j}{N}\right)\right)^{-1/2}\left\langle \left(L_{b,a} - \lambda^0_{2j-1}\right)f_{2j-1}(b,a), \overline{f_{2j-1}(b,a)}\right\rangle \tag{3.25}$$

where $\langle u,v \rangle = \sum u_j \overline{v_j}$ is the Hermitian product in $\mathbb{C}^{2N}$. Finally denote $z(b,a) = (z_1(b,a), \ldots, z_{N-1}(b,a))$ and $w(b,a) = (w_1(b,a), \ldots, w_{N-1}(b,a))$, and let Z be the map

$$(b,a) \mapsto Z(b,a) := (z(b,a), w(b,a)). \tag{3.26}$$

The map Ψ of Theorem 3.1 will be constructed by expressing Z as a function of the linear Birkhoff coordinates ξ, η.

The properties of the map Z are collected in the next lemma which constitutes the main technical step for the application of Kuksin–Perelman Theorem to the Toda lattice.

Lemma 3.8. *The map Z, defined by (3.26), is well defined for $(b,a) \in B^{\mathcal{C}^{s,\sigma}}\left(\frac{\epsilon_*}{N^2}\right)$. If b, a are real valued and fulfill $\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \frac{\epsilon_*}{N^2}$, then, for every $1 \leq j \leq N-1$, the following properties are also fulfilled:*

(Z1) $\overline{z_j(b,a)} = w_j(b,a)$;

(Z2) $\gamma_j^2 = \frac{2}{N}\omega\left(\frac{j}{N}\right)|z_j(b,a)|^2 = \frac{2}{N}\omega\left(\frac{j}{N}\right)|w_j(b,a)|^2$;

(Z3) $z_j(0,0) = w_j(0,0) = 0$; moreover the linearizations of z_j and w_j at $(b,a) = (0,0)$ are given by

$$dz_j(0,0)[(B,A)] = \left(2\omega\left(\tfrac{j}{N}\right)\right)^{-1/2}\left(\hat{B}_j - 2e^{ji\pi/N}\hat{A}_j\right),$$

$$dw_j(0,0)[(B,A)] = \left(2\omega\left(\tfrac{j}{N}\right)\right)^{-1/2}\left(\hat{B}_{N-j} - 2e^{-ji\pi/N}\hat{A}_{N-j}\right). \tag{3.27}$$

The map $dZ(0,0) = (dz(0,0), dw(0,0))$ is in the class $\mathcal{L}(\mathcal{C}^{s,\sigma}, \mathcal{P}^{s,\sigma})$. Its adjoint $dZ(0,0)^*$ is in the class $\mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s+1,\sigma})$. Finally there exist constants $C_{Z_1}, C_{Z_2} > 0$,

independent of N, such that for any $s \geq 0$ and $\sigma \geq 0$

$$\|\underline{dZ}(0,0)\|_{\mathcal{L}(\mathcal{C}^{s,\sigma}, \mathcal{P}^{s,\sigma})} \leq C_{Z_1}, \quad \|\underline{dZ}(0,0)^*\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s+2,\sigma})} \leq C_{Z_2} N^2 . \quad (3.28)$$

(Z4) *For any $s \geq 0$, $\sigma \geq 0$, there exist constants $C_{Z_3}, C_{Z_4}, \epsilon_* > 0$, independent of N, such that for every $0 < \epsilon \leq \epsilon_*$ the map $Z^0 := Z - dZ(0,0) \in \mathcal{N}_{\epsilon/N^2}\left(\mathcal{C}^{s,\sigma}, \mathcal{P}^{s+1,\sigma}\right)$ and the map $[dZ^0]^* \in \mathcal{N}_{\epsilon/N^2}\left(\mathcal{C}^{s,\sigma}, \mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s+2,\sigma})\right)$. Moreover*

$$\sup_{\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \epsilon/N^2} \left\|\underline{Z^0}(b,a)\right\|_{\mathcal{P}^{s+1,\sigma}} \leq C_{Z_3} \frac{\epsilon^2}{N^2},$$

$$\sup_{\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \epsilon/N^2} \left\|\underline{dZ^0}(b,a)^*\right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s+2,\sigma})} \leq C_{Z_4} N\epsilon. \quad (3.29)$$

The proof of the lemma is very technical, and is postponed in Appendix E.

Remark 3.9. In the limit of infinitely many particles, the linearization $dz_j(0,0)(b,a)$ at the different edges of the spectrum are given by

$$dz_j(0,0)(B, A) \approx \frac{\hat{B}_j - 2\hat{A}_j}{\sqrt{2\omega(j/N)}} \quad \text{if } j/N \ll 1$$

$$dz_j(0,0)(B, A) \approx \frac{\hat{B}_j + 2\hat{A}_j}{\sqrt{2\omega(j/N)}} \quad \text{if } 1 - j/N \ll 1 . \quad (3.30)$$

The existence of two different sequences is in agreement with the works [5,4], in which the spectrum of the Lax operator associated to the Toda lattice is approximated, up to a small error, by the spectrum of *two* Sturm–Liouville operators associated to *two* KdV equations. More explicitly, in [5] the following result is proved: take $\alpha, \beta \in C^\infty(\mathbb{T})$ such that $\int_{\mathbb{T}} \alpha = \int_{\mathbb{T}} \beta = 0$, $a_j = 1 + \frac{1}{N^2}\alpha(j/N)$ and $b_j = \frac{1}{N^2}\beta(j/N)$. Then the spectrum of the Lax matrix (3.5) with a_j, b_j as elements can be approximated at the two edges by the spectrum of the two Sturm–Liouville operators $L = -\frac{d^2}{dx^2} + (\beta \pm 2\alpha)$ on $C^\infty(\mathbb{T})$.

We are ready to define the map Ψ of Theorem 3.1: let

$$\Psi : \mathcal{P}^{s,\sigma} \to \mathcal{P}^{s,\sigma}, \qquad (\xi,\eta) \mapsto (\phi(\xi,\eta), \psi(\xi,\eta)) \quad (3.31)$$

defined by

$$\Psi = -Z \circ \Theta_\Xi; \quad \text{i.e.} \quad \phi = -z \circ \Theta_\Xi, \qquad \psi = -w \circ \Theta_\Xi. \quad (3.32)$$

We show now that Ψ satisfies the properties $(\Psi1)$–$(\Psi4)$ claimed in Theorem 3.1.

Proof of Theorem 3.1. Properties $(\Psi1)$ and $(\Psi2)$ follow by $(Z1)$ and $(Z2)$ respectively. We prove now $(\Psi3)$. By $(\Theta1)$ and $(Z3)$ one has $\Psi(0,0) = (0,0)$. In order to compute $d\Psi(0,0) = (d\phi(0,0), d\psi(0,0))$ note that

$$d\phi(0,0) = -dz(0,0)\, d\Theta_\Xi(0,0) = -(dz(0,0)\mathcal{F}^{-1}) \circ (\mathcal{F}d\Theta_\Xi(0,0)) \ .$$

Let $(\hat{B}, \hat{A}) = \mathcal{F}d\Theta_\Xi(0,0)(\xi, \eta)$. Then (3.27) and (3.13) imply that, for $1 \leq j \leq N-1$,

$$
\begin{aligned}
d\phi_j(0,0)(\xi,\eta) &= -\frac{1}{\sqrt{2\omega(j/N)}}\left(\hat{B}_j - 2e^{i\pi j/N}\hat{A}_j\right) \\
&= \frac{1}{\sqrt{2\omega(j/N)}}\left(\sqrt{\frac{\omega(j/N)}{2}}(\xi_j + \eta_{N-j}) - i\frac{2e^{i\pi j/N}\varpi_j}{\sqrt{2\omega(j/N)}}(\xi_j - \eta_{N-j})\right) \equiv \xi_j \ ,
\end{aligned}
$$

where we used that $2e^{i\pi j/N}\varpi_j = i\omega\left(\frac{j}{N}\right)$. One verifies analogously that $d\psi_j(0,0)(\xi,\eta) = \eta_j$.

We prove now property ($\Psi 4$), which is a consequence of the fact that the space of normally analytic maps is closed by composition (see Lemma A.1). Fix $s \geq 0$ and $\sigma \geq 0$. Let $0 < \epsilon \leq \frac{\epsilon_*}{C_{\Theta_5}}$, where C_{Θ_5} is the constant in (3.16). Since $Z = dZ(0,0) + Z^0$ and $\Theta_\Xi = d\Theta_\Xi(0,0) + \Theta_\Xi^0$, one gets that

$$\Psi^0 = -Z^0 \circ \Theta_\Xi - dZ(0,0) \circ \Theta_\Xi^0 \ . \tag{3.33}$$

Thus properties ($Z3$), ($\Theta 2$) and estimate (3.16) imply that there exists a constant $C > 0$, independent of N, such that

$$\left|\underline{\Psi^0}\right|_{\epsilon/N^2} \equiv \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2} \left\|\underline{\Psi^0}(\xi,\eta)\right\|_{\mathcal{P}^{s+1,\sigma}} \leq \frac{C\,\epsilon^2}{N^2} \ ,$$

which proves the first estimate of ($\Psi 4$). We study now the adjoint map $d\Psi^0(\xi,\eta)^*$. Writing $d\Theta_\Xi = d\Theta_\Xi(0,0) + d\Theta_\Xi^0$ one gets that

$$
\begin{aligned}
d\Psi^0(\xi,\eta)^* &= -d\Theta_\Xi(0,0)^*\, dZ^0(\Theta_\Xi(\xi,\eta))^* - d\Theta_\Xi^0(\xi,\eta)^*\, dZ^0(\Theta_\Xi(\xi,\eta))^* \\
&\quad - d\Theta_\Xi^0(\xi,\eta)^*\, dZ(0,0)^* \\
&= I + II + III.
\end{aligned}
$$

We estimate each term in the expression displayed above. In the following, if $A \in \mathcal{N}_\rho(\mathcal{P}^{s,\sigma}, \mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{P}^{s+1,\sigma}))$, we denote by

$$\left|\underline{A}\right|_\rho \equiv \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2} \left\|\underline{A}(\xi,\eta)\right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma},\,\mathcal{P}^{s+1,\sigma})} \ .$$

We begin by estimating I:

$$
\begin{aligned}
\left|\underline{I}\right|_{\epsilon/N^2} &\leq \frac{C_{\Theta_2}}{N} \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2} \left\|\underline{dZ^0}(\Theta_\Xi(\xi,\eta))^*\right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma},\,\mathcal{C}^{s+2,\sigma})} \\
&\leq \frac{C_{\Theta_2}}{N} C_{Z_4} C_{\Theta_5} N\,\epsilon \leq C\epsilon,
\end{aligned}
$$

where in the first inequality we used the second estimate of (3.14) and in the second inequality we used the second estimate in (3.29). Now we study II:

$$|\underline{II}|_{\epsilon/N^2} \leq \frac{C_{\Theta_4}\epsilon}{N^2} \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2} \left\| dZ^0(\underline{\Theta_\Xi}(\xi,\eta))^* \right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \, \mathcal{C}^{s+2,\sigma})}$$

$$\leq \frac{C_{\Theta_4}\epsilon}{N^2} C_{Z_4} C_{\Theta_5} N\epsilon \leq \frac{C\epsilon^2}{N},$$

where we used the second estimate in (3.15) and again (Z4). Finally, using again (Θ2) and the second estimate of (3.28), one has

$$|\underline{III}|_{\epsilon/N^2} \leq \frac{C_{\Theta_4}\epsilon}{N^2} \|\underline{dZ}(0,0)^*\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \, \mathcal{C}^{s+2,\sigma})} \leq \frac{C_{\Theta_4}\epsilon}{N^2} C_{Z_2} N^2 \leq C\epsilon \,.$$

Collecting the estimates above one gets

$$\left| [\underline{d\Psi^0}]^* \right|_{\epsilon/N^2} \equiv \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2} \left\| \underline{d\Psi^0}(\xi,\eta)^* \right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \, \mathcal{P}^{s+1,\sigma})} \leq 3C\epsilon,$$

and (Ψ4) follows.   □

Proof of Corollary 1.6. Provided $0 < R < R'_{s,\sigma}$ is small enough, one has that $w_0 := \Phi_N^{-1}(v_0)$ fulfills

$$\|w_0\|_{\mathcal{P}^{s,\sigma}} \leq \frac{R}{N^2}(1 + CR) \,,$$

and, denoting by $w(t)$ the solution in Birkhoff coordinates, one has $\|w_0\|_{\mathcal{P}^{s,\sigma}} = \|w(t)\|_{\mathcal{P}^{s,\sigma}}$. Thus, provided $0 < R < R'_{s,\sigma}$ is small enough one has

$$\|v(t)\|_{\mathcal{P}^{s,\sigma}} = \|\Phi_N(w(t))\|_{\mathcal{P}^{s,\sigma}} \leq \frac{R}{N^2}(1 + C'R)$$

which implies the thesis.   □

3.2. Proof of Theorem 1.7

The proof is based on the construction of the first terms of the Taylor expansion of Φ_N through Birkhoff normal form (following [22]). To this end we work with the complex variables (ξ,η) (defined in (2.3)) and will eventually restrict to the real subspace $\mathcal{P}_{\mathbb{R}}^{s,\sigma}$.

Remark 3.10. Consider the Taylor expansion of Φ_N at the origin, one has

$$\Phi_N = \mathbb{1} + Q^{\Phi_N} + O(\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}}^3) \,,$$

then Q^{Φ_N} is a bounded quadratic polynomial. Furthermore, since Φ_N is canonical, Q^{Φ_N} is a Hamiltonian vector field, i.e. there exists a cubic complex valued polynomial χ_{Φ_N} s.t. Q^{Φ_n} is the Hamiltonian vector field of χ_{Φ_N}.

We need a preliminary result about a uniqueness property of the transformation introducing Birkhoff coordinates (called below Birkhoff map).

Lemma 3.11. *Let Φ_N and Ψ_N be Birkhoff maps for H_{Toda}, analytic in some neighborhood of the origin; assume that $d\Phi_N(0,0) \equiv d\Psi_N(0,0) = \mathbb{1}$ and denote by χ_{Φ_N} and χ_{Ψ_N} the Hamiltonian functions corresponding to Q^{Φ_N} and Q^{Ψ_N} respectively, then one has*

$$\{H_0; \chi_{\Phi_N} - \chi_{\Psi_N}\} = 0 \,, \tag{3.34}$$

where H_0 is defined in (1.6).

Proof. By a standard computation of the Taylor expansion one has

$$H_{Toda} \circ \Phi_N = H_0 + \{H_0, \chi_{\Phi_N}\} + H_1 + \text{h.o.t.}$$

where H_1 is the function

$$H_1(q) = \sum_{j=0}^{N-1} \frac{(q_j - q_{j+1})^3}{6} \,.$$

Since Φ_N is a Birkhoff map, namely a map introducing Birkhoff coordinates, it follows that $H_{Toda} \circ \Phi_N$ is a function of the actions $(\xi_j \eta_j)_j$ alone, so in particular its Taylor expansion contains only terms of even degree. Thus the cubic terms in the expansion above must vanish: $\{H_0, \chi_{\Phi_N}\} + H_1 = 0$. The same argument holds also for the map Ψ_N, thus the thesis follows. $\square$

Remark 3.12. Writing as usual

$$\chi_{\Phi_N}(\xi, \eta) = \sum_{|K|+|L|=3} \chi_{K,L} \xi^K \eta^L \,,$$

one gets that, since

$$\{H_0, \chi_{\Phi_N}\} = -\sum_{|K|+|L|=3} i\omega \cdot (K - L)\chi_{K,L}\, \xi^K \eta^L \,,$$

eq. (3.34) implies that, if for some K, L one has $\omega \cdot (K - L) \neq 0$, then $\chi_{K,L}$ is unique and coincides with $\frac{H_{K,L}}{i\omega \cdot (K-L)}$ with an obvious definition of $H_{K,L}$.

Lemma 3.13. *In terms of the variables* (ξ, η) *one has*

$$H_1(\xi, \eta)$$

$$= \frac{1}{12\sqrt{2N}} \left[\sum_{\substack{k_1+k_2+k_3=0 \bmod N \\ 1\le k_1,k_2,k_3\le N-1}} (-1)^{\frac{k_1+k_2+k_3}{N}} \sqrt{\omega_{k_1}}\sqrt{\omega_{k_2}}\sqrt{\omega_{k_3}} \left(\xi_{k_1}\xi_{k_2}\xi_{k_3} + \eta_{k_1}\eta_{k_2}\eta_{k_3}\right) \right.$$

$$\left. + 3 \sum_{\substack{k_1+k_2-k_3=0 \bmod N \\ 1\le k_1,k_2,k_3\le N-1}} (-1)^{\frac{k_1+k_2-k_3}{N}} \sqrt{\omega_{k_1}}\sqrt{\omega_{k_2}}\sqrt{\omega_{k_3}} \left(\xi_{k_1}\xi_{k_2}\eta_{k_3} + \eta_{k_1}\eta_{k_2}\xi_{k_3}\right) \right]$$

Proof. First remark that

$$q_j - q_{j+1} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{q}_k \left(1 - e^{-\frac{2\pi i k}{N}}\right) e^{-\frac{2\pi i j k}{N}} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} i\omega_k e^{-\frac{i\pi k}{N}} \hat{q}_k e^{-\frac{2\pi i j k}{N}} \ ,$$

so that

$$\frac{1}{6}\sum_{j=0}^{N-1}(q_j - q_{j+1})^3 = \frac{i^3}{6N^{3/2}} \sum_{k_1,k_2,k_3} \omega_{k_1}\hat{q}_{k_1}\omega_{k_2}\hat{q}_{k_2}\omega_{k_3}\hat{q}_{k_3} e^{-\frac{i\pi}{N}(k_1+k_2+k_3)} \sum_{j=0}^{N-1} e^{\frac{2\pi i j}{N}(k_1+k_2+k_3)}$$

$$= \frac{i^3}{6N^{1/2}} \sum_{k_1+k_2+k_3=0 \bmod N} (-1)^{\frac{k_1+k_2+k_3}{N}} \omega_{k_1}\hat{q}_{k_1}\omega_{k_2}\hat{q}_{k_2}\omega_{k_3}\hat{q}_{k_3} \ .$$

Substituting

$$\omega_k \hat{q}_k = \sqrt{\omega_k}\frac{\xi_k - \eta_{N-k}}{i\sqrt{2}}$$

and reorganizing the terms one gets the thesis. $\square$

Lemma 3.14. *For any* $s \ge 0$, $\sigma \ge 0$, *there exists* $C > 0$ *s.t. one has*

$$\left\| Q^{\phi_N}(\bar{v}) \right\|_{\mathcal{P}^{s,\sigma}} \ge CN^2 \left\| \bar{v} \right\|_{\mathcal{P}^{s,\sigma}}^2 \ , \tag{3.35}$$

where $\bar{v} = ((\xi_1, 0, 0, \ldots, 0), (\bar{\xi}_1, 0, 0, \ldots, 0)) \in \mathcal{P}_{\mathbb{R}}^{s,\sigma}$.

