Ann. Inst. H. Poincaré
Anal. Non Linéaire 40 (2023), 1009–1049
DOI 10.4171/AIHPC/55

Dynamics of nonlinear Klein–Gordon equations in low regularity on $\mathbb{S}^2$

Joackim Bernier, Benoît Grébert, and Gabriel Rivière

Abstract. We describe the long-time behavior of small nonsmooth solutions to the nonlinear Klein–Gordon equations on the sphere $\mathbb{S}^2$. More precisely, we prove that the low harmonic energies (also called super-actions) are almost preserved for times of order ε^{-r}, where $r \gg 1$ is an arbitrarily large number and $\varepsilon \ll 1$ is the norm of the initial datum in the energy space $H^1 \times L^2$. Roughly speaking, it means that, in order to exchange energy, modes have to oscillate at the same frequency. The proof relies on new multilinear estimates on Hamiltonian vector fields to put the system in Birkhoff normal form. They are derived from new probabilistic bounds on products of Laplace eigenfunctions that we obtain using Levy's concentration inequality.

1. Introduction

The linear Klein–Gordon equation classically appears as a natural first candidate to describe a relativistic version of quantum mechanics [13, Chap. 1] and it can be written on the sphere as

$$\partial_t^2 \Phi(t,x) = \Delta\Phi(t,x) - \mu\Phi(t,x),$$

where $\mu > 0$ is an external parameter referred to as the *mass* (although physically speaking, μ is rather the square of the mass, up to taking $c = 1$ and $\hbar = 1$), $x \in \mathbb{S}^2$ (the Euclidean unit sphere of $\mathbb{R}^3$), $t \in \mathbb{R}$, $\Phi(t,x) \in \mathbb{R}$ and Δ denotes the Laplace–Beltrami operator on the sphere. As usual, we rewrite this evolution equation as a first-order system

$$\partial_t \begin{pmatrix} \Phi \\ \partial_t \Phi \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \Delta - \mu & 0 \end{pmatrix} \begin{pmatrix} \Phi \\ \partial_t \Phi \end{pmatrix}$$

and the change of variable

$$u := (\mu - \Delta)^{1/4}\Phi + i(\mu - \Delta)^{-1/4}\partial_t \Phi \tag{1}$$

makes the linear Klein–Gordon equation diagonal,

$$i\partial_t u = \sqrt{\mu - \Delta} u.$$

2020 Mathematics Subject Classification. Primary 37K55; Secondary 35Q40, 35Q75, 37K45.
Keywords. Birkhoff normal forms, low regularity, Hamiltonian PDE, Klein–Gordon, random Hilbertian basis.

Indeed, it is well known that the spherical harmonics (i.e. the restriction to $\mathbb{S}^2$ of homogeneous harmonic polynomials on $\mathbb{R}^3$) make the Laplace–Beltrami operator diagonal:

$$L^2(\mathbb{S}^2;\mathbb{R}) = \overline{\bigoplus_{\ell\in\mathbb{N}} E_\ell}, \quad \text{where } E_\ell = \mathrm{Ker}(\Delta + \ell(\ell+1)\mathrm{Id}_{L^2}) \simeq \mathbb{R}^{2\ell+1} \tag{2}$$

is the space of spherical harmonics of degree ℓ. In other words, the linear Klein–Gordon equation can be rewritten as

$$\forall \ell \in \mathbb{N}, \quad i\partial_t \Pi_\ell u = \omega_\ell \Pi_\ell u, \quad \text{where } \omega_\ell := \sqrt{\ell(\ell+1)+\mu}$$

and Π_ℓ denotes the orthogonal projector on E_ℓ.

On the one hand, it is relevant to note that the following quantities are constants of motion for the linear Klein–Gordon equation:

$$I_v(u(t)) = \left| \int_{\mathbb{S}^2} u(t,x)v(x)\mathrm{dvol}_{\mathbb{S}^2}(x) \right|^2, \quad \text{with } \ell \in \mathbb{N},\ v \in E_\ell.$$

Actually, they describe accurately its dynamics (up to the exact values of the frequencies ω_ℓ). However, they are too sharp to survive to perturbations of the linear Klein–Gordon equation. Indeed, due to the multiplicities of the eigenvalues of the Laplace–Beltrami operator (E_ℓ is of dimension $2\ell+1$), one could design spectral perturbations commuting with its vector field but destroying completely these constants of the motion (and so a fortiori we also expect the same phenomenon in the nonlinear case as in [36, 37]).

On the other hand, the harmonic energies (also called *super-actions*)

$$J_\ell(u(t)) := \|\Pi_\ell u(t)\|_{L^2}^2 =: \mathcal{E}_\ell(\Phi(t), \partial_t\Phi(t))$$

are much more robust constants of motion because they do not describe the energy exchanges inside the clusters E_ℓ. They only encode the energy preservation of each cluster. Note that they can be rewritten (in the original variables $(\Phi, \partial_t\Phi)$) as

$$\begin{aligned}\mathcal{E}_\ell(\Phi(t), \partial_t\Phi(t)) := {} & (\ell(\ell+1)+\mu)^{1/2}\|\Pi_\ell\Phi(t)\|_{L^2}^2 \\ & + (\ell(\ell+1)+\mu)^{-1/2}\|\Pi_\ell\partial_t\Phi(t)\|_{L^2}^2.\end{aligned} \tag{3}$$

In this paper we address the question of their preservation by a nonlinear perturbation of the linear Klein–Gordon equation. More precisely, we consider the nonlinear Klein–Gordon equation

$$\partial_t^2\Phi(t,x) = \Delta\Phi(t,x) - \mu\Phi(t,x) + g(x)(\Phi(t,x))^{p-1}, \tag{KG}$$

where $p \geq 3$ is an integer and $g \in L^\infty(\mathbb{S}^2;\mathbb{R})$ is a given factor making the equation possibly inhomogeneous. The equation is naturally equipped with initial data $\Phi^{(0)} \in H^1(\mathbb{S}^2;\mathbb{R})$ and $\dot{\Phi}^{(0)} \in L^2(\mathbb{S}^2;\mathbb{R})$, i.e.

$$\forall x \in \mathbb{S}^2, \quad \Phi(0,x) = \Phi^{(0)}(x) \quad \text{and} \quad \partial_t\Phi(0,x) = \dot{\Phi}^{(0)}(x).$$

Focusing only on small solutions, $\varepsilon := \|\Phi^{(0)}\|_{H^1} + \|\dot{\Phi}^{(0)}\|_{L^2} \ll 1$, (KG) is a perturbation of the linear Klein–Gordon equation and the question of the preservation of the harmonic energies (3) makes sense.

Since (KG) is locally well posed (see Section 6.1 for details), the dynamics of (KG) remain close to the dynamics of the linearized equation for times of order $\varepsilon^{-(p-2)}$. As a consequence, on such a timescale the super-actions are almost preserved. However, their conservation on longer timescales is nontrivial. Actually, there exist counterexamples for similar systems: the cubic wave equation on $\mathbb{T}^2$ [33] and the cubic Klein–Gordon equation on $\mathbb{S}^3$ with a unit mass [12, 20]. Nevertheless, they are closely related to the existence of resonances (i.e. the frequencies ω_ℓ have to be rationally linked) which only hold for exceptional values of the mass μ.

For generic values of the mass μ, in [3] , Bambusi, Delort, Grébert and Szeftel prove the almost preservation, for very long times, of the harmonic energies of the nonlinear Klein–Gordon equations on Zoll manifolds (which include $\mathbb{S}^d$ for all $d \geq 2$). Nevertheless, their result only holds for very smooth solutions (in particular g has to be smooth). More precisely, they prove[1] that for all $r \gg 1$ chosen arbitrarily large, there exists $s_0(r)$ such that for all $s \geq s_0(r)$, provided that ε (the norm of the initial datum $(\Phi^{(0)}, \dot{\Phi}^{(0)})$ in $H^{s+1/2} \times H^{s-1/2}$) is small enough, while $|t| < \varepsilon^{-r}$, the solution to the nonlinear Klein–Gordon equation exists and it satisfies

$$|t| \leq \varepsilon^{-r} \quad \Rightarrow \quad \sum_{\ell \in \mathbb{N}} \langle \ell \rangle^{2s} |\mathcal{E}_\ell(\Phi(t), \partial_t \Phi(t)) - \mathcal{E}_\ell(\Phi^{(0)}, \dot{\Phi}^{(0)})| \lesssim \varepsilon^p. \tag{4}$$

The main flaw of this result is the smoothness assumption $s \geq s_0(r)$. Indeed, in their construction, the smoothness parameter $s_0(r)$ grows at least linearly with respect to r. In other words, the longer the time during which they prove the preservation of the super-actions is, the smoother the solutions have to be. This smoothness assumption is crucial in their proof and is systematically used to prove similar results – see e.g. [2, 4, 5, 14, 19, 22, 23, 38]. Nevertheless, on simpler models, numerical experiments strongly suggest that this assumption is irrelevant (i.e. $s_0(r)$ should not depend on r); see e.g. [18, 19] for discussions about (KG) on $\mathbb{T}$.

Actually, in [3] the authors are interested in the preservation of super-actions because they aim to prove the *almost global well-posedness* of the equation (i.e. well-posedness for times of order ε^{-r} with r arbitrarily large). Roughly speaking, since

$$\|u(t)\|_{H^s}^2 = \sum_{\ell \in \mathbb{N}} \langle \ell \rangle^{2s} \mathcal{E}_\ell(\Phi(t), \partial_t \Phi(t)),$$

they proceed by bootstrap: assuming that $\|u(t)\|_{H^s}^2 \leq 2\|u(0)\|_{H^s}^2 \simeq \varepsilon^2$, they control the variations of the super-action using (4) and, as a corollary, they deduce the sharper estimate

$$\|u(t)\|_{H^s}^2 = \|u(0)\|_{H^s}^2 + \mathcal{O}(\|u(0)\|_{H^s}^p).$$

---

[1] Actually, they only prove an ℓ^∞ instead of ℓ^1 estimate (4) (see [3, Rem. 3.21]). Indeed, since they are only really interested in the variations of the H^s-norm, they have not written a sharp estimate on the variation of the super-actions. Nevertheless, estimate (4) would be a direct corollary of their proof.

However, in low dimensions ($d \leq 2$), it is well known that smoothness is not required to obtain solutions for very long times. Indeed, the preservation of the Hamiltonian

$$\mathcal{H}(\Phi, \partial_t \Phi) = \int_{\mathbb{S}^2} \frac{|\nabla \Phi(x)|^2}{2} + \mu \frac{(\Phi(x))^2}{2} + \frac{(\partial_t \Phi(x))^2}{2} - \frac{g(x)(\Phi(x))^p}{p} \mathrm{dvol}_{\mathbb{S}^2}(x) \quad (5)$$

provides an a priori global control of the energy norm ($H^1 \times L^2$) of small solutions (see Lemma 6.1). Hence, one can derive the global well-posedness of the Cauchy problem associated with (KG) (provided that the initial data are small enough; see Proposition 6.2 for details). Therefore, it is all the more natural to try to remove the smoothness assumption $s \geq s_0(r)$ of [3] to control the variations of the harmonic energies.

In the following theorem, which is the main result of this paper, we control, without regularity assumption, the variations of the low super-actions:

Theorem 1.1. *For all $r \geq p$, all $\nu > 0$ and almost all $\mu > 0$, there exist $\varepsilon_0 > 0$, $C > 0$ and $\alpha_r > 0$ (depending only on r) such that, provided $\varepsilon := \|\Phi^{(0)}\|_{H^1} + \|\dot{\Phi}^{(0)}\|_{L^2} < \varepsilon_0$, the global solution to* (KG) *satisfies*

$$|t| < \varepsilon^{-r} \quad \Rightarrow \quad \forall \ell \in \mathbb{N}, \ \ |\mathcal{E}_\ell(\Phi^{(0)}, \dot{\Phi}^{(0)}) - \mathcal{E}_\ell(\Phi(t), \partial_t \Phi(t))| \leq C \langle \ell \rangle^{\alpha_r} \varepsilon^{p-\nu}.$$

Let us compare this result with that of [3] (i.e. (4)). For low super-actions (i.e. $\ell \simeq 1$), Theorem 1.1 is much better as it provides the same control on the variations of the super-actions (up to the $\varepsilon^{-\nu}$ loss) without requiring any smoothness assumption. Conversely, contrary to (4), due to the $\langle \ell \rangle^{\alpha_r}$ loss, our result does not provide any information about the variation of the very high super-actions (i.e. $\ell \gg \varepsilon^{-(p-2)/\alpha_r}$). Nevertheless, since the loss with respect to ℓ is polynomial, Theorem 1.1 provides a nontrivial control of the variations of some "quite high" super-actions (i.e. $1 \ll \ell \ll \varepsilon^{-(p-2)/\alpha_r}$).

Using this optimization and the a priori control on the energy norm of the solutions, we derive the following corollary, which can be viewed as a kind of weak orbital stability result.

Corollary 1.2. *For all $r \geq p$, $s < 1/2$ and almost all $\mu > 0$, there exist $\varepsilon_0 > 0$, $C > 0$ and $\delta > 0$ (which does not depend on μ) such that, provided $\varepsilon := \|\Phi^{(0)}\|_{H^1} + \|\dot{\Phi}^{(0)}\|_{L^2} < \varepsilon_0$, the global solution of* (KG) *satisfies*

$$|t| < \varepsilon^{-r} \quad \Rightarrow \quad \Big\| u(t) - \sum_{\ell \in \mathbb{N}} e^{-iH_\ell(t)} \Pi_\ell u(0) \Big\|_{H^s} \leq C \varepsilon^{1+\delta},$$

where $H_\ell(t)\colon E_\ell \otimes \mathbb{C} \to E_\ell \otimes \mathbb{C}$ are Hermitian maps and $u \in C^0(\mathbb{R}; H^{1/2})$ is defined by (1).

Further bibliographical comments

The question of the stability of the linear dynamics makes sense for most nonlinear partial differential equations on confined domains. In high regularity, Birkhoff normal forms lead to many important successes in proving the stability of several other interesting systems: [4, 10, 11, 14, 28, 30, 31, 34, 46] in the nonresonant case and [1, 6, 7, 15] in the resonant case.

For Klein–Gordon, the papers [2, 4, 5, 14, 19, 22, 23, 38] provide results similar to that of Bambusi, Delort, Grébert and Szeftel [3] (i.e. preservation of the super-actions up to times of order ε^{-r} with r arbitrarily large) but hold on other manifolds or with quasi-linear perturbations. The works [21, 24–27, 29] only reach shorter times of stability but improve that given by the local well-posedness (i.e. they get stability for $|t| < \varepsilon^{-q}$ with $q > p-2$ but not arbitrarily large). On some manifolds, for high modes, due to the quasi-resonance (i.e. when the small divisors are too small), some of these timescales seem so far to be optimal. We also mention the recent works [9, 35] about the existence of KAM tori for the nonlinear Klein–Gordon equations.

Very recently, in [8], the first two authors have introduced a new way of performing Birkhoff normal forms for Hamiltonians PDEs which, contrary to the previous results, allows nonsmooth solutions to be dealt with. As in Theorem 1.1, they prove almost-conservation, for very long times, in low regularity, of the low (super-)actions of several nonlinear dispersive PDEs on tori or boxes (including nonlinear Klein–Gordon equations on $[0, \pi]$ with homogeneous Dirichlet boundary conditions). Nevertheless, as discussed below, to be extended to more general domains (like spheres), this result requires nontrivial multilinear vector field estimates. The derivation and the proof of these estimates on the sphere $\mathbb{S}^2$ are the main technical novelties of this paper (see Sections 2 and 4).

Comments about the results

- The arbitrarily small loss $\varepsilon^{-\nu}$ in Theorem 1.1 is the same as that of [8, Thm. 1.21] (about nonlinear Schrödinger equations on $\mathbb{T}^2$). It is due to the fact that, in dimension 2, H^1 is not an algebra.
- Reasoning as in [8, Cor. 1.14], we could prove that Corollary 1.2 holds in the critical case $s = 1/2$ provided that the initial data are a little smoother: $\varepsilon = \|\Phi^{(0)}\|_{H^{1+\eta}} + \|\dot{\Phi}^{(0)}\|_{H^{\eta}}$ for some $\eta > 0$ (and δ would depend on η).
- We could consider much more general nonlinearity in (KG) (e.g. nonlocal or nonpolynomial). Actually, we chose $g(x)(u(x))^{p-1}$ for simplicity.
- We are quite confident that our results could be extended to Zoll surfaces. Nevertheless, it would generate a lot a technicalities. It seems to us that we could adapt our multilinear estimates by considering clusters of quasi-modes (as in [3]) but the cohomological equations would be much harder to solve (because they would not be diagonal). Moreover, it would raise several interesting questions which deserve further investigation. For example, is it possible to prove the preservation of the low actions (i.e. not only the super-actions) for very long times on a generic Zoll manifold and with a generic mass? Somehow, it would be one way to prove the stability of the linear dynamics.

- Conversely, it is not clear whether a similar result could be proven in a higher dimension (for example on $\mathbb{S}^3$). First, the equation would not necessarily be well posed. Moreover, our method is strongly related to the fact that H^1 is an algebra (or almost an algebra like on $\mathbb{S}^2$). Indeed, roughly speaking, the Birkhoff normal procedure generates vector fields of arbitrarily large order which are somehow similar to $(\Phi, \partial_t \Phi) \mapsto \Phi^n$ with $p \leq n \leq r + p$. Hence, the requirement that the energy space is an algebra looks unavoidable.

Comments about the proof

The proof of our results follows the new Birkhoff normal form strategy introduced by the first two authors (see [8, §1.4] for an informal description of this new strategy). Roughly speaking, compared with [2, 4], it consists in removing terms which are usually small thanks to the smoothness assumption (and so which are unsolved in that case) using a stronger nonresonance condition. More precisely, we need that the small divisors are controlled by the smallest index instead of the third largest. Even if this new Diophantine condition may seem too restrictive, it is typically satisfied for (KG) since the eigenvalues of $\sqrt{\mu - \Delta}$ accumulate polynomially fast on $\mathbb{Z} + 1/2$, which is an affine lattice. Actually it is a quite direct application of [8, Prop. 2.1] as explained in Section 3.

Nevertheless, as usual, the implementation of a normal form procedure requires some structures on the nonlinear part of the vector field of the equation: it has to belong to a class of vector fields which is stable by Lie brackets, resolution of cohomological equations and whose vector fields enjoy good multilinear estimates in the energy space (here $H^{1/2}$ with respect to the variable u defined by (1)). In [8], such classes have been developed to deal with Hamiltonian PDEs on tori (or boxes) in low regularity. Unfortunately, it seems hopeless to adapt them in more general domains like spheres as they strongly rely on the exceptionally good algebraic properties of the eigenfunctions of the Laplace operator (which are the complex exponentials). On spheres (and more generally on compact Riemannian manifolds), Delort and Szeftel have developed powerful classes of vector fields (see e.g. [25, 26]) on which most of the Birkhoff normal form results are based. Unfortunately, these classes are unsuitable to work in low regularity as they require a lot of smoothness and it seemed unlikely to us that they could be adapted in low regularity. Hence, we chose to follow a slightly different route, relying on probabilistic tools referred to as Levy's concentration inequalities [39] (see Theorem 2.8), in order to build the Hamiltonian classes adapted to our problem. See Section 2 for the probabilistic estimates and Section 4 for the multilinear vector field estimates.

Notation

It is natural (and usual) to index eigenvectors of the Laplace–Beltrami operators on $\mathbb{S}^2$ by points in a discrete triangle. As a consequence, for all $M \in (0, \infty]$, we define

$$\mathcal{T}_M := \{(\ell, m) \in \mathbb{N} \times \mathbb{Z} \mid 0 \leq \ell \leq M \text{ and } -\ell \leq m \leq \ell\}.$$

We warn the reader that, as usual, we adopt the following convenient abuse of notation: being given $M > 0$, $k \in \mathcal{T}_M$, $\sigma \in \{-1, 1\}$ and $u = (u_{k'})_{k' \in \mathcal{T}_M} \in \mathbb{C}^{\mathcal{T}_M}$, we set

$$u_k^\sigma = u_k \quad \text{if } \sigma = 1 \quad \text{and} \quad u_k^\sigma = \overline{u_k} \quad \text{if } \sigma = -1.$$

If $\mathfrak{p}$ is a parameter or a list of parameters and $x, y \in \mathbb{R}$ then we write $x \lesssim_{\mathfrak{p}} y$ if there exists a constant $c(\mathfrak{p})$, depending continuously on $\mathfrak{p}$, such that $x \lesssim c(\mathfrak{p}) y$. Similarly, we write $x \gtrsim_{\mathfrak{p}} y$ if $y \lesssim_{\mathfrak{p}} x$ and $x \approx_{\mathfrak{p}} y$ if $x \lesssim_{\mathfrak{p}} y \lesssim_{\mathfrak{p}} x$.

