ON LOCAL RIGIDITY OF PARTIALLY HYPERBOLIC AFFINE $\mathbb{Z}^k$ ACTIONS

DANIJELA DAMJANOVIĆ[1] AND BASSAM FAYAD [2]

ABSTRACT. The following dichotomy for affine $\mathbb{Z}^k$ actions on the torus $\mathbb{T}^d$, $k, d \in \mathbb{N}$ is shown to hold : i) the linear part of the action has no rank-one factors, and then the affine action is locally rigid; ii) the linear part of the action has a rank-one factor, and then the affine action is locally rigid in a probabilistic sense if and only if the rank-one factors are trivial. Local rigidity in a probabilistic sense means that rigidity holds for a set of full measure of translation vectors in the rank-one factors.

1. INTRODUCTION

1.1. **Local rigidity of $\mathbb{Z}^k$ actions.** A smooth $\mathbb{Z}^k$ action ρ on a smooth manifold M is given by a homomorphism ρ from $\mathbb{Z}^k$ into the group Diff(M) of C^∞ diffeomorphisms of M.

An action ρ is said to be *locally rigid* if there exists a neighborhood $\mathcal{U}$ of ρ in the space $\mathcal{A}(k, M)$ smooth $\mathbb{Z}^k$ actions on M, such that every $\eta \in \mathcal{U}$ is C^∞ conjugate to ρ via a conjugacy which is C^1 close to identity. We recall that the action η is said to be C^∞ conjugate to ρ if there is a conjugacy $h \in$ Diff(M) of M such that $h(g) \circ \rho(g) \circ h(g)^{-1} = \eta(g)$ for every $g \in \mathbb{Z}^k$.

We say that a $\mathbb{Z}^k$ action ρ has a rank one factor if $k = 1$ or if ρ factors to a $\mathbb{Z}^k$ action which is (up to a finite index subgroup of $\mathbb{Z}^k$) generated by a single diffeomorphism. If a $\mathbb{Z}^k$ action has no rank-one factors, we will call it a *higher rank action* .

1.2. **Local rigidity of higher rank $\mathbb{Z}^k$ linear actions on the torus.** A $\mathbb{Z}^k$ linear action on the torus is given by a homomorphism from $\mathbb{Z}^k$ to the group of automorphisms of the torus.

For linear $\mathbb{Z}^k$ actions on the torus, it is proved in [22] that the action is higher rank if and only if:

[1] Supported by NSF grant DMS-0758555 [2] Supported by ANR-15-CE40-0001 and by the project BRNUH. .

(HR) The $\mathbb{Z}^k$ action contains a subgroup L isomorphic to $\mathbb{Z}^2$ such that every element in L, except for identity, is ergodic with respect to the standard invariant measure obtained from the Haar measure.

The local picture for higher-rank $\mathbb{Z}^k$ actions on the torus by toral automorphisms is fairly well understood. The condition *(HR)* is a necessary and sufficient condition for local rigidity ([3] and references therein).

1.3. **KAM rigidity in rank-one dynamics.** It is widely believed that local rigidity for actions with rank one factor never holds. The only known situation in rank-one dynamics where some form of local rigidity happens, is for Diophantine toral translations. In this case translation vectors with respect to invariant probability measures serve as moduli of smooth conjugacy. More precisely, if a diffeomorphism of a torus has a Diophantine average displacement vector α with respect to some invariant probability measure and if it is sufficiently close to the translation T_α by vector α, then it is smoothly conjugated to T_α, [15]. This phenomenon is a consequence of KAM theory (after Kolmogorov, Arnol'd and Moser) on the stability of Diophantine quasi-periodic motion in close to integrable systems [13], and we call it *KAM rigidity*.

Another point of view on KAM rigidity is given by parametric families of diffeomorphisms. If a torus translation is perturbed into a parametric family of diffeomorphisms and if the translation vectors relative to invariant measures satisfy an adequate transversality condition along the family, then for a large set of parameters the diffeomorphisms of the family are smoothly conjugate to translations. A typical example is given by Arnol'd family of circle diffeomorphisms [1] : $f_t(\theta) = \theta + t + \varepsilon \sin(2\pi\theta)$, $\theta \in \mathbb{R}/\mathbb{Z}$ which are smoothly conjugate to a rotation for a set of $t \in [0,1]$ of measure converging to 1 as ε goes to 0. Transversality in the parameter t ensures that the rotation number of f_t is Diophantine for a large set of the parameters $t \in [0,1]$, which implies the smooth conjugation result.

1.4. **Local rigidity and KAM rigidity of higher rank $\mathbb{Z}^k$ affine actions on the torus.** In this paper we consider *affine* volume preserving $\mathbb{Z}^k$ actions on the torus. Each element of such an action is a diffeomorphism of the torus which is a composition of a translation on the torus and an automorphism of the torus. Therefore every such action has the translation part, and the linear part which is generated by automorphisms. We denote by ρ_L the linear part of an affine action ρ on the torus.

We give now an informal statement that summarizes the main findings of this paper. The precise statement will be made in the next section.

Theorem. *The following dichotomy holds for an affine $\mathbb{Z}^k$ actions ρ on the torus $\mathbb{T}^d$, $k, d \in \mathbb{N}$: i) ρ_L has no rank-one factors, and then the affine action is locally rigid; ii) ρ_L has a rank-one factor and ρ is KAM locally rigid if and only if the rank-one factors are trivial.*

The first part of the dichotomy was proved in [3, Section 1.3.1], where it was observed that the (HR) condition on ρ_L is sufficient for local rigidity of the affine action ρ. It is not hard to see that (HR) is also necessary for local rigidity of affine actions on the torus, see for example Section 2 in [10].

What we prove here is part ii) that complements the local rigidity picture for $\mathbb{Z}^k$ affine actions on the torus by showing that some kind of rigidity, namely KAM rigidity, holds for affine actions having a rank-one factor provided this rank-one factor is trivial.

To illustrate this specificity of affine actions, take for example the $\mathbb{Z}^2$ action on $\mathbb{T}^{d+1}$ generated by $\bar{A} = A \times \mathrm{Id}$, $\bar{B} = B \times \mathrm{Id}$, where A and B are commuting automorphisms $\mathbb{T}^d$ and (A, B) is (HR) (or $d = 0$). Of course the $\mathbb{Z}^2$ action $(\bar{A}, \bar{B})$ does not satisfy the (HR) assumption. But in the affine setting, consider the $\mathbb{Z}^2$ action generated by $A \times R_\alpha$ and $B \times R_\beta$, where R_α and R_β are two circle rotations such that $1, \alpha, \beta$ are rationally independent. All non-trivial elements of this action are ergodic, while its linear part is not ergodic for any element of the action. This action is clearly not locally rigid. We can for example change the frequencies α and β to produce a nearby action which is not conjugate to the initial action. But even with fixed frequencies satisfying fast approximations by rational pairs, Anosov-Katok Liouville constructions show that we can perturb $R_\alpha \times R_\beta$ into a non linearizable commuting pair of circle diffeomorphisms having rotation numbers α and β [4].

In this paper, we show that actions of the type discussed above are actually KAM rigid in the following sense. If (α, β) satisfy some arithmetic condition of full measure and if the generators of the action $A \times R_\alpha$ and $B \times R_\beta$ are perturbed into a commuting pair (f, g) that is sufficiently small (as a function of A, B and the arithmetic condition on (α, β) and that keep the same mean rotation numbers in the $d + 1$ direction of $\mathbb{T}^{d+1}$ for a pair of invariant measures for f and g then f and g are simultaneously conjugated to $A \times R_\alpha$ and $B \times R_\beta$.

Note that if $d = 0$, this is due to a result by Moser on smooth linearization of commuting pairs of circle diffeomorphisms [16] (see also [4] for a global result). This is the content of Theorem 1. In Theorem 2 we prove a parametric form of KAM rigidity for affine torus actions similar to the phenomenon described above for Arnol'd families.

In fact, the two types of affine actions we just described, with linear parts (HR) or having trivial rank one factors, are the only ones for which local rigidity or KAM rigidity happen. For all other affine actions, local rigidity and KAM rigidity do not hold. The reason is that if the affine action does not belong to one of the two classes described above, then its linear part has a non-trivial rank one factor. In Theorem 2, we show how KAM rigidity does not hold for any parametric family of $\mathbb{Z}^k$ actions having a non-trivial rank one factor in their linear part. Namely, we show that any such a family can be perturbed so that no element of the perturbed family of actions is conjugated to an affine action.

The question of global rigidity. Global rigidity has been proved for abelian Anosov (HR) actions on tori in [20] (see references therein for prior works). Without the Anosov condition, it is not expected that for $\mathbb{Z}^k$ higher rank actions global rigidity holds. An exception are circle actions, where Herman?s global Theory on circle diffeomorphisms applies ([23, 4]). In the case of a partially hyperbolic abelian (HR) action with one dimensional central direction, it may be possible to combine the above global rigidity results of hyperbolic (HR) actions on the torus and the global rigidity of circle actions. This is clearly possible if the action is assumed to be a direct product of a hyperbolic (HR) action on the torus and an action on the circle. For the general case, one may first investigate whether a (HR) partially hyperbolic $\mathbb{Z}^k$-action with a one dimensional central direction is necessarily a product action. We do not pursue this question here and stay focused on local rigidity of affine actions without any dimension restrictions.

To finish this introduction, we mention some earlier works on local rigidity of affine group actions.

1.5. **Local rigidity of affine actions by higher rank groups.** Affine actions have been discussed by Hurder in [9]. Local rigidity of affine actions of higher rank non abelian groups was extensively studied (see for example the survey [6]). In [5] Fisher and Margulis provide a complete local picture for affine actions by higher rank lattices in semisimple Lie groups. The methods they use are very different from ours and are specific to groups with Property (T).

Prior to [5], the question about local rigidity of perturbations of product actions of large higher rank groups has been addressed in [17], [18], [21]; the actions considered there are products of the identity action and actions that generalize the standard $\mathrm{SL}(n, \mathbb{Z})$ action on $\mathbb{T}^n$. Local rigidity and deformation rigidity are obtained for such actions. We note that the actions we consider in this paper even though they belong to families of actions, are not deformation rigid in the sense of [8].

Local rigidity results for algebraic abelian Anosov actions were obtained by Katok and Spatzier in [12], including the case of affine actions on tori and nilmanifolds. For higher rank non-Anosov linear and affine $\mathbb{Z}^k$ actions on nilmanifolds, the classification of perturbations is still an open problem. It is not known if KAM rigidity holds for some $\mathbb{Z}^k$ affine actions on nilmanifold whose linear part is a trivial action, or has a trivial action in the factor. There are examples of families of $\mathbb{R}^2$ actions on 2-step nilmanifolds which display a weak form of rigidity for perturbed families [2], but it is not known if KAM rigidity holds or not for the corresponding families of $\mathbb{Z}^2$ actions.

2. Statements

2.1. **KAM-rigidity for Diophantine affine actions.** We denote by (f, g) a $\mathbb{Z}^2$ action generated by two commuting diffeomorphisms f and g. Let (f_0, g_0) be the action generated by

$$f_0 = (A + a) \times (\mathrm{Id}_{\mathbb{T}^{d_2}} + \varphi), g_0 = (B + b) \times (\mathrm{Id}_{\mathbb{T}^{d_2}} + \psi). \tag{2.1}$$

Here, $d_2 \in \mathbb{N}^*$ and $\varphi, \psi \in \mathbb{R}^{d_2}$. Also, $d_1 \in \mathbb{N}$ and either $d_1 = 0$ or A and B are two commuting toral automorphisms of $\mathbb{T}^{d_1}$ and $a, b \in \mathbb{R}^{d_1}$. We denote by $(\bar{A}, \bar{B})$ the linear part of (f_0, g_0).

Let (f, g) be a perturbation of the action (f_0, g_0) on $\mathbb{T}^d$, where $d = d_1 + d_2$.