Proof. In this proof, for clarity we denote $\eta_1 := \bar{\xi}_1$, and similarly for the other variables. We are going to compute the ξ_2 component $[Q^{\Phi_N}(\bar{v})]_{\xi_2}$ of $Q^{\Phi_N}(\bar{v})$ and exploit the inequality

$$\left\| Q^{\Phi_N}(\bar{v}) \right\|_{\mathcal{P}^{s,\sigma}} \ge \frac{1}{\sqrt{N}} 2^s e^{2\sigma} \omega_2^{1/2} \frac{1}{\sqrt{2}} \left| \left[Q^{\Phi_N}(\bar{v})\right]_{\xi_2} \right| = \frac{2^s e^{2\sigma} \omega_2^{1/2}}{\sqrt{2N}} \left| \frac{\partial \chi_{\Phi_N}}{\partial \eta_2}(\bar{v}) \right| \ ; \tag{3.36}$$

the only monomials in χ_{Φ_N} contributing to such a quantity are quadratic in (ξ_1, η_1) and linear in η_2, but due to the selection rule $k_1 \pm k_2 \pm k_3 = lN$ with a plus for the ξ's and a minus for the η's the only monomial contributing to the r.h.s. of (3.36) is $\chi_{\bar{K},\bar{L}} \xi^{\bar{K}} \eta^{\bar{L}}$ with $\bar{K} := (2, 0, \ldots, 0)$, and $\bar{L} = (0, 1, 0, 0, \ldots, 0)$.

Since

$$\omega \cdot (K - L) = 2\omega_1 - \omega_2 = 4\sin\frac{\pi}{N} - 2\sin\frac{2\pi}{N} = \frac{2\pi^3}{N^3} + O\left(\frac{1}{N^5}\right) \neq 0 , \quad (3.37)$$

such a coefficient is uniquely defined and, for the χ_{Φ_N} corresponding to *any* Birkhoff map, one has

$$\chi_{\bar{K},\bar{L}} = \frac{1}{4\sqrt{2N}} \frac{\omega_1 \omega_2^{1/2}}{\mathrm{i}(2\omega_1 - \omega_2)} . \quad (3.38)$$

Inserting in (3.36) one has that its r.h.s. is equal to

$$\frac{2^s e^{2\sigma} \omega_2^{1/2}}{\sqrt{2N}} \left|\chi_{\bar{K},\bar{L}}\right| |\xi_1|^2 = \frac{C''}{N} \frac{\omega_1 \omega_2}{|2\omega_1 - \omega_2|} |\xi_1|^2 = C' \frac{\omega_2}{|2\omega_1 - \omega_2|} \|\bar{v}\|_{\mathcal{P}^{s,\sigma}}^2 \geq CN^2 \|\bar{v}\|_{\mathcal{P}^{s,\sigma}}^2 ,$$

where C, C' and C'' are numerical constants independent of N and we used the expansions of ω_1, ω_2 in $1/N$ as well as equation (3.37). $\quad\square$

Proof of Theorem 1.7. The thesis immediately follows taking $\|\bar{v}\|_{\mathcal{P}^{s,\sigma}} = R/N^\alpha$ and imposing the inequality (1.11). $\quad\square$

Proof of Corollary 1.9. By Cauchy inequality and assumption (1.12) Q^{Φ_N} fulfills

$$\left\|Q^{\Phi_N}(\bar{v})\right\|_{\mathcal{P}^{s,\sigma}} \leq \frac{R'}{N^{\alpha'}} \frac{N^{2\alpha}}{R^2} \|\bar{v}\|_{\mathcal{P}^{s,\sigma}}^2 . \quad (3.39)$$

Comparing this inequality with (3.35), one gets

$$\frac{R'}{R^2} N^{2\alpha - \alpha'} \geq C'' N^2 ,$$

which in particular implies the thesis. $\quad\square$

4. FPU packet of modes: proofs

In this section we prove the results stated in Subsection 1.2 about the persistence of the metastable packet in the FPU system.

To clarify the procedure, we distinguish here between the (ξ, η) variables and the variables (p, q). Thus, we denote by $T : (\xi, \eta) \to (p, q)$ the change of coordinates of the

phase space introducing the linear Birkhoff variables (ξ, η) defined in (2.3). Furthermore it is useful to use for the (p, q) variables the following norms

$$\|q\|_{s,\sigma}^2 := \frac{1}{N} \sum_{k=0}^{N-1} \max(1, [k]_N^{2s}) e^{2\sigma[k]_N} |\hat{q}_k|^2 , \tag{4.1}$$

and

$$\|(p, q)\|_{\mathcal{P}^{s,\sigma}} := \left\| T^{-1}(p, q) \right\|_{\mathcal{P}^{s,\sigma}} . \tag{4.2}$$

Lemma 4.1. *Fix $s \geq 1$, $\sigma \geq 0$, then there exist constants $C_1, C_2 > 0$, independent of N, such that for all $(\xi, \eta) \in \mathcal{P}^{s,\sigma}$ and $\forall l \geq 2$ one has*

$$\|X_{H_l \circ T}(\xi, \eta)\|_{\mathcal{P}^{s,\sigma}} \leq \frac{C_1^l}{(l+1)!} \|(\xi, \eta)\|_{\mathcal{P}^{s,\sigma}}^{l+1} , \tag{4.3}$$

$$\|X_{H_l \circ T}(\xi, \eta)\|_{\mathcal{P}^{s-1,\sigma}} \leq \frac{C_2^l}{N(l+1)!} \|(\xi, \eta)\|_{\mathcal{P}^{s,\sigma}}^{l+1} . \tag{4.4}$$

Proof. Define the difference operators by

$$S_\pm : \{q_j\}_{0 \leq j \leq N-1} \mapsto \{q_j - q_{j\pm 1}\}_{0 \leq j \leq N-1} , \qquad \text{where } q_N \equiv q_0 , \tag{4.5}$$

and the operator $[S_+(q)]^l$ by

$$\left\{ [S_+(q)]^l \right\}_j := (q_j - q_{j+1})^l ,$$

so that

$$X_{H_l \circ T}(\xi, \eta) = \frac{1}{(l+1)!} T^{-1} \left(S_- [S_+(T(\xi, \eta))]^l , 0 \right) . \tag{4.6}$$

By Lemma B.3 and Remark B.5 in Appendix B, there exists a constant $C_{s,\sigma} > 0$, independent of N, such that for every integer $n \geq 1$

$$\left\| [S_\pm(q)]^{l+1} \right\|_{s,\sigma} \leq C_{s,\sigma}^{l+1} \|S_\pm(q)\|_{s,\sigma}^{l+1} \leq C_{s,\sigma}^{l+1} \|(\xi, \eta)\|_{\mathcal{P}^{s,\sigma}}^{l+1} , \tag{4.7}$$

where for the last inequality we have identified the couple $(0, q)$ with the corresponding (ξ, η) vector.

Then the thesis follows just remarking that $\left\| T^{-1}(q, 0) \right\|_{\mathcal{P}^{s,\sigma}} = \|q\|_{s,\sigma}$, and that S_- is bounded as an operator from $\mathcal{P}^{s,\sigma}$ to itself, while one has

$$\|(S_-(q), 0)\|_{\mathcal{P}^{s-1,\sigma}} \leq \frac{C}{N} \|q\|_{s,\sigma} . \qquad \square$$

Introducing the Birkhoff coordinates and using the standard formulae for the pull back of vector fields[3] one has the following

Corollary 4.2. *Fix $s \geq 1$ and $\sigma \geq 0$, then there exist constants $R_{s,\sigma}, C_1, C_2 > 0$, independent of N, such that for all $w \equiv (\phi, \psi) \in B^{s,\sigma}(R_{s,\sigma}/N^2)$ one has*

$$\left\| X_{H_l \circ T \circ \Phi_N}(w) \right\|_{\mathcal{P}^{s,\sigma}} \leq \frac{C_1^l}{(l+1)!} \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^{l+1} \ , \tag{4.8}$$

$$\left\| X_{H_l \circ T \circ \Phi_N}(w) \right\|_{\mathcal{P}^{s-1,\sigma}} \leq \frac{C_2^l}{N(l+1)!} \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^{l+1} \ . \tag{4.9}$$

Remark 4.3. Write

$$\tilde{H}_{FPU} \equiv H_{FPU} \circ T \circ \Phi_N = \tilde{H}_{Toda} + \tilde{H}_P \ , \tag{4.10}$$

where

$$\tilde{H}_{Toda} := H_{Toda} \circ T \circ \Phi_N \ , \quad \tilde{H}_P := (\beta - 1)H_2 \circ T \circ \Phi_N + H^{(3)} \circ T \circ \Phi_N \ , \tag{4.11}$$

then, provided R is small enough the vector field of $\tilde{H}_P$ fulfills the following estimates

$$\left\| X_{\tilde{H}_P}(w) \right\|_{\mathcal{P}^{s,\sigma}} \leq C \left[|\beta - 1| \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^3 + C \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^4 \right] \ , \tag{4.12}$$

$$\left\| X_{\tilde{H}_P}(w) \right\|_{\mathcal{P}^{s-1,\sigma}} \leq \frac{C}{N} \left[|\beta - 1| \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^3 + C \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^4 \right] \ , \tag{4.13}$$

for all $w \in B^{s,\sigma}(R/N^2)$.

In the following we denote by $v(t) \equiv (\xi(t), \bar{\xi}(t))$ the solution of the FPU model in the original Cartesian coordinates (we restrict to the real subspace). We denote by $w(t) := \Phi_N^{-1}(v(t))$ the same solution in Birkhoff coordinates.

Lemma 4.4. *Fix $s \geq 2$ and $\sigma \geq 0$. Then there exist $R'_{s,\sigma}, T, C_2 > 0$ such that $v_0 \in B_{\mathbb{R}}^{s,\sigma}\left(\frac{R}{N^2}\right)$ with $R \leq R'_{s,\sigma}$ implies $v(t) \in B_{\mathbb{R}}^{s,\sigma}\left(\frac{4R}{N^2}\right)$ for*

$$|t| \leq \frac{T}{R^2 \mu^4 [|\beta - 1| + C_2 R \mu^2]} \ . \tag{4.14}$$

---

[3] Namely

$$[\Phi_N^* X](x) = d\Phi_N^{-1}(\Phi_N(x))X(\Phi_N(x))$$

which gives the vector field of the transformed Hamiltonian due to the fact that Φ_N is canonical.

Proof. First consider $w_0 := \Phi_N^{-1}(v_0)$ and remark that (provided $R'_{s,\sigma}$ is small enough) one has $w_0 \in B_\mathbb{R}^{s,\sigma}\left(\frac{2R}{N^2}\right)$. Denote by $M(w) := \|w\|_{\mathcal{P}_\mathbb{R}^{s,\sigma}}^2$. Since $\left\{M, \tilde{H}_{Toda}\right\} \equiv 0$, one has

$$M(w(t)) = M(w_0) + \int_0^t \left\{M; \tilde{H}_P\right\}(w(s))ds . \tag{4.15}$$

Denoting $\bar{M}(t) := \sup_{|s| \leq t} M(w(s))$, one has

$$\bar{M}(w(t)) \leq M(w_0) + \int_0^t \left|\left\{M; \tilde{H}_P\right\}(w(s))\right| ds \tag{4.16}$$

$$\leq M(w_0) + \int_0^t \left(C \|w(s)\|_{\mathcal{P}_\mathbb{R}^{s,\sigma}}^4 |\beta - 1| + C \|w(s)\|_{\mathcal{P}_\mathbb{R}^{s,\sigma}}^5\right) ds$$

$$\leq M(w_0) + \int_0^t C\bar{M}(t)^2 \left(|\beta - 1| + C\bar{M}(t)^{1/2}\right) ds$$

$$\leq M(w_0) + |t| C\bar{M}(t)^2 \left(|\beta - 1| + C\bar{M}(t)^{1/2}\right) , \tag{4.17}$$

where, in order to prove the second inequality we used $\left\{M; \tilde{H}_P\right\} := dM X_{\tilde{H}_P}$ and

$$\|dM(w)\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \mathbb{C})} \leq C \|w\|_{\mathcal{P}^{s,\sigma}} ,$$

which follows from an explicit computation. Taking t as in the statement of the lemma we have that (4.16)–(4.17) ensures $\bar{M}(t) \leq 9M(w(0))/4$, which implies $w(t) \in B_\mathbb{R}^{s,\sigma}\left(\frac{3R}{N^2}\right)$ from which the thesis immediately follows. □

Proof of Theorem 1.16. Inequality (1.21) is a direct consequence of Lemma 4.4. To prove inequality (1.22) remark that $\dot{I}_k = \{I_k, \tilde{H}_P\} = x_k \frac{\partial \tilde{H}_P}{\partial y_k} - y_k \frac{\partial \tilde{H}_P}{\partial x_k}$. Thus

$$\frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s-2} e^{2\sigma[k]_N} \omega\left(\frac{k}{N}\right) \left|\{I_k, \tilde{H}_P\}\right|$$

$$= \frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s-2} e^{2\sigma[k]_N} \omega\left(\frac{k}{N}\right) \left|y_k \frac{\partial \tilde{H}_P}{\partial y_k} - x_k \frac{\partial \tilde{H}_P}{\partial x_k}\right|$$

$$\leq \left(\frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s-2} e^{2\sigma[k]_N} \omega\left(\frac{k}{N}\right) (y_k^2 + x_k^2)\right)^{1/2}$$

$$\times \left(\frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s-2} e^{2\sigma[k]_N} \omega\left(\frac{k}{N}\right) \left(\left|\frac{\partial \tilde{H}_P}{\partial y_k}\right|^2 + \left|\frac{\partial \tilde{H}_P}{\partial x_k}\right|^2\right)\right)^{1/2}$$

$$\leq 2 \left\| w \right\|_{\mathcal{P}_{\mathbb{R}}^{s-1,\sigma}} \left\| X_{\tilde{H}_P}(w) \right\|_{\mathcal{P}_{\mathbb{R}}^{s-1,\sigma}} \leq \frac{C}{N} \left[|\beta - 1| \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^4 + C \left\| w \right\|_{\mathcal{P}^{s,\sigma}}^5 \right],$$

where in the last inequality we used (4.13). Using that $|I_k(w(t)) - I_k(w(0))| \leq \int_0^t \left| \{ I_k, \tilde{H}_P \}(w(s)) \right| ds$, one gets

$$\frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s-2} e^{2\sigma[k]_N} \omega\left(\frac{k}{N}\right) |I_k(w(t)) - I_k(w(0))|$$

$$\leq \frac{|t|C}{N} \sup_{|s| \leq t} \left[|\beta - 1| \left\| w(s) \right\|_{\mathcal{P}^{s,\sigma}}^4 + C \left\| w(s) \right\|_{\mathcal{P}^{s,\sigma}}^5 \right],$$

which, using $w(t) \in B_{\mathbb{R}}^{s,\sigma}\left(\frac{3R}{N^2}\right)$ immediately implies the thesis. $\square$

Acknowledgments

Appendix A. Properties of normally analytic maps

In this section we study the properties of the space $\mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$ and $\mathcal{A}_{w^1,\rho}^{w^2}$ defined in Section 2, with weights $w^1 \leq w^2$. In particular, we consider the operations on germs defined in [30] and perform quantitative estimates.