2. A good orthonormal basis

Recall that

$$E_\ell = \mathrm{Ker}(\Delta + \ell(\ell + 1)\mathrm{Id}_{L^2(\mathbb{S}^2,\mathbb{R})}) \simeq \mathbb{R}^{2\ell+1}, \tag{6}$$

and we will denote by $\mathcal{B}_\ell$ the set of orthonormal bases of the Euclidean space E_ℓ. More generally, we denote by $\mathcal{B}$ the set of orthonormal bases of $L^2(\mathbb{S}^2; \mathbb{R})$:

$$\mathcal{B} := \big\{ b = (b_\ell)_{\ell \in \mathbb{N}} : \forall \ell \geq 0,\ b_\ell \in \mathcal{B}_\ell \big\}.$$

Hence, an element in $\mathcal{B}_\ell$ is an orthonormal basis of E_ℓ that we will denote by $b_\ell = (e_{\ell,m})_{-\ell \leq m \leq \ell}$ and an element of $\mathcal{B}$ can be represented as

$$b = (b_\ell)_{\ell \in \mathbb{N}} = (e_k)_{k \in \mathcal{T}_\infty} = (e_{(\ell,m)})_{(\ell,m) \in \mathcal{T}_\infty}.$$

When representing vector fields in a Hilbertian basis $b = (e_k)_{k \in \mathcal{T}_\infty} \in \mathcal{B}$ (which seems natural to perform Birkhoff normal forms), it is classical to end up with estimating quantities of the form

$$\int_{\mathbb{S}^2} e_{k_1}(x) \cdots e_{k_p}(x) \, \mathrm{dvol}_{\mathbb{S}^2}(x),$$

where $(k_1, \ldots, k_p)$ is some fixed element in $\mathcal{T}_\infty^p$. In the case of the round sphere, an orthonormal basis in $\mathcal{B}$ can be identified with a basis of homogeneous harmonic polynomials on $\mathbb{R}^3$ and one can make use of this structure to get good estimates. For instance, following [25, Ex. 4.2], we can verify that

$$\exists\, 1 \leq j_0 \leq r \text{ such that } \sum_{j \neq j_0} \ell_j < \ell_{j_0} \Rightarrow \int_{\mathbb{S}^2} e_{(\ell_1,m_1)}(x) \cdots e_{(\ell_p,m_p)}(x) \, \mathrm{dvol}_{\mathbb{S}^2}(x) = 0. \tag{7}$$

See also [26, Prop. 1.2.1] for related results on more general manifolds. However, without any assumption on the relative size of the ℓ_j, it seems that the best one can expect for a general orthonormal basis is to apply Hölder's inequality:

$$\left| \int_{\mathbb{S}^2} e_{(\ell_1,m_1)}(x) \cdots e_{(\ell_p,m_p)}(x) \, \mathrm{dvol}_{\mathbb{S}^2}(x) \right| \leq \|e_{(\ell_1,m_1)}\|_{L^p} \cdots \|e_{(\ell_p,m_p)}\|_{L^p}.$$

Then a classical result on Laplace eigenfunctions [42] states that, for any $(\ell,m)\in\mathcal{T}_\infty$, $\|e_{(\ell,m)}\|_{L^p}\leq C_p\langle\ell\rangle^{\delta(p)}$ with $\delta(p)=\max\{\frac{1}{4}-\frac{1}{2p},\frac{1}{2}-\frac{2}{p}\}$. Moreover, these bounds on L^p-norms are known to be sharp along certain sequences of the standard basis of spherical harmonics [43]. Despite these a priori bounds and thanks to spectral degeneracies, there is some flexibility in the choice of the orthonormal basis $b\in\mathcal{B}$ we are working with. Following [16, Thm. 6] (see also [41, 45] or [47, Thm. 18.5]), one can in fact prove that there exist many elements b in $\mathcal{B}$ (in fact almost all) for which the L^p-norms are uniformly bounded. Thus, for such a basis b, one can find a constant $C_b>0$ such that, for every $(k_1,\ldots,k_p)\in\mathcal{T}_\infty^p$,

$$\left|\int_{\mathbb{S}^2}e_{k_1}(x)\cdots e_{k_p}(x)\,\mathrm{dvol}_{\mathbb{S}^2}(x)\right|\leq C_b. \tag{8}$$

Unfortunately, this information does not seem to be enough to handle Birkhoff normal forms for data with low regularity, as we are aiming to do. Hence, we need to work a little more. As we will see in the upcoming sections, the missing information to handle our Birkhoff normal form procedure is to construct an orthonormal basis in $\mathcal{B}$ for which these integrals have enough decay when there exists an index $1\leq j_0\leq p$ such that

$$(\ell_j,m_j)=(\ell_{j_0},m_{j_0})\quad\Rightarrow\quad j=j_0.$$

To that aim, we will prove the following theorem which is the main result of this section:

Theorem 2.1. *Let $g\in L^\infty(\mathbb{S}^2;\mathbb{R})$ and let $p\geq 3$. Then there exist a constant $C_{g,p}>0$ and an orthonormal basis $b=(e_k)_{k\in\mathcal{T}_\infty}\in C^\infty(\mathbb{S}^2;\mathbb{R})^{\mathcal{T}_\infty}$ of $L^2(\mathbb{S}^2;\mathbb{R})$ such that, for all $\mathbf{k}=(k_1,\ldots,k_p)\in\mathcal{T}_\infty^p$, we have*

$$\left|\int_{\mathbb{S}^2}e_{k_1}(x)\cdots e_{k_p}(x)g(x)\,\mathrm{dvol}_{\mathbb{S}^2}(x)\right|\leq C_{g,p}\min\left\{1,\frac{\log^p(2+|\ell|_\infty)}{\sqrt{\Upsilon(\mathbf{k})}}\right\}, \tag{9}$$

where $|\ell|_\infty=\max_{1\leq j\leq p}\ell_j$ and

$$\Upsilon(\mathbf{k}):=\max\{1\}\cup\left\{\langle\ell_j\rangle:\forall j'\neq j,\ k_{j'}\neq k_j\right\}. \tag{10}$$

Moreover, $b\in\mathcal{B}$, i.e. for all $k=(\ell,m)\in\mathcal{T}_\infty$, we have

$$\Delta e_{\ell,m}=-\ell(\ell+1)e_{\ell,m}.$$

This theorem complements the properties given by (7) and (8) in the sense that it shows that the integrals of interest are small even if all the ℓ_j are of the same order. The only condition is that at least one of the eigenvectors appears with multiplicity 1 in the integral. Note that the decay property we obtain is not that small but it will be enough for our argument. We do not expect that the decay can be much increased except in higher dimensions where the denominator should be $\langle\ell_j\rangle^{\frac{d-1}{2}}$ rather than $\langle\ell_j\rangle^{\frac{1}{2}}$. We emphasize that, contrary to (7), this is not valid for any orthonormal basis but only for a generic one as (8) is. In order to prove this result, we will in fact refine the probabilistic approach used to prove (8).

Remark 2.2. As a corollary of the proof, we could also get a similar basis enjoying (9) for a countable set of degrees p and functions g (but not uniformly).

2.1. Probabilistic setup

We start with a short review on Haar measures which will be used to define natural probability measures on the orthogonal group of E_ℓ. Then we explain how to use these measures to define probability measures on $\mathscr{B}$ and how they are related to the normalized volume measure on the unit sphere S_ℓ of E_ℓ.

2.1.1. Background on Haar measures. Recall that, given a compact group G, there exists a Radon measure $\mathfrak{m}_G$ on G such that for every Borel subset $U \subset G$ and for every $g \in G$, $\mathfrak{m}_G(gU) = \mathfrak{m}_G(U)$ [32, Thm. 2.10]. This is called a (left-invariant) *Haar measure* on G and for any nonempty open set U, one has $\mathfrak{m}_G(U) > 0$ [32, Prop. 2.19]. Moreover, if we fix $\mathfrak{m}_G(G) = 1$, then this measure is unique [32, Thm. 2.20]. The main example we will use in the following is the orthogonal group $O(d)$ of $\mathbb{R}^d$ (with $d \in \mathbb{N}^*$) or more generally, the orthogonal group $O(E)$ of some Euclidean space E of dimension d.

Remark 2.3. For the sake of concreteness, let us give an explicit expression of $\mathfrak{m}_{O(d)}$ in terms of measures on spheres. Given an orthonormal family $(X_1, \dots, X_k)$ in $(\mathbb{R}^d)^k$, we denote by $\nu_{d-k-1}^{X_1,\dots,X_k}$ the normalized volume measure on $\mathbb{S}^{d-1} \cap \operatorname{Span}\{X_1, \dots, X_k\}^\perp$ induced by the Euclidean structure on $\mathbb{R}^{d-1}$. Equivalently,

$$\nu_{d-k-1} := \frac{\operatorname{vol}_{\mathbb{S}^{d-1}\cap \operatorname{Span}\{X_1,\dots,X_k\}^\perp}}{\operatorname{vol}_{\mathbb{S}^{d-k-1}}(\mathbb{S}^{d-k-1})}.$$

With these conventions at hand and writing $R = (X_1, \dots, X_d) \in O(d)$, one can verify using the invariance of ν_j by rotation that

$$\begin{aligned}&\int_{O(d)} f(R)\,\mathrm{d}\mathfrak{m}_{O(d)}(R)\\ &= \int_{(\mathbb{S}^{d-1})^d} f(X_1, \dots, X_d)\,\mathrm{d}\nu_0^{X_1,\dots,X_{d-1}}(X_d) \cdots \mathrm{d}\nu_{d-2}^{X_1}(X_2)\,\mathrm{d}\nu_{d-1}(X_1).\end{aligned}$$

In particular, if $f(R) = f(X_1, \dots, X_d) = g(X_1)$, then

$$\int_{O(d)} f(R)\,\mathrm{d}\mathfrak{m}_{O(d)}(R) = \int_{\mathbb{S}^{d-1}} g(X)\,\mathrm{d}\nu_{d-1}(X).$$

If we now fix some compact subgroup H of G, it also has a unique left-invariant probability measure $\mathfrak{m}_H$. This measure is naturally related to $\mathfrak{m}_G$ as follows. We define $G/H := \{[g] = gH : g \in G\}$ as the set of (left) cosets and according to [32, Thm. 2.51, Cor. 2.53], there exists some G-invariant measure $\mu_{G/H}$ such that, for every continuous function on G, one has

$$\int_G f(g)\,\mathrm{d}\mathfrak{m}_G(g) = \int_{G/H} \left(\int_H f(gh)\,\mathrm{d}\mathfrak{m}_H(h) \right) \mathrm{d}\mu_{G/H}([g]),$$

or more compactly

$$\mathfrak{m}_G = \int_{G/H} g_*(\mathfrak{m}_H)\, \mathrm{d}\mu_{G/H}([g]). \tag{11}$$

Remark 2.4. Again, we will use this disintegration of the measure in the case of the orthogonal group $G = O(E)$ and of a subgroup $H = O(V)$, where V is a linear subspace (with the same Euclidean structure) of E. Here, an element $R \in O(V)$ is identified with an element of $O(E)$ by letting $R|_{V^\perp} = \mathrm{Id}_{V^\perp}$.

2.1.2. Probability measures on an orthonormal basis. The measure $\mathfrak{m}_{O(E_\ell)}$ induces a probability measure $\mathbb{P}_\ell$ on the set $\mathcal{B}_\ell$ of orthonormal bases of E_ℓ through the map

$$R \in O(E_\ell) \mapsto (R\Phi_{(\ell,m)})_{-\ell\leq m\leq \ell},$$

where $(\Phi_{(\ell,m)})_{-\ell\leq m\leq \ell}$ is a fixed orthonormal basis of E_ℓ, e.g. the one given by the standard (real-valued) spherical harmonics. More generally, using the Kolmogorov extension theorem [44, Thm. 2.4.3], we define on the set $\mathcal{B}$ of orthonormal bases of Laplace eigenfunctions, the product measure

$$\mathbb{P} = \bigotimes_{\ell=0}^{+\infty} \mathbb{P}_\ell.$$

If we fix some (nonempty) subset L of $\mathbb{N}$, we can define the map

$$\pi_L : b = (b_\ell)_{\ell\in\mathbb{N}} \in \mathcal{B} \mapsto (b_\ell)_{\ell\in L} \in \mathcal{B}_L := \prod_{\ell\in L} \mathcal{B}_\ell.$$

The pushforward $\mathbb{P}_L := (\pi_L)_*\mathbb{P}$ is defined as

$$\int_{\mathcal{B}_L} f\, \mathrm{d}\mathbb{P}_L := \int_{\mathcal{B}} f \circ \pi_L\, \mathrm{d}\mathbb{P},$$

and it can be written as

$$\mathbb{P}_L = \bigotimes_{\ell\in L} \mathbb{P}_\ell.$$

Remark 2.5. When $L = \{\ell\}$, we just write $\mathbb{P}_{\{\ell\}} = \mathbb{P}_\ell$ as we did before. We will in fact mostly work with $\mathbb{P}_L$ for some finite set L.

We can also use the decomposition (11) in that context. For instance, one can fix a subset $\mathcal{M}$ of $\{-\ell, \ldots, \ell-1, \ell\}$ and define

$$V_{\ell,\mathcal{M}} := \mathrm{Span}\{\Phi_{\ell,m} : m \in \mathcal{M}\}.$$

Then, given an integrable function f on $\mathcal{B}_\ell$, one can write

$$\int_{\mathcal{B}_\ell} f(b_\ell)\, \mathrm{d}\mathbb{P}_\ell(b_\ell) \tag{12}$$
$$= \int_{O(E_\ell)/O(V_{\ell,\mathcal{M}})} \left(\int_{O(V_{\ell,\mathcal{M}})} f((RR_1\Phi_{\ell,m})_m)\, \mathrm{d}\mathfrak{m}_{O(V_{\ell,\mathcal{M}})}(R_1) \right) \mathrm{d}\mu_{O(E_\ell)/O(V_{\ell,\mathcal{M}})}([R]).$$

Remark 2.6. As $R_1\Phi_{\ell,m} = \Phi_{\ell,m}$ for $m \notin \mathcal{M}$ and for $R_1 \in O(V_{\ell,\mathcal{M}})$, the integral

$$\int_{O(V_{\ell,\mathcal{M}})} f((RR_1\Phi_{\ell,m})_m)\, \mathrm{d}\mathfrak{m}_{O(V_{\ell,\mathcal{M}})}(R_1)$$

can be identified with an integral on the set of orthonormal bases $\mathcal{B}_{\ell,\mathcal{M}}$ of $V_{\ell,\mathcal{M}}$ as we did above.

2.1.3. Induced measures on spheres. On the one hand, as we aim to find an orthonormal basis E_ℓ with good properties via probabilistic means, it is natural to work with the Haar measure on the corresponding orthogonal group $O(E_\ell)$. On the other hand, our main probabilistic ingredient will be a result on the concentration of the volume measure on spheres of large dimensions as the unit sphere S_ℓ of E_ℓ is when $\ell \to +\infty$. As already witnessed from Remark 2.3, the Haar measure is naturally related to such measures and, in view of our applications, we now make this connection slightly more precise in our context.

Fix $k = (\ell, m)$ in $\mathcal{T}_\infty$ and define the map

$$\pi_{(\ell,m)} : b_\ell = (e_{(\ell,m')})_{-\ell\leq m'\leq \ell} \in \mathcal{B}_\ell \mapsto e_{(\ell,m)} \in S_\ell,$$

where S_ℓ is the unit sphere (for the L^2-norm) in E_ℓ. The measure $\mathbb{P}_\ell$ induces a measure on the Euclidean sphere S_ℓ as follows:

$$\forall f \in \mathcal{C}^0(S_\ell), \quad \int_{S_\ell} f\, \mathrm{d}\nu_{2\ell} := \int_{\mathcal{B}_\ell} f \circ \pi_{(\ell,m)}\, \mathrm{d}\mathbb{P}_\ell. \tag{13}$$

By invariance of the Haar measure through orthogonal transformations, this measure does not depend on the choice of m. Still by definition of the Haar measure, one can also check that it is invariant under orthogonal transformations. Thus, by uniqueness of uniformly distributed measures on the sphere [40, Thm. 3.4], it can be identified with the *normalized* volume measure $\nu_{2\ell}$ on the 2ℓ-dimensional sphere $S_\ell \simeq \mathbb{S}^{2\ell}$ of $E_\ell \simeq \mathbb{R}^{2\ell+1}$.

Remark 2.7. In order to lighten the notation, rather than writing $\pi_{(\ell,m)} \circ \pi_\ell$, we will also denote by $\pi_{(\ell,m)}$ the map from $\mathcal{B}$ to S_ℓ that associates to $b = (e_{(\ell',m')})_{(\ell',m')\in\mathcal{T}_\infty}$ the eigenvector $e_{(\ell,m)}$. The induced measure on S_ℓ remains the same by construction.

2.2. The key probabilistic ingredient

The key ingredient in the proof of (8) and of our proof of Theorem 2.1 is the following property [39, Eq. 2.6]:

Theorem 2.8 (Levy's inequality). *Let $d \geq 1$ and let ν_d be the normalized volume measure on $\mathbb{S}^d$ induced by the Euclidean structure on $\mathbb{R}^{d+1}$. Let $F\colon \mathbb{S}^d \to \mathbb{R}$ be a continuous function. Then, for every $\delta > 0$,*

$$\nu_d\big(\big\{|F - m_F| \geq \omega_F(\delta)\big\}\big) \leq 2e^{-\delta^2 \frac{d-1}{2}},$$

where m_F *is a median of* F*, i.e. a real number such that*

$$\nu_d(\{F \geq m_F\}) \geq \frac{1}{2} \quad \textit{and} \quad \nu_d(\{F \leq m_F\}) \geq \frac{1}{2},$$

and where $\omega_F(\delta)$ *is the modulus of continuity of* F*:*

$$\omega_F(\delta) := \sup\big\{|F(u) - F(v)| : d_{\mathbb{S}^d}(u, v) \leq \delta\big\},$$

with $d_{\mathbb{S}^d}$ *the geodesic distance.*

In other words, this theorem states that functions with small oscillations on spheres of large dimensions are almost constant. Following [16, 41, 47], let us illustrate how to use this theorem when $F_q(u) := \|u\|_{L^q(\mathbb{S}^2)}$ with $2 \leq q < \infty$. Here u belongs to S_ℓ that we identify with $\mathbb{S}^{2\ell}$ by fixing some orthonormal basis $(\Phi_{(\ell,m)})_{-\ell\leq m\leq \ell}$ of E_ℓ. One has

$$\begin{aligned} |F_q(u) - F_q(v)| &\leq \|u - v\|_{L^q(\mathbb{S}^2)} \leq \|u - v\|_{L^2(\mathbb{S}^2)}^{\frac{2}{q}} \|u - v\|_{L^\infty(\mathbb{S}^2)}^{1-\frac{2}{q}} \\ &\leq \|u - v\|_{L^2(\mathbb{S}^2)}^{\frac{2}{q}} \left(\sup_{x\in\mathbb{S}^2} \left| \sum_{m=-\ell}^{\ell} \langle u - v, \Phi_{(\ell,m)}\rangle_{L^2} \Phi_{(\ell,m)}(x) \right| \right)^{1-\frac{2}{q}} \\ &\leq \|u - v\|_{L^2(\mathbb{S}^2)} \left(\sup_{x\in\mathbb{S}^2} \left\{ \sum_{m=-\ell}^{\ell} \Phi_{(\ell,m)}(x)^2 \right\} \right)^{\frac{1}{2}-\frac{1}{q}}. \end{aligned}$$

Now observing that the sum is the Schwartz kernel of the spectral projector $\mathbf{1}_{\ell(\ell+1)}(-\Delta)$ evaluated on the diagonal and that this is a spherical-invariant quantity, we deduce that these sums are independent of $x \in \mathbb{S}^2$ and thus equal to $2\ell + 1$. Hence, there exists some constant $c_0 > 0$ such that, for every $\ell \geq 1$ and for every $2 \leq q < \infty$,

$$|F_q(u) - F_q(v)| \leq \|u - v\|_{L^2(\mathbb{S}^2)} (2\ell + 1)^{\frac{1}{2}-\frac{1}{q}} \leq c_0 d_{\mathbb{S}^{2\ell}}(u, v)(2\ell + 1)^{\frac{1}{2}-\frac{1}{q}},$$

from which we infer the existence of $c_1 > 0$ (independent of ℓ and q) such that

$$\forall\delta > 0, \quad \nu_{2\ell}\big(\big\{u \in S_\ell : |\|u\|_{L^q} - m_{F_q}| \geq \delta\big\}\big) \leq 2e^{-c_1\delta^2\ell^{\frac{2}{q}}}.$$