We define for any pair μ_1, μ_2 of invariant probability measures for f and g respectively, the translation vectors along the $\mathbb{T}^{d_2}$ direction corresponding to these measures as follows

$$\alpha = \rho_{\mu_1}(f) = \int_{\mathbb{T}^d} \pi_2(f(x) - x) d\mu_1(x),$$
$$\beta = \rho_{\mu_2}(g) = \int_{\mathbb{T}^d} \pi_2(g(x) - x) d\mu_2(x)$$

Here π_2 denotes the projection on the $\mathbb{T}^{d_2}$ variable. We say that $(\alpha, \beta) \in \mathbb{T}^{d_2} \times \mathbb{T}^{d_2}$ is simultaneously Diophantine with respect to a pair of numbers (λ, μ) if there exists $\tau, \gamma > 0$ such that

$$\max(|\lambda - e^{i2\pi(k,\alpha)}|, |\mu - e^{i2\pi(k,\beta)}|) > \frac{\gamma}{|k|^\tau}$$

where $\|\cdot\|$ denotes the closest distance to the integers, and we denote this property by $(\alpha, \beta) \in \mathrm{SDC}(\tau, \gamma, \lambda, \mu)$. We say that $(\alpha, \beta) \in \mathrm{SDC}(\tau, \gamma, \bar{A}, \bar{B})$ if given any pair of eigenvalues (λ, μ) of $(\bar{A}, \bar{B})$, it holds that $(\alpha, \beta) \in \mathrm{SDC}(\tau, \gamma, \lambda, \mu)$. Observe that SDC pairs of vectors relatively to any pair $(\bar{A}, \bar{B})$ form a set of full Haar measure in $\mathbb{T}^{d_2} \times \mathbb{T}^{d_2}$.

We call such an action generated by (f, g) Diophantine. We have the following KAM rigidity statement for Diophantine actions.

THEOREM 1. *Let $d_1 \in \mathbb{N}$, $d_2 \in \mathbb{N}^*$ and $d = d_1 + d_2$. If $d_1 > 0$, let A and B be commuting automorphisms of $\mathbb{T}^{d_1}$ such that the $\mathbb{Z}^2$ action (A, B) satisfies the condition (HR).*

For any $\tau, \gamma > 0$, there exist $r(\tau) > 0$ and $\varepsilon = \varepsilon(\tau, \gamma) > 0$ such that: given (f, g)-a smooth (C^∞) $\mathbb{Z}^2$ action on $\mathbb{T}^d$, and probability measures μ_1, μ_2 invariant by f and g, respectively, if

$$(\alpha, \beta) = (\rho_{\mu_1}(f), \rho_{\mu_2}(g)) \in SDC(\tau, \gamma, \bar{A}, \bar{B}),$$

and if

$$\|f - (A + a) \times T_\alpha\|_r \leq \varepsilon, \ \ \|g - (B + b) \times T_\beta\|_r \leq \varepsilon,$$

for some $a, b \in \mathbb{R}^{d_1}$ and translations T_α and T_β of $\mathbb{T}^{d_2}$, then the action (f, g) is smoothly conjugate to the affine action. Namely, there exists $h \in \mathrm{Diff}(\mathbb{T}^{\mathrm{d}})$ such that

$$h \circ f \circ h^{-1} = (A + a) \times T_\alpha, \ \ h \circ g \circ h^{-1} = (B + b) \times T_\beta.$$

In the case $d_2 = 1$, the SDC condition is reminiscent of the condition used by Moser to prove local rigidity of commuting circle diffeomorphisms with this condition on their rotation numbers [16]. The ingredients of the proof of Theorem 1 are indeed a mixture of the ingredients used in the higher rank rigidity of toral automorphisms [3] and the KAM rigidity in the quasi-periodic setting as in [1] and [16]. Note that the case $d_1 = 0$ corresponds to a generalization to higher dimensional tori of Moser's result on local rigidity of commuting circle diffeomorphisms of [16].

2.2. **Parametric KAM-rigidity for affine actions.** Similar to the theory of circle diffeomorphisms, we have the parametric version of KAM rigidity for affine $\mathbb{Z}^k$ actions.

Let (f_t, g_t), $t \in [0,1]$, be a family of affine $\mathbb{Z}^2$ actions on $\mathbb{T}^d$ which is of class C^d in the parameter t. By continuity, all the actions in the family must have a common linear part $(\bar{A}, \bar{B})$. The generators f_t, g_t of the $\mathbb{Z}^2$ action (f_t, g_t) are then given by $\bar{A} + \bar{a}_t$ and $g_t = \bar{B} + \bar{b}_t$, where $\bar{a}_t, \bar{b}_t \in \mathbb{R}^d$ are translation vectors that are of class C^d in the parameter t.

We denote this family by $(f_\cdot, g_\cdot)$. Denote by $\|\cdot\|_{l,r}$ the combination of C^l norm in parameter t and C^r norm in the torus variable.

To state a parametric version of KAM rigidity, we need some transversality on the frequencies along the elliptic factor of the action. We will use a Pyartli [19] type condition (although other common transversality conditions in KAM theory may be applied as well).

Definition 1. We say that a function $\rho \in C^r([0,1], \mathbb{R}^d)$, $r \geq d$, satisfies a Pyartli condition if for any $t \in [0,1]$ we have that the first d derivatives of ρ are linearly independent. There exists then a constant $\nu > 0$ such that

$$|\det(\rho', \rho'', \ldots, \rho^{(d)})| \geq \nu, \quad \|\rho\|_d \leq \nu^{-1} \tag{2.2}$$

The Pyartli condition is clearly a generic condition in $C^r([0,1], \mathbb{R}^d)$, for any $r \geq d$.

Now we are ready to state the KAM rigidity theorem for families of actions.

THEOREM 2. *Let $(f_t, g_t) = (\bar{A} + \bar{a}_t, \bar{B} + \bar{b}_t)$, $t \in [0,1]$, be a family of affine $\mathbb{Z}^2$ actions on $\mathbb{T}^d$ which is of class C^d in the parameter t. Then the following trichotomy holds relative to the pair $(\bar{A}, \bar{B})$:*

(1) $(\bar{A}, \bar{B})$ satisfies (HR) and every action in the family is locally rigid.

(2) $(\bar{A}, \bar{B}) = (A \times \mathrm{Id}_{\mathbb{T}^{d_2}}, B \times \mathrm{Id}_{\mathbb{T}^{d_2}})$ with $d_2 > 0$, $d_1 + d_2 = d$ and if $d_1 \neq 0$, (A, B) is a higher rank action on $\mathbb{T}^{d_1}$. Then, the family (f_t, g_t) is KAM rigid for a generic choice of $\bar{a}_t$ and $\bar{b}_t$. More precisely, if $\bar{a}_t = (a_t, \varphi(t)) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, and if φ satisfies a Pyartli condition with some $\nu > 0$, then there exists $r_0(A, B, d_2) > 0$ and for any $\eta > 0$ there exists $\varepsilon = \varepsilon(\eta, \nu, \|a\|_d, \|b\|_d) > 0$ such that: if the action (f_t, g_t) is perturbed into $(\tilde{f}_t, \tilde{g}_t)$ such that $\|\tilde{f}_\cdot - f_\cdot\|_{d,r_0} \leq \varepsilon$, $\|\tilde{g}_\cdot - g_\cdot\|_{d,r_0} \leq \varepsilon$, then the set of parameters t for which $(\tilde{f}_t, \tilde{g}_t)$ is smoothly conjugated to an affine action is larger than $1 - \eta$.

(3) $(\bar{A}, \bar{B})$ is not as in (1) or (2). The family (f_t, g_t) is not KAM rigid: it can be perturbed into a family of actions so that no element of the perturbed family is conjugated to any affine action.

REMARK 1. The Pyartli condition on φ can be of course repalced by a Pyartli condition on the affine part of any generator of the action. Namely, let $\bar{b}_t = (b_t, \psi(t)) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then the condition is that for some $(n, m) \in \mathbb{Z}^2$ the function $n\varphi(t) + m\psi(t)$ satisfies a Pyartli condition with some $\nu > 0$. The same consequence in Theorem 2(2) then holds with r_0 and ε depending additionally on n and m.

Part (1) of Theorem 2 is proved in [3].

Part (2) is the main result of this paper. As for the Part (3), this is when the affine action has a non trivial rank one factor, and where KAM rigidity does not hold. We will prove part (3) in the next section.

For the clarity of the exposition, the proof of part (2) of Theorem 2 will be first carried in detail only in the case $d_2 = 1$. The generalization to any d_2 is explained in Section 5. Since the proof of Theorem 1 follows essentially the same lines as the proof of Theorem 2 and is in fact easier (no dependance on a parameter t must be taken into consideration in the KAM algorithm), we will only give a detailed proof of the latter and explain in Section 6 the main differences required for the proof of the former.

2.3. **Proof of part (3) of Theorem 2.**

To exclude alternative (1) in Theorem 2, assume that the linear part $(\bar{A}, \bar{B})$ of the given family of actions (f_t, g_t) is not (HR). Since $(\bar{A}, \bar{B})$ is a $\mathbb{Z}^2$ action by toral automorphisms, it preserves the standard Haar measure on the torus. It is proved in [22] that in this case there is a proper $(\bar{A}, \bar{B})$-action invariant subtorus $\mathbb{T}^l$ of $\mathbb{T}^d$ on which the action $(\bar{A}, \bar{B})$ is rank-one. This means that up to a change of coordinates in the acting group $\mathbb{Z}^2$ we may assume that $(\bar{A}, \bar{B})$ acts on $\mathbb{T}^l$ as a $\mathbb{Z}^2$ action generated by (Id, C) where C is some toral automorphism. If for any rank-one factor of $(\bar{A}, \bar{B})$ the map C is the identity, then we are in the alternative (2) of Theorem 2.

This leaves the case when C is not the identity. This is precisely the alternative (3) of Theorem 2. So consider now an affine $\mathbb{Z}^2$ action with linear part (Id, C). Such an action is generated by two commuting maps: $Id + \phi$ and $C + \psi$, where ϕ and ψ are translation vectors. The commutativity condition implies that ϕ is an eigenvector for C with eigenvalue 1.

(a) C has an ergodic factor.

By passing to a factor we may assume without loss of generality that C is ergodic. Then no eigenvalues of C are roots of 1 (see for example [22]). This implies on one hand that $\phi(t) \equiv 0$, and on the other hand that $C + \psi(t)$ has periodic points $x_{t,p}$ of any period p. To create a non-linearizable perturbation of the family $C + \psi(t)$, fix the period p and consider the family of periodic points $x_{t,p}$. Next, it suffices to perturb for every t the eigenvalues of the derivative map at $x_{t,p}$ to avoid C^1 linearizability (this is folklore, see for example Section 2 in [10]). Because the map $t \mapsto x_{t,p}$ is as smooth as ψ, and the eigenvalues of $C + \psi(t)$ do not depend on t, it is clear that the perturbed family can be obtained with the dependence on t as smooth as ψ and as small as we want.

(b) C has no ergodic factors and C is not the identity. By passing to a finite index subgroup of the acting group, we may assume that C has all eigenvalues equal to 1. Assume that C acts on some torus $\mathbb{T}^l$. Then in some basis of $\mathbb{T}^l$ the map C has matrix representation which is upper triangular with 1's on the diagonal and at least one non-zero entry off the diagonal. Recall that $\phi(t)$ is such that $C\phi(t) = \phi(t)$ and we have to perturb $C + \psi$ into a non-linearizable $C_\epsilon + \psi(t)$ that still satisfies $C_\varepsilon\phi(t) = \phi(t)$. This is easy to do as we now illustrate in the case $l = 2$. In this case $C(x,y) = (x + ay, y)$ for some $a \in \mathbb{R}$ and all $(x,y) \in T^2$, and $\phi(t) = (\phi_1(t), 0)$. For any $\epsilon > 0$, define

$$C_\epsilon(x,y) = (x + ay + \epsilon \sin y, y),$$

and observe that for every $t \in [0,1]$ we have that $C_\epsilon + \psi(t)$ is non-linearizable (because $x \mapsto x + ay + \epsilon \sin y + \psi(t)$ cannot be smoothly conjugated to a rotation for every y!). On the other hand, we clearly have $C_\epsilon\phi(t) = \phi(t)$ for every t, hence the family $(Id + \phi(t), C_\varepsilon + \psi(t))$ is indeed a perturbation of the action $(Id + \phi(t), C + \psi(t))$. The higher dimensional case can be dealt with exactly in the same way.

REMARK 2. We mention here that the essential difference between the case when C is parabolic (as in the part (b) above) and when C is the identity, is in the space of invariant distributions of the map $C + \psi$. When C is the identity, whenever ψ is irrational, the space of invariant distributions under the map $C + \psi$ is one-dimensional: it is generated by the Lebesgue measure. In contrast to this, when C is parabolic and ψ is Diophantine, the space of $C + \psi$ invariant distributions is infinite dimensional. This is proved in Section 3 in [11] in case $C + \psi$ acts on T^2, and it can be generalised to any T^l.