Lemma A.1. *Let $w^1 \leq w^2 \leq w^3$ be weights. Let $G \in \mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$ with $|\underline{G}|_\rho \leq \sigma$ and $F \in \mathcal{N}_\sigma(\mathcal{P}^{w^2}, \mathcal{P}^{w^3})$. Then $F \circ G \in \mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{P}^{w^3})$ and $|\underline{F \circ G}|_\rho \leq |\underline{F}|_\sigma$.*

Proof. Exploiting the obvious inequality $\underline{F \circ G}(|v|) \leq \underline{F} \circ \underline{G}(|v|)$ (cf. [30]), one has

$$|\underline{F \circ G}|_\rho \equiv \sup_{v \in B^{w^1}(\rho)} \left\| \underline{F \circ G}(|v|) \right\|_{w^3} \leq \sup_{v \in B^{w^1}(\rho)} \left\| \underline{F}(\underline{G}(|v|)) \right\|_{w^3}$$

$$\leq \sup_{u \in B^{w^2}(\sigma)} \left\| \underline{F}(|u|) \right\|_{w^3} \equiv |\underline{F}|_\sigma. \quad \square$$

Lemma A.2. *Let $F \in \mathcal{N}_\rho(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$, $F = O(v^2)$ and $|\underline{F}|_\rho \leq \rho/e$. Then the map $\mathbb{1} + F$ is invertible in $B^{w^1}(\mu\rho)$, μ as in (2.28). Moreover there exists $G \in \mathcal{N}_{\mu\rho}(\mathcal{P}^{w^1}, \mathcal{P}^{w^2})$, $G = O(v^2)$, such that $(\mathbb{1} + F)^{-1} = \mathbb{1} - G$, and*

$$|\underline{G}|_{\mu\rho} \leq \frac{|\underline{F}|_\rho}{8}. \tag{A.1}$$

Proof. We look for G in the form $G = \sum_{n \geq 2} G^n$, with the homogeneous polynomial G^n to be determined at every order n. Note that the equation defining G can be given in the form $F(v - G(v)) = G(v)$, which can be recasted in a recursive way giving the formula

$$G^n(v) = \sum_{r=2}^{n} \sum_{k_1 + \cdots + k_r = n} \tilde{F}^r \Big(G^{k_1}(v), \cdots, G^{k_r}(v) \Big), \qquad \forall\, n \geq 2 \,. \tag{A.2}$$

In the formula above $k_1, \ldots, k_r \in \mathbb{N}$, and we write $F = \sum_{r \geq 2} F^r$, where F^r is a homogeneous polynomial of degree r and $\tilde{F}^r$ is its associated multilinear map (see (2.7)). Moreover we write $G^1(v) := v$. We show now that the formal series $G = \sum_{n \geq 2} G^n$ with G^n defined by (A.2) is normally analytic in $B^{w^1}(\mu\rho)$. Note that

$$\underline{G^n}(|v|) \leq \sum_{r=2}^{n} \sum_{k_1 + \cdots + k_r = n} \underline{\tilde{F}^r} \Big(\underline{G^{k_1}}(|v|), \ldots, \underline{G^{k_r}}(|v|) \Big). \tag{A.3}$$

In order to prove that the series $\sum_{n \geq 2} \underline{G^n}$ is convergent in $B^{w^1}(\mu\rho)$, we prove that there exists a constant $A > 0$ such that

$$\left\| \underline{G^n}(|v|) \right\|_{w^2} \leq \frac{|F|_\rho}{8Sn^2} A^n \left\| v \right\|_{w^1}^n, \qquad \forall\, n \geq 2. \tag{A.4}$$

The proof is by induction on n. We will use in the following the chain of inequalities

$$\left\| \underline{\tilde{F}^r} \right\| \leq e^r \left\| \underline{F^r} \right\| \leq e^r \left| F \right|_\rho / \rho^r \qquad \forall\, r \geq 1 \,,$$

see [32]. For $n = 2$, by (A.2) it follows that $G^2(v) = \tilde{F}^2(v, v)$. Since

$$\left\| \underline{G^2}(|v|) \right\|_{w^2} \leq \left\| \underline{\tilde{F}^2} \right\| \left\| v \right\|_{w^1}^2 \leq e^2 \frac{|F|_\rho}{\rho^2} \left\| v \right\|_{w^1}^2 \,,$$

it follows that (A.4) holds for $n = 2$ with $A = \frac{e(32S)^{1/2}}{\rho}$. We prove now the inductive step $n - 1 \rightsquigarrow n$. Assume therefore that (A.4) holds up to order $n - 1$. Then one has

$$\left\| \underline{G^n}(|v|) \right\|_{w^2} \leq \sum_{r=2}^{n} \sum_{k_1 + \cdots + k_r = n} \left\| \underline{\tilde{F}^r} \right\| \left\| \underline{G^{k_1}}(|v|) \right\|_{w^2} \cdots \left\| \underline{G^{k_r}}(|v|) \right\|_{w^2}$$

$$\leq A^n \left\| v \right\|_{w^1}^n \sum_{r=2}^{n} \sum_{k_1 + \cdots + k_r = n} e^r \frac{|F|_\rho}{\rho^r} \frac{|F|_\rho^r}{8^r S^r k_1^2 \cdots k_r^2}$$

$$\leq \frac{|F|_\rho}{4Sn^2} A^n \left\| v \right\|_{w^1}^n \sum_{r=2}^{\infty} \left(\frac{e |F|_\rho}{2\rho} \right)^r \leq \frac{|F|_\rho}{8Sn^2} A^n \left\| v \right\|_{w^1}^n$$

where in the first inequality we used the fact that $w^1 \leq w^2$, in the second the inductive assumption and in the last we used the hypothesis $|\underline{F}|_\rho \leq \rho/e$. Finally to pass from the second to the third line we used the following inequality, proved in Lemma A.5 below:

$$n^2 \sum_{k_1 + \cdots + k_r = n} \frac{1}{k_1^2 \cdots k_r^2} \leq (4S)^{r-1}, \qquad n \geq 1 . \tag{A.5}$$

Hence, choosing $\mu\rho = 1/A = \rho/e(32S)^{1/2}$ one proves (A.1). $\quad\square$

Now it is easy to prove the following lemma, giving closedness of the class $\mathcal{A}_{w^1,\rho}^{w^2}$ under different operations.

Lemma A.3. *Let $w^1 \leq w^2$ be weights and let μ be as in (2.28). Then the following holds true:*

i) *Let $F \in \mathcal{A}_{w^1,\rho}^{w^2}$ and $G \in \mathcal{A}_{w^1,\mu\rho}^{w^2}$ with $\|G\|_{\mathcal{A}_{w^1,\mu\rho}^{w^2}} < \frac{\mu\rho}{e}$. Then $H(v) := F(v + G(v))$ is of class $\mathcal{A}_{w^1,\mu\rho}^{w^2}$ and*

$$\|H\|_{\mathcal{A}_{w^1,\mu\rho}^{w^2}} \leq 2\|F\|_{\mathcal{A}_{w^1,\rho}^{w^2}} .$$

ii) *Let $F \in \mathcal{A}_{w^1,\rho}^{w^2}$ and $\|F\|_{\mathcal{A}_{w^1,\rho}^{w^2}} \leq \rho/e$. Then $(\mathbb{1} + F)^{-1} = \mathbb{1} + G$, with $G \in \mathcal{A}_{w^1,\mu\rho}^{w^2}$. Moreover one has*

$$\|G\|_{\mathcal{A}_{w^1,\mu\rho}^{w^2}} \leq 2\|F\|_{\mathcal{A}_{w^1,\rho}^{w^2}} . \tag{A.6}$$

iii) *Let $F \in \mathcal{A}_{w^1,\rho}^{w^2}$, then the function $H(v) := dF(v)v$ is in the class $\mathcal{A}_{w^1,\mu\rho}^{w^2}$ and*

$$\|H\|_{\mathcal{A}_{w^1,\mu\rho}^{w^2}} \leq 2\|F\|_{\mathcal{A}_{w^1,\rho}^{w^2}} .$$

iv) *Let $F^0, G^0 \in \mathcal{A}_{w^1,\rho}^{w^2}$ with $\|F^0\|_{\mathcal{A}_{w^1,\rho}^{w^2}} \leq \frac{\rho}{e}$. Denote $F = \mathbb{1} + F^0$. Then $H(v) := dG^0(v)^*(F(v))$ is in the class $\mathcal{A}_{w^1,\mu\rho}^{w^2}$ and*

$$\|H\|_{\mathcal{A}_{w^1,\mu\rho}^{w^2}} \leq 2\|G^0\|_{\mathcal{A}_{w^1,\rho}^{w^2}} .$$

Proof.

i) Since $\underline{H}(|v|) \leq \underline{F}(|v| + \underline{G}(|v|))$ it follows that $|\underline{H}|_{\mu\rho} \leq |\underline{F}|_{2\mu\rho} \leq |\underline{F}|_\rho$. Furthermore, since $dH(v) = dF(v + G(v))(\mathbb{1} + dG(v))$ one gets that $\underline{dH}(|v|) \leq \underline{dF}(|v| + \underline{G}(|v|)) + \underline{dF}(|v| + \underline{G}(|v|))\underline{dG}(|v|)$, which implies that $\mu\rho|\underline{dH}|_{\mu\rho} \leq |\underline{dF}|_\rho(\mu\rho + \mu\rho|\underline{dG}|_{\mu\rho}) \leq |\underline{dF}|_\rho\mu\rho(1 + 1/e)$. The adjoint $dH(v)^*$ is estimated analogously, thus the claimed estimate follows.

ii) It follows from the formula $dG(v) = [\mathbb{1} - dF(v - G(v))]^{-1} dF(v - G(v))$, arguing as in item *i*).

iii) It follows from $dH(v)u = dF(v)u + d^2F(v)(u, v)$, arguing as in item *i*).

iv) To estimate $\underline{H}(|v|)$ and $\underline{dH}(|v|)$ one proceeds as in item *i*). In order to estimate $\underline{dH}(|v|)^*$ remark that (see [30]) $dH(v)^*u = (dF^0(v)^* + \mathbb{1})dG^0(v)u + d_v(dG^0(v)^*u)(F(v))$, thus

$$\underline{dH}(|v|)^*|u| \leq (\underline{dF^0}(|v|)^* + \mathbb{1})\underline{dG^0}(|v|)|u| + d_{|v|}(\underline{dG^0}(|v|)^*|u|)(\underline{F}(|v|)) \; .$$

The claimed estimate follows easily. □

Now we analyze the flow generated by a vector field of class $\mathcal{A}_{w^1,\rho}^{w^2}$. Given a time dependent vector field $V_t(v)$, consider the differential equation

$$\begin{cases} \dot{u}(t) = V_t(u(t)) \\ u(0) = v \; . \end{cases} \tag{A.7}$$

We will denote by $\phi^t(v)$ the corresponding flow map whose existence and properties are given in the next lemma.

Lemma A.4. *Assume that the map* $[0,1] \ni t \mapsto V_t \in \mathcal{A}_{w^1,\rho}^{w^2}$ *is continuous and furthermore fulfills* $\sup_{t \in [0,1]} \|V_t\|_{\mathcal{A}_{w^1,\rho}^{w^2}} \leq \rho/e$; *then for each* $t \in [0,1]$, $\phi^t - \mathbb{1} \in \mathcal{A}_{w^1,\mu\rho}^{w^2}$ *with* μ *as in* (2.28). *Furthermore one has*

$$\left\| \phi^t - \mathbb{1} \right\|_{\mathcal{A}_{w^1,\mu\rho}^{w^2}} \leq 2 \sup_{t \in [0,1]} \|V_t\|_{\mathcal{A}_{w^1,\rho}^{w^2}} \; . \tag{A.8}$$

Proof. We look for a solution $u(t,v) = \sum_{j \geq 1} u^j(t,v)$ in power series of v, with $u^j(t,v)$ a homogeneous polynomial of degree j in v. Expanding the vector field $V_t(v) = \sum_{r \geq 2} V_t^r(v)$ in Taylor series, one obtains the recursive formula for the solution

$$u^1(t,v) = v,$$

$$u^n(t,v) = \sum_{r=2}^{n} \sum_{k_1 + \cdots + k_r = n} \int_0^t \tilde{V}_s^r(u^{k_1}(s,v), \ldots, u^{k_r}(s,v)) \, ds \quad \forall n \geq 2, \tag{A.9}$$

where $\tilde{V}_s^r$ is the multilinear map associated to V_s^r (see (2.7)). Arguing as in the proof of (A.2) one gets the bounds

$$\|\underline{u}^n(t,v)\|_{w^2} \leq \frac{\sup_{t \in [0,1]} |V_t|_\rho}{8Sn^2} A^n \|v\|_{w^1}^n \quad \forall n \geq 2, \tag{A.10}$$

with $A = \frac{e}{\rho}(32S)^{1/2}$, from which it follows that $\left| \phi^t - \mathbb{1} \right|_{\mu\rho} \leq \sup_{t \in [0,1]} \left| \underline{V}_t \right|_\rho / 8$.

We come to the estimate of the differential of $u(t,v)$ and of its adjoint. We differentiate equation (A.9) getting the recursive formula

$$du^n(t,v)\xi = \sum_{r=2}^{n} \sum_{k_1+\cdots+k_r=n} \int_0^t \left[\tilde{V}_s^r(du^{k_1}(s,v)\xi,\ldots,u^{k_r}(s,v)) + \cdots \right.$$
$$\left. + \tilde{V}_s^r(u^{k_1}(s,v),\ldots,du^{k_r}(s,v)\xi)\right] ds . \tag{A.11}$$

To estimate such an expression remark that, defining $E_t(v) := dV_t(v)$ (where the differential is with respect to the v variable only), one has

$$d^{r-1}E_s(u^{k_2}(s,v),\ldots,u^{k_r}(s,v))\xi = \tilde{V}_s^r(\xi,u^{k_2}(s,v),\ldots,u^{k_r}(s,v))$$

which allows to write formula (A.11) as

$$du^n(t,v)\xi = \sum_{r=2}^{n} \sum_{k_1+\cdots+k_r=n} \int_0^t \left[d^{r-1}E_s(u^{k_2}(s,v),\ldots,u^{k_r}(s,v))du^{k_1}(s,v)\xi + \ldots \right.$$
$$\left. \ldots + d^{r-1}E_s(u^{k_1}(s,v),\ldots,u^{k_{r-1}}(s,v))du^{k_r}(s,v)\xi\right] ds . \tag{A.12}$$

This formula allows to proceed exactly as in the estimate of $\underline{u}^n$, namely making the inductive assumption that

$$\left\|\underline{du}^n(t,v)\right\|_{\mathcal{L}(\mathcal{P}^{w^1},\mathcal{P}^{w^2})} \le \frac{\sup_{t\in[0,1]}\left|dV_t\right|_\rho}{8Sn^2}A^n\left\|v\right\|_{w^1}^n$$

and proceeding as above one gets the thesis. Finally one has to estimate $[\underline{du}^n]^*$, but again equation (A.12) allows to obtain a formula whose estimate is obtained exactly as the estimate of $\underline{du}$. $\square$

We prove now a useful inequality.

Lemma A.5. *(See [37].) Let $r \in \mathbb{N}$ be fixed and $S = \sum_{k\geq 1}\frac{1}{k^2}$. Then for every $n \in \mathbb{N}$ it holds that*

$$n^2 \sum_{\substack{k_1,\ldots,k_r\in\mathbb{N} \\ k_1+\cdots+k_r=n}} \frac{1}{k_1^2\cdots k_r^2} \le (4S)^{r-1} .$$

Proof. The proof is by induction, the case $n = 1$ being trivial. For $n > 1$ one gets

$$n^2 \sum_{k_1+\cdots+k_r=n} \frac{1}{k_1^2\cdots k_r^2} = \sum_{k_1+j=n} \frac{n^2}{k_1^2 j^2} \sum_{k_2+\cdots+k_r=j} \frac{j^2}{k_2^2\cdots k_r^2} \le \sum_{k_1+j=k} \frac{n^2}{k_1^2 j^2}(4S)^{r-2}$$

by the induction assumption. Now it is enough to note that

$$\sum_{k_1+j=n} \frac{n^2}{k_1^2 j^2} = \sum_{k_1+j=n} \frac{n^2}{k_1^2(n-k_1)^2} \leq 2\sum_{k_1=1}^{n-1}\left(\frac{1}{k_1^2}+\frac{1}{(n-k_1)^2}\right) \leq 4\sum_{k_1=1}^{n-1}\frac{1}{k_1^2} \leq 4S. \qquad \square$$

Appendix B. Discrete Fourier transform

In this section we collect some well-known properties of the discrete Fourier transform (DFT). For $u \in \mathbb{C}^N$, $N \in \mathbb{N}$, the DFT of u is the vector $\hat{u} \in \mathbb{C}^N$ whose kth component is defined by

$$\hat{u}_k = \frac{1}{\sqrt{N}}\sum_{j=0}^{N-1} u_j e^{2\pi i jk/N}, \qquad \forall\, 0 \leq k \leq N-1. \tag{B.1}$$

When the DFT is considered as a map, it will be denoted by $\mathcal{F}$, i.e. $\mathcal{F}: u \mapsto \hat{u}$.

For any $s \geq 0$ and $\sigma \geq 0$ we endow $\mathbb{C}^N$ with the norm $\|\cdot\|_{s,\sigma}$ defined in (4.1). Such a space will be denoted by $\mathbb{C}^{s,\sigma}$.

Remark B.1. Let j be an integer such that $0 \leq j \leq N-1$. Then

$$\sum_{k=0}^{N-1} e^{i2\pi jk/N} = \begin{cases} 0 & \text{if } j \neq 0 \\ N & \text{if } j = 0 \end{cases} \quad \text{and}$$

$$\sum_{k=0}^{2N-1} u_k\, e^{i\pi kj/N} = \begin{cases} 2\sqrt{N}\,\hat{u}_l, & j \text{ even}, j = 2l \\ 0 & j \text{ odd} \end{cases} \tag{B.2}$$

Remark B.2. Fix $s > \frac{1}{2}$ and $\sigma \geq 0$. Then there exists a constant $C_{s,\sigma} > 0$, independent of N, such that for every $u \in \mathbb{C}^N$ the following estimate holds:

$$\sup_{0 \leq j \leq N-1} |u_j| \leq C_{s,\sigma} \|u\|_{s,\sigma} .$$

For $u, v \in \mathbb{C}^N$, we denote by $u \cdot v$ the component-wise product of u and v, namely the vector whose jth component is given by the product of the jth components of u and v:

$$(u \cdot v)_j := u_j v_j, \qquad 0 \leq j \leq N-1 . \tag{B.3}$$

We denote by $u * v$ the convolution product of u and v, a vector whose jth component is defined by

$$(u * v)_j := \sum_{k=0}^{N-1} u_k v_{j-k}, \qquad 0 \leq j \leq N-1 , \tag{B.4}$$

where in the summation above u and v are extended periodically defining $v_{k+lN} \equiv v_k$ for $l \in \mathbb{Z}$. The DFT maps the component-wise product in convolution:

Lemma B.3. *For $s > \frac{1}{2}$ and $\sigma \geq 0$ there exists a constant $C_{s,\sigma} > 0$, independent of N, such that the following holds:*

- *(i)* $\widehat{u \cdot v} = \frac{1}{\sqrt{N}} \hat{u} * \hat{v}$;
- *(ii)* $\|u \cdot v\|_{s,\sigma} \leq C_{s,\sigma} \|u\|_{s,\sigma} \|v\|_{s,\sigma}$;
- *(iii)* *the map $X : u \mapsto u^2$, has bounded modulus w.r.t. the exponentials, and $\|\underline{X}(u)\|_{s,\sigma} \leq C_{s,\sigma} \|u\|_{s,\sigma}^2$.*

Proof. Item (i) is standard and the details of the proof are omitted.