Finally, the constant m_{F_q} can be estimated precisely through explicit calculations [16, Thm. 6]. For our purpose, we will only use the existence of a constant $c_2 > \sqrt{2}$ such that, for every $2 \leq q < \infty$, $1 \leq m_{F_q} \leq c_2\sqrt{q}$ [16, Thm. 4]. In particular, there exists a constant $c_1 > 0$ such that, for every $\Lambda \geq 2c_2\sqrt{q}$, for every $\ell \geq 1$ and for every $2 \leq q < \infty$, one has

$$\nu_{2\ell}\big(\big\{u \in S_\ell : \|u\|_{L^q} \geq \Lambda\big\}\big) \leq 2e^{-c_1(\Lambda - c_2\sqrt{q})^2\ell^{\frac{2}{q}}}. \tag{14}$$

This quantitative estimate will be useful in our construction of a good orthonormal basis. Yet, besides these already known results, we will also need to apply Levy's inequality one more time directly to the integrals we are interested in. In order to clarify the upcoming argument, let us give another simple application of Levy's inequality that will be in the spirit of our proof. We fix some $h \in L^2(\mathbb{S}^2)$ and we consider the map

$$F: u \in S_\ell \mapsto \int_{\mathbb{S}^2} u(x)h(x)\, \mathrm{dvol}_{\mathbb{S}^2}(x).$$

By symmetry, the median of this function is equal to 0 and one has, thanks to the Cauchy–Schwarz inequality,

$$|F(u) - F(v)| \le \|u - v\|_{L^2}\|h\|_{L^2} \le c_0\|h\|_{L^2} d_{\mathbb{S}^{2\ell}}(u, v).$$

Hence, we deduce from Levy's inequality applied with $\delta = \frac{\log\langle\ell\rangle}{\sqrt{\langle\ell\rangle}}$ that

$$\nu_{2\ell}\Big(\Big\{u \in S_\ell : |F(u)| \ge \frac{\log\langle\ell\rangle}{\sqrt{\langle\ell\rangle}}\Big\}\Big) \le 2e^{-c_1 \frac{\log^2\langle\ell\rangle}{\|h\|_{L^2}}}.$$

From that, we infer that

$$\sum_{k=(\ell,m)\in\mathcal{T}_\infty} \mathbb{P}\Big(\Big\{b \in \mathcal{B} : \Big|\int_{\mathbb{S}^2} e_k(x)h(x)\, \mathrm{dvol}_{\mathbb{S}^2}(x)\Big| \ge \frac{\log\langle\ell\rangle}{\sqrt{\langle\ell\rangle}}\Big\}\Big) \le 2\sum_{\ell\in\mathbb{N}}(2\ell+1)e^{-c_1 \frac{\log^2\langle\ell\rangle}{\|h\|_{L^2}}}$$
$$< \infty.$$

In particular, thanks to the Borel–Cantelli lemma, we can derive that, given $h \in L^2$ and for $\mathbb{P}$-a.e. $b \in \mathcal{B}$, there exists a constant $C_b > 0$ such that

$$\forall k \in \mathcal{T}_\infty, \quad \Big|\int_{\mathbb{S}^2} e_k(x)h(x)\, \mathrm{dvol}_{\mathbb{S}^2}(x)\Big| \le C_b \frac{\log(1+\langle\ell\rangle)}{\sqrt{\langle\ell\rangle}}.$$

This is exactly the kind of decay we are looking for in Theorem 2.1, except that h is a product of eigenfunctions inside b (rather than a fixed element h in L^2). In order to handle this problem, we will make use of the fact that most eigenfunctions have their L^q-norm uniformly bounded and that this control on the L^q-norm can be made quantitative thanks to (14). Due to the multiple and nested applications of Levy's inequality, this turns out to be a fairly tedious task. Yet the decay phenomenon we obtain is the same as the one we have just described in this elementary calculation.

2.3. Proof of Theorem 2.1

For the sake of simplicity, it is convenient to endow $\mathcal{T}_\infty$ with the lexicographic order, namely

$$k_1 = (\ell_1, m_1) \preccurlyeq k_2 = (\ell_2, m_2) \quad \Leftrightarrow \quad \ell_1 < \ell_2 \text{ or } (\ell_1 = \ell_2 \text{ and } m_1 \le m_2). \tag{15}$$

We will now estimate the probability that an orthonormal basis in $\mathcal{B}$ does not satisfy the conclusion of Theorem 2.1 for a fixed $\mathbf{k} = (k_1, \dots, k_p) \in \mathcal{T}_\infty^p$ with

$$k_1 = (\ell_1, m_1) \preccurlyeq \cdots \preccurlyeq k_p = (\ell_p, m_p). \tag{16}$$

Indeed, since the estimate of Theorem 2.1 is invariant by the action of the permutation group on $\mathbf{k}$, we can assume without loss of generality that $k_1, \dots, k_p$ are ordered.

In order to lighten the notation, we also define

$$A(\mathbf{k}) := \{k \in \mathcal{T}_\infty : \exists\, 1 \leq j \leq p \text{ such that } k = k_j\},$$

which is a set of cardinality $\leq p$ so that

$$\begin{aligned} F_{\mathbf{k}}(b) &:= \int_{\mathbb{S}^2} e_{k_1}(x) \cdots e_{k_p}(x) g(x)\, \mathrm{dvol}_{\mathbb{S}^2}(x) \\ &= \int_{\mathbb{S}^2} \prod_{k \in A(\mathbf{k})} e_k(x)^{\alpha_k} g(x)\, \mathrm{dvol}_{\mathbb{S}^2}(x), \end{aligned}$$

where $1 \leq \alpha_k \leq p$ for every $k \in A(\mathbf{k})$. We always suppose in the following that g is not identically 0.

2.3.1. Applying Levy's inequality. We suppose that there exists $1 \leq j_0 \leq p$ such that

$$(\ell_j, m_j) = (\ell_{j_0}, m_{j_0}) \quad \Rightarrow \quad j = j_0.$$

In that case, we say that $\mathbf{k}$ satisfies property (S). We denote by j_+ the largest index in $\{1, \dots, p\}$ with this property. In particular, $\alpha_{(\ell_{j_+}, m_{j_+})} = 1$. We begin by treating the case of multi-indices verifying (S) and we also suppose for the moment that $\ell_{j_+} \geq p$.

Following the above calculation, we aim to apply Levy's inequality to the map

$$F_+ \colon e_{(\ell_{j_+}, m_{j_+})} \in S_{\ell_{j_+}} \mapsto \int_{\mathbb{S}^2} e_{k_1}(x) \cdots e_{k_p}(x) g(x)\, \mathrm{dvol}_{\mathbb{S}^2}(x),$$

with $(e_{k_j})_{1 \leq j \neq j_+ \leq p}$ fixed. By symmetry, the median m_{F_+} of F_+ is equal to 0. Moreover, by the Hölder inequality, this is a Lipschitz map:

$$\begin{aligned} |F_+(u) - F_+(v)| &\leq \|g\|_{L^\infty} \|u - v\|_{L^2} \left(\int_{\mathbb{S}^2} \prod_{j \neq j_+} |e_{k_j}(x)|^2\, \mathrm{dvol}_{\mathbb{S}^2}(x) \right)^{\frac{1}{2}} \\ &\leq c_0 \|g\|_{L^\infty} d_{\mathbb{S}^2}(u, v) \left(\int_{\mathbb{S}^2} \prod_{j \neq j_+} |e_{k_j}(x)|^2\, \mathrm{dvol}_{\mathbb{S}^2}(x) \right)^{\frac{1}{2}} \\ &\leq c_0 \|g\|_{L^\infty} d_{\mathbb{S}^2}(u, v) \prod_{k \in A(\mathbf{k}) \setminus \{k_{j_+}\}} \|e_k\|_{L^{2(p-1)}}^{\alpha_k}. \end{aligned}$$

Remark 2.9. Note that these two properties also hold true for $F_+ \circ R$ where $R \in O(E_{\ell_{j_+}})$.

In order to apply Levy's inequality, we would at least need that the L^{2p-2}-norms appearing in the Lipschitz constant are uniformly bounded. To that aim, we set, for $\Lambda > 0$,

$$B_\Lambda(\mathbf{k}) := \{b \in \mathcal{B} : \forall k \in A(\mathbf{k}) \setminus \{k_{j_+}\},\ \|\pi_k(b)\|_{L^{2(p-1)}} \leq \Lambda\}.$$

In particular, for $b \in B_\Lambda(\mathbf{k})$, the Lipschitz constant of F_+ is bounded by $c_0\|g\|_{L^\infty}\Lambda^{p-1}$. Moreover, using (14), one finds that the complementary set of $B_\Lambda(\mathbf{k})$ is small. More precisely, for $\Lambda \geq 4c_2\sqrt{p}$, one has

$$\begin{aligned}\mathbb{P}(B_\Lambda(\mathbf{k})^c) &\leq \sum_{k\in A(\mathbf{k})\setminus\{k_{j_+}\}} \mathbb{P}\big(\{b \in \mathcal{B} : \|\pi_k(b)\|_{L^{2p-2}(\mathbb{S}^2)} \geq \Lambda\}\big)\\ &\leq 2\sum_{j=1,j\neq j_+}^{p} e^{-c_1(\Lambda-2c_2\sqrt{p})^2\ell_j^{\frac{1}{p-1}}}.\end{aligned}$$

Now fix some $\Lambda \geq 4c_2\sqrt{p}$ and some $\delta > 0$. For $L \subset \mathbb{N}$, we set

$$B_{\Lambda,L}(\mathbf{k}) := \big\{b \in \mathcal{B}_L : \forall k \in L \times \mathbb{Z} \cap (A(\mathbf{k}) \setminus \{k_{j_+}\}),\ \|\pi_k(b)\|_{L^{2(p-1)}} \leq \Lambda\big\}$$

so that we can write

$$\begin{aligned}&\mathbb{P}\big(\{b \in \mathcal{B} : |F_\mathbf{k}(b)| \geq \delta\}\big)\\ &\quad\leq \mathbb{P}\big(\{b \in B_{\Lambda,\mathbb{N}}(\mathbf{k}) : |F_\mathbf{k}(b)| \geq \delta\}\big) + 2\sum_{j=1,j\neq j_+}^{p} e^{-c_1(\Lambda-2c_2\sqrt{p})^2\ell_j^{\frac{1}{p-1}}}\\ &\quad\leq \int_{B_{\Lambda,\mathbb{N}\setminus\{\ell_{j_+}\}}(\mathbf{k})} \mathbb{P}_{\ell_{j_+}}\big(\{b_{\ell_{j_+}} \in B_{\Lambda,\ell_{j_+}}(\mathbf{k}) : |F_\mathbf{k}(b', b_{\ell_{j_+}})| \geq \delta\}\big)\,\mathrm{d}\mathbb{P}_{\mathbb{N}\setminus\{\ell_{j_+}\}}(b')\\ &\qquad+ 2\sum_{j=1,j\neq j_+}^{p} e^{-c_1(\Lambda-2c_2\sqrt{p})^2\ell_j^{\frac{1}{p-1}}}.\end{aligned}$$

Hence, b' being fixed in $\mathcal{B}_{\mathbb{N}\setminus\{\ell_{j_+}\}}$, we are left with estimating, uniformly for $b' \in B_{\Lambda,\mathbb{N}\setminus\{\ell_{j_+}\}}(\mathbf{k})$,

$$\mathbb{P}_{\ell_{j_+}}\big(\{b_{\ell_{j_+}} \in B_{\Lambda,\ell_{j_+}}(\mathbf{k}) : |F_\mathbf{k}(b', b_{\ell_{j_+}})| \geq \delta\}\big), \tag{17}$$

which can be analyzed using (12). Expressed in terms of the orthogonal group of $E_{\ell_{j_+}}$, (17) can in fact be rewritten as

$$\mathfrak{m}_{O(E_{\ell_{j_+}})}\big(\{R : (R\Phi_{\ell_{j_+},m})_m \in B_{\Lambda,\ell_{j_+}}(\mathbf{k}) \text{ and } |F_\mathbf{k}(b', (R\Phi_{\ell_{j_+},m})_m)| \geq \delta\}\big). \tag{18}$$

We are now exactly in position to apply the disintegration formula (12) with $\ell = \ell_{j_+}$, $m_+ = m_{j_+}$ and

$$\mathcal{M} = \big\{k = (\ell, m) \notin A(\mathbf{k}) : \ell = \ell_{j_+}\big\} \cup \{(\ell_{j_+}, m_{j_+})\},$$

where we note that $|\mathcal{M}| \geq 2(\ell_{j_+} + 1) - p$. From this and as the condition on $B_{\Lambda,\ell_{j_+}}(\mathbf{k})$ only concerns indices m not belonging to $\mathcal{M}$, we infer that (18) (and thus (17)) can be rewritten as

$$\begin{aligned}&\int_{O(E_{\ell_{j_+}})/O(V_{\ell_{j_+},\mathcal{M}})} \mathbb{1}_{\{[R]:(R\Phi_{\ell_{j_+},m})_m\in B_{\Lambda,\ell_{j_+}}(\mathbf{k})\}}([R]) \qquad (19)\\ &\quad\times \mathfrak{m}_{O(V_{\ell_{j_+},\mathcal{M}})}\big(\{R_1 : |F_\mathbf{k}(b', (RR_1\Phi_{\ell_{j_+},m})_m)| \geq \delta\}\big)\,\mathrm{d}\mu_{O(E_{\ell_{j_+}})/O(V_{\ell_{j_+},\mathcal{M}})}([R]).\end{aligned}$$

In order to estimate (17) and thus $\mathbb{P}(\{b \in \mathcal{B} : |F_{\mathbf{k}}(b)| \geq \delta\})$, we are left with determining an upper bound on

$$\mathfrak{m}_{O(V_{\ell_{j_+}},\mathcal{M})}\big(\{R_1 : |F_{\mathbf{k}}(b', (RR_1\Phi_{\ell_{j_+},m})_m)| \geq \delta\}\big),$$

uniformly for $b' \in B_{\Lambda,\mathbb{N}\backslash\{\ell_{j_+}\}}(\mathbf{k})$ and for $[R]$ such that $(R\Phi_{\ell_{j_+},m})_m \in B_{\Lambda,\ell_{j_+}}(\mathbf{k})$. Equivalently, as in (13), one gets in terms of measures on spheres,

$$\begin{aligned}\mathfrak{m}_{O(V_{\ell_{j_+}},\mathcal{M})}&\big(\{R_1 : |F_{\mathbf{k}}(b', (RR_1\Phi_{\ell_{j_+},m})_m)| \geq \delta\}\big)\\ &= \nu_{|\mathcal{M}|-1}\big(\{u \in \mathbb{S}^{|\mathcal{M}|-1} : |F_+(Ru)| \geq \delta\}\big),\end{aligned}$$

where R is a fixed element in $E_{\ell_{j_+}}$ and where the function F_+ is defined using a fixed orthonormal family $\{e_{k_j} : 1 \leq j \neq j_+ \leq p\}$ verifying $\|e_{k_j}\|_{L^{2(p-1)}} \leq \Lambda$ for every $j \neq j_+$. Hence, using Levy's inequality and recalling from Remark 2.9 that $F_+ \circ R$ is Lipschitz and that its median is 0, we obtain

$$\mathfrak{m}_{O(V_{\ell_{j_+}},\mathcal{M})}\big(\{R_1 : |F_{\mathbf{k}}(b', (RR_1\Phi_{\ell_{j_+},m})_m)| \geq \delta\}\big) \leq 2e^{-\delta^2 \frac{|\mathcal{M}|-2}{c_0^2\|g\|_{L^\infty}^2 \Lambda^{2p-2}}}.$$

Gathering these bounds, we get

$$\mathbb{P}\big(\{b \in \mathcal{B} : |F_{\mathbf{k}}(b)| \geq \delta\}\big) \leq 2e^{-\delta^2 \frac{|\mathcal{M}|-2}{c_0^2\|g\|_{L^\infty}^2 \Lambda^{2p-2}}} + 2\sum_{j=1, j\neq j_+}^{p} e^{-c_1(\Lambda - 2c_2\sqrt{p})^2 \ell_j^{\frac{1}{p-1}}}.$$

Note that, for $\ell_{j_+} \geq p$, one has $|\mathcal{M}| - 2 \geq 2\ell_{j_+} - p \geq \ell_{j_+}$.

In summary, we end up with the existence of two positive constants $c_1, c_2 > 0$ (depending only on g, on p and on the geometry of $\mathbb{S}^2$) such that, for every $\delta > 0$ and for every $\Lambda \geq 4c_2\sqrt{p}$,

$$\mathbb{P}\big(\{b \in \mathcal{B} : |F_{\mathbf{k}}(b)| \geq \delta\}\big) \leq 2e^{-c_1 \frac{\delta^2 \ell_{j_+}}{\Lambda^{2p-2}}} + 2\sum_{j=1, j\neq j_+}^{p} e^{-c_1(\Lambda - 2c_2\sqrt{p})^2 \ell_j^{\frac{1}{p-1}}}, \tag{20}$$

whenever $\mathbf{k}$ verifies (S) and $\ell_{j_+} \geq p$. Taking $\Lambda = \log\langle \ell_p \rangle$ (and thus ℓ_p large enough), we can deduce the existence of a constant $c_{p,g} \geq 1$ such that, for every $\delta > 0$ and for every $\mathbf{k} \in \mathcal{T}_\infty^p$ with $k_1 \preccurlyeq \cdots \preccurlyeq k_p = (\ell_p, m_p)$ verifying (S),

$$\mathbb{P}\big(\{b \in \mathcal{B} : |F_{\mathbf{k}}(b)| \geq \delta\}\big) \leq c_{p,g} e^{-c_{p,g}^{-1} \frac{\delta^2 \langle \ell_{j_+} \rangle}{\log^{2(p-1)}\langle \ell_p \rangle}} + c_{p,g} e^{-c_{p,g}^{-1} \log^2\langle \ell_p \rangle}.$$

Thus, we obtain

$$\mathbb{P}\left(\left\{b \in \mathcal{B} : |F_{\mathbf{k}}(b)| \geq \frac{\log^p\langle \ell_p \rangle}{\sqrt{\langle \ell_{j_+} \rangle}}\right\}\right) \leq 2c_{p,g} e^{-c_{p,g}^{-1} \log^2\langle \ell_p \rangle}. \tag{21}$$

2.3.2. The conclusion. Given $\mathbf{k} \in \mathcal{T}_\infty^p$ with $k_1 \preccurlyeq \cdots \preccurlyeq k_p = (\ell_p, m_p)$ verifying property (S) and $\ell_{j+} \geq p$, we define the following probabilistic events:

$$\Omega(\mathbf{k}) := \left\{ b \in \mathcal{B} : |F_{\mathbf{k}}(b)| \geq \frac{\log^p \langle \ell_p \rangle}{\sqrt{\langle \ell_{j+} \rangle}} \right\}.$$

Applying (21), one has

$$\sum_{k_1 \preccurlyeq \cdots \preccurlyeq k_r : (S) \text{ holds and } \ell_{j+} \geq p} \mathbb{P}(\Omega(\mathbf{k})) \leq C_{p,g} \sum_{\ell=1}^{+\infty} \ell^{2p} e^{-C_{p,g}^{-1} \log^2 \langle \ell \rangle} < \infty.$$

In particular, thanks to the Borel–Cantelli lemma, we can conclude that, for $\mathbb{P}$-a.e. $b \in \mathcal{B}$, one has $b \in \Omega(\mathbf{k})^c$ except for finitely many $\mathbf{k}$ verifying (S) and $\ell_{j+} \geq p$. This yields the conclusion of the theorem for indices verifying these two properties. Recall now from[2] [16, Thm. 6] that, for $\mathbb{P}$-a.e. $b \in \mathcal{B}$, there exists a constant $C_b > 0$ such that, for every $\mathbf{k} = (k_1, \ldots, k_p) \in \mathcal{T}_\infty^p$,

$$\left| \int_{\mathbb{S}^2} e_{k_1}(x) \cdots e_{k_p}(x) g(x) \, \mathrm{dvol}_{\mathbb{S}^2}(x) \right| \leq C_b.$$

This last inequality yields the conclusion of the theorem whenever $\mathbf{k}$ does not satisfy (S) or $\ell_{j+} \geq p$. Hence, taking an element in the intersection of these two subsets of full measure concludes the proof of Theorem 2.1.