2.4. **Reduction to actions which are linear transversally to the elliptic factor.** The subsequent sections contain the proof of part (2) of Theorem 2 in the case when the unperturbed action is purely linear transversally to the elliptic factor. Namely, assume that $f_t = A \times R_{\varphi(t)}$ and $g_t = B \times R_{\psi(t)}$, where $R_{\varphi(t)}$ and $R_{\psi(t)}$ denote translation maps on the circle.

The same arguments extend to the case when the unperturbed action is affine transversally to the elliptic factor. In this case the action is generated by $(A + a_t) \times R_{\varphi(t)}$ and $(B + b_t) \times R_{\psi(t)}$. The proof in this case are identical to the purely linear case, except that certain constants which appear in estimates will be slightly different, without any incidence on the proof. This is explained in Remark 3 at the end of Section 2.

2.5. **Exact statement of part (2) of Theorem 2 in the case of a one dimensional elliptic factor.** Let $d_1 \geq 0$ and, if $d_1 > 0$, let A and B be two commuting toral automorphisms satisfying the (HR) condition. For $\varphi, \psi \in \mathrm{Lip}(I_0, \mathbb{R})$, $I_0 = [0,1]$, let

$$\begin{cases} f_{\varphi(t)}(x,\theta) & = (Ax, R_{\varphi(t)}(\theta)) \\ g_{\psi(t)}(x,\theta) & = (Bx, R_{\psi(t)}(\theta)). \end{cases} \tag{2.3}$$

For $I \in \mathbb{R}$, we denote by $C^{lip,\infty}(I, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ the set of families of smooth maps in the $\mathbb{T}^{d+1}$ variable and Lipschitz in the parameter $t \in I$. We denote by $C_0^{lip,\infty}(I, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ the subset of maps $f \in C^{lip,\infty}(I, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ such that if we write $f_t(z) = (f_t^1(z), f_t^2(z)) \in \mathbb{T}^d \times \mathbb{T}$, then $\int_{\mathbb{T}^{d+1}} f_t^2(z)dz = 0$ for $t \in I$.

Consider

$$\begin{cases} \tilde{f}_t(x,\theta) & = f_{\varphi(t)}(x,\theta) + \Delta f_t(x,\theta) \\ \tilde{g}_t(x,\theta) & = g_{\psi(t)}(x,\theta) + \Delta g_t(x,\theta), \end{cases} \tag{2.4}$$

with $\Delta f, \Delta g \in C_0^{lip,\infty}(I_0, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ and such that $\tilde{f}_t$ and $\tilde{g}_t$ commute for all $t \in I_0$. For $f \in C_0^{lip,\infty}(I_0, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$, we use the notation $\|f\|_{lip(I),r} = \max_{|\iota| \leq r} \mathrm{Lip}(f^{(\iota)})$, where $\mathrm{Lip}(f)$ is the maximum of the supnorm of f and its Lipschitz constant over all $t \in I$. Here $|\iota|$ is the maximal coordinate of the multi-index $\iota \in \mathbb{N}^{d+1}$. We will also use the notation $\|v\|_{0(I),r}$ for the supremum of the usual C^r norms of $v(t)$ as $t \in I$.

Assume that φ satisfies the following transversality condition for some $M > 0$

$$(*) \qquad \max(\|\varphi\|_{lip(I_0)}, \|\psi\|_{lip(I_0)}) \leq M, \quad \inf_{t\in I_0} \varphi'(t) \geq \frac{1}{M}.$$

THEOREM 3. *Let the family (f_t, g_t) be as in* (2.3) *and* $(*)$. *There exists $r_0(A, B) \in \mathbb{N}$ such that for any η there exists $\epsilon_0(A, B, M, \eta) > 0$ such that if* $\max(\|\Delta f\|_{lip(I),r_0}, \|\Delta g\|_{lip(I),r_0}) \leq \epsilon_0$, *then the set of parameters t for which the pair $(\tilde{f}_t, \tilde{g}_t)$ as in* (2.4) *is simultaneously smoothly conjugate to an affine action, has measure larger than* $1 - \eta$.

Note that in case $d_1 = 0$, Theorem 3 becomes a version of KAM rigidity of the Arnol'd family of circle diffeomorphisms, in the context of commuting actions. The result is then an immediate consequence of the KAM rigidity within the Arnol'd family. Indeed, the commutation relation implies that if f_t is conjugated to an irrational rotation, then the same conjugacy linearizes g_t.

2.6. **Plan of the paper.** Sections 3 and 4 below are devoted to the proof of Theorem 3. Sections 5 and 6 explain how this proof is modified to give the proof of part (2) of Theorem 2 and of Theorem 1.

Acknowledgments. The authors are grateful to Artur Avila, Hakan Eliasson, Anatole Katok and Raphael Krikorian for fruitful discussions and suggestions.

3. THE INDUCTIVE STEP

The goal of this section is to prove Proposition 1, that is the main ingredient in the proof of Theorem 3. It consists of the inductive step of the KAM algorithm, or quadratic scheme, that will allow to establish linearizability on a large measure set of parameters of the perturbed action. It essentially states that a perturbation of order ε of an affine action, of which the translation part satisfies Diophantine property up to some order N, can be conjugated to a new affine action plus a perturbation of order ε^2 times some power of N and a rest controlled by inverse powers of N. Indeed, N is the order of truncation that will be applied to the perturbation terms before they are eliminated at first order through solving the linearized equations.

In Section 4, the order of truncation N_n is chosen at each step of the inductive scheme to guarantee the convergence of the algorithm and the linearization on a large measure set of the parameters.

Definition 2. Let $\mathcal{E}(A)$ be the set of eigenvalues of A union 1 (if $d_1 = 0$ then $\mathcal{E}$ is just the set $\{1\}$). For $N \in \mathbb{N}$, define

$$\mathcal{D}(N,A) = \{\alpha \in I_0 \,/\, |\lambda - e^{i2\pi k\alpha}| \geq N^{-3}, \quad \forall \lambda \in \mathcal{E}(A), \forall 0 < |k| \leq N\}.$$

PROPOSITION 1. *Let the family* (f_t, g_t) *be as in* (2.3) *and* $(*)$, *and* $(\tilde{f}_t, \tilde{g}_t)$ *be as in* (2.4). *There exists* $\sigma(A,B)$ *such that if* $N \in \mathbb{N}$ *and* I *is an interval such that* $I \subset \{t \in I_0 \,/\, \varphi(t) \in \mathcal{D}(N,A)\}$, *then there exist* $\tilde{\varphi}, \tilde{\psi} \in \mathrm{Lip}(I,\mathbb{R})$ *and* $h, \widetilde{\Delta f}, \widetilde{\Delta g} \in C_0^{lip,\infty}(I, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ *such that if we write* $H = \mathrm{Id} + h$ *we have that*

$$\begin{aligned} H \circ \tilde{f} &= (f_{\tilde{\varphi}} + \widetilde{\Delta f}) \circ H \\ H \circ \tilde{g} &= (g_{\tilde{\psi}} + \widetilde{\Delta g}) \circ H, \end{aligned} \tag{3.1}$$

with

$$\begin{aligned} \Delta S &\leq C_0 N^{\sigma} \Delta_0 \\ \|h\|_{lip(I),r+1} &\leq C_r S N^{\sigma} \Delta_r + C_r S N^{\sigma} \Delta_0 \Delta_r \\ \widetilde{\Delta}_r &\leq C_r S N^{\sigma} \Delta_0 \Delta_r + C_{r,r'} N^{\sigma + r - r'} \Delta_{r'}, \end{aligned}$$

where:

$$\begin{aligned} S &= \max(\|\varphi\|_{lip(I)}, \|\psi\|_{lip(I)}) \\ \Delta S &= \max(\|\varphi - \tilde{\varphi}\|_{lip(I)}, \|\psi - \tilde{\psi}\|_{lip(I)}) \\ \Delta_r &= \max(\|\Delta f\|_{lip(I),r}, \|\Delta g\|_{lip(I),r}) \\ \widetilde{\Delta}_r &= \max(\|\widetilde{\Delta f}\|_{lip(I),r}, \|\widetilde{\Delta g}\|_{lip(I),r}). \end{aligned}$$

We will reduce the proof of Proposition 1 to the solution of a set of linear equations in the coordinates of h. These equations are solved using Fourier series and part of the solution is obtained with the higher rank techniques as in [3]. Another part is obtained from solving linear equations above a circular rotation. This requires parameter exclusion to insure that the parameters that are kept satisfy adequate arithmetic conditions that allow to control the small divisors.

3.1. **Reduction of the conjugacy step to linear equations.** By substituting $H = id + h$, the first equation in (3.1) becomes:

$$\Delta f - (Df_{\tilde{\varphi}} h - h \circ f_{\varphi}) = f_{\tilde{\varphi}} - f_{\varphi} + \widetilde{\Delta f}(id + h) + E_{L,A}, \tag{3.2}$$

where $E_{L,A} = f_{\tilde{\varphi}}(Id + h) - f_{\tilde{\varphi}} - Df_{\tilde{\varphi}} h - h(f_{\varphi} + \Delta f) + h f_{\varphi}$. The map $Df_{\tilde{\varphi}}$ actually does not depend on $\tilde{\varphi}$, in fact it is the map $\bar{A} = (A, Id)$, where A acts on $\mathbb{R}^d$ and Id acts on $\mathbb{R}$. The second equation in (3.1) is linearized in the same way, so the linearization of (3.1) is the system of equations in h:

$$\begin{aligned} \bar{A}h - h \circ f_\varphi &= \Delta f \\ \bar{B}h - h \circ g_\psi &= \Delta g, \end{aligned} \tag{3.3}$$

where $\bar{B} = (B, Id)$ and $E_{L,B} := g_{\tilde{\psi}}(Id + h) - g_{\tilde{\psi}} - Dg_{\tilde{\psi}}h - h(g_\psi + \Delta g) + hg_\psi$.

Given a pair of commuting automorphisms $\bar{A}$ and $\bar{B}$ we call (λ, μ) a pair of eigenvalues of $(\bar{A}, \bar{B})$ if λ and μ are eigenvalues of $\bar{A}$ and $\bar{B}$ for the same eigenvector.

If A and B are semisimple, then by choosing a proper basis in $\mathbb{R}^d$ in which A and B simultaneously diagonalize, the system (3.3) breaks down into several systems of the following form

$$\begin{aligned} \lambda h - h \circ f_\varphi &= v \\ \mu h - h \circ g_\psi &= w, \end{aligned} \tag{3.4}$$

where λ and μ are a pair of eigenvalues of $A \times \mathrm{Id}$ and $B \times \mathrm{Id}$ and $v, w \in C^{lip,\infty}(I \times \mathbb{T}^{d+1}, \mathbb{R})$. If A and B have non-trivial Jordan blocks, then instead of (3.7), for each Jordan block we would get a system of equations. Lemma 4.4 in [3] shows that this system of equations can be solved inductively in finitely many steps (the number of steps equals the size of a Jordan block), starting from equation of the form (3.7). We will not repeat the argument here, instead we assume throughout that A and B are semisimple and we refer to Lemma 4.4 in [3] for the general case.

3.2. **Reduction of the commutativity relation.** Since f_φ and g_ψ commute and are linear, the equation $(f_\varphi + \Delta f) \circ (g_\psi + \Delta g) = (g_\psi + \Delta g) \circ (f_\varphi + \Delta f)$ reduces to:

$$\bar{A}\Delta g - \Delta g(f_\varphi + \Delta f) = \bar{B}\Delta f - \Delta f(g_\psi + \Delta g).$$

If we push the terms linear in Δf and Δg to the left and all the non-linear terms to the right hand side we obtain

$$\bar{A}\Delta g - \Delta g \circ f_\varphi - \bar{B}\Delta f - \Delta f \circ g_\psi = \Phi, \tag{3.5}$$

where

$$\Phi = \Delta g(f_\varphi + \Delta f) - \Delta g \circ f_\varphi - (\Delta f(g_\psi + \Delta g) - \Delta f \circ g_\psi). \tag{3.6}$$

Similarily to section 3.1, if A and B are semisimple, the equation (3.5) reduce to several equations of the form:

$$(\lambda w - w \circ f_\varphi) - (\mu v - v \circ g_\psi) = \phi. \tag{3.7}$$

3.3. **An approximate solution to** (3.4)**.** The main ingredient in the proof is that the system of linear equations (3.4) can be solved up to an error term that is controlled by Φ which is quadratically small in the perturbation terms $\Delta f, \Delta g$.