We prove now item (ii). To begin, note that, by periodicity, one has

$$\|u\|_{s,\sigma}^2 = \frac{1}{N} \sum_{k \in K_N^0} [k]^{2s} e^{2\sigma|k|} |\hat{u}_k|^2 \ ,$$

where the set

$$K_N^0 := \{k \in \mathbb{Z} : -(N-1)/2 \leq k \leq (N-1)/2\} \cup \{\lfloor N/2 \rfloor\}, \tag{B.5}$$

while $[k] := \max(1, |k \bmod N|)$. By item (i), one has that

$$\|u \cdot v\|_{s,\sigma}^2 = \frac{1}{N} \sum_{k \in K_N^0} [k]^{2s} e^{2\sigma|k|} |\widehat{(u \cdot v)}_k|^2 = \frac{1}{N^2} \sum_{k \in K_N^0} [k]^{2s} e^{2\sigma|k|} \left| \sum_{l=0}^{N-1} \hat{u}_l \hat{v}_{k-l} \right|^2 . \tag{B.6}$$

Introduce now the quantities

$$\gamma_{k,l} := \frac{[k]^s}{[l]^s [k-l]^s} \cdot \frac{e^{\sigma|k|}}{e^{\sigma|l|} e^{\sigma|k-l|}} \ .$$

For $s > \frac{1}{2}$ and $\sigma \geq 0$, it holds that $\gamma_{k,l}^2 \leq 4^s \frac{([k-l]^{2s}+[l]^{2s}) \, e^{2\sigma(|k-l|+|l|)}}{[k-l]^{2s} \, [l]^{2s} \, e^{2\sigma|l|} \, e^{2\sigma|k-l|}} \leq 4^s \left(\frac{1}{[l]^{2s}} + \frac{1}{[k-l]^{2s}} \right)$, from which it follows that there exists a constant $C_{s,\sigma} > 0$, independent of N, such that

$$\sup_{0 \leq k \leq N-1} \sum_{l=0}^{N-1} \gamma_{k,l}^2 \leq C_{s,\sigma}^2 \ . \tag{B.7}$$

By Cauchy–Schwartz one has

$$[k]^s e^{\sigma|k|} \sum_{l=0}^{N-1} |\hat{u}_l| \, |\hat{v}_{k-l}| = \sum_{l=0}^{N-1} \gamma_{k,l} \, [l]^s \, e^{\sigma|l|} \, |\hat{u}_l| \, [k-l]^s \, e^{\sigma|k-l|} \, |\hat{v}_{k-l}|$$

$$\leq \left(\sum_{l=0}^{N-1} \gamma_{k,l}^2 \right)^{1/2} \left(\sum_{l=0}^{N-1} [l]^{2s} \, e^{2\sigma|l|} \, |\hat{u}_l|^2 \, [k-l]^{2s} \, e^{2\sigma|k-l|} \, |\hat{v}_{k-l}|^2 \right)^{1/2} .$$

Inserting the inequality above in (B.6), one has

$$\|u \cdot v\|_{s,\sigma} \leq \frac{C_{s,\sigma}}{N} \left(\sum_{l=0}^{N-1} [l]^{2s} \, e^{2\sigma|l|} \, |\hat{u}_l|^2 \right)^{1/2} \left(\sum_{k=0}^{N-1} [k-l]^{2s} \, e^{2\sigma|k-l|} \, |\hat{v}_{k-l}|^2 \right)^{1/2}$$

$$\leq C_{s,\sigma} \, \|u\|_{s,\sigma} \, \|v\|_{s,\sigma} .$$

We prove now item (iii). Consider $\widehat{X} := \mathcal{F} X \mathcal{F}^{-1}$. By item (i) one has $\widehat{X} : \{\hat{u}_j\}_{j \in \mathbb{Z}} \mapsto \{\frac{1}{\sqrt{N}} \sum_l \hat{u}_l \hat{u}_{j-l}\}_{j \in \mathbb{Z}}$. Thus $\underline{\widehat{X}} \equiv \widehat{X}$ and the claim follows. $\square$

Remark B.4. Let $S_\pm$ be the difference operators defined in (4.5). Let $\hat{\omega}_\pm$ be the vectors whose kth components are given by $\hat{\omega}_{\pm,k} := 1 - e^{\mp 2\pi i k/N}$. Then the following holds:

(i) the map $\widehat{S}_\pm := \mathcal{F} S_\pm \mathcal{F}^{-1}$ is a multiplication by the vector $\hat{\omega}_\pm$: $\widehat{S}_\pm : \hat{u} \mapsto \hat{\omega}_\pm \cdot \hat{u}$.
(ii) $\left| \widehat{S}_\pm(\hat{u}) \right| \leq \omega \cdot |\hat{u}|$, where $\omega \equiv \{\omega\left(\frac{k}{N}\right)\}_{k=1}^{N-1}$ is the vector of the linear frequencies.

Remark B.5. Consider $q = q(\xi, \eta)$ as a function of the linear Birkhoff variables defined in (2.3). Then one has $\left\| S_\pm(q) \right\|_{s,\sigma} \leq \|(\xi, \eta)\|_{\mathcal{P}^{s,\sigma}}$.

Appendix C. Proof of Proposition 3.3

We prove now property (Θ1). Let $T : (\xi, \eta) \mapsto (p, q)$ be the map introducing linear Birkhoff coordinates. Explicitly $(p, q) = T(\xi, \eta)$ iff $(\hat{p}_0, \hat{q}_0) = (0, 0)$ and

$$(\hat{p}_k, \hat{q}_k) = \left(\sqrt{\frac{1}{2}\omega\left(\frac{k}{N}\right)} \, (\xi_k + \eta_{N-k}), \ \frac{1}{i\sqrt{2\omega\left(\frac{k}{N}\right)}} (\xi_k - \eta_{N-k}) \right), \ 1 \leq k \leq N - 1 .$$

Then $\Theta_\Xi \equiv \Theta \circ T$ and in particular $d\Theta_\Xi(0,0) = d\Theta(0,0)T$. Using the formula above and the fact that $d\Theta(0,0)(P, Q) = \left(-P, \frac{1}{2}S_+(Q)\right)$, where S_+ is defined in (4.5), one obtains easily formula (3.13). The estimate of $\left\| d\Theta_\Xi(0,0) \right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \, \mathcal{C}^{s,\sigma})}$ is trivial, and is omitted.

We prove now the estimate for $\left\| d\Theta_\Xi(0,0)^* \right\|_{\mathcal{L}(\mathcal{C}^{s+1,\sigma}, \, \mathcal{P}^{s,\sigma})}$. Using the explicit formula (3.13), one computes that $(\xi, \eta) = d\Theta_\Xi(0,0)^*(B, A)$ iff

$$(\xi_k, \ \eta_k) = \left(-\sqrt{\frac{1}{2}\omega\left(\frac{k}{N}\right)} \hat{B}_k + \frac{\varpi_k}{i\sqrt{2\omega\left(\frac{k}{N}\right)}} \hat{A}_k, \ -\sqrt{\frac{1}{2}\omega\left(\frac{k}{N}\right)} \hat{B}_{N-k} - \frac{\varpi_k}{i\sqrt{2\omega\left(\frac{k}{N}\right)}} \hat{A}_{N-k} \right)$$

for $1 \leq k \leq N - 1$. Thus there exist constants $C, C_{\Theta_2} > 0$, independent of N, such that

$$\left\|\underline{d\Theta_\Xi}(0,0)^*(B,A)\right\|_{\mathcal{P}^{s,\sigma}} \leq C \left(\frac{1}{N} \sum_{k=1}^{N-1} [k]_N^{2s} e^{2\sigma[k]_N} \omega\left(\tfrac{k}{N}\right)^2 \left(|\widehat{B}_k|^2 + |\widehat{A}_k|^2\right)\right)^{1/2}$$

$$\leq \frac{C_{\Theta_2}}{N} \left\|(B,A)\right\|_{\mathcal{C}^{s+1,\sigma}} ,$$

where we used that $\left|\omega\left(\tfrac{k}{N}\right)\right|^2 \leq \frac{\pi^2 [k]_N^2}{N^2}$. Thus the second of (3.14) is proved.

We prove now property (Θ2). Denote by Θ_b the map $p \mapsto -p$ and by Θ_a the map $q \mapsto \exp\left(\tfrac{1}{2} S_+(q)\right) - 1$. Then $(b,a) = \Theta(p,q) \equiv (\Theta_b(p), \Theta_a(q))$. Introduce on $\mathbb{C}^N$ the norm $\|\cdot\|_{s,\sigma}$ defined in (4.1). Then $\|\Theta(p,q)\|_{\mathcal{C}^{s,\sigma}}^2 \equiv \|\Theta_b(p)\|_{s,\sigma}^2 + \|\Theta_a(q)\|_{s,\sigma}^2$. The analyticity of $p \mapsto \Theta_b(p)$ is obvious. Consider now the map $q \mapsto \Theta_a(q)$. Expand Θ_a in Taylor series with center at the origin to get

$$\Theta_a(q) = \sum_{r\geq 1} \Theta_a^r(q), \qquad \Theta_a^r(q) := \frac{1}{r!\, 2^r} (S_+(q))^r, \qquad \forall r \geq 1. \tag{C.1}$$

Consider q as a function of the linear Birkhoff variables ξ, η. Then Lemma B.3 and Remark B.5 imply that for any $s \geq 0$, $\sigma \geq 0$

$$\left\|\underline{\Theta_a^r}(q)\right\|_{s+1,\sigma} \leq C_1^r \left\|\underline{S_+}(q)\right\|_{s+1,\sigma}^r \leq C_2^r \|(\xi,\eta)\|_{\mathcal{P}^{s+1,\sigma}}^r$$

$$\leq C_3^r N^r \|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}}^r , \quad \forall r \geq 2, \tag{C.2}$$

where $C_1, C_2, C_3 > 0$ are positive constants independent of N. Therefore for $\epsilon < \frac{1}{C_3}$ one has

$$\sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2} \left\|\underline{\Theta_\Xi^0}(\xi,\eta)\right\|_{\mathcal{C}^{s+1,\sigma}} \leq \sum_{r\geq 2} \sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2} \left\|\underline{\Theta_\Xi^r}(\xi,\eta)\right\|_{\mathcal{C}^{s+1,\sigma}}$$

$$\leq \sum_{r\geq 2} C_3^r N^r \frac{\epsilon^r}{N^{2r}} \leq \frac{2C_3^2 \epsilon^2}{N^2} .$$

This proves the first estimate in (Θ2). We show now that for any $s \geq 0$, $\sigma \geq 0$ one has $[d\Theta_\Xi^0]^* \in \mathcal{N}_{\epsilon/N^2}(\mathcal{P}^{s,\sigma}, \mathcal{L}(\mathcal{C}^{s+2,\sigma}, \mathcal{P}^{s+1,\sigma}))$. Note that $d\Theta_\Xi(\xi,\eta)^* = T^* d\Theta(T(\xi,\eta))^*$. Using the explicit expression of T, one verifies that $(\xi,\eta) = T^*(P,Q)$ iff

$$(\xi_k, \eta_k) = \left(\sqrt{\tfrac{1}{2}\omega\left(\tfrac{k}{N}\right)}\widehat{P}_k + \frac{1}{i\sqrt{2\omega\left(\tfrac{k}{N}\right)}}\widehat{Q}_k, \ \sqrt{\tfrac{1}{2}\omega\left(\tfrac{k}{N}\right)}\widehat{P}_{N-k} - \frac{1}{i\sqrt{2\omega\left(\tfrac{k}{N}\right)}}\widehat{Q}_{N-k}\right)$$

$$\tag{C.3}$$

for $1 \leq k \leq N-1$. Thus one has that for any $s \geq 0$, $\sigma \geq 0$

$$\|T^*(0,Q)\|_{\mathcal{P}^{s,\sigma}} \leq \|Q\|_{s,\sigma} . \tag{C.4}$$

Using (C.1) one verifies that $d\Theta^r(p,q)(P,Q) = \frac{1}{(r-1)!\,2^r}\left(0,\,S_+(q)^{r-1}\cdot S_+(Q)\right), \forall r \geq 2$, from which it follows that

$$d\Theta^r(p,q)^*(B,A) = \frac{1}{(r-1)!\,2^r}\left(0,\,\overline{S_+(q)}^{\,r-1}\cdot S_-(A)\right), \qquad \forall r \geq 2 .$$

Thus, using estimate (C.4), there exists a constant $C_4 > 0$, independent of N, such that

$$\left\|\underline{d\Theta^r_{\underline{\underline{\Xi}}}}(\xi,\eta)^*(B,A)\right\|_{\mathcal{P}^{s+1,\sigma}} \leq C_4^r \left\|\underline{S_+(\underline{q}(\xi,\eta))}\right\|_{s+1,\sigma}^{r-1}\left\|\underline{S_-(A)}\right\|_{s+1,\sigma}$$

$$\leq C_4^r N^{r-2}\,\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}}^{r-1}\,\|(B,A)\|_{\mathcal{C}^{s+2,\sigma}} .$$

Then there exist $C_5, \epsilon_0 > 0$, independent of N, such that $\forall\, 0 < \epsilon \leq \epsilon_0$

$$\sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2}\left\|\underline{d\Theta^0_{\underline{\underline{\Xi}}}}(\xi,\eta)^*\right\|_{\mathcal{L}(\mathcal{C}^{s+2,\sigma},\,\mathcal{P}^{s+1,\sigma})}$$

$$\leq \sum_{r\geq 2}\sup_{\|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \leq \epsilon/N^2}\left\|\underline{d\Theta^r_{\underline{\underline{\Xi}}}}(\xi,\eta)^*\right\|_{\mathcal{L}(\mathcal{C}^{s+2,\sigma},\,\mathcal{P}^{s+1,\sigma})}$$

$$\leq \sum_{r\geq 2}C_4^r N^{r-2}\frac{\epsilon^{r-1}}{N^{2(r-1)}} \leq \frac{C_5\epsilon}{N^2}.$$

Appendix D. Proof of Lemma 3.6 and Corollary 3.7

Proof of Lemma 3.6. Since the map $(b,a) \mapsto L_p(b,a)$ is linear, it is enough to prove that it is continuous from $\mathcal{C}^{s,\sigma}$ to $\mathcal{L}(\mathbb{C}^{2N})$. In particular we will prove that

$$\|L_p\|_{\mathcal{L}(\mathbb{C}^{2N})} \leq \sup_{0\leq j\leq N-1}\left(|b_j| + 2\sup_j|a_j|\right). \tag{D.1}$$

This estimate, together with Lemma B.2, proves (3.20). In order to prove (D.1), write $L_p = D + A^+ + A^-$, where D is the diagonal part of L_p and $A^{\pm}$ are defined by

$$A^+ = \begin{pmatrix} 0 & a_0 & & \\ & 0 & \ddots & \\ & & 0 & a_{N-1} \\ a_{N-1} & & & 0 \end{pmatrix}, \quad A^- = \begin{pmatrix} 0 & & & a_{N-1} \\ a_0 & 0 & & \\ & \ddots & 0 & \\ & & a_{N-2} & 0 \end{pmatrix}.$$

To estimate the norms of D, A^+ and A^- is enough to observe that for every $x \in \mathbb{C}^{2N}$ one has

$$\|Dx\|_{\mathbb{C}^{2N}}^2 := \sum_{j=0}^{2N-1}|b_j x_j|^2 \leq \left(\sup_{0\leq j\leq N-1}|b_j|\right)^2\|x\|_{\mathbb{C}^{2N}}^2 ,$$

$$\left\| A^{\pm} x \right\|_{\mathbb{C}^{2N}}^2 \leq \left(\sup_{0 \leq j \leq N-1} |a_j| \right)^2 \|x\|_{\mathbb{C}^{2N}}^2 ,$$

where $\|\cdot\|_{\mathbb{C}^{2N}}$ is the standard Euclidean norm on $\mathbb{C}^{2N}$. Thus (D.1) follows. □

Proof of Corollary 3.7. Item (i) follows by standard perturbation theory, and the details are omitted. We prove now item (ii). Let Γ_j be the circle defined by $\Gamma_j :=$ $\left\{ \lambda \in \mathbb{C} : \left| \lambda_{2j}^0 - \lambda \right| = \frac{1}{2N^2} \right\}$, counter-clockwise oriented. By item (i), for any $\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \frac{\epsilon_*}{N}$, $\lambda_{2j}(b,a)$ and $\lambda_{2j-1}(b,a)$ are inside the ball enclosed by Γ_j. Write $L_{b,a} - \lambda = L_0 - \lambda + L_p = (L_0 - \lambda) \left(1 + (L_0 - \lambda)^{-1} L_p \right)$; its inverse

$$(L_{b,a} - \lambda)^{-1} = \left(\sum_{n=0}^{\infty} \left(-(L_0 - \lambda)^{-1} L_p \right)^n \right) (L_0 - \lambda)^{-1} \tag{D.2}$$

is well defined as a Neumann operator when $\left\| (L_0 - \lambda)^{-1} L_p \right\|_{\mathcal{L}(\mathbb{C}^{2N})} < 1$. Since $L_0 - \lambda$ is diagonalizable with $\{(\lambda_j^0 - \lambda)\}_{0 \leq j \leq 2N-1}$ as eigenvalues, the norm of its inverse is bounded by the inverse of the smallest eigenvalue:

$$\sup_{\lambda \in \Gamma_j} \left\| (L_0 - \lambda)^{-1} \right\|_{\mathcal{L}(\mathbb{C}^{2N})} \leq \sup_{\substack{\lambda \in \Gamma_j \\ 0 \leq k \leq 2N-1}} \left| \frac{1}{\lambda_k^0 - \lambda} \right| < 2N^2 \tag{D.3}$$

where the last estimate is due to the form of Γ_j. Therefore for $0 < \epsilon \leq \epsilon_*$ and $\|(b,a)\|_{\mathcal{C}^{s,\sigma}} < \frac{\epsilon}{N^2}$ one gets, using (3.20),

$$\left\| (L_0 - \lambda)^{-1} L_p \right\|_{\mathcal{L}(\mathbb{C}^{2N})} \leq \|L_p\|_{\mathcal{L}(\mathbb{C}^{2N})} \left\| (L_0 - \lambda)^{-1} \right\|_{\mathcal{L}(\mathbb{C}^{2N})}$$

$$\leq C_{s,\sigma} \|(b,a)\|_{\mathcal{C}^{s,\sigma}} 2N^2 < 2 C_{s,\sigma} \epsilon_* ,$$

which proves the convergence of the Neumann series (D.2) for $\epsilon_* \leq \frac{1}{2 C_{s,\sigma}}$.