Remark 2.10. We note that we proved something slightly stronger than what was stated in Theorem 2.1 as the conclusion holds true for $\mathbb{P}$-a.e. orthonormal basis in $\mathcal{B}$ (with a constant that depends on the choice of b).

3. A good mass

In this section we prove that, for almost all mass $\mu > 0$, the frequencies of (KG) are nonresonant and thus well suited to proceed to a Birkhoff normal form reduction. The frequencies of (KG) are defined by

$$\forall k = (\ell, m) \in \mathcal{T}_\infty, \quad \omega_k := \sqrt{\ell(\ell+1) + \mu}. \tag{22}$$

They are the eigenvalues of the operator $\sqrt{\mu - \Delta}$ (see (2)).

The Birkhoff normal form process involves small divisors of the form

$$\Omega(\sigma, \mathbf{k}) = \sigma_1 \omega_{k_1} + \cdots + \sigma_r \omega_{k_r}, \tag{23}$$

---

[2] This is in fact a rather direct consequence of (14) combined with Hölder's inequality and the Borel–Cantelli lemma.

with $r \geq 3$, $\sigma \in \{-1,1\}^r$ and $\mathbf{k} \in \mathcal{T}_\infty^r$. Of course there may be cancellations in these small divisors (the same term could appear both with a plus sign and with a minus sign). Therefore it is useful to define the *smallest effective index* by

$$\kappa(\sigma,\mathbf{k}) = \min\left\{\langle \ell_j \rangle \mid 1 \leq j \leq r \text{ and } \textstyle\sum_{\ell_i = \ell_j} \sigma_i \neq 0\right\} \cup \{+\infty\}, \tag{24}$$

where, for all $i \in [\![1,r]\!]$, we have set $(\ell_i, m_i) := k_i$. The following proposition provides a quite uniform lower bound for the small divisors of (KG).

Proposition 3.1. *For almost all $\mu > 0$ and all $r \geq 2$, there exist $\gamma_r, \alpha_r > 0$ such that for all $\mathbf{k} \in \mathcal{T}_\infty^r$, all $\sigma \in \{-1,1\}^r$, we have either*

$$|\Omega(\sigma,\mathbf{k})| \geq \gamma_r \kappa(\sigma,\mathbf{k})^{-\alpha_r} \tag{25}$$

or $\kappa(\sigma,\mathbf{k}) = +\infty$, i.e. r is even and there exists ρ in the symmetric group $\mathfrak{S}_r$ such that

$$\forall j \in [\![1, r/2]\!], \quad \sigma_{\rho_{2j-1}} = -\sigma_{\rho_{2j}} \text{ and } \omega_{k_{\rho_{2j-1}}} = \omega_{k_{\rho_{2j}}}.$$

Moreover, α_r does not depend on μ.

As already explained in the introduction, the key observation here is that the small divisors that will appear in our normal formal procedure (see the proof of Theorem 5.1) are controlled by the smallest effective index rather than the third largest index as for instance in [3, Prop. 3.16]. This will allow us to remove many more terms when solving cohomological equations.

Proof of Proposition 3.1 . First we note that the frequencies accumulate polynomially fast on lattice $\mathbb{Z} + \frac{1}{2}$:

$$\omega_{(\ell,m)} = \sqrt{\ell(\ell+1)+\mu} = \ell\sqrt{1+\frac{1}{\ell}+\frac{\mu}{\ell^2}} \underset{\ell\to+\infty}{=} \ell + \frac{1}{2} + \mathcal{O}\left(\frac{1}{\ell}\right).$$

Moreover, it is well known (see e.g. [25, Prop. 4.8] and [2, Thm. 6.5]) that Proposition 3.1 holds if (25) is replaced by the weaker estimate

$$\forall y \in \mathbb{Z}, \quad \left|\frac{y}{2} + \Omega(\sigma,\mathbf{k})\right| \geq \gamma_r \left(\max_{j=1}^{r} \langle k_j \rangle\right)^{-\alpha_r}.$$

Therefore, Proposition 3.1 is a consequence of [8, Prop. 2.1, p. 11] which only requires the two above ingredients. ∎

4. Hamiltonian formalism

We now introduce new families of norms on real-valued and homogeneous polynomials on $\mathbb{C}^{\mathcal{T}_M}$ that are well behaved with respect to the canonical symplectic structure on $\mathbb{C}^{\mathcal{T}_M}$ and thus well adapted to our initial PDE problem after diagonalization of Δ.

4.1. Functional setting

We use the standard functional setting to deal with Hamiltonian systems. Nevertheless, to avoid any possible confusion we recall it precisely (and we refer to [8, Sect. 3.1] for more comments and details).

We consider $M \in (0,\infty)$ as a fixed parameter and we note that $\mathbb{C}^{\mathcal{T}_M}$ is a real finite-dimensional vector space. We always consider this space as a Euclidean space for the ℓ^2 scalar product

$$\forall u, v \in \mathbb{C}^{\mathcal{T}_M}, \quad (u,v)_{\ell^2} := \Re \sum_{k\in\mathcal{T}_M} u_k \overline{v_k}.$$

As a consequence, if $H\colon \mathbb{C}^{\mathcal{T}_M} \to \mathbb{R}$, we have the relation

$$\forall k \in \mathcal{T}_M, \quad \frac{(\nabla H)_k}{2} = \partial_{\overline{u_k}} H =: \frac{1}{2}(\partial_{\Re u_k} H + i\partial_{\Im u_k} H).$$

As usual, we implicitly equip $\mathbb{C}^{\mathcal{T}_M}$ with the symplectic form $(i\cdot,\cdot)_{\ell^2}$. Therefore, a smooth map $\tau\colon \mathcal{D} \to \mathbb{C}^{\mathcal{T}_M}$, where $\mathcal{D}$ is an open set of $\mathbb{C}^{\mathcal{T}_M}$, is *symplectic* if

$$\forall u \in \mathcal{D},\ \forall v, w \in \mathbb{C}^{\mathcal{T}_M}, \quad (iv, w)_{\ell^2} = (i\,\mathrm{d}\tau(u)(v), \mathrm{d}\tau(u)(w))_{\ell^2}.$$

Moreover, if $H, K\colon \mathbb{C}^{\mathcal{T}_M} \to \mathbb{R}$ are two smooth functions, the *Poisson bracket* of H and K is defined by

$$\{H, K\}(u) := (i\nabla H(u), \nabla K(u))_{\ell^2}.$$

Note that, as usual, it can be checked that we have

$$\{H, K\} = \sum_{k\in\mathcal{T}_M} \partial_{\Re u_k} H \partial_{\Im u_k} K - \partial_{\Im u_k} H \partial_{\Re u_k} K = 2i \sum_{k\in\mathcal{T}_M} \partial_{\overline{u_k}} H \partial_{u_k} K - \partial_{u_k} H \partial_{\overline{u_k}} K.$$

For all $s \in \mathbb{R}$, we define the h^s-norm on $\mathbb{C}^{\mathcal{T}_M}$ by

$$\forall u \in \mathbb{C}^{\mathcal{T}_M}, \quad \|u\|_{h^s}^2 := \sum_{k=(\ell,m)\in\mathcal{T}_M} \langle \ell \rangle^{2s} |u_k|^2.$$

4.2. Multilinear estimates

In this paragraph we establish multilinear estimates for Hamiltonians which are homogeneous polynomials on $\mathbb{C}^{\mathcal{T}_M}$.

Definition 4.1 (Space $\mathcal{H}_M^r$). Given $M \geq 0$ and $r \geq 2$, $\mathcal{H}_M^r$ denotes the space of real-valued homogeneous polynomials of degree r on the real vector space $\mathbb{C}^{\mathcal{T}_M}$.

Remark 4.2. By definition, every homogeneous polynomial $H \in \mathcal{H}_M^r$ admits a unique decomposition of the form

$$H(u) = \sum_{\sigma\in\{-1,1\}^r} \sum_{\mathbf{k}\in\mathcal{T}_M^r} H_{\mathbf{k}}^{\sigma} u_{k_1}^{\sigma_1} \cdots u_{k_r}^{\sigma_r},$$

where $(H^{\sigma}_{\mathbf{k}})_{(\mathbf{k},\sigma)\in\mathcal{T}_M^r\times\{-1,1\}^r}$ is a sequence of complex numbers satisfying the reality condition

$$H^{-\sigma}_{\mathbf{k}} = \overline{H^{\sigma}_{\mathbf{k}}} \tag{26}$$

and the symmetry condition

$$\forall\phi\in\mathfrak{S}_r,\quad H^{\sigma_1,\dots,\sigma_r}_{k_1,\dots,k_r} = H^{\sigma_{\phi_1},\dots,\sigma_{\phi_r}}_{k_{\phi_1},\dots,k_{\phi_r}}. \tag{27}$$

We endow this space of polynomials with two unusual norms $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{C}}$. Roughly speaking, in our Birkhoff normal form process, the terms of the Taylor expansion of the Hamiltonian are controlled with the $\mathcal{H}$-norm, whereas the solutions to cohomological equations are controlled with a $\mathcal{C}$-norm (because they enjoy better properties).

Definition 4.3 (Norms $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{C}}$). Let $M\geq 0$, $r\geq 2$ and $H,\chi\in\mathcal{H}^r_M$; we set

$$\|H\|_{\mathcal{H}} := \max_{\sigma\in\{-1,1\}^r}\ \max_{\mathbf{k}\in\mathcal{T}_M^r} |H^{\sigma}_{\mathbf{k}}|\sqrt{\langle\ell_1\rangle\cdots\langle\ell_r\rangle}\sqrt{\Upsilon(\mathbf{k})} \tag{28}$$

and

$$\|\chi\|_{\mathcal{C}} := \max_{\sigma\in\{-1,1\}^r}\ \max_{\mathbf{k}\in\mathcal{T}_M^r} |\chi^{\sigma}_{\mathbf{k}}|\langle\sigma_1\ell_1+\dots+\sigma_r\ell_r\rangle\sqrt{\langle\ell_1\rangle\cdots\langle\ell_r\rangle}\sqrt{\Upsilon(\mathbf{k})}, \tag{29}$$

where $k_j =: (\ell_j, m_j)$ for all $j\in[\![1,r]\!]$ and Υ is defined by (10).

As we will see in this section, these nonstandard norms are well behaved with the symplectic operations (Poisson bracket, gradient) that are used when performing a Birkhoff normal form procedure in Theorem 5.1. One reason for these nice properties is the fact that they involve an extra regularity factor $\Upsilon(\mathbf{k})$ which only depends on the largest simple index $k_j = (\ell_j, m_j)$ of $\mathbf{k}$. Despite their unusual definition, these norms can be implemented in our normal form argument as this exponent appears naturally in the multilinear estimate of Theorem 2.1. See for instance (68) below.

Let us now turn to the nice properties enjoyed by these norms. They provide the following continuity estimate for the Poisson bracket:

Proposition 4.4. *Let $r,r'\geq 2$ and $M\geq 2$. For all $H\in\mathcal{H}^{r'}_M$ and all $\chi\in\mathcal{H}^r_M$, their Poisson bracket $\{\chi,H\}$ is a homogeneous polynomial of degree $r+r'-2$ (i.e. $\{\chi,H\}\in\mathcal{H}^{r+r'-2}_M$) enjoying the bound*

$$\|\{\chi,H\}\|_{\mathcal{H}} \lesssim_{r,r'} \log M\,\|H\|_{\mathcal{H}}\|\chi\|_{\mathcal{C}}.$$

Proof. By definition of the Poisson bracket, we have

$$\{\chi,H\}(u) = 2i\sum_{\mathfrak{K}\in\mathcal{T}_M} \partial_{\bar{u}_{\mathfrak{K}}}\chi(u)\partial_{u_{\mathfrak{K}}}H(u) - \partial_{u_{\mathfrak{K}}}\chi(u)\partial_{\bar{u}_{\mathfrak{K}}}H(u). \tag{30}$$

Since the coefficients of H and K are symmetric (i.e. satisfy (27)), we have

$$\partial_{\bar{u}_{\mathfrak{K}}}\chi\partial_{u_{\mathfrak{K}}}H = rr' \sum_{\substack{\sigma\in\{-1,1\}^{r-1}\\ \sigma'\in\{-1,1\}^{r'-1}}}\ \sum_{\substack{\mathbf{k}\in\mathcal{T}_M^{r-1}\\ \mathbf{k}'\in\mathcal{T}_M^{r'-1}}} \chi^{\sigma,-1}_{\mathbf{k},\mathfrak{K}} u^{\sigma_1}_{k_1}\cdots u^{\sigma_{r-1}}_{k_{r-1}} H^{\sigma',1}_{\mathbf{k}',\mathfrak{K}} u^{\sigma'_1}_{k'_1}\cdots u^{\sigma'_{r'-1}}_{k'_{r'-1}}. \tag{31}$$

Obviously, $\{\chi, H\}$ defines a homogeneous polynomial of degree $r + r' - 2$. Hence, we need to verify the reality condition (26) and the upper bound on the $\mathcal{H}$-norm. For the latter, we begin by estimating $\sum_{\mathfrak{K}} \chi^{\sigma,-1}_{\mathbf{k},\mathfrak{K}} H^{\sigma',1}_{\mathbf{k}',\mathfrak{K}}$. By (28) and (29), denoting $\mathbf{k} \in \mathcal{T}_M^{r-1}$, $\mathbf{k}' \in \mathcal{T}_M^{r'-1}$, $\mathbf{k}'' = (\mathbf{k}, \mathbf{k}')$ and $r'' = r + r' - 2$, we have

$$\sum_{\mathfrak{K}\in\mathcal{T}_M} |\chi^{\sigma,-1}_{\mathbf{k},\mathfrak{K}} H^{\sigma',1}_{\mathbf{k}',\mathfrak{K}}| \le \frac{\|H\|_{\mathcal{H}} \|\chi\|_{\mathcal{C}}}{\sqrt{\langle \ell_1\rangle \cdots \langle \ell_{r-1}\rangle \langle \ell'_1\rangle \cdots \langle \ell'_{r'-1}\rangle}} \times \sum_{\mathfrak{K}=(\mathfrak{l},\mathfrak{m})\in\mathcal{T}_M} \frac{1}{\langle \mathfrak{l}\rangle \langle \sigma_1\ell_1 + \cdots + \sigma_{r-1}\ell_{r-1} - \mathfrak{l}\rangle \sqrt{\Upsilon(\mathbf{k},\mathfrak{K})\Upsilon(\mathbf{k}',\mathfrak{K})}}. \tag{32}$$

We claim that for all $\mathfrak{K} \in \mathcal{T}_M$ we have

$$\Upsilon(\mathbf{k}, \mathbf{k}') \le \Upsilon(\mathbf{k}, \mathfrak{K})\Upsilon(\mathbf{k}', \mathfrak{K}). \tag{33}$$

Indeed, if $\Upsilon(\mathbf{k}, \mathbf{k}') = 1$ the inequality is trivial so we can assume that

- either there exists $1 \le i \le r - 1$ such that $\Upsilon(\mathbf{k}, \mathbf{k}') = \langle \ell_i\rangle$, $k_j \ne k_i$ for $1 \le j \le r - 1$ with $j \ne i$ and $k'_{j'} \ne k_i$ for $1 \le j' \le r' - 1$,
- or there exists $1 \le i' \le r' - 1$ such that $\Upsilon(\mathbf{k}, \mathbf{k}') = \langle \ell'_{i'}\rangle$, $k'_{j'} \ne k_{i'}$ for $1 \le j' \le r' - 1$ with $j' \ne i'$ and $k_j \ne k'_{i'}$ for $1 \le j \le r - 1$.

By symmetry of the problem, let us assume the former and let $\mathfrak{K} = (\mathfrak{l}, \mathfrak{m}) \in \mathcal{T}_M$.

If $\Upsilon(\mathbf{k}, \mathfrak{K}) \ge \langle \ell_i\rangle = \Upsilon(\mathbf{k}, \mathbf{k}')$ then (33) holds true trivially. So let us assume that $\Upsilon(\mathbf{k}, \mathfrak{K}) < \langle \ell_i\rangle$. This implies that $\mathfrak{K} = k_i$ (if not $\Upsilon(\mathbf{k}, \mathfrak{K})$ is the maximum of a list of numbers including $\langle \ell_i\rangle$). But then, if $\Upsilon(\mathbf{k}', \mathfrak{K}) \ge \langle \mathfrak{l}\rangle$, we deduce $\Upsilon(\mathbf{k}', \mathfrak{K}) \ge \langle \ell_i\rangle = \Upsilon(\mathbf{k}, \mathbf{k}')$, which in turn implies (33). Thus it remains to consider the case $\Upsilon(\mathbf{k}', \mathfrak{K}) < \langle \mathfrak{l}\rangle$, which leads to the existence of $1 \le j' \le r' - 1$ such that $k_{j'} = \mathfrak{K}$ (if not $\Upsilon(\mathbf{k}', \mathfrak{K})$ is the maximum of a list of numbers including $\langle \mathfrak{l}\rangle$). Therefore $k_i = k_{j'}$ which contradicts the definition of i.

Implementing (33) in (32) and denoting $a = \sigma_1\ell_1 + \cdots + \sigma_{r-1}\ell_{r-1}$, one is left with estimating

$$\begin{aligned}\sum_{\mathfrak{K}=(\mathfrak{l},\mathfrak{m})\in\mathcal{T}_M} \frac{1}{\langle \mathfrak{l}\rangle \langle \sigma_1\ell_1 + \cdots + \sigma_{r-1}\ell_{r-1} - \mathfrak{l}\rangle} &\le 4 \sum_{\mathfrak{l}=0}^{M} \frac{1}{\sqrt{1 + (a - \mathfrak{l})^2}} \\ &\le 4 \sum_{j=-a}^{M-a} \frac{1}{\sqrt{1 + j^2}} \\ &\le 8 \sum_{j=0}^{M} \frac{1}{\sqrt{1 + j^2}} \lesssim \log M,\end{aligned} \tag{34}$$

independently of the value of a.

Inserting (33) and (34) in (32), we get uniformly with respect to σ, σ', k, k',

$$\sum_{\mathfrak{K}\in\mathcal{T}_M} |\chi^{\sigma,-1}_{\mathbf{k},\mathfrak{K}} H^{\sigma',1}_{\mathbf{k}',\mathfrak{K}}| \lesssim \log M \frac{\|H\|_{\mathcal{H}} \|\chi\|_{\mathcal{C}}}{\sqrt{\Upsilon(\mathbf{k},\mathbf{k}')} \sqrt{\langle \ell_1\rangle \cdots \langle \ell_{r-1}\rangle \langle \ell'_1\rangle \cdots \langle \ell'_{r'-1}\rangle}}. \tag{35}$$

Then, denoting $r'' = r + r' - 2$, $\mathbf{k}'' = (\mathbf{k}, \mathbf{k}')$ and $\sigma'' = (\sigma, \sigma')$, we define

$$M_{\mathbf{k}''}^{\sigma''} := 2irr' \sum_{\mathfrak{K} \in \mathcal{T}_M} \chi_{\mathbf{k},\mathfrak{K}}^{\sigma,-1} H_{\mathbf{k}',\mathfrak{K}}^{\sigma',1} - \chi_{\mathbf{k},\mathfrak{K}}^{\sigma,1} H_{\mathbf{k}',\mathfrak{K}}^{\sigma',-1} \quad \text{and} \quad P_{\mathbf{k}''}^{\sigma''} = \frac{1}{r''!} \sum_{\rho \in \mathfrak{S}_{r''}} M_{\mathbf{k}'' \circ \rho}^{\sigma'' \circ \rho}.$$

By definition, $P(u) = \{\chi, H\}(u)$ and the estimate (35) proves that

$$\|P\|_{\mathcal{H}} \lesssim rr' \log M \|H\|_{\mathcal{H}} \|\chi\|_{\mathcal{C}}.$$

Finally, the coefficients of P are obviously symmetric and, by a direct calculation, we verify that they satisfy the reality condition (26). ■

We now study the vector field on $\mathbb{C}^{\mathcal{T}_M}$ associated with a Hamiltonian in $\mathcal{H}_M^r$.