LEMMA 1. *For $v, w, \phi \in C^{lip,\infty}(I \times \mathbb{T}^{d+1}, \mathbb{R})$ satisfying* (3.7)*, and $\lambda \neq 1, \mu \neq 1$, if $N \in \mathbb{N}$ and I is an interval such that $I \subset \{t \in I_0 \ / \ \varphi(t) \in \mathcal{D}(N, A)\}$, then there exists $h \in C^{lip,\infty}(I \times \mathbb{T}^{d+1}, \mathbb{R})$ such that:*

$$\|h\|_{lip(I),r+1} \leq C_r S N^\sigma \|v\|_{lip(I),r} + C_r S N^\sigma \|\phi\|_{lip(I),r-2}$$

$$\|v - (\lambda h - h \circ f_\varphi)\|_{lip(I),r} \leq C_{r,r'} N^{d+r-r'} \|v\|_{lip(I),r'} + C_r S N^\sigma \|\phi\|_{lip(I),r-2}$$

$$\|w - (\mu h - h \circ g_\psi)\|_{lip(I),r} \leq C_{r,r'} N^{d+r'-r} \|w\|_{lip(I),r'} + C_r S N^\sigma \|\phi\|_{lip(I),r-2}$$

for all $r' > r \geq 0$ and $\sigma = \sigma(A, B, \lambda, \mu, d)$. The same holds true for $(\lambda, \mu) = (1, 1)$ provided $v, w \in C_0^{lip,\infty}(I \times \mathbb{T}^{d+1}, \mathbb{R})$.

3.4. **Proof of Proposition 1.** Before we give the proof of Lemma 1, we show how it implies Proposition 1.

By applying Proposition 3 from the Appendix to the equation (3.6), we have that

$$\|\Phi\|_{lip(I),r-2} \leq C_r \Delta_0 \Delta_r. \tag{3.8}$$

Since $\Delta f, \Delta g \in C_0^{lip,\infty}(I \times \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ we can apply Lemma 1 to all the coordinates in (3.3) and get h such that

$$\|h\|_{lip(I),r+1} \leq C_r S N^\sigma \Delta_r + C_r S N^\sigma \Delta_0 \Delta_r$$

$$\|\Delta f - (\bar{A} h - h \circ f_\varphi)\|_{lip(I),r} \leq C_r S N^\sigma \Delta_0 \Delta_r + C_{r,r'} N^{d+r-r'} \Delta_{r'}$$

$$\|\Delta g - (\bar{B} h - h \circ g_\psi)\|_{lip(I),r} \leq C_r S N^\sigma \Delta_0 \Delta_r + C_{r,r'} N^{d+r'-r} \Delta_{r'},$$

where the new constant σ is d times the constant σ from Lemma 1. For the bound on h we use Lemma 1 and (3.8) with $r' = r$. In light of (3.2), we take

$$\begin{aligned} \tilde{\varphi} &:= \varphi + Ave(E_{L,A}^2 \circ (\mathrm{Id} + h)^{-1}) \\ \tilde{\psi} &:= \psi + Ave(E_{L,B}^2 \circ (\mathrm{Id} + h)^{-1}). \end{aligned} \tag{3.9}$$

Let

$$\widetilde{\Delta f} = \big((\Delta f_0 - (Ah - h \circ f_\varphi)) - E_{L,A}\big) \circ (\mathrm{Id} + h)^{-1} + f_\varphi - f_{\tilde{\varphi}},$$

and define $\widetilde{\Delta g}$ similarily.

Then we have that $\tilde{\varphi}, \tilde{\psi}$, h and $\widetilde{\Delta f}, \widetilde{\Delta g}$ satisfy the conclusion of Proposition 1. □

The rest of Section 3 is devoted to the proof of Lemma 1.

3.5. **Proof of Lemma 1.**

3.5.1. **Description of the obstructions to the solution of the linearized conjugacy equations.** We describe now the obstructions for solving a single coboundary equation in (3.4). For a fixed $t \in I$ the first equation in (3.4) becomes:

$$\lambda h_t - h_t \circ f_{\varphi(t)} = v_t, \tag{3.10}$$

where $h_t = h(t, \cdot)$ and similarly for v and w. By reducing to Fourier coefficients, for every $(n, m) \in \mathbb{Z}^d \times \mathbb{Z}$ we have:

$$\lambda \sum_{(n,m)} h_{n,m,t}\chi_{n,m}(x,\theta) - \sum h_{n,m,t}\chi_{n,m}(Ax, \theta + \varphi(t)) = \sum v_{n,m,t}\chi_{n,m}(x,\theta)$$

$$\sum_{(n,m)} (\lambda h_{n,m,t} - h_{A^*n,m,t}e^{2\pi i m\varphi(t)}\chi_{n,m}(x,\theta)) = \sum v_{n,m,t}\chi_{n,m}(x,\theta),$$

where $h_{n,m,t}$ denotes the (n, m)-Fourier coefficient of the function h_t, $\chi_{n,m}(x,\theta) = e^{2\pi i(n\cdot x + m\theta)}$, and $A^* = (A^t)^{-1}$. Thus for every (n, m)

$$\lambda h_{n,m,t} - h_{A^*n,m,t}e^{2\pi i m\varphi(t)} = v_{n,m,t}.$$

By denoting: $\lambda_{m,t} := e^{-2\pi i m\varphi(t)}\lambda$ and $v'_{n,m,t} := e^{-2\pi i m\varphi(t)}v_{n,m,t}$, we have

$$\lambda_{m,t}h_{n,m,t} - h_{A^*n,m,t} = v'_{n,m,t}. \tag{3.11}$$

For a fixed m and $n \neq 0$ and for a fixed t the equation (3.11) has been discussed in [3]; the obstructions are precisely defined as well as the construction which allows for removal of all the obstructions (Lemma 4.5 in [3]). The obstructions are:

$$O^A_{n,m}(v_t) = \sum_{k\in\mathbb{Z}} \lambda_{m,t}^{-(k+1)} v'_{A^k n,m,t}, \tag{3.12}$$

where we abuse notation a bit by using A^k to denote the k-th iterate of the dual map A^*. The proof of Lemma 1 relies on two claims. In the first one we solve a system of the type (3.4) provided a set of obstructions computed with the right hand side vanish. In the second claim, we show how the commutation relation allows to modify the right hand side in (3.4) to set the obtructions to 0. Moreover, the modification will be of the order of the "commutation error" Φ in (3.5).

3.5.2. **Solving the linearized equations if all the obstructions vanish.** The goal of this section is to prove the following claim.

Claim 1. *Let v be in $C^{lip(I),\infty}(I,\mathbb{T}^{d+1},\mathbb{R})$ such that for all $t\in I$ and $|m|>N$, $v_{0,m,t}=0$. If for all n,m, and $t\in I$, $n\neq 0$, $O^A_{n,m}(v_t)=0$, and $\mathrm{ave}(v_t)=0$ in the case $\lambda=1$, then there exists a solution h to the equation $\lambda h-h\circ f_\varphi=v$ in $C^{lip(I),\infty}(I\times\mathbb{T}^{d+1},\mathbb{R})$ satisfying*

$$\|h\|_{lip(I),r}\leq C_r S N^3\|v\|_{lip(I),r+\sigma} \tag{3.13}$$

for all $r\geq 0$, where $\sigma=\sigma\{\lambda,d,A\}$. Moreover, if h and v are smooth maps with $h_{0,m,t}=v_{0,m,t}=0$ for $|m|>N$, and with averages zero, such that $\lambda h-h\circ f_\varphi=v$ on $I\times\mathbb{T}^{d+1}$, then h satisfies the estimate (3.13).

Proof of Claim 1. The proof is similar to the proof of the Lemma 4.2 in [3], except that here one extra (isometric) direction causes somewhat greater loss of regularity.

Solution h is defined via its Fourier coefficients $h_{n,m,t}$, each of which can be defined, in case $n\neq 0$, by using one of the two forms:

$$h_{n,m,t}=\sum_{k=0}^{\infty}\lambda_{m,t}^{-(k+1)}v'_{A^kn,m,t}=-\sum_{k=-\infty}^{-1}\lambda_{m,t}^{-(k+1)}v'_{A^kn,m,t}. \tag{3.14}$$

One can use one or the other form to obtain an estimate for the size of $h_{n,m,t}$ depending on whether a non-trivial n is largest in the expanding or in the contracting direction for A. This is completely the same as in [3] and it automatically gives an estimate of the size of $h_{n,m,t}$ with respect to the norm of n. In order to obtain here the full estimate for the C^r norm of h we need to estimate the size of $h_{n,m,t}$ with respect to the norm of (n,m) and this is the only additional detail needed here. But this is not a big problem: since n is non-trivial, after approximately $\log m$ iterations of n by A, the resulting vector will surely be larger than m. We have:

$$\begin{aligned}|h_{n,m,t}|&\leq\sum_{k=0}^{\infty}|\lambda_{m,t}^{-(k+1)}||v'_{A^kn,m,t}|\\&=\sum_{k=0}^{\infty}|\lambda|^{-(k+1)}|v_{A^kn,m,t}|\\&\leq\|v\|_{0(I),r}\sum_{k=0}^{\infty}|\lambda|^{-(k+1)}\|(A^kn,m)\|^{-r}.\end{aligned} \tag{3.15}$$

Take the norm in $\mathbb{Z}^N \times \mathbb{Z}$ to be $\|(n,m)\| = \max\{\|n\|, |m|\}$, where for $n \in \mathbb{Z}^N$, $\|n\|$ is the maximum of euclidean norms of projections of n to expanding, contracting and the neutral directions for A. Let n_{exp} denote the projection of n to the expanding subspace for A. Due to ergodicity of A this projection is non-trivial. For example we say that n is largest in the expanding if $\|n_{exp}\| \geq C\|n\|$ where C is a fixed constant ($C = 1/3$ works). Similarily, we say that n is largest in the contracting (resp. neutral) direction if the projection of n to the contracting (resp. neutral) direction is greater than constant times the norm of n.

Then if ρ denotes the expansion rate for A in the expanding direction for A, we have by the Katznelson Lemma (See for example Lemma 4.1 in [3]):

$$\begin{aligned}\|(A^k n, m)\| &\geq \max\{\|A^k n_{exp}\|, |m|\} \geq \max\{\rho^k \|n_{exp}\|, |m|\} \\ &\geq \max\{C\rho^k \|n\|^{-d}, |m|\} \geq \max\{C\rho^{k-k_0}\rho^{k_0}\|n\|^{-d}, |m|\}.\end{aligned}$$

Since $\rho^k\|n\|^{-d} \geq \|(n,m)\|$ for all $k \geq \frac{d+1}{\ln\rho}\ln\|(n,m)\|$, we have:

$$\|(A^k n, m)\| \geq C\rho^{k-k_0}\|(n,m)\|,$$

for all $k > k_0 = [\frac{d+1}{\ln\rho}\ln\|(n,m)\|]$.

Now if n is largest in the expanding direction for A then for $0 \leq k \leq k_0$: $\|(A^k n, m)\| \geq C\|(n,m)\|$. If n is largest in the neutral direction for A, then for $0 \leq k \leq k_0$: $\|(A^k n, m)\| \geq C(1+k)^{-d}\|(n,m)\|$.

Thus for all $t \in I$ (in the worst case scenario, when $|\lambda| < 1$):

$$\begin{aligned}|h_{n,m,t}| &\leq \|v\|_{0(I),r}\Big(\sum_{k=0}^{k_0}|\lambda|^{-(k+1)}\|(n,m)\|^{-r} + \sum_{k=k_0}^{\infty}|\lambda|^{-(k+1)}(C\rho^k\|(n,m)\|)^{-r}\Big) \\ &\leq \|v\|_{0(I),r}(k_0|\lambda|^{-(k_0+1)}\|(n,m)\|^{-r} + |\lambda|^{k_0}(C\rho^{k_0}\|(n,m)\|)^{-r}\sum_{k=0}^{\infty}|\lambda|^{-k}\rho^{-kr} \\ &\leq C_r\|v\|_{0(I),r}(\|(n,m)\|^{\frac{d+1}{ln\rho}}\log\|(n,m)\|)\|(n,m)\|^{-r} + (C\rho^{k_0}\|(n,m)\|)^{-r}) \\ &\leq C_r\|v\|_{0(I),r}\|(n,m)\|^{-r+\sigma}.\end{aligned}$$

Here $\sigma = 2 + d + a + \delta$, $\delta > 0$, and $a = a(\lambda) = \frac{d+1}{ln\rho} > 0$ in general depends only on the eigenvalues of A. Note that for the convergence of the sum $\sum_{k=0}^{\infty}|\lambda|^{-k}\rho^{-kr}$ it suffices to assume that the regularity r of v is greater than a constant $-\frac{\ln|\lambda|}{\ln\rho}$, which in general depends on eigenvalues of A. We recall that the norm $\|v\|_{0(I),r}$ denotes the supremum of the usual C^r norms of $v(t)$ as $t \in I$.