Substituting (D.2) in (3.18) we get, for $1 \leq j \leq N-1$,

$$P_j(b,a) = P_{j0} - \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma_j} \left(\sum_{n=1}^{\infty} \left(-(L_0 - \lambda)^{-1} L_p \right)^n \right) (L_0 - \lambda)^{-1} \, \mathrm{d}\lambda. \tag{D.4}$$

Since the series inside the integral is absolutely and uniformly convergent for $(b,a) \in B^{\mathcal{C}^{s,\sigma}} \left(\frac{\epsilon}{N^2} \right)$, $(b,a) \mapsto P_j(b,a)$ is analytic as a map from $B^{\mathcal{C}^{s,\sigma}} \left(\frac{\epsilon}{N^2} \right)$ to $\mathcal{L}(\mathbb{C}^{2N})$. Estimate (3.22) follows easily from (D.4).

We prove now item (iii). Properties $(U1)$–$(U3)$ are standard [29]. The analyticity of the map $(b,a) \mapsto U_j(b,a)$ follows from item (ii). Indeed, in order for $U_j(b,a)$ to be defined as a Neumann series one needs $\|P_j(b,a) - P_{j0}\|_{\mathcal{L}(\mathbb{C}^{2N})} < 1$, which follows from (3.22). Estimate (3.23) follows by expanding (3.19) in power series of $P_j(b,a) - P_{j0}$. □

Appendix E. Proof of Proposition 3.8

Denote by $D : \mathbb{C}^{N-1} \to \mathbb{C}^{N-1}$ the diagonal operator

$$D : \{\xi_j\}_{1 \le j \le N-1} \mapsto \{D_j \xi_j\}_{1 \le j \le N-1}, \quad \text{where} \quad D_j := \left(\tfrac{2}{N}\omega\left(\tfrac{j}{N}\right)\right)^{-1/2} . \tag{E.1}$$

Proof of properties $(Z1)$–$(Z3)$. Property $(Z1)$ follows from formula (3.25), since[4]:

$$
\begin{aligned}
\overline{z_j(b,a)} &= D_j \overline{\left\langle \left(L_{b,a} - \lambda_{2j}^0\right) U_j f_{2j,0}, \overline{U_j f_{2j,0}} \right\rangle} = D_j \left\langle \overline{U_j f_{2j,0}}, \left(L_{b,a} - \lambda_{2j}^0\right) U_j f_{2j,0} \right\rangle \\
&= D_j \left\langle U_j f_{2j-1,0}, \left(L_{b,a} - \lambda_{2j}^0\right) \overline{U_j f_{2j-1,0}} \right\rangle = D_j \left\langle \left(L_{b,a} - \lambda_{2j}^0\right) f_{2j-1}, \overline{f_{2j-1}} \right\rangle \\
&= w_j(b,a).
\end{aligned}
$$

We prove now $(Z2)$. Using Lemma 3.7(iv) and the fact that $\overline{f_{2j,0}} = f_{2j-1,0}$, decompose $f_{2j,0}$ and f_{2j} in real and imaginary parts:

$$
\begin{aligned}
f_{2j,0} &= e_{j,0} + i h_{j,0}, \quad f_{2j} = e_j + i h_j \\
f_{2j-1,0} &= e_{j,0} - i h_{j,0}, \quad f_{2j-1} = e_j - i h_j,
\end{aligned}
$$

where

$$
\begin{aligned}
e_{j,0} &:= \operatorname{Re} f_{2j,0}, \quad h_{j,0} := \operatorname{Im} f_{2j,0}, \quad\quad \text{and} \\
e_j &:= \operatorname{Re} f_{2j} = U_j e_{j,0}, \quad h_j := \operatorname{Im} f_{2j} = U_j h_{j,0}.
\end{aligned}
$$

The vectors $\{e_j, h_j\}$ form a real orthogonal basis for $E_j(b,a)$. Let $M_j(b,a)$ be the matrix of the selfadjoint operator $L_{b,a} - \lambda_{2j}^0\big|_{E_j(b,a)}$ with respect to this basis:

$$M_j(b,a) = \begin{pmatrix} \alpha_j & \sigma_j \\ \sigma_j & \beta_j \end{pmatrix}.$$

The eigenvalues of M_j are obviously $\lambda_{2j} - \lambda_{2j}^0$ and $\lambda_{2j-1} - \lambda_{2j}^0$, hence

$$
\begin{aligned}
\operatorname{Tr} M_j &= \alpha_j + \beta_j = \left(\lambda_{2j} - \lambda_{2j}^0\right) + \left(\lambda_{2j-1} - \lambda_{2j}^0\right), \\
\operatorname{Det} M_j &= \alpha_j \beta_j - \sigma_j^2 = \left(\lambda_{2j} - \lambda_{2j}^0\right)\left(\lambda_{2j-1} - \lambda_{2j}^0\right).
\end{aligned}
$$

Now observe that

$$
\begin{aligned}
z_j(b,a) &= D_j \left\langle \left(L_{b,a} - \lambda_{2j}^0\right)(e_j + i h_j), (e_j - i h_j) \right\rangle \\
&= D_j \left\langle \left(L_{b,a} - \lambda_{2j}^0\right) e_j, e_j \right\rangle - D_j \left\langle \left(L_{b,a} - \lambda_{2j}^0\right) h_j, h_j \right\rangle
\end{aligned}
$$

---

[4] To simplify the notation, we write $f_j \equiv f_j(b,a)$ and $U_j \equiv U_j(b,a)$.

$$+ 2\mathrm{i}D_j \left\langle \left(L_{b,a} - \lambda_{2j}^0 \right) e_j, h_j \right\rangle$$
$$= \left(\tfrac{2}{N} \omega \left(\tfrac{j}{N} \right) \right)^{-1/2} (\alpha_j - \beta_j + \mathrm{i}2\sigma_j).$$

Finally one computes

$$(\lambda_{2j} - \lambda_{2j-1})^2 = (\mathrm{Tr}\ M_j)^2 - 4\mathrm{Det}\ M_j = (\alpha_j + \beta_j)^2 - 4\alpha_j\beta_j + 4\sigma_j^2$$
$$= (\alpha_j - \beta_j)^2 + 4\sigma_j^2 = (\mathrm{Re}\ z_j)^2 + (\mathrm{Im}\ z_j)^2 = \left(\tfrac{2}{N} \omega \left(\tfrac{j}{N} \right) \right) |z_j(b,a)|^2.$$

We prove now $(Z3)$. The first order terms of z_j and w_j in $(b,\ a)$ are given by

$$dz_j(0,0)(b,a) = D_j \left\langle L_p f_{2j,0}, \overline{f_{2j,0}} \right\rangle, \quad dw_j(0,0)(b,a) = D_j \left\langle L_p f_{2j-1,0}, \overline{f_{2j-1,0}} \right\rangle,$$
$$1 \le j \le N - 1 \ .$$

Using the explicit formula for $f_{2j,0}$ in Lemma 3.4, one computes

$$\left\langle L_p f_{2j,0}, \overline{f_{2j,0}} \right\rangle = \frac{1}{2N} \sum_{l=0}^{2N-1} b_l e^{\mathrm{i}2\rho_j l} + a_{l-1} e^{\mathrm{i}2\rho_j (l-1)} e^{\mathrm{i}\rho_j} + a_l e^{\mathrm{i}2\rho_j l} e^{\mathrm{i}\rho_j}$$
$$= \frac{1}{2N} \sum_{l=0}^{2N-1} b_l e^{\mathrm{i}2\pi j l/N} + a_{l-1} e^{\mathrm{i}2\pi (l-1)j/N} e^{\mathrm{i}\rho_j} + a_l e^{\mathrm{i}2\pi l j/N} e^{\mathrm{i}\rho_j}$$
$$= \frac{1}{\sqrt{N}} \left(\hat{b}_j + 2e^{\mathrm{i}\rho_j} \hat{a}_j \right) = \frac{1}{\sqrt{N}} \left(\hat{b}_j - 2e^{\mathrm{i}\pi j/N} \hat{a}_j \right). \tag{E.2}$$

The formula for $dz_j(0,0)(b,a)$ immediately follows. The one for $dw_j(0,0)(b,a)$ is proved in the same way and the details are omitted.

The estimate (3.28) for $\underline{dZ}(0,0)$ follows immediately. We estimate now the norm of $\underline{dZ}(0,0)^*$. One checks that $(B,A) = dZ(0,0)^*(\xi,\eta)$ iff $\hat{B}_0 = \hat{A}_0 = 0$ and for $1 \le k \le N-1$

$$(\hat{B}_k, \hat{A}_k) = \left(\frac{1}{\sqrt{2\omega \left(\tfrac{k}{N} \right)}} (\xi_k + \eta_{N-k}),\ \frac{2}{\sqrt{2\omega \left(\tfrac{k}{N} \right)}} (e^{\mathrm{i}\pi k/N} \xi_k + e^{\mathrm{i}\pi (N-k)/N} \eta_{N-k}) \right)\ .$$

Thus there exist constants $C, C', C_Z > 0$, independent of N, such that

$$\|\underline{dZ}(0,0)^*(\xi,\eta)\|_{C^{s+2,\sigma}}^2 \le \frac{C'}{N} \sum_{k=1}^{N-1} [k]_N^{2s}\, e^{2\sigma[k]_N} \omega \left(\tfrac{k}{N} \right) \frac{[k]_N^4}{\omega \left(\tfrac{k}{N} \right)^2} \left(|\xi_k|^2 + |\eta_k|^2 \right)$$
$$\le C_Z^2 N^4 \|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}}^2$$

where in the last inequality we used that $[k]_N^4 / \omega \left(\tfrac{k}{N} \right)^2 \le C'' N^4$ for some constant $C'' > 0$ independent of N. Thus the second of (3.28) is proved.

Proof of property $(Z4)$. We will prove that Z is normally analytic. Recall that, as mentioned in the discussion before Proposition 3.3, the map Z is said to be normally analytic if $\check{Z} := Z\mathcal{F}$ is normally analytic. With an abuse of notations, we omit the "check" from Z.

We begin by expanding the components of Z, denoted by $Z_j(b,a) := (z_j(b,a), w_j(b,a))$, in Taylor series with center at $(b,a) = (0,0)$. The first two terms of the expansions are given by

$$z_j(b,a) = D_j \langle L_p f_{2j,0}, \overline{f_{2j,0}} \rangle + D_j \langle L_p \left(L_0 - \lambda_{2j}^0 \right)^{-1} \left(\mathbb{1} - P_{j0} \right) L_p f_{2j,0}, \overline{f_{2j,0}} \rangle + O((b,a)^3),$$

$$w_j(b,a) = D_j \langle L_p f_{2j-1,0}, \overline{f_{2j-1,0}} \rangle + D_j \langle L_p \left(L_0 - \lambda_{2j}^0 \right)^{-1} \left(\mathbb{1} - P_{j0} \right) L_p f_{2j-1,0}, \overline{f_{2j-1,0}} \rangle$$
$$+ O((b,a)^3). \tag{E.3}$$

To perform the Taylor expansion at every order it is convenient to proceed in the following way. Write $z_j(b,a) = z_{j,1}(b,a) + z_{j,2}(b,a)$ and $w_j(b,a) = w_{j,1}(b,a) + w_{j,2}(b,a)$ where

$$z_{j,1}(b,a) = D_j \left\langle \left(L_0 - \lambda_{2j}^0 \right) f_{2j}(b,a), \overline{f_{2j}(b,a)} \right\rangle,$$

$$z_{j,2}(b,a) = D_j \left\langle L_p f_{2j}(b,a), \overline{f_{2j}(b,a)} \right\rangle, \tag{E.4}$$

while $w_{j,1}(b,a)$ and $w_{j,2}(b,a)$ are defined as in (E.4), but with $f_{2j-1}(b,a)$ replacing $f_{2j}(b,a)$.

Expand $z_{j,\varsigma}(b,a)$, $\varsigma = 1,2$, in Taylor series with center at $(b,a) = (0,0)$: $z_{j,\varsigma}(b,a) = \sum_{n \geq 1} z_{j,\varsigma}^n(b,a)$, with $z_{j,\varsigma}^n$ a homogeneous polynomial of degree n in b, a. We write an analogous expansion for $w_{j,\varsigma}(b,a)$. Therefore one has

$$Z_j^n(b,a) := (z_j^n(b,a), w_j^n(b,a)) \equiv \left(z_{j,1}^n(b,a) + z_{j,2}^n(b,a), w_{j,1}^n(b,a) + w_{j,2}^n(b,a) \right).$$

In order to write explicitly $z_{j,\varsigma}^n(b,a)$ as a function of b and a, one needs to expand the vectors $f_{2j}(b,a)$ and $f_{2j-1}(b,a)$ in Taylor series of b, a. Rewrite (3.19), (3.24) as

$$f_{2j}(b,a) = U_j(b,a) f_{2j,0} = \left(\mathbb{1} - (P_j(b,a) - P_{j0})^2 \right)^{-1/2} \left(\mathbb{1} + (P_j(b,a) - P_{j0}) \right) f_{2j,0}$$

and expand the r.h.s. above in power series of $P_j(b,a) - P_{j0}$, getting:

$$f_{2j}(b,a) = \sum_{m=0}^{\infty} c_m \left(P_j(b,a) - P_{j0} \right)^m f_{2j,0},$$

$$f_{2j-1}(b,a) = \sum_{m=0}^{\infty} c_m \left(P_j(b,a) - P_{j0} \right)^m f_{2j-1,0}, \tag{E.5}$$

where the c_m's are the coefficients of the Taylor series of the function $\phi(x) = \frac{1+x}{(1-x^2)^{1/2}}$. Note that $c_{2k+1} = c_{2k} \equiv (-1)^k \binom{-1/2}{k}$, where $\binom{-1/2}{k} := -\frac{1}{2} \left(-\frac{1}{2} - 1 \right) \cdots \left(-\frac{1}{2} - k + 1 \right)$ is

the product of k negative terms, thus $(-1)^k \binom{-1/2}{k} \geq 0$, $\forall k \geq 0$, and therefore $c_m \geq 0$, $\forall m$.

By Corollary 3.7 (see also formula (D.4)) one has, in the ball $B^{\mathcal{C}^{s,\sigma}}(\epsilon_*/N^2)$,

$$P_j(b,a) - P_{j0} = \frac{i}{2\pi} \sum_{n=1}^{\infty} (-1)^n \oint_{\Gamma_j} T^n(b,a,\lambda) (L_0 - \lambda)^{-1} \, d\lambda \qquad \text{(E.6)}$$

where the Γ_j's are defined as in equation (3.18), and

$$T(b,a,\lambda) := (L_0 - \lambda)^{-1} L_p .$$

Substituting (E.6) in (E.5) we get that

$$f_{2j}(b,a) = f_{2j,0} + \sum_{n \geq 1} \sum_{1 \leq m \leq n} c_m \sum_{\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^m, \, |\alpha| = n} f_{2j,m}^{\alpha}(b,a),$$

$$f_{2j,m}^{\alpha}(b,a)$$

$$:= \left(\frac{i}{2\pi}\right)^m (-1)^{|\alpha|} \oint_{\Gamma_j} \cdots \oint_{\Gamma_j} T^{\alpha_1}(b,a,\lambda_1) (L_0 - \lambda_1)^{-1} \cdots$$

$$\times T^{\alpha_m}(b,a,\lambda_m) (L_0 - \lambda_m)^{-1} f_{2j,0} d\lambda_1 \ldots d\lambda_m. \qquad \text{(E.7)}$$

An analogous expansion holds for $f_{2j-1}(b,a)$, with $f_{2j-1,0}$ substituting $f_{2j,0}$ in the integral formula above. In order to write explicitly the expression inside the integral, one needs to compute the iterated terms $T^n(b,a,\lambda)f_{2j,0}$ and $T^n(b,a,\lambda)f_{2j-1,0}$. The computation turns out to be simpler if we express $L_p f_{2j,0}$ in the basis of the eigenvectors of L_0. To simplify the notations we relabel the eigenvectors of L_0 in the following way:

$$g_0 := f_{00}, \quad g_N := f_{2N-1,0}, \quad g_j := f_{2j,0}, \quad g_{-j} := f_{2j-1,0}, \quad \text{for } 1 \leq j \leq N-1$$

and the eigenvalues of L_0 as

$$\hat{\lambda}_0 := \lambda_0^0, \quad \hat{\lambda}_N := \lambda_{2N-1}^0, \quad \hat{\lambda}_j := \lambda_{2j}^0, \quad \hat{\lambda}_{-j} := \lambda_{2j-1}^0, \quad \text{for } 1 \leq j \leq N-1.$$