Lemma 4.5. *Let $M \geq 2$ and $r \geq 2$. For all $H \in \mathcal{H}_M^r$, H is a real-valued smooth map on $\mathbb{C}^{\mathcal{T}_M}$ which enjoys the bounds*

$$\forall u \in \mathbb{C}^{\mathcal{T}_M}, \quad \|\nabla H(u)\|_{h^{-1/2}} \lesssim_r (\log(M))^{r/2} \|H\|_{\mathcal{H}} \|u\|_{h^{1/2}}^{r-1}.$$

Proof. As a polynomial (of finitely many variables), any Hamiltonian $H \in \mathcal{H}_M^r$ is a smooth map on $\mathbb{C}^{\mathcal{T}_M}$. We aim to bound the norm by duality. To that aim, we fix $v \in \mathbb{C}^{\mathcal{T}_M}$ and we need to estimate $|(\nabla H(u), v)_{\ell^2}|$. Since the coefficients of H are symmetric, we then write

$$\begin{aligned}
|(\nabla H(u), v)_{\ell^2}| &\leq r\|H\|_{\mathcal{H}} \sum_{\sigma \in \{-1,1\}^r} \sum_{\mathbf{k} \in \mathcal{T}_M^r} \frac{|u_{k_1}^{\sigma_1}|}{\langle \ell_1 \rangle^{\frac{1}{2}}} \cdots \frac{|v_{k_r}^{\sigma_r}|}{\langle \ell_r \rangle^{\frac{1}{2}}} \\
&\leq r2^r \|H\|_{\mathcal{H}} \sum_{\mathbf{k} \in \mathcal{T}_M^r} \frac{\langle \ell_1 \rangle^{\frac{1}{2}} |u_{k_1}|}{\langle \ell_1 \rangle} \cdots \frac{\langle \ell_r \rangle^{\frac{1}{2}} |v_{k_r}|}{\langle \ell_r \rangle} \\
&\leq r2^r \|H\|_{\mathcal{H}} \|u\|_{h^{1/2}}^{r-1} \|v\|_{h^{1/2}} \Big(\sum_{k=(\ell,m) \in \mathcal{T}_M} \frac{1}{\langle \ell \rangle^2} \Big)^{r/2} \\
&\lesssim_r (\log(M))^{r/2} \|H\|_{\mathcal{H}} \|u\|_{h^{1/2}}^{r-1} \|v\|_{h^{1/2}}.
\end{aligned}$$

Then by duality we obtain

$$\|\nabla H(u)\|_{h^{-1/2}} \lesssim_r (\log(M))^{r/2} \|H\|_{\mathcal{H}} \|u\|_{h^{1/2}}^{r-1}.$$ ■

The $\mathcal{C}$-norm provides a better estimate of the gradient:

Lemma 4.6. *Let $M \geq 2$, $r \geq 2$. For all $\chi \in \mathcal{H}_M^r$ and all $u \in \mathbb{C}^{\mathcal{T}_M}$, we have the bounds*

$$\|\nabla \chi(u)\|_{h^{1/2}} \lesssim_r (\log(M))^{(r-1)/2} \|\chi\|_{\mathcal{C}} \|u\|_{h^{1/2}}^{r-1} \tag{36}$$

and

$$\|\mathrm{d}\nabla \chi(u)\|_{\mathcal{L}(h^{1/2})} \lesssim_r (\log(M))^{(r-1)/2} \|\chi\|_{\mathcal{C}} \|u\|_{h^{1/2}}^{r-2}. \tag{37}$$

Proof. Without loss of generality, we assume that $\|\chi\|_{\mathcal{C}} = 1$. We aim to prove (36) by duality i.e. for every $v \in \mathbb{C}^{\mathcal{T}_M}$, we want to estimate $|(\nabla\chi(u), v)_{\ell^2}|$. We denote $\tilde{u}_k = \langle\ell\rangle^{\frac{1}{2}}|u_k|$ and $\tilde{v}_k = \langle\ell\rangle^{-\frac{1}{2}}|v_k|$ for all $k = (\ell, m) \in \mathcal{T}_M$ in such a way that $\|\tilde{u}\|_{\ell^2} = \|u\|_{h^{1/2}}$ and $\|\tilde{v}\|_{\ell^2} = \|v\|_{h^{-1/2}}$. Since the coefficients of χ are symmetric, we have

$$(\nabla\chi(u), v)_{\ell^2} = r \sum_{\sigma\in\{-1,1\}^r} \sum_{\mathbf{k}\in\mathcal{T}_M^r} \chi_{\mathbf{k}}^{\sigma} u_{k_1}^{\sigma_1} \cdots u_{k_{r-1}}^{\sigma_{r-1}} v_{k_r}^{\sigma_r}. \tag{38}$$

Then, by applying the triangular inequality, we get

$$\begin{aligned}&|(\nabla\chi(u), v)_{\ell^2}| \\ &\quad\leq 2r \sum_{\sigma\in\{-1,1\}^{r-1}} \sum_{\mathbf{k}\in\mathcal{T}_M^r} \frac{1}{\langle\sigma_1\ell_1 + \cdots + \sigma_{r-1}\ell_{r-1} - \ell_r\rangle\sqrt{\Upsilon(\mathbf{k})}} \frac{\tilde{u}_{k_1}}{\langle\ell_1\rangle} \cdots \frac{\tilde{u}_{k_{r-1}}}{\langle\ell_{r-1}\rangle} \tilde{v}_{k_r}.\end{aligned}$$

At this stage, we notice that, for all $\mathbf{k} \in \mathcal{T}_M^r$, we have $\Upsilon(\mathbf{k}) \geq \Upsilon'(\mathbf{k})$, where $\Upsilon'(\mathbf{k}) = 1$ except when $k_j \neq k_r$ for all $j = 1, \ldots, r-1$ and in that case $\Upsilon'(\mathbf{k}) = \langle\ell_r\rangle$. Thus

$$\begin{aligned}&|(\nabla\chi(u), v)_{\ell^2}| \\ &\quad\leq 2r \sum_{\sigma\in\{-1,1\}^{r-1}} \sum_{\mathbf{k}\in\mathcal{T}_M^r} \frac{1}{\langle\sigma_1\ell_1 + \cdots + \sigma_{r-1}\ell_{r-1} - \ell_r\rangle\sqrt{\langle\ell_r\rangle}} \frac{\tilde{u}_{k_1}}{\langle\ell_1\rangle} \cdots \frac{\tilde{u}_{k_{r-1}}}{\langle\ell_{r-1}\rangle} \tilde{v}_{k_r} \\ &\qquad + 2r \sum_{\sigma\in\{-1,1\}^{r-1}} \sum_{\substack{\mathbf{k}\in\mathcal{T}_M^r \\ \exists\, 1\leq i\leq r-1 : k_r = k_i}} \frac{1}{\langle\sigma_1\ell_1 + \cdots + \sigma_{r-1}\ell_{r-1} - \ell_r\rangle} \frac{\tilde{u}_{k_1}}{\langle\ell_1\rangle} \cdots \frac{\tilde{u}_{k_{r-1}}}{\langle\ell_{r-1}\rangle} \tilde{v}_{k_r} \\ &\quad= 2r(\Sigma_1 + \Sigma_2).\end{aligned}$$

First we estimate Σ_1:

$$\Sigma_1 = \sum_{\sigma\in\{-1,1\}^{r-1}} \sum_{\mathbf{k}\in\mathcal{T}_M^r} \frac{\tilde{u}_{k_1}}{\langle\ell_1\rangle} \cdots \frac{\tilde{u}_{k_{r-1}}}{\langle\ell_{r-1}\rangle} \frac{\tilde{v}_{k_r}}{\langle\sigma_1\ell_1 + \cdots + \sigma_{r-1}\ell_{r-1} - \ell_r\rangle\langle\ell_r\rangle^{\frac{1}{2}}}.$$

We notice that

$$\sum_{k=(\ell,m)\in\mathcal{T}_M} \frac{1}{\langle\ell + a\rangle^2\langle\ell\rangle} = \sum_{\ell=0}^{M} \frac{2\ell+1}{\langle\ell\rangle} \frac{1}{\langle\ell + a\rangle^2} \leq \sum_{j\in\mathbb{Z}} \frac{4}{\langle j\rangle^2} \lesssim 1$$

uniformly with respect to $a \in \mathbb{R}$ and

$$\sum_{k=(\ell,m)\in\mathcal{T}_M} \frac{1}{\langle\ell\rangle^2} \lesssim \log(M).$$

Thus by Cauchy–Schwarz we get

$$\Sigma_1 \lesssim_r \|u\|_{h^{1/2}}^{r-1} (\log(M))^{(r-1)/2} \|v\|_{h^{-1/2}}.$$

It remains to estimate Σ_2. We can assume without lost of generality, but paying an extra factor r, that $k_{r-1} = k_r$. Then, by Cauchy–Schwarz, we get

$$\begin{aligned}\Sigma_2 &\leq r2^{r-1} \sum_{k_{r-1}\in\mathcal{T}_M} \tilde{u}_{k_{r-1}} \tilde{v}_{k_{r-1}} \sum_{k=(\ell,m)\in\mathcal{T}_M^{r-2}} \frac{\tilde{u}_{k_1}}{\langle \ell_1\rangle} \cdots \frac{\tilde{u}_{k_{r-2}}}{\langle \ell_{r-2}\rangle} \\ &\lesssim_r \|u\|_{h^{1/2}}^{r-1} (\log(M))^{(r-2)/2} \|v\|_{h^{-1/2}}.\end{aligned}$$

Putting together the estimates of Σ_1 and Σ_2 we conclude that, for all $v \in \mathbb{C}^{\mathcal{T}_M}$,

$$|(\nabla\chi(u), v)| \lesssim_r (\log(M))^{(r-1)/2} \|v\|_{h^{-1/2}} \|u\|_{h^{1/2}}^{r-1},$$

which in turn implies (36).

To prove (37) we just notice that since $\nabla\chi(u)$ is a homogeneous polynomial, it can be viewed as the trace of an $(r-1)$-linear map on $\mathbb{C}^{\mathcal{T}_M}$: $\nabla\chi(u) = F(u, \dots, u)$ with F that can be expressed using (38). Thus, following the above proof, F satisfies

$$\|F(u^{(1)}, \dots, u^{(r-1)})\|_{h^{1/2}} \lesssim_r (\log(M))^{(r-1)/2} \|u^{(1)}\|_{h^{1/2}} \cdots \|u^{(r-1)}\|_{h^{1/2}}.$$

Then, since $d\nabla\chi(u)(v) = F(v, u, \dots, u) + \cdots + F(u, \dots, u, v)$, we deduce (37). ■

Thanks to a standard duality argument, we rewrite estimate (37) in a negative Sobolev space.

Corollary 4.7. *Let $M \geq 2$, $r \geq 2$. For all $\chi \in \mathcal{H}_M^r$ and $u \in \mathbb{C}^{\mathcal{T}_M}$, we have*

$$\|\mathrm{d}\nabla\chi(u)\|_{\mathcal{L}(h^{-1/2})} \lesssim_r (\log(M))^{(r-1)/2} \|\chi\|_{\mathcal{C}} \|u\|_{h^{1/2}}^{r-2}. \tag{39}$$

Proof. By duality we have

$$\sup_{\substack{v\in\mathbb{C}^{\mathcal{T}_M}\\ \|v\|_{h^{-1/2}}\leq 1}} \|\mathrm{d}\nabla\chi(u)(v)\|_{h^{-1/2}} = \sup_{\substack{v\in\mathbb{C}^{\mathcal{T}_M}\\ \|v\|_{h^{-1/2}}\leq 1}} \sup_{\substack{w\in\mathbb{C}^{\mathcal{T}_M}\\ \|w\|_{h^{1/2}}\leq 1}} (w, \mathrm{d}\nabla\chi(u)(v))_{\ell^2}.$$

Then by applying the Schwarz theorem we have

$$\begin{aligned}(w, \mathrm{d}\nabla\chi(u)(v))_{\ell^2} &= \mathrm{d}[(w, \nabla\chi(u))_{\ell^2}](v) = \mathrm{d}[\mathrm{d}\chi(u)(w)](v) = \mathrm{d}^2\chi(u)(w)(v) \\ &= \mathrm{d}^2\chi(u)(v)(w) = \mathrm{d}[(v, \nabla\chi(u))_{\ell^2}](w) = (v, \mathrm{d}\nabla\chi(u)(w))_{\ell^2}.\end{aligned}$$

Therefore

$$\begin{aligned}\sup_{\substack{v\in\mathbb{C}^{\mathcal{T}_M}\\ \|v\|_{h^{-1/2}}\leq 1}} \|\mathrm{d}\nabla\chi(u)(v)\|_{h^{-1/2}} &= \sup_{\substack{w\in\mathbb{C}^{\mathcal{T}_M}\\ \|w\|_{h^{1/2}}\leq 1}} \sup_{\substack{v\in\mathbb{C}^{\mathcal{T}_M}\\ \|v\|_{h^{-1/2}}\leq 1}} (v, \mathrm{d}\nabla\chi(u)(w))_{\ell^2} \\ &= \sup_{\substack{w\in\mathbb{C}^{\mathcal{T}_M}\\ \|w\|_{h^{1/2}}\leq 1}} \|\mathrm{d}\nabla\chi(u)(w)\|_{h^{1/2}} \\ &= \|\mathrm{d}\nabla\chi(u)\|_{\mathcal{L}(h^{1/2})}.\end{aligned}$$

As a consequence, (39) is just a corollary of estimate (37). ■

Finally, we define the flow associated with a Hamiltonian in $\mathcal{H}_M^r$:

Proposition 4.8. *Let $M \geq 2$, $r \geq 3$ and $\chi \in \mathcal{H}_M^r$. There exist*

$$\varepsilon_0 \gtrsim_r \big((\log(M))^{(r-1)/2}\|\chi\|_{\mathcal{C}}\big)^{-1/(r-2)} \tag{40}$$

and a smooth map

$$\Phi_\chi\colon \begin{cases} [-1,1] \times B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0,\varepsilon_0) \to \mathbb{C}^{\mathcal{T}_M}, \\ \qquad\qquad\qquad\qquad (t,u) \mapsto \Phi_\chi^t(u), \end{cases}$$

solving the equation

$$-i\partial_t \Phi_\chi = (\nabla\chi)\circ\Phi_\chi, \tag{41}$$

and such that for all $t \in [-1,1]$, Φ_χ^t is symplectic, close to the identity

$$\forall u \in B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0,\varepsilon_0), \quad \|\Phi_\chi^t u - u\|_{h^{1/2}} \leq \Big(\frac{\|u\|_{h^{1/2}}}{\varepsilon_0}\Big)^{r-2}\|u\|_{h^{1/2}}, \tag{42}$$

invertible

$$\|\Phi_\chi^t(u)\|_{h^{1/2}} < \varepsilon_0 \quad\Rightarrow\quad \Phi_\chi^{-t}\circ\Phi_\chi^t(u) = u. \tag{43}$$

Moreover, its differential enjoys the estimate

$$\forall u \in B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0,\varepsilon_0),\ \forall\sigma\in\{-1,1\}, \quad \|\mathrm{d}\Phi_\chi^t(u)\|_{\mathcal{L}(h^{\sigma/2})} \leq 2. \tag{44}$$

Proof. We note that (41) is an ODE associated with the smooth vector field $X_\chi = i\nabla\chi$ and therefore we deduce from the Cauchy–Lipschitz theorem that the flow $\Phi_\chi^t(u)$ is locally well defined for every $u \in \mathbb{C}^{\mathcal{T}_M}$ on some maximal interval $(T_-(u), T_+(u))$ containing 0. Let us first show that, if $\|u\|_{h^{1/2}} = \varepsilon$ is small enough, then the solution is defined up to time 1, equivalently $T_+(u) \geq 1$. To see this, we set

$$t_0 := \sup\big\{t \in [0, T_+(u)) : \forall 0 \leq s \leq t,\ \|\Phi_\chi^s(u)\|_{h^{1/2}} < 2\varepsilon\big\} > 0.$$

In the case where $T_+(u) < \infty$, we note that $t_0 < T_+(u)$ by the maximality of the interval of definition and we can verify that $t_0 \geq 1$ provided ε is chosen small enough. Indeed, if $t_0 < 1$, then we can write

$$\begin{aligned} \varepsilon \leq \|\Phi_\chi^{t_0}(u) - u\|_{h^{1/2}} &\leq \int_0^{t_0} \|(\nabla\chi)\circ\Phi_\chi^s(u)\|_{h^{1/2}}\,\mathrm{d}s \\ &\leq C_r^{-(r-2)} t_0 (\log(M))^{(r-1)/2}\varepsilon^{r-1}\|\chi\|_{\mathcal{C}}, \end{aligned}$$

for some constant $0 < C_r \leq 1$, depending only on r coming from (36). From this, we infer

$$\varepsilon^{-1}\big((\log(M))^{(r-1)/2}\|\chi\|_{\mathcal{C}}\big)^{-\frac{1}{r-2}} \leq C_r^{-1}|t_0|^{\frac{1}{r-2}}.$$

Thus, as long as $\varepsilon \leq C_r((\log(M))^{(r-1)/2}\|\chi\|_{\mathcal{C}})^{-\frac{1}{r-2}}$, we find that $t_0 \geq 1$ and that the flow is well defined up to time $t = 1$. The same holds in negative times. We now fix

$$\varepsilon_0 := \frac{C_r}{2}\left((\log(M))^{(r-1)/2}\|\chi\|_{\mathcal{C}}\right)^{-\frac{1}{r-2}}$$

so that $t_0 \geq 1$ for every $\|u\|_{h^{1/2}} = \varepsilon < \varepsilon_0$. Since $\Phi_\chi^t(u)$ is the flow associated with a Hamiltonian vector field, it is symplectic and invertible and we are left with the proof of (42) and (44). For the former, we write as above, for $-1 \leq t \leq 1$,

$$\begin{aligned}\|\Phi_\chi^t(u) - u\|_{h^{1/2}} &\leq \left|\int_0^t \|(\nabla\chi)\circ\Phi_\chi^s(u)\|_{h^{1/2}}\,\mathrm{d}s\right| \\ &\leq C_r^{-(r-2)}\|\chi\|_{\mathcal{C}}(\log(M))^{(r-1)/2}\|u\|_{h^{1/2}}^{r-1} \\ &\leq \left(\frac{\|u\|_{h^{1/2}}}{\varepsilon_0}\right)^{r-2}\|u\|_{h^{1/2}}.\end{aligned}$$

It now remains to prove (44). Up to decreasing the value of ε_0 a little bit (by a factor depending only on r), we can proceed as above by appealing to (37) and (39) and by writing

$$\mathrm{d}\Phi_\chi^t(u) = \mathrm{Id} + \int_0^t \mathrm{d}\nabla\chi(\Phi_\chi^s(u))\circ\mathrm{d}\Phi_\chi^s(u)\,\mathrm{d}s. \qquad ■$$

5. Birkhoff normal form

In this section, we aim to describe a procedure that, close to $u = 0$, allows Hamiltonians on $\mathbb{C}^{\mathcal{T}_M}$ that are of the form

$$H(u) := \frac{1}{2}\sum_{k\in\mathcal{T}_M}\omega_k|u_k|^2 + P(u),$$

where $P \in \mathcal{H}_M^p$, to be simplified. In other words, we will write a Birkhoff normal form for H which means that, up to conjugation by a symplectomorphism and up to a small remainder term, P can be replaced by a term Poisson commuting with the super-actions composing the leading part of H:

$$\forall \ell \geq 0, \quad J_\ell(u) = \sum_{m=-\ell}^{\ell}|u_{(\ell,m)}|^2.$$

This will be used in Section 6 to put (KG) into a Birkhoff normal form and to prove our main theorem. From now on, we fix an integer $p \geq 3$ (the degree of the nonlinearity of (KG)) and $\mu > 0$ (the mass of (KG)) making the frequencies ($\omega_{(\ell,m)} = \sqrt{\ell(\ell+1)+\mu}$) nonresonant (in the sense of Proposition 3.1). Our precise Birkhoff normal form statement reads as follows:

Theorem 5.1. *Let $a > 0$, $C_p > 0$ and $r \geq 1$. Then there exist $\beta > 1$ (independent of the choice of μ) and $C > 1$ such that the following holds.*