When n is largest in the contracting direction for A then just as in [3] we repeat the above estimates using the expression $h_{n,m,t} = -\sum_{k=-\infty}^{-1} \lambda_{m,t}^{-(k+1)} v'_{A^k n,m,t}$ for the coefficients $h_{n,m,t}$ instead to obtain the same bound: $|h_{n,m,t}| \leq C_r \|v\|_{0(I),r} \|(n,m)\|^{-r+\sigma}$. The constant σ is now slightly different (changed by a constant) to include the eigenvalues for A in the contracting directions.

Now in the case $n = 0$, and for any non-zero m the equation (3.11) implies:

$$\lambda h_{0,m,t} - h_{0,m,t} e^{2\pi i m \alpha} = v_{0,m,t},$$
$$\mu h_{0,m,t} - h_{0,m,t} e^{2\pi i m \beta} = w_{0,m,t},$$
$$\lambda h_{0,m,t} - h_{0,m,t} e^{2\pi i m \varphi(t)} = v_{0,m,t}.$$

Therefore in this case

$$h_{0,m,t} = \frac{v_{0,m,t}}{\lambda - e^{2\pi i m \varphi(t)}}. \tag{3.16}$$

Thus for $|\lambda| \neq 1$ we have that for all $t \in I$:

$$|h_{0,m,t}| \leq (|\lambda| - 1)^{-1} \|v\|_{0(I),r} |m|^{-r}.$$

In the case $|\lambda| = 1$ this is a small divisor problem. When $t \in I$ we have $\varphi(t) \in \mathcal{D}(N)$ and thus for $|m| \leq N$ we have:

$$|h_{0,m,t}| \leq N^3 |v_{0,m,t}| \leq \|v\|_{0(I),r} N^3 |m|^{-r}.$$

Since for $|m| > N$, $v_{0,m} = 0$, we define $h_{0,m} = 0$ for $|m| > N$.

Accumulating all the estimates, we have for all $t \in I$:

$$|h_{n,m,t}| \leq C_r \|v\|_{0(I),r} N^3 \|(n,m)\|^{-r+\sigma}.$$

All this implies that the function h defined via its Fourier coefficients $h_{n,m,t}$ satisfies the equation $\lambda h - h \circ f_\psi = v$ and the estimate:

$$\|h\|_{0(I),r} \leq C_r N^3 \|v\|_{0(I),r+\sigma}, \tag{3.17}$$

for all $r > r_0$. Here σ is a fixed constant, $\sigma = d + 2 + \max\{|\lambda|, |\lambda|^{-1}\}$, which in our set-up depends only on the eigenvalues of A and the dimension of the torus.

We estimate now h in the direction of the parameter t. First we can characterise $x \in C^{lip,\infty}(I, \mathbb{T}^{d+1}, \mathbb{R})$ by a property of Fourier coefficients of x. Let $\Delta x := x_t - x_{t'}$, and similarly $\Delta x_{n,m} = x_{n,m,t} - x_{n,m,t'}$. Namely, $x \in C^{lip,s}(I, \mathbb{T}^{d+1}, \mathbb{R})$ implies not only that that $x_{n,m,t}$ decay faster than $\|(n,m)\|^{-s}$ but also from $\|\Delta x^{(s)}\|_0 \leq L_s |t - t'|$ we get that $|\Delta x_{n,m}| \leq C_s\|(n,m)\|^{-s}|t-t'|$ for some constant C_s. It is then easy to check that $|\Delta x_{n,m}| \leq C_s\|(n,m)\|^{-s-d-1}|t-t'|$ suffices for $x \in C^{lip,s}(I, \mathbb{T}^{d+1}, \mathbb{R})$.

By using (3.14) (denote for simplicity by $\Sigma^{\pm}$ positive or negative sum in (3.14)) we have for $n \neq 0$:

$$|\Delta h_{n,m}| = |\Sigma^{\pm}\lambda^{-(k+1)}(e^{2\pi i k m \varphi(t)} v_{A^k n,m,t} - e^{2\pi i k m \varphi(t')} v_{A^k n,m,t'})|$$
$$= |\Sigma^{\pm}\lambda^{-(k+1)}((e^{2\pi i k m \varphi(t)} - e^{2\pi i k m \varphi(t')}) v_{A^k n,m,t} + e^{2\pi i k m \varphi(t')} \Delta v_{A^k n,m,t})|$$
$$\leq (2\pi\|\varphi\|_{lip(I)}\|v\|_{0(I),r} + \|v\|_{lip(I),r})|t-t'|\Sigma^{\pm}|\lambda|^{-(k+1)}|k|\|(A^k n,m)\|^{-r+1}.$$

From the discussion following (3.15) we have that for every (n,m), $n \neq 0$, either the positive or the negative sum in the last expression above can be bounded by $C_r\|n,m\|^{-r+\sigma+1}$. When $n=0$ from (3.16) and for $t,t' \in I$ it is clear that $\Delta h_{0,m} \leq CN^3\Delta v_{0,m}$. This gives the bound for the Lipschitz constant for any r-th derivative of h which combined with (3.17) implies $\|h\|_{lip(I),r-\sigma-2-d} \leq C_r N^3 S\|h\|_{lip(I),r}$.

For the second part of the claim, if h and v are smooth and satisfy $\lambda h - h \circ f_\varphi = v$ for $t \in \mathcal{I}$ then for $n \neq 0$ the obstructions $O^A_{n,m}(v_t)$ are all zero. Thus if v satisfies in addition that $v_{0,m,t} = 0$ for $|m| > N$ then by the first part of the Claim 1 there exists h' such that $\lambda h' - h' \circ f_\varphi = v$ on I and satisfies the estimate (3.13). Then for $h'' = h - h'$, $\lambda h'' = h'' \circ f_\varphi$ on I. But this implies $h'' = 0$ in case $\lambda \neq 1$, or it is constant in case $\lambda = 1$. However, by construction, h' has average 0, and so does h by assumption, so in any case $h = h'$ on I. This implies that h satisfies the estimate (3.13). Claim 1 is thus proved. □

3.5.3. **Removing the obstructions.** We show here how to exploit the commutation relation to modify the right hand side in (3.4) to set the obtructions to 0. Moreover, the modification will be of the same "quadratic" order of the "commutation error" Φ in (3.5). This is the content of the following claim. We first need to define, following [3], a special integer on the orbit of any integer that will be used to zero the obstructions $O^A_{n,m}(v_t)$ defined in (3.12).

Definition 3 ([3]). For every $n \in \mathbb{Z}^d$ there exists a point n^* on the orbit $\{A^k n\}_{k\in\mathbb{Z}}$ such that the projection of n to the contracting subspace of A is larger than the projection to the expanding subspace of A, and for An the opposite holds: projection of An to the contracting subspace of A is smaller than the projection to the expanding subspace of A. For each n choose an n^* on the orbit of n with this property.

Claim 2. *Assume that for all $t \in I$ the following holds:*

$$(\lambda w_t - w_t \circ f_{\varphi(t)}) - (\mu v_t - v_t \circ g_{\psi(t)}) = \phi_t \tag{3.18}$$

and $v_{0,m,t} = w_{0,m,t} = \phi_{0,m,t} = 0$ for $|m| > N$. Define $\tilde{v}_t$ by

$$\tilde{v}_{n,m,t} = \begin{cases} O^A_{n,m}(v_t),\ n \neq 0, n = n^* \\ 0,\ otherwise. \end{cases}$$

Then:

(1) For $n \neq 0$, $O^A_{n,m}(v_t - \tilde{v}_t) = 0$.

(2) $\|\tilde{v}\|_{lip(I),r} \leq C_r N^3 S \|\phi\|_{lip(I),r+\sigma}$, where $\sigma = \sigma(A, B, \lambda, \mu, d)$ and $r \geq 0$.

Proof of claim 2.

(1) This is immediate from the definition of $O^A_{n,m}$ and $\tilde{v}_t$.

(2) In Fourier coefficients (3.18) becomes:

$$(\lambda w_{n,m,t} - w_{An,m,t} e^{2\pi i m \varphi(t)}) - (\mu v_{n,m,t} - v_{Bn,m,t} e^{2\pi i m \psi(t)}) = \phi_{n,m,t}.$$

This implies that for non-zero n the obstructions $O^A_{n,m}$ for

$$(\mu v_{n,m,t} - v_{Bn,m,t} e^{2\pi i m \psi(t)}) + \phi_{n,m,t}$$

are trivial. Therefore $O^A_{n,m}(v_t)$ satisfies the equation:

$$\mu O^A_{n,m}(v_t) - e^{2\pi i m \psi(t)} O^A_{Bn,m}(v_t) = O^A_{n,m}(\phi_t), \tag{3.19}$$

where $O^A_{n,m}(v_t)$ and $O^A_{n,m}(\phi_t)$ are defined as in (3.12). From this, by backward and forward iteration by B, one obtains two expressions for $O^A_{n,m}(v_t)$:

$$\begin{aligned} O^A_{n,m}(v_t) &= \sum_{l \geq 0} \mu_{m,t}^{-(l+1)} e^{-2\pi i m \psi(t)} O^A_{B^l n,m}(\phi_t) \\ &= -\sum_{l < 0} \mu_{m,t}^{-(l+1)} e^{-2\pi i m \psi(t)} O^A_{B^l n,m}(\phi_t), \end{aligned}$$

where $\mu_{m,t} := e^{-2\pi i m \psi(t)} \mu$.

It is proved in Lemma 4.5 in [3] that if every $A^k B^l$ for $(k,l) \neq (0,0)$ is ergodic, and if $n = n^*$ then either for $l > 0$ for $l < 0$, the term $\|(B^l A^k n, m)\|$ has exponential growth in (l,k) for $\|(l,k)\|$ larger than some $C \log |n|$, and polynomial growth for $\|(l,k)\|$ less than $C \log |n|$. Hence, for $n = n^*$, it follows exactly as in Lemma 4.5 [3], that either one or the other sum above are comparable to the size of $\|\phi_t\|_r\|(n,m)\|^{-r+\sigma}$. Here σ is a constant which depends only on A, B and the dimension d. Therefore in case $n \neq 0$ for all $t \in I$

$$|\tilde{v}_{n,m,t}| = |O^A_{n,m}(v_t)| \leq C_r\|\phi\|_{0(I),r}\|(n,m)\|^{-r+\sigma}. \tag{3.20}$$

This implies the $\|\cdot\|_{0(I),r}$-norm estimate for $\tilde{v}$. To obtain the estimate in the t direction, just as in the Claim 1, we look at $\Delta\tilde{v}_{n,m}$. For $n \neq 0, n = n^*$:

$$\Delta\tilde{v}_{n,m} = O^A_{n,m}(v_t - v'_t) =$$
$$\Sigma^{\pm}_l\Sigma_k\mu^{-(l+1)}\lambda^{-(k+1)}e^{-2\pi i m(l\psi(t)+k\varphi(t))}(\phi_{B^lA^kn,m,t} - \phi_{B^lA^kn,m,t'})$$
$$+ \Sigma^{\pm}_l\Sigma_k\mu^{-(l+1)}\lambda^{-(k+1)}(e^{-2\pi i m(l\psi(t)+k\varphi(t))} - e^{-2\pi i m(l\psi(t')+k\varphi(t'))})\phi_{B^lA^kn,m,t'}.$$

If φ and ψ are Lipschitz and ϕ is in $C^{lip(I),r}$, we have:

$$|\Delta\tilde{v}_{n,m}| \leq \|\phi\|_{lip(I),r}|t-t'|\Sigma^{\pm}_l\Sigma_k|\mu|^{-(l+1)}|\lambda|^{-(k+1)}\|(B^lA^kn,m)\|^{-r}$$
$$+ 2\pi S|t-t'|\|\phi\|_{0(I),r}\Sigma^{\pm}_l\Sigma_k\mu^{-(l+1)}\lambda^{-(k+1)}|k||l|\|(B^lA^kn,m)\|^{-r+1}.$$