For every $1 \leq j \leq N-1$ one has that $\overline{g_j} = g_{-j}$, formally, one can also write $g_{j+2N} = g_j$, $\hat{\lambda}_j = \hat{\lambda}_{-j}$ and $\hat{\lambda}_{j+2N} = \hat{\lambda}_j$, as one verifies using the explicit expressions of the g_j's and $\hat{\lambda}_j$'s. In this notation, for $\lambda \neq \hat{\lambda}_{\pm j}$, one has $(L_0 - \lambda)^{-1} g_{\pm j} = g_{\pm j}/(\hat{\lambda}_{\pm j} - \lambda)$. With a computation analogous to the one in (E.2) (using also the second formula in (B.2)), one verifies that the projection of $L_p g_j$ on the vector g_k is given by

$$\langle L_p\, g_j, g_k \rangle = \frac{1}{\sqrt{N}} \left(\hat{b}_{\frac{j-k}{2}} - 2\cos\left(\frac{k\pi}{N}\right) \hat{a}_{\frac{j-k}{2}} \right) \delta_{(j-k;\text{even})}, \tag{E.8}$$

where $\delta_{(j-k;\text{even})} = 1$ if $j - k$ is an even integer, and equals 0 otherwise. Formula (E.8) implies that $L_p g_j$ is supported only on the vectors g_k whose index k satisfies $k = j - 2l$ for some integer l. Therefore we can write

$$T(b,a,\lambda) g_j = \sum_{l \in K_N^0} \frac{x_j^l}{\hat{\lambda}_{j-2l} - \lambda} g_{j-2l},$$

$$x_j^l := \langle L_p g_j, g_{j-2l} \rangle = \frac{1}{\sqrt{N}} \left(\hat{b}_l - 2\cos\left(\frac{(j-2l)\pi}{N}\right) \hat{a}_l \right), \tag{E.9}$$

where K_N^0 is the set of indexes defined in (B.5). Note that $|x_j^l| \leq \frac{2}{\sqrt{N}} \left(|\hat{b}_l| + |\hat{a}_l| \right)$ uniformly in j, and $x_j^{l+N} = x_j^l$. Iterating (E.9) one gets

$$T^n(b,a,\lambda)\,(L_0 - \lambda)^{-1}\, g_j = \sum_{i_1,\cdots,i_n \in K_N^0} \frac{x_j^{i_1}\, x_{j-2i_1}^{i_2} \cdots x_{j-2i_1-\cdots-2i_{n-1}}^{i_n}}{(\hat{\lambda}_j - \lambda) \prod_{l=1}^n \left(\hat{\lambda}_{j-2\sum_{m=1}^l i_m} - \lambda \right)} g_{j-2i_1-\cdots-2i_n}.$$

More generally, for a vector $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^m$ with $|\alpha| = n$ and $\lambda_1, \cdots, \lambda_m \in \Gamma_j$, one has

$$T^{\alpha_m}(b,a,\lambda_m)\,(L_0 - \lambda_m)^{-1} \cdots T^{\alpha_1}(b,a,\lambda_1)\,(L_0 - \lambda_1)^{-1}\, g_j$$

$$= \sum_{i_1,\ldots,i_n \in K_N^0} \frac{x_j^{i_1}\, x_{j-2i_1}^{i_2} \cdots x_{j-2i_1-\cdots-2i_{n-1}}^{i_n}}{(\hat{\lambda}_j - \lambda_1) \prod_{l=1}^n \left(\hat{\lambda}_{j-2\sum_{m=1}^l i_m} - \mu_l \right) \prod_{l=1}^{m-1} \left(\hat{\lambda}_{j-2\sum_{h=1}^{\alpha_1+\cdots+\alpha_l} i_h} - \lambda_{l+1} \right)}$$

$$\times\, g_{j-2i_1-\cdots-2i_n} \tag{E.10}$$

where

$$\mu_l = \lambda_1 \ \text{for} \ 1 \leq l \leq \alpha_1, \quad \text{and} \quad \mu_l = \lambda_k \ \text{for} \ \sum_{h=1}^{k-1}\alpha_h + 1 \leq l \leq \sum_{h=1}^{k}\alpha_h, \quad 2 \leq k \leq m \ . \tag{E.11}$$

To obtain the explicit expression of $z_{j,\varsigma}^n$ and $w_{j,\varsigma}^n$, $\varsigma = 1, 2$, in terms of the Fourier variables $\hat{b}$, $\hat{a}$, we substitute (E.10) in (E.7) and the obtained result in (E.4). By (E.7), $z_{j,1}^n$ is a sum of terms of the form $\left\langle (L_0 - \lambda_{2j}^0)\, f_{2j,p_1}^\alpha, \overline{f_{2j,p_2}^\beta} \right\rangle$ over $(p, \alpha, \beta) \in \mathbb{N}^2 \times \mathbb{N}^{p_1} \times \mathbb{N}^{p_2}$ with $|p| = p_1 + p_2 \leq n$ and $|\alpha| + |\beta| = n$. For $|\alpha| = r$, $|\beta| = n - r$ one gets

$$\left\langle \left(L_0 - \hat\lambda_j \right) f^\alpha_{2j,p_1}, \overline{f^\beta_{2j,p_2}} \right\rangle$$

$$= \left(\frac{\mathrm{i}}{2\pi} \right)^{|p|} (-1)^n \oint_{\Gamma_j} \cdots \oint_{\Gamma_j} \kappa^{p,\alpha,\beta}_{j,1}(i) x^{i_1}_j x^{i_2}_{j-2i_1} \cdots x^{i_r}_{j-2i_1-\cdots-2i_{r-1}}$$

$$\times x^{i_n}_j x^{i_{n-1}}_{j-2i_n} \cdots x^{i_{r+1}}_{j-2i_n-\cdots-2i_{r+2}} \left\langle g_{j-2i_1-\cdots-2i_r}, \overline{g_{j-2i_{r+1}-\cdots-2i_n}} \right\rangle \, \mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_{|p|},$$

$$\text{(E.12)}$$

where, writing $\mathbf{i} = (i_1, \cdots, i_n)$,

$$\kappa^{p,\alpha,\beta}_{j,1}(\mathbf{i})$$

$$:= \frac{\left(\hat\lambda_{j-2\sum_{m=1}^r i_m} - \hat\lambda_j \right)}{(\hat\lambda_j - \lambda_1) \prod_{l=1}^r \left(\hat\lambda_{j-2\sum_{m=1}^l i_m} - \mu_l \right) \prod_{l=1}^{p_1-1} \left(\hat\lambda_{j-2\sum_{h=1}^{\alpha_1+\cdots+\alpha_l} i_h} - \lambda_{l+1} \right)}$$

$$\times \frac{1}{(\hat\lambda_j - \lambda_{p_1+1}) \prod_{l=r+1}^n \left(\hat\lambda_{j-2\sum_{m=l}^n i_m} - \tilde\mu_l \right) \prod_{l=1}^{p_2-1} \left(\hat\lambda_{j-2\sum_{h=1}^{\beta_1+\cdots+\beta_l} i_h} - \lambda_{l+1} \right)}$$

$$\text{(E.13)}$$

and the $\tilde\mu_l$'s are defined as in (E.11), but with the multi-index β replacing α. Similarly, the term $z^n_{j,2}$ is a sum of terms of the form $\left\langle L_p f^\alpha_{2j,p_1}, \overline{f^\beta_{2j,p_2}} \right\rangle$ over $(p,\alpha,\beta) \in \mathbb{N}^2 \times \mathbb{N}^{p_1} \times \mathbb{N}^{p_2}$ with $|p| \le n$ and $|\alpha| + |\beta| = n-1$. The term $\left\langle L_p f^\alpha_{2j,p_1}, \overline{f^\beta_{2j,p_2}} \right\rangle$ has an expression similar to (E.12), and for $|\alpha| = r$ and $|\beta| = n-1-r$ the kernel $\kappa^{p,\alpha,\beta}_{j,2}(\mathbf{i})$ is given by

$$\kappa^{p,\alpha,\beta}_{j,2}(\mathbf{i})$$

$$:= \frac{1}{(\hat\lambda_j - \lambda_1) \prod_{l=1}^r \left(\hat\lambda_{j-2\sum_{m=1}^l i_m} - \mu_l \right) \prod_{l=1}^{p_1-1} \left(\hat\lambda_{j-2\sum_{h=1}^{\alpha_1+\cdots+\alpha_l} i_h} - \lambda_{l+1} \right)}$$

$$\times \frac{1}{(\hat\lambda_j - \lambda_{p_1+1}) \prod_{l=r+2}^n \left(\hat\lambda_{j-2\sum_{m=l}^n i_m} - \tilde\mu_l \right) \prod_{l=1}^{p_2-1} \left(\hat\lambda_{j-2\sum_{h=1}^{\beta_1+\cdots+\beta_l} i_h} - \lambda_{l+1} \right)}.$$

$$\text{(E.14)}$$

Using the explicit form of the eigenvectors $\{g_k\}_{-(N-1)\le k \le N}$ (see Lemma 3.4), one verifies that

$$\left\langle g_{j-2i_1-\cdots-2i_r}, \overline{g_{j-2i_{r+1}-\cdots-2i_n}} \right\rangle = \delta\left(j, \sum_{m=1}^n i_m \right),$$

$$\left\langle g_{N-j-2i_1-\cdots-2i_r}, \overline{g_{N-j-2i_{r+1}-\cdots-2i_n}} \right\rangle = \delta\left(-j, \sum_{m=1}^n i_m \right).$$

This is used to simplify the last term in (E.12). Moreover, using $j = \sum_{m=1}^{n} i_m$ and the identity $\hat{\lambda}_j = \hat{\lambda}_{-j}$, one gets that

$$\hat{\lambda}_{j-2i_n} = \hat{\lambda}_{j-2\sum_{m=1}^{n-1} i_m}, \quad \ldots, \quad \hat{\lambda}_{j-2i_n-2i_{n-1}-\cdots-2i_{r+1}} = \hat{\lambda}_{j-2\sum_{m=1}^{r} i_m}. \quad \text{(E.15)}$$

Recalling the definition of the coefficients x_j^l (formula (E.9)), we can write, for $\varsigma = 1, 2$,

$$z_{j,\varsigma}^n(\hat{b}, \hat{a}) = \frac{1}{N^{n/2}} \left(\tfrac{2}{N} \omega \left(\tfrac{j}{N} \right) \right)^{-1/2} \sum_{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta^n} \mathcal{K}_{j,\varsigma}^n(\mathbf{i}, \boldsymbol{\iota}) \, u_{i_1, \iota_1} \ldots u_{i_n, \iota_n} \quad \text{(E.16)}$$

where the set

$$\Delta^n := \left\{ (\mathbf{i}, \boldsymbol{\iota}) \in \mathbb{Z}^n \times \mathbb{N}^n : i_l \in K_N^0, \quad \iota_l \in \{1, 2\}, \quad \forall 1 \le l \le n \right\},$$

the variables $u = (u_{i_1, \iota_1}, \cdots, u_{i_n, \iota_n})$ are defined by

$$u_{i_r, 1} := \hat{b}_{i_r}, \qquad u_{i_r, 2} := \hat{a}_{i_r},$$

the kernels $\mathcal{K}_{j,\varsigma}^n(\mathbf{i}, \boldsymbol{\iota})$ are defined for $(\mathbf{i}, \boldsymbol{\iota}) \in \Delta^n$ by

$$\mathcal{K}_{j,\varsigma}^n(\mathbf{i}, \boldsymbol{\iota}) := \tilde{\mathcal{K}}_{j,\varsigma}^n(\mathbf{i}) \prod_{\{1 \le l \le n\}} \left(-2 \cos \left(\frac{(j - 2i_i - \cdots - 2i_l)\pi}{N} \right) \right)^{\iota_l - 1}, \quad \text{(E.17)}$$

$$\tilde{\mathcal{K}}_{j,\varsigma}^n(\mathbf{i}) = \sum_{\substack{r+s=n-(\varsigma-1) \\ p=(p_1, p_2) \in \mathbb{N}^2, \, |p| \le n}} c_{p_1} c_{p_2} \sum_{\substack{(\alpha, \beta) \in \mathbb{N}^{p_1} \times \mathbb{N}^{p_2} \\ |\alpha|=r, \, |\beta|=s}} \mathcal{S}_{j,\varsigma}^{p,\alpha,\beta}(\mathbf{i}) \quad \text{(E.18)}$$

and finally

$$\mathcal{S}_{j,\varsigma}^{p,\alpha,\beta}(\mathbf{i}) = \delta \left(j, \sum_{m=1}^{n} i_m \right) \left(\frac{\mathrm{i}}{2\pi} \right)^{|p|} (-1)^n \oint_{\Gamma_j} \ldots \oint_{\Gamma_j} \kappa_{j,\varsigma}^{p,\alpha,\beta}(\mathbf{i}) \, \mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_{|p|}. \quad \text{(E.19)}$$

An analogous expansion holds also for $w_{j,1}^n$ and $w_{j,2}^n$.

We need now to get estimates of the kernels $\mathcal{K}_{j,\varsigma}^n$, which will follow from estimates on the denominators of $\kappa_{j,\varsigma}^{p,\alpha,\beta}$.

Lemma E.1. *Let* $\mu \in \Gamma_j := \left\{ \lambda \in \mathbb{C} : |\lambda - \lambda_{2j}^0| = \min \left(\frac{\langle j \rangle}{2N^2}, \frac{\langle N-j \rangle}{2N^2} \right) \right\}$, *where* $\langle j \rangle = \left(1 + |j|^2 \right)^{1/2}$. *Then there exists a constant* $R > 0$, *independent of* N, *such that for every* $-(N-1) \le k \le N$ *one has*

$$\left| \hat{\lambda}_k - \mu \right| \ge \begin{cases} R \langle j-k \rangle \langle j+k \rangle / N^2, & \text{if } 0 \le |j| \le \lfloor N/2 \rfloor \\ R \langle j-k \rangle \langle (N-j) + (N-k) \rangle / N^2, & \text{if } \lfloor N/2 \rfloor + 1 \le |j| \le N \end{cases} \quad \text{(E.20)}$$

Proof. Consider first the situation in which both the eigenvalues $\hat{\lambda}_j$ and $\hat{\lambda}_k$ are in the low half of the spectrum, namely $0 \leq |j|, |k| \leq \lfloor N/2 \rfloor$. In this case one has

$$|\hat{\lambda}_k - \hat{\lambda}_j| \equiv |\lambda^0_{2|k|} - \lambda^0_{2|j|}| = 2 \left| \cos\left(\tfrac{|k|\pi}{N}\right) - \cos\left(\tfrac{|j|\pi}{N}\right) \right|$$

$$= 2 \left| \cos\left(\tfrac{k\pi}{N}\right) - \cos\left(\tfrac{j\pi}{N}\right) \right| \geq \frac{4|j^2 - k^2|}{N^2}.$$

Therefore, for $k \neq j$, there exists a positive constant R_1 such that for $\forall \mu \in \Gamma_j$

$$\left| \hat{\lambda}_k - \mu \right| \geq \left| \hat{\lambda}_k - \hat{\lambda}_j \right| - \frac{\langle j \rangle}{2N^2} \geq \frac{4|j^2 - k^2|}{N^2} - \frac{\langle j \rangle}{2N^2} \geq R_1 \frac{\langle j - k \rangle \langle j + k \rangle}{N^2}, \quad \text{(E.21)}$$

where we used the inequality $\langle j \rangle \leq 2 \langle j - k \rangle \langle j + k \rangle$, which holds since j, k are integers. If $k = j$, then the claimed estimate follows trivially since $|\hat{\lambda}_k - \mu| = \langle j \rangle / 2N^2$.

Consider now the case when $\hat{\lambda}_j$ is in the low half of the spectrum, while $\hat{\lambda}_k$ is in the high half, i.e. $0 \leq |j| \leq \lfloor N/2 \rfloor$, while $\lfloor N/2 \rfloor < |k| \leq N$. In this case the distance of the eigenvalues $\hat{\lambda}_j$ and $\hat{\lambda}_k$ is of order $\frac{1}{N}$, therefore the estimate (E.20) holds as well. More precisely, using $\cos x \geq 1 - \tfrac{2}{\pi} x$ for $0 \leq x \leq \pi/2$, one has

$$|\hat{\lambda}_k - \hat{\lambda}_j| = |\lambda^0_{2|k|} - \lambda^0_{2|j|}| = 2 \left| \cos\left(\tfrac{(N-|k|)\pi}{N}\right) + \cos\left(\tfrac{j\pi}{N}\right) \right|$$

$$\geq \frac{4(|k| - |j|)}{N} \geq \frac{\langle j - k \rangle \langle j + k \rangle}{N^2},$$

where the last inequality holds since $\langle l \rangle / N \leq 4$, $\forall |l| \leq 2N$. The inequality above implies that

$$\left| \hat{\lambda}_k - \mu \right| \geq |\hat{\lambda}_k - \hat{\lambda}_j| - \frac{\langle j \rangle}{2N^2} \geq \frac{\langle j - k \rangle \langle j + k \rangle}{N^2} - \frac{\langle j \rangle}{2N^2}$$

$$\geq R_2 \frac{\langle j - k \rangle \langle j + k \rangle}{N^2}, \quad \text{(E.22)}$$

for some $R_2 > 0$. Thus the first of (E.20) is proved.

The proof of the second inequality of (E.20) follows by symmetry and is omitted. $\quad\square$

We can now estimate the kernels $\mathcal{K}^n_{j,\varsigma}$ defined in (E.17).