For every $M \geq 2$, $N \geq 1$ and every polynomial Hamiltonian of the form $H : \mathbb{C}^{\mathcal{T}_M} \to \mathbb{R}$,

$$H = Z_2 + P^{(p)} \quad \textit{where } Z_2(u) = \frac{1}{2} \sum_{k \in \mathcal{T}_M} \omega_k |u_k|^2,\ P^{(p)} \in \mathcal{H}^p_M,\ \|P^{(p)}\|_{\mathcal{H}} \leq C_p B^a,$$

with $B = \max(\log M, N)$, one can find $\varepsilon_2 \geq (CB^\beta)^{-1}$ and two smooth symplectic maps $\tau^{(0)}$ and $\tau^{(1)}$ making the following diagram commute:

$$\begin{tikzcd} B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0, \varepsilon_2) \arrow[r, "\tau^{(0)}"] \arrow[rr, bend right=10, "\mathrm{id}_{\mathbb{C}^{\mathcal{T}_M}}"'] & B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0, 2\varepsilon_2) \arrow[r, "\tau^{(1)}"] & \mathbb{C}^{\mathcal{T}_M}, \end{tikzcd} \tag{45}$$

and close to the identity

$$\forall \nu \in \{0, 1\}, \quad \|u\|_{h^{1/2}} < 2^\nu \varepsilon_2 \ \Rightarrow\ \|\tau^{(\nu)}(u) - u\|_{h^{1/2}} \leq \left(\frac{\|u\|_{h^{1/2}}}{2^\nu \varepsilon_2} \right)^{p-2} \|u\|_{h^{1/2}}, \tag{46}$$

such that, on $B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0, 2\varepsilon_2)$, $H \circ \tau^{(1)}$ admits the decomposition

$$H \circ \tau^{(1)} = Z_2 + Q^{\leq N}_{\mathrm{res}} + R, \tag{47}$$

where $Q^{\leq N}_{\mathrm{res}} : \mathbb{C}^{\mathcal{T}_M} \to \mathbb{R}$ is a polynomial of degree $r + p - 1$ commuting with the low super-actions

$$\forall \ell \in \mathbb{N}, \quad \langle \ell \rangle \leq N \ \Rightarrow\ \{J_\ell, Q^{\leq N}_{\mathrm{res}}\} = 0. \tag{48}$$

Moreover, the remainder term R is a smooth function on $B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0, 2\varepsilon_2)$ satisfying

$$\|\nabla R(u)\|_{h^{-1/2}} \leq C B^\beta \|u\|^{r+p-1}_{h^{1/2}},$$

and, for all $\nu \in \{0, 1\}$, we have the bounds

$$\|\mathrm{d}\tau^{(\nu)}(u)\|_{\mathcal{L}(h^{1/2})} \leq 2^r \quad \textit{and} \quad \|\mathrm{d}\tau^{(\nu)}(u)\|_{\mathcal{L}(h^{-1/2})} \leq 2^r. \tag{49}$$

Proof. The proof is similar to that of [8, Thm. 4.1]. Nevertheless, here, we have a weaker control of the remainder term ($h^{-1/2}$ instead of $h^{1/2}$ in [8]) and the vector field and Poisson bracket estimates of Section 4 generate new constants we have to track. As usual, we proceed by induction. More precisely, we choose $n \in [\![p, r + p]\!]$ as induction index and assume that Theorem 5.1 holds if

- we replace (47) by

$$H \circ \tau^{(1)} = Z_2 + \sum_{j=p}^{r+p-1} Q^{(j)} + R$$
$$\text{where } Q^{(j)} \in \mathcal{H}^j_M \text{ satisfies } \|Q^{(j)}\|_{\mathcal{H}} \leq C B^\beta, \tag{50}$$

- we replace (48) by

$$\forall \ell \in \mathbb{N},\ \forall j \in [\![p, n-1]\!], \quad \langle \ell \rangle \le N \ \Rightarrow\ \{J_\ell, Q^{(j)}\} = 0, \tag{51}$$

- we replace (49) by

$$\|\mathrm{d}\tau^{(\nu)}(u)\|_{\mathcal{L}(h^{1/2})} \le 2^{n-p} \quad \text{and} \quad \|\mathrm{d}\tau^{(\nu)}(u)\|_{\mathcal{L}(h^{-1/2})} \le 2^{n-p}. \tag{52}$$

Even if we do not write it explicitly, we note that each polynomial $Q^{(j)}$ depends implicitly on n as well as R, ε_2 and $\tau^{(\nu)}$. Moreover, we suppose that R verifies the quantitative estimates of the theorem and that each $Q^{(j)}$ enjoys the same norm estimate as $P^{(p)}$ up to increasing the value of the constant C_p (in a way that depends only on (n, μ, a)) and up to increasing the value of a and β (in a way that depends only on (n, a)). If $n = p$, there is nothing to do: it is in fact enough to choose $\tau^{(0)} = \tau^{(1)} = \mathrm{id}_{\mathbb{C}^{\mathcal{T}_M}}$, $R = 0$, $Q^{(p)} = P^{(p)}$, $Q^{(j)} = 0$ for $j > p$ and $\beta = a$. For the sake of clarity, we will denote with a symbol $\sharp$ the objects we are going to introduce at the step $n+1$ (e.g. $\tau_\sharp^{(0)}$, $\beta_\sharp$). Before entering the details of the proof, recall that one goes formally from step n to $n+1$ by conjugating the normal form (50) by the time 1 map of the Hamiltonian flow of some well-chosen function χ. The function χ is chosen in such a way that the terms of $Q^{(n)}$ that do not commute with the expected super-actions are canceled out by solving a certain cohomological equation.

Decomposition of $Q^{(n)}$. We split the polynomial $Q^{(n)}$ as $Q = L + U$, the Hamiltonians $L, U \in \mathcal{H}_M^n$ being defined by

$$L_{\mathbf{k}}^{\sigma} = \begin{cases} (Q^{(n)})_{\mathbf{k}}^{\sigma} & \text{if } \kappa(\sigma, \mathbf{k}) \le N, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad U_{\mathbf{k}}^{\sigma} = \begin{cases} 0 & \text{if } \kappa(\sigma, \mathbf{k}) \le N, \\ (Q^{(n)})_{\mathbf{k}}^{\sigma} & \text{otherwise,} \end{cases}$$

where $\kappa(\sigma, \mathbf{k})$ is defined in (24) and denotes the smallest effective index of the small divisor $\Omega(\sigma, \mathbf{k})$ defined in (23). Observe that, since these Hamiltonians are extracted from $Q^{(n)}$, they enjoy the same norm estimates.

***U* commutes with the low super-actions.** Indeed, a direct computation shows that if $\langle \ell \rangle \le N$, we have

$$\begin{aligned} \{J_\ell, U\} &= 2i \sum_{\sigma \in \{-1,1\}^n} \sum_{\mathbf{k} \in \mathcal{T}_M^n} (\sigma_1 \mathbb{1}_{\omega_{k_1} = \omega_{(\ell,0)}} + \cdots + \sigma_n \mathbb{1}_{\omega_{k_n} = \omega_{(\ell,0)}}) U_{\mathbf{k}}^{\sigma} u_{k_1}^{\sigma_1} \cdots u_{k_n}^{\sigma_n} \\ &= 2i \sum_{\sigma \in \{-1,1\}^n} \sum_{\mathbf{k} \in \mathcal{T}_M^n} \Bigg(\sum_{j : \exists m, k_j = (\ell, m)} \sigma_j \Bigg) U_{\mathbf{k}}^{\sigma} u_{k_1}^{\sigma_1} \cdots u_{k_n}^{\sigma_n}. \end{aligned}$$

However, since $\langle \ell \rangle \le N$, by definition of U and κ (see (24)), either $\sum_{j : \exists m, k_j = (\ell, m)} \sigma_j$ vanishes or $U_{\mathbf{k}}^{\sigma}$ vanishes. Consequently, U and J_ℓ commute: $\{J_\ell, U\}(u) = 0$. We emphasize that the definition of κ as the smallest effective index is crucial here. Without it, we would need some smoothness assumption on u to control these commutators. As a result, we will have many more terms to solve in the upcoming cohomological equation but we will be able to handle these extra factors thanks to the control of the small divisors given by Proposition 3.1.

The cohomological equation. The mass μ has been fixed to make the frequencies strongly nonresonant (according to Proposition 3.1). Therefore, there exist $\gamma \in (0, 1)$ (depending only on (n, μ)) and $\alpha > 1$ (depending only on n) such that

$$\kappa(\sigma, \mathbf{k}) \leq N \quad \Rightarrow \quad |\Omega(\sigma, \mathbf{k})| \geq \gamma N^{-\alpha} =: \delta. \tag{53}$$

Therefore we set $\chi \in \mathcal{H}_M^n$ to be the Hamiltonian defined by

$$\chi_{\mathbf{k}}^{\sigma} := \frac{L_{\mathbf{k}}^{\sigma}}{i\,\Omega(\sigma, \mathbf{k})} \text{ if } \kappa(\sigma, \mathbf{k}) \leq N \quad \text{and} \quad \chi_{\mathbf{k}}^{\sigma} = 0 \text{ otherwise.}$$

A direct computation shows that χ is a solution of the cohomological equation

$$\{\chi, Z_2\} + L = 0. \tag{54}$$

Let us now verify that we have a good control of the $\mathcal{C}$-norm of χ. First, the bounds

$$\forall y \geq 0, \quad |\langle y\rangle - y| \leq 1 \quad \text{and} \quad |\sqrt{y(y+1)+\mu} - y| \leq \mu + 1$$

and the decomposition

$$\left\langle \sum_{j=1}^{n} \sigma_j \ell_j \right\rangle = \left(\left\langle \sum_{j=1}^{n} \sigma_j \ell_j \right\rangle - \sum_{j=1}^{n} \sigma_j \ell_j \right) + \sum_{j=1}^{n} \sigma_j (\ell_j - \omega_{k_j}) + \Omega(\sigma, \mathbf{k}),$$

where $k_j = (\ell_j, m_j)$ for all $j \in [\![1, n]\!]$, provide the estimate

$$\langle \sigma_1 \ell_1 + \cdots + \sigma_n \ell_n \rangle \leq (n+1)(\mu+1) + |\Omega(\sigma, \mathbf{k})|.$$

Therefore, as a consequence of (53) (since $\delta < 1$) we have the bound

$$|\chi_{\mathbf{k}}^{\sigma}| \leq (n+2)(\mu+1)\delta^{-1} \frac{|L_{\mathbf{k}}^{\sigma}|}{\langle \sigma_1 \ell_1 + \cdots + \sigma_n \ell_n \rangle}$$

and so

$$\|\chi\|_{\mathcal{C}} \lesssim_{n,\mu} \delta^{-1} \|L\|_{\mathcal{H}} \lesssim_{n,\mu} \delta^{-1} \|Q^{(n)}\|_{\mathcal{H}} \lesssim_{n,\mu} \delta^{-1} C B^{\beta}.$$

The new variables. As usual, we have to compose the change of variables τ at step n with the Hamiltonian flow of χ (see (58) below). Since they are only defined locally, we have to pay attention to their domains of definition. Even though the overall strategy is clear, it is a little tedious to check.

Since $\|\chi\|_{\mathcal{C}} \lesssim_n \delta^{-1} C B^{\beta}$ and $\gamma N^{-\alpha} =: \delta$, applying Proposition 4.8, we get a constant $K > 0$ depending only on (n, C, μ), an exponent $b > 0$ depending only on (n, β) such that, setting $\varepsilon_1 = (K B^b)^{-1/(n-2)}$, χ generates a smooth map

$$\Phi_\chi \colon \begin{cases} [-1, 1] \times B_{h^{1/2}(\mathbb{C}^{\mathcal{T}_M})}(0, \varepsilon_1) \to \mathbb{C}^{\mathcal{T}_M}, \\ \qquad\qquad (t, u) \mapsto \Phi_\chi^t(u), \end{cases}$$

solving the equation $-i\partial_t\Phi_\chi = (\nabla\chi)\circ\Phi_\chi$, and such that for all $t\in[-1,1]$, Φ_χ^t is symplectic, close to the identity

$$\|u\|_{h^{1/2}} < \varepsilon_1 \quad\Rightarrow\quad \|\Phi_\chi^t u - u\|_{h^{1/2}} \le \Big(\frac{\|u\|_{h^{1/2}}}{\varepsilon_1}\Big)^{n-2}\|u\|_{h^{1/2}}, \tag{55}$$

invertible

$$\|\Phi_\chi^{-t}(u)\|_{h^{1/2}} < \varepsilon_1 \quad\Rightarrow\quad \Phi_\chi^t\circ\Phi_\chi^{-t}(u) = u. \tag{56}$$

Moreover, the map $u\mapsto \mathrm{d}\Phi_\chi^t(u)$ is continuous and we have the estimates

$$\|u\|_{h^{1/2}} < \varepsilon_1 \quad\Rightarrow\quad \|\mathrm{d}\Phi_\chi^t(u)\|_{\mathcal{L}(h^{1/2})} \le 2 \quad\text{and}\quad \|\mathrm{d}\Phi_\chi^t(u)\|_{\mathcal{L}(h^{-1/2})} \le 2. \tag{57}$$

As usual, we aim to define, for a proper choice of $\varepsilon_2^\sharp$,

$$\tau_\sharp^{(1)} := \tau^{(1)}\circ\Phi_\chi^1 \ \text{ on } B_{h^{1/2}}(0,2\varepsilon_2^\sharp) \quad\text{and}\quad \tau_\sharp^{(0)} := \Phi_\chi^{-1}\circ\tau^{(0)} \ \text{ on } B_{h^{1/2}}(0,\varepsilon_2^\sharp). \tag{58}$$

To ensure that such a definition makes sense, we have to choose $\varepsilon_2^\sharp$ in such a way that

$$2\varepsilon_2^\sharp \le \varepsilon_1 \quad\text{and}\quad (\|u\|_{h^{1/2}} < 2\varepsilon_2^\sharp \ \Rightarrow\ \|\Phi_\chi^1(u)\|_{h^{1/2}} < 2\varepsilon_2). \tag{59}$$

$$\varepsilon_2^\sharp \le \varepsilon_2 \quad\text{and}\quad (\|u\|_{h^{1/2}} < \varepsilon_2^\sharp \ \Rightarrow\ \|\tau^{(0)}(u)\|_{h^{1/2}} < \varepsilon_1). \tag{60}$$

Let us analyze these conditions. First, we focus on (59). Provided that $\|u\|_{h^{1/2}} < 2\varepsilon_2^\sharp \le \varepsilon_1$, since Φ_χ^1 is close to the identity (see (55)), we have $\Phi_\chi^1(u) \le 2\|u\|_{h^{1/2}} < 4\varepsilon_2^\sharp$. Therefore, to get (59) it is enough to have $2\varepsilon_2^\sharp \le \min(\varepsilon_2,\varepsilon_1)$. Similarly, since $\tau^{(0)}$ is close to the identity (see (46)), to get (60) it is enough to ensure that $2\varepsilon_2^\sharp \le \varepsilon_1$ and $\varepsilon_2^\sharp \le \varepsilon_2$.

Before fixing $\varepsilon_2^\sharp$, let us only assume that $2\varepsilon_2^\sharp \le \min(\varepsilon_2,\varepsilon_1)$ and investigate which conditions $\varepsilon_2^\sharp$ has to satisfy to ensure that $\tau_\sharp^{(1)}$ and $\tau_\sharp^{(0)}$ enjoy the properties described in Theorem 5.1 (close to the identity, invertible, ...).

First, let us note that $\tau_\sharp^{(1)}$ and $\tau_\sharp^{(0)}$ are obviously symplectic and their differentials enjoy the bounds (52) thanks to (57) (with $n\to n+1$). Hence, it remains to prove that $\tau_\sharp^{(0)}$ and $\tau_\sharp^{(1)}$ are close to the identity in the sense of (46). To that aim, if $\|u\|_{h^{1/2}} < \varepsilon_2^\sharp$, since both Φ_χ^{-1} and $\tau^{(0)}$ are close to the identity, then we have

$$\begin{aligned}\|\tau_\sharp^{(0)}(u) - u\|_{h^{1/2}} &\le \Big(\frac{\|\tau^{(0)}(u)\|_{h^{1/2}}}{\varepsilon_1}\Big)^{n-2}\|\tau^{(0)}(u)\|_{h^{1/2}} + \Big(\frac{\|u\|_{h^{1/2}}}{\varepsilon_2}\Big)^{p-2}\|u\|_{h^{1/2}}\\ &\le \Big(\frac{2\|u\|_{h^{1/2}}}{\varepsilon_1}\Big)^{n-2}2\|u\|_{h^{1/2}} + \Big(\frac{\|u\|_{h^{1/2}}}{\varepsilon_2}\Big)^{p-2}\|u\|_{h^{1/2}}.\end{aligned}$$

Therefore, since $n\ge p$ and $2\|u\|_{h^{1/2}} < 2\varepsilon_2^\sharp \le \varepsilon_1$, we deduce that

$$\|\tau_\sharp^{(0)}(u) - u\|_{h^{1/2}} \le \Big(\frac{\|u\|_{h^{1/2}}}{\varepsilon_2^\sharp}\Big)^{p-2}\|u\|_{h^{1/2}}\Big[\frac{2(\varepsilon_2^\sharp)^{p-2}}{\varepsilon_1^{p-2}} + \frac{(\varepsilon_2^\sharp)^{p-2}}{\varepsilon_2^{p-2}}\Big].$$

Moreover, since $p \geq 3$, if $3\varepsilon_2^\sharp \leq \min(\varepsilon_2, \varepsilon_1)$, we deduce that both $(\varepsilon_2^\sharp)^{p-2}/\varepsilon_1^{p-2}$ and $(\varepsilon_2^\sharp)^{p-2}/\varepsilon_2^{p-2}$ are bounded by $1/3$. As a consequence, if $3\varepsilon_2^\sharp \leq \min(\varepsilon_2, \varepsilon_1)$ then $\tau_\sharp^{(0)}$ is close to the identity. It can be proven, with a similar decomposition, that if $6\varepsilon_2^\sharp \leq \min(\varepsilon_2, \varepsilon_1)$ then $\tau_\sharp^{(1)}$ is also close to the identity.

Finally, we also note that if $\tau_\sharp^{(0)}$ is close to the identity, then it takes values in $B_{h^{1/2}}(0, 2\varepsilon_2^\sharp)$. Thus, as Φ_χ^1 is invertible (see (56)), diagram (45) associated with $\tau_\sharp^{(0)}$ and $\tau_\sharp^{(1)}$ commutes.

To conclude this paragraph, we fix $\varepsilon_2^\sharp$ as large as possible to get all the properties of $\tau_\sharp^{(0)}$ and $\tau_\sharp^{(1)}$, i.e.

$$\varepsilon_2^\sharp = \frac{1}{6}\min(\varepsilon_2, \varepsilon_1).$$

We note that, therefore, we have $\varepsilon_2^\sharp \geq \frac{1}{6}\min((KB^b)^{-1/(n-2)}, (CB^\beta)^{-1}) \geq (C_\sharp B^{\beta_\sharp})^{-1}$ provided that $C_\sharp \geq 6\max(K^{1/(n-2)}, C)$ and $\beta_\sharp \geq \max(b/(n-2), \beta)$ (these constants will be determined at the end of the proof).

The new Hamiltonian. We aim to describe the Taylor expansion of $H \circ \tau_\sharp^{(1)}$. Since $t \mapsto \Phi_\chi^t$ is a smooth function solving the equation $-i\partial_t \Phi_\chi = (\nabla\chi) \circ \Phi_\chi$, realizing a Taylor expansion in $t = 0$ (on $B_{h^{1/2}}(0, 2\varepsilon_2^\sharp)$) gives

$$\begin{aligned}
H \circ \tau_\sharp^{(1)} = H \circ \tau^{(1)} \circ \Phi_\chi^1 = Z_2 \circ \Phi_\chi^1 + \sum_{j=p}^{r+p-1} Q^{(j)} \circ \Phi_\chi^1 + R \circ \Phi_\chi^1 \\
= Z_2 + \sum_{j=p}^{r+p-1} Q^{(j)} + \{\chi, Z_2\} + \sum_{h=1}^{m_n} \frac{1}{(h+1)!}\mathrm{ad}_\chi^{h+1} Z_2 \\
+ \sum_{j=p}^{r+p-1} \sum_{h=1}^{m_j} \frac{1}{h!}\mathrm{ad}_\chi^h Q^{(j)} + R \circ \Phi_\chi^1 \\
+ \int_0^1 \Big(\frac{(1-t)^{m_n+1}}{(m_n+1)!} (\mathrm{ad}_\chi^{m_n+2} Z_2) \circ \Phi_\chi^t \\
+ \sum_{j=p}^{r+p-1} \frac{(1-t)^{m_j}}{m_j!} (\mathrm{ad}_\chi^{m_j+1} Q^{(j)}) \circ \Phi_\chi^t \Big)\, \mathrm{d}t,
\end{aligned}$$

where m_j denotes the largest integer such that $j + m_j(n-2) < r + p$ and $\mathrm{ad}_\chi := \{\chi, \cdot\}$.