Now the same argument as above (based on Lemma 4.5 [3]) implies that for every $n = n^*$ one of the sums (for $l > 0$ or $l < 0$) $\Sigma^{\pm}_l\Sigma_k\mu^{-(l+1)}\lambda^{-(k+1)}|k||l|\|(B^lA^kn,m)\|^{-r+1}$ can be bounded by $\|(n,m)\|^{-r+\sigma}$, where σ is a constant depending on A, B, λ, μ and d. This implies

$$|\Delta\tilde{v}_{n,m}| \leq CS\|\phi\|_{lip(I),r}\|(n,m)\|^{-r+\sigma}|t-t'|.$$

Taking into account all the estimates above, we have:

$$\|\tilde{v}\|_{lip(I),r} \leq C_rN^3S\|\phi\|_{lip(I),r+\sigma},$$

with σ fixed depending only on A, B, λ and d. Claim 2 is thus proved. $\square$

3.5.4. We proceed now with the proof of Lemma 1. Given v, w such that $(\lambda w - w \circ f_\varphi) - (\mu v - v \circ g_\psi) = \phi$, first truncate v_t to T_Nv_t for all $t \in I$. We choose the same N for all $t \in I$. The truncation and the residue satisfy the following estimates for every t and $r \leq r'$

$$\begin{aligned}\|T_Nv_t\|_{r'} &\leq C_{r,r'}N^{r'-r+d}\|v_t\|_r\\ \|R_Nv_t\|_r &\leq C_{r,r'}N^{r-r'+d}\|v_t\|_{r'}.\end{aligned} \tag{3.21}$$

Since the same truncation is used for all t, it is easy to check that

$$\|T_Nv\|_{lip(I),r'} \leq C_{r,r'}N^{r'-r+d}\|v\|_{lip(I),r}$$
$$\|R_Nv\|_{lip(I),r} \leq C_{r,r'}N^{r-r'+d}\|v\|_{lip(I),r'}.$$

Now the Claim 2 applies to $T_N v$. It gives $\widetilde{T_N v}$ such that for $T_N v - \widetilde{T_N v}$ the obstructions $O^A_{n,m}(T_N v_t - \widetilde{T_N v_t})$ vanish for $n \neq 0$ and

$$\|\widetilde{T_N v}\|_{lip(I),r} \leq C_r N^3 S \|T_N \phi\|_{lip(I),r+\sigma}.$$

Notice that $\widetilde{T_N v_t}$ by construction has all $(0, m, t)$-Fourier coefficients equal to zero for $|m| > N$. Thus the Claim 1 can be applied to $T_N v - \widetilde{T_N v}$. Therefore there exists $h \in C^\infty(\mathcal{A} \times \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ as in Claim 1 such that for all $t \in \mathcal{A}$:

$$T_N v_t - \widetilde{T_N v_t} = \lambda h_t - h_t \circ f_t$$

and

(3.22)
$$\begin{aligned}\|h\|_{lip(I),r+1} &\leq C_r N^3 S \|T_N v - \widetilde{T_N v}\|_{lip(I),r+1+\sigma} \\ &\leq C_r N^3 S (\|T_N v\|_{lip(I),r+1+\sigma} + C_r N^3 \|T_N \phi\|_{lip(I),r+1+2\sigma}) \\ &\leq C_r S N^{4+\sigma} \|v\|_{lip(I),r} + C_r S N^{6+2\sigma} \|\phi\|_{lip(I),r-2}.\end{aligned}$$

Also

$$\begin{aligned}\|v - (\lambda h - h \circ f)\|_{lip(I),r} &= \|R_N v + \widetilde{T_N v}\|_{lip(I),r} \\ &\leq \|R_N v\|_{lip(I),r} + C_r S N^3 \|T_N \phi\|_{lip(I),r+\sigma} \\ &\leq C_{r,r'} N^{r-r'+d} \|v\|_{lip(I),r'} + C_r S N^{5+\sigma} \|\phi\|_{lip(I),r-2}.\end{aligned}$$

Now to estimate $w - (\mu h - h \circ g)$ we use:

$$\begin{aligned}&(\lambda w - w \circ f) - (\mu v - v \circ g) = \phi \\ &(\lambda w - w \circ f) - (\mu T_N v - T_N v \circ g) - (\mu R_N v - R_N v \circ g) = \phi \\ &(\lambda w - w \circ f) - (\mu (T_N v - \widetilde{T_N v}) - (T_N v - \widetilde{T_N v}) \circ g) \\ &- (\mu \widetilde{T_N v} - \widetilde{T_N v} \circ g) - (\mu R_N v - R_N v \circ g) = \phi \\ &(\lambda w - w \circ f) - (\mu (\lambda h - h \circ f) - (\lambda h - h \circ f) \circ g) - (\mu \widetilde{T_N v} - \widetilde{T_N v} \circ g) \\ &- (\mu R_N v - R_N v \circ g) = \phi \\ &\lambda (w - (\mu h - h \circ g)) - (w - (\mu h - h \circ g)) \circ f = \\ &\phi + (\mu \widetilde{T_N v} - \widetilde{T_N v} \circ g) - (\mu R_N v - R_N v \circ g).\end{aligned}$$

This implies:

$$\lambda(T_N w - (\mu h - h\circ g)) - (T_N w - (\mu h - h\circ g))\circ f =$$
$$\phi + (\mu\widetilde{T_N v} - \widetilde{T_N v}\circ g) - (\mu R_N v - R_N v\circ g) - (\mu R_N w - R_N w\circ g) =$$
$$T_N\phi + (\mu\widetilde{T_N v} - \widetilde{T_N v}\circ g).$$

Since both $T_N w - (\mu h - h\circ g)$ (by construction of h) and $T_N\phi + (\mu\widetilde{T_N v} - \widetilde{T_N v}\circ g)$ (by construction of $\widetilde{T_N v}$), satisfy that their $(0, m, t)$ Fourier coefficients are zero for $|m| > N$, the second part of the Claim 1 applies and gives an estimate for $T_N w - (\mu h - h\circ g)$:

$$\begin{aligned}\|T_N w - (\mu h - h\circ g)\|_{lip(I),r} &\leq C_r S N^3 \|T_N\phi + (\mu\widetilde{T_N v} - \widetilde{T_N v}\circ g)\|_{lip(I),r+\sigma}\\ &\leq C_r S N^{5+2\sigma}\|\phi\|_{lip(I),r-2}.\end{aligned}$$

Therefore:

$$\begin{aligned}\|w - (\mu h - h\circ g)\|_{lip(I),r} &\leq C_r S N^{5+2\sigma}\|\phi\|_{lip(I),r-2} + \|R_N w\|_{lip(I),r}\\ &\leq C_r S N^{5+2\sigma}\|\phi\|_{lip(I),r-2} + C_{r,r'} N^{d+r'-r}\|w\|_{r'}.\end{aligned}$$

Finally we can redefine the constant σ by $\sigma := 6 + 2\sigma$. This completes the proof of Lemma 1. □

REMARK 3. In case when the unperturbed actions are affine transversally to the elliptic factor, rather than linear, they are generated by $(A + a_t) \times R_{\varphi(t)}$ and $(B + b_t) \times R_{\psi(t)}$. In this case, Lemma 1 implies in the same way the Proposition 1. The only difference appears in the beginning of the proof of Lemma 1. Namely, the number $\lambda_{m,t}$ should be replaced with $\lambda_{m,n,t} = e^{-i2\pi(m\varphi(t)+\langle n,a_t\rangle)}\lambda$. This change does not effect in a substantial way any of the subsequent arguments in Lemma 1. It also does not affect substantially any estimates obtained in Lemma 1 because the two constants $\lambda_{m,t}$ and $\lambda_{m,n,t}$ have the same absolute value.

4. THE KAM SCHEME

The goal of this section is to derive Theorem 3 using a quadratic KAM like scheme, with parameter exclusion. More precisely, we apply inductively Proposition 1 to conjugate the perturbed family action closer and closer to an affine one. Doing so, we have to discard some parameters at each step n of the induction, to guarantee the validity of the Diophantine condition up to an order N_n that will be chosen below. The following simple lemma will allow to us to control the measure of the remaining set of parameters after successive exclusions.

LEMMA 2. *Let $M > 0$. There exists $N_0(M)$ such that if $N > N_0$ and $\tilde{N} = N^{3/2}$ and if I is an interval of size $1 \geq |I| \geq 1/(2MN^2)$ and if $M^{-1} < \varphi'(t) < M$ for every $t \in I$, then there exists a union of disjoint intervals $\mathcal{U} = \{\tilde{I}_j\}$ such that $\varphi(\tilde{I}_j) \in \mathcal{D}(\tilde{N}, A)$ and $\tilde{I}_j \subset I$ and $|\tilde{I}_j| \geq 1/(2M\tilde{N}^2)$ and $\sum |\tilde{I}_j| \geq (1 - 2dM^2\tilde{N}^{-1})|I|$.*

Proof. We just observe that the set of $t_k \in I$ such that $\lambda + e^{i2\pi\varphi(t)} = 0$ with $\lambda \in \mathcal{E}(A)$ and $k \leq \tilde{N}$ consists of at most $d([M\tilde{N}^2|I|] + 2)$ points separated one from the other by at least $1/(M\tilde{N}^2)$. Excluding from I the intervals $[t_k - M/\tilde{N}^3, t_k + M/\tilde{N}^3]$ leaves us with a collection of intervals of size greater than $1/(2M\tilde{N}^2)$ of total length $|I| - d([M\tilde{N}^2|I|] + 2)M/\tilde{N}^3 \geq (1 - 2dM^2\tilde{N}^{-1})|I|$. □

We now describe the inductive scheme that we obtain from an iterative application of Proposition 1. For the new translation frequencies φ_n and ψ_n that will appear during the induction, we will require the following transversality condition:

$$\text{(C1)} \quad \max(\|\varphi_n\|_{lip(\mathcal{A}_n)}, \|\psi_n\|_{lip(\mathcal{A}_n)}) \leq 2M, \quad \inf_{t \in \mathcal{A}_n} \varphi_n'(t) \geq \frac{1}{2M}.$$

Recall that by our hypothesis $(*)$, we have that φ and ψ satisfy (C1) with M instead of $2M$. During the induction, we will see that φ_n and ψ_n, where they are defined, remain very close to the original φ and ψ, and this will guarantee inductively the validity of (C1). We will come back to this below. But we first introduce the truncation order N_n and the notations for the parameter exclusion.

Let $N_0 \geq N_0(M)$ of Lemma 2 and define for $n \geq 1$,

$$N_n = N_{n-1}^{\frac{3}{2}}.$$

Observe that Lemma 2 implies that if at step n, we have a set $\mathcal{A}_n \subset [0, 1]$ that is a collection of intervals of sizes greater than $1/(4MN_n^2)$ and φ_n and ψ_n are functions satisfying (C1) on $\mathcal{A}_n$, then there exists $\mathcal{A}_{n+1}$ that is a collection of intervals with sizes greater than $1/(4MN_{n+1}^2)$ such that $\varphi_n(\mathcal{A}_{n+1}), \psi_n(\mathcal{A}_{n+1}) \subset \mathcal{D}(N_{n+1})$ and $\lambda(\mathcal{A}_{n+1}) \geq (1 - 8dM^2N_{n+1}^{-1})\lambda(\mathcal{A}_n)$.

At step n we have $f_n = f_{\varphi_n} + \Delta f_n$,$g_n = g_{\psi_n} + \Delta g_n$ defined for $t \in \mathcal{A}_n$, with $\mathcal{A}_{-1} = [0, 1]$. We denote $\varepsilon_{n,r} = \max(\|\Delta f_n\|_{lip(\mathcal{A}_n),r}, \|\Delta g_n\|_{lip(\mathcal{A}_n),r})$. We obtain h_n and φ_{n+1} and ψ_{n+1} defined on $\mathcal{A}_{n+1}$ such that

$$H_n f_n H_n^{-1} = f_{\varphi_{n+1}} + \Delta f_{n+1}$$
$$H_n g_n H_n^{-1} = g_{\psi_{n+1}} + \Delta g_{n+1}$$

with $\Delta f_{n+1}, \Delta g_{n+1} \in C_0^{lip(\mathcal{A}_{n+1}),\infty}(I, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$. Also, if we denote $\xi_{n,r} = \|h_n\|_{lip(\mathcal{A}_{n+1}),r+1}$ and $\nu_n = \max(\|\varphi_{n+1} - \varphi_n\|_{lip(\mathcal{A}_{n+1})}, \|\psi_{n+1} - \psi_n\|_{lip(\mathcal{A}_{n+1})})$, we have from Proposition 1 that

$$\xi_{n,r} \leq C_r \gamma_n N_n^{\sigma} \varepsilon_{n,r} \tag{4.1}$$

$$\nu_n \leq \varepsilon_{n,0} \tag{4.2}$$

$$\varepsilon_{n+1,r} \leq C_r \gamma_n N_n^{\sigma} \varepsilon_{n,0} \varepsilon_{n,r} + C_{r,r'} \gamma_n N_n^{\sigma+r-r'} \varepsilon_{n,r'} \tag{4.3}$$

with $\gamma_n = (1 + S_n + \varepsilon_{n,0})^{\sigma}$.