Lemma E.2. *There exists a constant $R > 0$, independent of N, such that $\mathcal{K}^n_{j,\varsigma}(\mathbf{i}, \boldsymbol{\iota})$, $\varsigma = 1, 2$, satisfy, for every $n \geq 2$ and $1 \leq j \leq \lfloor N/2 \rfloor$, the estimates*

$$\left| \mathcal{K}^n_{j,\varsigma}(\mathbf{i}, \boldsymbol{\iota}) \right| \leq R^n N^{2(n-1)} \delta\left(j, \sum_{l=1}^{n} i_l \right) \frac{1}{\prod_{l=1}^{n-1} \left\langle \sum_{k=1}^{l} i_k \right\rangle \left\langle \sum_{k=1}^{l} i_k - j \right\rangle},$$

$$\left| \mathcal{K}^n_{N-j,\varsigma}(\mathbf{i}, \boldsymbol{\iota}) \right| \leq R^n N^{2(n-1)} \delta\left(-j, \sum_{l=1}^{n} i_l \right) \frac{1}{\prod_{l=1}^{n-1} \left\langle \sum_{k=1}^{l} i_k \right\rangle \left\langle \sum_{k=1}^{l} i_k - j \right\rangle}. \quad \text{(E.23)}$$

Proof. We start by estimating $\kappa_{j,\varsigma}^{p,\alpha,\beta}(\mathbf{i})$, defined in (E.13) and (E.14). For every $-(N-1) \le k \le N$ and $\mu \in \Gamma_j$ one has $\left|\hat{\lambda}_k - \mu\right| \ge \left|\hat{\lambda}_j - \mu\right| \ge \min\left(\frac{\langle j \rangle}{2N^2}, \frac{\langle N-j \rangle}{2N^2}\right)$, therefore

$$\left|(\hat{\lambda}_j - \lambda_1)\prod_{l=1}^{p_1-1}\left(\hat{\lambda}_{j-2\sum_{h=1}^{\alpha_1+\cdots+\alpha_l} i_h} - \lambda_{l+1}\right)(\hat{\lambda}_j - \lambda_{p_1+1})\prod_{l=1}^{p_2-1}\left(\hat{\lambda}_{j-2\sum_{h=1}^{\beta_1+\cdots+\beta_l} i_h} - \lambda_{l+1}\right)\right|$$

$$\ge \left[\min\left(\frac{\langle j \rangle}{2N^2}, \frac{\langle N-j \rangle}{2N^2}\right)\right]^{|p|}.$$

Let now $1 \le j \le \lfloor N/2 \rfloor$. By Lemma E.1, formula (E.15) and the inequality $\frac{|\hat{\lambda}_{j-2\sum_{m=1}^r i_m} - \hat{\lambda}_j|}{|\hat{\lambda}_{j-2\sum_{m=1}^r i_m} - \mu|} \le 2$ (which is used to estimate just $\kappa_{j,1}^{p,\alpha,\beta}(\mathbf{i})$), it follows that, for $\varsigma = 1, 2$,

$$\left|\kappa_{j,\varsigma}^{p,\alpha,\beta}(\mathbf{i})\right| \le \frac{2}{\left[\min\left(\frac{\langle j \rangle}{2N^2}, \frac{\langle N-j \rangle}{2N^2}\right)\right]^{|p|}\prod_{l=1}^{n-1}\left|\hat{\lambda}_{j-2\sum_{m=1}^l i_m} - \mu_l\right|}$$

$$\le \frac{2\,a_j(i_1, \cdots, i_{n-1})}{\left[\min\left(\frac{\langle j \rangle}{2N^2}, \frac{\langle N-j \rangle}{2N^2}\right)\right]^{|p|}}$$

where

$$a_j(i_1, \cdots, i_{n-1}) := \frac{R^{n-1}N^{2(n-1)}}{\prod_{l=1}^{n-1}\left\langle\sum_{k=1}^l i_k\right\rangle\left\langle\sum_{k=1}^l i_k - j\right\rangle}.$$

To estimate $\mathcal{S}_{j,\varsigma}^{p,\alpha,\beta}$ consider (E.19). The $\mathcal{S}_{j,\varsigma}^{p,\alpha,\beta}$'s are defined by integrating the kernels $\kappa_{j,\varsigma}^{p,\alpha,\beta}$ over Γ_j $|p|$-times. Since $|\Gamma_j| = 2\pi\min\left(\frac{\langle j \rangle}{2N^2}, \frac{\langle N-j \rangle}{2N^2}\right)$, one gets

$$\left|\mathcal{S}_{j,\varsigma}^{p,\alpha,\beta}(\mathbf{i})\right| \le \left[\min\left(\frac{\langle j \rangle}{N^2}, \frac{\langle N-j \rangle}{N^2}\right)\right]^{|p|}\delta\left(j, \sum_{l=1}^n i_l\right)\left|\kappa_{j,\varsigma}^{p,\alpha,\beta}(\mathbf{i})\right|$$

$$\le 2\delta\left(j, \sum_{l=1}^n i_l\right)a_j(i_1, \cdots, i_{n-1}).$$

Finally consider $\mathcal{K}_{j,\varsigma}^n$. From (E.17) one has $\left|\mathcal{K}_{j,\varsigma}^n(\mathbf{i},\iota)\right| \le 2^n\left|\tilde{\mathcal{K}}_{j,\varsigma}^n(\mathbf{i})\right|$, and from (E.18)

$$\left|\tilde{\mathcal{K}}_{j,\varsigma}^n(\mathbf{i})\right| \le \delta\left(j, \sum_{l=1}^n i_l\right)a_j(i_1, \cdots, i_{n-1})\sum_{\substack{r+s=n-(\varsigma-1)\\p=(p_1,p_2)\in\mathbb{N}^2,\,|p|\le n}}c_{p_1}c_{p_2}\sum_{\substack{(\alpha,\beta)\in\mathbb{N}^{p_1}\times\mathbb{N}^{p_2}\\|\alpha|=r,\,|\beta|=s}}1$$

$$\le C^n\delta\left(j, \sum_{l=1}^n i_l\right)a_j(i_1, \cdots, i_{n-1}),$$

thus the first estimate of (E.23) follows. The proof of the second one is similar, and is omitted. □

Define now $\mathcal{K}_j^n := \mathcal{K}_{j,1}^n + \mathcal{K}_{j,2}^n$. Then

$$z_j^n(\hat{b}, \hat{a}) = \frac{D_j}{N^{n/2}} \sum_{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta^n} \mathcal{K}_j^n(\mathbf{i}, \boldsymbol{\iota})\, u_{i_1, \iota_1} \dots u_{i_n, \iota_n},$$

$$w_j^n(\hat{b}, \hat{a}) = \frac{D_j}{N^{n/2}} \sum_{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta^n} \mathcal{H}_j^n(\mathbf{i}, \boldsymbol{\iota})\, u_{i_1, \iota_1} \dots u_{i_n, \iota_n}, \tag{E.24}$$

where $\mathcal{H}_j^n(\mathbf{i}, \boldsymbol{\iota}) = \overline{\mathcal{K}_j^n(-\mathbf{i}, \boldsymbol{\iota})}$. The second formula holds since for b, a real one has $w^n(b, a) = \overline{z^n(b, a)}$.

Corollary E.3. *Let $\Delta_j^n := \{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta^n : \sum_{l=1}^n i_l = j\}$. Then for $1 \le j \le \lfloor N/2 \rfloor$ one has* supp $\mathcal{K}_j^n \subseteq \Delta_j^n$ *and* supp $\mathcal{K}_{N-j}^n \subseteq \Delta_{-j}^n$. *Moreover*

$$\left\| \mathcal{K}_j^n \right\|_{\Delta_j^n}, \quad \left\| \mathcal{K}_{N-j}^n \right\|_{\Delta_{-j}^n} \le \frac{R^n N^{2(n-1)}}{\langle j \rangle^{n-1}}, \tag{E.25}$$

where $\left\| \mathcal{K}_j^n \right\|_{\Delta_j^n}^2 := \sup_{\iota_1, \dots, \iota_n \in \{1,2\}} \sum_{i_1 + \dots + i_n = j} \left| \mathcal{K}_j^n(\mathbf{i}, \boldsymbol{\iota}) \right|^2.$

Proof. Just remark that $\frac{\langle j \rangle^2}{\langle k \rangle^2 \langle k-j \rangle^2} \le 4 \left(\frac{1}{\langle k \rangle^2} + \frac{1}{\langle k-j \rangle^2} \right).$ □

We prove now bounds on the map $\underline{Z}^n(\hat{b}, \hat{a}) := (\underline{z}^n(\hat{b}, \hat{a}), \underline{w}^n(\hat{b}, \hat{a}))$.

Lemma E.4. *There exists a constant $C > 0$, independent of N, such that for any $s \ge 0$ and $\sigma \ge 0$*

$$\left\| \underline{Z}^n(|\hat{b}|, |\hat{a}|) \right\|_{\mathcal{P}^{s+1, \sigma}} \le C^n N^{2(n-1)} \|(b, a)\|_{\mathcal{C}^{s, \sigma}}^n, \qquad \forall n \ge 2. \tag{E.26}$$

Proof. By formula (E.16) one has that for $1 \le j \le \lfloor N/2 \rfloor$

$$\left| \underline{z}_j^n(|\hat{b}|, |\hat{a}|) \right| \le \frac{D_j}{N^{n/2}} \sum_{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta_j^n} \left| \mathcal{K}_j^n(\mathbf{i}, \boldsymbol{\iota}) \right| |u_{i_1, \iota_1}| \dots |u_{i_n, \iota_n}|,$$

$$\left| \underline{z}_{N-j}^n(|\hat{b}|, |\hat{a}|) \right| \le \frac{D_j}{N^{n/2}} \sum_{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta_{-j}^n} \left| \mathcal{K}_{N-j}^n(\mathbf{i}, \boldsymbol{\iota}) \right| |u_{i_1, \iota_1}| \dots |u_{i_n, \iota_n}|. \tag{E.27}$$

Introduce $\Lambda(\mathbf{i}) := [i_1] \cdots [i_n]$, where $[i_r] = \max(1, |i_r|)$ $\forall 1 \le r \le n$, and remark that for some constant $R > 0$ one has

$$\sup_{i_1+\cdots+i_n=j} \Lambda(\mathbf{i})^{-1} \leq \frac{R^n}{\langle j \rangle}, \qquad \forall j \in \mathbb{Z}.$$

Therefore, by Corollary E.3,

$$\left| \underline{z}_j^n(|\hat{b}|, |\hat{a}|) \right|^2$$

$$\leq \frac{1}{N^n} D_j^2 \left\| \mathcal{K}_j^n \right\|_{\Delta_j^n}^2 \left(\sup_{i_1+\cdots+i_n=j} \Lambda(\mathbf{i})^{-2s} \right) \sum_{(\mathbf{i},\boldsymbol{\iota}) \in \Delta_j^n} [i_1]^{2s} |u_{i_1,\iota_1}|^2 \cdots [i_n]^{2s} |u_{i_n,\iota_n}|^2,$$

$$\left| \underline{z}_{N-j}^n(|\hat{b}|, |\hat{a}|) \right|^2$$

$$\leq \frac{1}{N^n} D_j^2 \left\| \mathcal{K}_{N-j}^n \right\|_{\Delta_{-j}^n}^2 \left(\sup_{i_1+\cdots+i_n=-j} \Lambda(\mathbf{i})^{-2s} \right) \sum_{(\mathbf{i},\boldsymbol{\iota}) \in \Delta_{-j}^n} [i_1]^{2s} |u_{i_1,\iota_1}|^2 \cdots [i_n]^{2s} |u_{i_n,\iota_n}|^2.$$

Use now inequalities (E.25), the definition of D_j, the fact that

$$e^{2\sigma|j|} \leq e^{2\sigma|i_1|} \cdots e^{2\sigma|i_{n-1}|} e^{2\sigma|j-i_1-\cdots-i_{n-1}|},$$

and the bounds $|u_{l,\iota_l}| \leq |\hat{b}_l| + |\hat{a}_l|$, to deduce that, for any $n \geq 2$,

$$\frac{1}{N} \sum_{j=1}^{\lfloor N/2 \rfloor} [j]^{2(s+1)} e^{2\sigma|j|} \omega\left(\tfrac{j}{N}\right) \left(\left| \underline{z}_j^n(|\hat{b}|, |\hat{a}|) \right|^2 + \left| \underline{z}_{N-j}^n(|\hat{b}|, |\hat{a}|) \right|^2 \right)$$

$$\leq N^{4(n-1)} \frac{C^n}{N^n} \sum_{j=1}^{\lfloor N/2 \rfloor} [j]^{2(2-n)} e^{2\sigma|j|} \sum_{(\mathbf{i},\boldsymbol{\iota}) \in \Delta_{\pm j}^n} [i_1]^{2s} |u_{i_1,\iota_1}|^2 \cdots [i_n]^{2s} |u_{i_n,\iota_n}|^2$$

$$\leq N^{4(n-1)} C^n \left\| (b,a) \right\|_{\mathcal{C}^{s,\sigma}}^{2n}.$$

Since $w^n(\hat{b}, \hat{a})$ satisfies the same inequality, estimate (E.26) holds. $\quad\square$

Consider now the map $(\hat{b}, \hat{a}) \mapsto dZ^n(\hat{b}, \hat{a})^*$, where $dZ^n(\hat{b}, \hat{a})^*$ is the adjoint of the differential of Z^n. Explicitly, if ξ, η are vectors in $\mathbb{C}^{N-1}$ and h, g are vectors in $\mathbb{C}^N$ such that $(h, g) \equiv dZ^n(\hat{b}, \hat{a})^*(\xi, \eta)$, then the jth components of h and g are given by

$$(h_j, g_j) = \left(\sum_{k=1}^{N-1} \left(\overline{\frac{\partial z_k^n}{\partial \hat{b}_j}(\hat{b}, \hat{a})} \xi_k + \overline{\frac{\partial w_k^n}{\partial \hat{b}_j}(\hat{b}, \hat{a})} \eta_k \right), \right.$$
$$\left. \sum_{k=1}^{N-1} \left(\overline{\frac{\partial z_k^n}{\partial \hat{a}_j}(\hat{b}, \hat{a})} \xi_k + \overline{\frac{\partial w_k^n}{\partial \hat{a}_j}(\hat{b}, \hat{a})} \eta_k \right) \right). \tag{E.28}$$

Denote by $\underline{h}, \underline{g}$ the vectors of $\mathbb{C}^N$ whose components are given by

$$(\underline{h_j},\, \underline{g_j}) = \left(\sum_{k=1}^{N-1} \left(\underline{\frac{\partial z_k^n}{\partial \hat{b}_j}}(|\hat{b}|,|\hat{a}|)|\xi_k| + \underline{\frac{\partial w_k^n}{\partial \hat{b}_j}}(|\hat{b}|,|\hat{a}|)|\eta_k| \right), \right.$$
$$\left. \sum_{k=1}^{N-1} \left(\underline{\frac{\partial z_k^n}{\partial \hat{a}_j}}(|\hat{b}|,|\hat{a}|)|\xi_k| + \underline{\frac{\partial w_k^n}{\partial \hat{a}_j}}(|\hat{b}|,|\hat{a}|)|\eta_k| \right) \right). \tag{E.29}$$

We begin to study the case $n = 2$.

Lemma E.5. *There exists a constant $R > 0$, independent of N, such that $\forall s \geq 0$, $\sigma \geq 0$ one has*

$$\left\| \underline{dZ^2}(|\hat{b}|,|\hat{a}|)^*(|\xi|,|\eta|) \right\|_{\mathcal{C}^{s+2,\sigma}} \leq R N^3 \, \|(b,a)\|_{\mathcal{C}^{s,\sigma}} \, \|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}} \ . \tag{E.30}$$

Proof. By (E.3), one computes that the second order terms $Z^2 = (z^2, w^2)$ are given by

$$z_k^2(\hat{b},\hat{a}) = \frac{D_k}{N} \sum_{l \neq 0} \left(\hat{b}_l - 2\cos(\tfrac{(k-2l)\pi}{N})\hat{a}_l \right) \left(\hat{b}_{k-l} - 2\cos(\tfrac{k\pi}{N})\hat{a}_{k-l} \right) / (\lambda_{2(k-2l)}^0 - \lambda_{2k}^0)$$

$$w_k^2(\hat{b},\hat{a}) = \frac{D_k}{N} \sum_{l \neq 0} \left(\hat{b}_{N-l} - 2\cos(\tfrac{(k-2l)\pi}{N})\hat{a}_{N-l} \right) \left(\hat{b}_{l-k} - 2\cos(\tfrac{k\pi}{N})\hat{a}_{l-k} \right) / (\lambda_{2(k-2l)}^0 - \lambda_{2k}^0).$$

Let $\underline{h}$, $\underline{g}$ be as in (E.29) with $n = 2$. Using the explicit expressions for z_k^2 and w_k^2, one computes that for $0 \leq j \leq \lfloor N/2 \rfloor$

$$|\underline{h_j}| \leq \frac{1}{N} \sum_{k=1}^{N-1} \frac{\left(|\hat{b}_{k-j}| + 2|\hat{a}_{k-j}| \right) D_k (|\xi_k| + |\eta_k|)}{|\lambda_{2(k-2j)}^0 - \lambda_{2k}^0|}$$

$$\leq N \sum_{k=1}^{\lfloor N/2 \rfloor} \frac{\left(|\hat{b}_{k-j}| + 2|\hat{a}_{k-j}| \right) D_k (|\xi_k| + |\eta_k|)}{\langle k-j \rangle \langle j \rangle}$$

$$+ N \sum_{k=\lfloor N/2 \rfloor + 1}^{N-1} \frac{\left(|\hat{b}_{k-j}| + 2|\hat{a}_{k-j}| \right) D_k (|\xi_k| + |\eta_k|)}{\langle N-k+j \rangle \langle j \rangle}$$

$$\leq N \sum_{k=1}^{\lfloor N/2 \rfloor} \frac{\left(|\hat{b}_{k-j}| + 2|\hat{a}_{k-j}| \right) D_k (|\xi_k| + |\eta_k|)}{\langle k-j \rangle \langle j \rangle}$$

$$+ \frac{\left(|\hat{b}_{N-k-j}| + 2|\hat{a}_{N-k-j}| \right) D_k (|\xi_{N-k}| + |\eta_{N-k}|)}{\langle k+j \rangle \langle j \rangle}$$

$$\leq \frac{N^2}{\langle j \rangle} \sum_{k=1}^{\lfloor N/2 \rfloor} \frac{\left(|\hat{b}_{k-j}| + 2|\hat{a}_{k-j}| \right) \langle k \rangle^{1/2} (|\xi_k| + |\eta_k|)}{\langle k-j \rangle \langle k \rangle}$$

$$+ \frac{\left(|\hat{b}_{N-k-j}| + 2|\hat{a}_{N-k-j}| \right) \langle k \rangle^{1/2} (|\xi_{N-k}| + |\eta_{N-k}|)}{\langle k+j \rangle \langle k \rangle}$$

where in the last inequality we used that $D_k \leq N/\langle k \rangle^{1/2}$. With analogous computations, one verifies that

$$
|\underline{h_{N-j}}| \leq \frac{N^2}{\langle j \rangle} \sum_{k=1}^{\lfloor N/2 \rfloor} \frac{\left(|\hat{b}_{k+j}| + 2|\hat{a}_{k+j}| \right) \langle k \rangle^{1/2} (|\xi_k| + |\eta_k|)}{\langle k+j \rangle \langle k \rangle}
$$
$$
+ \frac{\left(|\hat{b}_{j-k}| + 2|\hat{a}_{j-k}| \right) \langle k \rangle^{1/2} (|\xi_{N-k}| + |\eta_{N-k}|)}{\langle k-j \rangle \langle k \rangle}.
$$

Proceeding as in the proof of Lemma B.3, one obtains that there exist constants $C, C' > 0$, independent of N, such that

$$
\frac{1}{N} \sum_{j=0}^{\lfloor N/2 \rfloor} [j]^{2(s+2)} e^{2\sigma|j|} (|\underline{h_j}|^2 + |\underline{h_{N-j}}|^2)
$$
$$
\leq C N^3 \left(\sum_{k=0}^{N-1} [k]_N^{2s} e^{2\sigma[k]_N} (|\hat{b}_k|^2 + |\hat{a}_k|^2) \right) \left(\sum_{l=1}^{N-1} [l]_N^{2s} e^{2\sigma[l]_N} [l]_N (|\xi_l|^2 + |\eta_l|^2) \right)
$$
$$
\leq C' N^6 \, \|(b,a)\|_{\mathcal{C}^{s,\sigma}}^2 \, \|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}}^2 \tag{E.31}
$$

where in the last inequality we used that $[l]_N \leq N\omega\left(\frac{l}{N}\right)$ for l integer. One verifies that $\underline{g}$ satisfies the same inequality as (E.31). Thus estimate (E.30) follows from the following inequality:

$$
\left\| \underline{dZ^2}(|\hat{b}|,|\hat{a}|)^*(|\xi|,|\eta|) \right\|_{\mathcal{C}^{s+2,\sigma}}^2
$$
$$
\leq \frac{1}{N} \sum_{j=0}^{N-1} [j]_N^{2s+4} e^{2\sigma[j]_N} \left(|\underline{h_j}|^2 + |\underline{g_j}|^2 \right). \qquad \square \tag{E.32}
$$

We study now $dZ^n(\hat{b},\hat{a})^*$ for $n \geq 3$.