In order to pool these terms by packets, we recall that by construction $\{\chi, Z_2\} = -L$ is of order n, that $\chi \in \mathcal{H}_M^n$ is of degree n and that the Poisson bracket of two homogeneous polynomials of degrees r_1 and r_2 is of degree $r_1 + r_2 - 2$. Therefore, we set

$$Q_\sharp^{(j)} = Q^{(j)} \quad \text{if } j < n, \qquad Q_\sharp^{(n)} = Q^{(n)} + \{\chi, Z_2\} = Q^{(n)} - L = U,$$

$$Q_\sharp^{(j)} = \sum_{j_\star + h(n-2) = j} \frac{1}{h!}\mathrm{ad}_\chi^h Q^{(j_\star)} - \sum_{n + h(n-2) = j} \frac{1}{(h+1)!}\mathrm{ad}_\chi^h L \quad \text{if } j > n,$$

$$R_\sharp = R \circ \Phi^1_\chi - \int_0^1 \Big(\frac{(1-t)^{m_n+1}}{(m_n+1)!} (\mathrm{ad}_\chi^{m_n+1} L) \circ \Phi^t_\chi \\ + \sum_{j=p}^{r+p-1} \frac{(1-t)^{m_j}}{m_j!} (\mathrm{ad}_\chi^{m_j+1} Q^{(j)}) \circ \Phi^t_\chi \Big) \, \mathrm{d}t,$$

where h and $j_\star$ are the indices on which the sums hold in the definition of $Q^{(j)}_\sharp$.

If $j \leq n$, $Q^{(j)}_\sharp \in \mathcal{H}^j_M$ commutes with the low super-actions[3] and we have

$$\|Q^{(j)}_\sharp\|_{\mathcal{H}} \leq \|Q^{(j)}\|_{\mathcal{H}} \leq CB^\beta.$$

If $j > n$, we have $Q^{(j)}_\sharp \in \mathcal{H}^j_M$ and we apply Proposition 4.4 to estimate its norm. Indeed, if $j_\star + h(n-2) = j$, we can use our estimate on $\|\chi\|_{\mathcal{C}}$ to derive that

$$\|\mathrm{ad}^h_\chi Q^{(j_\star)}\|_{\mathcal{H}} \lesssim_r (\log M)^h \|\chi\|^h_{\mathcal{C}} \|Q^{(j_\star)}\|_{\mathcal{H}} \lesssim_r (\gamma^{-1} N^\alpha \log M)^h (CB^\beta)^{h+1} \\ \lesssim_r \gamma^{-h} C^{h+1} B^{h(\alpha+1)+(h+1)\beta},$$

where we recall that $B := \max(\log M, N)$. Similarly, L enjoying the same bound as $Q^{(n)}$, if $n + h(n-2) = j$, we have $\|\mathrm{ad}^h_\chi L\|_{\mathcal{H}} \lesssim_r \gamma^{-h} C^{h+1} B^{h(\alpha+1)+(h+1)\beta}$. As a consequence, since $h \leq r + p$, provided that $C_\sharp \gtrsim_r \gamma^{-r-p} C^{r+p+1}$ and $\beta_\sharp \geq (r+p)(\alpha+1) + (r+p+1)\beta$, we have $\|Q^{(j)}_\sharp\|_{\mathcal{H}} \leq C_\sharp B^{\beta_\sharp}$ for $j > n$.

Control of the remainder term. Now we are left with controlling $\nabla R_\sharp$ in $h^{-1/2}$. We fix $u \in \mathbb{C}^{\mathcal{T}_M}$ such that $\|u\|_{h^{1/2}} < 2\varepsilon_2^\sharp$. First we focus on $R \circ \Phi^1_\chi(u)$. By composition, we have

$$\nabla(R \circ \Phi^1_\chi)(u) = (\mathrm{d}\Phi^1_\chi(u))^* (\nabla R) \circ \Phi^1_\chi(u),$$

where $(\mathrm{d}\Phi^1_\chi(u))^* \in \mathcal{L}(\mathbb{C}^{\mathcal{T}_M})$ denotes the adjoint of $\mathrm{d}\Phi^1_\chi(u)$. Moreover, by duality, we have $\|(\mathrm{d}\Phi^1_\chi(u))^*\|_{\mathcal{L}(h^{1/2})} = \|\mathrm{d}\Phi^1_\chi(u)\|_{\mathcal{L}(h^{-1/2})} \leq 2$. Therefore, since $\|\nabla R(u)\|_{h^{-1/2}} \leq CB^\beta \|u\|^{r+p-1}_{h^{1/2}}$ and $\|\Phi^1_\chi(u)\|_{h^{1/2}} \leq 2\|u\|_{h^{1/2}}$, we have

$$\|\nabla(R \circ \Phi^1_\chi)(u)\|_{h^{-1/2}} \leq 2^{r+p} CB^\beta \|u\|^{r+p-1}_{h^{1/2}}.$$

Now we focus on $(\mathrm{ad}_\chi^{m_j+1} Q^{(j)}) \circ \Phi^t_\chi(u)$ where $p \leq j \leq r + p - 1$ and $t \in [0, 1]$. Arguing as above and using Proposition 4.4 to estimate the norm of the Poisson brackets and Lemma 4.5 to estimate the norm of the gradient, we have

$$\|\nabla((\mathrm{ad}_\chi^{m_j+1} Q^{(j)}) \circ \Phi^t_\chi)(u)\|_{h^{-1/2}} \\ \leq 2\|(\nabla(\mathrm{ad}_\chi^{m_j+1} Q^{(j)})) \circ \Phi^t_\chi(u)\|_{h^{-1/2}} \\ \lesssim_{r,\mu} (\delta^{-1} \log M)^{m_j+1} (CB^\beta)^{m_j+2} (\log M)^{r_j/2} \|\Phi^t_\chi(u)\|^{r_j-1}_{h^{1/2}},$$

---

[3] Note that U has been designed to get this property.

where $r_j = j + (m_j + 1)(n - 2) \in [\![r + p, 2(r + p)]\!]$ (by definition of m_j). Thus, provided that

$$C_\sharp \gtrsim_{r,\mu} \gamma^{-r-p-1} C^{r+p+2} \quad \text{and} \quad \beta_\sharp \geq (\alpha + 1)(r + p + 1) + \beta(r + p + 2) + r + p,$$

we have $\|\nabla((\mathrm{ad}_\chi^{m_j+1} Q^{(j)}) \circ \Phi_\chi^t)(u)\|_{h^{-1/2}} \leq C_\sharp B^{\beta_\sharp} \|u\|_{h^{1/2}}^{r+p-1}$. As above, the argument works as well for the term involving L as it enjoys the same norm estimate as $Q^{(n)}$.

Hence if, moreover, $\beta_\sharp \geq \beta$ and $C_\sharp \gtrsim_r C$ (to control $R \circ \Phi_\chi^1(u)$), we have

$$\|\nabla R_\sharp(u)\|_{h^{-1/2}} \leq C_\sharp B^{\beta_\sharp} \|u\|_{h^{1/2}}^{r+p-1}.$$

Choice of $C_\sharp$ and $\beta_\sharp$. To conclude our induction step (and thus the proof), we just have to pick the smallest constants enjoying all the constraints (and to note that they do not depend on B):

$$\begin{aligned} &\beta_\sharp = (\alpha + 1)(r + p + 1) + \beta(r + p + 2) + r + p, \\ &C_\sharp \simeq_r \max(\gamma^{-r+p-1} C^{r+p+2}, K^{1/(n-2)}). \end{aligned}$$

■

6. Proofs of the main results

This final section is devoted to the proof of Theorem 1.1 and of its Corollary 1.2.

6.1. On the global well-posedness of (KG)

In dimension 2, the Sobolev norm H^1 controls all the Lebesgue norms L^q, $2 \leq q < \infty$. Therefore, a standard fixed point argument (which does not require any kind of Strichartz estimate) provides the local well-posedness of the nonlinear Klein–Gordon equation (KG) on the sphere $\mathbb{S}^2$ in the energy space $H^1 \times L^2$ (see e.g. [17, Thm. 6.2.2, p. 83]).

This nonlinear equation is Hamiltonian because it can formally be written as

$$\partial_t \begin{pmatrix} \Phi \\ \partial_t \Phi \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \nabla \mathscr{H}(\Phi, \partial_t \Phi), \tag{61}$$

where the Hamiltonian $\mathscr{H}$ is given by (5). Therefore, $\mathscr{H}$ is a constant of the motion of (KG) (see e.g. [17, Prop, 6.2.3, p. 83]). It is especially useful since, as stated in the following lemma, it is uniformly elliptic in a neighborhood of the origin:

Lemma 6.1. *For all $g \in L^\infty(\mathbb{S}^2; \mathbb{R})$ and all $\mu > 0$, there exist $C > 1$ and $\varepsilon_0 > 0$ such that for all $(\Phi, \Psi) \in H^1 \times L^2(\mathbb{S}^2; \mathbb{R})$, provided that $\|\Phi\|_{H^1} + \|\Psi\|_{L^2} \leq \varepsilon_0$, we have*

$$C^{-1}(\|\Phi\|_{H^1} + \|\Psi\|_{L^2})^2 \leq \mathscr{H}(\Phi, \Psi) \leq C(\|\Phi\|_{H^1} + \|\Psi\|_{L^2})^2.$$

Proof. It follows directly from the Sobolev embedding $H^1 \hookrightarrow L^p$ and from the fact that $p \geq 3$. ■

As a consequence, as stated in the following proposition we get the global well-posedness of (KG) in a neighborhood of the origin in $H^1 \times L^2$ (see e.g. [17, Prop. 6.3.3, p. 84]).

Proposition 6.2. *For all $\mu > 0$ and all $g \in L^\infty$, there exist $\varepsilon_1 > 0$ and $K > 1$ such that, as soon as $\varepsilon := \|\Phi^{(0)}\|_{H^1} + \|\dot{\Phi}^{(0)}\|_{L^2} \leq \varepsilon_1$, there exists a unique $\Phi \in C^0(\mathbb{R}; H^1) \cap C^1(\mathbb{R}; L^2) \cap C^2(\mathbb{R}; H^{-1})$ solution to (KG). Moreover, it enjoys the bound*

$$\forall t \in \mathbb{R}, \quad \|\Phi(t)\|_{H^1} + \|\partial_t \Phi(t)\|_{L^2} \leq K\varepsilon.$$

6.2. Proof of Theorem 1.1

One more, we fix the mass $\mu > 0$ (in a set of full measure) to make the frequencies $(\omega_{(\ell,m)} = \sqrt{\ell(\ell+1) + \mu})$ nonresonant in the sense of Proposition 3.1. The strategy is the following. Using the above a priori estimates, we prove that the high super-actions are under control as long as $N = \langle \ell \rangle \gtrsim \varepsilon^{-\frac{p-2}{\alpha_r + 1}}$ for an arbitrary $\alpha_r > 1$. Thus, we only have to deal with the low super-actions that we handle using the Birkhoff normal form of Theorem 5.1. This requires making a truncation of the frequency up to a certain level M in order to reduce to the finite-dimensional situation of this theorem. In order to ensure that all the remainder terms are small in this reduction to finite dimension, we need to take M of order ε^{-r}. Then the conclusion follows by combining our a priori estimates on the solution with the normal form of Theorem 5.1 and by taking α_r larger than the exponent β appearing in the remainder terms of that statement.

(KG) as a Schrödinger equation. We consider $(\Phi^{(0)}, \dot{\Phi}^{(0)}) \in H^1 \times L^2$, satisfying $\varepsilon := \|\Phi^{(0)}\|_{H^1} + \|\dot{\Phi}^{(0)}\|_{L^2} < \varepsilon_0 \leq \varepsilon_1$, where ε_0 will be determined at the end of the proof and ε_1 is given by Proposition 6.2. Thanks to this proposition, one obtains a global solution Φ to (KG). Then, in order to diagonalize the linear part of (KG), we set (as usual)

$$u := \Lambda \Phi + i \Lambda^{-1} \partial_t \Phi, \quad \text{where } \Lambda := (\mu - \Delta)^{1/4}.$$

Indeed, u belongs to $C^0(\mathbb{R}; H^{1/2}) \cap C^1(\mathbb{R}; H^{-1/2})$ and solves the equation

$$i \partial_t u = \Lambda^2 u - \Lambda^{-1}(g[\Lambda^{-1} \Re u]^{p-1}). \tag{62}$$

It is relevant to note that the harmonic energies $\mathcal{E}_\ell$ (defined by (3)), that we aim to control in Theorem 1.1, satisfy

$$\forall \ell \in \mathbb{N}, \quad \mathcal{E}_\ell(\Phi(t)) = \|\Pi_\ell u(t)\|_{L^2}^2 := J_\ell(u(t)),$$

where Π_ℓ is the orthogonal projection on the eigenspace E_ℓ as defined in (2). Moreover, as a consequence of Proposition 6.2, there exists a constant $K' > 1$ depending only on μ such that

$$\forall t \in \mathbb{R}, \quad \|u(t)\|_{H^{1/2}} \leq K'\varepsilon. \tag{63}$$

The N-truncation. The control of the high super-actions is a direct consequence of the a priori bound (63). Indeed, applying the triangular inequality, we have

$$|J_\ell(u(t)) - J_\ell(u(0))| \le J_\ell(u(t)) + J_\ell(u(0)) \le 2\langle \ell\rangle^{-1}\|u\|^2_{L^\infty_t H^{1/2}_x} \le 2\langle \ell\rangle^{-1}(K')^2\varepsilon^2.$$

Being given $\alpha_r > 1$ (depending only on r), which will be optimized at the end of the proof, we set

$$N^{(\max)} := \varepsilon^{-\frac{p-2}{\alpha_r+1}}.$$

As a consequence, for all $t \in \mathbb{R}$ we have

$$\langle \ell\rangle \ge N^{(\max)} \quad\Rightarrow\quad |J_\ell(u(t)) - J_\ell(u(0))| \lesssim_{r,\mu} \langle \ell\rangle^{\alpha_r}\varepsilon^p. \tag{64}$$

Hence, from now on we will only focus on the variations of the low super-actions. More precisely, we fix $\ell_\star \in \mathbb{N}$ and $N \in \mathbb{R}$ such that

$$N := \langle \ell_\star\rangle < N^{(\max)}$$

and we aim to estimate the variations of $J_{\ell_\star}(u)$.

The M-truncation. In order to reduce ourselves to the finite-dimensional situation of our Birkhoff normal form Theorem 5.1, we are going to prove that the high enough modes (larger than $M \gg 1$) do not play any role in the dynamics for very long times (in $H^{-1/2}$). Let $M \ge 2N^{(\max)}$ be a constant that will be optimized later with respect to ε and $\Pi_{\le M}$ be the orthogonal projection on $\bigoplus_{\ell\le M} E_\ell$, i.e.

$$\Pi_{\le M} := \sum_{\ell\le M}\Pi_\ell \quad\text{and}\quad \Pi_{>M} := \mathrm{Id}_{L^2} - \Pi_{\le M}.$$

We set

$$F^{(>M)}(t) := \Pi_{\le M}[\mathcal{N}(\Pi_{\le M}u(t)) - \mathcal{N}(u(t))] \quad \text{where } \mathcal{N}(u) := \Lambda^{-1}(g[\Lambda^{-1}\Re u]^{p-1}).$$

Since u solves equation (62), $u^{(\le M)} := \Pi_{\le M}u(t)$ solves the *nonautonomous* equation

$$i\partial_t u^{(\le M)} = \Lambda^2 u^{(\le M)} - \Pi_{\le M}\mathcal{N}(u^{(\le M)}) + F^{(>M)}(t). \tag{65}$$

We note that, since $M \ge 2N^{(\max)}$, we have $M > \ell_\star$ and so

$$J_{\ell_\star}(u^{(\le M)}) = J_{\ell_\star}(u). \tag{66}$$

We aim to prove that the nonautonomous part of (65) (i.e. $F^{(>M)}(t)$) is negligible provided that M is large enough. Indeed, as a consequence of the Sobolev embeddings $H^1 \hookrightarrow L^{6(p-2)} \hookrightarrow L^{3/2} \hookrightarrow H^{-1}$, by Hölder and the mean value inequality, we have (uniformly with respect to t)

$$\begin{aligned}
\|F^{(>M)}\|_{H^{-1/2}} &\lesssim_\mu \|g\Phi^{p-1} - g(\Pi_{\le M}\Phi)^{p-1}\|_{H^{-1}}\\
&\lesssim_{\mu,g} \|\Phi^{p-1} - (\Pi_{\le M}\Phi)^{p-1}\|_{L^{3/2}}\\
&\lesssim_{\mu,g} \|(\Pi_{>M}\Phi)(|\Pi_{\le M}\Phi|^{p-2} + |\Phi|^{p-2})\|_{L^{3/2}}\\
&\lesssim_{\mu,g} \|\Pi_{>M}\Phi\|_{L^2}(\|(\Pi_{\le M}\Phi)^{p-2}\|_{L^6} + \|\Phi^{p-2}\|_{L^6})\\
&\lesssim_{\mu,g} M^{-1}\|\Phi\|^{p-1}_{H^1} \lesssim_{\mu,g} M^{-1}\varepsilon^{p-1}.
\end{aligned}$$

Therefore, from now, we assume that $M \geq \varepsilon^{-r}$, and we get

$$\forall t \in \mathbb{R}, \quad \|F^{(>M)}(t)\|_{H^{-1/2}} \lesssim_\mu \varepsilon^{r+p-1}.$$

Discretization. Thanks to Theorem 2.1, we get a basis $(e_k)_{k\in\mathcal{T}_\infty}$ of L^2 which diagonalizes the Laplace–Beltrami operator Δ and enjoys nice algebraic properties. In particular, thanks to this basis, we identify $\bigoplus_{\ell\leq M} E_\ell$ with $\mathbb{R}^{\mathcal{T}_M}$ (and the usual Sobolev norms with the discrete ones).

We use this basis to rewrite the autonomous part of (65) as a Hamiltonian system:

$$i\partial_t u^{(\leq M)} = \nabla H(u^{(\leq M)}) + F^{(>M)}(t), \tag{67}$$

where

$$H = Z_2 + P^{(p)} \quad \text{with } Z_2(u) = \frac{1}{2}\sum_{k\in\mathcal{T}_M} \omega_k |u_k|^2$$

and $P^{(p)} \in \mathcal{H}_M^p$ is defined, for all $\mathbf{k} = (k_1, \dots, k_p) \in \mathcal{T}_M^p$ and $\sigma \in \{-1, 1\}^p$, by

$$(P^{(p)})_{\mathbf{k}}^\sigma = -\frac{1}{p2^p}\left(\prod_{j=1}^p \frac{1}{(\ell_j(\ell_j+1)+\mu)^{1/4}}\right)\int_{\mathbb{S}^2} e_{k_1}(x)\cdots e_{k_p}(x) g(x)\, \mathrm{dvol}_{\mathbb{S}^2}(x).$$

Thanks to Theorem 2.1, the basis $(e_k)_{k\in\mathcal{T}_\infty}$ has been chosen such that

$$\|P^{(p)}\|_{\mathcal{H}} \lesssim (\log(M))^p. \tag{68}$$

Note that the choice of the orthonormal basis of Theorem 2.1 is crucial here. With the standard basis of spherical harmonics we would not get such good control on the nonlinearity.

Change of variables. Now we apply Theorem 5.1 (i.e. our Birkhoff normal form result) to simplify the Hamiltonian part of (67). More precisely, we get some transformations $\tau^{(0)}$, $\tau^{(1)}$, some Hamiltonians $Q_{\mathrm{res}}^{\leq N}$ and R, some constants C, β and ε_2 such that the statement of Theorem 5.1 holds. We recall that B is defined by $B = \max(N, \log(M))$.