If during the induction we can insure that $\sum \varepsilon_{n,0} < M/100$ we can conclude from (4.2) that for all n, φ_n and ψ_n satisfy on $\mathcal{A}_n$ the inductive condition (C1). Lemma 2 then insures that $\mathcal{A}_{n+1}$ is well defined and $\lambda(\mathcal{A}_{n+1}) \geq (1 - 8dM^2 N_{n+1}^{-1})\lambda(\mathcal{A}_n)$. To be able to apply the inductive procedure we also have to check that H_n is indeed invertible. This is insured if during the induction we have

$$\xi_{n,0} < \frac{1}{2}. \tag{C2}$$

We call the two conditions (C1) and (C2), the inductive conditions.

The proof that the scheme (4.1)–(4.3) converges provided an adequate control on $\varepsilon_{0,0}$ and $\varepsilon_{r_0,0}$ for a sufficiently large r_0, is classical, but we provide it for completeness.

LEMMA 3. *Let $\alpha = 4\sigma + 2$, $\beta = 2\sigma + 1$, and $r_0 = [8\sigma + 5]$. If $S_n, \xi_{n,r}, \varepsilon_{n,r}$ satisfy (4.1)–(4.3), there exists $\bar{N}_0(\sigma)$ such that if $N_0 = \bar{N}_0 M$ and*

$$\varepsilon_{0,0} \leq N_0^{-\alpha}, \quad \varepsilon_{0,r_0} \leq N_0^{\beta},$$

then for any n the inductive conditions (C1) and (C2) are satisfied and in fact $\varepsilon_{n,0} \leq N_n^{-\alpha}$, $\xi_{n,0} \leq N_n^{-\sigma}$, and for any $s \in \mathbb{N}$, there exists $\bar{C}_r$ such that $\max(\varepsilon_{n,s}, \xi_{n,s}) \leq \bar{C}_s N_n^{-1}$.

Proof. We first prove by induction that for every n, $\varepsilon_{n,0} \leq N_n^{-\alpha}$ and $\varepsilon_{n,r_0} \leq N_n^{\beta}$, provided $\bar{N}_0(\sigma)$ is chosen sufficiently large.

Assuming the latter holds for every $i \leq n$, the inductive hypothesis (C1) and (C2) can be checked up to n immediately from (4.1) and (4.2). Now, (4.3) applied with $r = 0$ and $r' = r_0$ yields

$$\begin{aligned}\varepsilon_{n+1,0} &\leq C_0 N_n^{\sigma}(2 + M)^{\sigma} N_n^{-2\alpha} + C_{0,r_0} N_n^{\sigma - r_0} N_n^{\beta} \\ &\leq N_{n+1}^{-\alpha}\end{aligned}$$

provided $\bar{N}_0(\sigma)$ is sufficiently large.

On the other hand, applying (4.3) with $r' = r = r_0$ yields

$$\begin{aligned}\varepsilon_{n+1,r_0} &\leq C_{r_0} N_n^{\sigma}(2+M)^{\sigma} N_n^{-\alpha} N_n^{\beta} + C_{r_0,r_0} N_n^{\sigma} N_n^{\beta} \\ &\leq N_{n+1}^{\beta}\end{aligned}$$

provided $\bar{N}_0(\sigma)$ is sufficiently large.

To prove the bound on $\varepsilon_{n,s}$ we start by proving that for any s, there exist $\tilde{C}_s$ and n_s such that for $n \geq n_s$ we have that $\varepsilon_{n,s} \leq \tilde{C}_s N_n^{\beta}$. Let indeed n_s be such that $N_{n_s}^{-1/10}((1+M)^{\sigma} C_s + C_{s,s}) < 1$. Let $\tilde{C}_s$ be such that $\varepsilon_{n_s,s} \leq \tilde{C}_s N_{n_s}^{\beta}$. We show by induction that $\varepsilon_{n,s} \leq \tilde{C}_s N_n^{\beta}$ for every $n \geq n_s$. Assume the latter true up to n and apply (4.3) with $r = r' = s$ to get

$$\begin{aligned}\varepsilon_{n+1,s} &\leq C_s N_n^{\sigma}(1+M)^{\sigma} N_n^{-\alpha} \varepsilon_{n,s} + C_{s,s} N_n^{\sigma} \varepsilon_{n,s} \\ &\leq N_n^{\sigma+1/10} \varepsilon_{n,s} \\ &\leq \tilde{C}_s N_n^{\sigma+1/10+\beta} \leq \tilde{C}_s N_{n+1}^{\beta}.\end{aligned}$$

We will now bootstrap on our estimates as follows. Let $s'(s) = s + [\sigma + \beta + \frac{3}{2}(\sigma+1)] + 1$, and define $\tilde{n}_s = \max(n_s, n_{s'})$. Let $\bar{C}_s$ be such that $\varepsilon_{\tilde{n}_s,s} \leq \bar{C}_s N_n^{-\sigma-1}$. We will show by induction that for any $n \geq \tilde{n}_s$ we have that $\varepsilon_{n,s} \leq \bar{C}_s N_n^{-\sigma-1}$. Indeed, apply (4.3) with $r = s$ $r' = s'$ to get

$$\begin{aligned}\varepsilon_{n+1,s} &\leq \bar{C}_s C_s N_n^{\sigma}(1+M)^{\sigma} N_n^{-\alpha} N_n^{-\sigma-1} + C_{s,s'} \tilde{C}_{s'} N_n^{\beta} N_n^{\sigma+s-s'} \\ &\leq \bar{C}_s N_{n+1}^{-\sigma-1}\end{aligned}$$

if n_s was chosen sufficiently large.

Finally, (4.1) yields that for $n \geq \tilde{n}_s$, $\xi_{n,s} \leq C'_s N_n^{-1}$. □

Proof of Theorem 3. The sets $\mathcal{A}_n$ are decreasing and we let $\mathcal{A}_\infty = \liminf \mathcal{A}_n$. The result of Lemma 3 implies that

$$\lambda(\mathcal{A}_\infty) \geq \Pi(1 - 8dM^2 N_{n+1}^{-1}) \geq 1 - \eta$$

if $N_0 \geq N_0(\eta)$. On $\mathcal{A}_\infty$, φ_n and ψ_n converge in the Lipschitz norm and the maps $H_n \circ \ldots \circ H_1, H_n^{-1} \circ \ldots \circ H_1^{-1}$ converge in the $C^{lip,\infty}$ norm to some G, G^{-1} such that $G f_\varphi G^{-1} = f_{\varphi_\infty}$, $G g_\psi G^{-1} = g_{\psi_\infty}$, where $(\varphi_\infty, \psi_\infty) = \lim_{n\to\infty}(\varphi_n, \psi_n)$. □

5. PROOF OF THEOREM 2 IN THE CASE OF HIGHER DIMENSIONAL ELLIPTIC FACTORS ($d_2 > 1$)

Define instead of the set $\mathcal{D}(N, A)$ of Section 3 the following

$$\mathcal{D}(N, A) = \{\alpha \in \mathbb{T}^{d_2} \ / \ |\lambda + e^{i2\pi(k,\alpha)}| \geq N^{-b}, \\ \forall \lambda \in \mathcal{E}(A), \forall k \in \mathbb{Z}^{d_2} - \{0\}, \|k\| \leq N\}$$

where $b = 30d_2^2$. Instead of Lemma 2 we have the following more general statement.

LEMMA 4. *Let $\nu > 0$. There exists $N_0(\nu, d_2)$ such that if $N > N_0$ and if I is an interval of size $1 \geq |I| \geq 1/N^a$, $a = 4d_2 + 20$, and if $\varphi : I \to \mathbb{T}^{d_2}$ satisfies a Pyartli condition with constant ν, then for $\tilde{N} = N^{3/2}$, there exists a union of disjoint intervals $\mathcal{U} = \{\tilde{I}_j\} \subset I$ such that $\tilde{I}_j \in \mathcal{D}(\tilde{N}, A)$ and $|\tilde{I}_j| \geq 1/\tilde{N}^a$ and $\sum |\tilde{I}_j| \geq (1 - \tilde{N}^{-1})|I|$.*

Proof. The proof is a direct consequence of the Pyartli condition and a repeated application of the intermediate value theorem. We just deal with case $\lambda = 1$ the other cases being similar. More precisely, for any fixed k, $\|k\| \leq \tilde{N}$, after excluding d_2 intervals of size $1/N^{2a}$ from I we get that $|(k, \varphi')| \geq N^{-2a(d_2+1)}$. After further excluding $\mathcal{O}(\tilde{N})$ intervals of size $N^{2a(d_2+1)}\tilde{N}^{-b}$ we remain with intervals on which $\|(k, \varphi)\| \geq \tilde{N}^{-b}$. We then apply this procedure for every $k \in \mathbb{Z}^{d_2}$ such that $0 < \|k\| \leq \tilde{N}$, then further eliminate all the remaining intervals that are smaller than $\tilde{N}^{-a}$, and finally observe that the remaining part of I is a union of intervals satisfying the conditions of the lemma. □

The effect of changing the exponent in the definition of $\mathcal{D}(N, A)$ just modifies $\sigma(A, B)$ of Proposition 1 to make it $\sigma(A, B, d_2)$. This is because in (3.16) the small divisor (in the case $|\lambda| = 1$) becomes $\frac{1}{|\lambda - e^{2\pi i(m,\varphi(t))}|} \leq N^b$ if $m \in \mathbb{Z}^{d_2}$ is such that $|m| \leq N$. The rest of the proof of Proposition 1 is identical to the case $d_2 = 1$, except that everywhere the Lipschitz norm in the parameter direction should be replaced by the C^{d_2} norm. The control of the C^{d_2} in the parameter direction is required to maintain the Pyartli condition during the KAM algorithm. Indeed, since φ satisfies an initial Pyartli condition with constant ν, then similarly to what was done in the case $d_2 = 1$, we can insure in the KAM scheme that a Pyartli condition with a fixed constant $\nu/2$ is satisfied by the functions φ_n, provided the perturbation ε is sufficiently small.

6. Proof of Theorem 1

Let A, B, α, β and f, g be as in the statement of Theorem 1. Let us momentarily assume that $\alpha \in \mathrm{DC}(\tau, \gamma, A)$ that is $|\lambda - e^{i2\pi(k,\alpha)}| > \frac{\gamma}{|k|^\tau}$ for every non zero vector $k \in \mathbb{Z}^{d_2}$ and every $\lambda \in \mathcal{E}(A)$. This clearly plays a similar role to $\varphi(t) \in \mathcal{D}(N, A)$ and the same proof as that of Proposition 1 yields a conjugacy $H = \mathrm{Id} + h$ such that

$$\begin{aligned} H \circ f &= (\tilde{f}_0 + \widetilde{\Delta f}) \circ H \\ H \circ g &= (\tilde{g}_0 + \widetilde{\Delta g}) \circ H. \end{aligned} \tag{6.1}$$

Here $\tilde{f}_0 = A \times R_{\tilde{\alpha}}, \tilde{g}_0 = B \times R_{\tilde{\beta}}$, and $h, \widetilde{\Delta f}, \widetilde{\Delta g}$ satisfy estimates as in Proposition 1. Now, the fact that $(\rho_{\mu_1}(f), \rho_{\mu_2}(g)) = (\alpha, \beta)$ implies that $(\rho_{H_*\mu_1}(H \circ f \circ H^{-1}), \rho_{H_*\mu_2}(H \circ g \circ H^{-1})) = (\alpha, \beta)$. In conclusion we can replace $\tilde{f}_0, \tilde{g}_0$ by $A \times R_\alpha, B \times R_\beta$ in (6.1) and include $\tilde{\alpha} - \alpha$,$\tilde{\beta} - \beta$ inside the error terms, without changing the quadratic nature of the estimates.