Lemma E.6. *There exists a constant $R > 0$, independent of N, such that for every $s \geq 0$, $\sigma \geq 0$ and $n \geq 3$*

$$
\left\| \underline{dZ^n}(|\hat{b}|,|\hat{a}|)^*(|\xi|,|\eta|) \right\|_{\mathcal{C}^{s+2,\sigma}} \leq R^n N^{2n-1} \|(b,a)\|_{\mathcal{C}^{s,\sigma}}^{n-1} \|(\xi,\eta)\|_{\mathcal{P}^{s,\sigma}}. \tag{E.33}
$$

Proof. Let h, g be as in (E.29). We concentrate on h only, the estimates for g being analogous. Write $h_j = \sum_{k=1}^{N-1} \frac{\partial z_k^n}{\partial \hat{b}_j} \xi_k + \sum_{k=1}^{N-1} \frac{\partial w_k^n}{\partial \hat{b}_j} \eta_k =: h_{j,1} + h_{j,2}$. By (E.24) one gets that

$$
h_{j,1} = \frac{1}{N^{n/2}} \sum_{l=1}^n A_j^{n,l}(D\xi, u, \ldots, u), \qquad h_{j,2} = \frac{1}{N^{n/2}} \sum_{l=1}^n B_j^{n,l}(D\eta, u, \ldots, u)
$$

where D is defined in (E.1), the multilinear map $A_j^{n,l}$ is defined by

$$A_j^{n,l}(h, u, \ldots, u) = \sum_{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta^n} \mathcal{A}_j^{n,l}(\mathbf{i}, \boldsymbol{\iota}) u_{i_1, \iota_1} \ldots h_{i_l} \ldots u_{i_n, \iota_n},$$

$B_j^{n,l}$ is defined analogously but with kernel $\mathcal{B}_j^{n,l}(\mathbf{i}, \boldsymbol{\iota})$, and finally $\mathcal{A}_j^{n,l}$ and $\mathcal{B}_j^{n,l}$ are defined for $1 \le j \le \lfloor N/2 \rfloor$ by

$$\mathcal{A}_j^{n,l}(\mathbf{i}, \boldsymbol{\iota}) := \mathcal{K}_{i_l}^n \Big((i_1, \ldots, i_{l-1}, j, i_{l+1}, \ldots, i_n), (\iota_1, \ldots, \iota_{l-1}, 1, \iota_{l+1}, \ldots, \iota_n) \Big),$$

$$\mathcal{A}_{N-j}^{n,l}(\mathbf{i}, \boldsymbol{\iota}) := \mathcal{K}_{i_l}^n \Big((i_1, \ldots, i_{l-1}, -j, i_{l+1}, \ldots, i_n), (\iota_1, \ldots, \iota_{l-1}, 1, \iota_{l+1}, \ldots, \iota_n) \Big),$$

while $\mathcal{B}_j^{n,l}(\mathbf{i}, \boldsymbol{\iota}) = \overline{\mathcal{A}_j^{n,l}(-\mathbf{i}, \boldsymbol{\iota})}$ and $\mathcal{B}_{N-j}^{n,l}(\mathbf{i}, \boldsymbol{\iota}) = \overline{\mathcal{A}_{N-j}^{n,l}(-\mathbf{i}, \boldsymbol{\iota})}$, see (E.24). By Corollary E.3 it follows that

$$\begin{aligned}
\operatorname{supp} \mathcal{A}_j^{n,l} &= \operatorname{supp} \mathcal{B}_{N-j}^{n,l} \\
&\equiv \{ (\mathbf{i}, \boldsymbol{\iota}) : i_1 + \cdots + i_{l-1} - i_l + i_{l+1} + \cdots + i_n = -j, \, \iota_l = 1 \} \subseteq \Delta_{-j}^n, \\
\operatorname{supp} \mathcal{A}_{N-j}^{n,l} &= \operatorname{supp} \mathcal{B}_j^{n,l} \\
&\equiv \{ (\mathbf{i}, \boldsymbol{\iota}) : i_1 + \cdots + i_{l-1} - i_l + i_{l+1} + \cdots + i_n = j, \, \iota_l = 1 \} \subseteq \Delta_j^n.
\end{aligned}$$

Proceeding as in the proof of Corollary E.3, one proves that there exists a constant $R > 0$, independent of N, such that (see [30])

$$\max_{1 \le l \le n} \left(\left\| \mathcal{A}_j^{n,l} \right\|_{\Delta_{-j}^n}, \left\| \mathcal{A}_{N-j}^{n,l} \right\|_{\Delta_j^n}, \left\| \mathcal{B}_j^{n,l} \right\|_{\Delta_j^n}, \left\| \mathcal{B}_{N-j}^{n,l} \right\|_{\Delta_{-j}^n} \right) \le \frac{R^n N^{2(n-1)}}{\langle j \rangle^2}, \quad \forall n \ge 3 .$$

$$(\text{E.34})$$

Thus $\underline{h}$, defined in (E.29), satisfies

$$|\underline{h_j}| \le \frac{1}{N^{n/2}} \sum_{l=1}^n \left(\underline{A_j^{n,l}}(|D\xi|, |u|, \ldots, |u|) + \underline{B_j^{n,l}}(|D\eta|, |u|, \ldots, |u|) \right),$$

where $\underline{A_j^{n,l}}(h, u, \ldots, u) = \sum_{(\mathbf{i}, \boldsymbol{\iota}) \in \Delta^n} \left| \mathcal{A}_j^{n,l}(\mathbf{i}, \boldsymbol{\iota}) \right| u_{i_1, \iota_1} \ldots h_{i_l} \ldots u_{i_n, \iota_n}$, and $\underline{B_j^{n,l}}$ is defined in analogous way. Then, using (E.34) and arguing as in the proof of Lemma E.4, one proves the estimate

$$\frac{1}{N} \sum_{j=0}^{N-1} [j]_N^{2(s+2)} e^{2\sigma[j]_N} |\underline{h_j}|^2$$

$$\le R^n N^{4n-5} \|(b, a)\|_{\mathcal{C}^{s,\sigma}}^{2(n-1)} \left(\frac{1}{N} \sum_{l=1}^{N-1} [l]_N^{2s} e^{2\sigma[l]_N} D_l^2 (|\xi_l|^2 + |\eta_l|^2) \right)$$

$$\leq R^n N^{4n-2} \left\| (b,a) \right\|_{\mathcal{C}^{s,\sigma}}^{2(n-1)} \left\| (\xi,\eta) \right\|_{\mathcal{P}^{s-1,\sigma}}^2 \ ,$$

where in the last inequality we used that $D_l^2 \leq \frac{N^3}{[l]_N^2} \omega \left(\frac{l}{N} \right)$. One verifies that $\underline{g}$ satisfies the same inequality, thus estimate (E.33) follows. $\square$

We can finally prove property $(Z4)$. Let $s \geq 0$, $\sigma \geq 0$ be fixed. By Lemmas E.4, E.5 and E.6, there exist $C_1, C_2, \epsilon_* > 0$, independent of N, such that for every $0 < \epsilon \leq \epsilon_*$ it holds that

$$\sup_{\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \epsilon/N^2} \left\| \underline{Z}^0(b,a) \right\|_{\mathcal{P}^{s+1,\sigma}} \leq \sum_{n \geq 2} \sup_{\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \epsilon/N^2} \left\| \underline{Z}^n(b,a) \right\|_{\mathcal{P}^{s+1,\sigma}}$$

$$\leq \sum_{n \geq 2} R^n N^{2(n-1)} \frac{\epsilon^n}{N^{2n}} \leq \frac{C_1 \epsilon^2}{N^2},$$

$$\sup_{\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \epsilon/N^2} \left\| \underline{dZ}^0(b,a)^* \right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s+2,\sigma})} \leq \sum_{n \geq 2} \sup_{\|(b,a)\|_{\mathcal{C}^{s,\sigma}} \leq \epsilon/N^2} \left\| \underline{dZ}^n(b,a)^* \right\|_{\mathcal{L}(\mathcal{P}^{s,\sigma}, \mathcal{C}^{s+2,\sigma})}$$

$$\leq \sum_{n \geq 2} R^n N^{2n-1} \frac{\epsilon^{n-1}}{N^{2(n-1)}} \leq C_2 N \epsilon \ .$$

References

[1] D. Bambusi, B. Grébert, Birkhoff normal form for partial differential equations with tame modulus, Duke Math. J. 135 (3) (2006) 507–567.
[2] Dario Bambusi, Antonio Ponno, On metastability in FPU, Comm. Math. Phys. 264 (2) (2006) 539–561.
[3] Dario Bambusi, Thomas Kappeler, Thierry Paul, De Toda à KdV, C. R. Math. Acad. Sci. Paris 347 (17–18) (2009) 1025–1030.
[4] D. Bambusi, T. Kappeler, T. Paul, Dynamics of periodic Toda chains with a large number of particles, arXiv:1309.5441 [math.AP], September 2013.
[5] D. Bambusi, T. Kappeler, T. Paul, From Toda to KdV, arXiv:1309.5324 [math.AP], September 2013.
[6] D. Bättig, B. Grébert, J.-C. Guillot, T. Kappeler, Fibration of the phase space of the periodic Toda lattice, J. Math. Pures Appl. (9) 72 (6) (1993) 553–565.
[7] G. Benettin, A. Ponno, Time-scales to equipartition in the Fermi–Pasta–Ulam problem: finite-size effects and thermodynamic limit, J. Stat. Phys. 144 (4) (2011) 793–812.
[8] G. Benettin, H. Christodoulidi, A. Ponno, The Fermi–Pasta–Ulam problem and its underlying integrable dynamics, J. Stat. Phys. 152 (2) (2013) 195–212.
[9] Luisa Berchialla, Luigi Galgani, Antonio Giorgilli, Localization of energy in FPU chains, Discrete Contin. Dyn. Syst. 11 (4) (2004) 855–866.
[10] Luisa Berchialla, Antonio Giorgilli, Simone Paleari, Exponentially long times to equipartition in the thermodynamic limit, Phys. Lett. A 321 (3) (2004) 167–172.
[11] A. Bloch, F. Golse, T. Paul, A. Uribe, Dispersionless Toda and Toeplitz operators, Duke Math. J. 117 (1) (2003) 157–196.
[12] A. Carati, An averaging theorem for Hamiltonian dynamical systems in the thermodynamic limit, J. Stat. Phys. 128 (4) (2007) 1057–1077.
[13] Andrea Carati, Alberto Mario Maiocchi, Exponentially long stability times for a nonlinear lattice in the thermodynamic limit, Comm. Math. Phys. 314 (1) (2012) 129–161.
[14] L.H. Eliasson, Normal forms for Hamiltonian systems with Poisson commuting integrals—elliptic case, Comment. Math. Helv. 65 (1) (1990) 4–35.
[15] W.E. Ferguson Jr., H. Flaschka, D.W. McLaughlin, Nonlinear normal modes for the Toda chain, J. Comput. Phys. 45 (2) (1982) 157–209.

[16] E. Fermi, J. Pasta, S. Ulam, Studies of non linear problems, in: Enrico Fermi Collected Papers, vol. II, University of Chicago Press/Accademia Nazionale dei Lincei, Chicago/Roma, 1965, pp. 977–988.
[17] H. Flaschka, The Toda lattice. I. Existence of integrals, Phys. Rev. B (3) 9 (1974) 1924–1925.
[18] H. Flaschka, D.W. McLaughlin, Canonically conjugate variables for the Korteweg–de Vries equation and the Toda lattice with periodic boundary conditions, Progr. Theoret. Phys. 55 (2) (1976) 438–456.
[19] Antonio Giorgilli, Simone Paleari, Tiziano Penati, Extensive adiabatic invariants for nonlinear chains, J. Stat. Phys. 148 (6) (2012) 1106–1134.
[20] Ernst Hairer, Christian Lubich, On the energy distribution in Fermi–Pasta–Ulam lattices, Arch. Ration. Mech. Anal. 205 (3) (2012) 993–1029.
[21] M. Hénon, Integrals of the Toda lattice, Phys. Rev. B (3) 9 (1974) 1921–1923.
[22] Andreas Henrici, Thomas Kappeler, Birkhoff normal form for the periodic Toda lattice, in: Integrable Systems and Random Matrices, in: Contemp. Math., vol. 458, Amer. Math. Soc., Providence, RI, 2008, pp. 11–29.
[23] Andreas Henrici, Thomas Kappeler, Global action–angle variables for the periodic Toda lattice, Int. Math. Res. Not. IMRN 11 (2008) 52, Art. ID rnn031.
[24] Andreas Henrici, Thomas Kappeler, Global Birkhoff coordinates for the periodic Toda lattice, Nonlinearity 21 (12) (2008) 2731–2758.
[25] Andreas Henrici, Thomas Kappeler, Resonant normal form for even periodic FPU chains, J. Eur. Math. Soc. (JEMS) 11 (2009) 1025–1056.
[26] Thomas Kappeler, Fibration of the phase space for the Korteweg–de Vries equation, Ann. Inst. Fourier (Grenoble) 41 (3) (1991) 539–575.
[27] Thomas Kappeler, Jürgen Pöschel, KdV & KAM, Ergeb. Math. Grenzgeb. (3) (Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics), vol. 45, Springer-Verlag, Berlin, 2003.
[28] T. Kappeler, B. Schaad, P. Topalov, Qualitative features of periodic solutions of KdV, Comm. Partial Differential Equations 38 (9) (2013) 1626–1673.
[29] Tosio Kato, Perturbation Theory for Linear Operators, Grundlehren Math. Wiss., vol. 132, Springer-Verlag New York, Inc., New York, 1966.
[30] Sergei Kuksin, Galina Perelman, Vey theorem in infinite dimensions and its application to KdV, Discrete Contin. Dyn. Syst. 27 (1) (2010) 1–24.
[31] A. Maiocchi, D. Bambusi, A. Carati, An averaging theorem for FPU in the thermodynamic limit, J. Stat. Phys. 155 (2) (2014) 300–322.
[32] Jorge Mujica, Complex Analysis in Banach Spaces, North-Holl. Math. Stud., vol. 120, North-Holland Publishing Co., Amsterdam, 1986, Holomorphic functions and domains of holomorphy in finite and infinite dimensions, Notas de Matemática [Mathematical Notes], vol. 107.
[33] N.V. Nikolenko, The method of Poincaré normal forms in problems of integrability of equations of evolution type, Uspekhi Mat. Nauk 41 (5(251)) (1986) 109–152, 263.
[34] B. Rink, Symmetry and resonance in periodic FPU chains, Comm. Math. Phys. 218 (3) (2001) 665–685.
[35] Guido Schneider, C. Eugene Wayne, Counter-propagating waves on fluid surfaces and the continuum limit of the Fermi–Pasta–Ulam model, in: International Conference on Differential Equations, vols. 1, 2, Berlin, 1999, World Sci. Publ., River Edge, NJ, 2000, pp. 390–404.
[36] M. Toda, Vibration of a chain with nonlinear interaction, J. Phys. Soc. Jpn. 22 (February 1967) 431.
[37] François Trèves, An abstract nonlinear Cauchy–Kovalevska theorem, Trans. Amer. Math. Soc. 150 (1970) 77–92.
[38] Pierre van Moerbeke, The spectrum of Jacobi matrices, Invent. Math. 37 (1) (1976) 45–81.
[39] J. Vey, Sur certains systèmes dynamiques séparables, Amer. J. Math. 100 (3) (1978) 591–614.