We will optimize the constants in such a way that we have

$$K'\varepsilon < (CB^\beta)^{-1},$$

where K' has been defined in (63). As a consequence, we have

$$\forall t \in \mathbb{R}, \quad \|u^{(\leq M)}(t)\|_{h^{1/2}} \leq K'\varepsilon < (CB^\beta)^{-1} \leq \varepsilon_2.$$

Therefore, it makes sense to define

$$v := \tau^{(0)} \circ u^{(\leq M)}.$$

Moreover, since the diagram (45) commutes we have

$$u^{(\leq M)} = \tau^{(1)} \circ v.$$

As a consequence, since $\tau^{(0)}$ is symplectic and $(\mathrm{d}\tau^{(0)}(u^{(\leq M)}))^{-1} = \mathrm{d}\tau^{(1)}(v)$, we have

$$i\partial_t v(t) = \nabla(Z_2 + Q_{\mathrm{res}}^{\leq N})(v(t)) + W(t), \tag{69}$$

where W is the new remainder term defined by

$$W(t) := \nabla R(v(t)) + \mathrm{d}\tau^{(0)}(u^{(\leq M)}(t))(F^{(>M)}(t)).$$

Let us estimate W. On the one hand, since $\tau^{(0)}$ is close to the identity in the sense of Theorem 5.1, we have

$$\begin{aligned}\|v(t)\|_{h^{1/2}} &\leq \|u^{(\leq M)}(t)\|_{h^{1/2}} + \|v(t) - u^{(\leq M)}(t)\|_{h^{1/2}} \\ &\leq 2\|u^{(\leq M)}(t)\|_{h^{1/2}} \leq 2K'\varepsilon \lesssim_\mu \varepsilon.\end{aligned} \tag{70}$$

Hence, thanks to Theorem 5.1, we get $\|\nabla R(v(t))\|_{h^{-1/2}} \lesssim_{r,\mu} B^\beta \varepsilon^{r+p-1}$. On the other hand, since $\mathrm{d}\tau^{(0)}(u^{(\leq M)}(t))$ is controlled in $\mathcal{L}(h^{-1/2})$ (by 2^r), we deduce that

$$\|\mathrm{d}\tau^{(0)}(u^{(\leq M)}(t))(F^{(>M)}(t))\|_{h^{-1/2}} \lesssim_{r,\mu} \varepsilon^{r+p-1}.$$

Therefore, we have

$$\|W(t)\|_{h^{-1/2}} \lesssim_{r,\mu} B^\beta \varepsilon^{r+p-1}. \tag{71}$$

Finally, let us note that, since $\tau^{(0)}$ is close to the identity in the sense of Theorem 5.1 and $(CB^\beta)^{-1} \leq \varepsilon_2$, we have

$$\|u^{(\leq M)}(t) - v(t)\|_{h^{1/2}} \lesssim_{r,\mu} \varepsilon^{p-1} B^{\beta(p-2)}. \tag{72}$$

Control of the low super-actions. As a consequence of (66), (72) and (70), we have

$$\begin{aligned}|J_{\ell_\star}(u(t)) - J_{\ell_\star}(v(t))| &\leq \|u^{(\leq M)}(t) - v(t)\|_{\ell^2}(\|u^{(\leq M)}(t)\|_{\ell^2} + \|v(t)\|_{\ell^2}) \\ &\lesssim_{r,\mu} \varepsilon^p B^{\beta(p-2)}.\end{aligned}$$

Hence, by the triangular inequality, we have

$$|J_{\ell_\star}(u(t)) - J_{\ell_\star}(u(0))| \lesssim_r |J_{\ell_\star}(v(t)) - J_{\ell_\star}(v(0))| + \varepsilon^p B^{\beta(p-2)}.$$

However, since v solves (69), we have

$$\partial_t J_{\ell_\star}(v(t)) = \{J_{\ell_\star}, Z_2 + Q_{\mathrm{res}}^{\leq N}\}(v(t)) + (i\nabla J_{\ell_\star}(v(t)), W(t))_{\ell^2}.$$

By construction, since $\langle \ell_\star \rangle = N$, $Z_2 + Q_{\mathrm{res}}^{\leq N}$ and $J_{\ell_\star}$ commute, i.e. $\{J_{\ell_\star}, Z_2 + Q_{\mathrm{res}}^{\leq N}\} = 0$. As a consequence, using estimate (71) on W we have

$$\begin{aligned}|\partial_t J_{\ell_\star}(v(t))| \leq |(i\nabla J_{\ell_\star}(v(t)), W(t))_{\ell^2}| &\leq \|\nabla J_{\ell_\star}(v(t))\|_{h^{1/2}} \|W(t)\|_{h^{-1/2}} \\ &\leq 2\|v(t)\|_{h^{1/2}} \|W(t)\|_{h^{-1/2}} \lesssim_{r,\mu} B^\beta \varepsilon^{r+p}.\end{aligned}$$

Consequently, while $|t| \leq \varepsilon^{-r}$, we have

$$|J_{\ell_\star}(u(t)) - J_{\ell_\star}(u(0))| \lesssim_{r,\mu} \varepsilon^p B^{\beta(p-2)} \lesssim_{r,\mu,\nu} \langle \ell_\star \rangle^{\alpha_r} \varepsilon^{p-\nu} \tag{73}$$

provided that $B^{\beta(p-2)} \lesssim_{r,\mu,\nu} N^{\alpha_r} \varepsilon^{-\nu}$ where $\nu > 0$.

Conclusion. As we wanted, in (64) and (73), we have controlled the variations of the super-actions. Nevertheless, to get these results we have made some assumptions on our parameters. Hence, to conclude, we have to check their compatibility and optimize them.

More precisely, we have to prove that there exists $\alpha_r > 1$ and $\varepsilon_0 \leq \varepsilon_1$ such that for all $\varepsilon < \varepsilon_0$ and all $N < N^{(\max)} = \varepsilon^{-\frac{p-2}{\alpha_r+1}}$, there exists $M \geq 2$ satisfying

$$\text{(i)}\ \ B^{\beta(p-2)} \lesssim_{r,\mu,\nu} N^{\alpha_r}\varepsilon^{-\nu}, \qquad \text{(ii)}\ \ K'\varepsilon < (CB^{\beta})^{-1},$$
$$\text{(iii)}\ \ M \geq \varepsilon^{-r}, \qquad \text{(iv)}\ \ M \geq 2N^{(\max)},$$

where $B = \max(N, \log(M))$. First, we set $M = \varepsilon^{-r}$ (so (iii) is satisfied). Then we set $\alpha_r = \beta(p-2)$ and we note that estimate (i) holds. Finally, since $p \leq r$, we note that (ii) and (iv) are clearly satisfied provided that ε_0 is small enough.

6.3. Proof of Corollary 1.2

For all $t \in \mathbb{R}$, let $w(t) \in H^{1/2}(\mathbb{S}^2;\mathbb{C})$ be defined, for all $\ell \in \mathbb{N}$, by

$$\Pi_\ell w(t) = \sqrt{\frac{J_\ell(u(0))}{J_\ell(u(t))}}\Pi_\ell u(t) \ \text{ if } J_\ell(u(t)) \neq 0 \quad \text{and} \quad \Pi_\ell w(t) = \Pi_\ell u(0) \ \text{ otherwise.}$$

Indeed, recalling that $J_\ell = \|\Pi_\ell \cdot\|_{L^2}^2$, this function satisfies $\|w(t)\|_{H^{1/2}} = \|u(0)\|_{H^{1/2}}$ and

$$\forall \ell \in \mathbb{N}, \quad J_\ell(w(t)) = J_\ell(u(0)) \quad \text{and} \quad \sqrt{J_\ell(w(t)-u(t))} = |\sqrt{J_\ell(u(t))} - \sqrt{J_\ell(w(t))}|.$$

As a consequence, applying Theorem 1.1 (with $\nu = 1/2$), while $|t| < \varepsilon^{-r}$, for all $\ell \in \mathbb{N}$, we have

$$J_\ell(u(t)-w(t)) \leq |J_\ell(u(t)) - J_\ell(w(t))| = |J_\ell(u(t)) - J_\ell(u(0))| \lesssim_{\mu,r} \langle \ell \rangle^{\alpha_r}\varepsilon^{p-1/2}.$$

Therefore, we have

$$\|u(t)-w(t)\|_{H^{-\alpha_r/2}} \lesssim_{\mu,r} \varepsilon^{(2p-1)/4}.$$

Consequently, since $s < 1/2$, setting $\theta = \min(1, \frac{1-2s}{1+\alpha_r})$, by interpolation and using Proposition 6.2, we get

$$\|u(t)-w(t)\|_{H^s} \lesssim_{r,s} \|u(t)-w(t)\|_{H^{1/2}}^{1-\theta}\|u(t)-w(t)\|_{H^{-\alpha_r/2}}^{\theta} \lesssim_{r,s,\mu} \varepsilon^{1+\delta},$$

where $\delta := \theta((2p-1)/4-1) > 0$ (because $p \geq 3$). Finally, to see that there exist some Hermitian operators $H_\ell(t)\colon E_\ell \otimes \mathbb{C} \to E_\ell \otimes \mathbb{C}$ such that

$$\forall \ell \in \mathbb{N}, \quad \Pi_\ell w(t) = e^{iH_\ell(t)}\Pi_\ell u(0),$$

it is enough to note that the unitary group of $E_\ell \otimes \mathbb{C}$ acts transitively on the spheres and that every unitary transform is the exponential of a skew-Hermitian operator (indeed, since $J_\ell(w(t)) = J_\ell(u(0))$, $\Pi_\ell w(t)$ and $\Pi_\ell u(0)$ belong to the same sphere).

Acknowledgments. We thank Nicolas Burq for helpful discussions on global well-posedness.

Funding. During the preparation of this work the authors benefited from the support of the Centre Henri Lebesgue ANR-11-LABX-0020-0 and by ANR-15-CE40-0001-01 “BEKAM”. The third author was also supported by the Agence Nationale de la Recherche through the PRC grants ODA (ANR-18-CE40-0020) and ADYCT (ANR-20-CE40-0017).

References

[1] D. Bambusi, Nekhoroshev theorem for small amplitude solutions in nonlinear Schrödinger equations. *Math. Z.* **230** (1999), no. 2, 345–387 Zbl 0928.35160 MR 1676714

[2] D. Bambusi, Birkhoff normal form for some nonlinear PDEs. *Comm. Math. Phys.* **234** (2003), no. 2, 253–285 Zbl 1032.37051 MR 1962462

[3] D. Bambusi, J.-M. Delort, B. Grébert, and J. Szeftel, Almost global existence for Hamiltonian semilinear Klein-Gordon equations with small Cauchy data on Zoll manifolds. *Comm. Pure Appl. Math.* **60** (2007), no. 11, 1665–1690 Zbl 1170.35481 MR 2349351

[4] D. Bambusi and B. Grébert, Birkhoff normal form for partial differential equations with tame modulus. *Duke Math. J.* **135** (2006), no. 3, 507–567 Zbl 1110.37057 MR 2272975

[5] J. Bernier, E. Faou, and B. Grébert, Long time behavior of the solutions of NLW on the d-dimensional torus. *Forum Math. Sigma* **8** (2020), Paper No. e12 Zbl 1441.35169 MR 4073882

[6] J. Bernier, E. Faou, and B. Grébert, Rational normal forms and stability of small solutions to nonlinear Schrödinger equations. *Ann. PDE* **6** (2020), no. 2, Paper No. 14 MR 4171851

[7] J. Bernier and B. Grébert, Long time dynamics for generalized Korteweg-de Vries and Benjamin-Ono equations. *Arch. Ration. Mech. Anal.* **241** (2021), no. 3, 1139–1241 Zbl 1476.37083 MR 4284525

[8] J. Bernier and B. Grébert, Birkhoff normal forms for Hamiltonian PDEs in their energy space. *J. Éc. polytech. Math.* **9** (2022), 681–745 Zbl 1490.35392 MR 4412513

[9] M. Berti and P. Bolle, *Quasi-periodic solutions of nonlinear wave equations on the d-dimensional torus*. EMS Monogr. Math., EMS Publishing House, Berlin, 2020 Zbl 1454.37003 MR 4321994

[10] M. Berti and J.-M. Delort, *Almost global solutions of capillary-gravity water waves equations on the circle*. Lect. Notes Unione Mat. Ital. 24, Springer, Cham, 2018 Zbl 1409.35002 MR 3839269

[11] L. Biasco, J. E. Massetti, and M. Procesi, An abstract Birkhoff normal form theorem and exponential type stability of the 1d NLS. *Comm. Math. Phys.* **375** (2020), no. 3, 2089–2153 Zbl 1441.35217 MR 4091501

[12] P. Bizoń, B. Craps, O. Evnin, D. Hunik, V. Luyten, and M. Maliborski, Conformal flow on S^3 and weak field integrability in AdS_4. *Comm. Math. Phys.* **353** (2017), no. 3, 1179–1199 Zbl 1378.35300 MR 3652487

[13] J. D. Bjorken and S. D. Drell, *Relativistic quantum mechanics*. McGraw-Hill, New York-Toronto-London, 1964 Zbl 0184.54201 MR 0187641

[14] J. Bourgain, Construction of approximative and almost periodic solutions of perturbed linear Schrödinger and wave equations. *Geom. Funct. Anal.* **6** (1996), no. 2, 201–230 Zbl 0872.35007 MR 1384610

[15] J. Bourgain, On diffusion in high-dimensional Hamiltonian systems and PDE. *J. Anal. Math.* **80** (2000), 1–35 Zbl 0964.35143 MR 1771522

[16] N. Burq and G. Lebeau, Injections de Sobolev probabilistes et applications. *Ann. Sci. Éc. Norm. Supér. (4)* **46** (2013), no. 6, 917–962 Zbl 1296.46031 MR 3134684

[17] T. Cazenave and A. Haraux, *An introduction to semilinear evolution equations*. Oxford Lecture Ser. Math. Appl. 13, The Clarendon Press, Oxford University Press, New York, 1998 Zbl 0926.35049 MR 1691574

[18] D. Cohen, E. Hairer, and C. Lubich, Conservation of energy, momentum and actions in numerical discretizations of non-linear wave equations. *Numer. Math.* **110** (2008), no. 2, 113–143 Zbl 1163.65066 MR 2425152

[19] D. Cohen, E. Hairer, and C. Lubich, Long-time analysis of nonlinearly perturbed wave equations via modulated Fourier expansions. *Arch. Ration. Mech. Anal.* **187** (2008), no. 2, 341–368 Zbl 1145.35087 MR 2366141

[20] B. Craps, O. Evnin, and V. Luyten, Maximally rotating waves in AdS and on spheres. *J. High Energy Phys.* (2017), no. 9, Paper No. 59 Zbl 1382.83083 MR 3714125

[21] J.-M. Delort, On long time existence for small solutions of semi-linear Klein-Gordon equations on the torus. *J. Anal. Math.* **107** (2009), 161–194 Zbl 1184.35211 MR 2496403

[22] J.-M. Delort, A quasi-linear Birkhoff normal forms method. Application to the quasi-linear Klein-Gordon equation on $\mathbb{S}^1$. *Astérisque* (2012), no. 341 Zbl 1243.35123 MR 2952065

[23] J.-M. Delort, Quasi-linear perturbations of Hamiltonian Klein-Gordon equations on spheres. *Mem. Amer. Math. Soc.* **234** (2015), no. 1103 Zbl 1315.35003 MR 3288855

[24] J.-M. Delort and R. Imekraz, Long-time existence for the semilinear Klein-Gordon equation on a compact boundary-less Riemannian manifold. *Comm. Partial Differential Equations* **42** (2017), no. 3, 388–416 Zbl 1365.37056 MR 3620892

[25] J.-M. Delort and J. Szeftel, Long-time existence for small data nonlinear Klein-Gordon equations on tori and spheres. *Int. Math. Res. Not.* (2004), no. 37, 1897–1966 Zbl 1079.35070 MR 2056326

[26] J.-M. Delort and J. Szeftel, Long-time existence for semi-linear Klein-Gordon equations with small Cauchy data on Zoll manifolds. *Amer. J. Math.* **128** (2006), no. 5, 1187–1218 Zbl 1108.58023 MR 2262173

[27] D. Fang and Q. Zhang, Long-time existence for semi-linear Klein-Gordon equations on tori. *J. Differential Equations* **249** (2010), no. 1, 151–179 MR 2644131

[28] E. Faou, L. Gauckler, and C. Lubich, Sobolev stability of plane wave solutions to the cubic nonlinear Schrödinger equation on a torus. *Comm. Partial Differential Equations* **38** (2013), no. 7, 1123–1140 Zbl 1274.35357 MR 3169740

[29] R. Feola, B. Grébert, and F. Iandoli, Long time solutions for quasi-linear Hamiltonian perturbations of Schrödinger and Klein-Gordon equations on tori. 2022, arXiv:2009.07553

[30] R. Feola and F. Iandoli, A non-linear Egorov theorem and Poincaré-Birkhoff normal forms for quasi-linear pdes on the circle. 2020, arXiv:2002.12448

[31] R. Feola and F. Iandoli, Long time existence for fully nonlinear NLS with small Cauchy data on the circle. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* **22** (2021), no. 1, 109–182 Zbl 1475.37081 MR 4288651

[32] G. B. Folland, *A course in abstract harmonic analysis*. 2nd edn., Textb. Math., CRC Press, Boca Raton, FL, 2016 Zbl 0857.43001 MR 3444405

[33] F. Giuliani, M. Guardia, P. Martin, and S. Pasquali, Chaotic-like transfers of energy in Hamiltonian PDEs. *Comm. Math. Phys.* **384** (2021), no. 2, 1227–1290 Zbl 1472.37079 MR 4259386

[34] B. Grébert, R. Imekraz, and É. Paturel, Normal forms for semilinear quantum harmonic oscillators. *Comm. Math. Phys.* **291** (2009), no. 3, 763–798 Zbl 1185.81073 MR 2534791

[35] B. Grébert and E. Paturel, KAM for the Klein Gordon equation on $\mathbb{S}^d$. *Boll. Unione Mat. Ital.* **9** (2016), no. 2, 237–288 MR 3502160

[36] B. Grébert and L. Thomann, Resonant dynamics for the quintic nonlinear Schrödinger equation. *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **29** (2012), no. 3, 455–477 MR 2926244

[37] B. Grébert and C. Villegas-Blas, On the energy exchange between resonant modes in nonlinear Schrödinger equations. *Ann. Inst. H. Poincaré C Anal. Non Linéaire* **28** (2011), no. 1, 127–134 MR 2765514

[38] R. Imekraz, Existence en temps grand pour des équations de Klein-Gordon à petite donnée initiale sur une structure de Toeplitz. *J. Funct. Anal.* **264** (2013), no. 6, 1480–1524 Zbl 1290.35226 MR 3017272

[39] M. Ledoux, *The concentration of measure phenomenon*. Math. Surveys Monogr. 89, American Mathematical Society, Providence, RI, 2001 MR 1849347

[40] P. Mattila, *Geometry of sets and measures in Euclidean spaces*. Cambridge Stud. Adv. Math. 44, Cambridge University Press, Cambridge, 1995 Zbl 0819.28004 MR 1333890

[41] B. Shiffman and S. Zelditch, Random polynomials of high degree and Levy concentration of measure. *Asian J. Math.* **7** (2003), no. 4, 627–646 Zbl 1060.32012 MR 2074895

[42] C. D. Sogge, Concerning the L^p norm of spectral clusters for second-order elliptic operators on compact manifolds. *J. Funct. Anal.* **77** (1988), no. 1, 123–138 Zbl 0641.46011 MR 930395

[43] C. Sogge, Problems related to the concentration of eigenfunctions. *Journ. Équ. Dériv. Partielles* (2015), article no. 9

[44] T. Tao, *An introduction to measure theory*. Graduate Studies in Mathematics 126, American Mathematical Society, Providence, RI, 2011 Zbl 1231.28001 MR 2827917

[45] J. M. VanderKam, L^∞ norms and quantum ergodicity on the sphere. *Internat. Math. Res. Notices* (1997), no. 7, 329–347 Zbl 0877.58056 MR 1440572

[46] X. Yuan and J. Zhang, Long time stability of Hamiltonian partial differential equations. *SIAM J. Math. Anal.* **46** (2014), no. 5, 3176–3222 Zbl 1405.37085 MR 3262600

[47] S. Zelditch, Local and global analysis of eigenfunctions on Riemannian manifolds. In *Handbook of geometric analysis. No. 1*, pp. 545–658, Adv. Lect. Math. (ALM) 7, International Press, Somerville, MA, 2008 Zbl 1176.58017 MR 2483375

Received 3 September 2021; accepted 24 January 2022.

Joackim Bernier
Laboratoire de Mathématiques Jean Leray, Université de Nantes, UMR CNRS 6629,
2 rue de la Houssinière, 44322 Nantes Cedex 03, France; joackim.bernier@univ-nantes.fr

Benoît Grébert
Laboratoire de Mathématiques Jean Leray, Université de Nantes, UMR CNRS 6629,
2 rue de la Houssinière, 44322 Nantes Cedex 03, France; benoit.grebert@univ-nantes.fr

Gabriel Rivière
Laboratoire de Mathématiques Jean Leray, Université de Nantes, UMR CNRS 6629,
2 rue de la Houssinière, 44322 Nantes Cedex 03; and Institut Universitaire de France, Paris,
France; gabriel.riviere@univ-nantes.fr