For the general case $(\alpha, \beta) \in \mathrm{SDC}(\tau, \gamma, A, B)$ one cannot use just one of the frequencies α or β to solve the linearized equations of (3.4). Indeed, both α and β may be Liouville vectors and the small divisors that appear in (3.16) may be too large. Actually the linearized system (3.4) will not be solved as in Claim 1 but just up to an error term that is quadratic as in Lemma 1. The idea goes back to Moser [16] who observed the following: if for each m one of the small divisors $\lambda - e^{2\pi i m\alpha}$ or $\mu - e^{2\pi i m\beta}$ is not too small, as stated in the SDC condition, then the relation implied by the commutation (3.7)

$$(\lambda w - w \circ f_\varphi) - (\mu v - v \circ g_\psi) = \phi$$

insures that (3.4) can be solved up to an error term of the order of ϕ, that is a quadratic error term as in (3.8).

The rest of the proof of Theorem 1 is identical to that of Theorem 3. □

7. Appendix

In this Appendix, we give references and proofs for the estimates used in the proofs of Lemma 1 and Proposition 1.

7.1. **Convexity estimates.**

Proposition 2. *Let $f, g \in C^{lip,\infty}(I, \mathbb{T}^d, \mathbb{R})$. Then*

(i)

$$||f||_{lip(I),s} \leq C_{s_1,s_2} ||f||^{a_1}_{lip(I),s_1} ||f||^{a_2}_{lip(I),s_2}$$

for all non-negative numbers a_1, a_2, s_1, s_2 such that

$$a_1 + a_2 = 1, \quad s_1 a_1 + s_2 a_2 = s.$$

(ii)

$$||fg||_{lip(I),s} \leq C_s(||f||_{lip(I),s}||g||_{lip(I),0} + ||f||_{lip(I),0}||g||_{lip(I),s})$$

for all non-negative numbers s.

Proof. (i) One way to show interpolation estimates in the scale of $C^{lip,s}$ norms is to derive them from the existence of smoothing operators and from the norm inequalities for the smoothing operators. This is done in [24] for spaces $C^{\alpha,s}$ where $0 < \alpha \leq 1$, which includes the case of $C^{lip,s}$. Another elementary proof for interpolation without going through smoothing operators can be found in [14].

(ii) Immediate corollary of the interpolation estimates is the following fact:

$$\|f\|_{lip(I),i}\|g\|_{lip(I),j} \leq C(\|f\|_{lip(I),k}\|g\|_{lip(I),l} + \|f\|_{lip(I),m}\|g\|_{lip(I),n})$$

if (i,j) lies on the line segment joining (k,l) and (m,n). (See Corollary 2.2.2. in [7]). The statement (ii) in the Proposition follows from this by using the product rule on derivatives (see Corollary 2.2.3. in [7]) and the following inequality:

$$\begin{aligned} Lip(fg) &= \sup_{x\neq y} \frac{|(fg)(x) - (fg)(y)|}{|x-y|} \\ &\leq \sup\left(\frac{|f(x) - f(y)||g(x)|}{|x-y|} + \frac{|g(x) - g(y)||f(y)|}{|x-y|}\right) \\ &\leq L_f\|g\|_0 + \|f\|_0 L_g \end{aligned}$$

where L_f and L_g are Lipshitz constants for f and g, respectively. □

7.2. **Composition.**

PROPOSITION 3. *Let $f, g \in C^{lip,\infty}(I, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$.Then*

(i) $h(x) = f(x + g(x)) - f(x)$ verifies

$$\|h\|_{lip(I),s} \leq C_s(\|f\|_{lip(I),0}\|g\|_{lip(I),s+1} + \|f\|_{lip(I),s+1}\|g\|_{lip(I),0}).$$

(ii) $k(x) = f(x + g(x)) - f(x) - Dfg(x)$ verifies

$$\|k\|_s \leq C_s(\|f\|_{lip(I),0}\|g\|_{lip(I),s+2} + \|f\|_{lip(I),s+2}\|g\|_{lip(I),0})$$

Proof. In the proof we shorten the notation $\|\cdot\|_{lip(I),s}$ to $\|\cdot\|_{lip,s}$.

(i) It suffices to prove the estimates for the coordinate functions of f, so in what follows we assume that f denotes a coordinate function of f. Let D_i^1 denote partial derivation in one of the basis directions and let g_j denote coordinate functions of g. Since $D_i^1 h = D_i^1(f(x + g(x)) - f(x)) = \sum_j D_j^1 f D_i^1 g_j$, we can apply part (ii) of the previous proposition to $D_j^1 f D_i^1 g_j$:

$$\begin{aligned}
\|D^1 h\|_{lip,s} &\leq C \max_j \|D_j^1 f D_i^1 g_j\|_{lip,s} \\
&\leq C_s \max_j (\|D_j^1 f\|_{lip,s} \|D_i^1 g_j\|_{lip,0} + \|D_j^1 f\|_{lip,0} \|D_i^1 g_j\|_{lip,s}) \\
&\leq C_s \max_j (\|f\|_{lip,s+1} \|g_j\|_{lip,1} + \|f\|_{lip,1} \|g_j\|_{lip,s+1}) \\
&\leq C_s'(\|f\|_{lip,s+2} \|g\|_{lip,0} + \|f\|_{lip,0} \|g\|_{lip,s+2}).
\end{aligned}$$

Here we invoked part (ii) of the previous proposition again to obtain the last line of estimates above. Since for the $lip, 0$-norm we have:

$$\|h\|_{lip,0} = \|f(x + g(x)) - f(x)\|_{lip,0} \leq L_f \|g\|_0 \leq \|f\|_{lip,0} \|g\|_{lip,0},$$

the claim follows.

(ii) Again by reducing to coordinate functions we look at one coordinate function of k and f (which we denote by k and f as well), so we have $k = f(x + g(x)) - f - \sum_i D_i^1 f g_i$, where D_i^1 denotes $\partial / \partial x^i$. Then: $D_j^1 k = -\sum_i D_j^1 D_i^1 f g_i$, where g_i denotes coordinate functions of g. This implies (by using (ii) of Proposition 2) the following estimate for the first derivatives:

$$\begin{aligned}
\|D_j^1 k\|_{lip,s} &\leq \sum_i \|D_j^1 D_i^1 f g_i\|_{lip,s} \\
&\leq C_s(\|D_j^1 D_i^1 f\|_{lip,s} \|g_i\|_{lip,0} + \|D_j^1 D_i^1 f\|_{lip,0} \|g\|_{lip,s}) \\
&\leq C_s(\|f\|_{lip,s+2} \|g_i\|_{lip,0} + \|f\|_{lip,2} \|g\|_{lip,s}) \\
&\leq C_s'(\|f\|_{lip,s+2} \|g\|_{lip,0} + \|f\|_{lip,0} \|g\|_{lip,s+2}).
\end{aligned}$$

For the $lip, 0$-norm we have:

$$\|k\|_{lip,0} \leq L_f \|g\|_0 + \max_i \{\|D_i^1 f g_i\|_{lip,0}\} \leq C \|f\|_{lip,1} \|g\|_{lip,0}$$

which together with the estimates above implies the claim. □

7.3. **Inversion.**

PROPOSITION 4. *Let $h \in C^{lip,\infty}(I, \mathbb{T}^{d+1}, \mathbb{R}^{d+1})$ and assume that*

$$\|h\|_{lip(I),1} \leq \frac{1}{2}.$$

Then

$$f : \mathbb{T}^{d+1} \to \mathbb{T}^{d+1}, x \mapsto H(x) = x + h(x)$$

is invertible and if we write $H^{-1}(x) = x + \bar{h}(x)$ *then*

$$\|\bar{h}\|_{lip(I),s} \le C_s \|h\|_{lip(I),s}$$

for all $s \in \mathbb{N}$.

Proof. For C^s norms this is proved for example in Lemma 2.3.6. in [7]. The proof uses induction and interpolation estimates, and it is general to the extent that it applies to any sequence of norms on C^∞ which satisfy interpolation estimates. Thus the claim follows from part (i) of the Proposition 2 and Lemma 2.3.6. in [7]. □

Acknowledgments. We are grateful to the referee for his suggestions that helped us make several improvements to the initial version of this paper.

References

[1] V. Arnol'd, *Small denominators. I. Mapping the circle onto itself*, Izv. Akad. Nauk SSSR Ser. Mat., 1961, Volume 25, Issue 1, Pages 2186

[2] D. Damjanovic, Actions with globally hypoelliptic leafwise Laplacian and rigidity, accepted for publication in Journal d'Analyse Mathematique, 2016.

[3] D. Damjanović, A. Katok, *Local Rigidity of Partially Hyperbolic Actions. I. KAM method and Z^k actions on the Torus*, Annals of Mathematics, Vol. 172 (2010) no 3, 1805–1858.

[4] B. Fayad, K. Khanin, *Smooth linearization of commuting circle dieomorphisms*, Annals of Mathematics 170 (2009), 961–980.

[5] D. Fisher, G. Margulis, *Local rigidity of affine actions of higher rank groups and lattices.* Ann. of Math. (2) 170 (2009), no. 1, 67122.

[6] D. Fisher, *Local rigidity of group actions: past, present, future,* in Dynamics, Ergodic Theory, and Geometry, Math. Sci. Res. Inst. Publ. 54, Cambridge Univ. Press, Cambridge, 2007, pp. 4597.

[7] R. Hamilton, *The inverse function theorem of Nash and Moser*. Bull. Amer. Math. Soc. (N.S.) 7 (1982), no. 1, 65222.

[8] S. Hurder, *Rigidity for Anosov actions of higher rank lattices*, Ann. Math. 135 (1992) p. 361–410.

[9] S. Hurder, Michigan Math. J. 40 (1993), no. 3, 561–575.

[10] A. Katok, B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, 1996.

[11] A. Katok, E. A. Robinson, Jr., Cocycles, cohomology and combinatorial constructions in ergodic theory, Smooth Ergodic Theory and its applications, Proc. Symp. Pure Math., 69 (2001), 107-173;

[12] A. Katok, R. Spatzier, *Differential rigidity of Anosov actions of higher rank Abelian groups and algebraic lattice actions*, Dynamical systems and related topics, (Volume dedicated to D.V. Anosov), Proc. Steklov Math. Inst , 216 (1997), 287-314

[13] R. de la Llave, Tutorial on KAM theory, Smooth Ergodic Theory and its Applications (Seattle, WA, 1999), Proc. Sympos. Pure Math., Vol 69, Amer. Math. Soc., Providence, RI (2001), 175-292.

[14] R. de la Llave, R. Obaya, *Regularity of the composition operator in spaces of Hölder functions*. Discrete Contin. Dynam. Systems 5 (1999), no. 1, 157184.

[15] J. Moser, A rapidly convergent iteration method and non-linear partial differential equations. I. Ann. Scuola Norm. Sup. Pisa (3) 20 1966 265315.

[16] J. Moser, *On commuting circle mappings and simoultaneous Diophantine approximations Math. Z.* **205**(1990) , 105-121.

[17] V. Nitica, A. Török, *Cohomology of dynamical systems and rigidity of partially hyperbolic actions of higher rank lattices*, Duke Math. J., 79 no.3 (1995), 751-810.

[18] V.Nitica, A.Török, *Local rigidity of certain partially hyperbolic actions of product type*. Ergodic Theory Dynam. Systems 21 (2001), no. 4, 1213 1237.

[19] A. S. Pyartli, *Diophantine approximations on submanifolds of euclidean space*, Functional Analysis and Its Applications 3 (1969), no. 4, 303306.

[20] F. Rodriguez-Hertz, Z. Wang, *Global rigidity of higher rank abelian Anosov algebraic actions. Inventiones mathematicae*, 2014, Volume 198, Issue 1, pp 165-209

[21] A.Török, *Rigidity of partially hyperbolic actions of property (T) groups*. Discrete Contin. Dyn. Syst. 9 (2003), no. 1, 193208.

[22] A. Starkov *The first cohomology group, mixing, and minimal sets of the commutative group of algebraic actions on a torus*. Dynamical systems. 7. J. Math. Sci. (New York), 95 (1999), no. 5, 2576–2582.

[23] J.-C. Yoccoz, *Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition Diophantienne*, Ann. Sci. Ec. Norm. Sup. 17 (1984), 333–361.

[24] E. Zehnder, *Generalized implicit function theorems with applications to some small divisor problems*. I. *Comm. Pure Appl. Math.* **28**, (1975)